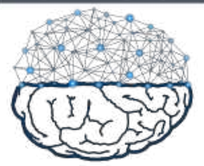


International Summer School on Deep Learning

July 1st-5th, 2019 in Gdansk, Poland

www.2019.dl-lab.eu

Convolutional Neural Networks for Predicting and Hiding Personal Traits from Face Images

Sebastian Raschka

Assistant Professor of Statistics at

University of Wisconsin-Madison

http://pages.stat.wisc.edu/~sraschka/

WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

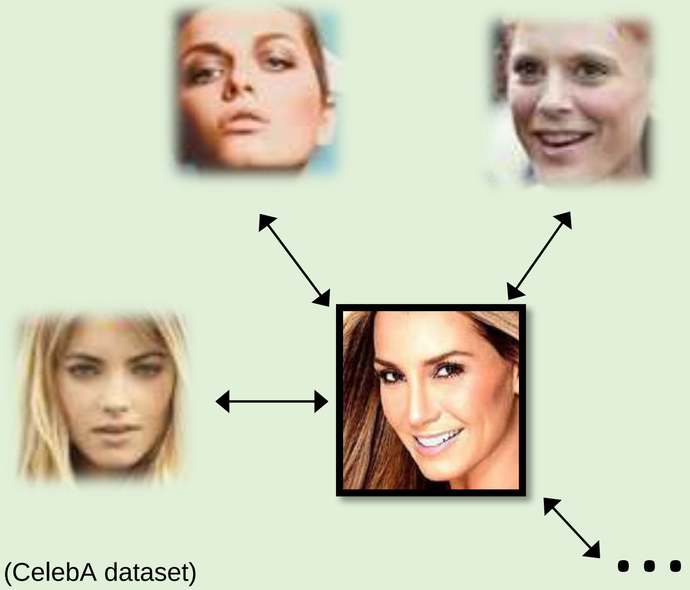
BIOMEDICAL ENGINEERING DEPARTMENT

GDAŃSK UNIVERSITY OF TECHNOLOGY

Biometric (Face) Recognition

A. Identification

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



(CelebA dataset)

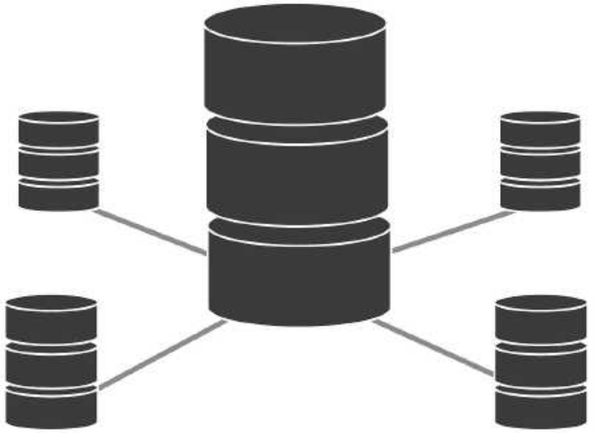
B. Verification

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

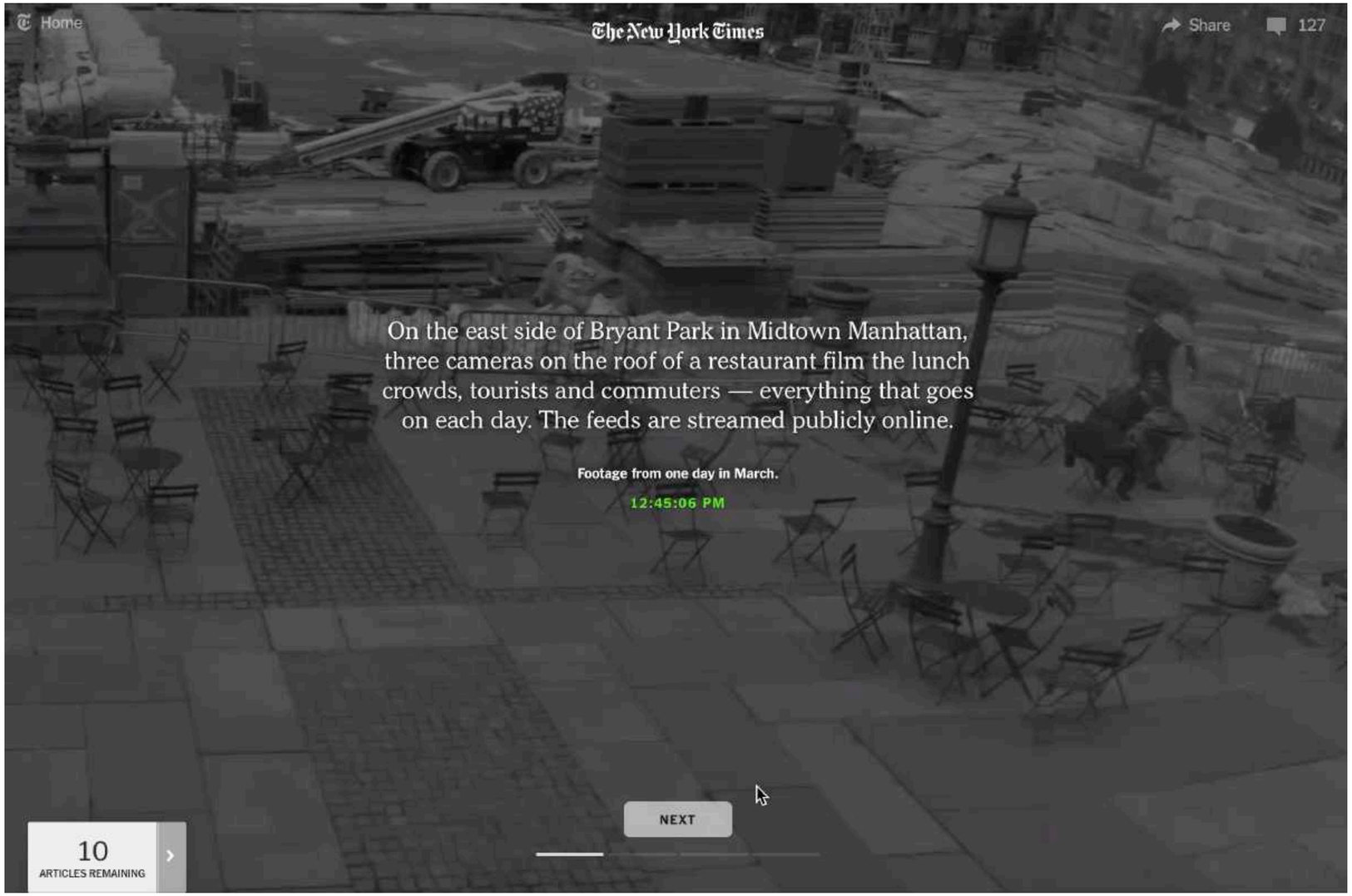
(MUCT dataset)

Applications of Biometric (Face) Recognition

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



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



https://www.nytimes.com/interactive/2019/04/16/opinion/facial-recognition-new-york-city.html

Soft-Biometrics

Identity	John Doe

Gender	Male
Age	65
Race	Caucasian
Medical	Healthy

SOFT BIOMETRIC ATTRIBUTES

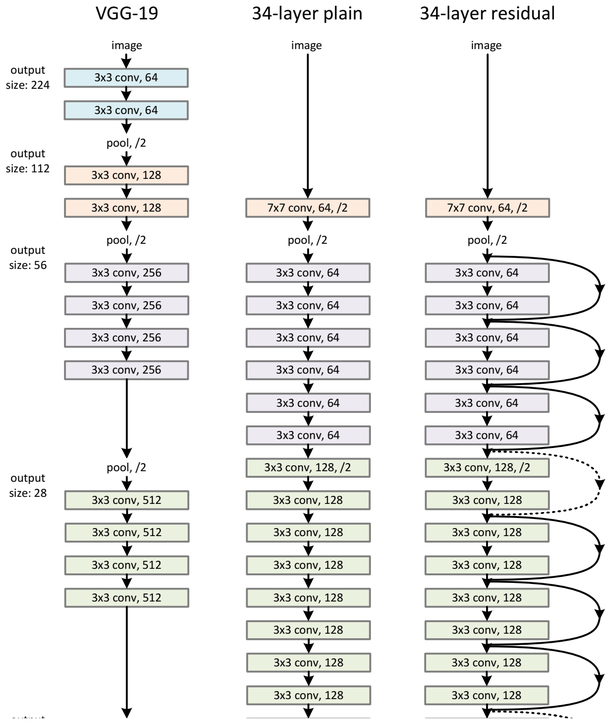
Part I: Extracting Soft-Biometric Attributes from Face Images

Identity	John Doe

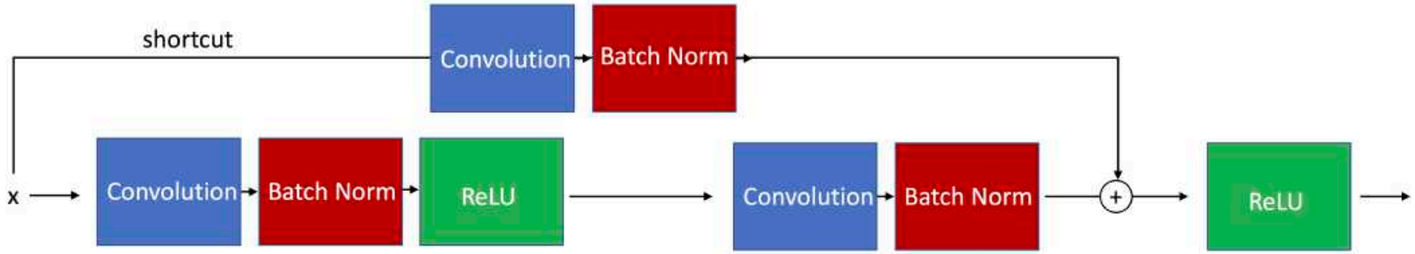
Gender	Male
Age	65
Race	Caucasian
Medical	Healthy

SOFT BIOMETRIC ATTRIBUTES

ResNet-101 Applied to Gender Classification



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




```
Epoch: 010/010 | Batch 1250/1272 | Cost: 0.0127
Epoch: 010/010 | Train: 99.183% | Valid: 98.017%
Time elapsed: 121.09 min
Total Training Time: 121.09 min
```

Evaluation

```
In [16]: with torch.set_grad_enabled(False): # save memory during inference
             print('Test accuracy: %.2f%%' % (compute_accuracy(model, test_loader, d
evice=DEVICE)))
```

```
Test accuracy: 97.52%
```

https://github.com/rasbt/deeplearning-models/blob/master/pytorch_ipynb/cnn/cnn-resnet101-celeba.ipynb

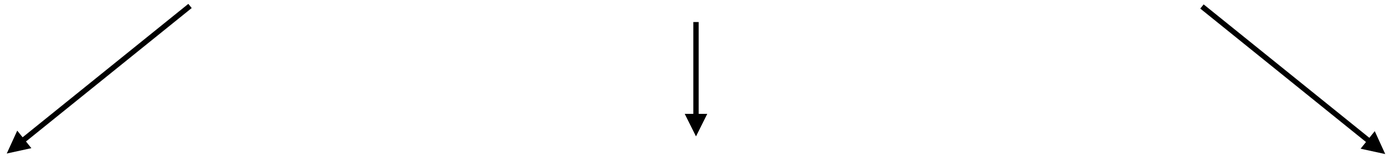
Identity	John Doe

Gender	Male
Age	65
Race	Caucasian
Medical	Healthy

SOFT BIOMETRIC ATTRIBUTES

Types of Labels in Supervised Learning Tasks

Color	Size	Price
green	M	10.1
red	L	13.5
blue	XXL	15.3



Nominal type
Task: classification

Ordinal type
Task: ordinal regression

Continuous
Task: metric regression

Ordinal Regression

Ordinal regression, also called *ordinal classification* or *ranking*
(although ranking is a bit different)

Order dependence like in metric regression,
but no metric distance

discrete values like in classification,
but order dependence/information

$$r_K \succ r_{K-1} \succ \ldots \succ r_1$$

E.g., movie ratings: *great* $\succ$ *good* $\succ$ *okay* $\succ$ *for genre fans* $\succ$ *bad*

★★★★★ ★★★★ ★★★ ★★ ★

Supervised Learning: Ordinal Regression

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

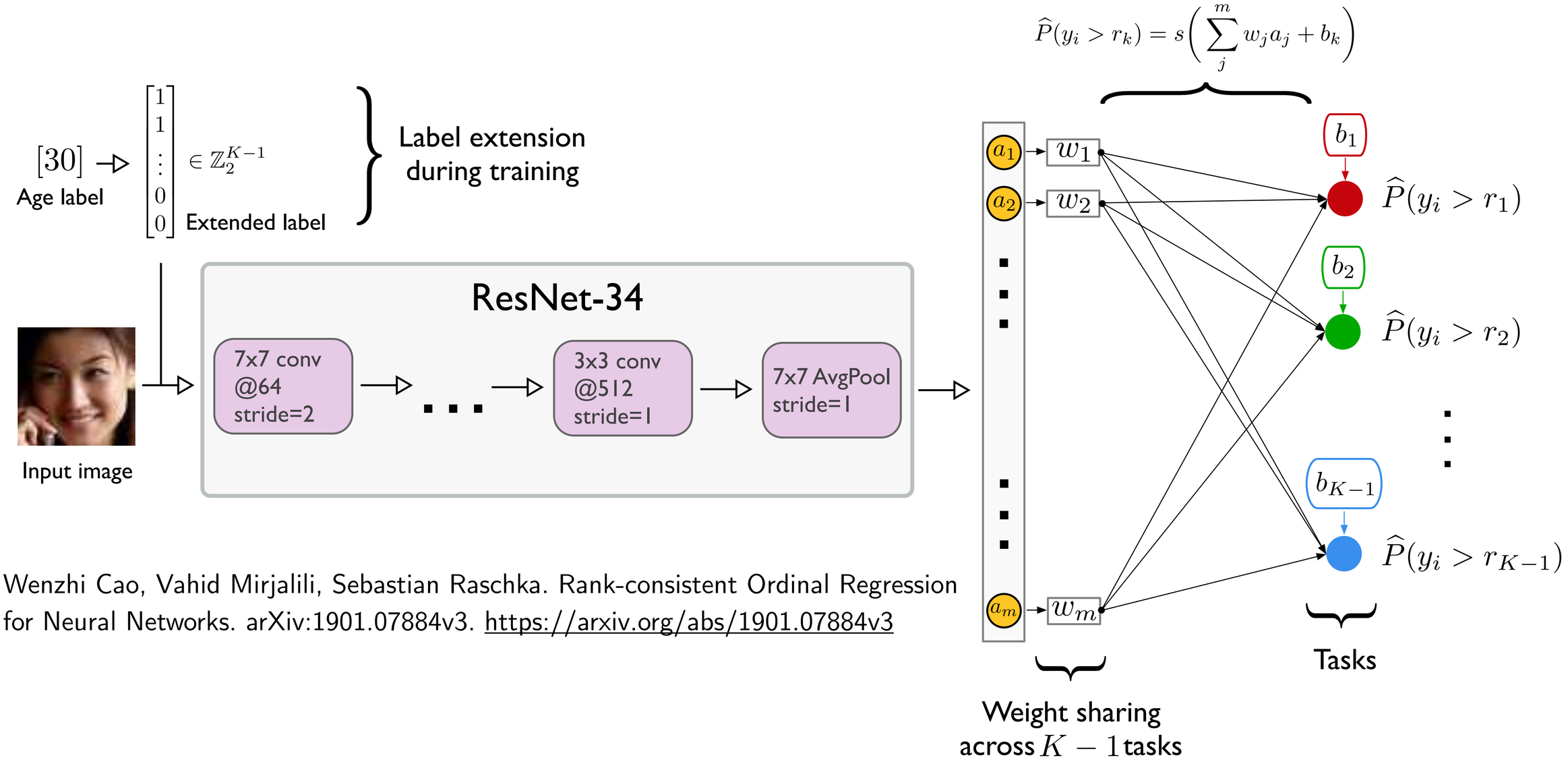
- **Ordinal Regression:** Correct label matters (as well)
  (E.g., age of a person in years; here, regard aging as a non-stationary process)

18 ≻ 29 ≻ 41

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

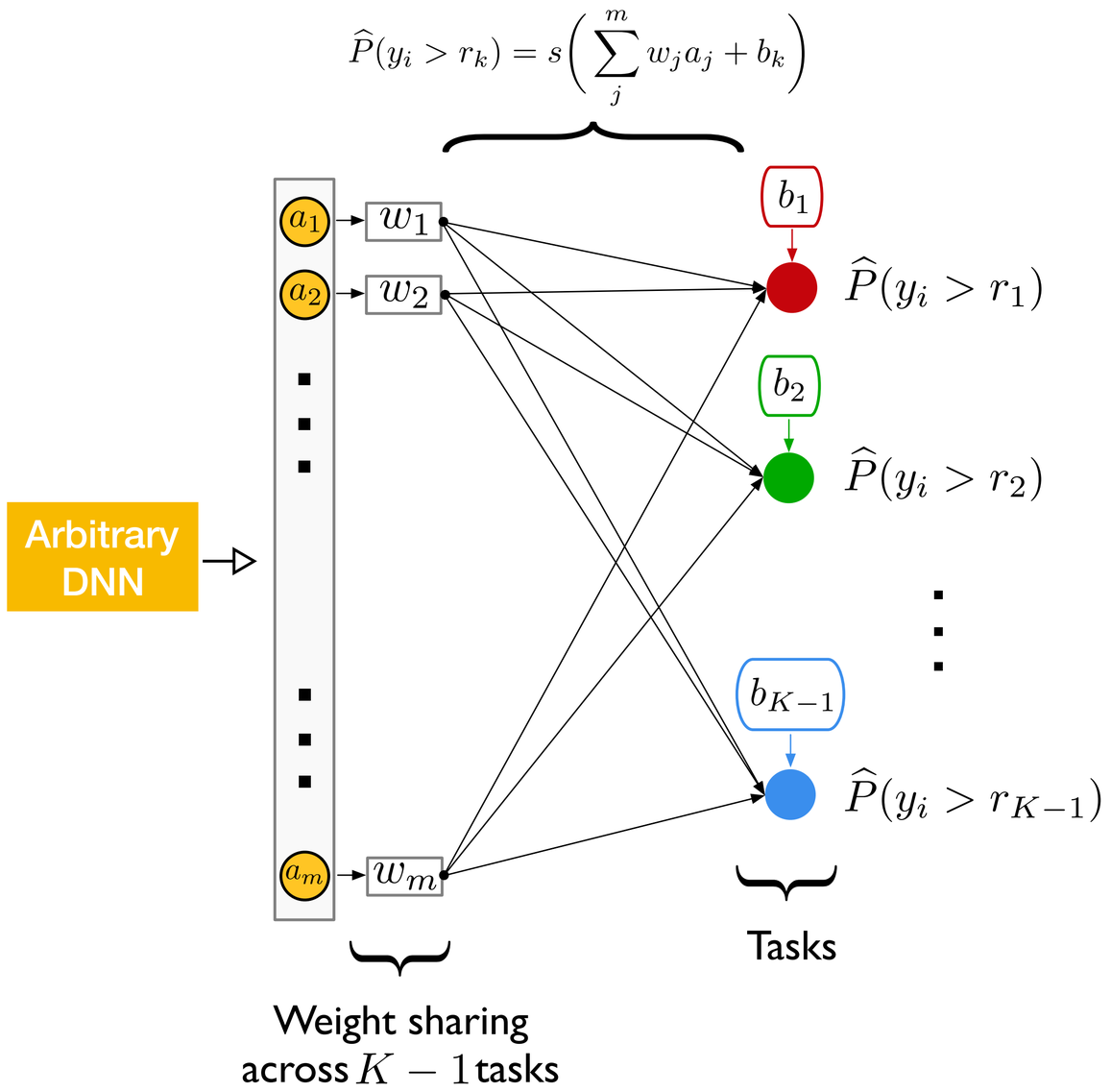
We will work with this dataset
in the hands-on tutorial this afternoon!

Consistent Rank Logits (CORAL) for Ordinal Regression with Convolutional Neural Networks Framework



Wenzhi Cao, Vahid Mirjalili, Sebastian Raschka. Rank-consistent Ordinal Regression for Neural Networks. arXiv:1901.07884v3. https://arxiv.org/abs/1901.07884v3

Predicting the Rank/Ordinal Label



Step 1: Converting the estimated probability of each task into a binary label (0/1)

$$f_k\left(\mathbf{x}_i\right) = 1\left\{\widehat{P}\left(y_i^{(k)} = 1\right) > 0.5\right\}$$

Step 2: Summing the K-1 binary labels

$$h\left(\mathbf{x}_i\right) = r_q$$

$$q = 1 + \sum_{k=1}^{K-1} f_k\left(\mathbf{x}_i\right)$$

Hypothesis: Rank Consistency Improves Predictive Performance

Desired property: probability estimates for the *K-1* tasks are decreasing

$$\widehat{P}\left(y_i^{(1)} = 1\right) \geq \widehat{P}\left(y_i^{(2)} = 1\right) \geq \ldots \geq \widehat{P}\left(y_i^{(K-1)} = 1\right) \quad \text{[Rank consistency]}$$

where the predicted empirical probability for task k is defined as

$$\widehat{P}\left(y_i^{(k)} = 1\right) = s\left(g\left(\mathbf{x}_i, \mathbf{W}\right) + b_k\right)$$

Hypothesis: Rank Consistency Improves Predictive Performance

Desired property: probability estimates for the *K-1* tasks are decreasing

$$\widehat{P}\left(y_i^{(1)} = 1\right) \geq \widehat{P}\left(y_i^{(2)} = 1\right) \geq \ldots \geq \widehat{P}\left(y_i^{(K-1)} = 1\right)$$

Rank Consistency can be satisfied by ordered bias units due to the weight constraint

$$b_1 \geq b_2 \geq \cdots \geq b_{K-1}$$

Theorem 1 (ordered biases)**.** *By minimizing loss function defined in Eq.* (4)*, the optimal solution* $(\mathbf{W}^*, \mathbf{b}^*)$ *satisfies* $b_1^* \geq b_2^* \geq \ldots \geq b_{K-1}^*$.

(Detailed proof provided in our paper)

Wenzhi Cao, Vahid Mirjalili, Sebastian Raschka. Rank-consistent Ordinal Regression for Neural Networks. arXiv:1901.07884v3. https://arxiv.org/abs/1901.07884v3

Niu, Z., Zhou, M., Wang, L., Gao, X., & Hua, G. (2016). Ordinal Regression with Multiple Output CNN for Age Estimation. CVPR.

Cao, W., Mirjalili V., Raschka S. (2019). Rank-consistent Ordinal Regression for Neural Networks. arXiv: 1901.07884v3. https://arxiv.org/abs/1901.07884v3

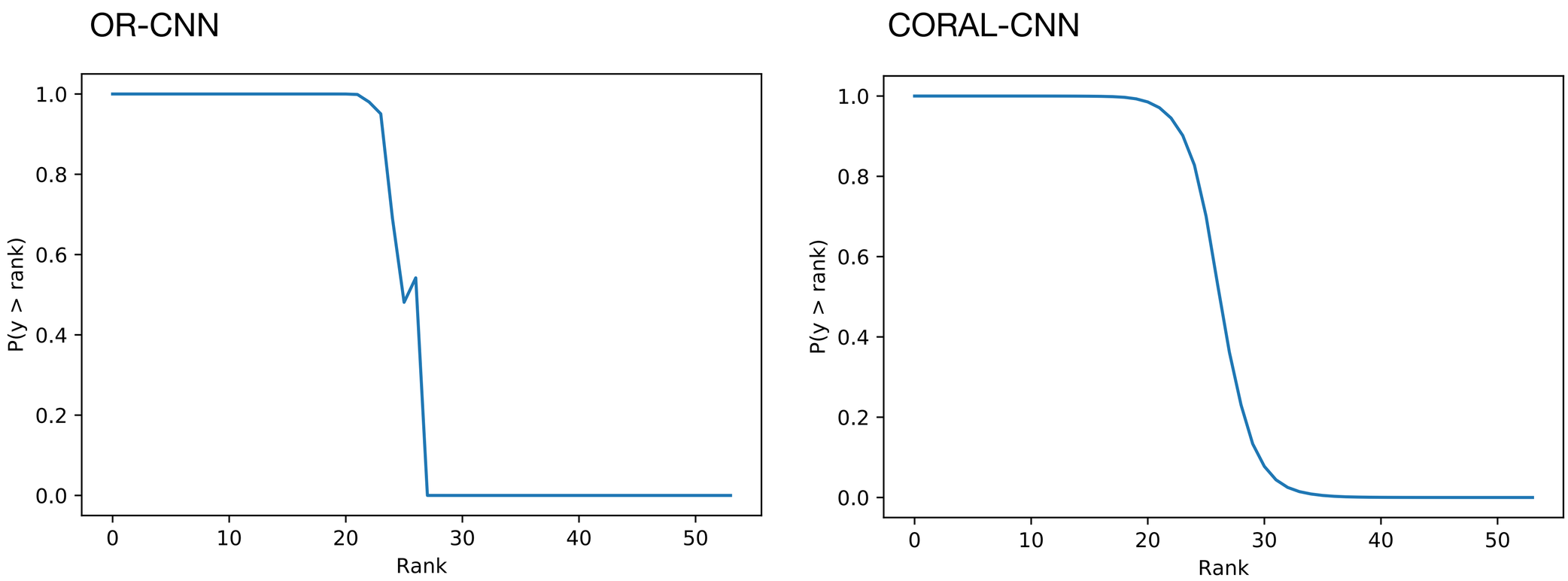
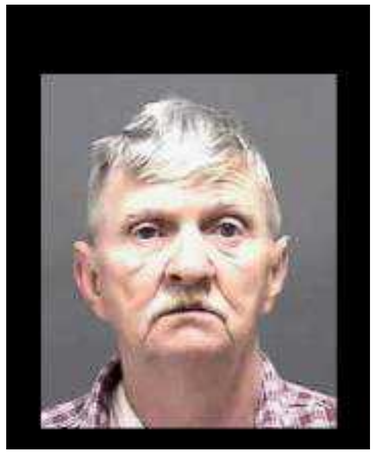


Figure S2: Plots show graphs of the predicted probabilities for each binary classifier task on one test data point in the MORPH dataset by OR-CNN (left subpanel) and CORAL-CNN (right subpanel). In this example, the ordinal regression CNN has an inconsistency at rank 26. The CORAL-CNN does not suffer from inconsistencies such that the rank prediction is a cumulative distribution function.

Datasets

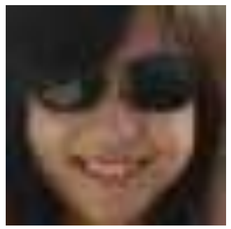


MORPH-2

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

AFAD

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



CACD

- 159,449 face images
- age range: 14-62 years

UTKFace

- 16,434 face images
- age range: 21-60 years

Test Results

$$\text{MAE} = \frac{1}{N}\sum_{i=1}^{N}\left|y_i - h(\mathbf{x}_i)\right| \qquad \text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(y_i - h(\mathbf{x}_i)\right)^2}$$

Table 1. Age prediction errors on the test sets *without* task importance weighting. All models are based on the ResNet-34 architecture.

Method	Random Seed	MORPH-2		AFAD		UTKFace		CACD	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
CE-CNN	0	3.40	4.88	3.98	5.55	6.57	9.16	6.18	8.86
	1	3.39	4.87	4.00	5.57	6.24	8.69	6.10	8.79
	2	3.37	4.87	3.96	5.50	6.29	8.78	6.13	8.87
	AVG ± SD	3.39 ± 0.02	4.89 ± 0.01	3.98 ± 0.02	5.54 ± 0.04	6.37 ± 0.18	8.88 ± 0.25	6.14 ± 0.04	8.84 ± 0.04
OR-CNN (Niu et al., 2016)	0	2.98	4.26	3.66	5.10	5.71	8.11	5.53	7.91
	1	2.98	4.26	3.69	5.13	5.80	8.12	5.53	7.98
	2	2.96	4.20	3.68	5.14	5.71	8.11	5.49	7.89
	AVG ± SD	2.97 ± 0.01	4.24 ± 0.03	3.68 ± 0.02	5.13 ± 0.02	5.74 ± 0.05	8.08 ± 0.06	5.52 ± 0.02	7.93 ± 0.05
CORAL-CNN (ours)	0	2.68	3.75	3.49	4.82	5.46	7.61	5.56	7.80
	1	2.63	3.66	3.46	4.83	5.46	7.63	5.37	7.64
	2	2.61	3.64	3.52	4.91	5.48	7.63	5.25	7.53
	AVG ± SD	**2.64 ± 0.04**	**3.68 ± 0.06**	**3.49 ± 0.03**	**4.85 ± 0.05**	**5.47 ± 0.01**	**7.62 ± 0.01**	**5.39 ± 0.16**	**7.66 ± 0.14**

MORPH-2

55,608 face images

age range: 16-70 years

AFAD

165,501 face images

age range: 15-40 years

UTKFace

16,434 face images

age range: 21-60 years

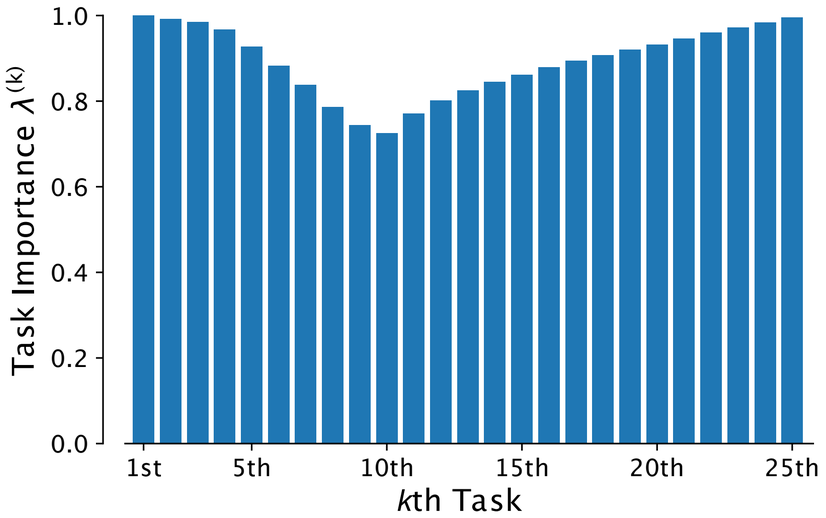
CACD

159,449 face images

age range: 14-62 years

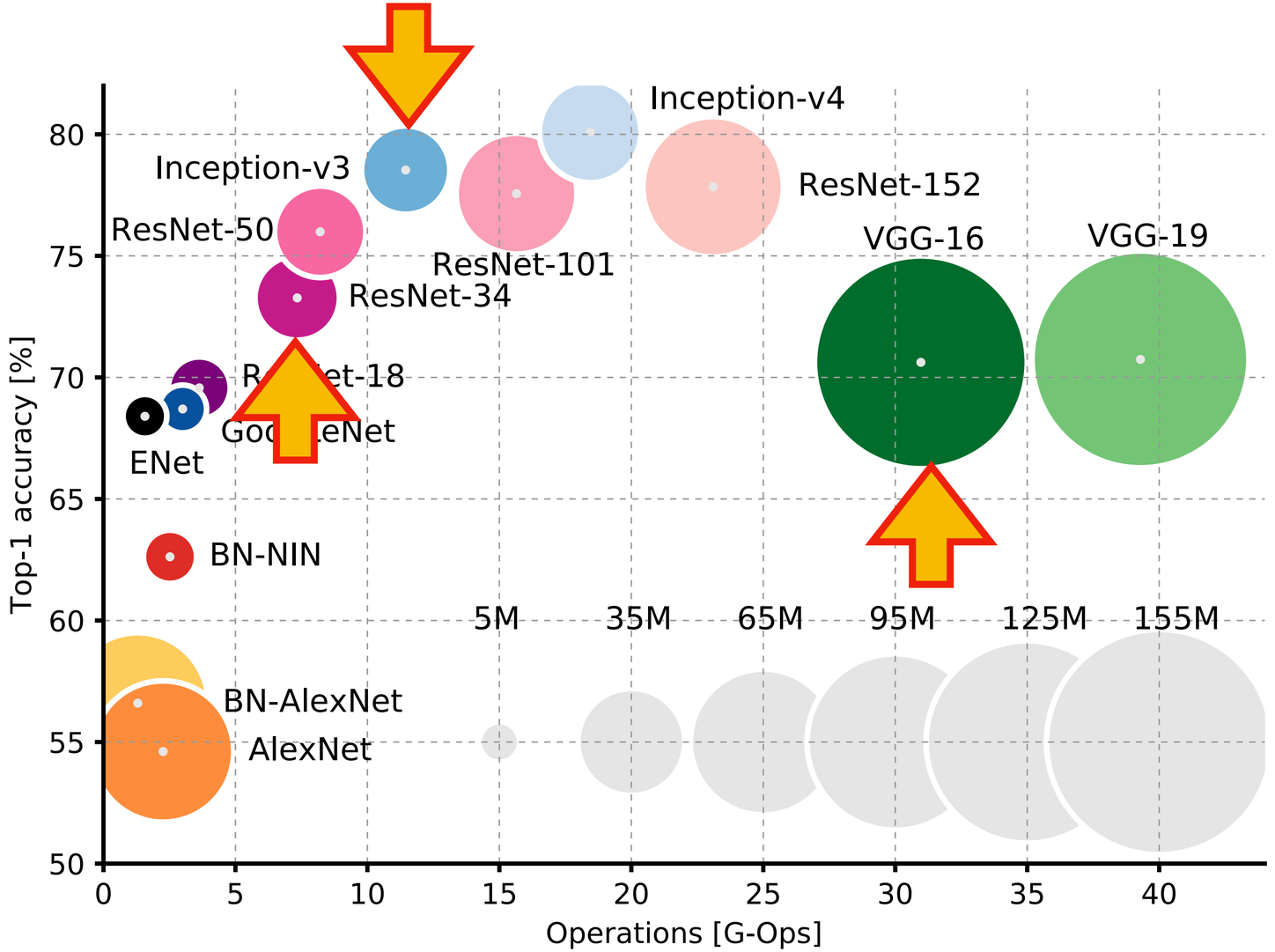
Test Results

With optional **task importance** weighting



Method	Weight	MORPH-2	AFAD	UTKFace	CACD
OR-CNN (Niu et al., 2016)	NO	2.97 ± 0.01	3.68 ± 0.02	5.74 ± 0.05	5.52 ± 0.02
OR-CNN (Niu et al., 2016)	YES	2.91 ± 0.02	3.65 ± 0.03	5.76 ± 0.19	5.49 ± 0.02
CORAL-CNN (ours)	NO	2.64 ± 0.04	3.49 ± 0.03	5.47 ± 0.01	5.39 ± 0.16
CORAL-CNN (ours)	YES	$\mathbf{2.59 \pm 0.03}$	$\mathbf{3.48 \pm 0.03}$	$\mathbf{5.39 \pm 0.07}$	$\mathbf{5.35 \pm 0.09}$

Figure 1. Example of the task importance weighting according to Eq. (7) shown for the AFAD dataset (Section 4.1).



Alfredo Canziani, Adam Paszke, and Eugenio Culurciello. "An analysis of deep neural network models for practical applications." *arXiv preprint arXiv:1605.07678* (2016).

VGG-16

Method	Random Seed	MORPH-2		AFAD		UTKFace		CACD	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
CE-CNN	0	14.07	17.54	3.84	5.38	6.32	8.72	5.73	7.82
	1	14.07	17.54	3.87	5.39	6.26	8.55	5.66	7.68
	2	3.71	5.15	3.93	5.45	6.33	8.8	5.81	7.89
	AVG ± SD	10.62 ± 5.98	13.41 ± 7.15	3.88 ± 0.05	5.41 ± 0.04	6.30 ± 0.04	8.69 ± 0.13	5.73 ± 0.08	7.80 ± 0.11
OR-CNN [13]	0	2.75	3.82	3.53	4.95	6.42	8.60	5.31	7.47
	1	2.92	4.08	3.55	5.00	6.25	8.33	5.28	7,47
	2	2.95	4.14	3.72	5.23	6.50	8.81	5.39	7.52
	AVG ± SD	2.87 ± 0.11	4.01 ± 0.17	3.60 ± 0.10	5.06 ± 0.15	6.39 ± 0.13	8.58 ± 0.24	5.33 ± 0.06	7.49 ± 0.03
CORAL-CNN (ours)	0	2.76	3.73	3.45	4.78	5.95	8.28	5.25	7.49
	1	2.79	3.74	3.39	4.72	5.59	7.6	5.21	7.42
	2	2.87	3.94	3.4	4.75	5.96	8.22	5.28	7.48
	AVG ± SD	**2.81 ± 0.06**	**3.80 ± 0.12**	**3.41 ± 0.03**	**4.75 ± 0.03**	**5.83 ± 0.21**	**8.03 ± 0.38**	**5.25 ± 0.04**	**7.46 ± 0.04**

Inception-v3

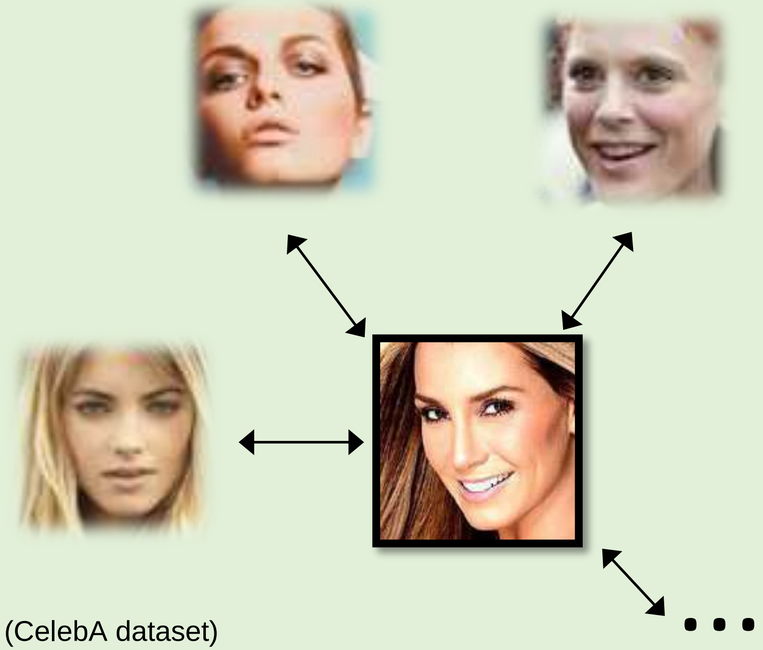
Method	Random Seed	MORPH-2		AFAD		UTKFace		CACD	
		MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
CE-CNN	0	3.07	4.39	3.78	5.30	6.73	9.47	5.52	8.18
	1	3.00	4.35	3.77	5.31	6.52	9.08	5.46	8.09
	2	3.00	4.36	3.79	5.33	6.81	9.4	5.44	8.04
	AVG ± SD	3.02 ± 0.04	4.37 ± 0.02	5.31 ± 0.02	3.78 ± 0.01	6.69 ± 0.15	9.32 ± 0.21	5.47 ± 0.04	8.10 ± 0.07
OR-CNN [13]	0	2.52	3.59	3.42	4.84	5.74	7.89	4.98	7.43
	1	2.57	3.69	3.45	4.87	5.49	7.58	4.93	7.37
	2	2.51	3.60	3.36	4.75	5.41	7.46	4.94	7.33
	AVG ± SD	2.53 ± 0.03	3.63 ± 0.06	3.41 ± 0.05	4.82 ± 0.06	5.55 ± 0.17	**7.64 ± 0.22**	4.95 ± 0.03	7.38 ± 0.05
CORAL-CNN (ours)	0	2.45	3.41	3.28	4.59	5.57	7.72	4.92	7.16
	1	2.41	3.36	3.32	4.63	5.26	7.3	4.91	7.21
	2	2.43	3.39	3.20	4.59	5.76	7.95	4.87	7.11
	AVG ± SD	**2.43 ± 0.02**	**3.39 ± 0.03**	**3.27 ± 0.06**	**4.60± 0.02**	**5.53 ± 0.25**	7.66 ± 0.33	**4.90 ± 0.03**	**7.16 ± 0.05**

Part II: Hiding Soft-Biometric Attributes from Face Images

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)

Identity	John Doe

Gender	Male
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

Goal: Selective Privacy

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

Face Matcher

100101010101111
101010101001000
001101010101010
101001100101010
010111010001010
101110001010101
101001010100101

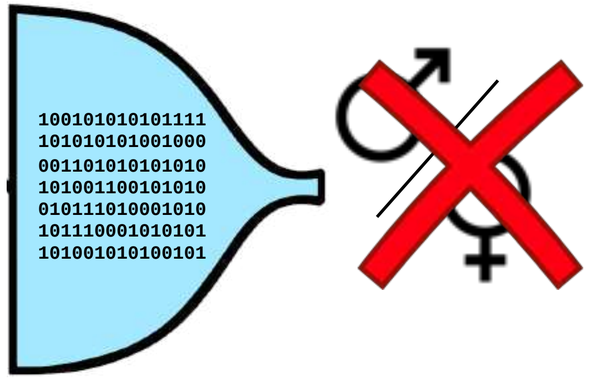
p("same person")

100101010101111
101010101001000
001101010101010
101001100101010
010111010001010
101110001010101
101001010100101

p("male")

Gender Classifier

Face Matcher



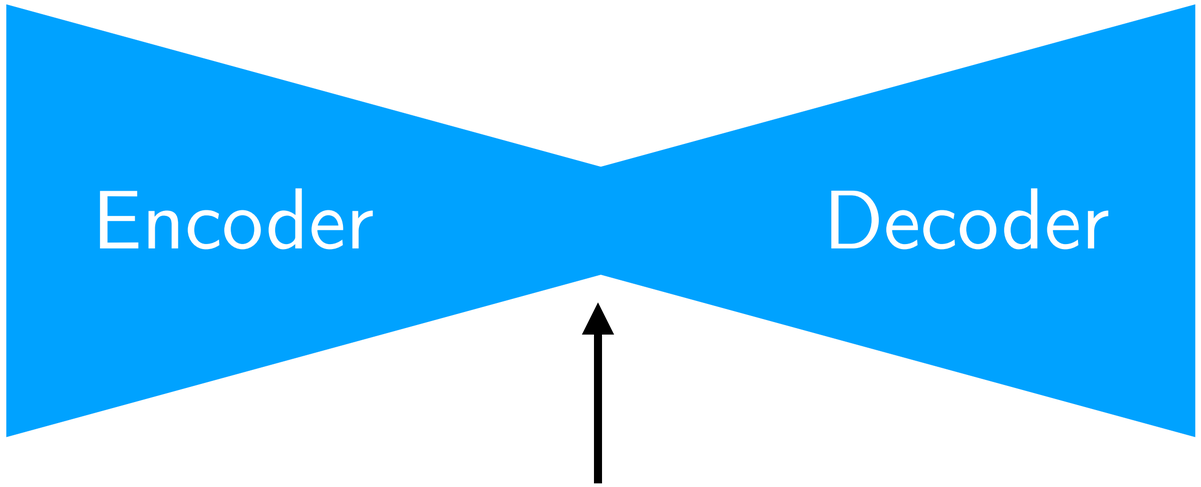
Gender Classifier

Autoencoders

Unsupervised Learning:

Representation Learning/Dimensionality Reduction

Source: https://3.bp.blogspot.com/-OUd11VBJNAM/VsFacR_YhBI/AAAAAAAABh0/ZKfKAnRj3x0/s1600/cannot%2Bresist.jpg



Latent representation/feature embedding

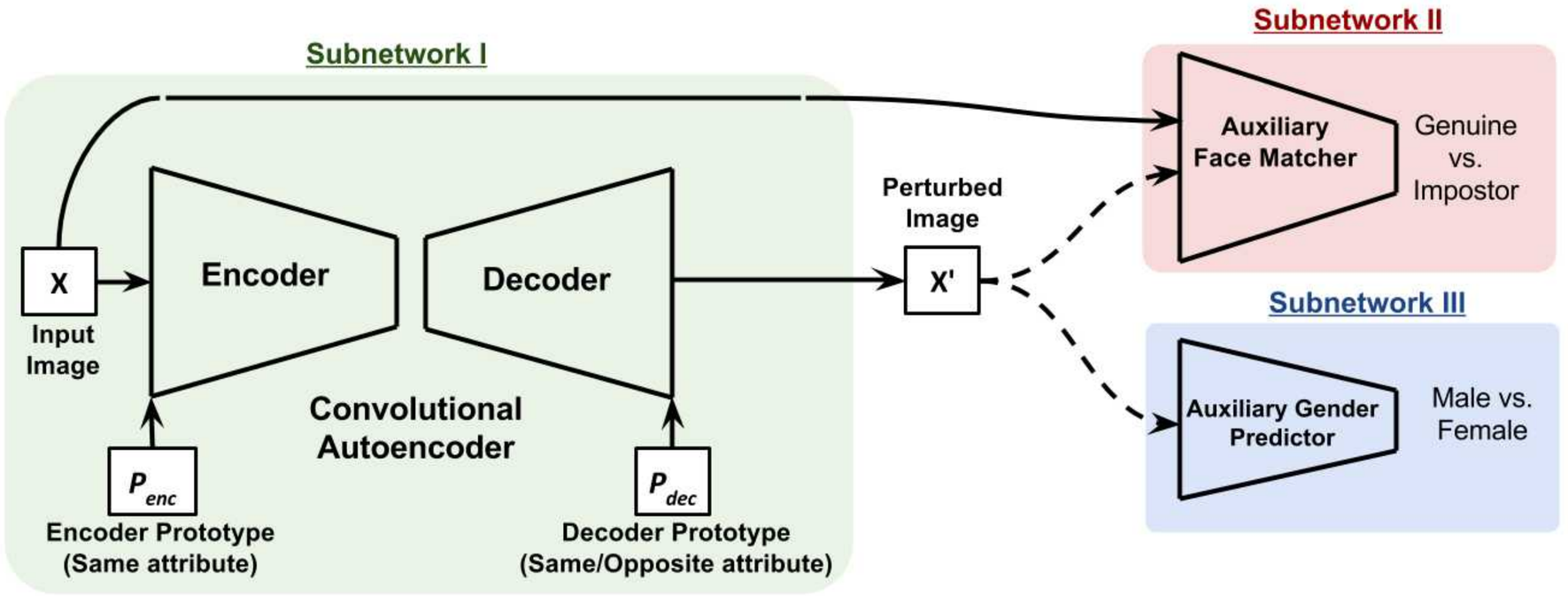
Autoencoder to perturb image

$$\phi(\mathbf{X}) = \mathbf{X}'$$

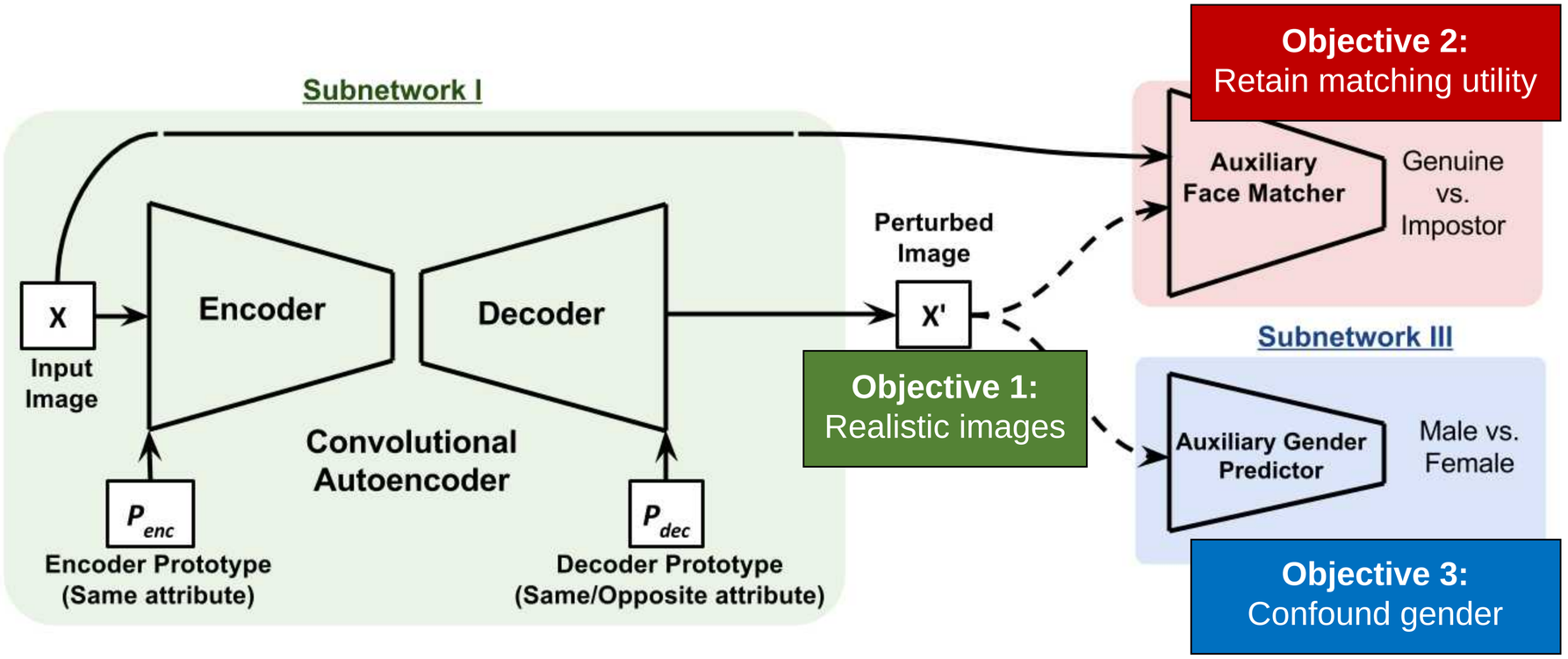
Face Matcher

Gender Classifier

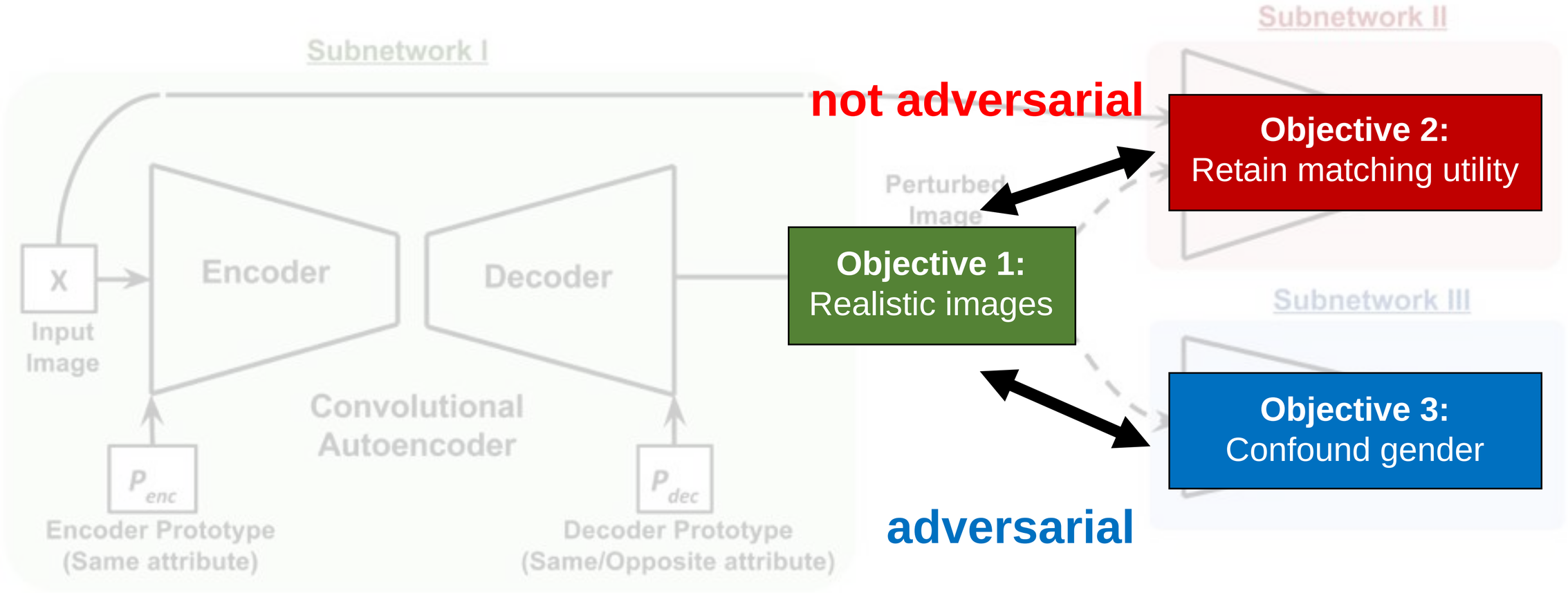
General architecture of the semi-adversarial network (SAN)



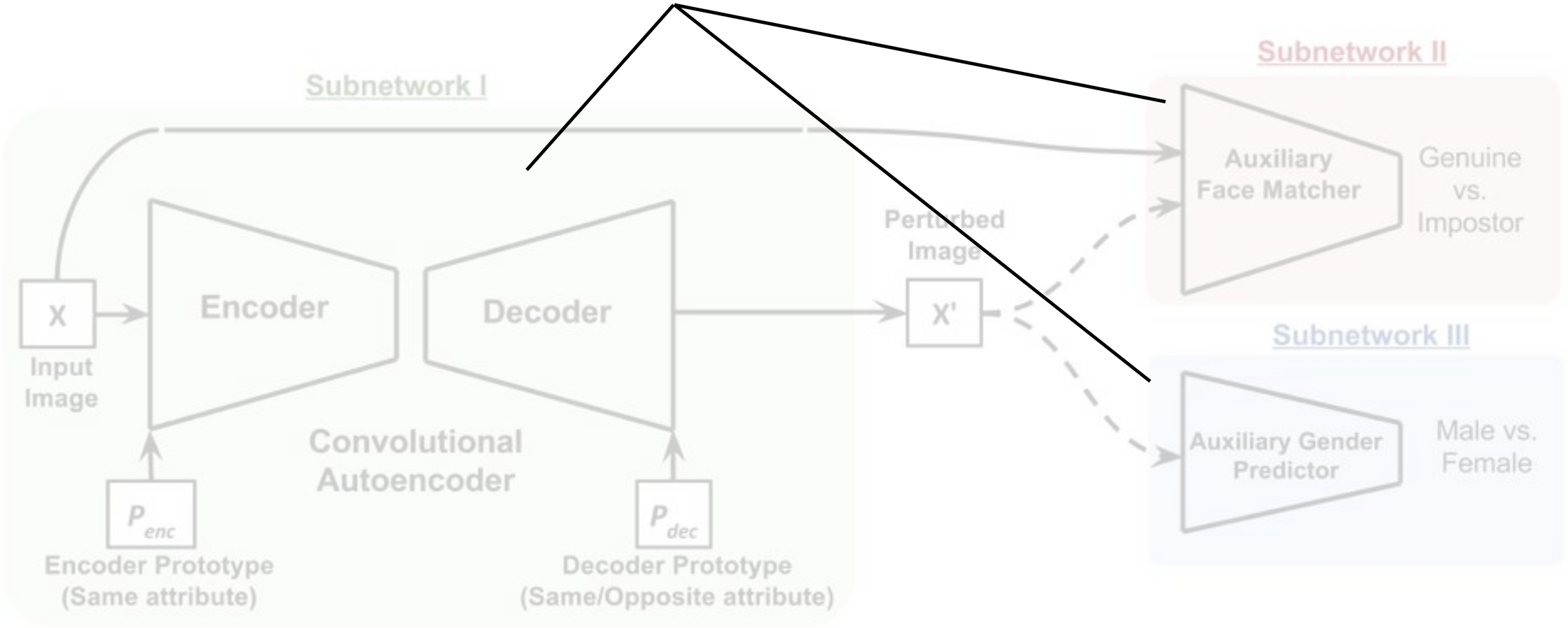
General architecture of the semi-adversarial network (SAN)



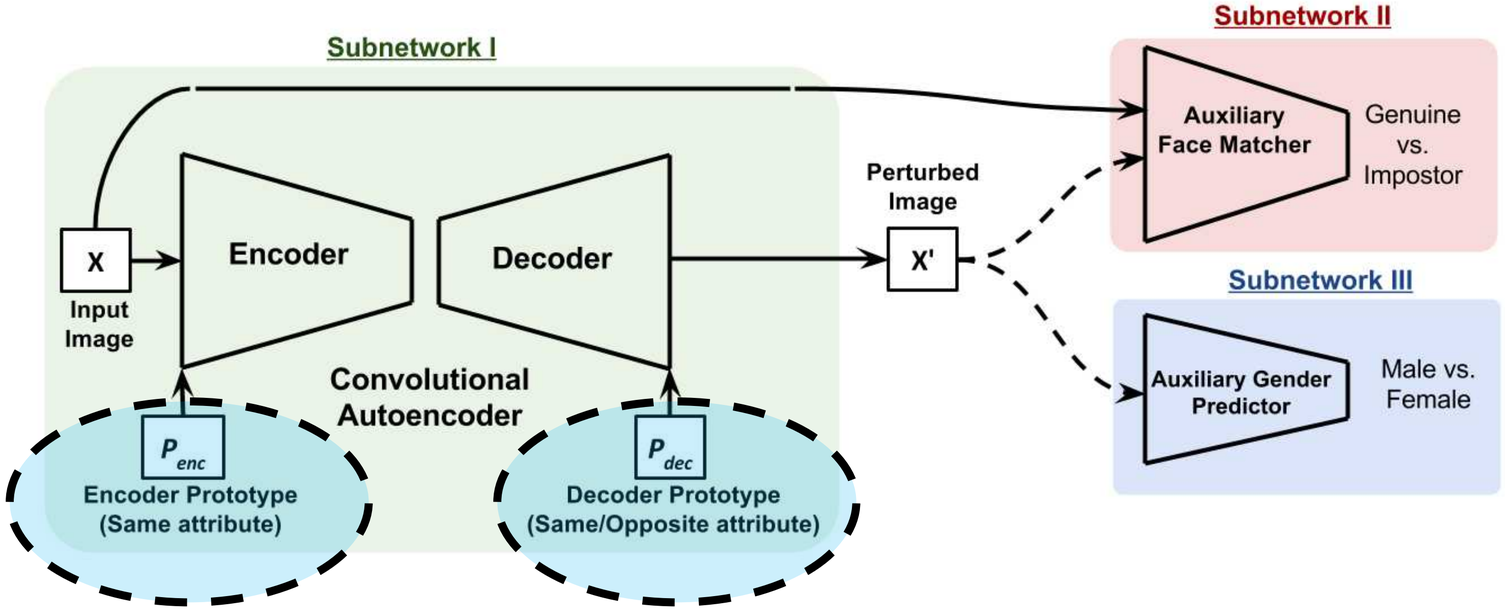
Semi-adversarial network



Convolutional neural networks



Gender prototypes



Class labels

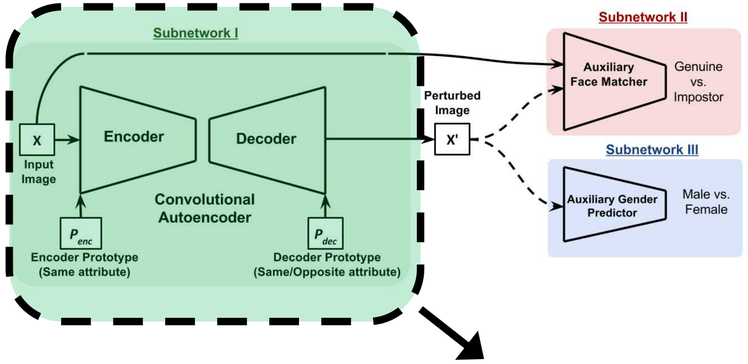
$y \in \{0, 1\}$,

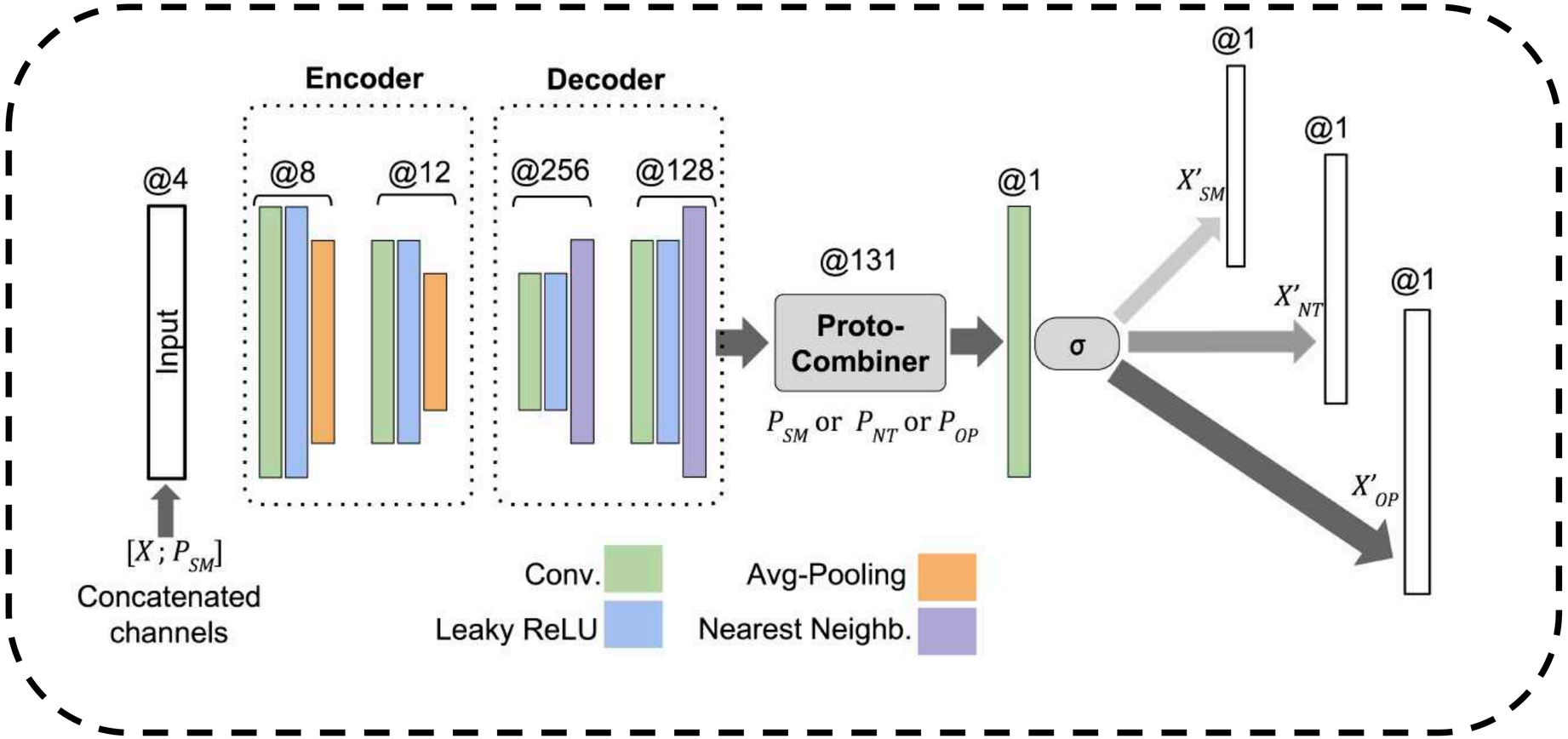
where $0 = \text{female}, 1 = \text{male}$

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

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

Convolutional autoencoder architecture





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

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

SAN Examples

Original Inputs

Male: 99%

Female: 98%

Male: 97%

Male: 100%

Outputs

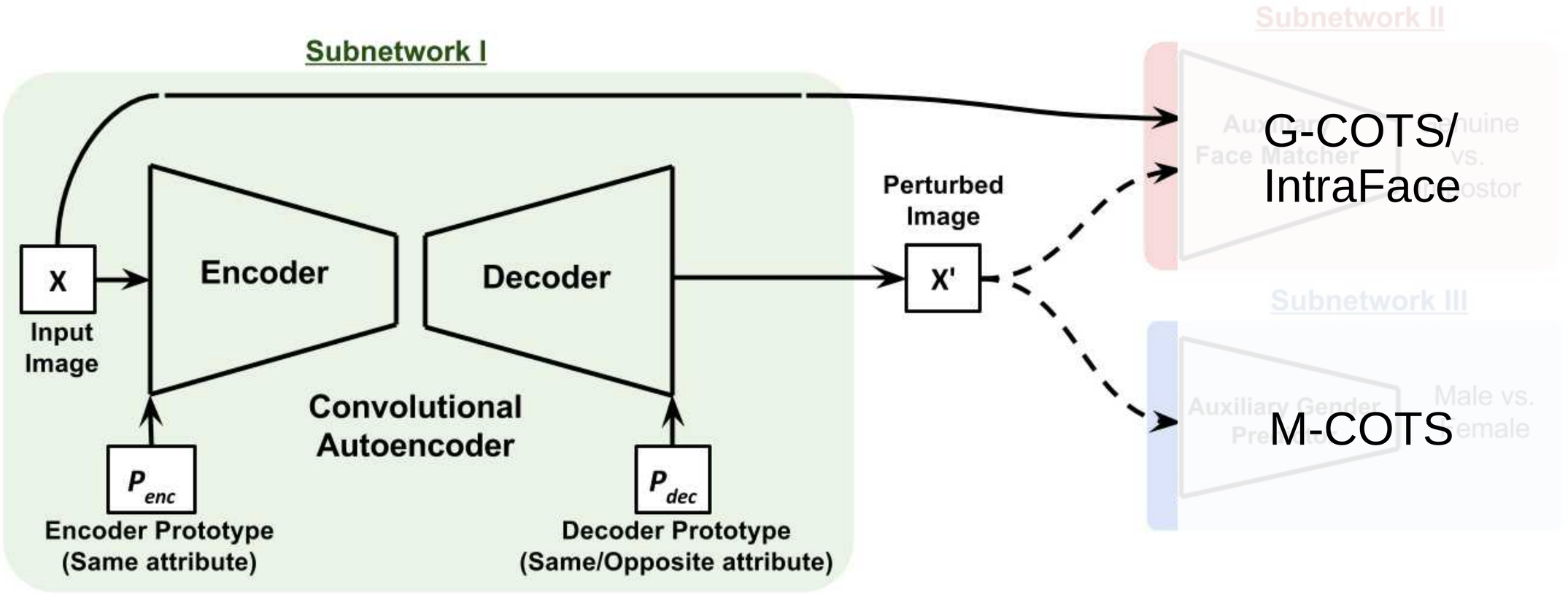
Female: 69%

Male: 99%

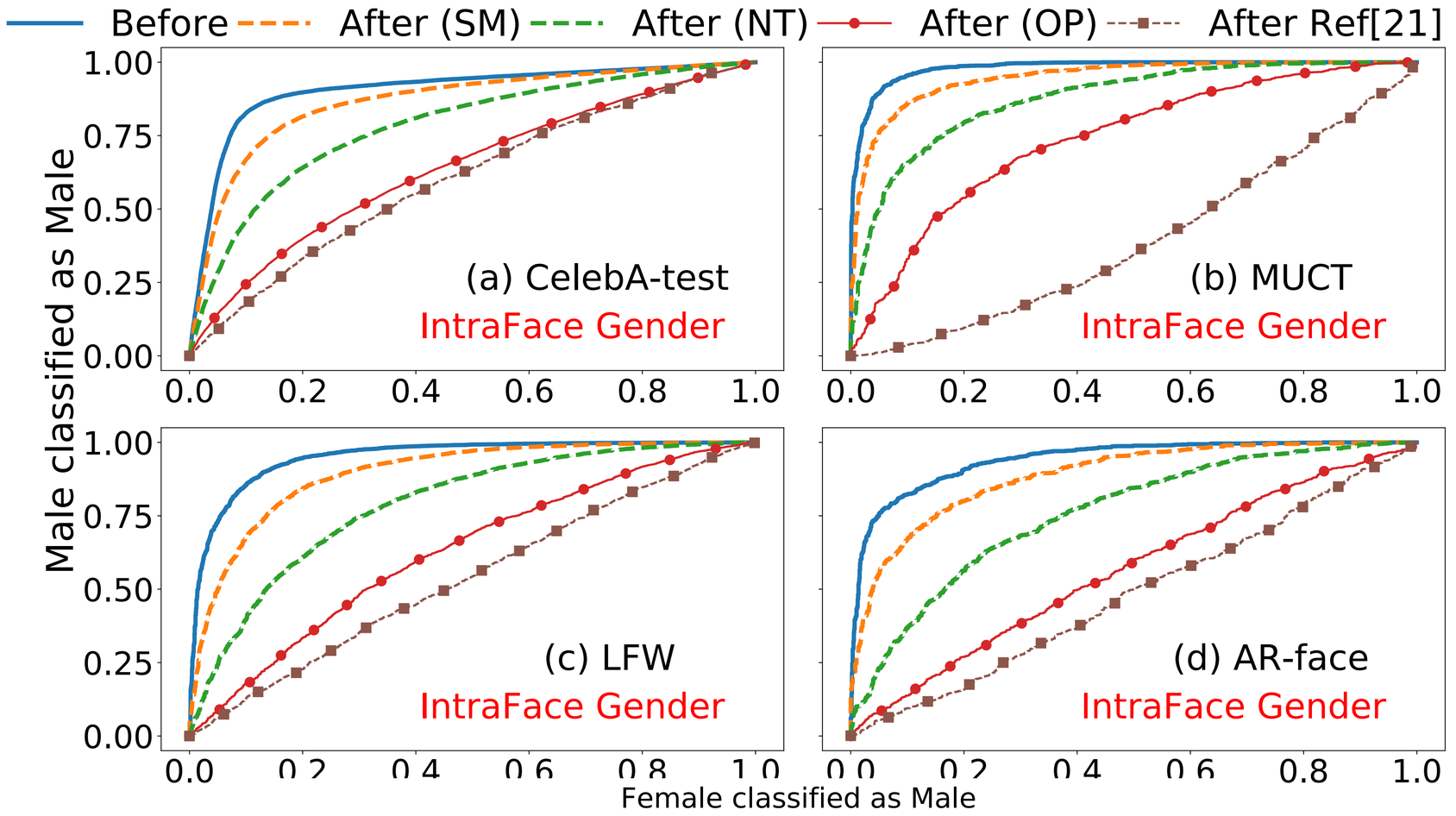
Female: 71%

Female: 58%

Replacing Detachable Parts for Evaluation



IntraFace Gender Classifier Performance

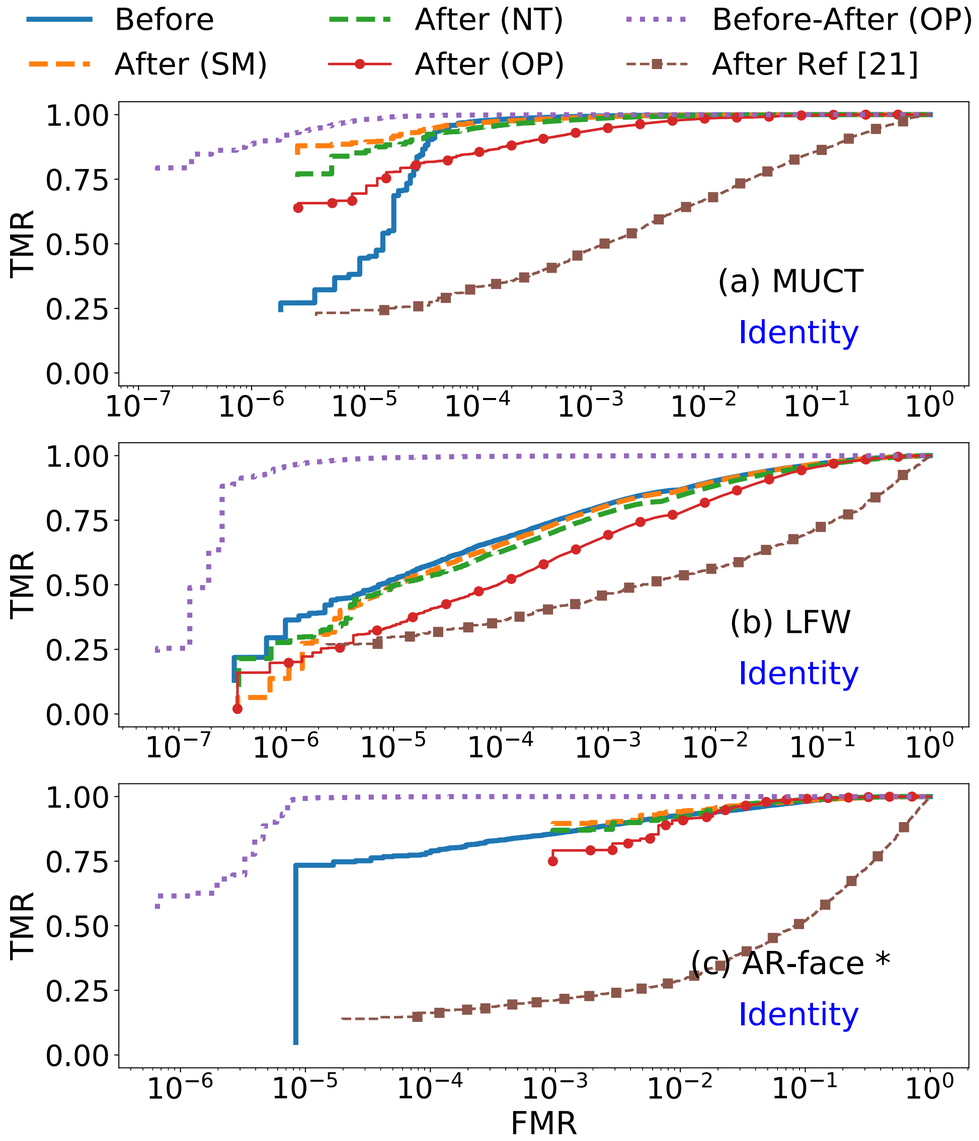


[21] A. Othman and A. Ross. Privacy of facial soft biometrics: Suppressing gender but retaining identity. In *European Conference on Computer Vision Workshop*, pages 682–696. Springer, 2014.

Face matching performance

Multi-subject comparisons

Before
After (SM)
After (NT)
After (OP)
After Ref [1]



[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 via ensembling

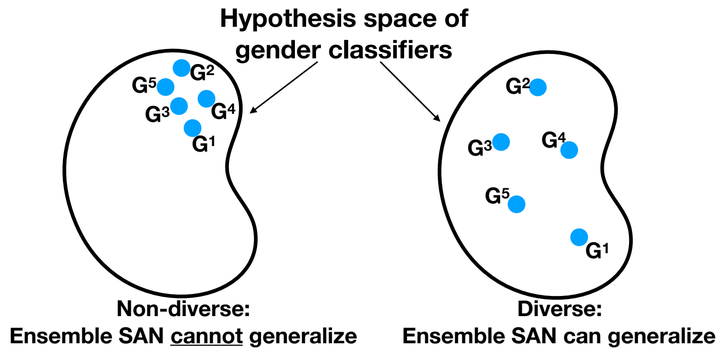


Figure 1: Diversity in an ensemble SAN can be enhanced through its auxiliary gender classifiers (see Figure 2). When the auxiliary gender classifiers lack diversity, ensemble SAN cannot generalize well to arbitrary gender classifiers.

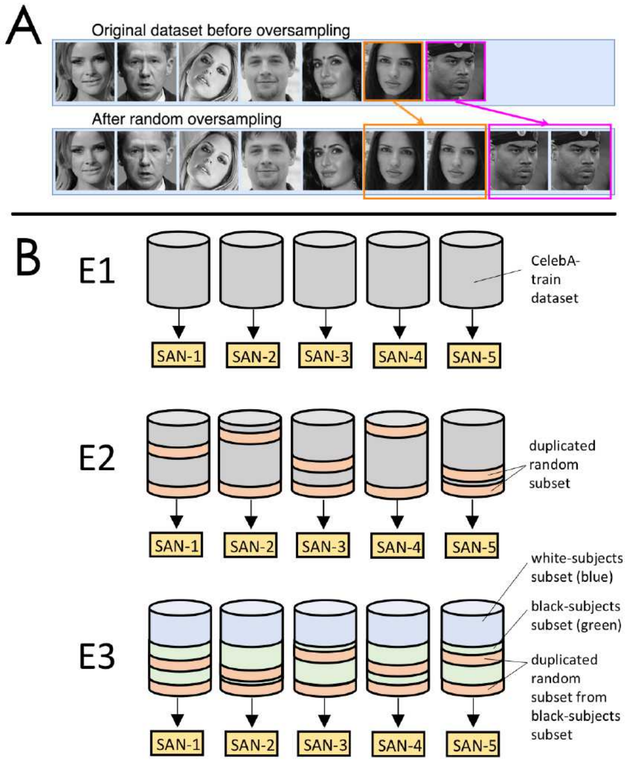


Figure 5: An example illustrating the oversampling technique used for enforcing diversity among SAN models in an ensemble. A: A random subset of samples are dupli-

Vahid Mirjalili, Sebastian Raschka, and Arun 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)

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 via ensembling

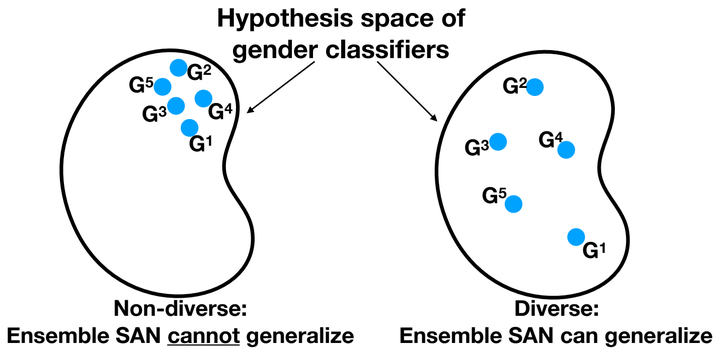


Figure 1: Diversity in an ensemble SAN can be enhanced through its auxiliary gender classifiers (see Figure 2). When the auxiliary gender classifiers lack diversity, ensemble SAN cannot generalize well to arbitrary gender classifiers.

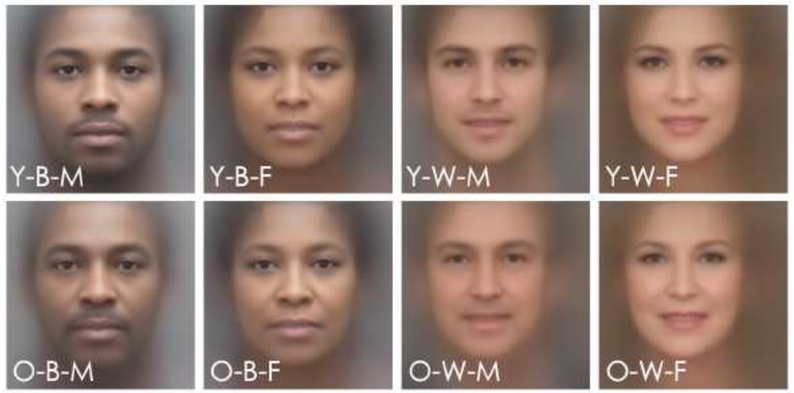
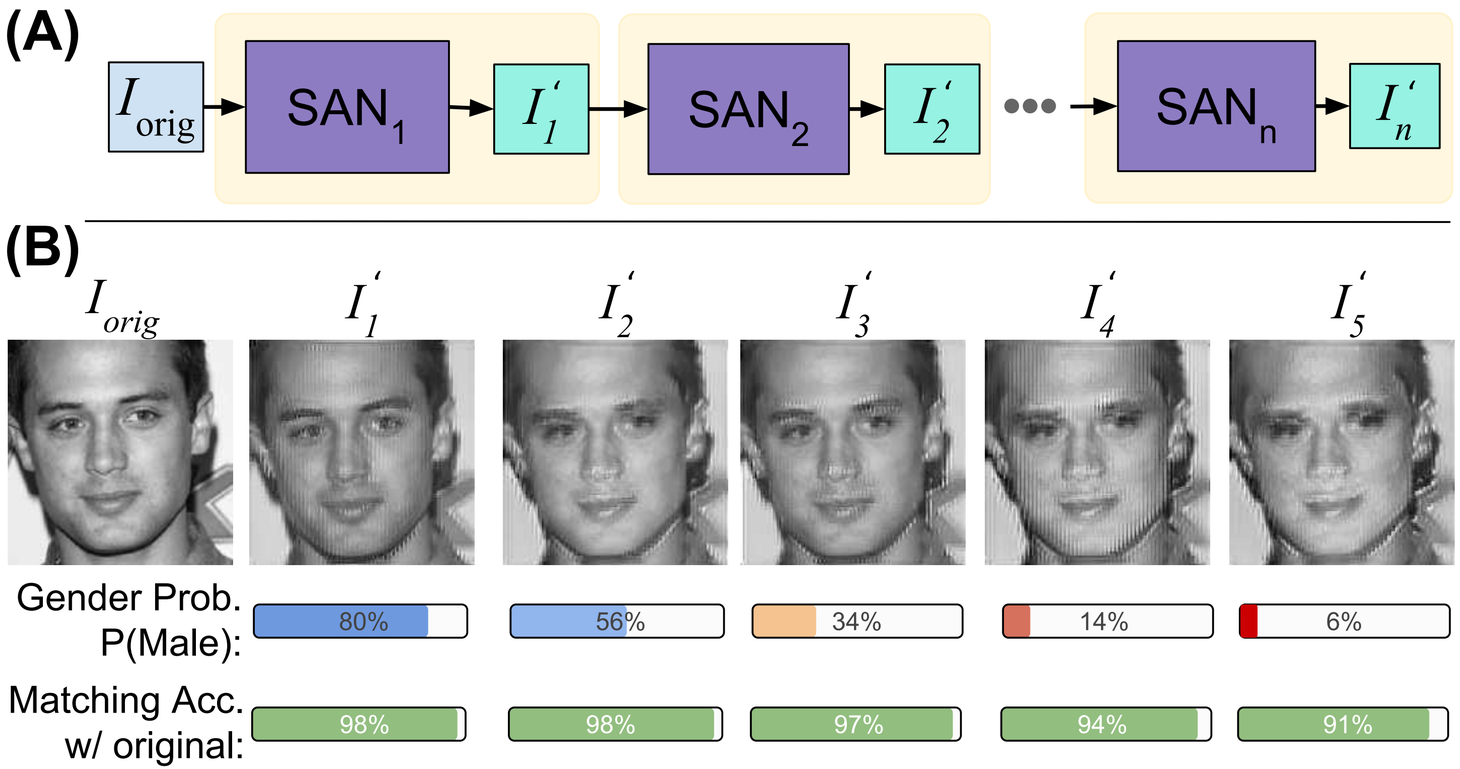


Figure 4: Face prototypes computed for each group of attribute labels. The abbreviations at the bottom of each image refer to the prototype attribute-classes, where Y=young, O=old, M=male, F=female, W=white, B=black.

Vahid Mirjalili, Sebastian Raschka, and Arun 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)

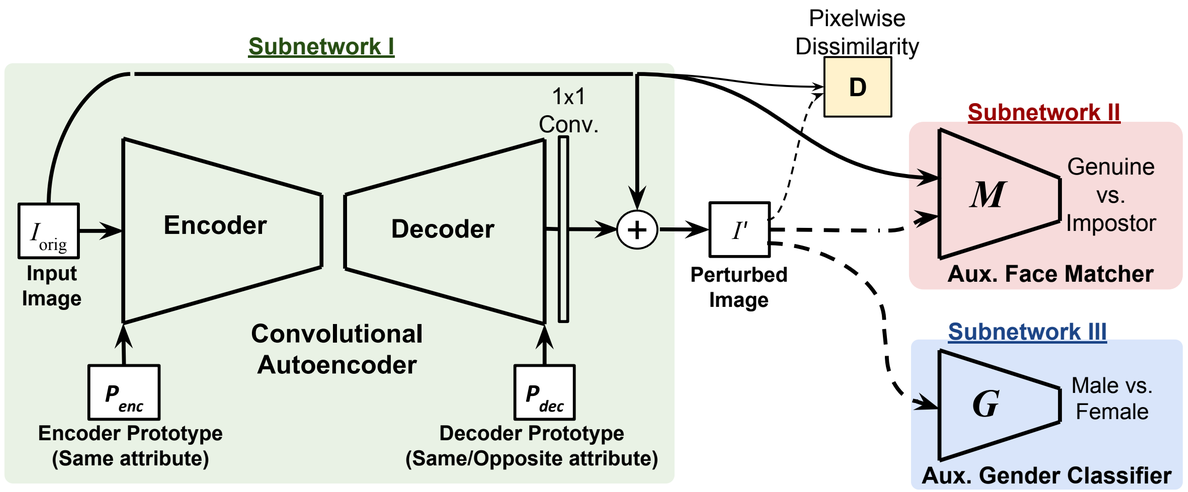
FlowSAN: Privacy-enhancing Semi-Adversarial Networks to Confound Arbitrary Face-based Gender Classifiers



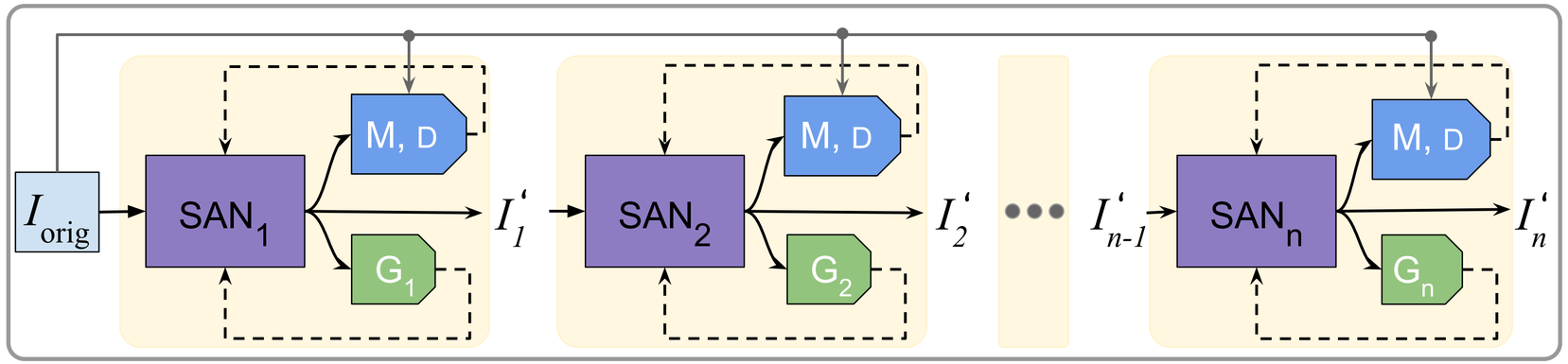
Improvements to better control the perturbations and enhance the removal of soft-biometric information

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

SAN base architecture

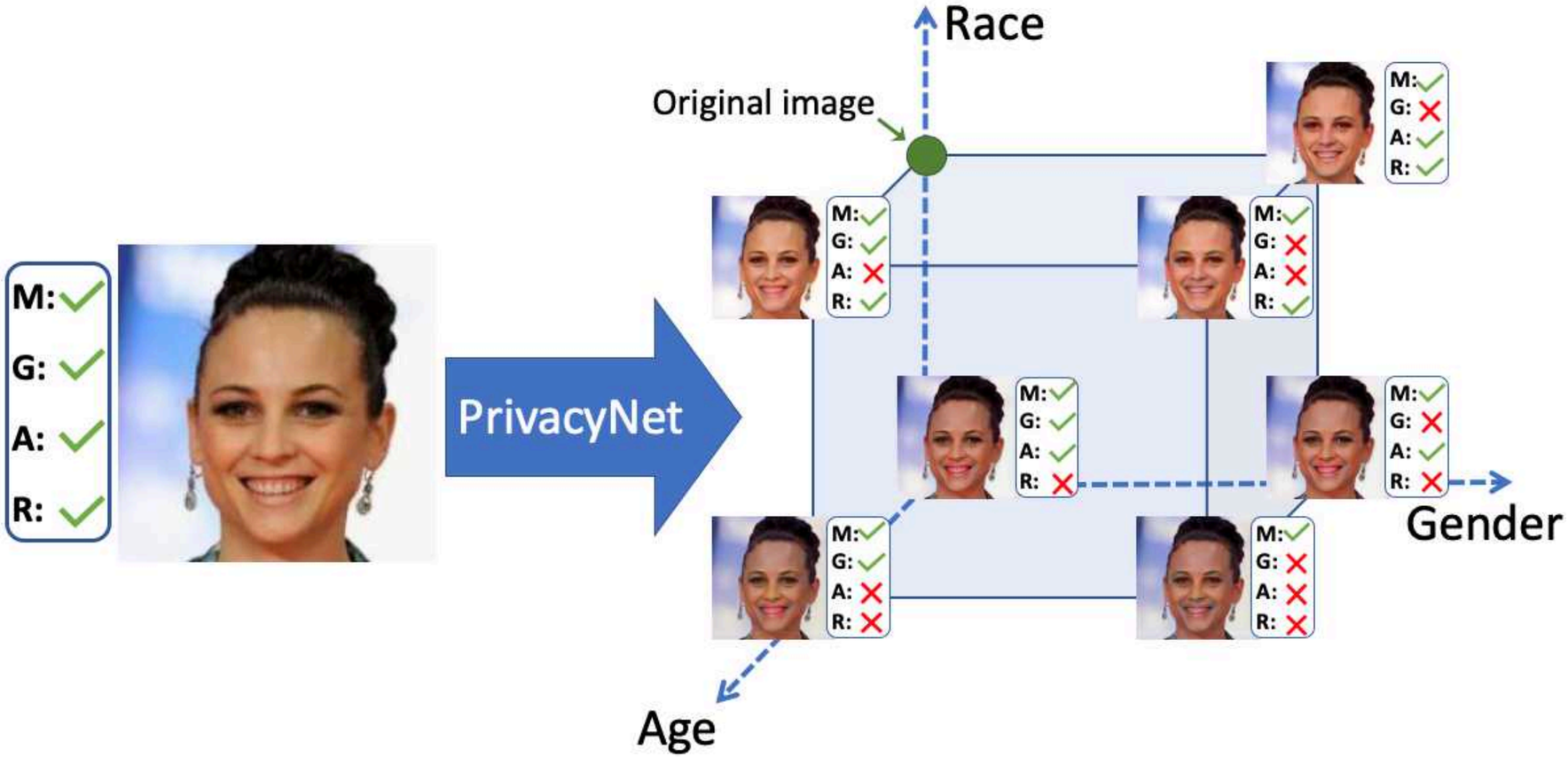


FlowSAN: Privacy-enhancing Semi-Adversarial Networks to Confound Arbitrary Face-based Gender Classifiers



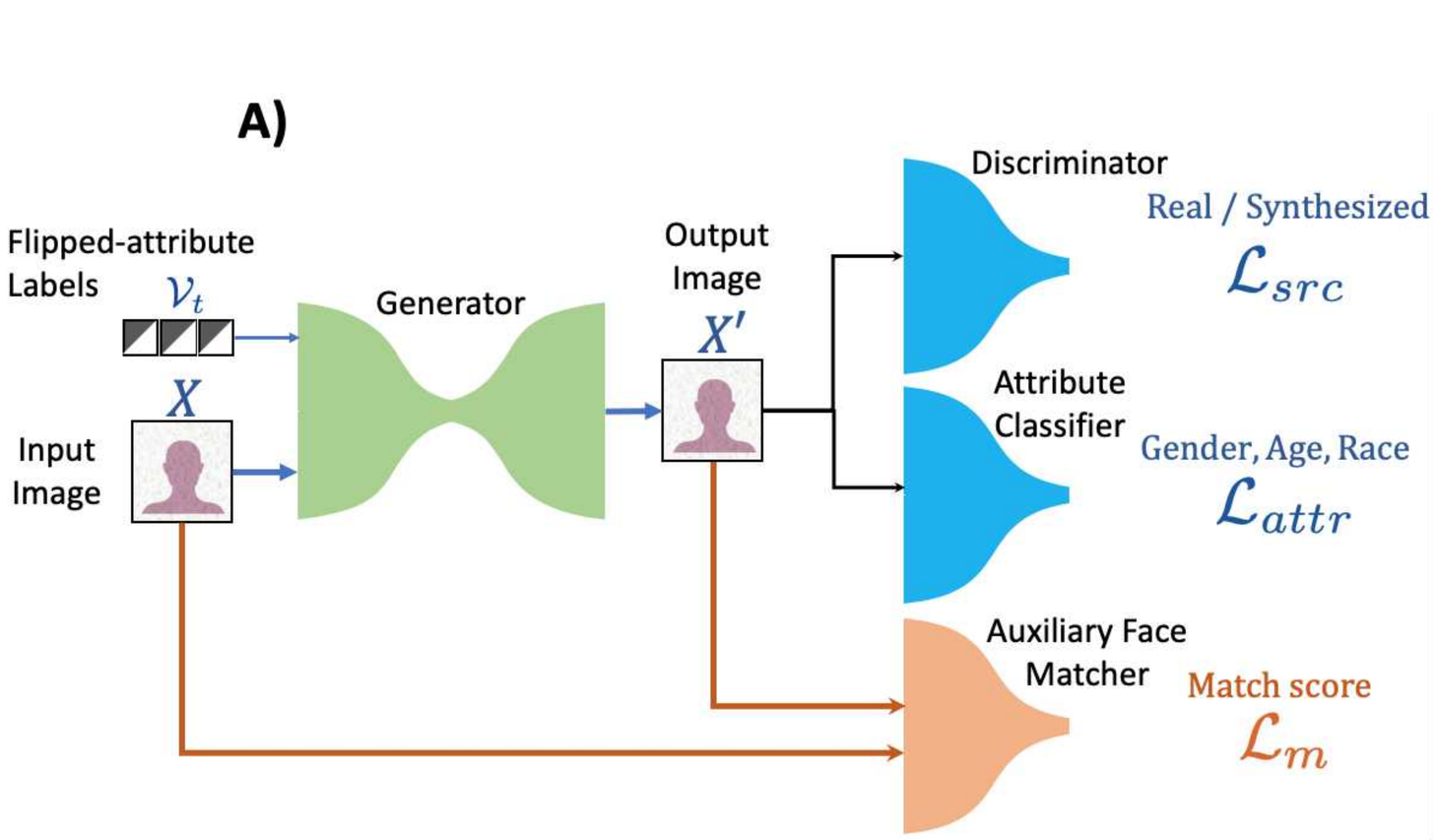
FlowSAN

PrivacyNet: Semi-Adversarial Networks for Multi-attribute Selective Privacy

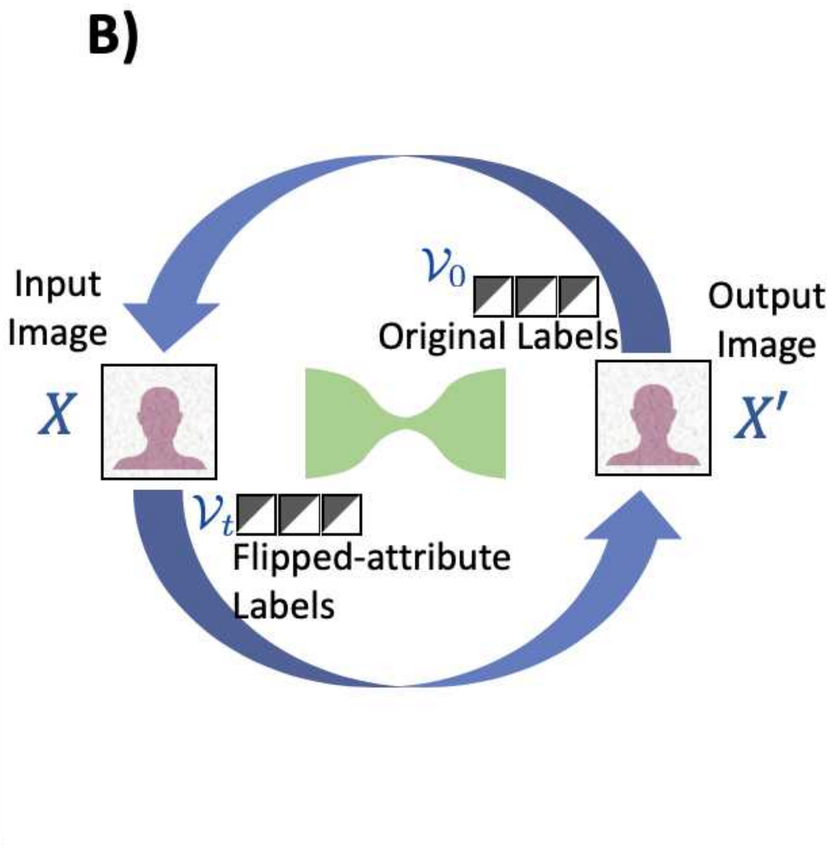


Vahid Mirjalili, Sebastian Raschka, and Arun Ross (2019) *PrivacyNet: Semi-Adversarial Networks for Multi-attribute Differential Privacy* (Submitted)

PrivacyNet: Semi-Adversarial Networks for Multi-attribute Selective Privacy



Architecture



Cycle-consistency constraint

Thank You!