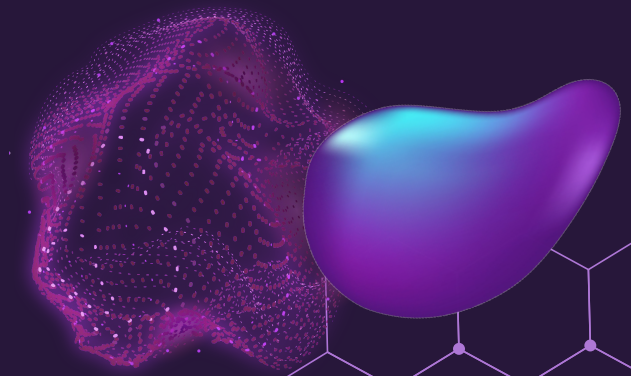
# Trends and Applications of Computer Vision: face swapping at the limit

Giovanni Lorenzini
Diego Planchenstainer
Riccardo Sassi
Riccardo Ziglio



# OUTLINE

| | | |
|---|---|---|
| 01 | **Introduction** | Why the topic is relevant? |
| 02 | **Theoretical background** | DF generation |
| 03 | **Project Overview** | What we have done and final results |
| 04 | **Outlook** | Future works and improvements |

# 01

# INTRODUCTION

Why our topic is relevant

# Nowadays...

- Modern life has made live, online video interactions fundamental and they are increasing. This creates fertile grounds for novel social-engineering attacks and online fraud [1].

- To prevent and reduce this kind of events, national government agencies require a practice known as KYC (Know Your Customer), specifically created to verify an individual's identity online [8].

# Which threats?

Increasing of online criminal activities (i.e. scams, frauds and identity thefts), in scenarios like job interview, signing contract online, official meetings, online government services, social activities.

- Substitution of person in delicate positions (CEO, Politicians ecc..)
- Stealing information (of people, companies, governments...)
- Cheating in tests
- Other presentation attacks

# How do we mitigate these attacks?

- Remote identity proofing (RIDP) analyzes biometric data such as facial features, fingerprints and other liveness parameters [6].

- KYC steps: ID verification, face matching and liveness verification [8].

# Why is this topic relevant?

- DF achieve realistic results and can be applied in real time, making them difficult to be recognized and they are able to fool the liveness detection systems used by banks to verify the user ID online [9].
- Since more and more people work online this has become an important security problem → integrity of video interaction.
- FBI classified deepfakes as an emergent risk [5].

# What do people think?

According to a survey held by iProov company* [7]:

- Only 13% are sure to know what a deepfake is.
- 75% would be more likely to use an online service that could prevent deepfake.
- 72% believe the need to authenticate identity is more important than ever before.

* leader in Genuine Presence Assurance to organizations around the world, their goal is to verify that an online user is a real person and not an imposter.

# DeepFakes detection

- Open source powerful deepfake generator → powerful deepfake detector.
- Offline deepfake detection:
  - Analyze specific regions.
- Online deepfake detection:
  - Propose (active or passive) challenges to the subject -> **GOTCHA**.

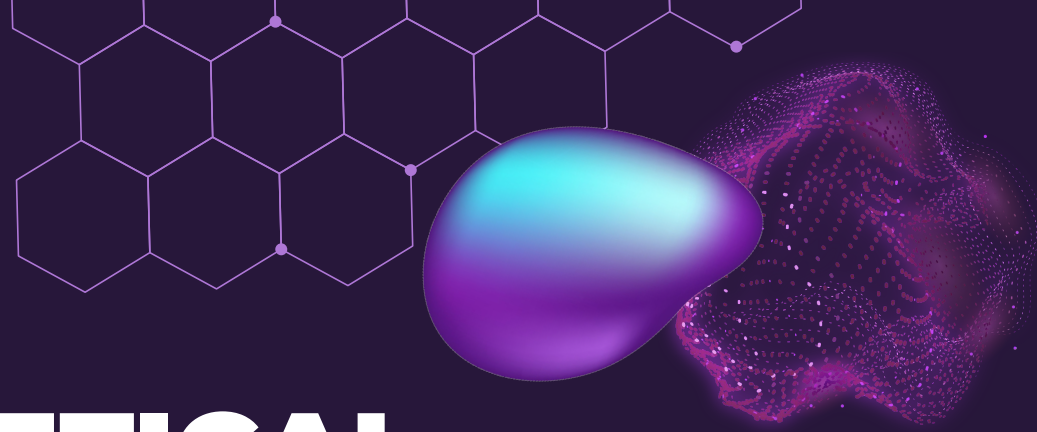
# GOTCHA [1]

- **Set of challenges** is present to a suspected RTDF.
- Interfere in DF pipeline.
- Cumulative score, compared with a threshold T.
- **Quality metric to evaluate inconsistencies**:
  - Difficult to compute;
  - incoherences between artifacts and scores in different RTDF techniques.
- **Active challenges**:
  - Rapid movements and expressions;
  - Occlusion;
  - Face profile;
  - Facial distortions;
  - Changing illuminance.

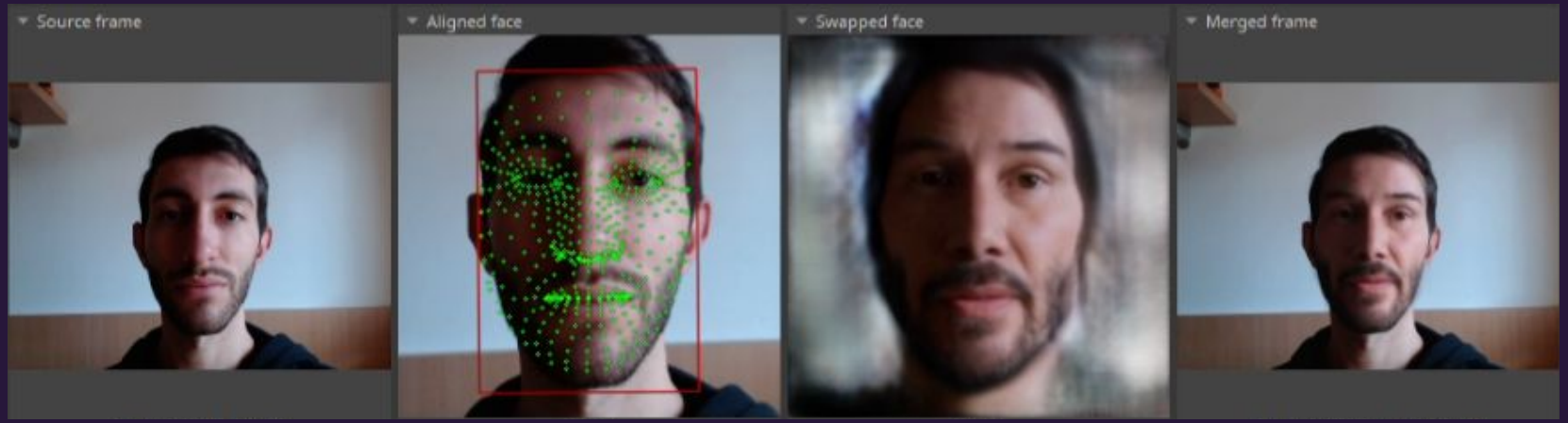
# 02

# THEORETICAL BACKGROUND

Deepfakes generation

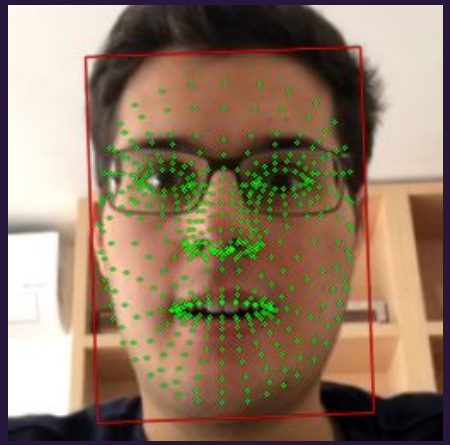


# DeepFakes - Pipeline



# Face Detection

Goal: find the target face in the given data (image or video), which will be then used to perform the face swapping.



# Landmark Detection

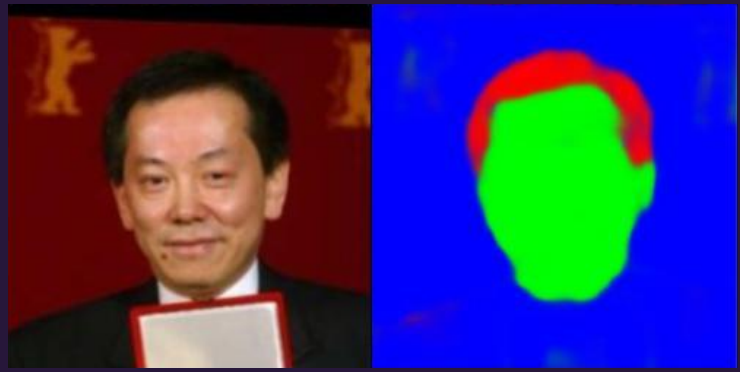
Identify all the relevant facial information of the target, such as eyes, mouth, nose, etc. These information will be used to correctly swap the source and target face.

# Face Alignment

Goal: vertically align the source and target face, important for a robust face swapping prediction.

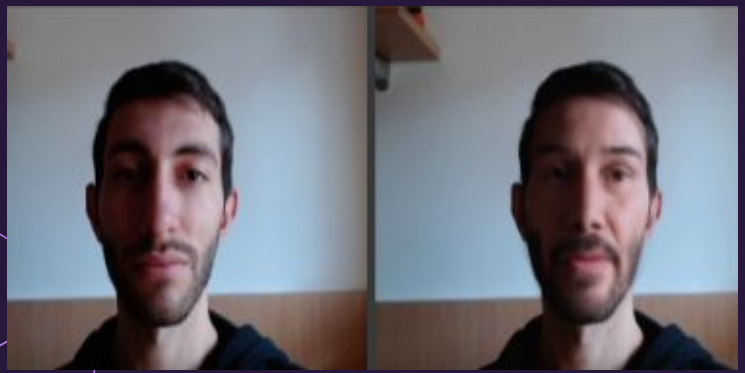

# Segmentation

Goal: determine which pixels belongs to the face and which do not (i.e. background, hair, fingers, glasses, etc.). Useful to perform exact face location and contour cropping.



# Face swapping

Swap face using facial landmarks, occlusions, and lighting information to predict the appearance of the source's face.



# Corrections

Blending: join together the inner and outer part of the face (involving some combination of blurring, scaling, etc.).

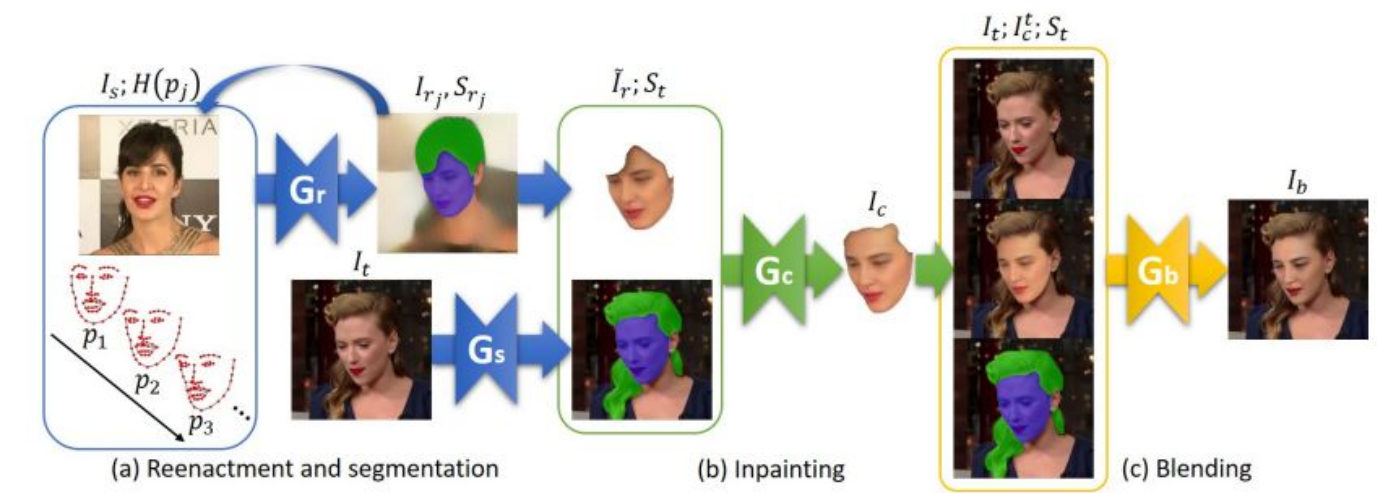
Color correction: sample color from outside face region and adjust the inner swapped face accordingly.

# Tested DF models

- FSGAN;
- DeepFaceLive;
- GHOST.

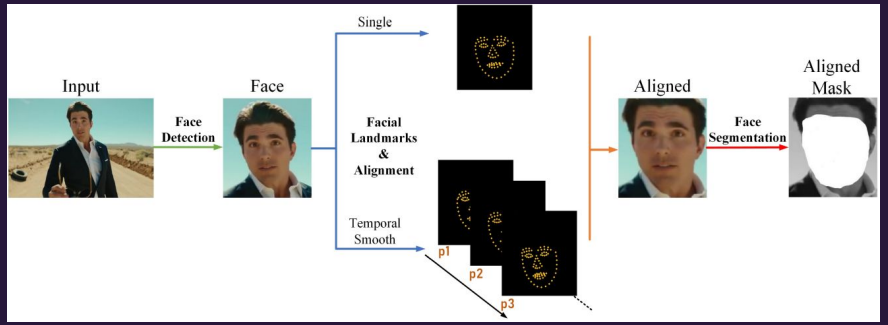
# FSGAN [3]

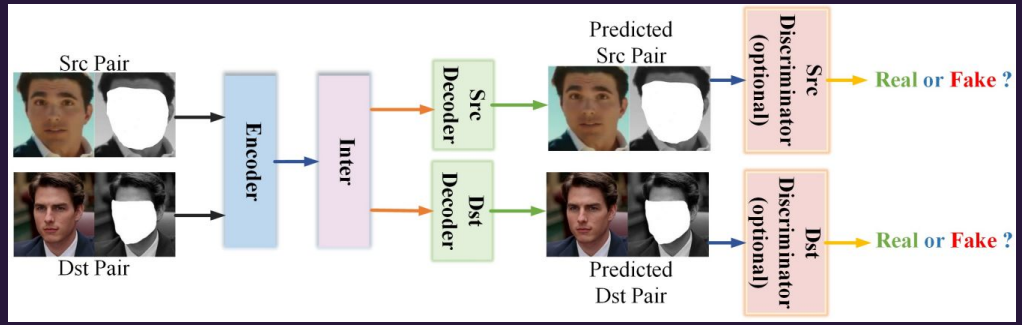
- Subject-agnostic scheme for face swapping and reenactment.
- Source and target faces: short videos;
- Swap time: 30 minutes (colab notebook).

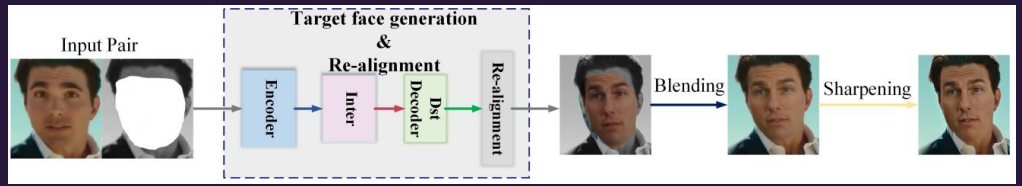


# DeepFaceLive [2]

- Pretrained model of DeepFaceLab;
- Works in real time;
- DeepFaceLab is more powerful (XSeg, one-to-one, ...).

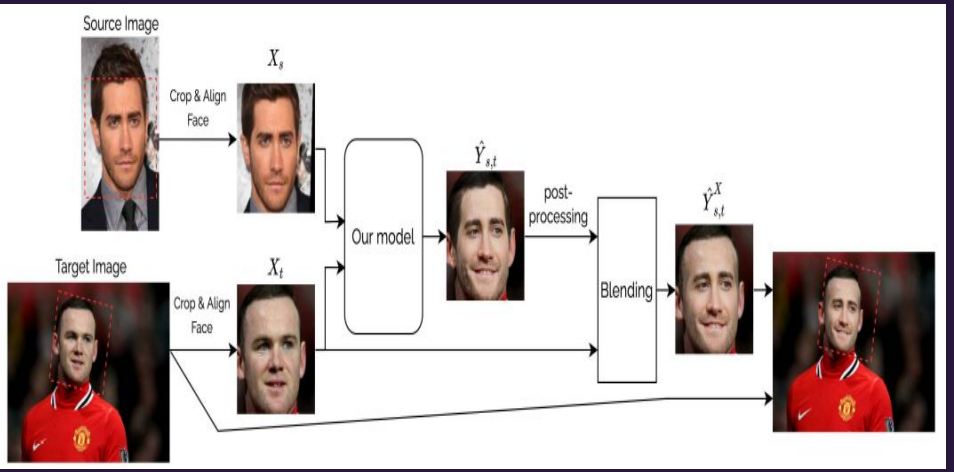


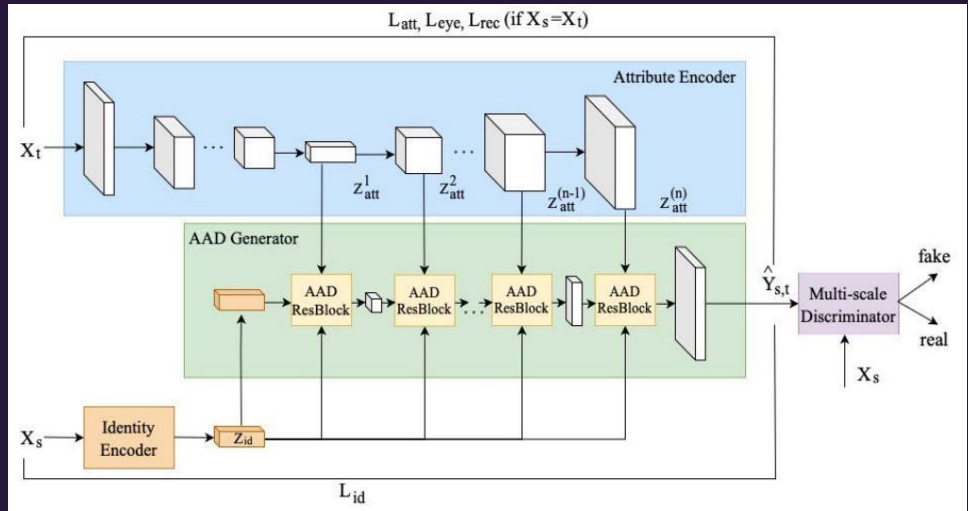



# GHOST [4]

- Single-shot solution: use a single source image to swap the target face in an image or a video.
- Swap time: 5 minutes (colab notebook).



# 03

# PROJECT OVERVIEW

What we have done and final results

# Project overview

- 8 challenges for 3 systems.
- 4 target faces: Diego, Giovanni, Riccardo S., Riccardo Z.
- 4 source faces: Keanu Reeves (Diego), Vin Diesel (Giovanni), Robert Downey Jr. (Riccardo S.), Ryan Reynolds (Riccardo Z.).

- More than 100 videos.

# Metrics


Challenge grade:

- We manually assigned a grade between 0 and 10 to each challenge (higher score → better disrupts the DF generator).


Results score:

- Defining a metric for an automatic system is difficult. The one in GOTCHA does not work well.
- Manually assigned score (higher score → better DF performance).

# 10 Challenges: head rotation

- Move the head with strong rotation as 90° horizontally or face up towards the ceiling.
- Assess whether the model can generate unseen views.
- All models cope badly with this challenge.

DeepFaceLive FSGAN GHOST

2

# Challenges: expression

- Perform different expressions.
- Assess whether the landmark detector and face alignment network can track this changes.
- All model are able to track facial movements, even GHOST trained with only one image.

DeepFaceLive FSGAN GHOST

# 9 Challenges: covering face

- Occlude the face hovering hands in front of it.
- Assess segmentation network capabilities.
- Only DFL is capable to cope with this challenge, even though some parts of the fingers are blended within the face.

DeepFaceLive FSGAN GHOST

# 7 Challenges: poking cheek

- Poke cheek to induce morphological changes in the face.
- Assess whether the landmark detector and face alignment network can track this changes.
- FSGAN performs the worst, GHOST generates some visible artifacts but maintains the finger, DFL blend the finger but can track morphological changes.

DeepFaceLive FSGAN GHOST

5

# Challenges: speaking

- Speak to generate lips movement.
- Assess whether the landmark detector and face alignment network can track this changes.
- FSGAN is not able to reproduce lips movements even though teeth and mouth motion is preserved, DFL excel in this task.

DeepFaceLive FSGAN GHOST

# Challenges: stand up

- Stand up in order to hide the face from the camera.
- Assess whether the face detector is capable of re-detecting a face that was hidden/gone out of the scene.
- FSGAN has trouble when the face is not detected, GHOST works but generate some artifact when face is only partially visible, DFL stops the stream if face not detected.

DeepFaceLive FSGAN GHOST

8*

# Challenges: tongue out

- Stick a portion of the tongue out.
- Assess whether the segmentation network is capable of detecting the tongue.
- All the models perform poorly, FSGAN is the best without achieving satisfying results.

* sticking the tongue out could be a task not appropriate when requested to an high profile person.

DeepFaceLive FSGAN GHOST

6

# Challenges: close up

- Move the face close to the camera.
- Assess whether the face detector is capable of detecting partially cut faces and inspect if more artifacts are visible.
- GHOST displays true face if face not detected, FSGAN is more resistant , DFL performs reasonably good even though artifacts appear if too close to camera.

DeepFaceLive FSGAN GHOST

# Lighting conditions

- Tested different lighting conditions (frontal or from the side).
- Adds difficulty to the system (shadows).
- DFL performs best even with scarce light, while FSGAN and GHOST not so much.

DeepFaceLive FSGAN GHOST

# FSGAN results

| | Diego | Giovanni | Riccardo S. | Riccardo Z. |
|---|---|---|---|---|
| **Head rotation** | 0 | 0 | 0 | 2 |
| **Tongue out** | 2 | 2 | 4 | 9 |
| **Hand on face** | 3 | 2 | 0 | 2 |
| **Close up** | 2 | 4 | 7 | 6 |
| **Standup** | 0 | 0 | 0 | 0 |
| **Poke cheek** | 0 | 0 | 0 | 0 |
| **Expression** | 8 | 7 | 6 | 7 |
| **Speaking** | 6 | 7 | 7 | 6 |
| **Realism** | **6** | **4** | **4** | **5** |

# GHOST results

| | Diego | Giovanni | Riccardo S. | Riccardo Z. |
|---|---|---|---|---|
| **Head rotation** | 1 | 0 | 7 | 3 |
| **Tongue out** | 0 | 4 | 3 | 1 |
| **Hand on face** | 2 | 3 | 3 | 2 |
| **Close up** | 0 | 0 | 9 | 7 |
| **Standup** | 8 | 7 | 9 | 7 |
| **Poke cheek** | 2 | 10 | 4 | 5 |
| **Expression** | 8 | 8 | 9 | 4 |
| **Speaking** | 7 | 8 | 5 | 4 |
| **Realism** | **8** | **7** | **8** | **7** |

# DeepFaceLive results

| | Diego | Giovanni | Riccardo S. | Riccardo Z. |
|---|---|---|---|---|
| **Head rotation** | 6 | 5 | 4 | 4 |
| **Tongue out** | 0 | 1 | 0 | 1 |
| **Hand on face** | 4 | 4 | 9 | 4 |
| **Close up** | 8 | 7 | 8 | 8 |
| **Standup** | 5 | 5 | 5 | 5 |
| **Poke cheek** | 3 | 8 | 3 | 6 |
| **Expression** | 9 | 8 | 9 | 8 |
| **Speaking** | 9 | 9 | 9 | 9 |
| **Realism** | **9** | **9** | **9** | **9** |

# Networks comparison

| | FSGAN | GHOST | DeepFaceLive |
|---|---|---|---|
| **Head rotation** | 0.5 | 2.8 | **4.8** |
| **Tongue out** | **4.2** | 2 | 0.5 |
| **Hand on face** | 1.8 | 2.5 | **5.3** |
| **Close up** | 5 | 4 | **7.8** |
| **Standup** | 0 | **7.8** | 5 |
| **Poke cheek** | 0 | **5.3** | 5 |
| **Expression** | 7 | 7.3 | **8.5** |
| **Speaking** | 6.5 | 6 | **9** |
| **Average** | **3.1** | **4.7** | **5.7** |
| **Realism** | **4.8** | **7.5** | **9** |
| **Lighting conditions** | **2** | **6** | **9** |

# 04

# OUTLOOK

Future works and improvements

# Outlook

- (Improve) Defining (the) a metric (from GOTCHA) to evaluate the response to the challenges;
- Use a face re-identification network to detect if before and during the challenge the face is of the same person (with a segmentation network as preprocessing);
- Implementing an automatic system that proposes the challenges and gives a result by evaluating them.

# 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] Alexander Groshev et al.: GHOST: A New Face Swap Approach for Image and Video Domains. https://ieeexplore.ieee.org/abstract/document/9851423

[5] Martin Anderson: To Uncover a DeepFake Video Call, Ask the Caller to Turn Sideways. metaphysic.ai. https://metaphysic.ai/to-uncover-a-deepfake-video-call-ask-the-caller-to-turn-sideways/

[6] Deepfakes attacks in Remote Identification and Countermeasures. https://antispoofing.org/Deepfake_Attacks_in_Remote_Identification_and_Countermeasures

# References II


[7] iProov: The threat of Deepfakes. https://www.iproov.com/wp-content/uploads/2021/05/iProov-Deepfakes-Report.pdf

[8] Deepfakes vs biometric KYC verification. https://sensity.ai/blog/deepfake-detection/deepfakes-vs-kyc-biometric-verification/

[9] James Vincent, Liveness tests used by banks to verify ID are 'extremely vulnerable' to deepfake attacks, The Verge, https://www.theverge.com/2022/5/18/23092964/deepfake-attack-facial-recognition-liveness-test-banks-sensity-report

[10] European Union Agency for Cybersecurity, Remote Identity Proofing - Attacks & Countermeasures, https://www.enisa.europa.eu/publications/remote-identity-proofing-attacks-countermeasures