Data Analytics Project:
Facial Recognition Using Siamese Network

By –
Sameer Hussain, Harishankar Sekhar,
Ravi Teja Seera, & Aditya Gaitonde

Introduction

The project delves into the dynamic and challenging world of image recognition, a cornerstone technology in diverse fields such as security, health diagnostics, and digital interfaces. This project critically examines the inherent limitations in existing deep learning approaches, particularly their dependency on extensive and varied datasets and significant computational demands. In response, the project pioneers the development of a model based on Siamese networks, which is ingeniously designed to perform accurate and efficient image recognition with substantially less training data. This innovative solution is not just a technical advancement but also a strategic response to real-world scenarios where data availability is limited and where rapid, reliable image analysis is essential. The project, therefore, stands at the forefront of technological innovation, offering a promising solution to longstanding challenges and opening new avenues for application in various industries.

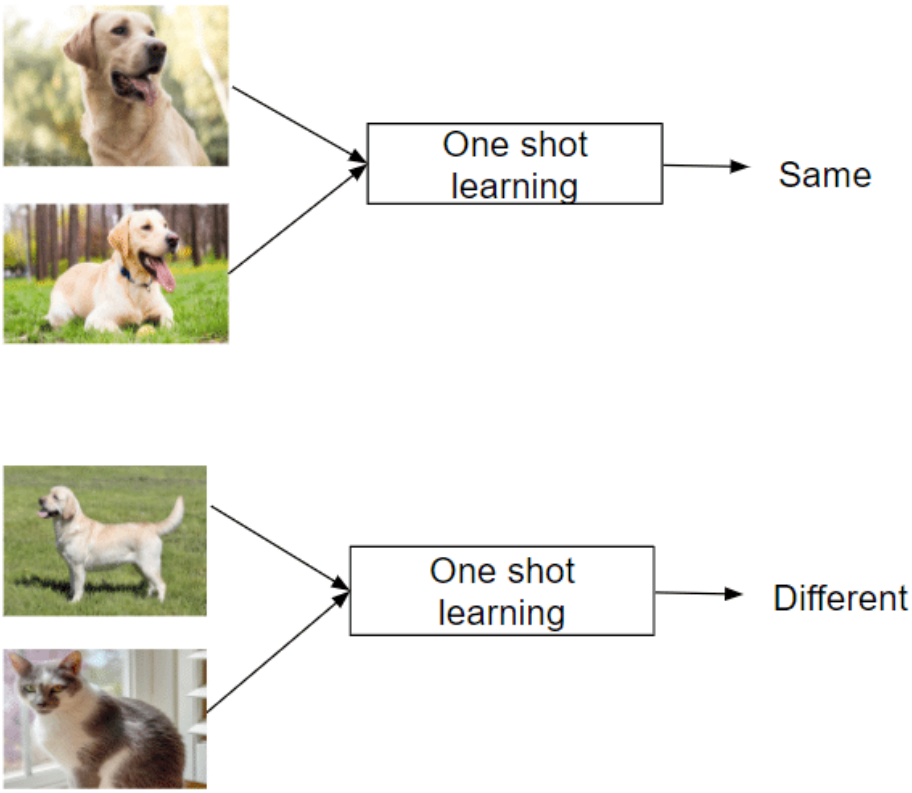

Figure 1: One-shot learning. Source Google images

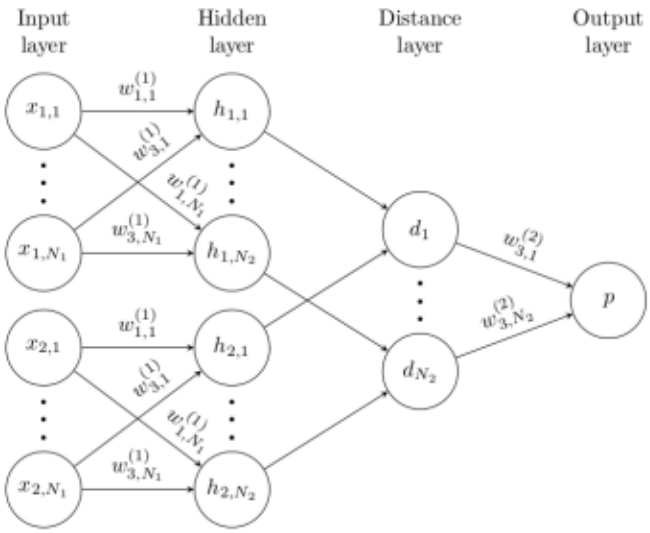

Figure 2: A simple 2 hidden layer Siamese network for binary classification with logistic prediction. Source: Siamese Neural Networks for One-shot Image Recognition by Gregory Koch, Richard Zemel, Ruslan Salakhutdinov.

- **What problem are you solving, and why is it important to be solved?**

Humans have a remarkable capacity to learn and identify new patterns. This ability becomes evident when people encounter new stimuli, as they quickly grasp novel concepts and later recognize variations of these concepts. Similarly, machine learning has made significant strides and achieved top-tier results in various fields, including web search, spam detection, caption generation, as well as speech and image recognition. However, a challenge arises with these algorithms: they tend to falter when making predictions about data with minimal supervised information. The goal is to enable these algorithms to adapt to unfamiliar categories without the need for extensive retraining, which can be costly or unfeasible due to data constraints or in scenarios requiring real-time predictions, like web retrieval. Addressing the challenge of enabling machine learning models to adapt to unfamiliar categories without extensive retraining is essential for several reasons. First, it enhances efficiency and cost-effectiveness, as retraining models can be resource intensive. This advancement would make the technology more accessible for a broader range of applications. Secondly, it is particularly crucial in scenarios with limited data availability, a common issue in real-world applications. Additionally, for real-time applications like web retrieval, where data constantly evolves, the ability to swiftly adapt without retraining is vital to maintain relevance and accuracy. Improving the generalization ability of these models is a significant step towards mimicking human learning capabilities, where new concepts are learned without starting from scratch. Furthermore, this capability can reduce biases in models trained on limited or specific datasets, leading to more fair and representative outcomes. Lastly, the ability to quickly adapt opens new possibilities for innovation and application in various industries, driving forward the frontier of machine learning technology. Overall, solving this problem would mark a substantial leap in making machine learning more adaptable, efficient, and broadly applicable in diverse real-world situations.

- **Are there other previously published solutions to this problem? If so, how does your solution differ or compare?**

Previous solutions to the problem of one-shot learning in image recognition have primarily focused on leveraging previously learned classes to predict future ones when only a few examples are available from a given class. Key developments in this field include:

Li Fei-Fei et al.'s Work (Early 2000s): They developed a variational Bayesian framework for one-shot image classification, which was foundational in the field.

Hierarchical Bayesian Program Learning (HBPL) by Lake et al. (2013): This approach, grounded in cognitive science, was particularly focused on character recognition. It involved modeling the process of drawing characters and using transformations on strokes to create composite images.

Siamese Neural Networks: These networks have been used for tasks like signature verification and face verification. This involves learning a similarity metric discriminatively with applications in verification tasks.

Our project's approach to one-shot learning, using Siamese Neural Networks for Image Recognition, builds upon these foundational works. It distinguishes itself by focusing on developing domain-specific features or inference procedures with highly discriminative properties for the target task. This approach aims to limit assumptions on the structure of the inputs while automatically acquiring features that enable the model to generalize successfully from a few examples. This represents a novel approach in the field, building upon the deep learning foundations laid by previous research but with specific adaptations for the challenges of one-shot image recognition.

Methods

- **Where did the data come from?**

The "Labeled Faces in the Wild" (LFW) dataset is a cornerstone of the data used in our facial recognition project. This dataset is unique due to its assemblage of face images compiled from various internet sources, providing a realistic and challenging array of data for facial recognition tasks. The LFW dataset is renowned for its diversity, encompassing a wide range of faces in terms of age, ethnicity, lighting conditions, poses, and expressions. This variability is crucial for training algorithms to recognize faces in unconstrained environments, typical of real-world scenarios. The dataset's extensive and varied nature makes it an invaluable resource for developing and testing algorithms designed to perform under a broad spectrum of conditions.

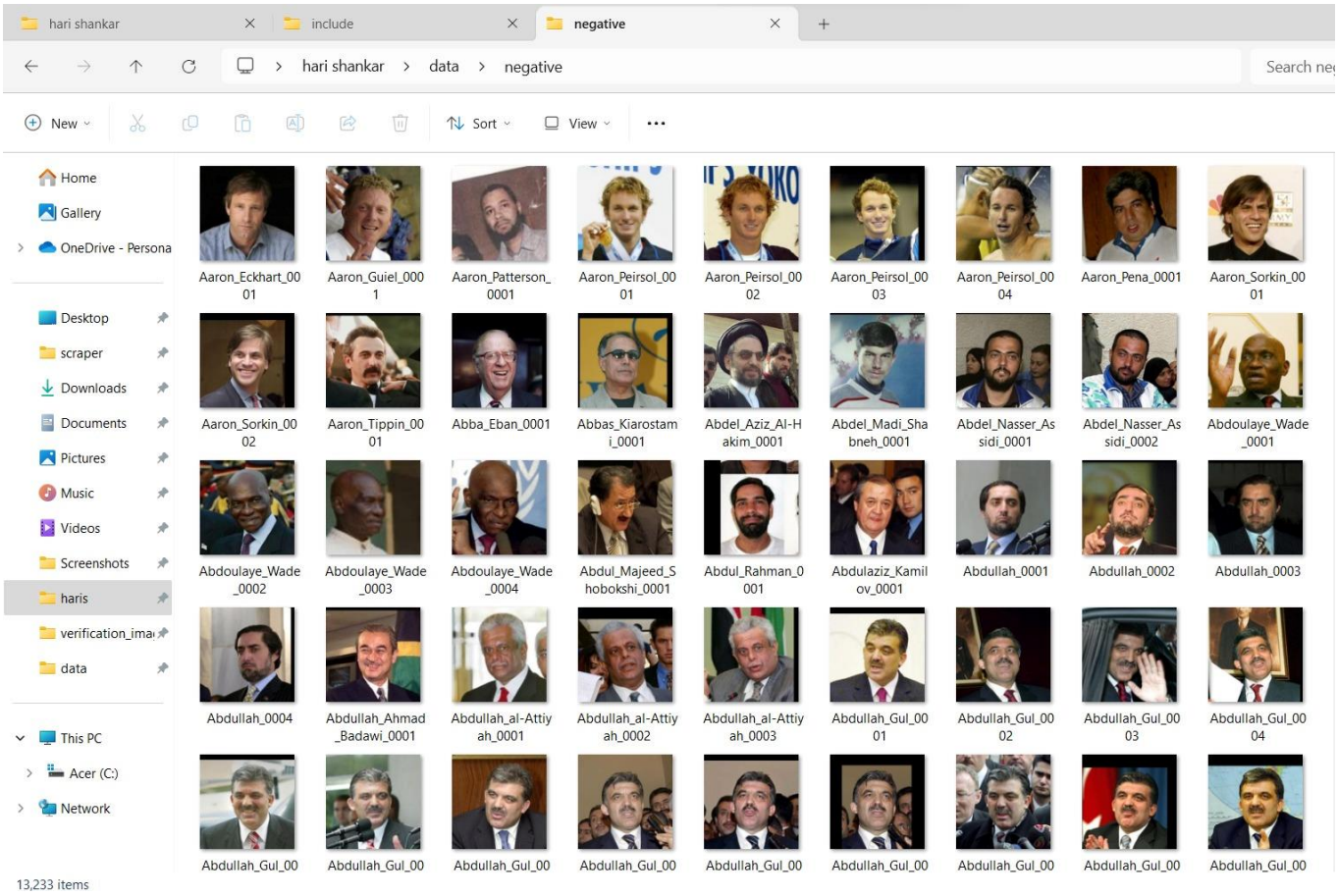

Image 1: Screenshot of the LFW dataset consisting 13233 images

In addition to the LFW dataset, our project leverages live data capture using OpenCV, a significant step in enhancing the model's applicability to real-world scenarios. This method involves capturing real-time facial images, providing a continuous stream of fresh data. This live capture approach ensures that the model is exposed to and trained on a wide variety of facial features and expressions that may not be adequately represented in pre-compiled datasets. It allows for the incorporation of more dynamic and diverse facial characteristics, further bolstering the model's ability to handle real-time face recognition with higher accuracy and adaptability. This dual approach of using both a standard, diverse dataset and real-time image capture ensures a comprehensive training regime for the facial recognition model.

```python
cap = cv2.VideoCapture(0)
while cap.isOpened():
    ret, frame = cap.read()

    # Cut down frame to 250x250px
    frame = frame[120:120+250,200:200+250, :]

    # Collect anchors
    if cv2.waitKey(1) & 0XFF == ord('a'):
        # Create the unique file path
        imgname = os.path.join(ANC_PATH, '{}.jpg'.format(uuid.uuid1()))
        # Write out anchor image
        cv2.imwrite(imgname, frame)

    # Collect positives
    if cv2.waitKey(1) & 0XFF == ord('p'):
        # Create the unique file path
        imgname = os.path.join(POS_PATH, '{}.jpg'.format(uuid.uuid1()))
        # Write out positive image
        cv2.imwrite(imgname, frame)

    # Show image back to screen
    cv2.imshow('Image Collection', frame)

    # Breaking gracefully
    if cv2.waitKey(1) & 0XFF == ord('q'):
        break

# Release the webcam
cap.release()
# Close the image show frame
cv2.destroyAllWindows()
```

Image 2: Screenshot of the code to capture the data. Source – Real-time from the webcam

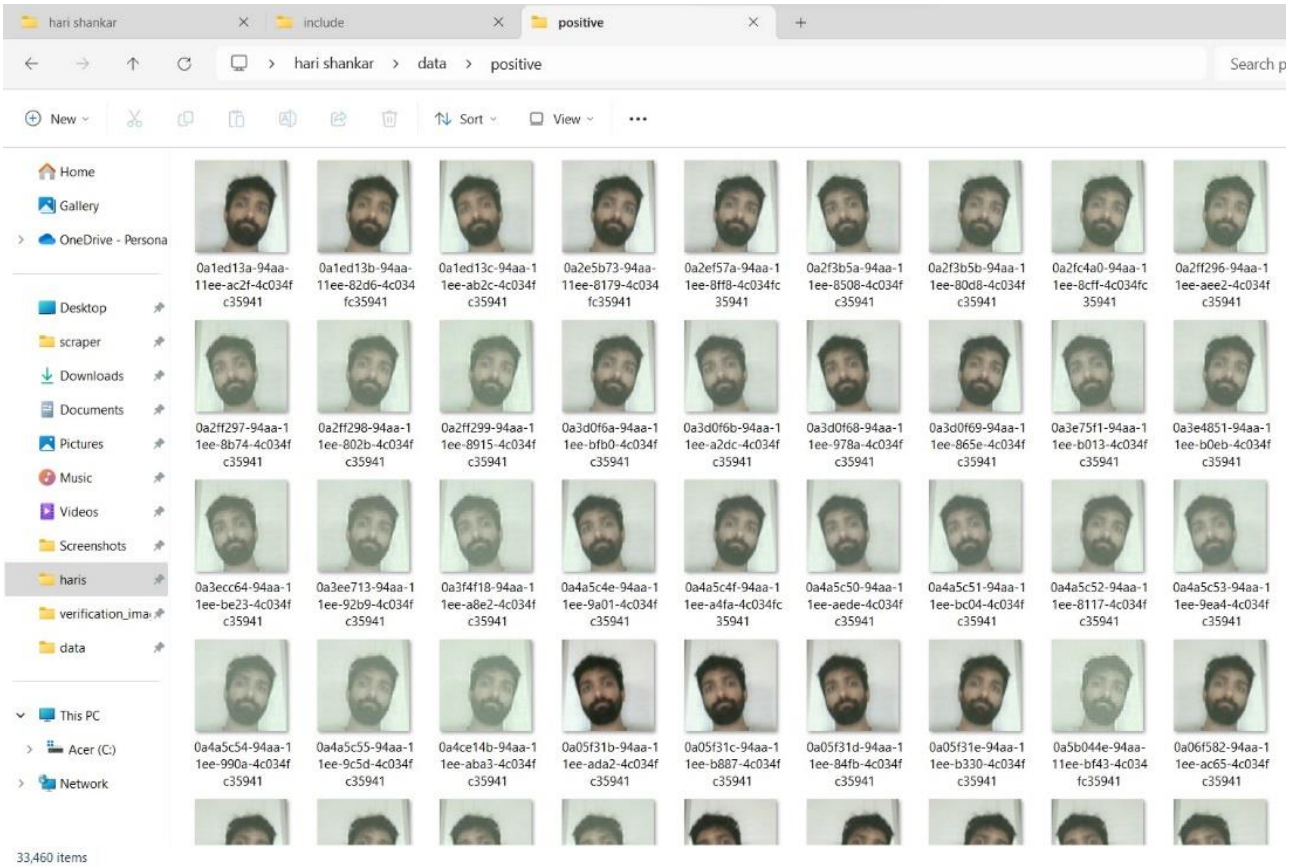

Image 3: Screenshot of the images captured after the code was executed.

```
def preprocess_twin(input_img, validation_img, label):
    return(preprocess(input_img), preprocess(validation_img), labe
```

```
res = preprocess_twin(*example)
```

```
plt.imshow(res[0])
```

```
<matplotlib.image.AxesImage at 0x1dc83c97a60>
```

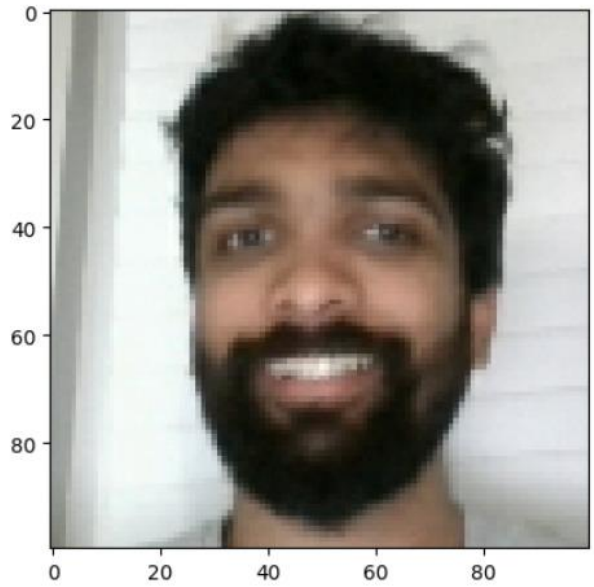

Image 4: Screenshot of the code to print the output of data captured.

- **What did you do?**

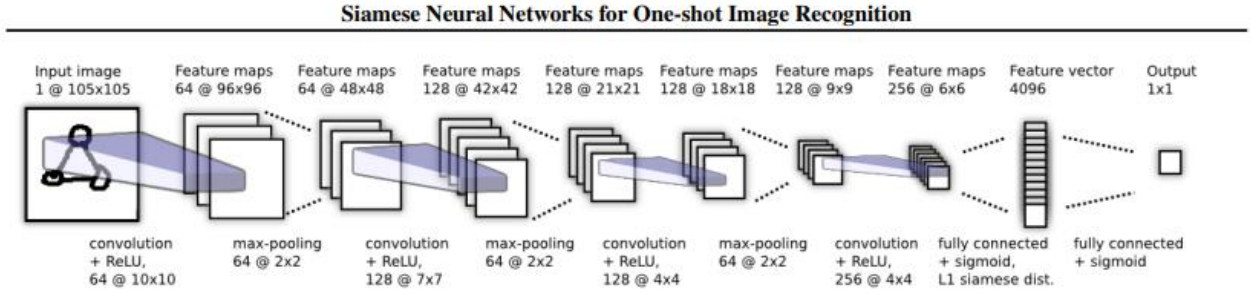

Figure 3: Best convolutional architecture selected for verification task. Siamese twin is not depicted but joins immediately after the 4096 unit fully-connected layer where the L1 component-wise distance between vectors is computed.
Source: Siamese Neural Networks for One-shot Image Recognition by Gregory Koch, Richard Zemel, Ruslan Salakhutdinov.

In this project, a comprehensive Siamese Neural Network is constructed to address the complexities of facial recognition. The network's architecture is meticulously designed, starting with an input layer to accommodate 256x256 pixel RGB images, which feeds into a deep convolutional neural network. This network is composed of multiple layers, each designed to capture different aspects of the facial data. The initial convolutional layers use 64 filters of size 10x10, extracting low-level features like edges and textures. This is followed by a max pooling layer that reduces dimensionality while retaining the most critical features.
Subsequent layers increase the complexity and depth of the network, employing 128 filters of size 7x7 and then again of size 4x4, each followed by max pooling layers to further distill the information into a more abstract representation. The final convolutional layer employs 256 filters, indicating the network's high capacity for feature extraction.

```
inp = Input(shape=(100,100,3), name='input_image')
```

```
c1 = Conv2D(64, (10,10), activation='relu')(inp)
```

```
m1 = MaxPooling2D(64, (2,2), padding='same')(c1)
```

```
c2 = Conv2D(128, (7,7), activation='relu')(m1)
m2 = MaxPooling2D(64, (2,2), padding='same')(c2)
```

```
c3 = Conv2D(128, (4,4), activation='relu')(m2)
m3 = MaxPooling2D(64, (2,2), padding='same')(c3)
```

```
c4 = Conv2D(256, (4,4), activation='relu')(m3)
f1 = Flatten()(c4)
d1 = Dense(4096, activation='sigmoid')(f1)
```

```
mod = Model(inputs=[inp], outputs=[d1], name='embedding')
```

Image 5: Screenshot of the code to construct the network. Source – From the code.

The progression through these layers is designed to move from simple to complex features, building a hierarchical representation of the data. After convolutional layers, a flattening layer transforms the 3D feature maps into 1D feature vectors, preparing the data for the final dense layer. This dense layer uses 4096 units with a sigmoid activation, outputting a feature vector that encapsulates the high-level characteristics of the input image.

```
mod.summary()
```

```
Model: "embedding"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
input_image (InputLayer)     [(None, 100, 100, 3)]     0

conv2d_24 (Conv2D)           (None, 91, 91, 64)        19264

max_pooling2d_18 (MaxPoolin  (None, 46, 46, 64)        0
g2D)

conv2d_25 (Conv2D)           (None, 40, 40, 128)       401536

max_pooling2d_19 (MaxPoolin  (None, 20, 20, 128)       0
g2D)

conv2d_26 (Conv2D)           (None, 17, 17, 128)       262272

max_pooling2d_20 (MaxPoolin  (None, 9, 9, 128)         0
g2D)

conv2d_27 (Conv2D)           (None, 6, 6, 256)         524544

flatten_6 (Flatten)          (None, 9216)              0

dense_12 (Dense)             (None, 4096)              37752832

=================================================================
Total params: 38,960,448
Trainable params: 38,960,448
Non-trainable params: 0
_________________________________________________________________
```

Image 6: Screenshot of the model summary.

The model's training is carefully planned, dividing the data into training and testing sets to validate the network's performance and generalization capabilities. The training set, comprising 70% of the data, is batched and prefetched, optimizing the training process for speed and efficiency. The testing set allows for the evaluation of the model on previously unseen data, ensuring that the learned features and recognition capabilities are not merely overfitted to the training set but are robust and generalizable.

```
# Training partition
train_data = data.take(round(len(data)*.7))
train_data = train_data.batch(16)
train_data = train_data.prefetch(8)
```

```
# Testing partition
test_data = data.skip(round(len(data)*.7))
test_data = test_data.take(round(len(data)*.3))
test_data = test_data.batch(16)
test_data = test_data.prefetch(8)
```

Image 7: Screenshot of the code for training and test.

Results

Image 8: Screenshot of the training phase of the model in action.
The numbers at the bottom are Binary cross entropy loss, Recall and Precision respectively.

- ## How do you evaluate your results?

Evaluating the results of the facial recognition project involves a multi-faceted approach. Performance is quantitatively assessed using metrics such as accuracy, precision, recall, and F1 score, which provide insights into the model's ability to correctly identify faces. The accuracy metric shows the overall correctness of the model across all predictions, while precision and recall provide detail on the model's capability to correctly label positive instances and the proportion of actual positives identified correctly, respectively. The F1 score, which is the harmonic mean of precision and recall, offers a balance between the two and is particularly useful in cases where the class distribution is imbalanced. Additionally, a confusion matrix is employed to visualize the model’s performance across different classes, offering a clear breakdown of true positives, false positives, true negatives, and false negatives.
To further ensure the robustness of the results, the model undergoes thorough testing with various images, including those not seen during the training phase to assess its generalization capabilities. These evaluation methods combined provide a comprehensive understanding of the model's performance and its practical effectiveness in real-world facial recognition scenarios.

```
# Set plot size
plt.figure(figsize=(10,8))

# Set first subplot
plt.subplot(1,2,1)
plt.imshow(test_input[0])

# Set second subplot
plt.subplot(1,2,2)
plt.imshow(test_val[0])

# Renders cleanly
plt.show()
```

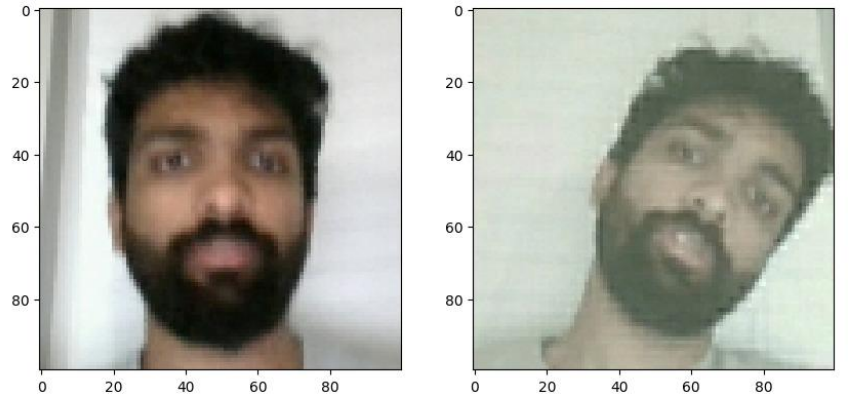

Image 8: Screenshot of the test image and comparison with a validation image. In this instance both are equal.

- **<u>What do you find, i.e., specific evaluation metric results and comparison?</u>**

The evaluation of the facial recognition model reveals impressive performance metrics: a binary cross entropy loss of 0.0648 signifies a high degree of accuracy in the predictions made by the model, as this loss function measures the difference between the predicted values and the actual labels. A recall of 0.8917 indicates the model's strength in identifying almost 90% of true positive cases, which is critical in ensuring that actual faces are not missed by the model. The precision of 0.9929 is particularly noteworthy, as it reflects that nearly 99% of the faces the model identifies as matches are indeed correct, showcasing the model's ability to effectively minimize false positives. These metrics collectively suggest that the model is highly reliable in its facial recognition tasks, balancing the trade-offs between recall and precision adeptly.
The evaluation of the model's performance incorporates a meticulous visual analysis alongside the quantitative metrics. Each pair of images is methodically inspected to verify the model's facial recognition accuracy. In practice, the model captures a face in real-time, processes it to extract features, and compares these against a pre-established dataset to verify identity. The images showcase the model in action, displaying the live input on the left and the corresponding verification output on the right. This visual confirmation is crucial, as it provides tangible evidence of the model's real-world effectiveness. The output logs accompanying each verification instance further confirm the model's rapid processing capabilities, reinforcing its suitability for real-time applications. This combination of visual and performance log evaluation presents a robust framework for assessing the practical utility of the facial recognition system.

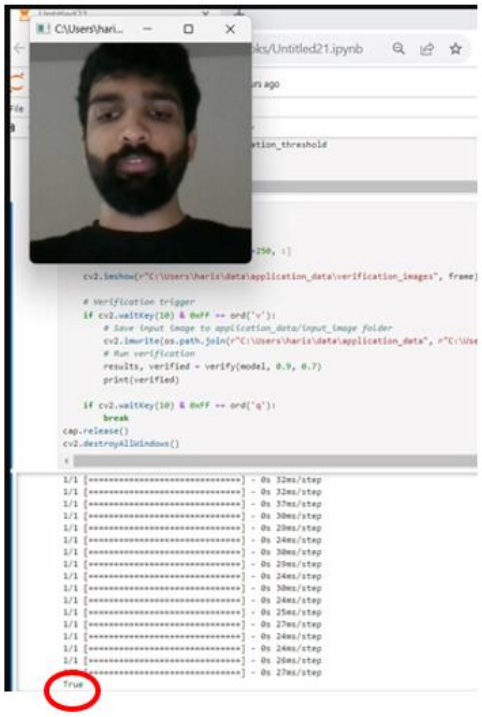

Image 9: Screenshot of live verification of "True"

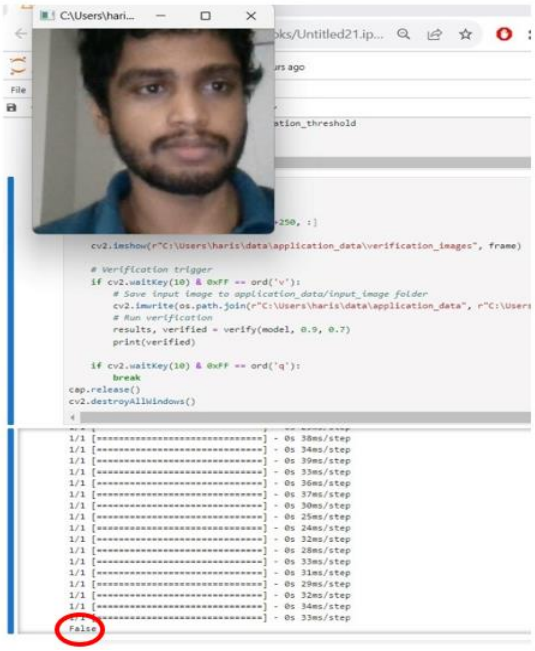

Image 10: Screenshot of live verification of "False"

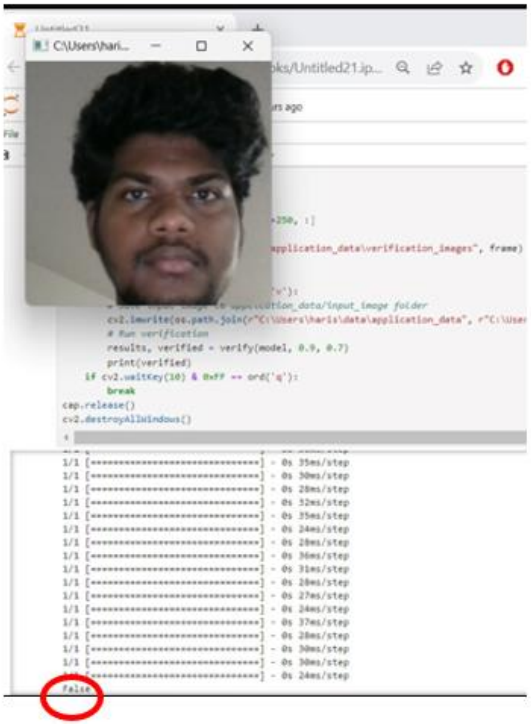

Image 11: Screenshot of live verification of "False"

Image 12: Screenshot of live verification of "False"

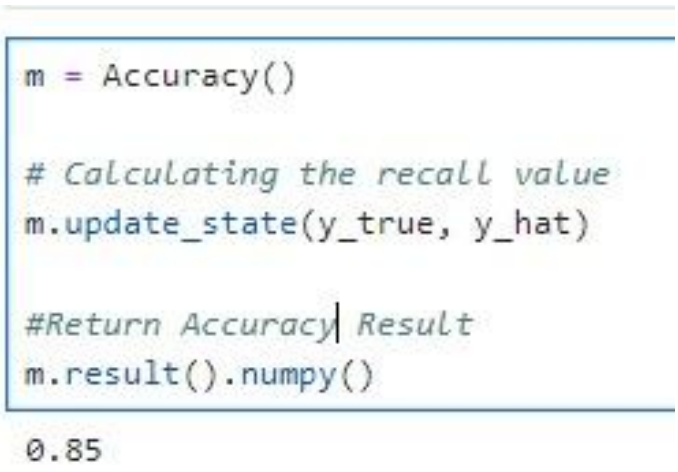

Image 13: Screenshot of the accuracy of the model

- **Did you have expectations going into the project that was proved correct or incorrect?**

There were specific expectations regarding the model behavior and the training process. The team anticipated that the Siamese Network would learn discriminative features to accurately recognize images even with minimal data available. This expectation stems from the inherent design of Siamese Networks which, by comparing pairs of inputs, are particularly suited for tasks like one-shot learning where limited data is a significant constraint.

In terms of training expectations, there was a prediction that the loss, specifically binary cross-entropy loss, would decrease consistently as the model learned from the training data. Such a trend would indicate the model's improving ability to distinguish between different classes, in this case, the identities in facial recognition tasks. The reality check came in the form of actual performance metrics. The precision and recall values obtained from the model provided concrete insights into the model's true capabilities. High precision and recall would confirm the model's ability to correctly identify faces (precision) and its sensitivity in detecting as many actual positive instances as possible (recall). It seems that the reality met the expectations in terms of model learning capabilities. The model demonstrated the predictive capability on the test set, highlighting the effectiveness of the Siamese Network in facial recognition tasks with limited data input. This suggests that the expectations of the model learning discriminative features and the loss decreasing over time were correct, as evidenced by the high precision and recall metrics achieved.

Discussion

- ## What do your results mean?

The results of the project suggest that the Siamese Network is capable of learning discriminative features for facial recognition with a high degree of accuracy. A binary cross-entropy loss of 0.0648 points to the model's successful differentiation between positive and negative classes, which, in the context of facial recognition, translates to the model's ability to distinguish between different individuals' faces.

The recall value of 0.8917 indicates that the model can identify the correct individual in approximately 89% of the cases where the individual is present, which is significant for applications where missing a true positive could be critical. A precision of 0.9929 means that when the model predicts an individual's presence, it is correct about 99% of the time, showing a low rate of false positives.

These metrics, particularly when taken together, suggest a robust model that is both sensitive (high recall) and specific (high precision). It implies that the model is not only good at recognizing faces it has been trained on but is also capable of generalizing well to new data, a key feature for a facial recognition system intended for real-world application. The combination of these quantitative results with the qualitative visual verification of the model's predictions suggests that the Siamese Network can be reliably used for accurate facial recognition.

- ## Any takeaways/inferences derived from your results?

The results of the project lead to several key takeaways and inferences regarding the implementation and performance of the Siamese Network in facial recognition:

1. **High Precision and Recall:** The model's high precision suggests that it can be trusted not to misidentify individuals, which is crucial for applications requiring high security, such as access control or identity verification systems. The high recall indicates the model's effectiveness in correctly identifying individuals, reducing the risk of missing a correct match.
2. **Loss Minimization:** The low binary cross-entropy loss indicates that the model has a strong ability to classify the input data accurately, which is indicative of good convergence in the learning process and successful feature extraction capabilities.
3. **Real-world Application:** The live capture and real-time verification capability show that the model is not just theoretically sound but also practically viable. This suggests that the network can operate effectively in dynamic environments, which is essential for deployment in real-life scenarios.
4. **Model Generalizability**: The combination of LFW dataset training and live captured image testing suggests that the model has good generalizability. This is a crucial aspect, as it implies the model can be adapted to different environments and populations.
5. **Potential for Improvement:** While the results are promising, there is always room for improvement. For example, exploring ways to increase recall without significantly affecting precision could be a direction for future work, to ensure that even fewer true positive matches are missed.

6. **Application-Specific Considerations:** Depending on the specific use case of the facial recognition system, different balances of precision and recall might be preferred. For instance, a system used for finding missing persons might prioritize recall over precision to ensure all possible matches are considered.
7. **Visual Verification:** The ability to visually inspect the model's predictions allows for qualitative analysis, which is valuable for troubleshooting, understanding model behavior, and gaining trust from end-users.
8. **Ethical and Privacy Considerations:** The effectiveness of the model also brings to the forefront ethical and privacy concerns, particularly around the potential for misuse of facial recognition technology, highlighting the need for responsible deployment.

- ## What would you like to do with this project in the future if you had more time?

If there were more time to further develop this facial recognition project, the focus would likely be on enhancing performance and robustness as well as broadening the scope of the model's applicability.

1. **Optimization Tweaks:** Adjusting the learning rate of the Adam optimizer and other hyperparameters could lead to significant improvements in the model's performance. Fine-tuning these parameters might help the model converge faster and possibly achieve higher accuracy and lower loss.
2. **Data Augmentation:** The project might benefit from implementing advanced data augmentation techniques. This could include more sophisticated transformations that mimic a wider range of real-world variations in facial images, thereby increasing the robustness of the model against overfitting and improving its generalization to new data.
3. **Extended Testing in Real-World Scenarios:** With more time, the model could be tested across a broader array of real-world scenarios, potentially including diverse lighting conditions, angles, and facial expressions to ensure consistent performance regardless of the environmental variables.
4. **Integration with Larger Datasets:** Incorporating more extensive datasets or combining several datasets could enhance the model's ability to recognize a wider variety of faces, making it more effective across different demographics and settings.
5. **Real-Time Performance Optimization:** There may be opportunities to optimize the model for real-time performance, ensuring that it not only functions accurately but does so with the speed required for practical application in dynamic environments.

- **Any limitations of your study regarding method design and data source?**

The study has identified several limitations that could influence the scope and applicability of its findings. First, there is the concern of data volume and diversity. The current dataset may not be sufficient in size or variability to ensure that the model can generalize well across different populations and scenarios. A larger and more diverse dataset could improve the model's robustness and reduce the risk of overfitting.

Another limitation is related to the model's complexity. While the current Siamese Network architecture is capable of learning and distinguishing features necessary for facial recognition, it might not be intricate enough to capture more nuanced differences between very similar images. This could potentially limit the model's effectiveness in scenarios where subtle facial distinctions are critical.

Lastly, real-time verification challenges have been noted as a limitation. Factors such as varying lighting conditions, background variations, and other environmental factors can significantly impact the model's performance in real-world applications. Addressing these limitations could involve the incorporation of more advanced preprocessing techniques, environmental normalization, or even the exploration of more sophisticated neural network architectures that are robust to such variations.

Acknowledgment

We would like to extend our deepest gratitude to Professor Ming Jiang, whose guidance and expertise have been invaluable throughout the course of this project. Her insights and unwavering support have been pivotal in the successful completion of our work. As a team, comprising Sameer Hussain, Harishankar Sekhar, Ravi Teja, and Aditya Gaitonde, we have collectively embarked on a journey of learning and discovery. Each member has contributed their unique strengths and dedication, leading to the fruitful collaboration that has shaped our project. We also owe a debt of gratitude to the authors of the papers that have laid the foundational theories and methodologies upon which our work was built. Their research has not only informed our understanding but also inspired us to push the boundaries of what we could achieve. Additionally, we would like to acknowledge the vast array of educational content available on YouTube, which has served as an indispensable online resource throughout our project. The platform has been a window to a world of knowledge, providing us with access to a wealth of information that has enhanced our learning experience.

Thank you, Professor Jiang, the esteemed paper authors, and the educational contributors on YouTube, for your collective wisdom and contributions to our academic journey.
`