TailorGAN: Making User-Defined Fashion Designs

Lele Chen, Justin Tian, Guo Li, Cheng-Haw Wu, Erh-Kan King, Kuan-Ting Chen, Shao-Hang Hsieh, and **Chenliang Xu**
{lchen63@ur., jtian@pas., gli27@ur., chenliang.xu@}rochester.edu
{jason.wu, jeff.king, dannie.chen, shao}@viscovery.com

WACV 2020 March 1-5, Aspen, CO

Code, dataset, and pretrained model

Introduction

- **Motivation**: Attribute editing has become an important and emerging topic of computer vision. However, image syntheses using generative adversarial network (GAN) models usually involve highly-entangled attributes and may fail in editing target-specific attributes/objects. Also, it is impossible to build a dataset that is large enough to approximate the distributions of all attributes, such as garment texture and design. Hence, proposing a novel learning paradigm is expected to solve those challenges.
- **Objective**: We consider the task: given an arbitrary garment image, we want to change a specific shape design of a particular region with guidance. To deal with this task, we propose a novel self-learning algorithm with self-attention mask and attribute-aware losses, which overcomes the difficulty of lack of paired dataset and forces the model to learn attribute guided synthesize.

Fig.1 Our synthesized results with querying an unseen fashion design.

TailorGAN Network Structure

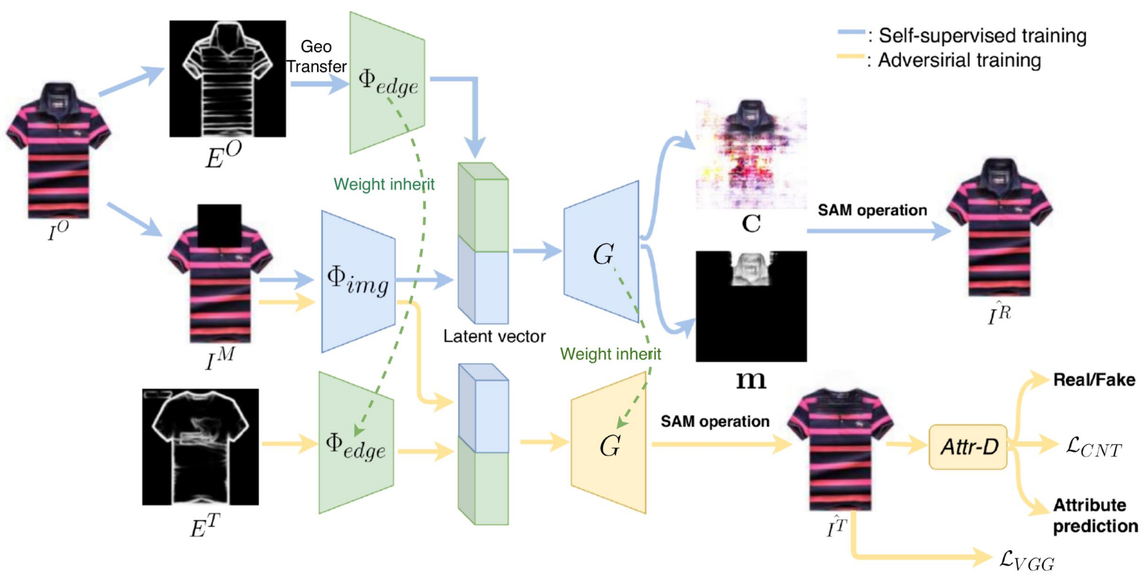

Fig.2 Overall network structure

- **Edge Encoder & Texture Encoder** extracts edge feature and texture feature from input images.
- **Decoder/Generator** uses concatenated reference edge feature and original texture feature to generate synthesized garment image.
- **Attribute-aware Discriminator** applies multiple different loss functions to better guide the generator.

Dataset

Fig.5 A new GarmentSet is proposed, which contains 12 collar types (9636 images) and 2 sleeve types (8616 images).

Algorithm & Losses

- **Self-supervised Reconstruction Training Step:** Using the edge image and the original masked image to train reconstruction with self-attention mask operation (**blue line** in the network structure).
- **Adversarial Training Step:** Design an attribute-aware discriminator and train with reference edge guidance and original masked image based on parameters inherited from the reconstruction step (**yellow line** in the network structure).
- **Objective function:**

$$\mathcal{L}_{\mathrm{G}} = (1 - \mathrm{D}(\mathrm{G}(E^T, I^M))^2 + \mathcal{L}_{CNT}(\hat{I}^T, I^T) + \lambda_1 * \mathcal{L}_{ATT}(\hat{I}^T, \vec{V}^T) + \lambda_2 * \mathcal{L}_{VGG}(\hat{I}^T, I^O)$$

- **Additional perceptual-level loss** measures the feature output of the convolution layer of the discriminator.

$$\mathcal{L}_{CNT}(\hat{I}^T, I^T) = ||D_{conv}(I^T) - D_{conv}(\hat{I}^T)||_1$$

- **Attribute-aware loss** forces the model to synthesize images with a correct attribute using cross-entropy loss on the attribute type.

$$\mathcal{L}_{ATT}(I, \vec{V}) = -\vec{V} * log(f(I)) - (1 - \vec{V}) * log(1 - f(I))$$

Qualitative Results

Fig.3 Collar synthesize results

Fig.4 Sleeve synthesize results

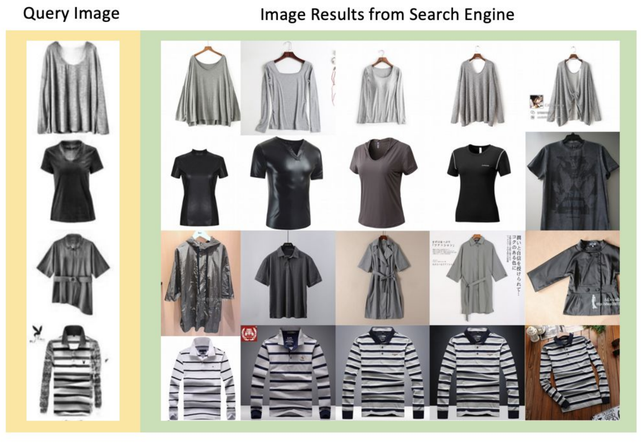

Fig.6 Searching results using synthesized images.

Acknowledgement: This work was partly supported by a research gift from Viscovery.