International Information and Engineering Technology Association
Advancing the World of Information and Engineering

Traitement du Signal
Vol. 40, No. 2, April, 2023, pp. 629-637
Journal homepage: http://iieta.org/journals/ts

Face Mask Segmentation Method Combining Salient Features and Gender Constraints

Li Xu*, Dechun Zheng

School of Electronic and Information Engineering, Ningbo University of Technology, Ningbo 315211, China

Corresponding Author Email: xuli@nbut.edu.cn

https://doi.org/10.18280/ts.400221

Received: 27 December 2022
Accepted: 15 March 2023

Keywords:
salient features, gender constraints, face mask segmentation

ABSTRACT

With the global pandemic of COVID-19, masks have become essential items in public places, posing challenges to security and convenience facilities based on facial recognition technology, such as access control systems and payment systems. In existing solutions, gender constraints help improve the accuracy of face mask segmentation, but in some special cases, such as transgender people and individuals with ambiguous gender expressions, it may lead to gender misjudgment, affecting the segmentation results. Deep learning methods may increase computational complexity, impacting real-time performance. In scenarios where a large number of images need to be processed quickly, these methods may not meet real-time requirements. Therefore, this paper studies the face mask segmentation method combining salient features and gender constraints. To enable the model to perform real-time face detection on hardware platforms, we introduce depthwise separable convolution to optimize the multi-task cascaded convolutional neural network structure, accomplishing the face detection task that combines salient features and gender constraints. The extraction of the face mask region is completed, and the technical steps for face mask extraction based on spectral features are provided. Experimental results verify the effectiveness of the constructed model.

1. INTRODUCTION

With the global pandemic of COVID-19, masks have become essential items in public places, and many people wear masks in their daily lives to protect themselves and others [1-5]. This poses challenges to security and convenience facilities based on facial recognition technology, such as access control systems and payment systems. Traditional face recognition technologies face many challenges in recognizing faces wearing masks, such as partial facial feature occlusion and light reflection caused by masks [6-11]. To address this issue, it is necessary to study a method that can effectively recognize and segment faces wearing masks [12-20]. By researching the face mask segmentation method that combines salient features and gender constraints, we can improve the face recognition accuracy under mask occlusion conditions, thus playing an important role in public safety, financial payment, access control systems, artificial intelligence monitoring, and health management.

Under today's specific conditions, due to the pandemic, a person must wear a mask for daily activities. In some cases, people are forced to remove their masks during performances. For example, consider the traditional method for facial recognition systems used for attendance monitoring, where individuals are forced to remove their masks, which is not obvious in the current situation. Priya et al. [21] established a system that helps individuals mark their attendance without the need to remove their masks. This model uses the Caffe model for face detection, a CNN model for recognition, and a Haar Cascade classifier. This model was created for use in real-time applications. The construction accuracy of the system is 90%. Rao et al. [22] proposed FMRS-CFR (Face mask recognition system-Centerface Resnet), a pandemic mask recognition system based on multi-algorithm fusion, to adapt to multi-scenario applications. In this work, the center face keypoints detection and Resnet50 classification model were used. A system that dynamically maintains multiple adaptations with external scenes is constructed, and the system is ported to the Atlas 200 development kit, with quantitative evaluations on videos in more than a dozen different scenes. Experimental results show that the FMRS-CFR system can achieve a recognition accuracy of 99.88%, greatly improving the recognition rate of maskless or correctly worn masks to some extent, achieving the purpose of effectively assisting in pandemic prevention and control. Kaliappan et al. [23] classified people into three categories, such as wearing masks, not wearing masks, and incorrectly positioned masks. The dataset was tested using three different object detection model variants, namely YOLOv4, Tiny YOLOv4, and YOLOv5. Experimental results show that the performance of the YOLOv5 model is better than the other two models, with the highest mAP value of 99.40%. Ramachandra and Marcel [24] were the first to study the use of a 3D face-customized silicone mask as a means of generating face deformation attacks. A systematic study was proposed to measure the attack potential of mask deformation (digital) attacks on commercial and academic FRS. To this extent, a new dataset was constructed using eight customized 3D silicone masks and corresponding real face images taken with three different smartphones. Mask deformation was done using a landmark-based method, and the newly constructed dataset includes 635 real, 1034 face mask, and 613 mask deformation face images. Extensive experiments were conducted to benchmark the attack potential and detection of mask deformation attacks on FRS.

Based on existing research results, it is known that current face mask segmentation methods have improved the accuracy of face recognition for people wearing masks to some extent. However, there are still some shortcomings and deficiencies. Under different lighting conditions, the extraction of significant features may be affected, leading to a decline in recognition performance. Gender constraints help improve the accuracy of face mask segmentation, but in some special cases, such as transgender individuals or those with ambiguous gender expression, gender misjudgment may occur, affecting the segmentation results. Deep learning methods may increase computational complexity, resulting in real-time performance being affected. In scenarios requiring rapid processing of large numbers of images, this method may not meet real-time requirements. Therefore, this paper investigates a face mask segmentation method that combines significant features and gender constraints.

In order to enable the model to achieve real-time face detection on hardware platforms, section 2 of this paper optimizes the multi-task cascaded convolutional neural network structure by introducing depthwise separable convolution to complete the face detection task that combines significant features and gender constraints. Section 3 completes the extraction of face mask regions, providing the technical steps for face mask extraction based on spectral features. Experimental results validate the effectiveness of the constructed model.

2. FACIAL LANDMARK FEATURES AND GENDER ATTRIBUTE RECOGNITION

To realize real-time facial detection on hardware platforms, this paper optimizes the multi-task cascade convolutional neural network structure by introducing depthwise separable convolution. Depthwise separable convolution is a computationally efficient and effective convolution method that separates spatial convolution and channel convolution, achieving lower parameter and computational complexity while still effectively extracting image features. This enables the algorithm to achieve higher recognition accuracy at a lower computational cost in the facial detection task with the fusion of landmark features and gender constraints. The reduction of model parameters makes it easier to deploy the model on hardware platforms, such as embedded devices and mobile devices. This allows the facial detection method with the fusion of landmark features and gender constraints to be widely applied on various hardware platforms. The multi-task cascade convolutional neural network structure adopted in this paper can handle multiple tasks simultaneously, such as mask detection and gender recognition. This allows the algorithm to achieve the recognition of multiple attributes in the facial detection task with the fusion of landmark features and gender constraints, improving the comprehensiveness of recognition.

The following formula gives the calculation of the parameter amount M_t and M_d in standard convolution, assuming the side length of the convolution kernel is represented by C_G, and the input and output channel numbers are represented by D_{in} and D_{out}, respectively:

$$M_t = C_G \times C_G \times D_{in} \times D_{out} \tag{1}$$

$$M_d = F_{out} \times Q_{out} \times D_{out} \times C_G \times C_G \times D_{in} \tag{2}$$

Figure 1 shows the calculation process of Mt and Md in standard convolution. In contrast, the formula for the parameter amount M_t and computation amount M_d in *DepthWis* of the adopted depthwise separable convolution is as follows:

$$M_t^{DW} = C_G \times C_G \times D_{in} \tag{3}$$

$$M_d^{DW} = C_G \times C_G \times D_{in} \times F_{out} \times Q_{out} \tag{4}$$

The following formula gives the calculation of the parameter amount and the computation amount in *PointWise*:

$$M_t^{PW} = 1 \times 1 \times D_{in} \times D_{out} \tag{5}$$

$$M_d^{PW} = F_{out} \times Q_{out} \times 1 \times 1 \times D_{in} \tag{6}$$

By adding the above formulas in series, the total parameter amount and the total calculation of the depthwise separable convolution can be obtained, and the calculation process is given by the following formula:

$$M_t = M_t^{DW} + M_t^{PW} \tag{7}$$

$$M_d = M_d^{DW} + M_d^{PW} \tag{8}$$

From the above derivation process, it can be seen that depthwise separable convolution significantly compresses the model parameter amount and computation amount based on traditional standard convolution. Assuming the ratio of the total parameter amount of depthwise separable convolution to standard convolution is represented by β, and the ratio of the total computation amount is represented by γ, the following calculation formula is obtained:

$$\beta = \frac{M^{CQ} + TQ}{M} = \frac{C_G \times C_G \times D_{in} + D_{in} \times D_{out}}{C_G \times C_G \times D_{in} \times D_{out}} = \frac{C_G^2 + D_{out}}{C_G^2 \times D_{out}} \tag{9}$$

$$\gamma = \frac{M^{DW} + TQ}{M} = \frac{C_G^2 + D_{out}}{D_{out} \times C_G^2} = \frac{1}{D_{out}} = \frac{1}{C_G^2} \tag{10}$$

Training facial landmark features and gender attributes separately requires the separate deployment of facial landmark feature recognition networks and gender recognition networks. The disadvantage of this approach is that hardware needs to read the weights of both networks, which doubles the overall parameter amount compared to a single network. Therefore, this paper designs a facial multi-attribute recognition network with the fusion of landmark features and gender constraints using *VGG-11* as the backbone network. By integrating the landmark features and gender constraints into a single network, the total parameter amount can be reduced. This can ease the burden on hardware and improve computational efficiency. Integrating landmark features and gender constraints into a single network also allows for better utilization of shared feature representations within the network. This helps to improve the performance of multi-attribute recognition tasks and avoid redundancy due to separate training. Additionally, the *VGG-11* network, compared to other deeper convolutional neural networks (such as *VGG-16*, *VGG-19*, etc.), has a lower computational complexity, which can maintain high recognition accuracy while reducing runtime.

To further reduce the parameter amount and obtain a smaller input size, this paper provides an optimization scheme for the original *VGG-11* network. Figure 2 shows the optimized network structure. By removing the *Conv8*, *FC*, and *FC2* layers, the model's parameter amount is significantly reduced, easing the burden on hardware resources, reducing computational costs, and increasing computation speed. Adjusting the input image size to 224×224 can reduce the computation amount and memory requirements. This allows the network to achieve high performance even under lower computational resource conditions. The optimized network structure is more compact, facilitating improved model deployment flexibility on different devices. By adjusting the size of the *FC3* layer to 1×1×22, the network can adapt to facial multi-attribute recognition tasks. This enables the network to handle multiple tasks simultaneously, such as landmark features and gender constraints, enhancing the comprehensiveness and efficiency of recognition.

This paper attempts to combine facial salient features and gender in a single network structure for output. The specific representation and combination method are given below: Suppose there are *N* salient features (such as keypoints, facial expressions, etc.), which can be numbered in order from 1 to *N*. For gender, we can use the same encoding method as described above, i.e., 0 for females and N+1 for males. Next, the facial salient feature values and gender values are added together to obtain a (*N*+1)*2-dimensional output space. In the last fully connected layer of the *VGG* network, the output size is adjusted to 1×1×((*N*+1)*2), and the probability values are obtained through softmax. In the output processing phase, the maximum probability value can be found from the (*N*+1)*2 probability values and the corresponding salient feature and gender classification results can be calculated. For example, suppose there are 10 salient features, so the gender encoding is 0 (female) and 1 (male), resulting in 22 outputs with a size of 1×1×22. By processing the output probability values, a classification result containing both salient features and gender attributes can be obtained, and the specific calculation method is shown in the following formula. Assuming that the network's probability of predicting females is represented by T_{FE}, the network's probability of predicting males is represented by T_{MA}, and the network's probability of predicting age *l* is represented by $T_{age}(l)$, then:

$$T_{FE}=\sum\nolimits_{i=1}^{11}T_i \tag{11}$$

$$T_{MA}=\sum\nolimits_{i=12}^{22}T_i \tag{12}$$

$$T_{age}\left(l\right)=T_l+T_{l+11} \tag{13}$$

In this paper, *N* salient features are numbered from 1 to *N*, and gender is encoded using 0 (female) and *N*+1 (male). Next, the facial salient feature values and gender values are added together to obtain a (*N*+1)*2-dimensional output space. For females (gender value 0), the probability of having a salient feature value of *i* is the probability of the i-th output of the *softmax* layer. For males (gender value *N*+1), the probability of having a salient feature value of *i* is the probability of the (*N*+1+i)-th output of the *softmax* layer. Therefore, the total probability of having a facial salient feature value of *i* is the sum of the probabilities of the i-th output and the (*N*+1+*i*)-th output of the *softmax* layer. To obtain the probability values for all facial salient feature classifications, this process can be applied to all salient feature values (1 to *N*). First, initialize a probability value array *P* of length *N*. For each salient feature value, calculate *P*[*i*]=*softmax_output*[*i*]+*softmax_output* [*N*+1+*i*]. Finally, the array *P* will contain the probability values for all facial salient feature classifications.

In the optimization of the network, the gender loss function and facial salient feature loss function are cross-entropy loss function and *L*1 loss function, respectively. In the facial salient feature classification task, this paper treats facial salient feature classification as a probability distribution problem, using Gaussian distribution to represent the probability distribution of facial salient features, and introducing prior knowledge, i.e., the pre-set Gaussian distribution, through the constraint of *KL* divergence. By comparing the predicted facial salient feature probability distribution of the model with the pre-set Gaussian distribution, the difference between the two distributions is minimized to improve the prediction accuracy. At the same time, the *KL* divergence loss function has good interpretability, making the performance of the model during the prediction process easier to understand and evaluate. Assuming that the true distribution is represented by *t*(*a*), the predicted distribution is represented by *w*(*a*), and the matching degree between the predicted distribution and the true distribution is represented by C_{KL}, the following formula gives the calculation formula of KL divergence:

$$C_{KL}\left(t\|w\right)=\sum\nolimits_{i=1}^{M}t\left(a_i\right)\log\left(\frac{t\left(a_i\right)}{w\left(a_i\right)}\right) \tag{14}$$

Assuming that *L*1 loss function is represented by $LOSS_1$, *KL* divergence loss function is represented by $LOSS_2$, and cross-entropy loss function is represented by $LOSS_3$. The loss function of the constructed multi-attribute recognition network can be calculated through the following formula:

$$LOSS=0.4LOSS_1+0.6LOSS_2+xLOSS_3 \tag{15}$$

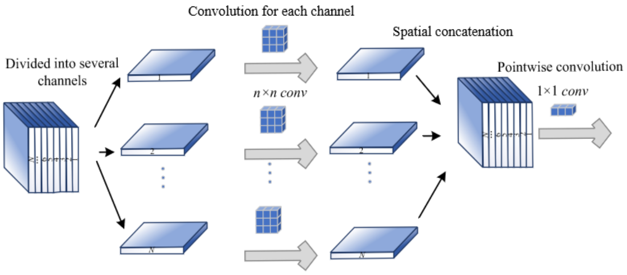


Figure 1. The calculation process of M_t and M_d in standard convolution

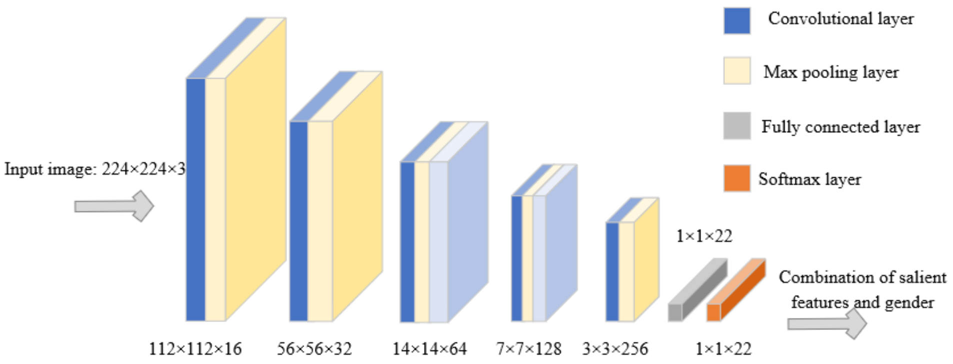


Figure 2. Optimized network structure

3. FACIAL MASK AREA EXTRACTION

In facial mask segmentation tasks, masks usually have significantly different colors from facial skin. Facial skin typically has a stronger red component and a weaker blue component, while some common mask colors (such as blue) have stronger blue components and weaker red components. The significance of the Normalized Blue-Red Index (*NBRI*) lies in measuring the relative strength of blue and red components in an image. Therefore, using the *NBRI* can help distinguish between facial skin and mask areas. The index is defined as follows:

$$NBRI = \frac{Y - S}{Y + S} \tag{16}$$

The blue component and red component are represented by *Y* and *S*, respectively. In facial mask segmentation tasks, larger *NBRI* values may indicate mask areas in the image, especially when the mask is a color that is clearly different from the skin, such as blue. Smaller *NBRI* values may indicate skin areas in the image, as facial skin usually has higher red components and lower blue components.

In the HSI color space, we can define the Normalized Mask Region Index (*NMRI*) based on the features of hue (*H*), saturation (*S*), and intensity (*I*). Since the HSI color space can better reflect the human eye's perception of color, it may have better performance in facial mask segmentation tasks. Assuming that saturation and intensity in HIS are represented by *F* and *I*, respectively, we define the Normalized Mask Region Index (*NMRI*) as follows:

$$NMRI = \frac{F - I}{F + I} \tag{17}$$

where, *S* and *I* represent the saturation and intensity components in the image. In facial mask segmentation tasks, larger *NMRI* values may indicate mask areas in the image, as masks usually have higher saturation and lower intensity. Smaller *NMRI* values may indicate skin areas in the image, as facial skin usually has higher intensity and lower saturation.

By utilizing the Normalized Blue-Red Index and the Normalized Mask Region Index, effective facial mask segmentation can be achieved, distinguishing between masked regions and exposed skin areas. Combining the two indices can provide better segmentation results under different color and brightness conditions, especially when the HSI color space can better distinguish mask and skin features.

This paper combines suppression algorithms with improved gray world algorithms to compensate for the facial mask area, increasing the difference between the facial mask area and other facial areas. The suppression algorithm helps to reduce the interference of high-frequency noise and non-important features in facial mask segmentation, making the mask edge and shape features more prominent. The improved gray world algorithm can automatically adjust the color balance and contrast of the image under different lighting conditions, making the difference between masks and skin easier to recognize. Under different lighting conditions and mask colors, this compensation method can effectively improve the accuracy of facial mask segmentation.

First, preprocess the input image using the suppression algorithm. The suppression algorithm can weaken high-frequency noise and non-important features in the image, thereby retaining more prominent structural features. In facial mask area segmentation tasks, this can help highlight the edge and shape features of the mask. Then, apply the improved gray world algorithm to balance color and enhance contrast in the suppressed image. The improved gray world algorithm adjusts the *RGB* components of the image so that their average approaches a predetermined gray value, thereby achieving automatic white balance. This method helps to reduce the impact of changes in lighting conditions, making the color difference between masks and skin more evident.

Divide the facial image into mask area *R*(*a*, *b*) and non-mask area *NS*(*a*, *b*) images, suppress the blue component of *R*(*a*, *b*) first, and then estimate the illumination intensity and color of the mask area and non-mask area respectively. The following formula expresses the blue component suppression:

$$Y^{'} = \mu Y \tag{18}$$

Assuming that the mask area illumination color is represented by *L*(*R*), the non-mask area illumination color is represented by *L*(*MR*), the mask area image is represented by *R*(*a*, *b*), the non-mask area image is represented by *MR*(*a*, *b*), the fixed constant is represented by *x*, and the exponent parameter in Minkowski norm is represented by *t*. Based on the calculation results, compensate the mask area as follows:

$$l(R) = x\left\{\frac{\iint\left(R^{\partial}(a,b)^{t}\,dadb\right)}{\iint dadb}\right\}^{\frac{1}{t}} \tag{19}$$

$$l\left(MR\right)=x\left\{\frac{\iint\left(MR^{\partial}\left(a,b\right)^{t}dadb\right)}{\iint dadb}\right\}^{\frac{1}{t}} \tag{20}$$

To correct the mask area image based on the following formula, the purpose of mask compensation can be achieved:

$$\bar{R}\left(a,b\right)=R\left(a,b\right)\times l(MR)/l(R) \tag{21}$$

In order to achieve the separation of facial mask regions from non-mask regions in images, this paper adopts a geometric-weighted connectivity analysis model to analyze the candidate facial mask regions. This model takes into account both the geometric and texture information of the facial mask region, enhancing the ability to characterize different region features. Firstly, the local variance of the image is extracted as texture information. Texture information can reflect the variations in pixel values within local areas, which helps to distinguish the characteristics of different regions. The geometric features are calculated using an improved aspect ratio index, which helps to measure the shape features of regions and better distinguish between mask and non-mask regions. By combining texture information and geometric features, a geometric-weighted connectivity analysis model is constructed. The facial mask candidate regions are analyzed, and the final facial mask region is determined based on the output of the geometric-weighted connectivity analysis model. Assuming that the diagonal length of the bounding box of the connected region object is represented by *K*, the area of the connected region object is represented by *R*, the maximum connected object area is represented by X_{max}, the area of the i-th connected object is represented by x_i, and the variance of the i-th connected object is represented by ε_i. This model must satisfy the thresholds P_1 and P_2.

$$HQ=(Mi>P_1)\cup(Po<P_2) \tag{22}$$

$$Mi=(K^2/R) \tag{23}$$

$$Po=\frac{X_{\max}}{x_i}\cdot\varepsilon_i \tag{24}$$

The following are the technical steps for facial mask extraction based on spectral features:

(1) Pre-processing: Perform noise removal, edge enhancement, and other processing on the input face image to improve image quality and create a more accurate basis for subsequent steps.

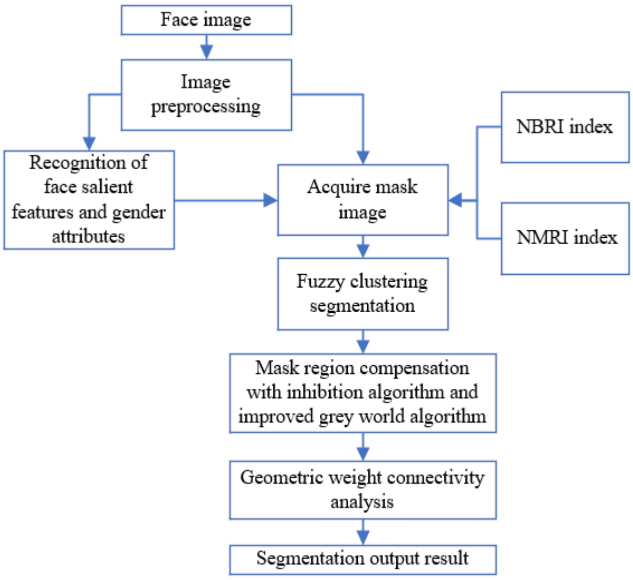
(2) Obtain mask image: Use the normalized blue-red component index or other methods, such as the normalized mask region index based on the *HIS* color space, to extract potential facial mask regions from the original image and generate the corresponding binary mask image.

(3) Fuzzy clustering segmentation: Use the fuzzy clustering algorithm to segment the mask image, obtaining a series of possible candidate regions.

(4) Region compensation: Combine suppression algorithms and improved gray world algorithms to compensate for the facial mask region, increasing the difference between the facial mask region and other facial regions for subsequent extraction.

(5) Candidate region acquisition: Analyze the compensated image to obtain candidate regions that may contain facial mask regions.

(6) Geometric-weighted connectivity analysis: Employ the geometric-weighted connectivity analysis model to analyze candidate regions, taking into account both geometric and texture information. Use local variance as texture features and the improved aspect ratio index as geometric features, and calculate area weights. By comprehensively analyzing these features, the true facial mask regions are identified and non-mask regions are eliminated. Figure 3 shows the technical flowchart of the algorithm proposed in this paper.



Face image
Image preprocessing
Recognition of face salient features and gender attributes
Acquire mask image
NBRI index
NMRI index
Fuzzy clustering segmentation
Mask region compensation with inhibition algorithm and improved grey world algorithm
Geometric weight connectivity analysis
Segmentation output result

Figure 3. Technical flowchart of the algorithm in this paper

4. EXPERIMENTAL RESULTS AND ANALYSIS

Observing the results of the saturation component grayscale histograms and brightness component grayscale histograms shown in Figures 4 and 5, we can draw the following conclusions. In the saturation component grayscale histogram, the vast majority of pixel points are concentrated between 135 and 140, indicating that the saturation of the face mask area is relatively low and the color is relatively uniform. In the brightness component grayscale histogram, the vast majority of pixel points are concentrated between 0 and 50, with virtually no pixel points distributed in other areas. This suggests that the distribution of the face mask area in the H component is relatively narrow, with little color variation. Combining these observations, we can analyze that by combining the suppression algorithm and the improved gray world algorithm, the color difference between the face mask area and other areas can be enhanced while maintaining the original texture and edge information, thus improving the accuracy and robustness of face mask segmentation.

Furthermore, this paper conducted a comparative analysis of the experimental results of face detection method optimization strategies, as shown in Table 1. The table shows the impact of various optimization strategies on feature extraction accuracy, gender recognition accuracy, *mIoU*, and *pAcc*. The original *VGG*-11 network performs well in feature extraction accuracy, gender recognition accuracy, and other aspects, but there is still room for improvement. Compared to the original *VGG*-11 network, deleting *Conv*8, *FC*, and *FC*2 layers improves gender recognition accuracy and *mIoU* but slightly reduces feature extraction accuracy and *pAcc*. This suggests that reducing network complexity to some extent helps with gender recognition and segmentation performance. Adjusting the input image size in this optimization strategy significantly improves gender recognition accuracy and *pAcc*, but slightly reduces feature extraction accuracy and *mIoU*. This indicates that adjusting the input image size contributes to improving the accuracy of gender recognition. Adjusting the size of the *FC*3 layer significantly improves feature extraction accuracy and gender recognition accuracy, and also increases *mIoU* and *pAcc*. This demonstrates that adjusting the size of the *FC*3 layer can effectively improve network performance. Combining salient features with gender improves feature extraction accuracy, gender recognition accuracy, *mIoU*, and *pAcc*. This proves that combining salient features with gender helps improve network performance. Compared to the original *VGG*-11 network, the final model shows significant improvements in feature extraction accuracy, gender recognition accuracy, *mIoU*, and *pAcc*. This demonstrates that the integration of the aforementioned optimization strategies can effectively improve the performance of face detection methods. By comparing various optimization strategies and their results, the conclusion can be drawn that the face detection method optimization strategy combining salient features and gender constraints effectively improves feature extraction accuracy, gender recognition accuracy, *mIoU*, and *pAcc*, thus achieving more accurate face detection.

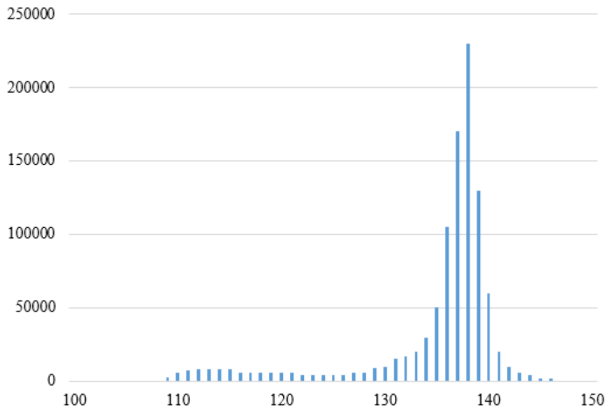


Figure 4. Histogram of saturation component in face mask area

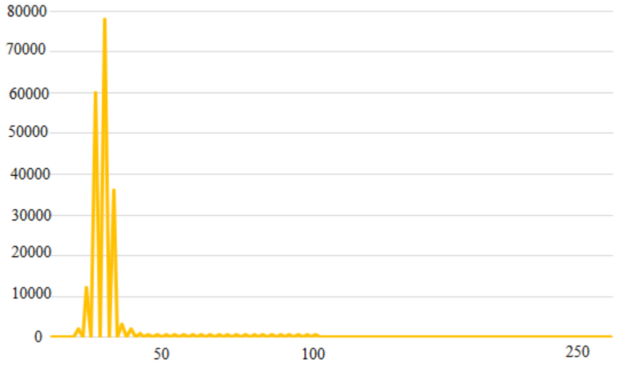


Figure 5. Histogram of brightness component in face mask area

Table 1. Comparative analysis of experimental results of face detection method optimization strategies

Method	Feature extraction accuracy	Gender recognition accuracy	*mIoU*	*pAcc*
Original *VGG*-11 network	0.95015	0.82131	0.74646	0.82451
Before deleting *Conv*8, *FC*, and *FC*2 layers	0.95416	0.84416	0.79156	0.81566
Before adjusting input image size	0.95311	0.86414	0.81515	0.87741
Before adjusting the size of *FC*3 layer	0.9864	0.85616	0.80515	0.87156
Before combining salient features with gender	0.98153	0.86155	0.80521	0.87515
Final model	0.96415	0.8761	0.81502	0.87514

Table 2. Loss function experiment results of the multi-attribute recognition network

Method	Feature extraction accuracy	Gender recognition accuracy	*mIoU*	*pAcc*
Original *VGG*-11 network	0.91546	0.85156	0.76481	0.82545
Introducing *L*1 loss function	0.96151	0.8458	0.81651	0.86741
Introducing cross-entropy loss function	0.97544	0.95615	0.81565	0.86165
Introducing *KL* divergence loss function	0.98484	0.96651	0.81677	0.89566

Table 3. Comparison of experimental results of different face mask region compensation methods

Method	Sample set 1	Sample set 2
Inhibition algorithm	0.816	0.875
Grey world algorithm	0.883	0.941
Histogram equalization	0.803	0.823
Bilateral filtering	0.901	0.919
Retinex algorithm	0.846	0.865
Gaussian pyramid fusion	0.651	0.775
Laplacian pyramid fusion	0.586	0.814
Color balance	0.762	0.764
Inverse projection	0.872	0.926
The combined algorithm of this paper	0.914	0.967

Table 4. Comparison of experimental results of different face mask region segmentation methods (*mIoU*)

Method	Sample set 1	Sample set 2
Threshold segmentation	0.6515	0.8715
Fuzzy *c*-means clustering	0.8264	0.9461
Edge detection	0.8365	0.9484
Level set method	0.8463	0.9488
Graph cut algorithm	0.7464	0.9064
U-Net	0.7994	0.9075
The proposed model	0.8187	0.9525

Table 2 presents the experimental results of the loss functions for the multi-attribute recognition network. From the table, we can see the impact of various loss functions on feature extraction accuracy, gender recognition accuracy, *mIoU*, and *pAcc*. The original *VGG*-11 network performs well in terms of feature extraction accuracy, gender recognition accuracy, etc., but there is still room for improvement. Compared with the original *VGG*-11 network, introducing the *L*1 loss function strategy significantly improves feature extraction accuracy, *mIoU*, and *pAcc*, but slightly decreases gender recognition accuracy. This indicates that the *L*1 loss function can improve the accuracy of feature extraction, but has little effect on gender recognition. Introducing the cross-entropy loss function significantly improves feature extraction accuracy and gender recognition accuracy, but slightly decreases *mIoU* and *pAcc*. This suggests that the cross-entropy loss function has a significant impact on feature extraction and gender recognition. Introducing the *KL* divergence loss function significantly improves feature extraction accuracy, gender recognition accuracy, and *pAcc*, and also increases *mIoU*. This indicates that the *KL* divergence loss function can further improve the accuracy of feature extraction and gender recognition. By comparing various loss functions and their results, we can draw the conclusion: in the constructed multi-attribute recognition network, introducing *L*1 loss function, cross-entropy loss function, and *KL* divergence loss function all contribute to improving feature extraction accuracy and gender recognition accuracy. In particular, introducing the KL divergence loss function has the most significant improvement on overall model performance. Therefore, the multi-attribute recognition network with these loss functions has high effectiveness in improving recognition accuracy.

When building a dataset for face mask segmentation, it is essential to ensure the dataset is diverse and representative to train a model with high generalization ability. This paper constructs a public scenario dataset and a specific scenario dataset. When building the dataset, make sure the sample size is sufficient and balanced among categories to avoid training bias caused by class imbalance. At the same time, when dividing the training set, validation set, and test set, ensure that the sample distribution in each subset is roughly the same to more accurately assess model performance. If the face mask segmentation model needs to be applied in a specific industry, more relevant scene photos can be included in the dataset to train a more targeted model.

Figure 6 shows the loss accuracy curve of network training. It can be seen that the constructed network model converges quickly during training and tends to be stable in the later stage. The accuracy of facial salient feature and gender attribute recognition reaches 92% and later exceeds 96%. These indicators show that the network model has good performance. Combining this information, it can be verified that the network model constructed in this paper has high accuracy in face mask segmentation tasks based on salient feature and gender constraint fusion. This indicates that the model can effectively recognize facial salient features and gender attributes, thereby improving the accuracy of face mask segmentation. The model converges quickly during training and is stable in the later stage, indicating that the model can learn effective feature representations during training, thus providing better prediction results in the testing phase.

According to Table 3, the experimental results of different face mask region compensation methods on two sample sets can be seen. It can be seen that in the two sample sets, the combined algorithm adopted in this paper achieves the best segmentation effect, with *mloU* of 0.914 and 0.967 respectively. For other methods, the gray world algorithm, bilateral filtering and inverse projection perform relatively well on the two sample sets. This shows that these methods have a certain robustness in handling different types of sample sets. The Gaussian pyramid fusion and Laplacian pyramid fusion methods perform worse in sample set 1, but have a significant improvement in sample set 2. This may indicate that these two methods have better adaptability to specific types of images in sample set 2, but weaker generalization ability in sample set 1. Histogram equalization and color balance methods perform relatively stable on the two sample sets, but the overall effect is not as good as other methods. This shows that these methods may not have sufficient advantages for the task of face mask region segmentation.

According to Table 4, the experimental results of different face mask region segmentation methods on the two sample sets can be seen. In the two sample sets, the segmentation effect of the model proposed in this paper is the best, with *mloU* values of 0.8187 and 0.9525 respectively. Among other methods, fuzzy c-means clustering, edge detection and level set methods perform relatively well on the two sample sets. This shows that these methods have a certain robustness in handling different types of sample sets. Threshold segmentation, graph cut algorithm and *U-Net* perform worse in sample set 1 but have significant improvement in sample set2. This may indicate that these methods have better adaptability to specific types of images in sample set 2 but weaker generalization ability in sample set 1. Although the proposed model achieved good results on both sample sets, there is still room for further optimization to improve its applicability in different scenarios. In summary, face mask region segmentation experiments on different types of sample sets can clearly show that the proposed model has better performance and stronger generalization ability.

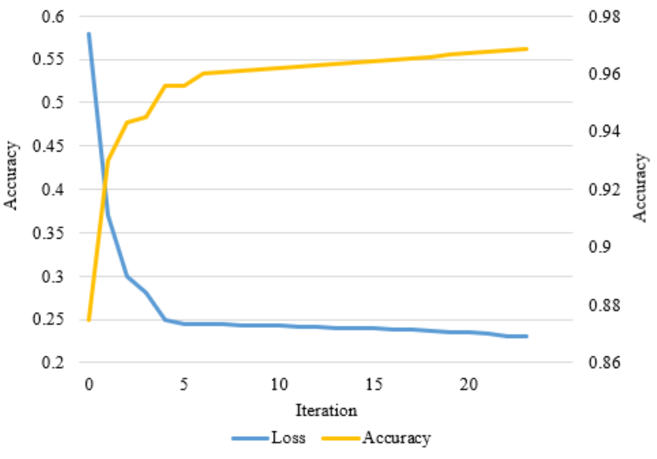


Figure 6. The curve of network loss and accuracy

5. CONCLUSION

This paper studied a face mask segmentation method that integrates salient features and gender constraints. In order to realize real-time face detection on hardware platforms, a multi-task cascade convolutional neural network structure is optimized by introducing depth separable convolution to complete the face detection task of integrating salient features and gender constraints. Face mask region extraction is completed, and technical steps of face mask extraction based on spectral features are given. The experimental results verify the effectiveness of the established model. The saturation component gray histogram and the brightness component gray histogram are given, and the experimental results of the face detection optimization strategy are analyzed and compared to effectively improve the feature extraction accuracy, gender recognition accuracy, mloU and pAcc, thus realizing more accurate face detection. The experimental results of the loss function of the multi-attribute identification network are given, and introducing KL divergence loss function improves the overall performance of the model the most significantly. The curve of network loss and accuracy in training is given, which verifies that the model can learn effective feature representation during training so that better prediction results can be provided in the test stage. The experimental results of different face mask region compensation methods and different face mask region segmentation methods are compared and verified that the combined algorithm adopted in this paper achieves the best segmentation effect.

ACKNOWLEDGMENT

This research was supported by Zhejiang Provincial Natural Science Foundation of China (Grant No.: LY19F020008); Zhejiang Province Commonweal Projects of China (Grant No.: LGG19F03007).

REFERENCES

[1] Miah, M.J., Pei, J., Kim, H., Sharma, R., Jang, J.G., Ahn, J. (2023). Property assessment of an eco-friendly mortar reinforced with recycled mask fiber derived from COVID-19 single-use face masks. Journal of Building Engineering, 105885. https://doi.org/10.1016/j.jobe.2023.105885

[2] Nam, J.Y., Lee, T.R., Tokmurzin, D., Park, S.J., Ra, H.W., Yoon, S.J., Moon, J.H., Lee, J.G., Lee, D.H., Seo, M. W. (2023). Hydrogen-rich gas production from disposable COVID-19 mask by steam gasification. Fuel, 331: 125720. https://doi.org/10.1016/j.fuel.2022.125720

[3] Javed, I., Butt, M. A., Khalid, S., Shehryar, T., Amin, R., Syed, A.M., Sadiq, M. (2023). Face mask detection and social distance monitoring system for covid-19 pandemic. Multimedia Tools and Applications, 82(9): 14135-14152. https://doi.org/10.1007/s11042-022-13913-w

[4] Jayaswal, R., Dixit, M. (2022). AI-based face mask detection system: A straightforward proposition to fight with Covid-19 situation. Multimedia Tools and Applications, 82(9): 13241-13273. https://doi.org/10.1007/s11042-022-13697-z

[5] Chen, X., Yan, H. F., Zheng, Y. J., Karatas, M. (2023). Integration of machine learning prediction and heuristic optimization for mask delivery in COVID-19. Swarm and Evolutionary Computation, 76: 101208. https://doi.org/10.1016/j.swevo.2022.101208

[6] Biswas, A., Paudel, B., Sarkar, N. (2022). Smart face recognition with mask/no mask detection. In advances in communication, devices and networking: Proceedings of ICCDN 2020, 776: 143-149. https://doi.org/10.1007/978-981-16-2911-2_15

[7] Goto, T., Hongo, M. (2022). Improvement of face recognition accuracy for mask wearers. In 2022 IEEE International Conference on Consumer Electronics-Taiwan, 301-302. https://doi.org/10.1109/ICCE-Taiwan55306.2022.9869023

[8] Le-Anh, T., Nguyen-Van, B., Le-Trung, Q. (2022). An intelligent edge system for face mask recognition application. In Industrial Networks and Intelligent Systems: 8th EAI International Conference, INISCOM 2022, Virtual Event, April 21–22, 2022, Proceedings, 107-124. https://doi.org/10.1007/978-3-031-08878-0_8

[9] Bennur, S., Bharadwaj, D., Anand, P., Premananda, B.S. (2022). Face mask detection and face recognition of unmasked people in organizations. In 2022 3rd International Conference on Electronics and Sustainable Communication Systems (ICESC), 1212-1218. https://doi.org/10.1109/ICESC54411.2022.9885367

[10] Dai, W., Wang, J., Ren, T., Zhu, Z. (2022). Face mask recognition based on YOLOv3-tiny. In 2022 3rd International Conference on Electronic Communication and Artificial Intelligence (IWECAI), 507-511. https://doi.org/10.1109/IWECAI55315.2022.00104

[11] Li, X. (2022). an effective and efficient face mask recognition system for edge devices. In 2022 5th International Conference on Data Science and Information Technology (DSIT), 1-5. https://doi.org/10.1109/DSIT55514.2022.9943825

[12] Mohamed, M.M., Nessiem, M.A., Batliner, A., Bergler, C., Hantke, S., Schmitt, M., Baird, A., Mallol-Ragolta, A., Karas, V., Amiriparian, S., Schuller, B.W. (2022). Face mask recognition from audio: The MASC database and an overview on the mask challenge. Pattern Recognition, 122: 108361. https://doi.org/10.1016/j.patcog.2021.108361

[13] Kumar, G., Zaveri, M.A., Bakshi, S., Sa, P.K. (2022). Who is behind the mask: Periocular biometrics when face recognition fails. In 2022 Second International Conference on Power, Control and Computing Technologies (ICPC2T), 1-6. https://doi.org/10.1109/ICPC2T53885.2022.9777027

[14] Azraai, M.A., Rani, R., Qibtiah, R.M., Samian, H. (2022). Face mask wear detection by using facial recognition system for entrance authorization. International Journal of Advanced Computer Science and Applications, 13(6): 117-123.

[15] Manzoor, S., Kim, E.J., Joo, S. H., Bae, S.H., In, G.G., Joo, K.J., Choi, J.H., Kuc, T.Y. (2022). Edge deployment framework of guardbot for optimized face mask recognition with real-time inference using deep learning. Ieee Access, 10: 77898-77921. https://doi.org/10.1109/ACCESS.2022.3190538

[16] Luu, T.H., Phuc, P.N.K., Yu, Z.Q., Pham, D.D., Cao, H.T. (2022). Face mask recognition for Covid-19 prevention. Computers, Materials and Continua, 73(2): 3251-3262. https://doi.org/10.32604/cmc.2022.029663

[17] Arroyo-Rojas, U., Jimenez-Martinez, M., Benitez-Garcia, G., Olivares-Mercado, J., Takahashi, H. (2022). Twitter face image mining for recognition of different face mask types. Frontiers in Artificial Intelligence and Applications, 355: 298-309.

[18] Al-Rammahi, A.H.I. (2022). Face mask recognition system using MobileNetV2 with optimization function. Applied Artificial Intelligence, 36(1): 2145638. https://doi.org/10.1080/08839514.2022.2145638

[19] Yang, B., Wu, J., Ikeda, K., Hattori, G., Sugano, M.,

Iwasawa, Y., Matsuo, Y. (2022). Face-mask-aware facial expression recognition based on face parsing and vision transformer. Pattern Recognition Letters, 164: 173-182. https://doi.org/10.1016/j.patrec.2022.11.004

[20] Jayashree, N.C., Karigar, A., Anantapura, A. (2022). Convolution neural network based Covid-19 face mask detection and recognition. In MysuruCon 2022-2022 IEEE 2nd Mysore Sub Section International Conference.

[21] Priya, V.G., Sri, M.A., Sunithamani, S., Roy, S.M. (2023). Development of a face recognition system for registering attendance of students wearing mask. In 2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT), 239-244. https://doi.org/10.1109/IDCIoT56793.2023.10053511

[22] Rao, H., Chen, A., Gon, J., Yan, F. (2023). Face mask recognition system for epidemic prevention and control based on multi-algorithm fusion. In Advanced Manufacturing and Automation XII, 10-17. https://doi.org/10.1007/978-981-19-9338-1_2

[23] Kaliappan, V.K., Thangaraj, R., Pandiyan, P., Mohanasundaram, K., Anandamurugan, S., Min, D. (2023). Real-time face mask position recognition system using YOLO models for preventing COVID-19 disease spread in public places. International Journal of Ad Hoc and Ubiquitous Computing, 42(2): 73-82. https://doi.org/10.1504/IJAHUC.2023.128499

[24] Ramachandra, R., Marcel, S. (2022). MASK-MORPH: does morphing of custom 3D face masks threatens the face recognition systems. In 2022 18th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 1-8. https://doi.org/10.1109/AVSS56176.2022.9959348