

Towards Pose Invariant Face Recognition in the Wild

Jian Zhao^{1,2}, Yu Cheng³, Yan Xu⁴, Lin Xiong⁴, Jianshu Li¹, Fang Zhao¹, Karlekar Jayashree⁴, Sugiri Pranata⁴, Shengmei Shen⁴, Junliang Xing⁵, Shuicheng Yan^{1,6}, and Jiashi Feng¹

¹National University of Singapore, ²National University of Defense Technology, ³Nanyang Technological University

⁴Panasonic R&D Center Singapore, ⁵National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, ⁶Qihoo 360 AI Institute

{zhaojian90, jianshu}@u.nus.edu, chengyu996@gmail.com, {yan.xu, lin.xiong, karlekar.jayashree, sugiri.pranata, shengmei.shen}@sg.panasonic.com

jlxing@nlpr.ia.ac.cn, {eleyans, elefjia}@nus.edu.sg

Abstract

Pose variation is one key challenge in face recognition. As opposed to current techniques for pose invariant face recognition, which either directly extract pose invariant features for recognition, or first normalize profile face images to frontal pose before feature extraction, we argue that it is more desirable to perform both tasks jointly to allow them to benefit from each other. To this end, we propose a Pose Invariant Model (PIM) for face recognition in the wild, with three distinct novelties. First, PIM is a novel and unified deep architecture, containing a Face Frontalization sub-Net (FFN) and a Discriminative Learning sub-Net (DLN), which are jointly learned from end to end. Second, FFN is a well-designed dual-path Generative Adversarial Network (GAN) which simultaneously perceives global structures and local details, incorporated with an unsupervised cross-domain adversarial training and a “learning to learn” strategy for high-fidelity and identity-preserving frontal view synthesis. Third, DLN is a generic Convolutional Neural Network (CNN) for face recognition with our enforced cross-entropy optimization strategy for learning discriminative yet generalized feature representation. Qualitative and quantitative experiments on both controlled and in-the-wild benchmarks demonstrate the superiority of the proposed model over the state-of-the-arts.

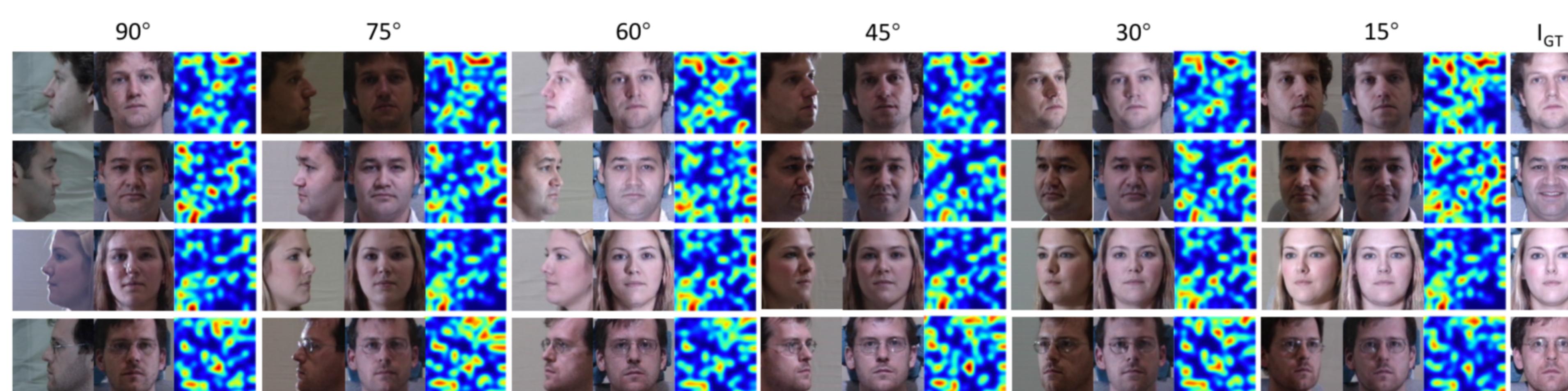


Fig. 1. Pose invariant face recognition in the wild. Each row shows a distinct identity under different poses along with other unconstrained factors (like expression, illumination, etc.), recovered frontal faces and learned facial representations (smoothed for better visualization, with blue indicating zero values) with our proposed PIM. The representations are extracted from the penultimate layer of PIM. The ground truth frontal face images are provided in the right-most column. These examples indicate that the facial representations learned by PIM are robust to pose variance, and the recovered frontal face images retain the intrinsic global structures and local details. Best viewed in color.

Motivation

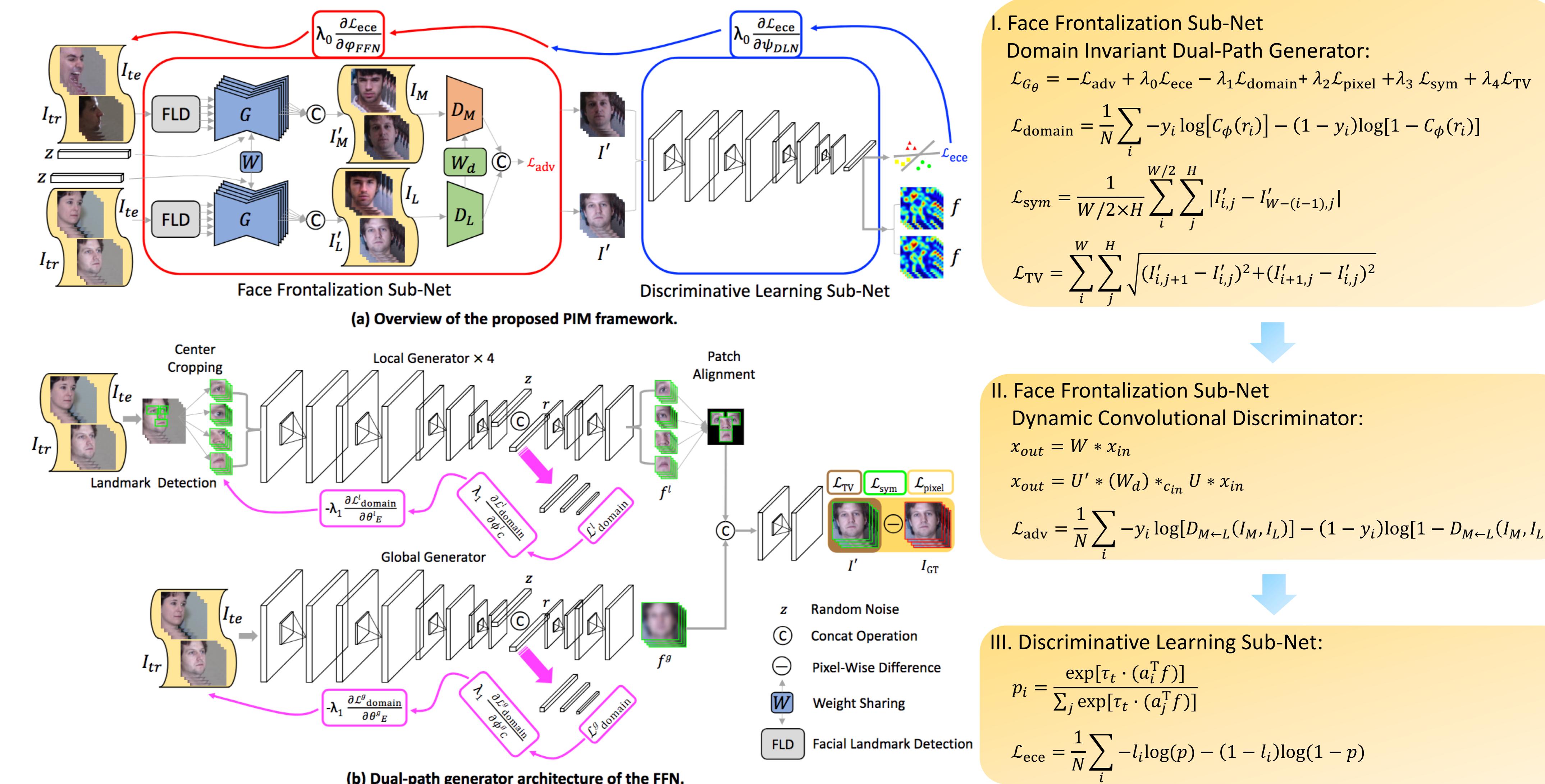


Fig. 2. Pose Invariant Model (PIM) for face recognition in the wild. The PIM contains an Face Frontalization sub-Net (FFN) and a Discriminative Learning sub-Net (DLN) that jointly learn end-to-end. FFN is a dual-path (i.e., simultaneously perceiving global structures and local details) GAN augmented by (1) unsupervised cross-domain (i.e., I_{tr} and I_{te}) adversarial training and (2) a siamese discriminator with a “learning to learn” strategy — convolutional parameters (i.e., W_d) dynamically predicted by the “learner” D_L of the discriminator and transferred to D_M . DLN is a generic Convolutional Neural Network (CNN) for face recognition optimized by Towards Pose Invariant Face Recognition in the Wild.

Contributions

- We propose a Pose Invariant Model (PIM) for face recognition in the wild. PIM is a novel and unified deep neural network containing a Face Frontalization sub-Net (FFN) and a Discriminative Learning sub-Net (DLN) that jointly learn in an end-to-end way to allow them to mutually boost each other.
- FFN is a carefully designed dual-path (i.e., simutaneously perceiving global structures and local details) Generative Adversarial Network (GAN) incorporating unsupervised cross-domain adversarial training and a “learning to learn” strategy using siamese discriminator with dynamic convolution for high-fidelity and identity-preserving frontal view synthesis.
- DLN is a generic Convolutional Neural Network (CNN) for face recognition with our proposed enforced cross- entropy optimization strategy for learning discriminative yet generalized feature representations with large intra- class affinity and inter-class separability.
- We develop effective and novel training strategies for FFN, DLN and the whole deep architecture, which generate powerful face representations.
- As a by-product, the recovered frontal face images by PIM can also be utilized by conventional descriptors and learning algorithms so as to eliminate the negative effects from unconstrained conditions.

Based on the above model innovations and technical contributions, we present a high-performance pose invariant face recognition system. It achieves state-of-the-art performance on Multi-PIE, CFP and LFW benchmark datasets.

Acknowledgement

The work of Jian Zhao was partially supported by China Scholarship Council (CSC) grant 201503170248.

The work of Junliang Xing was partially supported by the National Science Foundation of China 61672519.

The work of Jiashi Feng was partially supported by NUS startup R-263-000-C08-133, MOE Tier-I R-263-000-C21-112, NUS IDS R-263-000-C67-646 and ECRA R-263-000-C87-133.

Method	$\pm 90^\circ$	$\pm 75^\circ$	$\pm 60^\circ$	$\pm 45^\circ$	$\pm 30^\circ$	$\pm 15^\circ$
b1	18.80	63.80	92.20	98.30	99.20	99.40
b2	33.00	76.10	95.20	97.90	99.20	99.80
CPF	-	-	-	71.65	81.05	89.45
Hassner et al.	-	-	-	44.81	74.68	89.59
FV	24.53	45.51	68.71	80.33	87.21	93.30
HPN	29.82	47.57	61.24	72.77	78.26	84.23
FIP_40	31.37	49.10	69.75	85.54	92.98	96.30
C-CNN	47.26	60.66	74.38	89.02	94.05	95.97
TP-GAN	64.03	84.10	92.93	98.58	99.85	99.78
PIM1	71.60	92.50	97.00	98.60	99.30	99.40
PIM2	75.00	91.20	97.70	98.30	99.40	99.80

Tab. 1. Rank-1 recognition rates (%) across views, minor expressions and illuminations under Multi-PIE Setting-1.

Method	$\pm 90^\circ$	$\pm 75^\circ$	$\pm 60^\circ$	$\pm 45^\circ$	$\pm 30^\circ$	$\pm 15^\circ$
b1	15.50	55.10	85.90	97.10	98.40	98.60
b2	27.10	68.70	91.40	97.70	98.60	99.10
FIP	-	-	45.90	64.10	80.70	90.70
MVP	-	-	60.10	72.90	83.70	92.80
CPF	-	-	61.90	79.90	88.50	95.00
DR-GAN	-	-	83.20	86.20	90.10	94.00
TP-GAN	64.64	77.43	87.72	95.38	98.06	98.68
PIM1	81.30	92.70	96.60	97.30	98.40	98.80
PIM2	86.50	95.00	98.10	98.50	99.00	99.30

Tab. 2. Rank-1 recognition rates (%) across views, illuminations and sessions under Multi-PIE Setting-2.

Method	Frontal-Profile			Frontal-Frontal		
	Acc	EER	AUC	Acc	EER	AUC
Fv+DML	58.47 \pm 3.51	38.54 \pm 1.59	65.74 \pm 2.02	91.18 \pm 1.34	8.62 \pm 1.19	97.25 \pm 0.60
LBP+Sub-SML	70.02 \pm 2.14	29.60 \pm 2.11	77.98 \pm 1.86	83.54 \pm 2.40	16.00 \pm 1.74	91.70 \pm 1.55
HoG+Sub-SML	77.31 \pm 1.61	22.20 \pm 1.18	85.97 \pm 1.03	88.34 \pm 1.33	11.45 \pm 1.35	94.83 \pm 0.80
Fv+Sub-SML	80.63 \pm 2.12	19.28 \pm 1.60	88.53 \pm 1.58	91.30 \pm 0.85	8.85 \pm 0.74	96.87 \pm 0.39
Deep Features	84.91 \pm 1.82	14.97 \pm 1.98	93.00 \pm 1.55	95.40 \pm 0.69	3.48 \pm 0.67	99.43 \pm 0.31
Triple Embedding	89.17 \pm 2.35	8.85 \pm 0.99	97.00 \pm 0.53	96.93 \pm 0.61	2.51 \pm 0.81	99.68 \pm 0.16
Chen et al.	91.97 \pm 1.70	8.00 \pm 1.68	97.70 \pm 0.82	98.41 \pm 0.45	1.54 \pm 0.43	99.89 \pm 0.06
Light CNN-29	92.47 \pm 1.44	8.71 \pm 1.80	97.77 \pm 0.76	99.64 \pm 0.32	0.57 \pm 0.40	99.92 \pm 0.15
PIM	93.10 \pm 1.01	7.69 \pm 1.29	97.65 \pm 0.62	99.44 \pm 0.36	0.86 \pm 0.49	99.92 \pm 0.10
Human	94.57 \pm 1.10	5.02 \pm 1.07	98.92 \pm 0.46	96.24 \pm 0.67	5.34 \pm 1.79	98.19 \pm 1.13

Tab. 3. Face recognition performance (%) comparison on CFP.



Results

