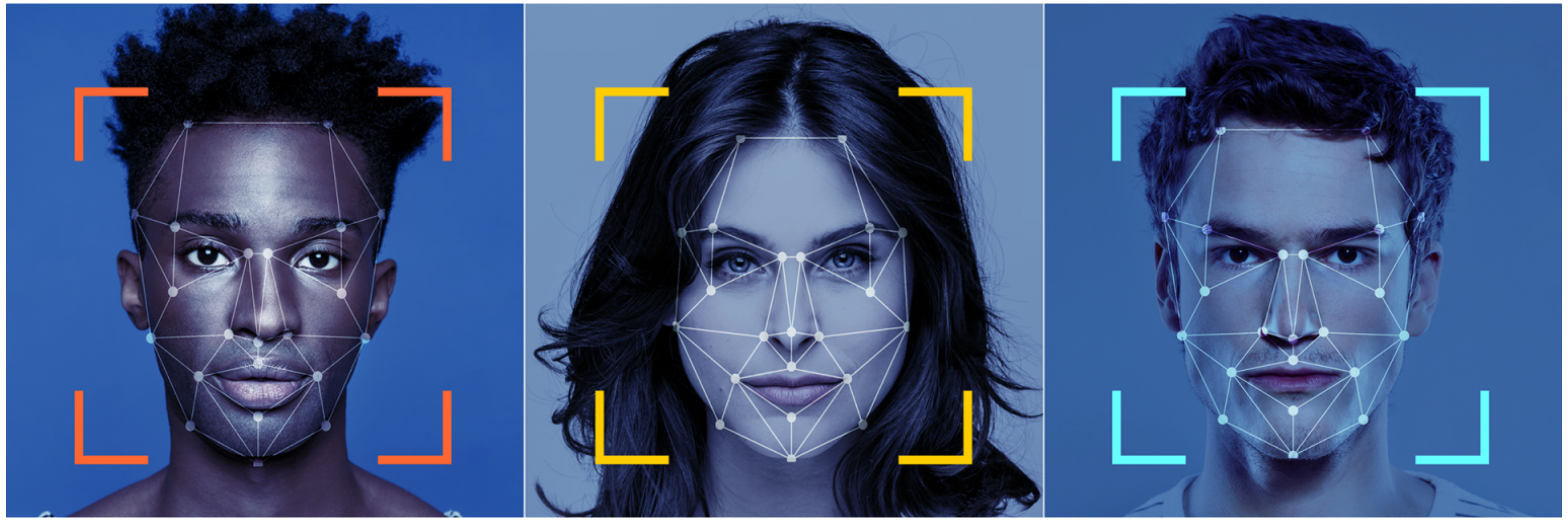
Designing Generative Adversarial Networks for Privacy-enhanced Face Recognition

Sebastian Raschka

Asst. Prof. of Statistics
@ U. of Wisconsin-Madison

https://sebastianraschka.com



14th International (Virtual) Conference on Human System Interaction (HSI 2021)
Gdansk University of Technology, Poland – July 08-10, 2021

http://hsi2021.welcometohsi.org

Topics

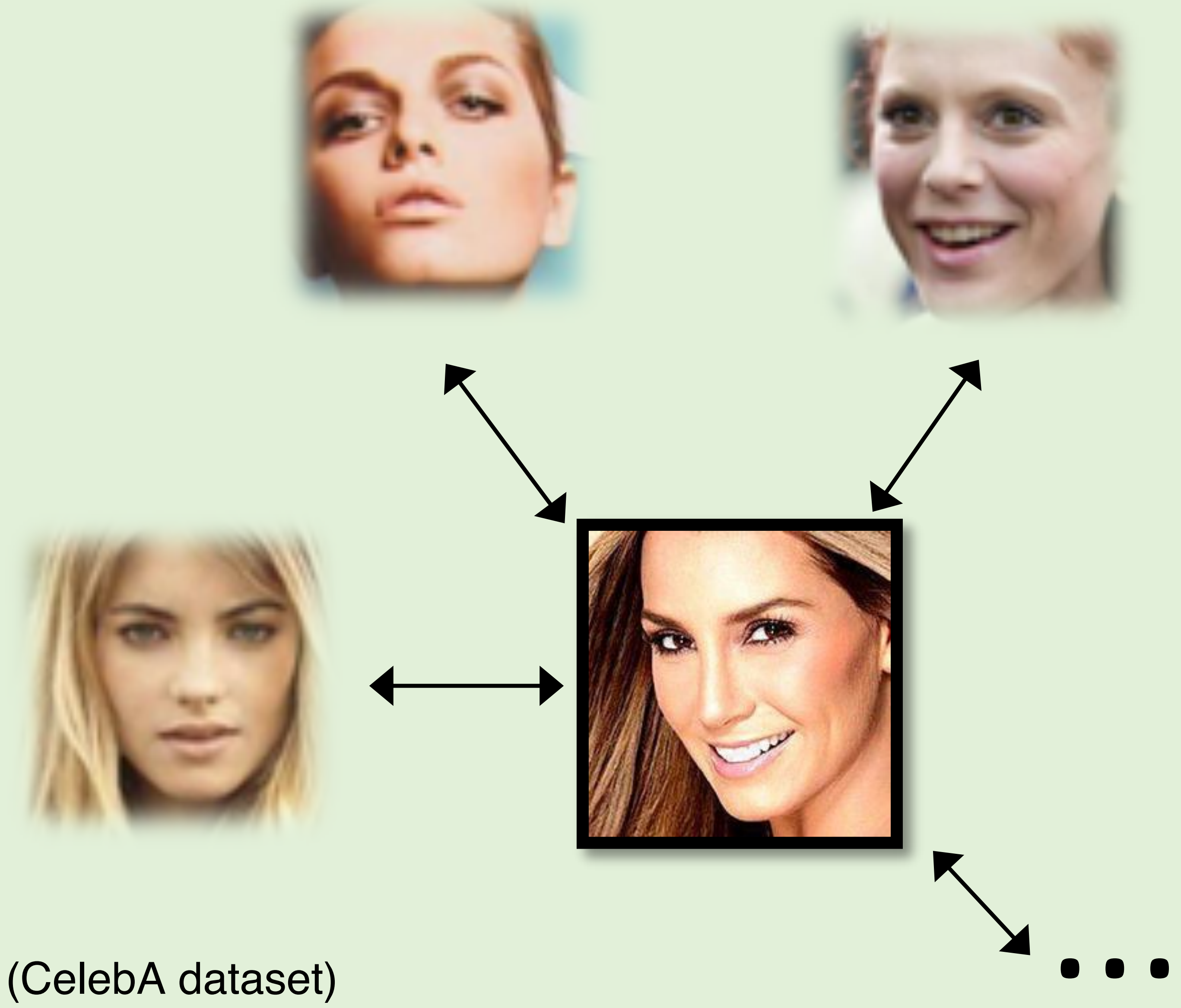
1. **Biometric Face Recognition**
2. Extracting Soft-Biometric Attributes from Face Images
3. Hiding Soft-Biometric Attributes in Face Images
4. PrivacyNet: GAN-based Multi-attribute Face Privacy

Biometric (Face) Recognition

A. Identification

Determine identity of an unknown person

1-to-*n* matching



(CelebA dataset)

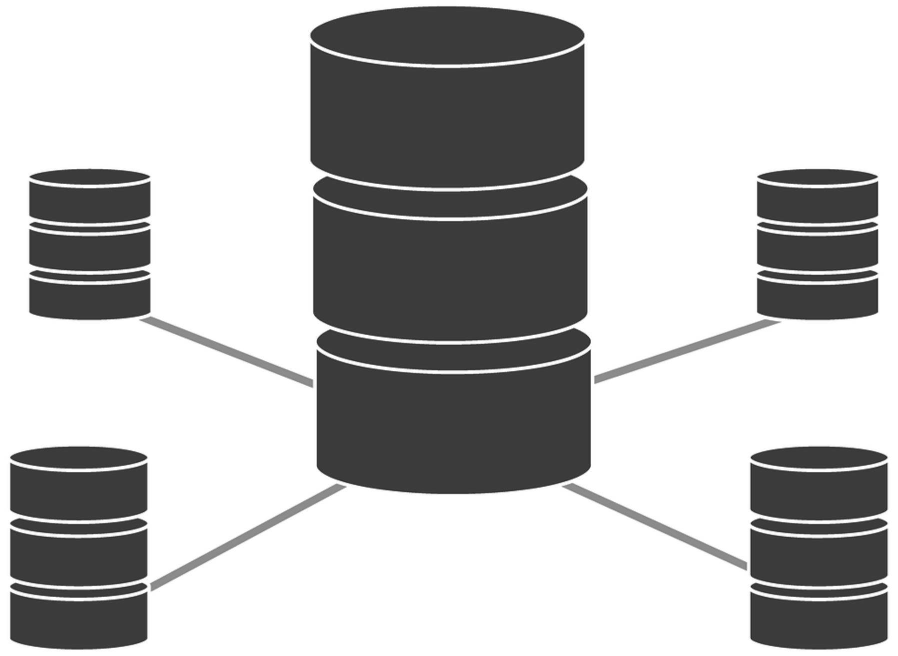
B. Verification

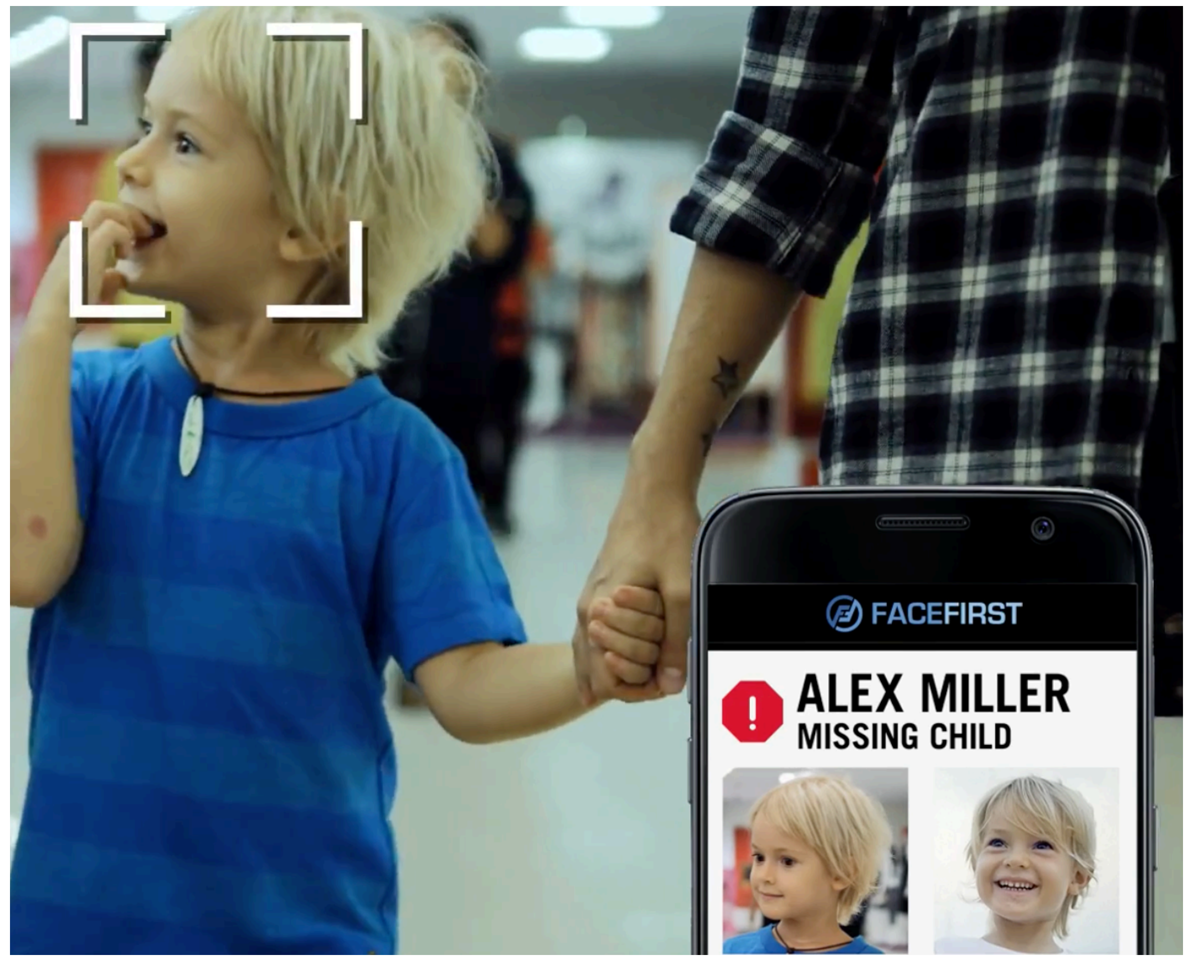
Verify claimed identity of a person

1-to-1 matching

(MUCT dataset)

Applications of Biometric (Face) Recognition





Soft-Biometric Attributes

Identity	Meryl Streep
Gender	Female
Age	72
Race	Caucasian
Medical	Healthy

Topics

1. Biometric Face Recognition

2. Extracting Soft-Biometric Attributes from Face Images

3. Hiding Soft-Biometric Attributes in Face Images

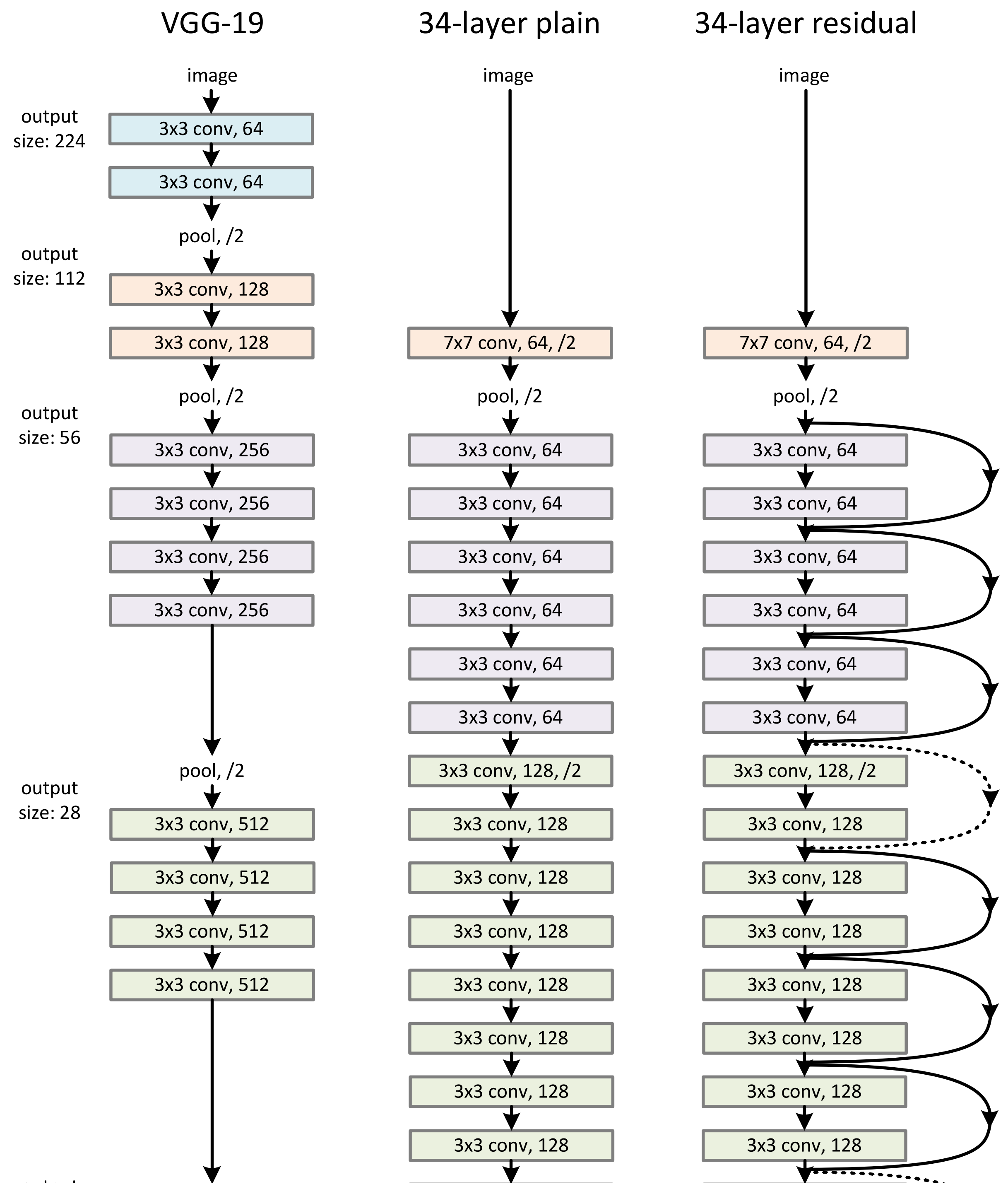
4. PrivacyNet: GAN-based Multi-attribute Face Privacy

Ex. 1: How difficult is it to extract **gender** information from face images?

Identity	Meryl Streep
Gender	Female
Age	72
Race	Caucasian
Medical	Healthy

Very Easy:

ResNet-50 Applied to Gender Classification



He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. **2016.**

```
Epoch: 010/010 | Batch 0000/0318 | Cost: 0.0238
Epoch: 010/010 | Batch 0050/0318 | Cost: 0.0251
Epoch: 010/010 | Batch 0100/0318 | Cost: 0.0144
Epoch: 010/010 | Batch 0150/0318 | Cost: 0.0133
Epoch: 010/010 | Batch 0200/0318 | Cost: 0.0441
Epoch: 010/010 | Batch 0250/0318 | Cost: 0.0358
Epoch: 010/010 | Batch 0300/0318 | Cost: 0.0277
Epoch: 010/010 | Train: 99.374% | Valid: 97.966%
Time elapsed: 37.70 min
Total Training Time: 37.70 min
```

Evaluation

```
with torch.set_grad_enabled(False): # save memory during inference
    print('Test accuracy: %.2f%%' % (compute_accuracy(model, test_loader,
```

```
Test accuracy: 97.40%
```

https://nbviewer.jupyter.org/github/rasbt/deeplearning-models/blob/master/pytorch_ipynb/cnn/cnn-resnet50-celeba-dataparallel.ipynb

Ex. 2: How difficult is it to extract **age** information from face images?

Identity	Meryl Streep
Gender	Female
Age	72
Race	Caucasian
Medical	Healthy

Ordinal Regression for ordinal data: integrating label order info

- **Ranking:** Predict Correct order
  (0 loss if order is correct, e.g., rank a collection of movies by "goodness")

$\prec$

$\prec$

- **Ordinal regression:** Predict correct (ordered) label
  (E.g., age of a person in years; here, regard aging as a non-stationary process)

Excerpt from the UTKFace dataset
https://susanqq.github.io/UTKFace/

18

$\prec$

29

$\prec$

41

Cao, Mirjalili, Raschka (2020)
Rank Consistent Ordinal Regression for Neural Networks with Application to Age Estimation
Pattern Recognition Letters. 140, 325-331

Age Prediction Datasets

MORPH-2

- 55,608 face images
- age range: 16-70 years

AFAD

- 165,501 face images
- age range: 15-40 years

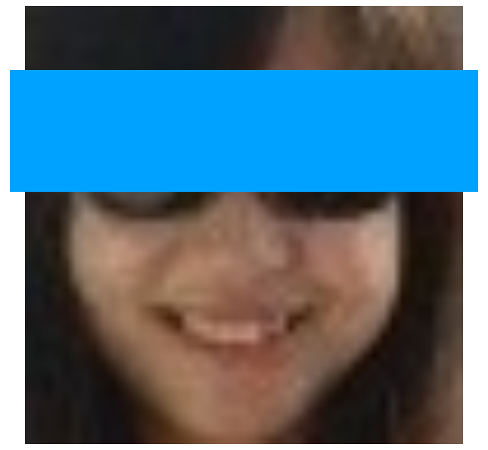


Table 1
Age prediction errors on the test sets. All models are based on the ResNet-34 architecture.

Method	Random seed	MORPH-2		AFAD	
		MAE	RMSE	MAE	RMSE
CE-CNN	0	3.26	4.62	3.58	5.01
	1	3.36	4.77	3.58	5.01
	2	3.39	4.84	3.62	5.06
	AVG ± SD	3.34 ± 0.07	4.74 ± 0.11	3.60 ± 0.02	5.03 ± 0.03
OR-CNN [16]	0	2.87	4.08	3.56	4.80
	1	2.81	3.97	3.48	4.68
	2	2.82	3.87	3.50	4.78
	AVG ± SD	2.83 ± 0.03	3.97 ± 0.11	3.51 ± 0.04	4.75 ± 0.06
CORAL-CNN (ours)	0	2.66	3.69	3.42	4.65
	1	2.64	3.64	3.51	4.76
	2	2.62	3.62	3.48	4.73
	AVG ± SD	**2.64** ± **0.02**	**3.65** ± **0.04**	**3.47** ± **0.05**	**4.71** ± **0.06**

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N} |y_i - h(\mathbf{x}_i)|$$

Age prediction only off by 2 ½ to 3 ½ years on average

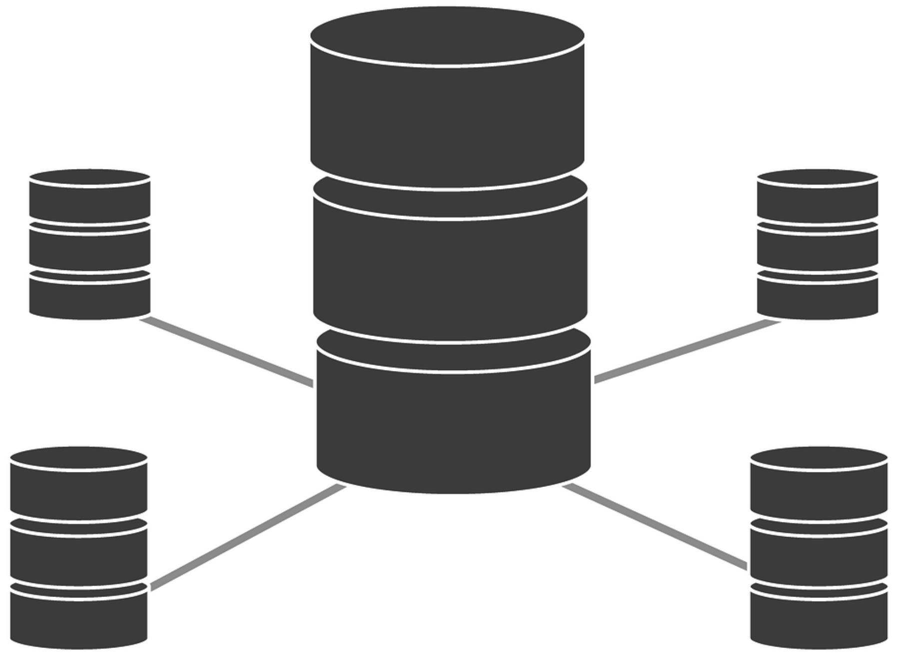
W Cao, V Mirjalili, and S Raschka (2020)
Rank Consistent Ordinal Regression for Neural Networks with Application to Age Estimation
Pattern Recognition Letters. 140, 325-331
https://www.sciencedirect.com/science/article/pii/S016786552030413X

Topics

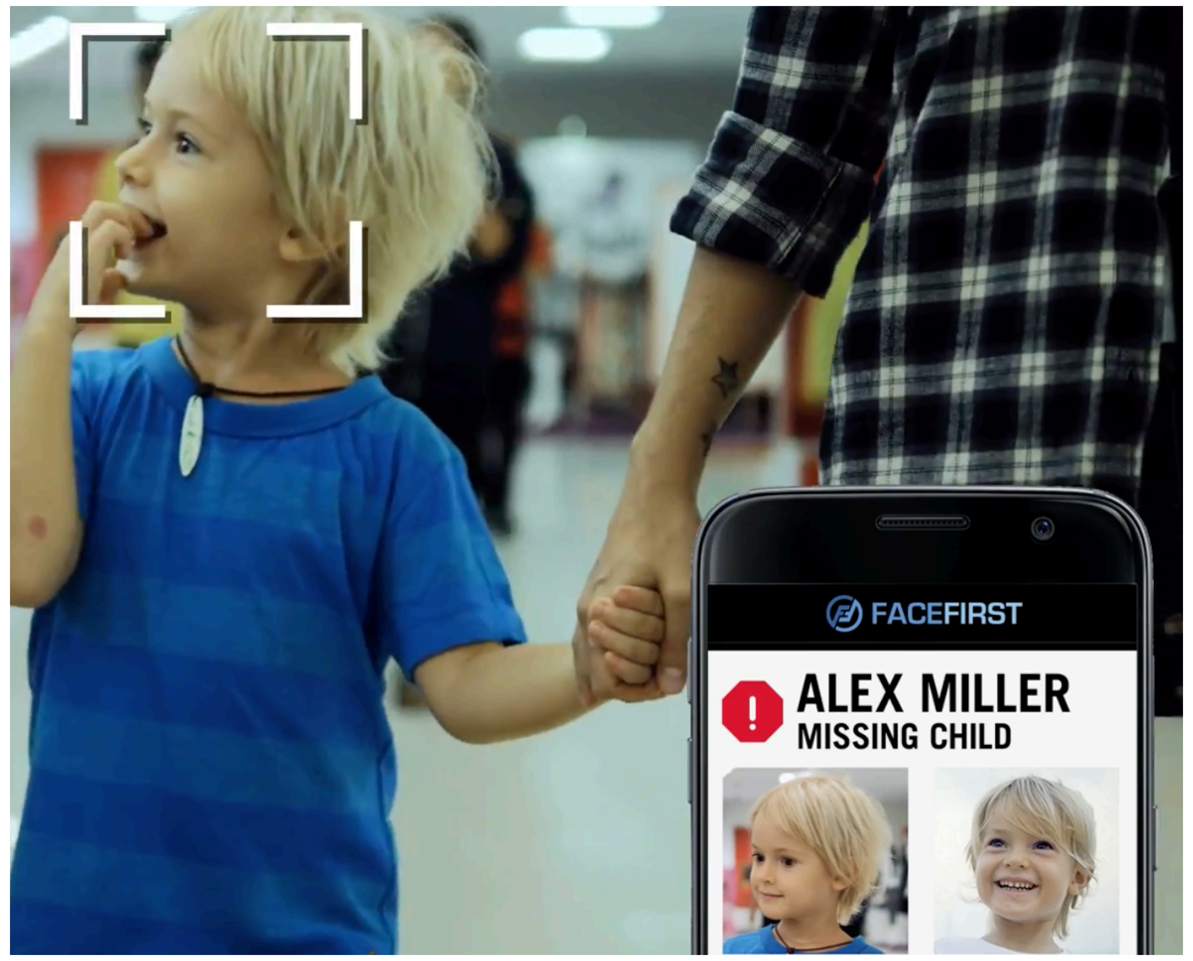
1. Biometric Face Recognition
2. Extracting Soft-Biometric Attributes from Face Images
3. **Hiding Soft-Biometric Attributes in Face Images**
4. PrivacyNet: GAN-based Multi-attribute Face Privacy

Biometric (Face) Recognition Can Be Useful

https://whyarg.com/wp-content/uploads/2017/06/videosurveillance.jpg



https://www.secureidnews.com/wp-content/uploads/2013/03/3m_autogate-300x259.jpg



Soft-Biometric Attribute Mining Can Be Problematic in Absence of Consent

Identity	Meryl Streep
Gender	Female
Age	72
Race	Caucasian
Medical	Healthy

Soft-biometric Attributes: Issues and Concerns

1. **Identity theft:** combining soft biometric info with publicly available data
2. **Profiling**: e.g., gender/race based profiling
3. **Ethics:** extracting data without users' consent (e.g., intentional or via database breaches)

Preventing Automatic Extraction

mail _at_ sebastianraschka .dot. com

Can/do we need to take similar measures to prevent soft-biometric attribute harvesting?

One solution:
Storing face representation vectors with sensitive information removed

1. Q. Xie, Z. Dai, Y. Du, E. Hovy, and G. Neubig: "Controllable invariance through adversarial feature learning," in Advances in Neural Information Processing Systems, 2017, pp. 585–596.

2. P. Terhorst, N. Damer, F. Kirchbuchner, and A. Kuijper, "Unsupervised privacy-enhancement of face representations using similarity-sensitive noise transformations," Applied Intelligence, pp. 1–18, 2019.

3. A. Morales, J. Fierrez, and R. Vera-Rodriguez, "SensitiveNets: Learning agnostic representations with application to face recognition," arXiv preprint arXiv:1902.00334,

4. P. C. Roy and V. N. Boddeti, "Mitigating information leakage in image representations: A maximum entropy approach," in IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 2586–2594.

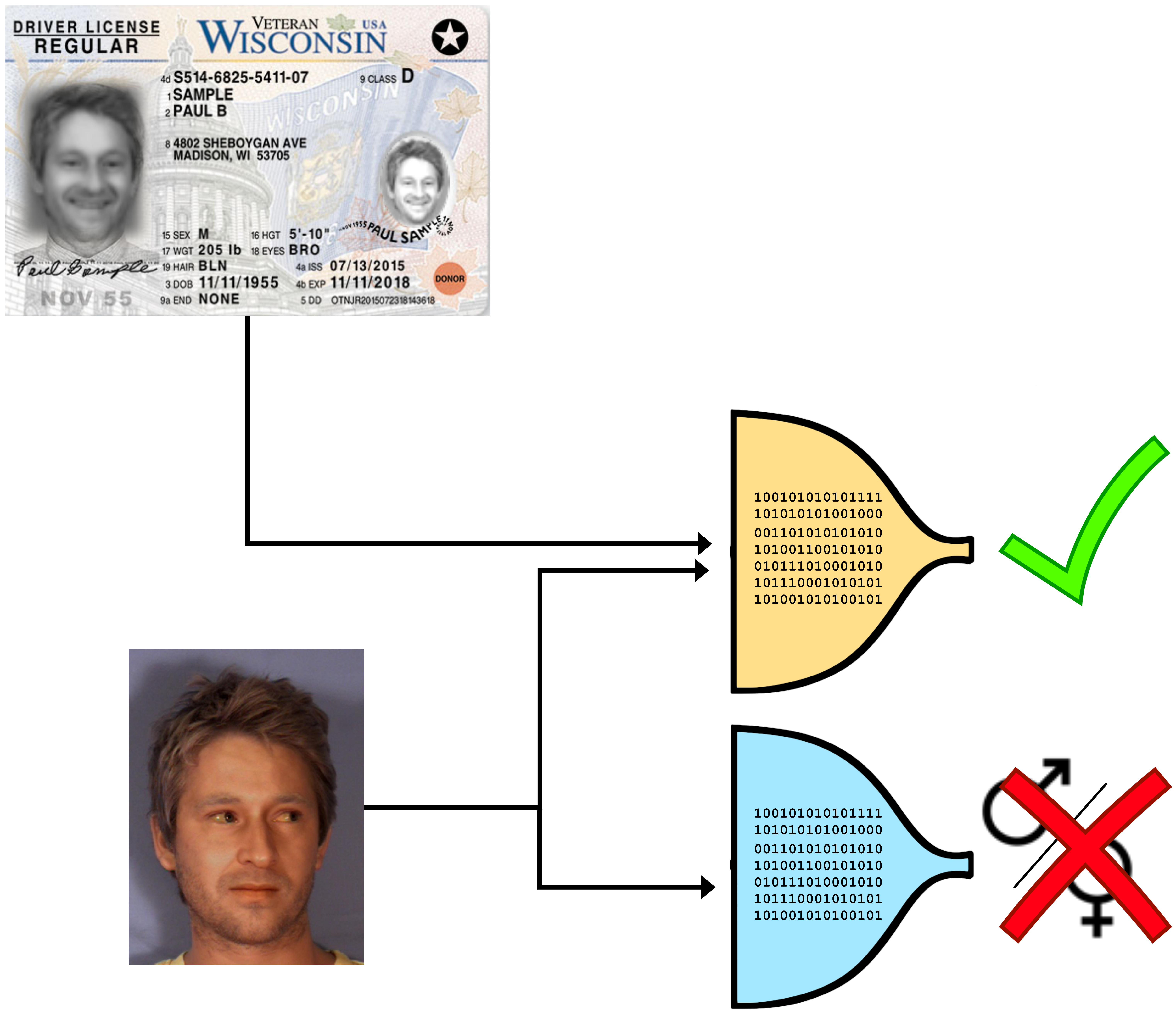
5. B. Sadeghi, R. Yu, and V. Boddeti, "On the global optima of kernelized adversarial representation learning," in Proceedings of the IEEE International Conference on Computer Vision, 2019, pp. 7971– 7979.

Very useful approach, but can have limitation for certain application domains, because

- not interpretable by humans
- not compatible with arbitrary face matching software

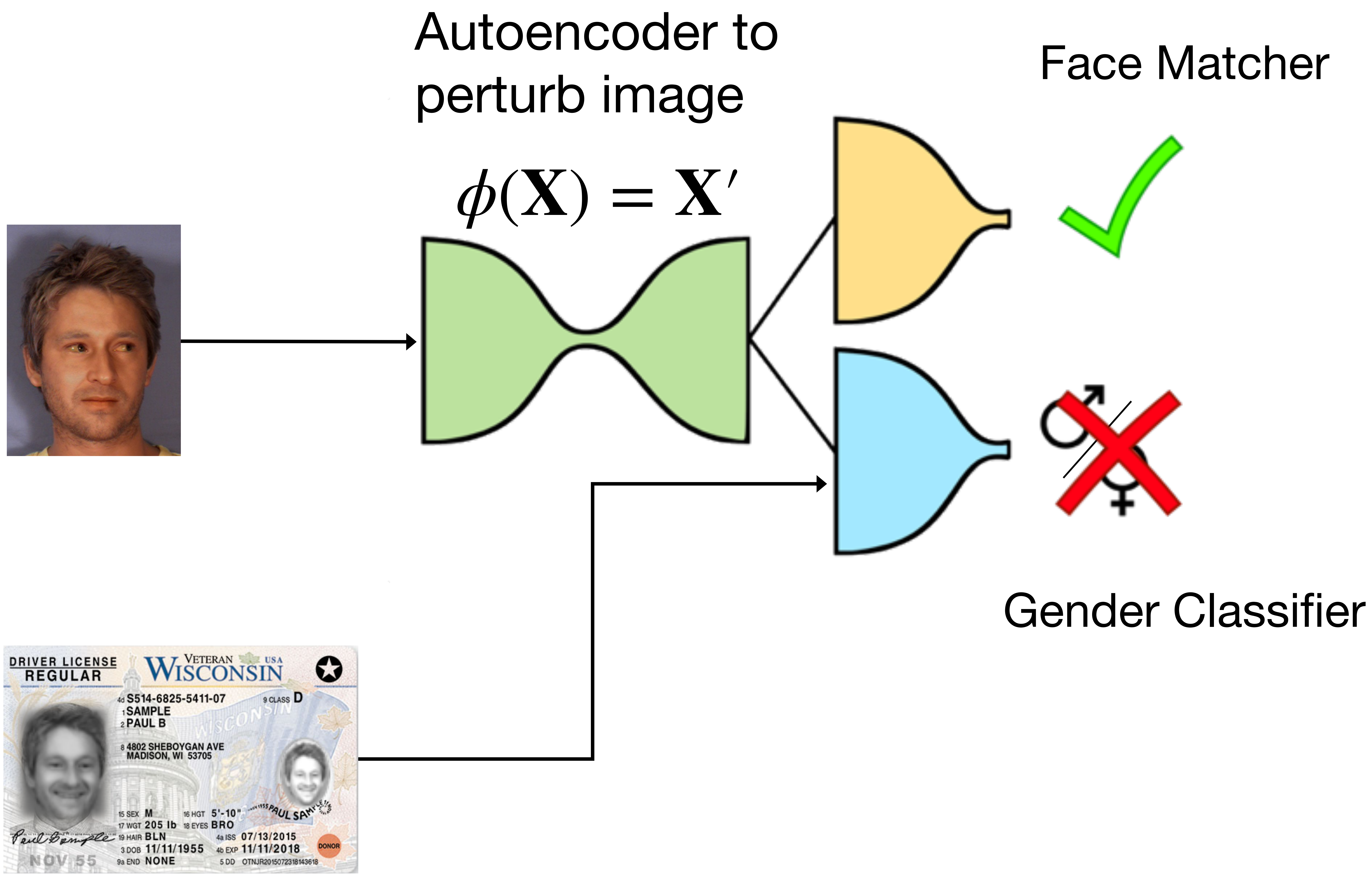
Goal: Selective Privacy

1. Perturb soft-biometric (e.g., gender) information
2. Ensure realistic face images
3. Retain biometric face recognition utility

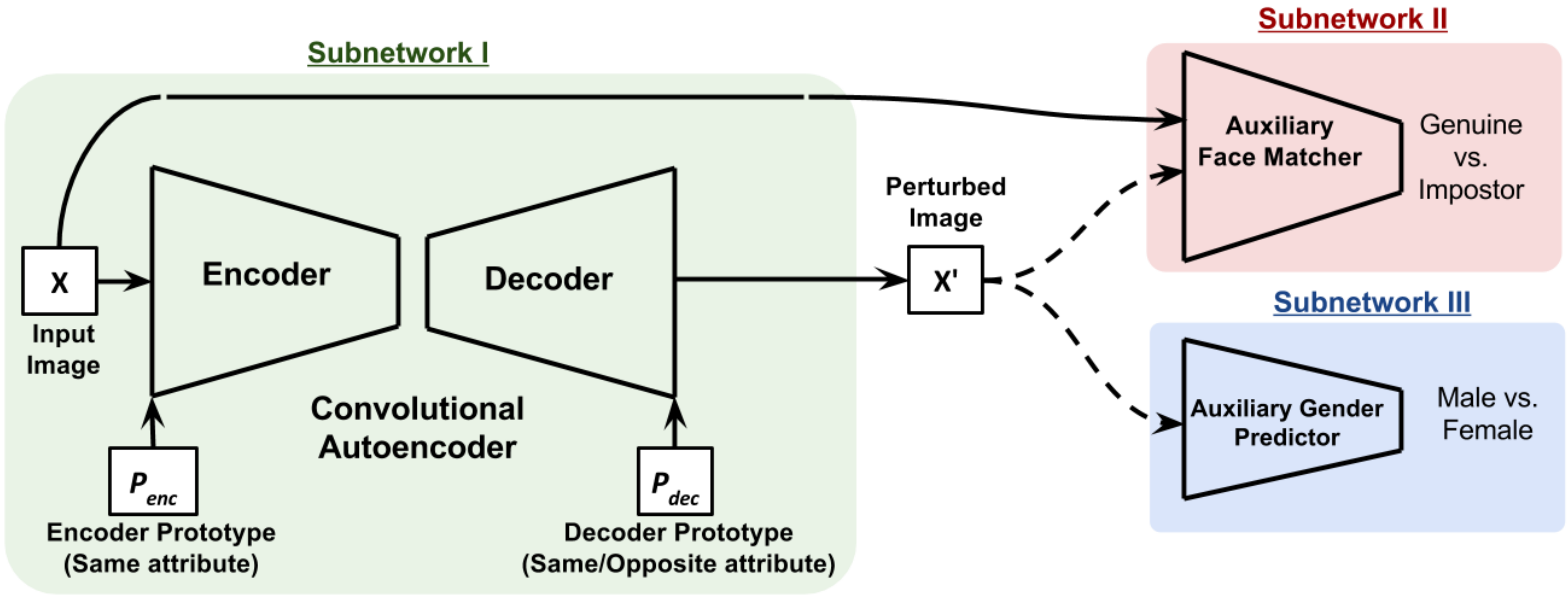


Face Matcher

Gender Classifier

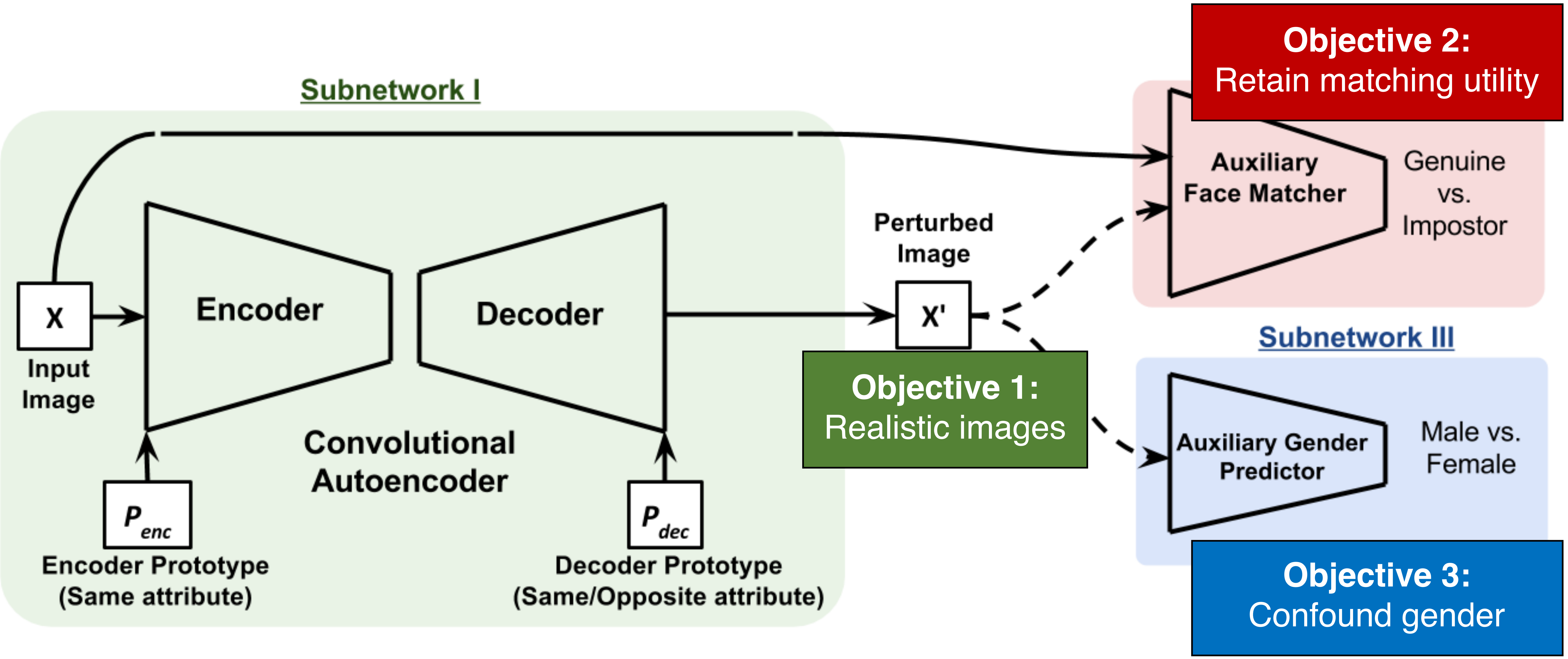

Autoencoder to
perturb image
$\phi(\mathbf{X}) = \mathbf{X}'$
Face Matcher
Gender Classifier
DRIVER LICENSE
REGULAR
VETERAN USA
WISCONSIN
4d S514-6825-5411-07
9 CLASS D
1 SAMPLE
2 PAUL B
8 4802 SHEBOYGAN AVE
MADISON, WI 53705
15 SEX M
16 HGT 5'-10"
17 WGT 205 lb
18 EYES BRO
19 HAIR BLN
4a ISS 07/13/2015
3 DOB 11/11/1955
4b EXP 11/11/2018
9a END NONE
5 DD OTNJR2015072318143618
DONOR
NOV 55

General architecture of the semi-adversarial network (SAN)

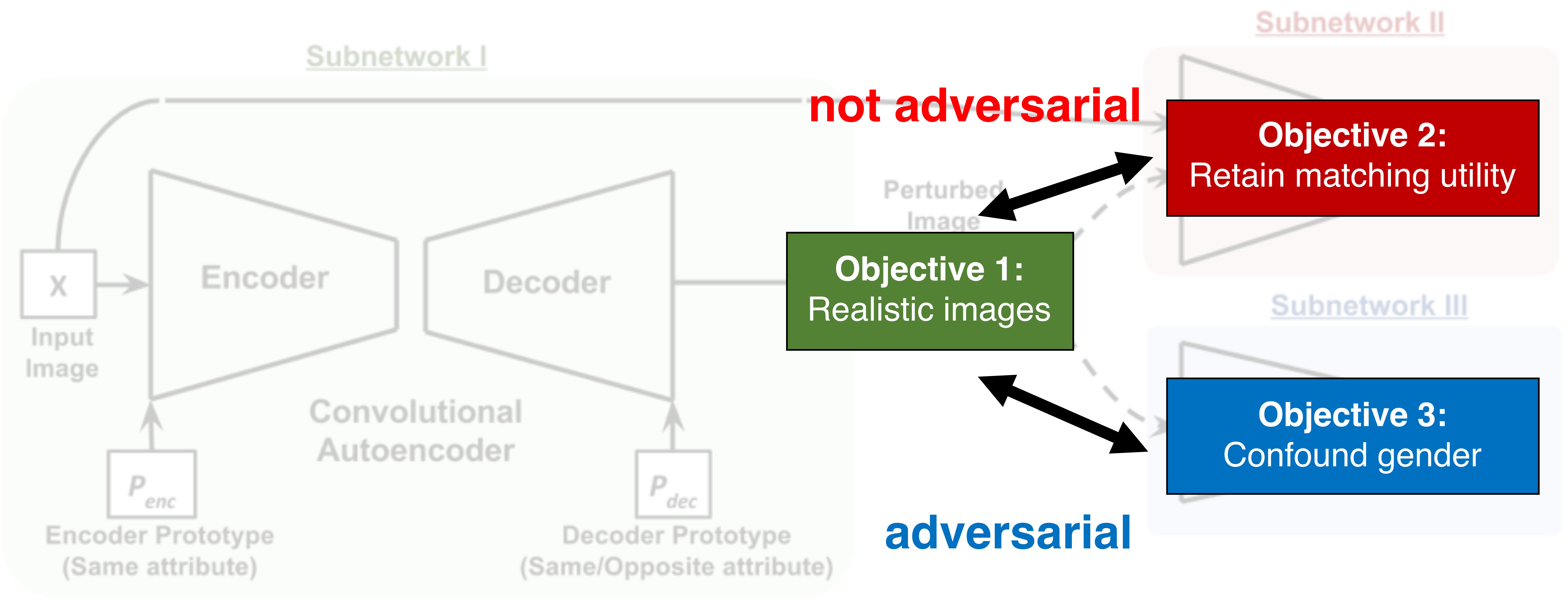


V Mirjalili, S Raschka, A Namboodiri, and A Ross (2018) *Semi-adversarial Networks: Convolutional Autoencoders for Imparting Privacy to Face Images.* Proc. of 11th IAPR International Conference on Biometrics (ICB 2018) https://ieeexplore.ieee.org/document/8411207/

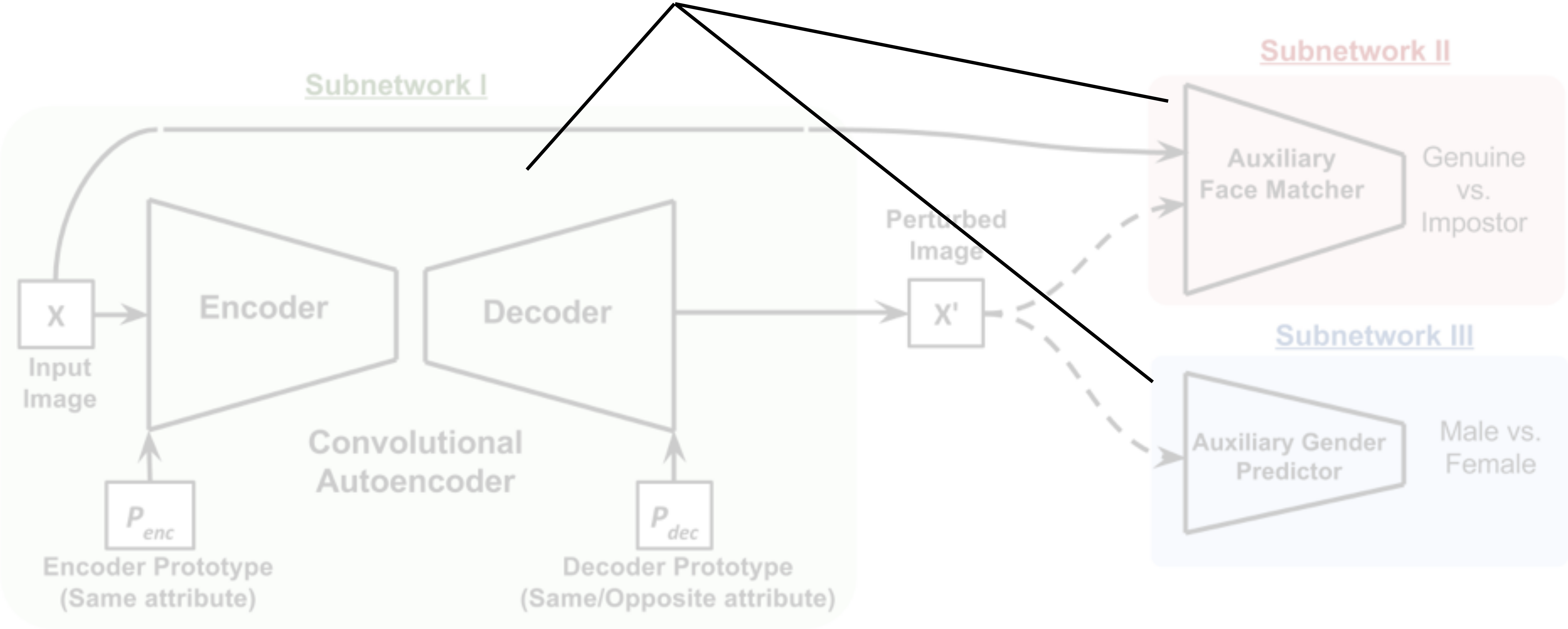
General architecture of the semi-adversarial network (SAN)



Semi-adversarial network



Convolutional neural networks



SAN Examples

Original Inputs


Male: 99%


Female: 98%


Male: 97%


Male: 100%

Outputs

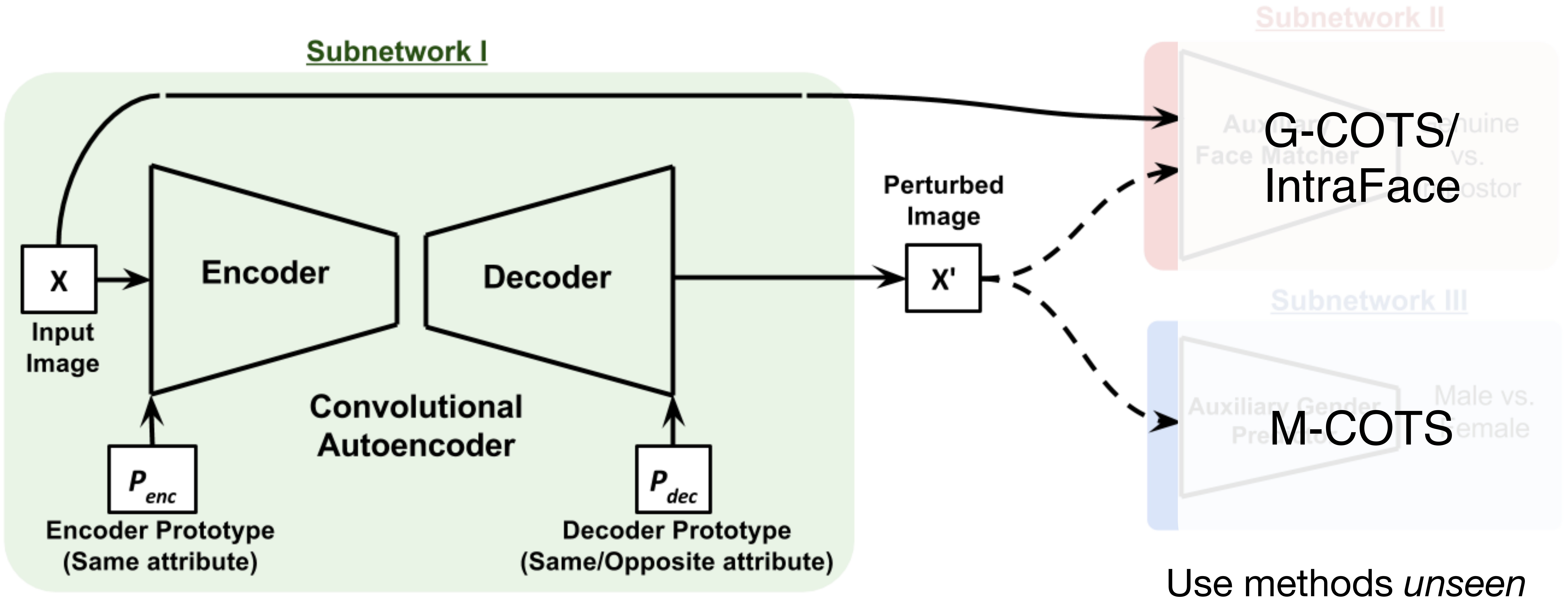

Female: 69%


Male: 99%


Female: 71%


Female: 58%

Replacing Detachable Parts for Evaluation



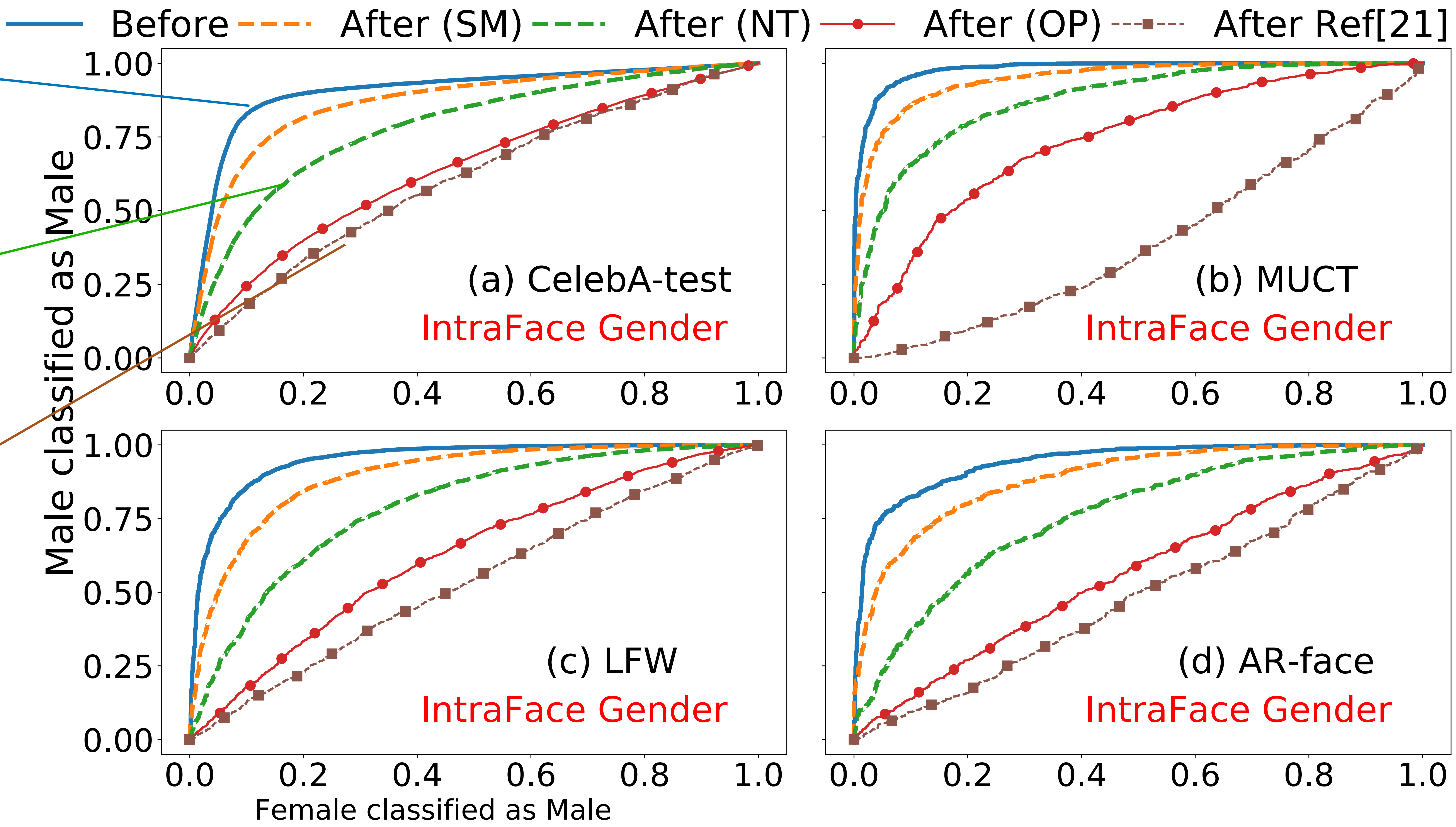
Use methods *unseen* during training

IntraFace Gender Classifier Performance on Different Datasets

No perturbation: Original gender classification performance

Strong perturbation via our SAN method

Very strong perturbation from Ref [21]



[21] A. Othman and A. Ross. Privacy of facial soft biometrics: Suppressing gender but retaining identity. In *European Conference on Computer Vision Workshop*, pages 682–696. Springer, 2014.

Face matching performance

Multi-subject comparisons

Before

After (SM)

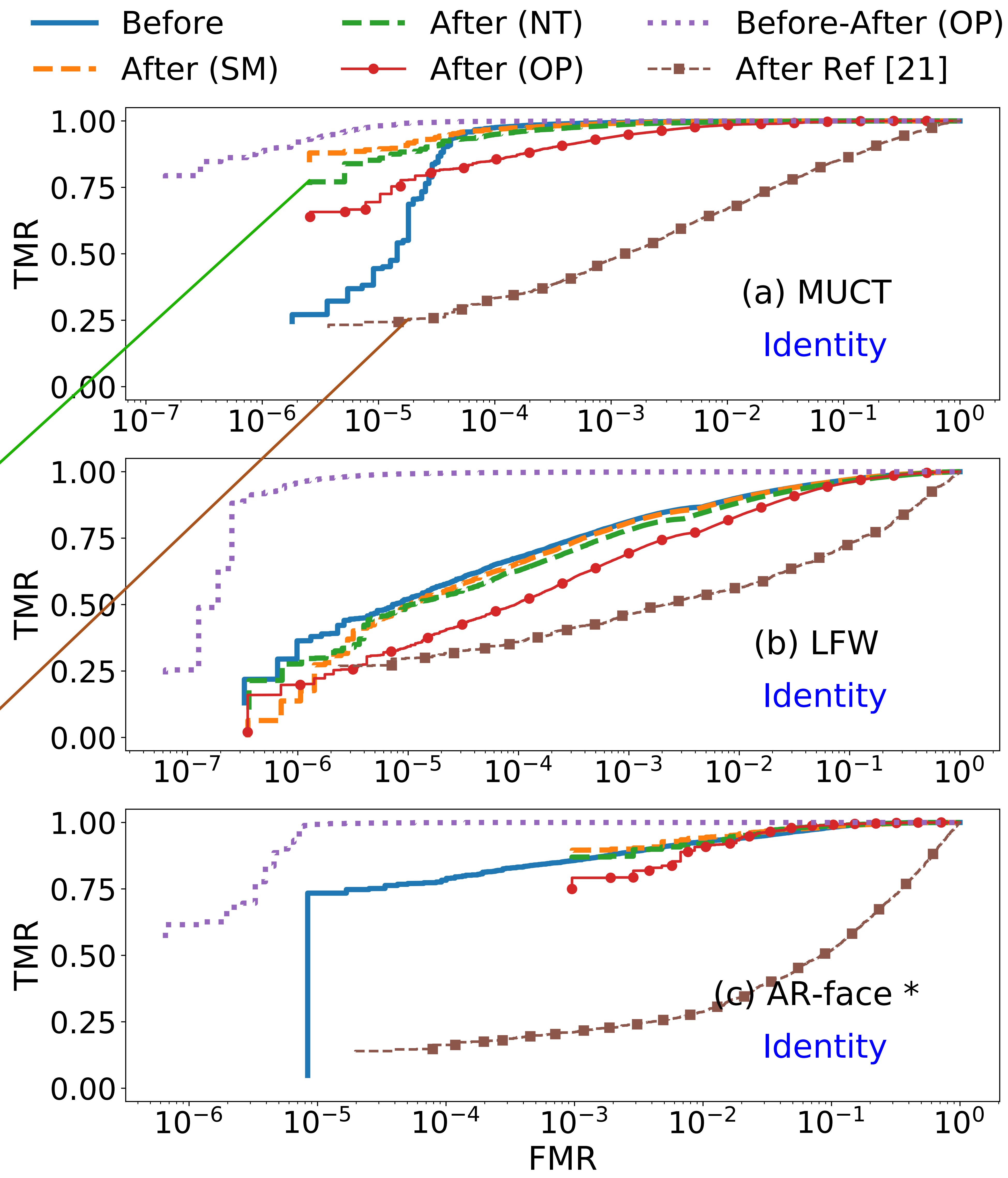
After (NT)

After (OP)

After Ref [1]

Strong perturbation
via our SAN method

Very very weak face recognition
after perturbation via Ref [21]



[21] A. Othman and A. Ross. Privacy of facial soft biometrics: Suppressing gender but retaining identity. In *European Conference on Computer Vision Workshop*, pages 682–696. Springer, 2014.

Gender Privacy: An Ensemble of Semi Adversarial Networks for Confounding Arbitrary Gender Classifiers

Improvements to construct a more diverse set of SAN models for better generalizability

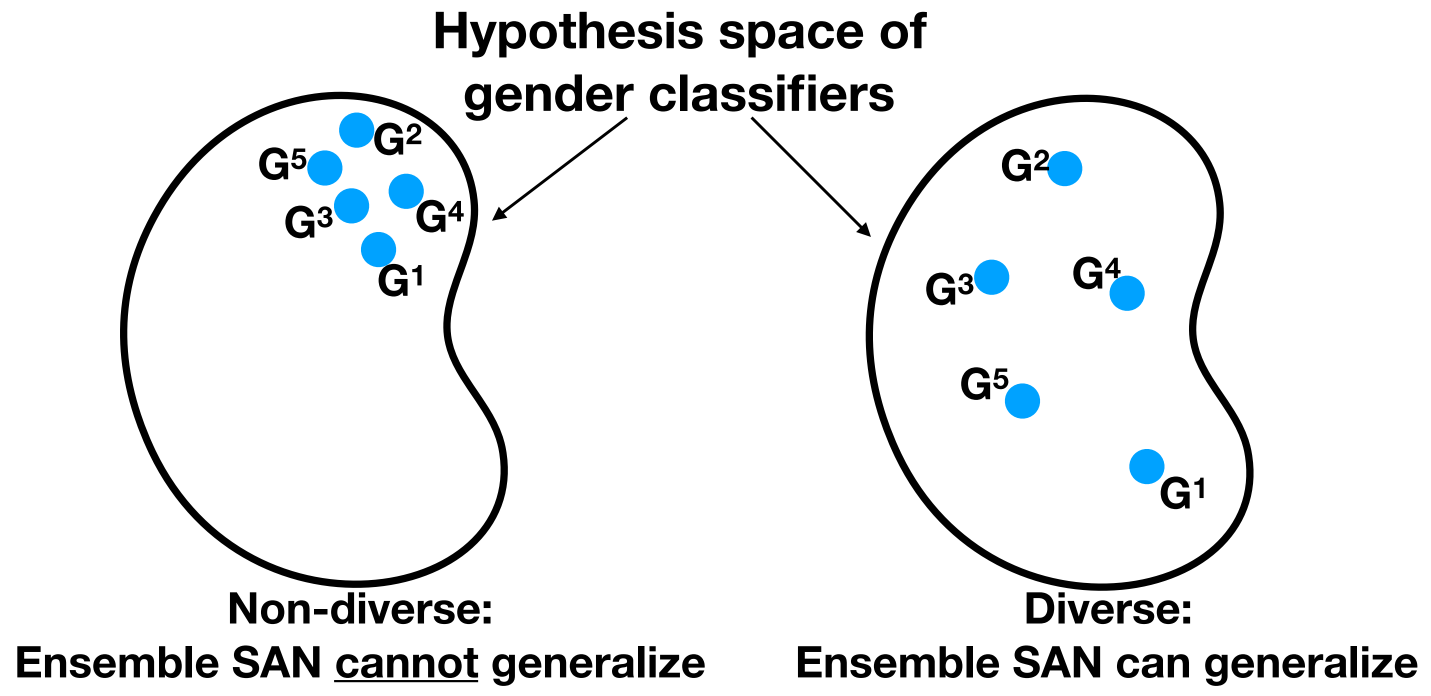


Figure 1: Diversity in an ensemble SAN can be enhanced through its auxiliary gender classifiers (see Figure 2). When the auxiliary gender classifiers lack diversity, ensemble SAN cannot generalize well to arbitrary gender classifiers.

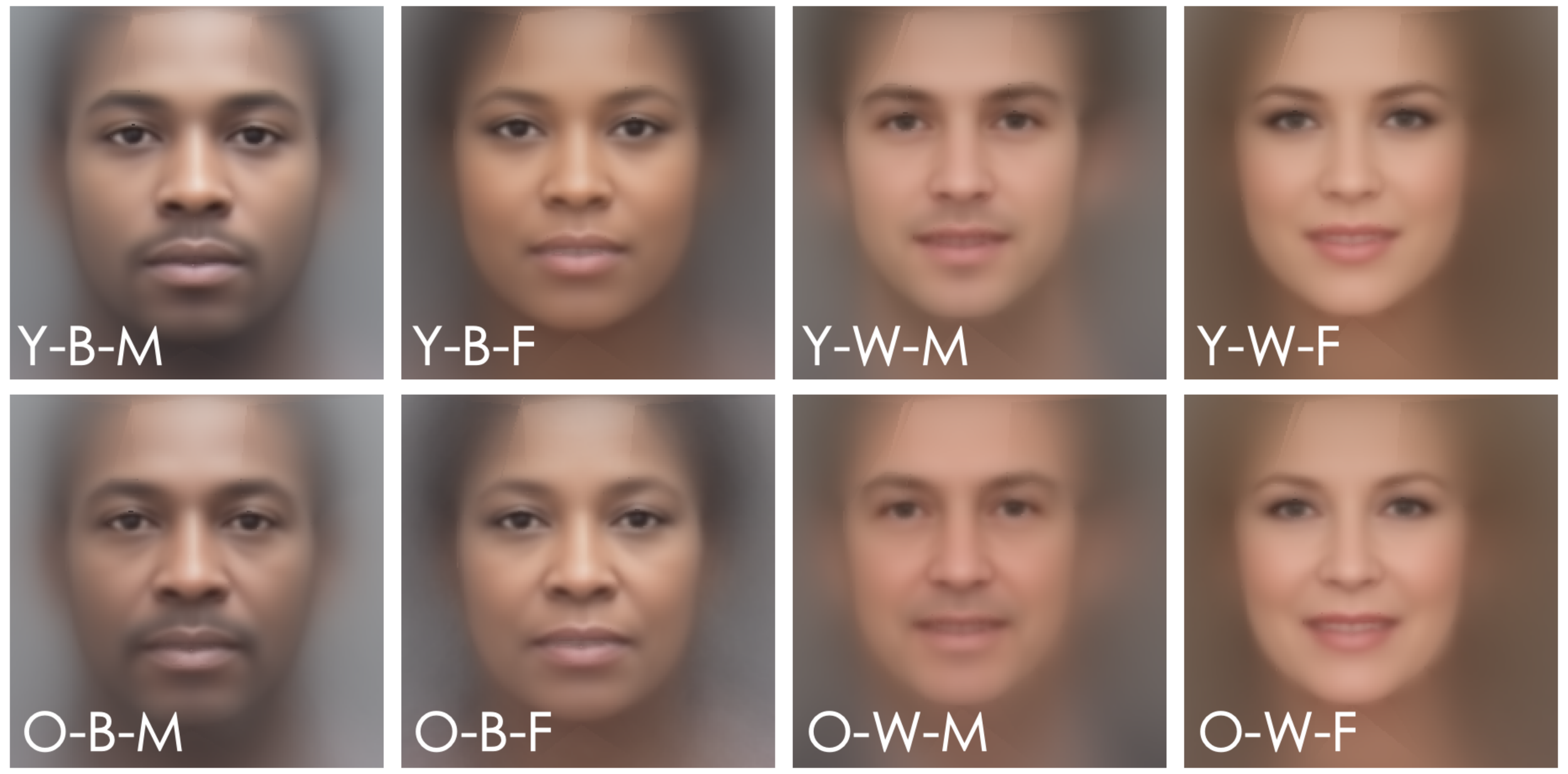
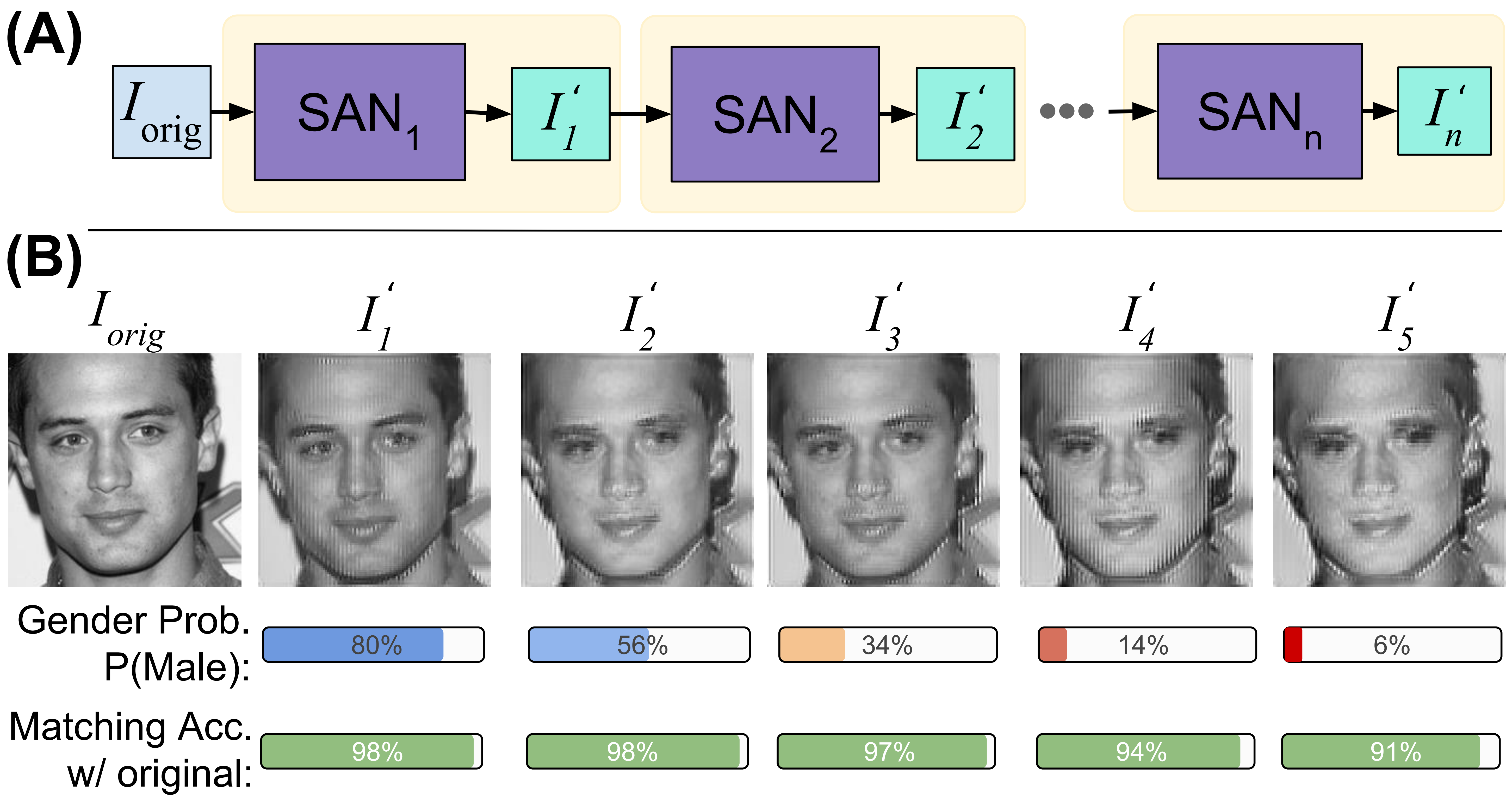


Figure 4: Face prototypes computed for each group of attribute labels. The abbreviations at the bottom of each image refer to the prototype attribute-classes, where Y=young, O=old, M=male, F=female, W=white, B=black.

V Mirjalili, S Raschka, and A Ross (2018) *Gender Privacy: An Ensemble of Semi Adversarial Networks for Confounding Arbitrary Gender Classifiers.* 9th IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS 2018)

FlowSAN: Privacy-enhancing Semi-Adversarial Networks to Confound Arbitrary Face-based Gender Classifiers



Improvements to better control the perturbations and enhance the removal of soft-biometric information

V Mirjalili, S Raschka, A Ross (2019)
FlowSAN: Privacy-enhancing Semi-Adversarial Networks to Confound Arbitrary Face-based Gender Classifiers
IEEE Access 2019, 10.1109/ACCESS.2019.2924619

Topics

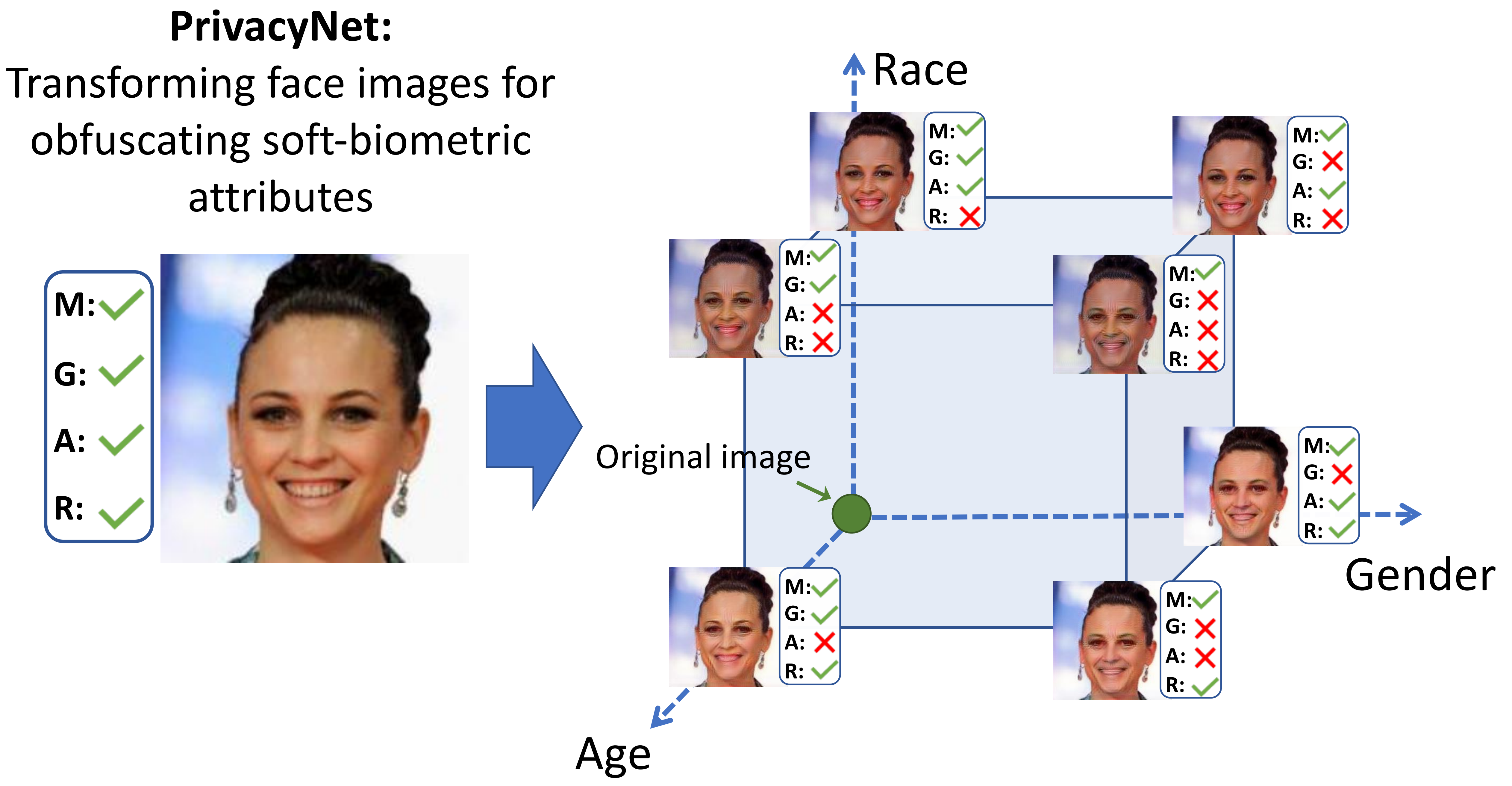
1. Biometric Face Recognition

2. Extracting Soft-Biometric Attributes from Face Images

3. Hiding Soft-Biometric Attributes in Face Images

4. **PrivacyNet: GAN-based Multi-attribute Face Privacy**

Selective and collective perturbations for imparting multi-attribute privacy to face images

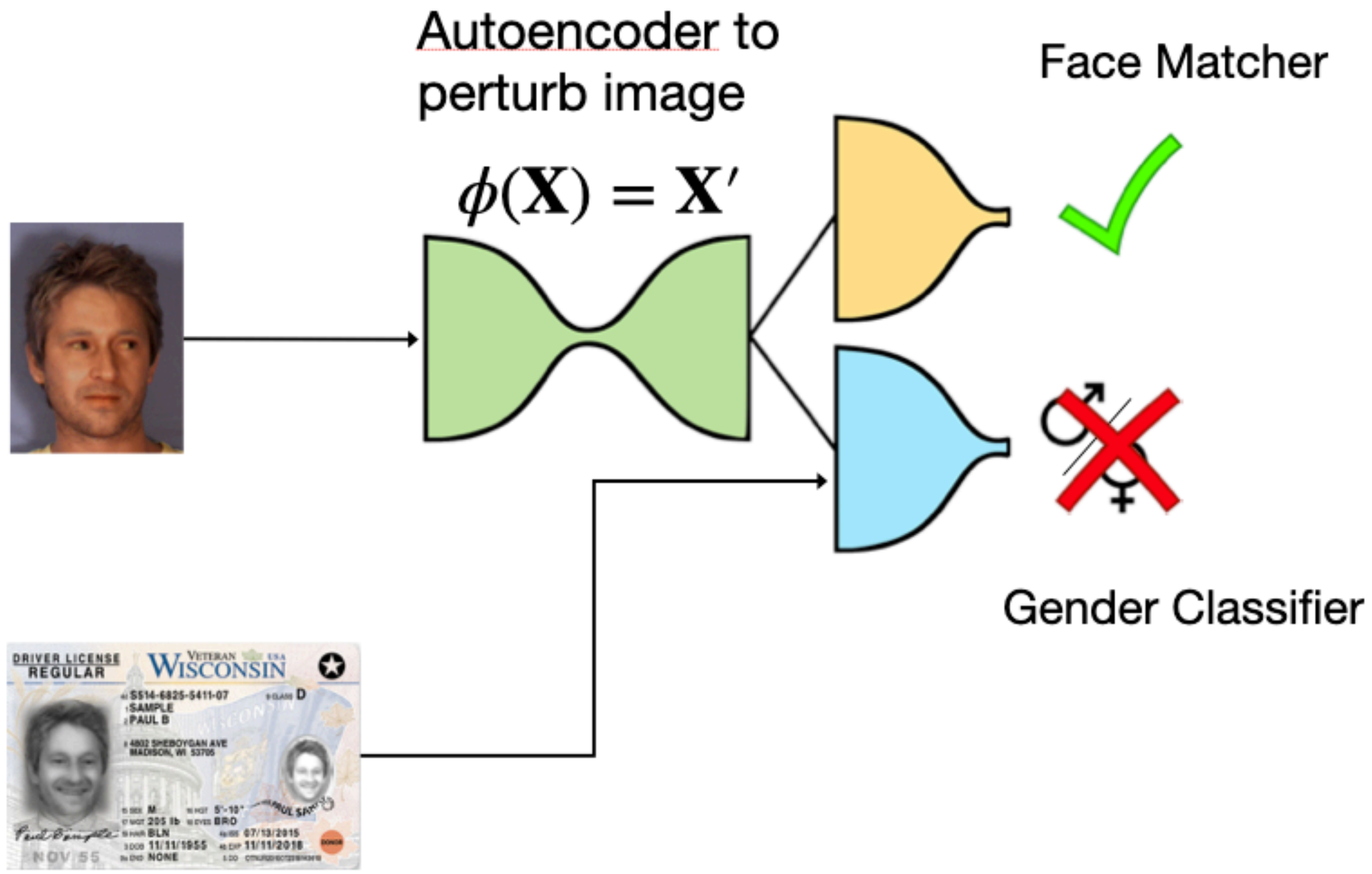


Selective = **which** attributes to conceal

Collective = **how many** attributes to conceal

V Mirjalili, S Raschka, and A Ross (2020)
PrivacyNet: Semi-Adversarial Networks for Multi-attribute Face Privacy
IEEE Transactions in Image Processing. Vol. 29, pp. 9400-9412, 2020.

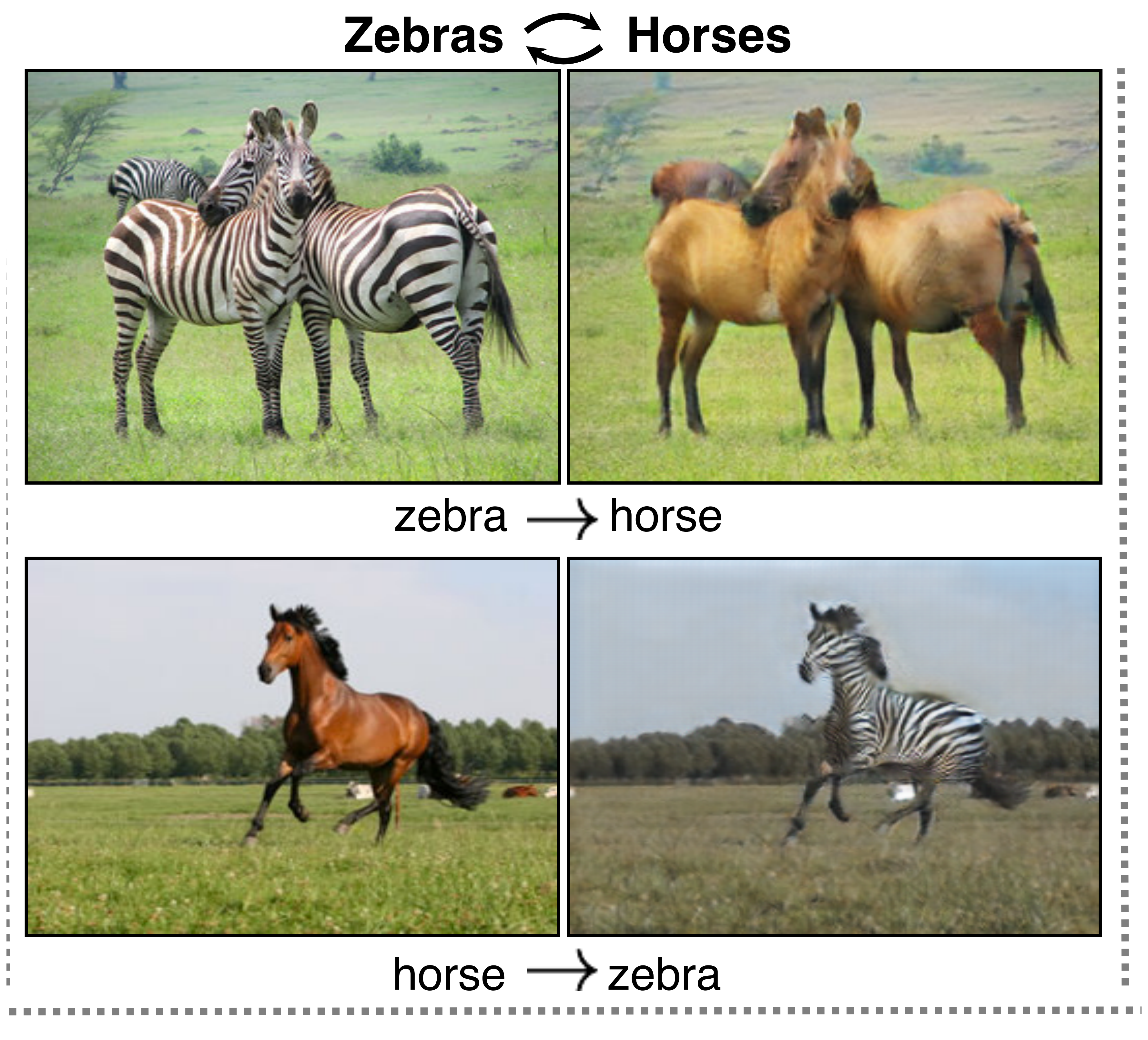
PrivacyNet replaces the convolutional autoencoder with a GAN-based model with cycle consistency loss



CycleGAN

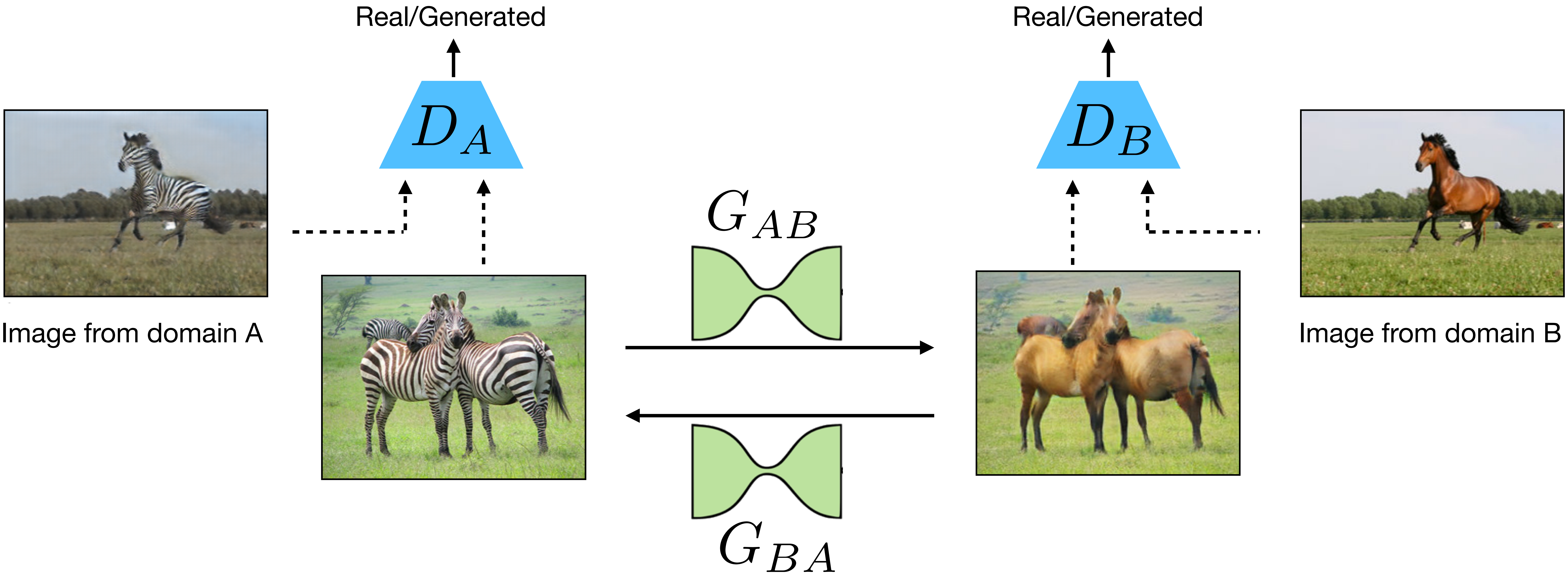
Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232). https://arxiv.org/abs/1703.10593

Does not require paired images from source and target domains

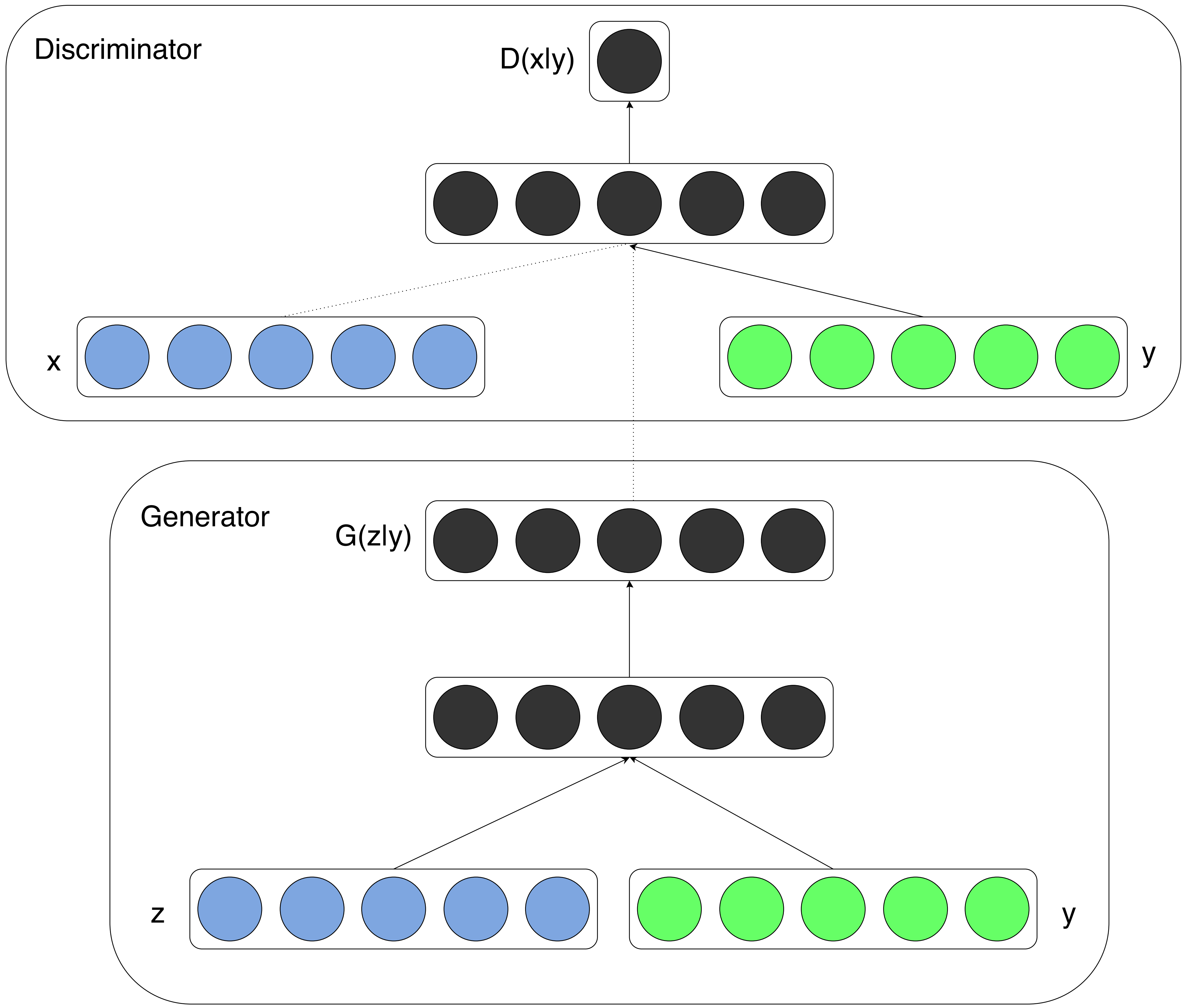


CycleGAN

Zhu, J. Y., Park, T., Isola, P., & Efros, A. A. (2017). Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision* (pp. 2223-2232). https://arxiv.org/abs/1703.10593

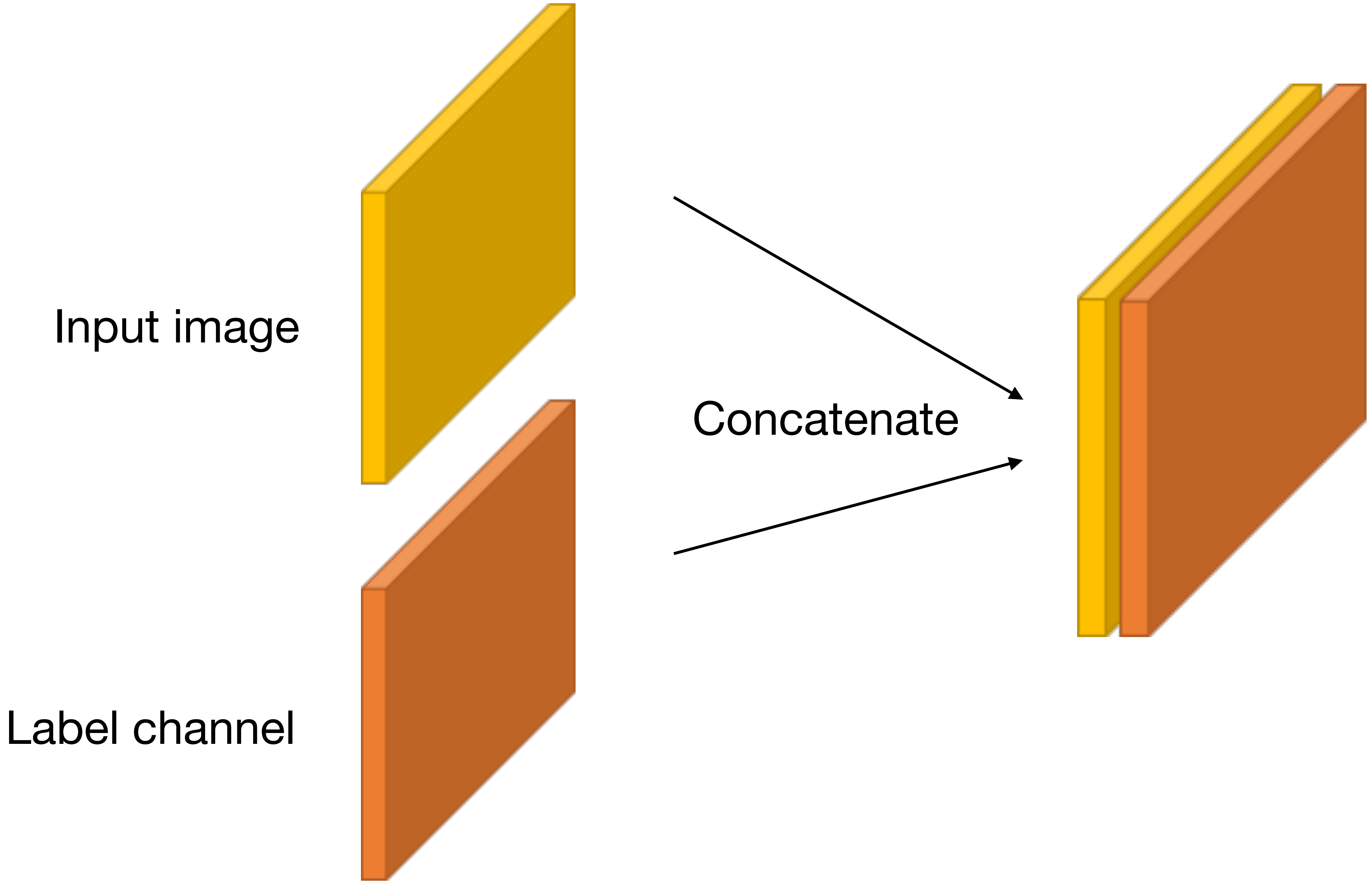


Conditional GAN



Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. https://arxiv.org/abs/1411.1784

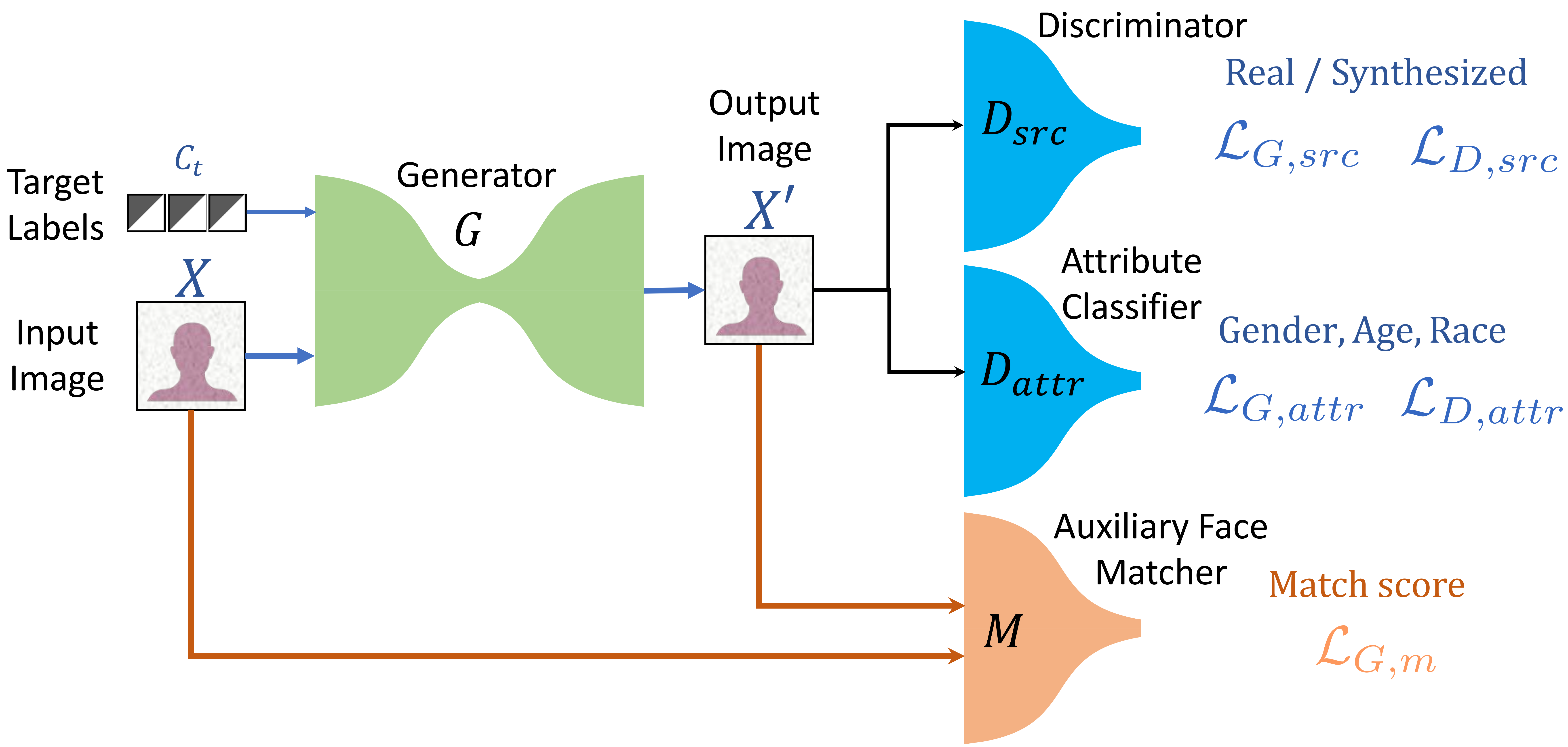
Conditional GAN



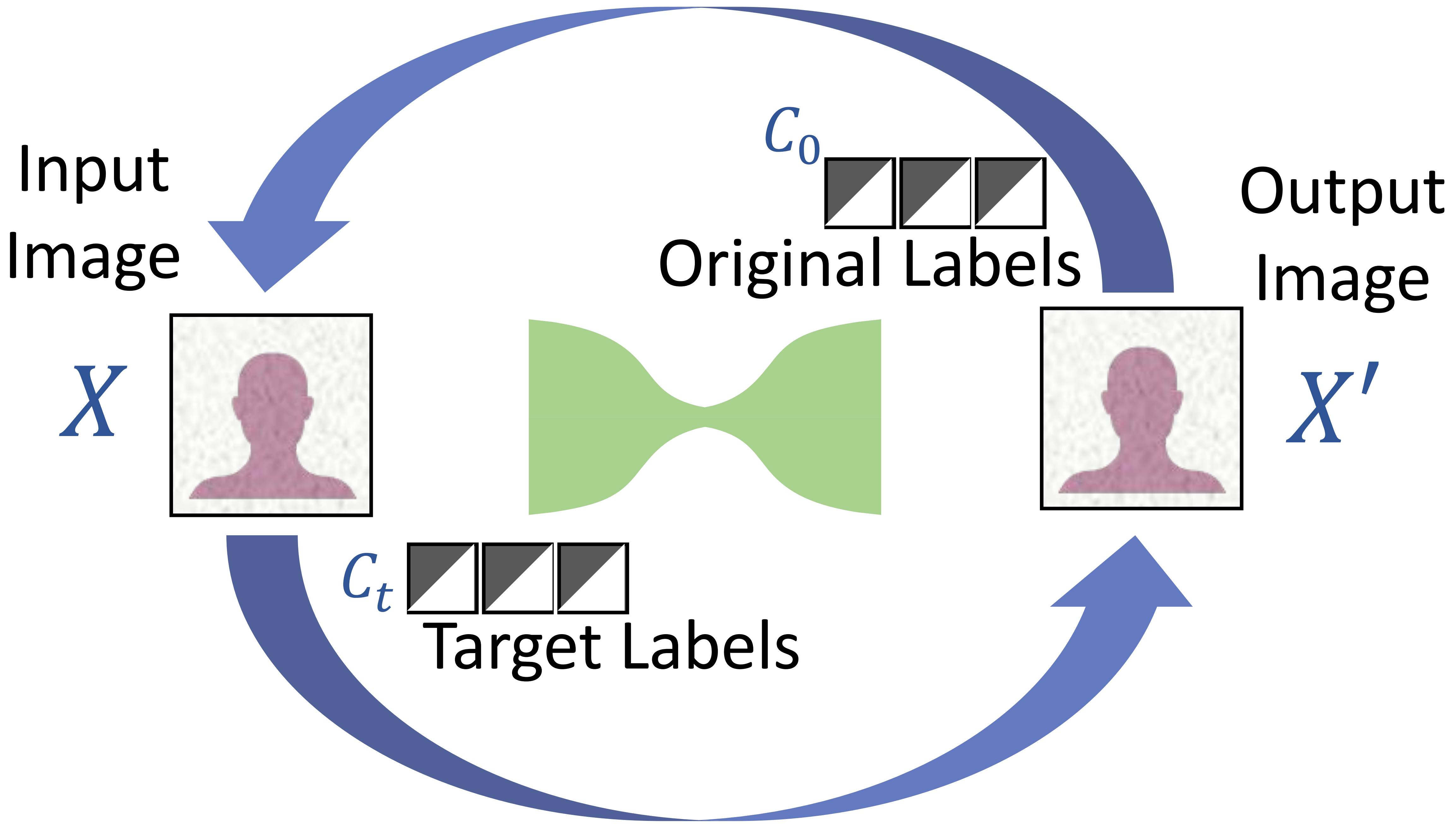
Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. https://arxiv.org/abs/1411.1784

PrivacyNet Architecture

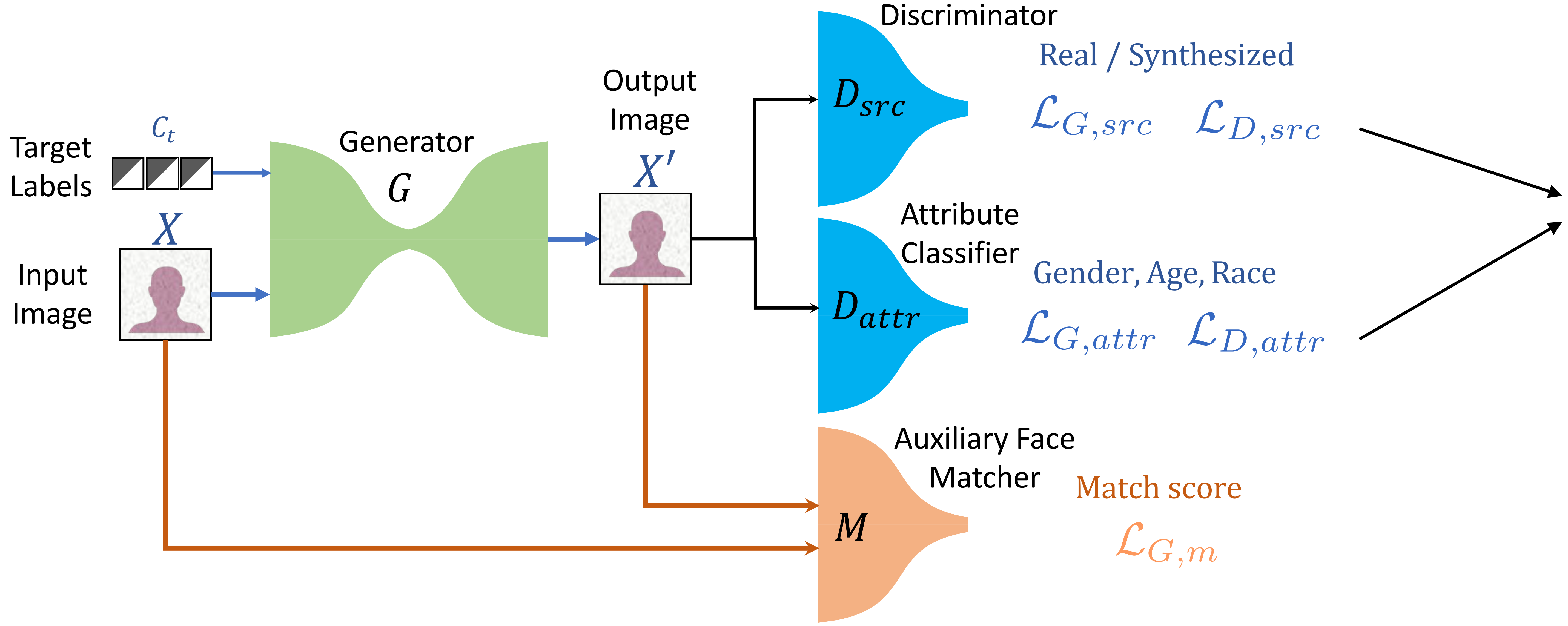
4 Subnetworks



PrivacyNet's Cycle Consistency

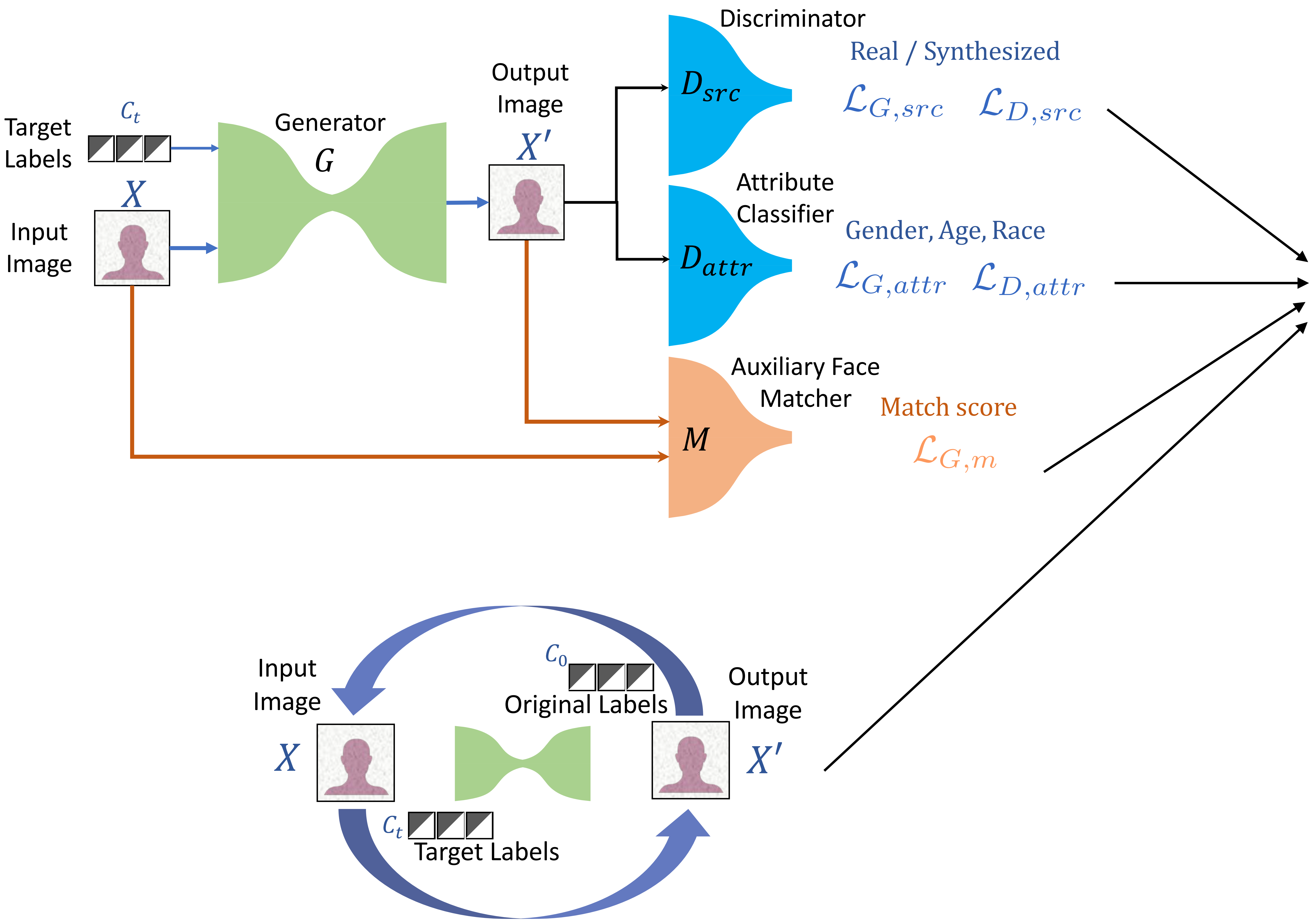


Discriminator Loss Function

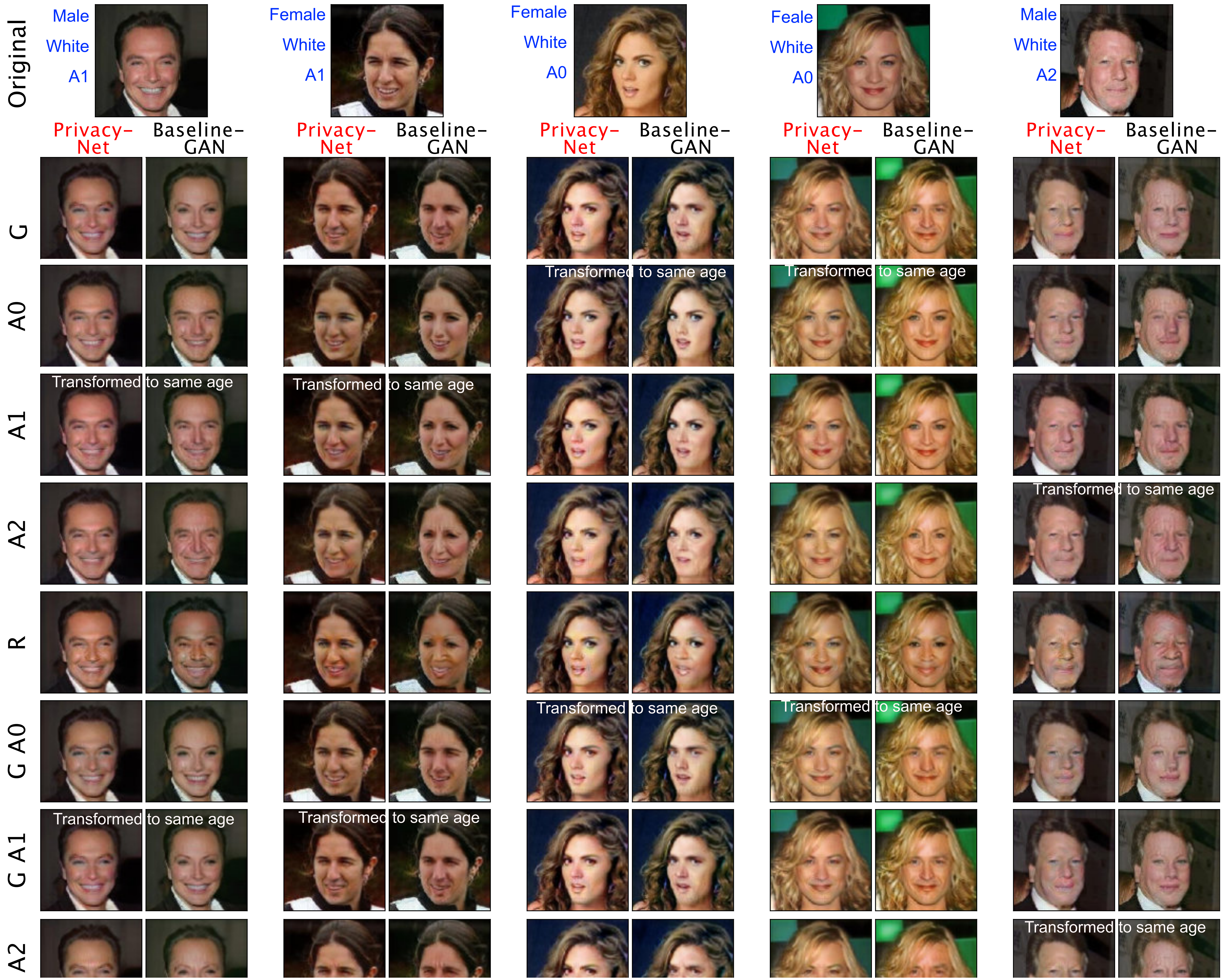


$$\mathcal{L}_{D,tot} = \mathcal{L}_{D,src} + \lambda_{D,attr}\mathcal{L}_{D,attr}$$

Generator Loss Function



$$\mathcal{L}_{G,tot} = \mathcal{L}_{G,src} + \lambda_{G,attr}\mathcal{L}_{G,attr} + \lambda_m \mathcal{L}_{G,m} + \lambda_{rec}\mathcal{L}_{G,rec}$$


Original
Male
White
A1
Female
White
A1
Female
White
A0
Feale
White
A0
Male
White
A2
Privacy-Net
Baseline-GAN
G
A0
A1
A2
R
G A0
G A1
A2
Transformed to same age

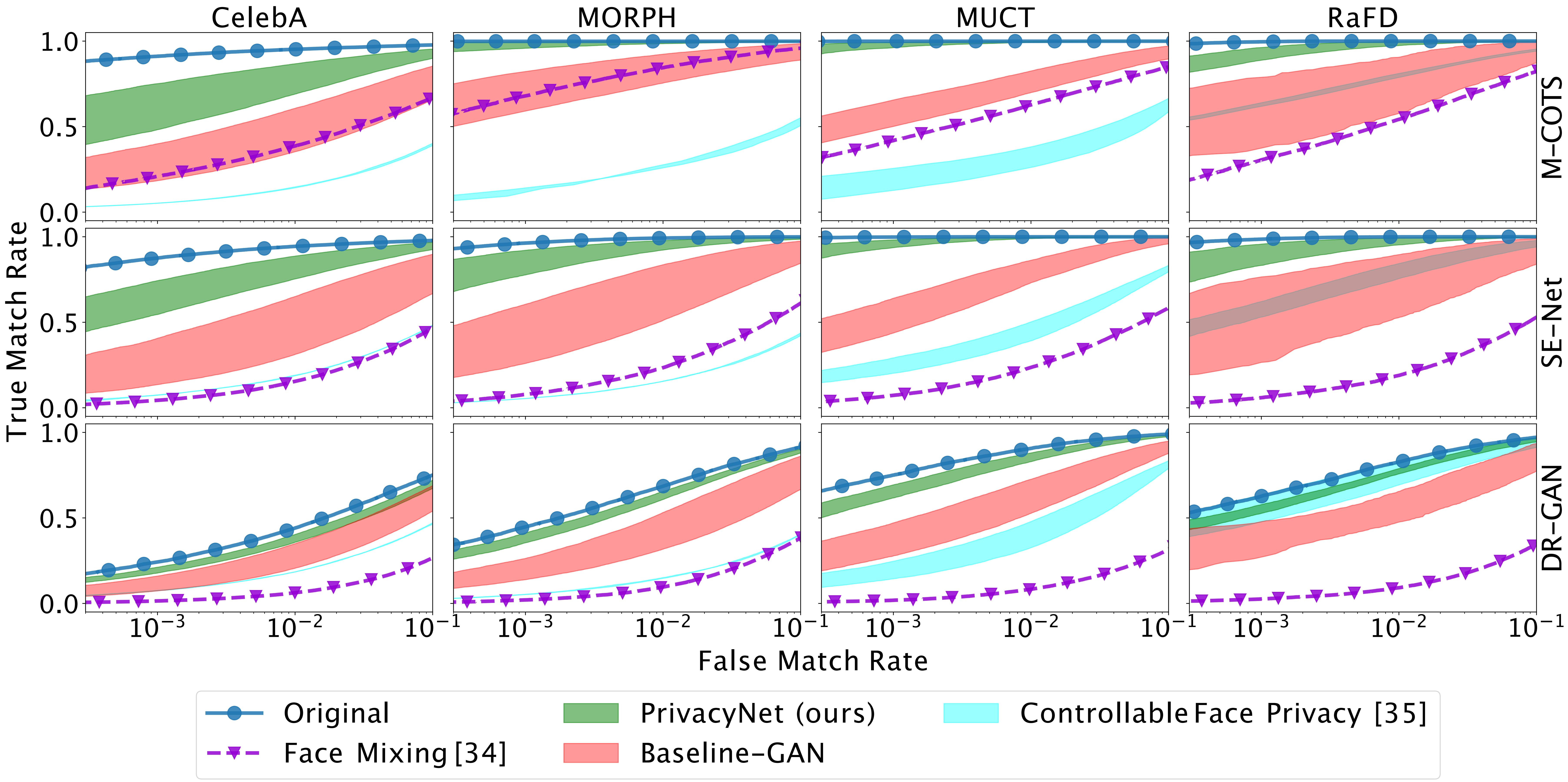


Fig. 8: ROC curves showing the performance of unseen face matchers on the original images for PrivacyNet, the baseline-GAN model, face mixing [34] approach and the controllable face privacy [35] method. The results show that ROC curves of PrivacyNet have the smallest deviation from the ROC curve of original images suggesting that the performance of face matching is minimally impacted, which is desired.

Limitation

$$y_{\text{age}} = \begin{cases} 0 & \text{age} \leq 30 \\ 1 & 30 < \text{age} \leq 45 \\ 2 & 45 < \text{age} \end{cases}$$

Suggested future solutions now that we can hide/change the age in face images ...

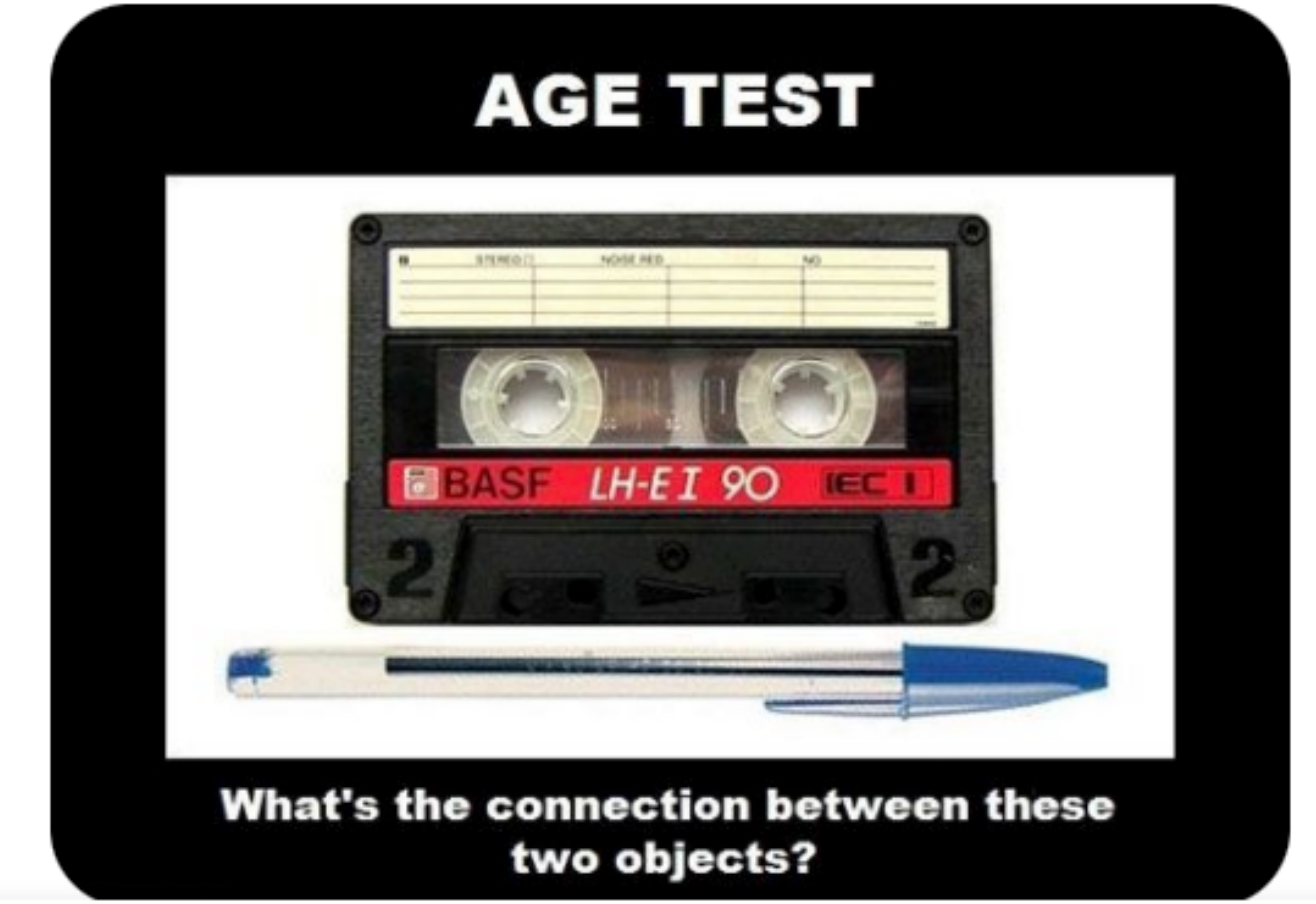


Image source: https://www.pinterest.cl/pin/211458144973743149/

Thank You!

https://sebastianraschka.com

@rasbt

Sebastian Raschka