

# Fashion-AttGAN: Attribute-Aware Fashion Editing with Multi-Objective GAN

Qing Ping, Bing Wu, Wanying Ding, Jiangbo Yuan  
Vipshop US Inc.  
San Jose, CA 95134, USA

{pingqing508,bingwuthu,dingwanying0820,yuanjiangbo}@gmail.com

## Abstract

*In this paper, we introduce attribute-aware fashion-editing, a novel task, to the fashion domain. We re-define the overall objectives in AttGAN [5] and propose the Fashion-AttGAN model for this new task. A dataset is constructed for this task with 14,221 and 22 attributes, which has been made publically available. Experimental results show effectiveness of our Fashion-AttGAN on fashion editing over the original AttGAN.*

## 1. Introduction

As we’re witnessing great booming of online fashion shopping, deep learning models such as GAN models[3] have been widely applied to fashion industry, such as virtual try-on[4], domain adaptation[1], text2image [8] and so on. In this paper, we introduce a novel task into the fashion domain, namely attribute-aware fashion-editing, which edits certain attribute(s) of the image of a fashion item and preserve other details as intact as possible. This task opens a new door of possibilities for user-driven fashion design, and is potentially beneficial to virtual try-on, outfit recommendation, visual search and so on. Although previous work on human face editing is relevant to our proposed task [7, 5, 2, 6], they are not directly applicable to fashion-editing, since now the scope of editing is not confined to a small area in human faces, but to much larger area of a fashion item, such as an entire sleeve or collar. To bridge these gaps, we present preliminary results of Fashion-AttGAN, a variant of face attribute-editing AttGAN. We define different overall objectives, and break the overly-strict constraint from the auto-encoder on the generator so that it can generate more “wild” samples with dramatic changes.

The contributions of our work are three folds: 1) We introduce the task of fashion-editing into the fashion domain. To our best knowledge, this is the first time such task is established and explored in this area. 2) We adapt the original AttGAN to our task, and improve the model by re-defining the overall objectives. 3) We build a large-scale fashion

dataset of 14,221 images and 22 attributes, which has been made publically available on *Github*.

## 2. Method

Our Fashion-AttGAN model is a variant of previous facial attribute-editing model AttGAN[5]. Both models include an encoder network  $E$ , a generator network  $G$ , a classification network  $C$  and discriminator network  $D$ , where  $C$  and  $D$  share parameters except the last layer. For details of the structure of the model, please refer to AttGAN [5]. The main differences between our model and AttGAN are the overall objectives: we define three objectives of different optimization scopes:

$$\min_{\theta_D, \theta_C} \mathcal{L}_{D,C} = \mathcal{L}_{adv_d} + \lambda_2 \mathcal{L}_C(x_{\hat{a}}) \quad (1)$$

$$\min_{\theta_E, \theta_G} \mathcal{L}_{E,G} = \mathcal{L}_{adv_g} + \lambda_1 \mathcal{L}_{rec} \quad (2)$$

$$\min_{\theta_G} \mathcal{L}_G = \mathcal{L}_C(x_{\hat{b}}) \quad (3)$$

In AttGAN, the loss objectives in equation 2 and 3 are of the same scope for both encoder and decoder networks. The training details of Fashion-AttGAN is summarized in Algorithm 1.

The intuition behind our objective functions is that by back-propagating the classification error  $\mathcal{L}_C(x_{\hat{b}})$  of attribute-edited generated sample  $x_{\hat{b}}$  to as back as the generator network, the model is empowered with more flexibility to generate “wild” samples related to different attribute edits, in contrast with propagating the errors back to encoders, which greatly limited the power of the generator network to generate novel samples, due to the auto-encoder path in AttGAN. This may not be a problem in face attribute-editing, since the editing is focused on small areas of human faces. However, when it comes to fashion-editing, original AttGAN does not allow generator sufficient flexibility to edit much larger areas, such as the length of the sleeves.



(a) Attribute-Editing Results of AttGAN



(b) Attribute-Editing Results of Fashion-AttGAN

Figure 1: Clothing Attribute-Editing Results. From left to right columns:(1) original image, (2) reconstructed image, (3-6) varied sleeve lengths, (7-24) varied colors.

### 3. Experiments & Results

#### 3.1. Dataset

In this study, we constructed a dataset based on the VITON dataset[4, 9] which is publically available. Each entry in the dataset consists of an image from VITON, and a vector of attributes, such as “no-sleeve”, “short-sleeve”, “red”, “blue” and so on. The attribute vector is predicted with our in-house classification model. The dataset includes 14,221 images, and 22 attribute values. In the future, we plan to publish a larger dataset with more images and more attributes.

#### 3.2. Experimental Results

The comparative results of fashion-editing between our Fashion-AttGAN and AttGAN are depicted in Fig.1. From the figures we can observe that: (1) Color edits: the original AttGAN can edit colors of clothes with lighter shades, but not so well for darker shades (row:(1,4,5), column:(7-22)). Whereas our Fashion-AttGAN can modify clothes of almost any shade to other colors as shown in Fig.1(b); (2) Sleeve length: original AttGAN does not present any sleeve-length changes even with careful parameter tuning (Fig.1(a),column:(3,4,5,6)).On the contrary, our

Fashion-AttGAN can generate vivid samples of different sleeve lengths and preserve the patterns in original samples (shapes,colors, logos, textures) as much as possible (Fig.1(b)).

---

#### Algorithm 1 The training pipeline of Fashion-AttGAN

---

Require:  $\theta_E, \theta_G, \theta_D, \theta_C$  are initial image encoder network, generator network, discriminator network and attribute classification network parameters.  $\ell(\cdot)$  is a binary cross-entropy loss for an attribute.

- 1: **while**  $\theta$  has not converged **do**
  - 2:   Sample  $x_a \sim p_x$  a batch from real image data;
  - 3:    $z \leftarrow E(x_a)$
  - 4:    $x_{\hat{a}} \leftarrow G(z, a)$
  - 5:    $x_{\hat{b}} \leftarrow G(z, b)$
  - 6:    $\mathcal{L}_C(x_{\hat{a}}) \sim \mathbb{E}_{x^a \sim p_{data}} [\sum_{i=1}^n \ell(a_i, C(x_{\hat{a}}))]$
  - 7:    $\mathcal{L}_C(x_{\hat{b}}) \sim \mathbb{E}_{x^a \sim p_{data}, b \sim p_{attr}} [\sum_{i=1}^n \ell(b_i, C(x_{\hat{b}}))]$
  - 8:    $\mathcal{L}_{adv_d} \sim -\frac{1}{m} \sum_{i=1}^m [\log D(x_a) + \log(1 - D(x_{\hat{b}}))]$
  - 9:    $\mathcal{L}_{adv_g} \sim \frac{1}{m} \sum_{i=1}^m [\log(1 - D(x_{\hat{b}}))]$
  - 10:    $\mathcal{L}_{rec} \sim \mathbb{E}_{x^a \sim p_{data}} \|\Phi(x_a) - \Phi(x_{\hat{a}})\|_2^2$
  - 11:    $\theta_D, \theta_C \leftarrow -\nabla_{\theta_D, \theta_C} (\mathcal{L}_{adv_d} + \mathcal{L}_C(x_{\hat{a}}))$
  - 12:    $\theta_E, \theta_G \leftarrow -\nabla_{\theta_E, \theta_G} (\mathcal{L}_{adv_g} + \mathcal{L}_{rec})$
  - 13:    $\theta_G \leftarrow -\nabla_{\theta_G} (\mathcal{L}_C(x_{\hat{b}}))$
-

## References

- [1] A. Almahairi, S. Rajeswar, A. Sordoni, P. Bachman, and A. Courville. Augmented cyclegan: Learning many-to-many mappings from unpaired data. *arXiv preprint arXiv:1802.10151*, 2018.
- [2] J. Bao, D. Chen, F. Wen, H. Li, and G. Hua. Cvae-gan: fine-grained image generation through asymmetric training. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2745–2754, 2017.
- [3] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [4] X. Han, Z. Wu, Z. Wu, R. Yu, and L. S. Davis. Viton: An image-based virtual try-on network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 7543–7552, 2018.
- [5] Z. He, W. Zuo, M. Kan, S. Shan, and X. Chen. Attgan: Facial attribute editing by only changing what you want. *arXiv preprint arXiv:1711.10678*, 2017.
- [6] Y. Lu, Y.-W. Tai, and C.-K. Tang. Attribute-guided face generation using conditional cyclegan. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 282–297, 2018.
- [7] G. Perarnau, J. Van De Weijer, B. Raducanu, and J. M. Álvarez. Invertible conditional gans for image editing. *arXiv preprint arXiv:1611.06355*, 2016.
- [8] N. Rostamzadeh, S. Hosseini, T. Boquet, W. Stokowiec, Y. Zhang, C. Jauvin, and C. Pal. Fashion-gen: The generative fashion dataset and challenge. *arXiv preprint arXiv:1806.08317*, 2018.
- [9] B. Wang, H. Zheng, X. Liang, Y. Chen, L. Lin, and M. Yang. Toward characteristic-preserving image-based virtual try-on network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 589–604, 2018.