Differential Morphed Face Detection Using Deep Siamese Networks*

Sobhan Soleymani[1 2 3][0000−0003−3541−0918], Baaria Chaudhary[1 2][0000−0002−9564−951X], Ali Dabouei[1][0000−0002−1084−6224], Jeremy Dawson[1][0000−0002−4539−7588], and Nasser M. Nasrabadi[1][0000−0001−8730−627X]

[1] West Virginia University, Morgantown WV 26506, USA
[2] Authors Contributed Equally.
[3] Corresponding Author
{ssoleyma, bac0062, ad0046}@mix.wvu.edu,
{jeremy.dawson, nasser.nasrabadi}@mail.wvu.edu

Abstract. Although biometric facial recognition systems are fast becoming part of security applications, these systems are still vulnerable to morphing attacks, in which a facial reference image can be verified as two or more separate identities. In border control scenarios, a successful morphing attack allows two or more people to use the same passport to cross borders. In this paper, we propose a novel differential morph attack detection framework using a deep Siamese network. To the best of our knowledge, this is the first research work that makes use of a Siamese network architecture for morph attack detection. We compare our model with other classical and deep learning models using two distinct morph datasets, VISAPP17 and MorGAN. We explore the embedding space generated by the contrastive loss using three decision making frameworks using Euclidean distance, feature difference and a support vector machine classifier, and feature concatenation and a support vector machine classifier.

Keywords: Differential morph detection · Siamese network · Contrastive loss.

1 Introduction

Biometric facial recognition systems have increasingly been integrated into border control and other security applications that utilize identification tasks, such as official identity cards, surveillance, and law enforcement. These systems provide high accuracy at a low operational cost. In addition, face capture is non-invasive and benefits from a relatively high social acceptance. People use their faces to unlock their phones and also to recognize their friends and family. Furthermore, facial recognition systems contain an automatic fail-safe: if the algorithm triggers a false alarm, a human expert on-site can easily perform the

* This work is based upon a work supported by the Center for Identification Technology Research and the National Science Foundation under Grant #1650474.

Table 1: Table 1: Differential morph algorithms.

Algorithm	Method	Database
Face Demorphing [27,1]	Image Subtraction	MorphDB, landmark-based
Mutli-algorithm fusion approach [21]	Feature vectors and feature difference	Landmark-based
Deep models [7]	Feature embeddings	GAN-based
Our Method – Deep Siamese Network	L_2 difference of embedding representations	landmarkbased, and GAN-based

verification. For these reasons, facial recognition systems enjoy a sizable advantage over other biometric systems. Consequently, the International Civil Aviation Organization (ICAO) has mandated the inclusion of a facial reference image in all electronic passports worldwide [16]. This means that the only biometric identifier present in passports globally is the face.

Although facial recognition systems are largely successful, they still are not impervious to attack. The mass adoption of automatic biometric systems in border control has revealed critical vulnerabilities in the border security scheme, namely the inability of these systems to accurately detect a falsified image. This vulnerability is further exacerbated by a loophole in the passport application process: the facial reference image, either digitally or as a physical print, is provided by the applicant at the time of enrollment. This opens a window for the applicant to potentially manipulate the image before application submission. One type of manipulation that is recently identified as a serious threat is the morph attack [10], in which a facial reference image can be verified as two or more separate identities. A successful morphing attack allows two or more people to utilize the same passport to travel.

Thus, a criminal attacker, who otherwise cannot travel freely, could obtain a valid passport by morphing his face with that of an accomplice [10]. Many morphing applications are not only freely available and easily accessible but also have no knowledge barrier [22]. As such, it is almost absurdly simple for a criminal to procure a legitimate travel document. There are only a few straightforward steps: (1) find an accomplice with similar facial features, (2) morph both faces together such that existing facial recognition systems would classify the resulting morphed face as either of the original individuals, (3) the accomplice applies for a passport with the morphed image. The resulting passport could then be used by both the criminal and the accomplice to travel as they wish. Currently, we are unaware of any system in the passport verification process that is designed specifically for the detection of these manipulations. Moreover, commercial off-the-shelf systems (COTS) have repeatedly failed to detect morphed images [27]. Likewise, studies show human recognizers are also unable to correctly differentiate a morphed facial image from an authentic one [27,26,3].

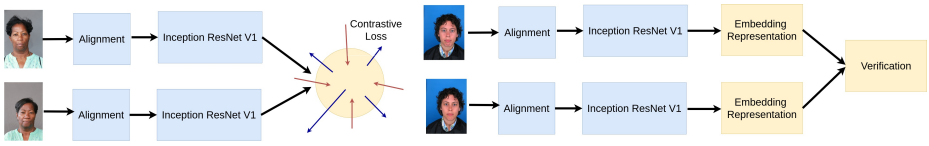

Fig. 1: Network architecture: Image pairs are fed into the MTCNN for face detection and alignment, then into the Inception ResNET v1,where contrastive loss is applied. From the feature embedding representation, verification is conducted by computing the L_2 distance between the feature vectors and a decision score is produced. Left) the training phase on WVU Twins Day dataset and the training portion of the morph dataset and right) the test phase.

We propose to develop a novel differential morphing attack detection algorithm using a deep Siamese network. The Siamese network takes image pairs as inputs and yields a confidence score on the likelihood that the face images are from the same person. We employ a pre-trained Inception ResNET v1 as the base network. The experiments are conducted on two separate morphed image datasets: VISAPP17 [21] and MorGAN [7,8]. Results show an D-EER of 5.6% for VISAPAP17 and an D-EER of 12.5% for MorGAN. In the following sections, we briefly summarize the related works in Section II, explain the methodology in Section III, and discuss our experiments and subsequent results in Section IV. Finally, conclusions are presented in Section V.

2 Related Works

The vulnerability of face recognition systems to morph attacks was first introduced by [10]. Since then, many morph detection algorithms have been proposed of two types: single (no reference) and differential. Single (no reference) morph attack detection algorithms rely solely on the potential morphed image to make their classification. Morphs are detected by extracting and analyzing features from the image in an attempt to identify the unique artifacts that indicate the face image was morphed or tampered with. On the other hand, differential morph attack detection algorithms rely on an additional trusted image, typically a live capture at border security, to compare the potential morphed image with. As such, differential morphing attack detection algorithms have more information at their disposal to make their classification and therefore perform significantly better than single morph detection algorithms [30].

The majority of the current research exists solely in the single morph attack detection domain. Many classical hand-crafted feature extraction techniques have been explored to solve this problem. The most well-performing of these general image descriptors is Binarized Statistical Image Features (BSIF) [17], used in [24], in which extracted BSIF features were classified using a Support Vector Machine (SVM). Deep learning methods have also been proposed [36] [33] [25]. In [25], complementary features from VGG-19 and AlexNet, both pre-trained on

ImageNet and additionally fine-tuned on a morph dataset, are concatenated and then used to train a Probabilistic Collaborative Representation-based Classifier (ProCRC). A multi-algorithm fusion approach that combines texture descriptors, keypoint descriptors, gradient estimators, and deep neural network methods has also shown promising results [29]. The authors extract BSIF and Local Binary Pattern (LBP) [18] features to obtain feature vectors. Other feature descriptors such as Scale Invariant Feature Transform (SIFT) [20] and Speeded-Up Robust Features (SURF) [2] are also used to extract keypoint descriptors and Histogram of Gradients (HOG) is used as a gradient estimator. Finally, deep feature embeddings from the OpenFace DNN are used as the last feature vector. All the above feature vectors are then used to train separate SVMs. In the end, score-level fusion is applied to obtain the final decision score for the potential morphed image.

There are a few papers also that address differential morph attack detection. Of these, reverting of a face morph or face demorphing [11] [12] has provided encouraging results. The demorphing algorithm subtracts the potential morphed image from the trusted image and uses the difference for classification. Feature extraction methods used in single morph attack detection can also be applied to the differential problem domain as well by taking the difference of the feature vectors of the potential morphed image and the trusted image. This difference vector along with the original feature vector for the potential morphed image were then used to train a difference SVM and a feature SVM, respectively. Score-level fusion is used to arrive at the final decision score. This method is explored in [30] using LBP, BSIF, SIFT, SURF, and HOG descriptors. Scherhag et. al [31] uses deep face representations from feature embeddings extracted from a deep neural network to detect a morph attack. The authors of [31] also emphasized the need for high-variance and constructed their morph image database using multiple morph algorithms. Several post-processing steps were also applied to emulate the actual compression methods used in storing passport photos in electronic passports, including reducing resolution, JPEG200 compression, and printing-and-scanning. Table 1 compares the existing differential morph algorithms.

Although these methods have shown some success, none are sufficiently robust. These algorithms train on morph datasets of very limited size and scope. When tested on additional datasets, they perform poorly, indicating the models overfit [28]. These results are notably evident in the NIST FRVT morph detection test [23], in which nearly all the algorithms submitted exhibited low performance on almost all tested morph datasets of varying quality and method. The NIST FRVT test is administered on multiple unseen datasets, ranging from automatically generated morphs to manually manipulated high quality morphs to print-and-scanned morphs. The deep learning method described in [31] outperforms the other models in the NIST test.

3 Method

The fundamental issue facing morph attack detection researchers is the lack of a large database of morphs with high variance. Many researchers create their own synthetic morph database, typically employing automated generative techniques, such as landmark manipulation [22] or General Adversarial Networks (GANs) [7]. Although commercial software such as Adobe Photoshop or GIMP 2.10 have also been used to manually construct morphs, these methods are often time-consuming, and it is difficult to generate the number of morphs required to train a model. Each method generates different artifacts in the image, such as ghosting, unnatural transition between facial regions, hair and eyelash shadows, blurriness in the forehead and color, among others.

As presented in Figure 1, the proposed architecture is a Siamese neural network [4], in which the subnetworks are instances of the same network and share weights. Contrastive loss [14][25] is the loss function for training the Siamese network. Contrastive loss is a distance-based loss function, which attempts to bring similar images closer together into a common latent subspace, whereas it attempts to distance the dissimilar ones even more. Essentially, contrastive loss emphasizes the similarity between images of the same class and underscores the difference between images of different classes. The distance is found from the feature embeddings produced by the Siamese network. The margin is the distance threshold that regulates the extent to which pairs are separated.

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

where I_1 and I_2 are the input face images, m is the margin or threshold as described above 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:

$$D(I_1, I_2) = ||\phi(I_1) - \phi(I_2)||_2 \tag{2}$$

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.

To streamline training, an Inception ResNET v1 architecture [35] is chosen as the base network, using weights pre-trained on the VGGFace2 dataset [5]. The network is then re-trained with the WVU Twins Day dataset [28] for the Siamese implementation. The model is optimized by enforcing contrastive loss on the embedding space representation of the genuine and imposter twin samples. The trained Siamese network is then additionally fine-tuned using the training portion of each morph database. To obtain a more discriminative embedding, the representations of the face image and its horizontal embedding are concatenated. The feature embeddings are taken from the last fully-connected layer and the L_2 distance between the two embeddings is calculated for the verification. As presented in Figure 2, in our experiments we consider two additional decision making algorithms to explore the embedding space constructed by the contrastive loss, where we augment the proposed framework with the

Table 2: The performance of the proposed framework on VISAPP17.

Method	APCER@BPCER			BPCER@APCER			D-EER
	5%	10%	30%	5%	10%	30%	
SIFT	45.12	37.89	17.94	65.11	43.28	17.91	0.221
SURF	55.57	42.72	20.76	72.58	50.74	20.89	0.225
LBP	23.88	19.40	1.58	23.88	20.65	13.43	0.187
BSIF	25.37	22.38	1.49	28.77	25.37	8.91	0.164
FaceNet	11.82	9.82	5.08	29.82	6.91	0.25	0.095
Ours	6.11	3.47	1.64	7.31	4.22	0.24	0.056
Ours+SVM (concat.)	5.78	3.29	1.52	6.67	3.95	0.21	0.054
Ours+SVM (difference)	5.29	3.17	1.43	6.12	3.71	0.19	0.052

Table 3: The performance of the proposed framework on MorGAN

Method	APCER@BPCER			BPCER@APCER			D-EER
	5%	10%	30%	5%	10%	30%	
SIFT	65.41	53.37	23.53	97.45	66.66	23.24	0.262
SURF	69.88	56.25	29.82	98.24	78.07	30.06	0.298
LBP	62.43	54.13	21.46	28.40	18.71	14.92	0.155
BSIF	39.85	31.26	16.97	14.22	8.64	7.40	0.101
FaceNet	36.72	30.15	18.49	38.38	26.67	10.51	0.161
Ours	31.85	25.61	13.21	14.32	12.11	5.49	0.125
Ours+SVM (concat.)	29.43	24.21	12.35	13.72	11.75	5.18	0.113
Ours+SVM (difference)	27.95	22.78	12.05	13.46	10.42	4.94	0.102

verification of the difference and concatenation of the embedding features of a pair using radial basis function kernel support vector machine (SVM) classifiers. These classifiers, which are learned using the training portion of each dataset, are utilized during the test phase to recognize genuine and imposter pairs.

3.1 Experimental Setup

The two morph image databases used in this experiment are VISAPP17 [21] and MorGAN [7] [8]. As presented in Figure 3, we purposefully employ two different morph databases, created from two different face image datasets that apply two different morphing techniques to investigate how our model generalizes. VISAPP17 is a collection of complete and splicing morphs generated from the Utrecht FCVP database [30]. The images are 900×1200 pixels in size.

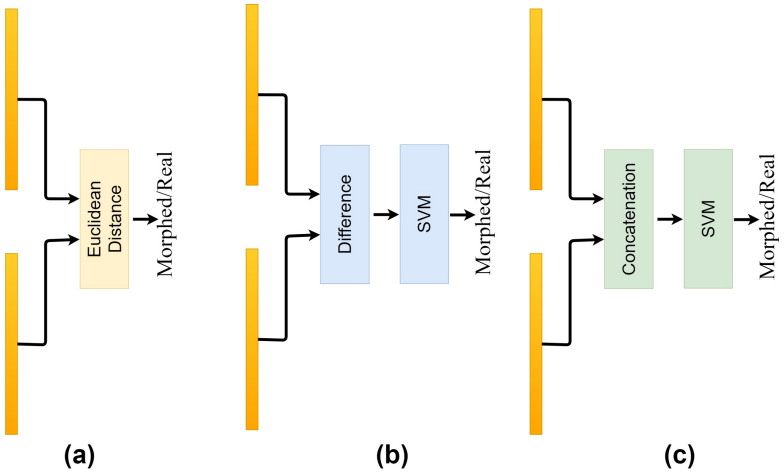

Fig. 2: During the test phase, three decision-making algorithms are considered to explore the embedding constructed embedding space: (a) Euclidean distance between the representation of the samples in a pair is considered to make the decision. (b) The difference and (c) the concatenation of the learned representations is fed to a SVM classifier.

This dataset is generated by warping and alpha-blending [37]. To construct this dataset, facial landmarks are localized, the face image is tranquilized based on these landmarks, triangles are warped to some average position, and the resulting images are alpha-blended, where alpha is set to 0.5. A subset of 183 high quality splicing morphs constructed by selecting morph images without any recognizable artifacts (VISAPP17-Splicing-Selected dataset) is used along with 131 real images for a total set of 314 images. The morphs were created using the splicing technique, where after landmark manipulation, the resulting morphed face is spliced into the face of one of the original images. This preserves the background and hair, which helps avoid the issue of blurry artifacts and ghosting that typically occurs in these regions. However, this also means that the resulting morph derives its face shape from only one of the contributing individuals.

The MorGAN database is constructed from a selection of full-frontal face images manually chosen from the CelebA dataset [19]. It consists of a custom morphing attack pipeline (MorGAN), created by the authors, that uses a GAN, inspired by inspired by learned inference model [9], to generate morphs. The database consists of 1500 bona fide probe images, 1500 bonafide references, and 1000 MorGAN morphs of 64×64 pixels in size. There is also an additional MorGAN database, in which the MorGAN morphs have been super-resolved to 128×128 pixels according to the protocol described in [6]. The faces are detected and aligned using the MTCNN framework [38]. The aligned images are then resized to 160×160 pixels to prepare the images for the Siamese network.

As the Siamese network expects pairs of images as input, the morph images are paired off into genuine face pairs and imposter face pairs, where a genuine pair consists of two trusted images and an imposter pair consists of a trusted image and a morph image. We employ the same train-test split provided by the authors of MorGAN to facilitate comparison of performance with other algorithms using this database. The train-test split is purposefully disjoint, with no overlapping morphs or contributing bonafides to morphs. This enables us to attain an accurate representation of the performance. For the VISAPP17 dataset, 50% of the subjects are considered for training, while the other 50% is used to evaluate the performance of the framework. For the train-test split we consider the same portions for male and female subjects. In addition, 20% of the training set of the morph datasets was used during model optimization as the validation set. Batch size of 64 pairs of images of size $160 \times 160 \times 3$ is used for training the model. Stochastic Gradient Descent (SGD) is the chosen optimizer. For the initial round of training with the Twins Day dataset, the initial learning rate is set to 0.1, multiplied by 0.9 every 5 epochs until the final value of 10^{-6}. When fine-tuning with morph datasets, the initial learning rate is set to 10^{-3}, then multiplied by 0.9 every 5 epochs until the final learning rate value of 10^{-6}. The input data is further augmented with vertical and horizontal flips to increase the training set and improve generalization.

4 Results

We study the performance of the proposed differential morph detection framework using VISAPP17 and MorGAN datasets. The performance is compared with state-of-the-art classical and deep learning frameworks.

4.1 Metrics

We apply the widely accepted metrics for morphing attack detection, APCER and BPCER, to our algorithm. The Attack Presentation Classification Error Rate (APCER) is the rate at which morph attack images are incorrectly classified as bonafide. Similarly, the Bonafide Probe Classification Error Rate (BPCER) is the rate at which bonafide images are incorrectly classified as morph attack presentations. In real-world applications, the BPCER is the measure by which individuals are inconvenienced with a false alarm. Hence, artificially regulating the BPCER rate by restricting it to fixed thresholds is recommended for face recognition systems [13]. We plot these rates in a Detection Error Tradeoff (DET) graph. D-EER is the detection Equal Error Rate or the decision threshold at which APCER and BPCER are equal. Additionally, we also present BPCER and APCER values for fixed APCER and BPCER values, respectively.

We employed BSIF [17] and LBP [18] as classical texture descriptors. In addition, we consider key-point detection frameworks, SURF [2] and SIFT [20]. These classical methods have shown promise in morph detection in the literature. However, due to the private nature of the databases used in the original papers,

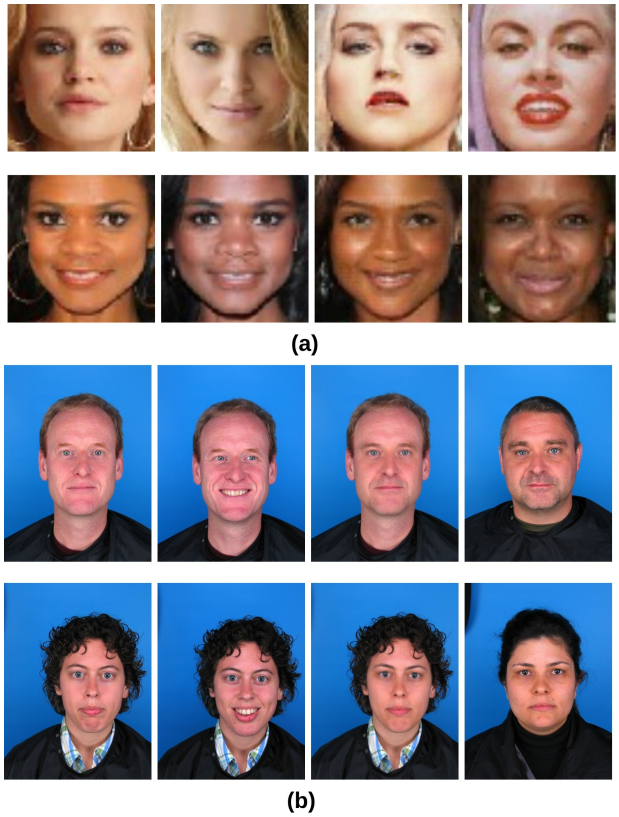

Fig. 3: Samples from (a) Morgan and (b) VISAPP17-Splicing-Selected datasets. For each dataset, the first and second faces are the gallery and probe bona fide images and the third face is the morph image construed from the first and forth face images. The original sizes for face images in these datasets are 64×64 and 1500×1200, respectively. All the faces are resized to 160×160 after detection and alignment using MTCNN.

we are unable to directly compare our results with theirs. Still, we employ the exact methodology described in the papers for best comparison. Our baseline models consist of these four frameworks in combination with an SVM classifier. The LBP feature descriptors are extracted according to patches of 3×3. The resulting feature vectors, normalized histograms of size 256, are the values of the LBP binary code. The SIFT and SURF are implemented using the default parameters [20] [2]. 8-bits BSIF feature vectors are constructed on a 3×3 filters. These filters are Independent Component Analysis filters provided by [15]. The feature vectors are then fed into a SVM with an RBF kernel. For all classical baseline models, we follow [30], where the feature representation of the image in question is subtracted from the feature representation of the trusted image before feeding it to the SVM classifier.

We also compare our Siamese network with FaceNet [32] implementation Inception-ResNET v1, where the distance between the embedding representations of the images in pair is considered to provide the decision. Tables 2 and 3 presents the results on VISAPP17 and MorGAN datasets, respectively. As presented in these tables, fine-tuning using the training portion of each morph dataset demonstrably provides better performance for both datasets. This can be interpreted as the Siamese network is learning the generative nature of the morph images in the dataset. For the VISAPP17 train-test split, fine-tuning on the training portion of the dataset results in the BPCER@APCER=5% dropping from 29.82% to 7.31%. For the MorGAN dataset, this fine-tuning helps lower the BPCER@APCER=5% from 38.38% to 14.32%. The great difference in BPCER for MorGAN and VISAPP17 can be attributed to the overall difficulty of the MorGAN dataset, particularly because it is of a lower resolution than VISAPP17. On the other hand, the proposed SVM-based decision making frameworks can further improve the performance of the proposed framework.

For both datasets texture descriptors outperform key-point based models which is consistent with the study in [10]. In addition, the proposed Siamese framework outperforms the FaceNet implementation since the proposed framework initially learns to distinguish between the images with very small differences through re-training on WVU Twins Day dataset and then it is fine-tuned on the training portion of the corresponding dataset. For the MorGAN dataset, our results follow the original paper [8] results on single image morph detection, where the texture descriptors outperform convolutional neural networks, i.e., FaceNet. This can be attributed to the generative nature of this dataset, which can be described using texture descriptors better than deep models. However, fine-tuning the deep model on the training portion of this dataset provides compatible results with BSIF.

Class activation maps [39] provide the attention of the decision with regard to regions of the face image. In Figure 4, we follow the implementation of gradient-weighted class activation maps [34]. Here, we present the differentiation of the contrastive distance with regard to the feature maps constructed by 'repeat_2' layer in Inception-ResNET v1. In this figure, we also report the average per pixel distance between the Grad-CAMs constructed for face images in each pair. As

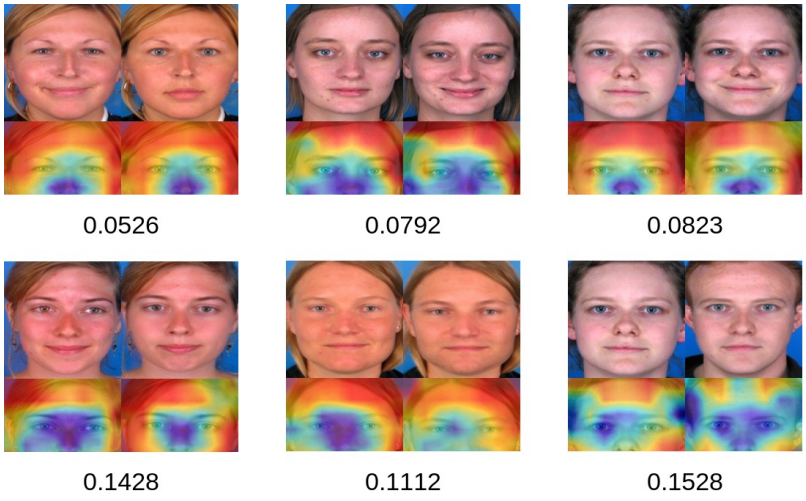

Fig. 4: Grad-CAMs for genuine (top) and imposter (bottom) pairs and the average per pixel distance between the images for each pair. For each imposter pair, the left and right images are real and morphed face images, respectively.

shown in these images, the difference between the activation maps for genuine pairs is smaller compared to the imposter pairs. It worth mentioning that since the datasets include neutral and smiley faces, while computing the distance between the activation maps, we do not consider the lower part of the faces. In addition, we can observe that the class activation maps for the two images in a genuine pair are roughly similar.

5 Conclusion

In this paper, we proposed a deep Siamese network architecture to detect morphed faces. Using contrastive loss and a pre-trained Inception-ResNET v1 on WVU Twins Day dataset, we demonstrate the performance of our Siamese model on two different morph datasets. Likewise, we compare our model's performance with baseline models constructed with common classical and deep methods employed in the literature, where our model outperforms the baseline models. This is attributed to the proposed framework learning to distinguish between images with small differences while training on WVU Twins Day dataset and learning the nature of the corresponding morph dataset by training on the training portion of the dataset.

References

1. Utrecht ecvp face database, `https://pics.stir.ac.uk/2D{_}face{_}sets.html`

2. Bay, H., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: European conference on computer vision. pp. 404–417. Springer (2006)
3. Bourlai, T.: Face recognition across the imaging spectrum. Springer (2016)
4. Bromley, J., Guyon, I., LeCun, Y., Säckinger, E., Shah, R.: Signature verification using a" siamese" time delay neural network. In: Advances in neural information processing systems. pp. 737–744 (1994)
5. Cao, Q., Shen, L., Xie, W., Parkhi, O.M., Zisserman, A.: Vggface2: A dataset for recognising faces across pose and age. In: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018). pp. 67–74. IEEE (2018)
6. Damer, N., Boutros, F., Saladie, A.M., Kirchbuchner, F., Kuijper, A.: 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)
7. Damer, N., Saladie, A.M., Braun, A., Kuijper, A.: Morgan: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network. In: 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS). pp. 1–10 (2018)
8. Debiasi, L., Damer, N., Saladié, A.M., Rathgeb, C., Scherhag, U., Busch, C., Kirchbuchner, F., Uhl, A.: On the detection of gan-based face morphs using established morph detectors. In: International Conference on Image Analysis and Processing. pp. 345–356. Springer (2019)
9. Dumoulin, V., Belghazi, I., Poole, B., Mastropietro, O., Lamb, A., Arjovsky, M., Courville, A.: Adversarially learned inference. arXiv preprint arXiv:1606.00704 (2016)
10. Ferrara, M., Franco, A., Maltoni, D.: The magic passport. In: IEEE International Joint Conference on Biometrics. pp. 1–7 (2014)
11. Ferrara, M., Franco, A., Maltoni, D.: Face demorphing. IEEE Transactions on Information Forensics and Security **13**(4), 1008–1017 (2017)
12. Ferrara, M., Franco, A., Maltoni, D.: Face demorphing in the presence of facial appearance variations. In: 2018 26th European Signal Processing Conference (EUSIPCO). pp. 2365–2369. IEEE (2018)
13. Frontex, R.: Best practice operational guidelines for automated border control (abc) systems. European Agency for the Management of Operational Cooperation, Research and Development Unit,. https://bit. ly/2KYBXhz Accessed **9**(05), 2013 (2012)
14. Hadsell, R., Chopra, S., LeCun, Y.: Dimensionality reduction by learning an invariant mapping. In: 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06). vol. 2, pp. 1735–1742. IEEE (2006)
15. Hyvärinen, A., Hurri, J., Hoyer, P.O.: Natural image statistics: A probabilistic approach to early computational vision., vol. 39. Springer Science & Business Media (2009)
16. ICAO, D.: 9303-machine readable travel documents-part 9: Deployment of biometric identification and electronic storage of data in emrtds. International Civil Aviation Organization (ICAO) (2015)
17. Kannala, J., Rahtu, E.: Bsif: Binarized statistical image features. In: Proceedings of the 21st international conference on pattern recognition (ICPR2012). pp. 1363–1366 (2012)
18. Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.Z.: Learning multi-scale block local binary patterns for face recognition. In: International Conference on Biometrics. pp. 828–837 (2007)

19. Liu, Z., Luo, P., Wang, X., Tang, X.: Large-scale celebfaces attributes (celeba) dataset. Retrieved August **15**, 2018 (2018)
20. Lowe, G.: Sift-the scale invariant feature transform. Int. J **2**, 91–110 (2004)
21. Makrushin, A., Neubert, T., Dittmann, J.: Automatic generation and detection of visually faultless facial morphs. In: International Conference on Computer Vision Theory and Applications. vol. 7, pp. 39–50 (2017)
22. Mallick, S.: Face morph using opencv–c++/python (2016)
23. Ngan, M., Grother, P., Hanaoka, K., Kuo, J.: Face recognition vendor test (frvt) part 4: Morph performance of automated face morph detection. National Institute of Technology (NIST), Tech. Rep. NISTIR **8292** (2020)
24. Raghavendra, R., Raja, K.B., Busch, C.: Detecting morphed face images. In: 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS). pp. 1–7 (2016)
25. Raghavendra, R., Raja, K.B., Venkatesh, S., Busch, C.: 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). pp. 1822–1830 (2017)
26. Raghavendra, R., Raja, K., Venkatesh, S., Busch, C.: Face morphing versus face averaging: Vulnerability and detection. In: 2017 IEEE International Joint Conference on Biometrics (IJCB). pp. 555–563. IEEE (2017)
27. Robertson, D.J., Kramer, R.S., Burton, A.M.: Fraudulent id using face morphs: Experiments on human and automatic recognition. PLoS One **12**(3), e0173319 (2017)
28. Scherhag, U., Rathgeb, C., Busch, C.: Performance variation of morphed face image detection algorithms across different datasets. In: 2018 International Workshop on Biometrics and Forensics (IWBF). pp. 1–6 (2018)
29. Scherhag, U., Rathgeb, C., Busch, C.: Morph deterction from single face image: A multi-algorithm fusion approach. In: Proceedings of the 2018 2nd International Conference on Biometric Engineering and Applications. pp. 6–12 (2018)
30. Scherhag, U., Rathgeb, C., Busch, C.: Towards detection of morphed face images in electronic travel documents. In: 2018 13th IAPR International Workshop on Document Analysis Systems (DAS). pp. 187–192 (2018)
31. Scherhag, U., Rathgeb, C., Merkle, J., Busch, C.: Deep face representations for differential morphing attack detection. arXiv preprint arXiv:2001.01202 (2020)
32. Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 815–823 (2015)
33. Seibold, C., Samek, W., Hilsmann, A., Eisert, P.: Detection of face morphing attacks by deep learning. In: International Workshop on Digital Watermarking. pp. 107–120 (2017)
34. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE international conference on computer vision. pp. 618–626 (2017)
35. Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A.: Inception-v4, inception-resnet and the impact of residual connections on learning. In: Thirty-first AAAI conference on artificial intelligence (2017)
36. Wandzik, L., Garcia, R.V., Kaeding, G., Chen, X.: Cnns under attack: On the vulnerability of deep neural networks based face recognition to image morphing. In: International Workshop on Digital Watermarking. pp. 121–135. Springer (2017)

37. Wolberg, G.: Image morphing: a survey. The visual computer **14**(8-9), 360–372 (1998)
38. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters **23**(10), 1499–1503 (2016)
39. Zhou, B., Khosla, A., Lapedriza, A., Oliva, A., Torralba, A.: Learning deep features for discriminative localization. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2921–2929 (2016)