

# FACE RECOGNITION ON POINT CLOUD WITH CGAN-TOP FOR DENOISING

*Junyu Liu\**

*Jianfeng Ren\*†*

*Hongliang Sun\**

*Xudong Jiang‡*

\*School of Computer Science, University of Nottingham Ningbo China, China

†Nottingham Ningbo China Beacons of Excellence Research and Innovation Institute,  
University of Nottingham Ningbo China, China

‡School of Electrical & Electronic Engineering, Nanyang Technological University

## ABSTRACT

Face recognition using 3D point clouds is gaining growing interest, while raw point clouds often contain a significant amount of noise due to imperfect sensors. In this paper, an end-to-end 3D face recognition on a noisy point cloud is proposed, which synergistically integrates the denoising and recognition modules. Specifically, a Conditional Generative Adversarial Network on Three Orthogonal Planes (cGAN-TOP) is designed to effectively remove the noise in the point cloud, and recover the underlying features for subsequent recognition. A Linked Dynamic Graph Convolutional Neural Network (LDGCNN) is then adapted to recognize faces from the processed point cloud, which hierarchically links both the local point features and neighboring features of multiple scales. The proposed method is validated on the Bosphorus dataset. It significantly improves the recognition accuracy under all noise settings, with a maximum gain of 14.81%.

**Index Terms**— Conditional Generative Neural Network on Three Orthogonal Planes, Linked Dynamic Graph Convolutional Neural Network, 3D Point Cloud, Face Recognition

## 1. INTRODUCTION

3D face recognition becomes increasing popular, as 3D data provide more information and are less affected by illumination [1–5]. Among all 3D models, point clouds are most commonly used [6]. Early methods for 3D face recognition often manually specify geometric properties for feature extraction [7]. 2D deep learning methods [8–10] project point clouds as depth maps and use 2D networks for recognition, which do not fully utilize the 3D information [11]. 3D geometric-based methods directly process the sparse point sets using Graph Convolutional Networks [12]. Most recent 3D face recognition models [1,2,13] utilize PointNet [12] or PointNet++ [14], in which sparse convolution kernels are directly applied on

each point. But these methods overemphasize the point features, and may not fully utilize the global information. Furthermore, most of them assume clean 3D face point clouds, limiting their practical usability.

To handle the noisy raw data in 3D face recognition, researchers attempted to directly learn a model from noisy data [15, 16], but these models can't flexibly handle point clouds with different types/levels of noise. Graph Convolutional networks [12, 14, 17] have been utilized to denoise regular objects, e.g., chairs, cars, and bottles, but they perform poorly on complex 3D face point clouds. Clean face models are usually thin surface-like models with one-point thickness, but the noisy face point clouds may drastically disturb the surface, challenging the design of networks. More importantly, we observe that most denoising methods only focus on producing visually clean point clouds and most recognition methods use mostly clean point clouds, leaving a loophole of an end-to-end face recognition method on the raw point clouds.

To tackle these challenges, an end-to-end framework for face recognition with denoising on 3D point clouds is proposed in this paper. Specifically, a Conditional Generative Adversarial Network on Three Orthogonal Planes (cGAN-TOP) is designed to reduce the noise in the point cloud, where in each plane, one of the 3D coordinates is treated as the feature (similar to the intensity level in an image) and the other two are kept as the coordinates. A cGAN [18] is designed for each plane to remove the noise in the feature (one of the coordinates) with the help of the other two coordinates. The three denoised coordinates are then combined to form the denoised point cloud. Finally, a Linked Dynamic Graph Convolutional Neural Network (LDGCNN) is adapted to comprehensively utilize both local and global point features for recognition.

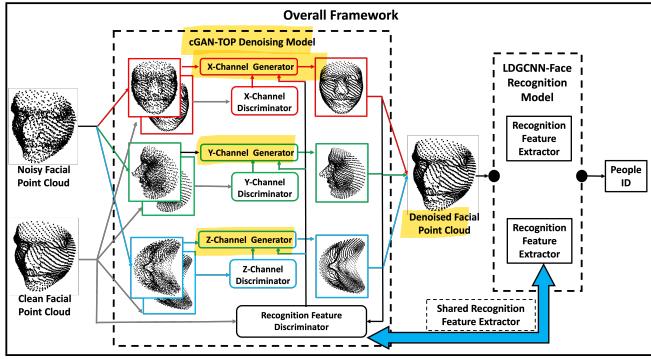
Our contributions are three-fold: 1) An end-to-end model for face recognition on noisy point clouds is proposed, which integrates the cGAN-TOP denoising model and the LDGCNN recognition model. 2) The proposed cGAN-TOP removes the noise in the point cloud targeted at a better subsequent recognition. 3) The proposed method significantly outperforms state-of-the-art models under challenging conditions on the Bosphorus dataset [19], with a maximum gain of 14.81%.

This work was supported in part by National Natural Science Foundation of China under Grant 72071116 and 71901115, and in part by Ningbo Municipal Bureau Science and Technology under Grant 2019B10026 and 2022Z173. Corresponding author: jianfeng.ren@nottingham.edu.cn.

## 2. PROPOSED METHOD

### 2.1. Overview of Proposed Method

Point clouds are sparse and unstructured [12] (unlike densely structured models such as 2D images), enabling them to represent complex models flexibly. But recognition using 3D point clouds faces two main challenges. 1) The sparsity poses a challenge to point cloud processing since traditional 2D and 3D CNNs can only process structured models like images and volumetric models [6]. 2) Raw point clouds often contain noise due to imperfect sensors [16], which is significantly different from the noise in images, depth maps, or other volumetric models. The 3D coordinates of a point cloud serve as both features and spatial coordinates, while other modalities have a distinct difference between features and spatial coordinates, *e.g.*, the spatial coordinates are commonly assumed noise-free while features may contain noise [20–22].



**Fig. 1.** Overview of the proposed end-to-end framework for 3D face point cloud denoising and recognition.

To tackle these challenges, the proposed method integrates noise removal and face recognition into one end-to-end framework. As shown in Fig. 1, the proposed model consists of a cGAN-TOP for denoising and an LDGCNN for recognition. To remove the noise in the point cloud, its 3D coordinates are mapped into three planes, *i.e.*,  $x - y$ ,  $y - z$  and  $z - x$  planes. In each of the planes, the point coordinate orthogonal to the plane is treated as the gray value, while the other two remain as the point spatial coordinates in the plane. The 3D point cloud is hence mapped into images of three orthogonal planes. Then a cGAN [18] is designed for image denoising from each of the planes. The denoised gray values in the three planes are then combined to form the denoised point cloud. We then adapt a latest general point cloud object recognition model LDGCNN [6] for 3D face recognition, which comprehensively extracts both global and local features using a Graph Neural Network.

### 2.2. cGAN-TOP denoising Model

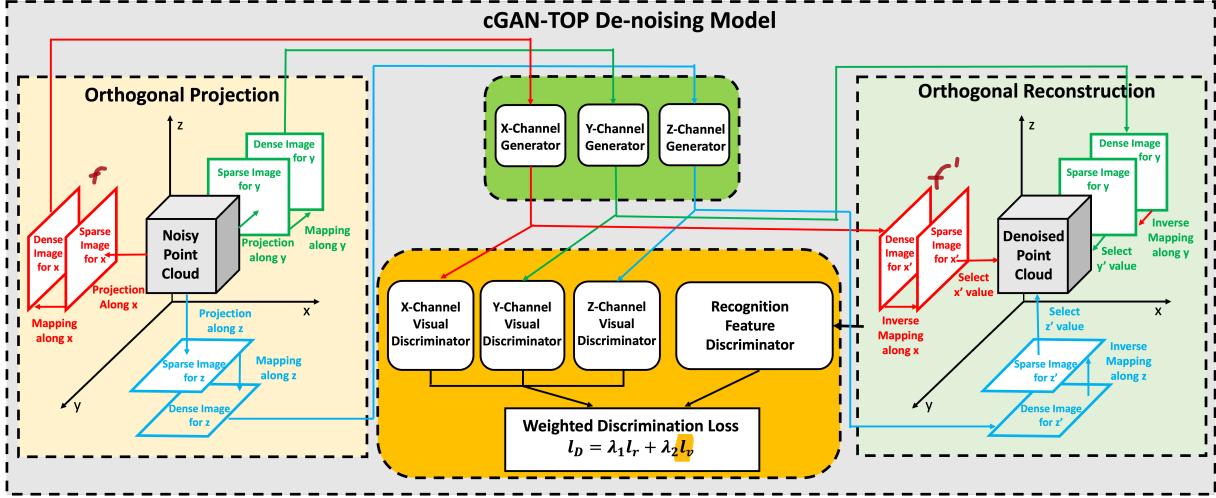
The proposed cGAN-TOP is shown in Fig. 2, which consists of three sets of generators and discriminators, one set for each of the three orthogonal planes. The motivations behind this design are two-fold: 1) Traditional 2D and 3D CNNs [23–25] can't directly handle sparse data structures like point clouds. 2) 3D Graph Convolutional Networks can't effectively process noisy 3D face models due to the thin surface-like characteristics of face point clouds. Therefore, we apply three orthogonal projections to decouple  $(x, y, z)$  values of a point cloud into three orthogonal planes and take the coordinate orthogonal to the plane as the gray value of the point in the 2D plane. Take projection along z-axis as an example, the 3D point cloud is first projected as a sparse image, where each pixel is represented by a gray value  $z$  with an associated coordinate  $(x, y)$ . We then map the coordinates of this sparse image into a structured image, *i.e.*,  $f_z : (x, y) \rightarrow (x', y')$ . After processing by cGAN, the three denoised images are mapped back to the sparse images with the inverse function  $f'$ , and the denoised coordinates are combined back to a point cloud.

The proposed cGAN-TOP contains three generators, one for each plane as shown in Fig. 3. Each generator is a UNet-like network [26] consisting of an encoder network and a decoder network. The encoder follows the general convolutional neural network [23] with repeated convolution-pooling units. The decoder uses symmetric processing with upsampling or de-convolution layers. The convolution kernels are of size  $5 \times 5$  and each convolution is followed by a rectified linear unit (ReLU) layer and a max-pooling layer for downsampling. The final image is generated using a convolution layer.

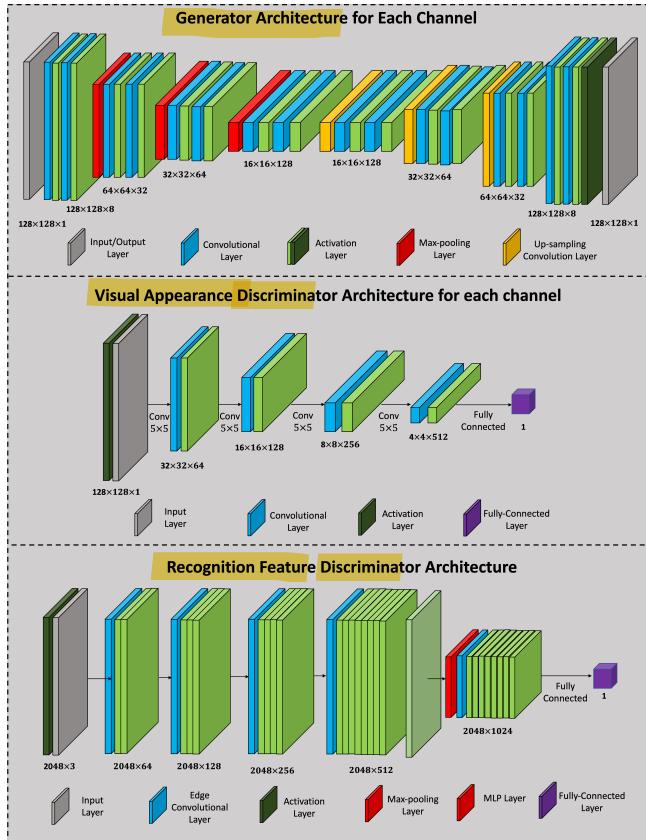
The discriminator network is designed to achieve the following. 1) The denoised point cloud should enable accurate face recognition. 2) The denoised point cloud should be visually and geometrically clean. Hence, two kinds of discriminators are designed. Each of the three Visual Appearance Discriminator (VAD) is designed similar to DCGAN [27] as shown in Fig. 3, which evaluates the visual cleanliness of the generated point cloud from the generator using convolution operations and returns the loss  $l_v$  for the VAD. Moreover, a Recognition Feature Discriminator (RFD) shown in Fig. 3 is designed to evaluate the similarity of the linked features with the clean clouds with a loss of  $l_r$ , to ensure that the denoised point cloud is suitable for face recognition. The final loss  $l_D$  for the discriminator is weighted by hyper-parameters  $\lambda_1$  and  $\lambda_2$  as  $l_D = \lambda_1 l_r + \lambda_2 l_v$ , which are tuned as  $\lambda_1 = 0.67$  and  $\lambda_2 = 0.33$  to achieve the optimal performance.

### 2.3. LDGCNN Face Recognition Model

An LDGCNN model [6] is adapted for face recognition, which hierarchically links both the local features and neighboring features of multiple scales in an elegant and high performance manner. As shown in Fig. 4, the Edge Convolution applies KNN and multi-layer perception with sharing

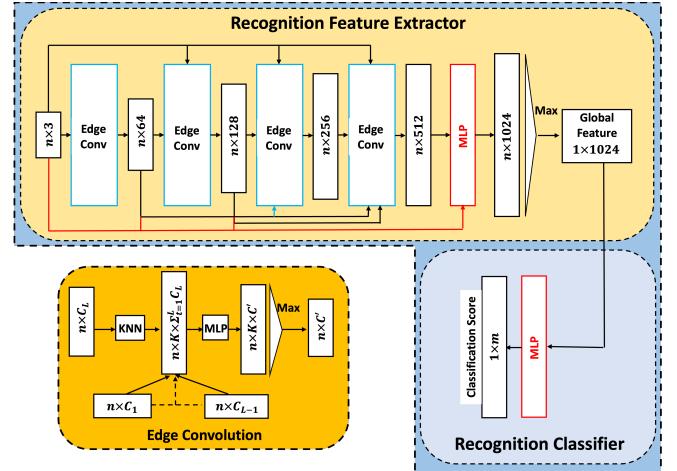


**Fig. 2.** Block diagram of cGAN-TOP for denoising. The point cloud is first projected into three orthogonal planes with the coordinate orthogonal to the plane as its gray value, and then processed by a cGAN in each plane. Three generators and discriminators are designed to generate denoised images. Each discriminator evaluates the visual cleanliness of  $x, y, z$  value respectively. The denoised point cloud is then reconstructed from these three planes. The Recognition Feature Discriminator evaluates the discriminant power of the generated point cloud for recognition.



**Fig. 3.** The structure of Generator and Visual Appearance Discriminator of each channel, and Recognition Feature Discriminator in cGAN-TOP.

parameters to extract the local and global features of different scales. The shortcuts between different layers are added to link the hierarchical features to derive the informative edge vectors. Slightly different from the original design, we modify the number of channels in a gradually increasing manner to better explore the features.



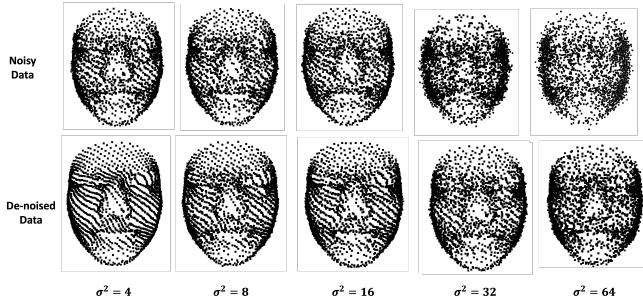
**Fig. 4.** LDGCNN face recognition model.

### 3. EXPERIMENTAL RESULTS

#### 3.1. Experimental Settings

The proposed method is evaluated on the Bosphorus dataset [19], a commonly used 3D face point cloud dataset. It con-

tains 45,000 3D faces of 150 individuals, including variations in expression, occlusion, and pose. We use two common settings [15]. 1) **Neutral Setting**: Faces with a neutral expression (60% of the dataset) are selected as the training data and the rest are test data. 2) **Random Setting**: The training data are randomly sampled from all expressions (60% of the dataset) and the rest for testing. We inject 5 different levels of Gaussian noise into point clouds with  $\sigma^2 = \{4, 8, 16, 32, 64\}$ , following the same experimental settings used in [16] and [15]. Some sample noisy point clouds are shown in Fig. 5.



**Fig. 5.** Sample noisy point clouds and denoised point clouds for  $\sigma^2 = \{4, 8, 16, 32, 64\}$ , respectively.

### 3.2. Comparisons with State-of-the-art Methods

We first compare the proposed method with state-of-the-art 3D face recognition models such as Led3D [16] and FER [15]. Table 1 shows the average recognition accuracy on noisy point clouds under both **Neutral Setting** and **Random Setting**. For all noise levels, the proposed model significantly outperforms the previous methods, with a maximum improvement of 14.81% when  $\sigma^2 = 64$  under the **Random Setting**.

**Table 1.** Comparison of recognition rate on noisy point cloud.

$\sigma^2$	Neutral Setting			Random Setting		
	LeD3D	FER	Ours	LeD3D	FER	Ours
4	95.08	96.88	<b>98.74</b>	90.07	94.87	<b>96.13</b>
8	92.87	95.42	<b>97.78</b>	84.33	90.85	<b>95.53</b>
16	87.54	91.72	<b>93.44</b>	72.77	81.09	<b>88.33</b>
32	74.66	82.33	<b>87.57</b>	50.51	61.32	<b>71.26</b>
64	55.37	60.37	<b>72.04</b>	28.73	36.80	<b>51.61</b>

Table 2 shows the comparisons of visual cleanliness with state-of-the-art denoising models, PCNet [28], DMR [29], and SB [30], where Chamfer distance (CD) and Point-to-Mesh distance (P2M) are used as evaluation metrics. The proposed method shows advantages at a high noise level and comparable performance to SB [30] at a low noise level.

To better understand the source of improvements, we conduct an ablation study on the two discriminators of the proposed cGAN-TOP: 1) Use Recognition Feature Discriminator

**Table 2.** Comparisons of visual cleanliness in terms of Chamfer distance (CD) and Point-to-Mesh distance (P2M).

$\sigma^2$	Metric	PCNet [28]	DMR [29]	SB [30]	Ours
4	CD	3.73	4.64	<b>2.87</b>	2.89
	P2M	1.49	1.83	0.46	<b>0.48</b>
8	CD	7.86	5.12	3.87	<b>3.82</b>
	P2M	4.16	2.35	1.23	<b>1.18</b>
16	CD	13.81	6.94	5.02	<b>4.83</b>
	P2M	6.19	3.27	2.14	<b>1.86</b>
32	CD	24.93	14.56	8.75	<b>6.83</b>
	P2M	12.37	7.25	4.28	<b>3.24</b>
64	CD	42.87	24.91	18.55	<b>13.26</b>
	P2M	20.35	12.11	9.67	<b>6.28</b>

only; 2) Use Visual Feature Discriminator only; 3) Use both discriminators (Full). The results shown in Table 3 indicate that: 1) Use Visual Feature Discriminator only significantly reduces the recognition accuracy as there lacks a supervision for recognition in this case; and 2) Use Recognition Feature Discriminator only also reduce the recognition accuracy since purely targeting at recognition ignores the noise in the point cloud. Thus, we conclude that the proposed method could simultaneously achieve high recognition accuracy on noisy point cloud and produce visually clean point clouds.

**Table 3.** Ablation studies of cGAN-TOP using different discriminators in terms of recognition rate.

$\sigma^2$	Neutral Setting			Random Setting		
	VFD	LRFD	Full	VFD	LRFD	Full
4	86.76	92.88	<b>98.74</b>	84.32	90.71	<b>96.13</b>
8	78.63	90.42	<b>97.78</b>	72.43	87.29	<b>95.53</b>
16	72.67	86.72	<b>93.44</b>	69.72	84.34	<b>88.33</b>
32	64.21	82.33	<b>87.57</b>	62.39	70.01	<b>71.26</b>
64	53.34	62.37	<b>72.04</b>	46.51	49.42	<b>51.61</b>

### 4. CONCLUSION

In this paper, we propose an end-to-end 3D face point cloud denoising and recognition model to provide both high recognition accuracy and visually clean point clouds. The proposed cGAN-TOP provides an efficient and effective way to directly remove the noise in the 3D coordinates of the point cloud by using cGAN on each of the three orthogonal planes. On each plan, the proposed cGAN-TOP strategically treats one of the coordinates as the feature and denoises it using cGAN. Two discriminators are designed to ensure both the high recognition rate and cleanliness of the point cloud. An LDGCNN is adapted for 3D face recognition. The proposed method is compared with state-of-the-art methods on the Bosphorus dataset. The proposed method significantly outperforms the compared methods in most of the settings.

## 5. REFERENCES

- [1] S. Gilani and A. Mian, "Learning from millions of 3D scans for large-scale 3D face recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1896–1905.
- [2] C. Jiang, S. Lin, W. Chen, F. Liu, and L. Shen, "PointFace: Point set based feature learning for 3D face recognition," in *International IEEE Joint Conference on Biometrics*, 2021, pp. 1–8.
- [3] J. Liu, H. Ding, A. Shahroudy, L. Duan, X. Jiang, G. Wang, and A. Kot, "Feature boosting network for 3D pose estimation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 42, no. 2, pp. 494–501, 2019.
- [4] X. Wang, X. Jiang, and J. Ren, "Blood vessel segmentation from fundus image by a cascade classification framework," *Pattern Recognition*, vol. 88, pp. 331–341, 2019.
- [5] Z. Xiong, J. Wang, W. Jin, J. Liu, Y. Duan, Z. Song, and X. Peng, "Face2statistics: user-friendly, low-cost and effective alternative to in-vehicle sensors/monitors for drivers," in *International Conference on Human-Computer Interaction*, 2022.
- [6] K. Zhang, M. Hao, J. Wang, C. W. de Silva, and C. Fu, "Linked dynamic graph CNN: learning on point cloud via linking hierarchical features," *CoRR*, vol. abs/1904.10014, 2019.
- [7] S. J. Belongie, J. Malik, and J. Puzicha, "Shape matching and object recognition using shape contexts," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 4, pp. 509–522, 2002.
- [8] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao, "3D ShapeNets: A deep representation for volumetric shapes," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1912–1920.
- [9] M. Yavartanoo, E. Kim, and K. M. Lee, "SPNet: Deep 3D object classification and retrieval using stereographic projection," in *Asian Conference on Computer Vision*, 2018, pp. 691–706.
- [10] S. Chen, L. Zheng, Y. Zhang, Z. Sun, and K. Xu, "VERAM: view-enhanced recurrent attention model for 3D shape classification," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 12, pp. 3244–3257, 2019.
- [11] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, "Dynamic graph CNN for learning on point clouds," *CoRR*, vol. abs/1801.07829, 2018.
- [12] A. Garcia-Garcia, F. Gomez-Donoso, J. G. Rodríguez, S. Orts-Escalano, M. Cazorla, and J. Azorín López, "PointNet: A 3D convolutional neural network for real-time object class recognition," in *International Joint Conference on Neural Networks*, 2016, pp. 1578–1584.
- [13] Z. Zhang, F. Da, and Y. Yu, "Learning directly from synthetic point clouds for "in-the-wild" 3D face recognition," *Pattern Recognit.*, vol. 123, pp. 108394, 2022.
- [14] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, "Pointnet++: Deep hierarchical feature learning on point sets in a metric space," in *Advances in Neural Information Processing Systems*, 2017, pp. 5099–5108.
- [15] Z. Zhang, C. Yu, S. Xu, and H. Li, "Learning flexibly distributional representation for low-quality 3D face recognition," in *AAAI Conference on Artificial Intelligence*, 2021, pp. 3465–3473.
- [16] G. Mu, D. Huang, G. Hu, J. Sun, and Y. Wang, "Led3D: A lightweight and efficient deep approach to recognizing low-quality 3D faces," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5773–5782.
- [17] Y. Wang, S. Wu, H. Huang, D. Cohen-Or, and O. Sorkine-Hornung, "Patch-based progressive 3D point set upsampling," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 5958–5967.
- [18] F. Yang, J. Ren, Z. Lu, J. Zhang, and Q. Zhang, "Rain-component-aware capsule-GAN for single image de-raining," *Pattern Recognition*, vol. 123, pp. 108377, 2022.
- [19] A. Savran, N. Alyüz, H. Dibeklioglu, O. Çeliktutan, B. Gökberk, B. Sankur, and L. Akarun, "Bosphorus database for 3D face analysis," in *Biometrics and Identity Management*, 2008, vol. 5372, pp. 47–56.
- [20] J. Ren, X. Jiang, and J. Yuan, "Noise-resistant local binary pattern with an embedded error-correction mechanism," *IEEE Transactions on Image Processing*, vol. 22, no. 10, pp. 4049–4060, 2013.
- [21] H. Chen, Z. Wei, X. Li, Y. Xu, M. Wei, and J. Wang, "Repcloudnet: Feature-aware recurrent point cloud denoising network," *International Journal of Computer Vision*, vol. 130, no. 3, pp. 615–629, 2022.
- [22] C. Shi, C. Wang, X. Liu, S. Sun, B. Xiao, X. Li, and G. Li, "Three-dimensional point cloud denoising via a gravitational feature function," *Applied Optics*, vol. 61, no. 6, pp. 1331–1343, 2022.
- [23] J. Zhang, Q. Zhang, J. Ren, Y. Zhao, and J. Liu, "Spatial-context-aware deep neural network for multi-class image classification," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 1960–1964.
- [24] S. Chen, W. He, J. Ren, and X. Jiang, "Attention-based dual-stream vision transformer for radar gait recognition," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2022, pp. 3668–3672.
- [25] W. He, J. Zhang, J. Ren, R. Bai, and X. Jiang, "Hierarchical convit with attention-based relational reasoner for visual analogical reasoning," in *AAAI Conference on Artificial Intelligence*, 2023.
- [26] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention*, 2015, vol. 9351, pp. 234–241.
- [27] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.
- [28] M. Rakotosaona, V. La Barbera, P. Guerrero, N. J. Mitra, and M. Ovsjanikov, "Pointcleannet: Learning to denoise and remove outliers from dense point clouds," in *Computer Graphics Forum*, 2020, vol. 39, pp. 185–203.
- [29] S. Luo and W. Hu, "Differentiable manifold reconstruction for point cloud denoising," in *ACM international conference on multimedia*, 2020, pp. 1330–1338.
- [30] S. Luo and W. Hu, "Score-based point cloud denoising," in *IEEE International Conference on Computer Vision*, 2021, pp. 4583–4592.