WarpGAN: Automatic Caricature Generation

CVPR2019 Oral

2019.09.24

발표자 박성현

DAVIAN
Data and Visual Analytics Lab

KOREA UNIVERSITY

1 Introduction

Caricature Generation

(a) Photo (b) WarpGAN (c) WarpGAN (d) Artist (e) Artist

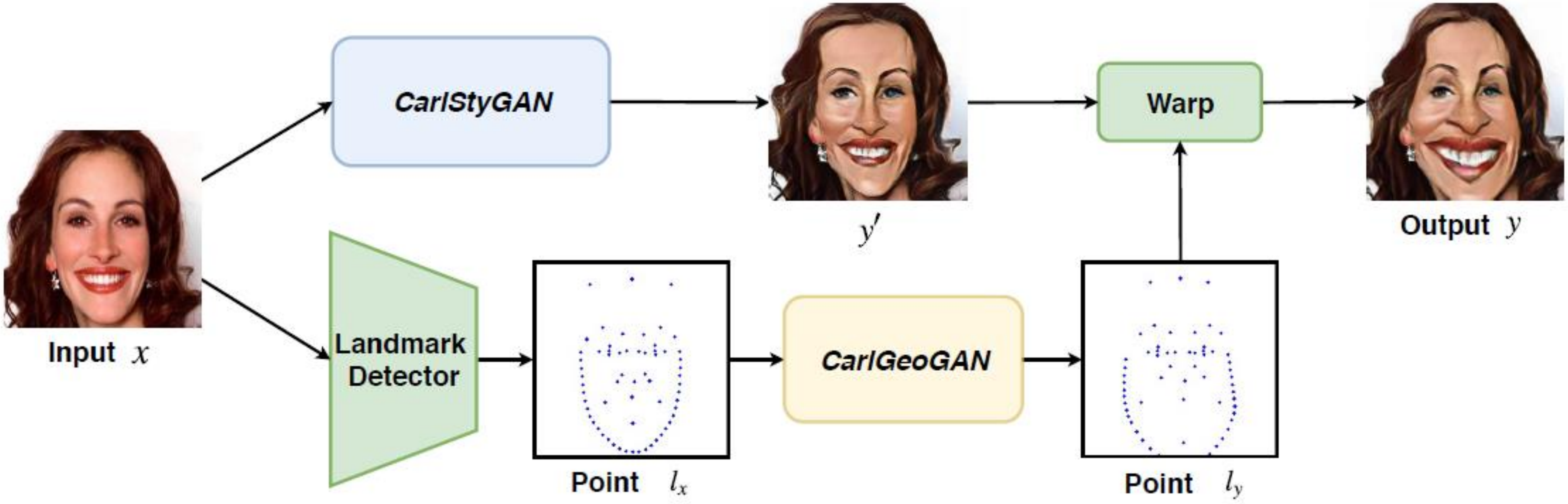
1 Introduction

Various Studies on Caricature Generation

Approach	Methodology			Examples
	Study	Exaggeration Space	Warping	
Shape Deformation	Brennan *et al.* [8] Liang *et al.* [4] CaricatureShop [9]	Drawing Line 2D Landmarks 3D Mesh	User-interactive User-interactive Automatic	 [8] [4] [9]
Texture Transfer	Zheng *et al.* [10] CariGAN [11]	Image to Image Image + Landmark Mask	None None	 [10] [11]
Texture + Shape	CariGANs [12] WarpGAN	PCA Landmarks Image to Image	Automatic Automatic	 [12] Ours

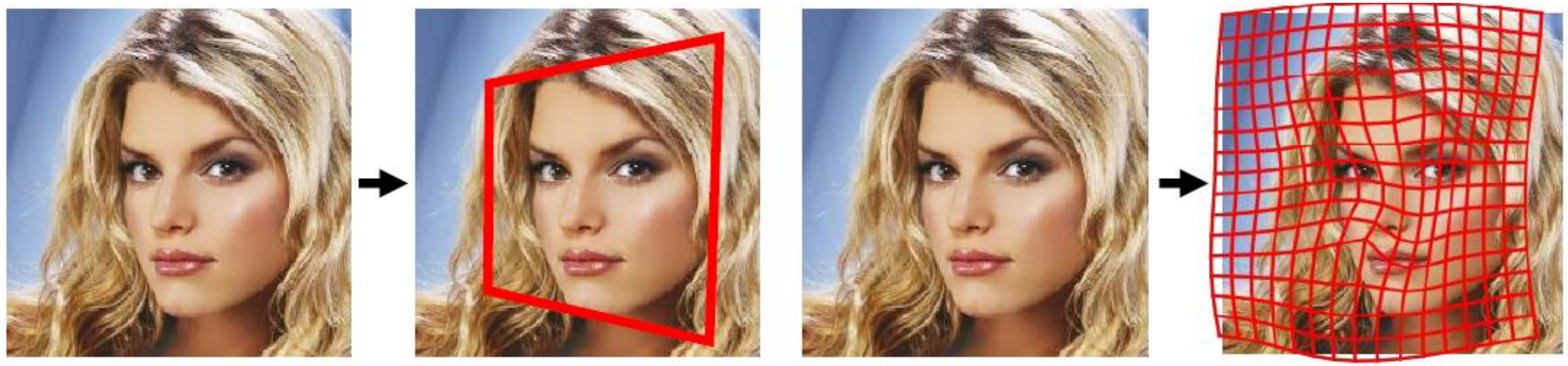
1 Introduction

CariGANs (SIGGRAPH 2018)

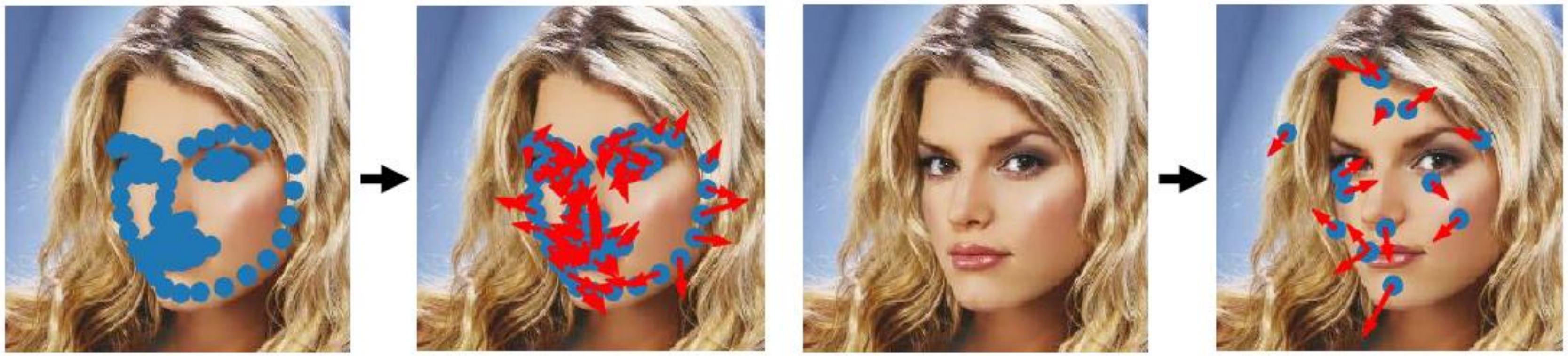


1 Introduction

Image Warping



(a) Global Parameters [14] [15] [16] (b) Dense Deformation Field [17]



(c) Landmark-based [18] (d) Control Points Estimating

2 Model

WarpGAN

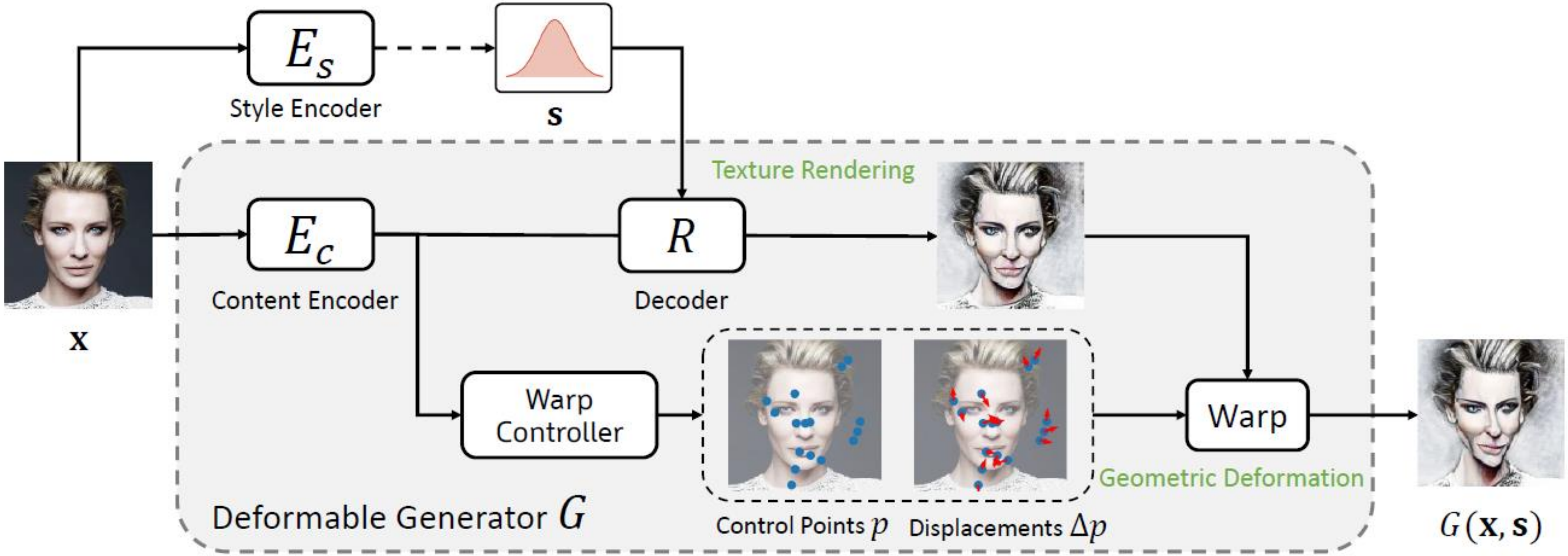


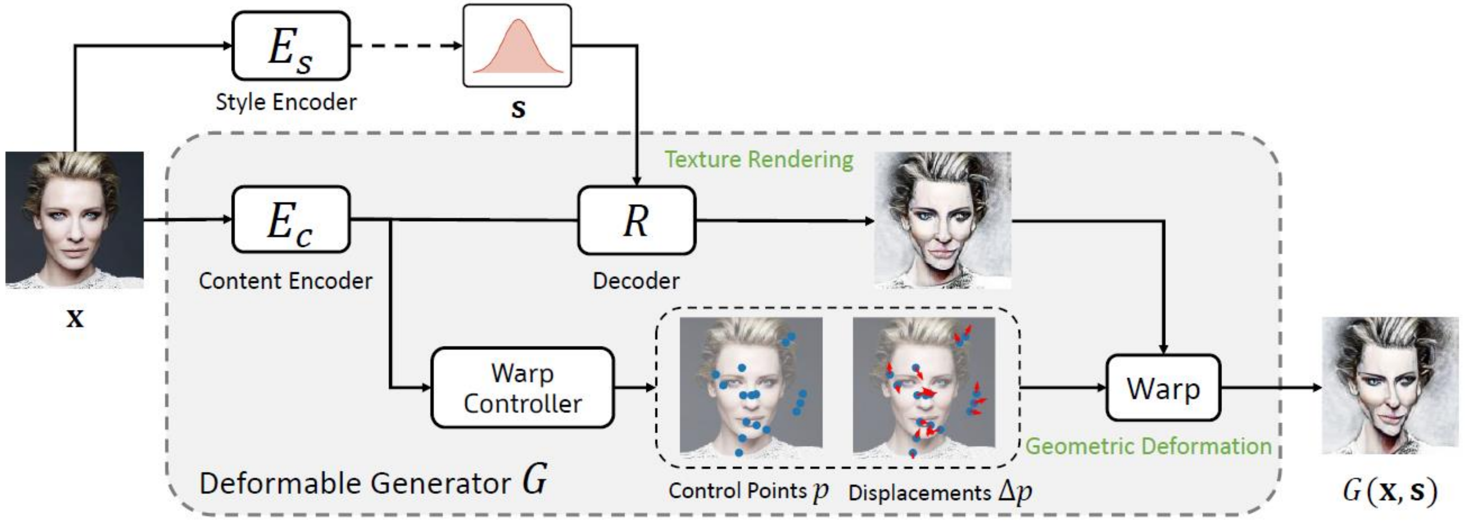
Figure 3: The generator module of WarpGAN. Given a face image, the generator outputs an image with a different texture style and a set of control points along with their displacements. A differentiable module takes the control points and warps the transferred image to generate a caricature.

2 Model

Generator - Texture Style Transfer

Name	Meaning	Name	Meaning
$\mathbf{x}_p$	real photo image	y^p	label of photo image
$\mathbf{x}_c$	real caricature image	y^c	label of caricature image
E_c	content encoder	R	decoder
E_s	style encoder	D	discriminator
p	estimated control points	Δp	displacements of p
M	number of identities	k	number of control points

Table 2: Important notations used in this paper.



$$\mathcal{L}^{p}_{idt} = \mathbb{E}_{\mathbf{x}_p \in \mathcal{X}_p}[\|R(E_c(\mathbf{x}_p), E_s(\mathbf{x}_p)) - \mathbf{x}_p\|_1]$$
$$\mathcal{L}^{c}_{idt} = \mathbb{E}_{\mathbf{x}_c \in \mathcal{X}_c}[\|R(E_c(\mathbf{x}_c), E_s(\mathbf{x}_c)) - \mathbf{x}_c\|_1]$$

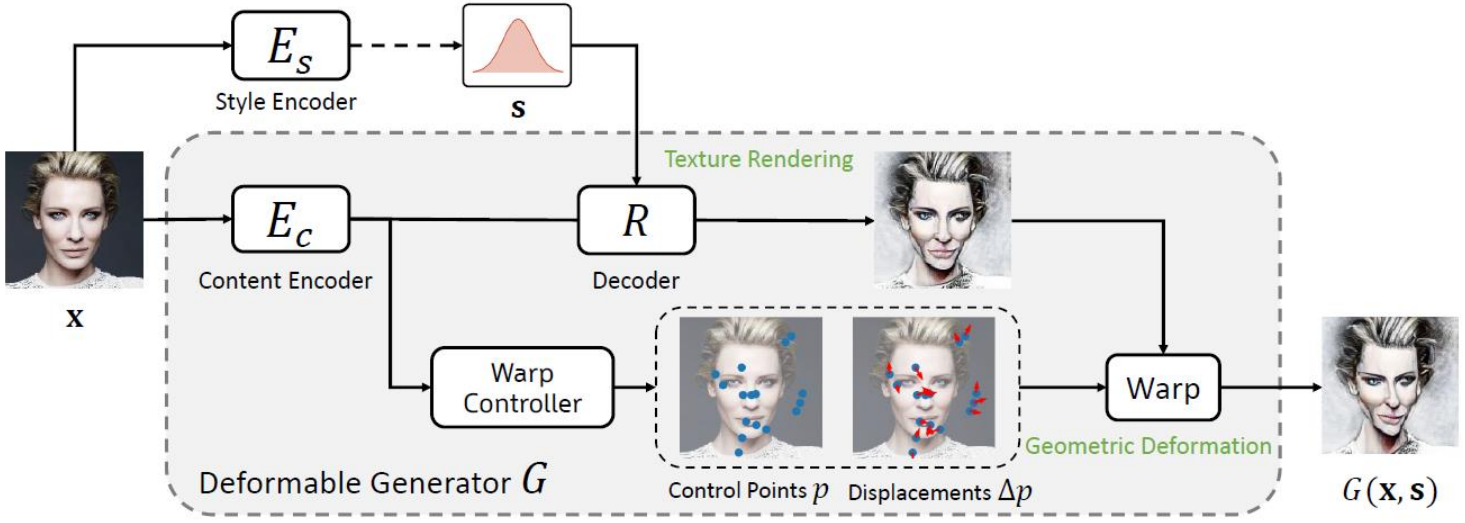
[Identity Loss]

2 Model

Generator - Automatic Image Warping

Name	Meaning	Name	Meaning
$\mathbf{x}_p$	real photo image	y^p	label of photo image
$\mathbf{x}_c$	real caricature image	y^c	label of caricature image
E_c	content encoder	R	decoder
E_s	style encoder	D	discriminator
p	estimated control points	Δp	displacements of p
M	number of identities	k	number of control points

Table 2: Important notations used in this paper.



$$p = \{\mathbf{p}_1, \mathbf{p}_2, ..., \mathbf{p}_k\}$$
$$\Delta p = \{\Delta\mathbf{p}_1, \Delta\mathbf{p}_2, ...\Delta\mathbf{p}_k\}$$
$$p' = \{\mathbf{p}'_1, \mathbf{p}'_2, ..., \mathbf{p}'_k\}$$
$$\mathbf{p}'_i = \mathbf{p}_i + \Delta\mathbf{p}_i$$

[Control points & Displacement vectors]

$$f(\mathbf{q}) = \sum_{i=1}^{k} w_i \phi(||\mathbf{q} - \mathbf{p}'_i||) + \mathbf{v}^T\mathbf{q} + \mathbf{b}$$

[TPS Transformation]

$$G(\mathbf{x}, \mathbf{s}) = \mathrm{Warp}\left(R(E_c(\mathbf{x}), \mathbf{s}), p, \Delta p\right)$$

[Generator]

2 Model

Appendix - Thin Plate Spline

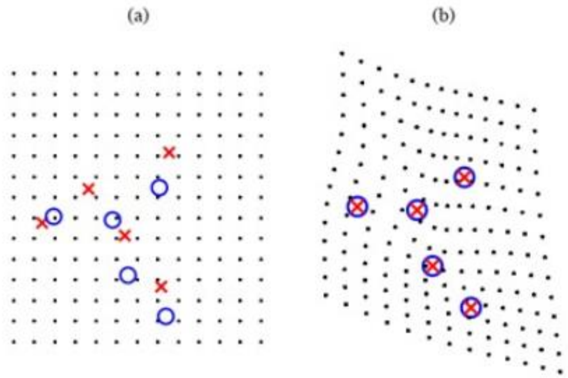
[Minimize the following function]

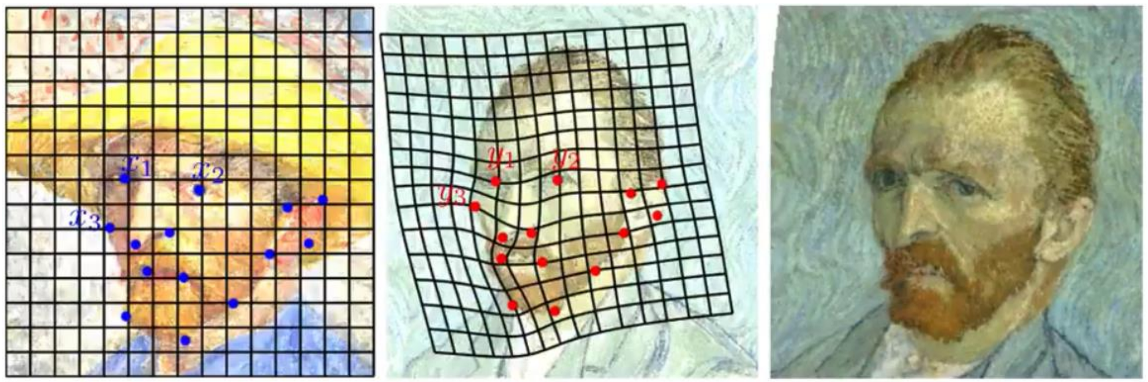
$$E_{\text{tps}}(f) = \sum_{i=1}^{K} \|y_i - f(x_i)\|^2$$

$$E_{\text{tps,smooth}}(f) = \sum_{i=1}^{K} \|y_i - f(x_i)\|^2 + \lambda \iint \left[\left(\frac{\partial^2 f}{\partial x_1^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x_1 \partial x_2} \right)^2 + \left(\frac{\partial^2 f}{\partial x_2^2} \right)^2 \right] \mathrm{d}x_1 \, \mathrm{d}x_2$$

[Radial Basis Function (RBF)]

$$f(x) = \sum_{i=1}^{K} w_i \varphi(\|x - c_i\|) \qquad \varphi(r) = r^2 \log r$$



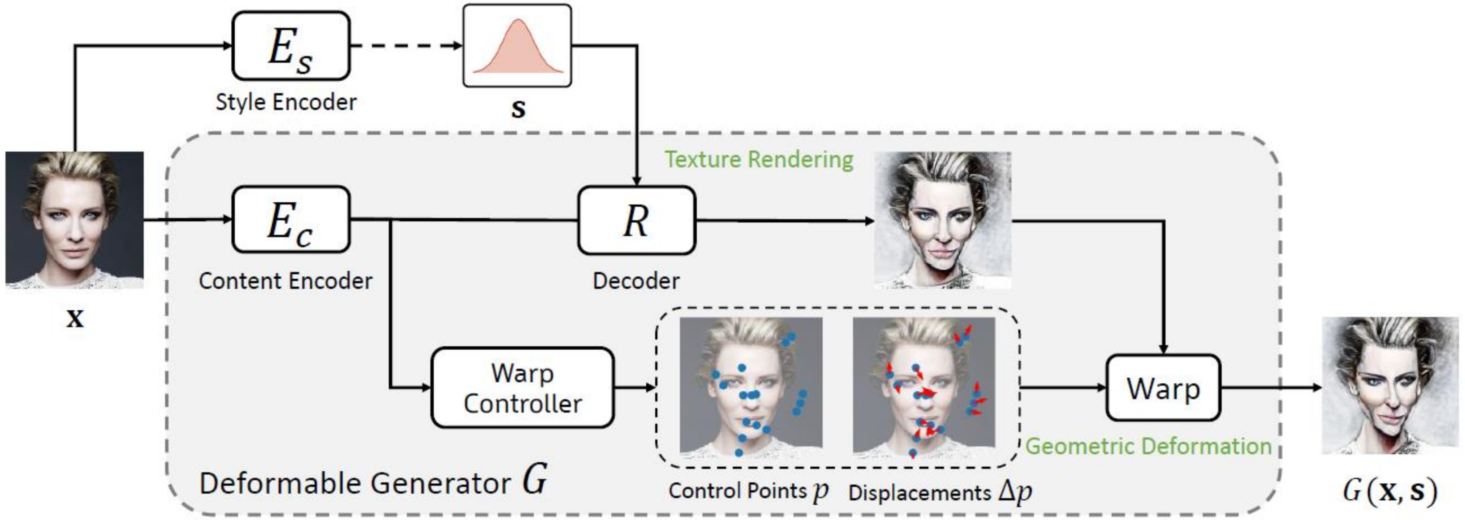


2 Model

Discriminator - Patch Adversarial Loss

Name	Meaning	Name	Meaning
$\mathbf{x}_p$	real photo image	y^p	label of photo image
$\mathbf{x}_c$	real caricature image	y^c	label of caricature image
E_c	content encoder	R	decoder
E_s	style encoder	D	discriminator
p	estimated control points	Δp	displacements of p
M	number of identities	k	number of control points

Table 2: Important notations used in this paper.



$$\mathcal{L}_p^G = -\mathbb{E}_{\mathbf{x}_p \in \mathcal{X}_p, \mathbf{s} \in \mathcal{S}}[\log D_1(G(\mathbf{x}_p, \mathbf{s}))]$$

$$\mathcal{L}_p^D = -\mathbb{E}_{\mathbf{x}_c \in \mathcal{X}_c}[\log D_1(\mathbf{x}_c)] - \mathbb{E}_{\mathbf{x}_p \in \mathcal{X}_p}[\log D_2(\mathbf{x}_p)] - \mathbb{E}_{\mathbf{x}_p \in \mathcal{X}_p, \mathbf{s} \in \mathcal{S}}[\log D_3(G(\mathbf{x}_p, \mathbf{s}))]$$

→ Patch discriminator is trained as a 3-class classifier

D_1 : Caricature / D_2 : Photos / D_3 : Generated Images

DAVIAN Data and Visual Analytics Lab

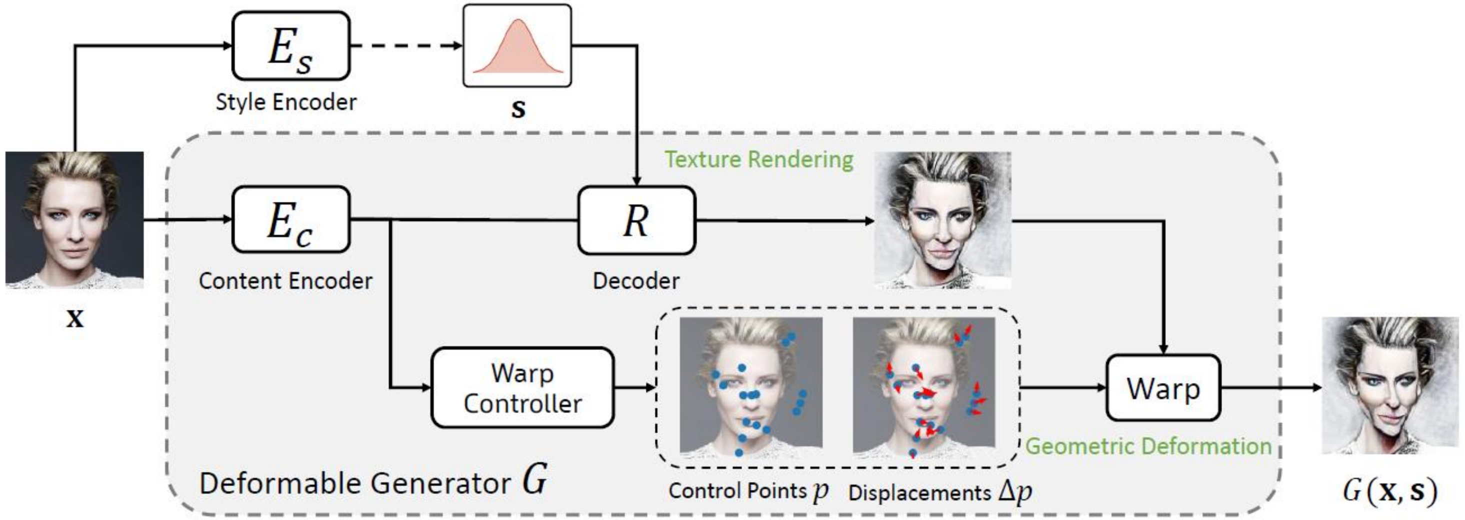
KOREA UNIVERSITY

2 Model

Discriminator - Identity-Preservation Adversarial Loss

Name	Meaning	Name	Meaning
$\mathbf{x}_p$	real photo image	y^p	label of photo image
$\mathbf{x}_c$	real caricature image	y^c	label of caricature image
E_c	content encoder	R	decoder
E_s	style encoder	D	discriminator
p	estimated control points	Δp	displacements of p
M	number of identities	k	number of control points

Table 2: Important notations used in this paper.



$$\mathcal{L}_g^G = -\mathbb{E}_{\mathbf{x}_p \in \mathcal{X}_p, \mathbf{s} \in \mathcal{S}}[\log D(y_p; G(\mathbf{x}_p, \mathbf{s}))]$$

$$\mathcal{L}_g^D = -\mathbb{E}_{\mathbf{x}_c \in \mathcal{X}_c}[\log D(y_c; \mathbf{x}_c)] - \mathbb{E}_{\mathbf{x}_p \in \mathcal{X}_p}[\log D(y_p + M; \mathbf{x}_p)] - \mathbb{E}_{\mathbf{x}_p \in \mathcal{X}_p, s \in \mathcal{S}}[\log D(y_p + 2M; G(\mathbf{x}_p, \mathbf{s}))]$$

→ Discriminator is trained as a 3M-class classifier (M is the number of identities)

2 Model

Overview of WarpGAN

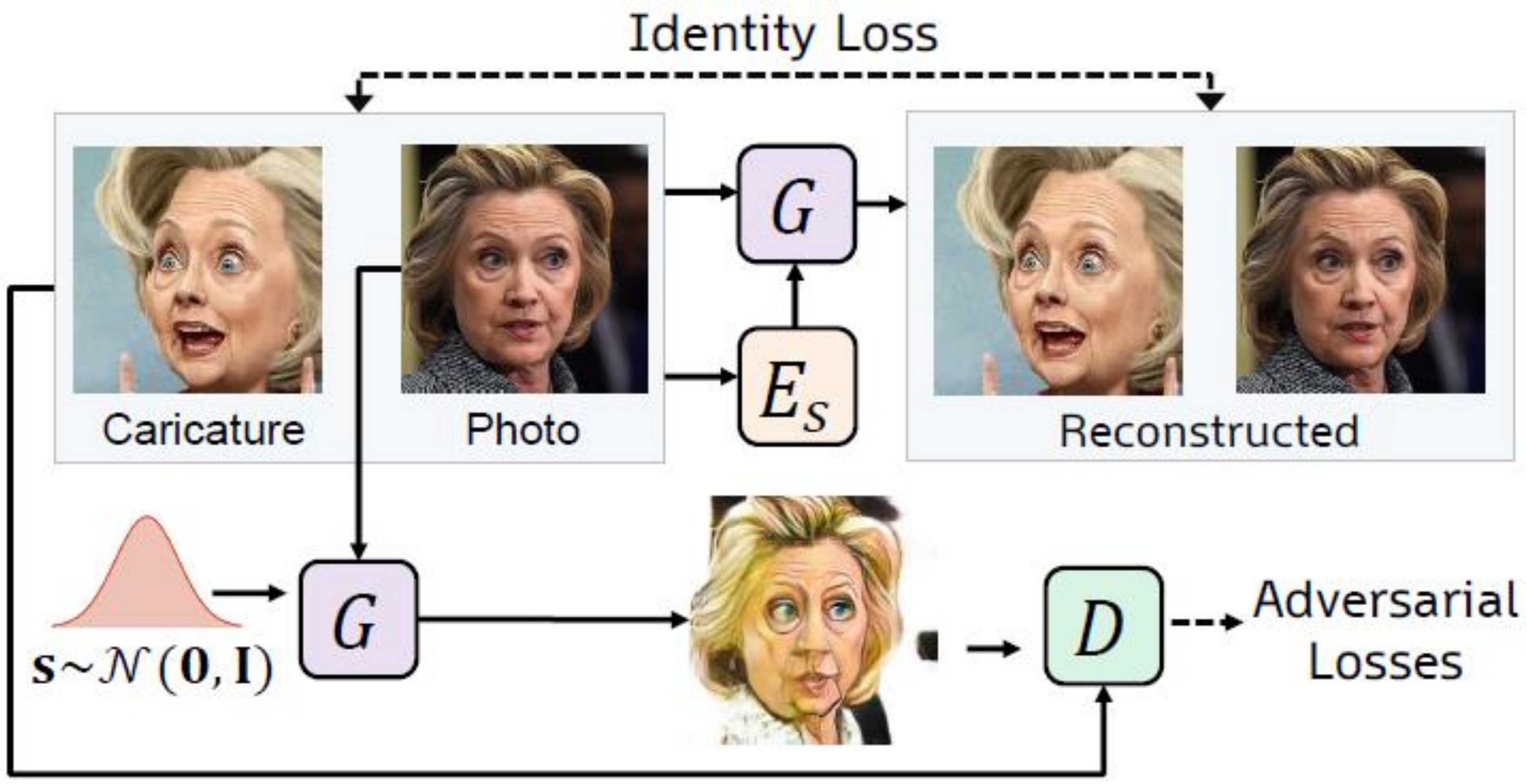


Figure 4: Overview of the proposed WarpGAN.

3

Experiments

Dataset

[Web Caricature Dataset]

3 Experiments

Comparison of Image Translation

Figure 5: Comparison of 3 different caricature styles from WarpGAN and four other state-of-the-art style transfer networks. WarpGAN is able to deform the faces unlike the baselines.

3 Experiments

Comparison of Caricature Generation

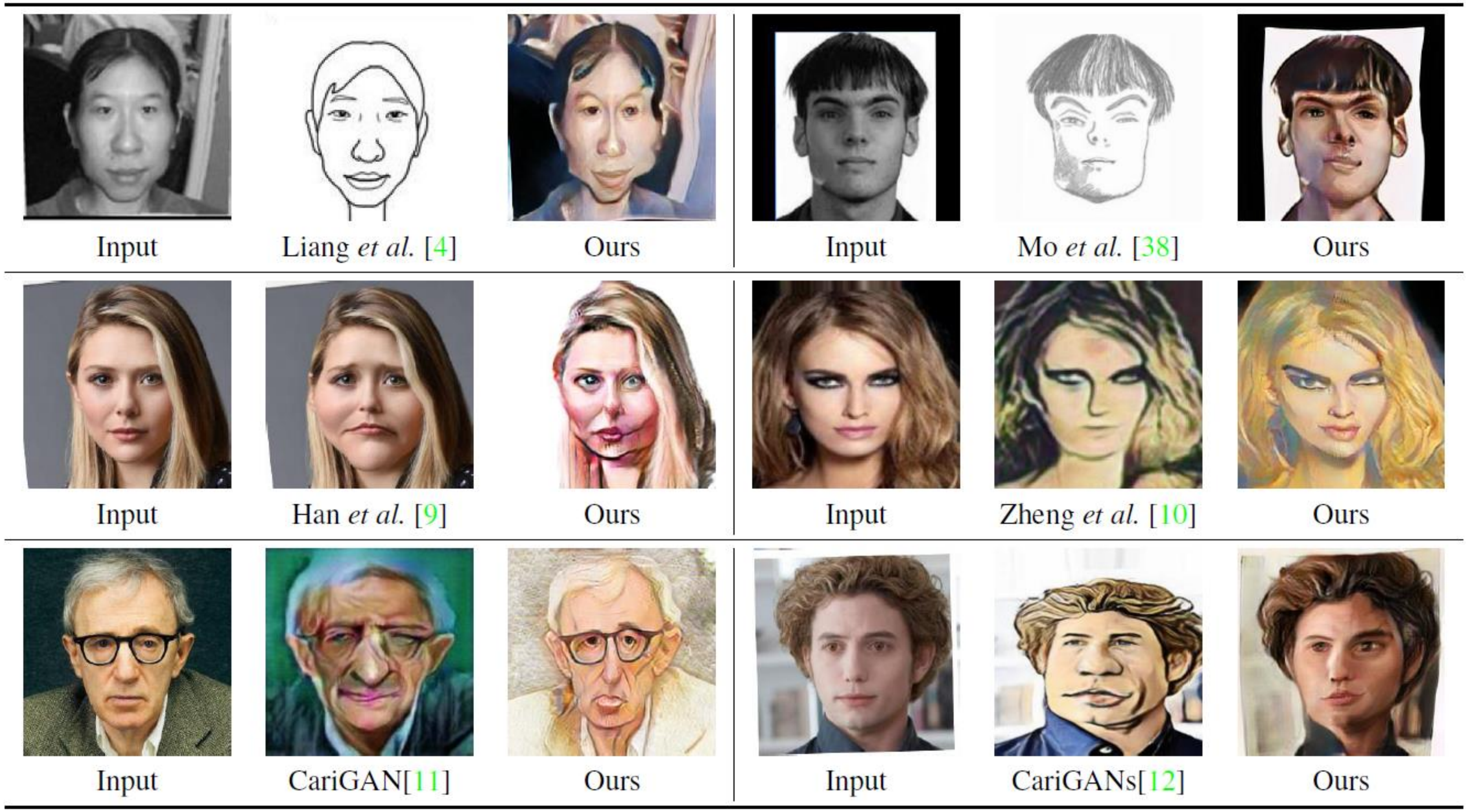


Figure 10: Comparison with previous works on caricature generation. In each cell, the left and middle images are the input and result images taken from the baseline paper, respectively. The right images are the results of WarpGAN.

3 Experiments

Ablation Study

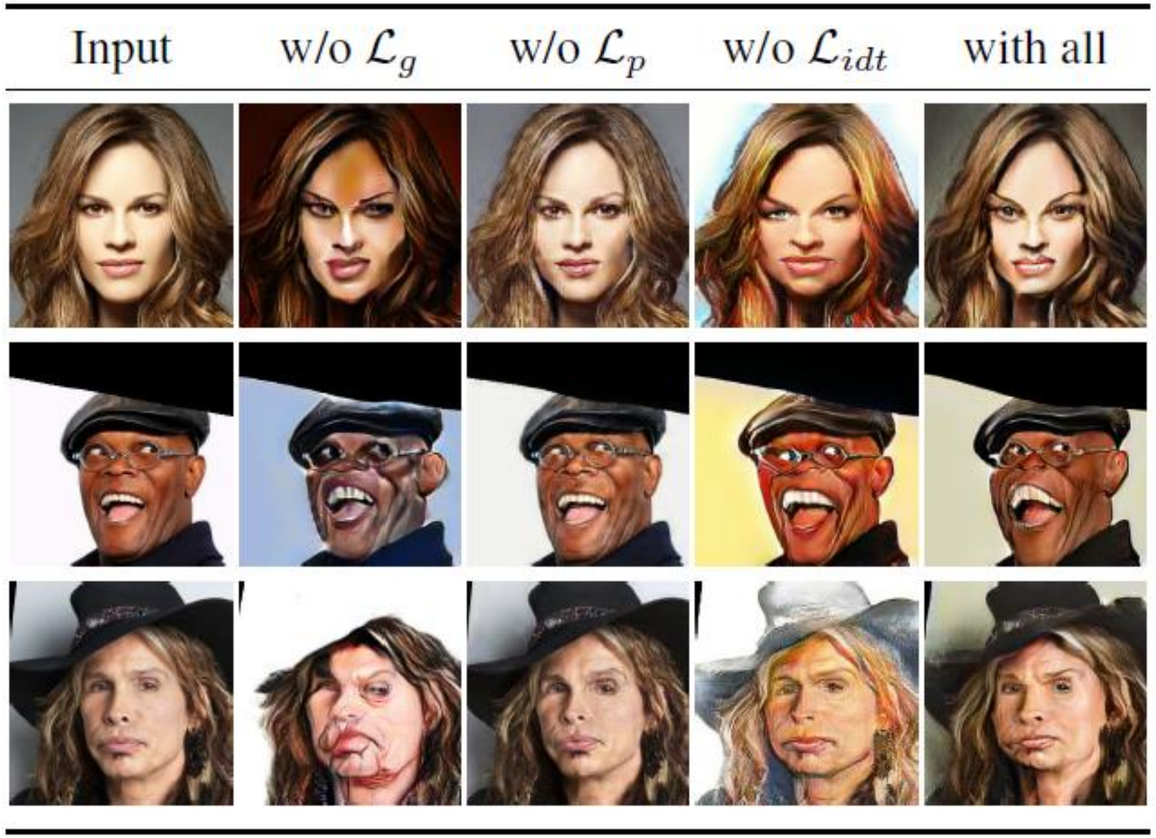


Figure 6: Different variants of the WarpGAN without certain loss functions.

3 Experiments

Shape Exaggeration Styles

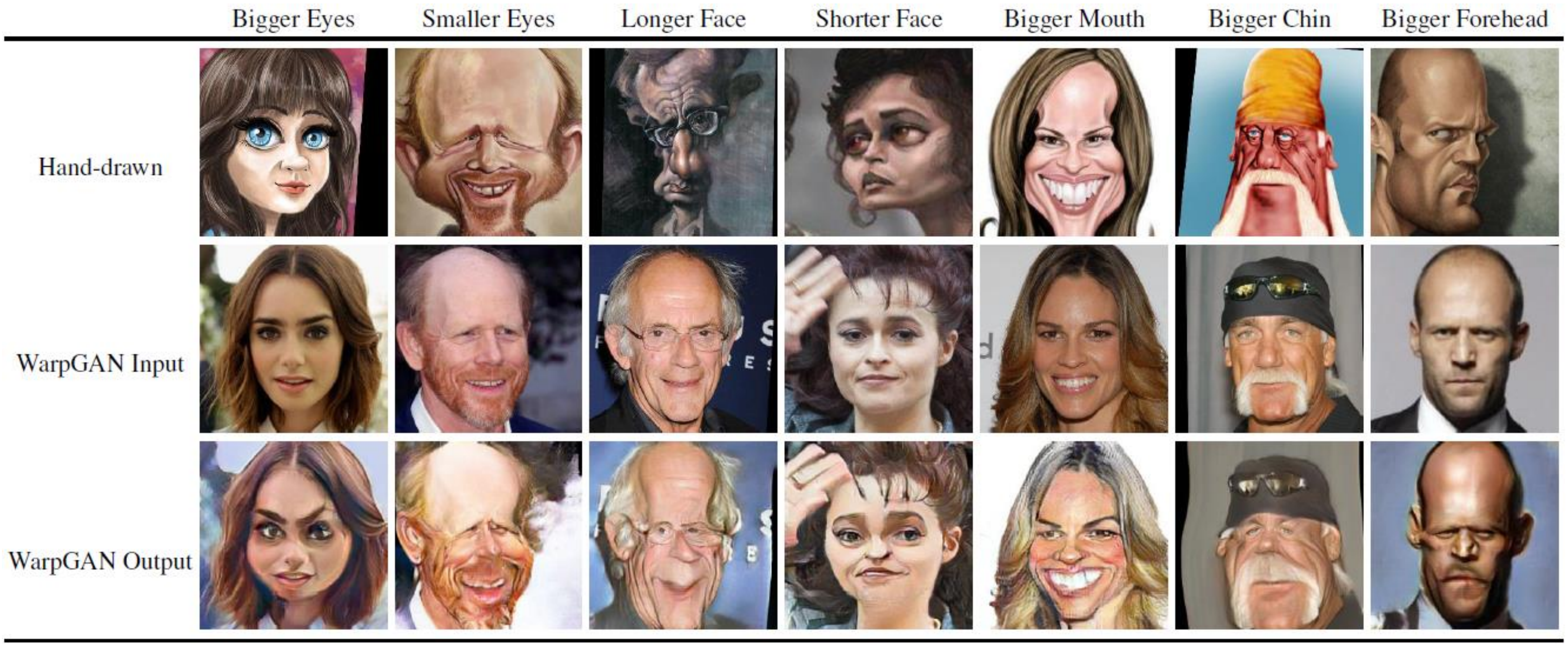


Figure 7: A few typical exaggeration styles learned by WarpGAN. First row shows hand-drawn caricatures that have certain exaggeration styles. The second and third row show the input images and the generated images of WarpGAN with the corresponding exaggeration styles. All the identities are from the testing set.

3 Experiments

Customizing the exaggeration

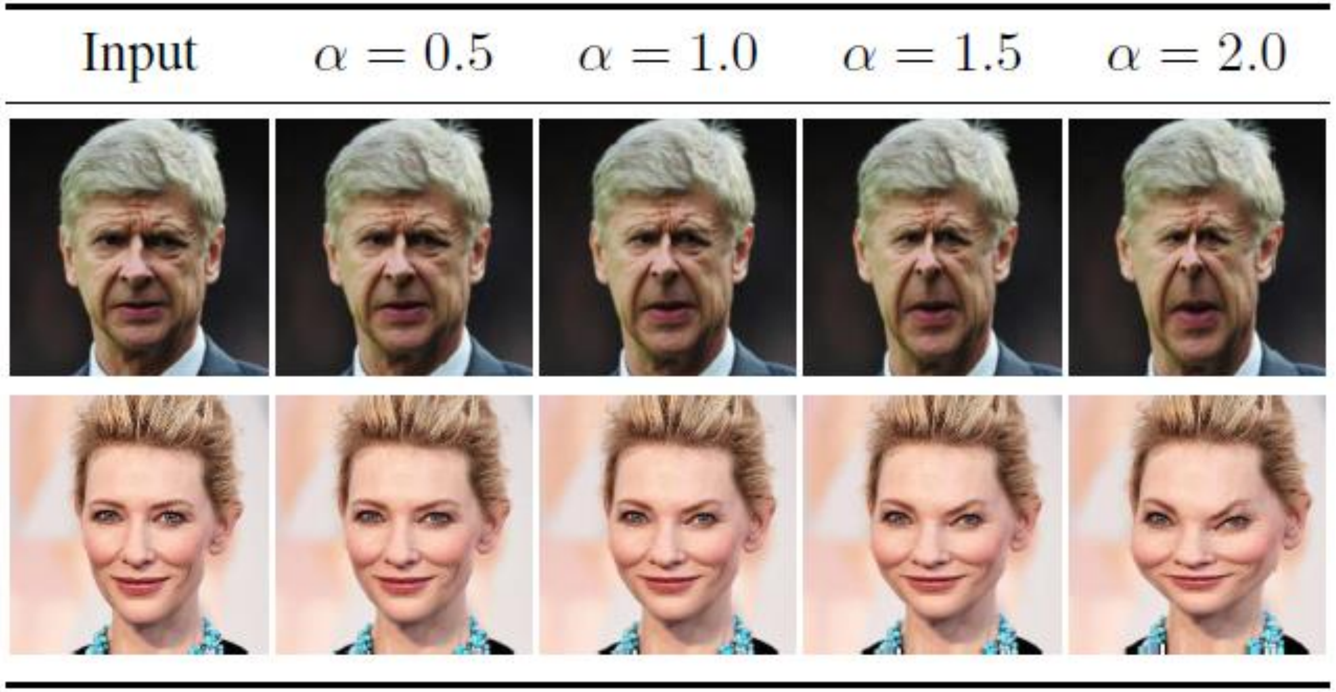


Figure 8: The result of changing the amount of exaggeration by scaling the Δp with an input parameter α.

$$\mathrm{p}'_i = \mathrm{p}_i + \Delta \mathrm{p}_i \quad \rightarrow \mathbf{p}'_\mathbf{i} = \mathbf{p}_\mathbf{i} + \boldsymbol{\alpha} * \Delta \mathbf{p}_\mathbf{i}$$

3 Experiments

Quantitative Analysis

[Face Recognition]

Method	COTS	SphereFace [35]
Photo-to-Photo	$94.81 \pm 1.22\%$	$90.78 \pm 0.64\%$
Hand-drawn-to-Photo	$41.26 \pm 1.16\%$	$45.80 \pm 1.56\%$
WarpGAN-to-Photo	$79.00 \pm 1.46\%$	$72.65 \pm 0.84\%$

Table 3: Rank-1 identification accuracy for three different matching protocols using two state-of-the-art face matchers, COTS and SphereFace [35].

[Perceptual Study]

Method	Visual Quality	Exaggeration
Hand-Drawn	7.70	7.16
CycleGAN [13]	2.43	2.27
MUNIT [7]	1.82	1.83
WarpGAN	**5.61**	**4.87**

Table 4: Average perceptual scores from 5 caricature experts for visual quality and exaggeration extent. Scores range from 1 to 10.

Figure 9: Example result images generated by the WarpGAN trained without texture/warping and with both.