Displays 69 (2021) 102028

Contents lists available at ScienceDirect

Displays

journal homepage: www.elsevier.com/locate/displa

Image inpainting based on deep learning: A review☆

Zhen Qin [a], Qingliang Zeng [b], Yixin Zong [c], Fan Xu [d,*]

[a] School of Design, Hunan University, Changsha, China
[b] Cognitive Computing Technology Joint Laboratory, Wave Group, Beijing, China
[c] Bureau of Frontier Sciences and Education, Chinese Academy of Sciences, Beijing, China
[d] Beijing Institute of Control Engineering, Beiing, China

ARTICLE INFO

Keywords:
Computer vision
Image inpainting
Variational autoencoder (VAE)
Generative adversarial networks (GAN)

ABSTRACT

Image inpainting aims to restore the pixel features of damaged parts in incomplete image and plays a key role in many computer vision tasks. Image inpainting technology based on deep learning is a major current research hotspot. To deeply understand related methods and technologies, this article combs and summarizes the latest research status in this field. Firstly, we summarize inpainting methods of different types of neural network structure based on deep learning, then analyze and study important technical improvement mechanisms. In addition, various algorithms are comprehensively reviewed from the aspects of model network structure and restoration methods. And we select some representative image inpainting methods for comparison and analysis. Finally, the current problems of image inpainting are summarized, and the future development trend and research direction are prospected.

1. Introduction

Image inpainting is a technology that aims to restore the damaged part of pixel features in the incomplete image, and then reconstruct and generate high-quality and deep semantic approximation to the original image. In recent years, the implementation of artificial intelligence scientific research and deep learning related technologies has achieved vigorous development along with the substantial increase in computer computing power, which has brought important promotion and improvement to science technology and the quality of human life. Image inpainting technology based on deep learning plays an important role in many computer vision applications [1] (such as target removal in image editing technology, old photo restoration, defective cultural relics and font restoration, facial restoration, etc.) and has become a major research hotspot in computer vision.

In traditional image inpainting technology, the related methods are mostly machine learning algorithms based on statistical probability. Marcelo Bertalmio et al. [2] proposed a Markov Random Field (MRF) image inpainting algorithm on the basis of structure migration mapping statistics and multi-directional features for large-scale damaged image restoration. The inpainting algorithm is mainly used for target removal, which can better maintain the continuity of repaired image structure and the consistency between adjacent pixels. Shen et al. [3] proposed an improved sparse representation inpainting algorithm in the light of similar matching blocks, which achieved good restoration effects in the inpainting of color damaged images with multiple damaged shapes in a small area. Tsai et al. [4] proposed a matrix completion method with automatic rank estimation based on low-rank decomposition is used to extract restored high-quality images from images with different degrees of low sampling rate. Considering the above, Bertalmio et al. [5] introduced conjugate gradient method based on riemannian manifold to optimize matrix completion and combined convolution neural network to preprocess sample images. The method of block processing is adopted to further save operation space and improve the quality of restored images. These methods have made improvements on traditional machine learning algorithms to promote image inpainting effects in different ways. However, compared with the deep learning image inpainting technology, the restored images generated by the traditional methods when processing large image data in damaged areas often lack semantic consistency and texture structure coherence.

Around 2014, with the rise of deep learning, image inpainting technology has been deeply applied in the field of computer vision. Many researchers have continuously carried out in-depth research on the problem of high-quality image inpainting at the level of generating

☆ This paper was recommended for publication by Prof G Guangtao Zhai.
* Corresponding author.
E-mail address: fan_walton@163.com (F. Xu).

https://doi.org/10.1016/j.displa.2021.102028
Received 8 March 2021; Received in revised form 15 April 2021; Accepted 19 May 2021
Available online 26 May 2021

semantic understanding [6–8], and then a large number of classical image inpainting methods based on deep learning have emerged [9–18], Some scholars have also summarized the work in this field. Recent work [1,19] summarized the image inpainting technology based on deep learning. Omar Elharrouss et al. [1] divided the image inpainting model methods proposed in some classic papers into three categories from a global perspective, namely, sequence-based methods, CNN-based methods and GAN-based methods. Qiang et al. [19] summarized the main image inpainting methods based on deep learning in recent years and classified the existing methods into three network structure types of image inpainting methods based on convolutional autoencoder network, generative adversarial network and recurrent neural network according to the inpainting network structure.

In the past few years, deep learning has made great breakthroughs in the field of image inpainting. A hybrid network model based on the combination of autoencoder and Generative Adversarial Network (GAN) [20–23], an improved autoencoder based on attention mechanism [24–29], and improved shared codec network layer based on coarse-to-fine network [30–35] have emerged, which gradually repair damaged images at the semantic level. Based on the above work, this paper makes a more comprehensive and detailed summary of image inpainting related network models based on deep learning in recent years, aiming to provide a more comprehensive and in-depth learning perspective for subsequent research in related fields.

2. Related works

2.1. Image inpainting tasks

Current image inpainting research mainly includes tasks such as repairing rectangular block mask, irregular mask, target removal, denoising, remove watermark, remove text, remove scratches, and coloring of old photos [20,26,35–38,79]. The example effects of above the 8 inpainting tasks are shown in Fig. 1:

2.2. Traditional image inpainting

Traditional image inpainting, mainly divided into diffusion-based

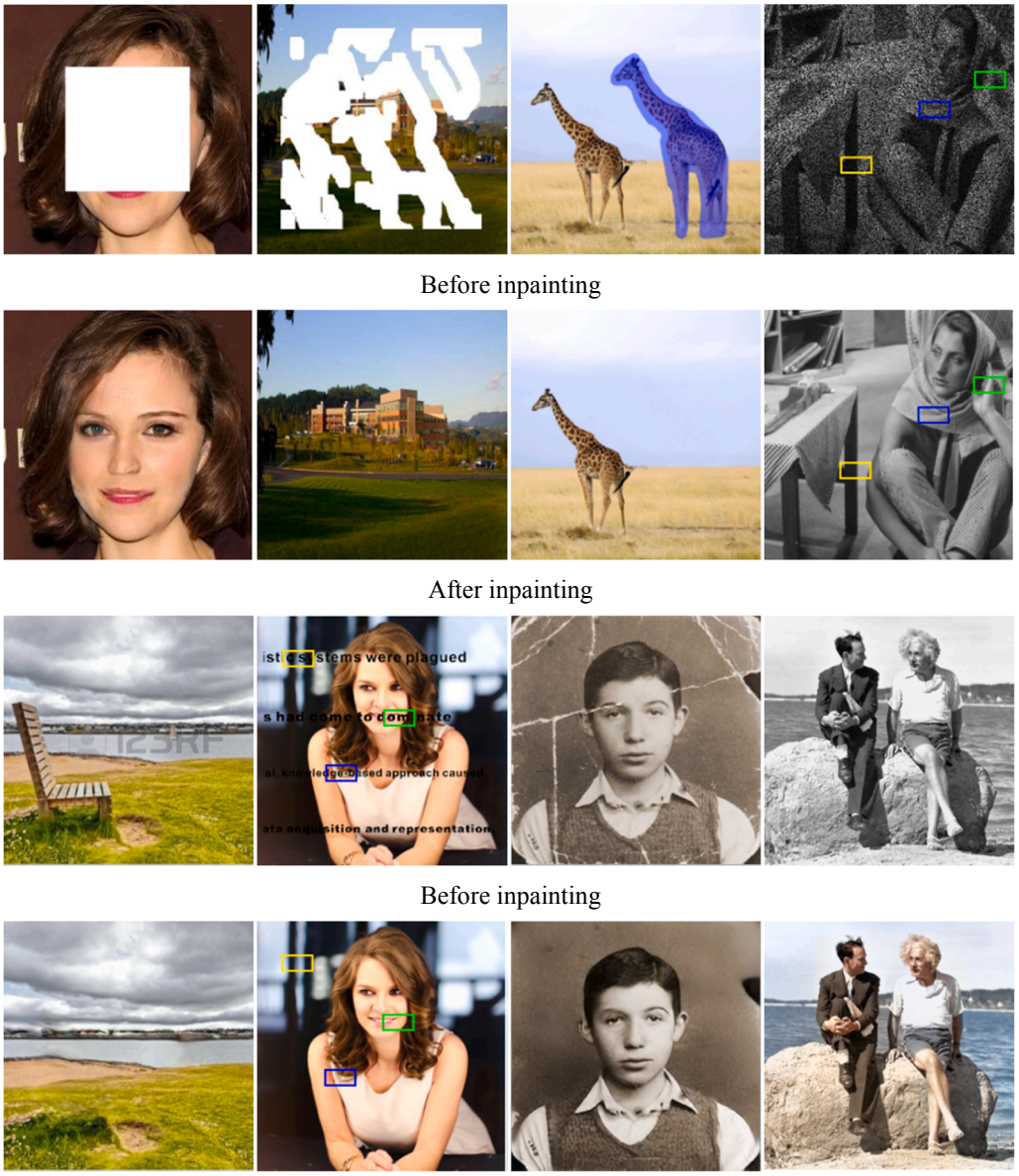


Fig. 1. Eight common image inpainting applications.

methods [2,39–41] and patch-based methods [42–46].

Diffusion-based image inpainting mainly spreads the pixel information around the damaged hole in the image gradually and synthesizes new textures to fill the hole. Reconstruction is usually restricted by the information around the hole, it is difficult to reasonably learn from distant information, and lack high-level semantic understanding of the image, making it difficult to restore meaningful texture structure in missing area. And as the diffusion distance of pixel information around the hole increases, the larger the hole is, the less effective pixel information will be obtained in the center. So the traditional diffusion-based method is more suitable for structure texture background inpainting and removal of small objects in the image, but the effectiveness of large-hole restoration for natural scene objects with complex textures in real life is limited.

The patch-based image inpainting assumes that damaged area and visible area of the image have similar content. It searches for the best matching similar patch in visible area of the image, and then copies the information to fill the missing area at pixel level. The traditional method [47] usually requires an enormous amount of computing power to calculate the similarity score between patches. PatchMatch [45] reduces high memory and computational cost in the search process by using a fast nearest neighbor field algorithm. And it shows a certain practical value in image editing applications. However, more often, the content of the image damage area may be a completely independent small individual or unstructured partial damage. At this time, the traditional patch-based method may become difficult to handle.

In special cases, it may not be possible to find a patch similar to the missing area in the image. Therefore, some researchers have proposed image inpainting between images, which mainly refers to searching for an image with similar semantics to the target damaged image in the existing image library, and then select appropriate patch information for porting and borrowing. For example, James Hays et al. [48] used a million-level data library to retrieve the most similar image of the target inpainting image, and then extracted corresponding area information to complete the damaged image. This method can better repair damaged images when there is a large amount of image data in a specific field, but it also often means that a large amount of field data collection and the best match search need to be carried out [49–51]. So in the real scene, such methods have fewer applicable scenarios, and application range is relatively limited.

2.3. Image inpainting based on generative network

With the emergence of two generative models such as VAE [52,53] and GAN [54,55], the image inpainting method based on generative model can fully learn the high and low frequency feature information of the damaged image visible area and learn the consistency of image structure and texture at the high-level semantic level by adding different constraints to generate novel and reasonable features to complement the damaged area. Therefore, in recent years, various deep learning image inpainting models based on generative network have been the hot direction for many researchers to make further improvement.

At first, works such as Satoshi et al. [20] and Liu et al. [21] were based on encoder-decoder architecture to repair small rectangular blocks and narrow defects. Many methods [26,29,31,56] were applied to handle random shaped irregular holes. Zheng et al. [57] and Zhao et al. [58] proposed multiple-solution inpainting method produces diverse image inpainting results. Yang et al. [12], Zeng et al. [34] and Yi et al. [35] were dedicated to high-resolution image inpainting.

In recent years, various deep learning image inpainting methods based on generative networks have been proposed to solve above needs, this paper further summarizes them into three categories: single-stage inpainting, progressive image inpainting, and inpainting based on prior knowledge. In the next chapter, the representative algorithms are analyzed and summarized.

3. Image inpainting methods based on deep learning

3.1. Single-stage inpainting

The approaches related to single-stage inpainting can classified into two categories: single result inpainting and pluralistic inpainting approaches.

3.1.1. Single result inpainting

(1) Context-encode

Pathak et al. [20] proposed an image inpainting network named context-encoder, which applies unsupervised feature learning driven by context-based pixel prediction to large-hole image inpainting. The model architecture is shown in Fig. 2. Overall architecture is a simple encoder-decoder. The encoder extracts feature representation of the input image, and decoder enlarges the compressed feature map step by step to restore to the size of the original picture. Since convolutional layers cannot directly connect all locations within a specific feature map, encoder composed of convolutional layers is no way for information to directly propagate from one corner of the feature map to another. A fully-connected layer with groups based on stride 1 convolution to propagate information cross channels method is proposed as the intermediate connection between the encoder and the decoder to propagate information within activities of each feature map.

Context-encoder adopts reconstruction loss (L2) and adversarial loss to handle both continuity within context and multiple modes in the output. The reconstruction loss is responsible for capturing overall structure of the repaired area and the consistency with the surrounding visible area, the adversarial loss makes the prediction of the repaired area looks real. The best possible inpainting results can be generated by maintaining the balance of them.

Context-encoder can understand semantics of image to a certain extent and predict pixels according to information around hole to generate new content. At that time, it was a very cutting-edge image inpainting technology. It lays a foundation for follow-up research work.

(2) Globally and Locally Consistent

Inspired by context-encoder [20], researchers have proposed a globally and locally consistent image completion method [21] to solve the defects of context-encoder. For example, only fixed low-resolution images can be processed, the mask area must be located in the center of the image, and the complete area cannot maintain local consistency with the surrounding area. The network uses two auxiliary context discriminator for training, where the global discriminator network takes entire image as input and the local discriminator network takes only a small region around the completed area as input ensure the global and local semantics of the restored image respectively. Dilated convolution [59] is used in the middle four layers of completion network to increase the receptive field of the extracted features. Images of any size can be completed, and new textures and objects can be generated according to local and global structural semantic information.

(3) Partial Convolutions

When using standard convolution network to repair damaged images, the average value of effective pixels and missing parts is usually used as filling, which easily makes the big hole inpainting area lack texture information, produces artifacts such as color difference and blur, seriously affects visual sensory. Liu et al. [26] proposed partial convolutions to solve the above problems. Under mask-update process, convolutional results depend on non-hole regions at every layer and binary mask corresponding to damaged area, through continuous updating of sufficient layers, finally only retain features which obtained by pixel

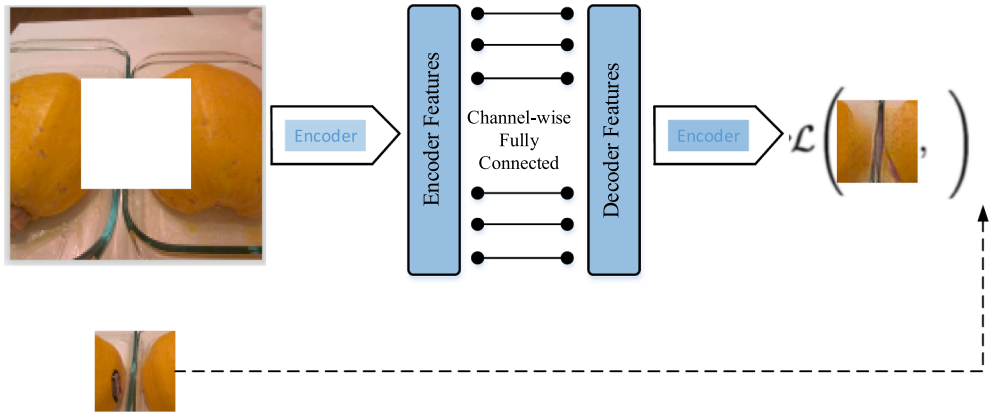


Fig. 2. Framework of context-encoder.

convolution of visible area. Therefore, a good missing part prediction independent of initialization value of the missing part and without any additional post-processing is realized, irregular missing part can be effectively repaired, so that image inpainting technology has a larger imagination space in production applications. The limitation is that when dealing with some sparsely structured images, repair effect is still limited like the previous method.

(4) Pyramid-context Encoder

Aiming at image level inpainting [44,45,60] generally lack understanding of the high-level semantics of the image by searching for similar patches in visible area of the image and copying them to damaged area for texture synthesis to fill damaged area, cann't produce semantically reasonable results, and although generative models [20,30] can enhance semantic consistency of repair region, stacked constructions and poolings to a certain extent cause image resolution details to be over-smooth, lack of visually-realistic etc. Zeng et al. [28] proposed Pyramid-context Encoder network (PEN-Net), assisted training by pyramid-context encoder, multi-scale decoder and an adversarial training loss, can fill missing regions at both image-level and feature-level for improving capability in image inpainting.

The main innovations are as follows:

An attention transfer network is introduced to learn the affinity in a high-level feature map between the damaged area and the visible area patch, and then convert visible area related features into low-level higher resolution feature maps according to the patch affinity weight to fill the missing content, thereby ensuring the visual and semantic coherence of image restoration.

A multi-scale decoder with deep supervision of pyramid loss and adversarial loss is proposed. Through skip connections, similar features learned by attention transfer network and latent features are decoded together to obtain repaired images, this design can not only make training converge quickly, but also can make the test more realistic.

(5) PRVS

PRVS (Progressive Reconstruction of Visual Structure) [61] introduced visual structure reconstruction (VSR) layer on the basis of partial constructions [26]. Two VSR layers are deployed in the encoder and decoder respectively to generate structural information of different scales. By gradually merging the structural information into features, a reasonable structured image is output based on the generative adversarial network, and transposed convolution is introduced into the original partial convolution layer of the sampling layer in the decoder to solve the limitations of existing module partial convolution. In the restoration process, partial convolution and bottleneck block are used to restore some edges in the missing area, and then the reconstructed edges are combined with the input image with holes to gradually reduce the size of holes by filling semantically meaningful content, and finally a fine image inpainting result is obtained.

(6) Recurrent Feature Reasoning

Aiming at [28,61] and other methods, which are easy to cause semantic ambiguity due to insufficient constraints to try to repair large defects. Li et al. [29] proposed a recurrent feature reasoning (RFR) module, which uses the correlation between adjacent pixels and enhances the constraint of estimating deep pixels, repeatedly infers the hole boundaries of convolution feature maps, and then uses them as clues for further inference. The module not only significantly improves network performance, but also bypasses some limitations of the progressive approach that the inputs and outputs of the network need to be represented in the same space. Knowledge consistent attention (KCA) module is proposed, which can adaptively combine scores from different loop processes, and ensure consistency between patch-swapping processes among recurring, leading to better results with exquisite details.

(7) Mutual Encoder-Decoder

Aiming at [20,21,62] and other methods that do not fully consider the consistency of structure and texture correlation, which lead to the problem that the inpainting results are prone to blur and artifacts. Hongyu Liu et al. [22] proposed a mutual encoder-decoder CNN for joint recovery of both structures and textures. Using deep and shallow CNN features from encoder to represent the structure and texture of the input image respectively. Encoder deep features are passed to the structure branch contain structure semantics, and shallow layers features are passed to the texture branch contain texture details. Each branch will use multiple scales of the CNN features to fill holes, concatenate both branches CNN features, then reweigh channel attentions first and use a bilateral propagation activation function to enable spatial equalization at different CNN feature levels, the decoder via skip connections generates a repaired image.

Its advantage is that can correlate filled structures with textures during image inpainting make it easier for the model to perform end-to-end training to generate more reasonable and refined structures and textures.

3.1.2. *Pluralistic inpainting approaches*

In recent years, diversified image generation has made important research progress in many image generation tasks. Pulse [63] proposed self-supervised photo upsampling, and generated high-resolution, realistic and diverse images in the field of face super-resolution by exploring the latent space. BicycleGAN [64], MUNIT [65] etc. by cross domain image translation gets diverse images.

Diversity image generation also has great practical application prospects in image inpainting tasks. Although many image inpainting methods have been able to produce visually realistic and semantically reasonable results on specific scene data, they only produce one result for each occluded image input. When the lost area is large and the image has complex texture attributes, even the repair experts in specific fields will produce different styles in detail when repairing the damaged cultural relics. Therefore, large-hole image inpainting in complex scenes should produce various reasonable restoration results.

(1) Pluralistic Image Completion

In order to produce variety of reasonable inpainting results, Zheng et al. [57] proposed a model using two parallel GAN networks. In the training phase, one of the reconstruction paths reconstructs the entire original image by using the real original image part of the mask region for confrontation training to obtain the prior distribution of the missing region. The other generation path regularizes the distribution obeyed by the encoder latent vector by using the prior distribution, which is equivalent to adding additional constraints to the encoder latent vector. It is this coupling design strategy that enables the generation path to obtain personalized complete images. In the test stage, the reconstruction path is discarded, and the generated path can repair the input mask image with limited conditional prior distribution to obtain diversified high-quality images.

(2) UCTGAN

Zhao et al. [58] proposed diversity image inpainting network UCTGAN based on unsupervised cross-space transformation, and the effect of generating inpainting images is shown in Fig. 3. The network consists of upper and lower branches as shown in Fig. 4. The main branch consists of diversity mapping module and generation module. The main branch is responsible for mapping the instance image space to the condition completion image space. The secondary branch acts as condition label in the network model and is mainly composed of condition encoder module. In this model, different inpainting images can be obtained by inputting different instance images, and several images with the best restoration effect can be returned by discriminator evaluation and comprehensive score ranking of multiple losses.

3.2. Progressive image inpainting

Given convolutional neural networks is weak in modeling long-term correlations between distant contextual information and damaged hole regions, some researchers proposed coarse-to-fine multi-stage network architecture for progressive image inpainting. The approaches related to coarse-to-fine image inpainting can be used for both conventional low-resolution image inpainting and high-resolution image inpainting.

3.2.1. Low-resolution image inpainting

(1) Contextual Attention

Many hallucination pixels will be generated in the process of image inpainting, so many plausible solutions will be generated for a specific hole. A plausible completed image may have very large patches or pixels level differences with the original image in the inpainting area. If only the original image and the repaired image are used to calculate the reconstruction loss, those differences are likely to mislead the training of convolution network. To solve this problem, Yu et al. [30] proposed Spatial discounted reconstruction loss to improve the visual quality of macroforamen restoration. Inspired by Iizuka et al. [21] works, it designed a two-stage network architecture from coarse to fine, similar to the function of residual learning [66] or deep supervision [67]. It is a feedforward fully convolution neural network without batch normalization layers [68]. It uses a thin and deep scheme to reduce the parameters of the network model and shorten the training time of the image inpainting network model. During the test, images with multiple holes of any size at any position can be repaired. The image inpainting network is divided into two stages:

As shown in Fig. 5, the holes filled with white pixels in the first stage and a binary mask of its corresponding size are used as input pairs of the coarse network in the first stage. Dilated convolution is also used in the coarse network to effectively increase the receptive field size, and reconstruction loss is used to stabilize training.

As shown in Fig. 6, the fine network in the second stage uses the coarse prediction in the first stage as input, uses improved WGAN-GP loss [69,70] for the global and local outputs in the fine network stage to enhance the global and local consistency, and jointly guides the model training in combination with the spatial attenuation reconstruction loss, thus learning finer image detail features than the coarse network. The fine network structure has two parallel encoders. The upper encoder introduces the contextual attention layer and uses the features of the

Fig. 3. Diverse inpainting results of UCTGAN.

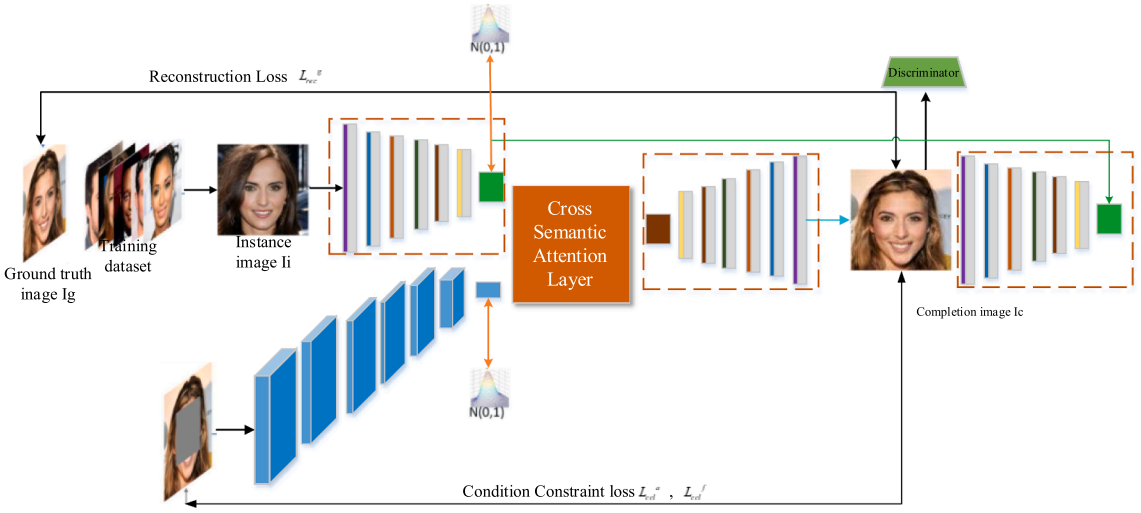


Fig. 4. Framework of UCTGAN.

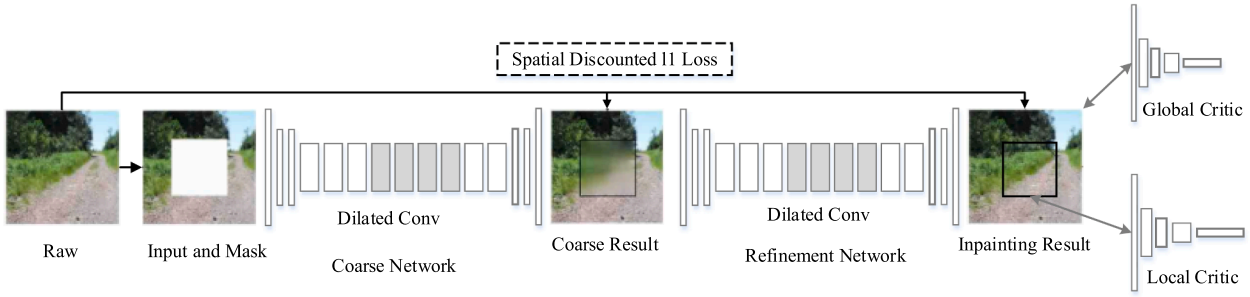


Fig. 5. Coarse-to-fine framework.

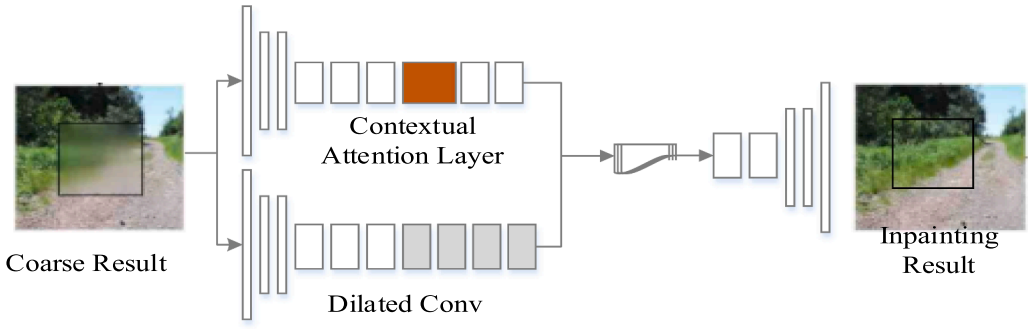


Fig. 6. Fine network structure.

visible area patches as convolutional filters to process the generated patches, focusing on extracting the background area of interest. The lower encoder imagines the content of missing area through the dilated conversion. After the outputs of the two encoders are aggregated, they are input into a decoder to reconstruct the restored image.

(2) Gated Convolution

Many scholars have made improvements and innovations on the basis of coarse-to-fine network architecture [30]. Yu et al. [31] made another step of innovation in their previous work, aiming at solving partial convolution [26] all channels in each layer share the same mask inflexibility, and when user-guided is added to the repaired image, it will be difficult to weigh whether the new user-guided information is regarded as valid pixels. Proposed improvements gated convolution replaces the vanilla convolution in the network, which better solves the problem of treating all inputs as legal pixels in the vanilla convolution, provides a learnable dynamic feature selection mechanism for each channel of each spatial location. Combined SN-patchGAN speeds up model training, and the addition of user-guided enables the new method to generate better quality and more flexible inpainting results than contextual attention [30].

(3) Coherent Semantic Attention

Liu et al. [32] used U-Net architecture [71] in both coarse and fine stages, and proposed a coherent semantic attention (CSA) layer focusing on semantic relevance and feature continuity of the hole region for the

fourth layer of the fine network encoder. Aiming at the problem that perceptual loss [72] has limited ability to optimize convolution layer in image inpainting, which may mislead the training of CSA layer, a consistency loss is introduced to solve the consistency between the corresponding layer feature maps in the encoding and decoding stage. And the feature patch discriminator combined with 70 × 70 patch discriminator [73] is introduced to accelerate and stabilize the model confrontation training, so that the refinement network synthesizes more means high frequency details.

(4) PEPSI

Sagong et al. [33] proposed a fast image inpainting method with parallel decoding structure. As shown in Fig. 7, PEPSI adopts a structure which composed of a shared coding network and a parallel decoding network with coarse path and inpainting path, can reduce the number of convolution operations and solve the problem of high occupation of computer resources for image inpainting with coarse and thin networks to a large extent. The traditional contextual attention module (CAM) [30] is improved. Euclidean distance is used instead of cosine similarities to calculate the similarity scores of foreground patches and background patches. Region ensemble discriminator (RED) is introduced to handle multiple feature regions individually, so as to solve the irregular hole that can deal with arbitrary locations, shapes, and sizes in real scenes.

3.2.2. High-resolution image inpainting

(1) Contextual Residual Aggregation

Previous image inpainting methods are often limited to low resolution image inpainting, typically smaller than 1 K. Yi et al. [35] proposed to repair 512 × 512 images at first, then perform inference on high-resolution images which bigger than 1 K. Proposed contextual residual aggregation mechanism and use attention transfer at multiple abstraction levels improved the inpainting quality, by fusing slim and deep layer configuration, light weight gated convolution (LWGC), attention score sharing and other technologies designed a light-weight model for irregular hole filling that can inference and repair on high-resolution image without occupying a lot of computing power.

(2) Iterative Confidence Feedback and Guided Upsampling

Yu and Lin et al. [34] proposed high-resolution image inpainting method to remove large target blocks without trace. The whole inpainting process is divided into two stages. In the first stage, the coarse inpainting result of low resolution image is obtained by using the cascade coarse to fine network structure. Then, in the fine inpainting stage, the confidence map of repair result is introduced to assist iterative correction of the unsatisfied area, so as to obtain the fine inpainting result. The second stage is with a guided inpainting upsampling network to generate a HR inpainting image given first stage LR inpainting result. Guided upsampling network consists of two shallow networks, one branch for learning patch similarity by patchGAN discriminator [73] and the other branch for image reconstruction.

3.3. Inpainting based on prior knowledge

Aiming at the shortcoming that single-stage inpainting and coarse-to-fine progressive inpainting methods often fail to make full use of the prior knowledge of the image visible region for accurate texture reasoning, some scholars have continuously put forward new improvement methods, which guide the network model to carry out more refined image inpainting by reasoning image contour, adding structural prior, mining image prior knowledge in GAN. So that the restored image generated by reconstruction has more reasonable texture structure and accurate semantic information [74–76].

The image inpainting methods based on prior knowledge are mainly divided into two categories: contour edge guided image inpainting and generative prior guided image inpainting.

3.3.1. Contour edge guided image inpainting

(1) FAII

FAII (Foreground-aware image inpainting) [77] is a foreground-aware image inpainting model. As shown in Fig. 8, the model first adopts DeepCut [78] to detect foreground objects in the image, then uses edge detectors to extract foreground contours, then apply coarse and thin networks to complete contours, and finally sends the completed contours and incomplete images to another coarse and thin network together, finally obtaining excellent repair results.

The innovation of this method is that the process of image structure inference and content completion is decoupled, and the natural contour of the target object is obtained, then the completed contour is used as the prior guidance of the incomplete image. It is proposed and verified that using structural prior to explicitly guide image inpainting task is a very meaningful research direction. This method is suitable for image inpainting scenes with overlapping foreground and background pixels of incomplete images and can make the restored images good and complete

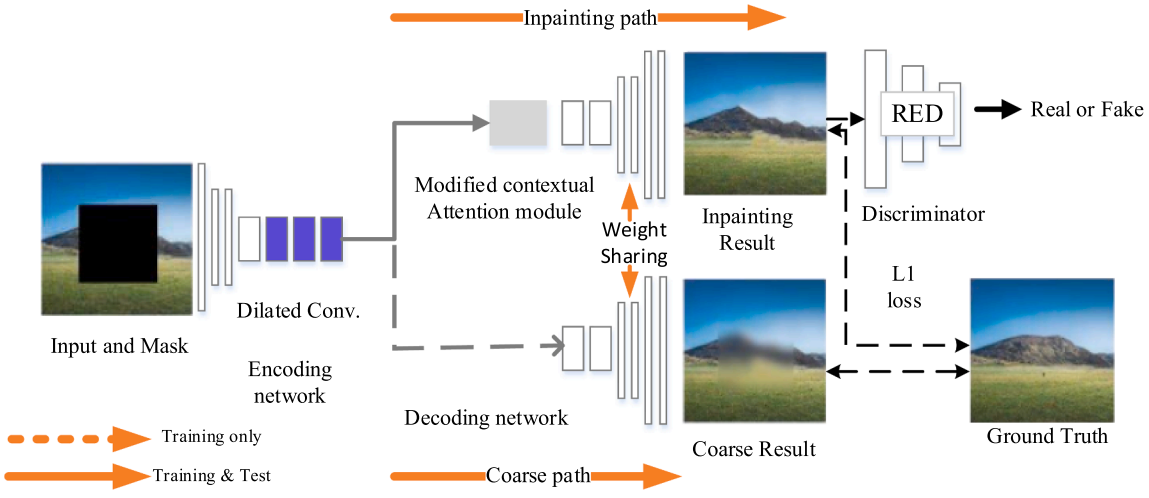


Fig. 7. Network framework of PEPSI.

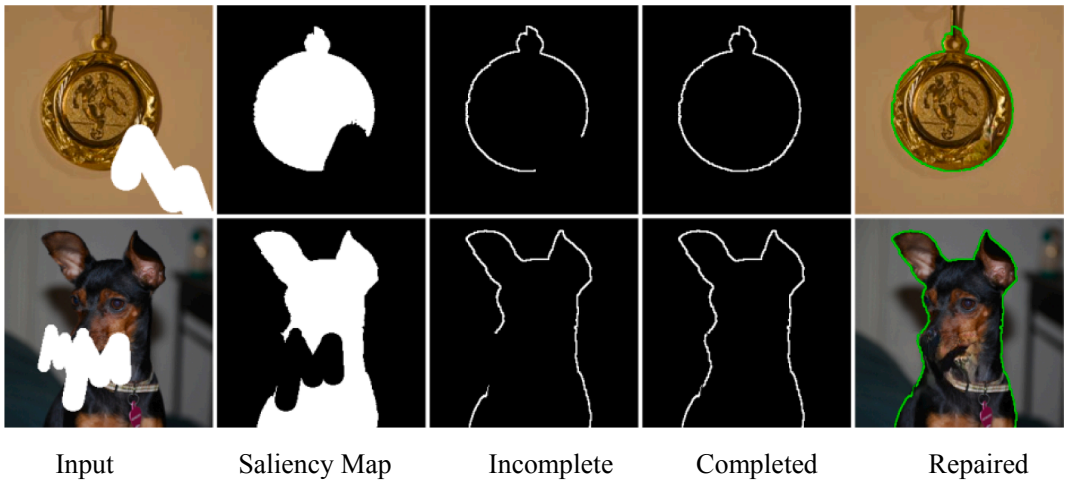


Fig. 8. Inpainting process of FAII.

edge structure information.

(2) EdgeConnect

Since the inpainting results generated by previous techniques are prone to excessive smoothness and blurring, Kamyar Nazeri et al. [56] proposed a generated image inpainting method EdgeConnect with adversarial edge learning. The method is divided into two stages of image edge detection and image completion. Mask, gray image with mask original image and edge image are the inputs of edge generator, which are used to predict the complete edge map. Edge map is used as prior knowledge and the original image with mask is used as the input of image completion network to obtain the repaired image. The inpainting effect is shown in Fig. 9. Compared with FAII [77] combining several different network models to finally obtain the completed image, EdgeConnect's network model structure design is more concise and reasonable, and easier to train.

EdgeConnect uses an edge generator to generate a rough outline in the lost area and provides a priori information of the image structure for the second-stage image completion network. The image completion network only needs to combine the prior fuzzy structure to fill and repair the details, so it can obtain a complementary image with good structure and texture, which is the innovation of the network. How to generate a reasonable edge of the lost area in the first stage will be a problem to be solved in the future of this method.

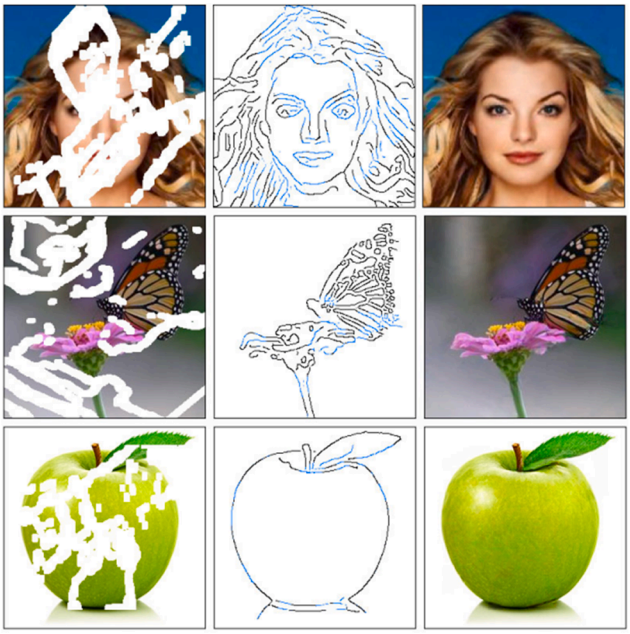


Fig. 9. Inpainting examples of EdgeConnect.

3.3.2. *Generative prior guided image inpainting*

(1) PGG

In PGG (Prior Guided GAN) [81], the best-matched damaged image corresponding to the predicted noise is extracted from the trained offline parameter model as a noise prior, sent to the generative model to reconstruct the natural image. Network is regularized by adding a priori of the structure of the target image. And then a recurrent network is proposed to help serialized reconstruction, and the model is further extended to high-pixel image inpainting and video restoration. Among them, image inpainting is regarded as a prior for perceiving the best matching latent code of the target image, and deep learning image inpainting is carried out from a new perspective, which is different from the inpainting method that directly trains the deep encoder-decoder driver on the damaged image.

(2) DGP

DGP (Deep Generative Prior) [82] utilizes a generative adversarial network trained on large-scale natural images in advance to capture rich image semantic information as a priori, which can obtain richer priori than from a single image, including color, spatial coherence, texture, high-level semantics, etc. By using the feature distance obtained by the discriminator to carry out regularization measurement and the progressive fine-tuning strategy of the generator, DGP better preserves the image statistical information learned by GAN, thus providing richer restoration and processing effects. Excellent and convincing inpainting results can be achieved in many image processing tasks such as image coloring, image completion and super-resolution reconstruction.

4. Image inpainting datasets

Currently, due to it is impossible to collect a large number of paired real damaged images, researchers often choose suitable image data set when performing image inpainting experiments, then add corresponding masks to the original data. The most widely used masks mainly include rectangular shaped hole and irregular mask, rectangular shaped hole usually added by experimenters in the center of the image or

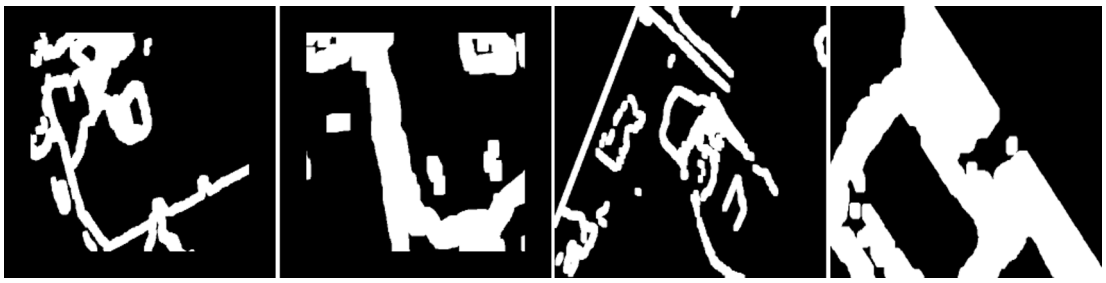


Fig. 10. Some masks of Irregular Mask Dataset Testing Set, the two on the left with border constraints, the two on the right without border constraints.

scattered with multiple small rectangular masks.

One of the most widely used is a testing set of Irregular Mask Dataset [26], contains 12, 000 masks and total six different hole-to-image area ratios, each category contains 1000 masks with border constraints (holes ensure at least 50 pixels from boundary) and 1000 masks without border constraints. Fig. 10 shows some sample data of this dataset.

In recent years, machine learning and deep learning have flourished in the field of computer vision research. Many academic institutions and companies have opened a large number of image data sets in various visual tasks to help more practitioners transform scientific research into production, providing better services for social production progress and people's lives. Based on current technology and computing power, it is still a great challenge to train a universal image inpainting model, so most model research is still based on specific types of data for training and testing. Table 1 lists several datasets most commonly used in image inpainting research. Where Fig. 11 shows some sample data in the cited datasets.

Among the dataset listed above, CelebA is mainly used for facial inpainting, ImageNet, Places2 and Paris Street View dataset are suitable for natural real scene image inpainting, and most of the data in the CMP Facade dataset and DTD dataset are more structured images.

5. Discussion and analysis

Since the birth of the two major generative models of VAE and GAN, various deep learning network models based on generative models have continuously emerged, leading the vigorous development of the entire computer vision [80,90,91,94–96]. Comparing and summarizing the above various types of representative image inpainting methods, we can find:

(1) In network selection, the image inpainting method based on convolution neural network is still the mainstream method of deep learning image inpainting application research at present, some scholars try to combine other types of networks such as variant RNN, hoping to make full use of image information at the time sequence level to achieve better image inpainting effect.

(2) As far as the generation network is concerned, VAE and GAN can effectively learn and model real data distribution from the training data. The training of image inpainting methods based on VAE is usually more stable, and the generation results are easy to be blurred. GAN-based image inpainting method can improve the quality of image inpainting generation, but it is difficult to train. Therefore, the image inpainting method based on the combination of VAE and GAN can better balance the shortcomings of the two method.

(3) At present, data feature can be better fitted by increasing the width and depth of the network. However, blindly expanding the depth and width of the network will cause model parameter explosion and training difficulties. Therefore, on the one hand, the current image inpainting network model adopts a thin and deep network structure to reduce and control the number of parameters, on the other hand, it will assist multi-scale feature layer or jump connection residual structure to help solve the gradient disappearance problem.

(4) In the task of image inpainting, vanilla convolution usually treats all input pixels as valid ones, which easy leads to artifacts such as color discrepancy and blurriness. Therefore, the introduction of partial convolution [26] or gated convolution [31] can alleviate this problem to a certain extent, so the improvement based on convolution is also a feasible breakthrough direction in the field of image inpainting.

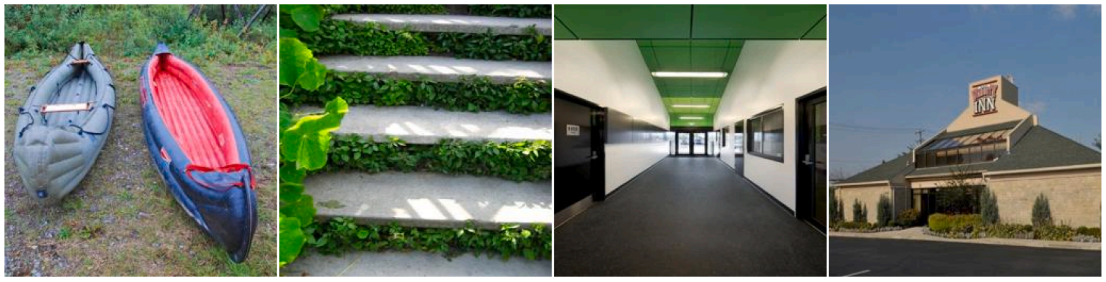
(5) When using coarse-to-fine network model for progressive image inpainting, it is easy to encounter the problems that the network model is too complex. Therefore, it is a meaningful attempt to study how to efficiently reuse the specific network layer of the encoder and decoder for controlling the number of model parameters and thus improving the training efficiency of the model. For example, the coarse path and the fine path share their weights to improve each other and both they use same encoder in PEPSI [33].

(6) Pluralistic inpainting can provide a variety of reasonable inpainting results when repairing seriously damaged

Table 1
Image inpainting datasets.

Dataset	Category	Description	Publisher
ImageNet[83]	Multiclass	A large scale visualization database for research of visual object recognition.	Stanford University
Places2 [84]	Multiclass	Places dataset contains more than 10 million images, including more than 434 unique scene categories and can be used for high-level visual understanding tasks.	Massachusetts Institute of Technology
CelebA [85,86]	Celebrity face	A large-scale face attributes dataset with 202,599 number of face images and 40 binary attributes annotations per image.	Chinese University of Hong Kong
Paris Street View [87]	Street view images	It is selected from Google Street View Dataset, which contains 15,000 high-quality Street View images and mainly focuses on the buildings in the city.	Carnegie Mellon University, Google
CMP façade [88]	Facade of buildings	It has 606 highly structured facades images from multiple cities buildings around the world.	Czech Technical University
DTD[89]	Texture images	It is divided into 47 categories, and each category has 120 images, a total of 5640 texture images.	University of Oxford

(a) ImageNet



(b) Places2

(c) CelebA

(d) Paris

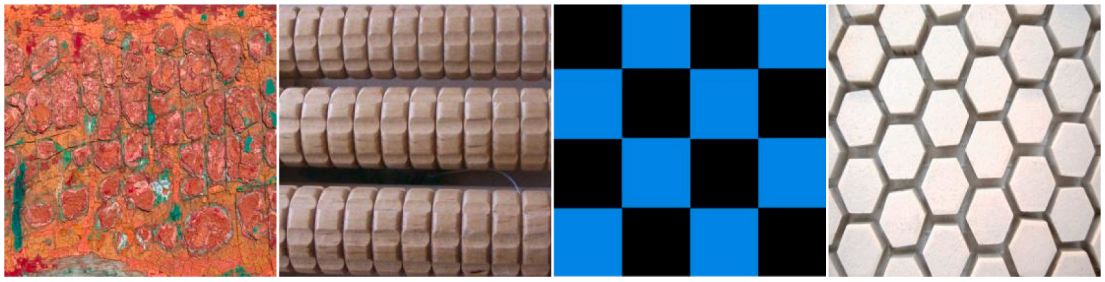
Fig. 11. Examples from image inpainting datasets.

unstructured complex texture objects and can provide users with richer choices and inpainting experiences in practical applications.

(7) Due to the limitation that it will occupy more resources when directly repairing high-resolution images. Therefore, the current mainstream method for high-resolution image inpainting will first repair the low-resolution image obtained by downsampling the original image. Then, the repaired area is up-sampled or the clarity technology of super resolution reconstruction is used to obtain the repaired area of the original image size level, and last the corresponding damaged area is replaced, thus indirectly completing the high-resolution image inpainting task.

(8) Compared with the pure data-driven deep learning image inpainting method, the inpainting method by adding inference image contour module or structure prior can make full use of the prior knowledge of the visible region to perform more accurate texture inference. Therefore, complex texture image inpainting based on prior knowledge will be a very meaningful research work when applied in specific scenes.

(e) Facade



(f) DTD

Fig. 11. (*continued*).

Table 2
Quantitative comparison on rectangular holes inpainting.

Method	PSNR	SSIM
SF [92]	25.9794	0.8835
CA [30]	24.2377	0.8671
CE [20]	26.1634	0.8910
CSA [32]	26.1920	**0.9021**
SN [93]	26.0732	0.8671
PIC [57]	24.4229	0.8692
UCTGAN [58]	**26.3833**	0.8862

The current image inpainting research mainly focuses on rectangular holes and irregular masks. Table 2 and Table 3 respectively show the quantitative comparison of some of the latest research methods on relevant data sets.

Table 2 shows quantitative comparison with the state-of-the-art approaches on CelebA-HQ face dataset which are trained on 128 × 128 rectangle holes in the center. The comparison is random selected 1000 test images of resolution 256 × 256, with quantitative measures of peak signal-to-noise ratio (PSNR), structural similarity (SSIM). CE, SN, PN and UCTGAN are single-stage methods, SF, CA, CSA are progressive inpainting methods. The methods with prior knowledge are not listed, because prior guidance is often used for irregular image inpainting. Table 2 experimental results are extracted from literature UCTGAN [58].

From the comparison results in Table 2, it can be seen that the

Table 3
Quantitative comparison on irregular mask inpainting.

Dataset		Places2			CelebA			Paris Street View		
Mask Ratio		10%–20%	30%–40%	50%–60%	10%–20%	30%–40%	50%–60%	10%–20%	30%–40%	50%–60%
SSIM	PIC[57]	0.932	0.786	0.494	0.965	0.881	0.672	0.930	0.785	0.519
	PConv[26]	0.934	0.803	0.555	0.977	0.922	0.791	0.947	0.835	0.619
	GatedConv [31]	–	–	–	0.973	0.914	0.767	0.953	0.849	0.621
	EdgeConnect [56]	0.933	0.802	0.553	0.975	0.915	0.759	0.950	0.849	0.646
	PRVS[61]	0.936	0.810	0.574	0.978	0.926	0.799	0.953	0.854	0.659
	RFR-Net [29]	**0.939**	**0.819**	**0.596**	**0.981**	**0.934**	**0.819**	**0.954**	**0.862**	**0.681**
PSNR	PIC	27.14	21.72	17.17	30.67	24.74	19.29	29.35	23.97	19.52
	PConv	27.29	22.12	18.29	32.77	26.94	22.14	30.76	25.46	21.39
	GatedConv	–	–	–	32.56	26.72	21.47	31.32	25.54	20.61
	EdgeConnect	27.17	22.18	18.35	32.48	26.62	21.49	31.19	26.04	21.89
	PRVS	27.41	22.36	18.67	33.05	27.24	22.37	31.49	26.17	22.07
	RFR-Net	**27.75**	**22.63**	**18.92**	**33.56**	**27.76**	**22.88**	**31.71**	**26.44**	**22.40**

current face inpainting based on rectangular holes still needs to be solved effectively. There is no obvious better method in single-stage methods and progressive inpainting methods.

Table 3 compares the experimental results of some typical irregular mask image inpainting methods on places2, celeba and Paris street view test sets. Here is a comparison of 6 methods, among which PIC, Pconv and RFR-Net are single-stage methods, GatedConv, PRVS are progressive inpainting methods, EdgeConnect is belonged inpainting based on prior knowledge. Table 3 experimental results are taken from literature RFR-Net [29].

From the results of Table 3, it can be seen that when the mask ratio of irregular image inpainting is quite different, the quantitative indicators such as PSNR and SSIM of image inpainting results are also obviously different. On the whole, when the mask ratio is low, current various methods can basically achieve excellent repair results. When the mask ratio is above 50%, most models are difficult to achieve satisfactory repair results. Relatively speaking, face-based image inpainting is easier to achieve satisfactory inpainting results, while image inpainting in natural scenes with complex textures still has much room for improvement.

6. Conclusions

At present, image inpainting technology has become an important branch in field of vision research. Deep learning image inpainting based on generation network gradually become mainstream method. Researchers have continuously innovated and made great progress in generation model selection, network structure design, introduction of prior guidance, discriminator optimization, loss function optimization, etc. However, the following problems still need to be solved urgently:

(1) The current image inpainting methods can achieve better inpainting results in processing regular structured data, small hole inpainting and low-resolution image inpainting. How to improve the inpainting effect of complex texture, large holes and high-resolution images has become the main focus in the follow-up research of image inpainting.
(2) Network selection and design, image inpainting based on GAN network, autoencoder and the combination of the two are currently main base framework. How to apply other deep learning models to image inpainting is worth exploring. In addition, generally speaking, the deeper network structure is, the better reconstruction and repair effect will be, but it will also lead to problems such as difficult training and convergence. How to balance the contradiction between the complexity of the network and the quality of the restored image is a problem that needs further study.
(3) How to combine domain knowledge, prior knowledge and deep learning framework in specific applications more effectively to improve the existing image restoration performance based on deep learning is a direction worth exploring. Domain and priori knowledge are correct knowledge and experience obtained by human beings through long-term research in related fields, which have important guiding significance. Deep learning can make full use of the big data training model, automatically select, learn useful features in repair task, learn nonlinear mapping from damaged images to latent code and then to repaired images. If we can make full use of domain and prior knowledge to guide deep learning model, not only can high-level semantic features rich in context information be extracted, and then more complex mapping relationship between damaged images and repaired images can be learned, but also the rationality of this mapping relationship can be ensured, thus further improving the reconstruction performance and interpretability of the image inpainting model based on deep learning.
(4) Video stream image inpainting, the current image inpainting methods based on deep learning benefit from the good spatial feature extraction ability of convolution neural network, and most of them use deep convolution neural network to build network layer. Recurrent neural network can mine semantic information at the time series feature level in data, and has a good application in the field of speech and natural language processing. How to effectively combine the two neural networks to deal with video stream image inpainting will be a very meaningful research direction.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgment

This work was supported by the National Science Foundation of China (Grant No. 61901436) and the Key Research Program of the Chinese Academy of Sciences (Grant No. XDPB22).

References

[1] O. Elharrouss, N. Almaadeed, S. Al-Maadeed, et al., Image inpainting: a review, Neural Process. Lett. 51 (2020) 2007–2028, https://doi.org/10.1007/s11063-019-10163-0.
[2] Marcelo Bertalmio, Guillermo Sapiro, Vincent Caselles, Coloma Ballester, Image inpainting, in: Proceedings of the 27th annual conference on Computer graphics and interactive techniques, 2000, 417–424. DOI: https://doi.org/10.1145/344779.344972.
[3] J.H. Shen, S.H. Kang, T.F. Chan, Euler's elastica and curvaturebased inpainting, SIAM J. Appl. Math. 63 (2) (2003) 564–592.
[4] A. Tsai, A. Yezzi, A.S. Willsky, Curve evolution implementation of the Mumford-Shah functional for image segmentation, denoising, interpolation, and magnification, IEEE Trans. Image Process. 10 (8) (2001) 1169–1186.
[5] M. Bertalmio, A.L. Bertozzi, G. Sapiro, Navier-stokes, fluid dynamics, and image and video inpainting, in: Proceedings of 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Kauai, HI, USA, IEEE, 2001, 355–362.
[6] X. Ning, W. Li, B. Tang, H. He, BULDP: biomimetic uncorrelated locality discriminant projection for feature extraction in face recognition, IEEE Trans Image Process 27 (5) (May 2018) 2575–2586, https://doi.org/10.1109/TIP.2018.2806229.
[7] Xin Ning, Ke Gong, Weijun Li, Liping Zhang, Xiao Bai, Shengwei Tian, Feature refinement and filter network for person re-identification, in: IEEE Transactions on Circuits and Systems for Video Technology, 2020, pp. 1–1.
[8] H. Qin, R. Gong, X. Liu, X. Bai, J. Song, N. Sebe, Binary neural networks: A survey, Pattern Recogn. 105 (2020), 107281.
[9] Zongyu Guo, Zhibo Chen, Tao Yu, Jiale Chen, Sen Liu, Progressive image inpainting with full-resolution residual network, 2019, 2496–2504.
[10] Rui Xu, Minghao Guo, Jiaqi Wang, Xiaoxiao Li, Bolei Zhou, Chen Change Loy, Texture memory-augmented deep patch-based image inpainting, 2020.
[11] Håkon Hukkelås, Frank Lindseth, Rudolf Mester, Image inpainting with learnable feature imputation, 2020.
[12] Chao Yang, Xin Lu, Zhe Lin, Eli Shechtman, Oliver Wang, Hao Li, High-resolution image inpainting using multi-scale neural patch synthesis, 2017, 4076–4084.
[13] D. Kim, S. Woo, J.Y. Lee, et al., Deep video inpainting, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020.
[14] Y. Wang, X. Tao, X. Qi, et al., Image inpainting via generative multi-column convolutional neural networks, in: Advances in Neural Information Processing Systems, 2018.
[15] Ya-Liang Chang, Zhe Liu, Kuan-Ying Lee, Winston Hsu, Learnable gated temporal shift module for deep video inpainting, 2019.
[16] U.S.M. Nadim, Y.J. Jung, Global and local attention-based free-form image inpainting, Sensors 20 (11) (2020) 3204.
[17] Avisek Lahiri, Arnav Jain, Prabir Biswas, Pabitra Mitra, Improving consistency and correctness of sequence inpainting using semantically guided generative adversarial network, 2017.
[18] Avisek Lahiri, Sourav Bairagya, Sutanu Bera, Siddhant Haldar, Prabir Biswas, Lightweight modules for efficient deep learning based image restoration, 2020.
[19] Z.P. Qiang, L.B. He, X. Chen, D. Xu, Survey on deep learning image inpainting methods, J. Image Graph. 24 (03) (2019) 0447–0463.
[20] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, Alexei A. Efros, Context encoders: Feature learning by inpainting, in: Proc. CVPR, 2016. 3, 2536–2544.

[21] Satoshi Iizuka, Edgar Simo-Serra, Hiroshi Ishikawa. Globally and locally consistent image completion, ACM TOG 36(4) (2017) 107.
[22] Hongyu Liu, Bin Jiang, Yibing Song, Wei Huang, Chao Yang, Rethinking image inpainting via a mutual encoder-decoder with feature equalizations, 2019MEDFE, 36(4):107, 2020.
[23] Y. Song, C. Yang, Z. Lin, X. Liu, Q. Huang, H. Li, C. Jay, Contextual-based image inpainting: Infer, match, and translate, ECCV (2018).
[24] X. Ning, P. Duan, S. Zhang, Real-time 3D face alignment using an encoder-decoder network with an efficient deconvolution layer, IEEE Signal Process Lett. 27 (2020) 1944–1948, https://doi.org/10.1109/LSP.2020.3032277.
[25] X. Ning, K.e. Gong, L.i. Weijun, L. Zhang, JWSAA: Joint weak saliency and attention aware for person re-identification, Neurocomputing (2020), https://doi.org/10.1016/j.neucom.2020.05.106.
[26] Guilin Liu, Fitsum A. Reda, Kevin J. Shih, Ting-Chun Wang, Andrew Tao, Bryan Catanzaro, Image inpainting for irregular holes using partial convolutions, in: Proc. ECCV, 2018. 3, 4, 6, 7, 85–100.
[27] Chaohao Xie, Shaohui Liu, Chao Li, Ming-Ming Cheng, Wangmeng Zuo, Xiao Liu, Shilei Wen, Errui Ding, Image inpainting with learnable bidirectional attention maps, in: Proc. ICCV, 2019. 1, 3, 8858–8867.
[28] Y. Zeng, J. Fu, H. Chao, B. Guo, Learning pyramid-context encoder network for high-quality image inpainting, in: CVPR, 2019.
[29] Jingyuan Li, Wang Ning, Lefei Zhang, Bo Du, Dacheng Tao, Recurrent feature reasoning for image inpainting, 2020, 7757–7765. 10.1109/CVPR42600.2020.00778.
[30] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, Thomas S. Huang, Generative image inpainting with contextual attention, in: Proc. CVPR, 2018. 1, 2, 3, 4, 8, 5505–5514.
[31] Jiahui Yu, Zhe Lin, Jimei Yang, Xiaohui Shen, Xin Lu, Thomas S. Huang, Free-form image inpainting with gated convolution, in: Proc. ICCV, 2019. 1, 3, 6, 7, 4471–4480.
[32] Hongyu Liu, Bin Jiang, Yi Xiao, Chao Yang, Coherent semantic attention for image inpainting, in: Proc. ICCV, 2019. 3, 7, 4170–4179.
[33] Min-cheol Sagong, Yong-goo Shin, Seung-wook Kim, Seung Park, Sung-jea Ko, Pepsi: Fast image inpainting with parallel decoding network, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019. 1, 11360–11368.
[34] Yu Zeng, Zhe Lin, Jimei Yang, Jianming Zhang, Eli Shechtman, Huchuan Lu, High-resolution image inpainting with iterative confidence feedback and guided upsampling, 2020.
[35] Zili Yi, Qiang Tang, Shekoofeh Azizi, Daesik Jang, Zhan Xu, Contextual residual aggregation for ultra high-resolution image inpainting, 2020. 10.1109/CVPR42600.2020.00753.
[36] Wenchao Du, Hu Chen, Hongyu Yang, Learning invariant representation for unsupervised image restoration, 2020, 14471–14480. 10.1109/CVPR42600.2020.01449.
[37] Ziyu Wan, Bo Zhang, Dongdong Chen, Pan Zhang, Dong Chen, Jing Liao, Fang Wen, Bringing old photos back to life, 2020, 2744–2754. 10.1109/CVPR42600.2020.00282.
[38] D. Ulyanov, A. Vedaldi, V. Lempitsky, Deep Image Prior, Int. J. Comput. Vision 128 (2020), https://doi.org/10.1007/s11263-020-01303-4.
[39] C. Ballester, M. Bertalmio, V. Caselles, G. Sapiro, J. Verdera, Filling-in by joint interpolation of vector fields and gray levels, IEEE Trans. Image Process. 10 (8) (2001) 1200–1211.
[40] Anat Levin, Assaf Zomet, Yair Weiss, Learning how to inpaint from global image statistics, in: null, 305. IEEE, 2003.
[41] S. Esedoglu, J. Shen, Digital inpainting based on the mumford–shah–euler image model, Eur. J. Appl. Math. 13(4) (2002) 353–370. 1, 2.
[42] M. Bertalmio, L. Vese, G. Sapiro, S. Osher, Simultaneous structure and texture image inpainting, IEEE Trans. Image Process. 12 (8) (2003) 882–889.
[43] Antonio Criminisi, Patrick Perez, Kentaro Toyama, Object removal by exemplar-based inpainting, in: Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on, volume 2, II– II. IEEE, 2003.
[44] A. Criminisi, P. Perez, K. Toyama,· Region filling and object removal by exemplar-based image inpainting, IEEE Trans. Image Process. 13 (9) (2004) 1200–1212.
[45] C. Barnes, E. Shechtman, A. Finkelstein, D.B. Goldman, Patchmatch: A randomized correspondence algorithm for structural image editing, ACM Trans. Graphics (ToG) 28 (24) (2009).
[46] J.-B. Huang, S.B. Kang, N. Ahuja, J. Kopf, Image completion using planar structure guidance, ACM Trans. Graphics (TOG) 33(4) (2014) 129. 1, 2.
[47] S. Darabi, E. Shechtman, C. Barnes, D.B. Goldman, P. Sen, Image melding: Combining inconsistent images using patch-based synthesis, ACM Trans. graphics (TOG) 31(4) (2012) 82–81. 1, 2.
[48] James Hays, Alexei A. Efros, Scene completion using millions of photographs, in: ACM Transactions on Graphics (TOG), volume 26, 4. ACM, 2007.
[49] X. Bai, C. Yan, L.u. Haichuan Yang, J.Z. Bai, Edwin Robert Hancock: Adaptive hash retrieval with kernel based similarity, Pattern Recogn. 75 (2018) 136–148.
[50] X. Bai, E.R. Hancock, R.C. Wilson, Graph characteristics from the heat kernel trace, Pattern Recogn. 42 (11) (2009) 2589–2606.
[51] X. Liu, J. Zhou, E.R. Hancock, Learning binary code for fast nearest subspace search, Pattern Recogn. 98 (2020).
[52] Diederik P. Kingma, Max Welling, Auto-encoding variational bayes, arXiv preprint arXiv:1312.6114, 2013.
[53] Romain Lopez, Jeffrey Regier, Nir Yosef, Michael Jordan, Information Constraints on Auto-Encoding Variational Bayes, 2018.
[54] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio, Generative adversarial nets, in: Advances in neural information processing systems, 2672–2680, 2014.
[55] Mehdi Mirza, Simon Osindero, Conditional Generative Adversarial Nets, 2014.
[56] Kamyar Nazeri, Eric Ng, Tony Joseph, Faisal Qureshi, Mehran Ebrahimi, Edgeconnect: Structure guided image inpainting using edge prediction, in: Proc. ICCV Workshops, 2019. 1, 3, 6.
[57] Chuanxia Zheng, Tat-Jen Cham, Jianfei Cai, Pluralistic image completion, in: Proc. CVPR, 2019, 1438–1447. 6.
[58] Lei Zhao, Qihang Mo, Sihuan Lin, Zhizhong Wang, Zhiwen Zuo, Haibo Chen, Wei Xing, Dongming Lu, UCTGAN: Diverse Image Inpainting Based on Unsupervised Cross-Space Translation, 2020, 5740–5749. 10.1109/CVPR42600.2020.00578.
[59] Fisher Yu, Vladlen Koltun, Multi-scale context aggregation by dilated convolution, in: International Conference on Learning Representations, 2016.
[60] Jian Sun, Lu Yuan, Jiaya Jia, Heung-Yeung Shum, Image completion with structure propagation, in: TOG, vol. 24, 2005. 2, 861–868.
[61] Jingyuan Li, Fengxiang He, Lefei Zhang, Bo Du, Dacheng Tao, Progressive reconstruction of visual structure for image inpainting, 2019, 5961–5970.
[62] Y. Li, S. Liu, J. Yang, M.H. Yang, Generative face completion, in: CVPR, 2017. 4.
[63] Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, Cynthia Rudin, PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models, 2020, 2434-2442 10.1109/CVPR42600.2020.00251.
[64] Jun-Yan Zhu, Richard Zhang, Deepak Pathak, Trevor Darrell, Alexei A. Efros, Oliver Wang, Eli Shechtman, Toward multimodal image-to-image translation, in: Advances in Neural Information Processing Systems, 2017, 465–476.
[65] Xun Huang, Ming-Yu Liu, Serge Belongie, Jan Kautz, Multimodal Unsupervised Image-to-Image Translation, 2018.
[66] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, 770–778.
[67] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, Z. Tu, Deeplysupervised nets, in: Artificial Intelligence and Statistics, 2015, 562–570.
[68] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, in: International Conference on Machine Learning, 2015, 448–456.
[69] M. Arjovsky, S. Chintala, L. Bottou, Wasserstein gan. arXiv preprint arXiv: 1701.07875, 2017.
[70] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. Courville. Improved training of wasserstein gans. arXiv preprint arXiv:1704.00028, 2017.
[71] Olaf Ronneberger, Philipp Fischer, Thomas Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation. LNCS. 9351, 2015, 234–241. 10.1007/978-3-319-24574-4_28.
[72] J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, ECCV, 2016.
[73] P. Isola, J.Y. Zhu, T. Zhou, A.A. Efros, Image-to-image translation with conditional adversarial networks, CVPR, 2017.
[74] Y. Yang, X. Guo, J. Ma, et al., LaFIn: Generative Landmark Guided Face Inpainting, 2019.
[75] J. Yang, Z. Qi, Y. Shi, Learning to Incorporate Structure Knowledge for Image Inpainting, 2020.
[76] Jinjin Gu, Yujun Shen, Bolei Zhou, Image processing using multi-code GAN Prior, 2020, 3009–3018. 10.1109/CVPR42600.2020.00308.
[77] W. Xiong, Z. Lin, J. Yang, et al., Foreground-aware Image Inpainting, in: IEEE/CVF Conference on Computer Vision, Pattern Recognition, IEEE, 2019. 5833-5841.
[78] Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam, Rethinking atrous convolution for semantic image segmentation, 2017.
[79] Xin Ning, Weijun Li, Wenjie Liu, A fast single image haze removal method based on human retina property, IEICE Transactions on information and system, 2017, E100. D(1):211–214. DOI:10.1587/transinf.2016edl8180.
[80] Xin Ning, Weijun Li, Jian Xu, The principle of homology continuity and geometrical covering learning for pattern recognition, Int. J. Pattern Recogn. Artificial Intelligence 32(12) (2018). DOI:10.1142/s0218001418500428.
[81] A. Lahiri, A.K. Jain, S. Agrawal, et al., Prior guided GAN based semantic inpainting, in: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2020,13693-13702.
[82] X. Pan, X. Zhan, B. Dai, et al., Exploiting Deep Generative Prior for Versatile Image Restoration and Manipulation, Springer, Cham, 2020.
[83] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet Large Scale Visual Recognition Challenge, Int. J. Comput. Vision (IJCV) 115 (3) (2015) 211–252.

[84] Bolei Zhou, Agata Lapedriza, Aditya Khosla, Aude Oliva, Antonio Torralba, Places: A 10 million image database for scene recognition, IEEE TPAMI 40(6) (2018) 1452–1464.
[85] Ziwei Liu, Ping Luo, Xiaogang Wang, Xiaoou Tang, Deep learning face attributes in the wild, in: ICCV, 3730–3738, 2015. 6, 10.
[86] Tero Karras, Timo Aila, Samuli Laine, Jaakko Lehtinen, Progressive growing of gans for improved quality, stability, and variation, in: ICLR, 2018. 6.
[87] Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, Alexei Efros, What makes paris look like paris? ACM TOG 31(4) (2012) 101. 6.
[88] Radim Tylecek, Radim Šára, Spatial pattern templates for recognition of objects with regular structure, in: GCPR, 364–374, 2013. 6, 10.
[89] Mircea Cimpoi, Subhransu Maji, Iasonas Kokkinos, Sammy Mohamed, Andrea Vedaldi, Describing textures in the wild, in: CVPR, 3606–3613, 2014. 6, 10.
[90] Ning Xin, Nan Fangzhe, Xu Shaohui, Yua Lina, Liping Zhang, Multi-view frontal face image generation: A survey, Concurrency and Computation: Practice and Experience, 2020: e6147. https://doi.org/10.1002/cpe.6147.
[91] Zhongpai Gao, Guangtao Zhai, Hongwei Deng, Xiaokang Yang, Extended geometric models for stereoscopic 3D with vertical screen disparity, 65, 2020. DOI: doi.org/10.1016/j.displa.2020.101972.
[92] Yurui Ren, Xiaoming Yu, Ruonan Zhang, Thomas H Li, Shan Liu, Ge Li, Structureflow: Image inpainting via structure-aware appearance flow, in: Proceedings of the IEEE International Conference on Computer Vision, 181–190, 2019.
[93] Zhaoyi Yan, Xiaoming Li, Mu Li, Wangmeng Zuo, Shiguang Shan, Shift-net: Image inpainting via deep feature rearrangement, in: Proceedings of the European Conference on Computer Vision (ECCV), 1–17, 2018.
[94] Jianming Zhang, Zhipeng Xie, Juan Sun, Xin Zou, A cascaded R-CNN with multiscale attention and imbalanced samples for traffic sign detection, IEEE Access 8 (2020) 29742–29754, https://doi.org/10.1109/ACCESS.2020.2972338.
[95] Jianming Zhang, Wei Wang, Lu. Chaoquan, Jin Wang, Arun Kumar Sangaiah, Lightweight deep network for traffic sign classification, Ann. Telecommun. 75 (7–8) (2020) 369–379, https://doi.org/10.1007/s12243-019-00731-9.
[96] Cheng Yan, Guangsong Pang, Xiao Bai, et. Beyond Triplet Loss, Person Re-identification with Fine-grained Difference-aware Pairwise Loss, IEEE Trans. Multimedia (2021).