Unsupervised Face Normalization with Extreme Pose and Expression in the Wild

CVPR 2019

이 정 수

DAVIAN
Data and Visual Analytics Lab

Motivation

Face recognition models have limited invariance to strong intra-personal variations

-> solution: face normalization

- challenges:

1) complex face variations (large pose, expression, lighting, self-occlusion….)

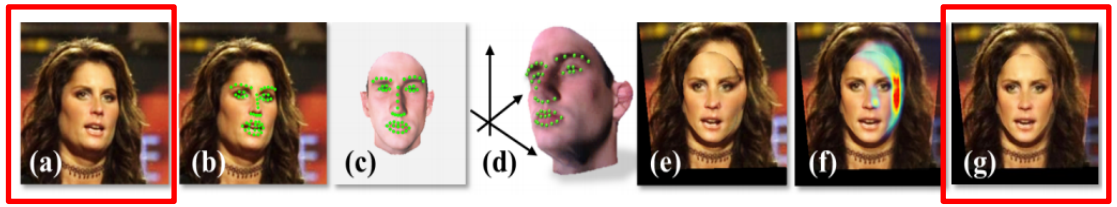
2) difficult to get supervision of target normalized face (front-facing, neutral expression)

: Face normalization output

Figure 1. Face normalization results under the same identity in unconstrained environment. Face images are under different views across pose, lighting, expression and background. FNM can keep a high-level consistency in preserving identity. On the right of the dashed line is a near-normal face of the same identity.

Previous work

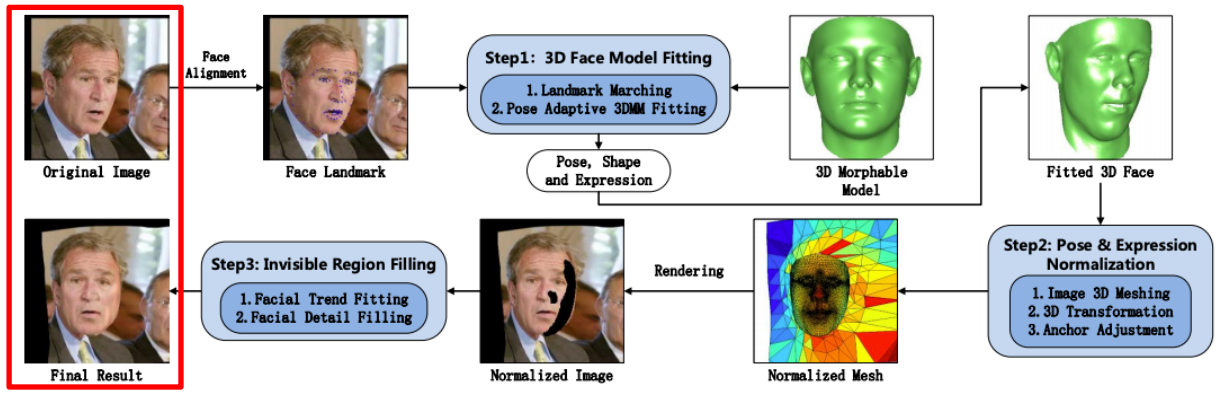
3D-based methods (old-outdated)



Original image

result

Hassner *et al* (CVPR 2015)



Zhu *et al* (CVPR 2015)

GAN

FF-GAN (ICCV 2017)

TP-GAN (ICCV 2017)

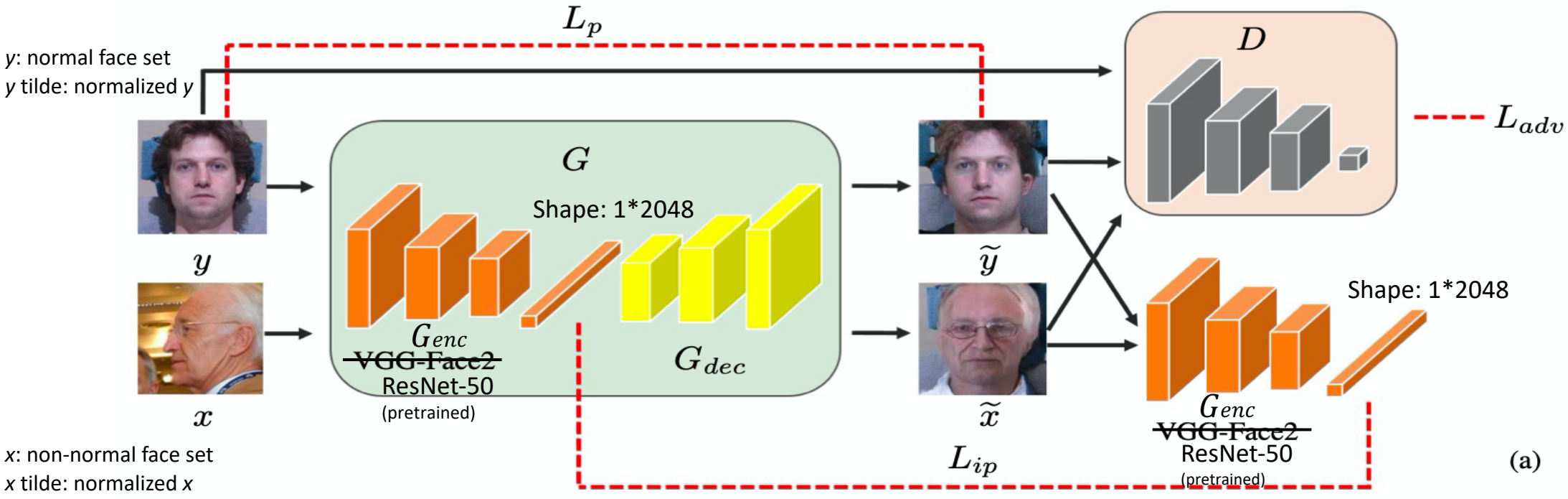
PIM (CVPR 2018)

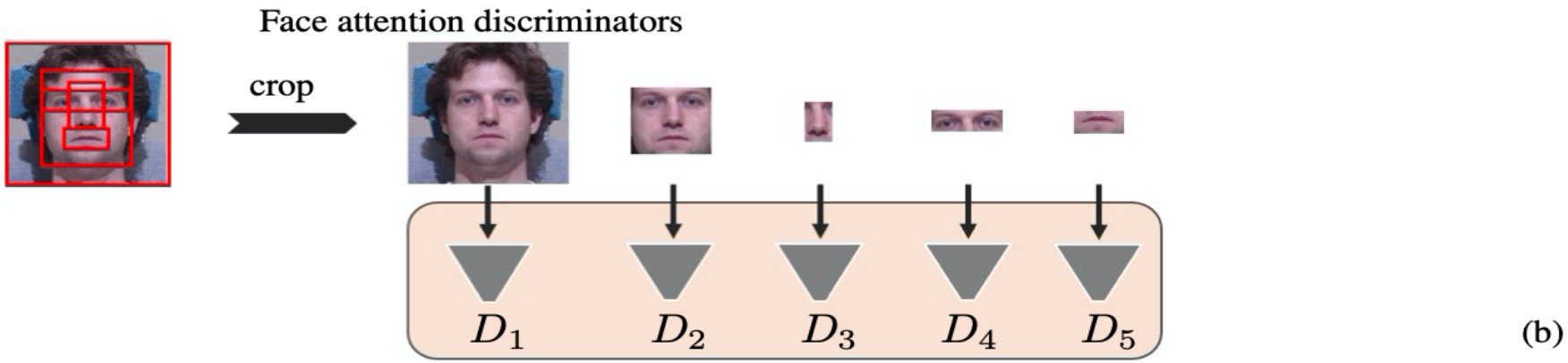
Method

Face Normalization Model (FNM)
- face synthesis model with unpaired data
- frontal/neutral/photorealistic expression
  -> invariant face normalization to pose/expression/lighting/occlusion

Contributions
- 1) incorporate *face expert network*: retain face identity -> pretrained ResNet50
- 2) ~~introduce pixel-wise loss (L1 loss) -> stabilize optimization process~~ (in face normalization work)
- 3) series of face attention discriminators to refine local textures
  - no 3D face model / no landmark localization
  - discriminators: fixed areas of generated normalized face
- 4) add-on module for face recognition model (pre-processing)

Method


L_p
y: normal face set
y tilde: normalized y
D
L_{adv}
G
Shape: 1*2048
y
$\widetilde{y}$
G_{enc}
~~VGG-Face2~~
ResNet-50
(pretrained)
G_{dec}
Shape: 1*2048
x
$\widetilde{x}$
G_{enc}
~~VGG-Face2~~
ResNet-50
(pretrained)
L_{ip}
x: non-normal face set
x tilde: normalized x
(a)


Face attention discriminators
crop
D_1
D_2
D_3
D_4
D_5
(b)

Method

Loss

$$L_{adv} = \sum_{k=1}^{5} D_k(\widetilde{x}_k) + \sum_{k=1}^{5} D_k(\widetilde{y}_k) - \sum_{k=1}^{5} D_k(y_k)$$

L_{ip} : identity perception loss

$$L_{ip} = ||G_{enc}(x) - G_{enc}(\widetilde{x})||_2^2 + ||G_{enc}(y) - G_{enc}(\widetilde{y})||_2^2$$

L_p : pixel-wise consistency
(L1 loss/reconstruction loss)

$$L_p = \frac{1}{W \times H \times C} \sum_{w,h,c}^{W,H,C} |y_{w,h,c} - \widetilde{y}_{w,h,c}|$$

$$\begin{cases} L_D = L_{adv}, \\ L_{G_{dec}} = -L_{adv} + \lambda_1 L_{ip} + \lambda_2 L_p. \end{cases}$$

Dataset

Normal Dataset
- MultiPIE Dataset

Non-normal Dataset
- CASIA-WebFace Dataset

Dataset description: http://www.cs.cmu.edu/afs/cs/project/PIE/MultiPie/Multi-Pie/Content.html

- 337 subjects / more than 750,000 images
- 308GB / need to pay / shipped with USB

Dataset url: https://search.wellspringsoftware.net/

Results

Dataset: IJB-A

Figure 3. Face normalization results on IJB-A [18] under extreme pose, express, lighting, occlusion and front view.

Results

Dataset: Multi-PIE

Figure 4. Synthesis results on Multi-PIE. Each pair presents profile (left), normalized face (middile) and ground truth normal face (right).

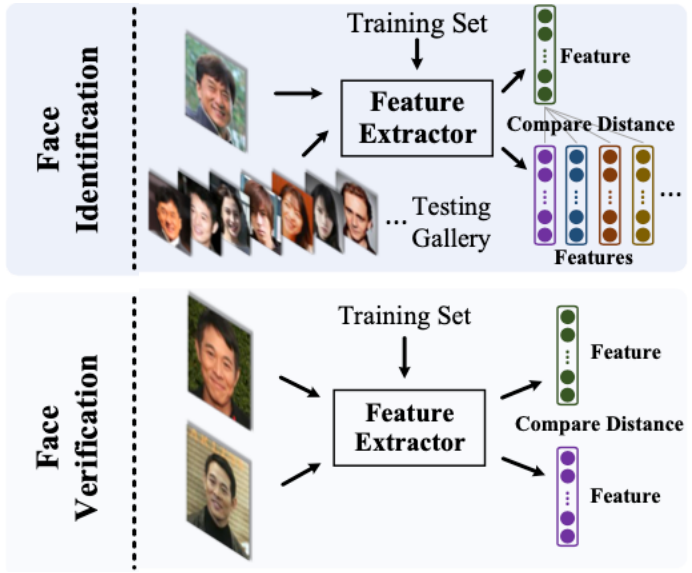
Results

Dataset: LFW

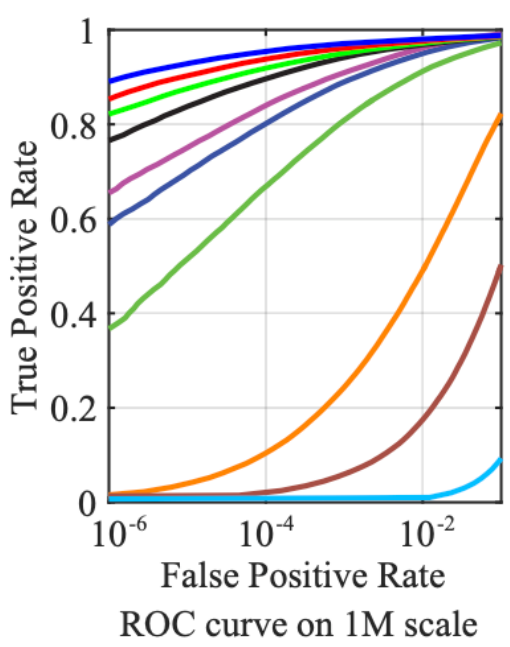
LFW PIM TP-GAN DR-GAN Hassner *et al.* FNM

Figure 6. Comparison of face frontalization on LFW[15].

*Experiment Results



face identification vs face verification



$$False\ Positive\ Rate = \frac{\sum False\ Positive\ 수}{\sum 정답의\ Negative\ 수}$$

= False Accept Rate (FAR)

Total Population	모델의 예측 Positive	모델의 예측 Negative
정답이 Positive	True Positive	False Negative
정답이 Negative	False Positive	True Negative

$$True\ Positive\ Rate = \frac{\sum True\ Positive\ 수}{\sum 정답의\ Positive\ 수}$$

= True Accept Rate (TAR)

Total Population	모델의 예측 Positive	모델의 예측 Negative
정답이 Positive	True Positive	False Negative
정답이 Negative	False Positive	True Negative

ROC Curve

Experiment Results

superiority of FNM on "recognition via generation"

- Dataset: IJB-A, Multi-PIE

Table 1. Performance comparison on IJB-A. The results are averaged over 10 testing splits. Symbol "-" implies that the result is not reported for that method. FNM is incorporated into two face recognition framework VGG-Face [24] and Light CNN [26] as a pre-processing procedure.

Method	Verification		Identification	
	@FAR=0.01	@FAR=0.001	@Rank-1	@Rank-5
OpenBR [18]	23.6±0.9	10.4±1.4	24.6±1.1	37.5±0.8
GOTS [18]	40.6±1.4	19.8±0.8	43.3±2.1	59.5±2.0
PAM [22]	73.3±1.8	55.2±3.2	77.1±1.6	88.7±0.9
DCNN [5]	78.7±4.3	-	85.2±1.8	93.7±1.0
DR-GAN [20]	77.4±2.7	53.9±4.3	85.5±1.5	94.7±1.1
FF-GAN [31]	85.2±1.0	66.3±3.3	90.2±0.6	95.4±0.5
VGG-Face [24]	86.8±1.8	68.4±3.3	92.8±0.8	97.9±0.6
FNM+VGG-Face	88.8±1.9	69.0±4.6	94.6±0.5	98.4±0.5
Light CNN [26]	82.7±2.0	67.4±2.2	84.5±1.7	92.6±0.9
FNM+Light CNN	**93.4**±0.9	**83.8**±2.6	**96.0**±0.5	**98.6**±0.3

Table 2. Rank-1 recognition rates (%) across poses and illuminations under Multi-PIE Setting-1. FNM is incorporated into two face recognition framework VGG-Face [24] and Light CNN [26] as a pre-processing procedure.

Method	±90°	±75°	±60°	±45°	±30°	±15°
HPN [8]	29.82	47.57	61.24	72.77	78.26	84.23
c-CNN [27]	47.26	60.7	74.4	89.0	94.1	97.0
TP-GAN [16]	64.0	84.1	92.9	98.6	99.9	99.8
PIM [33]	75.0	91.2	97.7	98.3	99.4	99.8
CAPG-GAN [29]	77.1	87.4	93.7	98.3	99.4	99.9
VGG-Face [24]	2.1	5.8	38.0	73.5	85.8	94.9
FNM+VGG-Face	41.1	67.3	83.6	93.6	97.2	99.0
Light CNN [26]	2.6	10.5	32.7	71.2	95.1	99.8
FNM+Light CNN	55.8	81.3	93.7	98.2	99.5	99.9

Ablation Study

Figure 5. The results produced by two variations of FNM. (a) Input face. (b) FNM without L_p. (c) FNM without face attention mechanism. (d) our FNM.

Table 3. Component analysis: rank-1 recognition rates (%) across poses and illuminations under Multi-PIE Setting-1. Light CNN [26] is choose as baseline.

Method	±90°	±75°	±60°	±45°	±30°	±15°
Light CNN [26]	2.6	10.5	32.7	71.2	95.1	99.8
w/o L_p	46.9	62.0	70.4	78.5	81.2	90.3
w/o attention	41.3	66.6	83.4	92.3	96.0	97.6
FNM+Light CNN	55.8	81.3	93.7	98.2	99.5	99.9

Further Discussion

Discussion

- accuracy on MegaFace Dataset

RANK	MODEL	ACCURACY	PAPER	CODE	RESULT	YEAR
1	ArcFace + MS1MV2 + R100 + R	98.35%	ArcFace: Additive Angular Margin Loss for Deep Face Recognition			2018
2	SV-AM-Softmax	97.2%	Support Vector Guided Softmax Loss for Face Recognition			2018
3	CosFace	82.72%	CosFace: Large Margin Cosine Loss for Deep Face Recognition			2018
4	SphereFace (3-patch ensemble)	75.766%	SphereFace: Deep Hypersphere Embedding for Face Recognition			2017
5	Light CNN-29	73.749%	A Light CNN for Deep Face Representation with Noisy Labels			2015
6	SphereFace (single model)	72.729%	SphereFace: Deep Hypersphere Embedding for Face Recognition			2017
7	FaceNet	70.49%	FaceNet: A Unified Embedding for Face Recognition and Clustering			2015

CVPR 2019

Previous

Method	IJB-B	IJB-C
VGG2, R50, ArcFace	0.898	0.921
MS1MV2, R100, ArcFace	**0.942**	**0.956**

Table 7. 1:1 verification TAR (@FAR=1e-4) on the IJB-B and IJB-C dataset.