Free-Form Mask in Image Inpainting with GC-PatchGAN for High-Resolution Natural Scenery

Xuanze Chen
College of Information and Technology & College of Artificial Intelligence, Nanjing Forestry University, Nanjing, China

Keywords: Image Inpainting, GC-PatchGAN, Gated Convolution, Contextual Attention.

Abstract: The research delves into the domain of image inpainting, an essential task in image processing with widespread applications. Image inpainting holds immense importance in restoring damaged or incomplete images, finding utility in photography, video processing, and other fields. The study aims to introduce an innovative approach that combines the Contextual Attention Layer, Gated Convolution and SN-PatchGAN (GC-PatchGAN) to enhance inpainting outcomes. The methodology incorporates the Contextual Attention Layer for strategic borrowing of feature information from known background patches. Gated Convolution dynamically selects and highlights pertinent features, leading to a substantial improvement in inpainting quality. Extensive experiments were conducted to assess the proposed method, utilizing the Kaggle dataset. The results consistently demonstrate exceptional performance across various scenarios, including those involving Free-Form masks and user-guided input. Gated Convolution plays a pivotal role in generating high-quality results consistently. In practical terms, this research contributes significantly to image restoration, facilitating the removal of distractions, layout modifications, watermark elimination, and facial editing within images. Additionally, it addresses the challenge of irregular objects obstructing landscapes in photographs. In conclusion, this study advances the field of image inpainting, holding considerable promise for enhancing image quality and editing capabilities across diverse industries reliant on image processing and restoration.

1 INTRODUCTION

Natural scenery images possess captivating beauty, but images may suffer from many imperfections such as object occlusions when people take photos by smartphone or professional camera or noise and stains on the photo. Aiming to recover damaged images or fill missing pixels of a picture, image inpainting, a rough task in computer vision, is immensely crucial and valuable to restore its unity and make it clear. It can be also used to process video streams. Addressing the challenge of free-form masks in restoration, utilizing the Spectral-Normalized Generative adversarial network (GAN) and Patch GAN (SN-PatchGAN) excels in image generation task, requiring the restoration of missing content from partially corrupted images rather than simple copy-paste operations.

The field of image impaiting has seen major breakthroughs over the past decade or so. The evolution of techniques has witnessed a blend of traditional and deep learning methods. Early approaches primarily relied on handcrafted features and graphical model, with a focal point on pixel-wise classification (Felzenszwalb and Huttenlocher 2008). Subsequently, Convolutional Neural Networks (CNN) emerged, capitalizing on their ability to capture hierarchical features (Krizhevsky et al 2012). Fully connected layers have been a pivotal component in traditional CNNs for classification tasks. Significantly, the work, Context Encoders, delved into the notion of feature acquisition via inpainting (Pathak et al 2016). This endeavor substantially enhanced the comprehension of semantics by embracing inpainting as a means of feature learning. Recent advancements have veered toward fully convolutional networks (FCNs), tailored for spatially dense predictions (Long et al 2015). Semantic segmentation techniques, such as DeepLab and U-Net, have harnessed dilated convolutions and skip connections for improved context integration and fine-grained boundary delineation (Chen et al 2018 & Ronneberger et al 2015). While initial methods leveraged limited contextual information, current state-of-the-art models emphasize contextual aggregation through aurous spatial pyramid pooling

Chen, X.
Free-Form Mask in Image Inpainting with GC-PatchGAN for High-Resolution Natural Scenery.
DOI: 10.5220/0012819300003885

In *Proceedings of the 1st International Conference on Data Analysis and Machine Learning (DAML 2023)*, pages 402-406
ISBN: 978-989-758-705-4

and non-local blocks (Chen et al 2017 & Wang et al 2018). These techniques exploit long-range dependencies for more precise predictions. However, addressing the computational intensity of deep architectures remains a challenge. Some approaches propose lightweight architectures like mobile networks (MobileNets) to balance performance and efficiency (Howard et al 2017).

The main objective of this study is to tackle the issue of traditional convolutions treating missing and intact pixels equally, resulting in artifacts around edges for arbitrary-shaped missing regions, and to improve upon the non-learnable hard gating mask characteristic of partial convolutions, so the proposed Gated Convolution and SN-PatchGAN (GC-PatchGAN) is GAN-based model for image inpainting. Specifically, first, the contextual attention layer has the capability to generate missing sections by inferring information from known background patches or by replicating feature data from specific locations. This layer is suitable for testing the effect of generation at different resolutions in deep learning due to its differentiability and fully convolutional nature. Second, Gated convolution is a study of each channel's spatial position, allowing the dynamic characteristic option mechanism to intelligently highlight crucial features, providing a more precise solution for image restoration and processing. Interestingly, the visualization of intermediate gating values demonstrates its capability to choose features based not only on background, masks, and sketches but also considering semantic segmentation of certain channels. It learns to emphasize mask regions and sketch information across various channels in deep layer, contributing to improved generation of inpainting results. Third, the predictive performance of the different models is analyzed and compared. In addition, the discriminator in SN-PatchGAN, characterized by discrete values aligned with feature map dimensions, effectively focuses on image details, mitigating noise occurrences in varied shapes across the image. This approach offers a solution to the problem of noise appearing arbitrarily within the image. The experimental findings substantiate the paramount importance of gated convolutions, revealing its pivotal role in achieving notable enhancements in inpainting outcomes. Notably, under diverse settings, incorporating user-guided input scenarios involving masks of arbitrary shapes, gated convolutions consistently excel, showcasing their effectiveness in generating high-quality inpainting results. The relevance of the research in this paper is to be able to solve the problem of irregular objects obscuring the landscape in landscape photographs.

2 METHODOLOGY

2.1 Dataset Description and Preprocessing

The dataset used in this study, called archive, is sourced from Kaggle (Dataset). Comprised of a diverse collection of high-resolution natural scenery images, each image is 720x960 or 960x720 pixels in size. It encompasses various landscapes, including forests, mountains, lakes, and urban environments. This dataset serves as a valuable resource for evaluating image inpainting algorithms, allowing for the removal of undesirable objects or elements from scenic images. To prepare the data for experimentation, standard preprocessing techniques were applied, including resizing to a uniform resolution, noise reduction, and color correction, ensuring consistent and high-quality input for the inpainting process.

2.2 Proposed Approach

The paper introduces GC-PatchGAN for image inpainting is to revolves around the innovative integration of contextual attention layer, Gated Convolution and SN-PatchGAN, with a specific emphasis on addressing scenarios involving free-form masks and user guidance. This method synergistically combines the spatial and contextual information enhancement of the Contextual Attention Layer, the dynamic feature selection capabilities of Gated Convolution, and the effective discrimination in inpainting tasks provided by SN-PatchGAN. Furthermore, an interactive component allows users to incorporate sketches as a form of guidance for the inpainting process. These technologies, when combined, effectively extract and utilize spatial, contextual, and detailed information from images, enhancing the overall inpainting performance. Figure 1 below illustrates the structure of the system.

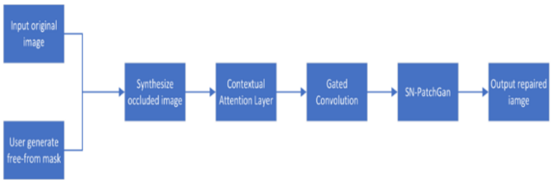
Input original image
User generate free-from mask
Synthesize occluded image
Contextual Attention Layer
Gated Convolution
SN-PatchGan
Output repaired iamge

Figure 1: The pipeline of the model (Original).

2.2.1 Contextual Attention Layer

The Contextual Attention Layer is a critical component of the model and plays a pivotal role in image inpainting. This layer is responsible for enhancing spatial and contextual understanding, ultimately improving the inpainting process by considering spatial relationships and context within the image. The layer begins by dividing the background region into patches and using them as convolution kernels on the foreground area. It then calculates cosine distances between foreground positions and background patches. Softmax function is applied along the channel dimension to compute attention values, which weigh the significance of various background patches for specific foreground positions. Finally, through a deconvolution operation, weighted sums of features are computed based on attention values, enriching the features for foreground positions. This process allows the model to better understand the spatial relationships and context within the image, resulting in more accurate and realistic inpainting results. Overall, the Contextual Attention Layer is a crucial module in the model, and its ability to enhance spatial and contextual understanding is essential for achieving high-quality inpainting results.

2.2.2 Gated Convolution

The Gated Convolution module is a pivotal component in our model, dynamically selecting and enhancing relevant features to improve inpainting. It introduces gating mechanisms within convolutional layers to assess feature importance at each spatial location. This adaptability allows it to emphasize pertinent information while suppressing less crucial details, thereby elevating the overall inpainting quality. Integrated into the model's architecture, Gated Convolution applies gating functions during implementation, dynamically adjusting feature contributions based on computed gating values. This significantly bolsters the model's ability to improve inpainting quality by adaptively highlighting contextually relevant features. Additionally, the module excels in learning dynamic feature selection, even considering semantic segmentation in specific channels, enhancing inpainting across various layers. Figure 2 below showing the whole structure of Gated Convolution described above.

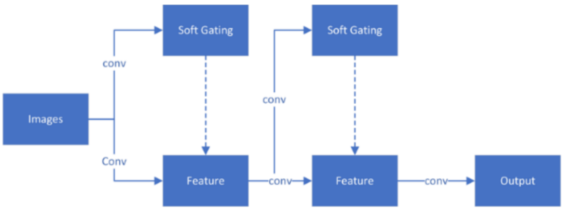

Figure 2: The pipeline of Gated Convolution (Original).

2.2.3 SN-PatchGAN

The purpose function of vanilla GAN is aim to make the generated data distribution as close as possible to the real data distribution, contributing to optimize the Jensen-Shannon divergence. However, a problem arises whereas the discriminator becomes better trained, the generator gradients tend to vanish. The Spectral Normalization technique introduces Lipschitz continuity constraints from the perspective of the spectral norm of the parameter matrices in each layer of the neural network. This imparts better robustness to input perturbations, making the neural network less sensitive, thus ensuring a more stable and convergent training process.

PatchGAN is a specific type of discriminator used in GAN, which stands for Generative Adversarial Network. Unlike a traditional GAN discriminator, which produces a single output representing the probability that the input is real, PatchGAN produces an N×N matrix of probabilities. In this matrix, each element corresponds to a small segment of the input picture. The final discriminator output is obtained by averaging these patch probabilities. This approach allows the model to consider the influence of different parts of the image, making it particularly useful for tasks that require high-resolution and fine-detail image generation. PatchGAN's smaller receptive field focuses on local image regions, making it more suitable for certain image-related tasks. The PatchGAN discriminator is trained to distinguish between real and fake images. During training, the generator produces fake images, and the discriminator evaluates them and provides feedback to the generator. The generator then adjusts its parameters to produce more realistic images, and the process continues until the generator produces images that are indistinguishable from real ones. PatchGAN has been used in a variety of image-related tasks, such as image-to-image translation, super-resolution, and image inpainting. In image-to-image translation, the generator takes an input image and produces an output image in a different style or with different attributes. In super-resolution, the generator produces high-resolution images from low-resolution ones. In image inpainting, the generator fills in missing parts of an image. In conclusion, PatchGAN is a powerful tool for image-related tasks that require high-resolution and fine-detail image generation. Its ability to consider the influence of different parts of the image makes it particularly useful for these tasks, and its smaller receptive field makes it more suitable for certain image-related tasks.

2.2.4 Loss Function

Selecting the appropriate loss function is of paramount importance in the training of deep learning models. The GC-Patch GAN loss function is a crucial optimization function in deep learning. It is defined by the provided mathematical expressions for the generator loss (L_G) and discriminator loss ($L_{D_{SN}}$). L_G focuses on minimizing the discrepancy between the generated data and the real data distribution, while $L_{D_{SN}}$ evaluates how well the discriminator distinguishes between real and generated data. These loss functions are fundamental in the training process, guiding the model towards convergence by adjusting the model parameters.

$$L_G = -E_{z\sim P_z(z)}[D_{SN}(G(z))] \tag{1}$$

$$L_{D_{SN}} = E_{x\sim P_{data}(x)}\left[ReLU\left(1 - D_{SN}(x)\right)\right] + E_{z\sim P_z(z)}\left[ReLU\left(1 + D_{SN}\left(G(z)\right)\right)\right] \tag{2}$$

The above formula denotes the GC-Patch GAN loss, where G takes incomplete image z with image inpainting network, D_{SN} represents spectral-normalized discriminator. Generator G and discriminator D_{SN} are trained simultaneously by solving $argmin_G L_G$ and $argmax_{D_{SN}} L_{D_{SN}}$. $G(z)$ symbolizes the image produced by the generator using noise z, while $D_{SN}(x)$ signifies the discriminator's output when assessing a real image x.

In formula (1), $E_{z\sim P_z(z)}$ denotes the expectation taken over the random noise input z to the generator. $P_z(z)$ represents the probability distribution of the noise z, typically a standard normal distribution or another predefined distribution.

In formula (2), $E_{x\sim P_{data}(x)}$ denotes the expectation taken over the real data distribution $P_{data}(x)$,standing for the probability distribution of real images. $E_{z\sim P_z(z)}$ denotes the expectation taken over the noise distribution $P_z(z)$, representing the probability distribution of the noise input z to the generator. $\left[ReLU\left(1 - D_{SN}(x)\right)\right]$: The ReLU (Rectified Linear Unit) function ensures that this term is zero when $D_{SN}(x)$ is greater than or equal to 1 (the discriminator correctly identifies a real image) and is positive when $D_{SN}(x)$ is less than 1 (the discriminator misclassifies a real image). $[ReLU(1 + D_{SN}(G(z)))]$: the ReLU function ensures that this term is zero when $D_{SN}(G(z))$ is less than or equal to -1 (the discriminator correctly identifies a fake image) and is positive when $D_{SN}(G(z))$ is greater than -1 (the discriminator misclassifies a fake image).

2.3 Implementation Details

In the implementation of this project, several key aspects were considered. First, in the background, the system was developed to address specific challenges related to image generation and manipulation. In order to enhance the diversity of the training dataset, Data modification techniques were applied, including operations like rotation, scaling, and brightness adjustments. Additionally, hyperparameters, such as learning rates and batch sizes, were carefully tuned to optimize training performance. These implementation details collectively contributed to the success of the system in generating high-quality images and achieving the project's objectives.

3 RESULTS AND DISCUSSION

In the conducted study, a hybrid model consisting of GC-Patch GAN is employed to repair high-resolution image inpainting tasks from a collection of over 1000 images with natural scenery.

Table 1: Comparison of Loss Among Contextual Attention, Gated Convolution and GC-Path GAN.

	Free-Form mask	
Method	L1 Loss	L2 Loss
Contextual Attention	0.1821	0.0486
Gated Convolution	0.1132	0.0197
GC-Patch GAN	0.0932	0.0145

As can be seen from the Table I, a comparative analysis reveals the performance of three distinct methods in Free-Form mask image restoration. The assessment employs both L1 Loss and L2 Loss metrics to gauge result quality. Notably, the Contextual Attention method exhibits relatively higher values in both L1 Loss (0.1821) and L2 Loss

(0.0486). This outcome suggests potential limitations in capturing fine-grained details when dealing with Free-Form masks, resulting in larger reconstruction errors. Consequently, in this specific task, Contextual Attention may not be the optimal choice. In contrast, both Gated Convolution and GC-Patch GAN methods demonstrate lower L1 Loss and L2 Loss values. Particularly, Gated Convolution stands out with impressive scores, recording 0.1132 for L1 Loss and 0.0197 for L2 Loss, while GC-Patch GAN exhibits even superior performance, with values of 0.0932 for L1 Loss and 0.0145 for L2 Loss.

These findings underscore the enhanced efficacy of GC-Patch GAN method in Free-Form mask image restoration tasks, as they excel in accurately reconstructing missing content. This advantage can be attributed to their ability to effectively handle Free-Form masks, enabling feature selection and highlighting, thus mitigating reconstruction errors. Consequently, opting for either of GC-Patch GAN can lead to superior restoration outcomes in practical applications.

4 CONCLUSION

This study has delved into the intricate domain of image inpainting, with a specific emphasis on tackling the complexities associated with Free-Form masks. The proposed methodology, which combines the Contextual Attention Layer, Gated Convolution, and SN-PatchGAN, offers a comprehensive framework to address these challenges effectively. The GC-PatchGAN shows impressive improvements by allowing to both capture strategic feature information borrowing from known background patches and selects and highlights relevant features which lead to notable enhancements in inpainting outcomes. Because of obtaining the continuity of image texture, the GC-PatchGAN network can be consistent with the overall chararteristics of the images. Moreover, the discriminator is enhanced by a refined focus on image details, successfully mitigating noise irregularities across varying shapes within the image. The generator ensures stability when training a large number of images, and therefore in conjunction with the discriminator can greatly reduce the rate of loss. Extensive experiments are meticulously conducted to evaluate the proposed method, consistently demonstrating its exceptional performance in diverse scenarios, including those involving Free-Form masks and user-guided input. These findings underscore the pivotal role of Gated Convolution in advancing inpainting outcomes, showcasing its effectiveness in generating high-quality results. The future research in image restoration will further explore the effectiveness of image restoration for complex objects in images from different scenarios and its applicability to high-resolution images before super resolution, aiming to refine and extend the model's capabilities to address an array of challenges within the realm of image restoration.

REFERENCES

P. F. Felzenszwalb, D. P. Huttenlocher, “Efficient Graph-Based Image Segmentation,” International Journal of Computer Vision, vol. 59, 2008, pp. 167-181.

A. Krizhevsky, I. Sutskever, G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” Advances in Neural Information Processing Systems (NIPS), vol. 25, 2012.

D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, A. A. Efros, “Context Encoders: Feature Learning by Inpainting,” In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 2536-2544.

J. Long, E. Shelhamer, T. Darrell, “Fully Convolutional Networks for Semantic Segmentation,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431-3440.

L. C. Chen, G. Papandreou, F. Schroff, H. Adam, “Rethinking Atrous Convolution for Semantic Image Segmentation,” arXiv: 2018, unpublished.

O. Ronneberger, P. Fischer, T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), 2015, pp. 234-241.

L. C. Chen, Y. Zhu, G. Papandreou, F. Schroff, H. Adam, “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation,” arXiv:2017, unpublished.

X. Wang, R. Girshick, A. Gupta, K. He, “Non-Local Neural Networks,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, pp. 7794-7803.

A. G. Howard, M. Zhu, B. Chen, D. Kalenichenko, W. Wang, T. Weyand, M. Andreetto, H. Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications,” arXiv:2017, unpublished.

Dataset, https://www.kaggle.com/datasets/arnaud58/landscape-pictures.