

文章编号: 1671-7449(2024)01-0034-07

基于上下文特征提取的边缘生成三阶段 图像修复算法

芮志超, 郭艳艳*

(山西大学 物理电子工程学院, 山西 太原 030006)

摘要: 对于具有较大不规则缺失区域的图像修复问题, 现有的基于深度学习的图像修复方法通常会生成具有模糊纹理和扭曲结构的内容。针对这个问题, 将修复问题分解为基于上下文特征的结构预测和图像补全三阶段模型。第一阶段, 通过空洞卷积编-解码网络, 利用周围图像特征来对缺失部分进行初步修复; 第二阶段, 将第一阶段粗修复结果进行边缘提取后, 输入到一个自注意力机制编-解码网络来预测缺失区域的纹理结构; 第三阶段, 将前两个阶段的输出一起输入到一个改进的U-net精修复网络中, 得到结构清晰、纹理细节丰富的图像。在公开数据集上将所提算法与现有经典算法进行对比, 实验表明, 所提方法在主观视觉和客观评价方面优于现有方法。

关键词: 深度学习; 图像修复; 自注意力机制; Res2net; 生成式对抗网络

中图分类号: TP391.41 **文献标识码:** A **doi:** 10.3969/j.issn.1671-7449.2024.01.005

引用格式: 芮志超, 郭艳艳. 基于上下文特征提取的边缘生成三阶段图像修复算法[J]. 测试技术学报, 2024, 38(1): 34-40.

RUI Zhichao, GUO Yanyan. Three-stage image inpainting algorithm by edge generation based on contextual feature extraction[J]. Journal of Test and Measurement Technology, 2024, 38(1): 34-40.

Three-Stage Image Inpainting Algorithm by Edge Generation Based on Contextual Feature Extraction

RUI Zhichao, GUO Yanyan*

(College of Physics and Electronic Engineering, Shanxi University, Taiyuan 030006, China)

Abstract: For an image of large irregular missing regions, recent deep learning-based image inpainting approaches often generate content with blurred textures and distorted structures. To solve these problems, a three-stage model that separates the inpainting problem into contextual feature-based structure prediction and image completion is proposed. In the first stage, our model utilizes surrounding image features to predict missing regions during a dilated convolutional encoder-decoder network training. In the second stage, an encoder-decoder network based on the self-attention mechanism takes edge features extracted from the first stage predictions as inputs and recovers the texture structure of the missing regions. In the third stage, the outputs of the first two stages are passed to a refined inpainting network using the improved U-net architecture to guide the repair process. The proposed algorithm is compared with the existing classic algorithm on the public datasets. Experiments show that our method outperforms existing methods in terms of subjective vision and objective evaluation.

Key words: deep learning; image inpainting; self-attention mechanism; Res2net; generative adversarial networks

收稿日期: 2022-11-20

作者简介: 芮志超(1997-), 男, 硕士生, 主要从事机器学习、人工智能和图像修复算法研究。E-mail: rzc1997@163.com。

* 通信作者: 郭艳艳(1976-), 女, 副教授, 主要从事人工智能、移动通信和信号处理研究。E-mail: guoyanyan@sxu.edu.cn。

0 引言

由于图像在传输和获取过程中受到人为或环境的影响会导致图像信息丢失,需要对图像进行修复处理,增强其视觉效果。传统图像修复是根据图像样本相似度、结构纹理一致性等方法,结合数学、物理理论构建算法模型修复破损图像,主要包括基于扩散的方法^[1]和基于补丁的方法^[2-3]。但是,传统的修复算法只能基于已有区域的信息来填充未知区域,未出现在已知区域的内容无法预知填充。

近年来,深度卷积网络和生成式对抗网络已被大量应用到图像修复中^[4-6]。Pathak D等^[7]在2016年提出了上下文编码器,该算法通过构建一个编码器-解码器结构完成缺损图像的编解码工作,并添加判别损失完成图像的精细修复,但该算法只能修复简单结构的破损图像,对于人脸图像修复远远不足;Yu J等^[8]在2019年提出利用门控卷积层替代传统卷积层,解决了传统卷积层将图像所有像素视为有效像素的缺点,该算法只适用于规则掩膜,而对于自由掩膜而言,图像修复效果差;Nazeri K等^[9]在2019年提出一种Edge-Connection模型,通过设计额外的辅助网络,利用缺损区域边缘信息辅助图像修复;Guo X等^[10]在2021年提出了一种结构约束纹理合成和纹理引导结构重建的耦合网络来完成整个修复过程,并设计了双向门控特征融合模块来实现结构和

纹理信息的融合。

综上所述,现有修复人脸图像的算法可以分为单阶段修复模型、渐进式两阶段模型和基于结构先验或者语义先验约束的图像修复模型。单阶段修复模型在修复背景复杂、纹理精细的图像时,容易产生模糊的纹理细节;渐进式两阶段模型虽然能生成纹理细节,但是图像整体的结构会发生扭曲;而基于结构先验或者语义先验约束的图像修复模型,结构与语义一致性有了保障,但修复图像纹理细节不贴近真实图像。

针对上述问题,本文提出了一种基于上下文特征提取的边缘生成三阶段的图像修复算法。首先,算法利用破损图像的上下文特征进行初步修复,并对初修复结果的边缘进行重建;然后,将重建后的边缘结构以及初修复结果一起作为先验信息,来约束最后阶段的精修复过程。并且,在边缘修复网络和精修复网络中加入Res2net模块^[11]和自注意力机制,提取多尺度特征以提升网络的表征能力,自适应计算远距离特征之间的依赖关系以解决图像修复区域和未缺失区域的边界衔接问题。

1 系统模型

如图1所示,系统主要采用生成式对抗网络框架^[12]。该架构由生成器和鉴别器组成,生成器部分主要包含3个网络结构:粗修复网络、边缘修复网络、精修复网络。鉴别器采用补丁鉴别器^[8,13]。

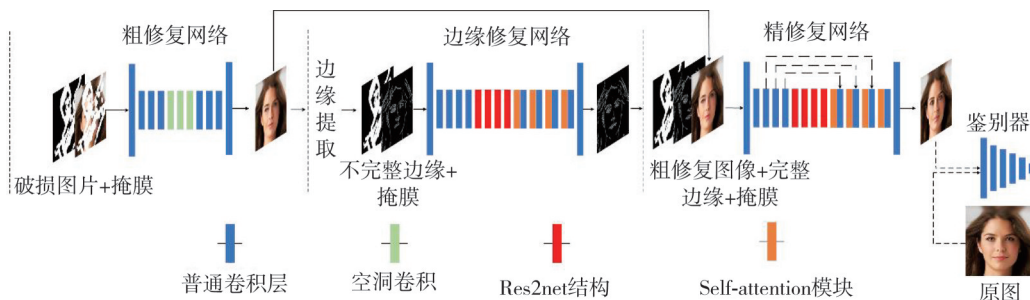


图1 三阶段修复算法结构图

Fig. 1 Structure diagram of three-stage inpainting algorithm

1.1 粗修复网络

为了能初步重建大面积不规则的缺失区域,粗修复网络结构参考文献[8]中的架构,由13个普通卷积层和4个空洞卷积层组成,其中普通卷积层包含两层卷积核大小为3,步长为2的下采样层,两层步长为1/2的上采样层,上采样通过结合下采样中提取的各层信息,还原图像缺失区

域;而4层空洞卷积层的空洞率大小分别为2,4,8,16,目的是为了扩大卷积核的感受野,提升网络性能。粗修复网络结构见表1所示。

用 I_{gt} 代表原始图像, $\bar{I}_{gt} = I_{gt} \otimes (1 - M)$ 表示破损图片, M 为随机生成的自由掩膜(缺失部分为1,背景为0),粗修复网络 G_c 生成的修复结果为

$$I_{pre}^c = G_c(\bar{I}_{gt}, M). \quad (1)$$

表1 粗修复网络结构

Tab. 1 Coarse inpainting network structure

每层名称	输入	输出	卷积核大小	步长	空洞率	激活函数
Conv_1	256*256*4	256*256*64	5	1	1	Elu
Down_Conv_2	256*256*64	128*128*128	3	2	1	Elu
Conv_3	128*128*128	128*128*128	3	1	1	Elu
Down_Conv_4	128*128*128	64*64*256	3	2	1	Elu
Conv_5	64*64*256	64*64*256	3	1	1	Elu
Conv_6	64*64*256	64*64*256	3	1	1	Elu
Dilated_Conv_1	64*64*256	64*64*256	3	1	2	Elu
Dilated_Conv_2	64*64*256	64*64*256	3	1	4	Elu
Dilated_Conv_3	64*64*256	64*64*256	3	1	8	Elu
Dilated_Conv_4	64*64*256	64*64*256	3	1	16	Elu
Conv_7	64*64*256	64*64*256	3	1	1	Elu
Conv_8	64*64*256	64*64*256	3	1	1	Elu
DeConv_9	64*64*256	128*128*128	4	1/2	1	Elu
Conv_10	128*128*128	128*128*128	3	1	1	Elu
DeConv_11	128*128*128	256*256*64	4	1/2	1	Elu
Conv_12	256*256*64	256*256*32	3	1	1	Elu
Conv_13	256*256*32	256*256*3	3	1	1	Tanh

1.2 边缘修复网络

利用 canny 算子对粗修复网络 G_c 输出的结果 I_{pre}^C 进行边缘提取得到 B_{canny} ，将 B_{canny} 和自由掩模 M 输入到边缘修复网络 G_e 中。

边缘修复网络由 13 个普通卷积层、4 个 Res2net 层和 4 个自注意力机制层组成^[14]，其中普通卷积层的结构与粗修复网络相同。Res2net 层^[15-16]以分层的

方式表示多尺度特征，输出包含不同感受野大小的组合，该组合结构扩大感受野大小的同时，有效解决网格伪影现象；而自注意力机制通过自适应调节目标特征与周围特征相关性权重的方式加强特征之间的联系，以保证前景和背景的一致性。边缘修复网络结构见表 2 所示。

边缘修复网络 G_e 输出边缘预测结果为

$$B_{pre} = G_e(B_{canny}, M)。 \quad (2)$$

表2 边缘修复网络结构

Tab. 2 Edge inpainting network structure

每层名称	输入	输出	卷积核大小	步长	空洞率	激活函数
Conv_1	256*256*2	256*256*64	5	1	1	Elu
Down_Conv_2	256*256*64	128*128*128	3	2	1	Elu
Conv_3	128*128*128	128*128*128	3	1	1	Elu
Down_Conv_4	128*128*128	64*64*256	3	2	1	Elu
Conv_5	64*64*256	64*64*256	3	1	1	Elu
Conv_6	64*64*256	64*64*256	3	1	1	Elu
Res2net_1	64*64*256	64*64*256				
Res2net_2	64*64*256	64*64*256				
Res2net_3	64*64*256	64*64*256				
Res2net_4	64*64*256	64*64*256				
Conv_7	64*64*256	64*64*256	3	1	1	Elu
Self_attention_1	64*64*256	64*64*256				
Conv_8	64*64*256	64*64*256	3	1	1	Elu
Self_attention_2	64*64*256	64*64*256				
DeConv_9	64*64*256	128*128*128	3	1/2	1	Elu
Self_attention_3	128*128*128	128*128*128				
Conv_10	128*128*128	128*128*128	3	1	1	Elu
Self_attention_4	128*128*128	128*128*128				
DeConv_11	128*128*128	256*256*64	3	1/2	1	Elu
Conv_12	256*256*64	256*256*32	3	1	1	Elu
Conv_13	256*256*32	256*256*1	3	1	1	Tanh

1.3 精修复网络

精修复网络结构基于 U-net 架构^[17], 由 13 个普通卷积层、4 个 Res2net 层、4 个自注意力机制层组成, 通过跳跃连接将编码器特征直接传输到

解码器中。其中普通卷积层与粗修复网络中的结构相同, Res2net 层、自注意力机制层与边缘修复网络中的结构相同; 增加的跳跃连接结构将一些低维特征直接传递到后面的卷积层中, 减少了特征损失。精修复网络结构见表 3 所示。

表 3 精修复网络结构

Tab. 3 Refinement inpainting network structure

每层名称	输入	输出	卷积核大小	步长	空洞率	激活函数
Conv_1	256*256*5	256*256*64	5	1	1	Elu
Down_Conv_2	256*256*64	128*128*128	3	2	1	Elu
Conv_3	128*128*128	128*128*128	3	1	1	Elu
Down_Conv_4	128*128*128	64*64*256	3	2	1	Elu
Conv_5	64*64*128	64*64*256	3	1	1	Elu
Conv_6	64*64*256	64*64*256	3	1	1	Elu
Res2net_1	64*64*256	64*64*256				
Res2net_2	64*64*256	64*64*256				
Res2net_3	64*64*256	64*64*256				
Res2net_4	64*64*256	64*64*256				
Conv_7	64*64*256	64*64*256	3	1	1	Elu
(Conv_5+Conv_7)		64*64*512				
Self_attention_1	64*64*512	64*64*512				
Conv_8	64*64*512	64*64*256	3	1	1	Elu
(Conv_4+Conv_8)		64*64*512				
Self_attention_2	64*64*512	64*64*512				
DeConv_9	64*64*512	128*128*128	3	1/2	1	Elu
(Conv_3+Conv_9)		128*128*256				
Self_attention_3	128*128*256	128*128*256				
Conv_10	128*128*256	128*128*128	3	1	1	Elu
Self_attention_4	128*128*128	128*128*128				
DeConv_11	128*128*128	256*256*64	3	1/2	1	Elu
Conv_12	256*256*64	256*256*32	3	1	1	Elu
Conv_13	256*256*32	256*256*3	3	1	1	Tanh

将粗修复结果 I_{pre}^C 、边缘修复结果 B_{pre} 和自由掩模 M 共同输入到精修复网络中。精修复网络 G_r 的输入分别为

$$I_{input}^R = I_{pre}^C \otimes M + I_{gt} \otimes (1 - M), \quad (3)$$

$$B_{input}^R = B_{pre} \otimes M + B_{gt} \otimes (1 - M), \quad (4)$$

式中: $I_{pre}^C \otimes M$ 为粗修复网络生成的缺失区域; $I_{gt} \otimes (1 - M)$ 为图像中未破损的部分; $B_{pre} \otimes M$ 为边缘修复网络生成的缺失边缘部分; B_{gt} 为原始图像边缘; $B_{gt} \otimes (1 - M)$ 为图像中未破损的边缘区域。精修复网络 G_r 输出的精修复结果 I_{pre}^R 为

$$I_{pre}^R = G_r(I_{input}^R, B_{input}^R, M). \quad (5)$$

1.4 损失函数

粗修复网络 G_c 、边缘修复网络 G_e 和精修复网络 G_r 的损失函数分别为

$$l_1^c = \|I_{pre}^C - I_{gt}^C\|_1, \quad (6)$$

$$l_1^b = \|B_{pre} - B_{gt}\|_1, \quad (7)$$

$$l_1^r = \|I_{pre}^R - I_{gt}^R\|_1. \quad (8)$$

生成器对抗损失 l_{adv}^G 定义为

$$l_{adv}^G = -E_{(I_{pre})} [D(I_{pre})]. \quad (9)$$

联合损失函数定义为

$$l_{total} = \lambda_1^c l_1^c + \lambda_2^b l_1^b + \lambda_3^r l_1^r + \lambda_{adv} l_{adv}^G, \quad (10)$$

式中: λ_1^c , λ_2^b , λ_3^r , λ_{adv} 分别为 l_1^c , l_1^b , l_1^r 和 l_{adv}^G 损失控制参数, 取 $\lambda_1^c = \lambda_2^b = \lambda_3^r = \lambda_{adv} = 1$ 。

鉴别器将生成的图像以及原图作为输入, 并判断输入的生成图像是真是假。通过最大化鉴别器对抗损失 l_{adv}^D 训练网络, 其中鉴别器 D 对抗损失定义为

$$l_{adv}^D = E_{(I_{gt})} [\max(0, 1 - D(I_{gt}))] + E_{(I_{pre})} [\max(0, 1 + D(I_{pre}))]. \quad (11)$$

2 实验与分析

实验配置为：64位 Windows 10 操作系统，Nvidia GeForce RTX 2080Ti，实验框架采用 TensorFlow 深度学习框架，编程语言使用 python3.6。本文采用 Celeba-HQ 人脸数据集进行训练和测试，训练集包含 20 000 张 256*256 大小的人脸图像，测试集包含 1 000 张人脸图像。在训练阶段，本文算法的训练时间为 62 h。在测试阶段，共耗时 30 min 12 s 完成测试，平均每张图像需耗时 1.812 s 完成修复。

为了验证本文算法的有效性，将本文算法与

PatchMatch 算法^[2]、Deepfill-v1 算法^[8]、Edge-Connect 算法^[9]和 CTSDG 算法^[10]进行对比实验，以主观视觉对比和客观指标分析两种评价方法进行实验对比分析。

2.1 主观评价

主观实验对比如图 2 所示。其中第 1, 2 行表示破损程度为[10%, 20%)的实验对比图，第 3, 4 行表示破损程度为[20%, 30%)的实验对比图，第 5, 6 行表示破损程度为[30%, 40%)的实验对比图，第 7, 8 行表示破损程度为[40%, 50%)的实验对比图。

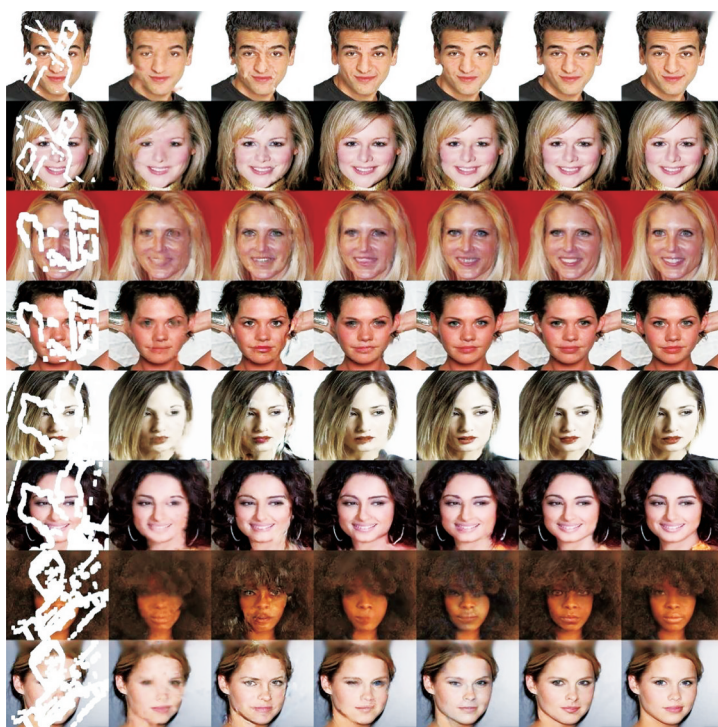


图 2 Celeba-HQ 测试集实验结果

Fig. 2 Celeba-HQ test set experimental results

图 2 中从左到右每一列分别表示原始破损图片，PatchMatch 算法、Deepfill-v1 算法、Edge-Connect 算法、CTSDG 算法，本文算法修复后的图片以及原始图像。

从修复对比图中第 2 列可以看出，PatchMatch 算法在处理小规模缺失图像时，具有良好的性能表现，但随着破损区域规模的增大，该算法会产生扭曲的结构和大面积的伪影，原因在于该算法仅能利用周围已知区域来填充未知区域，当缺失部分较大时，存在的已知信息不足以修复完整的图像。从修复对比图中第 3 列可以看出，Deepfill-v1 算法修复区域出现结构扭曲的现象，

原因在于 Deepfill-v1 算法只关注纹理修复，没有考虑结构的问题。从修复对比图中第 4 列第 3, 4 行可以看出，修复结果产生了模糊的眼眶结构，原因在于 Edge-Connect 算法会生成混合的边缘，通过使用边缘结构先验可能会导致错误的纹理结构。从修复对比图中的第 4 列第 7, 8 行可以看出，CTSDG 算法修复结果中脸颊部分明显存在色彩鲜艳度缺失的问题，原因在于 CTSDG 算法未注重缺失部分的色彩修复。相比于 PatchMatch 算法、Deepfill-v1 算法、Edge-Connect 算法和 CTSDG 算法，本文算法得到了更好的图像细节与边缘，更接近于真实图像。

2.2 客观评价

本文使用峰值信噪比 (Peak Signal to Noise Ratio, PSNR) 和结构相似性 (Structural Similarity, SSIM) 定量评估 5 种模型的修复结果。

PSNR 值越大, 代表图像的失真越少; SSIM 指标值越大, 代表和原始图像的差距越小, 即修复结果图像质量越好。在相同破损程度的情况下, 本文提出的修复算法的性能优于其他 4 种图像修复算法, 如表 4 所示。

表 4 各算法在 Celeba-HQ 测试集性能指标
Tab. 4 Celeba-HQ test set indicators of every algorithm

破损程度/%	SSIM				
	PatchMatch算法	Deepfill-v1算法	Edge-Connect算法	CTSDG算法	本文所提算法
[10,20)	0.865	0.901	0.935	0.939	0.940
[20,30)	0.801	0.845	0.893	0.904	0.915
[30,40)	0.766	0.811	0.858	0.877	0.903
[40,50)	0.741	0.744	0.820	0.827	0.836
破损程度/%	PSNR/%				
	PatchMatch算法	Deepfill-v1算法	Edge-Connect算法	CTSDG算法	本文所提算法
[10,20)	26.53	28.31	29.72	30.55	30.65
[20,30)	23.78	26.43	29.30	29.45	29.48
[30,40)	22.93	23.02	26.72	26.61	27.34
[40,50)	22.57	22.44	24.59	25.14	25.20

2.3 消融实验

为了说明本文所提网络各模块的有效性和必要性, 对其进行消融实验, 主要包括有无 Res2net 模块对比实验和有无自注意力机制对比实验。随机选取 Celeba-HQ 数据集中的两幅图进行消融实验对比, 并进行主观与客观指标评价。

降 2.45%; 在缺少自注意力机制情况下, PSNR 指标下降 3.32%, SSIM 指标下降 1.79%。由此证明算法中加入 Res2net 模块和自注意力机制会使修复图像的质量更好。

图 3 为消融实验的主观实验结果, 如图 3(b) 中所示, 在缺少 Res2net 模块的情况下, 恢复的人脸图像纹理、细节特征不清晰; 由图 3(c) 可知, 缺少自注意力机制将导致修复区域与已知区域缺乏一致性, 出现前景与背景不融洽的现象。

表 5 消融实验客观指标对比表

Tab. 5 Ablation experiment objective index comparison table

Res2net	✓	✗	✓
自注意力机制	✗	✓	✓
PSNR/%	24.48	24.39	25.20
SSIM	0.816	0.821	0.836

3 结束语

针对人脸图像的特征复杂多样, 细节和结构特征难以恢复等问题, 提出了一种基于上下文特征提取的边缘生成三阶段的图像修复算法。首先, 将待修复图像输入到粗修复网络中, 对图像缺损部分进行初步修复; 然后, 将初修复结果所提供的边缘结构作为边缘修复网络的输入, 完成边缘结构的细节重建; 最后, 将初修复结果和重建的边缘结构一起输入到精修复网络, 共同约束图像的精修复过程, 并在边缘修复网络和精修复网络中加入 Res2net 模块、自注意力机制, 增强了网络对于有限特征的提取能力, 改善了图像目标区域与周围区域边界不连贯的问题。通过与 4 种现有的修复方法在主观、客观方面的实验对比分析, 本文所提算法修复的人脸图像, 在



图 3 消融实现主观实验结果

Fig. 3 Subjective results of ablation experiments

表 5 数据表明本文算法在缺少 Res2net 模块情况下, PSNR 指标下降 2.94%, SSIM 指标下

SSIM和PSNR性能指标上优于现有算法,且在边缘细节的恢复中有着更好的视觉效果。

参考文献:

- [1] BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting [C]//Conference on Computer Graphics and Interactive Techniques, New Orleans, LA: ACM Press/Addison-Wesley Publishing Company, 2000: 417-424.
- [2] BARNES C, SHECHTMAN E, FINKELSTEIN A, et al. PatchMatch: a randomized correspondence algorithm for structural image editing[J]. ACM Transactions on Graphics, 2009, 28(3): 1-11.
- [3] CRIMINISI A, PATRICK P, TOYAMA K. Region filling and object removal by exemplar-based image inpainting[J]. IEEE Transactions on Image Processing, 2004, 13(9): 1200-1212.
- [4] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion[J]. ACM Transactions on Graphics, 2017, 36(4): 1-14.
- [5] LIU G L, REDA F A, SHIH K, et al. Image inpainting for irregular holes using partial convolutions [C]//European Conference on Computer Vision, Munich: ECCV, 2018: 89-105.
- [6] ZHOU X, LI J, WANG Z, et al. Image inpainting with contrastive relation network [C]//25th International Conference on Pattern Recognition (ICPR), 2021: 4420-4427.
- [7] PATHAK D, DONAHUE J, DARRELL T, et al. Context encoders: feature learning by inpainting [C]//IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas: IEEE, 2016: 2536-2544.
- [8] YU J, LIN Z, YANG J, et al. Generative image inpainting with contextual attention [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018: 5505-5514.
- [9] NAZERI K, NG E, JOSEPH T, et al. EdgeConnect: generative image inpainting with adversarial edge learning [C]// ICCV Workshops, Seoul: ICCV, 2019: 1-17.
- [10] GUO X, YANG H, HUANG D. Image inpainting via conditional texture and structure dual generation [C]//IEEE/CVF International Conference on Computer Vision (ICCV), 2021: 14114-14123.
- [11] 王贺, 韩磊. 基于Res2net和金字塔池化的图像去雾算法[J]. 测试技术学报, 2023, 37(5): 455-460.
- WANG He, HAN Lei. Image dehazing algorithm based on Res2net and pyramid pooling[J]. Journal of Test and Measurement Technology, 2023, 37(5): 455-460. (in Chinese)
- [12] GOODFELLOW I J, POUGETABADIE J, MIRZA M, et al. Generative adversarial nets[J]. Advances in Neural Information Processing Systems, 2014, 3(2): 2672-2680.
- [13] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks [C]// IEEE International Conference on Computer Vision (ICCV), IEEE, 2017: 1-9.
- [14] ZHANG H, GOODFELLOW I, METAXAS D, et al. Self-attention generative adversarial networks [C]//Proceedings of International Conference on Machine Learning, 2019, 97: 7354-7363.
- [15] GAO S, CHENG M M, ZHAO K, et al. Res2Net: a new multi-scale backbone architecture [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2019, 43(2): 652-662.
- [16] WANG Z, WU Y, XU J, et al. Res2Fusion: infrared and visible image fusion based on dense Res2Net and double nonlocal attention models[J]. IEEE Transactions on Instrumentation and Measurement. 2022, 71: 1-12.
- [17] RONNEBERGER O, FISCHER P, BROX T. U-net: convolutional networks for biomedical image segmentation [C]//International Conference on Medical Image Computing and Computer Assisted Intervention, 2015: 234-241.