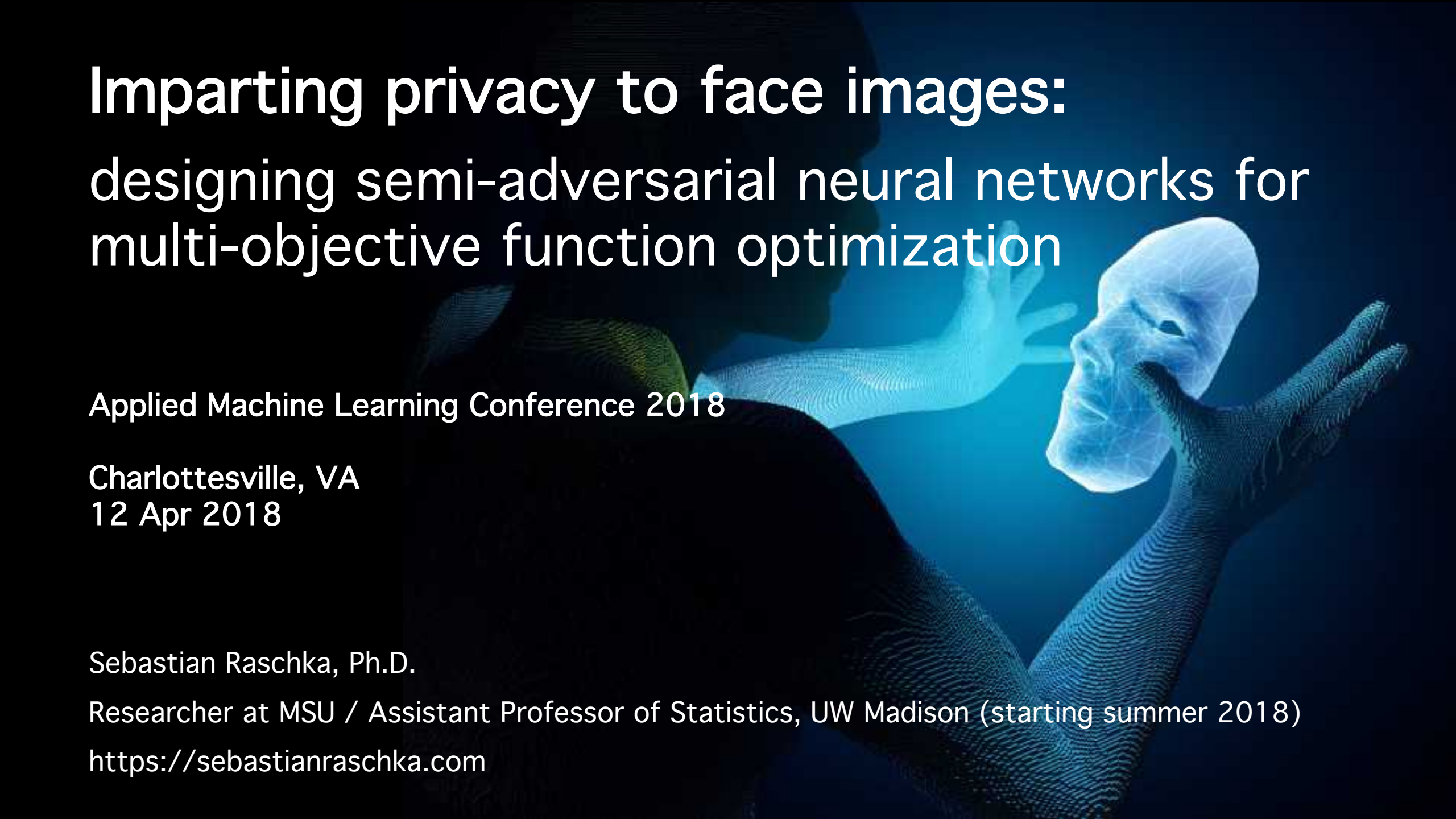
Imparting privacy to face images:

designing semi-adversarial neural networks for multi-objective function optimization



Applied Machine Learning Conference 2018

Charlottesville, VA
12 Apr 2018

Sebastian Raschka, Ph.D.

Researcher at MSU / Assistant Professor of Statistics, UW Madison (starting summer 2018)

https://sebastianraschka.com

Imparting privacy to face images:

designing semi-adversarial neural networks for multi-objective function optimization

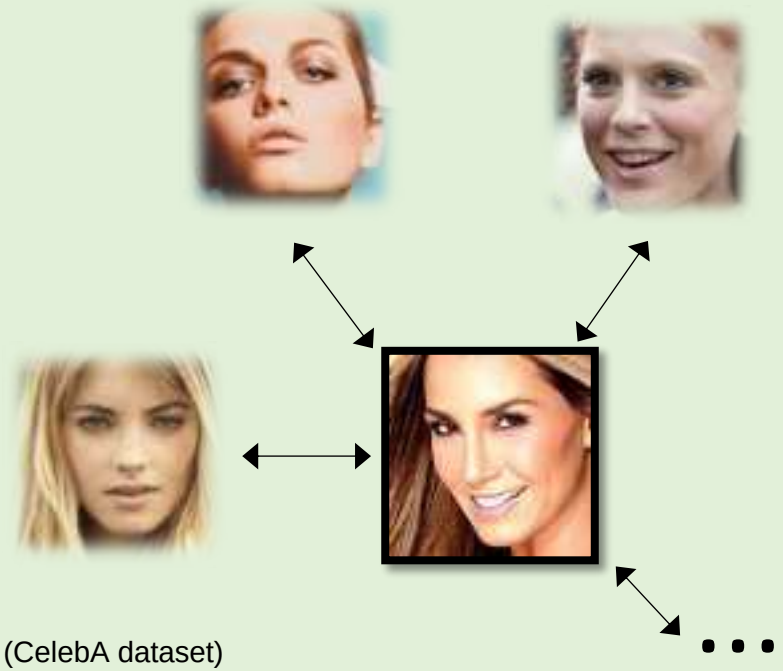
Mirjalili, Raschka, Namboodiri, Ross
"Semi-adversarial networks: Convolutional autoencoders for imparting privacy to face images."
The 11th IAPR International Conference on Biometrics, Gold Coast, Queensland, Australia (Feb 20th-23rd, 2018).
[manuscript version: https://arxiv.org/abs/1712.00321]

Best Paper Award @ ICB2018

Biometric (face) recognition

A. Identification

Determine identity of an unknown person
1-to-*n* matching



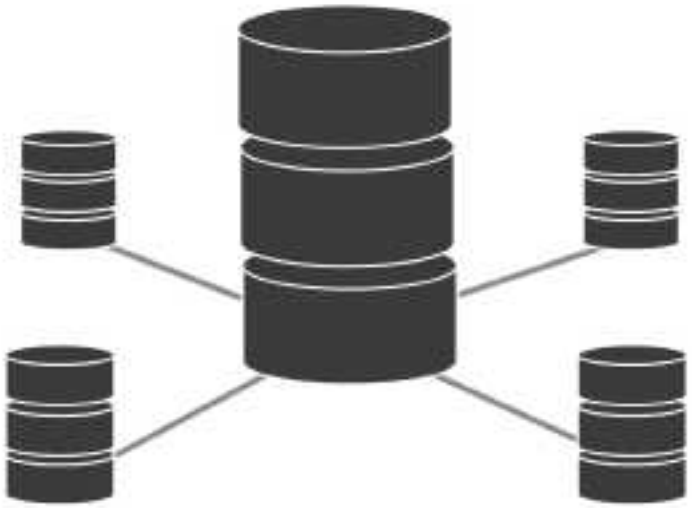
(CelebA dataset)

B. Verification

Verify claimed identity of a person
1-to-1 matching

(MUCT dataset)

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



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

Identity	John Doe

Age	65
Race	Caucasian
Medical	Healthy

SOFT BIOMETRIC ATTRIBUTES

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

Fully functional

Expand your software's capability by leveraging the ROC SDK. Face detection, search, verification, clustering, demographic estimation, and appearance classification are all available out of the box.

Any data, anywhere

Rank One algorithms excel under a multitude of facial variations. Whether faces are occluded, poorly lit or have a unique expression, your data can be used.

https://www.nytimes.com/2018/01/28/technology/europe-data-privacy-rules.html

Identity	John Doe
Age	65
Race	Caucasian
Medical	Healthy

SOFT BIOMETRIC ATTRIBUTES

Goal: differential privacy

1. Perturb gender information
2. Ensure realistic face images
3. Retain biometric face recognition utility

Maximize the performance with respect to one classifier while **minimizing** the performance of another.

Face matcher

100101010101111
101010101001000
001101010101010
101001100101010
010111010001010
101110001010101
101001010100101

P(same person)

100101010101111
101010101001000
001101010101010
101001100101010
010111010001010
101110001010101
101001010100101

P(male)

Gender classifier

Goal:

- perturbing gender
- retaining matching utility

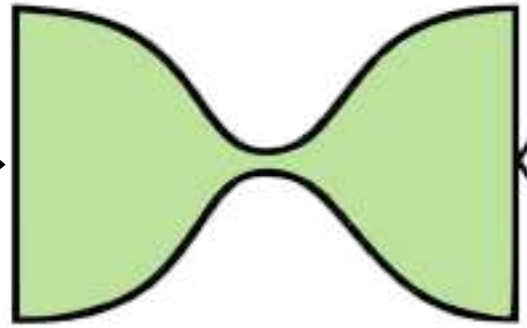
Face matcher

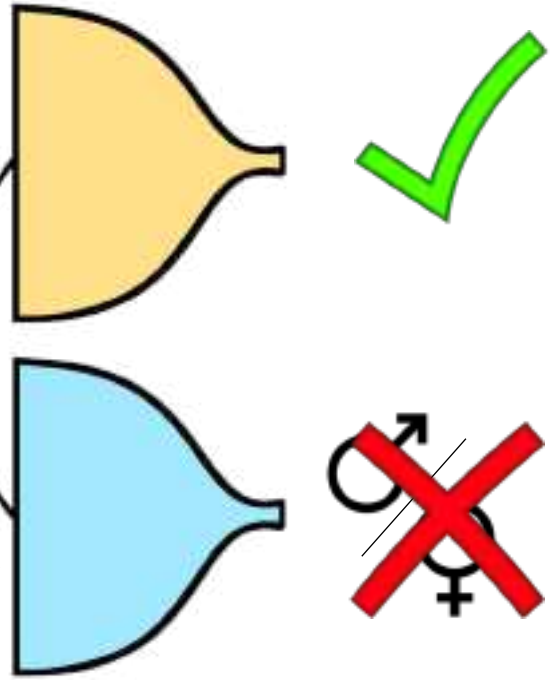
Gender classifier

Autoencoder to perturb image

$\phi(X) = X'$

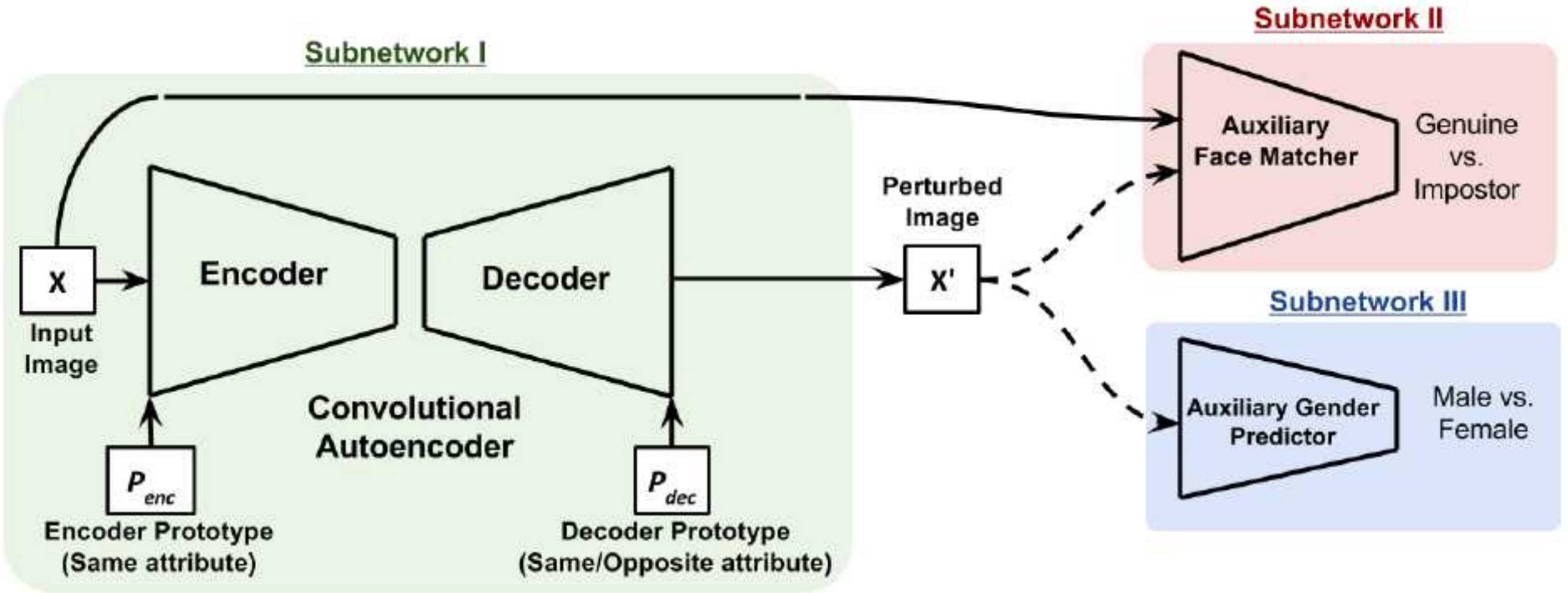
Face matcher



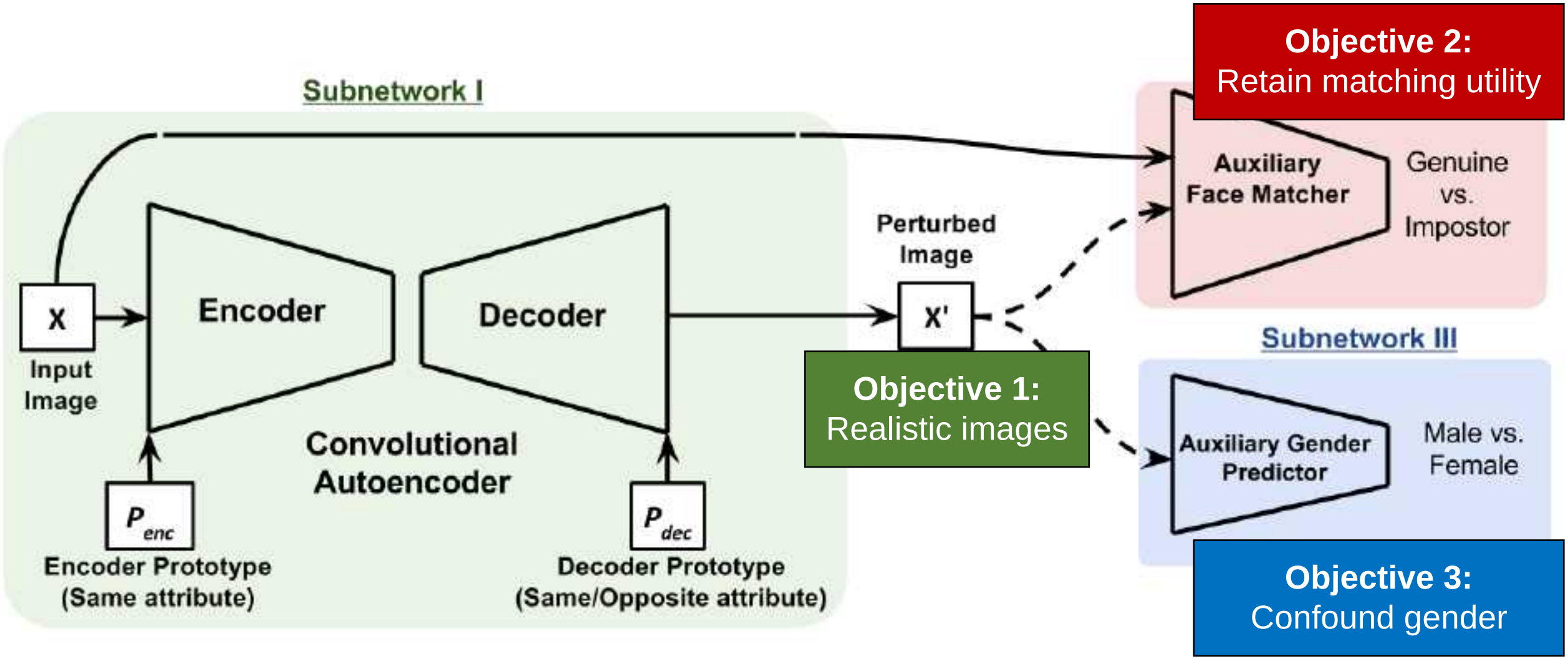


Gender classifier

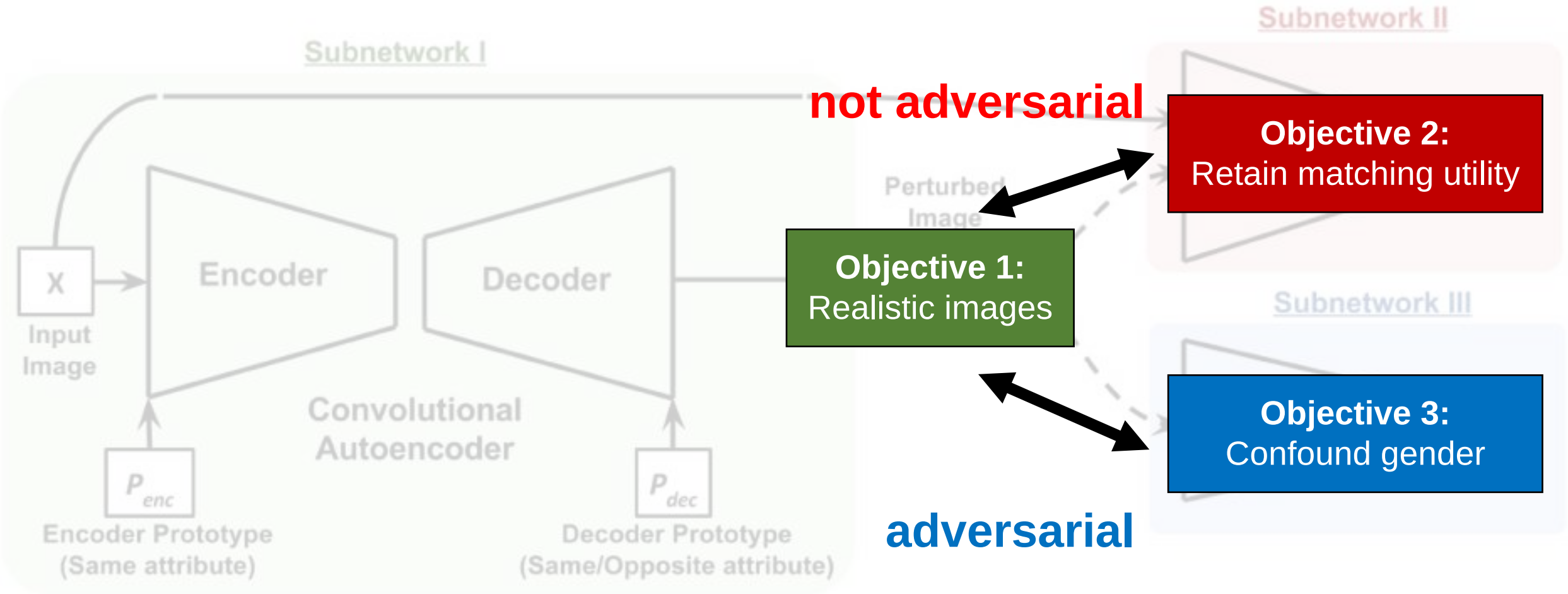
General architecture of the semi-adversarial network



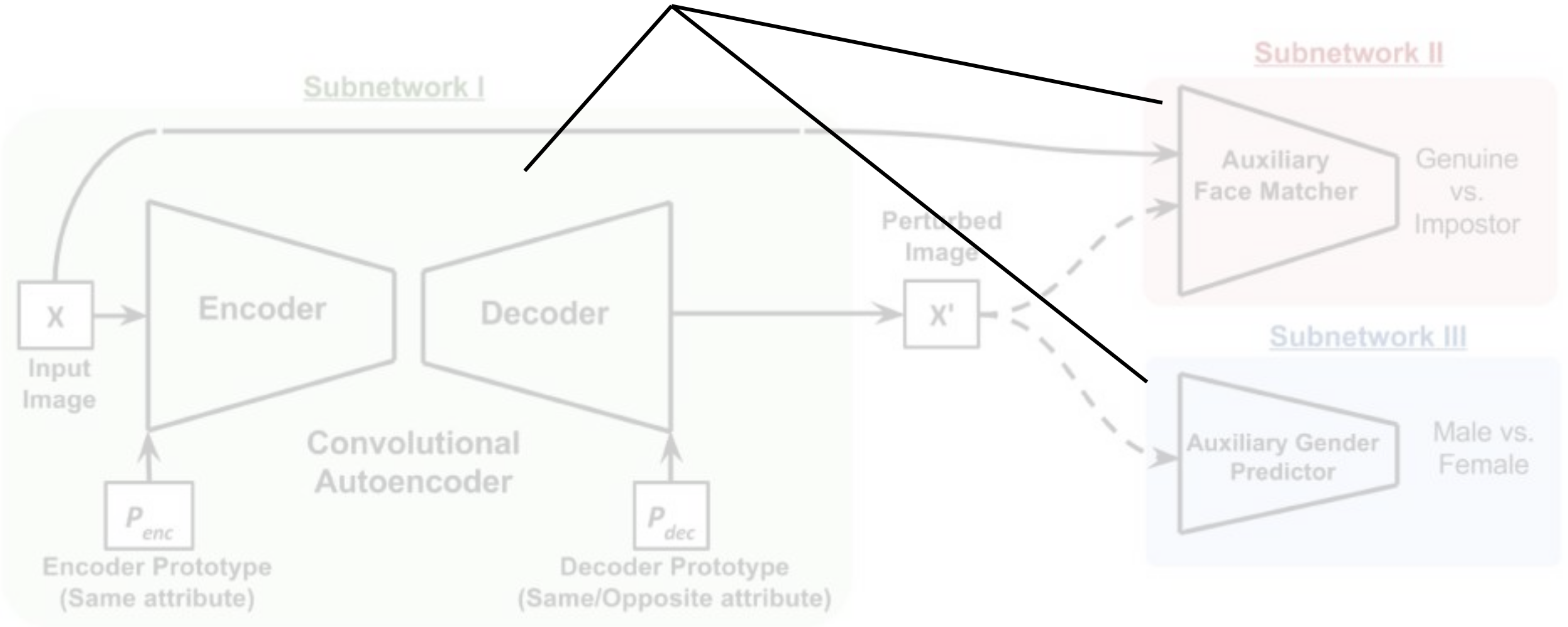
General architecture of the semi-adversarial network



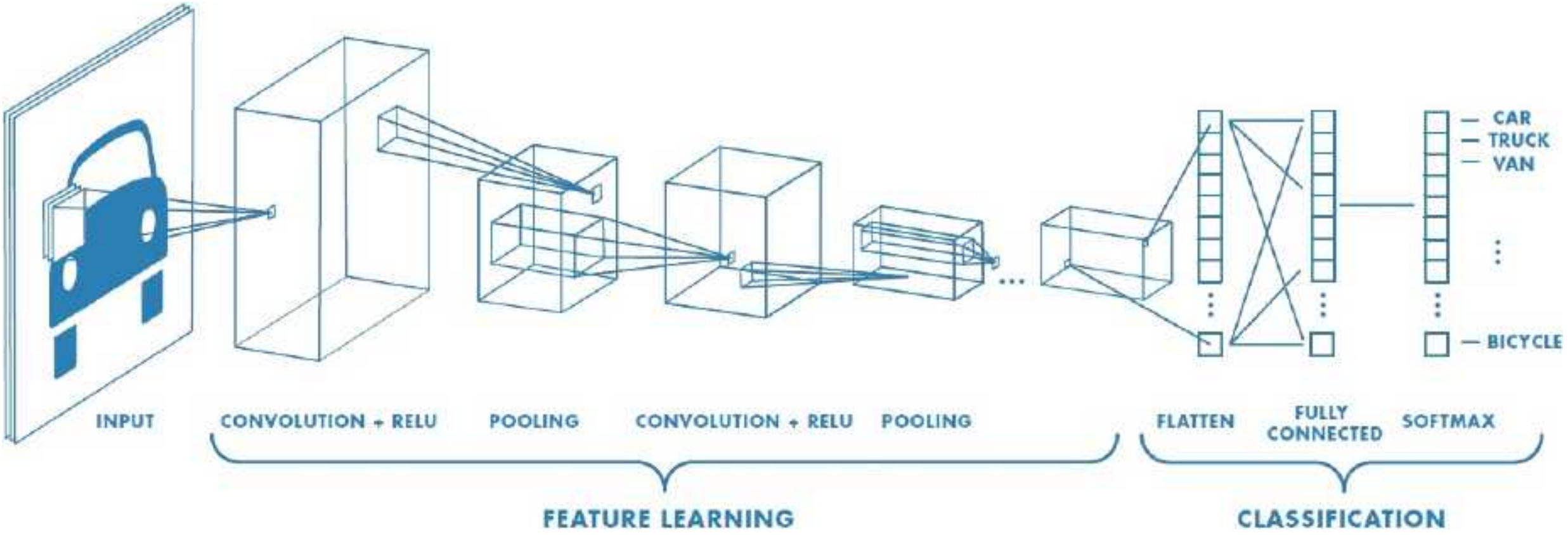
Semi-adversarial network



Convolutional neural networks

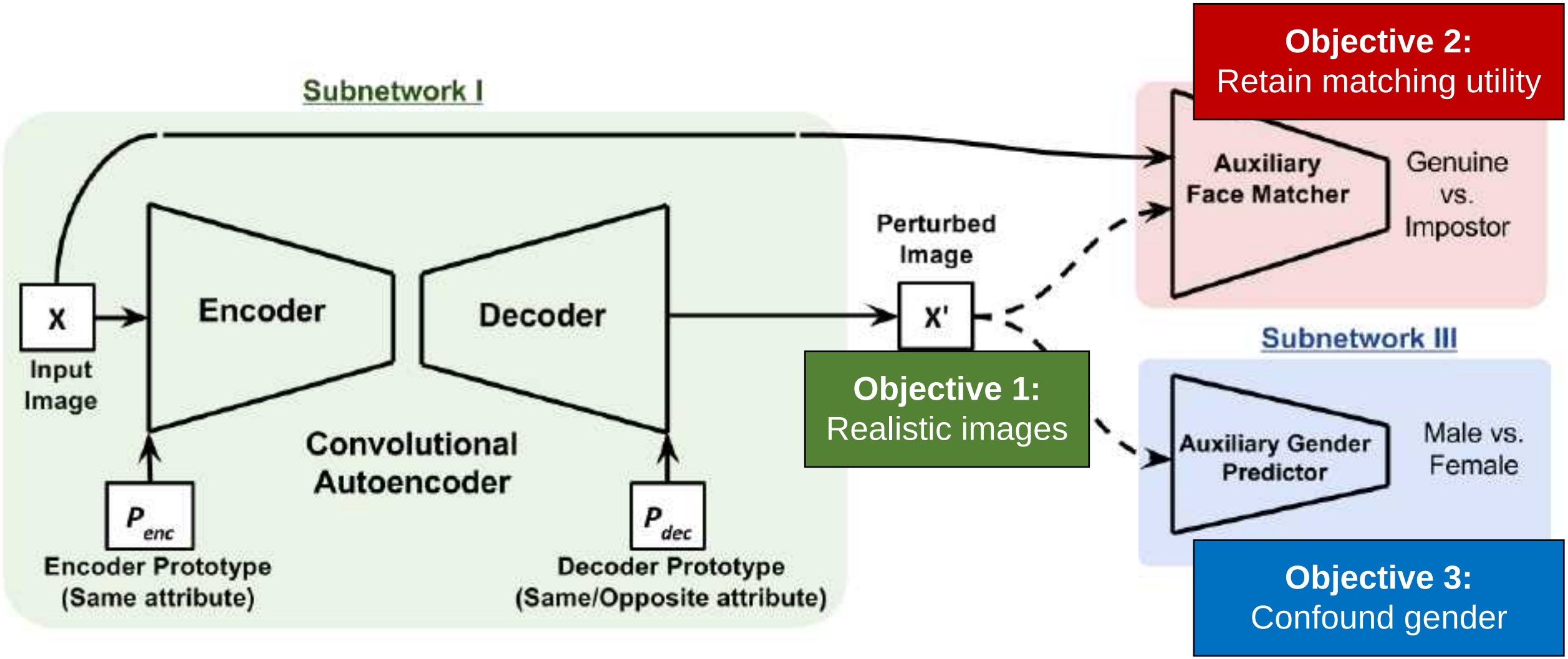


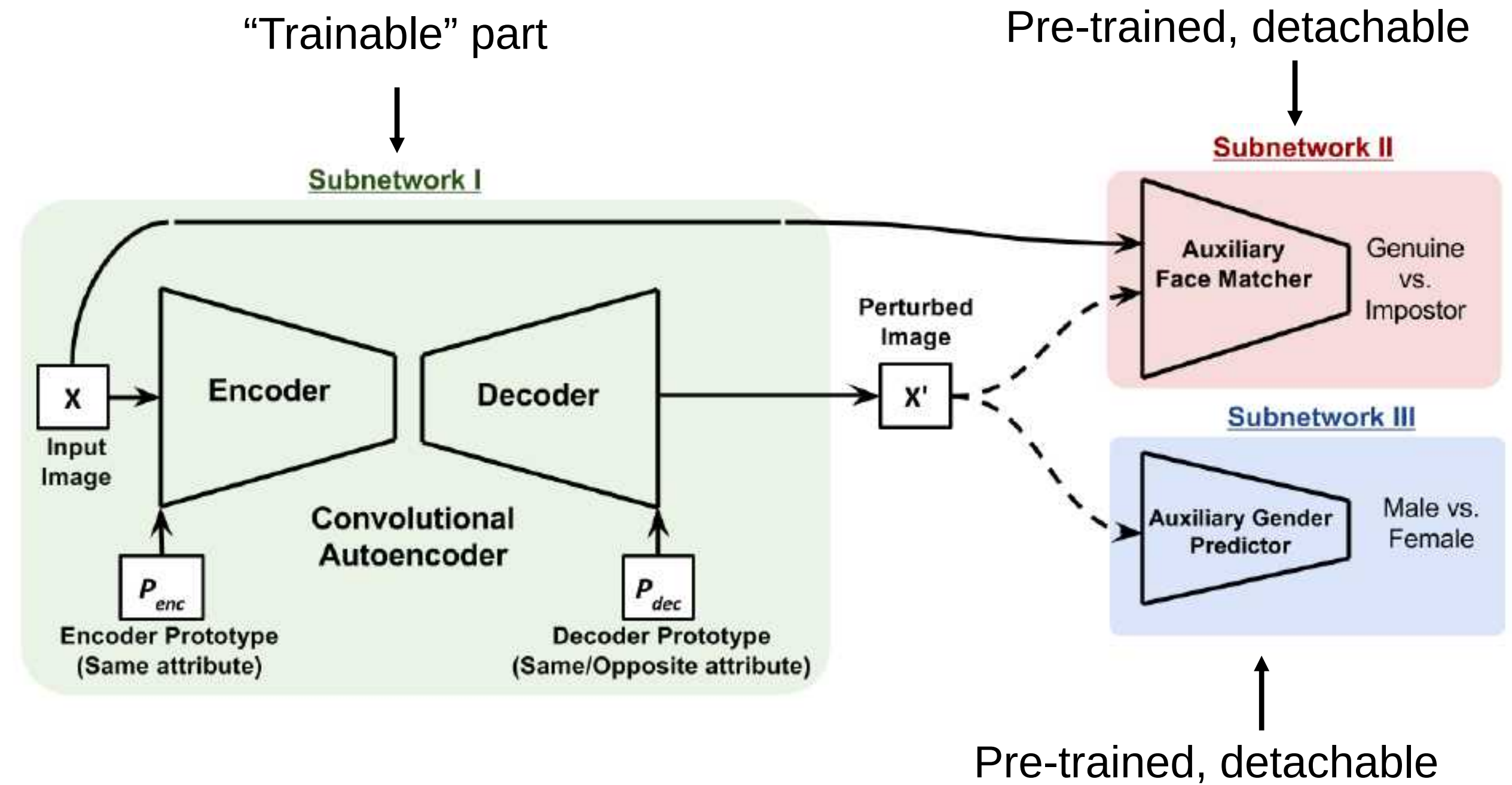
Convolutional neural network classifier



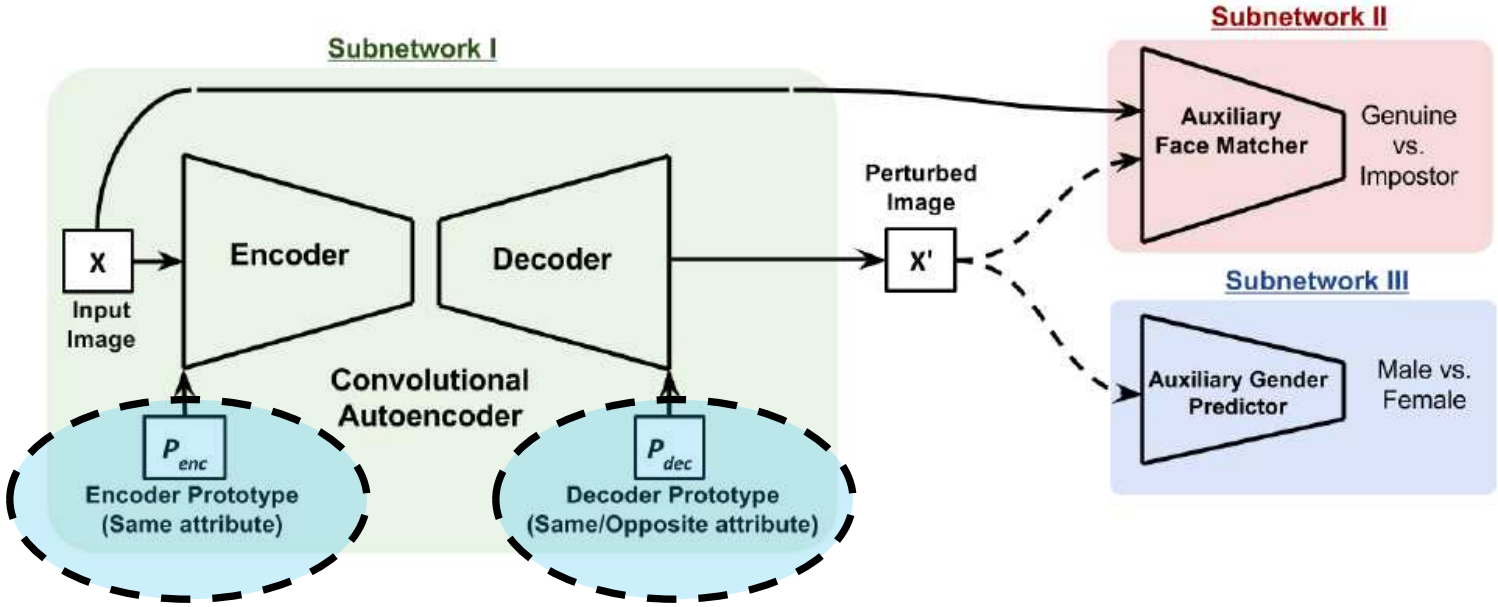
https://www.mathworks.com/content/mathworks/www/en/discovery/convolutional-neural-network/_jcr_content/mainParsys/image_copy.adapt.full.high.jpg/1517522275430.jpg

General architecture of the semi-adversarial network




"Trainable" part
Pre-trained, detachable
Subnetwork I
Subnetwork II
Auxiliary Face Matcher
Genuine vs. Impostor
Perturbed Image
X
Input Image
Encoder
Decoder
X'
Subnetwork III
Auxiliary Gender Predictor
Male vs. Female
Convolutional Autoencoder
P_{enc}
Encoder Prototype (Same attribute)
P_{dec}
Decoder Prototype (Same/Opposite attribute)
Pre-trained, detachable

Gender prototypes

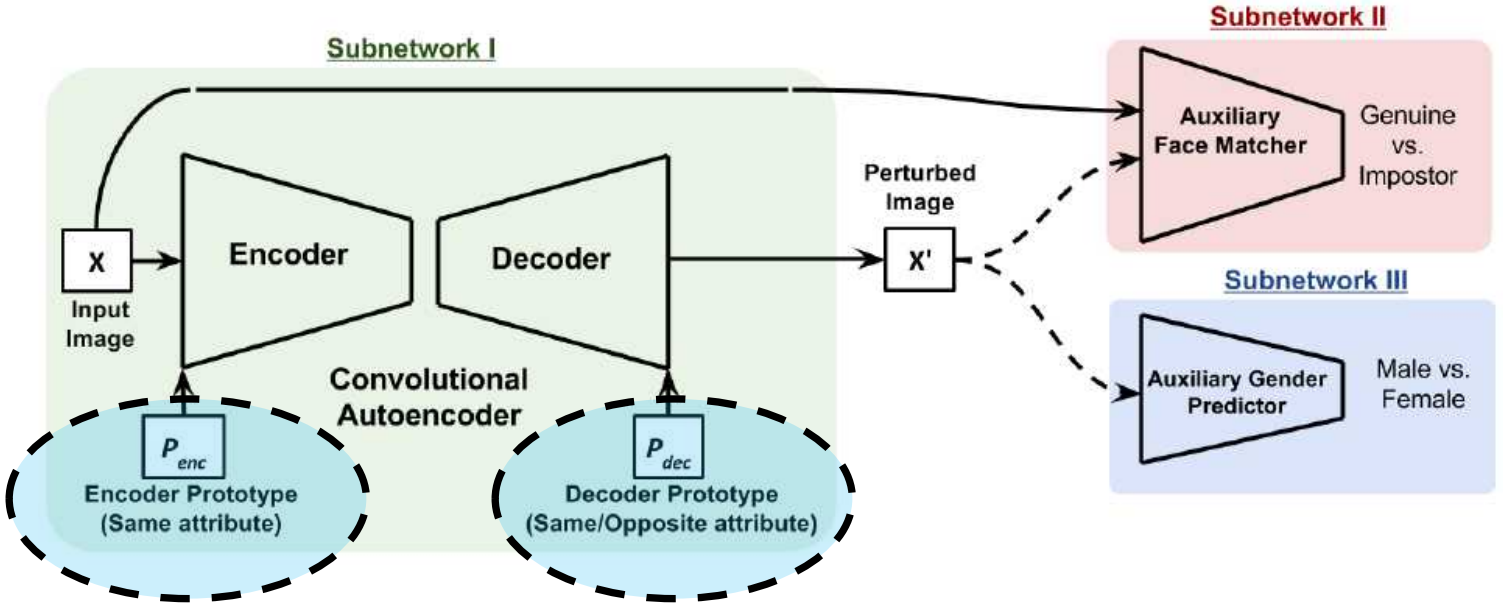


P_{Female}
average of all female images

$P_{neutral}$
weighted average P_{Male} and P_{Female}

P_{Male}
average of all male images

Gender prototypes



Class labels

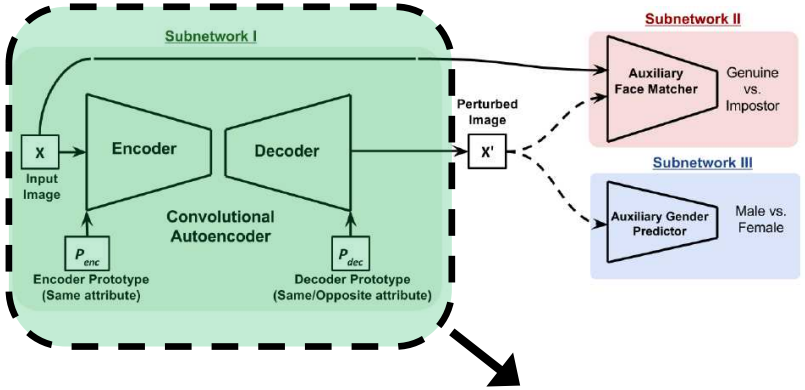
$y \in \{0, 1\}$,

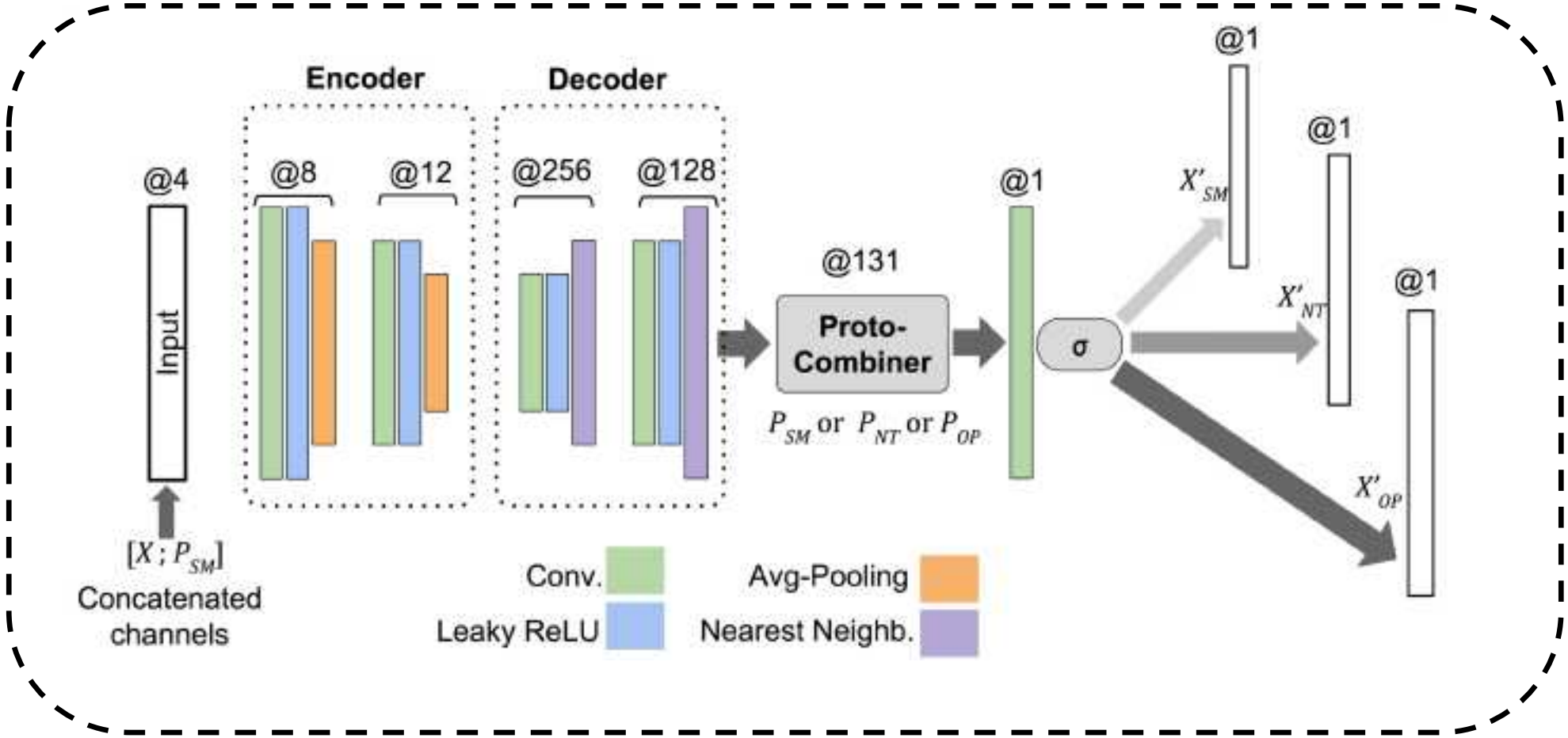
where $0 = \text{female}, 1 = \text{male}$

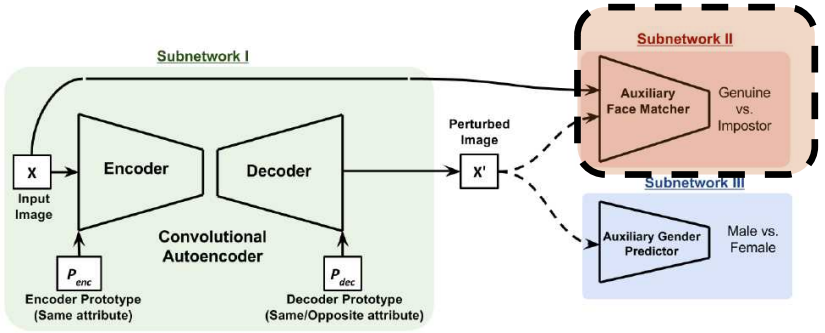
Same gender prototype: $P_{\mathrm{SM}}(y) = yP_{\mathrm{Male}} + (1-y)P_{\mathrm{Female}}$

Opposite gender prototype: $P_{\mathrm{OP}}(y) = (1-y)P_{\mathrm{Male}} + yP_{\mathrm{Female}}$

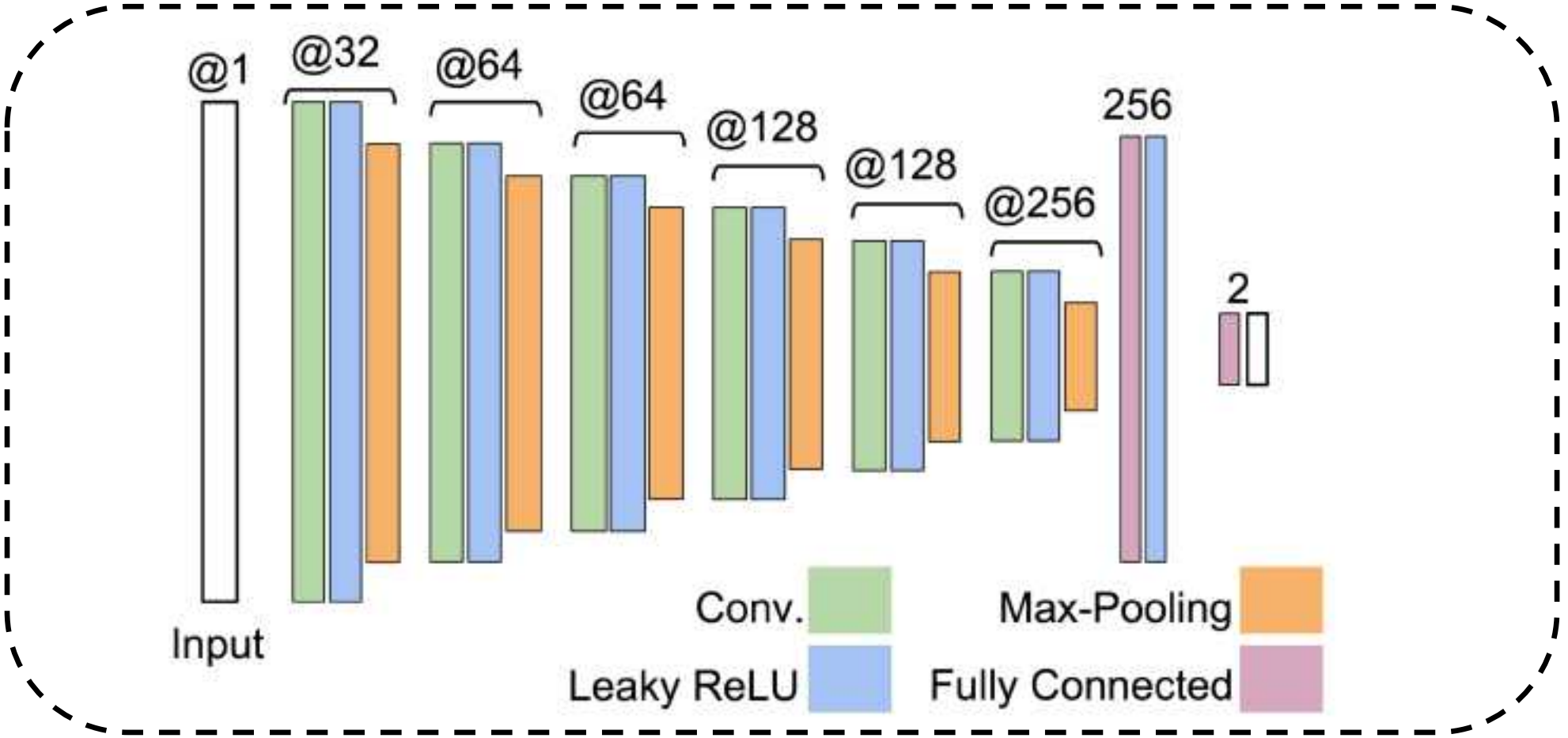
Convolutional autoencoder architecture

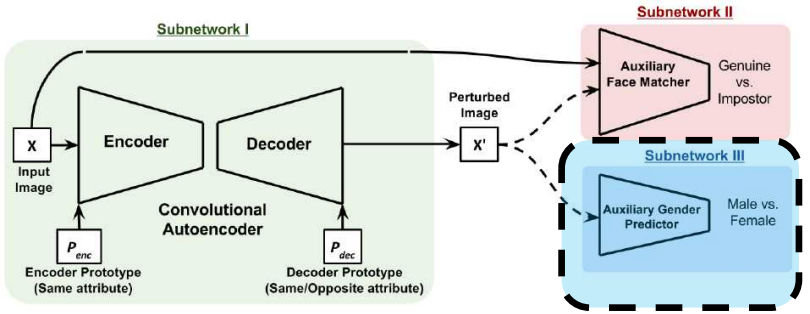




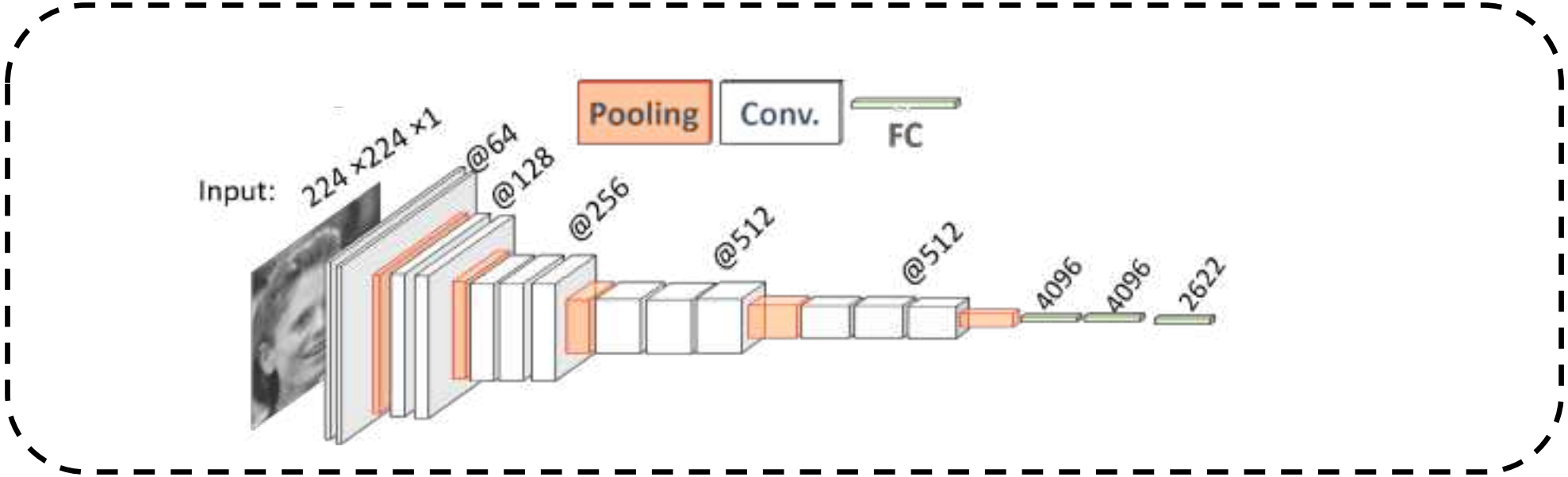


Gender classifier architecture



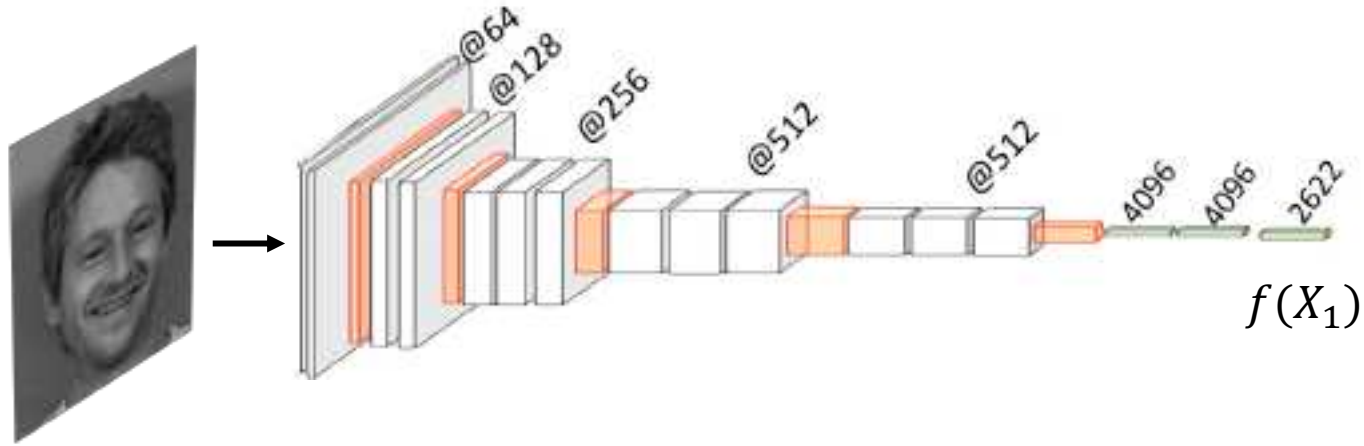


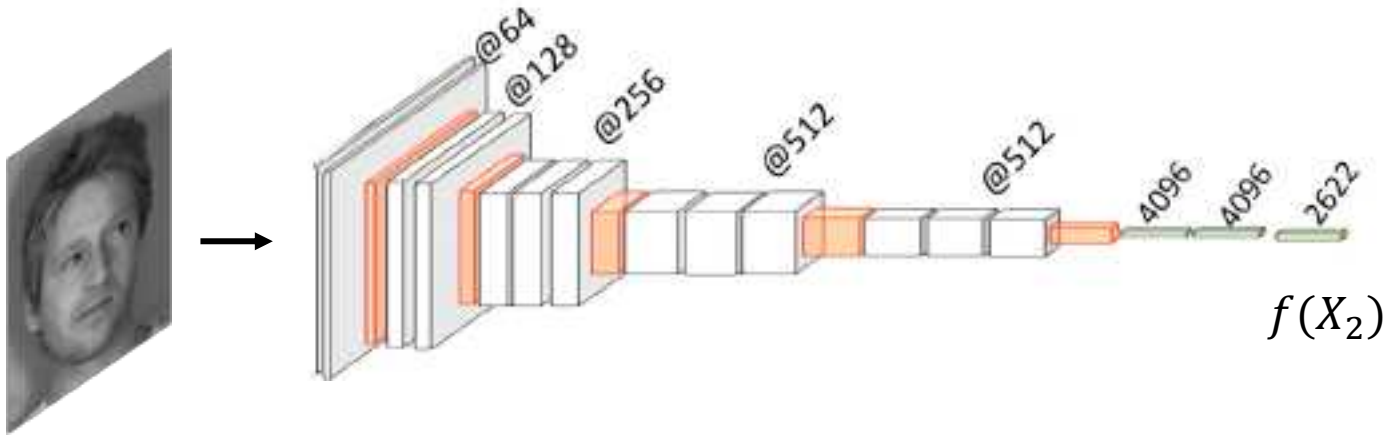
Face matcher architecture



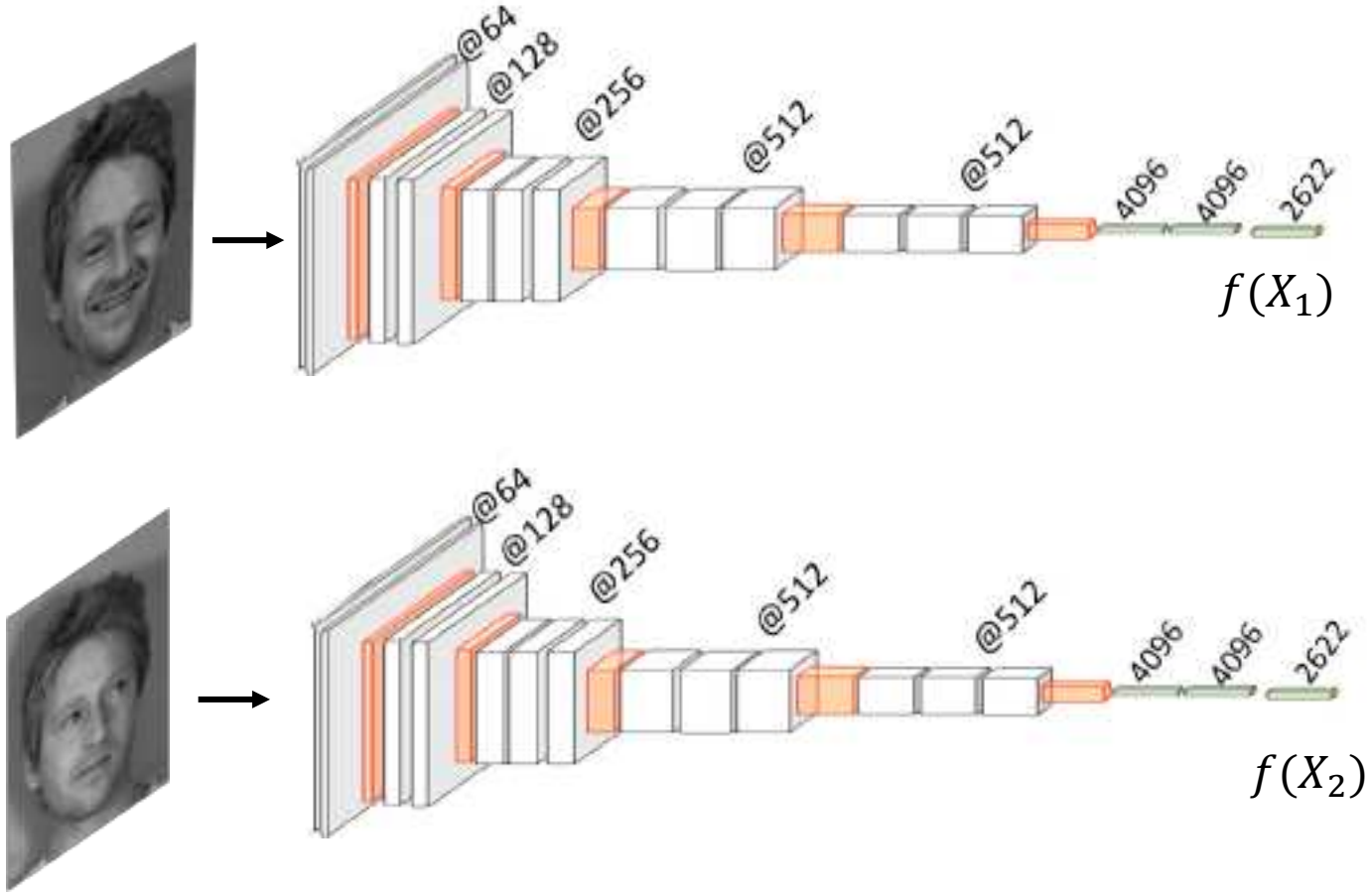
Architecture described in “Parkhi O., Vedaldi M., Zisserman A., “Deep Face Recognition”, BMVC, 2015.

Face matcher: Siamese network





Face matcher: Siamese network



$$d(X_1, X_2) = \|f(X_1) - f(X_2)\|_2^2$$

Training a face matcher: Triplet loss

Anchor

Positive

Want encodings to be very similar
(small distance)

Anchor

Negative

Want encodings to be very different
(large distance)

Training a face matcher: Triplet loss

Anchor

Positive

Anchor

Negative

$$d(A, P) \leq d(A, N)$$

$$\|f(A) - f(P)\|_2^2 \leq \|f(A) - f(N)\|_2^2$$

Training a face matcher: Triplet loss

Anchor

Positive

Anchor

Negative

$$d(A,P) + \alpha \leq d(A,N)$$

$$\|f(A) - f(P)\|_2^2 + \alpha \leq \|f(A) - f(N)\|_2^2$$

Training a face matcher: Triplet loss

Anchor

Positive

Anchor

Negative

$$d(A,P) + \alpha \leq d(A,N)$$

$$\|f(A) - f(P)\|_2^2 + \alpha \leq \|f(A) - f(N)\|_2^2$$

$$\|f(A) - f(P)\|_2^2 + \alpha - \|f(A) - f(N)\|_2^2 \leq 0$$

Training a face matcher: Triplet loss

Anchor

Positive

Anchor

Negative

$$J(A,P,N) = \max(\, \|f(A) - f(P)\|_2^2 + \alpha - \|f(A) - f(N)\|_2^2, 0 \,)$$

Training a face matcher: Triplet loss

Anchor

Positive

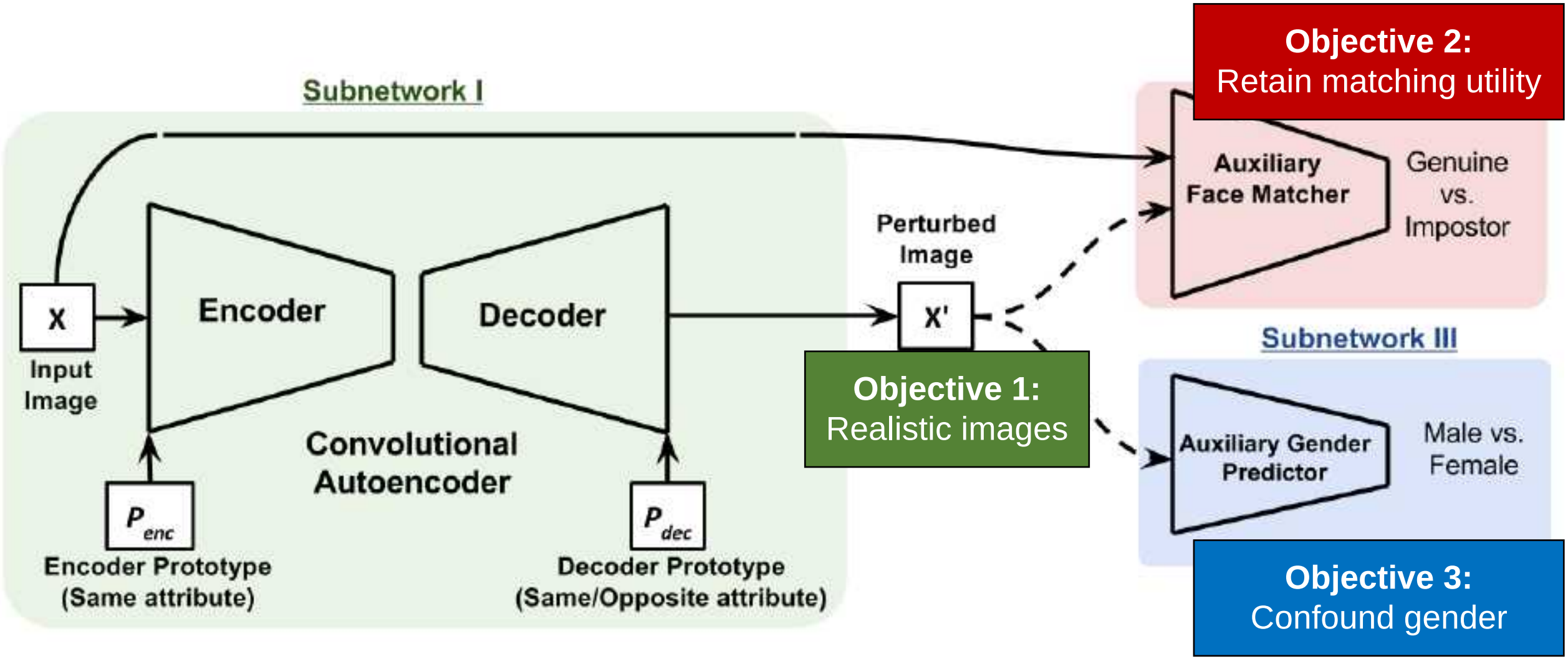
Anchor

Negative

$$J(A,P,N) = \max(\, \|f(A) - f(P)\|_2^2 + \alpha - \|f(A) - f(N)\|_2^2, 0 \,)$$

Side note: shortcoming off triplet loss
(as noted by Yann LeCun & Alfredo Canziani)

General architecture of the semi-adversarial network



Cost function for semi-adversarial learning

1. Pixel-wise similarity term
   - Only used during the pre-training of the autoencoder

$$J_D(X, X'_{SM}) = \sum_{i=1}^{224\times224} MSE\left(X^{(i)}, X'^{(i)}_{SM}\right)$$

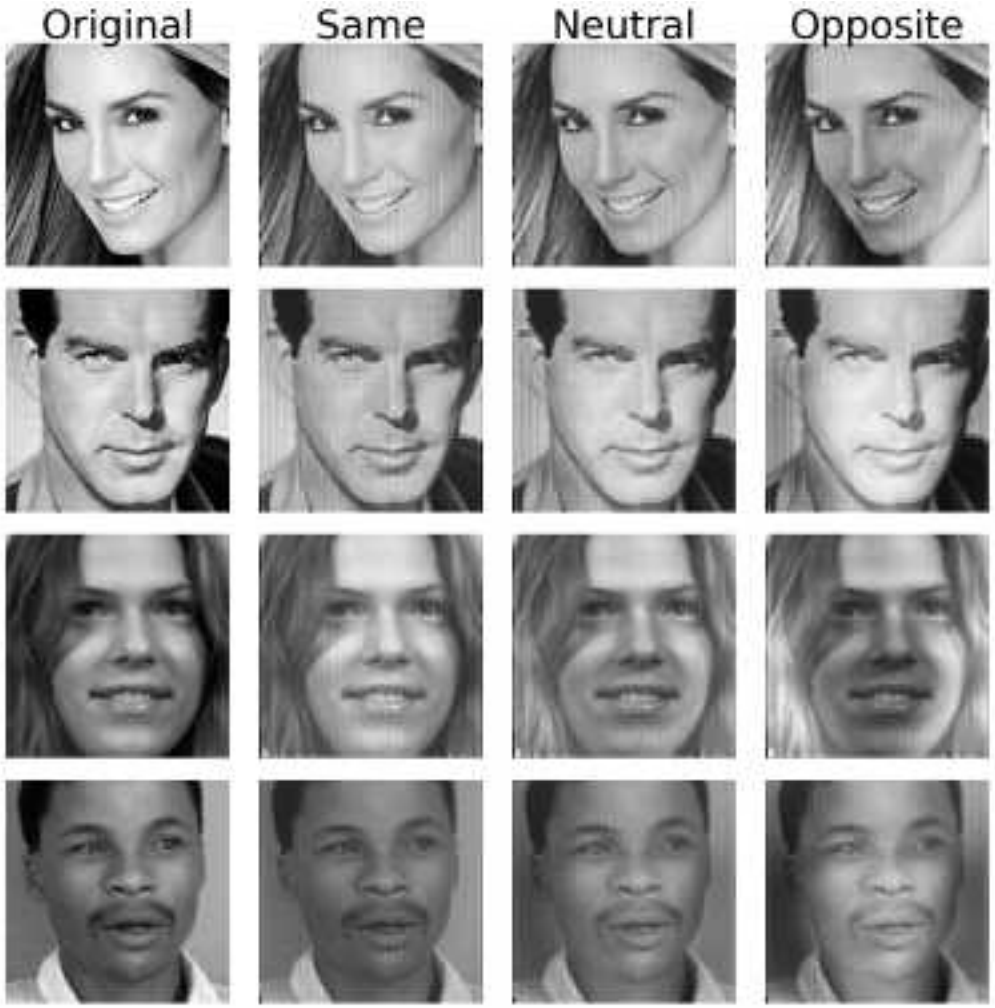
2. Loss term related gender attribute
   - Correctly predict gender of X'_{SM}
   - Flip the gender prediction of X'_{OP}

$$J_G(X, X'_{SM}, X'_{OP}, y; f_G) = S\left(y, f_G(X'_{SM})\right) + S\left(1 - y, f_G(X'_{OP})\right)$$

3. Loss related to matching

$$J_M(X, X'_{SM}; F_M) = \left|\left|F_M(X'_{SM}) - F_M(X)\right|\right|_2^2$$

Visual results



Visual results (improved)

Male: 99%

Female: 98%

Male: 97%

Male: 100%

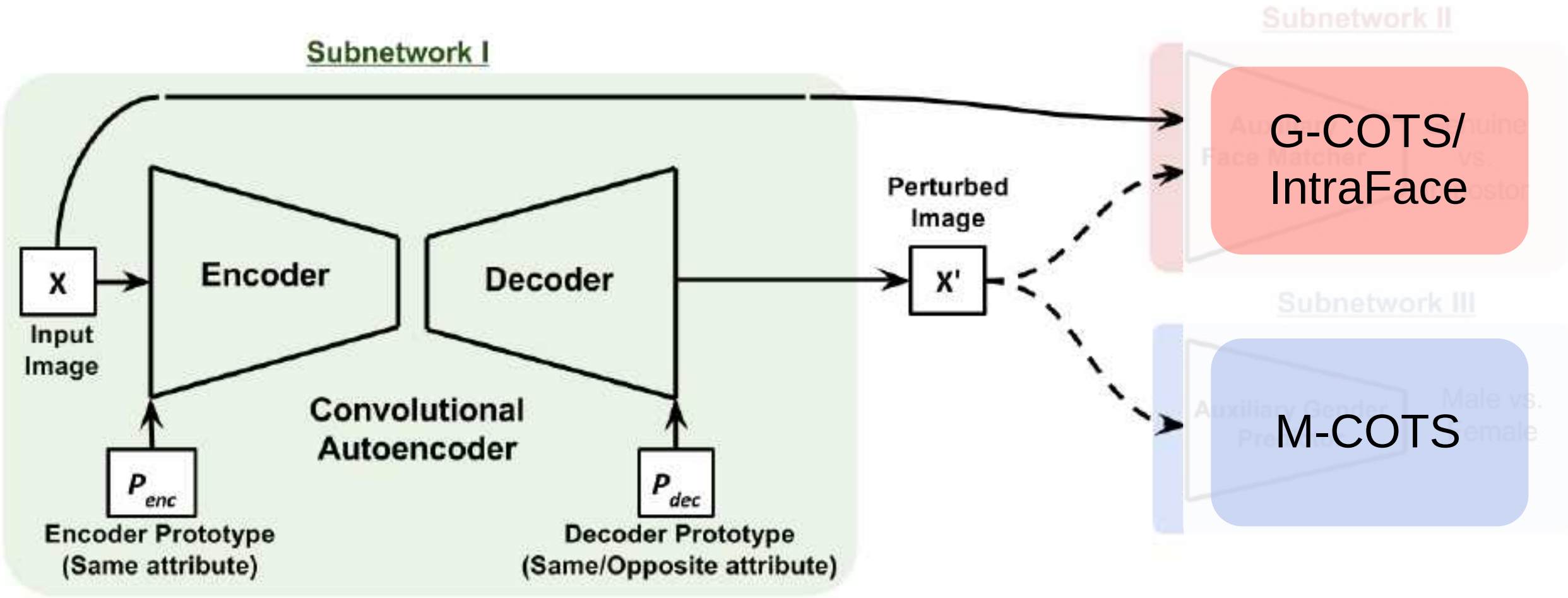
Female: 69%

Male: 99%

Female: 71%

Female: 58%

Replace detachable parts for evaluation



Datasets

	Dataset	Train	# Images	# Male	# Female
[1]	CelebA-train	yes	157,350	65,160	92,190
	CelebA-test	no	39,411	16,318	23,093
[2]	MUCT	no	3754	131	145
[3]	LFW	no	12,969	4205	1448
[4]	AR-face	no	3286	76	60

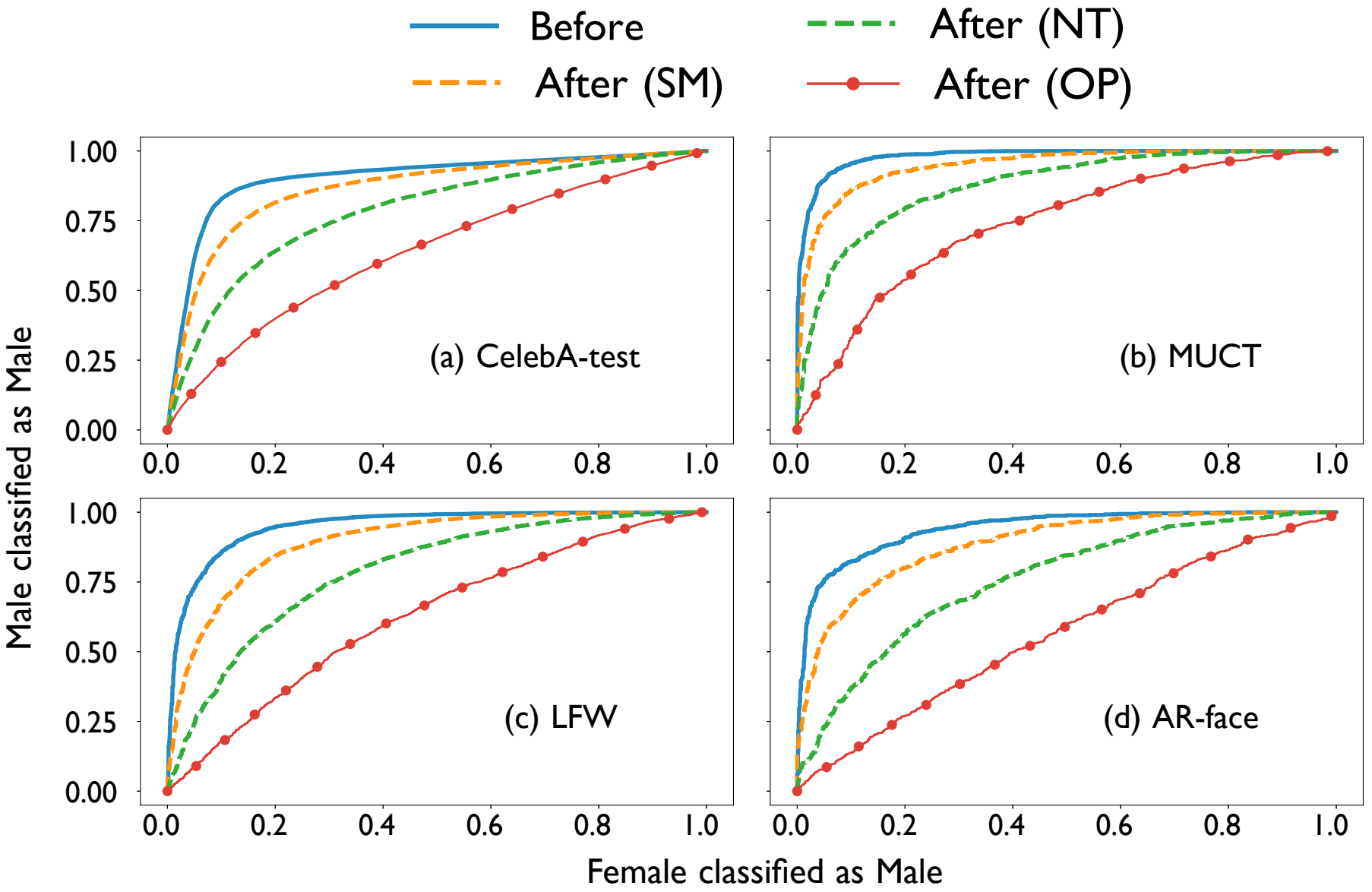
[1] Liu, Ziwei, et al. "Deep learning face attributes in the wild."
Proceedings of the IEEE International Conference on Computer Vision. 2015.

[2] Milborrow, Stephen, John Morkel, and Fred Nicolls. "The MUCT landmarked face database."
Pattern Recognition Association of South Africa 201.0 (2010).

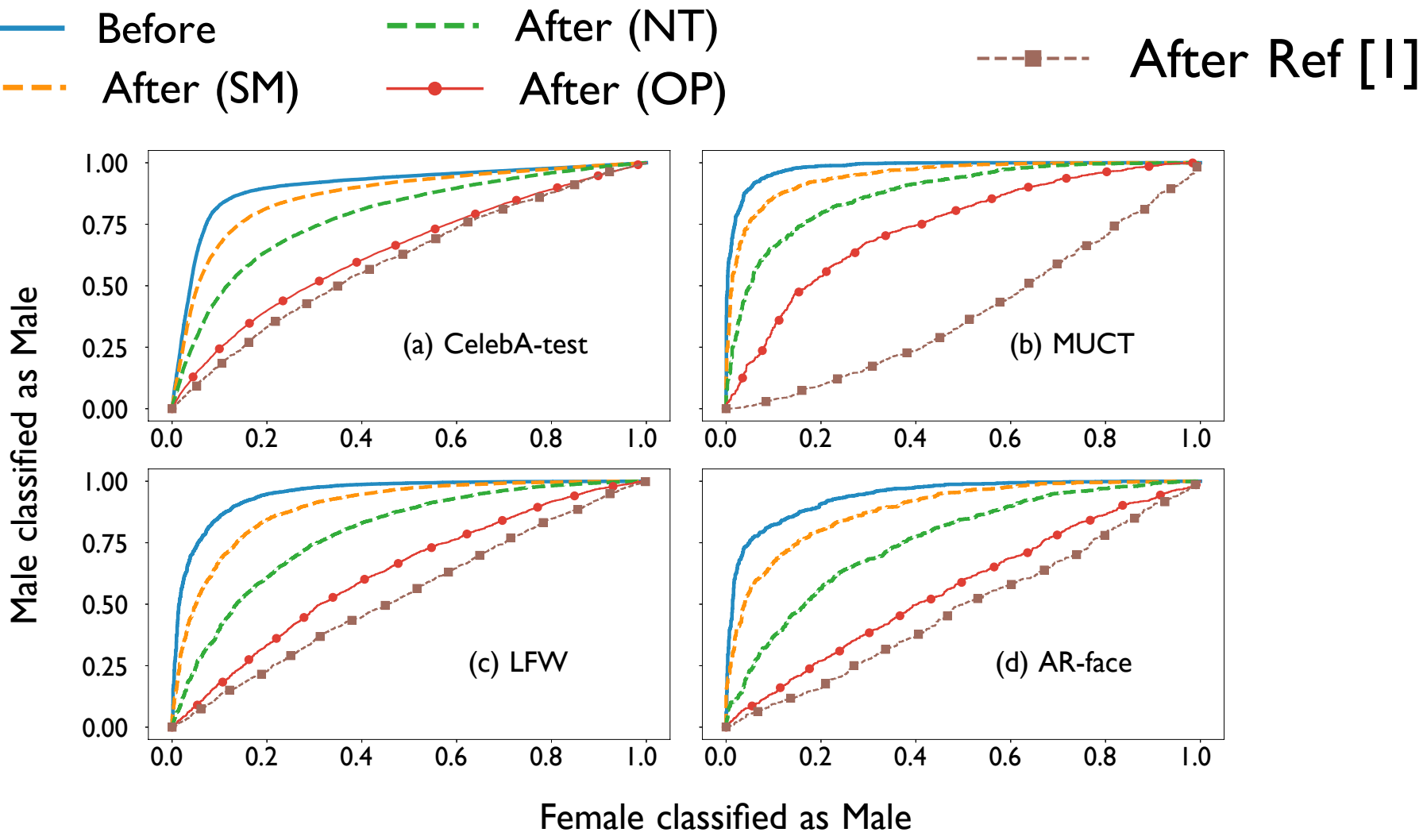
[3] Huang, Gary B., et al. *Labeled faces in the wild: A database for studying face recognition in unconstrained environments*.
Technical Report 07-49, University of Massachusetts, Amherst, 2007.

[4] Martinez, Aleix M. "The AR face database."
CVC Technical Report 24 (1998).

IntraFace gender classifier performance

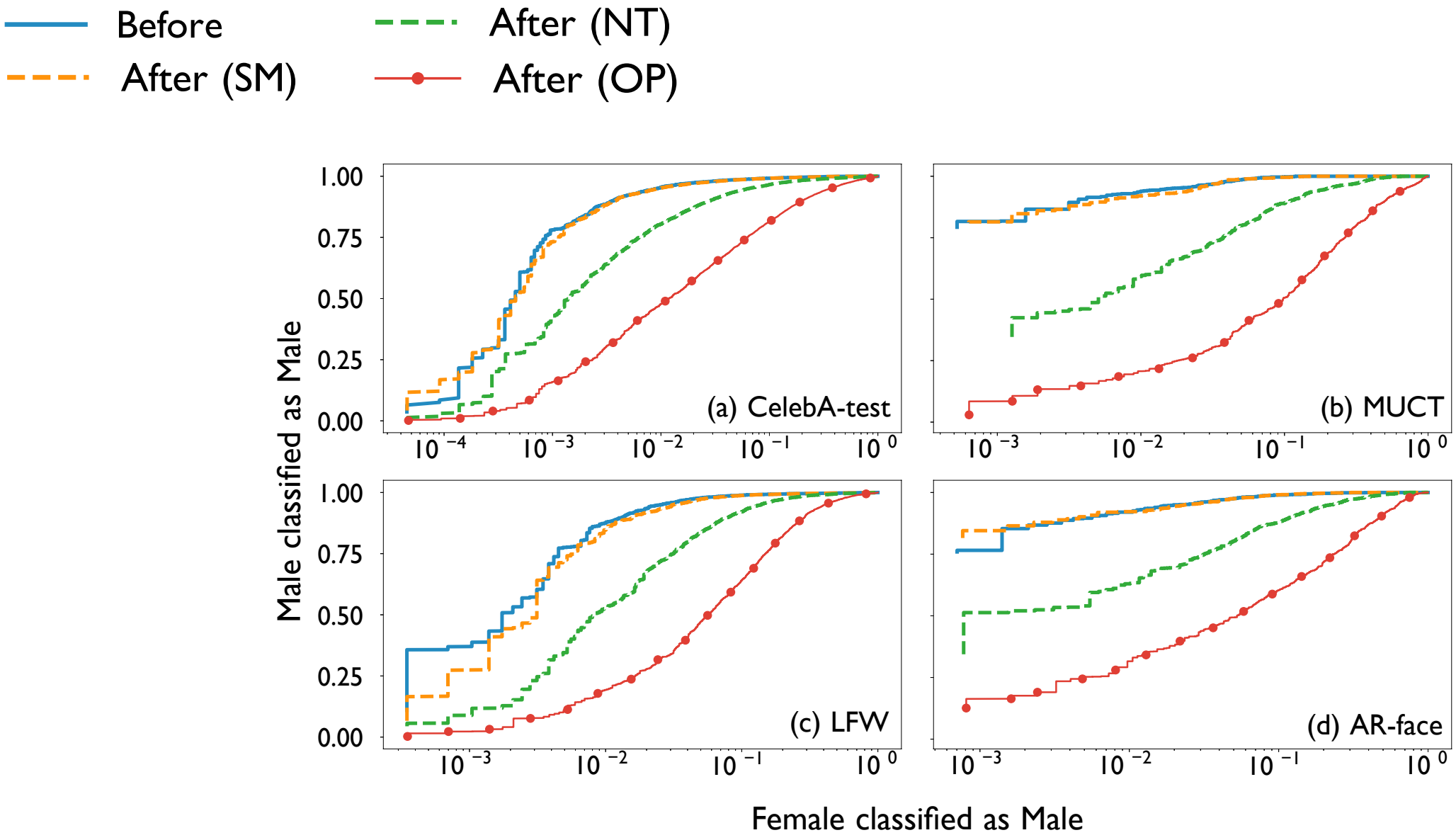


IntraFace gender classifier performance



[1] 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.

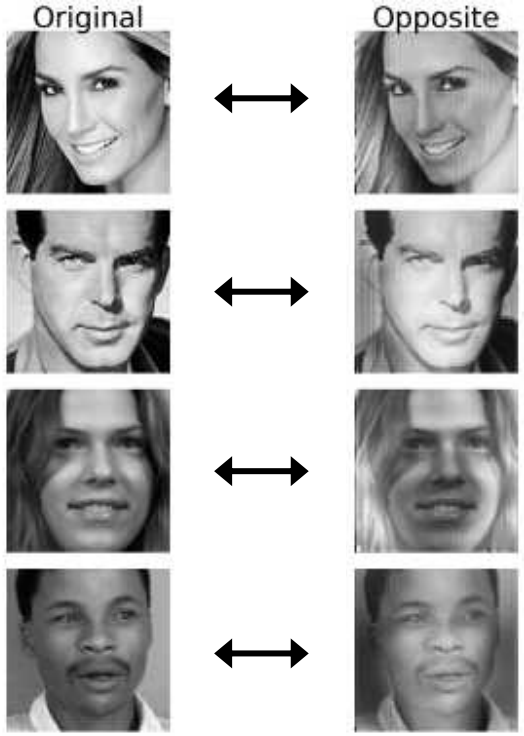
G-COTS gender classifier

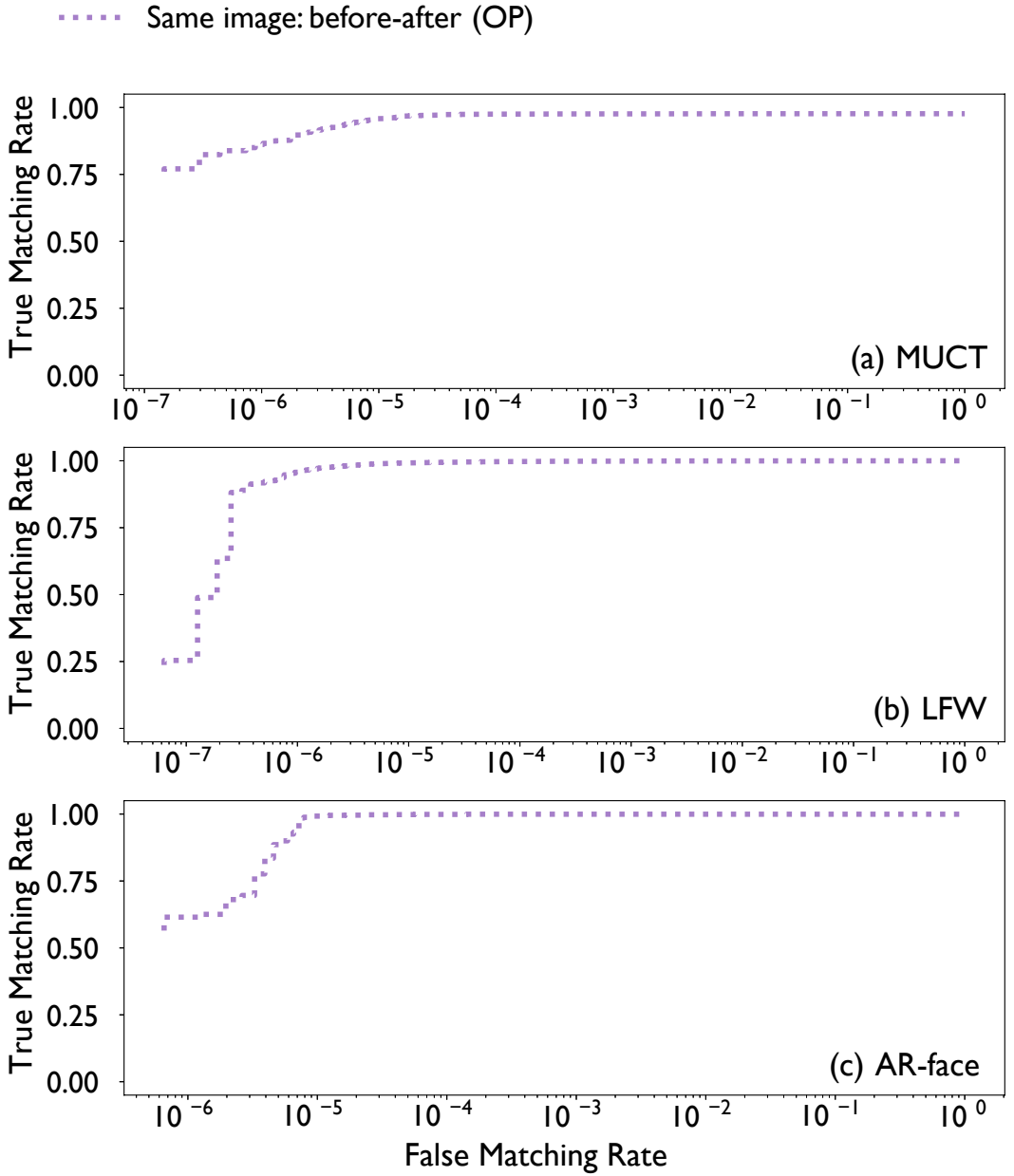


Gender classifier accuracy

Software	Dataset	Original (before)	Perturbed (after OP)
IntraFace	CelebA-test	19.7%	39.3%
	MUCT	8.0%	39.2%
	LFW	33.4%	72.5%
	AR-face	16.9%	53.8%
G-COTS	CelebA-test	2.2%	13.6%
	MUCT	5.1%	25.4%
	LFW	2.8%	18.8%
	AR-face	9.3%	26.9%

M-COTS face matcher performance





M-COTS face matcher performance

multi-subject comparisons

Before

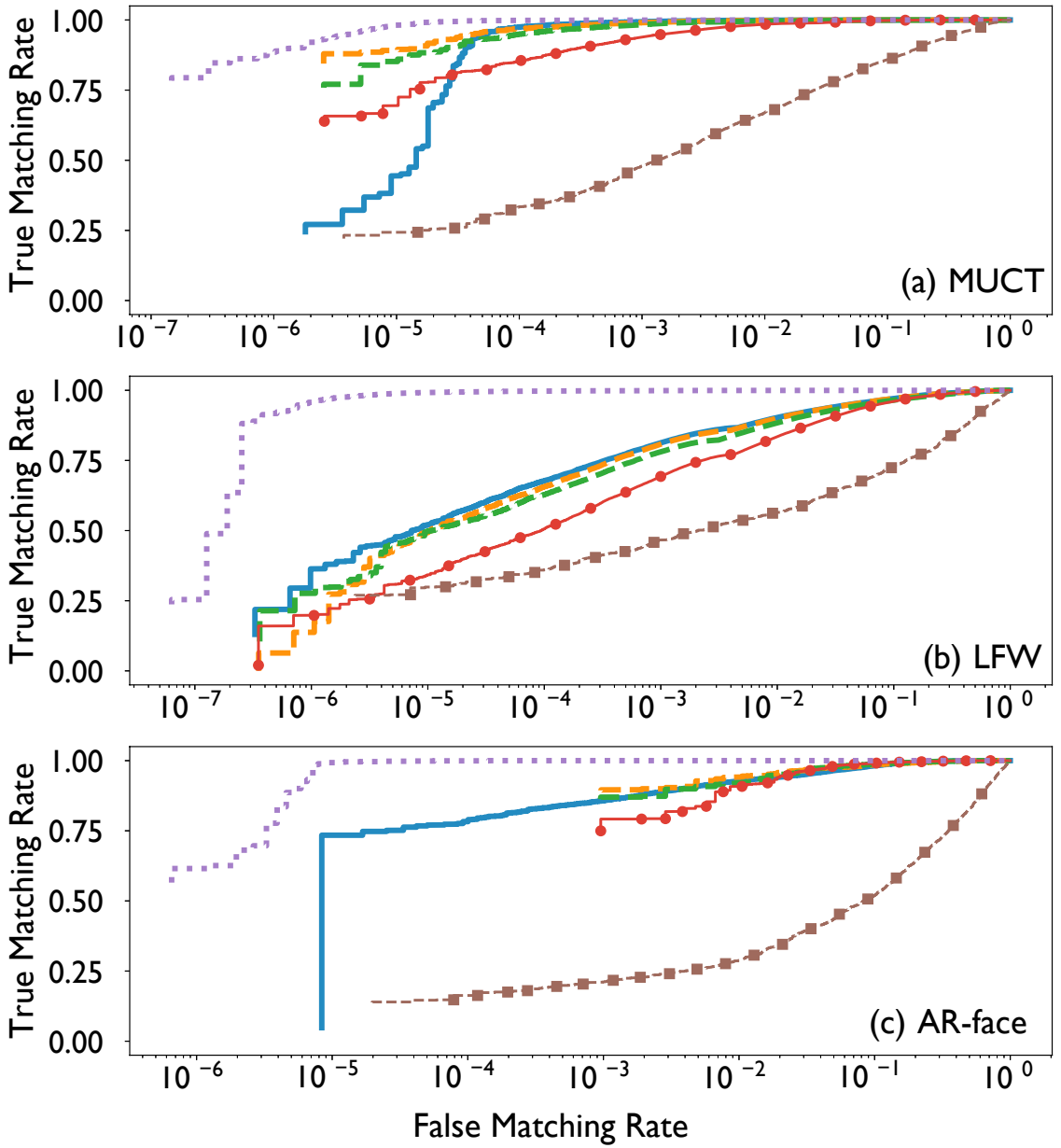
After (SM)

After (NT)

After (OP)

After Ref [1]

Same image: before-after (OP)



[1] 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.

M-COTS face matcher accuracy

Dataset	Original (before)	Perturbed (SM)	(NT)	(OP)
MUCT	99.88 %	99.79%	99.57%	98.44%
LFW	90.29%	90.02%	88.47%	83.45%
AR-face	94.97%	94.11%	91.95%	90.81%

Acknowledgements

Vahid Mirjalili

Anoop Namboodiri

Arun Ross

& MSU’s HPCC

Thanks for attending!

Questions?