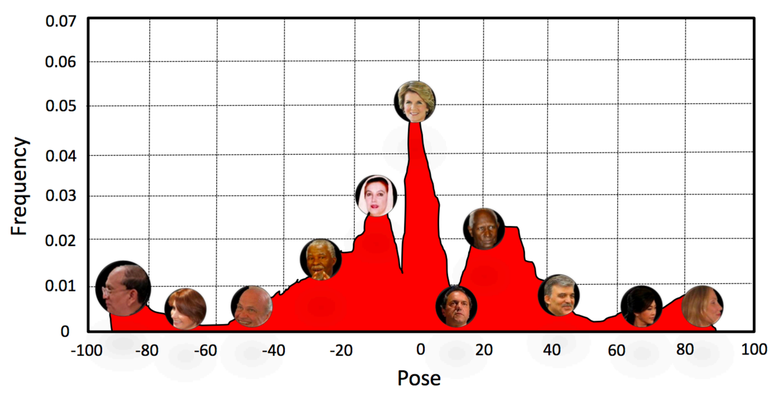
Dual-Agent GANs for Photorealistic and Identity Preserving Profile Face Synthesis

Jian Zhao[1,2], Lin Xiong[3], Karlekar Jayashree[3], Jianshu Li[1], Fang Zhao[1], Zhecan Wang[4], Sugiri Pranata[3], Shengmei Shen[3], Shuicheng Yan[1,5], and Jiashi Feng[1]
[1]National University of Singapore, [2]National University of Defense Technology,
[3]Panasonic R&D Center Singapore, [4]Franklin. W. Olin College of Engineering, [5]Qihoo 360 AI Institute

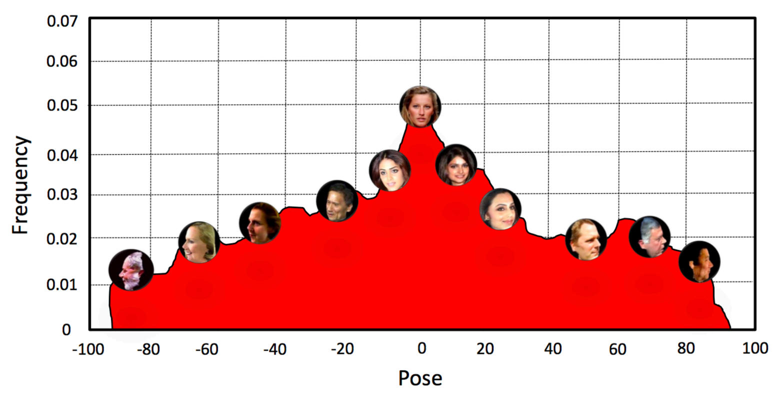
Abstract

Synthesizing realistic profile faces is promising for more efficiently training deep pose-invariant models for large-scale unconstrained face recognition, by populating samples with extreme poses and avoiding tedious annotations. However, learning from synthetic faces may not achieve the desired performance due to the discrepancy between distributions of the synthetic and real face images. To narrow this gap, we propose a **D**ual-**A**gent **G**enerative **A**dversarial **N**etwork (DA-GAN) model, which can improve the realism of a face simulator's output using unlabeled real faces, while preserving the identity information during the realism refinement. The dual agents are specifically designed for distinguishing real *v.s.* fake and identities simultaneously. In particular, we employ an off-the-shelf 3D face model as a simulator to generate profile face images with varying poses. DA-GAN leverages a fully convolutional network as the generator to generate high-resolution images and an auto-encoder as the discriminator with the dual agents. Besides the novel architecture, we make several key modifications to the standard GAN to preserve pose and texture, preserve identity and stabilize training process: (i) a pose perception loss; (ii) an identity perception loss; (iii) an adversarial loss with a boundary equilibrium regularization term. Experimental results show that DA-GAN not only presents compelling perceptual results but also significantly outperforms state-of-the-arts on the large-scale and challenging NIST IJB-A unconstrained face recognition benchmark. In addition, the proposed DA-GAN is also promising as a new approach for solving generic transfer learning problems more effectively. DA-GAN is the foundation of our submissions to NIST IJB-A 2017 face recognition competitions, where we won the 1st places on the tracks of verification and identification.

Motivation



(a) Extremely unbalanced pose distribution.

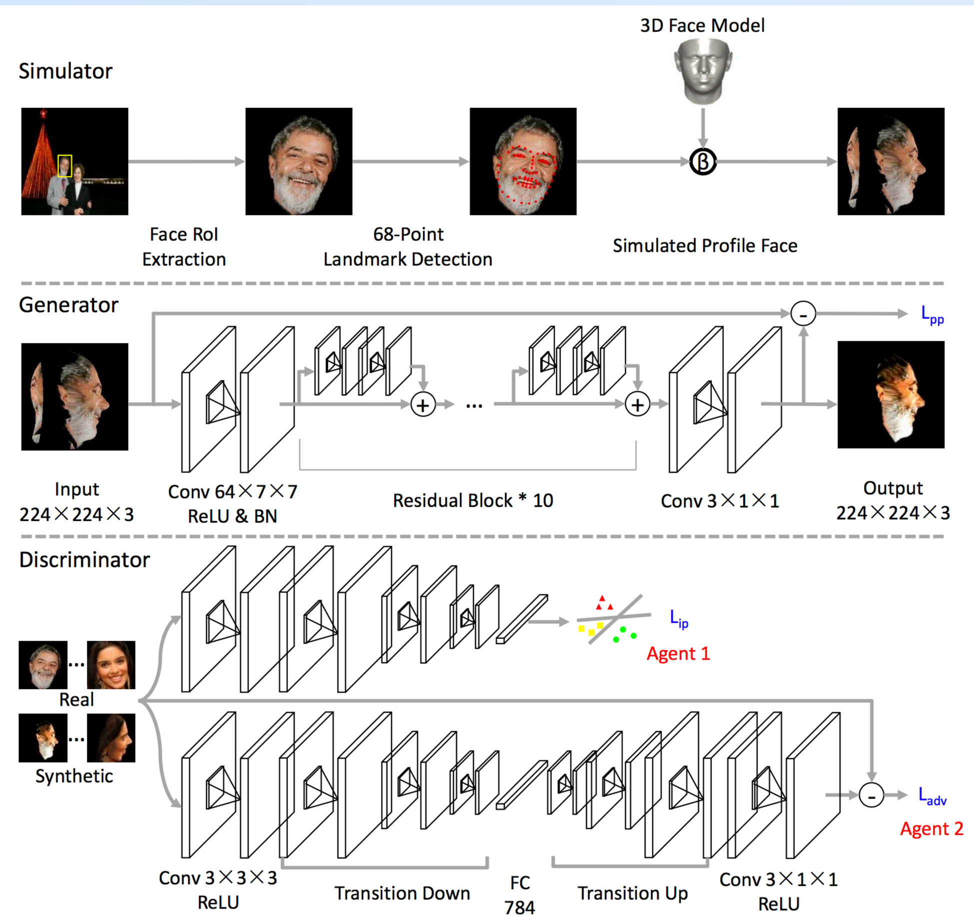


(b) Well balanced pose distribution with DA-GAN.

Contributions

1. We propose a novel **D**ual-**A**gent **G**enerative **A**dversarial **N**etwork (DA-GAN) for photorealistic and identity preserving profile face synthesis even under extreme poses.
2. The proposed dual-agent architecture effectively combines prior knowledge from data distribution (adversarial training) and domain knowledge of faces (pose and identity perception losses) to exactly recover the lost information inherent in projecting a 3D face into the 2D image space.
3. We present qualitative and quantitative experiments showing the possibility of a "recognition via generation" framework and achieve the top performance on the challenging NIST IJB-A unconstrained face recognition benchmark without extra human annotation efforts by training deep neural networks on the refined face images together with real images. To our best knowledge, our proposed DA-GAN is the first model that is effective for automatically generating augmented data for face recognition in challenging conditions and indeed improves performance. **DA-GAN won the 1st places on verification and identification tracks in the NIST IJB-A 2017 face recognition competitions. (We submitted our results for both verification and identification protocols to NIST IJB-A 2017 face recognition competition committee on 29th, March, 2017. We received the official notification on our top performance on both tracks on 26th, Apirl, 2017. The IJB-A benchmark dataset, relevant information and leaderboard can be found at https://www.nist.gov/programs-projects/face-challenges.)**

Method



1. Simulator: Face RoI Extraction -> 68-Point Landmark Detection -> 3D Face Model -> Simulated Profile Faces with Pre-Defined Yaw Angles
2. Generator:

$$\tilde{x} := G_\theta(x) \quad \mathcal{L}_{G_\theta} = (-\mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{ip}) + \lambda_2 \mathcal{L}_{pp} \quad \mathcal{L}_{pp} = \frac{1}{W \times H} \sum_i^W \sum_j^H |x_{i,j} - \tilde{x}_{i,j}|$$

3. Discriminator:

$$\mathcal{L}_{adv} = \sum_j |y_j - D_\phi(y_j)| - k_t \sum_i |\tilde{x}_i - D_\phi(\tilde{x}_i)| \quad \mathcal{L}_{D_\phi} = \mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{ip}$$

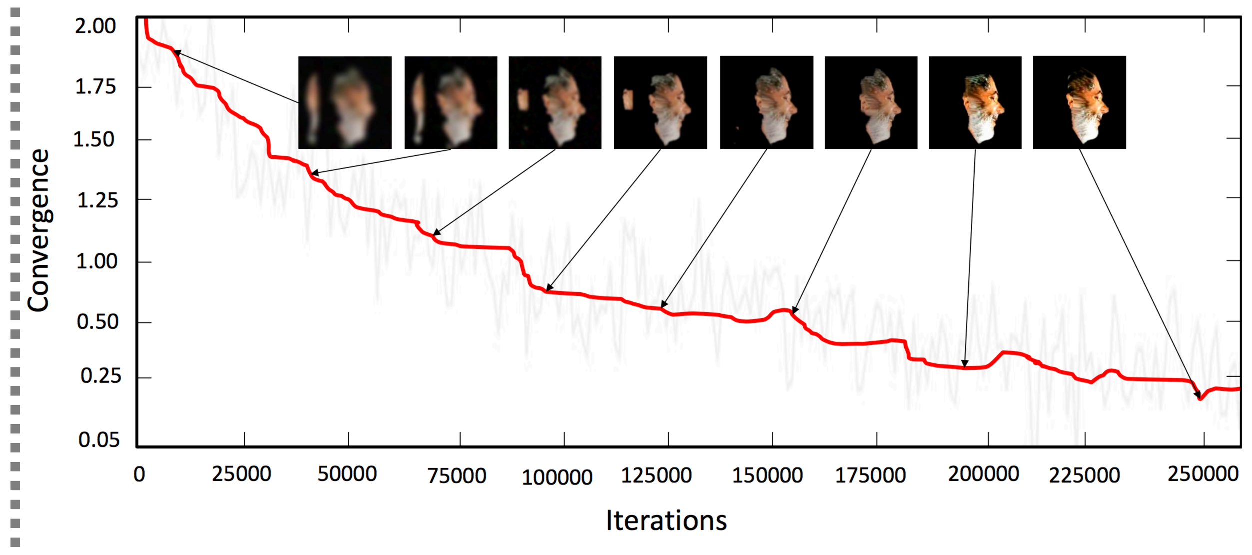
$$k_{t+1} = k_t + \alpha(\gamma \sum_j |y_j - D_\phi(y_j)| - \sum_i |\tilde{x}_i - D_\phi(\tilde{x}_i)|)$$

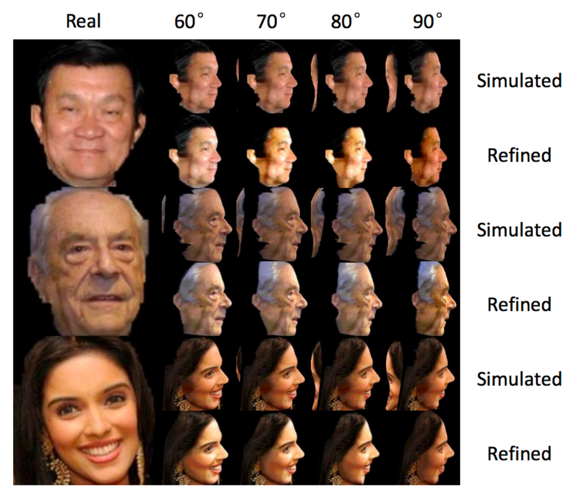
$$\mathcal{L}_{ip} = \frac{1}{N} \sum_j -(Y_j log(D_\phi(y_j)) + (1 - Y_j) log(1 - D_\phi(y_j)))$$
$$+ \frac{1}{N} \sum_i -(Y_i log(D_\phi(\tilde{x}_i)) + (1 - Y_i) log(1 - D_\phi(\tilde{x}_i))),$$

4. Optimization:

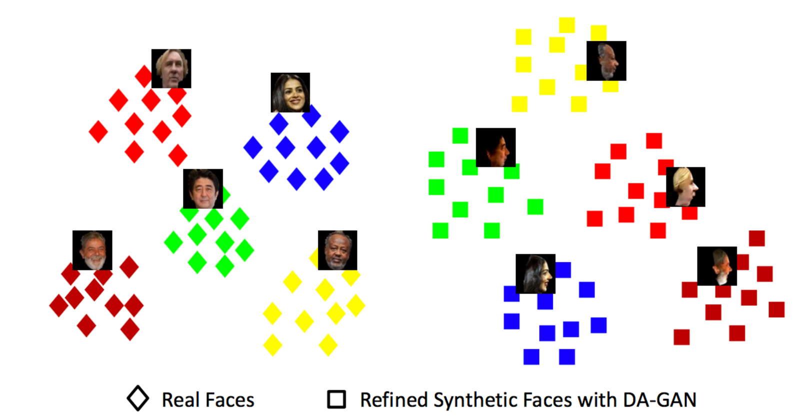
$$\begin{cases} \mathcal{L}_{D_\phi} = \mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{ip}, \\ \mathcal{L}_{G_\theta} = (-\mathcal{L}_{adv} + \lambda_1 \mathcal{L}_{ip}) + \lambda_2 \mathcal{L}_{pp}. \end{cases} \quad \mathcal{L}_{con} = \sum_j |y_j - D_\phi(y_j)| + |\gamma \sum_j |y_j - D_\phi(y_j)| - \sum_i |\tilde{x}_i - D_\phi(\tilde{x}_i)||$$

Results





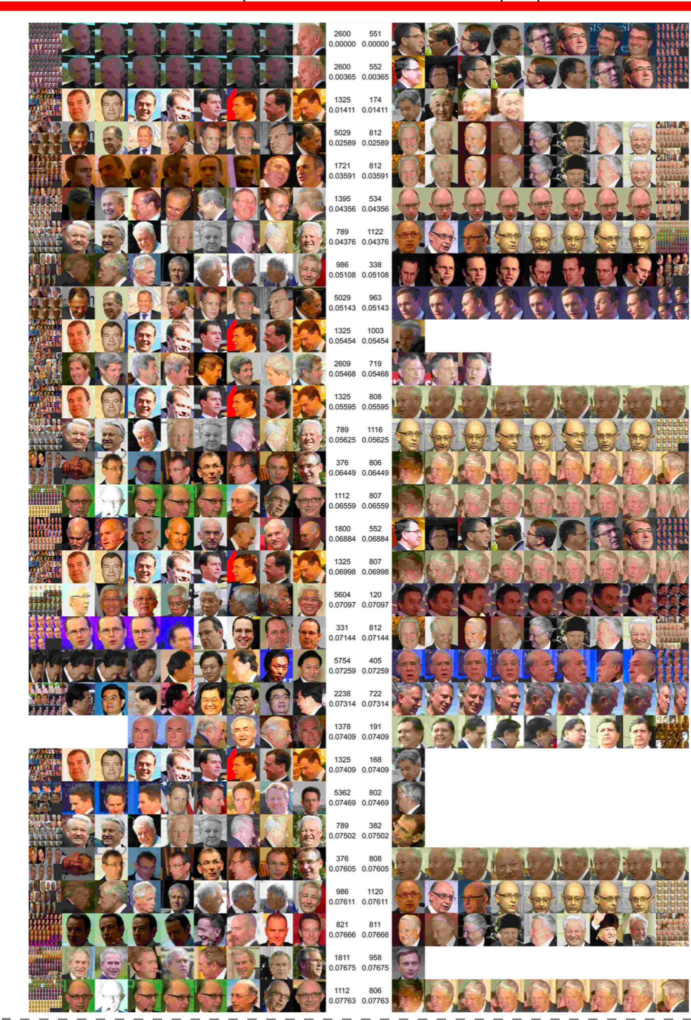
(a) Refined results of DA-GAN.



(b) Feature space of real faces and DA-GAN synthetic faces.

Method	Face verification		
	TAR @ FAR=0.10	TAR @ FAR=0.01	TAR @ FAR=0.001
OpenBR (15)	0.433 ± 0.006	0.236 ± 0.009	0.104 ± 0.014
GOTS (15)	0.627 ± 0.012	0.406 ± 0.014	0.198 ± 0.008
Pooling faces (11)	0.631	0.309	-
LSFS (30)	0.895 ± 0.013	0.733 ± 0.034	0.514 ± 0.060
Deep Multi-pose (1)	0.911	0.787	-
DCNN_{manual} (4)	0.947 ± 0.011	0.787 ± 0.043	-
Triplet Similarity (27)	0.945 ± 0.002	0.790 ± 0.030	0.590 ± 0.050
VGG-Face (23)	-	0.805 ± 0.030	-
PAMs (19)	0.652 ± 0.037	0.826 ± 0.018	-
DCNN_{fusion} (3)	0.967 ± 0.009	0.838 ± 0.042	-
Masi *et al.* (20)	-	0.886	0.725
Triplet Embedding (27)	0.964 ± 0.005	0.900 ± 0.010	0.813 ± 0.020
All-In-One (25)	0.976 ± 0.004	0.922 ± 0.010	0.823 ± 0.020
Template Adaptation (8)	0.979 ± 0.004	0.939 ± 0.013	0.836 ± 0.027
NAN (34)	0.978 ± 0.003	0.941 ± 0.008	0.881 ± 0.011
L_2-softmax (24)	0.984 ± 0.002	0.970 ± 0.004	0.943 ± 0.005
b-1	0.989 ± 0.003	0.963 ± 0.007	0.920 ± 0.006
b-2	0.978 ± 0.003	0.950 ± 0.009	0.901 ± 0.008
DA-GAN (ours)	$\mathbf{0.991 \pm 0.003}$	$\mathbf{0.976 \pm 0.007}$	$\mathbf{0.930 \pm 0.005}$

Method	Face identification			
	FNIR @ FPIR=0.10	FNIR @ FPIR=0.01	Rank1	Rank5
OpenBR (15)	0.851 ± 0.028	0.934 ± 0.017	0.246 ± 0.011	0.375 ± 0.008
GOTS (15)	0.765 ± 0.033	0.953 ± 0.024	0.433 ± 0.021	0.595 ± 0.020
B-CNN (7)	0.659 ± 0.032	0.857 ± 0.027	0.588 ± 0.020	0.796 ± 0.017
LSFS (30)	0.387 ± 0.032	0.617 ± 0.063	0.820 ± 0.024	0.929 ± 0.013
Pooling faces (11)	-	-	0.846	0.933
Deep Multi-pose (1)	0.250	0.480	0.846	0.927
DCNN_{manual} (4)	-	-	0.852 ± 0.018	0.937 ± 0.010
Triplet Similarity (27)	0.246 ± 0.014	0.444 ± 0.065	0.880 ± 0.015	0.950 ± 0.007
VGG-Face (23)	0.33 ± 0.031	0.539 ± 0.077	0.913 ± 0.011	-
PAMs (19)	-	-	0.840 ± 0.012	0.925 ± 0.008
DCNN_{fusion} (3)	0.210 ± 0.033	0.423 ± 0.094	0.903 ± 0.012	0.965 ± 0.008
Masi *et al.* (20)	-	-	0.906	0.962
Triplet Embedding (27)	0.137 ± 0.014	0.247 ± 0.030	0.932 ± 0.010	-
Template Adaptation (8)	0.118 ± 0.016	0.226 ± 0.049	0.928 ± 0.010	0.977 ± 0.004
All-In-One (25)	0.113 ± 0.014	0.208 ± 0.020	0.947 ± 0.008	-
NAN (34)	0.083 ± 0.009	0.183 ± 0.041	0.958 ± 0.005	0.980 ± 0.005
L_2-softmax (24)	0.044 ± 0.006	0.085 ± 0.041	0.973 ± 0.005	-
b-1	0.068 ± 0.010	0.125 ± 0.035	0.966 ± 0.006	0.987 ± 0.003
b-2	0.108 ± 0.008	0.179 ± 0.042	0.960 ± 0.007	0.982 ± 0.004
DA-GAN (ours)	$\mathbf{0.051 \pm 0.009}$	$\mathbf{0.110 \pm 0.039}$	$\mathbf{0.971 \pm 0.007}$	$\mathbf{0.989 \pm 0.003}$



Acknowledgement

The work of Jian Zhao was partially supported by China Scholarship Council (CSC) grant 201503170248.
The work of Jiashi Feng was partially supported by National University of Singapore startup grant R-263-000-C08-133, Ministry of Education of Singapore AcRF Tier One grant R-263-000-C21-112 and NUS IDS grant R-263-000-C67-646.

Contact: Jian Zhao, Ph.D. candidate, Email: zhaojian90@u.nus.edu, Phone: (65) 9610 7176, Homepage: https://zhaoj9014.github.io/ | Lin Xiong, Research Engineer, Email: lin.xiong@sg.panaosinic.com, Phone: (65) 83752875
Address: Vision and Machine Learning Lab, E4-#08-24, 4 Engineering Drive 3, National University of Singapore, Singapore 117583 | Panasonic R&D Center Singapore, Core Technology Group, 202 Bedok South Avenue 1 #02-11, Singapore 469332