WEP
Warwick
Evans
Publishing

Transactions on Computer Science and Intelligent Systems Research
ISSN: 2960-1800, eISSN: 2960-2238 | Volume 5
AIDML 2024

A Comparative Analysis Study of Deep Learning Methods in Cat And Dog Classification

Meng Lin *

School of Computer Science, Beijing Institute of Technology, Beijing, 100000, China

* Corresponding Author Email: 1120212825@bit.edu.cn

Abstract. In recent years, deep learning has significantly advanced image classification. However, as models grow more complex, their computational demands increase, prompting a shift toward lightweight models. The MobileNet series, known for its efficiency in resource-constrained environments, is a prime example. Despite its popularity, performance comparisons among different MobileNet versions for specific tasks like cat and dog image classification remain underexplored. This study addresses this gap by evaluating MobileNetV1, MobileNetV2, MobileNetV3Large, and MobileNetV3Small on a Kaggle dataset containing over 1000 images. The dataset underwent preprocessing before training and testing. The paper assessed the models' classification accuracy and convergence speed through comparative analysis. Results indicate that MobileNetV2 outperforms the others, with superior accuracy and faster convergence, making it the preferred choice. MobileNetV1 also showed stable performance, while MobileNetV3Large's larger size led to overfitting issues. In conclusion, MobileNetV2's exceptional performance in cat and dog image classification suggests its broad applicability in resource-limited scenarios. This study provides valuable insights for deploying image classification models on mobile devices and other constrained environments. Future research should focus on further optimizing these models to enhance their performance and generalization capabilities.

Keywords: Deep learning; Image classification; MobileNet; Cat and dog classification.

1. Introduction

In recent years, with the rapid development of deep learning technology, significant progress continues to be made in the field of image classification. Image classification is not only a fundamental problem in computer vision but also the foundation for complex tasks such as behavior detection, object recognition, and image segmentation. It is widely used in various fields, including medical diagnosis, security surveillance, and autonomous driving.

In the early days of image classification, Yann deployed CNNs for handwritten digit recognition, achieving excellent results [1]. Following 2012, representative deep network structures such as GoogLeNet and ResNet emerged successively [2, 3]. To achieve outstanding performance across various tasks, deep learning models continuously increase in parameter size. This leads to higher computational demands and prompts a shift towards lightweight research in deep learning. In recent years, researchers have proposed lightweight network structures such as MobileNetV1, MobileNetV2, and GhostNet, with the MobileNet series being particularly representative [4-7]. Liu selected MobileNetV2 as the backbone for their research, optimizing it using model distillation algorithms, achieving an accuracy of 97.9% on a 12-category garbage dataset with a model parameter size of 14M [8]. The MobileNet series, designed for resource-constrained environments like mobile devices, maintains accuracy while being smaller and faster, leading to the development of lightweight deep learning models [9].

In practical applications, image classification technology is widely used in many fields, including facial recognition, product identification, and image search. Within this domain, cat and dog image classification is a prominent issue due to its practical importance and its widespread use as a benchmark task for evaluating deep learning model performance.

As a representative of lightweight deep learning models, the performance of the MobileNet series in image classification tasks has received considerable attention. However, performance comparison studies among different versions of MobileNet for the specific task of cat and dog image classification are still lacking. This study aims to fill this gap by thoroughly investigating the performance differences of various MobileNet versions in cat and dog image classification, providing guidance and reference for image classification applications in resource-constrained environments like mobile devices.

This paper adopts a comparative analysis approach to evaluate the performance of MobileNetV1, MobileNetV2, MobileNetV3Large, and MobileNetV3Small in the task of cat and dog image classification. The specific content includes a comparative analysis of multiple indicators such as model convergence speed, classification accuracy, and F1 score. By comparing the performance of different versions of MobileNet in the cat and dog image classification task, this study explores their strengths and weaknesses, offering guidance and reference for image classification applications in resource-constrained environments.

2. Method

2.1. Dataset

The dataset used in this study is sourced from the cat and dog image classification dataset available on the Kaggle platform. This dataset contains over 1,000 images of cats and dogs. Additionally, duplicate images were removed to ensure the quality and diversity of the dataset, aiding the model's ability to generalize better. The image sizes range from approximately 100x100 pixels to 2000x1000 pixels, and the format is JPEG. Each image in the dataset is labeled as either a cat or a dog, serving as the labels for the classification task in this study. The dataset is downloaded over 15,000 times and receives 112 likes on Kaggle, indicating a certain level of recognition and reliability.

2.2. Data Preprocessing

A series of data preprocessing steps were employed in this study to provide high-quality input data for the model.

First, the raw images were loaded into memory and resized according to predefined dimensions to meet the input requirements of the model. This step helps maintain image uniformity, enabling effective processing by the model.

Next, the images were processed using the built-in preprocessing functions of the pre-trained deep learning models (MobileNet series models). This process includes normalizing and transforming the images to match the input data range of the utilized model. This approach leverages the pre-learned feature transformations to improve training efficiency and accelerate the convergence process.

The processed image data (in array form) and their corresponding labels were stored together for model training in the training dataset. The same processing is applied to the test dataset. To facilitate the final results presentation, a copy of the original images is also retained.

Finally, during the training process, the training set is split into training and validation sets according to a specific ratio to evaluate model performance and perform parameter tuning. This step helps prevent overfitting and improves the model's generalization capability.

2.3. Model

This study explores four different pre-trained models, each set as the foundation architecture for the deep learning model. These include MobileNetV1, MobileNetV2, MobileNetV3Large, and MobileNetV3Small, each possessing varying network depths and complexities.

To construct the classification model, one of the pre-trained models is chosen as the backbone network, to which a global average pooling layer and a series of fully connected layers were added. The fully connected layers utilized ReLU activation functions and batch normalization, which helps extract and learn high-level features from the image data while introducing a certain degree of non-linearity and generalization capability to the model.

A key highlight of this study is the use of dropout layers to prevent overfitting. By randomly disabling some neuron units during training, dropout helps improve the model's generalization ability and reduces the risk of overfitting the training data.

Additionally, the Adam optimizer and cross-entropy loss function were used to configure the model's training process. Notably, a custom learning rate scheduler was employed, adjusting the learning rate at different stages of training to enhance the model's convergence speed and performance.

Overall, the model construction in this study emphasizes the following highlights:

Transfer Learning: Utilizing pre-trained MobileNet series models as the base models allows for the learning of general image features, enhancing generalization ability and accelerating convergence speed [10]. This approach helps better initialize network parameters for similar tasks, speeding up the training process and improving model performance.

Adoption of Lightweight Structures: The MobileNet models are lightweight convolutional neural networks with fewer parameters and lower computational costs, making them suitable for deployment in resource-constrained environments.

Optimized Model Structure: The pre-trained models were augmented with global average pooling layers and multiple dense layers, effectively extracting and learning image features to achieve accurate image classification.

Learning Rate Adjustment Strategy: The model employed a learning rate scheduler that dynamically adjusted the learning rate based on the number of training iterations, ensuring smooth convergence during training and achieving better performance.

3. Experimental Results

3.1. MobileNetV1

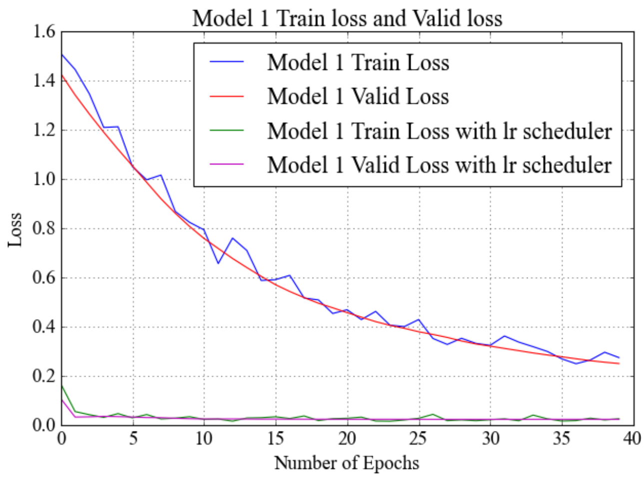

Fig. 1 Training and validation loss and accuracy over 40 epochs for the MobileNetV1 model (Photo/Picture credit: Original)

The horizontal and vertical coordinates in Fig. 1 represent the number of training epochs and the loss value, respectively. The four lines depict the changes in the loss values on the training set and validation set as the number of epochs increases. Overall, all four lines show a downward trend. After incorporating the learning rate scheduler, the loss values significantly decreased, accelerating the model's convergence.

3.2. MobileNetV2

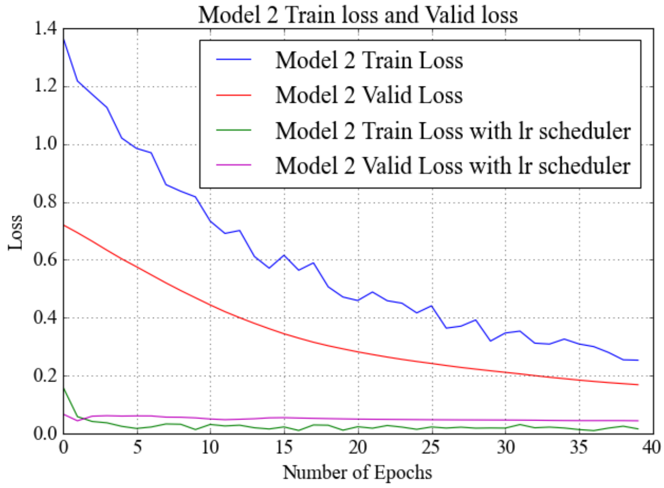

Fig. 2 Training and validation loss and accuracy over 40 epochs for the MobileNetV2 model (Photo/Picture credit: Original)

The horizontal and vertical coordinates in Fig. 2 represent the number of training epochs and the loss value, respectively. The four lines illustrate the changes in the loss values on the training set and validation set as the number of epochs increase. Similar to Fig. 1, all four lines show a downward trend. The incorporation of the learning rate scheduler led to a noticeable decrease in loss values, speeding up the model's convergence.

3.3. MobileNetV3_small

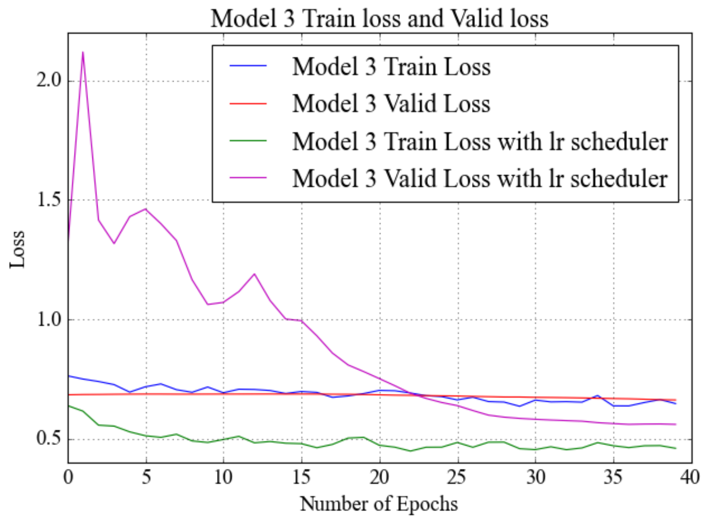

Fig. 3 Training and validation loss and accuracy over 40 epochs for the MobileNetV3_small model (Photo/Picture credit: Original)

The horizontal and vertical coordinates in Fig. 3 represent the number of training epochs and the loss value, respectively. The four lines show the changes in the loss values on the training set and validation set as the number of epochs increase. Unlike the trends observed in the previous in Fig. 1 and Fig. 2, the loss values of this model did not significantly decrease as the number of epochs increased. The addition of the learning rate scheduler resulted in only a slight acceleration of the convergence speed.

3.4. MobileNetV3_large

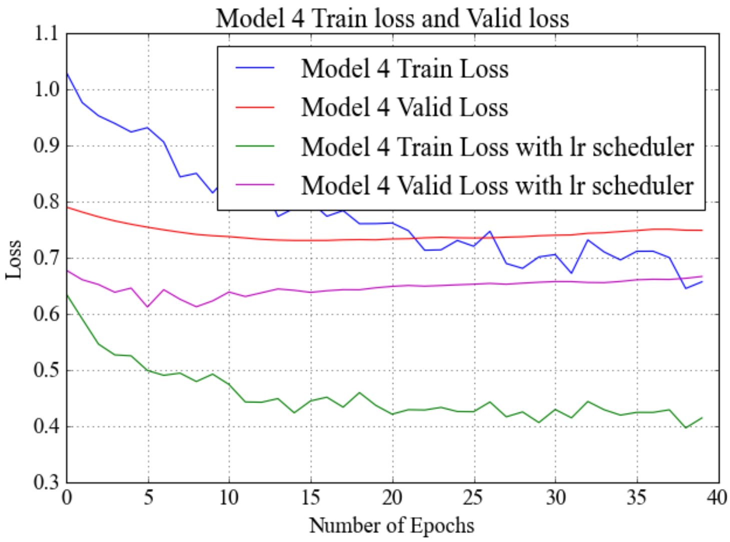

Fig. 4 Training and validation loss and accuracy over 40 epochs for the MobileNetV3_large model (Photo/Picture credit: Original)

The horizontal and vertical coordinates in Fig. 4 represent the number of training epochs and the loss value, respectively. The four lines depict the changes in the loss values on the training set and validation set as the number of epochs increase. The validation loss does not show a clear downward trend, while the training loss decreases significantly with the increase in epochs. After adding the learning rate scheduler, the loss values significantly decreased, speeding up the model's convergence.

Table 1. The F1-scores of the four models with and without learning rate schedulers on the cat and dog dataset.

	cats-f1-score	dogs-f1-score
MobileNetV1	0.88	0.9
MobileNetV1_lr	0.93	0.94
MobileNetV2	0.93	0.93
MobileNetV2_lr	0.95	0.95
MobileNetV3_Small	0.71	0.51
MobileNetV3_Small_lr	0.68	0.66
MobileNetV3_Large	0.57	0.57
MobileNetV3_Large_lr	0.71	0.63

Table 1 compares the F1 scores of the four models on the cat and dog dataset, both with and without the learning rate scheduler. The addition of the learning rate scheduler improved the F1 score by approximately 0.03, accelerating the model's convergence.

4. Discussion

This study compares and evaluates four deep learning models based on different pre-trained models to determine the best choice for the task of cat and dog image classification on a small dataset. Despite the relatively small scale of the dataset, containing only a few hundred images, performance differences among the models were observed, leading to several valuable conclusions.

Table 1 shows the F1 scores of the four models on the cat and dog dataset, both with and without the learning rate scheduler. It can be seen that the addition of the learning rate scheduler helps the models converge to the optimal state faster, resulting in higher F1 scores. From the table, it is evident that the MobileNetV2 model achieved the highest F1 score, indicating its superior performance on small datasets. The MobileNetV2 model features a lightweight structure and demonstrated faster convergence and higher classification accuracy during training compared to the other models. This is attributed to some design optimizations in MobileNetV2, such as the use of depthwise separable convolutions, which allow it to better adapt to the characteristics of our dataset [11].

The MobileNetV1 model, as an earlier deep learning model, performed slightly worse than MobileNetV2 but still achieved satisfactory results, outperforming the MobileNetV3Small model. The MobileNetV1 model has stable performance and good generalization ability, making it a viable alternative.

The MobileNetV3Large model showed somewhat inferior performance on the small dataset and also had a longer training time. Despite potentially learning-rich image features during the pre-training phase, its larger model size and more complex structure limited its adaptability to the small dataset, resulting in lower classification accuracy and slower training speed.

In summary, the MobileNetV2 model demonstrated the best performance on the small dataset, with its fast convergence speed and excellent classification accuracy making it the preferred choice. The MobileNetV1 model serves as a stable alternative, especially suitable for scenarios requiring strict performance standards. In contrast, the MobileNetV3Large model appears less suitable in this context, as its larger size may lead to overfitting issues. Future research can further optimize and adjust the parameter configurations and training strategies of different models to enhance their performance and generalization capabilities.

However, for larger datasets or more complex tasks, such as those involving an increased number of classification categories, MobileNetV2 may experience a decrease in accuracy due to higher image similarity [12]. In such cases, the MobileNetV3Large model might perform better. Future work will further explore and evaluate the performance of these models under different conditions to obtain more comprehensive comparative conclusions.

5. Conclusion

This study evaluated the performance of four pre-trained models—MobileNetV1, MobileNetV2, MobileNetV3Large, and MobileNetV3Small—on the task of cat and dog image classification through comparative analysis. The experimental results indicated that the MobileNetV2 model performed the best on the small dataset, exhibiting rapid convergence and high classification accuracy. Although the MobileNetV1 model was slightly inferior to MobileNetV2, it demonstrated stable performance and good generalization ability. In contrast, the MobileNetV3Large model tended to overfit on the small dataset, resulting in lower classification accuracy and longer training times. Therefore, the MobileNetV2 model shows great promise for applications in resource-constrained environments, particularly on mobile devices.

This study employed a lightweight model structure combined with transfer learning techniques, effectively enhancing the training efficiency and classification performance of the models. The use of a learning rate scheduler optimized the training process, further improving the convergence speed and stability of the models. The findings provide valuable guidance and reference for image classification applications in resource-constrained environments such as mobile devices.

Cat and dog image classification technology has widespread applications in various fields, including facial recognition, product identification, and image search. In the future, with the expansion of dataset sizes and advancements in model optimization techniques, lightweight deep-learning models are expected to demonstrate significant potential and extensive applicability in more complex tasks.

Continued exploration and improvement of these models will help advance the development of image classification technology.

References

[1] Lecun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[2] Szegedy C, Liu W, Yangqing J, et al. Going deeper with con-volutions. Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA, 2015: 1–9.

[3] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA, 2016: 770−778.

[4] Howard A G, Menglong Z, Chen B, et al. MobileNets: efficient convolutional neural networks for mobile vision applications. arXiv: 1704.04861, 2017.

[5] Sandler M, Howard A, Menglong Z, et al. MobileNetv2: inverted residuals and linear bottlenecks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA, 2018: 4510–4520.

[6] Han K, Wang Y, Tian Q, et al. GhostNet: More features from cheap operations. arXiv preprint arXiv: 1911.11907, 2019.

[7] Shuping G, Qingyuan Z, Xiaogang Q, et al. Research on an improved MobileNet image classification method. Journal of Intelligent Systems, 2021, 16(1): 11-20.

[8] Zhichao C, Haining J, Jie Y, et al. Garbage image classification algorithm based on improved MobileNet v2. Journal of Zhejiang University (Engineering Edition), 2021, 55(08): 1490-1499.

[9] Liu X, Wu Z Z, Zou L, et al. Lightweight neural network-based garbage image classification using deep mutual learning. 11th International Symposium on Parallel Architectures, Algorithms and Programming. Shenzhen: Springer, 2021: 212-223.

[10] Li Linhong, Yang Jie, Jiang Yanxuan, et al. Clothing image classification algorithm based on improved MobileNet v2. Modern Textile Technology, 2024, 32(04): 93-103.

[11] Wenjing G, Jie T, Baoqi L, et al. Underwater target classification method based on the fusion of acoustic and optical images using improved MobileNetV2 network. Applied Acoustics, 2022, 41(03): 462-470.

[12] Kang W. Research on garbage identification and classification based on deep learning. North University of China, 2024.