Manumorph: An Implementation of Deep Painterly Harmonization

David Abramov
Computational Media
University of California, Santa Cruz
dabramov@ucsc.edu

Sarah Frost
Computational Media
University of California, Santa Cruz
smfrost@ucsc.edu

Manu Mathew Thomas
Computational Media
University of California, Santa Cruz
mthomas6@ucsc.edu

Figure 1: Jacopo Pontormo's "Leda and the Swan" harmonized with a photo of Buzz Aldrin on the Moon.

ABSTRACT

Style transfer, the technique by which the style of one image is applied to the content of another, is one of the most popular and well-known uses of neural network algorithms. Deep Painterly Harmonization is an extension of style transfer, and includes a content object which is placed on the style image. The network then harmonizes the style and the content. We build on Deep Painterly Harmonization, originally implemented in Torch, and re-implement the paper's algorithm in Tensorflow. We extend the uses of the algorithm to explore different categories of visual media modification. We discuss the ramifications of style harmonization and style transfer on societal concepts of art, and we compare the results of the Tensorflow and Torch algorithms. Finally, we propose a design for a web application that will allow casual creators without a strong technical background to create new art using the algorithm. The repository for this project can be found at: https://github.com/sarahmfrost/manumorph.

Conference'19, July 2019, Copenhagen, Denmark

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM…$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

CCS CONCEPTS

• **Applied computing** → **Fine arts**; • **Computing methodologies** → *Image processing*.

KEYWORDS

neural networks, style transfer, memes

ACM Reference Format:
David Abramov, Sarah Frost, and Manu Mathew Thomas. 2019. Manumorph: An Implementation of Deep Painterly Harmonization. In *Proceedings of ACM Conference (Conference'19).* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

As artificial intelligence improves, it is increasingly possible to fool humans into thinking that computer-generated images are real. No longer can we believe everything we see. Deep fake videos [5], expertly rendered images [1], twitter bots [10] and artificial faces [8] are examples of new digital media that can fool people into thinking that images are real when they are not. While neural networks allow for novel ways to synthesize imagery that can be uncanny and malicious, there is interest in extending these algorithms for artists and content creators.

We build on the work of Gatys and Luan [4] [9] to combine the content of one image with the style and content of another image using convolutional neural networks. This project combines style transfer and content addition. The deep painterly harmonization algorithm was originally implemented in Torch, an optimized tensor library for deep learning using GPUs and CPUs [9]. We implement this project in Tensorflow, an open source library for numerical computation developed by Google [3]. Re-implementing this project will allow others to use this technology for their own creative endeavors.

We explore the use of neural networks for inserting new content into paintings and photos in the style of the original media. Our altered media is designed to amuse, provoke, and encourage casual creators to make art. We tested the results of the algorithm on style transfer with photographs. We explore the ability of this technology to convince people that our altered images are real, and offer this method as a way to augment the process of visual media creation.

Figure 2: A photograph of Sarah Frost's face harmonized with "Jane Fleming, later Countess of Harrington" by Joshua Reynolds.

2 RELATED WORK

Non-photorealistic rendering emerged in the 1990's at the intersection of art and computer graphics. A number of artistic rendering algorithms were developed to convert a two-dimensional image into a distinct artistic style. While some of these models can create compelling results, there are limitations in that a single model is limited to recreating a single style and can not be generalized. For example, there are stroke-based rendering algorithms that can recreate an image in the style of an oil or watercolor painting, but each painting style requires a different finely tuned algorithm. More recently, neural networks have been shown to be a powerful tool for style transfer that can be generalized to multiple artistic styles. Style transfer algorithms have been used, modified, and improved in interesting ways. For instance, there is the ability to control the brush stroke size when applying style transfer, semantic matching between content and style of images of similar objects, and audio style transfer for the synthesis of new sounds [7].

Deep Neural Networks for image recognition have been designed based on a model of how human's process visual stimulus. Gatys et al. present a style transfer algorithm using convolutional neural networks [4]. This style transfer algorithm is able to distinguish between content and style. The lower layers of the network reconstruct the pixel level of the original image, whereas in the higher layers more complex features emerge as content representation. Style feature spaces emerge simultaneously, which recognize multi-scale structures present in an image as style representation. Content and style are separable concepts in Convolutional Neural Networks [4].

Luan et al. extend the work of Gatys in the paper "Deep Painterly Harmonization" [9]. They develop an algorithm for determining the style aspects that will be transferred. Their goal is to make the addition of photographic content inconspicuous, to the point that the added content is not noticeable. They use a two pass method which allows them to begin with coarse harmonization, and in the second focus on fine visual quality. After the two passes are complete, they enact a two-step signal-processing approach - with chrominance denoising and patch synthesis. These methods improve the texture and smooth details. After running this algorithm, they conducted two user surveys that sought to quantify the quality of the results. The first user study investigated whether or not the harmonized element could be identified by a viewer. The second user study compared the quality of the Deep Painterly Harmonization algorithm to similar algorithms [9].

Figure 3: Style subtraction as demonstrated with a poster of the TV series "Friends".

3 DATASET

Luan et al. [9] exclusively use paintings as the source images and photos as content elements. Here we extend their method to harmonize a part of a painting with the style of a photo. We call this deep photoly harmonization. We also tested media in which multiple faces in photos were added to a painting, as well as adding a painting to a painting in a photo. To create inputs for our system, we used an online photo editing website, Photopea, and freepng to get images with transparent backgrounds. The objects are added to the source painting or photos to create the input image and then we use the added image layer function to create the mask image. The transparent background is filled with black.

4 METHODS

We designed a two-pass algorithm based on "Deep Painterly Harmonization" [9]. For each pass, a content image, a style image, and an intermediate image are used as inputs in the network, VGG 16. VGG 16 is a pre-trained image classifier network developed by a visual geometry group at Oxford [11]. This model is trained on an imagenet dataset and was the winner of imagenet competition in 2014. We use this network as a feature extractor for our style transfer application by removing the fully connected part of the network. There is no training of network involved in either of the passes. The optimization happens directly on the output image. The first pass gives us a coarse style transfer and the second pass harmonizes the output from the first pass. The style image is the target painting or photo and the content image consists of a component of a painting or a photo to be added to the style image. Unlike other style transfer methods we also use the mask image of the added object. The mask image is used to compute the loss functions only for that part of the image.

Pass 1

The content, style and mask images are passed to VGG 16 for 100 iterations. During each iteration we get the activation map from layer 1, 2, 3, 4 and 5 to compute the style loss. The lower layers identify abstract features, including style. Gram metric is calculated for the activation map from each layer and mean square error is computed for output and the style image. Gram metric is a spatially invariant metric used to extract feature distribution of images. The style losses are combined and weighted with a style loss constant. For the content loss, we compute the mean square error of features from the fifth layer of the output and content image. The high level features are extracted by the higher, or subsequent, convolutional layers in the network. For both losses, we pass the mask image through the network and multiply the activation of the mask image with the content and style activations. This ensures that we are only applying the loss around region of the added object. The combined loss function is shown below.

Pass 1: Loss function.

$$L = Wc * \text{Content Loss} + Ws * \text{Style Loss} \tag{1}$$

$$\text{Content Loss} = MSE(f(C) - f(G)) \tag{2}$$

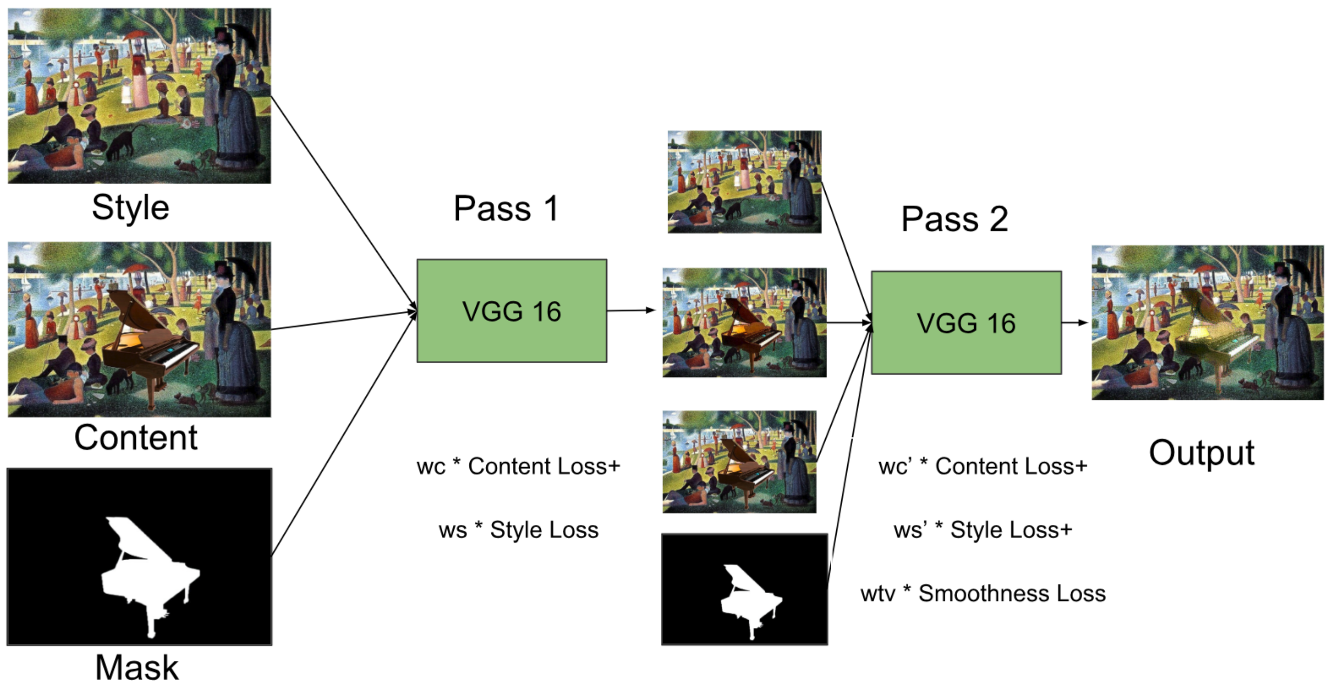

Figure 4: A diagram of our system architecture.

$$\text{Style Loss} = MSE(f(g(S)) - f(g(G)) \tag{3}$$

$$\text{Final Output} = \text{Blurred Mask} * \text{Output} + (1 - \text{Blurred Mask}) * \text{Style Picture} \tag{4}$$

Where Wc and Ws are weights associated with content and style loss. In our first pass, we have a higher content weight $Wc = 100$ and $Ws = 0.1$.

Pass 2

We pass the content image, style image, and the output from the first pass to the pre-trained VGG 16 network. We focus more on the style by putting a higher weight on the style loss. The content loss ensures that we are not losing any content details while transferring style. In addition to content and style loss, we use a smoothness loss to regularize colors in the adjacent pixels. We run this pass for 20 iterations. The combined loss function is shown below:

Pass 2: Loss function.

$$L = Wc' * \text{Content Loss} + Ws * \text{Style Loss} + Wtv * \text{Smoothness loss} \tag{5}$$

$$\text{Smoothness Loss} = \sum x, y[2(Ox, y - Ox - 1, y) + 2(x, y - Ox, y - 1)] \tag{6}$$

Where $Wc' = 1$ and $Ws' = 100$ and $Wtv = 0.5$.

5 RESULTS

We wanted to combine interesting content and style, make controversial new media, and push the harmonization process in new directions. Luan et al. use paintings as the style, with added content from photos [9]. We explored the reverse, a photo as style with added content from a painting, or deep photoly harmonization (fig. 1). We isolated the body of Leda in Jacopo Pontormo's "Leda and the Swan" and harmonized that with a photo of Buzz Aldrin on the Moon. We also harmonized two political works. One inserts the faces of Donald Trump, Kim Jong-un, and Vladimir Putin in Leonardo da Vinci's "Last Supper". In other image we combined the faces of Brett Kavanaugh and Mitch McConnell with a cartoon image of "The Frog and the Toad" (fig. 4). We removed content from a poster of the characters of the television series "Friends" and then passed that in as our style and content images (fig. 3). For our amusement, we added a photo of the face of one of the authors of this paper, David Abramov, to a movie poster of the 1989 film, "The Brave Little Toaster" (fig. 6). Manu Mathew Thomas was added as an apostle in Fra Angelico's "St. Peter Consecrates Stephen as Deacon" in a process the other authors of this paper are calling, "Deep Painterly Har-manu-ization" (fig. 6). Finally, Sarah Frost's face was added as the content to "Jane Fleming, later Countess of Harrington" by Joshua Reynolds (fig. 2).

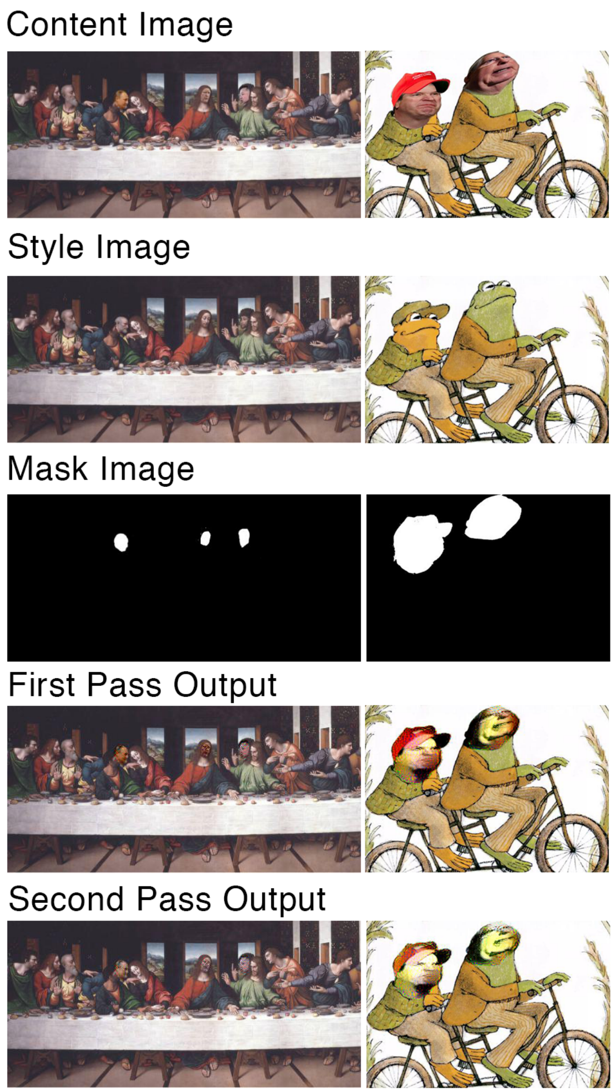

Figure 5: Two harmonized images with political content.

Feedback on created media

We demonstrated our work to a group of Computational Media graduate students at the University of California, Santa Cruz. We asked them if they would use this process to create art. Their response was generally positive, with one student saying, "[the process] could be used for a contemporary collage - with portions of famous artwork." Another student remarked that this process would allow her to "take something that's already been created and tweak it to make it my own." Much of the media we created was meant to appear harmonized, and draw attention to the added element. Therefore, we expected our viewers to easily identify the content we had added to the background piece. However, we asked the students to identify the almond blossom sections that we added to van Gogh's "Almond Blossoms" (fig. 6). Most were unable to identify the added branches after observing the painting for several seconds. The Torch and Tensorflow implementations follow the algorithm in "Deep Painterly Harmonization" [9] but the Torch version includes a mapping step that takes the spatially invariant style loss and improves the blending of the content image with the style behind it. Our Tensorflow implementation does not have this mapping step. This produces an abstracted style applied to the added content, which might be desirable to certain users.

6 DISCUSSION

This paper is motivated by our fascination with style transfer. Both style transfer and deep painterly harmonization effect how we view visual art. Many people view famous paintings as static and unchanging. Deep Painterly Harmonization allows us to re-conceptualize this art as changeable and more relevant to current events and popular culture. Our remix of da Vinci's Last Supper with Donald Trump as Jesus and Kim Jong-un and Vladimir Putin as apostles (fig. 5) is meant to draw attention to our current political climate and the adoration that these leaders command.

Similarly, this project has implications for its use as an artistic device. There is an interesting philosophical argument with regards to who should be considered the 'artist' when someone uses a neural network designed by someone else to produce and sell artwork. In December 2018, a piece of art created by an algorithm sold for $432,500 at Christie's. [2] The piece, "Portrait of Edmond Belamy" was created with a generative adversarial network that had been fed 15,000 portraits from the 14th to 20th century. The sale of this piece prompted discussion about whether or not algorithms can be credited with creating art. Hertzmann argues that algorithms are tools which will allow artists new pathways to create new art [6]. This project is undeniably a new pathway to create art.

7 FUTURE WORK

We would like to continue to experiment and test the boundaries of media creation using our Tensorflow implementation. We hope to test the algorithm with a black and white painting for style, while the content style is in color. We also would like continue investigating style subtraction and see if we can achieve more compelling results. This is an implementation of the algorithm in which a section of the style image is removed, and nothing is given to replace it. We would like to

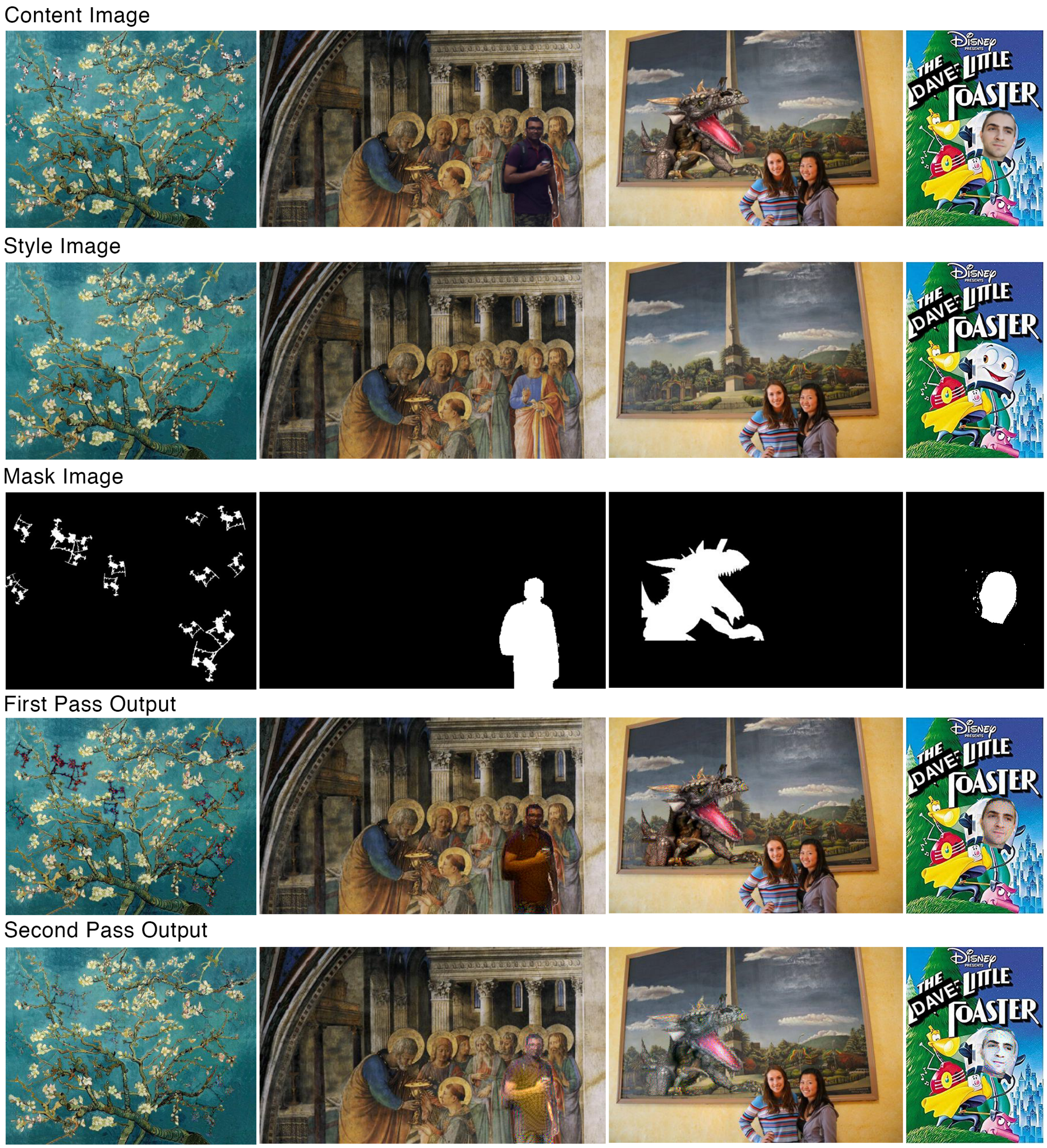

Figure 6: Examples of the media generated with our Tensorflow implementation of the Deep Painterly Harmonization algorithm.

investigate how the algorithm harmonizes the empty space with the style of the surrounding image.

In addition, we hope to build a web app that will allow users to upload a piece of art with style and content and harmonize the media with ease. This will allow casual creators to make media. Similar apps exist for style transfer (deepart.io, algorithmia, pikazo) but a style harmonization app does not currently exist. This web app would allow users to engage with the Deep Painterly Harmonization algorithm even if they lack the processing power of GPUs or the knowledge of recurrent neural networks. We have an initial idea of how we can implement this app. We will use openCV to automatically generate the mask image based on the content that is added. We would like to give users a way to tweak the combined loss function (the content and style loss) and the histogram loss function to allow the user to alter how much of the content style is retained in the final image.

REFERENCES

[1] [n. d.]. Chaos Group. https://www.chaosgroup.com/gallery

[2] 2018. Is artificial intelligence set to become art's next medium? https://www.christies.com/features/A-collaboration-between-two-artists-one-human-one-a-machine-9332-1.aspx

[3] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, et al. 2016. Tensorflow: A system for large-scale machine learning. In *12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16)*. 265–283.

[4] Leon A Gatys, Alexander S Ecker, and Matthias Bethge. 2015. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576* (2015).

[5] David Güera and Edward J Delp. 2018. Deepfake video detection using recurrent neural networks. In *2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*. IEEE, 1–6.

[6] Aaron Hertzmann. 2018. Can Computers Create Art?. In *Arts*, Vol. 7. Multidisciplinary Digital Publishing Institute, 18.

[7] Yongcheng Jing, Yezhou Yang, Zunlei Feng, Jingwen Ye, Yizhou Yu, and Mingli Song. 2017. Neural style transfer: A review. *arXiv preprint arXiv:1705.04058* (2017).

[8] Tero Karras, Samuli Laine, and Timo Aila. 2018. A style-based generator architecture for generative adversarial networks. *arXiv preprint arXiv:1812.04948* (2018).

[9] Fujun Luan, Sylvain Paris, Eli Shechtman, and Kavita Bala. 2018. Deep painterly harmonization. In *Computer Graphics Forum*, Vol. 37. Wiley Online Library, 95–106.

[10] Bjarke Mønsted, Piotr Sapieżyński, Emilio Ferrara, and Sune Lehmann. 2017. Evidence of complex contagion of information in social media: An experiment using Twitter bots. *PloS one* 12, 9 (2017), e0184148.

[11] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).