

FACE INVERSE RENDERING FROM SINGLE IMAGES IN THE WILD

Meng Wang^{1,3}, Wenjing Dai², Xiaojie Guo^{1,3*}, and Jiawan Zhang^{1,3}

¹School of Computer Software, Tianjin University, Tianjin, China

²Department of Management Engineering, Technical University of Denmark, Denmark

³Key Research Center for Surface Monitoring and Analysis of Cultural Relics, SACH, China

{autohdr, abigail.dai4, xj.max.guo} @gmail.com, jwzhang@tju.edu.cn

ABSTRACT

Face inverse rendering, an important and challenging task in computer vision and computer graphics, attempts to decompose face image into shape, reflectance, and illuminance. This problem becomes fundamentally difficult under non-laboratory conditions without controlled illumination. Though recent works have produced compelling results, most of these techniques rely on multiple lighting images captured under controlled lighting by complex equipment, such as Light Stage, which is not flexible and applicable to common users. In this paper, we propose a novel face inverse rendering framework, which neither relies on complex devices nor labeled training data. Instead, it learns reflectance, shape, and illuminance from its physical constraints. Extensive experiments on both synthetic and real image datasets demonstrate consistently superior performance of the proposed method. Our code will be made publicly available.

Index Terms— Face inverse rendering, Reflectance, Albedo, Normal, Shading

1. INTRODUCTION

The appearance of a face image depends on various factors, such as illumination, shape (normal) and material (albedo/reflectance). Face inverse rendering (FIR) aims to find a light function or decompose such a face image into albedo, normal and light. FIR can benefit many applications, such as face editing [1], relighting [2].

FIR is a classical problem and has been extensively investigated in the past decades [3, 1, 2]. To produce realistic relit face, multi-images based methods [4, 5] have been designed to acquire face images from different light conditions using controlled light equipment. These techniques are effective in diverse lighting, which indicates that equipment-based processes properly improve the quality. Unlike these methods focusing on complex equipment to get the accuracy of FIR

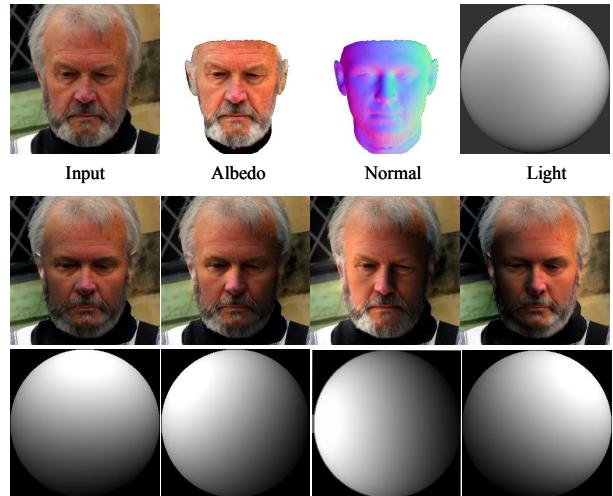


Fig. 1. Our algorithm takes a portrait image as input and produces FIR components, albedo, normal and light. The next are new relit faces generated from albedo, normal and target SH lighting in the last row.

components, we study FIR in a lightweight way to be suitable for consumer-level usage. We address this problem by learning a self-restraint of FIR components in the wild with an unsupervised manner.

To unsupervised learning, FIR without complex devices would be a non-trivial task because the diversity of face albedo and light in the real world. More recent unsupervised learning approaches [1, 6] explore the relationship of FIR components without using complex setups/ground truth. Motivated by this, we seek to take advantage of the relation among FIR components to constrain the decomposition.

In this paper, we present an unsupervised learning approach, which applies the self-restraint of FIR components to learn each component. To this end, we employ the relation between luminance and shading, intrinsic decomposition and physical-based inverse rendering to constrain our shading and light. With the synergy of these constraints, once the model is trained, it can produce each component only given a single

*Corresponding author. This work was supported by the National Key Research and Development Program of China under 2019YFC1521200, National Natural Science Foundation of China under Grant 62172295, and Grant 62072327, and TSTC under Grant 20JCQNJC01510.

image without losing high-frequency details.

To keep global illumination and speed up convergence, we introduce an independence shading network to deal with the global illumination. In summary, our contributions are: (1) We propose an end-to-end network for FIR, which can decompose a single face image into albedo, normal, light and shading in the wild without synthetic ground truth and complex equipment. (2) Our approach utilizes self-restraint of input and FIR components to model physical-based inverse rendering, and preserves high-frequency facial details on predicted albedo and normal. (3) Our methods achieves remarkably results on quality and qualitatively.

2. RELATED WORKS

In this section, we provide a brief overview of techniques relevant to inverse rendering work.

Intrinsic decompositions. Following Retinex theory [7], intrinsic decomposition has gained prominence. Priors have been widely used to constrain each component. For instance, non-local reflectance constraints [8, 3] based on chromatic and luminance. Rother et al. [8] propose a global potential on reflectances and model the reflectance as being drawn from a sparse set of basis colors. Barron et al. [3] assume that albedo is piecewise constancy. However, the estimated albedo would lose the details with smoothness priors. We train our model on shading and light constraint without albedo priors, which help preserve high-frequency details on albedo.

Device-based methods. The device-based techniques [4, 5] have been employed to address the unsuccessful problem. For instance, a specific hardware system is built to control the lighting and capture the reflectance fields of face under different illuminations [4]. The measurements-based reflectance model [5] applies custom-built devices to capture multi-images for modeling face reflectance. While complex equipment is not suitable for consumer-level usage.

Learning-based methods. To unsupervised methods, Shu et al. [1] build an end-to-end generative adversarial network to infer a face-specific disentangled representation for each component. Liu et al. [9] explore the independence between reflectance and shading and propose an unsupervised method with the help of domain invariant content constraint. Unsup3d [10] applies the principle of symmetric structure to learn the reflectance from a single-view image.

Another direction is supervised learning, InverseFaceNet [11], which is trained on synthetic data to estimate the equivalent components in real-world images without ground truth. In addition, SfSNet [2] learn low frequency from the labeled synthetic data and high-frequency details from real images. While they generate remarkable results, which is generally unavailable to in-the-wild images due to the distribution of synthetic data not matching the real world.

Priors have recently shown promising constraints on unsupervised inverse rendering. Inspired by [9, 10], we incorpo-

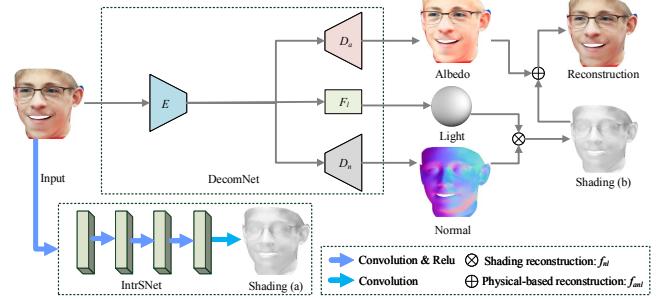


Fig. 2. Network architectures. Our network is composed of two sub-networks, DecomNet and IntrSNet. DecomNet decomposes input into albedo, normal, and light with IntrSNet. Shading (a) is directly predicted from the input by IntrSNet, Shading (b) is generated from estimated normal and light.

rate priors into our model to make the technique more flexible to users in the real world.

3. METHOD

Given an unconstrained collection of face images, we aim to decompose a single image without ground truth or a complex facility. For the sake of simplicity, in this paper, we assume that images are captured under Lambertian reflectance [12] and Retinex theory [7]. The FIR is represented as:

$$I(p) = f_{as}(A(p), S(p)) = f_{anl}(A(p), N(p), L), \quad (1)$$

where $I(p)$ is a pixel of the input image at the location p . f_{as} and f_{anl} are represented as the physical-based inverse rendering function under the lambertian model and Retinex theory, respectively. Reflectance $A(p)$ and shading $S(p)$ are intrinsic decomposition components. To lambertian reflectance, $A(p)$ is also the albedo, lighting L is a nine dimensional second order spherical harmonics coefficients, normal $N(p)$ and light L can be rendered as shading $S(p)$ by a function f_{nl} .

We only have images without ground truth at hand, additional constraints are indispensable for learning such a decomposition. Thus, normal initialization is firstly introduced into our framework, for coarse normal is easy to get from the modern methods, such as [13]. With the normal initialization, the problem can convert to albedo and lighting prediction, a ‘one-to-two’ problem significantly easier than ‘one-to-three’. Simply with normal initialization, it does not satisfy the task. Fortunately, shading always looks smooth, affected by lighting, light obeys statistical regularities and can be computed from shading and normal using least square optimization. Therefore, we design a network that further disentangles FIR components with the above constraints.

3.1. Architecture

Our architecture takes a source image I and outputs the FIR components, albedo A , shading S , normal N , and lighting L . The whole network can be seen in Figure 2. Our network adopts the architecture from Noise2noise [14], but rather than directly decoder shading from the whole network. Inspired by Kindling [15], our shading network separates from the entire network and individually predicts the shading. By separating the shading, we are able to guarantee the global illumination of the input images and allow fast to converge. In addition, image luminance is a combined effect of its albedo and illumination, and shading is generated from the normal (geometry) and the illumination. We assume that luminance and shading are certain similarities between them, which means the luminance would contribute to the shading. Thus, we also introduce the similarities between shading and luminance for the regularization of shading.

3.2. Training losses

The main loss function is reconstruction loss and adversarial loss. The reconstruction loss can reconstruct the input image from the decomposed components. The adversarial loss function can narrow the gap between input and reconstructed image. The loss function is represented as:

$$\mathcal{L}_R = \lambda_{recon} \mathcal{L}_{recon} + \lambda_{adv} \mathcal{L}_{adv}, \quad (2)$$

where $\mathcal{L}_{recon} = \|I - \hat{I}_{Ph}\|_1 + \|I - \hat{I}_{Re}\|_1$ is reconstruction loss consisted of physical-based rendering reconstruction \hat{I}_{Ph} and Rentinex reconstruction \hat{I}_{Re} , I is the input face image. $\mathcal{L}_{adv} = D(\hat{I}_{Ph})$, where D is a discriminator network.

As we all know, face normal is easy to get from previous research, such as [13]. Therefore, we regard the normal N as initialization to our network. It would keep our predicted normal close to plausible face normal during the training. We apply 3DMM normal [13] as our coarse initialization. We then introduce the following objective to \hat{N} :

$$\mathcal{L}_N = \lambda_n \|N - \hat{N}\|_2^2, \quad (3)$$

where λ_n is the weight to control the contribution of coarse normal in the whole loss.

Furthermore, following Rentinex theory [7], the shading should be piecewise smooth when affected by illumination. In order to better preserve the global illumination effects, we build a network similar to [15] to learn shading for regularization. The shading loss function is represented as:

$$\mathcal{L}_S = \lambda_s \left(\left\| \frac{\nabla \hat{S}_i}{\max(|\nabla \hat{S}_i|, \xi)} \right\|_1 + \left\| \frac{\nabla \hat{S}_j}{\max(|\nabla \hat{S}_j|, \xi)} \right\|_1 \right), \quad (4)$$

where $\nabla \hat{S}$ stands for derivative operator of $\nabla \hat{S}_x$ and $\nabla \hat{S}_y$ in the first order on estimated shading \hat{S} . In additional, $\xi =$

0.01, which is a small positive constant for avoiding zero denominator.

It is necessary to incorporate another regularizer for light to prevent the network from generating arbitrary light leading to albedo is similar to the input image. Similar to [1], coarse light \hat{l}' can be computed from normal \hat{N} and shading \hat{S} using least square optimization. In addition, we assume light is Gaussian distribution. Therefore, the light would follow a Gaussian distribution after predicting from our model, the light loss is:

$$\mathcal{L}_L = \lambda_l \|\hat{l}' - \hat{l}\|_1, \quad (5)$$

where \hat{l} is the predicted light, λ_l is the coarse light contributed to light.

Normal prediction is an ill-posed problem that suffers from bas-relief ambiguity unless given a shape prior. In our work, we exploit the gradient-preserving prior to normal prediction. Rather than predicting the normal by coarse normal initialization, which tends to lose edge details, we estimate the edge in a gradient-preserving manner. \mathcal{L}_G penalizes difference between gradient of estimated normal and the estimated albedo represented as:

$$\mathcal{L}_G = \lambda_g \|\nabla \hat{A} - \nabla \hat{N}\|_1. \quad (6)$$

It is worth noting that the gradient difference metric is more robust to improve normal accuracy, which further helps decompose other components.

To prevent the network generating bright shading (such as all pixels value are 1 in normalized shading) in the early training stages, we make the assumption that shading and luminance is a certain similarity. Therefore, we introduce a regularization term to shading and consequently avoid convergence of the generated albedo to the input image, the regularizer is represented as:

$$\mathcal{L}_{SL} = \lambda_{sl} \|\hat{S} - lum(I)\|_1, \quad (7)$$

where λ_{sl} is a weight to control the similarities contribution between shading \hat{S} and luminance $lum(I)$, which is computed from the input image I with a luminance function.

3.3. Implementation Details

The convolutional encoder E and decoder of albedo D_a and normal D_n is similar to Noise2noise [14]. The *IntSNet* is composed of five convolution layers with $64 \times 3 \times 3$, $128 \times 3 \times 3$, $64 \times 3 \times 3$, $32 \times 3 \times 3$ and $1 \times 3 \times 3$ filter sets. Each convolution is followed by a ReLU nonlinearity except the last one. Fully connected layer F_l is used to produce lighting coefficients.

We train our model about 200k iterations using a learning rate of 0.0001 and Adam optimizer with default parameters. The images is trained with 256×256 . During the training, we use $\lambda_{recon} = 0.5$, $\lambda_n = 0.5$, $\lambda_s = 0.01$, $\lambda_l = 0.1$, $\lambda_g = 0.01$ and $\lambda_{sl} = 1$. Besides, an adversarial loss with $\lambda_{adv} = 0.001$ is added to narrow the reconstruction errors between the input and reconstruction.

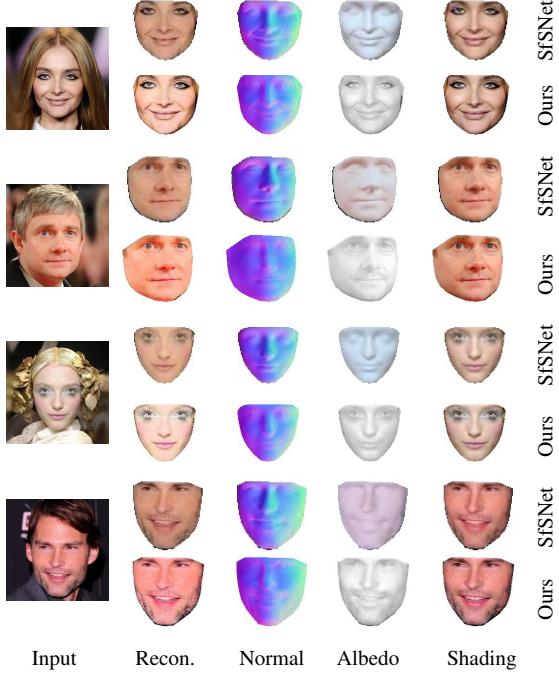


Fig. 3. Inverse rendering. Ours inverse rendering results compare with SfSNet [2]. It is worth noting that SfSNet [2] is a supervised learning method.

4. EXPERIMENTS

In this section, we compare our method with the state-of-the-art methods and evaluate it in quantitatively and qualitatively.

4.1. Datasets

We test our method on four datasets: FFHQ [16], Photoface [17] and CelebA [18]. Besides, we use lighting in DPR [19] and normal, albedo in SfSNet [2] to generate new data, which is used to evaluate the accuracy of our predictions.

4.2. Qualitative and quantitative evaluations

To compare with previous FIR methods, we first compare our method on the CelebA dataset with SfSNet [2], which can be regarded as the state-of-the-art method in this field. In Figure 3, the results show that our method is more applicable to faces with different skin tones, whereas SfSNet decomposes the albedo in favor of a color distribution, i.e. yellowish, due to the constraint of the synthetic dataset.

In Figure 4, we show some results with a wide range of ages and ethnicity on FFHQ [16]. Our technique can obviously maintain a more realistic albedo of the face and avoid inaccurate albedo and light decomposition due to the synthetic data set does not correspond to the real data distribution in the wild.

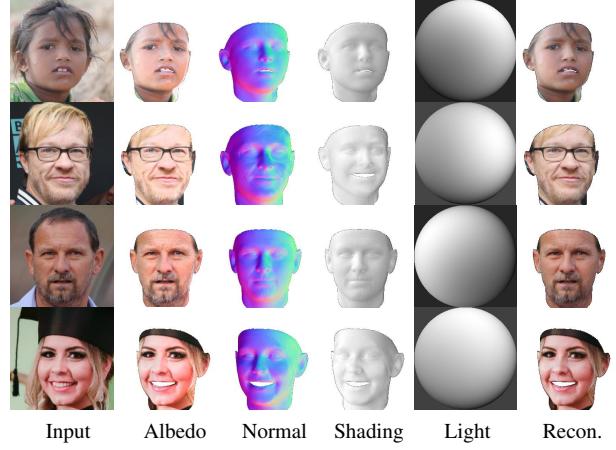


Fig. 4. FIR results on FFHQ [16]. The results show that our model can be applied to faces with diverse races and maintain high frequency details on albedo.

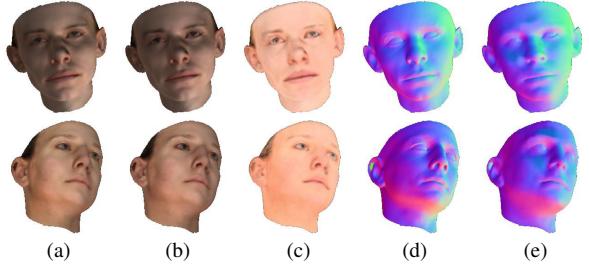


Fig. 5. Results on synthetic data. (a) Input, (b) Our reconstruction, (c) Albedo, (d) GT normal, (e) Our normal.

We compare our predicted normals on quality with the state-of-the-art methods from a single image. To this purpose, we evaluate our normal prediction using Photoface [17], which is computed from the photometric stereo and regarded as ground truth. Since the previous methods do not provide how to split the dataset, we first select all the data and then randomly split the data into training and test data. Similar to SfSNet [2], we use the mean angular error of the normal and the percentage of pixels at various angular error thresholds.

Table 1 shows that the accuracy of our method is not as good as SfSNet-ft [2] in estimating normal on real data, because their method relies on large amounts of synthetic data with albedo and light ground truth that help their model accurately decompose illumination and albedo, whereas our method only relies on the synthetic norml as the initialization and lacks constraints of illumination and albedo, which affects the accuracy of normal. It is worth mentioning that 3DMM [13] is not trained on this dataset. NiW [21] and SfSNet-ft [2] are trained on this dataset. The last two rows of the table, which are trained on our synthetic data. Compared with SfSNet, our method can greatly improve the accuracy of normal prediction due to the lighting distribution matches our light model. It also shows that an accurate light decomposition helps to improve the accuracy of other components, such

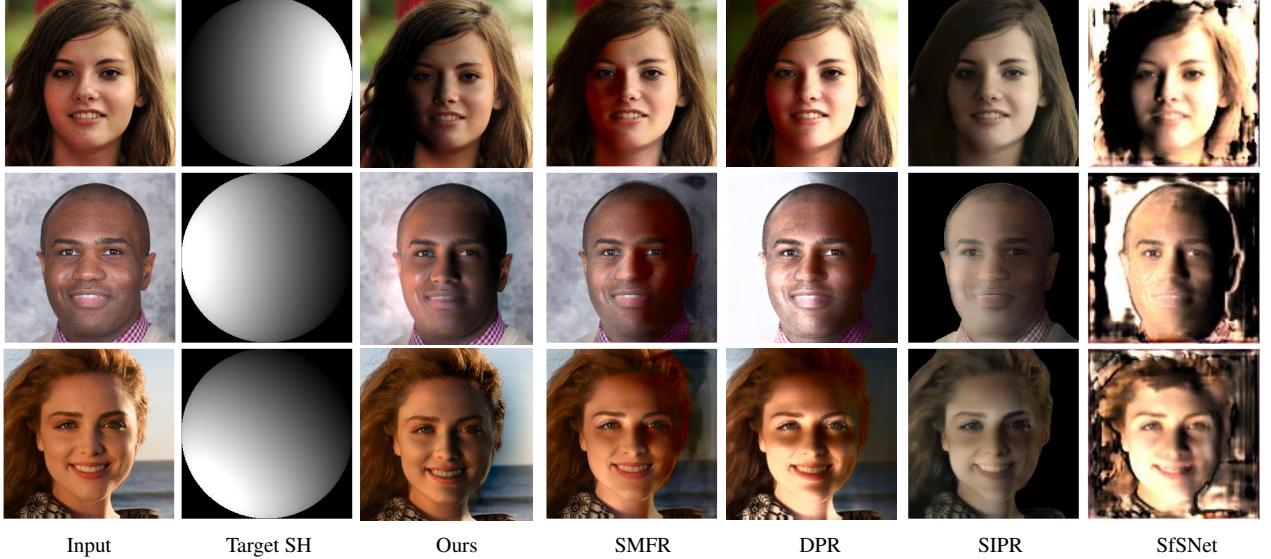


Fig. 6. Relighting comparison with the state-of-the-art methods on FFHQ [16]. Our model produces gradually shift shadows, which outperform others. We utilize Poisson blending [20] to our results with the original image.

| Method | Mean \pm std | $< 20^\circ$ | $< 25^\circ$ | $< 30^\circ$ |
|------------|----------------------------------|--------------|--------------|--------------|
| 3DMM | 26.3 ± 10.2 | 4.3% | 56.1% | 89.4% |
| NiW | 22.0 ± 6.3 | 36.6% | 59.8% | 79.6% |
| SfSNet-ft | 12.8 ± 5.4 | 83.7% | 90.8% | 94.5% |
| Ours | 8.9 ± 12.9 | 78.1% | 86.8% | 92.5% |
| SfSNet-syn | 10.6 ± 9.4 | 85.1% | 91.4% | 95.2% |
| Ours-syn | 7.8 ± 10.5 | 83.4% | 93.2% | 96.3% |

Table 1. Normal reconstruction error on Photoface. The data comes from SfsNet, except for ‘SfSNet-syn’ and ‘Ours-syn’ are trained on our synthetic data. The lower is better for mean error, and the higher is better for correct pixels at various thresholds.

as normal estimation.

In Figure 5, we provide our results on synthetic data, as expected that our method can remove lighting from the input and obtain the albedo. It can also find that our reconstruction and normal estimation are very close to the ground truth data.

4.3. FIR components and relighting comparison

To evaluate FIR components, we compare our results with SfsNet [2] on our synthetic data. In table 2, we show that our reconstruction error outperforms SfsNet, while our predicted albedo is not as good as SfsNet, due to our training in an unsupervised manner. It would be difficult to distinguish the ambiguity of lighting and albedo without ground truth. The ambiguity leads to poor results to albedo and also affects the light estimation.

In Figure 6, we compare the portrait relighting results

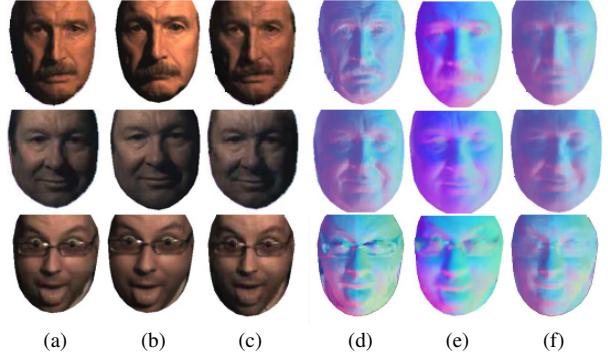


Fig. 7. Normal and reconstruction comparison on Photoface dataset. (a) Input, (b) Our reconstruction, (c) SfSNet reconstruction, (d) GT normal, (e) Our normal, (f) SfSNet normal. The same results used in SfSNet.

on FFHQ [16] to the state-of-the-art methods SMFR [6], DPR [19], SIPR [22], and SfSNet [2]. Our results provide a more realistic lighting effect on the face than SMFR because the light changes gradually on a face, which conforms to real natural lighting effects without high-contrast shadows. The shadow effect on the human face is generated by the interaction of face geometry and lighting, and ours is more in line with the real shadow effect under the physical-based model than SMFR [6] generated by the fitted shadow function. More results can be found in the supplementary.

5. ABLATION STUDIES

We study effects of different losses on quality of the reconstructed images. Namely, our training without reconstruction

| Methods | Albedo Error | | Recon. Error | | Light Acc. | | |
|---------|--------------|-------|--------------|------|------------|--------|--------|
| | MAE | RMSE | MAE | RMSE | Rank 1 | Rank 2 | Rank 3 |
| SfSNet | 10.54 | 13.47 | 0.31 | 1.08 | 79.12 | 90.31 | 93.21 |
| Ours | 23.02 | 41.2 | 0.44 | 1.43 | 73.34 | 82.0 | 86.78 |

Table 2. Albedo and reconstruction MAE and RMSE on synthetic datasets. We evaluate our predicted albedo and reconstruction with SfSNet.

loss of intrinsic decomposition, ‘w/o AS’; without gradient loss, ‘w/o Grad’; without constraint loss between luminance and intrinsic decomposition shading, ‘w/o Lumin’; without shading smoothness loss, ‘w/o Kind’; the loss between Intrinsic decomposition shading and physical-based shading, ‘w/o SNL’; with and without physical-based reconstruction loss, ‘w ANL’ and ‘w/o ANL’, respectively.

| Training | Albedo Error | | Recon. Error | | Normal Error | |
|-----------|--------------|-------|--------------|------|--------------|-------|
| | MAE | RMSE | MAE | RMSE | mean | std |
| w/o AS | - | - | - | - | - | - |
| w/o Grad | 28.86 | 51.28 | 0.46 | 1.08 | 10.22 | 6.73 |
| w/o Lumin | 47.83 | 75.63 | 0.27 | 0.94 | 8.43 | 5.9 |
| w/o Kind | 28.81 | 51.27 | 0.49 | 1.13 | 8.36 | 6.84 |
| w/o SNL | 29.41 | 55.12 | 0.53 | 1.26 | 8.74 | 7.02 |
| w ANL | 28.69 | 51.04 | 0.46 | 1.05 | 6.84 | 10.36 |
| w/o ANL | 23.02 | 41.2 | 0.44 | 1.43 | 9.34 | 6.22 |

Table 3. Ablation Experiments. We evaluate the performance effected by various loss on synthetic data.

As shown in Table 3, the model trained without ‘w/o AS’ is failed to FIR components decomposition due to white shading makes albedo is similar to the input. The training ‘w/o Grad’ would lose high-frequency details of normal. ‘w/o Lumin’, ‘w/o Kind’ and ‘w/o SNL’ lead to the ambiguity of albedo and light. Furthermore, ‘w/o ANL’ performs better than ‘w ANL’, for the loss back-forward twice, leading to a wobble problem and increasing the training difficulty. Thus, our model trained without physical-based reconstruction, ‘w/o ANL’ for balance.

6. CONCLUSIONS

In realizing that strong regularizations are a new path for unsupervised FIR components decomposition in the wild, this work learns FIR components from single-view unconstrained images. The model can produce high-frequency albedo and normal. The results show that our method significantly outperforms previous methods. Hopefully, with findings discussed in the paper, this work can be a step toward unlocking the possibility to capture details of FIR components.

7. REFERENCES

- [1] Zhixin Shu, Ersin Yumer, Sunil Hadap, Kalyan Sunkavalli, Eli Shechtman, and Dimitris Samaras, “Neural face editing with intrinsic image disentangling,” in *CVPR*, 2017, pp. 5541–5550.
- [2] Soumyadip Sengupta, Angjoo Kanazawa, Carlos D Castillo, and David W Jacobs, “Sfsnet: Learning shape, reflectance and illuminance of faces in the wild,” in *CVPR*, 2018, pp. 6296–6305.
- [3] Jonathan T Barron and Jitendra Malik, “Shape, illumination, and reflectance from shading,” *TPAMI*, vol. 37, no. 8, pp. 1670–1687, 2014.
- [4] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar, “Acquiring the reflectance field of a human face,” in *Computer graphics and interactive techniques*, 2000, pp. 145–156.
- [5] Tim Weyrich, Wojciech Matusik, Hanspeter Pfister, Bernd Bickel, Craig Donner, Chien Tu, Janet McAndless, Jinho Lee, Addy Ngan, Henrik Wann Jensen, et al., “Analysis of human faces using a measurement-based skin reflectance model,” *TOG*, vol. 25, no. 3, pp. 1013–1024, 2006.
- [6] Andrew Hou, Ze Zhang, Michel Sarkis, Ning Bi, Yiyi Tong, and Xiaoming Liu, “Towards high fidelity face relighting with realistic shadows,” in *CVPR*, 2021, pp. 14719–14728.
- [7] Edwin H Land and John J McCann, “Lightness and retinex theory,” *Josa*, vol. 61, no. 1, pp. 1–11, 1971.
- [8] Carsten Rother, Martin Kiefel, Lumin Zhang, Bernhard Schölkopf, and Peter Gehler, “Recovering intrinsic images with a global sparsity prior on reflectance,” *NeurIPS*, vol. 24, pp. 765–773, 2011.
- [9] Yunfei Liu, Yu Li, Shaodi You, and Feng Lu, “Unsupervised learning for intrinsic image decomposition from a single image,” in *CVPR*, 2020, pp. 3248–3257.
- [10] Shangzhe Wu, Christian Rupprecht, and Andrea Vedaldi, “Unsupervised learning of probably symmetric deformable 3d objects from images in the wild,” in *CVPR*, 2020, pp. 1–10.
- [11] Hyeongwoo Kim, Michael Zollhöfer, Ayush Tewari, Justus Thies, Christian Richardt, and Christian Theobalt, “Inversefacenet: Deep monocular inverse face rendering,” in *CVPR*, 2018, pp. 4625–4634.
- [12] David W Jacobs and Ronen Basri, “Lambertian reflectance and linear subspaces,” Feb. 8 2005, US Patent 6,853,745.
- [13] Volker Blanz and Thomas Vetter, “A morphable model for the synthesis of 3d faces,” in *SIGGRAPH*, 1999, pp. 187–194.
- [14] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila, “Noise2noise: Learning image restoration without clean data,” *arXiv preprint arXiv:1803.04189*, 2018.
- [15] Yonghua Zhang, Jiawan Zhang, and Xiaoqie Guo, “Kindling the darkness: A practical low-light image enhancer,” in *ACM MM*, 2019, pp. 1632–1640.
- [16] Tero Karras, Samuli Laine, and Timo Aila, “A style-based generator architecture for generative adversarial networks,” in *CVPR*, 2019, pp. 4401–4410.
- [17] Stefanos Zafeiriou, Mark Hansen, Gary Atkinson, Vasileios Argyriou, Maria Petrou, Melvyn Smith, and Lyndon Smith, “The photoface database,” in *CVPRW*, 2011, pp. 132–139.
- [18] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaou Tang, “Deep learning face attributes in the wild,” in *ICCV*, 2015, pp. 3730–3738.
- [19] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David W Jacobs, “Deep single-image portrait relighting,” in *ICCV*, 2019, pp. 7194–7202.
- [20] Patrick Pérez, Michel Gangnet, and Andrew Blake, “Poisson image editing,” in *SIGGRAPH*, pp. 313–318. 2003.
- [21] G Trigeorgis, P Snape, S Zafeiriou, and I Kokkinos, “Normal estimation for” in-the-wild” faces using fully convolutional networks,” in *CVPR*, 2017, vol. 2, p. 5.
- [22] Tiancheng Sun, Jonathan T Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul E Debevec, and Ravi Ramamoorthi, “Single image portrait relighting,” *TOG*, vol. 38, no. 4, pp. 79–1, 2019.