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

Yuhe Ding<sup>1,2</sup>, Xin Ma<sup>2,3</sup>, Mandi Luo<sup>2,3</sup>, Aihua Zheng<sup>1</sup>, Ran He<sup>2,3</sup> <sup>1</sup> School of Computer Science and Technology, Anhui University <sup>2</sup>NLPR & CEBSIT, CASIA <sup>3</sup> School of Artificial Intelligence, University of Chinese Academy of Sciences







• Photo-to-Caricature Translation



## Part Challenges & Contributions

- Challenges
  - 1. To exaggerate deformation and obtain plausible texture simultaneously.
  - 2. No paired data, no identity labels, and no geometric guidance.

Method	Component					
	Texture	Exaggeration	Diversity	Bidirectional	Unsupervised	
CycleGAN	$\checkmark$	×	×	$\checkmark$	$\checkmark$	
UNIT	$\checkmark$	×	×	$\checkmark$	$\checkmark$	
MUNIT	$\checkmark$	×	$\checkmark$	$\checkmark$	$\checkmark$	
StarGAN	$\checkmark$	×	×	$\checkmark$	$\checkmark$	
CariGAN	$\checkmark$	$\checkmark$	×	×	×	
CariGANs	$\checkmark$	$\checkmark$	×	$\checkmark$	$\checkmark$	
WarpGAN	$\checkmark$	$\checkmark$	$\checkmark$	×	×	
Ours	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	

- Contributions:
  - 1. Contrastive style loss: Enforce the similarity between the rendered photo's texture and caricatures, and enforce its discrepancy to the photos.
  - 2. Distortion Prediction Modules (DPM): predicts a set of displacements vectors without any guidance to warp the images in an unsupervised fashion.





Overview of our symmetric architecture.



The pipeline of photo-to-caricature transforming. The caricature-to-photo transforming can be designed in the same manner.



• Contrastive Style Loss

$$\begin{split} \mathcal{L}_{ctr} = &\alpha_1 Ctr(x_r, x, 0) + \alpha_2 Ctr(x_r, y, 1) + \alpha_3 Ctr(y_r, y, 0) + \alpha_4 Ctr(y_r, x, 1) \\ Ctr(i_1, i_2, l) = &\frac{1}{2} [l \cdot d(i_1, i_2)^2 + (1 - l)max(mg - d(i_2, i_1), 0)^2] \\ d(m, n) = &\frac{1}{4 * n_c * n_h * n_w} \sum_{1}^{n_c} (G_{ij}^m - G_{ij}^n)^2 \\ & *G \text{ is Gram Matrix} \end{split}$$

• Total Loss



• Qualitative Results

Deformation Diversity

**V** Compare with SOTA







• Qualitative Results

Ablation Study







• QuantitativeResults

		Face Decompition	7	
		Recognition _	Probe	Rank-1 accuracy
			Photos	100%
			Hand-drawnings	8.46%
User Study			WarpGAN [7]	34.18%
•			Ours	34.56%
Method	Proportion			
	21.20 %			

Method	Proportion		
UNIT [9]	24.39%		
WarpGAN [7]	24.56%		
Ours	51.05%		

