Written Report

1. **1. Split the training set, a validation set, and a test set using stratified sampling to ensure that there are the same number of images per person in each set. Provide your rationale for the split ratio [5 points]**

We split the dataset into training, validation, and test sets using stratified sampling. The chosen split ratio of 70% for training, 15% for validation, and 15% for testing ensures an even distribution of images per person in each set. Stratified sampling prevents biases, promoting a balanced representation of classes across sets.

**2. Discuss the rationale behind how your team preprocess the data. Include the mathematical equations used and any dimensionality reduction applied to the instanced and discuss its relevance to the problem at hand. *Note that your team will receive more points if you perform data preprocessing that help improve the eventual training process.* [20 points]**

**Rationale:**

* **Normalization:** Pixel values were normalized to the range [0, 1] using Xnormalized=X/255 to facilitate faster convergence during training.
* **Linear Discriminant Analysis (LDA):** Implemented LDA for dimensionality reduction, transforming the data to enhance class separability.

**3. Select a clustering technique taught in this course and apply it on the training instances. Provide the rationale behind your team’s choice of clustering technique and how your team tuned the parameters for the technique implemented. [30 points]**

**Rationale:**

* **DBSCAN (Density-Based Spatial Clustering of Applications with Noise):**
  + Chose DBSCAN for its ability to identify clusters of arbitrary shapes and adaptability to varying cluster densities.
  + Tuned hyperparameters (eps and min\_samples) through an iterative process, optimizing silhouette scores.

**4.** **Clearly discuss - with illustrations - the architecture your team has selected for training and predicting the test instances. Discuss the rationale behind your team’s choice of activation functions, loss function, and how you tuned the hyperparameters of the network model. [30 points]**

**Rationale:**

* **Architecture:** A feedforward neural network with an input layer, two hidden layers using ReLU activation, and an output layer with softmax activation was selected for its simplicity and effectiveness in image classification.
* **Activation Function:** ReLU was chosen for promoting non-linearity and faster convergence.
* **Loss Function:** Sparse categorical crossentropy, suitable for multi-class classification.

1. **Discuss the results of the trained system. [5 points]**

The trained system's performance was evaluated using accuracy, precision, recall, and the confusion matrix. These metrics provide a comprehensive overview of the model's ability to classify and distinguish between individuals.

1. **Present your system. Discuss the decisions your team made, the challenges your team encountered, how your team resolved the problems, and the results. [10 points]**

**Decisions Made:**

* Detailed preprocessing decisions, including normalization and dimensionality reduction.
* Explained the rationale behind the choice of clustering technique and its hyperparameters.
* Discussed the neural network architecture, activation functions, and loss functions.

**Challenges Encountered:**

* Challenges were faced in parameter tuning and model selection.
* Resolved challenges through an iterative process of experimentation.

**Results:**

* Presented the final results, including model performance metrics.
* Reflected on the success of the project and potential areas for improvement in future iterations.

Screen shots of the code

A screenshot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A computer screen shot of a computer program

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer screen

Description automatically generated

A screenshot of a computer

Description automatically generated

GRAPHS:

A graph with different colored lines

Description automatically generated

A graph of blue lines

Description automatically generated with medium confidence

A screen shot of a chart

Description automatically generated

A screen shot of a chart

Description automatically generated