Contrastive Data Learning for Facial Pose and Illumination Normalization

Gee-Sern Jison Hsu[1] (jison@mail.ntust.edu.tw), Chia-Hao Tang[1], Svetlana Yanushkevich[2], Marina L Gavrilova[2]
[1]National Taiwan University of Science and Technology, Taiwan, [2]University of Calgary, Canada.

Introduction

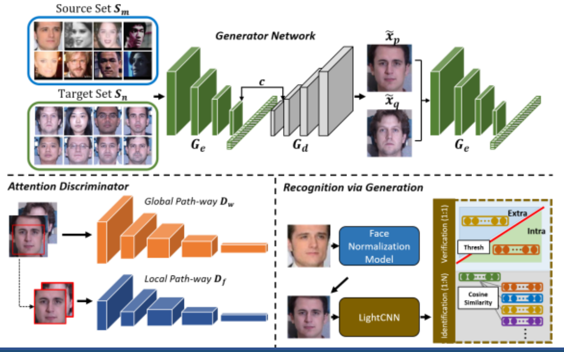
- We propose the Pose and Illumination (PIN) framework with contrast data learning for fac normalization.
- The PIN framework is designed to learn the transformation from a source set to a target set.
- The source set contains faces collected in the wild and thus covers a wide range of variation across illumination, pose, expression and other variables.
- The target set contains images taken under controlled conditions and all faces are in frontal pose and balanced in illumination.

Contributions

- Different from most previous work primarily focused on pose normalization, our proposed approach normalizes faces in both pose and illumination.
- The proposed approach offers a comprehensive study on the combination of various loss functions, extending the understanding of their importance on the generated images.
- The proposed approach is verified to be highly competitive to state-of-the-art methods for face recognition via normalization.

Proposed PIN Framework

- The encoder is made of the ArcFace recognition model and acts as a facial feature extractor.
- The decoder aims to transform an arbitrary face into a illumination and pose normalized face.
- The discriminators are trained to ensure the photo-realistic quality of the normalized face images generated by the decoder.

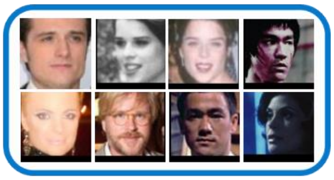


Objective Functions

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

Contrastive Data

Source set - CASIA-WebFace
Faces taken under unconstrained condition.

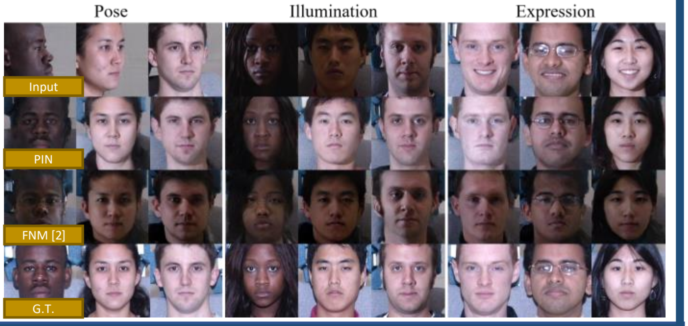


Target set - Multi-PIE
Faces taken under frontal and balance illumination.

Face Synthesis on IJB-C

Face Synthesis on Multi-PIE

Face normalization across pose, illumination and expression.



Performance Evaluation

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

Rank-1 Performance(%) 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

Conclusion

We the FNM with four components:

- Re-organized the contrastive data set by strictly keep the target set with front face with balanced illumination.
- Add in the symmetry loss to stabilize both target and source face optimization process.
- Determine the weights to emphasize the contributions of different losses.
- Incorporation of the SOTA ArcFace as our encoder which provide more discriminative prior knowledge to the decoder.

Experiments show that PIN framework is competitive to state-of-the-art approaches for handling both face recognition and face synthesis.