SYN-MAD 2022: Competition on Face Morphing Attack Detection Based on Privacy-aware Synthetic Training Data

Marco Huber[1,2,*], Fadi Boutros[1,2,*], Anh Thi Luu[1,*], Kiran Raja[3,*], Raghavendra Ramachandra[3,*], Naser Damer[1,2,*],
Pedro C. Neto[4,5,+], Tiago Gonçalves[4,5,+], Ana F. Sequeira[4,5,+], Jaime S. Cardoso[4,5,+], João Tremoço[6,+], Miguel Lourenço[6,+], Sergio Serra[7,+], Eduardo Cermeño[7,+], Marija Ivanovska[8,+], Borut Batagelj[8,+], Andrej Kronovšek[8,+], Peter Peer[8,+], Vitomir Štruc[8,+]
[1]Fraunhofer Institute for Computer Graphics Research IGD, Germany - [2]TU Darmstadt, Germany - [3]Norwegian University of Science and Technology, Norway - [4]INESC TEC, Portugal - [5]University of Porto, Portugal
[6]Yoonik, Portugal - [7]Vaelsys R&D, Spain - [8]University of Ljubljana, Slovenia
[*]Competition organizer [+]Competition participant
Email: marco.huber@igd.fraunhofer.de

2022 IEEE International Joint Conference on Biometrics (IJCB) | 978-1-6654-6394-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/IJCB54206.2022.10007950

Abstract

This paper presents a summary of the Competition on Face Morphing Attack Detection Based on Privacy-aware Synthetic Training Data (SYN-MAD) held at the 2022 International Joint Conference on Biometrics (IJCB 2022). The competition attracted a total of 12 participating teams, both from academia and industry and present in 11 different countries. In the end, seven valid submissions were submitted by the participating teams and evaluated by the organizers. The competition was held to present and attract solutions that deal with detecting face morphing attacks while protecting people's privacy for ethical and legal reasons. To ensure this, the training data was limited to synthetic data provided by the organizers. The submitted solutions presented innovations that led to outperforming the considered baseline in many experimental settings. The evaluation benchmark is now available at: https://github.com/marcohuber/SYN-MAD-2022.

1. Introduction

Current face recognition (FR) systems are vulnerable to different kinds of attacks, including so-called morphing attacks [22]. Morphing attacks combine two face images of two different individuals so that the resulting image can be, automatically or by human operators, matched to both individuals. If these morphed images are used for identity or travel documents, they would enable multiple subjects to be verified by the same documents, resulting in enormous security risks. To reduce this risk and counter these kinds of attacks, morphing attack detectors (MAD) [39] are developed to detect morphed images beside natural, unaltered bona fide images. These morphing attack detectors are often trained on morphed images based on real person images, raising the question of whether the privacy of the individual is respected as the given consent is often questioned [11]. Furthermore, the amount of morphs for training is often limited in terms of quantity and quality. There was previously a series of competitions on face presentation attacks that focused on spoofing attacks [6, 40]. However, this is the first competition on morphing attack detection and the first competition targeting any attack on face recognition system while limiting its development data to synthetic data.

The need for solutions to detect morphing attacks arose after uncovering the threat of morphing attacks [22] and the vulnerability of face recognition systems to these types of attacks [48, 13, 14]. As a result, a number of datasets for developing MAD systems were created [42, 44, 43, 15, 12, 8], which all used the privacy/legal/ethical sensitive face images of real individuals and were limited in quantity and variation. Most of these dataset consists of morph images based on interpolation of facial landmarks [42, 49, 23, 15]. More recently the use of datasets based on generative adversial networks (GAN) gained attention [13, 9, 16, 10, 56, 52].

Most of the recent face datasets have been collected from the web [45] which raises the question of consent of the individuals depicted in these images. However, due to legal and privacy issues [53, 41, 37] the use of face datasets collected from the web might not be possible in the future anymore. Privacy regulations such as the GDPR [53] assure individuals the right to withdraw their consent to use or store their private data, practically making the use and distribution of large face datasets impossible. These circumstances call for a solution that considers the privacy of individu-

als, which synthetically generated images can support. A recent work followed this motivation to take advantage of synthetic data to develop MADs in a privacy-friendly frame [11].

Driven by the legal and ethical concerns of using real images to develop MADs, and the need for well-generalized MADs developed on the basis of large and diverse datasets, we conducted the SYN-MAD 2022: Competition on Face Morphing Attack Detection Based on Privacy-aware Synthetic Training Data at the International Joint Conference on Biometrics 2022. The results and observations are summarized in this paper.

2. Datasets, Evaluation Criteria, and Participants

2.1. Provided Training Dataset

All participants were provided with the training set of the Synthetic Morphing Attack Detection Development (SMDD) dataset [11], and all solutions were allowed to use only this dataset to train the face morphing attack detector. The SMDD dataset was created using the official open-source implementation of StyleGAN2-ADA [30]. As detailed in [11], first, 500k images were generated using a random Gaussian noise vector drown from a normal distribution. These images were then randomly divided for training and test set, and then 50k images were selected based on their calculated face image quality. The face image quality was determined using CR-FIQA [3] and describes the utility of the images for face recognition, as defined in ISO/IEC 29794-1 [28]. The selected 50k images were again randomly split into two parts of equal size. The first part was considered the bona fide samples and was not used for morphing, while the second part was used to create the morphed images. 5k images of the second part were randomly chosen as the key morphing images and were mated with five randomly chosen images from the second part. For the morphing of the training images, the OpenCV/dlib morphing algorithm [36] is used for each pair.

In total, the training split of the SMDD dataset consists of 25k bona fide and 15k morphing images based on synthetic faces, which avoids using privacy-sensitive real face images. Furthermore, the bounding box and five facial landmark points of all the images were additionally provided to the participants.

2.2. Evaluation Benchmark

For the evaluation, a new MAD evaluation benchmark, MAD22, was created by the organizers as part of the competition and will be publicly released[1]. This dataset is based on the images of the Face Research Lab London (FRLL) dataset [17]. The FRLL dataset contains images of 102 different identities and provides high-quality close-up frontal face images created under uniform illumination with a wide range of ethnicities. All individuals in the dataset signed consent for their images to be "used in lab-based and web-based studies in their original or altered forms and to illustrate research (e.g., in scientific journals, news media or presentations)." The dataset itself has been released under the CC BY 4.0 license. For the morph generation, we limit the data to the frontal images of the dataset, consisting of "neutral_front" and "smiling_front" as most morphing approaches are optimized for frontal images.

2.2.1 Selection of the Morphing Pairs

For selecting the morphing pairs, we split the frontal images of FRLL depending on the provided gender label and retained the neutral and smiling split. We utilize ElasticFace-Arc [2] to generate embeddings of all the images and compare all possible pairs with cosine similarity. The pairs are then ordered depending on their achieved similarity score, and we select the 250 most similar pairs as the morphing pairs to obtain challenging, realistic morphs. Using the splits, we get a total of 1000 morphing pairs: 250 female neutral pairs, 250 female smiling pairs, 250 male neutral pairs, and 250 male smiling pairs.

2.2.2 Morphing Approaches

Starting from the 1000 image pairs selected by their similarity, we utilized five different morphing approaches, three landmark-based approaches, and two generative adversarial network (GAN)-based approaches to ensure diversity in the attacks.

The **OpenCV** algorithm is based on the "Face Morph Using OpenCV" tutorial [2] and uses the Dlib [34] implementation of the landmark detector proposed by Kazemi and Sullivan [33]. The detected landmarks are used to perform Delaunay triangulation on each image, and based on affine transformation, the triangles are wrapped and blended. 16 out of the 1000 morph images showed strong morphing artifacts (a black region on the mouth area). These images are removed, resulting in 984 morph images based on the OpenCV algorithm. The same morphing approach has also been used in the development dataset to create the morphs of the synthetic training dataset, the SMDD.

The **FaceMorpher** is also a landmark-based morphing tool. It is a commercial-of-the-shelf (COTS) face morphing tool. More details on the technical aspects of the approach are not released.

Webmorph is a landmark-based online tool[3] optimized for averaging and transforming faces. It can also be used to create morphs of two images. As the program often failed to create morphs without huge artifacts for images with smiling, we only created morphs of the neutral frontal images, resulting in just 500 morphed images based on the Webmorph face morphing approach.

[1] https://github.com/marcohuber/SYN-MAD-2022

[2] https://learnopencv.com/face-morph-using-opencv-cpp-python/

[3] https://webmorph.org/

For approaches based on GANs [31], we selected **MIPGAN-I** and **MIPGAN-II** [56]. MIPGAN utilizes a loss function that incorporates perceptual quality, identity, identity differences and structural visibility into StyleGAN, either StyleGAN [31] (MIPGAN-1) or StyleGAN2 [32] (MIPGAN-2). Both approaches produce high-resolution morphs with minimal artifacts. In one case, the detection of facial landmarks failed on a morphed image, which reduces the amount of morphs created with the MIPGAN-2 approach to 999.

The created MAD22 benchmark contains 4483 (984 OpenCV, 1000 FaceMorpher, 500 Webmorph, 1000 MIPGAN-I, 999 MIPGAN-II) morphed face images and 204 bona fide images from the FRLL dataset. Some example images of the synthetic training dataset, as well as the created MAD22 benchmark, are shown in Figure 1.

2.2.3 Baseline Approach

The baseline performance evaluation is based on MixFaceNet [11, 1] It has shown good performances in face presentation attack and morphing attack detection and possesses a low computational complexity while retaining high accuracy. The architecture also takes advantage of convolutional kernels of different sizes to capture different levels of face attack clues [21].

2.3. Evaluation Criteria

For the vulnerability analysis, we look at the False Non-Match Rate (FNMR) at two different, fixed False match rates (FMR), FMR100 and FMR1000, to show the strength of the attacks. The FMR100 refers to the operational point that achieves the lowest FNMR with an FMR $< 1.0\%$ and the FMR1000 to the operational point with FMR $< 0.1\%$. To investigate the vulnerability of the used face recognition models, we calculate the Mated Morph Presentation Match Rate (MMPMR) [48]. The MMPMR refers to the fraction of morphs whose similarity to both identities used to morph, are below the selected threshold relative to all morphs. In other words, the proportion of morphs that match both identities used to morph out of all morphs.

The evaluation of the morphing attack detector performance will be based on the ISO/IEC 30107-3 [27] standard and, therefore, presented as the Bona fide Presentation Classification Error Rate (BPCER) and the Attack presentation Classification Error Rate (APCER). The BPCER refers to the proportion of bona fide images incorrectly classified as attack samples, and the APCER refers to the proportion of attack images incorrectly classified as bona fide samples. Furthermore, we will report the APCER at a fixed BPCER to provide a more detailed analysis of the different approaches and their performance. To cover different operational points and to present comparative results, the submitted solutions are evaluated at three different fixed APCER (BPCER) values, 0.1%, 1.0%, 10%, and 20%, and the corresponding BPCER (APCER) is reported. The final ranking on the submitted solutions is based on the APCER at BPCER of 20%. As the bona fide samples are the same for all attack types, fixing the BPCER would produce the same operational threshold. This will allow us to analyze the detectability of the different attacks at the same operation threshold, and thus the APCER at fixed BPCER is considered for the most of the discussion and the final ranking.

2.4. Vulnerability Analysis

To investigate the success of the created morphing attacks, we performed a vulnerability analysis using two different state-of-the-art face recognition models, CurricularFace [26] and ElasticFace-Arc [2] [4]. To show the suitability of the chosen face recognition models, we also report their performance in terms of FNMR@FMR (in percentage) on the Labeled Faces in the Wild (LFW) [25] benchmark. On LFW, CurricularFace [26] achieved 0.3% at FMR100 and 0.33% at FMR1000. ElasticFace-Arc, in comparison, achieved 0.23% at FMR100 and 0.3% at FMR1000. Both models achieved top performance on mainstream benchmarks as reported in [2]. The FMR1000 and FMR100 decision thresholds used in the vulnerability study were calculated on the LFW benchmark. For the analysis of the vulnerability, we take the morphed image M_{A_n,B_n} of two images from two identities from one scenario (e.g. neutral), ID_{A_n} and ID_{B_n} and compare them to their counterpart in the other scenario (e.g. smiling), ID_{A_S} and ID_{B_s}. If the similarity of the morphed image M_{A_n,B_n} to the original images ID_{A_S} and ID_{B_s} is higher than the set threshold, the system is considered vulnerable to this attack. The results of the vulnerability analysis are summarized in Table 1. The results show the vulnerability (MMPMR) of the face recognition models to the morphed images, as in most cases, over 90% of the morphed images are matched with both identities used to create the morphed image, even at high similarity thresholds.

2.5. Competition Participants

The competition aimed at attracting participants, both from academia and industry, with a high geographic and activity variation. The call for participation was shared on the International Joint Conference on Biometrics 2022 website, on the competition's own website, on several social media platforms, and through private e-mailing lists. The call for participation has attracted 12 registered teams from both academia and industry. Out of these, six teams submitted a valid solution. These six teams have affiliations in four different countries. Three teams have academic affiliations, two teams have industry affiliations, and one team has mixed, both academic and industry backgrounds. Only one team has chosen to be anonymous, and one team submitted

[4] CurricularFace and ElasticFace utilized ResNet100 as backbone trained on MS1MV dataset

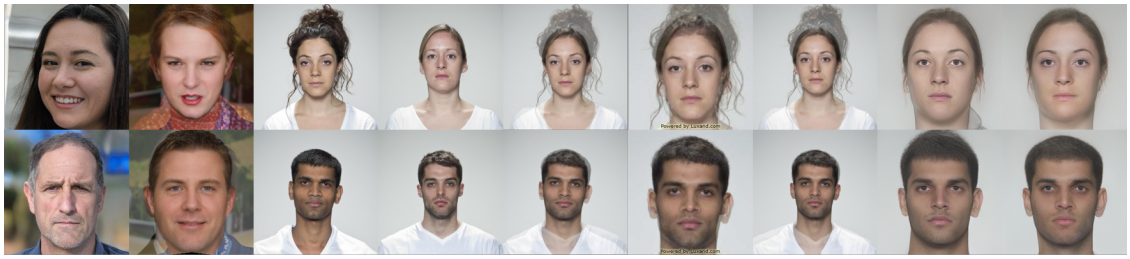

Figure 1. Examples of the used images during the competition: The synthetic bona fide and synthetic morphs were provided to the competition participants and are part of the training split of the SMDD dataset [11]. The bona fide images are taken from the FRLL dataset [17] and they also serve as the images used for the different morphing approaches to create our MAD22 benchmark. The evaluation morphs were created by the competition organizers and will be released to the public.

Morphing Approach	FR Model	T_{FMR100}	$T_{FMR1000}$
OpenCV	CF	0.996	0.986
	EF-Arc	0.997	0.980
FaceMorpher	CF	0.970	0.935
	EF-Arc	0.962	0.913
Webmorph	CF	0.988	0.988
	EF-Arc	0.990	0.986
MIPGAN-I	CF	0.962	0.890
	EF-Arc	0.980	0.845
MIPGAN-II	CF	0.953	0.832
	EF-Arc	0.953	0.778

Table 1. Vulnerability analysis in terms of MMPMR at two different operational points (FMR100 and FMR1000) on two face recognition (FR) models, CurricularFace (CF) and ElasticFace-Arc (EF-Arc). The results show that both face recognition models are vulnerable to the different morphing approaches as in most cases, even with high thresholds in terms of required similarity to be matched, the morphed images are matched with both identities used to create the morph in over 90% of the cases.

two solutions as each team was allowed to submit up to two solutions. The total number of validly submitted solutions is 7. A summary of the participating teams is presented in Table 2.

2.6. Submission and Evaluation Process

Each team had to request the synthetic data first and provide their team name and affiliation. Only the synthetic data, consisting of the synthetic morphs and synthetic bona fide images as well as text files containing information about the bounding box and facial landmarks for each image, has been provided. Each team was then requested to provide a Win32 or a Linux executable or, if wanted, provide their python script. The teams had to provide an executable to evaluate their approach on the classified evaluation benchmark and also an executable to re-train their model. The re-training executable is to be used by the organizing team to validate that solely the provided synthetic data is used to train the approach, and no other pre-trained weights of any kind are used. All the approaches have been evaluated on a restricted system without an internet connection to prevent any data leak. At no time the participating teams had any access or knowledge about the evaluation benchmark used during the competition, including any information about the morphing approaches used or the images source dataset.

3. Submitted Solutions

In total, 12 teams have registered for the competition. Each team was allowed to submit up to two submissions. Eventually, seven valid submissions from six different teams were received. Solution names, team members, affiliations, and type of the institution, i.e., academic, industry, or mixed, are summarized in Table 2. One team opted to keep their names and affiliations anonymized. A condensed summary of the details of the approaches (e.g., input size, architecture, loss function, etc.) is listed in Table 3. In the following, we provide a brief description of the valid submitted solutions:

Xception[5]: This approach utilizes the Xception-net architecture [7] modified by an additional 2-class classification layer on top of the network. The choice of Xception network is based on its high performance in detecting DeepFakes [46]. The approach is trained for 39 epochs with 2500 iterations per epoch utilizing Adam optimizer [35]. The learning rate is set to 0.0001, and a weight decay to 0.001 is applied. Instead of using the provided landmarks and bounding boxes, face detection is done based on RetinaFace detector [18]. During the development, the approach has been evaluated on FRLL morphs, FERET morphs, and FRGC morphs [47]. The input to the model is a 40% enlarged frame of the found face. The face image is then scaled to 299×299 pixels and randomly flipped horizontally. All images are then also normalized across all channels. More details on the approach are presented in [29].

MorphHRNet[5]: This approach is based on training an HRNet [51, 54] to perform two-class classification i.e, bona

[5]The research was supported in parts by the Slovenian Research Agency (ARRS) in the scope of the research project J2-1734 (B) "FaceGEN"

Solution	Team Members	Affiliations	Type
Xception	Borut Batagelj, Andrej Kronovšek, Peter Peer	Faculty of Computer and Information Science, University of Ljubljana	Academic
MorphHRNet	Marija Ivanovska, Borut Batagelj, Peter Peer, Vitomir Štruc	Faculty of Electrical Engineering, Faculty of Computer and Information Science, University of Ljubljana	Academic
E-CBAM@VCMI	Pedro C. Neto, Tiago Gonçalves, Ana F. Sequeira, Jaime S. Cardoso	INESC TEC, Faculty of Engineering, University of Porto	Academic
Con-Text Net (A/B)	João Tremoço, Miguel Lourenço	Yoonik	Industry
VaMoS	Sergio Serra, Eduardo Cermeño	Vaelsys R&D	Industry
Anonymous	Anonymous	Anonymous	mixed

Table 2. A summary of the valid submitted solutions, participating team members, affiliations, and type of institution. More details on the submitted algorithms are provided in Section 3.

Solution	**Input size (in pixels)**	**Architecture/ Model**	**Loss function**	**Evaluation data during training**	**Bounding box/ Landmarks?**
MorphHRNet	256×256	HRNet	Binary cross entropy	20% of provided data, FRLL, FERET and FRGC morphs	bounding box
Con-Text Net A	128×128 and 256×256	2xResNet-50	Weighted Focal	New Face Morphing Dataset	landmarks
Xception	299×299	Xception net	Cross-entropy	FRLL, FERET and FRGC morphs	other alignment
Con-Text Net B	128×128 and 256×256	2xResNet-50	Weighted Focal	New Face Morphing Dataset	landmarks
E-CBAM@VCMI	256×256	4xResNet-50 with CBAM	Cross-entropy	25% of the data for different models	no alignment
Anonymous	256×256	Hand-crafted+SVM	—	no other data	bounding box
VaMoS	112×112	ResNet-50	Additive Angular Margin	only provided data	bounding box + landmarks

Table 3. Basic details of the submitted approaches. ResNet-50 is the most common utilized model architecture in the competition and the used input size varies. More details on the submitted algorithms are provided in Section 3.

fide or morphing attack. The approach is trained using binary cross-entropy and optimized using the Adam optimizer [35]. During training, the model is evaluated on a combination of the FRLL morphs, FERET morphs, and FRGC morphs [47] and also 20% of the provided synthetic data. The best-performing model based on the validation dataset is selected. The model is trained for 20 epochs with a learning rate of 0.001. Furthermore, the provided bounding boxes have been used during pre-processing. As preprocessing, the images have been resized and normalized. For data augmentation, horizontal flipping and random rotations up to 5 degrees have been applied. More details on the approach are presented in [29].

Anonymous: The face morphing detector exploits the idea introduced by [38]. It is based on the idea of the change caused by JPEG compression in morphed images in comparison to bona fide images. In total, 46 features are extracted from the original images and the compressed image. On these extracted features, a support vector machine (SVM) is trained with the SMO solver. During the feature creation process, the provided bounding boxes are utilized. The team emphasizes that the approach will only work on very similar images to the synthetic training images, and no effort has been made to develop a generalizable approach.

E-CBAM@VCMI: This approach is based on training ResNet-50 architecture [24] with attention module [55] on the provided dataset. The provided dataset is divided into 4 splits with 25% of the data without overlapping. Four different models are then trained on 75% of the provided training dataset and are validated on a different split. For the evaluation, all the images are processed by the four different models for the score computation, and the mean value is taken as the final score. The models are optimized using cross-entropy loss and are trained for 20 epochs each. The learning rate was set to 0.0001. No bounding boxes or alignment has been used. Besides center cropping the images to 256×256, along with normalization, random affine transformation, and random horizontal splits have been performed.

Con-Text Net A/B: The approach consisted of two ResNet-50 backbones [24]. One backbone is trained using input images cropped to the size 128×128 using the provided landmarks. The other backbone is trained using the original image resized to 256×256. The linearized outputs of each backbone are 2048-dimensional feature vectors than are then fed to a classifier with three fully connected layers

	Morphing approach	OpenCV						
Rank	Solution	**APCER@ BPCER 20%**	APCER@ BPCER 10%	APCER@ BPCER 1%	BPCER@ APCER 20%	BPCER@ APCER 10%	BPCER@ APCER 1%	EER
-	Baseline	0.0376	0.065	0.3638	0.0345	0.0739	0.532	0.0833
1	MorphHRNet	**0.0163**	0.0376	0.6697	0.0147	0.0196	0.3382	0.0569
2	Xception	0.0254	0.0661	0.2175	0.0147	0.049	0.3529	0.0732
3	Con-Text Net A	0.1575	0.2652	0.7093	0.1618	0.3284	0.7402	0.1748
4	Con-Text Net B	0.2368	0.3445	0.5783	0.2353	0.4314	0.8235	0.2266
5	E-CBAM@VCMI	0.3303	0.4868	0.9858	0.3824	0.5049	0.7647	0.2754
6	Anonymous	0.4461	0.999	1.000	0.4363	0.5049	0.6078	0.3252
7	VaMoS	–	–	–	0.9118	0.9853	1.000	–

Table 4. The evaluation results on the landmark-based morph dataset utilizing the OpenCV morphing approach. The same morphing approach was used to create the synthetic morphs of the training dataset.

	Morphing approach	FaceMorpher						
Rank	Solution	**APCER@ BPCER 20%**	APCER@ BPCER 10%	APCER@ BPCER 1%	BPCER@ APCER 20%	BPCER@ APCER 10%	BPCER@ APCER 1%	EER
-	Baseline	0.029	0.036	0.055	0.0049	0.0049	1.000	0.046
1	Con-Text Net B	**0.000**	0.000	0.000	0.000	0.000	0.000	0.000
2	Xception	0.000	0.000	0.005	0.0147	0.0147	0.0147	0.006
3	Con-Text Net A	0.000	0.000	0.999	0.000	0.000	1.000	0.000
4	MorphHRNet	0.024	0.043	0.312	0.0147	0.0196	0.4804	0.059
5	E-CBAM@VCMI	0.628	0.928	1.000	0.8235	0.951	1.000	0.412
6	Anonymous	0.999	0.999	1.000	0.7255	0.7404	1.000	0.653
7	VaMoS	–	–	–	0.9069	0.9804	1.000	–

Table 5. The evaluation results on the landmark-based morphs created with the COTS system FaceMorpher. When investigating the details, it is noticeable that some systems find it very easy to recognize the morphs, resulting in no errors even at very restrictive thresholds.

with SiLU activations [20]. Both models are trained with a batch size of 24 with Adam optimizer [35]. The learning rate was initially set to 0.01 and decreased with an exponential learning rate scheduler with factor of 0.85. The utilized learning objective was weighted focal loss. During development, the Biometix "New Face Morphing Dataset (for vulnerability research)" [6] has been utilized to evaluate the approach. The Con-Text Net A has been trained for 18 epochs, and the Con-Text Net B has been trained for 22 epochs.

VaMoS: This approach is based on previous works on the concealment of morphing attacks [5] and solutions to prevent registration using fake images [4]. In comparison to the previous mentioned works, this work utilized ArcFace loss [19] instead of inception net trained with triplet loss (FaceNet) [50]. Furthermore, the neural network architecture has been modified to not longer require a morphing detector [43]. The model architecture is based on the ResNet-50 [24] architecture and is trained for 100 epochs with a learning rate of 0.001 with ArcFace loss [19]. The approach utilizes the provided bounding boxes and face landmarks.

[6]https://www.linkedin.com/pulse/new-face-morphing-dataset-vulnerability-research-ted-dunstone/

4. Results and Analysis

This section presents the evaluation results of the submitted solutions and the baseline approach on the five different morphing approach datasets presented in Section 2.2.2 in terms of the BPCER at a fixed APCER and the performance in terms of APCER at a fixed BPCER. We furthermore also report the detection Equal Error Rate (EER). We first present and discuss the results on the Landmark-based morphing approach datasets (OpenCV, FaceMorpher, and Webmorph) and then on the GAN-based morphing approach datasets (MIPGAN-I and MIPGAN-II). The final ranking of the submitted solutions is shown in Table 9. None of the submitted solutions consistently outperformed all other solutions on all benchmarks. Considering the average ranking on all benchmarks, MorphHRNet achieved the top-1 rank.

4.1. Landmark-based Morphing Approaches

The results on the landmark-based morphing approaches of the submitted solutions are shown in Table 4 for the morphs based on the OpenCV approach, in Table 5 for the FaceMorpher-based morphs, and in Table 6 for the morphs created using the Webmorph online tool.

The **MorphHRNet** and the **Xception** model outperform the baseline on all three landmark-based morph datasets

	Morphing approach	Webmorph						
Rank	Solution	**APCER@ BPCER 20%**	APCER@ BPCER 10%	APCER@ BPCER 1%	BPCER@ APCER 20%	BPCER@ APCER 10%	BPCER@ APCER 1%	EER
-	Baseline	0.176	0.24	0.74	0.1576	0.4138	1.000	0.182
1	MorphHRNet	**0.042**	0.112	0.902	0.0392	0.1078	0.5686	0.098
2	Xception	0.108	0.23	0.494	0.1176	0.2157	0.5392	0.146
3	Con-Text Net A	0.31	0.456	0.892	0.3186	0.4853	0.9314	0.262
4	Con-Text Net B	0.438	0.596	0.81	0.4412	0.5539	0.9461	0.314
5	E-CBAM@VCMI	0.468	0.868	0.99	0.3824	0.4706	0.6912	0.306
6	Anonymous	1.000	1.000	1.000	0.8627	0.8775	0.8971	0.806
7	VaMoS	–	–	–				–

Table 6. The evaluation results on the dataset consisting of morphs created with the online tool Webmorph. Most systems had greater problems recognizing the morphs in comparison to the other landmark-based approaches.

	Morphing approach	MIPGAN-I						
Rank	Solution	**APCER@ BPCER 20%**	APCER@ BPCER 10%	APCER@ BPCER 1%	BPCER@ APCER 20%	BPCER@ APCER 10%	BPCER@ APCER 1%	EER
-	Baseline	0.145	0.222	0.758	0.1281	0.33	1.000	0.167
1	Con-Text Net A	**0.081**	0.141	0.419	0.0637	0.1618	0.5931	0.123
2	MorphHRNet	0.13	0.219	0.898	0.1127	0.2402	0.7598	0.153
3	Con-Text Net B	0.398	0.53	0.715	0.4559	0.6029	0.8775	0.303
4	Xception	0.574	0.804	0.979	0.4902	0.5735	0.8627	0.369
5	E-CBAM@VCMI	0.602	0.849	0.999	0.4069	0.5343	0.7892	0.325
6	Anonymous	0.62	0.756	0.92	0.4804	0.4951	0.5	0.363
7	VaMoS	–	–	–	0.902	0.902	0.9853	–

Table 7. The evaluation results on the GAN-based MIPGAN-I morph dataset. For the most systems, recognizing GAN-based morphs seems to be harder than recognizing landmark-based morphs.

in terms of APCER@BPCER=20%. On the OpenCV and the Webmorph benchmarks (Table 4, 6) they are the top-2 performing submissions. On the Webmorph dataset, the MorphHRNet significantly outperformed the other submitted approaches. Both **Con-Text Net** models, **A** and **B** are top-performing solutions on the FaceMorpher dataset in terms of APCER@BPCER=20% (Table 5). On the other landmark-based morph datasets, Webmorph (Table 6) and OpenCV (Table 4), they achieved lower performance than MorphHRNet, Xception, and the baseline. The **E-CBAM@VCMI model** is ranked behind MorphHRNet, Xception, Con-Text Net A and B, and achieved lower performance than the baseline on the landmark-based morphing datasets as shown in Tables 4, 5, 6. The **Anonymous** model, based on hand-crafted features, shows interesting behavior, as it shows only some ability to distinguish between morphs and bona fide on the OpenCV dataset. This behavior has been predicted by the team, as their approach has been specialized based on the provided synthetic morphs, which were created using the OpenCV morphing approach. The evaluation of the **VaMoS** approach was impossible, especially when fixing BPCER. This is due to the fact that almost for all bona fide samples, and most of the attacks, the solution resulted in the same exact detection score, which makes setting up a threshold at a specific BPCER impossible. In such situations, the results in the relative tables are marked with "-". On the test data used by the submitting team, the organizers were able to reproduce the desired results. The team's assumption is strong over-fitting of the model, but without deeper insights into the specific details, the organizing team cannot make a more specific statement.

4.2. GAN-based Morphing Approaches

The results on the GAN-based morphing approaches of the submitted solutions are shown in Table 7 for the morphs based on the MIPGAN-I morphing approach and in Table 8 for the MIPGAN-II-based morphs.

MorphHRNet achieved the best performance on MIPGAN-II and ranked second on MIPGAN-I, as shown in Table 7, 8. Xception, which achieved very competitive results on landmark-based morphing datasets, achieved relatively very low performance on GAN-based morphing datasets. Similar behavior can be seen for the **Con-Text Net** models. While the Con-Text Net model trained on less epochs, **A**, outperforms the baseline on both datasets, the model trained for more epochs (**B**) struggles to reach the baseline performance. The performance of the **E-**

	Morphing approach	MIPGAN-II						
Rank	Solution	**APCER@ BPCER 20%**	APCER@ BPCER 10%	APCER@ BPCER 1%	BPCER@ APCER 20%	BPCER@ APCER 10%	BPCER@ APCER 1%	EER
-	Baseline	0.2062	0.3203	0.8158	0.2118	0.3941	1.000	0.2062
1	MorphHRNet	**0.0611**	0.1101	0.8418	0.0294	0.1127	0.6127	0.1041
2	Con-Text Net A	0.0861	0.1451	0.4344	0.0588	0.1961	0.5931	0.1291
3	E-CBAM@VCMI	0.3734	0.6466	0.996	0.3039	0.3039	0.5686	0.2593
4	Con-Text Net B	0.3914	0.5125	0.6767	0.4706	0.6176	0.9118	0.2943
5	Anonymous	0.7548	0.8539	0.975	0.4951	0.5000	0.5637	0.4334
6	Xception	0.7708	0.9249	0.995	0.5686	0.6765	0.902	0.4454
7	VaMoS	–	–	–	0.902	0.902	0.9314	–

Table 8. The evaluation results on the GAN-based MIPGAN-II morph dataset. Similar to the results on the MIPGAN-I dataset, recognizing GAN-based morphs seems to be harder than recognizing landmark-based morphs for the most face morphing detector systems.

	Ranks						
Solution	OpenCV	FaceMorpher	Webmorph	MIPGAN-I	MIPGAN-II	Avg	Final rank
MorphHRNet	1	4	1	2	1	1.8	1
Con-Text Net A	3	3	3	1	2	2.4	2
Xception	2	2	2	4	6	3.2	3
Con-Text Net B	4	1	4	3	4	3.2	3
E-CBAM@VCMI	5	5	5	5	3	4.6	4
Anonymous	6	6	6	6	5	5.8	5
VaMoS	7	7	7	7	7	7	6

Table 9. The final ranking of the submitted solutions based on the average rank of the performance on the five morphing datasets.

CBAM@VCMI approach, as well as the **Anonymous** hand-crafted approach, does not reach the baseline performance, although the E-CBAM@VMCI model outperforms the Xception and the Con-Text Net B model on the MIPGAN-II dataset. The **VaMoS** approach did again not allow an evaluation at APCER@BPCER=20% or any other operational points.

4.3. Comparison and Final Ranking

In this subsection, we briefly investigate the performance differences of the submitted solutions on landmark-based and deep learning-based morphing approaches and also investigate the final ranking based on the average rank achieved on all five morphing benchmarks. The final ranking is presented in Table 9.

For the **MorphHRNet** approach, we can observe that the model performs well independently of the used morphing approach, both, landmark-based and GAN-based face morphs are relatively well detected by the model. The **Con-Text Net A** model especially performs well on GAN-based face morphs and lacks some performance on landmark-based morphs. The **Xception** model, which showed good performance on the landmark-based approach, did not classify the GAN-based morphs well, as can be observed in Tables 7 and 8. Similar behavior can be observed for the **Con-Text Net B** model. For the **E-CBAM@VCMI** model, no specific change in performance was observable regarding being used for either type of morphing approaches, landmark, or GAN-based. No particular behavior can be observed with regard to the performance of **Anonymous** and **VaMoS** besides their lack of power to distinguish between morphs and bona fide face images.

In the final ranking, the **MorphHRNet** won the competition with an average rank of 1.8, the second place was achieved by the **Con-Text Net A** model (average rank 2.4), and the third place is shared between the **Xception** model and the **Con-Text Net B** model (both average rank 3.2), as detailed in Table 9.

5. Conclusion

In this paper, we summarized the results and observations of the SYN-MAD 2022: Competition on Face Morphing Attack Detection Based on Privacy-aware Synthetic Training Data. In total, 12 teams registered for participation, and six of them submitted seven valid submissions to tackle the problem of face morphing detection while considering the privacy and legal issues related to real development data. The evaluation focused on a wide range of different morphing approaches, including landmark- and GAN-based approaches, and various creative solutions have been evaluated, leading to enhanced performances in comparison to the considered baseline.

Acknowledgments

This research work has been funded by the German Federal Ministry of Education and Research and the Hessen State Ministry for Higher Education, Research and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.

References

[1] F. Boutros, N. Damer, M. Fang, F. Kirchbuchner, and A. Kuijper. Mixfacenets: Extremely efficient face recognition networks. In *IJCB*, pages 1–8. IEEE, 2021.

[2] F. Boutros, N. Damer, F. Kirchbuchner, and A. Kuijper. Elasticface: Elastic margin loss for deep face recognition. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1578–1587, June 2022.

[3] F. Boutros, M. Fang, M. Klemt, B. Fu, and N. Damer. CR-FIQA: face image quality assessment by learning sample relative classifiability. *CoRR*, abs/2112.06592, 2021.

[4] L. Cárabe and E. Cermeno. Methods to prevent registration using fake face images.

[5] L. Cárabe and E. Cermeño. Stegano-morphing: Concealing attacks on face identification algorithms. *IEEE Access*, 9:100851–100867, 2021.

[6] I. Chingovska, J. Yang, Z. Lei, D. Yi, S. Z. Li, O. Kähm, C. Glaser, N. Damer, A. Kuijper, A. Nouak, J. Komulainen, T. F. Pereira, S. Gupta, S. Khandelwal, S. Bansal, A. Rai, T. Krishna, D. Goyal, M. Waris, H. Zhang, I. Ahmad, S. Kiranyaz, M. Gabbouj, R. Tronci, M. Pili, N. Sirena, F. Roli, J. Galbally, J. Fiérrez, A. da Silva Pinto, H. Pedrini, W. S. Schwartz, A. Rocha, A. Anjos, and S. Marcel. The 2nd competition on counter measures to 2d face spoofing attacks. In *ICB*, pages 1–6. IEEE, 2013.

[7] F. Chollet. Xception: Deep learning with depthwise separable convolutions. In *CVPR*, pages 1800–1807. IEEE Computer Society, 2017.

[8] N. Damer, V. Boller, Y. Wainakh, F. Boutros, P. Terhörst, A. Braun, and A. Kuijper. Detecting face morphing attacks by analyzing the directed distances of facial landmarks shifts. In *GCPR*, volume 11269 of *Lecture Notes in Computer Science*, pages 518–534. Springer, 2018.

[9] N. Damer, F. Boutros, A. M. Saladie, F. Kirchbuchner, and A. Kuijper. Realistic dreams: Cascaded enhancement of gan-generated images with an example in face morphing attacks. In *BTAS*, pages 1–10. IEEE, 2019.

[10] N. Damer, J. H. Grebe, S. Zienert, F. Kirchbuchner, and A. Kuijper. On the generalization of detecting face morphing attacks as anomalies: Novelty vs. outlier detection. In *BTAS*, pages 1–5. IEEE, 2019.

[11] N. Damer, C. A. F. López, M. Fang, N. Spiller, M. V. Pham, and F. Boutros. Privacy-friendly synthetic data for the development of face morphing attack detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 1606–1617, June 2022.

[12] N. Damer, K. B. Raja, M. Süßmilch, S. Venkatesh, F. Boutros, M. Fang, F. Kirchbuchner, R. Ramachandra, and A. Kuijper. Regenmorph: Visibly realistic GAN generated face morphing attacks by attack re-generation. In *ISVC (1)*, volume 13017 of *Lecture Notes in Computer Science*, pages 251–264. Springer, 2021.

[13] N. Damer, A. M. Saladie, A. Braun, and A. Kuijper. Morgan: Recognition vulnerability and attack detectability of face morphing attacks created by generative adversarial network. In *BTAS*, pages 1–10. IEEE, 2018.

[14] N. Damer, A. M. Saladie, S. Zienert, Y. Wainakh, P. Terhörst, F. Kirchbuchner, and A. Kuijper. To detect or not to detect: The right faces to morph. In *ICB*, pages 1–8. IEEE, 2019.

[15] N. Damer, N. Spiller, M. Fang, F. Boutros, F. Kirchbuchner, and A. Kuijper. PW-MAD: pixel-wise supervision for generalized face morphing attack detection. In *ISVC (1)*, volume 13017 of *Lecture Notes in Computer Science*, pages 291–304. Springer, 2021.

[16] L. Debiasi, N. Damer, A. M. Saladie, C. Rathgeb, U. Scherhag, C. Busch, F. Kirchbuchner, and A. Uhl. On the detection of gan-based face morphs using established morph detectors. In *ICIAP (2)*, volume 11752 of *Lecture Notes in Computer Science*, pages 345–356. Springer, 2019.

[17] L. DeBruine and B. Jones. Face Research Lab London Set. 4 2021.

[18] J. Deng, J. Guo, E. Ververas, I. Kotsia, and S. Zafeiriou. Retinaface: Single-shot multi-level face localisation in the wild. In *CVPR*, pages 5202–5211. Computer Vision Foundation / IEEE, 2020.

[19] J. Deng, J. Guo, N. Xue, and S. Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *CVPR*, pages 4690–4699. Computer Vision Foundation / IEEE, 2019.

[20] S. Elfwing, E. Uchibe, and K. Doya. Sigmoid-weighted linear units for neural network function approximation in reinforcement learning. *CoRR*, abs/1702.03118, 2017.

[21] M. Fang, F. Boutros, A. Kuijper, and N. Damer. Partial attack supervision and regional weighted inference for masked face presentation attack detection. In *FG*, pages 1–8. IEEE, 2021.

[22] M. Ferrara, A. Franco, and D. Maltoni. The magic passport. In *IJCB*, pages 1–7. IEEE, 2014.

[23] M. Ferrara, A. Franco, and D. Maltoni. Face demorphing. *IEEE Trans. Inf. Forensics Secur.*, 13(4):1008–1017, 2018.

[24] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *CVPR*, pages 770–778. IEEE Computer Society, 2016.

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

[26] Y. Huang, Y. Wang, Y. Tai, X. Liu, P. Shen, S. Li, J. Li, and F. Huang. Curricularface: Adaptive curriculum learning loss for deep face recognition. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 5900–5909. Computer Vision Foundation / IEEE, 2020.

[27] International Organization for Standardization. ISO/IEC DIS 30107-3:2016: Information Technology – Biometric presentation attack detection – P. 3: Testing and reporting, 2017.

[28] ISO/IEC JTC1 SC37 Biometrics. ISO/IEC 29794-1:2016 Information technology - Biometric sample quality - Part 1: Framework. International Organization for Standardization, 2016.

[29] M. Ivanovska, A. Kronovšek, P. Peer, V. Štruc, and B. Batagelj. Face morphing attack detection using privacy-aware training data. In *Proceedings of the 31st International Electrotechnical and Computer Science Conference (under review)*, 2022.

[30] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen, and T. Aila. Training generative adversarial networks with limited data. In *NeurIPS*, 2020.

[31] T. Karras, S. Laine, and T. Aila. A style-based generator architecture for generative adversarial networks. In *CVPR*, pages 4401–4410. Computer Vision Foundation / IEEE, 2019.

[32] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of stylegan. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pages 8107–8116. Computer Vision Foundation / IEEE, 2020.

[33] V. Kazemi and J. Sullivan. One millisecond face alignment with an ensemble of regression trees. In *2014 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, Columbus, OH, USA, June 23-28, 2014*, pages 1867–1874. IEEE Computer Society, 2014.

[34] D. E. King. Dlib-ml: A machine learning toolkit. *J. Mach. Learn. Res.*, 10:1755–1758, 2009.

[35] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *ICLR (Poster)*, 2015.

[36] S. Mallick. Face morph using opencv — c++ / python. *LearnOpenCV*, 1(1), 2016.

[37] B. Meden, P. Rot, P. Terhörst, N. Damer, A. Kuijper, W. J. Scheirer, A. Ross, P. Peer, and V. Struc. Privacy-enhancing face biometrics: A comprehensive survey. *IEEE Trans. Inf. Forensics Secur.*, 16:4147–4183, 2021.

[38] T. Neubert. Face morphing detection: An approach based on image degradation analysis. In *IWDW*, volume 10431 of *Lecture Notes in Computer Science*, pages 93–106. Springer, 2017.

[39] M. Ngan, P. Grother, K. Hanaoka, and J. Kuo. Face Recognition Vendor Test (FRVT) Part 4: MORPH - Performance of Automated Face Morph Detection. *National Institute of Standards and Technology (NIST)*, 2021.

[40] S. Purnapatra, N. Smalt, K. Bahmani, P. Das, D. Yambay, A. Mohammadi, A. George, T. Bourlai, S. Marcel, S. Schuckers, M. Fang, N. Damer, F. Boutros, A. Kuijper, A. Kantarci, B. Demir, Z. Yildiz, Z. Ghafoory, H. Dertli, H. K. Ekenel, S. Vu, V. Christophides, D. Liang, G. Zhang, Z. Hao, J. Liu, Y. Jin, S. Liu, S. Huang, S. Kuei, J. M. Singh, and R. Ramachandra. Face liveness detection competition (livdet-face) - 2021. In *IJCB*, pages 1–10. IEEE, 2021.

[41] H. Qiu, B. Yu, D. Gong, Z. Li, W. Liu, and D. Tao. Synface: Face recognition with synthetic data. In *ICCV*, pages 10860–10870. IEEE, 2021.

[42] R. Raghavendra, K. B. Raja, and C. Busch. Detecting morphed face images. In *BTAS*, pages 1–7. IEEE, 2016.

[43] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch. Face morphing versus face averaging: Vulnerability and detection. In *IJCB*, pages 555–563. IEEE, 2017.

[44] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch. Transferable deep-cnn features for detecting digital and print-scanned morphed face images. In *CVPR Workshops*, pages 1822–1830. IEEE Computer Society, 2017.

[45] I. D. Raji and G. Fried. About face: A survey of facial recognition evaluation. *CoRR*, abs/2102.00813, 2021.

[46] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner. Faceforensics++: Learning to detect manipulated facial images. In *ICCV*, pages 1–11. IEEE, 2019.

[47] E. Sarkar, P. Korshunov, L. Colbois, and S. Marcel. Vulnerability analysis of face morphing attacks from landmarks and generative adversarial networks. *CoRR*, abs/2012.05344, 2020.

[48] U. Scherhag, A. Nautsch, C. Rathgeb, M. Gomez-Barrero, R. N. J. Veldhuis, L. J. Spreeuwers, M. Schils, D. Maltoni, P. Grother, S. Marcel, R. Breithaupt, R. Raghavendra, and C. Busch. Biometric systems under morphing attacks: Assessment of morphing techniques and vulnerability reporting. In *International Conference of the Biometrics Special Interest Group, BIOSIG 2017, Darmstadt, Germany, September 20-22, 2017*, volume P-270 of *LNI*, pages 149–159. GI / IEEE, 2017.

[49] U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch. Deep face representations for differential morphing attack detection. *IEEE Trans. Inf. Forensics Secur.*, 15:3625–3639, 2020.

[50] F. Schroff, D. Kalenichenko, and J. Philbin. Facenet: A unified embedding for face recognition and clustering. In *CVPR*, pages 815–823. IEEE Computer Society, 2015.

[51] K. Sun, B. Xiao, D. Liu, and J. Wang. Deep high-resolution representation learning for human pose estimation. In *CVPR*, pages 5693–5703. Computer Vision Foundation / IEEE, 2019.

[52] S. Venkatesh, H. Zhang, R. Ramachandra, K. B. Raja, N. Damer, and C. Busch. Can GAN generated morphs threaten face recognition systems equally as landmark based morphs? - vulnerability and detection. In *IWBF*, pages 1–6. IEEE, 2020.

[53] P. Voigt and A. v. d. Bussche. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer, 1st edition, 2017.

[54] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, W. Liu, and B. Xiao. Deep high-resolution representation learning for visual recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 43(10):3349–3364, 2021.

[55] S. Woo, J. Park, J. Lee, and I. S. Kweon. CBAM: convolutional block attention module. In *ECCV (7)*, volume 11211 of *Lecture Notes in Computer Science*, pages 3–19. Springer, 2018.

[56] H. Zhang, S. Venkatesh, R. Ramachandra, K. B. Raja, N. Damer, and C. Busch. MIPGAN - generating strong and high quality morphing attacks using identity prior driven GAN. *IEEE Trans. Biom. Behav. Identity Sci.*, 3(3):365–383, 2021.