

Say Yes To The Dress: Shape and Style Transfer Using Conditional GANs

Michael A. Hobley^[0000-0003-2294-5884] and Victor A. Prisacariu^[0000-0002-0630-6129]

University of Oxford
{mahobley,victor}@robots.ox.ac.uk

Abstract. Objects are defined by their shape and visual style. Previous work into image manipulation has generally altered the stylistic appearance of a whole image, while maintaining the image content and object shapes. In this paper we transfer both the shape and style of chosen objects between images, leaving the remaining areas unaltered. To tackle this problem, we propose a two stage method, where each stage contains a generative adversarial network, that will alter the shape and style of objects in a subject image to reflect a donor image. We demonstrate the effectiveness of our method by transferring clothing between images.

Keywords: Generative Adversarial Network · Conditional GAN · Style Transfer · Shape Transfer · Fashion · Clothing · Machine Learning

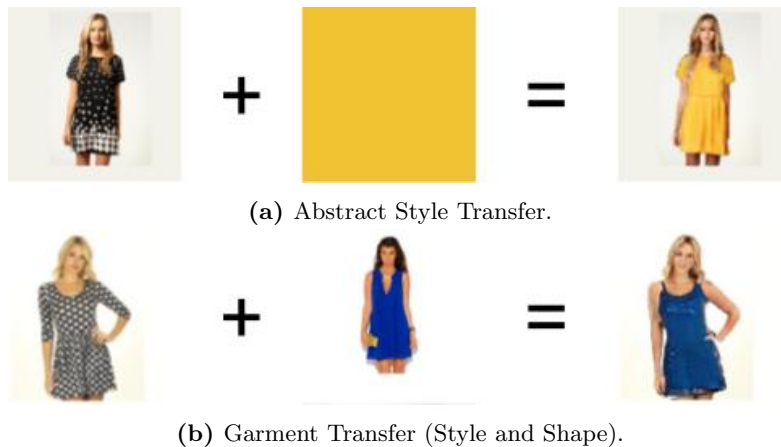


Fig. 1: Generated examples using the proposed system for garment style transfer and full garment transfer. The subject (*left*) and a donor (*middle*) are used to create an image with chosen features of the donor image while keeping the subject the same. For the full effect view in colour.

1 Introduction

There are countless applications where altering the appearance of specific objects in an image is needed and, due to the visual nature of image manipulation, example images are often more useful than textual descriptions or tags. An object is defined by both its shape and visual style, which includes its colour, texture and pattern. Therefore to produce compelling object alterations from image examples, we must reproduce the shape and style from the example in the subject image. Several researchers have addressed simple image translation and style transfer tasks [4,6,19,21]. The alterations are generally applied to the whole image [4,6,19], or a simply masked area [21]. All of these methods have focused on the style of an image, maintaining the shape and location of component objects. However, to realistically transfer the appearance of an object between images, both the shape and style must be transferred.

An ideal example of a selective object transfer task is swapping clothing between photos, as clothing is characterised by its distinctive shapes and styles. To convincingly swap garments between images, one must generate realistic clothes that reflect the donor garments, while still maintaining the pose, body shape and appearance of the subject. Garment shape transfer has previously been addressed using three dimensional methods, requiring anthropomorphic data [11,27]. We make no assumptions about the three dimensional geometry, instead we only use a single subject image to capture the appearance and pose of the person, and a single donor image to define the style and type of the generated garment.

To transfer the visual appearance of garments in two dimensions using single images, we propose a two-stage method. Each stage is comprised of an encoder and a conditional generative adversarial network [10]. Our first stage alters the shape of the chosen garments in a subject image to match a donor image. Our second stage selectively transfers the style of chosen garments of a donor image to the appropriate areas in the updated subject image.

The main contributions of this paper are: (i) we transfer style between certain selected objects in images, agnostic to their original appearance, as discussed in subsection 3.3; (ii) we manipulation the shape of objects based on their shape in a second image while maintaining their pose, as discussed in subsection 3.4; (iii) we create a complete method for compelling object transfer between images, demonstrated in section 4.

2 Related work

We discuss previous work related to the broad area of image generation, the use of images as an input to generative models, the previous methods that have been applied to garment related generation and how our approach differs from previous work.

2.1 Image Generation.

Image generation is a classic problem in computer vision, with the goal of creating realistic, high resolution images that match a set of user constraints.

The most popular current approach to image generation is the Generative Adversarial Network (GAN) [9]. GANs improve upon traditional convolution neural networks by learning a representation of a dataset rather than using a simple loss function. Promising research has been conducted into generating images in stages by using multiple GANs [2,6,20,28]. These approaches first generate the coarse structure of an object in the form of surface normals [20] or a low resolution image [28], then use this as the input to a GAN that generates a final, detailed image. These coarse-to-fine schemes are generally more stable during training, allowing for more detailed, complex, and higher resolution results [8].

2.2 Image Conditioning.

Conditional networks use a conditioning variable as the network’s input array rather than a random set of values. Once trained, the user can control the generated result by manipulating this variable. The generated result should be reflective of the conditioning variable as well as representative of the training dataset.

Early conditioning methods used relatively simple labels [15,16,24], or basic textual descriptions [17,28], but more recently entire images have been used as the conditioning variable in image generation [4,6,12,18,30,29]. Image conditioning has mainly been used for simple photo editing [18], full image style transfer [4], supervised domain transfer [6,30], and unsupervised domain transfer [12,29].

Domain transfer or image translation aims to alter the style of an image while maintaining the shape or location of individual objects. Basic image translation tasks include colourising greyscale images and converting a semantic map to a textured image. Traditionally, image translations alter the domain or style of an entire image indiscriminately, however work has recently been conducted into guiding the transfer using semantic channels as masks [14]. Image translation suffers from the same instability as GANs in unconditioned fields.

Wang et al. [19] combined methods of multi-stage GAN stabilisation [8,28] with work on image translation to create a high resolution domain invariant architecture. They also developed a method for varying the style of single instances in an image, however their results were simple and had minimal user control, limited to choosing from a discrete number of the most common styles found in the dataset.

2.3 Fashion.

Clothing related computer vision has generally been focused on the discriminative tasks of classification and recognition [1,3,26] to generate textual descriptions, labels, or tags from an image.

Generative tasks related to clothing have traditionally involved generating a three dimensional representation of the subject and their clothing using body

measurements [11,27]. The three dimensional clothing can then be mapped between subjects. Processing and generating three dimensional models is computationally intensive, and gathering the data for each instance is time consuming and expensive. Yang et al. [25] generated three dimensional models from a single photograph of a subject. However, this method still required the use of anthropometric data and accurate three dimensional human body models.

Purely two dimensional fashion manipulation was recently addressed by Zhu et al. [31], who created a framework to generate clothing for a subject, based on a textual description. Zhu et al. [31] generated garment shape to match the pose and body shape from a subject image, then applied the desired style. This framework was able to implement generation of simple changes to garment shape and could generate images with plain clothes. No complex or abstract pattern results were shown. Xian et al. [23] recently developed TextureGAN, a method for generating images from a sketch and a simple texture samples. The results from TextureGAN are simple, repeating the same pattern, and are generally independent of any image context or variation that may come with shape.

In this paper, in contrast to previous image translation work [6,12,29,30], we recognise the importance of shape, as well as style, to the visual appearance of an object. We aim to capture the shape and style of chosen garments from a complex image, and transfer this representation onto a second image, leaving the remaining areas untouched. We divide the generation into two sections, shape and style. Our method treats the style and shape of each garment between images independently, allowing for complete flexibility at generation time. In contrast to previous work into texture generation, for example TextureGAN [23], our style generation alters texture based on context and shape, creating accurate features like creases in the correct garments and places. We deconstruct the object shapes into the garment shapes and the wearers pose, we can then alter the garment shapes while maintaining the appearance of the subject. We enable transfer of styles between multiple sections independently by combining feature-wise style encodings of the subject and donor images.

3 Training and Generation Methodology

In order to represent a complete garment we separate its visual appearance into its shape, and its style, described by colour, pattern and texture. We use a dataset of fashion images and their associated garment segmentations. The segmentations divide the image into their component classes: background, hair, face, upper-body clothes, lower-body clothes, legs, and arms. We use the segmentation to define the shape of each garment.

We capture the representation of the shape and style independently and generate them separately in two separate networks, as in Figure 2. We can utilise the style generation stage alone to alter the style of garments in an image while maintaining the shape for *Garment Style Transfer* or alter both the shape and style of garments for *Complete Garment Transfer*. We condition the generators on the shape and style separately to ensure their independence.

In the following section, we train two encoder-generator pairs to generate realistic clothing and realistic part segmentations. We then manipulate the encodings to alter the style of the garments and the shape of the segmentations to reflect another image.

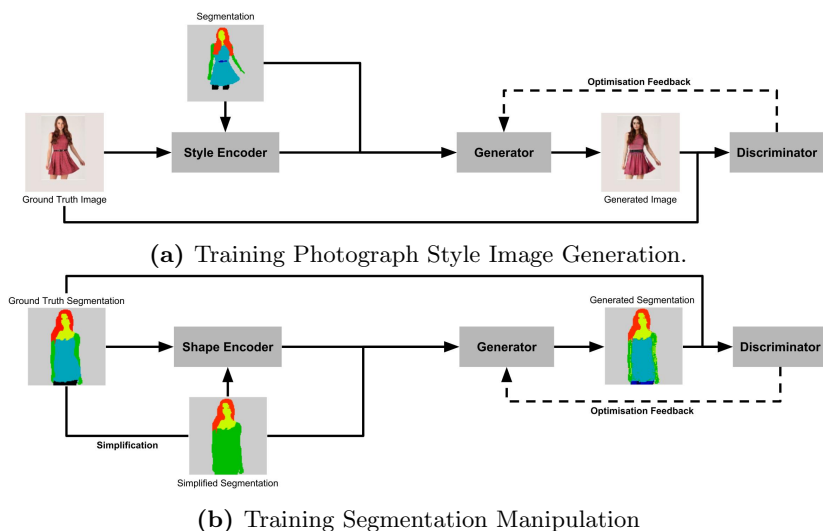


Fig. 2: Training models for the Image Generation stage and the Segmentation Manipulation Stage. The generated image and segmentation are both evaluated on the likelihood they were from the input dataset and on their similarity to the original input.

3.1 Style Training

Our method is a development of conditional GANs [10] which are themselves a development of the GAN [9]. In a GAN, two adversarial networks, a generator and discriminator, are trained simultaneously. The discriminator is trained to distinguish ground truth images from generated results, while the generator attempts to trick the discriminator by generating ever more realistic results. The discriminator learns a more complex and nuanced set of rules about what defines a realistic looking image, and what the conditioning variables represent, than could be specified manually. This complex set of rules in turn encourages the generator to learn to generate more realistic results.

In our method we incorporate a style encoder to condition the generator on the style of the input image, seen in Figure 2a. The style encoder generates a style encoding, \mathbf{e}_i^{img} , for each class, \mathbf{s}_i , in the segmentation, \mathbf{S}_0 , $1 \leq i \leq N$, where N is the number of classes in \mathbf{S}_0 . To ensure coherence within each class, we take

the mean value of \mathbf{e}_i^{img} . The style encodings are then grouped into a feature-wise encoding, \mathbf{E}^{img} , that represents the style of each garment. The encoder is trained with the generator and discriminator to encourage meaningful and specific encodings.

The image generator, G_{image} , uses the shape and style from the original image, represented by \mathbf{S}_0 and \mathbf{E}^{img} respectively, to create its estimate, \mathbf{I}^* , of the ground truth image \mathbf{I}_0 .

$$\mathbf{E}^{img} \leftarrow E_{image}(\mathbf{I}_0, \mathbf{S}_0) \quad (1)$$

$$\mathbf{I}^* \leftarrow G_{image}(\mathbf{S}_0, \mathbf{E}^{img}) \quad (2)$$

The image discriminator only appears in the training phase, shown in Figure 2a. The discriminator evaluates how realistic and representative \mathbf{I}^* is of the original dataset with a traditional unconditional GAN loss function. A perceptual loss term [7] in the discriminator gives a broader evaluation of texture and the wider context of the image by comparing activations at various layers in a pre-trained image classifier conditioned on either the generated or ground truth image. Our perceptual loss utilises a VGG-19 network, pretrained on the ImageNet dataset. A conditional loss function also evaluates how reflective \mathbf{I}^* is of \mathbf{E}^{img} . This form of conditional discrimination has been shown to incentivise consistency between conditions and generated results [17,31].

We use a network architecture inspired by Isola et al. [6], with two down-sampling layers and nine ResNet blocks [22] in the generator, and use a three value encoding for the style of each garment. The networks were trained for 50 epochs. These training conditions create a generator that can alter the style of clothing while maintaining the shape of the garments, allowing generation of realistic clothing with a given style.

3.2 Shape Training

As the style generation stage strictly maintains the shape of each garment, a segmentation stage is needed to alter garment shapes while maintain pose of the original subject. The generated segmentation can then be used as a conditioning input for the style generation.

The segmentation generation stage is a simplified version of the image generation stage. Both stages are trained in similar ways, depicted in Figure 2. Whereas in the style stage the input is an image of a person and a part segmentation, in this stage we use the part segmentation and another ‘simplified segmentation’. In our application, the ‘simplified segmentation’ contains four classes: background, hair, face, and potential clothing. Where potential clothing refers to areas that could be classified as garments in the generated segmentation, including garment classes, arms, and legs from the original image.

During training the segmentation encoder, $E_{segmentation}$, finds \mathbf{E}^{seg} , a set of segmentation feature encodings \mathbf{e}_i^{seg} , $1 \leq i \leq N^-$, where N^- is the number of classes in the simplified segmentation, \mathbf{S}^- . The segmentation generator then hallucinates \mathbf{S}^* , its attempt to reconstruct \mathbf{S}_0 , as shown in Figure 2b.

$$\mathbf{E}^{seg} \leftarrow E_{segmentation}(\mathbf{S}_0, \mathbf{S}^-) \quad (3)$$

$$\mathbf{S}^* \leftarrow G_{segmentation}(\mathbf{S}^-, \mathbf{E}^{seg}) \quad (4)$$

The segmentation discriminator is identical to that in the image generation stage. The discriminator incentivises the generator to replicate the pose of the original image present in the simplified segmentation and shape of each garment given by the complete segmentation encoding.

While more complex methods of discretisation could be used to insure class constraints, for example mean fields or CRFs, we are able to generate quality results by using a continuous spectrum while training and only later applying the class constraints.

The segmentation stage has a simpler architecture than the style stage to avoid modal collapse. The segmentation generator contains a single downsampling layer and three ResNet blocks. The smaller model and less complex problem meant the segmentation generation stage needed considerably less training than the image generation stage, achieving consistently high-quality results after only two epochs of training.

3.3 Style Transfer

Once trained, G_{image} can be used to transfer the style of garments from a donor image. During style transfer, a translated encoding is created as the combination of the subject and donor encodings as

$$\mathbf{e}_{trans, i}^{img} = \begin{cases} \mathbf{e}_{don, i}^{img} & \text{if } i \in T \\ \mathbf{e}_{sub, i}^{img} & \text{elsewhere} \end{cases} \quad (5)$$

where $1 \leq i \leq N_{trans}$, and N_{trans} is the number of classes in the segmentation, and T is the set of garments to transfer. These encodings are combined to create, \mathbf{E}_{trans}^{img} , where each class reflects the style of the subject or donor image. \mathbf{E}_{trans}^{img} and the original segmentation are then used as the conditioning variables for the image generator, which generates a realistic representation of the subject with garments that reflect the donor's.

3.4 Complete Garment Transfer

We can combine the two stages to create the complete garment transfer method shown in Figure 3. To transfer garments between images, we first generate the translated segmentation, \mathbf{S}_{trans} , from the simplified subject segmentation, \mathbf{S}_{sub}^- ,

as well as the donor segmentation encoding, \mathbf{E}_{don}^{seg} . As above, we combine the image style encodings to create \mathbf{E}_{trans}^{img} . Finally we generate the translated image, \mathbf{I}_{trans} , using \mathbf{S}_{trans} and \mathbf{E}_{trans}^{img} .

$$\mathbf{E}_{don}^{seg} \leftarrow E_{segmentation}(\mathbf{S}_{don}, \mathbf{S}_{don}^-) \quad (6)$$

$$\mathbf{E}_{sub}^{img} \leftarrow E_{image}(\mathbf{I}_{sub}, \mathbf{S}_{sub}) \quad \text{and} \quad \mathbf{E}_{don}^{img} \leftarrow E_{image}(\mathbf{I}_{don}, \mathbf{S}_{don}) \quad (7)$$

$$\mathbf{S}_{trans} \leftarrow G_{segmentation}(\mathbf{S}_{sub}^-, \mathbf{E}_{don}^{seg}) \quad (8)$$

$$\mathbf{E}_{trans}^{img} \leftarrow \text{Combine}(\mathbf{S}_{trans}, \mathbf{E}_{sub}^{img}, \mathbf{E}_{don}^{img}) \quad (9)$$

$$\mathbf{I}_{trans} \leftarrow G_{image}(\mathbf{S}_{trans}, \mathbf{E}_{trans}^{img}) \quad (10)$$

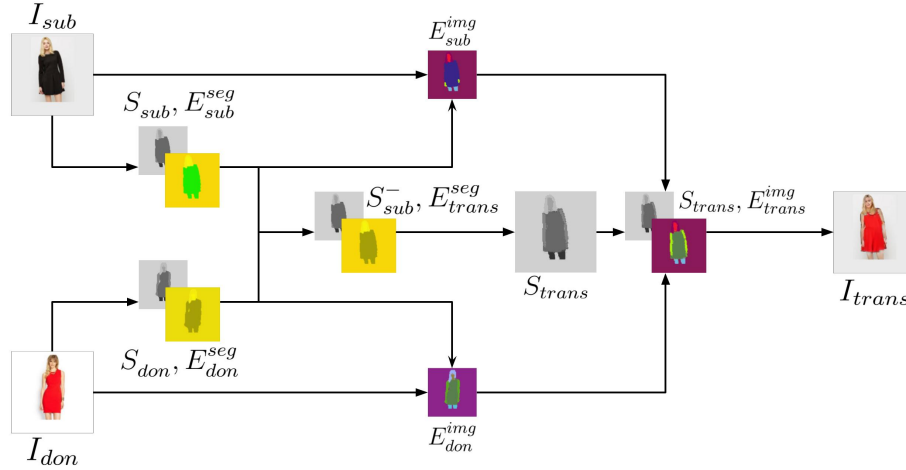


Fig. 3: Complete Garment Transfer schematic.

4 Evaluation and Results

After training the two stages independently on replication tasks, they were then both applied to transfer tasks. In this section we evaluate the style and shape transfer stages alone, before then evaluating the performance of the stages together. The first characterisation of performance is to replicate ground truth

images by using the same image as both the subject and the donor. This was done with both stages independently, and later together. After this, we evaluate the method’s ability to transfer style and shape as intended by replacing the garments in the subject image with those captured from the donor image.

It should be noted that a better representation of the subject would be achieved if the background, face and hair pixels were replaced with their original values. However, all the data presented is from the generator to demonstrate the ability to reproduce these features.

4.1 Dataset

While our method could be applied to any image domain we chose to generate clothing due to its complexity and diversity. We used the DeepFashion dataset [13], a large dataset of low resolution (128x128pixels) fashion images and their associated segmentations from Zhu et al. [31]. We used the same 70,000 training images as Zhu et al. [31].

4.2 Garment Style Transfer

As the two stages are independent it is possible to evaluate the quality of the style transfer stage alone. We evaluate the style transfer stage in two ways: we attempt to replicate a ground truth image by using the same image as both subject and donor, as in Figure 4, we also attempt to transfer styles of various sections of a single donor image onto the associated sections on a subject image, as demonstrated by Figure 5.

Our method is able to generate realistic and representative results, as can be seen by the results from replicating of ground truth images in Figure 4. The model is also able to transfer styles from both example images of garments and abstract style samples to create consistent and realistic images, as in Figure 5. The abstract style transfer pipeline is identical to that used for garment transfer, the only difference being that an entire abstract image is encoded instead of a selected area from a garment segmentation. Abstract style transfer would allow users to design and alter clothing without requiring an exact subject image matching their requirements.

With only three encoder values defining the style of each garment, the style encoding can produce complex patterns, as can be seen in the bottom row of Figure 5. These patterns and textures are significantly more complex than demonstrated in previous two dimensional garment generation work [31].

Complex garment features, such as creases, are convincingly generated in the correct garments and locations. The accurate location of creases in appropriate clothing shows that folds in clothing are correctly taken to be a function of garment shape and are ignored by the style encoder, this also explains the presence of realistic creases in loose garments generated using an abstract style as in Figure 5.

We have compared our method to the neural style transfer method of Gatys et al. [5] both qualitatively as shown in Figure 6 and quantitatively as is discussed

further below. The traditional style transfer method is only able to extract a single style from the whole donor image, meaning we could only compare the results for abstract style transfer. Gatys et al. [5] also focused on transfer of image texture, so could not transfer a single, flat colour. Their method is also unable to selectively apply this style to a subject image, so in order to evaluate their method for clothing, we masked the results and only perform numerical evaluation on the pixels in the areas of interest in the final image.

The traditional approach is unable to separate garment shape and pattern, so some of the original garment pattern is present in the final image. Our method, on the other hand is agnostic to the original garment style and presents no artifacts of the original pattern. Since the traditional style transfer method has not been trained to specifically generate realistic clothing it is unable to generate any realistic creases or similar features.

We compare the two methods quantitatively by translating the style of an image twice, first to an abstract style, then back to the original images style to generate an approximation of the original, ground truth image. We find both the average pixel difference between the original and ground truth images and the percentage of pixels within certain absolute pixel difference tolerances. These results are shown in Table 1.

Due to the nature of GANs and our goal, to create novel images of people in different outfits while maintaining their pose, complex analysis and evaluation is difficult, since we are creating an image that has not existed before. Comparison to other work is also difficult as we are unaware of any similar work on semantic style transfer. While the comparison above is not perfect, as it does not measure global perceptual appearance, we believe it sufficient to objectively show our method is superior to a standard approach. We also believe that because this evaluation only values pixel similarity it puts our more nuanced results at a disadvantage to a more similar pattern generation tool, as, for example, a crease may be generated in a different place which would be perceived as realistic but evaluated as a large error.

The poor quality of the DeepFashion dataset creates simple errors in our style transfer results. The dataset’s segmentations contains a significant number of misclassification errors. The effect of which can be seen in the hands in Figure 4. The dataset also has very limited classes, excluding less common clothing items such as jewellery, hats and belts so this information has to be captured in the style encoding.

Table 1: Quantitative evaluation of our method and traditional style transfer [5]. Values refer to absolute pixel value differences

	Mean difference	% within 15	% within 25	% within 50
Ours	24.5	45.93	64.59	86.548
Masked NST	92.64	10.75	17.92	33.84



Fig. 4: Style Replication Results. The same input image (*top*) is used as the subject and donor for style transfer. The ideal result would be for the generated image to be identical to the input image.



Fig. 5: Style Transfer Results. The style of a donor image (*left*) is transferred to chosen sections of subject images. The style of the dresses are transferred in all cases, the hair colour is also transferred in the case of the top row. Notice how the creases look natural and realistic in the first two rows. For full effect view in colour.

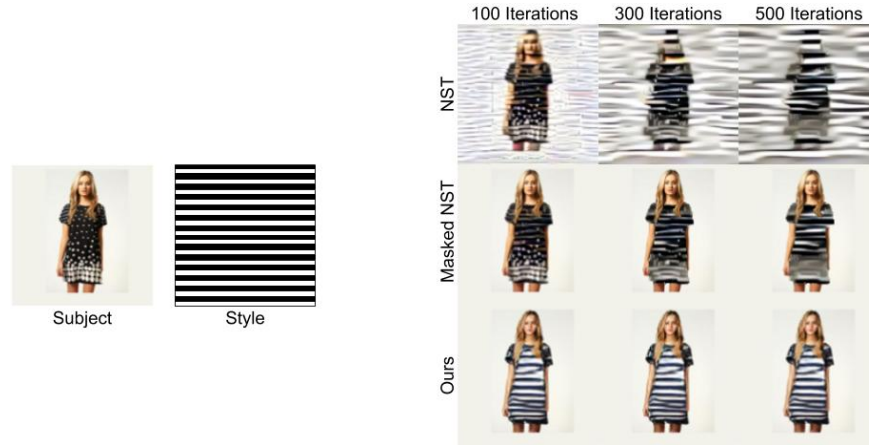


Fig. 6: Comparison between our method and Neural Style Transfer Gatys et al. [5].

4.3 Segmentation Transfer

Despite the poor quality and errors in the dataset, our segmentation generation model generates well defined, acceptable results. The garment shape is altered while the subject pose and body shape are maintained. This is clearly demonstrated in Figure 7a, where the sleeve length of a dress is dramatically changed.

This challenging shape alteration is achieved even when the original segmentation is in an obstructive pose, here the person with arms against their sides is wearing long sleeves.

4.4 Complete Garment Transfer

We combine the two stages to create a single garment generation method, which we evaluate similarly to the the style transfer stage, first by replicating ground truth images, as seen in the leading diagonal of Figure 7b, then by evaluating true garment transfers between images as demonstrated by the remainder of Figure 7b and Figure 8.

We found the generated results were realistic and reflected the shape and style of donor garments well. The bottom row of Figure 8 further demonstrates the encoder’s ability to capture, as well as replicate, the main colours and repetitive nature of a complex pattern with only three values in the style encoding.

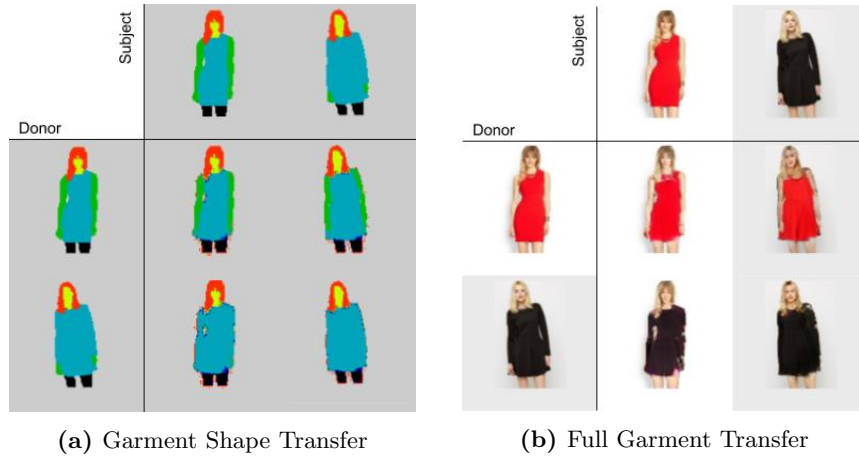


Fig. 7: Full Garment Swap. The replicated ground truth images (leading diagonal) are accurate, even with somewhat flawed segmentations.



Fig. 8: Garment Transfer Results. The dress from a donor image (*left*) is transferred to various subject images.

5 Conclusion and Discussion

We present a method which enables the transfer of object shape and style between images. Our method includes two separate GANs to alter the shape and style of an object independently. The shape alteration stage can maintain the original subject pose while being invariant to the original shape. The style of the generated object is only dependent on the style extracted from the donor image. The remainder of the image is unaffected by the object transfer. This method could be applied to a variety of two dimensional object transfer tasks. We apply this method to complex garment transfer tasks and achieve impressive, realistic clothing transfer.

Our current work has been limited by the lack of variation in the dataset we used. It would be valuable to explore a more diverse dataset containing more poses and body shapes. Here we used three values to define each garments shape. While this encoding has demonstrated the quality of the method experimentally, a more diverse dataset would likely require the use of a wider shape encoding to capture the full variation of the available data. It is likely too that the loss function in shape stage could be optimised to encourage low frequency results.

References

1. Bossard, L., Dantone, M., Leistner, C., Wengert, C., Quack, T., Van Gool, L.: Apparel classification with style. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **7727 LNCS**(PART 4), 321–335 (2013). https://doi.org/10.1007/978-3-642-37447-0_25
2. Denton, E., Chintala, S., Szlam, A., Fergus, R.: Deep Generative Image Models using a Laplacian Pyramid of Adversarial Networks. CoRR (jun 2015)
3. Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C.: LNCS 7574 - Computer Vision ... ECCV 2012 (2012)
4. Gatys, L.A., Ecker, A.S., Bethge, M.: A Neural Algorithm of Artistic Style (2015). <https://doi.org/10.1167/16.12.326>, <http://arxiv.org/abs/1508.06576>
5. Gatys, L.A., Ecker, A.S., Bethge, M.: Texture Synthesis Using Convolutional Neural Networks. The IEEE conference on computer vision and pattern recognition pp. 2414–2423 (2015). <https://doi.org/10.1109/CVPR.2016.265>
6. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017 **2017-Janua**, 5967–5976 (nov 2017). <https://doi.org/10.1109/CVPR.2017.632>, <http://arxiv.org/abs/1611.07004>
7. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) **9906 LNCS**, 694–711 (mar 2016). https://doi.org/10.1007/978-3-319-46475-6_43
8. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive Growing of GANs for Improved Quality, Stability, and Variation (oct 2017)
9. Kinoshita, S., Ogawa, T., Haseyama, M.: LDA-based music recommendation with CF-based similar user selection. 2015 IEEE 4th Global Conference on Consumer Electronics, GCCE 2015 pp. 215–216 (jun 2016). <https://doi.org/10.1109/GCCE.2015.7398561>

10. Larsen, A.B.L., Sønderby, S.K., Larochelle, H., Winther, O.: Autoencoding beyond pixels using a learned similarity metric. *Class Project for Stanford CS231N: Convolutional Neural Networks for Visual Recognition, Winter semester 2014* **5**, 2 (2015)
11. Lassner, C., Pons-Moll, G., Gehler, P.V.: A Generative Model of People in Clothing. *Proceedings of the IEEE International Conference on Computer Vision 2017-**Octob***, 853–862 (2017). <https://doi.org/10.1109/ICCV.2017.98>
12. Liu, M.Y., Breuel, T., Kautz, J.: Unsupervised Image-to-Image Translation Networks (mar 2017)
13. Liu, Z., Luo, P., Qiu, S., Wang, X., Tang, X.: DeepFashion: Powering Robust Clothes Recognition and Retrieval with Rich Annotations. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1096–1104. IEEE (jun 2016). <https://doi.org/10.1109/CVPR.2016.124>
14. Men, Y., Lian, Z., Tang, Y., Xiao, J.: A Common Framework for Interactive Texture Transfer. *CVPR* (2018)
15. Odena, A., Olah, C., Shlens, J.: Conditional Image Synthesis With Auxiliary Classifier GANs (oct 2016)
16. van den Oord, A., Kalchbrenner, N., Vinyals, O., Espeholt, L., Graves, A., Kavukcuoglu, K.: Conditional Image Generation with PixelCNN Decoders (jun 2016)
17. Reed, S., Akata, Z., Yan, X., Logeswaran, L., Schiele, B., Lee, H.: Generative Adversarial Text to Image Synthesis (may 2016)
18. Stroud, M.L., Stilgoe, S., Stott, V.E., Alhabian, O., Salman, K.: Vitamin D - A review. *Australian Family Physician* **37**(12), 1002–1005 (sep 2008). <https://doi.org/10.1177/1470320311410924>
19. Wang, T.C., Liu, M.Y., Zhu, J.Y., Tao, A., Kautz, J., Catanzaro, B.: High-Resolution Image Synthesis and Semantic Manipulation with Conditional GANs (nov 2017)
20. Wang, X., Gupta, A.: Generative image modeling using style and structure adversarial networks. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **9908 LNCS**, 318–335 (mar 2016). https://doi.org/10.1007/978-3-319-46493-0_20
21. Wells, A., Wood, J., Xiao, M.: Localized Style Transfer With Semantic Segmentation p. 2015 (2016)
22. Wu, S., Zhong, S., Liu, Y.: Deep residual learning for image steganalysis. *Multimedia Tools and Applications* pp. 1–17 (dec 2017). <https://doi.org/10.1007/s11042-017-4440-4>
23. Xian, W., Sangkloy, P., Agrawal, V., Raj, A., Lu, J., Fang, C., Yu, F., Hays, J.: TextureGAN: Controlling Deep Image Synthesis with Texture Patches (jun 2017)
24. Yan, X., Yang, J., Sohn, K., Lee, H.: Attribute2Image: Conditional image generation from visual attributes. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **9908 LNCS**, 776–791 (dec 2016). https://doi.org/10.1007/978-3-319-46493-0_47
25. Yang, S., Ambert, T., Pan, Z., Wang, K., Yu, L., Berg, T.L., Lin, M.C.: Detailed garment recovery from a single-view image. *CoRR* **abs/1608.01250** (2016)
26. Yang, W., Luo, P., Lin, L.: Clothing co-parsing by joint image segmentation and labeling. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition* pp. 3182–3189 (feb 2014). <https://doi.org/10.1109/CVPR.2014.407>

27. Zhang, C., Pujades, S., Black, M., Pons-Moll, G.: Detailed, accurate, human shape estimation from clothed 3D scan sequences. *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017* **2017-Janua**, 5484–5493 (mar 2017). <https://doi.org/10.1109/CVPR.2017.582>
28. Zhang, H., Xu, T., Li, H., Zhang, S., Wang, X., Huang, X., Metaxas, D.: StackGAN: Text to Photo-Realistic Image Synthesis with Stacked Generative Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision* **2017-October**, 5908–5916 (dec 2017). <https://doi.org/10.1109/ICCV.2017.629>
29. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision* **2017-October**, 2242–2251 (2017). <https://doi.org/10.1109/ICCV.2017.244>
30. Zhu, J.Y., Zhang, R., Pathak, D., Darrell, T., Efros, A.A., Wang, O., Shechtman, E.: Toward Multimodal Image-to-Image Translation (nov 2017)
31. Zhu, S., Fidler, S., Urtasun, R., Lin, D., Loy, C.C.: Be Your Own Prada: Fashion Synthesis with Structural Coherence. *Proceedings of the IEEE International Conference on Computer Vision* **2017-October**, 1689–1697 (oct 2017). <https://doi.org/10.1109/ICCV.2017.186>