# An Interactive System for Painterly Image Harmonization


Juwon Lee
*Content Research Division*
*ETRI*
Daejeon, Rep. of Korea
zacurr@etri.re.kr

Jung-Jae Yu
*Content Research Division*
*ETRI*
Daejeon, Rep. of Korea
jungjae@etri.re.kr

Wonyoung Yoo
*Content Research Division*
*ETRI*
Daejeon, Rep. of Korea
zero2@etri.re.kr


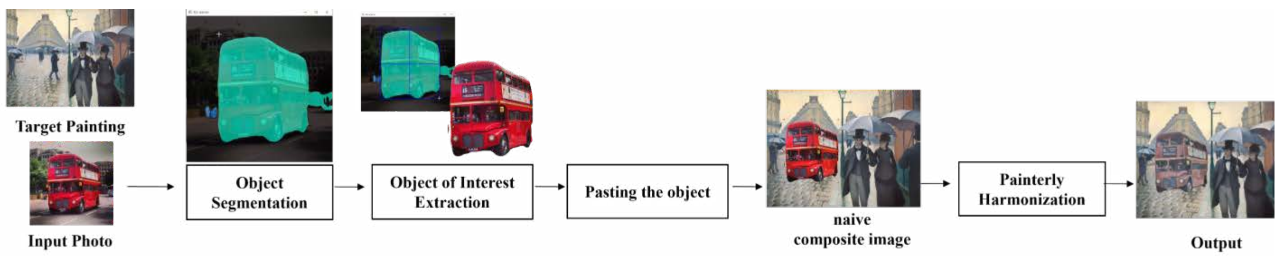


Fig. 1. The proposed system for painterly image harmonization


***Abstract*— Image harmonization is the process of adjusting the appearance of an object of interest that is copied from a source image and pasted onto a target image to make the composite image look seamless and natural. Painterly image harmonization represents a distinct category within image harmonization where the target image has a painterly style. To obtain the harmonized results, it involves several sub-tasks: segmenting the object of interest from a photograph, creating a naive composition that includes placing and scaling the object onto the target painting, generating a mask for the pasted object, and applying painterly image harmonization to the naive composite image. To facilitate these processes within a unified framework, we integrated object segmentation and painterly image harmonization methods, with a user-friendly graphical user interface. The proposed system enables users to easily create new artworks by placing objects from photos into paintings and harmonizing them according to their intention. In addition, we designed the harmonization pipeline employing a style transfer method and compared it with the conventional harmonization method, E2STN. Qualitative results show the effectiveness of our proposed approach.**



***Keywords—painterly harmonization, style harmonization, image segmentation, image composition, image manipulation***


## I. INTRODUCTION

Artistic style transfer [1-4] has gained a lot of popularity because it can generate novel and diverse artworks from common photos by applying the style of a painting to the content of a photo. While it transforms the entire photo into artworks, painterly image harmonization [5-9] aims to create a seamless and natural composite image by blending an object of interest in photo with painting in a harmonious way. However, the method requires a naive composite image as an input, which is hard to generate because it involves object segmentation and object placement onto the painting. To address this, we introduce an integrated framework that can easily create a naive composite image and apply a harmonization method to it by combining object segmentation and painterly image harmonization methods with a user-friendly GUI. This enables users to easily create artistic images through simple interactions, without requiring any prior knowledge or skills in image editing, and to swiftly obtain diverse results.

In addition, we designed the simple harmonization pipeline based on the fast arbitrary style transfer [4] and compared it with the tailored model for painterly image harmonization, E2STN (Element-Embedded Style Transfer Network) [7]. With the developed system, we conducted experiments on a variety of examples to validate the effectiveness of our method.

## II. RELATED WORK

### *A. Style Transfer*

Artistic style transfer aims to apply the visual characteristics and stylistic elements of a target style image, typically a painting, to a content image, typically a photograph. This process produces a new image that combines the content of the original photo with the artistic style of the reference painting. A pioneering work [1] in this field used an iterative optimization approach to adjust the content image, matching its style with that of the style image. The feed-forward methods [2-4] were developed to overcome the computational complexity of the iterative methods. Among them, AdaIN [3] and [4] enable real-time arbitrary style transfer.

Since these methods are designed for global stylization, it is known that the style transfer models produce unsatisfactory results [7-9] on the painterly image harmonization task.

### *B. Painterly Image Harmonization*

Painterly image harmonization [5-9], also referred to as style harmonization, is a more challenging task than photo-realistic image harmonization [10-12] because it involves composing images from two different domains (i.e., photo and painting) and it requires both the low-level attribute adjustment (e.g., brightness, contrast and illumination) and the high-level artistic style transfer such as brush strokes, color palette and texture. Despite the task's complexity, the landscape of related research remains relatively sparse. DPH (Deep Painterly Harmonization) [5] introduced a two-pass algorithm that ensures both spatial and inter-scale statistical consistency by mapping neural responses. However, this

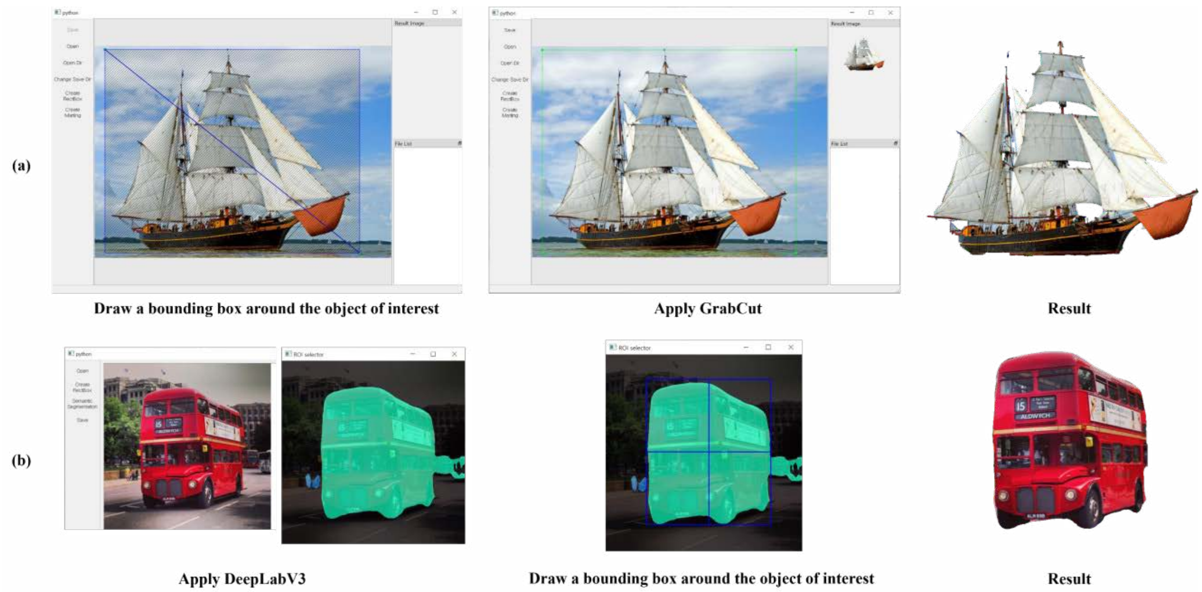


Fig. 2. Object of interest extraction using (a) GrabCut and (b) DeepLabV3

approach is time-consuming due to its iterative optimization process. E2STN [7] proposed a GAN-based feed-forward method. It employs a pre-trained VGG-16 as an encoder, extracting feature maps from both the input photo and the target painting. These two features are merged using the corresponding object mask obtained from the image matting module, producing a composite feature. Subsequently, the harmonized output is generated from this feature through the channel-aligned decoder. More recent advancements include PHDNet [8], which leverages a dual-domain model to better capture global style and periodic textures, and PHDiffusion [9], based on the stable diffusion model.

## III. Proposed System

In this section, we present our proposed system for interactive painterly image harmonization.

Fig. 1 illustrates the block diagram of the proposed system. Firstly, a user selects an input photo depicting an object of interest, and a target painting where this object will be pasted for image composition. Employing image segmentation methods and user interactions as shown in Fig. 2, the object of interest is extracted. Subsequently, the user positions and scales the extracted object onto the painting to align with their preferences. This yields the naive composite image, serving as the input for subsequent painterly image harmonization. Finally, the output image is generated through the style harmonization module. The details will be explained in the following sub-sections.

### A. *Object of Interest Extraction*

For image segmentation, either the semantic segmentation model, DeepLabV3 [13] or the foreground extraction model, GrabCut [14] was employed. There have been cases where DeepLabV3 produces the better results when GrabCut fails, and vice versa as shown in Fig. 3. Therefore, two different modes of segmentation were considered. To get better results, both methods can be used. As shown in Fig. 4, the result of DeepLabV3 can be refined using GrabCut.

*1) GrabCut:* Fig. 2 (a) illustrates the procedure of extracting the object of interest using GrabCut. First, a user

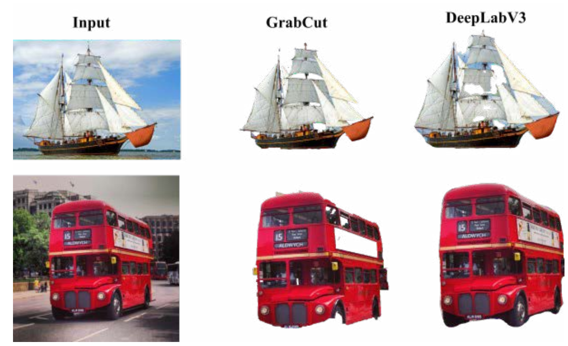


Fig. 3. Examplar results of GrabCut and DeepLabV3

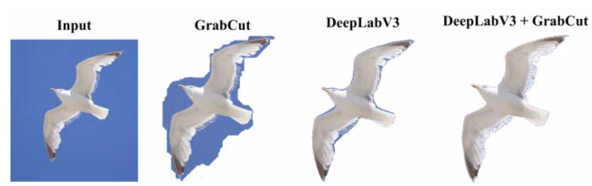


Fig. 4. An example of the refined result when both DeepLabV3 and GrabCut were used

draws a bounding box around the object of interest, and applys GrabCut for the image patch defined by the bounding box to get the result.

*2) DeepLabV3:* In contrast, a user applys DeepLabV3 first, and draws a bounding box around the object of interest to crop it as shown in Fig. 2 (b).

### B. *Object Pasting*

To obtain a naive composite image, which will be input to the style harmonization module, a user pastes the segmented object of interest onto the target painting. The location and size of the object can be adjusted corresponding to a user's preference. The user-friendly GUI enables users to easily place multiple objects as shown in Fig. 5.


Fig. 5. Naive Image Composition

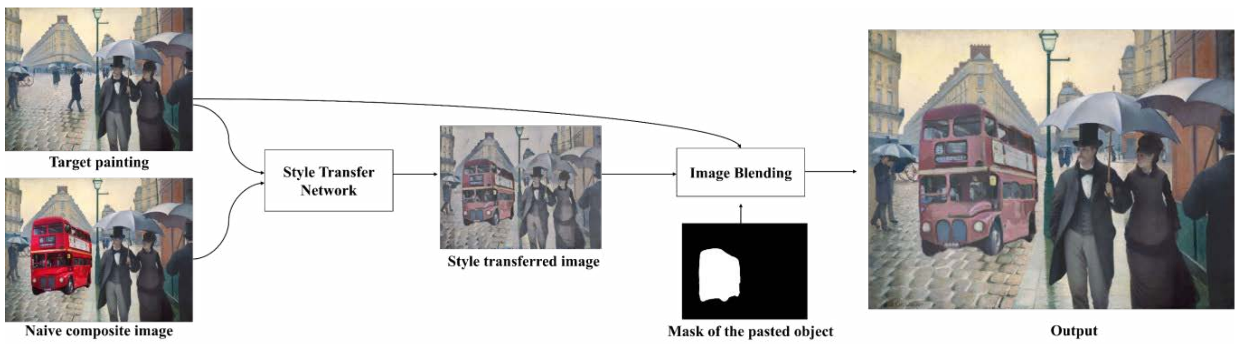


Fig. 6. Painterly Image harmonization procedure based on the style transfer model

### C. *Painterly Image Harmonization*

We designed a simple harmonization pipeline employing fast arbitrary style transfer model [4] as depicted in Fig. 6. The naive composite image is used as the content image, while the target painting serving as the style image. These are fed into the style transfer network to obtain the stylized output. We assumed that the style transfer model can focus on the object of interest because the large portion of the composite image overlaps with the style image. This is the reason why the composite image rather than the original input photo was served as content image. The mask of the pasted object was obtained by the difference of the naive composite image and the target painting. Using this mask, the stylized object is extracted and blended with the target painting, which produces the final style harmonized image. We tried the seamless cloning (i.e. Poisson image editing [15]) in image blending step, but the visual quality of results was degraded.

For the other configuration, the conventional painterly image harmonization model, E2STN [7] was utilized. We replaced the image matting module of E2STN with our object segmentation module. When extracting a composite feature, a naive composite image was used instead of input photo, as it is easier to extract the feature of a location and scale-adjusted object.

## IV. Experiments

We tested the proposed system on various paintings and photo pairs. The implementation details are as follows. For semantic segmentation, We used DeepLabV3 [13] with a ResNet-50 backbone, trained on a subset of COCO using Pascal VOC 20 classes. We utilized the OpenCV implementation for the GrabCut [14] algorithm. AST (Arbitrary Style Transfer) v2 [16] of Tensorflow-hub was used for style transfer model. In the case of the alternative configuration, we implemented E2STN and trained the model using COCO (for input photos and the corresponding masks) and WikiArt [17] (for target paintings) dataset. To facilitate user interactions, GUI was implemented using PyQT 5.

The results obtained from our proposed system are presented in Fig. 7. The user-friendly GUI enabled the creation of a variety of composite images with ease. In most cases, the results with AST well preserved object details while maintaining style consistency. However, E2STN tended to accentuate stylization, often leading to the loss of finer object details. E2STN exhibited difficulties in transferring the appropriate style, particularly in instances involving small pasted objects.

Contrary to conventional knowledge, the style transfer model has shown promising results in the context of style harmonization tasks. Our hypothesis posits that the use of a naive composite image as the content input for style transfer, rather than an original input photo, played a pivotal role in enhancing the achieved results. This assumption is grounded in two key factors:

- The substantial overlap between content and style images make the style transfer model to concentrate on the pasted object.
- During the style transfer procedure, the boundary between the pasted object and the background painting exhibited a more relaxed transition. In cases where an initial stylization of the input photo was followed by its placement onto the painting, a more pronounced boundary seam was likely to emerge

To demonstrate this, we are currently working on implementing an alternative harmonization configuration that involves initially transferring an input photo

## V. Conclusion

We have presented an interactive system for painterly image harmonization. The proposed system provides a unified framework by integrating object segmentation and painterly

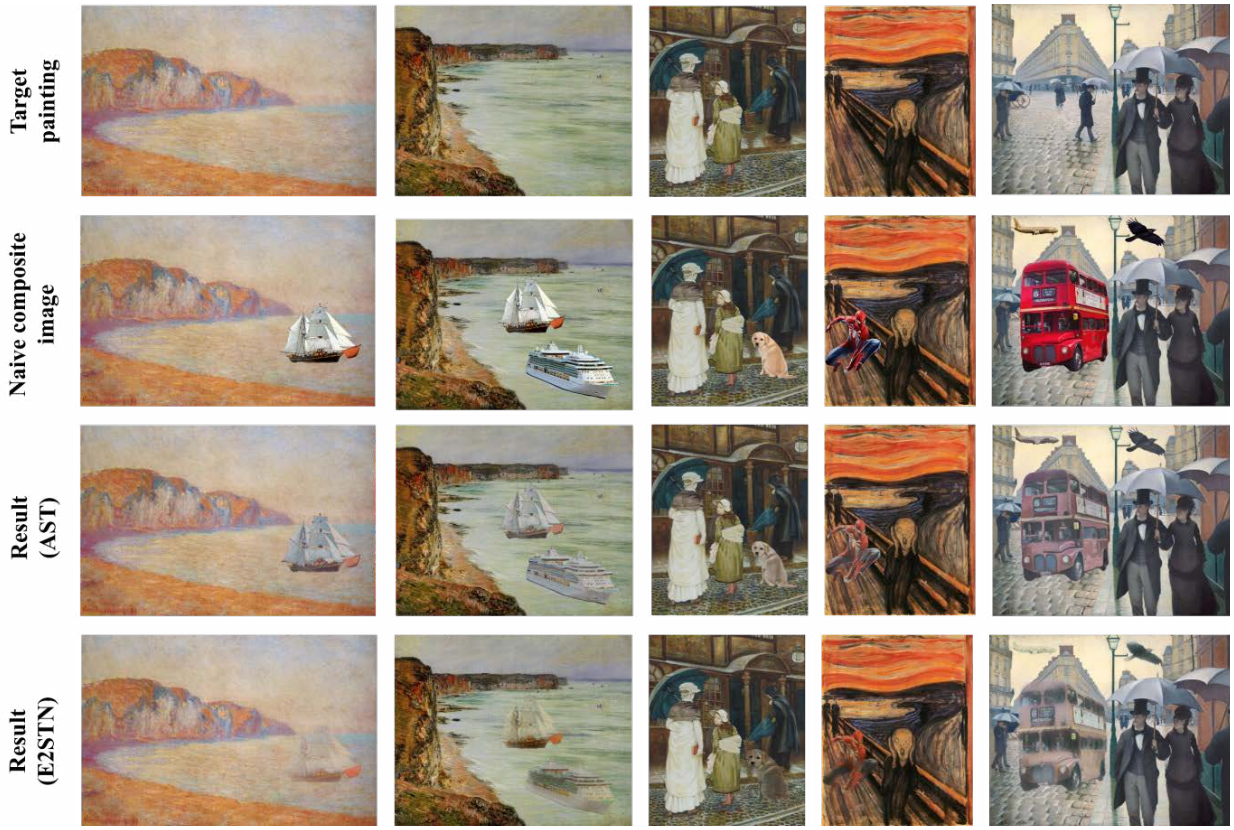


Fig. 7. Visualization of the results obtained by the proposed system. AST represents fast arbitrary style transfer [4].

image harmonization with a user-friendly GUI. Moreover, our design incorporates a straightforward harmonization pipeline empowered by an arbitrary style transfer model. The adoption of a naive composite image as the content for style transfer yields visually pleasing results, as it well preserves object details while upholding stylistic coherence.

Based on the results, we can conclude that the proposed system can effectively assist users in creating new artworks by placing objects from photos into paintings and harmonizing them according to their intention in a unified framework.


## ACKNOWLEDGMENT

This research is supported by Ministry of Culture, Sports and Tourism and Korea Creative Content Agency(Project Number: R2021040122, Project Name: Development of artificial intelligence-based photo/painting harmonization and content expansion technology)


## REFERENCES


[1] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," Conference on Computer Vision and Pattern Recognition (CVPR), pp.2414-2423, 2016.

[2] X. Huang and S. J. Belongie, "Arbitrary style transfer in real-time with adaptive instance normalization," International Conference on Computer Vision (ICCV), pp.1510-1519, 2017.

[3] J. Johnson, A. Alahi and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," European conference on computer vision (ECCV), pp.694-711, 2016.

[4] G. Ghiasi, H. Lee, M. Kudlur, V. Dumoulin, and J. Shlens, "Exploring the structure of a real-time, arbitrary neural artistic stylization network," British Machine Vision Conference (BMVC), 2017.

[5] F. Luan, S. Paris, E. Shechtman and K. Bala, "Deep painterly harmonization," In Computer graphics forum 37(4), pp.95-106, 2018.

[6] T.R. Shaham, T. Dekel and T. Michaeli, T, "SinGAN: Learning a generative model from a single natural image," International Conference on Computer Vision (ICCV), pp.4570-4580, 2019.

[7] H.J. Peng, C.M. Wang and Y.C.F. Wang, "Element-Embedded Style Transfer Networks for Style Harmonization," British Machine Vision Conference (BMVC), 2019.

[8] J. Cao, Y. Hong and L. Niu, "Painterly image harmonization in dual domains," AAAI Conference on Artificial Intelligence, 37(1), pp. 268-276, 2023.

[9] L. Lu, J. Li, J. Cao, L. Niu and L. Zhang, "Painterly Image Harmonization using Diffusion Model," ACM Multimedia, 2023.

[10] Y.H. Tsai, X. Shen, Z. Lin, K. Sunkavalli, X. Lu and M.H. Yang, "Deep image harmonization," Computer Vision and Pattern Recognition (CVPR), pp. 3789-3797, 2017.

[11] W. Cong, J. Zhang, L. Niu, L. Liu, Z. Ling, W. Li and L. Zhang "Dovenet: Deep image harmonization via domain verification," Computer Vision and Pattern Recognition (CVPR), pp. 8394-8403, 2020.

[12] X. Cun and C.M. Pun, "Improving the harmony of the composite image by spatial-separated attention module," IEEE Transactions on Image Processing 29, pp.4759-4771, 2020.

[13] L.C. Chen, G. Papandreou, F. Schroff and H. Adam, "Rethinking atrous convolution for semantic image segmentation," arXiv preprint arXiv:1706.05587, 2017.

[14] C. Rother, V. Kolmogorov and A. Blake, ""GrabCut": interactive foreground extraction using iterated graph cuts," In ACM transactions on graphics (TOG), 23(3), pp.309-314, 2004.

[15] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," In ACM Transactions on Graphics (TOG), vol. 22, pp. 313– 318, 2003.

[16] AST v2, https://tfhub.dev/google/magenta/arbitrary-image-stylization-v1-256/2 Accessed:2023-08-05

[17] WikiArt, https://www.wikiart.org/ Accessed: 2023-08-05