Unsupervised Contrastive Photo-to-Caricature Translation based on Auto-distortion

Yuhe Ding*, Xin Ma*, Mandi Luo, Aihua Zheng, Ran He

Introduction

Task

Exaggerate human faces, and re-rendering the facial texture to form a portrait.

Challenge

1. To exaggerate deformation and obtain plausible texture simultaneously.

2. No paired data, no identity labels, and no geometric guidance.

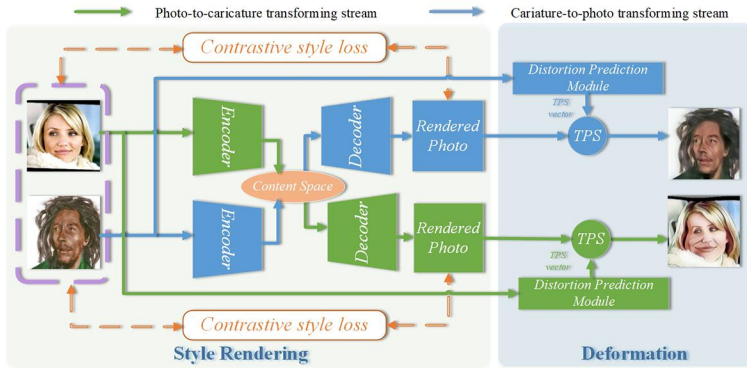
Contributions

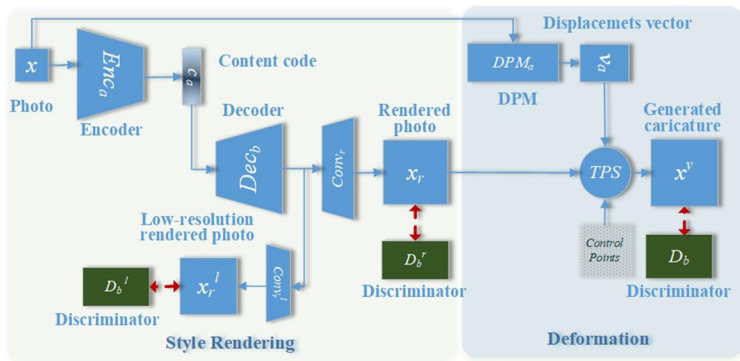
1. To reduce artifacts in rendered photos, we propose a novel contrastive loss by define a style to enforce the similarity between the rendered photo's texture and caricatures, and enforce its discrepancy to the photos, which can generate plausible textures in the rendered photos with more details.

2. To obtain exaggerating deformation in the unpaired setting, we propose a new symmetrical architecture with Distortion Prediction Modules (DPM), which predicts a set of displacements vectors without any guidance to warp the images in an unsupervised fashion.

3. Experiments on the benchmark caricature generation dataset WebCaricature compared to the state-of-the-art methods demonstrate our methods can synthesize caricatures with more hand-drawn like texture with diverse deformation without guidance in the unsupervised fashion. In addition, our method supports bidirectional translation due to the symmetrical architecture.

Architecture





Loss Functions

Contrasitive Style Loss

$$\mathcal{L}_{ctr} = \alpha_1 Ctr(x_r, x, 0) + \alpha_2 Ctr(x_r, y, 1) + \alpha_3 Ctr(y_r, y, 0) + \alpha_4 Ctr(y_r, x, 1)$$

$$Ctr(i_1, i_2, l) = \frac{1}{2}[l \cdot d(i_1, i_2)^2 + (1-l) max(mg - d(i_2, i_1), 0)^2]$$

$$d(m, n) = \frac{1}{4 * n_c * n_h * n_w} \sum_{1}^{n_c} (G_{ij}^m - G_{ij}^n)^2$$

Total Loss

Style Rendering	$\min_{Enc_a, Enc_b, Dec_a, Dec_b} = \lambda_r \mathcal{L}_{rec} + \lambda_K \mathcal{L}_{KL} + \lambda_a \mathcal{L}_{adv}^{styG} + \lambda_c \mathcal{L}_{cont} + \lambda_{ctr} \mathcal{L}_{ctr}$
	$\min_{D_a^l, D_a^r, D_b^l, D_b^r} = \lambda_a \mathcal{L}_{adv}^{styD}$
Deformation	$\min_{DPM_a, DPM_b} = \lambda_a \mathcal{L}_{adv}^{warpG} + \lambda_i \mathcal{L}_{idt}$
	$\min_{D_a, D_b} = \lambda_a \mathcal{L}_{adv}^{warpD}$

Results

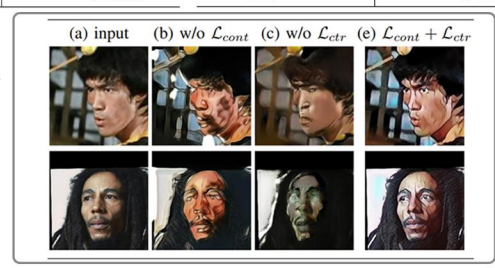
User Study

Probe	Rank-1 accuracy
Photos	100%
Hand-drawnings	8.46%
WarpGAN [7]	34.18%
Ours	34.56%

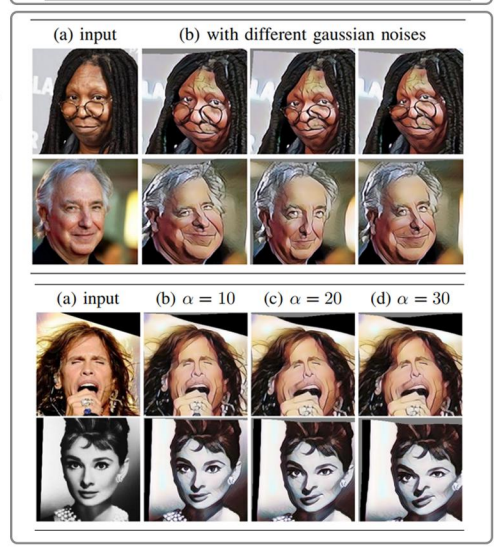
Face Recognition

Probe	Rank-1 accuracy
Photos	100%
Hand-drawnings	8.46%
WarpGAN [7]	34.18%
Ours	34.56%

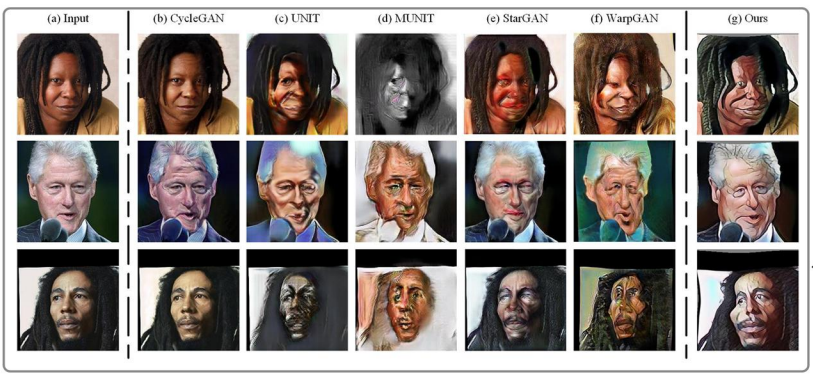
Ablation Study



Deformation Diversity



Compare with SOTA



*Equal Contribution