Democratize the creation of Pop Art using Neural Style Transfer

Ang Li

George School, PA, USA

Hzhc_wjdi@163.com

Keywords: pop art, Andy Warhol, Neural Style Transfer, art production, artificial intelligence (AI)

Abstract: The purpose of this essay is to explore how we can employ AI in the field of Art. Living in a world where technology rapidly developed, artificial intelligence becomes less and less peculiar for moderns. AI had been already applied to vast fields, including pharmaceutical production, automatic car, and etc. which benefits human life in a large extent. However, not until 2015, Artificial intelligence provided a way that can transfer the style of one picture to another picture while keeping its content. The technique of style transfer has achieved a mature level in recent years. However, comparing with the great achievement in this technique, the application and prospect for neural style transfer is still obscure and concerning. This essay examines the Neural Style Transfer and provides a possible application for this technique in art production. With the study of pop art master Andy Warhol's artworks and artistic techniques, the essay explores how artificial intelligence and art can be combined to refine and improve art production. Further, this essay also researches how artificial intelligence can benefit middle class and lower class's art production.

1. Introduction

Art is a way of expression. To some extent, it can be thought of as a more advanced language, which is more than a collection of symbols: colors, shapes, and any other tangible elements can play a part in this "synthetic language".

Art has always been the privilege of a few artists. It was not until the arrival of pop art, does art come to the life of ordinary people. However, even in pop art, people only learn to appreciate art as audience. In my project, I am introducing neural transfer as a new technique in order to democratize the process of art creation so that everyone is able to express himself freely through a variety of artistic method. My neural-style-transfer project aims to empower an individual to create, illustrate and demonstrate his artistic ideal freely without the need of much professional artistic training.

Any artistic creation can generally be divided into two components. One is content, while the other is style. Content to art is like media, which symbolizes existence and being. Style to art serves more as a mood, which delivers feeling, sensation and intuition. Through introducing neural-style-transfer method, one can offer Picasso's abstract painting an impressionism layer. One can also introduce Picasso's interpretation of art into Monet's Impression Sunrise. Through my project, barrier between different art genres gets teared off.

My project reminds me frequently art works by Andy Warhol. Using commodities as the primary content, Andy successfully makes his work appealing to an ordinary family. In his work, he could change colors, but not really styles, in order to bring variety into his artistic creation. Through my technique, this expression of freedom, democracy and popularity can be brought to another level, on which every single piece can stand for a unique stylistic representation.

2. Art Background

In the last century, one of the most famous art movement is the Pop Art movement initiated by Andy Warhol. Following the popularity of Abstract art, Pop art reintroduces recognizable imagery from media and culture such as advertising, cartoons and etc. It mainly creates artworks with modern commercial elements such as sculptures and of the famous actors, the poster of cans and Coca-Cola,

and etc. to make a fine art. Pop Artists aim to blur the boundary between so called the traditional “high” art and “low culture”, and that is the key concept of Pop Art: no hierarchy of culture. Contrasting to traditional masterpieces, which depict elegant themes like morality, mythology, religions and history, Pop Art emphasis on using commonplace objects in daily life to elevate popular culture in the level of fine art.

Andy Warhol, born on August 6, 1928, in Pittsburgh, Pennsylvania, was first a successful magazine and advertisement designer. He led the famous Pop Art movement in 1960s by creating symbolic works such as images of Marilyn Monroe, soup cans, and sensational newspaper stories. After 1965, although Warhol continues creating art pieces, the main focus of his career shifted to movie production.

In 1962, Andy Warhol published the paintings of Campbell's soup cans which caused a stir around the world. Compared to the traditional elegant art pieces, this new kind of art contains normal element, a Campbell’s soup which could be found everywhere, echoes with the public who desires for the practice of art but have limitations for approaching. He not only promoted the trend of Pop Art in to the topics of the art world and brought himself into the spotlight of upper class. Moreover, later, Warhol focus on film production, a genre of art which is a much more popular and vulgar art form. Compared with art pieces, films can better represent the spirit of pop art. And film completely transferred Warhol into a true celebrity among the public.

Further, not only Warhol’s pop art makes him an idol of the public, but his personal life, which is the representation of pop art. Born in an immigrant family from Eastern European, Warhol’s childhood was in serious dilemma. He contracted “St. Vitus's Dance”, a fatal chorea in the nervous systems at the age of eight. Father died when he was fourteen, living in the room full of cockroach, all these triadic events creates a striking comparison with his crowning achievement after 1960s. He not only brought pop art into the field of upper class, but also crafts himself as a public icon. He creates his own art studio “the Factory”, which later became the premier cultural hotspots where all celebrities gathered. For many pop art followers, his ambitions, a legendary experience, his prominent achievements which echoes with the true spirit of the pop art movement inspires them.

The most famous feature and style of Andy Warhol’s artwork is the using of silk screening. Shortly after Warhol entering the field of Pop Art, he found out that the traditional canvas drawing method was not fast enough to create mass of pieces. In 1962, he invented the process of silk screening which uses a certain prepared section of silk as a stencil, allowing one silk-screen to create similar patterns multiple times. The most famous application of this method is the duplicate of Marilyn Monroe.

One of the most famous art pieces of Warhol is the Gold Marilyn Monroe. He created the art shortly after the death of Marilyn Monroe. The creation of Marilyn Monroe series has multiple significance. One of them is that the creation of Marilyn Monroe series is the first time Warhol applies silk screening method in art production. It opens a new era of commercial art. From artistic views, compared to other Marilyn Monroe, the golden background seems the emitting light of Marilyn Monroe, implies that the admire of Warhol for Marilyn Monroe. Also, the golden background can be interpreted as the supreme status of Monroe in the field of art. Some art critics states that “The background is reminiscent of Byzantine religious icons that are the central focus in Orthodox faiths to this day. Only the instead of a god, we are looking at an image of a woman that rose to fame and died in horrible tragedy.” Since Warhol created this art shortly after the death of Monroe, this interpretation is profound and convincing. Another artistic critic for the background is that “Redolent of 1950s glamour, the face in Gold Marilyn Monroe is much like the star herself—high gloss, yet transient; bold, yet vulnerable; compelling, yet elusive. Surrounded by a void, it is like the fadeout at the end of a movie.” The emitting and fading light perfectly describes the life of Monroe, full of vanity.

Other than the glittering background, the main portray of Monroe was full of stains, blurs, which makes the whole art pieces looks faded contrast to the golden background. The smudges can be interpreted as the death of Monroe or other notorious affairs she had in her life. The turquoise eye shadows, the bright red lips, and the lemon-yellow hair all exhibits her fancy and popular life as an

actor. Besides the colors, the expression on Monroe's face is a forced, unnatural smile. Her dim and dark eyes create a vivid contrast with the fancy make up. In conclusion, the Golden Marilyn Monroe perfectly concluded Monroe's life, full of vanity and void.

Despite the great achievement Warhol had made in the field of art, there is still some limitations of his art methods that are contrary to his artistic principles. First of all, one of Warhol's aim is to introduce art into lower class. Indeed, Warhol introduced those lower-class elements into the field of "high" art. And he invented the silk-screening method which made it possible for lower classes to create their own art. However, people were not able to craft their personal stencils. The result of pop art movement turns out to be higher class crafting art pieces with lower class elements, art still remained inside higher classes. Even though the content changed, the creation of art still remains in higher classes.

Another limitation of Warhol's art is that even though pop art tries to blur the boundary between "high art" and "low culture", it does not mean everything can be art. Some of Warhol's works are just commercial advertisement without spirits, which are called "garbage's" by some critics. In Warhol's mind, everything that can create money are called art, but art is destined to design with spirits and values. The reason why Campbell's soup cans attract people's eyeballs is not because it is a commercial ad; it is the innovative methods that inspires people in that era. Imagine 20 years later, someone creates an art piece which just simply prints commercial brands, people would not even evaluate it as art. Therefore, it is Warhol's innovative idea made him successful, not those commercial brands contain no spirits.

3. Method

3.1 Original Neural style Transfer

Neural style transfer is to transfer the style of a picture to the target image while keeping the target picture's original information (which is the content of the picture) and therefore presenting an image with a different style but still maintaining its content.

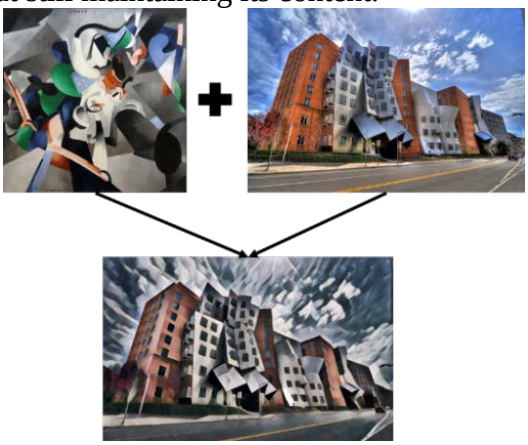

Figure 1. Example of Neural Style Transfer

The Neural style Transfer technique in this essay was based on the research A Neural Algorithm of Artistic Style published by Leon A. Gatys, Alexander S. Ecker, Matthias Bethge. The method illustrated in this essay is a modification and improvement of Leon's approach, which is the Fast Image Style Transfer with faster transfer speed and higher transfer quality. In the following passage, we will introduce how the basic Neural Network is applied to image transfer and how Neural style Transfer works. Research on neural networks has long existed and can be simplified into the following models.

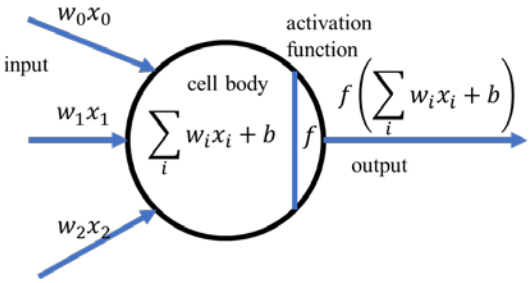

Figure 2. Single Neural model

With the development of neural networks, the concept of deep learning becomes more and more popular in recent years, especially the convolutional neural network (CNN). There are three characteristics of convolutional neural networks: local connections, weight sharing, and sub-sampling. The network effectively solves the shortcomings of excessive parameters in traditional neural networks applied to images. VGG Network is a typical and outstanding convolution neural network among all CNN networks. The basic network structure is shown in the following table.

Table 1. The Structure of VGG Network

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
Input(224×224 RGB image)					
Conv3-64	Conv3-64 LRN	Conv3-64 Conv3-64	Conv3-64 Conv3-64	Conv3-64 Conv3-64	Conv3-64 Conv3-64
Maxpool					
Conv3-128	Conv3-128	Conv3-128 Conv3-128	Conv3-128 Conv3-128	Conv3-128 Conv3-128	Conv3-128 Conv3-128
Maxpool					
Conv3-256 Conv3-256	Conv3-256 Conv3-256	Conv3-256 Conv3-256	Conv3-256 Conv3-256 Conv3-256	Conv3-256 Conv3-256 Conv3-256	Conv3-256 Conv3-256 Conv3-256 Conv3-256
Maxpool					
Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv3-512 Conv3-512
Maxpool					
Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv3-512	Conv3-512 Conv3-512 Conv3-512 Conv3-512
Maxpool					
FC-4096					
FC-4096					
FC-1000					
Soft-max					

When the convolutional neural network is trained, each filter kernel is equivalent to extracting features from the image. As for the characteristics of extraction, in a certain sense, this is something that humans cannot obtain through visual observation. Therefore, we use the trained VGG16 model

to obtain two types of information from the picture: image content information and image style information.

The image content mainly includes people or objects recognizable in the image. The image style information mainly includes styles such as colors and textures in the image. The purpose of the study is to synthetize both images to contain both the overall layout in the content image and the style in the style image without changing the overall layout.

3.2 Real Time Transfer Upgradation

Since the original model is with extremely low efficient and unacceptable qualities of products, it is necessary to improve the model. The upgraded model used in the paper is divided into two parts: Image Transformation Networks and Perceptual Loss Functions. More details are shown in Fig.3.

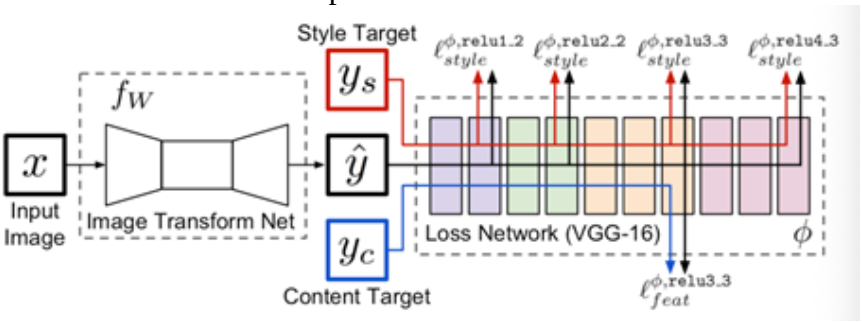

Figure 3. The structure of Model

The image transfer follows the structure provided by Radford's team. Instead of using pooling layers, the technique provided in the essay downsample and upsample with stripe and microstrip convolution to the network. The body of the network consists of five residual blocks. All non-residual convolutional layers follow spatial batch normalization and ReLU nonlinearity except the output layer. It uses a scaled tanh to ensure that the pixels of the output image are in the range [0, 255]. Other than the first and last layers which use 9 × 9 kernels, all convolutional layers use 3 × 3 kernels.

For style conversion, the network uses two convolutions with a stride of 2 to downsample the input, then several remaining blocks, and then use two convolutional layers spanning 1/2 to upsampling. Although the input and output have the same size, there are several advantages for the network to first downsample and then upsample.

The first is computational. With a simple implementation, a 3×3 convolution with C filters on an input of size $C \times H \times W$ requires $9HWC\text{^}2$ multiply-adds, which is the same cost as a 3 × 3 convolution with DC filters on an input of shape DC×H/D×W/D. After downsampling, we can therefore use a larger network for the same computational cost.

The second benefit has to do with effective receptive field sizes. High-quality style transfer requires changing large parts of the image in a coherent way; therefore, it is advantageous for each pixel in the output to have a large effective receptive field in the input. Without downsampling, each additional 3×3 convolutional layer increases the effective receptive field size by 2. After downsampling by a factor of D, each 3×3 convolution instead increases effective receptive field size by 2D, giving larger effective receptive fields with the same number of layers.

Residual Connections. He et al use residual connections to train very deep networks for image classification. They argue that residual connections make it easy for the network to learn the identify function; this is an appealing property for image transformation networks, since in most cases the output image should share structure with the input image. The body of our network thus consists of several residual blocks, each of which contains two 3 × 3 convolutional layers.

3.3 Perceptual Loss Functions

We define a loss function between the synthetize image and the content image to measure the extent to which the synthetize image loses image content.

$$Loss_{content}(\vec{p},\vec{x},\vec{l}) = \frac{1}{2}\sum_{i,j}(F_{ij}^{l} - P_{ij}^{l})^2 \quad (1)$$

In the Equation (1),Where $\vec{p}$donates the content image, $\vec{x}$donates the synthetize image, F_{ij}^{l} donates the activation of the i^{th} filter at position j in layer l, P_{ij}^{l} donates the activation of the i^{th} filter at position j in layer l.And $F_{ij}^{l}, P_{ij}^{l} \in \mathbb{R}^{N_l \times M_l}$, where N_l donates the number of feature maps,M_l donates the size of feature maps.

In the actual training process, the above formula is more difficult to converge. So, use the following formula instead.

$$Loss_{content}\left(\vec{p},\vec{x},\vec{l}\right) = \frac{1}{4\times P\_size}\sum_{i,j}(F_{ij}^{l} - P_{ij}^{l})^2 \quad (2)$$

$$P\_size = width \times height \times channel \quad (3)$$

In the experiment, we chose relu3_3 in VGG16 as the content representation.

Style Reconstruction

Similarly, we define the loss function of the image style:

$$E_l = \frac{1}{4N_l^2 M_l^2}\sum_{i,j}(G_{i,j}^{l} - A_{i,j}^{l})^2 \quad (4)$$

$$Loss_{style}(\vec{a},\vec{x}) = \sum_{l=0}^{L} w_l E_l \quad (5)$$

In the Equation (4)(5), Where G^l donates the gram matrix of the feature maps of the style image in layer l. A^l donates the gram matrix of the feature maps of the synthetize image in layer l. $\vec{a}$ donates the style image. Weighting the feature representations of multiple layers to obtain the total style loss.

In the actual experiment, I used relu1_1, relu2_2, relu3_3, relu4_3. We know that the higher layers extract the style better, so I used incremental weights in the code, which gives the deeper convolution layer a greater weight.

$$w_{conv1\_1} = 0.1 \quad w_{conv2\_1} = 0.15 \; w_{conv3\_1} = 0.25$$

$$w_{conv4\_1} = 0.2 \quad w_{conv5\_1} = 0.3 \quad (6)$$

3.4 Minimize the Total Loss

What's more, to encourage spatial smoothness in the output image $\hat{x}$, we follow prior work on feature inversion and super- resolution and make use of total variation regularizer $l_{TV}\hat{x}$.

Finally, we define the total loss function, which is weighted by content and style. Minimize the total loss and obtain the synthetize image.

$$Loss_{total}\left(\vec{p},\vec{a},\vec{x}\right) = \alpha Loss_{content}\left(\vec{p},\vec{x}\right)$$

$$+\beta Loss_{style}\left(\vec{a},\vec{x}\right) + \gamma l_{TV}\hat{x} \quad (7)$$

And the synthesize image $\hat{x}$ is generated by solving the problem.

4. Data Analysis

The experiment is run under the environment of ubuntu16.04, PyCharm2017, python3.6, and tensorflow1.10. The networks are trained on the Microsoft COCO dataset. We resize each of the 80k training images to 256 × 256 and train the networks with a batch size of 4 for 40,000 iterations, giving roughly two epochs over the training data. I use Adam with a learning rate of 0.001. The output images are regularized with total variation regularization with a strength of between 1×10^{-6} and 1×10^{-4}, chosen via cross-validation per style target. For all style transfer experiments, we compute feature reconstruction loss at layer relu2_2 and style reconstruction loss at

layers relu1_2, relu2_2, relu3_3, and relu4_3 of the VGG-16 loss network φ. It costs about 6 hours on a single GTX 1080Ti GPU to finish training.

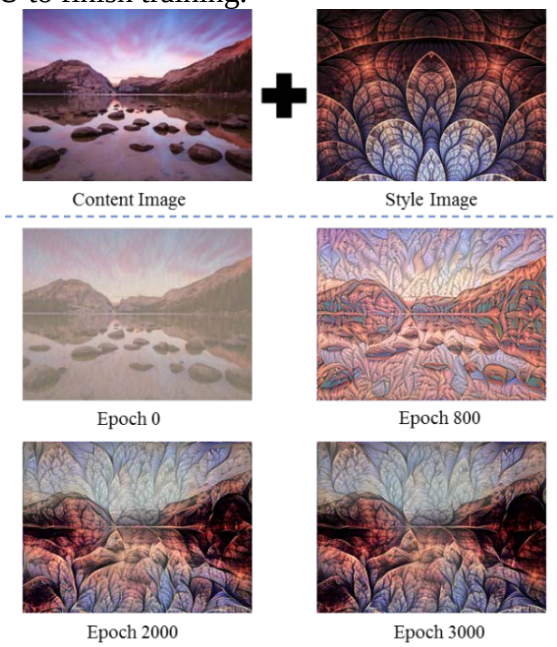

Figure 4. The Result of Experiment 1

As can be seen from the final synthetize image, the texture details of the leaves are well preserved.

We can use this technology to help designers design a variety of styles of posters in a batch, thus reducing the burden of their work.

Most importantly, the technology can use any style to create a film like Loving Vincent. As we all know, a video is composed of a large number of images. After training a specific style of the model, it costs about 300 ms to convert an image style transformation in the experimental environment. Therefore, we can quickly transfer each frame of the video and finally stitch it into a video.

In Experiment 4, I took a video introducing the work I did in this paper. And I transfer two styles into this video (3 mins, 5 frames per second). The result can be obtained on the GOOGLE hard drive. In this experimental environment, you can even call the camera for real-time video style transfer. Of course, this requirement for hardware devices is slightly higher.

5. Discussion And Conclusion

As we discussed when exploring the deficiency of Warhol's artistic style, the most serious deficiency is that even though Warhol want to introduce art into middle class and even lower class, the method of Warhol, silk screening, is an expensive method that makes it impossible for lower class even middle class to participate in art creation.

However, with the development of technology, intelligent devices such as phones and computers become common necessities for middle classes, even some lower classes. Therefore, with the help of neural style transfer technique, it is possible for nearly everyone to participate in air. The spirit of pop art, which is to bring art into the public rather than let it be the toy of elite can finally come true.

References

[1] Leon A. Gatys, Alexander S. Ecker, Matthias Bethge. A Neural Algorithm of Artistic Style. arXiv preprint arXiv: 1508.06576, 2015.

[2] Krizhevsky, A., Sutskever, I. & Hinton, G. E. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, 1097–1105, 2012.

[3] Gatys, L. A., Ecker, A. S. & Bethge, M. Texture synthesis and the controlled generation of natural stimuli using convolutional neural networks. arXiv: 1505.07376 [cs, q-bio], 2015.

[4] Guc¸lu, U. & Gerven, M. A. J. v. Deep Neural Networks Reveal a Gradient in the Complexity of Neural Representations across the Ventral Stream. The Journal of Neuroscience, 35, 10005–10014, 2015.

[5] Cadieu, C. F. et al. Deep Neural Networks Rival the Representation of Primate IT Cortex for Core Visual Object Recognition. PLoS Comput Biol 10, e1003963, 2014.

[6] Kummerer, M., Theis, L. & Bethge, M. Deep Gaze I: Boosting Saliency Prediction with Feature Maps Trained on ImageNet. In ICLR Workshop, 2015.

[7] Mahendran, A. & Vedaldi, A. Understanding Deep Image Representations by InvertingThem. arXiv: 1412.0035 [cs], 2014.

[8] Johnson J, Alahi A, Li F F. Perceptual Losses for Real-Time Style Transfer and Super-Resolution [J]. 2016:694-711.

[9] D Ulyanov, A Vedaldi, V Lempitsky.Instance normalization: The missing ingredient for fast stylization. arXiv preprint arXiv: 1607.08022, 2016.