IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Image Inpainting With Local And Global Refinement

[1]P. Pavan Sriram, [2]S. Gnana Chaitanya, [3]Dr. K. Siva Kumar

[1]Student, [2]Student, [3]Associate Professor

[1]Computer Science and Engineering,

[1]R.V.R & J.C College of Engineering, Guntur, India

Abstract: Image inpainting has made remarkable progress with recent advances in deep learning. Popular networks mainly follow an encoder-decoder architecture (sometimes with skip connections) and possess sufficiently large receptive field, i.e., larger than the image resolution. The receptive field refers to the set of input pixels that are path-connected to a neuron. For image inpainting task, however, the size of surrounding areas needed to repair different kinds of missing regions are different, and the very large receptive field is not always optimal, especially for the local structures and textures. In addition, a large receptive field tends to involve more undesired completion results, which will disturb the inpainting process. Based on these, a novel three-stage inpainting framework with local and global refinement is proposed.

Keywords – Image inpainting, neural networks, receptive field

I. INTRODUCTION

Image inpainting is a process that aims to complete missing regions in digital images by filling them with semantically reasonable and visually realistic content that aligns with the rest of the image. It finds applications in image editing, such as removing unwanted objects or restoring damaged regions in paintings. Traditional approaches to image inpainting include diffusion-based methods and patch-based methods, which have limitations in handling large missing regions and creating new structures do not present in the image.

Recent advances in deep learning, particularly with convolutional neural networks (CNNs) and generative adversarial networks (GANs), have shown promising results in image inpainting. However, most existing methods focus on using large receptive fields, which may not always be optimal for handling local structures and textures. These large receptive fields can also introduce undesired completion results, negatively affecting the inpainting process.

The primary objective of this work is to propose a more effective and efficient image inpainting method that addresses the limitations of the existing approaches. Our aim is to rethink the process of image inpainting from the perspective of receptive fields and design a three-stage inpainting framework that incorporates local and global refinement stages. The first stage involves using an encoder-decoder network with skip connections to achieve coarse initial results. The second stage introduces a shallow deep model with a small receptive field to perform local refinement, targeting the repair of missing regions with a focus on local structures and texture details.

II. LITERATURE REVIEW

C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman [1], "PatchMatch: A randomized correspondence algorithm for structural image editing." In this paper, the authors introduced a novel algorithm called PatchMatch, which is designed for structural image editing tasks. Structural image editing involves altering the content of an image while preserving its overall structure, textures, and details.

O. Ronneberger, P. Fischer, and T. Brox [2], "U-Net: Convolutional networks for biomedical image segmentation,". The research introduces a novel approach called "Image Melding," which aims to combine multiple inconsistent images into a cohesive and visually pleasing composite using patch-based synthesis. The main challenge addressed by Image Melding is the seamless integration of multiple input images with varying content and lighting conditions.

W. Luo, Y. Li, R. Urtasun, and R. Zemel [3], "Understanding the effective receptive field in deep convolutional neural networks," The effective receptive field of a neuron in a deep convolutional neural network (CNN) refers to the region in the input space that influences the neuron's activation. Understanding the effective receptive field is crucial for interpreting the behaviour and feature learning capabilities of individual neurons within the network.

Goodfellow et al. [4], "Generative adversarial nets,". The research introduces the concept of Generative Adversarial Networks (GANs), a groundbreaking framework for training generative models. GANs are composed of two neural networks, namely the generator and the discriminator, which are trained in an adversarial setting. The generator is responsible for producing synthetic data samples that resemble real data, while the discriminator tries to distinguish between real data and the generated data.

G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro [5], "Image inpainting for irregular holes using partial convolutions". The key contribution of this work is the introduction of partial convolutions, a novel technique that enables convolutional neural networks (CNNs) to handle regions with holes effectively. Traditional convolutions treat all pixels equally, which can lead to artifacts and blurriness when inpainting irregular holes. Partial convolutions, on the other hand, use a learnable mask that determines which pixels in the convolutional kernel are valid (non-hole) pixels.

Y. Zeng, J. Fu, H. Chao, and B. Guo [6], "Learning pyramid-context encoder network for high-quality image inpainting," introduce a novel approach aimed at advancing the state-of-the-art in image inpainting. The proposed method, termed the Pyramid-Context Encoder Network (PCEN), stands out for its ability to leverage multi-scale contextual information to inpaint missing or damaged regions in images with exceptional quality. Building upon deep learning techniques, PCEN encodes both the input image and its surrounding context, allowing it to effectively capture intricate details and maintain structural coherence during the inpainting process.

J. Johnson, A. Alahi, and L. Fei-Fei [7], "Perceptual losses for real-time style transfer and super-resolution". The paper addresses two important computer vision tasks: style transfer and super-resolution. Style transfer is the process of applying the visual appearance or "style" of one image to another, creating an output image that combines the content of one image with the artistic style of another. Super-resolution, on the other hand, is the task of generating a high- resolution image from a low-resolution input, enhancing image details and sharpness.

III. PROPOSED METHODOLOGY

LGNet is a novel inpainting network architecture proposed in this paper. It consists of two main components: the Local Refinement Network (NetL) and the Global Refinement Network (NetG). The architecture is designed to leverage different receptive fields to effectively inpaint missing regions in an image.

Local Refinement Network $(Net)_L$: The Local Refinement Network is responsible for handling "local inpainting" and repairing missing regions related to local structures and texture details. It employs a shallow deep network with a smaller receptive field. The main building block in this network is a two-layer residual block, which is represented by the purple block.

- Input: The input to NetL is the original incomplete image $I_{(in)}$ and the corresponding binary mask (M) indicating the locations of missing regions.
- Processing: The incomplete image $I_{(in)}$ and the output of the previous stage $I^{C}_{(out)}$ are merged, and the binary mask (M) is concatenated with this merged image. This combined input is then passed through the Local Refinement Network Net_L to produce the local refinement output $I^{L}_{(out)}$.
- Output: The local refinement output $I^{L}_{(out)}$ represents the inpainting results obtained by focusing on local structures and texture details. The missing regions are repaired based on information from surrounding local regions, preventing interference from long-distance failed completions.
- Merging: To obtain the merged image $I^{L}_{(mer)}$ regions from the original incomplete image $I_{(in)}$ are replaced with the corresponding regions from the local refinement output $I^{L}_{(out)}$ using a gray dotted line representation.

Global Refinement Network $(Net)_G$: The Global Refinement Network is designed to handle "global inpainting" and enhance the completion results by considering global information, particularly for large structures and long-distance texture patterns. The network includes three attention modules represented by green blocks, each operating at different resolutions: 16 × 16, 32 × 32, and 64 × 64.

- Input: The input to $(Net)_G$ is the merged image from the Local Refinement Network $I^{L}_{(mer)}$ and the same binary mask (M) used in the previous stage.
- Processing: The merged image $I^{L}_{(mer)}$ and the output of the previous stage $I^{L}_{(out)}$ concatenated with the binary mask (M). This combined input is then passed through the Global Refinement Network $(\mathrm{Net})\_\mathrm{G}$ to produce the global refinement output $I^{C}_{(out)}$.
- Output: The global refinement output $I^{C}_{(out)}$ represents the enhanced inpainting results obtained by leveraging global information. This stage is particularly effective in repairing large structures and long-distance texture patterns.
- Merging: To obtain the merged image $I^{C}_{(mer)}$ the valid (undamaged) regions from the original incomplete image $I_{(in)}$ are replaced with the corresponding regions from the global refinement output $I^{C}_{(out)}$ using a gray dotted line representation.

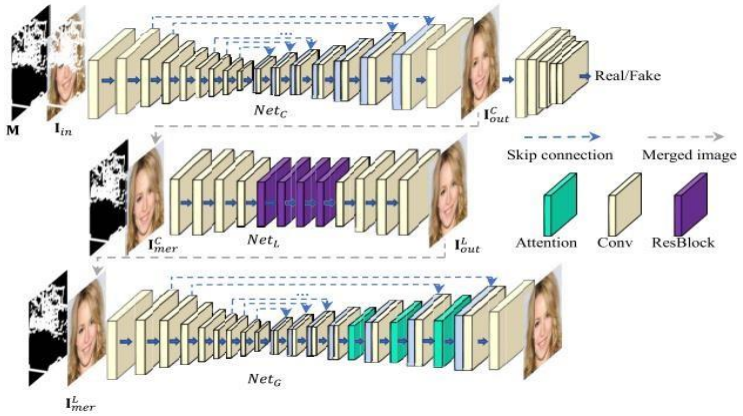

Fig. 1. Architecture of the proposed system

IV. RESULTS

	Masks	1-10%	10-20%	20-30%	30-40%	40-50%	50-60%
ℓ_1(%) ↓	PEN	0.80	2.15	3.88	5.83	8.02	11.77
	GConv	0.65	1.81	3.41	5.33	7.53	12.05
	MEDFE	1.02	2.15	3.68	5.51	7.65	11.67
	RFR	1.59	2.47	3.58	4.90	6.44	9.47
	MADF	0.47	1.30	2.40	3.72	5.26	8.43
	Ours	0.46	1.28	2.38	3.72	5.27	8.38
PSNR ↑	PEN	35.34	29.76	26.79	24.70	23.06	20.85
	GConv	37.14	31.02	27.57	25.03	23.10	20.22
	MEDFE	36.13	30.97	27.75	25.36	23.47	20.85
	RFR	36.39	31.87	29.07	26.87	25.09	22.51
	MADF	39.68	33.77	30.42	27.95	25.99	23.07
	Ours	40.04	33.99	30.54	27.99	26.01	23.12
SSIM ↑	PEN	0.988	0.965	0.933	0.894	0.849	0.764
	GConv	0.991	0.971	0.941	0.902	0.856	0.750
	MEDFE	0.990	0.971	0.943	0.908	0.865	0.775
	RFR	0.991	0.976	0.957	0.932	0.902	0.834
	MADF	0.995	0.984	0.967	0.945	0.917	0.848
	Ours	0.995	0.985	0.968	0.945	0.917	0.849
FID ↓	PEN	1.41	4.19	8.38	12.68	18.73	23.38
	GConv	0.78	2.05	3.93	5.86	8.64	12.75
	MEDFE	0.84	2.06	3.71	5.22	7.12	10.07
	RFR	0.86	1.68	2.67	3.77	5.21	7.60
	MADF	0.52	1.55	3.28	5.43	8.35	13.54
	Ours	0.39	1.06	2.08	3.16	4.61	7.07
LPIPS ↓	PEN	0.020	0.053	0.092	0.134	0.180	0.240
	GConv	0.012	0.034	0.061	0.091	0.125	0.181
	MEDFE	0.014	0.032	0.055	0.080	0.101	0.156
	RFR	0.015	0.028	0.042	0.060	0.081	0.118
	MADF	0.009	0.025	0.048	0.077	0.109	0.168
	Ours	0.006	0.017	0.031	0.048	0.069	0.108

Fig. 2. Comparisons with five state of art methods on CelebA-HQ Dataset

	Masks	1-10%	10-20%	20-30%	30-40%	40-50%	50-60%
ℓ_1(%) ↓	PEN	1.10	2.94	5.18	7.54	10.16	13.76
	GConv	1.16	3.03	5.30	7.66	10.28	14.24
	MEDFE	1.22	2.77	4.84	7.12	9.76	13.93
	RFR	0.83	2.20	3.93	5.83	7.96	11.37
	MADF	0.80	2.18	3.96	5.91	8.10	11.68
	Ours	0.68	1.89	3.51	5.33	7.41	10.86
PSNR ↑	PEN	33.42	27.90	25.09	23.21	21.74	20.07
	GConv	32.86	27.42	24.65	22.81	21.34	19.53
	MEDFE	34.08	29.05	25.92	23.78	22.07	19.93
	RFR	35.74	30.24	27.24	25.13	23.48	21.33
	MADF	36.17	30.37	27.17	25.00	23.31	21.10
	Ours	37.62	31.61	28.18	25.84	24.05	21.69
SSIM ↑	PEN	0.975	0.927	0.867	0.801	0.727	0.619
	GConv	0.968	0.917	0.856	0.792	0.722	0.610
	MEDFE	0.978	0.941	0.888	0.825	0.752	0.630
	RFR	0.983	0.952	0.911	0.862	0.805	0.699
	MADF	0.984	0.953	0.910	0.859	0.800	0.690
	Ours	0.988	0.963	0.925	0.878	0.823	0.714
FID ↓	PEN	4.60	11.65	20.78	31.12	45.72	60.43
	GConv	5.17	11.70	18.53	25.76	34.60	42.29
	MEDFE	3.59	8.76	15.12	22.15	30.43	40.72
	RFR	2.62	5.99	9.47	12.90	16.62	22.13
	MADF	2.15	5.58	9.20	13.08	17.36	24.42
	Ours	1.97	5.25	8.90	13.02	17.60	25.99
LPIPS ↓	PEN	0.035	0.093	0.160	0.226	0.295	0.365
	GConv	0.037	0.086	0.134	0.180	0.229	0.298
	MEDFE	0.028	0.063	0.105	0.150	0.201	0.268
	RFR	0.021	0.047	0.074	0.106	0.142	0.201
	MADF	0.014	0.038	0.068	0.102	0.141	0.209
	Ours	0.014	0.035	0.064	0.096	0.132	0.198

Fig.3. Comparisons with five state of art methods on Places2 Dataset

V. CONCLUSION

The proposed three-stage generative network for image inpainting based on the concept of receptive fields. The framework includes a coarse inpainting network with a large receptive field, a local refinement network with a small receptive field, and an attention-based global refinement network with a large receptive field. The experimental results demonstrate that the proposed method outperforms several state-of-the-art inpainting methods in terms of quantitative metrics and visual quality. It achieves visually realistic and accurate inpainting results on various datasets, including CelebA-HQ, Places2, and Paris StreetView. The method also shows robustness and versatility in real-world applications such as object removal, text editing, and logo removal, producing compelling results in each scenario.

VI. References

[1] C. Barnes, E. Shechtman, A. Finkelstein, and D. Goldman, "PatchMatch: A randomized correspondence algorithm for structural image editing," ACM Trans. Graph., vol. 28, no. 3, p. 24, 2009.

[2] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Medical Image Computing and Computer-Assisted Intervention—MICCAI 2015. Cham, Switzerland: Springer, 2015, pp. 234–241.

[3] W. Luo, Y. Li, R. Urtasun, and R. Zemel, "Understanding the effective receptive field in deep convolutional neural networks," in Proc. Adv. Neural Inform. Process. Syst., 2016

[4] I. Goodfellow et al., "Generative adversarial nets," in Proc. Adv. Neural Inform. Process. Syst., 2014.

[5] G. Liu, F. A. Reda, K. J. Shih, T.-C. Wang, A. Tao, and B. Catanzaro, "Image inpainting for irregular holes using partial convolutions," in Proc. Eur. Conf. Comput. Vis., Sep. 2018.

[6] Y. Zeng, J. Fu, H. Chao, and B. Guo, "Learning pyramid-context encoder network for high-quality image inpainting," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019.

[7] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in Proc. Eur. Conf. Comput. Vis., Oct. 2016.

[8] T. Miyato, T. Kataoka, M. Koyama, and Y. Yoshida, "Spectral normalization for generative adversarial networks," in Proc. Int. Conf. Learn. Represent., 2018.

[9] L. A. Gatys, A. S. Ecker, and M. Bethge, "Image style transfer using convolutional neural networks," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016.

[10] M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester, "Image inpainting," in Proc. ACM SIGGRAPH, 2000, pp. 417–424.

[11] D. Ding, S. Ram, and J. J. Rodríguez, "Image inpainting using nonlocal texture matching and nonlinear filtering," IEEE Trans. Image Process., vol. 28, no. 4, pp. 1705–1719, Apr. 2019.

[12] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 2536–2544.

[13] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, and T. Huang, "Free-form image inpainting with gated convolution," in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 4471–4480.

[14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. Int. Conf. Learn. Represent., 2015, pp. 1–14.