Instructions to Reproduce Final Image Dataset

The final dataset used to train our CNN-based Pet Breed Classifier is derived from a curated version of the Oxford-IIIT Pet Dataset, which includes 7,390 JPEG images of 35 distinct dog and cat breeds.

🔽 Step-by-Step Instructions

1. **Download Raw Images:**

   Navigate to the following link: https://www.robots.ox.ac.uk/~vgg/data/pets/

   a. Scroll to the "Downloads" section
   b. Download the images.tar.gz archive
   c. Extract the archive into your local project directory under DATA/

2. **Organize Breed Folders:**

   Use the script data_organization.py from the SCRIPTS/ folder to organize the raw images into breed-specific folders, run:

   python SCRIPTS/data_organization.py

3. **Preprocess Images for Model Input**
   To apply image augmentation and split into training and validation sets, run:

   python SCRIPTS/data_preprocessing.py

   This script will:

   - Resize all images to 224×224 pixels
   - Normalize pixel values
   - Apply random transformations (rotation, zoom, flip, brightness)
   - Reserve 20% of the dataset for validation
   - Output final structure to DATA/organized_images/

4. **Use in Colab Notebook**
   Once preprocessing is complete, upload the organized_images directory to your Google Drive and link it in Model_Training_and_Evaluation.ipynb. This notebook handles model training, evaluation, and result visualization.