Ahmedabad
University

Course Name: Computer Vision

Weekly Report: 4

Group Name: XYZ

Submitted to faculty:

Mehul Raval

Date of Submission:

22 Mar 2025

Student Details

Roll No.	Name of the student	Name of the program
AU2240106	Meet Rathi	B.Tech in CSE
AU2240160	Harsh Panchal	B.Tech in CSE
AU2240153	Aditya Agarwal	B.Tech in CSE

Table of Contents.

WORK DONE THIS WEEK

1. Dataset Preparation & Preprocessing

- Processed Flickr and ICFG-PDES PRS datasets.
- The system applied unique ID numbers to each image collection under multiple caption groupings.
- An organized data storage system received cleaned data which contained filenames along with captions alongside their respective IDs.
- The application included an API for loading data images and transforming captions before tokenization.

2. Model Development

The retrieval system featured precisely the following embedding system design elements:

- Image Encoder: ResNet50 (2048-dimensional embeddings)
- Text Encoder: DistilBERT (768-dimensional embeddings)
- Projection Head: Mapped embeddings into a shared 256-dimensional space.

CLIP Model serves as a combination of image and text encoders for performing similarity analysis through them.

3. Training & Optimization

The training system achieved high efficiency through the application of optimized loss functions together with learning rate strategies and batch processing techniques.

Loss Function:

- Used Contrastive Cross-Entropy Loss with cosine similarity.
- The algorithm functions to produce high similarities between corresponding image-text pairs but assigns low similarity values to unrelated pairs.

Optimization Strategy:

- AdamW functions as the optimizer because it utilizes adaptive weight decay for stability purposes.
- The learning rate scheduler adopts ReduceLROnPlateau which automatically decreases learning rate when validation loss reaches a stable point.

- The model operates with Batch Size set at 64 after making modifications for the memory capacity of Jetson.
- An evaluation period of 3 epochs was used to find the best model before final selection.

Feature Vector Storage:

- The model processed all training and validation images by creating feature vectors with 256 dimensions.
- During inference the system can access stored embeddings through a structured data retrieval system.

WORK TO BE DONE NEXT WEEK

Code Optimization for Jetson Compatibility

- The model needs a review followed by performance optimization to run properly on Jetson Orin AGX.
- The system needs to reach both reduced latency time and better execution performance.

The deployment process of the model takes place on Jetson Orin AGX.

- Apply and test the person retrieval system on the edge hardware platform.
- Run proper tests to determine the actual operational performance while using the restricted hardware resources.

TensorRT Conversion for Inference Optimization

- The conversion of the model to TensorRT will optimize its speed when running inference operations on Jetson Orin AGX.
- Benchmark performance improvements in terms of latency and power efficiency.

Fine-Tune Model Hyperparameters

- A series of efficiency improvements can be achieved through tuning learning rate and batch size and epoch number settings.

Evaluate Performance Metrics on Jetson

- Measure speed, power consumption, and accuracy post-deployment.

Optimize Resource Scheduling

- During inference operations it is important to apply methods which optimize CPU-GPU workload interaction to prevent system performance decreases from bottlenecks.