

Leveraging Transfer Learning for Face Recognition in Disguise

¹Dr. Abdul Rasool MD, ²Mohammed Wajahath Ali Zafar, ³Isha Mirza, ⁴Furquan Mohammed Saleem

¹Associate Professor, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

^{2,3,4}B.E Student, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

rasool.501@gmail.com, wajahata785@gmail.com, isha.work30@gmail.com, mfurquan.sal@gmail.com

Abstract: Facial recognition is a device getting to know method for spotting gadgets in a photo or video. Humans have a memory that lets in them to recognize different humans and certain topics which include animals, flowers, residing things and non-dwelling things. However, how can a laptop try this even if it has memory? Machine studying is a manner or pc vision machine that may be used in order that laptop structures can apprehend one individual's face for another person who is in a photo or video. In this newsletter, the author counselled checking out some famous convolution neural community (CNN) version architectures to look which one is higher at recognizing the disguised human face dataset. The writer uses the "Recognizing Disguised Faces" dataset to distinguish 75 lessons of faces, after which attempts to educate and test whether it is able to be identified by the device, with a purpose to be useful to every person who desires to discover and growth the shape of deep studying. This article is anticipated to make contributions to the field of tool learning associated set of rules this is used to remedy the picture category hassle. The experimental results display that there may be a significant improvement because of the changes of the VGG model. We then conclude that the satisfactory weighting of Image Net is used for face recognition the use of the VGG version.

Keywords- face recognition, transfer learning, deep learning, machine learning

I. INTRODUCTION

In the sector of Machine Learning, there are numerous locations and methods to recognize some aspect, whether or not or now not the detection is inanimate (matters) or residing (people, animals, plant life). Attempt). For biometric authentication, someone's face is regularly used to

discover a few different individual. We human beings can pick out with others due to the truth we've got the memory and thoughts to gadget our thoughts. But the machine cannot do that itself, so the region that creates automated questioning, that is system learning, changed into began with the resource of Arthur Samuel [1].

In the last decade, many techniques were followed to perceive human faces, specifically Eigen faces [2] and most important aspect assessment (PCA) [3], in addition to hand neural community (CNN) [4] which then will increase the potential of the face.

Transfer studying is a way utilized in device studying wherein the primary training project creates a version, then we do a second take a look at using the version from the number one training challenge. Transfer mastering differs from traditional gadget gaining knowledge of as it includes the use of a pre-educated version as a springboard to initialize the second one assignment [5].

Thanks to many advantages of the use of CNN, including transfer gaining knowledge of as an example, CNN has been drastically used in lots of researches. That is photo distribution [6], search [7], view [8], and video

Analysis [9], food seek [10] and face recognition [4], [11].

In this text, we examine a few pre-studying CNN layout fashions developed via Keras, an open deliver neural network library written in Python [12]. The structure we use is VGG16, VGG19, ResNet50, ResNet152 v2, InceptionV3 and Inception-Res Net V2. Then we divide it into additives: use vector to expose the

category model and observe the accuracy and performance price of the kind model.

From this research, we want to discover the superb pre-schooling version with the best diploma of accuracy and the lowest price in the hyper parameter consensus kingdom. The creator uses the "Recognizing Disguised Faces" dataset [13], that is a database of seventy five pixy of human faces the use of disguised gear which includes bandana, mask, fake beard , fake beard, glasses, and lots of others. Each character within the database typically gets 7-eight hidden pixy and the final 2 are their real faces.

II PROPOSED METHODS

A. Artificial Neural Network (ANN)

The artificial neural community (ANN) is a machine that performs many functions similar to the neural networks in our mind. McCulloch & Pitts first advanced the ANN in 1943. The ANN become evolved as a mathematical model of the growth of neurobiology or human cognition, and then primarily based on this speculation:

- Neurons are a place wherein information is processed.
- ◆ each link has a weight that signal transmission flow.

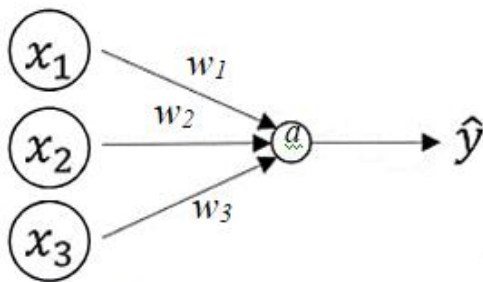


Fig. 1. Single neuron network.

◆ Connections between neurons referred to as synapses that transmit indicators.

◆ to decide the output of each neuron, it uses a

Make a feature that is frequently non-linear at the enter of its network.

There are many elements of operation in ANN, frequently referred to as neurons, devices, cells or nodes. Each neuron is attached with the aid of a communication hyperlink and interacts with the burden. The weight represents the records so one can be used to remedy the problem, and inside the neural networks which might be widely used, the implementation entails the competition of the class model.

Let's see the easy neural community in Figure 1, this is, the output neurons \hat{y} , it then gets input from the activation function neuron a. And respectively, get the input from the 3 extra capabilities x_1, x_2 and x_3 . Which are marked by way of X_1, X_2 and X_3 for the name in their neurons? Also, the weights connected to X_1, X_2 and X_3 for the activation feature a rare w_1, w_2 and w_3 .

Therefore, the calculation of the output can be described as (1).

$$\hat{y} = a = w_1x_1 + w_2x_2 + w_3x_3 \tag{1}$$

After that, we are able to calculate the loss feature community above. The loss function is a measure of the difference between the prediction of $y^{(i)}$, and the actual fee (floor reality), in other phrases, it's far an mistakes calculation for one level of training. This feature may be visible (2).

$$L(y^{(i)}, y^i) = \frac{1}{2} (y^{(i)} - y^i)^2 \tag{2}$$

In this research, we use express move-entropy as a feature, because we need to classify everybody in line with their face. This characteristic will evaluate the distribution of face estimates, from actual to fake, which is about to at least one for proper and zero for fake. The actual class of the character's face is represented by way of a temperature-encoded vector, which we lessen if the output vector model is near the real class. Unemployment is as follows:

$$L(X_i, Y_i) = - \sum_{j=1}^c y_{ij} * \log(p_{ij}) \tag{3}$$

$$y_{ij} = \begin{cases} 1, & \text{if } i_{th} \text{ element is in class } j \\ 0, & \text{if } i_{th} \text{ element is not in class } j \end{cases}$$

B. Convolution Neural Network (Conv Net)

Therefore, the most green deep community structure may have one operating network connection, that means that the 512 outputs will don't have any connection to each other [16]. Google Net has advanced a module called Inception module which includes the identical CNN structure with production. Since only a small portion of neurons are used as mentioned earlier than, the width/range of kernel-sized filters stays small. This version additionally makes use of multi-length convolution to seize content material at a couple of scales (5x5, 3x3, and 1x1).

4) Res Net

As mentioned up to now, for instance, to improve the accuracy of the community, one ought to growth the intensity of the layer, furnished that it can be better. However, increasing network depth does not just upload layers. Deep blending is difficult to exercise because of the problem of disappearing gradients, wherein gradients are repeated at the previous layer, the repetition could make the gradient much less very seen. As a end result, whilst the community grows, performance turns into saturated or maybe starts to degrade speedy.

Created in 2015, the principle idea of Res Net (Residual Network) is to show so-referred to as “non-public connections”

that cross one or greater layers [17], as shown in Figure five.

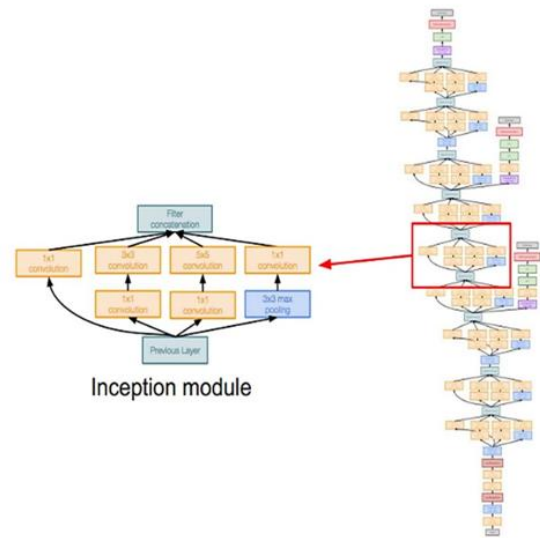


Fig. 4. Network architecture of Google Net/Inception.



Fig. 5. Example of used dataset.

III EXPERIMENT AND RESULTS

In this text, the author will assessment 75 instructions of photograph kind the use of Convolution Neural Network (Conv Net), we can differentiate the model with a specific technique called in advance that. The writer plays the take a look at the use of a laptop with particular instructions:

- Dual processor with Intel(R) Xeon(R) processor at 2.30 GHz.
- thirteen GB reminiscence
- Tesla K80 GPU, PCI card ID: 0000:00:04.Zero, compute capability:

three.7.

◆ Keras model 2.3.1

The creator uses the "Recognizing Disguised Faces" dataset; it truly is a database of seventy five pictures of human faces using disguised tools which includes bandana, mask, fake beard, moustache faux, glasses, and many others. We will divide the validation based totally at the character's actual face and the assessment or check statistics we use is the same as the validation facts. In this paper, we only use the greyscale photo format of the schooling device.

A. Training and trying out

We have used an easy format model to layer and add access to the previous gaining knowledge of models we used in these studies, which incorporates: VGG16, VGG19, ResNet50, ResNet152 v2, InceptionV3 and Inception – Res Net V2. Then we use the pre-weighted "Image Net" for our perceptual model which may be used for rework studying. We used VGG16 in our first check and did just like our comments version.

In each configuration, we use 30 times for every training model. And all models after thawing 4 layers pass via specific adjustments, besides the InceptionV3 model, that is shown in Table I.

B. Configuration method

Before education the model, we use a data

pre-processing method to avoid version over fitting. The device consists of resizing all images to 224 x 224 pixels, flipping and rotating them. The ensuing photo from pre-processing can be visible in Figure 7.

We use 2 configurations for education, that's freezing all layers of the CNN version (configuration 1) and unfreezing the remaining 4 layers (configuration 2). Freezing each layer technique we use the weights from the previous training for the CNN version. In our case, we trained Keras on the Image Net dataset which incorporates many labels. Configuration 2 approach that we freeze all the layers however now not the closing four layers, we educate the closing four layers to get an super accuracy of the records we use, this manner is regularly referred to as trade getting to know, which corresponds to the version tested in Figure 8. The high-quality consequences for this training are confirmed in Figure nine.

C. Leave all the layers inside the CNN version (configuration 1)

In this education setup, we want to realise the number one impact of applying the weighting earlier that we use at the CNN version, that's Image Net. From training (or change in Keras), we get the outcomes in Table II.

D. Unfreeze the closing four layers of the CNN version (configuration 2)

In this training technique, we first freeze all of the layers besides the remaining 4 layers in our pre-CNN model; because of this we first-rate train the remaining four layers of the CNN version. We practice this to the complete version, ignoring the layers. Then we healthy our version over again. We received the effects supplied in Table I.



Fig. 6. Output image after preprocessing.

TABLE I. TRAINABLE PARAMETER OF CNN MODEL

CNN Model	Freezing all layer (parameter)	Unfreezing last 4 layer (parameter)
VGG16	8,466,507	15,545,931
VGG19	8,466,507	15,545,931
ResNet50	102,838,347	103,893,067
ResNet152 v2	102,838,347	103,893,067
Inception v3	52,506,699	52,506,699
Inception-ResNet v2	100,741,195	103,937,611

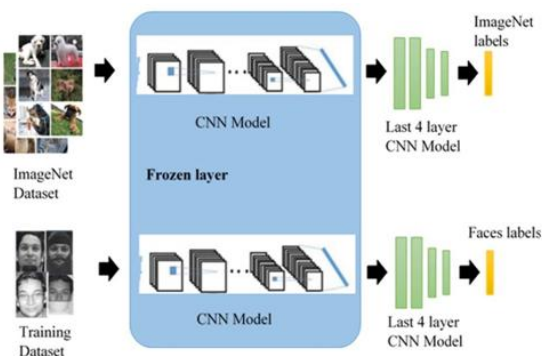


Fig. 7. Training model for transfer learning.

TABLE II. MEASURE PERFORMANCE ON FREEZE ALL LAYER

CNN Model	Loss Value	Accuracy (%)	Validation Loss Value	Validation accuracy (%)
VGG16	3.41	20.49	3.14	45.24
VGG19	3.91	11.93	3.81	26.19
ResNet50	1.74	53.6	4.66	1.6
ResNet152 v2	1.57	61.0	6.16	1.6
Inception v3	3.51	16.3	3.43	3.9
Inception-ResNetv2	4.01	7.3	4.15	1.6

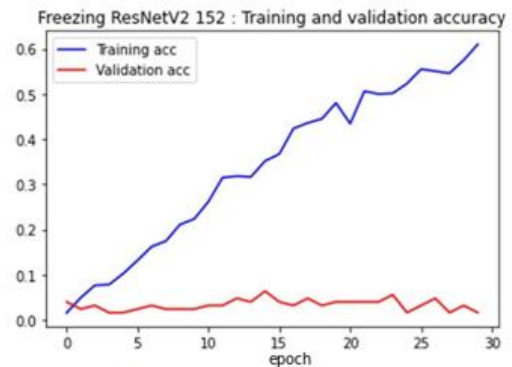


Fig. 8. Accuracy on ResNet152 v2.

In easy terms, we will end that the schooling configuration ultimately offers higher accuracy in some fashions and results in a decrease loss price in particular within the VGG16 model. But there may be a trend inside the second set that is Inception V3. The overall performance difference is shown in Table IV for higher knowledge. Value subtracted from the average array, that's reduced thru the value after the second one configuration is finished, and vice versa.

From Table IV, Figures 10 and 11, we conclude that the VGG16 model can achieve higher effects after acting the second spherical of training and validation. Unlike the Res Net version, this version completed well pleasant in the education

system, no longer inside the validation gadget. Then, in the testing method, we use Keras's prediction command to expect the facial data inside the educate and configure validation. And the end result is nearly the same: training has stepped forward. The results are offered in Table V and Table VI. And Figure 12 is the prediction of VGG16 model with set 2.

TABLE III. DIFFERENCE PERFORMANCE AFTER UNFREEZING LAST 4 LAYER

CNN Model	Performance Setup 2 – Performance Setup 1			
	Loss Value	Accuracy (%)	Validation Loss Value	Validation accuracy (%)
VGG16	-1.24	22.63	-2.16	34.92
VGG19	-1.52	25.99	-2.42	41.28
ResNet50	-0.16	4.4	0.20	0.0
ResNet152 v2	-0.13	1.0	-0.16	3.1
Inception v3	0.05	-3.3	0.11	2.4
Inception-ResNet v2	-0.50	7.7	-0.02	3.9

VGG16 W epoch=30[preproc]size=224 : Training and validation accuracy

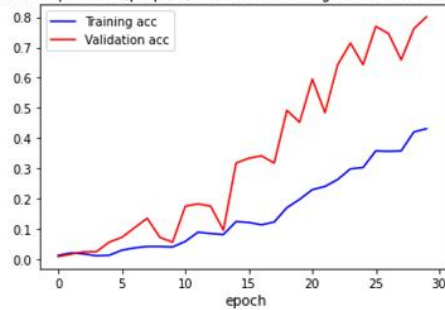


Fig. 9. Loss and accuracy graph on VGG16 model on setup 2.

```

Found 126 images belonging to 75 classes.
Found 654 images belonging to 75 classes.
66/65 [=====] - 4s 67ms/step
13/12 [=====] - 1s 65ms/step
----- VGG16 W epoch=30[preproc]size=224 -----
----- predict on train set -----
Total photos success prediction = 495 of 654
Total photos fail prediction   = 159 of 654
accuracy = 75.688 %
----- predict on validation set -----
Total photos success prediction = 101 of 126
Total photos fail prediction   = 25 of 126
accuracy = 80.159 %
Original label:P10, Prediction :P10, confidence : 0.208 Original label:P19, Prediction :P54, confidence : 0.105
    
```



Fig. 10. Running predict command example on CNN Model VGG16.

IV CONCLUSION

In this article, we provide a evaluation of 6 well-known CNN models for spotting blind humans' faces the usage of "Disguised Face Recognition" datasets, and the outcomes offer an reason behind how the studying machine will alternate. The manner to apply in face evidence trouble. The schooling outcomes display that the VGG model has stability amongst training and validation, and as an alternative, the ResNet152 v2 model has higher accuracy than VGG in education. But the test effects display that the VGG model plays higher than other CNN models. We then conclude that Image Net weights can be used for adaptive getting to know to recognize faces the usage of the VGG version. The fulfilment of this convolution neural community is likewise the precept reason why CNN Deep Learning is a warm subject matter in today's years.

REFERENCES

1. A. Samuel, "Some Studies in Machine Learning Using the Game of Checkers," *IBM Journal of Research and Development*, vol. 3, no. 3, p. 210–229, 1959.
2. M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," *IEEE Conference on Computer Vision and Pattern Recognition*, 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-based face 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, "Deep face: Closing the gap to human-level performance in face verification," *CVPR*, p. 1701–1708, 2014.
5. J. West, D. Ventura and S. Warnick, "A Theoretical Foundation for Inductive Transfer," Spring Research Presentation, 2007.
6. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama and a. T. Darrell, "Caffe: Convolution architecture for fast feature embedding," in *ACM MM*, 2014.
7. D. Tomé, F. Monti, L. Baroffio, L. Bondi, M. Tagliasacchi and S. Tubaro, "Deep Convolution Neural Networks for pedestrian detection," *Signal Processing: Image Communication*, vol. 47, pp. 482-489, 2016.
8. Z. Deng, H. Sun, S. Zhou, J. Zhao and H. Z. Lin Lei, "Multi-scale object detection in remote sensing imagery with convolution neural networks," *ISPRS Journal of Photogrammetric and Remote Sensing*, vol. 145, no. Part A, 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. Yogaswara and A. D. Wibawa, "Comparison of Supervised Learning Image Classification Algorithms for Food and Non-Food Objects," in *CENIM*, Surabaya, 2018.
11. Prasadu Peddi and Dr. Akash Saxena

(2014), "EXPLORING THE IMPACT OF DATA MINING AND MACHINE LEARNING ON STUDENT PERFORMANCE", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.1, Issue 6, page no.314-318, November-2014, Available: <http://www.jetir.org/papers/JETIR1701B47.pdf>

12 Prasadu Peddi and Dr. Akash Saxena (2015), "The Adoption of a Big Data and Extensive Multi-Labeled Gradient Boosting System for Student Activity Analysis", International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 3, Issue 7, pp:68-73.