Enchanted Wings: Marvels of Butterfly Species

1. Introduction

Butterflies are essential indicators of biodiversity and ecosystem health. However, traditional identification methods require domain expertise and manual efforts. This project, titled "Enchanted Wings: Marvels of Butterfly Species," aims to automate butterfly species identification through transfer learning-based image classification.

The system utilizes pre-trained convolutional neural networks (CNNs) to extract meaningful visual features, reducing training time and improving classification accuracy. It is designed to serve multiple domains including biodiversity monitoring, ecological research, and citizen science education.

2. Objectives

- Build a deep learning model capable of classifying butterfly species.
- Use transfer learning to leverage pre-trained CNN architectures.
- Minimize computational requirements while maximizing performance.
- Enable real-time and mobile-friendly species identification.
- Deploy in field, lab, or educational environments.

3. Dataset Description

Total Images: 6,499

Total Classes: 75 butterfly species

Data Split: 70% training, 15% validation, 15% testing

Sources: Open-source biodiversity image datasets (e.g., Kaggle, iNaturalist)

Preprocessing: Resized to 224x224 px, normalization, data augmentation (rotation, flipping, zoom, brightness)

4. Methodology

Transfer learning applies the knowledge from models trained on large datasets (e.g., ImageNet) and adapts them to the butterfly classification task.

Workflow:

Enchanted Wings: Marvels of Butterfly Species

1. Data loading and preprocessing
2. Choose a base CNN model (e.g., ResNet50)
3. Freeze initial layers, fine-tune deeper layers
4. Add custom classification layers
5. Train the model on butterfly images
6. Validate and test performance

5. Model Architecture

Pre-trained Models Evaluated: ResNet50, EfficientNetB0, MobileNetV2

Custom Layers:

GlobalAveragePooling2D -> Dense(128 units, ReLU) -> Dropout(0.5) -> Dense(75 units, Softmax)

Training Configuration:

Optimizer: Adam

Loss Function: Categorical Crossentropy

Epochs: 2550 (early stopping enabled)

Batch Size: 32

6. Evaluation Metrics

- Accuracy
- Precision
- Recall
- F1 Score
- Confusion Matrix
- Training/Validation Loss and Accuracy curves

7. Application Scenarios

Scenario 1: Biodiversity Monitoring

Used in the field to identify butterfly species in real-time for habitat monitoring and species inventory.

Enchanted Wings: Marvels of Butterfly Species

Scenario 2: Ecological Research

Automated cameras track species behavior, migration, and environmental responses over time.

Scenario 3: Citizen Science & Education

Interactive mobile apps engage users in species identification and learning about butterfly ecology.

8. Tools & Technologies

- Python, TensorFlow, Keras, OpenCV, NumPy, Pandas
- Matplotlib, Seaborn, Jupyter Notebook, Google Colab
- TensorFlow Lite for mobile deployment

9. Results

Model	Validation Accuracy	Test Accuracy
ResNet50	91.2%	89.8%
EfficientNetB0	93.5%	92.1%
MobileNetV2	89.3%	88.0%

10. Challenges and Solutions

- Limited data per species: Data augmentation, transfer learning
- Class imbalance: Class weights, synthetic oversampling
- Similar-looking species: Higher-resolution input, fine-tuning
- Resource constraints: Lightweight models (MobileNetV2)

11. Future Work

- Add caterpillar and chrysalis identification
- Integrate GPS and timestamp tagging
- Expand dataset and classes
- Build real-time mobile app with offline support
- Deploy a public butterfly tracking dashboard

Enchanted Wings: Marvels of Butterfly Species

12. Conclusion

The "Enchanted Wings" project demonstrates the successful application of transfer learning to butterfly species classification. It provides a scalable, efficient, and accessible method for researchers, educators, and the public to contribute to biodiversity conservation.

13. References

- He, K. et al. (2016). Deep Residual Learning for Image Recognition.
- Tan, M., & Le, Q. (2019). EfficientNet: Rethinking Model Scaling.
- TensorFlow Documentation: https://www.tensorflow.org
- Kaggle & iNaturalist Datasets