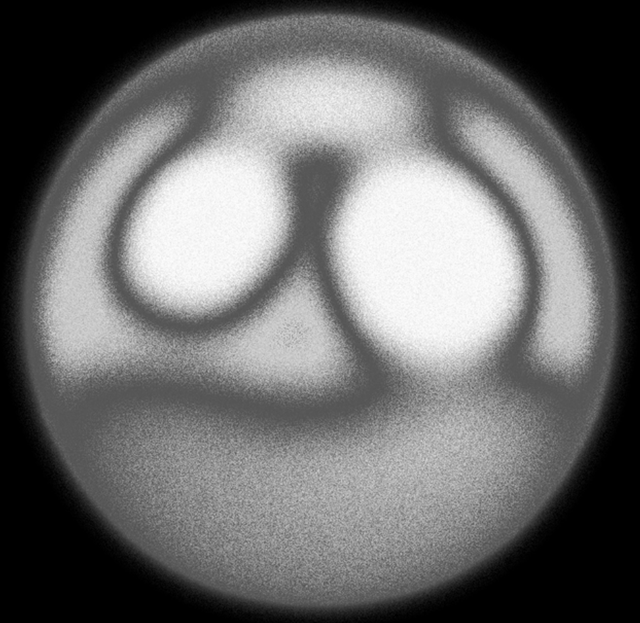
# DeepFake

Advanced Computer Vision
Lisa Golla 2023 Summer term

# Agenda



**What are DeepFakes?**
- Motivation
- Definiton
- Different types

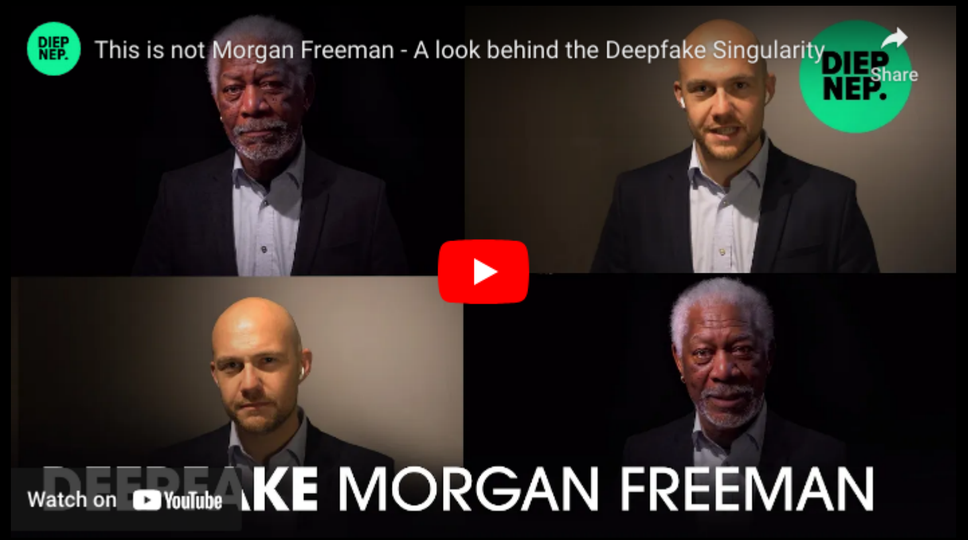
**How to generate a DeepFake?**
- State of the art approaches
- StyleGAN as an example model
- Limitations and challenges

**How to detect DeepFakes?**
- Can you detect the DeepFake?
- Different detection techniques
- Limitations and challenges

**Application and Implication**
- (Mis)use of DeepFakes
- Future directions
- Open discussion?

# Definition

= manipulated or synthetic media, created using deep learning techniques




manipulation of public opinion

spread of misinformation

https://konradweber.ch/2018/11/08/deep-fake-videos-verifizieren/

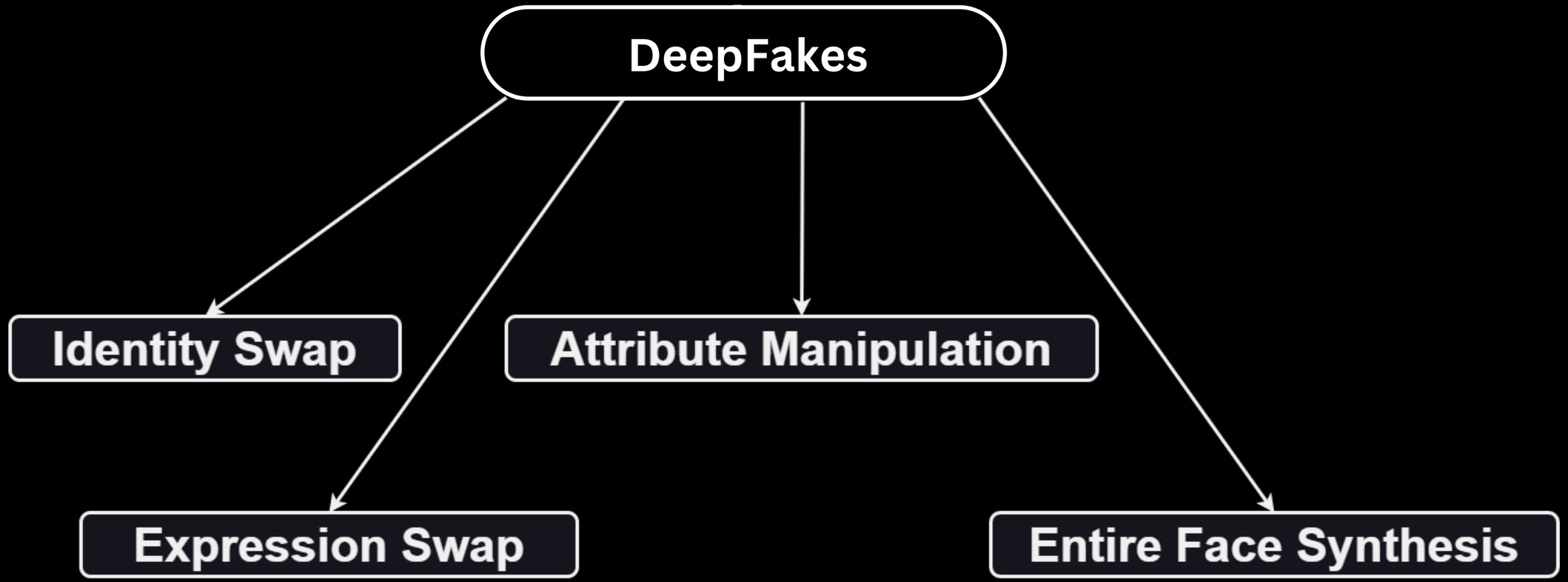
identity theft

defamation

https://www.br.de/radio/bayern2/sendungen/zuendfunk/fail-of-the-week-ki-papst-problem-100.html

[1] [5]

# Types of DeepFakes

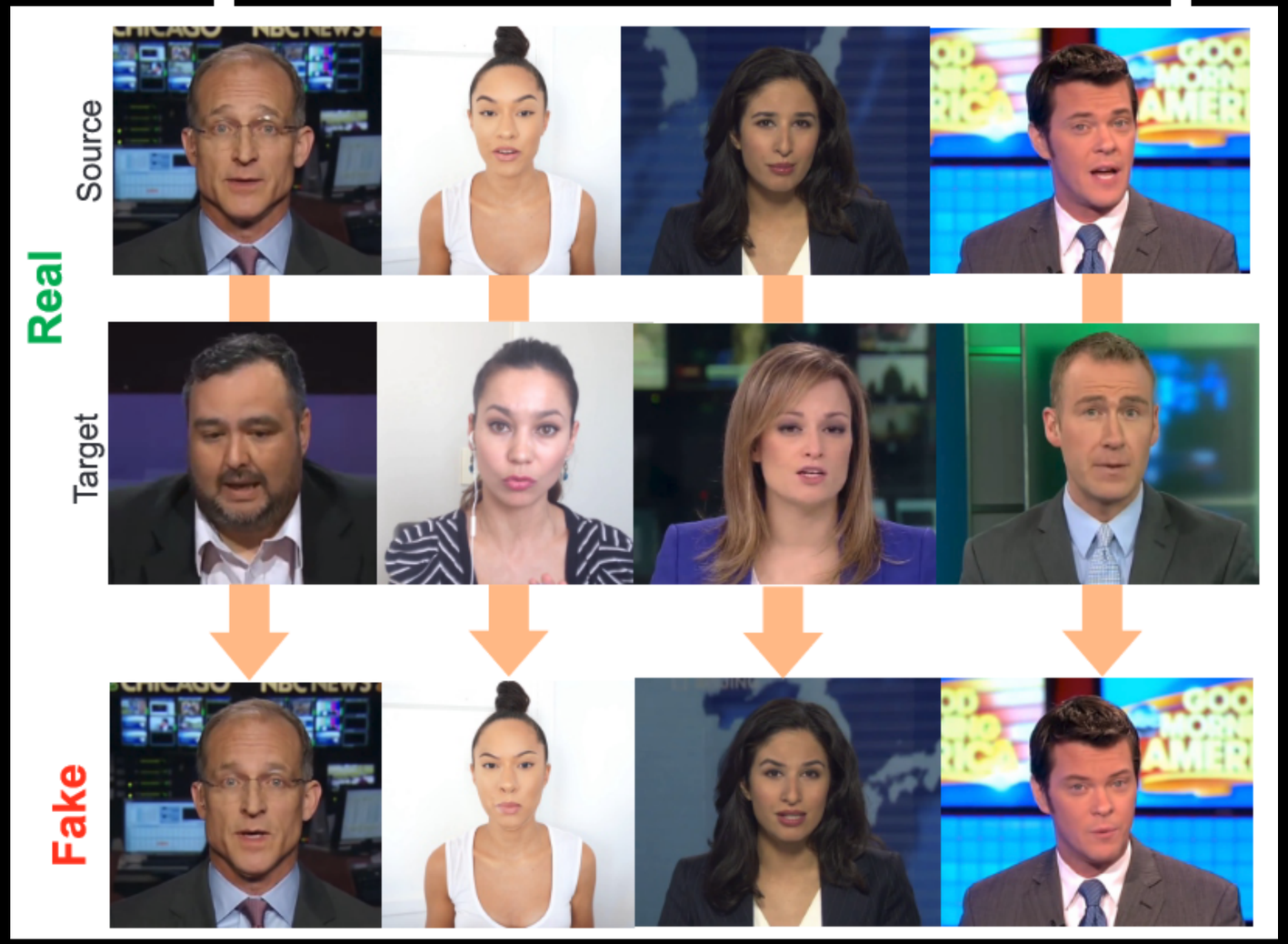

DeepFakes
Identity Swap
Attribute Manipulation
Expression Swap
Entire Face Synthesis

DeepFake

# Identity Swap

DeepFake

# Expression Swap




[8]

DeepFake

# Attribute Manipulation


[8]

DeepFake

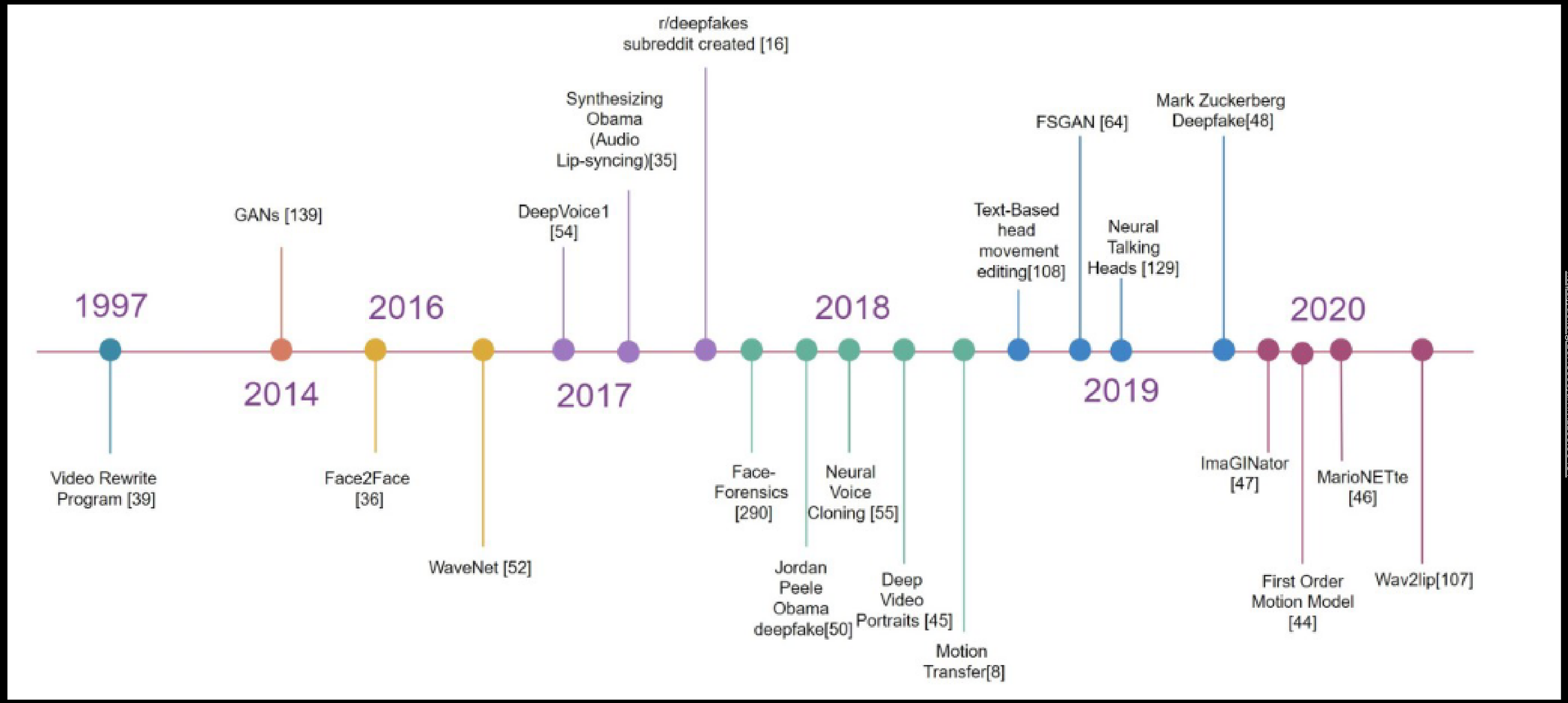
# Face Synthesis

https://thispersondoesnotexist.com/

https://thispersondoesnotexist.com/

[3]

# Evolution




https://doi.org/10.48550/arXiv.2103.00484

Which advancements in CV enabled the progress of deep fakes?

- Computer graphics
- Image and Video Synthesis Techniques
- Facial Landmark detection
- DL techniques
- GAN
- Autoencoders

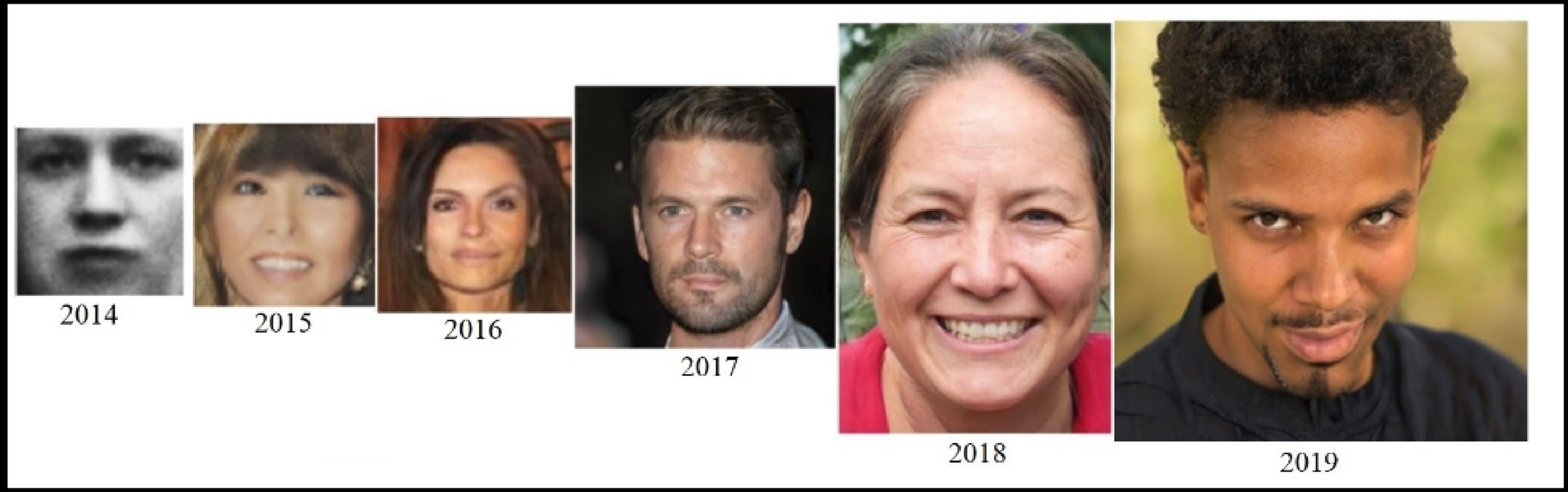
[1]


Artists create Zuckerberg 'deepfake' video
Share
Watch on YouTube


Video Rewrite (1997) - First Deepfake video
Share
Watch on YouTube


Very realistic Tom Cruise Deepfake | AI Tom Cruise
Share
Watch on YouTube

# Increasing improvements in the quality of synthetic faces, as generated by variations of GANs.




https://doi.org/10.48550/arXiv.2103.00484

[1]

# How to generate deep fakes

The method may change depending on the type of deep fake that is created

BUT generally speaking you can say these are the usual steps

### Training generative Model

During training, the model learns to capture visual patterns, features, and representations of relevant characteristics

### Encoding

The generative model is used to encode the features of the source person's face or other relevant attributes into a latent space

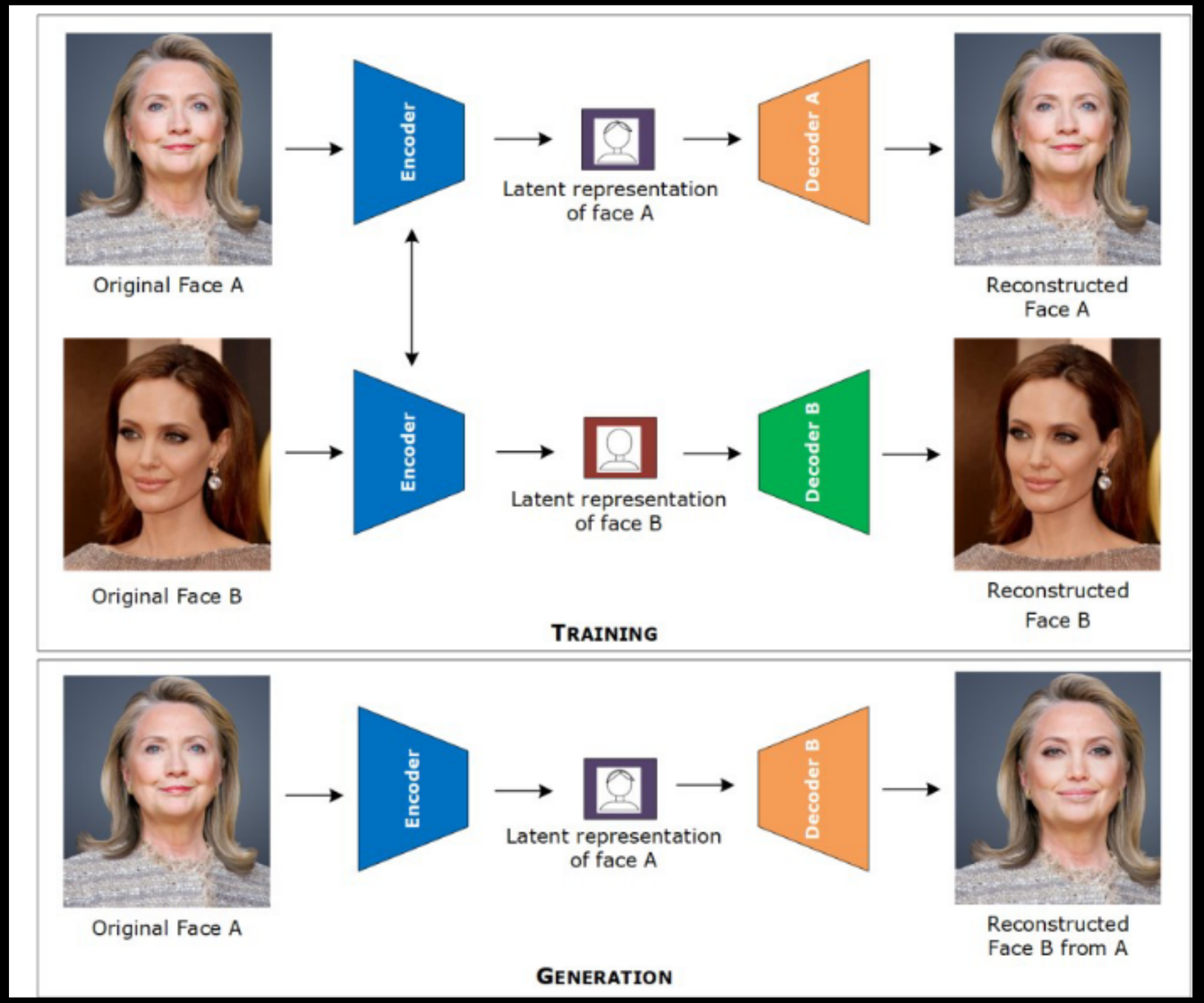
### Manipulation

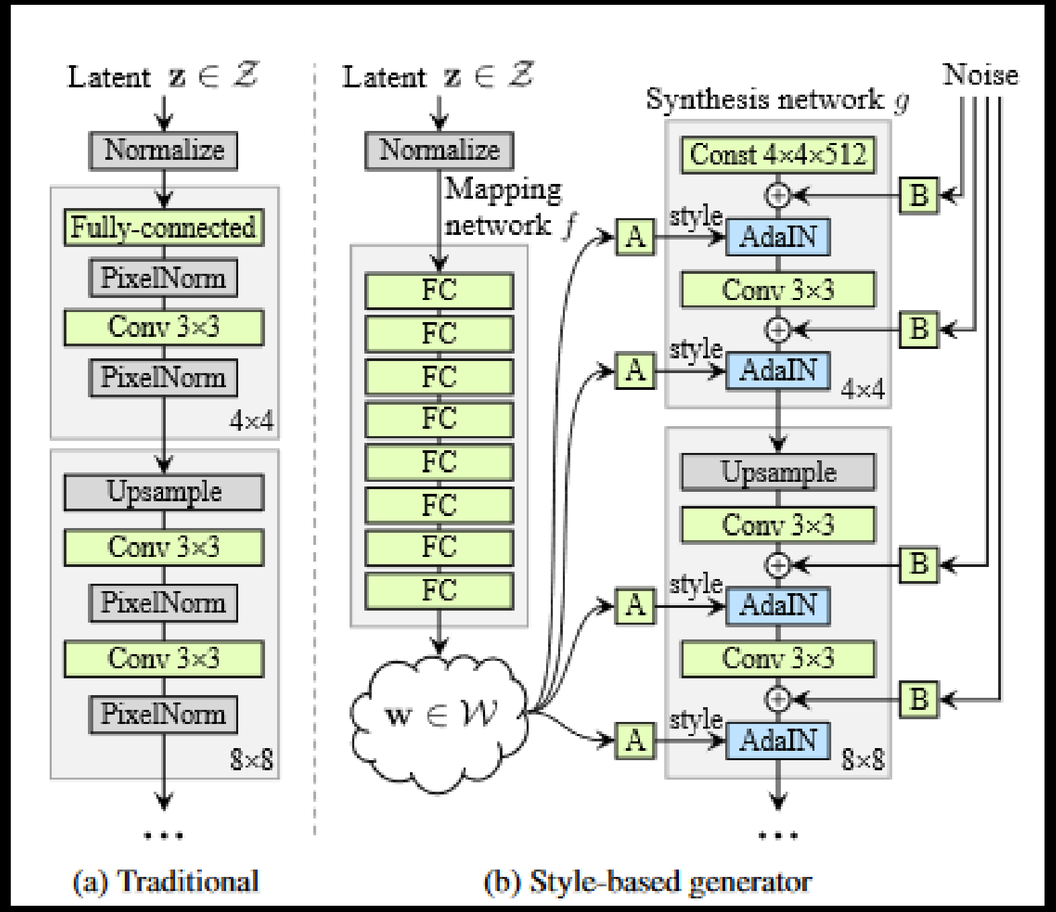
The latent representation is modified or manipulated to introduce desired changes or attributes. For example, facial expressions can be altered, identity can be swapped.

### Decoding and synthesis

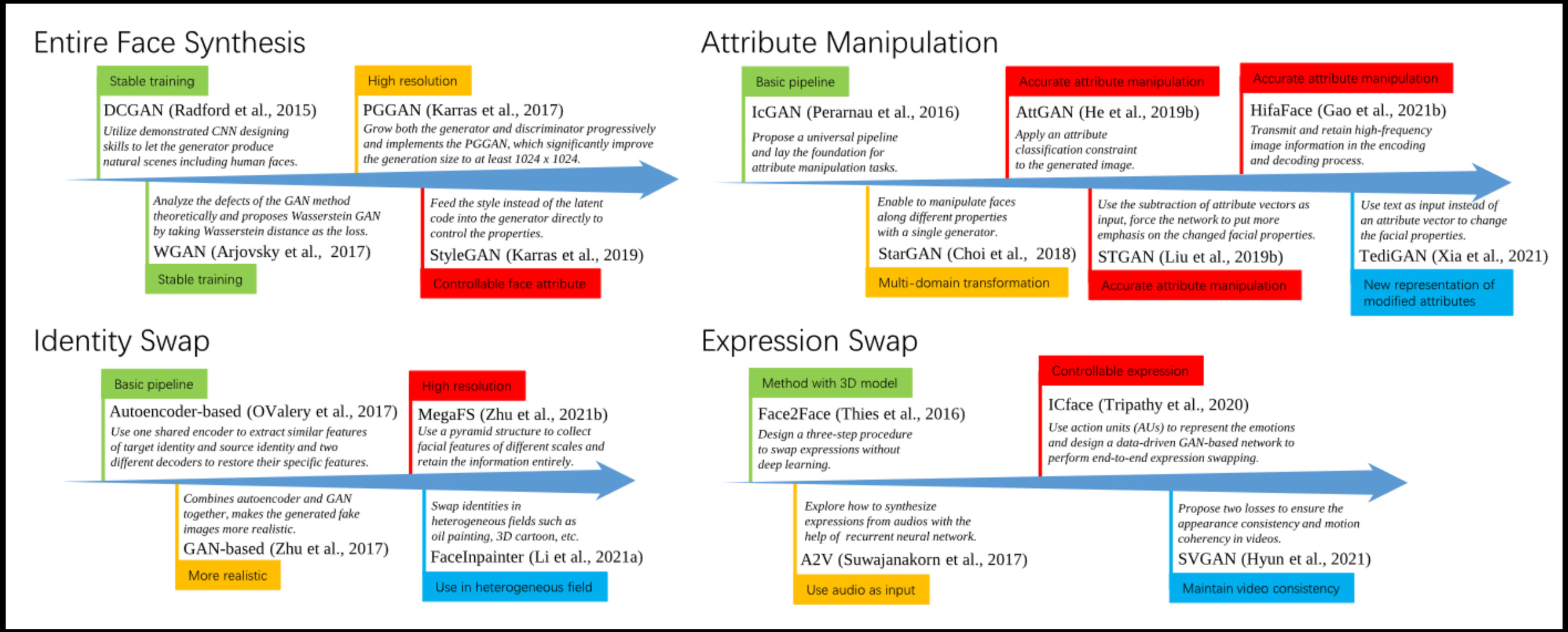
The manipulated latent representation is fed into the generative model, which decodes it to generate a new image or video that reflects the desired modifications. The generative model synthesizes content that resembles the target person while incorporating the desired changes learned during training.

[1]

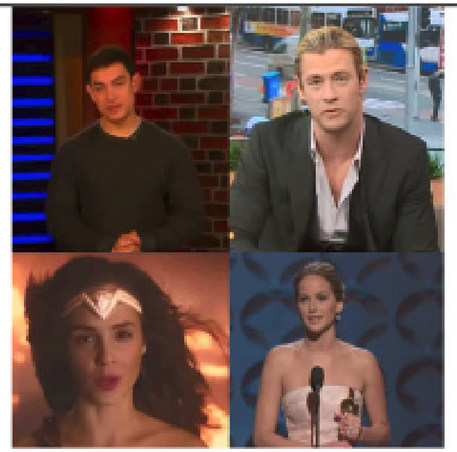
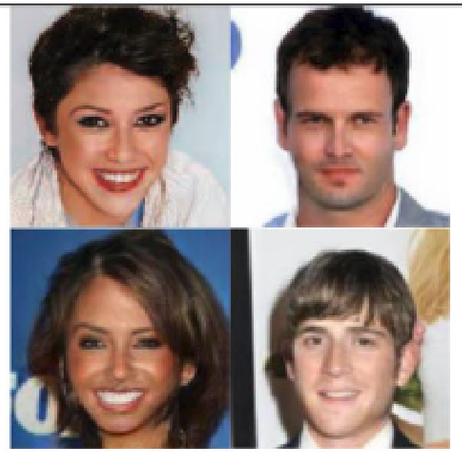
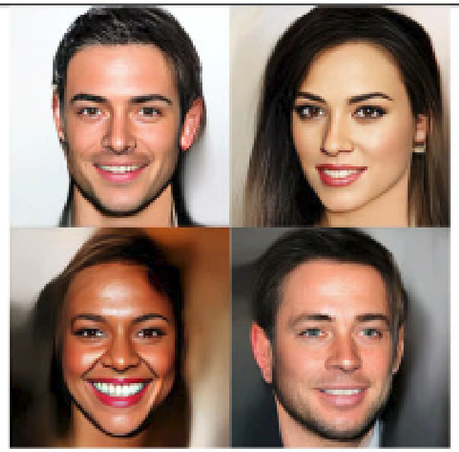

Original Face A
Encoder
Latent representation of face A
Decoder A
Reconstructed Face A
Original Face B
Encoder
Latent representation of face B
Decoder B
Reconstructed Face B
TRAINING
Original Face A
Encoder
Latent representation of face A
Decoder B
Reconstructed Face B from A
GENERATION


Latent z ∈ Z
Normalize
Fully-connected
PixelNorm
Conv 3×3
PixelNorm
4×4
Upsample
Conv 3×3
PixelNorm
Conv 3×3
PixelNorm
8×8
(a) Traditional
Latent z ∈ Z
Normalize
Mapping network f
FC
FC
FC
FC
FC
FC
FC
FC
w ∈ W
Noise
Synthesis network g
Const 4×4×512
B
style
AdaIN
A
Conv 3×3
B
style
AdaIN
A
4×4
Upsample
Conv 3×3
B
style
AdaIN
A
Conv 3×3
B
style
AdaIN
A
8×8
(b) Style-based generator

[3] [1]


Entire Face Synthesis
Stable training
DCGAN (Radford et al., 2015)
Utilize demonstrated CNN designing skills to let the generator produce natural scenes including human faces.
High resolution
PGGAN (Karras et al., 2017)
Grow both the generator and discriminator progressively and implements the PGGAN, which significantly improve the generation size to at least 1024 x 1024.
Analyze the defects of the GAN method theoretically and proposes Wasserstein GAN by taking Wasserstein distance as the loss.
WGAN (Arjovsky et al., 2017)
Stable training
Feed the style instead of the latent code into the generator directly to control the properties.
StyleGAN (Karras et al., 2019)
Controllable face attribute
Attribute Manipulation
Basic pipeline
IcGAN (Perarnau et al., 2016)
Propose a universal pipeline and lay the foundation for attribute manipulation tasks.
Accurate attribute manipulation
AttGAN (He et al., 2019b)
Apply an attribute classification constraint to the generated image.
Accurate attribute manipulation
HifaFace (Gao et al., 2021b)
Transmit and retain high-frequency image information in the encoding and decoding process.
Enable to manipulate faces along different properties with a single generator.
StarGAN (Choi et al., 2018)
Multi-domain transformation
Use the subtraction of attribute vectors as input, force the network to put more emphasis on the changed facial properties.
STGAN (Liu et al., 2019b)
Accurate attribute manipulation
Use text as input instead of an attribute vector to change the facial properties.
TediGAN (Xia et al., 2021)
New representation of modified attributes
Identity Swap
Basic pipeline
Autoencoder-based (OValery et al., 2017)
Use one shared encoder to extract similar features of target identity and source identity and two different decoders to restore their specific features.
High resolution
MegaFS (Zhu et al., 2021b)
Use a pyramid structure to collect facial features of different scales and retain the information entirely.
Combines autoencoder and GAN together, makes the generated fake images more realistic.
GAN-based (Zhu et al., 2017)
More realistic
Swap identities in heterogeneous fields such as oil painting, 3D cartoon, etc.
FaceInpainter (Li et al., 2021a)
Use in heterogeneous field
Expression Swap
Method with 3D model
Face2Face (Thies et al., 2016)
Design a three-step procedure to swap expressions without deep learning.
Controllable expression
ICface (Tripathy et al., 2020)
Use action units (AUs) to represent the emotions and design a data-driven GAN-based network to perform end-to-end expression swapping.
Explore how to synthesize expressions from audios with the help of recurrent neural network.
A2V (Suwajanakorn et al., 2017)
Use audio as input
Propose two losses to ensure the appearance consistency and motion coherency in videos.
SVGAN (Hyun et al., 2021)
Maintain video consistency

[6]

| | | | | |
|---|---|---|---|---|
|  |  |  |  |  |
| FaceForensics++ (Rossler et al., 2019) | PGGAN (Karras et al., 2017) | Celeb-DF (Li et al., 2020e) | StarGAN (Miyato et al., 2018) | StyleGAN (Karras et al., 2019) |
|  |  |  |  |  |
| DFDC (Dolhansky et al., 2020) | GDWCT (Cho et al., 2019) | Glow (Kingma and Dhariwal, 2018) | SC-FEGAN (Jo and Park, 2019) | SAN (Dai et al., 2019) |

[6]

**Table 9: An overview of synthetic facial deepfake generation techniques**

| Reference | Technique | Features | Dataset | Output Quality | Limitations |
|---|---|---|---|---|---|
| Liu et al. [142] | CoGAN | Deep Features | CelebA | 64×64 or 128×128 | ▪ Generate low-quality samples |
| Karras et al. [143] | ProGAN | Deep Features | CelebA | 1024×1024 | ▪ Limited control on the generated output |
| Karras et al. [145] | StyleGAN | Deep Features | ▪ ImageNet | 1024×1024 | ▪ Blob-like artifacts |
| Huang et al. [146] | TP-GAN | Deep Features | ▪ LFW | 256x256 | ▪ Lack fine details<br>▪ Lack semantic consistency |
| Zhang et al. [147] | SAGAN | Deep Features | ▪ ImageNet2012 | 128×128 | ▪ Unwanted visible artifacts |
| Brock et al. [148] | BigGAN | Deep Features | ▪ ImageNet | 512×512 | ▪ Class-conditional image synthesis<br>▪ Class leakage |
| Zhang et al. [149] | StackGAN | Deep Features | ▪ CUB<br>▪ Oxford<br>▪ MS-COCO | 256×256 | ▪ Lack semantic consistency |

# State of the art approaches

Example: face synthesis

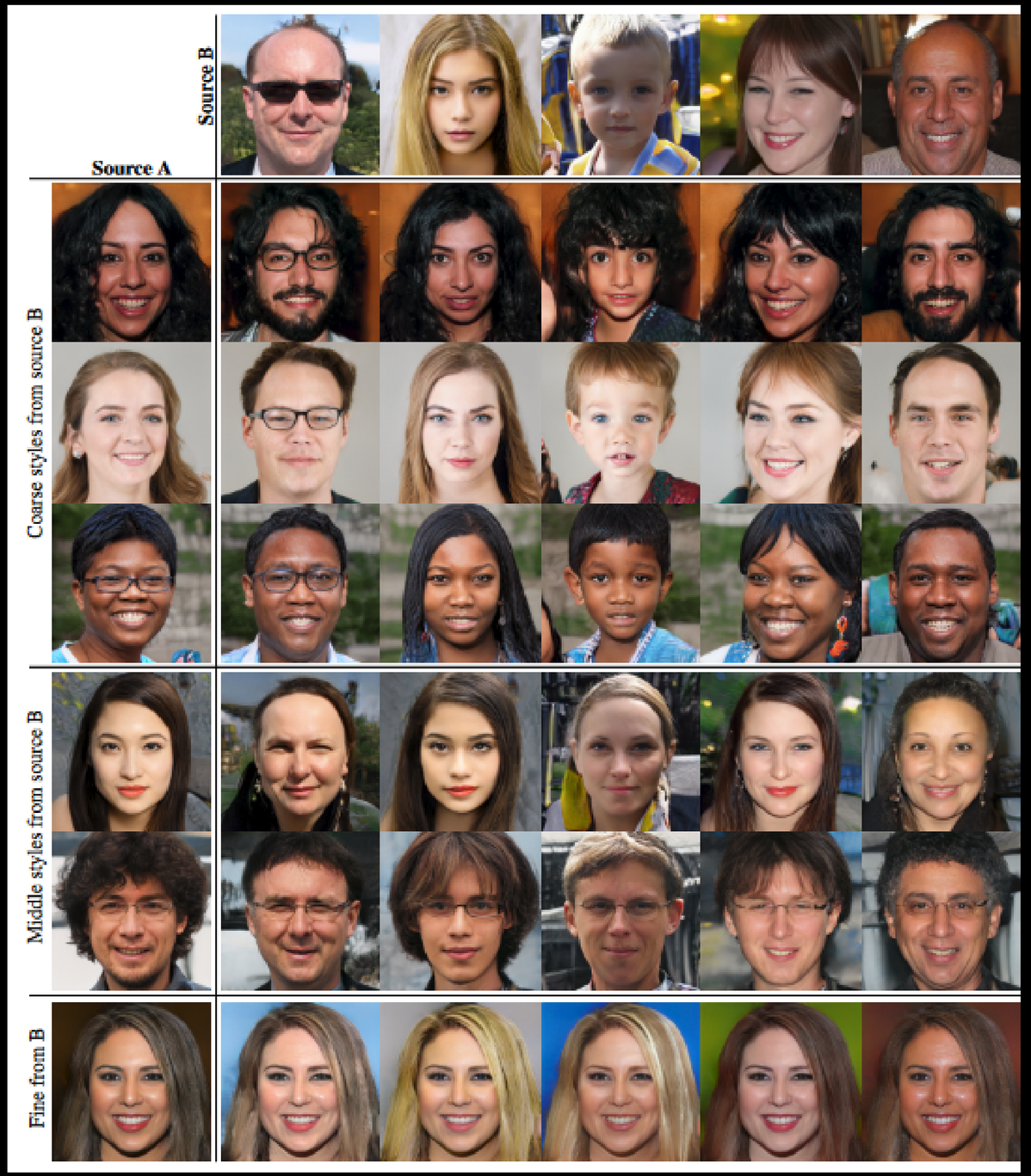
If you want to get to know about more techniques e.g. concerning face swap look into the cited paper

[1]

# StyleGAN

A model for Face synthesis explained in detail

- automatic, unsupervised separation of high-level attributes (e.g. pose, identity) from stochastic variation (e.g., freckles, hair) in the generated images
- enables intuitive scale-specific mixing and interpolation operations



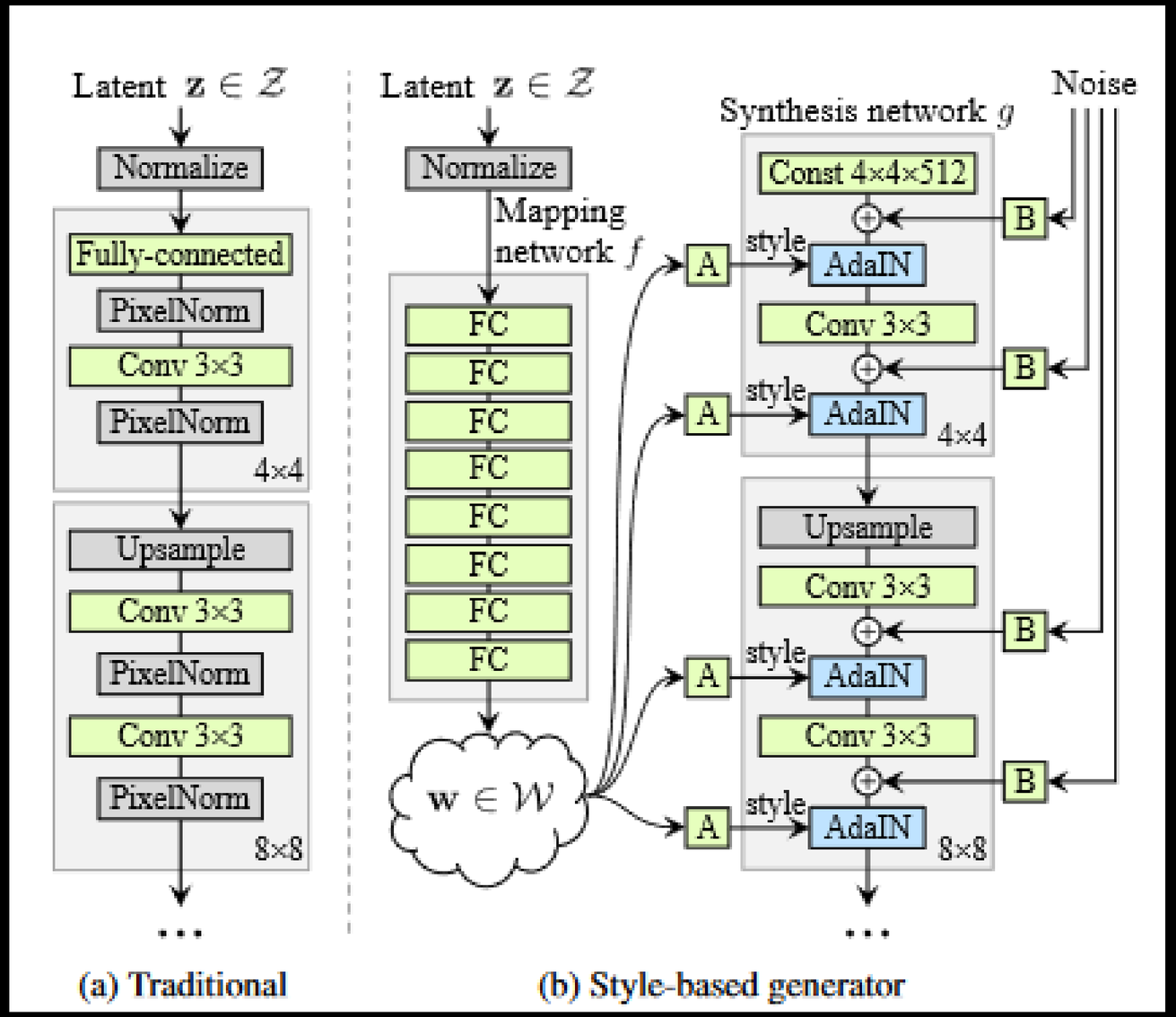

[3]

# Style-based generator

## Latent space

- map the input to an intermediate latent space W
  - adaptive instance normalization (AdaIN) at each convolution layer.
  - Learned affine transformations specialize w to styles
- Gaussian noise is added after each convolution, before evaluating the nonlinearity

"A" = learned affine transform
"B" = learned per-channel scaling factors to the noise input

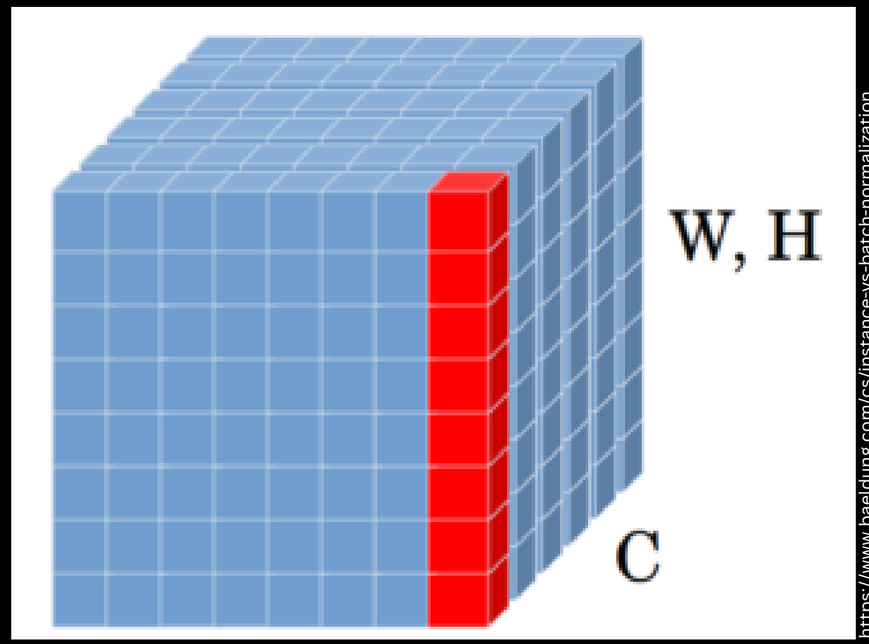



[3]

# Instance Normalization

**Instance Normalization (IN)**

- normalize the activations of each instance (sample) independently
- reducing the style variations across different instances and brings them to a similar distribution




https://www.baeldung.com/cs/instance-vs-batch-normalization

$$IN(X) = \frac{X - \mu(X)}{\sigma(X)}$$

[3]

# Adaptive instance normalization (AdaIN)

- incorporating adaptive adjustment of the style parameters.
- employs learnable affine transformations to adaptively scale and shift the normalized features according to the desired style attributes

Learned affine transformations A specialize w to styles $y = (y_s, y_b)$ that control adaptive instance normalization (AdaIN) operations after each convolution layer of the synthesis network g

$$AdaIN(x_i, y) = y_{s,i} \frac{x_i - \mu(x_i)}{\sigma(x_i)} + y_{b.i}$$

[3]

# Stochastic variation

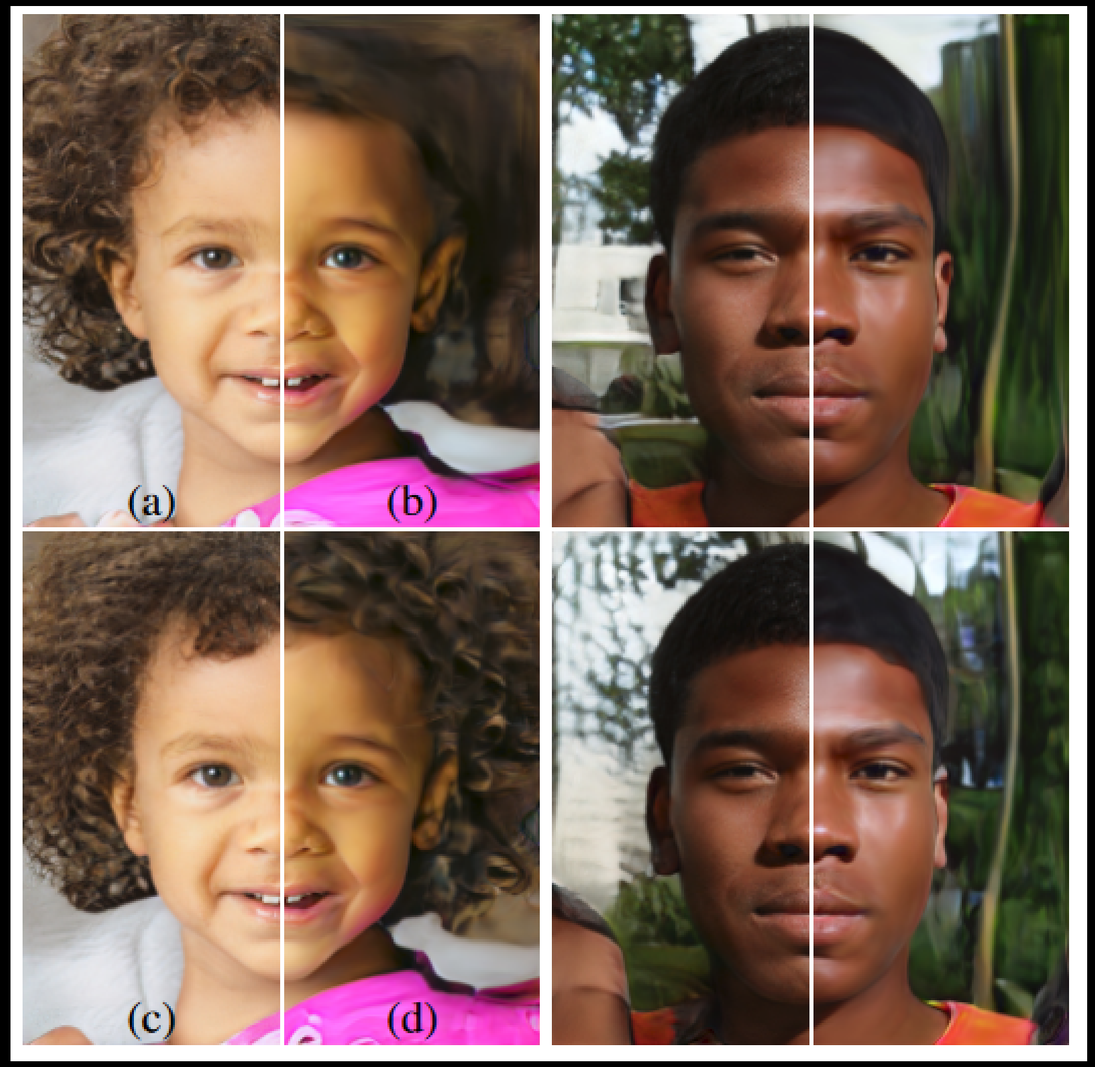
## Traditional Generator

- only input to network is through input layer
- network needs to invent a way to generate spatially varying pseudorandom numbers from earlier activations
- hiding periodicity of generated signal is difficult

## Style-based generator

- adding per-pixel noise after each convolution.
- noise affects only the stochastic aspects, leaving the overall composition and high-level aspects such as identity intact

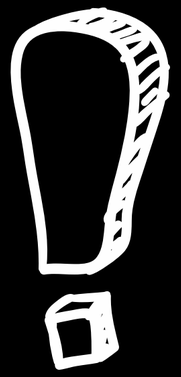
[3]



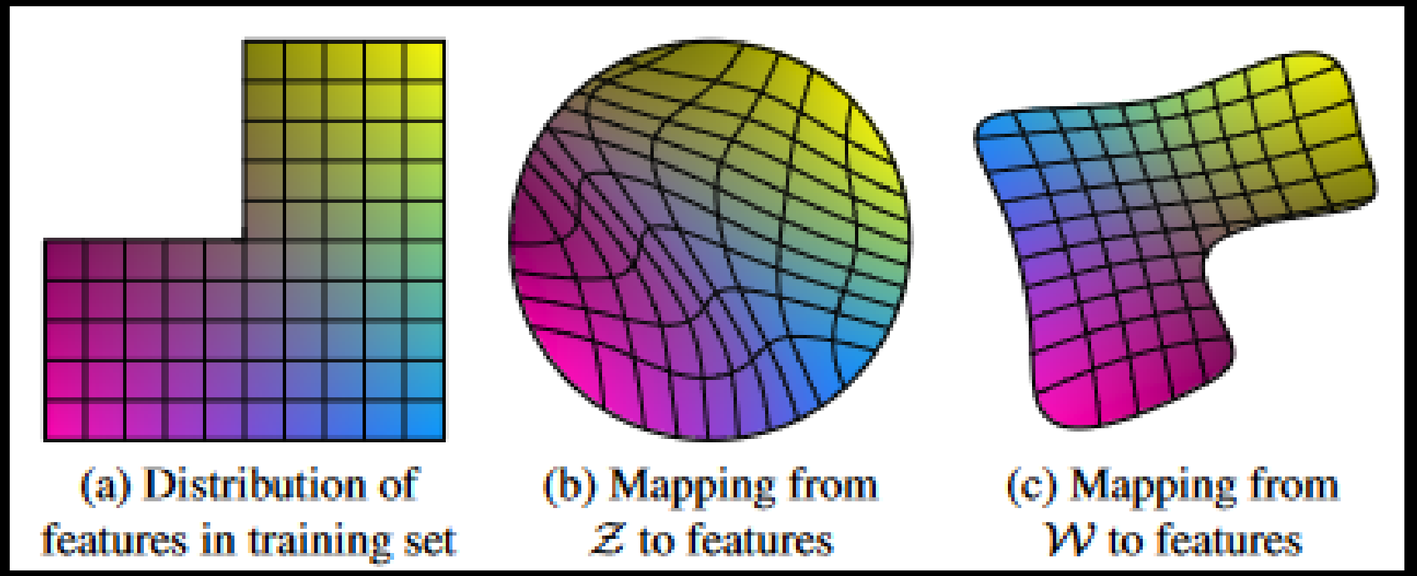

Effect of noise inputs at different layers of our generator
(a) Noise is applied to all layers
(b) No noise
(c) Noise in fine layers only (642 – 10242)
(d) Noise in coarse layers only (42 – 322)

# Disentanglement

**GOAL: latent space consisting of linear subspaces each of which controls one factor of variation**



Sampling probability of each combination of factors in Z needs to match the corresponding density in the training data




example with two factors of variation image features masculinity and hair length

The Sampling density of the intermediate latent space W is induced by the **learned piecewise continuous mapping f(z).** This mapping can be adapted to "unwarp" W so that the factors of variation become more linear.

[3]

[3] [4]

# Quantifying Disentanglement

## Perceptual path length

### Idea

- measure the interpretability and controllability of the latent space
- degree of linearity and predictability in the mapping between points in the latent space and the corresponding generated images.
- A smaller perceptual path length indicates that small changes in the latent code result in perceptually smooth and visually coherent changes in the generated images.

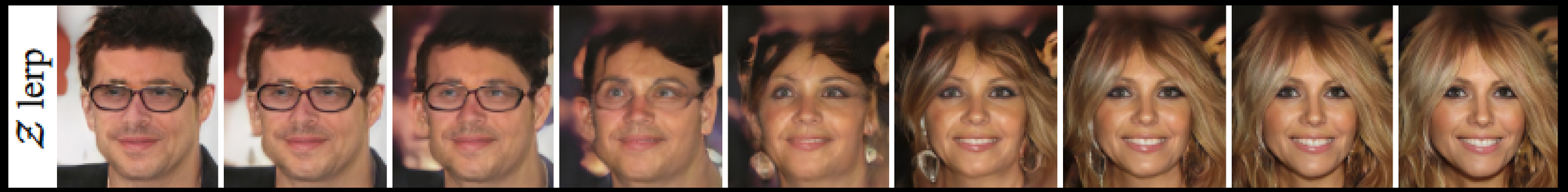
### Compute

- define a reference point in the latent space, usually chosen as the average latent code across a large number of training samples.
- generate pairs of images by linearly interpolating between the reference point and other random latent codes
- The perceptual path length is calculated by measuring the average change in the perceptual space across these image pairs.

$$l_z = \mathbb{E}[\frac{1}{e^2}\ d(G(slerp(z_1, z_2; t)), G(slerp(z_1, z_2; t + \epsilon)))]$$

$$l_w = \mathbb{E}[\frac{1}{e^2}\ d(g(lerp(f(z_1), f(z_2); t)), g(lerp(f(z_1), f(z_2); t + \epsilon)))]$$

# Example




interpolation of latent-space vectors may yield surprisingly non-linear changes in the image. For example, features that are absent in either endpoint may appear in the middle of a linear interpolation path. This is a sign that the latent space is entangled and the factors of variation are not properly separated

[3][4]

# Quantifying Disentanglement

## Linear Separability

### Idea

- if latent space is disentangled find direction vectors that consistently correspond to individual factors of variation
- measuring how well the latent-space points can be separated into two distinct sets via a linear hyperplane, so that each set corresponds to a specific binary attribute of the image
- train auxiliary classification networks for a number of binary attributes

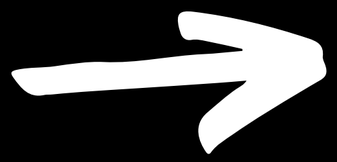
### Compute

- To measure the separability of one attribute, we generate 200,000 images with z  P(z)
- and classify them using the auxiliary classification network.
- sort samples : only retain best half of classifier confidence outcomes (labeled latent-space vectors)
- For each attribute, we fit a linear SVM to predict the label based on the latent-space point—z for traditional and w for style-based—and classify the points by this plane

[3]

# Linear Separability

$$conditional\ entropy\ H(Y|X)$$



This tells how much additional information is required to determine the true class of a sample

X are classes predicted by SVM
Y are classes determined by pre-trained classifier

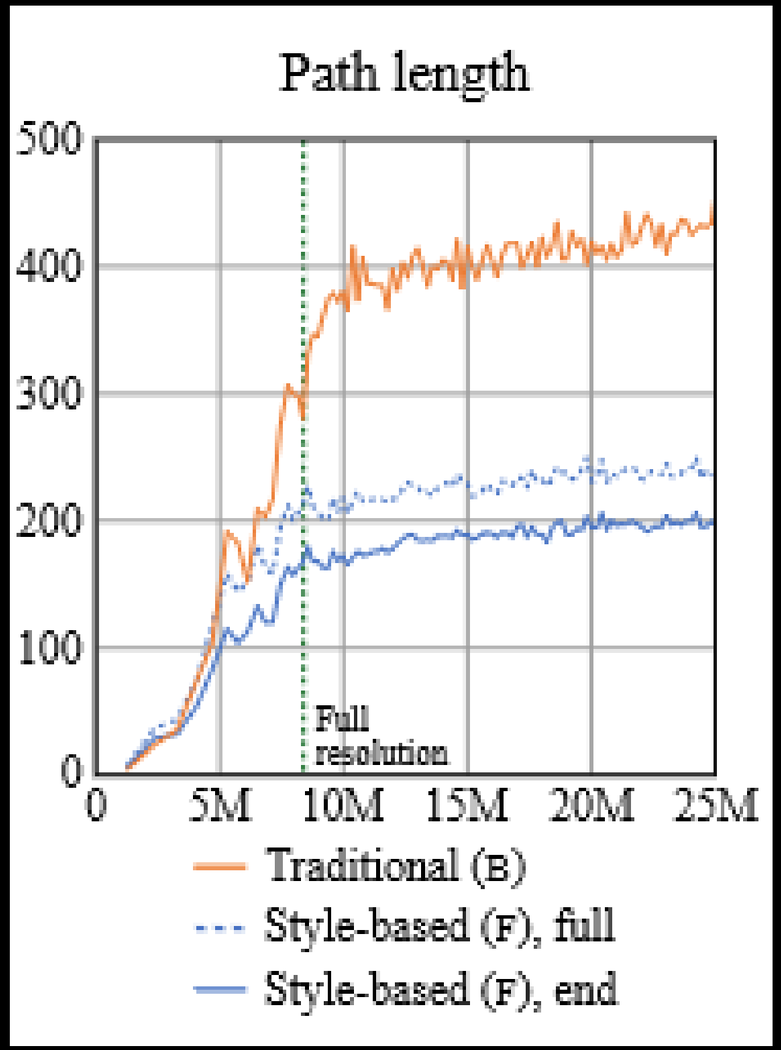
A low value suggests consistent latent space directions for the corresponding factor(s) of variation
exponentiation brings the values from logarithmic to linear domain so that they are easier to compare.

$$exp(\sum_{i} H(Y_i/X_i))$$

[3]

# Results

| | Method | | Path length full | Path length end | Separability |
|---|---|---|---|---|---|
| B | Traditional generator | $\mathcal{Z}$ | 412.0 | 415.3 | 10.78 |
| D | Style-based generator | $\mathcal{W}$ | 446.2 | 376.6 | 3.61 |
| E | + Add noise inputs | $\mathcal{W}$ | **200.5** | **160.6** | 3.54 |
| | + Mixing 50% | $\mathcal{W}$ | 231.5 | 182.1 | **3.51** |
| F | + Mixing 90% | $\mathcal{W}$ | 234.0 | 195.9 | 3.79 |




[3]

# Result pictures of StyleGAN


[3]

# Challenges faced in creating deep fakes

- Generalization
  - lack of sufficient training data
- Identity leakage
  - preservation of target identity can be a problem concerning specific methods, e.g. reenactment
- Pose variations and distance from camera
  - quality of deep fakes decreases when frontal view is not granted
- Controlled environment
  - application with data from the "wild" is not feasible since current methods highly depend on given lighting settings etc.
- Occlusion
  - still the current algorithm have issues coping with occlusion which leads to inconsistent facial features
- Lack of realism in synthetic audio
  - main challenges of audio-based DeepFake are the lack of natural emotions, pauses, breathiness, and the pace at which the target speaks.

[1]

# How to detect deep fakes?

# Can you detect the deepfake?

Click on the person who is real.

**Which Face Is Real?**

Your new friend on Facebook. Your next Tinder match. A potential employee. Sure you've seen their picture — but do they even exist? Learn how to tell.

callin_bull

# ... How can we detect DeepFakes?


https://www.whichfaceisreal.com/learn.html


https://www.whichfaceisreal.com/learn.html


https://www.whichfaceisreal.com/learn.html


https://www.whichfaceisreal.com/learn.html


https://www.whichfaceisreal.com/learn.html


https://www.whichfaceisreal.com/learn.html

## Water sploches

some GANS are producing water sploches between background and hair which is a clear indicator

## Background problems

it can happen that the background is generated weirdly, as well as other people in the background

## Asymmetries

Are the glasses symmetric? How about facial hair, facial expressions or earrings?

## Teeth & Hair

Teeth & Hair are not easy to render, they might me assymetric, or weird looking

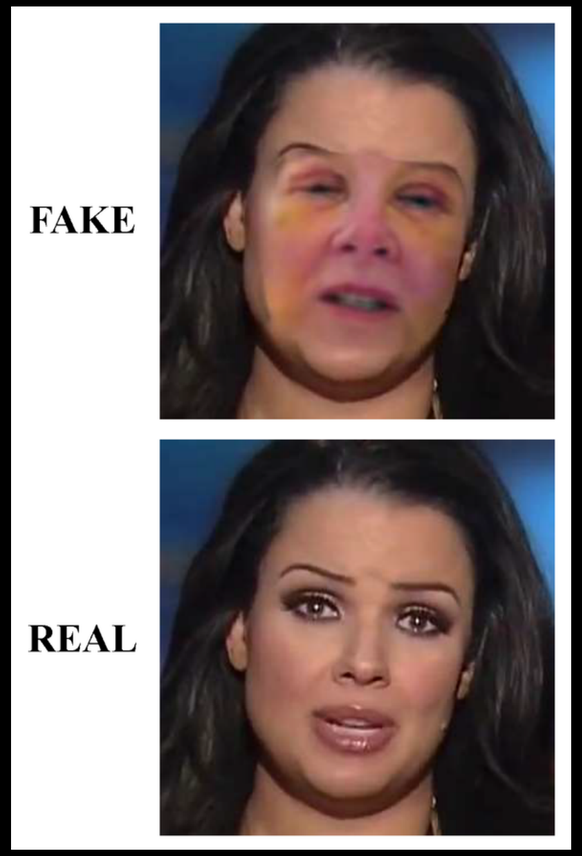
# How can algorithms detect DeepFakes?

# Spatial-based detection

- Observe visible and invisible artifacts in the spatial domain
- Mostly on pixel-level
- Currently most popular technique for DeepFake detection

Criticism:

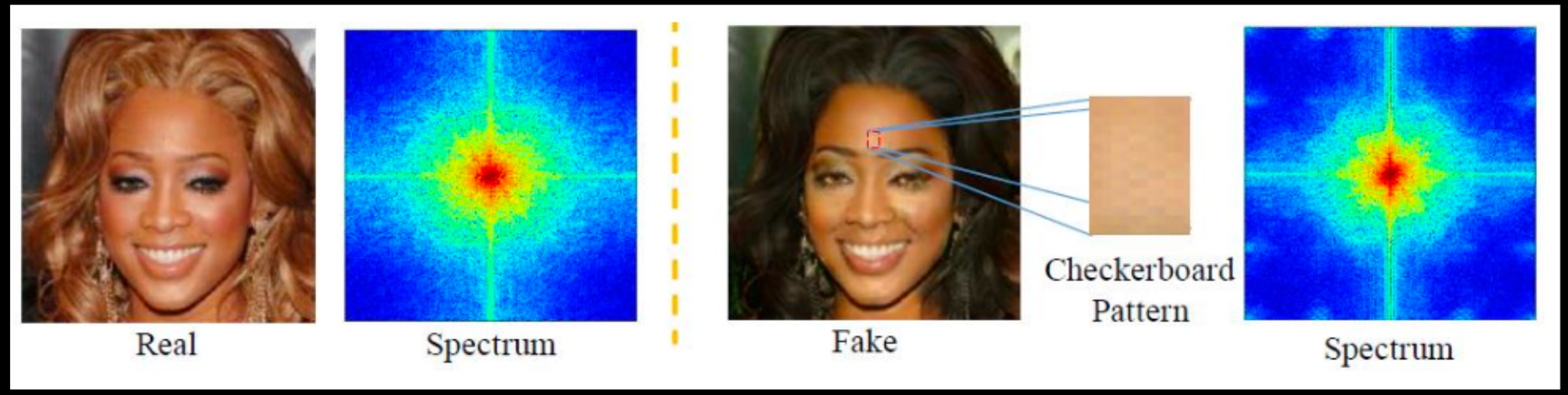
- Poor generalization against unknown generation techniques
- Low robustness  to adversarial noise attacks and permutation attacks




[6]

# Frequency-based detection

- Exploits...
  - artifacts on the frequency level introduced by GANs (GAN fingerprint)
  - difference between frequency domain features of real and fake faces
- Criticism:
  - Fingerprint can be destroyed by simple permutation attacks (blur or JPEG compression)

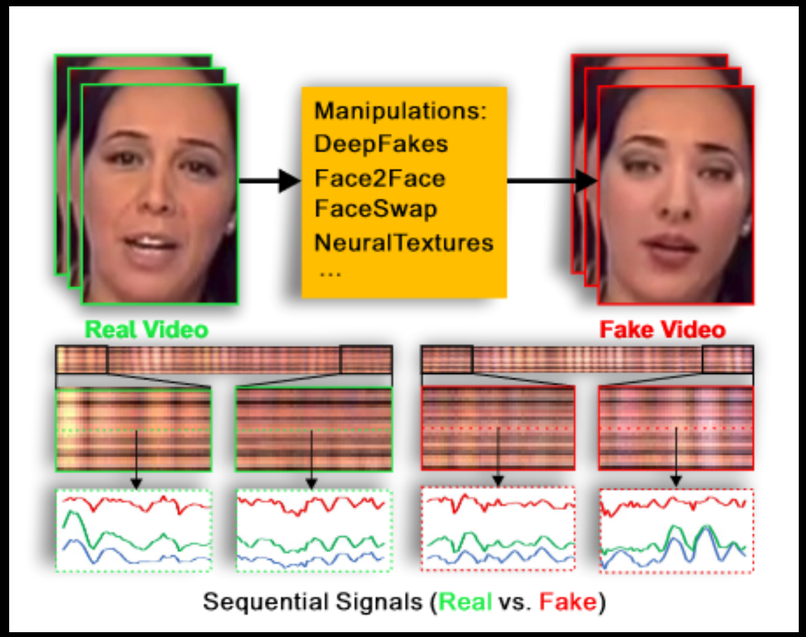



[6]

# Biological signal-based detection

- Visual inconsistency
  - Non-natural properties of synthesized faces (shape, facial features, ...)
  - Repetitive actions tend to disappear in DeepFakes (e.g. eye blinking)
  - Subtle biological characteristics tend to be corrupted in DeepFakes
    - E.g. heart rate (difficult to extract from a video)

- Criticism:
  - A lot of biological signals could be improved in future GANs
  - Many of them work only for videos



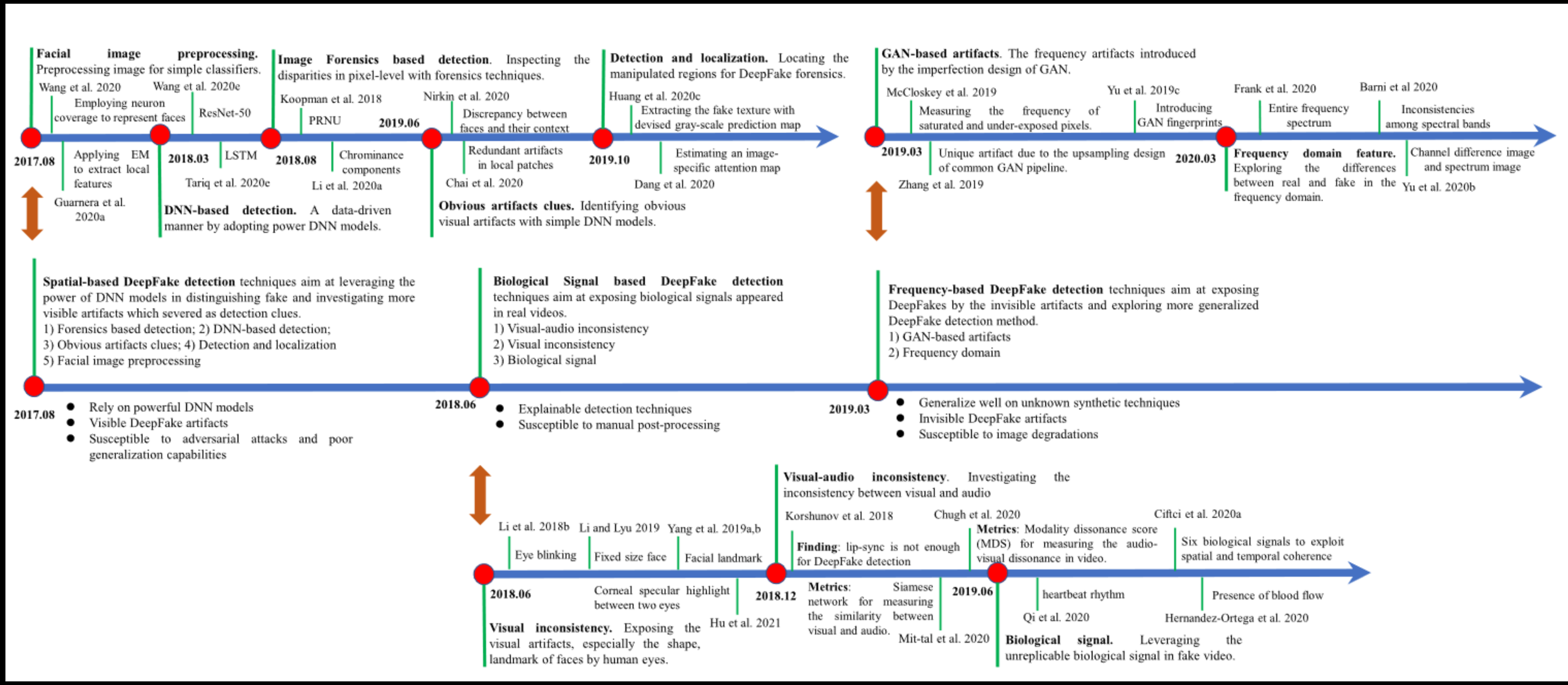

[7] [6]

# DeepFake detection of algorithms - different approaches

| Detection technique | Description | Benefits | Limitations |
|---|---|---|---|
| Detection based on visual artifacts. | Closed observation of deepfakes may reveal minor differences and inconsistencies between background and foreground. | Visible artifacts methods may detect inconsistency in the blending boundary of a modified object, such as image resolution. | These visible artifacts are rapidly diminishing as deepfake algorithms advance, demanding to exploit more intrinsic properties. |
| Detection based on GAN fingerprints. | Synthesized faces generated by GANs often have GAN generated fingerprints. | By the use of deep features, fingerprints identified in the GAN generated fake images. | Because of different versions of GANS, no universal fingerprint metric can be adopted. |
| Detection based on biological signals, such as eye blinking. | There can be abnormalities in the eye blinking frequency in deepfakes. | Synthetic biological signals are easier to detect, such as eye blinking and heartbeat. | Advanced versions of deepfakes face generation very precisely model the biological signals, making it harder to detect. |
| Detection based on adjacent video frames' continuity. | Deepfakes may have flickering, jittering, and different face positions due to the discontinuity among adjacent video frames. | Temporal consistency-based detection methods can recognize discontinuity in adjacent video frames. | Poor performance on low-quality videos as continuity between adjacent frames is affected by video compression. |
| Detection based on face emotions. | Alignment of facial emotions is improper on swapped faces in deepfakes. | Siamese network-based architecture can detect non-alignment of facial emotions by facial and audio features extraction. | This technique fails if the video has no emotions. |
| Detection based on out of lip-synced videos. | A deepfake video with synthesized audio may have out-of-sync lips. | The difference between visemes (mouth shapes) and spoken phenomes (utterances) is used for out-of-sync lips detection. | Improved GANs can now generate proper lip-synced deepfakes. Also, any out-of-sync video does not need to be deepfake. |
| Multimodal detection technique. | Deepfakes created by swapping the face and audio, same as Type IV in Figure 1, are known as multimodal deepfakes. | Multimodal detection techniques first detect swapped faces, then use lip-syncing techniques to identify manipulated speech. | Accuracy suffers as detection techniques extract audio from different datasets instead of the same deepfake video. |

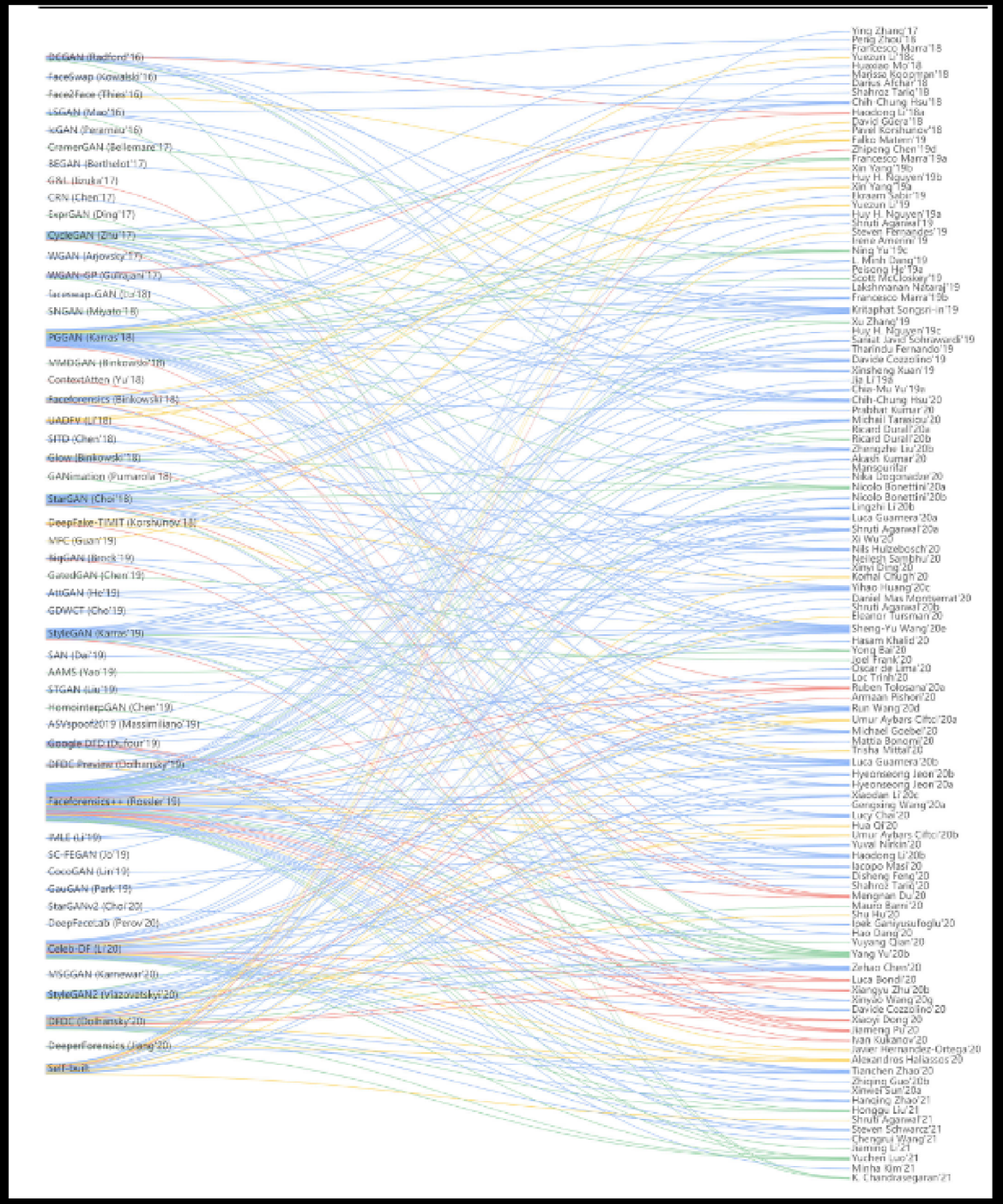
[5]

https://ieeexplore.ieee.org/document/10077834


**Facial image preprocessing.** Preprocessing image for simple classifiers.
Wang et al. 2020
Employing neuron coverage to represent faces
Wang et al. 2020e
ResNet-50
2017.08
Applying EM to extract local features
2018.03
LSTM
Tariq et al. 2020c
Guarnera et al. 2020a
**DNN-based detection.** A data-driven manner by adopting power DNN models.
**Image Forensics based detection**. Inspecting the disparities in pixel-level with forensics techniques.
Koopman et al. 2018
PRNU
2018.08
Chrominance components
Li et al. 2020a
2019.06
Nirkin et al. 2020
Discrepancy between faces and their context
Redundant artifacts in local patches
Chai et al. 2020
**Obvious artifacts clues.** Identifying obvious visual artifacts with simple DNN models.
**Detection and localization.** Locating the manipulated regions for DeepFake forensics.
Huang et al. 2020c
Extracting the fake texture with devised gray-scale prediction map
2019.10
Estimating an image-specific attention map
Dang et al. 2020
**GAN-based artifacts**. The frequency artifacts introduced by the imperfection design of GAN.
McCloskey et al. 2019
Measuring the frequency of saturated and under-exposed pixels.
2019.03
Unique artifact due to the upsampling design of common GAN pipeline.
Zhang et al. 2019
Yu et al. 2019c
Introducing GAN fingerprints
2020.03
Frank et al. 2020
Entire frequency spectrum
Barni et al 2020
Inconsistencies among spectral bands
**Frequency domain feature.** Exploring the differences between real and fake in the frequency domain.
Channel difference image and spectrum image
Yu et al. 2020b
**Spatial-based DeepFake detection** techniques aim at leveraging the power of DNN models in distinguishing fake and investigating more visible artifacts which severed as detection clues.
1) Forensics based detection; 2) DNN-based detection;
3) Obvious artifacts clues; 4) Detection and localization
5) Facial image preprocessing
2017.08
Rely on powerful DNN models
Visible DeepFake artifacts
Susceptible to adversarial attacks and poor generalization capabilities
**Biological Signal based DeepFake detection** techniques aim at exposing biological signals appeared in real videos.
1) Visual-audio inconsistency
2) Visual inconsistency
3) Biological signal
2018.06
Explainable detection techniques
Susceptible to manual post-processing
**Frequency-based DeepFake detection** techniques aim at exposing DeepFakes by the invisible artifacts and exploring more generalized DeepFake detection method.
1) GAN-based artifacts
2) Frequency domain
2019.03
Generalize well on unknown synthetic techniques
Invisible DeepFake artifacts
Susceptible to image degradations
Li et al. 2018b
Eye blinking
Li and Lyu 2019
Fixed size face
Yang et al. 2019a,b
Facial landmark
2018.06
Corneal specular highlight between two eyes
Hu et al. 2021
**Visual inconsistency.** Exposing the visual artifacts, especially the shape, landmark of faces by human eyes.
**Visual-audio inconsistency**. Investigating the inconsistency between visual and audio
Korshunov et al. 2018
**Finding**: lip-sync is not enough for DeepFake detection
2018.12
Chugh et al. 2020
**Metrics**: Modality dissonance score (MDS) for measuring the audio-visual dissonance in video.
**Metrics**: Siamese network for measuring the similarity between visual and audio.
Mit-tal et al. 2020
2019.06
Ciftci et al. 2020a
Six biological signals to exploit spatial and temporal coherence
heartbeat rhythm
Qi et al. 2020
Presence of blood flow
Hernandez-Ortega et al. 2020
**Biological signal.** Leveraging the unreplicable biological signal in fake video.

[6]

# 'Battleground' diagram between DeepFake detection and generation




[6]

# Limitations and challenges of DeepFake detection

- **Lack of explainability**
  - black box problem becomes relevant in terms of  court proceedings
- **DeepFake Detection Evasion**
  - adversarial perturbation attack
  - elimination of manipulation traces in the frequency domain
  - employing image filtering to mislead detectors

- **Performance evaluation**
  - replace binary classification by multilevel classification
- **Generalization**
  - wide variety of potential manipulations : difficult to build detection models that can effectively detect all types of deep fakes
- **Time sensitivity**
  - social media platforms often require real-time or near real-time detection : need for fast and efficient detection algorithms

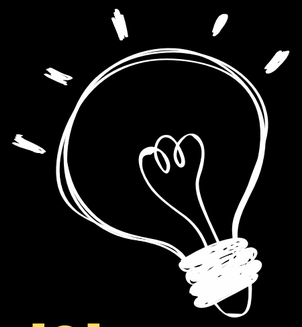
[1]

**Will there ever be a DeepFake detection method that can recognize every type of DeepFake? In which way will the generation techniques develop?**

# Applications and Implications

- Entertainment and Creative Arts
- Digital avatars and virtual characters
- Cross cultural video distribution
- Create personalized synthetic voice

- Political manipulation
- Revenge Porn and Non-consensual Content
- Misinformation and fake news
- De-age actors to a comparative level of costly CGI effects
- Tourist attractions



It is important to recognize and address the ethical, legal, and societal implications of deep fakes. Stricter regulations, awareness campaigns, and the development of robust detection methods are being pursued to mitigate the potential negative consequences associated with deep fakes while promoting their responsible and ethical use.

[9]

# Future directions

Regulation and Policy

- criminalizing malicious use
- guidelines for responsible use
- protecting individual privacy and rights

Example: California

DeepFakes in healthcare and therapy?

- virtual patient avatars for medical training
- combining with AR/ VR
- virtual therapists
- improving realism
- personalized avatars for online shopping?

→ What do you think is approaching us?

# Conclusion

- DeepFakes are closely connected to social phenomena
- There is a potential misuse of DeepFakes so that a need emerges to address the ethical, legal, and societal implications of DeepFakes
- DeepFake generations are mainly based on GAN or Autoencoder structures
- DeepFake generation and detection are closely connected to each other
- There are still a lot of open challenges for generation and detection methods
- It remains open whether there will be once a 'perfect' generation or detection technique

# Sources


[1] Masood, M., Nawaz, M., Malik, K., Javed, A., Irtaza, A. (2021). "DeepFakes Generation and Detection: State-of-the-art, open challenges, countermeasures, and way forward". doi: https://doi.org/10.48550/arXiv.2103.00484

[2]Negi, S., Jayachandran, M., Upadhyay, S. (2021). "DeepFake : An Understanding of Fake Images and Videos". doi: https://doi.org/10.32628/CSEIT217334

[3] Karras, T., Laine, S., Aila , T. (2018). "A Style-Based Generator Architecture for Generative Adversarial Networks". doi: https://doi.org/10.48550/arXiv.1812.04948

[4] Laine, S. (2018). "FEATURE-BASED METRICS FOR EXPLORING THE LATENT SPACE OF GENERATIVE MODELS". https://openreview.net/pdf?id=BJslDBkwG

[5] Salman, S., Shamsi, J., Qureshi, R. (2023). "DeepFake Generation and Detection: Issues, Challenges, and Solutions". doi: 10.1109/MITP.2022.3230353

[6] Juefei-Xu, F., Wang, R., Huang, Y., Guo4, Q., Ma, L., Liu, Y. (2022)."Countering Malicious DeepFakes: Survey, Battleground, and Horizon" doi: https://doi.org/10.48550/arXiv.2103.00218

[7] Qi, H., Guo, Q., Juefei-Xu, F., Xie, X., Ma, Lei., Feng, W. (2020).DeepRhythm: Exposing DeepFakes with Attentional Visual Heartbeat Rhythms. doi: https://doi.org/10.48550/arXiv.2006.07634

[8]Tolosana, R., Vera-Rodriguez, R., Fierrez, J., Morales, A. (2020). DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection. doi: https://doi.org/10.48550/arXiv.2001.00179

[9] Whittaker, L., Kietzmann, T., Kietzmann, J., Dabirian, A.(2020) “All Around Me Are Synthetic Faces”: The Mad World of AI-Generated Media". doi: 10.1109/MITP.2020.2985492

# Links for demonstration

Deciding which face is real

- https://www.whichfaceisreal.com/index.php

This person does not exist

- https://this-person-does-not-exist.com/en

Can you spot the audio deep fake?

- https://deepfake-demo.aisec.fraunhofer.de/?

# Thank you for your attention!

Questions?