

NeAT: Learning Neural Implicit Surfaces with Arbitrary Topologies from Multi-view Images – Supplemental Materials –

Xiaoxu Meng Weikai Chen Bo Yang

Digital Content Technology Center, Tencent Games

{xiaoxumeng, weikaichen, brandonyang}@global.tencent.com

1. Overview

In the main paper, we introduce a novel neural rendering framework, NeAT, that can learn implicit surfaces with arbitrary topologies from multi-view images. This supplemental material consists of:

- Section 2 provides the derivation of the continuous opacity density ρ (Section 2.1), the derivation of the discrete opacity density α (Section 2.2), and the proofs of the unbiasedness and occlusion-awareness of our proposed volume weight function (Section 2.3).
- Section 3 provides implementation details of NeAT-Net.
- Section 4 provides additional qualitative results of multi-view reconstruction.

2. Derivation of Opacity Density

In this section, we first derive the continuous opacity density ρ with probability density defined as $\phi_s(f(\mathbf{p}) \cdot \gamma(\mathbf{p}))$ for the SDF representation. Then, we discretize ρ to derive the discrete opacity density α with the SDF representation. Finally, we prove the unbiasedness and occlusion-awareness of the weight function for arbitrary surfaces w_β .

2.1. Derivation of the Continuous Opacity Density

Given a sampled pixel \mathbf{o} on an input image, we project it to the 3D space and get the sampled 3D points emitting from the pixel $\{\mathbf{p}(t) = \mathbf{o} + t\mathbf{v} | t \geq 0\}$, where \mathbf{v} is the unit direction vector of the ray. The rendered pixel color is

$$I_{pred}(\mathbf{o}, \mathbf{v}) = \int_0^{+\infty} \exp(-\int_0^t \rho(u)du) \cdot \rho(t)c(t)dt. \quad (1)$$

To derive the opaque density $\rho(t)$, we firstly consider a simple case where there is only one plane in the sampling space. Then we generalize $\rho(t)$ to the general case of multiple surface intersections.

Ray Intersects with One Surface In the simple case of a single plane, the signed distance function could be written as

$$f(\mathbf{p}(t)) = \cos(\theta)(t - t^*), \quad (2)$$

where $f(\mathbf{p}(t^*)) = 0$, and θ is the angle between ray \mathbf{v} and the surface normal \mathbf{n} . Because the surface is assumed locally, $\cos(\theta)$ is a constant value. Similarly, as there is one single plane with a constant face orientation, $Sign(\mathbf{v} \cdot \mathbf{n})$ is a constant value as well. We denote this constant value as γ .

By plugging Equation 2 into Equation 4 in the main paper, we have

$$\begin{aligned} w(t) &= \frac{\phi_s(f(\mathbf{p}(t)) \cdot \gamma)}{\int_{-\infty}^{+\infty} \phi_s(f(\mathbf{p}(u)) \cdot \gamma)du} \\ &= \frac{\phi_s(\gamma f(\mathbf{p}(t)))}{\int_{-\infty}^{+\infty} \phi_s(\cos(\theta)(u - t^*)) \cdot \gamma du} \\ &= \frac{\phi_s(\gamma f(\mathbf{p}(t)))}{(\cos(\theta) \cdot \gamma)^{-1} \int_{-\infty}^{+\infty} \phi_s((u - t^*))du} \\ &= \frac{\phi_s(\gamma f(\mathbf{p}(t)))}{(\cos(\theta) \cdot \gamma)^{-1} \cdot 1} \\ &= \phi_s(\gamma f(\mathbf{p}(t))) \cdot \gamma \cdot \cos(\theta) \\ &= \phi_s(\gamma f(\mathbf{p}(t))) \cdot \gamma \cdot \frac{d}{dt}(f(\mathbf{p}(t))) \\ &= \frac{d}{dt}[\Phi_s(\gamma f(\mathbf{p}(t)))] \end{aligned} \quad (3)$$

By denoting $T(t) = \exp(-\int_0^t \rho(u)du)$, we can rewrite Equation 5 in the main paper as:

$$w(t) = T(t)\rho(t) = \frac{dT}{dt}(t) \quad (4)$$

Plug Equation 4 into Equation 3, we have

$$\frac{dT}{dt}(t) = \frac{d}{dt}\Phi_s(\gamma f(\mathbf{p}(t))) \quad (5)$$

Integrating both sides, we have

$$T(t) = \Phi_s(\gamma f(\mathbf{p}(t))) \quad (6)$$

Taking the logarithm and then differentiating both sides, we have

$$\int_{\infty}^t \rho(u) du = -\ln(\Phi_s(f(\mathbf{p}(t)) \cdot \gamma)) \quad (7)$$

$$\Rightarrow \rho(t) = \frac{-\frac{d\Phi_s}{dt}(f(\mathbf{p}(t)) \cdot \gamma)}{\Phi_s(f(\mathbf{p}(t)) \cdot \gamma)} \quad (8)$$

We can generalize the opaque density to the general case where there are multiple surface intersections.

Ray Intersects with Multiple Surfaces We denote $Sign(\mathbf{v} \cdot \mathbf{n})$ as $\gamma(\mathbf{p}(t))$. Following the occlusion-aware weight function (Equation 5 in the main paper), we have

$$\begin{aligned} w(t) &= \exp\left(-\int_0^t \rho(u) du\right) \rho(t) \\ &= \exp\left(-\int_0^{t_i} \rho(u) du\right) \exp\left(-\int_{t_i}^t \rho(u) du\right) \rho(t) \\ &= T(t_i) \exp\left(-\int_{t_i}^t \rho(u) du\right) \rho(t) \\ &= T(t_i) \exp\left(-(-\ln(\Phi_s(f(\mathbf{p}(t))\gamma(\mathbf{p}(t))))\right. \\ &\quad \left.+ \ln(\Phi_s(f(\mathbf{p}(t_i))\gamma(\mathbf{p}(t_i))))\right) \rho(t) \\ &= T(t_i) \frac{\Phi_s(f(\mathbf{p}(t))\gamma(\mathbf{p}(t))) - \frac{d}{dt}\Phi_s(f(\mathbf{p}(t))\gamma(\mathbf{p}(t)))}{\Phi_s(f(\mathbf{p}(t_i))\gamma(\mathbf{p}(t_i))) - \frac{d}{dt}\Phi_s(f(\mathbf{p}(t_i))\gamma(\mathbf{p}(t_i)))} \\ &= T(t_i) \frac{\phi_s(f(\mathbf{p}(t))\gamma(\mathbf{p}(t)))\gamma(\mathbf{p}(t))\frac{d}{dt}f(\mathbf{p}(t))}{\Phi_s(f(\mathbf{p}(t_i))\gamma(\mathbf{p}(t_i)))} \\ &= \frac{T(t_i)\gamma(\mathbf{p}(t))\frac{d}{dt}f(\mathbf{p}(t))}{\Phi_s(f(\mathbf{p}(t_i))\gamma(\mathbf{p}(t_i)))} \phi_s(f(\mathbf{p}(t))\gamma(\mathbf{p}(t))) \end{aligned} \quad (9)$$

Suppose that the local surface is tangentially approximated by a sufficiently small planar patch with its outward unit normal vector denoted as \mathbf{n} . Then we can write the signed distance function $f(\mathbf{p}(t))$ as

$$\begin{aligned} f(\mathbf{p}(t)) &= \frac{d}{dt}f(\mathbf{p}(t)) \cdot \mathbf{v} \cdot t + O(t^2) \\ &= \mathbf{n} \cdot \mathbf{v} \cdot t + O(t^2) \end{aligned} \quad (10)$$

Then the first-order approximation of $w(t)$ can be written as

$$w(t) = \frac{T(t_i)\gamma(\mathbf{p}(t))\mathbf{n} \cdot \mathbf{v}}{\Phi_s(f(\mathbf{p}(t_i))\gamma(\mathbf{p}(t_i)))} \phi_s(f(\mathbf{p}(t))\gamma(\mathbf{p}(t))) \quad (11)$$

Here $\frac{T(t_i)\gamma(\mathbf{p}(t))\mathbf{n} \cdot \mathbf{v}}{\Phi_s(f(\mathbf{p}(t_i))\gamma(\mathbf{p}(t_i)))}$ can be considered as a constant. Hence, $w(t)$ attains a local maximum when $f(\mathbf{p}(t)) = 0$ because $\phi_s(x)$ is a unimodal density function attaining the maximum value at $x = 0$.

Therefore, we can construct a first-order unbiased and occlusion-aware opacity density

$$\rho(t) = \frac{-\frac{d\Phi_s}{dt}(f(\mathbf{p}(t)) \cdot \gamma(\mathbf{p}(t)))}{\Phi_s(f(\mathbf{p}(t)) \cdot \gamma(\mathbf{p}(t)))}$$

2.2. Derivation of the Discrete Opacity Density for SDF α

For a set of sampled points along the ray $\{\mathbf{p}_i = \mathbf{o} + t_i \mathbf{v} | i = 1, \dots, n, t_i < t_{i+1}\}$, the rendered pixel color is

$$I_{pred}(\mathbf{o}, \mathbf{v}) = \sum_{i=1}^n \prod_{j=1}^{i-1} (1 - \alpha_j) \alpha_i c_i \quad (12)$$

where α_j is the discrete opacity value, which can be derived from Equation 13.

$$\begin{aligned} \alpha_i &= 1 - \exp\left(-\int_{t_i}^{t_{i+1}} \rho(t) dt\right) \\ &= 1 - \exp\left(-\int_{t_i}^{t_{i+1}} \frac{-\frac{d\Phi_s}{dt}(f(\mathbf{p}(t)) \cdot \gamma(\mathbf{p}(t)))}{\Phi_s(f(\mathbf{p}(t)) \cdot \gamma(\mathbf{p}(t)))} dt\right) \\ &= 1 - e^{(\ln(\Phi_s(f(\mathbf{p}(t_{i+1})) \cdot \gamma(\mathbf{p}(t_{i+1})))) - \ln(\Phi_s(f(\mathbf{p}(t_i)) \cdot \gamma(\mathbf{p}(t_i))))} \\ &= 1 - \frac{\Phi_s(f(\mathbf{p}(t_{i+1})) \cdot \gamma(\mathbf{p}(t_{i+1})))}{\Phi_s(f(\mathbf{p}(t_i)) \cdot \gamma(\mathbf{p}(t_i)))} \\ &= \frac{\Phi_s(f(\mathbf{p}(t_i)) \cdot \gamma(\mathbf{p}(t_i))) - \Phi_s(f(\mathbf{p}(t_{i+1})) \cdot \gamma(\mathbf{p}(t_{i+1})))}{\Phi_s(f(\mathbf{p}(t_i)) \cdot \gamma(\mathbf{p}(t_i)))} \end{aligned} \quad (13)$$

2.3. Unbiasness and Occlusion-awareness of w_β

Given the discrete opacity density α_i for the SDF representation and the validity probability $\mathcal{V}(\mathbf{p}(t_i))$. For simplicity, we denote $\mathcal{V}(\mathbf{p}(t_i))$ as \mathcal{V}_i .

We then calculate the discrete opacity density β for the ‘‘real’’ arbitrary surfaces. A physically reasonable solution is

$$\beta_i = \alpha_i \cdot \mathcal{V}_i,$$

i.e., if the 3D point $\mathbf{p}(t_i)$ is a valid point, $\mathcal{V}_i = 1$, the discrete opacity density for the ‘‘real’’ arbitrary surfaces remains unchanged

$$\beta_i = \alpha_i \cdot \mathcal{V}_i = \alpha_i.$$

While if the 3D point $\mathbf{p}(t_i)$ is an invalid point, $\mathcal{V}_i = 0$, the discrete opacity density for the ‘‘real’’ arbitrary surfaces is zero

$$\beta_i = \alpha_i \cdot \mathcal{V}_i = 0.$$

Then we prove that w_β is unbiased and occlusion-aware. In this proof, we denote $w_{\alpha,i} = \prod_{j=1}^{i-1} (1 - \alpha_j) \alpha_i$ and denote $w_{\beta,i} = \prod_{j=1}^{i-1} (1 - \beta_j) \beta_i$.

Unbiasness We first prove that w_β is unbiased. Given a camera ray

$$\{\mathbf{p}_i = \mathbf{o} + t_i \mathbf{v} | i = 1, \dots, n, t_i < t_{i+1}\}$$

and unbiased SDF weight

$$\{w_{\alpha,i} | i = 1, \dots, n\},$$

we prove that if w_{α,k^*} attains a local maximal weight at a valid surface intersection point $\mathbf{p}(t_{k^*})$, w_{β,k^*} also attains a local maximum.

We randomly sample a 3D point $\mathbf{p}(t_k)$ in the local valid region around $\mathbf{p}(t_{k^*})$, we only need to prove that $w_{\beta,k^*} > w_{\beta,k}$. We first prove that if $k > k^*$, $w_{\beta,k^*} > w_{\beta,k}$.

Since $\mathbf{p}(t_{k^*})$ and $\mathbf{p}(t_k)$ are distributed in one valid local region, for any point $\mathbf{p}(t_l)$ distributed between $\mathbf{p}(t_{k^*})$ and $\mathbf{p}(t_k)$ in the valid local region, i.e. $k^* \leq l \leq k$, we have $\mathcal{V}_i = 1$.

Since $w_{\alpha,k^*} > w_{\alpha,k}$, we have

$$\begin{aligned}
& w_{\alpha,k^*} > w_{\alpha,k} \\
\Rightarrow & \prod_{l=1}^{k^*-1} (1 - \alpha_l) \alpha_{k^*} > \prod_{l=1}^{k-1} (1 - \alpha_l) \alpha_k \\
\Rightarrow & \prod_{l=1}^{k^*-1} (1 - \alpha_l) \alpha_{k^*} > \prod_{l=1}^{k^*-1} (1 - \alpha_l) \prod_{l=k^*}^{k-1} (1 - \alpha_l) \alpha_k \\
\Rightarrow & \alpha_{k^*} > \prod_{l=k^*}^{k-1} (1 - \alpha_l) \alpha_k \\
\Rightarrow & \alpha_{k^*} \mathcal{V}_{k^*} > \prod_{l=k^*}^{k-1} (1 - \alpha_l \mathcal{V}_l) \alpha_k \mathcal{V}_k \\
\Rightarrow & \prod_{l=1}^{k^*-1} (1 - \alpha_l \mathcal{V}_l) \alpha_{k^*} \mathcal{V}_{k^*} \\
& > \prod_{l=1}^{k^*-1} (1 - \alpha_l \mathcal{V}_l) \prod_{l=k^*}^{k-1} (1 - \alpha_l \mathcal{V}_l) \alpha_k \mathcal{V}_k \\
\Rightarrow & \prod_{l=1}^{k^*-1} (1 - \alpha_l \mathcal{V}_l) \alpha_{k^*} \mathcal{V}_{k^*} > \prod_{l=1}^{k-1} (1 - \alpha_l \mathcal{V}_l) \alpha_k \mathcal{V}_k \\
\Rightarrow & w_{\beta,k^*} > w_{\beta,k}
\end{aligned} \tag{14}$$

If $k \leq k^*$, we can prove that $w_{\beta,k^*} > w_{\beta,k}$ in a similar way. Therefore, w_{β,k^*} also attains a local maximal value at a surface intersection point $\mathbf{p}(t_{k^*})$.

Occlusion-awareness Now we prove that w_β is occlusion-aware. Given a camera ray

$$\{\mathbf{p}_i = \mathbf{o} + t_i \mathbf{v} | i = 1, \dots, n, t_i < t_{i+1}\}$$

and occlusion-aware SDF weight

$$\{w_{\alpha,i} | i = 1, \dots, n\}.$$

For any $j, k \in \{1, 2, \dots, n\}$ with the same opacity value $\alpha_j = \alpha_k$, $w_{\alpha,j} > 0$, $w_{\alpha,k} > 0$, and $j < k$, we have

$$w_{\alpha,j} > w_{\alpha,k},$$

i.e., $\mathbf{p}(t_j)$ and $\mathbf{p}(t_k)$ have the same opacity value and the point $\mathbf{p}(t_j)$ closer to the view point has a larger contribution to the final output color than that of the other point $\mathbf{p}(t_k)$.

We then prove that $w_{\beta,j} > w_{\beta,k}$.

If either $\mathbf{p}(t_j)$ or $\mathbf{p}(t_k)$ is a invalid surface point, neither point need to satisfy the occlusion-aware rule. Therefore, we only need to prove that when both $\mathbf{p}(t_j)$ and $\mathbf{p}(t_k)$ are valid surface points, i.e. $\mathcal{V}_j = \mathcal{V}_k = 1$, $w_{\beta,j} > w_{\beta,k}$.

We suppose $w_{\beta,j} < w_{\beta,k}$, we have

$$\begin{aligned}
& w_{\beta,j} < w_{\beta,k} \\
\Rightarrow & \prod_{l=1}^{j-1} (1 - \alpha_l \mathcal{V}_l) \alpha_j \mathcal{V}_j < \prod_{l=1}^{k-1} (1 - \alpha_l \mathcal{V}_l) \alpha_k \mathcal{V}_k \\
\Rightarrow & \prod_{l=1}^{j-1} (1 - \alpha_l \mathcal{V}_l) \alpha_j < \prod_{l=1}^{k-1} (1 - \alpha_l \mathcal{V}_l) \alpha_k \\
\Rightarrow & \prod_{l=1}^{j-1} (1 - \alpha_l \mathcal{V}_l) < \prod_{l=1}^{k-1} (1 - \alpha_l \mathcal{V}_l) \\
\Rightarrow & \prod_{l=1}^{j-1} (1 - \alpha_l \mathcal{V}_l) < \prod_{l=1}^{j-1} (1 - \alpha_l \mathcal{V}_l) \prod_{l=j}^{k-1} (1 - \alpha_l \mathcal{V}_l) \\
\Rightarrow & 1 < \prod_{l=j}^{k-1} (1 - \alpha_l \mathcal{V}_l)
\end{aligned} \tag{15}$$

Since $0 \leq \alpha_l \leq 1$, Equation 15 is not true. Therefore, when both $\mathbf{p}(t_j)$ and $\mathbf{p}(t_k)$ are valid surface points, $w_{\beta,j} > w_{\beta,k}$. Q.E.D.

3. Implementation of NeAT-Net

We implement NeAT-Net as follows.

SDF-Net As shown in Figure 1 (left), we leverage the implementation of the SDF network in DeepSDF [5], which consists of 8 layers with hidden layers of width 512, and a single skip connection from the input to the middle layer. We initialize the parameters of the MLP with geometric initialization [2]. For the single-view reconstruction experiment, we use the concatenation of the embedded position and the image feature as the input.

Validity-Net As shown in Figure 1 (right), the MLP consists of 8 layers with Xavier initialization. We use the *ReLU* activation between hidden layers and *Sigmoid* for the output. For the single-view reconstruction experiment, we use an additional MLP to encode the image features as shown in the units marked with the dotted lines in Figure 1 (right). In each layer, we use the addition of the outputs from the two MLPs as the input to the activation unit.

Color-Net As shown in Figure 2, we use the implementation of the renderer MLP in IDR [8], which consists of 4 layers, with hidden layers of width 512. We apply positional encoding [4] to improve the learning of high-frequency details. For the single-view reconstruction experiment, we use an additional MLP to encode the image features as shown in the units marked with the dotted lines in Figure 2. In each layer, we use the addition of the outputs from the two MLPs as the input to the activation unit.

4. Additional Results

We show additional qualitative results on the Deep Fashion 3D Dataset [3] and Multi-Garment Net Dataset [1]. Figure 3 shows additional visual comparisons with baseline methods on the Deep Fashion 3D Dataset [3]. Figure 4

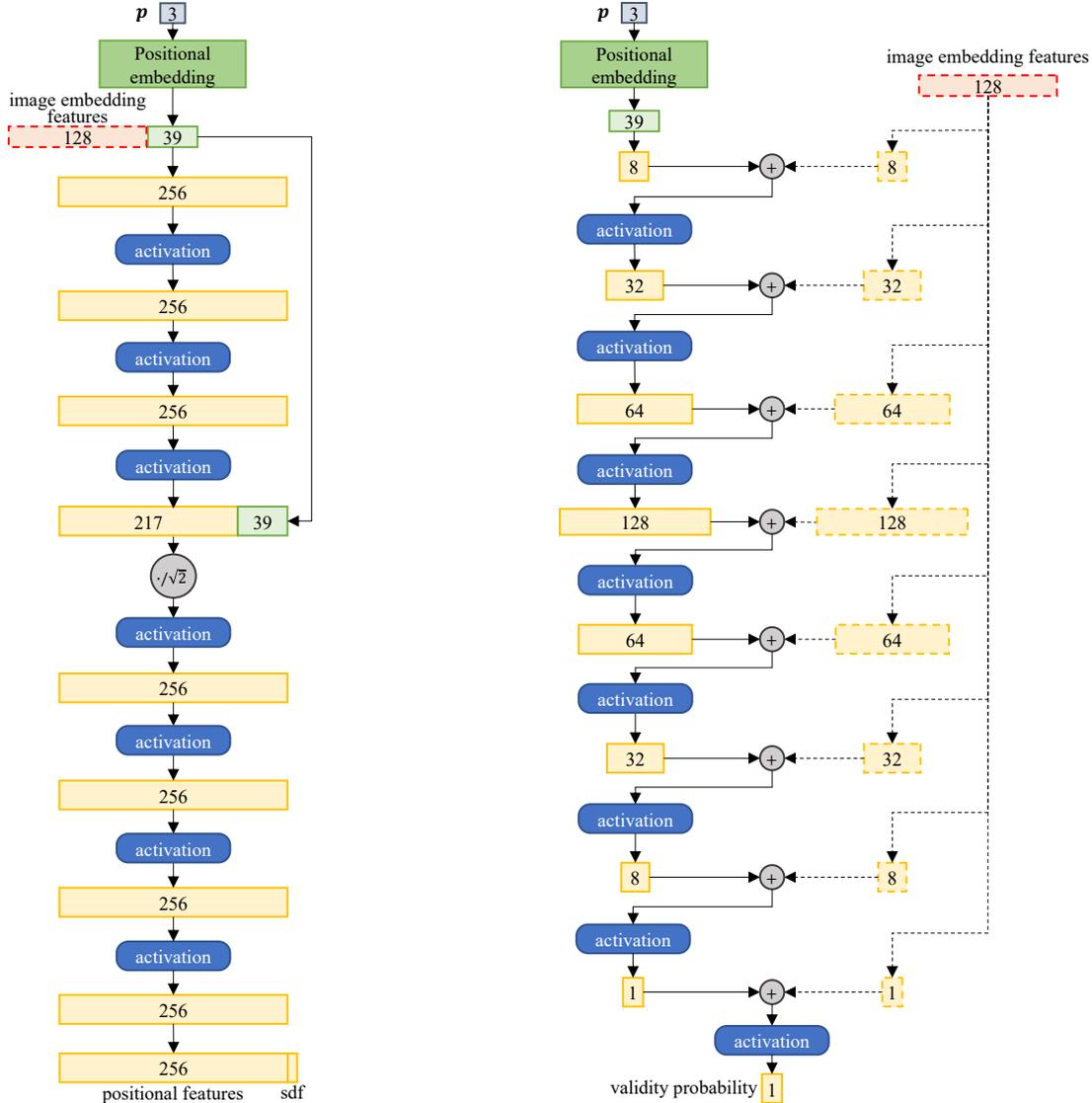


Figure 1. Left: the network structure of **SDF-Net**. Right: the network structure of **Validity-Net**. The units marked with solid lines are used in the multi- and single-view reconstruction experiment. The units marked with dotted lines are only used in the latter.

shows additional visual comparisons with baseline methods on the Multi-Garment Net Dataset [1]. Our method is able to generate high-fidelity results.

5. Additional Evaluation Metrics

We compare our approach with the state-of-the-art volume rendering based methods – NeuS [6], HFS [7], and surface rendering based method – IDR [8]. We report the F-score (FS) on five examples for each category from Deep Fashion 3D Dataset [3] and two examples for each category from Multi-Garment Net Dataset [1] in Table 1.

References

- [1] Bharat Lal Bhatnagar, Garvita Tiwari, Christian Theobalt, and Gerard Pons-Moll. Multi-garment net: Learning to dress 3d people from images. In *IEEE International Conference on Computer Vision (ICCV)*. IEEE, oct 2019. 3, 4, 5, 7
- [2] Amos Gropp, Lior Yariv, Niv Haim, Matan Atzmon, and Yaron Lipman. Implicit geometric regularization for learning shapes. In *Proceedings of Machine Learning and Systems 2020*, pages 3569–3579. 2020. 3
- [3] Zhu Heming, Cao Yu, Jin Hang, Chen Weikai, Du Dong, Wang Zhangye, Cui Shuguang, and Han Xiaoguang. Deep fashion3d: A dataset and benchmark for 3d garment reconstruction from single images. In *Computer Vision – ECCV 2020*, pages 512–530. Springer International Publishing, 2020. 3, 4, 5, 6
- [4] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik,

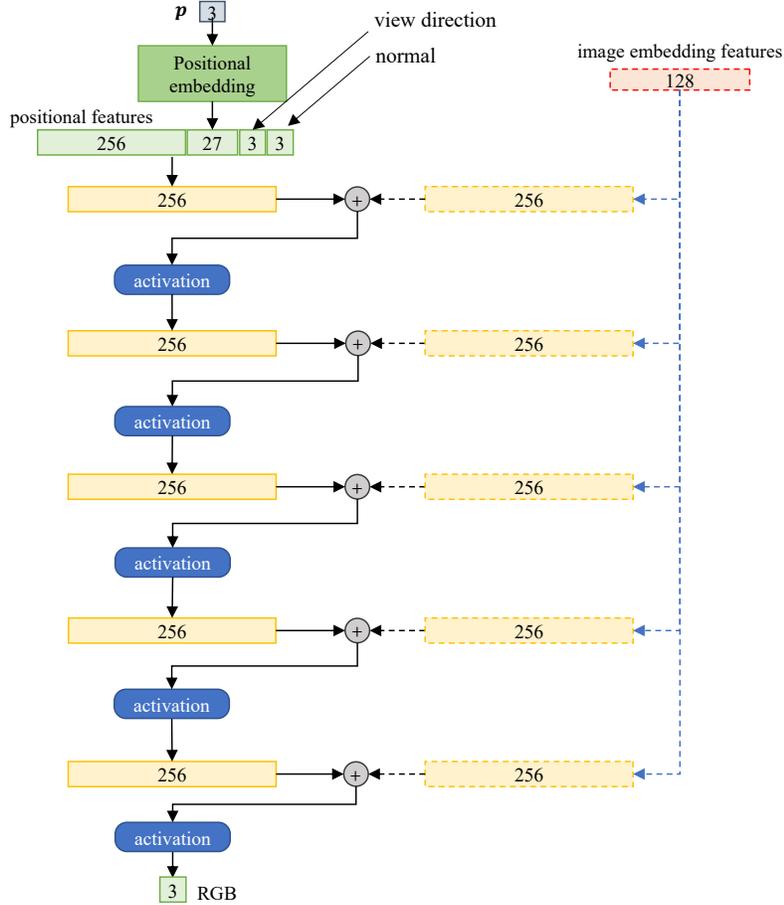


Figure 2. The network structure of **Color-Net**. The units marked with solid lines are used in the multi- and single-view reconstruction experiment. The units marked with dotted lines are only used in the latter.

| | FS \uparrow | Ours | NeuS [6] | IDR [8] | HFS [7] |
|-----|--------------------|--------------|----------|---------|---------|
| D3D | long sleeve upper | 0.563 | 0.140 | 0.169 | 0.200 |
| | short sleeve upper | 0.575 | 0.214 | 0.333 | 0.257 |
| | no sleeve upper | 0.693 | 0.333 | 0.228 | 0.176 |
| | long sleeve dress | 0.630 | 0.237 | 0.199 | 0.276 |
| | short sleeve dress | 0.647 | 0.170 | 0.190 | 0.230 |
| | no sleeve dress | 0.693 | 0.235 | 0.276 | 0.233 |
| | pants | 0.530 | 0.083 | 0.110 | 0.104 |
| | dress | 0.680 | 0.277 | 0.253 | 0.163 |
| | average | 0.626 | 0.211 | 0.220 | 0.205 |
| MGN | LongCoat | 0.145 | 0.110 | 0.110 | 0.074 |
| | TShirtNoCoat | 0.077 | 0.060 | 0.048 | 0.031 |
| | ShirtNoCoat | 0.226 | 0.157 | 0.112 | 0.139 |
| | ShortPants | 0.016 | 0.011 | 0.009 | 0.003 |
| | Pants | 0.041 | 0.052 | 0.037 | 0.024 |
| | average | 0.101 | 0.078 | 0.063 | 0.054 |

Table 1. Quantitative evaluation on *Deep Fashion 3D Dataset* (D3D) [3] with F-score (FS) averaged over five examples per category, and *Multi-Garment Net Dataset* (MGN) [1] with chamfer distance averaged on two examples per category.

Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. In *European conference on computer vision*, pages

405–421. Springer, 2020. 3

[5] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous



Figure 3. Comparisons on open surface reconstruction of the Deep Fashion 3D Dataset [3]. NeAT is able to reconstruct high-fidelity open surfaces while NeuS [6], HF-NeuS [7] and IDR [8] fail to recover the correct topologies.

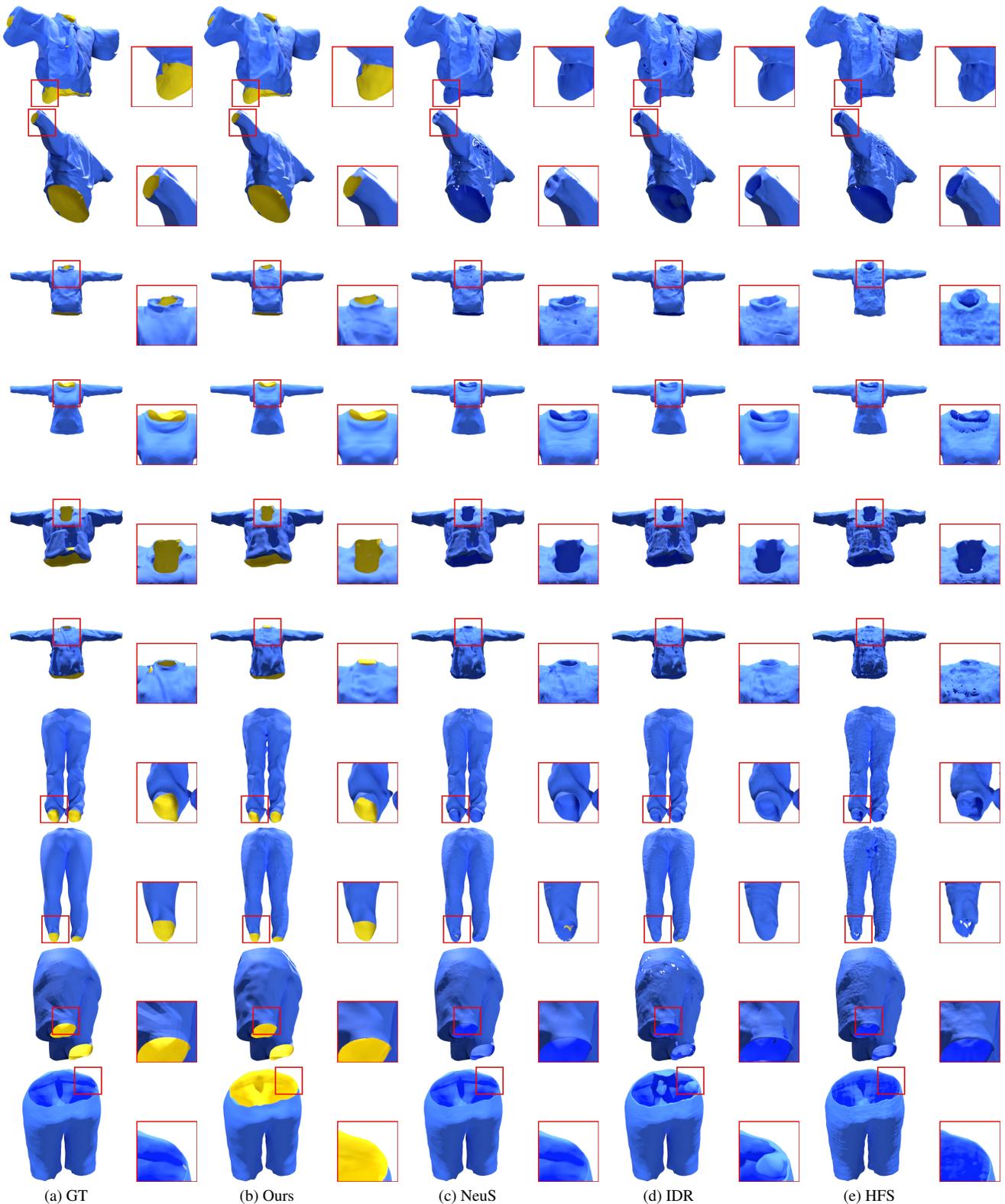


Figure 4. Comparisons on open surface reconstruction of the Multi-Garment Net Dataset [1]. NeAT is able to reconstruct high-fidelity open surfaces while NeuS [6], HF-NeuS [7] and IDR [8] fail to recover the correct topologies.

signed distance functions for shape representation. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019. 3

- [6] Peng Wang, Lingjie Liu, Yuan Liu, Christian Theobalt, Taku Komura, and Wenping Wang. Neus: Learning neural implicit surfaces by volume rendering for multi-view reconstruction. *NeurIPS*, 2021. 4, 5, 6, 7
- [7] Yiqun Wang, Ivan Skorokhodov, and Peter Wonka. Hf-neus: Improved surface reconstruction using high-frequency details. *arXiv preprint arXiv:2206.07850*, 2022. 4, 5, 6, 7
- [8] Lior Yariv, Yoni Kasten, Dror Moran, Meirav Galun, Matan Atzmon, Basri Ronen, and Yaron Lipman. Multiview neural surface reconstruction by disentangling geometry and appearance. *Advances in Neural Information Processing Systems*, 33:2492–2502, 2020. 3, 4, 5, 6, 7