Project Name:

Person Re-identification

Course Name: CSE641 Computer Vision
Professor Name: Prof Mehul Raval
University: Ahmedabad University

Team Members:

1. **Shlok Shelat** - AU2240025
2. **Shrey Salvi** - AU2240033
3. **Rushi Moliya** - AU2240020

Summary of Work Done This Week:

Tasks Accomplished:

Preprocessed the Flickr30k and RSTP-ReID datasets.Resized the images to 224×224 pixels.Tokenized the captions using the DistilBERT tokenizer (64 tokens for Flickr30k, 200 tokens for RSTP-ReID).

Defined the model architecture, including :-

ResNet-50 and FastViT-T8 as two alternate image encoders.- DistilBERT as the text encoder.- Combination of self-attention, cross-attention and sparse attention modules.- Specification of correct hyperparameters, such as learning rates, batch sizes, dropout rates and projection dimensions.

Areas of improvement:-

Making sure the longer captions in RSTP-ReID had token alignments.- Fine tuning the learning rates for pretrained encoders/running the risk of overfitting.

Courses of action:-

Begin training on Flickr30k using different attention mechanisms.- Compare model output performance with Recall@K, validation loss.