Comparative Analysis of Convolutional Neural Networks on The Stability and Performance for Micro-Expression Recognition

Bangwei Tang

School of Computer Science and Engineering, University of Electronic Science and Technology of China, Chengdu, Sichuan, 610000, China

2021080905012@std.uestc.edu.cn

Abstract. Micro-expression recognition can reveal many aspects of a person, such as emotions and mental states. They have been widely used in various fields, including computer vision and natural language processing. As deep learning has become increasingly popular recently, it has started to be used for micro-expression recognition to further improve its performance. In this paper, the research focuses on three typical convolutional neural networks for micro-expression recognition: VGG16, VGG19, and MobileNetV3-Small. This research aims to compare the stability of these three convolutional neural networks on uneven micro-expression datasets and to evaluate their real performance in micro-expression recognition. The research utilizes the publicly available real-world Affective Faces (RAF-DB) dataset to train the three convolutional neural networks. Two data preprocessing methods are employed in the study: one is adjusting the class weights to reduce the impact of uneven micro-expression datasets, and the other is using data augmentation to increase the size of the micro-expression datasets. When examining the stability of the three convolutional neural networks on uneven micro-expression datasets, the accuracy rates before and after adjusting the weights were 78% and 76% for VGG16, 76% and 76% for VGG19, and 75% and 72% for MobileNetV3-Small, respectively. Regarding the real performance of the three convolutional neural networks, the accuracy rates for VGG16, VGG19, and MobileNetV3-Small were 78%, 80%, and 75%, respectively. The results indicate that VGG16 and MobileNetV3-Small have stronger stability in dealing with uneven micro-expression datasets, while VGG19 performs better when the micro-expression datasets are sufficiently large.

Keywords: Micro-expression recognition; convolutional neural networks; data augmentation.

1. Introduction

Facial expressions play a very important role in human communication. They have a strong relationship with human minds, moods, attitudes, and mental states. What people see in daily life is a common type of facial expression called a macro-expression [1]. Macro-expressions last on the face for a long time, are very obvious, and can be easily recognized. However, people tend to disguise their macro-expressions to conceal their real emotions [2]. Therefore, macro-expressions may sometimes mislead people into making wrong judgments. Another type of facial expression is called a micro-expression. Micro-expressions differ from macro-expressions because they only last for a very short time, between 1/25 and 1/5 of a second [3]. Also, people tend to generate micro-expressions unconsciously and can hardly control or disguise them. As a result, micro-expressions usually reveal a person's true emotions. In addition, micro-expressions are universal across races and cultures because each micro-expression is connected to the specific movements of certain muscles, which means every micro-expression has its own distinct features. Therefore, research into facial expression recognition, especially micro-expression recognition, is very meaningful. Nowadays, micro-expression recognition spans many fields, including computer vision, natural language processing, physiology, acoustics, and neuroscience [4]. It is applied in various industries such as fraud detection, crime investigation, psychiatry, and education. Deep learning further improves the performance of micro-expression recognition [3]. Convolutional neural networks are an excellent type of neural network for image classification tasks [5]. For images, convolutional neural networks use two-dimensional convolutional kernels to extract height and width features for image prediction [6]. For

fixed-length videos, convolutional neural networks use three-dimensional convolutional kernels to better extract both temporal and spatial information in the video. For videos of flexible duration, another type of neural network called a recurrent neural network is used to process time series data [7].

Micro-expression recognition is a type of image classification task. Many convolutional neural networks, such as VGG16, VGG19, ResNet50, and MobileNetV3, are widely used for micro-expression recognition. While micro-expression recognition has achieved some progress, there are still some challenges. For example, limited and uneven micro-expression datasets make it difficult for convolutional neural networks to extract features of each micro-expression class, which limits the accuracy of micro-expression recognition [8, 9]. In addition, each convolutional neural network has its own strengths and weaknesses. Higher accuracy requires more computing resources. To achieve better performance, there is still room for improvement.

The main contributions of this research are as follows: 1) Comparing the performance of VGG16, VGG19, and MobileNetV3-Small on uneven micro-expression datasets by adjusting the class weights. VGG16 and MobileNetV3-Small exhibit the best stability with uneven micro-expression datasets. 2) Comparing the real performance of VGG16, VGG19, and MobileNetV3-Small on micro-expression datasets after data augmentation. VGG19 demonstrates the best performance in terms of accuracy.

2. Methodology

2.1. Dataset and Data Preprocessing

The research was conducted on the dataset called Real-world Affective Faces (RAF-DB) collected from Kaggle. RAF-DB is a real-world micro-expression dataset, which is made up of 29672 facial images. Each image is labeled by 40 labelers as single-label image or double-label image. Subjects in RAF-DB vary in age, gender, race. All the subjects are between 0 and 70 years old, and the percentage of male, female and uncertain gender is 52%, 43% and 5% respectively. Images in RAF-DB are classified into single-label images and double-label images. Single-label images are divided into seven emotion classes, including anger, disgust, fear, happiness, neutral, sadness, surprise. Double-label images are divided into twelve compound-emotion classes [10].

The research only used single-label images as the dataset. There are 15339 single-label images in the dataset. 12271 single-label images were allocated to training set and 3068 single-label images were allocated to testing set. The research also split 2452 single-label images in the training set for validation.

In RAF-DB, the training set and testing set have uneven class sizes. Classes like happiness and neutral have much more images than classes like anger, disgust and fear. In addition, the total number of training set and testing set is not large enough. The similar defects also exist in other publicly available micro-expression datasets such as Facial Expression Recognition 2013 Dataset (FER2013) and Expression in-the-Wild (ExpW) Dataset [10]. Therefore, the research preprocessed the dataset.

The data preprocessing included two aspects. One aspect is adjusting the weight of each class [11]. The research adjusted the weight of each class according to the percentage of the class in the dataset. First, the research divided the total number of the images in RAF-DB by the number of the images in each class to get each class weight. Then the research normalized the class weights. Through this method, the weight of the minority classes was set higher to get more attention from the model during the training. The weight of each class is shown in Table 1.

Table 1. Weight of Each Class

Class	anger	disgust	fear	happiness	neutral	sadness	surprise
Weight	0.171	0.169	0.431	0.026	0.048	0.061	0.094

The other aspect is data augmentation [12]. There are a lot of kinds of data augmentation, including geometric transformations, random erasing, kernel filters and so on [12]. The research used geometric

transformations to increase the scale of the dataset, including flipping, cropping, rotation, shifting and zooming. The concrete data augmentation is shown below: 1) Randomly rotate the image by an angle between -45 degrees and 45 degrees. 2) Randomly shift horizontally the image by a distance between -0.2 times the image width and 0.2 times the image width. 3) Randomly shift vertically the image by a distance between -0.2 times the image height and 0.2 times the image height. 4) Randomly shear the image between -0.2 radians and 0.2 radians. 5) Randomly zoom the image by a size between 0.8 times the image size and 1.2 times the image size. 6) Horizontally flip the image.

2.2. Models

2.2.1 VGG16

VGG16 is a kind of convolutional neural network called visual geometry group. There are totally 13 convolution layers and 3 fully-connected layers in VGG16 [3], which are divided into 5 convolution blocks and 1 fully-connected block. The construction of VGG16 is relatively simple so that VGG16 has been used widely in different kinds of computer vision tasks like image classification and image generation.

2.2.2 VGG19

VGG19 is also a kind of visual geometry group. What is different from VGG16 is that VGG19 has 16 convolution layers. But both of them have 3 fully-connected layers [3]. Compared to VGG16, VGG19 has more layers and parameters, which means it can perform better when dealing with complicated tasks.

2.2.3 MobileNetV3

The series of mobile net are lightweight convolutional neural networks. There are three versions of mobile net and MobileNetV3 is the newest version with both higher efficiency and higher accuracy at the same time. MobileNetV3 has two versions: MobileNetV3-Large and MobileNetV3-Small. The difference between them is that MobileNetV3-Large has four more layers than MobileNetV3-Small, which means MobileNetV3-Large needs more resources and can provide better performance than MobileNetV3-Small. Since many convolutional neural networks need a lot of computing and hardware resources for image classification which can't fit mobile devices' limited resources, MobileNetV3 is often used to solve image classification problems in mobile devices. Some research shows that MobileNetV3 can achieve a great balance between efficiency and a accuracy for image classification tasks in mobile devices [13]. The research chose MobileNetV3-Small for comparison.

3. Experiments and Results

The research used the publicly available micro-expression dataset called RAF-DB with seven micro-expression classes. Accuracy, precision, recall and F1 score were used as evaluation indexes in the research. First, the research studied the stability of VGG16, VGG19, MobileNetV3-Small on uneven dataset RAF-DB. Then the research used data augmentation to enlarge the scale of the dataset and compared the performance of VGG16, VGG19, MobileNetV3-Small in terms of accuracy, precision, recall and F1 score, using the split ratio of 8:2.

3.1. Training Details

The research used PyCharm with python programming language. In addition, several python libraries including TensorFlow, Keras, NumPy, Pandas and Sklearn were used in the research. The hardware consisted of an NVIDIA GeForce RTX 3070 Laptop GPU, 40GB disk storage and 16GB memory.

Hyperparameters of VGG16, VGG19, MobileNetV3-Small in the research were adjusted to optimize their performance. The research used learning rate decay strategy to reduce the risk of overfitting. The final hyperparameters are shown in Table 2. Patience means that the model's accuracy on the validation set does not increase after a specified number of epochs. Factor means that

when validation set accuracy doesn't improve by patience epochs, the learning rate will be multiplied by this factor.

Table 2. Hyperparameter for convolutional neural network

	Epoch	Learning Rate	Batch Size	Optimizer	Activation Function	Image Size
VGG16	100	0.001(patience:10, factor:0.9)	32	Adam	Relu, Softmax	48*48*3
VGG19	100	0.001(patience: 10, factor: 0.9)	32	Adam	Relu, Softmax	48*48*3
MobileNetV3-Small	100	0.01(patience: 10, factor: 0.9)	32	Adam	Relu, Hard-Swish, Softmax	224*224*3

3.2. Experiment 1: Stability Comparison of Convolutional Neural Networks in Uneven Dataset

Initially, the research found that micro-expression datasets RAF-DB had uneven class sizes, which might limit the final performance and cause overfitting problems. Therefore, as to each convolutional neural network, the research compared the performance before adjusting the class weights and after adjusting the class weights which could help solve uneven datasets problem. The aim of this experiment is to study whether these convolutional neural networks have the ability to deal with uneven micro-expression datasets on their own. The evaluation performance is shown in Table 3.

Table 3. Evaluation performance of experiment 1

	Accuracy	Precision	Recall	F1 Score
VGG16(before)	78%	70%	66%	68%
VGG16(after)	76%	68%	65%	66%
VGG19(before)	76%	67%	65%	66%
VGG19(after)	76%	69%	66%	67%
MobileNetV3-Small(before)	75%	70%	57%	59%
MobileNetV3-Small(after)	72%	63%	59%	61%

From Table 3, it can be observed that as to VGG16 and MobileNetV3-Small, performance before adjusting the class weights was better than that after adjusting the class weights in terms of accuracy. The percentage increase of accuracy of VGG16 and MobileNetV3-Small before adjusting the class weights is 2% and 3% respectively. As to VGG19, performances before adjusting the class weights and after adjusting the class weights in terms of accuracy were the same. Through the results, it can be concluded that VGG16 and MobileNetV3-Small have the strong stability to keep their performances away from the impact of uneven micro-expression datasets. VGG19 has the weaker but still strong stability to uneven micro-expression datasets.

3.3. Experiment 2: Performance Comparison of Convolutional Neural Networks after Data Augmentation

Experiment 1 shows that all three convolutional neural networks have the ability to deal with uneven micro-expression datasets. As a result, experiment 2 didn't adjust the class weights. Moreover, since the initial micro-expression datasets RAF-DB didn't have enough images, experiment 2 used data augmentation to enlarge the scale of the micro-expression datasets. Then experiment 2 compared the performances of three convolutional neural networks based on enlarged micro-expression datasets. The aim of this experiment is to study the real performance of VGG16, VGG19, MobileNetV3-Small for micro-expression recognition without the interference of datasets' scale. The confusion matrix heat map and evaluation indexes are shown below.

As seen in Table 4 and Fig. 1, it can be observed that VGG19 outperformed VGG16 and MobileNetV3-Small in terms of accuracy, precision, recall, F1 score and MobileNetV3-Small performed relatively the worst. The percentage increase of accuracy of VGG19 is 2% when compared

with VGG16 while this increase is 5% as compared to MobileNetV3-Small. Through the result, it can be concluded that without the interference of datasets' scale, VGG19 has the best performance for micro-expression recognition among the three convolutional neural networks.

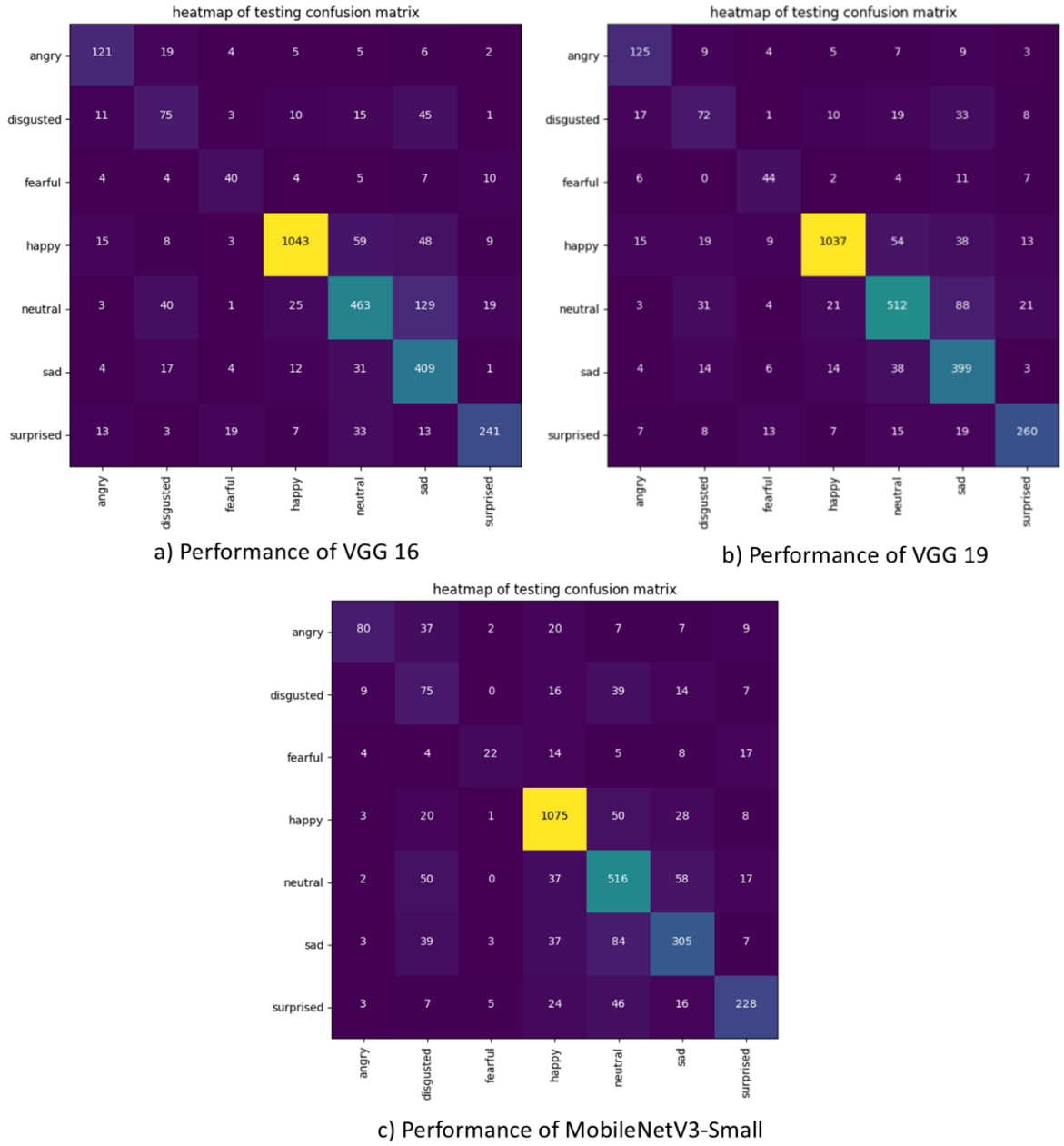


Fig. 1 Confusion matrixes of VGG16, VGG19, and MobileNetV3-Small models (Figure Credit: Original).

Table 4. Three Scheme comparing

Models	Accuracy	Precision	Recall	F1 Score
VGG16	78%	70%	70%	70%
VGG19	80%	71%	72%	71%
MobileNetV3	75%	69%	61%	63%

4. Discussion

The research shows that VGG19 has the best performance for micro-expression recognition. The first reason for this is that VGG19 has three more convolution layers than VGG16. More network layers improve VGG19's ability to learn more complicated and abstract features [14]. By extracting

more micro-expression features, VGG19 performs better that VGG16. The second reason is that MobileNetV3-Small is a lightweight convolutional neural network. It uses much less network layers and parameters than VGG19 for higher efficiency and less resource usage. As a result, MobileNetV3-Small has a relatively poor ability to learn complicated and abstract features than VGG19. VGG19 performs better than MobileNetV3-Small.

In this research, VGG19 had the best accuracy for micro-expression recognition. As a result, it is suggested that researchers use VGG19 to do micro-expression recognition among these three convolutional neural networks if they pursue high accuracy. In other researches, more convolutional neural networks like ResNet50 are introduced for image classification. ResNet50 has a better performance than VGG19 and VGG16 in terms of accuracy [7]. However, it has a more complicated construction and consumes more time and resources. Therefore, if researchers have enough time and resources and want higher accuracy, they can try some more complicated convolutional neural networks like ResNet50.

However, there are also some points to be improved. First of all, this research only compared VGG16, VGG19, MobileNetV3-Small in terms of accuracy, precision, recall and F1 score. If researchers also want to study the efficiency of neural networks for micro-expression recognition, they can measure the time and hardware resources which neural networks consume. Second, this research only compared three convolutional neural networks VGG16, VGG19, MobileNetV3-Small. Researchers can also use other convolutional neural networks like ResNet50, DenseNet-121 for performance comparison in the future. Third, uneven micro-expression datasets RAF-DB was used in this research. If researchers want to achieve a neural network with better performance, they can search for other evener and larger micro-expression datasets. In addition, limited by the hardware resources, this research changed the image size to 48*48*3 when training VGG16 and VGG19. In the future, researchers can change it back to 224*224*3 for further comparison if possible.

5. Conclusion

This research compared the performance of three typical convolutional neural networks—VGG16, VGG19, and MobileNetV3-Small—for micro-expression recognition. Since many publicly available micro-expression datasets have uneven class sizes, this research first compared the stability of these three neural networks on uneven micro-expression datasets. It was found that the accuracy of VGG16, VGG19, and MobileNetV3-Small before and after adjusting weights was 78% and 76%, 76% and 76%, and 75% and 72%, respectively. This indicated that VGG16 and MobileNetV3-Small have strong stability with uneven micro-expression datasets, while VGG19 has relatively weaker but still good stability. The research then used data augmentation to increase the scale of the datasets. The results showed that the accuracy of VGG16, VGG19, and MobileNetV3-Small increased to 78%, 80%, and 75%, respectively. This demonstrated that VGG19 performs better than VGG16 and MobileNetV3-Small, and MobileNetV3-Small performs the worst when the micro-expression datasets are large enough.

For future research, it is suggested that studies be extended to datasets with more complex micro-expression images, such as images with multiple human faces, images with incomplete or occluded faces, and images with complex backgrounds. Additionally, other convolutional neural networks like ResNet-50 and DenseNet-121 could be tested for micro-expression recognition to compare their performance. Finally, other evaluation metrics such as computational resources and processing time could be used to assess the performance of convolutional neural networks.

References

[1] Li Yante, Wei Jinsheng, Liu Yang, et al. Deep learning for micro-expression recognition: A survey. IEEE Transactions on Affective Computing, 2022, 13(4): 2028-2046.

[2] Yildirim Suleyman, Meshack Sandra Chimeumanu, Zeeshan A. Rana. The influence of micro-expressions on deception detection. Multimedia Tools and Applications, 2023, 82(19): 29115-29133.

[3] Perveen Gulnaz, Ali Syed Farooq, Ahmad Jameel, et al. Multi-Stream Deep Convolution Neural Network with Ensemble Learning for Facial Micro-Expression Recognition. IEEE Access, 2023, 11: 118474-118489.

[4] Varghese Ashwini Ann, Jacob P. Cherian, Jubilant J. Kizhakkethottam. Overview on emotion recognition system, international conference on soft-computing and networks security (ICSNS). IEEE, 2015: 1-5.

[5] Li Shan, Deng Weihong. Deep facial expression recognition: A survey. IEEE transactions on affective computing, 2020, 13(3): 1195-1215.

[6] Hara Kensho, Kataoka Hirokatsu, Satoh Yutaka. Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet? Proceedings of the IEEE conference on Computer Vision and Pattern Recognition. 2018: 6546-6555.

[7] Medsker Larry R., Lakhmi Jain. Recurrent neural networks. Design and Applications, 2001, 5(64-67): 2.

[8] Xie Hong-Xia, Lo Ling, Shuai Hong-Han, et al. An overview of facial micro-expression analysis: Data, methodology and challenge. IEEE Transactions on Affective Computing, 2022, 14(3): 1857-1875.

[9] Fuyuan Luo. Micro-expression Recognition Based on Dataset Balance and Local Connected Bi-branch Network, Southwest University Master's Thesis (Chinese), 2023.

[10] Lian Zheng, Li Ya, Tao Jian-Hua, et al. Expression analysis based on face regions in real-world conditions. International Journal of Automation and Computing, 2020, 17: 96-107.

[11] Le Wang, Meng Han, Xiaojuan Li, et al. Review of Classification Methods for Unbalanced Data Sets, Journal of Computer Engineering & Applications (Chinese), 2021, 57(22): 42-52.

[12] Shorten Connor, Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. Journal of big data, 2019, 6(1): 1-48.

[13] Qian Siying, Chenran Ning, Yuepeng Hu. MobileNetV3 for image classification. IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE). IEEE, 2021: 490-497.

[14] Mascarenhas Sheldon, Mukul Agarwal. A comparison between VGG16, VGG19 and ResNet50 architecture frameworks for Image Classification. International conference on disruptive technologies for multi-disciplinary research and applications. IEEE, 2021, 1: 96-99.