Multiple Styles Transfer Emphasizing Visual Saliency

Suho Cho
Dept. Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology (GIST)
Gwangju, Republic of Korea
thecho7@gist.ac.kr

Kuk-jin Yoon
Dept. Electrical Engineering and Computer Science
Gwangju Institute of Science and Technology (GIST)
Gwangju, Republic of Korea
kjyoon@gist.ac.kr

***Abstract*— An image style transfer is known as an artistic work with computer vision. Conventional approaches in the style transfer have focused on texture extraction and transfer. These approaches developed a successful style transfer but not the content preservation. Damage occurs due to the style transfer of the entire image. When we try to do the style transfer strongly, the contents which have to be preserved are distorted and spoiled. Here we suggest the multiple styles transfer to preserve the content which we aim to do that. The algorithm contains three main techniques – Style transfer with deep learning, Visual saliency and Semantic segmentation. With these computer vision techniques, we apply the multiple styles into a single image successfully. Our work provides the improvement of visual saliency and the better artistic work with the style transfer. Furthermore, it can contribute new sight with visual saliency in philosophy.**

Keywords—Style Transfer, Visual Saliency, Image Segmentation, Deep Learning

I. INTRODUCTION

An intriguing study which is based on Convolutional Neural Network has been released [1]. It shows a possibility that a deep learning can collaborate with artistic works. They named it 'A Neural Algorithm of Artistic Style' and we abbreviate it as 'Neural Style'. This algorithm makes the resulting image more natural than the other method [2, 3, 4] by utilizing broad level of image features. Despite it achieves great results for certain stylistic transfers, there is a limitation to the current method as the style transfer can be applied to the whole image only. As shown in Figure 1, it looks like the style is transferred successfully. However, people in the image are excessively distorted so that we cannot preserve the original contents. Furthermore, saliency of the image would be removed.

Figure 1. Original style transfer. A style of 'The Starry Night – Vincent Van Gogh (1889)' is applied to the content image.

In our work, we emphasize the visual saliency while applying the multiple styles in a single image. The brightness, texture gradient and style weight control help saliency more clear. It shows multiple artistic works without disharmony in a single image.

II. BACKGROUND

A. Neural Style

To separate the style of the image, Neural Style [1] use a convolutional neural network. The model they used is 19-layer VGG network which contains 16 convolutional layers and 5 pooling layers. The VGG-19 network extracts the feature representation and the neural network they made transfers it.

B. Visual Saliency

Within our line of sight, there are always some stuffs that stand out more than the others. The point we strongly gaze is a visual salient point and we call this property as visual saliency. The reason that a particular stimulus has such salience may be due to visual contrast.

Some works focus on the objectness to obtain saliency in the image. Chang et al. [5] combines low-level features and the objectness measurement. Similar works [6, 7, 8] have been implemented by objectness-based object proposals to estimate object-level saliency.

C. Semantic Segmentation

In computer vision, image segmentation is the process of partitioning an image into multiple segments. The goal of segmentation is to simplify and change the representation of an image into something that is more meaningful and easier to analyze. It is typically used to find the objects and boundaries in image.

Especially, semantic segmentation attempts to partition the image into semantically meaningful parts, and classify each part into one of the pre-determined classes. Mottaghi et al. [9] proposed a deformable part-based model which exploits both local context around each candidate detection as well as global context at the level of the scene. It provides 520 plenty of additional classes for semantic segmentation.

III. METHOD

As mentioned from the above, our method needs three kinds of techniques. To applying the style transfer partially, it is necessary to confirm which segment we have to choose. Using a semantic segmentation is a good choice to acquire the result without feeling awkward. Since doing semantic segmentation is kind a hard work, we use a semantic segmentation dataset which is already well segmented. A total segmented image is strongly recommended because saliency is not limited in certain object even though we estimate saliency by using objectness. The study [9] provides well fully-segmented images and its annotations of PASCAL VOC 2010.

Our work uses saliency map to obtain the clear regions of saliency. To obtain the saliency map, we generate the object proposals by utilizing the concept of objectness [5]. The object proposals have different size and objectness score. The scores are marked along the edges of the boxes and cumulated. We calculate and plot the saliency score map for several samples.

An intensity is marked on every pixel in the image. After that, we compute the average saliency score of every segment.

$$\Phi = \frac{\sum saliency\ score\ in\ segment}{area\ of\ segment} \tag{1}$$

We conclude that the most salient segment has the maximum Φ among the segments in an image. We denote Φ^* as a segment we select.

$$\Phi^* = segment|\max(\Phi) \tag{2}$$

To highlight saliency, we perform the multiple styles transfer into a single image. Since the difference of textures can bring saliency, the style difference between salient region and non-salient region can make saliency strong. We copy the content image and the number of the copy is the same as the number of styles we try to apply. After that, we apply each of styles to each content copy, respectively. Then, combine all the copies we made into a single resulting image. In the combine process, we paste only the segments we chose from the different style images. As a result, it is possible to apply the styles to the salient segment and the background respectively. We present our procedure in figure 2.

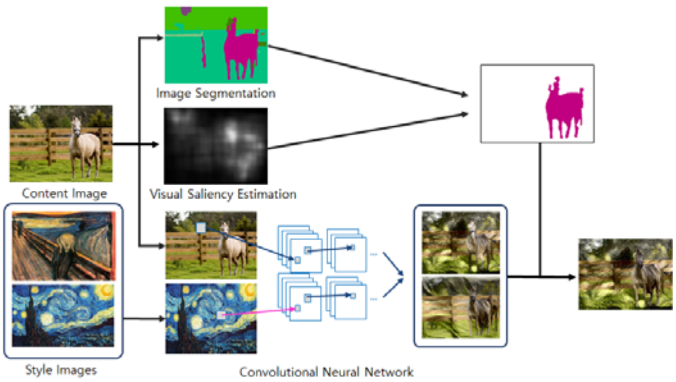

Figure 2. Entire procedure for our proposal. The output image has n number of styles according to saliency.

IV. Result

In figure 3, We did the multiple styles transfer to the salient object and the background with different style images respectively. Only the objects are selected as saliency, and these objects are highlighted by the style transfer.

Figure 3. The resulting images of the multiple styles transfer and Neural Style. According to saliency, the chosen segments are painted in different style to the background style (Ours).

From the comparison between our work and Neural Style, it is clear that we achieve transferring multiple styles in a single image successfully. In addition, by selecting proper style weights, we preserve the texture of salient contents while the style is applied to the background strongly. The objects in the resulting images are pretty clear to recognize what is it. Even though the contour of the images is cut by force, the resulting images look natural around the boundary of the objects.

Since what we aim to do is to emphasize saliency, we compare the saliency map result between the original image and the output image.

Figure 4. Comparison result of saliency maps between ours and the original images. Saliency is focused on the objects. The left figures of the saliency maps are the results of our method. The right figures are the saliency maps of the original images.

From figure 4, we can obtain the improved saliency maps. A redundant region of the saliency map is clearly diminished and the salient region which is focused on the object is became brighter than the original image.

The multiple styles transfer makes a discontinuous boundary between the segments. It causes a sharp texture change and makes saliency stronger. Also, since the style transfer carries not only the texture of the style image but also the brightness of the style image, our method gives a clear brightness contrast of the segment boundary. Comparing to the single style transfer, the multiple style transfer provides a distinct saliency. In fact, the saliency score in the salient region is actually increased. We calculate the proportion of saliency score between the salient region and the background.

$$\gamma = \frac{\sum score\ in\ the\ chosen\ segment}{\sum score\ in\ the\ background} \times 100(\%) \tag{3}$$

We test the actual change of the saliency score (3). If saliency of the chosen segment is increased after the multiple style transfer, $\Delta\gamma$ ($=\gamma_{ours} - \gamma_{original}$) should be a positive value. Table 1 is the saliency calculation result of the figures in Figure 4.

TABLE I. SALIENCY SCORE CALCULATION AND GAMMA COMPARISON BETWEEN OURS AND NEURAL STYLE.

	Original	Ours	The Starry Night	Femme nue assise	The shipwreck of the Minotaur
a	38.9	**41.6**	38.1	38.9	40.3
b	18.9	**19.0**	18.4	17.4	17.4

According to table 1, our method is far better than the single style transfer in many styles. Most of results show the clear saliency boundary with less noise. Furthermore, it is important to select a proper set of the style images. If the two style images are similar to those texture and brightness, the multiple style transfer cannot be such an effective method to emphasize saliency. According to the saliency score, we also calculate γ to examine whether saliency has been focused or not. The tested images are the images in figure 4.

V. DISCUSSION

In the case of 336×500 pixels image, the time spent for transferring the styles is about 4 minutes where the procedure is implemented on a Nvidia Titan X. But when we try to perform on the bigger size images, the time spent will increase considerably. Moreover, the loss value could be diverged in the beginning of procedure. In deep learning, a loss value which exceeds a certain threshold cannot be calculated. After that, training and learning do not work anymore. Therefore, the size of an input image is matter. If the size of an input image is too large, the style transfer can be halted during the procedure.

Meanwhile, we plan to extract the value in Gram matrix to compare the similarity between the styles. When we do the multiple styles transfer, it is important to choose a proper set of styles. The different combination of styles gives the different feeling and saliency. Since Gram matrix can represent the texture and brightness of the image, comparing Gram matrices of the style images could be an effective approach.

VI. CONCLUSION

This paper proposes a novel approach to emphasize saliency of an image with the style transfer. We implemented our work by image segmentation, saliency detection and the style transfer. These three works make the clear style transfer result and the better saliency result. According to our work, the style transfer is not only performed for fun but also gives improvement both in philosophy and computer vision. In artistic respect, it shows a novel work of the style transfer in general image. While it preserves the texture of the salient content, it exhibits the multiple styles at once. It can be expected to contribute to the flourish of artistic exhibition. In philosophy, the improved saliency can contribute to the analysis of human mind. Additionally, the multiple styles in an image may hide a potential function. Finally, in computer vision, we combine the different techniques to solve a problem in the former research, while improving a visual saliency. Furthermore, since it uses a convolutional neural network, the entire procedure can be done automatically.

ACKNOWLEDGMENT

This work is supported by Computer Vision Laboratory, Gwangju Institute of Science and Technology (GIST).

REFERENCES

[1] Gatys, L. A., Ecker, A. S., & Bethge, M. (2016). Image Style Transfer using Convolutional Neural Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 2414-2423).

[2] N. Ashikhmin. Fast texture transfer. IEEE Computer Graphics and Applications, 23(4):38–43, July 2003.

[3] A. Efros and T. K. Leung. Texture synthesis by nonparametric sampling. In Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on, volume 2, pages 1033–1038. IEEE, 1999.

[4] A. A. Efros and W. T. Freeman. Image quilting for texture synthesis and transfer. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pages 341–346. ACM, 2001.

[5] Chang, K. Y., Liu, T. L., Chen, H. T., & Lai, S. H. (2011, November). Fusing Generic Objectness and Visual Saliency for Salient Object Detection. In 2011 International Conference on Computer Vision (pp. 914-921). IEEE.

[6] Alexe, B., Deselaers, T., & Ferrari, V. (2012). Measuring the Objectness of Image Windows. IEEE Transactions on Pattern Analysis and Machine Intelligence, 34(11), 2189-2202.

[7] Cheng, M. M., Zhang, Z., Lin, W. Y., & Torr, P. (2014). BING: Binarized Normed Gradients for Objectness Estimation at 300fps. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 3286-3293).

[8] Srivatsa, R. S., & Babu, R. V. (2015, September). Salient Object Detection via Objectness Measure. In Image Processing (ICIP), 2015 IEEE International Conference on (pp. 4481-4485). IEEE.

[9] Mottaghi, R., Chen, X., Liu, X., Cho, N. G., Lee, S. W., Fidler, S., ... & Yuille, A. (2014). The Role of Context for Object Detection and Semantic Segmentation in the Wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 891-898).