Unsupervised Structure Confidence Sampling for Image Inpainting

Xinrong Hu[1,2], Tao Wang[1,2], Jinxing Liang[1,2]*, Junjie Jin[1,2], Junping Liu[1,2], Tao Peng[1,2], Yuanjun Xia[1,2]
1. Engineering Research Center of Hubei Province for Clothing Information
2.School of Computer Science & Artificial Intelligence, Wuhan Textile University
E-mail: hxr@wtu.edu.cn, wtadota@163.com, jxliang@wtu.edu.cn,junjie.jin@qq.com, ljp@wtu.edu.cn, pt@wtu.edu.cn, xiaujun@163.com

***Abstract*—Context: Current image inpainting methods show great effects in different applications such as image editing, object removal, art creation and soon, but lack of editability of the inpainting results and convincing unsupervised features. Objective: To improve the existing methods, an optimized framework for image inpainting purpose is proposed based on hierarchical variational auto-encoder (VAE) as well as some optimization strategies. Method: Firstly, the VAE is used to extract the distribution of the features of the masked image in different scales, however, it will cause the distribution offset of extracted features which is unfavorable for image inpainting. Therefore, an optimal strategy that sampling the effective feature and invalid feature separately to avoid the offset of feature distribution of the masked image is integrated into the framework. To further improve the formulation of the proposed framework, the same encoder is used to realize the conversion from two domains to the same domain, which is a benefit to enhance the extraction of effective feature regions. In addition, we also introduce the cycle consistency constraints and GAN constraints into the framework to supervise the inpainting process. Result: Experimental results on the available image dataset demonstrate the effectiveness and superiority of the proposed framework.**

Keywords- Image inpainting; Auto-encoder; Self-supervision;

I. INTRODUCTION

Image inpainting is always a fundamental challenge in the field of computer vision. The key problem of image inpainting is how to ensure the integrity and consistency of the filled area to the adjacent, including the content, color attribute as well as tone of the image. Methods that producing incomplete filled effects or artificial effects between filled area and surrounding area are no good ones. Image inpainting has been widely used in many fields such as image editing, object removal, art creation and other tasks [1-5]. However, most of the existing methods that rely on GAN-based [11] image generation of which lose the adaptation to a variety of applications. Fortunately, in recent years VAE-based [16] image generation strategy based on probability introduce the new path to the current researches on image inpainting.

Previously, the GAN-based image inpainting can be divided into the following two categories: the one-stage inpainting method and the progressive method. For the one-stage method [6-10], its hypothesis that all globally valid image information can be obtained at once for image reconstruction. Although they can ensure the consistency of generated information and context semantics, these methods suffer from the problem of pixel discontinuity and semantic gap of the inpainting result, which can be found in the presence of many missing regions [17]. The reason comes from the large pixel difference between the known and missing regions, which leads to a weak correlation and further produces the hole regions.

Different from utilizing prior available features from input, the progressive methods [23, 25-28] consider that missing regions can not be filled completion at once. Therefore, these kinds of methods gradually reasoning feature value in holes region until the missing region is filled. However, these methods fail to consider the difference and correlation between filled areas and some other certain regions. Most importantly, because the image inpainting is iteratively conducted at the image level, the computational cost is very expensive. These kinds of methods always need more efficient computing environment.

Significantly different from the GAN-based method, currently the VAE-based image generation methods [12-15] can be able to generate novel and diverse image samples by mapping the noise of normal distribution to the image. However, if without some optimization and improvement, these methods cannot be directly used for inpainting of diversified scene images, the reasons are listed as follows. Firstly, when applied these kinds of method to inpainting of diversified scene images, the condition label is the masked image itself and there are no paired training images in the training dataset for each condition label. It will lead to there are no conditional training datasets that can explicitly express the condition distribution for the diversity masked images. Secondly, there are strong constraints for inpainting of diversified scene images, which means that the repaired images should keep integrity and consistency in color and texture with the masked image, therefore, it is more vulnerable to suffer from mode collapse than typical image generation.

Based on the limitations of the currents methods described above, we propose an unsupervised image inpainting framework in this paper based on NVAE [14]. The proposed framework relies on the assumption of implicit space sharing of the three domains and is based on domain transformation and differentiated sampling for finer generation effects. For the original NVAE, the feature information decoded by the upper sampling is more complete, the sampling points of the lower sampling will become unavailable due to the presence of the holes. Different from the original NVAE, the proposed inpainting framework firstly fill the hole area and then combine and derive the posterior distribution based on the feature matching strategy. After that, we realize the sharing of the same encoder within two of the domains in the form of

* Corresponding author.
DOI reference number: 10.18293/SEKE2022-026

additional weights since the existence of ternary domains brings too much coding space. At last, the patch discriminator is used to guide image generation to refine image texture. The main innovations of the proposed framework can be summarized as follows.

The sharing of the same encoder has been realized in different domains through weighting the encoder in the proposed framework.

Different sampling methods has been performed on holes region and mask image respectively based on differentiated sampling.

The GAN constraints has been used to refine the image and generate texture.

The remainder of this paper is organized as follows. Section II analyzes the related works, Section III reports our research methodology and loss parameters, Section IV provides our experimental procedure and results, Finally, Section V concludes this study.

II. RELATED WORK

A. One-stage inpainting

Context encoder [6] is firstly introduced into image inpainting for learning semantic content. Global and local [7] discriminators are commonly used to distinguish generated image in global and local regions while enforcing the consistency of generated image in missing regions. Yu et al. [9] firstly introduce patch match in deep feature for filling missing holes. Liu et al. [22] devise a partial convolution to express different weights for different holes region. Liu et al. [17] design a strategy to limit the hole filling characteristics and the relationship between adjacent and outer regions. By introducing contour constraint, Nazeri et al. [23] propose that contour repair can be carried out gradually to fill the global region. Yu et al. [24] normalize pixels from the corrupted and uncorrupted regions separately based on the original inpainting mask to solve the mean and variance shift problem. These methods have improved the image inpainting accuracy somehow but lack effective constraints on the hole center and lack effective semantic reasoning ability in some complex scenarios.

B. Progressive inpainting

Li et al. [25] propose to leverage a shared module to gradually repair the edge of the hole to enhance the constraint on the center of the hole. Yang et al. [26] devise a pyramid structure loss to supervise structure learning and embedding for additional structural constraints. Yi et al. [27] design a contextual residual aggregation module as the residuals of generated features so that the incremental generation of target features ensures the detailed texture of generated results. Zeng et al. [28] propose a deep generation model with a feedback mechanism, which outputs the feature map as well as the result of the repair feature. In their method, the highly trusted feature pixel will be used as the valid information in the next iteration. All these methods infer subsequent features by appending the predicted features as prior knowledge, while the repaired features depend on the features of the previously filled area of the hole, the essence of their methods is that the following inferred features come from the initial effective features. This strong correlation makes it impossible to decouple the effective regions from the holes region.

C. VAE-based inpainting

Zhao et al. [18] use an effective reference image as the image inpainting style and further couple the hole image with the reference image based on cross-attention. Peng et al. [19] develop a structural attention module based on a hierarchical vectorized variational auto-encoder to capture the distance relationship, which allows for a variety of repair results. Wan et al. [20] train two variational auto-encoders to transform old photos and clean photos into two latent spaces, respectively. And the translation between these two latent spaces is learned with synthetic paired data. This method generalizes well to real photos because the domain gap is closed in the compact latent space. Du et al. [21] introduce discrete disentangling representation and adversarial domain adaption into general domain transfer framework, aided by extra self-supervised modules including background and semantic consistency constraints, learning robust representation under dual-domain constraints (for example the feature and image domains).

The goal of VAEs is to train a generative model in the form of $p(x,z)=p(z)p(x\mid z)$ where $p(z)$ is a prior distribution over latent variables z and $p(x\mid z)$ is the likelihood function or decoder that generates data X given variable z. Since the true posterior distribution $p(z\mid x)$ is in general intractable, the generative model is trained with the aid of an approximate posterior distribution or encoder $q(z\mid x)$.

In deep hierarchical VAEs, to increase the expressiveness of both the approximate posterior and prior, the latent variables are partitioned into disjoint groups, z={z1, z2,...,zl}, where L is the number of groups. Then, the prior is presented by $p(z)=\prod_l p(z_l\mid z_{<l})$ and the approximate posterior by $q(z\mid x)=\prod_l q(z_l\mid z_{<l},x)$ where each conditional in the prior $p(z_l\mid z_{<l})$ and the approximate posterior $q(z_l\mid z_{<l},x)$ are represented by factorial Normal distributions. We can write the variational lower bound $L_{vae}(x)$ on $\log(p(x))$ as :

$$L_{vae}(x)=E_{q(z|x)}[\log\ p(x\mid z)]-\sum_l KL(q(z_l\mid x,z_{<l})\parallel p(z_l\mid z_{<l})) \quad (1)$$

The objective is trained using the reparameterization trick.

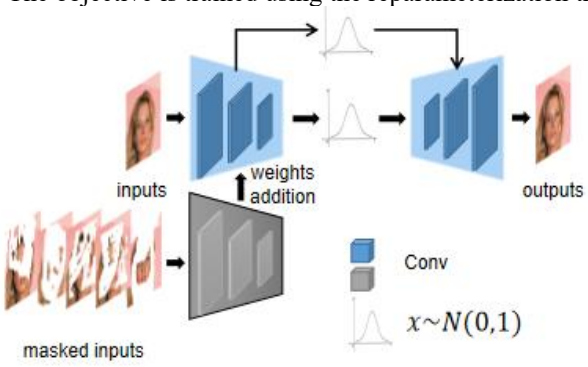

Figure 1. pipeline for image inpainting. The two domains are resolved into a normally distributed space by the same weighted encoder. Multi-level point sampling is used to obtain the distribution relations at different levels, which is conducive to the reliability of the results.

III. PROPOSED FRAMEWORK

For the proposed framework of image inpainting, We firstly introduce assumptions in section A. Then, we introduce in detail the cycle-consistency constraint introduced in the framework in section B. The implementation of the conversion on different domains through a common encoder is described in section C. In section D, the differential sampling method is devised to avoid ill-posed sampling. In section E, various loss functions are proposed, and numerical values are obtained by calculation. Finally, we illustrate in detail the composition of the loss function of the proposed image inpainting framework.

A. Assumption of the proposed framework

Let χ_t and $\chi_m(m=1,2,...)$ be true image domain and m-th mask image domain respectively. In unsupervised image-to-image translation, we are given samples drawn from the marginal distribution $P_{\chi_t(x_t)}$ and $P_{\chi_m(x_m)}$. The goal is to formulate a mapping from p1 to p2 to fill the image with holes. Since an infinite set of possible joint distributions can yield the given marginal distributions, we could infer nothing about the joint distribution from the marginal samples without additional assumptions [14].

Just as the experts can associate the masked image with the true image and consider them as the same image, we hypothesize that they have the shared-latent space. Explicitly, we formulate for any given pair of images x_t and $x_m(m=1,2,...)$, there exists a shared latent code z in a shared latent space, such that we can recover both images from this code, and we can compute this code from each of the two images. That is we postulate there exists functions E, G such that, given a pair of corresponding images (x_t, x_m) from the joint distribution, we have $z = E_t^*(x_t) = E_m^*(x_m)$ and conversely $x_t = G^*(z)$. In this model the function $x_t = F_{m->t}^*(x_m)$ that maps from $F_{m->t}^*(x_m) = G^*(E_m^*(x_m))$, which is a many-to-one mapping. Therefore, we can reconstruct the input image by translating it back to the translated input image. In other words, the proposed shared latent space assumption supports the cycle-consistency assumption, but not vice versa.

B. Self-cycle consistency

Feature learning of missing images deviates from the real image distribution is an important cause of learning failure, and the learning of real valid features will fix the distribution ambiguity caused by this ill deviation. Our goal is to focus on learning the best $F_{m->t}^*$ .The migration feature of ill-conditioned x_m will make the adaptive relationships within the model become locally valid. In the process of learning and training, the local adaptation relationship is further enhanced, and the encoder can't learn the distribution scheme that follows the whole real sample. This is fatal for re-parameterized sampling prior to decoding due to the offset of the distribution. Therefore, we can apply self-cycle consistency constraints in the proposed framework to further regularize the ill-posed unsupervised image inpainting problem. Formally, $x_t := G^*(E_t^*(x_t))$. This ensures complete representation learning within the image range.

C. Encoder-sharing

Based on the assumption of the shared implicit space proposed above, the true image and the masked image are distributed in different domains, so it is a good scheme to use two different encoders to map different domains. However, it notes that because two different encoders learn different spatial features, it does not transfer the learned mapping relationship to the masked image. Correspondingly, we use the same encoder to realize the transformation of two different domains to the same domain through the dynamic weighting method. Formally, we specify the input feature as X and the mask as M. we can get $E_m^* := M \otimes E_t^*$. Similarly, the mask is updated by the following rules:

$$m' = \begin{cases} 1, if\ (sum(X \otimes M)) > 0 \\ 0, otherwise \end{cases} \tag{2}$$

To implement the shared latent space assumption, we further assume a shared intermediate representation h such that the process of generating a pair of corresponding images admits a form of.

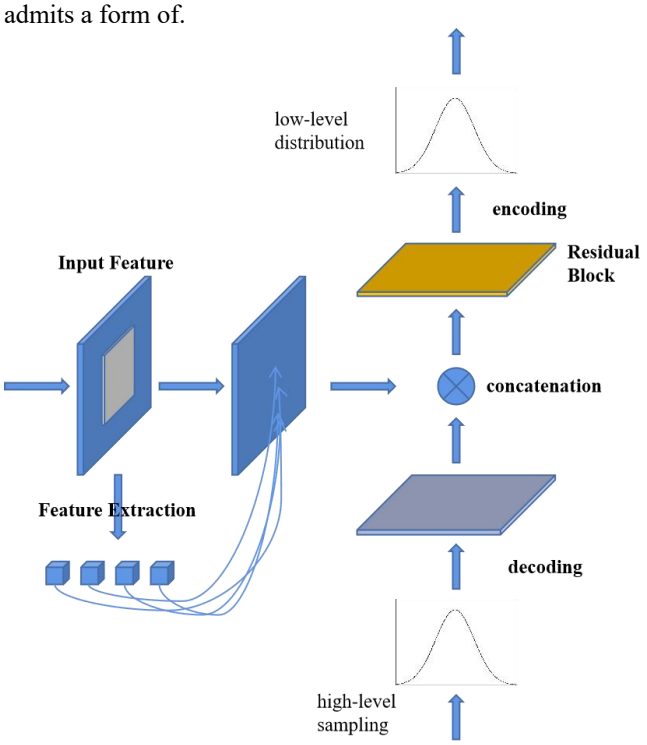

Figure 2. sampling module. The feature distribution obtained by the encoder is firstly filled with the hole features by the block matching strategy. Secondly, the decoded features from the high-level distribution sampling are concatenated to the repaired features to further obtain the joint feature distribution.

D. Distinctive Sampling

The absence of content leads to the ill-conditioned representation of true distribution features, and this ill-conditioned representation further limits sampling errors

due to global content sampling. Correspondingly, the sample errors further lead to the blurring and artificial imprinting of inpainting areas. The fundamental reason for this phenomenon is the use of consistent feature re-parameterized sampling for both valid and invalid feature regions, while ignoring the undesirable bias of invalid feature regions. Therefore, we conduct differentiated sampling in this region, which will no longer sample the features as a whole alone, but focus on the inpainting of sample bias in invalid feature regions. Moreover, it can not be ignored that the characteristics of invalid areas of different sizes and locations correspond to the different distribution of environmental characteristics, so we perform the sampling based on the joint distribution of the environment, regardless of the size and location of the region of the holes. Actually, we formulate an equation $f^i = q(z_i \mid x)$. To obtain the feature distribution of the hole region, we match the similarity of the patch region.

$$sim^i_{x,y,x',y'} = < \frac{f^i_{x,y}}{\| f^i_{x,y} \|}, \frac{f^i_{x',y'}}{\| f^i_{x',y'} \|} > \quad (3)$$

where $sim^i_{x,y,x',y'}$ indicates the similarity between the feature at the location of (x, y) and (x', y') . Further, the softmax function is used to calculate the attention score $score^i_{x,y,x',y'} = soft\max(sim^i_{x,y,x',y'})$. Eventually, attention scores are used to reconstruct the feature distribution.

$$\hat{f}^i_{x,y} = \sum score^i_{x,y,x',y'} f^i_{x,y} \quad (4)$$

The results of the filled feature distribution are linked to the prior distribution from the upper-level, and the joint probability distribution $q(z_i \mid x, z_{i-1})$ is obtained from the existing residual blocks.

E.Loss Function

The loss function of the proposed framework for image inpainting includes the sum of the three sub-loss functions, the KL divergence loss, the reconstruction lass as well as the GAN constraints. Details of the loss functions are described as follows.

1) KL Divergence Loss

Inspired by NAVE hierarchical sampling, we adopt a similar strategy for deeper-levels sampling. Practically, we examine the KL term in L_{vae} , as illustrated in equation

$$\mathrm{KL}(q(z_i|x)||p(z_i)) = \frac{1}{2}\left(\frac{\Delta\mu_i^2}{\sigma_i^2} + \Delta\sigma_i^2 - log\Delta\sigma_i^2 - 1\right) \quad (5)$$

where $\Delta\mu_i$ and $\Delta\sigma_i$ are the relative location and scale of the approximate posterior concerning the prior

2) Reconstruction Loss

Our network translates instance images into completion images in an unsupervised way. At times, the instance image is different from the corresponding completion image in pixel level. It is desired that the instance image is the same as the corresponding completion image in low-dimensional visible space. Therefore, the latent space loss is defined as:

$$L^Z_{rec} = ||E^*_t(x_t) - E^*_t\left(G\left(E^*_m(x_m)\right)\right)||_1 \quad (6)$$

For each masked image x_m there is only one ground true image x_t corresponding to it. When its corresponding ground truth image x_t is used as the guided instance image, the output of the generation module is x_t . Therefore, an identical reconstruction constraint is needed, which is defined as follows:

$$L^g_{rec} = ||x_t - G(E^*_m(x_m))||_1 \quad (7)$$

3) GAN Constraints

To refine our generation effects, we use a patch discriminator to activation the mapping of different range. The adversarial loss is defined as:

$$L_{adv} = \min_{G,E^*_m} \max_D [E_{x_t \sim Pdata} \log D(x_t) + E_{x_m \sim Pdata} \log(1 - D(G(E^*_m(x_m))))] \quad (8)$$

4) Total Loss

As is shown in equation (9), the total loss L_{total} of our method consists of three groups of component losses.

$$L_{total} = \lambda_{vae} L_{vae} + \lambda_{rec}\left(L^Z_{rec} + L^g_{rec}\right) + \lambda_{adv} L_{adv} \quad (9)$$

IV. EXPERIMENTS

The effectiveness and the superiority of the proposed framework for image inpainting are tested on the CelebA dataset [29]. The CelebA dataset contains more than 180,000 training images of face images. All images are re-sized and cropped to 256×256 for training and testing. Our framework is trained using the Adam [30] optimizer with the batch size of 6. We use the initial learning rate as 1e-4 to train the framework, and the learning rate is fine-tuned with learning rate of 5e-5 and decay rate of 0.02. We compare our method with several state-of-the-art methods based on the above-mentioned dataset. The qualitative comparisons between proposed framework and other methods are presented in Figure 3. Compared to other methods, the proposed framework shows more consistency with the true face effect in general and the result produced by the proposed framework with more detailed texture, but the image with some blur.

In addition, we also perform the quantitative comparison between the proposed framework with the NVAE method. It should be note that since the first three methods cannot generate diverse results for image inpainting, therefore these methods have only single output in the numerical comparison with the proposed framework, so the quantitative comparison is only performed to the NVAE method. The metrics of the quantitative comparison is summarized in Table 1 for the CelebA dataset. It can be seen that the proposed framework has better image reconstruction accuracy than NVAE for all the tested metrics of SSIM, PSNR, and MEAN l1. More detailed results are presented as follows.

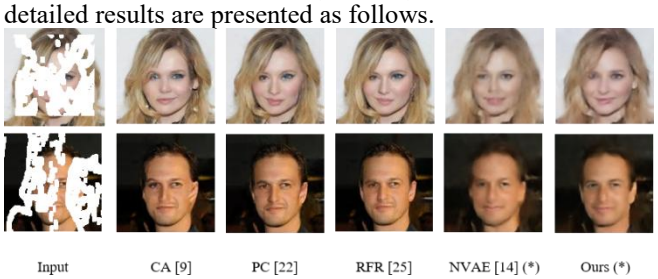

Figure 3. Qualitative comparisons the inpainting effect of the proposed framework (Ours) and existing methods on the CelebA dataset.

A.Diverse Generation

As is reported in section III, since our framework covers multiple levels of sampling, and the sampling parameters can be customized for each level of sampling deviation, of which ensures a variety of results generated. Figure 4 shows the inpainting results when setting different level of sampling parameters. It is showed that under different sampling parameters, the images generated by the guidance have different local features such as mouth shape. Visually compared to the NAVE method, the proposed framework produces more fine textures for the repaired images.

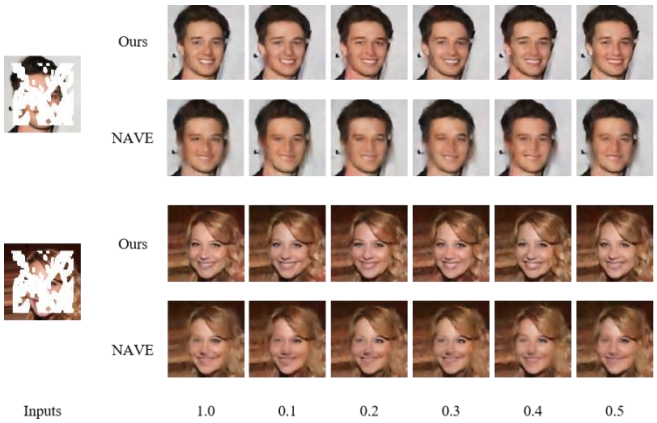

Figure 4. Under different sampling deviations, our method produces better texture characteristics than other method.

	SSIM	PSNR	MEAN l1 (%)
NVAE	0.7055	25.3406	4.0404
Ours	0.7649	27.7054	3.2793

Table 1. Quantitative comparison the inpainting accuracy of the proposed framework (Ours) and NAVE method on CelebA dataset.

B. Effectiveness of Distinctive Sampling

In the proposed framework, we introduce the differentiated sampling module. The differentiated sampling module is essentially a feature repair stack, which is conducive to obtaining more accurate feature distribution information. In order to verify the validity of this module, we test the image inpainting effects with and without the corresponding module in the proposed framework. The comparison is presented in Figure 5. We can see that with the distinctive sampling module we get more pleased inpainting result, where the evaluation metrics of SSIM, PSNR, MEAN L1 get improvement with the distinctive sampling module.

Figure 5. Comparison of image inpainting results with and without the corresponding module in the proposed framework (ssim/psnr/mean l1). The effectiveness of distinctive sampling can be inferred from metrics of SSIM, PSNR and MEAN l1.

V.CONCLUSIONS

In this paper, we propose an unsupervised image inpainting framework, of which can ensures the generation of credible results through domain sharing hypothesis with KL and refactoring constraints. We come up the strategy of encoder sharing to greatly simplifies the proposed framework. In addition, in order to ensure the repaired image have great consistent texture features with ground truth in the hole areas, we also devise a differentiated sampling strategy to distinguish the sampling parameters of the mask region and the effective region. Experimental results show the effectiveness and superiority of the proposed framework, which verifies the hypothesis behind the proposed method. However, there still have some defects for the proposed framework in current stage, for example, the blur phenomenon in the generated images. On one hand, this phenomenon comes from the limitation of VAE-based image generation. On the other hand, some features of the input image are missing, which will inevitably lead to the deviation of the overall features of the image. Although we have filled the missing features in the generation process of the proposed framework, however, this deviation is irreversible. In the next step of our research, we are going to address this issue and further improve the image inpainting effect.

REFERENCES

[1] Connelly Barnes, Eli Shechtman, Adam Finkelstein, and Dan B Goldman. 2009. PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing. In ACM SIGGRAPH 2009 Papers (New Orleans, Louisiana) (SIGGRAPH '09). Association for Computing Machinery, New York, NY, USA, Article 24, 11 pages.

[2] A. Criminisi, P. Perez, and K. Toyama. 2004. Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on Image Processing 13, 9(2004), 1200–1212.

[3] Zhan, X., Pan, X., Dai, B., Liu, Z., Lin, D., & Loy, C. C. 2020. Self-supervised scene de-occlusion. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 3784-3792.

[4] Rakshith Shetty, Mario Fritz, and Bernt Schiele. 2018. Adversarial Scene Editing: Automatic Object Removal from Weak Supervision. In Proceedings of the 32nd International Conference on Neural Information Processing Systems. Curran Associates Inc., Red Hook, NY, USA, 7717–7727.

[5] Linsen Song, Jie Cao, Lingxiao Song, Yibo Hu, and Ran He. 2019. Geometry-aware face completion and editing. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 33. 2506–2513.

[6] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros. 2016. Context Encoders: Feature Learning by Inpainting. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2536–2544.

[7] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. 2017. Globally and Locally Consistent Image Completion. ACM Trans. Graph. 36, 4, Article 107 (July 2017), 14 pages.

[8] Y. Zeng, J. Fu, H. Chao, and B. Guo. 2019. Learning Pyramid-Context Encoder Network for High-Quality Image Inpainting. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 1486–1494.

[9] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. 2018. Generative Image Inpainting with Contextual Attention. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. 5505–5514.

[10] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang. 2019. Free-Form Image Inpainting With Gated Convolution. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 4470–4479.

[11] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. 2020. Generative Adversarial Networks. Commun. ACM 63, 11 (Oct. 2020), 139–144.

[12] Neural Discrete Representation Learning.Oord, A. V. D., Vinyals, O., & Kavukcuoglu, K. 2017. Neural discrete representation learning. arXiv preprint arXiv:1711.00937.
[13] Razavi, A., van den Oord, A., & Vinyals, O. 2019. Generating diverse high-fidelity images with vq-vae-2. In Advances in neural information processing systems. 14866-14876.
[14] Vahdat, A., & Kautz, J. 2020. Nvae: A deep hierarchical variational autoencoder. arXiv preprint arXiv:2007.03898.
[15] Liu, H., Wan, Z., Huang, W., Song, Y., Han, X., & Liao, J. 2021. PD-GAN: Probabilistic Diverse GAN for Image Inpainting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 9371-9381.
[16] Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. arXiv preprint arXiv:1312.6114.
[17] H. Liu, B. Jiang, Y. Xiao, and C. Yang. 2019. Coherent Semantic Attention for Image Inpainting. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV). 4169–4178.
[18] Zhao, L., Mo, Q., Lin, S., Wang, Z., Zuo, Z., Chen, H., ... & Lu, D. 2020. Uctgan: Diverse image inpainting based on unsupervised cross-space translation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 5741-5750.
[19] Du, W., Chen, H., & Yang, H. 2020. Learning invariant representation for unsupervised image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 14483-14492.
[20] Guilin Liu, Fitsum A Reda, Kevin J Shih, Ting-Chun Wang, Andrew Tao, and Bryan Catanzaro. 2018. Image inpainting for irregular holes using partial convolutions. In Proceedings of the European Conference on Computer Vision (ECCV). 85–100.
[21] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi. 2019. EdgeConnect: Structure Guided Image Inpainting using Edge Prediction. In 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). 3265–3274.
[22] Tao Yu, Zongyu Guo, Xin Jin, Shilin Wu, Zhibo Chen, Weiping Li, Zhizheng Zhang, and Sen Liu. 2020. Region Normalization for Image Inpainting. In AAAI. 12733–12740.
[23] J. Li, N. Wang, L. Zhang, B. Du, and D. Tao. 2020. Recurrent Feature Reasoning forImage Inpainting. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 7757–7765.
[24] Jie Yang, Zhiquan Qi, and Yong Shi. 2020. Learning to incorporate structure knowledge for image inpainting. In Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34. 12605–12612.
[25] Z. Yi, Q. Tang, S. Azizi, D. Jang, and Z. Xu. 2020. Contextual Residual Aggregation for Ultra High-Resolution Image Inpainting. In 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). 7505–7514.
[26] Yu Zeng, Zhe Lin, Jimei Yang, Jianming Zhang, Eli Shechtman, and Huchuan Lu. 2020. High-resolution image inpainting with iterative confidence feedback and guided upsampling. In European Conference on Computer Vision. Springer, 1–17.
[27] Z. Liu, P. Luo, X. Wang, and X. Tang. 2015. Deep Learning Face Attributes in the Wild. In 2015 IEEE International Conference on Computer Vision (ICCV). 3730–3738.
[28] Diederik Kingma and Jimmy Ba. 2014. Adam: A Method for Stochastic Optimization. International Conference on Learning Representations.