Inpainting GAN-Based Image Blending with Adaptive Binary Line Mask

Thanh Hien Truong[1], Tae-Ho Lee[1*], Viduranga Munasinghe[1],
Tae Sung Kim[2], Jin-Sung Kim[2], Hyuk-Jae Lee[1]

Abstract

Image blending is a scheme for image composition to make the composite image looks as natural and realistic as possible. Image blending should ensure that the edges of the object look seamless and do not distort colors. Recently, numerous studies investigated image blending methods adopting deep learning-based image processing algorithms and contributed to generating natural blended images. Although the previous studies show remarkable performance in many cases, they suffer from quality drop when blending incompletely cropped object. This is because partial loss and unnecessary extra information on the cropped object image interferes with image blending. This paper proposes a new scheme that significantly reduce the unnatural edges and the color distortion. First, to detect and handle the incompletely cropped region, an adaptive binary line mask generation utilizing color difference checking algorithm (CDC) is proposed. The generated mask is exploited to improve image blending performance by isolating incompletely cropped image edges from image blending. Second, in order to perform inpainting the missing or masked area of the object image and image blending together, the inpainting generative adversarial model is adopted. Experimental results show that the blended images are not only more natural than those of the previous works but the color information is also well preserved.

Key Words: Image Blending, Image Inpainting, Composite Image, Generative Adversarial Network.

I. INTRODUCTION

In the field of image blending, the first step involves taking an object out of the source image using methods like image segmentation or matting. After the object is cropped, it is carefully added into another target image to create a blended image. The primary objective of image blending is to alleviate the presence of artificial boundaries that may arise between the object and the target image. In prior research, the methodologies employed for image blending can be categorized into two distinct approaches: the conventional approach and the deep learning-based approach.

First, the traditional methodologies for image blending primarily emphasized achieving a seamless transition between the object and the target image. One such technique is alpha blending, which involves the manual selection of alpha values and offers a simple and efficient approach [1]. Another method called Laplacian pyramid blending constructs multi-scale Laplacian pyramids and applies alpha blending at each level [2]. Additionally, the Poisson image blending technique enforces visual consistency by preserving gradient information [3]. However, it is worth noting

that poisson image blending often yields undesired outcomes, including color distortion, loss of fine details, and the appearance of ghost artifacts within the blended image [3].

Second, the recent approaches to image blending can be categorized as novel frameworks that incorporate machine learning. One such framework is the Gaussian-Poisson Generative Adversarial Network (GP-GAN), which combines the strengths of both GANs and gradient-based blending techniques [4]. Additionally, a two-stage deep learning algorithm for image blending, independent of any training data, has been proposed as an alternative to GP-GAN [5]. However, both of these methods tend to introduce color distortions in the blended regions. Another state-of-the-art scheme is a deep learning-based framework specifically designed for portrait image compositing, including object segmentation and mask refinement networks [6]. Although the previous studies show remarkable blended image quality, they suffer from image quality degradation when blending incompletely cropped object. In real-world situations, it's common to not cut out objects perfectly due to errors in the image segmentation process. Fig. 1 illustrates an example of incomplete object cropping and its subsequent blending

outcome. In Fig. 1(a), a source image and a target image are depicted. Fig. 1(b) displays a cropped result, where the green squares indicate instances of partial loss of the object, while the red squares represent cropped object images that include background elements from the source image. Fig. 1 (c) shows cases an image blending result by simply copying and pasting the cropped object image onto the target image. It is evident that the presence of loss and unintended additional background elements introduces an unnatural boundary between the object and the target image. Despite applying state-of-the-art image blending techniques, the unnatural boundary remains unresolved. In this paper, we refer to this artificial boundary as a "visual gap".

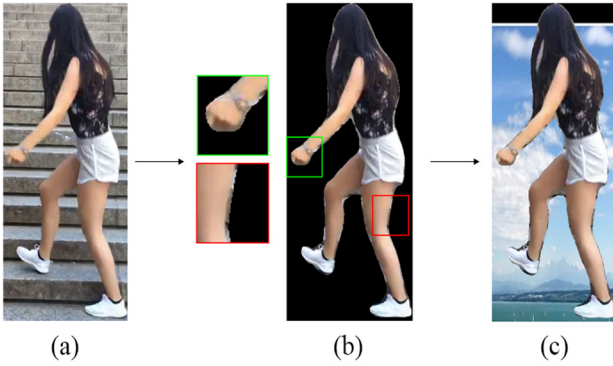This paper introduces a framework, depicted in Fig. 2,

Fig. 1. Example of an incomplete object cropping result and its effect on the blending result: (a) the source image, (b) the object image cropped by a segmentation mask, and (c) the blending result. Several pixels that belong to the background are included in the object image as shown in the red-bordered sample. On the other hand, the fingertip of the person is incomplete as shown on the green-bordered sample. Because of these reasons, the blending result is negatively affected, where a lot of artifacts can be noticed easily.

designed with the explicit purpose of minimizing the visual gap in image blending. Within this context, our contributions stand out prominently in the enhancement of seamless blending:

- We present an inventive approach for the generation of adaptive binary line masks. Employing a sophisticated color difference checking algorithm, our method can detect and accommodate variations in visual gap thickness. This empowers us to meticulously process blending boundaries, leading to the creation of natural and unobtrusive transitions.
- Our framework showcases the integration of two key elements: inpainting and blending. Leveraging a state-of-the-art inpainting GAN-based technique, we address the task of image restoration and merging with unparalleled finesse. By harnessing the inpainting GAN's prowess, we effectively restore absent or impaired sections within cropped images. Simultaneously, this mechanism operates as an image blender, orchestrating the seamless amalgamation of objects into target images. As a testament to its success, the blended output exhibits flawlessly smooth transitions, all while preserving the authentic colors of both the subject and the target image.

The remainder of this paper is organized as follows. Section II provides an overview of related works, highlighting the existing research and methodologies related to visual gap reduction in image blending. Section III presents the proposed scheme aimed at mitigating the visual gap. This section outlines the methodology and techniques utilized to address the challenges associated with blending boundaries. Section IV presents the experimental results obtained from applying the proposed scheme. The outcomes and perfor-
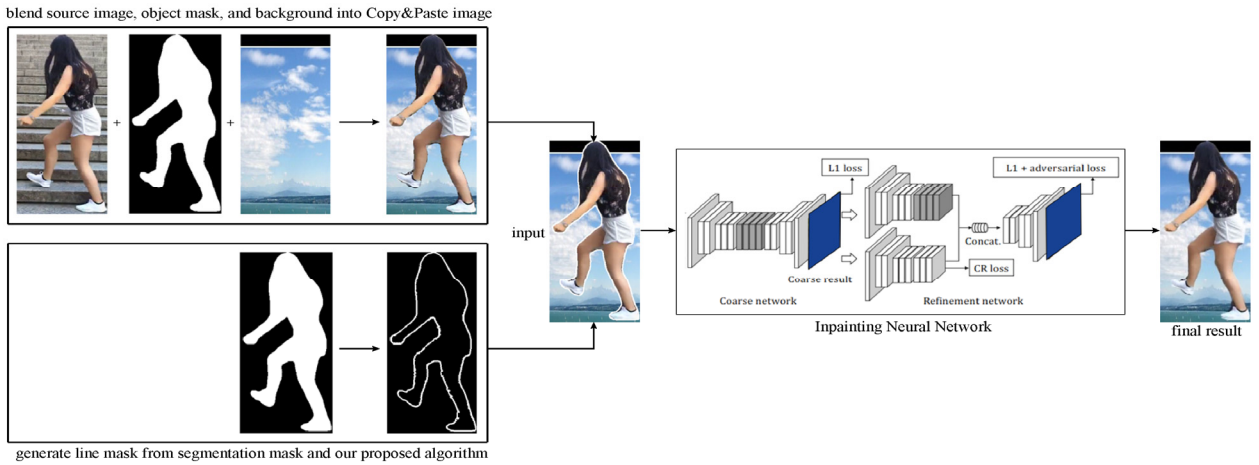

Fig. 2. The proposed framework. In the first step (top left block), we isolate the object from its source image using a segmentation mask. The object image can be blended on top of a target background image to generate a naïve Copy&Paste image. In the second step (bottom left block) then, a line mask which defines the region of the Copy&Paste image that needs adjustment is generated from the segmentation mask following our proposed algorithm. In the next step, the Copy&Paste image and line mask are fed as the input to an inpainting neural network model to refill the defined region. The inpainting model replaces the artifacts by appropriate pixels so that the visual gap is reduced at the cutting edge of the two components. The final result can be seen seamlessly blended.

mance of the framework are analyzed and discussed in detail. Finally, in Section V, the paper concludes by summarizing the key findings and contributions of the study, as well as suggesting future work.

II. RELATED WORKS

2.1. Image Blending using Generative Adversarial Network

Generative Adversarial Networks (GANs) [13] are a type of neural network commonly employed for generative modeling tasks. A generative model aims to generate new samples that resemble the existing dataset while introducing specific variations or differences. GANs consist of two main neural network models: the "generator" or "generative network" and the "discriminator" or "discriminative network". The generator model is responsible for generating plausible samples, while the discriminator model learns to distinguish between generated examples and real examples from the dataset. These two models engage in a competitive process during training, where the generator tries to produce samples that fool the discriminator, while the discriminator strives to accurately identify real and generated samples. GANs have found applications in various domains. For example, they have been used to generate new human poses, perform inpainting and blending of images, and generate synthetic examples for image datasets. By leveraging the adversarial training process, GANs have proven to be effective in generating realistic and diverse samples in a variety of generative modeling tasks.

The framework of GP-GAN [4] takes advantages of both GANs and gradient-based image blending methods while Zhang et al [5] proposed a two-stage deep-learning blending algorithm which does not rely on any training data as GP-GAN. However, both methods distort the colors around blending boundary.

2.2. Image Inpainting

Image inpainting is a fundamental task in Computer Vision that involves reconstructing missing regions within images. Traditional inpainting methods typically rely on borrowing pixels from the surrounding regions of the image that are intact and not missing. While these techniques work well for filling in background areas, they struggle in scenarios where the surrounding regions lack the necessary information or when the missing parts require the inpainting system to infer the properties of the objects that should be present. However, with the advent of deep learning and the availability of large-scale datasets, modern approaches employ deep neural networks to predict missing parts of an image. These deep learning-based methods enable generating missing pixels with improved global consistency and

local fine textures. The abundance of paired training data can be automatically generated by intentionally corrupting images and using the original, uncorrupted images as ground-truth. The EdgeConnect scheme leverages salient edge detection to guide the inpainting process [9]. DeepFill proposed by Yu et al., incorporates contextual attention, which refers to surrounding image features to make more accurate pixel predictions for the holes [8]. Recent advancements in deep generative methods [8-10], particularly those based on Generative Adversarial Networks (GANs), have showcased impressive perfor-mance in image completion and inpainting tasks.

III. PROPOSED METHOD

3.1. Preliminary

Given a source raw image x_{raw}, a background image x_{bg} and a segmentation mask image x_{mask}, using the copying-and-pasting strategy, a composite Copy-Paste image x_{comp} can be obtained by equation (1), where * is element-wise multiplication operator. The goal of conditional image generation is to generate a well-blended image that is semantically similar to the composited image x_{comp} but looks more realistic and natural with resolution unchanged.

$$x_{comp} = x_{raw} * x_{mask} + x_{bg} * (1 - x_{mask}). \tag{1}$$

3.2. Line Mask

The segmentation mask generated for the object may not be perfect, leading to some inaccuracies where certain points that should belong to the background are mistakenly included in the mask, while other points that should be part of the object are excluded. This issue is illustrated in Fig. 1. To address this and achieve more accurate blending of the object image, we propose the utilization of an additional mask called the Line Mask, which is used to fine-tune the segmentation mask.

The line mask is essentially a line drawn along the contours of the object, with varying thickness at different points. Its primary purpose is to identify and mark noise pixels that should not be included in the segmentation mask. During the subsequent blending procedure, these noise pixels will be replaced to enhance the visual quality of the final result. While the line mask primarily serves to identify noise pixels, it can also include some pixels that belong to the actual object, specifically at the very edge of the object. However, the values of these pixels can be modified during the blending algorithm to better align with the new background, ensuring a seamless and visually pleasing integration of the object into the composite image.

To generate the line mask, we introduce a Color differ-

ence checking (CDC) algorithm. This algorithm is primarily applied to the pixels located along the contours of the isolated object. The CDC algorithm leverages color differences to classify which pixels are considered noise and which are not. This classification assumes that the segmentation mask closely adheres to the object boundaries.

The CDC algorithm incorporates several hyperparameters, which we optimize based on the performance of the segmentation model and the principles of color differentiation. By fine-tuning these hyperparameters, we aim to achieve the best possible results in terms of accurately identifying and marking noise pixels in the line mask. The optimization process involves assessing the performance of the segmentation model and incorporating knowledge of color differentiation theory to determine the most suitable values for the hyperparameters.

3.3. Color Difference Checking Algorithm

The color difference checking (CDC) algorithm, presented in Table 1, is utilized to generate the line mask, which plays a crucial role in accurately separating the contour points of the isolated object into actual object and noise. By considering the color values obtained from the original image and the initial object/background classification provided by the binary mask, the algorithm examines

each contour point. It compares the color difference between a contour point and the average color of the object to determine if it should be classified as noise or part of the object in the line mask. Points that deviate significantly in color from the object's color distribution are marked as noise, while others are marked as part of the object. This process ensures the line mask effectively identifies and distinguishes noise pixels, contributing to the quality and precision of the image blending procedure.

The algorithm in general is an iterative comparison. For each point p in the contour C of object, we compare its color value to that of a reference point, which is k pixel away following a provided direction di from it (for example, if the direction di is left, the reference point is to the left of p). The distance k starts at 2, being the first comparison of p with its neighbor pixel. Until k reaches its maximum allowed value, we keep changing the reference point and comparing the colors. When there is a reference point that is actually different from p following our criteria, no more checking of p is necessary. In this case, we can say that p is a noise pixel, and other pixels from p to right before the reference point in the selected direction are also noise. Otherwise, p is considered an actual object point. All the results are tracked, including the query point p and the distance k.

The algorithm can be repeated with each of four directions: left, right, top, and bottom. With the gradual increment of k, non-convex segmentation mask is no problem since the algorithm can always detect an invalid reference point and stop early. One example of iterative comparison is illustrated in Fig. 3. In Fig. 3, the algorithm is allowed to reach up to k=5, but it will likely stop at k=3 as the threshold has been surpassed. This information is provided by the segmentation mask.

The selection of the threshold value (T_d) in the CDC algorithm is based on Weber's Law of Just Noticeable Differences, which states that stimuli must differ by a minimum percentage to be perceived as different. This principle applies to various domains, including light and color perception. Specifically, it has been demonstrated that a minimum difference of 8% is required for color discrimination.

In the case of using 8-bit color, where the intensity range is from 0 to 255, the color-difference threshold can be determined using the equation (1), which calculates the 8% of the maximum intensity value. Thus, the color-difference threshold is calculated as 255×8%=20. Accordingly, we adopt this value as the threshold in the CDC algorithm.

By selecting a threshold value based on Weber's Law, we ensure that the algorithm can effectively differentiate between color variations that are perceptually significant and those that are not. This threshold determination process aids in achieving accurate classification of noise and object points in the line mask, contributing to the overall quality

Table 1. Color difference checking.

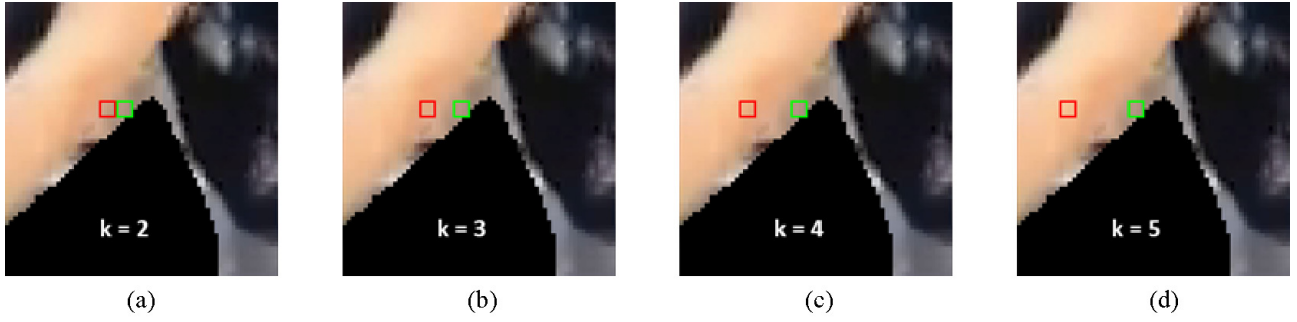|	Data: Image **Im**, binary mask **M**, contours **C**, direction **di**, max distance k_{max}, color threshold T_d = 20
1	Initialize track of background thickness **TH** for each point in **C**
2	**for** point p in **C do**
3	k = 2
4	**while** k ≤ k$_{max}$ **do**
5	Find point q: distance p to q in direction **di** is k
6	**if** q not exist **then**
7	**break**
8	**end**
9	**if** M[q] == 1 **then**
10	d = | **Im**[p] − **Im**[q] |
11	**if** d > T_d **then**
12	**TH**[p] = k
13	**Break**
14	**end**
15	$k = k$ + 1
16	**end**
17	**end**
18	**end**

Fig. 3. Illustration of iterative comparison in CDC algorithm. The selected direction is left. A contour point **p** is marked green and reference points are marked red. Reference points are selected by the distance **k**, which iteratively takes the value of (a) 2, (b) 3, (c) 4, and (d) 5. In each iteration, the difference of color, or pixel value, is computed and compared to a pre-defined threshold. In a real scenario, the algorithm may early stop at **k**= 3 if the threshold is surpassed (The cases where **k**=4 and **k**=5 are then never visited.).

and fidelity of the image blending process.

3.4. Generation of Line Mask

Fig. 4 illustrates the stages of the line mask generation process. This process involves three distinct steps. It starts with the segmentation mask and its contours as illustrated in Fig. 4(a). In the first step, the objective is to identify object points within the contours of the initial isolated object obtained from the segmentation mask. By utilizing the CDC algorithm, points that belong to the object region are identified, as highlighted in white in Fig. 4(b).

In the second step, the inverted segmentation mask is used to determine which points initially classified as background should actually be a part of the object region. The

CDC algorithm, which can assess both noise points and the proximity of neighboring points, aids in this classification process. As a result, the segmentation mask expands, encompassing additional object points as shown in Fig. 4(c).

The third step involves the consideration of a new mask that includes the newly discovered object points from the second step. Again, the CDC algorithm is employed to separate noises and adjust the contours of the new mask. An example result is illustrated in Fig. 4(d). After that, post-processing techniques, such as Gaussian Blur and thresholding, are applied to smooth out the thickness map to form the final line mask as shown in Fig. 4(e).

Throughout the line mask generation, color difference is asserted between the reference pixel and the query pixels in
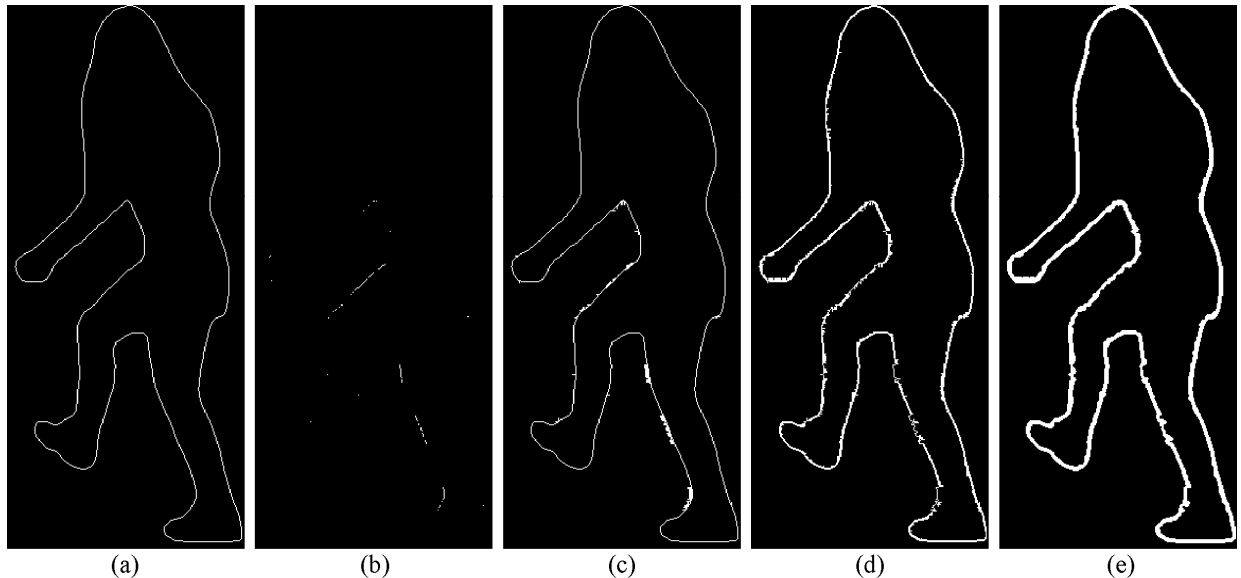

Fig. 4. Stage-by-stage generation of line mask following the proposed CDC algorithm. (a) First, we highlight the contours (white pixels) of the segmentation mask. Because the mask is imperfect, it can be expanded (from the contours) outward to include more pixels that should be belong to the main object, or inward to remove the pixels that should belong to the background. (b) By applying the CDC algorithm on the contours, the points which belong to the main object are highlighted in white. The second step is to expand outward from these points, resulting in a new border for the segmentation mask. (c) This border is also known as thickness map in the CDC algorithm. The third step is to expand inward from the non-highlighted contours to find the noises which should be removed from the segmentation mask. (d) This step results in a new thickness map. (e) Post-processing on this thickness map results in the final line mask. White pixels on the line mask denote the region that should be retouched by an inpainting neural network model later.

four directions, up, down, left, and right, ensuring comprehensive coverage. Notably, in the second step, the inverted segmentation mask is used to facilitate proper utilization of the CDC implementation. Additionally, the maximum distance, k_{max} is determined as 1% of the image height, considering the performance of the YOLACT-550 model, which yields an average Intersection-over-Union (IoU) of 90% when generating the segmentation mask.

3.5. Image Inpainting Model

The field of computer vision has witnessed remarkable advancements, leading to innovative solutions for image restoration and enhancement. One such breakthrough is the CR-FILL [7] inpainting model, a sophisticated algorithm designed to intelligently fill in missing or corrupted pixels within an image. This cutting-edge technology has demonstrated its prowess in seamlessly reconstructing image regions obscured by masks, significantly enhancing the visual quality and coherency of the final output. Building upon the success of image inpainting techniques, we propose an approach that leverages an inpainting model, CR-FILL. In Fig. 2, we illustrate the stages involved in utilizing the generator network of the CR-FILL model to obtain realistic composite images.

The first stage involves the coarse network of CR-FILL, which takes an incomplete image as input, with missing pixels represented as zeros, along with a line mask that indicates the regions requiring inpainting. The coarse network generates an initial prediction based on this input. Subsequently, in the refinement stage, the initial prediction produced by the coarse network is passed as input to the refinement network. The refinement network further processes the initial prediction and produces the final inpainting result. By incorporating the CR-FILL inpainting model, the proposed framework can achieve seamlessly blended images at the desired region while preserving the originality at the others.

IV. EXPERIMENTAL RESULTS

This section presents the experimental results of our proposed method. We conduct several experiments on different image datasets, comparing our method against various baseline approaches. To assess the performance of our framework, we utilize the Tiktok database [12], which consists of 2,615 human photos as raw images and an equal number of different background images as target images. The mask image used in the experiments is obtained from the YOLACT-550 [11] segmentation model applied to the raw images.

In our evaluation, we compare our method with several intuitive and strong baseline techniques. The naïve Copy

232

&Paste approach produces results with noticeable artificial boundaries, which detracts from the visual realism. The GP-GAN [4] method is capable of generating a smooth blending boundary, but it often leads to color distortion between the blending region and the background. Our previous work, Combined GP-GAN [14], addresses the color distortion issue but fails to achieve a truly seamless boundary when the raw and background images have significantly different color tones.

4.1. Quantitative Comparison

These experiments are aimed at demonstrating the advantages of the proposed method over baseline approaches. Various quantitative evaluation methods are employed to show the superior performance and capabilities of the proposed framework. By comparing the performance on the Tiktok database, we provide compelling evidence of the effectiveness and quality of the proposed method in achieving seamless and visually pleasing image blending.

For the quantitative evaluation, standard metrics such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM) are utilized to assess the compositing quality of the generated images. These metrics serve as verification tools and require ground-truth images with segmentation masks provided by the dataset author.

The detailed quantitative results are presented in Table 2. The Copy&Paste method, which is a naïve blending method and does not involve any additional adjustment, preserves the color of the original foreground well, leading to higher PSNR and SSIM scores, 72.329 dB and 0.928, compared to GP-GAN and Combined GP-GAN. By achieving PSNR and SSIM scores of 72.887 dB and 0.9321 respectively, which are superior than those of Copy&Paste, it is proven that the proposed method is also able to preserve the color scheme of the object. The proposed method focuses on retouching the cutting edge of the foreground and background using the line masks, resulting in minimal color distortion.

The proposed method exhibits an average SSIM value that is 0.072 higher than that of GP-GAN, and the average PSNR value is 7.754 higher than that of GP-GAN. It is similar when the proposed method is compared to Combined GP-GAN, demonstrating that our method effectively fills

Table 2. Comparison of various blending methods using SSIM, PSNR, and user survey.

Method	SSIM	PSNR (dB)	User votes
Copy&Paste	0.928	72.329	549
GP-GAN [4]	0.860	65.133	380
Combined GP-GAN [14]	0.912	69.698	550
Proposed method	**0.932**	**72.887**	**771**

and adjusts the boundary with an appropriate texture while improving the overall visual quality of the blended image. These results provide concrete evidence that the proposed method outperforms previous approaches and establishes a new state-of-the-art for the blending task.

4.2. Qualitative Comparison

Fig. 5 provides a visual comparison among all the methods, showing the differences in the blended results of 3 samples. Aligning with the quantitative results, it can be observed that the proposed method does not alternate the color scheme of the main object while GP-GAN and Combined GP-GAN failed this goal. In all 3 cases, GP-GAN blended the objects and the backgrounds with alpha-like matting method. This causes the objects to look partially transparent on top of the background which is unnatural. The Copy &Paste results, no matter which kind of segmentation mask is used, include noises in the main objects, unexpectedly

(a) Copy&Paste with default mask (GT)   (b) Copy&Paste deep learning mask   (c) GP-GAN deep learning mask   (d) Combined GP-GAN deep learning mask   (e) **Our method**
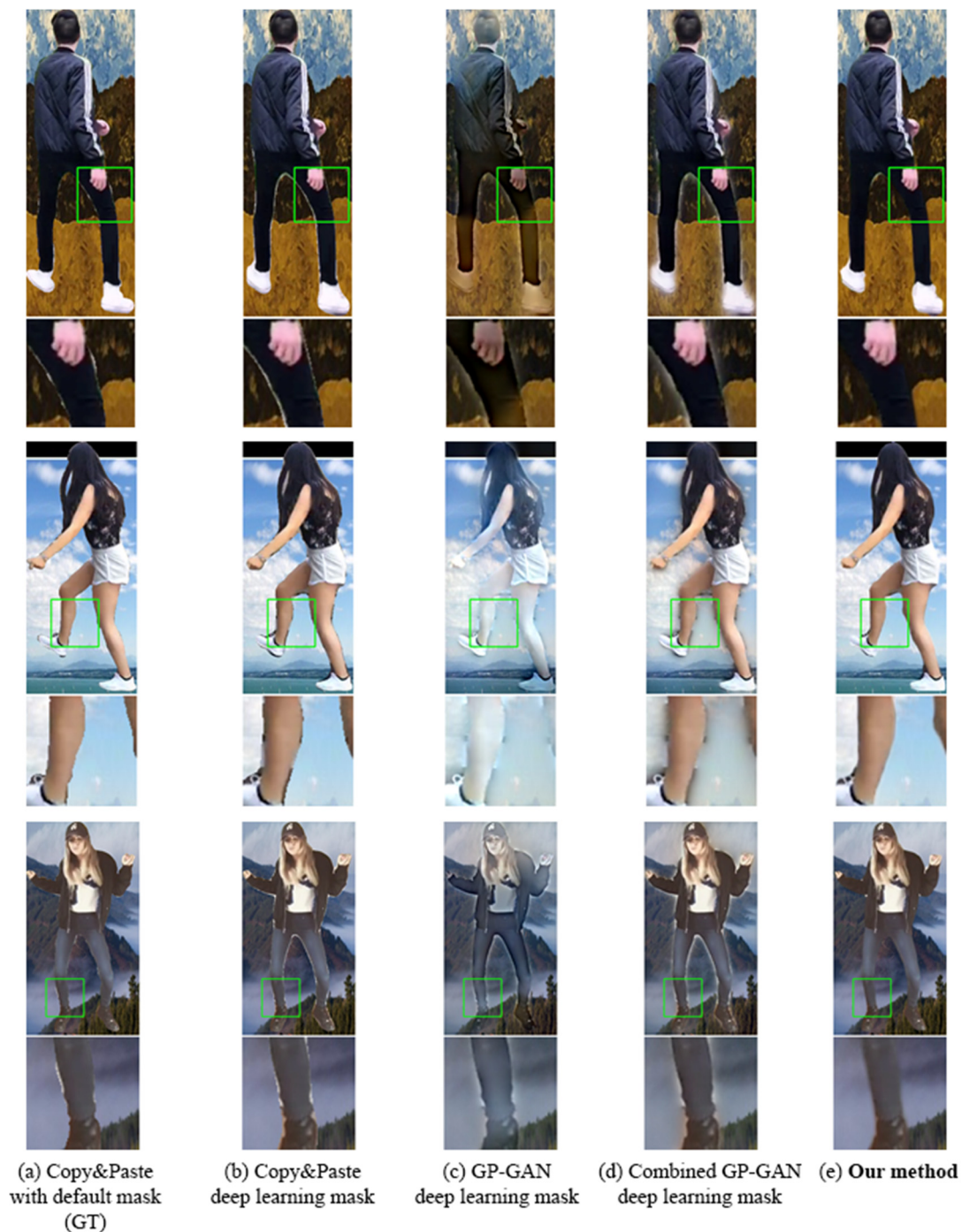
Fig. 5. Visual comparison of five different methods applied on three samples. From left to right: (a) Copy&Paste with default mask, (b) Copy &Paste with deep learning mask, (c) GP-GAN [4] with deep learning mask, (d) Combined GP-GAN [14] with deep learning mask, and (e) Proposed method with deep learning mask.

233

creating visually noticeable edges between the objects and the backgrounds. Despite being able to preserve the object color scheme, the quality of the results is not good. Meanwhile, the results offered by Combined GP-GAN show visible halo artifacts that severely degrade the visual quality as well. Only the proposed method consistently offers the cleanest blended images in all 3 cases. We chose the backgrounds which are perceptually and contextually different to the object; thus, the blended image is well recognized as unnatural. However, if we look at the patches in Fig. 5 (under each blended image, which is the zoomed-in version of the green-bordered samples), the matching of the objects and the backgrounds is acceptable with our proposed method.

We conducted a survey with forty-five participants to obtain subjective opinions on the realism of the blended images generated by the four algorithms. Each participant was asked to select the most realistic blended image out of four randomly generated options. The results, presented in Fig. 6 and Table 2, clearly indicate that the proposed method received the highest number of votes, 771, which outperformed the second best one by a large margin, Combined GP-GAN with 550 votes. This survey confirms that the proposed method has a superior performance in terms of quantitative assessment and subjective perception. Furthermore, the low number of votes, 380, that GP-GAN, received is due to its non-preservation of color. This problem can be solved by the proposed method.
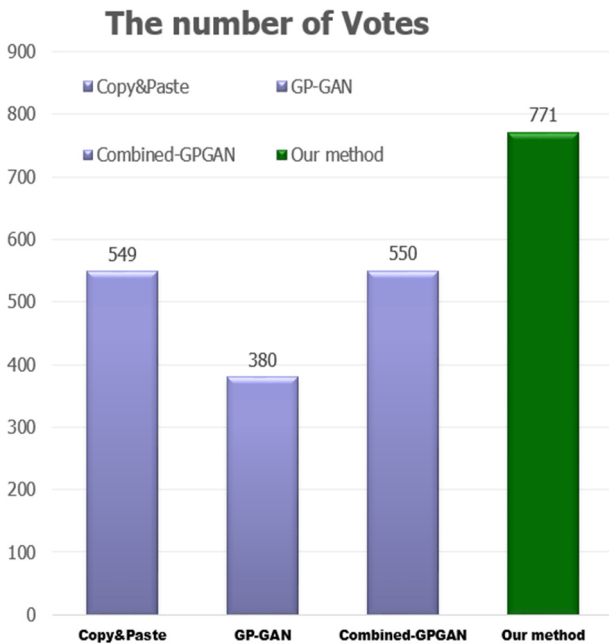

Fig. 6. User study results. From left to right: (1) Copy&Paste with deep learning mask, (2) GP-GAN, (3) Combined GP-GAN, and (4) Proposed method.

V. CONCLUSION

In this work, we present a novel image blending framework that leverages inpainting techniques to generate realistic and natural composite images without introducing artifacts or color distortion. The proposed method involves compositing a foreground and a background image using a binary mask to define the blending region. To refine the blending boundary, we introduce a line mask generated through an algorithm that adjusts the original mask by including relevant pixels and excluding others. We employ the CR-FILL generator network to inpaint the region indicated by the line mask, ensuring a seamless blending of the two images. Importantly, the proposed method preserves the color scheme of the main foreground texture, which is crucial for applications such as fashion image blending with human models.

The effectiveness of the proposed method is validated through both a user study and quantitative experiments. The user study reveals that the blended images obtained by the proposed method are consistently voted as having the highest visual quality compared to images generated by three previous methods. The quantitative experiments show that the proposed method achieves higher PSNR and SSIM scores than the alternative approaches, establishing a new state-of-the-art for image blending using inpainting techniques.

Notably, the proposed method offers a practical solution that is simple yet efficient to implement. It provides a robust framework for generating high-quality blended images while preserving the integrity of the original foreground texture.

REFERENCES

[1] T. Porter and T. Duff, "Compositing digital images," in *Proceedings of the 11th Annual Conference on Computer Graphics and Interactive Techniques*, 1984, pp. 253-259.

[2] P. J. Burt and E. H. Adelson, "A multiresolution spline with application to image mosaics," *ACM Transactions on Graphics*, vol. 2, no. 4, pp. 217-236, 1983.

[3] P. Pérez, M. Gangnet, and A. Blake, "Poisson image editing," in *ACM SIGGRAPH 2003,* 2003, pp. 313-318.

[4] H. Wu, S. Zheng, J. Zhang, and K. Huang, "Gp-gan: Towards realistic high-resolution image blending," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 2487-2495.

[5] L. Zhang, T. Wen, and J. Shi, "Deep image blending," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2020, pp. 231-240.

[6] H. Zhang, J. Zhang, F. Perazzi, Z. Lin, and V. M. Patel,

"Deep image compositing," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 365-374.

[7] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv Prep. arXiv:1511.06434*, 2015.

[8] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang, "Free-form image inpainting with gated convolution," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 4471-4480.

[9] K. Nazeri, E. Ng, T. Joseph, F. Z. Qureshi, and M. Ebrahimi, "Edgeconnect: Generative image inpainting with adversarial edge learning," *arXiv Prep. arXiv: 1901.00212,* 2019.

[10] Y. Zeng, Z. Lin, H. Lu, and V. M. Patel, "Cr-fill: Generative image inpainting with auxiliary contextual reconstruction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 14164-14173.

[11] D. Bolya, C. Zhou, F. Xiao, and Y. J. Lee, "Yolact: Real-time instance segmentation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2019, pp. 9157-9166.

[12] K. Roman, "Segmentation full body TikTok dancing dataset," 2023; https://www.kaggle.com/datasets/tapakah68/segmentation-full-body-tiktok-dancing-dataset

[13] I. Goodfellow, P. A. Jean, M. Mehdi, X. Bing, W. F. David, and O. Sherjil, et al., "Generative adversarial nets," in *Advances in Neural Information Processing Systems (NIPS 2014)*, 2014, pp. 2672–2680.

[14] T. H. Truong, T. H. Lee, V. Munasinghe, and H. J. Lee, "Combination of GP-GAN and content loss for realistic image blending," in *Korean Society of Electronics Engineers Conference*, 2022, pp. 1486-1489.

AUTHORS

Thanh Hien Truong received the B.S., degree in electronics and telecommunications from Hanoi University of Science and Technology, Hanoi, Vietnam in 2019 and the M.S. degree in electrical and computer engineering from Seoul National University, Seoul, Korea in 2023.

Her research interests include computer vision.

Tae-Ho Lee received the B.S., M.S., degrees in electronic, electrical, control & instrumentation engineering from the Hanyang University, in 2001 and 2003, respectively. and He received the Ph.D. degrees in electrical & computer engineering from the Seoul National University, Seoul, Korea, in 2018. From 2018 to 2021, he was a Post-Doctoral Researcher with Seoul National University. From 2021 until now, he is a Visiting Assistant Professor in the Next-Generation Semiconductor Convergence and Open Sharing System, Seoul National University. Fields of interest are deep learning object detection, hand gesture recognition, GAN based medical image registration for augmented reality applications.

Viduranga Munasinghe has received his B.S. degree in the Department of Electrical and Computer Engineering from Seoul National University, Korea, in 2018. In 2019, he joined Computer Architecture and Parallel Processing Lab for pursuing his Ph.D. degree at Seoul National University.

His research interests include image processing, deep learning object detection and computer vision.

Tae Sung Kim received the B.S degree in electrical electronic engineering from Pusan National University, Pusan, South Korea, in 2010 and the M.S. and Ph.D. degrees in electrical and computer engineering from Seoul National University, Seoul, Korea, in 2013 and 2017, respectively. He is currently a Staff Engineer with the System LSI Division, Samsung Electronics Corporation, Hwaseong, Korea. His research interests include the algorithm and organization design of H.264/AVC and HEVC and next-generation video coding systems.

Jin-Sung Kim received his BS and MS degrees and his Ph.D. degrees in electrical engineering and computer science from Seoul National University, Seoul, Korea, in 1996, 1998, and 2009, respectively. From 1998 to 2004, and from 2009 to 2010, he was with Samsung SDI Ltd., Cheonan, Korea, as a senior researcher, where he was involved in driver circuits and discharge waveform research. From 2010 to 2011, he was a post-doctoral researcher at Seoul National University. In 2011, he joined the Department of Electronic Engineering, Sun Moon University, Asan, Korea, where he is currently a Professor. His current research interests include deep learning, algorithms and architectures for video compression and computer vision.

Hyuk-Jae Lee (Member, IEEE) received the B.S. and M.S. degrees in electronics engineering from Seoul National University, Korea, in 1987 and 1989, respectively, and the Ph.D. degree in electrical and computer engineering from Purdue University, West Lafayette, IN, USA, in 1996. From 1996 to 1998, he was with the Faculty of the Department of Computer Science, Louisiana Tech University, Ruston, LA, USA. From 1998 to 2001, he was with the Server and Workstation Chipset Division, Intel Corporation, Hillsboro, OR, USA, as a Senior Component Design Engineer. In 2001, he joined the School of Electrical Engineering and Computer Science, Seoul National University, Korea, where he is currently a Professor. He is the Founder of Mamurian Design, Inc., a fabless SoC design house for multimedia applications. His research interests include the areas of computer architecture and SoC design for multimedia applications.