Butterfly Image Classification: Results and Evaluation

March 17, 2025

1 Introduction

Butterfly image classification is a deep learning task where a model is trained to identify and classify butterfly species from images. The goal is to develop a model that can recognize different species based on their visual features such as wing color, patterns, and shapes.

2 Importance of Butterfly Classification

- **Biodiversity Conservation:** Helps researchers track and monitor butterfly species, some of which are endangered.
- **Ecological Studies:** Butterflies are bio-indicators, meaning their presence or absence can indicate environmental changes.
- **Citizen Science & Education:** Useful for educational tools and citizen science projects where people can contribute to butterfly identification.
- **Automation in Taxonomy:** Reduces the manual effort required to classify species by entomologists.

3 Dataset for Butterfly Classification

A dataset for butterfly classification typically contains:

- Images of various butterfly species.
- Labels corresponding to species names.
- Metadata (e.g., location, time of capture, photographer details).

3.1 Example Datasets

- Butterfly Image Dataset (from Kaggle)
- iNaturalist Dataset
- Custom-collected datasets from field studies

4 Preprocessing Steps

Before training a model, images undergo preprocessing:

- **Resizing:** Standardize image dimensions (e.g., 224x224 pixels for CNN models).
- **Normalization:** Scale pixel values (e.g., between 0 and 1) to improve training stability.
- **Data Augmentation:** Apply transformations like rotation, flipping, and brightness adjustments to improve generalization.

5 Model Selection for Classification

Common deep learning models for butterfly classification:

- **CNNs (Convolutional Neural Networks):** Extract spatial features from images.
- **Pretrained Models (Transfer Learning):** Models like ResNet, VGG16, EfficientNet, and InceptionV3 can be fine-tuned for classification.
- **Vision Transformers (ViTs):** Advanced architectures for image classification.

6 Training the Model

- **Loss Function:** Cross-entropy loss is used for multi-class classification.
- **Optimizer:** Adam, SGD, or RMSprop optimizes weight updates.
- **Epochs & Batch Size:** Typically trained for 10–50 epochs with batch sizes like 32 or 64.

7 Model Evaluation Metrics

To measure performance, we use:

- **Accuracy:** Percentage of correctly classified images.
- **Precision, Recall, and F1-score:** Measures class-specific performance.
- **Confusion Matrix:** Helps visualize misclassification patterns.

8 Future Enhancements

- **Multi-label Classification:** Some species have similar traits, requiring more detailed identification.
- **Explainability:** Use Grad-CAM to highlight which parts of the image influenced classification.
- **Self-supervised Learning:** Leverage unlabeled data to improve model performance.

9 Results for ResNet-50

- **Training Accuracy:** Increased from 38.76% to 92.50% over 10 epochs.
- **Validation Accuracy:** Started at 77.08% and peaked at 92.54% (Epoch 7).
- **Overfitting Considerations:** The validation loss reached its lowest at Epoch 6 (0.2547), while validation accuracy peaked at Epoch 7. After that, the validation loss increased slightly, suggesting some overfitting.

10 Results for EfficientNetB0

- **Training Accuracy:** Improved from 26.27% to 90.17% in 10 epochs.
- **Validation Accuracy:** Reached 93.38% at the last epoch, which is higher than the previous run.
- **Validation Loss:** Dropped consistently to 0.2408, showing better generalization.
- **Improvements Over the Last Run:**
  - Better Initial Convergence: The first epoch had lower accuracy (26.27% vs. 38.76%) but caught up quickly in later epochs.
  - Stronger Generalization: Validation accuracy kept increasing till the last epoch (93.38%). Validation loss improved more consistently compared to the previous run.

11 Results for VGGNet-16

- **Training Accuracy:** Improved from 19.39% to 81.69% (slower compared to previous runs).
- **Validation Accuracy:** Reached 87.00%, which is lower than your best model (93.38%).
- **Validation Loss:** Dropped to 0.4842 (Epoch 8) but slightly increased afterward, indicating overfitting.

- **Key Observations & Issues:**
  - **Slower Convergence at the Start:** First epoch accuracy (19.39%) is lower than before, which means the model struggled more in early training. Could be due to suboptimal initialization, learning rate, or batch size.
  - **Validation Loss Inconsistency:** Loss increased after Epoch 8, suggesting overfitting. Validation accuracy stagnated from Epoch 7 onwards.
  - **Underperformance Compared to Previous Runs:** In the last run, accuracy reached 90.17%, and validation accuracy hit 93.38%. Here, the final validation accuracy is only 87.00

12 Conclusion

In conclusion, the models tested for butterfly image classification—ResNet-50, EfficientNetB0, and VGGNet-16—show varied performances. ResNet-50 demonstrated strong improvement over epochs, with high training and validation accuracy, though some overfitting was observed. EfficientNetB0 exhibited the best generalization performance, with consistent validation accuracy increase and better convergence compared to ResNet-50. On the other hand, VGGNet-16 showed slower convergence and underperformed compared to the other models, indicating that further fine-tuning or architectural changes might be needed. Overall, EfficientNetB0 emerged as the top-performing model, making it the ideal candidate for future work in butterfly image classification tasks.