Task 2 Report

In this task there are two folders, server and client. In the server side the models are loaded and server is developed by FastAPI and waits for client to request for verifying two images' similarities. In the client side, the datasets are loaded and the evaluation process evaluates the models and their results are gathered and calculated and saved in a json file.

In the server/app folder there are two main files, models.py and main.p. In the models.py, two InsightFace models (buffalo_s and buffalo_l) are loaded at the server startup to avoid loading them multiple times and a verify function is developed to compute the similarity between two images by comparing face embeddings in them and returns their cosine similarity.

In the main.py the verify function is developed that waits for clients post requests and checks the request parameters and returns the images two similarity if the request and input are valid, otherwise return error to the client.

In the client folder, there other files for evaluation and data loading. Since each dataset has its own folder names and information, there three different dataloaders, one for each. Also, in the evaluate.py, the pairs are loaded and each model and dataset pair is evaluated by making 6k requests to the server and collecting and calculating the results. In addition, the results are saved after every model/dataset evaluation. The results are depicted in table 1.

Dataset	Model	Accuracy	FMR	FNMR
LFW	buffalo_l	97.9	0	0.042
CALFW	buffalo_l	54.3	0	0.147
CPLFW	buffalo_l	84.8	0	0.304
LFW	buffalo_s	97.9	0	0.042
CPLFW	buffalo_s	84.8	0	0.304
CALFW	buffalo_s	54.3	0	0.147

Table 1. Models' evaluation results on different datasets

Both models perform very well on standard verification (LFW, 97.9%), but accuracy drops on challenging datasets: cross-age (CALFW) reaches ~54%, and cross-pose (CPLFW) ~85%. FMR is zero in all cases, showing the models avoid false acceptances, while FNMR is higher on more difficult conditions.

Average inference time per pair is 0.62–0.83 seconds, with the smaller model slightly faster. The system is robust under standard conditions and conservative in matching.