A Joint Gated Convolution Technique and SN-PatchGANn Model Applied in Oil Painting Image Restoration

Kuntian Wang
Computer Science, University of Nottingham, Nottingham, U.K.

Keywords: Image Inpainting, Oil Painting Restoration, Gated Convolutions, SN-PatchGAN.

Abstract: Image inpainting, which involves completing absent areas within an image, is a critical technique for enhancing image quality, preserving cultural heritage, and restoring damaged artworks. Traditional convolutional networks often struggle with irregular masks and multi-channel inputs in inpainting tasks. To address these challenges, this study presents a method that combines gated convolutions and a novel Spectral-Normalized Markovian Discriminator Generative Adversarial Network (SN-PatchGAN). Gated convolutions facilitate dynamic feature selection, ensuring color uniformity and high-quality inpainting. SN-PatchGAN, drawing inspiration from perceptual loss, Generative Adversarial Networks (GANs) driven globally and locally, Markovian Generative Adversarial Networks (MarkovianGANs), and Spectral-Normalized Generative Adversarial Networks (SN-GANs), efficiently handles arbitrary hole shapes. This study is conducted on the Oil Painting Images dataset and the outcomes from the experiment demonstrate the effectiveness of this method compared to two other traditional image inpainting methods. More importantly, it significantly improves the realism and quality of inpainted results, offering new possibilities for oil painting restoration and contributing to various societal aspects like conserving cultural heritage.

1 INTRODUCTION

The concept of image inpainting, which is also referred to as image completion or hole-filling, involves generating realistic and coherent content in missing regions of an image. This technique allows us to eliminate distracting objects, or repair areas in photos that are not needed, thereby enhancing image quality. It extends to tasks like image compression, super-resolution, rotation, stitching, and more (Yu et al 2019). Oil painting, as an art form, possesses a rich history and cultural value. However, over time, oil paintings may suffer damage, fading, or aging, which diminishes their aesthetic and preservation value. In this case, through image inpainting techniques, these damaged works can be restored to their original artistic charm. Restoring oil paintings not only preserves cultural heritage but also allows people to better understand the context of history and art development.

In the field of computer vision, there are two main strategies for image inpainting: one involves utilizing low-level image features for patch matching, while the other employs deep convolutional networks for generative models. The former technique is capable of generating plausible textures but often struggles with complex environments, facial features, and objects (Dewan and Thepade 2020). The latter technique leverages semantic knowledge learned from large datasets to generate content seamlessly within non-stationary images (Iizuka et al 2017 & Song et al 2018). However, conventional deep generative models using standard convolutions are not well-suited for image hole completion since these models consider all input pixels or features as equally valid, regardless of whether they are part of the missing regions. For inpainting, input at each network layer consists of both valid and invalid pixels in masked regions. The use of identical filters on all these pixels results in visible distortions like color mismatches, blurriness, and noticeable edge reactions around holes. This issue is particularly evident when dealing with irregular free-form masks (Iizuka et al 2017).

To overcome these limitations, this study proposes a gated convolution technique tailored for oil painting image restoration. It learns adaptive feature choice mechanisms for every channel and spatial position (e.g., internal or external to the mask, or red, green and blue primary colors (RGB) channels, or user-guided channels). This approach is easy to implement and demonstrates significant superiority

Wang, K.
A Joint Gated Convolution Technique and SN-PatchGAN Model Applied in Oil Painting Image Restoration.
DOI: 10.5220/0012804800003885

In *Proceedings of the 1st International Conference on Data Analysis and Machine Learning (DAML 2023)*, pages 492-497
ISBN: 978-989-758-705-4

in cases where the mask has an arbitrary shape and the input is no longer just RGB channels and the mask. In terms of network architecture, this research stacks gated convolutions to build an encoder-decoder network. Furthermore, without sacrificing performance, the author has significantly simplified the training objectives to include only pixel-level recovery loss and adversarial loss. This adjustment is specifically tailored for oil painting image restoration. Due to the potential presence of holes with any shape anywhere in oil painting images, global and local Generative Adversarial Networks (GANs) are not suitable, this paper introduces a modification of generative adversarial networks called Spectral-Normalized Markovian Discriminator Generative Adversarial Network (SN-PatchGAN), which is driven by GANs driven locally and globally, Markovian generative adversary networks (MarkovianGANs), perceptual lossand advances in Spectral-Normalized Generative Adversarial Networks (SN-GANs) (Li and Wand 2016 & Johnson et al 2016 & Miyato et al 2018). This addresses the limitations of previous GAN-based methods when dealing with arbitrary hole shapes. In conclusion, the proposed oil painting image restoration approach based on gated convolutions and SN-PatchGAN demonstrates significant improvements in realism and synthesized result quality. This method overcomes the limitations of traditional standard convolutions when dealing with different pixel states and multi-channel inputs, introducing new possibilities for oil painting image restoration tasks. Moreover, since oil painting image restoration has significant real-world implications in preserving cultural heritage, driving the art market, advancing education and research, maintaining personal memories, and fostering artistic innovation, the public can breathe new life into damaged oil paintings and bring positive impacts to various facets of society by using this technology.

2 METHODOLOGY

2.1 Dataset Description and Preprocessing

The dataset used in this study called Oil Painting Images, is sourced from Kaggle (Dataset). It is a collection of oil painting images, which contains a variety of topics including figures, animals, landscapes, etc. The dataset is divided into two distinct sizes: 256×256 and 128×128 pixels, both containing the same oil paintings, totaling 1763 examples in each size category. All models have been trained using images with a resolution of 256×256 pixels and a maximum hole size of 128×128 pixels. It's worth noting that performance may degrade when working with larger resolutions or hole sizes beyond these specifications. To provide a visual representation, Figure 1 offers a glimpse of some of the paintings within the dataset.

Figure 1: Images from the Oil Painting Images dataset (Picture credit: Original).

2.2 Proposed Approach

The focus of this proposed method for oil painting images inpainting revolves around the innovative fusion of Feature-wise Gating, Gated Convolution, and SN-PatchGAN. Employing gated convolution in order to acquire an adaptive feature choice process for every channel and spatial point throughout all network layers substantially enhances the uniformity of color and the quality of inpainting for free-form masks and inputs. Moreover, a more functional GAN discriminator based on patch, SN-PatchGAN, is presented. This approach is straightforward, speedy, and yields top-notch inpainting outcomes. Figure 2 below illustrates the structure of the system.

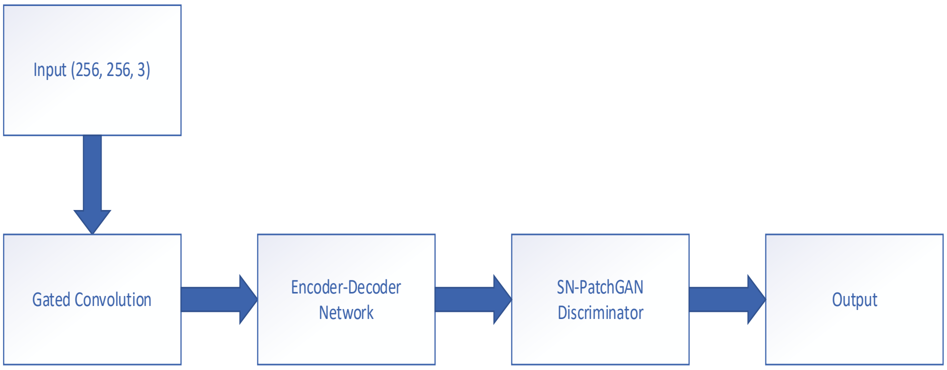

Figure 2: The pipeline of the model (Picture credit: Original).

2.2.1 Gated Convolution

The convolutional base of this model is established by employing gated convolution for the image inpainting network. Unlike partial convolution which employs rigid rules to update the mask with hard gating, gated convolutions autonomously learn a soft gating from the data based on update rules. It is formulated as:

$$Gating_{y,x} = \sum\sum W_g \cdot I \quad (1)$$

$$Feature_{y,x} = \sum\sum W_f \cdot I \quad (2)$$

$$O_{y,x} = \phi(Feature_{y,x}) \odot \sigma(Gating_{y,x}) \quad (3)$$

where σ is the sigmoid function. Therefore, the output values range between zero and one. ϕ is an activation function and could be any type. For example, Rectified Linear Unit (ReLU) and LeakyReLU are common and practical activation functions. W_g and W_f are two distinct convolutional filters. The former is a convolutional kernel that acts on the input to generate a soft mask (i.e. gating), while the latter is a convolutional kernel that acts on the input to generate a feature map.

Gated convolution utilizes an adaptive feature choice mechanism acting on the feature map for each channel and spatial position, as shown in Figure 3. The mask of the image and the incomplete image are initially processed through convolution to generate a soft mask and feature map. Then, the two are subjected to element-wise multiplication to obtain the output, which is then used as input to continue the same operation.

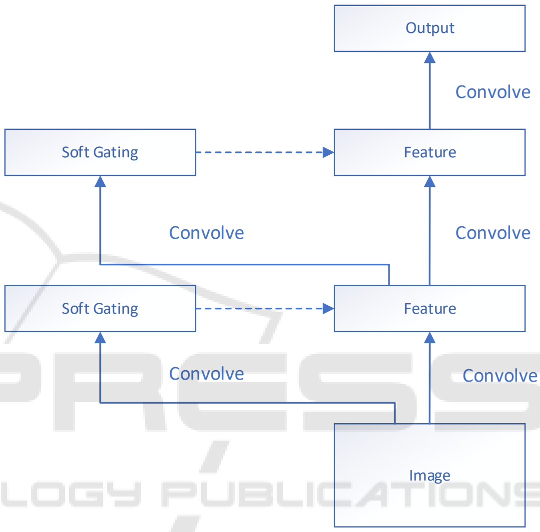

Figure 3: Gated convolution (Picture credit: Original).

2.2.2 SN-PatchGAN

In the context of free-form image restoration which involves multiple holes of various shapes and positions, prior methods relied on local GANs to enhance results when filling single rectangular holes. However, inspired by the perceptual loss, GANs driven globally and locally, MarkovianGANs, and the latest advancements in SN-GANs, a straightforward yet efficient GAN loss called SN-PatchGAN during the training of networks specialized in oil painting image inpainting introduced (Iizuka et al 2017, Li and Wand 2016, Johnson et al 2016 & Miyato et al 2018). In this approach, a convolutional network serves as the discriminator, taking as input the image and mask and producing a three-dimensional feature map. This feature map is created by stacking six stridden convolutions using a 5×5 kernel and a 2-step stride to grasp the feature statistics of Markovian patches. GANs are then employed on every feature element within this map, resulting in a multitude of GANs focusing on distinct spatial positions and semantic aspects of the input image. It's worth noting that the coverage area of each neuron in the result map covers the whole input image during training, making a global discriminator unnecessary. This approach also incorporates spectral normalization to enhance GAN training stability, following the fast approximation algorithm described in SN-GANs (Dataset).

By using SN-PatchGAN, the inpainting network trains more efficiently and robustly compared to the baseline model (Yu et al 2018). Notably, the use of perceptual loss is omitted since comparable information at the patch level is already embedded in SN-PatchGAN. In contrast to methods like Parial Convolution (Partial-Con), which involve multiple loss terms and hyperparameter balancing, the goal function for filling in the image now involves only pixel-wise L1 reconstruction error and SN-PatchGAN loss, with a default balancing hyperparameter for loss weighting set at 1:1 (Liu et al 2018).

2.2.3 Loss Function

The role of the loss function is pivotal during the training of deep learning models. In the case of this image inpainting task, the hinge loss function is deemed optimal. The goal of this loss is to simultaneously optimize two opposing objectives, namely maximize $\mathcal{L}_D$ and minimize $\mathcal{L}_G$.

For generator:

$$\mathcal{L}_G = -\mathrm{E}_{z\sim\mathbb{P}_z(z)}\,[D(G(z))] \quad (4)$$

For discriminator:

$$\mathcal{L}_D = \mathrm{E}_{x\sim\mathbb{P}_{data}}\big[\max\big(0, 1 - D(x)\big)\big] + \mathrm{E}_{z\sim\mathbb{P}_z(z)}[\max\,(0, 1 + D(G(z)))] \quad (5)$$

where D means the SN discriminator, and G is the network designed for image inpainting, which takes an incomplete image as input z. For D, only negative samples with D (G (z))>-1 and positive samples with D (x)<1 will influence the results. Consequently, this method can make the training more stable.

2.3 Implementation Details

In the implementation of the model, several crucial aspects are highlighted. Firstly, the model is trained using TensorFlow v2.4.0, CUDNN v8.2.4, and CUDA v11.4. For testing, regardless of the hole size, it takes 0.6 seconds per image to run on an NVIDIA GeForce MX350 GPU and 2.5 seconds on an Intel(R) Core(TM) i7 – 1065G7 CPU @ 1.5GHz. This applies to images sized at 256 × 256 pixels. The system is focused on training a generative model for image inpainting, specifically using a mixture of discriminator and generator networks. This technique is commonly used in the domains like image processing and computer vision. Then, regarding hyperparameters: the learning rate is configured at 0.0001. Additionally, a batch size of 16 is employed. Other than that, the model undergoes training for a total of 100 epochs. The choice of optimizer is the Adam optimizer, selected for its efficient handling of gradient descent in high-dimensional spaces. The betas for the optimizer are 0.5 (beta1) and 0.999 (beta2). In addition, data augmentation techniques are applied to the dataset to augment its size and diversity while mitigating overfitting. The specific data augmentation techniques used include random horizontal flipping.

3 RESULTS AND DISCUSSION

This section evaluates the free-form image inpainting system on the Oil Painting Images dataset.

Quantitative evaluation of image inpainting is challenging due to the lack of well-established metrics. Nonetheless, in this study, some evaluation results are provided, including average L1 error and average L2 error, on validation images for free-form masks. As depicted in Table 1, the techniques based on learning outperform the traditional Patch Match method regarding both average L1 error and average L2 error. This indicates that the approach achieves better performance in image inpainting tasks, accurately restoring missing image information. Furthermore, the use of partial convolutions within the same framework yields poorer performance. This might be attributed to partial convolution methods relying on rule-based gating, which may not adapt as effectively to various image inpainting scenarios, especially when dealing with free-form masks.

Table 1: Quantitative Comparisons.

Approach	L1 error	L2 error
Patch Match	12.6%	2.9%
Contextual Attention	18.4%	5.1%
Partial Convolution	11.6%	2.1%
Gated Convolution	10.2%	1.8%

This section compares the method introduced in this paper with previous cutting-edge approaches (Yu et al 2018 & Liu et al 2018). Figure 4 displays the results of automatic restoration on several representative images. Displayed from left to right: the raw image, the masked image, and the result of the model using spatial convolution and gated convolution.

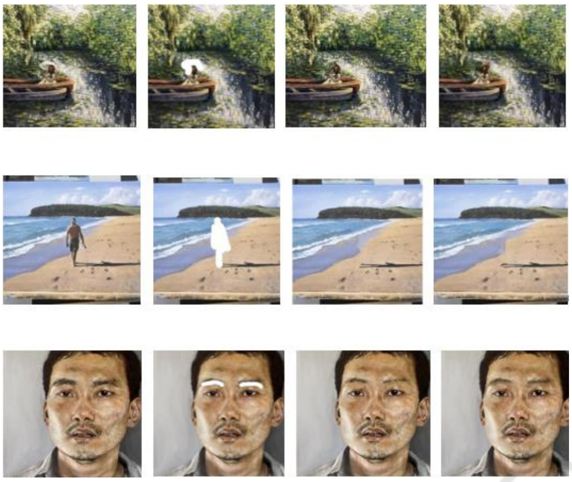

Figure 4: Comparison of the outcome of spatial and gated convolution approaches (Picture credit: Original).

By comparing these two approaches, it is noticeable that partial convolution produces better results but still exhibits observable color discrepancies. However, the approach based on gated convolutions achieves more visually pleasing results without significant color inconsistencies.

To summarize, by quantitatively comparing several commonly used image restoration methods with the approach introduced in this paper, it can be concluded that the model using gated convolution has significantly lower loss rates than other methods. Regarding the image restoration results, partial convolution, unlike vanilla convolution, does not exhibit obvious visual artifacts and edge responses within holes or around holes, but it still shows noticeable color discrepancies. In contrast, the gated convolution-based approach largely overcomes this issue, producing more realistic output results.

4 CONCLUSION

This study presents a groundbreaking approach to the restoration of oil painting images by integrating gated convolutions and the SN-PatchGAN discriminator. Traditional inpainting methods have long struggled with limitations when dealing with diverse hole shapes and multi-channel inputs, often yielding unrealistic or subpar results. However, this innovative technique offers a solution to these challenges, enabling the restoration of oil paintings with remarkable realism and high quality.

Gated convolutions are at the core of this approach, introducing dynamic feature selection mechanisms for each channel and spatial position. This significantly enhances color uniformity and inpainting performance, ensuring that the restored images are both faithful to the original artwork and aesthetically pleasing. This is a crucial advancement as it addresses a critical issue in image restoration, particularly when dealing with free-form masks that are common in the world of art conservation.

The SN-PatchGAN discriminator complements the process by streamlining the training phase, making it more efficient and robust. It simplifies the loss function, resulting in a more straightforward yet effective approach. The combination of gated convolutions and SN-PatchGAN is a novel technique in the field of image restoration. It not only significantly improves inpainting quality but also opens up new possibilities for various oil painting restoration tasks.

This research plays a vital role in preserving cultural heritage, revitalizing the art market, advancing educational and research endeavors, and safeguarding personal memories. Furthermore, it fosters artistic innovation by providing artists and restorers with powerful tools to breathe new life into old artworks.

Looking ahead, this research can serve as a foundation for further exploration within the realm of image restoration, inspiring new approaches and innovations to meet the evolving needs of art conservation and digital image processing.

REFERENCES

J. Yu, Z. Lin, J. Yang, et al. “Free-form image inpainting with gated convolution,” Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 4471-4480.

J. H. Dewan, S. D. Thepade, “Image retrieval using low level and local features contents: a comprehensive review,” Applied Computational Intelligence and Soft Computing, 2020, pp. 1-20.

S. Iizuka, E. Simo-Serra, H. Ishikawa, “Globally and locally consistent image completion,” ACM Transactions on Graphics (ToG), vol. 36, 2017, pp. 1-14.

Y. Song, C. Yang, Z. Lin, et al. “Contextual-based image inpainting: Infer, match, and translate,” Proceedings of the European conference on computer vision (ECCV), 2018, pp. 3-19.

C. Li, M. Wand. “Precomputed real-time texture synthesis with Markovian generative adversarial networks,” Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, 2016, pp. 702-716.

J. Johnson, A. Alahi, L. Fei-Fe, “Perceptual losses for real-time style transfer and super-resolution,” Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, 2016, pp. 694-711.
T. Miyato, T. Kataoka, M. Koyama, et al. “Spectral normalization for generative adversarial networks,” arXiv: 2018, unpublished.
Dataset https://www.kaggle.com/datasets/herbulaneum/oil-painting-images
J. Yu, Z Lin, J. Yang, et al. “Generative image inpainting with contextual attention,” Proceedings of the IEEE conference on computer vision and pattern recognition. 2018, pp. 5505-5514.
G. Liu, F.A. Reda, K.J. Shih, et al. “Image inpainting for irregular holes using partial convolutions,” Proceedings of the European conference on computer vision (ECCV), 2018, pp. 85-100.