2021

Single and Differential Morph Attack Detection

Baaria Chaudhary
West Virginia University, bac0062@mix.wvu.edu

Single and Differential Morph Attack Detection

Baaria Chaudhary

Thesis submitted to the
Benjamin M. Statler College of Engineering and Mineral Resources
at West Virginia University in partial fulfillment of the requirements
for the degree of

Master of Science in Computer Science

Nasser M. Nasrabadi, Ph.D., Chair
Matthew Valenti, Ph.D
Jeremy Dawson, Ph.D

Lane Department of Computer Science and Electrical Engineering

Morgantown, West Virginia
2021

Keywords: Face Recognition, Morph Attack Detection, Deep Learning, Wavelet Sub-bands

Abstract

Single and Differential Morph Attack Detection

Baaria Chaudhary

Face recognition systems operate on the assumption that a person's face serves as the unique link to their identity. In this thesis, we explore the problem of morph attacks, which have become a viable threat to face verification scenarios precisely because of their inherent ability to break this unique link. A morph attack occurs when two people who share similar facial features morph their faces together such that the resulting face image is recognized as either of two contributing individuals. Morphs inherit enough visual features from both individuals that both humans and automatic algorithms confuse them. The contributions of this thesis are two-fold: first, we investigate a morph detection methodology that utilizes wavelet sub-bands to differentiate bona fide and morph images. Second, we investigate the usefulness of morphing identical twins to train a network robustly.

Although not always discernible in the image domain, many morphing algorithms introduce artifacts in the final image that can be leveraged for morph attack detection. Because wavelet decomposition allows us to separately examine low and high frequency data, we can identify and isolate these morphing artifacts in the spatial frequency domain. To this end, a wavelet-based deep learning approach to detect morph imagery is proposed and evaluated. We examine the efficacy of wavelet sub-bands for both single and differential morph attack detection and compare performance to other methods in the literature.

Finally, experiments are done on a large scale morph dataset created using twins. This high quality morph twins dataset is used to train a single morph detector. The details of this detector are explained and the resulting morph detector is submitted to the NIST FRVT test for objective evaluation, where our detector exhibited promising results.

Acknowledgments

This work was completed with the guidance and contributions of many people. I would like here to specially thank them for their help.

I would like to thank my research advisor, Dr. Nasrabadi, for his guidance and insight, as well as my committee members, Dr. Dawson and Dr. Valenti. I also would like to thank Sobhan Soleymani for his help with the algorithm implementations as well as my other fellow research students: Poorya Aghdaie, Kelsey O'Haire, and Samuel Price for their contributions to this work.

Additionally, I also want to thank my husband, Mirza Nayyar Ahmad, for his endless patience and consideration while I worked on this thesis. Ultimately, I am thankful to God for this opportunity and the help He provided me to complete it.

Publications

1. Sobhan Soleymani, **Baaria Chaudhary***, Ali Dabouei, Jeremy Dawson, and Nasser M. Nasrabadi. "Differential Morphed Face Detection Using Deep Siamese Networks." Published in *ICPRW Multimedia Forensics in the Wild Workshop (MMforWild)*, 2020.

2. Poorya Aghdaie, **Baaria Chaudhary**, Sobhan Soleymani, Jeremy Dawson, and Nasser M. Nasrabadi. "Detection of Morphed Face Images using Discriminative Wavelet Sub-bands." Published in *9th IEEE International Workshop on Biometrics and Forensics (IWBF)*, 2021.

3. **Baaria Chaudhary**, Poorya Aghdaie, Sobhan Soleymani, Jeremy Dawson, and Nasser M. Nasrabadi. "Differential Morph Face Detection using Discriminative Wavelet Sub-bands." Published in *Conference on Computer Vision and Pattern Recognition (CVPR) Biometrics Workshop*, 2021.

4. Poorya Aghdaie, **Baaria Chaudhary**, Sobhan Soleymani, Jeremy Dawson, and Nasser M. Nasrabadi. "Attention Aware Wavelet-based Detection of Morphed Face Images." Published in *International Joint Conference on Biometrics (IJCB)*, 2021.

5. Kelsey O'Haire, Sobhan Soleymani, **Baaria Chaudhary**, Poorya Aghdaie, Jeremy Dawson, Nasser M. Nasrabadi. "Adversarially Perturbed Wavelet-based Morphed Face Generation." Published in *IEEE International Conference on Automatic Face and Gesture Recognition (FG)*, 2021.

6. Poorya Aghdaie, **Baaria Chaudhary**, Sobhan Soleymani, Jeremy Dawson, and Nasser M. Nasrabadi. "Morph Detection Enhanced By Structured Group Sparsity." Published in *Winter Conference on Applications of Computer Vision (WACV)*, 2022.

***equal contribution**

Contents

1 Introduction **1**

1.1 The Face as A Biometric Authenticator 3

1.2 Morph Attacks in Border Control 4

1.3 Thesis Contributions 6

1.4 Thesis Organization 7

2 Literature Review **9**

2.1 Morph Image Generation 10

2.2 Morph Attack Detection 14

2.3 Morph Detection Algorithm Evaluation 15

3 Discriminative Wavelet Sub-band Selection **17**

3.1 2D Discrete Wavelet Transform 19

3.2 Sub-band Selection using KL Divergence 20

3.3 Ablation Study on Sub-band Selection 23

3.4 Summary 25

4 Wavelet-Based Morph Detection **26**

4.1 Network Architecture 26

4.1.1 Siamese Network 27

4.2 Datasets and Pre-processing 28

4.2.1 Preprocessing 30

4.3 Experimental Setup 31

4.4 Metrics 31

4.5 Experiments Overview 32

4.5.1 Wavelet-based Morph Detection Comparison 33

4.5.2 State of the Art 37

4.6 Summary 40

5 Twins for Morph Attack Detection **41**
5.1 The Difficulty of Finding Look-Alikes 42
5.2 Datasets 44
5.3 Experiments 45
5.4 NIST Submission 46
5.5 Summary 51

6 Conclusion **52**
6.1 Limitations 52
6.2 Future Work 53
6.3 Conclusion 53

Bibliography **55**

List of Tables

4.1 Datasets used in training. 29
4.2 Single morph attack performance of the proposed framework and baselines. 33
4.3 Differential Performance of the proposed framework and baselines. With the exception of RGB-66 testing on MorGAN, BW-22 exhibits superior performance. 34
4.4 Single Performance Comparison of Proposed Framework. All algorithms trained with the Universal dataset. 38
4.5 Differential Performance Comparison of Proposed Framework. All algorithms trained with the Universal dataset. 39

5.1 Morph datasets created using twins. 43
5.2 Performance of our twins-trained morph detector on some select databases. 45

List of Figures

1.1 An example of a morph attack: bona fide identity #1 (left), morphed image(middle), bona fide identity #2 (right). 3
1.2 Abstract representation of two similar identities in the embedding space. The morph is designed to be within the decision boundary for the two identities. 4

2.1 Example of ghosting and artifacts resulting from the morphing process. 11
2.2 Morphs created using different techniques with the FRLL database. . 13

3.1 Essential overview of wavelet band selection algorithm. 18
3.2 **Selected sub-bands.** The selected sub-bands are shown with regards to their location in wavelet decomposition. Most of the informative sub-bands chosen by KL divergence are those that have been filtered with the HH filter. 21
3.3 Performance comparison between training on 48 sub-bands (left) and 22 sub-bands (right) shown by their Detection Error Tradeoff (DET) curves. 23

4.1 Single and differential network architecture. 27
4.2 DET curves for all protocols, tested on the (a) MorGAN and (b) LMA test sets. 35

5.1 Bona Fide Twins (far left and far right) with morphs created using different techniques with the Twins Day Dataset. 42
5.2 Performance of our detector, wvusingle_001, on the Local Morph Colorized Average morphs [28]. 47
5.3 Performance of our detector, wvusingle_001, on the MIPGAN-II morphs [28]. 48
5.4 Performance of our detector, wvusingle_001, on the Visa-Border morphs [28]. 49

5.5 Performance of our detector, wvusingle_001, on the Print and Scanned morphs [28]. 49
5.6 DET curve of our morph detector, wvusingle_001, on the impact of image resolution. 50

Chapter 1

Introduction

Face morph attacks have become a serious security concern for deployed face recognition systems, which rely on the assumption that the face modality serves as the unique link to a person's identity. Face morphing refers to the image manipulation process in which the face images of two individuals who already share similar facial features are morphed together. Because of the high degree of resemblance the morph contains to the two contributing individuals, both human inspectors and automatic face recognition algorithms verify the resulting morph face image as coming from either of the two original people. See Figure 1.1 for an example of a morph attack from the AMSL database [27].

Morph attacks expose vulnerabilities in many security applications. Face morph attacks are especially dangerous for the border control scenario because a morphed passport photo means that a person who is not previously authorized to enter a country can cross borders undetected. The loophole in the passport enrollment process opens a window for a criminal to morph his face with that of an accomplice, who then applies for a passport with the morphed face image. Because the morphed face image resembles the accomplice, it is approved and the passport is issued. The criminal then receives a legitimate travel document that, although fraudulently obtained, allows him to travel across borders and access restricted areas that otherwise would be

closed to him. For border control, morph passports can lead to illegal immigration, human trafficking, and getting around no-fly lists. Realizing this problem, several countries have invested in research in how to mitigate these attacks both to detect morphs currently in the system and prevent morph attacks in the future.

Successful morphs are not visually perceptible, which makes them especially difficult to detect. As can be seen in Figure 1.2, if we look at the embedding space representation of two identities with similar features, we see that morphed samples are intentionally crafted to be within the discriminating boundaries of the two identities, which are already close in the embedding space because they look alike. This is why morphs can be verified against both real subjects: they contain enough visual similarity to both contributing individuals. This is a trend that will only continue to be exacerbated as morphing technologies become more sophisticated. Therefore, it is imperative to find an alternative solution to simple visual confirmation.

Although not always visually obvious to the human eye, many automatic face morphing algorithms, such as landmark manipulation and GAN generation, introduce artifacts in the final image that indicate an image was morphed. These morphing artifacts mainly reside in the high frequency spectrum. Even though these artifacts are not typically discernible in the image domain, they can be leveraged for morph attack detection using wavelet decomposition. Because wavelet decomposition allows us to separately examine low frequency and high frequency data, we can identify and isolate these morphing artifacts in the spatial frequency domain. To this end, a wavelet-based deep learning approach to detect morph imagery is proposed. Our method leverages selected informative sub-bands as features to train a deep morph detector.

Figure 1.1: An example of a morph attack: bona fide identity #1 (left), morphed image(middle), bona fide identity #2 (right).

1.1 The Face as A Biometric Authenticator

The face is an incredibly powerful biometric modality that has found purpose in a variety of security applications: from verification to surveillance, and from passports to personal phone lock protections. Unlike key-based security systems, it is impossible to lose or forget your face the way you lose or forget your keys or passwords. Face recognition's ease of use and comparatively low operational cost puts it above all other biometric modalities for real world applications. Furthermore, if the face recognition algorithm goes offline or triggers a false alarm, a human inspector on site can easily perform the verification with minimal training. This makes it especially attractive for border security crossings, specifically in areas where access to advanced technologies may be limited and resource constraints must be considered. It is for these reasons that since 2002 the International Civil Aviation Organization (ICAO) has mandated the inclusion of a facial reference passports in all passports [15]. This means that the face is the only biometric modality universally recorded in every passport worldwide.

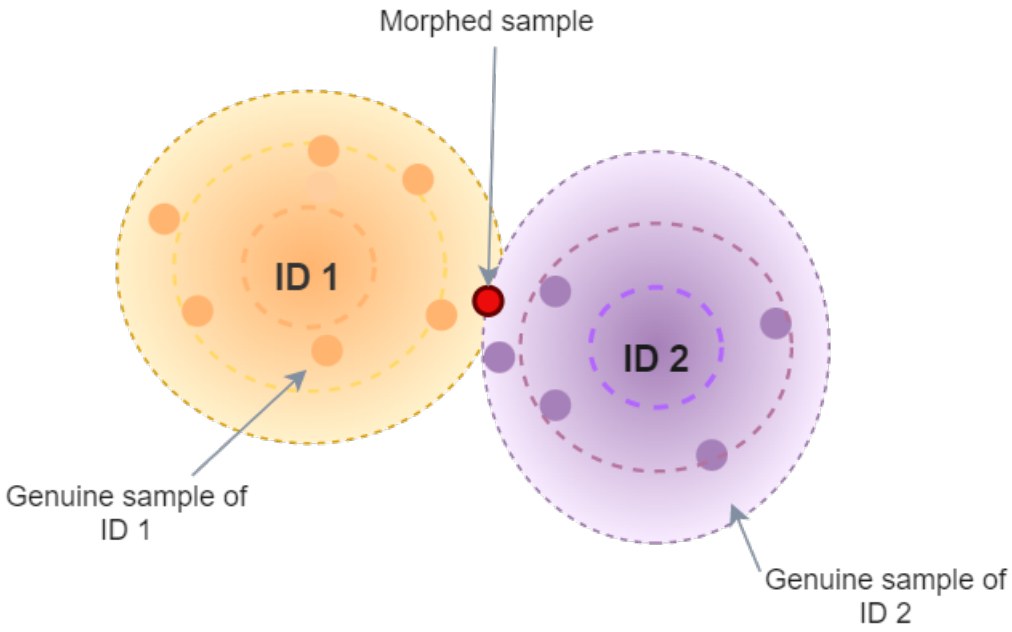

Figure 1.2: Abstract representation of two similar identities in the embedding space. The morph is designed to be within the decision boundary for the two identities.

1.2 Morph Attacks in Border Control

The steps of the passport biometric system, as outlined by ICAO, are enrollment, template creation, identification, and verification. Morph attacks target the enrollment step of the border security biometric system pipeline. The passport enrollment process is vulnerable because the applicant is requested to provide the photo at the time of application. This means that the actual taking and submitting of the photo is an unsupervised process, allowing a window for a malicious actor to manipulate the photo. Hence, it is during the passport enrollment process that an image that has been tampered with is submitted and assigned to an identity.

The identity verification process operates on the assumption that the face serves as the unique link between a person and their intrinsic identity, meaning that the face is the distinguishing characteristic that differentiates one person from the other. For the most part, this is true. Face recognition systems have shown that they can differentiate one individual from another with a high amount of accuracy. However,

face recognition system's vulnerability and, for lack of a better word, gullibility to a misleading or forged image is a major security concern. Morph images threaten the integrity of face recognition systems because of their inherent ability to break this unique link: successful morph images inherit enough features from both contributing individuals to be within the threshold of acceptance for face recognition algorithms. This means that the principle of unique link is violated whenever a morph image is entered in the passport database during the enrollment process. Moreover, face recognition algorithms are intentionally designed to tolerate a large intra class variance to account for the significant changes in facial appearance that naturally occur in the 5- to 10- year life cycle of a passport.

The threat of morphs attacks to border security is further exacerbated by the fact that morph attacks are relatively easy technically to execute. There is no complex forgery of passport technology but rather a straightforward manipulation of the passport photo at time of enrollment. Many morphing algorithms are not only freely accessible but also have no technical knowledge barrier. With some photo editing skills, a criminal could obtain a travel document that, although fraudulently obtained, is legitimate and genuine. Indeed, the more difficult part is actually finding a suitable look-alike. Morph attacks are appealing to criminals precisely because of how comparatively simple it is to execute and how effective morph attacks can be at fooling face recognition systems. In vulnerability terms, this means one person can enter the country and another person can leave the country using the same identity. Governments issue passports on the assumption that one passport is for one specific person only and the passports are used to keep track of who enters and who leaves the country. As these technologies become more accessible and more advanced and no morph detection algorithm is instituted to detect these manipulated images, national security is at stake.

1.3 Thesis Contributions

In this thesis, we examine the problem of face morph attacks, which significantly subvert the integrity of face recognition systems. The contributions of this thesis are two-fold: first, we investigate a new morph detection methodology that utilizes wavelet sub-bands to differentiate bona fide and morph images. Second, we investigate the power of using twins to create morph databases that are of high quality to train a network robustly. We submit our twins-trained network to NIST for objective evaluation where it performs reasonably well.

We propose a highly discriminative morph attack detection algorithm that utilizes undecimated discrete wavelet transform. As mentioned earlier, the morphing process introduces involuntary artifacts in the image, mainly those in the high frequency spectrum. We can utilize wavelet transform to extract and analyze these artifacts in an effort to detect morphs. The core of our framework is that morphing artifacts that can be readily identified in the spatial frequency domain, more so than in the image domain. As such, we decompose each image into its respective wavelet sub-bands and utilize Kullback-Liebler divergence (KLD) as a feature selection method to determine which sub-bands contain the most differentiation between a bona fide and morphed image. These discriminative sub-bands are then used to train a deep neural network for both single morph (classification) and differential morph (verification) scenarios. We examine the efficacy of discriminative wavelet sub-bands for morph attack detection and demonstrate that a wavelet-based morph detection can accurately identify morph imagery. Extensive experiments are conducted on three different morph image datasets and the performance is bench marked with several state-of-the-art techniques. The major contributions of this work are as follows:

- a novel morph attack detection algorithm for both single and differential morph detection that employs wavelet transform to extract and select discriminative

wavelet sub-bands, which are then used to train a morph detector.

- an ablation study on the sub-band selection process using KL divergence to show the efficacy of our method and reasoning behind the sub-bands selected for final training.

- an evaluation of discriminative wavelet sub-bands by comparing its performance to several baselines in the image domain, including the original (RGB, 3-channel) image.

- We compare our model's performance to several state-of-the-art models: BSIF, LBP, SIFT, and SURF. Deep learning method FaceNet is also used. Additionally, cross-dataset performance is evaluated using AMSL [27].

- We explore the efficacy of using a large-scale high quality morphed twins database to train a single morph detector, which we submit to NIST for evaluation.

1.4 Thesis Organization

The remainder of this thesis is organized as follows:

- Chapter 2 describes existing work related to the morph attack problem. Both morph generation techniques and morph detection algorithms are mentioned. The current standard in morph evaluation is also discussed.

- Chapter 3 discusses the overall wavelet sub-band selection methodology: wavelet decomposition and subsequent sub-band selection. An in-depth discussion of the ablation study and the effectiveness of sub-band selection is also explored.

- Chapter 4 presents the experimental setup details for the wavelet based morph detector for both single and differential scenarios. The metrics used to evaluate performance and dataset preprocessing information is also included. These

methods are also explored in relation to several comparative baselines including training using the original RGB image to validate the effectiveness of selected wavelet sub-bands for morph detection. Additionally, the method's performance is also benchmarked in comparison to several state-of-the-art methods.

- Chapter 5 discusses using morphed twins to train a morph detector robustly. We submitted this detector to NIST FRVT Morph test and the results from the test are discussed in detail.

- Chapter 6 serves as the conclusion with an overview of thesis contributions, the limitations confronted within the thesis, and future work in this field.

Chapter 2

Literature Review

The vulnerability of face recognition systems to morph attacks was first introduced by Ferrara et. al. [12], who describes the step-by-step process a criminal takes to morph an image and obtain a valid passport. This was the first time the issue was discussed seriously as a threat in a research setting. However, morph attacks are not simply theoretical and in fact are a real world issue.

As there is no morph attack detection algorithm in place at most border security crossings, most of these fraudulent passports go undetected. An activist group in Germany was able to take advantage of this loophole and successfully obtain a passport using a morphed image as part of a political media campaign [48]. When morphs are discovered, it is mostly by pure luck. For example, one asylum seeker who was traveling through Europe with a morphed Dutch passport was stopped at German border upon entry [18]. Some countries have already planned to heighten security measures in response to the threat of morphing. In Germany, plans were made to make people take their passport photos in official government photo booths that would immediately transfer the taken passport photo directly to the office. However, it was scrapped due to cost and protests from private photography studios that would lose business to the initiative [14].

Even if the passport enrollment process is eventually changed, closing the window of opportunity to manipulate the passport photo, there are still a countless number of morph passports in circulation now. Furthermore, this change would only stop morph attacks at border security. Morph attacks pose a threat to all identification scenarios: private, corporate, and governmental.

In every case, it is crucial to the success of a morph that it is able to deceive a human observer. If automatic face recognition algorithms exhibit weakness in detecting morphs, humans are much worse [36] [32] [3]. Studies show that even after training on how to identify the clues that indicate a morphed image, algorithms perform better than humans significantly [24] [19].

This being the case, both morph image generation and morph attack detection have become active areas of research in recent years. This chapter will first explore morph generation techniques, followed by a discussion of existing morph attack detection algorithms as well as existing methods to benchmark results.

2.1 Morph Image Generation

As a technique, image morphing has been around for decades. It is widely used in the film industry to create special effects. Although morphing tools exist for both legitimate business and personal entertainment purposes, this technology can easily be abused with malicious intent.

The morph generation process involves finding two people who look alike and morphing their faces together. This first step is crucial – especially in the border security scenario, where the face image must be able to fool both border inspectors and the algorithm and the consequences of discovery are dire. There has been some research in how to find suitable look-alikes [7] [37]. However, this is still a growing research area.

(a) Bona Fide (b) Morph

Figure 2.1: Example of ghosting and artifacts resulting from the morphing process.

Once look-alikes have been found, the second step is to generate a high quality morph image. If the image is of low quality and quite clearly has artifacts, as can be seen in Figure 2.1, a real world criminal would not risk discovery by using it to apply for a passport. But a database of actual morphed passports deployed by actual criminals is impossible to get. Instead, researchers must resort to creating their own synthetic morph image databases, typically employing automatic generation techniques. The best quality morph images are those created using manual manipulation techniques, which allow for a custom merging of two individuals. Although manual manipulation techniques such as Adobe Photoshop or GIMP 2.10 produce high quality morphs, using such methods in a research setting to create a database satisfactorily large enough to train an algorithm optimally is expensive, time consuming, and ultimately unfeasible.

The automatic face morphing generation techniques employed by researchers typ-

ically fall into two categories: landmark-based (manipulating the face in the image domain) and GAN-based (manipulating the morph in the latent space domain). Figure 2.2 demonstrates examples from various morphing techniques. The following paragraphs delve into detail about these morphing techniques.

The most commonly used automatic generation technique is landmark manipulation (LMA) [25]. At the center of many automatic morphing algorithms, landmark manipulation consists of finding the landmarks, or edges, of the two faces, warping and alpha-blending the landmarks together to create the morphed image. The LMA pipeline consists of three key steps: (1) correspondence, or identifying the landmarks (keypoints) on the face images (typically around the nose, eye, and mouth regions), (2) warping, or applying Delaunay Triangulation and warping each triangle to the mesh to align the images, and (3) blending the aligned faces according to some alpha value to produce the final morph. Landmark manipulated morphs are shown to be difficult to detect and preserve the identity well, their weakness is in the ghosting artifacts that occur around the edges of the face, especially the hair region, which show the image is morphed. OpenCV [25], FaceMorpher [31], and Webmorph [9] are a few of several morphing techniques employed by researchers that use landmark manipulation at their center.

Although landmark manipulation morphs are challenging for morph detectors, they can sometimes be visually obvious to the human eye due to the artifacts that result from the warping process, typically around the eyes and ghosting around the hairline. To counteract these obvious morph indicators, typically landmark manipulation morphing pipelines include a series of post-processing steps intended to eliminate the presence of these artifacts, such as image smoothing, manual retouching, image sharpening, and image enhancement to improve any noticeable change in brightness and contrast. A splicing technique, used in the creation of the VISAPP database [23], morphs the landmarks in the inner face region only. The resulting face is then splic-

Figure 2.2: Morphs created using different techniques with the FRLL database.

ing into one of the contributing individuals. In this way, the ghosting and artifacts that typically occur around the hair regions and edges of the face are avoided. An extension of the splicing technique is the Combined Morph [27] in which the faces are aligned before the morphing process.

Generative Adversarial Networks (GANS) have also recently gained traction for morph generation. GAN generation is more of an automatic end-to-end generation technique in comparison to landmark-based techniques, which sometimes require post-processing. The encoder in a GAN can transform images to a latent space and when two latent spaces related to two different subjects are combined, a morph image is synthesized. MorGAN [5] [8] creates morphs using an AliGAN inference model. StyleGAN [17] has also been used for morph generation. A GAN network based on identity prior generation that builds off of StyleGAN with the use of identity and perceptual loss has also been proposed [54]. GAN generated morphs also avoid the artifacts and ghosting that occurs in other morphing techniques, which has contributed greatly to their popularity. However, it should be noted that studies show that GAN generated morphs are typically weaker than landmark manipulation techniques because they retain less identity information [51]. To counter this effect and still use the power of GANs, the authors of [6] propose a morph generation algorithm that utilizes both landmark and GAN architectures. The two face images are first morphed together used landmark manipulation and then the resulting morph is sent to a GAN network to clean up the artifacts.

2.2 Morph Attack Detection

There are two scenarios researchers broadly consider for morph attack detection: single (no reference) and differential morph detection. Single (no reference) morph detection is when the algorithm bases its classification result only on the potential morphed image. In the border security scenario, this correlates to morph detection at time of passport enrollment, when the applicant submits the passport photo in either digital or physical form. On the other hand, differential morph detection uses an additional trusted image, typically a live capture at border control, to compare to the potential morphed image to make its decision. This scenario correlates to detecting a morph attack at border security when a person is actually attempting to use the morphed passport photo for identification. Differential morph detection utilizes the extra information present in the secondary image to make its decision, hence it is more comparative than single morph detection.

Many classical hand-crafted feature descriptors have been used to detect morphs. Of these, typically the better performing is Binarized Statistical Image Features (BSIF) [16] [33]. Local Binary Patterns (LBP) [20] , Scale-Invariant Feature Transform (SIFT) [22], Speeded-Up Robust Features (SURF) [1], and Histogram of Gradients (HOG) [4] have also been explored for morph detection. Methods employing these classical features overwhelmingly pair them with SVM as the classifier. For the differential scenario, the difference vector between the two feature vectors is obtained first and both the feature vector and difference vector are sent to separate SVMs and score-level fusion is applied to arrive at the final decision score. BSIF, LBP, SIFT, SURF, and HOG have also been explored in a multi-algorithm fusion approach [39].

Deep learning-based techniques have also shown promising results in detecting morph images and have consistently outperformed most classical methods for morph attack detection [45] [44]. Complementary VGG-16 and AlexNet features have been concatenated and sent to a Probablistic Collaborate Representation Classifier (PCRC)

for classification [34]. A Siamese disentangling network that disentangles the landmarks and the appearance of the two images being compared has also been proposed [47]. Similarly, a Siamese network has also been explored in the image domain [46]. A double Siamese network framework that utilizes two Siamese networks and merges their output for its decision has also been proposed [2]. ArcFace embeddings trained on many generated morphs is used in [41]. An interesting differential approach is face demorphing [10] [11]. Essentially, the algorithm subtracts the trusted image from the potential morph and a low similarity score indicates the image is a morph. This method of demorphing has also been explored using a GAN framework [30].

Methods employed in forgery detection have also found relevance in morph detection, such as Photo Response Non-Uniformity (PRNU) analysis [42]. PRNU relies on using the residual noise artifacts in the morphing process to detect morphs. In [26], the authors design a face morphing detector by combining spatial and frequency feature descriptors from an image. Fuzzy LBP in color channels of HSV and YCbCR color spaces are investigated in [35]. Additionally, studying the residual noise computed on color channels using deep CNN-based denoising has also been presented for reliable face morphing detection [53] [52]. This paper aggregates several denoised instances of an image in the wavelet domain.

2.3 Morph Detection Algorithm Evaluation

As mentioned earlier, because of the lack of large, publicly available morph datasets, researchers often resort to generating their own synthetic morph images, typically employing large public face image databases to find look-alikes to morph together. These databases suffer from a number of limitations. Some possess obvious visual morphing artifacts. Others are limited in size and number of faces to select look-alikes to morph together. And yet others are created using manual morphing techniques that though

high quality and good enough to fool a network are limited in number and impossible to train a morph detector robustly with. Additionally, because of the distribution restrictions on these large face image databases used in the creation of morphs, the sharing capability of the resulting morph image datasets is also limited. This severely restricts any meaningful comparison of the state-of-the-art in this field. It is difficult for researchers to replicate and validate one another's results, not to mention difficult to benchmark their own results in relation to the literature. Hence, many of these implementations of these algorithms remain strictly experimental, meaning none are sufficiently robust to meet real world applicability standards. In this work, we try to emphasize cross-dataset analysis, testing on multiple publicly available morph datasets, so that our detector is adaptable to all scenarios.

To counteract this limitation, researchers can submit their algorithms to the United States National Institute of Standards and Technology (NIST) Face Recognition Vendor Test (FRVT) Morph Test [28] for objective evaluation. NIST tests each algorithm on several morph datasets of varying quality and method to evaluate an algorithm's general ability to detect morphs. The NIST FRVT Morph test is for both single and differential morph attack detection algorithms. Likely due to the limited dataset available to researchers, most algorithms submitted to NIST exhibit poor generalization, failing to satisfying detect many different kinds of morphs, each generated using a different method and of different quality. Slowly, we are seeing more algorithms submitted to NIST that are getting better at detecting a variety of morphs, mainly algorithms that utilize an underlying deep learning architecture, especially as training on morph datasets gets easier.

Chapter 3

Discriminative Wavelet Sub-band Selection

Many of the methods described in Chapter 2 can introduce artifacts in the resulting image, particularly in key areas where proper landmarking is difficult: the eyes and hair regions. Furthermore, the morphs generated typically lose the high frequency fine-grained information that the original contributing individual has such as appearance of pores, wrinkles, and acne scars [35]. This so-called smoothing effect means that morphs contain less high frequency information than morphs. We can differentiate morphs from bona fides in the high frequency spectrum. It is easiest to isolate these high frequency features in the wavelet domain which allows us to access local discriminative information.

Our morph attack detection framework focuses on applying undecimated 2D discrete wavelet transform (DWT) to the images and selecting only the most informative sub-bands for network training. We can localize these most discriminative sub-bands using KL Divergence [29]. KL divergence is a highly useful metric to measure the similarity (or dissimilarity) between two probability distributions. It comes in handy in our case to identify the sub-bands that show the most differentiation between a

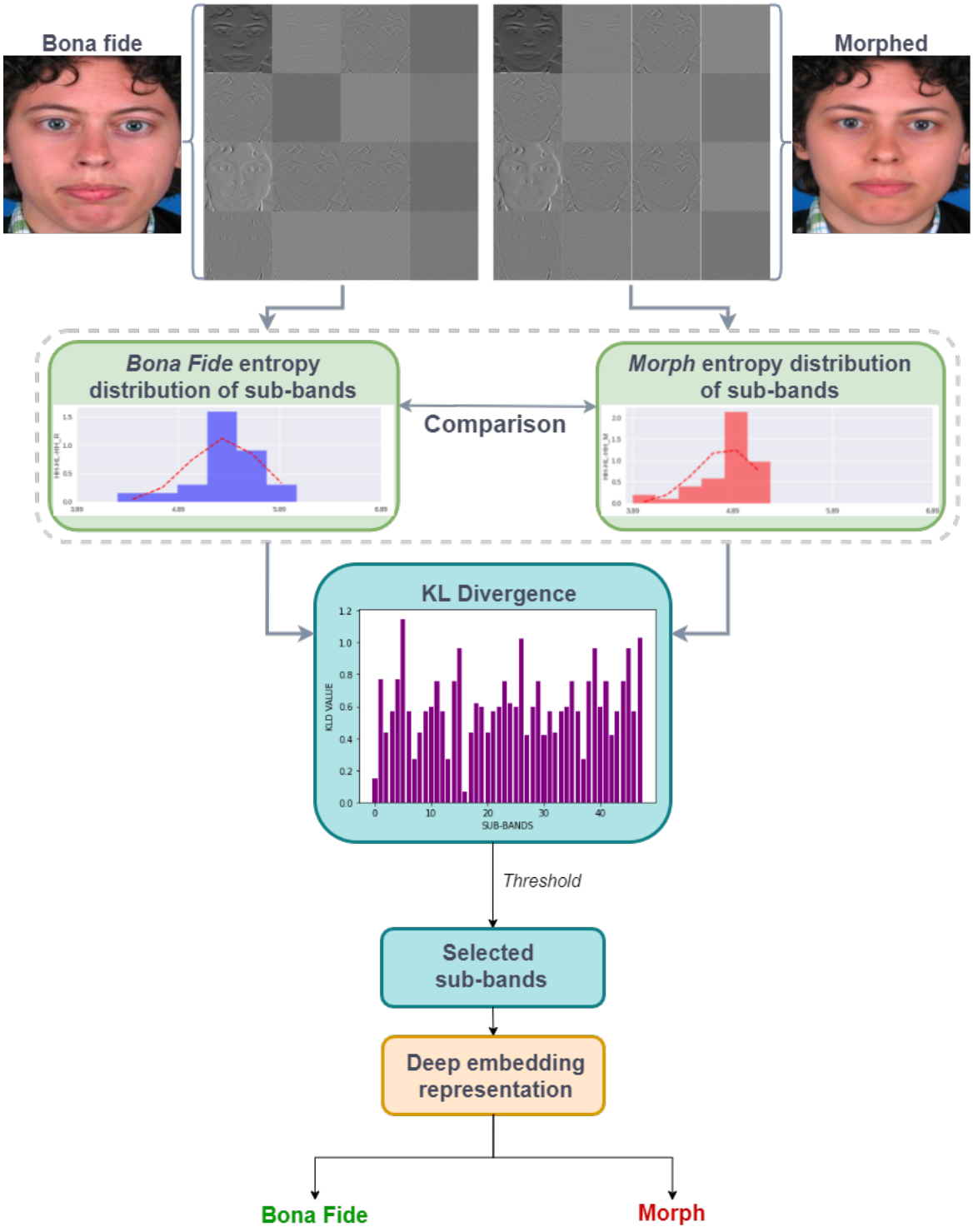

Figure 3.1: Essential overview of wavelet band selection algorithm.

morphed and a bona fide image. We obtain the KLD values for the probability distributions of each sub-band and instituting a threshold on the KLD values, resulting in 22 sub-bands. These 22 sub-bands, which have the highest KLD values, are then used to train our deep morph detectors: one for single morph detection and one for differential morph detection.

The following sections will briefly discuss wavelet transform and our sub-band selection method. The effectiveness of our sub-band selection method will also be evaluated through a data-driven ablation study.

3.1 2D Discrete Wavelet Transform

Wavelet transform is a useful and well-established tool. It has uses in many applications, especially in the image domain. A 2D wavelet transform decomposes an image in the spatial frequency domain and is considered especially powerful because of its ability to capture different frequencies at different resolutions. For most intents and purposes, it is a reversible process. 2D DWT extracts low pass and high pass image content in both horizontal and vertical directions. The low pass information yield approximation data whereas the high pass filter brings into focus the detail image content.

Because wavelet transform translates this data into individual sub-bands, we can separately examine the approximation and detail data in an image. For our problem, this means that wavelet transform allows us to effectively isolate the frequencies we wish to study and discard the ones we don't. We can accurately localize the sub-bands where the morph artifacts are most prominent and make those sub-bands the sole focus of our study. Furthermore, as most morphing artifacts reside in the high frequency spectrum and wavelet decomposition allows us to selectively choose the desired sub-bands only, using specific wavelet sub-bands in the place of the original

image is highly justified for our problem. Still, the performance comparison between the original RGB images and the selected wavelet sub-bands is substantiated in our experiments in 4.5.1.

After one level of wavelet decomposition, we obtain the Low-Low (LL), Low-High (LH), High-Low (HL), and High-High (HH) sub-bands. Further levels of decomposition occur by decomposing each of the above sub-bands separately. As mentioned earlier, since most morphing artifacts reside in the high frequency spectrum and since morphed images–like the LL sub-bands–could be considered close approximations of the original, we do not consider the LL sub-band informative for the morph attack detection problem and discard the LL sub-band entirely after the first level of decomposition. This is further substantiated by our exploration of the KLD values, where the high frequency sub-bands exhibited the most differentiation between the morphed and bona fide images, while the opposite could be said for the low frequency sub-bands. As can be seen in Figure 3.2, our chosen wavelet sub-bands have all been decomposed with the high pass filter. Thus, we only decompose the LH, HL, and HH sub-bands down to the third level, resulting in 48 mid- and high-frequency uniform wavelet sub-bands per image.

3.2 Sub-band Selection using KL Divergence

Even though the subtle discrepancies between a bona fide and a morph image can be localized with wavelet sub-bands, 48 mid- and high-frequency sub-bands does not completely isolate the sub-bands that actually contribute to the morph classification result. Thus, utilizing some sort of sub-band selection method can be a powerful and efficient way to find the most discriminating decision boundary between a morph and a bona fide image. To do this, we apply Shannon Entropy and Kullback-Liebler divergence to rank the sub-bands from most discriminative for morph detection to

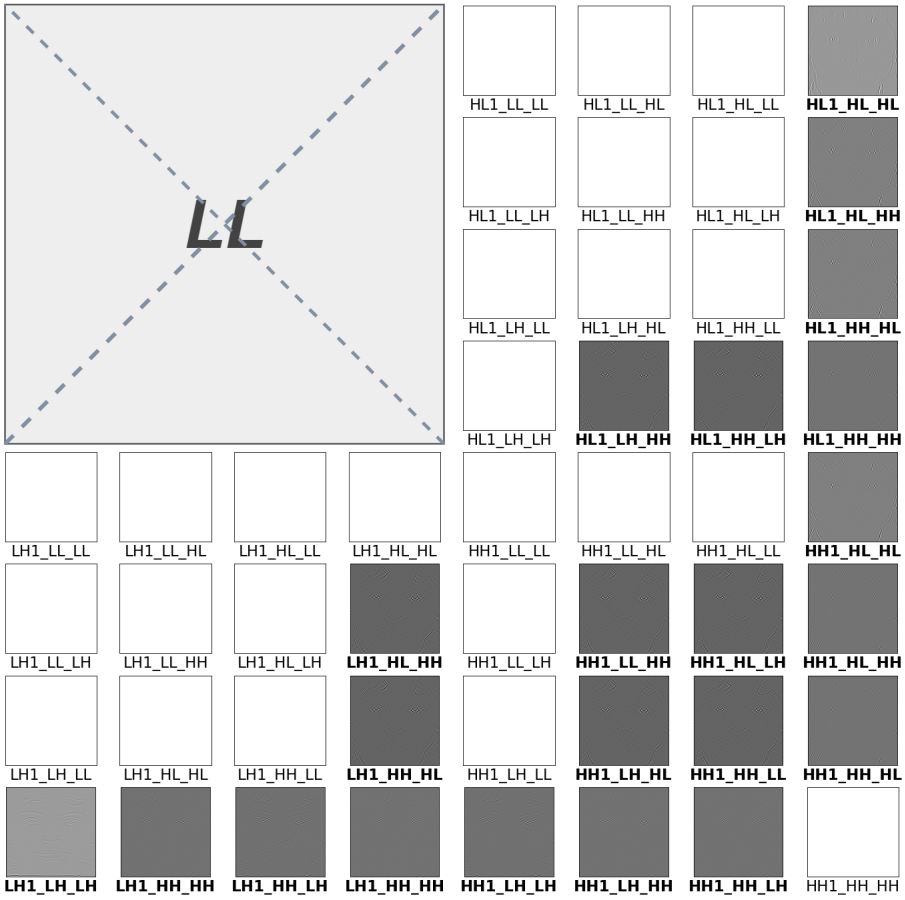

Figure 3.2: **Selected sub-bands.** The selected sub-bands are shown with regards to their location in wavelet decomposition. Most of the informative sub-bands chosen by KL divergence are those that have been filtered with the HH filter.

least discriminative.

Our band selection algorithm first computes the probability distributions for the morph sub-bands and the bona-fide sub-bands separately for all 48 sub-bands. For each sub-band, the difference between the bona fide entropy distribution and the morph entropy distribution is found to obtain a value for the sub-band. KL divergence of the entropy distributions is calculated. Because KLD measures the difference between the morph and bona fide entropy distributions, the higher the KLD value, the more discriminative the sub-band is. As such, we want to focus on the sub-bands that contain the most difference between a morphed and bona fide image. Therefore,

the sub-bands are sorted from highest KLD value to lowest and the sub-bands with the highest KLD value, corresponding to a larger divergence between a bona fide and a morph image, are chosen for network training. In total, 22 of the highest ranked sub-bands are used. The purpose of this method is to extract the sub-bands containing the discriminative features and use them to classify the given image as bona fide or morphed.

In detail, the method for finding the KLD values is as follows: after the entropy distributions of each sub-band are found, we find the histograms of entropy for all 48 sub-bands for both bona fide and morph images. Accordingly, 96 normal distributions (48 bona fide and 48 morph) are estimated using these histograms. Then, dissimilarity of the two probability distributions is calculated for all sub-bands and the KLD value is computed for each relative entropy distribution.

It is important to note here that the KLD values vary by morph dataset as each dataset is created using a different morphing technique. Although high frequency artifacts are a common result of every morphing technique, the variance of dataset still makes a difference in the KLD value. Therefore, we focus on selecting the sub-bands that are discriminative across the different morphing techniques. The KL divergence values of each of the training sets is normalized by removing the mean. The normalized KLD values are averaged for each sub-band for each of the three training sets. After sorting the normalized average KLD values from highest to lowest, we institute a threshold for selecting the sub-bands for network training. By choosing the sub-bands based on highest normalized average KLD value, we can find the sub-bands that are discriminative across morph techniques, not just for one specific morphing technique. In this way, we hope to avoid our model only learning one type of morphing technique. Ultimately, after instituting a threshold on the KLD values, we choose the top-22 sub-bands as our chosen input to our morph detector. The ablation study in the next section will delve more into why 22 of the top-most ranked sub-bands are

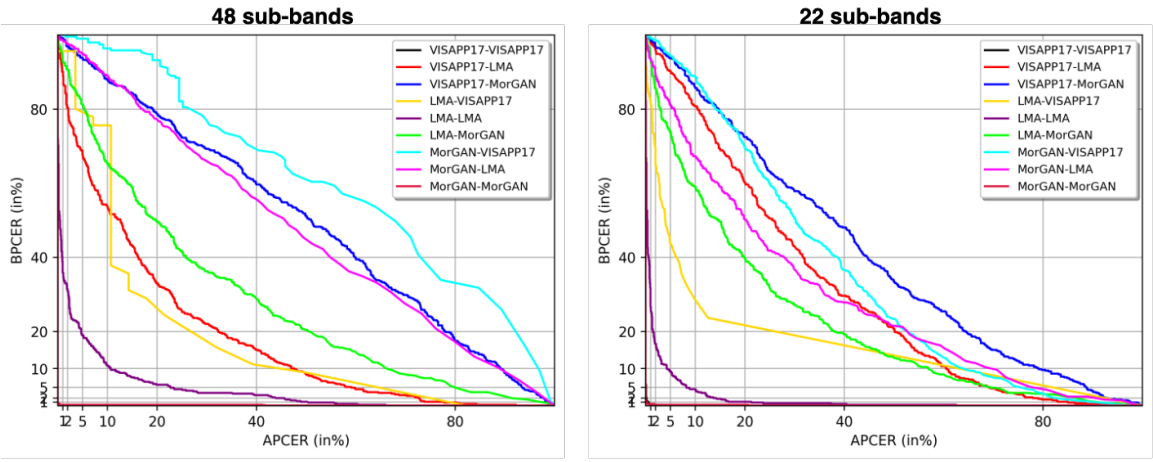

Figure 3.3: Performance comparison between training on 48 sub-bands (left) and 22 sub-bands (right) shown by their Detection Error Tradeoff (DET) curves.

optimal for morph attack detection.

3.3 Ablation Study on Sub-band Selection

After ranking the sub-bands by KLD value, the next step is to determine how many ranked sub-bands are actually needed for morph detection. Is the top-most discriminative sub-band enough for network convergence or do we need the top-20 (out of 48)? For this, we perform an ablation study on the number of ranked sub-bands using a data-driven approach. Different thresholds for different number of sub-bands are studied.

After ranking the sub-bands from highest KLD value to lowest, we select the subsets of the sub-bands according to the following criterion: the top-5 ranked sub-bands, the top-10 ranked sub-bands, the top-15 ranked sub-bands and so on to the full 48 sub-bands. We train several instances of the Inception Resnet v1 using the training portions of the MorGAN, LMA, and VISAPP morph datasets (described in 4.2) for each subset of number of ranked sub-bands. A DNN with input channel size of five is trained on the top-5 sub-bands and so forth. We assess the performance of

the networks using the validation portion of the dataset. Essentially, we evaluate the number of ranked sub-bands or the input channel size of the network for the ideal classification result. Given that each additional channel is a wavelet sub-band, this can also be seen as an ablation study on the amount of information required for a morph detector to have optimal results. Performance is mainly evaluated using the Area Under the Curve (AUC) metric.

We find the optimal performance to be of the 22 ranked sub-bands. After 22-sub-bands, any further addition does not increase the classification result meaningfully. In fact, 22 sub-bands is ideal in terms of network efficiency as well. Training on 48 sub-bands is very large and time consuming. Not to mention it can be costly to train on the machine, resource-wise. But sub-band selection lowers the dimensionality of the data. Cutting down on to top-22 sub-bands helps achieve network convergence quicker and reduces complexity. To highlight the merit of sub-band selection, we show that the network trained on 48 sub-bands exhibits inferior performance in comparison to the 22 discriminative sub-band scheme, especially in the case of cross-dataset performance, indicating that 22 sub-bands is better for generalization.

Figure 3.3 shows the Detection Error Tradeoff curves for morph detectors trained with 48 sub-bands and morph detectors trained with 22 sub-bands. Three detectors are trained on the MorGAN, LMA, and VISAPP datasets respectively and evaluated on each of the datasets as well for both 48 sub-bands and 22 sub-bands. The difference in training is most evident when it comes to cross-dataset training, where 22 sub-bands is provably shown to generalize better. The DET curves show that the cross-dataset performance for 22 sub-bands has smaller DET curves, indicating the morph detector is able to detect more unknown morph techniques than when trained on all 48 sub-bands.

3.4 Summary

In this chapter, we present an overview of our sub-band selection scheme. The merits of wavelet decomposition are discussed, particularly its usefulness for the morph detection problem. The details of our sub-band selection algorithm are explained and the effectiveness of sub-band selection is evaluated through a data-driven ablation study. In the following chapter, we explain how we used these discriminative wavelet features to train morph detectors to accurately detect morph imagery. The details of our detector and the subsequent experimental results are shared.

Chapter 4

Wavelet-Based Morph Detection

In this chapter, an overview of the wavelet-based deep morph detection algorithm that uses discriminative sub-bands as input is presented. The following sections will delve into detail of our experimental setup for both single and differential scenarios, introduce our datasets and dataset training protocol, and the metrics used to evaluate our experimental performance. Finally, the results of our experiments are tabulated and discussed.

4.1 Network Architecture

The applicability of discriminative wavelet sub-bands as features is explored both in the single and differential morph attack detection domains. The Inception Resnet v1 is employed as the base underlying architecture for both single and differential scenarios. In the single morph implementation, the Inception Resnet v1 is pretrained with weights from VGGFace2. Pretraining with face images helps prevent overfitting and reduces training time when we train on our significantly smaller morph dataset. We retrain the network then on the 22 selected discriminative wavelet sub-bands.

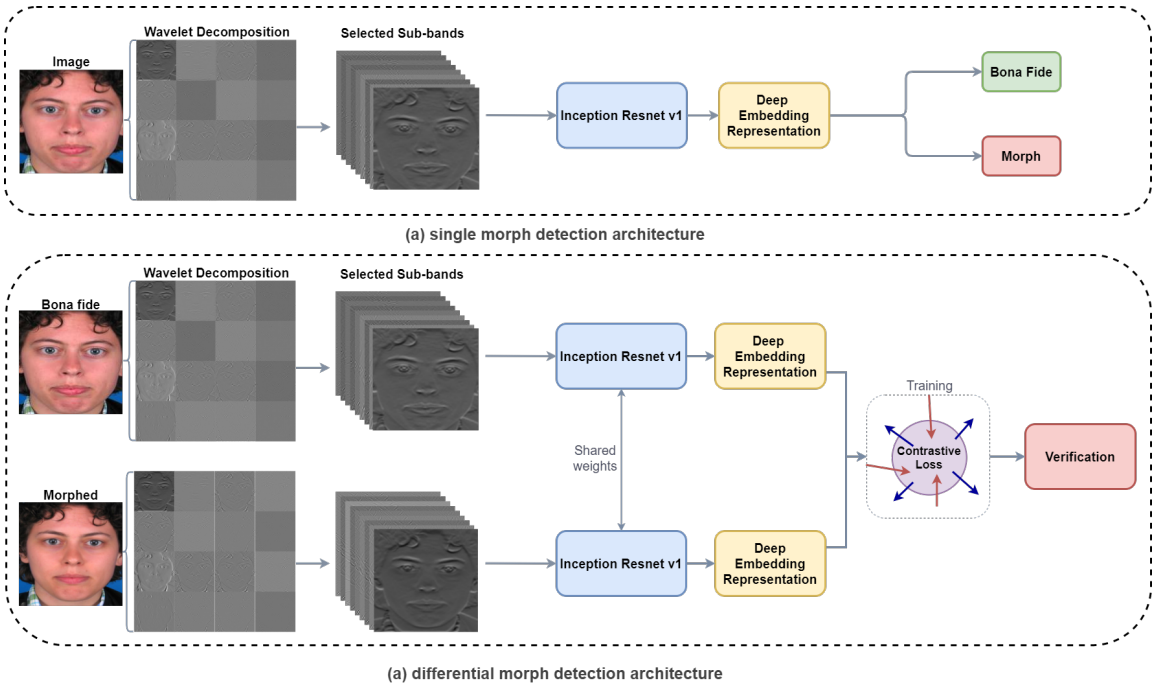

Figure 4.1: Single and differential network architecture.

4.1.1 Siamese Network

For differential morph attack detection, we employ a Siamese network architecture that uses the Inception Resnet v1 as the base network. The weights trained on the 22 sub-bands from the single morph implementation are used to initialize the Siamese implementation.

Siamese networks are ideal for this type of scenario as they are designed to facilitate comparison between two inputs. For this reason, they are popular to use in face verification tasks. Contrastive loss is used in conjunction with the Siamese network. Contrastive loss measures the similarities between the outputs of the two sub-nets of the Siamese network using Euclidean distance. In simplistic terms, contrastive loss is a distance-based loss function that forces similar samples together in the embedding space while pushing apart samples that are different even further. In essence, contrastive loss emphasizes the similarity between two samples of the same class and

exaggerates the difference between samples of different classes.

$$L_c = (1 - y_g)D(I_1, I_2)^2 + y_g \max(0, m - D(I_1, I_2))^2, \tag{4.1}$$

where I_1 and I_2 are the input face images, m is the margin or distance threshold to control the separation and y_g is the ground truth label for a given pair of training images and $D(I_1, I_2)$ is the L_2 distance between the feature vectors:

$$D(I_1, I_2) = ||\phi(I_1) - \phi(I_2)||_2. \tag{4.2}$$

Here, $\phi(.)$ represents a non-linear deep network mapping image into a vector representation in the embedding space. According to the loss function defined above, y_g is 0 for genuine image pairs and y_g is 1 for imposter (morph) pairs.

4.2 Datasets and Pre-processing

In our training, we combine three different datasets: VISAPP [23], MorGAN [5] [8], and LMA [5] that each utilize a different morphing technique to maximize network performance and generalization. We combine the datasets as they are relatively small and training on one will no doubt lead to overfitting. Furthermore, in a real world scenario, the morphing technique a criminal attacker would employ is unknown. Therefore, it makes sense to train a network on a variety of morph algorithms so that it learns to generalize to a morph image. Additionally, we also test on different so-called unseen datasets to obtain an accurate representation of how the network performs on morph techniques it has not seen before. For this cross-dataset evaluation, we employ AMSL Face Morph dataset [27]. This dataset is used strictly to evaluate cross dataset performance.

VISAPP [23] is a collection of splicing morphs generated using face images from

Table 4.1: Datasets used in training.

Dataset	Morph Method	Training		Testing	
		Bona Fide	Morph	Bona Fide	Morph
MorGAN	GAN-based	1495	500	1499	500
LMA	Standard LMA	1495	500	1499	500
VISAPP	Splicing LMA	78	198	78	116

the Utrecht dataset [50]. The splicing technique applies landmark morphing only to the inner region of the face image and then splices the resulting morph into the face region of one of the contributing individuals. In this way, the ghosting that typically occurs around the hair region and edges of the face that occur in Figure 2.1 are avoided. The images in this dataset are 900×1200 in size. We use a subset of this dataset, namely Visapp-Splicing Selected, which consists of 183 high quality morphs morphs generated as our training set. Our VISAPP dataset has 314 images total, 131 bona fide (neutral and smiling) images and 183 morphs.

MorGAN [5] was generated using a GAN model to synthesize morph images. A subset of the CelebA dataset [21] of manually selected full frontal images is used to create the morphs. The authors of MorGAN also generated the LMA dataset [5] that consists of morphs created using the standard landmark manipulation technique [25]. Both LMA and MorGAN use the same contributing individuals. Both MorGAN and LMA have 1,500 bona fide reference images from which 1000 morphs are created. In addition, the dataset includes 1,500 bona fide probe images for comparison. MorGAN morph images are of size 64×64 and LMA morph images are originally of size 128×128. Although a lower resolution than what is ICAO compliant, training on a variety of techniques and resolutions could help identify morphs at all levels, especially when the bona fide images are of a lower quality as well.

Finally, a publicly available morph dataset is used to measure cross dataset performance, namely AMSL. This dataset is strictly used as a test dataset. The Advanced Media Security Lab (AMSL) Face Morph Dataset is created using the Combined

Morph tool [27]. AMSL consists of 204 bona fide (neutral and smiling) images and 2175 morph images. All face images from AMSL are ICAO compliant and were additionally compressed using the JPEG2000 algorithm according to the morphing process described in [27]. This is important to note as the intention of the compression was to erase or hide most of the artifacts that are introduced in morphing.

4.2.1 Preprocessing

All images are pre-processing according to the MTCNN framework [55], which utilizes three convolutional neural networks to predict face and landmark locations. Faces are detected, aligned, and resized to 160×160 for network input as is standard for FaceNet. All training portions are further augmented with horizontal flips. Images are then converted to grayscale before wavelet decomposition is applied. The 22 informative sub-bands are selected as input per the scheme describe in 3.2 and the remaining sub-bands discarded. Thus, for each image, the network input is $22 \times 160 \times 160$.

50% of the subjects are considered for training and 50% of the subjects is used for testing. Additionally, 15% of the test set is selected during model optimization as the validation set. By design, the train-test split is disjoint, with no overlapping morphs or contributing bona fides to morphs. This is done to ensure an accurate representation of performance. Furthermore, to offset the class imbalance between the bona fide and morph classes, the batch generator is weighted to ensure that the network sees an equal number of morph and bona fide samples at every iteration. Additionally, for the differential implementation, the images are paired off into morph pairs (one bona fide image and one morph image) and genuine pairs (two bona fide images of the same individual).

4.3 Experimental Setup

Both single and differential implementations employ a similar experimental setup, albeit for some slight differences on account of their purpose. For both implementations, the batch size is 32 and Adam is chosen as the optimizer with an initial learning rate 0.0001. Validation loss is monitored continuously as a measure of how well the network trains and at every newly-achieved low, the best weights are saved. If the validation loss plateaus for 35 epochs, the learning rate is divided by ten, the best weights are re-loaded and training starts again from the last saved lowest validation loss down to a learning rate of 1e-06. After that, early stopping is implemented if the network plateaus. For the Siamese network, the margin m for contrastive loss is set to 1. Training is accelerated by the use of three Titan X (PASCAL) 12 GB GPUs.

We train our morph detectors using the combined training portions of VISAPP, MorGAN, and LMA. We called this combined dataset the universal dataset. As mentioned previously, several morph attack detection algorithms suffer from overfitting on a small morph dataset, especially if the dataset is only created using one morph generation technique. Our purpose in combining the training sets is to train a network that generalizes to morph artifacts in the hopes that the performance will remain robust when tested on a different morph dataset the detector has not seen before.

4.4 Metrics

We use the standard measures for morph attack detection to evaluate network performance, namely the APCER, BPCER, and EER rates. APCER stands for Attack Presentation Classification Error Rate or the rate at which morphs are erroneously classified as bona fides. This is the rate at which morphs pass undetected. Conversely, BPCER, which stands for Bonafide Presentation Classification Error Rate is the rate at which bona fide images are incorrectly classified as morphs. THE APCER

and BPCER rates correspond to Type 1 and Type 2 error, or the false positive and negative rates. BPCER can also be considered the false alarm rate. In a real world application, a false alarm is the rate at which an individual is inconvenienced and is considered expensive in terms of resources required and as such this rate is controlled according to FRONTEX guidelines [13]. Hence, artificially regulating the BPCER rate by restricted it to fixed thresholds is highly recommended. When the BPCER rate is artificially controlled, the decision boundary moves to allow more morphs to be accepted as genuine individuals. However, even with a more rigidly controlled BPCER rate and resulting higher morph miss rate, the percentage of morphs detected is still better than no morph detection algorithm at all [28]. Equal Error Rate (EER) is the point where BPCER and APCER are equal. Additionally, the rates at a controlled threshold are reported for morph detection, typically to control the false alarm rate. APCER5 is the APCER rate where BPCER is 5%. Similarly, APCER10 is the rate when BPCER is 10%. These rates are plotted in a Detection Error Tradeoff (DET) curve.

4.5 Experiments Overview

Next, there will be a discussion of experiments and their subsequent results. We assess the performance of using discriminative wavelet sub-bands as features using the test sets of VISAPP, MorGAN, and LMA. We also use the so-called universal test set, which comprises of all three individual test sets, and acts as an average indicator of how the universally trained networks perform. In addition, we test on AMSL as our unseen dataset. In this way, we fully evaluate both same dataset and cross dataset performance.

In the following sections, we will briefly discuss comparison of our wavelet based morph detector with a variety of comparative baselines, mostly in the image domain

Table 4.2: Single morph attack performance of the proposed framework and baselines.

Testing	Method	APCER@BPCER		BPCER@APCER		D-EER
		5%	10%	5%	10%	%
MorGAN	BW images	23.5	9.07	12.6	9.0	9.0
	RGB images	5.7	1.55	4.70	2.74	4.69
	LL-removed BW images	18.68	6.29	9.93	6.28	7.5
	LL-removed RGB images	2.55	1.14	1.98	0.99	3.58
	BW-22 wavelets	1.20	0.60	0.73	0.20	2.47
	RGB-66 wavelets	**1.0**	**0.86**	**0.66**	**0.20**	**2.33**
LMA	BW images	25.28	15.28	24.20	15.1	12.6
	RGB images	12.79	6.89	13.0	6.16	7.8
	LL-removed BW images	41.1	28.8	35.4	24.2	16.6
	LL-removed RGB images	9.17	5.29	8.5	4.36	6.14
	BW-22 wavelets	**10.8**	**5.0**	**11.2**	**6.0**	**7.47**
	RGB-66 wavelets	8.6	5.39	12.5	3.8	7.47
VISAPP	BW images	0.0	0.0	0.0	0.0	0.0
	RGB images	0.0	0.0	0.0	0.0	0.0
	LL-removed BW images	0.0	0.0	0.0	0.0	0.0
	LL-removed RGB images	0.0	0.0	0.0	0.0	0.0
	BW-22 wavelets	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
	RGB-66 wavelets	0.0	0.0	0.0	0.0	0.0
UNIVERSAL	BW images	18.50	10.60	14.40	9.38	9.55
	RGB images	6.75	2.92	6.08	2.98	5.57
	LL-removed BW images	30.4	13.62	19.83	11.40	10.59
	LL-removed RGB images	5.8	1.89	4.5	2.98	4.52
	BW-22 wavelets	**1.26**	**0.65**	**1.77**	**0.65**	**2.70**
	RGB-66 wavelets	3.3	1.97	2.84	0.97	4.48

to show its effectiveness. We also compare to several state of the art techniques, both classical feature-based and deep learning based morph attack detection techniques.

4.5.1 Wavelet-based Morph Detection Comparison

As wavelet transform is a reversible process, there are questions as to whether it really is better to use wavelet sub-bands for morph detection over the original RGB image, which is considered the standard input for CNN based classifiers. Therefore, we compare our wavelet-based morph detector with a standard RGB image-based morph detector. We conduct these experiments for both single and differential morph detection scenarios. The purpose of this exercise is to investigate whether discriminative wavelet sub-bands really do add more information for the classification result or if some other method in the image domain is actually better. For this, we choose several

Table 4.3: Differential Performance of the proposed framework and baselines. With the exception of RGB-66 testing on MorGAN, BW-22 exhibits superior performance.

Testing	Method	APCER@BPCER		BPCER@APCER		D-EER
		5%	10%	5%	10%	%
MorGAN	BW images	7.88	6.17	13.1	3.1	5.57
	RGB images	4.5	3.3	3.22	1.74	4.17
	LL-removed BW images	5.5	3.14	4.5	3.28	5.53
	LL-removed RGB images	3.66	2.98	1.58	0.79	3.55
	BW-22 wavelets	3.71	1.85	3.06	0.26	3.89
	RGB-66 wavelets	**0.86**	**0.0**	**0.37**	**0.37**	**1.62**
LMA	BW images	22.7	14.3	36.5	15.1	11.6
	RGB images	11.1	6.68	12.2	5.62	8.8
	LL-removed BW images	25.9	14.4	19.0	11.5	11.5
	LL-removed RGB images	15.75	7.4	12	6.48	8.06
	BW-22 wavelets	**4.95**	**2.67**	**4.38**	**1.46**	**4.52**
	RGB-66 wavelets	10.53	5.39	9.44	4.72	7.36
VISAPP	BW images	5.97	0.0	0.0	0.0	3.17
	RGB images	1.32	0.08	0.0	0.0	0.0
	LL-removed BW images	1.57	0.08	5.63	4.22	0.0
	LL-removed RGB images	2.98	0.8	0.0	0.0	3.25
	BW-22 wavelets	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
	RGB-66 wavelets	0.0	0.0	0.0	0.0	0.0
UNIVERSAL	BW images	15.0	8.95	14.4	7.5	8.53
	RGB images	6.65	4.01	5.22	2.5	5.63
	LL-removed BW images	19.1	6.74	10.872	7.78	8.45
	LL-removed RGB images	10.9	3.53	5.52	4.56	5.52
	BW-22 wavelets	**3.25**	**1.69**	**3.01**	**0.65**	**3.93**
	RGB-66 wavelets	6.4	2.67	5.15	2.57	5.15

baselines to compare. For each baseline, we train a separate network using the same train-test split and experimental setup for maximum comparison.

Wavelet transform is traditionally in grayscale yet the original images are in RGB. Noting this, we also wish to investigate the importance of color information for morph detection. Thus, we also obtain the wavelets for the color image by obtaining the 48 sub-bands for each color channel separately, resulting in 144 sub-bands total after the LL sub-band is removed from each color channel and before the sub-band selection is applied. We select 22 sub-bands from each color channel, for a total of 66 sub-bands, which we call this baseline RGB-66 in correlation with BW-22, which are our original 22 wavelet sub-bands. The 22 sub-bands chosen for each color channel are identical to the 22 sub-bands of BW-22. The experimental setup is also identical to our wavelet sub-bands.

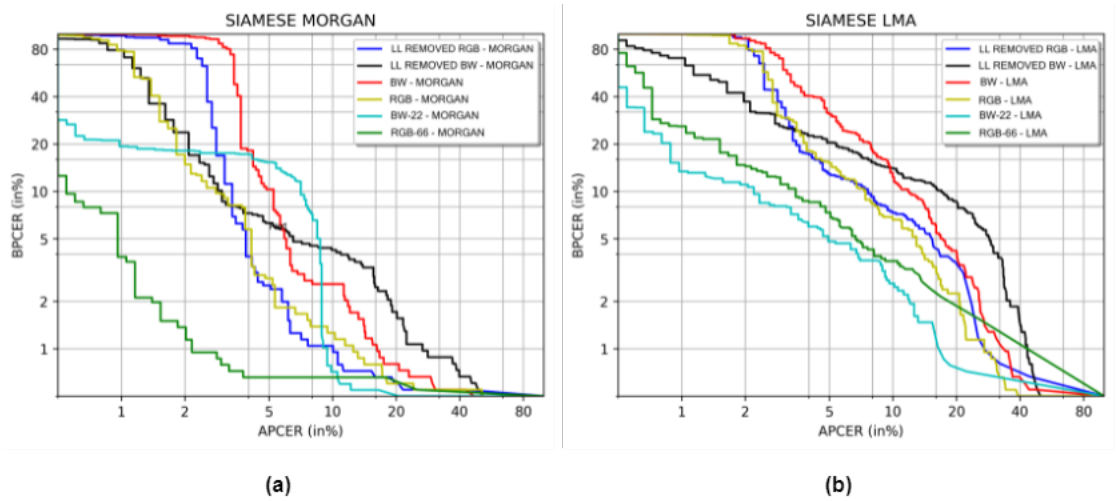

Figure 4.2: DET curves for all protocols, tested on the (a) MorGAN and (b) LMA test sets.

The first image-based baseline we consider is the original RGB image. The second is the grayscale image, which corresponds to the image right before wavelet decomposition is applied. In a similar fashion, we also investigate the images where LL is removed. To mirror the removal of the approximation data in the image domain and take advantage of the reconstructable feature of wavelet transform, the LL is removed by decomposing the image in the wavelet domain, removing the LL sub-band, and then performing Inverse Wavelet Transform (IWT) to get the LL-removed image. We apply the LL removal process to both the grayscale and color images, obtaining the baselines LL-removed BW and LL-removed RGB. This is roughly equivalent to 48 sub-bands of BW-22 and the 144 sub-bands of RGB-66 before sub-band selection occurs. We then train a separate Inception Resnet v1 architecture using all of the above baselines using the same training protocol and compare the test results. We do this for both single morph detector and the Siamese differential morph detector for each of the above scenarios: BW-22, RGB-66, RGB images, BW images, LL-RGB, and LL-BW.

The results show that wavelet-based deep morph detection does exhibit supe-

rior performance, significantly outperforming the other baselines. This can be seen clearly in Figure 4.2. We note that for almost all scenarios, RGB performs better than grayscale images, suggesting that there is utility in the color information for morph detection. Especially for the MorGAN dataset, the performance really improves with the use of color, indicating synthesized GAN-generated morphs have more color information than landmark based methods such as LMA or VISAPP. This is more clearly noted with the single morph detection results tabulated in Table 4.2 where there is a stark difference between the RGB baselines and the grayscale baselines, suggesting that color information is important for the classification result. This difference is more subtle in the differential domain, suggesting that the comparison of the two images dulls the need for color information. This difference overall is muted, however, in the wavelet domain, where the original BW-22 selected wavelets outperform the RGB-66 wavelets.

Interestingly enough, for single morph detection, the performance increases when the LL sub-band is removed from the RGB images, substantiating our claim that it is more useful to remove the LL sub-band altogether. However, for the differential implementation, the opposite can be seen, where the original RGB images perform better. It seems that when the detector only sees one image at a time, it is useful to remove the LL sub-band. Note that overall for each baseline if the single morph detection results are compared with the Siamese morph detection results in Table 4.3, the Siamese version of the morph detector performs better. This difference can be most clearly seen in the comparison of the EER rates. This is on par with what is assumed about differential morph detection. Because differential morph detection has the additional image and thus additional information to compare the potential morph with, generally speaking the performance is also higher.

It is also notable that VISAPP's performance is consistently high for each scenario, regardless of single or differential. This could be due to the small nature of the dataset

that the networks are able to easily fit to it. Furthermore, the performance of the LMA morphs is significantly lower for each baseline in comparison to MorGAN and VISAPP. This is also consistent with what is in the literature. Landmark-manipulated morphs are historically more difficult for deep morph detectors to detect than GAN-generated morphs [51].

4.5.2 State of the Art

In addition to evaluating the strength of wavelets in comparison to baselines in the image domain, we also compare the strength of wavelets to other state of the art techniques mentioned in the literature. Here, it is important to reiterate that due to the private nature of most morph image datasets, direct comparison is very limited. Therefore, the results shown below follow as much of the methodology as mentioned in the paper but they are trained with our datasets. The techniques we explore are both classical hand-crafted techniques and deep learning methods. Due to the hybrid nature of our method where classical wavelets meet deep learning and we use KLD-based wavelet feature extraction, evaluating performance in comparison to both other classical methods and deep learning methods is highly justified.

As far as classical hand-crafted techniques go, we employ BSIF, LBP, SURF, and SIFT to compare our network with. All four of these techniques have enjoyed popularity in the literature for morph detection. BSIF and LBP are feature descriptors. For BSIF, we choose the 8-bit BSIF features vectors that are constructed using 3×3 filters. The LBP feature descriptors are extracted according to patches of 3×3. The resulting feature vectors, normalized histograms of 256, are the values of the LBP binary code.

SURF and SIFT are keypoint descriptors. Keypoint descriptors, especially, are useful for morph detection because the morph should produce less keypoints than the bona fide images on the account of the smoothing effect of the morph operation. Each

Table 4.4: Single Performance Comparison of Proposed Framework. All algorithms trained with the Universal dataset.

Testing	Method	APCER@BPCER		D-EER
		5%	10%	%
MorGAN	SURF	76.0	70.0	38.67
	SIFT	93.3	88.6	47.6
	LBP	90.13	82.2	41.6
	BSIF	91.3	84.78	50.0
	BW-22 wavelets	**1.20**	**0.73**	**2.47**
LMA	SURF	74.50	62.70	33.40
	SIFT	67.70	50.00	28.31
	LBP	61.50	51.20	29.0
	BSIF	70.42	57.60	30.0
	BW-22 wavelets	**10.8**	**5.83**	**7.47**
VISAPP	SURF	79.4	70.1	31.0
	SIFT	83.2	70.9	27.0
	LBP	72.5	59.5	37.67
	BSIF	67.2	59.0	35.0
	BW-22 wavelets	**0.0**	**0.0**	**0.0**
AMSL	SURF	79.89	70.65	34.0
	SIFT	79.89	66.3	34.7
	LBP	94.02	85.8	49.0
	BSIF	91.3	84.78	50.0
	BW-22 wavelets	**33.82**	**26.96**	**19.54**

of the above four frameworks is paired with a Support Vector Machine (SVM) with an rbf kernel as a classifier. The SVM is chosen as the classifier as this is the most commonly used classifier in conjunction with these techniques used in the literature. As we cannot compare directly, it is important to facilitate comparison by following as much of the methodology employed by others.

For the differential implementation of these classical models, we follow [40] where the feature vector of the potential morph image is subtracted from the feature vector of the trusted image. This difference vector is then fed into an SVM classifier for differential morph detection. To compare our detector to deep learning models, we investigate FaceNet [43] performance with standard image input for the differential scenario.

The single morph detector outperforms the other classical techniques. When training on all three datasets or the universal dataset, this becomes even clearer as can be seen in 4.4. We also test our universal single morph detector on the AMSL dataset

Table 4.5: Differential Performance Comparison of Proposed Framework. All algorithms trained with the Universal dataset.

Testing	Method	APCER@BPCER		D-EER
		5%	10%	%
MorGAN	SURF	86.8	70.11	46.1
	SIFT	57.6	47.7	27.3
	LBP	90.13	82.2	41.6
	BSIF	86.8	71.6	31.7
	FaceNet	36.80	31.15	22.25
	BW-22 wavelets	**0.86**	**0.0**	**1.62**
LMA	SURF	81.1	63.69	51.1
	SIFT	63.2	55.8	36.7
	LBP	91.1	83.4	40.5
	BSIF	86.5	75.0	36.4
	FaceNet	43.70	40.90	30.35
	BW-22 wavelets	**4.95**	**2.67**	**4.52**
VISAPP	SURF	94.1	90.3	47.8
	SIFT	91.1	84.7	52.2
	LBP	31.1	19.5	16.0
	BSIF	30.6	22.73	16.4
	FaceNet	25.0	15.8	15.5
	BW-22 wavelets	**0.0**	**0.0**	**0.0**
AMSL	SURF	96.7	91.3	53.0
	SIFT	94.65	84.9	38.0
	LBP	91.0	72.9	43.0
	BSIF	91.0	82.0	41.3
	FaceNet	38.6	31.35	19.86
	BW-22 wavelets	**33.78**	**23.61**	**16.4**

in an effort to objectively evaluate the benefits of cross dataset performance.

For the Siamese morph detector, results can be compared in Table 4.5. Our discriminative wavelet sub-bands consistently outperform the other methods, including FaceNet. This is underlined by the performance of the detector on AMSL, which is used to evaluate cross dataset performance and is used for testing only. Our Siamese morph detector performs better than FaceNet. For the differential scenario, the framework achieves an EER of 3.93% on the universal test set, significantly better than the other baselines. Furthermore, the framework performs well on an unseen morph dataset, AMSL, that uses a different morphing technique than our training set, achieving an EER of 16.4%.

4.6 Summary

In the last two chapters, using discriminative wavelet sub-bands as features for morph attack detection is discussed in detail. The step-by-step process of obtaining the desired sub-bands is explained and the sub-band selection process is evaluated using a data-driven ablation study on the number of ranked sub-bands. Furthermore, the effectiveness of sub-band selection is substantiated. The performance gains of using selective wavelet sub-bands over the original RGB image is further explored in detail. Our method is also compared to other classical feature-based methods used in the literature. Overall, using discriminative wavelet features has shown to be a powerful technique, even when training on a smaller dataset. We have demonstrated that a wavelet-based morph detector can accurately identify morph imagery. In the following chapter, after realizing some of the weaknesses of training on a smaller dataset, we will explore using twins to train a morph detector accurately and robustly.

Chapter 5

Twins for Morph Attack Detection

Following our experiments in the wavelet domain, it became clear that to obtain optimal results, training on a large dataset of high quality morph images was needed. This is an ongoing problem faced by members of the research community. Luckily, as a WVU research student, we have access to the WVU Twins Day Dataset [49], a large-scale high-resolution dataset of twins collected on the annual Twins Day at Twinsburg, Ohio over 2010 to 2019. The twins are used to generate large scale morph databases using different automatic generation techniques.

Therefore, in this chapter, we explore the impact of training on large-scale high quality morph images generated from twins on an Inception Resnet v1. We employ several datasets generated by fellow research students from the Twins Day dataset in our study: Twins Landmark, Twins StyleGAN, and Twins Landmark Perturbed. We also investigate the effect of training on compressed versions of the Twins Landmark images. In the end, we train a robust morph detector and submit our detector to NIST FRVT Morph test for evaluation.

In the following sections, I will go over briefly the difficulty of finding suitable look-alikes, the datasets employed in this study, the experiments conducted, and discuss the results from NIST regarding our twin-morph-trained network.

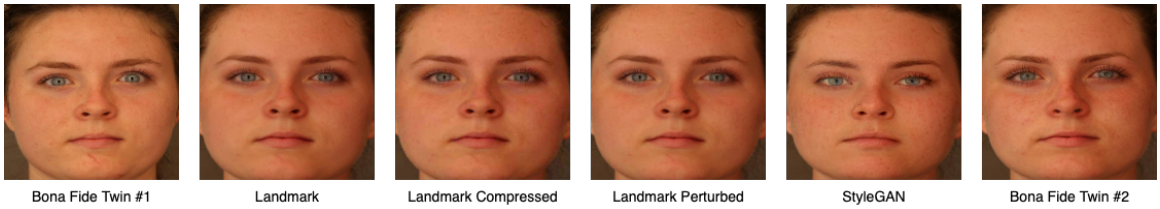

Figure 5.1: Bona Fide Twins (far left and far right) with morphs created using different techniques with the Twins Day Dataset.

5.1 The Difficulty of Finding Look-Alikes

Most current morph datasets are lax with which individuals they morph together. This is partly due to the fact that there is no standard procedure to identify and measure similarity. It is also due to the small size of the source public face image datasets available. Even if the source face image dataset is substantially large, there is no guarantee there will be a look-alike for each face in the dataset. Simply put, there isn't a big enough pool of faces to begin with. Researchers often resort to morphing individuals together that are only similar on the surface level of sharing gender and ethnicity. This is a very superficial way to morph people together and often results in lower-quality morphs. This can be a passable way to create morphs to train a single morph detector since the detector only sees the potential morphed image. However, for differential morph detection and for inspection by a human eye, these kinds of morphs are often very obvious to detect as the difference between the bona fide and the morph can sometimes be still quite large.

The problem of finding suitable look-alikes has been studied to some extent [7] [37]. FaceNet and SSIM have been proposed as ways to find people who look similar. Although these methods have shown some promise in assessing similarity, there is no standard method found yet agreed by the research community. In response to this problem, we propose using twins as our look-alikes to morph together.

Identical twins are already classified as a known and important security challenge.

Table 5.1: Morph datasets created using twins.

Dataset	Training		Testing	
	Bona Fide	Morph	Bona Fide	Morph
Twins Landmark	4526	6495	860	1240
Twins StyleGAN	4526	5506	1488	450
Twins Compressed	4526	6495	860	1240
Twins Perturbed	4526	6495	860	1240

Like morphs, differentiating between twins is also difficult for humans. For the purpose of training a robust morph detector, a morph dataset based on twins is ideal. There is no manual and time-consuming process of finding suitable look-alikes. Twins already share a similar face structure and already fool recognition systems. A twin morph lacks ghosting and obvious differences between the morph and bona fide image, hence it can fool a human border inspector. By morphing twins, we can generate high quality morph images that contain both the ideal visual similarity as well as the underlying morphing artifacts so our network can be trained on both distinguishing one twin from another and also distinguishing a morphed image from a genuine image using image artifacts that result from the morph process. This way we can exploit twin data characteristics and use a large scale dataset with variance.

In a real world scenario, where there is a very real risk of detection and subsequent criminal punishment, a criminal would first ensure they find someone with similar features to avoid detection. Hence, it is important for morph detection algorithms to be trained on high quality look-alikes. This is why we use morphs generated from the Twins Day dataset. Not only will the resulting morph contain artifacts and residual image clues that indicate the image was tampered with but the resulting face will be incredible similar to the original two people. Both the visual and inherent morph qualities are satisfied.

5.2 Datasets

The Twins Day dataset contains images from twins taken over 2010 to 2019 at the annual Twins Day Festival in Twinsburg, Ohio. For some twins, there are images taken over years. For each twin, there are multiple high resolution images in multiple poses, sometimes taken over multiple days. This is important, especially for the differential scenario, as we have multiple bona fide probes in addition to the twin photo used for morphing. There are also neutral frontal face images with a neutral background that are ideal for mimicking a passport photo. Some twins are excluded (i.e. fraternal twins of one female, one male) and after some dataset cleaning, over 1250 pairs of twins are used to create the morphs.

The first twins morph database is Twins Landmark, which was generated using the standard landmark manipulation method [31]. For each twin pair, there are three morphs with a corresponding alpha value of 0.3, 0.4, and 0.5 respectively. Following this, there is a morph database of twins generated using a GAN framework: Twins StyleGAN. This contains one morph image per twin pair. With Twins Landmark and Twins StyleGAN, we broadly cover the two common automated morphing techniques employed in the literature.

Following these two methods, we further augment the Twins Landmark morphs with two post-processing techniques: compression and adversarial perturbation. First, we compress the Twins Landmark morphs using JPEG compression to create Twins Landmark Compressed. Compression was used to help robustness against lower resolution images. Compression also helps mimic the passport enrollment process wherein a submitted passport photo is compressed before being put in the issued passport. Finally, we also investigate the performance after perturbing the Twins Landmark images. Adversarial perturbation is added to each morph such that a morph detector is fooled to classify the morph images as bona fide but not so much so that the perturbation is visually distinguishable.

Table 5.2: Performance of our twins-trained morph detector on some select databases.

Testing	APCER @ BPCER			D-EER
	1%	5%	10%	%
Twins Landmark	10.17	3.4	2.43	3.64
FERET OpenCV	42.91	31.7	19.86	13.0
FERET StyleGAN	43.02	26.99	17.49	15.09
FERET Facemorpher	37.21	33.95	18.08	13.3

The datasets are further augmented using horizontal flips. In addition, for each twin, there is an additional ground truth image that serves as our Bona Fide Probe. After all augmentation, our total training set, consisting of morphs from all twins morph datasets (Landmark, StyleGAN, Compressed, Perturbed) is 44,152 morph images and 20,751 real images. All images are pre-processed using the MTCNN framework and resized to 512×512 for network training. This is notably larger than the 160×160 used by the wavelet detector. We are aiming for a high quality, robust morph detector and this higher resolution is more closer to passport-sized images. Furthermore, some research has shown that training on a higher resolution helps with detecting artifacts in the image [28].

5.3 Experiments

We train several versions of our twins morph detector before we decide on the one to submit to NIST. Our motivation was to create a detector that was robust to most morphing algorithms. Since NIST tests on sequestered datasets on which there is limited information on the morphing techniques, it was important for our detector to also be robust to the unknown. This way our network generalizes and doesn't overfit to only one specific morph technique. This is important as in the real world application, the morphing technique a criminal would use would be unknown to the morph detector.

We employ several test sets to benchmark performance. We use AMSL as well

as the morphs generated by [38] using the FRLL and FERET databases. The FRLL morphs are generated using four methods: FRLL OpenCV, FRLL StyleGAN, FRLL Webmorph, and FRLL Facemorpher. The FERET morphs consist of FERET OpenCV, FERET StyleGAN, and FERET Facemorpher. These are all datasets used in our experiments for testing only to evaluate cross dataset performance, much in the same way NIST will evaluate our final twins morph detector.

As our purpose is to create a robust morph-technique-resistant detector, we employ cross dataset, cross morph technique training, at each level, adding more morph techniques to the training set. First, we train only on Twins Landmark and evaluate performance. Then we investigate training with compressed versions of Twins Landmark alongside the original Twins Landmark dataset. The purpose of this exercise was to see if compression could help with lower resolution images and act as a substitute to print-and-scanned images, which was a database we did not have. Finally, we train on all four datasets: Twins Landmark, Twins StyleGAN, Twins Landmark Compressed, and Twins Landmark Perturbed. Figure 5.2 shows the performance of our all-twins-morph-trained detector on Twins Landmark (same-dataset performance) as well as on the OpenCV, StyleGAN and Facemorpher morphs created from the FERET database (cross-dataset performance).

5.4 NIST Submission

Following our experiments, we train our final twins morph detector on all our datasets and submit it to NIST for objective evaluation. As mentioned earlier, the NIST FRVT Morph test is ongoing independent testing of morph attack detectors on a variety of morphing techniques: low quality (visible artifacts), automated morph techniques (i.e. MipGAN) and high quality (i.e. Manual). Our algorithm achieved very good performance, especially being the highest performing on multiple datasets, including

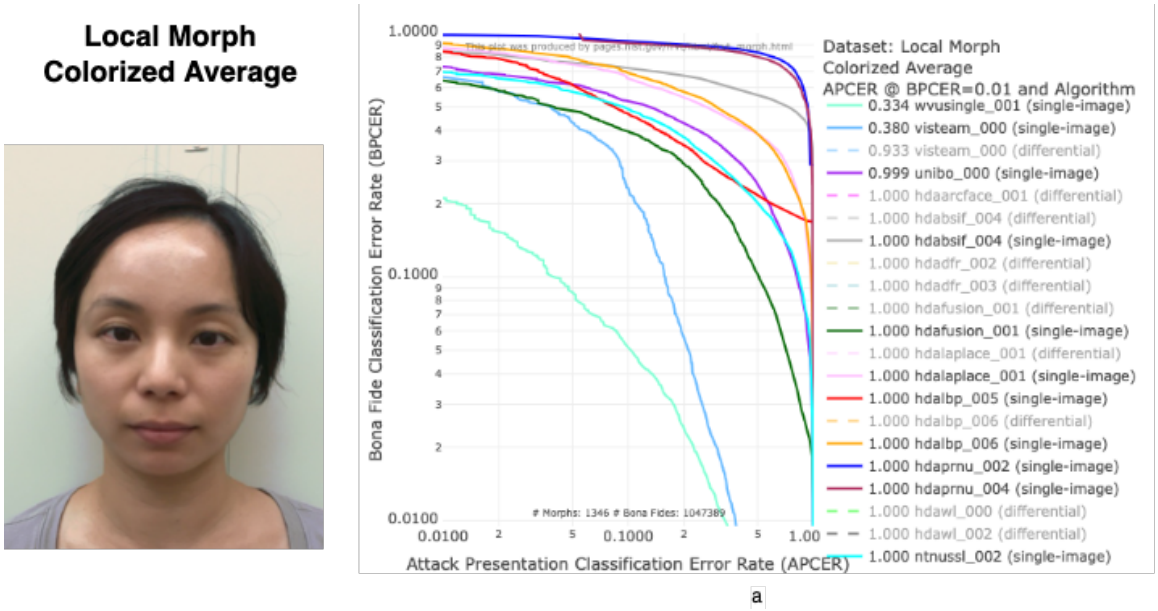

Figure 5.2: Performance of our detector, wvusingle_001, on the Local Morph Colorized Average morphs [28].

print-and-scanned.

The submission process for the NIST Morph test requires our algorithm to be in a Linux dynamically-linked library file in C++. First, we had to convert our experimental code, which was written in Python PyTorch, to C++. Furthermore, NIST only employs CPUs in its testing and the final executable has to be able to be forked into different processes to run concurrently on multiple machines at once. Once our algorithm was successfully running their validation package, we submitted it for objective evaluation.

As the NIST FRVT Morph tests a multitude of morph techniques, in this paper, we are only sharing a select few. The rest of the results can be viewed on their website or in the published paper.

In Figure 5.2, the Local Morph Colorized Average dataset, a landmark-based technique, is tested on each of the submitted morph detectors. This is a Tier 2 dataset, according to the methodology of NIST, meaning it is an automated morph technique used in academic research with no obvious artifacts. Our morph detector

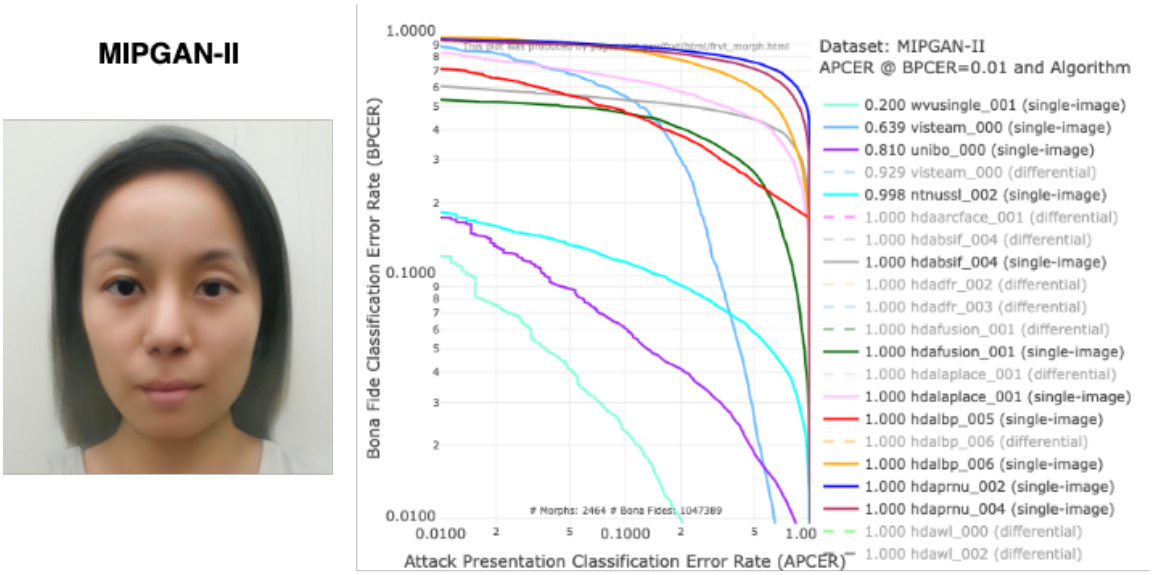

Figure 5.3: Performance of our detector, wvusingle_001, on the MIPGAN-II morphs [28].

performs extremely well on this dataset as can be seen in the DET curve. On the side, the APCER at a controlled threshold of 1% is listed. This means the APCER or morph miss rate at a contorlled threshold of 1 in 100 people being inconvenienced with a false alarm. Our detector has the lowest rate at 33.4% morphs being missed. This is still remarkable given the infancy of this research field however it does illustrate how much more progress needs to be made. This means that 1 in 3 morphs pass when 1 in 100 bona fides are inconvenienced with a false alarm. From an operational reliability standpoint, this is still a very high rate.

Next, in Figure 5.3, we see our detector doing well on a GAN-based technique, MIPGAN-II, which is a modified version of StyleGAN 2 with additional identity losses. In Figure 5.4, we see our detector does not do so well. These Visa-Border morphs were created using passport-like bona fides and the probes were live capture webcam bona fides. It is possible the low performance is due to the variance with pose angle and illumination in the border crossings. In the future, using lower resolution bona fides may help improve performance on this type of morph.

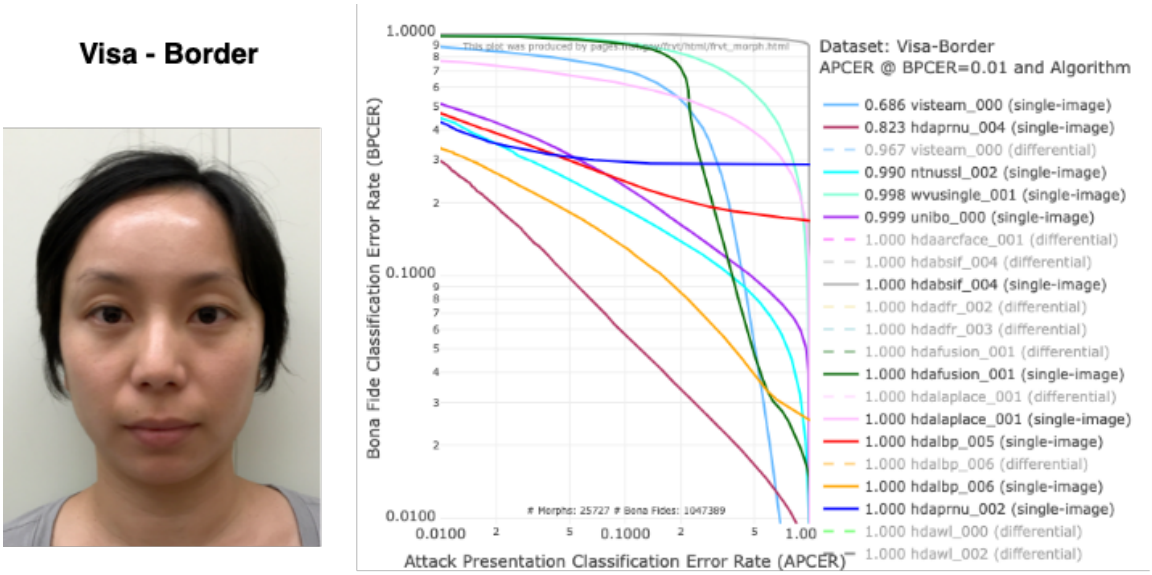

Figure 5.4: Performance of our detector, wvusingle_001, on the Visa-Border morphs [28].

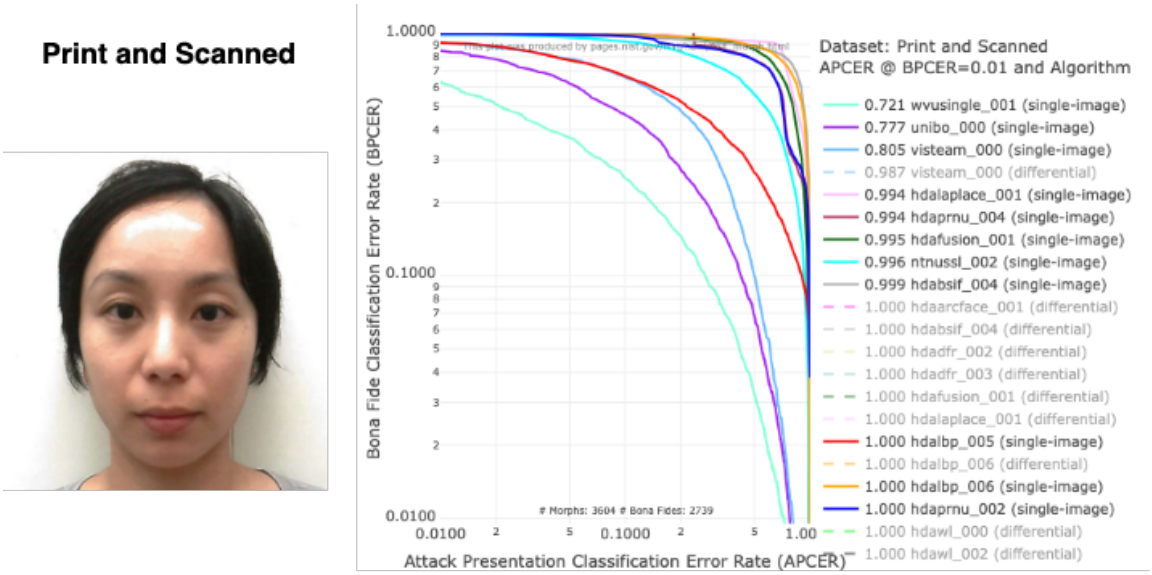

Figure 5.5: Performance of our detector, wvusingle_001, on the Print and Scanned morphs [28].

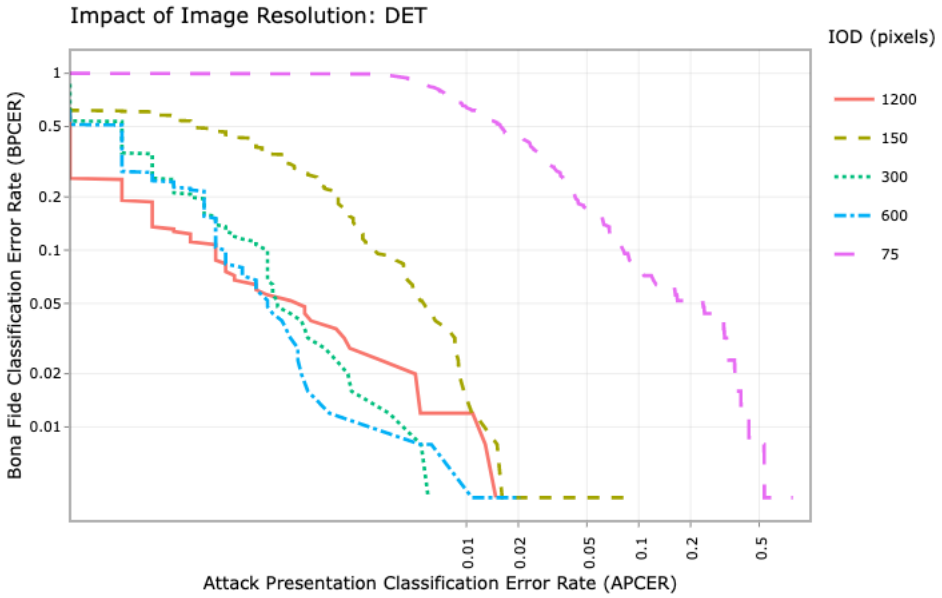

Figure 5.6: DET curve of our morph detector, wvusingle_001, on the impact of image resolution.

In Figure 5.5, the performance on the print-and-scan dataset is shown. Printing and scanning a morph image helps hide the artifacts present in a morphed image. This is done also to mimic the process of submitting a printed photo at time of passport application and the inevitable re-scanning of the photo before it appears on a passport. The images were printed on 2 inches by 2 inches photo paper and scanned at 300 ppi. Interestingly enough, a subset of the VISA-BORDER morphs were used for this print-and-scan dataset yet our performance for the print-and-scan Visa-Border is higher than the original Visa-Border images. It is possible that our detector picks up on the artifacts of the actual print-and-scan process and mistakes them for morphing artifacts, thus using those artifacts to determine that the images are morphed.

These are just a few of the notable results shown from the NIST FRVT Morph Test. Full results with all datasets and comparison to all other current state of the art methods is available in [28] as well as online [1]

[1] https://pages.nist.gov/frvt/html/frvt_morph.html

5.5 Summary

Overall, the network submitted to NIST generalizes well, especially on the automated morph techniques. This shows that using twins is a useful alternative to the time-consuming process of finding suitable look-alikes and it is effective for training a deep morph detector robustly. However, the detector is still weak on images of lower resolution, especially when the faces have an inter-ocular distance (IOD) of only 75 pixels or 150 pixels as can be seen in Figure 5.6. This is something to train on in the future. This was our first foray into NIST evaluation, however we hope to eventually submit more morphing algorithms, including in the wavelet domain in the future.

Chapter 6

Conclusion

In the scope of this thesis, morphing attack detection is studied in detail. Although there is progress towards robust morph detection, this still remains a challenging problem. Part of the problem is that morphed images are getting more realistic as technology improves. Like how morph detection algorithms are becoming more advanced, morph generation techniques are advancing too. Most solutions have not matured enough to be viable in the real world scenario.

The following sections explain some of the conclusions that can be drawn from this thesis, address the limitations we faced in creating this framework, and a discussion of the future work involved in this field.

6.1 Limitations

This work too was particularly limited by the availability of morph datasets of sufficient size and quality. More researchers have stepped up to the plate to address this problem, creating databases that can be distributing in a research setting. It is certain that future research in this area will soon take great strides due to the increasing amount of public databases available. This will also help create more reproducible works and facilitate open source benchmarking that researchers can compare their

work to. This author notes that since the beginning of this thesis, at least three morph datasets have become available for researchers: (1) the morphs generated from FRLL, FERET, and FRGC [38], (2) MipGAN [54], and (3) LMA DRD [6].

6.2 Future Work

There are many other methods that can be proposed for wavelet sub-band selection. The method we chose is definitely hand-crafted with KLD at its core, however deep learning based methods such as group sparsity and attention could also be helpful in selecting the most discriminative sub-bands.

We already trained an image-based twins morph detector and submitted it to NIST for evaluation. Our detector has promising results but also some faults. Noting these issues, we hope to submit a better morph detector in the future for evaluation. It would also be interesting to see if how a twins-morph-trained detector performs in the differential morph attack scenario. In the future, the next step would be train our wavelet-based detector too on a large database of morphs with high resolution and high variance, perhaps exploring a twins morph trained wavelet detector.

6.3 Conclusion

In this work, we introduce a framework to detect morph face images using discriminative wavelet sub-bands selected using KLD. The core of our method is the ability to identify morph artifacts in the wavelet domain, leveraging the most informative sub-bands for morph detection. We demonstrate the performance of our morph detector on four different datasets. Likewise, we compare our model's performance with baseline models constructed with common classical and deep methods employed in the literature. We show how discriminative wavelet sub-bands for morph detection compare to the original image and show how effective it is to translate the morph

from the image domain to the spatial frequency domain and harness discriminative wavelet sub-bands for morph attack detection.

Additionally, we explore the impact of training on a large scale dataset on morphs generated from identical twins. We also demonstrate that using twins as an alternative to the time-consuming process of finding look-alikes for generating a large-scale dataset is quite effective for training a morph detector robustly. The twins-morph-trained detector exhibits good performance in comparison to other current state-of-the-art methods submitted to the NIST FRVT Morph evaluation.

Bibliography

[1] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. "Surf: Speeded up robust features". In: *European conference on computer vision*. Springer. 2006, pp. 404–417.

[2] Guido Borghi et al. "A Double Siamese Framework for Differential Morphing Attack Detection". In: *Sensors* 21.10 (2021). ISSN: 1424-8220. DOI: 10.3390/s21103466. URL: https://www.mdpi.com/1424-8220/21/10/3466.

[3] Thirimachos Bourlai. *Face recognition across the imaging spectrum*. Springer, 2016.

[4] 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. 2005, 886–893 vol. 1. DOI: 10.1109/CVPR.2005.177.

[5] Naser Damer et al. "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)*. 2018, pp. 1–10. DOI: 10.1109/BTAS.2018.8698563.

[6] Naser Damer et al. "PW-MAD: Pixel-wise Supervision for Generalized Face Morphing Attack Detection". In: *CoRR* abs/2108.10291 (2021). arXiv: 2108.10291. URL: https://arxiv.org/abs/2108.10291.

[7] Naser Damer et al. “Realistic dreams: Cascaded enhancement of gan-generated images with an example in face morphing attacks”. In: *10th IEEE International Conference on Biometrics Theory, Applications and Systems, BTAS 2019*. 2019.

[8] Luca Debiasi et al. “On the detection of gan-based face morphs using established morph detectors”. In: *International Conference on Image Analysis and Processing*. Springer. 2019, pp. 345–356.

[9] Lisa DeBruine. *debruine/webmorph: Beta release 2*. Version v0.0.0.9001. Jan. 2018. DOI: 10.5281/zenodo.1162670. URL: https://doi.org/10.5281/zenodo.1162670.

[10] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. “Face Demorphing”. In: *IEEE Transactions on Information Forensics and Security* 13.4 (2018), pp. 1008–1017. DOI: 10.1109/TIFS.2017.2777340.

[11] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. “Face demorphing in the presence of facial appearance variations”. In: *2018 26th European Signal Processing Conference (EUSIPCO)*. IEEE. 2018, pp. 2365–2369.

[12] Matteo Ferrara, Annalisa Franco, and Davide Maltoni. “The magic passport”. In: *IEEE International Joint Conference on Biometrics*. IEEE. 2014, pp. 1–7.

[13] RDU Frontex. “Best practice operational guidelines for Automated Border Control (ABC) systems”. In: *European Agency for the Management of Operational Cooperation, Research and Development Unit,. https://bit. ly/2KYBXhz Accessed* 9.05 (2012), p. 2013.

[14] Justin Huggler. “German photography studios protest government’s new planned passport rules”. In: *Telegraph* (Jan. 2020). URL: https://www.telegraph.co.uk/news/2020/01/08/german-photography-studios-protest-governments-new-planned-passport/.

[15] Doc ICAO. "9303-Machine Readable Travel Documents-Part 9: Deployment of Biometric Identification and Electronic Storage of Data in eMRTDs". In: *International Civil Aviation Organization (ICAO)* (2015).

[16] Juho Kannala and Esa Rahtu. "Bsif: Binarized statistical image features". In: *Proceedings of the 21st International Conference on Pattern Recognition (ICPR2012)*. IEEE. 2012, pp. 1363–1366.

[17] Tero Karras, Samuli Laine, and Timo Aila. "A style-based generator architecture for generative adversarial networks". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2019, pp. 4401–4410.

[18] Fons Knopjes. *State of the art of Morphing Detection*.

[19] Robin Kramer et al. "Face morphing attacks: Investigating detection with humans and computers". In: *Cognitive Research: Principles and Implications* 4 (Dec. 2019). DOI: 10.1186/s41235-019-0181-4.

[20] Shengcai Liao et al. "Learning multi-scale block local binary patterns for face recognition". In: *International Conference on Biometrics*. 2007, pp. 828–837.

[21] Ziwei Liu et al. "Deep Learning Face Attributes in the Wild". In: *Proceedings of International Conference on Computer Vision (ICCV)*. Dec. 2015.

[22] David G. Lowe. "SIFT-the scale invariant feature transform". In: *International Journal of Computer Vision* 2 (2004), pp. 91–110.

[23] Andrey Makrushin, Tom Neubert, and Jana Dittmann. "Automatic Generation and Detection of Visually Faultless Facial Morphs". In: *Proceedings of the 12th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 6 VISAPP: VISAPP, (VISIGRAPP 2017)*. INSTICC. SciTePress, 2017, pp. 39–50. ISBN: 978-989-758-227-1. DOI: 10.5220/0006131100390050.

[24] Andrey Makrushin, Tom Neubert, and Jana Dittmann. "Humans Vs. Algorithms: Assessment of Security Risks Posed by Facial Morphing to Identity Verification at Border Control". In: *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 4: VISAPP,* INSTICC. SciTePress, 2019, pp. 513–520. ISBN: 978-989-758-354-4. DOI: 10.5220/0007378905130520.

[25] Satya Mallick. *Face morph using opencv–c++/python*. 2016.

[26] Tom Neubert, Christian Kraetzer, and Jana Dittmann. "A Face Morphing Detection Concept with a Frequency and a Spatial Domain Feature Space for Images on eMRTD". In: *Proceedings of the ACM Workshop on Information Hiding and Multimedia Security*. 2019, pp. 95–100.

[27] Tom Neubert et al. "Extended StirTrace benchmarking of biometric and forensic qualities of morphed face images". In: *IET Biometrics* 7.4 (2018), pp. 325–332.

[28] Mei Ngan et al. "Face Recognition Vendor Test (FRVT) Part 4: MORPH Performance of Automated Face Morph Detection". In: *National Institute of Technology (NIST), Tech. Rep. NISTIR* 8292 (2021).

[29] Timo Ojala, Matti Pietikainen, and David Harwood. "Performance evaluation of texture measures with classification based on Kullback discrimination of distributions". In: *Proceedings of 12th International Conference on Pattern Recognition*. Vol. 1. IEEE. 1994, pp. 582–585.

[30] Fei Peng, Le-Bing Zhang, and Min Long. "FD-GAN: Face-demorphing generative adversarial network for restoring accomplice's facial image". In: *CoRR* abs/1811.07665 (2018). arXiv: 1811.07665. URL: http://arxiv.org/abs/1811.07665.

[31] Alyssa Quek. *Facemorpher*. June 2019. URL: https://github.com/alyssaq/face_morpher.

[32] R Raghavendra et al. "Face morphing versus face averaging: Vulnerability and detection". In: *2017 IEEE International Joint Conference on Biometrics (IJCB)*. IEEE. 2017, pp. 555–563.

[33] Ramachandra Raghavendra, Kiran B. Raja, and Christoph Busch. "Detecting morphed face images". In: *2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. 2016, pp. 1–7. DOI: 10.1109/BTAS.2016.7791169.

[34] Ramachandra Raghavendra et al. "Transferable deep-CNN features for detecting digital and print-scanned morphed face images". In: *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*. IEEE. 2017, pp. 1822–1830.

[35] Raghavendra Ramachandra et al. "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)*. 2019, pp. 1–8.

[36] David J Robertson, Robin SS Kramer, and A Mike Burton. "Fraudulent ID using face morphs: Experiments on human and automatic recognition". In: *PLoS One* 12.3 (2017), e0173319.

[37] Alexander Röttcher, Ulrich Scherhag, and Christoph Busch. "Finding the Suitable Doppelgänger for a Face Morphing Attack". In: *2020 IEEE International Joint Conference on Biometrics (IJCB)*. 2020, pp. 1–7. DOI: 10.1109/IJCB48548.2020.9304878.

[38] Eklavya Sarkar et al. "Vulnerability Analysis of Face Morphing Attacks from Landmarks and Generative Adversarial Networks". In: *CoRR* abs/2012.05344 (2020). arXiv: 2012.05344. URL: https://arxiv.org/abs/2012.05344.

[39] Ulrich Scherhag, Christian Rathgeb, and Christoph Busch. "Morph Detection from Single Face Image: A Multi-Algorithm Fusion Approach". In: *Proceedings of the 2018 2nd International Conference on Biometric Engineering and Applications*. ICBEA '18. Amsterdam, Netherlands: Association for Computing Machinery, 2018, pp. 6–12. ISBN: 9781450363945. DOI: 10.1145/3230820.3230822. URL: https://doi.org/10.1145/3230820.3230822.

[40] Ulrich Scherhag, Christian Rathgeb, and Christoph Busch. "Towards detection of morphed face images in electronic travel documents". In: *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*. 2018, pp. 187–192.

[41] Ulrich Scherhag et al. "Deep face representations for differential morphing attack detection". In: *arXiv preprint arXiv:2001.01202* (2020).

[42] Ulrich Scherhag et al. "Detection of Face Morphing Attacks Based on PRNU Analysis". In: *IEEE Transactions on Biometrics, Behavior, and Identity Science* 1.4 (2019), pp. 302–317. DOI: 10.1109/TBIOM.2019.2942395.

[43] Florian Schroff, Dmitry Kalenichenko, and James Philbin. "Facenet: A unified embedding for face recognition and clustering". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015, pp. 815–823.

[44] Clemens Seibold et al. "Accurate and robust neural networks for security related applications exampled by face morphing attacks". In: *arXiv preprint arXiv:1806.04265* (2018).

[45] Clemens Seibold et al. "Detection of Face Morphing Attacks by Deep Learning". In: July 2017, pp. 107–120. ISBN: 978-3-319-64184-3. DOI: 10.1007/978-3-319-64185-0_9.

[46] Sobhan Soleymani et al. *Differential Morphed Face Detection Using Deep Siamese Networks*. 2020. arXiv: 2012.01541 [cs.CV].

[47] Sobhan Soleymani et al. “Mutual Information Maximization on Disentangled Representations for Differential Morph Detection”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 2020, pp. 1731–1741.

[48] Sophie Spelsberg. *Peng!-Kollektiv fälscht Passbilder: Zwei Gesichter, ein Dokument*. Sept. 2018. URL: https://taz.de/Peng-Kollektiv-faelscht-Passbilder/!5534868/.

[49] West Virginia University. “Twins Day Dataset”. In: (2010-2019). URL: https://biic.wvu.edu/data-sets/twins-day-dataset-2010-1015.

[50] *Utrecht ECVP Face Database*. URL: %7Bhttps://pics.stir.ac.uk/2D%7B%5C_%7Dface%7B%5C_%7Dsets.html%7D.

[51] Sushma Venkatesh et al. “Can GAN Generated Morphs Threaten Face Recognition Systems Equally as Landmark Based Morphs? - Vulnerability and Detection”. In: *2020 8th International Workshop on Biometrics and Forensics (IWBF)*. 2020, pp. 1–6.

[52] Sushma Venkatesh et al. “Detecting Morphed Face Attacks Using Residual Noise from Deep Multi-scale Context Aggregation Network”. In: *2020 IEEE Winter Conference on Applications of Computer Vision (WACV)*. 2020, pp. 269–278. DOI: 10.1109/WACV45572.2020.9093488.

[53] Sushma Venkatesh et al. “Morphed Face Detection Based on Deep Color Residual Noise”. In: *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. 2019, pp. 1–6. DOI: 10.1109/IPTA.2019.8936088.

[54] Haoyu Zhang et al. “MIPGAN - Generating Robust and High Quality Morph Attacks Using Identity Prior Driven GAN”. In: *CoRR* abs/2009.01729 (2020). arXiv: 2009.01729. URL: https://arxiv.org/abs/2009.01729.

[55] Kaipeng Zhang et al. “Joint face detection and alignment using multitask cascaded convolutional networks”. In: *IEEE Signal Processing Letters* 23.10 (2016), pp. 1499–1503.