Unsupervised Disentangling of Viewpoint and Residues Variations by Substituting Representations for Robust Face Recognition

ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

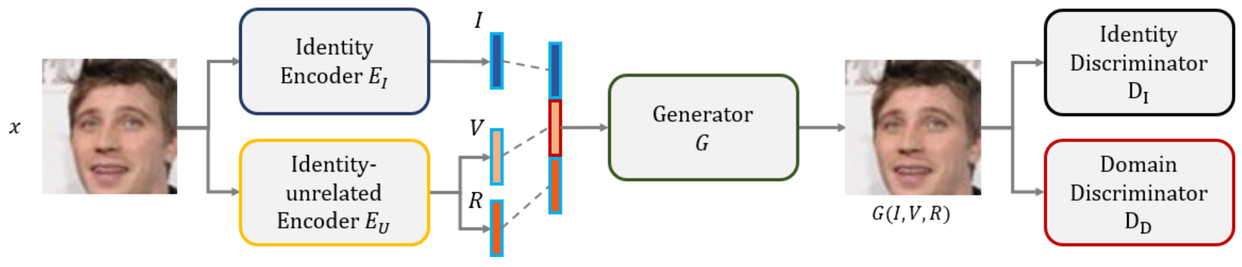
Minsu Kim Joanna Hong Junho Kim Hong Joo Lee Yong Man Ro*
Image and Video Systems Lab, School of Electrical Engineering, KAIST, South Korea

Problem Definition

- Face is generally captured along with diverse factors of variations such as identity, viewpoint, and illumination. These variations pose challenges in face recognition methods in having robust performance in a wild environment.
- In order to handle this challenge, several works have proposed disentangling methods that achieve robust performance in a wild environment by disentangling identity and non-identity variations (i.e., viewpoint and illumination)
- However, they need annotations of non-identity variations such as viewpoint and illumination. It is not easy to collect such pose or illumination information for all subjects in facial databases
- In this paper, we propose a learning method of disentangling identity and viewpoint representations **without any auxiliary supervision of the variations**.
- Furthermore, we disentangle not only the identity and viewpoint but also residues (e.g., illumination and color variations) that inevitably exists in a face.
- By disentangling the non-identity variations from a face, we set a new state-of-the-art face recognition mehod on CFP and Multi-PIE datasets that have large pose variations.

Visual comparison of identity, viewpoint, and residues representations.
The three rows of images are synthesized by interpolating each representationfrom source to target image.
From top to bottom, identity, viewpoint, and residues are represented.

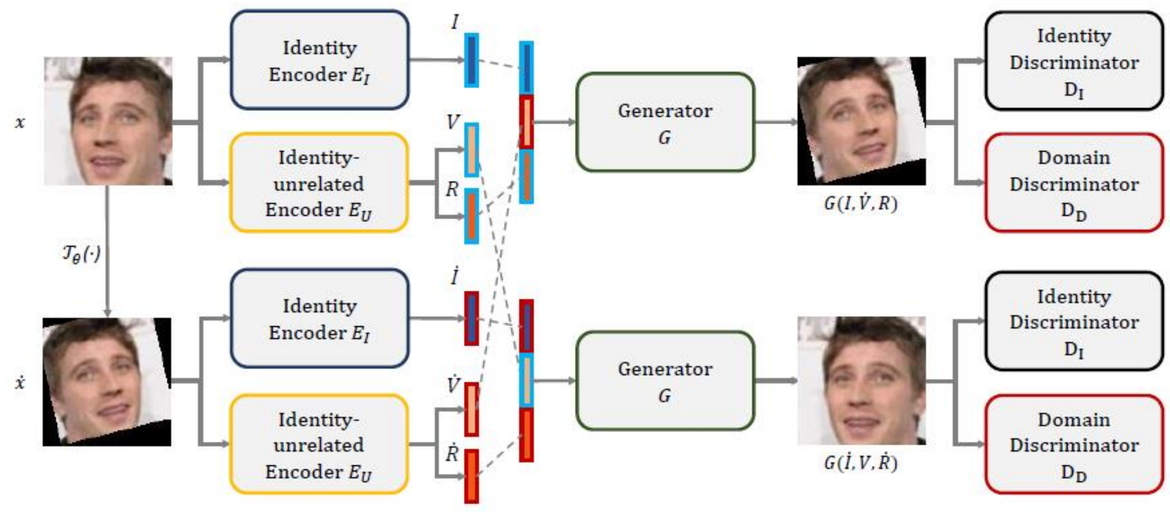
Proposed Method



- We design the learning problem as **finding a generative function** which is conditioned on the three representations (i.e., identity, viewpoint, and residues) that **wield independent effects on the output**.
- To this end, we propose two learning schemes.
  - **Viewpoint substitution**
  - **Identity substitution**
- Also we suggest **a disentangling loss function using distance covariance**.

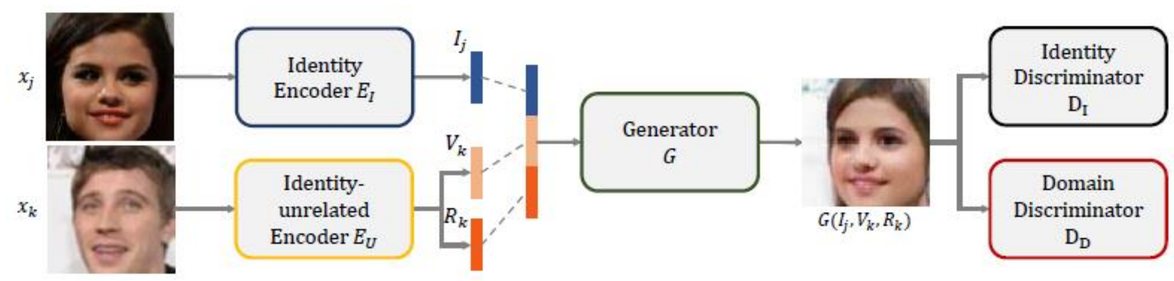
1.Learning the viewpoint representation

- we propose to use **a simple transformation** that changes the viewpoint while maintaining the identity of a face image.
- The transformation can be affine, perspective, or thin plate spline transformation.
- By using the transformation, we can access the pair of images that contain different viewpoints.



- The transformed image differs only in the viewpoint aspect from the input image.

2.Learning the Identity representation



- The generated image should contain the identity of the source identity image while other representations keep remaining.

3.Disentangling loss

- The disentangled representations should contain different information from one another. To guarantee the independency between learned representations, we use distance covariance as a disentangling loss function.
- Distance covariance is a metric that measures dependency between random vectors and becomes zero when the two random vectors are independent from each other.

Experimental Results

PERFORMANCE COMPARISON ON IJB-A.

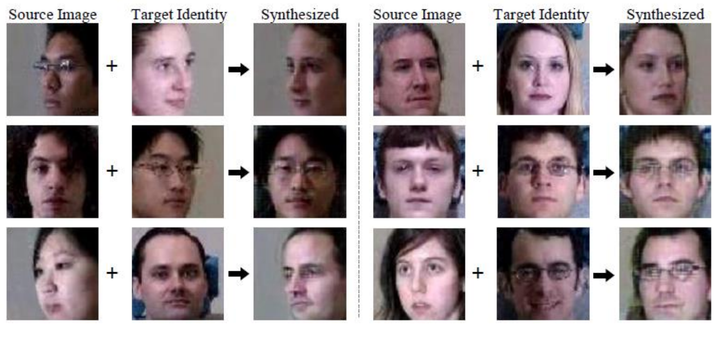
Method	Verification Accuracy	
	@0.01 FPR	@0.001 FPR
Wang et al. [37]	72.9	51.0
PAM [38]	73.3	55.2
DCNN [39]	78.7	-
DR-GAN [8]	77.4	53.9
Ours	**81.0**	**64.4**

VERIFICATION ACCURACY COMPARISON ON CFP.

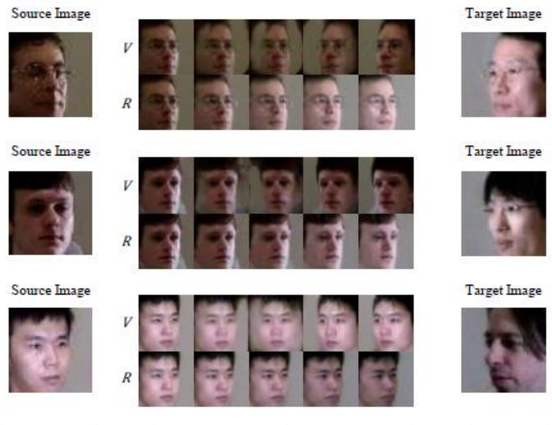
Method	Frontal-Frontal	Frontal-Profile
Sengupta et al. [35]	96.40	84.91
Sankarana et al. [36]	96.93	89.17
Chen et al. [34]	**98.67**	91.97
DR-GAN [8]	97.84	93.41
Peng et al. [2]	**98.67**	93.76
Human	96.24	94.57
Ours	98.66	**94.03**

IDENTIFICATION ACCURACY COMPARISON ON MULTI-PIE.

Method	0°	15°	30°	45°	60°	Average
Zhu et al. [40]	94.3	90.7	80.7	64.1	45.9	72.9
Zhu et al. [13]	95.7	92.8	83.7	72.9	60.1	79.3
Yim et al. [41]	**99.5**	**95.0**	88.5	79.9	61.9	83.3
DR-GAN [8]	97.0	94.0	90.1	86.2	83.2	89.2
Ours	94.9	92.9	**92.4**	**88.8**	**85.5**	**90.5**



Synthesized images by substituting the identity representation of target identity image



Synthesized images by interpolating viewpoint and residues representation from source to target image.

- In this paper, we introduced a novel framework to learn disentangled representations for robust face recognition
- In particular, we propose two learning schemes, veiwpoint substitution and identity substitution.
- Also we show that using distance covariance as a disentangling loss enforces the disentanglement.
- By disentangling identity, viewpoint, and residues representation, we set a new state-of-the-art on benchmark databases in terms of face recognition performance.