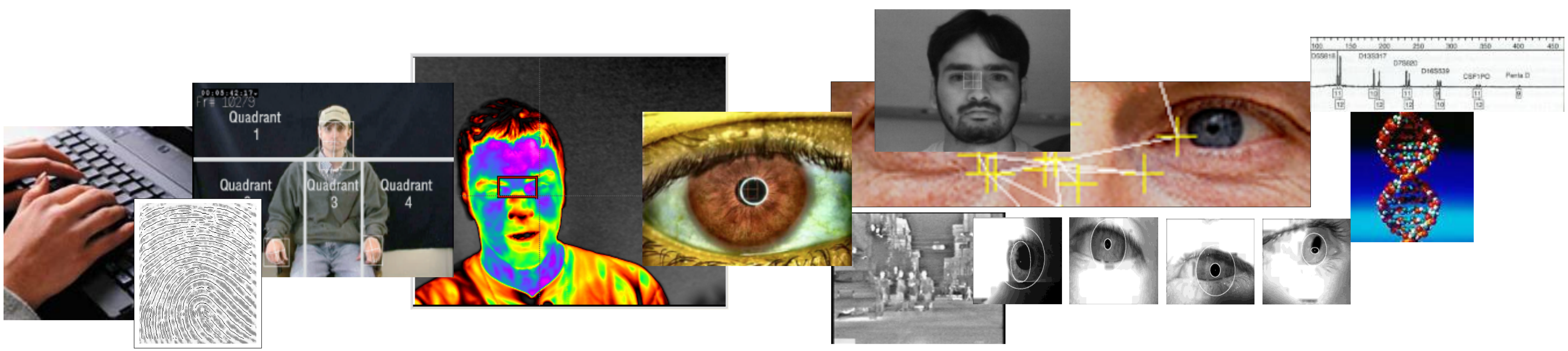
Center for Identification Technology Research (CITeR) Morph Attack Detection and Mitigation Projects

Center for Identification Technology Research (CITeR)

A National Science Foundation (NSF) Industry/University Cooperative Research Center (IUCRC)

Working in partnership with our government and industry stakeholders to advance the state of the art in human identification capabilities through coordinated university research



Research providing insight into the future of ID Technology

DHS Special Projects

- **Adversarial Learning Based Approach Against Face Morphing Attacks - Clarkson University**
- **Detecting Morphed Faces Using a Deep Siamese Network – West Virginia University**
- **Detecting Face Morphing: Dataset Construction and Benchmark Evaluation – West Virginia University**
- **FMONET: FAce MOrphing with adversarial NETworks and Challenge – University at Buffalo**

Adversarial Learning Based Approach Against Face Morphing Attacks

Zander Blasingame (Clarkson University)

Chen Liu (Clarkson University)

Stephanie Schuckers (Clarkson University)

Sebastian Marcel (Idiap Research Institute)

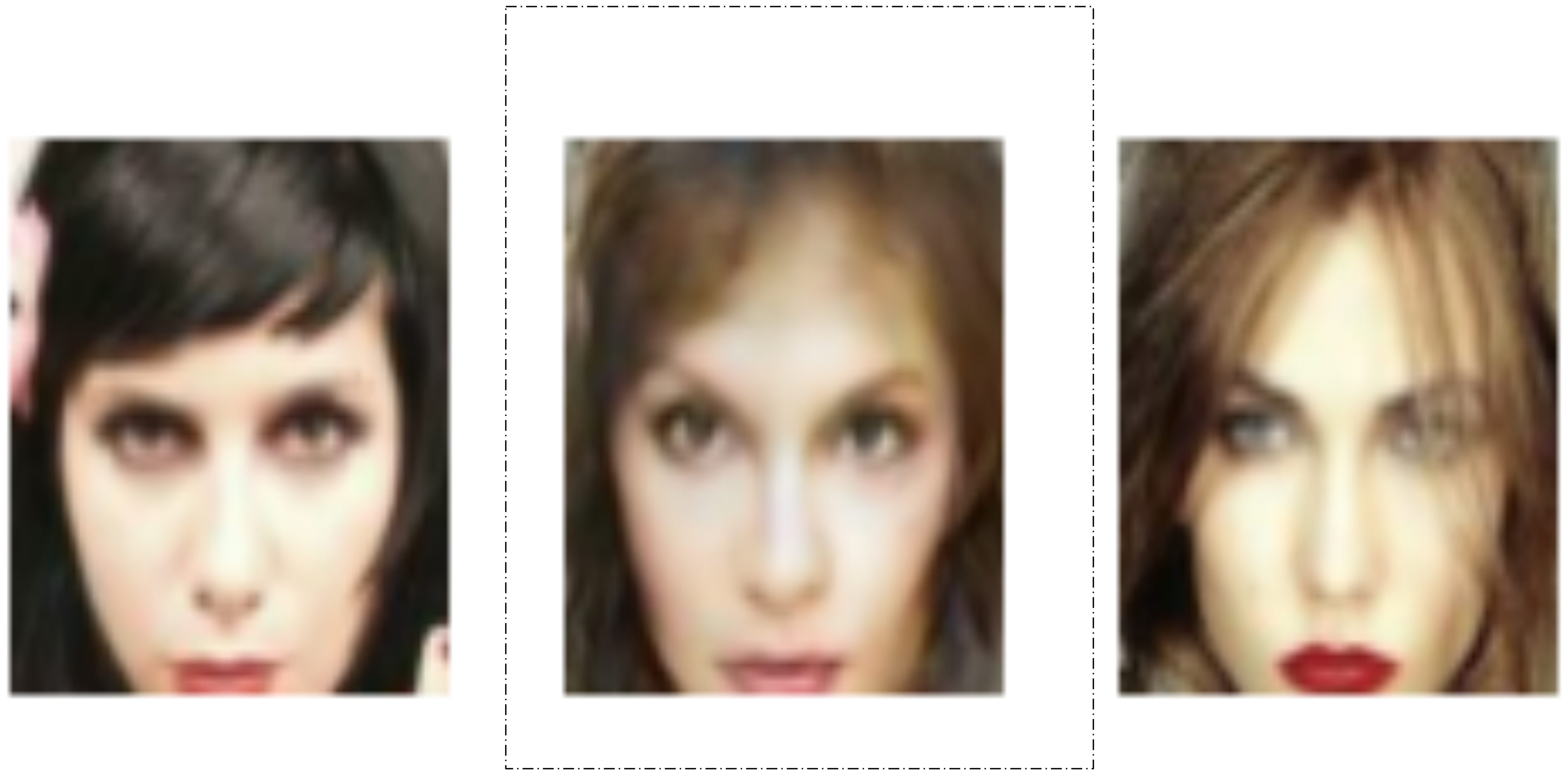
{blasinzw, cliu, sschucke}@clarkson.edu

Presentation at IFPC 2020

October 28, 2020

Motivation and State of the Art

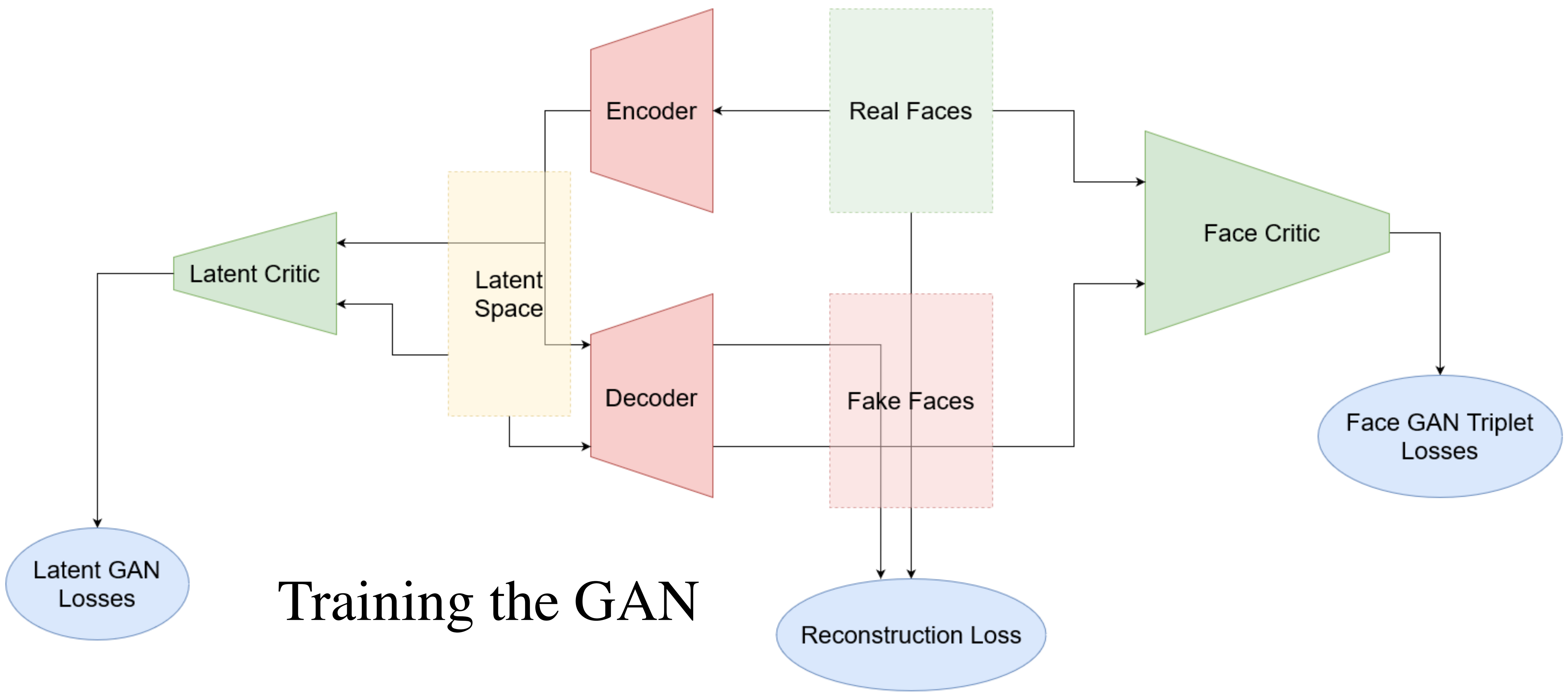
- Motivation
  - Malicious agents can create morphed faces that confuse both FR systems and human agents
- State of the Art
  - Attack: We propose a novel strategy for morphed face attacks using an additional network on the latent space of the GAN model
  - Defense: We propose an adversarial-based method for the detection of morphed faces



Example of morphed face
(featured in the center column)

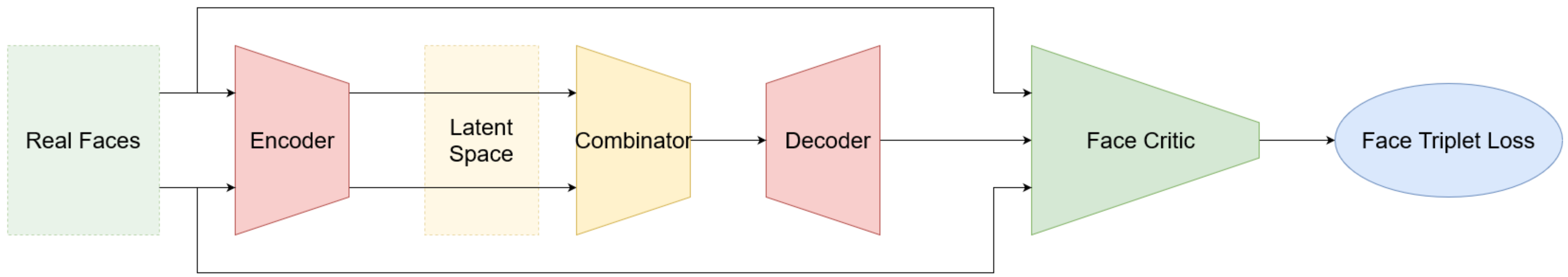
Objective and Approach

- Objective
  - Create realistic morphed faces
  - Have face critic be useful for detection



Training the GAN

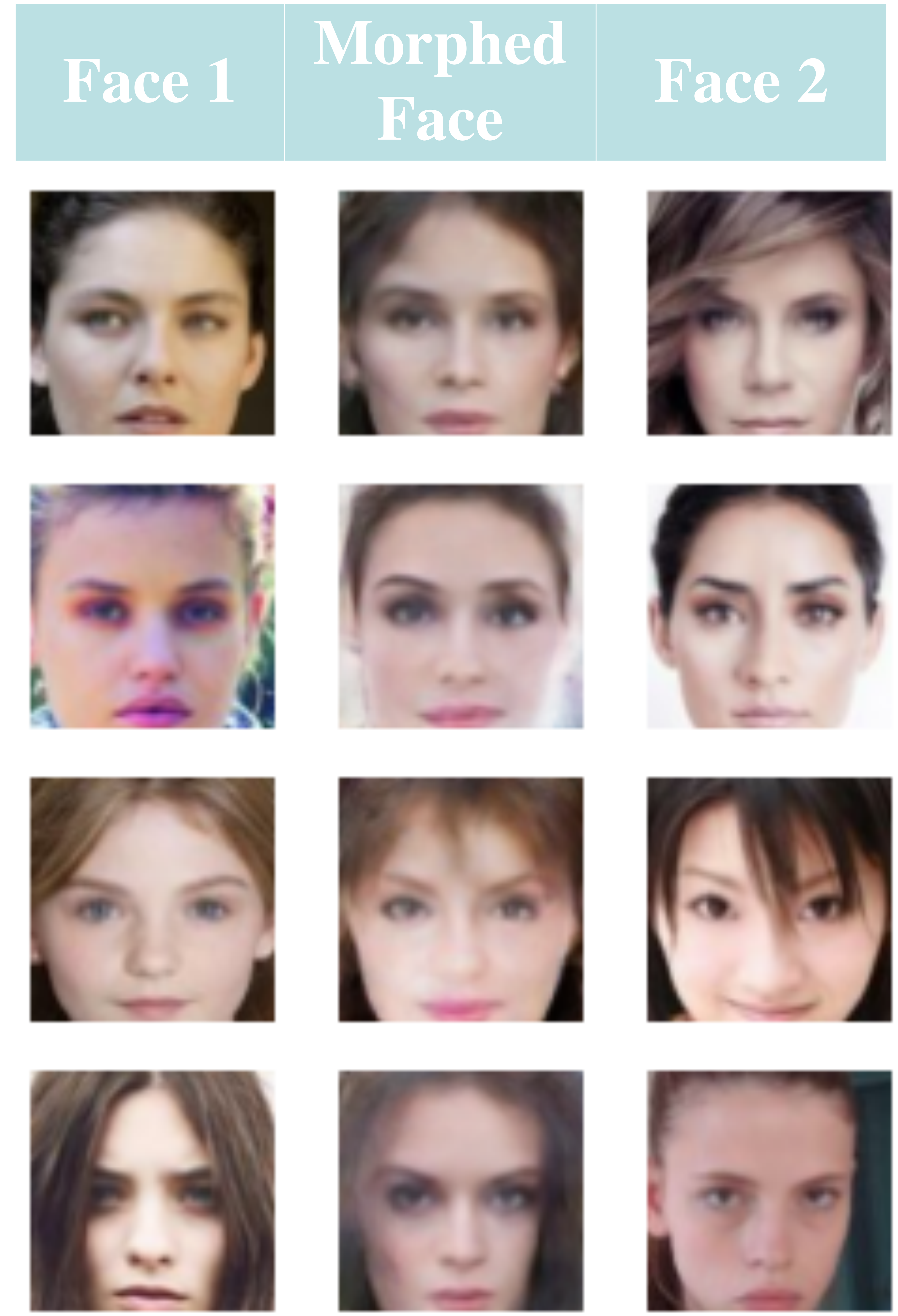
- Approach
  - Train cyclic GAN with triplet loss
  - Train combinator



Training the Combinator

Accomplishments

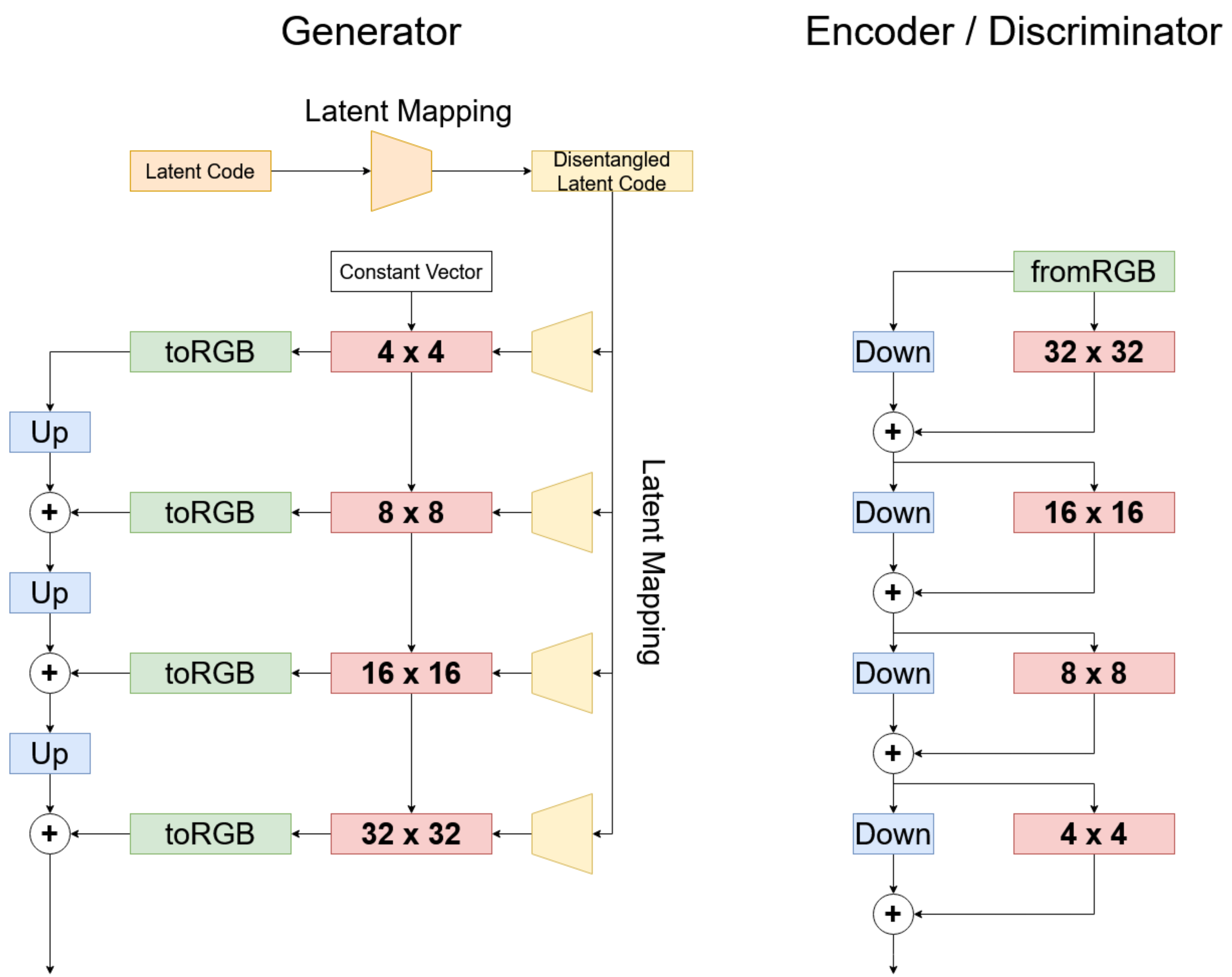
- GAN model
  - Created cyclic GAN with encoder, decoder, latent critic, and face critic
- Combinator for morphed faces
  - Created combinator model for morphed faces using the two encoded latent codes as input



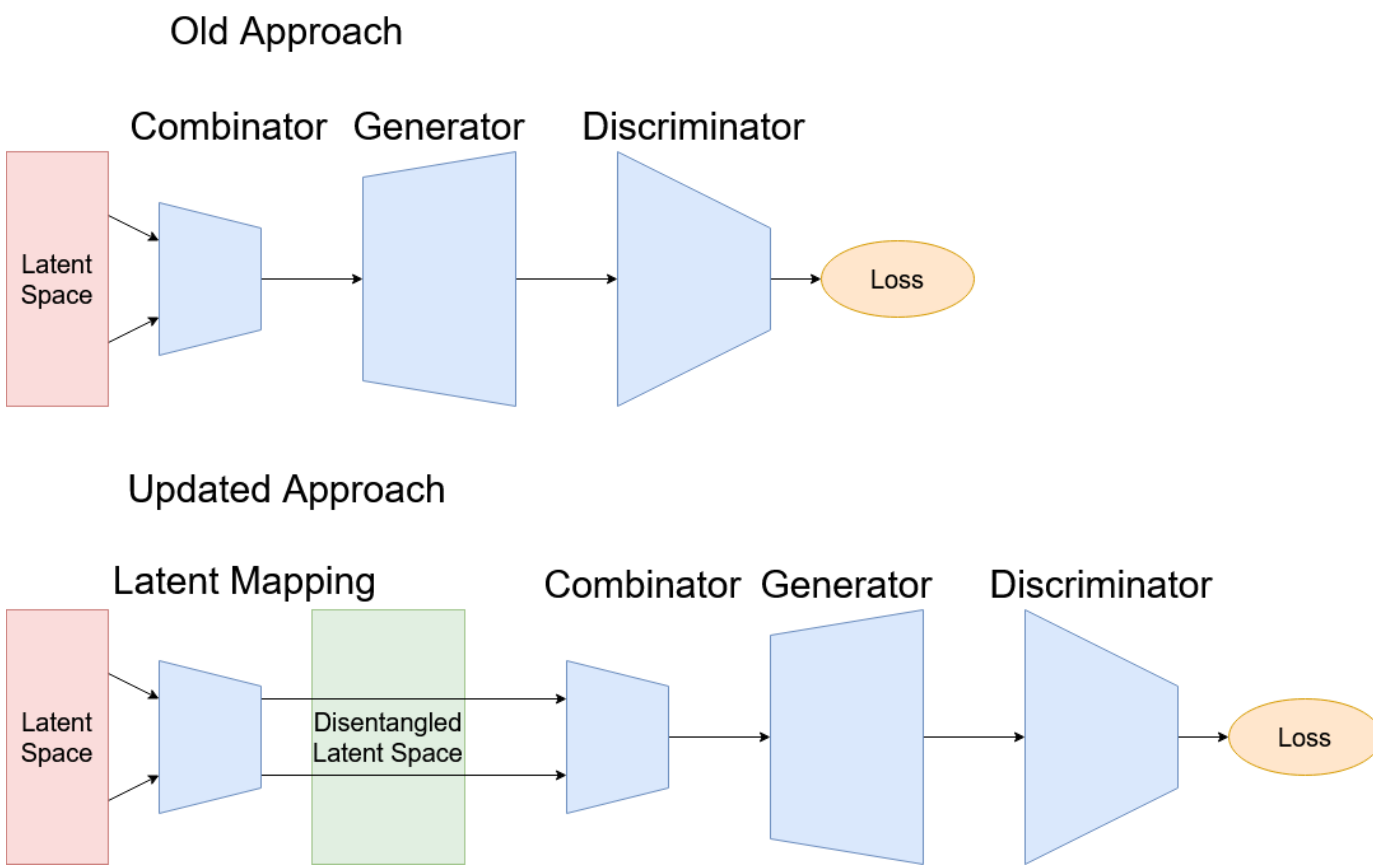
We used Large-scale CelebFaces Attributes (CelebA) Dataset for this work.

Updated GAN Approach

- Objective
  - Provide extensibility to higher resolution images
  - Restructure latent space to be more meaningful for morphing



Updated Combinator Approach



- New Combinator Training Technique

Next Steps

- Attack-wise: Use feature layer of face critic as additional input into the combinator model to increase strength of morphed attack

- Defense-wise: Develop face morphing detection model by using transfer learning from face critic of detector

- Validate performance of complete system (attack and defense)

Detecting Morphed Faces Using a Deep Siamese Network

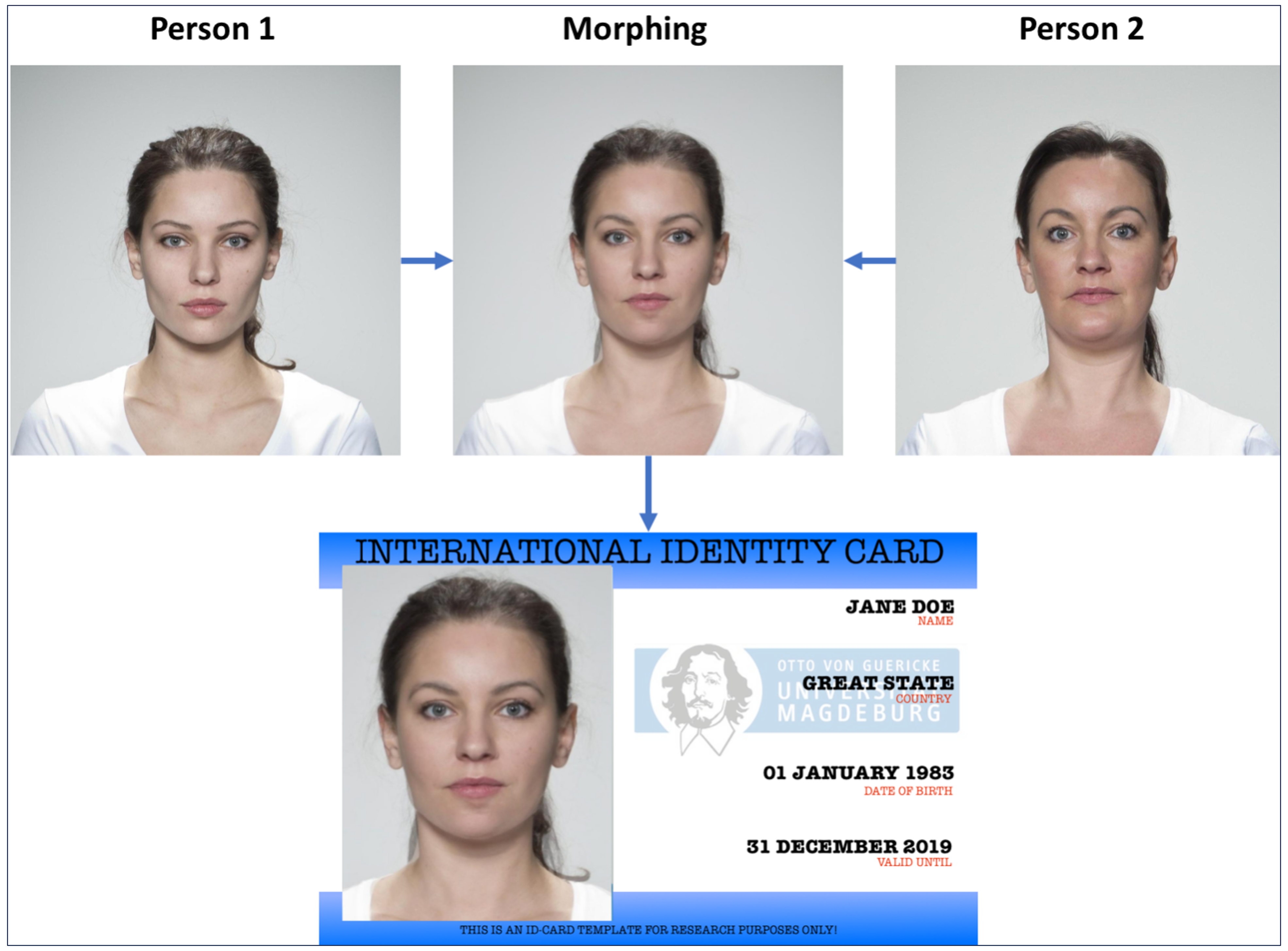
S. Soleymani, J.M. Dawson, and N. Nasrabadi

West Virginia University
Statler College of Engineering and Mineral Resources
Lane Department of Comp. Sci. & Elec. Engr.
PO Box 6109 Morgantown, WV 26554
jeremy.dawson@mail.wvu.edu
(304) 293-4028

Differential Morphing Face Attack Detection

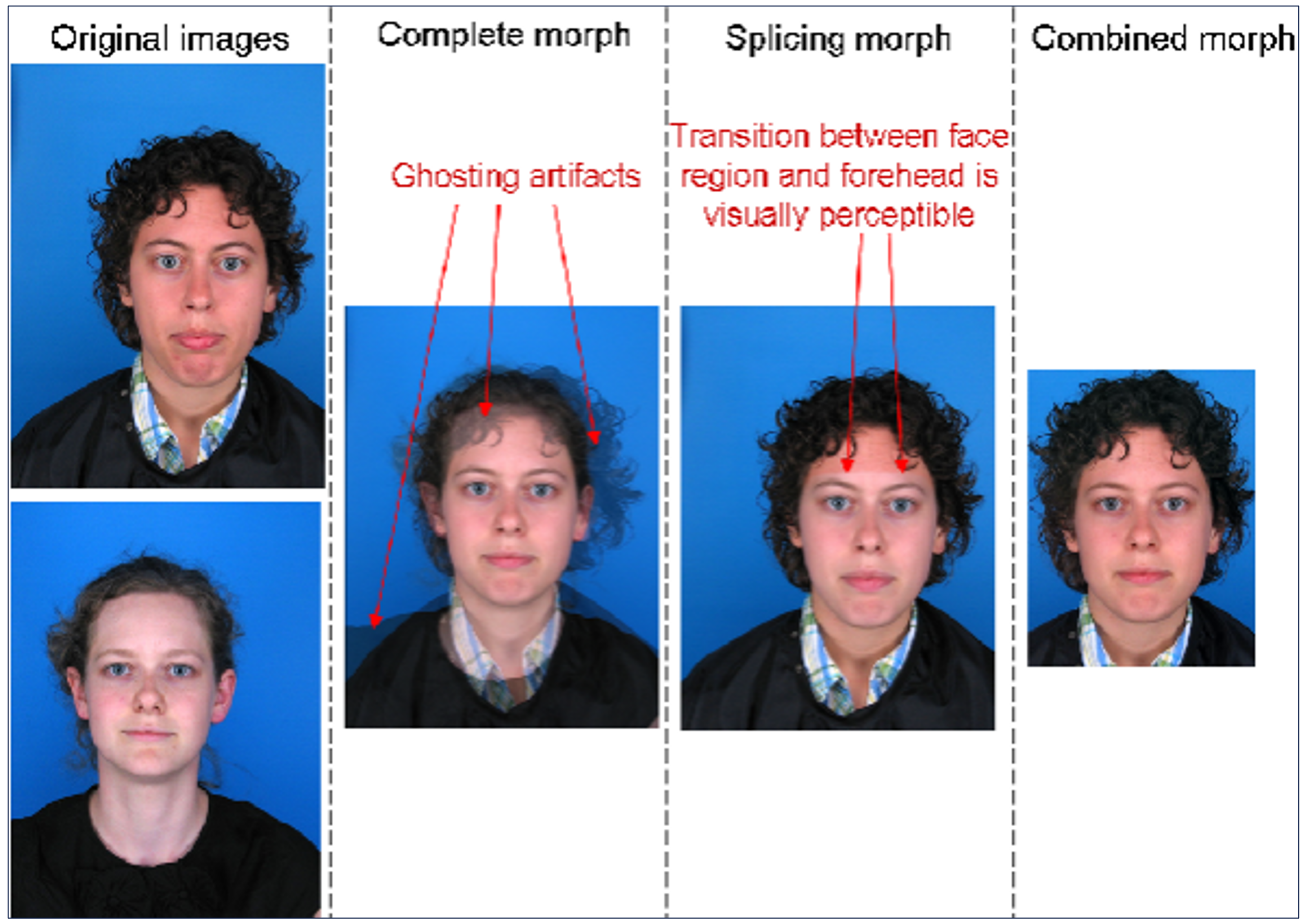
Motivation: Face morphing techniques allow any attacker to combine two different images from two subjects to get a single (composite) image.

- We are developing a novel universal **differential** morphing attack detection algorithm to distinguish between a morphed photo and one from an individual traveler with government ID



Challenges in Morphed Face Detection:

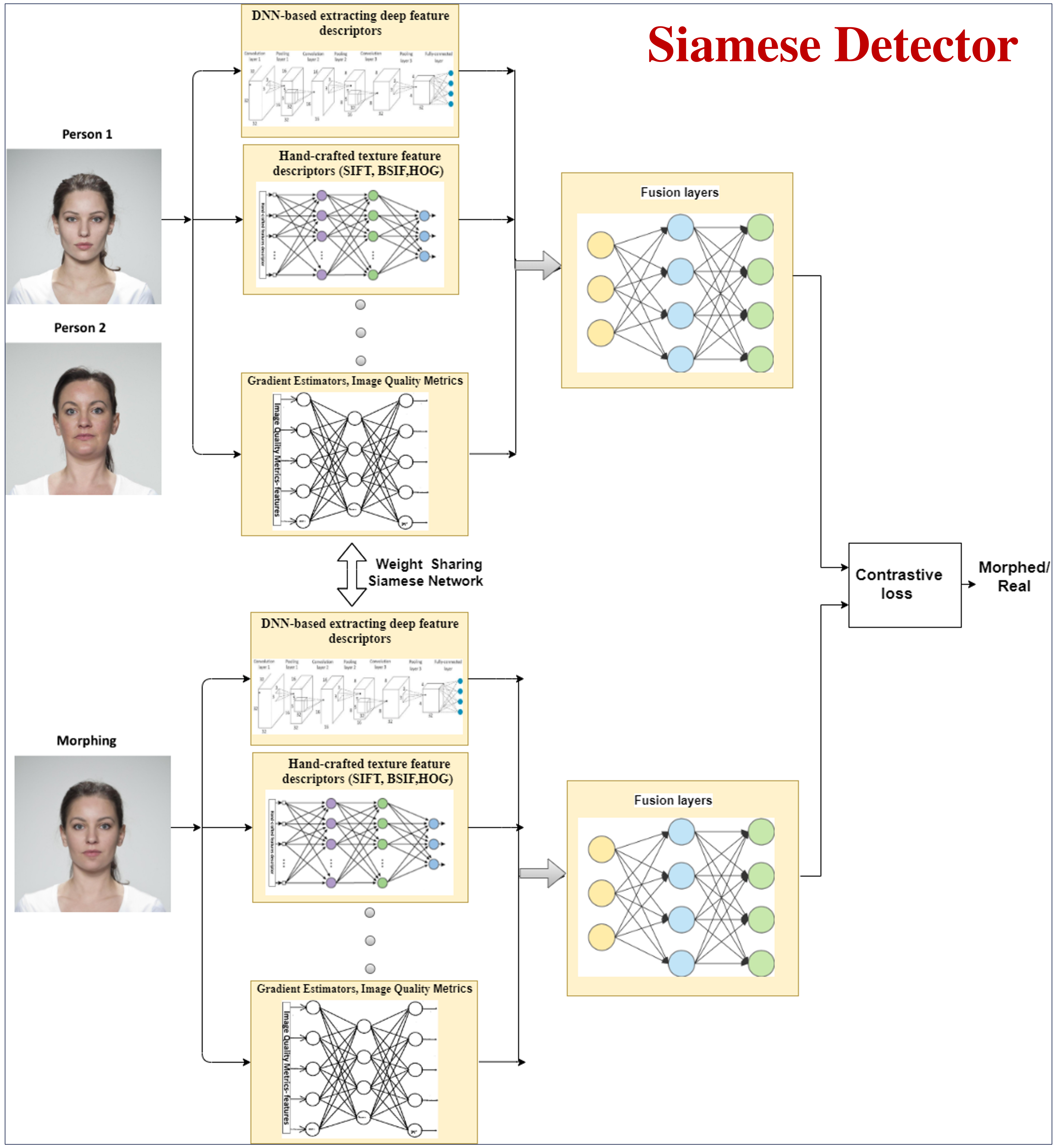
- There are many face morphing techniques, mainly based on landmarks+trangulation+warping or MorGAN architectures, and each one introduces a different facial morphing artifactMorphing artifact:
  - Ghosting, transition between facial regions, hair+eyelash shadows, deformed facial regions, blurriness in forehead, color and etc.



What features to use? How to compare two similar faces?

Objectives and Approach

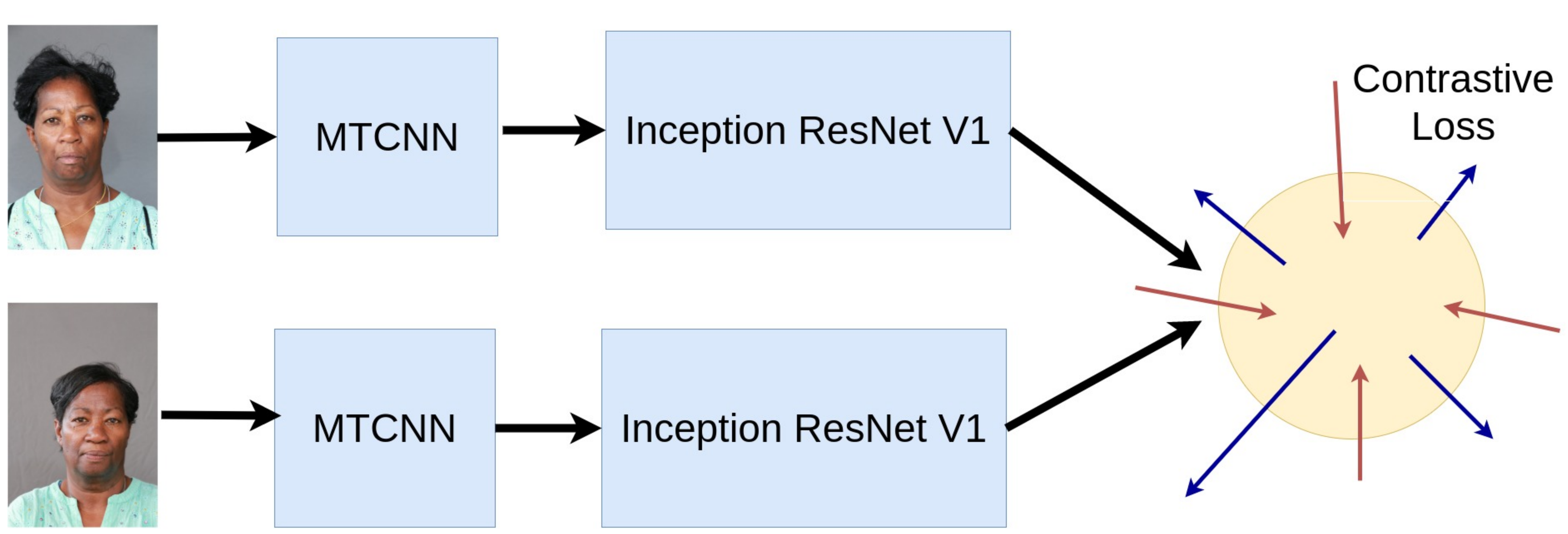
- Design a Siamese network distinguishing between genuine (non-morphed) and imposter (morphed) pairs.
  1. Use *Contrastive Loss Function* to jointly optimize a set of dedicated DNNs, each operating on different feature descriptors, as well as the input image, followed by a number of *Fusion Layers* and a Euclidean distance measure to make the final decision
  2. Our detector will be evaluated using the NIST report evaluation procedure, which is based on *morph miss rate* and *false detection rates*

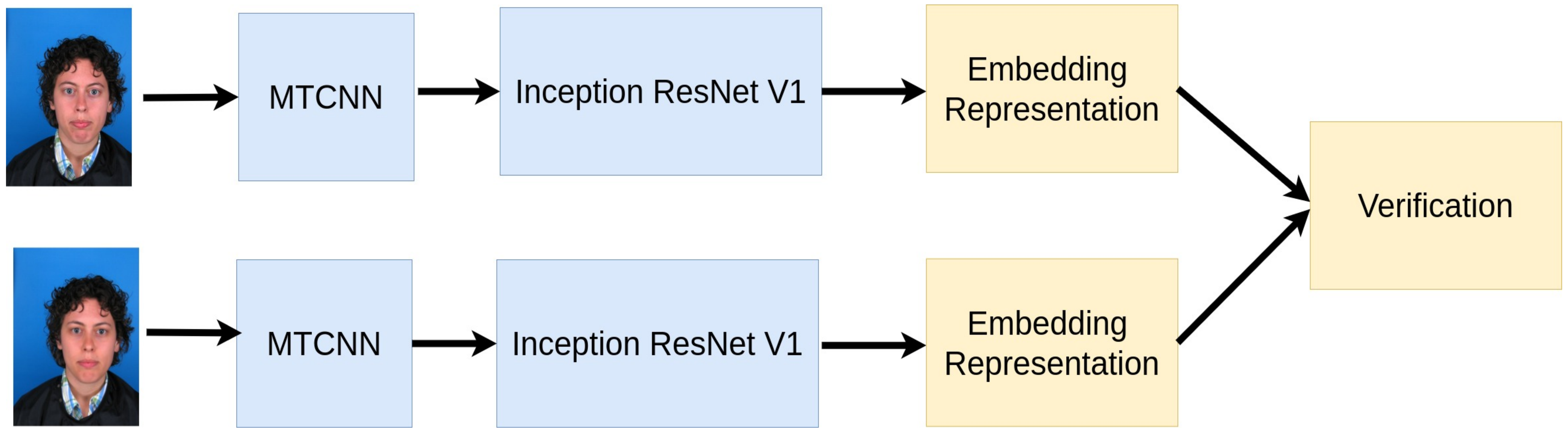


- What is Innovative and New?
  1. A Siamese-based differential face morphing detector
  2. A universal detector: Our proposed detector uses a large number of feature descriptors dedicated for different morphing artifacts

Progress

- Preprocessing differential morph datasets
  - 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
  - The faces are detected and aligned using MTCNN framework
- Training a Siamese network using a twins' face dataset
  - Our Siamese network is an Inception ResNet v1 initialized with weights pre-trained on VGGFace2
  - 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





Next Steps

- Employ different architectures to study the effectiveness of the proposed framework
- Compare the performance with state-of-the-art methods to demonstrate the effectiveness of our proposed framework
- Study multi-scale filters and their application in morph detection
- Fusion of hand-crafted and deep features for morphed face detection
- Attention maps to improve the discrimination between morphed and non-morphed images

Detecting Face Morphing: Dataset Construction and Benchmark Evaluation

Jacob Dameron, Guodong Guo, and Xin Li

(West Virginia University)

Problem

Motivation

- Face morphing attacks on facial recognition systems

  National Concern: https://pages.nist.gov/frvt/html/frvt_morph.html

- There is a lack of publicly available face morph datasets

  Currently, it's difficult to test algorithms without creating a whole new dataset. This makes training and comparing new algorithms more time consuming.

Value

[1] StyleGAN based morph.

- Provide our new dataset to the BIOM community

  Accelerate the work of other researchers with our dataset

- Benchmark cases to compare w/ new models.

[1] R. Abdal, Y. Qin, and P. Wonka, "Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space?"

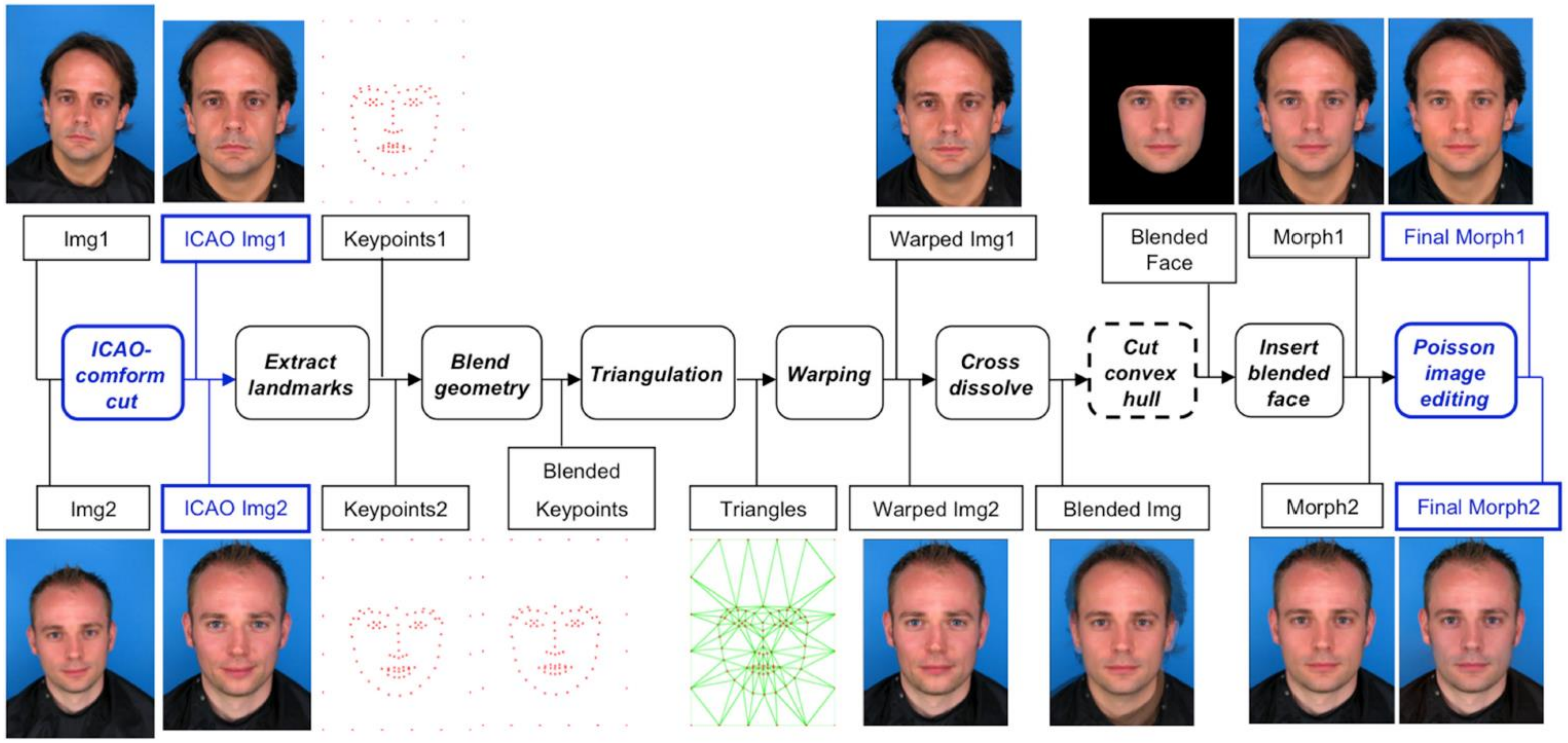
Landmark Morphing Attacks

- State of the Art

  [2] Standard "pipeline" for Landmark based Morphing Attacks (LMA) adopted as a framework. Some variants add steps before and after.

- General Pipeline

  1. Preprocessing
  2. Get Landmarks
  3. Blend Landmarks
  4. Triangulation
  5. Warping
  6. Alpha Blending
  7. Postprocessing



[2] Morphing Pipeline

[2] T. Neubert, et al. "Extended StirTrace benchmarking of biometric and forensic qualities of morphed face images."

Generative Adversarial Networks

- State of the Art

  [1] Outlined an embedding algorithm and approach used to generate morphs using StyleGAN.

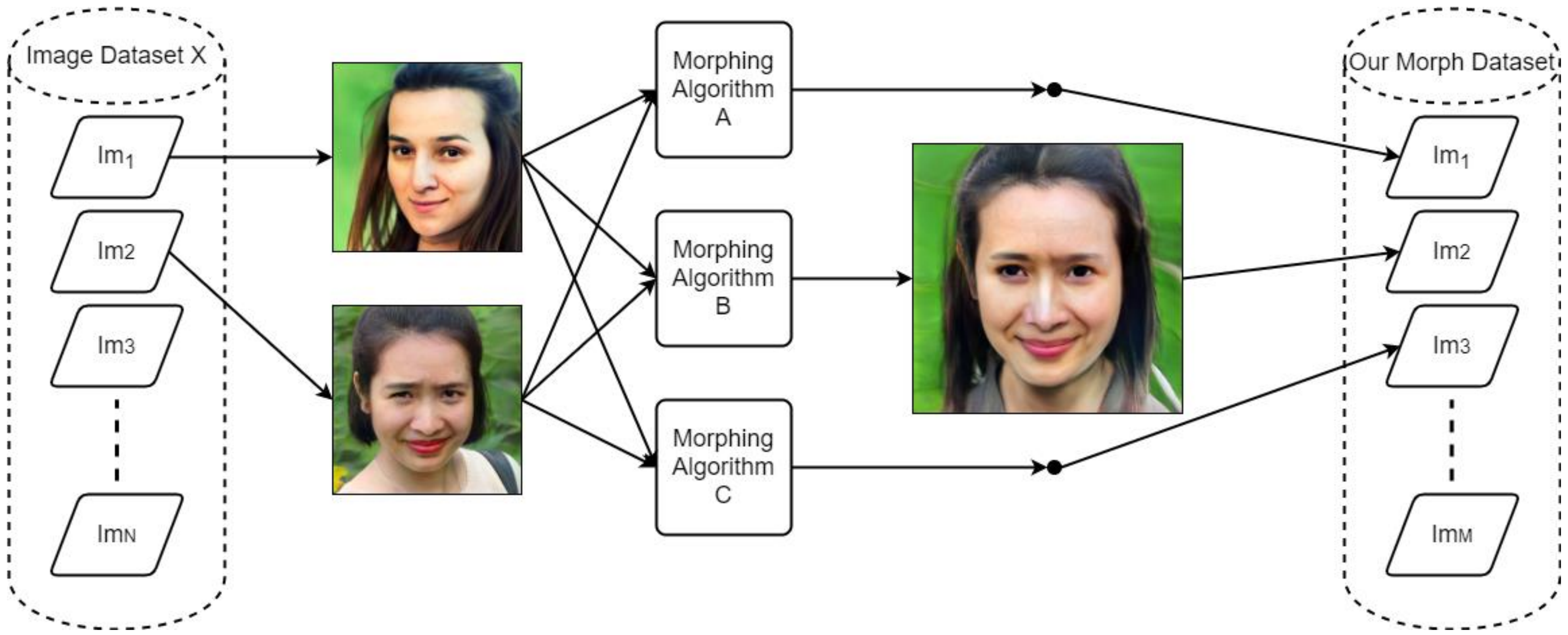
- StyleGAN Morphs
  1. Embed source images into StyleGAN latent space.
  2. Morph the two images in the latent space (average).
  3. Generate morphed image from latent space representation.

[1] GAN morph examples

[1] R. Abdal, Y. Qin, and P. Wonka, “Image2StyleGAN: How to Embed Images Into the StyleGAN Latent Space?”

Objective and Approach

Our Face Morph Dataset



- Objective

  Produce a dataset of face morphs.

- Approach

  Use image pairs belonging to publicly available datasets as sources.

  Use different algorithms to produce morphs of varying quality.

Our Benchmark Survey

- Objective

  Benchmark the state of the art in morphing attack detection (MAD).

- Approach

  Test popular morphing attack detection models against our dataset.

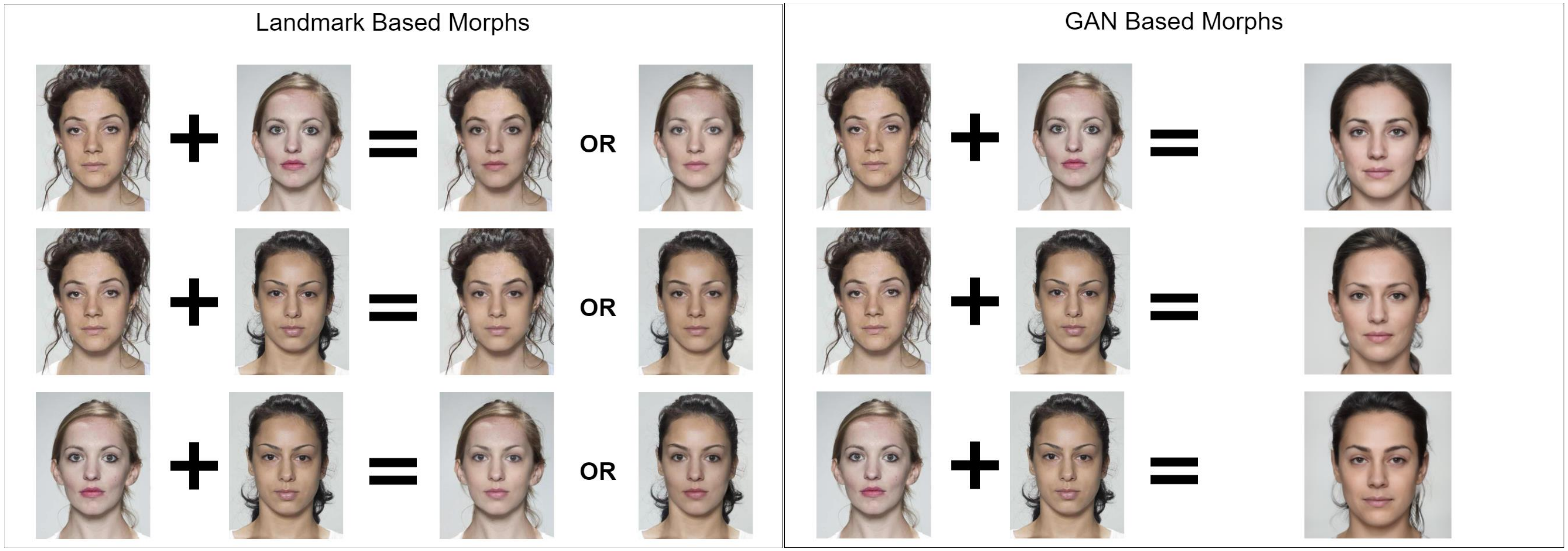
  Report performance in APCER and BPCER (ISO/IEC 30107-3).

Progress

- Morph Comparison w/ Other Datasets

  Using AMSL-Morph [3,4] source images as inputs to our morph attacks.

  Compare resulting morphs from our attacks against their morphs.



[3] AMSL Research Group. [n. d.]. 2019 AMSL Face Morph Image Data Set. ([n. d.]). https://omen.cs.uni-magdeburg.de/disclaimer/index.php
[4] https://figshare.com/articles/Face_Research_Lab_London_Set/5047666

Next Steps

- Phase 1: Dataset Construction
    1. Filter public datasets by pose and background to collect source images for face morphs.
    2. Test more LMA methods for inclusion in our dataset including [5].
    3. Large scale morph production using compiled methods and images.
- Phase 2: Benchmark Study
    1. Collect and verify software implementations of existing face morphing detection methods in the literature.
    2. Modify algorithms, such as [6], to detect morphed images.
    3. Test MAD methods against our dataset and report their performance in terms of APCER at specific BPCER values.

[5] M. Ferrara, et al. "Decoupling texture blending and shape warping in face morphing," in *BIOSIG*, 2019.
[6] Li, Yuezun, and Siwei Lyu. "Exposing deepfake videos by detecting face warping artifacts."(2018).

FMONET: FAce MOrphing with adversarial NETworks and Challenge

David Doermann, Srirangaraj Setlur
(University at Buffalo)

Motivation

- The need to stimulate face-specific manipulation detection research through a challenge seeded with combinations evolving GAN based face-to-face manipulations to counter security risks posed by documents presented to human and automated systems

Objectives

- Incorporate a range of quality and source image similarity measure quantitatively
- Provide enhanced generation and morph capabilities based on GANs

Accomplishments

- Dataset Collection
  ~10K images generated
- GAN Implementation
- Face Image Selection
- Morph Generation
  CafeGan: Conditional Attribute Face Editing
- Implemented Key Detectors
- Completed End to End Pipeline