



UNIVERSITÀ
DI PAVIA

UNIVERSITY OF PAVIA

FACULTY OF ENGINEERING

DEPARTMENT OF ELECTRICAL, COMPUTER AND BIOMEDICAL
ENGINEERING

MASTER'S DEGREE IN COMPUTER ENGINEERING

MASTER THESIS

**AI-Based Generation of Tactile Bas-Reliefs from
Pictorial Art**

**Generazione di bassorilievi tattili da opere pittoriche
mediante tecniche di intelligenza artificiale**

Candidate

Reza Ghorbani

Supervisor:

Prof. Marco Porta

Co-supervisor:

Prof. Wolfgang Nejdl

A.Y. 2025/2026

Abstract

Generating a tactile bas-relief from a pictorial artwork is not simply a problem of recovering three-dimensional geometry from a two-dimensional image. The resulting surface must preserve the composition and viewpoint of the original artwork while compressing the inferred geometry into a shallow physical depth range. At the same time, locally important structures must remain geometrically distinguishable without introducing unsupported shape, discontinuities, or distortions. These requirements are particularly challenging for paintings, where shading, texture, contrast, and artistic stylization do not necessarily correspond directly to physical surface geometry.

This thesis investigates a controlled, baseline-preserving framework for image-driven bas-relief generation and local geometric refinement. A stable image-aligned representation is first established through monocular depth estimation and relief-oriented depth remapping. For the selected case study, Depth Anything 3 is retained as the global depth source after comparison with alternative monocular depth approaches. Starting from this fixed baseline, increasingly structured sources of local geometric information are investigated, including external three-dimensional reconstruction, semantic and face-aware depth processing, parametric anatomical priors, independently estimated surface normals, confidence weighting, anatomical landmarks, curvature cues, and explicit mesh-space processing. Representative configurations are reconstructed under fixed Blender displacement conditions to allow controlled comparison of the resulting shallow surfaces.

The experiments show that increasing the amount or complexity of geometric information does not automatically produce a more effective bas-relief. External three-dimensional geometry introduced registration, projection, and compatibility constraints. Semantic processing provided direct spatial control while preserving the baseline, but did not independently specify internal anatomical geometry. Parametric and landmark-guided approaches enabled more explicit control of facial structures, whereas independently estimated normals provided complementary orientation and confidence information. Nevertheless, increasingly strong or complex local modifications frequently produced diminishing changes in the final reconstructed surface.

A central result of the investigation is that numerical modification in the depth domain is not directly proportional to the resulting modification of the shallow relief. The spatial scale, support, transition behaviour, and subsequent depth compression of a correction strongly influence whether it remains significant after displacement reconstruction.

Overall, the experimental evidence supports a hierarchical strategy in which a stable image-derived surface is retained as the global geometric reference and additional information is introduced through localized, bounded, and smoothly weighted corrections. The contribution of the thesis is therefore a systematic analysis of the geometric representations, constraints, failure modes, and trade-offs involved in controlled local refinement of image-driven bas-reliefs.

The study focuses primarily on a single painting and mainly on facial regions. Evaluation is computational and based on controlled digital reconstruction; no formal tactile user study is conducted. The results therefore concern geometric behaviour and refinement rather than experimentally validated improvements in tactile perception.

Contents

1	Introduction	9
1.1	Background and Context	9
1.2	Research Motivation	11
1.3	Research Questions	12
1.4	Objectives of Research	12
1.5	Scope and Limitations	13
1.6	Methodological Overview	14
1.7	Case Study	14
1.8	Evaluation Strategy	16
1.9	Contributions of the Thesis	17
1.10	Thesis Organization	18
1.11	Summary of Chapter	19
2	General Techniques for 2D-to-3D Conversion	20
2.1	Introduction	20
2.2	The 2D-to-3D Reconstruction Problem	21
2.3	Depth Maps and Height-Field Representations	23
2.4	Classical Geometric Recovery	24
2.4.1	Stereo Vision	24
2.4.2	Structure from Motion	24
2.4.3	Shape from Shading	24
2.5	Monocular Depth Estimation Based on Learning	25
2.5.1	Relative Depth	25
2.5.2	Metric Depth	25
2.5.3	Global and Local Depth Information	26
2.6	Modern Foundation Models for Depth and Geometry	26
2.7	Depth Estimation by Diffusion	27
2.8	Preserving Fine Structure for Depth Estimation	28
2.9	Geometric Interpretation of Surface Normals	28
2.10	Single-Image 3D Reconstruction	29
2.11	Point-Map Feed-Forward and Multi-View Geometry	30
2.12	Semantic Information in Geometric Estimation	30
2.13	Depth to 3D Surface	31
2.14	Bas-Relief Representation	32
2.15	Classical Relief Processing	32

2.15.1	Processing in Gradient Domain	32
2.15.2	Adaptive and Nonlinear Compression	33
2.15.3	Multi-Scale Processing	33
2.15.4	Detail-Preserving Processing	33
2.15.5	Normal-Based Relief Processing	34
2.16	Local and Global Geometric Information	34
2.17	Depth Compression and Nonlinear Remapping	35
2.18	Boundary Preservation and Surface Regularization	35
2.19	Geometric Priors and Anatomical Constraints	36
2.20	From Visual Geometry to Tactile-Oriented Geometry	36
2.21	Evaluation of Depth-Driven Relief Geometry	37
2.21.1	Depth-Domain Evaluation	37
2.21.2	Gradient and Structural Evaluation	37
2.21.3	Final Surface Evaluation	38
2.21.4	Physical and Tactile Evaluation	38
2.22	Comparison of the Main Technique Families	38
2.23	Implications for Image-Driven Bas-Relief Generation	40
2.24	Chapter Summary	40
3	Related Work	42
3.1	Introduction	42
3.2	Classical Bas-Relief Generation	43
3.2.1	Adaptive Depth Compression	43
3.3	Detail-Preserving Bas-Relief Methods	43
3.4	Normal-based Relief Synthesis	44
3.5	Tactile Reproduction of Pictorial Art	44
3.6	Reconstruction of Portraits and Human Bas-Reliefs	45
3.7	General-Purpose Monocular Depth Estimation	46
3.7.1	Metric3D v2	46
3.7.2	Depth Anything 3	47
3.7.3	Marigold	47
3.8	Fine-structure and Boundary-aware Depth Estimation	47
3.9	Depth and Surface-Normal Consistency	48
3.10	Feed-Forward 3D Reconstruction	48
3.11	Semantic and Region-Aware Geometric Estimation	49
3.12	External Geometry and Local Fusion	49
3.13	Comparison of Representative Approaches	50
3.14	Synthesis of the Related Work	52
3.15	Research Gap	52
3.16	Positioning of the Thesis	53
3.17	Boundary of the Claimed Contribution	53
3.18	Chapter Summary	54

4	Establishment of the Depth-Driven Baseline	55
4.1	Introduction	55
4.2	Baseline Requirements	56
4.2.1	Global Geometric Stability	56
4.2.2	Image Alignment	56
4.2.3	Local Structural Coherence	57
4.2.4	Compatibility with a Height Field	57
4.2.5	Post-Processing Stability	57
4.3	Input Artwork	57
4.4	Candidate Monocular Depth Models	58
4.4.1	Depth Anything 3	58
4.4.2	Marigold	59
4.4.3	Metric3D v2	59
4.5	Qualitative Comparison of Candidate Depth Representations	60
4.5.1	Global Scene Structure	60
4.5.2	Local Facial Structure	61
4.6	Application-Specific Model Comparison	61
4.7	Selection of Depth Anything 3 as the Global Depth Source	61
4.8	DA3 Configuration Study	63
4.9	Inference Resolution Analysis	63
4.10	Pre-Alignment and Custom PPD-Inspired Depth Refinement	64
4.11	Depth Normalization	64
4.12	Relief-Oriented Depth Remapping	65
4.13	Raw DA3 and DA3+Remap	65
4.14	High-Precision Depth Representation	66
4.15	Controlled Blender Reconstruction	66
4.16	Blender Validation of the Remapped Baseline	67
4.17	Local Analysis of the Baseline	67
4.17.1	David’s Face	68
4.17.2	Goliath’s Head	68
4.18	Baseline Strengths	69
4.19	Baseline Constraints and Implications	69
4.20	Definition of the Fixed Experimental Baseline	69
4.21	Chapter Summary	70
5	Hybrid Geometry Fusion and External 3D Priors	72
5.1	Introduction	72
5.2	Objective and Experimental Hypothesis	73
5.3	Fixed Inputs and External Geometry	73
5.4	Why Direct Geometry Fusion Is Difficult	74
5.4.1	Scale Mismatch	74
5.4.2	Translation Mismatch	75
5.4.3	Region-of-Interest Mismatch	75
5.4.4	Projection Mismatch	75
5.5	Preliminary Hybrid Exploration: Exp28–Exp31	75

5.5.1	Exp28: Baseline Reference	75
5.5.2	Exp29: Initial DA3 + Pixal3D Fusion	76
5.5.3	Exp30: Alignment-Aware Fusion	76
5.5.4	Exp31: Semantic Mask Refinement	76
5.6	Registration Refinement: Exp32–Exp35	76
5.6.1	Exp32: Bounding-Box Alignment	76
5.6.2	Exp33: Centroid Alignment	77
5.6.3	Exp34: Scale Normalization	77
5.6.4	Exp35: Manual Refinement	78
5.7	From Registration to Geometry Fusion	78
5.8	External Geometry Representation and EXR Processing	78
5.9	Fusion Strategies	78
5.9.1	Direct Local Injection	79
5.9.2	Residual-Based Injection	79
5.9.3	Curvature-Oriented Cues	79
5.9.4	Detail-Layer Transfer	79
5.9.5	Local Volumetric Inflation	80
5.10	Representative Fusion Experiments: Exp41–Exp47	80
5.10.1	Exp41: Clean EXR Fusion	80
5.10.2	Exp42: Residual Injection	80
5.10.3	Exp43: Curvature Cue Injection	80
5.10.4	Exp44: Detail-Layer Injection	80
5.10.5	Exp45: Local Volume Injection	80
5.10.6	Exp46: Large-Scale Volume Inflation	80
5.10.7	Exp47: Aggressive Fusion	81
5.11	Summary of the Hybrid Experiments	81
5.12	Representative Blender Comparison	82
5.13	What Improved	83
5.14	What Remained Unresolved	83
5.15	Methodological Interpretation	84
5.16	Chapter Summary	84
6	Semantic and Face-Aware Local Depth Refinement	85
6.1	Introduction	85
6.2	Experimental Objective	86
6.3	Semantic Supports	86
6.4	Early Semantic Enhancement	87
6.4.1	Exp48: Scene-Aware Foreground–Background Enhancement	87
6.4.2	Exp57: Mask-Guided Local Depth Enhancement	87
6.4.3	Exp58: Face-Specific Nonlinear Remapping	87
6.4.4	Exp69: Semantic Local Contrast Enhancement	88
6.5	Local Face-Aware Enhancement: Exp71–Exp73	88
6.5.1	Local Inflation	88
6.5.2	Gaussian Enhancement	88
6.5.3	Image-Guided Enhancement	88

6.6	Semantic Depth Remapping: Exp74–Exp80	88
6.6.1	Exp74A: Initial Semantic Depth Remapping	89
6.6.2	Exp75A–B: Semantic Region Depth Redistribution	89
6.6.3	Exp76A–B: Foreground-Aware Compression	89
6.6.4	Exp77A: Semantic-Aware Transition Smoothing	89
6.6.5	Exp78A–B: Boundary-Safe Semantic Enhancement	89
6.6.6	Exp79A–B: Semantic Interior Enhancement	90
6.6.7	Exp80A: Safe Interior Enhancement	90
6.7	Further Refinement: Exp81–Exp84	91
6.8	Z-Aware Semantic Compression: Exp85	91
6.8.1	Exp85A: Semantic-Aware Nonlinear Compression	91
6.8.2	Exp85B: Further Compression Variant	91
6.9	Camera-Relative Local Reconstruction: Exp90	92
6.10	Representative Experiment Overview	92
6.11	Comparison with External Geometry Fusion	94
6.12	Synthesis and Remaining Limitation	94
6.13	Chapter Summary	95
7	Parametric and Anatomy-Guided Facial Geometry Refinement	96
7.1	Introduction	96
7.2	Objectives and Design Constraints	97
7.3	Working Inputs and Geometric Supports	97
7.4	From Semantic Remapping to Structural Cues	98
7.4.1	Additional Global and Semantic Compression: Exp91–92	98
7.5	Feature-Line Extraction and Structural Localization	98
7.5.1	Automatic Feature-Line Extraction: Exp93	98
7.5.2	Cleaned Feature-Line Maps: Exp93A	99
7.6	Feature-Guided and Multi-Scale Local Processing	99
7.6.1	Feature-Guided Depth Gradients: Exp94	99
7.6.2	Multi-Source Semantic Detail Cue: Exp95	100
7.6.3	Multi-Scale Depth Re-Injection: Exp96	100
7.6.4	Image- and Curvature-Guided Micro-Relief: Exp97–98	101
7.7	Macro-Scale Facial Support	101
7.8	Landmark Localization	101
7.8.1	Automatic Landmark Strategies: Exp100A–B	101
7.8.2	Semi-Manual Anatomical Landmarks: Exp100C	101
7.9	Parametric Facial Priors	102
7.9.1	Initial Amplitude Study: Exp101A	102
7.9.2	Controlled Amplitude Increase: Exp101B	103
7.9.3	Wider Nose Support: Exp101C	103
7.10	Low-Frequency Support and Surface Relaxation	103
7.10.1	Low-Frequency Face Envelope: Exp102	103
7.10.2	Constrained Surface Relaxation: Exp103A	103
7.10.3	Continuous Diffused Target: Exp103B	104
7.10.4	Continuous Midface Patch: Exp104A	104

7.11	Signed Anatomical Templates	105
7.11.1	Signed Midface Template: Exp105A	105
7.11.2	Template-Weight Increase: Exp106A	105
7.11.3	Reference-Amplitude Increase: Exp107A	105
7.12	Numerical Comparison	105
7.13	Representative Experimental Summary	107
7.14	Controlled Blender Assessment	109
7.15	Discussion	111
7.16	Chapter Summary	111
8	Geometry-Aware and Surface-Normal-Guided Facial Refinement	113
8.1	Introduction	113
8.2	Objectives and Fixed Experimental Conditions	114
8.3	Input Assets	114
8.4	Overall Geometry-Aware Workflow	114
8.5	Independent Surface-Normal Analysis	116
8.5.1	Independent DSINE Normal Estimation: Exp108A	116
8.5.2	Depth-Normal Consistency: Exp108B	117
8.5.3	Bias Compensation: Exp108B2	117
8.5.4	Confidence Estimation: Exp108C	118
8.6	Weak Normal-Guided Refinement	119
8.7	Landmark- and Curvature-Guided Anatomical Refinement	119
8.7.1	Adaptive Anatomical Supports: Exp112	119
8.7.2	Curvature-Guided Refinement: Exp113	119
8.7.3	Sparse Landmark-Centered Refinement: Exp114	120
8.7.4	Confidence-Relaxed Anatomical Refinement: Exp115	120
8.8	Mesh-Space Facial Refinement	123
8.8.1	Mesh Preparation: Exp116A–B and Exp117A	123
8.8.2	Direct Local Mesh Displacement: Exp118A–B	124
8.8.3	Anatomically Constrained Surface Relaxation: Exp119A	124
8.8.4	Parametric Mesh Template: Exp120A	124
8.9	Normal-Guided Surface Reconstruction	124
8.9.1	Confidence-Gated Surface Reconstruction: Exp121A	124
8.9.2	Depth-to-Mesh Scale Validation: Exp121B	124
8.10	Component Contribution and Ablation	125
8.10.1	Component Ablation: Exp122A	125
8.10.2	Strict Component Contribution: Exp122B	125
8.11	Representative Experimental Summary	126
8.12	Discussion	127
8.13	Chapter Summary	128
9	Overall Experimental Comparison and Discussion	129
9.1	Introduction	129
9.2	Evolution of the Research Strategy	129
9.3	Cross-Family Methodological Synthesis	130

9.4	Cross-Family Comparison	131
9.5	Representative Cross-Stage Comparison	133
9.6	Final Representative Visual Comparison	133
9.7	Stability versus Local Modification	135
9.8	Numerical Change versus Final Relief Change	136
9.9	Role of the Different Geometric Cues	136
9.10	Approaches with Limited Advantage under the Evaluated Conditions	136
9.11	Overall Interpretation	137
9.12	Remaining Limitations	137
9.13	Principal Findings Across the Complete Investigation	138
9.14	Chapter Summary	138
10	Conclusions and Future Work	140
10.1	Overall Conclusion	140
10.2	Representative Final Pipeline	140
10.2.1	Automatic and Manually Controlled Components	143
10.2.2	Rationale for the Representative Configuration	143
10.3	Answers to the Research Questions	144
10.3.1	Can a stable shallow bas-relief be generated from a single painting using image-derived depth?	144
10.3.2	Can external three-dimensional geometry improve locally weak facial regions?	144
10.3.3	Can semantic information improve local refinement while preserving global geometry?	144
10.3.4	Can anatomical priors provide more structured local control than semantic masks alone?	145
10.3.5	Do independently estimated normals provide useful complementary geometry?	145
10.3.6	Do landmarks and mesh-space processing provide additional advantages?	145
10.3.7	What general strategy was most consistent across the experiments?	145
10.4	Main Contributions	145
10.5	Limitations	146
10.6	Future Work	146
10.6.1	Broader Evaluation and Generalization	147
10.6.2	Automated and Relational Anatomical Refinement	147
10.6.3	Fabrication- and Tactile-Oriented Optimization	147
10.6.4	Improved Multi-Cue Geometry Fusion	147
10.6.5	Learning-Based Bounded Local Refinement	148
10.6.6	Physical Tactile Validation	148
10.7	Final Remarks	148
A	Reproducibility and Experimental Parameters	150
A.1	Common Input and Depth Representation	150
A.1.1	PPD-Inspired Multistage Depth Refinement	151
A.2	Depth Normalization and Relief-Oriented Remapping	152

A.3	Blender Reconstruction Configuration	155
A.3.1	Physical Interpretation of Normalized Depth Residuals	156
A.3.2	Relation to Fabrication Resolution and Tactile Perception	160
A.3.3	Cross-Stage Quantitative Comparison	161
A.4	Semantic Mask Generation	168
A.4.1	GroundingDINO–SAM Face and Head Supports	168
A.4.2	Foreground and Face Exclusion Supports in Exp80A	169
A.4.3	Final Polygon-Based Facial Supports	169
A.4.4	Foreground-Mask Standardization	170
A.5	Phase-2B Semantic Reinjection and Phase-2C Intermediate Facial Baseline . . .	170
A.5.1	Phase-2B Multiscale Semantic Reinjection	171
A.5.2	Phase-2C Localized Goliath Mouth Correction	172
A.6	Anatomical Landmark Definition	174
A.6.1	Semi-Manual Landmarks for Parametric Facial Refinement	174
A.6.2	MediaPipe Landmarks for Geometry-Aware Refinement	174
A.7	Representative Experimental Configurations	175
A.7.1	Exp46: Large-Scale External-Geometry Fusion	176
A.7.2	Exp80A: Safe Interior Body and Drapery Refinement	177
A.7.3	Exp101B: Parametric Facial Prior	177
A.7.4	Exp105A: Signed Anatomical Midface Template	178
A.7.5	Exp115 R1: Confidence-Relaxed Landmark-Guided Refinement	178
	Statement on the Use of AI Tools	181

Chapter 1

Introduction

1.1 Background and Context

The conversion of visual artworks into tactile representations provides an alternative means of accessing pictorial content when visual perception is limited or unavailable [4].

Tactile representations offer an alternative for blind and visually impaired users to access spatial information contained in visual material. One of the possible representations, the bas-relief, is especially relevant as it turns the visual content into a shallow physical surface, while maintaining a good part of the original 2D composition.

A bas-relief does not attempt to reproduce the full geometry of a scene as a full three-dimensional reconstruction would do. Rather, the depth is strongly compressed along the viewing direction, while the spatial layout of the source image is preserved.

This thesis was also developed in the context of an ongoing collaboration with Italian partners involved in the exploration of accessible representations of pictorial art. The original idea for investigating tactile bas-relief representations in this context emerged from previous work and discussions involving Marco Lorenzin (Entracte) and photographer Flavio Chiesa. Their experience provided an important application context for the computational investigation developed in this thesis.

A computational bas-relief can therefore be approximated as a height field

$$H(x, y), \tag{1.1}$$

where the image coordinates (x, y) determine position on the relief plane and H determines the local surface displacement.

This representation is especially useful for image-driven relief generation, as it preserves a direct correspondence between the source image and the generated surface.

However, producing such a representation directly from a painting is difficult.

The input does not contain any explicit 3D geometry, and the structure of the scene cannot be uniquely determined from a single image. Pictorial artworks may also contain non-photographic shading, stylized proportions, strong contrast, texture, and deliberate departures from physical image formation.

Recent advances in monocular depth estimation and general-purpose image-based geometry provide increasingly powerful tools for estimating spatial structure from visual input [18, 16, 13].

In particular, modern monocular depth models can produce dense image-aligned depth rep-

representations from a single RGB image without requiring multi-view capture or an existing three-dimensional model.

However, transforming an estimated depth map into a useful shallow relief is a separate problem.

A depth representation that is visually plausible in image space might include:

- large variation in global depth;
- weakly represented local structures;
- uncertain object edges;
- unstable geometry at fine scales;
- lack of separation between adjacent anatomical regions.

These limitations may be more pronounced when the entire depth range is compressed into a shallow physical surface.

The problem tackled in this thesis is therefore located at the interface of monocular geometry estimation and bas-relief processing.

The goal is not simply to recover as much three-dimensional geometry as possible. Rather, the aim is to generate a stable image-aligned shallow surface and to investigate whether locally important structures can be enhanced without destabilizing the global geometry of the relief.

The construction of a shallow bas-relief from a single painting is a sequence of coupled geometric problems.

The first problem is the lack of explicit three-dimensional information. Given an RGB image $I(x, y)$, the associated scene depth

$$D(x, y) \tag{1.2}$$

is unknown and must be estimated. Monocular depth estimation can provide such a representation, but the resulting depth is not specifically designed or optimized for the geometric constraints of shallow-relief generation.

The second problem is the compression of depth.

The estimated scene geometry may span a depth range much larger than the range available in a bas-relief.

A transformation

$$D'(x, y) = T(D(x, y)) \tag{1.3}$$

is therefore needed to redistribute the estimated depth in a limited physical interval.

A simple global scaling would preserve large relationships in the scene, but shrink small local variations to near zero values.

The third problem is the local geometric structure.

Facial features such as the nose, lips, eye sockets, cheeks, and chin may occupy only a small fraction of the total depth range. Their relative geometry can therefore remain weak after global compression.

The fourth problem concerns geometric stability.

Too much local enhancement, or insufficient spatial control, can introduce new artifacts. Potential failure modes are:

- artificial mounds;
- mask-shaped projections;
- boundary escarpments;
- haloes;
- artificial spikes;
- disproportion in the face;
- mismatch with the surrounding geometry.

The central problem of the thesis can therefore be formulated as follows:

How can one produce a shallow bas-relief from a stable image-based global surface with controlled refinement of locally relevant structures, without generating unsupported geometry and destabilizing the original composition?

This formulation deliberately factors two requirements:

$$\text{global geometric stability} \tag{1.4}$$

and

$$\text{local geometrical refinement.} \tag{1.5}$$

The experimental work considers how these requirements can be balanced.

1.2 Research Motivation

The motivation for this work stems from the challenge of translating complex pictorial information into a shallow geometric representation.

A further motivation is the broader objective of offering a more complete multisensory experience when engaging with pictorial and audiovisual works of art. Introducing a tactile component can complement visual and auditory access for a wide range of audiences, while being particularly relevant for people with low vision or blindness. In this context, tactile bas-reliefs provide one possible means of translating spatial and structural information from visual artworks into a physical form that can be explored through touch.

A painting can contain many depth cues that are readily perceived visually but are not directly represented as physical geometry. Shading can imply volume where little physical depth variation exists, colour contrast can separate neighbouring regions even when their surfaces lie at similar depths, and texture or brushwork can create strong visual structure without corresponding three-dimensional variation.

When the artwork is translated into a physical bas-relief, these purely visual cues are not necessarily represented by corresponding variations in surface geometry.

This motivates the investigation of the geometric arrangement of the generated relief, rather than considering the monocular depth prediction as the last output of the pipeline.

At the same time, arbitrary geometric enhancement is not desirable.

The relief created should be related to the original artwork and preserve:

- the original composition;
- the relative position of subjects;
- the perspective of the source;
- the general foreground–background organization;
- the identity of the major structures depicted.

This motivates a conservative approach, where one first establishes a stable global geometry and then introduces modifications locally and under explicit constraints.

Hence, the thesis investigates not only whether additional geometric information can be introduced, but also when such information becomes incompatible with the baseline.

1.3 Research Questions

The thesis is framed around the following research questions:

1. Can a stable shallow bas-relief be generated from a single painting with image-derived depth, while preserving the global composition and viewpoint of the source?
2. Can externally reconstructed three-dimensional geometry provide useful local refinement without introducing registration, projection, or geometric-compatibility problems?
3. Can semantic information spatially constrain local depth modifications while preserving the global baseline geometry?
4. Can parametric and anatomical priors provide more structured local facial control than semantic masks alone?
5. Can independently estimated surface normals be used to provide complementary geometric information to improve local refinement without destabilizing the baseline?
6. How do landmark-guided and mesh-space methods affect localized facial refinement in the fixed shallow-relief representation?
7. Which general refinement strategy provides the most consistent tradeoff between local geometric modification and preservation of the global shallow-relief surface?

1.4 Objectives of Research

The thesis sets out the following objectives based on the research questions:

1. generate a stable depth representation from images for shallow-relief reconstruction;
2. determine the effect of depth normalization and nonlinear remapping on the final relief geometry;
3. evaluate whether external 3D reconstruction can provide useful local geometry while remaining compatible with the globally image-aligned surface;

4. investigate semantic masks as spatial constraints for local depth change;
5. develop increasingly localized facial-refinement strategies based on anatomical regions, landmarks and parametric priors;
6. evaluate surface normals estimated independently as an additional source of local geometric information;
7. study confidence, curvature and landmark information for steering the location of geometric refinement;
8. assess the impact and practical constraints of explicit mesh-space processing versus depth-domain modification;
9. compare the main experimental families with a common reconstruction configuration;
10. identify recurrent failure modes and practical limits of local geometric refinement for image-driven shallow relief.

1.5 Scope and Limitations

The limits of the thesis are deliberately narrow.

The detailed experimental investigation is focused mainly on a single painting, Caravaggio’s *David with the Head of Goliath*.

The same artwork is used in the experimental development to allow for controlled comparison between a large number of geometric refinement strategies. However, the observations obtained should not be interpreted as a demonstration of generalization across different paintings, artistic styles or image domains.

The experimental investigation is specifically concerned with the facial regions of David and Goliath since they contain compact and structurally complex geometry that is useful for evaluation of local refinement.

Other structures, such as hands, drapery, weapons, and larger areas of the scene, are given less attention.

Some later experiments use semi-manual anatomical landmarks.

These landmarks provide a controlled local spatial support for the selected artwork, but reduce the level of automation of the corresponding methods.

The evaluation is mainly computational.

Candidate depth maps are analyzed numerically and reconstructed in a controlled displacement pipeline in Blender.

No formal tactile user study with blind or visually impaired participants is conducted within the experimental work of this thesis. However, user studies involving real users have already been carried out in the broader context of the associated accessibility work, and further perceptual evaluations are planned once sufficiently mature relief configurations are available.

Accordingly, the present thesis does not claim experimentally demonstrated improvements in tactile recognition, tactile perception, or user performance. Within this work, tactile readability is treated as a design motivation and a geometric objective rather than as a directly validated perceptual outcome.

The thesis also does not claim:

- physical accuracy in reconstruction of the original painted scene;
- complete recovery of missing 3D facial anatomy;
- a universal optimal bas-relief representation;
- extrapolation of the experimental results to all pictorial artworks.

Therefore the results should be understood mainly as a controlled methodological investigation of image-driven shallow-relief geometry.

1.6 Methodological Overview

The experimental investigation applies a progressive baseline-preserving strategy.

The general methodological development is:

$$\begin{aligned}
 &\text{depth-model evaluation} \\
 &\rightarrow \text{initial depth preparation and remapping} \\
 &\rightarrow \text{external geometry integration} \\
 &\rightarrow \text{semantic local refinement} \\
 &\rightarrow \text{parametric anatomical refinement} \\
 &\rightarrow \text{normal-, landmark-, and mesh-guided refinement} \\
 &\rightarrow \text{cross-family comparison.}
 \end{aligned} \tag{1.6}$$

In the first step, a stable working global geometric representation is built from the source painting.

We compare alternative monocular depth approaches and select a stable working baseline, which we adapt to the shallow-relief representation by means of a controlled remapping.

In the second stage, we investigate whether richer external three-dimensional geometry can improve selected local regions.

In stage three, semantic localization is used to determine whether the local depth can be changed without affecting the surrounding baseline.

The fourth stage substitutes broad semantic modification with more explicit anatomical priors.

The fifth stage further introduces geometric representations, such as independently estimated surface normals, confidence maps, curvature, landmarks and explicit mesh-space processing.

Finally, the representative methods are compared within the complete experimental sequence.

Each stage is informed by the limitations identified in the preceding stage.

The goal of this progression is not therefore just to increase algorithmic complexity, but to determine which forms of additional geometric information are compatible with a stable shallow-relief surface.

1.7 Case Study

The experimental investigation is centered on Caravaggio’s *David with the Head of Goliath*, shown in Figure 4.1.

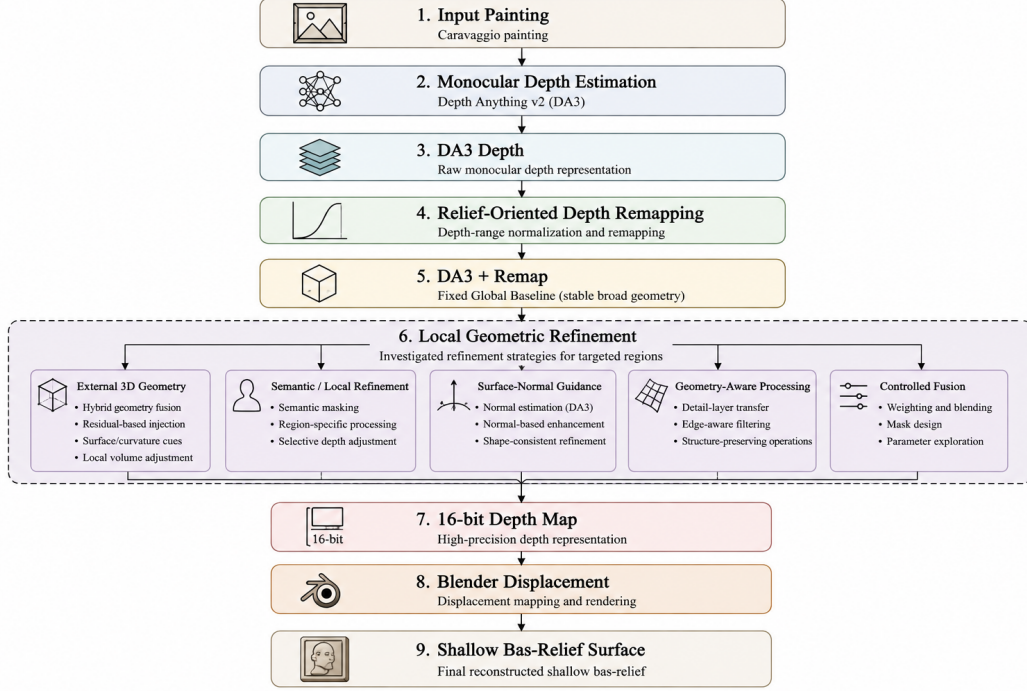


Figure 1.1: High-level overview of the image-driven shallow bas-relief pipeline investigated in this thesis. We first build a global image-derived depth representation, then study controlled local geometric refinement, and finally surface reconstruction.

This painting provides a useful controlled case study for image-driven bas-relief generation because it combines several geometric and appearance-related challenges within a single composition.

It includes:

- multiple human figures and body regions;
- two visually important faces at different scales;
- strong chiaroscuro;
- partially obscured anatomy;
- complex relationships between foreground and background;
- fine anatomical details and drapery structures;
- painterly rather than strictly photographic appearance.

The two principal facial regions provide complementary conditions for local geometric analysis.

David’s face is small relative to the entirety of the composition, and the nose, cheeks, eyes, lips, and chin are all gently blended.

Goliath’s head differs from David’s face in scale, orientation, shading, contrast, and local appearance.

These differences enable the same processing principles to be evaluated under two distinct local geometric conditions within the same globally fixed artwork.

Furthermore, the use of a single case study allows the experimental variables to remain controlled.

Individual geometric strategies can be interchanged while fixing the original image, working resolution, baseline depth representation, masks, reconstruction pipeline and evaluation configuration.

1.8 Evaluation Strategy

The experimental evaluation is carried out at two complementary levels.

We first numerically analyze candidate depth representations.

Depending on the experiment, analysis may include:

- depth range;
- percentile statistics;
- local residuals;
- region-specific depth differences;
- gradient behaviour;
- surface-normal consistency;
- confidence distributions;
- magnitude of geometric displacement.

These measurements quantify how strongly an operation modifies the underlying depth representation.

Numerical modification, however, is not the sole factor in determining the final shallow surface appearance.

Therefore, the second evaluation level reconstructs the corresponding depth maps according to a fixed Blender displacement configuration.

This stage enables the investigation of depth-domain modifications after conversion into the final shallow surface. This distinction is important in the investigated pipeline because measurable differences in the depth representation do not necessarily lead to equally noticeable differences after displacement reconstruction.

Representative comparisons thus retain the same reconstruction configuration wherever possible, including:

- mesh topology;
- displacement configuration;
- camera configuration;
- viewpoint;
- lighting;

- rendering conditions.

The evaluation is still geometric and computational.

Qualitative Blender comparisons are used to explore the behavior of the final surface, but are not interpreted as results of a perceptual user study.

1.9 Contributions of the Thesis

The contributions of this thesis are mainly methodological and experimental, and can be summarized in four major points.

1. A Controlled Image-to-Bas-Relief Experimental Framework

We have constructed a complete workflow for converting pictorial input to a depth representation and then to a shallow displacement-based bas-relief.

The framework comprises monocular depth estimation, depth normalization and remapping, high-fidelity depth export, and controlled Blender reconstruction.

This common framework enables the comparison of alternative geometric representations and refinement strategies, while fixing the various conditions for the final reconstruction.

2. A systematic study of local geometric refinement

Multiple sources of local geometric information were progressively tested including:

- external 3D reconstruction;
- semantic support;
- face-aware local modification;
- parametric face priors;
- signed anatomical templates;
- independently estimated surface normals;
- confidence maps;
- landmarks;
- curvatures;
- explicit mesh-space representations.

The experiments not only examine whether these representations change the depth map, but also whether the changes are still useful after shallow-surface reconstruction.

3. Understanding of practical tradeoffs and failure modes

The experimental progression reports a number of recurring limitations including:

- registration and projection mismatch in external-geometry fusion;
- boundary artifacts caused by strong local modification;
- limited survival of very small depth changes after displacement reconstruction;
- ambiguity between image appearance and physical geometry;

- instability of fully automatic feature localization in difficult image regions;
- diminishing returns from increasingly complex anatomical priors;
- marginal qualitative improvement despite increasing representation complexity.

These observations provide practical information on the conditions under which local geometric refinement becomes ineffective or destabilizing.

4. Controlled local refinement with a baseline-preserving formulation

The complete experimental study suggests a hierarchical strategy that preserves a stable image-derived global surface while adding further geometric information only through localized and constrained refinement.

In this formulation, external geometry, semantics, anatomical priors, normals, landmarks, and mesh-space operations are considered auxiliary sources of local guidance, not automatic replacements for the baseline geometry.

1.10 Thesis Organization

The remainder of this thesis is organized as follows.

Chapter 2 introduces the technical background required for the study, including 2D-to-3D conversion, monocular depth estimation, surface normals, geometric reconstruction, depth compression, and classical bas-relief processing.

Chapter 3 reviews the state of the art in image-driven bas-relief generation. It compares representative approaches, identifies the research gap, positions the thesis in relation to the existing literature and defines the boundary of the claimed contribution.

The working geometric baseline is described in Chapter 4. We explore alternative monocular depth representations and adapt the chosen depth representation by means of normalization, remapping, high precision export and controlled Blender reconstruction.

Chapter 5 investigates the utility of external three-dimensional reconstruction for local geometric refinement. Particular attention is given to registration, scale, projection, and compatibility between the locally reconstructed geometry and the global image-aligned baseline.

In Chapter 6 we study semantic-aware and face-aware local depth modification. This chapter studies the use of semantic localization to control the edits in selected regions, while preserving the surrounding global surface.

In Chapter 7 we gradually add more explicit anatomical control through feature localization, facial landmarks, parametric priors, and signed anatomical templates.

Chapter 8 discusses independent surface-normal information, depth-normal consistency, confidence-aware refinement, landmark and curvature guidance, explicit mesh-space processing, constrained surface reconstruction, and component-level analysis.

In Chapter 9 the representative experimental families in the common baseline-preserving framework are compared and the main methodological observations obtained during the investigation are synthesized.

Lastly, Chapter 10 answers the research questions, summarizes the main contributions and limitations of the study, and identifies future work.

1.11 Summary of Chapter

In this chapter we have posed the problem of generating shallow bas-relief geometry from a single pictorial image.

Unlike standard three-dimensional reconstruction, the problem is that the output must preserve the source composition and viewpoint while compressing the inferred geometry into a limited physical depth range.

Modern monocular depth estimation provides a convenient starting point; however, the resulting depth is not specifically designed for shallow-relief representation.

Global compression may degrade small local structures, and aggressive local enhancement may introduce unsupported geometry and surface artifacts.

The thesis is therefore based on a baseline-preserving experimental methodology.

First, a stable working global geometry is obtained from image-derived depth. Then more information is progressively introduced using external reconstruction, semantic localization, parametric anatomical priors, surface normals, confidence, landmarks, curvature, and mesh-space processing.

The investigation aims at identifying forms of local geometric guidance that are compatible with the global shallow-relief surface.

The contribution is primarily methodological and experimental.

The thesis does not claim physically exact reconstruction or perceptually validated tactile improvement. Rather, it provides a systematic framework for studying the geometric design, constraints, and trade-offs involved in translating pictorial content into a shallow relief.

Chapter 2 introduces the technical foundations required to understand the geometric representations and processing strategies applied throughout the experimental investigation.

Chapter 2

General Techniques for 2D-to-3D Conversion

2.1 Introduction

The recovery of 3D information from 2D visual input is a general research problem in computer vision, computer graphics, image processing, and geometric modelling.

The desired output depends on the application and may take several forms:

- a depth map;
- a surface-normal field;
- a point cloud;
- a polygonal mesh;
- an implicit three-dimensional representation;
- a shallow height field.

Such representations are similar but not interchangeable.

A method designed to reconstruct a complete 3D object addresses a different problem from one designed to estimate image-aligned depth. Both differ from bas-relief generation, where the geometry must remain highly compressed along the viewing direction.

This distinction is particularly important for the problem considered in this thesis. The input is a single image and the output is a shallow, viewpoint-preserving geometric representation.

Thus the overall transformation can be viewed as a sequence of related operations:

$$2D \text{ Image} \rightarrow \text{Geometric Estimation} \rightarrow \text{Geometric Representation} \rightarrow \text{Depth Adaptation} \rightarrow \text{Relief Surface.} \quad (2.1)$$

This chapter presents the main technical concepts needed to understand these stages.

This chapter does not aim to determine the most appropriate method for the proposed pipeline; instead, it provides the technical background needed to understand the geometric representations and processing operations considered in this thesis. Chapter 3 critically examines the research literature most directly related to the proposed problem, while the controlled experimental comparison and selection of the working depth baseline are presented in Chapter 4.

2.2 The 2D-to-3D Reconstruction Problem

A two-dimensional image is a projection of a three-dimensional scene onto an image plane.

According to the simple pinhole perspective camera model, a 3D point

$$\mathbf{P} = (X, Y, Z)^\top \tag{2.2}$$

is projected to image coordinates as

$$u = f_x \frac{X}{Z} + c_x, \quad v = f_y \frac{Y}{Z} + c_y, \tag{2.3}$$

where f_x and f_y are the focal-length parameters and (c_x, c_y) is the principal point.

The projection inevitably removes part of the original spatial information. In particular, different three-dimensional configurations may produce similar or even identical two-dimensional projections. Inferring scene geometry from a single image is therefore an under-constrained inverse problem.

An image contains many cues that can be supportive of geometric inference such as:

- viewpoint;
- relative object size;
- occlusion;
- shading;
- texture gradients;
- object contours;
- semantic information;
- learned priors for common objects and scenes.

However, none of these cues alone guarantees unique three-dimensional reconstruction.

This ambiguity is especially important for pictorial art.

A painting need not conform to the photometric assumptions of a physical camera. Shading may be deliberately manipulated, perspective may be stylized, texture may be represented through brushwork rather than physical material, and object proportions may be altered for artistic purposes.

The aim of image-based geometry estimation should thus be separated from the more ambitious aim of recovering a physically accurate scene.

For shallow bas-relief generation, a stable and image-aligned representation may be more useful than a complete but viewpoint-incompatible reconstruction.

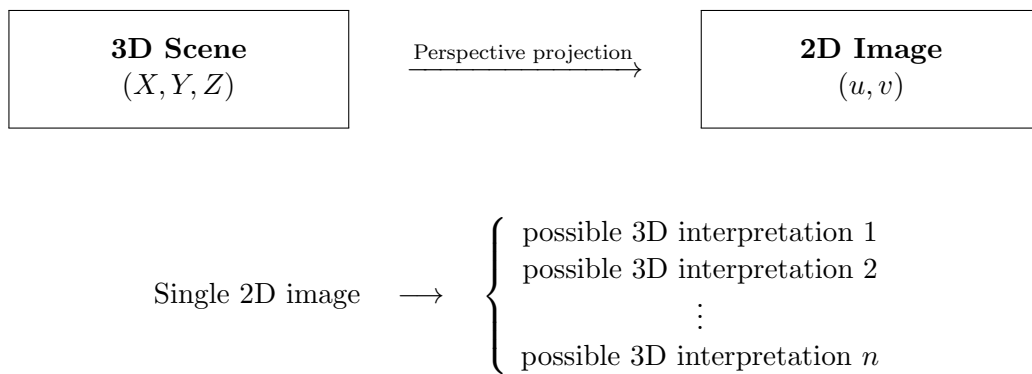


Figure 2.1: Conceptual illustration of perspective projection and the ambiguity of recovering three-dimensional structure from a single two-dimensional image. A three-dimensional scene can be projected onto a two-dimensional image, whereas a single image may be compatible with multiple three-dimensional interpretations.

2.3 Depth Maps and Height-Field Representations

A depth map assigns a scalar value to every pixel in an image:

$$D : \Omega \rightarrow \mathbb{R}. \quad (2.4)$$

where Ω is the image domain.

The value $D(x, y)$ encodes the depth, or relative depth ordering, of the visible surface at image location (x, y) . Depending on the estimation method, such a quantity may correspond to metric distance or a relative depth value defined up to an unknown scale or transformation.

Depth maps are especially convenient for image-driven bas-relief because each image point can be directly associated with a single displacement value on the relief surface.

For displacement-based bas-relief reconstruction, an image-aligned depth representation can be interpreted as a height field:

$$\mathbf{S}(x, y) = \begin{bmatrix} x \\ y \\ sD(x, y) \end{bmatrix}, \quad (2.5)$$

where (x, y) define coordinates on the relief plane and s controls the displacement scale applied to the depth representation.

This representation has several practical advantages:

- direct correspondence with the source image;
- straightforward conversion to mesh geometry;
- compatibility with semantic masking;
- efficient storage;
- compatibility with displacement-based reconstruction.

Its main limitation is that it is single-valued along the viewing direction.

A height field cannot directly represent:

- overhangs;
- hidden surfaces;
- self-intersections;
- rear-side geometry;
- multiple depth values at the same image location.

Although this is a restriction for full three-dimensional reconstruction, it can be advantageous for viewpoint-preserving bas-relief because the desired surface is deliberately shallow and image-aligned.

2.4 Classical Geometric Recovery

Before the rise of learning-based monocular models, the most common way to recover geometry was by explicit geometric or photometric constraints.

Many classical approaches are still helpful in understanding the assumptions behind modern systems.

2.4.1 Stereo Vision

Stereo Reconstruction estimates depth from two or more images taken from different view-points [23].

The relation between the depth and the disparity for a calibrated and rectified stereo pair is given by

$$Z = \frac{fB}{d}, \quad (2.6)$$

f is the focal length, B is the camera baseline and d is the disparity between corresponding image points.

With stereo methods, geometric depth can be obtained when good correspondences and camera calibration are available.

The principal limitation for the present application is input availability, since a historical painting provides only a single pictorial view.

2.4.2 Structure from Motion

Structure from Motion is a method to estimate camera motion and three dimensional structure from multiple overlapping views [24].

Features are matched across frames and jointly used to estimate camera poses and scene geometry.

The output of such a reconstruction is generally a set of estimated camera parameters and a sparse 3D structure of points. Further stages of multi-view reconstruction can then be used to obtain denser point clouds, depth maps or surface geometry.

The method is based essentially on multiple observations of the same scene, such as stereo vision.

It thus does not directly address the single-painting problem considered in this thesis.

2.4.3 Shape from Shading

Shape-from-shading techniques derive surface orientation or shape from image intensity variations [33].

Assuming a simplified Lambertian reflectance, image intensity can be written as a function of the relationship between the surface normal $\mathbf{n}$ and the illumination direction $\mathbf{l}$:

$$I(x, y) \propto \rho(x, y) \max(0, \mathbf{n}^\top \mathbf{l}), \quad (2.7)$$

where ρ is the surface reflectance.

The method shows the relationship between the appearance of the image and the orientation of the surface.

However, these assumptions are difficult to guarantee for paintings because artistic shading, pigment, texture and compositional choices can affect image intensity.

This limitation is important later if image-derived detail cues are used for local relief refinement.

2.5 Monocular Depth Estimation Based on Learning

Modern monocular depth estimation infers depth from a single RGB image using learned visual priors [5, 13, 18].

A generic representation for a model is

$$D = F_{\theta}(I), \quad (2.8)$$

where I is the input image, F_{θ} is a trained model and D is the predicted depth.

Unlike classical geometric reconstruction, monocular depth estimation does not require multiple views at inference time.

Prediction without explicit multi-view geometric constraints relies on visual regularities and geometric priors learned from training data.

2.5.1 Relative Depth

Relative-depth models predict the order and relative separation of visible surfaces, but not necessarily their physical distance from the camera.

For two image locations p_i and p_j the relevant information can be expressed as an ordering relation,

$$D(p_i) < D(p_j), \quad (2.9)$$

where the meaning of larger or smaller values is a function of the depth convention of the particular representation.

What matters is preserving the relative arrangement of the scene, not recovering the absolute distance in metric units.

Because the depth field is later compressed and rescaled into a controlled physical height interval, absolute metric distance is not always required for bas-relief generation. However, the relative distribution of depth remains important because it determines how the available geometric structure is distributed before relief compression.

2.5.2 Metric Depth

Metric-depth estimation recovers the depth of a scene in physically meaningful units.

Conceptually, the predicted value can be related to physical scene depth as

$$D(x, y) \approx Z_{\text{physical}}(x, y). \quad (2.10)$$

Metric prediction is especially relevant for applications where absolute scene scale is important such as robotics, measurement, navigation and augmented reality.

2.5.3 Global and Local Depth Information

For shallow-relief generation, it is useful to distinguish between global and local information in the predicted depth representation.

Global information describes:

- scene-level depth organization;
- foreground–background relationships;
- separation between large objects;
- broad surface orientation.

Local information describes:

- sharp boundaries;
- small depth variations;
- facial features;
- local curvature;
- details of thin objects.

This global–local distinction is fundamental for shallow bas-relief generation.

Large scene variations may occupy most of the available relief depth range, while important small structures may occupy only a small fraction of that range.

2.6 Modern Foundation Models for Depth and Geometry

Recently, large-scale learning-based models have extended the limits of image-based geometric estimation beyond narrowly trained, application-specific systems.

Depending on the model and the way the task is formulated, modern geometric estimation systems may predict different representations: relative or metric depth, surface orientation, camera information or more explicit 3D geometry.

Metric3D v2 and the Depth Anything family are pertinent examples for the later experimental investigation.

Metric3D v2 explores monocular metric depth estimation, surface normal prediction, and depth-normal consistency in a unified geometric framework [13].

The Depth Anything family targets the development of large-scale general-purpose monocular depth estimation, and later developments extend image-based prediction toward broader geometric representations [18].

These advances are pertinent to the current thesis because they offer image-derived geometric information without the need for an existing three-dimensional model of the depicted scene.

However, they do not directly address the shallow bas-relief generation problem considered in this work.

Their outputs must still be evaluated in terms of image alignment, depth compression, local structural behaviour, and compatibility with the final shallow-surface representation.

The suitability of specific depth representations for the chosen artwork is therefore assessed experimentally in Chapter 4, and not in this technical background chapter.

2.7 Depth Estimation by Diffusion

Diffusion-based approaches provide an alternative formulation for monocular depth estimation where the depth representation is generated through an iterative denoising process.

Marigold [16] is a representative example of this approach. It uses a pretrained latent diffusion model for monocular depth estimation and predicts depth through iterative denoising in the latent representation.

Diffusion-based estimation follows a different computational procedure from direct feed-forward depth prediction, as image-aligned depth is obtained through an iterative inference process.

The importance of such a method for shallow bas-relief generation is not only based on the visual plausibility of the predicted depth. The resulting representation also needs to be considered in terms of spatial organization, local geometric structure, depth compression and behavior after displacement-based surface reconstruction.

Therefore, diffusion-based depth estimation is treated as one candidate family rather than as an inherently superior representation for bas-relief generation.

Within the experimental pipeline considered in this thesis, the relevant sequence can be summarized as

$$\text{prediction} \rightarrow \text{normalization} \rightarrow \text{compression} \rightarrow \text{surface displacement}. \quad (2.11)$$

The behaviour of candidate depth representations within this pipeline is examined experimentally in Chapter 4.

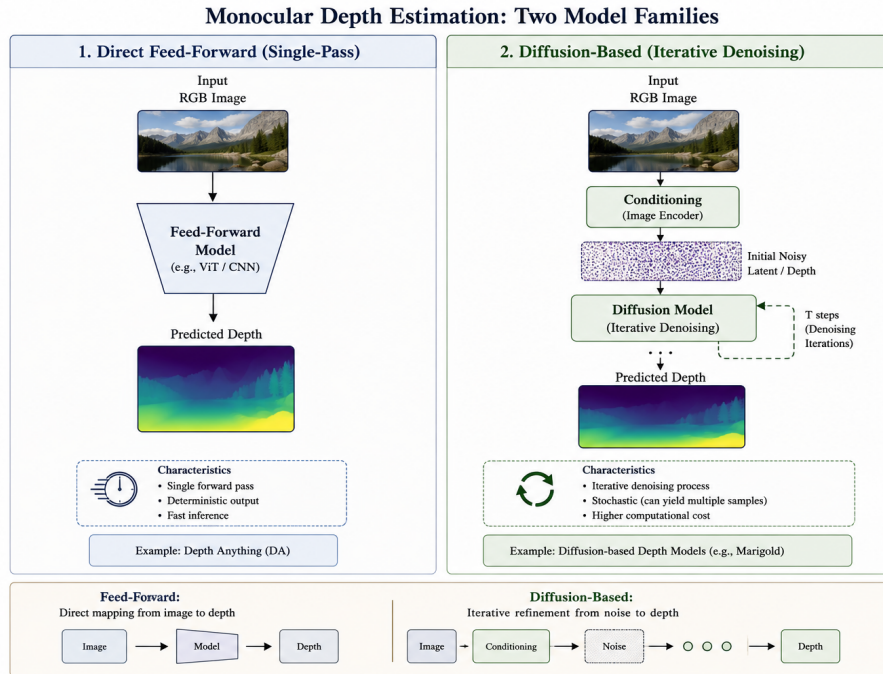


Figure 2.2: Conceptual distinction between direct feed-forward monocular depth estimation and diffusion-based depth prediction.

2.8 Preserving Fine Structure for Depth Estimation

A globally consistent depth map might still be weakly or insufficiently spatially detailed for small structures.

This limitation is especially relevant as narrow boundaries or facial structures cover only a small fraction of the image.

Fine-structure-aware depth methods therefore aim to preserve:

- thin structures;
- sharp object boundaries;
- local discontinuities;
- small geometric transitions.

Pixel-Perfect Depth [30] represents one approach in which high-level semantic information and pixel-level prediction are combined to preserve fine spatial structure in monocular depth estimation.

This distinction is important from the point of view of the generation of bas-reliefs, because an error that looks small in the depth image can turn into a physical ridge or depression after displacement.

In contrast, a good image feature that is weakly represented in depth may be lost when the depth field is compressed to a shallow physical range.

Thus, quality in terms of fine structure should be considered in relation to the final reconstructed surface and not only to the predicted image-space depth map.

2.9 Geometric Interpretation of Surface Normals

The surface position is encoded by depth and the local orientation is encoded by surface normals.

For a unit normal

$$\mathbf{n} = (n_x, n_y, n_z). \quad (2.12)$$

the vector gives the direction normal to the local surface.

For an image-aligned height field $z = D(x, y)$ under a simplified orthographic parameterization, a local surface normal can be approximated as

$$\tilde{\mathbf{n}} = \begin{bmatrix} -\frac{\partial D}{\partial x} & -\frac{\partial D}{\partial y} \\ 1 \end{bmatrix}, \quad (2.13)$$

and normalization:

$$\mathbf{n} = \frac{\tilde{\mathbf{n}}}{\|\tilde{\mathbf{n}}\|_2}. \quad (2.14)$$

Normals can reveal local geometric properties that are difficult to infer directly from the absolute depth.

For example, two surfaces can be in similar depth ranges but have different local orientations.

This makes surface normals useful for

- analysis of local shape;

- curvature processing;
- normal-guided reconstruction;
- consistency assessment;
- structure/detail decomposition.

However, normals do not provide absolute depth independently.

Predicted normal fields may contain inconsistencies, and reconstructing a surface from normals requires an integration process.

Recent joint depth–normal estimation approaches further show the complementary geometric information between depth and surface orientation and the importance of their mutual consistency when the two representations are used together [13].

Hence, depth and normals should be regarded as complementary representations, not as direct substitutes.

2.10 Single-Image 3D Reconstruction

Single-image 3D reconstruction aims to recover a more explicit 3D representation than an image-aligned depth map.

Depending on the method, the output may be:

- a complete mesh;
- a point-based representation;
- an implicit surface;
- a deformable model;
- a canonical 3D object representation.

Such methods may use learned or model-based priors to limit the ambiguity of single-image reconstruction.

For a limited number of object classes, e.g. human face and human body, prior geometric knowledge can provide structural information that cannot be directly recovered from the visible image-aligned depth.

Such priors can thus aid estimation of three-dimensional structure that goes beyond the directly visible image-aligned surface.

However, full reconstruction introduces several additional considerations for bas-relief generation:

- camera calibration;
- scale matching;
- compatibility of pose;
- shape compatibility related to the subject;
- conversion to a shallow image-aligned surface.

A reconstructed object could be plausible in three dimensions but still be inconsistent with the particular projection in the original painting.

This distinction is crucial when the experimental investigation of external geometry is made later in the thesis.

2.11 Point-Map Feed-Forward and Multi-View Geometry

Recent learning-based reconstruction methods have proposed feed-forward frameworks that estimate dense geometric representations from visual input.

For example, DUST3R predicts dense point maps and performs three-dimensional reconstruction without relying entirely on the conventional pipeline of camera calibration, feature matching, and triangulation [27].

In VGGT, this general direction is extended to predict multiple geometric quantities in a feed-forward manner, including camera and spatial information [26].

These models present a significant advance in image-based geometry estimation:

$$\text{RGB input} \rightarrow \text{joint geometric representation} \quad (2.15)$$

Such representations can encode more geometric information than a single image-aligned depth map, including explicit 3D point structure and camera-related information.

However, an explicit 3D representation is not equivalent to the image-aligned shallow geometry required for bas-relief generation.

For a relief the last representation still has to be:

- viewpoint-preserving;
- shallow;
- single-valued along the displacement direction;
- spatially aligned with the original image.

These constraints show that full three-dimensional reconstruction and shallow bas-relief generation should be regarded as related but distinct problems.

2.12 Semantic Information in Geometric Estimation

Semantic information about the content of an image can also be used for geometric processing.

Semantic information can identify:

- foreground objects;
- people;
- faces;
- body parts;
- semantic classes;

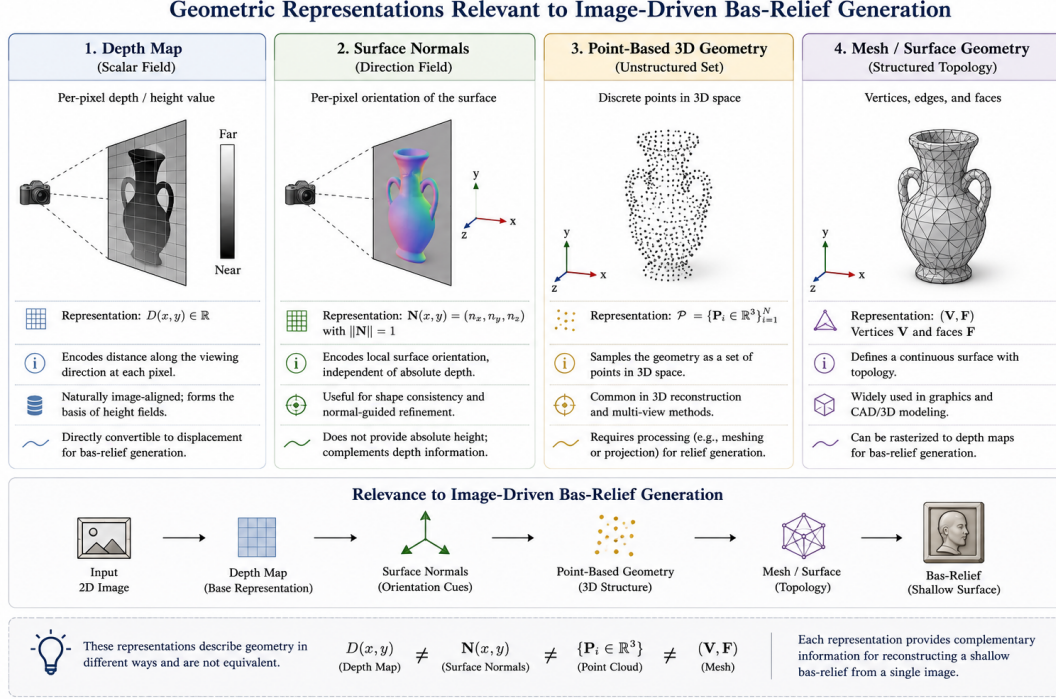


Figure 2.3: Conceptual comparison of geometric representations relevant to image-driven bas-relief generation.

- spatial relationships.

Semantic-aware depth methods have explored the utilization of high-level image information for assisting region consistency and local boundary estimation [31, 19].

Not all regions in an image are equally important; therefore, semantic information can be useful for relief generation.

A local operation can be generally written as

$$D'(x, y) = D(x, y) + W_S(x, y)\Delta D(x, y), \quad (2.16)$$

where W_S denotes a semantic spatial weight.

However, semantic localization and geometric specification are distinct tasks.

A face mask can identify where a face is located, but it does not specify the desired three-dimensional relationships among the nose, cheeks, eyes, lips, and chin.

The experimental chapters make use of this conceptual distinction in a crucial way.

2.13 Depth to 3D Surface

Once a depth map is obtained, it must be converted into a geometric representation.

For an image-aligned height-field representation, each sampled depth value can be associated with a mesh vertex:

$$\mathbf{v}_{ij} = \begin{bmatrix} x_i \\ y_j \\ sD(x_i, y_j) \end{bmatrix}. \quad (2.17)$$

Neighbouring vertices can then be connected to form triangles.

Alternatively, a depth image can be used directly as a displacement texture on an existing plane.

The displacement formulation is

$$\mathbf{v}' = \mathbf{v} + sD(\mathbf{u})\mathbf{n}, \quad (2.18)$$

$\mathbf{u}$ is the texture coordinate, $\mathbf{n}$ is the local displacement direction, and s controls the displacement magnitude.

This approach has two major advantages when used for bas-relief applications.

First, the mapping from image to surface is preserved.

Second, the same topology of the mesh and the same displacement configuration can be used to compare different depth candidates.

This renders the displacement-based reconstruction particularly convenient for the controlled experimental comparisons performed in this thesis.

2.14 Bas-Relief Representation

Bas-relief is a sculptural method that represents three-dimensional form within a restricted range of depth.

If the depth range of an original scene is

$$Z_{\min} \leq Z \leq Z_{\max}, \quad (2.19)$$

the corresponding relief may have to fit in a much smaller physical range

$$h_{\min} \leq h \leq h_{\max}. \quad (2.20)$$

A simple linear compression is given by

$$h = h_{\min} + \frac{Z - Z_{\min}}{Z_{\max} - Z_{\min}} (h_{\max} - h_{\min}). \quad (2.21)$$

While simple, linear compression treats all depth variation equally.

For scenes with large foreground-background separation and small facial details, local structures may become relatively shallow after global scaling.

This is an important geometric consideration for the image-driven shallow-relief problem dealt with in this thesis.

2.15 Classical Relief Processing

Classical computational bas-relief methods generally assume the availability of reliable three-dimensional geometry.

Their main goal is to compress that geometry, while preserving important local shape.

2.15.1 Processing in Gradient Domain

Gradient domain methods do not work directly on absolute depth, but on local surface variations.

If $H(x,y)$ is a height field, its gradient is given by

$$\nabla H = \left(\frac{\partial H}{\partial x}, \frac{\partial H}{\partial y} \right). \quad (2.22)$$

Large gradients can be compressed while smaller gradients are retained or selectively amplified.

Then a new height field is reconstructed from the modified gradient field.

Weyrich et al. studied a gradient-domain formulation to compress 3D scenes into shallow reliefs [29].

The fundamental principle is to control local surface variation independently of the global compression of absolute depth.

2.15.2 Adaptive and Nonlinear Compression

Instead of a global scale factor, nonlinear mappings can reassign the available height range.

A generic nonlinear transformation can be written as

$$H' = T(H), \quad (2.23)$$

where the derivative of T is a function of depth.

Adaptive compression methods attempt to redistribute the available relief height so that different ranges or structures are not necessarily compressed in the same way.

Sun et al. studied adaptive relief generation by nonlinear redistribution of geometry [25].

2.15.3 Multi-Scale Processing

A surface can be decomposed into a number of spatial frequency components:

$$H = H_{\text{base}} + H_{\text{detail}}. \quad (2.24)$$

The base component describes broad geometry, and the detail component describes smaller local variation.

The two parts can then be compressed differently.

This idea is particularly relevant for shallow relief where broad scene depth and small surface features compete for the same limited physical height range.

2.15.4 Detail-Preserving Processing

In detail-preserving methods the goal is to retain small geometric structures after aggressive global compression.

Bian and Hu investigated relief processing based on the preservation of gradient- and Laplacian-related detail [2].

The general principle is not to let the compression of the global geometry remove the local surface information.

However, detail-preserving processing remains limited by the geometric information already present in the input. It cannot independently reconstruct structures that are missing from the original depth map or mesh.

2.15.5 Normal-Based Relief Processing

Surface normals provide an alternative means of describing broad surface structure and fine local orientation. Wei et al. [28] studied a formulation that decomposes surface-normal information into base and detail components for relief processing. This representation enables local orientation information to be manipulated independently of absolute depth compression.

However, as with other classical relief-processing methods, this approach generally assumes that reliable source geometry or normal information is already available.

2.16 Local and Global Geometric Information

The distinction between global and local geometry appears repeatedly across depth estimation, surface reconstruction, and relief processing. Global geometry describes the broad spatial organization of the scene, including:

- subject ordering;
- large depth layers;
- scene composition;
- broad surface shape.

Local geometry instead describes smaller structures and surface variations, including:

- small protrusions;
- local curvature;
- facial structures;
- fine surface transitions;
- narrow boundaries.

These two spatial scales are not necessarily preserved equally after relief compression. Large scene-level depth differences may occupy most of the available physical height range, leaving comparatively little variation for smaller but structurally important features. The relief pipeline must therefore distribute the available physical depth across multiple spatial scales.

This does not imply that local detail should simply be amplified. Excessive enhancement can introduce:

- spikes;
- artificial ridges;
- mask-shaped bumps;
- exaggerated surface transitions.

The problem is therefore one of controlled redistribution of geometric information rather than simple maximization of local depth variation.

2.17 Depth Compression and Nonlinear Remapping

Depth predictions often occupy a numerical range that is not directly suitable for shallow physical displacement. Depth remapping therefore modifies the distribution of the predicted values before surface reconstruction.

A normalized depth field may first be expressed as

$$D_n = \frac{D - D_{\min}}{D_{\max} - D_{\min}}. \quad (2.25)$$

Nonlinear transformations can subsequently be applied:

$$D_r = T(D_n). \quad (2.26)$$

Possible transformations include:

- gamma mapping;
- percentile normalization;
- piecewise nonlinear compression;
- region-dependent mapping;
- contrast redistribution.

The objective is not necessarily to improve the accuracy of the original monocular prediction. Instead, the depth distribution can be adapted to the physical and geometric requirements of the shallow-relief representation.

2.18 Boundary Preservation and Surface Regularization

Depth refinement can introduce artifacts when neighboring image regions are subjected to different transformations. A discontinuous local update may, for example, generate an artificial step:

$$D'(x, y) = \begin{cases} T_1(D), & (x, y) \in R, \\ T_2(D), & (x, y) \notin R. \end{cases} \quad (2.27)$$

If T_1 and T_2 differ strongly, the boundary of the modified region can appear as an artificial geometric cliff. Smooth blending can reduce this effect:

$$D' = (1 - W)D_1 + WD_2, \quad 0 \leq W \leq 1. \quad (2.28)$$

A simple smoothness regularization term may also penalize excessive local variation:

$$E_{\text{smooth}} = \sum_{(x, y) \in \Omega} \|\nabla D'(x, y)\|_2^2. \quad (2.29)$$

Boundary stability is particularly important in bas-relief generation because even a small discontinuity in the depth representation becomes an explicit physical surface transition after displacement.

2.19 Geometric Priors and Anatomical Constraints

Image-derived depth is not the only possible source of geometric information.

Additional priors may include:

- object templates;
- facial landmarks;
- body models;
- parametric shape primitives;
- externally reconstructed geometry;
- surface-normal information.

A prior can be introduced as a local target:

$$D' = D + \alpha WP, \quad (2.30)$$

where P represents the prior geometry and W defines its spatial support.

Geometric priors can inject structural information that is not readily apparent in the baseline depth map.

The main risk is that the prior may not correspond accurately to the subject depicted in the image. This creates a fundamental trade-off: increasing the influence of a geometric prior may produce stronger structural modification, but it may also increase the risk of incompatibility with the subject, viewpoint, or local geometry represented in the pictorial input.

Therefore, prior-based modification calls for careful control of alignment, spatial support, modification strength and compatibility with the baseline geometry.

2.20 From Visual Geometry to Tactile-Oriented Geometry

A visually plausible depth map is not automatically an effective tactile representation [22].

Visual perception can use:

- colour;
- shading;
- texture;
- perspective;
- occlusion;
- fine luminance variation.

A tactile relief is based primarily on physical surface geometry. Consequently, visually strong information may be lost during geometric conversion, while depth artifacts that appear minor in the image may become physically prominent.

Some properties are useful for computational analysis:

- depth range;
- local surface variation;
- transition continuity;
- minimum feature scale;
- surface regularity;
- geometric separation between neighbouring regions;
- fabrication compatibility.

These properties can be quantified computationally; however, such geometric analysis should not be confused with direct evidence of tactile perception. Tactile recognition depends on perceptual factors that cannot be established from geometric measurements alone [12].

Claims of tactile recognition or readability require physical fabrication and evaluation with users.

In this thesis, tactile considerations therefore serve as design motivations and geometric constraints rather than as perceptually validated performance metrics.

2.21 Evaluation of Depth-Driven Relief Geometry

Evaluation of an image-driven relief can occur at several levels.

2.21.1 Depth-Domain Evaluation

Depth-domain analysis may include:

- minimum and maximum depth;
- percentile range;
- local contrast;
- residual maps;
- region-specific statistics.

These measures indicate how a transformation modifies the numerical representation.

2.21.2 Gradient and Structural Evaluation

Depth gradients describe local surface variation.

A gradient magnitude can be computed as

$$G = \sqrt{\left(\frac{\partial D}{\partial x}\right)^2 + \left(\frac{\partial D}{\partial y}\right)^2}. \quad (2.31)$$

Statistics on the gradients can tell whether a method enhances or dampens local structure.

However, high gradient magnitude is not necessarily desired as it can also correspond to noise or sudden artifacts.

2.21.3 Final Surface Evaluation

The final mesh or displaced surface should also be inspected directly.

Relevant characteristics include:

- global shape preservation;
- local surface organization;
- unwanted ridges or valleys;
- boundary discontinuities;
- surface smoothness;
- geometric stability after displacement.

This level of evaluation is important as changes in numerical depth may not be equally visible after the conversion into the final surface.

2.21.4 Physical and Tactile Evaluation

The strongest form of validation requires fabrication of the relief and evaluation of the physical object.

Possible evaluation dimensions include:

- printability;
- survival of small features;
- tactile distinguishability;
- subject localization;
- recognition of meaningful structures.

Direct claims about tactile performance would require such perceptual assessment.

The experiments in this thesis are mainly computational and geometric, and this limitation is clearly stated in the interpretation of the results.

2.22 Comparison of the Main Technique Families

The techniques described in the previous sections address different aspects of the image-to-bas-relief problem and therefore provide different advantages and limitations.

Classical image-based methods give direct and relatively controllable transformations, but their geometric interpretation is often limited due to their dependence on intensity, gradient or manually defined assumptions.

Monocular depth estimation offers a more explicit representation of scene structure while preserving direct correspondence with the source image. However, the resulting depth maps are not specialized for shallow bas-relief fabrication and may represent locally important geometric structures only weakly.

Although explicit 3D reconstruction methods can offer richer geometric information, they also bring additional requirements in camera estimation, registration, projection consistency and conversion of free-form geometry into a stable height-field representation.

Semantic and geometry-aware refinement methods offer more fine-grained control to selected local regions, but their effectiveness is limited by the reliability of the underlying geometric representation and the availability of suitable spatial/structural cues.

Different technique families must therefore not be considered as directly competing solutions under the same conditions. Instead they provide different kinds of geometric information, which can be useful at different stages of an image-driven bas-relief pipeline.

Within the framework of single-image bas-relief generation, their main characteristics are summarized in Table 2.1. This comparison is intended as a methodological overview rather than a quantitative ranking, because the different families of techniques are typically evaluated on different datasets, objectives, and experimental conditions.

Table 2.1: General comparison of technique families relevant to image-driven bas-relief generation.

Technique family	Typical input	Representation	Principal strength	Main limitation for single-image relief
Stereo / multi-view geometry	Multiple images	Depth / point cloud	Geometrically constrained re-construction	Requires multiple views of the scene.
Monocular depth	Single RGB image	Depth map	Direct image-aligned geometry	Local geometric structure may be represented unevenly or weakly.
Diffusion-based depth	Single RGB image	Depth map	Diffusion-based learned representation	Iterative inference and no direct guarantee of relief-specific geometry.
Surface-normal estimation	RGB or geometry	Normal field	Local surface orientation	Does not provide absolute surface position independently.
Single-image 3D reconstruction	Single RGB image	Mesh / implicit 3D	Can provide richer object geometry	May introduce pose, projection, and subject-compatibility mismatch.
Feed-forward 3D geometry	One or more images	Point maps / cameras / depth	Rich geometric representation	Requires conversion to a fixed shallow height field.
Semantic processing	RGB + geometry	Region-aware geometry	Provides region-specific spatial localization	Semantics do not independently specify local surface shape.

Continued on next page

Technique family	Typical input	Representation	Principal strength	Main limitation for single-image relief
Classical bas-relief processing	Existing geometry	Height field / mesh	Controlled depth compression	Usually assumes reliable geometry already exists.
Parametric geometric priors	Depth + anatomical support	Local correction field	Explicit structural control	Depends on prior compatibility and localization.

2.23 Implications for Image-Driven Bas-Relief Generation

The technique families discussed in this chapter provide complementary capabilities for the 2D-to-3D problem. Monocular depth estimation offers a direct image-aligned route from RGB input to a dense geometric representation, while surface normals provide information about local orientation. Single-image and feed-forward 3D reconstruction methods can provide richer geometry but require additional representation conversion and alignment. Semantic information supports spatial localization, whereas classical bas-relief processing provides strategies for compressing existing geometry into a limited physical height range.

These complementary capabilities show that the overall problem cannot be reduced to the selection of a single three-dimensional reconstruction model.

For the target application, a number of operations have to be taken into account together:

stable global geometry
+ controlled depth compression
+ local structural preservation
+ stable geometric transitions.

At the same time, increasing geometric richness or algorithmic complexity does not automatically guarantee a better relief.

Any additional geometric information must remain compatible with:

- the source viewpoint;
- the image-aligned baseline;
- the shallow height-field constraint;
- the surrounding surface geometry.

The following chapter narrows this broad technical background to the research most directly related to these requirements.

2.24 Chapter Summary

This chapter presented the main technical concepts involved in translating two-dimensional visual input into three-dimensional or relief-like representations. Classical geometric methods

recover depth through multi-view correspondence, camera motion, or photometric assumptions, but these approaches generally require information or constraints that are unavailable or difficult to enforce for a single painting.

Learning-based monocular depth estimation provides a direct mapping from a single RGB image to an image-aligned depth field. Relative and metric depth serve different geometric purposes, while modern large-scale models extend image-derived estimation toward surface normals, camera information, point-based geometry, and other representations. Diffusion-based depth estimation provides an alternative inference strategy, and fine-structure-aware methods aim to preserve small image-aligned geometric features.

Surface normals complement scalar depth by describing local orientation, whereas single-image and feed-forward 3D reconstruction methods can provide richer geometric information at the cost of additional alignment, projection, compression, and representation-conversion requirements. Semantic information can localize important regions but does not independently determine their desired geometry.

Classical bas-relief processing addresses limitations of simple global depth scaling through gradient-domain, nonlinear, multi-scale, detail-preserving, and normal-based operations. Together, these directions show that image-driven tactile-oriented bas-relief generation lies at the intersection of depth estimation, geometric reconstruction, local refinement, and controlled relief compression.

For the problem studied in this thesis, the desired representation should preserve stable global geometry while providing limited physical depth, controlled local structure, and smooth surface transitions. The following chapter therefore focuses on the research most directly related to these requirements and identifies the methodological gap addressed by the experimental investigation.

Chapter 3

Related Work

3.1 Introduction

Chapter 2 introduced the main technical families relevant to 2D-to-3D conversion and shallow-relief generation.

The aim of the present chapter is different: rather than discussing these technique families in general, it reviews representative work directly related to the specific problem addressed in this thesis:

the generation of a shallow, viewpoint-preserving bas-relief from a single pictorial image while maintaining stable global geometry and enabling controlled refinement of locally important structures.

The relevant literature can be broken down into several partially overlapping directions:

- classical bas-relief generation from existing geometry;
- detail-preserving and normal-based relief processing;
- tactile reproduction of pictorial artworks;
- portrait and human bas-relief reconstruction;
- general-purpose monocular depth estimation;
- fine-structure and boundary-aware depth prediction;
- depth–normal geometric consistency;
- modern feed-forward three-dimensional reconstruction;
- semantic and region-aware geometric estimation;
- external local geometry and hybrid fusion.

Each direction addresses part of the target problem but starts from different assumptions and optimizes different objectives.

The experimental comparison and selection of a working depth baseline are deliberately omitted. These decisions are part of the original experimental work and are discussed in Chapter 4.

This chapter focuses on the methodological relevance and limitations of the reviewed approaches and on the research gap that motivates the experimental investigation.

3.2 Classical Bas-Relief Generation

Classical computational bas-relief methods generally presume that valid three-dimensional geometry exists.

Hence, the primary challenge is not to infer geometry from an image but to convert existing geometry into a shallow height range.

Weyrich et al. described the creation of digital bas-reliefs in the gradient domain [29].

Rather than compressing the absolute depth uniformly, the method modifies the gradient field to control large-scale geometric variation while preserving some local surface information. This expression gives an important principle:

$$\text{local shape information} \neq \text{global depth range} \quad (3.1)$$

This distinction is particularly relevant to the present thesis because a complete painting may contain a large range of scene depth, while important facial structures occupy only a small fraction of that range.

However, a limitation of classical gradient domain bas-relief methods for the present problem is that they assume the source geometry is available a priori.

The present problem begins with a single pictorial image; geometry must therefore first be estimated before relief-specific compression can be applied.

3.2.1 Adaptive Depth Compression

Sun et al. investigated adaptive relief generation based on nonlinear redistribution of the available height range [25].

Such methods do not apply a single global scale factor but distribute the different parts of the source depth range in a non-uniform way.

This is important for the present application, otherwise the global scene depth might hide small local geometric differences.

However, the method operates on the geometry that already exists for the problem at hand, and does not address the preceding task of single-image geometry estimation.

3.3 Detail-Preserving Bas-Relief Methods

Even if the large-scale surface is recognizable, strong depth compression can suppress fine geometric structure.

Bian and Hu [2] proposed a detail-preserving framework based on the gradient-domain compression and Laplacian-related local information.

They explicitly separate the larger shape and the finer detail so that the latter can be restored after strong compression.

This direction is relevant because the target relief has to preserve small but meaningful surface structures in a strongly restricted height range.

However, detail-preserving relief processing still depends on geometric information already present in the source representation.

If a facial feature is missing or weak in the estimated depth map, the decomposition and re-injection alone does not provide an independent source of the missing geometry.

Another painting-specific approach was proposed by Fu et al. [6], who generate high-relief geometry from a single brush painting by explicitly extracting brushstrokes and constructing corresponding relief proxies. Rather than attempting to recover an exact scene-depth map, their method preserves characteristic painted structures through stroke decomposition, inflation, and displacement. This is relevant to the present thesis because it demonstrates that painterly information can itself guide relief construction. However, the target representation differs substantially: their method focuses on texture-mapped high relief and is particularly suited to paintings with explicit brushstroke structure, whereas the present work investigates a shallow, viewpoint-preserving height field derived from monocular geometry and controlled local refinement.

3.4 Normal-based Relief Synthesis

Another way to describe the local geometry is by using surface normals.

Wei et al. [28] proposed a normal-based relief processing framework for the separation of base and detail normal components.

This formulation allows independent treatment of large-scale surface orientation and smaller local variations.

This approach is relevant for the present thesis because the orientation of the surface may contain local shape information which is not as apparent in scalar depth values.

However, normal-based relief methods usually assume that the source geometry provides a suitable normal field.

The normals computed independently from a painting must be verified to be consistent with the depth representation before they are used for geometric refinement.

This issue is examined experimentally later in the thesis.

3.5 Tactile Reproduction of Pictorial Art

A research direction particularly relevant to the present thesis concerns the translation of pictorial artworks into tactile 2.5D or bas-relief representations for blind and visually impaired users.

Furferi et al. [10] proposed a semi-automatic methodology for converting paintings into tactile 2.5D models. Their work explicitly addressed the difficulties associated with recovering relief geometry from pictorial input and considered the accessibility requirements of blind and visually impaired users. The resulting tactile models were evaluated with end users, providing an important example of bas-relief generation in which the objective extends beyond geometric reconstruction to the tactile accessibility of artworks.

Related work by Furferi, Governi, Vanni, and Volpe [9] investigated the semi-automatic generation of tactile bas-reliefs from single-point-perspective paintings. Their approach combined computer-assisted geometric reconstruction with interactive modelling and was specifically motivated by reducing the sensory barrier between visually impaired users and intrinsically two-dimensional works of art.

More recently, Furferi [8] reviewed artificial-intelligence approaches for generating three-dimensional models from two-dimensional artworks for tactile exploration. The study discusses

several learning-based reconstruction paradigms, including convolutional neural networks, generative models, and zero-shot approaches, and emphasizes the intrinsic ambiguity of recovering three-dimensional geometry from a single painting. The reported case studies also indicate that AI-generated geometry may still require post-processing and user interaction before it becomes suitable for reliable tactile reproduction.

The work most closely related to the present thesis is the recent study by Furferi et al. [11], which directly compares AI-based strategies for 2.5D reconstruction from modern artwork paintings. Their evaluation considers monocular depth estimation, large multimodal or language-model-based generation, and large reconstruction models. The comparison is not limited to visual plausibility: it also considers geometric consistency, integrability, topological robustness, and manufacturability for tactile reproduction. Their results highlight complementary strengths and weaknesses among the different reconstruction paradigms. In particular, monocular depth estimation can preserve broad depth organization while attenuating fine details, whereas more generative approaches may provide richer semantic or three-dimensional structure at the cost of geometric or topological inconsistencies. The authors therefore identify hybrid strategies as an important direction for future tactile reconstruction from paintings.

This recent comparison is particularly relevant because it addresses essentially the same high-level transformation considered in this thesis: converting a single painting into a physically meaningful 2.5D representation. The methodological emphasis, however, is different. Furferi et al. [11] compare alternative AI reconstruction paradigms and evaluate their suitability for 2.5D tactile reproduction. The present thesis instead fixes an image-aligned DA3-derived baseline and studies, in a controlled experimental sequence, how additional geometric information can be introduced through external geometry, semantic supports, anatomical priors, surface normals, confidence weighting, landmarks, curvature, and mesh-space processing while preserving the global shallow-relief surface.

A further difference is the role of local refinement. Rather than selecting a single reconstruction paradigm according to overall output quality, the present work investigates how local corrections propagate through a fixed shallow displacement pipeline. This makes it possible to analyze global preservation, boundary disturbance, local geometric contrast, and the physical magnitude of individual modifications. The objective is therefore not to replace the broader comparative evaluation of AI reconstruction methods, but to complement it with a detailed analysis of baseline-preserving local geometric refinement.

Taken together, this line of research establishes tactile bas-relief as a practical representation for increasing access to pictorial artworks while also demonstrating that the problem cannot be reduced to generating visually plausible three-dimensional geometry. The reconstructed surface must also be geometrically stable, sufficiently shallow, manufacturable, and compatible with tactile exploration. In the present thesis, these requirements motivate the geometric analysis, while direct claims about tactile effectiveness remain outside the scope of the computational evaluation.

3.6 Reconstruction of Portraits and Human Bas-Reliefs

Another related line of research concerns human and facial geometry. Methods operating in the portrait domain can exploit class-specific anatomical priors that are unavailable to generic scene-reconstruction methods.

A classical example of such a class-specific facial prior is the 3D Morphable Model introduced by Blanz and Vetter [3]. In this formulation, facial shape and texture are represented using a statistical model derived from registered three-dimensional face examples, allowing a plausible 3D face to be fitted to image observations. Such models demonstrate how prior knowledge can constrain the otherwise ambiguous reconstruction of facial geometry from limited visual information. However, their assumptions of statistical facial shape, correspondence, and model fitting do not directly guarantee compatibility with stylized painted faces or with the strongly compressed height-field representation required for tactile bas-relief.

Liu et al. explored the reconstruction of facial bas-relief from a single portrait with the help of template geometry and image details [20].

This approach is relevant to the local facial-refinement problem because it introduces explicit geometric structure (which is missing in a broad semantic face mask).

However, these methods rely on several assumptions:

- the input resembles the facial domain used by the method;
- the facial template is compatible with the subject;
- pose and viewpoint can be aligned sufficiently;
- the reconstructed geometry can be transformed into the desired shallow relief.

Such assumptions may be more difficult to satisfy for painted faces, considering that stylization, partial occlusion, dramatic illumination, and non-photographic appearance may not correspond to the conditions represented by conventional portrait imagery.

Yang et al. studied the combination of the guide geometry and the local surface-orientation information for human reconstruction [32].

This provides a relevant example of combining global geometric guidance with local surface-orientation information.

However, the method relies on a compatible guide representation and is not tailored to generate fixed viewpoint relief from paintings.

3.7 General-Purpose Monocular Depth Estimation

Recent monocular depth models directly learn a mapping from RGB input to dense image-aligned geometry.

This eliminates the need for an existing 3D scene, making such models particularly interesting for image-driven bas-relief generation.

Several representative model families were considered during the development of this thesis.

3.7.1 Metric3D v2

Metric3D v2 [13] predicts metric depth and surface-normal information in a unified geometric framework.

Its relevance to the present problem lies in the availability of both global depth information and local surface-orientation cues.

However, the model is optimized for general geometric estimation, not specifically for shallow relief compression or local tactile-oriented control.

The availability of richer geometric outputs therefore does not negate the requirement of relief-specific processing.

3.7.2 Depth Anything 3

Depth Anything 3 is a general-purpose image-derived geometry framework that can produce dense spatial information from one or more images [18].

Its output is relevant to the present problem because it provides dense image-derived geometry while preserving spatial correspondence with the source image, which is important for a height-field relief in which image location maps directly to surface displacement.

However, the model is not specialized for bas-relief generation.

Its depth output must therefore still be adapted to:

- restricted physical relief height;
- local structural preservation;
- boundary stability;
- final displacement reconstruction.

The suitability of this model relative to alternative depth estimators is evaluated experimentally in Chapter 4.

3.7.3 Marigold

Marigold [16] formulates monocular depth estimation using diffusion-model priors.

This method exploits generative diffusion priors for monocular depth estimation from a single RGB image.

The relevance of this work to the thesis is methodological: it provides an alternative depth-estimation paradigm based on iterative diffusion inference instead of direct feed-forward prediction.

Its experimental behaviour after relief-oriented processing is evaluated rather than inferred from the depth-estimation formulation itself.

3.8 Fine-structure and Boundary-aware Depth Estimation

General monocular depth models may be more explicit in modeling large-scale scene organization than narrow boundaries or small local structures.

Pixel-Perfect Depth addresses the problem of preserving fine spatial structure in monocular depth estimation by combining high-level semantic information with pixel-space prediction [30].

This direction is relevant to tactile-oriented relief design because local surface transitions are still explicit geometric features after displacement-based reconstruction.

A small depth artifact may become an explicit physical ridge after displacement, and a weak facial feature may vanish after global depth compression.

However, a method designed to improve depth fidelity does not explicitly define how a certain anatomical region should be changed for a shallow relief.

Thus, fine-structure preservation and anatomy-aware local refinement remain distinct problems.

3.9 Depth and Surface-Normal Consistency

Surface normals and depth provide complementary representations of surface geometry. Surface normals encode local surface orientation, whereas depth represents spatial position.

More recent learning-based approaches have studied these representations jointly rather than treating them as completely independent geometric quantities. For instance, Metric3D v2 explicitly enforces depth-normal consistency in the normal estimation and jointly optimizes depth and normals [13].

Other works have studied the relationship between depth and surface-normal prediction. These methods show that information can be exchanged between the two representations, but also that their geometric relation has to be taken into account when they are combined [7, 21].

This is especially true when depth and normals are obtained through independent estimation processes. Disagreement between the two fields does not necessarily mean one representation is right and the other is wrong. Rather, the difference might be:

- complementary local geometric information;
- systematic orientation bias;
- prediction uncertainty;
- domain mismatch;
- local failure of one or both representations.

Therefore, an independently predicted normal field should not be used to automatically correct an existing depth map by using it as ground-truth geometry. Its compatibility with the depth-derived surface must therefore be evaluated first.

This distinction is particularly pertinent for the methodology developed later in this thesis. Instead of directly replacing the orientation implied by the DA3+Remap baseline, the geometry-aware experiments first evaluate the depth-normal agreement and then use the independent normal information as an auxiliary local geometric constraint.

3.10 Feed-Forward 3D Reconstruction

Recent research has moved toward increasingly unified feed-forward geometric reconstruction.

DUSt3R predicts dense point maps and performs global alignment without requiring the full traditional sequence of calibration, correspondence estimation and triangulation [27].

VGGT goes one step further by jointly predicting geometric quantities such as depth, point maps and camera information [26].

These systems provide more explicit 3D representations than a single image-aligned depth map, such as point-based and camera-related geometric information.

Their outputs, however, are not yet the needed representation for a shallow bas-relief.

A full point or camera-aware reconstruction still needs to be converted into a surface that is:

- image-aligned;

- viewpoint-preserving;
- single-valued along the relief direction;
- strongly compressed in depth.

The main issue is not only what and how much geometric information can be reconstructed, but whether this information is still compatible with the intended relief representation.

3.11 Semantic and Region-Aware Geometric Estimation

Semantic-depth approaches address the use of high-level scene information as an additional cue to geometric prediction.

Approaches that fuse semantic information with local and global depth cues have explored region consistency and object-boundary estimation [31, 19].

This is important for the present thesis because different image regions do not contribute equally to the final relief.

Semantic information can help identify:

- faces;
- human figures;
- foreground structures;
- locally important objects.

However, semantic localization itself does not determine the desired internal surface.

Knowing the spatial extent of a face does not specify the correct relative geometry of the nose, cheeks, eye sockets, lips, or chin.

This distinction is crucial for the experimental method developed below:

$$\text{semantic localization} \neq \text{anatomical geometry.} \quad (3.2)$$

3.12 External Geometry and Local Fusion

Another possible strategy is to combine a stable global depth representation with a separate source of local 3D geometry.

Modern feed-forward reconstruction methods, such as DUS3R and VGGT, yield explicit three-dimensional representations directly from visual input [27, 26]. However, using such geometry within a fixed shallow-relief representation introduces an additional compatibility problem.

The motivation is straightforward:

$$\text{global image-aligned geometry} + \text{local reconstructed geometry} \rightarrow \text{hybrid surface.} \quad (3.3)$$

Such a strategy provides a possible way to introduce additional local geometric information in areas that are poorly represented by the monocular depth.

However, hybrid fusion introduces several practical requirements:

- coordinate-system alignment;
- scale normalization;
- viewpoint compatibility;
- ROI correspondence;
- boundary continuity;
- conversion from full 3D geometry to a shallow height field.

Thus, in the present problem, the availability of additional local three-dimensional geometry is not enough by itself.

The local prior must also remain consistent with the global representation.

This problem motivates the hybrid-geometry experiments discussed in Chapter 5.

3.13 Comparison of Representative Approaches

The reviewed research addresses complementary parts of the target problem.

Table 3.1 compares representative approaches according to their input, geometric representation, treatment of local structure, and principal limitation with respect to the requirements of this thesis.

Table 3.1: Comparison of representative approaches relevant to image-driven tactile-oriented bas-relief generation. The final column describes limitations specifically with respect to the requirements of this thesis rather than general weaknesses of the original methods.

Method	Input	Representation	Relevant characteristics	Limitation for the present problem
Weyrich et al. [29]	3D scene	Height / gradient field	Gradient-domain compression with controlled treatment of local surface variation.	Requires reliable pre-existing geometry and does not infer geometry from pictorial input.
Sun et al. [25]	Height field / 3D geometry	Depth / height field	Adaptive nonlinear redistribution of the available relief range.	Addresses compression of known geometry rather than single-image geometry estimation.
Bian and Hu [2]	3D model	Gradient and Laplacian detail	Explicit preservation and restoration of fine surface detail.	Remains dependent on local structures represented in the source geometry.
Wei et al. [28]	3D model / normals	Base and detail normals	Separates broad surface structure from fine orientation detail.	Assumes reliable source geometry or normal information.

Continued on next page

Method	Input	Representation	Relevant characteristics	Limitation for the present problem
Liu et al. [20]	Single portrait image	Template geometry + local detail	Introduces facial geometry using portrait-specific structural priors.	Depends on facial-template compatibility, alignment, pose, and portrait assumptions.
Yang et al. [32]	Single photograph	Guide geometry + normals	Combines guide geometry with local surface-orientation information.	Requires compatible guide geometry and is oriented toward photographic human inputs.
Metric3D v2 [13]	Single RGB image	Metric depth + normals	Provides metric depth together with surface-normal information.	Optimized for general geometric estimation rather than shallow-relief compression.
Marigold [16]	Single RGB image	Depth	Uses diffusion priors for monocular depth prediction.	Does not directly optimize relief-specific depth allocation or anatomical local control.
Pixel-Perfect Depth [30]	Single RGB image	Relative depth	Targets fine spatial structure through semantic and pixel-level information.	Improved depth fidelity does not itself define anatomy-aware relief modification.
Depth Anything 3 [18]	One or more images	Depth / spatial geometry	Provides dense general-purpose image-derived geometry.	Requires relief-oriented remapping and local control for the target application.
DUSt3R [27]	One or more images	Dense point maps	Provides dense point-based geometry within a feed-forward reconstruction framework.	Output is not directly compatible with a fixed shallow height-field representation.
VGGT [26]	One or more images	Depth, point maps, cameras	Jointly predicts multiple geometric quantities within a unified feed-forward framework.	General 3D reconstruction does not directly satisfy viewpoint-preserving relief constraints.
Semantic depth methods [31, 19]	RGB image	Depth + semantic features	Use semantic information to support region and boundary-aware depth estimation.	Semantics provide localization but do not prescribe anatomy-specific local surface geometry.

3.14 Synthesis of the Related Work

The literature contains methods that address many of the individual components of the problem considered in this thesis. Classical bas-relief research provides established strategies for compressing known geometry, while monocular depth estimation recovers image-aligned scene geometry from a single RGB image. Portrait-specific reconstruction methods introduce class-specific anatomical priors, surface-normal approaches provide local orientation information, and modern point-map and feed-forward reconstruction methods generate more explicit three-dimensional representations. Semantic approaches, in turn, provide spatial localization. These capabilities are generally optimized separately rather than under the combined constraints required for image-driven shallow-relief generation.

The target problem requires these capabilities to operate together under a set of constraints that is not simultaneously addressed by the reviewed methods.

In particular, the required system should:

1. infer geometry from a single pictorial image;
2. preserve the original composition and viewpoint;
3. compress the geometry into a shallow height range;
4. preserve stable global scene organization;
5. allow controlled treatment of locally important structures;
6. avoid unsupported geometric replacement;
7. maintain smooth transitions between modified and unmodified regions.

The literature thus provides the main building blocks for the problem, but no complete solution under the constraints investigated in this thesis.

3.15 Research Gap

This thesis addresses the research gap at the intersection of monocular geometry estimation, bas-relief compression and localized semantic or anatomical refinement.

Existing bas-relief methods usually assume that valid three-dimensional geometry is already available. Modern monocular depth and geometry models can generate image-derived geometric estimates without requiring pre-existing scene geometry, but they are not specifically designed for shallow tactile-oriented relief generation. Portrait-specific methods can introduce local anatomical structure but often rely on template compatibility and photographic assumptions, while semantic methods provide spatial localization without defining the desired internal geometry. Modern three-dimensional reconstruction methods offer richer geometric representations, but they introduce alignment and representation challenges when converted into a fixed image-aligned height field.

The remaining research question can therefore be formulated as:

How can a globally stable surface derived from images be maintained, and locally important structures be refined, using additional geometric or semantic information,

without introducing viewpoint inconsistency, unsupported geometry or unstable surface transitions?

This gap motivates the experimental pipeline developed in the following chapters.

3.16 Positioning of the Thesis

The thesis does not propose an alternative to monocular depth estimation or a new full three-dimensional reconstruction model.

Instead, the work is framed as a controlled study of adapting and refining existing geometric representations for shallow bas-relief generation from pictorial input.

A stable global image-derived baseline is central to the experimental framework of this thesis. Additional information is then introduced progressively through:

- external local geometry;
- semantic region constraints;
- parametric facial priors;
- signed anatomical templates;
- independently estimated normals;
- confidence weighting;
- landmarks;
- curvature;
- explicit mesh-space processing.

The main methodological aim is to investigate how these cues affect local geometry and whether their use remains compatible with the stability of the overall shallow relief.

The work therefore occupies a position between:

$$\text{image-based geometry estimation} \tag{3.4}$$

and

$$\text{relief-specific local geometric refinement.} \tag{3.5}$$

3.17 Boundary of the Claimed Contribution

The scope of the contribution is explicitly bounded as follows.

The thesis does not claim:

- physically exact reconstruction of the original painted scene;
- complete recovery of missing three-dimensional facial anatomy;

- a universally optimal bas-relief representation;
- demonstrated tactile recognition improvement;
- generalization across all paintings or artistic styles.

The experimental contribution is instead limited to:

- establishing and evaluating a controlled image-to-bas-relief pipeline;
- comparing alternative global geometric representations;
- investigating the integration of local external geometry;
- evaluating semantic and anatomy-guided local depth modification;
- analyzing independent normal, landmark, curvature, and mesh cues;
- identifying practical trade-offs and failure modes under a fixed experimental configuration.

This boundary is important because the experimental evaluation is primarily computational and Blender-based; no formal tactile user study is used to support perceptual claims.

3.18 Chapter Summary

This chapter reviewed the research most directly related to the problem investigated in this thesis. Classical bas-relief methods provide strategies for controlling global and local geometry during strong depth compression, but they generally assume that suitable source geometry is already available. Detail-preserving and normal-based methods can preserve or manipulate local structure, yet they cannot independently reconstruct geometric information that is absent from the input.

Portrait and human-reconstruction methods introduce class-specific anatomical priors but depend on alignment, template compatibility, and assumptions that may be difficult to satisfy for pictorial art. Modern monocular depth models provide direct image-aligned geometry from a single image, while feed-forward reconstruction systems offer richer point- and camera-based geometric representations. Semantic approaches provide spatial localization, and surface normals offer complementary information about local orientation.

The reviewed literature therefore provides a set of complementary geometric tools, but no single approach simultaneously satisfies the requirements of stable global geometry, relief-specific depth compression, controlled local refinement, and compatibility with pictorial input. This gap motivates the experimental strategy of this thesis: first, a stable image-based baseline is established; then, increasingly localized geometric information is introduced to evaluate its effect on the final shallow surface and its compatibility with the global representation.

The controlled experimental investigation begins in Chapter 4.

Chapter 4

Establishment of the Depth-Driven Baseline

4.1 Introduction

A stable geometric reference is required before any local refinement can be evaluated within the experimental investigation.

This is particularly important because the later stages of the thesis employ different sources of geometric information, including external 3D reconstruction, semantic masks, parametric anatomical priors, surface normals, landmarks, curvature, and explicit mesh-space processing.

Without a fixed baseline, observed differences between later experimental runs may be attributed to changes in the underlying global geometry rather than the local refinement strategy under study.

The objective of this chapter is therefore to define the depth-driven baseline used in the subsequent experimental work.

The process consists of five principal stages:

1. comparison of candidate monocular depth representations;
2. selection of a stable global depth source;
3. alignment to the common working grid and custom PPD-inspired depth refinement;
4. relief-oriented normalization and remapping;
5. validation through a fixed Blender displacement pipeline.

Three representative monocular depth approaches were considered during baseline selection:

- Depth Anything 3 (DA3) [18];
- Marigold [16];
- Metric3D v2 [13].

The models were not evaluated with the aim of establishing a general benchmark of monocular depth estimation; instead, they were studied specifically in relation to the requirements of the present shallow-relief pipeline.

The selected baseline had to provide sufficiently stable global geometry while preserving the spatial arrangement of the original painting and remaining compatible with subsequent local processing.

The chapter also explicitly distinguishes two representations that are used throughout the rest of the thesis:

$$\boxed{\text{DA3 Raw} \neq \text{DA3+Remap}} \tag{4.1}$$

DA3 Raw refers to the direct depth prediction obtained from the selected DA3 configuration.

DA3+Remap refers to the processed depth representation obtained after alignment to the common working grid, custom PPD-inspired depth refinement, normalization, distribution adjustment, and relief-oriented depth remapping.

Unless otherwise stated, the term *baseline* in the subsequent experimental chapters refers to the DA3+Remap representation.

4.2 Baseline Requirements

The baseline should preserve a consistent representation of the overall spatial organization of the painting.

The depth representation should remain suitable after strong compression and conversion into a shallow height field for the target application.

The following requirements were therefore taken into account.

4.2.1 Global Geometric Stability

The baseline should preserve the broad spatial organization of the painting, including:

- foreground–background separation;
- relative placement of the principal subjects;
- large-scale body geometry;
- broad surface orientation;
- overall composition.

This requirement is important because in later experiments we intentionally preserve the global surface while modifying only selected local regions.

4.2.2 Image Alignment

Each depth value is expected to correspond directly to the image location.

This allows for a direct transfer of semantic masks, facial landmarks, local support functions and other image-derived constraints to the depth domain.

4.2.3 Local Structural Coherence

The baseline should contain sufficient coherent local geometry to support subsequent refinement. It is not required to reconstruct every facial feature perfectly; however, critical regions should remain sufficiently stable to permit controlled local modification.

4.2.4 Compatibility with a Height Field

The final bas-relief is represented as a single-valued displacement surface.

The selected geometry should therefore be convertible into

$$H(x, y), \tag{4.2}$$

without requiring complex visibility reasoning or multiple depth values at the same image location.

4.2.5 Post-Processing Stability

The depth representation should tolerate:

- normalization;
- percentile adjustment;
- nonlinear remapping;
- local masking;
- high-precision export;
- displacement-based reconstruction.

This criterion was especially important because the raw monocular prediction was not meant to be used directly as the final relief.

4.3 Input Artwork

The principal experimental input is Caravaggio’s *David with the Head of Goliath*.

The working RGB image has a resolution of

$$3000 \times 2333 \tag{4.3}$$

pixels.

The painting provides a challenging case for monocular geometry estimation because it contains:

- strong chiaroscuro;
- multiple overlapping anatomical regions;
- large dark image areas;

- facial structures at different scales;
- painterly rather than strictly photographic appearance;
- complex foreground–background relationships.

Two facial regions became particularly important during later analysis:

- David’s face;
- Goliath’s head.

The baseline therefore needed to provide both coherent global scene organization and sufficient local stability in these regions.



Figure 4.1: Input painting used throughout the controlled experimental investigation: Caravaggio’s *David with the Head of Goliath*.

4.4 Candidate Monocular Depth Models

Three representative monocular depth approaches were examined before establishing the working baseline. The comparison was not intended to reproduce the original evaluation procedures of the individual models; instead, their outputs were analyzed according to the requirements of the common bas-relief pipeline.

4.4.1 Depth Anything 3

Depth Anything 3 was investigated as a general-purpose source of dense image-aligned geometry [18].

For the present application, several properties were particularly relevant:

- dense image-derived geometric prediction;
- preservation of the image coordinate system;
- compatibility with single-image inference;
- straightforward conversion into a height field.

The direct model output is referred to throughout the thesis as

$$D_{\text{DA3,raw}}. \quad (4.4)$$

This raw representation provides the initial global geometry but is not yet the final working relief baseline.

4.4.2 Marigold

Marigold was assessed as a diffusion-based alternative for monocular depth estimation [16].

Compared to direct feed-forward depth prediction, its formulation estimates monocular depth via a diffusion-based inference process.

For the current study, Marigold provided an alternative diffusion-based depth representation that could be examined within the same relief-oriented pipeline.

The comparison focused particularly on:

- broad scene organization;
- local facial structure;
- foreground separation;
- behaviour in dark regions;
- suitability for shallow depth compression.

4.4.3 Metric3D v2

Metric3D v2 was also examined because it provides a geometry-oriented monocular formulation with metric-depth capability and complementary surface-normal information [13].

These outputs made the model relevant as an additional geometric reference for the comparison.

However, metric reconstruction is not the objective of this thesis.

The main question was whether the resulting image-aligned geometry provided a stable basis for:

$$\text{depth} \rightarrow \text{compression} \rightarrow \text{displacement} \rightarrow \text{shallow relief}. \quad (4.5)$$

Metric3D v2 was therefore evaluated according to the same relief-oriented criteria as the other candidate models.

4.5 Qualitative Comparison of Candidate Depth Representations

The three candidate models produced different geometric interpretations of the same pictorial input, which were examined at both global and local scales.

The comparison considered:

- global subject separation;
- depth ordering;
- continuity of the human figures;
- stability of dark regions;
- local facial organization;
- preservation of image-aligned boundaries;
- suitability for subsequent remapping.

This comparison was conducted on an application-specific basis.

A depth map may perform well on a traditional monocular-depth benchmark, but still be less useful for shallow-relief production.

In contrast, a representation in which the physically meaningful metric scale is missing may still be useful if its relative geometry is sufficiently stable to be redistributed in the available relief height.



Figure 4.2: Representative depth outputs produced by the three candidate monocular approaches for the same input painting. The comparison is used to examine their suitability as global geometric inputs to the shallow-relief pipeline rather than as a general benchmark of monocular depth accuracy.

4.5.1 Global Scene Structure

At the global level, the principal question was whether the depth representation maintained a stable organization of the overall composition.

A useful baseline had to separate the major subjects and broad scene layers without generating severe discontinuities that would later dominate the relief.

The candidate models differed in the distribution of their predicted depth and in the degree of separation between individual scene structures. These differences became particularly important after normalization, because the complete predicted depth range had to be compressed into a much smaller relief interval.

4.5.2 Local Facial Structure

The facial regions provided a more demanding comparison.

A globally stable depth map could still exhibit:

- weak nose projection;
- limited geometric separation in the eye region;
- limited lip separation;
- broadly smoothed cheek surfaces.

This observation was important for the subsequent experimental design. The baseline was not expected to reconstruct complete facial anatomy; rather, it needed to provide a sufficiently stable facial surface on which controlled local refinement could subsequently be evaluated.

4.6 Application-Specific Model Comparison

Table 4.1 summarizes the application-specific comparison.

The entries should be interpreted as observations for the selected artwork and the target pipeline rather than as general rankings of the three models.

4.7 Selection of Depth Anything 3 as the Global Depth Source

DA3 was selected as the principal global depth source based on the application-specific qualitative comparison and was retained throughout the subsequent experimental investigation.

This decision should not be taken as a claim that DA3 is universally better than Marigold, Metric3D v2 or other monocular depth models.

The selection was specific to:

- the chosen painting;
- the desired image-aligned representation;
- the fixed-viewpoint shallow-relief objective;
- the subsequent local-refinement pipeline.

DA3 was retained primarily because its global representation was sufficiently stable for the selected artwork and processing pipeline.

Subsequent experiments required a reference surface that could be held fixed while local modifications were introduced. Keeping the global depth source fixed across the experimental

Table 4.1: Application-specific comparison of the candidate monocular depth representations. The comparison concerns suitability for the shallow-relief pipeline used in this thesis and should not be interpreted as a general benchmark of the models.

Model	Relevant strengths	Observed limitations	Role in this thesis
Depth Anything 3	Image-aligned global geometry with relative scene organization suitable for the subsequent processing pipeline.	Several local facial structures remained comparatively weak after direct prediction, motivating relief-oriented redistribution and later refinement.	Selected as the principal global depth source.
Marigold	Provides a diffusion-based monocular depth representation that can be processed within the same relief-oriented pipeline.	For the selected artwork and evaluation criteria, the observed output did not provide a clear application-specific reason to replace DA3 as the working global reference.	Retained as an alternative depth candidate for baseline comparison.
Metric3D v2	Provides geometry-oriented depth estimation together with complementary normal information.	Its metric-depth objective was not itself an application-specific advantage for the strongly compressed shallow-relief representation considered here.	Used as an alternative geometric reference; its normal-related capabilities also motivated later geometry-aware analysis.

phases therefore made it easier to attribute observed changes to the local refinement strategy itself.

Thus, DA3 was retained as the fixed global depth reference for the remainder of the experimental investigation.

4.8 DA3 Configuration Study

After DA3 was selected as the global depth source, several processing choices were examined before fixing the working baseline.

These included:

- inference resolution;
- depth orientation;
- alignment to the common working grid;
- custom PPD-inspired depth refinement;
- normalization;
- percentile range;
- nonlinear compression;
- high-precision export.

The aim was not to change local anatomy at this point.

All processing was kept global so that the baseline could be established prior to the introduction of local semantic or anatomical interventions.

4.9 Inference Resolution Analysis

The spatial resolution of the monocular inference affects both the computational cost and the amount of local structure in the predicted depth field.

Inference at lower resolutions may fail to preserve narrow image structures.

However, increasing the inference resolution does not guarantee that additional image detail corresponds to valid geometric detail.

The resolution study therefore considered the balance between:

- global stability;
- local continuity;
- preservation of important boundaries;
- computational requirements;
- compatibility with later processing.

The selected configuration was retained consistently throughout the subsequent experiments.

4.10 Pre-Alignment and Custom PPD-Inspired Depth Refinement

After the 1536 processing-resolution configuration was selected, its DA3 depth representation was aligned to the common 1022×798 working grid associated with the 1024-reference representation. The resulting pre-aligned depth was then processed using the custom PPD-inspired multistage refinement before relief-oriented remapping.

The term *PPD-inspired* denotes the custom deterministic depth-domain post-processing used in this thesis and should not be interpreted as execution of the complete Pixel-Perfect Depth method. The refinement stabilizes the numerical range, removes small isolated support components, preserves strong depth boundaries during smoothing, and regularizes relatively flat regions. Its exact six-stage implementation and parameters are reported in Appendix A.1.1.

This stage remained global and did not introduce semantic or anatomy-specific modification. Its role was to prepare the selected DA3 representation for the subsequent relief-oriented depth remapping while retaining the common image-aligned working grid.

4.11 Depth Normalization

The PPD-refined DA3 representation cannot be interpreted directly as the final physical relief height. The next processing stage therefore normalizes the refined depth representation.

A basic min-max normalization can be written as

$$D_n(x, y) = \frac{D(x, y) - D_{\min}}{D_{\max} - D_{\min}}. \quad (4.6)$$

However, global extrema can be strongly influenced by a small number of pixels.

An alternative that reduces the direct dependence on the global minimum and maximum is the use of percentile-based limits.

Let

$$P_l = \text{percentile}(D, p_l) \quad (4.7)$$

and

$$P_h = \text{percentile}(D, p_h). \quad (4.8)$$

The clipped depth is

$$D_c = \text{clip}(D, P_l, P_h), \quad (4.9)$$

and the normalized representation becomes

$$D_p = \frac{D_c - P_l}{P_h - P_l}. \quad (4.10)$$

This operation reduces the influence of extreme values and provides greater control over the effective depth interval.

4.12 Relief-Oriented Depth Remapping

Normalization does not solve the problem of depth allocation.

A painting can exhibit large scene-level depth variation and very small local variation.

When the normalized depth is linearly mapped into the available relief height, large global differences can dominate the physical surface.

A nonlinear transformation was therefore applied:

$$D_r = T(D_p), \quad (4.11)$$

where T redistributes the normalized depth values.

The purpose of this transformation was not to reconstruct additional 3D geometry, but to redistribute the existing depth so that the overall scene could be represented more effectively within a shallow displacement range.

The resulting representation is referred to as:

$$\boxed{D_{\text{baseline}} = D_{\text{DA3+Remap}}} \quad (4.12)$$

This distinction is maintained throughout the thesis: the raw DA3 prediction serves as a reference for the original model output, whereas the remapped representation provides the baseline against which subsequent local-refinement experiments are evaluated.

4.13 Raw DA3 and DA3+Remap

The difference between the two representations can be summarized as follows.

Table 4.2: Distinction between the raw DA3 prediction and the working DA3+Remap baseline.

Property	DA3 Raw	DA3+Remap
Source	Direct DA3 prediction	DA3 prediction after pre-alignment, custom PPD-inspired refinement, and relief-oriented remapping
Primary objective	General image-derived geometry	Stable shallow-relief baseline
Depth distribution	Original model distribution	Redistributed depth range
Local anatomy	Determined directly by DA3	No explicit anatomical reconstruction; existing geometry is redistributed
Use in later experiments	Reference / diagnostic input	Principal experimental baseline
Blender role	Used for baseline comparison	Used as the fixed reference for subsequent refinement



(a) DA3 Raw.



(b) DA3+Remap.

Figure 4.3: Comparison between the direct DA3 prediction and the relief-oriented remapped representation used as the working baseline.

4.14 High-Precision Depth Representation

The processed depth map was exported using a 16-bit representation.

For normalized depth

$$D_r \in [0, 1], \quad (4.13)$$

the corresponding unsigned 16-bit representation can be written as

$$D_{16} = \text{round}(65535 D_r). \quad (4.14)$$

Using a 16-bit representation provides

$$2^{16} = 65536 \quad (4.15)$$

possible quantization levels.

This higher numerical resolution is useful because the final relief may contain small local depth differences that would be represented less accurately with an 8-bit map.

Therefore, 16-bit TIFF was adopted as the standard exchange format between the depth-processing pipeline and Blender.

4.15 Controlled Blender Reconstruction

A fixed Blender reconstruction pipeline was set up for the evaluation of candidate depth maps as actual shallow surfaces.

The depth representation was used as a displacement map on a common surface.

Conceptually, the displaced vertex position can be written as

$$\mathbf{v}' = \mathbf{v} + sD(\mathbf{u})\mathbf{n}, \quad (4.16)$$

where:

- $\mathbf{v}$ is the original mesh vertex;
- $\mathbf{u}$ is the corresponding image coordinate;
- $D(\mathbf{u})$ is the depth-map value;
- $\mathbf{n}$ is the displacement direction;
- s is the fixed displacement scale.

The same reconstruction configuration was retained across comparative experiments whenever possible. The fixed configuration included:

- mesh topology;
- displacement configuration;
- camera configuration;
- viewpoint;
- lighting;
- rendering conditions.

A common configuration had to be kept to avoid misinterpreting changes in the Blender environment as improvements in the underlying depth geometry.

4.16 Blender Validation of the Remapped Baseline

The DA3 Raw and DA3+Remap representations were reconstructed under the same Blender configuration.

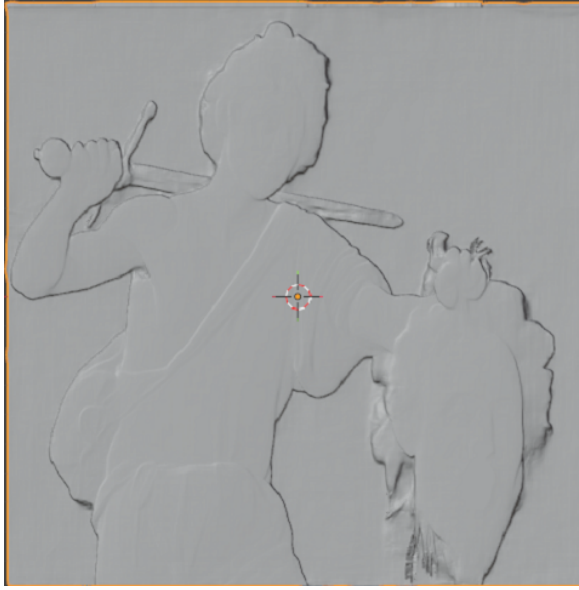
The comparison made it possible to examine the effect of the complete global baseline- preparation pipeline, including pre-alignment, custom PPD-inspired refinement, and relief-oriented depth redistribution, without introducing any new local anatomical geometry.

Under the fixed Blender configuration, the remapped representation redistributed depth variation within the restricted relief range while preserving the broad composition of the original DA3 prediction.

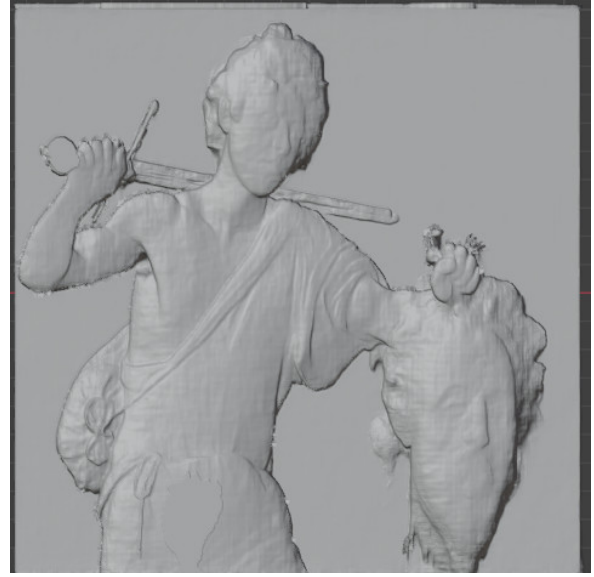
This result was therefore adopted as the fixed baseline for the subsequent experimental phases. The remapping step should not, however, be interpreted as a solution to the problem of local facial geometry, because several facial structures remained weak. Its significance was instead to provide a stable global surface from which these local limitations could be investigated systematically.

4.17 Local Analysis of the Baseline

Once the global baseline was established, the two principal facial regions were examined in greater detail. The purpose of this analysis was not to reassess the validity of the baseline, but to identify the local geometric constraints that motivated the subsequent experimental chapters.



(a) DA3 Raw.



(b) DA3+Remap.

Figure 4.4: Controlled Blender comparison of the raw DA3 depth and the DA3+Remap representation under the same reconstruction configuration.

4.17.1 David’s Face

David’s face occupies a relatively small region of the complete image.

The DA3+Remap baseline maintained the overall facial position and broad volumetric structure present in the global depth representation.

However, several internal structures remained comparatively subtle, including:

- nose projection;
- nose–cheek transition;
- eye-socket structure;
- lip separation;
- chin transition.

These limitations motivated subsequent experiments with semantic refinement, anatomical supports, landmarks and surface normal information.

4.17.2 Goliath’s Head

Goliath’s head differs from David’s face in scale, orientation, and appearance.

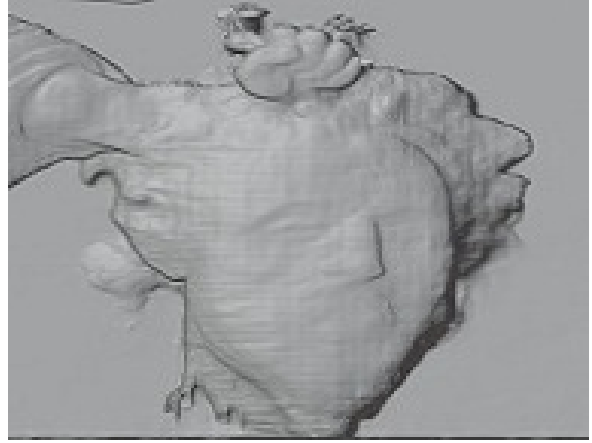
The baseline maintained its broad head shape and overall volumetric organization.

However, in an area of strong painterly contrast and complex surface appearance, some local facial structures were still weak.

Thus this region provided another test of whether a local refinement strategy could work under different geometric and appearance conditions without a change in the global baseline.



(a) David's facial region.



(b) Goliath's head region.

Figure 4.5: Close-up Blender views of the David and Goliath facial regions reconstructed from the DA3+Remap baseline. The local geometric representation of these regions motivated the subsequent investigation of localized refinement strategies.

4.18 Baseline Strengths

The selected baseline offered several practical advantages for the experimental investigation. First, it preserved one-to-one pixel-to-surface correspondence with the source image, allowing masks, landmarks, and support functions to be defined in image coordinates and transferred directly to the depth representation. Second, the global geometry provided a consistent reference for subsequent comparative experiments. Third, the height-field representation remained directly compatible with the fixed Blender displacement pipeline.

The 16-bit representation also provided substantially finer numerical quantization than an 8-bit representation for encoding small depth variations. Finally, the baseline remained sufficiently simple and interpretable for local modifications to be measured directly relative to the fixed DA3+Remap reference.

4.19 Baseline Constraints and Implications

The selected baseline nevertheless retained several constraints. The monocular depth remained an inferred representation rather than physically verified three-dimensional geometry, and some local facial structures remained weak after relief-oriented remapping. The predicted geometry could also be influenced by painterly shading and image texture. Moreover, the remapping stage redistributed existing geometric information rather than recovering missing anatomical structure, while the height-field representation could not encode multiple depths at the same image location.

These constraints were deliberately accepted because the baseline was not intended to solve every aspect of the target problem. Its role was to provide a stable global reference against which additional local geometric strategies could be evaluated.

4.20 Definition of the Fixed Experimental Baseline

At the end of this stage, the following components were fixed for the subsequent experiments:

- source painting;
- DA3 as the principal global depth source;
- alignment to the common working grid;
- custom PPD-inspired depth refinement;
- relief-oriented DA3 depth remapping;
- working depth-map dimensions;
- 16-bit TIFF output;
- image-to-depth correspondence;
- global composition;
- Blender mesh topology;
- displacement configuration;
- camera and viewpoint;
- representative facial regions.

The working baseline is therefore defined as:

$$\boxed{\begin{aligned} \text{Baseline} = & \text{DA3} + \text{Pre-Alignment} \\ & + \text{Custom PPD-Inspired Refinement} + \text{Relief-Oriented Remapping} \end{aligned}} \quad (4.17)$$

or, in the terminology used throughout the experimental chapters:

$$\boxed{\text{Baseline} = \text{DA3} + \text{Remap}.} \quad (4.18)$$

Subsequent experiments modify this representation locally while attempting to preserve its global geometry.

4.21 Chapter Summary

This chapter established the fixed depth-driven baseline used throughout the subsequent experimental investigation. Three representative monocular depth approaches—Depth Anything 3, Marigold, and Metric3D v2—were considered under application-specific criteria related to image alignment, global geometric stability, local structural coherence, and compatibility with shallow-relief reconstruction.

DA3 was retained as the principal global depth source for the selected artwork and processing pipeline. Its selected 1536-resolution prediction was aligned to the common working grid, processed by the custom PPD-inspired multistage refinement, and then subjected to relief-oriented normalization and remapping, producing the DA3+Remap representation. This processing prepared and redistributed the available depth range without introducing explicit new anatomical

geometry, and the result was exported using a 16-bit representation and evaluated under a fixed Blender displacement configuration.

Although the resulting baseline preserved the broad composition and provided a stable image-aligned reference, several local facial structures remained weak. These local constraints motivated the subsequent refinement stages. From this point onward, DA3+Remap is therefore treated as the fixed global reference against which additional local geometric information is evaluated.

Chapter 5

Hybrid Geometry Fusion and External 3D Priors

5.1 Introduction

Chapter 4 established DA3+Remap as the fixed global geometric baseline for the subsequent experimental investigation.

The baseline retained the general composition and overall geometry of the scene, but several local structures, especially in the facial areas, were still geometrically compressed.

This observation motivated the first major refinement strategy studied in the thesis:

Can a stable global monocular depth representation be combined with richer local three-dimensional geometry?

The motivation was the complementary nature of the available representations. DA3 provided a stable global geometry aligned with the source image, while external 3D reconstruction could potentially provide richer local volumetric information.

The initial hypothesis was therefore:

$$\text{stable global depth} + \text{richer local geometry} \rightarrow \text{improved hybrid relief.} \quad (5.1)$$

In practice, however, the two geometric representations were created independently and did not naturally share the same coordinate system, scale, projection, or region of interest.

The experiments in this chapter therefore progressed from direct fusion toward registration-aware and residual-based integration.

The chapter outlines the experimental evolution from the first hybrid feasibility tests to the more controlled geometry-fusion experiments.

The main outcome of this phase was not the identification of a universally effective fusion technique.

Instead, the experiments identified the conditions under which external geometry becomes difficult to incorporate into a fixed shallow height-field representation.

5.2 Objective and Experimental Hypothesis

The main goal of this stage was to evaluate whether local three-dimensional geometry could compensate for weak local volumetric structure in the DA3+Remap baseline without displacing or destabilizing the global scene geometry.

The baseline was therefore kept fixed:

$$D_B(x, y) = D_{\text{DA3+Remap}}(x, y). \quad (5.2)$$

An external local geometry prior is denoted by

$$D_L(x, y). \quad (5.3)$$

One simplified additive formulation for introducing local external geometry can be expressed as

$$D_H(x, y) = D_B(x, y) + \alpha M(x, y) D_L(x, y), \quad (5.4)$$

where:

- D_H is the hybrid depth representation;
- D_B is the fixed DA3+Remap baseline;
- D_L represents the local external geometry;
- M restricts the modification to the target region;
- α controls the contribution of the external geometry.

This intentionally simplified formulation highlights the sensitivity of direct fusion to the spatial and geometric correspondence between the two representations.

5.3 Fixed Inputs and External Geometry

The global input remained unchanged from Chapter 4.

The principal fixed components were:

- the original Caravaggio painting;
- the DA3+Remap baseline;
- the image coordinate system;
- the target facial regions;
- the fixed Blender reconstruction configuration.

The principal external geometry used in the initial hybrid experiments was obtained with Pixa3D [17]. The method provides an image-to-3D representation intended to preserve correspondence between image observations and generated three-dimensional geometry.

In this thesis, Pixa3D was used only as an external source of local geometric information. Its output was not treated as ground-truth geometry, and no assumption was made that it provided an anatomically correct reconstruction of the painted subjects. Instead, the hybrid experiments investigated whether geometric information obtained from this independent reconstruction could be aligned with and selectively transferred to the stable DA3+Remap baseline.

Subsequent experiments also used EXR representations of the reconstructed geometry for local depth analysis and transfer.

The two principal target regions remained:

- David’s face;
- Goliath’s head.

These regions were selected because the baseline provided stable broad geometry while leaving several local facial structures comparatively weak.

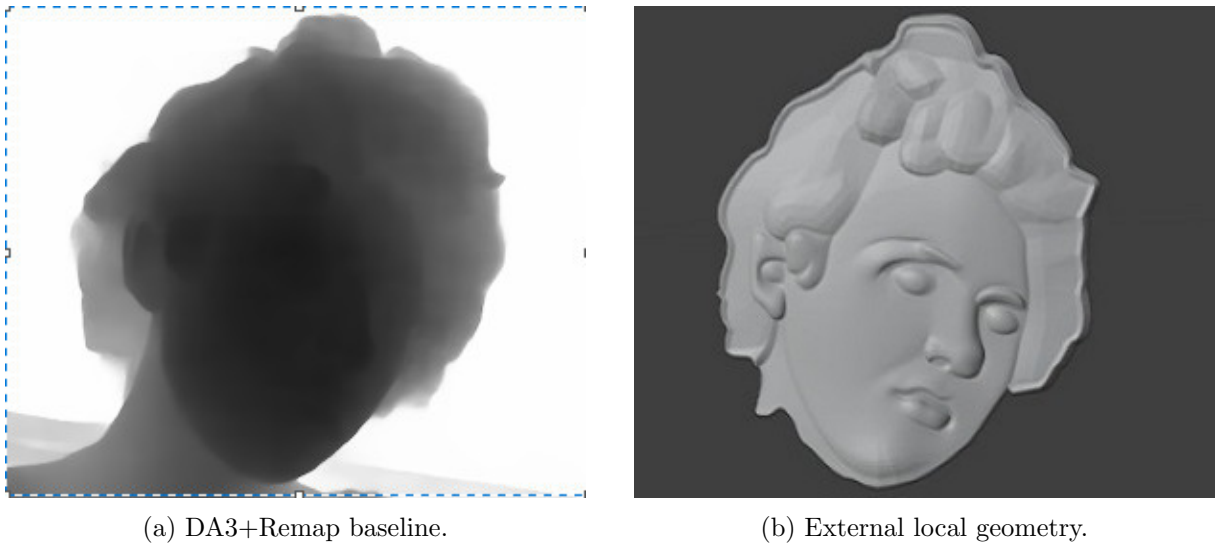


Figure 5.1: Representative global baseline and external local geometry considered during the hybrid reconstruction experiments. The two representations were generated independently and therefore required registration before controlled fusion.

5.4 Why Direct Geometry Fusion Is Difficult

The initial experiments showed that the challenge was not simply the availability of local three-dimensional information. The DA3 baseline and the external reconstruction were defined in different geometric representations and were not directly compatible. Four principal types of mismatch were identified.

5.4.1 Scale Mismatch

The reconstructed local geometry did not necessarily cover the same spatial scale as the corresponding region in the DA3 depth map.

A local coordinate transformation therefore required a scale factor

$$s \tag{5.5}$$

such that

$$\mathbf{p}' = s\mathbf{p}. \tag{5.6}$$

5.4.2 Translation Mismatch

The reconstructed geometry could still be displaced relative to the intended facial region even after approximate scaling.

Translation therefore had to be estimated:

$$\mathbf{p}'' = \mathbf{p}' + \mathbf{t}. \tag{5.7}$$

5.4.3 Region-of-Interest Mismatch

The support of the reconstructed geometry did not necessarily coincide exactly with the desired semantic region for modification.

This resulted in cases where valid reconstructed geometry extended outside the desired region or did not cover parts of the desired region.

5.4.4 Projection Mismatch

The most challenging mismatch was related to projection compatibility.

A reconstructed three-dimensional surface may correspond to a camera model or projection that differs from that implied by the original painting.

Consequently, a plausible 3D face does not necessarily remain compatible with the painted facial structure when converted into an image-aligned depth map.

The registration problem can therefore be summarized as:

$$\boxed{\text{registration} = \text{scale} + \text{translation} + \text{ROI correspondence} + \text{projection compatibility}.} \tag{5.8}$$

5.5 Preliminary Hybrid Exploration: Exp28–Exp31

The first set of experiments investigated the computational feasibility of hybrid reconstruction, prior to the implementation of more advanced registration procedures.

5.5.1 Exp28: Baseline Reference

Exp28 established the remapped DA3 depth as the reference representation.

The baseline provided:

- scene-wide composition;
- subject positioning;
- broad depth continuity.

At the same time, facial regions remained comparatively compressed in the depth direction. Exp28 therefore served as the fixed reference rather than as a local refinement experiment.

5.5.2 Exp29: Initial DA3 + Pixal3D Fusion

Exp29 introduced the first direct combination of the DA3 baseline and Pixal3D-derived local geometry.

The primary objective was to test computational feasibility. Although the two representations could be combined, the resulting surface revealed substantial correspondence problems.

The presence of additional three-dimensional geometry did not guarantee correct position or scale relative to the painting.

5.5.3 Exp30: Alignment-Aware Fusion

Exp30 introduced a more controlled fusion procedure.

Valid local geometry was extracted and then inserted into the baseline representation.

This reduced unintended interaction between the external reconstruction and unrelated image regions.

However, the observed correspondence errors led to the explicit treatment of alignment as a dedicated processing stage rather than a minor preprocessing step.

5.5.4 Exp31: Semantic Mask Refinement

Exp31 improved the spatial support used for local fusion, with particular attention to Goliath’s head.

The refined mask improved the spatial definition of the target region.

However, better semantic localization did not automatically resolve scale or projection mismatch in the external geometry.

This distinction became important later in the thesis:

$$\text{correct ROI} \neq \text{correct local geometry.} \quad (5.9)$$

5.6 Registration Refinement: Exp32–Exp35

The next step after the first fusion experiments was explicitly devoted to registration.

The aim was to determine whether improved geometric correspondence would make the external geometry more useful for local refinement.

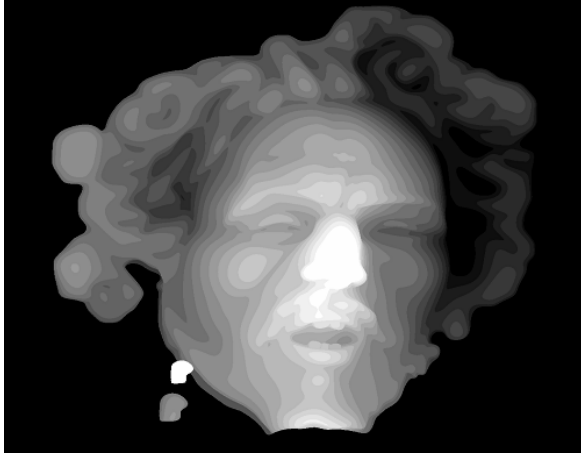
5.6.1 Exp32: Bounding-Box Alignment

Bounding-box alignment utilized the size of the target region and the reconstructed geometry to compute an initial spatial transformation.

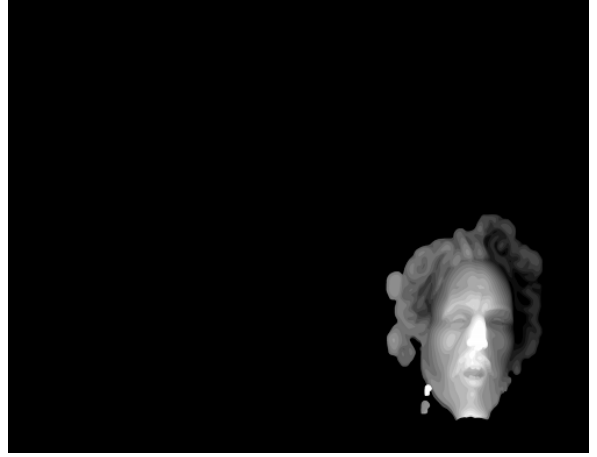
If the source geometry occupies a bounding box

$$B_s = (x_s, y_s, w_s, h_s) \quad (5.10)$$

and the target region occupies



(a) Initial correspondence.



(b) Alignment-aware placement.

Figure 5.2: Representative alignment diagnostic from the preliminary hybrid experiments. The experiments identified spatial registration as a principal limitation of direct external-geometry fusion.

$$B_t = (x_t, y_t, w_t, h_t), \quad (5.11)$$

approximate scaling can be estimated from

$$s_x = \frac{w_t}{w_s}, \quad s_y = \frac{h_t}{h_s}. \quad (5.12)$$

This improved coarse spatial correspondence but did not guarantee anatomical alignment.

5.6.2 Exp33: Centroid Alignment

The next experiment aligned the center of the reconstructed region with the center of the target support.

Given source and target centroids

$$\mathbf{c}_s \quad \text{and} \quad \mathbf{c}_t, \quad (5.13)$$

the required translation can be approximated as

$$\mathbf{t} = \mathbf{c}_t - \mathbf{c}_s. \quad (5.14)$$

This reduced gross translation error but again could not compensate for projection differences.

5.6.3 Exp34: Scale Normalization

Exp34 focused more directly on scale consistency.

The objective was to prevent unrealistic local volume caused solely by differences between the numerical depth ranges of the external geometry and the baseline.

The local prior was normalized before fusion.

Conceptually:

$$\hat{D}_L = \frac{D_L - \min(D_L)}{\max(D_L) - \min(D_L)}. \quad (5.15)$$

The normalized geometry could then be mapped into a controlled local range.

5.6.4 Exp35: Manual Refinement

Following the automatic alignment steps, manual adjustment was introduced.

This step helped determine whether the remaining limitation was primarily caused by registration error or by incompatibility in the external geometry itself.

Manual refinement provided greater control over spatial correspondence. However, even visually acceptable alignment did not necessarily produce a more compatible final bas-relief. This observation marked an important turning point in the hybrid-geometry investigation.

5.7 From Registration to Geometry Fusion

After the registration experiments, the focus shifted from whether the external geometry could be aligned to whether the aligned geometry provided useful local surface information after shallow-relief reconstruction. Even a geometrically aligned prior could still contain incompatible proportions, projection-dependent distortions, excessive depth range, unsupported surface structure, or details that became weak after shallow compression. Several fusion strategies were therefore investigated rather than directly replacing the baseline surface.

5.8 External Geometry Representation and EXR Processing

External geometry was converted into depth representations amenable to local analysis and fusion.

EXR files were particularly useful because they preserved floating-point depth information before final quantization.

The geometry-processing stage included:

- extraction of valid geometry;
- removal of invalid or background regions;
- normalization;
- spatial alignment;
- local depth-range adjustment;
- conversion into a baseline-compatible representation.

These operations allowed the external geometry to be analyzed independently before being introduced into the complete depth representation.

5.9 Fusion Strategies

Several strategies were evaluated for transferring information from the external geometry into the baseline.

5.9.1 Direct Local Injection

The simplest strategy introduced normalized external geometry directly inside the target region:

$$D_H = (1 - M)D_B + MD_L. \quad (5.16)$$

where $M(x, y) \in [0, 1]$ denotes the local fusion mask. For a binary mask, the formulation corresponds to direct local replacement, while intermediate mask values provide spatial blending near the transition region.

This approach directly replaced or blended the baseline geometry within the selected region and therefore introduced a comparatively high risk of incompatibility with the surrounding baseline surface.

5.9.2 Residual-Based Injection

A more conservative approach used the external geometry as a residual rather than as a replacement.

A local residual can be written as

$$R_L = D_L - \mathcal{S}(D_L), \quad (5.17)$$

where $\mathcal{S}$ denotes a smoothing or low-frequency extraction operator.

The hybrid surface then becomes

$$D_H = D_B + \alpha MR_L. \quad (5.18)$$

This formulation attempts to preserve the broad DA3 geometry while transferring only selected local variation.

5.9.3 Curvature-Oriented Cues

Another approach focused on local surface variation rather than absolute external depth.

The motivation was that curvature-related information could encode local geometric transitions without requiring the complete reconstructed surface to replace the baseline.

5.9.4 Detail-Layer Transfer

Fine geometric components were transferred from the external prior to the baseline.

This approach further separated:

$$\text{global shape} \quad (5.19)$$

from

$$\text{local detail}. \quad (5.20)$$

The global shape remained primarily controlled by DA3, while the external reconstruction acted as a source of local geometric variation.

5.9.5 Local Volumetric Inflation

Later experiments explored stronger low-frequency local volume modifications.

Rather than simply conveying fine detail, additional volume was applied to the target face region.

The objective was to test whether stronger broad-scale modifications would survive the final shallow-relief reconstruction more effectively than fine-detail transfer.

5.10 Representative Fusion Experiments: Exp41–Exp47

The principal fusion strategies were evaluated through Exp41–Exp47.

5.10.1 Exp41: Clean EXR Fusion

Exp41 used the cleaned and aligned EXR geometry as a local fusion source. The experiment demonstrated that the preprocessed external geometry could be transferred into the baseline representation; however, visible geometric transfer alone was not sufficient to establish that the resulting surface was preferable to the baseline.

5.10.2 Exp42: Residual Injection

Exp42 reduced the influence of the broad external shape and focused more strongly on geometric residuals, providing a more conservative strategy in which the DA3 baseline retained control over the global surface.

5.10.3 Exp43: Curvature Cue Injection

Exp43 investigated curvature-related information to determine whether local geometric transitions could be strengthened without importing the complete external depth field.

5.10.4 Exp44: Detail-Layer Injection

Exp44 transferred a detail-oriented component of the external geometry to test whether fine local information could remain effective after integration into the shallow baseline.

5.10.5 Exp45: Local Volume Injection

Exp45 increased the magnitude of the local geometric contribution by introducing broader local volume rather than emphasizing only fine detail.

5.10.6 Exp46: Large-Scale Volume Inflation

Exp46 further emphasized broad local volumetric structure. In qualitative comparison, these low-frequency modifications remained more visually apparent after shallow-relief reconstruction than the finer external-detail transfers tested earlier. However, the resulting geometric differences remained limited and did not resolve the fundamental compatibility issue between the external reconstruction and the baseline.

5.10.7 Exp47: Aggressive Fusion

Exp47 further increased the influence of the external geometry. The experiment showed that greater fusion strength produced more apparent modification, but also increased the risk of distortion and incompatibility with the surrounding baseline.

5.11 Summary of the Hybrid Experiments

Table 5.1 summarizes the principal experiments in this chapter.

Table 5.1: Summary of the principal hybrid-geometry experiments.

Exp.	Strategy	Main purpose	Main observation
Exp28	DA3+Remap baseline	Establish the fixed global reference.	Consistent global composition and broad depth continuity; local facial volume remained limited.
Exp29	Direct DA3 + Pixal3D fusion	Test computational feasibility of hybrid geometry.	Fusion was computationally feasible, but substantial registration mismatch became evident.
Exp30	Alignment-aware fusion	Restrict external geometry to valid regions before placement.	Provided greater control over placement while exposing registration as a continuing bottleneck.
Exp31	Mask refinement	Improve semantic support, particularly around Goliath’s head.	Provided a more restricted ROI definition but did not solve geometric correspondence.
Exp32	Bounding-box alignment	Improve coarse spatial correspondence.	Reduced large placement mismatch but did not guarantee anatomical correspondence.
Exp33	Centroid alignment	Correct translation mismatch.	Improved positioning but projection differences remained.
Exp34	Scale normalization	Improve numerical and spatial scale compatibility.	Normalized the numerical depth range of the local prior before fusion.
Exp35	Manual refinement	Test whether additional registration could resolve remaining mismatch.	Provided additional manual alignment control, but better registration did not automatically produce a more compatible relief.

Continued on next page

Exp.	Strategy	Main purpose	Main observation
Exp41	Clean EXR fusion	Transfer cleaned external geometry.	Established geometric transfer after preprocessing, without establishing superiority over the baseline.
Exp42	Residual injection	Preserve baseline while transferring local variation.	Provided a more conservative alternative to direct geometry replacement.
Exp43	Curvature cue injection	Transfer local transition information.	Local geometric modification was possible, but the resulting local geometric effect remained limited after reconstruction.
Exp44	Detail-layer injection	Transfer fine-scale external structure.	Fine modifications did not consistently remain prominent after shallow reconstruction.
Exp45	Local volume injection	Increase broader facial volume.	Broader modification produced more visible geometric change.
Exp46	Large-scale volume inflation	Test stronger low-frequency local enhancement.	Produced one of the more stable results in this experimental family, although the resulting geometric difference remained limited.
Exp47	Aggressive fusion	Explore the upper limit of external-geometry influence.	Greater modification increased the risk of distortion and incompatibility with the baseline.

5.12 Representative Blender Comparison

Differences in depth space do not necessarily translate directly into comparable differences in the final shallow surface. Representative hybrid candidates were therefore reconstructed using the same Blender configuration as the DA3+Remap baseline.

Several fusion methods produced visible local depth modifications, but the final shallow surface attenuated many of the smaller-scale changes. Stronger low-frequency volumetric modifications remained more apparent after reconstruction, whereas excessive fusion increased the risk of local surface inconsistency.

Within this experimental family, Exp46 provided one of the more stable compromises between visible modification and baseline compatibility. This observation should not be interpreted as evidence of improved tactile perception; rather, Exp46 is treated as a representative, relatively stable configuration within the evaluated external-geometry fusion strategies.

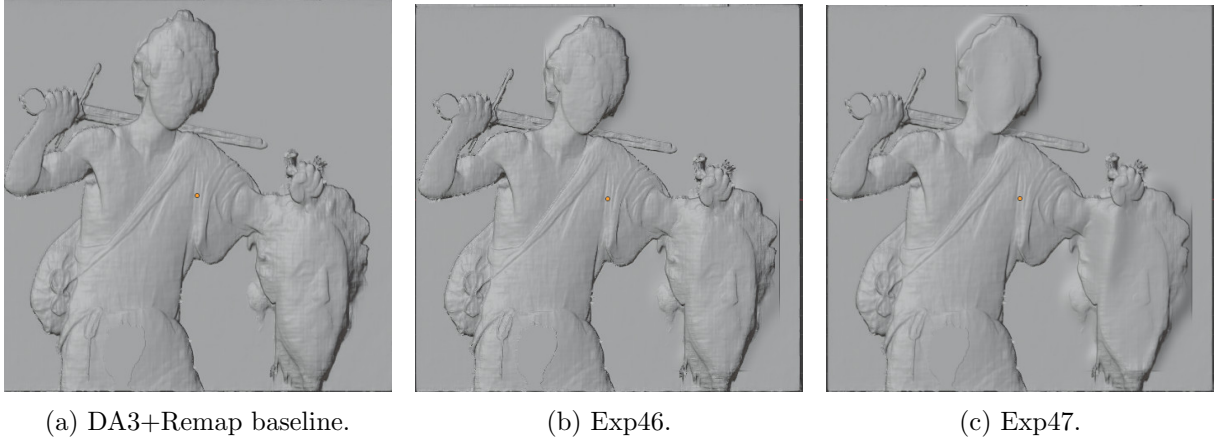


Figure 5.3: Representative Blender comparison of the DA3+Remap baseline and two late external-geometry fusion configurations. Exp46 provided one of the more stable results within this experimental family, while stronger fusion increased the risk of local geometric distortion.

5.13 What Improved

The hybrid-geometry experiments produced several methodological improvements. First, they established a processing sequence for extracting, normalizing, aligning, and transferring external local geometry into the DA3 depth space. Registration was then progressively refined through bounding-box alignment, centroid alignment, scale normalization, ROI refinement, and manual correction.

The experiments also showed that direct replacement was generally less controllable than selectively transferring components of the external geometry while preserving the baseline surface. In addition, broad volumetric modifications remained more visually apparent after shallow-relief reconstruction than finer external-detail transfers within this experimental family.

5.14 What Remained Unresolved

Despite improved alignment, external geometry did not provide a consistently superior local relief. This revealed an important distinction:

$$\boxed{\text{alignment quality} \neq \text{geometry quality} \neq \text{relief compatibility.}} \quad (5.21)$$

A prior may be spatially aligned yet remain geometrically incompatible with the painting, while a plausible three-dimensional reconstruction may contain structures that do not transfer effectively into a strongly compressed height field. The remaining issues included projection mismatch, uncertain correspondence between reconstructed and painted anatomy, inconsistent local proportions, weak survival of fine geometric detail, boundary artifacts under stronger fusion, and increasing distortion as the influence of external geometry increased.

The dominant bottleneck therefore shifted from registration to geometric compatibility as the experiments progressed.

5.15 Methodological Interpretation

The experiments yielded an important methodological result for the remainder of the thesis.

The presence of additional or more explicit 3D geometry does not imply that it should replace a stable image-derived baseline.

A useful local prior for the present problem should satisfy several conditions simultaneously:

1. it must correspond spatially to the intended region;
2. it must be compatible with the viewpoint of the painting;
3. its depth range must be compatible with the shallow relief;
4. its local geometry must remain consistent with the surrounding baseline;
5. its contribution must survive the final displacement reconstruction.

These requirements constrained direct geometry fusion and motivated a change in strategy. Rather than asking how complete external geometry could be inserted into the baseline, the subsequent investigation focused on how local regions could be modified while preserving the geometry already provided by the stable baseline. This shift led directly to semantic-aware local refinement.

5.16 Chapter Summary

This chapter investigated whether external three-dimensional geometry could be integrated into the fixed DA3+Remap baseline to improve local geometric structure. The preliminary experiments established that DA3-based depth and Pixa3D-derived geometry could be combined computationally, while also revealing substantial mismatch in scale, translation, region-of-interest correspondence, and projection.

Subsequent registration experiments improved spatial correspondence through bounding-box alignment, centroid alignment, scale normalization, and manual refinement. However, improved registration did not necessarily produce compatible local relief geometry. The investigation therefore shifted toward more conservative transfer strategies, including EXR-based fusion, residual injection, curvature cues, detail-layer transfer, and broader volumetric modification.

Among the evaluated configurations, Exp46 provided one of the more stable results within the external-geometry fusion family, although its geometric effect remained limited. The principal outcome of this phase was the distinction between spatial registration and geometric compatibility: even a well-aligned three-dimensional prior could remain inconsistent with the painted anatomy, source viewpoint, shallow depth range, or surrounding baseline.

These findings motivated the subsequent transition to semantic refinement, in which local regions are modified directly within the established DA3+Remap representation rather than replaced by independently reconstructed geometry.

Chapter 6

Semantic and Face-Aware Local Depth Refinement

6.1 Introduction

The external-geometry experiments presented in Chapter 5 showed that additional local three-dimensional information could be introduced into the DA3+Remap baseline, but also exposed persistent problems related to registration, projection compatibility, local scale, and geometric consistency. These limitations motivated a more conservative strategy in which the image-aligned DA3+Remap representation was retained and modified only within selected semantic regions.

The central hypothesis became:

$$\boxed{\text{stable global baseline} + \text{localized semantic modification}} \quad (6.1)$$

rather than

$$\boxed{\text{stable global baseline} + \text{external geometry replacement.}} \quad (6.2)$$

The principal research question addressed in this chapter is therefore:

Can semantic and face-aware local depth processing improve the geometric differentiation of important facial regions while preserving the global stability of the DA3+Remap baseline?

The investigation progressed from simple mask-guided enhancement toward increasingly controlled depth redistribution, boundary-safe processing, interior-region enhancement, Z-aware semantic compression, and finally camera-relative local reconstruction. All results in this chapter are interpreted geometrically; no formal tactile user study was conducted, and the experiments are therefore not presented as evidence of improved tactile recognition.

The experimental sequence covered four broad stages:

1. early semantic enhancement;
2. local face-aware enhancement;
3. semantic depth remapping and boundary-safe refinement;

4. Z-aware compression and local reconstruction.

6.2 Experimental Objective

The objective of this phase was to increase local geometric differentiation while retaining the established global baseline.

Let

$$D_B(x, y) \tag{6.3}$$

denote the DA3+Remap baseline.

A semantic region can be represented by a support function

$$M(x, y) \in [0, 1]. \tag{6.4}$$

A general conceptual formulation of a local additive modification can be expressed as

$$D'(x, y) = D_B(x, y) + M(x, y)E(x, y), \tag{6.5}$$

where $E(x, y)$ represents a local enhancement function.

This formulation provides an important property:

$$M(x, y) = 0 \quad \Rightarrow \quad D'(x, y) = D_B(x, y) \tag{6.6}$$

Under this conceptual additive formulation, pixels for which $M(x, y) = 0$ are not modified by the local enhancement term. Some later variants additionally included global normalization or filtering, so exact pixelwise equality outside the semantic support was not guaranteed in every exported result. The design objective nevertheless remained to localize the intended geometric modification while retaining the DA3-derived representation as the global reference.

6.3 Semantic Supports

The principal semantic supports represented David’s face and Goliath’s head, while additional foreground information was used in experiments addressing broader foreground–background organization. The exact mask definitions evolved during the experimental sequence, but their role remained the same: they were treated as spatial constraints rather than geometric models. They specified where modification was allowed, but did not determine how the anatomy inside the selected region should be shaped.

$$\boxed{\text{semantic mask} = \text{where to modify}} \tag{6.7}$$

but

$$\boxed{\text{semantic mask} \neq \text{how the anatomy should be shaped.}} \tag{6.8}$$

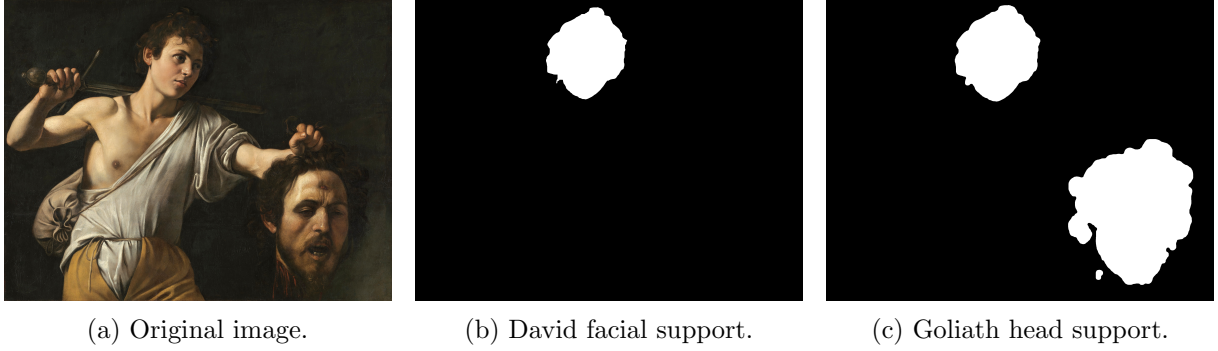


Figure 6.1: Representative semantic supports used to restrict local depth modification. The masks determine the spatial support of the refinement but do not prescribe the internal three-dimensional geometry of the facial regions.

6.4 Early Semantic Enhancement

The first semantic experiments tested whether spatial localization alone could provide useful local control, progressing from scene-level separation to mask-guided and nonlinear facial depth modification.

6.4.1 Exp48: Scene-Aware Foreground–Background Enhancement

Exp48 investigated broad foreground–background depth separation through nonlinear depth adjustment. The global surface organization remained close to the baseline, while changes in small facial structures were limited. This indicated that scene-level depth separation and local facial refinement should be treated as distinct objectives.

6.4.2 Exp57: Mask-Guided Local Depth Enhancement

Exp57 applied local modification directly within semantic facial masks. The resulting changes were modest but spatially controlled, and the surrounding DA3 geometry remained stable. In contrast to external-geometry fusion, this formulation did not require registration between independent coordinate systems.

A simplified formulation is

$$D' = D_B + \alpha MF(D_B), \quad (6.9)$$

where F represents a local depth-enhancement operation.

6.4.3 Exp58: Face-Specific Nonlinear Remapping

Exp58 applied nonlinear redistribution directly to the existing depth values inside the facial ROIs. This changed the local depth statistics without replacing the surrounding representation, although the resulting geometric differences remained moderate.

Rather than introducing new geometry, the local depth distribution was transformed:

$$D'_{\text{ROI}} = T(D_{\text{ROI}}). \quad (6.10)$$

6.4.4 Exp69: Semantic Local Contrast Enhancement

Exp69 applied local depth-contrast enhancement inside the semantic supports and produced a more visible local modification than the preceding semantic variants. Its significance was not the reconstruction of new facial anatomy, but the demonstration that local differentiation could be increased without introducing an independently reconstructed geometric prior.

6.5 Local Face-Aware Enhancement: Exp71–Exp73

The Exp71–Exp73 branch examined increasingly localized facial operations, including direct inflation, Gaussian-supported enhancement, ROI refinement, and image-guided modification.

6.5.1 Local Inflation

The simplest strategy increased depth inside a facial support.

For a local mask M , a basic inflation model can be written as

$$D' = D_B + \alpha M. \quad (6.11)$$

This operation increases the overall prominence of the selected facial region but does not differentiate individual anatomical structures; the nose, cheeks, lips, eye region, and chin may therefore be displaced together.

6.5.2 Gaussian Enhancement

A smoother local enhancement can be produced using a Gaussian support

$$G(x, y) = \exp \left[-\frac{(x - x_c)^2}{2\sigma_x^2} - \frac{(y - y_c)^2}{2\sigma_y^2} \right]. \quad (6.12)$$

The modified depth becomes

$$D' = D_B + \alpha G. \quad (6.13)$$

Gaussian support produced smoother spatial transitions than a hard mask. Across the Exp71–Exp73 experiments, the resulting changes in facial prominence and surface smoothness were incremental and remained strongly dependent on the support location and spatial scale.

6.5.3 Image-Guided Enhancement

Image-derived cues were also evaluated as guides for local enhancement. Although intensity variation can indicate potentially relevant structures, painterly shading and texture do not necessarily correspond to physical geometry. Consequently, the resulting facial modifications were spatially variable and remained limited.

6.6 Semantic Depth Remapping: Exp74–Exp80

The next branch shifted from adding geometric volume to redistributing the depth already present in semantically important regions. This reduced dependence on external priors and

made the modification directly compatible with the established image-aligned baseline.

The general formulation can be written as

$$D'(x, y) = \begin{cases} T(D_B(x, y)), & (x, y) \in \Omega_S, \\ D_B(x, y), & \text{otherwise,} \end{cases} \quad (6.14)$$

where Ω_S denotes a semantic region and T is a local depth transformation.

6.6.1 Exp74A: Initial Semantic Depth Remapping

Exp74A evaluated region-specific depth redistribution as an alternative to geometry injection. The resulting depth field showed stronger regional differentiation while retaining the broad DA3 structure, and the modification was easier to constrain spatially because it operated directly on the baseline representation.

6.6.2 Exp75A–B: Semantic Region Depth Redistribution

Exp75A and Exp75B progressively increased the strength of semantic depth redistribution. Stronger settings increased regional separation but also increased the risk of abrupt local transitions, revealing the trade-off

$$\text{stronger local differentiation} \leftrightarrow \text{transition stability}. \quad (6.15)$$

6.6.3 Exp76A–B: Foreground-Aware Compression

Exp76A–B compressed the surrounding depth range while preserving important foreground structures. This moderately increased foreground–background separation, but improvements in small facial structures remained limited.

6.6.4 Exp77A: Semantic-Aware Transition Smoothing

$$M_s = \mathcal{S}(M), \quad (6.16)$$

where $\mathcal{S}$ denotes spatial smoothing.

The modified surface becomes

$$D' = (1 - M_s)D_B + M_s D_E, \quad (6.17)$$

where D_E is the enhanced local depth.

The resulting surface showed less abrupt transition behaviour near the modified semantic boundary, motivating more explicit boundary management in the following experiments.

6.6.5 Exp78A–B: Boundary-Safe Semantic Enhancement

The Exp78 branch restricted strong modification near semantic boundaries.

Let $B(x, y)$ denote a boundary-risk map.

A safe support can be defined conceptually as

$$M_{\text{safe}} = M(1 - B). \quad (6.18)$$

Enhancement is then applied as

$$D' = D_B + \alpha M_{\text{safe}} E. \quad (6.19)$$

Restricting modification near semantic contours reduced boundary sensitivity, while the refined variant further softened the transition between the semantic interior and the surrounding baseline.

6.6.6 Exp79A–B: Semantic Interior Enhancement

The Exp79 experiments concentrated modification toward the interior of the semantic region. If

$$M_{\text{int}} \quad (6.20)$$

denotes an eroded or interior-weighted mask, enhancement becomes

$$D' = D_B + \alpha M_{\text{int}} E. \quad (6.21)$$

This interior weighting reduced modification near the mask contour and produced a more spatially concentrated depth change with less visible halo formation.

6.6.7 Exp80A: Safe Interior Enhancement

Exp80A combined semantic localization, interior-focused weighting, smooth support transitions, and conservative depth modification while retaining the DA3-derived representation as the global geometric reference. It represented one of the more stable configurations in this semantic-remapping branch because the local modification remained spatially controlled and boundary disturbance was comparatively low.

Its significance should nevertheless be interpreted conservatively. Exp80A did not reconstruct missing internal facial anatomy; rather, it demonstrated a more reliable compromise between local depth redistribution and preservation of the inherited global relief.

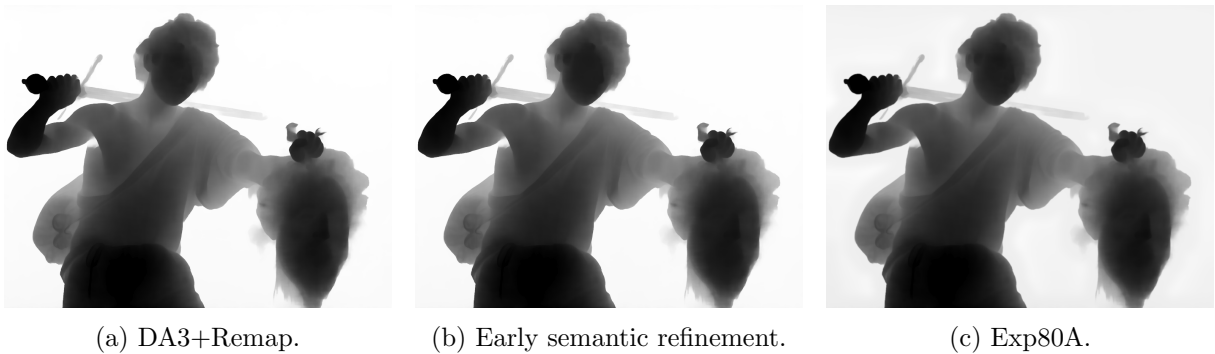


Figure 6.2: Representative progression from the fixed DA3+Remap baseline through early semantic enhancement to the boundary-aware interior refinement represented by Exp80A. The progression primarily improves spatial control of the modification rather than introducing independent anatomical geometry.

6.7 Further Refinement: Exp81–Exp84

The Exp81–Exp84 branch further investigated boundary suppression, local contrast, smoothing, and surface continuity. These experiments clarified that artifact suppression and geometric enhancement are different optimization objectives: a smoother surface with fewer visible boundary artifacts does not necessarily provide stronger differentiation of internal facial structures, while stronger local modification can increase artifact sensitivity.

The main contribution of this branch was therefore improved control of the modification rather than a fundamentally new source of facial geometry.

6.8 Z-Aware Semantic Compression: Exp85

The Exp85 branch extended semantic processing by making depth redistribution dependent on both semantic support and the existing depth position, rather than treating enhancement as uniform local lifting.

A general Z-aware formulation can be written as

$$D'(x, y) = T(D_B(x, y); S(x, y)), \quad (6.22)$$

where $S(x, y)$ represents semantic support.

6.8.1 Exp85A: Semantic-Aware Nonlinear Compression

Exp85A applied nonlinear depth compression while preserving semantically important regions. The resulting surface maintained global continuity and smooth transitions, but remained qualitatively close to Exp80A. It should therefore be interpreted as a refinement of the semantic-remapping strategy rather than a fundamentally different geometric solution.

6.8.2 Exp85B: Further Compression Variant

Exp85B tested stronger depth redistribution, but the additional changes remained limited and did not resolve the remaining weakness in internal facial volume.

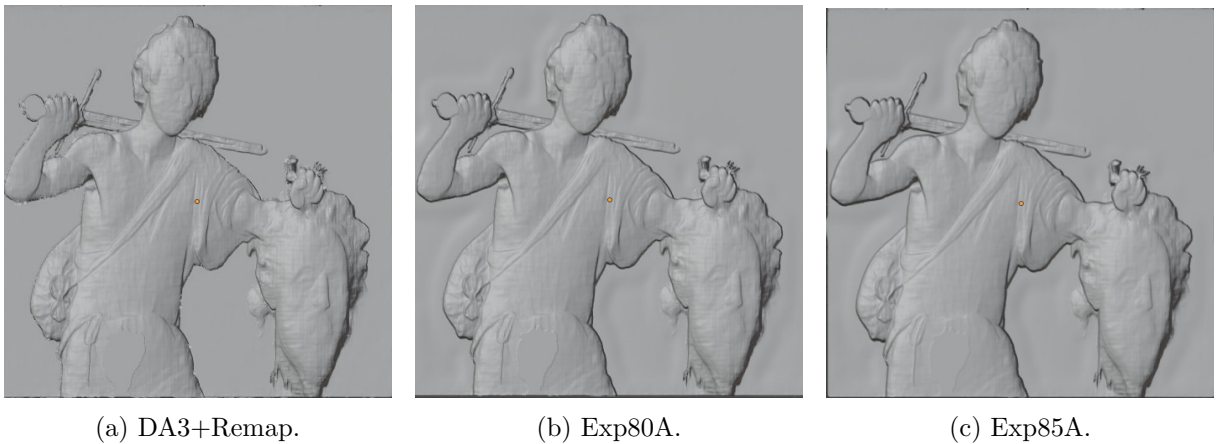


Figure 6.3: Controlled Blender comparison of the DA3+Remap baseline with representative late-stage semantic-remapping configurations Exp80A and Exp85A.

6.9 Camera-Relative Local Reconstruction: Exp90

Exp90 revisited explicit three-dimensional reconstruction using local facial crops to test whether a restricted field of view could reduce the compatibility problems observed in Chapter 5. Separate reconstructions were generated for David’s face and Goliath’s head.

The results remained inconsistent. David’s reconstruction exhibited weak facial coherence and irregular local shape, while Goliath’s reconstruction was more coherent but still contained limited usable facial detail. Painterly texture, ambiguous three-dimensional evidence, reconstruction instability, and compatibility with the fixed global relief remained limiting factors.

Exp90 therefore provided no consistent qualitative advantage over direct semantic depth processing for the selected artwork. This result reinforced the decision to retain the image-aligned baseline and focus subsequent work on structured local modification rather than independent local reconstruction.

6.10 Representative Experiment Overview

The progression of the principal experiments is summarized in Table 6.1.

Table 6.1: Summary of representative semantic and face-aware refinement experiments.

Exp.	Strategy		Main objective	Main observation
Exp48	Scene-aware enhancement	en-	Improve foreground–background separation	Broad global surface organization remained visually consistent with the baseline, while changes in small facial structures remained limited.
Exp57	Mask-guided depth enhancement	depth	Increase local prominence	Produced modest local modification while maintaining the global baseline.
Exp58	Face-specific nonlinear remapping		Redistribute depth inside facial ROIs	Changed the local depth distribution, but the geometric effect remained moderate.
Exp69	Semantic local contrast	local con-	Increase facial differentiation without external geometry	Produced a comparatively visible local depth modification without requiring external geometry registration.
Exp71–73	Local face enhancement	enhance-	Evaluate inflation, Gaussian, ROI, and image-guided strategies	Local prominence and surface smoothness changed slightly, but the geometric effect remained incremental.

Continued on next page

Exp.	Strategy		Main objective	Main observation
Exp74A	Semantic remapping	depth	Redistribute existing local depth	Provided spatially constrained regional depth modification directly within the baseline representation.
Exp75A–B	Semantic redistribution		Increase regional separation	Stronger separation was possible, but aggressive settings increased transition risk.
Exp76A–B	Foreground-aware compression		Improve depth allocation around foreground structures	Produced broader foreground–background separation, while local facial effects remained modest.
Exp77A	Transition smoothing		Reduce semantic-boundary discontinuities	Produced smoother visible transitions near semantic boundaries.
Exp78A–B	Boundary-safe enhancement		Reduce contour artifacts	Reduced modification near semantic boundaries while keeping the local enhancement comparatively subtle.
Exp79A–B	Interior enhancement		Reduce halo and boundary sensitivity	Produced more controlled modification by concentrating changes away from boundaries.
Exp80A	Safe interior semantic enhancement		Balance local modification and boundary stability	Provided controlled interior modification with reduced boundary sensitivity; did not independently reconstruct facial anatomy.
Exp81–84	Semantic refinement		Improve contrast, smoothing, and artifact control	Further reduced visible artifacts, but did not fundamentally resolve weak internal facial geometry.
Exp85A	Z-aware semantic compression		Combine semantic support with nonlinear depth allocation	Preserved continuity and produced behaviour close to Exp80A; represented a refinement of the same semantic-remapping strategy.
Exp85B	Further compression	Z-aware	Test stronger depth redistribution	Additional changes remained limited.

Continued on next page

Exp.	Strategy	Main objective	Main observation
Exp90	Camera-relative local reconstruction	Re-evaluate explicit local three-dimensional reconstruction	Local reconstruction remained inconsistent and offered no clear geometric advantage over baseline-preserving semantic refinement.

6.11 Comparison with External Geometry Fusion

The semantic-refinement experiments differed fundamentally from the external-geometry approach investigated in Chapter 5. External fusion required correspondence between independently generated geometric representations and was therefore sensitive to scale, camera, projection, and registration mismatch. Semantic refinement instead operated directly on the established DA3+Remap representation, avoiding replacement of the stable baseline and providing more direct spatial control over where changes were introduced.

This advantage did not solve the complete geometric problem. A semantic mask identifies a region in which modification is allowed, but it does not independently specify the internal organization of structures such as the nose bridge, nose tip, cheeks, eye sockets, lips, and chin. Semantic refinement therefore reduced the correspondence problem of external geometry, but exposed a different limitation: spatial localization alone is insufficient to define anatomically structured relief.

The methodological progression of Chapters 5 and 6 can therefore be summarized as a shift from *geometry replacement* to *baseline-preserving localization*. The remaining requirement was to move from region-level control toward feature-level geometric control.

6.12 Synthesis and Remaining Limitation

Across the semantic-refinement experiments, three findings were consistent. First, operating directly on the DA3+Remap representation provided substantially simpler spatial control than external-geometry fusion because no independent geometric coordinate system had to be registered. Second, boundary handling was as important as the nominal enhancement strength: interior-weighted and smoothly varying supports produced more stable transitions than uniform full-mask modification. Third, increasing local depth contrast or prominence did not by itself create anatomically structured geometry.

This final limitation is fundamental:

$$\boxed{\text{semantic localization} + \text{generic depth enhancement} \neq \text{anatomically structured refinement.}} \quad (6.23)$$

A face can therefore become more prominent in depth without improving the relative geometric organization of the nose, cheeks, eye sockets, lips, and chin. Similarly, increasing local contrast does not determine which anatomical structures should move forward or backward relative to one another.

The semantic phase should consequently be interpreted as solving the *where* component of local refinement more effectively than the *how*. Exp80A and Exp85A provided representative examples of controlled baseline-preserving processing, but the remaining problem required explicit structure-level information rather than stronger generic semantic enhancement.

6.13 Chapter Summary

This chapter shifted the investigation from externally reconstructed geometry to baseline-preserving semantic refinement. The central result was that working directly within the DA3+Remap representation provided substantially more reliable spatial control: local modifications could be restricted to selected regions, semantic boundaries could be managed explicitly, and the global image-aligned geometry could be retained.

The experimental progression also showed that spatial control and anatomical control are different problems. Early mask-guided operations demonstrated local modification but provided limited internal structure. Boundary-safe, interior-weighted, and Z-aware depth redistribution improved stability and reduced the dependence on independent geometry, with Exp80A and Exp85A representing controlled examples of this strategy. Reintroducing local explicit three-dimensional reconstruction in Exp90 did not provide a consistent advantage.

The principal conclusion of this phase is therefore not that semantic processing reconstructs missing facial geometry, but that it provides an effective mechanism for determining where baseline modification can be applied safely. Semantic masks alone do not specify the relative geometry of individual facial structures. This distinction motivates Chapter 7, which introduces feature-level localization, anatomical landmarks, and signed geometric priors to move from region-level toward structure-level control.

Chapter 7

Parametric and Anatomy-Guided Facial Geometry Refinement

7.1 Introduction

The semantic-refinement experiments presented in Chapter 6 established that local modification could be spatially constrained while retaining the DA3-derived representation as the global geometric reference. However, semantic masks primarily addressed *where* modification should occur and did not specify the relative geometry of individual facial structures such as the nose, cheeks, eye sockets, lips, chin, and jaw.

The next experimental phase therefore introduced progressively more structured geometric guidance. The objective was not to reconstruct an independent three-dimensional face, but to determine whether structural cues, anatomical landmarks, signed parametric priors, and constrained local optimization could provide more explicit feature-level control while remaining compatible with the established shallow-relief representation.

The experiments in this chapter can be organized into five methodological families:

1. additional depth redistribution and structural-cue extraction;
2. feature-guided and multi-scale local processing;
3. landmark localization;
4. parametric facial priors;
5. constrained anatomical templates and surface relaxation.

The complete experimental sequence extends from Exp91 to Exp107A.

The central question was:

Can simplified anatomical and parametric geometry provide more structured local facial control than semantic-region processing alone while preserving the stable global relief?

7.2 Objectives and Design Constraints

The experimental design was guided by four limitations observed in the preceding semantic phase. Semantic enhancement did not explicitly distinguish neighbouring anatomical structures; small local depth changes often became difficult to distinguish after shallow-relief reconstruction; image-derived features were ambiguous because they could represent geometry, lighting, texture, hair, beard, or painterly appearance; and broader facial elevation was more visible in the reconstructed surface but increased the risk of halo-like or mask-shaped deformation.

The methods investigated in this chapter were therefore required to satisfy the following constraints:

- retain the DA3-derived depth lineage as the global geometric reference;
- remain localized to selected facial regions;
- allow different anatomical structures to receive different geometric roles;
- support both positive and negative geometric modification;
- avoid hard mask boundaries;
- remain compatible with the fixed Blender evaluation pipeline.

7.3 Working Inputs and Geometric Supports

All experiments in this chapter remained within the same DA3-derived depth lineage, but the original DA3+Remap TIFF was not the direct input to every stage. The Exp91 onward branch was initialized from the retained intermediate facial baseline

`phase2C_v4_R5_E2_balanced_elliptical_mouth_refined_uint16.tiff,`

which was itself derived from DA3+Remap through the selected Phase-2B R5 semantic reinjection and the Phase-2C E2 localized mouth correction. The exact Phase-2B and Phase-2C formulations and parameters are reported in Appendix A.5.

Subsequent parametric experiments inherited progressively refined facial baselines derived from this depth lineage. Within each controlled experimental family, the corresponding native input baseline and all non-tested parameters were kept fixed so that the effect of the tested geometric modification could be isolated.

The working depth resolution was

$$W = 1022, \quad H = 798. \quad (7.1)$$

All internal processing used normalized floating-point depth values:

$$D(x, y) \in [0, 1]. \quad (7.2)$$

The principal spatial supports remained David’s face, Goliath’s head, their combined facial support, and selected foreground regions where required.

Final candidates were exported as 16-bit TIFF depth maps and evaluated under the same Blender reconstruction conditions.

This provided a controlled basis for comparing the geometric effects of the refinement methods without introducing changes in the reconstruction configuration.

7.4 From Semantic Remapping to Structural Cues

The first experiments, Exp91–98, investigated whether additional processing of the existing depth or image structure could provide stronger facial information before explicit parametric priors were introduced.

Rather than treating each experiment as an independent method, this family can be understood as a sequence of tests addressing three questions:

1. Is further global or semantic compression still useful?
2. Can facial feature lines provide reliable geometric localization?
3. Can depth, image, or curvature detail be re-injected effectively?

7.4.1 Additional Global and Semantic Compression: Exp91–92

Exp91 evaluated additional global Z-aware compression using several power-law settings. Mild and intermediate transformations produced limited visible differences, whereas stronger compression increased the risk of artificial redistribution without introducing new facial structure. Exp92 extended this strategy by applying different nonlinear mappings to different semantic regions.

Although both experiments produced measurable changes in the depth representation, their final Blender surfaces remained close to the established baseline. This indicated that further global or region-specific gamma-based redistribution offered limited additional benefit once the preceding semantic-remapping stages had already established the working facial baseline.

7.5 Feature-Line Extraction and Structural Localization

The next branch investigated whether image-space facial structures could provide finer localization than semantic masks.

7.5.1 Automatic Feature-Line Extraction: Exp93

The processing pipeline included:

1. grayscale conversion;
2. facial ROI extraction;
3. Gaussian smoothing;
4. edge detection;
5. connected-component filtering;
6. morphological cleanup;

7. semantic-mask restriction.

The extracted maps contained both potentially useful structural responses and substantial painterly texture. David’s face produced comparatively coherent feature lines, whereas Goliath’s region showed stronger interference from hair, beard, forehead texture, and local image contrast. The experiment therefore demonstrated that image-space feature extraction could improve localization but could not be interpreted directly as facial geometry.

7.5.2 Cleaned Feature-Line Maps: Exp93A

Exp93A introduced stronger connected-component filtering and revised detection thresholds, retaining larger and more spatially coherent feature structures while suppressing part of the high-frequency painterly response. These cleaned maps were retained as spatial guidance signals rather than interpreted as direct surface geometry.



Figure 7.1: Cleaned facial feature-line maps for David and Goliath within their respective semantic regions of interest. The white feature lines were used as structural localization cues, while the green contours indicate the spatial support of the corresponding facial regions. These cues were used for guidance rather than as direct geometric reconstruction.

7.6 Feature-Guided and Multi-Scale Local Processing

Exp94–98 investigated whether the extracted structural cues could be combined with information already present in the baseline or source image.

7.6.1 Feature-Guided Depth Gradients: Exp94

A softened feature-line gate was combined with the gradient magnitude of the baseline depth.

If L denotes the structural-line support, a soft feature gate was obtained through dilation and smoothing:

$$G_f = \mathcal{G}(\text{Dilate}(L)). \quad (7.3)$$

The depth-gradient magnitude was

$$G_D = \sqrt{\left(\frac{\partial D}{\partial x}\right)^2 + \left(\frac{\partial D}{\partial y}\right)^2}. \quad (7.4)$$

The local cue was then formed as

$$E = G_f G_D. \quad (7.5)$$

The resulting depth correction remained spatially controlled but produced only subtle differences after reconstruction, indicating that depth-gradient magnitude could support localization but did not provide a strong independent source of missing facial geometry.

7.6.2 Multi-Source Semantic Detail Cue: Exp95

Exp95 combined:

- feature-line proximity;
- image gradient;
- depth gradient;
- semantic support;
- boundary suppression.

The resulting multi-source control field provided useful spatial localization, but the underlying geometric information remained weak. Exp95 therefore reinforced the distinction between identifying where a correction may be useful and determining what three-dimensional shape that correction should have.

7.6.3 Multi-Scale Depth Re-Injection: Exp96

The baseline depth was decomposed into several spatial scales:

$$S_s = \mathcal{G}_2(D), \quad S_m = \mathcal{G}_5(D), \quad S_l = \mathcal{G}_{12}(D). \quad (7.6)$$

The corresponding components were:

$$D_{\text{fine}} = D - S_s, \quad (7.7)$$

$$D_{\text{mid}} = S_s - S_m, \quad (7.8)$$

$$D_{\text{large}} = S_m - S_l. \quad (7.9)$$

The components were re-injected under semantic control, but the final Blender differences remained small. The experiment therefore showed that multi-scale decomposition cannot recover strong facial geometry when the corresponding structure is already weak in the source depth representation.

7.6.4 Image- and Curvature-Guided Micro-Relief: Exp97–98

Exp97 used Difference-of-Gaussians image detail as a local micro-relief cue, whereas Exp98 used a curvature-related signal derived from the Laplacian of a smoothed depth field. Both methods produced localized numerical modifications, but their final reconstructed effects remained limited.

Taken together with Exp94–96, these experiments showed that feature lines, gradients, image detail, and curvature were more useful as localization or weighting cues than as independent sources of additional facial shape.

7.7 Macro-Scale Facial Support

Exp99 tested the opposite spatial regime by increasing broad low-frequency facial support rather than re-injecting fine detail. The resulting modification remained more visible after Blender displacement, but stronger broad-scale elevation also produced a mask-related halo around Goliath.

The experiment therefore exposed a spatial-scale trade-off:

$$\text{greater spatial support} \rightarrow \text{greater final-surface visibility}, \quad (7.10)$$

but also:

$$\text{greater spatial support} \rightarrow \text{greater boundary-artifact risk}. \quad (7.11)$$

Exp99 was not retained as a final configuration. Its main contribution was to demonstrate that increasing the spatial extent of a depth correction improves its visibility in the final surface but simultaneously increases sensitivity to support boundaries.

7.8 Landmark Localization

The limitations of feature maps motivated a transition from distributed image structures to explicit anatomical anchors.

7.8.1 Automatic Landmark Strategies: Exp100A–B

Exp100A derived candidate landmarks from feature-map centroids, while Exp100B used connected components of the cleaned feature-line representation. Although many resulting points remained inside the facial masks, they did not consistently correspond to meaningful anatomy and were frequently attracted to hair, beard, texture, or facial contours. Both automatic strategies were therefore rejected as primary anatomical anchors.

7.8.2 Semi-Manual Anatomical Landmarks: Exp100C

Semi-manual landmarks were introduced in Exp100C to provide anatomically interpretable anchors for the eyes, nose, mouth, cheeks, chin, and face centre. Local Gaussian primitives centred on these points provided substantially more predictable spatial control than the automatic feature-map strategies.

The remaining limitation was representation: isolated Gaussian primitives could behave as individual bumps rather than continuous anatomical structures. Exp100C therefore provided the transition from landmark localization to continuous parametric facial geometry.

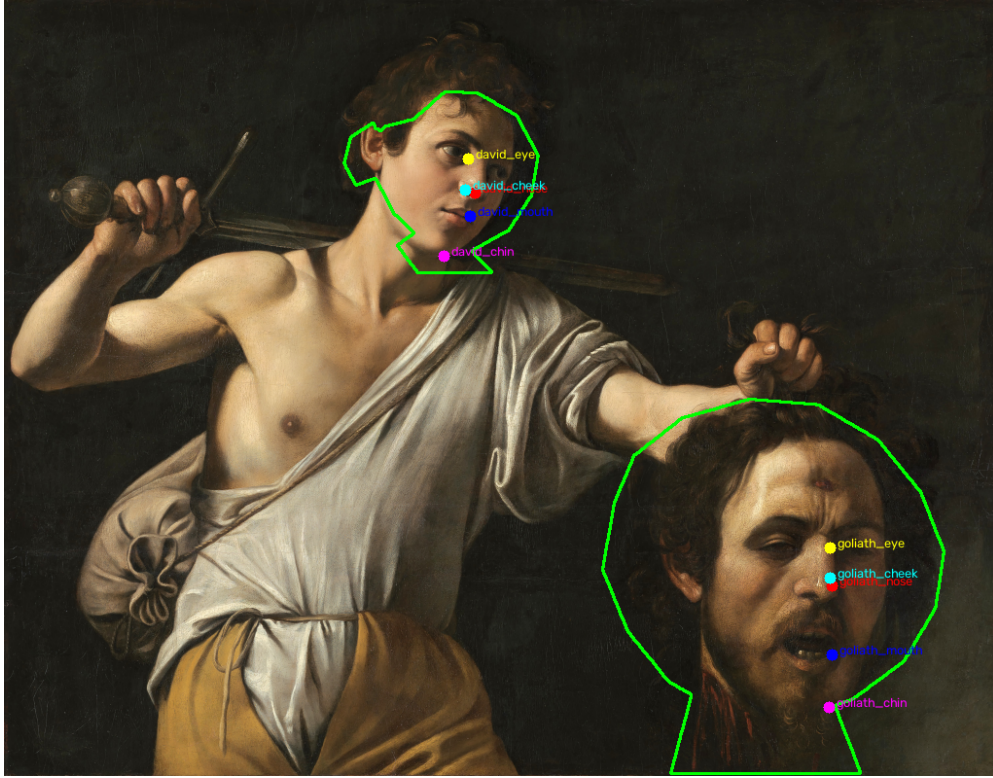


Figure 7.2: Semi-manually defined anatomical landmarks for David and Goliath. The selected reference points identify key facial structures, including the eyes, nose, cheeks, mouth, chin, and face centre, and provide spatial anchors for the subsequent construction of local parametric facial priors.

7.9 Parametric Facial Priors

Exp101 introduced the first continuous parametric facial representation.

Instead of modifying the face using a single mask or isolated point bumps, separate anatomical components were defined.

The prior combined continuous positive components for the nose bridge, nose tip, lips, cheeks, chin, and jaw with negative components representing the mouth opening and eye sockets.

A signed prior can be represented as

$$P = 0.80P_{\text{bridge}} + 1.00P_{\text{nose}} + 0.46P_{\text{lips}} - 0.18P_{\text{mouth}} + 0.36P_{\text{cheeks}} - 0.30P_{\text{eyes}} + 0.26P_{\text{chin}} + 0.22P_{\text{jaw}}. \quad (7.12)$$

The use of signed components was important because facial geometry contains both protruding and recessed structures.

7.9.1 Initial Amplitude Study: Exp101A

Three global amplitudes were evaluated:

Table 7.1: Exp101A parametric-prior amplitude variants.

Variant	Amplitude
Safe	0.025
Visible	0.050
Strong	0.080

The Safe variant ($\alpha = 0.025$) produced little visible change, whereas the Visible configuration ($\alpha = 0.050$) produced a clearer central-face modification without a severe visible boundary artifact. Increasing the amplitude to 0.080 further increased the numerical difference but did not produce a proportional change in the reconstructed surface. The 0.050 configuration was therefore retained as the reference for the subsequent controlled amplitude study.

7.9.2 Controlled Amplitude Increase: Exp101B

Exp101B increased the global prior amplitude from 0.050 to 0.065 while retaining the remaining formulation. The correction remained localized and did not introduce a severe visible boundary artifact. The most apparent difference occurred around the central face and nose, although the overall reconstructed change remained modest. Exp101B was therefore retained as the reference configuration for the subsequent parametric comparisons.

7.9.3 Wider Nose Support: Exp101C

Exp101C increased the width of the nose-bridge support. The resulting prior was broader and smoother, but the Blender reconstruction changed only slightly, indicating that bridge width alone was not the principal limitation.

7.10 Low-Frequency Support and Surface Relaxation

The next family, Exp102–104, investigated whether broader and more continuous target surfaces could produce a stronger result than direct parametric addition.

7.10.1 Low-Frequency Face Envelope: Exp102

Exp102 added a broad elliptical facial envelope to increase low-frequency support. The target became smoother and more spatially continuous, but the reconstructed surface remained close to Exp101B. Additional low-frequency elevation alone was therefore insufficient to produce a substantial new facial structure.

7.10.2 Constrained Surface Relaxation: Exp103A

Exp103A replaced direct additive editing with a constrained local optimization.

The target was defined as

$$D_{\text{target}} = D + \alpha P. \quad (7.13)$$

The relaxed solution approximately minimized:



Figure 7.3: Combined signed parametric facial prior used in Exp101C for David and Goliath. Bright and dark regions represent localized positive and negative depth modifications, respectively, encoding protruding and recessed facial structures.

$$E(D') = w_d \|D' - D\|^2 + w_c W \|D' - D_{\text{target}}\|^2 + w_s \|\nabla D'\|^2. \quad (7.14)$$

where w_d , w_c , and w_s control the data-fidelity, target-consistency, and smoothness terms, respectively, while W denotes the spatial weighting field used to restrict the target-consistency term to the selected facial support.

The optimization produced a smooth local surface without severe seam, halo, or boundary discontinuity. However, the reconstructed result remained close to Exp101B, showing that a more sophisticated application mechanism could improve target smoothness without overcoming limitations in the target geometry itself.

7.10.3 Continuous Diffused Target: Exp103B

Spatial diffusion of the parametric target produced a smoother field, but little additional surface change relative to Exp103A. This indicated that the relaxation formulation already provided substantial smoothing.

7.10.4 Continuous Midface Patch: Exp104A

Exp104A introduced a broader continuous midface support. Although spatially coherent, the target behaved mainly as a smooth height perturbation and produced little additional Blender difference. Broad positive support therefore remained insufficient to encode the desired anatomical differentiation.

7.11 Signed Anatomical Templates

Exp105A replaced the broad positive target with a signed midface template containing protruding components for the nose bridge, nose tip, cheeks, and central-face support, and recessed components for the eye sockets and nasolabial transitions.

7.11.1 Signed Midface Template: Exp105A

A representative formulation was:

$$T = 0.72T_{\text{bridge}} + 0.92T_{\text{nose}} + 0.38T_{\text{cheeks}} + 0.22T_{\text{support}} - 0.34T_{\text{eyes}} - 0.20T_{\text{nasolabial}}. \quad (7.15)$$

The resulting residual showed clearer spatial separation between the nose, cheeks, and eye-socket regions than the preceding broad-support configurations. The corresponding Blender surface also exhibited a slightly more organized central-face modification, although the overall difference remained modest. Exp105A was therefore retained as the reference for the subsequent strength tests.

7.11.2 Template-Weight Increase: Exp106A

Exp106A doubled the signed-template contribution from 0.50 to 1.00. The numerical modification increased, particularly in David’s facial region, but the reconstructed surface changed only modestly. This indicated diminishing returns from template-weight escalation.

7.11.3 Reference-Amplitude Increase: Exp107A

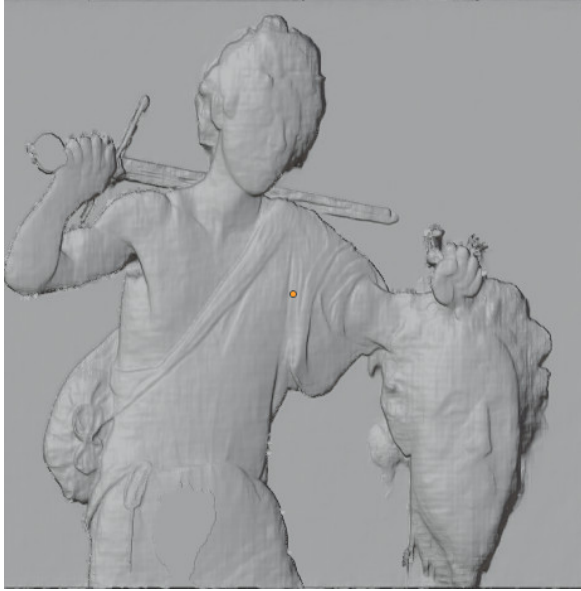
Exp107A increased the reference target amplitude from 0.065 to 0.100. The target field became stronger numerically, but the Blender reconstruction remained close to Exp106A, with only a marginal additional increase around the nose. Further amplitude escalation was therefore not pursued because the numerical increase was no longer accompanied by a proportional surface-level change.

7.12 Numerical Comparison

The numerical residuals of selected parametric and relaxation-based experiments are summarized in Table 7.2. Each value measures the modification relative to the corresponding native input baseline of that experiment and should therefore not be interpreted as a common-reference distance from DA3+Remap.

For Exp101B, Exp101C, and Exp102, the regional values were recomputed directly from the final unsigned 16-bit TIFF files used for Blender reconstruction. For the later relaxation-based experiments, the reported values correspond to the residual statistics recorded during the respective experimental runs.

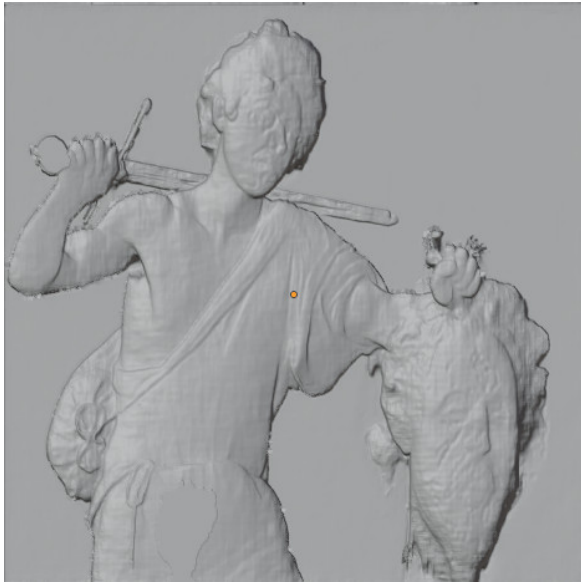
The regional values show that the modifications were substantially more concentrated inside the selected facial supports than suggested by the global image-wide mean. For example, Exp101B produced a global mean absolute residual of only 0.00117320, while the mean absolute residual inside David’s facial support was 0.01075478 and that inside Goliath’s support was 0.00700978.



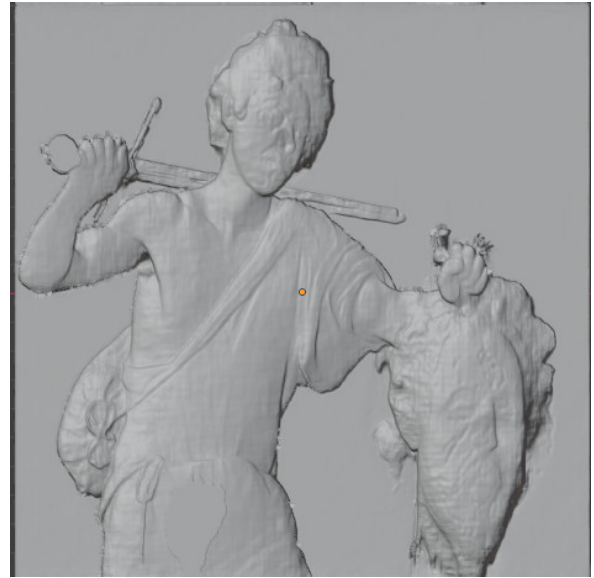
(a) DA3+Remap common reference.



(b) Exp101B.



(c) Exp105A.



(d) Exp107A.

Figure 7.4: Controlled Blender comparison relative to the common DA3+Remap reference. Exp101B, Exp105A, and Exp107A represent progressively structured parametric and anatomy-guided stages. The later configurations were generated from stage-specific DA3-derived facial baselines; therefore, this figure provides a common visual reference rather than a direct native-baseline ablation.

Table 7.2: Magnitude of selected parametric and relaxation-based depth modifications in normalized depth units.

Experiment	Global Mean Abs.	David Mean Abs.	Goliath Mean Abs.
Exp101B	0.00117320	0.01075478	0.00700978
Exp101C	0.00122996	0.01116960	0.00737464
Exp102	0.00151419	0.01350678	0.00913822
Exp103B	0.00037505	0.00345693	0.00223634
Exp104A	0.00036950	0.00347213	0.00218701
Exp105A	0.00046898	0.00450144	0.00275280
Exp106A	0.00061450	0.00602243	0.00357673

A similar pattern was observed for Exp101C and Exp102. The addition of the low-frequency face envelope in Exp102 increased the mean absolute modification to 0.01350678 inside David’s mask and 0.00913822 inside Goliath’s mask.

For Exp101B, Exp101C, and Exp102, recomputation from the final exported 16-bit depth maps showed that the residual outside the union of the David and Goliath facial supports was zero at the exported depth resolution. The corresponding modifications therefore preserved the surrounding depth field.

The very small differences between the recomputed global values for Exp101B, Exp101C, and Exp102 and the previously recorded floating-point statistics arise from quantization during unsigned 16-bit TIFF export. The values reported in Table 7.2 for these three configurations correspond directly to the exported depth maps used for Blender reconstruction.

Across the experimental sequence, larger normalized residuals did not necessarily correspond to proportionally larger visible changes in the reconstructed surface. Residual magnitude should therefore be interpreted as a measure of geometric modification rather than as a direct measure of perceptual or tactile improvement.

A physical interpretation of these normalized residuals, together with a cross-stage comparison using common-reference metrics, difference maps, boundary measurements, and anatomical height profiles, is provided in Appendix A.3.1 and Appendix A.3.3.

7.13 Representative Experimental Summary

Rather than reproducing every individual parameter variant in the main narrative, Table 7.3 summarizes the methodological progression.

Table 7.3: Representative progression of the parametric and anatomy-guided refinement phase.

Exp.	Method family	Observed behaviour		Decision / interpretation	
Exp91–92	Additional compression	Produced depth but limited Blender benefit.	measurable redistribution additional	Further remapping was not pursued.	gamma-based

Continued on next page

Exp.	Method family	Observed behaviour	Decision / interpretation
Exp93A	Feature-line extrac- tion	Retained larger structural lines while suppressing part of the painterly tex- ture.	Useful as spatial localization rather than geometry.
Exp94– 95	Feature-guided pro- cessing	Provided controlled spa- tial cues, but the under- lying geometric signal re- mained weak.	Retained as supporting lo- calization methods.
Exp96	Multi-scale depth re-injection	Produced very small final-surface differences.	Did not recover weak source geometry.
Exp97– 98	Image- and curvature-guided micro-relief	Generated localized nu- merical modification with limited final-surface ef- fect.	Not retained as standalone geometry sources.
Exp99	Macro facial sup- port	Produced more visible modification but intro- duced mask-related halo risk.	Rejected as a final strategy.
Exp100A– B	Automatic land- marks	Feature-based points were not reliably anatomical.	Rejected.
Exp100C	Semi-manual land- marks	Provided more anatom- ically interpretable and predictable spatial con- trol.	Used as the basis for para- metric geometry.
Exp101A– B	Signed parametric prior	Produced localized mod- ification without severe visible boundary ar- tifacts; Exp101B was retained as the refer- ence initial parametric configuration.	Retained as the reference configuration for subsequent controlled comparisons.
Exp101C	Wider bridge sup- port	Produced a broader tar- get with little additional Blender effect.	Not clearly superior.
Exp102– 104	Broad and relaxed targets	Produced smoother or more continuous target behaviour but remained close to Exp101B in Blender.	Useful methodological tests, but not retained as stronger candidates.

Continued on next page

Exp.	Method family	Observed behaviour	Decision / interpretation
Exp105A	Signed anatomical template	Produced a more structured central-face residual and modest additional visible change.	Retained for strength validation.
Exp106A–107A	Template-weight / amplitude increase	The prescribed modification increased while the additional visible Blender differences showed diminishing returns.	Further strength escalation was not justified.

7.14 Controlled Blender Assessment

Representative parametric and anatomy-guided configurations were reconstructed using the same Blender settings in order to compare their final-surface behaviour under controlled conditions.

The comparison focused on the following qualitative aspects:

- preservation of the global composition;
- visibility of local facial modification;
- behaviour of the nose and central-face region;
- organization of the eye and cheek regions;
- continuity of local transitions;
- presence or absence of halos, seams, or abrupt surface artifacts.

A numerical ranking is not used here because the evaluation was not based on a standardized perceptual protocol or a formal tactile user study.

Instead, representative configurations are compared qualitatively according to the behaviour observed in the reconstructed surface.

Across the representative configurations, the global composition remained visually consistent and no severe boundary artifact dominated the reconstructed surfaces. The principal differences were therefore localized to the facial supports.

The comparison shows a progression from direct parametric modification toward increasingly structured signed targets. Exp101B established a controlled reference configuration; Exp102–103 primarily improved continuity; and Exp105A introduced clearer signed anatomical organization. Increasing the template weight in Exp106A and the reference amplitude in Exp107A increased the prescribed numerical modification but produced only limited additional surface-level change.

The main result of this comparison is therefore not that stronger parametric targets produced progressively better reliefs. Rather, the experiments demonstrate that explicit anatomical organization improved control over the location and sign of the correction, while the magnitude of the visible reconstructed benefit reached diminishing returns within the tested range.

Table 7.4: Qualitative Blender comparison of representative parametric and anatomy-guided facial-refinement experiments. The descriptors summarize the behaviour observed under the fixed reconstruction configuration and should not be interpreted as standardized perceptual scores.

Configuration	Local facial effect	Central-face structure	Transition behaviour	Principal observation
Exp101B	Moderate and localized	Slightly increased central-face prominence	No severe visible boundary artifact	Served as the controlled reference configuration for the initial parametric-prior family.
Exp102	Broad but subtle	Increased low-frequency support	Smooth	Produced a smoother facial envelope but only limited additional visible benefit.
Exp103A	Controlled and subtle	Similar to Exp101B	Smooth, without severe visible boundary artifacts	Surface relaxation produced smoother target behaviour but did not substantially strengthen the final facial geometry.
Exp105A	Localized and more structured	More explicit separation between positive and recessed anatomical components	No severe visible boundary artifact	Signed anatomical structure provided more explicit organization of the midface.
Exp106A	Stronger numerical modification	Central-face effect slightly stronger than Exp105A	No severe visible boundary artifact	Increasing template weight produced diminishing visible returns.
Exp107A	Strongest parametric target	Marginal additional increase in nose prominence	No severe visible boundary artifact	Further amplitude increase did not produce a proportional improvement in the reconstructed surface.

7.15 Discussion

The experiments in this chapter clarify the distinction between localization, target geometry, and modification strength. Image-derived feature lines, gradients, and curvature provided useful spatial cues but did not independently define missing facial shape. Fine-scale corrections were often difficult to distinguish after shallow-relief reconstruction, whereas broader modifications remained more visible but became increasingly sensitive to semantic boundaries and halo-like deformation.

Introducing semi-manual landmarks improved anatomical localization substantially, at the cost of reduced automation. More importantly, signed parametric priors allowed protruding and recessed components to be represented separately, providing more explicit structural control than uniform positive lifting. Nevertheless, these primitives remained simplified geometric approximations rather than subject-specific three-dimensional anatomical models.

The relaxation experiments further showed that smoother optimization cannot compensate for an insufficient geometric target. Exp103A–104A improved continuity but did not substantially strengthen the reconstructed facial geometry. Conversely, Exp105A demonstrated that changing the organization of the target itself—by introducing signed anatomical structure—was more informative than simply increasing smoothing or spatial extent.

Finally, Exp106A and Exp107A showed diminishing returns with increasing target strength. Larger prescribed residuals did not produce proportional changes in the reconstructed surface. The practical limitation of this phase was therefore not only correction magnitude, but the information content of the procedural target representation.

The phase remained limited by semi-manual landmark placement, simplified anatomical primitives, evaluation on a single painting, and the absence of a formal tactile user study. The results therefore support conclusions about geometric localization, surface continuity, and reconstruction behaviour, but not about tactile recognition or perceptual effectiveness.

7.16 Chapter Summary

This chapter moved from region-level semantic control toward explicit feature-level facial geometry. The early experiments showed that additional compression, feature-line processing, multi-scale depth decomposition, image detail, and curvature cues could improve localization or redistribute existing depth, but did not provide a strong independent source of missing facial shape.

The introduction of anatomical landmarks and signed parametric priors represented the more important methodological change. Semi-manual landmarks provided reliable spatial anchors, while signed components allowed protruding and recessed facial structures to be represented separately. This improved the organization and controllability of local corrections without requiring replacement of the DA3-derived global geometry.

The later relaxation and strength studies exposed the principal limitation of this approach. Smoother optimization improved continuity but could not compensate for limited target geometry, while increasing template weight or target amplitude produced progressively smaller surface-level returns. The experiments therefore indicate that the structure and information content of the local geometric prior were more important than simply increasing correction strength.

These findings motivated Chapter 8, which investigates whether independently estimated surface orientation, confidence, curvature, and dense landmark information can provide additional geometric guidance beyond the simplified parametric primitives used here.

Chapter 8

Geometry-Aware and Surface-Normal-Guided Facial Refinement

8.1 Introduction

The parametric facial-refinement experiments presented in Chapter 7 showed that increasingly structured anatomical priors improved control over the location and sign of local corrections, but stronger target amplitudes did not produce proportional changes in the reconstructed surface. This motivated the final geometry-aware phase of the thesis.

Rather than further increasing the strength of procedural facial priors, this phase examined whether complementary geometric information could provide additional local guidance. The investigated cues included independently estimated surface normals, depth-normal consistency, confidence, facial landmarks, curvature, explicit mesh-space representation, and constrained surface reconstruction.

The DA3-derived depth lineage remained the global geometric reference throughout this phase, although later experiments did not necessarily operate directly on the original DA3+Remap TIFF. Instead, some inherited progressively refined facial baselines derived from the same global representation.

The objective was therefore not to replace the established relief geometry, but to determine whether these additional cues could provide more explicit local geometric control while preserving composition, viewpoint, global depth ordering, surface continuity, and compatibility with the fixed Blender reconstruction pipeline.

The complete experimental sequence from Exp108A to Exp122B can be organized into five principal methodological families:

1. independent normal estimation and validation;
2. confidence-aware normal-guided refinement;
3. landmark- and curvature-guided anatomical refinement;
4. mesh-space facial refinement;
5. constrained surface reconstruction and component analysis.

8.2 Objectives and Fixed Experimental Conditions

The geometry-aware phase addressed the following questions:

1. Do independently predicted surface normals contain useful local geometric information beyond the baseline depth field?
2. How consistent are those normals with the surface orientation already implied by the baseline?
3. Can confidence be used to restrict geometric guidance to locally reliable regions?
4. Do sparse anatomical landmarks provide more effective local control than broad semantic supports?
5. Does changing from depth space to explicit mesh space improve the effectiveness of local facial refinement?
6. Can constrained surface reconstruction integrate independent orientation information without destabilizing the global surface?

Within each controlled experimental family, the native input baseline and the Blender reconstruction settings were held fixed, while later families could inherit progressively refined facial baselines from the same DA3-derived lineage. This distinction was particularly important because several candidate corrections were small. Camera configuration, displacement scale, lighting, subdivision, mesh topology, and all non-tested processing parameters were therefore kept unchanged within each direct comparison.

8.3 Input Assets

The geometry-aware experiments used several complementary sources of information.

8.4 Overall Geometry-Aware Workflow

The experimental progression can be summarized as:

$$\begin{aligned} &\text{independent surface-normal estimation} \\ &\rightarrow \text{depth-normal consistency analysis} \\ &\rightarrow \text{bias compensation and confidence} \\ &\rightarrow \text{weak normal-guided refinement} \\ &\rightarrow \text{landmark / curvature-guided refinement} \\ &\rightarrow \text{mesh-space refinement} \\ &\rightarrow \text{normal-guided reconstruction} \\ &\rightarrow \text{component contribution analysis.} \end{aligned} \tag{8.1}$$

The sequence was deliberately conservative: each stage evaluated an additional source of geometric information before allowing it to influence the inherited depth representation.

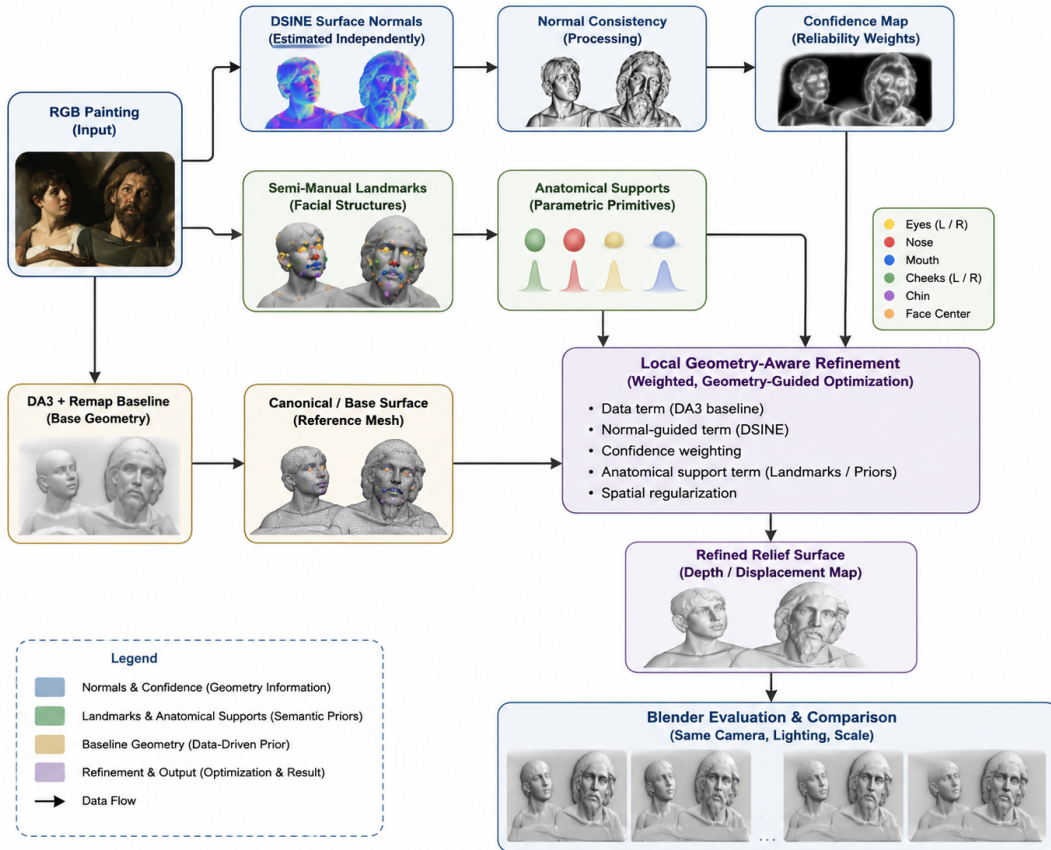


Figure 8.1: Overview of the geometry-aware refinement pipeline. The DA3-derived depth lineage provides the global geometric reference, while independently estimated DSINE normals, normal-consistency-based confidence, and automatically detected MediaPipe facial landmarks provide complementary geometric and anatomical guidance. These cues are combined through localized geometry-aware refinement and subsequently evaluated under consistent Blender reconstruction conditions.

Table 8.1: Principal input assets used during the geometry-aware refinement phase.

Input Asset	Role
DA3-derived depth lineage	Provides the global geometric reference; later experiments may use stage-specific refined facial baselines derived from this representation.
David facial mask	Restricts local refinement to David’s facial region.
Goliath head mask	Restricts local refinement to Goliath’s head region.
DSINE normal field [1]	Provides an independent estimate of local surface orientation.
Confidence maps	Control the influence of uncertain local geometric information.
Facial landmarks	Provide sparse anatomical anchors for localized refinement.
Curvature cues	Identify regions of existing local surface variation.
Canonical triangular mesh	Allows explicit mesh-space manipulation of the baseline surface.
Blender reconstruction pipeline	Provides a fixed environment for final-surface evaluation.

8.5 Independent Surface-Normal Analysis

The first family of experiments evaluated independently predicted surface orientation before allowing it to modify the inherited depth representation. Surface normals provide local orientation rather than absolute depth.

For a unit normal

$$\mathbf{n} = (n_x, n_y, n_z), \quad \|\mathbf{n}\|_2 = 1, \quad (8.2)$$

the corresponding local height-field gradients can be approximated as

$$p = -\frac{n_x}{n_z}, \quad q = -\frac{n_y}{n_z}. \quad (8.3)$$

This representation provided an orientation signal complementary to the scalar depth field.

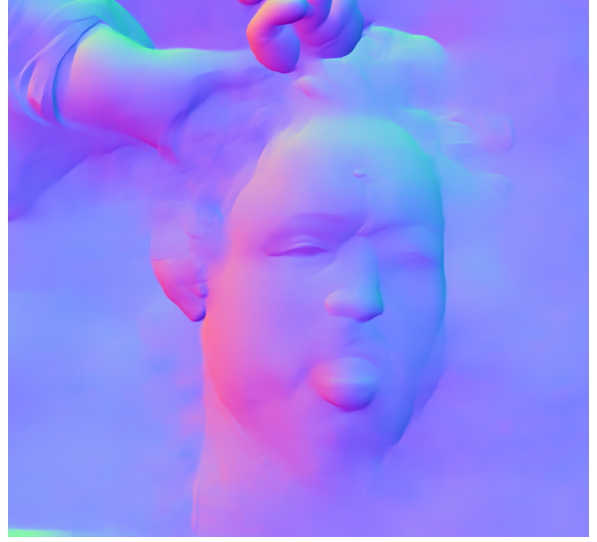
8.5.1 Independent DSINE Normal Estimation: Exp108A

Surface normals were independently estimated from the original RGB painting using DSINE [1] and resized to the working depth resolution. David and Goliath were evaluated separately. The resulting normal fields contained spatially structured orientation variation across the main facial regions.

No depth modification was introduced in Exp108A. Its purpose was to determine whether the independent orientation field contained sufficient structured variation to justify subsequent depth–normal consistency analysis.



(a) David.



(b) Goliath.

Figure 8.2: Independent DSINE surface-normal estimates for the David and Goliath facial regions. The RGB encoding represents the predicted per-pixel surface orientation and reveals structured directional variation across the main facial features.

8.5.2 Depth–Normal Consistency: Exp108B

Normals were also derived from the baseline depth field.

For depth $D(x, y)$, an approximate depth-derived normal can be written as

$$\tilde{\mathbf{n}}_D = \begin{bmatrix} -s \frac{\partial D}{\partial x} \\ -s \frac{\partial D}{\partial y} \\ 1 \end{bmatrix}, \quad (8.4)$$

where s converts normalized depth gradients into the effective local slope scale.

The normalized vector is

$$\mathbf{n}_D = \frac{\tilde{\mathbf{n}}_D}{\|\tilde{\mathbf{n}}_D\|_2}. \quad (8.5)$$

Angular disagreement between depth-derived normal $\mathbf{n}_D$ and DSINE normal $\mathbf{n}_S$ was computed as

$$\theta = \cos^{-1} \left[\text{clip} \left(\mathbf{n}_D^\top \mathbf{n}_S, -1, 1 \right) \right]. \quad (8.6)$$

The comparison revealed broad orientation agreement together with non-negligible local angular disagreement.

The independent normal field thus contained information that was not numerically identical to the baseline surface orientation.

However, disagreement alone did not establish that either representation was geometrically more accurate.

8.5.3 Bias Compensation: Exp108B2

Some of the observed disagreement was found to result from systematic orientation bias.

A global orientation compensation was therefore applied before interpreting the remaining local disagreement.

The median angular disagreement decreased from

$$27.01^\circ \rightarrow 21.55^\circ \quad (8.7)$$

for David, and

$$17.56^\circ \rightarrow 16.25^\circ \quad (8.8)$$

for Goliath.

The reduction indicated that part of the initial disagreement resulted from systematic orientation bias. The remaining difference was therefore interpreted conservatively as a mixture of potentially complementary orientation information, normal-estimation uncertainty, painterly appearance ambiguity, and residual geometric disagreement.

8.5.4 Confidence Estimation: Exp108C

Agreement and spatial support of the independent normal field were not uniform across the facial regions.

A soft confidence representation was therefore constructed.

The general confidence model can be expressed as

$$C(x, y) = C_n(x, y)C_b(x, y)C_a(x, y), \quad (8.9)$$

Here, C_n represents local depth-normal agreement, C_b suppresses uncertain boundary regions, and C_a represents anatomical or landmark-based support. The confidence field was continuous, allowing uncertain regions to contribute less without introducing a new hard spatial boundary.

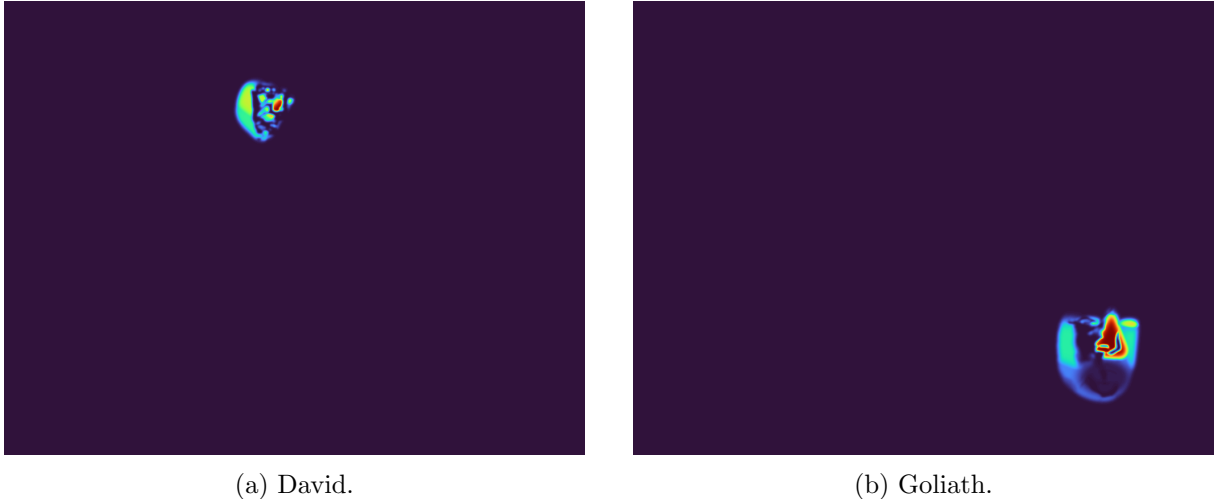


Figure 8.3: Final soft confidence fields for David and Goliath used to spatially weight the subsequent geometry-aware refinement. Higher confidence values indicate facial regions receiving stronger support, while low-confidence regions are progressively suppressed rather than excluded by a hard binary mask.

8.6 Weak Normal-Guided Refinement

After the independent normal field had been characterized, Exp109A investigated whether normal-derived orientation could be introduced as a weak local constraint.

The baseline gradients were intentionally not replaced directly.

An example formulation for energy is:

$$E(D') = \lambda_d E_{\text{data}} + \lambda_n E_{\text{normal}} + \lambda_s E_{\text{smooth}}, \quad (8.10)$$

where

$$E_{\text{data}} = \sum (D' - D)^2 \quad (8.11)$$

preserves the baseline,

$$E_{\text{normal}} = \sum C \|\nabla D' - \mathbf{g}_N\|^2 \quad (8.12)$$

encourages agreement with the normal-derived local gradient, and

$$E_{\text{smooth}} = \sum \|\nabla D'\|^2 \quad (8.13)$$

Exp109A applied the normal-derived constraint conservatively and only to David. The resulting depth changes were measurable and spatially localized, while the broad facial surface and the corresponding Blender reconstruction remained close to the inherited baseline. The experiment therefore established the feasibility of weak normal-guided correction but did not demonstrate a substantial additional surface-level effect.

8.7 Landmark- and Curvature-Guided Anatomical Refinement

The limited surface-level effect of normal guidance motivated a shift toward explicit anatomical localization. The Exp112–115 family progressively combined landmark-centred supports, curvature, confidence weighting, and multiple anatomical regions, moving from broad geometric guidance toward feature-specific correction.

8.7.1 Adaptive Anatomical Supports: Exp112

Exp112 introduced separate support functions for major facial structures in order to localize the correction more explicitly than broad normal guidance. The resulting modification was more feature-specific and spatially restricted, although increased anatomical localization did not by itself produce a substantially stronger reconstructed effect.

8.7.2 Curvature-Guided Refinement: Exp113

Exp113 incorporated local surface variation already present in the baseline.

A Difference-of-Gaussians curvature proxy was used:

$$C_\kappa = |\mathcal{G}_{\sigma_1}(D) - \mathcal{G}_{\sigma_2}(D)|. \quad (8.14)$$

The curvature signal was used as a spatial weight rather than as an independent source of geometry. It concentrated the correction where the inherited depth field already exhibited local variation, but the additional reconstructed effect remained moderate.

8.7.3 Sparse Landmark-Centered Refinement: Exp114

Exp114 further restricted the correction to a small set of anatomical structures, including the nose bridge, nose tip, upper and lower lips, and eye sockets.

A general local weighting formulation was:

$$W_f = W_{\text{landmark}} W_{\text{confidence}} W_{\text{curvature}}. \quad (8.15)$$

The resulting changes were tightly localized and the surrounding face remained close to the baseline. The Blender difference, however, was still subtle, showing that increased spatial specificity alone did not guarantee a larger surface-level modification.

8.7.4 Confidence-Relaxed Anatomical Refinement: Exp115

Exp115 investigated whether the confidence signal could be used as a soft attenuation mechanism rather than as a strict acceptance threshold. The experiment combined dense MediaPipe landmarks, landmark-based confidence, baseline curvature, and localized anatomical support functions.

Unlike the earlier stages of the pipeline, Exp115 operated on the progressively refined facial baseline

`phase2C_v4_R5_E2_balanced_elliptical_mouth_refined_uint16.tiff`,

which itself remained derived from the common DA3-based depth lineage.

Before constructing the local supports, the David and Goliath masks were eroded by five pixels:

$$M_s = \text{erode}(M, 5), \quad (8.16)$$

thereby preventing local corrections from reaching the external mask boundary.

The confidence field was normalized independently inside each valid facial support using its 99th percentile:

$$C_n = \text{clip} \left(\frac{C}{P_{99}(C \mid M_s) + \varepsilon}, 0, 1 \right). \quad (8.17)$$

A relaxed confidence gate was then defined as

$$C_r = f + (1 - f)C_n^\gamma, \quad (8.18)$$

where f specifies a non-zero confidence floor.

For David,

$$f_D = 0.35, \quad \gamma_D = 0.70, \quad (8.19)$$

whereas for Goliath,

$$f_G = 0.48, \quad \gamma_G = 0.55. \quad (8.20)$$

Consequently, low-confidence regions were attenuated rather than eliminated completely. A Difference-of-Gaussians curvature proxy was computed from the baseline depth:

$$K = |\mathcal{G}_{1.5}(D) - \mathcal{G}_{5.0}(D)|. \quad (8.21)$$

The curvature cue was normalized within the safe facial mask using its 97.5th percentile:

$$K_n = \text{clip} \left(\frac{K}{P_{97.5}(K \mid M_s) + \varepsilon}, 0, 1 \right), \quad (8.22)$$

and converted into a soft weighting term:

$$W_K = 0.72 + 0.28K_n. \quad (8.23)$$

Thus,

$$0.72 \leq W_K \leq 1.00. \quad (8.24)$$

For the localized anatomical components, the common support gate was

$$G_c = C_r W_K M_s. \quad (8.25)$$

Gaussian anatomical supports were constructed around landmark-derived centres for:

- nose bridge;
- nose tip;
- left and right cheeks;
- upper lip;
- lower lip;
- chin;
- left and right eye sockets;
- a weak global facial component.

The inter-eye distance was used as the principal relative scale for these support functions, allowing the spatial extent of the components to adapt to the apparent size of each face.

Three strength configurations, R1, R2, and R3, were evaluated. R1 was retained as the representative conservative configuration.

Its signed component amplitudes were

$$\begin{aligned}
A_{\text{global}} &= 0.00010, \\
A_{\text{bridge}} &= 0.00130, \\
A_{\text{nose}} &= 0.00170, \\
A_{\text{cheeks}} &= 0.00075, \\
A_{\text{upperlip}} &= 0.00055, \\
A_{\text{lowerlip}} &= 0.00070, \\
A_{\text{chin}} &= 0.00065, \\
A_{\text{eyes}} &= -0.00042.
\end{aligned} \tag{8.26}$$

The negative eye-socket amplitude produced a shallow recessed contribution, while the nose, cheeks, lips, and chin received positive local support.

The complete Goliath correction was scaled by

$$a_G = 0.78, \tag{8.27}$$

whereas David used

$$a_D = 1.00. \tag{8.28}$$

After summing the signed anatomical components, a support-weighted offset correction was applied to suppress unnecessary mean elevation. The result was then smoothed with a Gaussian filter using

$$\sigma = 1.10. \tag{8.29}$$

The final local modification was bounded by

$$|\Delta D| \leq a 0.00175. \tag{8.30}$$

The David and Goliath corrections were combined and explicitly set to zero outside the union of the two facial masks:

$$\Delta D(x, y) = 0, \quad (x, y) \notin M_D \cup M_G. \tag{8.31}$$

The refined depth was finally obtained as

$$D' = \text{clip}(D + \Delta D, 0, 1). \tag{8.32}$$

No additional global normalization was performed after the local correction.

Exp115 R1 therefore represents a bounded, confidence-relaxed, landmark-guided local correction rather than a new global depth reconstruction. Because no global renormalization was applied and the correction was explicitly zero outside the David and Goliath facial supports, the surrounding native depth field was preserved.

The physical magnitude of the correction is important for interpreting this result. Under the displacement mapping used in this thesis, the active-pixel mean absolute modification is approximately $1.09 \mu\text{m}$, the 99th percentile is approximately $12.26 \mu\text{m}$, and the maximum is ap-

proximately $22.30\text{ }\mu\text{m}$ (0.02230 mm). These values are all well below the approximately 0.1 mm engineering reference scale discussed in Appendix A.3.2. Consequently, Exp115 R1 should be interpreted as a conservative demonstration that confidence-, curvature-, and landmark-guided corrections can be tightly localized while preserving the inherited depth field. Its native-stage residuals do not constitute evidence of a fabrication-relevant or tactile improvement.

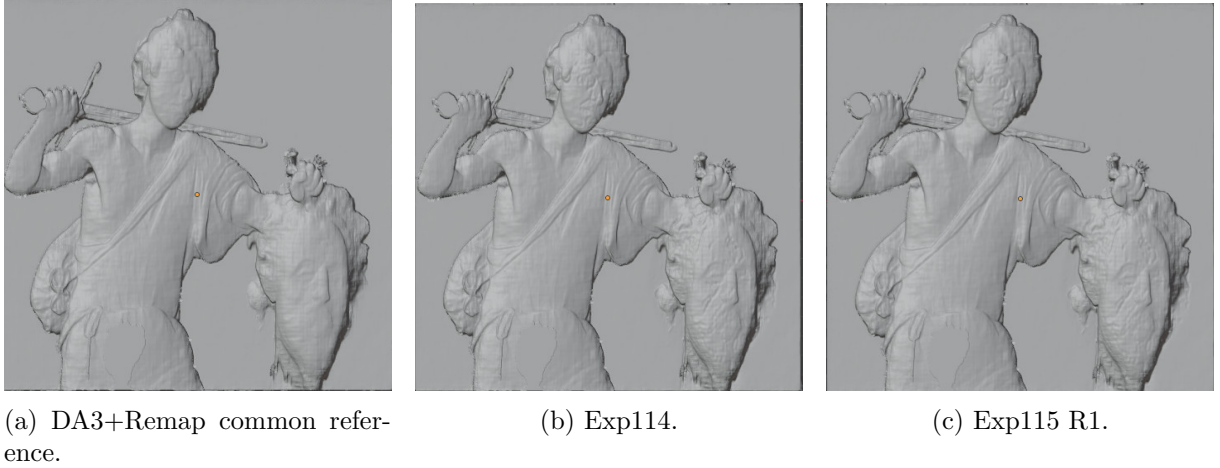


Figure 8.4: Controlled Blender comparison using DA3+Remap as a common visual reference, together with sparse landmark-guided refinement (Exp114) and confidence-relaxed anatomical refinement (Exp115 R1). Exp115 R1 was computed from the later Phase-2C facial baseline; therefore, this figure should be interpreted as a common visual comparison rather than a direct native-baseline ablation.

8.8 Mesh-Space Facial Refinement

The Exp116–120 family investigated whether changing the representation from raster depth to an explicit triangular mesh could improve local control. The motivation was representational: mesh space permits direct vertex-level manipulation, but does not by itself introduce new information about the desired facial geometry.

$$\mathbf{v}_{ij} = \begin{bmatrix} x_i \\ y_j \\ sD(x_i, y_j) \end{bmatrix}, \quad (8.33)$$

where s is the depth-to-mesh displacement scale.

8.8.1 Mesh Preparation: Exp116A–B and Exp117A

Exp116A transferred facial landmarks to the canonical mesh coordinates, while Exp116B transferred continuous anatomical support functions. Exp117A then validated the canonical depth-to-mesh conversion and established the triangular surface used in subsequent mesh-space experiments. These stages verified the representation and correspondence pipeline but did not yet introduce a stronger geometric target.

8.8.2 Direct Local Mesh Displacement: Exp118A–B

Exp118 applied direct local displacement to selected facial mesh vertices. Increasing the strength and support produced stronger localized deformation without severe visible surface discontinuities, demonstrating that direct mesh manipulation was technically feasible.

8.8.3 Anatomically Constrained Surface Relaxation: Exp119A

Exp119A propagated the local deformation to neighbouring vertices through iterative relaxation. This improved surface continuity, but the controlled Blender comparison did not show a consistent advantage over the strongest depth-domain anatomical refinements.

8.8.4 Parametric Mesh Template: Exp120A

Exp120A introduced a parameterized nose template directly in mesh space. Although the experiment demonstrated the flexibility of explicit mesh editing, the additional visible effect remained small. The mesh-space family therefore increased representational freedom without providing a new source of geometric knowledge.

8.9 Normal-Guided Surface Reconstruction

After the mesh-space experiments, Exp121 revisited the independent normal field through a more explicit reconstruction objective. Rather than applying a local additive correction, the surface was optimized jointly with baseline-fidelity, confidence-weighted normal, and smoothness constraints.

8.9.1 Confidence-Gated Surface Reconstruction: Exp121A

A representative least-squares objective can be written as

$$\begin{aligned} E(D') = & \lambda_d \sum (D' - D)^2 \\ & + \lambda_g \sum C \left[\left(\frac{\partial D'}{\partial x} - p \right)^2 + \left(\frac{\partial D'}{\partial y} - q \right)^2 \right] \\ & + \lambda_s \sum \|\nabla D'\|^2, \end{aligned} \tag{8.34}$$

where p and q are gradients derived from the DSINE normals.

Several normal guidance strengths were tested.

The optimization converged to stable surfaces across the tested guidance strengths. Stronger normal weighting increased the numerical modification but did not produce a consistent additional change in the controlled Blender reconstruction. This again separated numerical geometric modification from visible surface-level effect.

8.9.2 Depth-to-Mesh Scale Validation: Exp121B

Exp121B explicitly tested the depth-to-mesh mapping because Exp121A produced measurable numerical changes with limited visible effect. Amplified diagnostic variants confirmed that the depth correction was transferred correctly through the tested mesh and displacement pipeline.

The limited Exp121A difference was therefore not primarily explained by an incorrect depth-to-mesh scale; the applied corrections themselves remained small relative to the complete facial surface.

8.10 Component Contribution and Ablation

Exp115 combined multiple weak anatomical supports.

Finally, Exp122A–B examined whether the observed behaviour was driven by a single facial component or distributed across multiple structures.

8.10.1 Component Ablation: Exp122A

Individual support groups were removed while the remainder of the refinement remained fixed.

The tested groups included:

- global facial support;
- cheeks;
- lips;
- chin;
- eye sockets;
- nose.

Removing individual groups produced only limited changes in the complete computed refinement. No single tested component dominated the full correction field.

8.10.2 Strict Component Contribution: Exp122B

Selected anatomical groups were then reconstructed independently.

The tested combinations included:

- nose only;
- eye sockets only;
- cheeks only;
- nose + eye sockets;
- nose + cheeks;
- nose + eye sockets + cheeks.

No tested isolated component or combination reproduced the complete Exp115 correction. Within the algorithmic refinement field, the result was therefore distributed across several weak anatomical supports rather than dominated by one component. This finding concerns the computed geometric correction and should not be interpreted as evidence of tactile contribution.

8.11 Representative Experimental Summary

Table 8.2 summarizes the methodological progression of the geometry-aware phase.

Table 8.2: Representative progression of the geometry-aware facial-refinement experiments.

Experiment	Method family	Primary objective	Main observation
Exp108A	Normal estimation	Estimate independent local orientation	Produced spatially organized DSINE normal variation suitable for consistency analysis.
Exp108B/B2	Depth-normal consistency	Measure and compensate disagreement	Showed broad orientation agreement and reduced part of the systematic angular disagreement after compensation.
Exp108C	Confidence modelling	Estimate local reliability	Produced soft spatial confidence fields for later geometry-aware processing.
Exp109A	Weak normal guidance	Introduce independent orientation conservatively	Produced measurable local depth changes but limited additional visible Blender effect.
Exp112–113	Adaptive anatomical support	Combine landmarks and curvature	Improved spatial targeting while final-surface changes remained moderate.
Exp114	Sparse landmark refinement	Restrict modification to selected facial structures	Produced highly localized modification with limited visible disturbance outside the selected supports.
Exp115	Confidence-relaxed refinement	Test bounded landmark-, confidence-, and curvature-guided correction	Produced a tightly localized and baseline-preserving correction, but its physical magnitude remained well below the approximately 0.1 mm engineering reference scale.
Exp116–117	Mesh preparation	Establish image-to-mesh anatomical correspondence	Validated canonical mesh construction and continuous mesh supports.
Exp118–120	Mesh-space refinement	Test direct deformation, relaxation, and templates	Established explicit mesh-space editing capability but no clear qualitative advantage over depth-space refinement.

Continued on next page

Experiment	Method family	Primary objective	Main observation
Exp121A	Normal surface reconstruction	Integrate normal-derived gradients through constrained optimization	Produced smooth reconstructed surfaces with limited additional visible change.
Exp121B	Scale validation	Verify depth-to-mesh mapping	Supported that the limited visible difference was not attributable to the tested depth-to-mesh scale factor.
Exp122A–B	Component analysis	Measure anatomical contribution	Showed that the computed Exp115 correction was distributed across multiple weak anatomical components rather than dominated by a single tested support.

8.12 Discussion

The geometry-aware experiments clarify the difference between complementary geometric information and directly recoverable surface shape. Independently estimated DSINE normals contained structured orientation information that differed from the orientation implied by the depth field, but depth–normal disagreement could not be interpreted automatically as missing or incorrect geometry. Bias compensation and confidence weighting were therefore necessary before using the independent orientation field as a local constraint.

Normal-derived guidance was most useful when applied conservatively. Both weak normal refinement and constrained normal-based reconstruction produced stable and measurable changes, but their reconstructed surface effects remained limited. Curvature behaved similarly: it was useful as a spatial weighting cue for existing local structure, but did not independently determine missing facial anatomy.

Landmarks provided more explicit anatomical localization than broad orientation cues. The Exp112–115 progression showed that feature-specific supports could be bounded and confined to selected facial structures while preserving the surrounding inherited depth field. Exp115 R1 is the clearest example of this control, but its physical magnitude was extremely small: even its maximum native-stage displacement-equivalent residual was only approximately 0.0223 mm. It should therefore be interpreted as evidence of localization and baseline preservation rather than tactile or fabrication-scale enhancement.

Moving from depth space to mesh space increased representational flexibility but did not add independent knowledge about the correct facial geometry. Similarly, stronger optimization or normal integration could not compensate for weak or uncertain target information. The limiting factor was therefore not simply the representation in which the correction was applied, but the reliability and geometric content of the target signal itself.

The component analysis further indicated that the computed landmark-based correction was distributed across multiple weak supports rather than dominated by one anatomical feature. Taken together, these experiments reinforce a thesis-wide observation: greater algorithmic

complexity or larger numerical modification does not necessarily produce a proportionally larger or more meaningful shallow-relief change.

The geometry-aware phase remained limited by the use of normal estimation on pictorial input, uncertainty in interpreting depth–normal disagreement, simplified landmark-guided supports, evaluation on a single painting, and the absence of physical tactile validation. The results therefore support conclusions about geometric localization, bounded correction, surface continuity, and pipeline stability, but not about tactile recognition or perceptual effectiveness.

8.13 Chapter Summary

This chapter investigated whether geometric information beyond the inherited depth field could provide additional control over local facial refinement. Independently estimated surface normals supplied complementary orientation information, but the depth–normal analysis showed that such disagreement must be interpreted cautiously because systematic bias, estimation uncertainty, and painterly appearance can contribute to the observed differences.

Confidence, curvature, and anatomical landmarks progressively improved the spatial control of the correction. The principal methodological gain was therefore the ability to restrict and organize local modification rather than the recovery of substantial new facial geometry. Exp115 R1 demonstrated particularly strict localization and exact preservation outside the selected facial supports, but its native-stage physical magnitude remained far below the approximately 0.1 mm engineering reference scale. It consequently serves as evidence of controlled geometry-aware correction rather than fabrication- or tactile-relevant improvement.

The mesh-space and constrained surface-reconstruction experiments reached a related conclusion. Changing the geometric representation or increasing optimization complexity improved flexibility and continuity, but did not supply missing information about the correct local surface. Numerical modification and final reconstructed effect therefore remained distinct quantities throughout this phase.

Overall, the geometry-aware experiments showed that reliable local refinement depends more strongly on the quality and geometric meaning of the available guidance than on the complexity of the representation used to apply it. Chapter 9 therefore compares these findings with the preceding experimental families and identifies the methodological tendencies that persisted across the complete thesis.

Chapter 9

Overall Experimental Comparison and Discussion

9.1 Introduction

The preceding experimental chapters evaluated progressively more structured strategies for generating and refining a shallow tactile-oriented bas-relief from a single painting. The investigation began with an image-aligned monocular depth representation and subsequently tested external three-dimensional geometry, semantic localization, parametric anatomical priors, independently estimated surface normals, landmarks, confidence, curvature, and explicit mesh-space processing.

The purpose of this chapter is not to introduce another refinement method, but to synthesize the evidence across these experimental families. The comparison focuses on three recurring questions: how reliably the global image-derived geometry was preserved, how explicitly local facial structures could be controlled, and whether increased geometric complexity produced a meaningful additional effect in the reconstructed shallow relief.

A recurring observation across the thesis was that numerical modification and final-surface change were not directly proportional. The following comparison therefore considers both the geometric correction itself and its behaviour after displacement-based reconstruction.

9.2 Evolution of the Research Strategy

The experimental progression can be viewed as a gradual transition from general image-derived geometry toward increasingly localized and structured geometric control.

The principal stages were:

$$\begin{aligned}
&\text{monocular depth estimation} \\
&\rightarrow \text{depth preparation and remapping} \\
&\rightarrow \text{external geometry integration} \\
&\rightarrow \text{semantic local refinement} \\
&\rightarrow \text{parametric anatomical priors} \\
&\rightarrow \text{surface-normal and confidence analysis} \\
&\rightarrow \text{landmark-guided refinement} \\
&\rightarrow \text{mesh-space and constrained reconstruction.}
\end{aligned} \tag{9.1}$$

The progression was driven by limitations observed experimentally rather than by a desire to increase algorithmic complexity. The initial depth stage established a consistent image-aligned global reference. External geometry tested whether additional three-dimensional information could recover weak local structure, but exposed registration and compatibility problems. Semantic processing therefore shifted the emphasis from geometry replacement to localized baseline modification. Parametric priors and landmarks subsequently introduced structure-level control, while surface normals, confidence, curvature, and mesh-space processing examined whether additional geometric representations could provide further guidance.

The resulting progression can therefore be interpreted as a systematic shift from *adding more geometry* toward *controlling more carefully where and how additional geometry is allowed to influence the inherited surface*.

9.3 Cross-Family Methodological Synthesis

The experimental families differed primarily in the source and role of the additional geometric information. DA3+Remap provided the common image-aligned global reference from which the baseline-preserving investigation developed. Its main strength was stable image-to-depth correspondence and direct compatibility with the shallow-relief reconstruction pipeline, while its principal limitation was insufficient differentiation of some small facial structures after depth compression.

External geometry addressed this limitation by introducing independently reconstructed three-dimensional information. Although this increased the amount of available local geometry, it also introduced scale, projection, camera, registration, and representation compatibility requirements. The resulting experiments demonstrated that successful spatial alignment was not sufficient to guarantee relief-compatible local geometry:

$$\boxed{\text{registration quality} \neq \text{relief compatibility.}} \tag{9.2}$$

Semantic refinement avoided this correspondence problem by operating directly within the image-aligned depth lineage. It provided explicit control over where modification was allowed and enabled boundary-aware and interior-weighted processing. Its limitation was complementary to that of external geometry: semantic information localized the correction but did not specify the internal anatomy of the selected region,

$$\boxed{\text{semantic localization} \neq \text{anatomical specification.}} \tag{9.3}$$

Parametric priors and anatomical landmarks therefore shifted the problem from region-level toward structure-level control. Signed components allowed protruding and recessed facial structures to be treated separately, but increasing target strength or representation complexity produced diminishing surface-level returns. Surface normals introduced independent orientation information and were useful for consistency analysis, confidence estimation, and weak geometric guidance, but did not independently specify absolute facial shape.

Finally, explicit mesh-space processing increased representational flexibility through direct vertex manipulation, relaxation, and template deformation. It did not, however, provide new information about what the correct local facial geometry should be. Across these families, the main methodological transition was therefore from independent geometry replacement toward baseline-preserving, increasingly localized geometric control.

9.4 Cross-Family Comparison

Table 9.1 summarizes the principal characteristics observed across the main experimental families.

The descriptors are qualitative and refer specifically to the behaviour observed within the experimental setup of this thesis.

They should not be interpreted as standardized perceptual scores.

Table 9.1: Descriptive cross-family comparison of the principal experimental strategies. The entries summarize the behaviour observed within the experimental setup of this thesis and do not represent standardized perceptual or quantitative scores.

Experimental family	Global-reference behaviour	Type of local control	Main integration requirement	Principal observation
DA3+Remap	Used as the common global geometric reference	Limited intrinsic control over individual facial structures	No external geometric registration required	Provided consistent global image-depth correspondence within the evaluated pipeline, while several local facial structures remained weak after shallow-relief compression.
External geometry fusion	Dependent on successful integration with the baseline	Independent local three-dimensional geometry	Scale, translation, projection, camera correspondence, and depth-range alignment	Introduced additional geometric information, but registration and representation mismatch limited consistent integration with the fixed baseline.

Continued on next page

Experimental family	Global-reference behaviour	Type of local control	Main integration requirement	Principal observation
Semantic refinement	DA3-derived geometry retained as the global reference while modification was localized through semantic supports	Region-level local depth control	Semantic-mask definition, boundary handling, and local depth transformation	Provided direct spatial control over selected regions, but semantic localization alone did not specify their internal anatomical geometry.
Parametric anatomical priors	Baseline retained while procedural local structures were added	Structure-level control through signed anatomical components	Definition and tuning of anatomical primitives, supports, and amplitudes	Allowed protruding and recessed facial structures to be treated separately, but increasing prior strength produced diminishing visible returns.
Normal-guided refinement	Baseline retained as the principal depth reference	Orientation-guided local refinement	Depth-normal consistency, bias compensation, confidence weighting, and gradient scaling	Provided independent surface-orientation information and useful diagnostic guidance, while the additional visible final-surface effect remained limited.
Landmark-guided refinement	Baseline retained while modification was restricted to localized anatomical supports	Sparse anatomy-specific local control	Landmark placement, support design, curvature weighting, and confidence control	Provided explicit anatomical localization and tightly bounded correction; the later Exp115 R1 configuration was physically very small and should not be interpreted as evidence of tactile improvement.

Continued on next page

Experimental family	Global-reference behaviour	Type of local control	Main integration requirement	Principal observation
Mesh-space refinement	Baseline depth converted into an explicit canonical mesh	Direct vertex- and surface-level control	Image-to-mesh correspondence, mesh construction, deformation, and relaxation	Enabled explicit surface manipulation and greater representation flexibility, but did not show a consistent qualitative advantage over the controlled depth-domain methods.

9.5 Representative Cross-Stage Comparison

A smaller set of representative configurations was used to illustrate the principal methodological transitions. These configurations are compared descriptively rather than ranked numerically because no standardized perceptual or tactile evaluation was performed.

The comparison considers preservation of the inherited global geometry, localization and magnitude of the correction, transition behaviour, and the principal limitation of each stage. Table 9.2 summarizes these observations.

The representative stages show a progressive change in the role of additional geometric information. External reconstruction increased geometric richness but introduced compatibility requirements. Semantic processing reduced these integration problems by modifying the existing depth lineage directly, while parametric and landmark-guided stages provided increasingly explicit local control.

The later stages should not, however, be interpreted as a monotonic sequence of improving reliefs. Increasing anatomical specificity or numerical correction did not consistently produce a proportional surface-level effect. Exp115 R1 is an especially conservative example: its value lies in demonstrating bounded localization rather than in establishing a perceptually meaningful geometric improvement.

The recurring trade-off was therefore

$$\boxed{\text{increasing local geometric control} \leftrightarrow \text{retaining global surface consistency.}} \quad (9.4)$$

The experiments progressively improved control over where and how correction was applied, while the magnitude of reliable additional geometry remained constrained by the available image evidence and the shallow-relief representation.

9.6 Final Representative Visual Comparison

Figure 9.1 provides a common visual comparison of representative stages: DA3 Raw, DA3+Remap, external-geometry fusion, semantic refinement, parametric anatomical refinement, and Exp115

Table 9.2: Qualitative comparison of representative configurations from the principal experimental stages. The descriptors summarize behaviour observed within the present experimental setup and should not be interpreted as standardized perceptual scores.

Method	Global-reference behaviour	Local facial effect	Transition behaviour	Principal limitation
DA3+Remap	Used as the common global geometric reference	Limited intrinsic differentiation of individual facial structures	Broad surface continuity retained	Several local facial structures remain comparatively compressed.
External geometry	Dependent on successful integration with the baseline	Potentially stronger local geometric modification	Sensitive to registration, projection, and fusion mismatch	Registration, projection, and geometric compatibility with the baseline remain difficult to control.
Exp80A Exp85A	/ DA3-derived representation retained as the global reference while semantic modification was localized to selected supports	Moderate region-level facial modification	Smooth local transitions with limited visible boundary disturbance	Semantic localization provides spatial control but does not define internal anatomical geometry.
Parametric refinement	Baseline retained while local anatomical priors are introduced	Localized and structure-specific modification	Transitions controlled through smooth support functions	Increasing prior strength produces diminishing visible returns and may introduce artificial local structure.
Exp115 R1	Inherited DA3-derived facial baseline preserved outside the selected supports	Tightly localized correction distributed across several anatomical supports	Smoothly weighted and explicitly bounded local transitions	Maximum native-stage displacement-equivalent residual was approximately 0.0223 mm; therefore, the experiment demonstrates localization and baseline preservation rather than a fabrication- or tactile-relevant improvement.

R1. All renders use the same Blender visualization and displacement configuration so that rendering differences do not confound the geometric comparison.

Because several later experiments inherited stage-specific DA3-derived facial baselines, the figure should be interpreted as a common-reference visual comparison rather than as a single native-baseline ablation sequence.

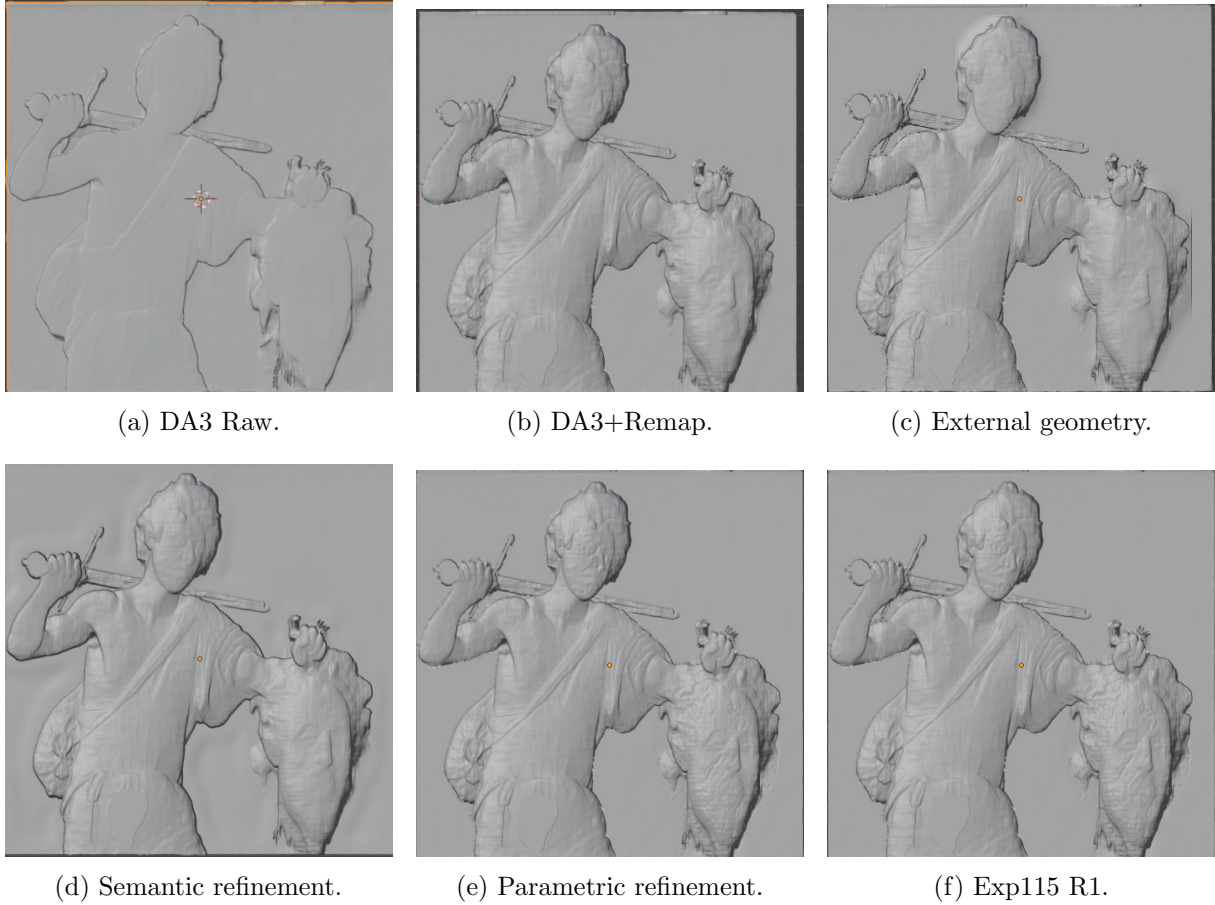


Figure 9.1: Common-reference Blender comparison of representative stages in the experimental progression: DA3 Raw, DA3+Remap, external-geometry fusion, semantic refinement, parametric anatomical refinement, and Exp115 R1. All panels use the same visualization and displacement configuration. The later configurations may inherit stage-specific DA3-derived baselines; therefore, the comparison summarizes the experimental progression rather than representing a direct native-baseline ablation.

9.7 Stability versus Local Modification

A recurring trade-off across the external-geometry, semantic, parametric, landmark-guided, and mesh-space experiments concerned the magnitude of local modification relative to the stability of the surrounding surface. Weak corrections generally preserved the inherited geometry, whereas stronger or broader corrections became more susceptible to bumps, depressions, halos, abrupt transitions, or disagreement with neighbouring surface structure.

This behaviour was not specific to one algorithmic family. Within the evaluated pipeline it reflected the general difficulty of introducing additional local geometry into an already established and strongly depth-compressed surface. Increasing correction strength was therefore not,

by itself, a reliable strategy for improving the final relief.

9.8 Numerical Change versus Final Relief Change

Across the experimental chapters, numerical residuals were useful for verifying whether a method modified the intended region and for quantifying the magnitude of that modification. They did not, however, provide a direct measure of final-relief quality or tactile relevance.

The reconstructed effect also depended on the spatial extent and frequency of the correction, local surface slope and curvature, transition width, displacement mapping, mesh resolution, and visualization conditions. Broad low-frequency changes could remain visible after displacement even when their normalized amplitude was moderate, whereas narrow high-frequency corrections could produce measurable residuals while remaining difficult to distinguish on the reconstructed surface.

Exp115 R1 illustrates the importance of this distinction. Although the correction was spatially controlled and measurable, its active-pixel mean displacement-equivalent magnitude was approximately 0.00109 mm, its 99th percentile approximately 0.01226 mm, and its maximum approximately 0.02230 mm. These values are all below the approximately 0.1 mm engineering reference scale used in this thesis. The Exp115 result therefore supports conclusions about bounded localization and preservation of the surrounding depth field, not fabrication-scale or tactile improvement.

Numerical residuals should consequently be interpreted together with their spatial support, physical mapping, and final reconstructed behaviour rather than as standalone evidence of improvement.

9.9 Role of the Different Geometric Cues

The evaluated cues played complementary rather than interchangeable roles.

9.10 Approaches with Limited Advantage under the Evaluated Conditions

Several negative results were methodologically important. Strong external-geometry replacement was limited by registration and representation compatibility; aggressive local inflation increased the risk of artificial bumps; hard semantic boundaries produced discontinuities; stronger parametric priors produced diminishing visible returns; direct normal integration was sensitive to bias and uncertainty; and explicit mesh-space processing increased flexibility without demonstrating a consistent surface-level advantage over the controlled depth-domain pipeline.

These results progressively shifted the investigation away from stronger geometric substitution and toward bounded, baseline-preserving correction. Their importance lies in identifying which additional forms of complexity did not resolve the central local-geometry problem under the evaluated conditions.

Table 9.3: Functional role and principal limitation of the geometric cues used in the thesis.

Cue	Primary role	Principal limitation
Depth	Global image-aligned surface and depth ordering	Small anatomical structures may remain weak after shallow-relief compression.
Semantic supports	Region-level localization of allowed modification	Do not specify internal anatomical geometry.
Parametric priors	Explicit hypotheses for protruding and recessed local structures	Depend on simplified, manually or heuristically defined target geometry.
Surface normals	Independent local orientation and consistency information	Do not provide absolute surface position and may be affected by pictorial ambiguity.
Landmarks	Sparse anatomical localization for structure-specific supports	Localize features but do not independently determine their correct three-dimensional shape.
Curvature	Weighting of existing local surface variation	Reflects the inherited surface rather than reconstructing missing anatomy.
Confidence	Controls the influence of uncertain auxiliary cues	Modulates other information but does not generate geometry itself.

9.11 Overall Interpretation

The complete investigation indicates that the principal challenge was not simply obtaining more geometric information, but determining which information could be integrated without destabilizing an already useful image-aligned global surface. Each auxiliary representation addressed a different part of this problem: external geometry supplied independent shape, semantics supplied region-level localization, parametric priors supplied explicit geometric hypotheses, normals supplied orientation, landmarks supplied anatomical anchors, and mesh space supplied representational flexibility.

None of these sources independently resolved the missing local-geometry problem. The experimental progression therefore became increasingly conservative and supports the general strategy

retain the global baseline geometry + introduce weak, localized, controlled geometric corrections.

(9.5)

This formulation should not be interpreted as recovery of the true three-dimensional anatomy. It summarizes the strategy that most consistently preserved the inherited shallow-relief surface while permitting controlled local experimentation within the present case study.

9.12 Remaining Limitations

The investigation remains constrained by the ambiguity of single-image geometry and by the use of pictorial rather than photographic input. Painterly shading, texture, artistic simpli-

fication, and occlusion can influence depth, normal, and feature estimation, while the true three-dimensional surface is unavailable as ground truth.

The final relief also requires strong depth compression, so locally measurable geometric changes may become very small after displacement conversion. This issue is particularly important for late-stage conservative refinements such as Exp115 R1. The present evaluation therefore supports claims about geometric localization, continuity, residual magnitude, and pipeline stability, but not about improved tactile recognition.

In addition, the detailed experiments were concentrated on a single challenging painting and no formal tactile user study or systematic fabrication-based evaluation was performed. Generalization to other artistic styles, compositions, subjects, fabrication processes, and tactile users requires a broader experimental dataset and physical validation.

9.13 Principal Findings Across the Complete Investigation

The cross-family comparison supports five principal findings:

1. A consistent image-aligned global reference was essential for controlled experimentation. Within the present pipeline, the DA3-derived depth lineage provided this reference while later stages introduced increasingly localized modifications.
2. Additional geometric information was useful only when it was compatible with the inherited representation. External geometry exposed registration and projection problems, while semantics and landmarks improved localization without independently determining complete local anatomy.
3. Greater geometric specificity did not imply proportionally greater surface-level benefit. Parametric, normal-guided, landmark-guided, and mesh-space experiments repeatedly showed diminishing returns when correction strength or representational complexity increased.
4. Numerical residual magnitude and final reconstructed effect must be interpreted separately. A measurable correction can remain physically small or visually subtle after displacement; Exp115 R1 provides a clear example of this distinction.
5. The most robust methodological tendency was baseline preservation: retain the global image-derived geometry and introduce additional information only through localized, smoothly weighted, and explicitly bounded corrections.

9.14 Chapter Summary

This chapter synthesized the experimental evidence across the complete refinement pipeline. The progression from monocular depth to external geometry, semantic localization, parametric priors, normals, landmarks, and mesh-space processing did not reveal a monotonic relationship between algorithmic complexity and final-relief improvement.

Instead, the experiments exposed three persistent requirements: the need for a stable image-aligned global reference, the need to localize additional geometric information carefully, and the need to distinguish numerical modification from physically meaningful surface change. More explicit geometric representations generally increased control over where and how a correction

could be applied, but they did not independently resolve the ambiguity of the desired local three-dimensional shape.

The resulting methodological conclusion is therefore baseline-preserving rather than replacement-oriented. Additional information is most defensibly used as constrained local guidance while the inherited DA3-derived depth lineage remains responsible for the global composition and viewpoint. This conclusion concerns geometric control and reconstruction stability; the thesis does not establish improved tactile recognition without physical fabrication and dedicated user evaluation.

Chapter 10 consolidates these findings, states the contributions supported by the experimental evidence, and identifies the principal directions for future work.

Chapter 10

Conclusions and Future Work

10.1 Overall Conclusion

This thesis investigated how image-derived geometry can be transformed into a shallow, viewpoint-preserving bas-relief of pictorial art while allowing controlled local refinement of geometrically important structures. The central difficulty was not simply the estimation of more three-dimensional information, but the integration of that information into a strongly depth-compressed surface without disturbing the global composition and image alignment.

The experiments showed that geometric richness, numerical modification magnitude, and usefulness in the final shallow relief are distinct properties. Independently reconstructed geometry could provide additional three-dimensional structure but introduced registration and representation-compatibility problems. Semantic, parametric, normal-, landmark-, and mesh-based methods progressively increased local control, yet greater geometric specificity or algorithmic complexity did not produce a proportional increase in final-surface change.

Across the investigated families, the most consistent methodological tendency was therefore baseline preservation:

retain the global image-derived geometry + introduce localized, constrained refinement.

(10.1)

Within the present case study, the DA3-derived depth lineage provided the global geometric reference, while auxiliary information was most useful when treated as local guidance rather than as an unrestricted replacement for the surface. This conclusion concerns geometric control, reconstruction stability, and compatibility with the shallow-relief representation. It does not establish recovery of the true three-dimensional anatomy or improved tactile recognition, both of which would require additional physical and perceptual validation.

10.2 Representative Final Pipeline

The complete experimental investigation produced a representative baseline-preserving computational pipeline for the present case study. Its purpose is to document the sequence of operations that led to the retained final configuration, rather than to claim that every late-stage correction is perceptually or fabrication-relevant.

The pipeline retains the DA3-derived depth representation as the global geometric reference,

applies the selected semantic facial refinement, and finally introduces the conservative Exp115 R1 correction. The latter is included because it demonstrates bounded geometry-aware localization and exact preservation outside its selected supports; its physical magnitude is too small to constitute evidence of tactile or fabrication-scale improvement.

Algorithm 1: Representative baseline-preserving bas-relief generation pipeline

Require: RGB painting I

- 1: Estimate monocular depth using DA3-LARGE-1.1 with processing resolution 1536 and `upper_bound_resize`.
- 2: Align the selected 1536-resolution DA3 depth representation to the 1022×798 working grid associated with the 1024-reference representation.
- 3: Apply the custom PPD-inspired multistage depth refinement to this pre-aligned 1536-to-1024 depth representation.
- 4: Apply robust relief-oriented remapping using `P_LOW`= 0.5, `P_HIGH`= 99.5, `HIST_BINS`= 256, and `SMOOTH_K`= 21.
- 5: Restrict the search for the dominant intermediate-depth histogram mode to $[0.18, 0.82]$.
- 6: Apply the adaptive mid-range redistribution using `BAND_SIGMA`= 0.12, `BAND_STRENGTH`= 0.42, `BASE_GAIN`= 5.5, `BAND_GAIN`= 10.5, and `CENTER_BIAS`= -0.015 .
- 7: Re-inject controlled local detail using `DETAIL_SIGMA`= 1.0, `DETAIL_WEIGHT`= 0.05, and `DETAIL_MAX`= 0.05.
- 8: Apply the final gamma transformation using `FINAL_GAMMA`= 0.97.
- 9: Generate and validate the semantic supports for David’s face and Goliath’s head, and align them to the working depth resolution.
- 10: Apply the selected Phase-2B semantic reinjection using the `R5_balanced` configuration with

$$a_{\text{fine}} = 0.26, \quad a_{\text{mid}} = 0.32, \quad a_{\text{large}} = 0.11, \quad a_{\text{contrast}} = 0.055, \quad a_{\text{ridge}} = 0.030.$$

- 11: Apply the selected Phase-2C Version-4 localized Goliath-mouth refinement to the Phase-2B R5 output using the `E2_balanced` configuration.
- 12: Define the Goliath mouth-support ellipse relative to the Goliath head-mask bounding box:

$$c_x = x_1 + 0.61w, \quad c_y = y_1 + 0.65h, \quad a_x = 0.16w, \quad a_y = 0.10h.$$

- 13: For the Phase-2C E2 balanced configuration, use mouth-region suppression 0.40 and local smoothing strength 0.18.
- 14: Use the resulting `phase2C_v4_R5_E2_balanced_elliptical_mouth_refined_uint16.tiff` as the input to the final geometry-aware refinement.
- 15: Detect dense facial landmarks using MediaPipe Face Landmarker.
- 16: Construct the landmark-based confidence field used by the geometry-aware refinement.
- 17: Erode the David and Goliath facial masks by 5 pixels to define the safe supports M_s .

18: Normalize the confidence field independently inside each safe support:

$$C_n = \text{clip} \left(\frac{C}{P_{99}(C \mid M_s) + \varepsilon}, 0, 1 \right).$$

19: Construct the relaxed confidence gate:

$$C_r = f + (1 - f)C_n^\gamma.$$

20: Use $f_D = 0.35$ and $\gamma_D = 0.70$ for David, and $f_G = 0.48$ and $\gamma_G = 0.55$ for Goliath.

21: Compute the Difference-of-Gaussians curvature proxy:

$$K = |\mathcal{G}_{1.5}(D) - \mathcal{G}_{5.0}(D)|.$$

22: Normalize curvature within each safe support using its 97.5th percentile:

$$K_n = \text{clip} \left(\frac{K}{P_{97.5}(K \mid M_s) + \varepsilon}, 0, 1 \right).$$

23: Compute the curvature weight:

$$W_K = 0.72 + 0.28K_n.$$

24: Form the common local support gate:

$$G_c = C_r W_K M_s.$$

25: Construct landmark-centred anatomical supports using the inter-eye distance as the relative spatial scale.

26: Apply the Exp115 R1 component amplitudes:

$$\begin{aligned} A_{\text{global}} &= 0.00010, \\ A_{\text{bridge}} &= 0.00130, \\ A_{\text{nose}} &= 0.00170, \\ A_{\text{cheeks}} &= 0.00075, \\ A_{\text{upperlip}} &= 0.00055, \\ A_{\text{lowerlip}} &= 0.00070, \\ A_{\text{chin}} &= 0.00065, \\ A_{\text{eyes}} &= -0.00042. \end{aligned}$$

27: Apply the person-dependent scales $a_D = 1.00$ and $a_G = 0.78$.

28: Apply support-weighted offset removal to suppress unnecessary mean elevation.

29: Smooth the combined local correction using a Gaussian filter with $\sigma = 1.10$.

30: Bound the final local modification by

$$|\Delta D| \leq a \, 0.00175.$$

- 31: Explicitly set $\Delta D(x, y) = 0$ outside the union of the David and Goliath facial supports.
- 32: Compute the final depth representation:

$$D' = \text{clip}(D + \Delta D, 0, 1).$$

- 33: Do not apply any additional global normalization after the final local correction.
- 34: Export the resulting depth map as an unsigned 16-bit TIFF.
- 35: Reconstruct and inspect the bas-relief using the fixed Blender displacement configuration reported in Appendix A.

The Exp115 R1 stage should be interpreted as a bounded geometry-aware correction rather than as a fabrication- or tactile-relevant enhancement. Under the adopted displacement mapping, the active-pixel mean displacement-equivalent residual is approximately 0.00109 mm, the 99th percentile is approximately 0.01226 mm, and the maximum is approximately 0.02230 mm, all below the approximately 0.1 mm engineering reference used in this thesis.

10.2.1 Automatic and Manually Controlled Components

Most stages of the representative pipeline are deterministic once the required semantic supports have been defined. These include DA3 inference, pre-alignment to the working grid, the custom PPD-inspired refinement, relief-oriented remapping, Phase-2B/Phase-2C processing, MediaPipe landmark detection, confidence and curvature computation, bounded Exp115 R1 correction, and depth export.

The principal manually controlled component in the present implementation is the definition and validation of the final semantic facial supports. The final Version-6 David and Goliath masks used in the retained pipeline were manually controlled polygon-based supports, developed after the preceding automatic-segmentation experiments. They therefore provide stable spatial constraints rather than fully automatic anatomical segmentations.

The dense landmarks used in the final Exp115 R1 configuration are obtained automatically using MediaPipe Face Landmarker. In the implementation used in this thesis, the face crops are derived from the corresponding semantic masks, and the detected landmarks are mapped back to the complete image coordinate system before being rescaled to the working depth resolution. The earlier semi-manual landmarks used by the parametric experiments are therefore not required by the final Exp115 R1 refinement.

The Phase-2C mouth correction is defined relative to the Goliath head-mask bounding box. Consequently, once the semantic support has been established, the location and scale of the elliptical mouth region are determined from the mask geometry rather than by manual point placement.

The final Blender reconstruction is performed using the fixed reconstruction configuration reported in Appendix A. Its role is to provide a consistent surface-reconstruction and inspection environment rather than to introduce an additional optimization stage.

10.2.2 Rationale for the Representative Configuration

The representative configuration reflects the principal methodological conclusion of the thesis: auxiliary geometric information was most controllable when introduced as a local constraint

rather than as a replacement for the complete image-derived surface.

External geometry exposed registration and compatibility problems; semantic processing improved spatial localization but did not define internal anatomy; parametric priors provided more explicit structural hypotheses but showed diminishing returns; and normal- and mesh-based processing increased geometric guidance or representational flexibility without consistently producing stronger final-surface changes.

The retained pipeline therefore combines an image-aligned DA3-derived baseline with the selected Phase-2B R5 / Phase-2C E2 facial refinement and the conservative Exp115 R1 geometry-aware correction. Exp115 R1 is retained as a reproducible example of tightly bounded local control, not because its displacement magnitude demonstrates fabrication-scale or tactile benefit.

This pipeline should therefore be interpreted as the representative configuration supported by the present case study rather than as a universally optimal bas-relief-generation method.

10.3 Answers to the Research Questions

The experimental findings allow the research questions introduced in Chapter 1 to be addressed directly.

10.3.1 Can a stable shallow bas-relief be generated from a single painting using image-derived depth?

Within the investigated case study, yes. The selected DA3-derived representation provided a consistent image-aligned global surface that preserved the composition and major spatial relationships of the painting and was compatible with the fixed displacement-based reconstruction pipeline. The principal remaining limitations concerned local geometric differentiation rather than the broad global organization.

10.3.2 Can external three-dimensional geometry improve locally weak facial regions?

External geometry can introduce additional local three-dimensional structure, but the experiments did not show that this information could be integrated consistently enough to justify unrestricted geometric replacement. Registration, projection, scale, and representation compatibility remained critical constraints.

10.3.3 Can semantic information improve local refinement while preserving global geometry?

Semantic supports provided effective region-level localization and enabled boundary-aware modification within the inherited depth representation. They addressed where refinement should occur, but did not independently specify the desired internal geometry of the selected facial regions.

10.3.4 Can anatomical priors provide more structured local control than semantic masks alone?

Yes, in the sense of geometric controllability. Signed and landmark-centred anatomical priors allowed individual protruding and recessed structures to be treated separately. However, greater prior strength did not produce proportional surface-level improvement, and the procedural priors remained simplified geometric hypotheses rather than recovered anatomy.

10.3.5 Do independently estimated normals provide useful complementary geometry?

The DSINE normal field provided complementary local orientation information and was useful for consistency analysis, confidence estimation, and weak geometric guidance. After bias compensation, remaining disagreement was treated conservatively because it could reflect both additional orientation information and estimation uncertainty. Normal guidance did not independently produce substantial additional facial structure.

10.3.6 Do landmarks and mesh-space processing provide additional advantages?

Landmarks provided more explicit anatomical localization, while mesh-space processing provided greater representational flexibility. Neither source independently determined the correct missing facial geometry, and the mesh-space experiments did not demonstrate a consistent advantage over controlled depth-domain refinement.

10.3.7 What general strategy was most consistent across the experiments?

The most consistent methodological tendency was to retain the image-derived global geometry and introduce auxiliary information only through localized, smoothly weighted, and bounded corrections. This conclusion concerns geometric stability and controllability rather than validated tactile effectiveness.

10.4 Main Contributions

The contributions of this thesis are methodological and experimental rather than claims of perceptually validated tactile improvement. Four contributions are supported by the experimental evidence:

1. **A controlled image-to-bas-relief experimental framework.** The thesis establishes an image-aligned workflow connecting monocular depth estimation, custom depth preparation and remapping, unsigned 16-bit depth representation, and fixed Blender displacement conditions for controlled geometric comparison.
2. **A systematic investigation of complementary local geometric cues.** External three-dimensional reconstruction, semantic supports, parametric and signed anatomical priors, surface normals, confidence, landmarks, curvature, and mesh-space operations were evaluated within the same baseline-preserving research framework.

3. **Characterization of recurring failure modes and trade-offs.** The experiments identify registration and projection mismatch, semantic-boundary artifacts, painterly ambiguity, diminishing returns from stronger local priors, disagreement between numerical and reconstructed-surface change, and the limited benefit of representation complexity when target geometry remains uncertain.
4. **A baseline-preserving formulation for local geometric refinement.** The complete experimental progression supports a strategy in which an image-derived global surface is retained and auxiliary information is introduced through localized, smoothly weighted, and bounded corrections rather than unrestricted geometric substitution.

These contributions define an experimentally grounded framework for studying controlled geometric refinement of shallow bas-reliefs. Their validation is geometric and computational; formal tactile effectiveness remains a subject for future physical evaluation.

10.5 Limitations

The findings should be interpreted within the scope of a single primary artwork and a predominantly digital experimental evaluation. The use of one challenging painting enabled controlled comparison across many experimental variants but limits generalization to other styles, compositions, subject scales, poses, and levels of abstraction.

The geometric estimates were obtained from pre-trained models not specifically designed for pictorial art or tactile-relief synthesis. Depth, normals, landmarks, and external geometry can therefore be affected by painterly shading, texture, occlusion, stylization, and domain shift. The true three-dimensional geometry of the depicted scene was not available as ground truth.

Most detailed refinement experiments focused on David’s face and Goliath’s head. Other tactilely relevant structures, including hands, clothing folds, hair, weapons, and broader object boundaries, were not investigated with the same level of detail. Several earlier parametric experiments also relied on manually or semi-manually controlled spatial supports, reducing the automation of those stages.

A further limitation concerns scale. Many algorithmically measurable corrections became small after conversion to the shallow physical displacement range. Exp115 R1 is the clearest example: its maximum native-stage displacement-equivalent residual was approximately 0.02230 mm, well below the approximately 0.1 mm engineering reference used in this thesis. Such results should therefore be interpreted as evidence of geometric localization rather than physical perceptibility.

Finally, no formal tactile user study with blind or visually impaired participants was performed, and fabrication parameters were not systematically optimized across the complete experimental sequence. The thesis therefore does not establish improved tactile recognition, tactile readability, or fabrication-level superiority.

10.6 Future Work

The experimental findings suggest several directions for further research.

10.6.1 Broader Evaluation and Generalization

The pipeline should be evaluated on a substantially larger and more diverse artwork dataset, including different artistic styles, compositions, subject scales, poses, levels of abstraction, and foreground-background organizations. Such an evaluation would help distinguish artwork-specific behaviour from methodological trends that generalize beyond the present case study.

10.6.2 Automated and Relational Anatomical Refinement

Future systems should reduce reliance on manually controlled supports by combining semantic segmentation, facial landmarks, uncertainty estimation, and image-to-template registration in a domain-aware localization pipeline for painted subjects.

The geometric objective should also move beyond isolated feature amplitude toward the relationships between neighbouring structures. Relevant examples include the nose bridge relative to the cheeks, nose tip relative to the upper lip, eye sockets relative to the forehead, and lips relative to the chin. Optimizing these relationships may be more useful than uniformly increasing the prominence of individual features.

10.6.3 Fabrication- and Tactile-Oriented Optimization

Most geometry-estimation models are optimized for visual or geometric accuracy rather than for tactile representation. A future tactile-oriented objective could explicitly combine anatomical separability, transition quality, preservation of the inherited global surface, geometric regularity, and fabrication constraints:

$$E_{\text{tactile}} = \lambda_{\text{sep}} E_{\text{separation}} + \lambda_{\text{tr}} E_{\text{transition}} + \lambda_{\text{pres}} E_{\text{preservation}} + \lambda_{\text{reg}} E_{\text{regularity}} + \lambda_{\text{fab}} E_{\text{fabrication}}. \quad (10.2)$$

Here, $E_{\text{separation}}$ encourages distinguishability between neighbouring geometric regions, $E_{\text{transition}}$ controls local surface transitions, $E_{\text{preservation}}$ retains the stable global surface, $E_{\text{regularity}}$ suppresses spikes and unstable curvature, and $E_{\text{fabrication}}$ represents physical manufacturing constraints.

Fabrication-aware optimization should explicitly account for minimum printable height and width, maximum local slope, wall thickness, printer resolution, material-dependent smoothing, and post-processing. Representative configurations should then be fabricated under identical physical conditions to determine which digital differences survive the transition from depth representation to physical object.

10.6.4 Improved Multi-Cue Geometry Fusion

Future work should investigate multi-scale allocation of the available physical relief range, separating global composition, regional organization, and local anatomical structure. A conceptual decomposition is

$$D = D_{\text{global}} + D_{\text{regional}} + D_{\text{local}}. \quad (10.3)$$

Depth, normals, curvature, semantics, and confidence could then contribute at different spatial scales rather than through a single local correction mechanism. Depth-normal fusion

should also incorporate uncertainty, integrability, multi-scale consistency, and region-dependent weighting so that independent orientation information is used only where it provides reliable complementary evidence.

10.6.5 Learning-Based Bounded Local Refinement

A longer-term extension could learn a bounded local correction function conditioned on the RGB image, baseline depth, semantic support, surface normals, confidence, and anatomical landmarks:

$$\Delta D = F_{\theta}(I, D, S, N, C, L). \quad (10.4)$$

The learned model should predict a constrained local correction rather than an unconstrained replacement depth field. Such an approach would require tactile- or fabrication-aware supervision that is not available in the present study.

10.6.6 Physical Tactile Validation

The most important validation step is a controlled tactile study using physically fabricated reliefs and blind or visually impaired participants. Representative baseline, semantic, parametric, and geometry-aware configurations should be fabricated using identical dimensions, material, and printer settings.

Evaluation should investigate recognition of major subjects, facial orientation, neighbouring facial structures, foreground–background organization, and undesirable surface discontinuities. Such a study is necessary to connect the computational geometric criteria of this thesis with actual tactile perception.

The same framework should also be extended beyond faces to hands, fingers, clothing folds, hair masses, held objects, object contours, architectural structures, and foreground–background contact regions, since these may require different spatial supports and fabrication-aware transition constraints.

10.7 Final Remarks

The generation of tactile-oriented bas-reliefs from paintings is not equivalent to recovering the most detailed possible three-dimensional scene. The target representation must preserve the composition and viewpoint of the artwork within a severely constrained physical depth range, while remaining geometrically continuous and compatible with fabrication.

The experiments in this thesis show that the most difficult problem is therefore not simply obtaining additional geometry, but deciding which geometric information is sufficiently reliable, spatially compatible, and physically meaningful to justify modifying an established image-derived surface. Greater numerical change or greater representational complexity did not consistently provide a greater final-relief effect.

Within the present case study, the most defensible strategy was to retain a stable image-aligned global depth representation and introduce auxiliary information through localized, bounded corrections. The final late-stage geometry-aware correction demonstrates the controllability of

this formulation, but its small physical magnitude also shows why future work must connect computational refinement directly to fabrication constraints and tactile perception.

The next step is therefore not simply a more complex geometry model. It is an evaluation framework in which geometric evidence, physical manufacturability, and human tactile perception are optimized and validated together.

Appendix A

Reproducibility and Experimental Parameters

This appendix reports the principal implementation parameters required to reproduce the baseline construction and the representative refinement experiments discussed in the thesis. It complements the methodological descriptions in the main chapters with concrete model configurations, the custom PPD-inspired preprocessing stage, depth-processing parameters, semantic supports, the Phase-2C intermediate facial baseline, anatomical landmark definitions, and surface-reconstruction settings.

The experiments share the same source artwork and a common DA3-derived depth lineage, but not all refinement families operate directly on the same intermediate depth file. Earlier configurations such as Exp46 and Exp80A were applied directly to the DA3+Remap representation, whereas later parametric configurations used progressively refined facial baselines and Exp115 R1 used the retained intermediate baseline obtained from the selected Phase-2B R5 semantic reinjection followed by the Phase-2C E2 localized mouth correction. Within each controlled experimental family, the input depth, masks, reconstruction configuration, and all non-tested parameters were kept fixed so that differences could be attributed primarily to the refinement variable under investigation.

A.1 Common Input and Depth Representation

The experimental pipeline was constructed around a fixed pictorial input and an image-aligned depth representation. The source image was Caravaggio’s *David with the Head of Goliath*, with an original RGB resolution of 3000×2333 pixels.

Depth Anything 3 (DA3) was used as the principal monocular depth-estimation framework. The selected large-model inference configuration used the **DA3-LARGE-1.1** checkpoint with a processing resolution of 1536 and the **upper_bound_resize** preprocessing strategy.

During the resolution analysis, processing resolutions of 504, 1024, 1536, and 2048 were compared. The 1024-resolution result served as the reference representation for the resolution-scaling analysis. The 1536 configuration retained high agreement with this reference, with a normalized cross-correlation of approximately 0.984, while maintaining slightly higher mean gradient magnitude.

The relief-remapping stage did not operate directly on the raw inference array. Its input was

the PPD-inspired multistage-refined 16-bit representation associated with the 1536-resolution DA3 pipeline and aligned to the 1024 reference grid. The resulting working depth dimensions were

$$1022 \times 798 \text{ pixels.} \tag{A.1}$$

Depth-processing operations were performed internally using floating-point arrays, while depth maps transferred between major stages and into Blender were stored as unsigned 16-bit TIFF images.

Table A.1: Common input and depth-representation parameters.

Parameter	Value / Configuration
Source artwork	Caravaggio, <i>David with the Head of Goliath</i>
Original RGB resolution	3000×2333 pixels
Principal monocular depth framework	Depth Anything 3
Selected DA3 model	DA3-LARGE-1.1
Selected processing resolution	1536
DA3 processing-resize strategy	<code>upper_bound_resize</code>
Resolution configurations evaluated	504, 1024, 1536, and 2048
Reference resolution for scaling analysis	1024
Working aligned depth dimensions	1022×798 pixels
Input to relief-oriented remapping	PPD-inspired multistage-refined 16-bit depth
Internal processing representation	32-bit floating-point arrays
Depth exchange format	Unsigned 16-bit TIFF
Quantization range	0–65535

A.1.1 PPD-Inspired Multistage Depth Refinement

Before relief-oriented remapping, the selected DA3 depth estimate was processed using a custom PPD-inspired multistage refinement procedure implemented in

`run_ppd_multistage.py`.

The term *PPD-inspired* is used here to describe this custom depth-domain post-processing procedure; it should not be interpreted as an implementation of the complete Pixel-Perfect Depth method. The purpose of the stage was to stabilize the numerical range, remove small isolated support components, preserve strong depth boundaries during smoothing, and regularize locally flat regions before the relief-specific remapping stage.

The procedure was applied to DA3 depth maps rescaled to the common 1024-reference grid for the 504, 1536, and 2048 processing-resolution comparisons. The retained input for the subsequent pipeline was the refined output associated with `cmp_1536_to_1024`, corresponding to the selected DA3-LARGE-1.1 inference at processing resolution 1536.

The refinement consisted of six stages:

1. **Range stabilization.** The input depth was clipped to the 0.5th and 99.5th percentiles and min-max normalized to $[0, 1]$.
2. **Support cleanup.** A binary support was defined by $D > 0.01$. Eight-connected components with area smaller than 64 pixels were removed, and depth values outside the retained support were set to zero.
3. **Boundary detection.** Horizontal and vertical derivatives were computed using 3×3 Sobel operators. The normalized gradient magnitude was thresholded at its 85th percentile. The resulting boundary mask was then dilated once and morphologically closed using a 3×3 elliptical structuring element.
4. **Non-boundary smoothing.** A bilateral filter with $d = 7$, $\sigma_{\text{color}} = 0.06$, and $\sigma_{\text{space}} = 7$ was applied, followed by a 3×3 median filter.
5. **Boundary-aware blending.** Outside the detected boundary support, the smoothed representation was retained. Inside the boundary support, the refined value was computed as

$$D_{\text{blend}} = 0.80D_{\text{orig}} + 0.20D_{\text{smooth}}, \quad (\text{A.2})$$

thereby preserving a larger contribution from the original depth near strong transitions.

6. **Local surface regularization.** Pixels with normalized gradient magnitude below 0.20 were treated as relatively flat regions. In those regions, a Gaussian-smoothed representation with $\sigma = 0.8$ was blended with weight 0.25:

$$D_{\text{reg}} = D_{\text{blend}}(1 - 0.25M_f) + \mathcal{G}_{0.8}(D_{\text{blend}})(0.25M_f), \quad (\text{A.3})$$

where $M_f = 1$ for gradient magnitude below 0.20 and zero elsewhere. The result was then again stabilized using the 0.5th and 99.5th percentiles.

The final normalized output was quantized to unsigned 16-bit depth and stored as

`ppd_multistage_refined_16bit.tiff`.

Preview images for the normalized input, boundary mask, gradient map, smoothed intermediate representation, and final refined map were also stored, together with `ppd_multistage_stats.json`, which records summary statistics for the principal processing stages.

A.2 Depth Normalization and Relief-Oriented Remapping

The PPD-inspired representation was converted into the relief-oriented working baseline through robust percentile normalization, adaptive mid-range redistribution, mild detail re-injection, and a final gamma transformation. The implementation was performed by the script

Table A.2: Parameters of the custom PPD-inspired multistage refinement.

Parameter	Value / Configuration
Range-stabilization percentiles	0.5, 99.5
Support threshold	$D > 0.01$
Minimum connected-component area	64 pixels
Gradient operator	Sobel, 3×3
Boundary threshold	85th percentile of normalized gradient magnitude
Boundary morphology	one dilation + closing, 3×3 elliptical kernel
Bilateral-filter diameter	7
Bilateral σ_{color}	0.06
Bilateral σ_{space}	7
Median-filter kernel	3×3
Boundary blend	0.80 original + 0.20 smoothed
Flat-region gradient threshold	0.20
Flat-region Gaussian σ	0.8
Flat-region regularization weight	0.25
Final exchange format	Unsigned 16-bit TIFF

`marco_curve_remap_v2_range.py`.

Let D denote the input depth map. The lower and upper robust normalization values were obtained from the 0.5th and 99.5th percentiles:

$$l = P_{0.5}(D), \quad h = P_{99.5}(D). \quad (\text{A.4})$$

The normalized depth was computed as

$$z = \text{clip} \left(\frac{D - l}{h - l + \varepsilon}, 0, 1 \right), \quad (\text{A.5})$$

where ε is a small constant used to prevent numerical division by zero.

A histogram of z was then computed using 256 bins. The histogram was smoothed using a one-dimensional averaging kernel of length 21. To reduce the influence of dominant background values, the search for the dominant intermediate-depth mode was restricted to

$$0.18 \leq z \leq 0.82. \quad (\text{A.6})$$

The location of the largest smoothed histogram value within this range defines the mid-band centre c .

A Gaussian band-support function was defined as

$$b(z) = \exp \left(-\frac{(z - c)^2}{2\sigma_b^2 + \varepsilon} \right), \quad (\text{A.7})$$

where

$$\sigma_b = 0.12. \quad (\text{A.8})$$

The local sigmoid gain was computed as

$$g(z) = g_0 + \alpha g_b b(z), \quad (\text{A.9})$$

with

$$g_0 = 5.5, \quad \alpha = 0.42, \quad g_b = 10.5. \quad (\text{A.10})$$

The corresponding adaptive sigmoid centre was

$$c_m(z) = c + \beta b(z), \quad (\text{A.11})$$

with

$$\beta = -0.015. \quad (\text{A.12})$$

The nonlinear remapping was therefore

$$y = \frac{1}{1 + \exp[-g(z)(z - c_m(z))]} \quad (\text{A.13})$$

After the sigmoid transformation, y was min-max normalized again to $[0, 1]$.

A mild detail component was subsequently extracted. A Gaussian-smoothed representation was computed using

$$y_s = \mathcal{G}_{1.0}(y), \quad (\text{A.14})$$

and the residual detail was

$$d = y - y_s. \quad (\text{A.15})$$

When the standard deviation of d exceeded the numerical tolerance, the detail signal was standardized by its standard deviation. Its magnitude was then limited to

$$[-0.05, 0.05]. \quad (\text{A.16})$$

Thus,

$$\Delta = \text{clip}\left(\frac{d}{\text{std}(d)}, -0.05, 0.05\right), \quad (\text{A.17})$$

for the non-degenerate case.

The clipped detail component was reintroduced with weight 0.05:

$$y_b = \text{clip}(y + 0.05\Delta, 0, 1). \quad (\text{A.18})$$

Finally, a gamma transformation with exponent 0.97 was applied:

$$D_r = y_b^{0.97}. \quad (\text{A.19})$$

The result was clipped to $[0, 1]$ and quantized to unsigned 16-bit depth:

$$D_{16} = \text{uint16}(65535 D_r). \quad (\text{A.20})$$

The output generated by this stage was

`depth_marco_curve_v2_range_16bit.tiff`.

Table A.3: Parameters of the relief-oriented depth-remapping stage.

Parameter	Final value
Lower percentile p_l	0.5
Upper percentile p_h	99.5
Histogram bins	256
Histogram smoothing kernel	21
Mid-band search interval	$[0.18, 0.82]$
Band width σ_b	0.12
Band strength α	0.42
Base sigmoid gain g_0	5.5
Band gain g_b	10.5
Centre bias β	-0.015
Detail Gaussian sigma	1.0
Detail re-injection weight	0.05
Maximum normalized detail magnitude	0.05
Final gamma exponent	0.97
Normalized output range	$[0, 1]$
Final output format	Unsigned 16-bit TIFF

A.3 Blender Reconstruction Configuration

Representative depth candidates were reconstructed using a controlled Blender configuration so that differences in the resulting surfaces reflected changes in the depth representations rather than changes in the reconstruction procedure.

The Blender scene used the Metric unit system, millimetres as the working length unit, and a scene unit scale of 0.001. Object scale was applied before displacement.

The geometric support consisted of a square plane with physical dimensions

$$200 \times 200 \text{ mm}. \quad (\text{A.21})$$

No inset boundary was introduced in the baseline reconstruction. The displacement texture therefore affected the complete available surface.

The modifier stack was applied in the following order:

1. Subdivision Surface;
2. Displace;
3. Solidify.

The Subdivision Surface modifier used Catmull–Clark subdivision with viewport level 4 and render level 6. Optimal Display was enabled.

The corresponding unsigned 16-bit TIFF depth map was used as the displacement texture. Texture coordinates were supplied through UVMap. The texture was interpreted as Non-Color data, linear interpolation was used, and the extension mode was set to **Extend**.

Displacement was applied along the surface-normal direction with

$$\text{Midlevel} = 0.500 \tag{A.22}$$

and

$$\text{Strength} = -0.015. \tag{A.23}$$

The negative strength compensates for the orientation convention of the DA3-derived depth representation. The magnitude was empirically calibrated to obtain an effective relief height of approximately 7 mm.

No smoothing modifier was applied in the quantitative baseline in order to avoid introducing an additional filtering stage during geometric comparison.

Finally, a structural backing was added using the Solidify modifier in Simple mode. The backing thickness was 2 mm, with an offset of -1 and Fill Rim enabled. The resulting model therefore had an approximate total thickness of 9 mm.

No manual sculpting or non-procedural mesh editing was applied when establishing the geometric baseline.

A.3.1 Physical Interpretation of Normalized Depth Residuals

The local-refinement experiments report geometric modifications primarily in normalized depth units. To provide a physically interpretable representation of these residuals, the following conversion is based on the displacement mapping adopted in the later physical-perceptibility validation, in which the displacement-strength magnitude was 0.015 and the validated scale was 1 Blender Unit = 1 metre.

This physical interpretation is therefore reported separately from the scene-display unit settings listed in Table A.4. It represents a displacement-equivalent conversion under the validated local-correction mapping and should not be interpreted as a direct consequence of the Blender display-unit configuration or as a direct measurement of the fabricated relief.

For two normalized depth maps D and D' , the displacement midlevel is identical and therefore cancels when the difference between the two reconstructed surfaces is considered. Using the adopted displacement magnitude of 0.015 m per normalized depth unit, the absolute displacement difference is

$$\Delta z_m = 0.015 |\Delta D| \quad [\text{m}], \tag{A.24}$$

Table A.4: Fixed Blender reconstruction parameters.

Parameter	Value / Configuration
Unit system	Metric
Length unit	Millimetres
Scene unit scale	0.001
Base geometry	Plane
Plane dimensions	200×200 mm
Object rotation	0°
Applied object scale	(1, 1, 1)
Inset margin	None
Subdivision method	Catmull–Clark
Subdivision viewport level	4
Subdivision render level	6
Optimal Display	Enabled
Displacement texture	Unsigned 16-bit TIFF
Texture coordinates	UV
UV map	UVMap
Texture color space	Non-Color
Texture interpolation	Linear
Texture extension	Extend
Displacement direction	Normal
Displace midlevel	0.500
Displace strength	−0.015
Additional smoothing modifier	Not applied
Solidify mode	Simple
Solidify thickness	2 mm
Solidify offset	−1
Fill Rim	Enabled
Approximate relief height	7 mm
Approximate total thickness	9 mm

where

$$\Delta D = D' - D. \quad (\text{A.25})$$

Equivalently, the normalized residual can be expressed in millimetres as

$$\Delta z_{\text{mm}} = 15 |\Delta D|, \quad (\text{A.26})$$

or in micrometres as

$$\Delta z_{\mu\text{m}} = 15000 |\Delta D|. \quad (\text{A.27})$$

For an unsigned 16-bit representation, one integer depth level therefore corresponds to approximately

$$\frac{15}{65535} \approx 0.0002289 \text{ mm} = 0.2289 \mu\text{m}. \quad (\text{A.28})$$

This conversion describes the displacement-equivalent magnitude of a depth-map difference under the validated local-correction displacement mapping. It should not be interpreted as the total thickness or total relief height of the reconstructed object, which additionally depends on the occupied range of the particular depth map.

Table A.5 reports the physical interpretation of the residuals obtained for representative experiments from the parametric and anatomy-guided refinement phase.

Table A.5: Physical interpretation of selected normalized depth residuals using the validated 15 mm-per-normalized-unit local-correction displacement mapping.

Experiment	Global mean	David mean	Goliath mean
Exp101B	17.60 μm	161.32 μm	105.15 μm
Exp101C	18.45 μm	167.54 μm	110.62 μm
Exp102	22.71 μm	202.60 μm	137.07 μm
Exp103B	5.63 μm	51.85 μm	33.55 μm
Exp104A	5.54 μm	52.08 μm	32.81 μm
Exp105A	7.03 μm	67.52 μm	41.29 μm
Exp106A	9.22 μm	90.34 μm	53.65 μm

The regional values are substantially larger than the corresponding image-wide means because the modifications are spatially localized. For example, Exp101B produced an image-wide mean displacement difference of approximately 17.60 μm , while the mean displacement difference inside David’s facial support was approximately 161.32 μm and that inside Goliath’s support was approximately 105.15 μm .

Exp102 produced the largest regional values among the three direct parametric configurations reported above, reaching approximately 202.60 μm for David and 137.07 μm for Goliath. Nevertheless, the Blender reconstruction remained visually close to the preceding configurations. This supports the observation that increasing the numerical magnitude of a localized depth modification does not necessarily produce a proportional increase in visible surface change.

For Exp101B, Exp101C, and Exp102, recomputation from the final unsigned 16-bit TIFF

files showed a mean residual of zero outside the union of the David and Goliath facial supports. These configurations therefore preserved the surrounding depth field at the exported depth-map resolution.

Exp115 R1 requires a separate interpretation because the statistics originally generated by its implementation were evaluated only over pixels receiving a non-zero local correction.

The experiment contained

$$101600 \tag{A.29}$$

active pixels, including

$$20187 \tag{A.30}$$

active pixels in the David region and

$$81413 \tag{A.31}$$

active pixels in the Goliath region.

The mean absolute residual evaluated over these active pixels was

$$0.0000725428, \tag{A.32}$$

corresponding to approximately

$$1.09 \mu\text{m}. \tag{A.33}$$

The corresponding active-pixel percentile statistics are reported in Table A.6.

Table A.6: Normalized and physically interpreted active-pixel displacement statistics for Exp115 R1. Percentile and mean values are evaluated over pixels receiving a non-zero local correction.

Statistic	Normalized depth	Predicted displacement
Mean absolute residual	0.0000725428	1.09 μm
90th percentile	0.0001630153	2.45 μm
95th percentile	0.0002604302	3.91 μm
99th percentile	0.0008175200	12.26 μm
Maximum absolute residual	0.0014867431	22.30 μm

When the final Exp115 R1 TIFF is instead compared with its native baseline over the complete 1022×798 image, including unchanged pixels, the image-wide mean absolute residual is approximately

$$0.00000912, \tag{A.34}$$

which corresponds to approximately

$$0.137 \mu\text{m}. \tag{A.35}$$

The difference between the active-pixel mean of $1.09\,\mu\text{m}$ and the whole-image mean of $0.137\,\mu\text{m}$ therefore results from the evaluation support rather than from any discrepancy in the generated depth map.

The active-pixel statistic describes the magnitude of the correction where Exp115 actually modified the surface, whereas the whole-image statistic additionally includes the large unchanged portion of the depth map. Pixels outside the two facial supports were preserved exactly.

These values show that the later Exp115 R1 configuration was substantially more conservative than the earlier parametric experiments. In physical units, the active-pixel mean residual is approximately 0.00109 mm , the 99th percentile is approximately 0.01226 mm , and the maximum observed residual is approximately 0.02230 mm . All of these values are well below the approximately 0.1 mm fabrication-scale reference used later in this appendix.

Exp115 R1 should therefore be interpreted primarily as evidence that confidence-, curvature-, and landmark-guided corrections can be introduced in a tightly localized and bounded manner while preserving the surrounding depth field. Its native-stage residuals do not constitute evidence of a fabrication-relevant or tactile improvement. The reported physical quantities are displacement-equivalent geometric differences under the adopted Blender mapping rather than direct measurements of tactile perceptibility.

A.3.2 Relation to Fabrication Resolution and Tactile Perception

The displacement-equivalent residuals can be interpreted further by comparing their magnitude with characteristic scales reported for additive manufacturing and tactile perception. This comparison is intended only as an order-of-magnitude interpretation: neither the printing layer height nor a psychophysical threshold constitutes a strict minimum feature size for the present bas-relief geometry.

In the controlled Blender reconstruction used in this thesis, the occupied relief height was approximately 7 mm . The residual values discussed below, however, represent local displacement-equivalent differences under the separately validated $15\text{ mm-per-normalized-unit}$ mapping described in Section A.3.1. They should therefore be interpreted as local geometric modifications rather than as the total height of the relief.

For fused-deposition modelling (FDM), layer thicknesses in the range of approximately $0.1\text{--}0.2\text{ mm}$ have been reported as favourable for the dimensional accuracy of PLA and ABS parts [34]. This range provides a useful vertical fabrication-scale reference for interpreting the geometric modifications reported in Table A.5. Layer thickness should not, however, be interpreted as a strict minimum printable feature height, since fabrication accuracy also depends on geometry, material, process parameters, and print orientation.

Under the adopted displacement-equivalent mapping, several of the earlier parametric configurations produced facial-region mean residuals within or close to this scale. For example, Exp101B corresponds to approximately 0.161 mm in David’s facial support and 0.105 mm in Goliath’s support, while Exp102 reaches approximately 0.203 mm and 0.137 mm , respectively. In contrast, Exp105A produces approximately 0.068 mm for David and 0.041 mm for Goliath. These values illustrate that nominally small normalized depth differences can correspond to substantially different physical scales after displacement conversion.

A complementary reference is provided by psychophysical work on tactile elevation. Jehoel et al. [15] reported that tactile-map symbols with elevations of approximately $0.040\text{--}0.080\text{ mm}$ could

be identified by sighted and visually impaired participants. These values are relevant as an order-of-magnitude comparison, but they should not be interpreted as a direct detection threshold for the present bas-relief. The stimuli, exploration task, spatial extent, surface material, and geometric context differ substantially from the facial structures evaluated in this thesis.

More recent work on the haptic perception of 2.5D surface features further demonstrates that perceptual height thresholds depend on feature geometry. Hwang et al. [14] found that both difference and absolute height thresholds were affected by the lateral dimensions of dome-shaped surface features, with the difference threshold behaviour also related to the ratio between feature height and base diameter. Consequently, there is no single universal height threshold below which a geometric modification can be classified as imperceptible.

These comparisons provide an engineering context for the experimental results rather than a direct prediction of tactile performance. Modifications substantially below 0.1 mm are smaller than a typical 0.1–0.2 mm FDM layer increment and should therefore not be assumed, solely from their numerical magnitude, to correspond to a robustly distinguishable fabricated or tactile feature. Conversely, spatially coherent local modifications approaching approximately 0.1 mm or more are comparable in magnitude to a commonly used vertical fabrication scale and therefore provide stronger candidates for subsequent physical and tactile validation.

Applied specifically to Exp115 R1, this comparison is restrictive: even the maximum native-stage residual of approximately 0.02230 mm remains well below the 0.1 mm engineering reference, while the active-pixel mean is only approximately 0.00109 mm. Therefore, the value of Exp115 R1 lies in controlled localization and preservation of the inherited baseline rather than in the demonstrated magnitude of its physical surface change.

Accordingly, a practical design rule emerging from the present experiments is to distinguish between numerical depth modification and fabrication-relevant geometric modification. For future refinement stages, local changes should be evaluated in physical units and considered together with the selected fabrication parameters and the spatial extent of the modified anatomical feature. Changes whose magnitude approaches the vertical fabrication scale over a spatially coherent region may be prioritized for physical evaluation, whereas substantially smaller residuals should not be interpreted as evidence of tactile improvement without further validation. This criterion is an engineering screening rule rather than a psychophysical tactile threshold; final tactile effectiveness must still be established through fabrication and user evaluation.

A.3.3 Cross-Stage Quantitative Comparison

A unified quantitative comparison was performed across representative stages of the refinement pipeline in order to distinguish large-scale geometric modification from localized facial refinement. The selected configurations were DA3+Remap, Exp46, Exp80A, Exp101B, Exp105A, and Exp115 R1.

Two complementary reference definitions were used. First, all final candidates were compared against DA3+Remap using a common reference. This comparison measures the accumulated geometric difference of each final surface relative to the original DA3-derived relief. Second, each experiment was compared against its own native input baseline. This isolates the geometric modification introduced by the experiment itself.

This distinction is particularly important for Exp101B, Exp105A, and Exp115 R1, because these experiments operate on stage-specific refined input baselines rather than all operating

directly on the original DA3+Remap depth map.

Table A.7: Mean absolute displacement-equivalent residuals relative to each method’s native input baseline. Values are evaluated over the complete image or the indicated facial support.

Method	Global (μm)	David (μm)	Goliath (μm)
Exp46	157.58	1200.53	727.78
Exp80A	392.27	183.85	290.80
Exp101B	17.60	161.32	105.15
Exp105A	7.03	67.52	41.29
Exp115 R1	0.14	1.19	0.83

These native-baseline values quantify the modification introduced by each selected experiment itself. They should therefore be distinguished from the common-reference metrics, which measure the accumulated difference of the final candidate relative to DA3+Remap. This distinction is especially important for Exp101B, Exp105A, and Exp115 R1, because these experiments operate on stage-specific refined input baselines rather than all operating directly on the original DA3+Remap depth map.

Common-Reference Local Surface Metrics

To complement the whole-face comparison, a feature-level quantitative analysis was performed for three anatomically relevant facial regions: the nose, lips, and eye regions. The evaluation regions were derived from the same landmark-based anatomical support definitions used by the later geometry-aware refinement.

For each subject, the MediaPipe landmarks were first mapped to the 1022×798 working resolution. The anatomical regions were then defined from the Gaussian support functions associated with the Exp115 formulation. A binary evaluation region was obtained from the one-standard-deviation contour,

$$S(x, y) \geq \exp(-1/2), \quad (\text{A.36})$$

and intersected with the corresponding semantic facial support. The nose region was formed from the union of the nose-bridge and nose-tip supports, the lip region from the upper- and lower-lip supports, and the eye region from the two eye-socket supports. These evaluation regions should therefore be interpreted as reproducible support-aligned regions of interest rather than as mutually exclusive anatomical segmentations.

Using the physical displacement-equivalent mapping introduced in Section A.3.1, normalized depth was converted to

$$Z = 15D, \quad (\text{A.37})$$

where Z is expressed in millimetres.

Local relief contrast within region R was evaluated as

$$C_R = P_{95}(Z_R) - P_5(Z_R), \quad (\text{A.38})$$

while local surface variation was measured using the mean gradient magnitude

$$G_R = \text{mean}_R \left(\sqrt{\left(\frac{\partial Z}{\partial x}\right)^2 + \left(\frac{\partial Z}{\partial y}\right)^2} \right), \quad (\text{A.39})$$

reported in millimetres per pixel.

Tables A.8 and A.9 report the resulting feature-level values. Each table entry is written as C_R/G_R , where the first value is relief contrast in millimetres and the second is the mean gradient magnitude in millimetres per pixel.

Table A.8: Feature-level geometric metrics for David. Each entry reports local relief contrast C_R in mm followed by mean gradient magnitude G_R in mm/px.

Method	Nose C_R/G_R	Lips C_R/G_R	Eyes C_R/G_R
DA3+Remap	0.992 / 0.058	1.095 / 0.049	2.038 / 0.043
Exp46	0.748 / 0.043	0.695 / 0.032	1.255 / 0.033
Exp80A	0.945 / 0.055	1.045 / 0.046	1.944 / 0.040
Exp101B	1.712 / 0.088	1.125 / 0.057	2.440 / 0.063
Exp105A	1.148 / 0.075	0.977 / 0.053	2.320 / 0.058
Exp115 R1	1.155 / 0.072	0.939 / 0.052	2.133 / 0.055

For David, Exp46 reduced the relief contrast of all three evaluated anatomical regions relative to DA3+Remap, with decreases of approximately 24.6% for the nose, 36.5% for the lips, and 38.4% for the eye region. Exp80A produced smaller reductions of approximately 4–5%.

The parametric Exp101B configuration produced the largest increase in David’s local anatomical variation. Nose contrast increased by approximately 72.7%, while the corresponding mean-gradient magnitude increased by approximately 51.6%. The eye-region contrast increased by approximately 19.7% and its mean gradient by approximately 46.3%.

The later configurations were more conservative. Exp105A increased David’s nose contrast by approximately 15.8% and eye-region contrast by approximately 13.8%, while Exp115 R1 produced corresponding changes of approximately 16.5% and 4.6%. The lip contrast did not increase consistently: it decreased by approximately 10.8% for Exp105A and 14.2% for Exp115 R1. Thus, the feature-level measurements do not indicate uniform enhancement of all facial structures, but rather show that the strongest retained effects are concentrated primarily in the nose and local surface-gradient organization.

A similar trend was observed for Goliath. Exp46 reduced the feature-level contrast by approximately 23–26% across the three evaluated regions, whereas Exp80A produced reductions of approximately 4–6%.

Exp101B again generated the strongest local variation, increasing nose contrast by approximately 49.1% and eye-region contrast by approximately 22.6%. Exp105A retained a smaller increase of approximately 12.7% in nose contrast and 7.1% in eye-region contrast.

Exp115 R1 produced substantially smaller contrast changes for Goliath: approximately +1.8% for the nose, +1.5% for the lips, and –1.2% for the eye region. Nevertheless, its mean-gradient magnitude increased by approximately 6.5%, 7.3%, and 26.2% in the nose, lip, and eye regions, respectively. This indicates that the final configuration primarily reorganized localized

Table A.9: Feature-level geometric metrics for Goliath. Each entry reports local relief contrast C_R in mm followed by mean gradient magnitude G_R in mm/px.

Method	Nose C_R/G_R	Lips C_R/G_R	Eyes C_R/G_R
DA3+Remap	1.225 / 0.049	0.879 / 0.026	1.042 / 0.019
Exp46	0.919 / 0.036	0.652 / 0.019	0.804 / 0.014
Exp80A	1.151 / 0.046	0.840 / 0.024	0.995 / 0.018
Exp101B	1.827 / 0.056	0.886 / 0.029	1.278 / 0.028
Exp105A	1.381 / 0.051	0.834 / 0.028	1.116 / 0.026
Exp115 R1	1.247 / 0.052	0.892 / 0.028	1.030 / 0.024

surface transitions rather than strongly increasing the overall relief range.

Global Preservation and Boundary Disturbance

Global preservation was quantified using the mean absolute displacement-equivalent difference outside the union of the final David and Goliath facial supports,

$$E_{\text{outside}} = \text{mean}_{(x,y) \notin M_D \cup M_G} |Z'(x,y) - Z_B(x,y)|, \quad (\text{A.40})$$

where Z_B denotes the DA3+Remap reference surface.

Boundary disturbance was evaluated in a five-pixel band surrounding each semantic support. The absolute difference between the candidate and baseline gradient magnitudes was computed inside this band, and its 95th percentile was retained as a conservative boundary-artifact indicator.

Table A.10: Global-preservation and boundary-disturbance metrics relative to DA3+Remap. Outside-support change is the mean absolute displacement-equivalent difference outside the union of the two final facial supports. Boundary values are the 95th percentile of absolute gradient-magnitude change within a five-pixel band around each support.

Method	Outside change (μm)	David boundary P_{95} ($\mu\text{m/px}$)	Goliath boundary P_{95} ($\mu\text{m/px}$)
DA3+Remap	0.000	0.000	0.000
Exp46	39.224	106.821	53.257
Exp80A	414.129	79.787	53.011
Exp101B	0.207	16.798	16.544
Exp105A	0.207	17.021	16.814
Exp115 R1	0.207	16.255	16.747

The global-preservation metric clearly separates the early semantic and external-geometry experiments from the later localized facial refinements. Exp80A produced the largest outside-support change, approximately 414 μm , which is consistent with its intended body- and drapery-refinement objective. Exp46 also modified regions outside the final facial-support definition, with a mean common-reference difference of approximately 39 μm .

By contrast, Exp101B, Exp105A, and Exp115 R1 all produced a common-reference outside-support residual of only approximately $0.207\ \mu\text{m}$. This value should not be interpreted as evidence of substantial modification leakage from those individual stages. These later candidates do not all operate directly on DA3+Remap. Therefore, the common-reference measurement also includes differences inherited from their respective input baselines, together with small export and quantization effects, rather than representing only the modification introduced by the individual experiment. Their native-stage formulations constrain the intended local facial modifications to the corresponding supports.

The boundary metric shows the same general transition. Exp46 produced a David boundary disturbance of approximately $106.8\ \mu\text{m}/\text{px}$, whereas Exp101B, Exp105A, and Exp115 R1 remained close to $16\text{--}17\ \mu\text{m}/\text{px}$. Goliath showed the same pattern, with the later configurations remaining near $16\text{--}17\ \mu\text{m}/\text{px}$ compared with approximately $53\ \mu\text{m}/\text{px}$ for Exp46 and Exp80A.

These quantities are geometric proxy measures rather than direct tactile-performance scores. Nevertheless, they indicate that the later refinement stages are associated with more spatially localized modifications and lower measured boundary disturbance, while retaining measurable feature-level geometric changes.

Difference Maps Relative to DA3+Remap

Figure A.1 presents difference maps computed relative to the same DA3+Remap reference. A common symmetric colour scale was used for all candidates so that the magnitude and spatial extent of the modifications could be compared directly.

The common-scale maps demonstrate that Exp46 introduced strong large-scale modifications within both head regions. Exp80A affected a substantially broader foreground region because its objective was body and drapery enhancement rather than isolated facial refinement. Exp101B, Exp105A, and Exp115 R1 instead remained concentrated primarily around the two facial supports.

The common-scale representation is intentionally conservative: its scale is dominated by the larger Exp46 and Exp80A differences. Consequently, the much smaller local changes of the later configurations are less visually prominent, even though they remain measurable.

Native-Baseline Anatomical Height Profiles

A common-reference comparison alone cannot determine how much modification was introduced by an individual experiment when that experiment inherited a previously refined baseline. For this reason, anatomical height-difference profiles were also evaluated relative to each method's native input depth map.

Two standardized paths were evaluated for each subject. The vertical path traversed the eye region, nose, mouth, and chin, while the horizontal path traversed the left cheek, nose, and right cheek.

The native-baseline profiles reveal a large difference in modification scale between the experimental families. Exp46 introduced broad millimetre-scale variation along the facial profiles. Exp101B produced a strongly localized positive modification around the nose and central face, with peak profile differences approaching the millimetre scale despite a much smaller image-wide mean residual.

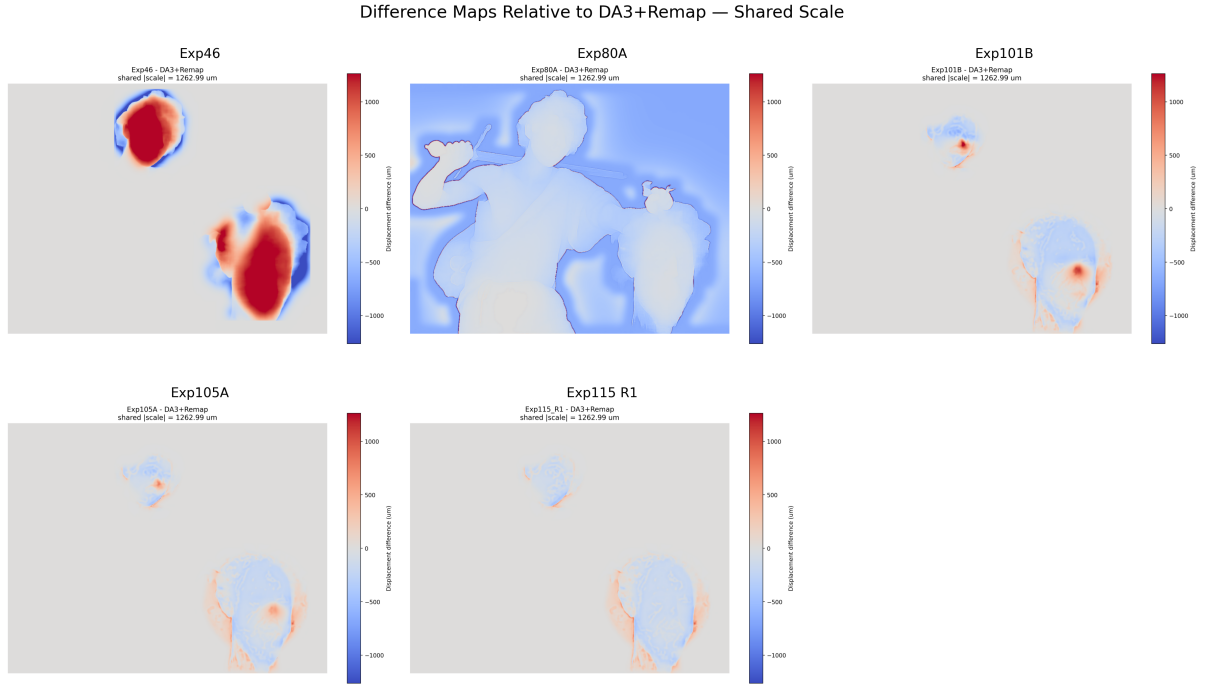
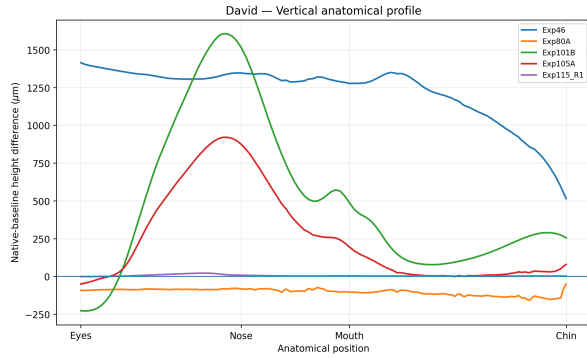
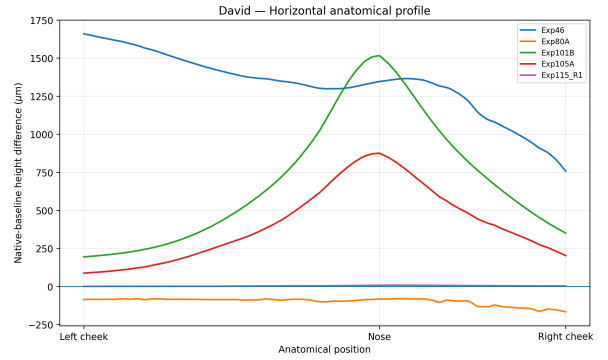


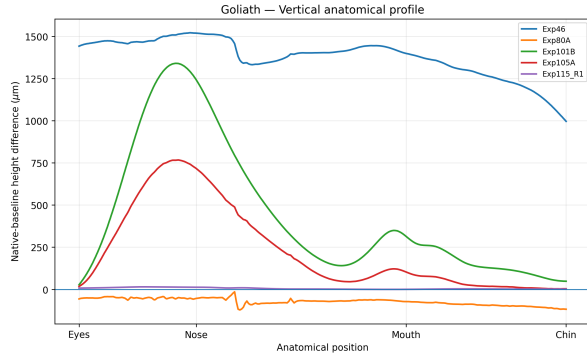
Figure A.1: Depth-difference maps for representative refinement stages relative to DA3+Remap using a common symmetric displacement scale. Red and blue indicate positive and negative displacement-equivalent differences, respectively. Exp46 produces large localized volumetric changes, while Exp80A modifies a broader foreground region. The later facial-refinement configurations produce substantially more localized changes. Because all panels share the same scale, small differences in Exp101B, Exp105A, and Exp115 R1 appear comparatively faint.



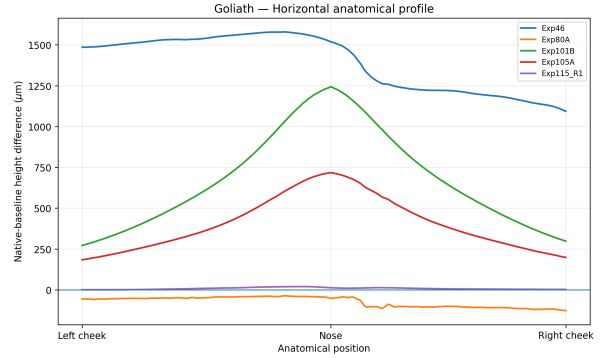
(a) David: eyes–nose–mouth–chin



(b) David: left cheek–nose–right cheek



(c) Goliath: eyes–nose–mouth–chin



(d) Goliath: left cheek–nose–right cheek

Figure A.2: Native-baseline anatomical height-difference profiles for the representative refinement stages. Each curve measures only the modification introduced relative to the actual input baseline of that experiment. The vertical profiles traverse the eyes, nose, mouth, and chin, while the horizontal profiles traverse the cheeks and nose.

Exp105A retained a similar spatial pattern but with reduced amplitude, consistent with its more constrained signed anatomical template and surface-relaxation formulation.

By contrast, the Exp115 R1 profiles remain close to zero on the same axes. This is consistent with the active-pixel and whole-image statistics reported previously: Exp115 R1 was designed as a conservative, bounded refinement rather than a large-scale reshaping operation.

Exp80A exhibits a comparatively small and predominantly negative displacement along the facial paths. This behaviour should not be interpreted as failure of the method, since Exp80A was designed primarily to modify body and drapery structure while suppressing direct enhancement within the facial regions.

Taken together, the residual statistics, difference maps, boundary measurements, and anatomical profiles show a progression from broad and comparatively strong geometric manipulation toward increasingly localized and conservative facial refinement. The later methods modify less of the inherited depth field and reduce measured boundary disturbance. For Exp115 R1 in particular, the native-stage correction is extremely small in physical terms; its role is therefore best understood as demonstrating controlled, support-bounded refinement and baseline preservation rather than stronger fabrication-scale relief. Common-reference feature metrics for the final Exp115 candidate include geometric differences inherited from the preceding Phase-2C baseline and should not be attributed solely to the Exp115 R1 step.

A.4 Semantic Mask Generation

The semantic supports evolved during the experimental sequence. Consequently, the representative experiments did not all use one identical set of masks. Three principal support configurations were employed: the initial GroundingDINO–SAM facial masks, the foreground/body supports used in Exp80A, and the later manually controlled v6 facial supports used by the parametric and geometry-aware experiments.

A.4.1 GroundingDINO–SAM Face and Head Supports

The masks used by the earlier external-geometry experiments were initially generated through a GroundingDINO–SAM pipeline.

GroundingDINO used the Swin-T OGC configuration and the text prompt

```
face . head .
```

on the original 3000×2333 image. The box-confidence and text-confidence thresholds were both

$$0.25. \tag{A.41}$$

Each retained GroundingDINO bounding box was supplied to the Segment Anything Model using the ViT-B checkpoint

```
sam_vit_b_01ec64.pth.
```

SAM produced multiple candidate masks, from which the highest-scoring mask was retained.

The initial masks were cleaned by applying morphological closing with a 25×25 kernel, opening with a 7×7 kernel, and binary hole filling.

For Goliath, the support was subsequently expanded to recover missing hair and ear regions. An elliptical 35×35 dilation was followed by an elliptical 25×25 closing operation and hole filling. The resulting mask was stored as `goliath_final_mask.v2.png`.

These supports were used directly by Exp46.

A.4.2 Foreground and Face Exclusion Supports in Exp80A

Exp80A used a foreground mask together with separate David and Goliath facial exclusion supports.

The foreground mask defined the body/drapery region from which a boundary-safe interior support was derived. The David support was obtained from

`exp73_precheck/david_mask_resized.png`,

whereas the Goliath support was

`exp73_precheck/goliath_face_mask_manual.v2_filled.png`.

The foreground support was eroded before modification, while the union of the facial masks was dilated and removed from the direct body-enhancement region. This prevented the frequency-band correction itself from being applied directly within the excluded facial supports. However, the resulting depth map was subsequently globally normalized and, after the final 3×3 Gaussian filtering step, normalized again. Consequently, the exported facial depth values were not mathematically guaranteed to remain identical to the original DA3+Remap values.

A.4.3 Final Polygon-Based Facial Supports

Later parametric and geometry-aware experiments used deterministic masks generated by

`create_masks_refined.v6.py`.

These masks were initially defined in the original 3000×2333 coordinate system using explicit polygonal regions.

For David, the mask incorporated:

- the main visible facial region;
- the visible ear and side-face region;
- a small upper-neck region below the chin.

Explicit subtraction polygons removed excessive upper-hair coverage and possible shoulder/chest leakage. The resulting mask was smoothed with a 7×7 Gaussian kernel and thresholded at 68.

For Goliath, the polygonal mask covered the head, face, hair, and a limited lower neck/cut region. Additional subtraction polygons removed excessive lower-left and lower-right background support. The result was smoothed with a 9×9 Gaussian kernel and thresholded at 68.

Both full-resolution masks were then resized to the DA3 working dimensions using nearest-neighbour interpolation and re-binarized. Their final aliases were stored as

`phase0_masks_final/david_face_mask.png`

and

`phase0_masks_final/goliath_head_mask.png`.

These masks were subsequently used by Exp101B, Exp105A, the DSINE-based analysis, and Exp115 R1.

The file identities were additionally verified against the corresponding `phase0_masks_refined_v6` outputs using SHA-256 hashes.

A.4.4 Foreground-Mask Standardization

The final foreground support was resized to the DA3 working dimensions with nearest-neighbour interpolation when necessary. Its polarity was determined automatically from the border pixels. If more than 50 interpreted as inverted and complemented.

The final convention was

$$M_{\text{fg}}(x, y) = \begin{cases} 1, & \text{foreground,} \\ 0, & \text{background.} \end{cases} \quad (\text{A.42})$$

Table A.11: Semantic supports used by the representative experiments.

Experiment	Support source	Relevant processing
Exp46	GroundingDINO–SAM David/Goliath masks	Goliath v2 refinement; masks re-sized to the working depth map; 91×91 Gaussian soft-mask generation
Exp80A	Foreground support and separate face-exclusion masks	55×55 foreground erosion; 45×45 face dilation; faces removed from the body-enhancement zone
Exp101B Exp105A	/ Final polygon-based v6 facial masks	Full-resolution polygon construction; Gaussian edge smoothing; nearest-neighbour DA3 alignment
Exp115 R1	Final polygon-based v6 facial masks	Same DA3-aligned masks; additional 5-pixel erosion used internally to define a safe support

A.5 Phase-2B Semantic Reinjection and Phase-2C Intermediate Facial Baseline

The later geometry-aware experiments did not operate directly on the original DA3+Remap depth map. Their native input was an intermediate facial baseline produced in two successive stages: Phase-2B multiscale semantic reinjection and Phase-2C localized correction around Goliath’s mouth.

A.5.1 Phase-2B Multiscale Semantic Reinjection

Phase-2B was implemented in

`phase2B_multiscale_semantic_reinjection.py`.

The input depth was

`Da3+remap.tiff`,

which was min-max normalized to $[0, 1]$. The David and Goliath semantic masks were loaded from `phase0_masks_final`. Two feature-line maps were used for each subject: a structure map and a higher-recall detail map.

For a binary semantic mask M , structure map S , and detail map Q , the feature gate was defined as

$$G = \text{norm} [\mathcal{G}_{2.0} ((w_s S + w_d Q) M)], \quad (\text{A.43})$$

where $\text{norm}(\cdot)$ denotes normalization by the maximum non-zero value. The structure/detail weights were

$$(w_s, w_d) = \begin{cases} (0.65, 0.35), & \text{David,} \\ (0.80, 0.20), & \text{Goliath.} \end{cases} \quad (\text{A.44})$$

A soft semantic support M_s was constructed by eroding the binary face/head mask with an 11×11 elliptical structuring element, applying Gaussian smoothing with $\sigma = 8$, and normalizing the result to $[0, 1]$.

Three depth-frequency components were extracted from the normalized baseline D :

$$D_2 = \mathcal{G}_2(D), \quad D_6 = \mathcal{G}_6(D), \quad D_{14} = \mathcal{G}_{14}(D), \quad (\text{A.45})$$

$$F = \text{clip}(D - D_2, -0.05, 0.05), \quad (\text{A.46})$$

$$M = \text{clip}(D_2 - D_6, -0.08, 0.08), \quad (\text{A.47})$$

and

$$L = \text{clip}(D_6 - D_{14}, -0.10, 0.10), \quad (\text{A.48})$$

where F , M , and L denote the fine-, mid-, and large-scale components, respectively.

For a parameter set $(a_f, a_m, a_l, a_c, a_r)$, the multiscale detail term was

$$T_d = a_f F + a_m M + a_l L. \quad (\text{A.49})$$

This term was modulated by both the feature gate and the soft semantic support:

$$T_g = T_d (0.45 + 0.75G) M_s. \quad (\text{A.50})$$

A local-contrast term was also included. A second soft support was generated using 13×13 erosion and Gaussian smoothing with $\sigma = 9$. Let μ_M denote the mean depth over pixels for which this support exceeded 0.2. The local contrast field was

$$C = \text{clip}(D - \mu_M, -0.18, 0.18) M_c, \quad (\text{A.51})$$

where M_c denotes the corresponding soft contrast support. The contrast contribution was then

$$T_c = a_c C M_s. \quad (\text{A.52})$$

A soft ridge cue was constructed from the feature gate:

$$R = \text{clip}(\mathcal{G}_{1.2}(G) - \mathcal{G}_{5.0}(G), -0.25, 0.25), \quad (\text{A.53})$$

with contribution

$$T_r = a_r R M_s. \quad (\text{A.54})$$

The refined depth for each semantic region was therefore

$$D' = \text{clip}(D + T_g + T_c + T_r, 0, 1). \quad (\text{A.55})$$

David was processed first and Goliath second using the same selected parameter set. Three principal configurations were evaluated:

Table A.12: Exact Phase-2B multiscale semantic-reinjection parameters.

Variant	Fine	Mid	Large	Contrast	Ridge
R4 Safe Plus	0.18	0.22	0.08	0.035	0.018
R5 Balanced	0.26	0.32	0.11	0.055	0.030
R6 Strong	0.34	0.42	0.14	0.075	0.045

R4 remained conservative and produced relatively weak visible modification. R6 produced stronger local variation but increased the risk of noisy or exaggerated behaviour, especially around the lower part of Goliath’s face. R5 provided the most balanced qualitative behaviour among the tested variants and was therefore retained. Its exported depth map was

`phase2B_R5_balanced_depth_uint16.tiff`.

A.5.2 Phase-2C Localized Goliath Mouth Correction

Phase-2C was implemented in

`phase2C_v4_auto_elliptical_mouth_refinement.py`.

This stage used two separately min–max-normalized depth maps: the original `Da3+remap.tiff` representation, denoted D_B , and the selected Phase-2B R5 depth, denoted D_{R5} . The R5 difference field was

$$\Delta_{R5} = D_{R5} - D_B. \quad (\text{A.56})$$

The objective was to reduce excessive R5 enhancement around Goliath’s mouth while preserving the remainder of the semantic refinement. The final localization used the bounding box of the Goliath head mask. Let (x_1, y_1) denote the upper-left corner of the bounding box, with width w and height h . The retained Version-4 ellipse was defined by

$$c_x = x_1 + 0.61w, \quad c_y = y_1 + 0.65h, \quad (\text{A.57})$$

with semi-axes

$$a_x = 0.16w, \quad a_y = 0.10h. \quad (\text{A.58})$$

The filled ellipse was intersected with the Goliath head mask and converted to a soft mouth support M_{mouth} using Gaussian smoothing with $\sigma = 8$, followed by normalization to $[0, 1]$.

For suppression coefficient s , the R5 correction was attenuated inside the mouth support as

$$D_s = D_B + \Delta_{R5} (1 - sM_{\text{mouth}}). \quad (\text{A.59})$$

A second localized smoothing step was then applied. Let

$$D_{\text{sm}} = \mathcal{G}_{1.0}(D_s). \quad (\text{A.60})$$

For smoothing strength q , the final Phase-2C depth was

$$D_{\text{Phase2C}} = D_s (1 - qM_{\text{mouth}}) + D_{\text{sm}} (qM_{\text{mouth}}), \quad (\text{A.61})$$

followed by clipping to $[0, 1]$.

Three suppression/smoothing combinations were tested:

Table A.13: Exact Phase-2C Version-4 mouth-correction parameters.

Variant	Suppression s	Smoothing strength q
E1 Light	0.25	0.12
E2 Balanced	0.40	0.18
E3 Strong	0.55	0.22

E1 produced only limited correction. E3 suppressed a larger fraction of the lower-face variation and risked excessive flattening around the mouth. E2 provided the selected balance between artifact reduction and preservation of the surrounding R5 facial enhancement.

The retained Phase-2C output was

`phase2C_v4_R5_E2_balanced_elliptical_mouth_refined_uint16.tiff`.

This file became the native input baseline for the subsequent geometry-aware experiments, including Exp115 R1. In the reproducibility repository, both `phase2B_multiscale_semantic_reinjection.py` and `phase2C_v4_auto_elliptical_mouth_refinement.py` should be retained together with the corresponding masks, feature-line gates, and intermediate 16-bit depth maps.

A.6 Anatomical Landmark Definition

Two different landmark strategies were used during the experimental progression. The parametric experiments used a compact set of semi-manually specified anatomical reference points, whereas the later geometry-aware experiments used dense landmarks obtained from MediaPipe Face Landmarker.

A.6.1 Semi-Manual Landmarks for Parametric Facial Refinement

Semi-manual anatomical landmarks were introduced in Exp100C after automatic image-feature strategies were found to provide unreliable anatomical localization on the painting.

The landmarks were originally specified in a 2048×1592 coordinate system and converted to the 1022×798 working depth-map resolution.

The expanded landmark set subsequently used by Exp101 and inherited by Exp101B and Exp105A is listed in Table A.14.

Table A.14: Semi-manual landmark coordinates at the 1022×798 working resolution.

Landmark	David (x, y)	Goliath (x, y)
Left eye	(454, 155)	(811, 561)
Right eye	(499, 170)	(886, 564)
Nose tip	(484, 198)	(851, 602)
Mouth left	(464, 221)	(831, 672)
Mouth right	(494, 223)	(871, 674)
Mouth centre	(479, 222)	(851, 673)
Chin	(452, 263)	(848, 727)
Face centre	(474, 196)	(848, 594)
Left cheek	(445, 202)	(808, 620)
Right cheek	(505, 211)	(892, 618)

The cheek centres were parametrically positioned according to the observed face structure. The landmark coordinates were used to locate smooth Gaussian and polyline-based primitives; they did not directly prescribe the final depth values.

The same landmark set was retained across the controlled Exp101 variants.

A.6.2 MediaPipe Landmarks for Geometry-Aware Refinement

For the later geometry-aware experiments, dense facial landmarks were obtained using MediaPipe Face Landmarker with

`face_landmarker.task`.

For each person, a crop was constructed from the corresponding final face/head mask. Relative crop-padding ratios of 0.35 for David and 0.22 for Goliath were used.

The following MediaPipe thresholds were fixed to 0.25:

- minimum face-detection confidence;

- minimum face-presence confidence;
- minimum tracking confidence.

Detected crop-space landmarks were transformed back into the coordinate system of the original 3000×2333 image and saved as

`david_landmarks_full_float32.npy`

and

`goliath_landmarks_full_float32.npy`.

Exp115 subsequently rescaled these coordinates from the original-image coordinate system to the working depth resolution.

Anatomical centres were computed by averaging selected MediaPipe landmark groups rather than using a single point. The principal groups are reported in Table A.15.

Table A.15: MediaPipe landmark groups used by Exp115.

Anatomical structure	MediaPipe indices
Left eye	33, 133, 159, 145
Right eye	362, 263, 386, 374
Nose bridge	6, 168, 197
Nose tip	1, 4, 5
Left cheek	50, 101, 117, 123
Right cheek	280, 330, 346, 352
Upper lip	0, 13, 37, 267
Lower lip	14, 17, 84, 314
Chin	152, 175, 199

The inter-eye distance was used as the primary relative spatial scale for the anatomical support functions, allowing the same parameterization to adapt to the different apparent sizes of David’s and Goliath’s faces.

A.7 Representative Experimental Configurations

The five configurations in Table A.16 represent the principal stages selected for reproducible cross-family comparison.

Table A.16: Representative refinement configurations and their principal parameters.

Experiment	Family	Principal parameters	Role
Exp46	Pixal3D-guided large-scale volume fusion	David blend $\alpha_D = 0.18$; Goliath blend $\alpha_G = 0.25$; mask blur = 91; volume blur = 121; Pixal3D clipping P_2 – P_{98} ; cue inversion; local DA3 range matching P_5 – P_{95}	Representative external-geometry fusion configuration
Exp80A	Safe interior body/-drapery refinement	Mid-scale weight = 0.055; low-frequency weight = 0.025; foreground erosion = 55×55 ; face dilation = 45×45 ; support blur = 71×71 ; low-frequency $\sigma = 45$; mid-scale $\sigma = 10$; reference low-pass $\sigma = 90$; final Gaussian kernel = 3×3	Representative boundary-safe semantic refinement
Exp101B	Parametric facial prior	Global amplitude $\alpha = 0.065$; David scale = 1.00; Goliath scale = 0.78; bridge = 0.80; nose = 1.00; lips = 0.46; mouth valley = -0.18 ; cheeks = 0.36; eye sockets = -0.30 ; chin = 0.26; jaw = 0.22	Representative balanced parametric facial configuration
Exp105A	Signed anatomical midface template	Reference amplitude = 0.065; template weight = 0.50; iterations = 250; data weight = 8.0; smoothness weight = 3.0; constraint weight = 5.0; David scale = 1.00; Goliath scale = 0.78; bridge = 0.72; nose tip = 0.92; cheeks = 0.38; broad midface = 0.22; eye sockets = -0.34 ; nasolabial valleys = -0.20	Representative structured signed-anatomical configuration
Exp115 R1	Confidence-relaxed landmark-guided refinement	Radius scale = 1.00; global face = 0.00010; nose bridge = 0.00130; nose tip = 0.00170; cheeks = 0.00075; upper lip = 0.00055; lower lip = 0.00070; chin = 0.00065; eye sockets = -0.00042 ; maximum absolute delta = 0.00175; boundary erosion = 5 px; David floor/gamma = 0.35/0.70; Goliath floor/gamma = 0.48/0.55; Goliath scale = 0.78; curvature range = 0.72–1.00; final smoothing $\sigma = 1.10$	Representative conservative geometry-aware configuration; emphasizes bounded local control rather than fabrication-scale enhancement

A.7.1 Exp46: Large-Scale External-Geometry Fusion

Exp46 started directly from the DA3+Remap baseline. Pixal3D EXR depth cues were used for David’s face and Goliath’s head.

Valid Pixal3D values were clipped between the second and 98th percentiles and min–max normalized. The cue was then inverted so that regions interpreted as geometrically closer produced stronger relief.

To suppress small unstable geometric structures, the cue was filtered using a 121×121

Gaussian kernel. Its range was then mapped to the local DA3 depth interval defined by the fifth and 95th percentiles within the corresponding facial support.

If V denotes the normalized low-frequency external cue and $[b_{\min}, b_{\max}]$ the local DA3 percentile range, then

$$V' = V(b_{\max} - b_{\min}) + b_{\min}. \quad (\text{A.62})$$

Local fusion was performed as

$$L = (1 - \alpha)D + \alpha V', \quad (\text{A.63})$$

followed by soft-mask blending:

$$D_{\text{out}} = D(1 - M_s) + LM_s. \quad (\text{A.64})$$

The blend coefficients were 0.18 for David and 0.25 for Goliath.

A.7.2 Exp80A: Safe Interior Body and Drapery Refinement

Exp80A also operated directly on the DA3+Remap baseline.

The foreground mask was eroded using a 55×55 kernel to exclude the outer foreground boundary. The David and Goliath face masks were combined and dilated using a 45×45 kernel. The dilated facial support was then removed from the eroded foreground region.

The resulting body support was softened using a 71×71 Gaussian filter.

Two depth-frequency bands were extracted. The low-frequency component was based on

$$S_{\text{low}} = \text{robust}(G_{45}(D) - G_{90}(D)), \quad (\text{A.65})$$

whereas the intermediate component was

$$S_{\text{mid}} = \text{robust}(G_{10}(D) - G_{45}(D)). \quad (\text{A.66})$$

The modified depth was computed as

$$D' = D + 0.025 S_{\text{low}}W + 0.055 S_{\text{mid}}W. \quad (\text{A.67})$$

where W is the softened interior-body support.

The resulting depth map was subsequently min-max normalized. A final 3×3 Gaussian filter was then applied. Here, the value 3 refers to the kernel size rather than to a Gaussian standard deviation. After this filtering step, a second min-max normalization was applied before unsigned 16-bit export.

A.7.3 Exp101B: Parametric Facial Prior

Exp101B used the parametric prior defined in Exp101A and introduced only one controlled change: the global prior amplitude was increased from

$$0.050 \rightarrow 0.065. \quad (\text{A.68})$$

All other parameters, including the base depth, masks, landmarks, primitive geometry, component weights, boundary weighting, and mean-shift correction, were unchanged.

The signed facial prior was composed as

$$P = 0.80P_{\text{bridge}} + 1.00P_{\text{nose}} + 0.46P_{\text{lips}} - 0.18P_{\text{mouth valley}} + 0.36P_{\text{cheeks}} - 0.30P_{\text{eyes}} + 0.26P_{\text{chin}} + 0.22P_{\text{jaw}}. \quad (\text{A.69})$$

The modification was

$$D' = \text{clip}(D + 0.065P, 0, 1). \quad (\text{A.70})$$

The person-dependent prior scales were 1.00 for David and 0.78 for Goliath.

A.7.4 Exp105A: Signed Anatomical Midface Template

Exp105A retained the constrained relaxation framework of Exp103A but introduced an additional structured signed midface template.

The template was

$$T = 0.72T_{\text{bridge}} + 0.92T_{\text{nose}} + 0.38T_{\text{cheeks}} + 0.22T_{\text{midface}} - 0.34T_{\text{eyes}} - 0.20T_{\text{nasolabial}}. \quad (\text{A.71})$$

The signed-template contribution was applied with weight

$$0.50. \quad (\text{A.72})$$

The reference target amplitude was 0.065, and the relaxation solver used 250 iterations with

$$w_{\text{data}} = 8.0, \quad w_{\text{smooth}} = 3.0, \quad w_{\text{constraint}} = 5.0. \quad (\text{A.73})$$

The David person scale remained 1.00, while the Goliath person scale was 0.78.

A.7.5 Exp115 R1: Confidence-Relaxed Landmark-Guided Refinement

Exp115 R1 was applied to the later refined facial baseline

`phase2C_v4_R5_E2_balanced_elliptical_mouth_refined_uint16.tiff`,

rather than directly to the original DA3+Remap TIFF.

The David and Goliath facial masks were eroded by five pixels:

$$M_s = \text{erode}(M, 5). \quad (\text{A.74})$$

Confidence values inside this safe mask were normalized using their 99th percentile:

$$C_n = \text{clip}\left(\frac{C}{P_{99}(C \mid M_s) + \varepsilon}, 0, 1\right). \quad (\text{A.75})$$

A relaxed confidence gate was then constructed as

$$C_r = f + (1 - f)C_n^\gamma. \quad (\text{A.76})$$

where the David parameters were

$$f_D = 0.35, \quad \gamma_D = 0.70, \quad (\text{A.77})$$

and the Goliath parameters were

$$f_G = 0.48, \quad \gamma_G = 0.55. \quad (\text{A.78})$$

Thus, confidence acted as a soft attenuation mechanism rather than a binary threshold.

A curvature cue was computed from two Gaussian scales:

$$K = |G_{1.5}(D) - G_{5.0}(D)|. \quad (\text{A.79})$$

The cue was normalized within the valid facial support using its 97.5th percentile and mapped to the range $[0.72, 1.00]$:

$$W_K = 0.72 + (1.00 - 0.72)K_n. \quad (\text{A.80})$$

For the local anatomical components, the common spatial gate was

$$G_c = C_r W_K M_s. \quad (\text{A.81})$$

The global-face component was intentionally more strongly attenuated than the localized anatomical components.

The R1 amplitudes were

$$\begin{aligned} A_{\text{global}} &= 0.00010, \\ A_{\text{bridge}} &= 0.00130, \\ A_{\text{nose}} &= 0.00170, \\ A_{\text{cheeks}} &= 0.00075, \\ A_{\text{upperlip}} &= 0.00055, \\ A_{\text{lowerlip}} &= 0.00070, \\ A_{\text{chin}} &= 0.00065, \\ A_{\text{eyes}} &= -0.00042. \end{aligned} \quad (\text{A.82})$$

The person-dependent correction scales were

$$a_D = 1.00, \quad a_G = 0.78. \quad (\text{A.83})$$

After summing the signed anatomical components, a support-weighted offset correction was applied to suppress unnecessary mean elevation. The result was then smoothed with a Gaussian filter using

$$\sigma = 1.10. \quad (\text{A.84})$$

The final local modification was bounded by

$$|\Delta D| \leq a 0.00175, \quad (\text{A.85})$$

where $a = a_D$ for David and $a = a_G$ for Goliath.

All modifications were explicitly restricted to the union of the David and Goliath facial supports. Pixels outside these supports were copied unchanged from the input baseline.

The quantitative interpretation in Section A.3.1 shows that the resulting Exp115 R1 correction is extremely small at the adopted physical mapping: the active-pixel mean is approximately $1.09\,\mu\text{m}$, the 99th percentile is approximately $12.26\,\mu\text{m}$, and the maximum is approximately $22.30\,\mu\text{m}$ ($0.02230\,\text{mm}$). Consequently, Exp115 R1 is retained as a representative conservative geometry-aware stage because it demonstrates controlled anatomical localization and preservation of the inherited depth field, not because it establishes a fabrication-relevant or tactile improvement. Verification of perceptual benefit would require physical fabrication and a dedicated tactile user study.

Statement on the Use of AI Tools

Generative AI tools, primarily ChatGPT, were used during the preparation of this thesis to support language editing, summarisation, restructuring of draft text, and the drafting and debugging of code and LaTeX. AI-generated outputs were reviewed and verified against the experimental files, scripts, and reported results before being incorporated into the thesis. The author remains fully responsible for the final text, code, reported results, interpretations, and conclusions.

References

- [1] Gwangbin Bae and Andrew J. Davison. Rethinking inductive biases for surface normal estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9535–9545, 2024.
- [2] Zhe Bian and Shi-Min Hu. Preserving detailed features in digital bas-relief making. *Computer Aided Geometric Design*, 28(4):245–256, 2011.
- [3] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, pages 187–194, 1999.
- [4] Luis Cavazos Quero, Jorge Iranzo Bartolomé, and Jundong Cho. Accessible visual artworks for blind and visually impaired people: Comparing a multimodal approach with tactile graphics. *Electronics*, 10(3):297, 2021.
- [5] David Eigen, Christian Puhrsch, and Rob Fergus. Depth map prediction from a single image using a multi-scale deep network. In *Advances in Neural Information Processing Systems*, volume 27, pages 2366–2374, 2014.
- [6] Yunfei Fu, Hongchuan Yu, Chih-Kuo Yeh, Jianjun Zhang, and Tong-Yee Lee. High relief from brush painting. *IEEE Transactions on Visualization and Computer Graphics*, 25(9):2763–2776, 2019.
- [7] Zhiheng Fu, Siyu Hong, Mengyi Liu, Hamid Laga, Mohammed Bennamoun, Farid Bous-said, and Yulan Guo. Multi-stage information diffusion for joint depth and surface normal estimation. *Pattern Recognition*, 141:109660, 2023.
- [8] Rocco Furferi. Deep learning approaches for 3d model generation from 2d artworks to aid blind people with tactile exploration. *Heritage*, 8(1):12, 2025.
- [9] Rocco Furferi, Lapo Governi, Niccolò Vanni, and Yary Volpe. Tactile 3d bas-relief from single-point perspective paintings: A computer based method. *Journal of Information and Computational Science*, 11(16):5667–5680, 2014.
- [10] Rocco Furferi, Lapo Governi, Yary Volpe, Luca Puggelli, Niccolò Vanni, and Monica Carfagni. From 2d to 2.5d i.e. from painting to tactile model. *Graphical Models*, 76(6):706–723, 2014.
- [11] Rocco Furferi, Lapo Governi, Yary Volpe, Michaela Servi, and Francesco Buonamici. From pictorial space to tactile form: A comparative evaluation of ai-based 2.5d reconstruction from modern artwork paintings. *Frontiers in Computer Science*, 8:1821454, 2026.

- [12] Morton A. Heller. Tactile picture perception in sighted and blind people. *Behavioural Brain Research*, 135(1–2):65–68, 2002.
- [13] Mu Hu, Wei Yin, Chi Zhang, Zhipeng Cai, Xiaoxiao Long, Hao Chen, Kaixuan Wang, Gang Yu, Chunhua Shen, and Shaojie Shen. Metric3d v2: A versatile monocular geometric foundation model for zero-shot metric depth and surface normal estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):10579–10596, 2024.
- [14] Inwook Hwang, Sungryul Yun, and Jaeyoung Park. Haptic perception of 2.5d surface feature height. *Scientific Reports*, 16:12116, 2026.
- [15] Sandra Jehoel, Paul T. Sowden, Simon Ungar, and Annette Sterr. Tactile elevation perception in blind and sighted participants and its implications for tactile map creation. *Human Factors*, 51(2):208–223, 2009.
- [16] Bingxin Ke, Anton Obukhov, Shengyu Huang, Nando Metzger, Rodrigo Caye Daudt, and Konrad Schindler. Repurposing diffusion-based image generators for monocular depth estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9492–9502, 2024.
- [17] Dong-Yang Li, Wang Zhao, Yuxin Chen, Wenbo Hu, Meng-Hao Guo, Fang-Lue Zhang, Ying Shan, and Shi-Min Hu. Pixel3d: Pixel-aligned 3d generation from images. *arXiv preprint arXiv:2605.10922*, 2026.
- [18] Haotong Lin, Sili Chen, Jun Hao Liew, Donny Y. Chen, Zhenyu Li, Yang Zhao, Sida Peng, Hengkai Guo, Xiaowei Zhou, Guang Shi, Jiashi Feng, and Bingyi Kang. Depth anything 3: Recovering the visual space from any views. In *International Conference on Learning Representations (ICLR)*, 2026.
- [19] Yanlong Lin, Ping Wang, Jiyuan Liu, and Shouhui Zhang. Self-supervised monocular depth estimation via local-global semantic enhancement. In *2024 13th International Conference on Computing and Pattern Recognition (ICCPR)*, pages 233–239, 2024.
- [20] Yipeng Liu, Zhongping Ji, Yu-Wei Zhang, and Gang Xu. Example-driven modeling of portrait bas-relief. *Computer Aided Geometric Design*, 80:101860, 2020.
- [21] Yunze Man, Xinshuo Weng, Xi Li, and Kris Kitani. Groundnet: Monocular ground plane normal estimation with geometric consistency. In *Proceedings of the 27th ACM International Conference on Multimedia*, pages 2170–2178, 2019.
- [22] Mukhriddin Mukhiddinov and Soon-Young Kim. A systematic literature review on the automatic creation of tactile graphics for the blind and visually impaired. *Processes*, 9(10):1726, 2021.
- [23] Daniel Scharstein and Richard Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47:7–42, 2002.
- [24] Johannes L. Schönberger and Jan-Michael Frahm. Structure-from-motion revisited. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4104–4113, 2016.

- [25] Xianfang Sun, Paul L. Rosin, Ralph R. Martin, and Frank C. Langbein. Bas-relief generation using adaptive histogram equalization. *IEEE Transactions on Visualization and Computer Graphics*, 15(4):642–653, 2009.
- [26] Jianyuan Wang, Minghao Chen, Nikita Karaev, Andrea Vedaldi, Christian Rupprecht, and David Novotny. Vggt: Visual geometry grounded transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5294–5306, 2025.
- [27] Shuzhe Wang, Vincent Leroy, Yohann Cabon, Boris Chidlovskii, and Jerome Revaud. Dust3r: Geometric 3d vision made easy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20697–20709, 2024.
- [28] Mingqiang Wei, Yang Tian, Wai-Man Pang, Charlie C. L. Wang, Ming-Yong Pang, Jun Wang, Jing Qin, and Pheng-Ann Heng. Bas-relief modeling from normal layers. *IEEE Transactions on Visualization and Computer Graphics*, 25(4):1651–1665, 2019.
- [29] Tim Weyrich, Jia Deng, Connelly Barnes, Szymon Rusinkiewicz, and Adam Finkelstein. Digital bas-relief from 3d scenes. *ACM Transactions on Graphics*, 26(3):32:1–32:7, 2007.
- [30] Gangwei Xu, Haotong Lin, Hongcheng Luo, Xianqi Wang, Jingfeng Yao, Lianghui Zhu, Yuechuan Pu, Cheng Chi, Haiyang Sun, Bing Wang, Guang Chen, Hangjun Ye, Sida Peng, and Xin Yang. Pixel-perfect depth with semantics-prompted diffusion transformers. In *Advances in Neural Information Processing Systems*, volume 38, 2025.
- [31] Xianfa Xu, Zhe Chen, and Fuliang Yin. Multi-scale spatial attention-guided monocular depth estimation with semantic enhancement. *IEEE Transactions on Image Processing*, 30:8811–8822, 2021.
- [32] Zhenjie Yang, Beijia Chen, Youyi Zheng, Xiang Chen, and Kun Zhou. Human bas-relief generation from a single photograph. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):4558–4569, 2022.
- [33] Ruo Zhang, Ping-Sing Tsai, James Edwin Cryer, and Mubarak Shah. Shape from shading: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(8):690–706, 1999.
- [34] Beibit Zharylkassyn, Asma Perveen, and Didier Talamona. Effect of process parameters and materials on the dimensional accuracy of fdm parts. *Materials Today: Proceedings*, 44:1307–1311, 2021.