Enhancing Fine-Grained Cat Classification with Layer-Wise Transfer Learning

Rui Wang
Faculty of Applied Sciences, Macao Polytechnic University, Macau, 999078, China

Keywords: Deep Learning, Transfer Learning, Convolutional Neural Networks.

Abstract: In the field of machine learning, transfer learning is a frequently employed technique. It aims to improve a model's performance and learning efficiency by focusing on tasks through feature extraction or fine-tuning. Compared to training a model from scratch, transfer learning can more effectively leverage the knowledge from the source task to achieve superior outcomes on the target task. This paper presents two experiments. The first investigates the impact of unfreezing different numbers of convolutional layers on model performance. The second compares the outcomes of fine-tuning a portion of the convolutional layers after initially training the weights of the fully connected layers with those of unfreezing an equal number of convolutional and fully connected layers simultaneously. The experimental results suggest that adding more convolutional layers to the unfreezing sequence can typically help the model learn finer and more specific features, thereby enhancing performance metrics. It is crucial to find a balance between the model's generalization capability and its learning capacity, as excessive unfreezing can increase the risk of overfitting. Moreover, the more efficient approach of unfreezing the convolutional layers after first training the weights of the fully connected layers confirms the feasibility of transfer learning and the advantages of a step-by-step transfer learning strategy.

1 INTRODUCTION

It is generally straightforward for machines to identify and discriminate between distinct types of items. However, because of their strong resemblance, the work of categorization becomes more difficult when dealing with the same class of things. Fine-grained categorization is crucial in this situation (Xiangxia, 2001). Fine-grained classification, as the name implies, attempts to extract small information for more accurate categorization (Wei, 2019). Biological detection and medical image analysis are two areas that make extensive use of fine-grained categorization. It not only increases classification accuracy but also fosters technological innovation and a wide range of uses. Fine-grained categorization in medical image analysis can assist medical professionals in precisely identifying the features in the picture (Xu, 2024). It is possible to monitor and conserve endangered species more successfully in ecological conservation.

Deep learning is one of the most popular research area in the field of machine learning, which mimics the connections and working patterns of neurons in the human brain, analyzing data through the transmission of signals between neurons and generating outputs through a series of calculations (Guo, 2019). Unlike traditional machine learning, deep learning analyzes vast volumes of data and uses multi-layer neural networks for feature analysis and extraction. Through hierarchical structure, deep learning can gradually capture data features from general to more complex and nuanced (LeCun, 2015). In recent years, deep learning has developed rapidly and has already achieved remarkable success in multiple fields such as image, speech, natural language processing, and more. Deep learning can perform better than conventional techniques in a variety of tasks, handle more complicated data, and require less human interaction. While deep learning has shown great success in the field of image recognition, it still needs to improve on fine-grained classification problems, where many older methods are unable to handle subtle variations. These tasks frequently call for more intricate model architectures and more precise feature extraction.

Fine-grained classification is now advanced in several sectors. For example, the Stanford Dogs

Wang, R.
Enhancing Fine-Grained Cat Classification with Layer-Wise Transfer Learning.
DOI: 10.5220/0013246200004558

In *Proceedings of the 1st International Conference on Modern Logistics and Supply Chain Management (MLSCM 2024)*, pages 167-171
ISBN: 978-989-758-738-2

Datasets use Convolutional Neural Network (CNN) to classify over 120 dog breeds and achieves high accuracy on the model (Dataset, 2011). Also, there are some fine-grained visual classification works that classify specific objects types such as vehicle or plants with data argumentation and transfer learning techniques. Though these methods work out in performance overall, they are not as effective at fine-grained. The feature extraction system does not form a unified framework.

Unlike these research, this work focuses on identifying cat breed using the transfer learning approach and unfreezing the Visual Geometry Group (VGG) convolutional layers layer by layer, with the goal of evaluating the influence of different layer unfreezing classification on the outcome and performance (Simonyan, 2014). This strategy allows the model to be fine-tuned for cat breed-specific details while retaining pre-trained features, paying particular attention to subtle differences such as ear shape, coat color, and eye shape. By testing with different freezing layers (bottom to top) and fine-tuning their weights, this work determines the optimal number of unfrozen layers to improve classification accuracy while maintaining computing efficiency. This detailed analysis method is in sharp contrast to existing research and promotes new research progress in fine-grained classification, especially in the field of cat breed classification.

2 METHOD

2.1 Dataset

Dataset of cat breed classification was leveraged in this work (Diffran, 2022). This work picked up 5 different kinds of cat breed: Calico, Persian, Siamese, Tortoiseshell and Tuxedo. In this experiment, data cleaning methods were used to improve the quality and consistency of the data. Through the above simple processing, dirty data is processed into clean data, laying a solid foundation for subsequent analysis or modeling. To retrieve the dataset, the code first reads the CSV file using the pd.read_csv() function. In order to guarantee that only legitimate data was kept in the data set, missing values were then processed, and the rows that did not match the image were filtered out using the dropna() method. Also, the code transforms the id column to string format to prevent type mismatch issues when string matching is performed. To guarantee that each ID is unique, the code additionally eliminates duplicate IDs using the drop_duplicates() method. To verify the correctness of the file path, the code recursively searches through the picture folder, extracts the first eight characters of the image file name, and compares it with the ID in the CSV file. The aforementioned basic processing transforms filthy data into clean data, providing a strong basis for further analysis or modelling. Then this work splits the dataset into training set and validation set with the ratio of 8:2. In addition, data augmentation is applied to preprocess the data like: rescale, rotation, width shift, height shift, shear, zoom, fill mode, horizontal and so on. These data preprocessing methods are collectively applied within the training and some of the validation data generators, effectively increasing data diversity, enhancing the model's generalization ability, and enabling it to better handle various input scenarios, thereby boosting model performance.

2.2 Models

Many well-developed models for picture categorization are currently available in the field of computer vision. For instance, the VGG16 model is a classic deep convolutional neural network that consists of many pooling layers, three fully connected layers, and thirteen convolutional layers (Simonyan, 2014). With its straightforward and effective design, which makes extensive use of 3*3 convolutional layer filters, it is able to better extract more information for training.

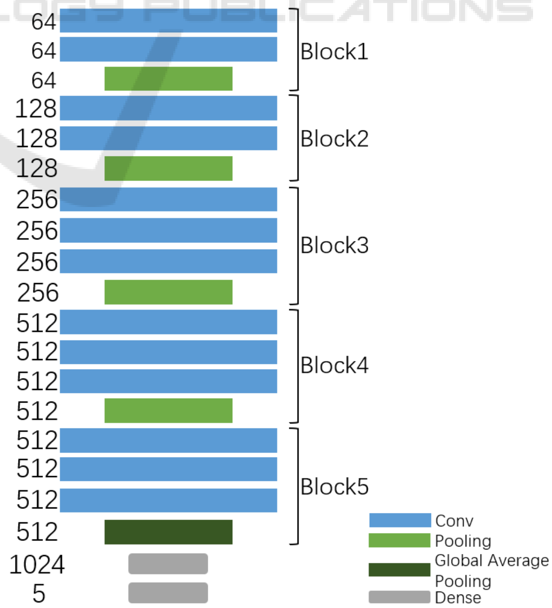

Figure 1: Architecture of VGG model (Figure Credits: Original).

Transfer learning strategy is used in this research. A new model was created by using the weight of the convolution layer that was pre-trained by VGG16 in the ImageNet dataset for feature extraction and customizing the classification layer to adapt the cat breed classification task. Its architecture is demonstrated in Figure 1 (Zhuang, 2020).

Unfreezing various convolutional layers may significantly affect the model's performance on new tasks since different convolutional layers extract picture information at varying levels of complexity and abstraction. This thorough approach to thaw training will offer deeper understanding of difficult problems like fine-grained categorization, assisting in the discovery of more efficient transfer learning techniques and useful advice. Also, the training methodology itself will also have an impact on the outcome. For instance, distinct approaches, such as immediate unfreezing of Fully Connected (FC) and convolutional layers in transfer learning and unfreezing of FC and convolutional layers in stages, may provide varying results. This work may further advance the state-of-the-art development in the field of image classification and make significant progress toward more precise and effective image classification technology by refining the model structure and increasing the sensitivity to minute characteristics.

2.3 Evaluation Matrices

This study utilizes the confusion matrix, F1 value, accuracy, recall, and precision as well as the roc curve indices for validation while measuring the performance of the model.

3 EXPERIMENT AND RESULT

3.1 Training Details

In the experiment, a lot of hyperparameters were selected. Initially, batch size is set to 32, meaning that there are 32 samples in each training batch. This has benefits for both training speed and memory usage. Then, target size is set to (224, 224), the input image will be resized to 224 x 224 pixels in order to comply with the neural network's need for a fixed size input. Furthermore, class mode is set to 'categorical,' which implies that the array of labels formed in one-hot encoding can be usefully returned for multi-class classification issues. Additionally, a binary vector representing each category is created, with only one element being 1 and the remaining elements being 0, indicating which category the sample falls into. In order to help the model better converge to the ideal solution, this work also utilized Adam as the optimizer, categorical crossentropy as the loss function and accuracy as the metric. Besides, early stopping, a callback function, is utilized to halt model training early in order to prevent overfitting. This work trained on Colab and utilized the GPU as a hardware resource to continuously optimize the convergence speed and accuracy of the model by constantly adjusting the learning rate and epoch count.

3.2 Quantitative Performance

There are 499 photos are chosen for each of the five cat breeds, for a total of 2495 photos, in order to assess the model's performance. The impact of varying the number of convolutional layers from deep to shallow unfreezing on the experimental outcomes is examined in Table 1 and Table 2.

This experiment examines, in Table 1, the effects of varying the number of unfreezing layers on model performance, ranging from 0 to 13 layers. And Table 2 displays the experiment and compares the two methods of layer unfreezing: sequential unfreezing (SEQ) and simultaneous unfreezing (SIM). Whereas the SEQ technique unfreezes the fully connected (FC) layer first and then unfreezes the last six convolutional layers, the SIM strategy unfreezes six convolutional layers and fully connected layers at once. Several performance metrics were employed to assess the results of these two experiments.

This study further validated the performance of the best trained model with 6 unfreezed convolutional layers using the confusion matrix, and obtained the following Figure 2 and Figure 3 to further evaluate and visualize the simultaneous unfreezing (SIM) and sequential unfreezing (SEQ) result in Table1.

Table 1: Performance using different number of unfreezing layers.

	ACC	recall	precision	F1	ROC
0	.56	.56	.59	.56	.84
3	.79	.79	.80	.80	.96
6	.86	.86	.87	.86	.98
9	.84	.84	.84	.84	.97
11	.84	.84	.84	.84	.97
13	.84	.84	.84	.84	.97

Table 2: Unfreezing 6 convolutional layers simultaneously (SIM) vs. unfreezing the FC layer first and then unfreezing the 6 convolutional layers sequentially (SEQ).

	ACC	recall	precision	F1	ROC
SIM	.86	.86	.87	.86	.98
SEQ	.89	.89	.89	.89	.99

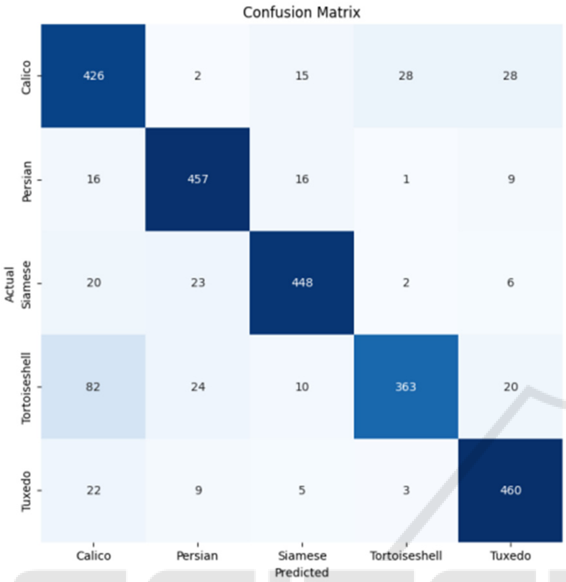

Figure 2: Confusion matrix of SIM (Figure Credits: Original).

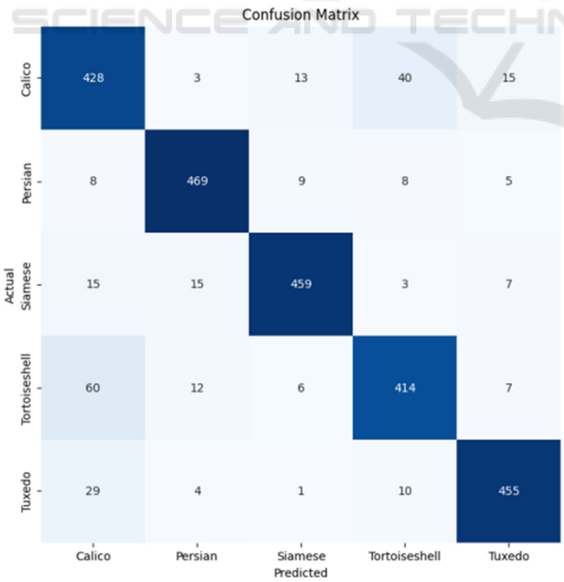

Figure 3: Confusion matrix of SEQ (Figure Credits: Original).

4 DISCUSSIONS

The starting learning rate and number of epochs were both set to 1e-4 in all experiments. To get the model to converge, the learning rate and the number of epochs were then progressively decreased. The experimental findings indicate that learning more precise and fine-grained characteristics is typically possible for the model by unfreezing additional convolutional layers, and a better development trend is indicated by a number of indicators. And with six convolutional layers unfrozen, it operates at its peak. However, after unfreezing an excessive number of convolutional layers, the author discovered that the model's performance had somewhat declined. This could be because there is a greater chance of overfitting when there are too many unfrozen convolutional layers since the model may overfit the noise in the training set. To create a model that performs well on untested data, it is therefore vital to balance the learning and generalization capabilities of the model while unfreezing varying numbers of convolutional layers.

This study the risk of overfitting in this experiment because the dataset used was moderate. The author attempt using a bigger data set in the next research to see if the author can enhance the model's performance any further. However, this experiment just employed the VGG network structure. In the future, the author will try to run comparison studies using various network architectures to see how the quantity of unfreezing layers affects them (Alzubaidi, 2021). To further enhance the model's performance, the author would also like to try dynamic unfreezing, unfreezing progressively, and adaptively changing the number of unfreezing layers based on the training procedure.

5 CONCLUSIONS

This work employs the transfer learning method to completely demonstrate how feature extraction or fine-tuning techniques constantly increase the model's performance and learning efficiency, which has a significant impact on the categorization of delicate images. A large number of studies have shown that transfer learning has achieved remarkable results in fields such as image recognition, natural language processing, and speech recognition. Transfer learning has the potential to greatly increase model accuracy, generalization capabilities, and data efficiency by utilizing models that have already been trained on pertinent tasks. Most notably, transfer learning can be used to get around the problem of scarce data by applying it to cross-domain tasks. Transfer learning is a common approach in real-world applications that offers a workable way to quickly design and implement machine learning models. The author aims to experiment with finer-grained classification tasks in the future, including those involving plants, animals, products, medical diagnostics, etc., in order to assist formalize domain knowledge, create chances for expert knowledge mining and expression, and support the establishment of a knowledge system.

REFERENCES

Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L., 2021. Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. *Journal of big Data*, *8*, 1-74.

Dataset, E., 2011. Novel datasets for fine-grained image categorization. In *First Workshop on Fine Grained Visual Categorization, IEEE Conference on Computer Vision and Pattern Recognition, 5*(1), 1-2.

Diffran N. C., 2022. Cat Breed Classification Computer Vi sion Project. URL: https://universe.roboflow.com/diffr an-nur-cahyo-egmuz/cat-breed-classification. Last Acc essed: 2024/08/26

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S., 2016. Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27-48.

LeCun, Y., Bengio, Y., & Hinton, G., 2015. Deep learning. *nature*, *521*(7553), 436-444.

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

Wei, X. S., Wu, J., & Cui, Q., 2019. Deep learning for fine-grained image analysis: A survey. *arXiv preprint arXiv:1907.03069*.

Xiangxia, L., Xiaohui, J., Bin, L., 2021. Deep Learning Method for Fine-Grained Image Categorization. *Journal of Frontiers of Computer Science & Technology*, *15*(10), 1-10.

Xu, Z., Yue, X., & Lv, Y., 2024. Trusted Fine-Grained Image Classification Based on Evidence Theory and Its Applications to Medical Image Analysis. In Granular Computing and Big Data Advancements, 269-288.

Zhuang, F., Qi, Z., Duan, K., Xi, D., Zhu, Y., Zhu, H., ... & He, Q., 2020. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, *109*(1), 43-76.