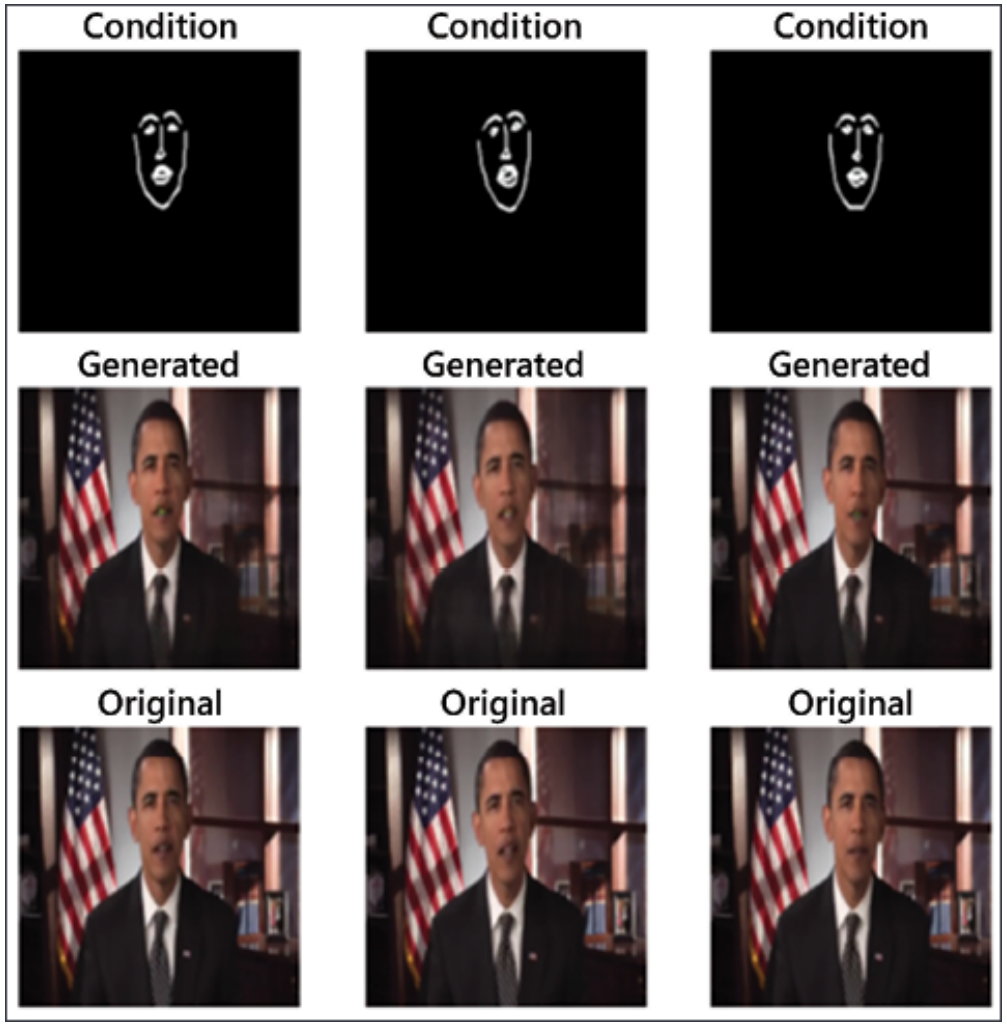
The Secret Ingredients to Train DeepFakes

Raghav Bali

Agenda

- About Me
- What is DeepFakes
- The Setup
- The DeepFakes Landscape
- Feature Set
- GANs and Obama
- Challenges
- Q/A



Raghav Bali

- Senior Manager Data Science at Optum(United HealthGroup)
- A decade's experience involving research & development of enterprise level solutions based on Machine Learning, Deep Learning and Natural Language Processing for real world use-cases.

Optum
2017 - Present

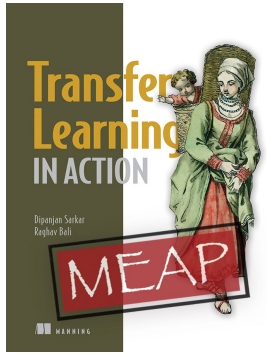
Intel
2015 - 2017

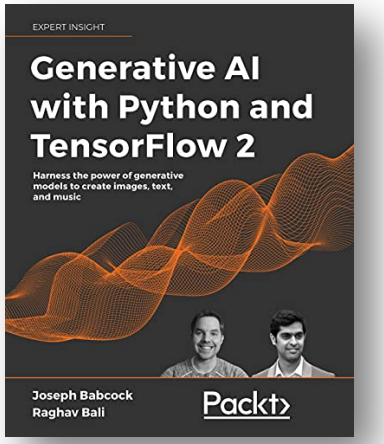
American Express
2014 - 2015

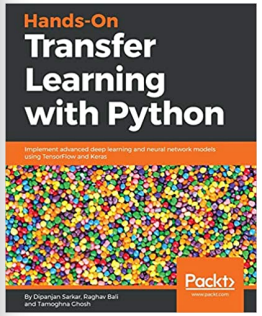
IIIT Bangalore
2012 - 2014

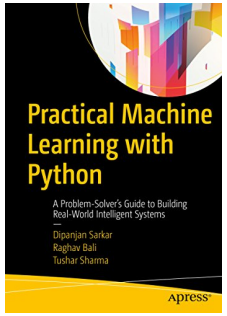
Raghav Bali : Publications

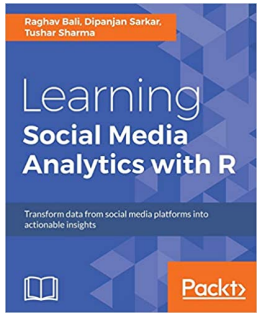
- Talks/Workshops at leading conferences/platforms such as Analytics Vidhya, SpringBoard, ODSC, etc.
- Patents: 7 patents in the field of healthcare, deep learning, machine learning and NLP
- Papers
  - CAIAC 2021, EASTER: Simplifying Text Recognition using only 1D Convolutions
  - CAIAC 2021, A Simple and Interpretable Predictive Model for Healthcare
  - Preprint, Exclusion and Inclusion--A model agnostic approach to feature importance in DNNs
  - IEEE SmartData 2016, Real Time Failure Prediction of Load Balancers and Firewalls
- Books

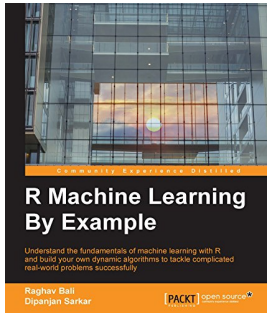


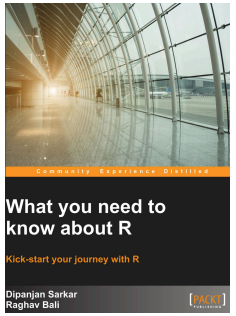












What is DeepFakes

Deepfakes is an all-encompassing term representing content generated using artificial intelligence that seems realistic and authentic to a human being.

Image Sources:

Ronald Reagan in Call of Duty (link) | https://thispersondoesnotexist.com/

Image Sources:

Salvador Dali Museum (link)

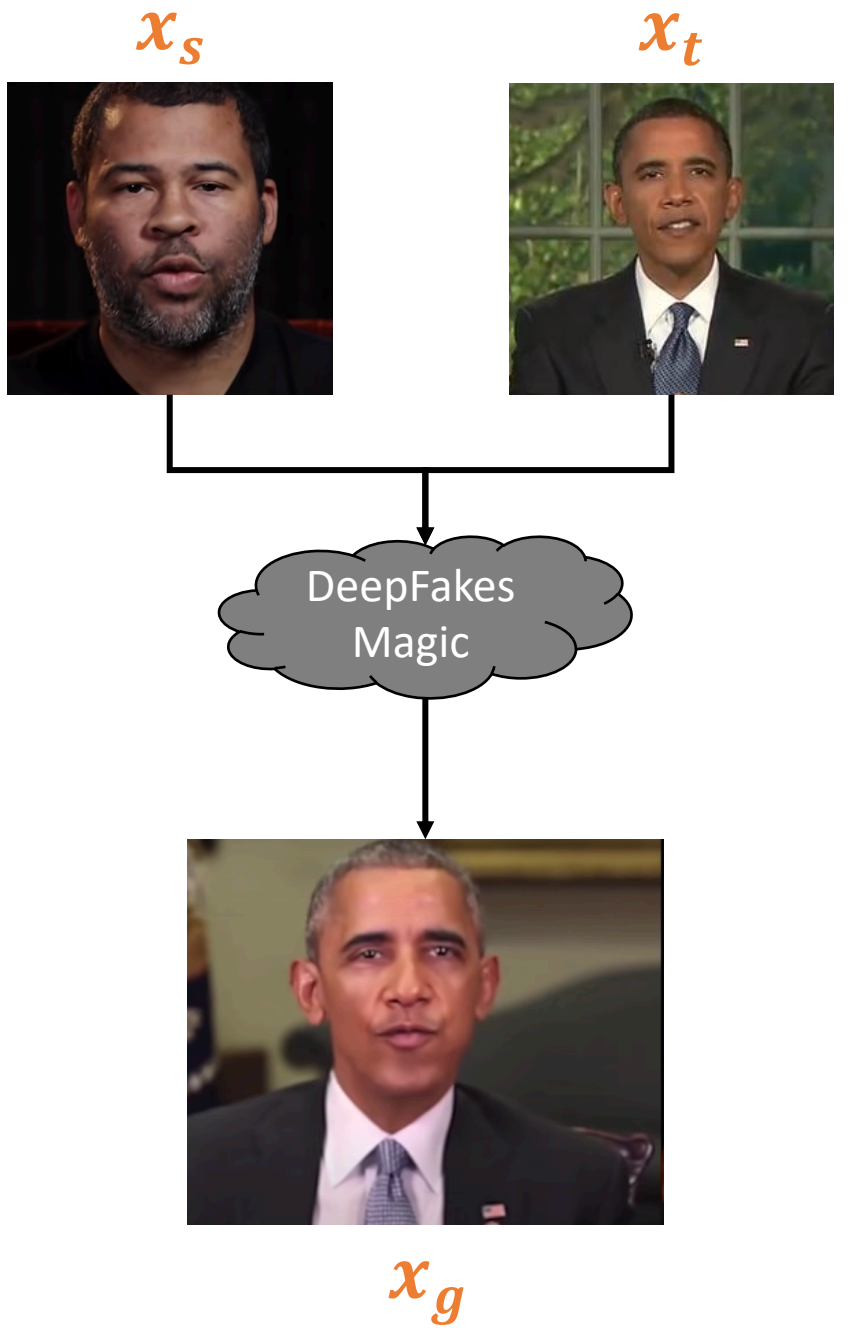
What is DeepFakes

- **Creative Use Cases:**
  - Recreating history and famous personalities
  - Movie translation
  - Fashion
  - Video game characters
  - Stock images
- **Malicious Use Cases:**
  - Pornography
  - Impersonation

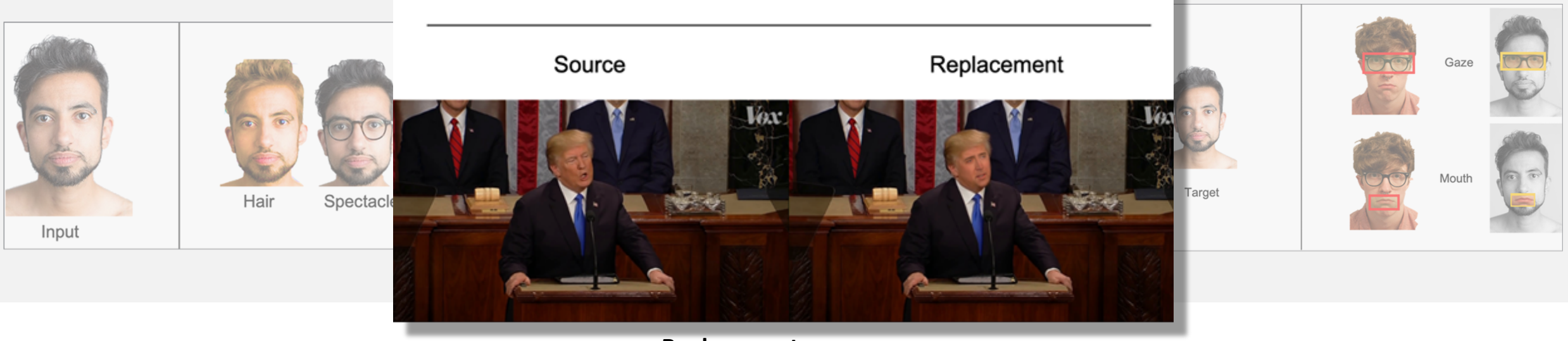
The Setup

The Setup

- A typical deepfake setup requires a **source**, **a target**, to get **the generated content**.
  - The source, denoted with **subscript *s***, is the driver identity that controls the required output
  - The target, denoted with **subscript *t***, is the identity being faked
  - The generated content, denoted with **subscript *g***, is the result following the transformation of the source to the target.

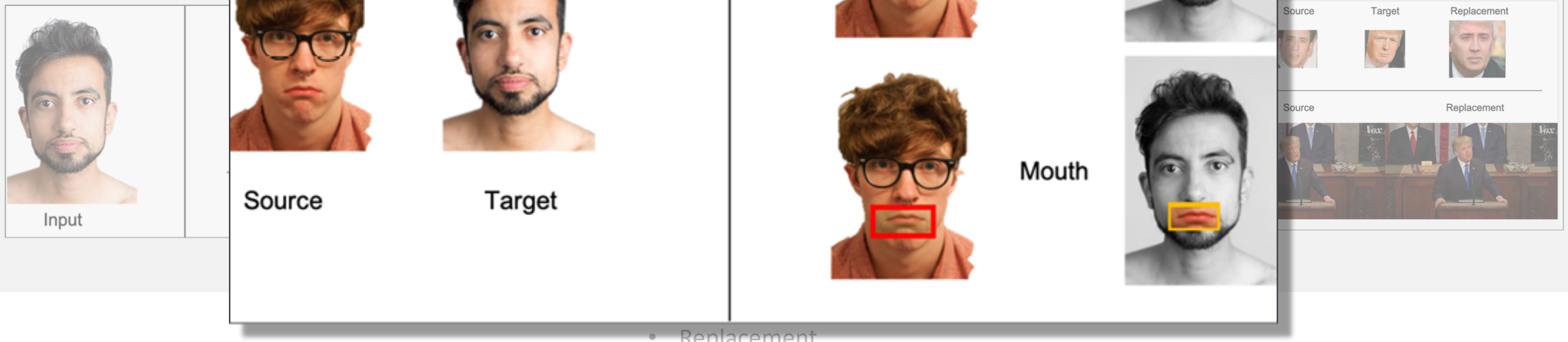


DeepFakes Landscape



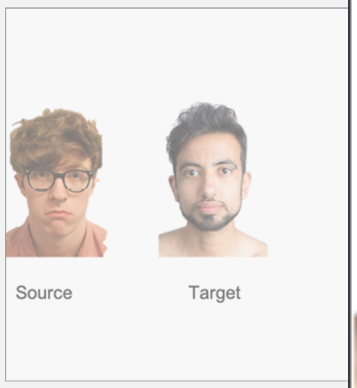
The DeepFakes Landscape

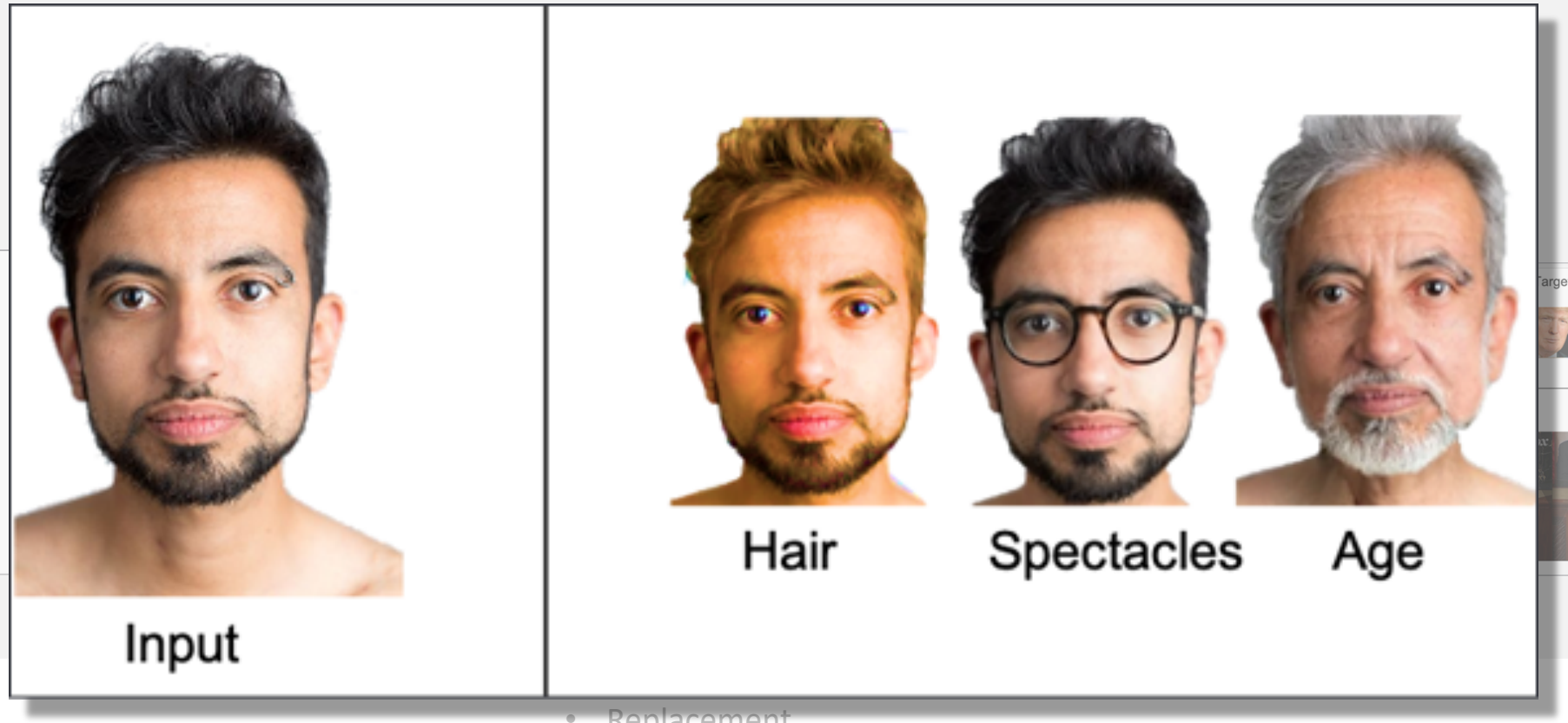
- **Replacement**
  The aim is to replace specific content of the target (x_t) with that from the source (x_s)
- Re-enactment
  Re-enactment methods are utilized to capture characteristics such as the pose, expression, and gaze of the target to improve upon the believability of the generated content.
- Editing
  Editing involves manipulation of clothing, age, ethnicity, gender, hair, and so on.

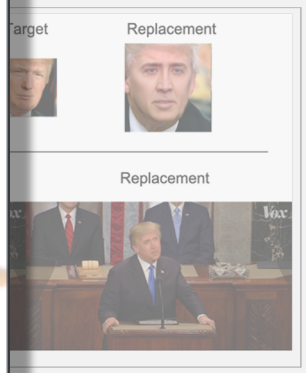


The DeepFakes Landscape

- Replacement
  The aim is to replace specific content of the target (xt) with that from the source (xs)
- **Re-enactment**
  Re-enactment methods are utilized to capture characteristics such as the pose, expression, and gaze of the target to improve upon the believability of the generated content.
- Editing
  Editing involves manipulation of clothing, age, ethnicity, gender, hair, and so on.







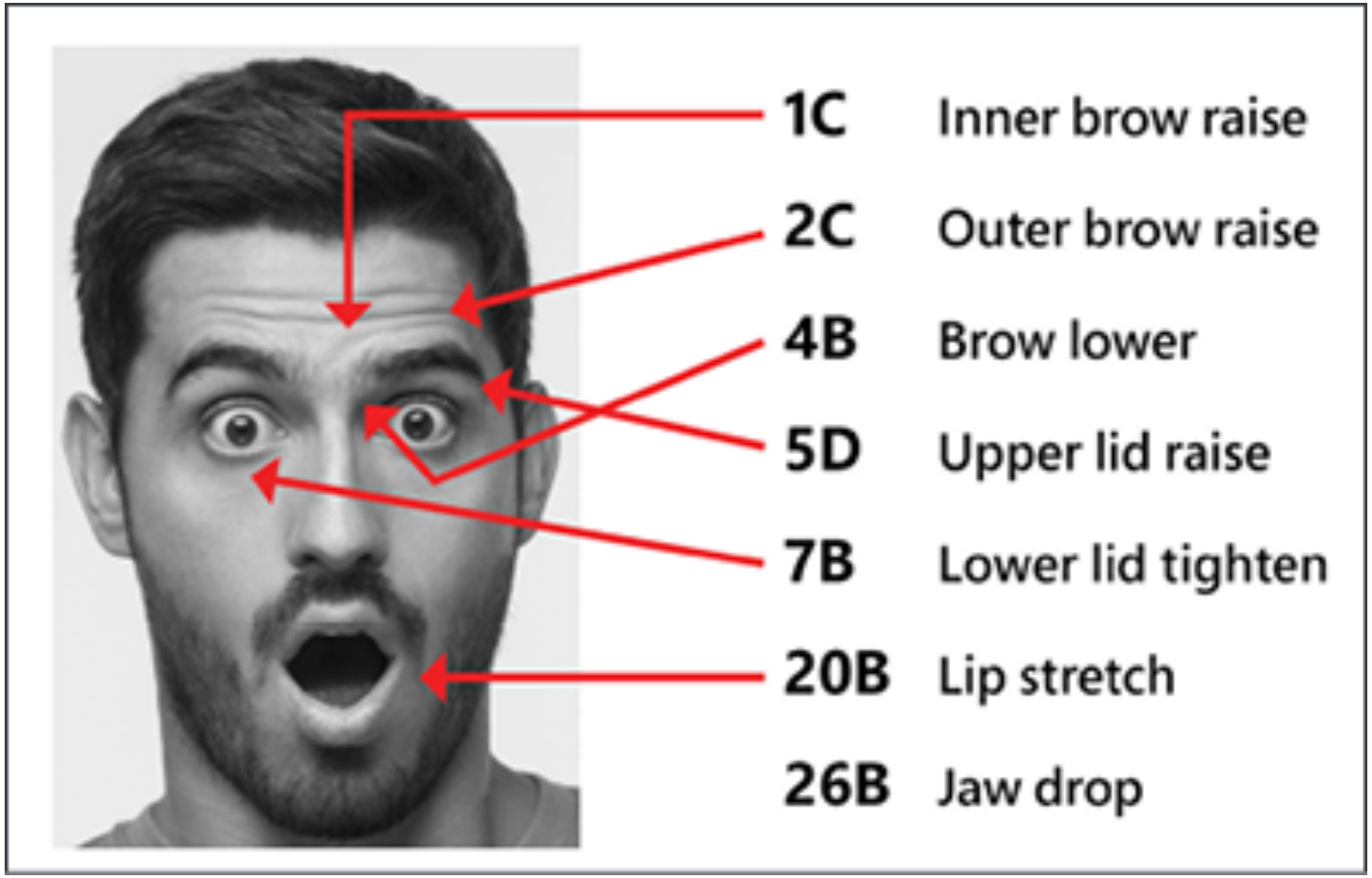
The DeepFakes Landscape

- Replacement
  The aim is to replace specific content of the target (xt) with that from the source (xs)
- Re-enactment
  Re-enactment methods are utilized to capture characteristics such as the pose, expression, and gaze of the target to improve upon the believability of the generated content.
- **Editing**
  Editing involves manipulation of clothing, age, ethnicity, gender, hair, and so on.

Feature Set

Feature Set

- **Facial Action Coding System (FACS)**
  Anatomy-based system for understanding facial movements. It is one of the most extensive and accurate coding systems for analysing facial muscles to understand expressions and emotions.
- 3D Morphable Model (3DMM)
- Facial Landmarks

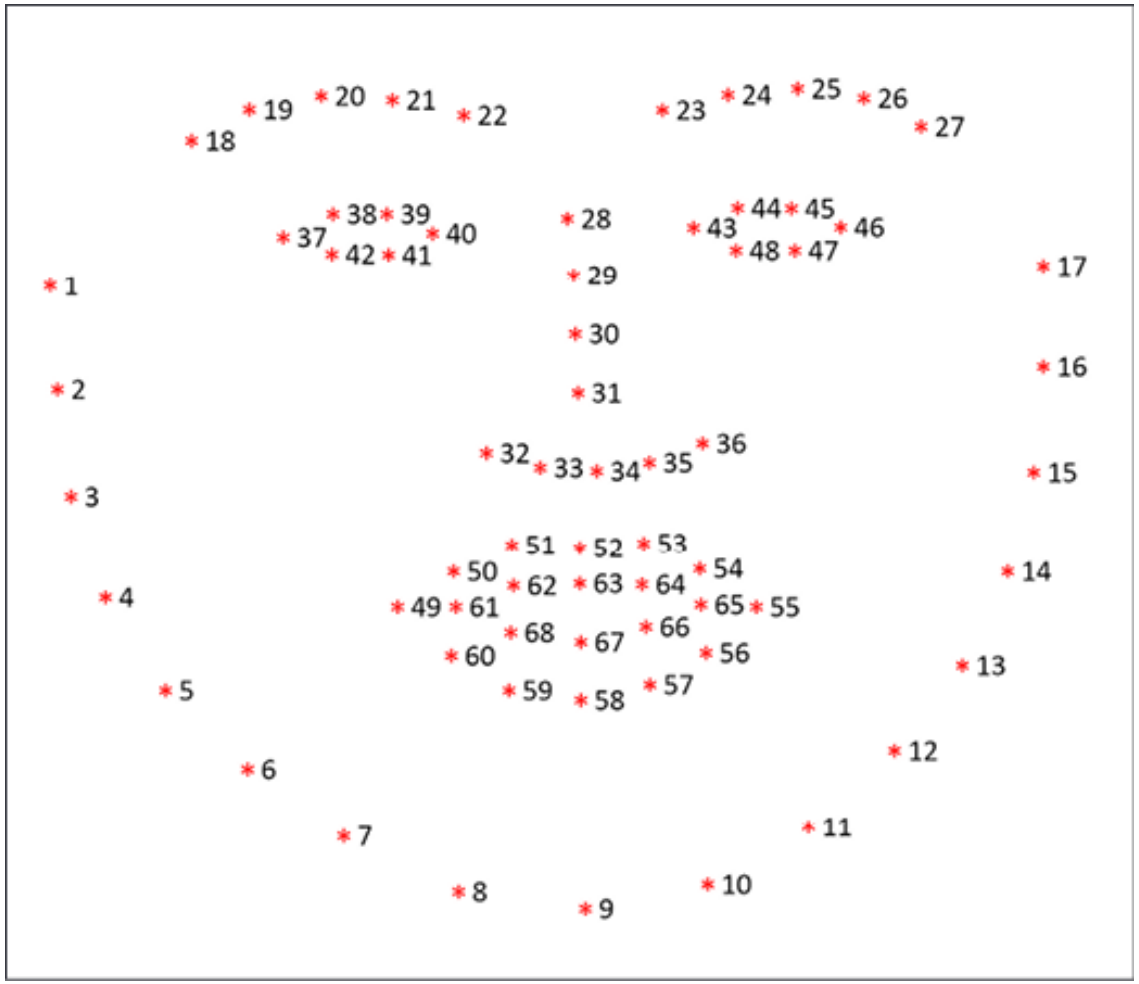


Feature Set

- Facial Action Coding System (FACS)
- **3D Morphable Model (3DMM)**
  A method of inferring a complete 3D facial surface from a 2D image. This is a powerful statistical method that can model human face shape and texture along with pose and illumination.
- Facial Landmarks

Feature Set

- Facial Action Coding System (FACS)
- 3D Morphable Model (3DMM)
- **Facial Landmarks**
  Facial landmarks are a list of important facial features, such as the nose, eyebrows, mouth, and the corners of the eyes. The goal is the detection of these key features using some form of a regression model.



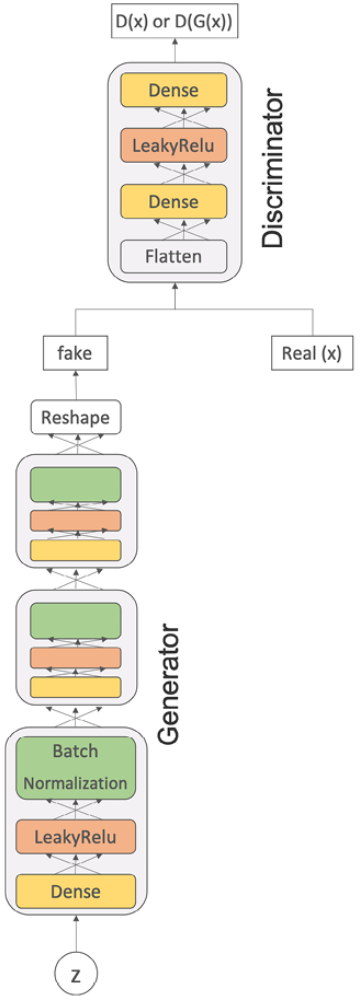
Key Feature Set

- Facial Landmarks
- Hands-on

Generative Adversarial Networks

GANs

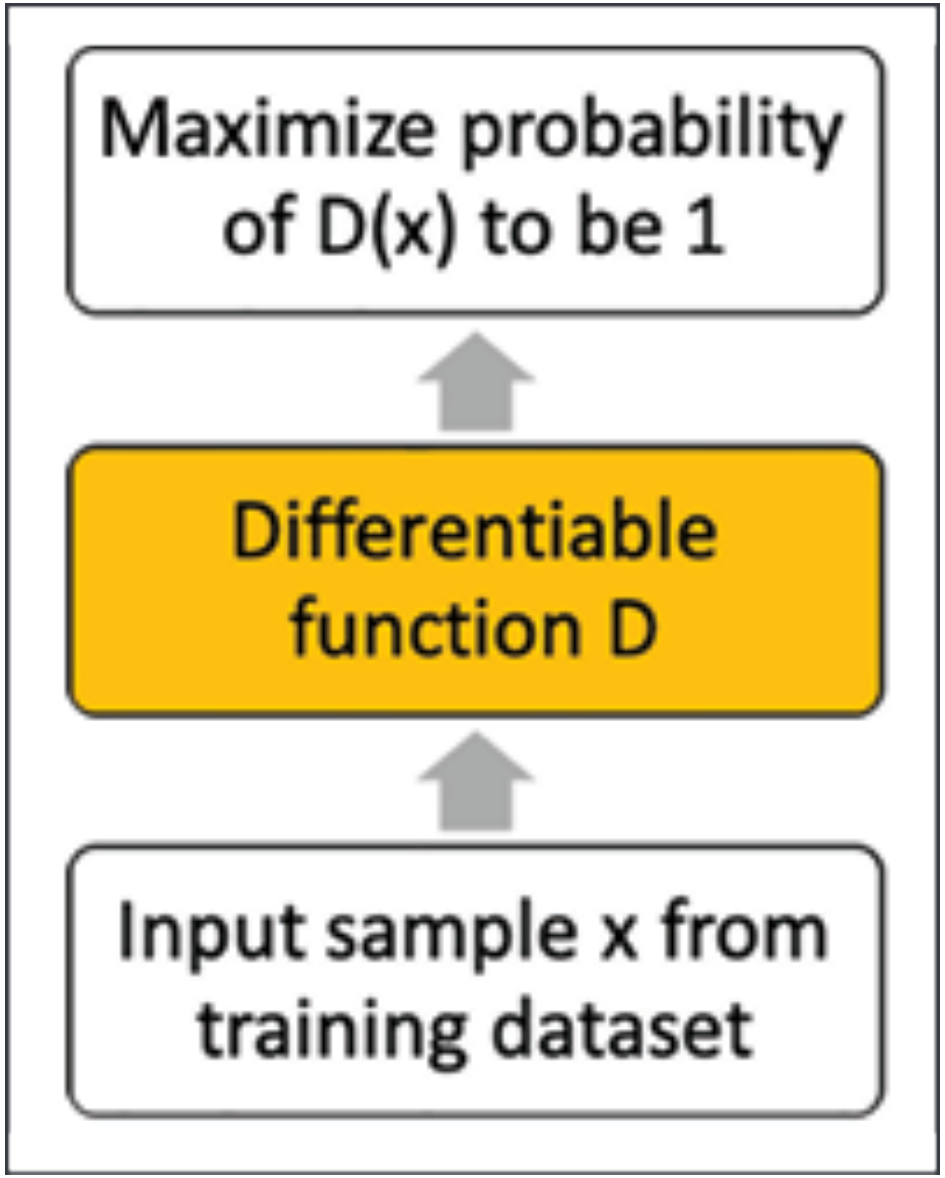
- GANs are implicit density functions which sample directly from the underlying distribution.
- They do this by defining a two-player game of adversaries.
- The adversaries compete against each other under well-defined reward functions and each player tries to maximize its rewards



"Discriminator Vs Generator"

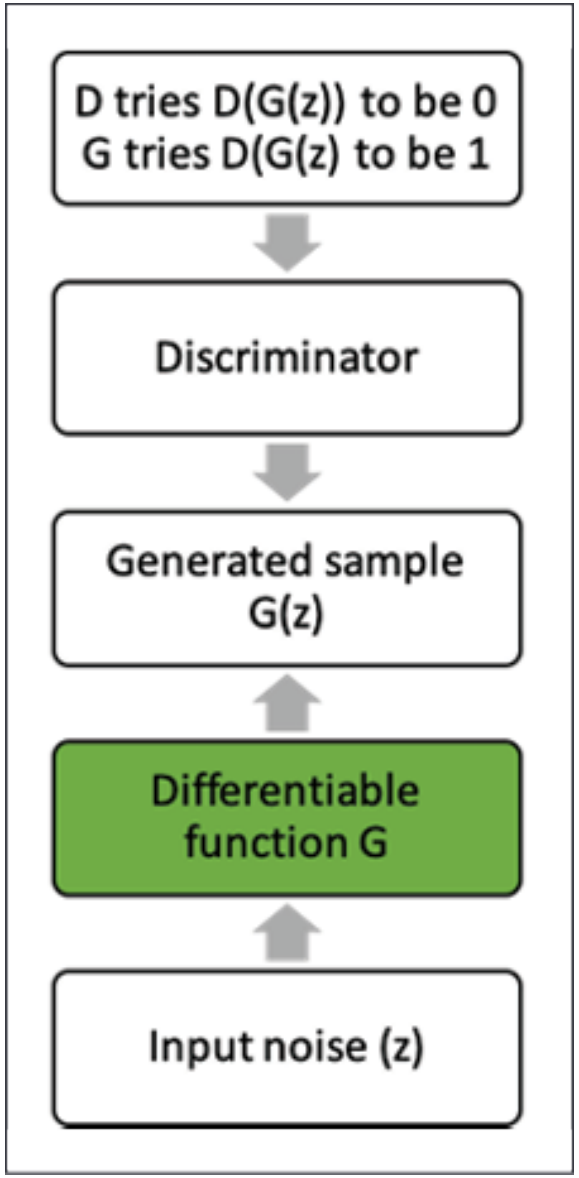
GAN: Discriminator

- This model represents a differentiable function that tries to maximize a probability of 1 for samples drawn from the training distribution.
- The discriminator is also used to classify whether the output from the generator is real or fake.
- We denote the discriminator model as *D* and its output as *D*(*x*).



GAN: Generator

- This model generates samples that are intended to resemble the samples from our training set.
- The model takes random unstructured noise as input (typically denoted as *z*) and tries to create a varied set of outputs.
- We denote the generator as *G* and its output as *G*(*z*).
- We typically use a lower-dimensional *z* as compared to the dimension of the original data, *x*, that is, $z_{dim} \leq x_{dim}$



GAN: Training

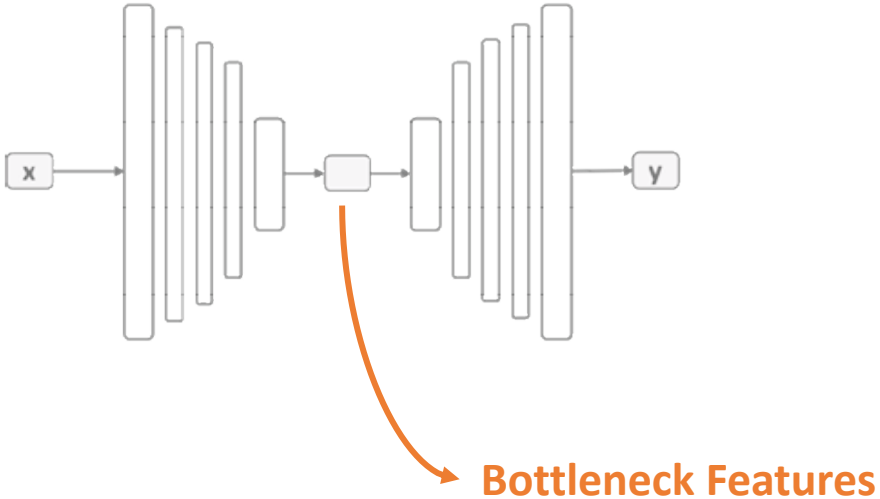
- Training a GAN is like playing this game of two adversaries.
- More formally, this is termed as the minimax game, where the value function *V*(*G*, *D*) is described as follows:

$$min_G max_D V(G,D) = E_{x \sim p_{data}} log\ log\ D(x) + E_{z \sim p_z} log\ log\ (1 - D(G(z)))$$

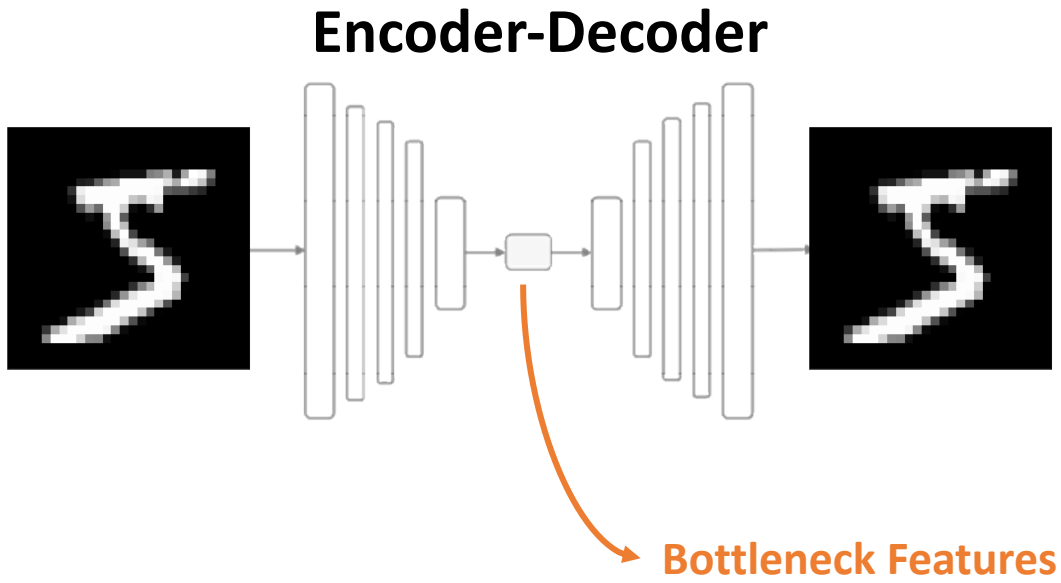
Discriminator Loss **Generator Loss**

GANs : U-Net

Encoder-Decoder

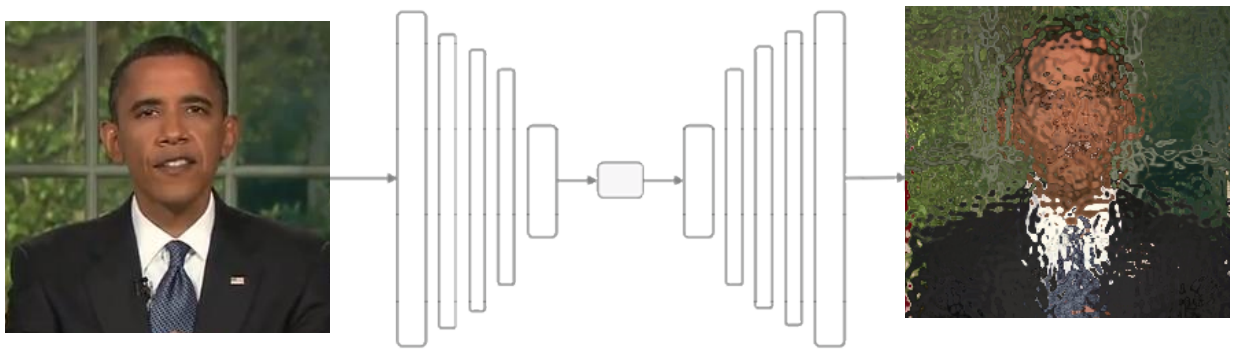


GANs : U-Net



GANs : U-Net

Encoder-Decoder

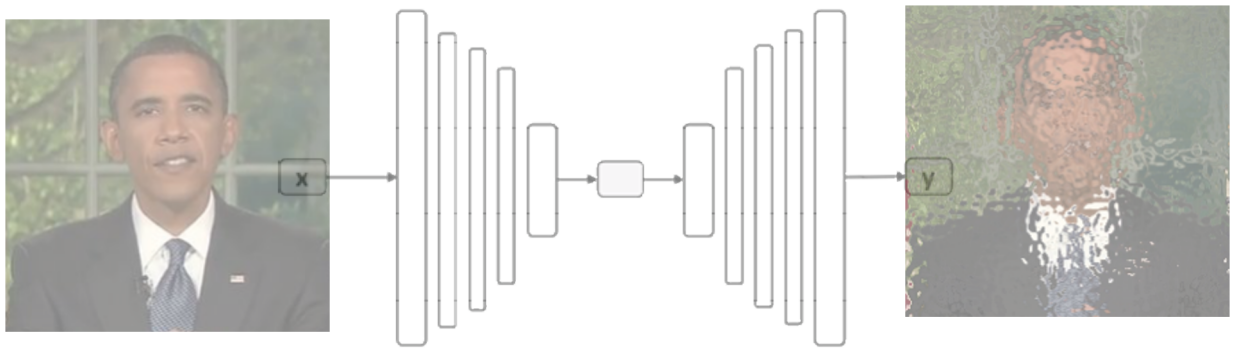


High Resolution Image

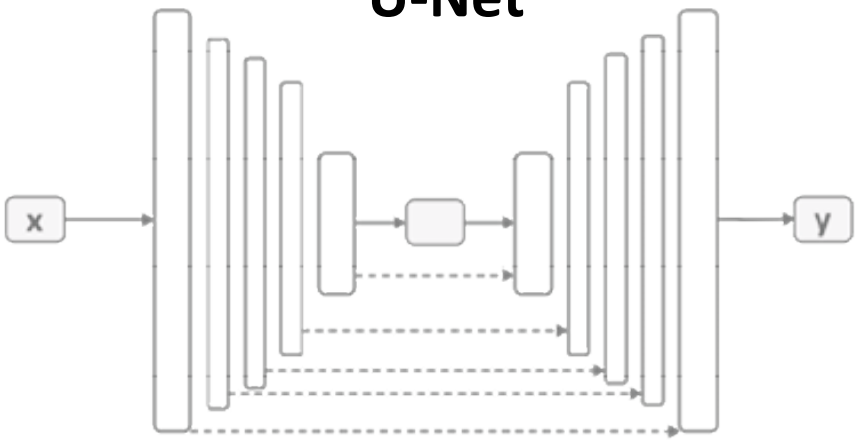
High Reconstruction Loss

GANs : U-Net

Encoder-Decoder



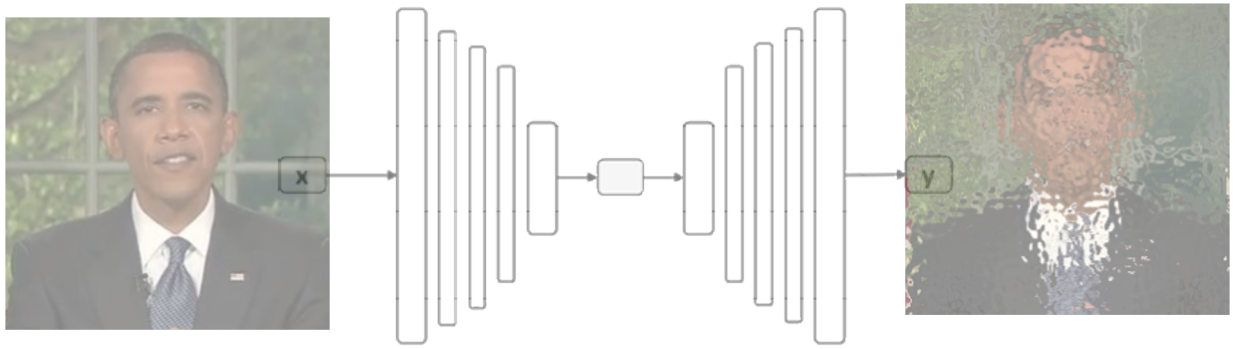
U-Net



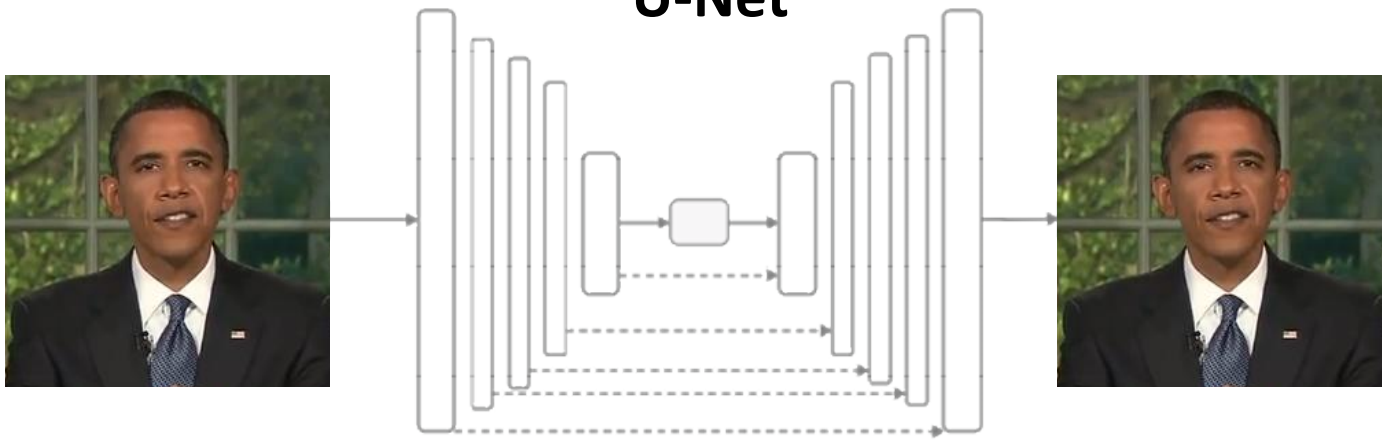
Skip Connections
Between Encoder
and Decoder Blocks

GANs : U-Net

Encoder-Decoder

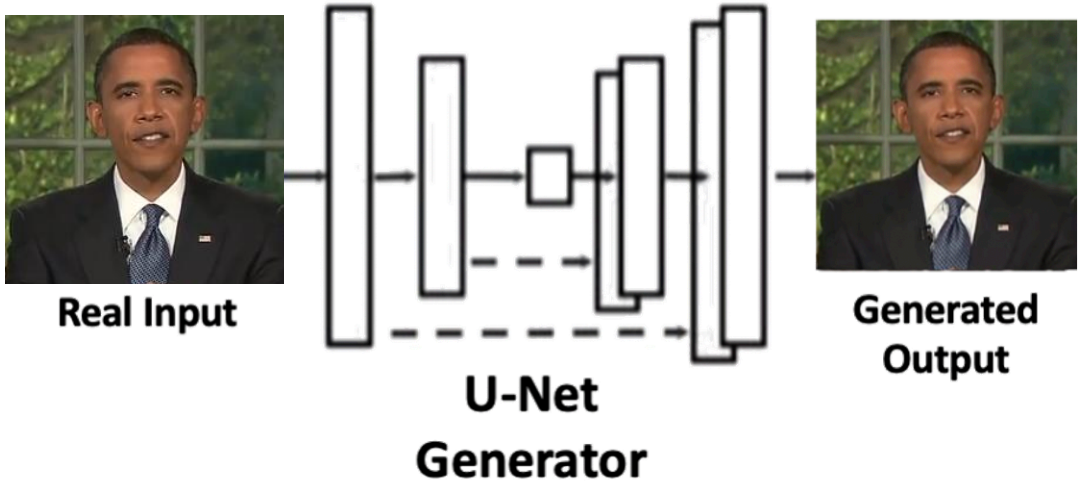


U-Net

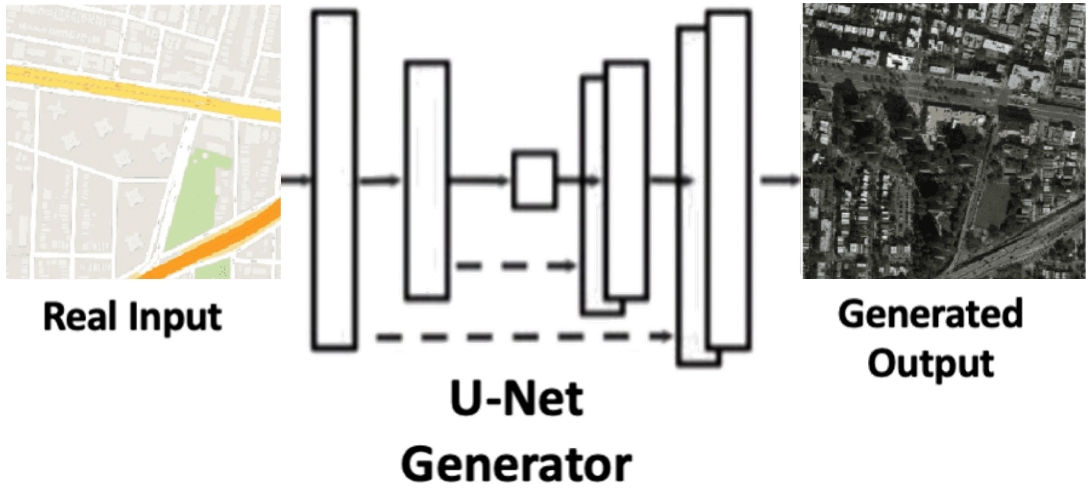


Skip Connections
Between Encoder
and Decoder Blocks

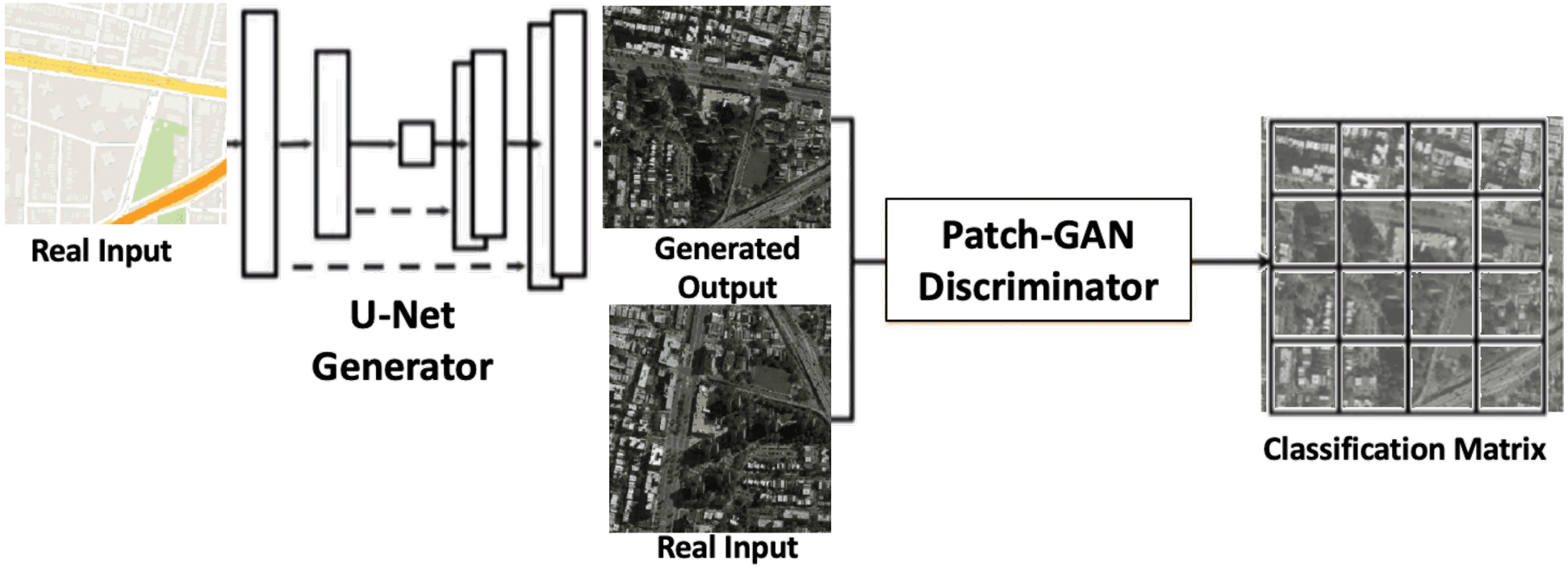
GANs : Pix2Pix



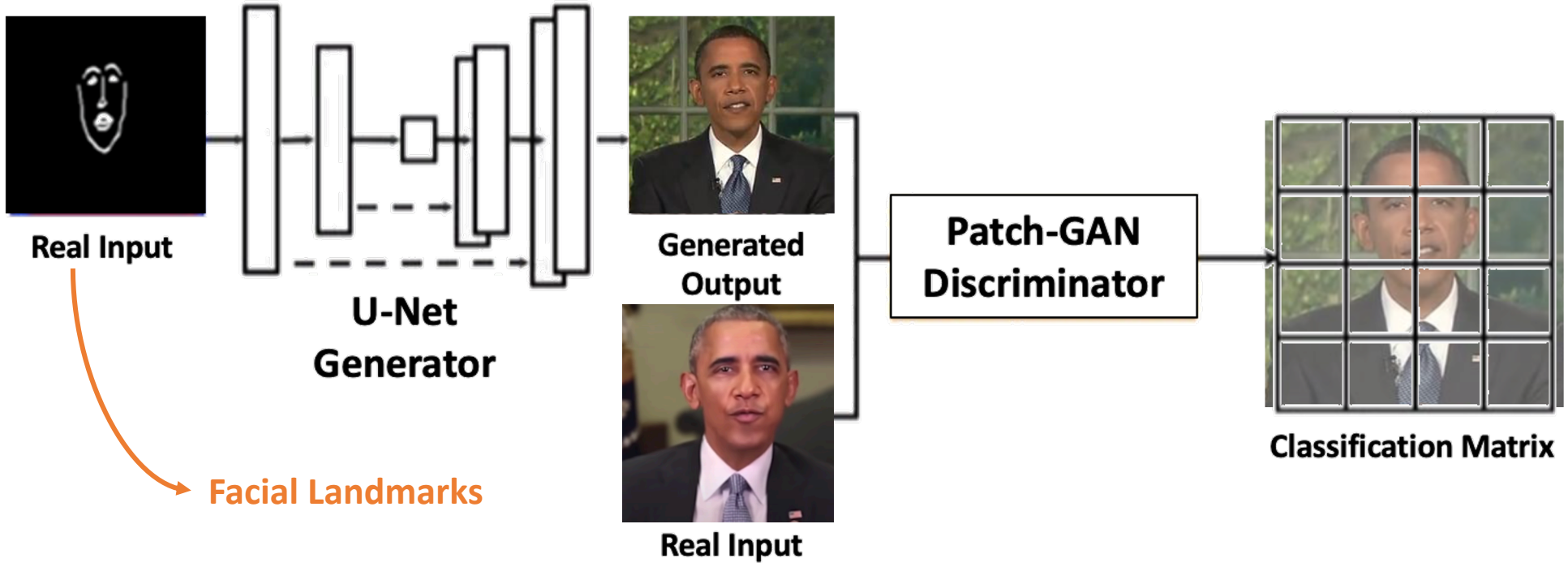
GANs : Pix2Pix

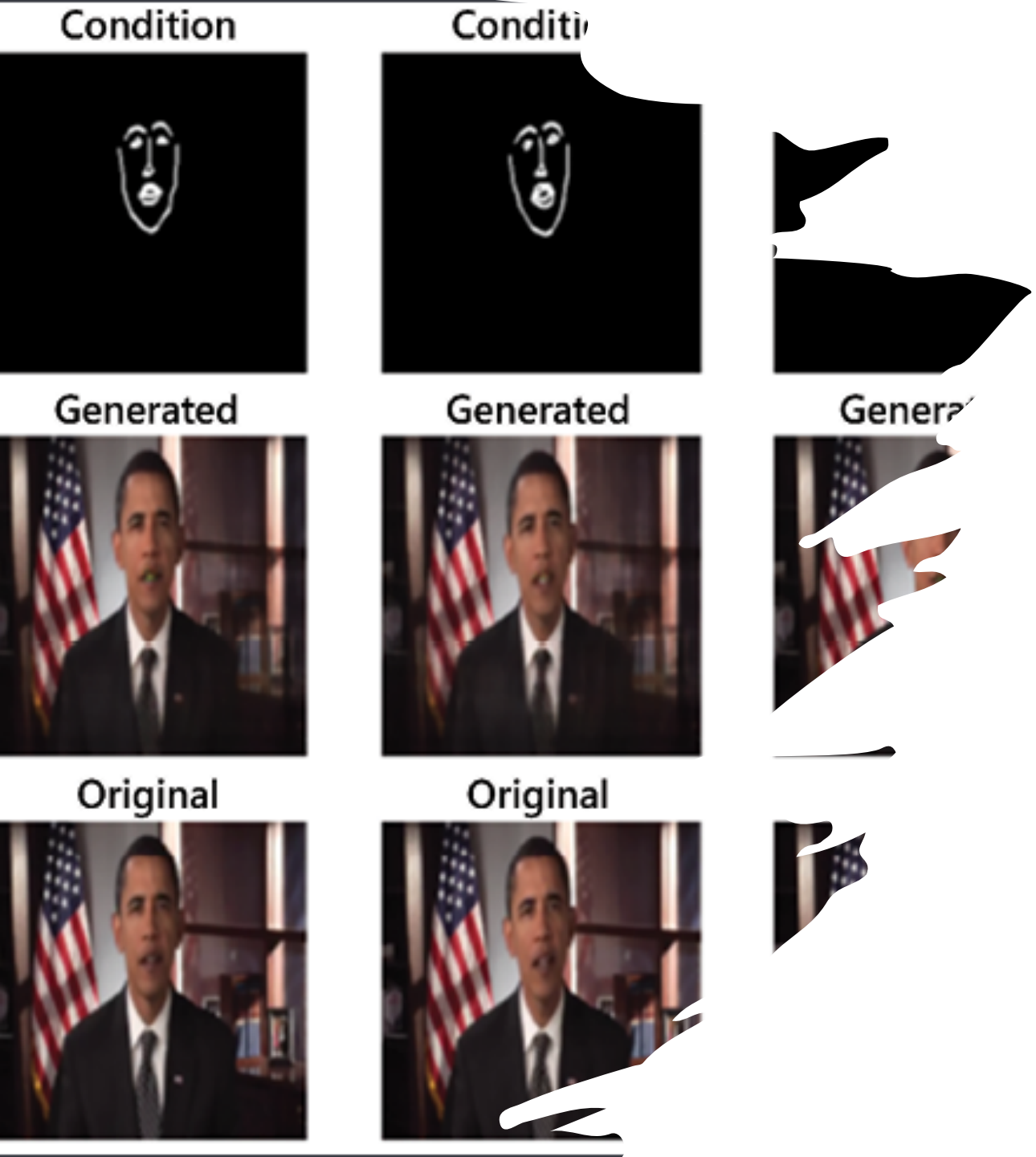


GANs : Pix2Pix



GANs : Pix2Pix





GANs and Obama

- Re-enactment based on pix2pix
- Hands-on

Challenges

- Ethical issues
- Technical challenges
- Generalization
- Occlusions
- Temporal issues

Q/A


KEEP CALM
BECAUSE THERE IS NO
CRET INGRED