INTERIOR DESIGN 2.0: STYLE TRANSFER BY STYLEMIXING OF REAL IMAGES USING STYLEGANS

A. Rayyan Ul Haq, B. Mudasir Hanif Shaikh, C. Syed Sameer Nadeem

Habib University
DSSE
Karachi

ABSTRACT

Style transfer is an unsupervised task that includes transferring the style of a source image to another image called a destination image. While many attempts have been made at getting realistic images from style transfers, there hasn't been much success apart from the facial feature mixing done by NVIDIA. In this paper, we attempt to apply a similar method to bedrooms in our effort to automate the interior designing process. To this end, we use a pre-trained generator of the StyleGAN and use real bedroom images to map them to the latent space of StyleGAN. In doing so, we get the latent vectors which allow us to do further style mixing and image morphing between any two images.

***Index Terms*—** Style Transfer, Generative Adversarial Network, StyleGANs, Latent Space, interior design

1. INTRODUCTION

Style transfer is an unsupervised task that involves transferring the style of a source image to another image called a destination image while retaining some essence of both images. The idea was first introduced by [1] that produced artistic images. While Neural Style Transfer produced artistic images, the state-of-the-art in realistic style transfers, that is, using real images as source and destination and getting a real image as output, is a StyleGAN. While attempts have been made at getting realistic images from style transfers, there hasn't been much success apart from the facial feature mixing done by NVIDIA. In this paper, we attempt to apply a similar method to a dataset of bedrooms in our effort to automate the interior designing process. To this end, we use a pre-trained generator of the StyleGAN and use real bedroom images to map them to the latent space of StyleGAN. In doing so, we get the latent vectors which allow us to do further style mixing and image morphing between any two images.

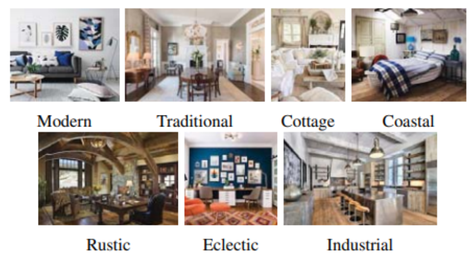


Fig. 1. Examples for room styles [2]

1.1. The Challenge

The challenge we aim to solve is that, given two images belonging to the interior of rooms, transfer the style of the first image to the second without changing the high level features of the first such as the size of the room and ideally the position of the objects. In essence transfer the theme of one room to another. This can be useful to users trying to see if recreating a particular style they observe would suit their setting. Rooms may have many different types of themes such as classical, modern, traditional or cultural ones such as Japanese or it could even simply be a color scheme that needs to be transferred. The approach we suggest will allow one to transfer details at varying levels of details and observe what their room would look like in that particular theme. Different themes are shown in 1

1.2. Dataset

The dataset that we are going to use is the popular LSUN Bedroom dataset[3]. The dataset includes 256x256 RGB images of different rooms. Although the name says "Bedrooms" the dataset contains other rooms and their styles as well. This is a public dataset and perfect for our use, as we can sample as many images as we want from varying styles and room designs. There are over three hundred thousands images in the dataset across different styles. Most of these style categories are shown in 1.

2. RELATED WORKS

2.1. Neural Style Transfer

Given the problem one may instantly recall the approach of Neural Style Transfer [1]. The method synthesizes a novel image by combining the content of one image with the style of another image (typically a painting) based on matching the Gram matrix statistics of pre-trained deep features. While it is true that high level features from the content image will not change, however, this does not solve our problem, firstly because the images generated are not realistic, and secondly, we have no control over how much style gets transferred without retraining the network and thirdly, when transferring style it is indiscriminate to the objects in the target image, that is, it may transfer style of closet to bed, wall to chair, and so on. Our primary focus, however, is to generate realistic images of an interior of a room given two pictures of interiors. We want to copy just the overall theme, color, style from the destination(style) image to the source(content) image.

2.2. Image-to-Image Translation

As the field of deep neural networks has evolved significantly over the last decade and with the advent of Generative Adversarial Networks (GANs) really inspires many works in Image-to-Image translation. Through Image-to-Image translation, we try to infer correspondences between a source image and another target image. This usually requires a set of paired images for training. Isola et al propose a conditional GAN called "pix2pix" model for a wide range of supervised Image-to-Image translation tasks [4]. However, paired data may be difficult or even impossible to obtain in many cases, as in our case because no such dataset is available that shows how a room would look like in a theme of a different room. A relatively recent model, CycleGAN [4] is proposed to tackle the unsupervised image-to-image translation problem by constraining two cross-domain translation models to maintain cycle consistency. CycleGans can easily solve the problem of generating realistic images and are almost perfect for our task - almost. CycleGans can successfully transfer images from one domain to another such as they can be trained to take an image of a room with a traditional theme and output what it would look like in the modern setting. Although, they generate very realistic images and are able to retain a wide level of features but in our case a different network would need to be trained for each pair of themes and they also offer no control over the amount of style transferred.

2.3. StyleGANs

Here, StyleGANs come in. Due to a lack of interior design style transfer using modern deep learning methods we first look at the older models that utilize probabilistic methods to generate designs from scratch. In fact, the use of StyleGANs for interior designing is a very recent application being looked into. In the older models, one study looked specifically at the colour, and defines it as an important aspect when considering interior design. They use a data driven approach and train a Bayesian Network [5]. The goal they keep is to create harmonious colours for the room and furniture based on some specified base furniture. They also use a real interior design dataset to work with the decorative styles and colours. The optimum colour scheme is found by maximizing the conditional joint probability of the colour assignment. There has also been work done on using a source image or video to transfer its style onto a 3D representation of a room [6]. The method involves combinatorial optimization to apply the material extracted from the guide source to the objects in the 3D constructed space. They further optimize it to fulfil some metrics, such as spatial material organization etc.

Another avenue taken in interior designing is the use of AR [7]. Projecting the design, furniture etc. onto the space of the room that is constructed beforehand to have a feel of it before actually investing is the idea of this paper. Aside from being classical methods they have one thing in common, that is they require a base structure to work on, such as the layout of the empty room, as in [5] and [7], or somehow extract the layout of the empty using geometric methods as in the case of [6] and then completely style the room from scratch, as well as introduce objects. Our target however is to allow the objects or whatever the level of detail the person deems sufficient to be extracted from the original image of the room and the style applied to it from a secondary image, thus in essence to vastly generalize this process and make it simpler. There has not been any focused work using deep learning methods in this domain, only the application of general approaches. This began with [2] where they used a pretrained vgg network and transfer learning to classify the image of various interior shots according to their themes such as modern, classic and so on. The paper [8] explored the performance of various methods of generating images from limited data including using conditional and unconditional GANs, choice of target and source domains (such as modern to traditional in our case), and use of pretrained models to accelerate learning. Although the paper is very wide, it also conducted experiments on LSUN Bedrooms (dataset of images of the interior of bedrooms). The direct in-applicability of this to our problem has already been discussed in the problems of a simple GAN above. The best it can do is to generate two images using random latent vectors and then perform style mixing of these two generated images using arithmetic operations on their latent vectors, whereas we aim to perform this on real images. However, it puts forward an interesting idea which is to use pre-trained models for GANs and implements it successfully on the bedroom dataset, so it may also be applicable in our case.

Finally, the paper with the work closest to ours and with good results uses deep photo style transfer [9] using a modified neural style transfer. Their idea is that since neural style is indiscriminate, as discussed, one should semantically label the images before performing style transfer and then modify the transfer so that styles are transferred to corresponding subregions in the target image. To solve the problem of images not looking realistic, they introduce an additional loss term that judges photo-realisticness of the image.

They achieve remarkable results and in comparison there are only two things left for us to improve: remove the need for labelling and still achieve sub-region style transfer (which the disentanglement of intermediate latent vector in styleGANs allows for) and control the degree of style transfer without retraining the network (which the insertion of this vector at multiple points in the network allows for).

3. METHODOLOGY

Given the problem, that is, to design such a system where on a destination image (i.e. the user's room image) a style of the source image is to be applied we tackle it as follows.

We start off by setting up the StyleGAN generator as made by T. Karras et al in [10] using pretrained weights for Karras bedrooms. Pretrained weights are opted for as we do not have access to resources to train the generator adequately ourselves. This was done with the help of pytorch as implementation in tensorflow for which the system was designed in [10] was not feasible. For this we referred to the workings from this github repository [11]. After the generator is set up we test it for working.

Then we moved onto styleMixing using real images. As the StyleGAN generator requires intermediate latent vectors as inputs, the real images need to be encoded into the latent space. The styleMixing occurs when of the x layers to which the intermediate latent vector for images is provided to the generator, n amount of layers get the latent vector of the source image and $x - n$ layers get the latent vector for the destination image. This required us to build a convolutional encoder. We tried various architectures and parameters, a deep convolutional network, resnet followed by dense layers and fully dense networks were tried. These encoders were trained but did not provide very promising results over a large database. We optimized for the images utilized to get our latent vectors. The intermediate latent vectors were required of length 512 and the model utilizes this in 14 places.

3.1. Mixing and encoding implementation

We tried out 3 procedures for the implementation.

Procedure 1 entailed using the encoder to generate the vector and optimize it such that when used in the styleGANs, it generates the encoded image back. This encoding image

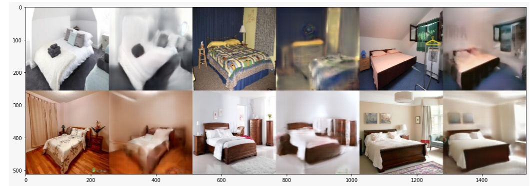


Fig. 2. Reconstruction of encoded image via procedure 1

Fig. 3. Resconstruction of encoded image via procedure 2

did not give good results as the images generated do not represent the original ones well. However, for the occasional good generation of image, the style mixing results are very satisfactory. This is done on real images. 4(iii)

Instead of giving the same latent vector at all the 14 places needed, in procedure 2 we generated an intermediate latent vector for all the insertions. This gave great encoding results, such that the decoded images from the styleGAN were close to the original. However, styleMixing through this procedure gave average results. As we can control the degree of style application, we varied it to see the results. Procedure 2 gave good results for a small degree of style application, but as the number of layers to which the destination (style) image latent vector is fed increased, the resulting image disfigured. 4(iv)

Lastly in procedure 3, we trained the encoder such that starting from a vector of zeros, the latent vector is produced such that the image generated by the StyleGAN generator gives us nearly the original image back.

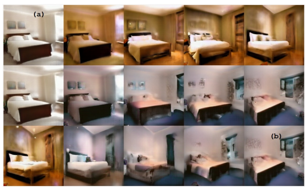
3.2. Loss functions

Three Loss functions as well as their combinations were used to train the models in the aforementioned procedures. They were namely, mean squared error (MSE), mean absolute error and perceptual/feature loss using VGG. For procedure 1, the best results were obtained using MSE and the 2nd conv layer of the VGG. Whereas, for procedure 2 the 12th layer of VGG provided the best results.

The best results in descending order were developed from the following combinations, procedure 1 with images generated by the GAN itself (generated images) 4(i), procedure 2 with generated images 4(ii), procedure 1 both source and destination being real images, followed by procedure 2, both real images.

(i) Procedure 1 via generated images

(ii) Procedure 2 via generated images



(iii) Procedure 1 via real images

(iv) Procedure 2 via real images

Fig. 4. StyleMixing results - (a) is the source image, and (b) is the destination image

4. CONCLUSION

Overall, styleMixing to get the interior design styles transferred is a promising domain. Our work gives us good results and can be improved by looking into better encoder training along with better architectures. The results obtained provide substantial validation of our methodology. Separate latent vectors for separate layer identification helps the model in stylemixing and can be a venue to further explore.

5. REFERENCES

[1] Leon Gatys, Alexander Ecker, and Matthias Bethge, "A neural algorithm of artistic style," *Journal of Vision*, vol. 16, no. 12, pp. 326, 2016.

[2] Esra Ataer-Cansizoglu, Hantian Liu, Tomer Weiss, Archi Mitra, Dhaval Dholakia, Jae-Woo Choi, and Dan Wulin, "Room style estimation for style-aware recommendation," *2019 IEEE International Conference on Artificial Intelligence and Virtual Reality (AIVR)*, 2019.

[3] Fisher Yu, Yinda Zhang, Shuran Song, Ari Seff, and Jianxiong Xiao, "Lsun: Construction of a large-scale image dataset using deep learning with humans in the loop," *arXiv preprint arXiv:1506.03365*, 2015.

[4] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.

[5] Guangming Chen, Guiqing Li, Yongwei Nie, Chuhua Xian, and Aihua Mao, "Stylistic indoor colour design via bayesian network," *Computers & Graphics*, vol. 60, pp. 34–45, 2016.

[6] Chuong H. Nguyen, Tobias Ritschel, Karol Myszkowski, Elmar Eisemann, and Hans-Peter Seidel, "3d material style transfer," *Computer Graphics Forum*, vol. 31, no. 2pt2, pp. 431–438, 2012.

[7] Viet Toan Phan and Seung Yeon Choo, "Interior design in augmented reality environment," *International Journal of Computer Applications*, vol. 5, no. 5, pp. 16–21, 2010.

[8] Yaxing Wang, Chenshen Wu, Luis Herranz, Joost Van De Weijer, Abel Gonzalez-Garcia, and Bogdan Raducanu, "Transferring gans: Generating images from limited data," *Computer Vision – ECCV 2018 Lecture Notes in Computer Science*, p. 220–236, 2018.

[9] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala, "Deep photo style transfer," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[10] Tero Karras, Samuli Laine, and Timo Aila, "A style-based generator architecture for generative adversarial networks," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[11] Piotr Bialecki and Thomas Viehmann, "Pytorch implementation of the stylegan generator," Feb 2019.