CYCLE – 2

I Build a small Convolutional Neural Network (CNN) model using any of deep libraries for:

a) Image Recognition/ Classification

b) Digit Identification

Code:

a)

```
import tensorflow as tf
from tensorflow.keras import layers, models

# Load CIFAR-10 dataset
(train_images, train_labels), (test_images, test_labels) = tf.keras.datasets.cifar10.load_data()

# Normalize pixel values to be between 0 and 1
train_images, test_images = train_images / 255.0, test_images / 255.0

# Define the CNN model
model = models.Sequential([
    layers.Conv2D(32, (3, 3), activation='relu', input_shape=(32, 32, 3)),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(64, (3, 3), activation='relu'),
    layers.MaxPooling2D((2, 2)),
    layers.Conv2D(64, (3, 3), activation='relu'),
    layers.Flatten(),
    layers.Dense(64, activation='relu'),
    layers.Dense(10, activation='softmax')
])

# Compile the model
model.compile(optimizer='adam',
              loss='sparse_categorical_crossentropy',
              metrics=['accuracy'])

# Train the model
model.fit(train_images, train_labels, epochs=10, batch_size=64, validation_data=(test_images,
test_labels))
```

b)

```
import tensorflow as tf
from tensorflow.keras import layers, models
```

```
# Load MNIST dataset
mnist = tf.keras.datasets.mnist
(train_images, train_labels), (test_images, test_labels) = mnist.load_data()

# Normalize pixel values to be between 0 and 1
train_images, test_images = train_images / 255.0, test_images / 255.0

# Define the CNN model for digit identification
model = models.Sequential([
layers.Conv2D(32, (3, 3), activation='relu', input_shape=(28, 28, 1)),
layers.MaxPooling2D((2, 2)),
layers.Conv2D(64, (3, 3), activation='relu'),
layers.MaxPooling2D((2, 2)),
layers.Conv2D(64, (3, 3), activation='relu'),
layers.Flatten(),
layers.Dense(64, activation='relu'),
layers.Dense(10, activation='softmax')
])

# Compile the model
model.compile(optimizer='adam',
loss='sparse_categorical_crossentropy',
metrics=['accuracy'])

# Reshape the input data to fit the CNN input shape
train_images = train_images.reshape(train_images.shape[0], 28, 28, 1)
test_images = test_images.reshape(test_images.shape[0], 28, 28, 1)

# Train the model
model.fit(train_images, train_labels, epochs=5, batch_size=64, validation_data=(test_images,
test_labels))
```

Output:

a)

```
Epoch 6/10

782/782 [==============================] - 4s 5ms/step - loss: 0.8482 - accuracy: 0.7076 -
val_loss: 0.9099 - val_accuracy: 0.6842
Epoch 7/10
782/782 [==============================] - 4s 5ms/step - loss: 0.7968 - accuracy: 0.7231 -
val_loss: 0.8801 - val_accuracy: 0.6972
Epoch 8/10
782/782 [==============================] - 4s 6ms/step - loss: 0.7487 - accuracy: 0.7395 -
val_loss: 0.8674 - val_accuracy: 0.6994
Epoch 9/10
```

782/782 [==============================] - 4s 5ms/step - loss: 0.7155 - accuracy: 0.7491 - val_loss: 0.8860 - val_accuracy: 0.6953
Epoch 10/10
782/782 [==============================] - 4s 5ms/step - loss: 0.6783 - accuracy: 0.7652 - val_loss: 0.8507 - val_accuracy: 0.7162
<keras.src.callbacks.History at 0x7e817c7c0b50>

b)

```
Epoch 1/5 938/938 [==============================] - 7s 5ms/step - loss: 0.1821
- accuracy: 0.9449 - val_loss: 0.0657 - val_accuracy: 0.9812 Epoch 2/5 938/938
[==============================] - 5s 5ms/step - loss: 0.0527 - accuracy:
0.9838 - val_loss: 0.0462 - val_accuracy: 0.9854 Epoch 3/5 938/938
[==============================] - 5s 5ms/step - loss: 0.0392 - accuracy:
0.9878 - val_loss: 0.0293 - val_accuracy: 0.9895 Epoch 4/5 938/938
[==============================] - 5s 5ms/step - loss: 0.0287 - accuracy:
0.9906 - val_loss: 0.0390 - val_accuracy: 0.9862 Epoch 5/5 938/938
[==============================] - 5s 6ms/step - loss: 0.0237 - accuracy:
0.9924 - val_loss: 0.0313 - val_accuracy: 0.9891
```

II How to use Pre-trained CNN models for feature extraction.

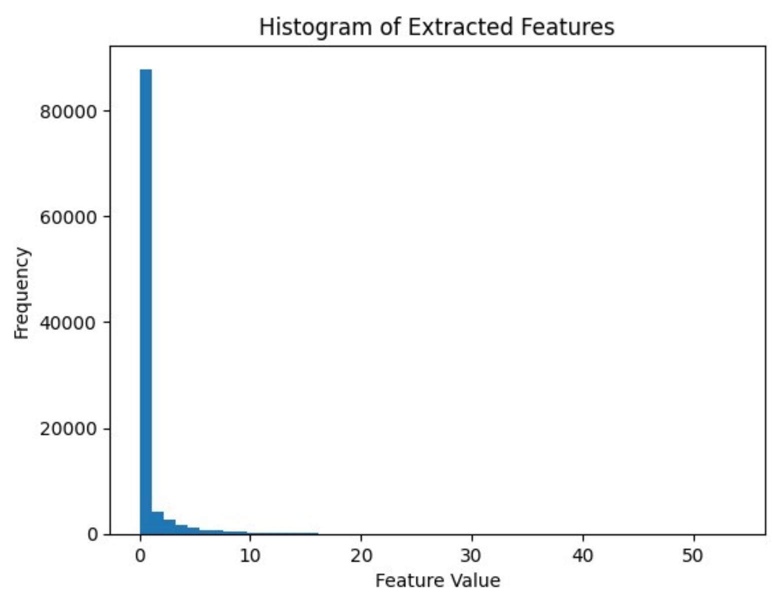
Code:

```
import numpy as np
from tensorflow.keras.applications import ResNet50
from tensorflow.keras.preprocessing import image
from tensorflow.keras.applications.resnet50 import preprocess_input
from tensorflow.keras.models import Model
import matplotlib.pyplot as plt

# Load pre-trained ResNet50 model
resnet50_model = ResNet50(weights='imagenet', include_top=False)

# Define a new model with ResNet50 base
model = Model(inputs=resnet50_model.input,
outputs=resnet50_model.get_layer('conv5_block3_out').output)

# Function to extract features from images using ResNet50
def extract_features(img_path):
img = image.load_img(img_path, target_size=(224, 224)) # ResNet50 expects images of size 224x224
x = image.img_to_array(img)
x = np.expand_dims(x, axis=0)
x = preprocess_input(x) # Preprocess the input data to align with the way ResNet50 was trained
features = model.predict(x)
```

```
return features.flatten() # Flatten the features to use them as input to another classifier

# Example usage
img_path = '/content/drive/MyDrive/cat.png'
features = extract_features(img_path)
print("Shape of extracted features:", features.shape)


# Function to plot histogram of extracted features
def plot_features_histogram(features):
plt.hist(features, bins=50)
plt.title('Histogram of Extracted Features')
plt.xlabel('Feature Value')
plt.ylabel('Frequency')
plt.show()

# Example usage
plot_features_histogram(features)
```

Output:

```
1/1 [==============================] - 2s 2s/step Shape of extracted features:
(100352,)
```



III Implementation of Pre-trained CNN models using transfer learning for classification/object detections.

a) AlexNet

b) VGG-16

Code:

a)

```
import torch
import torch.nn as nn
import torch.optim as optim
import torchvision.transforms as transforms
import torchvision.datasets as datasets
import torchvision.models as models
from torch.utils.data import DataLoader

# Set device (GPU if available, else CPU)
device = torch.device('cuda' if torch.cuda.is_available() else 'cpu')

# Define transformation for data augmentation and normalization
transform = transforms.Compose([
transforms.Resize((224, 224)),
transforms.RandomHorizontalFlip(),
transforms.ToTensor(),
transforms.Normalize(mean=[0.485, 0.456, 0.406], std=[0.229, 0.224, 0.225])
])

# Load CIFAR-10 dataset
train_dataset = datasets.CIFAR10(root='./data', train=True, download=True, transform=transform)
test_dataset = datasets.CIFAR10(root='./data', train=False, download=True, transform=transform)

# Create data loaders
train_loader = DataLoader(dataset=train_dataset, batch_size=64, shuffle=True)
test_loader = DataLoader(dataset=test_dataset, batch_size=64, shuffle=False)

# Load pre-trained AlexNet model
alexnet_model = models.alexnet(pretrained=True)

# Freeze parameters of pre-trained layers
for param in alexnet_model.parameters():
param.requires_grad = False

# Modify the last fully connected layer for CIFAR-10 classification (10 classes)
num_features = alexnet_model.classifier[6].in_features
alexnet_model.classifier[6] = nn.Linear(num_features, 10)
```

```
# Move model to device
alexnet_model = alexnet_model.to(device)

# Define loss function and optimizer
criterion = nn.CrossEntropyLoss()
optimizer = optim.Adam(alexnet_model.parameters(), lr=0.001)

# Train the model
num_epochs = 5
for epoch in range(num_epochs):
running_loss = 0.0
for i, (inputs, labels) in enumerate(train_loader):
inputs = inputs.to(device)
labels = labels.to(device)

optimizer.zero_grad()

outputs = alexnet_model(inputs)
loss = criterion(outputs, labels)
loss.backward()
optimizer.step()

running_loss += loss.item()

if (i+1) % 100 == 0:
print(f'Epoch [{epoch+1}/{num_epochs}], Step [{i+1}/{len(train_loader)}], Loss:
{running_loss/100}')
running_loss = 0.0

# Evaluate the model
alexnet_model.eval()
correct = 0
total = 0
with torch.no_grad():
for inputs, labels in test_loader:
inputs = inputs.to(device)
labels = labels.to(device)
outputs = alexnet_model(inputs)
_, predicted = torch.max(outputs.data, 1)
total += labels.size(0)
correct += (predicted == labels).sum().item()

accuracy = 100 * correct / total
print(f'Accuracy of the network on the 10000 test images: {accuracy}%')
```

b)

```
import tensorflow as tf
from tensorflow.keras.datasets import cifar10
from tensorflow.keras.applications import VGG16
from tensorflow.keras.layers import Dense, GlobalAveragePooling2D
from tensorflow.keras.models import Model
from tensorflow.keras.optimizers import Adam

# Load CIFAR-10 dataset
(x_train, y_train), (x_test, y_test) = cifar10.load_data()
x_train, x_test = x_train / 255.0, x_test / 255.0

# Load pre-trained VGG16 model (without top fully connected layers)
base_model = VGG16(weights='imagenet', include_top=False, input_shape=(32, 32, 3))

# Freeze the convolutional layers
for layer in base_model.layers:
layer.trainable = False

# Add custom classification layers
x = GlobalAveragePooling2D()(base_model.output)
x = Dense(256, activation='relu')(x)
predictions = Dense(10, activation='softmax')(x)

# Create the model
model = Model(inputs=base_model.input, outputs=predictions)

# Compile the model
model.compile(optimizer=Adam(), loss='sparse_categorical_crossentropy', metrics=['accuracy'])

# Train the model
model.fit(x_train, y_train, epochs=50, batch_size=64, validation_data=(x_test, y_test))
```

Output:

a)

```
Epoch [5/5], Step [100/782], Loss: 0.5746017411351204

Epoch [5/5], Step [200/782], Loss: 0.5881411558389664

Epoch [5/5], Step [300/782], Loss: 0.6102184489369392

Epoch [5/5], Step [400/782], Loss: 0.5950301320850849

Epoch [5/5], Step [500/782], Loss: 0.6262218597531318

Epoch [5/5], Step [600/782], Loss: 0.5968183997273445
```

```
Epoch [5/5], Step [700/782], Loss: 0.6252494990825653
Accuracy of the network on the 10000 test images: 82.13%
```

b)

```
Epoch 46/50
782/782 [==============================] - 11s 14ms/step - loss: 0.3186 -
accuracy: 0.8953 - val_loss: 1.7425 - val_accuracy: 0.5968
Epoch 47/50
782/782 [==============================] - 10s 13ms/step - loss: 0.3130 -
accuracy: 0.8953 - val_loss: 1.7340 - val_accuracy: 0.6060
Epoch 48/50
782/782 [==============================] - 10s 13ms/step - loss: 0.3026 -
accuracy: 0.8994 - val_loss: 1.7971 - val_accuracy: 0.5943
Epoch 49/50
782/782 [==============================] - 11s 14ms/step - loss: 0.2939 -
accuracy: 0.9028 - val_loss: 1.7883 - val_accuracy: 0.6012
Epoch 50/50
782/782 [==============================] - 11s 14ms/step - loss: 0.2841 -
accuracy: 0.9060 - val_loss: 1.8380 - val_accuracy: 0.6009
```

IV Practicing various strategies of fine tuning.

Code:

```
import tensorflow as tf

from tensorflow.keras.datasets import cifar10
from tensorflow.keras.applications import VGG16
from tensorflow.keras.layers import Dense, GlobalAveragePooling2D
from tensorflow.keras.models import Model
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.callbacks import ModelCheckpoint

# Load CIFAR-10 dataset
(x_train, y_train), (x_test, y_test) = cifar10.load_data()
x_train, x_test = x_train / 255.0, x_test / 255.0

# Load pre-trained VGG16 model
base_model = VGG16(weights='imagenet', include_top=False, input_shape=(32, 32, 3))
```

```
# Freeze layers except the last convolutional block
for layer in base_model.layers[:-4]:
layer.trainable = False

# Add custom classification layers
x = GlobalAveragePooling2D()(base_model.output)
x = Dense(256, activation='relu')(x)
predictions = Dense(10, activation='softmax')(x)

# Create the model
model = Model(inputs=base_model.input, outputs=predictions)

# Compile the model
model.compile(optimizer=Adam(lr=1e-4), loss='sparse_categorical_crossentropy',
metrics=['accuracy'])

# Print model summary
model.summary()

# Train the model
history = model.fit(x_train, y_train, epochs=10, batch_size=64, validation_data=(x_test, y_test))

# Evaluate the model
test_loss, test_acc = model.evaluate(x_test, y_test)
print("Test Accuracy:", test_acc)
```

Output:

```
Total params: 14848586 (56.64 MB)

Trainable params: 7213322 (27.52 MB)

Non-trainable params: 7635264 (29.13 MB)
```

```
Epoch 6/10

782/782 [==============================] - 16s 20ms/step - loss: 0.4759 -

accuracy: 0.8338 - val_loss: 0.8067 - val_accuracy: 0.7454

Epoch 7/10

782/782 [==============================] - 17s 22ms/step - loss: 0.4210 -

accuracy: 0.8527 - val_loss: 0.8570 - val_accuracy: 0.7322

Epoch 8/10

782/782 [==============================] - 16s 20ms/step - loss: 0.3660 -

accuracy: 0.8697 - val_loss: 0.9557 - val_accuracy: 0.7352
```

```
Epoch 9/10
782/782 [==============================] - 16s 21ms/step - loss: 0.3181 -
accuracy: 0.8870 - val_loss: 0.9987 - val_accuracy: 0.7391
Epoch 10/10
782/782 [==============================] - 16s 20ms/step - loss: 0.2770 -
accuracy: 0.9025 - val_loss: 1.0109 - val_accuracy: 0.7410 313/313
[==============================] - 3s 9ms/step - loss: 1.0109 - accuracy:
0.7410
```

V ImplementingGenerative Models:

a) Autoencoder for image reconstruction

b) Word Prediction using RNN

c) Image Captioning

Code:

a)

```
import numpy as np

import matplotlib.pyplot as plt
from tensorflow.keras.layers import Input, Dense, Conv2D, MaxPooling2D, UpSampling2D
from tensorflow.keras.models import Model
from tensorflow.keras.datasets import mnist

# Load MNIST dataset
(x_train, _), (x_test, _) = mnist.load_data()

# Normalize pixel values to be between 0 and 1
x_train = x_train.astype('float32') / 255.
x_test = x_test.astype('float32') / 255.
x_train = np.reshape(x_train, (len(x_train), 28, 28, 1))
x_test = np.reshape(x_test, (len(x_test), 28, 28, 1))

# Define the autoencoder model
input_img = Input(shape=(28, 28, 1))
x = Conv2D(16, (3, 3), activation='relu', padding='same')(input_img)
x = MaxPooling2D((2, 2), padding='same')(x)
x = Conv2D(8, (3, 3), activation='relu', padding='same')(x)
x = MaxPooling2D((2, 2), padding='same')(x)
x = Conv2D(8, (3, 3), activation='relu', padding='same')(x)
```

```
encoded = MaxPooling2D((2, 2), padding='same')(x)

# Decoder
x = Conv2D(8, (3, 3), activation='relu', padding='same')(encoded)
x = UpSampling2D((2, 2))(x)
x = Conv2D(8, (3, 3), activation='relu', padding='same')(x)
x = UpSampling2D((2, 2))(x)
x = Conv2D(16, (3, 3), activation='relu')(x)
x = UpSampling2D((2, 2))(x)
decoded = Conv2D(1, (3, 3), activation='sigmoid', padding='same')(x)

# Create autoencoder model
autoencoder = Model(input_img, decoded)
autoencoder.compile(optimizer='adam', loss='binary_crossentropy')

# Train the model
autoencoder.fit(x_train, x_train,
epochs=10,
batch_size=128,
shuffle=True,
validation_data=(x_test, x_test))

# Generate reconstructed images
decoded_imgs = autoencoder.predict(x_test)

# Plot some of the reconstructed images
n = 10 # Number of images to display
plt.figure(figsize=(20, 4))
for i in range(n):
# Display original images
ax = plt.subplot(2, n, i + 1)
plt.imshow(x_test[i].reshape(28, 28))
plt.gray()
ax.get_xaxis().set_visible(False)
ax.get_yaxis().set_visible(False)

# Display reconstructed images
ax = plt.subplot(2, n, i + 1 + n)
plt.imshow(decoded_imgs[i].reshape(28, 28))
plt.gray()
ax.get_xaxis().set_visible(False)
ax.get_yaxis().set_visible(False)
plt.show()
```

b)

```
import torch
import torch.nn as nn
import torch.optim as optim
from torch.utils.data import Dataset, DataLoader
import numpy as np
import random
import string
import re
from sklearn.model_selection import train_test_split

# Sample list of sentences
text_data = [
"I like to eat apples and oranges",
"Apples and oranges are delicious fruits",
"I prefer oranges over apples",
"Oranges are juicy and delicious"
]

# Preprocess and tokenize text data
def preprocess_text(text):
text = text.lower()
text = re.sub(r'\d+', '', text) # Remove numbers
text = text.translate(str.maketrans('', '', string.punctuation)) # Remove punctuation
return text.split()

# Preprocess and tokenize text data
tokens = []
for sentence in text_data:
tokens.extend(preprocess_text(sentence))

# Create sequences of input and target pairs
seq_length = 5
sequences = []
for i in range(len(tokens) - seq_length):
sequences.append((tokens[i:i+seq_length], tokens[i+seq_length]))

# Create word-to-index and index-to-word mappings
word_to_idx = {word: idx for idx, word in enumerate(set(tokens))}
idx_to_word = {idx: word for word, idx in word_to_idx.items()}
vocab_size = len(word_to_idx)

# Define a PyTorch Dataset
class TextDataset(Dataset):
def __init__(self, sequences):
self.sequences = sequences
```

```
def __len__(self):
return len(self.sequences)

def __getitem__(self, idx):
input_seq, target = self.sequences[idx]
input_idx = torch.tensor([word_to_idx[word] for word in input_seq], dtype=torch.long)
target_idx = torch.tensor(word_to_idx[target], dtype=torch.long)
return input_idx, target_idx

# Split data into train and test sets
train_data, test_data = train_test_split(sequences, test_size=0.2, random_state=42)

# Create DataLoader instances
train_loader = DataLoader(TextDataset(train_data), batch_size=64, shuffle=True)
test_loader = DataLoader(TextDataset(test_data), batch_size=64, shuffle=False)

# Define the LSTM model
class WordPredictionModel(nn.Module):
def __init__(self, vocab_size, embedding_dim, hidden_dim):
super(WordPredictionModel, self).__init__()
self.embedding = nn.Embedding(vocab_size, embedding_dim)
self.lstm = nn.LSTM(embedding_dim, hidden_dim, batch_first=True)
self.fc = nn.Linear(hidden_dim, vocab_size)

def forward(self, x):
embedded = self.embedding(x)
lstm_out, _ = self.lstm(embedded)
out = self.fc(lstm_out[:, -1, :])
return out

# Initialize the model, loss function, and optimizer
embedding_dim = 100
hidden_dim = 128
model = WordPredictionModel(vocab_size, embedding_dim, hidden_dim)
criterion = nn.CrossEntropyLoss()
optimizer = optim.Adam(model.parameters(), lr=0.001)

# Training the model
num_epochs = 10
for epoch in range(num_epochs):
running_loss = 0.0
for input_idx, target_idx in train_loader:
optimizer.zero_grad()
output = model(input_idx)
loss = criterion(output, target_idx)
loss.backward()
optimizer.step()
```

```
running_loss += loss.item()
print(f'Epoch [{epoch+1}/{num_epochs}], Loss: {running_loss / len(train_loader):.4f}')

# Function to generate next words
def generate_next_words(seed_text, next_words):
with torch.no_grad():
seed_seq = seed_text.split()
for _ in range(next_words):
input_idx = torch.tensor([word_to_idx[word] for word in seed_seq], dtype=torch.long).unsqueeze(0)
output = model(input_idx)
predicted_idx = torch.argmax(output, dim=1).item()
next_word = idx_to_word.get(predicted_idx, '<UNK>')
seed_seq.append(next_word)
return ' '.join(seed_seq)

# Generate and print next words
seed_text = "apples are delicious"
print(generate_next_words(seed_text, 5))
```

c)

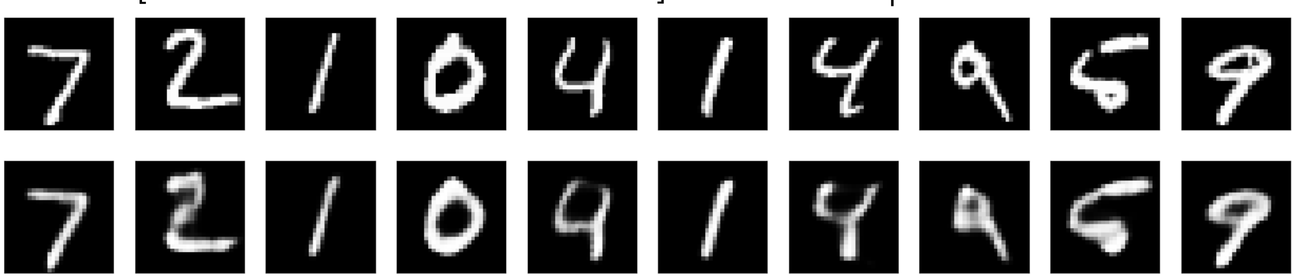
```
import matplotlib.pyplot as plt


import torch
from torchvision.transforms import transforms
from PIL import Image
from transformers import BlipProcessor, BlipForConditionalGeneration
import nltk
nltk.download('punkt')

# Load the pre-trained image captioning model
processor = BlipProcessor.from_pretrained("Salesforce/blip-image-captioning-base")
model = BlipForConditionalGeneration.from_pretrained("Salesforce/blip-image-captioning-base")

# Load and preprocess the image
image_path = "/content/cat.jfif"
image = Image.open(image_path).convert("RGB") # Convert to RGB
preprocess = transforms.Compose([
transforms.Resize((256, 256)),
transforms.ToTensor(),
])
input_tensor = preprocess(image).unsqueeze(0)

# Generate captions
with torch.no_grad():
```

```
captions = model.generate(pixel_values=input_tensor)

# Decode the generated captions
caption_text = processor.decode(captions[0], skip_special_tokens=True)

# Print the generated caption
print("Generated Caption:", caption_text)

plt.imshow(image)
plt.axis('off')
plt.show
```

Output:

a)

```
Epoch 6/10

469/469 [==============================] - 4s 8ms/step - loss: 0.1126 -

val_loss: 0.1102

Epoch 7/10

469/469 [==============================] - 3s 6ms/step - loss: 0.1108 -

val_loss: 0.1085

Epoch 8/10

469/469 [==============================] - 3s 6ms/step - loss: 0.1091 -

val_loss: 0.1070

Epoch 9/10

469/469 [==============================] - 3s 6ms/step - loss: 0.1079 -

val_loss: 0.1058

Epoch 10/10

469/469 [==============================] - 4s 8ms/step - loss: 0.1068 -

val_loss: 0.1052

313/313 [==============================] - 1s 2ms/step
```



b)

```
Epoch [6/10], Loss: 2.2166
Epoch [7/10], Loss: 2.1457
Epoch [8/10], Loss: 2.0729
Epoch [9/10], Loss: 1.9981
Epoch [10/10], Loss: 1.9212
apples are delicious fruits fruits prefer oranges apples
```

c)

```
Generated Caption: a group of cats sitting in a row
```