Face Recognition using Transfer Learning

Yati Saxena[1], Vaibhav Kumar[2], Prof. Mayuri.H.Molawade[3]
[1,2]Student, BVDUCOE Pune
[3]Professor, BVDUCOE Pune

***Abstract* - Convolutional neural networks (CNNs) are wont to achieve unprecedented facial accuracy on large labeled databases. However, smaller organizations wishing to provide secure biometric-based access to its members may find it challenging to train CNNs with minimal computer resources and small training databases. We address this issue by proposing a facial recognition approach and validation based on transfer learning that requires minimal retraining. We show that by adding a single layer of a single trained element with a small number of neurons in a pre-trained network to be able to function and validate authentication can be achieved in small data sets. In addition, new courses can be registered with high monitoring accuracy without properly adjusting the new feature layer.**

Index Terms - **Face recognition, transfer learning, vgg16.**

I.INTRODUCTION

FOR a small organization, a facial look like a biometric can be an attractive way to allow for different levels of control of different personnel. It has some advantages over other popular biometrics such as the iris, voice, signature, and fingers. It is expensive because it can work with less expensive cameras. It is clean because it is untouched.

In addition, it has a user experience that feels natural because it is part of our daily communication.

Facial expression can be used in one of two situations. In face recognition, the image of the test face should be matched to one of the topics contributing to the set of training photos. However, as the number of studies increased, recognition levels declined. With a large number of studies, facial validation can be used instead. In face verification, it should be determined whether the test image belongs to the desired person from the training database, where the claim can be made according to a second ID-like ID. Although, the accuracy of the verification also decreases with the number of articles, it is often higher than the accuracy of the detection.

For both face recognition and validation, high accuracy is critical to efficient operation, followed by low calculation and data retention requirements. Although the precise nature of the art in facial recognition and validation is achieved through deep convolutional neural networks (CNNs) [XX], it is unclear how this technology can be adopted by smaller organizations. CNNs typically have tens of parameters and are trained on hundreds of thousands of faces using expensive computer systems with large memory and graphics processing units (GPUs). Smaller organizations, they say, with less than a hundred employees, especially those not in the highly efficient computer business, will generally not want to invest in a program dedicated to training CNN in face recognition programs. In addition, the system will need training every time an employee leaves or joins an organization.

In this paper, we describe the results of the design and evaluation of a face recognition system and the verification that, in our opinion, is appropriate for smaller organizations. Our intentions were natural differs from recent standard face recognition tasks that attempt to increase test accuracy on large unencrypted face data sets without having to worry too much about computer training needs [XX]. Our goals were to achieve the following:

1) Very high accuracy (95% + average recognition up to 150 subjects,)

It provides a state of the art accuracy in small databases while also saves XX orders of magnitude in training statistics compared to methods that provide similar accuracy. We obtained these results by combining transfer readings with metrics.

Learning transfers are used to reduce the training data and computers needed for learning work in a new database or new domain. A popular way to transfer learning to CNN is to copy the lower layers of the network trained in the source database to the neural network that will be trained in the targeted database, instead of randomly launching those layers. The lower layers can be frozen or well prepared while the upper layers are trained from scratch. Even with weight loss, CNN needs tens of thousands of new training samples and billions of floating points to learn to recognize a new set of classes or human subjects. While also transferring instruments in the lower parts from pre-trained networks, we fulfill the state of the art with accuracy while keeping orders for size on training computers using a different art, purpose, and training method.

The main difference between our method and previous methods is the strong cold of CNN's lower layers, and training only the last two layers of the matrix instead of recognition. In matric study, the task is to explore how similar businesses are similar. Among the CNN structures Siamese networks have been successfully used in mathematical learning, and especially in facial recognition prior to [XX]. However, the previous work was not suitable for small organizations because it used a large database of images and trained the entire network from scratch so that there was a problem with general facial recognition in many subjects. Our goals were different from theirs, which led to a difference in our training that we wanted to achieve (a) the highest accuracy, (b) the smallest database, (c) the lowest cost of training, and not suitable for small organizations. Our goals should have been extremely accurate with very little training computers and In addition, we have found their costly function to provide the best performance in our face-to-face testing and validation in small data sets. Our approach uses the Siamese network in two different ways. First, unlike previous work, it does not train the entire network from scratch. Instead, it uses learning to transfer savings orders of magnitude in training statistics. Second, it uses a different cost function, which is better suited for smaller data sets.

In summary, the novel and the main points of the work are as follows:

- Introduces a unique combination of transfer learning and metric learning that achieves the facial recognition accuracy of small data sets with orders for a very small number of training statistics. In particular, we include only one layer of training over the expanded layers of layers with about ten thousand to forty parameters compared to millions of parameters.
- Our approach can incorporate new topics to some extent without properly processing CNN.
- We present results and insights based on comprehensive tests with multiple data sets, pre-trained networks, cost activities, and hyper-parameters settings. We also show the results of visualizing the impact of one proposed addition.

II.BACKGROUND AND RELATED WORK

People have the ability to understand and know new patterns or new categories based on their knowledge and memory. We can distinguish any new class if we only know the meaning. Suppose, for example, that we know that there is a car with two wheels in the front, one handle and the balance, we will be able to see the 'segway' in the future. Similarly there are two types of split tasks with a small database, single reading and unstructured reading. In a single-shot study, we are able to predict the test model section given for one training model. While, in zero-shot reading, the model does not look for any training model from the target category.

Koch et al. [7] used the Siamese network to use single-reading Omni-Glot data. The Omni-Glot database is handwritten in 50 languages worldwide. They train Siamese who study the common features of the image in image pairs and provide possible points that show the magnitude of the similarities of both images. The model is then used to test the test images, one per class of the novel, in the form of a pair against the test image. Matches with the highest scores calculated through the verification network are given the highest probability of a single job.

The convectional neural network can serve as a powerful general definition of nature. Razavian et al. has shown that features obtained through in-depth Nets can act as a right of appointment in vision functions [8]. They have extracted features from the OverFeat network and are considered to be the standard representation of various object integration functions, descriptive descriptions and visual scenes. The 4096x1 feature presentation is used in the SVM separator or simple line partition to get various viewing functions. They also suggested that the features, released using OverFeat trained over the

ImageNet database, could be used with a number of visual functions.

III.DATASET DESCRIPTION

A.Over the years, facial recognition skills have been tested on information from: AT&T Laboratories Cambridge, AR Face Database from Ohio, Facial Database from Group Essex Ideas, Cohn Kande A-Coded Facial Expression Database (FE), Verified Multi-Modal Verification for Teleservices and Security Services (XM2VTS) Database and Japan Female Facial Expression (JAFFE) Database etc. However, our work focuses on finding the accuracy of facial data with very few lessons that can be used for bio-metrics in small organizations.

B.Our experiment used six face-to-face data transfer sets called AT&T database, Essex 94, Essex 95, Essex 96, Essex Grimace and Georgia Tech database. All databases have a limited number of images (between 10-18 images per subject) with a number of titles ranging from 15 to 110. Images are copied to three channels (if the images are colorless) and then re-scaled to 299x299 and 224x224 Pixels you can use in Inception V3 or VGG face model respectively to extract the feature.

IV.FEATURE EXTRACTION

In the learning curve, a well-trained model of transmission equipment, well suited to the task is required. In very small databases we cannot train all convolutional neural networks from scratch. Therefore, we use pre-trained models to extract features and feed those features into our densely populated parts. The first layers of convolutional layers are able to extract.

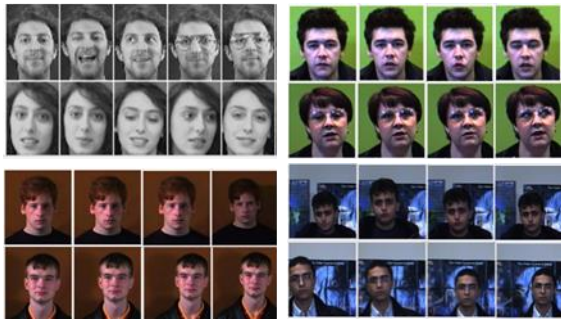

Fig. 1: Images of 2 subjects in database (a) At&t (b) Essex 94 (c) Essex 95 (d) Essex 96.

The generic features pertaining to the input images. The last layers are dataset specific and they contain class-details. We have extracted features using two trained models one trained over faces and other trained over natural images.

A. VGG-Face

CNN's definitions of VGG faces were calculated using CNN implementation based on the VGG-Very-Deep-16 CNN repository as described on and tested on face-labeled field and YouTube face database. VGG Face accepts 224x224 image input size.

VGG surface construction is basically a 16-layer VGG-16. Table II shows the formation of a feature in the central layers of FC. In our experiment, the features released in fc7 output provide excellent results.

TABLE II: VGG-face feature shape with layers

layer name	feature shape
fc6	4096x1
fc7	4096x1
fc8	2622x1

Inception V3 model, are trained on ImageNet which is a large visual database designed for use in visual object recognition software research having total number of images 14,197,122. The default input size for this model is 299x299. The pool layer 3, which gives the output feature of shape 2048x1, is used to extract features.

B. Visualization of Extracted Features

Figure 2 shows the green-drawn elements showing how the features of the same image differ from the contrasting image. At VGG-Face, Georgian technical features have been redesigned from 4096x1 to 64x64 in Figure 2a and Inception V3, At&T features have been redesigned from 2048x1 to 64x32 in 2b for display.

TABLE I: Image Databases Summary (* indicates the value used in experiments)

Databases	Resolution	Image type	Number of subjects	Images per subjects
AT & T	92x112	Gray	40	10
Essex94	180x200	RGB	110*	15*
Essex95	180x200	RGB	71*	11*
Essex96	196x196	RGB	149*	18*

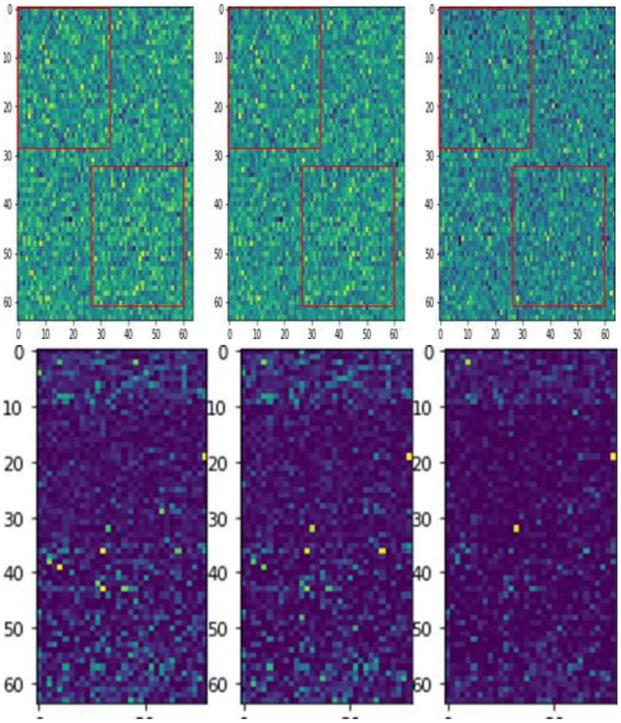

Fig. 2: Visualization of extracted features of triplets using (a) VGG-Face (b) Inception V3

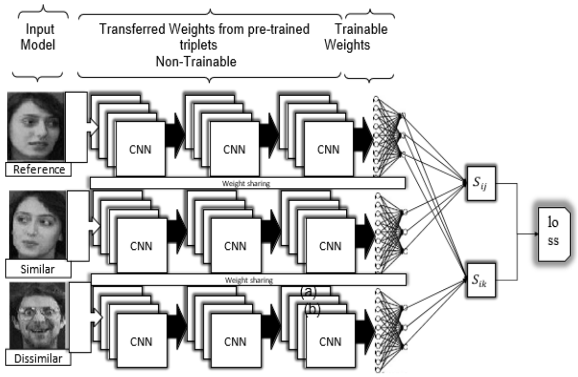

Fig. 6: Model Configuration

C. tMODEL DESCRIPTION

Our model is a Siamese network three times where we have three identical weight sharing structures. Each building has convolutional layers with layers that are completely connected above it. The convolution layers are here untrained and the FC layers are trained.

The initial resolution layers are transmitted in layers from VGG-surface and Inception-V3 with pre-trained weights as described above and serve as a feature output. Details of FC baseline training and loss function are explained in the latest sections. The structure of the complete model is shown in Figure 6.

1.Additional FC Layers

For each of the parallel network used in our model, one fully connected hidden layer is augmented on the top of the convolutional layers. The weights of each layer are shared between all three networks. Currently, all of our tests use a single hidden layer containing 10 neurons without functioning. If we use ReLU with this release of 10 neurons, the accuracy decreases. However, ReLU reconstruction has shown promise when increasing the number of neurons. But since our database is small, we cannot increase the number of output neurons because it will lead to a growing number of parameters.

1) Selection of the number of neurons in the FC layer

Fully connected layers can have an N-number of neurons. We attempted to modify the N and tested the accuracy of the AT&T database with the above parameters. The exact number of neurons can be summarized in Table III:

TABLE III: Testing Accuracy

N	Accuracy
5	98.6
10	99.36
15	99.93
20	99.76

Accuracy increases with increasing number of neurons. But we need to restrict it to lower value because higher N will lead to more number of parameters and will lead to over fitting.

2.Loss Function

For each triplet, we have three vectors of size 10x1. We have used cosine similarity metric. Sij is the normalized dot product of first and second vector denoting the similarity between similar images and Sik is the normalized dot product of first and third vector denoting the similarity between dissimilar images.

$$S_{lm} = \frac{a_l . a_m}{|a_l| . |a_m|}$$

Once the Sij and Sik are known, we want to train our network to learn such that it is able to distinguish similar image from dissimilar image. And for this purpose, we have used hinged loss with some margin η as learn-able parameter. The loss function is:

max(0, (η + Sik − Sij))

While minimizing this loss function, our network learns to find similarity value for similar images which will be greater than the similarity value of dissimilar images.

3.Embedding and Recognition

After network training, we take two of the three branches of art and calculate the cosine similarity in each image article where the training images form lines and tests or the verification images form columns. The association of matrix similarity with the kNN method. This image-generated embedding can be used to detect topics by creating a kNN over it or feeding the SVM separator as a custom kernel. For kNN, in each test image we will be looking at column intelligence and selecting top k images with the highest similarities and based on voting we can assign the subject. The hyper-parameter k is chosen with validation set images and with this k, kNN is performed for test images. Variation of validation accuracy with k can be seen in figure 8.

4.Triplet Formation

Triple training and testing of the Siamese network requires the construction of triplets. The procedures for the construction of the three are different and are described below. In later sections, we will discuss the selection of a number of neurons N in a layer that is fully connected to the problems associated with which one to choose from. When the N and triplets are ready, we will show the loss and pair the variation with precision with intelligence. Our model adopts three images as a [Batch size x 3 x Feature shape] triplet. Therefore, we need to build training and testing three times.

5.Training Triplet Formation

The first image is randomly selected and the second image is randomly selected from the same topic and the third image is randomly selected from any remaining subjects.

Databases	Possible No training triplets	Possible No validation triplets	Data split ratio per subject
AT&T	229,320	76,440	7:2
Yale	47,040	13,440	8:2
Essex94	7,253,950	1,450,790	11:3
Essex95	730,590	243,530	7:3
Essex96	22,360,728	3,726,788	13:4
Essex Grimace	481,950	68,850	15:3
Georgia Tech	1,102,500	245,000	10:2

TABLE IV: Triplet formation.

Table IV shows the number of possible threes, the number of photographs per subject and the number of photographs taken per caption for triple training and testing.

6.Validation Triplet Formation

The first image is chosen randomly from test set, then second image is chosen randomly from the same subject from training set and the third image is chosen randomly for any of the remaining subjects from training set.

V.IMPOSTOR RECOGNITION

In any bio-metrics system recognizing intruders makes an important component of the system. With our algorithm developed we are proposing an appropriate similarity threshold on which single stage and two stage imposter verification is based.

A. Experimental Setup

The datasets are divided in two ratio training and testing faces for every subject. Some subjects are held out beforehand to be treated as imposter. The model is trained with the triplets using VGG-Face extracted features. We add fully connected layers with 10 neurons and initialize the weights as truncated normal with standard deviation=0.1, bias term = 0.1 and margin parameter = 0.000001. We train our network in batches of 50 and randomly shuffle our training samples after every 10 epochs. Adam-optimizer is being used to train the network with learning rate ranging from 0.001 to 0.0001 (depending on dataset). We are using dropout of 0.1 in FC layer while training. After training is done, the similarity matrix is constructed with testing and imposter faces constitute rows and training faces constitute columns. Number of imposter and real faces and threshold is summarized in table V for 1-fold.

TABLE V: 1-fold impostor Recognition Setup

Databases	Real Subjects	Real Faces	impostor Subjects	impostor faces
AT & T	37	111	3	21
Essex94	105	315	5	55
Essex95	66	198	5	35
Essex96	145	435	5	65
Essex Grimace	16	64	2	30
Georgia Tech	46	184	4	36

VI.RECOGNITION RESULTS

The faces are recognized using the kNN method over the generated similarity embedding as explained in section VI-C. The recognition tasks have been performed. When we combine our fraud-based authentication limit with a single validated category (e.g. ID card) to identify the beat rate and the missed rate results are greatly improved. Any fraudulent face, if it does not exceed the maximum value of the maximum match, will be properly rejected even if he or she has an Id card for any older subjects that are already part of the program. However when his face crosses the line of similarity to one of the faces from the previous studies, he can only pass if he holds the Id card of the predicted face. This set gives us two equal measurements of the beat rate and the fraud rate of the fraud as follows was used to extract the features. We are using dropout of 0.1 in FC layer while training. For most of the cases we have fixed number of iterations ranging from 25-150. In this processes at the end loss becomes equal to 1e 07 indicating that network has learned to distinguish the dissimilar image from similar image.

Training part also includes generating all the embedding in the form of similarity matrix and searching the hyper- parameter ”k” used for kNN to recognize the subjects of test classes over all the six datasets

$$\text{Hit rate} = \frac{tp * \text{old subjects} + fn * (\text{old subjects} - 1)}{(tp + fn) * \text{old subjects}}$$

$$\text{Miss rate} = \frac{fn * 1}{(tp + fn) * \text{old subjects}}$$

A.Experimental Setup

1.Training Phase

Before embarking on model training, we transmit the whole image with a feature extractor eg passing through previously trained work. After that we did the training three times as described in the section 2.4. We insert fully connected layers containing 10 neurons and start the instruments as normal reduction with standard devia- tion = 0.1, bias term = 0.1 and margin parameter η = 0.000001. We train our network in groups of 50 and regularly discuss our training samples every 10 times. Adam-optimizer used for network training at a level of learning from 0.001 to 0.0001 (depending on the database). Batch adjustment is already installed in the original basic network.

2.Validation and Testing Phase

Each information is divided into training, validation, and testing sets. Also, three training and certifications were developed according to the criteria in table IV. We use n-fold cross verification. Separating details in training, validation and testing sets is repeated in a roundabout way and every time we produce only one image for each test topic this cross verification cross-section makes our results very difficult.

Our Model Result

We have performed the recognition accuracy for 7 different databases. We have quoted results for top-5 best accuracy and overall averaged accuracy from n-folded rigorous recognition. We have been able to beat the current state of the art results for most of the cases. The results can be summarized in table VIII.

3.SVM based recognition

For recognition tasks, one of the most common traditional techniques is based on support vector machines (SVM). For asserting that higher pairwise accuracy indeed leads to supe-rior classification accuracy, features extracted from the triplet network are used to train a variety of SVM networks, namely one v/s one and one v/s all.

1)SVM one v/s rest

Once again a final layer of a three-vector network that pulls out 10 vectors embedded is used for separation training. Suppose you have different N categories. Vs vs all will train one separator in each class in the total classifiers of N Class i will be considered i-labels as constructive and others as something. This classification division is found to offer consistent, but not state-of-the-art results and is subject to the KNN separator. Tests were performed on all the data used above and recorded.

2) One SVM v/s one

In the other v/s one you have to train a different distinction for each different label. This leads to the division of N (N - 1). This is very sensitive to problems with unequal data stocks, in contrast to the previous process. But it is more expensive for the computer. The results are very similar to a single break division. The results are presented in table IX.

VII.NEW ENROLLMENT

We provide effective algorithms for easy registration of new courses and identifying new courses if they are not part of the program. An absurd way to register new courses would be to make triplets and, train the entire network, create embedding and get the right k by doing kNN. But we offer completely new algorithms for new courses. We will no longer train the network from the beginning.

Once we have a very efficient model we will simply transfer the images through the newly trained network, split the images into set-up training and tests and produce embedding and additions with the old theme. Once we have the embedding metrix we do kNN as described in section VI-C.

Once again we do double cross validation with strategies to omit some of the subjects and we get very high accuracy in this new and simple way too. Results can be summarized in Table X.

Databases	Image Size	Pairwise accuracy	Our Model top 5	Our Model overall	Benchmark DL [15]	Benchmark Non-DL	PCA [2]	LDA [2]	LBP [2]	Gabor [2]
AT & T	92x112	99.5	98	96.5	93.75	96.3 [16]	–	–	–	–
Essex94	180x200	99.85	99.64	99.09	99.55	99.2 [17]	72.1	79.39	85.93	93.49
Essex95	180x200	99.6	98.59	97.43	97.84	99.5	69.87	76.61	80.47	89.76
Essex96	196x196	99.07	94.22	95.7	–	92.68 [2]	70.95	78.34	84.14	92.68
Essex grim	180x200	99.8e	100	99.25	99.11	99.5	74.79	81.93	86.45	96.91
Georgia Tech	640x480	99.4	99.2	96.61	–	92.57 [17]	–	–	–	–

TABLE VIII: Recognition results Summary

VIII.OTHER APPLICATIONS

A. The ability to transfer depends on the domain

Our entire model is based on transfer learning. The ability to transfer is highly dependent on the work being done. We have tried to transfer the instruments from VGG-face and Inception v3 trained facials and natural images respectively. We have noted that transferring instruments to the same domain is always better than transferring from another domain.

However in the absence of unusual class images transmitted from natural images can also be made. We have found a high degree of precision with remarkable accuracy even though we are trying to classify subjects according to the natural model trained by Inception V3. The comparison can be seen in Figure 9.

The highly trained At&T Inception v3 gives approximately 95% accuracy compared to 99.2% when trained over the face of VGG. The same behavior goes with the Yale database as well.

B. Time for training and testing makes sense

This particular method gives us a unique way of informing fraudsters. We have generated fraudulent embedding with pre-trained images. Our claim is that if we look smart with a column with an unpopular matrix and find that the highest similarity value from the cheat column is lower than the very low level of the highest values from the subjects that have been wisely trained in the column.

IX.DISCUSSIN AND CONCLUSION

According to the tests above, we provide ways to use in-depth learning strategies in small data using transfer learning and a triple siamese network. We have achieved a level of art recognition in some of the details. So far no one had included a case study for the transmission of facial expressions. We also discussed how a pre-trained network can be used as a feature release for a seamless class segmentation. Now we have used it for face recognition but this can get an application in the works including the division of race cancer, the recognition of rare species, one gun learns to say a few. Although we do not have a pre-trained network from a single domain, however these methods can be used to transfer devices from a network originally trained for a different job as well.

We also showed two different ways of viewing the feature. Before using any machine learning methods it is important to know whether our extracted features are separated or not. However you may not always be able to visualize but our visualization using redesigned features in the form of images and t-SNE sites is a powerful and effective way to test whether your feature will work or not.

We offer the most effective algorithms for enrolling new courses. For any in-depth learning methods you always need to train your network and new data. But our algorithm allows customization to a level where you can generate direct embedding and get to know new topics. And with the knowledge of fraudsters we offer a new way to set a clear limit.

We can also think of it this way as a kernel learning algorithm. Because the matched match matrix has a fixed size and can be used as a custom kernel for SVM-based information.

The accuracy of N-fold recognition makes our results more robust and includes all types of cases. Although we provide top-5 very good accuracy and because when we see only one image the task becomes very difficult as we have very little variability of contrast. The best accuracy of the Top-5 is suitable whenever such a system will be used wherever we can be sure that the training takes place in the best images. By quoting general general accuracy, we have some low recognition values that draw the results slightly lower, indicating that the training images were not very clear in that set.

REFERENCES

[1] 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.

[2] Gregory Koch, Richard Zemel, and Ruslan Salakhutdinov. ”siamese neural networks for one-shot image recognition ”. In ICML Deep Learning workshop, 2015.

[3] Ali Sharif Razavian, Hossein Azizpour, Josephine Sullivan, and Stefan Carlsson. ”CNN features off-the-shelf: an astounding baseline for recognition”. CoRR, abs/1403.6382, 2014.

[4] ”the database of faces,” att laboratories cambridge, (2002). [online] avail- able: http://www.cl.cam.ac.uk/research/dtg/attarchive/facedatabase.html.

[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] O. M. Parkhi, A. Vedaldi, and A. Zisserman. ”deep face recognition”. A In British Machine Vision Conference, 2015.

[8] G. B. Huang, M. Ramesh, and T. Berg. ”labeled faces in the wild: A database for studying face recognition in unconstrained environments.”. Technical Report, University of Massachusetts, Amherst, pages 07–49, 2007.