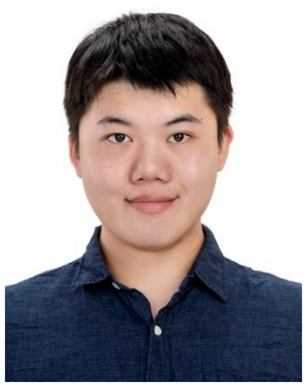
ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Contrastive Data Learning for Facial Pose and Illumination Normalization

Paper ID: 2284 Project Page: https://github.com/HaoRecog/pose-illumination-normalization

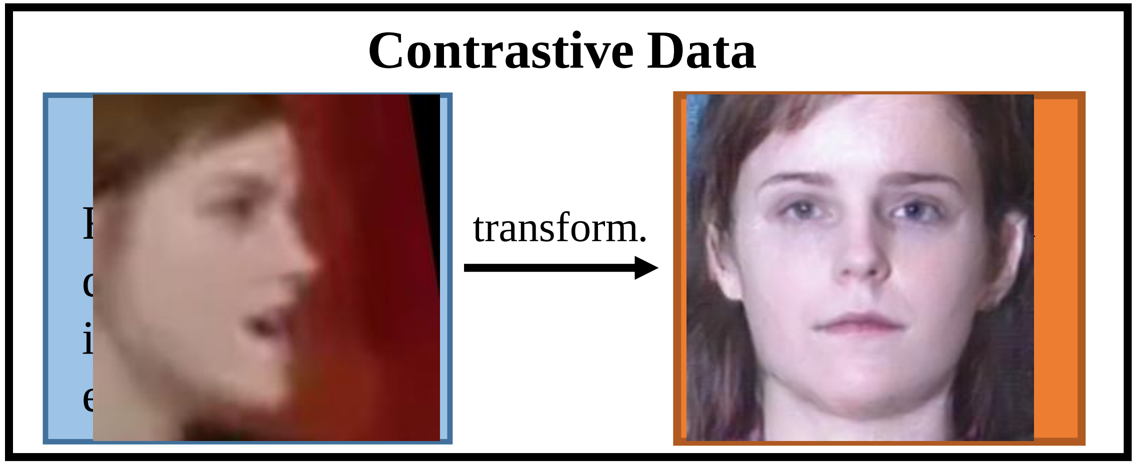



Gee-Sern Jison Hsu[1] Chia-Hao Tang[1] Svetlana Yanushkevich[2] Marina L Gavrilova[2]

[1]National Taiwan University of Science and Technology, [2]University of Calgary

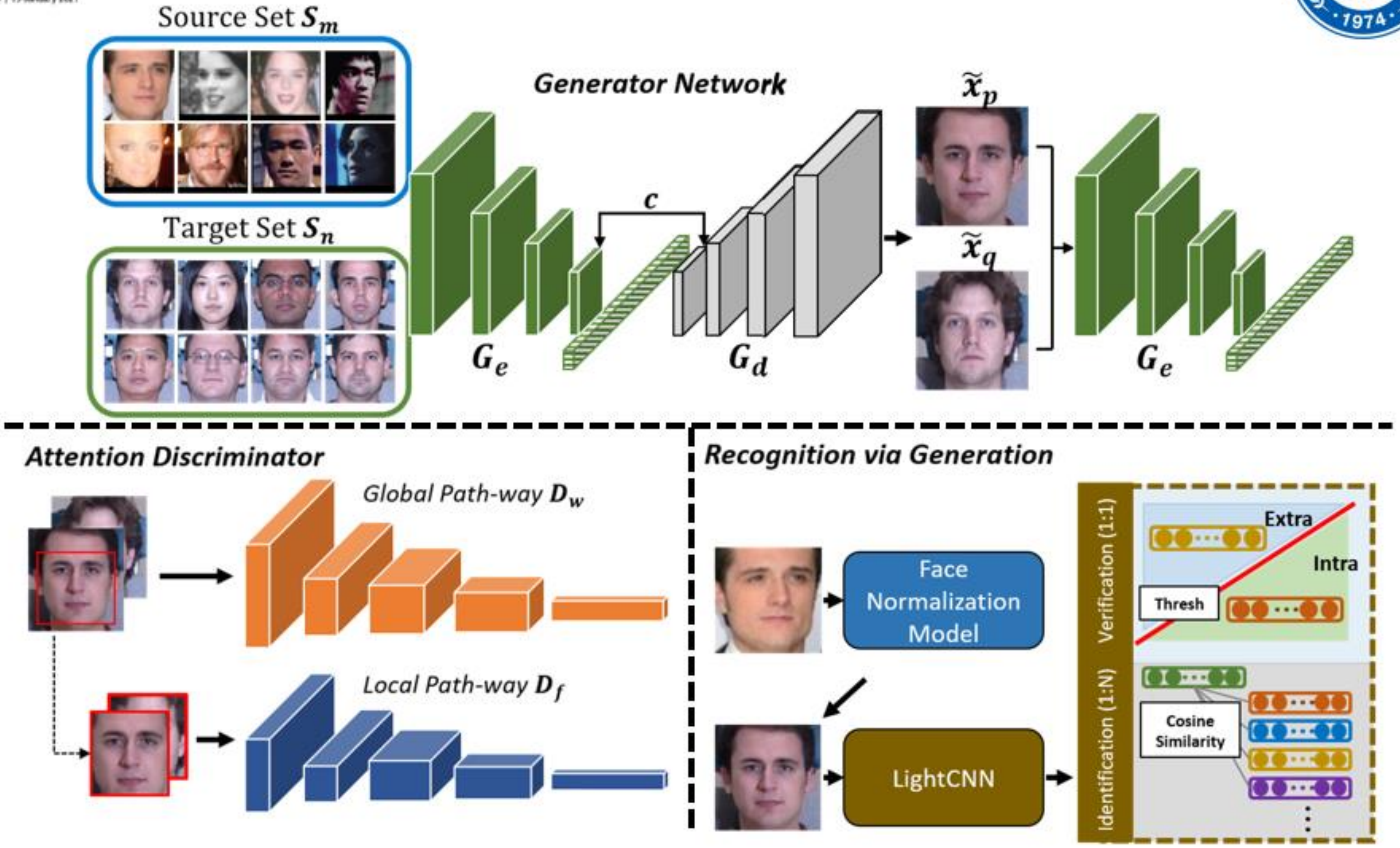
Problem statement

- **Motivation:** Most previous work [1, 2] primarily focused on frontal pose normalization only.
- **Task:** Learn the transformation from an arbitrary face into illumination and pose normalized face, in an unsupervised manner.
- **Key idea:** Follow the work in [2], we learn on the contrastive data (source and target), it can transform an arbitrary face into illumination and pose normalized face.



[1] Huang et al., "Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis." CVPR, (2017).
[2]Qian et al. "Unsupervised face normalization with extreme pose and expression in the wild.“ CVPR. (2019).

Pose and Illumination Normalization



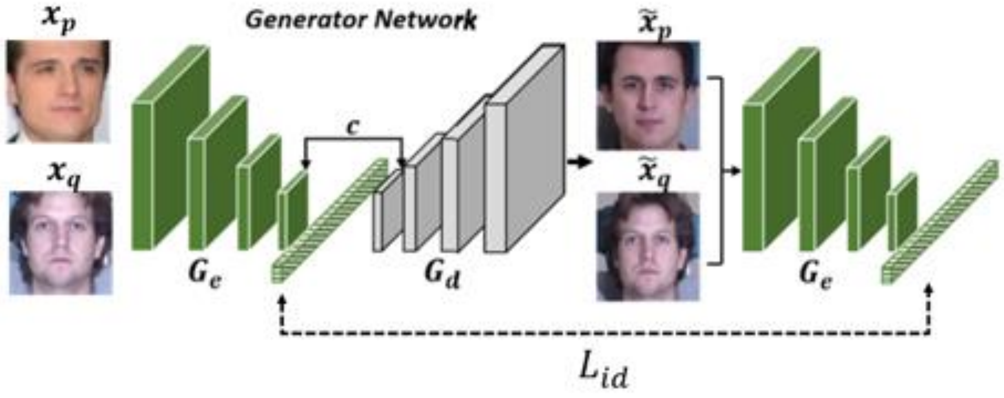
Objective Functions

Identity Loss	Adversarial Loss
Reconstruction Loss	Symmetry Loss

Objective Functions

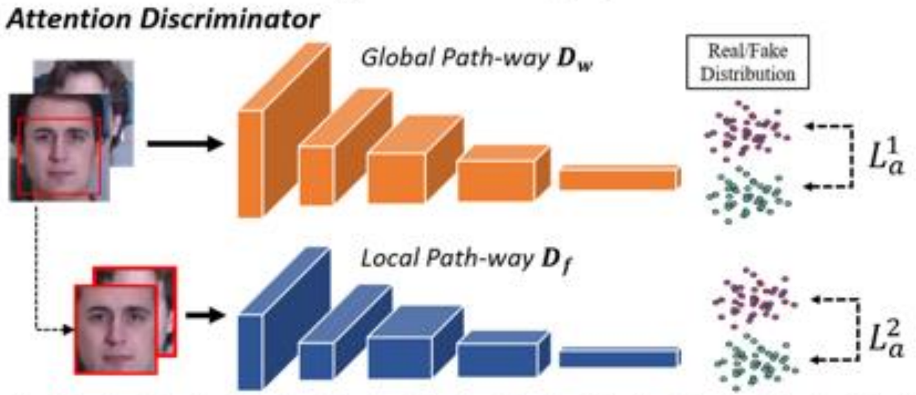
Identity Loss

$$L_{id} = \|G_e(x_p) - G_e(\tilde{x}_p)\|_2 + \|G_e(x_q) - G_e(\tilde{x}_q)\|_2$$



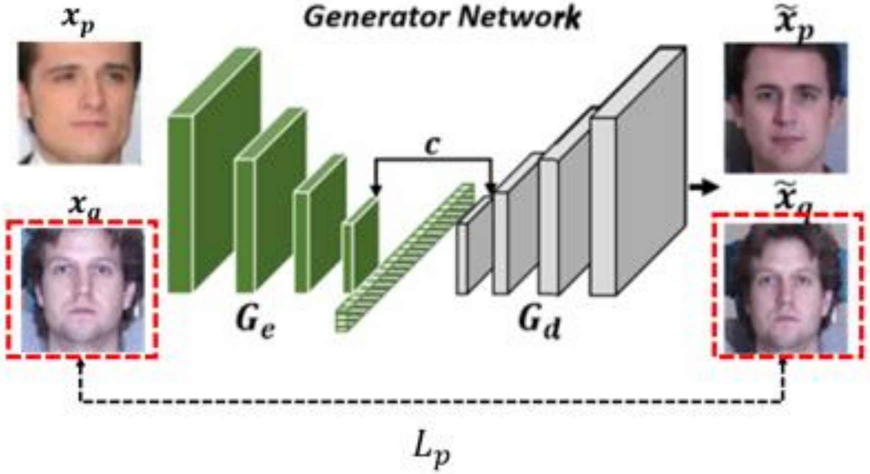
Adversarial Loss

$$L_a = \mathbb{E}[D_{\theta_i}(\tilde{x}_q)] + \mathbb{E}[D_{\theta_i}(\tilde{x}_p)] - \mathbb{E}[D_{\theta_i}(x_p)] + \lambda\mathbb{E}\left[\left(\|\nabla_{\tilde{x}} D_{\theta_i}(\tilde{x}) - 1\|_2\right)^2\right], i \in [1,2]$$



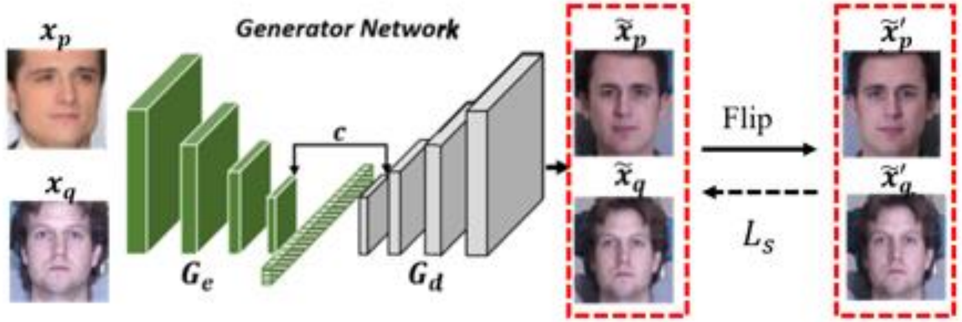
Reconstruction Loss

$$L_p = |\tilde{x}_p - x_p|$$



Symmetry Loss

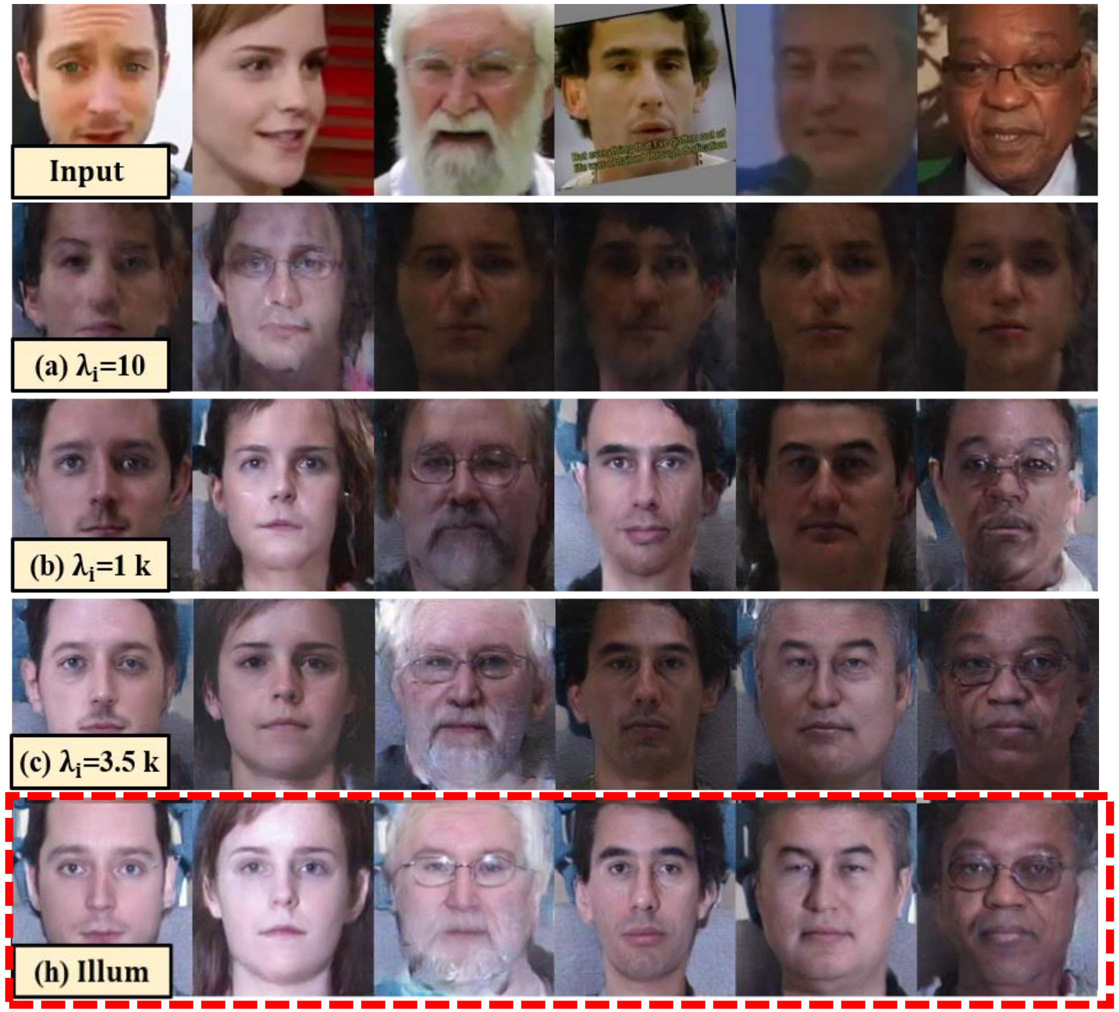
$$L_s = |\tilde{x}_p - \tilde{x}'_p| + |\tilde{x}_q - \tilde{x}'_q|$$



Experiments – Ablation study

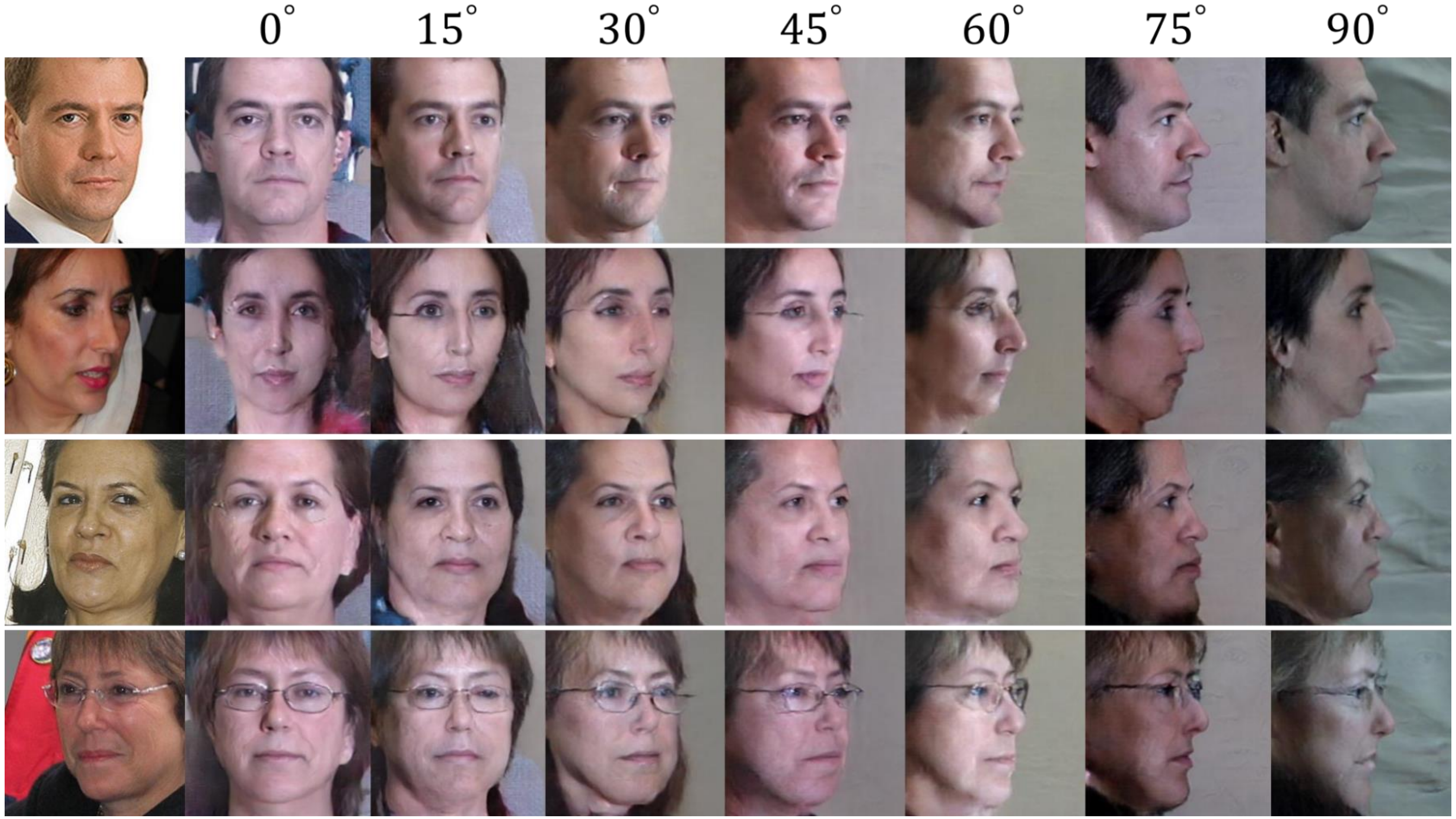
The ablation study on IJB-A.

Method	Verification @FAR=.01	Verification @FAR=.001	Identification @Rank-1	Identification @Rank-5
PIN ($\lambda_i = 10$)	7.5±1.5	2.8±1.2	12.8±0.4	28.9±2.5
PIN ($\lambda_i = 1000$)	87.8±2.1	72.8±0.6	91.7±2.1	95.9±1.1
PIN ($\lambda_i = 3500$)	92.6±0.9	85.5±1.4	95.1±1.4	96.9±0.5
PIN ($\lambda_i = 5500$)	90.9±0.4	80.7±1.0	93.2±1.7	96.8±0.9
PIN ($\lambda_a = 0$)	85.3±0.7	70.1±1.9	86.6±1.2	92.4±1.2
PIN ($\lambda_s = 0$)	86.2±1.6	78.9±1.1	88.1±1.0	93.9±1.7
PIN ($\lambda_p = 0$)	90.6±1.4	81.3±1.3	91.6±1.6	94.3±1.2
PIN (Illum)	93.9±0.6	87.0±0.4	96.7±0.9	98.3±0.7



Experiments – Ablation study

PIN (15°)	93.3±0.2	85.7±0.1	96.0±1.8	97.3±1.9
PIN (30°)	91.0±1.6	82.6±2.1	95.6±1.3	96.0±1.1
PIN$_m$	94.2±0.8	88.9±1.0	96.8±1.1	98.3±0.9
PIN$_c$	95.4±1.8	90.1±1.9	97.1±0.1	98.6±0.9
PIN (75°)	82.4±0.2	63.7±1.5	89.9±0.7	94.1±1.1
PIN (90°)	73.2±1.9	53.1±1.1	84.6±2.6	93.5±1.2

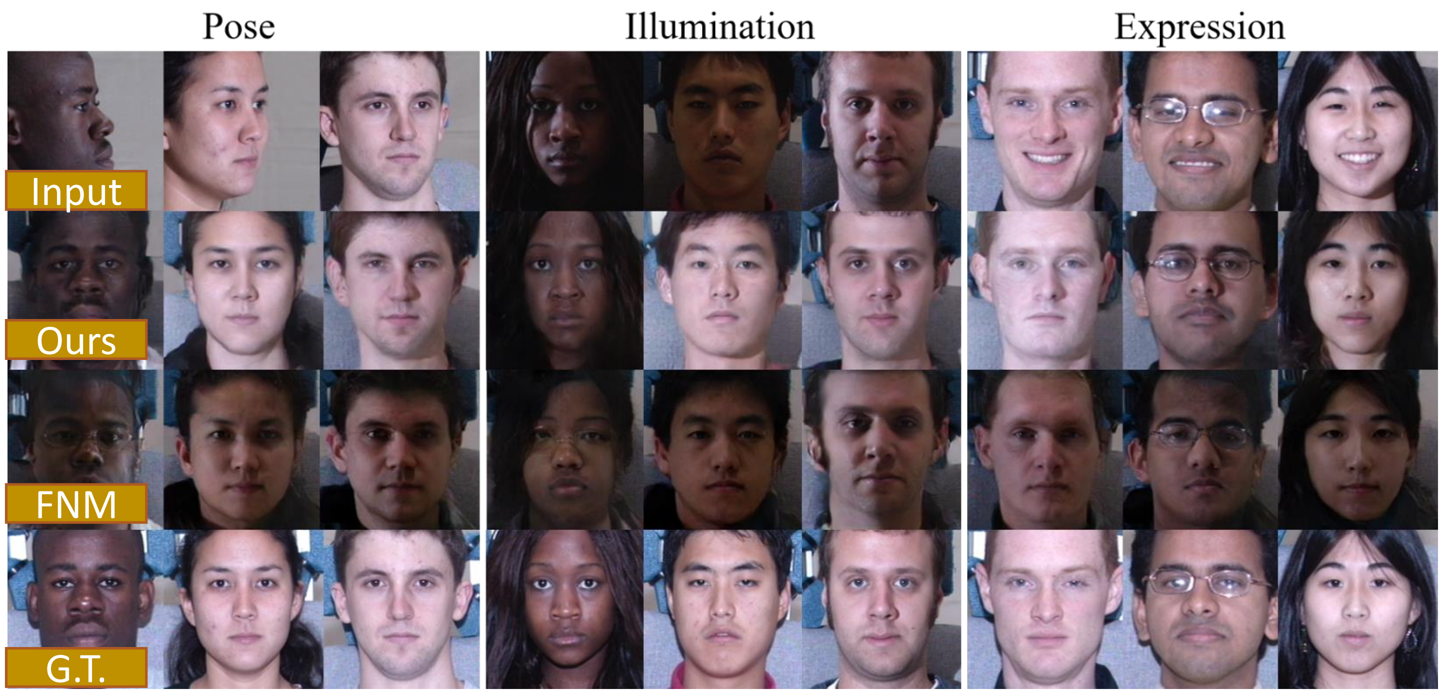


Experiments – Multi-PIE

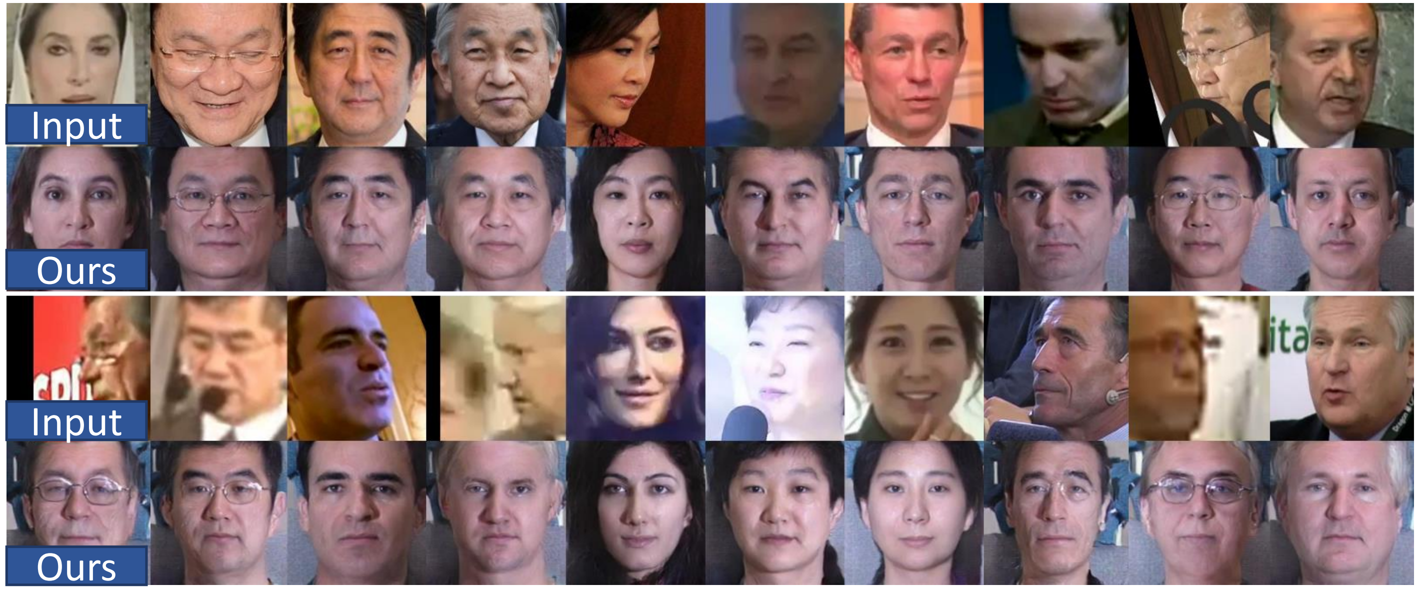
Rank-1 recognition rates (%) across views, illuminations on Multi-PIE.

Method	15°	30°	45°	60°	75°	90°
FF-GAN [14]	94.6	92.5	89.7	85.2	77.2	61.2
TP-GAN [1]	99.8	99.9	98.6	98.1	92.9	75.0
DR-GAN [5]	95.0	91.3	88.0	85.8	-	-
LightCNN [26]	99.2	98.0	97.7	95.5	73.3	20.7
CAPG-GAN [17]	99.9	99.4	98.3	93.7	87.4	77.1
PIM [2]	99.3	99.0	98.5	98.1	95.0	86.5
FNM [3]	99.9	99.5	98.2	93.7	81.3	55.8
LightCNN [26]	100	100	100	95.5	73.3	20.7
PIN + LightCNN	100	100	99.8	98.9	95.2	84.2
ArcFace [13]	100	100	100	96.5	83.1	40.5
PIN + ArcFace	100	100	100	99.2	96.9	86.5

PIN generated normalized faces compared with ground truth.



Experiments – IJB-A and IJB-C



Performance comparison on IJB-A.

	Verification		Identification	
Method	@FAR=.01	@FAR=.001	@Rank-1	@Rank-5
PAM [8]	73.3±1.8	55.2±3.2	77.1±1.6	88.7±0.9
DCNN [28]	78.7±4.3	-	85.2±1.8	93.7±1.0
FF-GAN [14]	85.2±1.0	66.3±3.3	90.2±0.6	95.4±0.5
FaceID-GAN [14]	87.6±1.1	69.2±2.7	-	-
DR-GAN [5]	87.2±1.4	78.1±3.5	92.0±1.3	96.1±0.7
FNM [3]	93.4±0.9	83.8±2.6	96.0±0.5	98.6±0.3
LightCNN [26]	91.2±1.1	84.4±0.8	92.4±1.7	95.4±0.8
PIN + LightCNN	95.4±1.8	90.1±1.9	97.1±0.1	98.6±0.9
ArcFace [13]	94.9±1.2	90.2±0.5	95.1±0.6	98.1±0.3
PIN + ArcFace	96.2±1.2	91.5±0.5	97.6±0.6	98.9±0.3

Performance comparison on IJB-C.

	Verification	
Method	@FAR=.01	@FAR=.001
FaceNet [7]	32.40	20.58
VGGFace [10]	45.60	26.18
DR-GAN [5]	88.2	73.6
VGGFace2 [11]	95.0	90.0
LightCNN[26]	90.63	84.32
PIN + LightCNN	91.49	86.56
ArcFace[13]	95.82	91.69
PIN + ArcFace	96.11	92.27

Conclusion

We improve the FNM with four components:

1) Re-organized the contrastive data set by strictly keep the target set with frontal face with balance illumination.
2) Add in the symmetry loss to stable both target and source face in optimization process
3) Determine the weights to emphasize the contributions of different losses
4) Incorporation of the ArcFace as our encoder which provide more discriminative prior knowledge to the decoder.

Experiments show that PIN framework is competitive to SOTA approaches.

Thanks for Watching