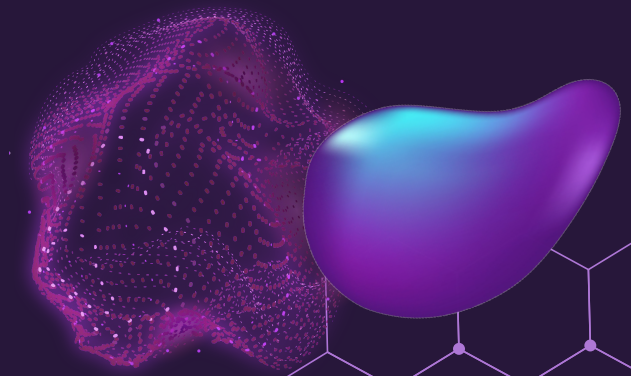
Trends and Applications of Computer Vision: face swapping at the limit

Giovanni Lorenzini
Diego Planchenstainer
Riccardo Sassi
Riccardo Ziglio



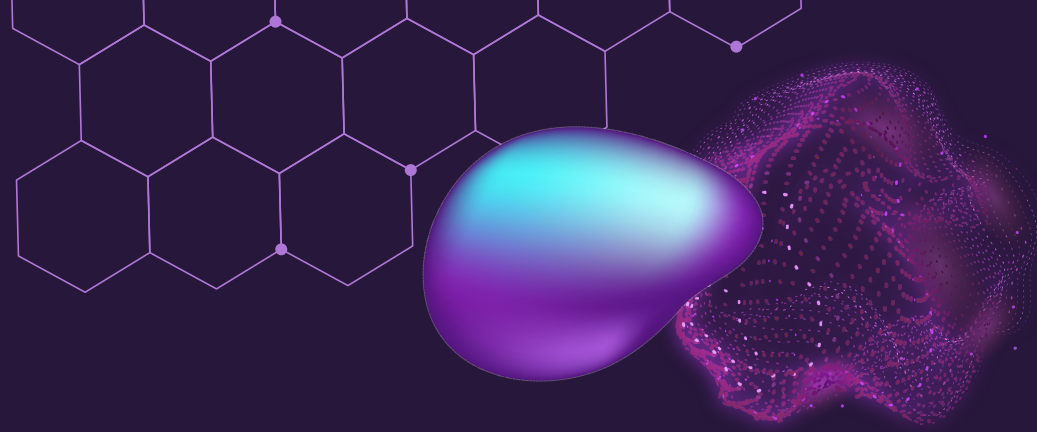
OUTLINE

01	FSGAN	deepfake generator used in our experiments
02	DeepFaceLab	deepfake generator
03	GOTCHA	system for real time deepfake detection
04	PROJECT STATUS	

01

FSGAN

Deepfake generator used in our experiments

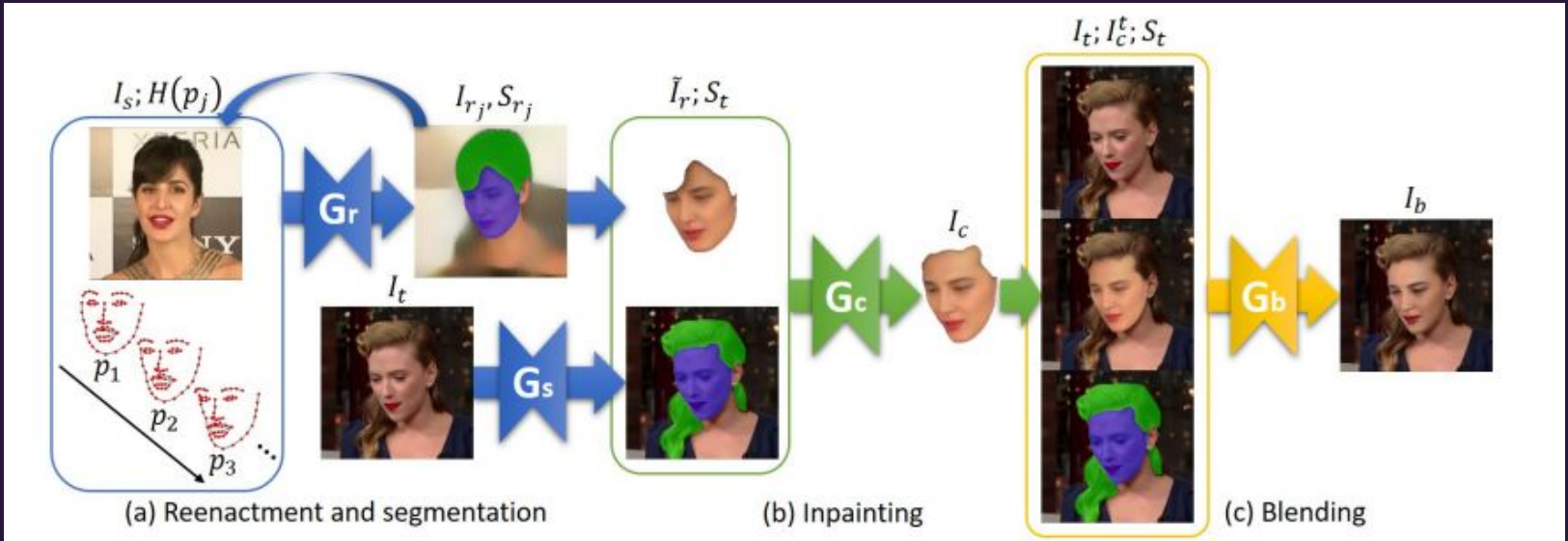


FSGAN

- Subject agnostic model for face swapping and face reenactment both in images and videos.
- Produces photo realistics, temporally coherent results.
- Goal: “synthesize a new image based on the target image (It), such that the target face (Ft) is seamlessly replaced by source face (Fs) while retaining the same pose and expression” [12].

FSGAN Pipeline

- Main components:
  - reenactment generator Gr, and the segmentation network Gs;
  - face inpainting generator Gc;
  - face blending generator Gb.



02

DeepFaceLab

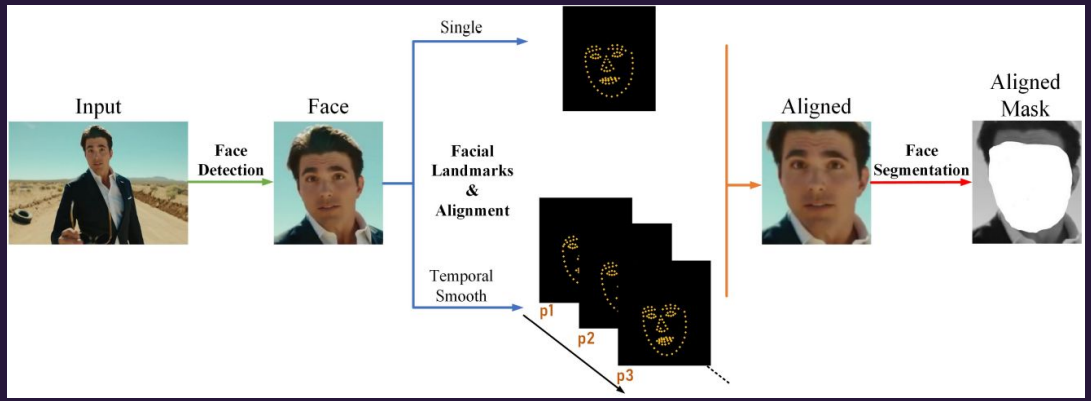
Deepfake generator

DeepFaceLab

- “DeepFaceLab [2] is arguably the most sophisticated deepfake generator” [1].
- Modular implementation: every block can be swapped with another (better) one.
- The better the data used to train, the more robust the forgery detection algorithm.
- This reference is important because:
  - state of the art face swapping network to defeat;
  - knowledge of its building blocks give us information about the attack surface;
  - is the technique that GOTCHA has more difficulties to detect.

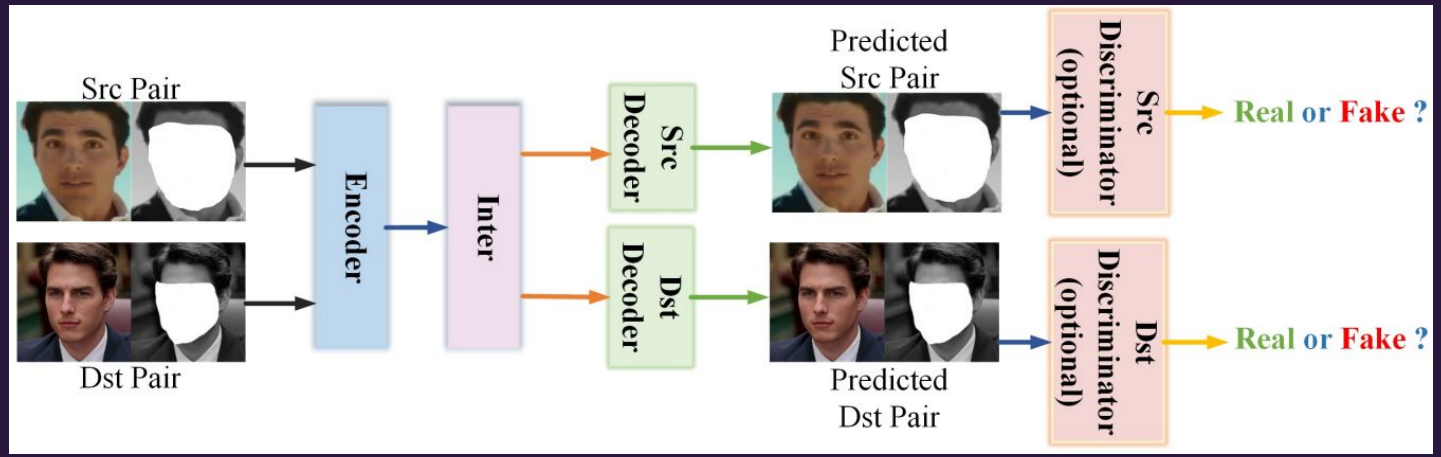
DeepFaceLab Pipeline

- Extraction:
  - Face Detection: find target face in given data. Uses S3FD [4] by default.
  - Face Alignment: find facial landmarks to calculate a similarity transformation matrix [5] used for face alignment. Uses 2DFAN [6] and PRNet [7] by default as landmarks extractors.
  - Face Segmentation: TernausNet [8] is used to do face segmentation but the result can be improved by employing XSeg [2].



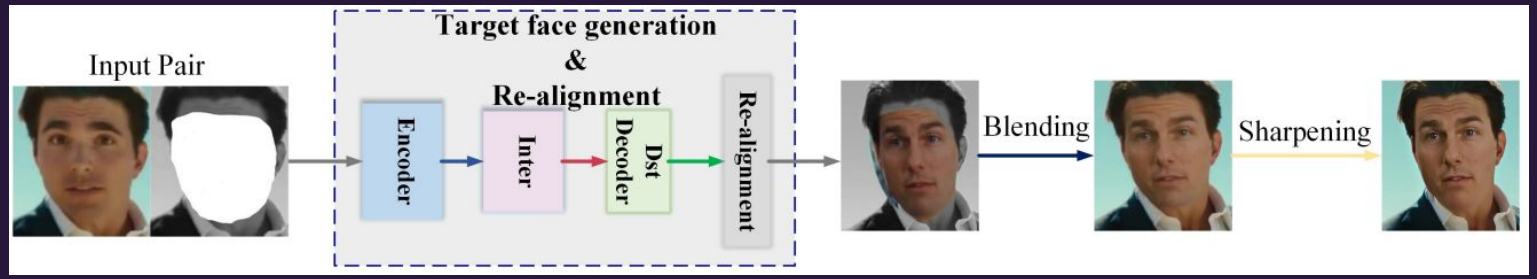
DeepFaceLab Pipeline

- Training:
  - Can choose between two architectures: DF and LIAE (more powerful).
  - Shared `encoder` and `inter` btw `src` and `dst` to deal with unmatched facial expressions.
  - Weight different part of the face in a different way.
  - Uses a mixed loss to improve speed and clarity.



DeepFaceLab Pipeline

- Conversion:
  - Can use many color transfer algorithms as RCT [9] and IDT [10].
  - Blends the faces by accounting for skin tones, face shapes, illumination conditions.
  - Implemented by using Poisson blending [11].
  - Sharpening of the output by a super-resolution neural network.



03

GOTCHA

System for real time deepfake detection

Overview

- Objective: disrupt the deep fake generator pipeline.
- Used in an Identification task or video-call, like job interview o meetings.
- sequence of challenges to a suspected Real-Time Deep Fake.
- Passive and active challenges to induce human-perceptible artifacts in the output.
- Through scores assignments, the video will be declared original or fake.

Which tasks?

Occlusion

- Active: occlude face with/without external objects.
- Passive: random facial cutouts, augmented reality (AR) filters, and stickers.

Facial expression

- Intentionally: specific emotional expressions, e.g., frowns or laughter. and lips movements
- Micro-expressions: involuntary facial expression humans

Facial distortion

- Active: include poking the cheek with a finger, and sticking out a small portion of the tongue.
- Passive: distorts the face with transformations such as affine, scaling, piece-wise affine, or warping.

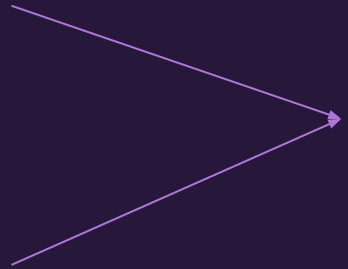
Surroundings

- Active: changing the ambience (e.g. illumination)
- Passive: synthetically changing the ambience (e.g. color filtering) or changing the background image

Quality metric

This quantity measures the quality of frames. It is assumed that:

- Exists Q(I_source), the quality metric of original frame
- Exists Q(I_imp), the quality metric of the corresponding manipulated frame



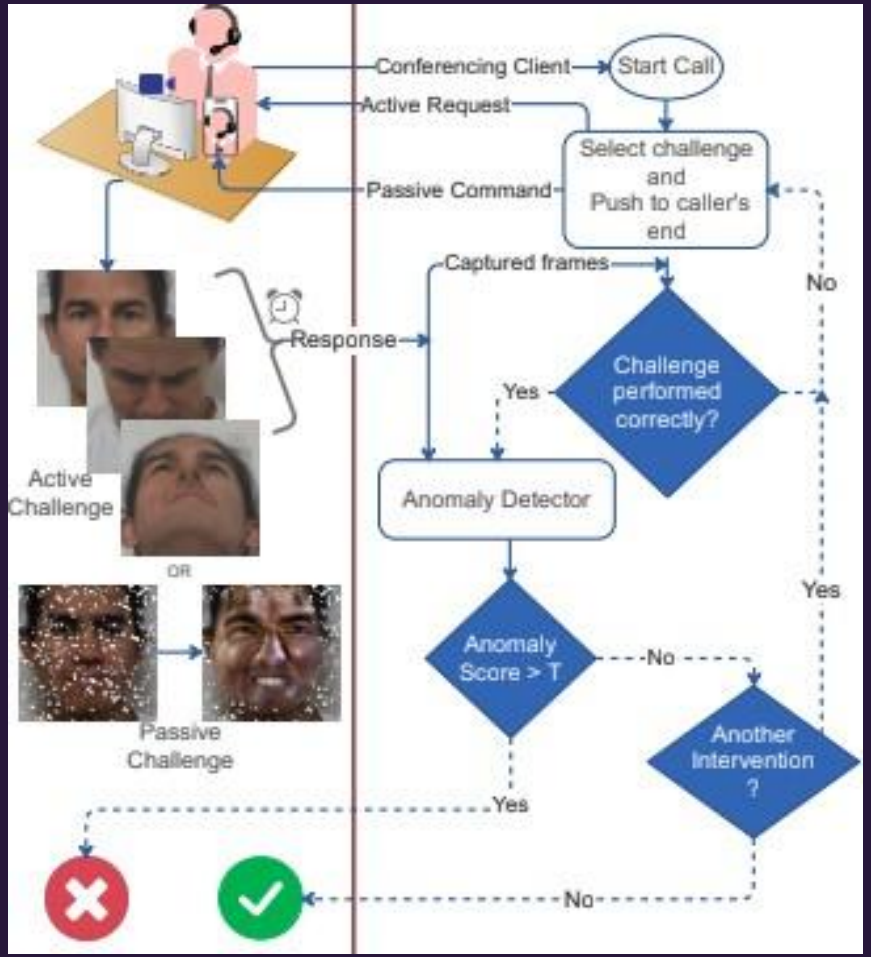
$$\Delta = Q(I_{source}) - Q(I_{imp}) > 0$$

Cumulative weighted score

ε indicates how much the video is fake for an Anomaly Detector

$$\varepsilon = \sum_{i=1}^{C} \log p_i * Q(I_{par,i})$$

If ε > T, the person is declared as impostor and the video as RTDF

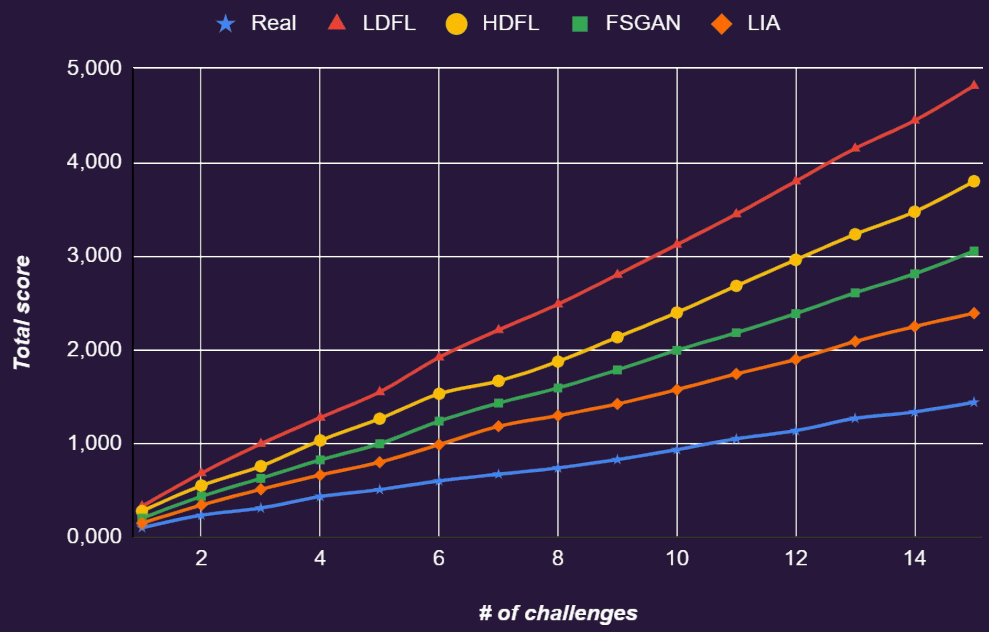


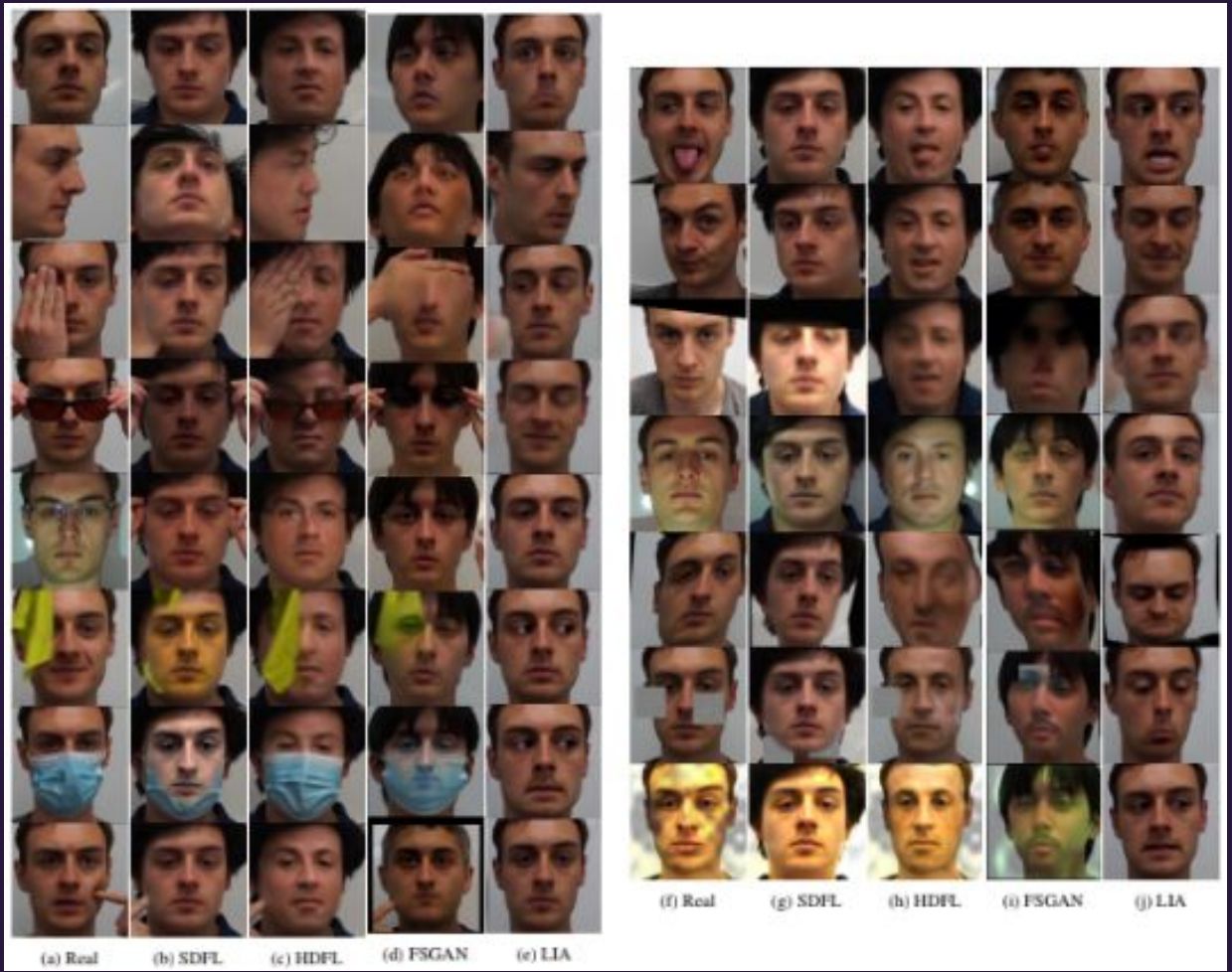
Steps:

1. Initialization
2. Select a challenge in a pre-selected list
3. Ask the person to complete the task
4. Capture the video frames
5. Is the task completed?
6. Score assignment using anomaly detector Q
7. Verify that total score < T
8. Other challenges to do?

Results

	Real	LDFL	HDFL	FSGAN	LIA
Angles	0,068	0,227	0,192	0,139	0,103
Head rotation	0,094	0,247	0,190	0,163	0,134
Hand on face	0,053	0,217	0,144	0,133	0,117
Sunglasses	0,085	0,195	0,191	0,137	0,106
Clear Glasses	0,051	0,190	0,161	0,120	0,095
Cloth	0,065	0,257	0,185	0,167	0,130
Facemask	0,049	0,203	0,095	0,134	0,136
Poke Cheek	0,047	0,191	0,144	0,112	0,079
Tongue out	0,062	0,219	0,181	0,134	0,087
Expression	0,074	0,223	0,183	0,146	0,106
Standup	0,081	0,226	0,198	0,130	0,117
Flash	0,060	0,244	0,193	0,142	0,108
Piece. Affine	0,092	0,243	0,190	0,154	0,133
Cutout	0,047	0,207	0,167	0,141	0,111
Color Filter	0,072	0,258	0,226	0,169	0,099
Average	0,067	0,223	0,176	0,142	0,110




(a) Real
(b) SDFL
(c) HDFL
(d) FSGAN
(e) LIA
(f) Real
(g) SDFL
(h) HDFL
(i) FSGAN
(j) LIA

04

PROJECT STATUS

Project status

Test the challenges and see how the system responds.
FSGAN to apply deepfakes requires:

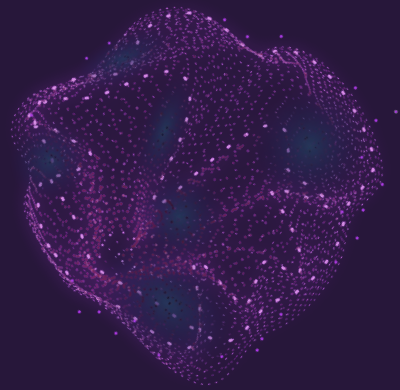
- source video (taken from youtube, in our case celebrities);
- target video (where the active challenges are tested).

Network weights shall be requested to the authors.
To download and cut videos we have used youtube-dl and ffmpeg.

Project status

Tested challenges:

- head rotation along all three axes;
- face occlusion with paper sheet;
- putting on a mask;
- display a colored image on the monitor;
- wear sunglasses;
- flip image upside down.



Giovanni → Bill Gates
Head rotation
Face mask

Diego → Keanu Reeves
Head rotation
Paper sheet

Diego → Robert Downey Jr

Talking

Sunglasses

Giovanni → Bill Gates
Image flipping

Initial results

- FSGAN is not robust to the challenges that we tested, also it is not in real time.
- Videos of source last around 15-30 s. Training time ~ 40 min.
- Quality is not great, but allows to apply face from any source to any impersonator.

- Inspect how to apply FSGAN in real time.
- DeepFaceLab [1] can be a more powerful network that can better hold up to these challenges.

References I

[1] Govind Mittal et al.: Gotcha: A Challenge-Response System for Real-Time Deepfake Detection. https://arxiv.org/abs/2210.06186

[2] Ivan Perov et al.: DeepFaceLab: Integrated, flexible and extensible face-swapping framework. https://arxiv.org/abs/2005.05535

[3] Yuval Nirkin et al.: FSGANv2: Improved Subject Agnostic Face Swapping and Reenactment. https://arxiv.org/abs/2202.12972

[4] Shifeng Zhang et al.: S^3FD: Single Shot Scale-invariant Face Detector. https://arxiv.org/abs/1708.05237

[5] Shinji Umeyama: Least-squares estimation of transformation parameters between two point patterns. https://ieeexplore.ieee.org/document/88573

[6] Adrian Bulat and Georgios Tzimiropoulos: How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). https://arxiv.org/abs/1703.07332

References II

[7] Yao Feng et al.: Joint 3D Face Reconstruction and Dense Alignment with Position Map Regression Network. https://arxiv.org/abs/1803.07835

[8] Vladimir Iglovikov et al.: TernausNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. https://arxiv.org/abs/1801.05746

[9] Erik Reinhard et al.: Color transfer between images. https://ieeexplore.ieee.org/document/946629

[10] Francois Pitie et al.: Automated colour grading using colour distribution transfer. https://www.sciencedirect.com/science/article/abs/pii/S1077314206002189

[11] Patrick Perez et al.: Poisson image editing. https://dl.acm.org/doi/10.1145/882262.882269

[12] Yuval Nurking et al.: FSGANv2: Improved Subject Agnostic Face Swapping and Reenactment. https://arxiv.org/abs/2202.12972