Introduction

The purpose of this project is to classify animal species using image data from the CIFAR-100 dataset. The dataset consists of labeled images with fine-grained and coarse-grained classifications. This appendix provides a detailed description of the datasets used, including their structure, variables, and summary statistics.

The datasets are divided into three main components:

1. **Metadata:** Provides label mappings for images in the training and testing datasets.
2. **Train Dataset:** Contains labeled image data used for training the classification model.
3. **Test Dataset:** Contains labeled image data used to evaluate the classification model.

Below is a description of each individual mentioned above:

Metadata

- Each row represents a label mapping for a specific class of images.

Variable Name	Data Type	Description	Example Value
Fine Label Name	String	Specific class of an image.	Butterfly
Coarse Label Name	String	Broader category of an image.	Insects
Fine Label ID	Integer	Integer representation of the fine label.	12
Coarse Label ID	Integer	Integer representation of the coarse label.	3

Summary Statistics:

- Total Fine Labels: 100 classes (e.g., Butterfly, Lion, Train).
- Total Coarse Labels: 20 superclasses (e.g., Insects, Large Carnivores, Vehicles 1).
- Example Mappings:
    - Fine Label: 12 → Name: Butterfly
    - Coarse Label: 8 → Name: Insects

Train Dataset

- Each row represents an individual image, its pixel data, and its fine and coarse labels.

Variable Name	Data Type	Description	Example Value

Image Array	Array	RGB pixel data for the image.	[255, 0, 0, ...]
Fine Label ID	Integer	ID for the specific class of the image.	12
Coarse Label ID	Integer	ID for the broader category of the image.	3

Summary Statistics:

- **Total Images:** 50,000
- **Distribution of Fine Labels:**
    - 500 images per fine label (e.g., Butterfly, Lion, Train)
- **Distribution of Coarse Labels:**
    - 2,500 images per coarse label (e.g., Insects, Large Carnivores, Vehicles 1).

Test Dataset

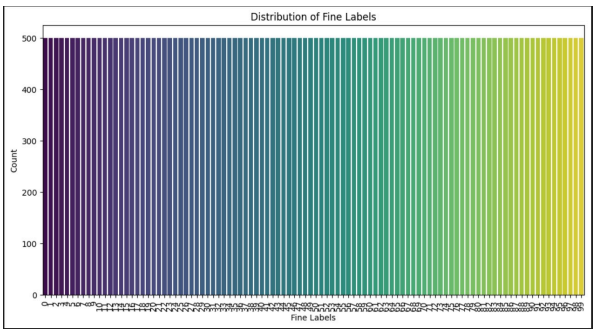
- Each row represents an individual image, its pixel data, and its fine and coarse labels.

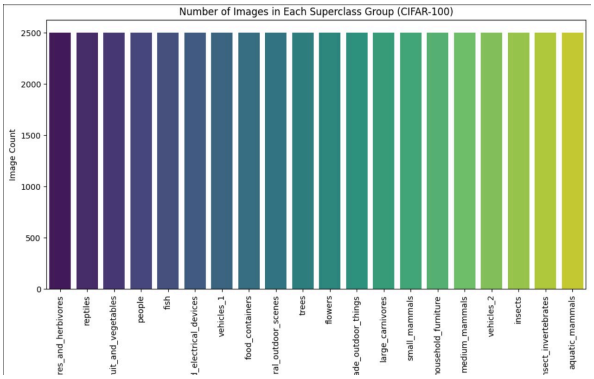
Variable Name	Data Type	Description	Example Value
Image Array	Array	RGB pixel data for the image.	[255, 0, 0, ...]
Fine Label ID	Integer	ID for the specific class of the image.	12
Coarse Label ID	Integer	ID for the broader category of the image.	3

Summary Statistics:

- **Total Images:** 10,000
- **Distribution of Fine Labels:**
    - 100 images per fine label.
- **Distribution of Coarse Labels:**
    - 500 images per coarse label.

Figures

1. **Label Distribution (Fine Labels):**
   - 

.
2. **Label Distribution (Coarse Labels):**
   - 

3. **Example image:**

Conclusion

This appendix provides a comprehensive overview of the CIFAR-100 dataset, its structure, and its role in the project. The metadata ensures proper labeling of image data, while the train and test datasets provide the foundation for model development and evaluation. Descriptive statistics and visualizations highlight key dataset characteristics, supporting reproducibility and transparency.