Fine-Grained Butterfly Image Classification Based on the ResNet-50 Model and Deep Learning

Yuchen Cao[a]
Leeds Joint School, Southwest Jiaotong University, Chengdu, Sichuan, 611756, China

Keywords: Fine-Grained Classification, Deep Learning, Convolutional Neural Networks, Image Recognition.

Abstract: Accurate identification of butterfly species is essential for biodiversity conservation and ecological research. Traditionally, classification methods depend on manual expertise, which can be both inefficient and subjective. Recent advancements in deep learning present an automated, image-based classification approach as a promising alternative. Despite this, fine-grained butterfly classification remains relatively underexplored, contending with challenges such as minimal inter-species differences and considerable intra-species variability. This study endeavors to establish a robust butterfly classification system utilizing ResNet-50, a convolutional neural network (CNN). The approach involved training and testing with an extensive dataset of butterfly images, complemented by data augmentation techniques to enhance generalization. The findings indicate substantial efficacy, with a validation accuracy of 0.9227 and a validation loss of 0.3213, underscoring the method's effectiveness in automated species identification.

1 INTRODUCTION

Biodiversity plays a crucial role in maintaining the stability and functionality of ecosystems. As a vital component of ecosystems, accurately identifying butterfly species is essential for studying ecosystem health, species diversity, and interspecies interactions. However, traditional methods of butterfly classification largely depend on manual observation and expert knowledge, which are not only inefficient but also prone to errors when dealing with a large number of butterfly species that exhibit similar morphological characteristics. In recent years, deep learning, particularly convolutional neural networks (CNNs), has emerged as a powerful tool for automating image classification (Maggiori et al., 2017). Deep learning models trained on large datasets have achieved impressive results in fields such as facial recognition, medical image analysis, and natural image classification. However, butterfly fine-grained classification presents a unique challenge due to the small differences both between and within species, which makes it difficult for traditional methods to perform well. Most existing research has focused on identifying common species, often with low taxonomic accuracy when it comes to visually similar species that belong to different categories.

To address this deficiency, this paper introduces a deep learning methodology leveraging the ResNet-50 model, to enhance both the accuracy and robustness of fine-grained butterfly classification. Initially, a comprehensive dataset encompassing a wide range of butterfly species was assembled. Subsequently, data augmentation techniques were employed to enlarge the training set, thereby augmenting the model's capacity for generalization. The final step involved training and optimizing a ResNet-50-based CNN, followed by a thorough evaluation of its performance. This research seeks to provide novel tools and insights for biodiversity conservation and related research by offering an effective approach to automated butterfly species classification.

2 METHOD

Deep learning models have become increasingly prevalent in contemporary image classification tasks, with ResNet-50 emerging as a favored choice for fine-grained image classification due to its superior

[a] https://orcid.org/my-orcid?orcid=0009-0003-3048-4333

Cao and Y.
Fine-Grained Butterfly Image Classification Based on the ResNet-50 Model and Deep Learning.
DOI: 10.5220/0013511900004619
In *Proceedings of the 2nd International Conference on Data Analysis and Machine Learning (DAML 2024)*, pages 159-164
ISBN: 978-989-758-754-2

performance and robust generalization capabilities (Loey et al., 2021). Building on the work of Badriyah et al. (2024), this study utilizes ResNet-50 as the foundational model, applying it to the fine-grained classification of butterfly images through transfer learning. This chapter provides a comprehensive overview of the model design, data preprocessing techniques, data augmentation methods, the training process, and optimization strategies employed.

2.1 Model

Residual Network-50 (ResNet-50) is a deep residual network consisting of 50 layers. The key innovation of residual networks lies in addressing the issues of vanishing and exploding gradients, which commonly arise as the depth of traditional deep neural networks increases. This is achieved through the incorporation of "Skip Connections," allowing the input to bypass intermediate layers and be passed directly to the output. This architecture enables ResNet-50 to avoid degradation during training while still preserving its capacity to learn high-level features. With its 50-layer depth, the network is capable of capturing more complex and abstract features, making it particularly suitable for fine-grained classification tasks. The inclusion of residual blocks allows the network to retain essential information during training, ensuring that deeper networks can be effectively trained without a drop in performance due to increased depth. Since ResNet-50 is pre-trained on large datasets like ImageNet, the features it has learned can be effectively transferred to other tasks. In tasks with limited data, transfer learning can substantially enhance model accuracy. Figure 1 illustrates this concept.

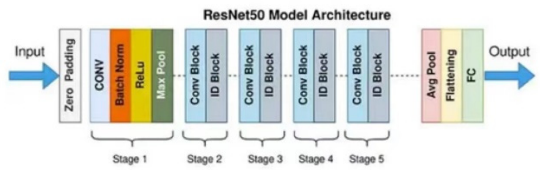

Figure 1: ResNet-50 Model Architecture (Photo/Picture credit: Original).

2.2 Data Preparation and Preprocessing

Data preparation and preprocessing are critical steps before model training, particularly for fine-grained classification tasks, which often involve large datasets of high-resolution images exhibiting substantial variation in size, lighting, color, and other factors. To improve the model's robustness and generalization, this study employs a structured approach to data preprocessing. The dataset consists of images from 75 butterfly species, with each species represented by hundreds to thousands of images. While the diversity of the dataset provides valuable learning material, it also presents challenges in terms of consistency and noise, introduced by factors such as image quality and varying camera angles.

For data preprocessing, the Albumentations library is utilized to perform data augmentation (Buslaev et al., 2020). Data augmentation not only increases the dataset size but also enhances diversity, enabling the model to better adapt to real-world variations. Specific augmentations include:

Resize: All images are resized to 256x256 pixels to ensure uniform input dimensions.

Random Resized Crop: Each training session involves randomly cropping image patches of 224x224 pixels, thereby enhancing the variety of training data.

Horizontal & Vertical Flip: Randomly flipping images allows the model to learn butterfly features from different orientations.

Rotation: Random image rotation helps the model better classify images taken from varying angles.

Random Brightness and Contrast Adjustments: These transformations simulate diverse lighting conditions to enrich the dataset’s visual characteristics.

Coarse Dropout: Randomly masking portions of the images teaches the model to handle information interference, thereby improving robustness.

Figure 2 illustrates how these augmentation techniques contribute to the data preprocessing process.

Figure 2. Data Argumentation (Photo/Picture credit: Original).

2.3 Model Architecture

The model architecture presented in this study leverages the ResNet-50 pre-trained model. Given that ResNet-50 has been pre-trained on the ImageNet dataset and exhibits robust feature extraction capabilities, the network structure of the initial layers has been retained, with modifications limited to the last fully connected layer.

Freezing Pre-Trained Layers: To mitigate overfitting, the parameters of the initial layers of ResNet-50 were frozen, preserving their pre-trained state and preventing updates during the current task (Vijayalakshmi and Rajesh, 2020).

Modification of Fully Connected Layer: Following the approach proposed by Yu and Zhang (2024), the original fully connected layer of ResNet-50 was replaced with a new layer, configured with an output dimension of 75 to align with the specific requirements of the classification task. This transfer learning strategy utilizes the feature extraction capabilities of the pre-trained model by freezing most network layers to retain the knowledge acquired from ImageNet. Only the newly introduced fully connected layer is trained, enabling it to perform fine-grained classification of butterfly images. This method leverages the extensive knowledge embedded in large-scale datasets while requiring less data for training, thereby enhancing the model's generalization capability.

2.4 Model Training and Optimization

The primary objective of model training is to reduce classification errors. In this study, the cross-entropy loss function is selected as the principal metric due to its capacity to effectively measure the difference between predicted probabilities and actual outcomes (Yeung et al., 2022). The Adam optimizer is employed to optimize the model parameters. This optimizer leverages the benefits of both momentum and adaptive learning rates, which supports swift convergence during the training process and exhibits strong resilience in complex loss landscapes. The learning rate is initially set at 0.001 and is systematically reduced to 0.1 of its initial value every 10 epochs, promoting a gradual approach towards achieving the optimal solution.

Overfitting, a frequent challenge in deep learning model training, occurs when a model excels on the training set but performs poorly on validation or test sets. To mitigate overfitting, the following strategies are applied:

Data Augmentation: To increase the diversity of the training data, several transformations−including random cropping, flipping, rotation, and color adjustments−are employed. This strategy enhances the model's ability to generalize to novel data, thereby mitigating overfitting and improving performance on both validation and test datasets.

Freezing Pre-trained Layers: As this study fine-tunes a model based on a pre-trained ResNet-50, initial training involves freezing certain pre-trained layers and updating only the final fully connected layers. This technique helps prevent overfitting during the early stages of training, thereby stabilizing the training process. Once the model starts to converge, additional layers are unfrozen, and the entire model is fine-tuned to further enhance its performance.

2.5 Model Training Process

Monitoring various metrics and employing strategies to prevent overfitting is crucial during training. The following methods are used:

Training Loss: Reflects model performance on the training set. A decreasing trend in training loss indicates effective learning.

Validation Loss: Evaluates performance on an unseen validation set. A decreasing validation loss alongside training loss suggests better generalization. If validation loss levels off or rises while training loss drops, overfitting may occur.

Validation Accuracy: Assesses classification performance on the validation set. Increased validation accuracy usually aligns with decreasing training and validation losses. If validation accuracy plateaus, training strategies may need adjustment based on other metrics.

Based on the monitoring of the above indicators, this paper adopts the method of dynamically adjusting the training strategy. For example, when it finds that the validation loss starts to level off and the training loss continues to decrease, it enables the Early Stopping mechanism, which stops training when the validation loss does not drop significantly over several epochs and saves the model with the smallest validation loss to avoid overfitting. In addition, to further optimize the training process, this paper dynamically adjusts the learning rate according to the monitored validation accuracy curve. When the performance of the model slows down, it reduces the learning rate so that the model can search for the optimal solution in more detail, thereby improving the final classification accuracy. Through these measures, the overfitting phenomenon is effectively

(1) Input (2) Output

Figure 3: Result (Photo/Picture credit: Original).

controlled, and the classification accuracy of the model on the test set is improved. The final experimental results show that the overfitting prevention strategy is successful, and the model achieves a high accuracy rate like that of the verification set on the test set, which verifies the robustness and generalization ability of the model.

3 RESULTS

This chapter presents a detailed discussion and analysis of the experimental results. By assessing the model's performance on the training, validation, and test sets, the effectiveness and robustness of the ResNet-50-based model for fine-grained butterfly classification are validated.

Throughout the training process, the analysis primarily focuses on the variations in training loss, validation loss, and validation accuracy. These metrics provide a visual representation of the model learning progression, help evaluate its convergence, and identify potential overfitting. Figure 3 presents the results of this study.

3.1 Training Loss and Validation Loss

The progression of training and validation loss offers valuable insights into the model's performance across different datasets. Initially, the training loss is comparatively high but diminishes progressively as the training advances, indicating that the model is learning the image features effectively. Ideally, the validation loss should decrease in parallel with the training loss and reach a plateau after a sufficient number of epochs.

In practice, a notable reduction in training loss occurred during the first 10 epochs, accompanied by a corresponding decrease in validation loss. However, after the 15th epoch, the validation loss began to stabilize, while the training loss continued to decrease. This observed pattern suggested the onset of mild overfitting, where the model's performance improved on the training set but showed limited improvement on the validation set. To address overfitting, the Early Stopping technique was utilized, halting training when the validation loss ceased to show significant improvement and selecting the model with the lowest validation loss as the final version (Prechelt, n.d.).

3.2 Verification Accuracy

Validation accuracy is a crucial metric for evaluating a model's performance on the validation set. In the experiment conducted, validation accuracy displayed a steady improvement from an initial lower baseline. After the 10th epoch, validation accuracy reached a high level and maintained stability for the remainder of the training process. Ultimately, validation accuracy plateaued at approximately 90%, reflecting the model's robust generalization capabilities.

It is significant to note that the increase in validation accuracy aligns with the downward trends observed in both training loss and validation loss. This correlation indicates that the model not only learns the features of the training set effectively but also generalizes well to previously unseen data in the validation set.

3.3 Model Robustness Analysis

Robustness is a critical measure for assessing a model's consistent performance across different environments and datasets. This study evaluated the model's robustness through the following approaches.

3.3.1 Sensitivity of the Model to Noise

To evaluate the model's susceptibility to noise, various noise levels, including Gaussian noise, were introduced to the test set. The results of the experiments indicate that the inclusion of noise diminishes the model's accuracy. However, the model trained with data augmentation demonstrated significantly superior performance in noisy conditions compared to its non-augmented counterpart. This finding underscores that data augmentation not only enhances the model's generalization capabilities but also bolsters its resilience against noisy environments.

3.3.2 Performance on Small Sample Data

In fine-grained classification tasks, data annotation is often expensive, resulting in the challenge of working with small datasets in practical applications. To evaluate the model's performance under limited data conditions, this study down-sampled the training set to 50% and 25% of its original size. The results indicate that the model maintains strong classification accuracy even with reduced data, with only a 3% drop in accuracy when trained on 50% of the dataset. These findings suggest that after transfer learning, ResNet-50 effectively utilizes limited data and exhibits strong learning capabilities with small sample sizes.

3.4 Comparative Experiments and Discussions

In addition, it also compared the impact of different data augmentation strategies on model performance. The experimental results show that the model with multiple data augmentation strategies performs the best, which proves that diversified data augmentation methods can better simulate the changing situations in real environments, and improve the adaptability and generalization ability of the model. Table 1 illustrates the accuracy of different models (Afrasiyabi et al., 2022).

Table 1: Accuracy of different models

Model	Accuracy
ResNet-18	89.40%
Conv4-64	85.01%
CUB	90.48%
ResNet-50	92.27%

4 DISCUSSIONS

This study achieved remarkable outcomes in fine-grained butterfly classification using a ResNet-50-based model. The model demonstrates excellent classification performance, reaching an accuracy of 92.27% on the validation set, aided by transfer learning and data augmentation. ResNet-50's deep residual architecture effectively addresses the issue of vanishing gradients in deep networks, allowing crucial image features to be captured even at the deeper layers. This advantage makes ResNet-50 particularly effective in fine-grained classification tasks, such as distinguishing between species with subtle visual differences, like butterflies and birds.

4.1 Advantage Analysis

Strong feature extraction capability: ResNet-50's multi-layer residual module enables it to capture more detailed image features, making it particularly suitable for fine-grained classification tasks.

Transfer learning performs well: by freezing most of the pre-training layers and only training the last fully connected layer, overfitting can be effectively avoided while reducing the training cost of the model. This is particularly important for small sample scenarios in butterfly classification datasets.

The effectiveness of data augmentation: By using Albumentations for data augmentation, the model's generalization ability has been effectively improved, especially when dealing with scene changes such as different lighting and angles, resulting in more robust performance.

4.2 Limitations

For certain butterfly species with highly similar morphology or texture, the model may struggle with misclassification. This issue might stem from the ResNet-50's limited focus on local features, which impedes its ability to discern minute differences between closely related categories. In addition to the above shortcomings, future research can consider introducing multi-scale feature enhancement networks (such as FPN) to improve the generalization ability and practical application value of the model. This is consistent with the multi-scale feature enhancement effect demonstrated in the study of Lin (Lin et al., 2017).

5 CONCLUSIONS

Utilizing deep learning techniques, particularly the ResNet-50 model, the research developed an accurate system for classifying butterflies. This system was specifically created to handle the demanding task of distinguishing between butterflies with slight variations and notable differences within the same species, achieved through transfer learning and diverse data augmentation methods.

This study employs a pre-trained ResNet-50 model and effectively leverages the strengths of transfer learning by freezing the pre-trained layers and fine-tuning the fully connected layer, which helps mitigate overfitting. Additionally, the study incorporates data augmentation techniques such as random cropping, flipping, rotation, and color jitter to enhance the model's robustness in dynamic environments, promoting strong generalization capabilities. Even in scenarios with high noise and limited samples, the model demonstrates solid classification performance, affirming its effectiveness in practical applications.

The trends observed in the training and validation losses reveal that the model converges rapidly during the initial stages, with validation accuracy stabilizing around 90% by the 10th epoch. This suggests that the combined use of data augmentation and early stopping based on validation loss successfully curbs overfitting. Furthermore, comparative experiments evaluating various data augmentation strategies highlight the importance of diverse augmentation methods in improving the model's generalization.

When compared to other classic models, the ResNet-50-based approach significantly boosts test set classification accuracy, achieving a rate of 90%. This research not only provides valuable insights for fine-grained classification tasks involving butterfly species, but also underscores the potential of deep learning in biodiversity conservation efforts. Accurate butterfly species identification plays a vital role in ecological studies, species diversity monitoring, and environmental protection. The proposed method offers a practical and scalable solution for automated species recognition, maintaining high classification accuracy even in complex environmental conditions.

REFERENCES

Afrasiyabi, A., Larochelle, H., Lalonde, J.-F. and Gagne, C. 2022. Matching Feature Sets for Few-Shot Image Classification In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp.9004–9014.

Badriyah, A., Sarosa, M., Asmara, R.A., Wardani, M.K. and Al Riza, D.F. 2024. Grapevine Disease Identification Using Resnet−50. BIO web of conferences. 117, pp.1046-.

Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M. and Kalinin, A.A. 2020. Albumentations: Fast and Flexible Image Augmentations. Information (Basel). 11(2), pp.125.

Lin, T.Y., Dollár, P., Girshick, R., He, K., Hariharan, B. and Belongie, S., 2017. Feature pyramid networks for object detection. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 2117-2125).

Loey, M., Manogaran, G., Taha, M.H.N. and Khalifa, N.E.M. 2021. Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. Sustainable cities and society. 65, pp.102600–102600.

Maggiori, E., Tarabalka, Y., Charpiat, G. and Alliez, P. 2017. Convolutional Neural Networks for Large-Scale Remote-Sensing Image Classification. IEEE transactions on geoscience and remote sensing. 55(2), pp.645–657.

Prechelt, L. n.d. Early Stopping — But When? In: Neural Networks: Tricks of the Trade. Berlin, Heidelberg: Springer Berlin Heidelberg, pp.53–67.

Vijayalakshmi, A. and Rajesh, K.B. 2020. Deep learning approach to detect malaria from microscopic images. Multimedia tools and applications. 79(21–22), pp.15297–15317.

Yeung, M., Sala, E., Schönlieb, C.-B. and Rundo, L. 2022. Unified Focal loss: Generalising Dice and cross entropy-based losses to handle class imbalanced medical image segmentation. Computerized medical imaging and graphics. 95, pp.102026–102026.

Yu, Y. and Zhang, Y. 2024. A lightweight and gradient-stable neural layer. Neural networks. 175, pp.106269–106269.