

SFDM: Robust Decomposition of Geometry and Reflectance for Realistic Face Rendering from Sparse-view Images

– Supplementary Material –

1. Overview

This supplementary material provides additional results and analysis that support the methodology of our main text. We provide more details on the implementation of our method in Sec. 2. To validate the effectiveness and robustness of our proposed method, we give additional results for subjects with divergent characteristics and experiments conducted on *H3DS* [15] in Sec. 3. In addition, we analyze the experimental results in detail in Sec. 4, including quantitative and qualitative metrics comparisons with other work. In Sec. 5, we discuss the relationship between our task and single-view reconstructions, highlighting the advantages and limitations of each. Sec. 6 contemplates the potential negative impact of our work and our responsibility to human subjects. Extensive visualization results can be found in Sec. 7.

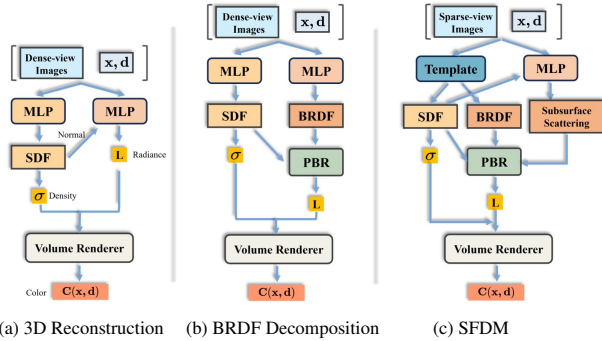


Figure 1. **Overview of different frameworks.** (a) Framework like VolSDF [20] accomplishes 3D reconstruction from dense images but does not decompose radiance-related factors. (b) Framework like NerFactor [23] introduces physically based rendering (PBR) to simulate real-world lighting. (c) We propose a novel geometry and reflectance decomposition framework for 3D face reconstruction from sparse views.

2. Implementation details

Network Architecture. In Stage 1, we aim to learn a general facial template with both geometry and reflectance attributes. For the geometry template, we slightly modify the

identity mini-nets and template mini-nets of ImFace [24]. Specifically, we introduce two additional MLPs as feature extraction branches for each part, together with the mini-nets, forming our deformation net F_{dfm} and template net T_{geo} , respectively. Regarding the reflectance template T_{rf} , we utilize an 8-layer MLP with 256 dimensions in each layer, and the BRDF offset module F_{ofs} employs the same architecture. The PBR module uses an MLP consisting of four layers, each with 128 dimensions, to compute the BRDF look-up-textures B_0 and B_1 . For the albedo gradient predictor \mathcal{G} , we also employ a four-layer MLP with 128 dimensions.

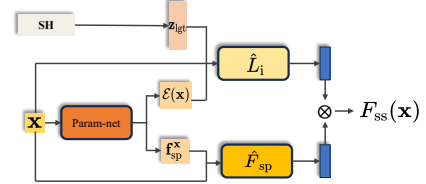


Figure 2. The network architecture of subsurface scattering offset.

In Stage 2, we introduce two additional modules: the displacement net F_{dis} and the subsurface scattering offset net F_{ss} . The displacement net is composed of a 4-layer MLP, with each layer having 256 dimensions. To capture high-frequency geometric details, we incorporate a positional encoding of 8 frequencies. The architecture of the subsurface scattering offset module is illustrated in Fig. 2. At first, we use a 4-layer MLP (Param-net) to obtain the scattering parameters $\mathbf{f}_{\text{sp}}^{\mathbf{x}}$ and a vector $\mathcal{E}(\mathbf{x})$ indicating the light integration region. Subsequently, we utilize two branches to integrate light and scattering reflectance. In the light integration branch, we initially compress the Spherical Harmonics (SH) weights into an ambient light embedding \mathbf{z}_{igt} by using a linear layer. Afterwards, we utilize a light integration network \hat{L}_i to obtain the light intensity of a small region on the surface as $\hat{L}_i(\mathbf{x}, \mathbf{z}_{\text{igt}}, \mathcal{E}(\mathbf{x}))$, where $\mathbf{x} \in \mathbb{R}^3$ are the coordinates. For scattering reflectance, the scattering reflectance integration network \hat{F}_{sp} integrates the scattering reflectance based on the scattering parameters $\mathbf{f}_{\text{sp}}^{\mathbf{x}}$ and coordinates \mathbf{x} , re-

sulting in the output of integrated reflectance. Both \hat{L}_i and \hat{F}_{sp} comprise a 4-layer MLP. We use sine activation to allow the model to more effectively handle Gaussian curves in scattering profiles. Finally, the 3-dimensional integrated light and scattering reflectance are multiplied to produce the offset term for diffuse components.

Volume rendering. Following most previous work [10], we calculate the color by integrating the radiance L_o and density $\sigma \in \mathbb{R}$ along the ray \mathbf{r} as:

$$\mathbf{C}(\mathbf{r}) = \int_{t'}^{t''} L_o(\mathbf{x}(t), \boldsymbol{\omega}_o) \sigma(t) T(t) dt, \quad (1)$$

where $\mathbf{x}(t)$ represents the coordinates of a point in the ray, with t' and t'' as the starting and ending points along the ray, respectively. $T(t) = \exp(-\int_{t_n}^{t'} \sigma(t) dt)$, which can be considered as the radiance attenuation rate caused by hitting particles along the ray.

Data preprocessing. We mainly use the *Facescape* [19] dataset, which contains 359 subjects in 20 different expressions. For each subject, *Facescape* has over 50 multi-view images, along with corresponding camera parameters and scanned meshes. We select 3, 5, and 10 frontal views for training, and around 10 views for testing. Before training the model, we first preprocess the images and camera parameters. Adopting the preprocessing approach of NeuFace [25] and ImFace [24], we crop the facial region from the *Facescape* head mesh and align the cropped mesh within the $[-1, 1]$ range. This cropped mesh is also used as ground truth for evaluating geometry accuracy. For camera parameters, we adjust the original *Facescape* camera settings to correspond with the mesh processing operations. Regarding the images, using the updated camera parameters, we can render the corresponding face masks, which then allows us to extract the facial regions from the original images.

Training details. During training, we use 10 subjects, each with 10 views images from *Facescape* to train the facial template in Stage 1. We tested 3, 6, 10, and 15 subjects for template training, finding that performance saturates between 6 and 10 subjects. We employ the pre-trained ImFace model to initialize the geometry template. To promote stable template learning, we apply a weight of 0.02 to the learning rate for template optimization, which effectively avoids severe variation. The total loss function for Stage 1 is:

$$\mathcal{L}_{st1} = \lambda_1 \mathcal{L}_{col} + \lambda_2 \mathcal{L}_{eik} + \lambda_3 \mathcal{L}_{light} + \lambda_4 \mathcal{L}_{spec} + \lambda_5 \mathcal{L}_{ofs} + \lambda_6 \mathcal{L}_{code} + \lambda_7 \mathcal{L}_G, \quad (2)$$

where $\lambda_1, \dots, \lambda_7$ are 1, $1e^{-1}$, $5e^{-3}$, $1.5e^{-2}$, $1e^{-3}$, 500, and $5e^{-3}$, respectively.

In Stage 2, we select subjects that are not used during Stage 1. We mainly select one frontal image and two profile images in the 3-view setting, so as to capture the most

comprehensive facial information of an individual. To adjust the contribution of each BRDF parameter from the reflectance template, we employ a set of learnable parameters $\mathbf{W}(\mathbf{x})$. Specifically, we initialize the weight of each albedo channel in \mathbf{w}_a to 0.4, while setting others to 0.9 due to the substantial variation in albedo. During the second training stage, we apply a weight of 0.02 to the learning rate for template optimization. This approach ensures that the majority of template information is retained, while still allowing for slight modifications to enhance the suitability of template parameters in Stage 2. The loss function for Stage 2 is based on Stage 1 with two additional items:

$$\mathcal{L}_{st2} = \mathcal{L}_{st1} + \lambda_8 \mathcal{L}_{dis} + \lambda_9 \mathcal{L}_{ss}, \quad (3)$$

where $\lambda_8 = 1e^{-3}$ and $\lambda_9 = 2e^{-3}$. Since the albedo gradient predictor \mathcal{G} is not trained during Stage 2, we set λ_7 to 0.

We conducted our training procedure on Tesla V100 GPUs. In Stage 1, we utilized 4 GPUs for approximately 20 hours, using a mini-batch size of 2048 rays. For Stage 2, we employed 2 GPUs with 3, 5, and 10 views, with each training session lasting around 2, 2.5, and 5 hours respectively. In the future, we plan to explore acceleration techniques such as hash encoding [11] and Plenotree [21] to enhance the efficiency of the decomposition and reconstruction processes.

Testing details. During testing, we apply calibration on all methods to mitigate the exposure differences as NeuFace [25]. This process helps eliminate the overall color disparities between the rendered images and the ground truth images caused by unknown illumination or exposure conditions. For example, in Fig. 10, the diffuse prediction of NeuFace appears yellowish, potentially due to a false assumption about the light being yellow. To minimize the impact of such erroneous assumptions, the calibration aligns the ground truth image by employing a 3×3 matrix on the radiance values.

In scenarios involving relighting, we convert the environment maps to SH weights following [4]. For specular editing, we adjust the specular intensity parameter in b_s to control the shininess of faces.

In the reflectance decomposition analysis (Sec. 5.4 of the main text), we utilize RefMM [6] to fit and generate pseudo-reflectance ground truth for all test images of five subjects. Subsequently, we calculate the SSIM values for the diffuse and specular results of each subject under the 3, 5, and 10 view settings.

3. Generalization on out-of-domain data

Different expressions. During template training, we use randomly selected subjects with the same “grin” expression to ensure that the facial template consists of the majority of

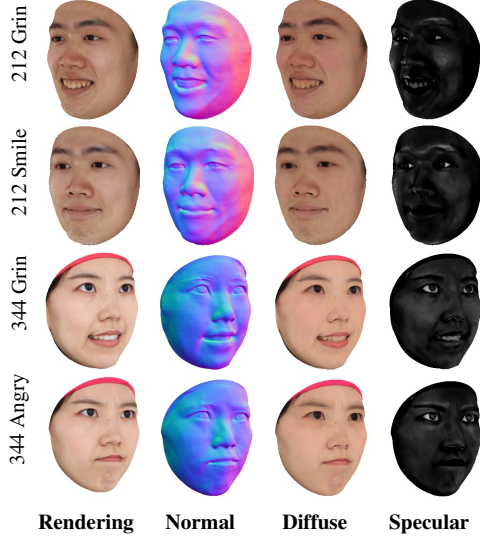


Figure 3. Results on different expressions.

facial features. As shown in Fig. 3, our method exhibits robustness in decomposing subjects with various expressions, since the “grin” face contains major elements on human faces (e.g. teeth and eyes).

Divergent characteristics. Fig. 4 shows the results of subjects with different skin colors and ages, which have large variations from our facial template. Thanks to our adaptable facial template, enriched with comprehensive human facial priors, SFDM adeptly navigates these challenging scenarios to achieve realistic face reconstructions.

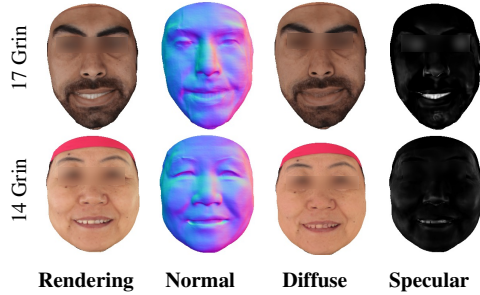


Figure 4. Results of subjects with large variations of skin colors and ages from the template.



Figure 5. **Albedo gradient visualization.** The albedo gradient can partially outline the sketches of human faces and provide additional guidance for albedo prediction.

H3DS. To further verify the generalization performance of

our method on out-of-domain data, we conduct additional experiments on the *H3DS* dataset [15]. We directly use the facial template trained on *Facescape* for Stage 1, which allows us to better assess the performance of the facial template on unseen data. Moreover, we carry out an ablation study on the components of Stage 1 to further validate the contribution of our template to the robustness of decomposition. As shown in Fig. 6, in the absence of the template and albedo gradient predictor, the model exhibits various degrees of decomposition errors in sparse view scenarios, such as mistakes in the nose area. In contrast, with a complete facial template, despite facing challenges—without training a new template on new data, imprecise facial segmentation in *H3DS*, and significant lighting differences across views (some with flash)—our method is still able to decompose the face. We also visualize the albedo gradient in Fig. 5, which resembles a sketch of the human face. The albedo gradient provides a more generalized representation than the absolute values of albedo, enhancing the model’s ability to generalize in albedo prediction. To further enhance the robustness of decomposition results for subjects with significant domain gaps, a more effective approach may involve training the facial template on a more extensive and diverse set of subjects.

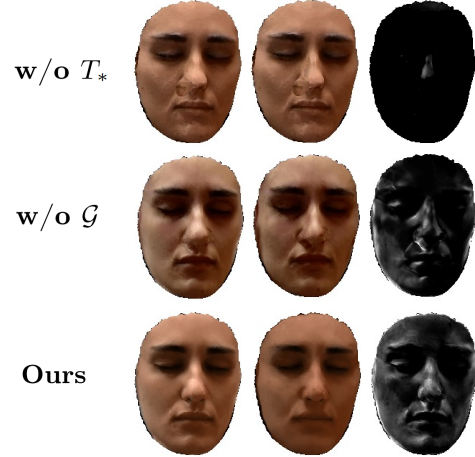


Figure 6. **Results on H3DS.** T_* presents the facial template and \mathcal{G} is the albedo gradient predictor. For each method showcased, the set of three images corresponds to rendering, diffuse, and specular, respectively.

4. Performance analysis

Quantitative results. We present the quantitative comparison results of all methods across different views in Tab.1 of the main text. Specifically, the evaluation of image synthesis quality is conducted using PSNR, SSIM, and LPIPS [22] metrics. PSNR and SSIM only consider pixel-wise color differences between predictions and ground truth, which may not adequately reflect the synthesis reality. To eval-

uate the reality of rendered images, LPIPS uses a neural network to quantify perceptual similarity. In the 3-view setting, our method has a lower PSNR than DeformHead [18]. However, it excels over DeformHead in terms of LPIPS as it allows for the capture of more intricate appearance details, such as facial highlights and spots. Although the successful reconstruction of these details has little contribution to overall pixel-wise error metrics such as PSNR, it can enhance the realism of the rendered images.

We use the Chamfer distance (CD) to evaluate the accuracy of geometry reconstruction. Compared to our method, geometry-oriented methods are more sensitive to geometry deformations, allowing them to capture intricate details such as teeth. However, the high sensitivity leads to false high-frequency geometry predictions, especially under sparse views. In Fig.4 of the main text and the figures in Sec. 7, the facial skin appears smooth, but there are noticeable bumps in the predictions of VolSDF and DeformHead. Hence, our method has better performance in the CD metric under sparse view settings. In a low-view setting of 10 views, our method tackles a more complex decomposition task but still exhibits a comparable performance (marginally lower) to pure 3D reconstruction methods.

In Tab.3 of the main text, we present a quantitative comparison of decomposition quality between our method and NeuFace. Leveraging a general reflectance facial template, **SFDM** achieves robust decomposition of reflectance terms, leading to significantly superior diffuse and specular reflectance results compared to NeuFace. Specifically, NeuFace often contributes all reflectance to the diffuse term since the optimization simply relies on fitting facial images without reflectance supervision. Consequently, its specular decomposition performance deteriorates notably under sparse-view settings.

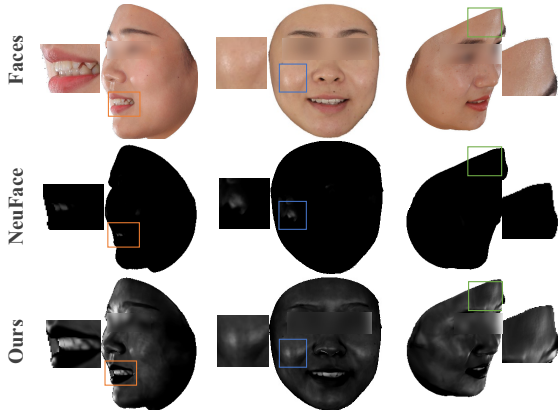


Figure 7. Comparison with NeuFace [25] in terms of specular reflectance decomposition.

Qualitative results. Our method outperforms other BRDF decomposition methods in terms of all metrics, demon-

strating superior rendering and geometry reconstruction effects. Besides, as shown in Fig. 7 and Sec. 7, our method can robustly decompose specular terms across various subjects and settings, benefiting from the comprehensive prior knowledge of the learned facial template. Considering that there is no ground truth for specular reflectance, we introduce a straightforward metric called specular failure rate, representing the percentage of entirely empty specular output. This metric facilitates a quantitative comparison between our method and NeuFace. According to the statistics, our method achieves a specular failure rate of 0%, whereas NeuFace’s specular failure rates are 30%, 25%, and 20% under 3 views, 5 views, and 10 views, respectively.

Furthermore, we observe that the TensoIR [7] exhibits noticeable lining effects, prompting us to investigate. We find these effects are caused by sparse view conditions. When we utilize all training views, these phenomena disappear.

5. Discussion with single-view methods

Our method focuses on the reconstruction from sparse views, and thus we do not discuss single-view methods in detail in the main text. However, this section will discuss our relationship with single-view approaches. Although sparse views offer limited perspectives, they can still provide rich 3D information when combined with camera angles. For instance, with at least two-view images, the 3D coordinates of a point corresponding in real space across both views can be calculated using its 2D coordinates and camera parameters. Conversely, single-view images contain only 2D information and lack 3D details, often relying on 3D estimation methods for reconstruction. Due to the limited information in a single 2D image, single-view approaches typically require extensive training data (1000+ images) across a wide variety of identities to obtain the ability to perform 3D estimation effectively. By contrast, multi-view methods can extract vast quantities of 3D information from images of a single identity, making them more efficient for tasks where multiple views are available.

One type of single-view approach involves GAN-based methods, such as EG3D [2] and StyleSDF [12], which trains models to generate 3D meshes and novel view images with extensive training data. However, this method faces challenges with data volume requirements and consistency issues in generated images (e.g., eyes fixating on the camera from all views as 8). Similarly, Relightify [13] employs another generative method, the diffusion model, which also requires a large number of training images.

Another approach is based on parametric face models, predicting facial parameters from a single-view image to reconstruct faces. For example, AlbedoMM [16] uses 3DMM [1], TRUST [5] is based on FLAME [9], and RefMM [6] utilizes BFM09 [14]. However, due to the

Method	Geometry	Reflectance	# views	Fine Details	Requiring Albedo GT
NeRF [10]	N	N	20+	Y	N
VolSDF [20]	Y	N	20+	Y	N
DeformHead [18]	Y	N	10	Y	N
TensoIR [7]	Y	Y	20+	Y	N
NeuFace [25]	Y	Y	20+	Y	N
AlbedoMM [16]	Y	Y	1	N	Y
TRUST [5]	Y	Y	1	N	Y
Relightify [3]	Y	Y	1	N	Y
Fitme [8]	Y	Y	1	N	Y
RefMM [6]	Y	Y	1	N	N
Ours	Y	Y	3	Y	N

Table 1. **Qualitative comparison with other methods.** We assess the capabilities of various methods in comparison to our own. Our approach can achieve detailed decomposition and reconstruction from three-view images, without the need for intricate lighting and ground truth albedo prerequisites.



Figure 8. **Results of StyleSDF [12].** The kid looks at the camera, regardless of the rendering view.

difficulty of face decomposition tasks, these methods often require higher quality training data, such as various environmental lighting settings (3D-RFE dataset [17] in AlbedoMM) and pseudo ground truth reflectance (utilizing texture maps in TRUST and diffuse maps in the 3D-RFE dataset). Additionally, the use of 3D morphable face models, while efficient in compressing facial expressions, tends to lose many details in face reconstruction. As shown in Fig. 9, our experiments with RefMM’s reconstruction results, although capable of generating geometry and reflectance from a single view, significantly lack the facial details and realism achieved by our SFDM.

6. Ethnic statement

Our method is primarily used for facial reconstruction, thus necessitating a focus on protecting individuals’ privacy. We utilized the *Facescape* dataset, of which only a subset of IDs are publishable. Consequently, we applied mosaic processing (approved by the publishers of *Facescape*) to all portraits not on the publishable list, similar to what is done in Sec. 7, to safeguard personal privacy. We encourage future SFDM users to conduct facial reconstruction and decomposition using facial images for which they have obtained authorization, thereby respecting people’s portrait rights.

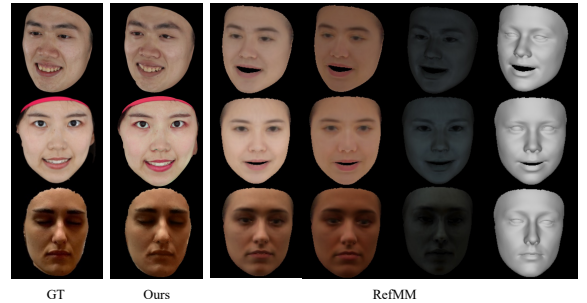


Figure 9. **Comparison with RefMM.** The first column presents the ground truth images, while the second column showcases our reconstruction results. Regarding RefMM’s outcomes, from left to right, the sequence includes the reconstruction result, diffuse shading, specular shading, and geometry.

7. Additional results

Here we provide more results to illustrate the effectiveness of our method. Fig. 10 - 14, Fig. 15 - 19, Fig. 20 - 22 are results under 3-view settings, 5-view settings, and 10-view (low-view) settings, respectively. Specifically, the rendering results are calibrated as Sec. 2, correcting the overall color deviations, such as the results of NeuFace [25] in Fig. 10. In the *Facescape* dataset, only four subjects (122, 212, 340, 344) are cleared for publication. For the remaining subjects, we have applied the mosaic technique to ensure their anonymity.

References

- [1] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *Seminal Graphics Papers: Pushing the Boundaries, Volume 2*, pages 157–164, 2023. 4
- [2] Eric R Chan, Connor Z Lin, Matthew A Chan, Koki Nagano, Boxiao Pan, Shalini De Mello, Orazio Gallo, Leonidas J Guibas, Jonathan Tremblay, Sameh Khamis, et al. Efficient

- geometry-aware 3d generative adversarial networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16123–16133, 2022. 4
- [3] Zhaoxi Chen and Ziwei Liu. Relighting4d: Neural relightable human from videos. In *European Conference on Computer Vision*, pages 606–623. Springer, 2022. 5
- [4] Zhang Chen, Anpei Chen, Guli Zhang, Chengyuan Wang, Yu Ji, Kiriakos N Kutulakos, and Jingyi Yu. A neural rendering framework for free-viewpoint relighting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5599–5610, 2020. 2
- [5] Haiwen Feng, Timo Bolkart, Joachim Tesch, Michael J Black, and Victoria Abrevaya. Towards racially unbiased skin tone estimation via scene disambiguation. In *European Conference on Computer Vision*, pages 72–90. Springer, 2022. 4, 5
- [6] Yuxuan Han, Zhibo Wang, and Feng Xu. Learning a 3d morphable face reflectance model from low-cost data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8598–8608, 2023. 2, 4, 5
- [7] Haian Jin, Isabella Liu, Peijia Xu, Xiaoshuai Zhang, Songfang Han, Sai Bi, Xiaowei Zhou, Zexiang Xu, and Hao Su. Tensor: Tensorial inverse rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 165–174, 2023. 4, 5
- [8] Alexandros Lattas, Stylianos Moschoglou, Stylianos Ploumpis, Baris Gecer, Jiankang Deng, and Stefanos Zafeiriou. Fitme: Deep photorealistic 3d morphable model avatars. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8629–8640, 2023. 5
- [9] Tianye Li, Timo Bolkart, Michael J Black, Hao Li, and Javier Romero. Learning a model of facial shape and expression from 4d scans. *ACM Trans. Graph.*, 36(6):194–1, 2017. 4
- [10] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021. 2, 5
- [11] Thomas Müller, Alex Evans, Christoph Schied, and Alexander Keller. Instant neural graphics primitives with a multi-resolution hash encoding. *ACM Transactions on Graphics (ToG)*, 41(4):1–15, 2022. 2
- [12] Roy Or-El, Xuan Luo, Mengyi Shan, Eli Shechtman, Jeong Joon Park, and Ira Kemelmacher-Shlizerman. Stylesdf: High-resolution 3d-consistent image and geometry generation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 13503–13513, 2022. 4, 5
- [13] Foivos Paraperas Papantoniou, Alexandros Lattas, Stylianos Moschoglou, and Stefanos Zafeiriou. Relightify: Relightable 3d faces from a single image via diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 8806–8817, 2023. 4
- [14] Pascal Paysan, Reinhard Knothe, Brian Amberg, Sami Romdhani, and Thomas Vetter. A 3d face model for pose and illumination invariant face recognition. In *2009 sixth IEEE international conference on advanced video and signal based surveillance*, pages 296–301. Ieee, 2009. 4
- [15] Eduard Ramon, Gil Triginer, Janna Escur, Albert Pumarola, Jaime Garcia, Xavier Giro-i Nieto, and Francesc Moreno-Noguer. H3d-net: Few-shot high-fidelity 3d head reconstruction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5620–5629, 2021. 1, 3
- [16] William AP Smith, Alassane Seck, Hannah Dee, Bernard Tiddeman, Joshua B Tenenbaum, and Bernhard Egger. A morphable face albedo model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5011–5020, 2020. 4, 5
- [17] Giota Stratou, Abhijeet Ghosh, Paul Debevec, and Louis-Philippe Morency. Effect of illumination on automatic expression recognition: a novel 3d relightable facial database. In *2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG)*, pages 611–618. IEEE, 2011. 5
- [18] Baixin Xu, Jiarui Zhang, Kwan-Yee Lin, Chen Qian, and Ying He. Deformable model-driven neural rendering for high-fidelity 3d reconstruction of human heads under low-view settings. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 17924–17934, 2023. 4, 5
- [19] Haotian Yang, Hao Zhu, Yanru Wang, Mingkai Huang, Qiu Shen, Ruigang Yang, and Xun Cao. Facescape: a large-scale high quality 3d face dataset and detailed riggable 3d face prediction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 601–610, 2020. 2
- [20] Lior Yariv, Jiatao Gu, Yoni Kasten, and Yaron Lipman. Volume rendering of neural implicit surfaces. *Advances in Neural Information Processing Systems*, 34:4805–4815, 2021. 1, 5
- [21] Alex Yu, Ruilong Li, Matthew Tancik, Hao Li, Ren Ng, and Angjoo Kanazawa. Plenotrees for real-time rendering of neural radiance fields. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5752–5761, 2021. 2
- [22] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. 3
- [23] Xiuming Zhang, Pratul P Srinivasan, Boyang Deng, Paul Debevec, William T Freeman, and Jonathan T Barron. Nerfactor: Neural factorization of shape and reflectance under an unknown illumination. *ACM Transactions on Graphics (ToG)*, 40(6):1–18, 2021. 1
- [24] Mingwu Zheng, Hongyu Yang, Di Huang, and Liming Chen. Imface: A nonlinear 3d morphable face model with implicit neural representations. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20343–20352, 2022. 1, 2
- [25] Mingwu Zheng, Haiyu Zhang, Hongyu Yang, and Di Huang. Neuface: Realistic 3d neural face rendering from multi-view images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16868–16877, 2023. 2, 4, 5

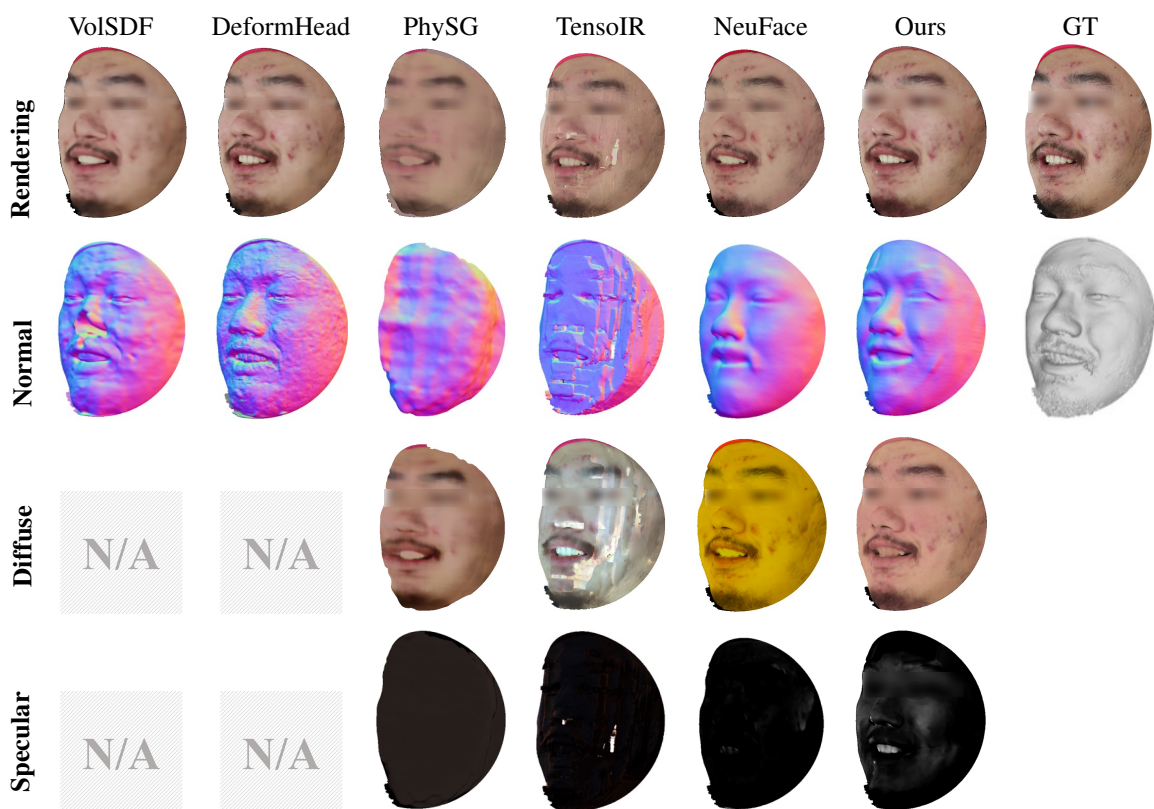


Figure 10. Comparison under a 3-view setting for Subject 12.

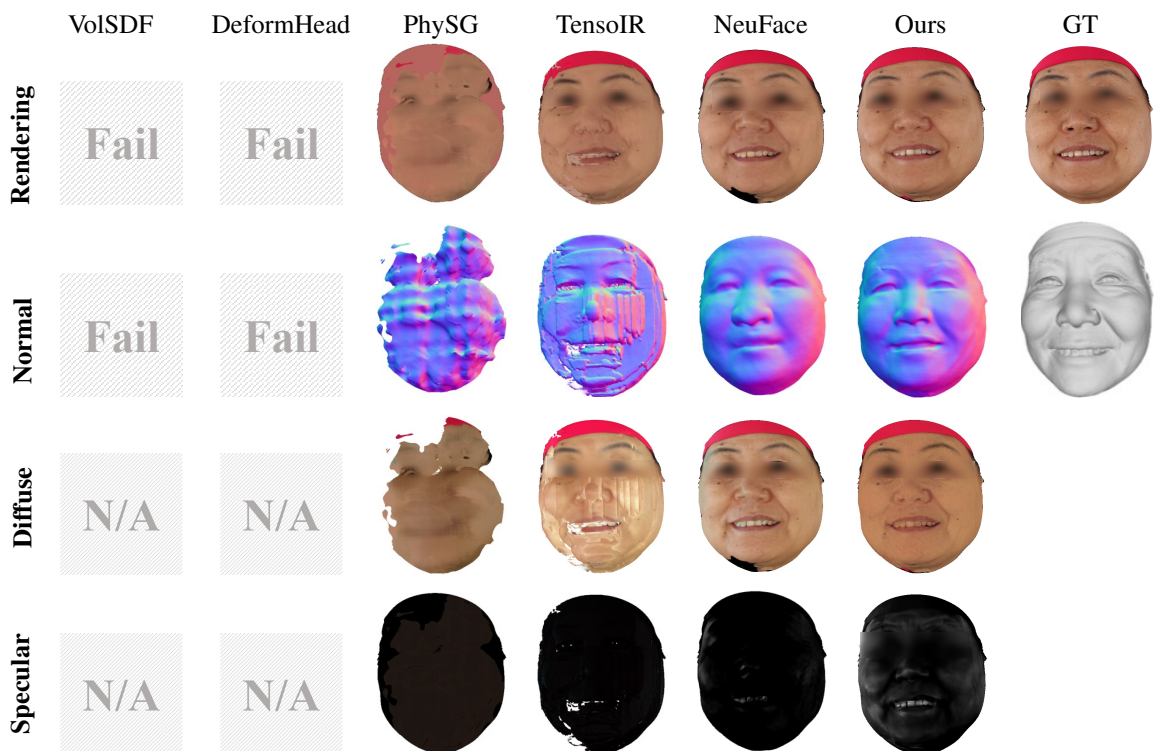


Figure 11. Comparison under a 3-view setting for Subject 14.

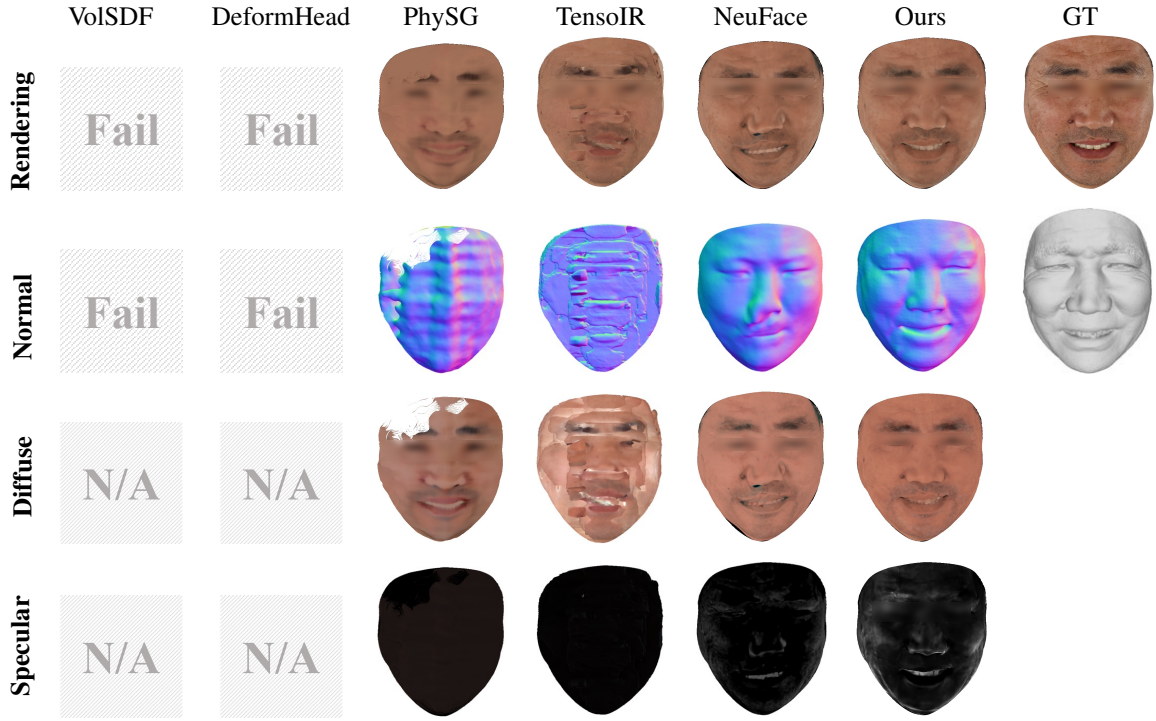


Figure 12. Comparison under a 3-view setting for Subject 15.

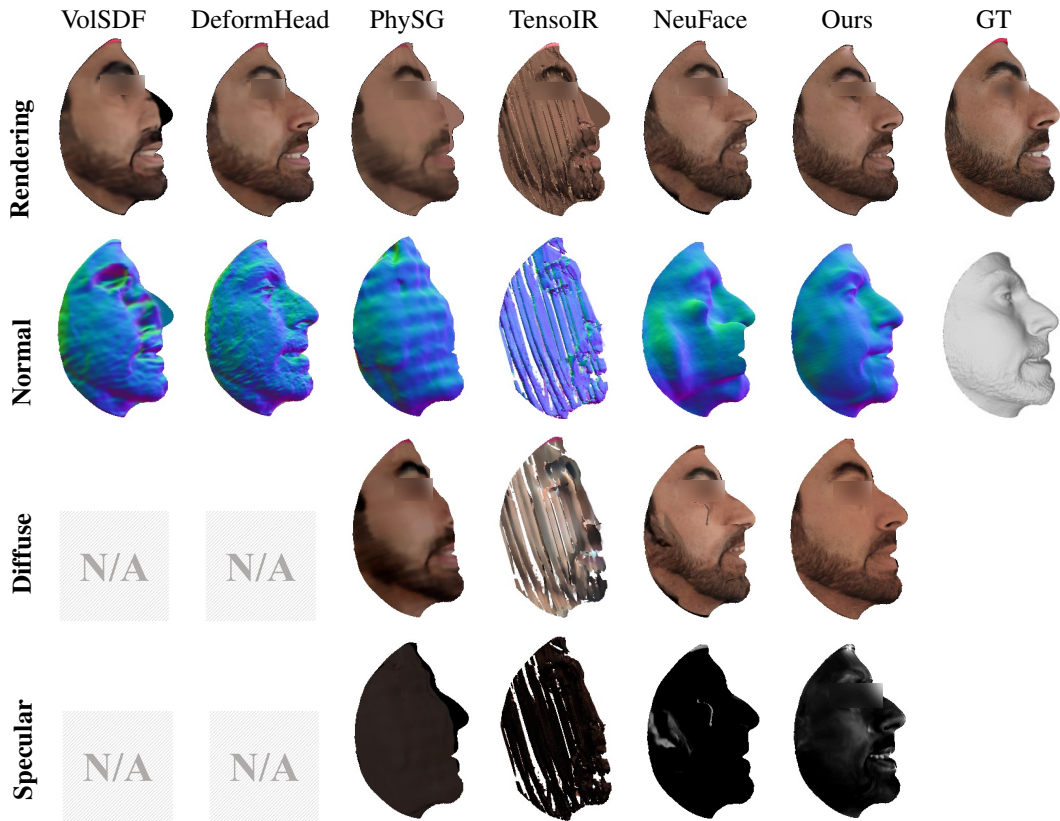


Figure 13. Comparison under a 3-view setting for Subject 17.

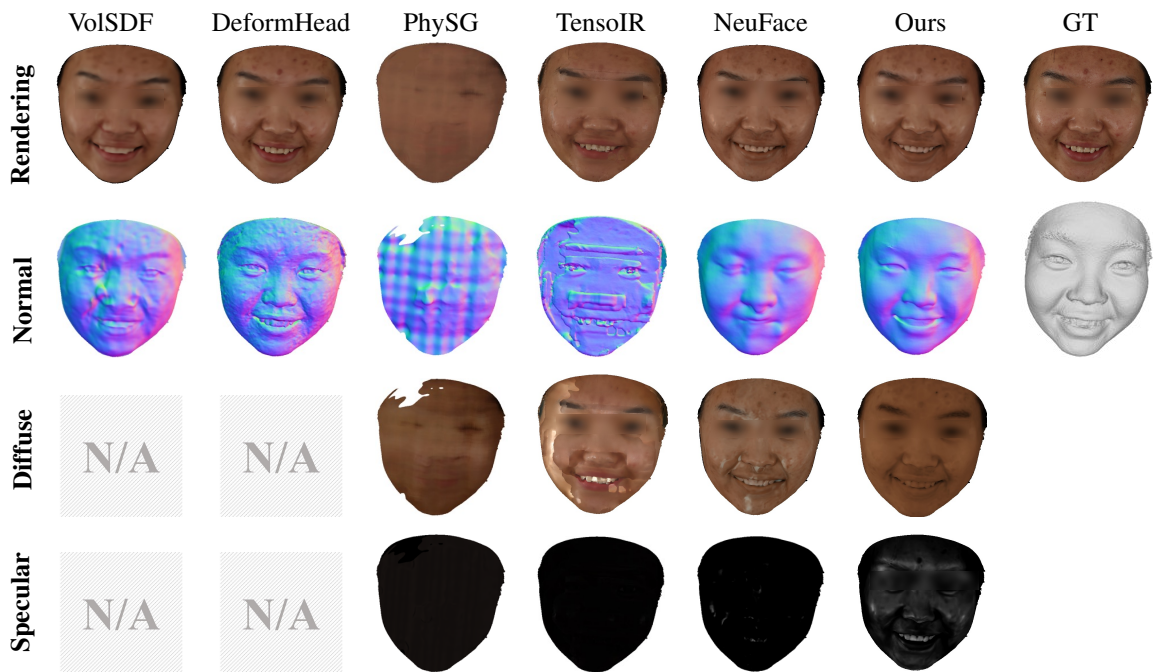


Figure 14. Comparison under a 3-view setting for Subject 342.

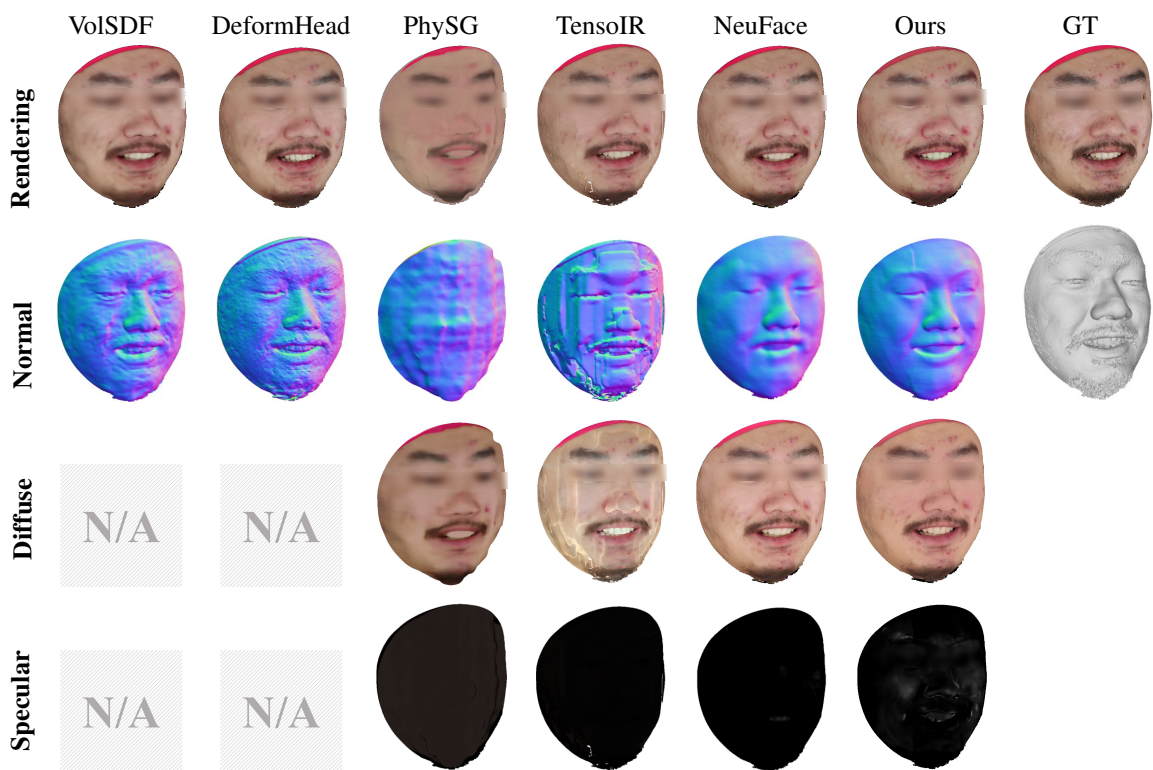


Figure 15. Comparison under a 5-view setting for Subject 12.

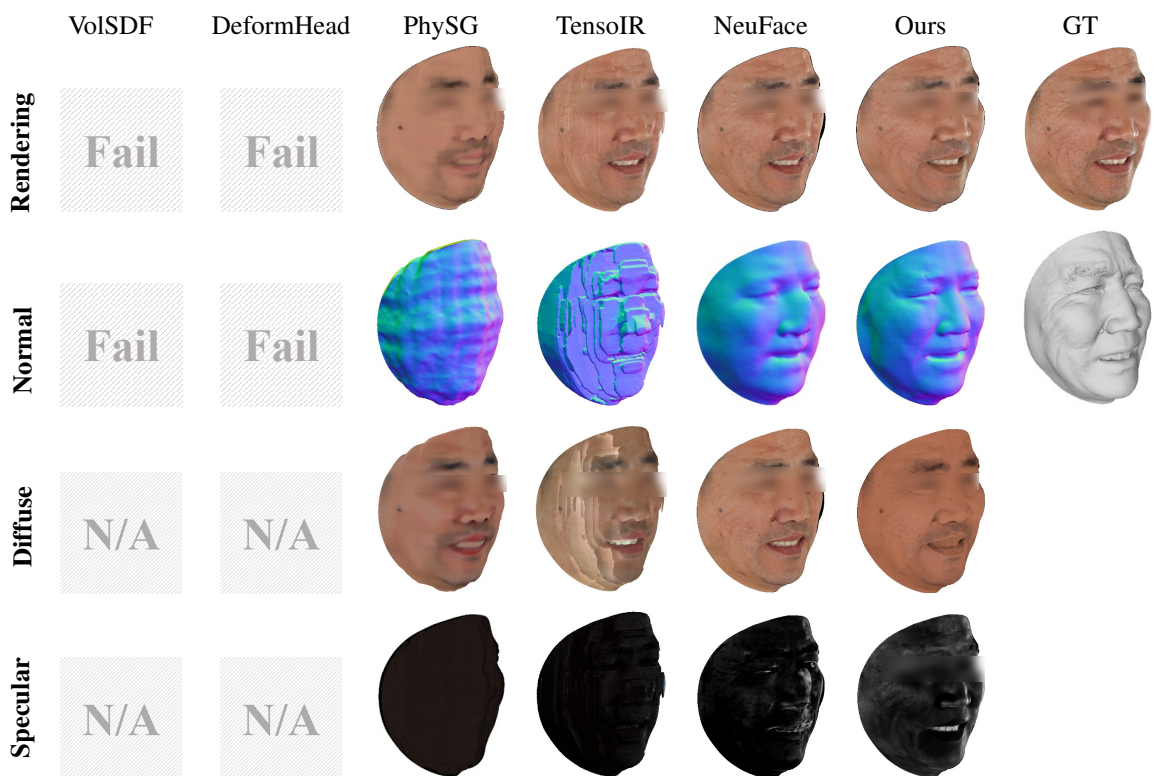


Figure 16. Comparison under a 5-view setting for Subject 15.

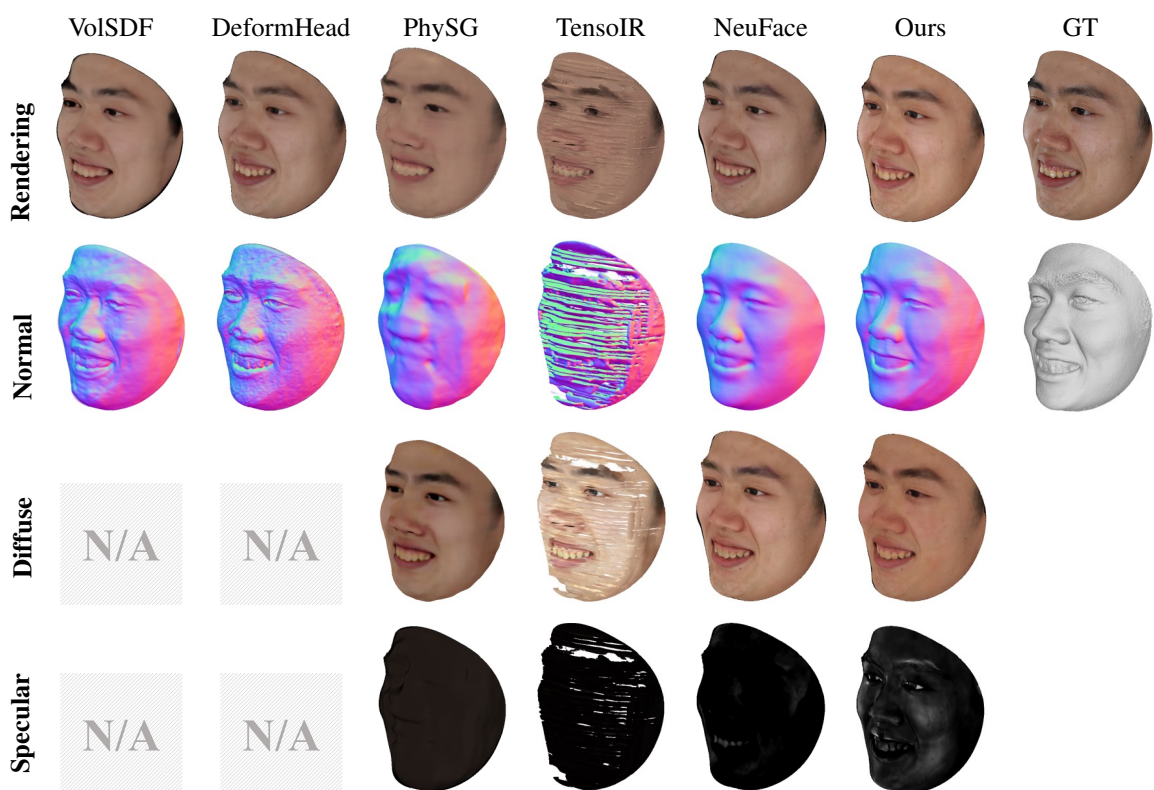


Figure 17. Comparison under a 5-view setting for Subject 212.

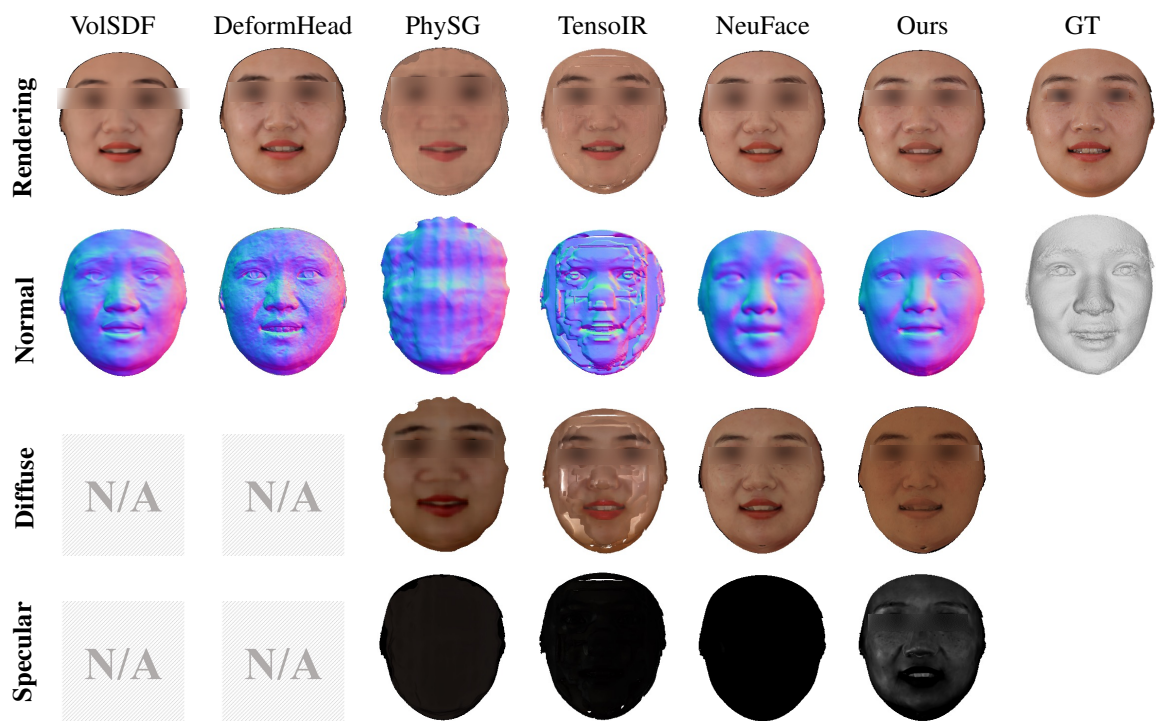


Figure 18. Comparison under a 5-view setting for Subject 327.

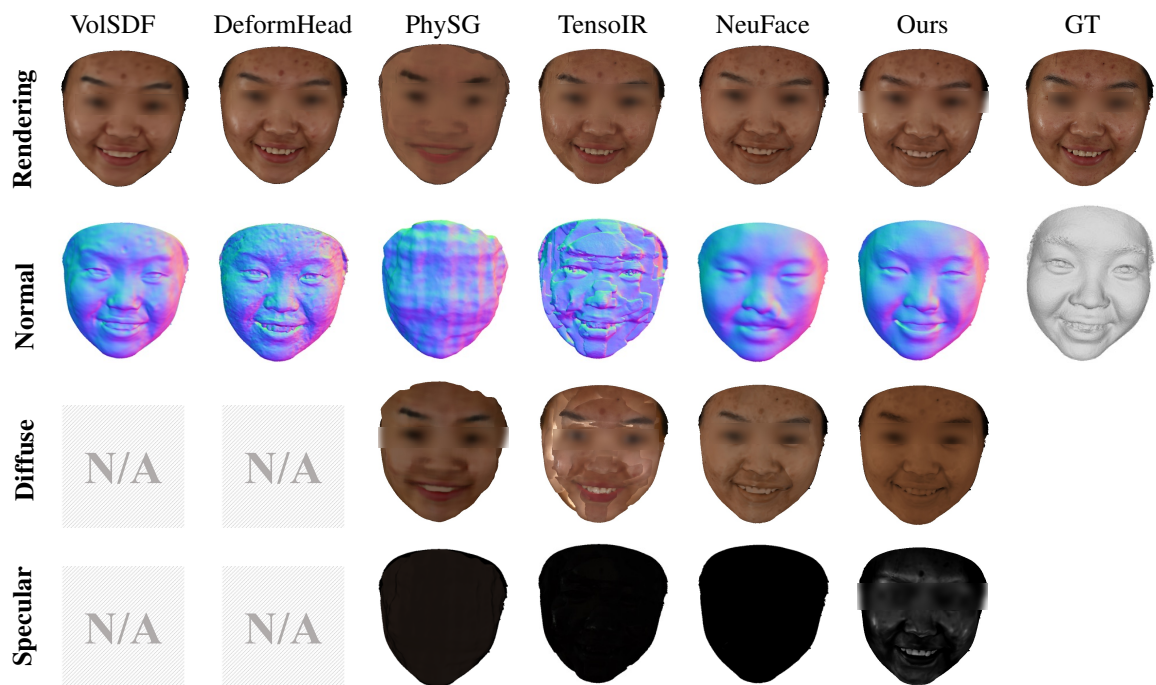


Figure 19. Comparison under a 5-view setting for Subject 342.

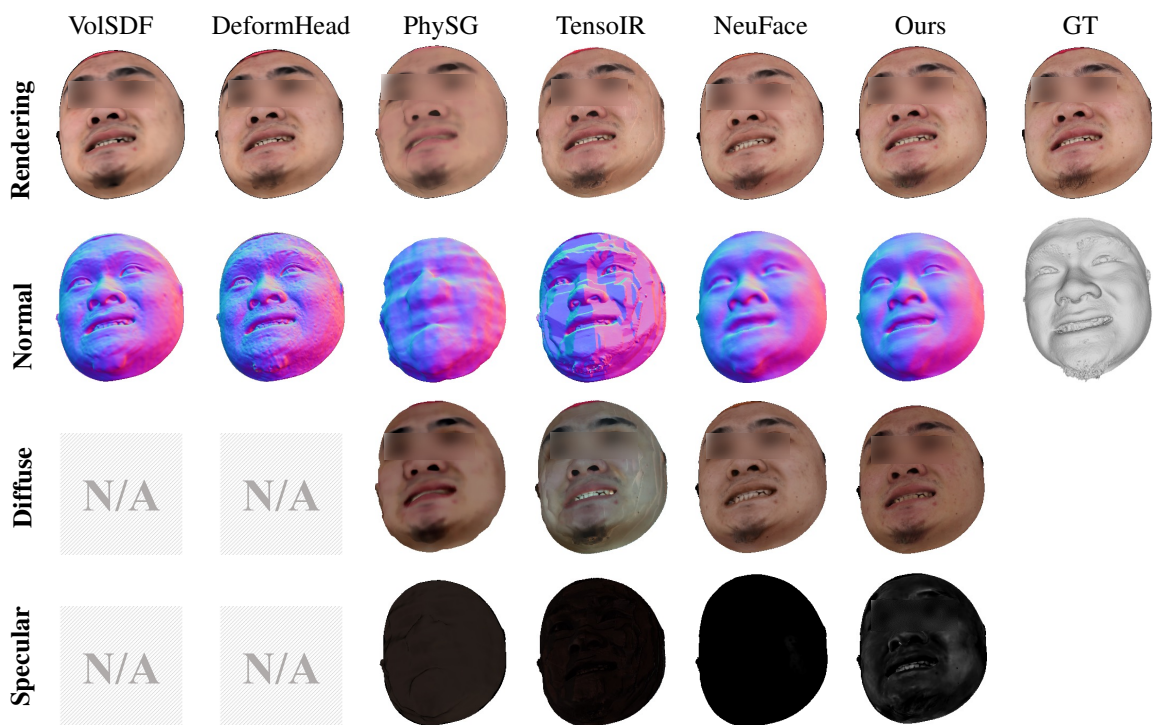


Figure 20. Comparison under a 10-view setting for Subject 13.

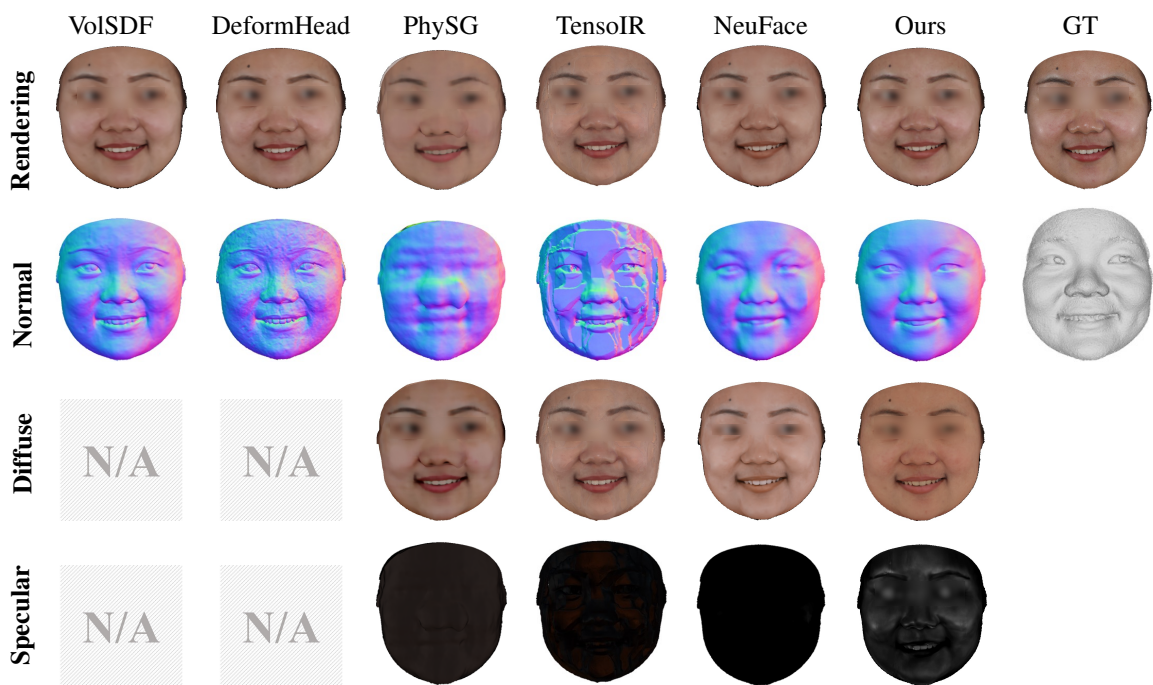


Figure 21. Comparison under a 10-view setting for Subject 139.

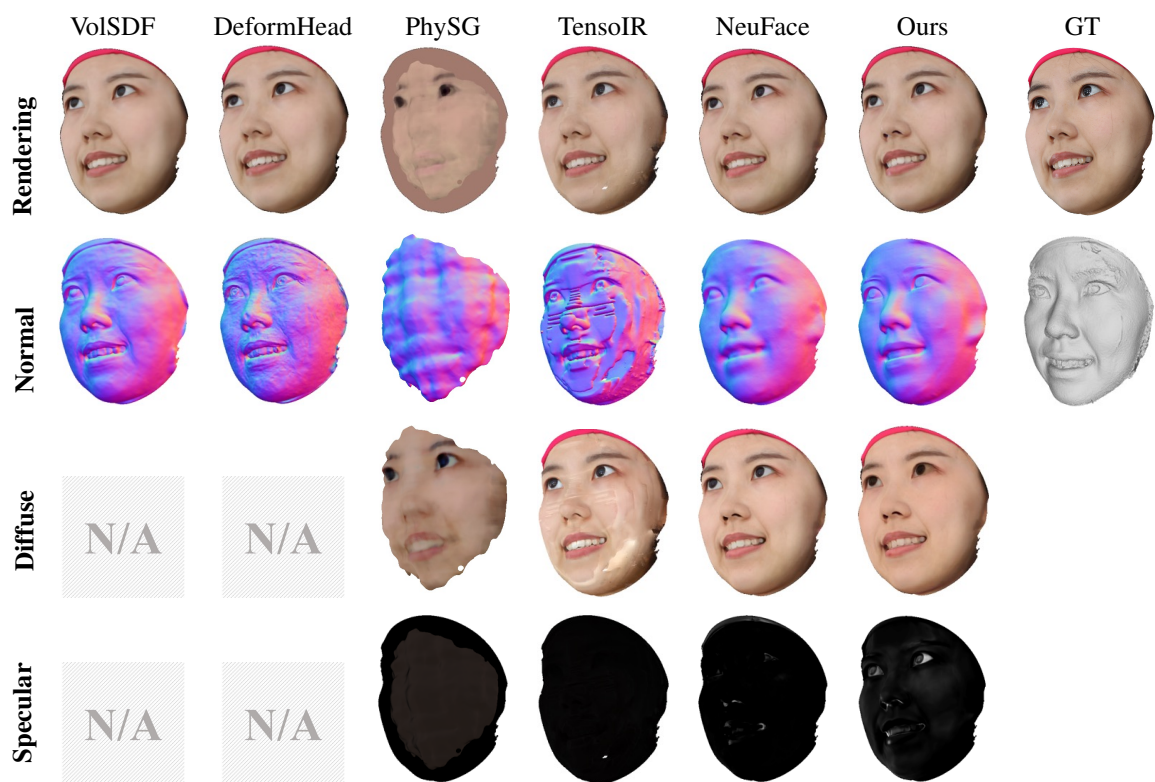


Figure 22. Comparison under a 10-view setting for Subject 344.