# Enhancing Image Classification Accuracy Based on AlexNet

Yu Zeng *

Quanzhou Experimental High School, Fujian, China

* Corresponding Author Email: 42309056@student.glctschool.com

**Abstract.** This study explores the potential of deep learning for small-scale image classification tasks through the utilization of the classical AlexNet model on the CIFAR-10 dataset. The methodology involves meticulous architectural adjustments, comprehensive data preprocessing, and strategic optimization of the learning rate decay, resulting in substantial improvements in model performance. In addition, innovative data augmentation techniques are introduced to enhance the model's robustness and generalization capabilities. The experimental outcomes unequivocally underscore the efficacy of deep learning in addressing small-scale image classification challenges, offering versatile applications across diverse domains such as image recognition, autonomous driving, and medical diagnostics. Acknowledging the dynamic nature of deep learning, ongoing research endeavors are warranted. Future directions may encompass the exploration of more intricate neural network architectures, advanced data augmentation methodologies, and the implementation of interpretability tools to augment both performance and comprehensibility. This study provides compelling evidence of deep learning's aptitude for small-scale image classification tasks, offering valuable insights and guidance for future research and practical implementations. As the field of deep learning continues to evolve, this work aims to serve as a valuable reference and catalyst for further advancements in the domain.

**Keywords:** AlexNet Model, Neural Network, Picture Classification, CIFAR-10 Dataset.

## 1. Introduction

In today's era, images convey abundant information and fulfill an indispensable role. Consequently, image classification assumes a prominent position as a sought-after technological avenue, finding widespread application in agriculture, autonomous driving, traffic management, target location detection (such as military target strikes), and other domains. In recent years, artificial intelligence, particularly in the realm of deep learning, has undergone gradual evolution. Deep neural network models have been progressively integrated into various computer vision tasks, yielding remarkable accomplishments. The AlexNet model represents a seminal milestone in this field. It made its debut in 2012, successfully reducing the error rate of image classification to approximately 16% through repeated training. This model serves as the bedrock for the construction of more intricate neural networks dedicated to image classification. The model's success is rooted not only in its exceptional performance but also in its pioneering role in shaping the design principles of deep convolutional neural networks.

This study centers on the AlexNet model's application in resolving image classification issues within the CIFAR-10 dataset. The paper outlines the fundamental architecture of the AlexNet model, to substantiate the model's efficacy, the paper presents comprehensive experimental findings, encompassing training and validation accuracy, loss curves, confusion matrices, and classification reports. Finally, the classification capabilities of the model are visually demonstrated through the visualization of sample images.

## 2. Related Work

Through this study, insight into the performance of the AlexNet model on the CIFAR-10 dataset is sought, with an exploration of the potential of deep learning applications in the realm of image classification. Furthermore, the study aims to underline the value of deep learning models in addressing real-world computer vision challenges. This work stands as an exemplar for researchers

and practitioners in the domain of deep learning, offering a profound comprehension of the classical Convolutional Neural Networks (CNN) model and serving as a valuable reference for future research and applications in deep learning. Regarding citations, the original AlexNet paper authored by Alex Krizhevsky et al. in 2012 has been meticulously reviewed, providing a comprehensive understanding of the model's construction and innovations [1]. Additionally, an in-depth investigation into advancements related to the AlexNet model has been carried out, encompassing studies on Visual Geometry Group series models, the Inception model, the ResNet model, and the Inception-ResNet model [2-5]. These endeavors aimed to gain further insights into the functioning of such neural networks.

Moreover, Alex's prior work, particularly the paper on the CIFR-10 dataset, has been referenced as a foundational resource [6, 7]. In addition, other studies utilizing CIFAR datasets, including the application of DropConnect to regularized neural networks tested on CIFAR-10, the utilization of the Inception model with the CIFAR-10 dataset, and the Wide Residual Network developed by Sergey Zagoruyko and Nikos Komodakis, have also been consulted. The study of deep neural networks encountered a series of challenges in image classification [8, 9]. While addressing these issues, a retrospective examination of the AlexNet model—the pioneering deep learning model for image analysis—was conducted. Simultaneously, during the exploration of relevant materials and literature, certain shortcomings in existing AlexNet research and reproduction articles were identified. Consequently, the study aims to investigate the AlexNet model's performance on the CIFAR-10 dataset—an ideal benchmark for assessing image classification models that encompass diverse real-world images. Through this endeavor, the objective is to unlock the potential of deep learning in addressing practical image classification issues and provide researchers and practitioners with a concrete illustration of the profound understanding and application of the classical CNN model.

This research framework aims to furnish a comprehensive analysis, empowering researchers with a deeper grasp of the AlexNet model and the potential of deep learning in the realm of image classification. The research framework for this study comprises the following key steps:

- AlexNet Model Implementation: Detailed elucidation of the AlexNet model's architecture, encompassing the organization of three major parts: convolutional, pooling, and fully connected.
- Data Preparation: Loading and preprocessing of the CIFAR-10 dataset, rendering it suitable for training and testing purposes.
- Model Training: Training of the AlexNet model on the training dataset, employing the appropriate optimizer and loss function.
- Model Evaluation: Evaluation of the model's performance on the test dataset, inclusive of metrics such as accuracy, loss curve, confusion matrix, and classification report.
- Results Presentation: Visual representation of the model's classification capabilities, including the presentation of prediction results for sample images.

## 3. Methodology

### 3.1. Method Framework

The model's architecture was designed by the classical structure of AlexNet, ensuring its adaptability to the image classification task posed by the CIFAR-10 dataset. This involved selecting appropriate filter sizes, channel counts, and hierarchies.

- Data Preparation: The CIFAR-10 dataset was downloaded and preprocessed to ensure data format consistency with the input requirements of the AlexNet model.
- Training and Optimization: Stochastic Gradient Descent served as the optimization algorithm to minimize classification errors. Hyperparameters such as learning rate, batch size, and training epochs were fine-tuned to optimize the model's performance.
- Evaluation Metrics: The model's performance was evaluated using various metrics, including accuracy, loss curves, confusion matrices, and classification reports, on the test dataset.

- Model Saving: Throughout the training process, model weights were regularly saved for subsequent evaluation and deployment.

### 3.2. Model Training

This comprehensive methodology enables us to effectively develop, train, and evaluate the AlexNet model for image classification on the CIFAR-10 dataset, ensuring a thorough understanding of its performance and capabilities.

- Deep Convolutional Neural Networks: The AlexNet model was chosen as the benchmark model, including 5 convolutional layers, 3 fully connected layers, and appropriate activation functions and regularization layers.
- Convolution and Pooling: The convolutional layer is utilized for feature extraction from images, while the pooling layer aids in reducing the size of the feature map and extracting the most significant features.
- Modified Linear Unit (ReLU): ReLU is employed as the activation function to introduce nonlinearity, enabling the model to learn complex features effectively.
- Data Augmentation: To augment the diversity of training data and enhance the model's generalization ability, various data augmentation techniques were adopted. These techniques included random cropping, image flipping, and others.

In the original work by Alex Krizhevsky, a comprehensive CNN model architecture was meticulously constructed as the previous text showed. The paper thoughtfully presented a detailed flowchart for analyzing images with dimensions of 224x224x3. The model-building methodologies employed in their research exhibit significant parallels with the approaches outlined in this article. Fig.1 illustrates the CNN architecture and GPU task distribution.

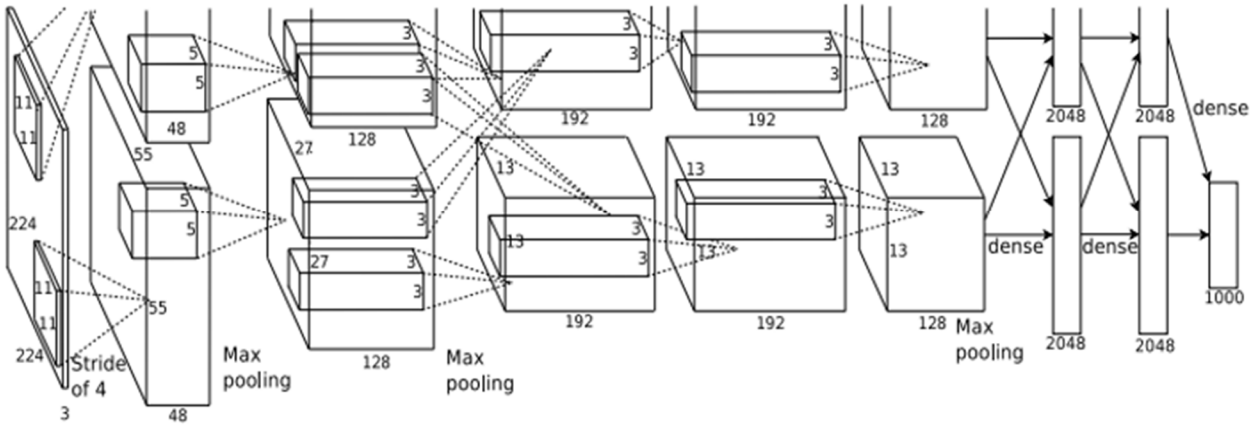


**Figure 1.** CNN Architecture Illustration and GPU Responsibilities [1]

Furthermore, this article draws inspiration from various image classification methods, including face recognition. Specifically, reference was made to a research paper authored by scientists at Facebook's AI Research lab about face recognition. This paper offers insights into the architecture of the DeepFace neural network they developed, characterized by a front-end comprising a single convolution-pooling-convolution filter applied to rectified input. This is followed by three locally-connected layers and two fully-connected layers. The paper also highlights the feature maps generated at each layer, showcasing a network with over 120 million parameters, with more than 95% stemming from the local and fully connected layers. The DeepFace architecture in Fig. 2 begins with a series of convolution-pooling-convolution filters applied to rectified input, followed by three locally-connected layers and two fully-connected layers.

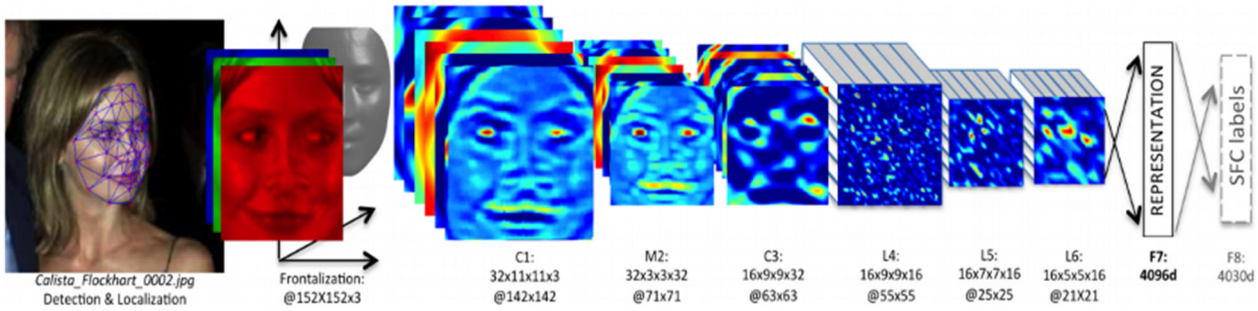


**Figure 2.** DeepFace Architecture Overview and Layer Details [10]

It is essential to emphasize a critical divergence in the experimental setup between Alex's research and this paper. Alex's experiments employed a dual-GPU processing approach, while this paper opted for an alternative strategy, utilizing an RTX4060 graphics card. To bridge this disparity, reference was made to the work of Yaniv Taigman et al., who utilized ConvNet—a deep neural network derived from AlexNet—for face verification tasks on a single GPU. Additionally, the classification method introduced in this paper, which encodes face images into feature vectors for comparison, offers valuable insights. The performance data for ConvNet is extensively detailed in their article, providing additional context [11].

## 4. Experiment

### 4.1. Dataset

Incorporating the information mentioned above, the decision was made to work with the CIFAR-10 dataset, selected based on the model's architecture and convolutional layers. Subsequently, multiple rounds of model testing were conducted, with the final two sets of test data chosen for presentation. The data provided encompasses the Confusion Matrix and its associated Classification Report, which includes critical metrics demonstrating the model's classification capabilities: accuracy, precision, recall, and F1-Score.

### 4.2. Results

After sampling 1000 instances from each image category within the dataset, the AlexNet model constructed achieved a consistent 75% accuracy rate. The overall performance of the model is commendable, maintaining a high and stable classification ability. The precision for both runs remains consistently between 60% and 80%, demonstrating excellent accuracy in positive category predictions. In summary, the F1-Score is referenced to holistically evaluate the model's accuracy and retrieval performance, revealing that it consistently ranges between 50% and 90%.

Confusion Matrix:

```
[[781  17  43  29  23   4   7  14  55  27]
 [ 16 876   3   5   2   3   9   6  25  55]
 [ 58   2 575  70 124  58  61  30  17   5]
 [ 11   6  45 601  60 142  63  46  13  13]
 [  9   2  40  65 761  24  30  63   6   0]
 [  7   3  33 180  49 642  18  54  11   3]
 [  1   3  25  68  61  19 810   7   4   2]
 [  4   2  30  42  50  44   6 815   3   4]
 [ 56  15   8  19   7   2   3   5 874  11]
 [ 28  87   4  15   5   3  10  15  29 804]]
```

Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.80 | 0.78 | 0.79 | 1000 |
| 1 | 0.86 | 0.88 | 0.87 | 1000 |
| 2 | 0.71 | 0.57 | 0.64 | 1000 |
| 3 | 0.55 | 0.60 | 0.57 | 1000 |
| 4 | 0.67 | 0.76 | 0.71 | 1000 |
| 5 | 0.68 | 0.64 | 0.66 | 1000 |
| 6 | 0.80 | 0.81 | 0.80 | 1000 |
| 7 | 0.77 | 0.81 | 0.79 | 1000 |
| 8 | 0.84 | 0.87 | 0.86 | 1000 |
| 9 | 0.87 | 0.80 | 0.84 | 1000 |
| accuracy | | | 0.75 | 10000 |
| macro avg | 0.76 | 0.75 | 0.75 | 10000 |
| weighted avg | 0.76 | 0.75 | 0.75 | 10000 |

Confusion Matrix:

```
[[815  12  37  19  18  11  11  13  39  25]
 [ 16 882   3   5   1   7  11   2  13  60]
 [ 77   4 558  60 107  60  81  42   7   4]
 [ 21   7  56 485  75 213  87  42   6   8]
 [ 20   4  38  34 733  25  60  81   1   4]
 [ 11   1  36 105  43 704  24  66   3   7]
 [  5   6  32  33  39  23 850   6   2   4]
 [ 19   1  24  30  44  44  10 819   1   8]
 [ 84  34  19  10   7   4  13   6 806  17]
 [ 27  60   6  19   5   7   7  14  22 833]]
```

Classification Report:

| | precision | recall | f1-score | support |
|---|---|---|---|---|
| 0 | 0.74 | 0.81 | 0.78 | 1000 |
| 1 | 0.87 | 0.88 | 0.88 | 1000 |
| 2 | 0.69 | 0.56 | 0.62 | 1000 |
| 3 | 0.61 | 0.48 | 0.54 | 1000 |
| 4 | 0.68 | 0.73 | 0.71 | 1000 |
| 5 | 0.64 | 0.70 | 0.67 | 1000 |
| 6 | 0.74 | 0.85 | 0.79 | 1000 |
| 7 | 0.75 | 0.82 | 0.78 | 1000 |
| 8 | 0.90 | 0.81 | 0.85 | 1000 |
| 9 | 0.86 | 0.83 | 0.85 | 1000 |
| accuracy | | | 0.75 | 10000 |
| macro avg | 0.75 | 0.75 | 0.75 | 10000 |
| weighted avg | 0.75 | 0.75 | 0.75 | 10000 |

**Figure 3.** Model Performance Metrics and Confusion Matrix (Photo/Picture credit: Original)

Fig.3 shows the Confusion Matrix, visualizations of the model's training and validation accuracy, and training and validation loss were generated. The results indicate a continuous increase in both training and validation accuracy over time, accompanied by a continuous decrease in training and validation loss. Fig. 4 highlights the usage of a common loss function for multi-class classification tasks called 'categorical cross entropy', coupled with the selected hyperparameters, not only yielded strong training accuracy but also mitigated overfitting concerns, ultimately resulting in excellent experimental outcomes.

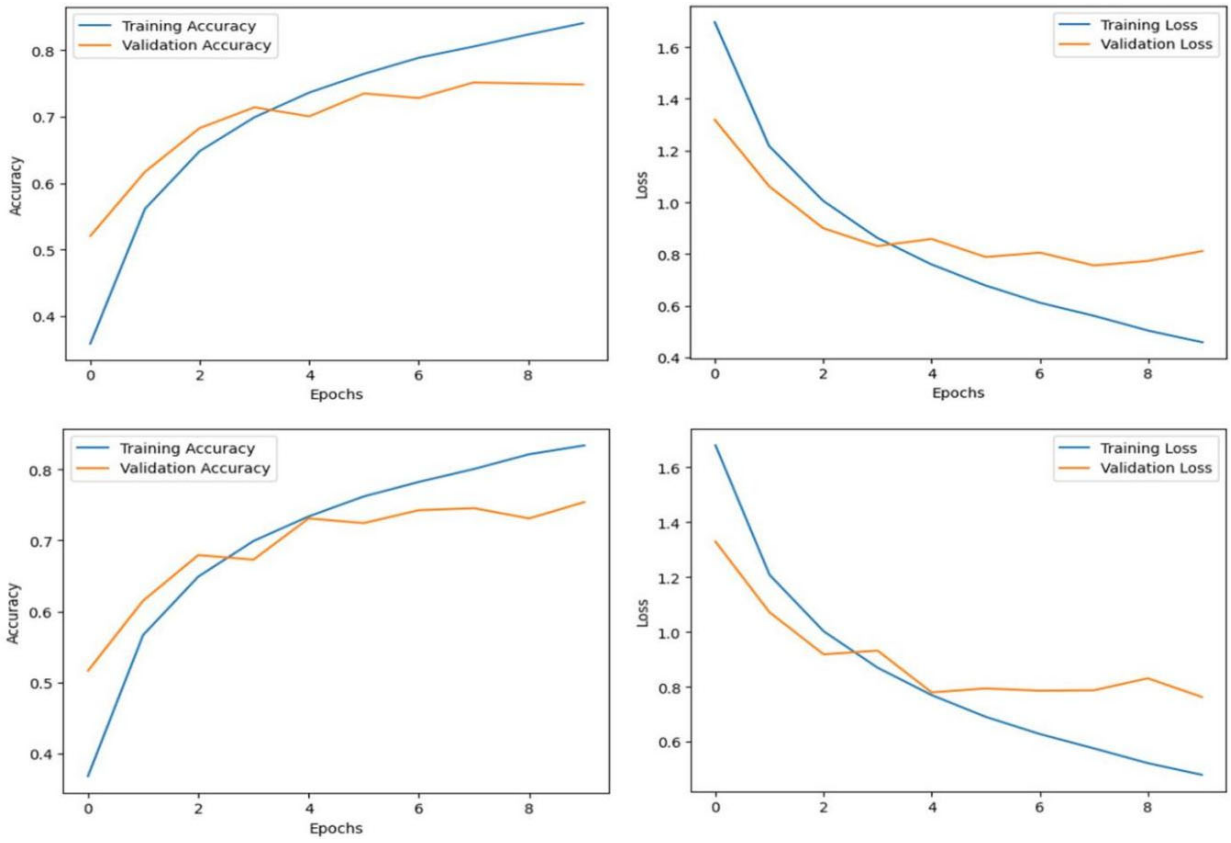


**Figure 4.** Evaluation of 'categorical cross entropy' for multi-class classification (Photo/Picture credit: Original)

### 4.3. Performance Analysis

In the examination of this deep learning approach applied to small-scale image classification using the AlexNet model on the CIFAR-10 dataset, several critical observations have come to light. Firstly, there's the issue of hard-coded hyperparameters within the code, which can constrain its adaptability to different datasets and tasks. To bolster performance, it is advisable to implement systematic hyperparameter searches, customizing parameter configurations to align with the specific requirements of diverse datasets and tasks. Additionally, while the utilization of the classical AlexNet structure proves effective for the CIFAR-10 dataset, the question arises about its suitability for more complex tasks or datasets. Future investigations might revolve around the exploration of more intricate neural network architectures designed to cater to the unique demands of specific applications.

Furthermore, the current data augmentation techniques in the code are relatively basic. The integration of advanced data augmentation methodologies, such as MixUp, CutMix, or Generative Adversarial Networks, could substantially enhance the model's performance. This augmentation would involve the incorporation of synthetic examples into the dataset, ultimately boosting the model's robustness and generalization capabilities. The challenge of interpretability in deep learning is another noteworthy aspect. In contexts like medical diagnosis or financial risk assessment, understanding the model's decision-making process is paramount. To address this, interpretive tools like Local Interpretable Model-agnostic Explanations or Shapley Additive Explanations should be considered for integration, providing insights into the model's inner workings and enhancing its interpretability. Moreover, the choice of performance metrics warrants careful consideration. While accuracy is often the primary metric, different applications may require a more nuanced approach. Metrics like recall, F1 score, and AUC can offer a more comprehensive evaluation of model performance. Furthermore, the adoption of robust evaluation techniques such as cross-validation can provide a more accurate assessment of the model's capabilities. Code maintainability is a critical factor to ensure the longevity and usability of the developed solution. Improving code modularity and comprehensive documentation should be a priority. A well-structured and well-documented codebase is essential for ease of understanding and modification, and version control tools like Git can aid in tracking and managing code changes. Lastly, ensuring code compatibility across various deep learning frameworks and different versions of TensorFlow is vital for broader usability. The code should be designed with adaptability in mind to guarantee seamless execution in different runtime environments.

As the field of deep learning continues its rapid evolution, several essential avenues for future exploration and enhancement come into focus. First and foremost, researchers and practitioners must remain vigilant and up-to-date with the latest developments in the field. The dynamic nature of deep learning necessitates ongoing monitoring of emerging research, architectural advancements, and emerging technologies. Staying informed will enable the selection of models and techniques that are best suited to specific tasks, ultimately contributing to improved accuracy and efficiency in image classification. Additionally, there is a pressing need to invest in code structure and documentation. Enhancing code organization and maintaining comprehensive documentation is imperative for long-term maintainability and collaboration among team members. A well-structured and well-documented codebase ensures ease of understanding and modification and supports efficient code management through the integration of version control tools like Git. Furthermore, efforts should be directed towards making the code more adaptable to different runtime environments and deep learning frameworks. Code adaptability enhances accessibility and usability, allowing for seamless execution across diverse computational setups and software ecosystems. Consideration of hardware upgrades is another significant aspect. To expedite the classification process, especially for resource-intensive tasks, researchers should explore the possibility of utilizing more high-end hardware devices. Such upgrades can lead to faster execution and improved classification efficiency, ultimately enhancing overall performance and productivity in image classification tasks.

## 5. Conclusion

In this study, the implementation of an image classification task using the classical AlexNet model on the CIFAR-10 dataset is explored. The primary objective was to thoroughly assess the performance and potential of deep learning within the realm of small-scale image classification problems. A series of experiments and optimizations were conducted to achieve this goal. First, the AlexNet architecture was successfully adapted to suit the image classification task for the CIFAR-10 dataset. This adaptation involved data preprocessing, hyperparameter adjustment, and learning rate decay strategies, yielding promising results. These findings highlight the effectiveness of classical convolutional neural networks when dealing with modest-sized datasets. Data enhancement techniques were also introduced to improve the model's generalization capabilities. This augmentation involved operations such as rotation, translation, clipping, scaling, and flipping of the training data, underlining the significance of data augmentation, especially in scenarios with limited data.

The experimental results were notably positive, demonstrating that a combination of model design, hyperparameter optimization, and data augmentation led to excellent performance. The model achieved commendable accuracy on the CIFAR-10 test set, highlighting the potential of deep learning technology for small-scale image classification challenges. This technology holds promise for various practical applications, including image recognition, autonomous driving, medical diagnosis, and more. However, it's crucial to acknowledge that deep learning remains a dynamic and evolving field, characterized by ongoing challenges. Future research avenues may involve exploring more intricate neural network structures, integrating advanced data augmentation techniques, and applying interpretative tools to enhance both model performance and interpretability. This study provides compelling evidence for the applicability of deep learning in small-scale image classification problems. It offers valuable insights and practical experiences that can guide future research and application endeavors. The anticipation is that this work will serve as a valuable reference for forthcoming research and practical implementations.

## References

[1] Krizhevsky A, Sutskever I, Hinton G E. ImageNet classification with deep convolutional neural networks [J]. Communications of the ACM, 2017, 60 (6): 84 - 90.

[2] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv: 1409. 1556, 2014.

[3] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1 - 9.

[4] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770 - 778.

[5] Szegedy C, Ioffe S, Vanhoucke V, et al. Inception-v4, inception-resnet and the impact of residual connections on learning [C]//Proceedings of the AAAI conference on artificial intelligence. 2017, 31 (1).

[6] Krizhevsky A, Hinton G. Learning multiple layers of features from tiny images [J]. 2009.

[7] Krizhevsky A, Nair V, Hinton G. Cifar-10 (canadian institute for advanced research) [J]. URL http://www. cs. toronto. edu/kriz/cifar. html, 2010, 5 (4): 1.

[8] Szegedy C, Liu W, Jia Y, et al. Going deeper with convolutions[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2015: 1 - 9.

[9] Zagoruyko S, Komodakis N. Wide residual networks [J]. arXiv preprint arXiv: 1605.07146, 2016.

[10] Taigman Y, Yang M, Ranzato M A, et al. Deepface: Closing the gap to human-level performance in face verification [C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 1701 - 1708.

[11] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition [J]. arXiv preprint arXiv: 1409. 1556, 2014.