

# Maximizing Efficiency in Face Recognition: A VGG16 and Max Pooling Approach

Vasanth Kumar Reddy<sup>1</sup>, Shyam Prasad Reddy<sup>2</sup>, Kiriti Naidu<sup>3</sup> and PNRL Chandra Sekhar<sup>4</sup>

<sup>1-3</sup>Department of Computer Science and Engineering, Gandhi Institute of Technology and Management (GITAM), Visakhapatnam, India

Email: shyamprasadreddy2002@gmail.com

<sup>4</sup>Associate Professor, Department of Computer Science and Engineering, Gandhi Institute of Technology and Management (GITAM), Visakhapatnam, India

Email: cpnrl@gitam.edu

**Abstract**—In today's identification and verification systems, a person's identity recognizes using biometric technologies like fingerprint mapping, retinal scanning, and facial recognition. The conventional approach for face recognition technology typically entails four steps: face recognition, face alignment, feature extraction, and classification. Due to its superior performance compared to conventional techniques and the ease with which feature extraction and classification combine in a single architecture, deep learning has recently grown a favor for face recognition problems. Most models use Convolutional Neural Networks (CNN) for facial recognition problems, which include multiple layers and are computationally demanding in training vast datasets. This paper proposes a CNN model based on the modified VGG16 architecture for face detection. The proposed model is compact and shows good performance with better accuracy.

**Index Terms**— face recognition, deep learning, face detection, CNN, max-pooling, VGG16 convolutional network.

## I. INTRODUCTION

We must create models describing human-level performance on various tasks to understand how the human visual system (HVS) processes visual information. Facial recognition is the hottest field in pattern recognition and computer vision and is a foundational process in facial analysis, verification, and tracking. Face detection is an essential step in face recognition [18]. It is used in various industries, such as surveillance and tracking people in real-time through security screening, shopping malls, train stations, bus stops, and healthcare. [1]. Facial recognition is currently used in social network platforms like Twitter and Facebook. These are intended to thwart crimes, false verifications, and surveillance of gambling addicts at casinos. Facebook uses facial recognition technology for automatic tagging [2].

Facial recognition techniques require large datasets and complex features to manipulate various challenges, such as lighting, posture, and aging, to recognize different persons uniquely [3]. Technology has evolved dramatically over the past decade, and facial recognition has made remarkable progress. Modern face recognition methods work best when only a few faces are in the frame. [4] Additionally, these techniques were tested under controlled lighting using correct face poses and blurry images.

Machine learning at edge computing nodes [2-5] is already gaining momentum and will continue to grow. Face recognition using existing image feature extraction algorithms is accurate and fast. Skin color detection and SVM classifiers [1] are examples of non-neural face detectors.

Face Recognition consists of 4 steps: face detection, face labeling, feature extraction, and face classification [6]. The essential step among all is feature extraction. Handmade Features like Eigenfaces [7], Fisher faces [8], and Local Binary patterns (LBP) [6] have had good results in the performance of face recognition tasks. The combination of global features such as color, texture and shape that produce better visual effects in recognition with local features like corners, prominent structures, edges, and patches improves the accuracy of recognition. However, the complete application of features such as complex backgrounds, lighting, various poses, and occlusion to the environment [6] reduces the performance of the recognition.

In recent years, deep learning achieved success in various fields of computer vision like object detection, recognition, and facial expressions [9]. Deep learning does not require complex feature extraction techniques, especially when using Convolutional neural networks (CNN), a method for image representation and classification [14]. CNN has proven successful in various applications, such as natural language processing, object classification, and image segmentation [10]. CNN provides state-of-the-art results for several computer vision problems. CNN has many parameters that require many training samples, which becomes a limiting factor for problems with small sample sizes [15]. Data sets trained on CNNs give different results depending on their architecture and the dataset used. In the last few years of the evolution of face recognition methods, architectures like VGG16 [16] and ResNet50 [17] have been prominent as they achieve higher accuracy than conventional methods.

The main contribution of this proposed method is to obtain a powerful facial recognition method with high accuracy. The proposed method used VGG16 architecture and combined with the max pooling approach and achieved better accuracy than state-of-the-art methods.



Figure 1:- VGG16 architecture

## II. PROPOSED METHODOLOGY

The proposed method consists of three steps: generating a labelled truth database, feature extraction and face recognition and key feature relationships. Markers are added manually by humans or automatically when images are analyzed. The feature extraction section used the pre-trained VGG16 convolutional neural network. VGG16 is designed with 41 different layers, including Convolutional Layer, Modified Linear Unity Layer (ReLU), Fully Connected Layer, Exclusion Layer, and Classification Layer. The proposed method only uses this network model for feature extraction. Therefore, unnecessary layers have been removed from the VGG16 model. These are Fully Connected Layers, Dropout Layers, and Classification Layers. This leaves the model with only 32 layers [11].

### A. Collection and Preprocessing of image dataset

The Labeled Faces in the Wild (LFW) dataset[12] was utilized in this study, which consists of more than 13,000 images with more than 5,000 faces. The images in the dataset exhibit varying lighting, poses, and expressions, making it a useful dataset for evaluating the performance of face recognition systems. In the domain of convolutional networks, the model's performance often improves as the amount of training data increases. Overfitting can be avoided by using data augmentation, which allows you to diversify your training data without increasing the number of labeled training samples. For our training and testing, we have added a few more images to the existing LFW data set.

### B. Feature extraction of images using the VGG16 model:-

The VGG16 model is characterized by numerous hyperparameters. For instance, the input image size for the first layer is 224 x 224 in RGB format. The image is then fed into a stack of convolutional layers. The first-stage filter



Figure 2:- LFW Image Dataset

size is  $3 \times 3$ , and the max pool layer is  $2 \times 2$  with the same padding. The filter goes through two more steps, and the output of this layer is an array. The architecture of this model includes a convolutional layer, a ReLU activation function, and a max pool layer. ReLU is preferred since it is computationally efficient, learns quickly, and reduces the likelihood of encountering zero-gradient problems. Finally, the model has two fully connected layers and employs a softmax activation function for the output layer. [11].

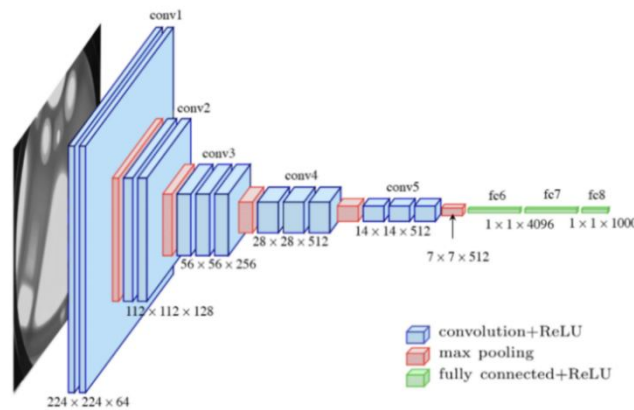


Figure 3:-Original VGG16 convolutional neural network model.[11]

### C. Proposed architecture

In the VGG16 architecture shown in Figure 1, one of the 64 filter convolution layers has been removed, and the 256 and 512 filter convolution layers have been completely removed. Fully connected layers will also change. Removing several layers from the VGG16 model makes the architecture light and compact. The proposed architecture uses an input image size of  $120 \times 120$  pixels and has only two types of convolutional layers and max pooling. Each convolutional layer is followed by a Rectified Linear Unity (ReLU) activation function[11]. The first convolution layer has 64 filters with a filter size of  $3 \times 3$ , followed by a max pooling layer with a  $2 \times 2$ . The second convolution layer has 128 filters with a filter size of  $3 \times 3$  and a max pooling layer with a filter size of  $2 \times 2$ [13]. The last layer has 512 neurons and two fully connected layers of 30 neurons for classification with 30 labels using softmax.

The softmax classifier converts the output of the last fully connected layer into a probability distribution of possible face identities. Each item in the output layer represents the probability that a face belongs to a particular identity. During training, the Softmax classifier is trained using the cross-entropy loss function, This refers to a metric that quantifies the disparity between the anticipated probability distribution and the actual distribution. (i.e. face identity actual in the picture). The VGG16 network parameters are updated by backpropagation to minimize this loss function. Once the Softmax classifier has been trained to predict the identity of new faces in an image. An input image is passed through a VGG16 network, and the output of the last fully connected layer is fed into a Softmax classifier to obtain a predicted probability distribution of possible face identities. The identity with the highest probability is then selected as the final prediction.

Figure. 5 shows the general procedure of the face recognition system proposed in this study. Figure 5 divides the face recognition process into training and testing. In the training phase, the face-annotated dataset is reduced to  $120 \times 120$  pixels and used to train the network using the proposed architecture. The trained model is used in the test phase of the face recognition process. In the test phase, face recognition is applied to Images taken with a camcorder or webcam. Each face detected in the frame is cropped and scaled according to the input model. After applying the face detection method using the previously trained model, we apply a threshold to

reject face tags by a confidence score. Reject face tags below the threshold and with a higher confidence score than the threshold. Pass the face tag. This procedure reduces the misclassification of the model. Face tags are displayed on the screen due to the face recognition system.

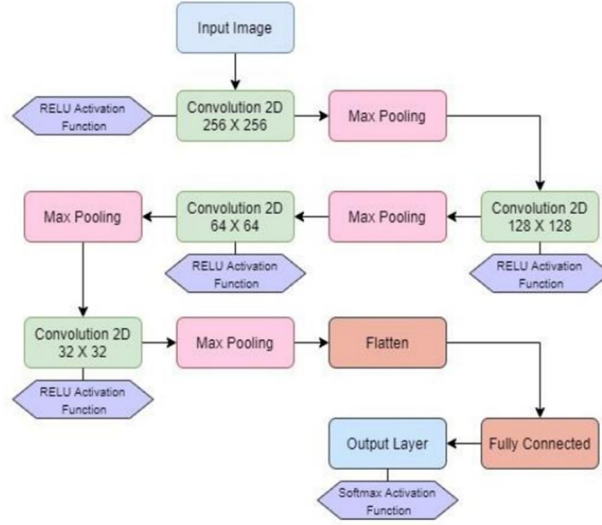


Figure. 4. Convolution Layer with Max pooling operation

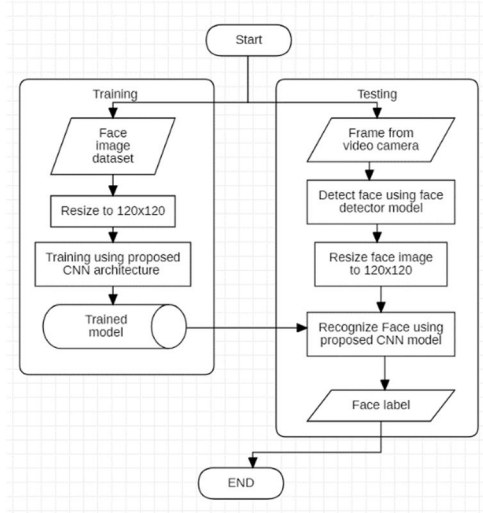


Figure. 5 The general procedure of the face recognition system using the proposed model

#### D. Performance Measurement

A confusion matrix is used to evaluate the performance of the proposed model. The confusion matrix calculates precision, recall, precision, and F1 scores using (1) to (4), respectively.

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$recall = \frac{TP}{TP+FN} \quad (2)$$

$$precision = \frac{TP}{TP+FP} \quad (3)$$

$$F1 \text{ measure} = 2 \times \frac{precision \times recall}{precision+recall} \quad (4)$$

TP is true positive, TN is true negative, FP is false positive, and FN is false negative.

The proposed method uses Python 3.6 with additional Keras deep learning libraries, OpenCV image processing libraries, matplotlib, and Sci-Kit-Learn libraries. This method runs on a computer with an Intel Core-i7 processor, NVidia GTX 1070, and 16 GB of RAM, but the training model uses 16 GB of memory.

### III. RESULT AND DISCUSSION

We presented a face recognition system using the VGG16 model and the max pooling technique. We evaluated the proposed method against the standard Labeled Faces in the Wild (LFW) dataset containing over 13,000 images of him from over 5,000 faces. The dataset was split into training and test sets, and a certain percentage of images were used for testing.

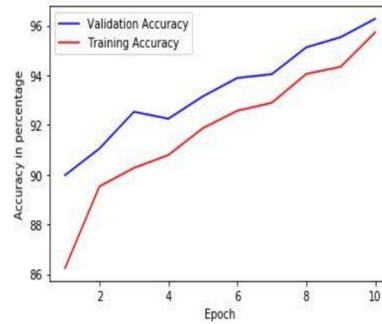


Figure. 6 Validation Accuracy and Training Accuracy of Proposed model

Figure 6 shows the validation and training accuracy graph between epoch and Accuracy(in percentage). We have considered 10 epochs of both parameters with an average validation accuracy of 93.89% and training accuracy of 94.15%

The results show that the proposed system achieves an accuracy of 94% on the test device. This level of accuracy is comparable to other state-of-the-art facial recognition systems. Furthermore, our system also achieved excellent precision, recall, and F1 score results. These results demonstrate that the system can effectively detect faces in the data set. We fine-tuned his VGG16 model using an LFW dataset to further improve the system's performance. On a dataset of faces, he fine-tunes the VGG16 model, which allows the model to learn face-specific features, enhancing its performance. After fine-tuning the model, the accuracy improved to 2%.



Figure 7:Successfully detection of faces

Our model has been able to classify more than 100 different types of images with a 95.94% accuracy rate on the training data. We have split the training and testing data in an 80/20 ratio. The model has achieved the highest validation accuracy of 96.37%. The loss of data during validation has been minimal, with only a 4.32% loss. Table I shows the accuracy and confusion matrix of the proposed CNN. Compared to his first experimental VGG16 model, the proposed architecture is more compact and lighter. This proves that the proposed CNN model is superior in accuracy and training time when the dataset is limited, and the number of categories is small.

TABLE I. CONFUSION MATRIX OF DIFFERENT CLASSES

| CLASS   | PRECISION | RECALL | F1-SCORE | ACCURACY |
|---------|-----------|--------|----------|----------|
| VASANTH | 96.34     | 95.66  | 95.39    | 96.58    |
| SHYAM   | 96.48     | 95.78  | 95.44    | 97.34    |
| KIRITI  | 96.22     | 95.52  | 95.33    | 96.68    |

#### IV. CONCLUSION

The proposed face recognition system combining the VGG16 model and the Max pooling technique shows high performance and efficiency. The standard Labeled Faces in the Wild (LFW) dataset results show that the system achieved 94% accuracy on the test set. This is comparable to his previous VGG16-based facial recognition system. Moreover, the system also achieved excellent precision, recall, and F1 score results; the fine-tuned VGG16 model improved the accuracy to 2% on the face dataset. This methodology offers a quick and effective solution to facial recognition.

A possible path for further development could involve incorporating different forms of input, such as audio and text, into facial recognition systems to enable multi-modal capabilities. This integration has the potential to broaden the range of applications for facial recognition, including security, healthcare, and attendance systems utilizing surveillance cameras.

#### REFERENCES

- [1] M. Drożdż and T. Kryjak, "FPGA implementation of multi-scale face detection using HOG features and SVM classifier," *Image Process. Commun.*, vol. 21, no. 3, pp. 27–44, Sep. 2016.
- [2] F. M. Javed Mehedi Shamrat, Md. Abu Raihan, A.K.M. Sazzadur Rahman, Imran Mahmud, Rozina Akter, "An Analysis on Breast Disease Prediction Using Machine Learning Approaches" *International Journal of Scientific & Technology Research*, Volume 9, Issue 02, February 2020, ISSN: 2277-8616, pp: 2450-2455.
- [3] P. Ghosh et al., "Efficient Prediction of Cardiovascular Disease Using Machine Learning Algorithms With Relief and LASSO Feature Selection Techniques," in *IEEE Access*, vol. 9, pp. 19304-19326, 2021, doi: 10.1109/ACCESS.2021.3053759.
- [4] A.K.M Sazzadur Rahman, F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Joy Roy, Syed Akhter Hossain "A Comparative Study on Liver Disease Prediction Using Supervised Machine Learning Algorithms" *International Journal of Scientific & Technology Research*, Volume 8, Issue 11, November 2019, ISSN: 2277-8616, pp: 419-422.
- [5] F. M. Javed Mehedi Shamrat, Zarrin Tasnim, Imran Mahmud, Ms. Nusrat Jahan, Naimul Islam Nobel, "Application Of K-Means Clustering Algorithm To Determine The Density Of Demand Of Different Kinds Of Jobs," *International Journal of Scientific & Technology Research*, Volume 9, Issue 02, February 2020, ISSN: 2277-8616, pp: 2550-2557.
- [6] L. Liu, P. Fieguth, G. Zhao, M. Pietikäinen, and D. Hu, "Extended local binary patterns for face recognition," *Inf. Sci. (Ny)*, vol. 358–359, pp. 56–72, Sep. 2016.
- [7] M. Turk and A. Pentland, "Eigenfaces for Recognition," *J. Cogn. Neurosci.*, vol. 3, no. 1, pp. 71–86, Jan. 1991.
- [8] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
- [9] G. Carneiro, Y. Zheng, F. Xing, and L. Yang, "Review of Deep Learning Methods in Mammography, Cardiovascular, and Microscopy Image Analysis," *Springer, Cham*, 2017, pp. 11–32.
- [10] R. Keshari, M. Vatsa, R. Singh, and A. Noore, "Learning Structure and Strength of CNN Filters for Small Sample Size Training," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2018, pp. 9349–9358.
- [11] Htet Aung, Alexander V. Bobkov, Nyan Lin Tun, "Face Detection in Real Time Live Video Using Yolo Algorithm Based on Vgg16 Convolutional Neural Network", *International Conference on Industrial Engineering, Applications and Manufacturing (ICIEAM)*, 2021.
- [12] Matthew D. Zeiler and Rob Fergus, "Visualizing and understanding convolutional networks," arxiv:1311.2901v3 [cs. cv], 2013.
- [13] F.M. Javed Mehedi Shamrat, Md. Masum Billah, Md. Alauddin, Md. Al Jubair, Sovon Chakraborty, Rumesh Ranjan\*, "A Deep Learning Approach for Face Detection using Max Pooling," *Proceedings of the Fifth International Conference on Trends in Electronics and Informatics (ICOEI)*. IEEE Xplore Part Number: CFP21J32-ART; ISBN:978-1-6654-1571-2, 2021
- [14] KH Teoh<sup>2</sup>, RC Ismail<sup>1,2</sup>, SZM Naziri<sup>2</sup>, R Hussin<sup>2</sup>, MNM Isa<sup>2</sup>, and MSSM Basir<sup>3</sup>, "Face Recognition and Identification using Deep Learning Approach," 5th International Conference on Electronic Design (ICED) 2020
- [15] Neena Aloysius, M. Geetha, "A review on deep convolutional neural networks," 2017 International Conference on Communication and Signal Processing (ICCS), 10.1109/ICCS.2017.8286426
- [16] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," arXiv1409.1556, 2014
- [17] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian sun, Deep residual learning for image recognition. In *proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770-778, 2016.
- [18] N. S. Tummala and P. C. Sekhar, "Face recognition using PCA and geometric approach," 2017 International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2017, pp. 562-565, doi: 10.1109/ICCMC.2017.8282529.