Butterfly and Moths Image Classification 100 species

Ananya Gandhi (20319)

CONTENTS

Introduction

This project analyzes a dataset of high-resolution images featuring 100 butterfly and moth species. The objective is to use machine learning to accurately classify each image into its respective species. The challenge lies in the diverse appearances and potential misidentification of similar species.

Dataset Overview

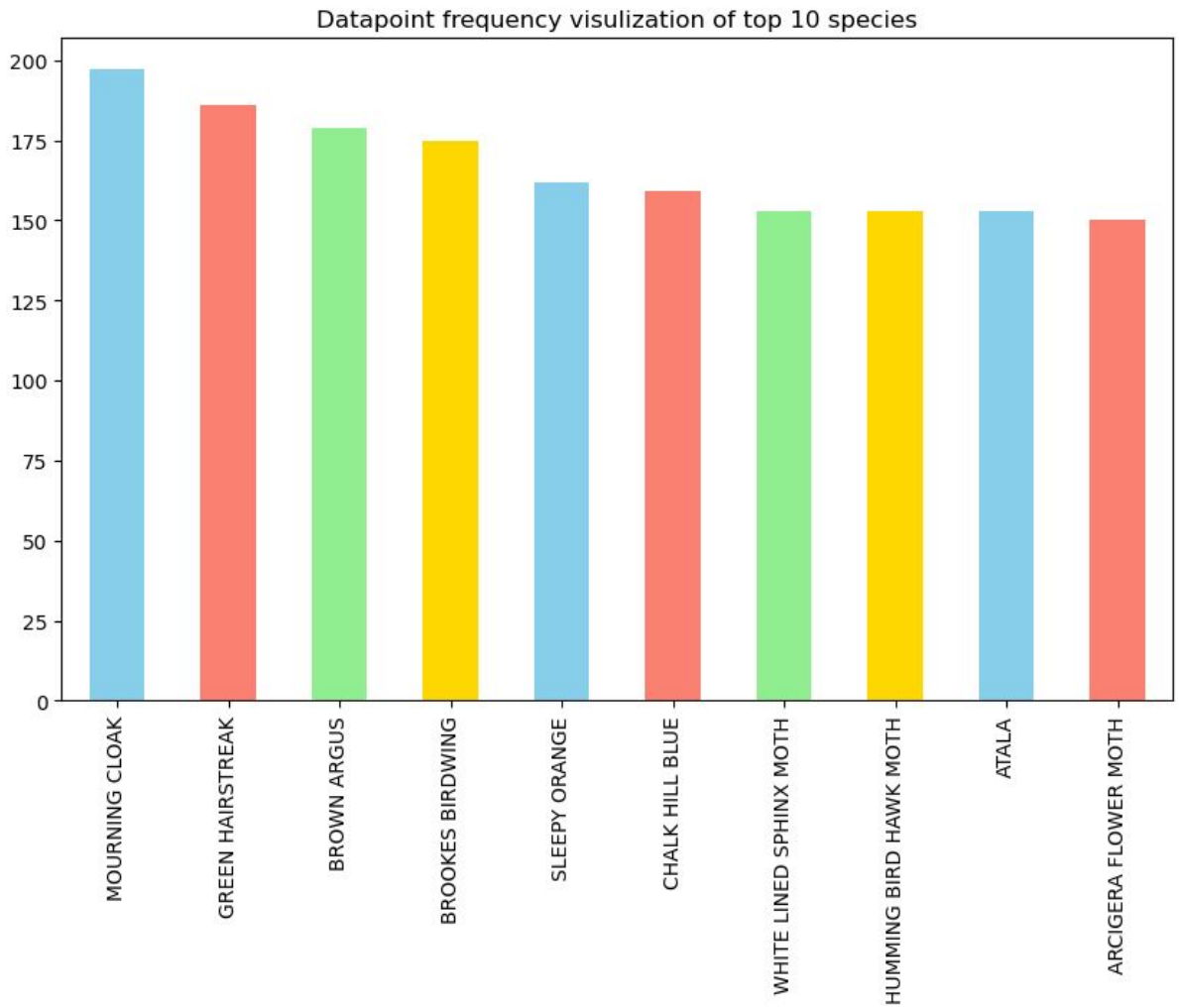
The dataset includes images of 100 butterfly or moth species in JPG format with dimensions of 224 x 224 x 3. It is divided into three parts:

- Train set: 12,594 images across 100 subdirectories.
- Test set: 500 images across 100 subdirectories with 5 test images per species.
- Validation set: 500 images across 100 subdirectories with 5 validation images per species.

Additionally, a CSV file accompanies the images, detailing class IDs, file paths, labels, and dataset categories.

Image and its Frequency Visualization

Image and its Frequency Visualization

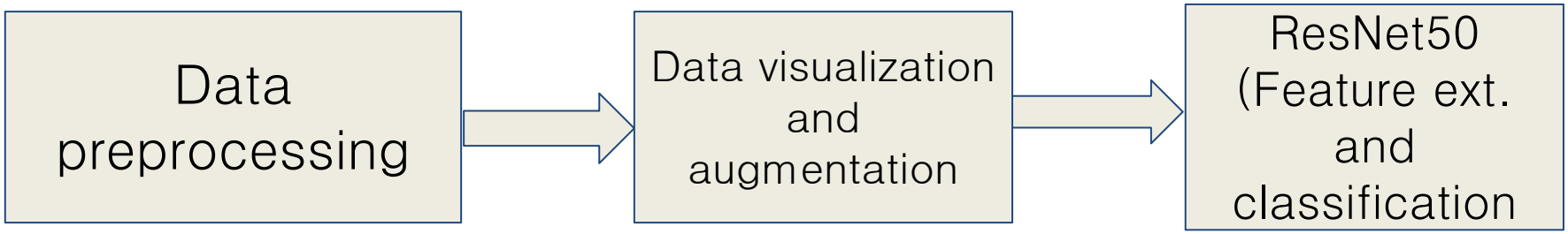


Libraries used

Utilized Python libraries for analysis:

- NumPy: Core for numerical operations, integral for efficient array processing.
- Matplotlib: Produced visualizations for exploratory data analysis, offering a rich suite of plotting tools.
- Pandas: Managed and manipulated the dataset, especially the CSV file, leveraging its robust data frame structure.
- OpenCV (cv2): Executed essential image processing tasks crucial for data preparation in model training.
- TensorFlow and Keras: Enabled seamless development, training, and validation of deep learning models. Keras provided a high-level API for TensorFlow and was instrumental in implementing the ResNet model.

Methodology



The pipeline involved:

1. Normalizing pixel values to the range [0,1] for model compatibility.
2. Augmenting the training data to improve model generalization.
3. Splitting the data into the designated training, testing, and validation sets.

ResNet-50

ResNet-50, short for Residual Network with 50 layers, is a deep convolutional neural network architecture that belongs to the ResNet family.

The key innovation in ResNet is the use of residual blocks, which contain shortcut connections (skip connections) that allow the network to learn residuals, making it easier to train deep networks.

ResNet-50 is often used for image classification tasks, including tasks with a large number of classes, as demonstrated in our butterfly and moth species classification project.

ResNet-50 can be implemented using TensorFlow and Keras, where Keras serves as a high-level API for TensorFlow, providing a user-friendly interface for building and training deep learning models.

Result

Model Performance:

The ResNet50 model achieved an accuracy of 77.80% on the validation set and 79.4% on the testing set. This high level of accuracy demonstrates the model's efficacy in classifying butterfly and moth species. It suggests that the preprocessing steps and the chosen architecture were appropriate for this task.

THANK YOU