UNIVERSIDADE D
COIMBRA

Joana Alves Pimenta

IMPACT OF IMAGE CONTEXT FOR DEEP LEARNING FACE MORPHING ATTACK DETECTION

Dissertation for the Master's degree in Biomedical Engineering under the scientific supervision of Prof. Nuno Gonçalves, MsC Iurii Medvedev and presented to the Faculty of Sciences and Technology of the University of Coimbra.

July, 2023

University of Coimbra

Impact of Image Context for Deep Learning Face Morphing Attack Detection

Joana Alves Pimenta

Dissertation for the Master's degree in Biomedical Engineering presented to the Faculty of Sciences and Technology of the University of Coimbra.

Supervisors:
Ph.D. Nuno Gonçalves
MsC Iurii Medvedev

INSTITUTE OF SYSTEMS AND ROBOTICS
UNIVERSITY OF COIMBRA

Coimbra, 2023

Agradecimentos

Gostaria de começar por agradecer ao Prof. Dr. Nuno Gonçalves, pela disponibilidade que teve ao longo destes meses, pelo profissionalismo e pela oportunidade que me deu em desenvolver e concretizar este trabalho.

Ao Iurii, quero agradecer sobretudo pela paciência para esclarecer todas as minhas dúvidas, pela orientação e apoio contínuo ao longo deste período e por todas as sugestões e ideias que foram fulcrais para o resultado deste trabalho.

À minha família, pelo apoio ao longo desta jornada académica, por me incentivarem e acreditarem no meu potencial, quando muitas vezes eu não acreditava. Em especial à Raquels por nesta fase final ter sido ela a irmã mais velha.

Por fim, mas não menos importante, aos meus amigos, que tiveram sempre as palavras certas ou necessárias para me dar, que me acompanharam ao longo de todo este percurso e que ainda aturaram as minhas frustrações e inseguranças. Bule, obrigada pelo teu otimismo e motivação e acima de tudo por teres estado sempre lá. Bia, obrigada por fazeres sempre com que tudo fosse mais leve. Filipa, Mari, Madalena obrigada por terem tornado estes últimos 5 anos mais felizes e cheios de boas memórias.

A todos, muito obrigado!

"Coimbra é o berço do conhecimento, onde a academia floresce e os sonhos ganham asas."

Resumo

A utilização do rosto como forma de identificar e verificar a identidade de um dado indivíduo tem impactado significativamente a expansão dos sistemas biométricos, em particular sistemas de reconhecimento facial, como medida de segurança. A face humana é por outro lado, extremamente suscetível a manipulações, tornando estes sistemas vulneráveis a ameaças e tentativas de ataque.

O *morphing* facial é um dos ataques mais preocupantes, na medida em que permite obter uma imagem de um indivíduo que aparenta ser real, mas que na verdade, não existe. Além disso, a imagem resultante é confundível com a face de dois ou mais indivíduos já que incorpora uma combinação das características faciais dos mesmos. Isto permite, por exemplo, que um atacante se faça passar por outra pessoa obtendo acesso não autorizado a informações ou sistemas sensíveis.

Por todos estes motivos a capacidade de detetar estes ataques é fundamental e tem sido alvo intensivo de estudo por parte de investigadores. Atualmente, a maioria das abordagens envolve o uso de algoritmos de aprendizagem profunda, que se têm mostrado eficazes em cenários mais realistas.

Nesta dissertação o objetivo principal passa por investigar a influência do contexto da imagem na deteção de ataques de *morphing* facial no caso particular de algoritmos de aprendizagem profunda. Para isso, propõe-se analisar o impacto das configurações de alinhamento da imagem na deteção. Isto é motivado pelo facto de o procedimento de alinhamento facial influenciar diretamente as interconexões entre o contorno do rosto e o contexto da imagem. Nesse sentido, a deteção eficaz pode ser alcançada através da obtenção de condições de alinhamento ótimas.

Palavras-chave: Reconhecimento facial, sistemas biométricos, ataque de *morphing* facial, algoritmos de aprendizagem profunda, alinhamento facial, deteção de ataques de *morphing* facial.

Abstract

The use of the face as a way to identify and verify an individual's identity has significantly impacted the expansion of biometric systems, particularly face recognition systems, as a security measure. However, the human face is extremely susceptible to manipulation, making the systems highly vulnerable to threats and attack attempts.

Face morphing is one of the most concerning attacks since it allows to obtain an image of an individual that appears to be real but, in fact, does not exist. Furthermore, the resulting image can be easily confused with the faces of two or more individuals, as it incorporates a combination of their facial characteristics. This allows, for instance, an attacker to impersonate another person and gain unauthorized access to sensitive information or systems.

For all these reasons, the ability to detect these attacks is crucial and has been the subject of intensive study by researchers. Currently, most of these techniques use deep learning algorithms, which have demonstrated effectiveness in realistic scenarios.

In this dissertation, the main goal is to investigate the influence of image context on the detection of face morphing attacks in the particular case of deep learning algorithms. In that regard, it is proposed to analyze the impact of the image alignment settings on the detection of these attacks. This is motivated by the fact that the face alignment procedure directly influences the interconnections between the face contour and image context. Thus, effective detection can be achieved by obtaining optimal alignment conditions.

Keywords: Face recognition, biometric systems, face morphing attack, deep learning algorithms, face alignment, face morphing detection.

Contents

List of Figures

List of Tables

List of Abbreviations

ABC Automatic Border Control. 4, 11, 20, 36
ABIS Automated Biometric Identification System. 12
AGIR Average of the Gradient Intensity Ratio. x, xi, 57, 65, 67, 69, 70
APCER Attack Presentation Classification Error Rate. ix, x, xi, 20, 21, 59, 60, 61, 62, 63, 65, 67, 68, 71, 72, 73, 74, 75, 76

BPCER BonaFide Presentation Classification Error Rate. ix, x, xi, 20, 21, 59, 60, 61, 62, 63, 65, 67, 68, 71, 72, 73, 74, 75, 76
BSIF Binarized Statistical Image Features. 33, 36

CI Complementary Image. 53
CNN Convolutional Neural Network. 25, 31, 34, 36, 40, 56, 58

DCNN Deep Convolutional Neural Network. 34
DeepID Deep Hidden IDentity features. 25
D-MAD Differential Morphing Attack Detection. viii, ix, xi, 18, 19, 35, 41, 53, 55, 68, 69, 70, 77, 78
DET Detection Error Trade-off. vii, viii, ix, 13, 14, 60, 61, 62, 63, 65, 66, 67, 68, 71, 72

EER Equal Error Rate. 14
eMRTD Electronic Machine Readable Travel Document. 2, 4, 30

FC Fully Connected. 26, 51
FD-GAN Face Demorphing GAN. 35, 36
FMR False Match Rate. 14
FNIR False Negative Identification Rate. 15
FNMR False Non Match Rate. 14
FNR False Negative Rate. 13, 14
FPIR False Positive Identification Rate. 14, 15
FPR False Positive Rate. 13, 14

FR Face Recognition. vii, 5, 6, 7, 9, 10, 11, 13, 22, 23, 24, 25, 26, 32, 34, 35, 36, 37, 40, 41, 43, 47, 70, 77
FRS Face Recognition System. iv, 2, 3, 4, 5, 7, 9, 10, 13, 22, 24, 25, 33, 40, 41
FRVT Face Recognition Vendor Test. xi, 34, 53, 70, 71, 77
FS-SPN Fourier Spectrum Of Sensor Pattern Noise. 34, 36

GAN Generative Adversarial Network. vii, 17, 31, 32, 33, 35, 41, 47
GPP Gabor Phase Pattern. 24
Grad-CAM Gradient-weighted Class Activation Mapping. viii, ix, 56, 64, 66, 67, 69, 70

HGPP Histogram of Gabor Phase Patterns. 24, 29
HOG Histogram Of Gradient Orientations. 24, 29

ICA Independent Component Analysis. 23
ICAO International Civil Aviation Organization. 2, 7, 43, 77
ID Identity Document. 2, 4, 11, 42, 78
iMARS Integrated Monitoring, Analysis and Response System. 33
INCM Imprensa Nacional-Casa da Moeda. 5, 33
ISR Instituto de Sistemas e Robótica. 5

k-NN k-Nearest Neighbor. 24
KYC Know Your Customer. 2

LBP Local Binary Patterns. 24, 29, 33
LDA Linear Discriminate Analysis. 23
LMCL Large Cosine Margin Loss. 27
LPQ Local Phase Quantization. 33

MAD Morphing Attack Detection. vii, x, 5, 6, 7, 9, 18, 33, 34, 36, 41, 42, 52, 53, 71, 77
MPS Max-margin Pairwise Score. 29
MTCNN Multi-Task Cascaded Convolutional Neural Network. 23, 48

NIST National Institute of Standards and Technology. xi, 5, 34, 53, 70, 71, 77

P-CRC Probabilistic Collaborative Representation. 34
P-LBP Pyramidal Local Binary Pattern. 34, 36
PCA Principal Component Analysis. 23
PRNU Photo Response Non-Uniformity. 33, 36

ReLU Rectified Linear Units. 24

ROC Receiver Operating Characteristic. vii, 13, 14

S-MAD Single Morphing Attack Detection. viii, ix, x, xi, 18, 19, 33, 41, 50, 52, 53, 59, 60, 61, 62, 63, 64, 65, 66, 67, 69, 70, 77
SOTAMD Securing Online Transactions against MitM Fraud. 33
SPN Sensor Pattern Noise. 34
SVM Support Vector Machine. 33

TPR True Positive Rate. 13, 14

1

Introduction

This chapter provides an introduction to this dissertation. In sections 1.1 and 1.2 an overview of the context and motivation that underlie the proposed dissertation are presented. The goals and contributions are defined in sections 1.4 and 1.5. Finally, in section 1.6, the structure of the dissertation is outlined.

1.1 Context

Modern society's security concerns have been on the rise due to technological advancements. Most security systems used today still rely on the use of passwords, usernames, and signatures. These traditional identification techniques have become less reliable due to their vulnerability to being forgotten, misplaced, duplicated, or stolen, thus compromising their intended security function.

The solution to this problem seems to be biometric approaches, which use physiological or behavioral characteristics to enhance the recognition process [8].

In greater detail, biometric image modalities such as fingerprints, iris patterns, and facial features are used as input in image processing algorithms, thereby improving the robustness and reliability of the recognition process when compared to previous methods.

Biometric systems typically involve two general steps: *enrollment* and *verification* [9]. During the *enrollment* stage, biometric information is captured and processed to generate a representative template of the unique biometric features of a given individual. In some cases, the template can be stored in a database for further comparison during the *verification* stage. However, there are also scenarios without persistent storage where the template is only used for real-time comparison.

During the *verification* stage, new biometric information is compared to this reference template to determine whether or not a match exists. The outcome is a similarity score, which indicates the degree of resemblance between the two templates.

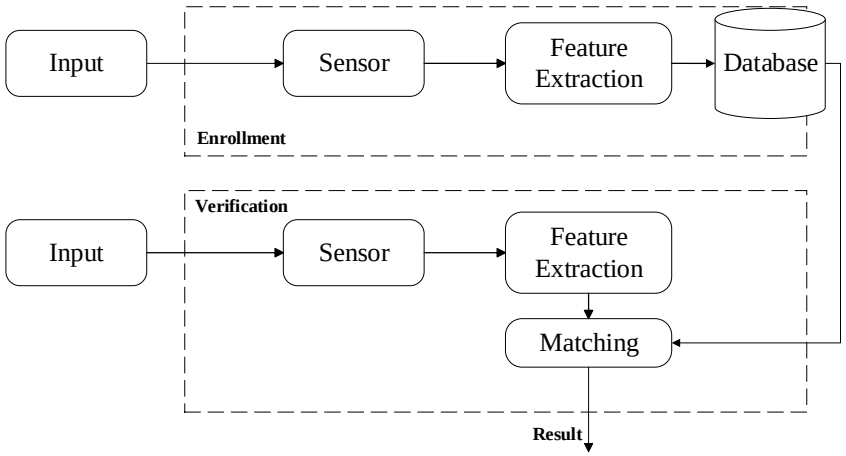

Figure 1.1: Schematic representation of the overall architecture of a biometric system.

The human face serves as a unique link to an individual's identity. Therefore, motivated by the simplicity of face image acquisition, recent advances in computer vision techniques, the non-invasive nature of the process, and the fact that it does not require contact, which is a critical factor after the recent pandemic, the use of the face as a biometric modality has been dominant in modern biometric applications.

One of the most notable examples of its application is the Face Recognition System (FRS), which utilizes facial traits for identification or verification purposes [10].

Currently, FRSs are used in a variety of applications, such as document security, border control systems, and policies such as Know Your Customer (KYC) in banks and financial institutions. This extensive usage was something unexpected during the period of initial work on automatic FRSs in the early 1990s.

Motivated by all this growth, many European countries include a face image in their civil Identity Documents (IDs) to use it as the main form of identification. Additionally, the International Civil Aviation Organization (ICAO) proposed to add a reference face image in the Electronic Machine Readable Travel Document (eMRTD) used in travel documents for the same purpose [11]. The face image must be *ICAO-compliant*, which means that it must meet specific standards and guidelines to ensure the interoperability and global acceptance of electronic passports and the biometric data contained in them.

Despite all of these advances, knowing how to deal with the high variability of the face (such as face expression, aging, lighting conditions, and head rotation) is still a challenge.

Nowadays, with the development of machine learning techniques, namely deep learning techniques, FRSs are becoming more sophisticated, thereby addressing some of these challenges. However, the general imperfection of FRSs and their probabilistic nature make them targets for threats, raising serious concerns about the use of face in biometric approaches [12].

From a general standpoint, the human face can undergo several modifications, including age or gender changes, face merging, and the introduction of perturbations, among others [13]. Such modifications significantly increase the vulnerability of FRSs since their primary purpose is to deceive and potentially expose them to fraud attempts and criminal attacks. Certain types of attacks allow falsifying data to achieve an illicit advantage, such as efficiently identifying one individual as another, *i.e.,* disguising the real identity (spoofing attack) [14].

Currently, one of the most powerful types of threats is the face morphing attack. The goal of this kind of attack is to combine the facial features of two (or more) images so that the resulting synthetic reference image incorporates characteristics from both faces [15]. Figure 1.2 depicts a morphed image created using two faces from different individuals.

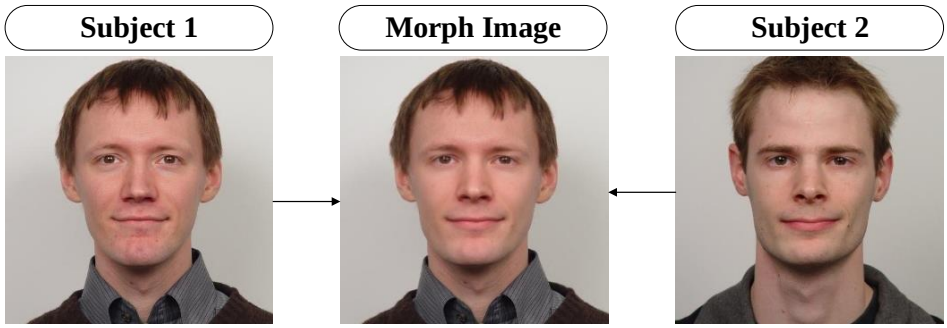

Figure 1.2: Schematic representation of face morphing between two subjects face images. Images from IMM dataset [1].

This smooth transition between the two faces causes great challenges for both humans and FRSs since the resulting image is simultaneously similar to the images of faces that gave rise to it, making it difficult to tell "where one face ends and the other begins".

One of the most popular face morphing applications happens during the passport application process. In this scenario, although live enrollment is preferred, it is not always possible, and in some countries (in Portugal, this is not the case), applicants often have to provide their images, either printed or via e-mail, or even use specialized state photographers for this service [16]. All these forms make the

submitted image highly susceptible to falsification, allowing morphing attacks to go unnoticed, *i.e.,* one individual can impersonate another, thus violating the principle of exclusive ownership. Summing up, it makes it possible to obtain a legitimate ID using false information.

As a consequence, during border control scenarios, both control agents and Automatic Border Control (ABC) gates become vulnerable to attack attempts since they are based on comparing a live image with an image stored in a ID or eMRTD that may be modified.

A real example was the case of an Albanian individual who attempted to pass through border control using a Slovenian passport [17]. The incident is depicted in figure 1.3, where the left side shows the accomplice of the individual, the right side shows the person attempting to pass through border control, and the center displays the transformed image that was presented on the document.

Figure 1.3: Real case example of a face morphing attack.

The challenges posed by face morphing attacks have made the detection of such attacks an important area of research. In consequence, significant efforts have been made to investigate and develop techniques that can detect and counteract the effects of these attacks.

1.2 Motivation

The ability to accurately recognize a person's face makes face recognition technology a predominant form of personal identification. Nevertheless, the susceptibility of FRS to attacks, specifically face morphing attacks, requires the development of effective detection techniques.

The main strength of a face morphing attack lies in its capability to obtain a face image of a subject that appears to be real but, in fact, does not exist. This opens up the potential for the creation of "multiple identities", where two (or more) different subjects can use the same ID to impersonate the same person.

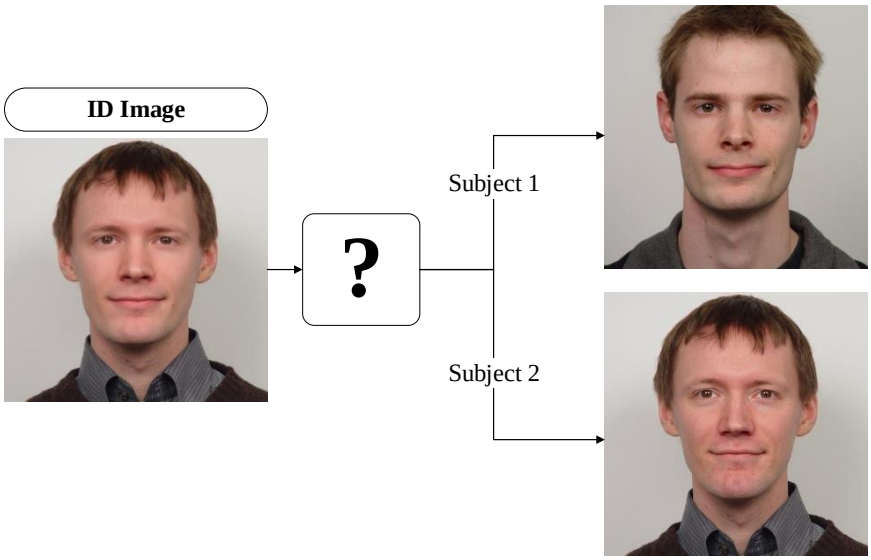

Figure 1.4: Schematic representation of face morphing problem.

For all these reasons, the ability to detect face morphing attacks is crucial to guaranteeing the secure operation of FRS.

In order to deal with that and following the *FACING* project [18] that ran from 2019 to 2021, the Computer Vision Team from the Instituto de Sistemas e Robótica (ISR) within the University of Coimbra, in partnership with the Imprensa Nacional-Casa da Moeda (INCM), is developing the *FACING2* project. The main goal is to create an automatic system for authenticating people through Face Recognition (FR) of photographs.

The *FACING2* project aims to study and develop methods to improve facial biometric technology while respecting privacy and security standards. In a more detailed view, the *FACING2* project intends to act in six areas:

- Improve the present implementations of the algorithms for evaluating *ICAO-compliant* [11] and image quality.
- Improve the face identification and verification algorithms, which will subsequently be tested against the National Institute of Standards and Technology (NIST) benchmark [19].
- Improve liveness detection (check whether a human being facing a camera is, in fact, real) algorithms to protect against presentation attacks.
- Implement a Morphing Attack Detection (MAD) system that can be incorporated into the *FACING2* project's overall system.
- Implement new protection algorithms for biometric templates, which now include features such as irreversibility, recognition quality, renewability, and non-

correlation.

- Improve the current *FACING* application to increase the security of the technology.

Taking into account the current scenario of FR and the specific context of face morphing and its detection, this dissertation aims to contribute to advances in the study of MAD more specifically by investigating whether or not image context influences the detection. Throughout this dissertation, the term "image context" refers to the background and surrounding elements.

This motivation arises from the understanding that the performance of face morphing detection can be influenced by several factors, including the alignment and pre-processing techniques applied to the input images. In the particular case of face alignment, the alignment settings can have an impact on the amount of contextual information captured in the input image. Consequently, this can hypothetically affect the performance of the detection algorithm.

1.3 Ethical Concerns

The development of face morphing technology raises serious ethical concerns, particularly when the use of morphed images is not limited to specific purposes, *i.e.,* closed environments.

As presented in the previous sections, the creation of manipulated images through face morphing has the potential to deceive both FR systems and people, requiring research.

In this sense, there is clearly a need to create morph images to train systems and improve models, making them more robust and able to detect these types of images and, consequently, prevent attacks. However, this also raises ethical issues since, in order to achieve this goal, it is necessary to improve or even create methods to generate these morphed images. The open publication of these approaches poses significant risks as it allows for their unrestricted and uncontrolled utilization.

To ensure a responsible and ethical use of face morphing technology, it is crucial to establish guidelines and restrict its application to the scope of research institutions, official state organizations, and their unions.

1.4 Goals

The overall objective of this dissertation, as stated in the section 1.2, is to relate image context to the detection of face morphing attacks, trying to verify its

influence, in particular:

- Finding the best context properties for detection, *i.e.,* defining optimal alignment settings for face morphing detection.
- Exploring how different face image alignment settings can impact the amount of context captured in the input image.

1.5 Contributions

To achieve the aforementioned goal, the following specific contributions are outlined below:

- Creation of a large dataset that adheres to the ICAO standards through the combination and pre-processing of multiple datasets.
- Generation of a morphed dataset using both landmark-based and StyleGAN-based approaches.
- Investigation of the relationship between image context and MAD to identify the most effective context properties for detection.
- Formulation and implementing several strategies for MAD.
- Development of more robust and reliable face morphing detection algorithms that are less susceptible to manipulation and errors, ultimately contributing to the creation of more secure systems for identifying individuals and preventing identity fraud.

The work developed in this dissertation also allowed the development and respective submission of a paper titled "Impact of Image Context for Single Deep Learning Face Morphing Detection" in the BIOSIG 2023 conference that will be included in the Appendix.

1.6 Outline of the Dissertation

This document contains five chapters beyond the introduction that are organized as follows:

- Chapter 2: presents background information related to FRS, face morphing, and its detection, as well as some other important concepts.
- Chapter 3: presents a review of the state of the art in FR, as well as face morphing techniques and the different approaches for detecting these attacks.
- Chapter 4: presents an overview of the methodology.
- Chapter 5: depicts the experimental results of the tests conducted during the

development of this dissertation.

- Chapter 6: presents the conclusions of the work performed.

2

Theoretical Background

Throughout this chapter, essential aspects and concepts concerning Face Recognition System (FRS), face morphing, and Morphing Attack Detection (MAD) will be presented.

2.1 Face Recognition System

Face Recognition (FR) is a complex area of study within the domains of Computer Vision and Biometrics that focuses on theoretical approaches and software tools enabling machines to recognize people based on their facial features, which are strongly linked to each individual's unique identity [10].

Despite the challenges presented by variations in many factors, such as illumination, pose, facial expressions, and aging, significant progress has been made in developing accurate and reliable face recognition algorithms.

Currently, FR approaches rely on deep learning methods [2], in which deep networks of several layers learn data representations with different feature extraction levels. The standard pipeline for these approaches is depicted in figure 2.1.

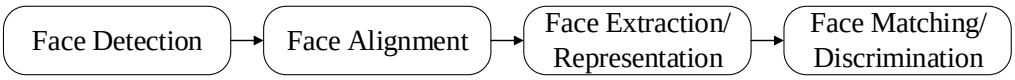

Figure 2.1: Standard pipeline for a deep-learning approach in FR [2].

In sequence, the first stage involves detecting the presence of a face in the image or video, followed by the alignment of that face.

There are several methods that can be employed for the alignment procedure. One commonly utilized approach is to define a bounding box around the face and utilize a predetermined set of facial landmarks, such as the eyes, the tip of the nose, and the mouth's corners, to determine the position and orientation of the face. Subsequently, the image can be transformed in accordance with predefined

coordinates and positions using rotation, scaling, and translation approaches.

In an ideal scenario, this alignment procedure should be able to handle face fluctuations, including variations in position, illumination, expression, and occlusion, that turn out to be frequent conditions in real-world contexts.

Subsequently, the aligned image is used as input to the system, resulting in a feature vector that represents it. This phase is also known as "Face Representation", since it involves the transformation of a face image into a learned feature space in a structured manner, simplifying the recognition task.

The features can be divided into: *Low-level* features, that represent the information extracted from the pixels using algorithms such as color histograms, gradient orientation, or texture descriptors; *High-level* features, that represent the interpretation and understanding of the image content from a more realistic perspective. In general, higher layers learn higher-level abstractions and are often more discriminative than lower-level features.

In the FR task, each distinctive feature vector can be used comparatively to evaluate similarity against another face representation or it can be used in identity classification tasks, depending on the scenario.

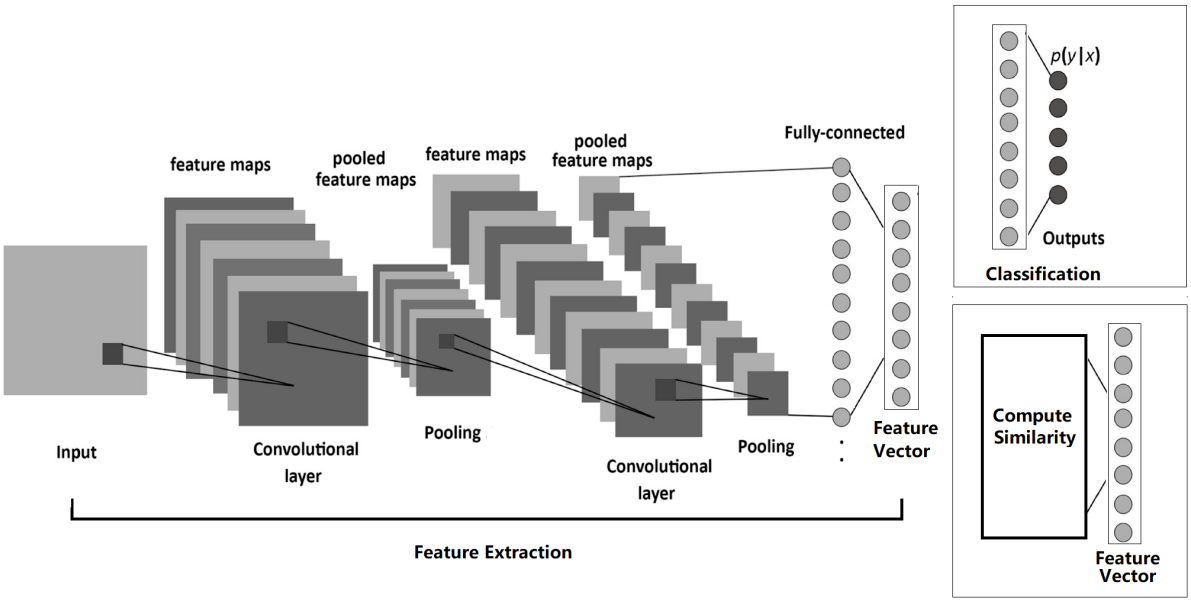

Figure 2.2: Diagram for a generic deep learning-approach FR for both scenarios.

2.2 Face Recognition Scenarios

From a general perspective, a FRS is usually used for two primary tasks [20]:

- Verification (*One-to-One Matching*)

- Identification (*One-to-Many Matching*)

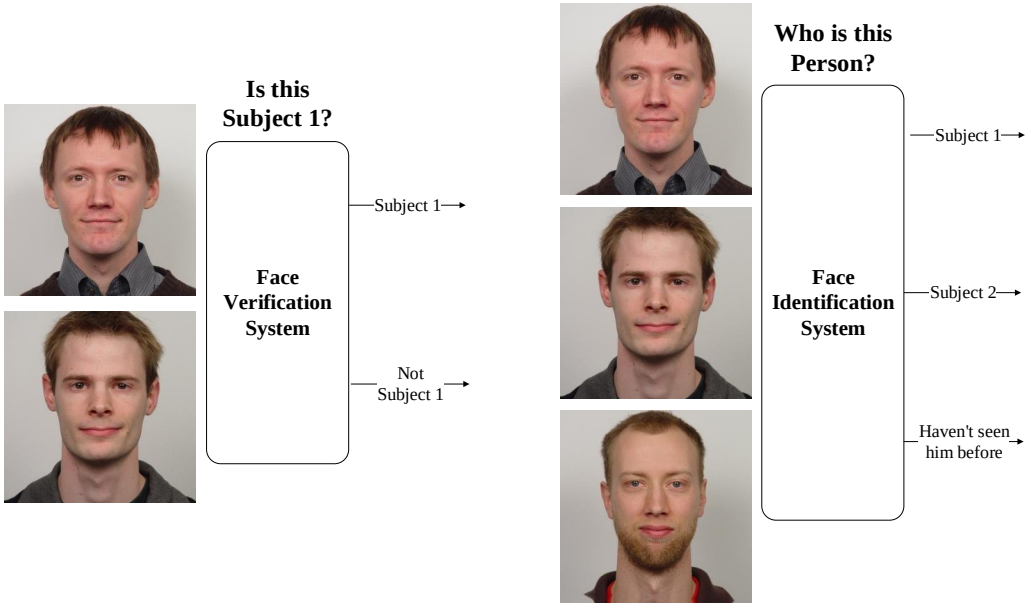

Figure 2.3: Schematic representation of face verification and face identification differences. Based on [3].

In both cases, the system starts with the common step of acquiring a face descriptor in the form of a feature vector and subsequently diverges into different paths, according to the scenario (figure 2.2).

2.2.1 Verification Scenario

Also known as *one-to-one matching*, the verification scenario process involves determining whether or not an individual is who he/she claims to be. In the context of FR, the main challenge is to verify whether two images depict the same person. This is a binary decision problem, with the outcome being a determination of whether the images match or do not match.

One of the most evident real-life scenarios of its application happens at Automatic Border Control (ABC) gates, where a live captured image is compared against a photo from an Identity Document (ID) or passport.

In practical terms, the process of face verification can be executed through the application of the following series of steps (figure 2.2):

- Obtain the face descriptor, which is a set of distinctive features.
- Match the face descriptor against a pre-existing representation of the individual's face, which may be stored in a database or extracted from an ID, to generate a similarity score.

- Based on a predetermined threshold value for the similarity score, the system can then verify or reject the individual's claimed identity.

2.2.2 Identification Scenario

Also known as *one-to-many matching*, the task of face identification is based on determining an individual's identity from a system database. The primary goal of this process is to answer the question: *who is the person?* or *is the person in the database?*

In real-world applications, this scenario is commonly seen in enterprises that conduct identification checks as part of the authorization process (Automated Biometric Identification System (ABIS)).

In the context of the identification scenario, it is possible to specify two particular cases: **open-set** and **close-set**.

The **close-set** case pertains to situations where the test identity corresponds to one of the pre-existing identities stored in the database. These cases can be approached as multi-class classification problems, being executed through the following series of steps (figure 2.2).

- Obtain the face descriptor, which is composed of a set of unique features.
- Process the face descriptor through fully connected layers.
- Perform classification on the extracted features to identify the individual in the input image (the closest matching identity).

In opposition, the case **open-set** denotes situations where the test identity may or may not correspond to one of the pre-existing identities stored in the database. This scenario is more representative of real-world authentication systems, which often have to reject unidentified subjects who are not registered in the system. In this particular case, face identification can be approached as a face verification problem, comparing the test face against all the identities stored in the database.

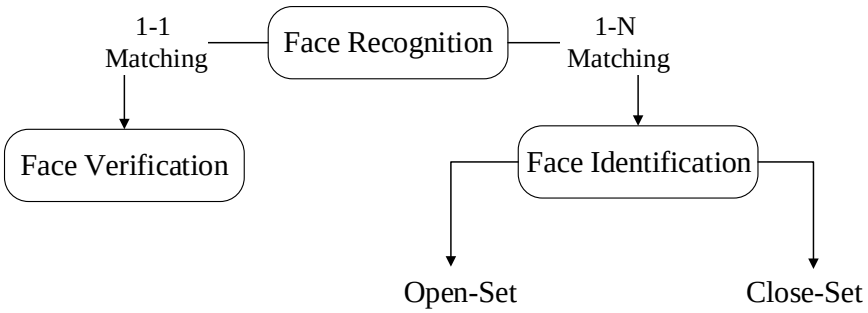

Figure 2.4: Summary diagram of FR scenarios.

2.3 Performance Evaluation in FRS

To evaluate the effectiveness and efficiency of a system, it is essential to compute performance metrics. In the case of FRSs these metrics differ based on recognition scenarios: identification and verification. All of the metrics are defined based on the concepts outlined in *ISO/IEC 19795-1 2006* [21].

2.3.1 ROC and DET Curves

Receiver Operating Characteristic (ROC) and Detection Error Trade-off (DET) curves are graphical representations that describe the trade-off between the False Positive Rate (FPR) and the False Negative Rate (FNR) of a binary classifier at different decision threshold levels [22].

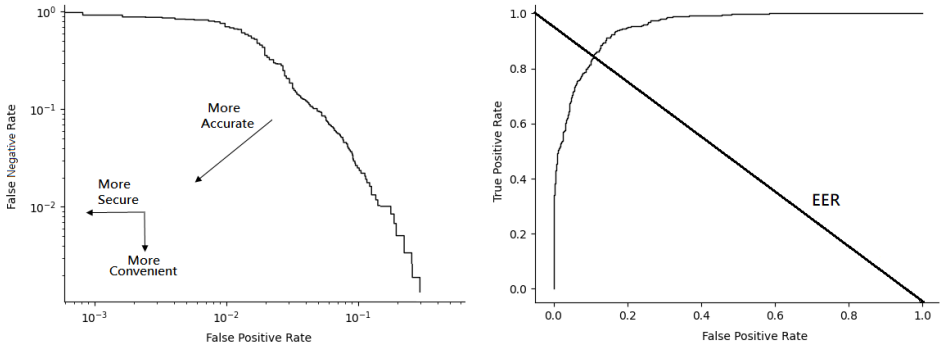

Figure 2.5: Schematic representation of the DET (left) and ROC (right) curves.

More specifically, ROC curves plot the FPR on the x-axis and the True Positive Rate (TPR) (1 - FNR) on the y-axis for different decision thresholds. In an ideal scenario, the ROC curve should be close to the upper left corner of the graph, reaching high TPR values while maintaining a low FPR across a range of decision

thresholds.

In the DET curves, it is common to represent both FPR and FNR in logarithmic scales. An optimal classifier will have a DET curve that is situated near the bottom left corner of the plot, which indicates low values for both the FNR and FPR across a range of decision thresholds.

Some other relevant metrics can be extracted from these curves, for instance: **Equal Error Rate (EER)**, which represents the point on the curve where FPR + TPR = 1, *i.e.,* where FPR is equal to FNR. It can be used to compare the performance of different systems, where a lower value is indicative of better performance.

2.3.2 Verification Scenario

In the verification scenario, two common metrics are used to evaluate the performance [22]:

- **False Match Rate (FMR)** - Represents the proportion of non-matching identities falsely identified as matching (security level).
- **False Non Match Rate (FNMR)** - Represents the proportion of matching identities falsely identified as non-matching (convenience measure).

Regarding the ROC and DET curves, it is possible to make the assumption that FMR is analogous to FPR, while FNMR is equivalent to FNR. The accuracy of a system's verification is usually presented as the value of FNMR fixing the FMR at certain thresholds (*e.g.* FNMR@FMR=0.01%).

2.3.3 Identification Scenario

In the context of identification, the choice of metrics depends on whether it is a **closed-set** or an **open-set** problem. Despite this difference, both approaches use a common metric known as the *identification rate at rank r*. This metric represents the probability within the list of top r matched identities during an identification attempt.

The identification rate at a particular rank level is an useful metric for evaluating the performance of a biometric system, as it provides an indication of the system's ability to accurately identify users. In general, the higher the value, the better the performance of the system.

In the specific case of **open-set** problem, another two metrics can be computed [22]:

- **False Positive Identification Rate (FPIR)** - Represents the proportion

of non-registered identities successfully attempted to be identified.

- **False Negative Identification Rate (FNIR)** - Represents the proportion of registered identities for which identification attempts failed.

A common way to report these metrics is by presenting the FNIR for a fixed FPIR at a specific *rank r*.

2.4 Face Morphing Generation

This section provides a detailed explanation of different methodologies for generating morphed images.

Revisiting the concept, face morphing occurs when the facial features of one or more individuals are blended into a single image. As a result, the resulting face image resembles the faces of the contributing individuals.

In the literature, two well-established ways of generating morphed face images are known:

- Landmark-based
- Deep-learning based

Landmark-based

The most common way to generate a face morphed image is through the use of landmark-based morphing methods. In this case, the face morphing procedure is performed by warping the two contributing images using a collection of related points (facial landmarks) defined on each image. The usual pipeline is shown in figure 2.6.

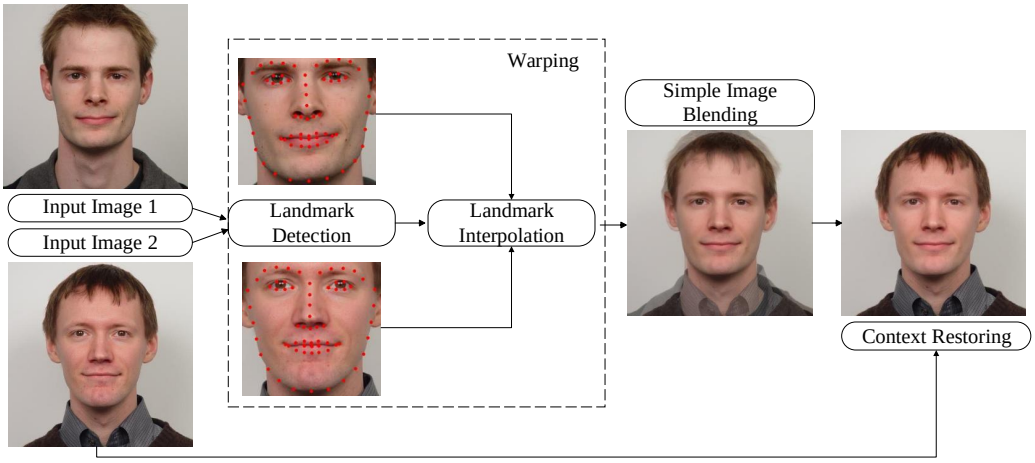

Figure 2.6: Landmark face morphing generation pipeline. Based on [4].

In the first step, a detector is used to locate and extract the facial landmarks of both faces. In most cases, a common approach is used, where 68 facial landmarks are selected [23]. These landmarks correspond to important facial points that define the face, such as the eyes, nose, mouth, and others (as shown in figure 2.7).

Next, the average of these two sets of landmarks is interpolated, and the result is used as a reference for aligning the facial features of the two images being morphed.

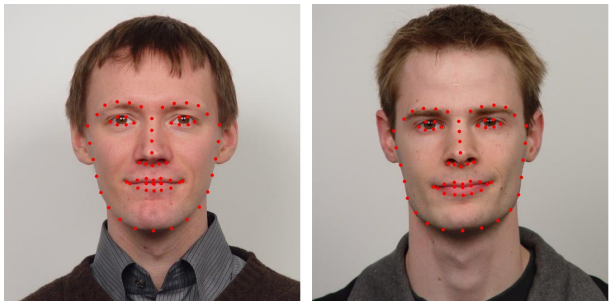

Figure 2.7: Representation of the 68 facial landmarks in each subject's face.

The images are warped in a process that relies on geometric transformations to modify those facial landmarks in the images in order to achieve correspondence between them.

The next step is blending, where the resulting image is obtained by calculating a weighted average of the warped images. Each image has an associated α factor (which goes from 0 to 1) that indicates the individual contribution of each subject to the resulting morph. In this sense, an α value of 0.5 makes the contributions of both faces comparable.

Finally, it is important to restore the context information/background, which is done using one of the original images.

Deep Learning-based

Currently, deep learning-based techniques have also been used in morph generation, namely Generative Adversarial Networks (GANs) [24]. The main underlying motivation is due to the increased resolution and quality of the resulting images.

From a broad perspective, the GAN approach consists of an "opposition game" between a generator and a discriminator in order to achieve dynamic balance. Considering an image-based domain, the discriminator learns to separate the input images into two outputs (real or fake), and at the same time, the generator is trained to trick the discriminator by creating fake inputs. Once trained, the discriminator is discarded, and the generator is used to generate realistic fake images.

Some of the approaches, such as StyleGAN [25] in order to improve the quality of generated images, also introduce a perceptual model. The goal is to evaluate the similarity between generated and authentic images based on their perceptual characteristics (color, texture, and structure), *i.e.,* minimize perceptual loss between reference and generated images in feature space. In short, this means that the network is trained to generate images that are not only visually similar to real images but also have similar perceptual characteristics.

In the particular face morphing context, the common pipeline is presented in figure 2.8.

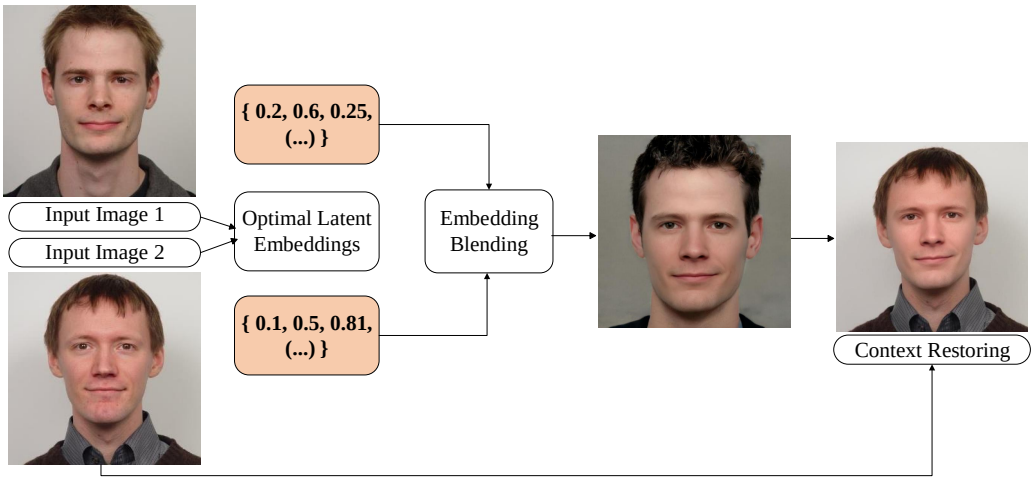

Figure 2.8: GAN face morphing generation pipeline.

Following the order, after aligning and resizing the input image accordingly to the discriminator network's requirements, the optimal latent face embeddings are extracted for both images. The term "latent face embedding" refers to the representation of the human face that encodes an individual's facial features into a

compact representation.

These embeddings are then used to generate the latent embedding of the morphed face through an interpolation process. Finally, the morph image is generated based on this embedding, and the context is restored given one of the original images.

2.5 Morphing Attack Detection

So far, techniques to generate morphed images have been covered, but the big challenge is being able to detect them. Throughout this section, some detection methodologies will be discussed.

Based on the processing scenario, MAD can be divided into two main approaches:

- Single Morphing Attack Detection (S-MAD)
- Differential Morphing Attack Detection (D-MAD)

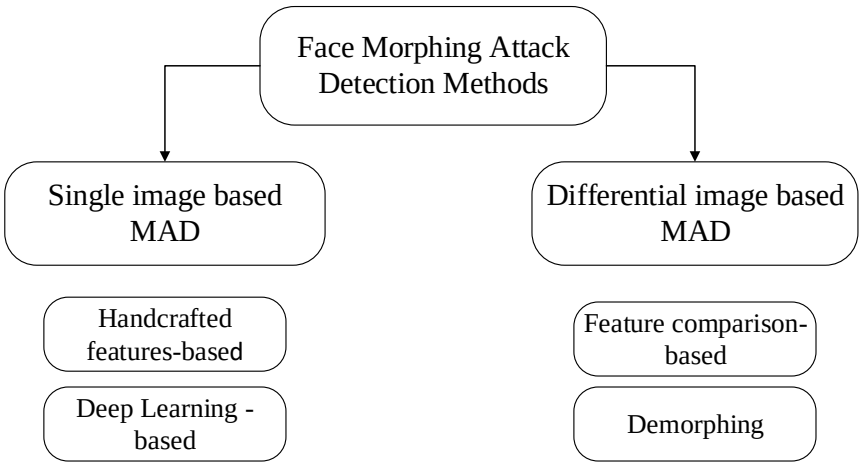

Figure 2.9: Summary diagram of MAD methods for the different processing scenarios. Based on [5].

Single Morphing Attack Detection

The S-MAD case relies on situations where the algorithm uses a single image for the purpose of detecting face morphing attacks. Specifically, determining whether an image has been modified (face morphing) without the need to use a reference (authentic) image as a comparison. Due to this, it is also referred to as the *no-reference* method.

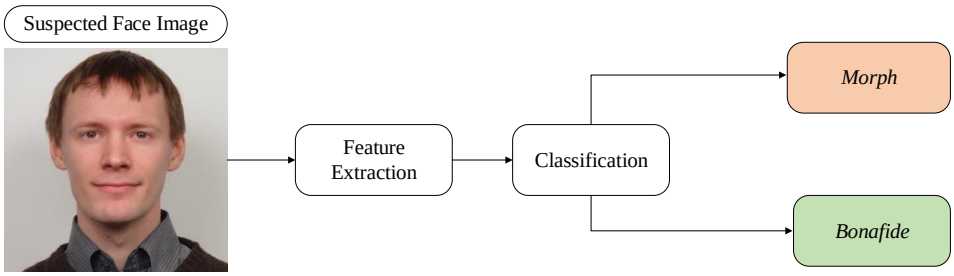

Figure 2.10: *No-Reference* morphing detection scheme. Image based on [6].

In real-life scenarios, this type of detection is visible, for instance, at the time of the initial passport application, when the applicant presents the photo in digital or physical format but there is no previous reference photo to use as a comparison.

Regarding the approaches, they can be divided into two major groups [5]: 1) Those that use handcrafted features, which are a set of manually designed and engineered features that represent and describe the characteristics of an image, such as texture-based, quality-based, and residual noise-based; 2) Those that are based on deep learning approaches. Examples of each case will be discussed separately in chapter 3.

Differential Morphing Attack Detection

Conceptually, D-MAD algorithms have an advantage over S-MAD due to their access to additional information. The method involves comparing the input image to a reference image (typically captured in a trusted environment) and looking for inconsistencies that may indicate an attempted attack. Due to this, it is also referred to as the *reference-based* method.

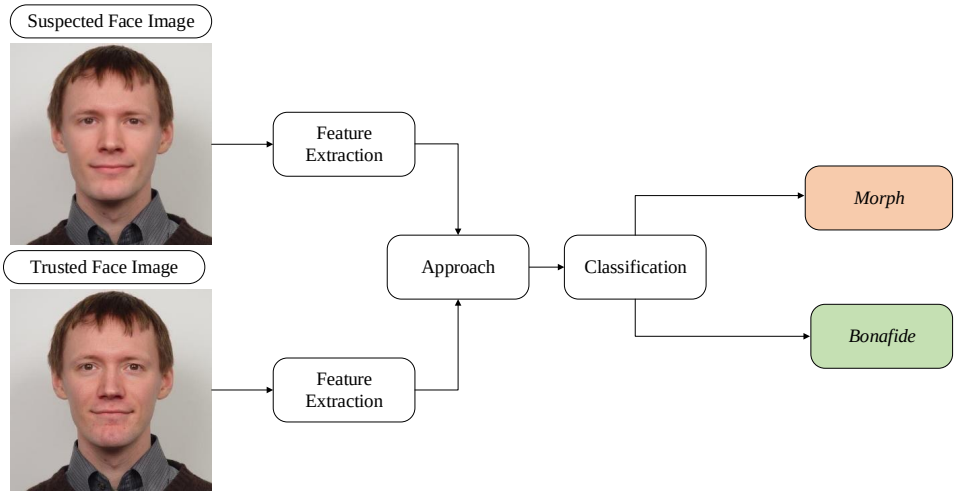

Figure 2.11: *Reference-based* morphing detection scheme. Image based on [6].

In real-life scenarios, this type of detection is commonly seen in border crossing control situations where the suspicious morph image (taken from the passport) can be compared with a trusted image, namely captured live at the ABC gates [5].

Regarding the approaches, they can be divided into two major groups [5]: 1) Those that use feature comparison-based approaches; 2) Those that use demorphing approaches. Examples of each case will be discussed separately in chapter 3.

Performance Evaluation

According to *ISO/IEC 30107-1:2016* [26] the performance evaluation can be measured with the metrics below.

- **Attack Presentation Classification Error Rate (APCER)**

Represents the proportion of fake samples mistakenly identified as genuine (system insecurity). In the particular case of face morphing, it represents the ratio of incorrectly classified morphed images (M) to the total number of morphed images (N_m) - *morph miss rate*.

$$APCER = \frac{M}{N_m} \tag{2.1}$$

- **BonaFide Presentation Classification Error Rate (BPCER)**

Represents the proportion of genuine samples that were misidentified as fake (user inconvenience). In the particular case of face morphing, represents the ratio of incorrectly classified *bonafide* images (B) to the total number of *bonafide* images (N_b) - *false detection rate*.

$$BPCER = \frac{B}{N_b} \tag{2.2}$$

These two metrics are typically presented as the value of the BPCER at a fixed APCER threshold (BPCER@APCER). The goal is to strike a balance between minimizing the BPCER while ensuring that the system maintains an acceptable APCER. In an ideal scenario, the *false detection rate* should be as low as possible (minimize BPCER values), since in most real attempts, *bonafide* photographs are often presented.

3

State of the art

The current state of the art in Face Recognition (FR) is briefly explained in section 3.1. Sections 3.2 and 3.3 describe current developments in face morphing as well as techniques for detecting corresponding attacks. Section 3.4 presents the available datasets used in training and benchmarking procedures. Finally, in section 3.5, a summary of the key state-of-the-art concepts along with a short discussion is provided.

3.1 Face Recognition

The dynamic nature of the human face poses significant challenges in the FR field. To address these challenges, several solutions have been suggested to enhance the robustness and accuracy of Face Recognition System (FRS).

Looking at early works, in the 1960s, Bledsoe [27] proposed the first semi-automatic FR approach, which focused on manually extracting elements from the face and using them to establish proper identification.

From a general standpoint, as described in section 2.1, the main purpose of a FRS is to accurately identify or verify a person's identity. To achieve this, a sequence of steps must be taken.

3.1.1 Face Detection

Face detection is usually the first stage in the FR task. As the name suggests, it relies on the detection of the face in an image or video.

One of the most seminal works was proposed by Viola and Jones [28]. Based on the research of Papageorgiou *et al.* [29], the researchers proposed an efficient real-time face detection system utilizing a set of features based on Haar-Basis functions. Furthermore, they made a simple modification to improve the system's performance by proposing the use of AdaBoost's learning cascade structure [30].

The simplicity of Haar's features proved to be a significant obstacle in achieving

meaningful performance, leading to the development of several techniques to improve its performance [31].

In recent times, most face detection approaches have been based on the use of deep learning techniques. These approaches outperform the traditional ones, especially in unconstrained scenarios with a wide range of face poses, viewing angles, occlusions, and lighting conditions.

Zhang *et al.* [32] proposed the Multi-Task Cascaded Convolutional Neural Network (MTCNN) approach for the detection and simultaneous alignment of the face. The use of a cascading algorithm makes it possible to quickly detect the face with more accuracy and simultaneously handle those unconstrained scenarios.

Once the face is detected, the subsequent stages involve obtaining a face descriptor/representation, highlighting the use of traditional and deep learning methods.

3.1.2 Traditional Approaches

From a broader perspective, traditional face recognition algorithms can be categorized into two main types: holistic and feature-based methods.

In holistic approaches, the whole face is mapped into a lower-dimensional subspace using linear or non-linear methods, such as Independent Component Analysis (ICA), and Linear Discriminate Analysis (LDA), while retaining the most relevant information. In other words, this approach involves treating the entire face as a single unit for recognition purposes.

Sirovich and Kirby [33] proposed one of the most popular approaches with holistic-based methods called Eigenface. Later, Turk and Pentland [34] improved the method specifically for the FR task by automating the process to perform eigendecomposition on multiple images.

In a nutshell, using Principal Component Analysis (PCA) the face images are projected into a lower-dimensional space called the eigenface subspace, which captures the most significant fluctuations in the data. To identify faces, the technique compares the weights of a new face, obtained through projection, with the weights of the training set, and the closest match is considered the recognized face.

On the other hand, feature-based methods involve identifying and extracting specific facial features from an image. In this method, descriptors play a crucial role in representing the characteristics of the extracted features, providing a compact representation of the appearance, texture, or geometric properties.

Ouarda *et al.* [35] explored the use of geometric features extracted from face images, including the positions and distances between significant facial landmarks,

to create a representation for each face.

Ahonen *et al.* [36] proposed a technique that uses shape and texture information to describe face images. Initially, the face image was segmented into smaller sections, from which Local Binary Patterns (LBP) [37] histograms were extracted. Next, the k-Nearest Neighbor (k-NN) classifier was used to quantify the disparity between the two facial descriptors.

The simplicity of computing LBP features and their independence from prior knowledge of face geometry make them a practical and convenient approach for real-world FR applications, as underlined by the authors.

Dalal and Triggs [38] suggested the use of locally normalized Histogram Of Gradient Orientations (HOG). The HOG method enables the extraction of discriminative feature vectors from images by counting the occurrences of specific gradient orientations. by proposing the use of AdaBoost's learning cascade structure [30][39, 40].

In recent years, the Gabor transformation has gained significant prominence as an effective element in image processing and pattern recognition tasks. Zhang *et al.* [41] proposed a feature descriptor called Histogram of Gabor Phase Patterns (HGPP), which combines the spatial histogram and Gabor Phase Pattern (GPP) to capture both local texture and shape information of the face. In the approach, Gabor filters were utilized to extract the phase information from the facial images, and the resulting descriptor was obtained by encoding this information based on the use of histograms.

Despite some of these works performing well in constrained scenarios, the same is not so evident in more realistic scenarios. As a result, at that time, FRSs had an unstable performance in such scenarios.

3.1.3 Deep-learning Approaches

Recent advancements in the FR domain have been driven by the integration of deep learning techniques. The motivation arises from the limitations that traditional FRS face in more realistic scenarios. As a solution, the use of features that can handle complex intra-personal variations has become more prevalent.

In 2012, AlexNet [42] won the ImageNet ILSVRC competition, highlighting the potential of deep learning approaches in solving complex image recognition problems. One of the key innovations of AlexNet was the use of Rectified Linear Units (ReLU) as activation functions, allowing faster training and better generalization when compared to traditional activation functions such as *sigmoid*.

Since then, there has been a remarkable growth in Convolutional Neural Networks (CNNs) based approaches, with several deeper and sophisticated architectures emerging, including: VGGNet [43], GoogLeNet [44], ResNet [45], MobileNet with three versions [46–48], DeepFace [49], among others.

Furthermore, in an attempt to achieve high performance while minimizing the number of parameters, innovative architectures like DenseNet [50] and EfficientNet [51] have also been developed.

Over time, the primary research focus of FRSs has undergone a shift. Presently, the emphasis is no longer on exploring novel architectures but rather on utilizing established architectures that produce optimal outcomes, simplifying the training and benchmarking procedures.

To achieve optimal performances, classification methods that use *softmax* are typically employed. The *softmax* function transforms the output of a neural network into a probability distribution over the predicted classes. The truth class labels are then compared to the predicted class probabilities obtained from the model using a cost function. The goal is to minimize the error, which in turn maximizes the probability of correctly identifying the truth class.

Sun *et al.* [52] implemented one of the classical FR approaches using *softmax*. The authors proposed the adoption of deep learning techniques to obtain a set of high-level feature representations called Deep Hidden IDentity features (DeepID). The model was trained on a large dataset, removing the *softmax* function, and utilized the output of the penultimate layer as the face representation.

In recent investigations, there has been an increasing interest in modifying or improving the loss functions used in deep learning pipelines in order to improve the discriminatory power of the extracted features. The loss function plays a crucial role in quantifying the dissimilarity between the predicted output of a model and the true output, thus allowing adjustment of the model parameters or weights during the training process.

Classification-based Approaches

The vast majority of classification-based approaches focus on the use of *softmax* loss function, whose main goal is to distinguish features, maximizing the posterior probability of the correct class.

Overall, it can be defined as the combination of the *softmax* activation function and the *cross-entropy* loss, resulting in the following equation:

$$L_{softmax} = \frac{1}{N}\sum_{i=1}^{N} -\log(\frac{e^{fy_i}}{\sum_{j=1}^{C} e^{fy_j}}) \tag{3.1}$$

where C is the number of classes of the classification problem, N is the number of training samples, and y_i is the corresponding label from the feature vector x_i. Considering a weight vector W_j and a bias $B_j = 0$, the activation of the last Fully Connected (FC) layer f_j can be defined as follows:

$$f_j = W_j^T x = \|W_j^T\| \, \|x\| \cos(\theta_j) \tag{3.2}$$

Despite its effectiveness in various classification tasks, the original *softmax* function has been found to be insufficiently discriminative for the practical FR task, resulting in lower performances when dealing with significant intra-class variations.

Initially, the modifications focused on redefining the *softmax* loss by incorporating margin-based alternatives. As a common starting point and based on the equation 3.2, both feature and weight vectors are normalized using $L2$ normalization, which ensures that the learned embedding features are distributed on a s-hypersphere, where s is a scale parameter.

In this case, the dot product is equivalent to the distance cosine, which strictly depends on the cosine of the angle between the two vectors, so that:

$$f_j = W_j^T x = s\cos(\theta_j) \tag{3.3}$$

Based on this similar formulation, the loss function was modified in different ways. In 2018, Liu *et al.* [53] introduced a deep hypersphere embedding for faces called SphereFace. Based on the idea that *softmax* loss has an intrinsic angular distribution, a parameter m was established on a hypersphere manifold to control the angular margin and improve the decision margin between the classes. This modification of the original *softmax* loss function is known as Angular *softmax* or A-*softmax* and ensures that the margin is appropriately adjusted, improving classification accuracy.

Considering equation 3.2 and the specific scenario of a binary problem, the decision boundary presented in figure 3.1 (for the particular case of class 1) can be described as follows:

$$s(\cos(m\theta_1) - \cos(\theta_2)) = 0 \tag{3.4}$$

The m factor makes the decision boundary more strict compared to the *softmax* original approach. It should be noted that scale s is not a constant in this case and

represents the $L2$-norm of the feature vector. The reformulated equation 3.1 is then defined as:

$$L_{SphereFace} = \frac{1}{N}\sum_{i=1}^{N} -\log(\frac{e^{s\cos(m\theta_{y_i})}}{e^{s\cos(m\theta_{y_i})} + \sum_{j\neq i} e^{s\cos(\theta_{y_j})}}) \tag{3.5}$$

In the same year, to overcome the limitations of existing methods in effectively distinguishing similar faces, Wang *et al.* [54] introduced the CosFace approach. To this end, the authors proposed the use of a loss function known as Large Cosine Margin Loss (LMCL), which incorporates the cosine margin penalty directly into the target logit, resulting in highly discriminative facial features. The objective of this approach is to maximize inter-class variation by reducing the angle between them while minimizing intra-class variance.

Considering equation 3.2 and the specific scenario of a binary problem, the decision boundary presented in figure 3.1 (for the particular case of class 1) can be described as follows:

$$s(\cos(\theta_1) - m - \cos(\theta_2)) = 0 \tag{3.6}$$

Unlike the SphereFace approach, in this case, the scale s value represents a constant. The reformulated 3.1 equation can be defined as follows:

$$L_{CosFace} = \frac{1}{N}\sum_{i=1}^{N} -\log(\frac{e^{s(\cos(\theta_{y_i})-m)}}{e^{\,s(\cos(\theta_{y_i})-m)} + \sum_{j\neq i} e^{s\cos(\theta_{y_j})}}) \tag{3.7}$$

One year later, Deng *et al.* [55] introduced the ArcFace approach, which utilizes an additive angular margin loss. Similarly to the previously mentioned approaches, ArcFace also takes advantage of the inherent angular distribution of the *softmax* function. However, it uses a different angular margin technique, where the penalty m is directly added to the angle.

Considering equation 3.2 and the specific scenario of a binary problem, the decision boundary presented in figure 3.1 (for the particular case of class 1) can be described as follows:

$$s(\cos(\theta_1 + m) - \cos(\theta_2)) = 0 \tag{3.8}$$

The scale s value, like in the CosFace approach, represents a constant. The reformulated 3.1 equation can be defined as follows:

$$L_{ArcFace} = \frac{1}{N}\sum_{i=1}^{N} -\log(\frac{e^{s\cos(\theta_{y_i}+m)}}{e^{\,s(\cos(\theta_{y_i}+m))} + \sum_{j\neq y_i} e^{s\cos\theta_j}}) \tag{3.9}$$

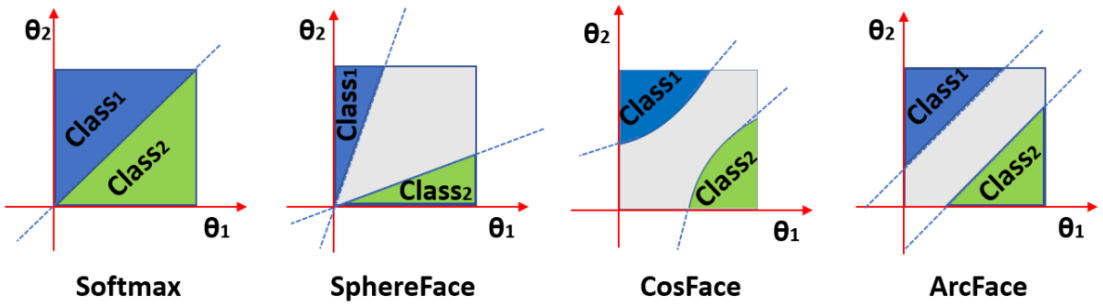

Figure 3.1: Decision boundaries across the different loss function approaches for the particular binary case. The grey areas represents the decision margins. Image from [7].

Despite the evolution and intensive use of margin-based loss functions, there has been a growing interest in adaptive loss functions. These techniques aim to incorporate adaptiveness into the margin to address the limitations of existing margin-based loss functions. That is, instead of using a fixed margin, adaptive loss functions allow for dynamic adjustment against the problem context.

Meng *et al.* [56] proposed the MagFace approach. This method optimizes the feature embedding using an adaptive margin and regularization based on the magnitude of the embedding. In general, the metric follows the ArcFace approach based on the cosine margin while reinforcing its direction and magnitude.

Another promising technique is AdaFace, proposed by Kim *et al.* [57]. The idea is to adapt the margin-based functions by using the feature embedding norm (an indicator of image quality) in order to control the gradient scale assigned to different image qualities.

Metric-based Approaches

Metric learning methods aim to optimize feature embeddings, *i.e.,* increase their discriminative power. To this end, the methods rely on learning a representation space that efficiently approximates similar samples while maintaining a clear separation between different samples. One of the drawbacks associated with these approaches is the need for extensive datasets and sophisticated sample mining strategies to provide consistent convergence.

One of the main examples was introduced by Schroff *et al.* [58] in FaceNet. The authors proposed a framework in which face images were mapped to a compact Euclidean space, where the $L2$ distances between those embeddings represent the face similarity measure.

In the approach, three sets of images are used during training: an anchor im-

age, a positive image, and a negative image. The objective is to reduce the distance between the anchor and the positive image while simultaneously increasing the distance between the anchor and the negative image (triplet loss function).

Shi and Jain [59] introduced a novel face recognition approach called DocFace, inspired by the FaceNet framework. The DocFace approach uses the Max-margin Pairwise Score (MPS) loss function, which aims to optimize the model by comparing pairs of instances.

All the papers cited in this section are summarized in table 3.1.

Table 3.1: Summary of the papers cited in section 3.1 regarding face recognition.

Approach	**Reference**	**Methodology**
Face Detection	Viola and Jones [28]	Real-time system based on Haar-Basis functions
	Tieu and Viola [30]	AdaBoost learning cascade structure
	Zhang *et al.* [32]	Deep cascaded multi-task framework (MTCNN)
Tradicional Approaches (Face Recognition)	Sirovich and Kirby [33]	EigenFace method
	Turk and Pentland [34]	EigenFace improvement
	Ahonen *et al.* [36]	LBP feature descriptor
	Zhang *et al.* [41]	HGPP feature descriptor
	Dalal and Triggs [38]	Normalized HOG feature descriptor
	Ouarda *et al.* [35]	Geometric facial features
Deep learning Approaches (Face Recognition)	Krizhevsky *et al.* [42] [60]	AlexNet architecture
	Simonyan and Zisserman [43]	VGGNet Architecture
	Szege *et al.* [44]	GoogLeNet architecture
	He *et al.* [45]	ResNet architectures
	Howard *et al.* [46–48]	MobileNet architecture
	Taigman *et al.* [49]	DeepFace architecture
	Huang *et al.* [50]	DenseNet architecture
	Tan and Le *et al.* [51]	EfficientNet architecture

Classification-based approaches	Sun *et al.* [52]	DeepID - *softmax* classification
	Liu *et al.* [53]	SphereFace: Margin-based loss function
	Wang *et al.* [54]	CosFace: Margin-based loss function
	Weng *et al.* [55]	ArcFace: Margin-based loss function
	Meng *et al.* [56]	MagFace: Adaptive loss function
	Kim *et al.* [57]	AdaFace: Adaptive loss function
Metric-learning approaches	Schroff *et al.* [58]	FaceNet: Triplet loss
	Shi and Jain [59]	DocFace: Max-margin Pairwise Score loss

3.2 Face Morphing Generation

Ferrara *et al.* [61] introduced the concept of generating a morphed face image by combining images of two (or more) subjects. The paper addresses the main problem of using machine-assisted identity confirmation in Electronic Machine Readable Travel Document (eMRTD) through face-to-face morphing attacks.

Currently, obtaining a morphed image is easily accessible to anyone through the use of tools such as MorphThing [62], FaceMorpher [63] and MagicMorph [64]. These user-friendly interfaces are developed using various methodologies that are continuously being researched and studied.

In the scope of research, landmark-based techniques are the most widely used for generating morphs [65, 66]. The general procedure is based on getting the facial landmarks of both faces, warping them in such a way that the corresponding landmarks are aligned, and finally performing blending.

During the warping stage, a triangulation procedure [67] is used to align the corresponding triangles of both images through a geometric transformation. One of the commonly used triangulation approaches is the Delaunay triangulation, which was introduced by Delaunay [68]. The technique is used to triangulate points on a surface, maximizing the minimum angle between adjacent triangles.

One of the main challenges faced by landmark morphing approaches is the

presence of visible artifacts in the resulting images. Seibold *et al.* [69] observed that morphed images often suffer from blurring and suboptimal image quality. To address this issue, the authors proposed the use of a conventional style transfer algorithm using a pre-trained CNN network to improve the quality of morphed images.

Motivated by the aforementioned challenges, the adoption of deep learning techniques has gained significance, leading to the emergence of approaches such as Generative Adversarial Networks (GANs). In these approaches, morphed images are generated by sampling two face images from the latent space rather than the pixel space, which means that in the end, the morphed images are generated at the representation level instead of the image level.

The concept of GAN was first introduced by Goodfellow *et al.* [70] in a method that involved training two multilayer perceptrons simultaneously in order to minimize the loss function of the generator and maximize the loss function of the discriminator.

Damer *et al.* [71] inspired by the work of Dumoulin *et al.* [72] proposed one of the most emblematic GAN approaches for generating morphed faces, called MorGAN. In this approach, the authors transformed the images into latent representations via a variational autoencoder, averaged these latent domain representations, and finally fed the generator with the resulting morph latent representation.

Later, building upon the MorGAN approach, Damer *et al.* [73] introduced a novel method to improve the realism of the generated images using cascading multiple GAN models. The motivation behind this approach arises from the observation that existing GAN models can generate realistic images but often lack fine detail and high-frequency components. By employing this approach, the authors demonstrate a significant improvement in the realism of the morphed images, making them more difficult to detect.

Venkatesh *et al.* [74] performed a comparison between two main approaches for face morphing: GAN and landmark-based methods. The authors observed that morphed faces generated by GAN did not exhibit blending artifacts, which was a limitation of landmark-based methods. However, the morphed faces generated by GAN were easier to detect due to the presence of characteristic noise. As a result, the authors concluded that, at that time, conventional methods had an advantage over GAN approaches.

In 2019, Karras *et al.* [25] introduced the StyleGAN architecture, which aimed to improve the quality and diversity of generated images by expanding the original GAN framework. The key idea of the StyleGAN approach is the utilization of a latent intermediate space, enabling controlled modifications in the image generation

process. In a more detailed manner, instead of generating images based on a fixed set of features, as in traditional GANs, StyleGAN generates images based on a set of learned styles, resulting in the generation of images that exhibit a wide range of styles and features.

Several new approaches have been inspired by the StyleGAN architecture. Abdal *et al.* [75] proposed the Image2StyleGAN approach, to map a given image into the latent space of StyleGAN. In the specific case of morphing purposes, Zhang *et al.* [76] introduced the MIPGAN approach. In this approach, morphed images were obtained using an architecture with a modified loss function. The main idea was the introduction of identity priors and perceptual quality information. These identity priors were learned through a pre-trained FR task.

Recently, Damer *et al.* [77] suggested the utilization of diffusion autoencoders to generate morphed images in a project called MorDIFF. The paper compares the proposed technique with a wide range of existing image-level and representation-level morphing methods. The motivation for the use of diffusion autoencoders arises from the limited reconstruction fidelity of GAN architectures during the interpolation process in the latent space.

All the papers cited in this section are summarized in table 3.2.

Table 3.2: Summary of the papers related to face morphing generation approaches cited in section 3.2

Approach	**Reference**	**Methodology**
Landmark-based Approach	Ferrara *et al.* [61]	Face morphing operation in a semiautomatic way
	Seibold *et al.* [69]	Improvement in landmark morph quality using using Style-transfer techniques
GAN-based Approach	Damer *et al.* [71]	MorGAN approach
	Damer *et al.* [73]	Enhancing the realism of GAN images using cascading models
	Venkatesh *et al.* [74]	Landmark vs GAN-based approaches
	Karras *et al.* [25]	StyleGAN approach
	Abdal *et al.* [75]	Image2StyleGAN approach
	Zhang *et al.* [76]	MIPGAN approach

	Damer *et al.* [77]	MorDIFF approach to overcome GAN limitations

3.3 Face Morphing Attack Detection

Several research projects relating to Morphing Attack Detection (MAD) have been supported by the European Union and National Research Council [5], such as Integrated Monitoring, Analysis and Response System (iMARS) [78], in which Imprensa Nacional-Casa da Moeda (INCM) is a partner, and Securing Online Transactions against MitM Fraud (SOTAMD) [79]. These programs emphasize the significance of researching and developing approaches for the detection of face morphing attacks, which can significantly impact the FRS's security.

Depending on the processing scenario, two methodologies can be distinguished: *non-reference* or *reference-based*, as presented in section 2.5.

3.3.1 Single Morphing Attack Detection

The Single Morphing Attack Detection (S-MAD) or *non-reference based* refers to techniques that can detect a morphed image without requiring a direct comparison with an authentic reference image. Instead, detection is based solely on the inherent characteristics of the transformed image itself, such as visual artifacts or inconsistencies.

Handcrafted Feature-based Approaches

One of the earlier approaches to dealing with S-MAD was proposed by Raghavendra *et al.* [80]. The goal was to utilize a descriptor based on Binarized Statistical Image Features (BSIF) and determine whether a face was morphed or *bonafide* using linear Support Vector Machine (SVM). Several other techniques based on texture features have been proposed using LBP [81] and Local Phase Quantization (LPQ) [82] image descriptors.

The quality of the image can also influence the detection, and in this sense, the quantification of image degradation can be performed. Scherhag *et al.* [83] explored the potential of utilizing Photo Response Non-Uniformity (PRNU) sensor noise for detecting morphed images. The method involves an analysis in both the spatial and frequency domains. The authors observed that in the spatial domain, the distribution of sensor noise in morphed images was compressed compared to

genuine images. Additionally, in the frequency domain, morphed images exhibited a reduced coverage of large magnitudes when compared to genuine images.

Zhang *et al.* [84] proposed a method motivated by image source identification using Fourier Spectrum Of Sensor Pattern Noise (FS-SPN). The idea behind this approach was based on the fact that *bonafide* and morphed face images were generated from different image acquisition pipelines. Therefore, by finding the difference in Sensor Pattern Noise (SPN), it was possible to detect the face morph image.

Venkatesh *et al.* [85] proposed the identification of morphed face images using residual color noise. In the method, the authors employed a deep CNN-based denoising network to obtain the residual noise and effectively quantify the noise patterns. Then the Pyramidal Local Binary Pattern (P-LBP) descriptor was used to extract distinctive features.

Deep Learning-based Approaches

In recent works, deep learning has been extensively utilized in the MAD field. Raghavendra *et al.* [86] proposed one of the pioneering approaches that used features extracted by pre-trained Deep Convolutional Neural Network (DCNN) in FR contexts, such as VGG-19 and AlexNet architectures [42]. Those features were subsequently employed for classification using the Probabilistic Collaborative Representation (P-CRC) method. The paper also reported successful results in detecting morphed face images, including digital and print-scanned samples.

Seibold *et al.* [87] proposed a deep learning morphing attack detection approach using morphed faces during the training in a different manner. The core of the approach was the adaptation of existing deep learning classification approaches to solve the specific problem of face morphing detection. The authors concluded from the work that the features learned for object classification were effective in detecting morphing attacks.

Neto *et al.* [88] proposed the OrthoMAD approach. The approach uses a new regularization term to incorporate the identity information existing in both contributing images and, in addition, proposes the creation of two orthogonal latent vectors.

Recently, Medvedev *et al.* [89] introduced the MorDeephy method, which employs a fused classification approach to generalize morphing detection to unseen attacks. The proposed methods reached the second position worldwide in the Face Recognition Vendor Test (FRVT) MORPH National Institute of Standards and Technology (NIST) benchmark test [90], and inspired by that, this dissertation will

follow the authors' methodology.

3.3.2 Differential Morphing Attack Detection

The Differential Morphing Attack Detection (D-MAD) or *reference-based* refers to techniques that involve a direct comparison between the suspicious face morphed image and a reference face image.

Feature Comparison-based Approaches

The basis of feature comparison-based approaches is to compare feature vectors generated from reliable live captures with vectors extracted from possible morphs.

Scherhag *et al.* [91] proposed a method that involves extracting facial landmarks from the input image and computing a similarity score between the landmarks of the input image and the landmarks of a reference/genuine image.

Damer *et al.* [92] proposed a method that by analyzing the directed distances of face landmark displacements attempts to detect face morphing attacks.

Overall, the extraction of the features can also be performed using deep learning approaches. Scherhag *et al.* [93] extracted feature embeddings from the ArcFace model trained for a FR task and then classified them as *bonafide* or morphed.

Soleymani *et al.* [94] explored the use of a deep Siamese network to obtain feature embeddings. The objective was to compute the distance between embeddings using contrastive loss, which tends to group similar images closer into a common latent subspace, while pushing dissimilar images further apart.

Demorphing Approaches

The face demorphing technique is based on reversing the morphing process.

Ferrara *et al.* [95] pioneered the introduction of the concept of face demorphing. In this approach, a potentially morphed image stored within the document is demorphed (reverted) as a way to determine the true identity of the document owner by comparison with the live (genuine) image. Subsequent studies evaluated the robustness of face demorphing by examining its effectiveness in the presence of various facial appearance variations [96].

More recently, Peng *et al.* [97] introduced a method called Face Demorphing GAN (FD-GAN) that utilizes GANs to demorph the accomplice's facial image. This approach employs an autoencoder architecture to effectively reverse the morphing process and generate a facial image that closely resembles the original appearance.

Ortega-Del-Campo *et al.* [98] also proposed a novel demorphing-based approach using a CNN to detect morphing presentation attacks in a real Automatic Border Control (ABC) system.

All the papers cited in this section are summarized in table 3.3.

Table 3.3: Summary of the papers cited in section 3.3 regarding MAD.

Approach	**Reference**	**Methodology**
Handcrafted feature-based (S-MAD)	Raghavendra *et al.* [80]	BSIF feature approach
	Scherhag *et al.* [83]	PRNU analysis
	Zhang *et al.* [84]	Image source identification using FS-SPN
	Venkatesh *et al.* [85]	P-LBP image descriptors
Deep learning-based (S-MAD)	Raghavendra *et al.* [86]	Deep features using a transfer CNN learning approach
	Seibold *et al.* [87]	DNN based morphing detecting
	Neto *et al.* [88]	OrthoMAD approach
	Medvedev *et al.* [89]	MorDeephy approach
Feature comparison-based (D-MAD)	Scherhag *et al.* [91]	Facial landmarks comparison between the genuine and morphed images
	Damer *et al.* [92]	Feature-based approach using distances between the landmarks.
	Scherhag *et al.* [93]	Feature-based approach using ArcFace model feature embeddings
	Soleymani *et al.* [94]	Feature-based approach using a Siamese network
Demorphing (D-MAD)	Ferrara *et al.* [95, 96]	Potential morph image reversion to identify the true identity of the owner of the document
	Peng *et al.* [97]	FD-GAN approach
	Ortega-Del-Campo *et al.* [98]	Demorphing using CNN

3.4 Available Datasets

Currently, a majority of state-of-the-art FR methods rely on data-driven approaches. In this sense, datasets are a critical factor in the development of new

methodologies and are also used as benchmarks for system validation. Many of them are accessible to the public and can be utilized for non-commercial research purposes. However, the vast majority are restricted to use by large companies that typically collect private datasets.

Table 3.4 schematically summarizes common datasets for FR training and benchmarking.

Table 3.4: Sample datasets for face recognition training and benchmarking.

Dataset Name	**Description**	**Is it public ?**
CASIA-WebFace [99]	**494.4K images of 10.5K identities** The dataset exhibits a wide range of variations in terms of age, gender, pose, expressions, and illumination. The images have high quality, good resolution, and minimal artifacts.	Yes
Labelled Faces in the Wild (LFW)[100]	**13.2K images of 5.7K identities** The dataset includes images with different lighting conditions, poses, expressions, and ages, as well as strong occlusions and low resolution.	Yes
YouTube faces (YTF) [101]	**3.4K images of 1.5K identities** The dataset includes images with a significant degree of variability in terms of pose, expression, and lighting. These images were extracted from videos in which individuals performed several activities.	Yes
CelebA [102]	**202.5K images of 10.1K identities** The dataset includes images taken from the web with a wide range of poses, expressions, and backgrounds.	Yes
VGGFace2 [103]	**3.3M images of 9.1K identities** The dataset includes images with very low label noise, high pose diversity, and age diversity. The images exhibit high quality, with good resolution and minimal artifacts.	Yes
Microsoft Celeb (MS-Celeb-1M) [104]	**10M images of 100K identities** The dataset contains a diverse set of different individuals, including people of different ages, genders, and ethnicities.	No

IMDB-Face [105]	**1.7M images of 59K identities** The dataset is noise controlled with variable resolutions and aspect ratios.	No
Face Recognition Grand Challenge (FRGC) [106]	**50K recordings** The dataset contains 2 uncontrolled still images and 4 controlled still images for each individual.	Yes
Extended M2VTS (XM2VTS) [107]	**13K images of 295 identities** The dataset includes images taken over a period of four months from recordings with speaking and a rotating head shot.	Yes
Notre Dame (ND) Twins [108]	**15K images of 1500 pairs of twins** The dataset contains variations in pose, expression, lighting, and image quality, along with the presence of glasses, hats, and other accessories.	No
Face Recognition Technology (FERET) [109, 110]	**11K images of 994 identities** The dataset includes different head poses with variations in lighting and facial expression.	Yes
AR Face [111]	**4k images of 126 identities (70 male and 56 female)** The dataset includes images with various facial expressions and facial occlusions caused by accessories such as sunglasses and scarves.	Yes
Psychological Image Collection of Stirling (PICS) [112]	Collection of several datasets. **Aberdeen**: 687 color faces of 90 individuals with some variations in lighting, 8 have varied viewpoint **Utrecht**: 131 images, 49 men, 20 women, usually a neutral and smiling face for each.	Yes
FEI Face [113]	**2.8K images of 200 identities** The dataset is gender balanced and presents a homogeneous white background where all subjects assume a frontal position with variations in profile rotation.	Yes

IMM Face [1]	**240 images of 40 identities (33 male and 7 female)** The dataset contains annotated monocular images of different frontal faces.	Yes
Georgia Tech Face (GTDB) [60]	**150 images of 50 identities** The dataset includes images with diverse, cluttered backgrounds, capturing subjects in various frontal poses with different facial expressions, lighting conditions, and scales.	Yes
Ethnic Facial Images of Ecuadorian People (EFIEP) [114]	**5.4K images of 180 identities** The dataset contains frontal images without facial occlusions. The faces are associated with various ethnic groups in Ecuador, and they have white backgrounds.	Yes
MIT-CBCL [115]	**10K images of 500 identities** The dataset includes both indoor and outdoor images with variations in lighting conditions, facial expressions, poses, and cluttered backgrounds.	Yes
Face Research Lab London (FRLL) Set [116]	**16K images of 102 identities** The dataset includes images of individuals of different ages, ethnicities and genders captured on white backgrounds with clearly visible frontal faces.	Yes
Young Labeled Faces in the Wild (YLFW) [117]	**10K images of 3K identities** The dataset contains images of children between the ages of 4 and 13, presenting a diverse representation of races and ethnicities.	Yes

As mentioned in the previous sections, the face morphing approach leads to the appearance of misleading images. For that reason, the creation of morphing-specific datasets often raises ethical concerns, making most existing datasets private and not accessible for non-commercial use.

Table 3.5: Sample morph datasets for face recognition training and benchmarking.

Dataset Name	**Description**	**Is it public?**

FRGC-MORPHS [118]	Dataset of morphed faces selected from the publicly available FRGC [106] dataset. The morphs were generated with the OpenCV [65], FaceMorpher [66], and StyleGAN2 [119] tools.	No
FRLL-MORPHS [120]	Dataset of morphed faces based on images selected from the FRLL dataset [116]. The morphs were generated using the following morphing tools: OpenCV [65], FaceMorpher [66], StyleGAN2 [119] and WebMorpher [121].	Yes
FERET-MORPHS [122]	Dataset of morphed faces selected from the FERET dataset [109, 110]. The morphs were generated with the OpenCV [65], FaceMorpher [66], and StyleGAN2 [119] tools.	No
Dunstone [123]	Dataset of standard morphed facial images based on the well-known FERET dataset [109, 110].	No

3.5 Discussion on the State of the Art

Over time, FR approaches have witnessed substantial advancements in terms of performance and accuracy. From a general point of view, a FRS can be represented in a generic pipeline that includes several essential steps, including face detection and alignment, as well as face representation and classification tasks.

Regarding face representation, initial approaches relied on the geometry, semantics, and texture of the faces (handcrafted features). However, the inherent variability of faces, including factors such as facial expressions, aging, lighting conditions, and head rotation, has resulted in an increased focus on developing technologies that can effectively handle unconstrained scenarios, leading to the introduction of deep learning approaches.

Currently, motivated by the need to address real-life situations, the vast majority of FRS have adopted the integration of CNN as part of their pipeline. The main purpose is to obtain a face representation that can handle complex variations within individuals, *i.e.,* learn highly discriminative face embeddings. For some time, the emphasis was on the search for optimal architectures, many of which are used

in most current approaches.

Currently, modifying and improving the loss function formulation has also been the focus in the development of FR techniques as a way to increase the discriminative power of features extracted from the deep networks.

In the case of classification approaches, methods based on *softmax-based* loss functions achieve the best results in terms of performance. However, margin-based loss functions have been introduced to improve accuracy by increasing inter-class variance and intra-class compactness. Additionally, adaptive loss functions have been developed to incorporate adaptiveness into the margin based on the quality of the input image.

From another perspective, the overall imperfection of FRS and their probabilistic nature make them targets for various types of attacks, in particular face morphing.

The morphing generation was introduced with the concept of landmark-based approaches, which focus on manipulating the key facial points of each contributing face. However, due to the presence of some imperfections, often associated with blending artifacts, over time they have been replaced by approaches based on the use of deep learning, in particular GAN. Despite the greater realism of the images, the process of image generation inadvertently introduces specific characteristics into the images, making them more susceptible to detection.

The problems associated with the typology of face morphing attacks reinforced the need to adopt detection techniques in order to prevent criminal attacks and fraud attempts on FRSs. As a result, there has been significant attention and interest in the field of MAD.

Depending on the security application scenario, MAD methods can be classified into two types: S-MAD and D-MAD, which generally differ by the presence of a reference image.

In S-MAD, the fact that there is no reference image for comparison makes the detection challenge greater. Algorithms typically rely on analyzing features of the image itself, such as visual artifacts or inconsistencies, to determine whether or not the image is morphed.

In the case of D-MAD, the approaches are based on comparing feature vectors extracted from the suspect image and a reference image or using reverse morphing (demorphing) techniques for detection purposes.

Despite the emergence of various approaches, MAD approaches still face several challenges and unresolved questions.

In general, during the training procedure, the majority of approaches utilize small datasets of faces that do not incorporate a significant number of morphed faces, which limits the performance of the models.

On the other hand, the testing phase is most often implemented in *closed-set* scenarios, where the models tend to perform well because the test and the train images share the same type of morphed data. However, such a scenario is unrealistic, and it is essential to extend MAD approaches to *open-set* scenarios where unseen data is used.

In the process of issuing identity credentials, such as passports or Identity Document (ID) cards, printed or scanned (re-digitized) face images are often used. These printed images can also be manipulated, *i.e.,* can be morphed images, and in that sense, investigating the performance of MAD algorithms in such cases is crucial.

All this highlights the need for further research and development to address these limitations and improve the effectiveness of detection methods.

4

Methodology

In this chapter, a detailed explanation of the procedures performed will be presented.

4.1 Source Data Curating

The academic community is currently facing a challenge due to the limited availability of large datasets that conform to the International Civil Aviation Organization (ICAO) guidelines. In response to this problem, a strategy was developed based on the aggregation of several datasets, both public and private.

During the dataset selection process, priority was given to datasets with a larger number of images per identity. As a result, the following datasets were chosen: FRGC [106], XM2VTS [107], ND Twins [108], FERET [109, 110], AR Face [111], PICS [112], FEI [113], IMMFD [1], and GTDB [60]. A summary of these datasets can be found in table 3.4.

Before combining all the datasets, a pre-processing step was required, taking into account the *suitability criteria* for application in face morphing. In other words, images that were considered unsuitable for performing the face morphing task, such as non-frontal images or images with evident face occlusions, were removed. In the specific case of the ND Twins dataset, only one of the twins was considered due to their striking resemblance, which remains an unsolved challenge in terms of Face Recognition (FR).

As a result of combining and pre-processing all these datasets, the ICMD dataset emerged, which contains over 50k images of more than 2500 individuals. Figure 4.1 shows representative images of each of the datasets.

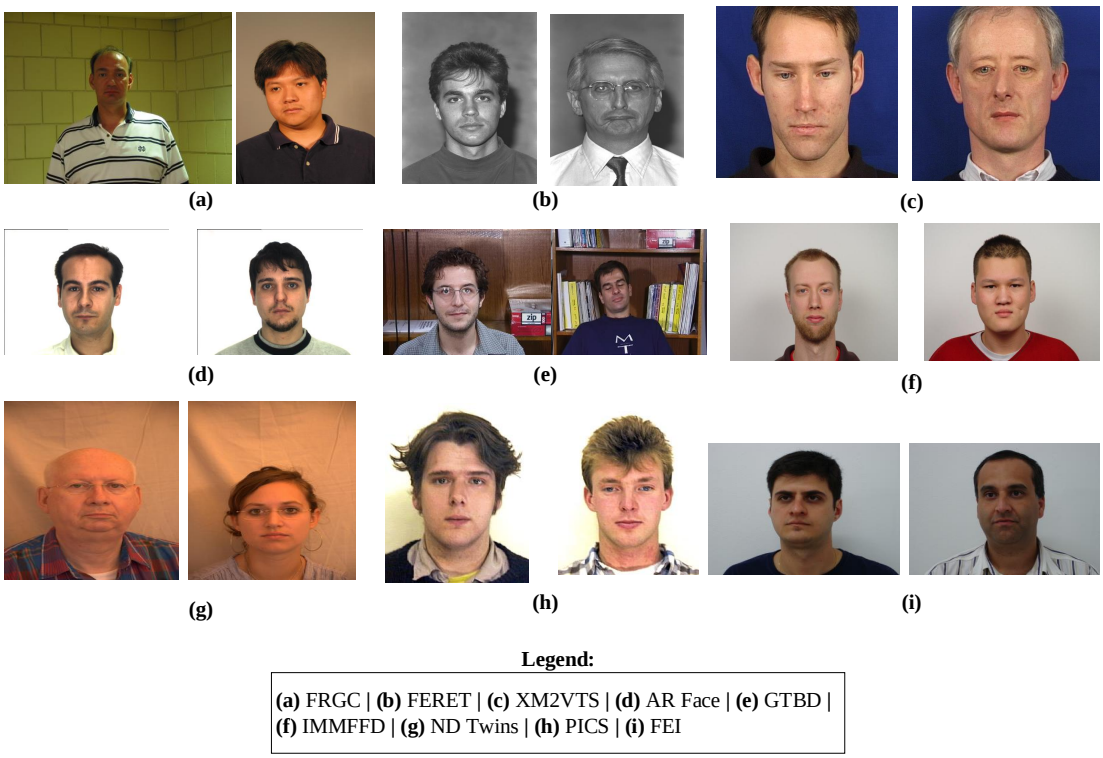

Figure 4.1: Sample images from each of the datasets used to generate the ICMD dataset.

4.2 Morphed Image Generation

In chapter 2, two methods for generating morphs were introduced: landmark-based and deep learning-based approaches. The idea of this dissertation is to use both. In the case of the deep learning approach, StyleGAN [25] is the chosen approach.

4.2.1 Landmark-based Approach

As presented in figure 2.7, the landmark-based approach involves using facial landmarks to establish a correspondence between specific points on two or more faces. These landmarks are then used to perform a triangulation process, creating a mesh that connects these points.

One problem associated with the triangulation step is that it is performed not only with the reference points, but also using the borders of the image. On images with a large margin of edges indicating that the face is not centrally located in the image, this inclusion of the edges can cause problems with context distortion, as shown in figure 4.2.

Consequently, a pre-processing step was performed on the images before generating the morphs.

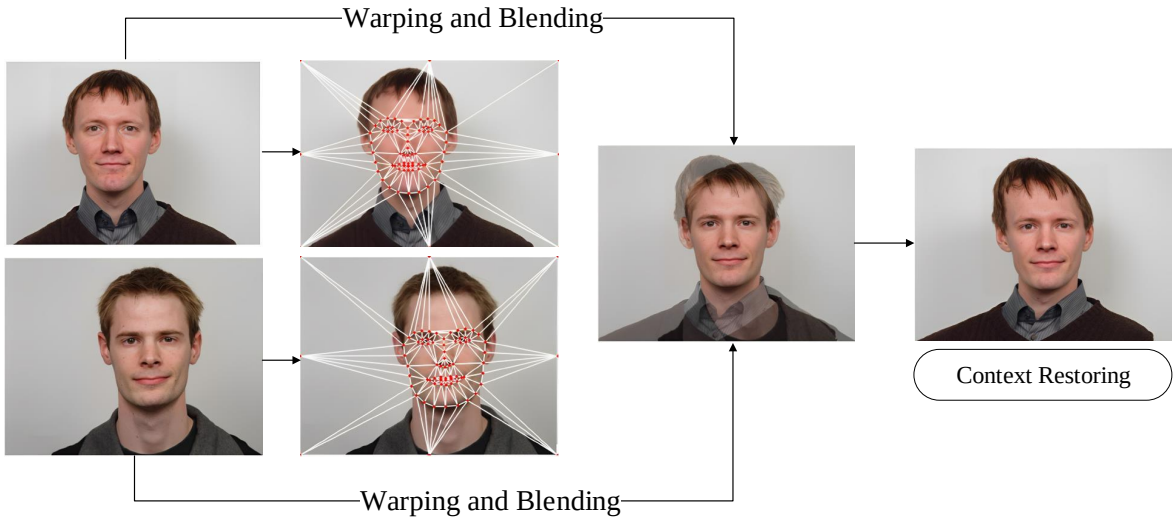

Figure 4.2: Distortion problem in landmark-based approach.

This pre-processing step consisted of cropping the original images, taking into account a well-defined proportion given by the "face contour/image" area, and considering the tip of the nose as the center of the image. The value chosen for the proportion was 1/8, which is used to correct the scale of the detected face region in the image to a more meaningful value. To obtain the new image dimensions, the square root of the resulting scaled area was applied. The face contour area is associated with the detection of the 68 landmarks [23].

Finally, to handle cases where the extracted region (crop) exceeded the image boundary, the image was padded using the reflection of the boundary pixels to ensure that the output size matched the desired dimensions.

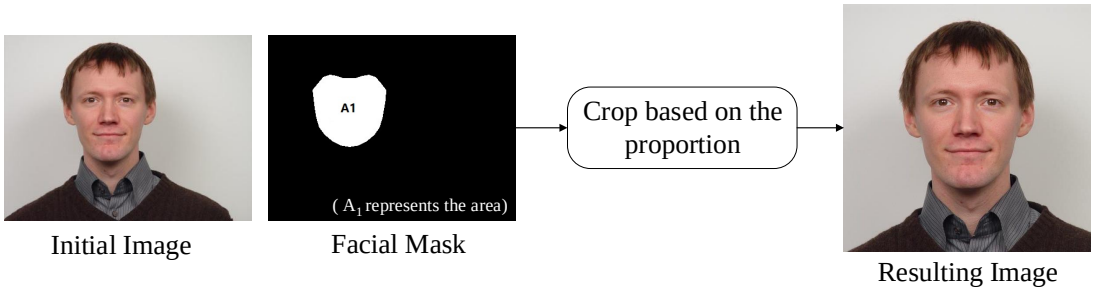

Figure 4.3: Pre-processing pipeline to deal with the distortion problem.

In the morphing generation process, it was necessary to take into consideration

the correct labeling of the classes (identities). The motivation comes from the fact that the resulting morphed image belongs to both source identities, *i.e.,* it simultaneously has two class labels associated with it. This would cause ambiguity about the correct labeling of the classes in the models used in this dissertation, which in turn will be described in the sections 4.4 and 4.5.

In that sense, and following the idea proposed by Medvedev *et al.* [89], the pre-processed ICMD dataset was split into two halves, and pairs of images were generated from each half. Then each generated image was labeled based on the corresponding sub-list for further classification.

Morphed images were then generated for each pair, resulting in a final dataset with about 49K images.

4.2.2 StyleGAN Approach

In the case of the StyleGAN approach, a similar split and pairing approach was followed, maintaining the same identity separation performed in the landmark-based approach but regenerating the paring to diversify the morphs. Morphed images were then generated for each pair, resulting in a final dataset with about 49K images.

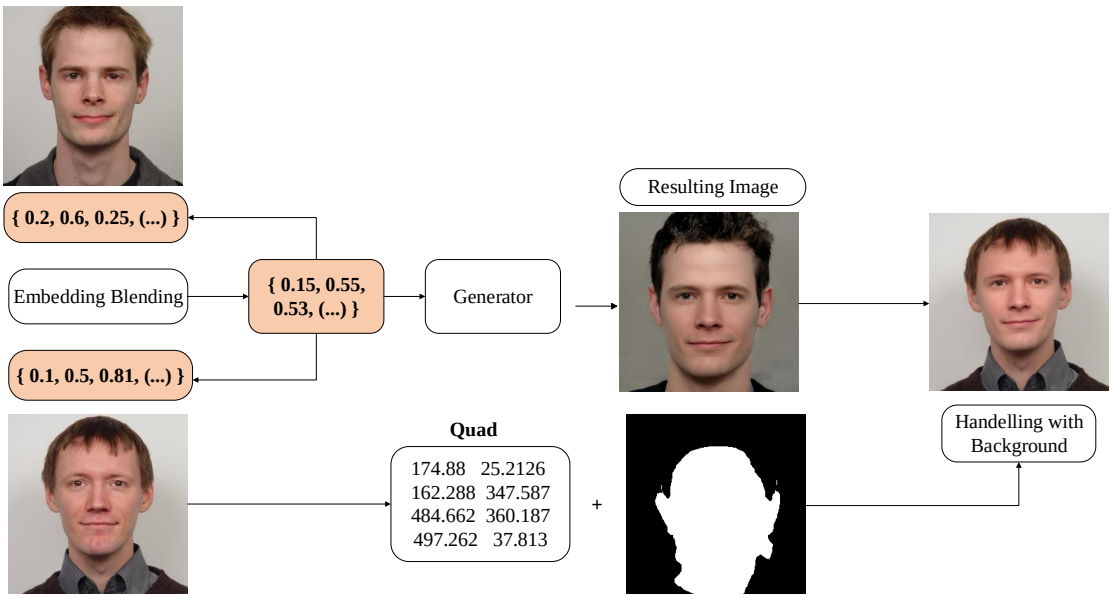

Figure 4.4: StyleGAN interpolation pipeline.

The process of generating a morphed image involved interpolating the embeddings into the latent domain, which represents a high-dimensional vector space where each dimension corresponds to different features or attributes of the generated data. These latent variables are commonly sampled from a probability distribution (Gaus-

sian) and are then used as inputs to the Generative Adversarial Network (GAN) generator.

During the interpolation procedure, a linear interpolation is calculated between the two latent vectors, which ensures that the interpolated embeddings lie in a straight line within the latent space. In this sense, the linearity of the latent domain is crucial to allowing a smooth transition between faces during the face morphing process.

After generating the image, a context restoring step was also performed, replacing the background of the generated image with the original. This was done using a mask to isolate the face region in the generated image and a vector known as *quad* representing the coordinates of a quadrilateral surrounding the face in the original image, which is used to preserve the original background region.

4.2.3 Selfmorph Approach

One of the drawbacks associated with landmark-based morphed images is the presence of artifacts, including ghosting, noise, and blur. These artifacts can have a significant impact on the quality and reliability of morph images, making them easier to identify for both humans and FR algorithms.

On the other hand, the training procedure becomes unrealistic and biased by focusing on learning these artifacts. Thus, to generalize the detection performance and reduce overfitting due to artifact detection, *selfmorphs* [124] were generated.

As the name suggests, *selfmorphs* are generated using images of the same individual, resulting in a final image that still represents that original individual. Therefore, the difference in the image is solely attributed to morphing artifacts and not due to different identities, as would be the case in traditional morphing scenarios.

Given this context and following the idea proposed by Medvedev *et al.* [89], in this dissertation *selfmorphs* were considered *bonafide* images. This allowed prioritizing detection based on the behavior of deep features over artifact detection, following the assumption that these discriminative features are maintained after the *selfmorphing* procedure.

Figure 4.5 schematizes a practical example of a morph image in each of the approaches mentioned above.

Figure 4.5: Samples of morph images for each approach. Order: StyleGAN-based, landmark-based (LDM), LDM *selfmorph* and finally StyleGAN *selfmorph* approach.

Analyzing the images, it can be observed that there are visible artifacts in the landmark approach, while the StyleGAN approach presents a more realistic image. In the case of *selfmorphs*, the final identity is preserved.

4.3 Alignment Settings

The main purpose of this dissertation is to evaluate the influence of an image's context on the detection of morphing attacks. In this sense, different alignment conditions were defined in order to vary the relationship between the face and the background of the image.

Figure 4.6: Facial image aligned according to the different alignment settings.

Related to the alignment procedure, initially, face detection was accomplished by utilizing the Multi-Task Cascaded Convolutional Neural Network (MTCNN) algorithm [125] which as input receives a face image and provides a set of facial landmarks ({left eye}, {right eye}, {nose}, {left mouth corner}, {right mouth corner}), a face bounding box, and a confidence score that reflects the level of certainty regarding the validity of the identified face points.

The face alignment was then performed by applying a rigid transformation to minimize the coordinate distance between those five facial landmarks (red points in figure 4.7) and a predefined set of target coordinates (blue points in figure 4.7), which for the resulting image size 112×112 can be defined as {{38.2, 41.7}, {73.5, 41.5}, {56.0, 61.7}, {41.5, 82.4}, {70.7, 82.2}}) [55].

The specific settings used in the alignment procedure involved scaling this target coordinate set using several scale factors, which are presented in table 4.1. It is important to highlight that, as a final result, all images were set to 300 × 300 pixels.

On the other hand, it was also necessary to take into account possible problems in the face detection process, where two typical scenarios can appear: 1) Face detection failure or 2) Multiple face detection.

In the first case, a crop was performed using the same scale factor in order to replace the conventional alignment. In the second case, the central face was selected, and the alignment followed the general procedure.

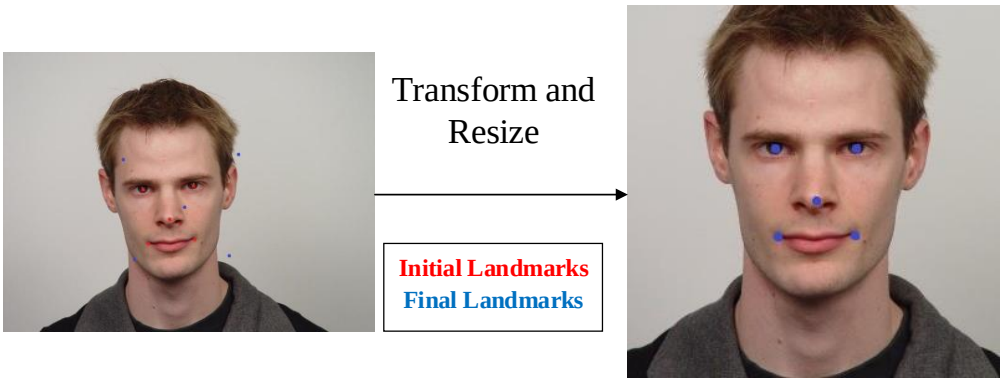

Figure 4.7: Alignment procedure pipeline.

For each selected scale factor, in table 4.1 the respective indicative ratio of the face's occupancy area in the image under each alignment condition is presented. This value represents the ratio of the face area, limited by a face contour (determined by the detection of 68 landmarks [23]) to the full image area.

Table 4.1: Summary table of all alignment conditions with their respective scale factors and ratios.

Alignments	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**
Scale Factor	1.65	1.40	1.10	1.00	0.90	0.85	0.80	0.75	0.70	0.65	0.60
Ratio	0.15	0.21	0.34	0.42	0.51	0.56	0.62	0.70	0.77	0.86	0.94

4.4 Single Image MAD Approach

In this dissertation, the *no-reference* face morphing detection was approached using two variations: the *fused classification* approach and the *binary classification* approach.

4.4.1 Fused Classification Model

To implement this approach, the method proposed by Medvedev *et al.* [89] was followed.

The pipeline involves training two networks simultaneously, which are specifically designed to learn high-level identity discriminative features by performing classification tasks. These features can then indicate the presence of face morphing.

In the section 4.2, a labeling strategy for morphs was described, which basically involves paring images between two different sub-lists. In this setup, each network labeled the morphs accordingly with that sub-list, *i.e., First* network considers the first source image from each pair (first sub-list) and the *Second* network the second. For *bonafides* images, the original label is duplicated.

Beyond that, the extracted features are compared using a similarity metric (based on the dot product), which represents the *morphing detection score*. The notations $\dot{value}$ and $\ddot{value}$ are used to simplify the representation for the *First* and *Second* networks, respectively.

The overall pipeline is depicted in figure 4.8.

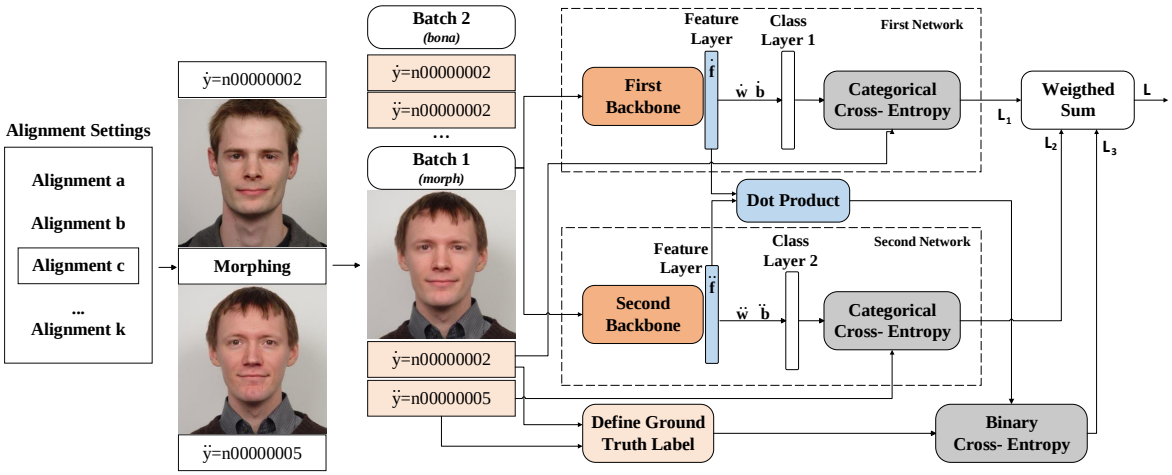

Figure 4.8: Single Morphing Attack Detection (S-MAD) model schema for *fused classification* approach. In order to simplify the visualization, a single image is shown per batch.

Revisiting the formula 3.1 the training process is regularized by the losses:

$$L_1 = -\frac{1}{N}\sum_{i}^{N}\log(\frac{e^{\dot{W}_{\dot{y}_i}^T\dot{f}_i+\dot{b}_{\dot{y}_i}}}{\sum_j^C e^{\dot{f}_{\dot{y}_j}}}) \tag{4.1}$$

$$L_2 = -\frac{1}{N}\sum_{i}^{N}\log(\frac{e^{\ddot{W}_{\ddot{y}_i}^T\ddot{f}_i+\ddot{b}_{\ddot{y}_i}}}{\sum_j^C e^{\ddot{f}_{\ddot{y}_j}}}) \tag{4.2}$$

where f_i represents the deep features of the i-th sample, y_i represents the class index of the i-th sample, W represents the weights, b represents the biases of the last Fully Connected (FC) layer, N represents the number of samples per batch, and C represents the total number of classes (identities).

The *morphing detection score* is computed by taking the dot product of the backbone outputs ($\dot{f} \cdot \ddot{f}$) and is then activated using the *sigmoid* function. The loss function is defined as binary cross entropy.

$$L_3 = -\frac{1}{N}\sum_{i}^{N} t\log\frac{1}{1+e^{-\ddot{f}\cdot\dot{f}}} + (1-t)\log\left(1-\frac{1}{1+e^{-\ddot{f}\cdot\dot{f}}}\right) \tag{4.3}$$

where t represents the ground truth label, which is obtained by comparing the input class labels ($\dot{y}_i$ and $\ddot{y}_i$).

$$t = 1 - |sgn(\dot{y}_i - \ddot{y}_i)| \tag{4.4}$$

Note that, for simplicity, the index i relative to the i-th sample has been omitted in ($\dot{f} \cdot \ddot{f}$).

The optimization process involves combining the individual losses (L_1, L_2, L_3) as a weighted sum, resulting in an overall loss (denoted as L). By minimizing this loss configuration, the model learns discriminative facial features for effective morphing detection.

$$L = \alpha_1 L_1 + \alpha_2 L_2 + \beta L_3 \tag{4.5}$$

The values of α_1, α_2 and β control the weight of each loss function in the final minimization.

4.4.2 Binary Classification Model

Assuming the mathematical formulation of the model described in subsection 4.4.1, a modification was implemented to perform *binary classification* (morph or non-morph) using a single network. From a general standpoint, this modification

involves removing the identity classification component present in the *fused classification* approach.

The training process is then regularized only by the binary cross-entropy loss $L = L_3$, depicted in the formula 4.3. However, in this particular case, the input for this loss is not the dot product of the backbone outputs ($\dot{f} \cdot \ddot{f}$) from both networks but rather the output of the single network f.

The model pipeline is presented in figure 4.9.

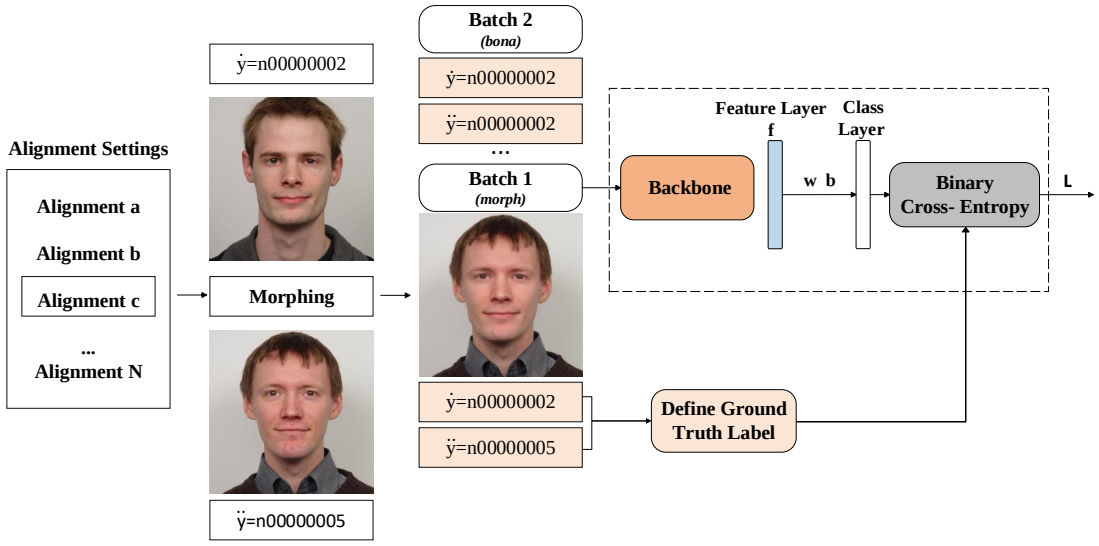

Figure 4.9: S-MAD model schema for *binary classification* approach. In order to simplify the visualization, a single image is shown per batch.

4.5 Differential MAD Approach

In a *reference-based* scenario, Morphing Attack Detection (MAD) was approached through a *fused classification* schema similar to the one presented in S-MAD case. However, in this particular case, the *First* and *Second* networks do not receive the same single image, and an image pair is used.

The *First* network receives the "enrolled image", which may or may not be a morphed image, while the *Second* network receives the "live capture image", which is always *bonafide*. The output of each backbone (feature vector) represents the respective input image. In relation to the mathematical formulation, it is identical to the one presented for the S-MAD case in subsection 4.4.1.

The model pipeline is presented in figure 4.10.

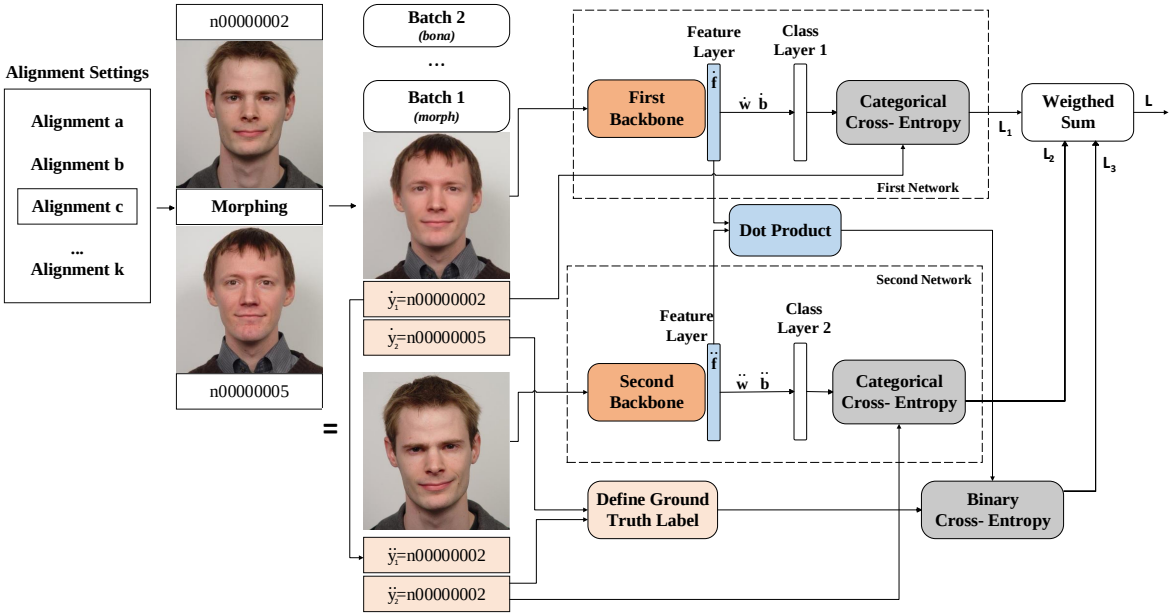

Figure 4.10: Differential Morphing Attack Detection (D-MAD) *fused classification* approach schema. In order to simplify the visualization, a single image pair is shown per batch. Note that each image has two identity labels y_1 and y_2.

To obtain the Complementary Image (CI), that feeds the *Second* Network, a list of matches was created based on label similarity and the requirement that only non-morph images are accepted ("live capture image"). Then, the CI was selected based on the first label, which should be identical to the first label of the image processed by the *First* network. Although the formulation is identical to the S-MAD *fused classification* case, it should be noted that the L_1 loss computation is done based on the first label of the image processed by the *First* network, and the L_2 loss is computed based on the second label of the image processed by the *Second* network.

4.6 Benchmarking

Evaluation benchmarks have the function of providing valuable insights about the performance of a particular model. However, the main problem in MAD field is that for research purposes, the existence of public benchmarking protocols is limited. For those that exist, such as the Face Recognition Vendor Test (FRVT) National Institute of Standards and Technology (NIST) MORPH [90] and the FVC-onGoing MAD [126], have a number of submission restrictions.

The idea for this dissertation was to use a series of protocols using various public datasets with the goal of performing robust benchmarking to evaluate and compare the performance of different models with different parameters. For that purpose, the open-source morphing benchmarking utilities [1] were employed with some modifications.

The default suggested protocols involved sharing images with our training data. Therefore, a modification was made by replacing the subset *bonafide* images with FRLL-Set [127], PICS-Utrecht [112], EFIEP [114] and MIT-CBCL [115]. Related to morphs FRLL-Morphs [120] and Dustone [123] were used.

The protocols are described schematically in table 4.2. It should be noted that all the protocols share the same list of *bonafide* images.

Table 4.2: Benchmark protocols for both single and differential cases.

Name	**S-MAD**	**D-MAD**
protocol-asml	∼ 2k morphs (FRLL-Morphs) ; <1k *bonafide*s (FRLL-Set + Utrecht + MIT-CBCL + EFIEP)	∼ 4.3k (FRLL-Morphs vs FRLL-Set) and ∼ 60 (Utrecht vs Utrecht) and ∼ 100 (FRLL-Set vs FRLL-Set)
protocol-opencv	∼ 1.3k morphs (FRLL-Morphs) ; <1k *bonafide*s (FRLL-Set + Utrecht + MIT-CBCL + EFIEP)	∼ 2.4k (FRLL-Morphs vs FRLL-Set) and ∼ 60 (Utrecht vs Utrecht) and ∼ 100 (FRLL-Set vs FRLL-Set)
protocol-real	∼ 3k morphs (Dustone + FRLL-Morphs) ; <1k *bonafide*s (FRLL-Set + Utrecht + MIT-CBCL + EFIEP)	
protocol-facemorpher	∼ 2k morphs (FRLL-Morphs) ; <1k *bonafide*s (FRLL-Set + Utrecht + MIT-CBCL + EFIEP)	∼ 2.4k morphs (FRLL-Morphs) and ∼ 60 (Utrecht vs Utrecht) and ∼ 100 (FRLL-Set vs FRLL-Set)

[1] https://github.com/iurii-m/MorDeephy.git

protocol-webmorph	∼ 1k morphs (FRLL-Morphs) ; <1k *bonafides* (FRLL-Set + Utrecht + MIT-CBCL + EFIEP)	∼ 2.4k morphs (FRLL-Morphs) and ∼ 60 (Utrecht vs Utrecht) and ∼ 100 (FRLL-Set vs FRLL-Set)
protocol-stylegan	∼ 2k morphs (FRLL-Morphs) ; <1k *bonafides* (FRLL-Set + Utrecht + MIT-CBCL + EFIEP)	∼ 2.4k morphs (FRLL-Morphs) and ∼ 60 (Utrecht vs Utrecht) and ∼ 100 (FRLL-Set vs FRLL-Set)

In the case of D-MAD, the definition of image pairs was necessary. It was determined that only the first image in each pair could be a morphed image.

The protocol names are based on how the morphs were generated, including approaches like Style-GAN2 [119] (*protocol-stylegan*), WebMorph [121] (*protocol-webmorph*), AMSL [128] (*protocol-asml*), FaceMorpher [66] (*protocol-facemorpher*) and OpenCV [65] (*protocol-opencv*).

In terms of their characteristics, *protocol-real* and *protocol-asml* include morphs that exhibit minimal visible blending artifacts, reflecting more realistic images for human perception, *protocol-facemorpher* and *protocol-opencv* include simple morphs that exhibit foreground and background artifacts. In *protocol-webmorph*, the artifacts are found dominantly in the background, which is possibly the most challenging.

Figure 4.11: Sample images from the benchmark protocols. The first row contains morph images.

4.7 Grad-CAM Approach

In addition to evaluating the model's performance using the $BPCER@APCER$ metric, an analysis was also conducted to investigate the influence of different regions of the input image on the final output prediction using the Gradient-weighted Class Activation Mapping (Grad-CAM) technique [129].

Detailed modeling and explanation of this technique are out of scope for this dissertation. However, from a general standpoint, Grad-CAM uses the feature maps produced by the last convolutional layer of a Convolutional Neural Network (CNN). Then, by projecting the weights of the output layer onto these feature maps, it is possible to highlight the important regions in the input image.

4.7.1 Heatmaps Computation

Following the Grad-CAM approach, the main objective was to obtain a final heatmap that highlights the important regions of the input image, providing valuable insights into the decision-making process of the model.

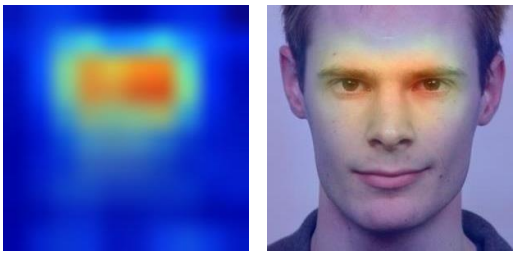

Figure 4.12: Grad-CAM sample heatmap and its overlaid sample image.

In the models with two network backbones, the gradient for real binary classification (morph and non-morph) was computed by taking the dot product between the feature embeddings with respect to the activations of the last convolutional layers of both networks. On the other hand, in the case of a single network, the model performs a regular classification task. In this case, the gradient was obtained directly by tracking the activations of the predicted class.

In both cases, the gradients are separated into morph and *bonafide* cases. Thus, in the end, for each benchmark protocol, two average heatmaps were obtained for both morph and *bonafide* cases.

By visualizing each heatmap and overlaying it on the face image, the regions

with more influence on the final result can be perceptually observed. However, to provide a more rigorous analysis, the *Average of the Gradient Intensity Ratio (AGIR) (foreground/background)* was computed in order to obtain a numerical value instead of relying solely on the visual perception of activations. This procedure was employed for all alignment conditions.

4.7.2 Average of the Gradient Intensity Ratio

The general idea was to overlay a mask on each heatmap, dividing it into foreground and background regions. This way, it was possible to associate a certain area of the heatmap with the respective part of the image and subsequently calculate the average intensity of the gradient in that region. The pipeline is presented in figure 4.13.

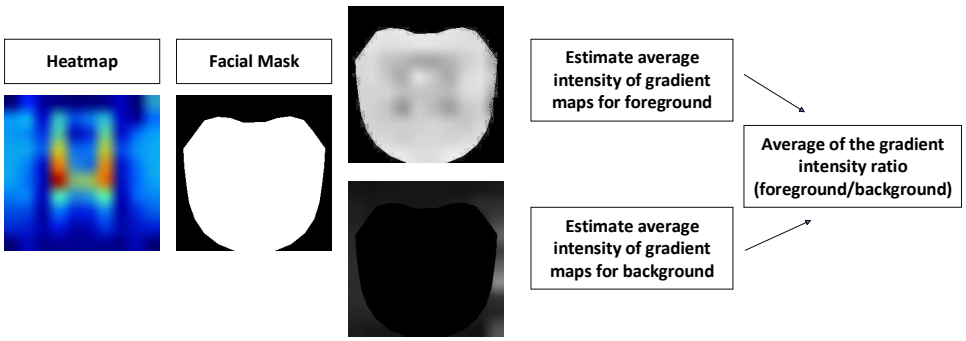

Figure 4.13: Schematic representation of the methodology to obtain the average intensity of gradient maps for the foreground and background and the respective ratio.

To determine the facial contour for each image, 68 landmarks were used. Then an average facial mask was computed for each benchmark protocol. It should be noted that this was done for both *bonafide* and morph cases, *i.e.,* as the final result for each alignment condition and respective protocol, two binary average facial masks were obtained.

After separating the gradient maps into two regions, the average gradient intensity for each region was computed using only the non-zero pixels, which correspond to the regions of interest defined by the respective masks. The final output is then obtained by performing a simple ratio between the two average intensities.

4.8 Architecture Choice

Given the objective of this dissertation, in training, an already well-defined CNN architecture was used as the backbone. Following that, one of the first stages was the choice of this architecture.

As depicted in chapter 3, *ResNets* are commonly used in most state-of-the-art approaches. However, for this dissertation, the choice of architecture did not follow the same path, since the one chosen was the *EfficientNet* [51], specifically version B3, which achieves state-of-the-art 81.6% top-1 accuracy on ImageNet [130].

The motivation stems from the fact that *EfficientNetB3* is an architecture that exhibits higher accuracy than other neural network architectures while using fewer computational resources. The superior performance of the network can be attributed to its optimized use of parameters and the compound scaling method it employs. Summing up, the *EfficientNetB3* model enables increased network efficiency by optimizing its depth, width, and resolution.

4.9 Implementation Utils

Related to the code, the implementation was done in Python 3.8 using TensorFlow 2.5. The training procedure was performed using NVIDIA RTX 3090 GPU. Related to the model formulation, the *EfficientNetB3* architecture was imported from the Keras library, along with all of the activation functions used, namely *sigmoid* and *softmax*.

5

Experiments and Results

5.1 Training Settings

In the work developed, one of the crucial steps was the selection of the appropriate hyperparameters in order to optimize the model. To accomplish this, several training sets were performed.

In all these experiments, as mentioned before, the backbone architecture used was *EfficientNetB3*, initialized with weights pre-trained on the ImageNet dataset, returning in the end 512 deep features. The batch size was set to 28 images. The optimizer employed was stochastic gradient descent with a momentum parameter of 0.9.

It should be noted that these experiments were performed for only one of the alignment settings and carried out specifically for the Single Morphing Attack Detection (S-MAD) *fused classification* approach. For the remaining models, the hyperparameters were set in a similar manner.

Revisiting equation 4.5, for effective convergence and further morphing detection, it was necessary to choose the appropriate balance among the components of the loss function (α_1, α_2, and β). In that sense, different proportional settings were tested. The performance evaluation was performed by analyzing the BonaFide Presentation Classification Error Rate (BPCER) value by setting the Attack Presentation Classification Error Rate (APCER) value to 0.1 and 0.01.

It is important to highlight that in the context of S-MAD *binary classification* approach, the weight loss parameters α and β are not relevant as they do not contribute to the computation of the loss.

Table 5.1: BPCER@APCER = (0.1, 0.01) of S-MAD *fused* model for various weight loss proportions in different protocols. Considering $\alpha = \alpha_1 = \alpha_2$.

Weight Loss	BPCER@APCER = δ											
	Protocol-asml		Protocol-facemorpher		Protocol-opencv		Protocol-stylegan		Protocol-webmorph		Protocol-real	
	δ =0.1	δ =0.01	δ=0.1	δ=0.01	δ=0.1	δ =0.01	δ=0.1	δ =0.01	δ = 0.1	δ =0.0	δ = 0.1	δ = 0.01
α/β = **0.1**	**0.364**	0.773	0.152	0.452	0.264	0.551	0.748	0.940	0.660	0.919	0.284	0.741
α/β =**0.2**	0.367	**0.723**	**0.128**	**0.277**	**0.183**	**0.433**	**0.384**	**0.683**	**0.479**	**0.822**	**0.309**	**0.664**
α/β =0.5	0.779	0.931	0.275	0.461	0.392	0.647	0.541	0.838	0.851	0.968	0.695	0.909
α/β =0.75	0.451	0.784	0.278	0.577	0.306	0.602	0.436	0.742	0.807	0.954	0.414	0.741
α/β =1	0.598	0.842	0.408	0.732	0.367	0.670	0.688	0.845	0.842	0.935	0.546	0.813

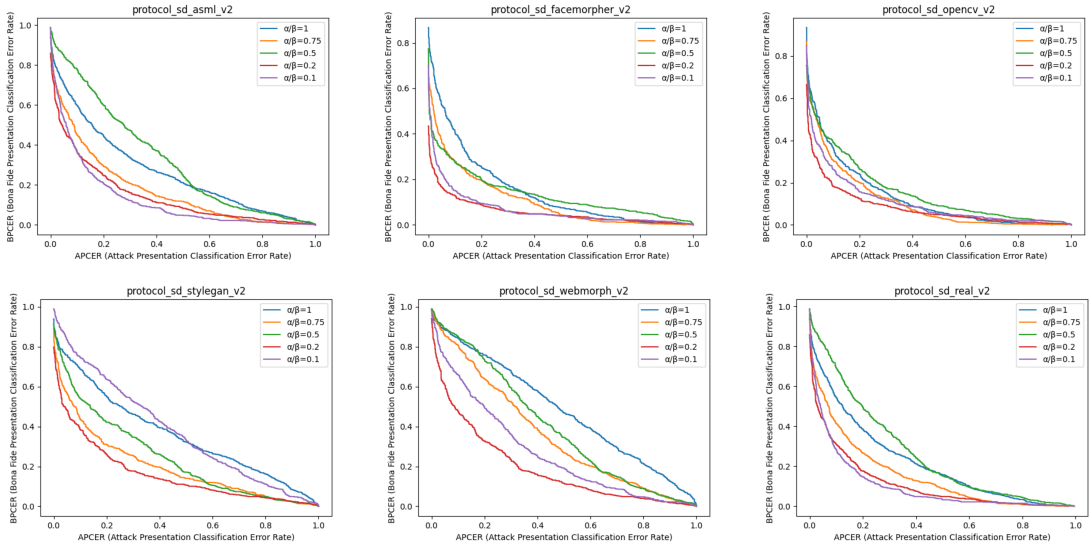

Figure 5.1: Detection Error Trade-off (DET) curves for various α/β values in the different protocols.

Based on table 5.1, the value α/β=0.2 outperforms all others in the different protocols, and is therefore the chosen value. The next step involved determining the optimal number of epochs for the model. For that, the value α/β=0.2 was set, and different epoch numbers were tested, with 5 being the chosen value (table 5.2).

Table 5.2: BPCER@APCER = (0.1, 0.01) of S-MAD *fused* model for various epoch numbers in different protocols, fixing the value α/β= 0.2 for weight loss.

Epochs Number	BPCER@APCER = δ											
	Protocol-asml		Protocol-facemorpher		Protocol-opencv		Protocol-stylegan		Protocol-webmorph		Protocol-real	
	δ =0.1	δ =0.01	δ=0.1	δ=0.01	δ=0.1	δ =0.01	δ=0.1	δ =0.01	δ = 0.1	δ =0.0	δ = 0.1	δ = 0.01
5	0.446	0.798	0.106	0.297	0.200	0.464	0.592	0.887	0.633	0.906	0.355	0.736
10	0.695	0.926	0.174	0.541	0.297	0.658	0.817	0.922	0.837	0.979	0.594	0.915
15	0.421	0.811	0.122	0.414	0.208	0.577	0.577	0.817	0.834	0.972	0.324	0.763

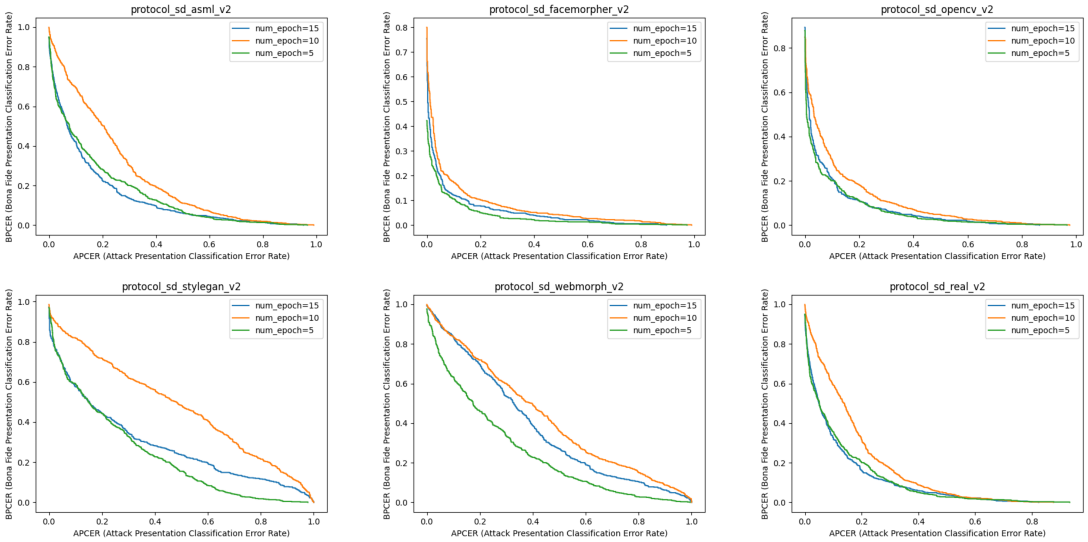

Figure 5.2: DET curves for various epoch values in the different protocols, fixing the value α/β= 0.2 for weight loss.

Similarly, after fixing both the previous values, different values for the initial learning rate were also tested. Based on table 5.3 values, the best result was 0.075, resulting in the choice of a learning rate that decays linearly from 0.075 to 1e-5.

Table 5.3: BPCER@APCER = (0.1, 0.01) of of S-MAD *fused* model for various learning rate values in different protocols, fixing the value α/β= 0.2 for weight loss and 5 for the number of epochs.

Learning Rate	BPCER@APCER=δ											
	Protocol-asml		Protocol-facemorpher		Protocol-opencv		Protocol-stylegan		Protocol-webmorph		Protocol-real	
	δ =0.1	δ =0.01	δ=0.1	δ=0.01	δ=0.1	δ =0.01	δ=0.1	δ =0.01	δ = 0.1	δ =0.0	δ = 0.1	δ = 0.01
0.1	0.130	0.502	0.082	0.345	**0.047**	0.225	0.732	0.918	0.475	0.791	0.081	0.329
0.05	0.262	0.629	0.275	0.664	0.275	0.664	0.887	0.976	0.482	0.803	0.196	0.592
0.075	**0.070**	**0.314**	**0.040**	**0.191**	0.053	**0.153**	0.592	0.853	**0.298**	**0.574**	**0.045**	**0.268**
0.001	0.324	0.594	0.144	0.314	0.215	0.493	**0.483**	**0.710**	0.331	0.654	0.330	0.623
1eˆ-5	0.773	0.966	0.820	0.9779	0.835	0.993	0.920	0.994	0.716	0.972	0.820	0.984

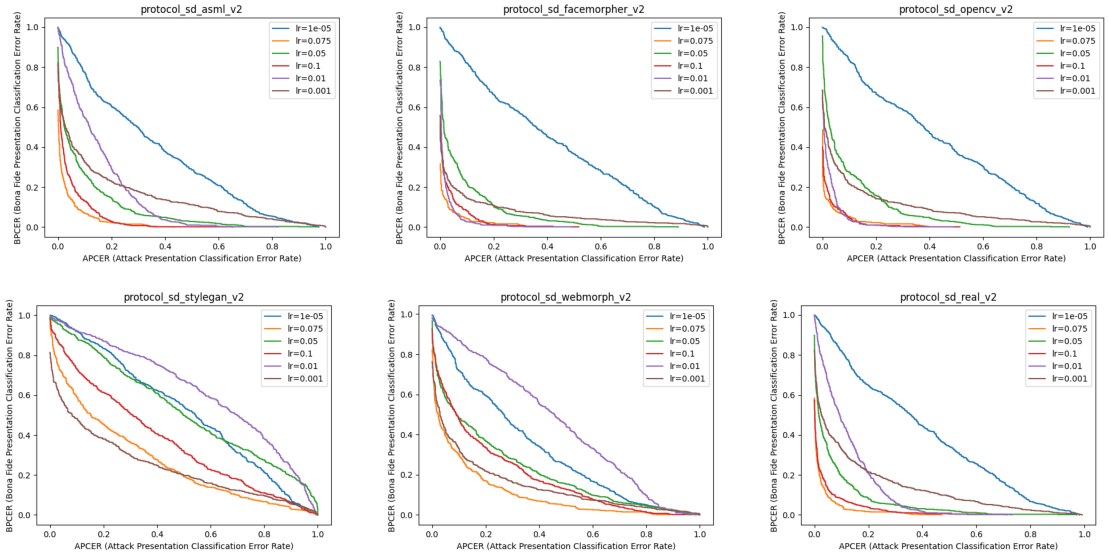

Figure 5.3: DET curves for various learning rate values in the different protocols, fixing the value α/β= 0.2 for weight loss and 5 for the number of epochs.

Once the hyperparameters were defined, separate training experiments were performed for each alignment setting (a to k) on the concatenated dataset, comprising original images (ICMD dataset), landmark-based morphed images, StyleGAN morphed images, and *selfmorphs*.

5.2 Benchmark Results

In this section, the performance results of both *no-reference* and *reference* models will be analyzed.

5.2.1 S-MAD Binary Classification Model

The S-MAD *binary classification* approach refers to the model where only a single network was used in a straight binary classification task. The results related to BPCER@APCER are presented in the following tables, as well as the respective DET curves in figure 5.4.

Table 5.4: BPCER@APCER = (0.1, 0.01) of S-MAD *binary classification* model across all the alignment settings for each benchmark protocol.

Alignments	BPCER@APCER=δ											
	Protocol-asml		**Protocol-facemorpher**		**Protocol-opencv**		**Protocol-stylegan**		**Protocol-webmorph**		**Protocol-real**	
	δ =0.1	**δ =0.01**	**δ =0.1**	**δ =0.01**	**δ =0.1**	**δ =0.01**	**δ =0.1**	**δ =0.01**	**δ =0.1**	**δ =0.01**	**δ =0.1**	**δ =0.01**
a	0.199	0.622	0.125	0.558	0.199	0.663	0.663	0.663	0.523	0.663	0.191	0.568
b	0.143	0.380	0.131	**0.387**	0.144	0.440	0.586	0.586	0.340	0.586	**0.144**	**0.396**
c	0.365	0.630	0.331	0.675	0.320	0.676	0.676	0.676	0.489	0.676	0.351	0.630
d	0.236	0.511	0.161	0.549	0.161	0.489	0.623	0.623	0.436	0.623	0.246	0.511
e	**0.141**	**0.348**	**0.102**	0.532	**0.080**	**0.424**	0.710	0.710	**0.321**	0.641	0.194	0.463
f	0.199	0.455	0.127	0.551	0.125	0.533	0.675	0.675	0.328	0.579	0.215	0.478
g	0.158	0.373	0.106	0.532	0.209	0.532	0.586	0.586	0.348	0.586	0.175	0.411
h	0.330	0.580	0.138	0.682	0.093	0.486	0.724	0.724	0.486	0.724	0.306	0.577
i	0.214	0.408	0.174	0.476	0.149	0.442	**0.573**	**0.573**	0.396	**0.573**	0.212	0.430
j	0.221	0.465	0.187	0.596	0.141	0.457	0.776	0.776	0.475	0.682	0.233	0.504
k	0.243	0.498	0.194	0.557	0.146	0.513	0.794	0.794	0.467	0.707	0.262	0.573

Table 5.5: Overall performance across all benchmark protocols for S-MAD *binary classification* approach.

BPCER@APCER=δ	a	b	c	d	**e**	**f**	**g**	h	i	j	k
δ=0.1	0.317	0.248	0.442	0.320	0.258	0.278	0.263	0.346	0.286	0.338	0.351
δ=0.01	0.623	0.463	0.660	0.551	0.520	0.545	0.503	0.629	0.484	0.580	0.607

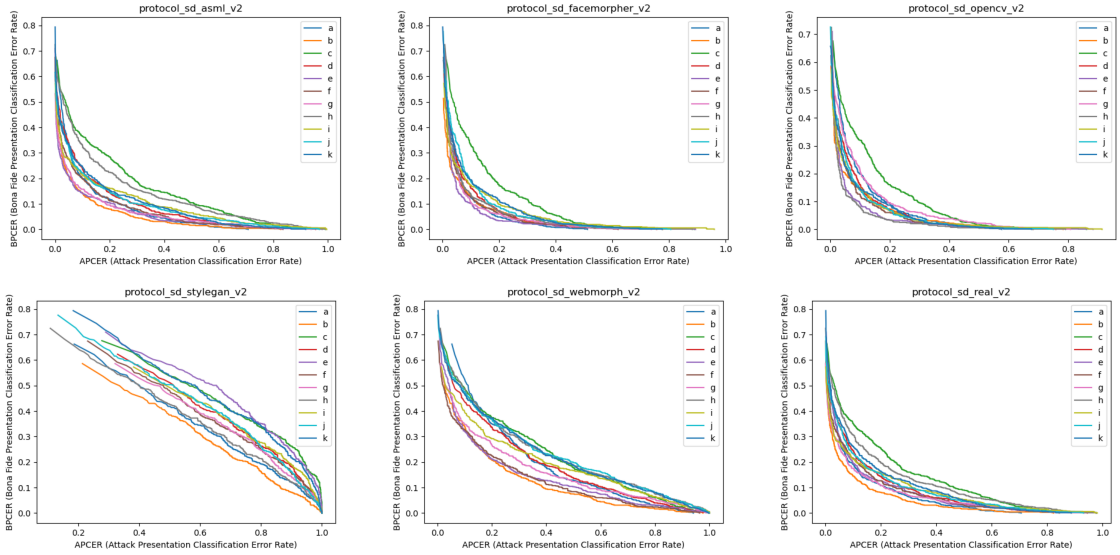

Figure 5.4: DET curves across the different alignment settings (a to k) for S-MAD *binary classification* approach. Each subplot represents one of the benchmark protocols.

Based on the results presented, the alignment settings between e and g appear to represent the optimal range of values across all the benchmark protocols, with the e alignment setting being the potential optimal case.

This optimal range of settings is reinforced based on the values presented in table 5.5. It is important to note that these values were obtained by calculating a simple average across all protocols, which is very simplistic in view of the complexity of the problem since it does not take into account the characteristics of each protocol or the differences and challenges that each one represents. In that sense, the alignment settings *b* and *i* at first sight also seem to be possible optimal candidates. However, analyzing the values presented in table 5.4, this final average value is possibly influenced by the performance values obtained in the *protocol-stylegan*, which are very discrepant (lower values, namely when compared to the optimal one *e* value (0.710)) influencing the final average value.

It is important to highlight that we observed instability in the training process across various alignment settings. Specifically, several cases, such as *b*, experienced difficulties due to poor initial convergence and were retrained. This inconsistency has affected the stability of our results. We believe that the trivial nature of the binary classification task, combined with the complexity of differentiating face morph features, may be contributing factors to this issue.

According to the explanation provided in section 4.7, the evaluation of the impact of different regions of the input image on the final output prediction was also performed. This was done by obtaining the Gradient-weighted Class Activation Mapping (Grad-CAM) heatmaps and analyzing the behavior for each alignment condition.

For the sake of compactness, we present those map results only for the *protocol-asml* (figure 5.5).

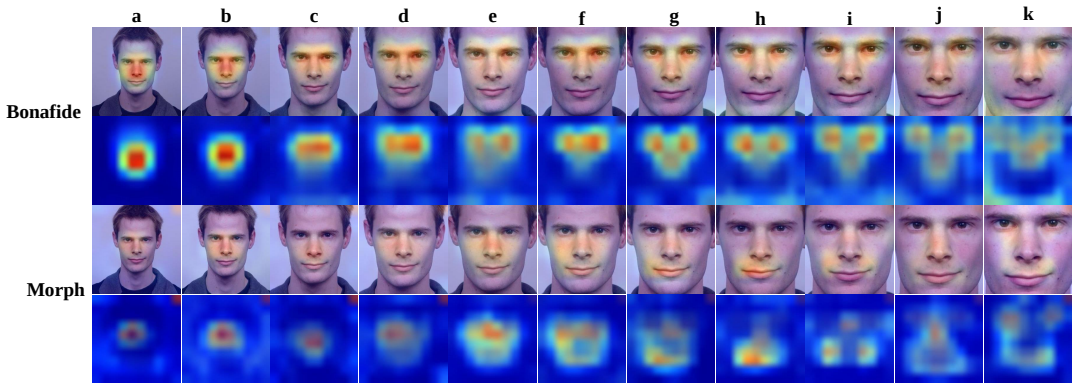

Figure 5.5: Grad-CAM heatmaps across all the alignment settings for S-MAD *binary classification* approach.

The face/foreground is mostly dominantly activated across all the alignment

Table 5.6: Summary table for the values of *Average of the Gradient Intensity Ratio (AGIR)* in the different protocols, as well as the average value for the morphs. Note that for *bonafide* only one column is shown since the values are the same for all protocols (*bonafide* set is similar). Values for S-MAD *binary classification* approach.

Alignments	**AGIR morph values**						**AGIR value for bona**	**Average AGIR value for morphs**
	asml	facemorpher	opencv	stylegan	webmorph	real		
a	1.754	2.824	2.394	0.546	0.776	1.841	4.481	1.689
b	1.844	3.094	3.298	0.687	1.107	2.114	3.354	2.024
c	1.797	3.352	3.790	0.879	1.574	2.129	2.017	2.254
d	2.164	3.610	3.725	1.140	1.238	2.469	1.811	2.391
e	3.586	4.458	4.796	1.443	2.060	3.587	1.589	3.322
f	2.775	3.675	4.393	1.835	1.902	2.957	1.397	2.922
g	2.067	3.726	3.526	1.419	1.246	2.459	1.774	2.407
h	1.765	3.393	4.405	2.079	1.376	2.079	1.534	2.516
i	2.337	3.717	4.281	1.461	1.770	2.579	1.221	2.691
j	2.475	3.550	3.868	1.369	1.581	2.532	1.238	2.563
k	1.629	2.067	2.586	1.270	1.313	1.857	0.951	1.787

settings for both morph and *bonafide cases*. The values presented in the table 5.6 also confirm that.

5.2.2 S-MAD Fused Classification Model

In a similar way, the results related to BPCER@APCER are presented in the following tables, as well as the respective DET curves, in figure 5.6.

Table 5.7: BPCER@APCER = (0.1, 0.01) of S-MAD *fused classification* model across all the alignment settings for each benchmark protocol.

Alignments	**BPCER@APCER=δ**											
	Protocol-asml		**Protocol-facemorpher**		**Protocol-opencv**		**Protocol-stylegan**		**Protocol-webmorph**		**Protocol-real**	
	δ=0.1	**δ=0.01**	**δ =0.1**	**δ=0.01**	**δ =0.1**	**δ=0.01**	**δ =0.1**	**δ=0.01**	**δ=0.1**	**δ=0.01**	**δ=0.1**	**δ=0.01**
a	0.159	0.689	0.187	0.517	0.239	0.599	0.842	0.946	0.606	0.885	0.137	0.608
b	0.063	0.495	0.072	0.646	0.099	0.658	0.671	0.946	0.702	0.964	0.081	0.427
c	0.125	0.467	0.215	0.588	0.240	0.566	0.694	0.884	0.541	0.859	0.167	0.455
d	0.040	0.374	0.102	0.558	0.103	0.568	0.574	0.835	0.305	0.781	0.113	0.421
e	0.162	0.580	0.149	0.582	0.177	0.602	0.566	0.767	0.605	0.870	0.138	0.549
f	0.184	0.530	0.180	0.488	0.175	0.451	0.582	0.788	0.517	0.785	0.158	0.479
g	**0.034**	**0.233**	**0.025**	0.701	**0.037**	0.701	0.487	0.875	**0.216**	0.788	**0.072**	**0.322**
h	0.168	0.642	0.168	0.535	0.165	0.599	0.536	0.850	0.542	0.854	0.138	0.594
i	0.046	0.255	0.036	**0.365**	0.044	**0.390**	**0.305**	**0.583**	0.246	**0.554**	0.094	0.365
j	0.287	0.630	0.268	0.585	0.262	0.564	0.844	0.959	0.697	0.907	0.228	0.574
k	0.193	0.652	0.253	0.745	0.262	0.792	0.825	0.953	0.674	0.915	0.178	0.611

Table 5.8: Overall performance across all benchmark protocols for S-MAD *fused classification* approach.

BPCER@APCER=δ	a	b	c	**d**	**e**	**f**	**g**	**h**	**i**	j	k
δ=0.1	0.361	0.281	0.330	0.206	0.299	0.299	0.145	0.286	0.129	0.431	0.398
δ=0.01	0.732	0.690	0.636	0.589	0.658	0.586	0.603	0.679	0.418	0.703	0.778

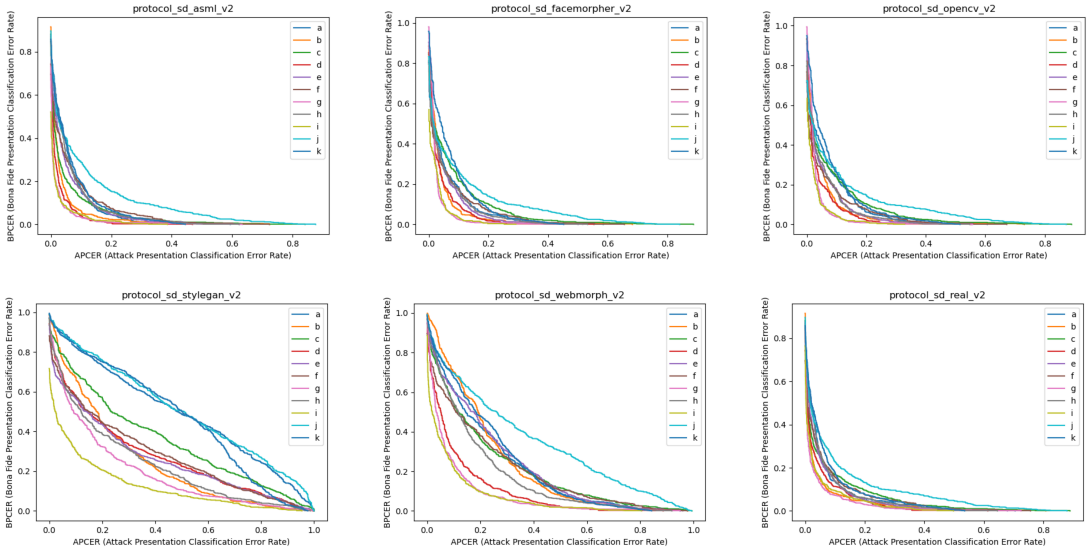

Figure 5.6: DET curves across the different alignment settings (a to k) for S-MAD *fused classification* approach. Each subplot represents one of the benchmark protocols.

Looking at the values, it seems that the optimal range of values lies between the alignment settings of d and i. Within this range, alignment setting g is still possibly the optimal case.

Similarly, looking at the values presented in table 5.8, the alignment definition i seems to be the potential optimal case rather than g. However, based on the values depicted in table 5.7 in this case the alignment condition i values are influenced by *protocol-stylegan*. On the other hand, condition g outperforms all the others for *protocol-asml* and *protocol-real*, which give the most realistic images for human perception out of the others.

At the same time, this detection technique allowed to achieve superior results in comparison to the *binary classification* case, namely comparing e (optimal in S-MAD binary) and g (optimal in S-MAD fused). This result is probably related to the optional regularization effect imposed by the face recognition task.

Making a visual analysis of Grad-CAM heatmaps, it is possible to observe that the detection focuses mainly on the face region and, in many cases, on the intersection regions between foreground and background, as shown in figure 5.7.

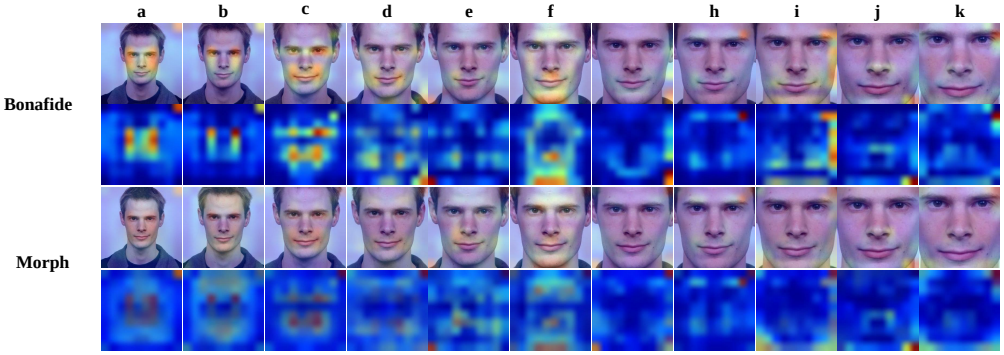

Figure 5.7: Grad-CAM heatmaps across all the alignment settings for S-MAD *fused classification* approach.

Table 5.9: Summary table for the values of *AGIR* in the different protocols, as well as the average value for the morphs. Note that for *bonafide* only one column is shown since the values are the same for all protocols (*bonafide* set is similar). Values for S-MAD *fused classification* approach.

Alignments	AGIR morph values						AGIR value for bona	Average AGIR value for morphs
	asml	facemorpher	opencv	stylegan	webmorph	real		
a	1.125	1.120	1.123	1.985	2.083	1.209	2.348	1.441
b	0.970	0.861	0.892	2.354	2.077	1.088	1.934	1.374
c	1.403	1.516	1.545	2.687	2.555	1.565	2.671	1.879
d	0.853	0.786	0.798	1.644	1.219	0.954	1.385	1.042
e	1.001	0.920	0.943	1.413	1.394	0.992	1.055	1.111
f	1.184	1.200	1.243	2.064	1.816	1.245	1.836	1.458
g	0.679	0.625	0.617	1.012	1.000	0.729	1.094	0.777
h	1.208	1.127	1.181	1.782	1.772	1.175	1.509	1.374
i	0.640	0.580	0.599	1.267	0.906	0.692	1.331	0.781
j	0.910	0.843	0.906	1.203	1.246	0.962	1.433	1.011
k	0.500	0.516	0.542	0.934	0.700	0.549	0.824	0.624

Based on a direct comparison with the S-MAD *binary classification* approach the background appears to have more influence on the results, especially in the morph case where the *AGIR* values are overall lower than the binary ones. This is also reinforced based on the visual observation of the maps, where the regions of greatest activation are more dispersed when compared to the binary approach and are often in the contour regions.

5.2.3 D-MAD Fused Classification Model

The BPCER@APCER values are presented in the following tables, as well as the respective DET curves in figure 5.8.

Table 5.10: BPCER@APCER = (0.1, 0.01) of Differential Morphing Attack Detection (D-MAD) *fused classification* model across all the alignment settings for each benchmark protocol.

Alignments	BPCER@APCER=δ									
	Protocol-asml		**Protocol-facemorpher**		**Protocol-opencv**		**Protocol-stylegan**		**Protocol-webmorph**	
	δ **=0.1**	δ **=0.01**	δ **=0.1**	δ **=0.01**	δ **=0.1**	δ **=0.01**	δ **=0.1**	δ **=0.01**	δ **=0.1**	δ **=0.01**
a	0.013	0.194	0.013	0.250	0.019	0.206	0.506	0.725	0.244	0.806
b	**0.000**	0.394	0.006	0.356	0.013	0.400	0.563	0.888	0.288	0.713
c	0.063	0.344	0.138	0.644	0.181	0.644	0.656	0.919	0.306	0.831
d	0.100	0.613	0.144	0.613	0.175	0.588	0.625	0.894	0.381	0.769
e	0.138	0.688	0.219	0.744	0.194	0.744	0.544	0.781	0.544	0.831
f	0.013	0.263	**0.000**	**0.206**	**0.006**	**0.206**	**0.188**	**0.525**	0.244	**0.569**
g	0.056	0.363	0.050	0.494	0.069	0.550	0.381	0.706	0.319	0.644
h	0.056	0.475	0.031	0.531	0.056	0.544	0.419	0.713	0.300	0.719
i	0.150	0.531	0.088	0.531	0.069	0.494	0.381	0.738	0.575	0.900
j	0.013	**0.144**	0.013	0.506	0.013	0.438	0.319	0.794	**0.163**	0.738
k	0.044	0.288	0.044	0.369	0.038	0.313	0.500	0.868	0.288	0.856

Table 5.11: Overall performance across all benchmark protocols for D-MAD *fused classification* approach.

BPCER@APCER=δ	a	b	c	d	e	**f**	**g**	**h**	i	j	k
δ=0.1	0.159	0.174	0.269	0.285	0.327	0.09	0.175	0.172	0.252	0.103	0.182
δ=0.01	0.436	0.550	0.676	0.695	0.757	0.350	0.551	0.596	0.639	0.524	0.538

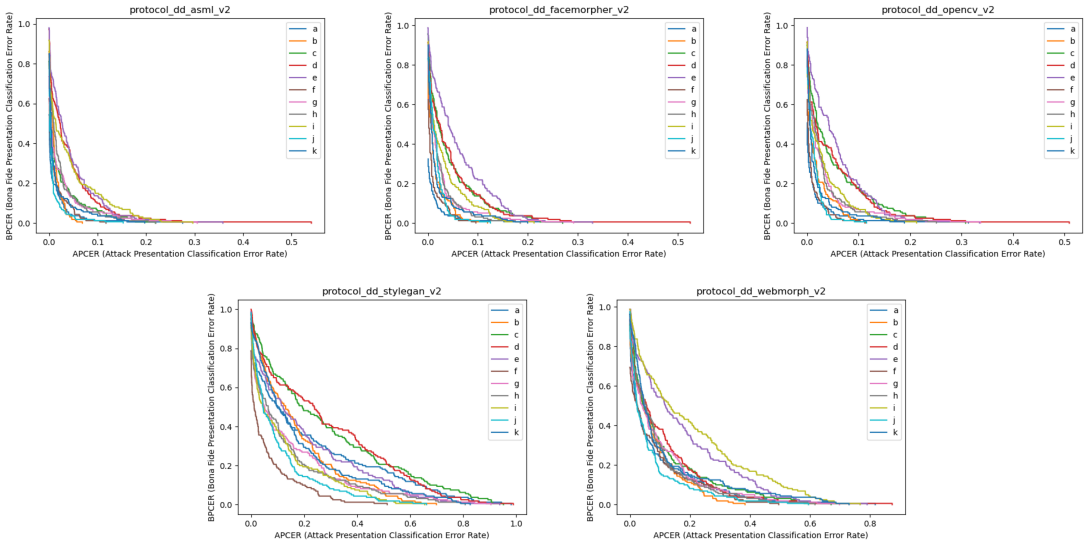

Figure 5.8: DET curves across the different alignment settings (a to k) for D-MAD *fused classification* approach. Each subplot represents one of the benchmark protocols.

Looking at table 5.11 values, it seems that the optimal range of values can be found between the alignment settings of f to h, and with alignment setting f being the potential optimal case.

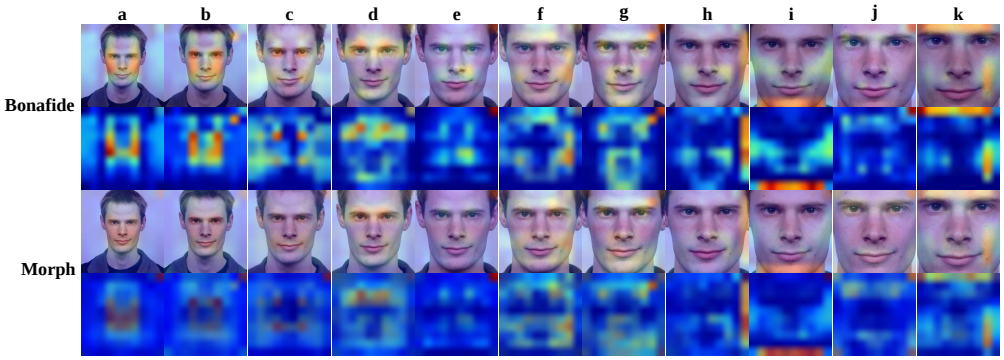

Figure 5.9: Grad-CAM heatmaps for all alignment settings for D-MAD *fused classification* approach.

Table 5.12: Summary table for the values of *AGIR* in the different protocols, as well as the average value for the morphs. Note that for *bonafide* only one column is shown since the values are the same for all protocols (*bonafide* set is similar). Values for D-MAD *fused classification* approach.

Alignments	***AGIR* morph value**					AGIR value for bona	Average AGIR value for morphs
	asml	facemorpher	opencv	stylegan	webmorph		
a	1.604	1.331	1.446	2.064	2.719	1.865	1.833
b	1.061	1.070	1.085	2.321	2.012	2.306	1.509
c	0.975	0.950	0.952	1.127	1.179	1.127	1.037
d	1.192	1.148	1.123	1.828	1.807	1.811	1.420
e	1.275	1.135	1.149	2.183	1.967	2.231	1.542
f	1.018	0.965	1.029	1.504	1.274	1.080	1.158
g	1.056	0.975	1.967	1.808	1.575	1.967	1.476
h	0.789	0.747	0.791	0.863	1.030	0.865	0.844
i	1.249	1.148	1.110	1.405	1.398	1.423	1.262
j	0.733	0.664	0.656	1.315	1.179	1.196	0.909
k	0.842	0.867	0.896	0.699	0.746	0.688	0.810

Making a visual analysis of Grad-CAM heatmaps, it is possible to observe that similar to S-MAD *fused classification*, the detection focuses primarily on the face region and, in many instances, on the intersections of the foreground and background, as presented in figure 5.9. The *AGIR* values presented in table 5.12 seem to prove it.

5.2.4 Discussion on the Results

In an overall analysis of the results, for all the trained models, there is possibly a region or a certain alignment condition where the results are more effective. On the other hand, in that range, there seems to be a correspondence throughout all the models, which translates into a certain area of occupancy of a face in the image. For instance, in the S-MAD *binary classification* approach, the optimal range varies between about 50% and 60%, in S-MAD *fused classification*, the value ranges

from about 42% and 77%, and finally, for D-MAD *fused classification* the optimal condition seems to happen when the face takes up 56% to 70% of the whole image area.

These conclusions are made based on the performance metrics of models. However, as stated in section 4.7, Grad-CAM heatmaps were used to get a more detailed view of which regions of the image are actually responsible for the final prediction. Thus, it is possible to assess the explainability of the training process in the final output.

Looking at the values in table 5.13 the answer seems to be immediate, the face is the predominant activated region regardless of the type of alignment used. Note that these values are taken from previous tables.

Table 5.13: Summary table of *Average of the Gradient Intensity Ratio (foreground/background)* (*AGIR*) values across all the alignment conditions for the different models used. In the case of the morphs, the average value across all the protocols is presented.

Scenario	Approach	Alignments	a	b	c	d	e	f	g	h	i	j	k
S-MAD	Binary Classification	AGIR value for bona	4.48	3.35	2.02	1.81	1.59	1.40	1.77	1.53	1.22	1.24	0.95
		Average AGIR value for morphs	1.69	2.02	2.25	2.39	3.32	2.92	2.41	2.52	2.69	2.56	1.79
	Fused Classification	AGIR value for bona	2.35	1.93	2.67	1.39	1.06	1.84	1.09	1.51	1.33	1.43	0.82
		Average AGIR value for morphs	1.44	1.37	1.88	1.04	1.11	1.46	0.78	1.37	0.78	1.01	0.62
D-MAD	Fused Classification	AGIR value for bona	1.87	2.31	1.13	1.81	2.23	1.08	1.97	0.87	1.43	1.20	0.69
		Average AGIR value for morphs	1.83	1.51	1.04	1.42	1.54	1.16	1.48	0.84	1.26	0.91	0.81

However, it is possible to observe that in both *fused classification* cases (S-MAD and D-MAD), the background seems to have more influence on detection when compared to the S-MAD *binary classification* approach. This is concluded from the fact that the values of *AGIR* values are in general lower, especially for morph cases (recall that the set of *bonafide* images is the same across all the protocols). These two *fused classification* approaches achieve the best performances, especially if we look at the optimal alignment condition in each scenario, which may indicate that the background of the image does influence the results to some extent.

5.3 FRVT MORPH Test Results

The Face Recognition Vendor Test (FRVT) is an ongoing series of evaluations conducted by the National Institute of Standards and Technology (NIST) to assess the performance of the Face Recognition (FR) algorithms. In the specific case of

FRVT MORPH, the purpose is to assess the effectiveness of those algorithms in detecting morphed images, providing insights into their potential and limitations in various contexts [90].

In this regard, the top-performing models obtained for *fused classification* approaches (S-MAD and D-MAD) were submitted. The performances were subsequently evaluated by comparison with state-of-the-art Morphing Attack Detection (MAD) approaches.

This test can be broadly categorized into single-image and differential cases, and it encompasses multiple datasets created using a diversity of methodologies. Table 5.14 provides a summary of some of these datasets.

Table 5.14: Summary table of some of the different datasets used in the FRVT NIST MORPH benchmark.

Dataset	Morphs	Bonafides	Tier
UNIBO Automatic Morphed Face Generation Tool v1.0	2464	1047389	2 - Automated Morphs
Visa-Border	25727		
Twente	2464		
MIPGAN-II	2464		
Manual	323		3 - High Quality Morphs
Print + Scanned	3604	2739	

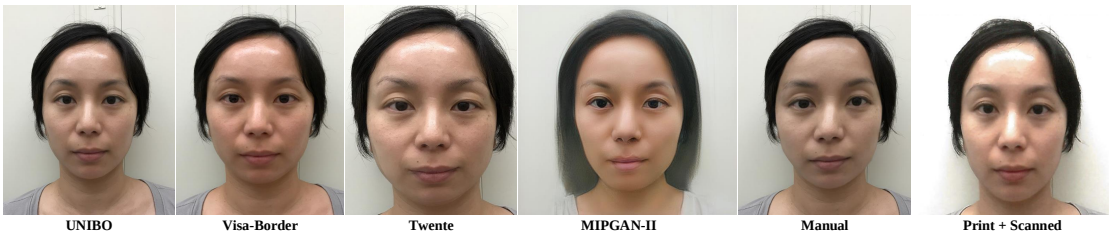

Figure 5.10: Morph Images samples for each dataset presented in table 5.14.

In order to evaluate the performance of morph detection, two commonly used metrics, namely the *morph miss rate* (also known as APCER) and the *false detection rate* (also known as BPCER), are computed, and the results will be presented in the following subsections along with the corresponding DET curves. All these results were taken from the updated FRVT MORPH report, published on June 20, 2023.

It should be noted that in all the following results, the name **visteamicao-000 (single-image)** corresponds to the top-performing model discussed in subsection 5.2.2 and the name **visteamicao-000 (differential)** is related to the top-performing model presented in subsection 5.2.3.

Regarding the remaining SOTA approaches, the results of **wvusingle** [131] by West Virginia University, **unibo** by University of Bologna, **wvudiff** [132] by West

Virginia University, Hochschule Darmstadt with **hdaarcface** [133], **hdaprnu** [134, 135] and **hdabsif**, **ntnussl** [136] by Norwegian University of Science and Technology and finally **visteam** [89] by Universidade de Coimbra will be presented.

5.3.1 Tier 2 - Automated Morphs Analysis

Returning to table 5.14, the results for four distinct datasets will be presented. These datasets consist of morphs created using automated morphing methods, ensuring adherence to academic research standards.

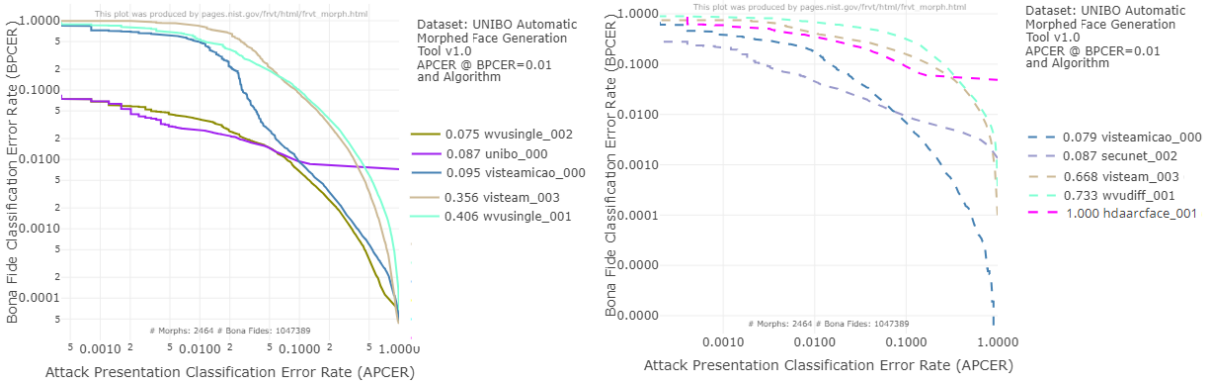

Figure 5.11: DET plot for **UNIBO Automatic Morphed Face Generation Tool v1.0 Dataset**. This chart plots BPCER as a function of APCER. The left side displays the results for the single-image approach, while the right side displays the results for the differential approach.

Table 5.15: APCER@BPCER =(0.1,0.01) for **UNIBO Automatic Morphed Face Generation Tool v1.0 Dataset** across different algorithms.

APCER@BPCER=δ							
Single-image	**Algorithm**	δ=0.1	δ=0.01	**Differential**	**Algorithm**	δ=0.1	δ=0.01
	wvusingle-002	0.000	0.075		**visteamicao-000**	0.014	0.079
	unibo-000	0.000	0.087		secunet-002	0.003	0.087
	visteamicao-000	0.027	0.095		visteam-003	0.171	0.668
	visteam-003	0.091	0.356		wvudiff-001	0.257	0.733
	wvusingle-001	0.101	0.406		hdaarcface-001	0.089	1.000

Related to the UNIBO Automatic Morphed Face Generation Tool v1.0 dataset, the results indicate that our model outperforms all others in the differential case. In the case of single-image presentations, the results were also noteworthy. Although our model did not achieve the lowest *morph miss rates*, it still performed competitively compared to the other algorithms evaluated. The results are presented in table 5.15, and the DET curves depicted in figure 5.11 also reinforce those conclusions, especially for the differential case (blue dashed line).

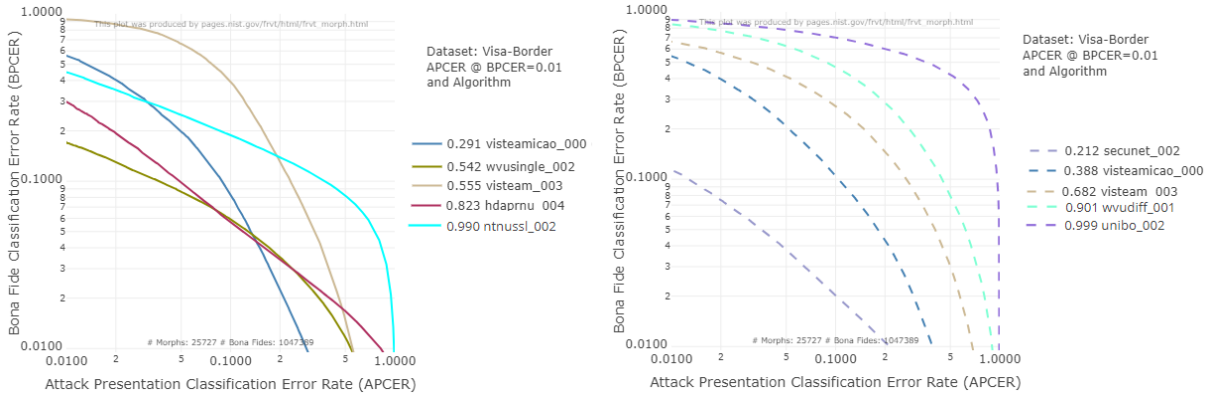

Figure 5.12: DET plot for **Visa-Border Dataset**. This chart plots BPCER as a function of APCER. The left side displays the results for the single-image approach, while the right side displays the results for the differential approach.

Table 5.16: APCER@BPCER =(0.1,0.01) for **Visa-Border Dataset** across different algorithms.

APCER@BPCER=δ							
Single-image	**Algorithm**	δ=0.1	δ=0.01	**Differential**	**Algorithm**	δ =0.1	δ =0.01
	visteamicao-000	0.089	0.291		**secunet-002**	0.013	0.212
	wvusingle-002	0.037	0.542		visteamicao-000	0.105	0.388
	visteam-003	0.232	0.555		visteam-003	0.271	0.682
	hdaprnu-004	0.049	0.823		wvudiff-001	0.447	0.901
	ntnussl-002	0.375	0.990		unibo-002	0.966	0.999

On the Visa-Border dataset, the results from table 5.16 shows that our model outperforms all other SOTA approaches in the single-image case, showing a *morph miss rate* of 29% at a *false detection* rate of 0.01. In the case of a differential approach, the results prove to be competitive as well. Overall, this is a very interesting result, since the single-image case is known to be more challenging than the differential case.

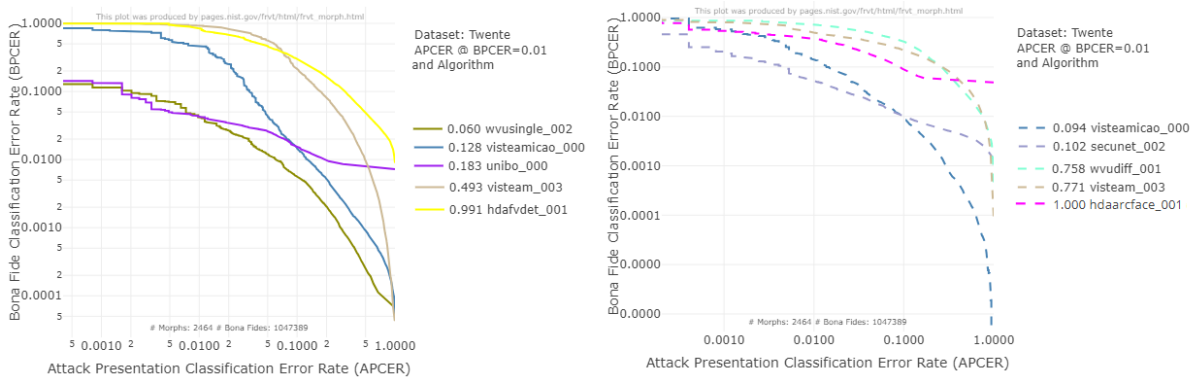

Figure 5.13: DET plot for **Twente Dataset**. This chart plots BPCER as a function of APCER. The left side displays the results for the single-image approach, while the right side displays the results for the differential approach.

Table 5.17: APCER@BPCER =(0.1,0.01) for **Twente Dataset** across different algorithms.

APCER@BPCER=δ							
Single-image	**Algorithm**	δ=0.1	δ=0.01	**Differential**	**Algorithm**	δ=0.1	δ=0.01
	wvusingle-002	0.002	0.060		**visteamicao-000**	0.014	0.094
	visteamicao-000	0.032	0.128		secunet-002	0.005	0.102
	unibo-000	0.002	0.183		wvudiff-001	0.262	0.758
	visteam-003	0.174	0.493		visteam-003	0.269	0.771
	hdafvdet-001	0.308	0.991		hdaaecface-001	0.090	1.000

Regarding the Twente dataset, when comparing with other approaches, the results also turn out to be very positive, achieving, for instance, a *morph error rate* of 9.4% at a *false detection rate* of 0.01 for the differential case.

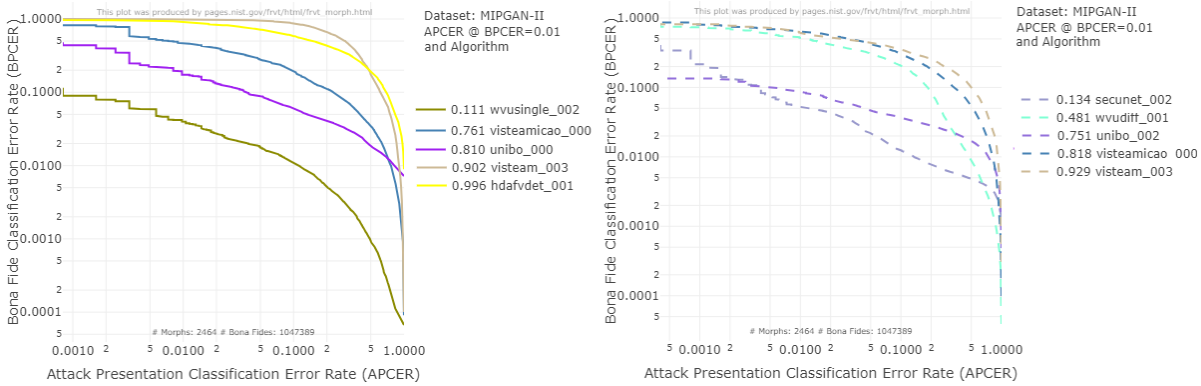

Figure 5.14: DET plots for **MIPGAN-II Dataset**. This chart plots BPCER as a function of APCER. The left side displays the results for the single-image approach, while the right side displays the results for the differential approach.

Table 5.18: APCER@BPCER =(0.1,0.01) for **MIPGAN-II Dataset** across different algorithms

APCER@BPCER=δ							
Single-image	**Algorithm**	δ=0.1	δ=0.01	**Differential**	**Algorithm**	δ=0.1	δ=0.01
	wvusingle-002	0.001	0.111		**secunet-002**	0.004	0.134
	visteamicao-000	0.227	0.761		wvudiff-001	0.182	0.481
	unibo-000	0.037	0.810		unibo-002	0.004	0.751
	visteam-003	0.608	0.902		visteamicao-000	0.332	0.818
	hdafvdet-001	0.695	0.996		visteam-003	0.505	0.929

The results obtained with the MIPGAN-II dataset did not demonstrate the same significant performances as seen in the previous examples. However, except for the *wvusingle-002* algorithm in the single case and *secunet-002* algorithm in the differential case, it turns out that these worst performances happen across all the algorithms, which is possibly influenced by the type of morphed images used in the dataset.

5.3.2 Tier 3 - High Quality Morphs Analysis

This type of analysis holds great significance as these datasets closely resemble real-life situations. The Manual dataset, specifically in the context of single-image approaches, stands out as being the most realistic. On the other hand, the Print+Scanned dataset provides a more comprehensive representation of differential scenarios.

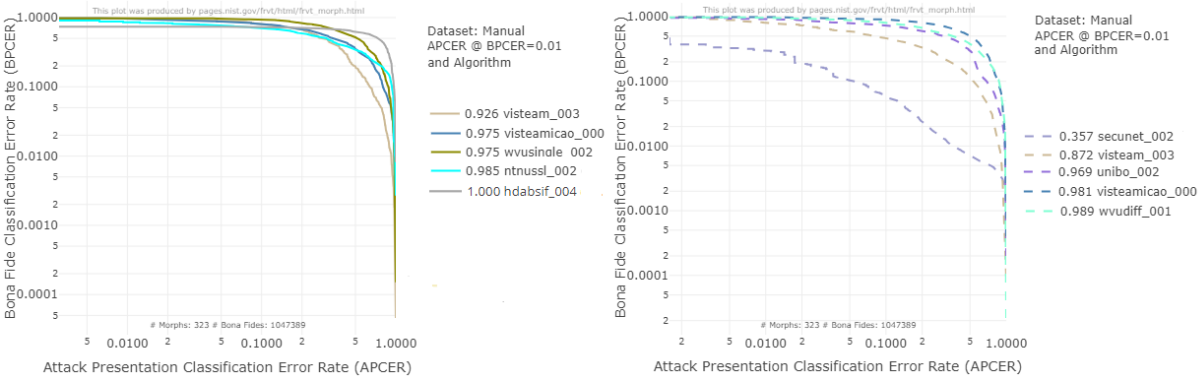

Figure 5.15: DET plot for **Manual Dataset**. This chart plots BPCER as a function of APCER. The left side displays the results for the single-image approach, while the right side displays the results for the differential approach.

Table 5.19: APCER@BPCER =(0.1,0.01) for **Manual Dataset** across different algorithms.

APCER@BPCER=δ							
Single-image	**Algorithm**	δ=0.1	δ=0.01	**Differential**	**Algorithm**	δ=0.1	δ=0.01
	visteam-003	0.641	0.926		**secunet-002**	0.055	0.357
	visteamicao-000	0.802	0.975		visteam-003	0.531	0.872
	wvusingle-002	0.879	0.975		unibo-002	0.689	0.969
	ntnussl-002	0.938	0.985		visteamicao-000	0.853	0.981
	hdabsif-004	0.969	1.000		wvudiff-001	0.873	0.989

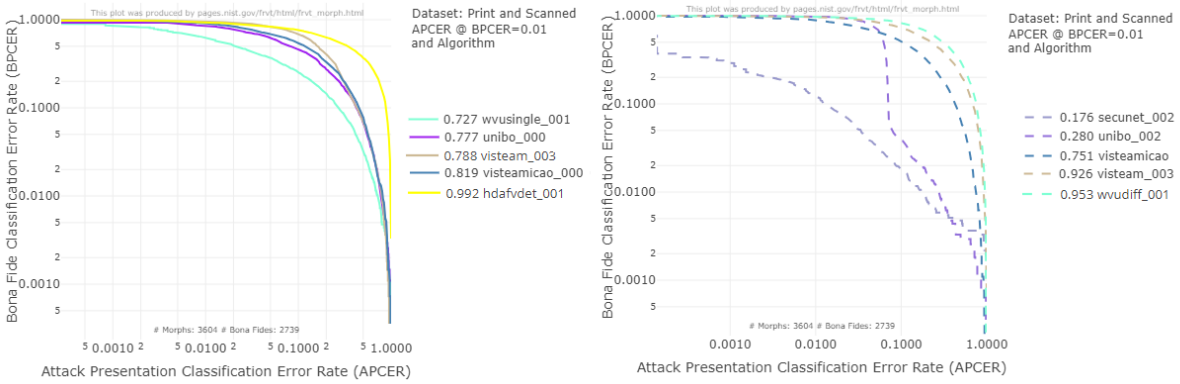

Figure 5.16: DET plot for **Print + Scanned Dataset**. This chart plots BPCER as a function of APCER. The left side displays the results for the single-image approach, while the right side displays the results for the differential approach.

Table 5.20: APCER@BPCER =(0.1,0.01) for **Print + Scanned Dataset** across different algorithms.

APCER@BPCER=δ							
	Algorithm	δ=0.1	δ=0.01		**Algorithm**	δ=0.1	δ=0.01
Single-image	**wvusingle-001**	0.271	0.721	**Differential**	**secunet-002**	0.012	0.176
	unibo-000	0.420	0.777		unibo-002	0.070	0.280
	visteam-003	0.424	0.788		visteamicao-000	0.426	0.751
	visteamicao-000	0.453	0.819		visteam-003	0.680	0.926
	hdafvdet-001	0.879	0.992		wvudiff-001	0.756	0.953

When confronted with these more realistic datasets, it becomes evident that the submitted models are not flawless, in particular when it comes to single-image morph detection. In those cases, the results show that the algorithms do not exhibit robust generalization across various unseen morphing techniques (tables 5.19 and 5.20).

On the other hand, the process of printing and scanning, or re-digitalization, is widely recognized as one of the most significant challenges in morph detection, and the values presented in table 5.20 reinforced that conclusion. Nevertheless, considering that this dataset represents the most realistic for the differential case, the results obtained for our model when compared to other SOTA approaches reached a competitive position.

6

Conclusion

In this dissertation, the main goal was to evaluate the influence of the image context in the detection of face morphing attacks. In this sense, based on the assumption that face alignment can have an influence on detection, we tried to propose an optimal alignment condition.

To accomplish this objective, the initial step involved creating an International Civil Aviation Organization (ICAO) compliant dataset by combining and pre-processing several datasets. As a final result, the ICMD dataset emerged. In addition, to have a dataset specifically for morphed images to train the models, morphs were generated using two main approaches: landmark-based and SyleGAN-based.

The concatenated dataset was then aligned under various alignment conditions. Throughout the different alignment conditions, the face's occupancy area in the image varies, and consequently, so does the context information.

For each one of these alignment settings, different models were trained in two major groups: Single Morphing Attack Detection (S-MAD) case, where the detection is performed based on a single image, and the Differential Morphing Attack Detection (D-MAD) case, where a genuine reference image serves as a comparative basis. Regarding the S-MAD, two approaches were performed: the *fused classification* approach, where the morphing detection task is regularized with a Face Recognition (FR) task and a simple *binary classification* approach (morph/non-morph).

Through extensive experiments, a possible alignment range has been determined at which Morphing Attack Detection (MAD) is most effective. However, the overall impact of image context on face morphing attack detection appears to be limited.

An interesting and encouraging result was obtained on the National Institute of Standards and Technology (NIST) Face Recognition Vendor Test (FRVT) MORPH benchmark, where a performance analysis was conducted comparing different state-of-the-art (SOTA) MAD approaches. The results of the presented models demonstrated high performance in several benchmarks. Reaching the SOTA level in some

of them.

To further investigate the impact of image context on the detection of morphing attacks, a potential future approach could involve removing the background from the image. In this way, the study would focus exclusively on the relevant features of the foreground (facial image), giving rise to different possible insights.

Alternatively, in the context of the D-MAD approach, it would also be interesting to study the case for a strictly *binary classification* approach, *i.e.,* without imposing the identity classification part from the *fused classification*. To do so, a potential approach could involve utilizing a Siamese network and, in a similar way, evaluating the similarity between the two feature embeddings followed by a *sigmoid* activation.

From the results presented, the greatest differentiating factor seems to be centered on the morph images. In that regard, and considering the common practice of printing and re-scanning Identity Document (ID) documents during the issuance process, it would also be important to generate a dataset that includes these morphs in order to train and evaluate the models more realistically, potentially leading to other outcomes.

Finally, it would also be interesting to explore other explainability tools or attention mechanisms in order to provide more accurate and realistic insights about the decision-making process.

Bibliography

[1] J. Fagertun and M. B. Stegmann, "The IMM frontal face database," 2005.

[2] M. Wang and W. Deng, "Deep face recognition: A survey," *Neurocomputing*, vol. 429, pp. 215–244, 2021.

[3] V. Gupta and S. Mallick, "Face recognition -introduction for beginners – nearly everything you need to know." Blog post, April 16 2019.

[4] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert, "Accurate and robust neural networks for face morphing attack detection," *Journal of Information Security and applications*, vol. 53, p. 102526, 2020.

[5] S. Venkatesh, R. Ramachandra, K. Raja, and C. Busch, "Face morphing attack generation & detection: A comprehensive survey," *IEEE Transactions on Technology and Society*, vol. PP, pp. 1–1, 03 2021.

[6] U. Scherhag, C. Rathgeb, and C. Busch, "Face morphing attack detection methods," in *Handbook of Digital Face Manipulation and Detection*, 2022.

[7] N. I. Kajla, M. M. S. Missen, M. M. Luqman, M. Coustaty, A. Mehmood, and G. S. Choi, "Additive angular margin loss in deep graph neural network classifier for learning graph edit distance," *IEEE Access*, vol. 8, pp. 201752–201761, 2020.

[8] S. Rahal, H. Aboalsamah, and K. Muteb, "Multimodal biometric authentication system - mbas," in *2006 2nd International Conference on Information and Communication Technologies*, pp. 1026–1030, 2006.

[9] J. Wayman, A. Jain, D. Maltoni, and D. Maio, "An introduction to biometric authentication systems," in *Biometric systems: Technology, design and performance evaluation*, pp. 1–20, Springer, 2005.

[10] L. Li, X. Mu, S. Li, and H. Peng, "A review of face recognition technology," *IEEE Access*, vol. 8, pp. 139110–139120, 2020.

[11] "ICAO-Uniting aviation. ICAO Doc 9303." https://www.icao.int/publications/pages/publication.aspx?docnum=9303. [Accessed 17-Nov-2022].

[12] M. Rusia and D. Singh, "A comprehensive survey on techniques to handle face identity threats: challenges and opportunities," *Multimedia Tools and Applications*, 06 2022.

[13] Z. Akhtar, D. Dasgupta, and B. Banerjee, "Face authenticity: An overview of face manipulation generation, detection and recognition," *SSRN Electronic Journal*, 2019.

[14] K. Jindal, S. Dalal, and K. K. Sharma, "Analyzing spoofing attacks in wireless networks," in *2014 Fourth International Conference on Advanced Computing and Communication Technologies*, pp. 398–402, Feb 2014.

[15] C. Busch, "Christoph Busch Projects: Morphing Attack Detection — christoph-busch.de." https://www.christoph-busch.de/projects-mad.html. [Accessed 15-Nov-2022].

[16] "Required documents — home-affairs.ec.europa.eu." https://home-affairs.ec.europa.eu/policies/schengen-borders-and-visa/visa-policy/required-documents_en. [Accessed 15-Nov-2022].

[17] M. Torkar, "Morphing cases in slovenia." Ministry of the Interior Police, Slovenia.

[18] VISTeam, "Facing project." https://visteam.isr.uc.pt/projects/facing, 2022. [Accessed 19-Nov-2022].

[19] "Nist benchmark test data." https://www.nist.gov/ambench/benchmark-test-data. [Accessed 19-Nov-2022].

[20] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, p. 399–458, dec 2003.

[21] ISO/IEC, "ISO/IEC 19795-1:2006. Information technology — Biometric performance testing and reporting — Part 1: Principles and framework." ISO/IEC JTC 1/SC 37 Biometrics, April 2006.

[22] A. K. Jain, "Biometric recognition: Overview and recent advances," in *Progress in Pattern Recognition, Image Analysis and Applications* (L. Rueda, D. Mery, and J. Kittler, eds.), (Berlin, Heidelberg), pp. 13–19, Springer Berlin Heidelberg, 2007.

[23] D. E. King, "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research*, vol. 10, pp. 1755–1758, 2009.

[24] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems* (Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Weinberger, eds.), vol. 27, Curran Associates, Inc., 2014.

[25] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4396–4405, 2019.

[26] ISO/IEC JTC 1/SC 37 Biometrics, "Information technology — Biometric presentation attack detection — Part 1: Framework." ISO/IEC Standard 30107-1:2016, 01 2016. Accessed on March 14, 2023.

[27] "Woodrow Bledsoe Originates Of Automated Facial Recognition : History Of Information." `https://www.historyofinformation.com/detail.php?entryid=2495`. [Accessed 15-Nov-2022].

[28] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, vol. 1, pp. I–I, 2001.

[29] C. Papageorgiou and T. Poggio, "Trainable pedestrian detection," in *Proceedings 1999 International Conference on Image Processing (Cat. 99CH36348)*, vol. 4, pp. 35–39 vol.4, 1999.

[30] K. Tieu and P. Viola, "Boosting image retrieval," in *Proceedings IEEE Conference on Computer Vision and Pattern Recognition. CVPR 2000 (Cat. No.PR00662)*, vol. 1, pp. 228–235 vol.1, 2000.

[31] T. Mita, T. Kaneko, and O. Hori, "Joint haar-like features for face detection," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, vol. 2, pp. 1619–1626 Vol. 2, Oct 2005.

[32] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, pp. 1499–1503, Oct 2016.

[33] L. Sirovich and M. Kirby, "Low-dimensional procedure for the characterization

of human faces.," *Journal of the Optical Society of America. A, Optics and image science*, vol. 43, pp. 519–24, 1987.

[34] M. Turk and A. Pentland, "Face recognition using eigenfaces," in *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–591, 1991.

[35] W. Ouarda, H. Trichili, A. M. Alimi, and B. Solaiman, "Face recognition based on geometric features using support vector machines," in *2014 6th International Conference SoCPaR*, pp. 89–95, 2014.

[36] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *European Conference on Computer Vision*, 2004.

[37] T. Ojala, M. Pietikäinen, and T. Mäenpää, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, p. 971–987, jul 2002.

[38] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 886–893 vol. 1, 2005.

[39] O. Déniz, G. Bueno, J. Salido, and F. De la Torre, "Face recognition using histograms of oriented gradients," *Pattern Recognition Letters*, vol. 32, no. 12, pp. 1598–1603, 2011.

[40] C. Shu, X. Ding, and C. Fang, "Histogram of the oriented gradient for face recognition," *Tsinghua Science and Technology*, vol. 16, no. 2, pp. 216–224, 2011.

[41] B. Zhang, S. Shan, X. Chen, and W. Gao, "Histogram of gabor phase patterns (hgpp): A novel object representation approach for face recognition," *IEEE Transactions on Image Processing*, vol. 16, pp. 57–68, Jan 2007.

[42] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems* (F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds.), vol. 25, Curran Associates, Inc., 2012.

[43] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings* (Y. Bengio and Y. LeCun, eds.), 2015.

[44] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1–9, June 2015.

[45] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, June 2016.

[46] A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, and H. Adam, "Mobilenets: Efficient convolutional neural networks for mobile vision applications," *arXiv preprint arXiv:1704.04861*, 2017.

[47] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," pp. 4510–4520, 06 2018.

[48] A. Howard, M. Sandler, G. Chu, L.-C. Chen, B. Chen, M. Tan, W. Wang, Y. Zhu, R. Pang, V. Vasudevan, *et al.*, "Searching for mobilenetv3," in *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1314–1324, 2019.

[49] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Deepface: Closing the gap to human-level performance in face verification," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1701–1708, June 2014.

[50] G. Huang, Z. Liu, and K. Q. Weinberger, "Densely connected convolutional networks," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261–2269, 2016.

[51] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in *Proceedings of the 36th International Conference on Machine Learning* (K. Chaudhuri and R. Salakhutdinov, eds.), vol. 97 of *Proceedings of Machine Learning Research*, pp. 6105–6114, PMLR, 09–15 Jun 2019.

[52] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1891–1898, June 2014.

[53] W. Liu, Y. Wen, Z. Yu, M. Li, B. Raj, and L. Song, "Sphereface: Deep hypersphere embedding for face recognition," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6738–6746, July 2017.

[54] H. Wang, Y. Wang, Z. Zhou, X. Ji, Z. Li, D. Gong, J. Zhou, and W. Liu, "Cosface: Large margin cosine loss for deep face recognition," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5265–5274, 2018.

[55] J. Deng, J. Guo, N. Xue, and S. Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4685–4694, 2019.

[56] Q. Meng, S. Zhao, Z. Huang, and F. Zhou, "MagFace: A universal representation for face recognition and quality assessment," in *CVPR*, 2021.

[57] M. Kim, A. K. Jain, and X. Liu, "Adaface: Quality adaptive margin for face recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[58] 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)*, pp. 815–823, June 2015.

[59] Y. Shi and A. K. Jain, "Docface: Matching id document photos to selfies," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–8, Oct 2018.

[60] "Georgia Tech Face Database." `http://www.anefian.com/face_reco.htm`. [Accessed November 29, 2022].

[61] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *IEEE International Joint Conference on Biometrics*, pp. 1–7, Sep. 2014.

[62] "Start Morphing - MorphThing.com — morphthing.com." `https://www.morphthing.com/morph`. [Accessed 15-Nov-2022].

[63] "FaceMorpher - Home — facemorpher.com." `http://www.facemorpher.com`. [Accessed 15-Nov-2022].

[64] "Magic Morph—Morphing Software — effectmatrix.com." `https://www.effectmatrix.com/morphing/`. [Accessed 15-Nov-2022].

[65] M. Satya, "Face morph using opencv." Website, 2016. `www.learnopencv.com/face-morph-using-opencv-cpp-python/`.

[66] A. Quek, "Facemorpher." `https://github.com/yaopang/FaceMorpher`, 2019.

[67] U. Scherhag, C. Rathgeb, and C. Busch, "Towards detection of morphed face images in electronic travel documents," in *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*, pp. 187–192, April 2018.

[68] B. Delaunay, "Sur la sphère vide," *Bulletin de l'Academie des Sciences de l'URSS. Classe des sciences mathematiques et na*, vol. 1934, no. 6, pp. 793–800, 1934.

[69] C. Seibold, A. Hilsmann, and P. Eisert, "Style your face morph and improve your face morphing attack detector," in *2019 International Conference of the Biometrics Special Interest Group (BIOSIG)*, pp. 1–6, 2019.

[70] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. C. Courville, and Y. Bengio, "Generative adversarial nets," in *NIPS*, 2014.

[71] N. Damer, A. M. Saladié, A. Braun, and A. Kuijper, "Morgan: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network," in *2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–10, Oct 2018.

[72] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, and A. Courville, "Adversarially learned inference," in *International Conference on Learning Representations*, 2017. Published as a conference paper at ICLR 2017.

[73] N. Damer, F. Boutros, A. M. Saladié, F. Kirchbuchner, and A. Kuijper, "Realistic dreams: Cascaded enhancement of gan-generated images with an example in face morphing attacks," in *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–10, Sep. 2019.

[74] S. K. Venkatesh, H. Zhang, R. Ramachandra, K. B. Raja, N. Damer, and C. Busch, "Can gan generated morphs threaten face recognition systems equally as landmark based morphs? - vulnerability and detection," *2020 8th International Workshop on Biometrics and Forensics (IWBF)*, pp. 1–6, 2020.

[75] R. Abdal, Y. Qin, and P. Wonka, "Image2stylegan: How to embed images into the stylegan latent space?," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4431–4440, 2019.

[76] H. Zhang, S. Venkatesh, R. Ramachandra, K. B. Raja, N. Damer, and C. Busch, "MIPGAN - generating robust and high quality morph attacks using identity prior driven GAN," *CoRR*, vol. abs/2009.01729, 2020.

[77] N. Damer, M. Fang, P. Siebke, J. N. Kolf, M. Huber, and F. Boutros, "Mordiff: Recognition vulnerability and attack detectability of face morphing attacks created by diffusion autoencoders," *2023 11th International Workshop on Biometrics and Forensics (IWBF)*, pp. 1–6, 2023.

[78] "Homepage — iMARS — imars-project.eu." `https://imars-project.eu/`, 2015. [Accessed 15-Nov-2022].

[79] "State of the art of Morphing Detection (SOTAMD) - NTNU — ntnu.edu." `https://www.ntnu.edu/iik/sotamd`. [Accessed 15-Nov-2022].

[80] R. Raghavendra, K. B. Raja, and C. Busch, "Detecting morphed face images," in *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*, pp. 1–7, Sep. 2016.

[81] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[82] V. Ojansivu and J. Heikkilä, "Blur insensitive texture classification using local phase quantization," ICISP '08, (Berlin, Heidelberg), p. 236–243, Springer-Verlag, 2008.

[83] U. Scherhag, L. Debiasi, C. Rathgeb, C. Busch, and A. Uhl, "Detection of face morphing attacks based on prnu analysis," *IEEE Transactions on Biometrics, Behavior, and Identity Science*, vol. 1, no. 4, pp. 302–317, 2019.

[84] L.-B. Zhang, F. Peng, and M. Long, "Face morphing detection using fourier spectrum of sensor pattern noise," in *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2018.

[85] S. Venkatesh, R. Ramachandra, K. Raja, L. Spreeuwers, R. Veldhuis, and C. Busch, "Morphed face detection based on deep color residual noise," in *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*, pp. 1–6, Nov 2019.

[86] R. Raghavendra, K. B. Raja, S. K. Venkatesh, and C. Busch, "Transferable deep-cnn features for detecting digital and print-scanned morphed face images," *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1822–1830, 2017.

[87] C. Seibold, W. Samek, A. Hilsmann, and P. Eisert, "Detection of face morph-

ing attacks by deep learning," in *International Workshop on Digital Watermarking*, 2017.

[88] P. C. Neto, T. Gonçalves, M. Huber, N. Damer, A. F. Sequeira, and J. S. Cardoso, "Orthomad: Morphing attack detection through orthogonal identity disentanglement," in *2022 International Conference of the BIOSIG*, pp. 1–5, 2022.

[89] I. Medvedev, F. Shadmand, and N. Gonçalves, "Mordeephy: Face morphing detection via fused classification," in *Proceedings of the 12th International Conference on Pattern Recognition Applications and Methods - Volume 1: ICPRAM,*, pp. 193–204, INSTICC, SciTePress, 2023.

[90] "Frvt morph." `https://pages.nist.gov/frvt/html/frvt_morph.html`. [Accessed 15-Nov-2022].

[91] U. Scherhag, D. Budhrani, M. Gomez-Barrero, and C. Busch, "Detecting morphed face images using facial landmarks," in *Image and Signal Processing* (A. Mansouri, A. El Moataz, F. Nouboud, and D. Mammass, eds.), (Cham), pp. 444–452, Springer International Publishing, 2018.

[92] N. Damer, V. Boller, Y. Wainakh, F. Boutros, P. Terhörst, A. Braun, and A. Kuijper, "Detecting face morphing attacks by analyzing the directed distances of facial landmarks shifts," in *German Conference on Pattern Recognition*, 2018.

[93] U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch, "Deep face representations for differential morphing attack detection," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3625–3639, 2020.

[94] S. Soleymani, B. Chaudhary, A. Dabouei, J. M. Dawson, and N. M. Nasrabadi, "Differential morphed face detection using deep siamese networks," *ArXiv*, vol. abs/2012.01541, 2020.

[95] M. Ferrara, A. Franco, and D. Maltoni, "Face demorphing," *IEEE Transactions on Information Forensics and Security*, vol. 13, pp. 1008–1017, April 2018.

[96] M. Ferrara, A. Franco, and D. Maltoni, "Face demorphing in the presence of facial appearance variations," in *2018 26th European Signal Processing Conference (EUSIPCO)*, pp. 2365–2369, Sep. 2018.

[97] F. Peng, L.-B. Zhang, and L. Min, "Fd-gan: Face de-morphing generative

adversarial network for restoring accomplice's facial image," *IEEE Access*, vol. PP, pp. 1–1, 06 2019.

[98] D. Ortega-Delcampo, C. Conde, D. Palacios-Alonso, and E. Cabello, "Border control morphing attack detection with a convolutional neural network de-morphing approach," *IEEE Access*, vol. 8, pp. 92301–92313, 2020.

[99] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Casia webface database: a large-scale face database for face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 490–497, 2014.

[100] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Technical Report 07-49, University of Massachusetts, Amherst, October 2007.

[101] B. Ni, C. Yan, Y. Li, H. Yang, and X. Yang, "The coupled cascaded mtl-cnn for face detection and alignment and its application to face recognition," *IEEE Transactions on Multimedia*, vol. 19, no. 8, pp. 1636–1651, 2017.

[102] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," in *Proceedings of International Conference on Computer Vision (ICCV)*, 2015.

[103] Q. Ma, J. Li, S. Yang, and X. Yang, "Vggface2: A dataset for recognising faces across pose and age," in *International Conference on Automatic Face and Gesture Recognition (FG)*, 2018.

[104] Y. Guo, L. Zhang, Y. Hu, X. He, and J. Gao, "Ms-celeb-1m: A dataset and benchmark for large-scale face recognition," in *Proceedings of European Conference on Computer Vision (ECCV)*, 2016.

[105] F. Wang, L. Chen, C. Li, S. Huang, Y. Chen, C. Qian, and C. C. Loy, "The devil of face recognition is in the noise," in *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 765–780, 2018.

[106] P. J. Phillips, P. J. Flynn, W. T. Scruggs, K. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. J. Worek, "Overview of the face recognition grand challenge," *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1, pp. 947–954 vol. 1, 2005.

[107] K. Messer, J. Matas, J. Kittler, K. Jonsson, J. Luettin, and G. Maître,

"Xm2vtsdb: The extended m2vts database," *Proc. of Audio- and Video-Based Person Authentication*, 04 2000.

[108] P. Phillips, P. Flynn, K. Bowyer, R. Vorder, P. Grother, G. Quinn, and M. Pruitt, "Distinguishing identical twins by face recognition," The Ninth IEEE International Conference on Automatic Face and Gesture Recognition (FG 2011), Santa Barbara, CA, 2011-03-21 2011.

[109] P. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The feret evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[110] P. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The feret database and evaluation procedure for face-recognition algorithms," *Image and Vision Computing*, vol. 16, no. 5, pp. 295–306, 1998.

[111] A. Martinez and R. Benavente, *The AR Face Database: CVC Technical Report, 24*. Jan. 1998.

[112] P. I. C. at Stirling (PICS), "Psychological Image Collection at Stirling (PICS)." `http://pics.stir.ac.uk/`. [Accessed November 29, 2022].

[113] "FEI- Face Database." `https://fei.edu.br/~cet/facedatabase.html`. [Accessed November 29, 2022].

[114] J. Avilés, H. Toapanta, P. Morillo, and D. Vallejo-Huanga, "Dataset of ethnic facial images of ecuadorian people," 2019.

[115] B. Heisele, T. Serre, M. Pontil, and T. Vetter, "Mit-cbcl face recognition database." `http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html`, 2001. Accessed on March 17, 2023.

[116] L. DeBruine and B. Jones, "Face research lab london set," 2017.

[117] I. Medvedev, F. Shadmand, and N. Gonçalves, "Young labeled faces in the wild (ylfw): A dataset for children faces recognition," *arXiv:2301.05776*, 2023.

[118] IDIAP Research Institute, "FRGC-Morphs dataset." `https://www.idiap.ch/en/dataset/frgc-morphs`, Year Accessed.

[119] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila, "Analyzing and improving the image quality of stylegan," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8110–8119, 2020.

[120] IDIAP Research Institute, "FRLL-Morphs dataset." https://www.idiap.ch/en/dataset/frll-morphs, Year Accessed.

[121] L. DeBruine, "webmorph," Jan. 2018.

[122] IDIAP Research Institute, "FERET-Morphs dataset." https://www.idiap.ch/en/dataset/feret-morphs, Year Accessed.

[123] T. Dunstone, "New face morphing database for vulnerability research." https://www.linkedin.com/pulse/new-face-morphing-dataset-vulnerability-research-ted-dunstone/, 2017. Accessed on March 17, 2023.

[124] G. Borghi, A. Franco, G. Graffieti, and D. Maltoni, "Automated artifact retouching in morphed images with attention maps," *IEEE Access*, vol. 9, pp. 136561–136579, 2021.

[125] J. Xiang and G. Zhu, "Joint face detection and facial expression recognition with mtcnn," in *2017 4th ICISCE*, pp. 424–427, 2017.

[126] B. at the University of Bologna, "FVC-onGoing benchmarks." https://biolab.csr.unibo.it/FvcOnGoing/UI/Form/Benchmarks.aspx, 2020. [Online; accessed March 16, 2023].

[127] L. DeBruine and B. Jones, "Face research lab london set." figshare, May 2017.

[128] T. Neubert, A. Makrushin, M. Hildebrandt, C. Krätzer, and J. Dittmann, "Extended stirtrace benchmarking of biometric and forensic qualities of morphed face images," *IET Biom.*, vol. 7, pp. 325–332, 2018.

[129] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization," *CoRR*, vol. abs/1610.02391, 2016.

[130] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision(IJCV)*, vol. 115, no. 3, pp. 211–252, 2015.

[131] P. Aghdaie, B. Chaudhary, S. Soleymani, J. M. Dawson, and N. M. Nasrabadi, "Attention aware wavelet-based detection of morphed face images," *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–8, 2021.

[132] H. Kashiani, S. M. Sami, S. Soleymani, and N. M. Nasrabadi, "Robust ensem-

ble morph detection with domain generalization," *2022 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–10, 2022.

[133] U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch, "Deep face representations for differential morphing attack detection," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 3625–3639, 2020.

[134] L. Debiasi, C. Rathgeb, U. Scherhag, A. Uhl, and C. Busch, "Prnu-based detection of morphed face images," in *Proceedings of 6th International Workshop on Biometrics and Forensics (IWBF 2018)*, 2018.

[135] L. Debiasi, C. Rathgeb, U. Scherhag, A. Uhl, and C. Busch, "Prnu variance analysis for morphed face image detection," in *Proceedings of 9th International Conference on Biometrics: Theory, Applications and Systems (BTAS 2018)*, 2018.

[136] R. Ramachandra, S. Venkatesh, K. Raja, and C. Busch, "Towards making morphing attack detection robust using hybrid scale-space colour texture features," in *2019 IEEE 5th International Conference on Identity, Security, and Behavior Analysis (ISBA)*, pp. 1–8, 2019.

A

Appendix

N. Damer, M. Gomez-Barrero, K. Raja, C. Rathgeb, A. Sequeira,
M. Todisco, and A. Uhl (Eds.): BIOSIG 2023,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2023

Impact of Image Context for Single Deep Learning Face Morphing Attack Detection

Abstract: The increase in security concerns due to technological advancements has led to the popularity of biometric approaches that utilize physiological or behavioral characteristics for enhanced recognition. Face recognition systems (FRSs) have become prevalent, but they are still vulnerable to image manipulation techniques such as face morphing attacks. This study investigates the impact of the alignment settings of input images on deep learning face morphing detection performance. We analyze the interconnections between the face contour and image context and suggest optimal alignment conditions for face morphing detection.

Keywords: Face morphing detection; face recognition, deep learning; convolutional neural networks; classification.

1 Introduction

The expansion of technological advancements in modern society has led to an increase in security concerns. Traditional identification methods have become less reliable due to their vulnerability to forgetfulness, loss, replication, or theft, thereby compromising their intended security function. As a solution to this issue, biometric approaches are gaining popularity as they utilize physiological or behavioral characteristics to enhance the recognition process.

Face image modality took one of the most important roles in modern biometric applications due to the simplicity of face image acquisition and recent advances in computer vision techniques. This led to the widespread use of Face Recognition Systems (FRSs) which utilize facial traits for the purpose of identification or verification [Li20]. Despite the fact that FRSs are currently used in various applications, they are still highly vulnerable to attacks due to the extensive range of image manipulation techniques that can be used to deceive the system.

One of the most important types of threats to FRSs is the face morphing attack. In this attack, facial features from two or more images are merged to create a synthetic image that incorporates features from both faces. The resulting image is similar to the images that gave rise to it, which allows one person to impersonate another, thereby violating the principle of self-ownership. That is why face morphing detection is a critical task in the era of digital manipulation and deep learning techniques. However, the performance of face morphing detection may depend on various factors, such as the alignment and preprocessing of input images. Specifically, the face image alignment setting can impact the amount of context included in the input image, which in turn can hypothetically affect the performance of the detection algorithm. We conduct our research to define optimal alignment settings for face morphing detection, exploring the possibility of using interconnections

between the face contour and image context to improve the performance of the detection algorithm.

Essentially, our purpose is to investigate the relationship between image context and MAD, with the aim of identifying the most effective context properties for detection. Throughout this paper, the term ”image context” refers to the background and surrounding elements in the image, i.e., the part of the image that does not contain the face.

As an additional contribution, we combined a dataset that adheres to the International Civil Aviation Organization (ICAO) guidelines for detecting face morphing.

2 Related Work

Face Recognition. Current advances in face recognition methods use deep learning techniques that employ deep neural networks, allowing the learning of deep facial features, which have high discriminative power.

Face recognition deep networks are commonly trained using classification-based tasks, employing softmax loss or its margin-based alternatives like ArcFace [De19]. The addition of a margin to the softmax loss is crucial because it significantly improves the discriminative power of the learned features. More recently, there has been a focus on incorporating adaptiveness into the margin based on the quality of the input image. For instance, MagFace [Me21] optimizes the feature embedding using an adaptive margin and regularization based on its magnitude. Another approach is AdaFace [KJL22], which proposes adapting the margin function based on the norm of the feature embedding.

Face Morphing Generation. Face morphing can be performed using landmark-based or deep learning-based approaches. Landmark-based methods employ a set of fiducial facial points, which are detected on all contributing face images, to generate a morph image by warping and bending procedures [FFM14].

Deep learning-based methods may employ encoder-decoder architectures, such as Generative Adversarial Networks (GANs) [Go14]. For example, the MorGAN [Da18] approach aims to make the generated images look similar to the real images while also encouraging the generators to produce diverse images that differ from each other. Karras et al. [KLA19] proposed the StyleGAN approach, which can be used to generate high-quality morphs. By projecting original images into the latent domain and interpolating latent embeddings, StyleGAN enables face morphing without the blending artifacts commonly observed when morphing is performed in the image domain.

The MIPGAN [H.21] approach revisits the StyleGAN by introducing an end-to-end optimization approach with a novel loss function that emphasizes preserving the identity of the generated morphed face images by incorporating identity priors. MorDIFF [Da23] proposes the use of diffusion autoencoders to generate high-fidelity and smooth face morphing attacks, which are highly vulnerable to state-of-the-art face recognition models. ReGenMorph [Da21] approach proposes to eliminate blending artifacts by combining image-level morphing and GAN-based generation, resulting in visibly realistic morphed images with high appearance quality.

Face Morphing Detection. Morphing attack detection (MAD) methods can be classified into two types, depending on the security application scenario: Single Morphing Attack Detection (S-MAD) and Differential Morphing Attack Detection (D-MAD).

S-MAD refers to techniques that can detect a morphed image without comparing it to an authentic reference image (*non-reference*). They are therefore based on the analysis of visual artifacts or inconsistencies in the morphed image itself. Many approaches rely on the analysis of handcrafted features like Binarized Statistical image features (BSIF) [RRB16], Local Binary Pattern (LBP) [OPH96], Local Phase Quantization (LPQ) [OH08] image descriptors, and Photo Response Non-Uniformity (PRNU) known as sensor noise [Sc19].

Recent works intensively uses deep learning for face morphing detection. OrthoMAD approach [Ne] proposes to use a regularization term for the creation of two orthogonal latent vectors that disentangle identity information from morphing attacks. MorDeephy method [MSG23] introduced fused classification to generalize morphing detection to unseen attacks. The formulation will be followed in this work. Tapia et al. [TB21] proposed a framework using few-shot learning with siamese networks and domain generalization. The framework includes a triplet-semi-hard loss function and clustering to assign classes to image samples.

In this work, we focus only on the S-MAD case to perform the analysis of image alignment settings.

3 Methodology

Source Data Curating. An initial challenge encountered in this research was the lack of a suitably extensive dataset that conformed to ICAO compliance requirements. To address this issue, we combined multiple datasets comprising compliant images, including both publicly available and privately obtained data. When selecting the datasets, we prioritized those that provided a larger number of images per identity and included the following ones: FRGC [Ph05], XM2VTS [Me00], ND Twins [Ph11], FERET [Ph00, Ph98], AR [MB98], PICS [Un99], FEI [Th06], IMMF [FS05] and GTDB [A.99].

Several selected components were filtered to remove non-compliant images, i.e., non-frontal images or other images not suitable for morphing. In the specific case of the ND twins dataset, only one twin from the pair was included due to their striking resemblance, which will be confusing for the methodology of this research. Our result dataset, which we call the ICMD dataset, comprises over 50k images of more than 2.5k individuals.

Morph Image Generation. To accompany our training data with face morph samples, we employed landmark-based and deep learning-based (specifically GAN-based) face morphing approaches. These samples are generated using the originals from the ICMD dataset, where pairing is performed following the [MSG23]. Namely, the identity list of the dataset is randomly split into two subsets, and the pairing is only made between identities from those subsets.

To generalize the detection performance and reduce overfitting for artifact detection, we have included *selfmorphs* for both LDM and StyleGAN approaches. *Selfmorphs* are gen-

erated using images of the same individual, resulting in morphed images that continue to represent that same individual but contain merging artifacts of a different kind. As a result, we can prioritize morphing detection based on the behavior of deep facial features.

Alignment settings Our search for the optimal amount of image context for morphing detection is based on selecting several different alignment settings and running identical experiments for each setting. The face alignment in academia is usually performed by a rigid transformation, which minimizes the coordinate distance between the five facial landmarks (detected by MTCNN [XZ17]) ({left eye}, {right eye}, {nose}, {left mouth corner}, {right mouth corner}) and the definite target list of coordinates (for the resulting image size of 112×112 - {{38.2, 41.7}, {73.5, 41.5}, {56.0, 61.7}, {41.5, 82.4}, {70.7, 82.2}}) [De19]. The particular list of settings that we used is based on the scaling of this target set of coordinates. The Table 1 presents a schematic correspondence of each alignment with the scale factor utilized, along with its respective indicative ratio of the face's occupancy area in the image. We estimate this face's occupancy as the ratio of face area (limited by a face contour detected using 68 landmarks [Ki09]) to the full image area.

Tab. 1: Summary table of all alignment conditions with their respective scale factors and ratios.

Alignments	**a**	**b**	**c**	**d**	**e**	**f**	**g**	**h**	**i**	**j**	**k**
Scale Factor	1.65	1.40	1.10	1.00	0.90	0.85	0.80	0.75	0.70	0.65	0.60
Ratio	0.15	0.21	0.34	0.42	0.51	0.56	0.62	0.70	0.77	0.86	0.94

S-MAD - Fused Classification. In our work, we approach *no-reference* face morphing detection in several ways. First, we follow the fused classification approach, where the morphing detection task is regularized with face recognition to generalize the performance to unseen attacks [MSG23]. The overall pipeline schematic is presented in Fig.1. Each sample is assigned two class labels: *morphs* inherit them from source identities; *bona fides* have a duplicated original label. The approach requires using two different networks for face recognition classification with two different sets of labels. The main motivation is learning high-level identity discriminative features, which can indicate the presence of face morphing. Such classification is regulated by the explicit binary classification of a dot product of those resulting high-level features.

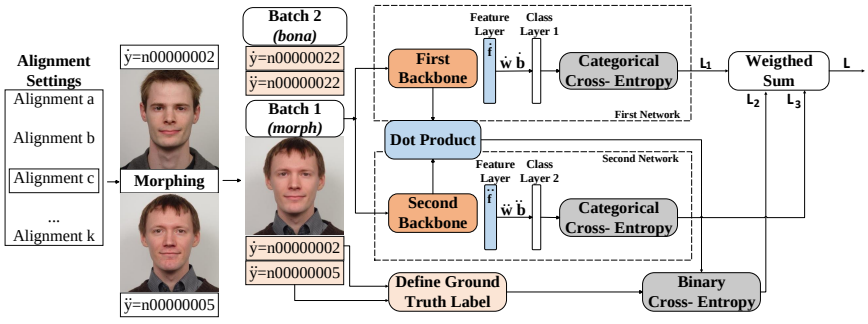

Fig. 1: S-MAD fused approach schematics. In order to simplify the visualization, a single image is shown per batch.

Mathematically, such a schematic is formulated as the weighted sum: $L = \alpha_1 L_1 + \alpha_2 L_2 + \beta L_3$, where L_1 and L_2 are face recognition components, and L_3 is a morphing detection component. Based on the common softmax formulation, each network is regularized by the respective losses:

$$L_1 = -\frac{1}{N}\sum_{i}^{N}\log(\frac{e^{\dot{W}_{\dot{y}_i}^T \dot{f}_i + \dot{b}_{\dot{y}_i}}}{\sum_j^C e^{\dot{f}_{\dot{y}_j}}}), \qquad L_2 = -\frac{1}{N}\sum_{i}^{N}\log(\frac{e^{\ddot{W}_{\ddot{y}_i}^T \ddot{f}_i + \ddot{b}_{\ddot{y}_i}}}{\sum_j^C e^{\ddot{f}_{\ddot{y}_j}}}), \tag{1}$$

where f_i are deep features of the i^{th} sample, y_i represents the class index of the i^{th} sample, and W and b denote the weights and biases of the last fully connected layer, respectively. N represents the batch size, while C represents the total number of classes.

Finally, in order to determine the similarity metric based on the ground truth authenticity label of the image, the morphing detection score is obtained by computing the dot product of the backbone outputs ($\dot{f} \cdot \ddot{f}$). This score is then passed through the *sigmoid* function and used to define the binary cross-entropy loss. As a final result, the corresponding loss is defined by:

$$L_3 = -\frac{1}{N}\sum_{i}^{N} t \log \frac{1}{1+e^{-\ddot{f}\cdot\dot{f}}} + (1-t)\log\left(1 - \frac{1}{1+e^{-\ddot{f}\cdot\dot{f}}}\right) \tag{2}$$

The optimization process involves combining the resulting losses as a weighted sum, resulting in L, with the goal of minimizing it. This is done to learn facial features that are discriminative and specifically regularized for the task of detecting morphing.

S-MAD - Binary Classification. Another approach for face morphing detection is indeed similar to the straightforward binary classification (morph/non-morph). To implement it, we removed the identity classification part from the fused approach and retained only a single deep network in the entire pipeline. The model schema is presented in Fig.2.

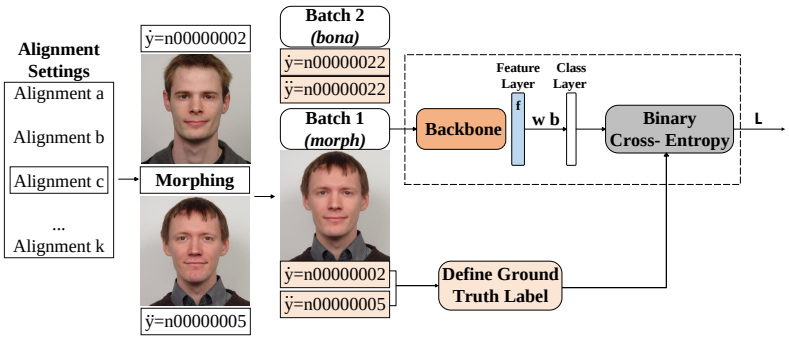

Fig. 2: S-MAD approach model schema for a single network. In order to simplify the visualization, a single image is shown per batch.

Benchmarking. For performance estimation, we employ the open-source morphing benchmarking utilities [4] and adopt them into our work. We replace the bona fide subset with the images from FRLL-Set [DJ17], Utrecht [Un99], MIT-CBCL [He01] and EFIEP [Av19]

[4] https://github.com/iurii-m/MorDeephy.git

(since the default suggested protocols share images with our training data). All protocols share the same list of bona-fide images and are only different in the content of morphs, which are taken from the FRLL-Morphs dataset [Sa22] (protocol names correspond to the FRLL-Morph subset names): *protocol-asml* with $\sim$ 2k morphs, *protocol-opencv* with $\sim$ 1.3k morphs, *protocol-facemorpher* with $\sim$ 2k morphs, *protocol-webmorph* with $\sim$ 1k morphs and *protocol-stylegan* with $\sim$ 2k morphs.

Heatmap Computation. We analyze the image context impact using the Gradient-Weighted Class Activation Mapping (Grad-CAM) technique and generate a heatmap that highlights the regions of the input image that have the most significant influence on the ground truth binary prediction.

4 Experiments and Results

Training Settings. As a baseline model in our work, we use EfficientNetB3 [TL19], which is pretrained on the ImageNet dataset. We trained our models for five epochs using a stochastic gradient descent (SGD) optimizer with a momentum of 0.9 and a learning rate linearly decaying from 0.075 to 1e-5 . The batch included 28 images. Separate training experiments are performed for each alignment case on concatenated datasets: original, LDM, StyleGAN morphs, and *selfmorphs*. Face morphs are generated with LDM and StyleGAN approaches. The parameters for the fused approach, which determine the appropriate balance between the different components of the loss function, are taken from the original work [MSG23]: $\alpha = \alpha_1 = \alpha_2$ and α/β=0.2.

Binary Classification. Based on the results presented in Table 2, the alignment range with optimal performance is observed between *e* to *g*, with *e* being the possible optimal case. Based on heatmaps, the face is the principal region for the detection decision, and the regions, which are prompt to contain morphing artifacts, are mainly activated (see Fig. 3).

Tab. 2: BPCER@APCER = (0.1, 0.01) of our S-MAD binary approach for various alignment settings

Alignments	**BPCER@APCER=δ**									
	Protocol-asml		**Protocol-facemorpher**		**Protocol-opencv**		**Protocol-stylegan**		**Protocol-webmorph**	
	δ=0.1	**δ=0.01**	**δ=0.1**	**δ=0.01**	**δ=0.1**	**δ=0.01**	**δ =0.1**	**δ=0.01**	**δ=0.1**	**δ=0.01**
a	0.199	0.622	0.125	0.558	0.199	0.663	0.663	0.663	0.523	0.663
b	0.143	0.380	0.131	0.387	0.144	0.440	0.586	0.586	0.340	0.586
c	0.365	0.630	0.331	0.675	0.320	0.676	0.676	0.676	0.489	0.676
d	0.236	0.511	0.161	0.549	0.161	0.489	0.623	0.623	0.436	0.623
e	**0.141**	**0.348**	**0.102**	0.532	**0.080**	**0.424**	0.710	0.710	**0.321**	0.641
f	0.199	0.455	0.127	0.551	0.125	0.533	0.675	0.675	0.328	0.579
g	0.158	0.373	0.106	0.532	0.209	0.532	0.586	**0.586**	0.348	0.586
h	0.330	0.580	0.138	0.682	0.093	0.486	0.724	0.724	0.486	0.724
i	0.214	0.408	0.174	0.476	0.149	0.442	**0.573**	0.573	0.396	**0.573**
j	0.221	0.465	0.187	0.596	0.141	0.457	0.776	0.776	0.475	0.682
k	0.243	0.498	0.194	0.557	0.146	0.513	0.794	0.794	0.467	0.707

Fused Classification. For this approach, the optimal range is observed at alignment settings from *d* to *i*, with *g* being possibly the optimal case. At the same time, this methodology allows for superior results in comparison to the binary classification case, which may be related to the regularization imposed by the face recognition task.

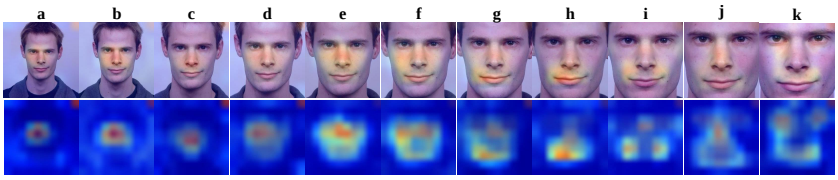

Fig. 3: Grad-CAM morph heatmaps for the S-MAD binary approach under different alignment conditions (Recall that bona-fide sets are equal across all the protocols).

Tab. 3: BPCER@APCER = (0.1, 0.01) of S-MAD fused approach for various alignment settings

Alignments	**BPCER@APCER=δ**									
	Protocol-asml		**Protocol-facemorpher**		**Protocol-opencv**		**Protocol-stylegan**		**Protocol-webmorph**	
	δ=0.1	**δ=0.01**	**δ=0.1**	**δ=0.01**	**δ=0.1**	**δ=0.01**	**δ =0.1**	**δ=0.01**	**δ=0.1**	**δ=0.01**
a	0.159	0.689	0.187	0.517	0.239	0.599	0.842	0.946	0.606	0.885
b	0.063	0.495	0.072	0.646	0.099	0.658	0.671	0.946	0.702	0.964
c	0.125	0.467	0.215	0.588	0.240	0.566	0.694	0.884	0.541	0.859
d	0.040	0.374	0.102	0.558	0.103	0.568	0.574	0.835	0.305	0.781
e	0.162	0.580	0.149	0.582	0.177	0.602	0.566	0.767	0.605	0.870
f	0.184	0.530	0.180	0.488	0.175	0.451	0.582	0.788	0.517	0.785
g	**0.034**	**0.233**	**0.025**	0.701	**0.037**	0.701	0.487	0.875	**0.216**	0.788
h	0.168	0.642	0.168	0.535	0.165	0.599	0.536	0.850	0.542	0.854
i	0.046	0.255	0.036	**0.365**	0.044	**0.390**	**0.305**	**0.583**	0.246	**0.554**
j	0.287	0.630	0.268	0.585	0.262	0.564	0.844	0.959	0.697	0.907
k	0.193	0.652	0.253	0.745	0.262	0.792	0.825	0.953	0.674	0.915

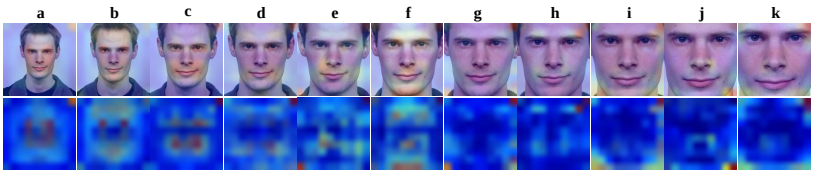

Fig. 4: Grad-CAM morph heatmaps for the S-MAD fused approach under different alignment conditions (Recall that bona-fide sets are equal across all the protocols).

Based on the heatmaps, the detection is mainly focused on the face region and, in many cases, on the regions of intersection between the foreground and background (see Figure 4).

NIST FRVT MORPH Results. We compare the results of our best model (fused case) with several state-of-the-art (SOTA) MAD approaches, tested on the FRVT NIST MORPH Benchmark [FR]. Each dataset from the benchmark has images generated through different protocols, with distinctions made in tiers such as Tier 2 - Automated Morph Analysis and Tier 3 - High-Quality Morph Analysis.

Regarding the Visa-Border dataset, our approach outperforms all other SOTA approaches, with a *morph miss rate* of 0.29 at a *false detection* rate of 0.01. In the Twente dataset, when comparing with other approaches, the results demonstrate a highly favorable outcome as well, with a *morph error rate* of 0.128 at a *false detection rate* of 0.01 (See table 4).

Tab. 4: Comparison with the SOTA S-MAD approaches using APCER@BPCER = (0.1, 0.01).

Algorithm	**Visa-Border (Tier 2)**		**Twente (Tier 2)**		**Manual (Tier 3)**	
	δ=0.1	**δ=0.01**	**δ=0.1**	**δ=0.01**	**δ=0.1**	**δ=0.01**
Our	0.089	**0.291**	0.032	0.128	0.802	0.975
Aghdaie et al. [Ag21]	0.037	0.542	0.002	**0.060**	0.879	0.975
Medvedev et al. [MSG23]	0.232	0.555	0.174	0.493	0.641	**0.926**
Ferrara et al. [FFM21]	0.477	0.999	0.002	0.183	0.938	0.985
Ramachandra et al. [Ra19]	0.375	0.990	0.304	0.998	0.938	0.985

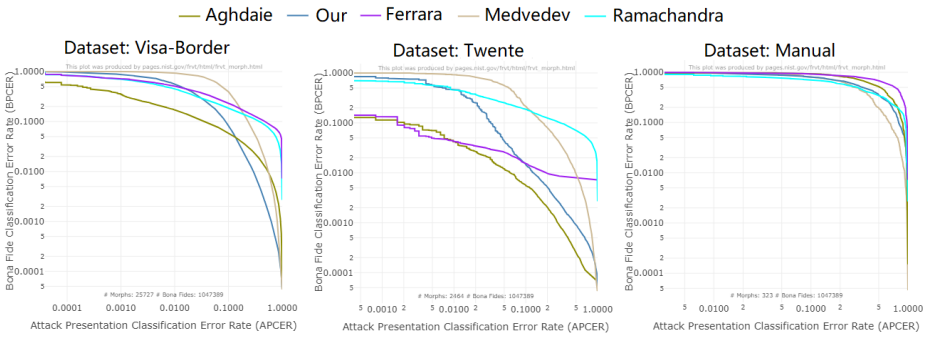

Fig. 5: Detection Error Trade-off curves for different SOTA approaches in different datasets (Visa-Border, Twente and Manual dataset).

Although not represented in the table, comparable results were achieved for other datasets, such as the UNIBO Automatic Morphed Face Generation Tool v1.0 and even MIPGAN-II with less dominant but still competitive performances.

It is important to take into consideration the influence of the dataset used, and this Tier 2 typology is generally less challenging. When faced with more realistic datasets (Manual dataset), it becomes apparent that overall SOTA approaches show poor generalization across various unseen morphing techniques. Even so, our model results achieved competitive results when compared to those approaches.

5 Conclusions

In this work, we aim to identify the context properties that are most effective for S-MAD. The extensive experiments allowed us to determine the alignment range where S-MAD is more effective. Moreover, in this range, there seems to be a certain correspondence between both fused and binary approaches, which translates into a similar area of face occupancy in the image. Despite that, our results also show that face is the most dominant activation region across all the alignment settings, and the impact of context on face morphing detection is limited. Our method achieved state-of-the-art comparable performances on some of the NIST FRVT MORPH benchmark protocols.

Our future work will be directed toward investigating similar properties in the differential scenario.

References

[A.99] A. V. Nefian: , Georgia Tech face database. http://www.anefian.com/research/face_reco.htm, 1999.

[Ag21] Aghdaie, P.; Chaudhary, B.; Soleymani, S.; Dawson, J.; Nasrabadi, N.: Attention aware wavelet-based detection of morphed face images. CoRR, abs/2106.15686, 2021.

[Av19] Avilés, J.; Toapanta, H.; Morillo, P.; Vallejo-Huanga, D.: Dataset of Ethnic Facial Images of Ecuadorian People. 2019.

[Da18] Damer, N.; Saladié, A. M.; Braun, A.; Kuijper, A.: MorGAN: Recognition Vulnerability and Attack Detectability of Face Morphing Attacks Created by GAN. In: 2018 IEEE 9th International Conference on BTAS). pp. 1–10, Oct 2018.

[Da21] Damer, N.; Raja, K. B.; Sussmilch, M.; Venkatesh, S. K.; Boutros, F.; Fang, M.; Kirchbuchner, F.; Ramachandra, R.; Kuijper, A.: ReGenMorph: Visibly Realistic GAN Generated Face Morphing Attacks by Attack Re-generation. In: ISVC. 2021.

[Da23] Damer, N.; Fang, M.; Siebke, P.; Kolf, J. N.; Huber, M.; Boutros, F.: , MorDIFF: Recognition Vulnerability and Attack Detectability of Face Morphing Attacks Created by Diffusion Autoencoders, 2023.

[De19] Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S.: ArcFace: Additive Angular Margin Loss for Deep Face Recognition. In: 2019 IEEE/CVF CVPR. pp. 4685–4694, 2019.

[DJ17] DeBruine, L.; Jones, B.: , Face Research Lab London Set, May 2017.

[FFM14] Ferrara, M.; Franco, A.; Maltoni, D.: The magic passport. In: IEEE IJCB. pp. 1–7, 2014.

[FFM21] Ferrara, Matteo; Franco, Annalisa; Maltoni, Davide: Face morphing detection in the presence of printing/scanning and heterogeneous image sources. IET Biometrics, 10, 02 2021.

[FR] FRVT MORPH. https://pages.nist.gov/frvt/html/frvt_morph.html.

[FS05] Fagertun, Jens; Stegmann, Mikkel Bille: , The IMM Frontal Face Database. http://www2.imm.dtu.dk/ aam/datasets/imm_frontal_face_db_high_res.zip, 2005.

[Go14] Goodfellow, I. J.; P.Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.l; Courville, A.; Bengio, Y.: Generative Adversarial Nets. In: NeurIPS. Curran Associates, Inc., 2014.

[H.21] H.Zhang, and Venkatesh, S.; Ramachandra, R.and Raja, Kiran; Damer, N.; Busch, C.: MIPGAN—Generating Strong and High Quality Morphing Attacks Using Identity Prior Driven GAN. IEEE T-BIOM, 3(3):365–383, 2021.

[He01] Heisele, B.; Serre, T.; Pontil, M.; Vetter, T.: , MIT-CBCL FR Database. http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html, 2001.

[Ki09] King, Davis E.: Dlib-ml: A Machine Learning Toolkit. Journal of Machine Learning Research, 10:1755–1758, 2009.

[KJL22] Kim, M.; Jain, A. K; Liu, X.: AdaFace: Quality Adaptive Margin for Face Recognition. In: Proceedings of the IEEE/CVF CVPR. 2022.

[KLA19] Karras, T.; Laine, S.; Aila, T.: A Style-Based Generator Architecture for Generative Adversarial Networks. In: 2019 IEEE/CVF CVPR. pp. 4396–4405, 2019.

[Li20] Li, L.; Mu, X.; Li, S.; Peng, H.: A Review of Face Recognition Technology. IEEE Access, pp. 139110–139120, 2020.

[MB98] Martinez, A.; Benavente, R.: , The AR Face Database: CVC Technical Report, 24. https://www2.ece.ohio-state.edu/ aleix/ARdatabase.html, 1998.

[Me00] Messer, K.; Matas, J.; Kittler, J.; Jonsson, K.; Luettin, J.; Maître, G.: Xm2vtsdb: The extended m2vts database. Proc. of Audio- and Video-Based Person Authentication, 04 2000.

[Me21] Meng, Q.; Zhao, S.; Huang, Z.; Zhou, F.: MagFace: A universal representation for face recognition and quality assessment. In: CVPR. 2021.

[MSG23] Medvedev, I.; Shadmand, F.; Gonçalves, N.: MorDeephy: Face Morphing Detection via Fused Classification. In: Proceedings of the 12th ICPRAM. SciTePress, pp. 193–204, 2023.

[Ne] Neto, P. C.; Gonçalves, T.; Huber, M.; Damer, N.; Sequeira, A. F.; Cardoso, J. S.: OrthoMAD: Morphing Attack Detection Through Orthogonal Identity Disentanglement. In: BIOSIG 2022. pp. 1–5.

[OH08] Ojansivu, V.; Heikkilä, J.: Blur Insensitive Texture Classification Using Local Phase Quantization. In: Springer-Verlag. ICISP '08, Berlin, Heidelberg, p. 236–243, 2008.

[OPH96] Ojala, T.; Pietikäinen, M.; Harwood, D.: A comparative study of texture measures with classification based on featured distributions. Pattern recognition, 29(1):51–59, 1996.

[Ph98] Phillips, P.; Wechsler, H.; Huang, J.; Rauss, P.: The FERET database and evaluation procedure for face-recognition algorithms. Image and Vision Computing, 16(5):295–306, 1998.

[Ph00] Phillips, P.J.; Moon, H.; Rizvi, S.A.; Rauss, P.J.: The FERET evaluation methodology for face-recognition algorithms. IEEE TPAMI, 22(10):1090–1104, 2000.

[Ph05] Phillips, P.; Flynn, P.; Scruggs, W.; Bowyer, K.; Chang, J.; Hoffman, K.; Marques, J.; Min, J.; Worek, W.: Overview of the Face Recognition Grand Challenge. In: IEEE CVPR. 2005.

[Ph11] Phillips, P.; Flynn, P.; Bowyer, K.; Vorder, R.; Grother, P.; Quinn, G.; Pruitt, M.: Distinguishing Identical Twins by Face Recognition. In: 9th IEEE FG 2011, Santa Barbara, CA. 2011-03-21 2011.

[Ra19] Ramachandra, R.; Venkatesh, S.; Raja, K.; Busch, C.: Towards making morphing attack detection robust using hybrid scale-space colour texture features. In: 2019 IEEE 5th ISBA. pp. 1–8, 2019.

[RRB16] Raghavendra, R.; Raja, K. B.; Busch, C.: Detecting morphed face images. In: 2016 IEEE 8th International Conference on BTAS. pp. 1–7, Sep. 2016.

[Sa22] Sarkar, E.; Korshunov, P.; Colbois, L.; Marcel, S.: Are GAN-based morphs threatening face recognition? In: ICASSP 2022. pp. 2959–2963, 2022.

[Sc19] Scherhag, U.; Debiasi, L.; Rathgeb, C.; Busch, C.; Uhl, A.: Detection of Face Morphing Attacks Based on PRNU Analysis. IEEE T-BIOM, 1(4):302–317, 2019.

[TB21] Tapia, J. E.; Busch, C.: Single Morphing Attack Detection Using Feature Selection and Visualization Based on Mutual Information. IEEE Access, 9:167628–167641, 2021.

[Th06] Thomaz, C. E.: , FEI- Face Database. https://fei.edu.br/ cet/facedatabase.html, 2006.

[TL19] Tan, Mingxing; Le, Quoc V.: EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks. arXiv, 2019.

[Un99] University of Stirling: , Psychological Image Collection at Stirling (PICS). http://pics.stir.ac.uk/, 1999.

[XZ17] Xiang, J.; Zhu, G.: Joint Face Detection and Facial Expression Recognition with MTCNN. In: 2017 4th ICISCE. pp. 424–427, 2017.