



# Average Neural Face Embeddings for Gender Recognition\*\*

Semiha Makinist<sup>1\*</sup>, Betül Ay<sup>2</sup>, Galip Aydın<sup>3</sup>

<sup>1</sup> Data Processing Department, Firat University, Elazig, Turkey (ORCID: 0000-0002-6636-7898)

<sup>2</sup> Computer Engineering, Firat University, Elazig, Turkey (ORCID: 0000-0002-3060-0432)

<sup>3</sup> Computer Engineering, Firat University, Elazig, Turkey (ORCID: 0000-0002-9564-3329)

(Conference Date: 5-7 March 2020)

(DOI: 10.31590/ejosat.araconf67)

**ATIF/REFERENCE:** Makinist, S., Ay, B & Aydın, G. (2020). Average Neural Face Embeddings for Gender Recognition. *Avrupa Bilim ve Teknoloji Dergisi*, (Özel Sayı), 522-527.

## Abstract

In recent years, with the rise of artificial intelligence and deep learning, facial recognition technologies have been developed that operate with high accuracy even in adverse conditions. However, extracting demographic information such as gender, age and race from facial features has been a hot research area. In this study, a new Average Neural Face Embeddings (ANFE) method that uses facial vectors of people for gender recognition is presented. Instead of training deep neural network from scratch, a simple, fast and effective solution has been developed that performs a distance calculation between the average gender vectors and the person's face vector. The method proposed as a result of the study carried out provided a high and successful recognition performance with with 96.47% of the males and 99.92% of the females.

**Keywords:** Face Embeddings, Face Detection, Average Embeddings, Gender Recognition, Deep Learning.

## 1. Introduction

Recently, the name of deep neural networks has been frequently heard both in image processing and natural language processing. Deep neural networks are based on the depth architecture of brain neurons. This model is also known as Deep Artificial Neural Networks in the academic world. In other words, it was discovered based on the idea of deepening artificial neural networks. In the early 2000s, serious studies were made in the field of deep learning and this period was accepted as a turning point for the field of artificial intelligence. In the mid-2000s, Geoffrey Hinton and Ruslan Salakhutdinov taught other researchers how to train multilayered forward and feedback neural networks through their publications [1]. Successful models developed to date have not exceeded two or three-layer network models. In 2006, Deep Belief Nets showed how multilayer neural networks will work and how undefined features are learned by the system. These new generation ANNs are named as Deep Net and the studies in this field are gathered under the title of Deep Learning. The development stages of facial recognition technology with deep learning are illustrated in the Figure 1.

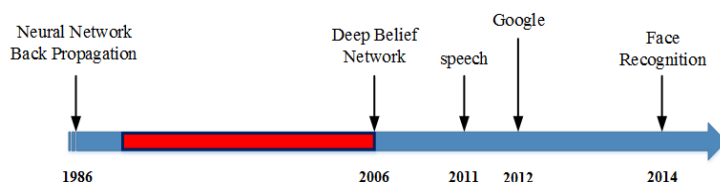


Figure 1. Face recognition timeline

\* Semiha Makinist: Data Processing Department, Firat University, Elazig, Turkey, ORCID: 0000-0002-6636-7898, [smakinist@firat.edu.tr](mailto:smakinist@firat.edu.tr)

\*\* This paper was presented at the *International Conference on Access to Recent Advances in Engineering and Digitalization (ARACONF 2020)*.

Deep learning is widely used in image, sound and text analysis. The major research areas in the field are face recognition and detection, age and gender detection. The success of the vanilla machine learning methods developed for the recognition of facial features such as age and gender detection remained between 75% and 80%. When the deep neural networks are used in the studies conducted in the same field, the success rate has exceeded 90%. In the studies examined, classical classifiers were generally used for face, gender or age recognition. In recognition and detection systems, the classifier reduces both system performance and success. In order to eliminate all these disadvantages, this paper is based on the idea that “With the developing technologies and methods can a face, age or sex be distinguished without using any classification algorithm?”. The starting point of the proposed method is a GitHub project shared by Geitgey [2] for face recognition. In his study, Geitgey showed how to identify a person through a single image using face embedding model of dlib library (<http://dlib.net/>). In this study, this method was made more specific and used for gender recognition. The greatest innovation in this new method proposed in the study was to show that gender recognition can be performed using 128-D average face vectors without using any classification algorithm.

Many studies have been conducted in the literature on face, age and gender recognition by using facial landmark with deep learning methods. Cha et al. [3] adopt a multi-task Deep Convolutional Neural Network (DCNN) method and performed face detection using facial landmarks for different face poses. They used the FDDB dataset [4] and as a result of the study it was observed that the method they proposed improved the other state-of-the-art methods by 3%. Sun et al. [5] designed a 3-level DCNN which cascades three levels of convolutional networks for facial point detection. They have obtained much more successful results in the detection of facial points than previous methods. But at the same time, the proposed method requires a complex cascade architecture of deep network. Based on this disadvantage, a new tasks-constrained deep convolutional network (TCDCN) reduces model complexity has been presented for facial point detection [6]. Eidinger et al. [7] has made age and gender prediction using unfiltered faces. Within the scope of the study, they formed their own dataset for age and gender prediction. They developed a dropout-SVM method for classifying data, inspired by the deep belief network's dropout learning technique. Hassner et al. [8] corrected the front view of the face by applying “frontalization” process to the face detected in unconstrained photos. They used important facial feature points in the infrastructure of their studies. As a result of the new image obtained, the percentage of success in face recognition and gender prediction systems was increased. Levi et al. [9] have designed a simple CNN that can work on a limited dataset and can predict age and gender. Ranjan et al. [10] proposed a deep multi-task learning framework called HyperFace that can perform simultaneously face detection, landmark localization, pose estimation, and gender recognition using CNN. Experimental results have shown that the proposed method can capture both global and local information on faces, and it performs far better than many algorithms for each of these four tasks. Rothe et al. [11] presented a model that can predict age and gender on a single image using the deep learning method. They used IMDB-WIKI dataset within the scope of the study. In the previous studies in this field, the images in the dataset used for training were not a single image, but the most important feature that distinguishes this study from the others is the use of a single image for training. Some convolutional layers in the VGG-16 architecture have been redesigned. For gender recognition, Mansanet et al. [12] proposed a Local Deep Neural Network named as Local-DNN. The proposed local-DNN model is based on the deep learning architecture and local features of the face. The model learns by using Feed Forward Networks in several layers and small overlapping regions in the visual fields. In another study using CNN architecture [13], face based gender estimation was performed. Xinga et al. [14] proposed a DNN model that can predict race and gender as well as age prediction using deep multi-task learning architecture. Moeini et al. [15] has performed gender detection using the features of face position and expression with gender dictionary learning. Qawaqneh et al. [16] have been designed a neural network model that can classify age and gender using DNNs. They also proposed a new cost function. Both speech data and face images were used in the study. Philip et al. [17] have been using both VGG19 and VGGface models, which were previously trained CNN-based deep neural networks. They have been studying transfer learning for model trainings. In order to increase system success, they have been changing the model parameters. they have achieved 98% success in gender recognition with their CNN-based models. Dhomne et al. [18] have proposed a VGGNet model based on D-CNN using facial images for gender recognition. Xu et al. [19] have been proposing Hierarchical Multi-task Network (HMTNet), a deep neural network that can identify both sex, race, and facial beauty from a person's portrait image.

## **2. Material and Metot**

### **2.1. Face Embeddings**

Embedding is the representation of a document, word or image in a 2D or 3D space. In other words, documents, words or pictures (objects, human, face, so on) are represented vectorically in two-Dimensional space. This representation of faces as numerical vector is called “face embeddings”. Different methods are used to create face embeddings. One of them is deep neural networks. There are two important studies in the literature that use deep neural networks to extract face embedding: Dlib [20] and Openface face recognition library [21]. dlib is written in C++ and has Python API. Openface uses the dlib library for basic operations such as face detection, while it uses a deep neural network model written in a Torch environment to extract face embedding. In these two important studies, the person uses face vector representation of 128-D to recognize. The core of the deep neural network used in the Dlib face recognition model consists of ResNet. The ResNet (Residual Networks) used is a 34-layer network developed by He, Zhang, Ren and Sun [22] for image recognition in 2016. FaceNet, Schroff et al. [23] from the Google team, is a deep neural network model for extracting face embedding vectors. In this model, faces are represented by 128D vectors. At least 3 pictures of each person for extracting face embeddings with FaceNet are required. Because FaceNet uses a triple based loss function used in the LMNN model [24]. FaceNet model consists of such layers like input, CNN for face detection, L2 distance to separate face vectors, create the face embedding, and the triple loss function, in which the error values.

dlib basically uses the ResNet-34 architecture. The number of filters and layers are reduced unlike ResNet-34 architecture. The number of filters on each layer is halved. Some layers were removed and the network was redesigned as 29 layers. Thus, the cost of

calculation has been reduced. A 128-D face embedding vector was obtained with the newly designed network. VGG and face scrub dataset was used for new network training.

**2.2. Average Neural Face Embeddings**

In this paper, we propose a novel and simple approach with Average Neural Face Embeddings (ANFE) to recognize gender without using any classifier method. The algorithm steps of the proposed method are given below:

- Step 1.** The training set is divided into two groups as male and female.
- Step 2.** 128-D embeddings for each face in two different groups in the training set are extracted with special python libraries.
- Step 3.** These 128-D embeddings of the individuals in each group within the training set are summed and then divided by this sample number. The mathematical representation of this structure is given in equation (1). In the equation, m represents the number of samples, ie the number of people. X is the face embeddings (feature vectors) extracted for each person.

$$average_{gender(x)} = \frac{\sum_{i=1}^n \sum_j^m X[j, i]}{m} \tag{1}$$

- Step 4.** The calculated ANFE values are saved in the MongoDB database.
- Step 5.** A controlled dataset was prepared to measure the success of the method. In test dataset, two different groups, male and female, were formed like the training dataset.
- Step 6.** The 128-D face embeddings of the images in each group in the test dataset were extracted and compared with the average gender vector of 128-D stored in the database. Euclidean distance was used to compare these two vectors (individual face embeddings and ANFE for each gender group). Mathematical representation of Euclidean Distance is given in Equation (2).

$$eu\_dist([x, y], [a, b]) = \sqrt{(x - a)^2 + (y - b)^2} \tag{2}$$

As a result of the query, the distance value of the face feature vector to the average female and male face feature vectors is returned. The minimum distance is selected from these calculated distance values. The gender with the minimum distance is the estimated value. The graphical representation of the proposed approach is presented in Figure2.

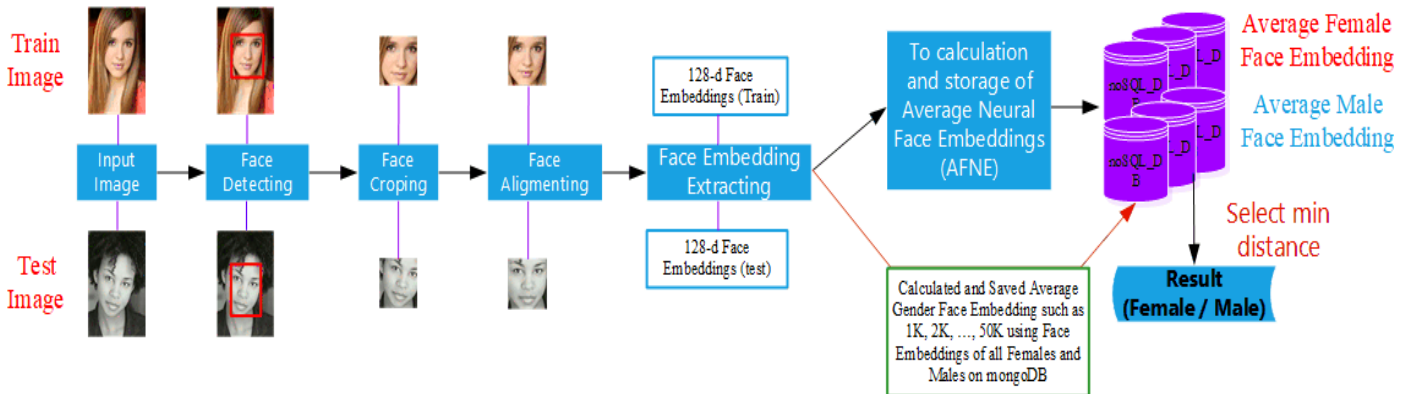


Figure 2. System architecture of ANFE

**2.2.1. Dataset**

While the training and test dataset were created, they were passed through different stages. These stages are as follows;

- Data harvesting,
- Automatic and manual clearing of collected data,
- Creation of dataset with different numbers of data to determine the most appropriate ANFE vector.

Web scraping software developed on java platform for data collection step. With this software, pictures of famous people from different movie sites that contain gender data such as www. filmweb.pl, www.listal.com were taken. The scrapped image data is automatically foldered by gender class. Another important step is clearing the images in the dataset. Because in the automatically downloaded images there are mislabeled images or images contain more than one face. These images were cleaned by passing through two stages. In the first step, multiple face and unreadable images are extracted from the dataset using Python's face detection libraries. In the second step, the remaining images are checked manually. Incorrectly labelled images are added to the appropriate folder. All these stages were performed for both training and test datasets. The data collected from filmweb.pl was allocated as training data and the data collected from listal.com as test data. After the pre-processing steps, dataset consist of 133,498 face images, 62,333 of which are female and 71,165 of which are male. ANFE were extracted using a maximum of 50K images for both men and women. Data collection step is illustrated in Figure 3.

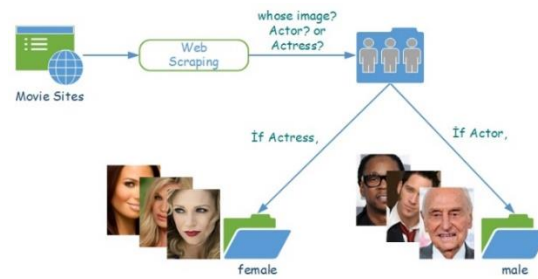


Figure 3. Data collection system with web scraping

Dataset is presented in the Figure 4.



Figure 4. Female and male dataset

Facescrub [25] benchmark dataset was used to evaluate the performance of the proposed system. This dataset was also pre-processed. With a preprocessing program written in Python script, images containing multiple faces and noisy faces were automatically removed from the dataset. Then, people between the ages of 0 and 16 were removed from the remaining pictures. Finally, the labels of the remaining images have been manually checked. As a result of all the preliminary procedures; 31,370 pictures remained in the male dataset and 26,631 pictures in the female dataset.

### 3. Research Results and Discussion

#### 3.1. Experimental Setup

Tests were performed on a desktop machine with 3.5 GHz CPU, 32GB RAM, 4GB NVIDIA card and 1TB HDD. All tests performed with this machine lasted 25 minutes in total.

#### 3.2. Experimental Tests

The system has been tested in two different ways. The first test is to find the most appropriate ANFE vector. The second test is to measure system success on the facescrub dataset. In the first performance test, the data taken from filmweb.pl as training were adjusted to be 2K (1K female, 1K male), 4K (2K female, 2K male), ..., 100K (50K female, 50K male). The images selected for the training data were taken randomly. ANFE vectors of each different data group were extracted and recorded in MongoDB database. In the database, a distinctive id is given to differentiate each different ANFE vector. For example, the id index of the 1K female and 1K male data group is g\_1000. The graphical result of gender recognition is presented in Figure 5.

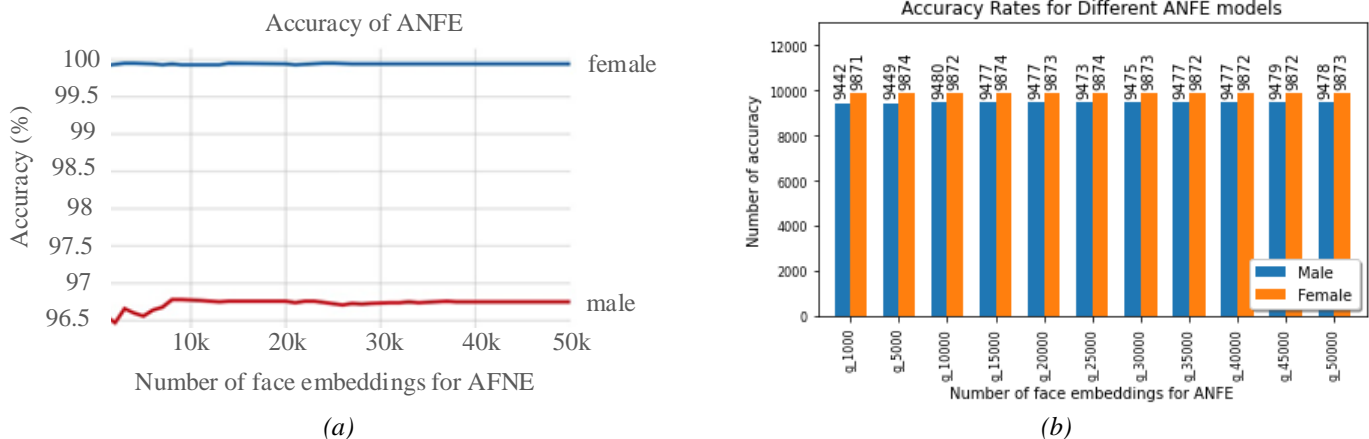


Figure 5. Recognition performance of ANFE

The same test images are used for each model shown in Figure 5. These models were tested on 20000 images, including 10000 women and 10000 men. When the Figure 5(a) is examined, it is observed that success of ANFE approach ranged between 96.5% and 99.92% according to gender. When all the average gender embeddings are examined, it is seen that the most appropriate study group is 10K (10K female and 10K male average gender embeddings). In the tests performed for this data set, the gender of 8 women (99.92%) and 353 men (96.47%) were incorrectly estimated. When the incorrect data were examined, it was found that unknown persons in males were generally Far Eastern. False negatives are given in the Figure 6.



Figure 6. False negatives

In the next step of the study, calculated distance values for correctly predicted female and male classes were investigated. As a result of the investigations, the mean values of the distance values from the images used in the tests to the ANFE vectors in their class were 0.6, and this value was observed as 0.65 on the wrongly estimated images (Figure 7).

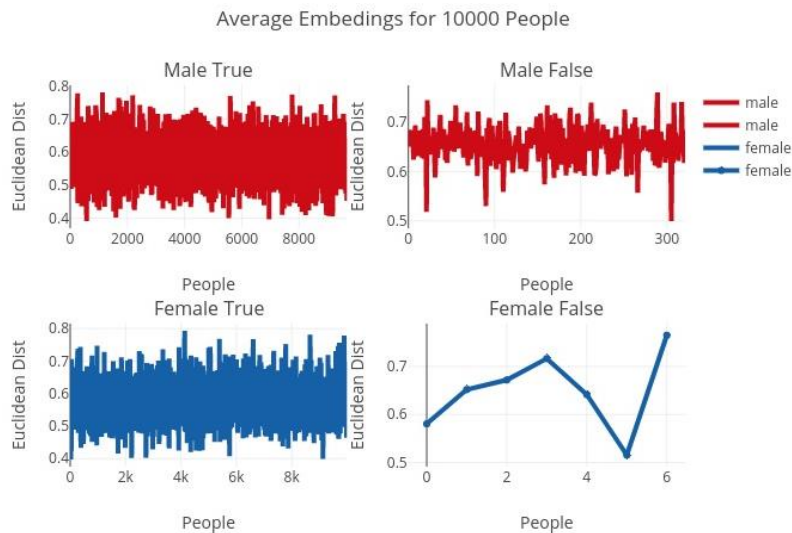


Figure 7. Suitable working range of ANFE

The second test was performed using 10K ANFE vectors determined at the end of the first test. For this test, facescrub (31.370 male and 26.631 female) data set were used. Of the 58,001 images, 228 (62 male, 166 female) pictures were incorrectly estimated. As a result of this test, it was observed that the system achieved 99.802% success in men and 99.376% success in women.

## 4. Conclusion

In this study, ANFE method which is a new, fast and simple approach instead of classical data classification methods for gender recognition is presented. The success of the proposed method was demonstrated by the tests performed.

All tests were examined and two races were identified that prevented success. One of them is black people and the other is Asian. The main reason for this is that the dlib model does not have enough picture of people related to these races. In the next study, it will be re-trained to address these shortcomings in the facial recognition model. A new study will be conducted with the new model trained.

## Acknowledgement

This study was carried under the project “Deep Learning and Big Data Analysis Platform (DEGIRMEN)” supported by Presidency of the Republic of Turkey, Presidency for Defence Industries (SSB).

## References

- [1] Hinton, G. E., Osindero, S., & Teh, Y. W. (2006). A fast learning algorithm for deep belief nets. *Neural computation*, 18(7), 1527-1554.
- [2] Geitgey, A., Available: [https://github.com/ageitgey/face\\_recognition](https://github.com/ageitgey/face_recognition)
- [3] Zhang, C., & Zhang, Z. (2014). Improving multiview face detection with multi-task deep convolutional neural networks. *In Applications of Computer Vision (WACV), 2014 IEEE Winter Conference on*, 036-1041.
- [4] Jain, V., & Learned-Miller, E. (2010). Fddb: A benchmark for face detection in unconstrained settings (Vol. 2, No. 6). UMass Amherst technical report.
- [5] Sun, Y., Wang, X., & Tang, X. (2013). Deep convolutional network cascade for facial point detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3476-3483).
- [6] Zhang, Z., Luo, P., Loy, C. C., & Tang, X. (2014, September). Facial landmark detection by deep multi-task learning. In *European conference on computer vision* (pp. 94-108). Springer, Cham.
- [7] Eidinger, E., Enbar, R., & Hassner, T. (2014). Age and gender estimation of unfiltered faces. *IEEE Transactions on Information Forensics and Security*, 9(12), 2170-2179.
- [8] Hassner, T., Harel, S., Paz, E., & Enbar, R. (2015). Effective face frontalization in unconstrained images. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4295-4304).
- [9] Levi, G., & Hassner, T. (2015). Age and gender classification using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops* (pp. 34-42).
- [10] Ranjan, R., Patel, V. M., & Chellappa, R. (2017). Hyperface: A deep multi-task learning framework for face detection, landmark localization, pose estimation, and gender recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 41(1), 121-135.
- [11] Rothe, R., Timofte, R., & Van Gool, L. (2018). Deep expectation of real and apparent age from a single image without facial landmarks. *International Journal of Computer Vision*, 126(2-4), 144-157.
- [12] Mansanet, J., Albiol, A., & Paredes, R. (2016). Local deep neural networks for gender recognition. *Pattern Recognition Letters*, 70, 80-86.
- [13] Antipov, G., Baccouche, M., Berrani, S. A., & Dugelay, J. L. (2017). Effective training of convolutional neural networks for face-based gender and age prediction. *Pattern Recognition*, 72, 15-26.
- [14] Xing, J., Li, K., Hu, W., Yuan, C., & Ling, H. (2017). Diagnosing deep learning models for high accuracy age estimation from a single image. *Pattern Recognition*, 66, 106-116.
- [15] Moeini, H., & Mozaffari, S. (2017). Gender dictionary learning for gender classification. *Journal of Visual Communication and Image Representation*, 42, 1-13.
- [16] Qawaqneha, Z., Mallouha, A.A. & Barkana, B.D. (2017, November). Age and gender classification from speech and face images by jointly fine-tuned deep neural networks. *Expert Systems with Applications*, 85, 76-86.
- [17] Smith, P., & Chen, C. (2018, December). Transfer Learning with Deep CNNs for Gender Recognition and Age Estimation. In *2018 IEEE International Conference on Big Data (Big Data)*, 2564-2571.
- [18] Dhomne, A., Kumar, R., & Bhan, V. (2018). Gender Recognition Through Face Using Deep Learning. *Procedia Computer Science*, 132, 2-10.
- [19] Xu, L., Fan, H., & Xiang, J. (2019, September). Hierarchical Multi-Task Network For Race, Gender and Facial Attractiveness Recognition. In *2019 IEEE International Conference on Image Processing (ICIP)*, 3861-3865.
- [20] King, D.E. (2009, July). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 1755-1758.
- [21] Amos, B., Ludwiczuk, B. & Satyanarayanan, M. (2016). Openface: A general-purpose face recognition library with mobile applications." *CMU School of Computer Science* 6.
- [22] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770-778.
- [23] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 815-823.
- [24] Weinberger, K.Q., & Saul, L.K. (2009). Distance metric learning for large margin nearest neighbor classification. *Journal of Machine Learning Research*, 10, 207-244.
- [25] Ng, H.W., & Winkler, S. (2014, October). A data-driven approach to cleaning large face datasets. In *2014 IEEE International Conference on Image Processing (ICIP)*, 343-347.