Parametric Shape Estimation of Human Body Under Wide Clothing

Aswin P T
Hariraj S

Introduction to the SHADER Paper

The Core Concept

The paper *"Parametric Shape Estimation of Human Body Under Wide Clothing"* proposes a method to estimate body shape even when hidden by loose clothes.

Key Idea: Compute a Silhouette Confidence Map that tells the network which pixels are reliable (body) and which are unreliable (clothing).

Evaluation Scope

The method is evaluated on a massive scale to ensure robustness:

- 300,000+ synthetic clothing variations
- Real images and 3D scans
- Integrations with SOTA regressors (SPIN, CMR, HMR)

What the Paper Actually Does

1. Synthetic Dataset

Generates thousands of avatars using CLO3D with naked/dressed pairs.

2. Confidence Map

Calculates confidence decay based on distance from the naked body.

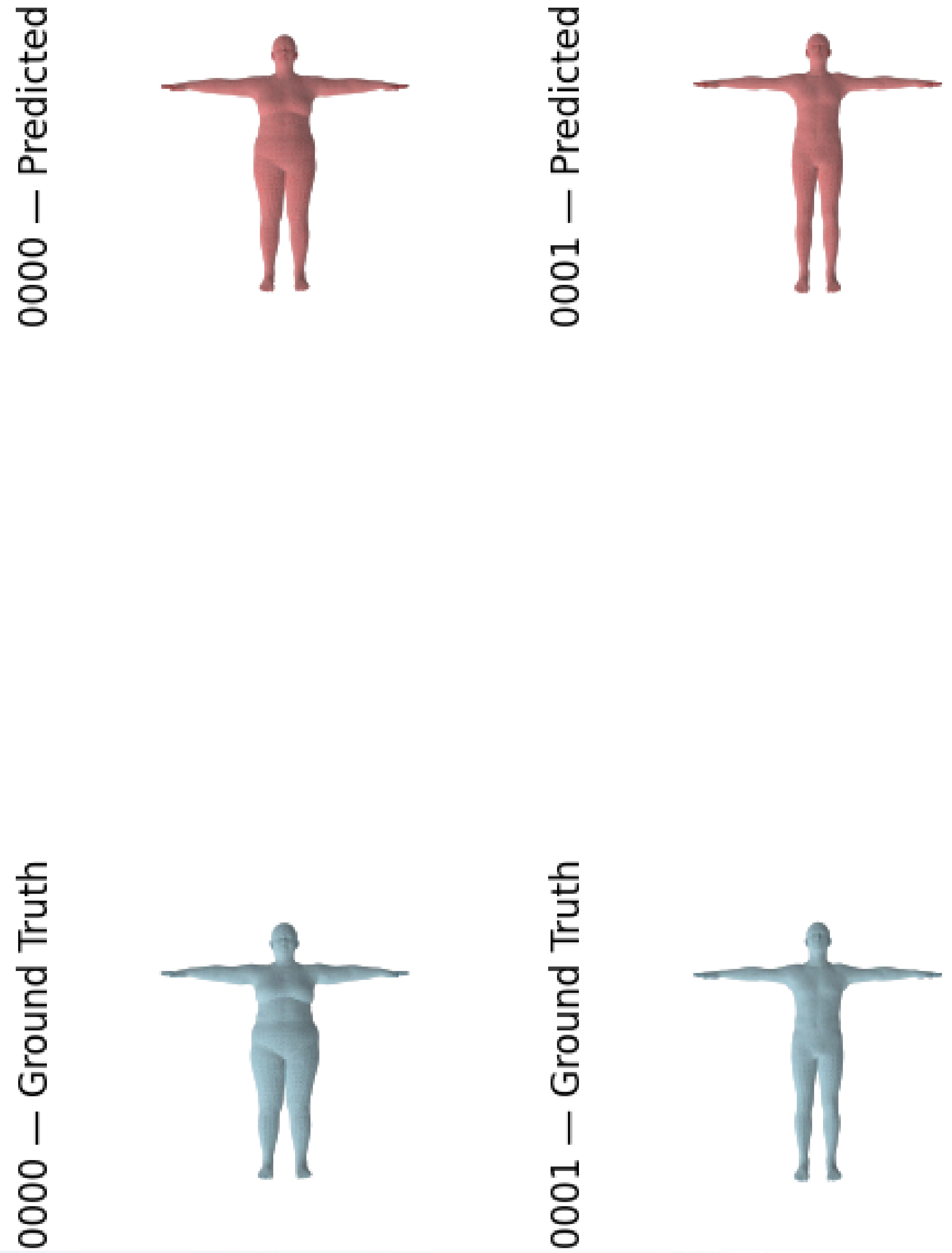
3. Hourglass CNN

Trains a 6-stack Hourglass network to predict confidence from RGB.

4. Integration

Multiplies RGB by confidence to guide shape regressors.

Problem SHADER Solves



The Clothing Ambiguity

Estimating accurate body shape under clothing is inherently difficult because clothing distorts the true silhouette.

The Solution: Confidence Maps

SHADER reduces this ambiguity using **Confidence Maps**.

- Assigns a probability score to each pixel.
- High confidence = Likely real body boundary.
- Low confidence = Likely loose clothing or noise.

SHADER Confidence Formulation

1. Distance Transform

Measures the "gap" between clothing and body.

$$d_{i,j} = \min_{p_k \in P_{body}} p_i - p_{k}{}_{1}$$

2. Gaussian Decay

Converts distance into a confidence score.

$$c_{i,j} = \exp - \frac{d_{i,j}^{\;2}}{2\sigma_c^2}$$

Our Implementation Pipeline

Preprocessing

Silhouette extraction and Distance Transform computation.

Confidence

Gaussian confidence generation and 2-channel Hourglass prediction.

Refinement

Confidence × RGB multiplication ($I' = I \odot C$).

Estimation

Feed to SMPL estimator and compute vertex/MSE error.

Training the Hourglass Network

Implementation: HourglassStackFixed

We built a multi-stack Hourglass network (6 stacks) to predict confidence maps from RGB inputs.

- **Architecture:** 6 Stacks with residual blocks.
- **Channels:** Input RGB (3) → Output Confidence (1).
- **Loss Function:** MSE against Ground Truth Confidence.

Training Details

- **Sample Size:** 9000 samples
- **Optimizer:** RMSProp
- **Resolution:** 224x224

Motivation for AMR (Adaptive Multi-Region)

⚠️Problem with SHADER

Uses a **single σ** for the whole body.

Reality: Clothing looseness varies!

- ✕ Upper body: Often tight
- ✕ Pelvis/Legs: Often loose/baggy

💡AMR Solution

A single Gaussian cannot capture regional clothing variations effectively.

Concept: Region-specific confidence decay.

AMR allows confidence to adapt region-wise, reducing over-confidence in loose areas like skirts or baggy pants.

What AMR Does

Region-Specific Equation

We introduce a region-dependent parameter σ_R.

$$c_{i,j}^{(R)} = \exp\left(-\frac{d_{i,j}^{2}}{2\sigma_R^2}\right)$$

This allows for more realistic modeling of clothing, preventing the model from assuming tight fits everywhere.

Assigned σ Values

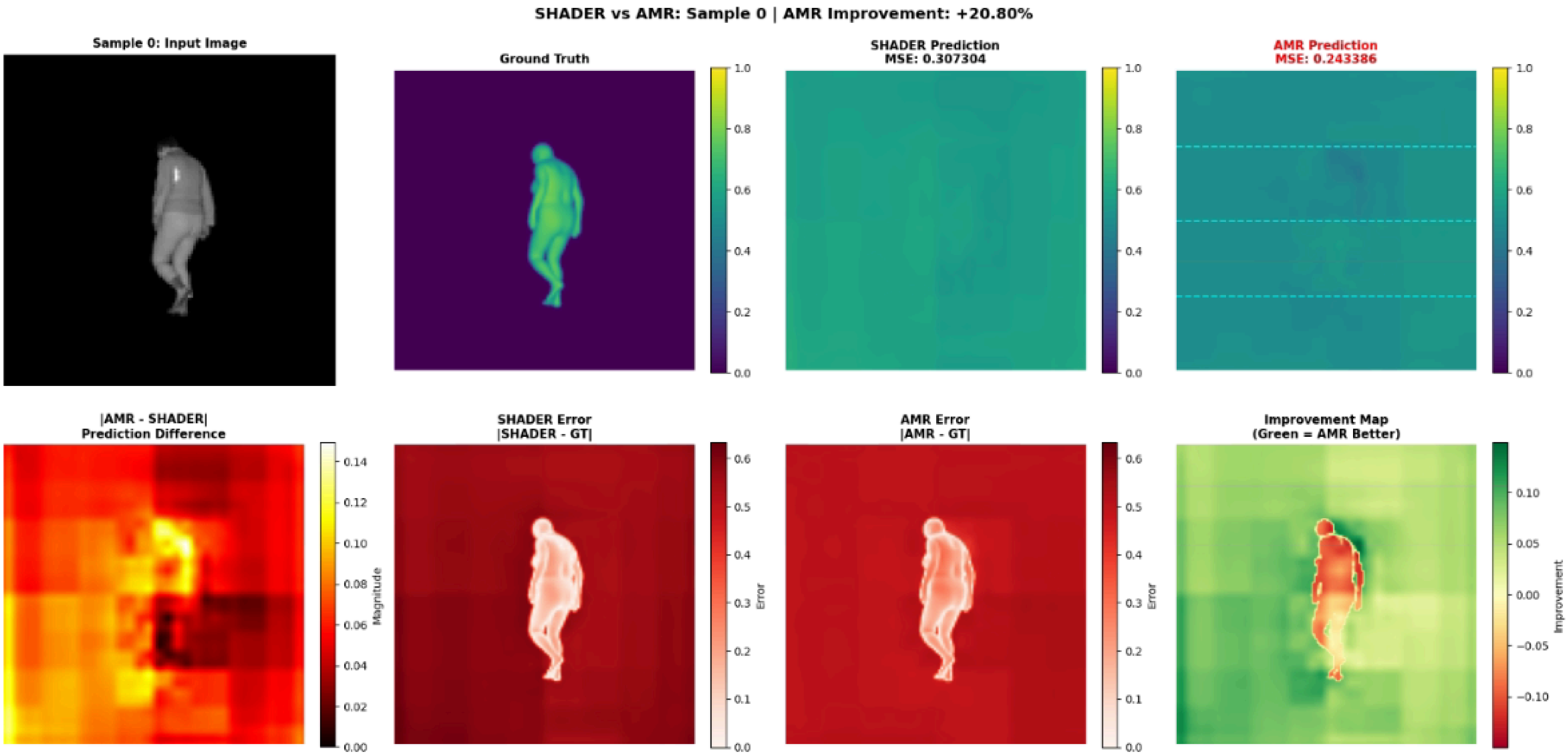
Body Region	σ Value	Clothing Type
Upper Body	**0.6**	Tight / Fitted
Torso	**0.8**	Moderate
Legs	**1.0**	Medium
Pelvis	**1.2**	Loose / Baggy

Results: SHADER vs AMR

Mean Squared Error (Lower is Better)

+23.5% Improvement

Statistically significant ($p < 0.0001$)


SHADER vs AMR: Sample 0 | AMR Improvement: +20.80%
Sample 0: Input Image
Ground Truth
SHADER Prediction
MSE: 0.307304
AMR Prediction
MSE: 0.243386
|AMR - SHADER|
Prediction Difference
Magnitude
SHADER Error
|SHADER - GT|
AMR Error
|AMR - GT|
Error
Improvement Map
(Green = AMR Better)
Improvement

Why SHADER Authors Didn't Use AMR

Simplicity

They aimed for a simple, stable Gaussian confidence model.

Synthetic Data

Their dataset had consistent clothing behavior, so a single sigma worked fine.

Network Learning

Assumed Hourglass CNN implicitly learns region variation.

Complexity

Multi-region sigma increases tuning difficulty and parameter overhead.

Why WE Used AMR

- **Realistic Variation:** For dataset contains diverse clothing styles not captured by a single sigma.
- **Measurable Gains:** Region-based sigma yielded a significant 23.5% error reduction.
- **Better Modeling:** Specifically improved estimation for hips, skirts, and baggy pants.
- **Proof of Concept:** Serves as a foundation for future dynamic confidence modeling.

Limitations of AMR

1. Fixed Region Slicing

AMR uses a static horizontal split, which fails when:

- 🚶 The subject changes pose (bending, sitting).
- ↕ Subject height varies significantly.
- 🔄 Camera angle is rotated or sideways.

2. Manual Tuning

Fixed sigma values require manual calibration and do not generalize to all clothing types automatically.

Future Work: Dynamic AMR

Compute σ from image characteristics instead of fixed values.

Proposed Methods

1. Gap-based:

$$\sigma = \mathrm{mean}(d_{cloth})$$

2. Texture-based:

$$\sigma \propto \mathrm{std}(\mathrm{Laplacian}(I))$$

3. Combined Model:

$$\sigma = a \cdot gap_{mean} + b \cdot tex_{std} + c \cdot int_{std}$$

Advantages

- **Totally Adaptive:** No need for manual region cutting or tuning.
- **Robust:** Works on any pose, height, or clothing style.

Final Conclusion

Key Achievements

- **Full Implementation:** Built SHADER training and inference pipeline from scratch.
- **Novel Extension:** Added AMR to handle regional clothing variance.
- **Proven Results:** Demonstrated +23% improvement in MSE.

Path Forward

- **Critical Analysis:** Identified limitations in fixed-region slicing.
- **Future Roadmap:** Proposed Dynamic-AMR for truly robust estimation.

Q & A

Thank you for your attention.

MSc Computer Science Project