Face Recognition Using Siamese Neural Network: A Comprehensive Review

Om Sai Vasireddy   
Department of CSE, *SRM University-AP, India*  
omsai\_vasireddy@srmap.edu.in

Vikram Muchumarri  
Department of CSE, *SRM University-AP, India*  
vikram\_muchumarri@srmap.edu.in Uma Maheswar Reddy Nelli  
Department of CSE, *SRM University-AP, India*  
umamaheswar\_nelli@srmap.edu.in

Mudassir Rafi  
 Department of CSE, *SRM University-AP, India*  
mudassir.r@srmap.edu.in Sai Charan Teja Janaki   
Department of CSE, *SRM University-AP, India*  
saicharanteja\_j@srmap.edu.in

.

**Abstract—Today the field of Face Recognition and its applications are experiencing rapid development. Facial Recognition plays a vital role in identifying individuals, even setting apart identical twins. It has become widely adopted in contemporary applications for authentications such as phone unlocking, and criminal identification. Identifying and extracting relevant information from the input data are crucial for training a machine learning model. The One-shot learning method, where we have only given one example of each new class and must make accurate predictions. To automatically prioritize inputs we use a technique for training *Siamese neural networks*. Once we fine-tune the network, we can use its strong features to make predictions not only on new data but also on completely new categories that it hasn’t seen before. Convolutional Neural Network (CNN) is one of the deep learning techniques used in face recognition. Using this convolutional architecture, we get better results that do better than other models, especially in tasks where we only have one example for each category.**

Keywords—Face Recognition; Machine Learning; Deep Learning; Siamese Neural Network; Convolutional Neural Network;

# Introduction

Face recognition is identifying a person by looking at their facial features. We can do this by a powerful technique,Deep learning .we use large set of photos to help the deep learning system to recognize faces and help it learn characteristics that make each face different .The system then saves a special description of each face in database .When recognizing a face, the system then compares the input face which is saved and find for the best face available for match .This paper mainly focuses on enhancing the performance of different types of convolutional neural networks (CNN), in terms of accuracy ,generalization and inference speed ,using several optimization methods such as state-of-art rectified Adam. Convolutional neural networks (CNNs) are widely employed to extract features for face recognition systems by using images. On the other hand deep learning techniques like Siamese networks are used to improve matching. [[1]](#_Vincent_Gitz,_Alexandre).

Deep learning is a kind of state-of-art method which can give high performance on face recognition. With the advancement of deep learning the primary method used in face recognition now relies on Convolutional Neural Networks(CNNs).This networks are mainly used for processing visual information. It consists of three layers: convolutional, pooling and fully connected. The convolutional and pooling layers are responsible for extracting features, while the fully connected layer is responsible for grouping them together [[2]](#_Albert_Khakimov,_Alisher).

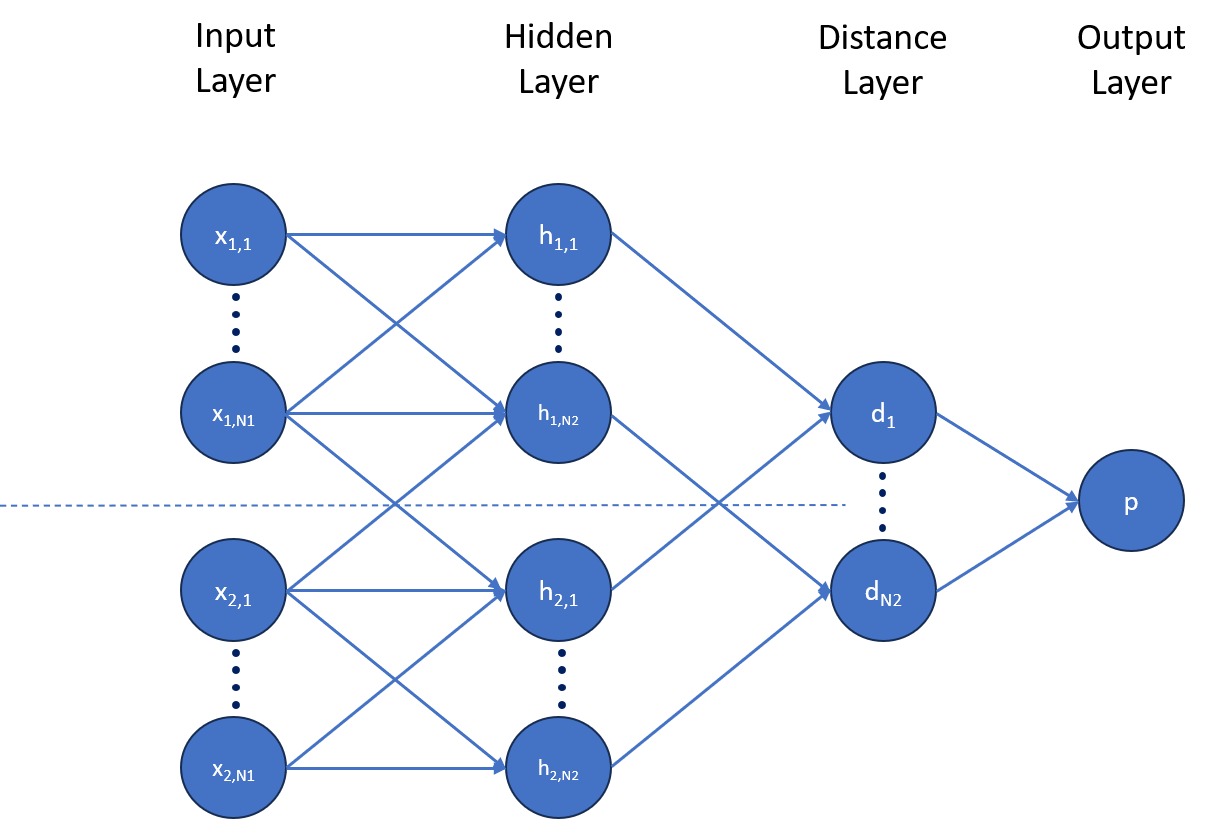
One-shot learning is a machine learning approach where we train a model to make accurate predictions or classifications based on just a single example of each class.[[3]](#_Role_of_Modern).

To create a model for one-shot image classification, our initial goal is to train a neural network that can differentiate between the class of image pairs. This is commonly known as verification task in image classification [[4]](The#_Jehoon_Sung,_). The verification model learns to tell if the two inputs belong to the same category or not. This model can be used to verify new images in a pairwise manner against the test image The pair that the verification network think is most similar gets the highest change of being correct for the task at hand.

The architecture that is designed to measure the similarity and dissimilarity between two input samples is Siamese neural networks. [[5]](#_Vishnu_S,_and). They consists of two identical subnetworks that share a same weights and parameters.[[6]](#_Lili_Li,_Shujuan).The L1 distance (also known as Manhattan distance) is used to take the absolute difference between corresponding elements of the two vectors and summing then up. Mathematically ,*L*1(*X*,*Y*)=∑*i*=1*n*​∣*Xi*​−*Yi*​∣ [[7]](Plant_Village_Dataset#_) The dataset we used is labelled faces in the wild home.[[8]](Plant_Pathology_2020-fgvc7#_)This research presents a strong and straight forward technique for face recognition.

# Related Work

The research on one-shot learning highlights the early stages of machine learning approach that aims to make accurate predictions or classifications with very limited training data.[[9]](#_Shima_Ramesh,_Ramachandra) The important early work on one-shot learning dates back to the early 2000’s with work by Li Fei-Fei et al. Lake and his team took a unique approach to one-shot learning for recognizing characters. They used a method called Hierarchical Bayesian Program Learning (HBPL). Their main focus is on breaking down images into smaller parts to figure out to see how the structure of these parts explained what we see in the image. However, the computational complexity of HBPL posed challenges in inference Maas and Kemp made significant contributions with their pioneering work in 2009, improving Bayesian networks to predict attributes related to Ellis island passenger data. Their research highlights the versatility of Bayesian approaches in handing complex datasets, beyond face recognition. Lims’s work in 2012 focused mainly on practical approach to handle limited datasets for certain classes. [[10]](#_Srdjan_Sladojevic,_Marko).

It involves the use of Convolutional Neural Networks(CNNs) for face recognition. [[11]](#_Pranesh_Kulkarni,_Atharva) CNNs has been applied by many researchers to solve various problems in face recognition (Li & Hua 2015; Parchami et al. 2017 )CNN is the a sort of feedforward neural networks with convolutional computation and deep learning architecture. With the help of RGB level co-occurrence matrix. we have extracted color, shape, and texture features from all the images in the dataset and accomplished feature selection task. with the series of Convolutional layers this architecture is designed . These channels use separate filters of different sizes,To enhance efficiency the quantity of convolutional filters is set as multiple of 16.This CNN consists of multiple sequence convolutional layers.

While following the convolutional layers the network applies a activation function ( ReLU ) Rectified Linear Unit in resulting feature maps. This helps in bringing non-linearity into the model, enabling it to capture complex relationships within the data. [[12]](#_Gnanavel_Sakkarvarthi,_Godfrey)

Following the design of the convolutional layers and activation function ,the model’s optimization process is facilitated by the Adam optimizer . The network ends with the 4096 fully connected layer, where L1 component distance between vectors is computed. [[13]](#_Gnanavel_Sakkarvarthi,_Godfrey) By computing the L1 distance ,the model can effectively compare and evaluate the level of similarity between different facial feature representations. .

# Datasets information

The main goal of our project is to build a face recognition model for a small-scale data set using Convolutional neural network and Siamese neural network models. We collected some labeled images from the data set Labeled Faces in the Wild Home and we also generated some images of ourselves using Opencv to train the model. The dataset mainly contains three kinds of images. The dataset comprises images that are anchor, negative, and positive where anchor images are the images that were fed into the model as input along with either negative or positive images. Negative images contain the images that we have collected from the dataset Labeled Faces in the Wild Home whereas the positive images consist of the user images

1. Details of the datasets

|  |  |
| --- | --- |
| Datasets | No of Images |
| Anchor Images | 804 |
| Negative Images | 13,233 |
| Positive Images | 647 |

# Methodology

In our research, we restrict ourselves to character recognition but this basic approach can be replicated for any other models. Convolutional neural networks and Siamese neural networks were widely used in the classification of persons based on their faces. To train and test our model we resize our data set images to 100x100 pixels to get a uniform flow in the entire process of both training and testing. Before splitting the data to training and testing we augment the data in such a way as to make the model more robust and make the training process easy. This data augmentation is the basis of our model training and evaluation. An anchor image, a negative image, and a label 0 associated with it were combined together similarly an anchor image, a positive image, and a label 1 associated with it were combined together. This is how we have augmented our data. This will help us in running our model with ease of effort where we send two images parallelly from two same kinds of networks and compute the difference between them at the output layers

Figure 1.Flow chart of a general Siamese model network

To develop an efficient and robust model we have used the Siamese networks which also include Convolutional Neural Networks (CNN). This architecture is designed with convolution layers, activation functions, feature maps, loss functions, and Euclidian distance. This model is rigorously trained using the data that we have augmented before to learn the different structures and shapes of the input images and classify them accordingly and the model's performance is evaluated using the reserved testing dataset. Generally, we learn how different images were represented through a supervised metric-based approach with Siamese networks, and then we are going to use the features of the network for one-shot learning without any retraining. We have used rectified linear units (ReLU) in the first L-2 Layers and sigmoidal units in the remaining layers. In our model we have applied the ReLU activation function to the output feature maps, The units which are there in the final layer of convolutional are flattened into a single vector. This convolutional layer is followed by a fully-connected layer which in turn followed by one more layer which computes the induced distance metric between each Siamese twin, which was finally given to a single output layer with sigmoid as it’s output function.

Ley y(x1(i) , x2(i)) be a length-M vector that contains the labels for a given mini-batch, where we assumed that the label is 1 whenever x1 and x2 are from same class and 0 when they are from a different class. We have imposed a regular cross-entropy objective on our binary classifier which looks like this:

L(x1(i) , x2(i)) = y(x1(i) , x2(i))logp(x1(i) , x2(i))+(1 - y(x1(i) , x2(i)))log (1 - p(x1(i) , x2(i))) + λ|w|2

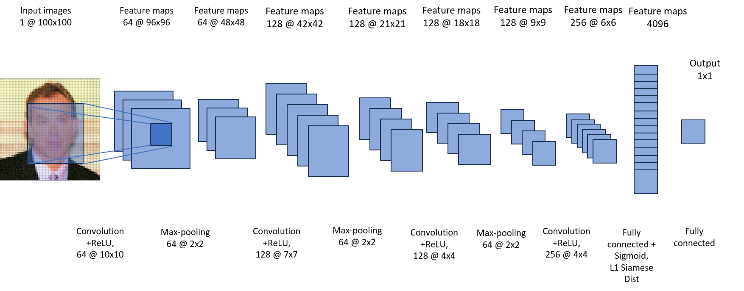
We have used a standard backpropagation algorithm where the gradient acts as an additive across the networks because of the tied weights in the twin networks. After completing optimising our Siamese network model for the verification task, we have to check the discriminative potential of the features which were learned at one-shot learning.

Figure 2. Convolutional Neural Network of our Model

## 3.A. DEFINITIONS

### 1. Learning Rate :

Learning Rate is a hyper parameter which was used to check the pace at which the algorithm learns or updates the values of parameter estimate. It denotes how frequent neural network refreshes the notion that machine has learned.

The amount of weights that are updated during training

is said to be “learning Rate”. learning rate indicates how quickly the model is adapted to the problem.

Too large learning rates can cause the model to converge quickly to a sub-optimal solution whereas low learning rates may cause the process to get stuck as it takes more time to learn for any given data set.

The new weights that are updated by using learning rate is given by the formula.

New weight = existing weight - learning rate \* gradient

*2. Siamese neural network:*

Siamese neural network often called as twin networks is a part of neural network architecture contains some identical networks. If the networks have same configuration with same parameters and weights then they are identical. Siamese networks have the ability to learn from very little data which made them popular in the recent years. Only few Images are used by Siamese network to get good predictions. The main objective of Siamese network is to identify whether data is similar or not. The Similarity score can be calculated by Binary cross entropy, Triplet loss.

*3. Convolutional Neural Networks (CNN):*

CNN is a network architecture used for deep learning algorithms mainly for image recognition and the tasks that involve the pixel data processing. It is made up of multiple layers mainly convolutional layers, pooling layers, fully connected layers.

The convolutional layers are the main component of CNN this layer is responsible for applying filters to input image to extract features such as shapes, textures. The output of the convolutional layer is passed through pooling layers which reduces the spatial dimensions retaining important data. The output of the pooling layers is passed through one or more fully connected layers these are used to classify a image or in making predictions.

CNN networks are good at picking up on designs like lines, circles eyes, and even faces in the input image this makes the CNN more robust for computer network.

*4.Max pooling:*

Max pooling is the pooling layer used in CNN is a operation which selects the maximum element in the region which is covered by the filter. The output of the max pooling operation will be featuring map contains all important features of the previous map.

*5.One shot learning:*

Observing a single example of each possible class before making any predictions about a test instance is called as one-shot learning. Unlike other computer vision projects that uses very large amount of images to detect from one frame to the next, One shot classification can operate it in limited amount of training data set to compare any two images.

*6.Loss function:*

Loss functions are responsible for fitting the data model to the given training data set. Loss function compares the target output with the predicted output it measures how the neural network models the training data. for efficient model loss between targeted and predicted outputs is minimum.

*7.Optimizer:*

Optimizers are used to minimize the loss function by adjusting the model parameters which measures how well the model can make predictions on a given dataset. Different optimization algorithms are available that impact the model's performance we have used Adam optimizer to make predictions for our model.

# Results

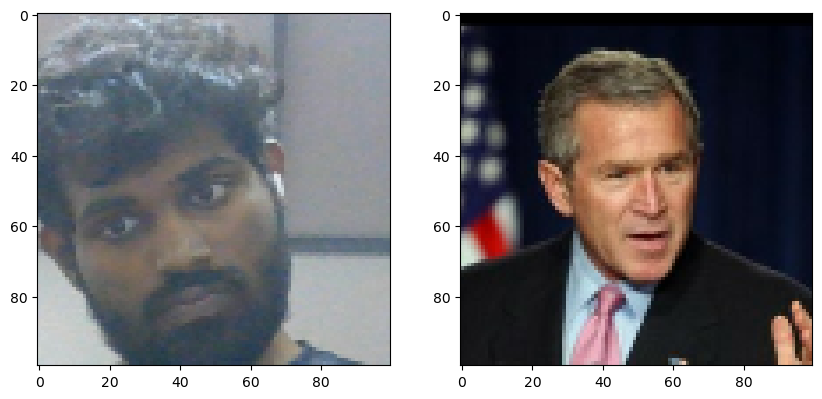
We collected the images for anchor and positive classes using Opencv through the webcam and wild dataset for negative images and preprocessed those images. In the preprocessing step, we scaled down and resized those images for better memory and time optimization of model. We built train and test partition and set the labels for every combinations of anchor images and every positive image as '1' and for every negative images '0'. We divided those combinatio of 7:3 for training partition and testing partition, after that created some layers which are convolutional and max-pooling and a distance layer to find the difference between those images.

Figure 3. A sample dataset of an input and  
a negative image

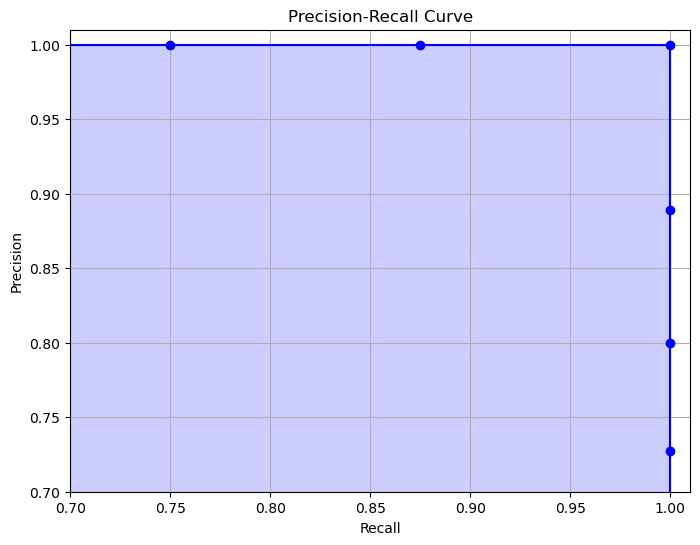
After that we went to built a siamese model and combined all those layers into one and embedded into siamese model and passing through all input images and validation images and passing those output distances from embedding layers into a dense layer using sigmoid activation to get single output either '1' or '0'.

Figure 4 Precision-Recall Curve

Then trained our model by building traning loop and passing our data through siamese model for some epochs in every epoch we are updating the weights in the CNN layers by calculating the data loss and optimizing those by using adam optimizer and saving checkpoints for every 10 epochs.

We evaluated our model using performance metrics which are Precision and Recall.

**Precision:** It is a common measure of finding accuracy of our siamese model for face recognition system. It checks of all positive face recognitions made by the model, how many are true postitives.

**Recall:** It is also called true positive rate, it means of all the actual positives face recognitions made by the model, how many did the model correctly identified?

After that we took a batch of 16 images from the test data as shown in below table, it contains postive or negative images as validation images and predicted labels for those images and using our siamese model and the labels which we got from predictions rounding them to '1' if the value is closer to '1' and rounding them to '0' if the value is closer to '0' to perform metrics to find how accurate of our predictions.

|  |  |  |
| --- | --- | --- |
| **VALIDATION IMAGE** | **LABEL** | **PREDICTED VALUES** |
| Negative | 0 | 1.0303144e-06 |
| Negative | 0 | 8.2426457e-05 |
| Positive | 1 | 9.9999893e-01 |
| Positive | 1 | 1.000000e+00 |
| Negative | 0 | 1.7540457e-05 |
| Negative | 0 | 1.6903167e-07 |
| Negative | 0 | 2.1551946e-06 |
| Positive | 1 | 9.8448521e-01 |
| Negative | 0 | 1.1103551e-05 |
| Positive | 1 | 9.9999082e-01 |
| Positive | 1 | 1.000000e+00 |
| Positive | 1 | 9.9933481e-01 |
| Positive | 1 | 1.000000e+00 |
| Negative | 0 | 1.4063612e-06 |
| Positive | 1 | 7.2294313e-01 |
| Negative | 0 | 5.8970704e-07 |

# Conclusion and Future Work

There are various fields using Siamese Networks. Face Recognition using deep learning is a classic computer vision application. While using CNN-based deep learning, the model requires so much time to train and also need so many input images. While Siamese neural network can work with less input images and can easily find the similarity scores between the two images.

Firstly we used convolutional siamese neural networks for verification, then we perform the one-shot classification. Using opencv and captured our own images to create dataset and worked on that dataset to verify Siamese Neural Network and One-shot classification and after that we worked on real-time dataset and improved the performance of our model by doing that. We also implemented differnt metrics to measure the performance of our model so that we can examine the model and how it can useful in other domains, mostly image classification.

Performing different layers and finding the distance error between all possible combinations of anchor images with positive and negative images and trying to minimize those errors to calculate the outcome or label for each possible combinations. By using these operations the model can extract the features from each image and adapted better and can give more accurate outcome.

Adding some more convolutional layers and max-pooling the output from each layer and it will give more precised distance error between the test image and validation image. Using Adam Optimizer to minimize those errors and updating every time, will gives us the better performance. And increasing the layers will always helps us to improve the performance and can be helped to get better outcome to recognize the facial structure in various situations.

# References

### Krizhevsky, Alex, Sutskever, Ilya, and Hinton, Geoffrey E. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pp. 1097–1105, 2012.

### Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv 2014, arXiv:1409.1556

### Palatucci, Mark, Pomerleau, Dean, Hinton, Geoffrey E, and Mitchell, Tom M. Zero-shot learning with semantic output codes. In Advances in neural information processing systems, pp. 1410–1418, 2009.

### Kuznetsova, A.; Rom, H.; Alldrin, N.; Uijlings, J.; Krasin, I.; Pont-Tuset, J.; Kamali, S.; Popov, S.; Malloci, M.; Duerig, T.; et al. The open images dataset v4: Uniﬁed image classiﬁcation, object detection, and visual relationship detection at scale. arXiv 2018, arXiv:1811.00982.

### Bromley, Jane, Bentz, James W, Bottou, Leon, Guyon,´ Isabelle, LeCun, Yann, Moore, Cliff, Sackinger, Ed-¨ uard, and Shah, Roopak. Signature verification using a siamese time delay neural network. International Journal of Pattern Recognition and Artificial Intelligence, 7 (04):669–688, 1993

### C. Zhang, and Z. Zhang, “Improving multiview face detection with multi-task deep convolutional neural networks,” IEEE Winter Conference on Applications of Computer Vision, 2014, pp. 1036-1041.

### Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; Rabinovich, A. Going deeper with convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 1–9

### Kim, B.-K.; Dong, S.-Y.; Roh, J.; Kim, G.; Lee, S.-Y. Fusing aligned and non-aligned face information for automatic affect recognition in the wild: A deep learning approach. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 48–57.

### Bengio, Yoshua. Learning deep architectures for ai. Foundations and Trends in Machine Learning, 2(1):1–127, 2009.

### Fe-Fei, Li, Fergus, Robert, and Perona, Pietro. A bayesian approach to unsupervised one-shot learning of object categories. In Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on, pp. 1134– 1141. IEEE, 2003

### K. Yan, S. Huang, Y. Song, W. Liu and N. Fan, "Face recognition based on convolution neural network," 2017 36th Chinese ControlConference (CCC), Dalian, 2017, pp. 4077-4081, doi:10.23919/ChiCC.2017.8027997

### Face Recognition Based on Convolution Neural Network. Kewen Yan,Shaohui Huang, Yaoxian Song1, Wei Liu1, Neng Fan.

### G. Koch, R. Zemel and R. Salakhutdinov, "Siamese neural networks for one-shot image recognition", Proceedings of International Conference on Machine Learning, vol. 2, 2015.