Abstract:

This study develop a human face recognition using Siamese CNN, a convolutional neural network architecture. The study found that the Siamese CNN architecture is both practical and accurate, with the best results achieved with a dataset of 9,000 face images, a buffer size of 10,000, and epochs of 5. The proposed method achieved a minimum loss of 0.002, recall of 0.996, precision of 0.999, and F1-score of 0.672. The Siamese CNN model was successfully implemented in Python.

Introduction:

Face and facial expression recognition have gained significant interest in academic studies worldwide over the past five years. Faces are considered essential body features and can communicate with different emotions. Face recognition systems, based on feature extraction and dimension reduction, are used to validate human identity. However, face recognition remains a challenging challenge in real-world applications, with no reliable solution available. The challenge is divided into two groups: face verification jobs, which involve one-to-one matching, and facial recognition tasks, which require humans to determine an individual’s identity. Despite the progress made in face recognition algorithms, no technique provides a reliable solution for the various conditions and applications faced by face recognition. This study aims to develop a Siamese CNN model for face recognition. The model creates encodings for input images and calculates encodings without changing network parameters. The model can be used for face recognition, signature verification, and object tracking in computer vision.

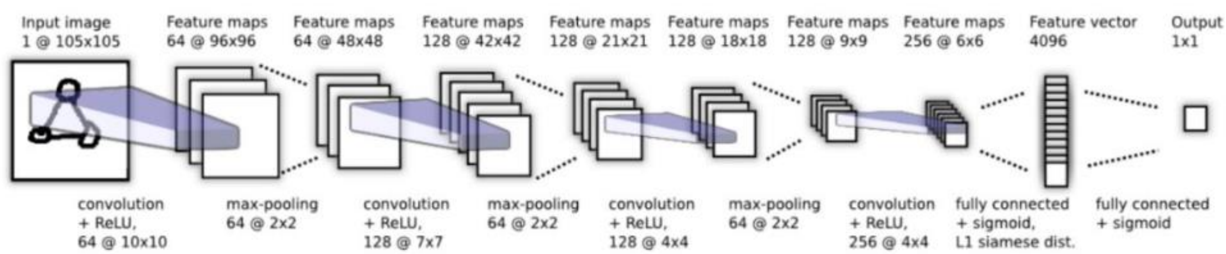
Method:

Image datasets:

The study used a.jpg file format to save three images: the anchor, positive, and negative images, which were used to train a face recognition model using Siamese CNN. The anchor image, which includes the face of person A, is the existing face. The positive image, which includes the face of person A, is the positive image. The negative image may be associated with person B, C, or Y. The anchor and positive images were captured using a 250x250 resolution camera, yielding 550 and 400 images, respectively. The labeled faces in the wild (LFW) face database was used for the negative image, which contains 13233 images of people’s faces labeled with their names.

Augmentation techniques:

We attempted five different augmentation techniques: random brightness, random contrast, random left-right flip, random jpeg quality, and random saturation. Figure 3 depicts their augmentations. This augmentation aims to replicate data to make the categorization process easier. After using the augmentation approach, the anchor and positive images' data increase to 5,590 and 4,000, respectively. Siamese convolutional neural network: The typical model is Siamese CNN with 'L' layers each with 'Nl' units, where 1,l denotes the hidden vector in layer l for the first twin, and 2,l means the same for the second twin. In the first L - 2 layers, we only employ rectified linear ReLU units, whereas the subsequent layers use sigmoidal units. The model comprises a series of convolutional layers, each of which employs a single channel with different-sized filters and a fixed stride of one. The number of convolutional filters is given as a multiple of 16 to improve performance. The resultant feature maps are subjected to a ReLU activation function, which is optionally followed by max-pooling with filter size and stride of 2.



Training:

Training for Siamese CNN is done in mini-batch sizes. In order to create an effective network train, and Siamese CNN training involves selecting image pairings randomly in mini-batch sizes to avoid imbalanced image pairs. The anchor in distinct classes is random, and the paired images are controlled as half the same class and half the different class. The weights are updated using an adaptive moment estimate optimizer (Adam) with an initial learning rate of 0.0001. The encodings of input images are computed in Siamese CNN, and the results perform the same task with the same network, calculating the encoding image of a different individual.

Encoding comparisons reveal that the photos belong to the same individual. The network's training uses an anchor image and compares it to positive and negative images, with a modest gap between anchor-positive and anchor-negative and a substantial gap between anchor-negative and anchor-positive this is called the triplet loss function, and it may be used to compute gradients.

Performance evaluation:

The model is evaluated using four metrics: accuracy, recall, precision, and F1-score. Accuracy is the percentage of forecasts that match actual data, while precision is the percentage of validated main face images. Recall, also known as true positive rate (TPR), is the percentage of verified main face photos correctly classified as positive. Specificity, also known as TNR, is the fraction of facial characteristics not in the primary face image classified as negative. The F1-score is the harmonic mean derived from the weighted average of precision and recall.

Results and Discussion:

The study tested the effectiveness of the proposed Siamese CNN model in face recognition instances using various approaches, data collection methods, buffer size, and epoch count.

Method I used 900 face photos without augmentation, while Methods II and III collected 9,000 with augmentation. The sample count influenced the training task, and the buffer size was increased to 10,000 from Method I's 1,024. The epoch influence was also examined, with

Methods I, II, and II using 50, 5, and 32 epochs, respectively. The results are presented in the Table below.

$$Accuracy = \frac{TP+TN}{(TP+TN+FP+FN)}$$

$$Precision = \frac{TP}{(TP+FP)}$$

$$Recall = \frac{TP}{(TP+FN)}$$

$$F1score = \frac{2(Recall \times Precision)}{Recall+Precision}$$

Table 1. Performance evaluation of our method without augmentation and with augmentation

	Anchor	Positive	Negative	Buffer size	Epoch	Loss	Recall	Precision	F1-Score
Method I (without augmentation)	300	300	300	1,024	50	1.788	1	1	0.997
Method II (with augmentation)	3,000	3,000	3,000	10,000	5	0.002	0.996	0.999	0.672
Method III (with augmentation)	3,000	3,000	3,000	10,000	32	0.694	1	0.506	0.997

Conclusion:

This paper proposes a method to enhance face detection performance using Siamese CNN. The augmentation technique results in superior results compared to non-augmentation methods.

The approach was tested on 9,000 face images, with a classification accuracy rate of 98 percent.

The Siamese CNN can be used for real-world face recognition, which effectively constructs and tests the proposed facial recognition method. The researchers aim to validate the method with various datasets and increase the number of images used to further enhance accuracy.