Structural Knowledge-Guided Feature Inference Network for Image Inpainting

Yongqiang Du
School of Information Engineering, Xinyang Agriculture and Forestry University
No.1 North Circular Road, Pingqiao District, Xinyang, 464000.
China

Received: June 21, 2021. Revised: December 15, 2021. Accepted: January 18, 2022. Published: January 27, 2022.

Abstract- Image inpainting is an essential task in image restoration field. Currently, most methods for image inpainting employ the encoder-decoder framework to restore degraded areas, and this often results in synthesizing wrong semantic structure due to the lack of guiding from effective prior information. In this paper, we propose a structural knowledge-guided framework for image inpainting, which predicts both the edge map and corrupted content at the same time. Our model captures structural knowledge in the structure estimation branch to guide the content inference in the latent feature space. By employing self-attention mechanism to aggregate known information and inferred structural knowledge, our model is able to synthesize more semantically reasonable content for the corrupted areas. Extensive experiments on three benchmark datasets demonstrate that our method outperforms most state-of-the-art methods for image inpainting in terms of the evaluation of both visual quality and quantitative metrics.

Keywords- Image inpainting, image restoration, feature representation, deep learning.

I. Introduction

Image inpainting gradually plays an essential role in the computer vision community due to its wide applications, including old image restoration[1], image edit[2], and so on. Main goal of image inpainting is to complete missing parts of image by inferring from known information. The crux of image inpainting stems from the difficulty of synthesizing plausible semantic content for the missing area. Thus, image inpainting is still extremely challenging in image restoration tasks.

Traditional methods for image inpainting mostly adopt exemplar-based strategy, and fill in content by diffusing pixels[3, 4] or similar patches replacement[5, 6]. Such traditional methods for image inpainting can well deal with corrupted images with repetitive textures due to the similarity of most patches. However, they fail to synthesize reasonable results while coping with complex semantic patterns, and usually fill in unreasonable structure or undesired artifacts, which decreases the image quality. This mainly because the content are reconstructed without high-level guidance, and the lack of semantic understanding limits the effectiveness of traditional methods for image inpainting.

Recently, image inpainting methods gradually make notable progress because of the fast develop of CNN-based deep generative models, such as the encoder-decoder based framework[7] and generative adversarial network[8]. By performing supervision on final output of deep model, their results have higher quality than traditional methods due to the excellent ability of distribution learning by deep models. To further enhance the restoration performance, some researches[9-12] employ attention mechanism[13] in deep models to strengthen the effectiveness of feature extraction. By weighting confidence for each pixel, deep models are able to cope with irregular corruption. Despite the quality of restoration results is improved, such methods fail to handle large corruption, and generate color discrepancy and undesired noise.

To solve above problems, some researches attempt to employ specific prior for improving performance of restoration. A prominent way[14, 15] is to employ high-frequency information, and obtain more reliable semantic boundary. In detail, it first predicts the edge map of restored images, and then puts the edge map and corrupted image together into the deep model. Although it can enhance the confidence of boundary information in some degree, the error of edge prediction might decrease the quality of final restored images. In addition, only leveraging the edge map might contribute little to the restoration due to the lack of explicit guide of feature inference from structural information. Thus, it's essential to infer corrupted content by effective structural knowledge.

In this paper, we propose a structural knowledge-guided feature inference network for image inpainting, and it is able to synthesize more reasonable content for the corrupted area. Specifically, we propose the structural knowledge-guided attention module to guide the content inference from structural knowledge in the latent feature space. The overall contribution of our method is as follows:

- We propose a novel framework for image inpainting, which employs the structural knowledge as guidance

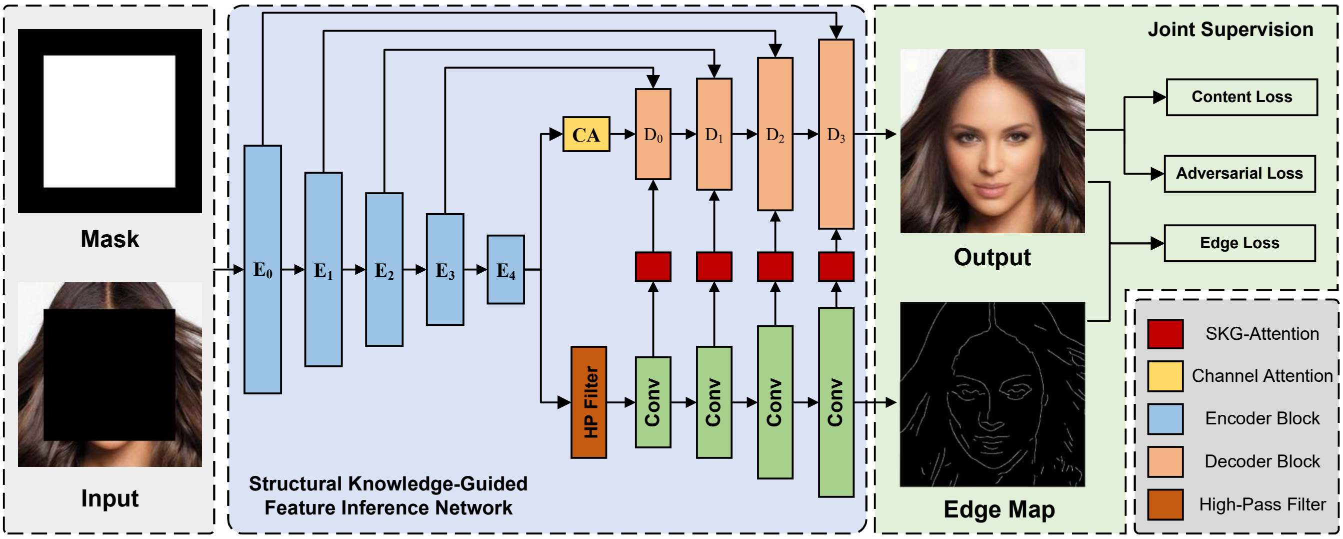

Fig. 1: Given an corrupted image, our *SK-FIN* infers the corrupted content by the cooperation of two types of feature inference: 1) pixel-level image reconstruction from the known pixels of the input image, and 2) structural reconstruction of the high-frequency constituents of the corrupted area. Our *SK-FIN* distills the structural knowledge from the branch of edge prediction. The learned structural knowledge are further used for pixel-level reconstruction by our designed Structural Knowledge-Guided(SKG) attention mechanism. Consequently, our *SK-FIN* is able to restore plausible image even dealing with large corruption ratio.

for inferring corrupted content. In addition, by employ the adversarial learning strategy with spectral normalization based discriminator, our method is able to synthesize more realistic content.

- We propose the structural knowledge-guided attention mechanism to facilitate restoration by aggregating structural information in latent feature space, and thus the model is able to generate more reasonable semantic boundary.
- Experiments on three benchmark datasets for image inpainting, including Paris Street View, CelebA-HQ, and Places2, demonstrate the effectiveness of our proposed method on both visual quality and quantitative metrics.

II. Related Work

There are a large mount of researches on image inpainting, and we select the most related and typical methods for reviewing. Methods for image inapainting can be thoroughly divided into two main types, non-learning based methods and learning based methods.

Traditional methods. Traditional methods for image inpainting include two main strategies, diffusion-based methods[3, 4, 16] and exemplar-based methods[5, 6]. The Diffusion-based methods expand neighboring pixels to fill in missing area. Some researches[16] attempt to use the isophote direction field to guide the pixel diffusion process. Then, Levin et al.[17] employ statistic histograms of local information to select the most similar pixels for inpainting damaged area. Such methods perform well on image smoothness and images with repetitive textures. However, the diffusion-based methods is only able to deal with images with small degradation such as scratches. In contrast, the patch-based methods can enhance the performance of image inpainting by computing the similarity for selecting the most proper patch and replacing the missing area. Efros et al.[18] first attempt to paste a image patch into a target image. Bertalmio et al.[19] propose patch search to find the most proper image patch. To accelerate the search efficiency, Barnes et al. propose the PatchMatch[20] methods for real-time image inpainting. However, the patch-based methods fail to work when the missing areas are larger due to the lack of semantic understanding for known content.

Deep learning based methods. With the fast development of deep learning[29, 39], CNN-based methods are able to extract semantic features to infer the structure of missing areas. And it facilitates to improve the performance of medical image processing[38]. The context encoder[21] is the first deep learning based method for image inpainting, which employs the encoder-decoder framework to construct the structure of missing area. Later, Iizuka et al.[22] leverage the dilated convolution to enlarge the receptive fields, and reconstruct the restored images. Liu et al.[9] propose the partial convolution, which normalizes the weights of convolutions, to infer missing pixels from the outside to the inside. For the excellent feature representation ability of attention mechanism, Yu et al.[10] propose the contextual attention to consider the consistency of known area and missing area. Then, they employ the attention mechanism to normalize deep features, and leverage the spectral normalization in discriminator to stabilize the adversarial learning.

Lately, some researches attempt to employ priors

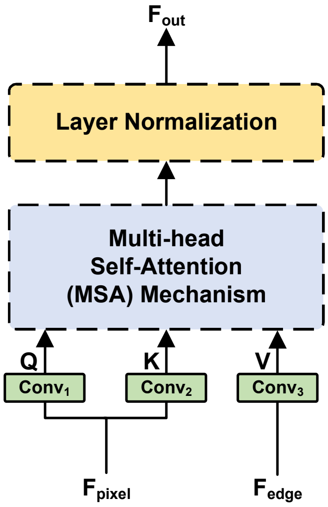

Fig. 2: The structure of our proposed structural knowledge guided self-attention mechanism.

for feature reasoning. Yeh et al.[23] trains a generative model as generative prior for image inpainting. By employing the variation auto-encoder[24], some researches[25] estimate the distribution of the complete image from known content. However, these methods suffer from the limitation of bluriness and undesired artifacts. To synthesize reasonable structure, some methods predict the structure information as structural prior to facilitate image inpainting. Nazeri et al.[14] first estimate the edge map, and then directly combine the edge map and input image together to reason the missing areas. Xiong et al.[26] propose a framework which infers the foreground objects of degraded images. Although this category of methods benefit from handcrafted priors, they cannot synthesize reasonable structure when the degradation areas are large. Thus, it's necessary to employ the structural information to guide the restoration process in the latent feature space to solve this problem.

III. Method

In this section, we introduce the architecture of the proposed structural knowledge-guided feature inference network for image inpainting. To begin with, we elaborate the main components in our framework. Then, we explain how the structural knowledge-guided attention mechanism leverages high-frequency information to guide the restoration process. Furthermore, we enumerate the loss functions during model training. Finally, we introduce the implementation details in our experiments.

A. Main Framework

Our proposed *SK-FIN* is a cooperative network, which includes two parallel branches in the decoder and infers the corrupted area by the guidance of structural information. As illustrated in Figure 1, the encoder $\mathcal{E}$ of our *SK-FIN* encodes the corrupted image into latent embeddings. In the structural estimation branch, we establish a simple decoder by stacking plain convolution layers due to the sparsity of high-frequency information. Then, by distilling the estimated structural knowledge of the corrupted area, the pixel-level reconstruction branch aggregates the known information and structural knowledge to accurately infer the content of corrupted area.

Formally, given an corrupted image I, our *SK-FIN* first encodes the known area into prior embeddings $\mathbf{e}$ by stacked encoder blocks:

$$\mathbf{e} = \sum_{i=1}^{N_e} \mathcal{E}_i(\mathrm{I}), \tag{1}$$

where N_e is the number of encoder blocks. To be specific, the encoder block of *SK-FIN* follows the structure of residual learning[35] to extract multi-level features from the input image:

$$\mathbf{F_{i+1}} = f(\mathbf{F_i}) + \mathbf{F_i}, \tag{2}$$

where $\mathbf{F_i}$ is the input feature of the i-th encoder block, and f denotes the convolution and ReLU layers. In addition, we downsample the size of features by stride-2 convolution.

Next, the *SK-FIN* separates the embeddings into two parallel branches, the structural knowledge estimation branch and pixel-level reconstruction branch. To extract the high-frequency information of input features, we employ the high-pass filter[37] to obtain the known structure $\mathbf{F}_{\text{high}}$:

$$\mathbf{F}_{\text{high}} = \mathcal{F}_{\text{high-pass}}(\mathbf{e}), \tag{3}$$

Due to the sparsity of structural knowledge, we stack plain convolution layers as decoder for estimating the edge map:

$$\mathbf{E} = \sum_{j=1}^{N_d} \mathcal{M}_j(\mathbf{F}_{\text{high}}), \tag{4}$$

where $\mathbf{E}$ is the predicted edge map, N_d is the number of decoder blocks, and $\mathcal{M}_j$ denotes the j-th decoder block in our *SK-FIN*.

B. Structure Knowledge Guided Attention

Differing from most methods that employ predicted edge maps to guide the pixel-level reconstruction, our *SK-FIN* proposes the structural knowledge guided(SKG) attention mechanism to aggregate the extracted features from input image and estimated structural knowledge. As illustrated in Figure 2, the SKG attention module captures feature correspondence between the known area and predicted structural information.

Following the transformer model[13], we crop the structural features F_{edge} into patches, and obtain the dependencies among patches by performing self-attention. Thus, the corrupted area can be reconstructed by the guidance of the predicted structural information in the latent feature space. In the pixel-level reconstruction phase, the SKG attention module first leverages different convolution layers to obtain the feature representations of two branches respectively:

$$\mathbf{Q}, \mathbf{K}, \mathbf{V} = f_1(\mathbf{F}_{\text{pixel}}), f_2(\mathbf{F}_{\text{pixel}}), f_3(\mathbf{F}_{\text{edge}}), \tag{5}$$

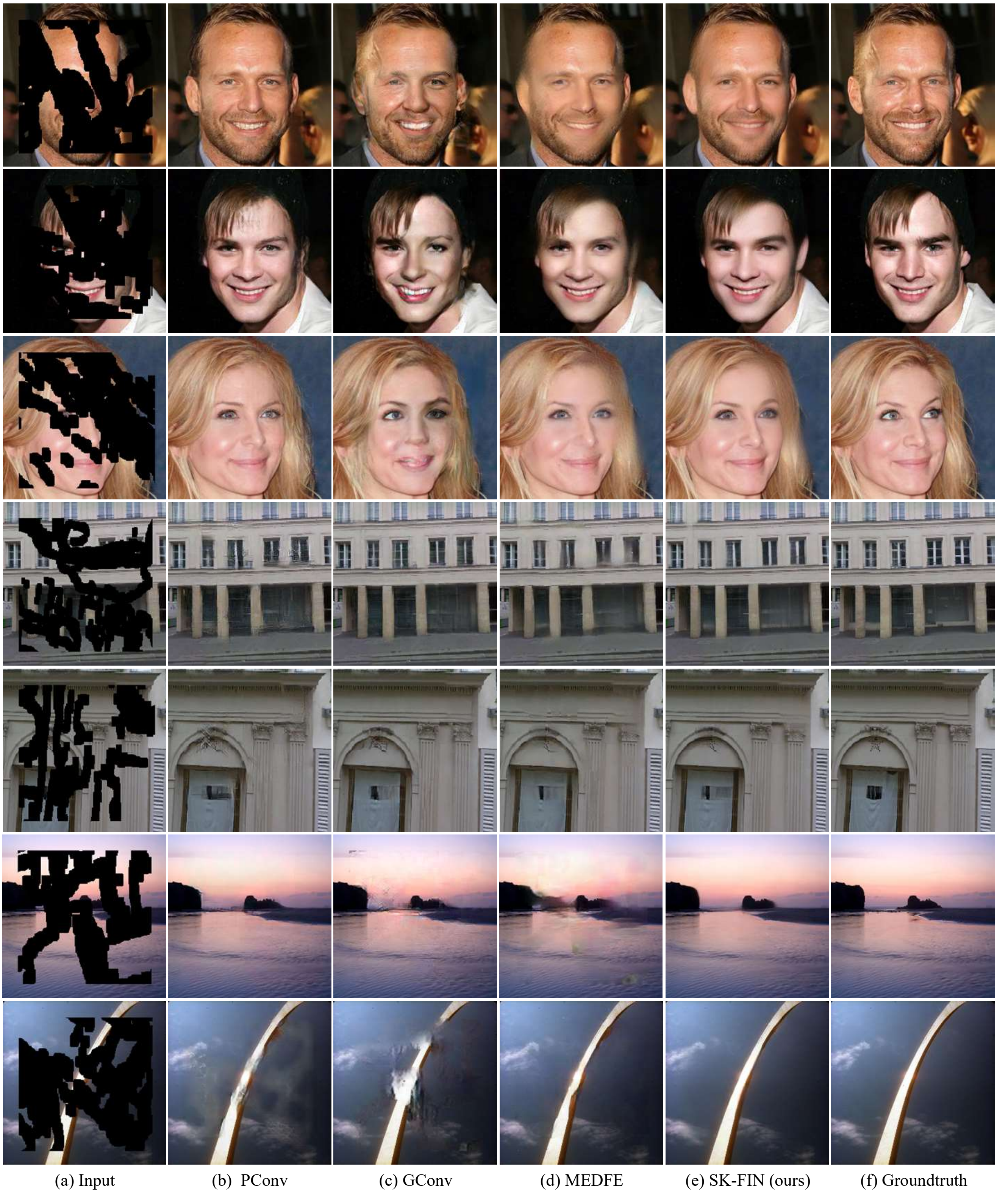

Fig. 3: Visualization of restoration results from four methods for image inpainting on seven randomly selected test images. Our *SK-FIN* is able to restore more reasonable results from corrupted images than most state-of-the-art methods.

where the $\mathbf{F}_{\text{pixel}}$ denotes the features from the pixel-level reconstruction branch, the $\mathbf{F}_{\text{edge}}$ is the features from the structural estimation branch, and f is the convolution layer. Then, we leverage the extracted feature representations to perform the multi-head self-attention(MSA) for capturing the long-term semantic dependencies in the image:

$$\mathbf{F}_{\text{out}} = \text{LN}(\text{MSA}(\mathbf{Q}, \mathbf{K}, \mathbf{V})), \tag{6}$$

where LN is the layer normalization[27] module, and $\mathbf{F}_{out}$ denotes the output features of SKG attention mechanism. Indeed, the decoder block of the pixel-level

Table 1: Quantitative comparison of the state-of-the-art methods for image inpainting and our *SK-FIN* on the Paris StreetView [30], CelebA[31], and Places[32] datasets. Best results are highlighted in bold. ↑ notes that higher is better, but ↓ notes lower is better.

	Method	Paris StreetView			CelebA			Places2		
		20-30%	30-40%	40-50%	20-30%	30-40%	40-50%	20-30%	30-40%	40-50%
PSNR ↑	PConv[9]	23.34	21.37	20.73	25.91	23.42	21.01	23.03	20.99	19.21
	GConv[11]	23.99	21.86	19.79	25.52	23.19	20.64	23.00	20.94	18.68
	EdgeConnect[14]	24.91	23.35	20.23	25.97	23.62	20.19	22.99	20.86	18.97
	CA[10]	22.96	21.32	19.26	26.92	25.01	21.07	23.99	21.96	19.48
	LBAM[33]	25.04	23.12	20.76	25.95	23.91	21.73	23.04	21.55	19.36
	MEDFE[34]	24.32	22.25	19.97	23.89	22.12	19.96	23.02	21.29	18.31
	Lizuka *et al.*[22]	24.69	22.43	21.02	25.76	24.32	21.24	23.33	21.52	19.39
	SK-FIN (ours)	**25.69**	**23.49**	**21.17**	**27.03**	**25.51**	**22.70**	**24.93**	**22.91**	**20.60**
SSIM ↑	PConv[9]	0.792	0.708	0.625	0.844	0.779	0.705	0.769	0.688	0.609
	GConv[11]	0.799	0.716	0.622	0.841	0.772	0.673	0.785	0.722	0.636
	EdgeConnect[14]	0.819	0.737	0.646	0.855	0.784	0.682	0.797	0.714	0.643
	CA[10]	0.821	0.745	0.643	**0.889**	0.814	0.713	0.812	0.734	0.649
	LBAM[33]	0.807	0.723	0.648	0.855	0.789	0.706	0.797	0.712	0.633
	MEDFE[34]	0.809	0.711	0.607	0.802	0.747	0.656	0.801	0.729	0.648
	Lizuka *et al.*[22]	0.809	0.725	0.647	0.846	0.813	0.722	0.813	0.729	0.649
	SK-FIN (ours)	**0.831**	**0.761**	**0.678**	**0.889**	**0.821**	**0.733**	**0.821**	**0.736**	**0.661**

branch also employs the residual block as core unit to reconstruct the image, and upsamples feature maps by bilinear interpolation:

$$\hat{\mathrm{I}} = \sum_{i}^{N_d} \mathcal{D}_i(\mathbf{F}_{\mathrm{out}}), \tag{7}$$

where N_d is the number of decoder layers, and $\hat{\mathrm{I}}$ is the restored image.

C. Loss Functions for Supervision Learning

We optimize the parameters of our *SK-FIN* in an end-to-end manner supervised by three types of loss functions.

Edge Reconstruction Loss. To accurately estimate the structural knowledge of corrupted area, we employ L_1 distance to optimize the predicted edge map $\hat{E}$ and the groundtruth:

$$L_{\mathrm{edge}} = ||E_{gt} - \hat{E}||_1. \tag{8}$$

In detail, we obtain the groundtruth of edge map by the canny algorithm of edge detection[28].

Content Loss. Our content loss includes two main parts, the L_1 distance of pixel-wise reconstruction L_{pixel} and the perceptual loss L_{perc}:

$$L_{\mathrm{content}} = L_{\mathrm{pixel}} + L_{\mathrm{perc}}. \tag{9}$$

The pixel-wise loss calculates the L_1 distance between restored image and groundtruth image in pixel level, and it is defined as follows:

$$L_{\mathrm{pixel}} = ||\mathrm{I}_{gt} - \hat{\mathrm{I}}||_1. \tag{10}$$

The perceptual loss[36] facilitate the deep model to learn semantic consistency between restored image and groundtruth image, and it is defined as follows:

$$\mathcal{L}_{\mathrm{perc}}(\hat{\mathrm{I}}, \mathrm{I}_{gt}) = \sum_{l=1}^{L} \|f^l_{\mathrm{vgg}}(\hat{\mathrm{I}}) - f^l_{\mathrm{vgg}}(\mathrm{I}_{gt})\|_1, \tag{11}$$

where $f^l_{\mathrm{vgg}}(\hat{\mathrm{I}})$ and $f^l_{\mathrm{vgg}}(\mathrm{I}_{gt})$ are the feature maps extracted by the l-th layer of pre-trained VGG-19[29].

Adversarial Loss. To improve the quality of synthesized content, we employ the adversarial learning strategy[8] by an additional discriminator:

$$\mathcal{L}_{\mathrm{adv}} = -\mathbb{E}_{\mathrm{I} \backsim \mathbb{P}_{\text{SK-FIN}}}[D(G(\mathrm{I}))], \tag{12}$$

where G and D are the *SK-FIN* and discriminator respectively. In this way, the distributions of restored images and groundtruth images are similar as much as possible.

In sum, the whole loss functions of our *SK-FIN* are defined as follows:

$$\mathcal{L} = \lambda_1 \mathcal{L}_{\mathrm{edge}} + \lambda_2 \mathcal{L}_{\mathrm{content}} + \lambda_3 \mathcal{L}_{\mathrm{adv}}, \tag{13}$$

where λ_1, λ_2, and λ_3 are hyper-parameters to balance between different losses. In our experiments, we empirically set λ_1=1, λ_2=0.1, and λ_3=0.01.

IV. Experiments

In this section, we conduct extensive experiments to evaluate the performance of our proposed *SK-FIN*. First, we elaborate the datasets to evaluate the performance of different methods for image inpainting. Then, we introduce the implementation details while training our model. Next, we perform quantitative and qualitative comparisons with the state-of-the-art methods for image inpainting, and analyzes the results. Finally, we conduct ablation study to check the effectiveness of each component of our *SK-FIN*.

Table 2: Performance comparison of four ablation variants of our *SK-FIN* with 30%-40% irregular corruption on two benchmark datasets in terms of *PSNR* and *SSIM*.

Dataset	Metrics	w/o Edge	w/o MSA	w/o L_{adv}	*SK-FIN*
Paris Street View[30]	PSNR	19.97	21.09	22.06	**23.49**
	SSIM	0.730	0.749	0.757	**0.761**
CelebA[31]	PSNR	22.13	23.22	24.17	**25.51**
	SSIM	0.793	0.806	0.816	**0.821**

A. Datasets

Three benchmark datasets are employed to evaluate the performance of different methods for image inpainting. **Paris Street View**[30], which is captured from the streetview of Paris includes altogether 14900 training images, and 100 test images. **CelebA**[31], which includes 30000 images of human faces. We random select 29000 images as training set and 1000 images as test set. **Places2**[32], which includes more than 2,000,000 images of different scenes. We random select three scenes altogether 120000 images. And we random select 117000 images as training set and 3000 images as test set. Furthermore, we evaluate the results of various methods by the PSNR and SSIM.

B. Implementation Details

We conduct our experiments under the PyTorch. We first resize the training data into 284×284, and then random crop into 256×256. We employ the Adam to optimize the parameters of our *SK-FIN*. In addition, the learning rate is decreasing from 5e-4 to 1e-4 during altogether 100 epochs. All experiments are conducted with 2 2080ti GPUs. The batchsize is 4, and augmentations are employed during training, including random flipping, resizing, and rotation.

C. Results Analysis

In this section, we analyze the results of different methods in the experiments for image inpainting to demonstrate the effectiveness of our proposed *SK-FIN*.

The Table 1 lists the quantitative comparison of our *SK-FIN* and other state-of-the-art methods[9, 10, 11, 14, 22, 33, 34] for image inpainting. In terms of PSNR, our method outperforms all competing methods for a large margin, even if the corruption ratio is high. In terms of SSIM metric, our *SK-FIN* is able to obtain higher performance than most state-of-the-art methods for image inpainting when the corruption area becomes larger. For the CelebA dataset, the CA[10] employs contextual attention in the spatial feature to obtain more accurate content inference, but our *SK-FIN* still outperforms CA's performance in large corruption ratio due to the guidance of structural knowledge in the edge map estimation branch.

As illustrated in Figure 3, our *SK-FIN* is able to generate the most reasonable visual results than other competing methods. This demonstrates again the advantage of our designed structural knowledge guidance than traditional encoder-decoder based framework. For instance, we synthesize more realistic human face in the third row in Figure 3, and accurate semantic structure with less artifacts than other image inpainting methods in the seventh row. In sum, our *SK-FIN* is able to restore corrupted images with large corruption raio. It keeps the color fidelity of generated image, and reduces the blurriness and undesired artifacts than most state-of-the-art methods.

D. Ablation Study

To further demonstrate the effectiveness of each functional component in our proposed *SK-FIN*, we conduct ablation study experiments: w/o Edge notes to train our *SK-FIN* without structural estimation branch; w/o MSA fuses the structural knowledge without multi-head self-attention mechanism; w/o L_{adv} denotes to train our *SK-FIN* without the adversarial loss function; And Complete is the full model of our proposed *SK-FIN*. Table 2 lists the performance of four variants of our *SK-FIN*, and the results are increasingly better. This demonstrates the necessity of all functional components in our *SK-FIN*.

V. Conclusion

In this paper, we propose a structural knowledge guided feature reasoning network for single image inpainting. Inspired by the multi-head attention mechanism in the transformers, we improve it to aggregate the structural information and pixel information to reconstruct the corrupted area. By distilling high-frequency features in the structure estimation branch, the pixel-level reconstruction branch is able to infer the corrupted content more accurately. Extensive experiments on three benchmark datasets demonstrate that our *SK-FIN* is able to restore more reasonable content than most state-of-the-art methods for image inpainting.

References

[1] Wan Z, Zhang B, Chen D, et al. Bringing old photos back to life[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 2747-2757.

[2] Xu N, Price B, Cohen S, et al. Deep image matting[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 2970-2979.

[3] Ballester C, Bertalmio M, Caselles V, et al. Filling-in by joint interpolation of vector fields and gray levels[J]. IEEE Transactions on Image Processing, 2001, 10(8): 1200-1211.

[4] Bugeau A, Bertalmio M. Combining Texture Synthesis and Diffusion for Image Inpainting[C]//VISAPP 2009-Proceedings of the Fourth International Conference on Computer Vision Theory and Applications. 2009: 26-33.

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

[6] Li Z, He H, Tai H M, et al. Color-direction patch-sparsity-based image inpainting using multidirection features[J]. IEEE Transactions on Image Processing, 2014, 24(3): 1138-1152.

[7] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[8] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.

[9] Liu G, Reda F A, Shih K J, et al. Image inpainting for irregular holes using partial convolutions[C]//Proceedings of the European Conference on Computer Vision. 2018: 85-100.

[10] Yu J, Lin Z, Yang J, et al. Generative image inpainting with contextual attention[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 5505-5514.

[11] Yu J, Lin Z, Yang J, et al. Free-form image inpainting with gated convolution[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 4471-4480.

[12] Yu T, Guo Z, Jin X, et al. Region normalization for image inpainting[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2020, 34(07): 12733-12740.

[13] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[C]//Advances in Neural Information Processing Systems. 2017: 5998-6008.

[14] Nazeri K, Ng E, Joseph T, et al. Edgeconnect: Structure guided image inpainting using edge prediction[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops. 2019: 0-0.

[15] Ren Y, Yu X, Zhang R, et al. Structureflow: Image inpainting via structure-aware appearance flow[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 181-190.

[16] Bertalmio M, Sapiro G, Caselles V, et al. Image inpainting[C]//Proceedings of the 27th annual conference on Computer graphics and interactive techniques. 2000: 417-424.

[17] Levin A, Zomet A, Weiss Y. Learning How to Inpaint from Global Image Statistics[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2003, 1: 305-312.

[18] Efros A A, Freeman W T. Image quilting for texture synthesis and transfer[C]//Proceedings of the 28th annual conference on Computer graphics and interactive techniques. 2001: 341-346.

[19] Bertalmio M, Vese L, Sapiro G, et al. Simultaneous structure and texture image inpainting[J]. IEEE transactions on image processing, 2003, 12(8): 882-889.

[20] Barnes C, Shechtman E, Finkelstein A, et al. PatchMatch: A randomized correspondence algorithm for structural image editing[J]. ACM Trans. Graph., 2009, 28(3): 24.

[21] Pathak D, Krahenbuhl P, Donahue J, et al. Context encoders: Feature learning by inpainting[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 2536-2544.

[22] Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion[J]. ACM Transactions on Graphics, 2017, 36(4): 1-14.

[23] Yeh R A, Chen C, Yian Lim T, et al. Semantic image inpainting with deep generative models[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 5485-5493.

[24] Kingma D P, Welling M. Auto-encoding variational bayes[J]. arXiv preprint arXiv:1312.6114, 2013.

[25] Peng J, Liu D, Xu S, et al. Generating Diverse Structure for Image Inpainting With Hierarchical VQ-VAE[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 10775-10784.

[26] Xiong W, Yu J, Lin Z, et al. Foreground-aware image inpainting[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 5840-5848.

[27] Ba J L, Kiros J R, Hinton G E. Layer normalization[J]. arXiv preprint arXiv:1607.06450, 2016.

[28] Canny J. A computational approach to edge detection[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1986 (6): 679-698.

[29] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[30] Doersch C, Singh S, Gupta A, et al. What makes paris look like paris?[J]. ACM Transactions on Graphics, 2012, 31(4).

[31] Karras T, Aila T, Laine S, et al. Progressive growing of gans for improved quality, stability, and variation[J]. arXiv preprint arXiv:1710.10196, 2017.

[32] Zhou B, Lapedriza A, Khosla A, et al. Places: A 10 million image database for scene recognition[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 40(6): 1452-1464.

[33] Xie C, Liu S, Li C, et al. Image inpainting with learnable bidirectional attention maps[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 8858-8867.

[34] Liu H, Jiang B, Song Y, et al. Rethinking image inpainting via a mutual encoder-decoder with feature equalizations[C]//Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16. Springer International Publishing, 2020: 725-741.

[35] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[36] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//European conference on computer vision. Springer, Cham, 2016: 694-711.

[37] Stephane M. A wavelet tour of signal processing[J]. 1999.

[38] Constance Barson, Reza Saatchi, Prasad Godbole, Shammi Ramlakhan. Infrared Thermal Imaging to Detect Inflammatory Intra-Abdominal Pathology in Infants[J]. WSEAS Transactions on Biology and Biomedicine, pp. 82-98, Volume 17, 2020

[39] Feng X, and Kan JM. "A pseudo entropy based self-organizing neural network for nonlinear system." International Journal of Circuits, Systems and Signal Processing 13 (2019): 266-272.