Video anonymization

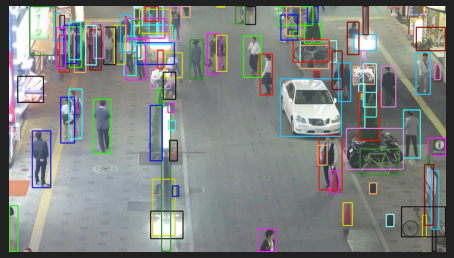
Prof. Dr. Laura Leal-Taixé

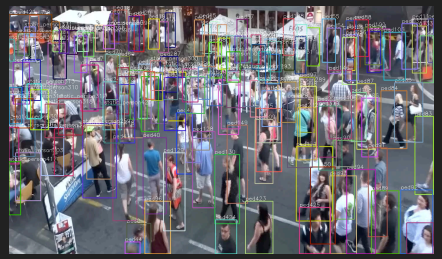
Technical University of Munich

"All human beings have three lives: public, private,and secret."
Gabriel García Márquez

Motivation

How I see my work





- Challenging
- Plenty of applications: autonomous driving, robot navigation

How others see my work



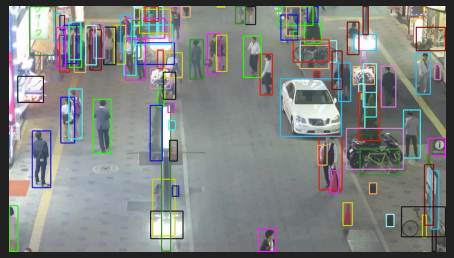
Big brother

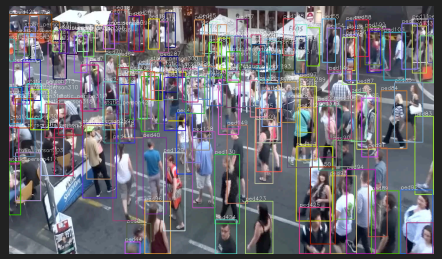


Data from www.motchallenge.net

Motivation

How I see my work





- Challenging
- Plenty of applications: autonomous driving, robot navigation

How others see my work



I do not care if this is Mark or John, I only use a label “person”

Data from www.motchallenge.net

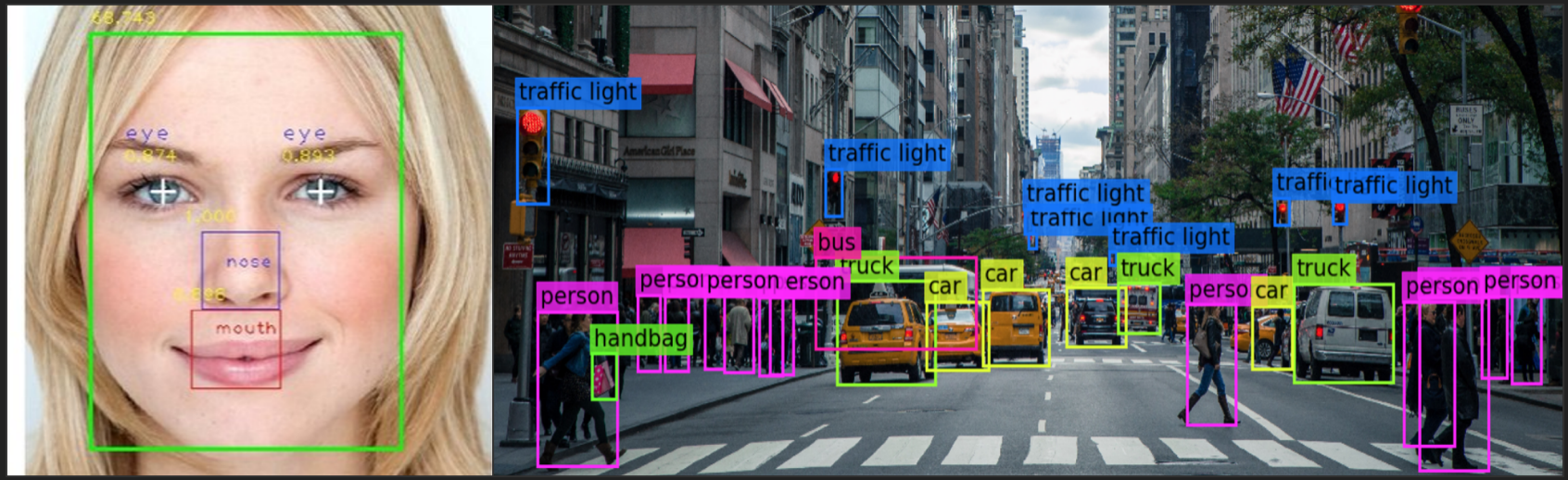
Motivation



Just remove a face using blur/square/mosaic

https://arxiv.org/abs/1803.11556 - Learning to Anonymize Faces for Privacy Preserving Action Detection

Motivation



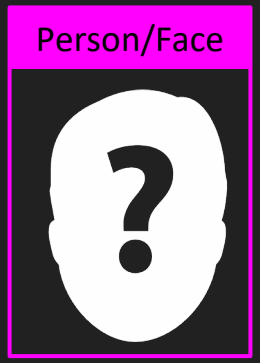
Detection and tracking performance is heavily affected

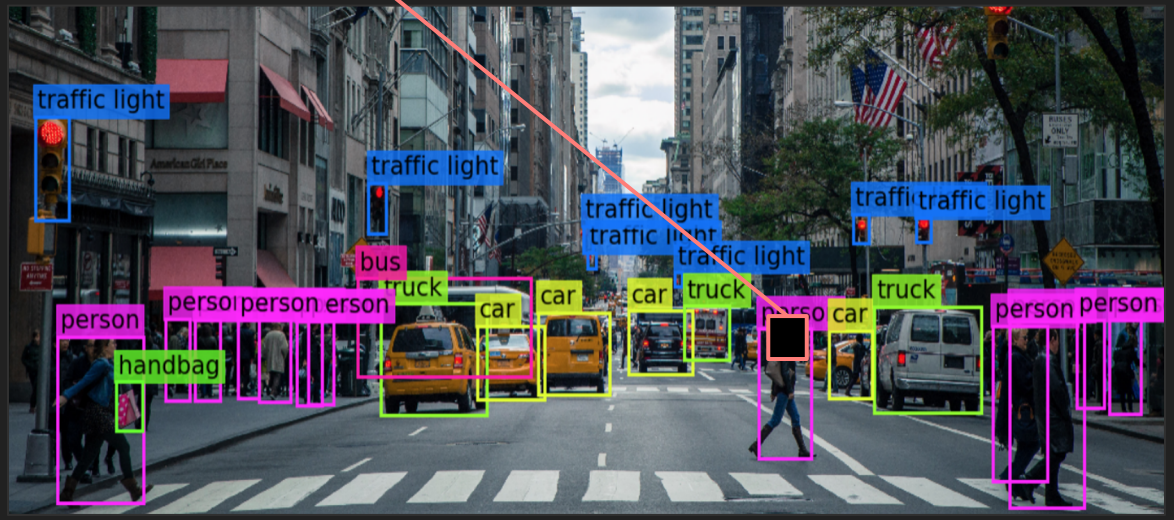
Images:
Left - https://www.researchgate.net/publication/308944615_A_Fast_Deep_Convolutional_Neural_Network_for_Face_Detection_in_Big_Visual_Data
Right - https://towardsdatascience.com/you-only-look-once-yolo-implementing-yolo-in-less-than-30-lines-of-python-code-97fb9835bfd2

Goals for anonymization









Properties:

- Anonymous
- Realistic (for a CV algorithm)
- New Identity
- Control
- Temporal Consistency

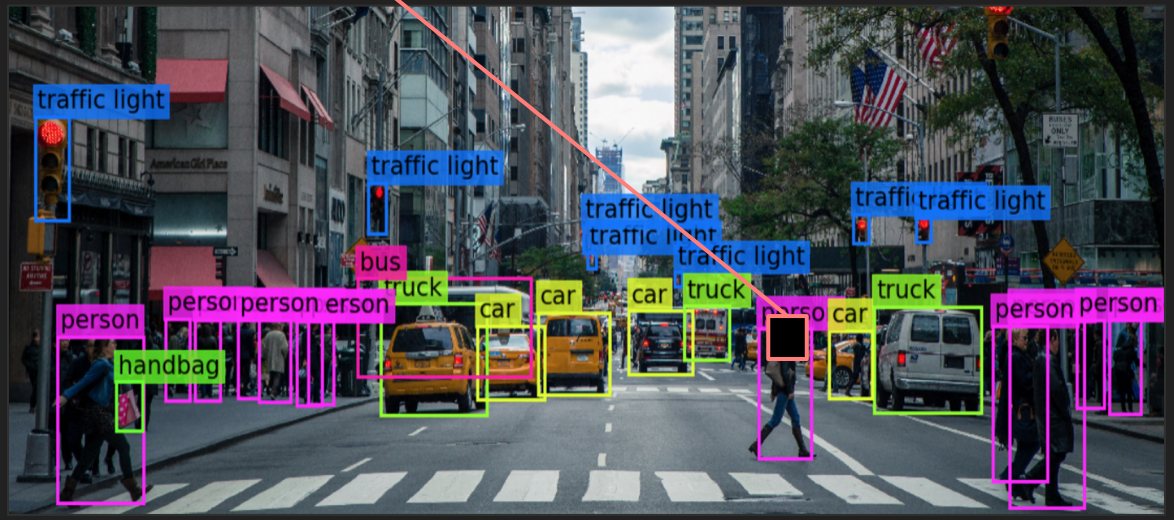
Reference:

Face swap









Properties:

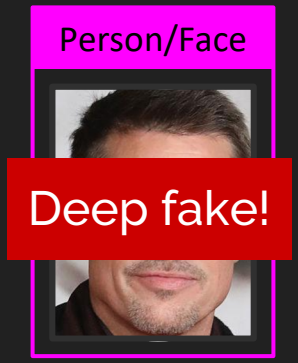
- Anonymous
- Realistic (for a CV algorithm)
- New Identity
- Control
- Temporal Consistency

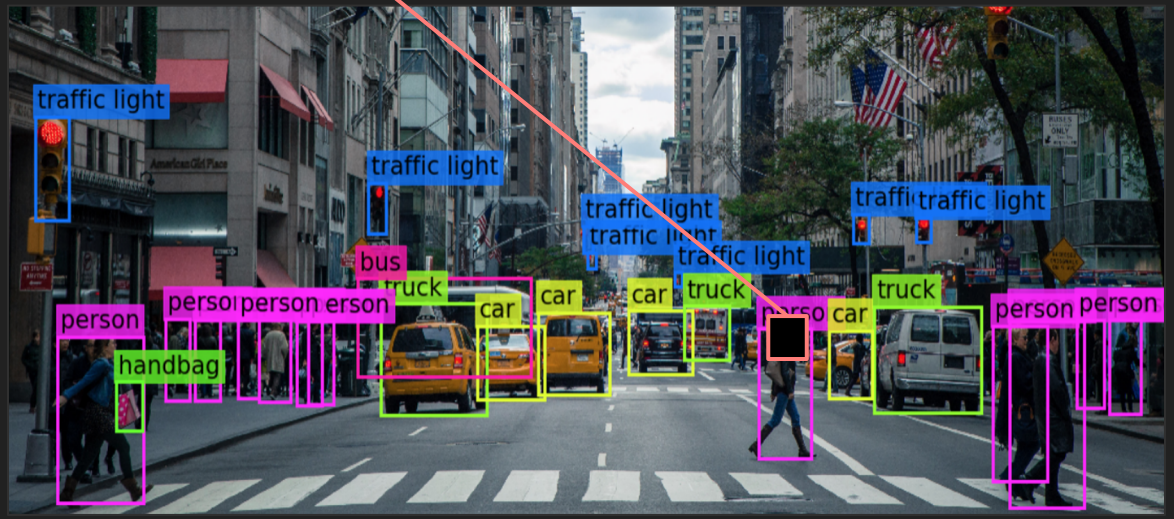
Reference:

Face swap









Properties:

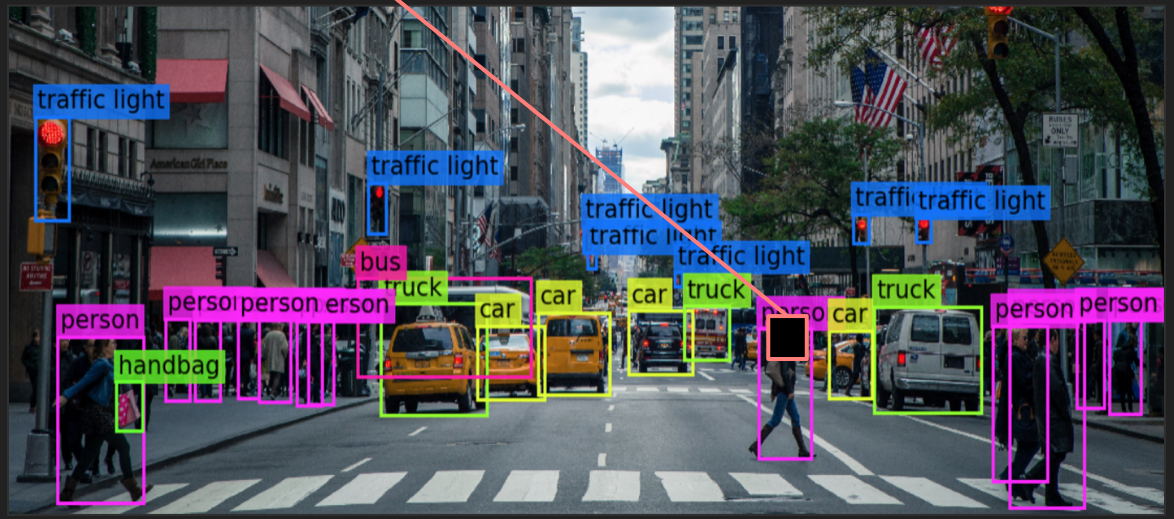
- Anonymous
- Realistic (for a CV algorithm)
- New Identity
- Control
- Temporal Consistency

Anonymization: previous work









Properties:

- Anonymous
- Realistic (for a CV algorithm)
- New Identity
- Control (one-to-many)
- Temporal Consistency

Who is he?

More anonymized ⟷ Less anonymized





Gafni et al. “Live face de-identification in video”. ICCV 2019

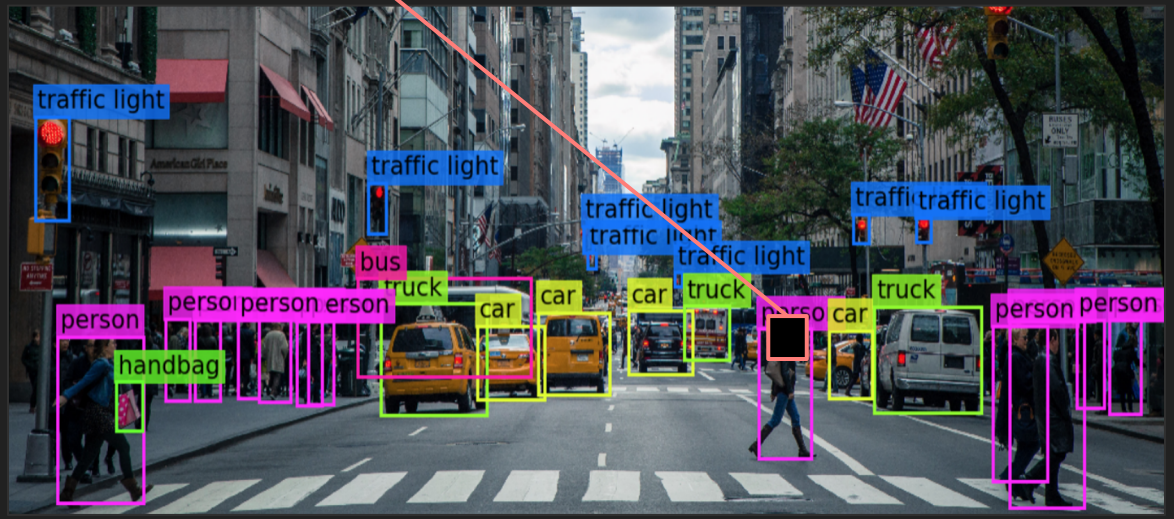
M. Maximov et al. „CIAGAN: Conditional Identity Anonymization Generative Adversarial Networks“. CVPR 2020

Anonymization: previous work









Properties:

- Anonymous
- Realistic (for a CV algorithm)
- New Identity
- Control (one-to-many)
- Temporal Consistency

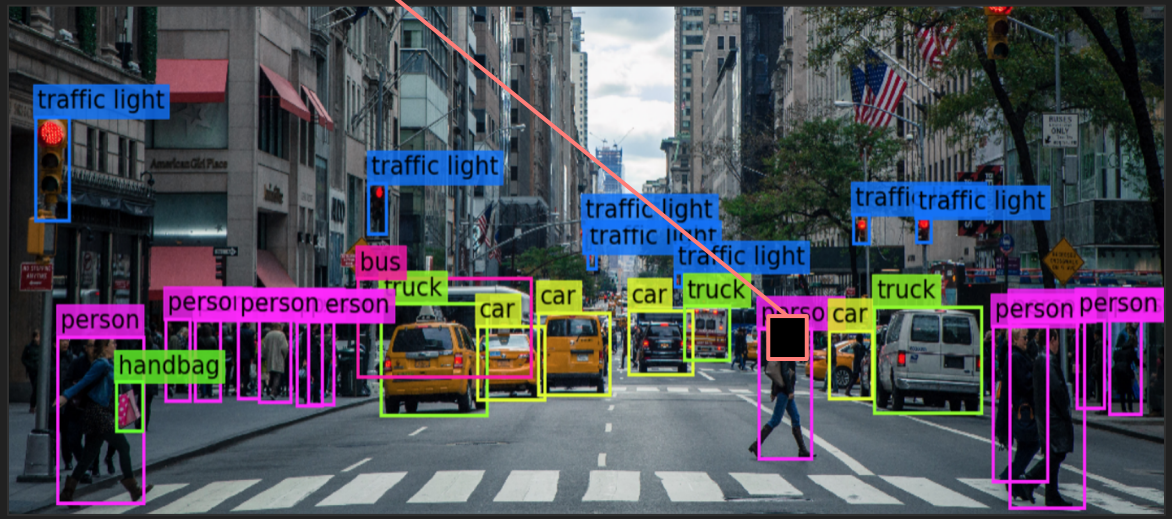
CIAGAN



CNN

Control over identity

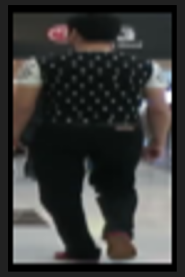


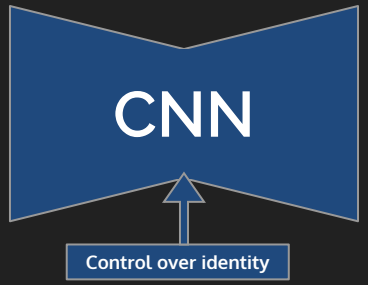


- Anonymous
- Realistic
- New Identity
- Control Temporal Consistency

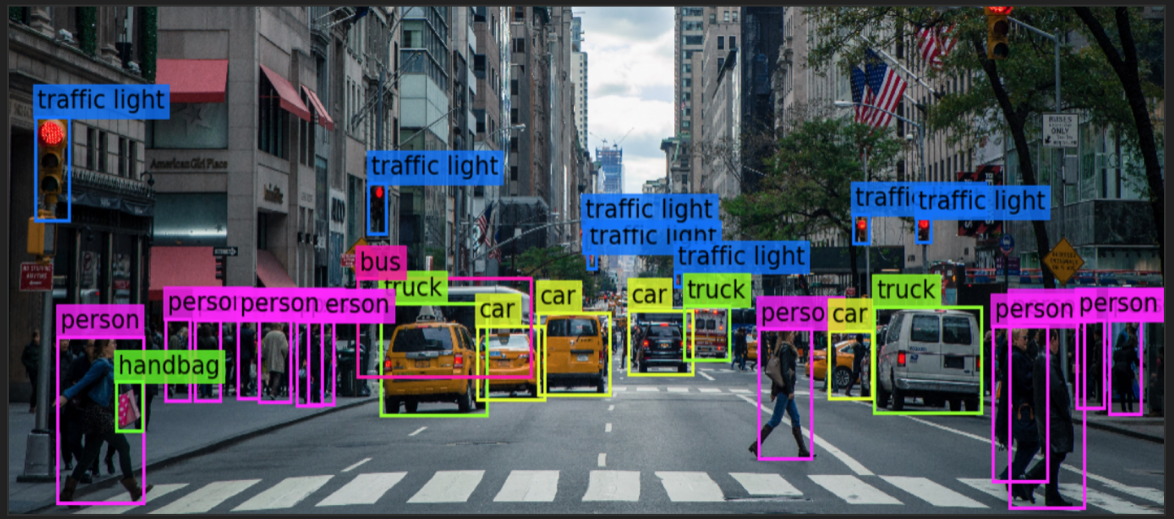
Reference:

CIAGAN







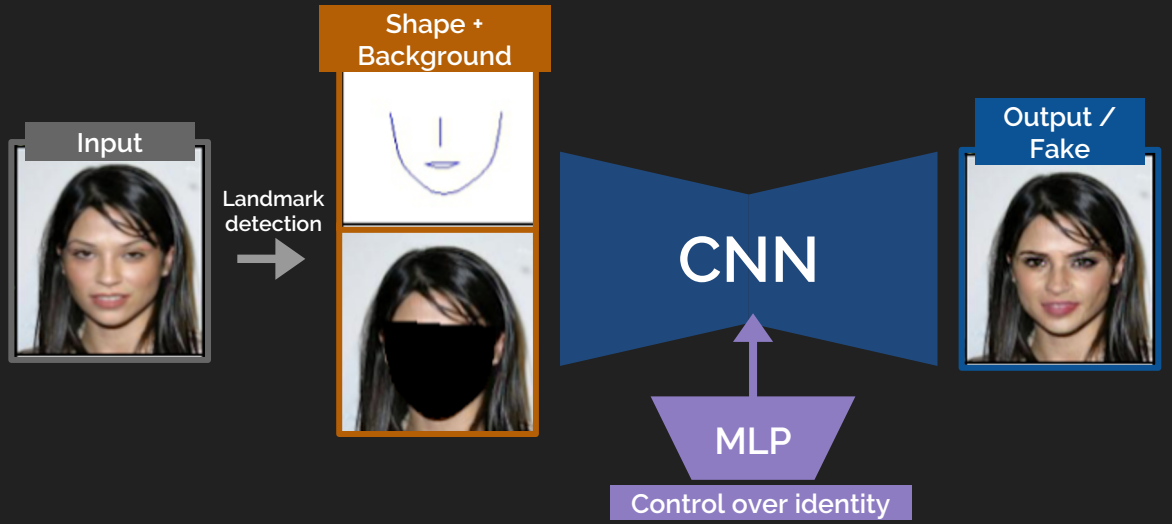


- Also works on full bodies!

Reference:

Methodology

Overview of CIAGAN

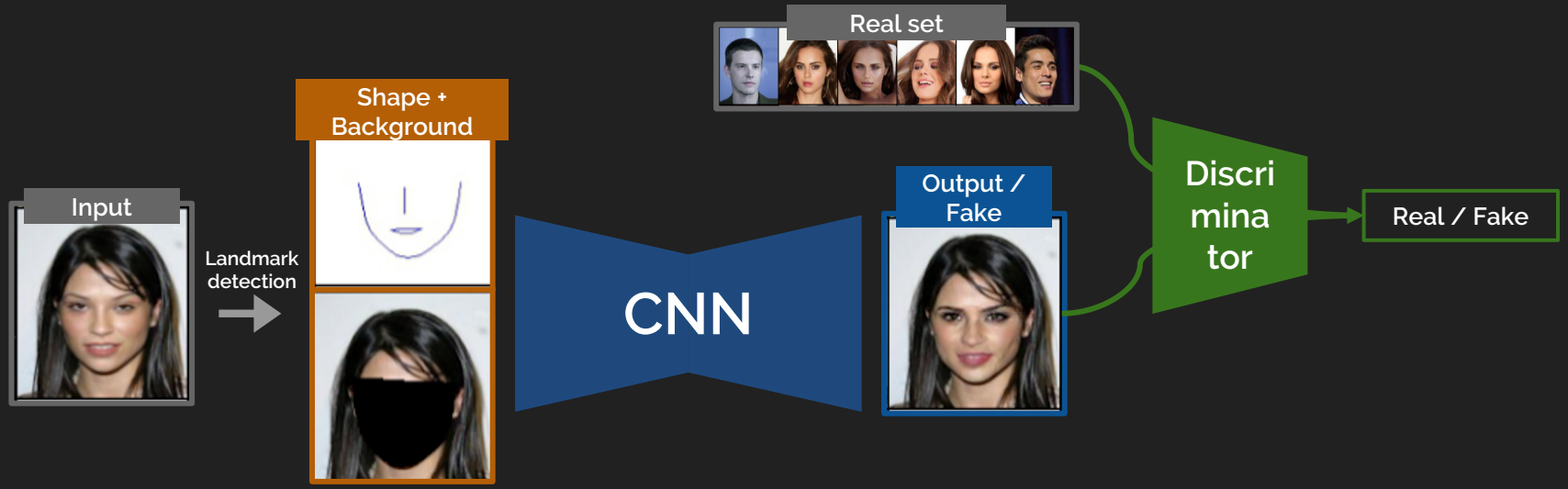


Inputs

- Partial Landmarks
  - We do not want appearance of the input face to “leak” to the new face
  - Mouth for expressions
  - Nose & Frame for orientation
  - “Free” temporal consistency
- Background Image
  - From Landmarks
  - For better blending of the face with the head and hair

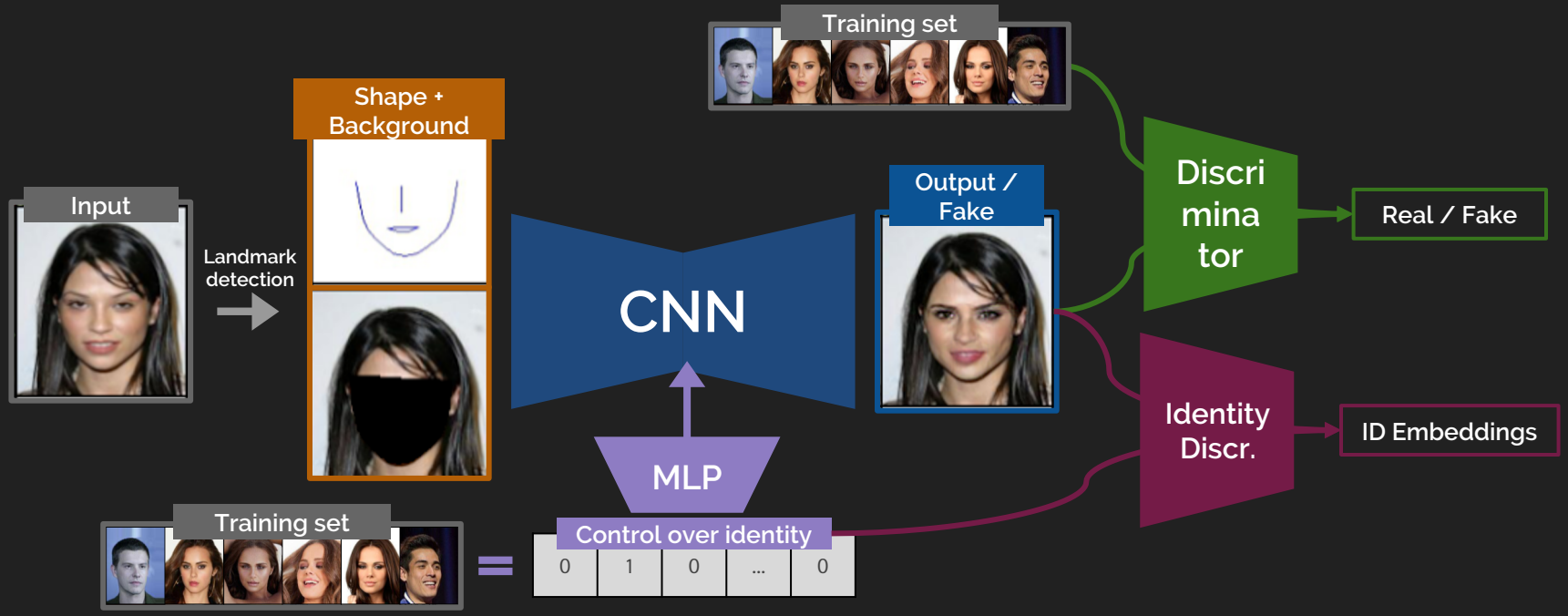


Losses 1: GAN Loss



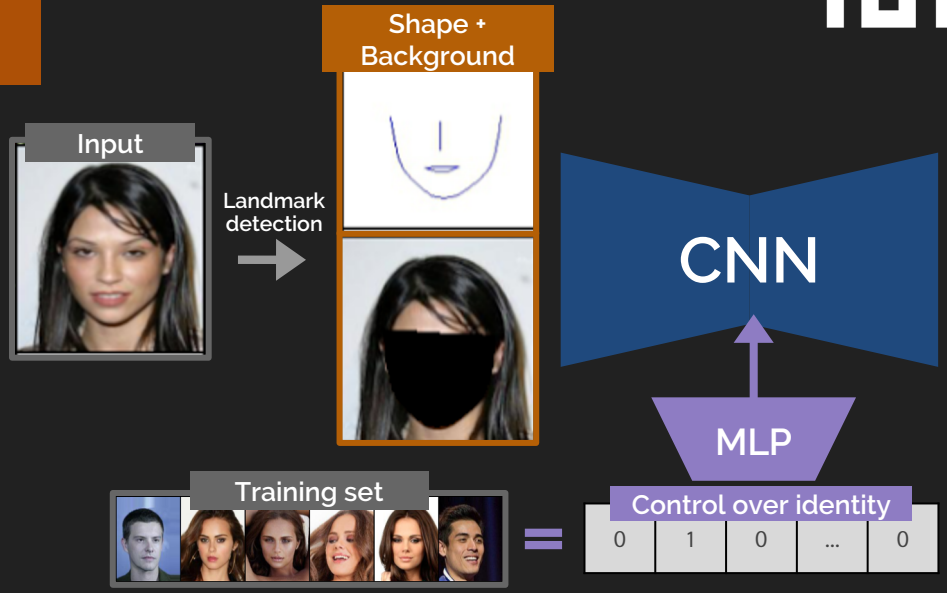
Without further losses, the network overfits and simply does reconstruction

Losses 2: ID Loss



Identity Guidance

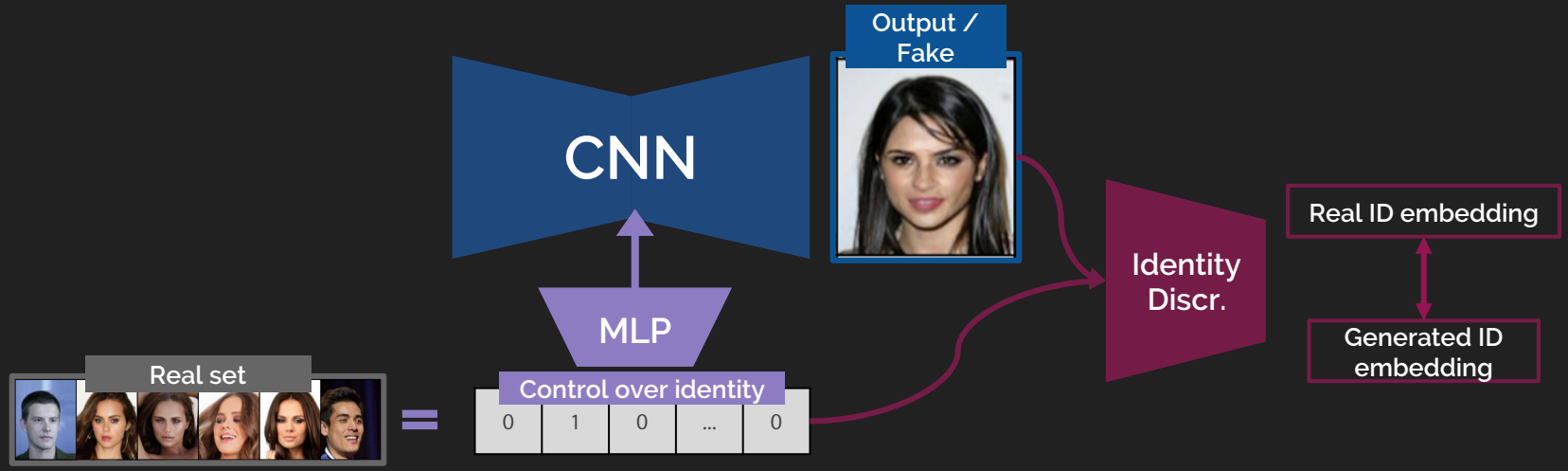
- Input:
  - One-hot vector encoding of a random ID of the training set
  - We pass it through an MLP and obtain a representation which is then concatenated at the bottleneck of the CNN
- Decoder:
  - Effectively uses the encoded information of the initial ID and mixes it with one of the random training IDs



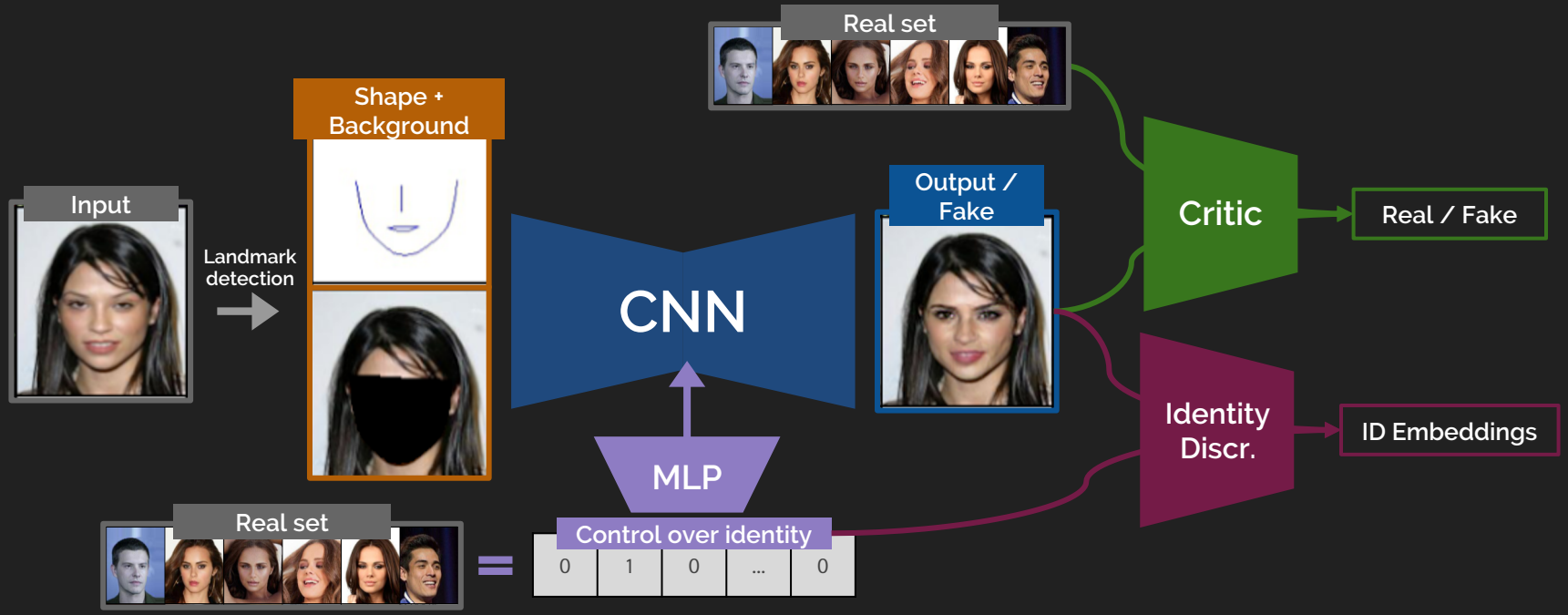
In how many ways can we anonymize an image?

Identity Discriminator

- Identity Discriminator
  - Pre-train for re-ID on real images with Proxy-NCA loss
  - Contrastive loss during GAN training: brings the embedding of the new ID closer to the real training ID embedding



Summary of CIAGAN



The identity discriminator is not used as *adversarial*, is it a *guidance* for the generator.

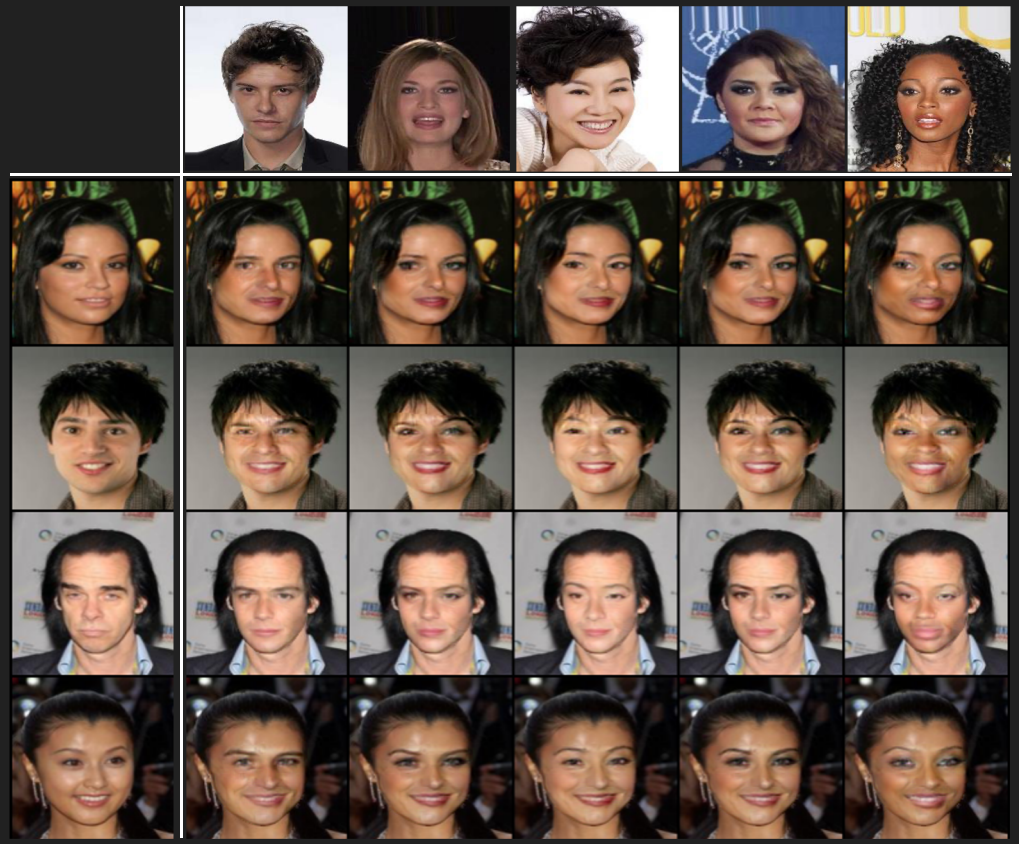
And for multi-object tracking?

- At each frame of a video:
  - We apply the *same transformation* to all pedestrians, so that we can perform tracking across frames.

- For a different camera
  - We apply the *a different transformation* to avoid long-term tracking and potential misuse of the data.

Results

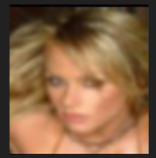
Qualitative results



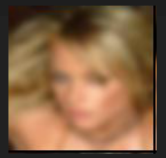
Detection & Identification

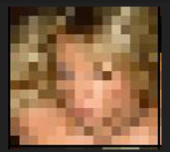
- Detection and identification on the CelebA dataset

Models	Detection (↑)		Identification (↓)	
	Dlib	SSH	PNCA	FaceNet
Original	100	100	70.7	65.1
Pixelization 16 by 16	0.0	0.0	**0.3**	**0.3**
Pixelization 8 by 8	0.0	0.0	0.4	**0.3**
Blur 9 by 9	90.6	38.6	16.9	57.2
Blur 17 by 17	68.4	0.3	1.9	0.5
Ours	**99.9**	**98.7**	1.3	1.0

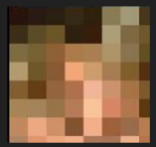


Blurring





Pixelization



Ablation studies

Face detection

Identification

Visual quality

Models	Detection (↑)	Recall@1 (↓)	FID (↓)
Siamese	**99.9**	1.3	**2.1**

Ablation studies

Face detection

Identification

Visual quality

Models	Detection (↑)	Recall@1 (↓)	FID (↓)
Siamese	**99.9**	1.3	**2.1**
Classification	64.6	**0.4**	63.2

- Classification of the Identity instead of Siamese training:
  - Identity recall goes down, mostly because the generated faces start to have artifacts → low detection rate and poor visual quality

Ablation studies

Face detection

Identification

Visual quality

Models	Detection (↑)	Recall@1 (↓)	FID (↓)
Siamese	**99.9**	1.3	**2.1**
Classification	64.6	**0.4**	63.2
Faces	98.3	1.1	6.5

- Input are full face images instead of landmarks.
  - Visual quality of the generated faces and detectability both decrease

Comparison with SOA

Two methods for face identification

De-ID method	VGGFace2 (↓)	CASIA (↓)
Original	0.986 ±0.010	0.965 ±0.016
Gafni et al.	0.038 ±0.015	0.035 ±0.011
Ours	**0.029 ± 0.012**	**0.026 ± 0.015**

- We are able to mask identities better
  - While also providing more diversity in the output and more control

Comparison with SOA

Source



Anonymization variations

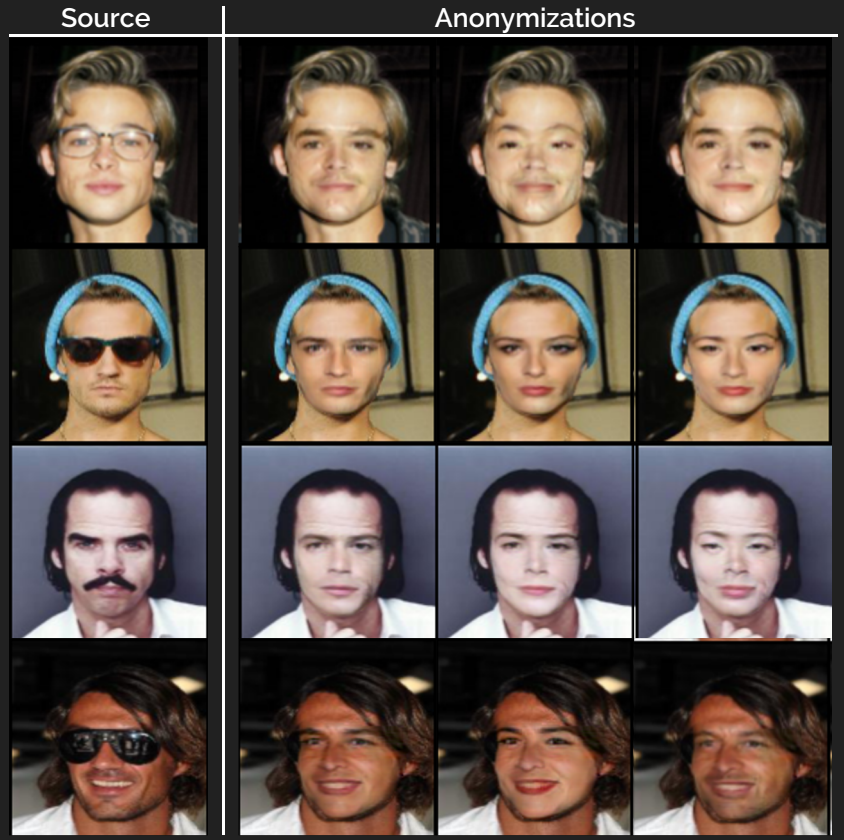
Gafni et al

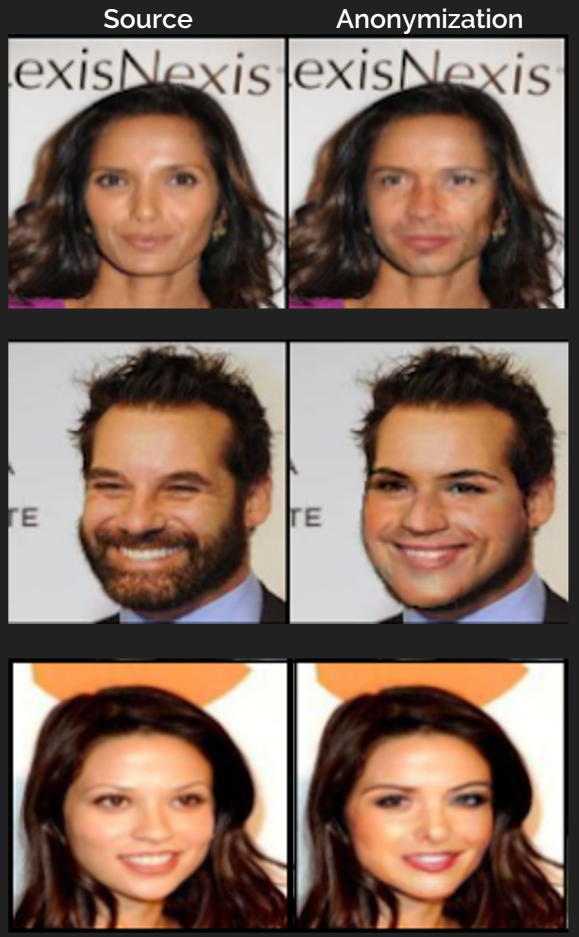
Ours



- We are able to mask identities better
  - While also providing more diversity in the output and more control

Glasses & Hair & Makeup



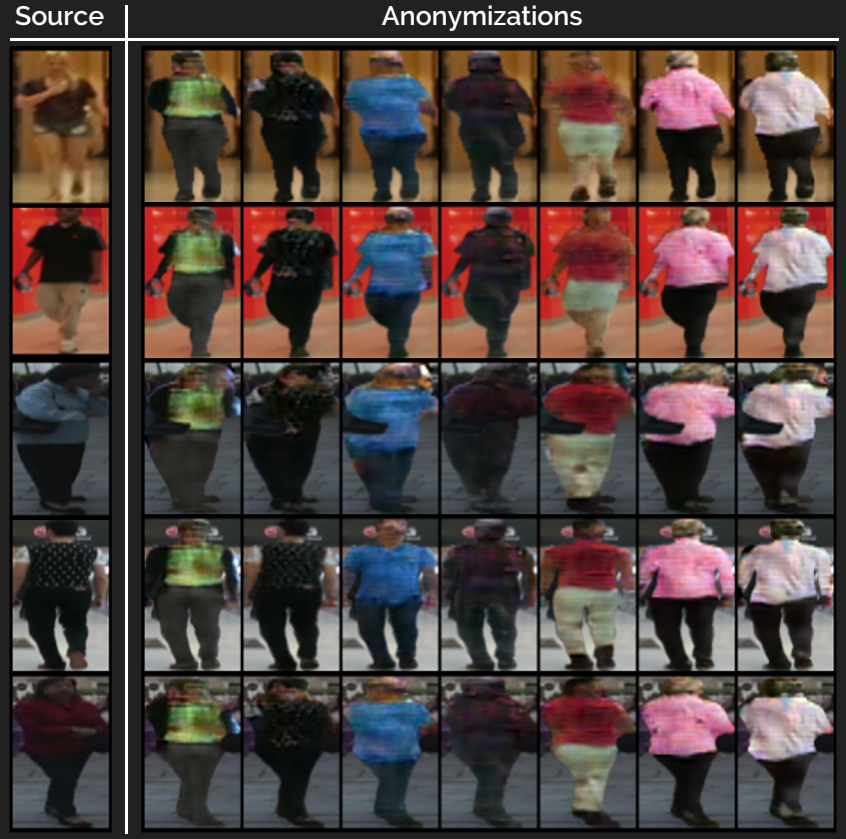


Results

Source Anonymizations



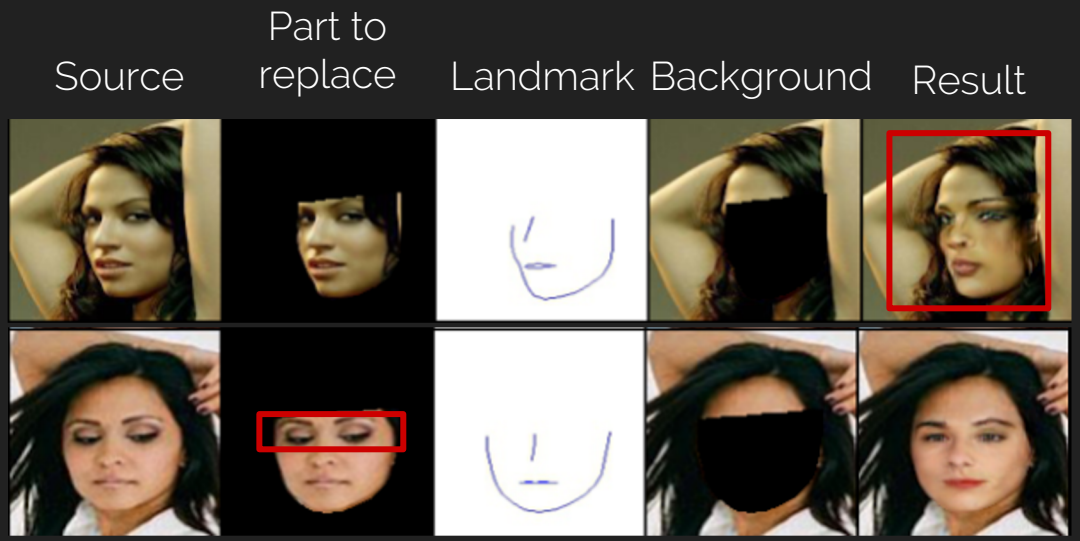
Different Domain



Video results



Limitations



Future Work

- Occlusions
- Different Domains
- Study the effect on multiple object tracking
- Do not depend on the output of the landmarks
- More realistic and high-definition images
- Work on explicit temporal consistency

The Team



Maxim Maximov



Ismail Elezi



Laura Leal-Taixé



Thank you

Prof. Dr. Laura Leal-Taixé

Technical University of Munich

"All human beings have three lives: public, private, and secret."

Gabriel García Márquez