

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/373204429>

Face recognition using open CV and VGG 16 transfer learning

Conference Paper in AIP Conference Proceedings · January 2023

DOI: 10.1063/5.0157084

CITATION

1

READS

445

6 authors, including:



[Mrs. Sheetal S. Patil](#)

Bharati Vidyapeeth Deemed University

52 PUBLICATIONS 64 CITATIONS

[SEE PROFILE](#)



[Avinash M. Pawar](#)


Bharati Vidyapeeth's College of Engineering For Women Pune

48 PUBLICATIONS 49 CITATIONS

[SEE PROFILE](#)

RESEARCH ARTICLE | AUGUST 17 2023

Face recognition using open CV and VGG 16 transfer learning

Mrunal Bewoor ; Sheetal Patil; Saumya Kushwaha; Sparsh Tandon; Siddharth Trivedi; Avinash Pawar



AIP Conference Proceedings 2890, 020019 (2023)

<https://doi.org/10.1063/5.0157084>



Export
Citation

CrossMark

Articles You May Be Interested In

A functional integral formalism for quantum spin systems

J. Math. Phys. (July 2008)

500 kHz or 8.5 GHz?
And all the ranges in between.

Lock-in Amplifiers for your periodic signal measurements



Find out more



Face Recognition Using Open CV and VGG 16 Transfer Learning

Mrunal Bewoor^{1, a)} Sheetal Patil^{1, b)} Saumya Kushwaha^{1, c)} ,Sparsh Tandon^{1, d)}
Siddharth Trivedi^{1, e)} Avinash Pawar^{2, f)}

¹ Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune, Maharashtra, India

² Bharati Vidyapeeth's College of Engineering for Women, Pune, Maharashtra, India

^{a)} Corresponding author: msbewoor@bvucoep.edu.in

^{b)} sspatil@bvucoep.edu.in

^{c)} saumyakushwaha14@gmail.com

^{d)} snowmansparsh4@gmail.com

^{e)} trivedi.siddharth04@gmail.com

^{f)} avinash.m.pawar@bharativedyapeeth.edu

Abstract- Large organizations have been successfully using facial recognition in security systems which were using Convolutional Neural Networks (CNNs). However, small organizations struggle to implement the same security measures like CNNs using smaller training databases and fewer computer resources. By applying transfer learning for facial recognition, which requires less retraining, The research paper provides a solution to resolve this problem. This can be demonstrated by adding a layer of a trained element with a minimal number of neurons in a network can be performed in tiny datasets, allowing for functional and legitimate authentication.

Keywords - Face recognition, transfer learning, vgg16.

INTRODUCTION

Face is one of the most widely accepted biometrics, and it has become the most common way to detect human use in their visual interactions. Problem with verification systems based on fingerprints, voice, iris and recent genetic structure (DNA fingerprints) data acquisition problem. For example, with fingerprints the person concerned should keep his or her finger in the correct position and direction and when the speaker is known the microphone should be kept in the correct position along with the speaker and distance. However, the process of getting facial images does not interfere with the face being used as a biometric encryption (where the user is unaware that they are being deceived) structure. The face is a normal human trait. Face recognition is important not only because of the power of your many potential applications in the field of research but also because of the power of your solution that can help solve other divisive issues such as object recognition.

In the face recognition system it automatically detects existing faces in photos and videos. It is divided into two modes:

- face verification (or authentication)
- face recognition (or sight)

In verifying a face or proving authenticity there is one similarity and that compares the question of the face image with the face image of the image sought by its ownership. In face recognition or eye contact there is one of the many comparisons that compares the face and question to all the image templates on the website to determine

the identity of the face of the question. Another type of facial recognition involves the checklist check, in which the question face is compared to a list of suspects (one-to-one matching).

As Small institutions are not able to fully utilize AI for face recognition in their security systems and it requires a huge database for training models.

We can solve this problem using **Transfer Learning** with **VGG16**.

In **transfer learning**, the knowledge of an already trained machine learning model is applied to a different but related problem. This will also help to achieve more accuracy with less amount of data.

VGG16 is a convolutional neural network (CNN) model which is a dataset of over 14 million images belonging to 1000 classes. As collected data is minimal to train the model, we use the concept of augmented images.

The Steps involved are:

- Step 1: Collect the dataset
- Step 2: Train the model using VGG16
- Step 3: Test and run the model
- Step 4: Model is ready to use

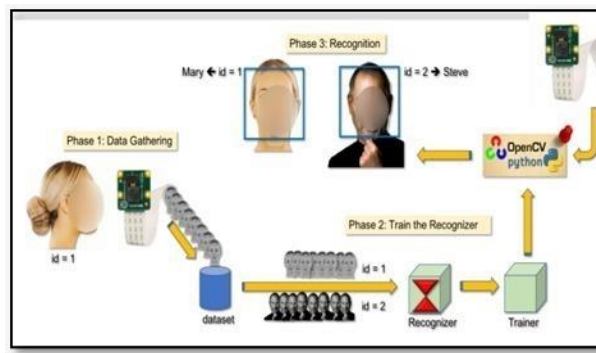


FIGURE 1: Face Recognition Process Flow

BACKGROUND AND RELATED WORK

People have the ability to understand and know new patterns or categories based on their existing knowledge and memories. If we just merely understand the meaning, we can distinguish any new class. If we assume that there is a vehicle with two wheels up front, one handle, and a balance. The two categories of divided jobs in mini databases are identical to those in large databases: single, unstructured reading. We can calculate the test model section for a single training model in a one-shot analysis. In contrast, the zero-shot reading model doesn't search for any training models in the necessary category.

The Siamese network was utilised by Koch et al. [7] to handle single-reading Omni-Glot data. 50 different languages are handwritten in the Omni-Glot database. They train Siamese to examine the similarities between images in pairs and to offer potential indicators of how closely the two images resemble one another. The test photos are then tested against the test image using the model, one against each class in the novel. The matches with the highest possibility of landing a single job are those with the highest scores determined by the verification network.

A potent all-encompassing definition of nature can be provided by the convectional neural network. According to Razavian et al. [8], information gleaned from extensive networks can be used as a right of appointment for visual functions. They are thought to be the accepted representation of various object integration functions, descriptive descriptions, and visual scenarios since they include features that were obtained through the OverFeat network. To obtain various viewing functions, the SVM separator or simple line split uses the 4096x1 feature presentation. Additionally, they indicated that a variety of visual functions may be applied with the features that were provided using OverFeat trained on the ImageNet database.

DATASET DESCRIPTION

Data from various sources, such as the AT&T Laboratories Cambridge, the AR Face Database from Ohio, the Facial Database from Group Essex Ideas, the Cohn Kanade AU-Coded Facial Expression Database (FE), the Verified Multi-Modal Verification for Teleservices and Security Services (XM2VTS) Database, the Japan Female Facial Expression (JAFPE) Database, etc. have been used to test facial recognition abilities over time. There aren't many conclusions that can be drawn from our research because it's mostly concerned with figuring out how accurate facial data is, which is a smaller company's biometrics.

In our experiment, six face-to-face data transfer sets were used, including the AT&T database(A), Essex 94(B), Essex 95(C), Essex 96(D), Essex Grimace(E), and Georgia Tech(f) database. Each database only contains a small number of photos between 15 and 110 titles total. Photos are scaled down to 299x299 and 224x224 pixels, which can be used in Inception V3 or the VGG face model, respectively, to extract the feature. If the images are colorless, they are replicated to three channels.

In this experiment we used 40 pictures of each member to train our model, of which 90% was used in training and 10% for testing. After identification of face, the model displays the corresponding name of the person.

FACE RECOGNITION APPROACHES

Facial recognition has drawn researchers from a variety of fields, including psychology, computer vision, neural networks, face patterns, and face recognition. Even though it is a difficult procedure, it is intriguing. Several face recognition techniques include

1. Holistic matching
2. Feature-based matching
3. Hybrid matching

1. Holistic matching methods

This method uses the entire face as a raw input for a facial recognition system. The holistic matching strategy can be divided into linear and non-linear projection methods. Examples of the linear projection appearance-based methodology include principal component analysis (PCA), independent component analysis (ICA), linear discriminate analysis (LDA), and linear regression classifier (LRC). Non-linear projection appearance-based techniques include kernel principal component analysis, kernel linear discriminate analysis, and locally linear embedding. The input image is translated into a higher-dimensional space using a nonlinear technique, and the face is then linearly and simply represented. So, conventional techniques Gray levels are used to capture a number of images for principal component analysis.

2. Feature based method

The eyes, nose, and mouth are located in this procedure, and features are extracted and sent into the structural classifier. Face restoration via feature-based methods is quite difficult. The system is unable to retrieve features because of significant variations. The various extraction techniques can be divided into

- Structural matching methods
- basic techniques based on the eyes, ears, and nose
- Feature-template based method

3. Hybrid method

It combines a feature-based matching method and a holistic method. This technique makes use of 3-D photographs. This enables the device to record features like cheek, nose, and eye curvature. This technique involves identifying the face by scanning or a real-time snapshot. The head's position, angle, and size are determined. After

measuring each curve and making a template, the area around the eye, inside the eye, and on the nose are concentrated. The template is then converted into code. By comparing it to the input image, this code is used to do recognition after being stored in the database. [9]

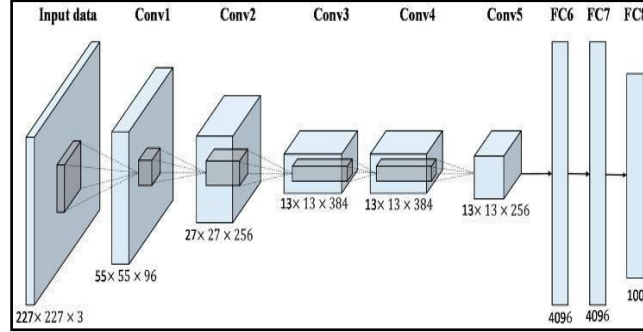


FIGURE 2: VGG - 16 Architecture

FEATURE EXTRACTION

A well-trained transmission equipment model that is suitable for the job is needed during the learning curve. We are unable to train any convolutional neural network from scratch in relatively tiny databases. As a result, we use pre-trained models to obtain information which we then feed into our closely packed portions..

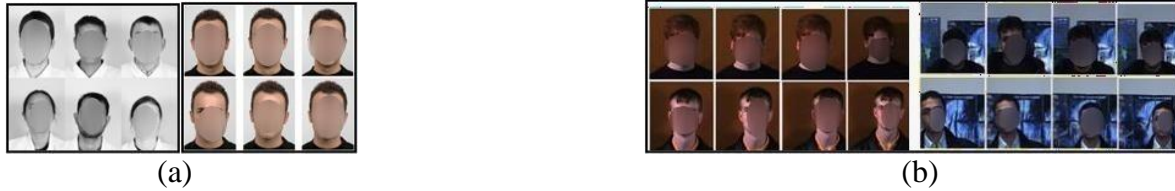


FIGURE 3: Two subject images in database (a) and (b)

The characteristics that the input images shared. The last, dataset-specific layers contain data that is pertinent to classes. Two trained models, one trained on faces and the other on real-world pictures, were used to retrieve properties.

VGG-Face

The VGG-Very-Deep-16 implementation used by CNN CNN's definitions of VGG faces were created using information from and evaluations of the face-labeled field and YouTube face database. Images up to 224x224 in size are supported by VGG Face.

A 16-layer VGG-16 is simply a VGG surface architecture. Table I demonstrates how a characteristic changed during the intermediate stages of the FC. The capabilities of fc7 output deliver good outcomes in our trial.

TABLE 1: Layered VGG-face feature form

Layer Name	Feature Shape
fc6	4094
fc7	4094
fc8	1000

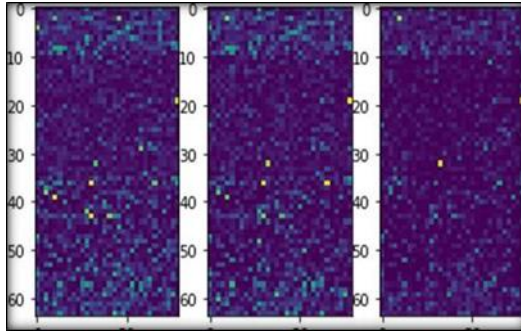
The Inception V3 model was trained using the 14,197,122-image ImageNet visual database, a sizable visual database created for use in research on visual object identification software. This model's standard input size is 299x299 pixels. To extract features, utilise the pool layer 3, which produces output features with the shape 2048x1

Feature Visualization from Extracted Data

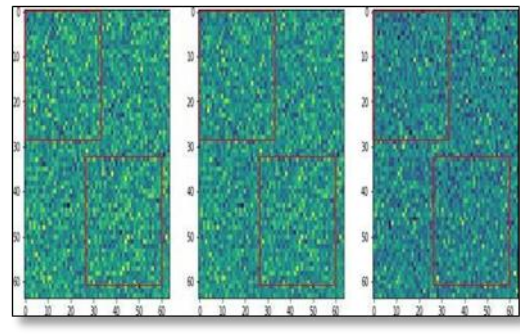
The green-drawn parts in Figure 2 illustrate how the characteristics of the same image differ from those of the opposing image. Figure 2a shows the redone Georgian technical features at VGG-Face from 4096x1 to 64x64, and Figure 2b shows the redone At&T technical featuresat Inception V3 from 2048x1 to 64x32 for display.

TABLE 2: Image Databases Summary

Databases	Resolution	Image type	Number of subjects	Images per subjects
A	92x110	Gray	40	10
B	160x200	RGB	110*	15*
C	160x200	RGB	71*	11*
D	196x196	RGB	149*	18*



(a)



(b)

FIGURE 4: Triplets' extracted features are visualised using (a) VGG-Face and (b) Inception V

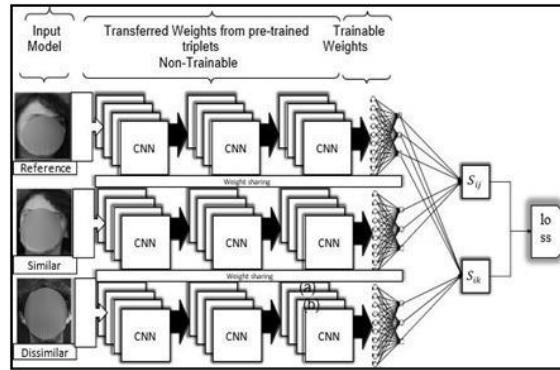


FIGURE 5: Model Configuration

MODEL DESCRIPTION

The weight-sharing structures in our model are triple identical Siamese networks. Convolutional layers that are completely connected to the layers above them make up each structure.

With the previously mentioned pre-trained weights, the initial resolution layers are output as features using layers using VGG-surface and Inception-V3. More information about FC baseline training and loss function is provided in the most current sections. The model's overall structure is displayed.

1. FC Layers

In our approach, one completely integrated hidden layer is added on top of the convolutional layers for each parallel network.

Currently, 10 dormant neurons are used in a single hidden layer for all of our studies. Accuracy falls off if ReLU is used with this discharge of 10 neurons. ReLU reconstruction, however, has demonstrated promise when adding more neurons. But because our database is tiny, we are unable to add more output neurons because doing so would result in an increasing amount of parameters.

1) Determining the number of neurons to add to the FC layer. An N-number of fully linked layers of neurons can be layered. We attempted to change the N and used the aforementioned standards to evaluate the reliability of the AT&T database. Table III illustrates the precise number of neurons:

TABLE 3: Evaluating Accuracy

N	Accuracy
5	97.4
10	98.57
15	99.67
20	99.81

As the number of neurons increases, accuracy rises. However, we must limit it to a lower amount because a bigger N will result in more parameters and over fitting.

$$S_{lm} = \frac{al \cdot am}{|al| \cdot |am|}$$

Figure 6 shows the precision of correctly classified face of different persons.

Number of Face Recognition			
Person 1	Person 2	Person 3	Result
15	1	2	Person 1
2	16	2	Person 2
3	3	16	Person 3

FIGURE 6: Confusion Matrix of Person Images

2. Lack of Function

For each triplet, we have three vectors, each measuring 10 by 1. We used the cosine similarity metric. Sil and Sim are the normalised dot products of the first and third vectors, respectively, and they represent the similarity of images that are not comparable to one another.

We want to teach our network how to learn once the (Sil and Sim) have been built so that it can tell the difference between similar and distinct photos.

And for this, we've used a hinged loss with some margin as a learnable parameter. $\text{Max}(0, (+ \text{Sil Sim}))$ is the loss function. Our network learns to find similarity values for like images that will be higher than the similarity values for unlike images while reducing this loss function.

3. Recognition and Incorporation

We choose two of the three categories of art after network training, and for each image article, we compute the cosine similarity, where the training images or images of verification form columns, and the tests or tests form lines. The connection between kNN and matrix similarity. By building a kNN over this image-generated embedding or by using the SVM separator as a custom kernel, it is possible to identify subjects. For kNN, we will examine column intelligence in each test image and choose the top k photos with the most similarity before assigning the topic based on voting. With the help of the validation set photos, the hyper-parameter k is selected, and for the test images, kNN is carried out using this k. Accuracy of validation can vary.

4. Three-way Formation

Construction of triplets is necessary for the Siamese network's triple training and testing. The three construction processes are different from one another and are explained below. After the N and triplets are ready, we will show the loss and combine the variance with intelligence and precision in the following section. A [Batch size x 3 x Feature shape] triplet is what our model uses to represent three photos. We must therefore construct training and testing three times.

5. Triplet Formation in Training

The first image is chosen at random, as is the second from the same subject, and the third is chosen at random from any remaining topics.

TABLE 4: NUMBER OF POSSIBLE OUTCOMES

Databases	No training triplets Possibilities	No validation triplets Possibilities	Data split ratio per subject
Yale	47,041	13,441	8:2
A	229,319	76,439	7:2
B	7,325,949	1,450,789	11:3
C	730,600	243,529	7:3
D	22,360,731	3,726,776	13:4
E	481,949	68,849	15:3
F	1,102,600	246,000	10:2

Triangular Formation For the triple training and testing, Table IV lists the number of possible threes, the number of photos taken per subject, and the number of photos taken each caption.

6. Validation Triangular Formation

The first image is randomly selected from the training set's test set, the second from the same subject's training set, and the third from any of the photos in the training sets of the other subjects.

RECOGNITION RESULT

The found resemblance The kNN technique is utilised to recognise the faces utilising embedding, as explained in section VI-C. The tasks requiring acknowledgment are finished. The results for determining the beat rate and the

missed rate are much improved when we combine our fraud-based authentication limit with a single validated category (such as an ID card). Regardless of whether the applicant has an ID card for an older subject already registered in the system, any fraudulent face will be properly rejected if it does not exceed the maximum value of the maximum match. He can only pass if he possesses the identity card for the predicted face when his face begins to resemble one of the faces from the earlier studies.

1.Training Phase

Before starting model training, we transmit the full image using a feature extractor, such as going through previously trained work. The training was then completed three times in accordance with clause 2 after that. 4. Completely connected layers receive 10 more neurons. Normal reduction instruments are begun, and the standard deviation, bias term, and margin parameter are set to 0.1, 0.1, and 0.000001, respectively. We train our network in groups of 50, and after every 10 training samples, we have a conversation. Network training using the Adam-optimizer at the 0.001 to 0.0001 learning level

2. Phase of Testing and Validation

Each piece of information is separated into sets for training, validating, and testing. Three training programmes and certifications were also created using the standards listed in table IV. N-fold cross verification is employed. Repeatedly separating the information in the training, validation, and testing sets results in the production of only one image for each test topic, which makes it exceedingly challenging to interpret the results.

3. Our Model's Outcome

We tested the recognition accuracy across 7 various databases. Results for the top five best accuracy and overall mean accuracy from n-fold rigorous recognition are given below. For the majority of cases, we were able to outperform the state of the art outcomes. Table VIII provides a summary of the findings.

4. Using SVM for recognition

One of the most popular conventional methods for recognising problems is based on support vector machines (SVM). To show that increased pairwise accuracy does, in fact, lead to better classification accuracy.

SVM 1 vs. the rest

For separation training, a three-vector network's last layer, which extracts 10 embedded vectors, is once more used. Think about having different N categories. One separator will be trained for each class out of the total N classifiers using Vs vs all. While others will be treated as something, Class I will be given positive labels.

The KNN separator is sensitive to this classification division, which is intended to produce reliable but not cutting-edge results. All of the information mentioned above was tested, and the outcomes were documented.

A SVM against an SVM

A separate differentiation must be trained for each unique label in the other v/s one. N is consequently divided ($N - 1$). This method is far more sensitive to problems with unequal data stocks than the previous one. The price of the PC is higher, though.

NEW OPENING

We offer efficient algorithms for quickly registering for new courses and assessing whether additional courses should be added to the curriculum. Building triplets, training the complete network, producing insertion, and obtaining the correct k by performing kNN would be absurd when registering for new courses. We do, however, offer fresh algorithms for new courses. The network won't need to be completely trained again.

It is now simply a matter of passing the photos through the recently trained network, splitting them into training and test sets, and then creating embedding and enhancements using the original theme after we have a very effective model. The kNN is carried out in accordance with section VI-C once we have the embedding matrix.

With ways to exclude some of the patients, we perform double cross validation once more, and we achieve very high accuracy using this brand-new, straightforward method as well. Table X provides a summary of the findings.

TABLE 5: RECOGNITION RESULTS SUMMARY

Databases	Image Size	Pair wise accuracy	Our Model top 5	Our Model overall	Benchmark DL	Benchmark Non-DL	PCA[2]	LDA [2]	LBP [2]	Gabor [2]
A	92x110	99.4	97	95.8	94.12	95.9 [16]	–	–	–	–
B	180x200	99.74	99.54	99.12	99.49	99.12 [17]	72.2	79.49	85.89	93.50
C	180x200	99.7	98.61	97.45	97.79	99.4	69.91	76.59	80.50	89.69
D	196x196	99.08	94.19	95.6	–	92.70 [2]	70.94	78.43	84.12	92.70
E	180x200	99.7	100	99.30	99.12	99.4	74.80	81.90	86.50	96.89
F	640x480	99.5	99.19	96.62	–	92.61 [17]	–	–	–	–

OTHER APPLICATIONS

The domain determines the transferability

Transfer learning is the foundation of our entire model. The quantity of work being done has a big effect on the transmission capacity. We tried to transfer the tools from real photos and trained facials from Inception v3 and VGG-face, respectively.

However, images communicated from natural images can also be created in the absence of uncommon class images. Even though we are attempting to categorize subjects using the natural model developed by Inception V3, we have discovered a high degree of precision and amazing accuracy. Figure 9 displays the comparison. When trained on the face of VGG, the highly trained At&T Inception v3 provides accuracy of about 95 percent as opposed to 99.2 percent. The Yale database exhibits the same tendency.

There is a need for training and testing time.

This strategy offers us a special means of alerting con artists. With the help of pre-trained photos, we have created fake embedding. According to our argument, cheating occurs when we use a column with an unpopular matrix to appear intelligent and later discover that the highest similarity value from the cheat column is lower than the incredibly low level of the highest values from the subjects who have been smartly schooled in the column.

DISCUSSION AND CONCLUSION

Employing transfer learning and a triple siamese network, we offer methods for using in-depth learning strategies in small data in accordance with the tests above. Some of the details have reached a level of art recognition. A case study on the transmission of facial expressions has not yet been added. A pre-trained network was also discussed as a feature release for fluid class segmentation. We presently use it for facial recognition, but it has other potential applications as well, including the classification of racial cancer and the identification of unusual animals, to mention a couple. Employing transfer learning and a triple siamese network, we offer methods for using in-depth learning strategies in small data in accordance with the tests above. Some of the details have reached a level of art recognition. A case study on the transmission of facial expressions has not yet been added.

We provide the best algorithms for adding new courses. You must constantly train your network and add new data in order to use any in-depth learning techniques. Our method, however, allows for sufficient customization so that you can create direct embedding and pick up new knowledge. Additionally, we offer a novel strategy for determining a specific limit by drawing on fraudsters' knowledge. Our findings are more reliable and cover a wider range of cases thanks to the accuracy of N-fold recognition. Despite the fact that we consistently rank in the top 5, our task becomes incredibly challenging when we only view one image because our contrast variability is so low. When such a system is utilized, the Top-5 accuracy is appropriate since we can be sure that training will take place

on the finest photos. By using general general accuracy, we have some low recognition values that somewhat lower the results and show that the training images in that collection were not very clear.

REFERENCES

1. Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. "siamese neural networks for one- shot image recognition ". In ICML Deep Learning workshop, 2015.
2. Maxime Oquab, Leon Bottou, Ivan Laptev, and Josef Sivic. "learning and transferring mid-level image representations using convolutional neural networks.". Computer Vision and Pattern Recognition, 2014.
3. "the database of faces," att laboratories cambridge, (2002). [online] avail- able: <http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html>.
4. Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. "CNN features off- the-shelf: an astounding baseline for recognition". CoRR, abs/[1403.6382](https://arxiv.org/abs/1403.6382), 2014.
5. "description of the collection of facial images" dr libor spacek [online] available: <http://cswww.essex.ac.uk/mv/allfaces/index.html>
6. "georgia tech face database" [online] available: <http://www.anefian.com/research/facereco.htm>.
7. G. B. Huang, M. Ramesh, and T. Berg. "labeled faces in the wild: A database for studying face recognition in unconstrained environments."
8. O. M. Parkhi, A. Vedaldi, and A. Zisserman. "deep face recognition". A In British Machine Vision Conference, 2015.
9. Michelle Araujo E Viegas , Richard Simoes , Nikisha Thanekar , Saptak Banerjee, Rahul Dicholkar, 2019, Different Approaches of Face Recognition, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 08, Issue 06 (June 2019).
10. Technical Report, University of Massachusetts, Amherst, pages 07–49, 2007.