Data Collection and Preprocessing Phase

Date	17 July 2025
Team ID	739816
Project Title	Galactic Gazetteer:A comprehensive Dataset of Planet Classification Images
Maximum Marks	6 Marks

Preprocessing Template

The images will be preprocessed by resizing, normalizing, augmenting, denoising, adjusting contrast, detecting edges, converting color space, cropping, batch normalizing, and whitening data. These steps will enhance data quality, promote model generalization, and improve convergence during neural network training, ensuring robust and efficient performance across various computer vision tasks.

Section	**Description**
Data Overview	The dataset used in this project consists of labeled images of planetary surfaces collected from satellite or telescope imagery. Each image represents a different type of planet or surface feature, such as rocky, gaseous, or icy textures. The images are of uniform size and format, ensuring consistency during preprocessing. Labels are provided for supervised learning, allowing the model to learn patterns and features specific to each class. The dataset is divided into training, validation, and test sets to evaluate model performance effectively. Preprocessing techniques like resizing, normalization, and augmentation are applied to improve accuracy and robustness.
Resizing	Resizing images involves changing their dimensions to a uniform size to ensure consistency across the dataset. In this project, all images are resized to a fixed resolution (e.g., 128x128 or 224x224 pixels) before being fed into the model. This helps reduce computational load, speeds up training, and ensures compatibility with convolutional neural network architectures that require fixed input sizes.

Normalization	Normalization is the process of scaling pixel values of images to a specific range, typically between 0 and 1. In this project, pixel values originally ranging from 0 to 255 are divided by 255 to normalize them. This helps the model train faster and improves convergence by maintaining numerical stability and ensuring all input features are on a similar scale.
Data Augmentation	Data augmentation in the Planet image classification project involves applying transformations to satellite images to improve model performance and generalization. Common techniques include random rotations, horizontal/vertical flips, brightness and contrast adjustments, cropping, resizing, and adding noise. These augmentations help the model learn robust features and reduce overfitting. Libraries like Albumentations and torchvision are commonly used for efficient augmentation pipelines.
Denoising	Denoising in the Planet image classification project involves applying filters to reduce noise in satellite images, enhancing image clarity and feature detection. Techniques like Gaussian blur, median filtering, and bilateral filtering are commonly used to smooth the image while preserving important edges. This preprocessing step helps improve model accuracy by removing unwanted artifacts and focusing on relevant patterns in the data.
Edge Detection	Edge detection in the Planet image classification project is used to highlight prominent boundaries and features within satellite images. Algorithms like Sobel, Canny, or Laplacian are applied to emphasize edges, helping the model better identify structures such as roads, rivers, and field boundaries. This preprocessing step enhances spatial details and can improve classification accuracy by making key features more distinguishable. .
Color Space Conversion	Color space conversion in the Planet image classification project involves transforming images from one color space to another, such as from RGB to HSV, LAB, or Grayscale. This helps emphasize specific features like intensity, hue, or texture, which may be more relevant for certain classification tasks. By analyzing images in different color spaces, the model can capture diverse and meaningful patterns, improving overall performance and robustness.
Image Cropping	Image cropping in the Planet image classification project involves trimming images to focus on regions that contain key

	objects or land features of interest. This reduces background noise and irrelevant data, allowing the model to concentrate on the most informative parts of the image. Cropping can improve classification accuracy by enhancing feature clarity and reducing computational load.
Batch Normalization	Batch normalization in the Planet image classification project is applied to the inputs of each neural network layer to stabilize and accelerate training. It normalizes the activations, reducing internal covariate shift and allowing for higher learning rates. This results in faster convergence, improved model performance, and better generalization on satellite image data.
Data Preprocessing Code Screenshots	
Loading Data	Code to load the dataset into the preferred environment (e.g., Python, R).
Resizing	Give the code snippet as an image (copy and paste the picture in this block).
Normalization	Give the code snippet as an image (copy and paste the picture in this block).
Data Augmentation	Give the code snippet as an image (copy and paste the picture in this block).
Denoising	Give the code snippet as an image (copy and paste the picture in this block).
Edge Detection	Give the code snippet as an image (copy and paste the picture in this block).
Color Space Conversion	Give the code snippet as an image (copy and paste the picture in this block).
Image Cropping	Give the code snippet as an image (copy and paste the picture in this block).

Batch Normalization	Give the code snippet as an image (copy and paste the picture in this block).