Disentangled Representation Learning GAN for Pose-Invariant Face Recognition [CVPR `17]

20189008 정병의

KAIST

Review

- A Zero-Shot Framework for Sketch Based Image Retrieval [ECCV `18] (Speaker. Doheon Lee)

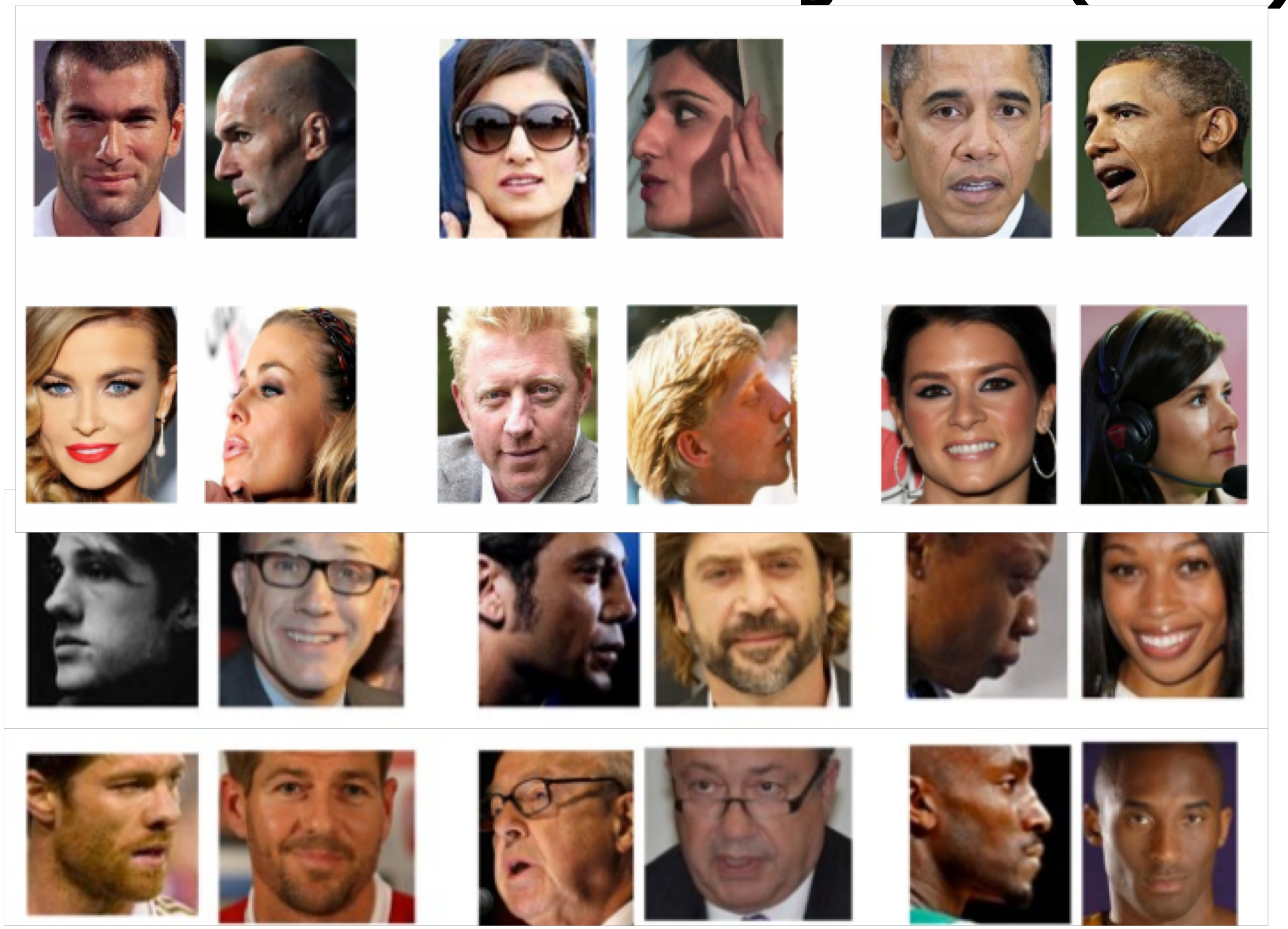
- Problems of Previous Works
  - SBIR is usually used for fine-grained IR.
  - They are focused on class-based retrieval.
  - Shape or attributed-based retrieval are important.

- Solution: Zero-shot learning for coarse-grained IR
  - Zero-shot Learning to recognize images of novel classes.
  - Proposed a new benchmark for zero-shot SBIR.
  - Proposed a generative approach for the SBIR task.

Table of Contents

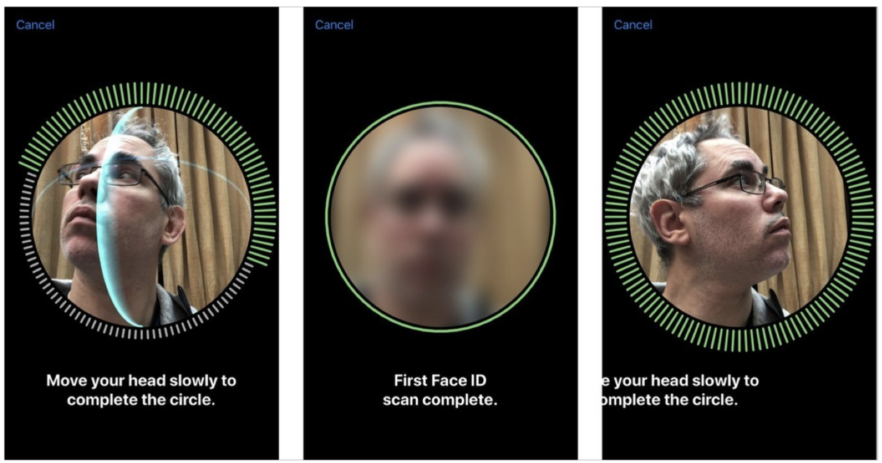
Introduction

- **Pose-Invariant Face Recognition (PIFR)**



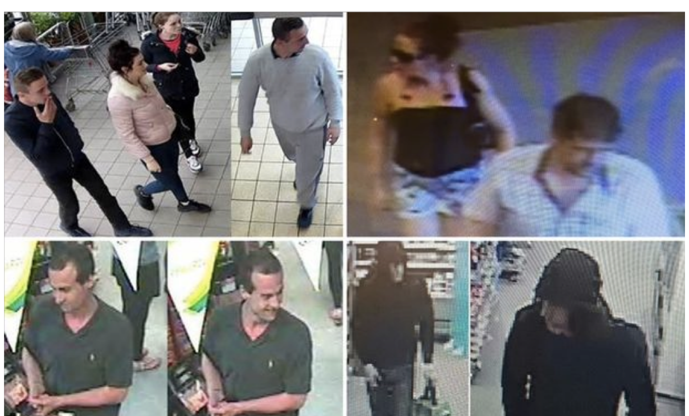
Introduction

- **Pose-Invariant Face Recognition (PIFR)**



Face ID & Fraud detection

Finding missing persons



Many face images are Not taken in frontally!

Previous Works

- **Frontal to Profile Face Verification in the Wild [IEEE '16]**
  - **Celebrities in Frontal-Profile (CFP) dataset.**
  - **State-of-the-art algorithms are degraded more than 10% from Frontal-Frontal to Frontal-Profile verification.**

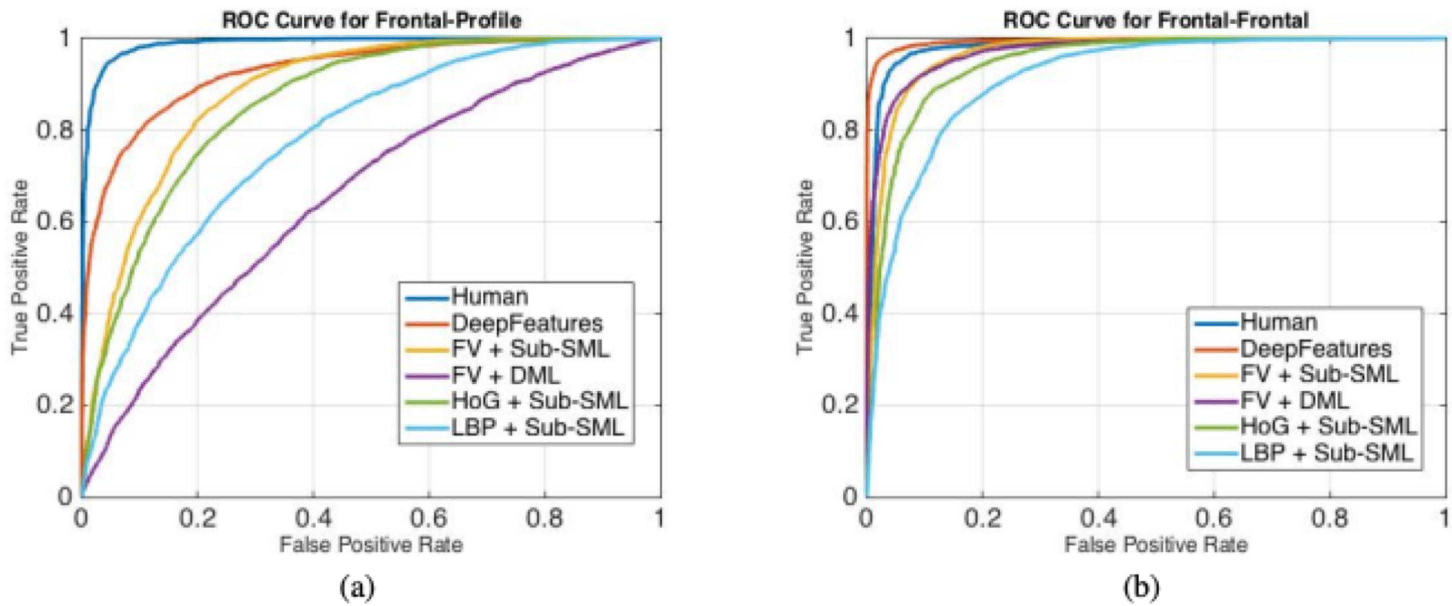
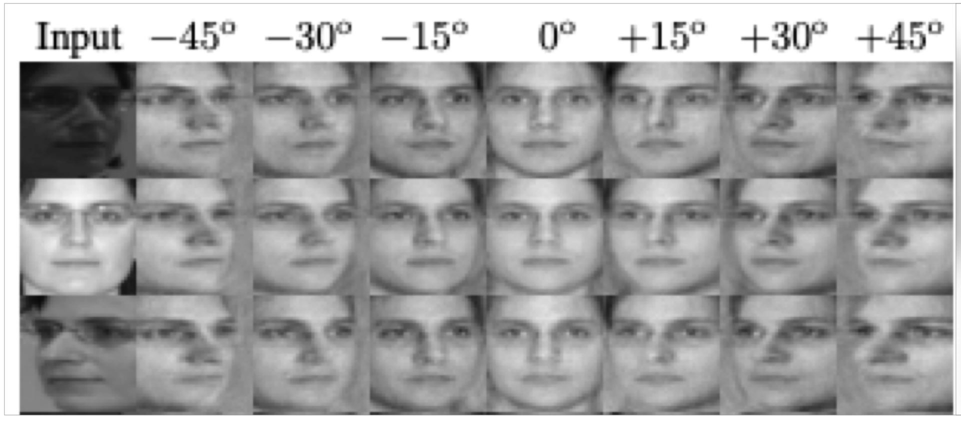


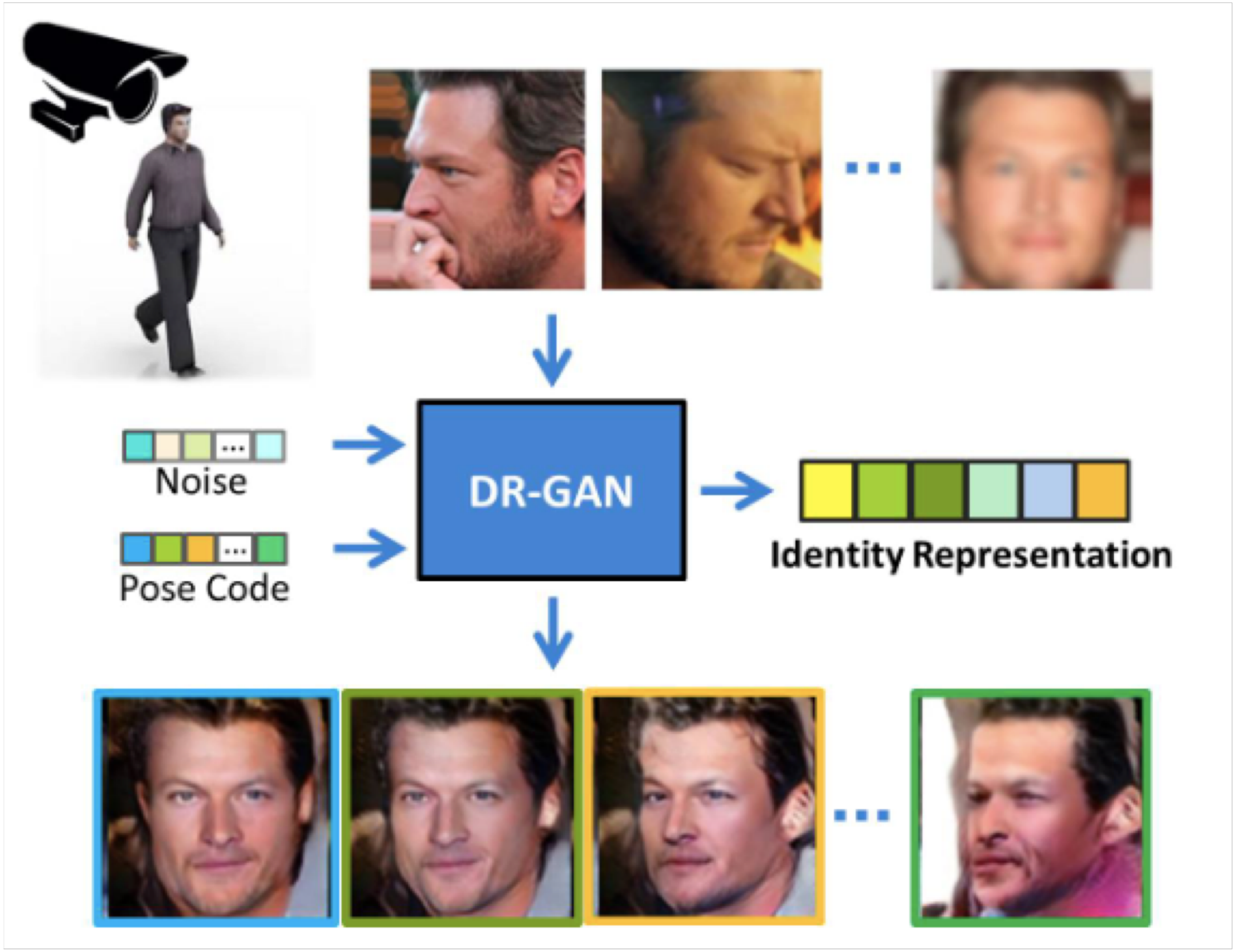
Figure 5. Roc curve for (a) Frontal-Profile and (b) Frontal-Frontal

Previous Works

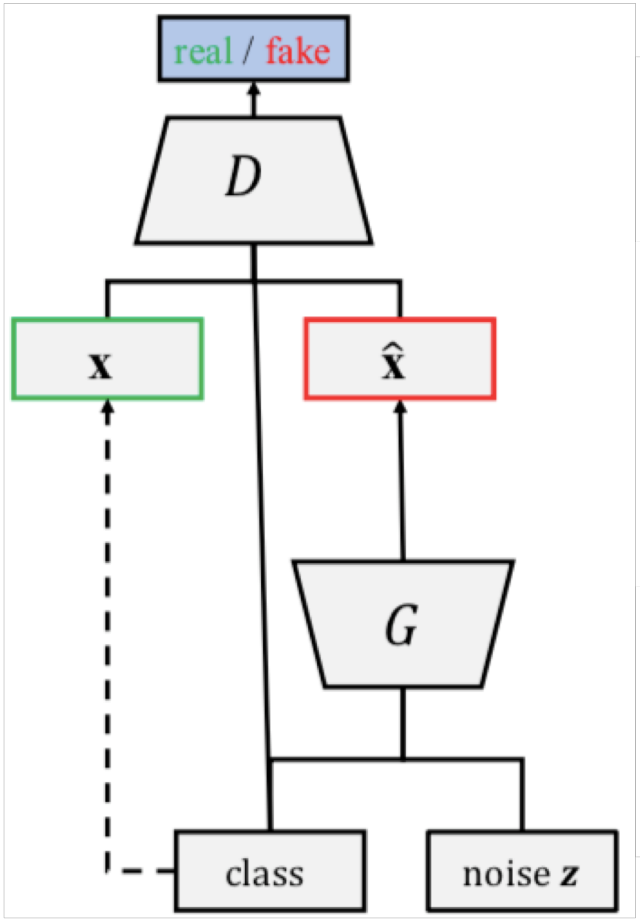
- **Face Frontalization**
  - **Limited in same scene face images.**
  - **Limited in near frontal images.**
  - **Not suitable for in-the-wild data.**
  - **Only handle single image.**



Overview



Conditional GAN



Generator G, Discriminator D

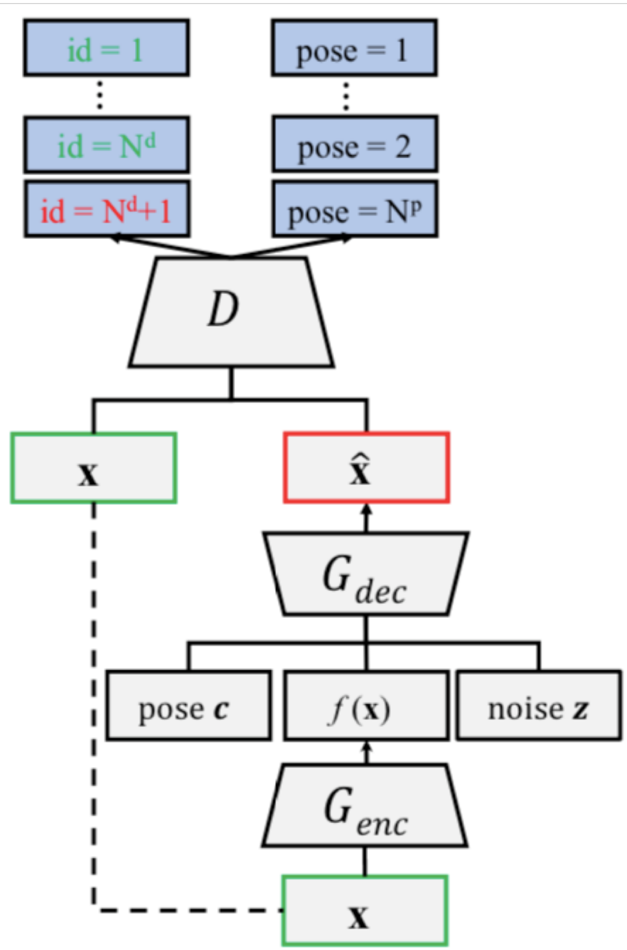
$$\min_G \max_D V(D,G) = E_{\mathbf{x}\sim p_d(\mathbf{x})}[\log D(\mathbf{x})] + E_{\mathbf{z}\sim p_z(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))]. \quad (1)$$

In practice,
maximizing log(D(G(z))) is better than
instead of minimizing log (1- D(G(z)))

$$\max_D V_D(D,G) = E_{\mathbf{x}\sim p_d(\mathbf{x})}[\log D(\mathbf{x})] + E_{\mathbf{z}\sim p_z(\mathbf{z})}[\log(1 - D(G(\mathbf{z})))], \quad (2)$$

$$\max_G V_G(D,G) = E_{\mathbf{z}\sim p_z(\mathbf{z})}[\log(D(G(\mathbf{z})))]. \quad (3)$$

Single-Image DR-GAN



Generator xˆ = G(x, c, z)
D attempts to classify xˆ as fake
Maximize the probability of xˆ being classified as a fake

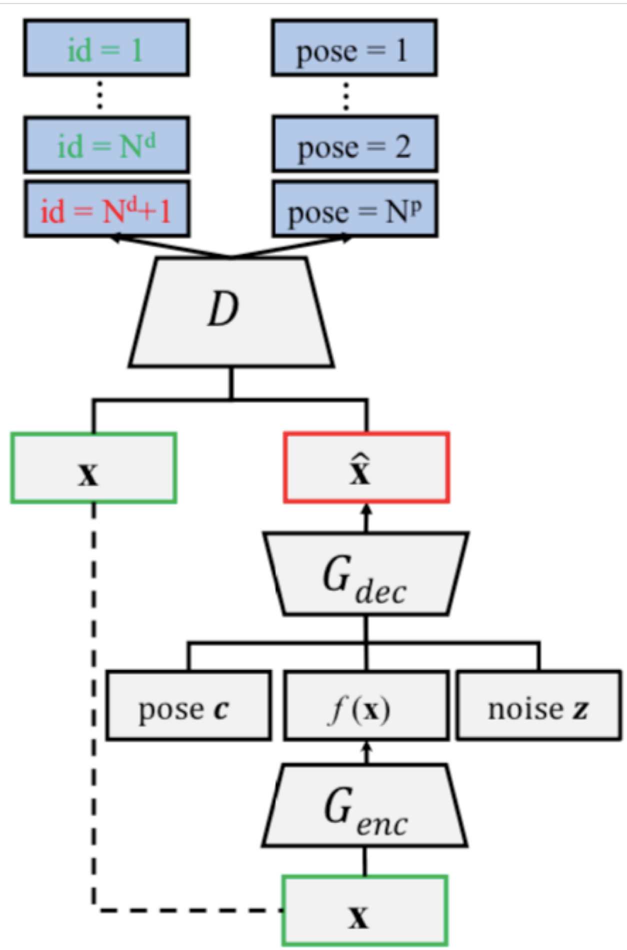
$$\max_D V_D(D,G) = E_{\mathbf{x},\mathbf{y}\sim p_d(\mathbf{x},\mathbf{y})}[\log D^d_{y^d}(\mathbf{x}) + \log D^p_{y^p}(\mathbf{x})] + E_{\substack{\mathbf{x},\mathbf{y}\sim p_d(\mathbf{x},\mathbf{y}),\\ \mathbf{z}\sim p_z(\mathbf{z}),\mathbf{c}\sim p_c(\mathbf{c})}}[\log(\underline{D^d_{N^d+1}}(G(\mathbf{x},\mathbf{c},\mathbf{z})))], \quad (4)$$

The goal of G is to fool D to classify xˆ

$$\max_G V_G(D,G) = E_{\substack{\mathbf{x},\mathbf{y}\sim p_d(\mathbf{x},\mathbf{y}),\\ \mathbf{z}\sim p_z(\mathbf{z}),\mathbf{c}\sim p_c(\mathbf{c})}}[\log(D^d_{y^d}(G(\mathbf{x},\mathbf{c},\mathbf{z}))) + \log(D^p_{y^t}(G(\mathbf{x},\mathbf{c},\mathbf{z})))]. \quad (5)$$

yd represents the label for identity
yp represents the label for pose

Multi-Image DR-GAN



$$f(\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_n) = \frac{\sum_{i=1}^{n} \omega_i f(\mathbf{x}_i)}{\sum_{i=1}^{n} \omega_i}. \quad (6)$$

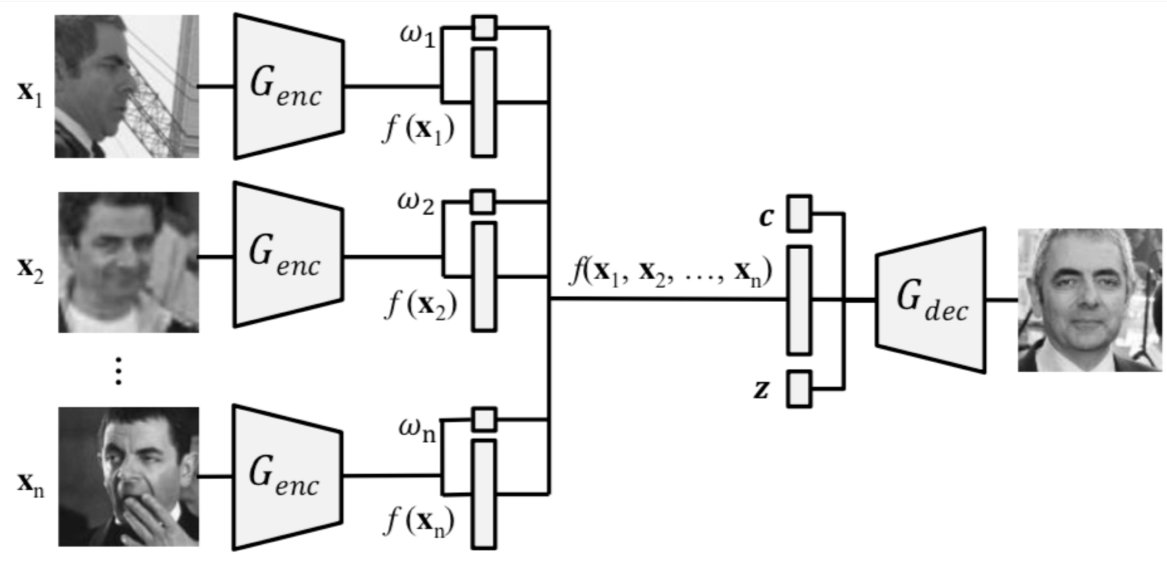
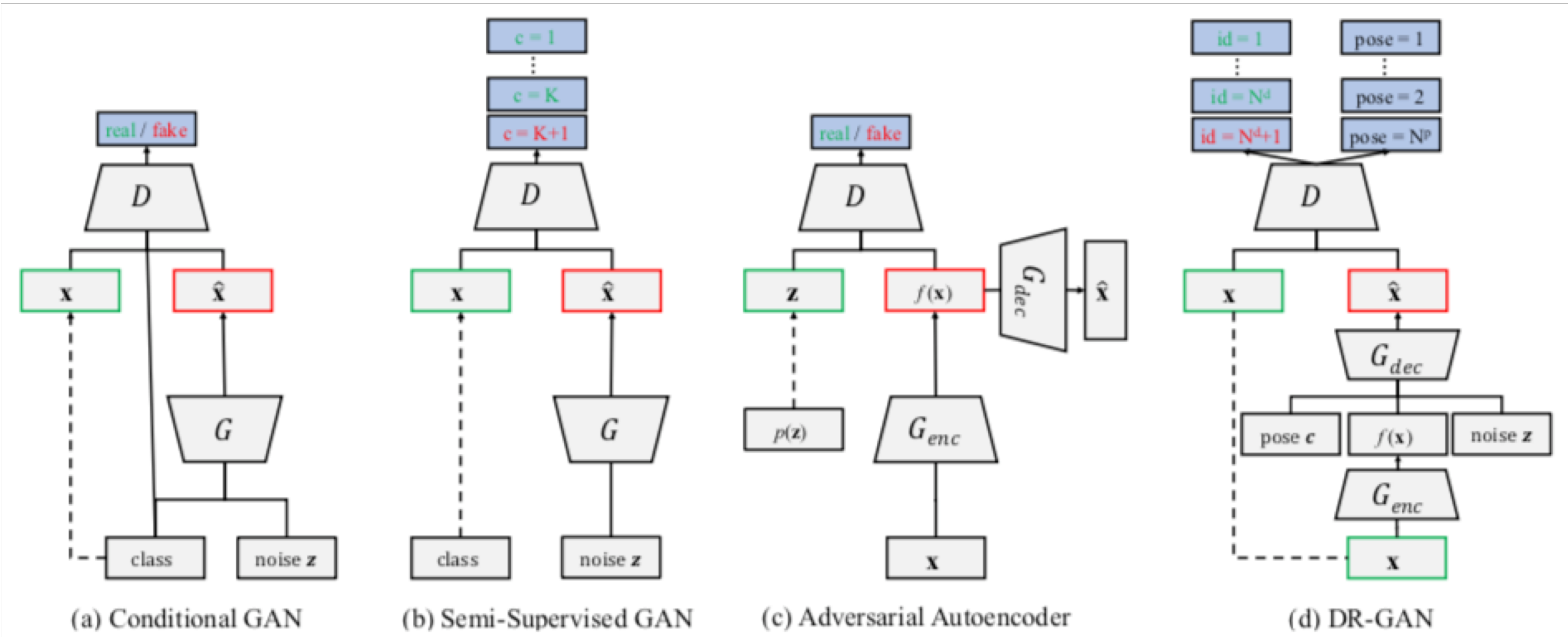


Figure 3: Generator in multi-image DR-GAN. From an image set of a subject, we can fuse the features to a single representation via dynamically learnt coefficients and synthesize images in any pose.

Comparison to Prior GANs



Experimental Result

- **Dataset**
  - **Multi-PIE:** The Largest database for evaluating face recognition under pose, illumination, and expression variations in controlled setting
    - 337 subjects with 9 poses within ±60 deg.
  - **CASIA-WebFace:**
    - 500,000 near-frontal faces of 10,000 subjects
  - **CFP (Celebrities in Frontal-Profile):**
    - 500 subjects each with 10 frontal and 4 profile
  - **IJB-A (IARPA Janus Benchmark A):**
    - 500 subjects from images and video frames

Experimental Result

- **Single vs. Multiple Training Images (CFP)**
  - **Face Identification Performance**

Table 2: Performance comparison on CFP.

Method	Frontal-Frontal	Frontal-Profile
Sengupta et al. [34]	96.40 ± 0.69	84.91 ± 1.82
Sankarana et al. [32]	96.93 ± 0.61	89.17 ± 2.35
Chen et al. [4]	$\mathbf{98.67} \pm 0.36$	91.97 ± 1.70
Human	96.24 ± 0.67	94.57 ± 1.10
DR-GAN: synthetic	97.08 ± 0.62	91.02 ± 1.59
DR-GAN: n=1	97.13 ± 0.68	90.82 ± 0.28
DR-GAN: n=4	97.86 ± 0.75	92.93 ± 1.39
DR-GAN: n=6	97.84 ± 0.79	$\mathbf{93.41} \pm 1.17$

N > 6 : limitation of computation capacity

Experimental Result

- **Result on Benchmark Datasets (Multi-PIE)**
  - **Face Identification Performance**

Table 4: Benchmark comparison on Multi-PIE.

Method	0°	15°	30°	45°	60°	Average
Zhu et al. [44]	94.3	90.7	80.7	64.1	45.9	72.9
Zhu et al. [45]	95.7	92.8	83.7	72.9	60.1	79.3
Yim et al. [40]	**99.5**	**95.0**	88.5	79.9	61.9	83.3
Using $L2$ loss	95.1	90.8	82.7	72.7	57.9	78.3
DR-GAN (n=6)	97.0	94.0	**90.1**	**86.2**	**83.2**	**89.2**

Experimental Result

- **Result on Benchmark Datasets (IJB-A)**
  - **Face Identification Performance**

Table 5: Performance comparison on IJB-A.

Method	Verification		Identification	
	@FAR=.01	@FAR=.001	@Rank-1	@Rank-5
OpenBR [16]	23.6 ± 0.9	10.4 ± 1.4	24.6 ± 1.1	37.5 ± 0.8
GOTS [16]	40.6 ± 1.4	19.8 ± 0.8	44.3 ± 2.1	59.5 ± 2.0
Wang et al. [36]	72.9 ± 3.5	51.0 ± 6.1	82.2 ± 2.3	93.1 ± 1.4
PAM [25]	73.3 ± 1.8	$\mathbf{55.2} \pm 3.2$	77.1 ± 1.6	88.7 ± 0.9
DCNN [3]	$\mathbf{78.7} \pm 4.3$	–	85.2 ± 1.8	93.7 ± 1.0
DR-GAN (avg.)	75.5 ± 2.8	51.8 ± 6.8	84.3 ± 1.3	93.2 ± 0.8
DR-GAN (fuse)	77.4 ± 2.7	53.9 ± 4.3	$\mathbf{85.5} \pm 1.5$	$\mathbf{94.7} \pm 1.1$

False Accept Rates

Experimental Result

- **Single vs. Multiple Testing Images.**

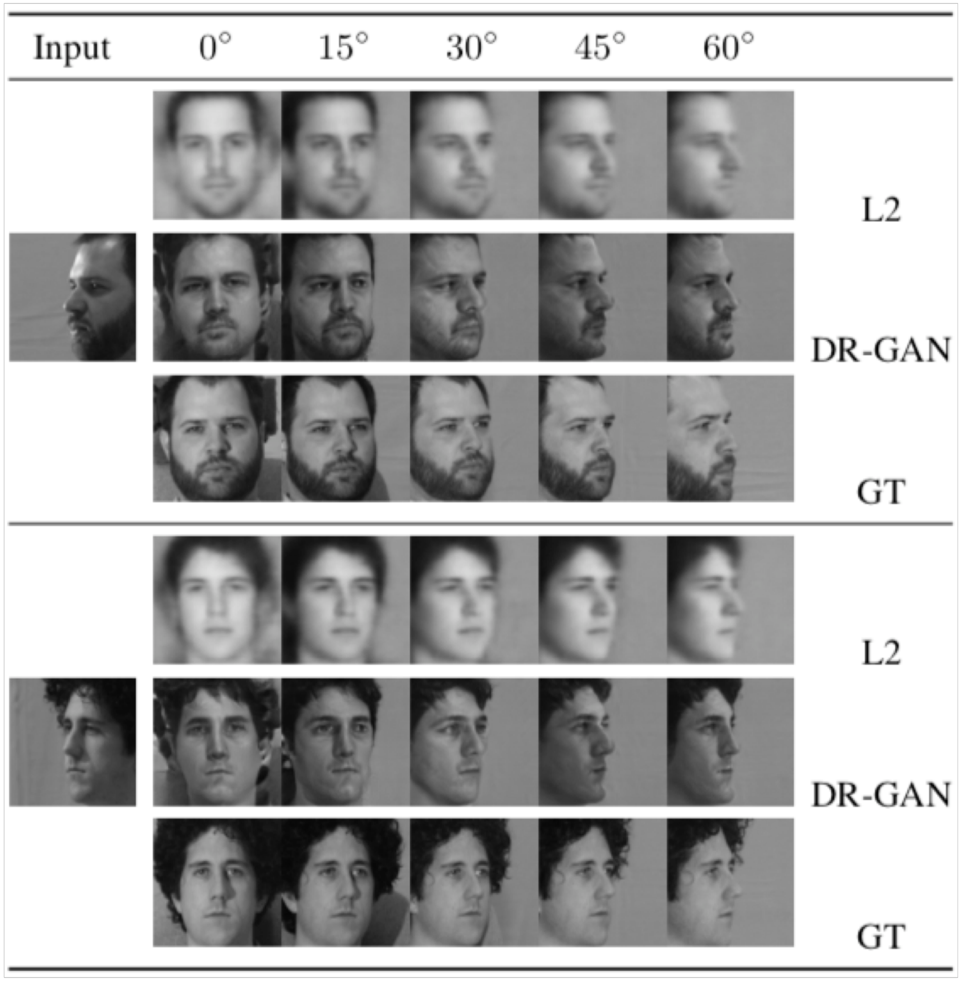
Table 3: Identification rates of three approaches on Multi-PIE.

	n_t	1	2	3	4	5
(n=6)	single-image (avg.)	84.6	91.8	94.1	95.3	95.8
(n=6)	multi-image (avg.)	85.9	92.4	94.5	95.5	95.9
(n=6)	multi-image (fuse)	85.9	92.8	95.1	96.0	96.5

cosine distances of representation

Experimental Result

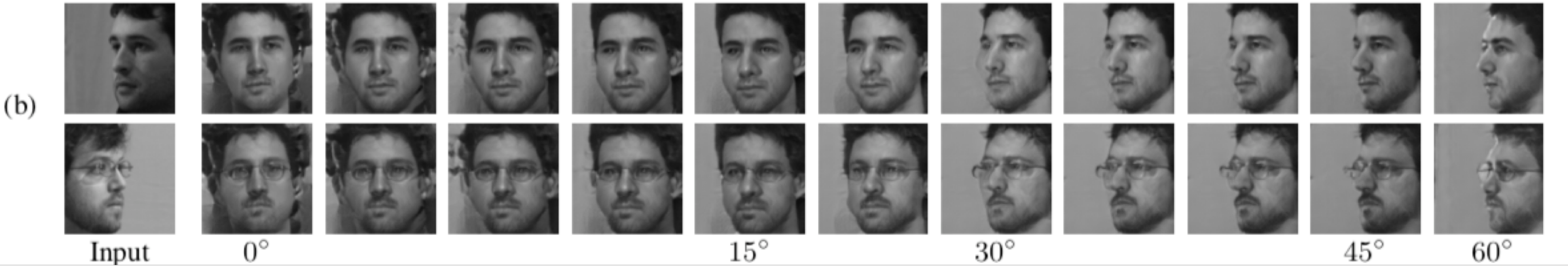
- **DR-GAN (Adversarial) Loss vs. L2 Loss**



Experimental Result

- **Interpolation of Representations (f(x))**

- **Interpolation of c (pose code, degree)**



Experimental Result

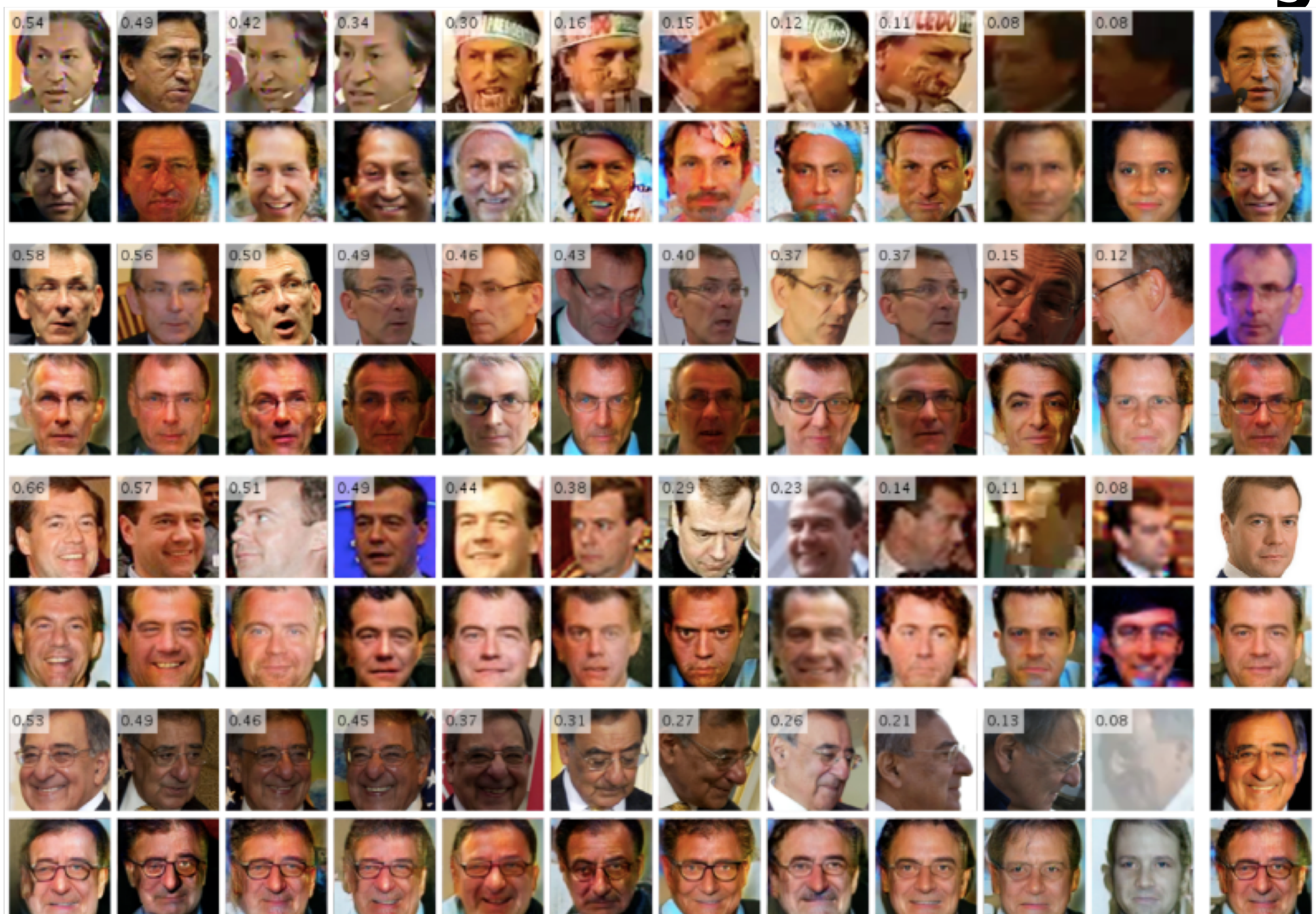
- **Face Rotation in CFP**

"To the best of our knowledge,
this is the first work that is able to *frontalizate a profile-view in-the-wild face image*."

Input
Frontalized
Ground Truth

Experimental Result

- **Face Rotation on IJB-A with Multi-Images**



Conclusion

- Authors proposed DR-GAN for pose-invariant face recognition and face synthesis.
- Their Representation learning.
  - Generative: Image synthesis
  - Discriminative: PIFR
- First work for extreme-pose in-the-wild face frontalization.
- Fusing multiple in-the-wild faces of the same subject into one representation.

THANK YOU