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) [11] 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.

1.1 Goals

The goal of this thesis is to improve facial identification techniques 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?
- Are these systems able to identify a person that is wearing a mask even when there are no masked people in the database? If they are not, how can we achieve it?

2. Racial bias:

- How do current systems perform on people of different colors? I would like to do separate evaluation on different groups.
- How to reduce this bias?

TO DO : WHY IS ANSWERING THESE QUESTIONS IMPORTANT...
-nieco ako motivaciu prace - How do current systems perform on people of differt colors -> Toto je hlavne dolezite preto ze mame viac socialneho bias voci people of color(ZDROJ) a tympadom su castejsie odsudeny na sudoch. prepojit s tym ako sa momentalne pouzvaja identifikacia tvarii.
+ ukazat ze sa tomu nevenuje dostatok studii, ze tieto otazky neboli dostatocne zodpovedane

maybe add short intro on how facial recognition works. Like in here [17]
TO DO: Dodat organizaciu prace. aka co sa kde nachadza.

Chapter 2

Facial recognition

This chapter provides a short description of facial recognition 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.

How it works

–this wants too be a short section that describes in short the three major steps of FRT–

– this is just an outline – 1. Face detection - crop just the face from the picture

2. Extracting features from the image - we no longer work with the image, feature vectors is

3. Face matching

During the recent years face identification gained a lot of popularity.

Compared to other methods of biometrics face recognition is non-invasive and non-contact. In fact it does not need any participation of the subject, 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.

Mention stuff from [5] : However, existing face recognition systems are presented with mostly non-occluded faces, which include primary facial features such as the eyes, nose, and mouth. During the COVID19 coronavirus epidemic, almost everyone wears a facial mask, which poses a huge challenge to existing face recognition systems. Traditional face recognition systems may not effectively recognize the masked faces. ... 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

Mask occlusion may lead to obstruction of the feature structure of the face as certain parts of the face are hidden; thus, detecting facial masks is an important step for effectively recognizing masked and occluded faces in the wild.Wang and Kim [4] trained a convolutional neural network in real and simulated data of masked and unmasked faces to alleviate the problem of facial-mask detection. A novel approach that addressed the problem of masked face recognition by extracting deep features from the unmasked regions of the face and then using the bag-of-features paradigm to the learned feature maps was proposed in [5]. Finally, the visual attention mechanism was also employed in [6] to enhance the recognition accuracy by focusing on the regions around the eyes. [18]

2.1 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. [5]

InsightFace track

For training they employ two existing datasets MS1M (in other sources also called MS1M-RetinaFace [6]) 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. [5]

WebFace260M track

2.2 Viola Jones

–in this section shortly mention Viola Jones algorithm and how it contributed to FRT. That even nowadays some parts of it are used–

2.3 Deep neural networks

–in short describe how FRT with DNN works, what are the main steps used. Different variations like reconising only front facing face, handeling rotations. Similarities and differences of models (most use triplet loss function are there some that don't?)
Problems and struggles that are still in question whit this methods. –

"In early times, research interests were mainly focused on face recognition under controlled conditions where simple classical approaches provided excellent performance. Today, the focus of research is on unconstrained conditions in which deep learning technology has gained more popularity as it offers strong robustness against the numerous variations that can alter the recognition process. In addition, many academics struggle to find robust and reliable data sets for testing and to evaluate their proposed method: finding an appropriate data set is an important challenge especially in 3D facial recognition and facial expression recognition." [1]

The goal of training is to maximize the probability of the correct class(identity). [16] Which is achieved by minimizing chosen loss function and backpropagation of error.

2.4 Loss functions

2.4.1 Cross-enthropy loss

used in DeepFace algorithm.

2.4.2 Similarity measures

Matched Background Similarity (MBGS). This similarity is shown to considerably improve performance on the benchmark tests. [8]

2.5 Data

–the inportance of data
what databases are available
what isn't availabe
my modification of data for this research
from DeepFace it seems that it's good to have a lot of images per person in training
my question why is it necessairy or even good to clean data?–

" 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. For instance, LFW has about 75% males, celebrities that were photographed by mostly professional photographers. As demonstrated in [5], training and testing within different domain distributions hurt performance considerably and requires further tuning to the representation (or classifier) in order to improve their generalization and performance. However, fitting a model to a relatively small dataset reduces its generalization to other datasets" [16]

The large number of images per person provides a unique opportunity for learning the invariance needed for the core problem of face recognition. [16]. == for training is good to have many images per person

Is it really that some particular algorithm outperforming others or access to big/good training data is the key to success? The ImageNet competition [24] showed that neural networks [15] approaches dominate, and tend to perform better as 1) deeper networks are developed and 2) more data is provided to accurately tune network weights. It is important, therefore, for a benchmark to provide big enough data for algorithms to be successful.... Large-scale training optimization considers large numbers of samples per class where batching and online approaches, e.g., stochastic gradient descent, are valuable [4]. [14]

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. [16]. 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 at least 158 have more than 10 pictures in the database. [8]

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. [19]

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. [16]. 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. [7]

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. [14]

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. [21]

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 $\times$ 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. [5]

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. [5]

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. –Intro about not having enough data with masked faces available–

One option is to use simple mask generation as suggested in [4], consisting of covering part of the face with a uniform color using facial features points (I probably used wrong words). Example result of this algorithm can be seen in figure 2.1.

Code for this: *GenerateFaceMask.py* (I don't remember what was this code based on. I definitely did not start from scratch.)

Figure 2.1: Simple face mask generation on a face from LFW

Second, more sophisticated option would be to use algorithm described in Masked Face Recognition Challenge [5] 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 2.2. Code for this: `https://github.com/JDAI-CV/FaceX-Zoo/tree/main/addition_module/face_mask_adding/FMA-3D`

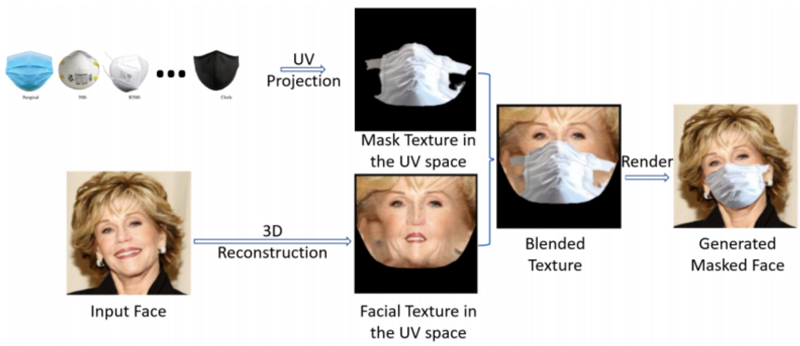

Figure 2.2: Mask generation using texture blending in the UV space.

Source: Masked Face Recognition Challenge: [5]

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 [15]. 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

–vsetky informacie a zdroje tu treba overit a podlozit este nejakymi dalsimi–

–fields of usage:

security

*phone/computer unlocking

*when issuing identity documents (probably not in SR)

*border checks - airport in Paris since 2018

*police checks - in US at least 26 states allow law enforcement to run searches against their databases of driver's license and ID photos. The FBI has access to driver's license photos of 18 states.

*surveillance in the public sector

**Find missing people

**Identify and track criminals.

**accelerate investigations

banking and retail (still kinda secutity)

*access to bank-account

*open a bank account

*facial recognition payment system (I am not sure if I belive this)

**Since 2017, KFC (American fast food), and Alibaba(Chinese retail and tech giant) have been testing a face recognition payment solution in Hangzhou, China.

** metro in Moscow(2021 -if they actually implemented it)

helth

*detecting some genetic diseases

*Patient check in and check out

*Access control
*Employee time clock - I hope this is not really being used
*Care-taking robots reading emotions (ofc emotions are complex so it's just matching the 6 basic emotions from classic emotoion theory)

=> corona has widened the usage of FRT
... TO BE CONTINUED...

3.1.1 In EU or SR

3.2 Regulations and data protection

3.3 Current models

–According to THALES:
1. Academia - **The GaussianFace algorithm** developed in 2014 by researchers at The Chinese University of Hong Kong - identification
2. Facebook - **DeepFace** - verification
3. Google -**FaceNet**
=> open source version **OpenFace**
4. Microsoft - "A study done by MIT (https://www.eurekalert.org/news-releases/587454) researchers in February 2018 found that Microsoft, IBM, and China-based Megvii (FACE++) tools had high error rates when identifying darker-skin women compared to lighter-skin men. At the end of June 2018, Microsoft announced that it had substantially improved its biased facial recognition technology in a blog post. " [17]–

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.

I will mention three of the most used algorithms: **DeepFace** developed by Facebook, **FaceNet** from Google and open-source version that arose from it, **OpenFace**.

3.3.1 DeepFace

3.3.2 FaceNet

3.3.3 OpenFace

3.3.4 YOLO, YOLOv3, YOLOv4, YOLOv5

3.3.5 DarkNet

3.4 Performance of current models on masked data

please check this out on Thales[17] later

March 2018 – The live testing done using more than 300 volunteers identified the best-performing facial recognition technologies.

More on performance benchmarks: The NIST (National Institute of Standards and Technology) report, published in November 2018, details recognition accuracy for 127 algorithms and associates performance with participant names.

The NIST Ongoing Face Recognition Vendor Test (FRVT) 3 performed at the end of 2019 provides additional results. See NIST report.

NIST also demonstrated that the best facial recognition algorithms have no racial or sex bias, as reported in January 2020 by ITIF. Critics were wrong.

In NIST's reports (August 2020 and March 2021) entitled "Face recognition accuracy with **face masks** using post-COVID-19 algorithms", we see how algorithms, in less than a year, are increasing their performance.

on general database that systems use

on racialy separated data

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] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

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

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

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

[12] 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.

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

[14] A. Nech and I. Kemelmacher-Shlizerman. Level playing field for mil-

lion scale face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

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

[16] 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.

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

[18] 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.

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

[20] 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.

[21] 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