# Registering Explicit to Implicit: Towards High-Fidelity Garment mesh Reconstruction from Single Images – Appendix

In this appendix, we provide more results and details in the following aspects: (1) more implementation details regarding the generation of explicit template mesh, network training, and explicit fitting; (2) more evaluation on the explicit fitting stage; (3) more results reconstructed from the in-the-wild images.

## A. Explicit Template

As is illustrated in Figure.1, the explicit garment template meshes $M_t$ in **ReEF** covers twelve common clothes categories. On top of the garment template meshes $M_t$, we defined nine types of garment boundaries $\{L_t^i\}$. Each type of garment boundary is corresponded to a boundary implicit boundary field $\{f_b^i\}$.

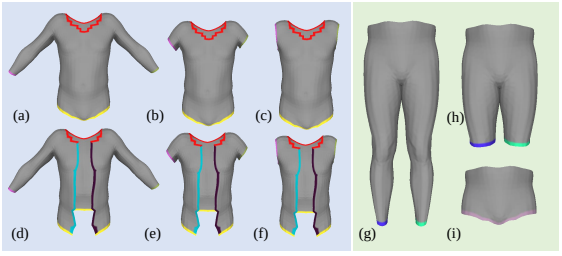


Figure 1. Garment templates supported by **ReEF**: (a) long-sleeve upper clothing and long-sleeve dress; (b) short-sleeve upper clothing and short-sleeve dress; (c) no-sleeve upper clothing and no-sleeve dress; (d) long-sleeve open coat; (e) short-sleeve open coat; (f) no-sleeve open coat; (g) long pants; (h) short pants; (e) skirt. Different kinds of garment boundaries are annotated with distinct colors.

## B. Network Training

In the main paper, we have briefly introduced the generation of the implicit target shape field $f_f$, implicit semantic fields $\{f_s^i\}$ and the implicit boundary fields $\{f_b^i\}$. This section will describe the detailed training settings for the implicit field generation networks.

### B.1. Attention map generation.

As mentioned in the main paper, generating implicit semantic fields $\{f_s^i\}$ and implicit boundary fields $\{f_b^i\}$ requires curve-aligned features $\phi_h(I, \pi(X))$ fetched from the predicted semantic and boundary attention maps. To this end, we generate ground truth boundary heat maps and semantic heat maps as the supervision to guide the network's training: Firstly, we project the points sampled on different semantic regions(or boundaries cylinders) to different image planes with a weak-perspective camera. Then, we generate Gaussian kernels centering at the projected positions with $\sigma$ set to 2. Finally, the semantic/boundary heat maps($\{H_s^i\}$ and $\{H_b^i\}$) can be obtained by fusing the Gaussian kernels with maximum operator on each image plane.

### B.2. Loss Functions.

As mentioned in the main paper, we jointly train the generation module for coarse shape field $f_c$, the semantic fields $\{f_s^i\}$ and the boundary fields $\{f_b^i\}$ with coarse occupancy loss $\mathcal{L}_{cocc}$, semantic attention loss $\mathcal{L}_{hms}$, boundary attention loss $\mathcal{L}_{hmb}$, boundary field loss $\mathcal{L}_b$ and semantic field loss $\mathcal{L}_s$ :

$$\mathcal{L} = \mathcal{L}_{cocc} + \mathcal{L}_{hms} + \mathcal{L}_{hmb} + \mathcal{L}_b + \mathcal{L}_s \tag{1}$$

where the loss for each component is the mean squared error(MSE) between the predicted value and the ground truth.

## C. Explicit Fitting

In the main paper, we have explained the loss functions adopted for deforming the explicit template progressively to fit the implicit shape. On this top, we will provide further details on the explicit fitting regarding the hyper-parameter settings and the post processing.

**Template initialization.** With the purpose of setting up a good initialization for the later stages, we optimized the SMPL body parameters $SMPL(\theta, \beta)$ to be aligned with the implicit clothed body and the predicted 2D joints $J_{gt}$:

$$\begin{aligned} V_{pred}, J_{pred} =& SMPL(\theta, \beta) \\ \mathcal{L}_{body} =& MSE(J'_{pred}, J_{gt}) + \eta_{reg} Reg(\theta) \\ &+ \eta_{shape} CD(V_{lres}, V_{pred}) \end{aligned} \tag{2}$$

where $\eta_{reg}$ is set to $1e^{-3}$ and $\eta_{shape}$ is set to 1.0.

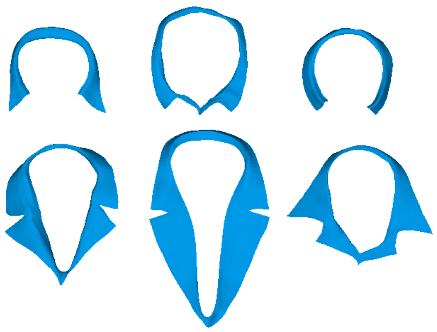

Figure 2. Selected collar templates from our collar warehouse.

**Boundary Fitting.** To further fit the initialized template to align with the garment boundaries of the implicit target, we may deform the boundaries by minimizing the following loss function:

$$\mathcal{L}_b = f_b^i(l_p^i) + \eta_{ea} Avg(e_b^i) + \eta_{ed} Var(e_b^i) \tag{3}$$

where $\eta_{ea}$ is set to $0.025$ and $\eta_{ed}$ is set to $2.5$.

**Shape Fitting** The shape fitting stage will further deform the boundary-aligned garment template to approach the implicit target guided by the following loss function:

$$\begin{aligned}\mathcal{L}_o = D_{act}(M_o) - \eta_{pen} TSDF(M_{smpl})(M_o) \\ +\eta_b \mathcal{L}_b + \eta_{lap} \mathcal{L}_{lap}\end{aligned} \tag{4}$$

where $\eta_{pen}$, $\eta_b$ and $\eta_{lap}$ are set to $0.1$, $0.1$ and $100$ respectively.

**Post Processing** As mentioned in the main paper, our method could recover the garment styles and surface details from an in-the-wild input image though it may fail to generate folded structure, i.e., the collars. Therefore, as Figure.2 illustrates, we firstly create a collar warehouse that covers ten common collars categories. A multi-layer perceptron is then adopted, which takes image features(for coarse shape field generation) sampled from the collar area to predict the type of the collar presents on the image.

## D. Ablation Study

In this section, we compile a set of ablation experiments to verify the effectiveness of each algorithmic component for our explicit fitting module. We provide qualitative comparisons between our proposed method and the alternatives that take other candidate settings: 1) Deform the garment template mesh to fit the implicit target without pose initialization, termed as **w/o Init**. 2) Deform the garment template mesh to fit the implicit target without boundary initialization, termed as **w/o Bound**. 3) Deform the garment template mesh to fit the implicit target without active area probing, termed as **w/o Probe**. 4) The proposed full model, termed as **Ours**.

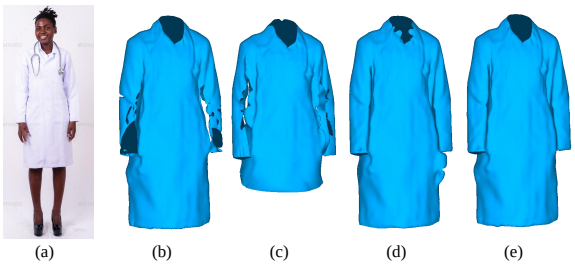


Figure 3. Qualitative comparison of the explicit garment meshes generated under different ablation settings. The input image (a) is followed by the garments generated with (b) **w/o Init**, (c) **w/o Bound**, (3) **w/o Bound** and (4) **Ours**.

Figure.3 and Table.1 demonstrate the qualitative and quantitative comparison between the proposed model and the design alternatives. As the garments are diversified shapes with varying geometrical details, it is inherently hard to strike a balance between the reconstruction accuracy and surface smoothness without proper initialization(**w/o Init** and **w/o Bound**). Although the results generated with **w/o Probe** can well reflect the garment styles and most surface details from the image, the reconstructed surface would be corrupted by non-relevant regions (e.g. the hands for this case). In contrast, **Ours** can produce high-quality garment meshes with accurate styles and surfaces details highly identical to the input image.

| Methods | **w/o Init** | **w/o Bound** | **w/o Probe** | **Ours** |
|---|---|---|---|---|
| Dist($\times 10^{-3}$) | 3.52109 | 70.3211 | 3.56883 | **3.41651** |

Table 1. Comparison on registration accuracy between the proposed method and the ablation alternatives.

## E. More results on in-the-wild images

This section provides more results generated by our method on in-the-wild images from the internet. As is shown in Figure.4, given an in-the-wild image as input, our model could produce high-quality garments with fine-grained details and correct garment styles.

Figure 4. The results generated by our method on in-the-wild images. Each image is followed by the reconstructed layered garment mesh.