COMENIUS UNIVERSITY BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND INFORMATICS

PERSON IDENTIFICATION WITH PARTIALLY OCCLUDED FACE

Diploma Thesis

2022

Bc. Anna Camara

COMENIUS UNIVERSITY BRATISLAVA
FACULTY OF MATHEMATICS, PHYSICS AND
INFORMATICS

PERSON IDENTIFICATION WITH PARTIALLY OCCLUDED FACE

Diploma Thesis

Study programme:	mAIN/k - Applied Computer Science (Conversion Programme)
Field of Study:	Computer Science
Department:	FMFI.KAI Departement of Applied Informatics
Supervisor:	RNDr. Zuzana Černeková, PhD.

Bratislava, 2022 Bc. Anna Camara

96678182

Comenius University in Bratislava
Faculty of Mathematics, Physics and Informatics

THESIS ASSIGNMENT

Name and Surname:	Bc. Anna Camara
Study programme:	Applied Computer Science (Conversion Programme) (Single degree study, master II. deg., full time form)
Field of Study:	Computer Science
Type of Thesis:	Diploma Thesis
Language of Thesis:	English
Secondary language:	Slovak

Title: Person identification with partially occluded face

Annotation: The goal of the thesis is to identify a person in case when face is partially occluded for example with sunglasses or face mask. Study the topic of person identification based on the face. Analyze the performance of the existing solutions published in the literature. Propose a new method based on a neural network, which can find and identify a person. Create a dataset for training and testing purposes. Evaluate the proposed method and draw the conclusions.

Supervisor:	RNDr. Zuzana Černeková, PhD.
Department:	FMFI.KAI - Department of Applied Informatics
Head of department:	prof. Ing. Igor Farkaš, Dr.
Assigned:	24.09.2018
Approved:	03.10.2018

prof. RNDr. Roman Ďurikovič, PhD.
Guarantor of Study Programme

....
Student

....
Supervisor

Čestne prehlasujem, že túto diplomovú prácu som vypracovala samostatne len s použitím uvedenej literatúry a za pomoci konzultácií s môjou školiteľkou.

Bratislava, 2023 ....

Bc. Anna Camara

Poďakovanie

Touto cestou by som sa chcel v prvom rade poďakovať môjmu školiteľovi .... za jeho cenné rady a usmernenia, ktoré mi veľmi pomohli pri riešení tejto diplomovej práce. Takisto sa chcem poďakovať mojím kolegom ...... za rady ohľadom implementácie a v neposlednom rade chcem tiež poďakovať .....

Abstract

Key words: facial identification, facial recognition,

Abstrakt

Kľúčové slová: tvárová identifikácia, rozpozanie tvárií.

Contents

Chapter 1

Introduction

More and more, facial identification is becoming part of our lives. We are hearing terms like facial identification, facial recognition, verification, biometric and others. First let's clarify what this terms mean and what is the difference between them.

The International Organization for Standardization (ISO) [13] provides following definitions:
Biometric Characteristic is a biological and behavioural characteristic of an individual from which distinguishing, repeatable biometric features can be extracted for the purpose of biometric recognition.
Biometric Recognition/Biometrics is an automated recognition of individuals (referring to only humans) based on their biological and behavioural characteristics. Biometric recognition encompasses *biometric verification* and *biometric identification.*
Biometric identification is a process of searching against a biometric enrolment database to find and return the biometric reference identifier(s) attributable to a single individual.

Biometric verification is a process of confirming a biometric claim through comparison.

In simpler words, when we speak of biometrics or biometric recognition we mean biological and behavioral measurements that can be used to identify individuals. This is a broad term for both verification and identification. In verification we are comparing (1:1) one input against one control point. Basically we are asking *"Is this the same person as the one saved control point?"* or *"Are you who you say you are?"*. In identification we are comparing (1:N) one input against a whole database. We are asking *"Who, from our database, is this?"* or *"Who are you?"*.
This of course includes more than recognition based on ones face. In current age we are able to identify a person from many sources some of which are fingerprints, voice, scan of retina and face.

Now when we have these definitions it is simple to clarify what is facial recognition. **Facial recognition** is biometric recognition based on persons face. This theses will focus on facial identification, specifically on facial identification with partially occluded face by face mask.

Goals

The goal of this thesis is to gather research on face recognition techniques that use deep learning and improve facial them to be able to identify masked faces. In this process I would like to focus on two groups of questions:

1. Face identification with facial mask:

    - What different techniques are currently used for facial identification?

- How well do facial identification models/softwares perform on masked faces?

2. Racial bias:

    - Do current face recognition have any skin colour bias or is this a thing of the past? I would like to separate evaluation on different ethnic groups.
    - If this bias exists, how to reduce it?

Chapter 2

Facial recognition

This chapter provides a short description of facial recognition (FR) to gain basic understanding of the topic.

As mentioned in chapter 1, facial recognition is biometric recognition based on persons face. It comprises both identification, comparing one identity to many(1:N) and verification, where we match one to one (1:1). To put it in plain words, facial recognition is an automated system for finding a person's identity based on an image of their face.

In **verification mode** a one-to-one comparison is performed to decide whether the identity claim is true or false. Face verification systems are classically assessed by the receiver operating characteristic (ROC) and the estimated mean accuracy (ACC). [1] This is used mainly in security features like faceID to unlock phone or access a bank account.

In **identification mode**, the system identifies an individual by searching for the best match between all faces stored in the database. Therefore, a one-against-all comparison is performed to determine this individuals identity or failure if that individual does not exist in the database. [1] For identification, two different test protocols may be used: open-set and closed-set. For **closed-set protocol**, all testing identities are predefined in training set. In this scenario, face verification is equivalent to performing identification for a pair of faces respectively. Therefore, closed-set FR can be well addressed as a classification problem, where features are expected to be separable[16]. We can directly map face to a known identity in the training set. Rank-N is a fundamental performance metric used in closed-set face identification to measure the model's accuracy, where the valid user identifier is returned within the N-Top matches. The primary measuring performance is recorded using correct identification rates on a cumulative match characteristics (CMC) curve [1]. For **open-set protocol**, the testing identities are usually not included in the training set, which makes FR more challenging yet closer to practice. In this scenario we need to map faces to a discriminative feature space [16]. To measure the model's accuracy we use measures such as the false negative identification rate (FNIR) and the false positive identification rate (FPIR) [1]. Desired features are expected to satisfy a criterion that the maximal intra-class distance is smaller than the minimal inter-class distance under a certain metric space [16]. In CNN approaches we usually achieve this by the choice of used loss function.

2.1 How it works

In general automated face recognition has three stages:

1. Face detection
2. Feature extraction
3. Classification

In face detection stage, the system determins whether a face is in the picture. If so, the face is segmented from the rest of the picture and aligned to some predefined canonical structure. In literature tight crops of the face area are referred to as thumbnails. Depending on the application we can choose to locate more than one face in the input image. Face recognition systems usually work with simple, every day hardware so the images are often distorted by various factors like light, noise, blur, face occlusion, disguises, make-up and other secondary factors that are very common in real life applications. Therefore this stage may incorporate some pre-processing as well. The feature extraction stage consists of extracting a feature vector from the detected face that will represent the face sufficiently. In classification stage we match the face image to identity. These stages are not necessarily separate processes. Face detection and feature extraction can run simultaneously or feature extraction can be part of the classification. Depending on the nature of feature extraction and classification we can divide 2D face recognition methods into several sub-classes [1]. This thesis is going to be focusing on deep-learning based methods.

2.2 Deep neural networks

If we consider a task of face recognition under controlled conditions (static pose, constant light, homogenous face expressions,...) simple classical approaches provide excellent performance. However, these conditions are hard to achieve. Real life applications aroused the need to focus research on FR under unconstrained conditions where deep neural networks have proven to be the best tool thanks to their strong robustness against numerous variations that can alter the recognition process.

Deep neural network (DNN) model is acting as a feature extractor and classifier. It extracts or transforms features from input using multiple nonlinear processing units arranged in layers with various levels of representation and abstractions. Similarly to a convolutional neural network (CNN) it consists of a few types of layers: Filtering layer, convolutional or fully connected. Pooling layer used for down sapling. Thresholding or activation layer which introduces nonlinearity, most commonly referred to as RELU because rectified linear activation function is the most widely used activation function. Normalization to make sure each feature contributes to the output in the same scale. And at the "bottom" of the network, there is a classification or loss layer that can be used to adjust learning parameters. Therefore, a properly constructed loss layer can be crucial to train an effective deep neural network model [26].

In order to produce distinct representations of faces, so-called embeddings, networks with various architectures are used for achieving high accuracy [26].

DNNs are a learning based method, they have to be trained for specific task. It is a data driven process, which means the network learns directly from the input, in our case the pixels of the face. The goal of training is to maximize the probability of the correct class(identity) [21]. Which is achieved by minimizing chosen loss function and backpropagation of error.

2.3 Loss functions

2.3.1 Triplet loss function

In FaceNet [19] triplets consist of two matching face thumbnails and a non-matching face thumbnail and the loss aims to separate the positive pair from the negative by a distance margin. They found out that the choice of triplets is very important for achieving good performance.

Self-restrained Triplet loss

Batch triplet loss

2.3.2 Cross-enthropy loss

used in DeepFace algorithm.

2.4 Similarity measures

Matched Background Similarity (MBGS). This similarity is shown to considerably improve performance on the benchmark tests [10].

2.5 Data

DNNs are data driven methods, that is, they are trained to solve one specific task determined by the data. Therefore a good dataset is very important. For the network to be able to generalize well, the training data must include all possible variations. This demand naturally leads to a very large amount of data. The ImageNet competition showed that models tend to perform better when more data is provided to accurately tune network weights [17]. For the task of face recognition we need examples of all ages, ideally also including data from different age stages of a life of the same people, all genders, all races, all poses and facial expressions, etc. Considering that the interest of this thesis is in the task of masked face recognition, all of these examples are needed both as masked and unmasked. It is nearly impossible to meet these demands, hence finding (or creating) an appropriate dataset is a challenge on its own. Moreover if the training set excludes some part of demographics by training on the target-domain's training set, one is able to fine-tune a feature vector (or classifier) to perform better within the particular distribution of the dataset [21].

2.5.1 Datasets

Labeled Faces in the Wild (LFW) (2008), which is de facto the benchmark dataset for face verification in unconstrained environments (information from 2014). It has about 75% males, celebrities that were photographed by mostly professional photographers [21]. This dataset was created in 2008 as an aid in studying the problem of unconstrained face recognition. The database contains of 13233 labeled images of 5749 people in large range of various conditions typically encountered in everyday life, such as pose, lighting, race, accessories, occlusions, and background. From these 5749 identities 1680 have two or more images and at least 158 have more than 10 pictures in the database [10].

YouTube Faces dataset (YTF) (2011) is database of labeled videos of faces in challenging, uncontrolled conditions. Many of these videos are produced by amateurs, typically under poor lighting conditions, difficult poses, and are often corrupted by motion blur. In addition, bandwidth and storage limitations may result in compression artifacts, making video analysis even harder. This database is a simple pair-matching benchmark, allowing for standard testing of similarity and recognition methods. It is constructed from a subset of identities from LFW dataset. It consists of 3425 videos of 1595 subjects, on average 2.15 videos per person [25].

Social Face Classification dataset (SFC) (2014) is a large colection of photos from Facebook. It includes 4.4 million labeled faces from 4,030 people each with 800 to 1200 faces. Since these are pictures collected from social media, face identities were labeled by its users, which typically comes with about 3% error. But on the plus size, these pictures have even larger variations than LFW or YTF in expression, image quality, lightning and other conditions and they consist of not just celebrities. This dataset was used to train (the first version of) DeepFace. The large number of images per person provides a unique opportunity for learning the invariance needed for the core problem of face recognition [21]. Unfortunately this database is private and belongs to Facebook research team.

MS-Celeb-1M (2016) consists of 10M images of 1M celebrities. It was designed specifically for face recognition at the web scale. Unlike other datasets this one also introduces a knowledge base with several informations about the person to avoid ambiguity. This research team provides an aditional training dataset that contains 10M images of top 100K celebrities selected as a subset from their celebrity list. For each of the image in the training data, the thumbnail of the original image and cropped face region, with or without alignment, are provided [9].

MegaFace2 (MF2) (2017) dataset was created from Flickr (utility photo-sharing platform) photos. Unlike other mentioned datasets this one consists mostly of non-celebs. This was done because by training only on celebrity photographs, we risk constructing a bias to particular photograph settings. This database is intended for training neural networks for face recognition tasks. It consists of 4.7M images of 672K identities [17].

RMFRD(2020) Real-world masked face recognition dataset was created as a reaction to global pandemic COVID-19 when face masks were worn by majority of the population in order to stop the spread of the corona virus. Since FR models were failing to recognize masked faces an improvement therfore a new database of this kind was a necessity. This dataset contains 95K front-face images of 525 public figures. 5K pictures with a mask and 90K without a mask. All images are croped to face area by semi-automated annotation tool [24]. Examples can be seen in Figure 2.1.

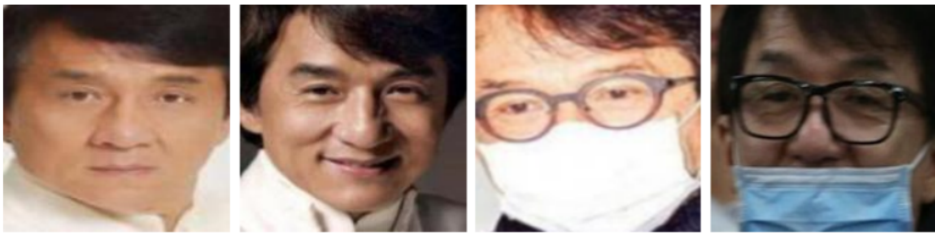

Figure 2.1: Examples of masked and unmasked faces from RMFRD

Source: Masked Face Recognition Dataset and Application [24]

SMFRD(2020) Simulated masked face recognition dataset was created together with RMFRD in order to expand the volume and diversity. This dataset was constructed by generating masks with software created by authors of [24] inspired by Dlib library on preexisting datasets, LFW and CASIA-Webface. It consists of 500,000 images of 10,000 subjects [24]. Examples can be seen in Figure 2.2.

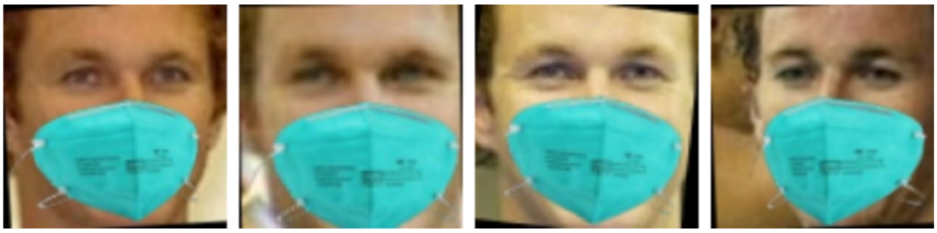

Figure 2.2: Example images with generated mask from SMFRD

Source: Masked Face Recognition Dataset and Application [24]

WebFace (2021) contains a million-scale face benchmark WebFace260M and a training data WebFace42M. It was created with the idea to close the gap between research and commercial face recognition networks owned by companies which have private access to large datasets like Google or Facebook. WebFace260M consists of 260M images of 4M celebrities. From these a training dataset WebFace42M was created by using Cleaning Automatically by Self Training pipeline which kept only high-quality images resulting in 42M images of 2M identities [28].

MS1M(MS1M-RetinaFace) is a training dataset cleaned from the MS-Celeb-1M. It contains 5.1M images of 93K identities. All images are preprocessed to the size of 112 × 112 by the five facial landmarks predicted by RetinaFace. Afterwards, a semi-automatic refinement is conducted by employing the pre-trained ArcFace model and ethnicity-specific annotators [7].

Glint360K is training dataset cleaned from the MS-Celeb-1M and Celeb-

500k dataset. It contains 17M images of 360K individuals, which is one of the largest and cleanest training datasets in academia. All face images are preprocessed to the size of 112 × 112 by the five facial landmarks predicted by RetinaFace. Then, an automatic refinement is conducted by employing the pre-trained ArcFace model for intra-class and inter-class cleaning [7].

2.5.2 Augmenting data to include masked faces

Even though available databases improved significantly over the years there is yet no publicly available large-scale masked face recognition training set available. Thus augmentation of data is necessary.

One option is to use simple mask generation as suggested in [5], consisting of covering part of the face with a uniform color using 68 facial key points. Example result of this algorithm can be seen in Figure 2.3.
Code for this: *GenerateFaceMask.py* For this concrete algorithm the face detection is done only for front facing faces.

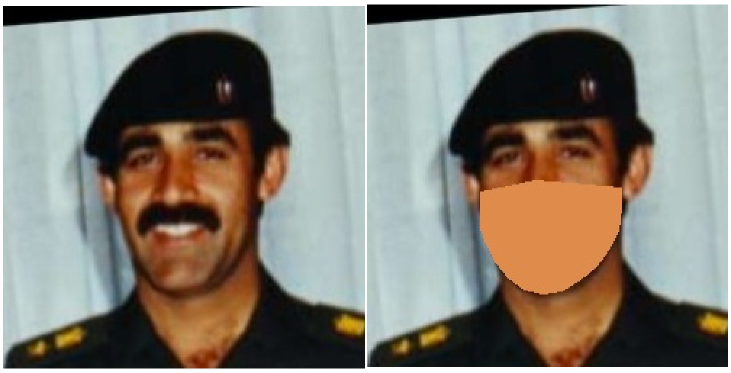

Figure 2.3: Simple face mask generation on a face from LFW

Second, more sophisticated option would be to use algorithm described in Masked Face Recognition Challenge [7] using texture blending in the UV space: Given an unmasked face and a real picture of a face mask, they perform 3D reconstruction of the face, obtain the UV texture map, the face geometry and the camera pose. Then they map the face mask onto the UV space, blend the textures and using the face geometry render the masked face back into 2D image. Illustration of this algorithm can be seen in Figure 2.4. Code for this: `https://github.com/JDAI-CV/FaceX-Zoo/tree/main/addition_module/face_mask_adding/FMA-3D`.

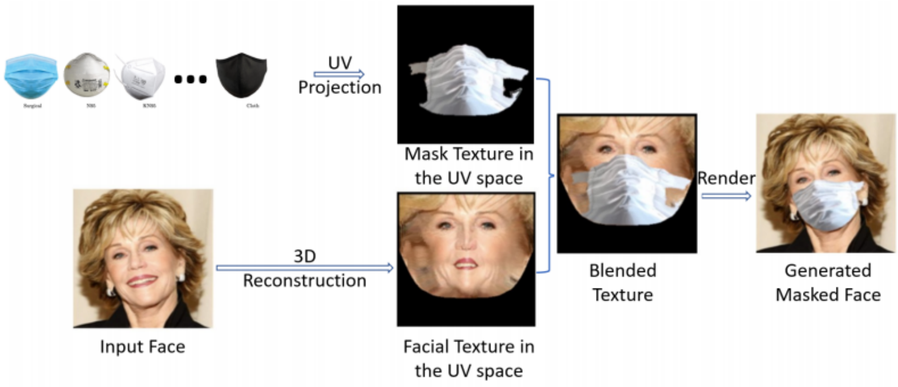

Figure 2.4: Mask generation using texture blending in the UV space.

Source: Masked Face Recognition Challenge: [7]

2.6 Masked Face Recognition

During the recent years, especially since the raise of artificial intelligence techniques and the broad use of digital cameras and smartphones, automated face identification gained a lot of popularity. Not only because it is the most natural way for human recognition, but compared to other methods of biometrics, face recognition is non-invasive and non-contact [1]. In fact it does not need any participation of the subject at all, which can be viewed both as a positive and a negative. On one hand subject does not need to make any effort to be identified but on the other hand it can be used without the subject's knowledge. Remarkably, both of these factors increase the popularity of face recognition. However, existing face recognition systems are trained to recognize non-occluded faces, which include primary facial features such as the eyes, nose, and mouth [7]. Thus, situations where people have part of their face covered, for example with a face mask, poses a huge challenge to existing face recognition systems.

Article about MFR Challenge [7] states that generally, there are two kinds of methods to overcome masked face recognition: (1) recovering unmasked faces for feature extraction and (2) producing direct occlusion-robust face feature embedding from masked face images. Additionally [23] claims that an important step for effectively recognizing masked and occluded faces is detecting if facial mask is present.

2.6.1 Masked Face Recognition Challenge

After the global pandemic COVID-19 in year 2019 when everyone needed to wear a face mask in public, the Masked Face Recognition (MFR) Challenge was introduced. This challenge consists of two main tracks: the InsightFace track and the WebFace260M track. Each of the two tracks has a collection of large-scale datasets for testing that include masked adults, children and a multi-racial test set. The goal of this challenge is to provide a comprehensive evaluation of CNN face recognition models. They introduce a new benchmark for masked face recognition as well as non-masked face recognition. By not allowing pre-trained model and giving rules on fixed training data and strict constraints on computational complexity and model size they enable fair performance comparison between different models. Data augmentation for the facial mask is allowed but the augmentation method needs to be reproducible [7].

InsightFace track

For training they employ two existing datasets MS1M (in other sources also called MS1M-RetinaFace [8]) and Glint360K.

For the test set, a large-scale set of real-life masked and unmasked faces with 7K identities was manually collected. In addition, they created a children test set including 14K identities and a multi-racial test set containing 242K identities.

The 1:1 face verification is employed as a evaluation metric. Each testing set is evaluated separately. For multi-racial test set accuracy is assessed by demographic groups [7].

WebFace260M track

[27]

Chapter 3

Current facial recognition technologies (FRT)

For us humans it is most natural to identify one another by recognizing the face of an individual (if we have sight). Therefore, even though it may be less precise compared to other biological triads, like fingerprints, it is our first choice. Thus the idea of face recognition has existed for a very long time. First, we could recognize people only from their physical presence, then paintings and other visualizations came along and later on with the invention of photography we started creating "databases" of identities. Not only for personal use, these collections have been used in forensic examinations, as referential databases, to find the identity of an individual [1]. A photography comparison as evidence in order to verify a person's identity was used in an English court as early as 1871 [18]. Needless to say, back then the comparison was done manually by humans. This was even before forensic techniques for this kind of face recognition were yet to be born. Since then technology completely changed the way we view facial recognition and widened the possibilities of its usage.

3.1 Usage of FRT

3.1.1 In EU or SR

3.2 Regulations and data protection

Naturally, this technology raises many concerns. The idea of widespread surveillance, fear of breach of data, privacy and personal freedom. Therefore regulations and data protection are a must as we discuss face recognition.

3.3 Current models

As shown in section 3.1, face recognition technologies have very wide usage, that is only growing, therefore there is no surprise that the the top algorithms on the market are developed by powerful companies like Facebook or Google. However these models are not available to the public.

3.3.1 DeepFace

Face recognition system DeepFace was introduced in 2014 by Facebook research team [21]. Their model was a revolution in FR technologies due to the fact that they were the first that managed to close the gap between human and machine accuracy in unconstrained face recognition.

First they address the face detection step, precisely face alignment. They employ explicit 3D face modeling in order to apply a piecewise affine transformation. In other words, they apply analytical 3D modeling of the face based on fiducial/key points, that is used to warp a detected facial crop to a 3D frontal mode, so called frontalization. They use simple fiductial point detector based on Support Vector Regressor (SVR) trained to predict point configurations from an image descriptor and apply it in several iterations to refine its output (Nowadays this can be done by DNN as well). Their alignment process has three steps: 2D alignment that starts by detecting 6 key points (centers of eyes, tip of the nose, corners and center of mouth) inside the cropped face image and then uses them to approximately scale, rotate and translate the image into six anchor locations. 3D alignment for handling all faces undergoing out-of-plane rotations. This is done by using a generic 3D shape model and a 3D affine camera to warp the 2D-aligned crop to the image plane of the 3D shape by adding an additional 67 fiducial points and manually matching them on the 3D shape. Lastly frontalization is achieved by a piece-wise affine transformation directed by the Delaunay triangulation derived from the 67 fiducial points.

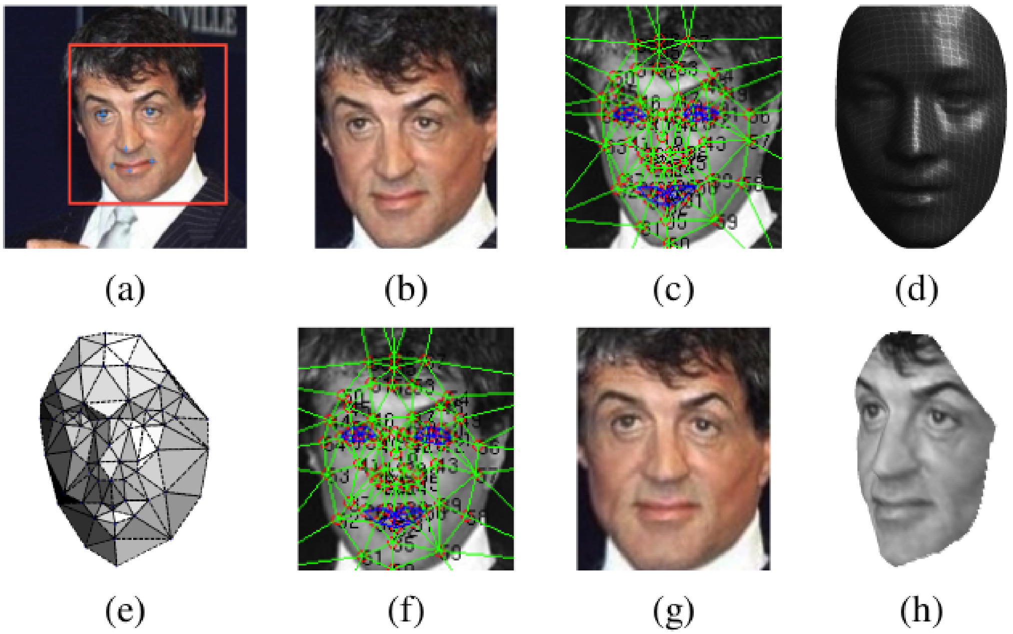

Figure 3.1: DeepFace alignment pipeline

Source: DeepFace [21]

To derive the face embeddings they use nine-layer deep neural network. The network architecture is based on the assumption that once the alignment is completed, the location of each facial region is fixed at the pixel level. Therefore the network learns directly from pixel RGB values. The inputs into their network structure are the 3D aligned face images of size 512x512. First they pass through a convolutional layer with 32 filters of size 11x11x3. The resulting 32 feature maps are then fed to a max-pooling layer which takes the max over 3x3 spatial neighborhoods with a stride of 2, separately for each channel. This is followed by another convolutional layer that has 16 filters of size 9x9x16. The purpose of these three layers is to extract low-level features, like simple edges and texture [21]. They are followed by three locally connected layers. These layers are very beneficial thanks to the fact that the input images are aligned; hence, the network can put more emphasis on parts of the face that are more important for distinguishing identity, like eyes. They can afford three large locally connected layers as a result of training on a large dataset of four million facial images belonging to more than 4,000 identities. The last two layers are fully connected. Afterwards the output goes into a K-way softmax which produces a distribution over the class labels. Learning to classify to the correct class is done by minimizing cross-entropy loss for each training sample. The network produces sparse face representations.

3.3.2 FaceNet

FaceNet was introduced in 2015 by Google researchers [19]. The main idea they present is to use DNN to learn the face embeddings in an Euclidean space such that the embedding itself is optimized and the squared L2 distances in the embedding space directly correspond to face similarity: faces of the same person have small distances and faces of different people have large distances. Thanks to this, the task of verification becomes simply thresholding of distances between embeddings and the task of identification becomes a k-NN classification problem.

They explore two different convolutional DNN architectures. The first one is based on the Zeiler&Fergus model with added 1×1×d convolution layers. It has 22 layers and a total of 140 million parameters and requires around 1.6 billion FLOPS per image. This model can be used in a data center where a lot of memory and processing power are available whereas the second one is much lighter therefor more suited for mobile devices. The second architecture is based on the Inception, GoogLeNet style models inspired by Szegedy et al. models. Compared to the first architecture this one is dramatically reduced in size. It has 20-times fewer parameters and up to 5-times fewer FLOPS.

To train they use a triplet based loss function based on Large margin nearest neighbor method. It trains the output to be compact 128-D embedding. The triplets they use are roughly aligned matching / non-matching face patches generated using an online triplet mining method which ensures consistently increasing difficulty of triplets as the network trains. The triplets consist of two matching face thumbnails and a non-matching face thumbnail and the loss aims to separate the positive pair from the negative by a distance margin. The thumbnails are simply tight crops of the face area with no additional alignment or transformation performed. The triplet loss tries to enforce a margin between each pair of faces from one person to all other faces. This allows the faces for one identity to live on a manifold, while still enforcing the distance and thus discriminability to other identities.

They achieved 99.63% accuracy in the task of face verification on the LFW dataset and 95.12% on YTF dataset.

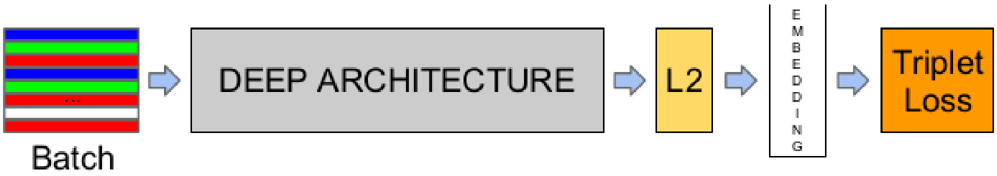

Figure 3.2: **FaceNet model structure.** The network consists of a batch input layer and a deep CNN followed by L_2 normalization, which results in the face embedding. This is followed by the triplet loss during training.

Source: FaceNet [19]

FaceNet also has an open source version called **OpenFace**.

3.3.3 YOLO, YOLOv3, YOLOv4, YOLOv5

3.3.4 DarkNet

3.4 Performance of current models on masked data

Chapter 4

Research

Bibliography

[1] I. Adjabi, A. Ouahabi, A. Benzaoui, and A. Taleb-Ahmed. Past, present, and future of face recognition: A review. *Electronics*, 9(8), 2020.

[2] M. Alghaili, Z. Li, and H. A. R. Ali. Facefilter: Face identification with deep learning and filter algorithm. *Scientific Programming*, 2020:7846264, Aug 2020.

[3] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper. Self-restrained triplet loss for accurate masked face recognition. *Pattern Recognition*, 124:108473, 2022.

[4] J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In S. A. Friedler and C. Wilson, editors, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pages 77–91. PMLR, 23-24 Feb 2018.

[5] A. Carragher, Daniel J.and Towler, V. R. Mileva, D. White, and P. J. B. Hancock. Masked face identification is improved by diagnostic feature training. *Cognitive Research: Principles and Implications*, pages 2365–7464, 2022.

[6] D. Cheng, Y. Gong, S. Zhou, J. Wang, and N. Zheng. Person re-

identification by multi-channel parts-based cnn with improved triplet loss function. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[7] J. Deng, J. Guo, X. An, Z. Zhu, and S. Zafeiriou. Masked face recognition challenge: The insightface track report. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, pages 1437–1444, October 2021.

[8] J. Deng, J. Guo, D. Zhang, Y. Deng, X. Lu, and S. Shi. Lightweight face recognition challenge. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops*, Oct 2019.

[9] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In B. Leibe, J. Matas, N. Sebe, and M. Welling, editors, *Computer Vision – ECCV 2016*, pages 87–102, Cham, 2016. Springer International Publishing.

[10] G. B. Huang, M. Mattar, T. Berg, and E. Learned-Miller. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Oct 2008.

[11] B. Institute. Biometrics institute - what is biometrics? `https://www.biometricsinstitute.org/what-is-biometrics/`.

[12] B. Institute. Types of biometrics: Face – use cases. `https://www.biometricsinstitute.org/types-of-biometrics-face-use-cases/`.

[13] ISO. Iso/iec 2382-37:2022(en) information technology — vocabulary — part 37: Biometrics. 2022.

[14] B. Klare, M. Burge, J. Klontz, R. Vorder Bruegge, and A. Jain. Face recognition performance: Role of demographic information. *Information Forensics and Security, IEEE Transactions on*, 7:1789–1801, 12 2012.

[15] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri. Face recognition systems: A survey. *Sensors*, 20(2), 2020.

[16] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song. Sphereface: Deep hypersphere embedding for face recognition. 04 2017.

[17] A. Nech and I. Kemelmacher-Shlizerman. Level playing field for million scale face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[18] G. Porter and G. Doran. An anatomical and photographic technique for forensic facial identification. *Forensic Science International*, 114(2):97–105, 2000.

[19] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 815–823, 2015.

[20] D. Sáez Trigueros, L. Meng, and M. Hartnett. Enhancing convolutional neural networks for face recognition with occlusion maps and batch triplet loss. *Image and Vision Computing*, 79:99–108, 2018.

[21] Y. Taigman, M. Yang, M. A. Ranzato, and L. Wolf. Deepface: Closing the gap to human-level performance in face verification. *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014.

[22] Thales. Facial recognition: top 7 trends (tech, vendors, use cases). `https://www.thalesgroup.com/en/markets/digital-identity-and-security/government/biometrics/facial-recognition`.

[23] M. Vrigkas, E.-A. Kourfalidou, M. E. Plissiti, and C. Nikou. Facemask: A new image dataset for the automated identification of people wearing masks in the wild. *Sensors*, 22(3), 2022.

[24] Z. Wang, G. Wang, B. Huang, Z. Xiong, Q. Hong, H. Wu, P. Yi, K. Jiang, N. Wang, Y. Pei, H. Chen, Y. Miao, Z. Huang, and J. Liang. Masked face recognition dataset and application. *CoRR*, abs/2003.09093, 2020.

[25] L. Wolf, T. Hassner, and I. Maoz. Face recognition in unconstrained videos with matched background similarity. In *CVPR 2011*, pages 529–534, 2011.

[26] H. W. F. Yeung, J. Li, and Y. Y. Chung. Improved performance of face recognition using cnn with constrained triplet loss layer. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pages 1948–1955, 2017.

[27] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, J. Zhu, T. Yang, J. Guo, J. Lu, D. Du, and J. Zhou. Masked face recognition challenge: The webface260m track report. *Creative Commons Attribution 4.0 International*, 2021.

[28] Z. Zhu, G. Huang, J. Deng, Y. Ye, J. Huang, X. Chen, J. Zhu, T. Yang, J. Lu, D. Du, and J. Zhou. Webface260m: A benchmark unveiling the power of million-scale deep face recognition. In *Proceedings of the*

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 10492–10502, June 2021.

List of Figures