Nicola Dall'Asen
University of Pisa
Multimedia and Human Understanding Group - University of Trento

Graph-based Generative Face Anonymisation with Pose Preservation

A journey between laws and deep learning (and extras)

Overview

Talk overview

- **Chapter 1:** Why do we need privacy-preserving ML?
- **Chapter 2:** Brief recap of Generative Models
- **Chapter 3:** Graph-based Generative Face Anonymisation with Pose Preservation
- **Chapter 4:** Current state-of-the-art and future work

Talk overview

- **Chapter 1:** Why do we need privacy-preserving ML?
- **Chapter 2:** Brief recap of Generative Models
- **Chapter 3:** Graph-based Generative Face Anonymisation with Pose Preservation
- **Chapter 4:** Current state-of-the-art and future work

Legal Landscape of AI

GDPR

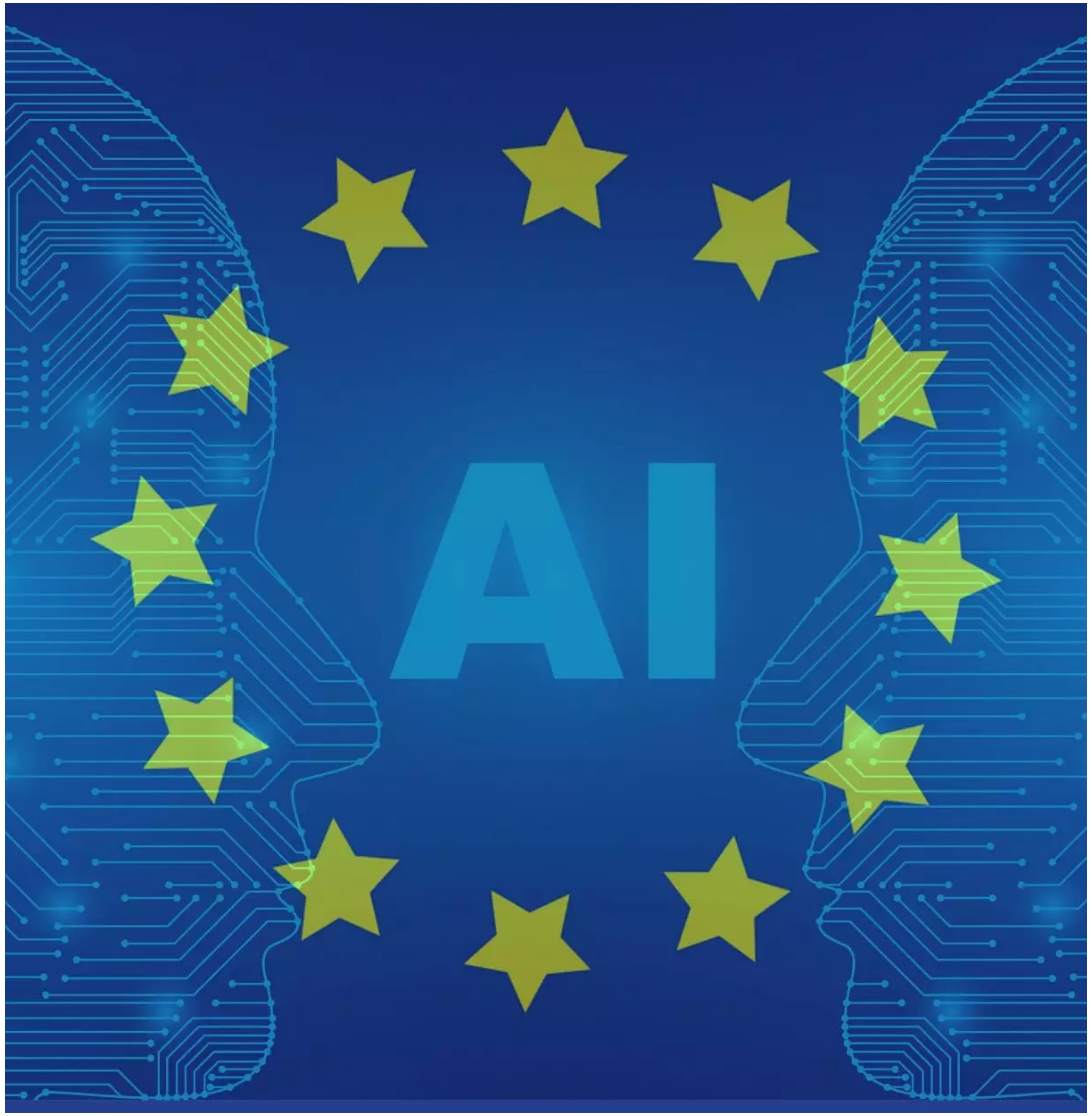
General Data Protection Regulation

- Regulates privacy and security of personal data on the European territory
- Provides mandatory rules for organisations on processing personal data
- Collection of data must have a defined purpose, they cannot be collected freely

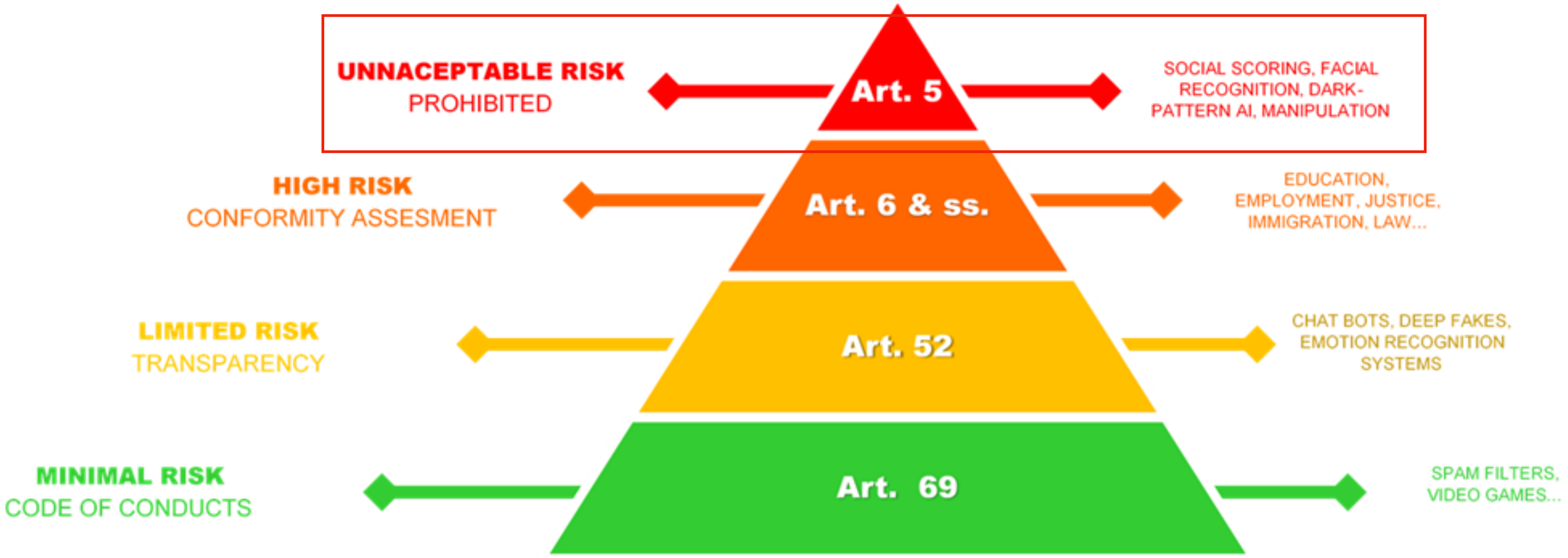
AI Act

Proposal for a regulation on Artificial Intelligence

- It tackles the risks of the abuse of AI systems, to foster ethic innovation
- It will cover any AI system
  - Machine Learning applications
  - Logic and knowledge-based approaches
  - Statistical methods



A risk-based approach

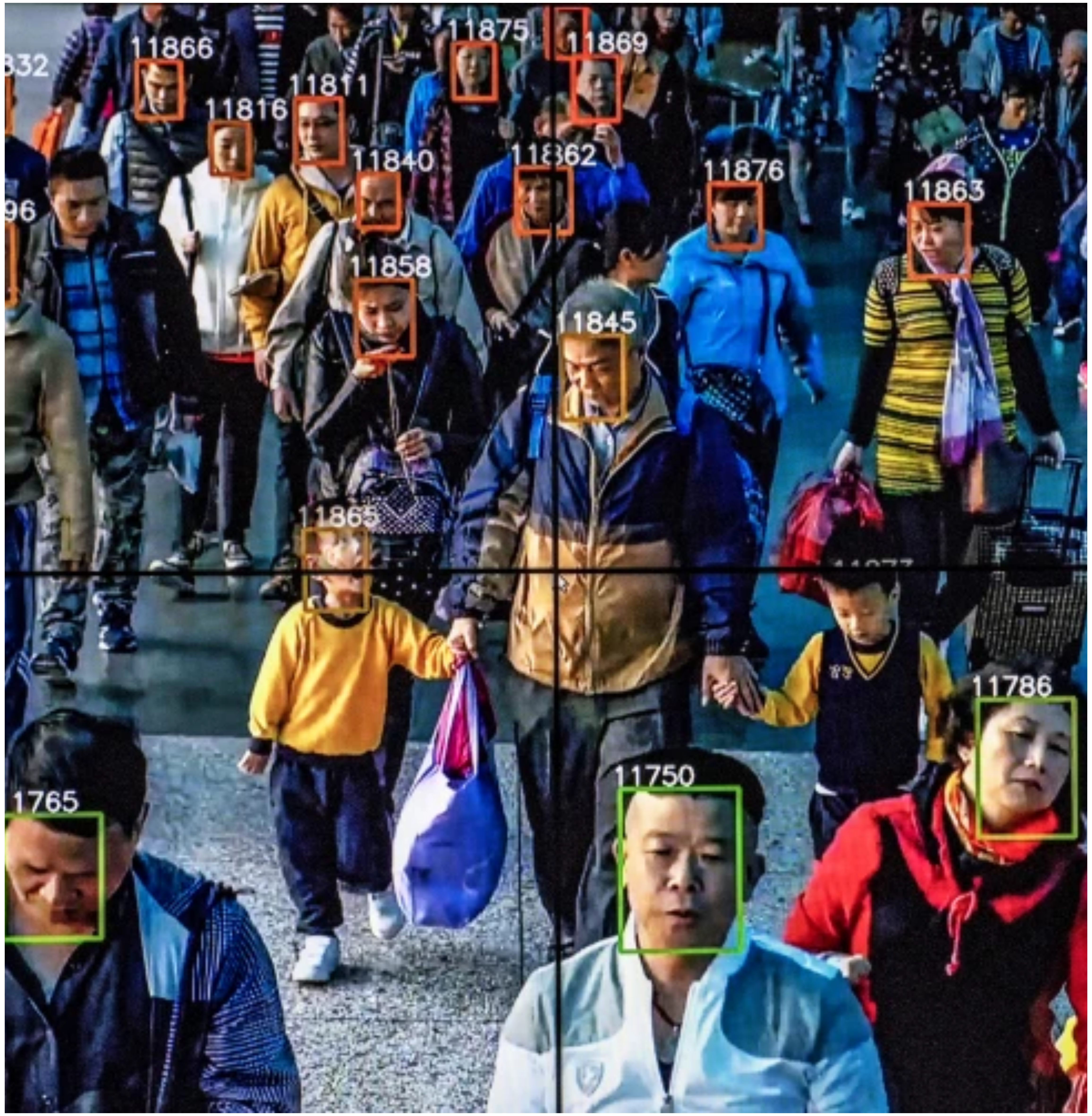


- Four classes of risk based on the data collected and their usage
- The most interesting is the ban on the *unacceptable* level:
  - Social scoring
  - Real-time biometric identification, e.g. facial recognition for identification in public spaces

Facial Recognition Debate

Can biometric data be used for law enforcement?

- GDPR grants some exceptions, for example, public interest
- Live identification is forbidden, except for three cases:
    - Specific investigations, e.g. missing children
    - Prevention of terrorist attacks
    - Identification of suspects of specific crimes, e.g. arson, human trafficking, […]



MARVEL & PROTECTOR

MARVEL

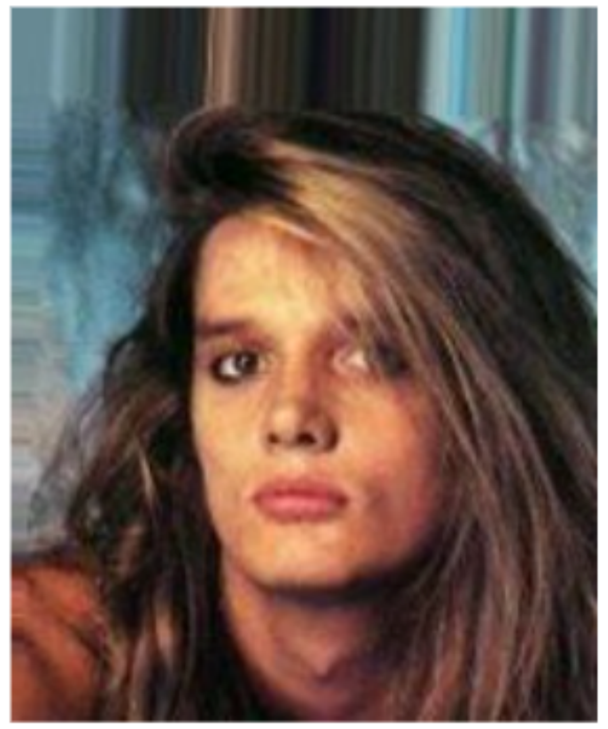
PROTECTOR

PROTECTing places of wORship

The problem to solve

Anonymisation before further processing

- Privacy by removing Personal Identifiable Information (PII), the output is anonymised and the use is lawful
- Retain non-personal information (pose and expression) used in emotion recognition and anomaly detection
- Perform this in a non-degradative way, classic techniques fail, no downstream task is possible
- Our approach performs landmark-based face swapping



Original

Blurring

Pixelation

CIAGAN

AnonyGAN (ours)

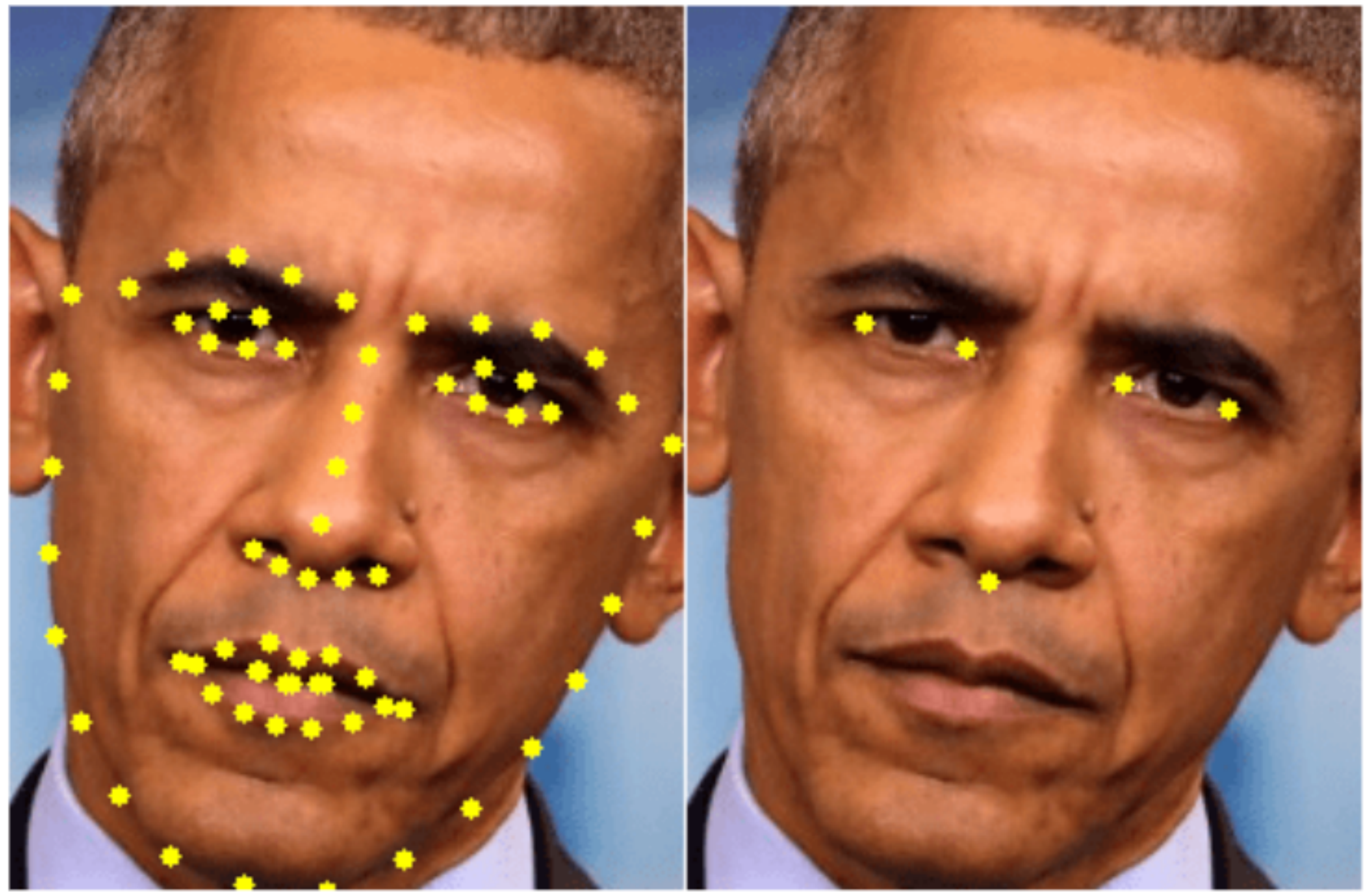
Terminology

Landmarks

- Facial landmarks represent salient regions of the face: eyes, nose, mouth, […]
- Different detectors use different numbers of landmarks
    - dlib extracts 68 points

Face swapping

- Transfer a face from a condition image to a source
- Preserve pose and expression



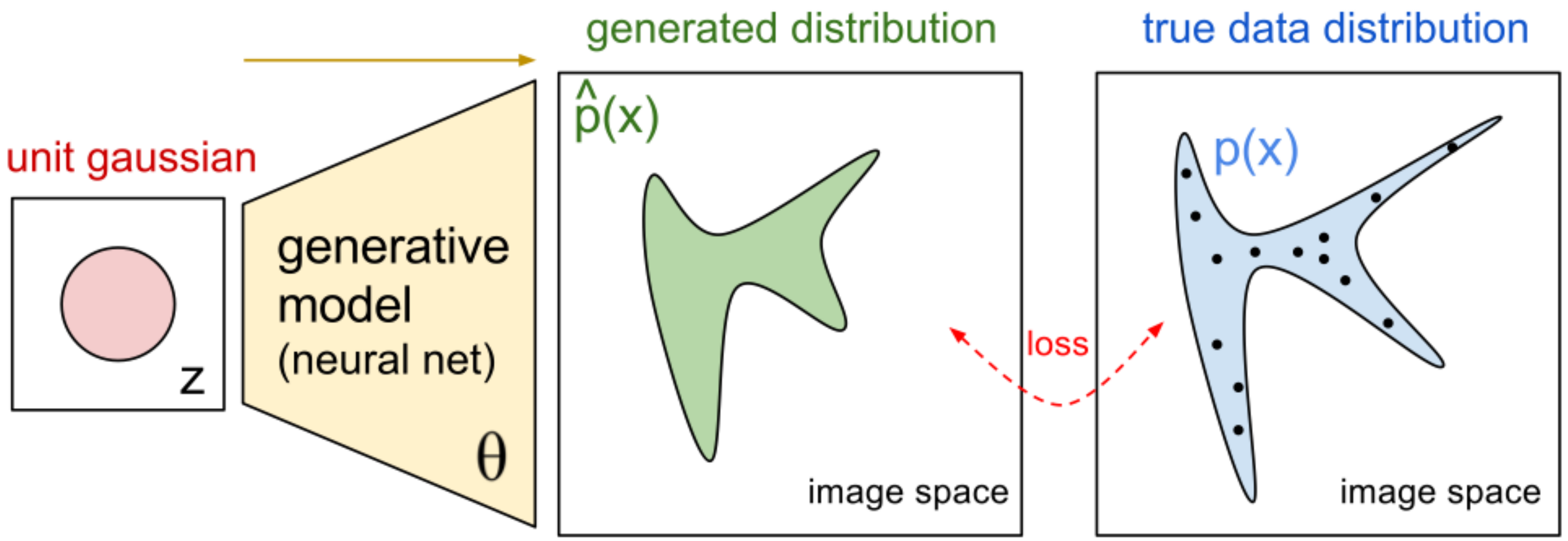
68 and 5 landmark representation for faces

FaceApp gender swap

Talk overview

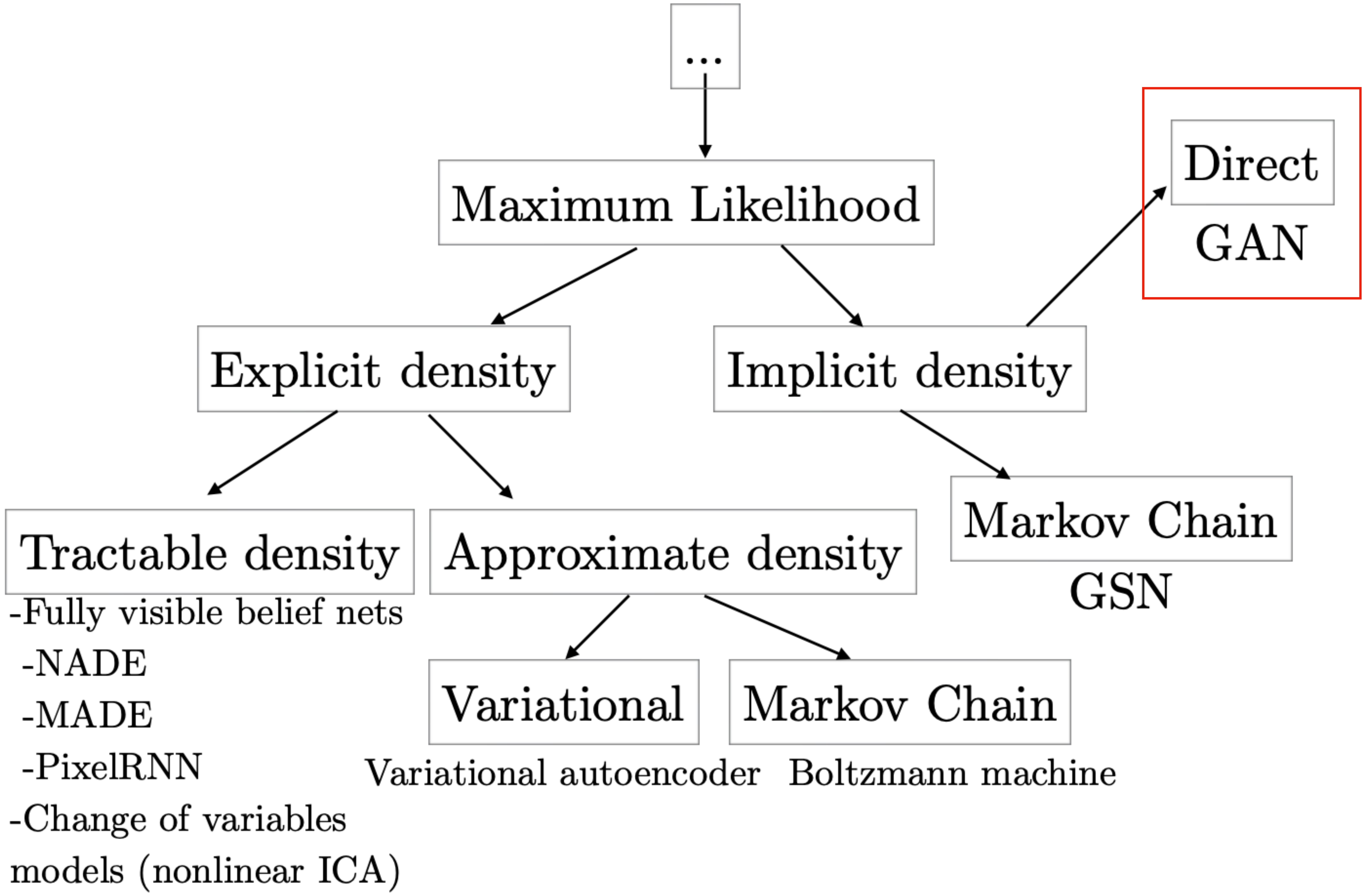
- **Chapter 1:** Why do we need privacy-preserving ML?
- **Chapter 2:** Brief recap of Generative Models
- **Chapter 3:** Graph-based Generative Face Anonymisation with Pose Preservation
- **Chapter 4:** Current state-of-the-art and future work

Deep Generative Models



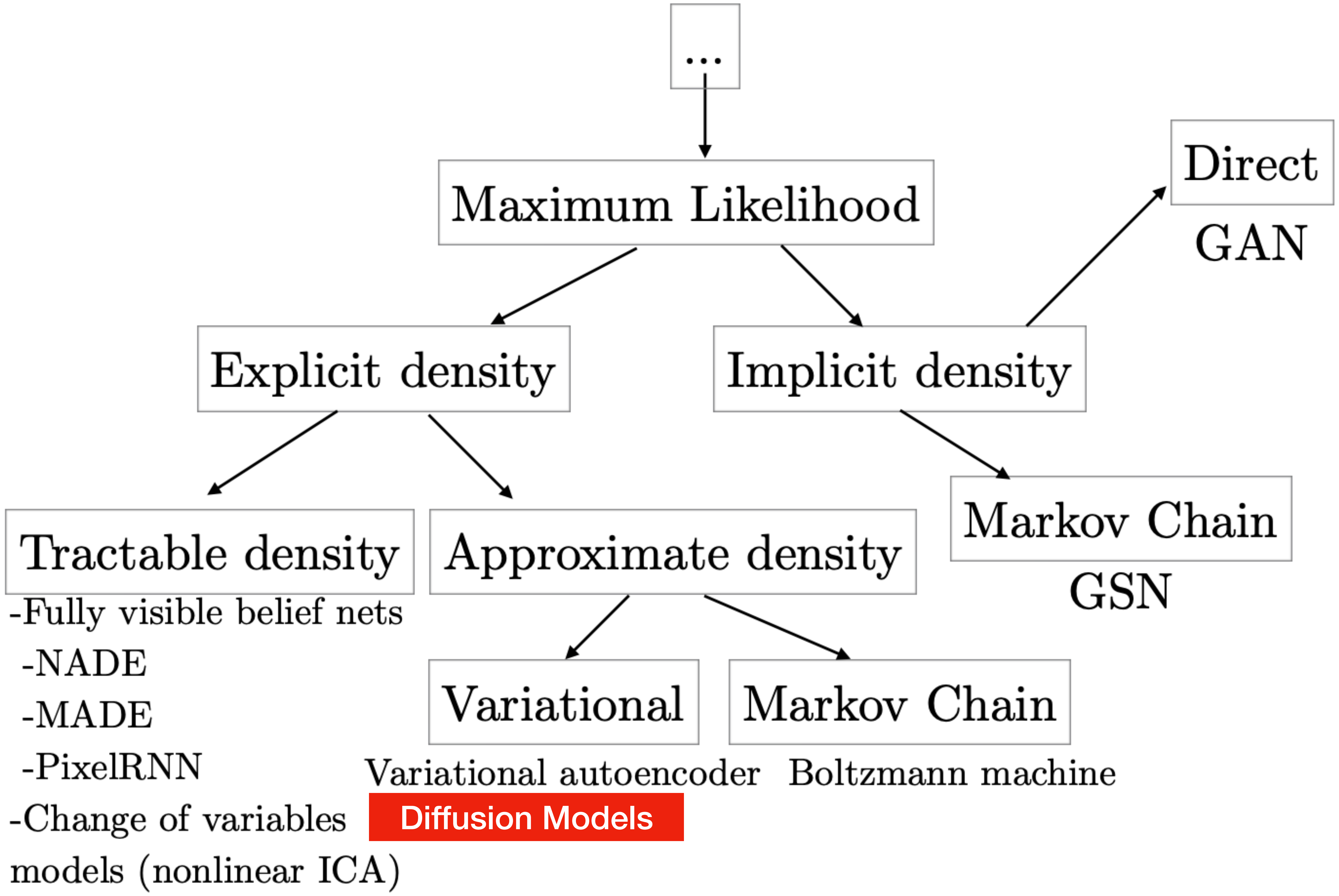
- Train a Generator G parametrised by θ from latent space to data space and approximate the real data distribution
- Training is difficult
  - Hyperparameter choice
  - Quantify similarity between sets
  - Choice of latent space

Deep Generative Models - Taxonomy



Goodfellow, Ian. "Nips 2016 tutorial: Generative adversarial networks." (2016).

Deep Generative Models - Taxonomy

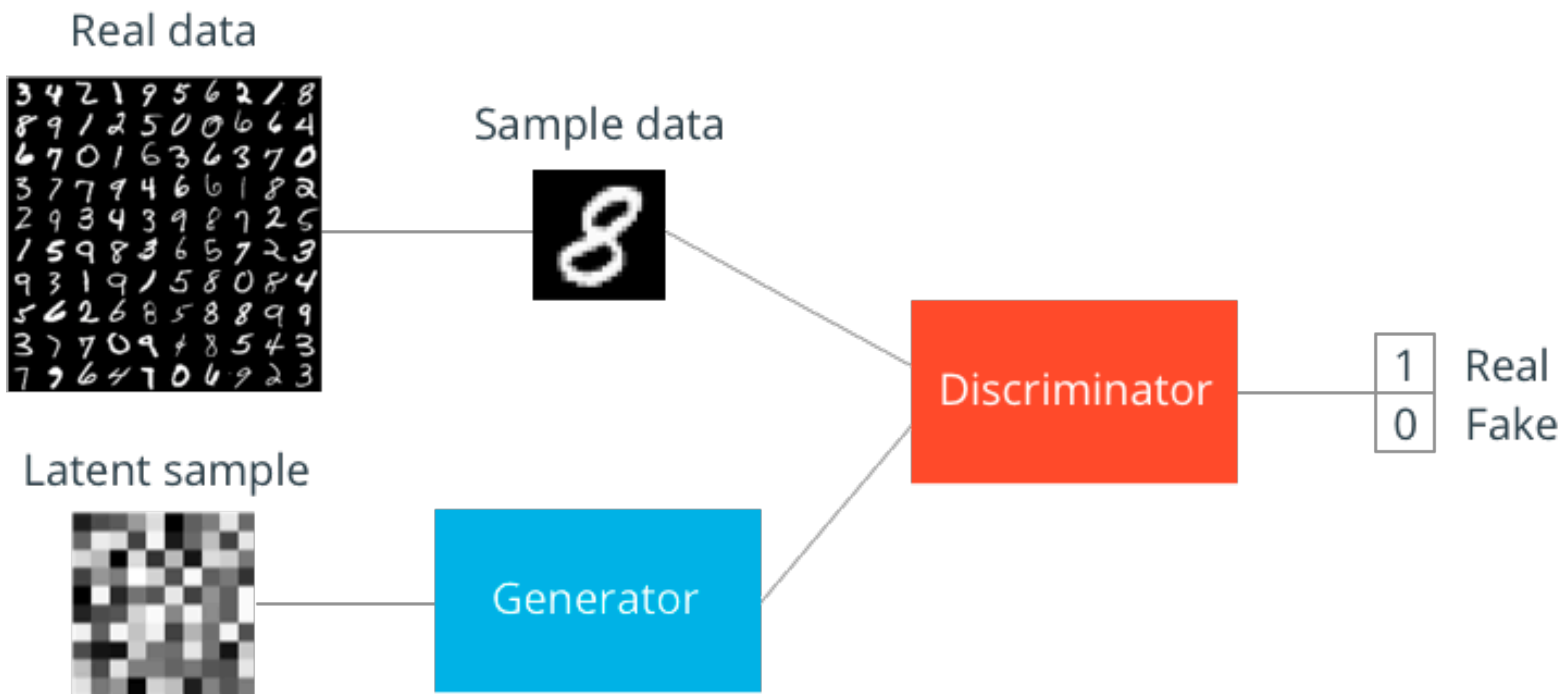


Goodfellow, Ian. "Nips 2016 tutorial: Generative adversarial networks." (2016).

Generative Adversarial Networks

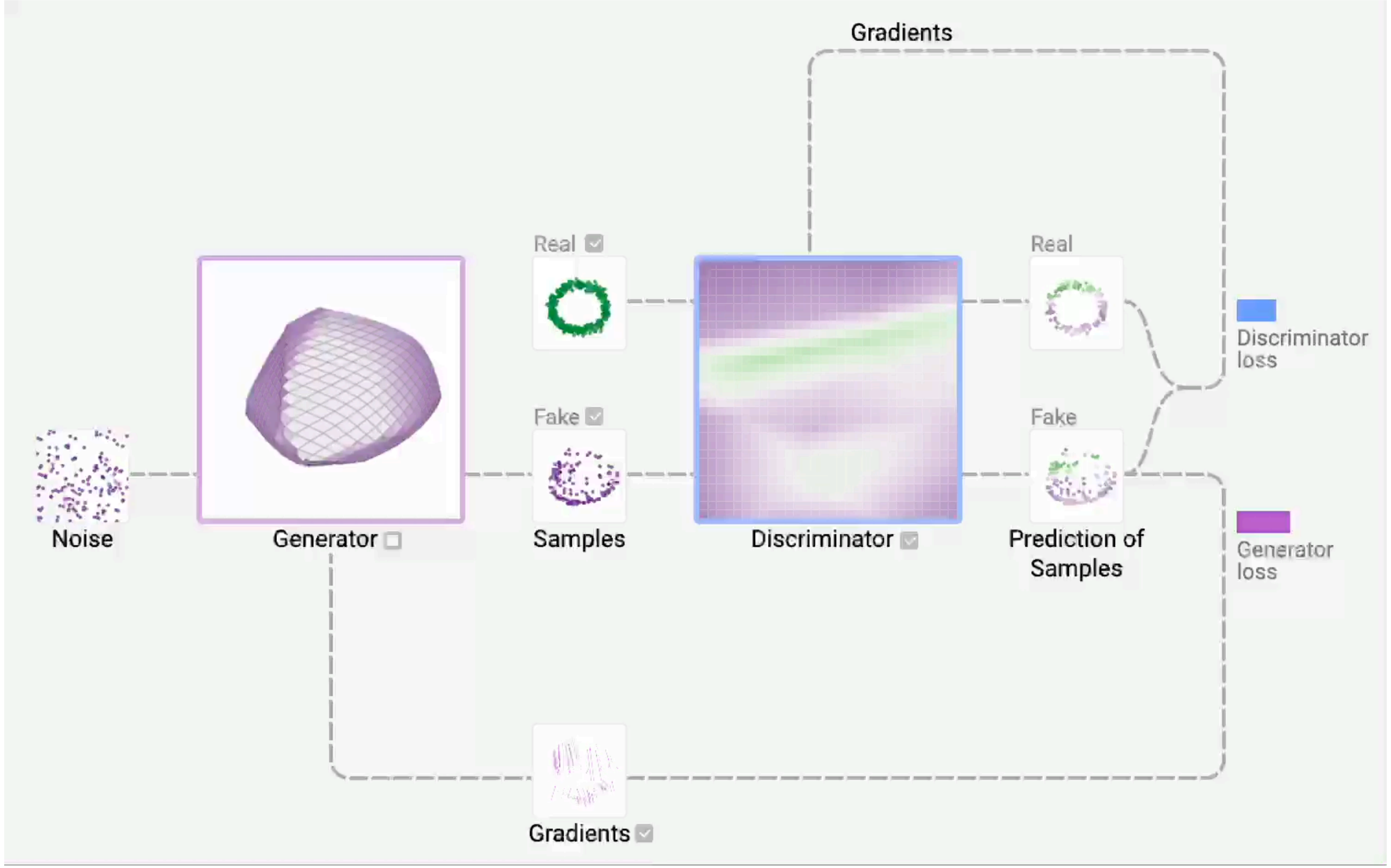
- Trained in data space, implicit density of latent space
- Similarity of generated space and real space is computed through a second network D
- Training goal is to find a saddle point, or Nash equilibrium, between the two networks
- Gradient propagation could be a problem, GAN training is known to be fickle

Framework



- Generator samples noise from a known latent distribution Z
  - Learns a mapping from latent space to data space
- Discriminator tries to distinguish between real and fake samples
- Training becomes a game between the generator and discriminator
  - G tries to fool D into thinking its samples are real

Framework - Nice viz



https://poloclub.github.io/ganlab/

Results

MNIST, Toronto Face Dataset, FC CIFAR-10, Con CIFAR-10 results

Face Generation

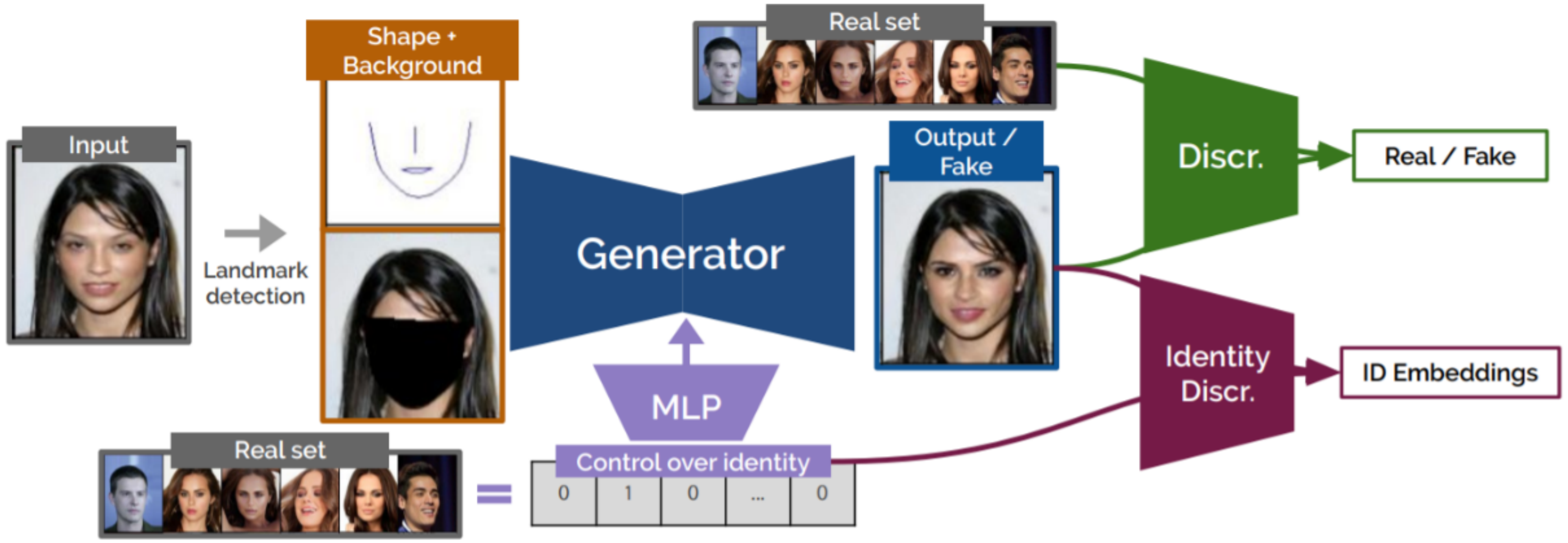
- Quality has greatly improved since Goodfellow's implementation
- StyleGAN family are able to produce realistic faces and to interpolate smoothly in the latent space
- There are still distinctive traits to tell if an image is artificially-generated (e.g. eyes shape, heartbeat)

Karras, Tero, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. "Alias-free generative adversarial networks." *Advances in Neural Information Processing Systems* 34 (2021).

Visual Anonymisation

Face swapping

- In literature there is a tradeoff between face quality, pose preservation and anonymisation performances
- CIAGAN covers same setting as ours, with poor face quality
- More recent works achieve good quality, but the setting is different



CIAGAN architecture

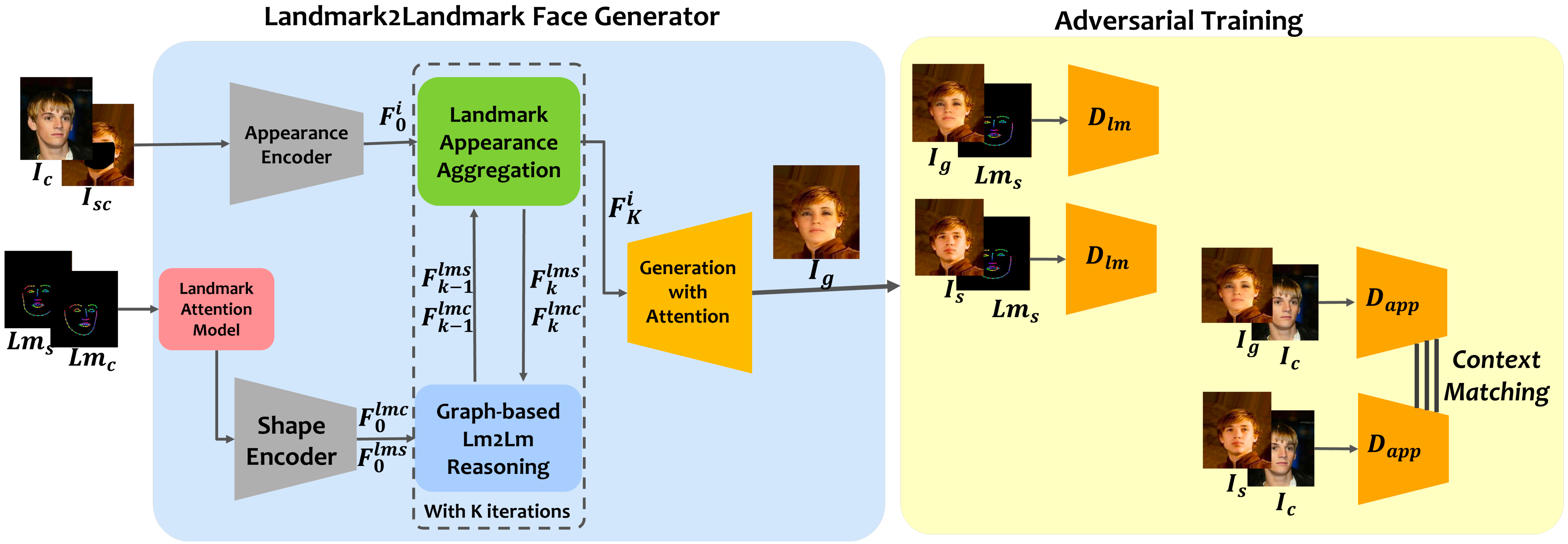
FaceShifter qualitative results

Maximov, Maxim, Ismail Elezi, and Laura Leal-Taixé. "Ciagan: Conditional identity anonymization generative adversarial networks." In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 5447-5456. 2020.
Li, Lingzhi, Jianmin Bao, Hao Yang, Dong Chen, and Fang Wen. "Faceshifter: Towards high fidelity and occlusion aware face swapping." *arXiv preprint arXiv:1912.13457* (2019).

Talk overview

- **Chapter 1:** Why do we need privacy-preserving ML?
- **Chapter 2:** Brief recap of Generative Models
- **Chapter 3:** Graph-based Generative Face Anonymisation with Pose Preservation
- **Chapter 4:** Current state-of-the-art and future work

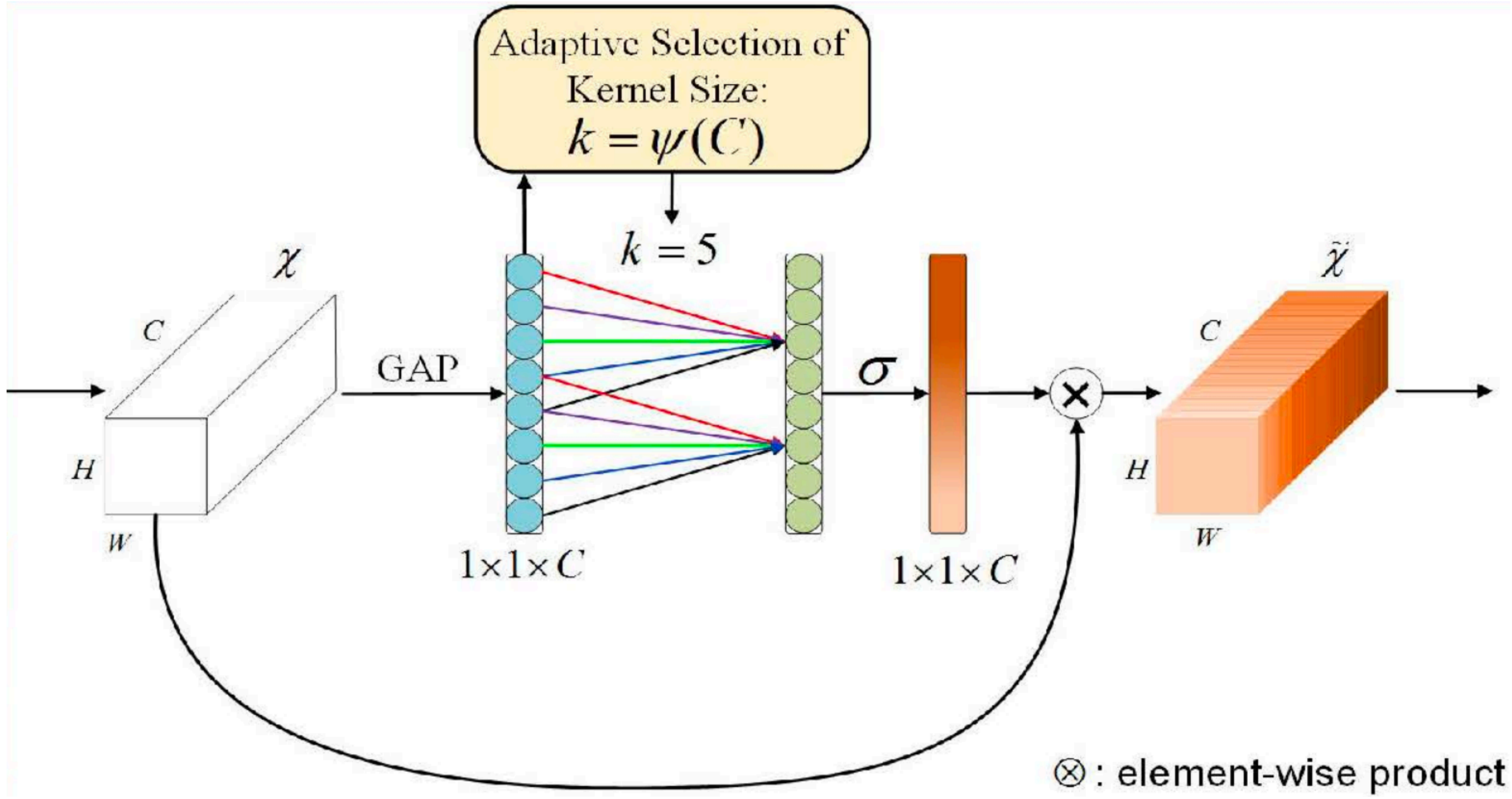
Architecture



- Condition and context source are encoded with the appearance encoder
- Source and condition landmarks are encoded with the shape encoder
- Shape codes are reasoned on a bipartite graph, and aggregated with the appearance
- Adversarial training with two discriminators covers appearance and pose

Dall'Asen, Nicola, Yiming Wang, Hao Tang, Luca Zanella, and Elisa Ricci. "Graph-Based Generative Face Anonymisation with Pose Preservation." In *International Conference on Image Analysis and Processing*, pp. 503-515. Springer, Cham, 2022.

Landmark Attention Model



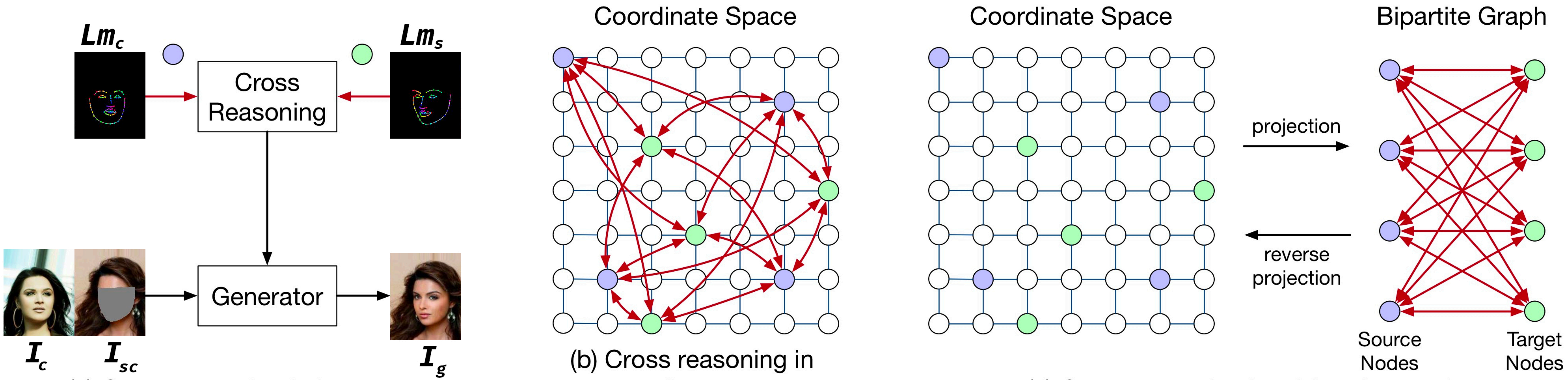
- Designed to learn the weighting strategy on the landmarks to strike the balance between visual naturalness, pose preservation, face detection and face de-identification.

$$\omega = \sigma(Conv_{1D_J}(GAP(Concat(Lm_c, Lm_s))))$$

where $\sigma(\cdot)$ is the sigmoid and $Conv_{1D_J}(\cdot)$ is 1-D convolution. The resulting $\omega$is a concatenation of two 68-element vector representing the importance of each input channel

Wang, Qilong, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. "Supplementary material for 'ECA-Net: Efficient channel attention for deep convolutional neural networks." In *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, Seattle, WA, USA*, 2020.

Landmark 2 Landmark Generator



(a) Cross reasoning in image space.

(b) Cross reasoning in coordinate space.

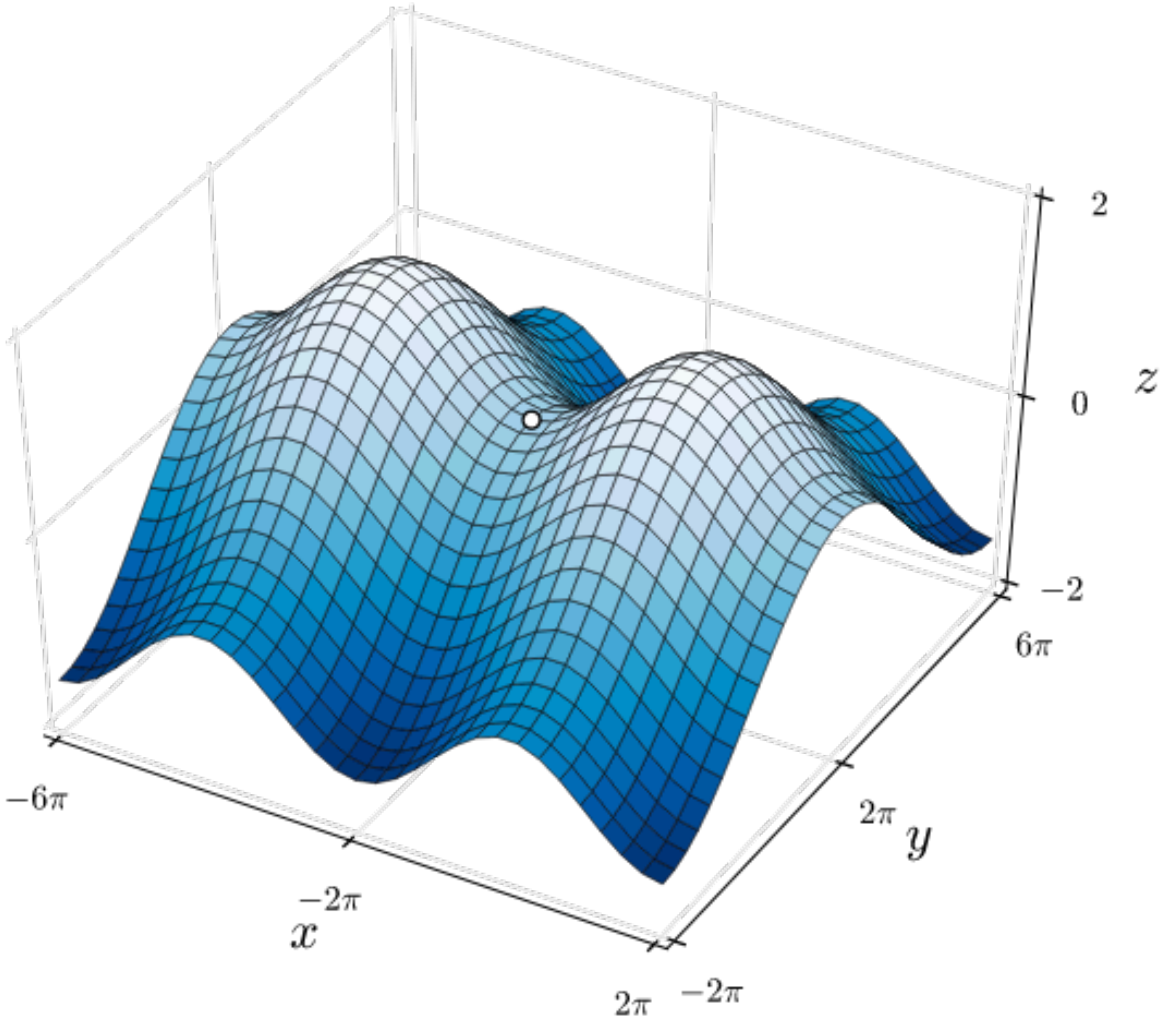
(c) Cross reasoning in a bipartite graph.

- Follows BiGraphGAN architecture to reason between the source landmarks and the condition ones
- Final aggregated feature is used for the landmark-guided face generation with the condition identity
- Landmarks are projected from coordinate space to the bipartite graph and reprojected after cross-reasoning the graph.

Tang, Hao, Song Bai, Philip HS Torr, and Nicu Sebe. "Bipartite graph reasoning gans for person image generation." (2020).

Adversarial Training

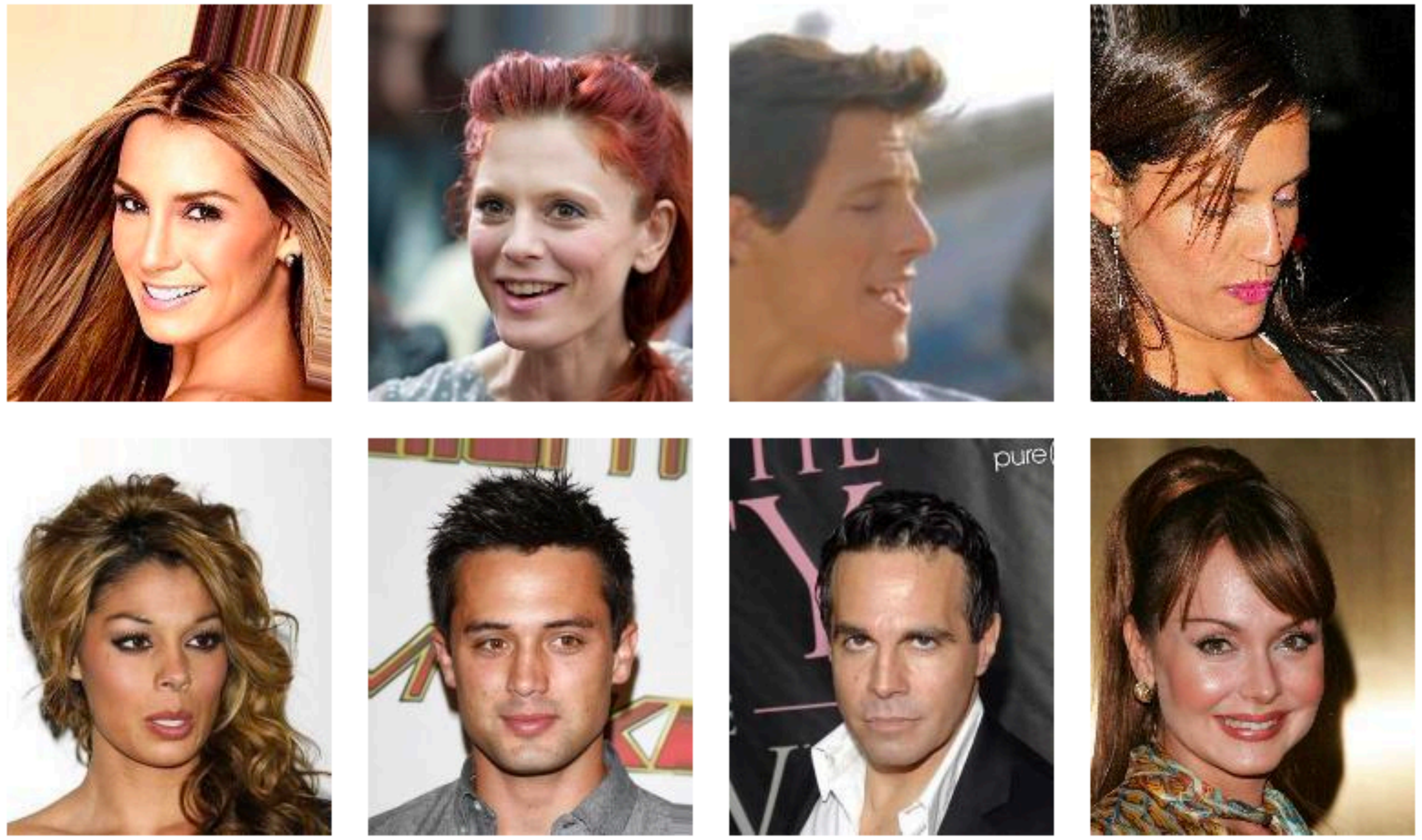
- Landmark discriminator forces the correct pose
- Appearance discriminator helps transferring the condition attributes to the context of the source
- Hybrid training to overcome the lack of supervised training possibility
- Context Matching is achieved by employing a weak Feature Matching loss on the appearance discriminator



Chen, Renwang, Xuanhong Chen, Bingbing Ni, and Yanhao Ge. "Simswap: An efficient framework for high fidelity face swapping." In *Proceedings of the 28th ACM International Conference on Multimedia*, 2020.

Experiments

Dataset



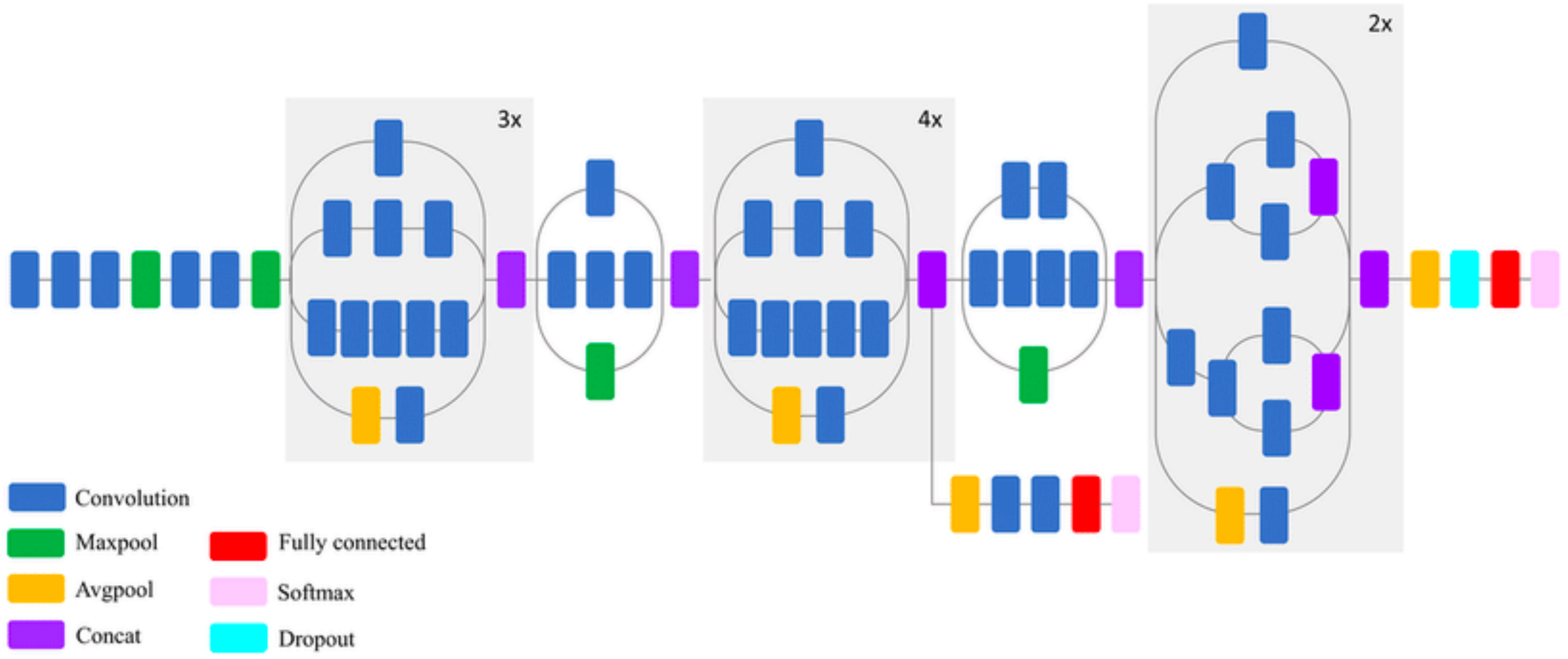
- Large scale Celebrity Faces Attributes (CelebA)
  - 202,599 images with different poses and backgrounds of 10,177 celebrities
- Labelled Faces in the Wild (LFW)
  - 13,233 images collected from the web of 5,749 identities

We follow the same train/test split as CIAGAN and evaluate on both datasets. Images are preprocessed with dlib

Liu, Ziwei, Ping Luo, Xiaogang Wang, and Xiaoou Tang. "Deep learning face attributes in the wild." In *Proceedings of the IEEE international conference on computer vision*, 2015.
Huang, Gary B., Marwan Mattar, Tamara Berg, and Eric Learned-Miller. "Labeled faces in the wild: A database forstudying face recognition in unconstrained environments." In *Workshop on faces in'Real-Life'Images: detection, alignment, and recognition*. 2008.

Metrics

- Image Quality
    - Fréchet Inception Distance (FID)
- Pose Preservation
    - Normalized distance between the detected and ground truth landmarks
- Face Detection
    - dlib
    - FaceNet
- Face re-identification
    - Percentage of anonymised faces mapped to the same identity as original face

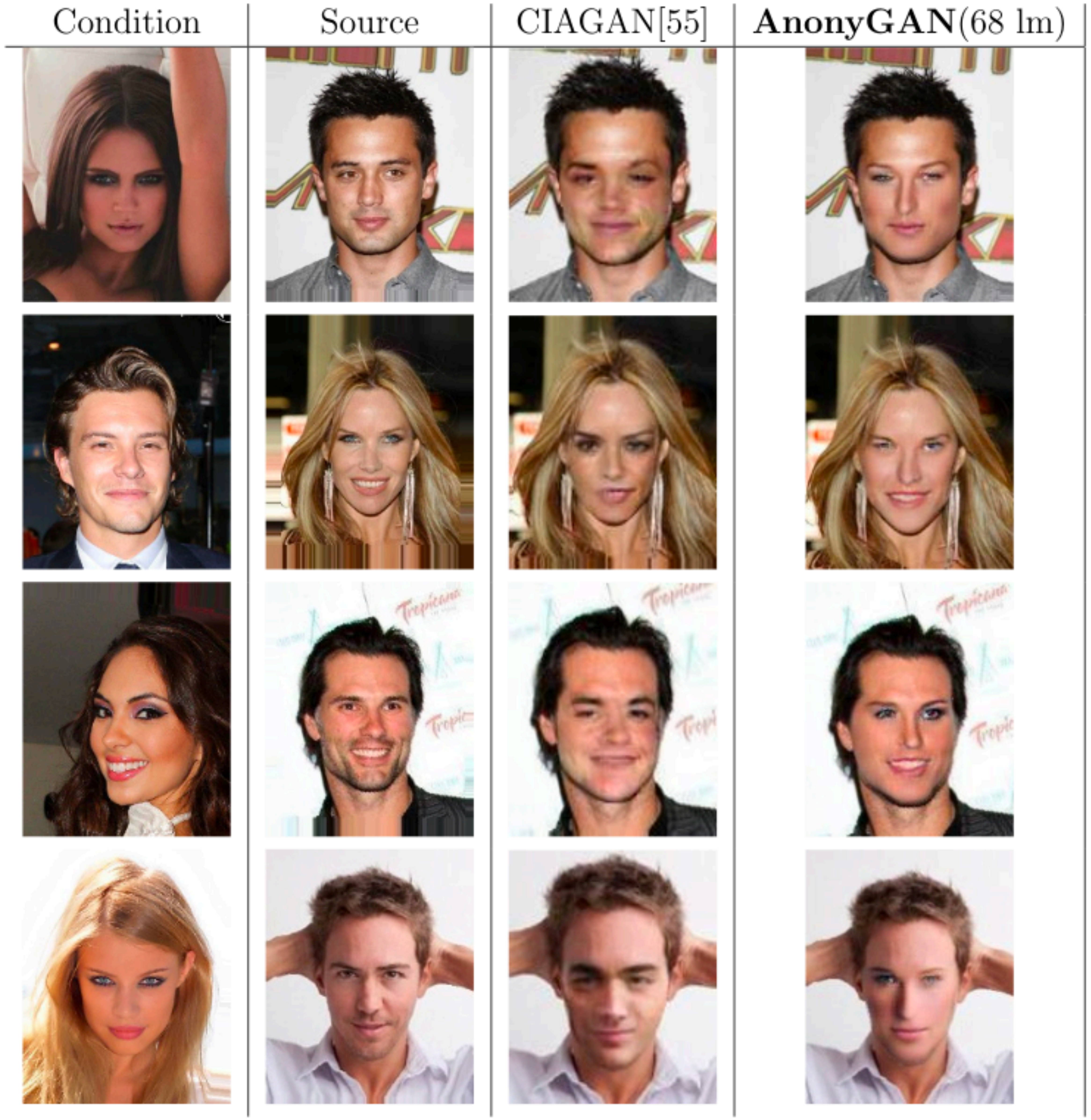


Inception v3 architecture

Face detection with FaceNet

Qualitative Results

- AnonyGAN faces are the most natural-looking
- AnonyGAN better transfers facial attributes of condition images to the context of the source
- AnonyGAN better preserves pose, proving the effect of graph reasoning and Landmark Attention



Quantitative Results

	FID↓	Dectection (dlib)↑	Detection (FaceNet)↑	Re-id (CASIA)↓	Re-id (VGG)↓	Pose↓
Blurring	95.13	4%	4%	**0.07%**	**0.02%**	-
Pixelation	59.82	1%	28%	0.28%	0.12%	-
CIAGAN	37.94	96%	**100%**	1.61%	0.51%	1.44
AnonyGAN-CM[1]-LA[2] (68lm)	43.99	**100%**	**100%**	2.63%	0.58%	**0.16**
AnonyGAN-CM-LA (29lm)	30.24	**100%**	**100%**	2.84%	0.66%	**0.16**
AnonyGAN-CM (68lm)	26.12	**100%**	**100%**	2.70%	0.91%	**0.16**
AnonyGAN (68lm)	**22.53**	**100%**	**100%**	3.52%	1.60%	**0.16**

Context Matching

Landmark Attention

Quantitative Discussion

- AnonyGAN generates better images than classical techniques and CIAGAN, given by the lower FID score
- AnonyGAN better preserves pose thanks to the graph-based reasoning among landmarks
- Landmark Attention enables the network to achieve higher quality without impacting pose preservation
- Context Matching allows for the most natural faces at a slight compromise in re-identification rates

Limitations and what didn’t work

A.k.a. why I am changing topic right now

- It requires precise localisation of landmarks
    - Which is impossible in most real-time scenarios
- No explicit time constraints for video anonymisation

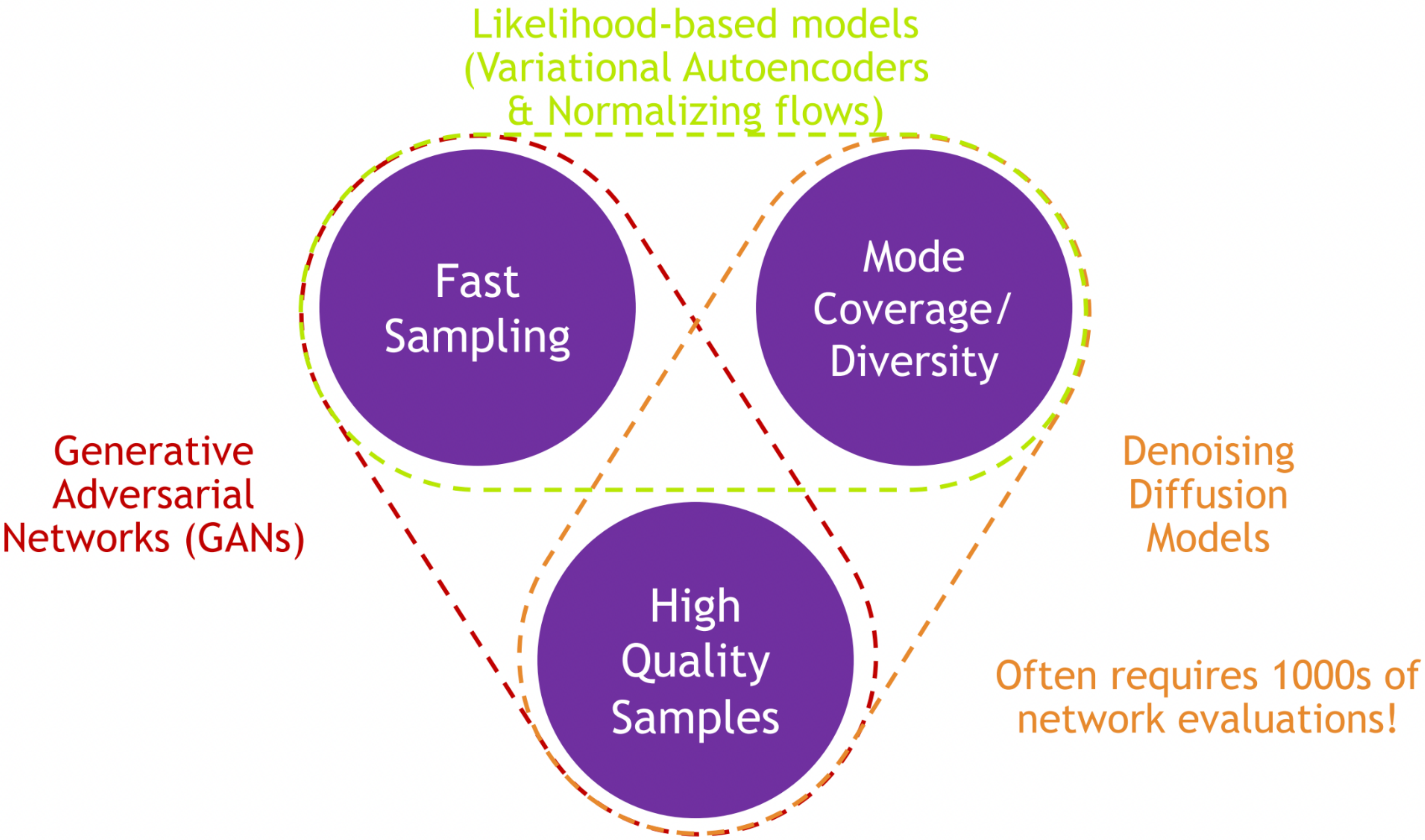
- We tried introducing Transformer to relax the landmark input but…
    - Difficult to train
    - Different mixing strategies didn’t solve the issue

Talk overview

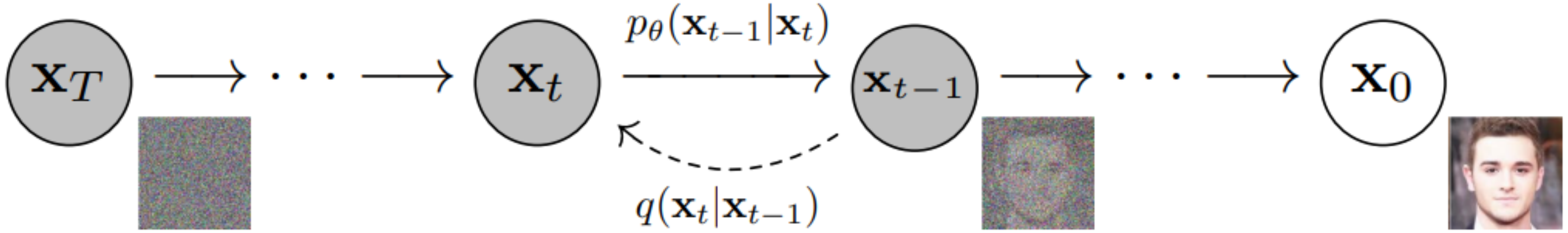
- **Chapter 1:** Why do we need privacy-preserving ML?
- **Chapter 2:** Brief recap of Generative Models
- **Chapter 3:** Graph-based Generative Face Anonymisation with Pose Preservation
- **Chapter 4:** Current state-of-the-art and future work

Generative Trilemma

What makes a good Generative Model?



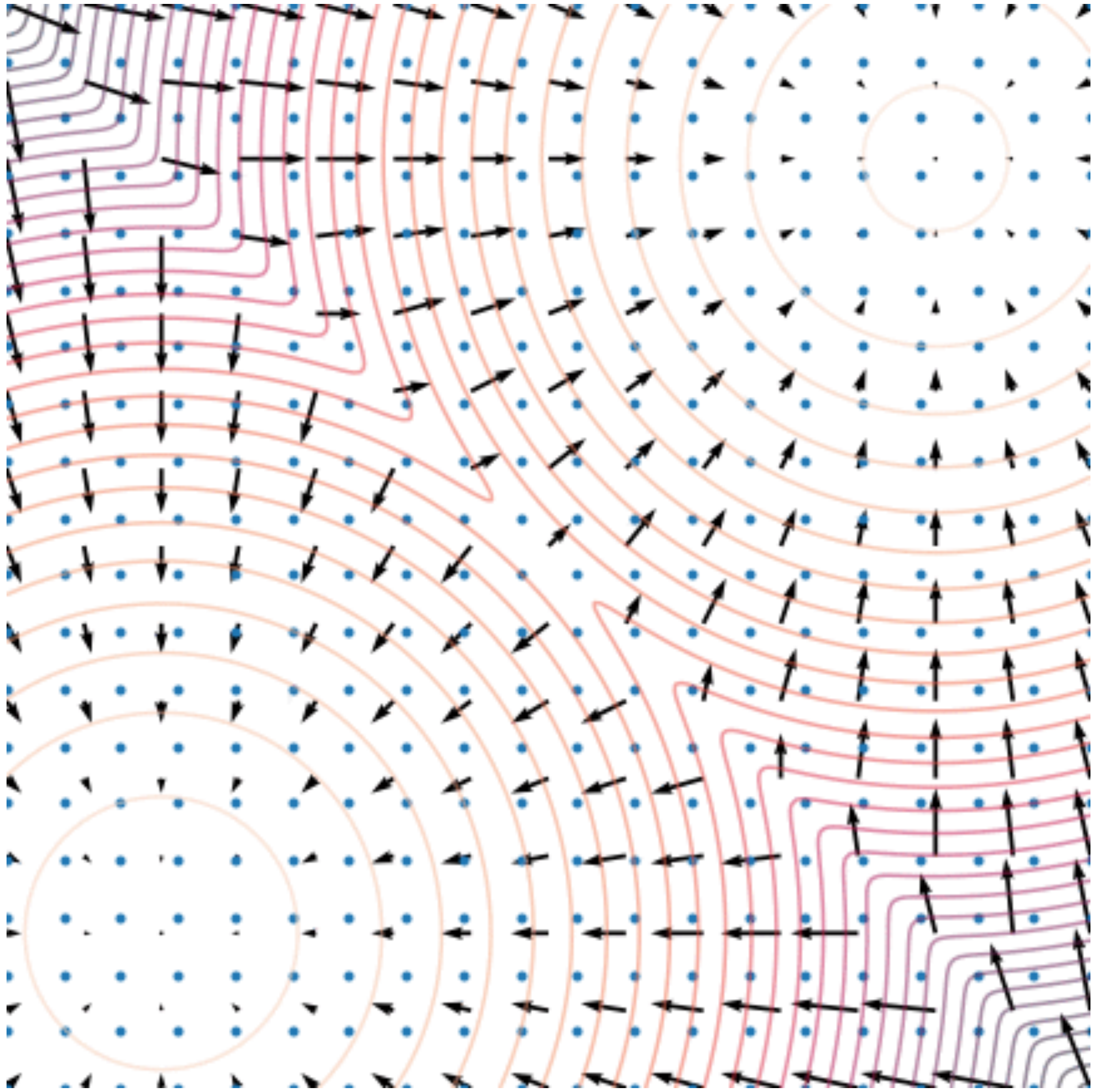
Diffusion Models



- Destroy the data in the forward process by adding Gaussian noise
- Learn to recover in the reverse process
- Sample after training by just feeding random noise

Alternative Formulation

- Alternative formulations seek the direction from noise to data through score matching
- Sample through Langevin dynamics given the score function
- Can be formulated in terms of SDE
  - This leads to other mathematical formulations and problems



Yang, and Stefano Ermon. "Generative modeling by estimating gradients of the data distribution." Advances in Neural Information Processing Systems 32 (2019).
Song, Yang, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. "Score-based generative modeling through stochastic differential equations." (2020).

Diffusion Model beat GANs

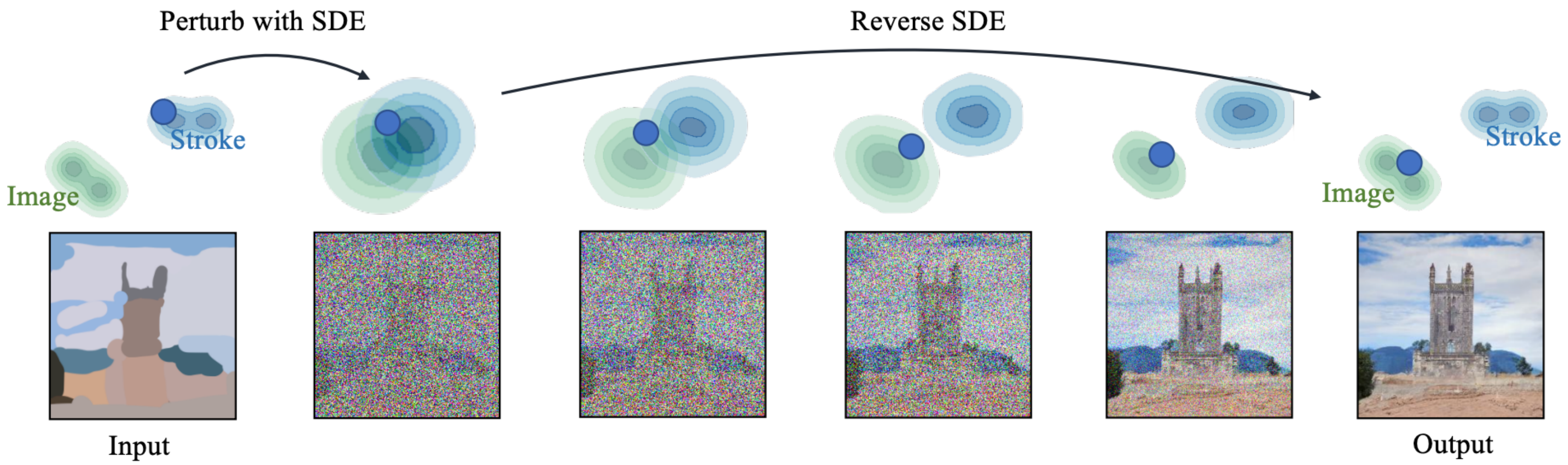
- Very quickly Diffusion Models reached performance of GANs and overcome them
- Most of the tricks used to beat GANs are engineering and not theoretical

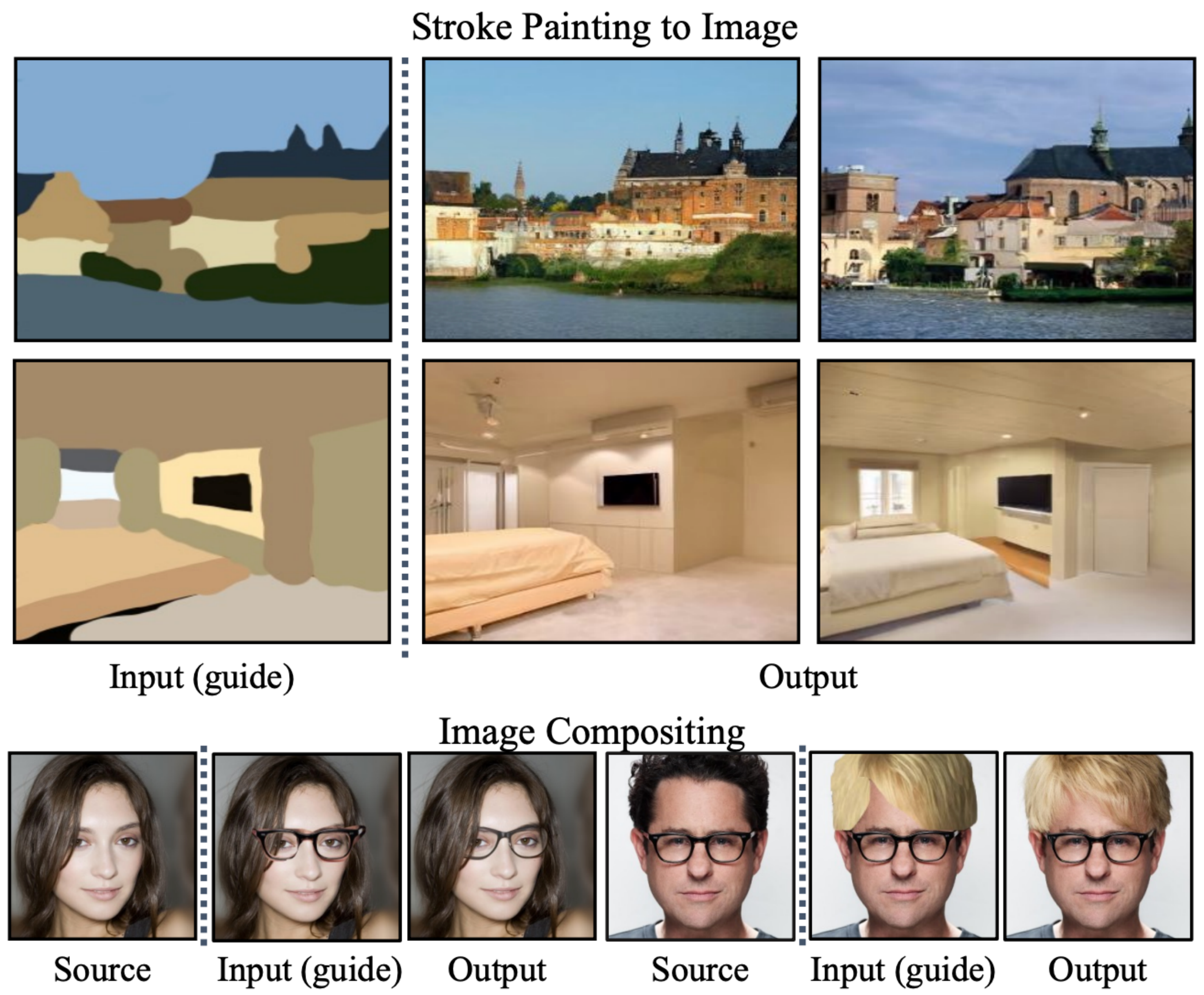
Model	FID	sFID	Prec	Rec
ImageNet 128×128				
BigGAN-deep [5]	6.02	7.18	**0.86**	0.35
LOGAN† [68]	3.36			
ADM	5.91	**5.09**	0.70	**0.65**
ADM-G (25 steps)	5.98	7.04	0.78	0.51
ADM-G	**2.97**	**5.09**	0.78	0.59
ImageNet 256×256				
DCTransformer† [42]	36.51	8.24	0.36	**0.67**
VQ-VAE-2†‡ [51]	31.11	17.38	0.36	0.57
IDDPM‡ [43]	12.26	5.42	0.70	0.62
SR3†‡ [53]	11.30			
BigGAN-deep [5]	6.95	7.36	**0.87**	0.28
ADM	10.94	6.02	0.69	0.63
ADM-G (25 steps)	5.44	5.32	0.81	0.49
ADM-G	**4.59**	**5.25**	0.82	0.52
ImageNet 512×512				
BigGAN-deep [5]	8.43	8.13	**0.88**	0.29
ADM	23.24	10.19	0.73	**0.60**
ADM-G (25 steps)	8.41	9.67	0.83	0.47
ADM-G	**7.72**	**6.57**	0.87	0.42

Dhariwal, Prafulla, and Alexander Nichol. "Diffusion models beat gans on image synthesis." *Advances in Neural Information Processing Systems* 34 (2021).

Diffusion Model Qualitative Results

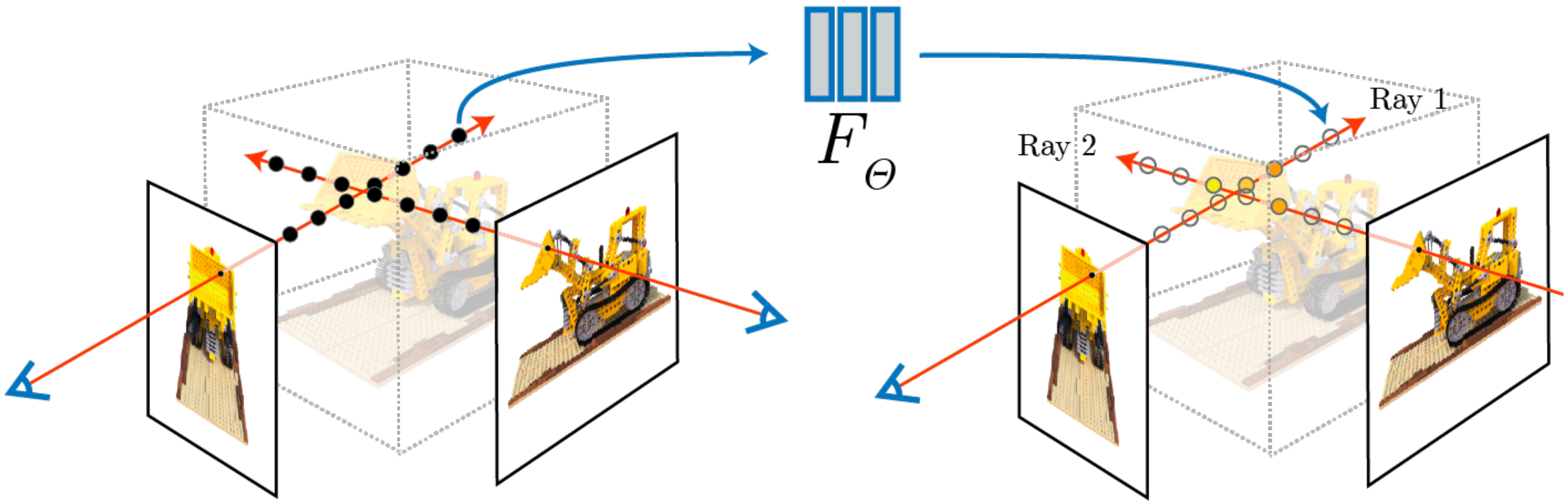
Figure 1: Selected samples from our best ImageNet 512×512 model (FID 3.85)



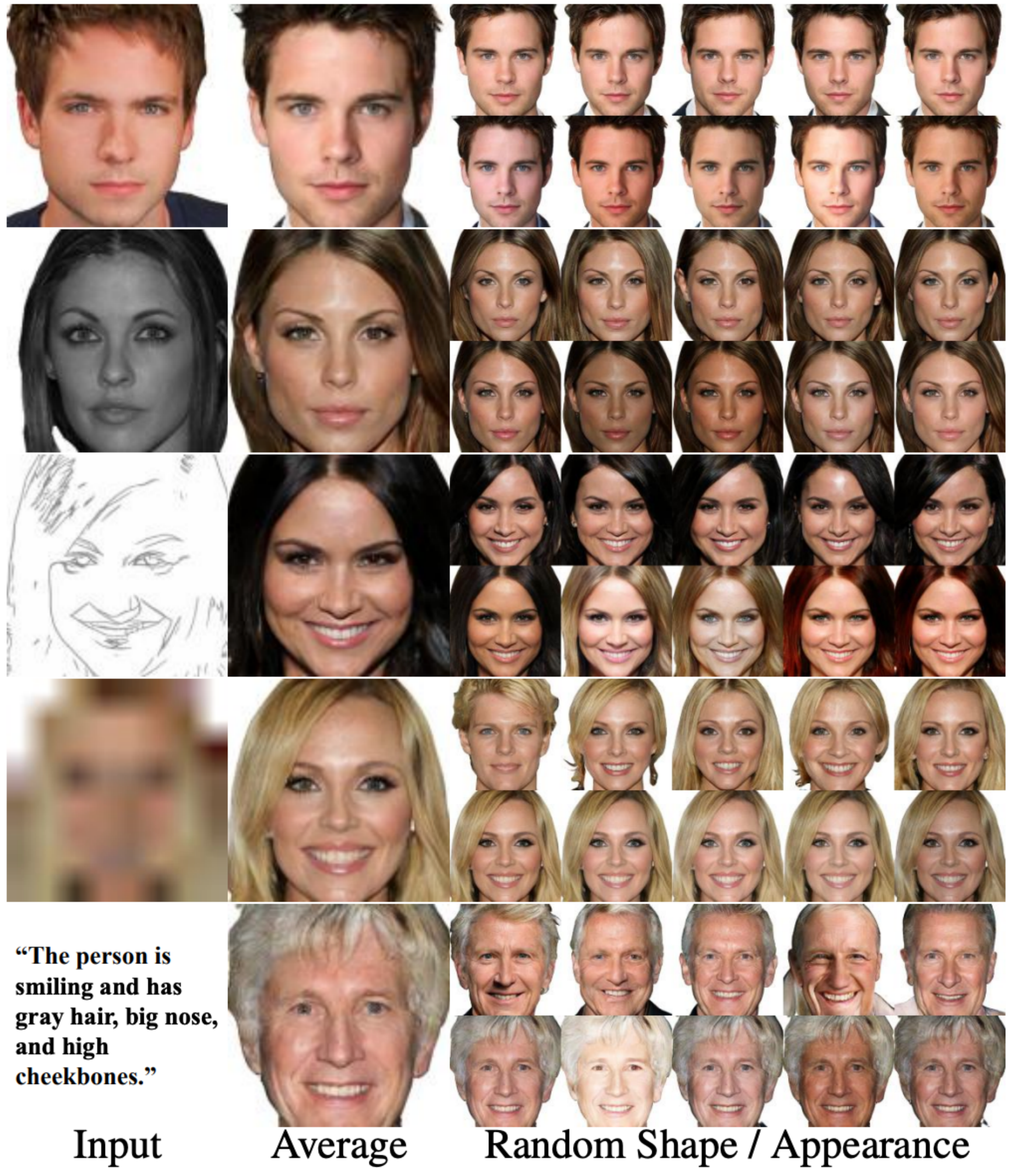


Meng, Chenlin, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. "Sdedit: Image synthesis and editing with stochastic differential equations." *arXiv preprint arXiv:2108.01073* (2021).
Dhariwal, Prafulla, and Alexander Nichol. "Diffusion models beat gans on image synthesis." *Advances in Neural Information Processing Systems* 34 (2021): 8780-8794.

Future Research I



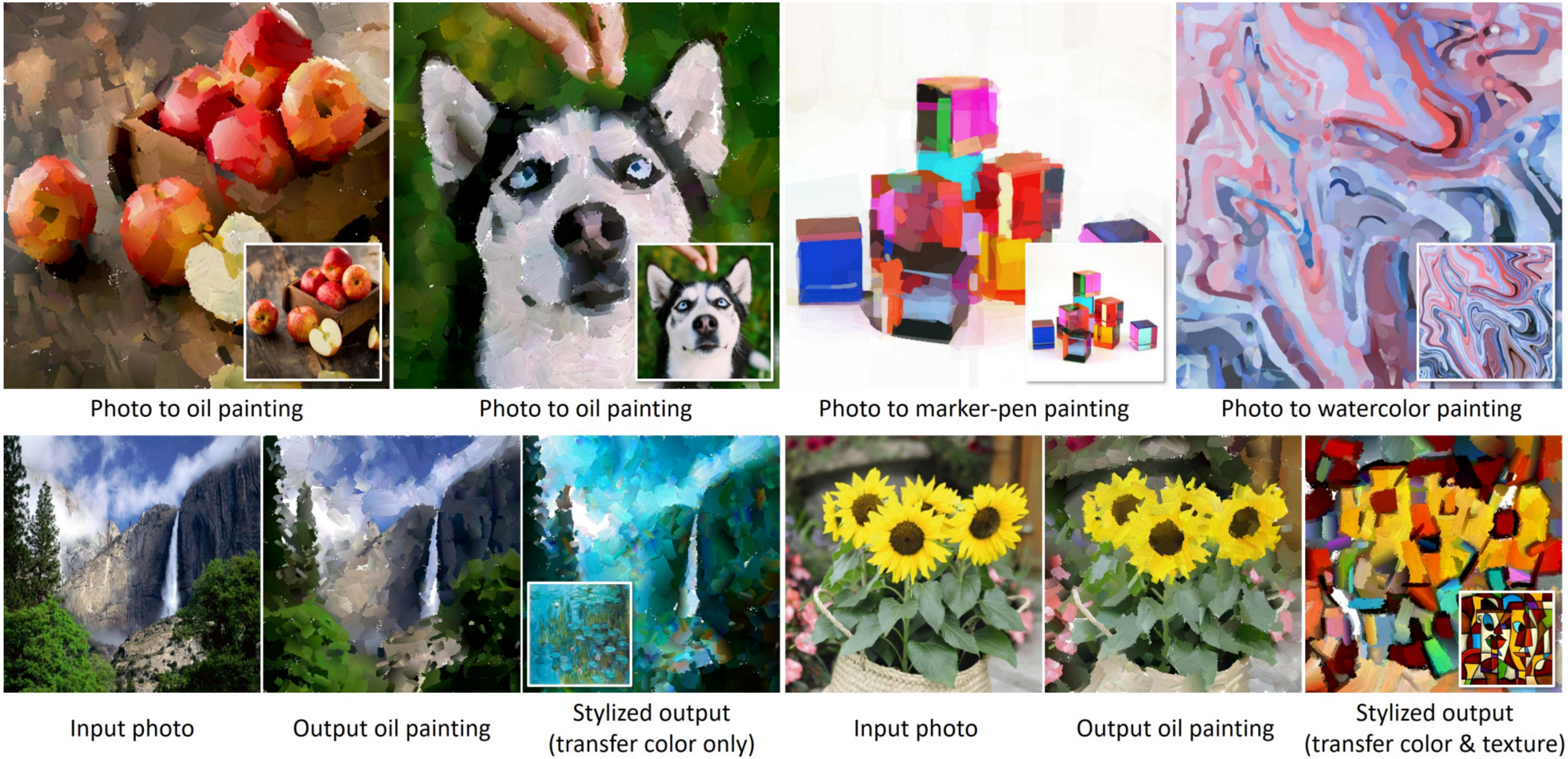
NERF



CG-NERF

Mildenhall, Ben, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. "Nerf: Representing scenes as neural radiance fields for view synthesis." In *European conference on computer vision*, pp. 405-421. Springer, Cham, 2020.
Jo, Kyungmin, Gyumin Shim, Sanghun Jung, Soyoung Yang, and Jaegul Choo. "Cg-nerf: Conditional generative neural radiance fields." arXiv preprint arXiv:2112.03517 (2021).

Future Research II



Zou, Zhengxia, Tianyang Shi, Shuang Qiu, Yi Yuan, and Zhenwei Shi. "Stylized neural painting." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15689-15698. 2021.