

Deep Image Blending

Lingzhi Zhang Tarmily Wen Jianbo Shi

University of Pennsylvania

{zlz, jshi}@seas.upenn.edu, went@sas.upenn.edu

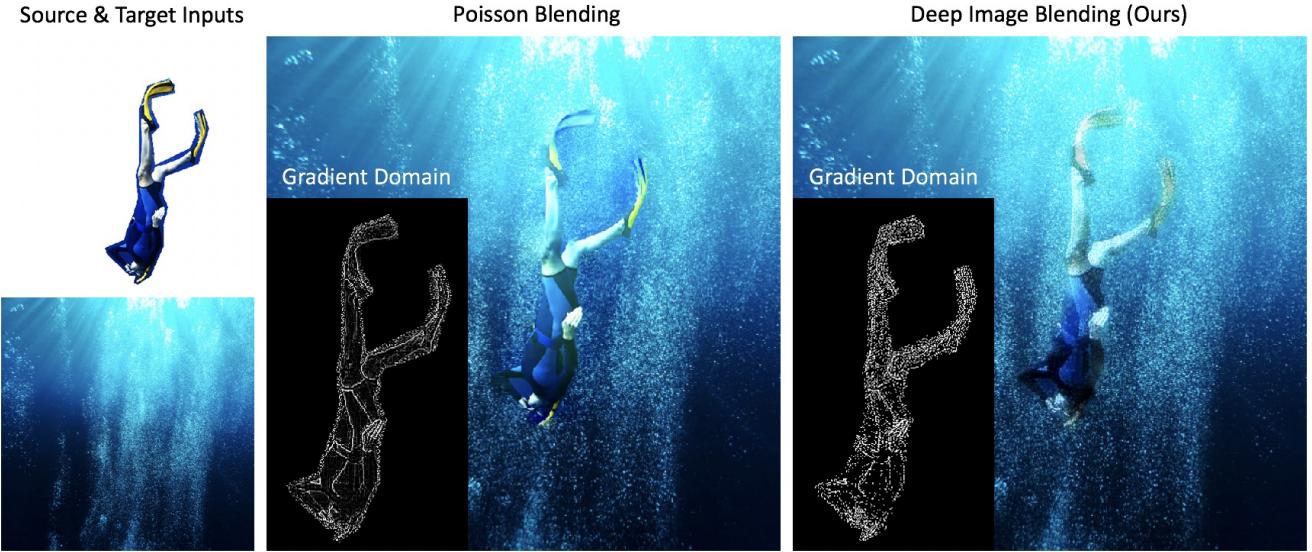


Figure 1: Given a coarsely cropped object in a source image, a target image and a blending location, our algorithm can blend the selected object onto a target image with seamless boundary and consistent style with respect to the target image.

Abstract

Image composition is an important operation to create visual content. Among image composition tasks, image blending aims to seamlessly blend an object from a source image onto a target image with lightly mask adjustment. A popular approach is Poisson image blending [23], which enforces the gradient domain smoothness in the composite image. However, this approach only considers the boundary pixels of target image, and thus can not adapt to texture of target background image. In addition, the colors of the target image often seep through the original source object too much causing a significant loss of content of the source object. We propose a Poisson blending loss that achieves the same purpose of Poisson image blending. In addition, we jointly optimize the proposed Poisson blending loss as well as the style and content loss computed from a deep network, and reconstruct the blending region by iteratively updating the pixels using the L-BFGS solver. In the blending image, we not only

smooth out gradient domain of the blending boundary but also add consistent texture into the blending region. User studies show that our method outperforms strong baselines as well as state-of-the-art approaches when placing objects onto both paintings and real-world images. Code is available at: <https://github.com/owenzlz/DeepImageBlending>

1. Introduction

Image blending is a method for image composition. It generally refers to cropping a certain region of a source image (usually an object) and placing it onto the target image at a specified location, where the goal is to make the composite image look as natural as possible. The challenge of this task is that the cropped region may not be precisely delineated. Therefore, the blending process needs to not only adjust the appearance of the cropped object to be compatible with the new background but also make the cropping boundary appear seamless.

The current most popular method for image blending is Poisson image editing [23]. The idea is to reconstruct pixels

in the blending region such that the blending boundary has smooth pixel transition or small gradients with respect to the boundary pixels in the target image. However, this method is difficult to combine with other reconstruction objectives because of its closed-form matrix solution. A recent work [33] combines the closed-form solution of Poisson equation with GAN loss to synthesize realistic blending images. However, this method requires a source region, a target image, and a corresponding well-blended image as training examples for supervised learning. Closely related to image blending is image harmonization, but the foreground object must be precisely delineated and thus the goal is to only adjust the illumination, color, and texture of the foreground to make it compatible with the new background.

In this work, we propose a novel two-stage blending algorithm. The algorithm first generates a seamless boundary for the source region, and then further refines the region with similar styles and textures with respect to the target image. In this algorithm, we propose a differentiable loss that enforces the equivalent purpose of the original objective of Poisson equation, and it can be easily combined with other reconstruction objective functions. Our algorithm works well not only for real-world target images but also stylized paintings by utilizing content and style loss [7] from deep features. In addition, our algorithm solves the reconstruction of image blending using only a single source image, a coarse mask, and a target image. Since our algorithm does not rely on any training data, it can generalize to any source and target images. Finally, we show the uniqueness and effectiveness of our algorithm compared to the state-of-the-arts methods through various testing cases.

2. Related Work

2.1. Image Blending

Image blending refers to cropping a certain region of a source image (usually an object) and placing it onto the target image at a specified location, where the goal is to make the composite image look as natural as possible. In contrast with image harmonization, an important characteristic of image blending is that it does not need precise object delineation for the blending mask. The default way of doing this task is to directly copy pixels from source image and paste them onto the target image, but this would generate obvious artifacts because of the abrupt intensity change in the compositing boundaries.

An early work, alpha blending [25], is the process of layering multiple images, with the alpha value for a pixel in a given layer indicating what fraction of the colors from lower layers are seen through the color at the given level. Although alpha blending performs much better than directly copy-and-pasting, it produces a ghost effect in the composite image as the contents in both source and target images exist in the same region.

Alternatively, the most popular image blending technique aims to inform gradient domain smoothness [23, 1, 6, 11, 13, 16, 29, 31]. The motivation of gradient domain blending is that the human visual system is very sensitive to the regions with abrupt intensity change, such as edges, and thus we want to produce an image with smooth transition over the blending boundary. The earliest work [23] proposes to reconstruct the pixels of the blending region in the target image by enforcing the gradient domain consistency with respect to the source image, where the gradient of the blending region is computed and propagated from the boundary pixels in the target image. With such gradient domain consistency, the blended image will have smooth transitions over the composite boundary even though the object mask are not precisely delineated. Our work is partially inspired by Poisson image editing [23], which will be further described in Section 3.

A recent approach GP-GAN [33] has leveraged the closed-form solution of the Gaussian-Poisson Equation [3] and Wasserstein Generative Adversarial Network (WGAN) [2] to produce photo-realistic blending results. However, this approach relies on supervised training, which requires paired data of a source image, target image, and corresponding well-blended image as ground-truth.

2.2. Other Image Editing Techniques

Other common image editing tools includes image denoising, image super-resolution, image inpainting, image harmonization, style transfer and so on. In recent years, Generative Adversarial Networks (GANs) [8] have been extensively applied to these tasks and produced very exciting results.

In super resolution [5, 14, 12, 37, 38, 39, 32], deep models learn the image texture prior to upsample a low-resolution image into high-resolution version. In image inpainting [15, 34, 22, 10, 35, 18, 36, 40], the network aims to fill the missing pixels in an image with the learned semantic information as well as real image statistics.

Closely related to our task is image harmonization, which extracts the foreground region in one image and combines it with the background of another image while adjusting the appearance of the foreground region to make it compatible with the new background. Early works [26, 24] use the global color statistics of the background image to adjust the appearance of the foreground image. Recently, [30] propose an automatic data acquisition approach and learn an end-to-end deep network to capture both the context and semantic information of the composite image during harmonization. Closest to our work, [20] propose a deep painterly harmonization technique to composite a real-world object onto a stylized paintings by adjusting the parameters of the transfer depending on the painting.

Another closely related task is style transfer [7, 12, 9, 19,

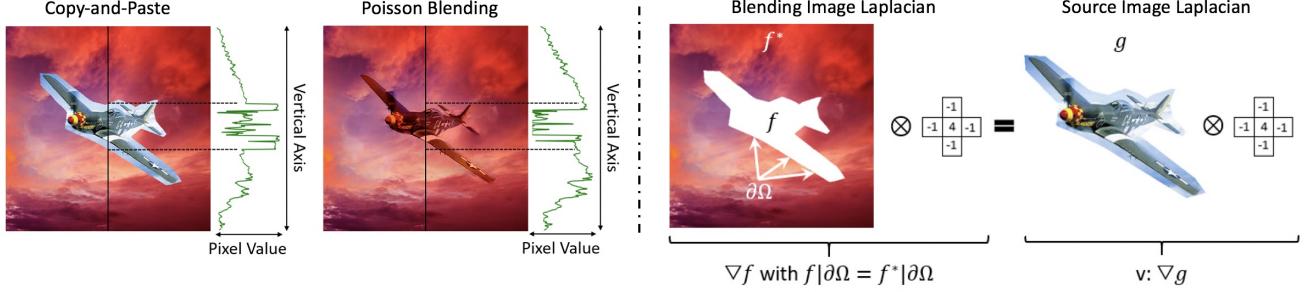


Figure 2: The two images on the left show the intensity change in a "copy-and-paste" image and a "Poisson Blended" image. The images on the right demonstrate the essential idea of poisson image editing, where the goal is to enforce the Laplacian gradient domain consistency between the source image and the blended image with a boundary constraint from target image. Specifically, we use a Laplacian filter to compute the second order gradient of images.

[17, 27], which aims to transform the style of an image into the style of another image. [7] first propose to transform a content image into the style of another image by jointly optimizing the transformed image's similarity of deep features with respect to the content image and similarity of gram matrices of deep features with respect to the style image.

3. Background - Poisson Image Editing

Directly copying a foreground object from source image and pasting it onto a target image can produce big intensity changes at the boundary, which creates obvious artifacts to human eyes. Therefore, the motivation of poisson image blending is to smooth the abrupt intensity change in the blending boundary in order to reduce artifacts. In Fig. (2), the left image shows the abrupt intensity change in the composite boundary in the copy-and-paste image and a smooth boundary transition using poisson blending [23].

In the original work [23], poisson image blending is formulated as an image interpolation problem using a guidance vector field.

$$\min_f \iint_{\Omega} |\nabla f - \mathbf{v}|^2 \text{ with } f|_{\partial\Omega} = f^*|_{\partial\Omega} \quad (1)$$

where $\nabla = [\frac{\partial}{\partial x}, \frac{\partial}{\partial y}]$ is the gradient operator, f is the function of the blending image, f^* is the function of the target image, \mathbf{v} is the vector field, Ω is the blending region and $\partial\Omega$ is the boundary of the blending region. In this case, the guidance field \mathbf{v} is the gradient field taken directly from source image g .

$$\mathbf{v} = \nabla g \quad (2)$$

We solve this minimization problem with boundary condition for each color channel independently to obtain the RGB image.

For images, the problem can be discretized using the underlying discrete pixel grid to obtain a quadratic optimiza-

tion problem.

$$\min_{f|_{\Omega}} \sum_{\langle p, q \rangle \cap \Omega \neq \emptyset} (f_p - f_q - v_{pq})^2, \text{ with } f_p = f_p^* \text{ for all } p \in \partial\Omega \quad (3)$$

where N_p is the set of 4-connected neighbors for pixel p , $\langle p, q \rangle$ denote a pixel pair such that $q \in N_p$, f_p is the value of f at p and $v_{pq} = g_p - g_q$ for all $\langle p, q \rangle$.

For discrete system, the solution can be converted into the following simultaneous linear equations. For $p \in \Omega$, the linear equations are as follows:

$$|N_p|f_p - \sum_{q \in N_p \cap \Omega} f_q = \sum_{q \in N_p \cap \partial\Omega} f_q^* + \sum_{q \in N_p} v_{pq} \quad (4)$$

For pixels p interior to Ω , there are no boundary pixels from the target image. Thus, the equation becomes the following:

$$|N_p|f_p - \sum_{q \in N_p} f_q = \sum_{q \in N_p} v_{pq} \quad (5)$$

We need to solve for f_p from the given set of simultaneous linear equations. Since Eq.(4) form a sparse symmetric positive-definite system, two classical iterative solvers Gauss-Seidel and V-cycle multigrid [23] are used to solve the linear system in the early works.

4. Methods

Our algorithm is a two-stage process, as shown in Fig. (3). In the first stage, a preliminary blended image is synthesized using the proposed Poisson gradient loss, style loss, and content loss. In the second stage, the preliminary blended image is further transformed to have a more similar style to match the target image. Here, we denote I_S as source image, I_T as target image, I_B as blending image, I_{BR} as refined blending image, and M as mask. Here, we assume the source image I_S has already been cropped out

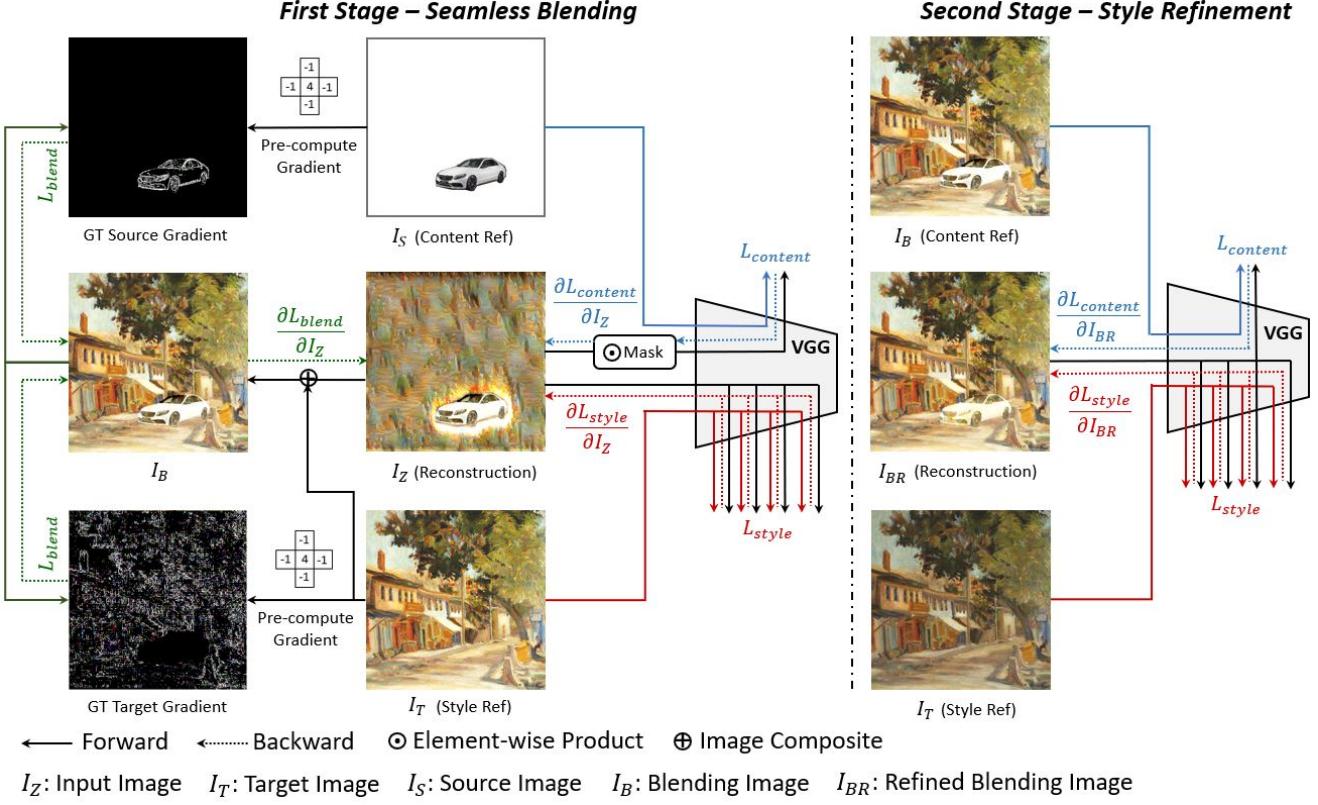


Figure 3: This figure shows our two-stage blending algorithm. In the first stage, an input image I_Z is randomly initialized and directly gets updated with respect to a gradient loss, content loss, and style loss. The gradient loss enforces the gradient of blending region to be the same as the gradient of the source object, the content loss enforces the semantic similarity between the blending region and the source object, and the style loss enforces the textural similarity between the blending region and the target image. In the second stage, the blending image from first stage is considered as an input image, and is further optimized with respect to the blending image and target image in terms of content and style, respectively.

using the coarse mask M . The size of I_S and I_T may or may not be the same, but they can be easily aligned using the user-provided offsets. For simplicity, we consider I_S and M are already aligned with I_T and thus have the same dimension in the following discussion. We further define an input image as I_Z , which represents the reconstructed pixels. During training, the joint loss back-propagates to I_Z in stage one or I_{BR} in stage two, and thus the optimization process essentially adjusts pixel values in I_Z or I_{BR} .

4.1. Poisson Gradient Loss

As we discuss in Section 3, Poisson image Eq. (1) proposed to reconstruct the pixels of the blending region by enforcing the gradient domain consistency between the blending image and the source image. In the meantime, the gradient of the blending image is initially computed from the boundary pixels of the target image and propagated toward the inside. Such consistency produce seamless boundary of the blending region, but it is solved using a well-designed matrix operation and is difficult to combine with other constraints for pixel reconstruction. Thus, we propose to convert this gradient domain constraint into a differentiable loss function as follows,

$$\mathcal{L}_{grad} = \frac{1}{2HW} \sum_{m=1}^H \sum_{n=1}^W [\nabla f(I_B) - (\nabla f(I_S) + \nabla f(I_T))]_{mn}^2 \quad (6)$$

In Eq. (6), ∇ represents the Laplacian gradient operator, and H and W are the width and height of image. The blending image is defined as $I_B = I_Z \odot M + I_T \odot (1 - M)$. This loss function is a close approximation to the Poisson Eq. (1). First, the reconstructed pixels of I_Z is directly combined with I_T to construct I_B , and then the Laplacian filter is operated on the whole I_B , which takes the boundary pixels of I_T into account. This part satisfies the boundary constraint in Poisson equation. Second, we directly minimize the difference between the gradient of I_B and the addition of gradients of I_S and I_T . Since the gradient of I_T is exactly the same as the gradient outside blending region in I_B ,

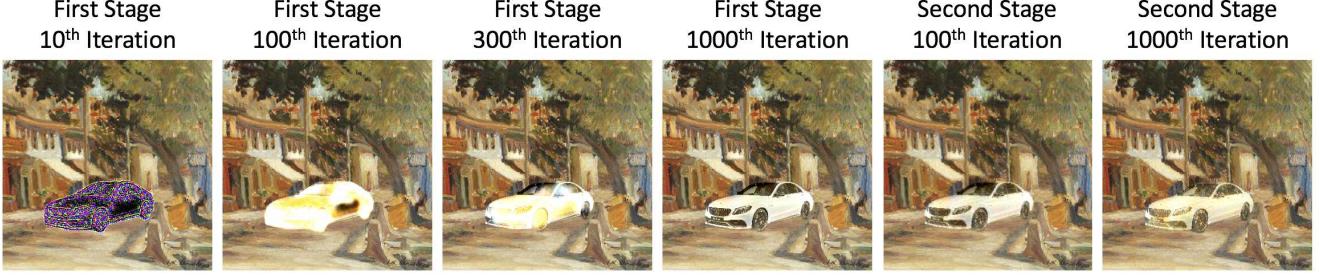


Figure 4: This is a demonstrate of the blending pixel reconstruction process at different iterations in stage one and stage two.

the loss is essentially computed within the blending region. The second part satisfies the gradient domain constraint in Eq. (1). An alternative way to implement this loss is to crop out the gradient of the blending region in I_B using M and compare it with ∇I_S . We think these two ways of implementation make little differences.

4.2. Style and Content Loss

In the original work [7], Gatys et al. proposed to transform the style of a source image using a style loss while preserving the content of the source using a content loss. In the first stage, the content and style losses are defined as follows,

$$\mathcal{L}_{cont} = \sum_{l=1}^L \frac{\alpha_l}{2N_l M_l} \sum_{i=1}^{N_l} \sum_{k=1}^{M_l} (F_l[I_Z] \odot M - F_l[I_S])_{ik}^2 \quad (7)$$

$$\mathcal{L}_{style} = \sum_{l=1}^L \frac{\beta_l}{2N_l^2} \sum_{i=1}^{N_l} \sum_{j=1}^{N_l} (G_l[I_Z] - G_l[I_T])_{ij}^2 \quad (8)$$

where \odot is the element-wise product, L is the number of convolutional layers, N_l is the number of channels in activation, M_l is the number of flattened activation values in each channel. $F_l[\cdot] \in \mathbb{R}^{N_l \times M_l}$ is an activation matrix computed from a deep network F at the l^{th} layer. $G_l[\cdot] = F_l[\cdot]F_l[\cdot]^T \in \mathbb{R}^{N_l \times N_l}$ denotes the Gram matrix of the corresponding activation matrix at the l^{th} layer. Intuitively, the Gram matrix captures the similarity relation between all pairs of channel features, which encodes the image style or texture and zero information about spatial structure. Finally, α_l and β_l are the weights that control the influence of each layer when computing the content and style loss.

In the second stage, the inputs to the content and style losses are different, which are defined as follows,

$$\mathcal{L}_{cont} = \sum_{l=1}^L \frac{\alpha_l}{2N_l M_l} \sum_{i=1}^{N_l} \sum_{k=1}^{M_l} (F_l[I_{BR}] - F_l[I_B])_{ik}^2 \quad (9)$$

$$\mathcal{L}_{style} = \sum_{l=1}^L \frac{\beta_l}{2N_l^2} \sum_{i=1}^{N_l} \sum_{j=1}^{N_l} (G_l[I_{BR}] - G_l[I_T])_{ij}^2 \quad (10)$$

where I_{BR} is the refined blending image, which is optimized with respect to I_B in terms of content and I_T in terms of style.

4.3. Regularization Loss

To stabilize the style transformation of blended region and encourage spatial smoothness, we further add a histogram loss proposed by [27] and a total variation loss proposed by [21] to regularize the generated image.

The histogram loss from [27] performs a histogram matching on each corresponding feature map at each output layer between the target image and the blended output. In this case we do it each iteration. Let $F_l(I_B)$ be the activation output for each layer of the blended image and $R_l(I_B)$ be the histogram matched activation between the blended output and the target image. This loss serves to stabilize the style transfer by matching the marginal distributions for each filter in each layer of the blended image with the marginal distribution of the target image. The activations we used for the histogram loss is the same activations as the style loss.

$$\mathcal{L}_{hist} = \sum_{l=1}^L \gamma_l \|F_l(I_B) - R_l(I_B)\|_F^2 \quad (11)$$

The total variation (tv) loss is used to remove the unwanted details while preserving the import information in the image. The loss objective is shown below.

$$\mathcal{L}_{tv} = \sum_{m=1}^H \sum_{n=1}^W |I_{m+1,n} - I_{m,n}| + |I_{m,n+1} - I_{m,n}| \quad (12)$$

4.4. Two-Stage Algorithm

In our algorithm, the first stage aims to seamlessly blend the object onto the background, and the second stage aims

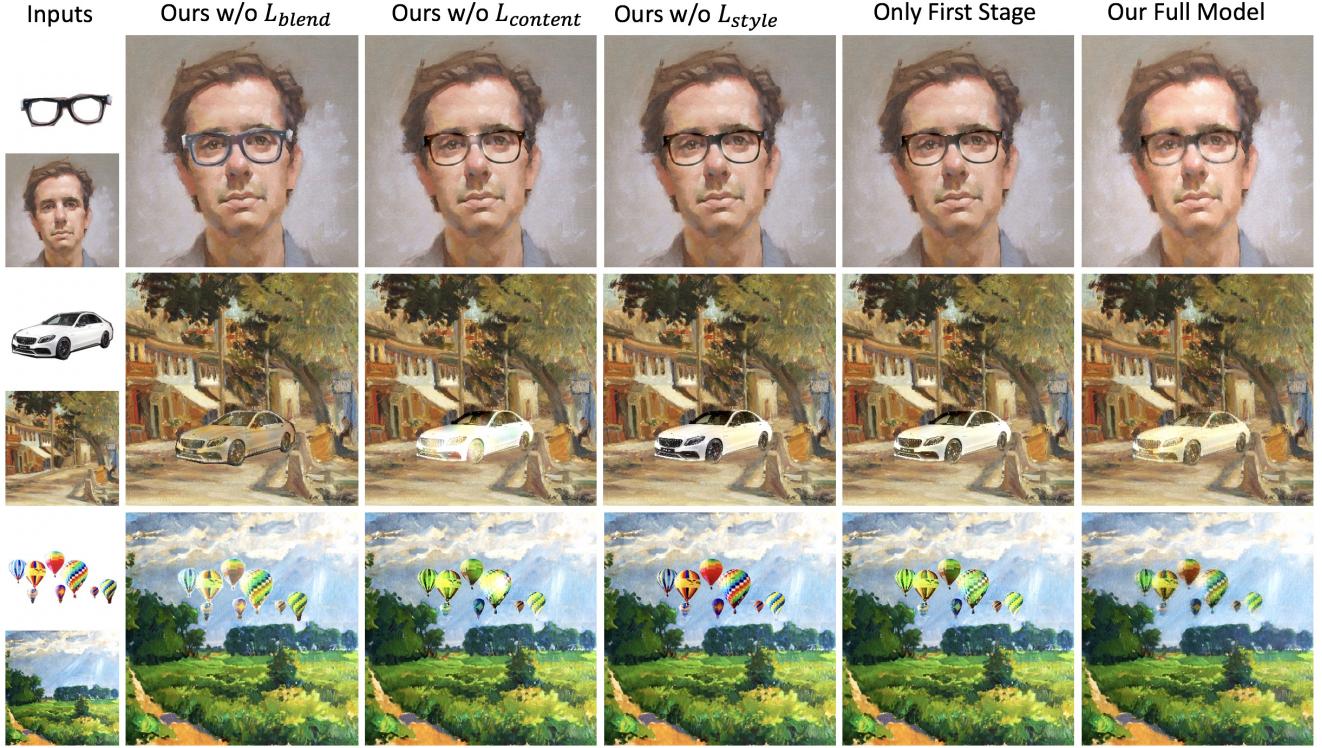


Figure 5: Ablation study of different loss functions as well as single-stage versus two-stage.

to further refine the texture and style of the blending region. The input to the first stage is a 2D random noise, while the input to the second stage is the final blending image from the first stage. We use the VGG-16 [28] network pretrained on ImageNet [4] to extract features for computing style and content losses. Regarding the gradient domain, we use Laplacian filter to compute the second order gradient of images to compute the gradient blending loss.

Algorithm 1 First Stage - Seamless Blending

Input: source image I_S , blending mask M , target image I_T
max iteration T , loss weights $\lambda_{grad}, \lambda_{cont}, \lambda_{style}, \lambda_{hist}, \lambda_{tv}$

Given: a gradient operator ∇ , a pretrained VGG network F

Output: blending image I_B

for $i \in [1:T]$ **do**

$$I_B = I_Z \odot M + I_T \odot (1 - M)$$

$$\mathcal{L}_{grad} = GradientLoss(I_B, I_S, I_T, \nabla) \text{ by Eq. 6}$$

$$\mathcal{L}_{cont} = ContentLoss(I_Z, M, I_S, F) \text{ by Eq. 7}$$

$$\mathcal{L}_{style} = StyleLoss(I_B, I_T, F) \text{ by Eq. 8}$$

$$\mathcal{L}_{hist} = HistogramLoss(I_B, I_T, F) \text{ by Eq. 11}$$

$$\mathcal{L}_{tv} = TVLoss(I_B) \text{ by Eq. 12}$$

$$L_{total} = \lambda_{grad} * \mathcal{L}_{grad} + \lambda_{cont} * \mathcal{L}_{cont} + \lambda_{style} * \mathcal{L}_{style} + \lambda_{hist} * \mathcal{L}_{hist} + \lambda_{tv} * \mathcal{L}_{tv}$$

$$I_Z \leftarrow L\text{-BFGS}\text{-Solver}(L_{total}, I_Z)$$

end

$$I_B = I_Z \odot M + I_T \odot (1 - M)$$

Algorithm 2 Second Stage - Style Refinement

Input: blending image I_B , target image I_T

max iteration T , loss weights $\lambda_{cont}, \lambda_{style}, \lambda_{hist}, \lambda_{tv}$

Given: a pretrained VGG network F

Output: refined blending image I_{BR}

$$I_{BR} = copy(I_B)$$

for $i \in [1:T]$ **do**

$$\mathcal{L}_{cont} = ContentLoss(I_{BR}, I_B, F) \text{ by Eq. 9}$$

$$\mathcal{L}_{style} = StyleLoss(I_{BR}, I_T, F) \text{ by Eq. 10}$$

$$\mathcal{L}_{hist} = HistogramLoss(I_{BR}, I_T, F) \text{ by Eq. 11}$$

$$\mathcal{L}_{tv} = TVLoss(I_{BR}) \text{ by Eq. 12}$$

$$L_{total} = \lambda_{grad} * \mathcal{L}_{grad} + \lambda_{cont} * \mathcal{L}_{cont} + \lambda_{style} * \mathcal{L}_{style} + \lambda_{hist} * \mathcal{L}_{hist} + \lambda_{tv} * \mathcal{L}_{tv}$$

$$I_{BR} \leftarrow L\text{-BFGS}\text{-Solver}(L_{total}, I_{BR})$$

end

We use VGG layers $conv_{1_2}$, $conv_{2_2}$, $conv_{3_3}$, $conv_{4_3}$ to compute style loss and $conv_{2_2}$ to compute content loss. We set maximum iteration to be 1,000 in both stages, and optimize the loss with L-BFGS solver. The runtime on a 512×512 image takes about 5 minutes on a single NVIDIA GTX 1080Ti. We set $\lambda_{blend} = 10e5$, $\lambda_{cont} = 1$, $\lambda_{style} = 10e5$, $\lambda_{hist} = 1$, $\lambda_{tv} = 10e-6$ in the second stage, and set $\lambda_{cont} = 1$, $\lambda_{style} = 10e7$, $\lambda_{hist} = 1$, $\lambda_{tv} = 10e-6$ in the second stage. More carefully tuning might give a better combination of hyper-parameters, but this is not our focus.



Figure 6: This figure shows the comparison between strong baseline approaches and ours on the paintings. **Poisson Blending** refers to Poisson Image Editing[23]. **Poisson Blending + Style Transfer** refers to first blend the object onto the target image and then run style transfer[7] on the blending image. **Style Transfer + Poisson Blending** refers to first run style transfer[7] on the source image and then blend it onto the target image.

5. Experimental Results

We conduct ablation, comparison and user studies to show the robustness of our method and results. In our ablation study, we show the importance of each loss function and our second stage style refinement. As shown in Fig. (5), our full model clearly outperforms other baseline variants. There exists obvious artifacts on the blending boundary without our proposed blending loss. Some visual contents are wiped out without the content loss. The blending region and target background have inconsistent illumination and texture without style loss. Finally, our two-stage algorithm adds more style and texture in the blending image compared to the single-stage baseline.

In our comparison study, we first conduct experiments on paintings as shown in Fig. (6). In this study, we compare with several intuitive and strong baselines. "copy-and-paste" produces results that have obvious artificial boundary. "Poisson Blending" [23] is able to produce a smooth blending boundary, but produce inconsistent style and texture between the blending region and the background. "Poisson Blending + Style Transfer" [23, 7] produce more consistent style and texture than "Poisson Blending" but produces unpleasant color illumination during

Poisson Blending. "Style Transfer + Poisson Blending" [7, 23] produces consistent style but lack some of the original source image content. In contrast, our method produces most consistent style and texture while maintaining the source content.

In the experiment of real-world images, we compare our algorithm with several state-of-the-art image composite algorithms, as shown in Fig. (7). "Deep Image Harmonization" [30] is a data-driven harmonization algorithm that aims to adjust illumination of composite the region using the learned image prior. As seen, it is able to adjust the illumination of the blending region but is not able to smooth out the boundary, and thus has an artificial border. "Poisson Blending"[23] generates seamless boundary, but the background color "blends through" the blending region. "GP-GAN"[33] is a recent proposed blending algorithm that leverages Poisson Blending with Generative Adversarial Network, and is trained in a supervised way. However, this method can hardly generalize to our test cases and produces unrealistic boundary and illumination. Finally, our algorithm produces the best visual results in terms of blending boundary, texture, and color illumination.

To quantify the performance of our method in compar-

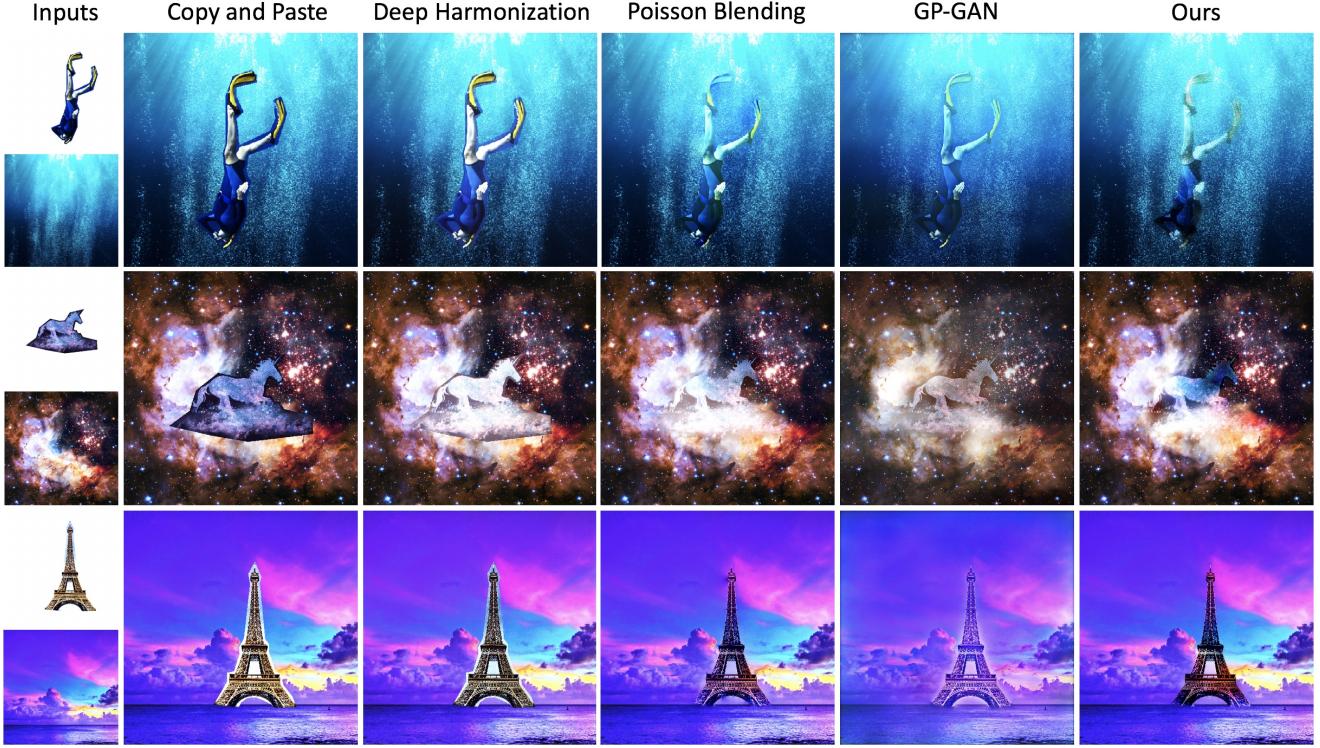


Figure 7: This figure shows the comparison between the state-of-the-art image composite approaches and ours on real-world images. **Deep Harmonization** refers to Deep Image Harmonization[26]. **Poisson Blending** refers to Poisson Image Editing[23]. **GP-GAN** refers to Gaussian-Poisson Generative Adversarial Network[33].

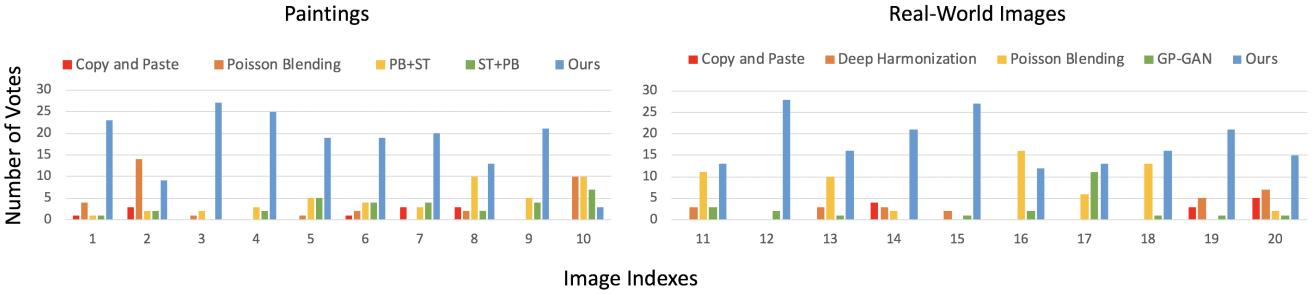


Figure 8: This figure shows our user study results. Two histograms on the left and right show the quantitative comparison in paintings and real-world images respectively. "PB+ST" denotes "Poisson Blending + Style Transfer" and "ST+PB" denotes "Style Transfer + Poisson Blending".

ison to other algorithms, we recruited thirty users for user studies using 20 sets of images, 10 real-world and 10 painting style target images. Each user is asked to pick one composite image they think is the most realistic from five different images generated by five different algorithms. From the histograms (8), we see that the images that have the highest votes are almost all images generated by our method. Specifically, nine out of ten images our method generated using real-world target images received the highest votes from the users, and eight out of ten images our method generated using painting-style target images received the highest votes from the users. The results indicate that our

method is more preferable than other methods 80 ~ 90% of the time.

6. Conclusion

In this paper, we propose a novel gradient blending loss and a two-stage algorithm that generates blending image on-the-fly with a L-BFGS solver. Our algorithm does not rely on any training data and thus can generalize to any real-world images or paintings. Through user study, our algorithm is proven to outperform strong baselines and state-of-the-arts approaches. We believe that our work could be applied in common image editing tasks and opens new possibility for users to easily compose artistic works.

Acknowledgement

We gratefully appreciate the support from Honda Research Institute Curious Minded Machine Program. We also gratefully acknowledge a GPU donation from NVIDIA.

References

- [1] A. Agarwala, M. Dontcheva, M. Agrawala, S. Drucker, A. Colburn, B. Curless, D. Salesin, and M. Cohen. Interactive digital photomontage. In *ACM Transactions on Graphics (ToG)*, volume 23, pages 294–302. ACM, 2004.
- [2] M. Arjovsky, S. Chintala, and L. Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223, 2017.
- [3] P. Burt and E. Adelson. The laplacian pyramid as a compact image code. *IEEE Transactions on communications*, 31(4):532–540, 1983.
- [4] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.
- [5] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.
- [6] R. Fattal, D. Lischinski, and M. Werman. Gradient domain high dynamic range compression. In *ACM transactions on graphics (TOG)*, volume 21, pages 249–256. ACM, 2002.
- [7] L. A. Gatys, A. S. Ecker, and M. Bethge. Image style transfer using convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2414–2423, 2016.
- [8] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *Advances in neural information processing systems*, pages 2672–2680, 2014.
- [9] X. Huang and S. Belongie. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 1501–1510, 2017.
- [10] S. Izuka, E. Simo-Serra, and H. Ishikawa. Globally and locally consistent image completion. *ACM Transactions on Graphics (ToG)*, 36(4):107, 2017.
- [11] J. Jia, J. Sun, C.-K. Tang, and H.-Y. Shum. Drag-and-drop pasting. *ACM Transactions on Graphics (TOG)*, 25(3):631–637, 2006.
- [12] J. Johnson, A. Alahi, and L. Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision*, pages 694–711. Springer, 2016.
- [13] M. Kazhdan and H. Hoppe. Streaming multigrid for gradient-domain operations on large images. In *ACM Transactions on Graphics (TOG)*, volume 27, page 21. ACM, 2008.
- [14] J. Kim, J. Kwon Lee, and K. Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.
- [15] R. Köhler, C. Schuler, B. Schölkopf, and S. Harmeling. Mask-specific inpainting with deep neural networks. In *German Conference on Pattern Recognition*, pages 523–534. Springer, 2014.
- [16] A. Levin, A. Zomet, S. Peleg, and Y. Weiss. Seamless image stitching in the gradient domain. In *European Conference on Computer Vision*, pages 377–389. Springer, 2004.
- [17] Y. Li, C. Fang, J. Yang, Z. Wang, X. Lu, and M.-H. Yang. Universal style transfer via feature transforms. In *Advances in neural information processing systems*, pages 386–396, 2017.
- [18] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro. Image inpainting for irregular holes using partial convolutions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 85–100, 2018.
- [19] F. Luan, S. Paris, E. Shechtman, and K. Bala. Deep photo style transfer. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4990–4998, 2017.
- [20] F. Luan, S. Paris, E. Shechtman, and K. Bala. Deep painterly harmonization. In *Computer Graphics Forum*, volume 37, pages 95–106. Wiley Online Library, 2018.
- [21] A. Mahendran and A. Vedaldi. Understanding deep image representations by inverting them. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5188–5196, 2015.
- [22] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros. Context encoders: Feature learning by inpainting. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2536–2544, 2016.
- [23] P. Pérez, M. Gangnet, and A. Blake. Poisson image editing. *ACM Transactions on graphics (TOG)*, 22(3):313–318, 2003.
- [24] F. Pitié and A. Kokaram. The linear monge-kantorovich linear colour mapping for example-based colour transfer. 2007.
- [25] T. Porter and T. Duff. Compositing digital images. In *ACM Siggraph Computer Graphics*, volume 18, pages 253–259. ACM, 1984.
- [26] E. Reinhard, M. Adhikhmin, B. Gooch, and P. Shirley. Color transfer between images. *IEEE Computer graphics and applications*, 21(5):34–41, 2001.
- [27] E. Risser, P. Wilmot, and C. Barnes. Stable and controllable neural texture synthesis and style transfer using histogram losses. *arXiv preprint arXiv:1701.08893*, 2017.
- [28] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [29] R. Szeliski, M. Uyttendaele, and D. Steedly. Fast poisson blending using multi-splines. In *2011 IEEE International Conference on Computational Photography (ICCP)*, pages 1–8. IEEE, 2011.
- [30] Y.-H. Tsai, X. Shen, Z. Lin, K. Sunkavalli, X. Lu, and M.-H. Yang. Deep image harmonization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3789–3797, 2017.

- [31] M. Uyttendaele, A. Eden, and R. Skeliski. Eliminating ghosting and exposure artifacts in image mosaics. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 2, pages II–II. IEEE, 2001.
- [32] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 0–0, 2018.
- [33] H. Wu, S. Zheng, J. Zhang, and K. Huang. Gp-gan: Towards realistic high-resolution image blending. *arXiv preprint arXiv:1703.07195*, 2017.
- [34] L. Xu, J. S. Ren, C. Liu, and J. Jia. Deep convolutional neural network for image deconvolution. In *Advances in Neural Information Processing Systems*, pages 1790–1798, 2014.
- [35] R. A. Yeh, C. Chen, T. Yian Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do. Semantic image inpainting with deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5485–5493, 2017.
- [36] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. Generative image inpainting with contextual attention. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5505–5514, 2018.
- [37] X. Yu and F. Porikli. Ultra-resolving face images by discriminative generative networks. In *European Conference on Computer Vision*, pages 318–333. Springer, 2016.
- [38] X. Yu and F. Porikli. Face hallucination with tiny unaligned images by transformative discriminative neural networks. In *AAAI*, volume 2, page 3, 2017.
- [39] X. Yu and F. Porikli. Hallucinating very low-resolution unaligned and noisy face images by transformative discriminative autoencoders. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3760–3768, 2017.
- [40] L. Zhang, J. Wang, and J. Shi. Multimodal image outpainting with regularized normalized diversification. *arXiv preprint arXiv:1910.11481*, 2019.