Project Report: The Rise of Artificial Intelligence & Facial Recognition

Team 53 – Ayush Pagaria, Jamelia Gordon, Xinyi Zhou, Zixiao Xu

1. Motivation & Data Understanding

A secure school setting fosters a sense of trust and well-being, allowing students to focus on their studies without the distraction or fear of potential threats. When students feel safe, they are more likely to engage in classroom activities, form positive relationships with peers and educators, and explore their intellectual curiosity. It is the foundation upon which a nurturing and thriving learning environment is built, and influences children in ways that shape the rest of their lives. However, according to the Annual Schooling in America Survey, 77% of American school parents are concerned that a violent intruder could enter child's school. Additionally, for guns, only 47% of public district school parents reported that their school handled it "very" or "extremely" well. In perspective, when it came to reasons for parents choosing their child's school "a safe environment" was the top response, while test scores came in 13th out of 14 options, underscoring the significance of a safe learning environment.

The rise of Artificial Intelligence and Facial Recognition has provided many new opportunities in the Security Industry. Therefore, to make school campuses and buildings safer, facial recognition may be adopted to identify and admit students, faculty and guests onto the campus and thereby restricting potential trespassers to enter the campus. While our focus is on enhancing security and safety on campus, it can also be implemented to take attendance and also apply other machine-learning applications such as mood-detection to identify and assess the mood of students to improve the associated environment and the learning experience.

To prepare our model, we downloaded a dataset that contained over 200,000 images of 10,177 celebrities from The CelebFaces Attributes Dataset (CelebA Dataset) from Papers with Code. The images come with different binary facial attributes as labels. For the scope of our project, we are only selecting 1,000 celebrities to conduct our analysis and model training.

2. Data Preparation

To prepare our data for further modeling, we extracted the images and the labels from the dataset. Due to the size of our dataset, we restricted the length of the dataset to contain 486 different identities across a total number of 7778 images.

Procedure:

Create Metadata: As we had different datasets for images and corresponding labels, we created a Metadata list using the annotations from the labels text file to uniquely match each image path from the image dataset to the corresponding label. The length of the Metadata was 65366 after the merging of the datasets.

Multiple Images of Common Identities: For further analysis, we then checked for identities with more than 14 images within our dataset, and it resulted in 486 such identities.

Transformed, Normalized and Center Cropped Images: Based on the Metadata, we transformed each of the images and normalized the pixel values of the images within a standard range using mean values of [0.485,0.456,0.406] with standard deviations of [0.229,0.224,0.225]. Additionally, we used central cropping to cut out a central portion of the image, to ensure all images are of the same size and are focused on the central part of the image.

Generated Facial embeddings using the VGG Face model

VGG Face:

The VGG Face Model adapts the 16-layer VGG 16 Model, known for image classification, for facial recognition. It's pre-trained on a vast face dataset to specialize in facial feature detection. In our application, we employed the VGG16 architecture with VGG Face weights but discarded the last SoftMax layer. This created a Descriptor Model that produces facial embeddings, enabling functions like face matching and verification critical for our use case.

Splitting our Data into Training, Testing and Validation Datasets

We then split our dataset into training, testing and validation datasets to check the effectiveness of the model generated. The training set is used to let the modern learn patterns within the data and once the model is applied on the testing set, it will generate new results on data that has not been trained yet. Additionally, the validation set will be used to check the efficacy of the parameters within the model. We decided to split the model using a test size of 0.1, where the testing data contains 10% of the dataset and the remaining data is split between training and validation in a ratio of 90% to 10% respectively. Using the images within our subset, we then loaded these images into data loaders with their corresponding labels that were encoded to fit within the index of (0, 485).

3. Modeling

Starting the modeling process, we established a baseline model that configures the base line model to logical return to the neural network. Subsequently, we built a more complex neural network, the SoftMax classifier model, which aims to enhance performance through its deeper architecture and other functions such as batch processing standardization and dropout.

Baseline model: Logistic regression

A logistic regression model is defined as our baseline model and we build the neural network where a fully connected linear layer transforms the input to 1064 dimensions. A ReLU (Rectified Linear Unit) is used as activation function, and another fully connected linear layer that reduces the dimensionality from 1064 to the number of classes specified by output. Then, the SoftMax function contributes to the probability output.

The loss function is defined as cross-entropy loss. It assigns a higher penalty to predictions and provides better gradient performances. If the model outputs a high probability for the incorrect class, the loss will be significantly higher, pushing the model to correct errors. A stochastic gradient descent optimizer is defined to update the model's parameters. The learning rate is set to 0.001, which can be used to minimize the cross-entropy loss. To stabilize the model learning process and better identify the complex facial recognition patterns, 50 epochs are used in the training process. For the baseline model, the average training loss decreased from 0.0235 to 0.0230. The test accuracy averaged .26% which we believed was a feature of having too little data points for each identity (approximately 15).

Optimization: Classifier model

Our SoftMax Classifier model, an advanced neural network, surpasses our initial baseline in complexity. This model is built with three fully connected layers. The first two layers are enhanced with batch normalization, an activation function, and dropout techniques to regularize and prevent the model from overfitting. Specifically, batch normalization is set with a momentum of 0.1, and dropout rates are 0.3 and 0.2 for the first and second layers, respectively. The initial

layer processes 2622 input features down to 100, while the second layer further narrows them to 10. The final layer then uses these 10 features to determine the likelihood of the 486 possible output classes through a SoftMax function. We trained the model for 200 epochs with a 0.001 learning rate to minimize cross-entropy loss, and our classifier model's accuracy was disappointingly similar to the baseline model's accuracy of .24%. This further confirmed our initial hypothesis that each identity had too few data points to effectively train the dataset.

4. Implementation

We faced significant challenges in downloading the entirety of the CelebA dataset. Consequently, we opted to work with a subset of the dataset, containing approximately 7,778 images spanning 486 identities on average, equating to around 15 images per identity. It became apparent that this multi-class problem lacked sufficient data points for an effective model training (refer to Appendix 3FinalProject.ipynb). To address this limitation, we introduced the Pins Face Recognition Dataset, which comprised 17,534 faces across 105 identities. For our model, we selected a subset of 18 identities for training and testing purposes.

The Pins Dataset

We followed a similar data preparation approach to transform the Pins dataset, which involved:

- Transforming each image by normalizing it using mean values of [0.485, 0.456, 0.406] and standard deviations of [0.229, 0.224, 0.225].
- Applying center cropping to achieve a dimension of (224, 224).
- Loading these processed images into a dataset using PyTorch's built-in ImageFolder class.

We then generated facial embeddings for each image, using the VGG face descriptor described above, and split the dataset into training, testing, and validation sets. Finally, we loaded these sets into data loaders for further processing.

Modelling

Using the same baseline model described above, the test accuracy impressively rose to 93.38% with a holdout sample accuracy of 86.10%

We made slight alterations to the SoftMax Classifier Model described above, including:

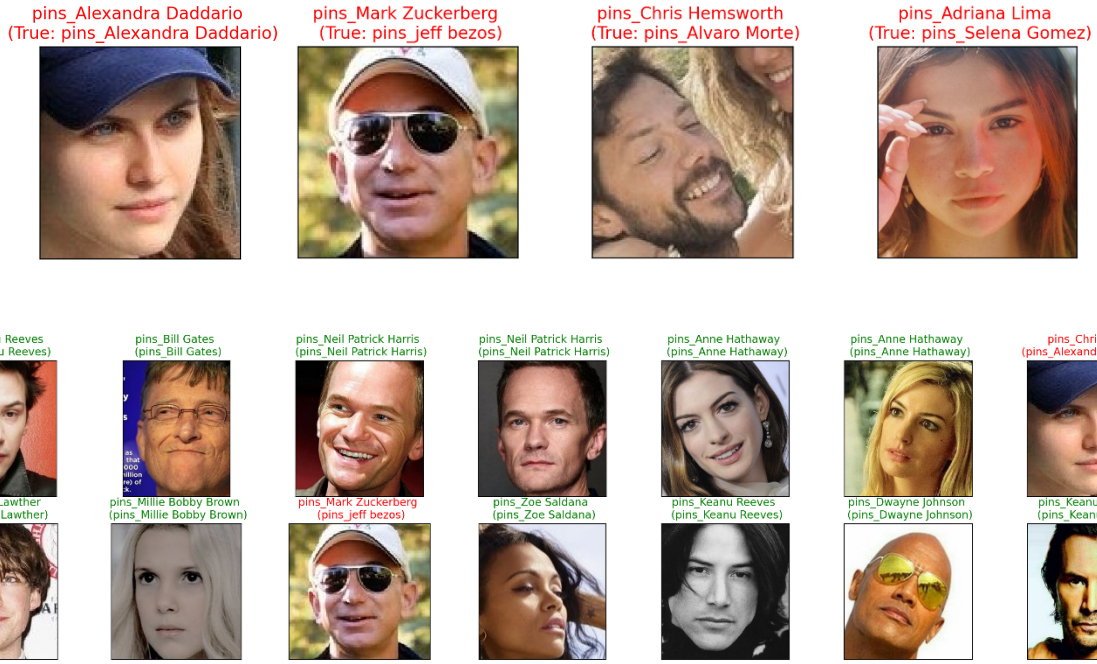
- Using three hidden layers, the system processes the initial 2622 embeddings, progressively reducing them to 1024, then further narrowing down to 512. The third layer further refines the features to 256, which are then employed to assess the likelihood of the 18 potential output classes using a SoftMax function.
- The first two layers are enhanced with batch normalization, ReLU activation functions, and dropout techniques to regularize and prevent the model from overfitting. Specifically, dropout rates are set to 0.3, 0.2 and 0.2 for the first, second and third layers, respectively.
- The training process involves 200 epochs with a learning rate of 0.001, aiming to minimize cross-entropy loss. Adam optimization is employed to update the weights after backpropagation.

The model resulted in a 92.68% test accuracy and a holdout sample accuracy of 91.89%.

5. Results and Evaluation

False Positives

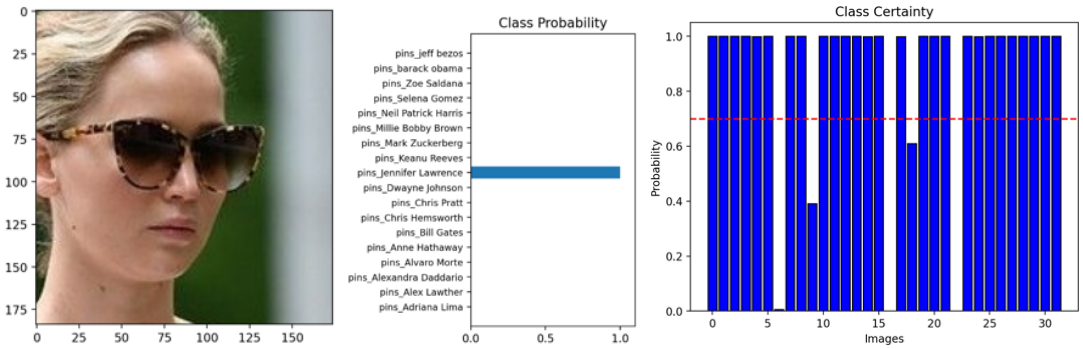
In evaluating the models, we placed significant emphasis on identifying false positives as false positives are more detrimental to this particular use scenario as they could potentially allow unauthorized individuals to enter secured spaces. In the validation dataset, we observed that out of the 259 images, there were four instances of false positives. In each of these instances, there were mitigating factors such as an object partially shielding the person's face, the individual wearing glasses, variations in lighting causing facial distortion, or the presence of another individual in the background. Addressing and minimizing false positives is critical to ensuring the reliability and effectiveness of the model in security applications.



Probability

When predicting classes, the model consistently shows high confidence, with probabilities often close to 100%. For instance, in recognizing Jennifer Lawrence, the model achieved a 99.9974% probability, even with the challenge of her wearing glasses. This underscores the robustness of

the model, indicating its ability to handle additional classes or identities and perform well with fewer data points per identity.



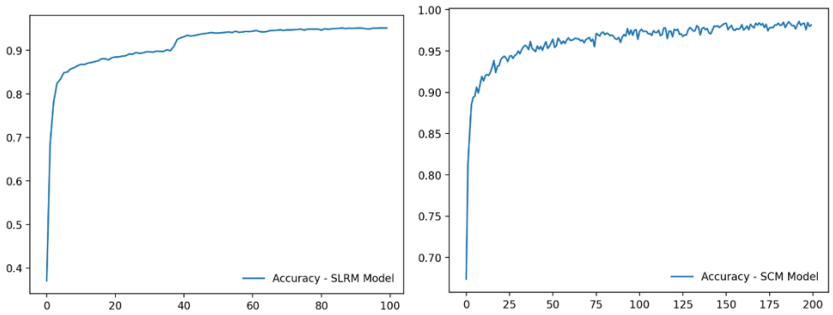
In the chart above most probabilities for the correct class in a batch are remarkably high. Notably, only 5 images out of 32 fell below a 70% threshold.

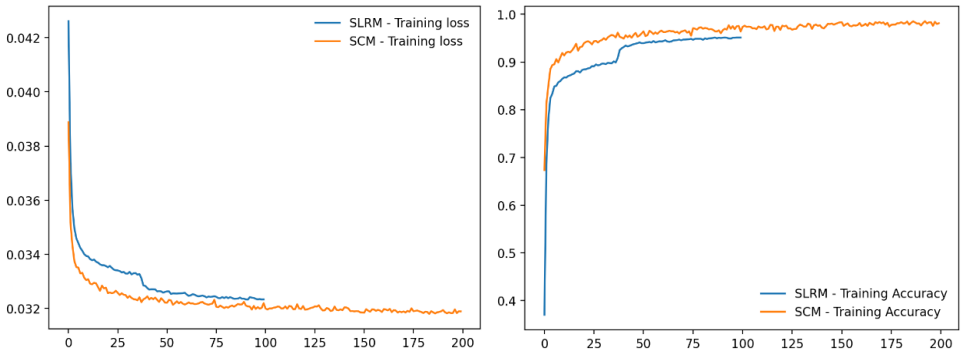
Accuracy

As mentioned earlier, both the Simple Logistic Regression Model (SLRM) and the Softmax Classifier Model (SCM) achieved high test accuracy, with SCM showing slightly superior performance. However, the exceptionally high accuracy raises questions about potential overcomplexity relative to the dataset. That is the dataset has a significant number of datapoints for each identity, yet the models are currently trained to recognize only 18 identities. This suggests the models could accommodate more identities or reduce the number of data points per identity, while maintaining a significant accuracy rate.

The oscillations in the loss function graph suggest that the model might be overly complex relative to the available data. It may attempt to fit noise in the training set, resulting in these oscillations. To address this, reconsidering model complexity, incorporating additional

regularization strategies, or increasing the number of classes could be necessary to achieve optimal performance.





6. Deployment

The Classifier Model designed to identify facial features based on images and labels will be deployed as a facial recognition model at educational institutions. It is aimed at enhancing security and safety on campuses of schools and other educational institutes, to prevent violent intruders from entering the campus and ensure that the school premises are safe and secure.

The Classifier Model tested on images resulted in an accuracy rate of 92.68% which signifies that the model can accurately identify students/faculty/guests within the database of images that are permitted within the school campus. It essentially means that it will be able to filter potential threats and alarm the school officials to prevent any violent intrusion or a similar situation.

However, there are a couple of issues in the deployment of the model. Firstly, the schools would need a large dataset to train the model and enhance its accuracy in successfully identifying students/faculty. Additionally, the school would have to label everyone to each person and any new addition would require further labeling and the model may have to be trained again, which may lead to an incremental effort and time consumption.

The model may be impractical at educational institutions that offer short-term programs, ideally 10-12 months, as they would have to change the images in short-time periods which may be so efficient.

Another roadblock would be that as children grow, especially between primary school to middle school, their facial features also change, and the model may not be able to accurately capture the changing facial features and would not correctly permit students to enter the premises. The model would have to be trained again with new images of the children as the grow. In terms of ethical considerations, it would be prudent to ensure that the images of the children/faculty are not shared outside of the implementation of the model due to privacy/security concerns.

We would try to mitigate the risks by ensuring that the model does not breach any privacy concerns and the images stored are encrypted and secured. There may be risks associated with the model not identifying students/faculty, and therefore, we would try to improve the accuracy rate and to also factor minor changes in features to improve the model.