Comparative Analysis of VGG16 and EfficientNet for Image-Based Cat Breed Classification

Hao Xu
Polytechnic Institute, Purdue University, West Lafayette, Indiana, 47907, U.S.A.

Keywords: Cat Bread Classification, Convolutional Neural Networks, Image Classification.

Abstract: Nowadays, Convolutional Neural Network (CNN) architectures are widely used to distinguish animal species. For example, they are used to differentiate between various types of sheep, dogs, fish, and so on. This greatly assists people in identifying their species and assessing their value. After all, it is challenging for individuals to differentiate these animals' species without extensive relevant experience and expertise. Although Deoxyribonucleic Acid (DNA) testing can be used for identification, it is time-consuming and costly, making it impractical. Utilizing machine learning methods for differentiation saves a significant amount of time and effort. However, different CNN architectures have distinct focuses and functionalities. This study compares the differences between Visual Geometry Group (VGG)16 and EfficientNetB0 by classifying cat breeds. The primary method is to train models using these two CNNs and then compare their performance, focusing on their accuracy, computational efficiency, and generalization capabilities. This study reveals the strengths and weaknesses of these two models, enabling you to understand which neural network is more suitable for use.

1 INTRODUCTION

Today, international organizations have confirmed that there are more than 70 cat species in the world. Even a cat lover cannot recognize all cat species (Zhang, 2020). The task of identifying cat breeds is not only a trivial pursuit for cat lovers, it also holds significant importance in various fields such as veterinary science, animal conservation, and even in the pet industry (Ramadhan, 2023).

With the advent of deep learning, the accuracy and efficiency of breed identification have seen remarkable improvements (Shrestha, 2019). Deep learning, particularly convolutional neural networks (CNNs), has revolutionized the way complex image classification tasks are approached, offering a powerful tool for distinguishing between different species and breeds (Li, 2021).

This paper presents a comparative analysis of two prominent CNN architectures, including VGG16 and EfficientNetB0, in the context of cat breed classification. This work trains and evaluates these models on a dataset of cat images, focusing on their performance metrics such as accuracy, computational efficiency, and generalization ability. The goal is to not only determine which model performs better in classifying cat breeds but also to understand the underlying reasons for their performance differences.

2 METHOD

2.1 Dataset

In this study, the dataset comes from Kaggle, named "Cat Breeds Dataset" (Ma, 2019). This dataset consists of hundreds of thousands of high-quality images, covering sixty cat breeds. Because there are too many cat breeds in this dataset. To facilitate processing and uploading and save training time, this work selected the five most common cat breeds for research, including Calico, Persian, Siamese, Tortoiseshell, and Tuxedo. The final filtered dataset has about 14,000 images.

2.2 Data Preprocessing

To ensure a proper balance between training and validation, this work randomly split the dataset into training and validation sets in an 8:2 ratio, where 80% of the data is used for training and 20% is used for validation. The validation set helps evaluate the

Xu, H.
Comparative Analysis of VGG16 and EfficientNet for Image-Based Cat Breed Classification.
DOI: 10.5220/0013297100004558

In *Proceedings of the 1st International Conference on Modern Logistics and Supply Chain Management (MLSCM 2024)*, pages 238-242
ISBN: 978-989-758-738-2

performance of the model on unseen data and more accurately reflects the generalization ability of the model rather than just its performance on the training data. By using the validation set, this work can detect whether the model is overfitting the training data. If the model performs well on the training set but poorly on the validation set, it may indicate overfitting. By fixing the random seed, the split is guaranteed to be reproducible, which is very useful for debugging and comparing model performance.

2.3 Data Augmentation

To enhance the model's learning capabilities, this work implemented a series of data augmentation techniques. Initially, pixel values were normalized from [0, 255] to [0, 1] to streamline model processing and expedite training convergence. The data augmentation included random rotations up to 40 degrees and translations of up to 20% in width or height, simulating varied object positions and orientations. Shearing was applied to introduce complex deformations, while random scaling up to about 20% helped the model recognize objects of different sizes as the same category. Horizontal flipping of images was also performed to expand the training dataset. Additionally, any blank spaces created by these transformations were filled with the nearest pixel value to preserve image integrity. Consistent with the training data, the validation set was normalized to maintain uniform data formatting, which is essential for effective model training and evaluation (Xu, 2023).

2.4 Model Architecture

VGG16 is a very early CNN model. When deep learning became popular, it was the best neural network architecture at the time (Simonyan, 2014). So, it is a very good control group, reference group. It deepens the network by stacking 3*3 convolutional layers. Its reference volume is relatively large, about 138Millions, and its design is relatively simple. So, for some basic image classification tasks, its performance is still relatively good. But it may be a bit difficult to distinguish the types of cats, because it needs to capture more complex and subtle features, and vgg16 is not as good as the new architecture in this regard.

EfficientNetB0 is a new type of neural network architecture (Tan, 2019). Its design has been optimized and is different from VGG16. It balances depth, width and resolution through compound scaling technology, greatly improving accuracy. It has a low reference size, only about 5.1 million parameters. So, from a design point of view, it is lighter and more efficient than VGG16. It is stronger than VGG16 in solving complex image classification tasks. This study is the results and performance analysis of the training model to compare the performance of VGG16and EfficientNetB0 when used to classify cat breeds. Their performance is analyzed and compared by using prediction rate, readiness rate, F1 score and confusion matrix (Hossin, 2015).

3 EXPERIMENT AND RESULTS

3.1 Training Details

This study used the following training techniques to optimize the model. They are ModelCheckpoint, EarlyStopping, and ReduceLROnPlateau.

ModelCheckpoint is a callback function that helps users to save the best performing model on the validation set during training. EarlyStopping is a callback function used to stop training early (Yao, 2007). If the performance on the validation set (e.g., validation loss) does not improve within a certain number of epochs, training will end early to avoid overfitting. ReduceLROnPlateau is a callback function used to reduce the learning rate when the model performance does not improve. It automatically reduces the learning rate when the validation set performance does not improve within a certain number of epochs. This helps the model adjust weights more smoothly when it is close to the optimal solution.

The model needs to be trained in two phases. The last few layers were selectively unfrozen for fine tuning after freezing the convolutional layer at the first stage, by freezing convolutional layers one can be sure that it has learned general low-level features and simple objects. This prevents overfitting. During the second phase, the author unfrozes most of these last layers only for some fine-tuning tasks specify ones to learn new features from this data while preserving all other learned general features previously. Applying a low learning rate, fine-tuning weighs the end layers to smoothly adjust them through while training on those new tasks.

3.2 Result Comparison

To compare the performance of the two trained models, this work evaluated the models in various ways. In addition to accuracy, it also includes precision, recall, F1 score and ROC curve. These indicators can help people better evaluate the performance of the model and compare the models. This work set up a test set of 2495 images, which are classified into five cat breeds: 'Calico', 'Persian', 'Siamese', 'Tortoiseshell', 'Tuxedo'. Then use this dataset to test the performance of the two models. The performance is of VGG16 and EfficientNet is demonstrated in Table 1 and Table 2, respectively, with their confusion matrixes in Figure 1 and Figure 2.

Table 1: Performance comparison using VGG16 model.

	Precision	Recall	F1-score
Calico	0.753	0.842	0.795
Persian	0.871	0.934	0.901
Siamese	0.910	0.874	0.892
Tortoiseshell	0.889	0.768	0.823
Tuxedo	0.909	0.896	0.902
Accuracy	0.863	0.863	0.863
Macro avg	0.866	0.863	0.863
Weighted avg	0.866	0.863	0.863

Table 2: Performance comparison using EfficientNetB0.

	Precision	Recall	F1-score
Calico	0.858	0.898	0.878
Persian	0.911	0.988	0.948
Siamese	0.956	0.962	0.959
Tortoiseshell	0.945	0.790	0.860
Tuxedo	0.934	0.960	0.947
Accuracy	0.919	0.919	0.919
Macro avg	0.921	0.919	0.918
Weighted avg	0.921	0.919	0.918

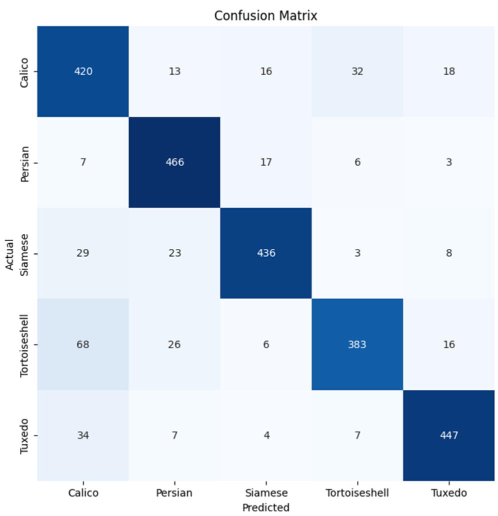

Figure 1: Confusion matrix result of VGG16 model (Figure Credits: Original).

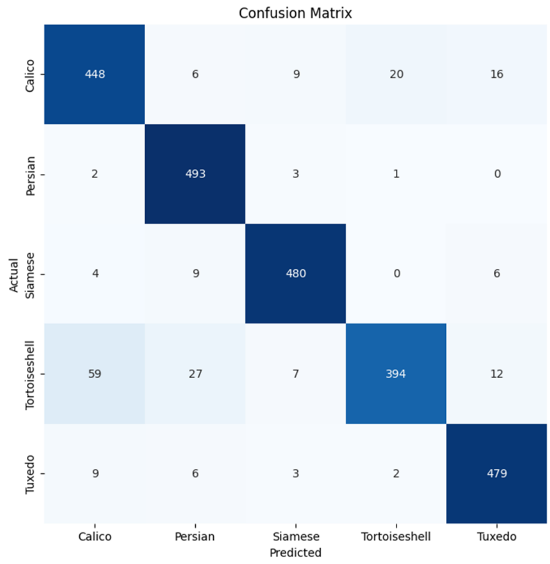

Figure 2: Confusion matrix result of EfficientNetB0 model (Figure Credits: Original).

For accuracy, the VGG16 model achieves 86.25% and EfficientNetB0 91.94%. It is obvious that EfficientNetB0 model surpasses VGG16 for general accuracy improvement around 5.69% VGG16 weighted average precision: 86.62%EfficientNetB0 weighted average precision: 92.08% Overall EfficientNetB0 has a better accuracy in almost all categories and fewer false positives than VGG16. Moreover, VGG16 has the weighted average precision of 86.25% and EfficientNetB0, it is increased to the weighted average recall of 91.94%. EfficientNetB0 shows a higher recall over VGG16 in all categories. VGG16 has a weighted average of 86.26% F1 score, and EfficientNetB0 got that number up to 91.83%. EfficientNetB0 has a better balanced between precision and recall compared to VGG16 with the higher F1 score. If analyzed by category, EfficientNetB0 has higher accuracy, precision, recall, and F1 score than VGG16 for every cat breed. From the confusion matrix, VGG16 had the most difficulty distinguishing between certain categories, such as "Tortoiseshell" and "Calico", as seen in the confusion between the two in the matrix. For example, 68 Tortoiseshell cats were incorrectly classified as Calico cats. EfficientNetB0 significantly reduced this confusion, although there were still some misclassifications (e.g., 59 Tortoiseshells were classified as Calico). EfficientNetB0 also showed better overall performance in identifying the "Siamese" and "Tuxedo" breeds, with fewer misclassifications than VGG16.

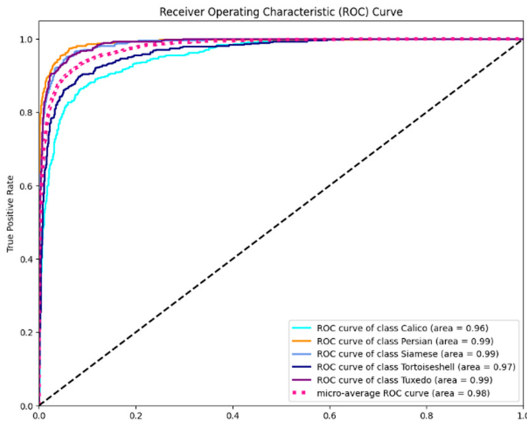

Figure 3: ROC curve of VGG16 model (Figure Credits: Original).

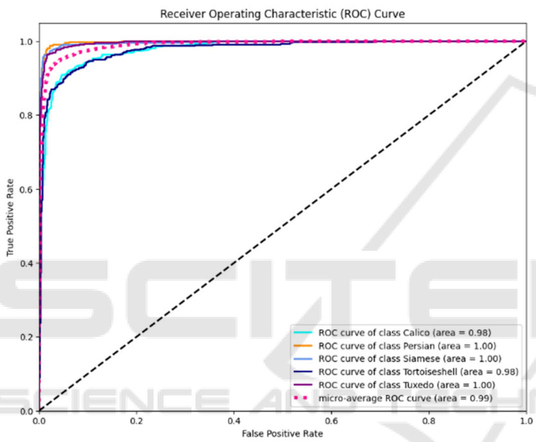

Figure 4: ROC curve of EfficientNetB0 model (Figure Credits: Original).

As demonstrated in Figure 3, the AUC values of the VGG16 model show that, except for Calico, all other categories have scores very close to 1.0 or they have high accuracy in this type, which tells that in most cases, VGG16 predicts animals as predicted, especially the Persian and Siamese cat types, which are even close to 100% accurate. In addition, as displayed in Figure 4, EfficientNetB0 has better AUC than VGG16 overall, with an AUC of 1.00 for Persian and Siamese cats. EfficientNetB0's Tuxedo is equal to 1 with VGG16's Calico and Tortoiseshell. Of course, intuitively, it could be observed that the curve of EfficientNetB0 is closer to the y-axis, and the curves of EfficientNetB0 are more concentrated with each other.

4 DISCUSSIONS

From the above experimental results, pictures and analysis, it could be observed that although VGG16 is a classic deep learning model and performs well in multiple classification tasks, its performance is obviously inferior to EfficientNetB0 in this experiment. This phenomenon may be related to the fact that VGG16 has too many parameters and has not been optimized. Its huge fully connected layer design is prone to overfitting when processing relatively small data sets, while increasing training time and computing resource consumption. In contrast, EfficientNetB0 is more sophisticated in design and uses compound scaling. While maintaining a low parameter volume, it can still achieve higher classification accuracy. Its performance on the three varieties of Persian, Siamese and Tuxedo is particularly outstanding, with an AUC of 1.00, almost eliminating misclassification. This shows that EfficientNetB0 not only has stronger discrimination ability when processing complex visual features, but also can better adapt to task requirements when the distribution differences between categories are large.

In terms of overall performance, EfficientNetB0 outperforms VGG16. But in fact, in these five categories, the AUC of EfficientNetB0 is only slightly higher than that of VGG16. This may be related to the selected dataset and the category of cats. The quality of the dataset in this study is not high, the images are not complex enough, and there are fewer fine features. And the five cats selected in this study look very similar in visual features. Therefore, the model sometimes misclassifies them.

Future work. This study can still be further improved. It could be found that some higher quality datasets to train the model. Works can also use some better data enhancement methods when processing data. These can further improve the performance of the model and its ability to accurately recognize. Secondly, more categories of cats could be introduced for research. Then more results and data will make the research more convincing. Third, several different neural network architectures could be applied in the research, obtain the results for analysis and comparison, to improve the level and quality of the research.

5 CONCLUSIONS

This study uses two neural network models, VGG16 and EfficientNetB0, to classify cat breeds.

EfficientNetB0 shows higher values in key indicators such as precision, recall, F1-score, and AUC values, especially in the classification of Persian, Siamese, and Tuxedo breeds, because its perfect classifier reaches 1.00. In addition, the micro-average AUC of EfficientNetB0 is 0.99, which is significantly higher than 0.98 of VGG16.

Compared with VGG16, EfficientNetB0 is more efficient and has significantly fewer parameters, which not only leads to better performance, but also lower computational complexity and training time compared with previous architectures. Based on the comparative analysis of ROC curves, EfficientNetB0 shows stronger classification capabilities, especially in the case of reducing misclassification. The result is that compared with VGG16, EfficientNetB0 is a more advantageous model in the task of cat breed classification, and its excellent classification performance and efficient computational performance make it have broad application prospects in similar image classification tasks.

Overall, this study reveals the significant advantages of EfficientNetB0 in performance and computational efficiency through data comparison between EfficientNetB0 and VGG16, and draws some reasonable analysis, results and conclusions. This also provides a good reference for those who attempt image classification tasks in the future.

REFERENCES

Hossin, M., & Sulaiman, M. N. 2015. A review on evaluation metrics for data classification evaluations. *International journal of data mining & knowledge management process*, *5*(2), 1.

Li, Z., Liu, F., Yang, W., Peng, S., & Zhou, J. 2021. A survey of convolutional neural networks: analysis, applications, and prospects. *IEEE transactions on neural networks and learning systems*, *33*(12), 6999-7019.

Ma. 2019. Cat breeds dataset. Kaggle. URL: https://www.kaggle.com/datasets/ma7555/cat-breeds-dataset. Last Accessed: 2024/09/14

Ramadhan, A. T., & Setiawan, A. 2023. Catbreedsnet: An Android Application for Cat Breed Classification Using Convolutional Neural Networks. *Jurnal Online Informatika*, *8*(1), 52-60.

Shrestha, A., & Mahmood, A. 2019. Review of deep learning algorithms and architectures. *IEEE access*, *7*, 53040-53065.

Simonyan, K., & Zisserman, A. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Tan, M. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*.

Xu, M., Yoon, S., Fuentes, A., & Park, D. S. 2023. A comprehensive survey of image augmentation techniques for deep learning. *Pattern Recognition*, *137*, 109347.

Yao, Y., Rosasco, L., & Caponnetto, A. 2007. On early stopping in gradient descent learning. *Constructive Approximation*, *26*(2), 289-315.

Zhang, Y., Gao, J., & Zhou, H. 2020. Breeds classification with deep convolutional neural network. In *Proceedings of the 2020 12th international conference on machine learning and computing*. 145-151.