Journal of Computers Vol. 37 No. 4, August 2026, pp. 191-209
https://doi.org/10.63367/19911599202608370401 2

A Novel Image Inpainting Model Based on Multi-scale Parallel Dense Connection Network

Jie Wang[1], Liyuan Zhang[1], Yibo Deng[2], and Xiaoxuan Ma[1*]

[1] School of Intelligence Science and Technology, Beijing University of Civil Engineering and Architecture, Beijing 102616, China

wangjie@bucea.edu.cn, zhangliyuan092@gmail.com, maxiaoxuan@bucea.edu.cn

[2] Beijing Huimin Digital Technology Co., Ltd, Beijing 100080, China

dengyibo@bfhg.com.cn

Received 17 April 2026; Revised 3 June 2026; Accepted 14 July 2026

Abstract. Existing image inpainting methods often fail to effectively exploit multi-scale feature information and lack sufficient interaction between shallow and deep representations, leading to unsatisfactory restoration performance in damaged regions. In addition, structural discontinuities and blurred details frequently occur near the boundaries of missing areas. To address these challenges, this paper proposes a novel image inpainting framework based on a Multi-Scale Parallel Dense Connection Network (MSPDCN). Specifically, holistically nested edge detection (HED) is first employed to extract structural priors and estimate edge information of missing regions, which provides guidance for subsequent reconstruction and alleviates boundary blurring. Subsequently, a multi-scale parallel dense connection module is integrated into the generation network to capture features with different receptive fields and to strengthen the interaction between hierarchical representations, thereby improving feature utilization. Extensive experiments are conducted to evaluate the proposed model against state-of-the-art methods, including GAN-based and diffusion-based approaches. Compared with the contextual attention (CA) model, the proposed method achieves improvements of 13.09% and 5.12% in PSNR and SSIM, respectively, while reducing LPIPS by 21.72%. The results demonstrate that the proposed MSPDCN not only enhances visual quality but also improves robustness to interference. Overall, the proposed approach consistently outperforms existing methods in image inpainting tasks.

Keywords: image inpainting, multi-scale feature extraction, parallel structure, dense connection, generative adversarial networks

1 Introduction

Image inpainting or image completion is the completion or restoration of pixel information in missing regions of damaged images, and its objective lies in ensuring semantic accuracy and maintaining visual coherence. Using image inpainting technology, the missing information of an image or video can be inpainted, and the inpainted results can further support subsequent image processing tasks.

Within the field of image inpainting, methods are mainly divided into traditional texture or diffusion-based methods and deep learning-based methods. Among them, traditional texture or diffusion-based methods are alternatively referred to as non-deep learning methods. For instance, Efros et al. and Kwatra et al. put forward nonparametric sampling [1] and texture optimization methods for texture synthesis [2]. Ding et al. utilized non-local texture matching and nonlinear filtering to perform image inpainting [3]. Erkan et al. employed the concept of pixel value similarity as prior knowledge [4]. Wang et al. used structure-consistent patch matching for image inpainting [5]. Song et al. proposed to use context information for image inpainting [6]. The texture, structure and other information in the image can be well utilized in these methods, but these traditional methods only use the shallow feature information in the image, and the inpainting effect for complex damaged images is not ideal.

The second category of image inpainting methods is driven by deep learning techniques, which aim to enhance representation capability by learning hierarchical features from large-scale data. With the rapid

* Corresponding Author

development of deep neural networks, a variety of inpainting architectures have been proposed to improve reconstruction quality. Among them, generative adversarial networks (GANs) [7] and diffusion models (DMs) [8] have become two dominant frameworks. The GAN-based framework is characterised by the presence of two key components: a generator and a discriminator. These components are subject to a process of joint optimisation through the utilisation of adversarial learning. This strategy enables the generator to produce more realistic and semantically consistent inpainting results. Based on this paradigm, numerous approaches have been developed, such as the context encoder [9] and the multi-column convolutional neural network [10]. Furthermore, Liu et al. introduced partial convolution to effectively handle irregular missing regions by utilizing only valid pixels during feature learning [11]. To further address the challenge of free-form masks, Yu et al. proposed a gated convolution mechanism that adaptively learns spatial features for more flexible image completion [12]. In contrast, diffusion-based methods formulate image inpainting as a progressive denoising process, where noise is gradually injected into the image and then removed through a reverse process to recover the missing content. Despite their strong generative capability, these approaches often suffer from semantic inconsistency between masked and unmasked regions, and they typically require high computational cost during training and inference. To mitigate these limitations, LatentPaint [13] performs forward–backward fusion in latent space to improve efficiency, while StrDiffusion [14] introduces structure guidance to better preserve image content during the denoising process. Overall, existing deep learning-based inpainting methods are able to reconstruct missing regions by leveraging learned feature representations. However, their performance is still limited when handling large missing areas or images with complex structures. In such cases, artifacts, structural inconsistencies, and boundary blurring are likely to occur, leading to unsatisfactory visual quality.

In order to better extract the feature information in the image, two optimization methods are proposed. The first method is often adopting the way of deepening the number of layers of the network or adopting a multi-stage inpainting architecture in the existing image inpainting models. But these will make the instability of model training and the risk of gradient disappearance and network degradation increase, the training process of the network becomes more challenging. To address the gradient disappearance problem in deep networks, He et al. proposed a deep residual learning network (ResNet) with a specialized skip connection approach [15]. Subsequently, Huang et al. proposed Dense Convolutional Network (DenseNet) [16], which has the capacity to further mitigate the risk of gradient vanishing during network training. Concurrently, dense connection will enhance the efficacy of feature transmission and ensure the comprehensive utilisation of feature information in images. The other method is that the middle layers of the generating network often introduce several hole convolution layers with different expansion rates for feature extraction based on the full convolution network structure. The use of hole convolutions with different expansion rates can extract features in different ranges, thus improving the accuracy of inpainting. GLCIC [17] uses four hole convolutions with expansion rates of 2, 4, 8 and 16 to extract features in series in the generating network part, and uses global and local discriminators to pay attention to global information and local information, respectively, to improve the image inpainting effect. EdgeConnect [18] uses the combination of hole convolutions and residual blocks to extract the middle layer features in series and proposes the combination of “image to edge” and “edge to image” for the first time to inpaint damaged images. MFMAM fused a multi-scale feature module based on dilated convolution and improved attention module, which can strengthen the ability of image semantic inpainting [19]. E2I [20] uses the series connection of four hole convolutions with different expansion rates to extract features in the two-stage network and adopts the strategy of “edge to image” in the process of image inpainting. In image inpainting, the strategy of “image to edge” is mainly used to solve the problem that the edge of damaged image inpainting results will appear as artifacts and blur. In addressing this challenge, it is customary to incorporate an edge detector into the image inpainting network. This enables the extraction of edge information from the original image, which functions as prior knowledge for image inpainting. This, in turn, facilitates the inpainting of damaged regions. It is evident that these edges serve to preserve the key points and unexpected details of the image. Furthermore, it can be posited that the judicious use of prior knowledge in the image can yield a superior image inpainting effect. LFang [21] introduces a Spatial-Transformer (ST) module consisting of a convolutional neural network (CNN) and transformer branches to fuse the structure and texture features to improve the inpainting results. SG-GAN uses self-prior guided to obtain reconstructed high-quality results [22]. However, different hole convolutions are connected in series, so that the latter layers will extract features on the basis of the previous layers, which will reduce the effect of extracting global and local information. At the same time, the latter layers depend on the features of the previous layers, which will reduce the accuracy of extracting feature information and the completeness of feature utilization.

In order to capitalise on the full potential of the feature information contained within the image, it is necessary to reduce the absolute dependence between the front and back layers in the feature extraction process.

Furthermore, the connection between different scales in the image must be strengthened. To this end, this paper proposes an image inpainting model based on a multi-scale parallel dense connection network (MSPDCN) on the basis of gated convolution network architecture. Firstly, the holistically nested edge detection (HED) is employed for preprocessing the damaged image, aiming to predict edge details in the missing areas and thereby assist in reducing artifacts during the subsequent inpainting process. Then, in the generating network part, multi-scale parallel connected hole convolution is used to extract features, which reduces the absolute dependence between the front and back layers. At the same time, the layers of hole convolution are densely connected, which improves the connection between shallow and deep feature information, makes better use of the feature information of different scales, and realizes more sufficient reuse of features between different layers of hole convolutions. Finally, the MSPDCN is trained on the standard CelebA and Places2 datasets and compared against state-of-the-art methods, including the GAN-based contextual attention (CA) [23], E2I, and gated convolution (GC), the transformer-based LFang, and the DM-based StrDiffusion. Experimental findings have demonstrated that the MSPDCN is capable of enhancing objective evaluation metrics and visual effects in instances of inpainting damaged images. It is possible to summarise the primary contributions of the paper as such.

- This paper introduces a multi-scale parallel dense connection network (MSPDCN) for damaged image inpainting, which fully leverages image prior knowledge and feature information to enhance inpainting performance.

- In order to enhance the quality of the inpainting, we propose a multi-scale parallel dense connection module (MSPDCM). This module functions to reduce the absolute dependence between the front and back layers in the feature extraction process. Furthermore, it serves to strengthen the connection between different scales in the images.

- We conduct a comprehensive evaluation of the MSPDCN by comparing its inpainting effects qualitatively and quantitatively with the state-of-the-arts. Testing on the CelebA and Places2 datasets shows that the MSPDCN not only yields improved inpainting results and visual effects but also optimizes the structural texture details and anti-interference ability of the inpainted areas.

The rest of this article is organized as follows. In Section 2, we introduce the related work of prior knowledge, dense connection mechanism and multi-scale parallel connection in image inpainting. Section 3 describes the proposed methods. The experimental results and analysis are given in Section 4. Finally, a summary of the thesis is presented in Section 5.

2 Related Work

2.1 Prior Knowledge for Image Inpainting

Inpainting the damaged area directly is a strategy used by image inpainting algorithms, and good inpainting results can be obtained for images with simple texture structure or less damage. With the development of computer technology, more and more image information is captured by the camera, and the information contained in the image is richer and the texture structure is more complex. It becomes more difficult to inpaint the damaged area only by relying directly, and there will be problems such as blur, artifacts or mismatch of texture structure after inpainting. In order to solve these problems, the prior knowledge in the image is used to assist the image inpainting, such as edge information, color, and prior knowledge of deformation, so as to improve the inpainting effect.

Incorporating edge information as prior knowledge has been widely recognized as an effective strategy for improving image inpainting performance, since structural cues can provide important guidance for reconstructing missing regions. Edge detection aims to identify meaningful boundaries and salient contours from natural images, which play a crucial role in preserving structural consistency while maintaining fine-grained details. In image inpainting tasks, extracted edge features are typically used as auxiliary information to guide the restoration process, helping to alleviate structural discontinuity and reduce boundary artifacts. With the rapid development of computer vision techniques, numerous edge detection methods have been proposed, ranging from classical approaches such as the Canny detector [24] to deep learning-based models, including BDCN [25], RCF [26], and HED [27]. More recently, Poma et al. employed DexiNed [28] to extract edge representations and further incorporated edge connectivity constraints into the inpainting process [29], leading to improved reconstruction quality. These studies demonstrate that introducing edge priors can effectively enhance structural coherence and contribute to more visually plausible inpainting results.

In order to address the issue of blurred and damaged edges, the image inpainting process incorporates prior knowledge. EdgeConnect firstly inpainted the extracted edge information and then used the inpainting results as prior knowledge to assist the subsequent image inpainting. In their seminal study, Chen et al. [30] proposed a novel fusion context transformation network (FCTN) driven by semantic priors for the image inpainting task. This innovative network employs a bespoke semantic prior generator that facilitates the transformation of the semantic features present in real images and the low-level features of damaged images into semantic priors. Li et al. proposed an image inpainting method based on effective feature fusion and processed it using a two-stage network [31]. The external and internal features are fused by using the dynamic memory network (DMN+), which generates an optimized version of the incomplete image.

2.2 Dense Connection

In order to extract more deep feature information from the image, the method of deepening the layers of the network model is often adopted. However, the deepening of network layers will bring huge parameters to the model, and the risk of gradient disappearance or over-fitting will also increase. In order to address the issues previously mentioned, arising from the parameter surge induced by increased network layers, the Dense Convolutional Network (DenseNet) was proposed by Huang et al. This new network is capable of further reducing the risk of gradient vanishing during network training. The schematic diagram of the dense connection mechanism is shown in Fig. 1.

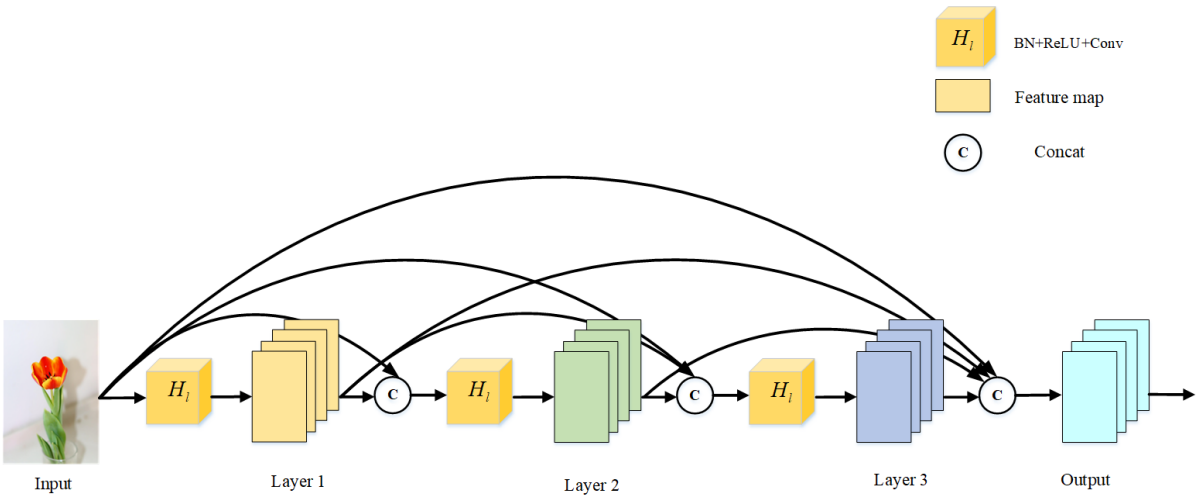

Fig. 1. Schematic diagram of dense connection mechanism

By establishing dense connections between upper and lower layers, each layer is directly linked to all preceding layers, ensuring efficient information flow across the network. As a result, the input of each subsequent layer consists of the feature maps generated by all previous layers, enabling strong interactions between earlier and later stages and promoting effective feature reuse. In addition, this connectivity pattern forms implicit links among convolutional layers, which helps improve training stability and alleviates optimization difficulties commonly observed in deep neural networks as the network depth increases.

This special dense connection structure has the following advantages:

(1) It enables more effective transmission of features and gradients, thereby simplifying the network training process.

(2) It promotes the efficient utilization of feature information in subsequent tasks.

2.3 Multi-Scale Parallel Connection

In order to extract the feature information of different positions and scales in the image, it is difficult to fully extract the feature information of the image only by using convolution kernels with the same size. In order to solve this problem, convolution layers of convolution kernels with different sizes are often used to extract features in the network. GLCIC takes four hole convolutions with different expansion rates in series to extract image features in the generating network part. Yu et al. used a series of hole convolutions with different expansion rates to extract image features in the two-stage inpainting model. Then, on the basis of this model, the learning mode of convolution is improved, and gated convolution is proposed to optimize the inpainting effect of irregular damaged areas. The hole convolution layers connected in series can reasonably use the feature information of different scales in the image and simplify the model construction, but it will make the latter layer extract feature on the basis of the previous layer in the extraction process, and the extraction effect of the latter layer will absolutely depend on the extraction effect of the previous layer, which will reduce the accuracy of actual feature information and the completeness of feature utilization.

In order to reduce the dependency between different convolution layers and improve the utilization of feature information of different scales in the image, a parallel feature extraction network is proposed, and then the extracted results are fused by feature fusion to optimize the effect of image inpainting. Li et al. proposed a damaged image inpainting algorithm based on multi-scale features based on GLCIC architecture, and the generating network used a parallel multi-scale hole convolution module to extract multi-scale features to improve the inpainting effect [32]. Jin et al. proposed an image inpainting algorithm based on multi-feature fusion and multi-scale generative adversarial network to improve the details and texture features of the inpainted image [33]. Jiang et al. proposed a new multi-scale sensing module composed of a multi-scale sensing block, intra-layer residual connection and convolution block attention module in order to obtain accurate content and clear details, aiming at effectively capturing scale diversity features [34]. Relatively independent parallel structure can indeed improve the image inpainting effect, but there is often a certain relationship between shallow and deep features in the image, and the relative relationship between features should be strengthened between layers.

3 Proposed Methods

The existing image inpainting methods are not optimally utilising the feature information of different scales in the image, and there is an absence of connection between shallow and deep features. This results in a suboptimal inpainting effect for damaged images. At the same time, discontinuous features and blurred details remain at the edge of the damaged area. In order to solve these problems, this paper proposes an image inpainting model based on the MSPDCN. The model in question is capable of reducing the absolute dependence between front and back layers during the process of feature extraction, strengthening the connection between various scales in the image, and making full use of feature information in the image itself. Section 3.1 mainly introduces the structure of the MSPDCN, Section 3.2 mainly introduces the process of inpainting damaged images, and Section 3.3 details the core multi-scale parallel dense connection module of the MSPDCN.

3.1 Multi-Scale Parallel Dense Connection Network Architecture

The comprehensive architectural framework of the image-inpainting model, grounded in the MSPDCN delineated in this study, is elucidated in Fig. 2.

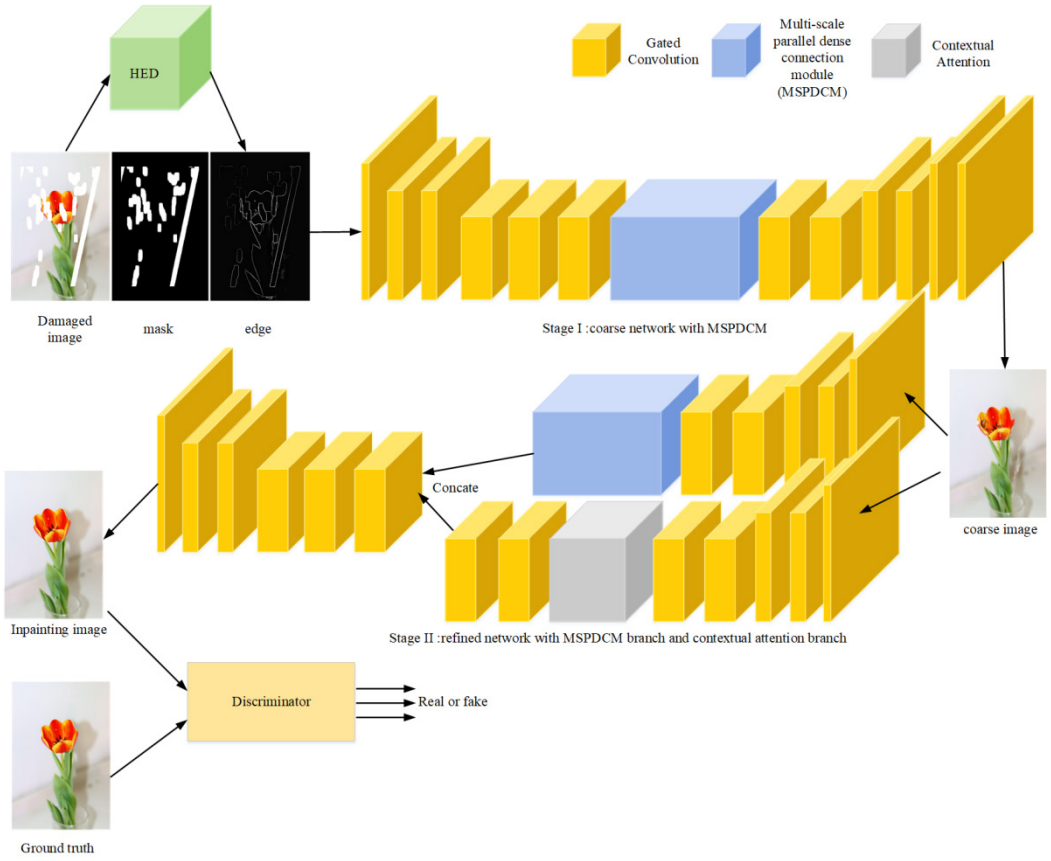

Fig. 2. Inpainting model architecture of multi-scale parallel dense connection network (MSPDCN)

The MSPDCN is built on the gated convolution network framework. The model is composed of three constituent elements: a preprocessing module (image preprocessing), a generator module (image inpainting), and a discriminator module (result discrimination). The primary structure of the preprocessing module is composed of HED. This end-to-end edge detection network has been engineered to extract edges in the damaged regions of the image, with the resulting edge information subsequently being employed as prior knowledge for the purpose of image inpainting, the purpose of this being the alleviation of artefacts and blurriness. The composition of the generator module comprises two networks, a coarse and a refined, which function to inpaint and complete the image where damage has occurred. Both networks adopt a typical encoder–decoder architecture. Among them, the coarse generating network is utilized to inpaint large-scale information of damaged areas. Based on the encoder–decoder structure, the multi-scale parallel dense connection module (MSPDCM) is used in the middle layer to further extract the feature information in the image, thus optimizing the image inpainting effect. The refined generating network focuses on generating detailed textures in images and consists of a contextual attention branch and an MSPDCM branch. The contextual attention mainly considers the importance of long-distance feature information to image inpainting, and the MSPDCM branch mainly uses multi-scale parallel dense connection module to extract feature information at different positions and then fuses the features of the two branches for up-sampling operation to generate an inpainting image. In this paper, the MSPDCM branch is added on the basis of gated convolution network architecture, so the principle and ablation experiment of contextual attention mechanism are not introduced in detail in this paper. The discriminator module is composed of SN-PatchGAN, which is mainly used to judge whether the inpainted image by the generator module is consistent with the images in the training dataset. Consistency denotes authenticity, while inconsistency indicates falseness—this judgment underpins the adversarial game between the discriminator and generator, which serves to optimize the generator's image inpainting performance. Other parts, including the loss functions, have not been changed on the basis of gated convolutional networks.

3.2 Image Inpainting Process

The detailed schematic illustrating the process for inpainting damaged images can be observed in Fig. 3.

Image Preprocessing Stage. In this stage, damaged images are made, data are enhanced and edge information is extracted. The damaged image x_{Di} can be obtained by superimposing the mask x_{mask} on the original complete image x_{oi}, wherein the masks come from the mask datasets disclosed by partial convolution. Because there are no public datasets of damaged photos in real scenes at present, damaged images in simulated real scenes are generated through this process:

$$x_{Di} = x_{oi} \oplus x_{mask} \tag{1}$$

where $\oplus$ represents concatenate operation. The obtained x_{Di} is enhanced by rotation, flipping, mirroring and adding noise, and then edge features are extracted by HED:

$$x_{edge} = H\left(x_{Di}\right) \tag{2}$$

where $H(\cdot)$ in Formula (2) represents the edge extraction operation of HED. Later, the original damaged image x_{Di}, the mask x_{mask} and the extracted edge information x_{edge} are fed into the generating network for image inpainting, as expressed in Formula (3):

$$x_{in} = x_{Di} + x_{mask} + x_{edge} \tag{3}$$

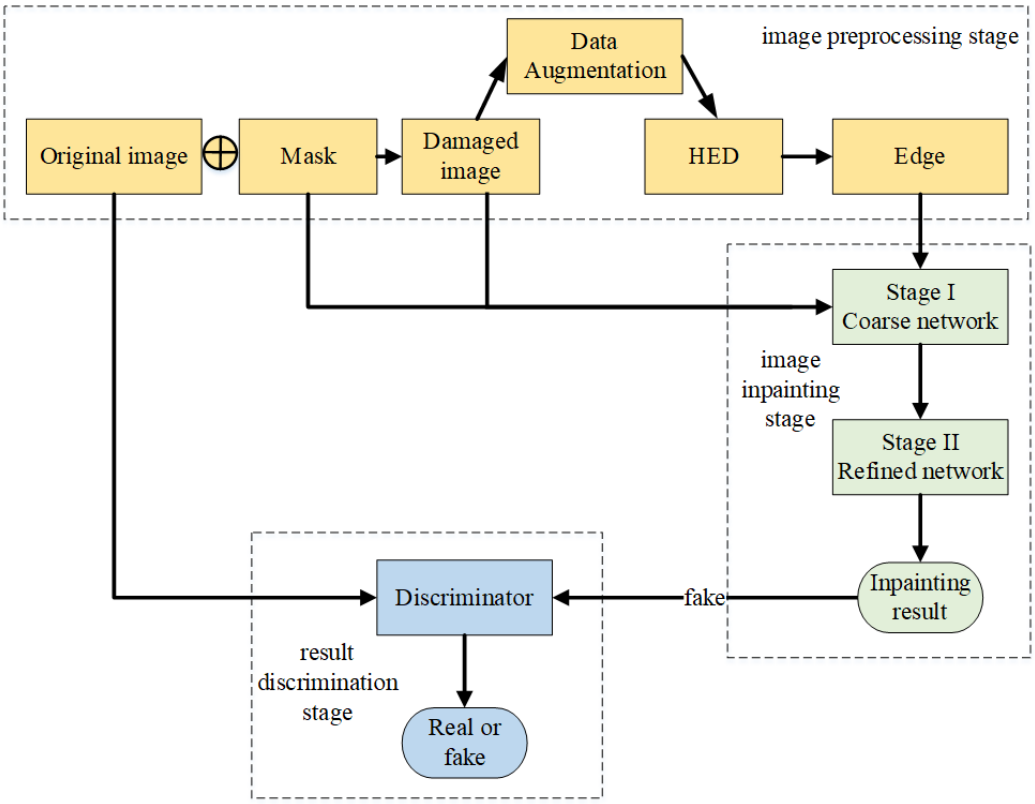

Fig. 3. Schematic diagram of damaged image inpainting process

Image Inpainting Stage. In this stage, the damaged area is inpainted mainly according to the input image information and prior knowledge. In the generating network part, firstly, the input image is roughly generated by $G_{coarse}(\cdot)$; then, the texture and edge details of the damaged area are refinedly inpainted by $G_{refine}(\cdot)$, and finally the inpainted result is obtained. The inpaint process can be expressed by Formula (4):

$$x_{out} = G_{refine}\left[G_{coarse}\left(x_{in}\right)\right] \tag{4}$$

Result Discrimination Stage. This stage mainly judges whether the inpainting image x_{out} is true or not and whether it is close to the original image, that is, whether the inpainting result is complete or not. The inpainted image x_{out} generated by the generator is input to the discriminator, which judges whether it is consistent with the images in the training dataset. The positive output of the discriminator indicates consistency with the original complete training image, while the negative output indicates inconsistency. The generator aims to produce images that the discriminator classifies as positive, and the discriminator focuses on distinguishing the authenticity of the generator's outputs. Through this adversarial game between the two modules, the inpainting performance of the generator is continuously improved.

3.3 Multi-Scale Parallel Dense Connection Module

The core of the MSPDCN is to use the MSPDCM to extract feature information in the generating network. The schematic diagram of the MSPDCM is shown in Fig. 4. This module mainly adopts multi-scale feature extraction strategy to extract the feature information of different scales in damaged images more fully, and at the same time, it adopts dense connection mechanism between layers to strengthen the connection between feature information of different scales, thus optimizing the inpainting effect.

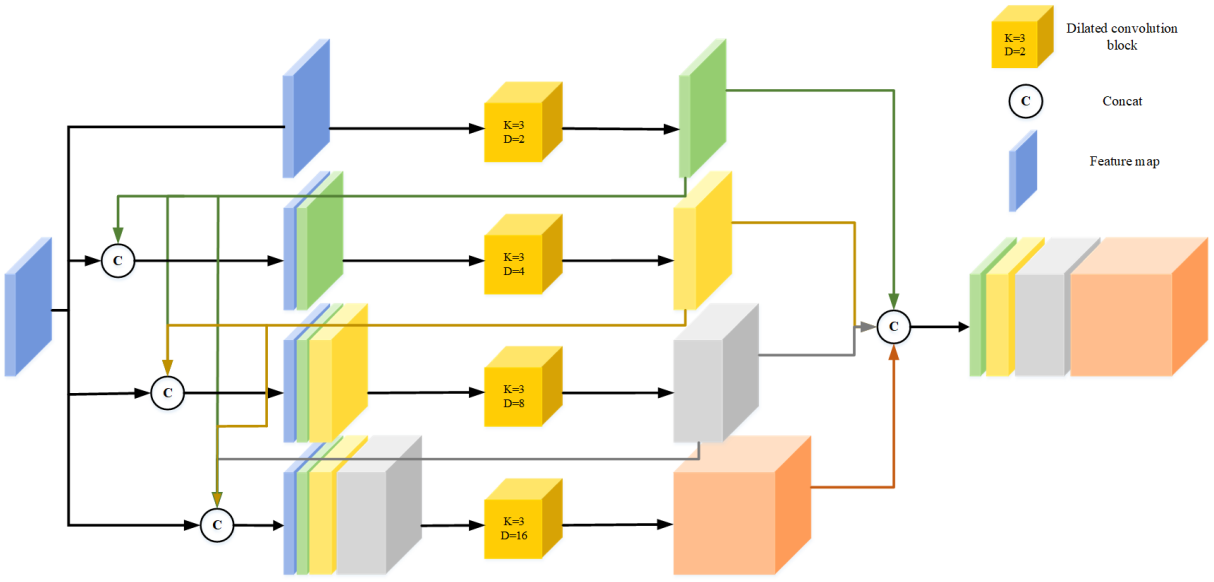

Fig. 4. Schematic diagram of multi-scale parallel dense connection module (MSPDCM)

It is evident that, in accordance with the principles of the gated convolutions network architecture, there is a requirement for the position of hole convolutions to be adjusted from a serial structure to a parallel connection, with a view to achieving the objective of reducing the absolute dependence between the front and back convolution layers. At the same time, the MSPDCM uses four hole convolutions with expansion rates of 2, 4, 8 and 16 (the hole convolution blocks are gated convolution with a convolution kernel size of 3*3, and the expansion rates are 2, 4, 8 and 16) to provide different receptive fields. Hole convolution with a smaller expansion

rate can pay more attention to image details, while larger ones can pay more attention to global information and get better semantic information. However, in the process of feature extraction with mutually independent parallel structures, larger receptive fields will lose some detailed information, while smaller receptive fields will lose global information in the process of feature extraction. In the process of feature extraction, the features x_{MSPDCM_out} obtained by the MSPDCM can be expressed by Formula (5):

$$x_{MSPDCM_out} = x_1 \oplus x_2 \oplus x_3 \oplus x_4 \tag{5}$$

where $\oplus$ represents concatenate feature fusion, and x_1, x_2, x_3 and x_4 represent the feature information obtained after four hole convolution blocks $DC(\cdot)$ with expansion rates of 2, 4, 8 and 16, respectively. x_1, x_2, x_3 and x_4 are obtained by the following formulas.

As shown in Formula (6), the feature x_{MSPDCM_in} in passes through a hole convolution block with an expansion rate of 2 to extract local information:

$$x_1 = DC\left(x_{MSPDCM_in}\right) \tag{6}$$

As shown in Formula (7), the feature x_{MSPDCM_in} in is fused with the feature x_1 extracted by Formula (6), and then the larger local information is extracted by the hole convolution block with an expansion rate of 4:

$$x_2 = DC\left(x_{MSPDCM_in} \oplus x_1\right) \tag{7}$$

As shown in Formula (8), the features x_{MSPDCM_in} is fused with the features obtained by x_1 and x_2, and then the smaller global information is extracted by the hole convolution block with an expansion rate of 8:

$$x_3 = DC\left(x_{MSPDCM_in} \oplus x_1 \oplus x_2\right) \tag{8}$$

As shown in Formula (9), the feature x_{MSPDCM_in} is fused with x_1, x_2, and x_3, which are obtained through feature extraction, and then the global information is extracted through a hole convolution block with an expansion rate of 16:

$$x_4 = DC\left(x_{MSPDCM_in} \oplus x_1 \oplus x_2 \oplus x_3\right) \tag{9}$$

The primary benefit of this approach is that it enables the acquisition of information across diverse ranges during the process of feature extraction. Concurrently, the loss of intricate details and local small-scale features can be mitigated during the extraction of convolution features from larger receptive fields. Consequently, this facilitates a more comprehensive utilisation of the profoundly embedded feature information present within the image. All the layers directly connected to the front are inspired by DenseNet; the principle diagram is shown in Fig. 1, and the principle can be shown by Formula (10):

$$x_l = Hl\left(x_0, x_1, x_2, \cdots, x_{l-1}\right) \tag{10}$$

where $Hl(\cdot)$ represents a series of combined operations: BN [35], ReLU [36] and convolution. The features x_l obtained through dense connection are composed of all the previous layers.

Dense connection is a special connection mechanism. The establishment of a "short circuit connection" between the upper layers and the lower layers is pivotal in ensuring direct connectivity between all layers, thereby facilitating optimal information transmission between network layers. This approach ensures that the input of the lower layers encompasses the outputs of the preceding layers, thereby fostering a comprehensive and integrated network architecture. This strategy has been demonstrated to enhance the transmission of features and to facilitate the optimised utilisation of the feature information contained within the image. Inspired by this,

the MSPDCN directly adopts dense connection between the blocks of hole convolution on the basis of parallel connection and fuses the feature information extracted from the smaller hole convolutions with all subsequent layers to extract the subsequent features, which can make the subsequent hole convolutions pay attention to the global information while not losing the local texture details.

3.4 Theoretical Analysis

To further explain the effectiveness of the proposed MSPDCM and to provide additional mathematical support for the architectural design. Let $X \in R^{C \times H \times W}$ denote the input feature map of the MSPDCM, and let the output of the i-th gated dilated convolution branch be denoted by

$$F_i = D_{ri}\left(X_i\right) \tag{11}$$

where $D_{ri}(\cdot)$ denotes a gated dilated convolution operation with dilation rate ri, and X_i denotes the corresponding branch input. For a convolution kernel of size $k \times k$, the effective receptive field size of a dilated convolution with dilation rate ri can be expressed as

$$R_i = \left(k-1\right)ri + 1 \tag{12}$$

When k = 3, the receptive field sizes corresponding to the four branches are 5, 9, 17, and 33, respectively. Therefore, the four branches naturally cover different spatial neighborhoods, ranging from fine-grained local textures to broader semantic contexts. After feature fusion, the output representation of the module can be written as

$$F_{MSPDCM} = Concat\left(F_1, F_2, F_3, F_4\right) \tag{13}$$

For any target pixel position p, the contextual region captured by the fused representation can be regarded as the union of the neighborhoods corresponding to different receptive field sizes:

$$N\left(p\right) = \bigcup_{i=1}^{4} N_{R_i}\left(p\right) \tag{14}$$

where $N_{R_i}(p)$ denotes the receptive neighborhood centered at p with effective size R_i. Equation (14) shows that the proposed module jointly captures short-range structural details and long-range contextual dependencies, which is crucial for image inpainting because the restoration of missing content requires both local texture continuity and global semantic consistency. Compared with a single-scale feature extractor, the proposed multi-scale design therefore provides a richer and more informative feature representation space.

Another important motivation of the proposed MSPDCM is to reduce the excessive cascading dependency commonly found in serial dilated convolution structures. In a conventional serial design, the output of the lll-th layer is given by

$$F_{serial}^{(l)} = D_{ri}\left(F_{(l-1)}^{serial}\right) \tag{15}$$

which means that the quality of the current layer representation fully depends on the output of the previous layer. If the previous layer output contains representation error, then that error may be propagated to deeper layers. Suppose the (l−1)-th layer output can be decomposed as

$$F_{serial}^{(l-1)} = F^{*(l-1)} + \varepsilon^{(l-1)} \tag{16}$$

where $F^{*(l-1)}$ denotes the ideal feature representation and $\varepsilon^{(l-1)}$ denotes the representation error. Then the lll-th layer output becomes

$$F_{serial}^{(l-1)} = D_{ri}(F^{*(l-1)} + \varepsilon^{(l-1)}) \tag{17}$$

Under the Lipschitz continuity assumption of the convolution operator, there exists a constant $L_l > 0$ such that

$$\left\| F_{serial}^{(l)} - D_{ri}\left(F^{*(l-1)}\right) \right\| \le L_l \left\| \varepsilon^{(l-1)} \right\| \tag{18}$$

Equation (18) indicates that the representation error introduced in one layer can directly affect the output of the subsequent layer, and such dependency may accumulate as the serial depth increases. In contrast, in the proposed MSPDCM, each branch retains access to the original input feature and is not generated solely from the output of the immediately preceding branch. Therefore, the feature extraction of one branch does not completely rely on a single former branch, which effectively reduces cascading dependency and improves the robustness of multi-scale representation learning.

In addition, the proposed dense inter-branch aggregation further strengthens feature reuse. According to the module design,

$$F_1 = D_2(X) \tag{19}$$

$$F_2 = D_4\left(Concat\left(X, F_1\right)\right) \tag{20}$$

$$F_3 = D_8\left(Concat\left(X, F_1, F_2\right)\right) \tag{21}$$

$$F_4 = D_{16}\left(Concat\left(X, F_1, F_2, F_3\right)\right) \tag{22}$$

Thus, the input of the lll-th branch can be expressed in a general form as

$$X_l = Concat\left(X, F_1, F_2, \ldots, F_{l-1}\right) \tag{23}$$

This formulation indicates that each deeper branch receives not only the original input feature but also all features extracted by the shallower branches.

Consequently, local texture cues captured by smaller-receptive-field branches can be explicitly passed to larger-receptive-field branches. If Φ_l denotes the information set available to the l-th branch, then the dense aggregation structure satisfies

$$\Phi_1 \subseteq \Phi_2 \subseteq \Phi_3 \subseteq \Phi_4 \tag{24}$$

Equation (24) shows that the available information increases monotonically with the branch depth. Therefore, later branches can model global semantics while preserving fine details already extracted by earlier branches. From the optimization perspective, dense aggregation also improves gradient propagation. Let L denote the total loss function. Since each branch output contributes to the fused representation, the gradient with respect to an early branch feature F_i can be written as

$$\frac{\partial L}{\partial F_i} = \sum_{j=i+1}^{4} \frac{\partial L}{\partial X_j} \frac{\partial X_j}{\partial F_i} + \frac{\partial L}{\partial F_{MSPDCM}} \frac{\partial F_{MSPDCM}}{\partial F_i} \tag{25}$$

This means that the earlier branches receive gradients through multiple direct and indirect paths, rather than through a single long serial chain, which is beneficial for alleviating gradient attenuation and stabilizing optimization.

Furthermore, the introduction of edge priors in the proposed MSPDCN provides an additional structural constraint for missing-region restoration. Let x_D, x_{mask}, and x_{edge} denote the corrupted image, the mask, and the extracted edge prior, respectively. Then the generator input can be expressed as

$$x_{in} = Concat\left(x_D, x_{mask}, x_{edge}\right) \tag{26}$$

and the inpainting process can be viewed as learning the conditional distribution

$$p\left(x_{out} \middle| x_D, x_{mask}, x_{edge}\right) \tag{27}$$

Compared with the setting without edge priors, the additional structural information narrows the feasible solution space of the restored output. Under the assumption that the extracted edge prior contains valid boundary and contour information correlated with the target image, the uncertainty of the restored result satisfies

$$H\left(x_{out} \middle| x_D, x_{mask}, x_{edge}\right) \le H\left(x_{out} \middle| x_D, x_{mask}\right) \tag{28}$$

where $H(\cdot)$ denotes conditional entropy. Equation (28) shows that the use of edge priors reduces the ambiguity of the reconstruction process, especially near mask boundaries where structural discontinuity is most likely to occur. Therefore, the proposed MSPDCM and the introduced edge-guided mechanism jointly contribute to improved multi-scale feature modeling, enhanced information reuse, more stable optimization, and better structural continuity in image inpainting.

4 Experiments

The proposed MSPDCN is trained and evaluated on the CelebA [37] and Places2 [38] datasets, respectively. Detailed descriptions of the training and testing procedures are provided in Section 4.1. In Section 4.2, the proposed model is compared with state-of-the-art methods, where PSNR, SSIM [39], and LPIPS [40] are adopted as objective metrics to quantitatively evaluate the quality of inpainting results. In Section 4.3, qualitative visual comparison is employed as a subjective evaluation to present the inpainting performance, and the corresponding results are analyzed. The implementation code and trained models are publicly available at: https://github.com/Bberber/MSPDCN.

4.1 Implementation Details

The MSPDCN is trained and tested separately on CelebA and Places2 datasets, trained and tested on the RTX 4090 server. In the training process, the batch size is set to 32. In the testing process, the MSPDCN takes an average of 3.37s per image.

In the preprocessing of CelebA and Places2 datasets, the mask dataset provided by partial convolution is selected to be superimposed with CelebA and Places2 datasets, respectively, to generate damaged results, and the results are enhanced by adding noise, rotating, flipping and mirroring to simulate the real damaged images and improve the generalization ability of the network model. At the same time, the size of the CelebA dataset after data enhancement is changed to 256*256, and the size of the Places2 dataset 256*256 remains unchanged.

Table 1. Objective metrics comparison: MSPDCN versus the state-of-the-arts

Damage degree		0.01-0.1	0.1-0.2	0.2-0.3	0.3-0.4	0.4-0.5	0.5-0.6
CA	PSNR	36.347	27.742	24.508	19.795	18.294	17.327
	SSIM	0.9715	0.8912	0.7998	0.7025	0.6790	0.6079
	LPIPS	0.0245	0.1210	0.1900	0.2314	0.2091	0.2656
E2I	PSNR	24.085	17.917	14.788	12.743	12.303	12.535
	SSIM	0.9506	0.8334	0.7051	0.6163	0.5992	0.5645
	LPIPS	0.1019	0.2697	0.3472	0.4133	0.3587	0.3512
GC	PSNR	36.151	30.136	25.790	22.465	20.590	18.605
	SSIM	0.9696	0.9071	0.8198	0.7500	0.7136	0.6473
	LPIPS	0.0303	0.0837	0.1519	0.1752	**0.1665**	**0.2028**
LFang	PSNR	38.043	31.315	26.962	22.286	20.000	21.616
	SSIM	0.9755	0.9141	0.8353	0.7690	0.7045	**0.7238**
	LPIPS	0.0178	0.0732	0.1334	0.1873	0.2248	0.2222
StrDiffusion	PSNR	29.098	25.634	23.507	23.180	20.458	17.360
	SSIM	0.9162	0.9156	**0.8807**	**0.8311**	0.6917	0.6204
	LPIPS	0.1109	0.1827	0.2547	0.3377	0.4164	0.4484
MSPDCN (Ours)	PSNR	**38.616**	**31.372**	**27.619**	**24.287**	**22.777**	**21.653**
	SSIM	**0.9884**	**0.9200**	0.8565	0.7621	**0.7218**	0.7111
	LPIPS	**0.0093**	**0.0648**	**0.1243**	**0.1742**	0.1890	0.2280

4.2 Objective Evaluation

The images with different damage degrees on CelebA and Places2 datasets are tested by the MSPDCN. The present study sets out to compare the inpainting effects of cutting-edge image inpainting models with those of the MSPDCN. In the aspect of objective comparison, peak signal-to-noise ratio (PSNR, the larger the PSNR, the better the effect of image inpainting; if it is lower than 20, it proves that there is basically no inpainting effect), structural similarity index (SSIM, ranging from 0 to 1, the closer to 1, the more similar to the original image) and learning to perceive the similarity of image blocks (LPIPS, which is more in line with human perception, and the lower the value, the more similar the two images are) are selected for evaluation and comparison. LPIPS learns to generate the inverse mapping from the original image, forces the generator to learn to reconstruct the inverse mapping from the false image, and gives priority to the perceptual similarity between them, so LPIPS is more in line with human perception than traditional methods (such as PSNR and SSIM).

Table 1 shows the comparison results of objective indexes of image inpainting effect of different image inpainting models for different datasets with different damage degrees and summarizes the image inpainting results with the same damage degree. Where "0.1-0.2" means that the damage degree is 10% to 20%. Among them, the number of images of each damage degree consists of 100, half of which comes from the Places2 dataset and half from the Celeba dataset. Each group of photos is randomly selected and randomly superimposed with the mask within the range of damage degree to generate damaged images. It can be seen from Table 1 that the MSPDCN is superior to the other image inpainting models in terms of PSNR, SSIM and LPIPS in images with a smaller damage degree (damage degree < 40%). For images with large damage degree, the LPIPS of GC will be superior to other models, including the MSPDCN, but the MSPDCN is still superior to other models in terms of PSNR and SSIM, and there are still some advantages. According to the qualitative analysis of the inpainting image, the MSPDCN proposed in this paper is still superior to several other representative image inpainting models under the condition of large damage degree (damage degree > 40%), and there is a certain improvement in the inpainting effect. After calculation, compared with the CA model on CelebA and Places2 datasets, the MSPDCN improves the PSNR and SSIM by 13.09% and 5.12%, respectively, and decreases the LPIPS by 21.72%, which shows that the image inpainting effect is improved to some extent.

Table 2 shows the comparison between the MSPDCN and SOTAs on the objective indexes of inpainting the same batch of damaged images with noise in different datasets, in which the noise is added as Gaussian noise. This batch of images consists of five random images, which are from two datasets. The PSNR, SSIM and LPIPS data each consists of two datasets; the former is the calculation result of the inpainting result of damaged image with noise and the original complete image without noise, and the latter is the calculation result of the inpainting result of damaged image without noise and the original complete image without noise. In Table 2, the best results

are shown in bold black, and in the average column, the best results are shown in bold italics. It can be seen from this table that the objective indexes of the MSPDCN and CA are not much different when inpainting damaged images of Places2 without adding noise and are even lower than the inpainting results of CA. After adding noise, the MSPDCN can surpass the CA algorithm and outperform other models in all indexes. It is worth noting that StrDiffusion performs well in inpainting images with added noise, but its effect on noise-free images is not as good as that of MSPDCN. The MSPDCN is superior to other models in inpainting damaged images with noise and without noise. It can be seen that the MSPDCN proposed in this paper has certain advantages in robustness and anti-interference ability under the condition of good inpainting effect.

Table 2. Comparison of inpainting effects of different models for noisy images

		Places2			CelebA		
	Image NO.	1	2	3	1	2	**Average**
CA	PSNR	15.445/**41.483**	15.285/**37.194**	***17.013/31.480***	13.324/25.350	14.729/28.671	15.159/32.676
	SSIM	*0.1992/0.9819*	0.1287/**0.9797**	0.2638/0.9150	0.1492/0.9784	0.1370/0.8871	0.1756/0.9484
	LPIP	0.5501/***0.0216***	0.7075/0.0350	0.5619/***0.0697***	0.7410/0.0340	0.7248/0.1221	0.6571/0.0565
	Time/s	3.23	3.57	3.78	3.65	3.4	***3.53***
E2I	PSNR	15.411/31.564	15.014/25.523	15.409/19.430	13.435/32.281	14.055/19.736	14.665/25.707
	SSIM	0.1977/0.9687	0.1259/0.9685	0.2409/0.8469	0.1520/0.9851	0.1428/0.8449	0.1719/0.9228
	LPIP	0.5546/0.0622	0.7011/0.0894	**0.5527**/0.2331	0.7411/0.0220	0.6914/0.2179	0.6482/0.1249
	Time/s	-	-	-	-	-	13.65
GC	PSNR	15.443/38.453	15.267/34.134	16.941/29.343	13.441/**37.564**	14.980/28.390	15.214/33.577
	SSIM	0.1992/0.9786	0.1284/0.9766	0.2625/0.9042	0.1495/**0.9887**	0.1438/0.9033	0.1767/**0.9503**
	LPIP	0.5506/0.0362	0.7044/0.0482	0.5550/0.0920	0.7402/**0.0090**	0.7162/**0.0842**	0.6533/**0.0539**
	Time/s	3.64	3.85	4.01	3.99	3.95	3.89
LFang	PSNR	15.441/38.711	15.293/34.477	16.267/29.318	13.444/35.826	15.112/**30.065**	15.111/**33.679**
	SSIM	0.1985/0.9752	0.1291/0.9586	**0.2668**/0.8947	0.1521/0.9886	**0.2536/0.9060**	**0.2000**/0.9446
	LPIP	**0.5497**/0.0385	0.7065/**0.0328**	0.5555/0.1072	0.7406/0.0142	0.5686/0.0878	0.6242/0.0561
	Time/s	-	-	-	-	-	3.64
StrDiffusion	PSNR	**16.180**/30.453	**16.013**/29.536	14.467/28.440	**14.784**/30.973	**15.520**/28.367	**15.393**/29.554
	SSIM	0.1831/0.9442	**0.1794**/0.9445	0.1863/**0.9367**	0.1836/0.9371	0.1854/0.8983	0.1836/0.9321
	LPIP	0.5604/0.0956	**0.5404**/0.1076	0.6920/0.1284	**0.4616**/0.1520	**0.4878**/0.2868	**0.5484**/0.1541
	Time/s	26.70	26.53	27.16	29.73	26.57	27.34
MSPDCN (Ours)	PSNR	**16.461/40.784**	**15.312/37.099**	**17.078/31.490**	**13.467/43.015**	**15.220/30.840**	**15.508/36.646**
	SSIM	**0.2002/0.9821**	**0.2534/0.9796**	**0.2784/0.9164**	**0.1568/0.9891**	**0.2546/0.9539**	**0.2287/0.9642**
	LPIP	**0.5496/0.0169**	**0.6236/0.0270**	**0.5142/0.0500**	**0.7306/0.0086**	**0.5632/0.0230**	**0.5962/0.0251**
	Time/s	3.24	3.39	3.61	3.22	3.39	**3.37**

The inpainting of a damaged image during the testing is all conducted on the same device, and the time taken to inpaint an image using different models is shown in Table 2. In the E2I inpainting process, batch inpainting is used, and other methods are used to inpaint images separately. E2I carries out edge extraction, rough reconstruction and thinning together, so it takes a long time. StrDiffusions used parse structure to guide the texture denoising network, adding and inpainting masks for each image one by one. Therefore, we only calculated the time for the entire completion process. Compared with other SOTA models, the MSPDCN takes less time in image inpainting than other centralized image inpainting models, and it also has some advantages in saving time.

4.3 Subjective Evaluation

Visual effect serves as the subjective criterion for evaluating image inpainting quality, and assessing inpainting quality through human subjective perception is more effective. An example of inpainting damaged images of different datasets by the MSPDCN proposed in this paper is shown in Fig. 5.

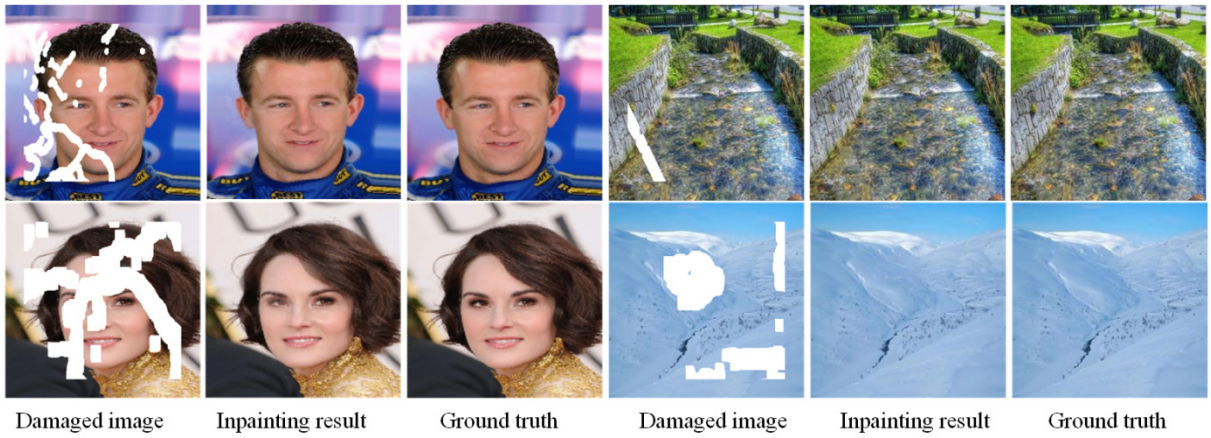

Fig. 5. Example of inpainting results of MSPDCN

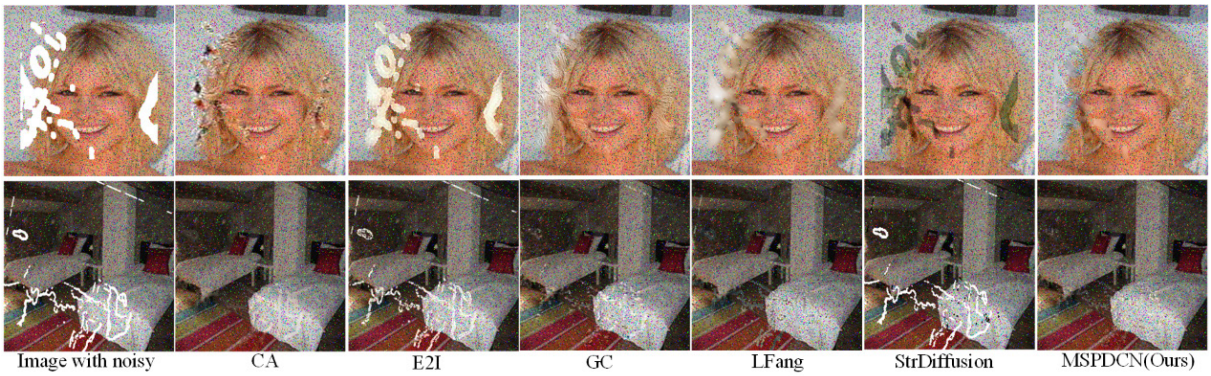

Fig. 6. Comparison of inpainting effects of noisy damaged images

As shown in Fig. 5, the MSPDCN can inpaint damaged images in different datasets and achieve certain results in texture details. There are four damaged images from two datasets. The inpainting images obtained by the MSPDCN are close to the original images, and the details are basically in line with people's subjective feelings.

Fig. 6 shows the comparison of the inpainting effects of SOTA models for the same noisy damaged image (the generation process of noisy image: the original image is super- imposed with a mask to generate a damaged image, and then Gaussian noise is added to obtain a noisy damaged image). From the inpainting effect, it can be seen that the MSPDCN proposed in this paper can still achieve better inpainting effect than other models. Combined with the comparison results of objective indicators of noisy image inpainting results in Table 2 of 4.2, the MSPDCN can improve the inpainting effect and the robustness of the network, which is beneficial to the training and testing of the model.

Fig. 7 shows the inpainting results of the MSPDCN and SOTA models for different datasets. The damaged areas are gradually enlarged from top to bottom, and from left to right are the damaged images, CA inpainting results, E2I inpainting results, GC inpainting results, LFang inpainting results, StrDiffusion inpainting results, MSPDCN inpainting results and original images. Horizontally, for the damage degree of small area, the MSPDCN proposed in this paper can better inpaint the damaged area and obtain better inpainting effect. In terms of vertical alignment, the extent of the damaged area is positively correlated with the inpainting effect of all models, though the MSPDCN demonstrates superior objective evaluation indexes and visual efficacy in comparison to alternative models. However, it can be seen from Fig. 7 that when the damaged area is large, the visual effect of the image inpainted by the MSPDCN is greatly reduced, and there is still room for improvement on how to further optimize the inpainting effect of large-area damaged areas.

Compared with several state-of-the-art image inpainting methods, the proposed MSPDCN achieves superior performance in terms of PSNR, SSIM, LPIPS, and computational efficiency. From a subjective perspective, the MSPDCN incorporates edge information as prior knowledge to guide the inpainting process, which effectively alleviates boundary blurring in damaged regions and improves the semantic consistency and completeness of the restored images. These results indicate that the inpainting performance can be further enhanced by integrating prior structural information, employing a multi-scale parallel connection design, and reinforcing the interaction between shallow and deep features through dense connectivity. In addition, the proposed architecture contributes to improved robustness and stronger resistance to noise and interference. Nevertheless, the MSPDCN still faces certain limitations when dealing with images containing large missing regions, which remains a challenging issue for future research.

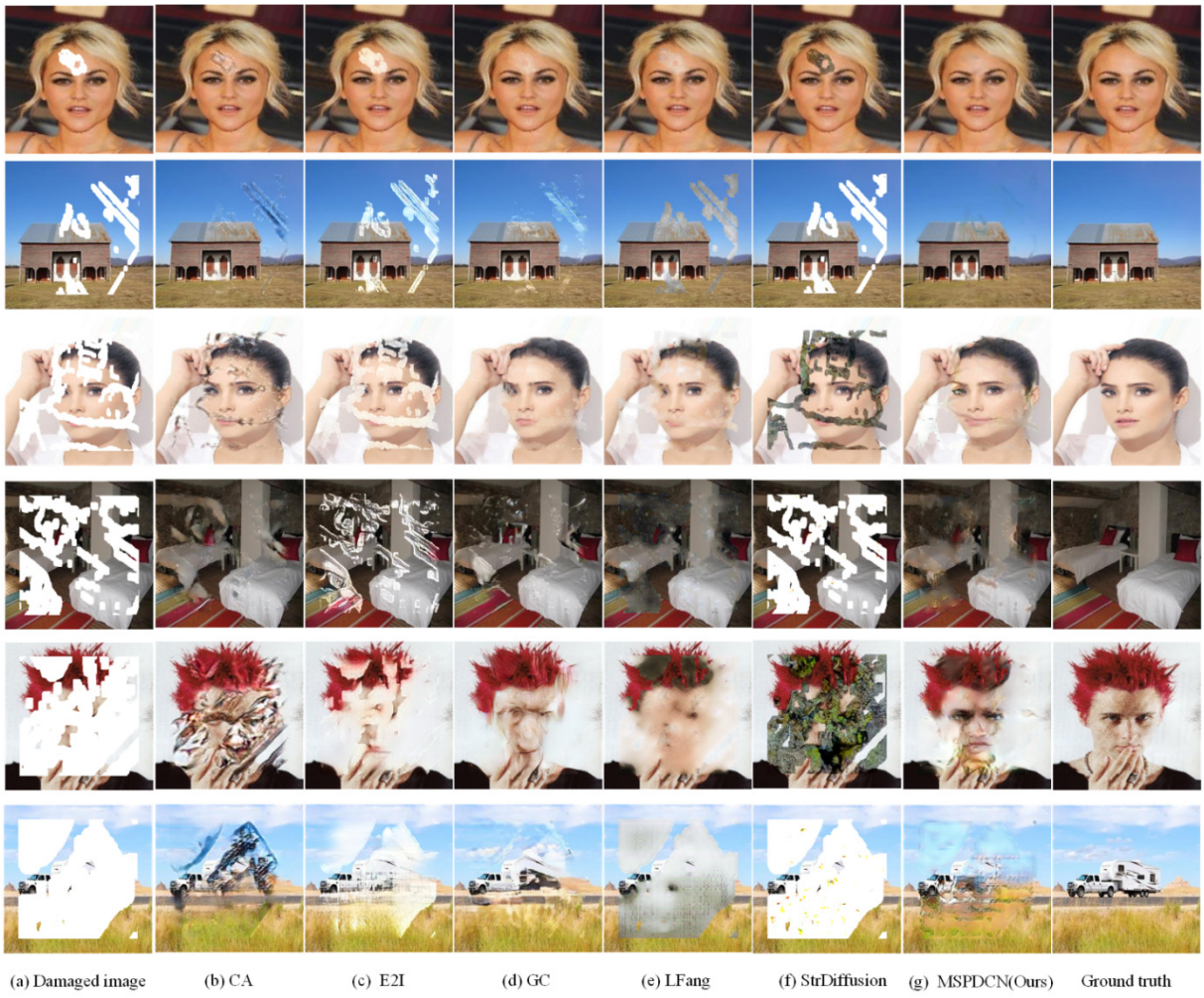

Fig. 7. Examples of inpainting effects of different inpainting models

5 Conclusion

In this paper, a novel image inpainting approach based on a multi-scale parallel dense connection network is presented. By employing a multi-scale parallel feature extraction framework, the model is able to capture feature representations at different spatial locations within the image. Meanwhile, the strong dependency between early and later layers is alleviated, leading to improved efficiency in feature extraction. Furthermore, the dense connections established among multi-scale parallel modules enhance the interaction between shallow and deep

features and promote more effective utilization of feature information, thereby improving the overall inpainting performance. Experimental comparisons with state-of-the-art methods demonstrate that the proposed approach achieves gains of 13.09% and 5.12% in PSNR and SSIM, respectively, while reducing LPIPS by 21.72% compared with the CA model. In addition, both visual quality and robustness to interference are improved. Overall, the MSPDCN produces more accurate and visually coherent inpainting results than existing approaches. The proposed model also shows potential in practical applications, particularly in the digital restoration of cultural relics and calligraphy. By collecting relevant image data, the MSPDCN can generate restoration results for damaged artifacts, which can serve as a useful reference for professional conservators in real-world restoration tasks.

Nevertheless, the image inpainting model proposed in this paper is not without its limitations, and significant challenges remain in the realm of inpainting tasks for images characterised by larger damaged areas. In addition, because the use of dense connections strengthens the feature relationship between convolution layers, the parameters of the model also increase, so the optimization of the parameters of the network model needs to be reconsidered. Faced with complex and diverse images, the issue of how to leverage the method proposed in this paper to generate more diverse and selectable inpainting results is also a promising research direction for the future.

References

[1] A.A. Efros, T.K. Leung, Texture synthesis by non-parametric sampling, in: Proc. 1999 IEEE International Conference on Computer Vision, 1999.
https://doi.org/10.1109/ICCV.1999.790383

[2] V. Kwatra, I. Essa, A. Bobick, N. Kwatra, Texture optimization for example-based synthesis, in: Proc. 2005 ACM SIGGRAPH, 2005.
https://doi.org/10.1145/1186822.1073263

[3] D. Ding, S. Ram, J.J. Rodríguez, Image inpainting using nonlocal texture matching and nonlinear filtering, IEEE Transactions on Image Processing 28(4)(2019) 1705–1719.
https://doi.org/10.1109/TIP.2018.2880681

[4] U. Erkan, S. Enginoğlu, D.N.H. Thanh, An iterative image inpainting method based on similarity of pixels values, in: Proc. 2019 IEEE 6th International Conference on Electrical and Electronics Engineering, 2019.
https://doi.org/10.1109/ICEEE2019.2019.00028

[5] H. Wang, L. Jiang, R. Liang, X.X. Li, Exemplar-based image inpainting using structure consistent patch matching, Neurocomputing 269(2017) 90–96.
https://doi.org/10.1016/j.neucom.2016.08.149

[6] Y. Song, C. Yang, Z. Lin, X. Liu, Q. Huang, H. Li, C.C.J. Kuo, Contextual-based image inpainting: Infer, match, and translate, in: Proc. 2018 European Conference on Computer Vision, 2018.
https://doi.org/10.1007/978-3-030-01216-8_1

[7] I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Proc. 2014 Advances in Neural Information Processing Systems, 2014.
https://papers.nips.cc/paper_files/paper/2014/hash/f033ed80deb0234979a61f95710dbe25-Abstract.html

[8] J. Ho, A. Jain, P. Abbeel, Denoising diffusion probabilistic models, in: Proc. 2020 Advances in Neural Information Processing Systems, 2020.
https://proceedings.neurips.cc/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html

[9] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, A.A. Efros, Context encoders: Feature learning by inpainting, in: Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016.
https://doi.org/10.1109/CVPR.2016.278

[10] Y. Wang, X. Tao, X. Qi, X. Shen, J. Jia, Image inpainting via generative multi-column convolutional neural networks, in: Proc. 2018 Advances in Neural Information Processing Systems, 2018.
https://proceedings.neurips.cc/paper_files/paper/2018/file/6f3ef77ac0e3619e98159e9b6febf557-Paper.pdf

[11] G. Liu, F.A. Reda, K.J. Shih, T.-C. Wang, A. Tao, B. Catanzaro, Image inpainting for irregular holes using partial convolutions, in: Proc. 2018 European Conference on Computer Vision, 2018.
https://doi.org/10.1007/978-3-030-01252-6_6

[12] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, T.S. Huang, Free-form image inpainting with gated convolution, in: Proc. 2019 IEEE/CVF International Conference on Computer Vision, 2019.
https://doi.org/10.1109/ICCV.2019.00457

[13] C. Corneanu, R. Gadde, A.M. Martinez, LatentPaint: Image inpainting in latent space with diffusion models, in: Proc. 2024 IEEE/CVF Winter Conference on Applications of Computer Vision, 2024.
https://doi.org/10.1109/WACV57701.2024.00428

[14] H. Liu, Y. Wang, B. Qian, M. Wang, Y. Rui, Structure matters: Tackling the semantic discrepancy in diffusion models for image inpainting, in: Proc. 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024. https://doi.org/10.1109/CVPR52733.2024.00768
[15] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proc. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016. https://doi.org/10.1109/CVPR.2016.90
[16] G. Huang, Z. Liu, L. van der Maaten, K.Q. Weinberger, Densely connected convolutional networks, in: Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017. https://doi.org/10.1109/CVPR.2017.243
[17] S. Iizuka, E. Simo-Serra, H. Ishikawa, Globally and locally consistent image completion, ACM Transactions on Graphics 36(4)(2017) 1–14. https://doi.org/10.1145/3072959.3073659
[18] K. Nazeri, E. Ng, T. Joseph, F.Z. Qureshi, M. Ebrahimi, EdgeConnect: Generative image inpainting with adversarial edge learning. <https://arxiv.org/abs/1901.00212>, 2019.
[19] Y. Chen, R. Xia, K. Yang, K. Zou, MFMAM: Image inpainting via multi-scale feature module with attention module, Computer Vision and Image Understanding 238(2024) 103883. https://doi.org/10.1016/j.cviu.2023.103883
[20] S. Xu, D. Liu, Z. Xiong, E2I: Generative inpainting from edge to image, IEEE Transactions on Circuits and Systems for Video Technology 31(4)(2021) 1308–1322. https://doi.org/10.1109/TCSVT.2020.3001267
[21] F. Long, Y.-G. Wang, Image inpainting with information loss reduction and texture-structure feature fusion, in: Proc. 2023 IEEE International Conference on Image Processing, 2023. https://doi.org/10.1109/ICIP49359.2023.10222164
[22] C. Shi, W. Liu, J. Meng, X. Jia, J. Liu, Self-prior guided generative adversarial network for image inpainting, The Visual Computer 41(4)(2025) 2939–2951. https://doi.org/10.1007/s00371-024-03578-x
[23] J. Yu, Z. Lin, J. Yang, X. Shen, X. Lu, T.S. Huang, Generative image inpainting with contextual attention, in: Proc. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018. https://doi.org/10.1109/CVPR.2018.00577
[24] J. Canny, A computational approach to edge detection, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-8(6)(1986) 679–698. https://doi.org/10.1109/TPAMI.1986.4767851
[25] J. He, S. Zhang, M. Yang, Y. Shan, T. Huang, Bi-directional cascade network for perceptual edge detection, in: Proc. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019. https://doi.org/10.1109/CVPR.2019.00395
[26] Y. Liu, M.-M. Cheng, X. Hu, K. Wang, X. Bai, Richer convolutional features for edge detection, in: Proc. 2017 IEEE Conference on Computer Vision and Pattern Recognition, 2017. https://doi.org/10.1109/CVPR.2017.622
[27] S. Xie, Z. Tu, Holistically-nested edge detection, in: Proc. 2015 IEEE International Conference on Computer Vision, 2015. https://doi.org/10.1109/ICCV.2015.164
[28] X.S. Poma, E. Riba, A. Sappa, Dense extreme inception network: Towards a robust CNN model for edge detection, in: Proc. 2020 IEEE/CVF Winter Conference on Applications of Computer Vision, 2020. https://doi.org/10.1109/WACV45572.2020.9093290
[29] M. Arimoto, J. Hara, H. Watanabe, An image inpainting method considering edge connectivity of defects, in: Proc. 2021 IEEE 10th Global Conference on Consumer Electronics, 2021. https://doi.org/10.1109/GCCE53005.2021.9621917
[30] Y. Chen, R. Xia, K. Zou, K. Yang, FFTI: Image inpainting algorithm via features fusion and two-steps inpainting, Journal of Visual Communication and Image Representation 91(2023) 103776. https://doi.org/10.1016/j.jvcir.2023.103776
[31] H. Li, Y. Song, H. Li, Z. Wang, Semantic prior-driven fused contextual transformation network for image inpainting, Journal of Visual Communication and Image Representation 91(2023) 103777. https://doi.org/10.1016/j.jvcir.2023.103777
[32] X. Ma, Y. Li, T. Yao, Image completion based on edge prediction and improved generator, Tehnički vjesnik 28(5)(2021) 1590–1596. https://doi.org/10.17559/TV-20210616090311
[33] Z.X. Jin, H. Qin, Generative adversarial network based on multi-feature fusion strategy for motion image deblurring, Journal of Computers 33(1)(2022) 31–41. https://doi.org/10.53106/199115992022023301004
[34] J. Jiang, X. Dong, T. Li, F. Zhang, H. Qian, G. Chen, Parallel adaptive guidance network for image inpainting, Applied Intelligence 53(1)(2023) 1162–1179. https://doi.org/10.1007/s10489-022-03387-6

[35] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: Proc. 2015 International Conference on Machine Learning, 2015.
https://doi.org/10.48550/arXiv.1502.03167
[36] V. Nair, G.E. Hinton, Rectified linear units improve restricted Boltzmann machines, in: Proc. 2010 International Conference on Machine Learning, 2010.
https://www.cs.utoronto.ca/~hinton/absps/reluICML.pdf
[37] Y. Zhang, Z. Yin, Y. Li, G. Yin, J. Yan, J. Shao, Z. Liu, CelebA-Spoof: Large-scale face anti-spoofing dataset with rich annotations, in: Proc. 2020 European Conference on Computer Vision, 2020.
https://doi.org/10.1007/978-3-030-58610-2_5
[38] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, A. Torralba, Places: A 10 million image database for scene recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 40(6)(2018) 1452–1464.
https://doi.org/10.1109/TPAMI.2017.2723009
[39] Z. Wang, A.C. Bovik, H.R. Sheikh, E.P. Simoncelli, Image quality assessment: From error visibility to structural similarity, IEEE Transactions on Image Processing 13(4)(2004) 600–612.
https://doi.org/10.1109/TIP.2003.819861
[40] R. Zhang, P. Isola, A.A. Efros, E. Shechtman, O. Wang, The unreasonable effectiveness of deep features as a perceptual metric, in: Proc. 2018 IEEE Conference on Computer Vision and Pattern Recognition, 2018.
https://doi.org/10.1109/CVPR.2018.00068