INTERNATIONAL JOURNAL OF COMPUTER ENGINEERING AND APPLICATIONS
VOLUME XVI, ISSUE X, Oct. 22, WWW.IJCEA.COM, ISSN 2321-3469

RECOGNIZING FACES IN DISGUISE USING TRANSFER LEARNING

P. Samson AnoshBabu[1], Y.Rajeswari[2],G.Vasudha[3] ,B.Sreenatha Reddy[4],G. Rakesh[5]
[1] Associate Professor, Krishna Chaitanya Institute of Technology & Sciences , Markapur, A.P, India
[2,3,4,5] Scholar, Krishna Chaitanya Institute of Technology &Sciences , Markapur, India

ABSTRACT:

Machine learning uses the technique of face recognition to identify items in a photo or video. Humans can remember other individuals and some items, including animals, plants, living things, and non-living things. This may be accomplished by computers utilizing the Computer Vision field's Machine Learning approach. Additionally, computers are capable of deciphering the faces of individuals in a picture or video. This study suggests putting three well-known Convolutional Neural Network (CNN) Model Architectures to the test to determine which one is best at recognizing a person's face while they are disguised. The "Recognizing Disguised Faces" dataset is used in this study to separate 75 groups of faces, after which it attempts to train and evaluate its models to determine their accuracy in computer recognition. This research is anticipated to advance the machine learning-related algorithm utilised to address the picture classification issue. Utilizing transfer learning in VGG Models significantly improves the experimental outcomes. In this study, face recognition using VGG Models works best when utilizing Image Net weights.

Keywords : Machine Learning, Face Recognition, Convolutional Neural Network, VGG Model

[1] INTRODUCTION

There are various areas and techniques in the field of machine learning that may be used to identify anything, whether it be an item or a live entity (human, animal, event plant). The face of a person is frequently used in biometric identification to identify someone else. Humans are able to distinguish one another from one another because we have memories and brains that can comprehend our thoughts. However, as machines are unable to think for themselves, a subject called machine learning, which was founded by Arthur Samuel [1] has emerged.

The capacity to detect faces improved when a number of methods, including Eigen faces [2], Principal Component Analysis (PCA) [3], and Convolutional Neural Networks (CNN) [4] were introduced in the previous decade.

Transfer learning is a method used in machine learning where a model is created during the first training assignment and then used during the second test. Transfer learning differs from conventional machine learning in that it starts a secondary task using a model that has already been learned [5].

Due to the numerous benefits of employing CNN, such as transfer learning, CNN has been extensively used in a variety of study domains. These include face recognition [4][11], object detection [8], video analysis [9], food detection [10], pedestrian detection [7], object detection [6], object detection [7], and object detection [8] and other studies discussed [18-25]. In this study, we contrast a few well-known Pre-Trained CNN Model Architectures offered by Kera's, an open-source Python neural network toolkit [12]. VGG16, VGG19, ResNet50, ResNet152 v2, InceptionV3, and Inception-ResNet V2 comprise the architecture that we employed. Next, we split the process into two parts: utilising the vector to train the classifier model and assessing its accuracy and cost function.

We anticipated that the best Pre-Trained Architecture model, with the highest degree of accuracy and the lowest cost function in the ideal hyper parameter state, would result from our effort. The "Recognizing Disguised Faces" dataset [13], which consists of 75 images of people's faces covered with disguised objects such bandanas, maskers, face moustaches, false beards, spectacles, etc., is used in this research. Each participant in the dataset typically receives 7-8 disguised photos, with the final two showing their true faces.
Due of the multiple changes that may be created using various disguises, disguised face identification (DFI) is a very difficult subject. In this study, a deep learning architecture is used to identify 14 facial key characteristics before performing disguised face detection. Due to the necessity of large annotated datasets for deep learning architecture training, two annotated face key-point datasets are presented. For each keypoint, the efficacy of the facial keypoint detection framework is demonstrated. Comparing the key-point detection framework to other deep networks also demonstrates its advantages. Comparisons with comparable face disguise classification techniques also show the efficacy of classification performance.

Face recognition is a significant and difficult issue [2, 4]. A wide range of changed physical characteristics, such as wearing a wig, changing one's hairdo or hair colour, donning eyeglasses, shaving or growing a beard, etc., can substantially mask one's identity [3]. Right et al[9] .'s conclusion was that purposeful face modifications, such as changes in clothing and haircuts, had a negative impact on face recognition performance; by adding wigs and eyeglasses.

Analyzing the structure of the face using facial key points is crucial for identifying the face. This problem has only been attempted to be solved a handful of times in the past. To better identify disguised face patches, Teja's et al. [3] suggested localized feature descriptors and used this knowledge to face identification performance. Classification of masked faces was based on texture-based characteristics. Facial key-point technology has lately grown in favour for uses including facial emotion categorization, facial alignment, tracking faces in movies, and other uses [13]. There have been many prior attempts to accomplish this aim, but the two primary cutting-edge approaches remain. The first kind uses Gabor features with texture- and shape-based feature extraction techniques to identify various face key-points [12]. The probabilistic graphical models are used in the second class of techniques.

In order to identify face key-points[6] captures the link between pixels and features. Deep networks are used for facial key-point recognition because to their better performance in a variety of computer vision applications [11][5]. In order to pinpoint the locations of key-points, Sun et al. [11] devised a three-layer architecture that captures the global high-level properties. Deep Belief Networks (DBN) using background knowledge

Haavisto et al. [5] employed a feed-forward neural network with a linear Gaussian output layer to find face important points.Since there is not enough annotated training data to train deep networks for this application, transfer learning has been used instead, making the usage of deep networks for this application hard. Although transfer learning frequently outperforms other methods, this is not always the

case since there may not be enough training data to fully optimise the pre-trained deep networks.

In this study, a framework for facial key-point detection for masked face recognition is introduced. The framework starts by identifying 14 face important points using a deep convolutional network. The facial points are as follows: Nose region (yellow): P11; Lip region (green): P12- Lip left corner, P13- Lip centre, P14- Lip right corner. Eyes region (cyan): P1- Left eyebrow outer corner, P2- Left eyebrow inner corner, P3- Right eyebrow inner corner, P4- Right eyebrow outer corner, P5- Left eye outer corner, P6- Left eye centre, P7- Left eye inner corner, P8- Right eye inner A few crucial components may be named and regarded as necessary for face recognition [13].A star-net structure is created by connecting the identified locations. A suggested classification framework uses the orientations between the linked points in the star-net structure to do the facial identification. Due to deep convolutional networks' dependency on big training datasets, the project additionally includes two annotated face disguise datasets to enhance training.

The main contributions of the work are stated below:

Disguised Face Identification (DFI) Framework: The proposed framework extracts 14 critical elements from the face that are thought to be crucial for describing the facial structure [13] using a Spatial Fusion deep convolutional network [8]. A star-net structure is created by connecting the extracted points. The suggested categorization framework for face identification makes advantage of the connections between the points' orientations.

Simple and Complex Face Disguise Datasets: A lot of data is needed for the training of the deep convolutional network that detects facial key-points. Transfer learning has been used by academics to identify facial key-points because such datasets are not readily available (small: AR [1] and Yale [7] face databases) [13]. Transfer learning frequently performs well, but it might also perform poorly if there is not enough data to optimise the pre-trained network. To be able to

Examples of photos with various disguises from the Simple and Complex face disguise (FG) databases are shown in the figure. As can be observed from the image, samples from the complex background dataset have a backdrop that is more complex than samples from the simple dataset. We developed two basic and sophisticated Face Disguise (FG) Datasets that researchers might utilise in the future to train deep networks for facial key-point identification in order to overcome the aforementioned problems. On the introduced datasets, face disguise identification is carried out using the suggested framework. For each key-point in both datasets, the average key-point detection accuracy is shown. A thorough comparison of the suggested pipeline and alternative key-point identification techniques is also provided. Finally, a comparison of the classification pipeline's efficacy with cutting-edge face disguise classification techniques is made. The portions of the work are as follows. While Section 3 gives the suggested Disguised Face Identification (DFI) Framework, Section 2 presented the datasets included in the work. The experimental findings are presented in Section 4, and conclusions are made in Section 5. A relatively small number of photos with very few disguise modifications, such as scarves and/or sunglasses, are present in the datasets often utilised for research on disguise (AR [1] and Yale [7] face databases). A vast number of photos with various combinations of disguises, such as persons wearing glasses, a beard, various haircuts, and a scarf or cap, are needed to train deep learning networks. As a result, we offer two face disguise (FG) datasets with 2000 photographs each with I Simple and (ii) Complex backdrops. These datasets include persons wearing a variety of disguises and are photographed against various backgrounds and lighting conditions.2000 photos were captured with male and female volunteers ranging in age from 18 to 30 years old for each suggested dataset (Simple and Complex). The dataset of masked faces was compiled using 25 participants, 8 distinct backdrops, and 10 different masked faces. The following are the disguises included in the dataset: I sunglasses (ii) caps/hats (iii), scarves (iv), beards (v), glasses and caps (vi), glasses and scarves (vii), glasses and beards (viii), caps and scarves (ix), caps and beards (x), and caps, glasses, and scarves.

[2] LITERATURE SURVEY

The game of checkers has been used to study two machine learning techniques in depth. A computer can be taught to learn to play checkers better than the person who built the software, and enough research has been done to support this claim. Furthermore, given only the game's rules, a general sense of direction, and a redundant and incomplete list of parameters thought to be related to the game but whose correct signs and relative weights are unknown and unspecified, it can learn to do this in an astonishingly short amount of time (8 or 10 hours of machine-playing time). These trials confirmed several machine learning principles, which are naturally applicable in a wide variety of different circumstances.

A method for the detection and recognition of human faces is provided, along with a functional, almost real-time face recognition system that monitors a subject's head and identifies them by comparing facial features to those of well-known people. Because faces are often upright, this method addresses face identification as a two-dimensional recognition issue, taking use of the fact that a limited number of 2-D features may accurately characterize faces. Typical perspectives. Face pictures are projected onto a feature space called "face space" that effectively stores the variance among existing face images. The "Eigen faces," or eigenvectors of the set of faces, define the face space; they are not always the same as isolated characteristics like the eyes, ears, and noses. The system enables unsupervised learning of new face recognition skills.

This research develops a novel method for representing images called two-dimensional principal component analysis (2DPCA). Contrary to PCA, 2DPCA bases its computations on 2D image matrices rather than 1D vectors, eliminating the necessity to convert the image matrix into a vector before feature extraction. In its place, an image covariance matrix is built directly from the original picture matrices, and the eigenvectors of this matrix are obtained for the purpose of extracting image features. Three face picture databases—ORL, AR, and Yale face databases—were used in a series of studies to test 2DPCA and assess its effectiveness. Comparing 2DPCA to PCA, the recognition rate was greater across all trials. The experimental findings also showed that utilising 2DPCA, picture feature extraction is computationally more efficient than PCA.

A daring attempt to replace people in the laborious process of filtering internet material has been addressed in recent literature as automated pornographic detection. Unfortunately, the state of the art can produce a lot of false alarms in scenarios with excessive skin exposure, including people tanning and wrestling. This research raises the standard for automated pornography identification with the use of motion information and deep learning architectures on the theory that include motion information in the models can solve the issue of mapping skin exposure to pornographic material. Deep Learning, particularly in the form of Convolutional Neural Networks, has produced impressive results in computer vision, but its promise for pornography detection through the use of motion data has not yet been completely realised.Using optical flow and MPEG motion vectors, we provide new techniques for fusing static (image) and dynamic (motion) data. We demonstrate that while the accuracy of both approaches is comparable, MPEG motion vectors allow for a more effective implementation. On a dataset of 800 difficult test cases, the best proposed technique achieves a classification accuracy of 97.9%, a decrease in error of 64.4% over the state of the art. Finally, we show and talk about the findings from a bigger, more difficult dataset.

In computer vision, object identification is a technique for identifying and recognizing objects in still or moving images. Humans are able to swiftly identify various objects, such as a car, bus, person, cat, food, and other visual artefacts, when they see images or movies. But how do we use it with computers? To recognize one thing from another in an image or video, a computer can utilize classification, a technique or approach in object recognition. This project's author suggests evaluating several widely-used image binary classification algorithms, along with each algorithm's performance matrix results. These include the Logistic Regression with Perceptron, Multi-Layer Perceptron (MLP), Deep Multi-Layer Perceptron, and Convolutional Neural Network (Conv Net). It will be helpful to anyone who needs to identify a food object using auto recognizing tools because the author uses the Food-5K dataset to differentiate between two classes of objects, namely food and non-food, and then tries to train and test how accurate the computer is in recognizing food and non-food objects. It is anticipated that this effort will advance the field of

computer vision. algorithm that is employed to address the issue of picture classification, with a validation accuracy level of above 90% and the state of optimal hyper parameter The test findings show that Conv Net has a considerable advantage over the general artificial neural network for the picture classification issue, with a level of testing accuracy using Conv Net reaching above 90% and a loss function less than 25%.

[3] SYSTEM ARCHITECTURE

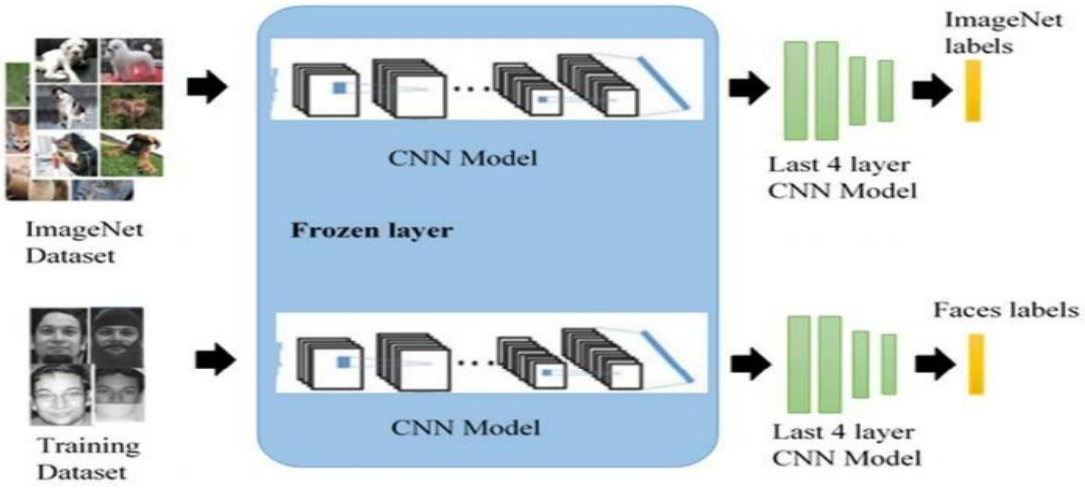


Fig.1 Training Model for Transfer Learning

[4] IMPLEMENTATION

MODULESDESCRIPTION

i) User: The user may gather the photographs by browsing the web. Training, testing, and validation photographs are provided in the data folder. These are divided into three categories: disguise, mask, and scarf. Now, the user must build the Vgg16 and Vgg19 objects and produce the call.h5 file with the model weights. This h5 file contains all of our models' weights. The los function, loss accuracy, accuracy, and test accuracy will be shown in the graph along with the execution time at the moment a h5 file is generated. A system camera or video file can be used to verify object detection by the user. This user needed a high configuration system to be processed. With the aid of the created model weights files, the user may test photos.

ii) **VGG1:** The VGG group at Oxford is the source of the Vgg16 architecture. By substituting certain 3x3 kernel filters for bigger kernel filters (11 and 5 in the first and second convolutional layers, respectively), VGG was designed to improve upon the Alex Net architecture. Small-sized kernels that are stacked are preferable to large-sized kernels for a given receptive field because they improve the depth of the network and enable the learning of more complicated features.

iii) **ResNet50:** According to what has been said so far, the network must deepen the layer in order to boost accuracy, as long as over-fitting is possible. However, just adding layers will not increase the depth of the network. Due to the issue of vanishing gradients, when gradients are repeatedly replicated to the preceding layer, the gradient can become very tiny and making deep networks challenging to implement. When a result, as the network expands, performance is soon saturated or even starts to decline. ResNet's (Residual Network) key concept is the introduction of a "identity shortcut link" that traverses one or more layers.

iv) **InceptionV3**:: VGG Used A GPU To Reach Exceptional Accuracy In The Imagenet Dataset, But Its Application Still Takes A Lot Of Work (Graphic Processing Unit). The Wide Convolutional Layer That Was Utilised Has Made This Ineffective. Googlenet Is Based On The Premise That The Connection Between Most Activations In Deep Networks Makes Them Either Unnecessary (Zero Value) Or

Excessive. Therefore, The 512 Output Channels Won't Be Connected To One Another In The Most Effective Deep Network Design Since Connections Between Activations Would Be Scarce. A Module Created By Googlenet Named Inception Had A Sturdy Architecture And Numbered Somewhat Like A Thin CNN.The Width/Number Of Convolutional Filters Of The Kernel Size Is Kept Modest Because, As Was Previously Indicated, Only A Tiny Portion Of The Neurons Are Functional. Convolutions Of Various Sizes Are Also Used In This Module To Capture Features At Various Scales.

4.1 Sample Screenshots

Fig. 2 VGG16 Confusion Matrix

Fig.3 VGG16 Accuracy

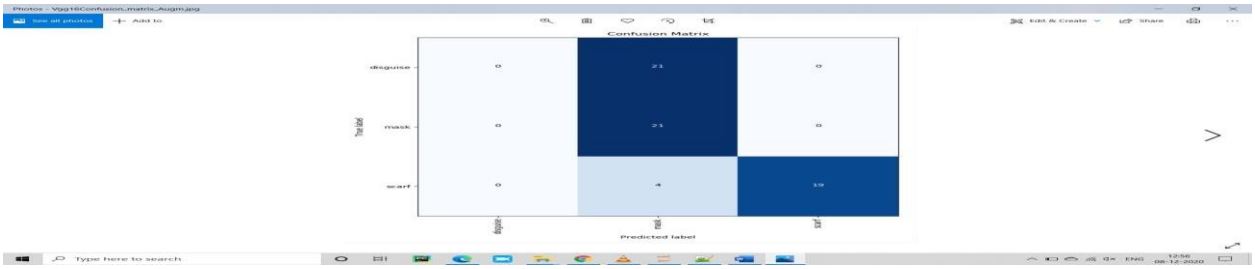


Fig.4 VGG16AugmConfusionmatrix

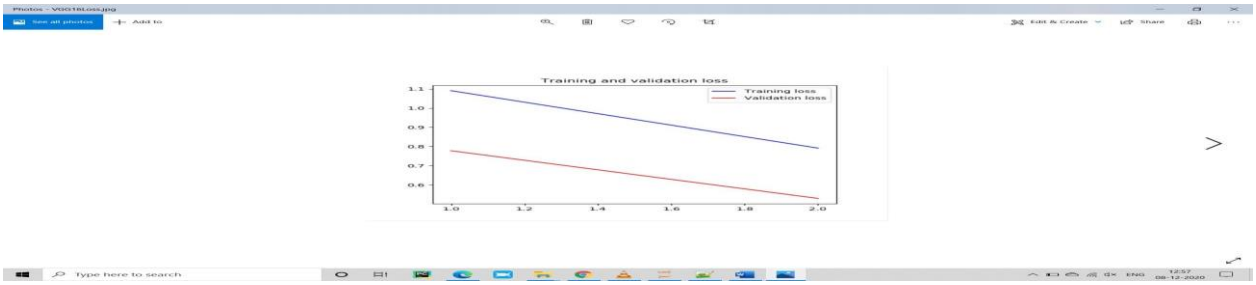


Fig.5 VGG16 Loss Accuracy

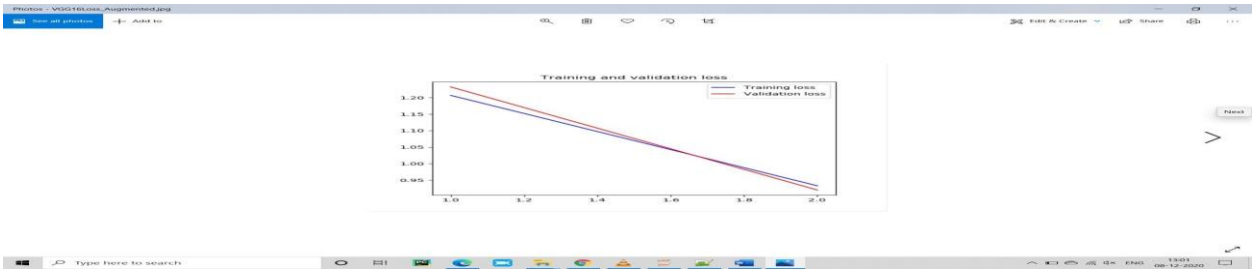


Fig.6 Training Loss Accuracy

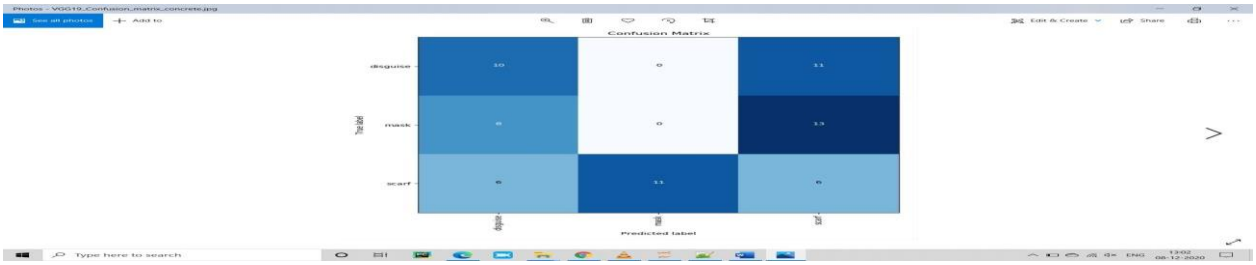


Fig.7 VGG19 Accuracy

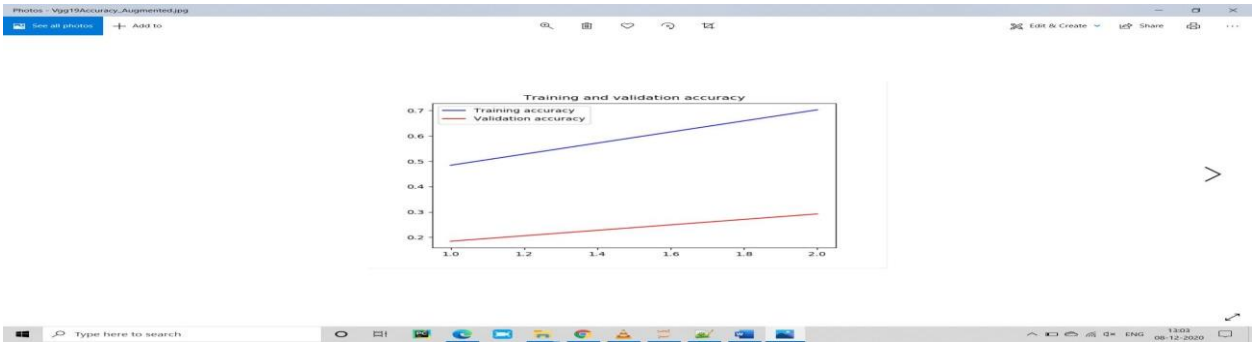


Fig. 8 VGG19TrainingandLossValidation

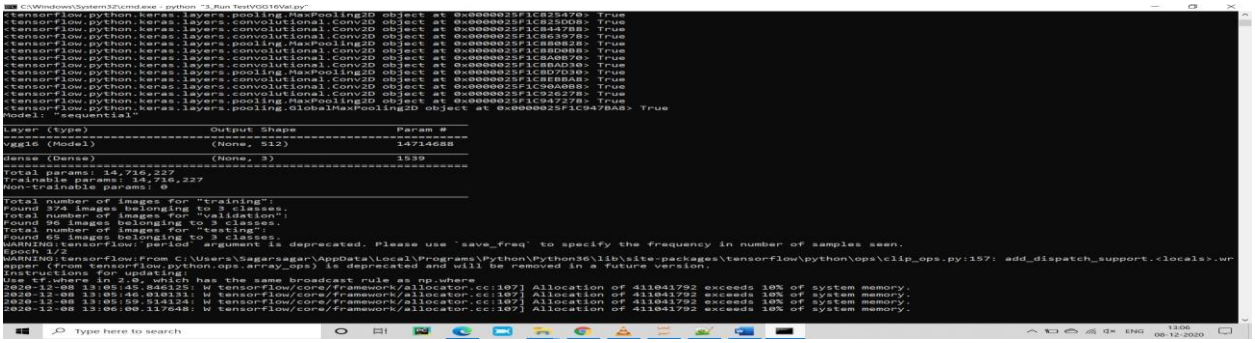


Fig.9 PreVGG16ModelLoaded

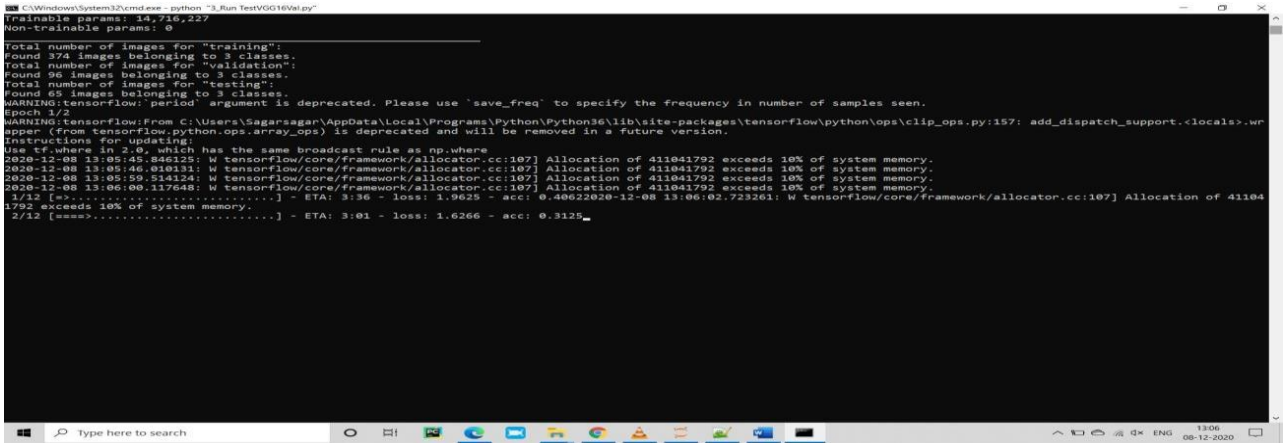


Fig. 10 Epoch started Execution

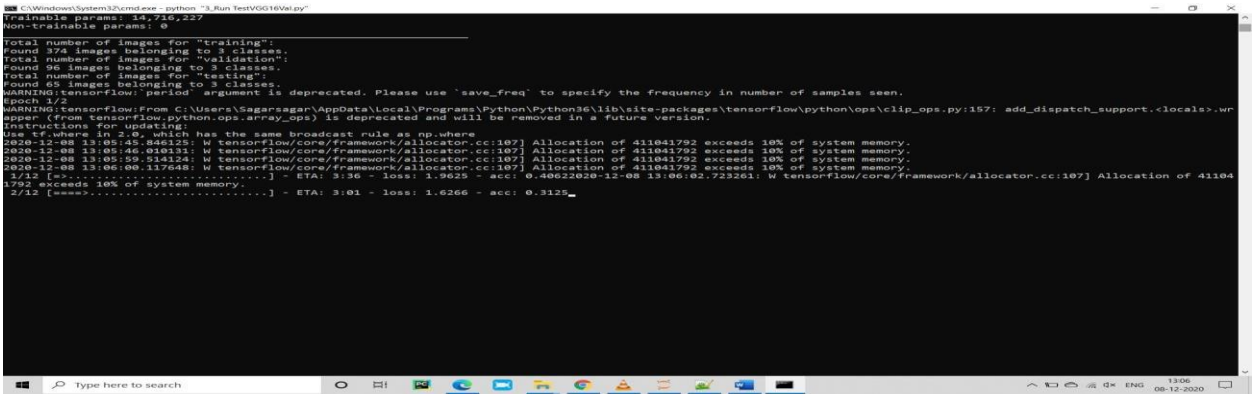


Fig.11 Epoch started Execution

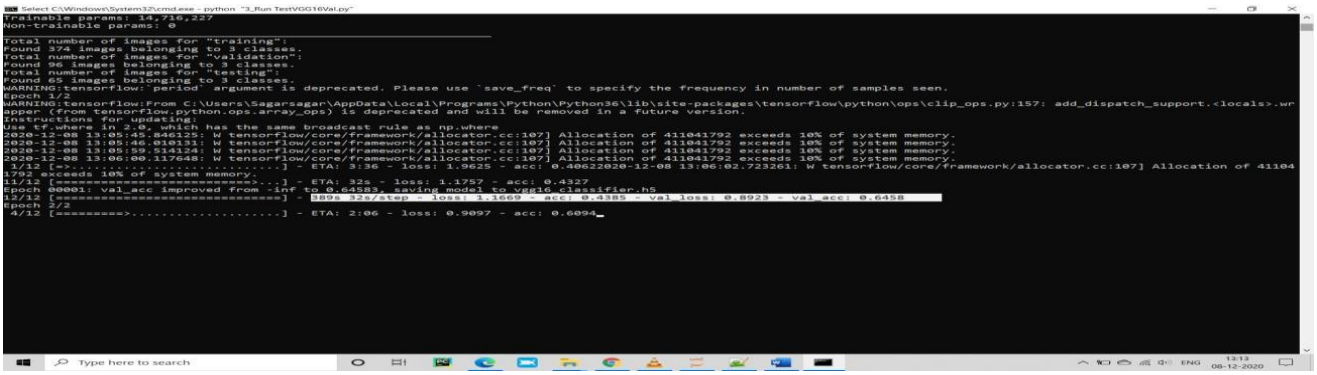


Fig. 12 Los fun, loss accuracy

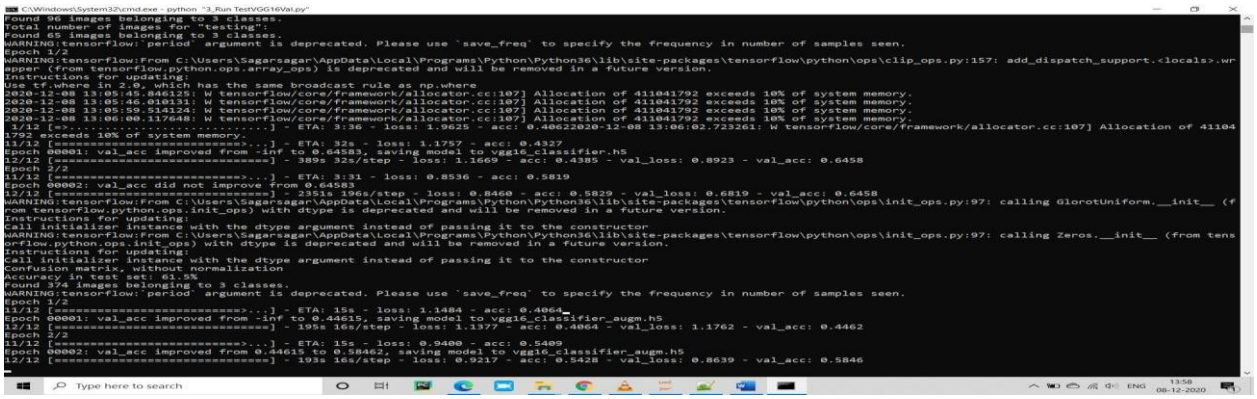


Fig. 13 Validation Accuracy

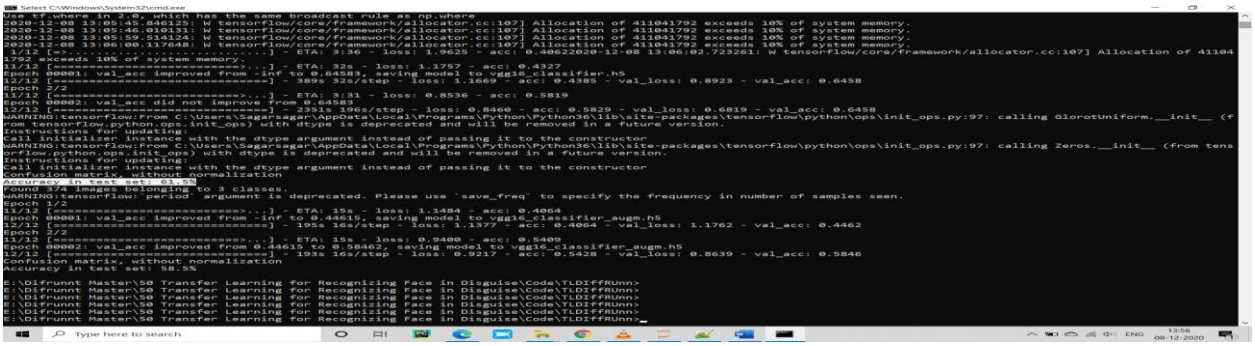


Fig. 14 Test Accuracy

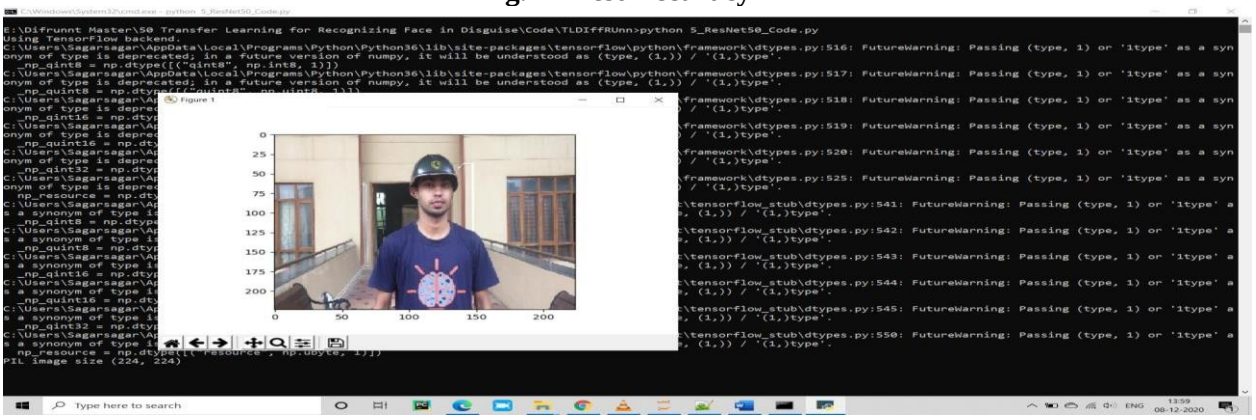


Fig. 15 ResNet50 Loading

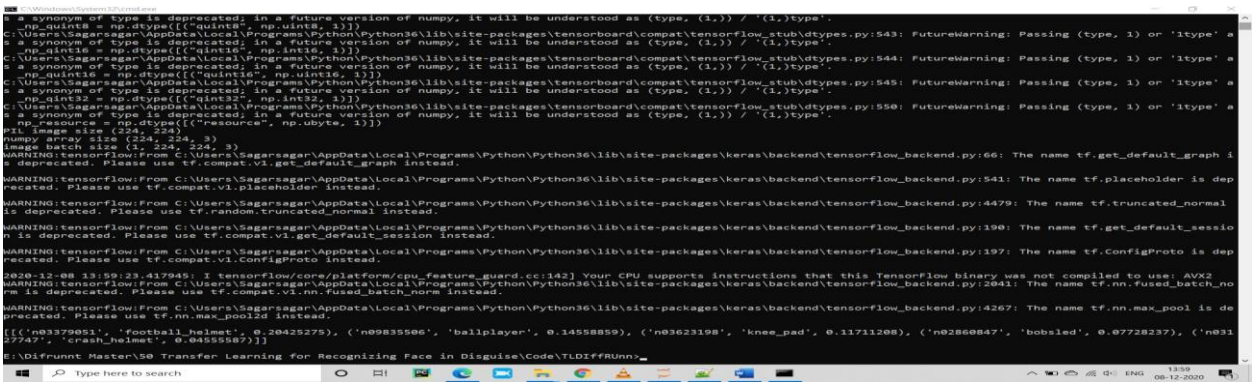


Fig.16 Prediction Results

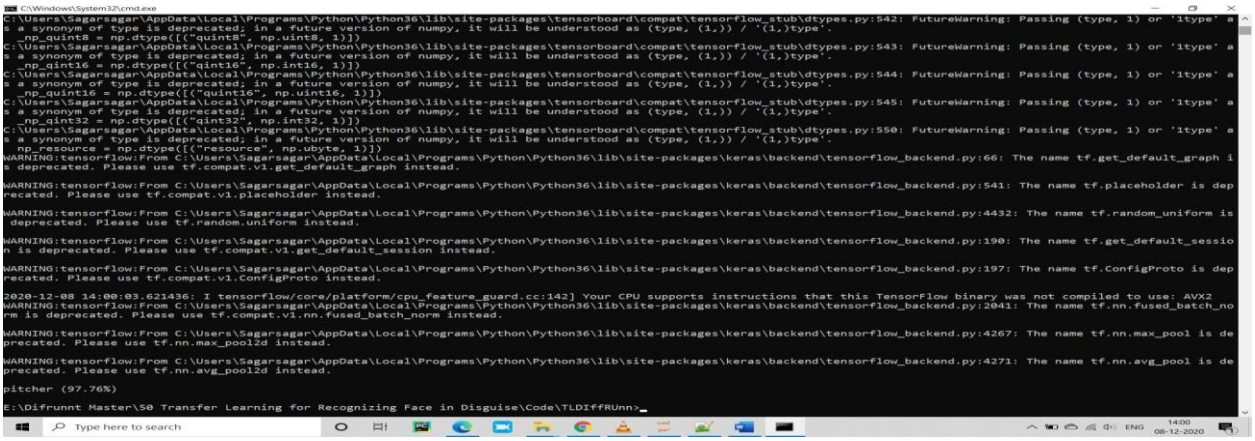
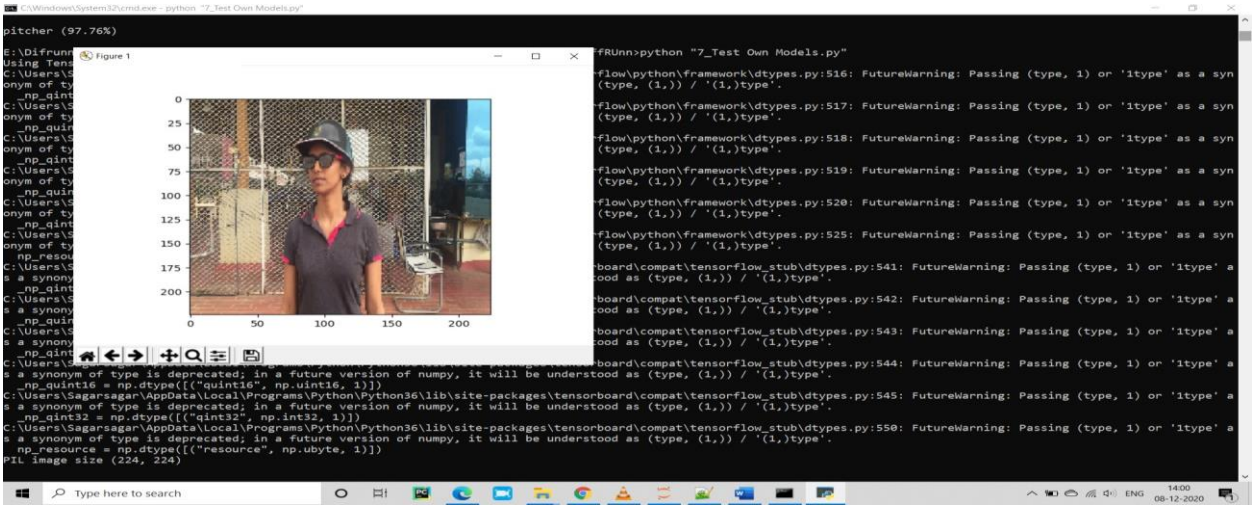


Fig. 17 Inception Loading



[5] CONCLUSION

In this study, we compare six well-known CNN models for identifying disguised faces using "Recognizing Disguised Faces" datasets. The results show how Transfer Learning may be used to the Face Verification issue. The ResNet152 v2 Model has a higher accuracy than the VGG model in the train set, whereas the training results for the VGG model indicate a balance between training and validation accuracy. However, the test results indicate that the VGG model performs better than other CNN Models. Finally, we draw the inference that Image Net weight may be applied to transfer learning for face recognition using a VGG model. Convolution neural network success is also the primary factor driving the popularity of deep learning CNN in recent years. We intend to encode and include the idea of familiarity in automated algorithms as a future research topic, which might enhance performance. Furthermore, we think that research into how masking particular facial features affects depictions of faces may help find more effective ways to lessen these differences.

REFERENCES

[1] A. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," IBM Journal of Research andDevelopment,vol.3, no.3,p.210–229,1959.
[2] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," IEEE Conference on Computer Visionand PatternRecognition, vol.1, p.pp.586–591, 1991.
[3] J. Yang, D. Zhang, A. F. Frangi and J. Yang, "Two dimensional PCA: a new approach to appearance-basedface representation and recognition," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26,no. 1,p.131–137, 2004.
[4] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, "Deepface: Closing the gap to human-level performance infaceverification," CVPR, p.1701–1708, 2014.
[5] J. West, D. Ventura and S. Warnick, "A Theoretical Foundation forInductive Transfer," Spring ResearchPresentation,2007.
[6] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S.Guadarrama and a. T. Darrell, "Caffe:Convolutionalarchitecture forfastfeatureembedding," inACMMM, 2014.
[7] D. Tomé, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi and S.Tubaro, "Deep Convolutional NeuralNetworksforpedestriandetection,"SignalProcessing:ImageCommunication,vol.47,pp.482-489,2016.
[8] Z. Deng, H. Sun, S. Zhou, J. Zhao and H. Z. Lin Lei, "Multi-scaleobject detection in remote sensing imagerywith convolutional neuralnetworks," ISPRS Journal of Photogrammetry and Remote Sensing,vol. 145, no. PartA,pp. 3-22, 2018.
[9] M. Perez, S. Avila, D. Moreira, D. Moraes, V. Testoni, E. Valle, S.Goldenstein and A. Rocha, "Video pornography detection through deep learning techniques and motion information,"Neurocomputing,vol.230,pp.279-293,2017.

[10] R.D.YogaswaraandA.D.Wibawa,"ComparisonofSupervisedLearningImageClassificationAlgorithmsforFoodandNon-FoodObjects," in CENIM, Surabaya,2018.
[11] Z.YangandR.Nevatia,"Amulti-scalecascadefullyconvolutionalnetworkfacedetector,"inInternationalConferenceonPatternRecognition(ICPR), Mexico, 2016.
[12] F.Chollet, Keras, GitHub: GitHubrepository,2015.
[13] T.I.Dhamecha,A.Nigam,R.SinghandM.Vatsa,"Disguisedetectionandfacerecognitioninvisibleandthermalspectrums,"inIEEE International Conferenceon Biometrics, pp. 1–8,2013.
[14] Q.V.Le,J.Ngiam,Z.Chen,D.Chia,P.KohandA.Y.Ng.,"TiledConvolutionalNeuralNetworks,"inNeuralInformationProcessingSystemsFoundation,2010.
[15] A.Krizhevsky,I.SutskeverandG.E.Hinton,"ImageNetClassificationwithDeepConvolutional,"2012.
[16] C.Szegedy,W.Liu,Y.Jia,P.SermanetandS.Reed,"Goingdeeperwithconvolutions,"in2015IEEEConferenceon Computer Vision and Pattern Recognition(CVPR), Boston, MA, USA, 2015.
[17] K.He,X.Zhang,S. RenandJ.Sun,"DeepResidualLearningforImageRecognition,"inIEEEConferenceon Computer Vision and Pattern Recognition(CVPR),LasVegas, NV,USA,2016.
[18] Ayesha Mariyam, SK Althaf Hussain Basha,, S Viswanadha Raju, “A Literature Survey On Recurrent Attention Learning For Text Classification”, 2nd International Conference on Machine Learning, Security and Cloud Computing (ICMLSC2020), Springer Conference 18th & 19th December2020.
[19] SK Althaf Hussain Basha, Ayesha Mariyam, and S Vishwanadha Raju "Applications of Multi- Label Classification", International Journal of Innovative Technology and Exploring Engineering(IJITEE),pp.86-89,ISSN:2278-3075,Volume-9, Issue-4S2, March 2020.
[20] Ayesha Mariyam, SK Althaf Hussain Basha Sk, and Viswanadha Raju S , “A Brief Literature Survey on Text Classification Applications”, International Conference on Devices, Intelligent Systems & Communications (DISC)2020.
[21] Venkata Pavan Kumar Savala, Sk Althaf Hussain Basha, Ranganath P, P V Ravi Kumar, “Information Inclusion: The Modern Rank and The Approach Forward”, International Journal of Computer Engineering and Applications(IJCEA), Volume 13, Issue 6 , December. 20, ISSN2321-3469.
[22] Sreedhar Jinka, Sk. Althaf Hussain Basha, Suresh Dara, Baijnath Kaushik, “Sequence Labelling for Three Word Disambiguation in Telugu Language Sentences”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT) Volume 2 , Issue 7 , pp.311-315, 2017, ISSN :2456-3307.
[23] Baijnath Kaushik, Sk. Althaf Hussain Basha, Sreedhar Jinka, D Praveen Kumar, “ Sequence Labelling for Two Word Disambiguation in Telugu Language Sentences”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT) Volume 2 , Issue 7 , pp.321-327, 2017, ISSN :2456-3307.
[24] Sreedhar Jinka, Sk. Althaf Hussain Basha, Baijnath Kaushik, D. Praveen Kumar, A. Jagan, “ Empirical Analysis of Context Sensitive Grammars and Parse Trees for Disambiguiting Telugu Language Sentences”, International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT) Volume 2 , Issue 7 , pp.328-331, 2017, ISSN :2456-3307.
[25] GNR Prasad , SK Althaf Hussain Basha, Mallikharjuna Rao K M GnanaVardhan “A Review of Predictive And Descriptive Data Mining Techniques In Higher Education Domain, International Journal of Computer Engineering and Applications(IJCEA),Volume 13, Issue 6, January. 21, ISSN2321-3469.
[26] Dr. G. N. R. PRASAD, “Identification of Bloom’s Taxonomy level for the given Question paper using NLP Tokenization technique”, Turkish Journal of Computer and Mathematics Education, Vol.12 No.13 (2021), 1872-1875.