SPRING 2025

Butterfly Image Classification With ResNet50

By
Juana Wong

CSc 44700 P
Introduction to Machine Learning

Agenda

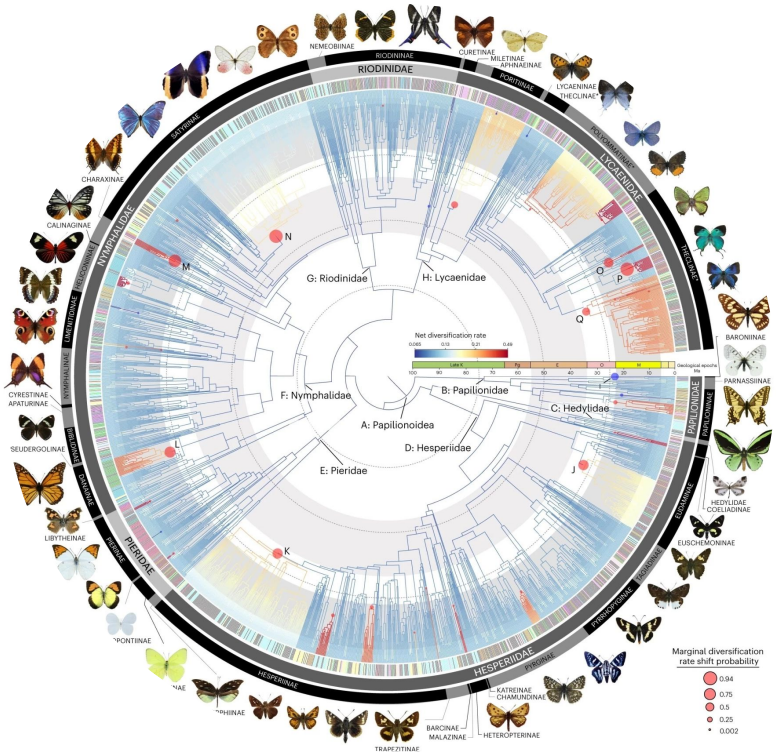
01 Problem & Dataset

02 Data Pipeline & Augmentation

03 Model Development & Training

04 Results, Evaluation & Discussion

01 Identify, Classify, Conserve, Support — Butterfly Populations



- approx. 17,500 to 20,000 butterfly species worldwide
- population crash!
  - U.S.A.: “107 species **declined** by **more than 50%**, and 22 species declined by **more than 90%**” from 2000–2020

Edwards, Collin B., et al. “Rapid Butterfly Declines across the United States during the 21st Century.” *Science*, vol. 387, no. 6738, 7 Mar. 2025, pp. 1090 – 1094, https://doi.org/10.1126/science.adp4671.

Rapidly and accurately identify butterflies to support large-scale **monitoring** of population trends & **conservation** initiatives.

Kawahara, Akito Y., et al. “A Global Phylogeny of Butterflies Reveals Their Evolutionary History, Ancestral Hosts and Biogeographic Origins.” *Nature Ecology & Evolution*, vol. 7, no. 7, 15 May 2023, pp. 1 – 11, https://doi.org/10.1038/s41559-023-02041-9.

01 “Butterfly Image Classification” (Version 2) Dataset

DePie

phucthaiv02 (he/him)

Data Explorer

Version 2 (237.31 MB)

- test
- train
- Testing_set.csv
- Training_set.csv

Thái Văn, Phúc (2024). Butterfly Image Classification (Version 2) [Dataset]. Kaggle. https://www.kaggle.com/datasets/phucthaiv02/butterfly-image-classification/data

Description

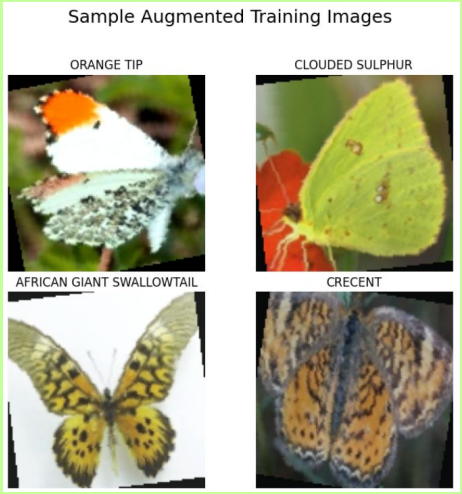
- 9,000+ sample images
- 75 predefined classes/labels
- ‘train’ set – labeled (.csv)
- ‘test’ set – unlabeled

original ‘train’ dataset:
(‘train’, ‘val’): 80–20 split

train: 5,199 images
val: 1,300 images
test: 2,786 images

02

Data Pipeline & Augmentation



Raw Images (JPG)	Data Loading	Data Preprocessing	Data Augmentation (Train)	DataLoader	Model Training
	read filepaths, csv				
	split to train/val/test				
		Resize (128x128)			
		Normalize using ImageNet mean/std			
			RandomResizedCrop		
			RandomHorizontalFlip		
			RandomRotation (15°)		
			ColorJitter		
				Batch, shuffle, load	
					Simple Butterfly CNN
					ResNet50

To increase data diversity, improve generalization.
To overcome limitations of a small dataset.

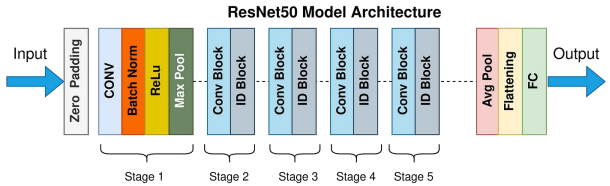
03 Baseline Model & Transfer Learning Approach

SimpleButterflyCNN

- CNN (4 conv layers, ReLU, max pooling), achieved ~73% val accuracy at 16/20 epochs
- Limitation: plateau from limited model capacity, underfitting

ResNet50

- Pretrained model on ImageNet (50 layers, 16 residual blocks)
- Final layer adapted for 75 classes
- Train/validation gap is small (96.7% vs. 91.2%), indicating strong generalization and minimal overfitting



--- Final Metrics (SimpleButterflyCNN) ---
Final Training Loss: 0.8904
Final Validation Loss: 0.9954
Final Training Accuracy: 72.88%
Final Validation Accuracy: 73.54%
Best Validation Accuracy: 73.54% at epoch 19

--- Final Metrics (ResNet50) ---
Final Training Loss: 0.0926
Final Validation Loss: 0.4293
Final Training Accuracy: 96.69%
Final Validation Accuracy: 91.15%
Best Validation Accuracy: 91.77% at epoch 18
Best Validation Loss: 0.4074 at epoch 10

=== Training Summary (ResNet50) ===
Model: ResNet50 (pretrained on ImageNet, final layer for 75 classes)
Epochs: 20
Optimizer: Adam
Batch size: 32
Learning Rate Scheduler: StepLR (step_size=7, gamma=0.1)
Initial Learning rate: 0.001
Final Learning rate after 20 epochs: 0.00001 (1e-05)
Loss function: CrossEntropyLoss
Data augmentation: RandomResizedCrop, Flip, Rotation, ColorJitter
Device: CPU

04 Baseline Model vs. ResNet50 – Training and Validation Progress

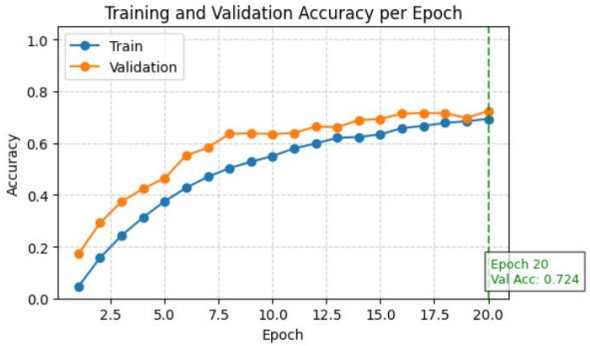
Accuracy & Loss

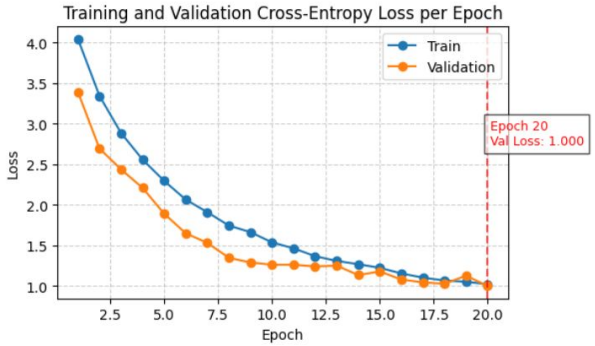
The SimpleButterflyCNN (baseline model) plateaued around 72% accuracy at epoch 20.

The ResNet50 model's accuracy and loss improved rapidly during the first 8–10 epochs, reaching over a 90% validation accuracy by epoch 10.

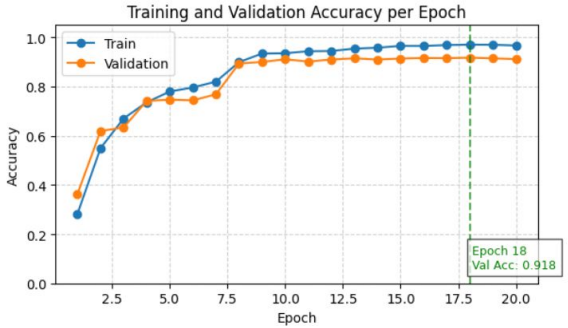
After that, both training and validation performance plateaued. Training beyond 10 epochs offered little benefit, so stopping earlier can be done to prevent overfitting.

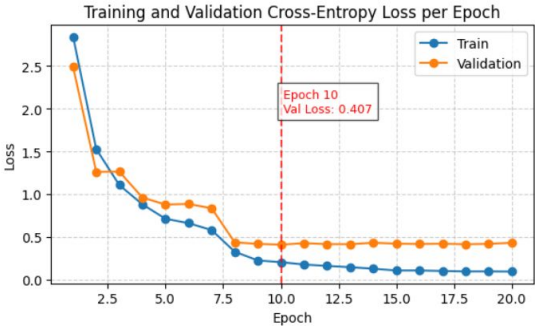
SimpleButterflyCNN



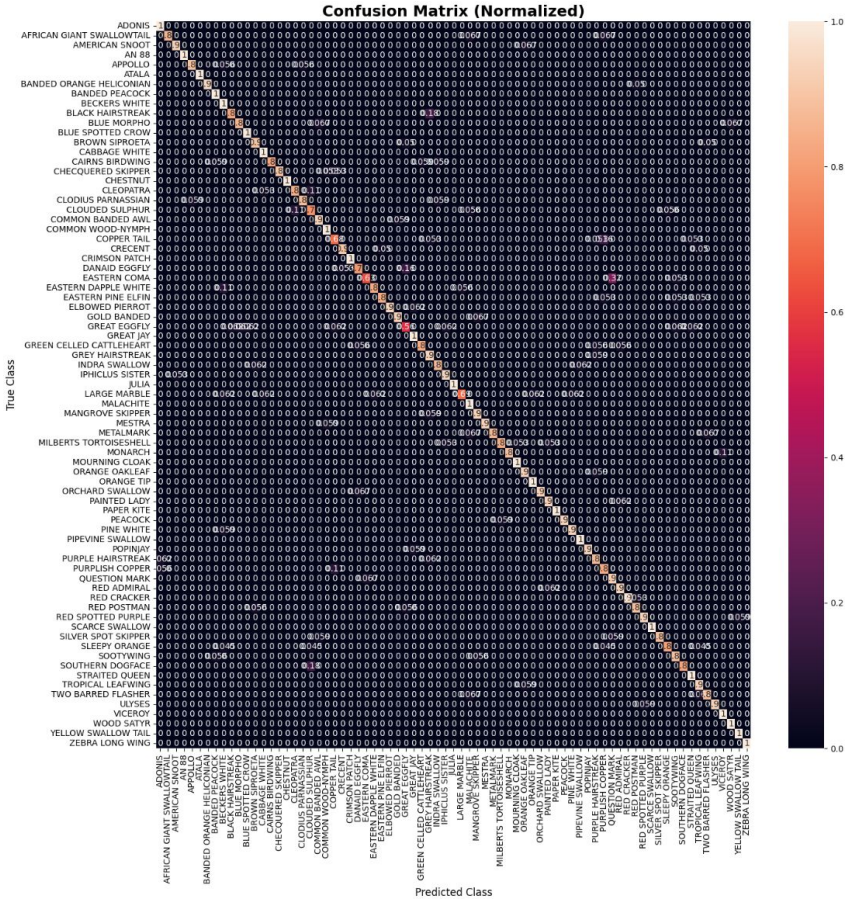


ResNet50





04 Normalized Confusion Matrix (Recall)



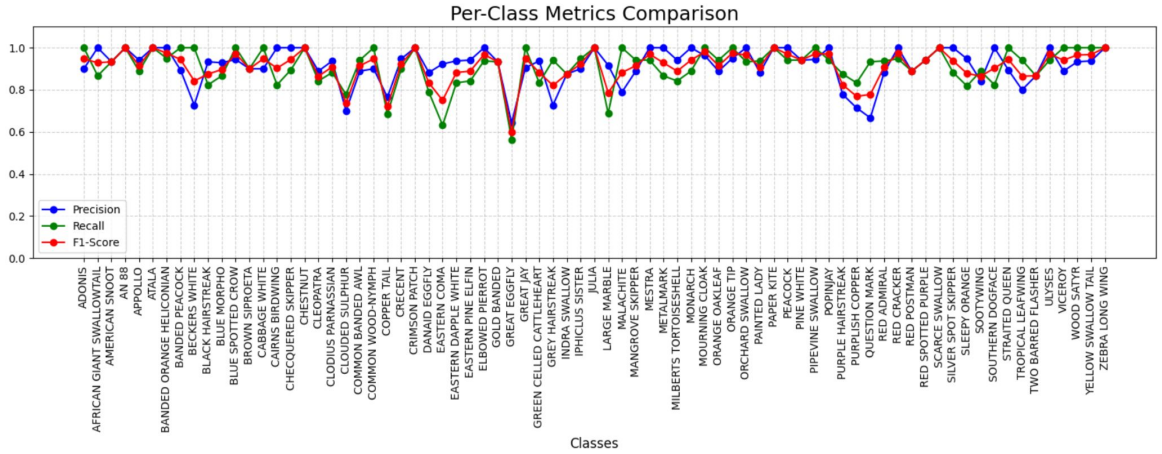
Normalized Confusion Matrix

Most butterfly classes show high recall (0.9+) – the model correctly identifies the majority of samples for most classes.

Several classes achieved perfect recall (1.0) – no misclassifications for those categories.

A few low-performing classes (<0.8) – the model struggles to distinguish these classes from others (misclassification).

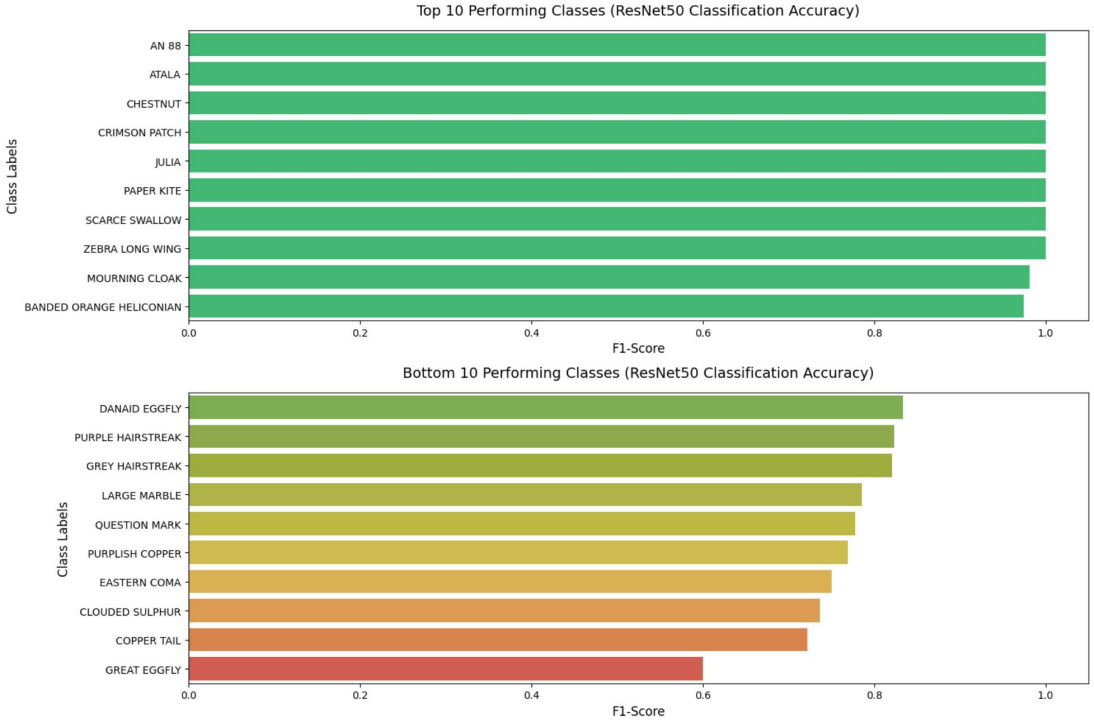
04 Per-Class Performance Metrics – Precision, Recall, F1-Score



ResNet50 Per-Metric Comparison (F1-Score)

Overall Accuracy: 0.91 (91%)
Highest Performing Classes (8): 1.0 (100%)
Lowest Performing Class (1): 0.6 (60%)
Average Classes: Between 0.85 and 0.97 – indicates strong and consistent performance across most categories.

04 Classification Performance/Accuracy Rankings



Top 10 & Bottom 10

Train: avg. 86 images per class

Validation: avg. 17 images per class

Some classes had perfect 100% classification accuracy.

Some classes had moderate accuracy when classifying species.

One class, the **Great Eggfly**, noticeably performed the worst.

04 Great Eggfly (0.6 F1–Score)

Great Eggfly

Precision: 0.64 – 64% of images predicted as Great Eggfly was correct

Recall: 0.56 – of all Great Eggfly images, 56% were correctly predicted

F1–Score: 0.60 – harmonic mean of precision and recall; moderate/average score – can be improved

Support: 16 – small sample size, 16 true images in the set

04 Great Eggfly (0.6 F1-Score)

Great Eggfly

Precision: 0.64 **Recall**: 0.56 **F1-Score**: 0.60 **Support**: 16

Incorrect Prediction

Sootywing

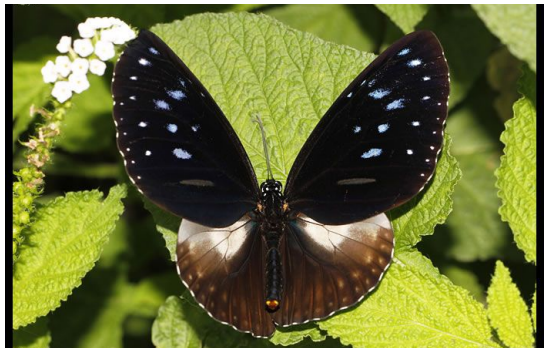
04 Great Eggfly (0.6 F1-Score)

Great Eggfly

Precision: 0.64 **Recall**: 0.56 **F1-Score**: 0.60 **Support**: 16

Incorrect Prediction

Blue Spotted Crow



04 Great Eggfly (0.6 F1-Score)

Great Eggfly

Precision: 0.64 **Recall**: 0.56 **F1-Score**: 0.60 **Support**: 16

Incorrect Prediction

Iphiclus Sister

04 Discussion

Butterfly Image Classification With ResNet50

Collect more data

Increase support, more diverse examples reduces false positives and false negatives

Adjust model values

Eg. adjust model to favor precision over recall, or vice versa

Data augmentation

Apply different transformations and adjustments to sample images, improve generalization

Improve model architecture

Adjust depth and width of CNN or try different architectures, eg. custom CNN, ResNet50, etc.

Hyperparameter tuning

Change learning rate, batch size, optimizer, number of epochs, etc.

Edwards, Collin B., et al. “Rapid Butterfly Declines across the United States during the 21st Century.” *Science*, vol. 387, no. 6738, 7 Mar. 2025, pp. 1090 – 1094, https://doi.org/10.1126/science.adp4671.

Kawahara, Akito Y., et al. “A Global Phylogeny of Butterflies Reveals Their Evolutionary History, Ancestral Hosts and Biogeographic Origins.” *Nature Ecology & Evolution*, vol. 7, no. 7, 15 May 2023, pp. 1 – 11, https://doi.org/10.1038/s41559-023-02041-9.

Thái Văn, Phúc (2024). Butterfly Image Classification (Version 2) [Dataset]. Kaggle. https://www.kaggle.com/datasets/phucthaiv02/butterfly-image-classification/data