© IJCIRAS | ISSN (O) - 2581-5334
June 2020 | Vol. 3 Issue. 1

NEW ART WORKS GENERATION USING DEEP LEARNING

Zin Mar Htun[1], Zar Ni Zln [2]

Technological University (Taunggyi), Lecturer, +095, Myanmar

Abstract

In this paper, the author propose a system that can create creative "art" works by combining Creative Adversarial Network (CAN) which is built over Generative Adversarial Network (GAN) and Neural Style Transfer Learning (NSTL). It uses image from CAN, up-sampled it with super-resolution algorithm and feed it into NSTL to generate new creative "art" work. The proposed system introduces a new perspective about creating "art" with Artificial Intelligence. It uses machine's ability to explore infinite creative space and combines it with human understands of shapes and forms about surrounding objects to create new kinds of "art" works that have never been seen before. The proposed system may also have other uses such as it can paint images provided by humans in new "art" styles which then can be used in many artistic affairs. A survey has been conducted on human subjects to compare the "art" work generated by the proposed system with "art" work generated by standalone CAN. The results show that human's rate higher on proposed system's "art" works than those generated by CAN.

Keyword: CAN, GAN and NSTL

1.INTRODUCTION

Imagine a world, where machines can be creative as humans. What kind of "art" works, will machine create? Will humans find them appealing ? or will they find them odious? Will machines' art works and humans' ones coexist ? Or even better, will these two entities co-operate and produce more creative 'art" works ?

In this paper, I proposed a system that will try to answer the last question. The proposed system is based on Creative Adversarial Network (CAN) [17 Elg] and Neural Style Transfer Learning (NSTL) [15 Leo]. It uses machine creativity from CAN and human creativity,(photographs provided by humans) from NSTL and combine them to create creative "art" works. Both CAN and NSTL are, in my opinion, complement for each other.

The proposed system is better than standalone CAN in the aspect that it allows humans to have more control over generated "art" work by allowing human to provide image whose content that they would like to be present in final image using content representation from NSTL. It can be rephrased as allowing humans to inject understanding of shapes and forms, which comes naturally for them, to CAN system.

It is better than standalone NSTL because the style images (See section later) fed into NSTL model are generated by CAN, which is the result of machine exploring creative art space and trying to generate image that increases arousal potential[67 Ber] for observers.

In other words, the proposed system uses image generated by CAN and fed into NSTL as style image, which is an entry point for machine's creativity to flow into the proposed system. NSTL uses both that image and content image provided by human, which allows human to have control over generated art work.

By allowing human to have control means human can provide any creative image as they like. Thus, this can be served as an entry point for human's creativity(understanding of shapes and forms) flows into system. By combining these two creativities, the proposed system produces creative and aesthetic "art" works which looks like human provided image painted in machine's generated style. Furthermore, one can also tune the system for how these two creativities are combined – which creativity will govern the final image.

2. LITERATURE REVIEWS

The CAN's approach is motivated from the theory suggested by D. E. Berlyne (1924-1976). Berlyne argued that the psychophysical concept of "arousal" has a great relevance for studying aesthetic phenomena [71 Ber]. "Level of excitement" quantifies how alert or energized a person is. The level of arousal varies from the lowest level, when a person is asleep or relaxed, to the highest level when s/he is violent, in a fury, or in a passionate situation [1].

Among different mechanisms of arousal, of particular importance and relevance to art are properties of external stimulus patterns [2]. The term "arousal potential" refers to the properties of stimulus patterns that lead to raising arousal. Besides other psychophysical and ecological properties of stimulus patterns, Berlyne emphasized that the most significant arousal-raising properties for aesthetics are novelty, surprising-ness, complexity, ambiguity, and puzzling-ness. He coined the term collative variables to refer to these properties collectively.

Martindale emphasized the importance of habituation in deriving the art-producing system [90 Mar]. In the event that specialists continue creating comparative works of expressions, this straightforwardly lessens the excitement potential and subsequently the attractive quality of that craftsmanship. Along these lines, anytime of time, the workmanship creating framework will attempt to expand the excitement capability of delivered craftsmanship. As such, habituation frames a steady strain to change workmanship. In any case, this expansion must be inside the base sum important to make up for habituation without falling into the negative gluttonous range, as indicated by Wundt bend findings ("boosts that are marginally as opposed to endlessly supernormal are liked"). Martindale called this the principle of "least effort". Therefore, there is an opposite pressure that leads to a graduated pace of change in art [3].

At that point the subsequent sign will strongly punish the generator for doing that. This is on the grounds that the subsequent sign pushes the generator to produce style uncertain works. Thusly, these two signals together should push the generator to investigate portions of the inventive space that lay near the circulation of workmanship (to augment the first objective), and simultaneously amplifies the vagueness of the created craftsmanship concerning how it fits in the domain of standard craftsmanship styles [4].

3. METHODOLOGY

This section discusses the basic knowledge of CAN and NSTL how to generate new art works from two different images.

3.1. Working Principles of CAN

Creative Adversarial Network (CAN) is based on Generative Adversarial Network (GAN) which is one of most successful image synthesis model in recent years. GAN is typically comprised of two neural networks – models, Generator (G) and Discriminator (D). These two models are trained simultaneously. The model G is trained to capture the data distribution, while the model D is trained to estimate the probability that a sample came from the training data rather than G. The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game. In other words, model G is trained to produce fake samples from data distribution and model D is trained to determine whether data are real or fake. The training procedure is similar to a two-player min-max game with the following objective function (Equation 3.1.)

$$min_G max_D V(D,G) = E_{x\sim p_{data}}[log\, D(x)] + E_{z\sim p_z}[log\left(1 - D\left(G(z)\right)\right)] \qquad (3.1)$$

where z is a noise vector sampled from distribution p z (e.g., uniform or Gaussian distribution) and x is a real image from the data distribution p data . In practice, the discriminator and the generator are alternatively optimized for every batch. The discriminator aims at maximizing Equation which improves the utility of the D as a fake vs. real image detector. Meanwhile, the generator aims at minimizing Equation by maximizing log(D(G(z)), which works better than −log(1 − D(G(z)) since it provides stronger gradients. CAN added style (art style) classification loss and style ambiguity loss to GAN. CAN aims to minimize the cross entropy between style class posterior entropy and uniform target distribution. This cross entropy will be minimized when all the classes are equiprobable. Therefore, using the cross-entropy results in a hefty penalty if the generated image is classier to one of the classes with high probability [5].

This in turn would generate very large loss, and hence large gradients if the generated images start to be classified to any of the style classes with high confidence. CAN objective function is defined as follow (Equation 3.2.),

$$min_G max_D V(D,G) = \mathbb{E}_{x,\hat{c}\sim p_{data}}[\log D_r(x) + \log D_c(c=\hat{c}|x)] + \mathbb{E}_{z\sim p_z}[\log\Big(1 - D_r\big(G(z)\big)\Big) - \sum_{k=1}^{K}\Big(\frac{1}{K}\log\big(D_c\big(c_k|G(z)\big) + \Big(1-\frac{1}{K}\Big)\log\Big(1 - D_c\big(c_k|G(z)\big)\Big)\Big)\Big)] \quad (3.2)$$

where z is noise vector sampled from p z (uniformed or Gaussian distribution), x and c are real image and corresponding style label from data distribution p data . D r (.) is the function that tries to discriminate between real art images and generated images, while D c (.) Is the function that tries to classify different style categories and estimates style class posteriors?

Training process: The weights are initialized from zero-centered Normal Distribution with standard deviation 0.02. I used a mini-batch of size 64 and Adam [9] optimization algorithm for gradient descent, which has a learning rate of 0.002 and decay rates, ß1 = 0.6 and ß2 = 0.999 and LeakyReLU for as activation function for all hidden layers in the model and trained for 100 epochs. To stabilize training, I used Batch normalization [10] for all outputs in hidden layers and input images for discriminator. Dataset information: Wikiart dataset is used for training. It has total 103250 images and has 147 style classes [7].

Algorithm 1 CAN training algorithm with step size α, using mini-batch SGD for simplicity.

Input: mini-batch images x, matching label $\hat{c}$, number of training batch steps S
for $n = 1$ **to** S **do**
  $z \sim \mathcal{N}(0,1)^Z$ {Draw sample of random noise}
  $\hat{x} \leftarrow G(z)$ {Forward through generator}
  $s_D^r \leftarrow D_r(x)$ {real image, real/fake loss }
  $s_D^c \leftarrow D_c(\hat{c}|x)$ {real image, multi class loss}
  $s_G^f \leftarrow D_r(\hat{x})$ {fake image, real/fake loss}
  $s_G^c \leftarrow \sum_{k=1}^{K} \frac{1}{K} log(p(c_k|\hat{x}) + (1-\frac{1}{K})(log(p(c_k|\hat{x}))$ {fake image Entropy loss}
  $\mathcal{L}_D \leftarrow \log(s_D^r) + \log(s_D^c) + \log(1 - s_G^f)$
  $D \leftarrow D - \alpha\partial\mathcal{L}_D/\partial D$ {Update discriminator}
  $\mathcal{L}_G \leftarrow \log(s_G^f) - s_G^c$
  $G \leftarrow G - \alpha\partial\mathcal{L}_G/\partial G$ {Update generator}
end for

Figure 1. Step-by-step Procedure of CAN

3.2. Principles of NSTL

Neural style transfer (NSTL) is an optimization technique used to take three images, a content image, a style reference image (in this case – image generated by CAN), and the input image (RGB noise image is used in proposed model) — and blend them together such that the input image is transformed to look like the content image, but "painted" in the style of the style image.

It uses typical Deep Convolutional Neural Networks, usually trained on object detection and recognition purpose, to separate and recombine content and style of arbitrary images, proviNSTL calculates GefZgfZ by extracting content representations from both content image and input image. When Convolutional Neural Networks (CNN) are trained on object recognition, they develop a representation of the image that makes object information increasingly explicit along the processing hierarchy. Therefore, along the processing hierarchy of the network, the input images are transformed into representations that increasingly care about the actual content of the image instead of locations of specific features and actual pixel of the image. Higher layers in the network capture the high-level content in terms of objects. Therefore, content representations (also called feature vectors) in higher layers of the networks are used. NSTL calculates hZijg by extracting style features from style image and input image across multiplelayers of CNN. The more layers involved in extracting style representation of images, the better style features(textured, color, any features that can represent style of the image) will be obtained. NSTL then uses back propagation to minimize hZijg. ding a neural algorithm for the creation of artistic images. 16 convolutional and 5 pooling layers of the 19 layer VGG Network without fully connected layers are used.

In higher layer or deeper layer of CNN, detailed pixel information is lost while the high-level content of the image is preserved. The style representation computes correlations between the different features in different layers of the CNN. NSTL reconstruct the style of the input image from style representations built on different subsets of CNN layers [6].

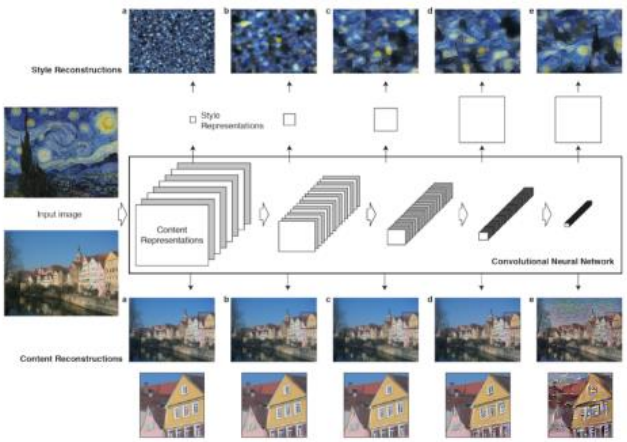


Figure 2. Content and Style Reconstruction of NSTL

4. PROPOSED SYSTEM DESIGN

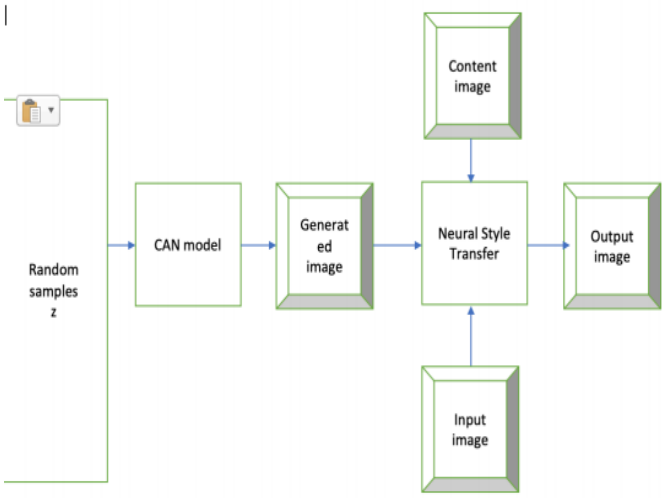


Figure 3. Proposed System Architecture

As shown in Fig 3.4, the proposed system first takes random sample from Normal Distribution (ranging from 0 to 1) and feeds into CAN model , as mentioned in section 3.1, with hyper-parameters also described in that section to generate new art style image. The proposed system then uses this style image and feeds it into Neural Style Transfer algorithm (NSTL) using VGG-network, which is also fed into content image from user and image from noise (Gaussian Distribution) in all three RGM channels. NSTL then extracts style representation from art style image, from CAN and content representation from content image from user and RGB noise image.

NSTL then extracts style representation from art style image, generated by CAN and content representation from content image, provided by human. It then transforms random noise image into an output image which looks like the content image painted in style of the style image. Hyper-parameters, α (content weight) and β (style weight), which are described in NSTL total cost function (Equation 3,9), are tuned as 106 and 102 respectively. It is done so to achieve proposed system's purpose of creating art. Much higher style weight β, (104 times higher than style weight used in standalone NSTL) is used because it wants to have image generated by CAN, style image in NSTL context, to overwhelm final output image, due to its creativity. Since, images from CAN are created by A.I, it does not have ability to percept objects like humans do, and its generated "art" images do not have clear sense of shapes or forms in it. For this reason, the proposed system used content image provided from human to inject a sense of forms and shapes to the system, using typically smaller content weight than standalone NSTL algorithm. We don't want content image to overwhelms final image, thus destroying A.I's creativity, either.

Since the proposed system is trying to integrate both A.I's creativity and human's creativity, it is crucial to give each of them a fair amount of involvement. Thus, the proposed system used random noise image as input image to NSTL instead of using content image or style image as input to NSTL which is typically seen in many standalone NSTL models to get rid of noise in output image. Due to the use of random noise image as input image, the generated "art" works contain noise in i.

5.IMPLEMETATION OF NSTL

The proposed system uses NSTL to extract content representation from images provided by humans and style representation from "art" style image generated by CAN. It then recombines them to produce an output image that looks like a content image painted in style image.

NSTL is trained on VGG-19 model [15 Kar] with its fully connected layers and output layers removed. Thus, VGG-19 used in NSTL is made up of five blocks of neural

networks. First two blocks have two convolutional layers and one max-pooling layer each. The remaining three blocks have four convolutional layers and one max-pooling layer each. The architecture of VGG-19 model is shown in following Fig 4.

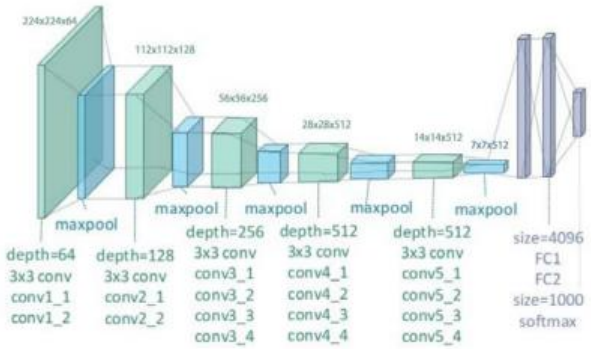


Figure 4. Architecture of NSTL

6. PERFORMANCE EVALUATION

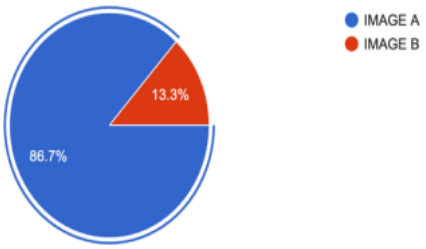
This experiment is to compare which image is more aesthetic and more likeable to human subjects. Human subjects are shown two images, image from CAN and image from proposed system, and asked to answer following questions. They are also asked to look at images at least 5 seconds before answering any questions.

Question #1 : Which art is more artistic ? The participant has to choose between two images Image A or Image B. This question is intended to compare aesthetic of two systems. In order to reduce bias towards any particular system, the survey showed two visually similar images to participants. First, sample image from CAN is obtained. As for the proposed system, the output image which obtained from feeding the same CAN image as style image and image from Fig 1.2 as content image. Following images are shown to human participants.

Figure 5. Image A: Generated by Proposed Model

Figure 6. Image B: Generated by CAN

Results are from 75 participants are collected summarized as following pie char in Figure 7.



7. CONCLUSION

By learning from existing art styles throughout the entire human history and generating new and creative art styles that have never seen before using CAN, the proposed system is proved to be creative. However, due to the machine inability to create typical figures or subject matters in CAN, the proposed system uses NSTL to allow humans to inject Figs, structures and shapes into final generated "art" work. By integrating NSTL with CAN also allows humans to have more control over generated "art" work. The survey responses from human participants also support the hypothesis the proposed system is based on, which is by injecting human familiar Figs and subject matters into "art" works generated by creative agents of A.I, these art works will be more aesthetic and more likeable by humans. Thus, one can say that the proposed system serves as the medium for collaboration of human's creativity and machine's creativity. It is also exciting to see what kinds of "art" works will be created when these two sources of creativity meet.

REFERENCES

[1] Himanshu Sharma. 2019. "Activation Functions : Sigmoid, ReLU, Leaky ReLU and Softmax basics for Neural Networks and Deep Learning.

[2] Tensorflow. 2018. "Neural Style Transfer: Creating Art with Deep Learning using tf.keras and eager execution".

[3] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, Yun Fu. 2018. "Residual Dense Network for Image Super-Resolution". arXiv 2018.

[4] Elgammal, Ahmed. 2017. "CAN: Creative Adversarial Networks, Generating "Art" by Learning About Styles and Deviating from Style Norms". On the eighth International Conference on Computational Creativity (ICCC).

[5] Kiri Nichol. 2016. "Wikiart Dataset". https://www.kaggle.com/c/painter-by-numbers.

[6] Diederik P. Kingma, Jimmy Lei Ba. 2015. "ADAM : A METHOD FOR STOCHASTIC OPTIMIZATION". In International Conference on Learning Representations 2015.

[7] Karen Simonyan, Andrew Zisserman. 2015. "VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION". arXiv 2015.