Model Optimization and Tuning Phase Template

Date	29 June 2025
Team ID	SWTID1750316859
Project Title	ASL - Alphabet Image Recognition
Maximum Marks	10 Marks

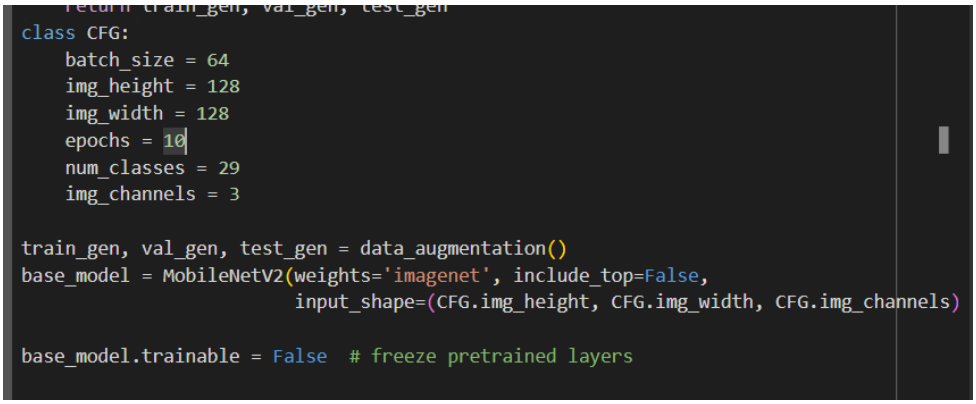
Model Optimization and Tuning Phase

The Model Optimization and Tuning Phase involves refining neural network models for peak performance. It includes optimized model code, fine-tuning hyperparameters, comparing performance metrics, and justifying the final model selection for enhanced predictive accuracy and efficiency.

Hyperparameter Tuning Documentation (8 Marks):

Model	Tuned Hyperparameters

MobileNetV2	class CFG:
	batch_size = 64 # Number of images processed before the model updates weights
	img_height = 128 # Height of each input image
	img_width = 128 # Width of each input image
	epochs = 10 # Total number of training cycles over the entire dataset
	num_classes = 29 # Total number of output classes (e.g., 26 alphabets + 3 extra signs)

	img_channels = 3 # Number of color channels (3 for RGB images)
	 
	`return train_gen, val_gen, test_gen`
	`class CFG:`
	`batch_size = 64`
	`img_height = 128`
	`img_width = 128`
	`epochs = 10`
	`num_classes = 29`
	`img_channels = 3`
	`train_gen, val_gen, test_gen = data_augmentation()`
	`base_model = MobileNetV2(weights='imagenet', include_top=False,`
	`input_shape=(CFG.img_height, CFG.img_width, CFG.img_channels)`
	`base_model.trainable = False # freeze pretrained layers`

VGG16	# Configuration for training
	class CFG:
	batch_size = 64 # How many images to process in one step
	img_height = 64 # Height of input image
	img_width = 64 # Width of input image
	epochs = 10 # Number of training rounds
	num_classes = 29 # Total ASL labels: A-Z + del + nothing + space
	img_channels = 3 # RGB image has 3 channels

```
# Configuration for training
class CFG:
    batch_size = 64           # How many images to process in one step
    img_height = 64           # Height of input image
    img_width = 64            # Width of input image
    epochs = 10               # Number of training rounds
    num_classes = 29          # Total ASL labels: A-Z + del + nothing + space
    img_channels = 3          # RGB image has 3 channels
```

```
# Set correct dimensions for VGG16
CFG.img_height = 224
CFG.img_width = 224
```

Final Model Selection Justification (2 Marks):

Final Model	Reasoning
MobileNet-V2	VGG16 is a deep convolutional network with ~138 million parameters, known for its simplicity but extremely large size and slow training. In this project, it took nearly **4 hours per epoch** and started with a very low accuracy of **0.03**, making it impractical despite being pretrained. MobileNetV2, in contrast, is a lightweight model (~3.4 million parameters) optimized for speed and efficiency. It trained much faster (within minutes per epoch) and started with a significantly better accuracy (~18%). Its use of depthwise separable convolutions and

	inverted residuals makes it ideal for tasks like ASL recognition, where quick iteration and lower resource usage are essential. Thus, VGG16 was discontinued, and MobileNetV2 was chosen for its superior performance, speed, and practicality.