Quality Judgment of 3D Face Point Cloud Based on Feature Fusion

GONG GAO, HONG LIU, AND HONGYU YANG
National Key Laboratory of Fundamental Science on Synthetic Vision, College of Computer Science, Sichuan University, Chengdu 610000, China
Corresponding author: Hongyu Yang (yanghongyu@scu.edu.cn)

This work was supported in part by the Sichuan Scientific and Technological Transformative Project under Grant 2019ZDZX0039.

ABSTRACT With the rapid development of face recognition, the 3D face has gradually become the mainstream, 3D face point cloud quality judgment was an important process. A Feature Fusion Network (FFN) was proposed to judge 3D face point cloud quality acquired by binocular CCD camera. Firstly, the 3D point cloud was preprocessed to cut out the face area, and the image obtained from the point cloud and the corresponding 2D plane depth map projection was used as the input. Secondly, Dynamic Graph Convolutional Neural Network (DGCNN) was trained for point cloud learning and ShuffleNet was trained for image learning. Then, the middle layer features of the two network modules were extracted and concat to fine-tune the whole network. Finally, three fully connected layers were used to realize the five-class classification of the 3D face point cloud (excellent ordinary, stripe, burr, deformation). The proposed FFN achieved the classification accuracy of 83.7%, which was 5.8% higher than that of ShuffleNet and 2.2% higher than that of DGCNN. The experimental results show that concat depth map features and point cloud features can achieve the complementary effect between different features.

INDEX TERMS Face point cloud, point cloud feature, depth map projection, weighted fusion, quality judgment.

I. INTRODUCTION
The point cloud data collected by 3D structured light may have some defects, such as stripes, burrs, and deformation [1]. 3D reconstruction system must be controlled within a certain range for the defect of the human face point cloud, otherwise, it will cause the model matching is not accurate and can not work properly, so the quality judgment of 3D face point cloud is of great significance to minimize the error of 3D reconstruction system.

Because the face is a kind of non-rigid body, different from other objects with the complex three-dimensional surface, the acquisition of face surface data will often appear as shadow, occlusion, or stripes: shadow, in the light, there will be shadows in the lower part of the face nose and chin, but the shadow part of the data can not be measured; occlusion, the face is often obscured by hair and collar parts, which will cause incomplete point cloud data of the face. When stripes, charge-coupled devices (CCD) scan, the adjustment parameters are too small, which will cause the view to be dark, accompanied by incomplete data scanning, too large parameters, and reverse.

The light phenomenon is obvious and the parameters are not adjusted properly, which will lead to the color stripe pattern [2].

II. RELATED WORK
The three-dimensional point cloud quality measurement can be divided into two categories: the point-based metrics and the depth map projection-based metric. The point-based measurement extracts the corresponding relationship between the original point cloud and the template point cloud, while the depth map projection-based metric maps the 3D point cloud to the more classical 2D plane. The point-to-point index of point cloud quality judgment [3] uses root-mean-square difference or Hausdorff distance to estimate the geometric error between the original point and the processed point cloud; the point-to-plane index [4] depends on the distance between a point and the similarity between the tangent face point.
Lee et al. [5] proposed a quality evaluation method based on the significant sum of visible vertices after projection. Based on the visual perception characteristics of human beings, the point cloud on the grid is weighted by gaussian curvature, and then the important areas on the grid are described. On this basis, Feixas et al. [6] extended the evaluation standard of visual point entropy and put forward the outline of viewpoint mutual information (VMI) from the point of view of information theory.

Lavoué et al. [7] proposed the concept of mesh structure distortion (MSDM) by modeling 2D images as 3D data, drawing on the principle of structural similarity of 2D images, and used MSDM to evaluate the distortion of 3D model digital watermarking algorithms with good results. Meynet et al. [8] proposed a point cloud geometric quality measure based on structural similarity based on local curvature statistics, which calculates the curvature of each point and compares the differences of the corresponding points. Javaheri et al. [9] proposed a depth map projection-based measurement method, which takes into account both geometric shape and color by using two-dimensional image quality indexes such as peak signal-to-noise ratio (PSNR), and structural similarity (SSIM) or video quality metric (VQM).

Some scholars propose PointNet which is based on the characteristics of the point cloud and combined with the existing deep learning methods. They explore the point cloud data as an input for feature extraction as one of the modules to achieve the human face point cloud quality judgment task. In 2017, Qi et al. [10] proposed PointNet network using 3D point cloud as input combined with deep learning method, which solved the disorder of point cloud and the invariance of rigid body transformation. In the same year, Qi et al. [11] proposed PointNet++ based on PointNet and combined with the network mode of 2D CNN, using sampling, grouping, and PointNet modules, and the classification accuracy was improved. In 2019, Liu et al. [12] proposed the relation shape convolution neural network (RS-CNN), which extends regular grid CNN to irregular point clouds. In the same year, Wang et al. [13] proposed a new PointNet++-based point cloud feature extraction network module, EdgeConv, which can capture local geometric information while ensuring invariance of point cloud rotation to achieve the highest classification accuracy on multiple point cloud datasets.

The existing 3D face cloud quality judgment only uses point cloud data or projected depth map data, missing two-dimensional or three-dimensional features, the accuracy is not high. ShuffleNet and DGCNN are used to extract image features and point cloud features, respectively. Two-dimensional features and three-dimensional features complement each other. This feature fusion network (FFN) improves the accuracy of quality judgment.

## III. 3D POINT CLOUDS CUT FACIAL AREAS AND CALCULATE CURVATURE

The 3D face point cloud is the most direct and primitive data information captured by the CCD camera. To reduce the influence of redundant data such as ear, neck, hair, and so on, it is necessary to cut out the facial area based on the tip of the nose which serves as the center of the ball. Because of the defects of the original data, such as point cloud noise, lack of sampling, cusp and hole, deformation, and so on [14], judging the quality of the 3D face point cloud plays an important role further.

### A. 3D HUMAN FACE DOTS CUT FACIAL AREAS

The point cloud information only includes three-dimensional coordinates (X, Y, Z). After using the improved nasal cusp location algorithm to locate the nasal cusp, the facial area was cut out with the nasal apex as the center of the ball and R=100 mm. The different quality of the face point cloud facial area is shown in Figure 1.

![Figure 1. Different quality 3D face point cloud.](image)

### B. CALCULATION OF GAUSSIAN CURVATURE OF 3D FACE

After a surface with infinite number of orthogonal curvatures at some point, there exists a curve such that the curvature of this curve is extreme, this curvature is the extreme value K_{max} and the curvature perpendicular to the surface of extreme curvature is the minimal value K_{min}. These two curvatures are the main curvature. The product of two curvatures is gaussian curvature, which reflects the total bending at a point [16].

For 3D face point cloud, the point cloud of K nearest neighbor around each point cloud is approximated to a local plane. In this paper, k=15 is selected, and then the implicit surface equation is obtained by least square fitting $F(x, y, z)$, then the height of p $(x, y, z)$ point is obtained.

The formula for calculating the curvature K is as follows:

$$K = \frac{\nabla F \times H^*(F) \times \nabla F^T}{|\nabla F|^4}$$

where: $\nabla F$ is the gradient corresponding to the implicit equation in x, y, z, $H(F)$ is a square matrix composed of the second-order partial derivatives of a multivariate function, the Hessian matrix describing the local curvature of the function, $H^*(F)$ is the adjoint matrix of the corresponding $H(F)$, and both gaussian and mean curvatures are computed from $\nabla F$, $H(F)$, and $H^*(F)$.

### C. FACE RECOGNITION AND FACE MODELING

The nasal cusp is located in the 3D face point cloud, and the facial area is intercepted with the nose tip as the center of the ball and R=100 mm as the radius. The number of the facial point cloud is 80000~90000 [15]. The 3D point cloud data is sampled by the curvature sampling of the farthest fusion (CSFF) method as the input of the dynamic graph convolutional neural network (DGCNN) module. The point...
cloud rotates at a certain angle to generate two-dimensional images as the input of the ShuffleNet module.

Where \( F \) is the gradient of the implicit equation corresponding to \( x, y, z \), \( H(F) \) is a square matrix composed of the second-order partial derivative of a multivariate function, the hessian matrix of the local curvature of the function is described. \( H^*(F) \) is the adjoint matrix of the corresponding \( H(F) \), and the gaussian curvature and the average curvature are calculated by \( \leq F, H(F) \) and \( H^*(F) \).

D. POINT CLOUDS PROJECTED INTO A TWO-DIMENSIONAL PLANE

As shown in figure 2(b), it is not easy to observe stripes, burrs and other problems from the depth map generated on the front. Only after rotating the point cloud at a certain angle can it be observed more clearly, and obvious stripes can be seen around the nose and the left side of the face, as shown in figure 2(a) and 2(c). Multiple angles of rotating positive face cloud are used to generate two-dimensional images, which can be better used for feature extraction.

Use 3D face point clouds to rotate on the x-axis and y-axis, respectively 2.5° and −2.5°, each three-dimensional point cloud produces 5 projected two-dimensional images, as shown in Figure 3.

IV. FEATURE EXTRACTION AND FUSION

Concate feature is one of the simplest and most widely used methods, which is to extract information from different data sources [17].

A. THREE-DIMENSIONAL FACE CLOUD QUALITY JUDGEMENT PROCESS

The process of 3D point cloud quality judgment is shown in Figure 4. The collected 3D face point cloud is cut out of the facial region, and then the 3D point cloud data and two-dimensional image data are trained in DGCNN and ShuffleNet v2 [18], then the trained model is extracted and feature fusion is carried out, and then the two network modules are fine-tuned. Finally, the trained model is tested with the testset data to obtain quality classification results, thus completing the quality judgment task.

B. FEATURE FUSION MODE

Inspired by ShuffleNet v2 and DGCNN, we propose a feature fusion network (FFN) for better recognition of 3D point cloud quality judgment tasks. Firstly, ShuffleNet is used to extract two-dimensional features, and DGCNN is used to extract three-dimensional features. Both networks obtain 1024 dimensional features and concate features, and then use a three-layer full connection layer to achieve classification, get category score, and achieve quality judgment task [19].

As shown in Figure 5, the DGCNN module network input number is \( N \), the coordinate dimension is 3, that is, the point cloud size is \( N \times 3 \). The DGCNN module uses the furthest point sampling method of fusion curvature to sample 1024 points, and the output is 1024-dimensional features. After the training of the DGCNN module, the ShuffleNet v2 module carries on two-dimensional feature extraction. The input of ShuffleNet is 224×224 two-dimensional images, and the output is 1024-dimension feature. In the feature fusion stage, we use feature aggregation, that is, the fusion of point
cloud features and 2D image features.

\[ g^F (x^i, \theta^F) = \text{concat} \left( g^F (x^i, \theta^P), g^F (x^i, \theta^l) \right) \]  \hspace{1cm} (2)

where \( x^i \) represents the first sample of the dataset and \( \theta^F \).

V. IMPROVED SAMPLING MODE

In the three-dimensional face point cloud, the feature points selected by the iterative farthest point sampling algorithm often get the edge outline or noise point of the face, which is useless for the face, and only a few of them are located in the nose, eyes and other important feature areas [20]. Based on the farthest point sampling method, the sampling-based on geometric features is integrated, which makes the extraction of face features better.

For given input face point cloud \( P \), use \( P = \{ P_i \}, \}

\( i = 1, 2, \ldots, n \) indicates that the set of points sampled by iterative furthest points can be expressed as \( G = \{ G_j \}, \}

\( j = 1, 2, \ldots, k \) performance. Assuming that the current sampling is \( j \) point, the next sampling point should be the \( m \) point farthest from the \( j \) point (\( m = G_k, 1 \leq m \leq k \)), and the point \( m \) should be full.

The following requirements are met:

\[ d (k, P) = \max_{j \in P} (d (j, P)) = \max_{j \in P} \left( \min_{1 \leq \alpha \leq n} (d (j, P)) \right) \]  \hspace{1cm} (3)

where \( d (j, P) \) represents the Euclidean distance between two points. It can be seen from formula (5) that the furthest point sampling mostly samples the edge points or noise points of the face, so the fusion of Euclidean distance and face curvature can sample the facial organs of the face. The formula \( d = d (k, P) + \alpha c_m \) is obtained, where \( c_m \) is the \( m \) curvature of point \( m \) and \( \alpha \) is the weight of the set equilibrium Euclidean distance and curvature. It can be seen from formula (5) that the distance of the points with large curvature is large and it is easier to be selected. The sampling points are concentrated in the places where the curvature changes greatly, and the sampling is greatly affected by curvature, so \( \alpha \) is used to control the influence of curvature on sampling.

As shown in Figure 6, Figure 6 (a) is the result of sampling as the original iterative farthest point when \( \alpha = 0 \); Figure 6 (b) is the result of combining curvature sampling with \( \alpha \) = 1, the Figure 6 (b) shows that there are too many sampling points in the range of organs; thus, \( \alpha = 0.1 \) is selected to move some sampling points without destroying the uniform sampling of points.

A. LOSS FUNCTION

The common loss function in the classification task is Softmax, which is to map the linear combination of features to (0, 1), and the probability of each classification is given by cross-entropy [21]. In order to measure the performance of the FFN method applied to 3D point cloud quality judgment, the loss function uses the cross-entropy function of Softmax, which includes the loss function of the DGCNN module, ShuffleNet module, and FFN method, and allocates the corresponding superparameters:

\[ \text{Loss} = \beta \text{loss} (\text{DGCNN}) + \delta \text{loss} (\text{ShuffleNet}) + \gamma \text{loss} (\text{FFN}) \]  \hspace{1cm} (4)

\( \beta, \delta, \lambda \) is the hyperparameter of three partial loss functions.
fusion training is carried out, and the learning rate is 0.0001. The superparameter settings of point cloud sampling and loss function include $\alpha, \beta, \delta, \lambda, \alpha = 0.1, \beta = 0.4, \delta = 0.4, \gamma = 0.2$, which are obtained through experiments.

A. DATA
The 3D face point cloud data set that we used can be divided into five categories, which are excellent quality, common quality, burr, striped, and deformed.

1) Excellent quality: all facial organs are clearly visible without any distortion; no obvious streaks, burrs or other defects are found after rotation at a certain angle.
2) Average quality: the entire face is complete, with a few parts missing; rotated to the front, no obvious defects or minor defects are found; rotated at multiple angles: undulations are found on the face.
3) Stripes: the stripes are not obvious and need to be rotated at multiple angles to be seen; the direction of rotation is frontal with slight stripes; rotating multiple angles reveals the stripes.
4) Burr: no deformation on the entire face surface, with burrs in some places. Rotated to the front side, no obvious defects were found; rotated multiple angles, multiple obvious burrs were found on the surface.
5) Deformation: including serious errors in facial features, blurred organs, etc.

Each category of the dataset has 1200 3D face point cloud images, a total of 6000 point cloud images. Each point cloud image is rotated from the front face to generate 5 two-dimensional images and 5 point clouds, with 30000 two-dimensional images and point clouds respectively.

B. RESULT
The quality judgment experiments of PointNet, PointNet++, RS-CNN, DGCNN, ShuffleNet v2 and FFN are carried out on the dataset. For the first time, PointNet uses deep learning to deal with disordered point clouds directly and then outputs global or partial classification tags for each input point. DGCNN as the use of EdgeConv, combined with global features and local features, feature extraction ability is better.

Two-dimensional images are used for quality judgment experiments. 4800 pieces are used for training and 1200 for testing in each category, and 5000 iterations are made. Using three versions of ShuffleNet v2 (0.5x, 1.0x, 1.5x) the network architecture is shown in Table 1, and the classification accuracy of ShuffleNet v2 decreases with the decrease of the number of channels. The experimental results are shown in Figure 8.

As can be seen from Figure 7, ShuffleNet v2 uses point-wise group revolution and channel shuffle to ensure accuracy while reducing network parameters. The accuracy rate of ShuffleNet v2 1.5x is 79.1%. Compared with ShuffleNet v2 1.5x, the accuracy rate of three versions of ShuffleNet v2 decreases by 1.3 percentage points and 7.8 percentage points respectively as the number of channels decreases.

DGCNN is used to extract the feature of point cloud data. Geometric sampling, random sampling, farthest point sampling, and farthest point fusion curvature sampling are used respectively (superparameter is set to $\alpha = 0.1$). 1024 points are trained 160 times to carry out 5 classification tests on 3D face point clouds of different quality. The experimental results are shown in Figure 8.

The random sampling method has the same sampling probability for each point, and the sampling efficiency is high, but the experimental results are unstable. Geometric sampling has a certain anti-noise ability when the curvature of the point cloud is larger. The farthest point sampling method uses Euclidean distance to sample the point cloud profile. The improved sampling method is based on the farthest point sampling combined with geometric sampling, and the point cloud with large facial curvature is sampled. It can be concluded from the experimental results of figure 8 that the classification accuracy of the random sampling method is significantly lower than that of the farthest point sampling.
TABLE 1. Overall architecture of ShuffleNet v2, for three different levels of complexities.

| Layer                  | Input size | Kernel size | Stride | Repeat | Output channels |
|------------------------|------------|-------------|--------|--------|-----------------|
| image                  | 224×224    |             |        |        | 3×3             |
| Convolution layer      |            |             |        |        | 2×2             |
| Maximum pool layer     | 112×112    | 3×3         |        | 1      | 24              |
| 56×56                  | 3×3         | 2           |        | 48     | 24              |
| Stage2                 | 28×28      | 2           |        | 48     | 176             |
| Stage3                 | 14×14      | 1           |        | 768    | 352             |
| Stage4                 | 7×7        | 2           |        | 1024   | 1024            |
| Global pooling layer   | 7×7        | 1×1         |        | 1024   | 1024            |
| Total parameters       | 1×1        | 7×7         |        | 1.4×10⁸| 2.3×10⁸         |

TABLE 2. Classification accuracy of different networks. P denotes point clouds.

| Network              | Input | Test Accuracy/\% |
|----------------------|-------|------------------|
| PointNet             | 1024 P | 76.1             |
| PointNet++           | 1024 P | 78.2             |
| RS-CNN               | 1024 P | 80.4             |
| DGCNN                | 1024 P | 81.9             |
| ShuffleNet v2 1.5x   | Image | 79.1             |
| FFN                  | 1024 P+Image | 83.7             |

method with fusion curvature. The classification accuracy of the furthest point fusion curvature is 81.5%. The farthest point sampling was 78.5%, the geometric sampling method was 74.2%, and the random sampling method was 75.1%. The classification accuracy of the furthest point sampling method with fusion curvature is 5.5% higher than that of the farthest point sampling.

DGCNN module and ShuffleNet v2 are trained respectively, then finetune the entire network, Table 2 is the experimental comparison results.

We use the cutting-edge quality judgment method based on 3D point cloud, and carry out experiments to obtain the classification accuracy. PointNet extracts the global features from the sampled face point cloud, and uses three full connection layers to classify the experimental quality, and obtains 76.1% accuracy. PointNet++ uses sampling, grouping, and feature extraction operations, and extracts global signs and local features at the same time. It has better feature extraction ability for point cloud, and its classification accuracy is 78.2%. RS-CNN designed a convolution operator based on relation learning, extended the regular network CNN to the irregular point cloud, and obtained 80.4% accuracy. DGCNN designed EdgeConv and stacked multiple EdgeConv modules to obtain multi-level and richer semantic features, with a classification accuracy of 81.9%. ShuffleNet 1.5x uses two-dimensional images for classification tasks, and the accuracy is 79.1%. In this paper, we use the FFN method to train DGCNN and ShuffleNet v2 1.5x modules respectively, and then fine-tune the whole network to achieve classification accuracy of 83.7%.

Overall, the accuracy of the FFN method is 5.8% higher than that of ShuffleNet v2 1.5x network and 2.2% higher than that of the DGCNN network. Experiments show that the fusion of three-dimensional features and two-dimensional features realizes feature complementarity, improves the classification accuracy, and further explains the effectiveness of feature fusion network for the quality judgment of three-dimensional point cloud.

VII. CONCLUSION

We propose a point cloud quality evaluation method based on feature fusion. DGCNN module and ShuffleNet v2 module are used to extract 1024 dimensional features respectively, and concat fusion is used to train a network, which improves the accuracy of 3D point cloud judgment. In this paper, the farthest point sampling combined with curvature is improved to solve the problem of too much face edge sampling in 3D face point cloud sampling. The judgment accuracy is obtained under different deep learning network models, which proves the feasibility of feature fusion. The experimental results show that the feature fusion scheme can significantly improve the judgment accuracy of point cloud quality. Since the point cloud contains a lot of information, we will use the point cloud with RGB information to further improve the FFN network.

REFERENCES

[1] Z. Ren, F. Fang, N. Yan, and Y. Wu, “State of the art in defect detection based on machine vision,” Int. J. Precis. Eng. Manuf.-Green Technol., vol. 9, no. 2, pp. 661–691, Mar. 2022.
[2] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan, “A practical approach to 3D scanning in the presence of interreflections, subsurface scattering and defocus,” Int. J. Comput. Vis., vol. 102, nos. 1–3, pp. 33–55, 2012.
[3] M. Gupta, A. Agrawal, A. Veeraraghavan, and S. Narasimhan, “A practical approach to 3D scanning in the presence of interreflections, subsurface scattering and defocus,” Int. J. Comput. Vis., vol. 102, nos. 1–3, pp. 33–55, 2012.
Gong Gao et al.: Quality Judgment of 3D Face Point Cloud Based on Feature Fusion

[4] R. Diniz, P. G. Freitas, and M. C. Q. Farias, “Point cloud quality assessment based on geometry-aware texture descriptors,” Comput. Graph., vol. 103, pp. 31–44, Apr. 2022.

[5] C. H. Lee, A. Varshney, and D. W. Jacobs, “Mesh saliency,” ACM Trans. Graph., vol. 24, no. 3, pp. 659–666, Jul. 2005.

[6] M. Feixas, M. Sbert, and F. González, “A unified information-theoretic framework for viewpoint selection and mesh saliency,” ACM Trans. Appl. Perception, vol. 6, no. 1, pp. 1–23, Feb. 2009.

[7] G. Lavoué, “A multiscale metric for 3D mesh visual quality assessment,” Comput. Graph. Forum, vol. 30, no. 5, pp. 1427–1437, 2011.

[8] G. Meynet, J. Digne, and G. Lavoué, “PC-MSDM: A quality metric for 3D point clouds,” in Proc. 11th Int. Conf. Quality Multimedia Exper. (QoMEX), Jun. 2019, pp. 765–770.

[9] A. Javaheri, C. Brites, F. Pereira, and J. Ascenso, “Subjective and objective quality evaluation of compressed point clouds,” in Proc. IEEE 19th Int. Workshop Multimedia Signal Process. (MMSP), Oct. 2017, pp. 1030–1035.

[10] R. Q. Charles, H. Su, M. Kaichun, and L. J. Guibas, “PointNet: Deep learning on point sets for 3D classification and segmentation,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 652–660.

[11] C. R. Qi, L. Yi, H. Su, and L. J. Guibas, “PointNet++: Deep hierarchical feature learning on point sets in a metric space,” in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 5099–5108.

[12] Y. Liu, B. Fan, S. Xiang, and C. Pan, “Relation-shape convolutional neural network for point cloud analysis,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 8887–8896.

[13] Y. Wang, Y. Sun, Z. Liu, S. E. Sarma, M. M. Bronstein, and J. M. Solomon, “Dynamic graph CNN for learning on point clouds,” ACM Trans. Graph., vol. 38, no. 5, pp. 1–12, Oct. 2019.

[14] S. I. Park and S.-J. Lim, “Template-based reconstruction of surface mesh animation from point cloud animation,” ETRI J., vol. 36, no. 6, pp. 1008–1015, Dec. 2014.

[15] X.-F. Han, J. S. Jin, M.-J. Wang, W. Jiang, L. Gao, and L. Xiao, “A review of algorithms for filtering the 3D point cloud,” Signal Process., Image Commun., vol. 57, pp. 103–112, Sep. 2017.

[16] A. Okano, T. Matsumoto, and T. Kato, “Gaussian curvature entropy for curved surface shape generation,” Entropy, vol. 22, no. 3, p. 353, Mar. 2020.

[17] N. N. Ma, X. Y. Zhang, H.-T. Zheng, and J. Sun, “ShuffleNet V2: Practical guidelines for efficient CNN architecture design,” in Proc. Eur. Conf. Comput. Vis. (ECCV), Sep. 2018, pp. 116–131.

[18] H. Li, P. Xiong, H. Fan, and J. Sun, “DFANet: Deep feature aggregation for real-time semantic segmentation,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 9522–9531.

[19] Q. Wei, “Saliency maps-based convolutional neural networks for facial expression recognition,” IEEE Access, vol. 9, pp. 76224–76234, 2021.

[20] V. Kocaman, O. M. Shir, and T. Back, “Improving model accuracy for balanced image classification tasks by adding a final batch normalization layer: An empirical study,” in Proc. 25th Int. Conf. Pattern Recognit. (ICPR), Jan. 2021, pp. 10404–10411.

Gong Gao was born in 1995. He received the M.S. degree. His research interests include computer vision and point cloud recognition.

Hong Liu received the Ph.D. degree from Sichuan University, Chengdu, China, in 2014. He has been working as a Teacher and a Researcher with Sichuan University, since 2007. He has been an Associate Professor, since 2015. His research interest includes machine learning.

Hongyu Yang received the Ph.D. degree from Sichuan University, Chengdu, China, in 2008. She is currently a Professor with Sichuan University. Her research interests include machine learning and image processing.