Differential Morphed Face Detection Using Deep Siamese Networks

MultiMedia FORensics in the WILD

Sobhan Soleymani, Baaria Chaudhary, Ali Dabouei,

Jeremy Dawson, Nasser M. Nasrabadi

West Virginia University

January 2021

Motivation

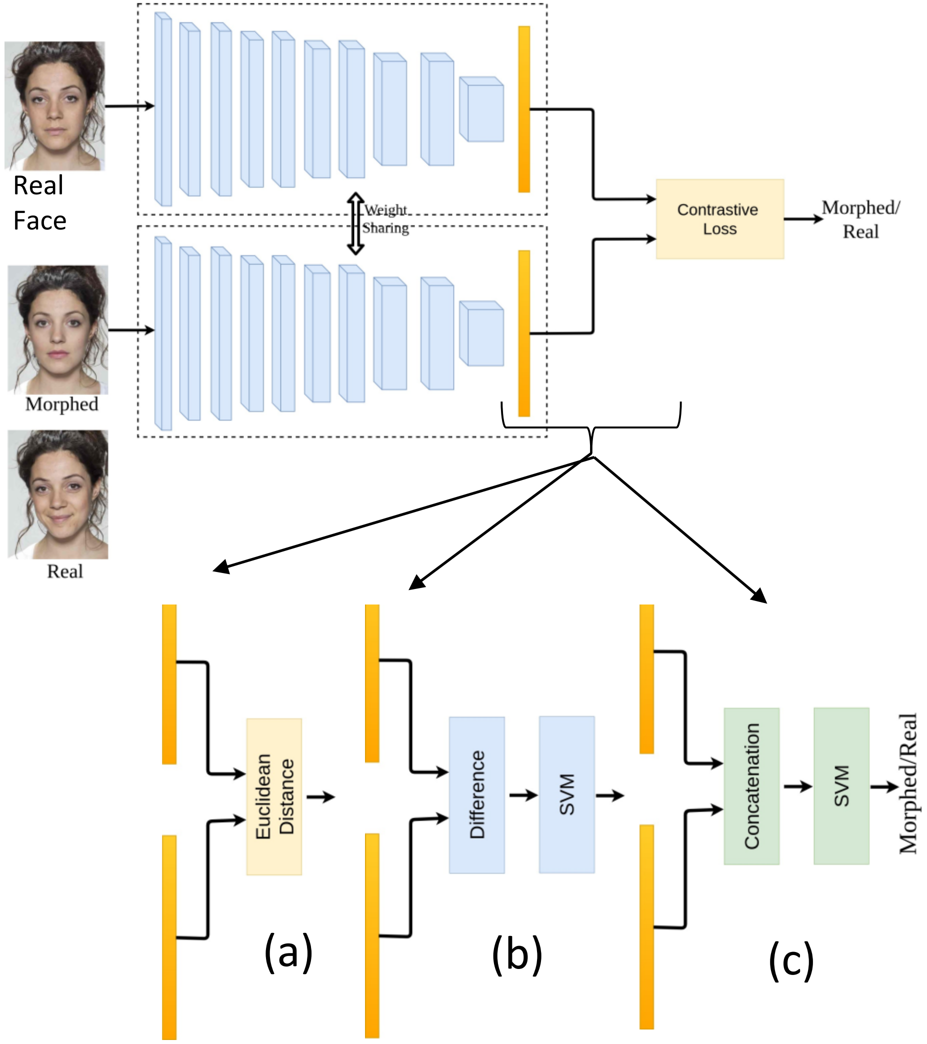
- Biometric facial recognition systems have increasingly been integrated into border control and other security applications that utilize identification tasks, such as official identity cards, surveillance, and law enforcement.
- Face morphing techniques allow any attacker to combine images from two subjects to get a single morph image.
- There are a large number of face morphing techniques, mainly based on landmarks, trangulation, and warpping or generative frameworks.
- Morphing artifacts can include: Ghosting, transition between facial regions, hair and eyelash shadows, deformed facial regions, and blurriness in forehead and color.
- We develop a differential morphing attack detection algorithm to distinguish between the morphed photo and one of individual travelers with the government ID.

MorGAN

VISAPP17

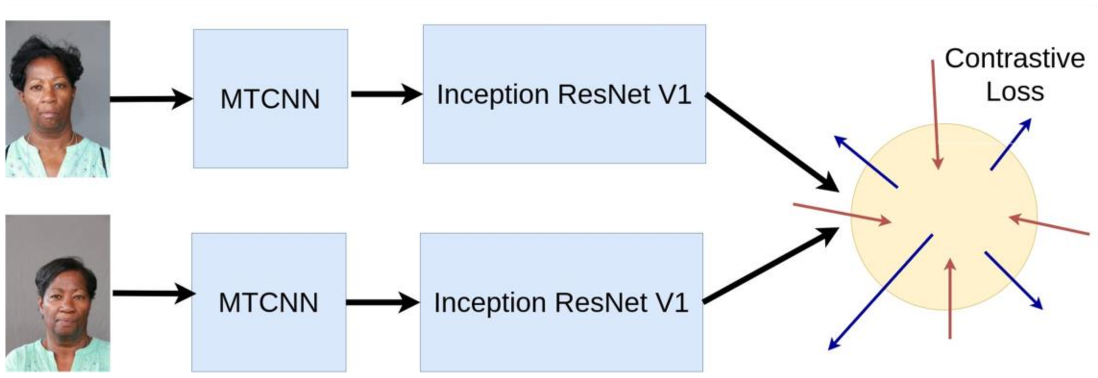
Approach

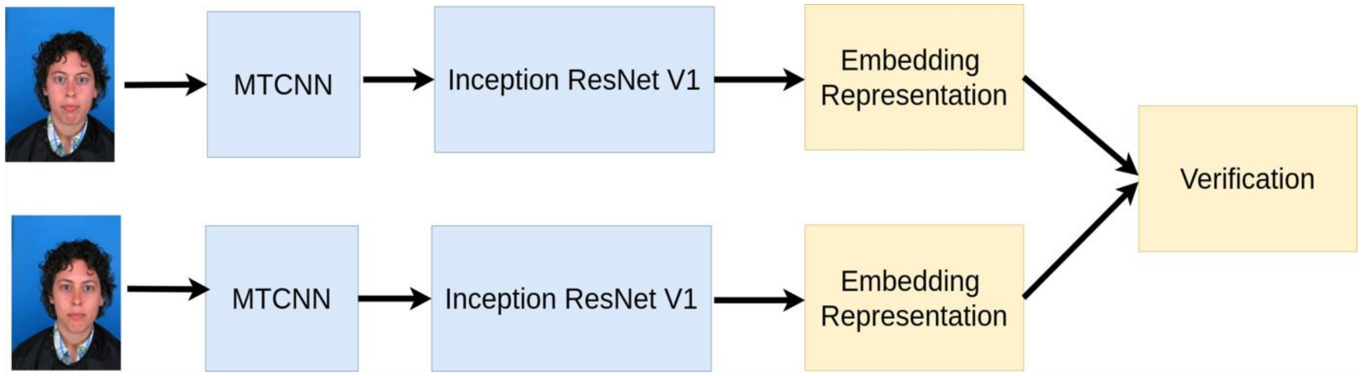
- Design a Siamese network distinguishing between genuine (non-morphed) and imposter (morphed) pairs.
- Use Contrastive Loss Function to optimize two identical DNNs (Siamese network) each operating on a different input image and use a Euclidean distance measure or an SVM classifier to make the final decision.
- Our detector is evaluated by comparing with texture-based and deep models reporting APCER, BPCER, and D-EER.
- APCER and BPCER are the proportion of morph samples and bona fide samples classified incorrectly, respectively.
- VISAPP17_selected morph samples (900x1200) are generated by geometrically warping the landmarks of the source image to the target image.
- MorGAN morph samples (64x64) are generated using a ALIGAN generative model.



Training the Detector

- The faces are detected and aligned using MTCNN framework.
- Our Siamese network is an Inception ResNet v1 initialized with weights pre-trained on VGGFace2.
- Training a Siamese network using WVU Twin dataset.
- The network is re-trained by enforcing contrastive loss on the embedding space representation of the genuine and imposter twin pairs.
- The trained Siamese network is then fine-tuned using the training portion of each morph dataset.
- The representations of the face image and its horizonal flipping are concatenated to provide a more distinguishable embedding.





Performance

- The performance of the proposed framework is compared with five morph detection models. The texture-based models are combined with an SVM classifier.
- We augment the proposed framework (Euclidian distance) with two other decision-making frameworks: Siamese (difference)+SVM and Siamese (concatenation)+SVM.

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

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

Class Activation Maps

- Class activation maps provide the attention of the decision with regards to regions of the face image.
- For this aim, we follow the implementation of gradient-weighted class activation maps.
- We present differentiation of the contrastive distance with regards to the feature maps constructed by ‘repeat_2’ layer.
- We report the normalized distance between the Grad-CAMs constructed for face images in a pair.