

Research Article

Multilevel and Multiscale Feature Aggregation in Deep Networks for Facial Constitution Classification

Er-Yang Huan  and Gui-Hua Wen 

School of Computer Science and Engineering, South China University of Technology, Guangzhou 510006, China

Correspondence should be addressed to Gui-Hua Wen; crghwen@scut.edu.cn

Received 4 July 2019; Revised 9 October 2019; Accepted 6 December 2019; Published 20 December 2019

Academic Editor: Anna Tsantili-Kakoulidou

Copyright © 2019 Er-Yang Huan and Gui-Hua Wen. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Constitution classification is the basis and core content of TCM constitution research. In order to improve the accuracy of constitution classification, this paper proposes a multilevel and multiscale features aggregation method within the convolutional neural network, which consists of four steps. First, it uses the pretrained VGG16 as the basic network and then refines the network structure through supervised feature learning so as to capture local image features. Second, it extracts the image features of different layers from the fine-tuned VGG16 model, which are then dimensionally reduced by principal component analysis (PCA). Third, it uses another pretrained NASNetMobile network for supervised feature learning, where the previous layer features of the global average pooling layer are outputted. Similarly, these features are dimensionally reduced by PCA and then are fused with the features of different layers in VGG16 after the PCA. Finally, all features are aggregated with the fully connected layers of the fine-tuned VGG16, and then the constitution classification is performed. The conducted experiments show that using the multilevel and multiscale feature aggregation is very effective in the constitution classification, and the accuracy on the test dataset reaches 69.61%.

1. Introduction

The constitution in TCM (traditional Chinese medicine) refers to the relatively stable body traits of the individual to the internal and external environment of the body. It is a morphological structural psychological state and physiological function formed on the basis of congenital inheritance, which is a system concept formed by combining the Chinese medical discourse on human physique phenomena and the understanding of physique in many disciplines and the purpose of medical research [1]. Constitution phenomenon is an important manifestation of human life phenomenon. It has the characteristics of individual difference group homogeneity and relative stability and dynamic variability [2, 3].

Constitution classification is the basis and core content of TCM constitution research. The purpose is to standardize human constitution categories, and then to give different personalized conditioning options for different constitution

types. Therefore, it is especially important for specific people to accurately identify their constitution categories. The commonly used identification method is based on the questionnaire. All questions are answered and then scored to determine the type of constitution according to the national standard “classification and determination of Chinese medicine constitution” [4]. This method has the following shortcomings [5, 6]:

- (1) It is influenced by the subjective factors. Individuals are not very familiar with some problems, so that it is difficult for them to accurately choose answers. Second, individuals have concerns about some private issues and are reluctant to choose real answers.
- (2) The number of questions to be answered is relatively large. It takes long time to answer these questions, easily making individuals lose patience to answer the problems located at the end of the questionnaire. These problems are often randomly selected, which

will inevitably affect the correct judgment of the constitution.

- (3) The calculation method of scores is much complex, so that constitution types of many people cannot be accurately calculated.

In order to solve these problems, many new methods have been proposed [7–9]. For example, Su et al. [7] studied the acoustic characteristics of eight different constitutions and applied them to constitution recognition. Hou et al. [8] extracted the color and texture features of the face and then classified the body constitution. Lu et al. [9] extracted the color and texture features of the tongue and performed feature fusion. These methods use traditional feature representations such as color, texture, histogram of oriented gradient (HOG), and so on. However, these methods extracted manually designed features focusing on the local pattern of the object while ignoring the semantic information, so that these features usually have limited performance. Recently, many scholars have applied machine learning algorithms to TCM constitution recognition [10, 11]. For example, Wang and Bai [10] applied the BP neural network to pulse diagnosis to classify the type of constitution and then demonstrated its rationality and superiority. Zhang et al. [11] proposed a dynamic classification model algorithm based on relevance, constructing feature indicators on face skin and then used the improved decision tree and the fuzzy naive Bayesian algorithm to classify the constitution. Moreover, with the rapid spread of CNN, many visual recognition tasks have achieved outstanding achievements, such as image classification [12, 13], image segmentation [14, 15], object detection [16, 17], and scene recognition [18, 19]. Instead of manually designing visual features, CNN provides an end-to-end feature learning framework that automatically learns deep representations of images from a global view. Some researchers have also applied CNN to constitution recognition. Hu et al. [20] applied the convolution neural network to the pulse diagnosis. In the case of feature ambiguity, the proposed method is superior to other well-known methods. Li et al. [21] used the convolution neural network to extract the features of the pulse and then classify the body constitution. The experimental results show that this method can obtain high accuracy. Huan et al. [22] proposed a constitution recognition algorithm based on the convolutional neural network, which trained a convolutional neural network model for constitution recognition on face data. Li et al. [23] proposed a constitution recognition algorithm based on the deep neural network, which first detected the tongue image and then determined the body constitution type. Hu et al. [24] proposed a classification algorithm based on the tongue, which uses the Inception v3 model for constitution classification. Zhou et al. [25] also proposed a physique classification algorithm based on the tongue. This method first cuts the tongue and then classifies its type, but it has only three categories. Ma et al. [26] proposed a complex perception-based algorithm for constitution recognition, whose dataset is tongue images. These methods are designed to predict the

constitution categories by learning high-level semantic features through a hierarchical architecture. As the depth of the network increases, more abstract and higher level features can be extracted. However, the above methods are highly concerned with high-level semantic features, while intermediate features are ignored in the final classification. Moreover, the CNN-based methods typically rely on large-scale, manually tagged training datasets, such as the ImageNet [27]. However, in the field of Chinese medicine, it is very expensive to label large amounts of data by Chinese medicine experts, which could be obtained in the short time. Training a convolutional neural network model from scratch is not easy, and it takes a long time, even requires some patience and expertise in training neural networks [28].

Considering the above challenges, in this paper, we propose a constitution classification algorithm based on pretrained convolutional neural networks for the aggregation of multilevel and multiscale features. Figure 1 shows an overview of the proposed algorithm. Our algorithm is divided into four steps, namely, network training, feature extraction and dimension reduction, feature aggregation, and constitution classification. First, the VGG16 [29] and NASNetMobile [30] network structures are fine-tuned by the transfer learning method. Second, the features of the different layers in the fine-tuned VGG16 are extracted and the PCA is performed to these features. Third, the previous layer features of the global average pooling layer in NASNetMobile are extracted and then performed by PCA. These dimensionally reduced features are aggregated with the fully connected layer features in the fine-tuned VGG16 so as to obtain aggregated features. In the final stage, the aggregated features are input into the classifier to perform the constitution classification.

In this article, we make the following innovative contributions:

- (1) An improved version of the VGG16 network, called VGG-CI, is proposed and shown in Figure 2, which added two modules: Conv Block and Inception V2 [31]. The Conv Block module contains three cascaded convolution modules, while the Inception V2 module increases the ability to represent features. By adding the Conv Block and Inception V2 modules, the depth and width of the network are further increased to improve the classification effect.
- (2) A fusion method of multiview features is proposed. First, the output of the different layers in the VGG-CI network represents features of different levels of abstraction. Based on the VGG-CI network, we extract the features of different layers, perform PCA to them, and then aggregate them with the fully connected layer to obtain the output V1. In order to continue to increase the feature representation capability, the NASNetMobile model is used to extract features from another view. By obtaining the previous layer features of the global average pooling layer, PCA also applied these features, which are then aggregated with V1.

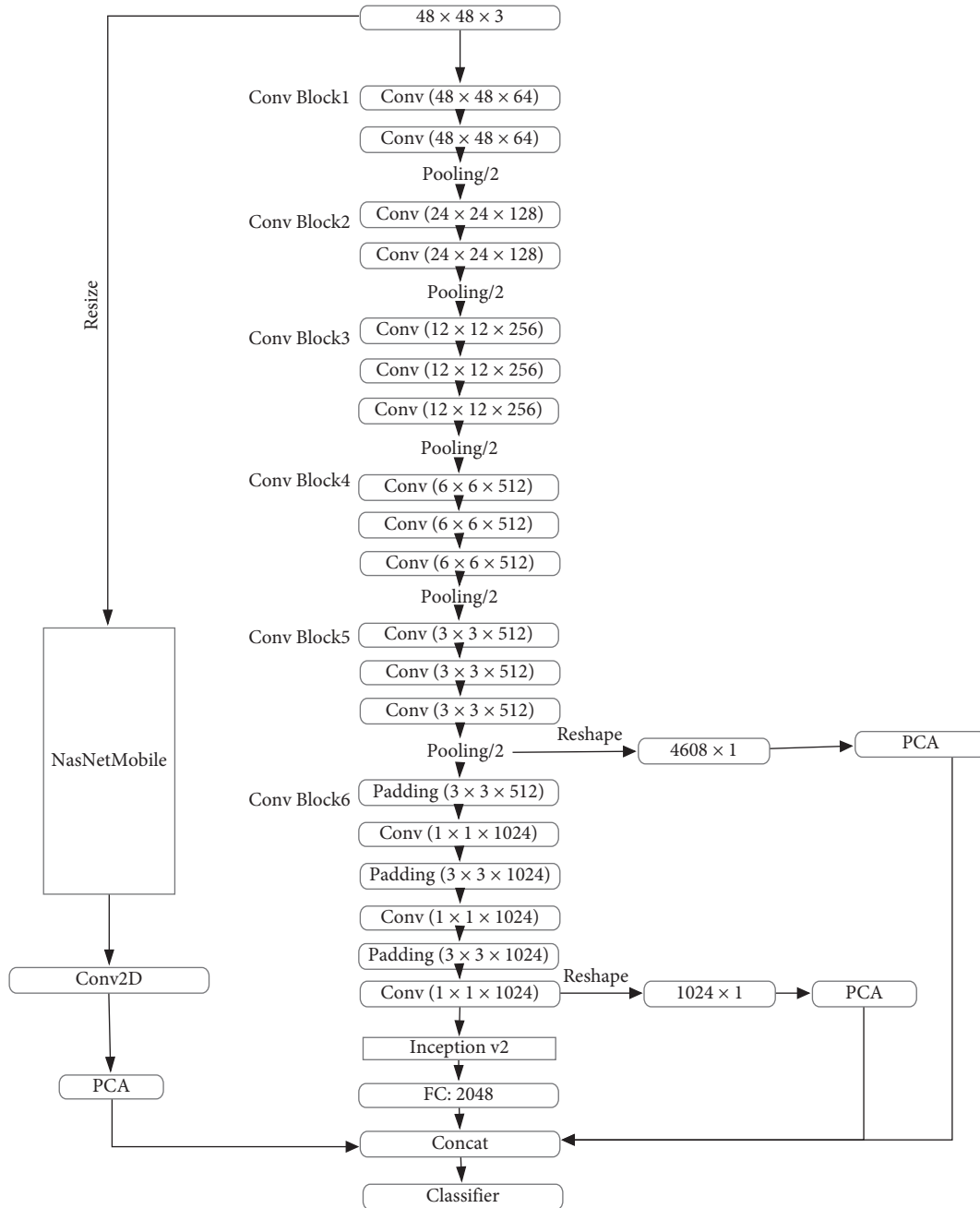


FIGURE 1: Overview of our proposed algorithm.

- (3) A large and high-quality database of clinical facial images is constructed, which can nicely support the research of facial constitution classification algorithm. At present, no institute provides a large number of high-quality clinical databases for facial constitution classification.

The structure of the paper is as follows: in Section 2, we present the method proposed. In Section 3, lots of experiments are conducted to validate the proposed method. The conclusion is presented in Section 4.

2. Proposed Approach

The overall architecture we propose is shown in Figure 1. In this section, we will detail our approach. First, the network for feature extraction is introduced. Second, the aggregation method of multilevel and multiscale features is introduced. Finally, the constitution classification algorithm is introduced.

2.1. Feature Extraction Network. In the case of limited medical image dataset, it is likely that the convolutional

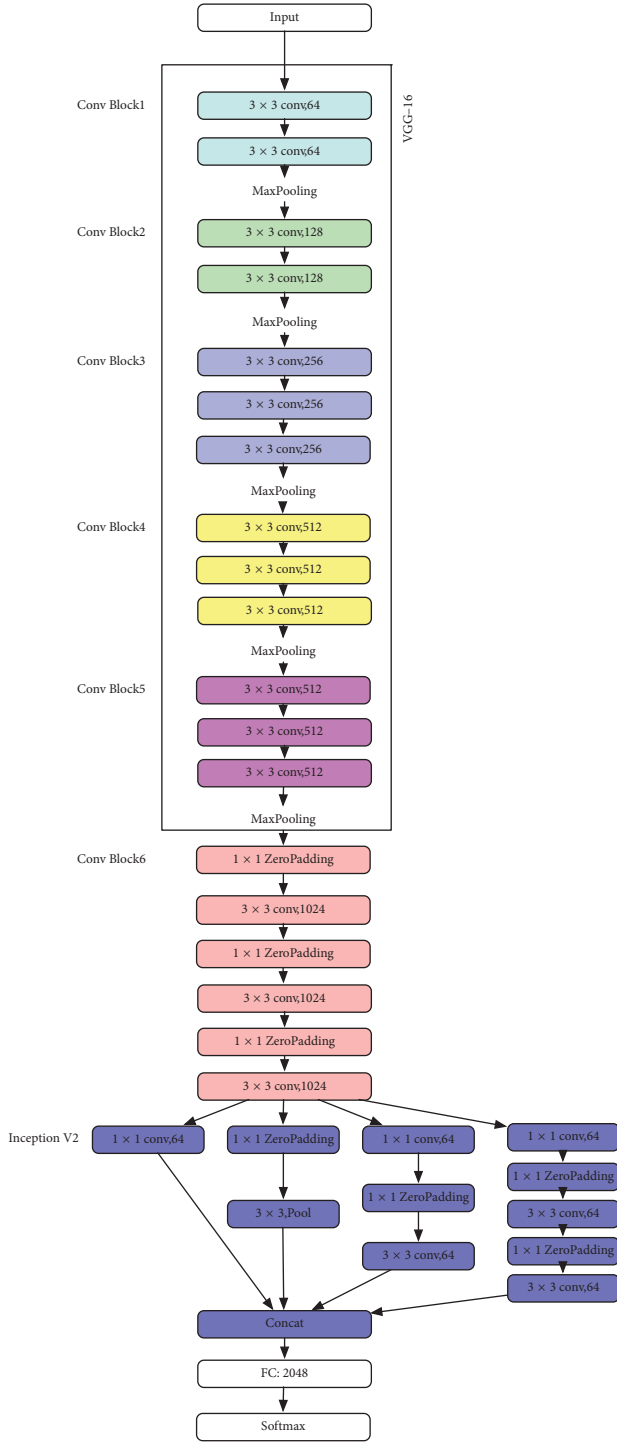


FIGURE 2: The architecture of CNN improved from VGG16.

neural network could not learn the features of the image well during the training process. Because the pretrained CNN network was trained on millions of different images, it contains powerful generic feature extraction filters. In order to extract the optimized features from the image, we used the well-verified CNN architecture, namely, VGG16 and NASNetMobile. These networks contain the max pooling and cascaded convolution layers. The total number of layers

depends on each network. VGG is known for its elegance and simplicity, while it has near state-of-the-art results in image classification and good versatility. The VGG16 is the runner-up in the 2014 large-scale visual recognition challenge. In our work, we used a pretrained VGG16 network with 16 weight layers, 13 convolution layers, and 3 fully connected layers. Then, we changed the top layer as we needed, by adding the Conv Block and Inception V2 modules. The Conv Block contains 3 ZeroPadding layers and 3 convolution layers. The Inception V2 module contains a 1×1 convolutional layer that reduces the amount of computation while increasing the network width. By adding the Conv Block and Inception V2 modules, the network has a stronger feature representation capability for the better classification. The network structure is shown in Figure 2. The image size entered in this network is 48×48 .

2.2. Multilevel and Multiscale Feature Aggregation. The pretrained CNN can be considered as a feature extractor. A single CNN model can extract features of different layers for the given input size. As shown in Figure 2, the network contains six Conv Block modules, one Inception V2 module, and one fully connected layer. We train the network on the training dataset and then extract the features of different layers. First, for a given 48×48 image, the feature map sizes of the 3rd, 6th, 10th, 14th, and 24th layers are 24×24 , 12×12 , 6×6 , 3×3 , and 1×1 , respectively. This article shows the feature map of the 3rd, 6th, 10th, and 14th layers, as shown in Figure 3, respectively. In this paper, we use the information of the middle layers, namely, the 14th and 24th layers are selected. However, the number of features for the 14th and 24th layers is (3, 3, 512) and (1, 1, 1024), respectively. In order to be able to carry out further feature dimensionality reduction, it is necessary to flat the dimensions of the 14th and 24th layers, leading to 4608 and 1024 features, respectively. Subsequently, PCA is applied to these features. It can be seen from Figure 3 that in the lower layer, the features are more and complicated, while features in the higher layers are simple, only the distinguishing features are needed. The abstraction levels of different layer features are different. Obviously, the abstraction leads to information loss, so that the features of different layers should be merged to compensate for the loss.

At the same time, we use the pretrained NASNetMobile network for feature extraction. The network consists of two modules: normal cell and reduction cell. The stacking of these two modules is then applied to form the whole network. In this network, we resized the image to 224×224 and then inputted it into the network. After training, it needs to extract the previous layer features of the global average pooling layer, in which PCA is applied to perform dimensionality reduction.

Finally, we aggregate the features of the 14th and 24th layers, features of the previous layer of the global average pooling layer, and features of the fully connected layer as shown in Figure 2 to obtain the aggregated features. The aggregated features are then entered into the classifier.

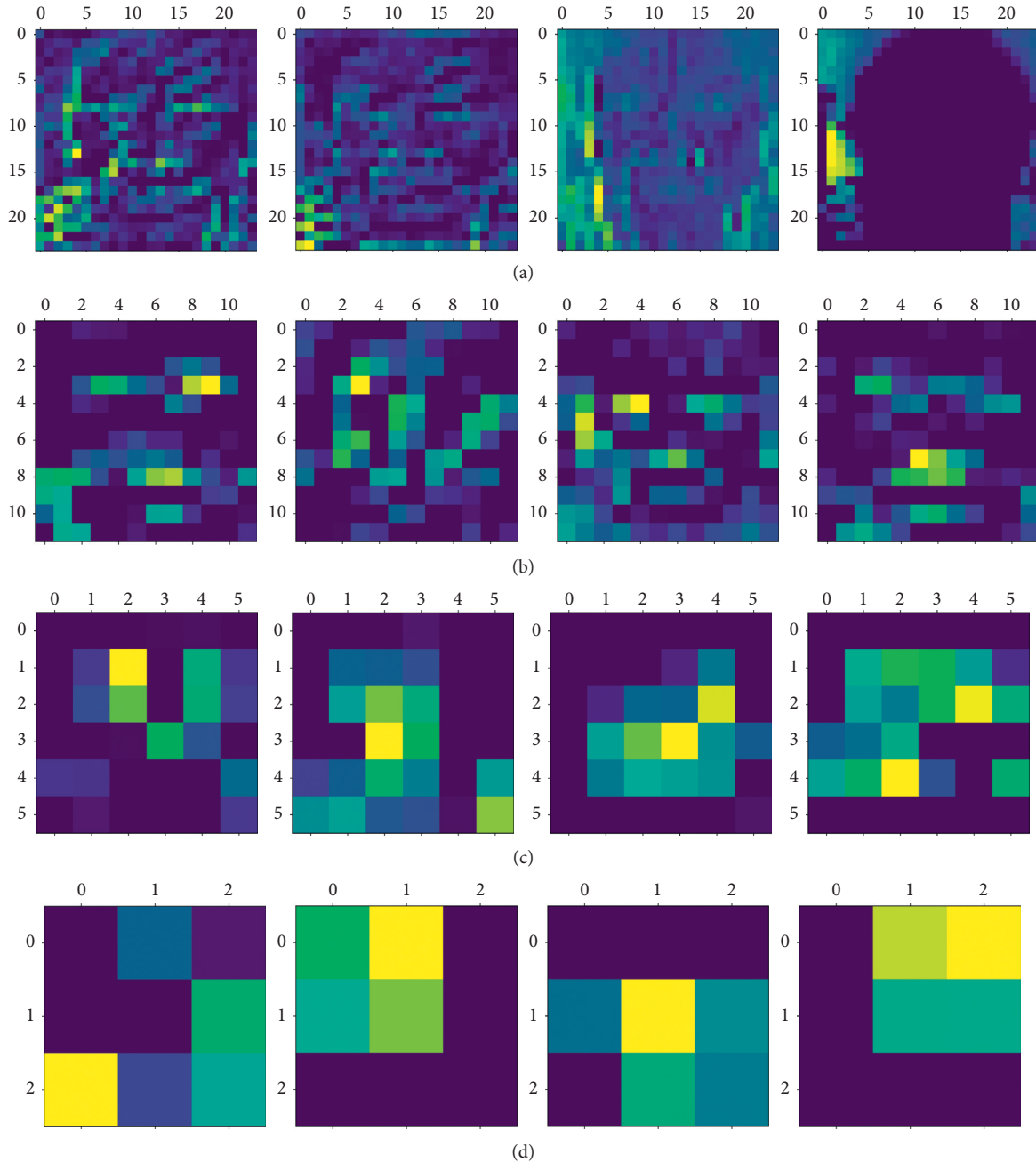


FIGURE 3: The feature maps of different layers: (a) layer 3, (b) layer 6, (c) layer 10, and (d) layer 14.

2.3. Constitution Classification. In this step, we predict the type of constitution based on the multilevel and multiscale aggregated features. There are lots of pattern classification algorithms, such as support vector machine(SVM) [32], K nearest neighbor(KNN) [33], Bayesian classifier(NB) [34], decision tree(DT) [35], logistic regression(LR) [36], and random forest algorithm(RF) [37]. Ensemble learning improves the effectiveness of machine learning by combining several models. This method can provide the better prediction results than a single model. This article also uses ensemble learning methods for classification, such as XGBoost [38], LightGBM [39], and CatBoost [40]. In our

experiments, we evaluated the classification effects of different classifiers.

3. Experiments

3.1. Dataset. The face data used in this paper have 21,150 pictures, which are obtained from the Chinese medicine clinic of the three hospitals, in which each facial image of the patient is assigned a constitution type by a professor of Chinese medicine. The identification of the constitution type is based on the national judgment criteria [41] for TCM constitution. Before collecting data, the standard is discussed

by nearly ten medical experts. Some agreed with this standard. Some professors were partially in favor of the standard. Some professors have a negative attitude on this standard. We chose three professors who were in favor of this standard. This means that they reached the consensus (agreement of standard) to determine the type of body constitution. Subsequently, they were in different hospitals to judge the patient's body constitution according to the standard. In this way, the impact of experience can be reduced as much as possible. Besides, these professors are well known and their ages are close and the personal experience is not greatly different. Finally, the body constitution type of the patient in the same hospital is determined by the same medical professor. The entire dataset is determined by three Chinese medicine professors from three different hospitals according to the abovementioned standard.

Therefore, all face images are taken by the same type of digital device and the patient's constitution type is specified by the doctor. The indoor environment is no sunshine, and lighting conditions are normal fluorescent lamps. In the face database, there are 8 kinds of constitution types, that is, gentleness, Qi-deficiency, Qi-depression, dampness-heat, phlegm-dampness, blood-stasis, Yang-deficiency, and Yin-deficiency. The number of samples with each constitutional type is given in Table 1. Samples with constitution types are shown in Figure 4. In the preprocessing process, the face detection algorithm is used to detect the acquired picture and the corresponding bounding box is obtained. Considering both time complexity and precision, this paper uses the OpenCV tool to complete the face detection. The test dataset used in this paper is the test dataset used in [22], and the training dataset does not overlap with the test dataset.

3.2. Data Augmentation. This article uses the data augmentation when training the VGG16 networks and NAS-NetMobile networks. In this paper, the width and height of each facial image are scaled proportionally and the image is zoomed in both length and width direction. This paper uses the Keras [42] tool to achieve data augmentation through the functions it contains. It just sets the values of `width_shift_range`, `height_shift_range`, and `zoom_range` in the `ImageDataGenerator` function. After data augmentation, it trains the network on these training samples through transfer learning.

3.3. Training Details. The tools used in this experiment are Keras, TensorFlow [43], Scikit-learn [44], and Scikit-image [45]. The GPU is NVIDIA GTX Titan X, the memory size is 12 GB, and the operating system is Ubuntu 14.04. The VGG16 and NASNetMobile networks are with the same setting. They are trained by the random gradient method. The learning rate is 0.0002, the momentum is set to 0.9, and the batch size is set to 30. In data augmentation processing, the values of `width_shift_range`, `height_shift_range`, and `zoom_range` are all set to 0.2.

3.4. Experimental Results and Discussion. In the previous works, many traditional feature extraction methods have

been applied to perform the constitution recognition. In order to show the superiority of deep learning methods to traditional feature extraction methods, lots of experiments are conducted to make comparison with them. Traditional facial feature extraction methods include color, texture, histogram of oriented gradient, and so on. Here, the color feature is represented by the HSV space and the texture feature is represented by local binary patterns. The classifiers in this article are described in Section 2.3. The settings of the classifier are as follows: RBF is selected in the support vector machine. The learning rate of the CatBoost classifier is 0.05, and the depth is 6. The learning rate is 0.05 in the Xgboost classifier, and the `max_depth` is 6. In the LightGBM classifier, the learning rate is 0.009, the `max_depth` is 8, the `lambda_l1` and `lambda_l2` are set to 0.9, and the `num_leaves` is 90. CatBoost, Xgboost, and LightGBM are abbreviated as Catb, Xgb, and Lgb in our paper, respectively. In order to make comparison among different feature extraction methods through experiments, we use the same classifier.

It can be seen from Table 2 that under the premise of the same classifier, the classification effect based on the VGG16-CI network extraction feature is better than that based on the single HSV, LBP, and HOG features. On the contrary, under the same feature extraction method, the classification effects of different classifiers are compared. Based on the single HSV feature and the LBP feature, SVM has the best classification accuracy. Based on the single HOG feature, Random Forest classification has the best performance. Based on the features extracted by the VGG16-CI network, the LightGBM has the best classification effect. Overall, the VGG16-CI network is far better than other feature extraction methods. At the same time, the confusion matrix of each classifier is expressed as follows. It can be seen from Tables 3 and 4 that the selected classifier has a good classification effect on the Qi-deficiency and a poor classification effect on the gentleness. This is because the quality of gentleness is affected by the Qi-deficiency. It can be seen from Table 5 that the classifier has a good classification effect on the Yin-deficiency. The effect on the gentleness is poor which is also affected by the Yin-deficiency. It can be seen from Table 6 that the classifier has a good classification effect on the Qi-deficiency. The effect on the gentleness is poor which is also affected by the Qi-deficiency. It can be seen from Table 2 that the classification of LightGBM is the best, whose confusion matrix is presented as given in Table 6. Therefore, in the following experiments, LightGBM is selected as the classifier for constitution recognition.

In this paper, the VGG-CI model is obtained by transfer learning based on VGG16. The features of different layers of the VGG-CI model are extracted to classify the facial images. In more detail, the features of the 14th layer, 24th layer, and fully connected layer are selected. Subsequently, the PCA method is used to reduce the dimensionality of features of the 14th layer and the 24th layer, respectively. The number of merged features is 100. Each is then merged with the features of the fully connected layers. Finally, these merged features are applied to perform the classification, in which the classifier is the LightGBM. The classification results are shown in Table 7. It can be seen that the classification

TABLE 4: The confusion matrix of the SVM algorithm based on the LBP feature.

	Qi-deficiency	Yin-deficiency	Yang-deficiency	Phlegm-dampness	Dampness-heat	Qi-depression	Blood-stasis	Gentleness
Qi-deficiency	75	0	0	0	0	0	0	0
Yin-deficiency	14	61	0	0	0	0	0	0
Yang-deficiency	25	0	35	0	0	0	0	0
Phlegm-dampness	15	0	0	60	0	0	0	0
Dampness-heat	23	0	0	0	52	0	0	0
Qi-depression	44	0	0	0	0	31	0	0
Blood-stasis	9	0	0	0	0	0	32	0
Gentleness	57	0	0	0	0	0	0	0

TABLE 5: The confusion matrix of the random forest algorithm based on the HOG feature.

	Qi-deficiency	Yin-deficiency	Yang-deficiency	Phlegm-dampness	Dampness-heat	Qi-depression	Blood-stasis	Gentleness
Qi-deficiency	44	17	0	9	3	2	0	0
Yin-deficiency	5	64	0	4	1	1	0	0
Yang-deficiency	16	2	35	4	1	1	0	1
Phlegm-dampness	7	4	0	63	0	1	0	0
Dampness-heat	11	3	0	5	53	3	0	0
Qi-depression	29	5	0	7	1	33	0	0
Blood-stasis	5	1	0	2	1	0	32	0
Gentleness	27	10	0	12	5	3	0	0

TABLE 6: The confusion matrix of the LightGBM algorithm based on the VGG-CI feature.

	Qi-deficiency	Yin-deficiency	Yang-deficiency	Phlegm-dampness	Dampness-heat	Qi-depression	Blood-stasis	Gentleness
Qi-deficiency	68	3	1	1	2	0	0	0
Yin-deficiency	3	66	0	2	3	1	0	0
Yang-deficiency	15	3	36	1	4	0	1	0
Phlegm-dampness	7	2	0	63	1	1	0	1
Dampness-heat	13	0	3	1	58	0	0	0
Qi-depression	13	3	1	3	6	41	0	8
Blood-stasis	7	1	0	0	0	1	32	0
Gentleness	29	6	2	5	4	7	2	2

TABLE 7: Classification results of different layers.

Layer	Accuracy (%)
14 th layer	53.65
24 th layer	66.97
FC layer	68.67
(14 th layer + PCA) + FC	68.48
(24 th layer + PCA) + FC	68.85
(14 th layer + PCA) + (24 th layer + PCA) + FC	69.04
(14 th layer + PCA) + (24 th layer + PCA) + (Conv2D + PCA) + FC	69.61

accuracy of the 14th layer is very low, and the classification effect reaches 68.67% in the fully connected layer. At the same time, the features of the 14th layer and the 24th layer are dimensionally reduced by PCA, respectively, which is then applied to merge with the features of the fully connected

layer. In such case, the classification effect is further improved.

On the contrary, we also use the transfer learning to further train the NASNetMobile network, extracting the features of the previous layer of the global average pooling layer. This layer is represented by Conv2D, whose features are flattened, and then PCA is applied to obtain fewer features. These features, the features of the 14th layer after PCA, the features of the 24th layer after PCA, and fully connected layers, are aggregated, and the classification accuracy reached 69.61%.

In order to visually analyze the performance of our method, t-SNE is applied to display the extracted features form on the training dataset, as shown in Figure 5. It can be seen from Figure 5 that as to the features of the 14th layer, the difference can be observed between the original features and dimensionally reduced features by PCA. It validates the

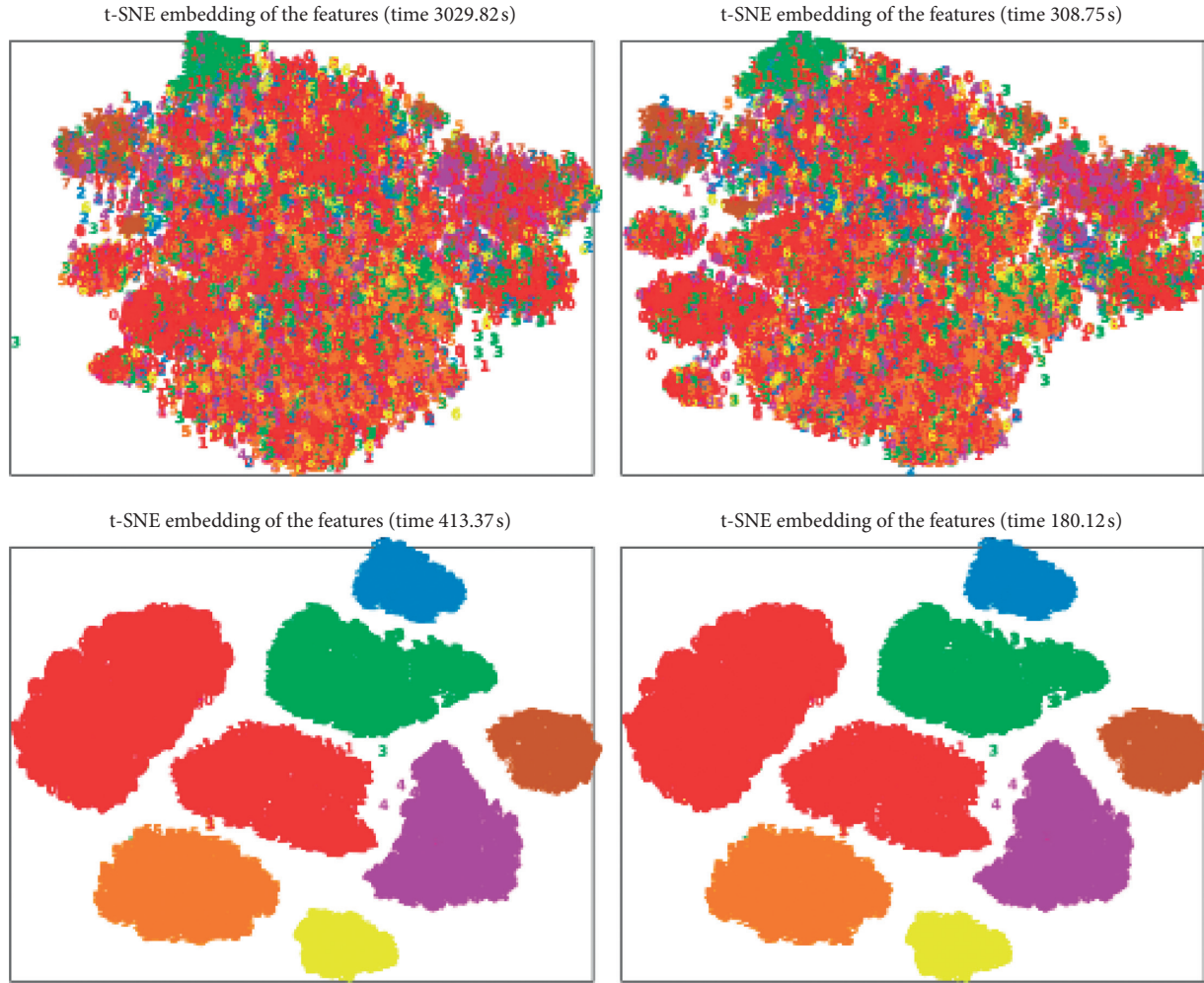


FIGURE 5: Visualization of the extracted features for the training data by t-SNE. The first row is for the 14th layer, and the second row is for the 24th layer. The left column is the original features, and the right column is the features after PCA.

usage of PCA. However, as to the 24th layer features, there is no significant change after the PCA is used. This is because the original features have been good enough. This illustrates that PCA should be applied appropriately.

Now, there are some well-known deep learning networks, such as VGG19 [29], Inception v3 [31], ResNet [46], InceptionResnet [47], Xception [48], MobileNet v1 [49], DenseNet [50], EfficientNets [51], and NASNet [30]. In order to verify the classification effect of the proposed algorithm, we use the transfer learning to train these networks with the same facial training dataset. The classifier is also the LightGBM that has the same setting. The experimental results are shown in Table 8. It is easy to see that the algorithm this paper proposed works best, and the accuracy rate is 69.61%. Moreover, in the case of the same test dataset, the method proposed in this paper is compared with the method of [22]. The literature [22] aggregated the features extracted from convolutional neural networks with the traditional color feature. Its classification accuracy was 65.29%. The algorithm proposed in this paper adopts the multiscale and multilayer feature aggregation

TABLE 8: Classification results of different models.

Model	Accuracy (%)
VGG19 [29]	59.47
Inception v3 [31]	64.17
ResNet-50 [46]	65.47
InceptionResnet [47]	62.48
MobileNet v1 [49]	63.98
Xception [48]	64.54
DenseNet-121 [50]	62.47
DenseNet-169 [50]	63.04
DenseNet-201 [50]	64.16
NASNetMobile [30]	62.85
EfficientNetsB0 [51]	63.23
Lit. [22]	65.29
Our proposed method	69.61

method, so that the better classification accuracy is obtained up to 69.61%. The confusion matrix of the proposed algorithm is shown in Table 9. It can be seen that the quality of the gentleness is still affected by the Qi-deficiency.

TABLE 9: The confusion matrix of the algorithm proposed in this paper.

	Qi-deficiency	Yin-deficiency	Yang-deficiency	Phlegm-dampness	Dampness-heat	Qi-depression	Blood-stasis	Gentleness
Qi-deficiency	68	3	1	1	2	0	0	0
Yin-deficiency	4	66	0	1	3	1	0	0
Yang-deficiency	12	4	36	2	4	1	1	0
Phlegm-dampness	7	2	0	63	1	1	0	1
Dampness-heat	13	0	3	1	58	0	0	0
Qi-depression	7	3	2	3	6	46	0	8
Blood-stasis	6	1	0	0	0	2	32	0
Gentleness	26	6	2	5	4	10	2	2

4. Conclusion and Future Work

In this paper, we propose a constitution classification algorithm based on a newly designed convolutional neural network, which makes full use of the information of different layers in the network. We also demonstrate the effectiveness of our approach by merging multilevel and multiscale features. Finally, the method proposed in this paper achieves the best results on the test set. In the future, we plan to improve our designed network by using the new learning method. At the same time, the impact of Qi-deficiency on gentleness should be considered.

Data Availability

The TCM data used in this study can be obtained by contacting the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

References

- [1] Y. W. Ding, "Development of constitution theory in TCM," *Yunnan Journal of Traditional Chinese Medicine and Materia Medica*, no. 2, pp. 71–75, 2010.
- [2] Q. Wang, "Three key issues in the study of TCM constitution (part I)," *Journal of Traditional Chinese Medicine*, vol. 47, no. 4, pp. 250–252, 2006.
- [3] Q. Wang, "Three key issues in the study of TCM constitution (part II)," *Journal of Traditional Chinese Medicine*, vol. 47, no. 5, pp. 329–332, 2006.
- [4] China Association of Chinese Medicine, "Classification and identification of constitution theory of TCM (ZYXXH/T157-2009)," *World Journal of Traditional Chinese Medicine*, vol. 4, pp. 303–304, 2009.
- [5] R. X. Yu, Q. Wang, J. Wang et al., "An analysis of the status quo of application of constitution identification," *Chinese Journal of Information on Traditional Chinese Medicine*, vol. 2, pp. 107–109, 2013.
- [6] X. Liu and Q. Wang, "Suggestion and analysis on revise of standard of classification and determination of constitution in TCM," *Beijing University of Chinese Medicine*, vol. 36, no. 5, pp. 300–304, 2013.
- [7] S.-Y. Su, C.-H. Yang, C.-C. Chiu, and Q. Wang, "Acoustic features for identifying constitutions in traditional Chinese medicine," *The Journal of Alternative and Complementary Medicine*, vol. 19, no. 6, pp. 569–576, 2013.
- [8] S. Hou, J. Zhang, P. Li et al., "Research on TCM constitution classification based on facial color and texture," *Biomedical Research*, vol. 28, no. 10, pp. 4645–4650, 2017.
- [9] G. L. Lu, Y. S. Huang, Q. Zhang et al., "The study of auxiliary TCM constitution identification model based on tongue image and physical features," *LISHIZHEN Medicine and Material Medica Research*, vol. 30, no. 1, pp. 244–246, 2019.
- [10] Y. C. Wang and L. N. Bai, "Classification of body constitution of pulse signal in TCM based on BP neural network," *Journal of Traditional Chinese Medicine*, vol. 55, no. 15, 2014.
- [11] H. Y. Zhang, S. Li, X. Y. Wang et al., "Dynamic classification model based on correlation recognition —an example of skin and Traditional Chinese Medicine constitution," *Advanced Engineering Sciences*, vol. 49, no. 3, pp. 137–143, 2017.
- [12] L. Song, J. Liu, B. Qian et al., "A deep multi-modal CNN for multi-instance multi-label image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 12, pp. 6025–6038, 2018.
- [13] M. Zhang, W. Li, and Q. Du, "Diverse region-based CNN for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2623–2634, 2018.
- [14] J. Dolz, K. Gopinath, J. Yuan et al., "HyperDense-Net: a hyper-densely connected CNN for multi-modal image segmentation," *IEEE transactions on medical imaging*, vol. 38, no. 5, pp. 1116–1126, 2018.
- [15] L. Chen, P. Bentley, K. Mori, K. Misawa, M. Fujiwara, and D. Rueckert, "DRINet for medical image segmentation," *IEEE Transactions on Medical Imaging*, vol. 37, no. 11, pp. 2453–2462, 2018.
- [16] Z. Cai and N. Vasconcelos, "Cascade R-CNN: delving into high quality object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 6154–6162, Salt Lake City, UT, USA, June 2018.
- [17] Y. Chen, W. Li, C. Sakaridis et al., "Domain adaptive faster R-CNN for object detection in the wild," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3339–3348, Salt Lake City, UT, USA, June 2018.
- [18] K. H. Jin and J. M. Frahm, "Hierarchy of alternating specialists for scene recognition," in *Proceedings of the European Conference on Computer Vision*, pp. 451–467, Munich, Germany, September 2018.
- [19] X. Cheng, J. Lu, J. Feng, B. Yuan, and J. Zhou, "Scene recognition with objectness," *Pattern Recognition*, vol. 74, pp. 474–487, 2018.
- [20] X. Hu, H. Zhu, J. Xu, D. Xu, and J. Dong, "Wrist pulse signals analysis based on deep convolutional neural networks," in *Proceedings of the IEEE Conference on Computational*

- Intelligence in Bioinformatics and Computational Biology*, pp. 1–7, Honolulu, HI, USA, 2014.
- [21] H. Li, B. Xu, N. Wang et al., “Deep convolutional neural networks for classifying body constitution,” in *Proceedings of the Springer International Conference on Artificial Neural Networks*, pp. 128–135, Barcelona, Spain, September 2016.
- [22] E. Y. Huan, G. H. Wen, S. J. Zhang et al., “Deep convolutional neural networks for classifying body constitution based on face image,” *Computational and Mathematical Methods in Medicine*, vol. 2017, Article ID 9846707, 9 pages, 2017.
- [23] H. H. Li, G. H. Wen, and H. B. Zeng, *Natural Tongue Physique Identification Using Hybrid Deep Learning Methods*, Springer, Berlin, Germany, 2018.
- [24] J. L. Hu, Y. T. Ding, and H. X. Kan, “Constitution classification based on machine learning through tongue image,” *Journal of Jiamusi University (Natural Science Edition)*, vol. 36, no. 5, pp. 709–713, 2018.
- [25] H. Zhou, G. Hu, and X. Zhang, “constitution identification of tongue image based on CNN,” in *Proceedings of the 2018 11th International Congress on Image and Signal Processing*, pp. 1–5, BioMedical Engineering and Informatics (CISP-BMEI), IEEE, Beijing, China, October 2018.
- [26] J. Ma, G. Wen, C. Wang, and L. Jiang, “Complexity perception classification method for tongue constitution recognition,” *Artificial Intelligence in Medicine*, vol. 96, pp. 123–133, 2019.
- [27] J. Deng, W. Dong, R. Socher et al., “Imagenet: a large-scale hierarchical image database,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, Miami, FL, USA, June 2009.
- [28] N. Tajbakhsh, J. Y. Shin, S. R. Gurudu et al., “Convolutional neural networks for medical image analysis: full training or fine tuning,” *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1299–1312, 2016.
- [29] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” 2014, <https://arxiv.org/abs/1409.1556>.
- [30] B. Zoph, V. Vasudevan, J. Shlens et al., “Learning transferable architectures for scalable image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 8697–8710, Salt Lake City, UT, USA, June 2018.
- [31] C. Szegedy, V. Vanhoucke, S. Ioffe et al., “Rethinking the inception architecture for computer vision,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2818–2826, Las Vegas, NV, USA, July 2016.
- [32] Y. Chen, G. Tao, H. Ren, X. Lin, and L. Zhang, “Accurate seat belt detection in road surveillance images based on CNN and SVM,” *Neurocomputing*, vol. 274, pp. 80–87, 2018.
- [33] Y. Ma, Q. Xie, Y. Liu, and S. Xiong, “A weighted KNN-based automatic image annotation method,” *Neural Computing and Applications*, pp. 1–12, 2019.
- [34] X. Liu, R. Zhang, Z. Meng, R. Hong, and G. Liu, “On fusing the latent deep CNN feature for image classification,” *World Wide Web*, vol. 22, no. 2, pp. 423–436, 2019.
- [35] R. B. Oliveira, J. P. Papa, A. S. Pereira, and J. M. R. S. Tavares, “Computational methods for pigmented skin lesion classification in images: review and future trends,” *Neural Computing and Applications*, vol. 29, no. 3, pp. 613–636, 2018.
- [36] D. Han, Q. Liu, and W. Fan, “A new image classification method using CNN transfer learning and web data augmentation,” *Expert Systems with Applications*, vol. 95, pp. 43–56, 2018.
- [37] C. S. Marquesan, M. Karuppiah, and T. Prathiba, “A multi resolution transform for thermal face recognition using random forest classifier,” *Journal of Computational and Theoretical Nanoscience*, vol. 15, no. 3, pp. 785–792, 2018.
- [38] T. Che and C. Guestrin, “Xgboost: a scalable tree boosting system,” in *Proceedings of the 22nd Acm Sigkdd international Conference on Knowledge Discovery and Data Mining*, pp. 785–794, ACM, San Francisco, CA, USA, August 2016.
- [39] G. Ke, Q. Meng, T. Finley et al., “LightGBM: a highly efficient gradient boosting decision tree,” *Advances in Neural Information Processing Systems*, pp. 3146–3154, 2017.
- [40] A. V. Dorogush, V. Ershov, and A. Gulin, “CatBoost: gradient boosting with categorical features support,” 2018, <https://arxiv.org/abs/1810.11363>.
- [41] Q. Wang, “Classification and diagnosis basis of nine basic constitutions in Chinese medicine,” *Journal-Beijing University of Traditional Chinese Medicine*, vol. 28, no. 4, pp. 1–8, 2005.
- [42] F. Chollet, “Keras,” 2015, <https://github.com/fchollet/keras>.
- [43] M. Abadi, P. Barham, J. Chen et al., “Tensorflow: a system for large-scale machine learning,” in *Proceedings of the 12th {USENIX} Symposium on Operating Systems Design and Implementation*, pp. 265–283, Savannah, GA, USA, November 2016.
- [44] F. Pedregosa, G. Varoquaux, A. Gramfort et al., “Scikit-learn: machine learning in python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [45] V. W. Stéfan, L. Johannes, J. N. Schönberger et al., “Scikit-image: image processing in python,” 2014.
- [46] K. He, X. Zhang, S. Ren et al., “Deep residual learning for image recognition,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778, Las Vegas, NV, USA, July 2016.
- [47] C. Szegedy, S. Ioffe, V. Vanhoucke et al., “Inception-v4, inception-resnet and the impact of residual connections on learning,” in *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, San Francisco, CA, USA, February 2017.
- [48] F. Chollet, “Xception: deep learning with depthwise separable convolutions,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1251–1258, Honolulu, HI, USA, July 2017.
- [49] A. G. Howard, M. Zhu, B. Chen et al., “Mobilenets: efficient convolutional neural networks for mobile vision applications,” 2017, <https://arxiv.org/abs/1704.04861>.
- [50] G. Huang, Z. Liu, V. D. Maaten et al., “Densely connected convolutional networks,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4700–4708, Honolulu, HI, USA, July 2017.
- [51] M. Tan and Q. V. Le, “EfficientNet: rethinking model scaling for convolutional neural networks,” 2019, <https://arxiv.org/abs/1905.11946>.