Received: 18 July 2022 | Revised: 22 August 2022 | Accepted: 24 August 2022 | Published online: 28 September 2022

RESEARCH ARTICLE

Journal of Computational and Cognitive Engineering
2022, Vol. 1(4) 174–180
DOI: 10.47852/bonviewJCCE2202319

Deep Generative Inpainting with Comparative Sample Augmentation

Boli Fang[1,*], **Miao Jiang**[1], **Jerry Shen**[2] **and Björn Stenger**[3]

[1]*Indiana University, USA*

[2]*University of Southern California, USA*

[3]*Rakuten Institute of Technology, Japan*

Abstract: Recent advances in deep learning techniques such as convolutional neural networks, recurrent neural network, and generative adversarial networks have achieved breakthroughs in many fields or in many real-world applications. For the problem of semantic image inpainting, the task of reconstructing meaningful missing pixels also demonstrates that deep neural networks can play an effective role. While more effective than conventional approaches, deep learning models demand high storage capacity, since the many layers and parameters involved in neural network construction incur great space complexity and the end-to-end system incorporating these inpainting modules need to reserve a significant storage to save the model alone. Additionally, neural network training often requires large datasets consisting of high dimensional images, thereby requiring intensive computational resources such as the GPU. Furthermore, one problem that we may need to consider is that the inpainting quality of the images may vary considerably across different contexts because the foundational training data where the inpainting model is trained on differ in size and diversity. To address these problems, we present an inpainting strategy called *comparative sample augmentation*, which enhances the quality of the training set by filtering irrelevant images and constructing additional images using information about the surrounding regions of the target image and this strategy managed to augment the datasets. Experiments on multiple datasets demonstrate that our method extends the applicability of deep inpainting models to training sets with varying levels of diversity, while enhancing the inpainting quality as measured by qualitative and quantitative metrics for a large range of class of deep models, with little need for model-specific consideration.

Keywords: comparative data augmentation, deep neural network, generative adversarial networks

1. Introduction

In the field of modern computer vision research, various applications such as computational photography and image restoration (Yu et al., 2018) have been increasingly important and popular. Image inpainting, in particular, has been an example of image restoration. Although there has been substantial progress in image inpainting, up to now, there are still great challenges in achieving the objective due to the inherent difficulty in synthesizing missing pixels that are visually and semantically consistent with surrounding pixels. For example, if there is an image where a boat is located in the middle of the ocean and our objective is to 'remove' the presence of the boat in such an image, there is no guarantee that the resulting picture will be idealistically consistent to the oceanic surroundings due to noises in the data. This issue will become especially apparent when the amount of available training data is limited, a common scenario in real-time applications where diverse images are hard to obtain. To overcome the barriers in inpainting challenges, hard coding is often involved to reflect human expert intervention. Such method invariably incurs huge expenses to retrieve enough data, creating a barrier for any potential progress in tasks such as image restoration.

To reduce the cost of human involvement, previous research has sought to design automated methods for inpainting. Current solutions to the inpainting problem mainly belong to two groups: traditional patch-based learning methods and deep learning-based methods. Traditional methods directly utilize background information by assuming that information of missing patches can be detected by checking the features and textures in background regions (Barnes et al., 2009). However, these methods often lead to poor performances when they are used to reconstruct complex non-repeating structure that requires capturing high-level semantics for two reasons. The major reason is that traditional methods of computation of surrounding parts of the missing region is similar based. In other words, traditional methods do not capture the inherent semantic/scenic features while processing digitizing the image, but instead rely on naive similarity value formula, which can be misleading for pixel-based inferences because outlier erratic pixels often lie around the area needing

*Corresponding author: Boli Fang, Indiana University, USA. Email: bfang@iu.edu

inpainting. Moreover, missing regions in the inpainting tasks will be very irregular. Unlike many other image-related tasks, the inpainting problems often involve very peculiar shapes and orientation. Such reality means that it is difficult to implement an algorithm that can run the patch-based algorithm in just one pass; even when multiple passes are run, the resulting images may still be distorted due to intrinsic challenges. Furthermore, the existing work emphasizes on finding a similar image from a database to replace the missing region. This method may have two problems. The first problem is that the similarity search could raise other issues since the similarity score could be different with different metrics. Second, if the database contains limited number of candidates, the performance of ranking and retrieval could be relatively poor due to the differences in underlying semantics. With the problems thus described above, patch-based algorithms are applicable for the problem only when one can observe a strong pattern of missing region and when these patterns are easily generalized through simple mathematical descriptions.

Deep learning-based methods, on the other hand, use representations of latent space of existing pixels and transform inpainting into a conditional pixel generation problem (van den Oord et al., 2016; Yu et al., 2018). These methods typically will apply a deep learning model, especially the generative adversarial networks (GANs) as a backbone and generate a pattern of the missing region. A GAN network (Arjovsky et al., 2017), which contains a discriminate networks and generative networks, will learn the representations of the missing region and the surrounding areas of the missing pixels. The latent space can be learnt by training the networks and the parameters can be stored after training. Then we can produce the missing region by generating the missing part to fill in.

While such methods are promising and current development methods do show a potential, they would work only when there is ample training data. For instance, Yu et al. (2018)'s networks have demonstrated a very strong result to show that the representations of surrounding parts can be generated as an attention to the generative model, which boosts the result. Also globally and locally consistent image completion (GLCIC) (Iizuka et al., 2017) also has demonstrated that loss function can control the learning of the missing parts.

While these approaches did produce images of significantly higher quality in most cases, there are still many notable problems that need to be addressed. First of all, as we mentioned, deep learning-based model generally requires a large amount of highly varied training data for model training, a requirement of which makes it impossible to apply these strategies when the set of available training data is limited. The problem will be extremely obvious when the dataset is not quite common. Additionally, some datasets are extremely difficulty to be collected, which may significantly drop the performance of deep learning-based method. Recent research (Goodfellow et al., 2015) also suggests that the performances of neural networks vary considerably in a variety of tasks when input images contain adversarial noise that potentially affects the latent space, showing that countering adversarial examples is a key in boosting deep learning models.

Finally, training time and training load of deep learning model are also a real work issue that we cannot escape. A typically deep learning-based model requires an expensive GPU with more than 3 hours' training. If the dataset is even larger, even more computational resources will be required, and the expenses incurred would dramatically limit access to deep learning methods to a small portion of research community who can afford enough RAM to support the computation. Therefore, it is imperative to reduce the size and training time and inference time of the model of inpainting tasks, so that these models can be practical.

To address these issues, we propose a simple strategy that can easily be adopted in existing generative frameworks without model-specific considerations and extend deep neural network strategies in the case of varying amounts of training data. The contributions of this paper can be summarized as follows:

- We present an effective strategy of selecting relevant samples by constructing a similarity measure based on color attributes of the target image and the training dataset, selecting the K most relevant images.
- Our algorithm also considers local image quality by adding images created by adding white noise to the target image.
- We conduct experiments using generative inpainting and demonstrate that our method achieves improved inpainting results when the available training data are limited.

2. Related Work

Image inpainting has been a difficult problem since the year of 2010. Patch-based algorithm (Bugeau et al., 2010) has been proposed to solve this problem. In the early stage, finding a similar pattern from surrounding textures is the main idea to solve this problem. This method requires a large database and it is very difficult to control the defects especially the database lacks of enough image to adapt to the original image. It is also unavoidable that the edge of the patch could be as smooth as possible.

With the increasing development of machine learning, learning-based image inpainting is becoming more popular. The prior researches on problem of learning-based image inpainting can be classified into two groups: work that either directly incorporates features to be learned within the image to be inpainted, and approaches that use learning methods in relation to other possibly available training images to extract representation of latent space. Traditional approaches usually involve algorithms to directly handle information from the background to the missing pixels (Yu et al., 2018). These methods borrow surrounding textures and achieve satisfactory performance when the inpainting area largely contains repetitive features, but they fail at more complex inpainting images such as human faces and natural scenery. Additionally, methods which extensively handle patch similarity (Simakov et al., 2008) tend to be computationally expensive, making them less applicable to cases where the training data and computational resources are both limited.

Generative neural networks (GAN) have been a very useful way to implement the task of inpainting due to the natural of model. Over the course of recent years, GAN has evolved from a very experimental idea with simple structure to an extremely complicated structure allowing for industrial applications. Goodfellow et al. (2014) have raised the first generation of GAN, which brings up the concept of adversarial training of two models or networks simultaneously. The two networks are generative model and discriminator model, respectively. While Goodfellow's work did not incorporate the development of deep neural network, it illustrates a mechanism that optimizes both networks using the framework of adversarial training (Qian et al., 2022). The parameters could be retrieved by finding optimal solution of minimax value of the loss function.

To further optimize the performances of GAN, Radford et al. (2015) proposed DCGAN, which can be treated as a GAN with better deep neural networks design. The working mechanism is same as the original GAN and the optimization function is still pixel based. The generative model is similar to a decoder in a variational autoencoder (Kingma & Welling, 2013) and the output of generative model will be input of a convolutional neural network (CNN)-based discriminator (Krizhevsky et al., 2012). Both of the two networks with deep convolutional layers can be optimized in the same way as Goodfellow et al. (2014). However, the larger number of weight parameters and training can be relatively computational expensive. In the following years, many other work of GAN has been produced in the same framework, but the difference is on the backbone of the deep neural network. ResNetGAN (Wang et al., 2017) is an example of evolved GANs from DCGAN. The difference is that the backbone of both of the generative model and adversarial models is replaced with Resnet model (He et al., 2015), which is beneficial from a skip mechanism.

In addition to the computation problem, the issue of mode collapse has been gradually observed by some researchers and engineers (Thanh-Tung et al., 2018) who follow the same research. Mode collapse means that the variety of output images is becoming much less than training data. The reason of model collapse is that the loss function of GAN is pixel based which means the calculation of the ground truth and predictions are from Euclidean distance between them. In this case, many types of images will never be produced from generative model because the optimizer will be stuck into some specific local minima even with adequate training data. In this case, a new expression of generative neural network can be a solution for this problem. WGAN (Arjovsky et al., 2017) tried to solve the problem by replacing the original loss function from pixel-based comparison with a calculation of Wasserstein distance (Stéphanovitch et al., 2022) between the predicted image and ground truth image. The benefit of applying Wasserstein distance is that the loss of predictor and real image are based on a difference of pixel distribution. In this case, two major problems of original loss functions can be avoided. First of all, pixel-based loss function will be very fragile to the coordination change of some important features. For example, the predictions will be very different if we just make an image with up-side-down effect. Second, different features could generate the similar loss value if it is calculated by a pixel-based loss function. In the famous MINIST dataset's experiment, many 9 are lost from predictors just because 6 and 9 are quite close. However, the Wasserstein function can avoid this problem.

There are additional GAN structures that are evolved with WGAN. WGAN-GP (Gulrajani et al., 2017) has been approved to avoid vanish gradients. Big-GAN, however, has pushed the performance of GAN into a new peak (Brock et al., 2018). Orthogonal regularization is used in Big-GAN's generator, making it amenable to a simple "truncation trick" (Brock et al., 2018). This GAN-based model can generate accurate images with high resolution. Also it has shown the robustness even when the training data are complex. However, it is often the case that model training is computationally intensive, preventing this model from being deployed for extensive applications.

Deep learning-based model sometimes needs to adapt to the task. The major issue is that the raw GAN model can only produce the whole image as a candidate so in this case, we can only replace a whole image to complete a task of inpainting. Partial filling is necessary to make image inpainting and some GAN structure has been proposed to deal with this problem. The proposed methods have proved that they can solve the problem from these two perspectives: first, the GAN model can generate a pattern that fits the whole image, and the second is that the generated part is also smooth on the boundary to fit the missing part.

GLCIC (Iizuka et al., 2017) is a method that fits both of the global environment and the local missing part. The trick of the model is that the authors have add another output of discriminator network. During the training of GLCIC model, there would be two outputs of discriminator. One output will be a comparison between the whole predicted image and the original training image. Another half part of the discriminator is to compute another loss for the missing part of original image and the filled part. The model can balance the loss between globally and locally, which can help generate an inpainting with smooth make up.

Yu et al. (2018) also have observed that the surrounding texture of the missing region is very important to generate the missing part of the image. In his work, he had explicitly utilized surrounding image features as references during the training of his deep neural network to make the better predictions. To implement this, the features of surrounding images need to be extracted so that the features could be captured and then fed forward to the generator network that generates the filled image.

On the other hand, deep learning-based methods using CNNs and GANs encode high-level and low-level features via an encoder-decoder network and proceed by using constructing objective optimizers which take consistency factors into consideration (Iizuka et al., 2017b). Such design indeed enables the model to generate more diverse content in structured images, but effective training for satisfactory performances of these models requires access to huge amounts of varied labeled data often unavailable in real-time applications.

It is still an issue that deep learning-based methods are very computational intensive. It can be observed in Yu et al. (2018) and Iizuka et al. (2017b) that training generative inpainting models requires high computational resources and training time up to weeks and months due to the high complexity of the current network structures. Some backbones can cost even more time. It is recorded that a training with Big-GAN (Brock et al., 2018) may cost even 2 months with 8 GPUs. It would be catastrophe if we need to rebuild or tune the model after the whole training cycle. It is also a headache if researchers cannot afford the expensive equipment.

3. Comparative Sample Augmentation

In this section, we propose our data augmentation method of *comparative sample augmentation*, which consists of two separate parts: a comparative augmenting filter which selects the most relevant samples from the training dataset, and an augmentation step, which adds noise masks to the original image to produce additional training images. Images chosen by these two procedures are then combined to form the dataset for subsequent neural network training. The first part of our algorithm takes global information about the training set into consideration, while the second part focuses on local features within the target image. All steps in our algorithm are easy to implement and can be adapted to a variety of generative adversarial models.

3.1. Comparative augmenting filter

One notable problem in many traditional inpainting methods (Barnes et al., 2009; Simakov et al., 2008) is the under-representation of contextual information present in similar images. Previous deep learning methods (Iizuka et al., 2017b; Yu et al., 2018) take background pixels into consideration, but the training time of weeks or months can make adoption impractical.

To counter such problems, we introduce a comparative augmenting filter on the training dataset before inpainting. Motivated by the l_1 reconstruction loss described in (Yu et al., 2018), we first define a distance between two discrete distributions P and Q as:

$$d(P, Q) = \frac{\|P - Q\|_1}{2\max(\|P\|_1, \|Q\|_1)}$$

This distance normalizes to [0,1] and becomes 0 if and only if $P = Q$ almost surely. Here we consider the distributions of RGB pixels in the two images. Given the training dataset and the image to be inpainted, we compute the distances between the color distribution of our image and those of other images and select the K closest images with respect to the distance. This strategy collects the most relevant images as measured by color and helps inpainting by filtering out other irrelevant images whose contrasting color may adversely effect inpainting results when the inpainting model uses global information to learn possible choices of missing images.

3.2. Noise augmentation

In addition to considerations on global information and consistency, we also bolster the robustness of the deep inpainting model by adding random noise to the target image. Our motivation is that the deep neural networks, due to the dependence on the training data, may omit useful latent features during training and yield unsatisfactory outputs for the given tasks when the input contains perturbations. Goodfellow et al. (2015), for instance, noted that neural networks may assign incorrect labels to images in classification tasks when noise is added to images. This can be attributed to the lack of adversarial instances and noise perturbations within the training dataset. Since most of the GANs are implemented with similar deep neural networks, such limitations of deep neural networks may translate to unsatisfactory generated images. Therefore, to achieve satisfactory inpainting results, models need to take random perturbations into consideration during training.

To address such concerns, we propose the idea of self-enrichment of the training dataset. We add 64×64-sized masks $\mathbf{r}_i$ to the original 64×64-sized inpainting images $\mathbf{I}$ in RGB form in our datasets. Each random mask $\mathbf{r}_i$ consists of pixel-level noise, each drawn from a normal distribution with mean 0 and fixed variance. This procedure produces a batch of images $\tilde{\mathbf{I}}_i = \mathbf{I} \oplus \mathbf{r}_i$, which reveals local contextual clues specific to the inpainting image $\mathbf{I}$ itself. Together with the previously chosen training images that reflect information about global similarity, the noise-added images complement the training outcome by providing additional information about the local contextual details involved in inpainting.

Once the two steps are finished, we proceed to train the GAN model using the selected and augmented data. Notice that such augmentation is independent from the choice of loss function used in generative models and can be easily implemented across different GAN architectures. Here we use the modified WGAN-GP objective (Yu et al., 2018) for training.

4. Experiments

We construct datasets of varied size and content from the commonly used CIFAR-10, CelebA, and Places image datasets to test our method on inpainting tasks with large and small numbers of available training data. To test the cases with complete datasets, we select the "Ocean," "Orchard," and "Pier" folders from the Places dataset, resulting in 15,000 images. To test the applicability of our method to cases of smaller, incomplete, datasets, we randomly sample 5,000 diverse images from the CIFAR-10 and CelebA datasets, forming the datasets reduced-CIFAR and reduced-CelebA, respectively. The number of images we use for training is less than 10% of all images in CIFAR-10 and less than 1% of all images in CelebA, respectively. For our inpainting task in both cases, we randomly select an image as the inpainting target by masking it with a black square in its center.

We experiment on the state-of-the-art WGAN-GP generative model, since it is the building block for high resolution generative frameworks in recent research (Iizuka et al., 2017b; Yu et al., 2018). Notice that the procedures in our algorithm do not rely on the specific details of generative models, and that our strategy can easily be used for any current generative inpainting architectures.

4.1. Qualitative evaluations

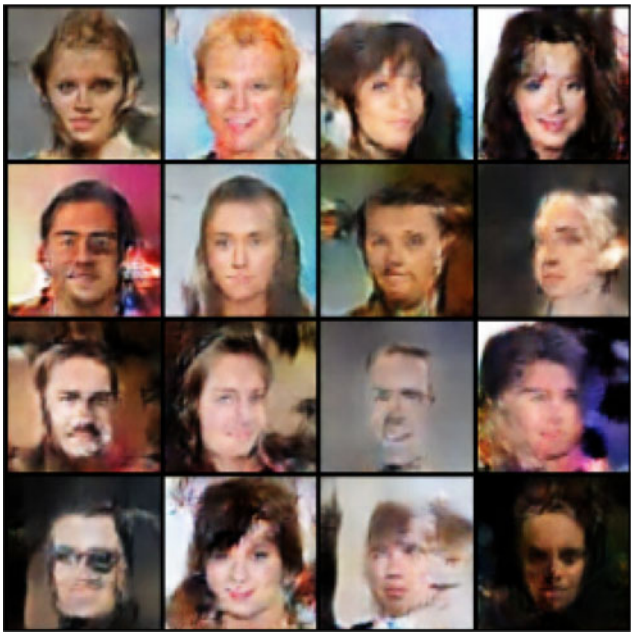
We present examples of generated images with and without comparative sample augmentation, all using the state-of-the-art WGAN-GP (Gulrajani et al., 2017). Figures 1 and 2 suggest that our sample augmentation strategy better captures prominent features of inpainted images such as eyes, nose, and mouth, while avoiding distortion that occurs in the unaugmented GAN when the training data contain highly diverse structured patterns. Figures 3 and 4 show inpainted examples with and without comparative sample augmentation from datasets with varying degrees of diversity. From these images, we can observe the relative improvements of inpainting quality, even for a naive inpainting network with only a single WGAN-GP building block.

4.2. Quantitative evaluation

Measuring the effectiveness of inpainting strategies remains a challenging task, since multiple solutions are plausible given the

Figure 1
Training samples with augmentation

Figure 2
Inpainting without augmentation



background of a target image. Common metrics for generative model evaluation such as inception score (Gurumurthy et al., 2017) and Frechet inception distance (Heusel et al., 2017) do not apply in our case since they measure the diversity across different images rather than the semantic coherence and authenticity of the inpainting results with respect to the background. Thus, we follow usual practices (Iizuka et al., 2017b; Yu et al., 2018) to report the mean l_1, l_2, total variation (TV) error, as well as the peak-signal-noise-ratio (PSNR) in Tables 1–4, evaluated on the respective test sets. The l_1, l_2, and TV errors measure the potential information loss incurred by our reconstruction, whereas PSNR indicates the

Table 1
l_1 reconstruction error

Datasets	Without augmentation (%)	With augmentation (%)
Ocean	11.6	9
Pier	12.2	11.2
Orchard	13.5	12.1
Reduced-Celeba	12.2	10.9
Reduced-CIFAR-10	12.2	12.2

Figure 3
Inpainting results on complete datasets. The augmentation method helps converge to a semantically plausible reconstruction, while inpainting without augmentation results in noticeable artifacts

original image masked image w augmentation w/o augmentation

Figure 4
Inpainting results on incomplete datasets. The inpainting result using the proposed augmentation method shows better convergence of color distributions and structure than without augmentation. Note that these results were obtained using a simple network architecture with a training time of 20 minutes on a laptop

original image masked image w augmentation w/o augmentation

Table 2
l_2 reconstruction errors

Datasets	Without augmentation (%)	With augmentation (%)
Ocean	16.6	12.6
Pier	15.5	14.9
Orchard	21.1	17.2
Reduced-CelebA	14.2	12.9
Reduced-CIFAR-10	27.6	25.3

Table 3
Peak-signal-noise-ratio

Datasets	Without augmentation	With augmentation
Ocean	36.01	37.09
Pier	36.06	36.46
Orchard	36.02	36.53
Reduced-CelebA	35.98	36.52
Reduced-CIFAR-10	24.92	34.98

Table 4
TV loss of inpainting

Datasets	Without augmentation (%)	With augmentation (%)
Ocean	25.3	21.2
Pier	31.1	26.4
Orchard	32.0	27.2
Reduced-CelebA	34.9	26.7
Reduced-CIFAR-10	48.3	44.9

information to noise ratio our reconstruction produces. Thus, higher PSNR and lower l_1,l_2, TV errors imply better performance.

4.3. Computational resources and cost

For all our experiments, we used a laptop equipped with an Intel Core 7700K CPU and an Nvidia GTX 1080 GPU. It was noted in previous research (Iizuka et al., 2017b; Yu et al. 2018) that inpainting GANs are computationally expensive to train, with the shortest reported training time being 2 weeks while having uncertainty on performances over smaller datasets. Our GAN with augmentation structure, on the other hand, requires less than 20 minutes of training time and obtains satisfactory results even for image sets with limited sizes and diversity factors.

5. Conclusion and Future Work

In sum, we introduce in this paper the strategy of comparative sample augmentation for deep generative inpainting models. Extracting the relevant images with color distances and then adding noise augmentation to promote the robustness of the resulting dataset before using the dataset for deep learning methods, our method not only avoids the distortion of images incurred by naive mathematical comparisons but also utilizes the developments in deep learning which allows for latent feature representation that leads to better image generation. We show by experiments on benchmark datasets that our method extends the deep learning-based inpainting methods to the datasets with varying levels of diversity and sizes, without any need for model-specific adjustments.

There are several potential avenues for future work. Since much recent work on generative inpainting (Iizuka et al., 2017b; Yu et al., 2018) incorporates multiple repetitive structures of generative models, we plan to take into consideration interactions between GAN units in inpainting models. Additionally, we will also consider other aspects such as gradient matching and feature encoders into the first part of our augmentation method to better approximate the distance between the inpainting image and the training images. Moreover, as suggested by Cubuk et al. (2018), better control of image augmentation in the second part of our method could be achieved using advancements in reinforcement learning. We also aim to explore interactions between our method and other generative paradigms such as Kingma & Welling (2013) and its variants.

One direction is that we can apply and develop the model to solve model compression problem. Current model compression can reduce the size of a deep learning model for around 90%. In this case, it is possible to train a complicated network but with small parameter size, which will further optimize the computational resources needed towards our task. Furthermore, the speed of the inference of the deep learning model can also be increased. Some possible methods include binarization (Hubara et al., 2016), quantization (Hubara et al., 2018), and model distilling, all of which can be applied to our model of inpainting so as to allow for training on devices with less computational resources. Finally, complicated GAN structures which have been shown to be effective, for instance the Big-GAN (Brock et al., 2018), can also be applied to our framework, so that one can anticipate a substantial improvement on the level of resolution in the images that are generated by our framework.

Supplementary Material

For supplementary material accompanying this paper visit https://doi.org/10.47852/bonviewJCCE2202319

Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

References

Arjovsky, M., Chintala, S., & Bottou, L. (2017). Wasserstein generative adversarial networks. In D., Precup, & Y. W., Teh (Eds.), *Proceedings of the 34th International Conference on Machine Learning (Proceedings of Machine Learning Research)*. PMLR.

Barnes, C., Shechtman, E., Finkelstein, A., & Goldman, D. B. (2009, August). PatchMatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics*, *28*(3), 1–11. https://doi.org/10.1145/1531326.1531330

Brock, A., Donahue, J., & Simonyan, K. (2018). Large scale GAN training for high fidelity natural image synthesis. *CoRR*, abs/1809.11096. https://doi.org/10.48550/arXiv.1809.11096

Bugeau, A., Bertalmío, M., Caselles, V., & Sapiro, G. (2010). A comprehensive framework for image inpainting. *IEEE Transactions on Image Processing*, *19*(10), 2634–2645. https://doi.org/10.1109/TIP.2010.2049240

Cubuk, E. D., Zoph, B., Mané, D., Vasudevan, V., & Le, Q. V. (2018). AutoAugment: Learning augmentation policies from data. *CoRR*, abs/1805.09501. https://doi.org/10.48550/arXiv.1805.09501

Goodfellow, I., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. In *International Conference on Learning Representations*. http://arxiv.org/abs/1412.6572

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., . . . Bengio, Y. (2020, November). Generative adversarial networks. *Communications of the ACM, 63*(11), 139–144. https://doi.org/10.48550/arXiv.1406.2661

Gulrajani, I., Ahmed, F., Arjovsky, M., Dumoulin, V., & Courville, A. C. (2017). Improved training of Wasserstein GANs. In *Advances in neural information processing systems* (Vol. *30*). Curran Associates, Inc.

Gurumurthy, S., Sarvadevabhatla, R. K., & Venkatesh Babu, R. (2017). DeLiGAN: Generative adversarial networks for diverse and limited data. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, abs/1512.03385. https://doi.org/10.48550/arXiv.1512.03385

Heusel, M., Ramsauer, H., Unterthiner, T., Nessler, B., & Hochreiter, S. (2017). GANs trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in neural information processing systems* (Vol. *30*). Curran Associates, Inc.

Hubara, I., Courbariaux, M., Soudry, D., El-Yaniv, R., & Bengio, Y. (2016). Binarized neural networks. In D. Lee, M. Sugiyama, U. Luxburg, I. Guyon, & R. Garnett (Eds.), *Advances in neural information processing systems* (Vol. *29*). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2016/file/d8330f857a17c53d217014ee776bfd50-Paper.pdf

Hubara, I., Courbariaux, M., Soudry, D., El-Yaniv, R., & Bengio, Y. (2018). Quantized neural networks: Training neural networks with low precision weights and activations. *Journal of Machine Learning Research*, *18*, 187–130. http://jmlr.org/papers/v18/16-456.html

Iizuka, S., Simo-Serra, E., & Ishikawa, H. (2017). Globally and locally consistent image completion. *ACM Transactions on Graphics*, *36*(4), 1–14. https://doi.org/10.5555/3122009.3242044

Kingma, D. P., & Welling, M. (2013). Auto-encoding variational bayes. https://doi.org/10.48550/arXiv.1312.6114

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. In F. Pereira, C. J. Burges, L. Bottou, & K. Q. Weinberger (Eds.), *Advances in neural information processing systems* (Vol. *25*). Curran Associates, Inc. https://proceedings.neurips.cc/paper/2012/file/c399862d3b9d6b76c8436e924a68c45b-Paper.pdf

Qian, Z., Huang, K., Wang, Q.-F., & Zhang, X.-Y. (2022). A survey of robust adversarial training in pattern recognition: Fundamental, theory, and methodologies. https://doi.org/10.48550/arXiv.2203.14046

Radford, A., Metz, L., & Chintala, S. (2016). Unsupervised representation learning with deep convolutional generative adversarial networks. In Y. Bengio & Y. LeCun (Eds.), *4th International Conference on Learning Representations, {ICLR} 2016*, San Juan, Puerto Rico, May 2–4, 2016, Conference Track Proceedings. http://arxiv.org/abs/1511.06434

Simakov, D., Caspi, Y., Shechtman, E., & Irani, M. (2008). Summarizing visual data using bidirectional similarity. In *CVPR*.

Stéphanovitch, A., Tanielian, U., Cadre, B., Klutchnikoff, N., & Biau, G. (2022). Optimal 1-Wasserstein distance for WGANs. https://doi.org/10.48550/arXiv.2201.02824

Thanh-Tung, H., Tran, T., & Venkatesh, S. (2018). On catastrophic forgetting and mode collapse in generative adversarial networks. *CoRR*, abs/1807.04015. https://doi.org/10.48550/arXiv.1807.04015

van den Oord, A., Kalchbrenner, N., Espeholt, L., Koray, K., Vinyals, O., & Graves, A. (2016). Conditional image generation with PixelCNN decoders. In *NIPS*.

Wang, M., Li, H., & Li, F. (2017). Generative adversarial network based on resnet for conditional image restoration. *CoRR*, abs/1707.04881. https://doi.org/10.48550/arXiv.1707.04881

Yu, J., Lin, Z., Yang, J., Shen, X., Lu, X., & Huang, T. S. (2018). Generative image inpainting with contextual attention. In *CVPR*.

How to Cite: Fang, B., Jiang, M., Shen, J., & Stenger, B. (2022). Deep Generative Inpainting with Comparative Sample Augmentation. *Journal of Computational and Cognitive Engineering* 1(4), 174–180, https://doi.org/10.47852/bonviewJCCE2202319