ReEnFP: Detail-Preserving Face Reconstruction by Encoding Facial Priors

In this supplementary file, we will provide more results to validate the effectiveness of our method. Section 1 introduces the implementation details. Section 2 demonstrates texture reconstruction performance, including an illustration of some intermediate results and comparison with previous [7, 8, 11, 16, 18] and recent [6, 12] studies. In section 3, we present geometry comparison with state-of-art methods [21, 5, 9, 3, 4, 17].

1. Texture Results

1.1. More Results on Texture Reconstruction

Experiment Settings. We present more intermediate results in Fig. 1 to further demonstrate representation power of our model. From left to right list the input image, reconstructed albedo and illumination, UV texture, transformed UV texture and its rendered image. Our model encodes the input image to the latent space of a fixed texture generator and produces its albedo and illumination. The illumination is represented as a single-channel ratio image to describe brightness as in previous relighting study [10]. Following a similar paradigm in [10], we multiply the Y channel of albedo by ratio image in YUV color space and convert it back to RGB space to obtain the UV texture. Since we employ a mesh topology [13] different from previous work [14], we convert back to Basel Face Model 2009 (BMF09) [14] by NRICP [1]. Note that our UV texture do not include eyeballs, so we rasterize the input eye part to UV of BFM09 as [13]. Finally, we show the rendered image at two different angles in the rightmost column.

Analysis of Experiment Results. As is shown in Fig. 1, our model achieves great robustness in reconstruction under varying facial appearance scenarios. Realistic texture with high-fidelity can be reconstructed for portraits of different skin colors and diverse facial features. Besides, all the albedo images share muted tones, which will be adjusted by ratio image to fit in real-world conditions. The ratio image is responsible for correctly increasing the brightness and contrast of albedo, making the lighting close to real-world environmental illumination. This practice reduces the difficulty of facial appearance learning, which is also validated in ablation study.

1.2. Texture Comparison with Existing Methods

Comparison Methods. We demonstrate our framework's capability to represent facial texture by providing more comparison results with state-of-art methods [7, 8, 11, 16, 18, 6, 12]. Works of Tran *et al.* [18], Jackson *et al.* [11] and Tewari *et al.* [16] are further introduced for comparison. Tran *et al.* [18] proposes a novel strategy to learn additional proxies as means to side-step strong regularizations and promote detailed shape/albedo. Their method improves the non-linear 3D morphable model in both learning objective and network architecture. Jackson *et al.* [11] follows non-parametric model strategy by directly regressing from input image to voxel coordinates with an hourglass structure. Without using statistical face models, their approach has large potential for exploring solution space. Tewari *et al.* [16] tries to extend 3DMM representation power by embedding 3DMM basis into DCNN. Their training scheme combines the advantages of 3DMM for regularization with out-of-space generalization of a learned corrective space.

Analysis of Comparison. Comparison with previous methods [7, 8, 11, 16, 18] is shown in Fig. 2. Although Tewari *et al.* [16] recovers more details than before, their method still struggles to recover high-level details because their training process involves strong regularization. The reconstructed surface of Jackson *et al.* [11] is not smooth due to high freedom of their model. Genova *et al.* [8] adopt an unsupervised scheme to regress 2D image to 3DMM coordinates where only unlabeled images and synthesized images are used. But their results still suffer from over-smoothed artifacts since their model is restricted to linear model. Targeting at improving on non-linear 3DMM, Tran *et al.* [18] captures higher-level details. Gecer *et al.* [7] achieves realistic texture reconstruction by optimizing the latent code of a fixed generator. But the identity preservation is not very well since their manually collected dataset is unable to cover diverse facial appearances. In contrast, our approach compares favorably to them in terms of both identity similarity and high-level detail reconstruction. Fig. 3 illustrates more comparison results with recent works(e.g. OSTeC [6] and AvatarMe [12]). Our performance is on par with OSTeC [6] but without huge amount of time consumption.

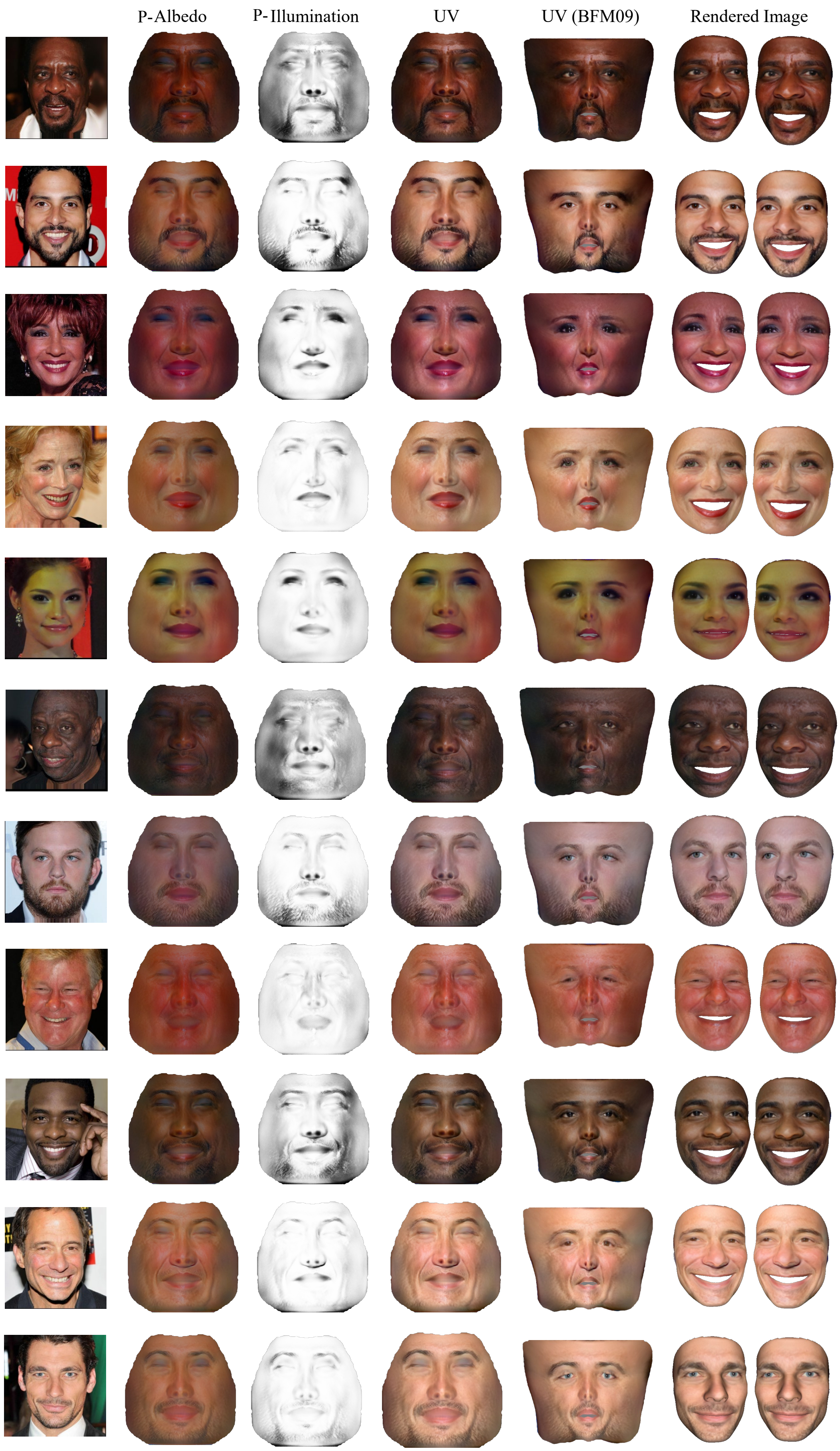

Figure 1: **More Texture Reconstruction Results.**

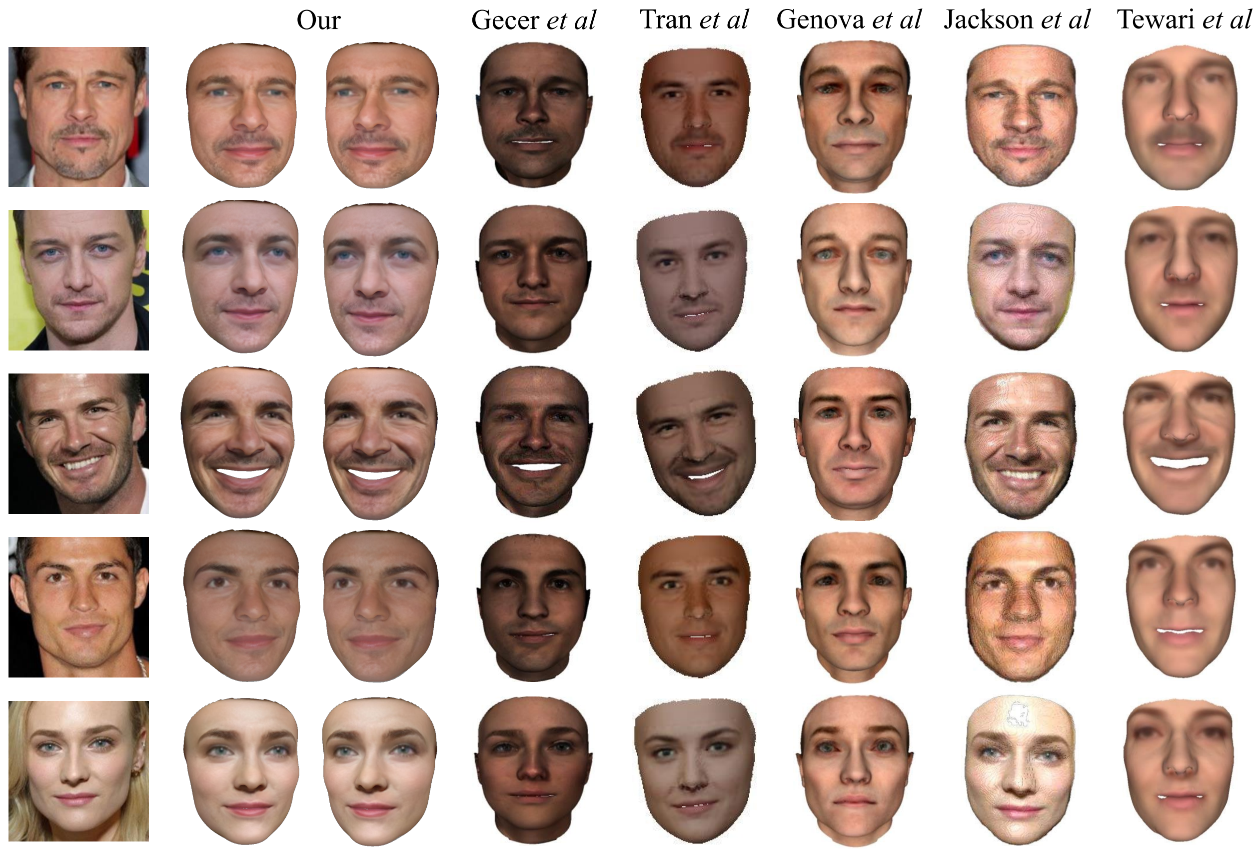

Figure 2: **Texture Comparison with Existing Methods** [7, 8, 11, 16, 18].

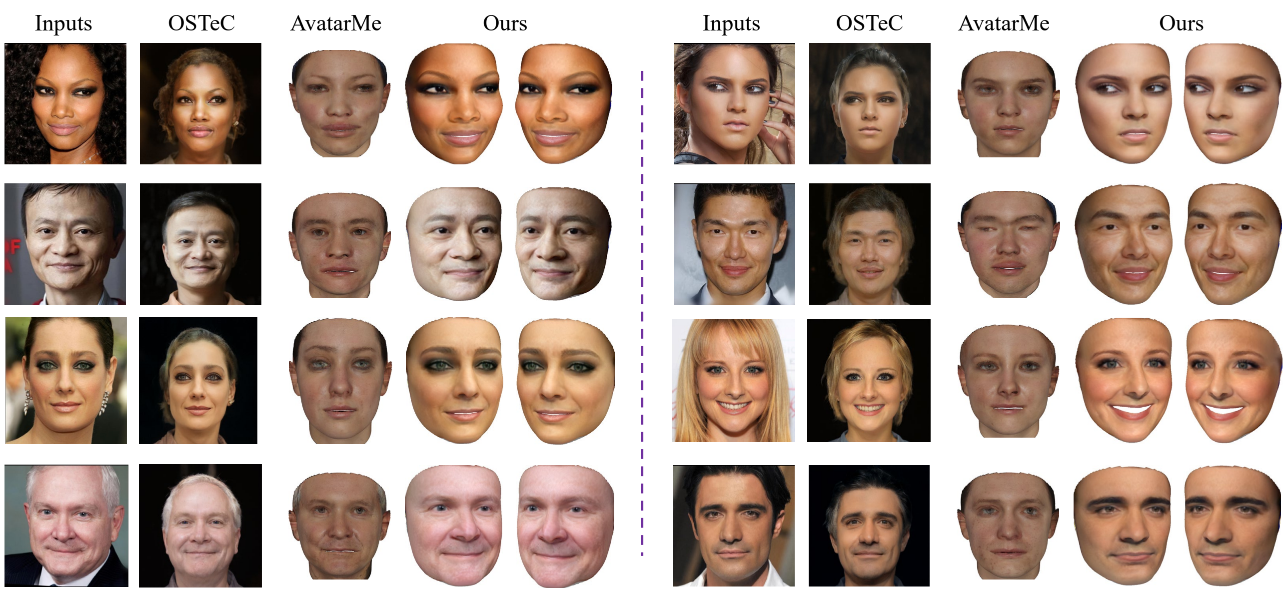

Figure 3: **Texture Comparison with Existing Methods** [6, 12].

2. Geometry Results

We provide more visual geometric comparison results with existing studies [21, 5, 9, 3, 4, 17] in Fig. 4. Fo-

Figure 4: **Geometry Comparison with Existing Methods** [21, 5, 9, 3, 4, 17].

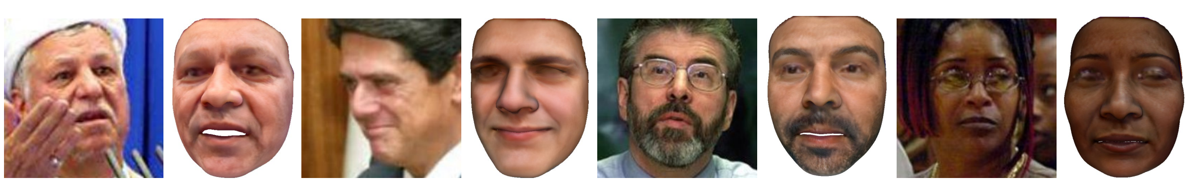

Figure 5: Our pipeline achieves high robustness in occlusion or large pose situation.

cusing on detail preservation, our method achieves better fine-grained geometric manifestation consistent with input image. For comprehensive comparison, we also evaluate geometric reconstruction results as previous works [19, 2, 15, 3]. Table. 1 depicts comparison of the point-to-planes distance of reconstructed meshes in MICC dataset, which suggests our framework achieve competitive results with state-of-art methods. Moreover, we achieve lower dis-

Table 1: **Comparison of Geometric Reconstruction Error.** Point-to-Plane distance under various scenarios in MICC dataset is demonstrated

	Cooperative		Indoor		Outdoor	
Method	Mean	Std	Mean	Std	Mean	Std
Tran *et al.*	1.93	**0.27**	2.02	0.25	1.86	0.23
Booth *et al.*	1.82	0.29	1.85	**0.22**	1.63	**0.16**
Piotraschke *et al.*	1.68	0.57	1.67	0.47	1.73	0.53
Deng *et al.*	1.60	0.51	1.61	0.44	1.63	0.47
Ours	**1.51**	0.55	**1.52**	0.42	**1.54**	0.48

Figure 6: **Demonstration of Relighting Application**. The input images and its inverted p-albedos are illustrated in the first row. The second and third row show rendering results under various illumination environment.

placement L1 error than [20] on validation split (12.20 vs. 14.34), which justifies our architecture design superior to their UNet [20] strategy.

3. Further Analysis

Robustness in Occlusion Scenario. As illustrated in Fig. 5, our methods demonstrate great robustness in reproduce textures under challenging conditions *e.g.* low-quality input, large pose and occlusion by hands or glasses.

Demonstration of Relighting Application. We depict the relighting results as shown in Fig. 6. Although our network is not designed for disentanglement of illumination and albedo, we here still illustrate some relighting applications based on synthesized pseudo-albedo. Since only weak supervision is utilized without other labels, some illumination information remains in the pseudo-albedo.

References

[1] Brian Amberg, Sami Romdhani, and Thomas Vetter. Optimal step nonrigid icp algorithms for surface registration. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.

[2] James Booth, Epameinondas Antonakos, Stylianos Ploumpis, George Trigeorgis, Yannis Panagakis, and Stefanos Zafeiriou. 3d face morphable models" in-the-wild". In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 48–57, 2017.

[3] Yu Deng, Jiaolong Yang, Sicheng Xu, Dong Chen, Yunde Jia, and Xin Tong. Accurate 3d face reconstruction with weakly-supervised learning: From single image to image set. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019.

[4] Yao Feng, Haiwen Feng, Michael J Black, and Timo Bolkart. Learning an animatable detailed 3d face model from in-the-wild images. *ACM Transactions on Graphics (TOG)*, 40(4):1–13, 2021.

[5] Yao Feng, Fan Wu, Xiaohu Shao, Yanfeng Wang, and Xi Zhou. Joint 3d face reconstruction and dense alignment with position map regression network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 534–551, 2018.

[6] Baris Gecer, Jiankang Deng, and Stefanos Zafeiriou. Ostec: One-shot texture completion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7628–7638, 2021.

[7] Baris Gecer, Stylianos Ploumpis, Irene Kotsia, and Stefanos Zafeiriou. Ganfit: Generative adversarial network fitting for high fidelity 3d face reconstruction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1155–1164, 2019.

[8] Kyle Genova, Forrester Cole, Aaron Maschinot, Aaron Sarna, Daniel Vlasic, and William T Freeman. Unsupervised training for 3d morphable model regression. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 8377–8386, 2018.

[9] Jianzhu Guo, Xiangyu Zhu, Yang Yang, Fan Yang, Zhen Lei, and Stan Z Li. Towards fast, accurate and stable 3d dense face alignment. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XIX 16*, pages 152–168. Springer, 2020.

[10] Andrew Hou, Ze Zhang, Michel Sarkis, Ning Bi, Yiying Tong, and Xiaoming Liu. Towards high fidelity face relighting with realistic shadows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14719–14728, 2021.

[11] Aaron S Jackson, Adrian Bulat, Vasileios Argyriou, and Georgios Tzimiropoulos. Large pose 3d face reconstruction from a single image via direct volumetric cnn regression. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1031–1039, 2017.

[12] Alexandros Lattas, Stylianos Moschoglou, Baris Gecer, Stylianos Ploumpis, Vasileios Triantafyllou, Abhijeet Ghosh, and Stefanos Zafeiriou. Avatarme: Realistically renderable 3d facial reconstruction" in-the-wild". In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 760–769, 2020.

[13] Jiangke Lin, Yi Yuan, and Zhengxia Zou. Meingame: Create a game character face from a single portrait. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 2021.

[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.

[15] Marcel Piotraschke and Volker Blanz. Automated 3d face reconstruction from multiple images using quality measures. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3418–3427, 2016.

[16] Ayush Tewari, Michael Zollhöfer, Pablo Garrido, Florian Bernard, Hyeongwoo Kim, Patrick Pérez, and Christian Theobalt. Self-supervised multi-level face model learning for monocular reconstruction at over 250 hz. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2549–2559, 2018.

[17] Anh Tuan Tran, Tal Hassner, Iacopo Masi, Eran Paz, Yuval Nirkin, and Gérard G Medioni. Extreme 3d face reconstruction: Seeing through occlusions. In *CVPR*, pages 3935–3944, 2018.

[18] Luan Tran, Feng Liu, and Xiaoming Liu. Towards high-fidelity nonlinear 3d face morphable model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1126–1135, 2019.

[19] Anh Tuan Tran, Tal Hassner, Iacopo Masi, and Gérard Medioni. Regressing robust and discriminative 3d morphable models with a very deep neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5163–5172, 2017.

[20] 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.

[21] Zhenyu Zhang, Yanhao Ge, Renwang Chen, Ying Tai, Yan Yan, Jian Yang, Chengjie Wang, Jilin Li, and Feiyue Huang. Learning to aggregate and personalize 3d face from in-the-wild photo collection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14214–14224, 2021.