Real Time Image Processing And Recognition

Maimouna Abderamane Tahir
mabderam@uncc.edu
ID 801350431

Yousef Al Hanbali
yalhanba@uncc.edu
ID 801284674

Ismatul Chowdhury
ichowdhu@uncc.edu
ID 801328679

Theodore Hunt
mjhunt@uncc.edu
ID 800116963

I. Introduction and Motivation

As technology advances, the need for security becomes even more critical. Technological advancements bring about numerous benefits and opportunities, but they also introduce new challenges and risks. There is a growing need to protect personal information like bank account details, website logins, streaming camera feeds, and much more. One such method of increased security includes facial recognition. Facial recognition can be a mean of security and preservation of privacy and personal information.

Facial identification methods have been recognized as one of the most secured ways of keeping personal data safe in electronics. For instance, most smart phones and laptops nowadays use face ID for authentication of identity to access cell contain. Incorporating facial recognition in our daily life would also make our life much safer and help the law enforcement sector to do their job more effectively as well. For example, with facial identification factor embedded in street cams would easily help the police keep criminals off of the road.

Additionally, certain leasing agencies and home owners these days use a smart system to lock and access their properties and housing. An example of such system is an app called SmartRent, which could be hacked or accessed by anyone who manages to put their hands on our cellphones or have a look on the app homepage, which contains the passcode. And that would hinder our safety in the sole places that are supposed to be our safe heavens. Therefore, the delicacy of our safety, privacy and chance of secure living require the incorporation of a higher encryption such as bio-metric authentication in such crucial systems. Hence, the objective of this project was to build classical and modern machine learning models to implement real time face recognition through image classification.

II. Approach

This project has 2 parts consisting of 3 respective approaches. The first part implies an implementation of both classical and modern ML models for facial recognition using the same dataset split into 80% training set and 20% validation set and a comparison between their performance evaluation outcomes. The classical ML model in this project consisted of PCA feature extraction that was passed through an SVM classifier. As for the modern ML modern, it was built based on residual network (ResNet) 10 architecture.

The second part involves a combination of the 4 project members' pictures to create a custom dataset that is fed into a model that was trained to differentiate between their specific identities and recognize them. This real time model execution combines transformer feature extraction with ResNet 18 classifier that perform in real time.

A. *PCA with SVM*

PCA was introduced to reduce the dimensionality of the dataset while keeping important information. This is accomplished by transforming the data into a lower-dimensional space by identifying principal components that capture the maximum variance in the dataset. SVM is utilized as a classification algorithm. It works by finding an optimal hyperplane that best separates different classes within the reduced-dimensional dataset obtained from PCA.

The combination of PCA for dimensionality reduction and SVM for classification constitutes a classical approach in machine learning. PCA reduces the feature space, making it computationally efficient, while SVM serves as a robust classifier capable of handling high-dimensional data effectively. This approach is widely used in various classification tasks, especially when dealing with large datasets or when computational resources are limited.

B. *ResNet10*

ResNet-10 is a smaller variant of the ResNet (Residual Network) architecture. ResNet10 are popular deep neural network architectures known for their use of residual connections, which help mitigate the vanishing gradient problem and enable the training of very deep networks.

C. *Real Time Face Verification with ResNet18 and Transformer*

For effective face recognition, a diverse dataset covering various lighting conditions, facial expressions, and orientations was collected among the group members, especially that the members are the prime individuals the model needs to recognize, and the labels of e (Yousef, Ted, Mai, Isma). Consistent pre-processing steps such as face alignment, resizing, and normalization standardize input images. Choosing a robust pre-trained model like ResNet for feature extraction and fine-tuning it on your dataset can significantly enhance its ability to recognize specific faces. Implementing continuous learning enables the model to adapt to changes in appearance over time.

Regarding the real time model, the group ensured that the model processed efficiently the face recognition in real-time, and recorded a video that was able to handle real time feed variations fairly effectively. Dynamic thresholding used for the real time approach allowed for flexible recognition sensitivity, and regular evaluation with accuracy metrics ensuring reliable performance of the model. An intuitive user interface and a feedback mechanism would be further developed into the model to achieve continuous model improvement.

III. Dataset and Training Setup

The initial data-set chosen was the Celeb-Faces Attributes (CelebA), which was stated in the project proposal and accessed from https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html , and it consists of 202,599 celebrity images with 40 large-scale face attributes and 10,177 identities.

Due to the size of the CelebA dataset, it proved difficult to utilize so the Labeled Faces in the Wild (LFW) dataset was chosen it its place for the presentation. The LFW dataset was created and maintained by researchers at the University of Massachusetts; it contains 13,233 images of 5,749 people.

Regarding the real time face recognition model, the major part of the dataset was from the image that was captured and made into a directory of the project group members and set as the gradient as a weight factor to determine the group members. the training setup went into using CNN ResNet18 and transformation for the training model which lead into getting the faster pre-processing state for the model and regular monitor for overfitting and underfitting. The group implemented a batch size of 16 due to the current capacity of the hardware specification that the group currently in position.

But for the classical and modern ML models, there was a dire need of a larger dataset was sought after, so the FaceScrub dataset with 100,000 face Images of 530 people was chosen. This dataset contains over 106,863 facial images of 530 people (265 female and 265). The dataset is released under a creative commons license from https://vintage.winklerbros.net/facescrub.html and is only available for use after filling out a form.

However, similar to the CelebA, the FaceScrub data-set proved difficult to manage, even with pre-processed subsets of the FaceScrub dataset, it was still very computationally consuming and not manageable given the group's limited time before before the final project report. Despite the numerous struggles encountered, the group had to improvise and build a different image classification project in both classical and modern ML, due to the time constraint. Thus, the CIFAR-100 dataset was chosen for the final deliverable since it could be more manageable as it only has 100 classes and allow the group to present a workable image dataset and better chance for a fair grade given the team's efforts towards the fulfillment of the project requirements.

IV. Results and Analysis

A. *PCA with SVM*

Metrics for the classical model running the CIFAR-100 dataset can be seen from Figure 1 and Figure 2 while Figure 3 contains the confustion matrix. It can be seen from the recall and accuracy that the model was balanced.

```
Precision: 0.236414
F1 score: 0.236414
Recall: 0.244500
Model Accuracy: 0.244500
```

Fig. 1. PCA SVM Metrics

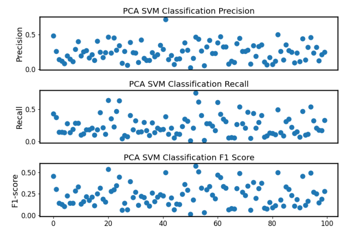

Fig. 2. PCA SVM Metrics Plots

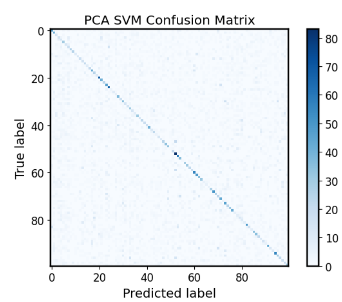

Fig. 3. PCA SVM Confusion Matrix

B. *ResNet10 Classifier*

The ResNet-10 architecture contains 10 resblock layers, including convolutional layers, pooling layers, and fully connected layers. Metrics for modern model using the CIFAR-100 dataset can be seen from Figure 4 and Figure 5 while Figure 6 displays the confustion matrix. Similar numbers between recall and accuracy once again display a balanced model.

```
Precision: 0.587021
F1 score: 0.587021
Recall: 0.588900
Model Accuracy: 0.588900
```

Fig. 4. ResNet 10 Metrics

C. *Real Time Face Verification with ResNet 18 and Transformer*

Final Project Real Time Face Recognition Video:
https://youtu.be/J5QdLOLGOco

The segment for the result was to understand how the model worked and from that, the constant learning from real time to real time implementation the model to understand that the faces for the individual based on skin color and

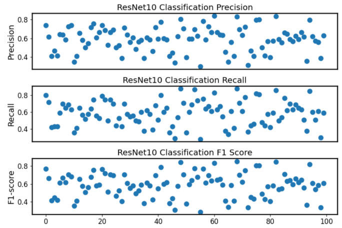

Fig. 5. ResNet 10 Metrics Plots

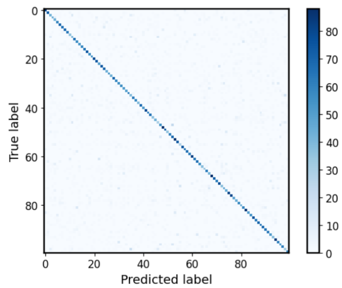

Fig. 6. ResNet 10 Confustion Matrix

facial expression and even the angle of the state that the face occurred. The resulted in the video was mainly on the fact that the face of 'Yousef' had a greater weight than the rest of the members which is an easy fix for the long run to mitigate the error and work on enhancing the model.

D. PCA SVM and ResNet 10 Models Performance Comparison

Table 1 displays the metric performance for both the classical and modern models. The modern outperformed the classical for all major categories using the same dataset. Both models appeared to be well-balanced.

V. Lesson Learned

1. The project is approach two part : the first part involves implementing classical (PCA with SVM) and modern (ResNet-10) ML models for facial recognition using different datasets, while the second part focuses on a real-time face verification system utilizing a custom dataset of the project members' pictures. The results showcase the performance metrics of implemented models, highlighting the superiority of modern models over classical ones in the same dataset. Despite challenges in dataset acquisition and pre-processing within tight time-frames, the project concludes with a real-time face recognition model displaying promising testing performance and reasonable precision and recall rates. The report reflects on the complexities of facial recognition projects, emphasizing the difficulties in obtaining and processing suitable datasets within project constraints. It acknowledges the collaborative efforts of the team members and expresses gratitude to mentors and contributors for their support.

Model	Classical	Modern
Accuracy	0.245	0.587021
Precision	0.2467	0.587021
Recall	0.2445	0.5889
F1 Score	0.2364	0.5889

TABLE I
Classical vs Modern ML models

2. New topics have been introduced in the real time model the group was able to develop using the concept of transformer, which is an upgrade to the traditional deep learning training methods, and how to utilize it in this project. Although the project ran on a small dataset for the real time face identification, the results were surprisingly accurate. Basis is a major factor into this equation as demonstrated in the video. The reason that the model was showing the name of 'Yousef' was due to the larger amount of the 'Yousef' sub-dataset as compared to 'Ted', 'Isma' and 'Mai' sub-datasets.

3. We learned that face recognition is a complex topic and conducting such a project requires much simpler dataset given our time constraint. However, the importance of the matter makes it worth exploring the context of classifying facial datasets. The main problem we faced was the fact that there is no public access to an image dataset that could have been more easily manageable and whose pre-processing would have been less time consuming. Because the prime dilemma in the completion of our project was the time and computational process required to get the available datasets to pre-processed and loaded into our code. The closest we came to succeeding in building facial recognition models was with the LFW dataset which was already pre-processed and the faceScrub dataset which took days even to be downsized to a subset comprising of 200 images in training size, 50 in validation size and 50 in test size. The code for the successful pre-processing of the faceScrub dataset is included in the appendix.

Contributions

Yousef Al Hanbali initially started the real time model and after the group was formed, he worked together with Maimouna Abderamane Tahir, Ismatul Jannat Chowdhury and Theodore Hunt to bring the work to fruition. The group members also contributed fairly equally in researching and pre-processing the datasets used in the assignment. Moreover, everyone worked collaboratively to build a classical and modern ML models to comply with the requirements of Dr. Tabkhi regarding the final project deliverables because the real time face recognition doesn't have model performance metrics as compared regular machine learning image classification.

Acknowledgment

We would like to extend our extreme gratitude to Dr. Tabkhi for his unwavering support and teaching throughout the semester. We are also utterly grateful to Fatema Jannat and Badrul Hassan Badhon for their assistance in the accomplishment of the course assignment during the term.

Appendix

Real Time Face Recognition GitHub Respository: https://github.com/YousefALH/MLHWCore/tree/main/Project%20Real%20Time
https://github.com/hurricane195/Intro-to-Machine-Learning/tree/Final-Project