Data Collection and Preprocessing Phase

Date	18 July 2024
Name	Sourabh Sanjay Dabhade
Project Title	Greenclassify: Deep Learning-Based Approach For Vegetable Image Classification
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 consists of images of vegetables organized into three main directories: Training, Testing, and Validation. Each directory contains images from 15 different classes (types of vegetables).
Resizing	The objective is to resize all input images to a consistent target size of 224x224 pixels. For Xception and Inception, resized to 299x299 pixels. This is a crucial preprocessing step in preparing the dataset for training, testing, and validating a neural network model.
Normalization	Normalizing pixel values ensures that all image pixels are scaled to a range of [0, 1]. This is accomplished using the 'rescale' parameter in the 'ImageDataGenerator' class from Keras, which divides each pixel value by 255.
Data Augmentation	----------

Denoising	----------

Edge Detection	----------
Color Space Conversion	----------
Image Cropping	----------
Batch Normalization	----------

Data Preprocessing Code Screenshots

Loading Data	

```
import tensorflow as tf
```

```
!mkdir -p ~/.kaggle
!cp kaggle.json ~/.kaggle
```

```
!kaggle datasets download -d misrakahmed/vegetable-image-dataset
```

```
Warning: Your Kaggle API key is readable by other users on this system! To fix this, you can run 'chmod 600 /root/.kaggle/kaggle.json'
Dataset URL: https://www.kaggle.com/datasets/misrakahmed/vegetable-image-dataset
License(s): CC-BY-SA-4.0
Downloading vegetable-image-dataset.zip to /content
100% 534M/534M [00:25<00:00, 23.6MB/s]
100% 534M/534M [00:25<00:00, 21.7MB/s]
```

```
!unzip '/content/vegetable-image-dataset.zip'
```

```
Streaming output truncated to the last 5000 lines.
  inflating: Vegetable Images/train/Radish/0001.jpg
  inflating: Vegetable Images/train/Radish/0002.jpg
  inflating: Vegetable Images/train/Radish/0003.jpg
  inflating: Vegetable Images/train/Radish/0004.jpg
```

```
# Read image folders (train, test, validation)
train_path = "/content/Vegetable Images/train"
test_path = "/content/Vegetable Images/test"
validation_path = "/content/Vegetable Images/validation"
```

Normalization	`train_gen = ImageDataGenerator(rescale = 1./255,# Normalise the data shear_range = 0.2, zoom_range = 0.2, horizontal_flip = True)`
	`test_gen = ImageDataGenerator(rescale=1./255)` `val_gen = ImageDataGenerator(rescale=1./255)`

Resizing	`train_data = train_gen.flow_from_directory( train_path, target_size=(224, 224), batch_size=64, class_mode='categorical',)`
	Found 15000 images belonging to 15 classes.
	`test_data = test_gen.flow_from_directory( test_path, target_size=(224, 224), batch_size=64, class_mode='categorical', shuffle = False )`
	Found 3000 images belonging to 15 classes.
	`validation_data = val_gen.flow_from_directory( validation_path, target_size=(224, 224), batch_size=64, class_mode='categorical', )`

Data Augmentation	----------
Denoising	----------
Edge Detection	----------

Color Space Conversion	----------
Image Cropping	----------
Batch Normalization	----------