Multi-level Discriminator and Wavelet Loss for Image Inpainting with Large Missing Area

Junjie Li and Zilei Wang(✉)

University of Science and Technology of China, Hefei, China
hnljj@mail.ustc.edu.cn, zlwang@ustc.edu.cn

Abstract. Recent image inpainting works have shown promising results thanks to great advances of generative adversarial networks (GANs). However, these methods would still generate distorted structures or blurry textures for the situation of large missing area, which is mainly due to the inherent difficulty to train GANs. In this paper, we propose a novel multi-level discriminator (MLD) and wavelet loss (WT) to improve the learning of image inpainting generators. Our method does not change the structure of generator and only works in the training phase, which thus can be easily embedded into sophisticated inpainting networks and would not increase the inference time. Specifically, MLD divides the mask into multiple subregions and then imposes an independent discriminator to each subregion. It essentially increases the distribution overlap between the real images and generated images. Consequently, MLD improves the optimization of GANs by providing more effective gradients to generators. In addition, WT builds a reconstruction loss in the frequency domain, which can facilitate the training of image inpainting networks as a regularization term. Consequently, WT can enforce the generated contents to be more consistent and sharper than the traditional pixel-wise reconstruction loss. We integrate WLD and WT into off-the-shelf image inpainting networks, and conduct extensive experiments on CelebA-HQ, Paris StreetView, and Places2. The results well demonstrate the effectiveness of the proposed method, which achieves state-of-the-art performance and generates higher-quality images than the baselines.

Keywords: Image inpainting · Generative Adversarial Network (GAN) · Multi-level discriminator · Wavelet transform · Loss function

J. Li—Student.

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-88010-1_5) contains supplementary material, which is available to authorized users.

H. Ma et al. (Eds.): PRCV 2021, LNCS 13021, pp. 53–65, 2021.
https://doi.org/10.1007/978-3-030-88010-1_5

1 Introduction

For image inpainting, it is unnecessary to require the generated image to be exactly same as the groundtruth image, but the generated content to fill the hole is expected to be natural and consistent with the surrounding content. Generative Adversarial Network (GAN) [7] is an excellent generative model and has shown powerful ability to generate natural and high-quality images. Therefore, a lot of works [9,11,19,20,24] introduce GAN to the image inpainting task. Typically, the networks in these methods consist of two main components, namely, generator and discriminator. Most of the methods [9,19,20,24] focus on how to construct a better generator, and rarely concern the design of discriminator. According to the structure of generators, we can divide these methods into two broad categories: one-stage model and two-stage model. The one-stage methods adopt one encoder-decoder as a generator, *e.g.*, CE [19] and GMCNN [20]. They usually cannot work well for complex images (such as the Places2 dataset [27]) due to limited capacity to express complicated patterns of image contents. Differently, the two-stage models, *e.g.*, EdgeConnect (EC) [9] and HighFill [23], employ two encoder-decoder networks to generate natural results. However, either one-stage or two-stage models still would suffer from serious artifacts, especially for the images with large missing area, as shown in Fig. 1.

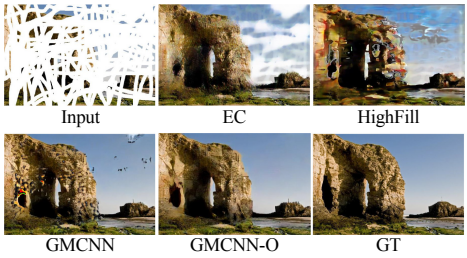

Fig. 1. Inpainting results of different methods. *-O means that our proposed method is applied. It can be seen that the current methods have great difficulties in the case of large area missing. On the contrary, our method can handle this situation well.

According to [2], an important reason why large-area missing images is difficult to repair is that the distribution of generated images and the distribution of real images are difficult to overlap, which leads to optimization difficulties. We will explain it in detail in Sect. 3.1. In addition, the pixel-wise ℓ_1 or ℓ_2 loss can not capture high-frequency information, which can easily produce blurred results. Based on the above two issues, we mainly focus on how to better learn a generator, *i.e.*, designing an method to more effectively train image inpainting networks. Our method is only involved in the training phase and thus would not increase the time of forward inference.

In this paper, we propose a new discriminator and loss function to address the aforementioned challenges of image inpainting. First, we propose a multi-level discriminator (MLD) to increase the overlap between the generated and real distributions so that the generator can be more effectively trained. MLD partitions the missing regions into multiple parts, and then we separately build the adversarial loss for each of them. Consequently, a discriminator for larger missing area is decomposed into multiple sub-discriminators with smaller missing area.

Second, we propose to use wavelet coefficients to represent image contents and then wavelet loss (WT) to enforce the content consistency between the generated image and raw image. Different from the ℓ_1 and ℓ_2 loss, WT is to compute the reconstruction loss in the frequency domain, in which each component represents the information of regions rather than pixels. For example, the low-frequency component can describe the global topology information and the high-frequency one can describe the local texture details [10]. Note that our proposed method in this work are orthogonal to existing image inpainting generators, and thus can be integrated into off-the-shelf networks. Particularly, we would investigate the combinations with GMCNN [20] and CA [24] in this paper due to their good performance. To our best knowledge, this work is the first attempt to handle the large-hole problem of image inpainting from the perspective of distribution overlap.

We experimentally evaluate the proposed method on three benchmark datasets: CelebA-HQ [13], Paris StreetView [4], and Places2 [27]. The results well demonstrate the effectiveness of our proposed method, which can produce much higher-quality images than the corresponding baseline.

The main contributions of this work are summarized as follows:

- We propose a multi-level discriminator, which increases the overlap between the generated and real distributions and thus can aid the generator networks to be more effectively trained.
- We propose wavelet loss for image inpainting networks to improve the constraints of content consistency. As a result, the artifact and blur of generated images can be greatly alleviated.
- We integrate the proposed method into representative inpainting networks, and experimentally show the effectiveness of our method on different types of images.

2 Related Work

2.1 Image Inpainting

There are two broad types of approaches in previous works, *i.e.*, traditional matching methods [3,17] and deep convolutional neural network (DCNN) based methods [9,11,16,19–22,25,26]. Traditional image inpainting methods like [3,5,6] work well for the images with small holes and consistent textures. However, these methods [5,6,14] are generally time-consuming due to iteratively searching for the most similar patch in surrounding areas. In sharp contrast, the DCNN-based method can better capture high-level information of images and efficiently synthesize images in an end-to-end manner. Wang *et al.* [20] proposed a Generative Multi-column Convolutional Neural Network (GMCNN) to capture different levels of information to generate more plausible synthesized content. CA [24] takes a coarse-to-fine structure with contextual attention module for image inpainting. Particularly, contextual attention can borrow the information from distant spatial location and thus overcomes the limitation of convolutional

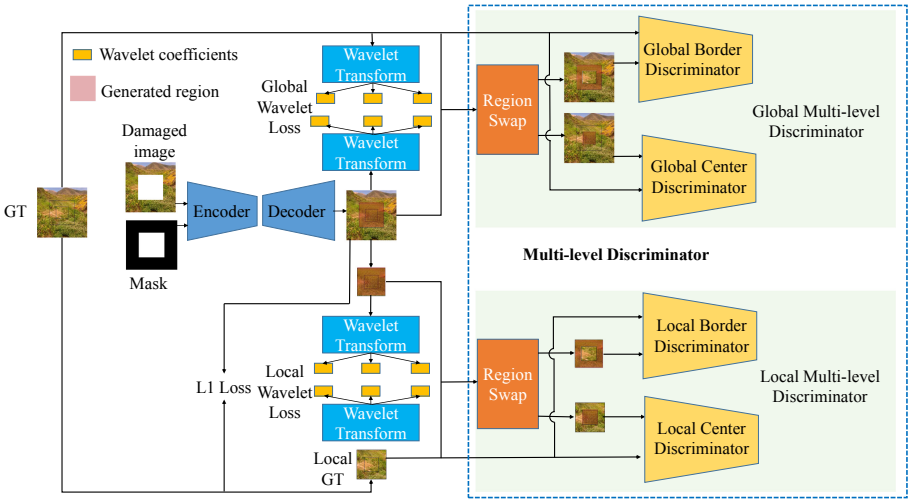

Fig. 2. Illustration of our proposed method. For an encoder-decoder image inpainting network, *multi-level discriminator* and *wavelet loss* are appended to the outputs of network to guide the training of generator. Here the global and local versions are used.

layers that only integrate local information. More recently, EdgeConnect (EC) [9] further improves the quality of generated images by a two-stage model, in which an edge generator is inserted before image completion network. However, these methods still suffer from serious artifacts, especially for the situation of large missing area. The work most similar to ours is PGN [25]. It proposes to progressively complete the masked area from edge to center, where multiple generators are introduced to take charge of generating different areas. Evidently, PGN would involve large memory consumption and high computational complexity, so even an image of 256×256 size is difficult to process. Our method only needs an additional discriminator during the training phase, so it does not increase the burden of memory too much.

2.2 Adversarial Training

Essentially, GAN acts as a min-max two-player game, which utilizes the adversarial loss to train the generator g_θ and the discriminator D alternatively [7]. Here the discriminator D is targeted to distinguish the generated images I^g and real images I^r, and the generator g_θ is targeted to produce an image I^g from a latent code z or a corrupted image I^c, to cheat the discriminator D. In practice, the training of GANs contains two main steps. The first one is to fix the generator g_θ and train a discriminator D with maximizing

$$L_D(I^r, I^g) = E_{x \sim P_r}[\log D(x)] + E_{x \sim P_g}[\log(1 - D(x))]. \tag{1}$$

The second one is to fix the discriminator D and train a generator g_θ with minimizing

$$L_g(I^g) = E_{g_\theta(x)\sim P_g}[\log(1 - D(g_\theta(x))]. \quad (2)$$

Here P_r and P_g represent the distributions of the real images and generated images, respectively. Ideally, after iterative training, the distributions of generated images and real images would be nearly same. However, GANs still face huge challenge in generating high-resolution images due to unstable training procedure [13]. A similar issue actually exists for image inpainting with large missing area [20], which consequently will generate sharp artifacts. In this work, we particularly propose a multi-level discriminator and wavelet loss to alleviate the unstable optimization of GANs and meanwhile capture local details.

3 Our Approach

In this work, we propose a novel multi-level discriminator and wavelet loss to facilitate the training of generator networks from two different perspectives. MLD is to improve the stability of adversarial training by increasing the distribution overlap between generated and real images, and the WT loss is to improve the regularization of content consistency and generator optimization by constructing a reconstruction loss in frequency domain. Figure 2 illustrates the overall architecture of our proposed method. Note that MLD and WT are only employed in the training phase and thus do not increase the inference time.

3.1 Multi-level Discriminator

Before elaborating on MLD, we first briefly explain the underlying reason of GANs to produce artifacts for image inpainting with large missing area. According to [2], when the generator g_θ is fixed, the optimal discriminator is

$$D^*(x) = \frac{P_r(x)}{P_r(x) + P_g(x)}. \quad (3)$$

As proven in [2], if P_g and P_r have no overlap (*i.e.*, are too easy to tell apart), the optimal discriminator D^* cannot provide effective gradients to guide the optimization of the generator g_θ. Here we provide a brief explanation. We define $U = \{x : P_r(x) > 0\}$, $\bar{U} = \{x : P_r(x) = 0\}$,

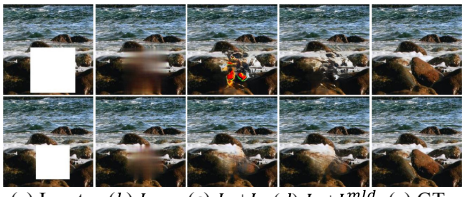

Fig. 3. Visual comparison of different loss functions to guide the learning of generator, where GMCNN is particularly used as the base model. L_1 and L_a represent the ℓ_1 loss and the global and local adversarial loss, respectively. L_a^{mld} denotes our proposed MLD adversarial loss. Compared to a small area missing, a large area missing is more likely to lead to sharp artifacts and MLD can make the generated images more natural.

$V = \{x : P_g(x) > 0\}$, and $\bar{V} = \{x : P_g(x) = 0\}$. Then D^* can be redefined as

$$D^*(x) = \begin{cases} 1, & \text{for} x \in U \cap \bar{V}, \\ 0, & \text{for} x \in \bar{U}, \\ \dfrac{P_r(x)}{P_r(x) + P_g(x)}, & \text{for} x \in U \cap V. \end{cases} \tag{4}$$

If P_g and P_r have no overlap, which means $U \cap V = \varnothing$, D^* would equal to either 1 or 0. Obviously, the gradient $\nabla_x D^*(x)$ would always be zero. According to the chain rule, we have

$$\nabla_\theta L_g = \nabla_D L_g \cdot \nabla_{g_\theta(x)} D^*(g_\theta(x)) \cdot \nabla_\theta g_\theta(x) = 0. \tag{5}$$

That is, the discriminator D^* cannot provide effective gradients to guide the optimization of the generator g_θ. Thus P_g and P_r need have enough overlap so that the generator can be effectively optimized.

For image inpainting, the difference of the generated image I^g and corresponding real image I^r is exactly the missing region represented by a mask. When the missing area is larger, therefore, it is difficult to effectively learn a generator since the gradients provided by the discriminator would have more or less random direction [2,18]. As a result, the generated images often present some artifacts, as shown in Fig. 3(c). In this work, we exploit the information of original images to improve the optimization of generators. The key idea is to increase the overlap of P_r and P_g by reducing the differential area between the generated image I^g and real image I^r. To be specific, we propose a multi-level discriminator, which decomposes the missing regions of an image into multiple parts and then separately imposes a discriminator for each part. Consequently, compared with the original discriminator over the whole missing regions, one single discriminator in MLD would possess an increased overlap between generated and real distributions due to smaller missing area to handle. Figure 3(d) shows the effect of MLD, which usually can generate more natural images, especially for the case of large missing area.

Formally, assume the generated image I^g corresponds to the real image I^r, and the missing mask is M. We divide the masked region into K subregions (*e.g.*, from border to center), as shown in Fig. 4. Then we build one discriminator for each subregion, *i.e.*, K discriminators would be constructed. Considering the computational complexity and inpainting performance, $K = 2$ is particularly used throughout our experiments. For such a setting, we define two virtual synthesized images $\bar{I}_c$ and $\bar{I}_b$ as

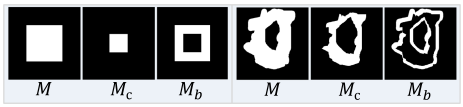

Fig. 4. Illustration of the MLD masks for different shapes. Here two levels are particularly used.

$$\bar{I}_c = M_c \odot I^g + (1 - M_c) \odot I^r, \tag{6}$$

$$\bar{I}_b = M_b \odot I^g + (1 - M_b) \odot I^r, \tag{7}$$

where M_c is the central mask (we set its area as a quarter of that of the original mask M in our implementation), M_b is the border mask with $M = M_b \vee M_c$, and $\odot$ denotes the spatially element-wise multiplication. Note that the division from border to center is particularly used for simplicity and we can also adopt other division strategy.

Evidently, the area of different regions between $\bar{I}_c(\bar{I}_b)$ and I^r is smaller than that between I^g and I^r, implying that the distributions of I^r and $\bar{I}_c(\bar{I}_b)$ have a larger overlap than that of I^g and I^r. We use $\bar{I}_c$ and $\bar{I}_b$ to replace I^g in the traditional adversarial loss separately. Then we can build a two-level discriminator. The adversarial loss functions become L_D^{mld} and L_g^{mld} for discriminator and generator, *i.e.*,

$$L_D^{mld} = L_D(I^r, \bar{I}_b) + L_D(I^r, \bar{I}_c), \tag{8}$$

$$L_g^{mld} = L_g(I^r, \bar{I}_b) + L_g(I^r, \bar{I}_c). \tag{9}$$

To better understand our proposed MLD, here we particularly compare MLD and GLD (Global and Local discriminator). The main difference is that GLD uses different discriminators to focus on multiple images with different scales, while MLD uses different discriminators to focus on multiple regions of the same image with different overlaps. To be specific, GLD targets to utilize multi-scale background contents, while MLD targets to increase the overlap of the distributions between generated and real images. From the perspective of implementation, GLD operates on the background regions to construct multi-scale contents, while MLD operates on the missing regions to construct different synthesized images. Therefore, GLD cannot alleviate the optimization difficulties caused by non-overlapping distributions. It often causes sharp artifacts in the case of large areas, as shown in Fig. 3(c). However, MLD can greatly alleviate this problem, as shown in Fig. 3(d). In addition, due to the existence of local discriminator, GLD is not applicable for irregular masks. But our MLD can still work well for such masks, as shown in Fig. 4.

3.2 Wavelet Loss

Wavelet transform can extract multi-frequency signals of images [1,10,15], and thus can model the consistency for different details. Here we introduce wavelet transform into image inpainting networks to explicitly model high frequency components, and it is shown in our experiments to work well in preventing image blurring. Note that we do not change the discriminator which still works in image space, and our proposed WT serves as a reconstruction loss to guide the optimization of generator together with adversarial loss. In our implementation, we particularly adopt the Harr wavelet for the sake of simplicity.

Formally, given an image I, wavelet transform decomposes it with L levels, and each level l consists of four types of coefficients with $C = \{C_1, C_2, C_3, C_4\}$ [10]. Then we define the wavelet loss as

$$L_{wt} = \sum_{l=1}^{L} \sum_{m \in C} \left\| \bar{w}_m^l - w_m^l \right\|_1, \tag{10}$$

where $\bar{w}_m^l$ and w_m^l represent the wavelet transform coefficients of the m-th component in the l-th level for the completed image and raw one, respectively.

4 Experiments

In this section, we evaluate the proposed method on three challenging datasets, including CelebA-HQ faces [13], Paris StreetView [4], and Places2 [27]. In particular, we adopt the one-stage model GMCNN [20] and two-stage model CA [24] as the baseline networks due to their excellent image inpainting performance. We directly add the WT and MLD modules to CA and GMCNN, resulting in

Table 1. Performance comparison of different methods on the Places2 dataset. Here ↓ means the lower is better, and ↑ means the higher is better. ∗ represents our retrained model.

		Thin Irregular		Thick Irregular		Center
	Method	30–40%	40–50%	30–50%	50–70%	Rectangle
PSNR↑	HighFill [23]	25.839	24.337	22.488	21.198	17.142
	CA [24]	–	–		–	18.145
	EC [9]	26.565	25.046	**23.707**	22.428	**19.069**
	GMCNN [20]	26.732	25.294	21.869	20.656	–
	CA* [24]	–	–		–	17.974
	CA-O	–	–	–	–	18.931
	GMCNN* [20]	26.665	24.793	22.291	20.751	17.998
	GMCNN-O	**27.146**	**25.569**	23.679	**22.503**	18.877
SSIM↑	HighFill [23]	0.877	0.835	0.802	0.740	0.459
	CA [24]	–	–	–	–	0.537
	EC [9]	0.883	0.847	0.824	0.771	0.548
	GMCNN [20]	0.891	0.860	0.813	0.764	–
	CA* [24]	–	–		–	0.523
	CA-O	–	–	–	–	0.553
	GMCNN* [20]	0.889	0.851	0.820	0.771	0.549
	GMCNN-O	**0.895**	**0.863**	**0.835**	**0.791**	**0.578**
FID↓	HighFill [23]	12.802	19.224	27.525	47.526	22.845
	CA [24]	–	–	–	–	**7.696**
	EC [9]	7.190	11.402	13.994	24.227	9.288
	GMCNN [20]	5.476	8.409	16.509	27.885	–
	CA* [24]	–	–		–	9.149
	CA-O	–	–	–	–	8.231
	GMCNN* [20]	5.293	9.087	18.336	33.918	8.958
	GMCNN-O	**5.122**	**7.872**	**11.703**	**18.560**	8.011

GMCNN-O and CA-O. Both GMCNN and CA use the global and local discriminators for the rectangular masks, and correspondingly we replace them with the global-MLD and local-MLD. CA does not support the irregular masks due to using local discriminator, while GMCNN supports the irregular masks by adopting PatchGAN [12] as discriminator. Hence we conduct the experiments on the irregular masks only using GMCNN. For the irregular masks, we use a 16×16 kernel to corrode the masks to produce the M_b and M_c, as shown in Fig. 4. For comparison and fairness, in all experiments, we keep the loss function of the original method unchanged.

4.1 Experimental Settings

4.2 Performance Evaluation

In this section, we evaluate our proposed method by comparing with five representative baseline methods, including HighFill [23], GMCNN [20], EC [9] and CA [24]. For fair comparison, we retrain the baseline CA and GMCNN using our data besides testing the models provided by authors. The retrained models are marked by $*$ in the experimental results.

Fig. 5. Compared to the thick masks, the center of the thin masks is closer to the existing regions.

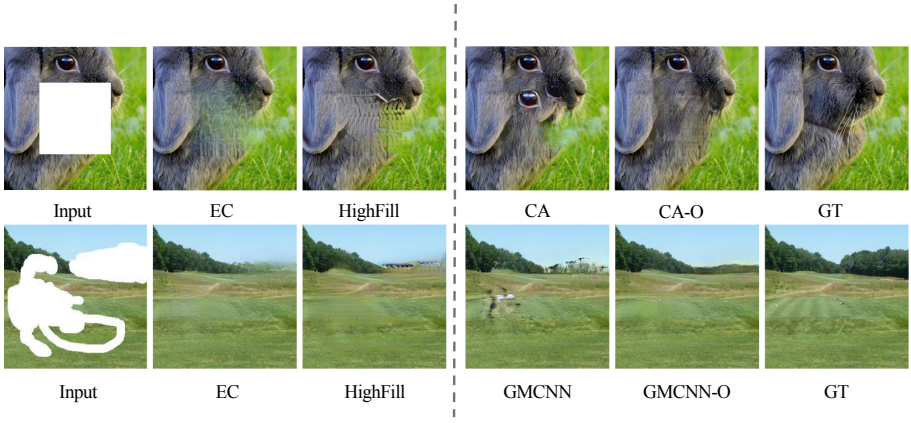

Fig. 6. Example inpainting results of different methods on Places2 with the rectangle masks.

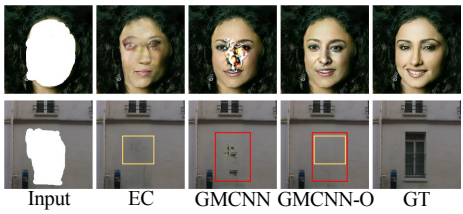

Fig. 7. Performance of different combinations of MLD and wavelet loss.

Quantitative Comparison. We measure the quality of inpainting results using the following metrics: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and Frchet Inception Distance (FID) [8]. For PSNR and SSIM, higher values indicate better performance, while for FID, the lower the better. For Places2, 10,000 images randomly selected from the validation set for test. For Paris StreetView and CelebA-HQ, we follow GMCNN [20] to construct test set.

In order to comprehensively evaluate various methods, we test on both the rectangular masks and irregular masks. All images are resized to 256×256. For the rectangular masks, all masked holes are placed at the center of images with the size of 128×128. For the irregular masks, as shown in Fig. 5, we use the thin masks and thick masks separately. For the thin mask, the hole-to-image ratio ranges from 0.3 to 0.5. For the thick mask, the hole-to-image ratio ranges from 0.3 to 0.7.

Table 1 gives the results of different methods on the Places2 dataset, where PSNR and SSIM are computed on the hole regions for the rectangular masks and on the whole images for the irregular masks. From the results, we have the following observations. First, compared with the baseline GMCNN* and CA* using the same data, our GMCNN-O and CA-O can bring significant performance improvement, especially for the thick masks. Because in this case, the problem of non-overlapping distribution is more serious, and it can better reflect the advantages of our method. Second, GMCNN-O achieves better performance than the current state-of-the-art methods including two-stage method EC and HighFill for almost all settings of irregular masks, which shows the ability of our method to boost the advanced methods. In particular, HighFill [23] is a good method to recover the image with large-missing area. However, the key for HighFill to reduce artifacts is that on the reduced image (256×256), the missing area is small enough to be easily repaired. Finally, a super-resolution module can be used to obtain a inpaited image of the original size. Therefore, it requires that the proportion of the missing area on the original image cannot be large (generally not more than 25%), otherwise the reduced missing area will still be large, resulting in artifacts, as shown in Fig. 6.

Qualitative Comparison. For the image inpainting task, there is no a comprehensive metric that can accurately measure the quality of completed images. Here we qualitatively compare our model with the baselines. Figure 6, Fig. 7 give some examples of different methods with different image masks. It can be seen that most of existing methods usually suffer from discontinuous structure and mutant color (black or white). Our proposed method can effectively reduce

the artifacts and improve the quality of the generated images. For more results, please refer to the supplementary material.

4.3 Ablation Study

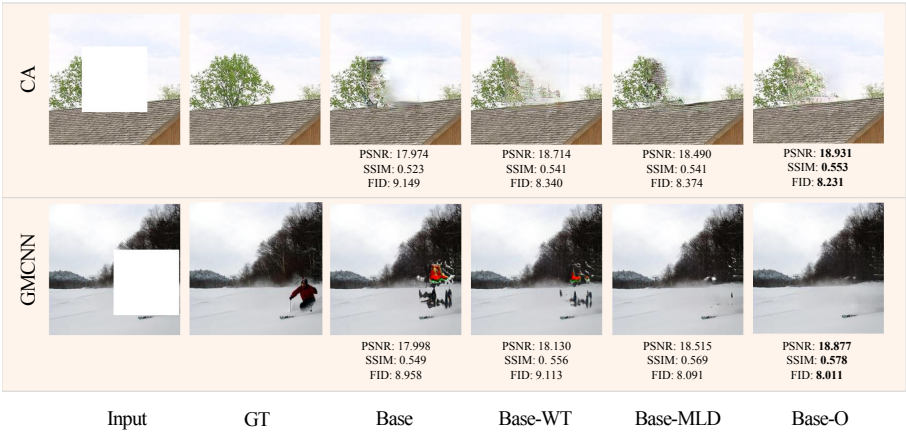

Fig. 8. Performance of different combinations of MLD and wavelet loss.

In this section, we investigate the effects of the main components of our proposed methods. Since the rectangle masks with a large hole are very challenging, we particularly use such masks with a size of 128 × 128 to conduct ablation experiments. We use the metrics PSNR, SSIM, and FID to quantitatively compare different methods. All the metrics are reported on the test set. For convenience, we put the SSIM, PSNR, and FID together with the example generated images. For fair comparison, we use retrained CA* and GMCNN* as the baseline. In the case of no ambiguity, we omit $*$ here.

Effect of MLD and Wavelet Loss. Here we evaluate the effect of our proposed MLD and wavelet loss by comparing their combinations with the baseline GMCNN and CA. To be specific, we separately impose MLD and wavelet loss to the baselines (denoted by *-MLD and *-WT respectively), and then use both of them (denoted by *-O). Particularly, we use the Place2 dataset due to its challenge. Figure 8 provides the results for different models. It can be seen that MLD and wavelet loss both contribute to the performance of image inpainting. In particular, GMCNN may produce some black and red regions, wavelet loss can provide more details, and MLD can make the images more natural. Finally, the combination of our proposed MLD and WT performs best, which can generate high-quality images.

Number of MLD Levels. Here we set the level of MLD K from $\{1, 2, 3, 4\}$ to analysis the impact of different MLD levels. We adopt CA with the rectangle

masks as the baseline model and CelebA-HQ is used for evaluation. Figure 9 give the results of different MLD levels. Compared with the original CA (*i.e.*, $K = 1$), the MLD can significantly increase the quantity of generated images. Considering the computational complexity, we particularly adopt $K = 2$ in our experiments that actually can produce satisfactory results.

Level of Wavelet Transform. In this section, we analyze the impact of each wavelet decomposition level. Here *w/o WT* represents the baseline model without wavelet loss, *w/o WT-n* represents the baseline model which uses 3 wavelet levels but removes the results of the n-th level, and *WT-all* represents the baseline model with all the 3 wavelet decomposition levels. Figure 10 gives the results of different settings on CelebA-HQ. It can be seen that each wavelet level contributes to the quality of inpainting results.

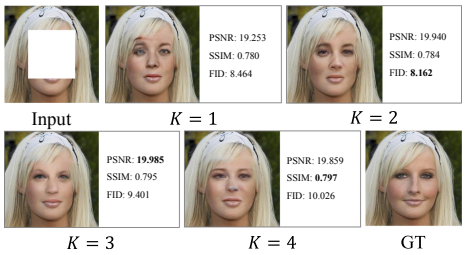

Fig. 9. Results of different MLD levels on CelebA-HQ.

5 Conclusion

In this paper, we proposed two novel techniques to improve the training of GANs for image inpainting, namely multi-level discriminator. Specifically, MLD can improve the stability of network training by increasing the distribution overlap between the generated images and real images. WT can achieve a good trade-off between the sharpness and naturalness of generated images by exploiting the frequency-domain information as the reconstruction loss. We experimentally verified the effectiveness of the proposed MLD and WT, which can generate high-quality images for both the rectangle and irregular masks.

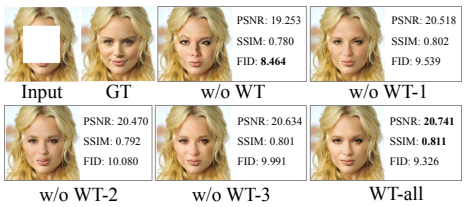

Fig. 10. Results of removing single wavelet level on CelebA-HQ.

Acknowledgments. This work is supported by the National Natural Science Foundation of China under Grant 61836008 and 61673362, Youth Innovation Promotion Association CAS (2017496).

References

1. Antonini, M., Barlaud, M., Mathieu, P., Daubechies, I.: Image coding using wavelet transform. TIP (1992)

2. Arjovsky, M., Bottou, L.: Towards principled methods for training generative adversarial networks. In: ICLR (2017)
3. Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B.: Patchmatch: a randomized correspondence algorithm for structural image editing. TOG (2009)
4. Doersch, C., Singh, S., Gupta, A., Sivic, J., Efros, A.A.: What makes paris look like paris? TOG (2012)
5. Efros, A.A., Freeman, W.T.: Image quilting for texture synthesis and transfer. In: SIGGRAPH (2001)
6. Ghorai, M., Samanta, S., Mandal, S., Chanda, B.: Multiple pyramids based image inpainting using local patch statistics and steering kernel feature. TIP (2019)
7. Goodfellow, I., et al.: Generative adversarial nets. In: NeurIPS (2014)
8. Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., Hochreiter, S.: Gans trained by a two time-scale update rule converge to a local nash equilibrium. In: NeurIPS (2017)
9. Hu, J., Shen, L., Sun, G.: Edgeconnect: Generative image inpainting with adversarial edge learning. In: ICCV Workshop (2019)
10. Huang, H., He, R., Sun, Z., Tan, T.: Wavelet-srnet: a wavelet-based cnn for multi-scale face super resolution. In: ICCV (2017)
11. Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. TOG (2017)
12. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: ICCV (2017)
13. Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. In: ICLR (2018)
14. Kwatra, V., Essa, I., Bobick, A., Kwatra, N.: Texture optimization for example-based synthesis. In: TOG (2005)
15. Lewis, A.S., Knowles, G.: Image compression using the 2-d wavelet transform. TIP (1992)
16. Liu, G., Reda, F.A., Shih, K.J., Wang, T.C., Tao, A., Catanzaro, B.: Image inpainting for irregular holes using partial convolutions. In: ECCV (2018)
17. Liu, J., Yang, S., Fang, Y., Guo, Z.: Structure-guided image inpainting using homography transformation. TMM (2018)
18. Odena, A., Olah, C., Shlens, J.: Conditional image synthesis with auxiliary classifier gans. In: ICML (2017)
19. Pathak, D., Krahenbuhl, P., Donahue, J., Darrell, T., Efros, A.A.: Context encoders: Feature learning by inpainting. In: CVPR (2016)
20. Wang, Y., Tao, X., Qi, X., Shen, X., Jia, J.: Image inpainting via generative multi-column convolutional neural networks. In: NeurIPS (2018)
21. Xie, J., Xu, L., Chen, E.: Image denoising and inpainting with deep neural networks. In: NeurIPS (2012)
22. Yang, Y., Guo, X.: Generative landmark guided face inpainting. In: PRCV (2020)
23. Yi, Z., Tang, Q., Azizi, S., Jang, D., Xu, Z.: Contextual residual aggregation for ultra high-resolution image inpainting. In: CVPR (2020)
24. Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., Huang, T.S.: Generative image inpainting with contextual attention. In: CVPR (2018)
25. Zhang, H., Hu, Z., Luo, C., Zuo, W., Wang, M.: Semantic image inpainting with progressive generative networks. In: ACM MM (2018)
26. Zheng, C., Cham, T.J., Cai, J.: Pluralistic image completion. In: CVPR (2019)
27. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: a 10 million image database for scene recognition. TPAMI (2018)