

文章编号: 2095-2163(2022)06-0141-05

中图分类号: TP391

文献标志码: A

基于改进注意力机制的生成对抗网络图像修复研究

张剑飞, 张 洒, 夏万贵

(黑龙江科技大学 计算机与信息工程学院, 哈尔滨 150022)

摘要: 针对受损区域修复存在语义不连贯, 纹理不清晰的问题, 本文提出了一种基于改进注意力机制的生成对抗网络图像修复方法。用 U-Net 作为生成器主干, 为充分学习图像特征, 提高编码器中特征的利用率, 在编码阶段引入改进的通道注意力模块, 同时为了克服长距离对于信息的依赖, 在跳跃连接层中添加转移连接层, 保持图像信息的连贯性。此外在原本的重构损失函数和对抗损失函数中, 添加指导损失函数与风格损失函数, 增加了整个网络的稳定性。实验结果表明在 Celeb A 和 Places2 数据集上, 本文方法取得了较好的修复效果。

关键词: 图像修复; 通道注意力; 转移连接层; 生成对抗网络

Research on images restoration of generative adversarial network based on improved attention mechanism

ZHANG Jianfei, ZHANG Sa, XIA Wangui

(School of Computer & Information Engineering, Heilongjiang University of Science & Technology, Harbin 150022, China)

[Abstract] Aiming at the problems of semantic incoherence and unclear texture in damaged area repair, a generative adversarial network images restoration method based on improved attention mechanism is proposed in this paper. U-net is used as the backbone of the generator. In order to fully learn images features and improve the utilization of features in the encoder, an improved channel attention module is introduced in the coding stage. At the same time, for the purpose of overcoming the dependence on information over a long distance, a transfer connection layer is added to the jump connection layer to maintain the coherence of images information. In addition, in the original reconstruction loss function and countermeasure loss function, guidance loss function and style loss function are added to increase the stability of the whole network. The experimental results show that the proposed method has achieved good restoration results on Celeb A and Places2 data sets.

[Key words] images restoration; channel attention; transfer connection layer; generative adversarial network

0 引言

图像修复是利用破损图像和训练图像, 获取数据中的结构纹理信息, 将破损图像进行填充。近年来被广泛地应用于诸多专业技术领域, 目前已成为智能图像处理^[1-5]、神经网络^[6-10]和计算机视觉等方面的研究热点。传统图像修复方法主要有基于纹理和基于样本块, 经典模型有基于偏微分方程的 BSCB^[11-12]模型、填充的 Criminisi^[13]算法模型等。但这些方法大多存在缺少高级语义信息且与原图结构相似度不高等问题。近年来, 深度学习的图像修复方法取得了许多突破性进展, Goodfellow 的上下文自动编码器(Content Encoder, CE)^[14], 首次将生成对抗网络^[15](Generative Adversarial Networks, GAN) 应用于图像修复, 对图像上下文语义信息做出预测, 但修复的图像存在明显伪影。为了更好地

获取高级语义信息, 在全卷积神经网络^[14](Fully Convolutional Network, FCN) 式的特征值逐点相加的基础上, 出现了使用 U-Net^[16]网络, Yan 等人^[17]在 U-Net 的解码器中添加了一个转移连接层, 即网络为 Shift-Net, 有效地结合图像中相隔较远的特征进行图像修复, 但当破损孔洞过大时, 存在细节纹理不清晰的问题。Hu 等人^[18]提出了挤压和激励网络(Squeeze-and-Excitation Network, SENet), 扩大感受野, 将重要的特征进行强化来提高准确率权重。

针对目前图像修复中存在语义不连贯、纹理不清晰的现象, 本文构建了一个以 U-Net 为基础模型添加转移连接层和改进的通道注意力进行精细修复的图像修复方法, 旨在获得具有高级语义和清晰纹理的修复图像。

基金项目: 国家自然科学基金(61803148)。

作者简介: 张剑飞(1978-), 女, 博士, 副教授, 主要研究方向: 人工智能、计算机视觉; 张 洒(1993-), 女, 硕士研究生, 主要研究方向: 计算机视觉; 夏万贵(1993-), 男, 硕士研究生, 主要研究方向: 人工智能。

收稿日期: 2021-12-20

1 基于改进注意力机制的生成对抗网络图像修复

注意力机制分为空间注意力^[19] (Spatial attention, SA) 和通道注意力^[20] (Channel attention, CA), 2种注意力机制对于资源分配的级别不一样。SA 定位感兴趣区域进行变换获取权重, 而 CA 则是在于分配各个卷积通道之间的资源, 两者对于不同部分均有侧重, 为了结合两者优点, 本文使用了空间注意力与通道注意力结合的注意力机制, 空间上采用转移连接层, 改变了原本修复只能从破损图片周围点像素进行补充, 跨越空间限制, 寻找与之最相似的点。通道上采用改进的 SE block, 首先通过对资源分配不同比重, 然后专注于图像的待修复区域, 借助于通道和空间注意力的结合, 更好获取图像全局和局部特征。

1.1 模型框架

为了使图像修复结果具有更好的语义表达和更

精细的结构纹理, 本文提出了基于改进注意力机制的生成对抗网络两阶段图像修复方法。网络整体采用2阶段修复, 生成器以 U-Net 网络架构为基础, 编码器、解码器都采用步长为 2、4×4 的卷积。为充分学习图像特征, 提高编码器中特征的利用率, 在编码阶段引入多尺度卷积与通道注意力结合, 进行通道特征权重的重新标定, 同时为了克服长距离对于信息的依赖, 将第 i 层和第 $L-i$ 层之间通过跳跃连接后再传递给下一层, 对图像进行空间特征的重排, 保持图像信息的连贯性。至此得到修复粗糙图像, 同时计算重建损失函数和指导损失函数。将粗糙网络修复的图像和真实图片输入 VGG16 网络中进行特征提取分析, 然后通过鉴别器 DCGAN 鉴定图像的真假, 若为假, 给生成器反馈重新进行图像修复, 通过设置学习率、迭代次数和损失函数来约束生成器不断重复进行图像学习, 直至鉴别器无法确认生成器输入图像的真假, 即完成了图像修复过程。本文的网络模型框架如图 1 所示。

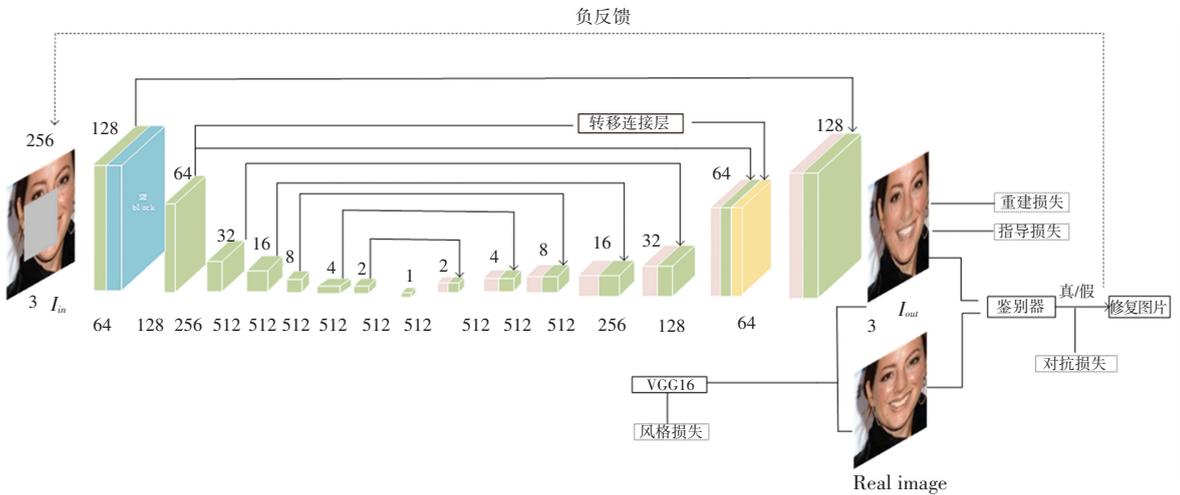


图 1 本文的网络模型框架

Fig. 1 The network model framework of this paper

1.2 改进 SE block 模型

为了增加在不同尺度卷积得到的特征, 同时不增加参数数量, 减少计算时间, 所以在原来 SE block 的基础上同时进行卷积和扩张卷积, 用大小为 3、5、7 的卷积核, 但使用 2 个 3×3 的卷积核来模拟 5×5 的卷积核, 用 3 个 3×3 的卷积核来模拟 7×7 的卷积核。为了扩大感受野, 增加特征图的均衡性, 采用膨胀卷积, 卷积采用膨胀系数为 2、步长为 2、同样用 2 个 3×3 的膨胀卷积来模拟 5×5 的膨胀卷积, 2 个 3×3 的膨胀卷积来模拟 7×7 的膨胀卷积, 对于不同尺度卷积结果级连。同时为了增加特征的全局和局部

一致性, 对图像进行 3×3、5×5、7×7 的卷积, 这里的 3×3、5×5、7×7 的膨胀卷积, 通过 $eRelu$ 函数分别得到不同局部之间的关系为: $R_1, R_2, R_3, R_4, R_5, R_6$ 。与此同时, 为了得到图像特征之间更加均衡有效的依赖关系, 通过左侧基础 SE block 进行通道注意力操作, 并从全局池化 Global Average Pooling 中得到通道特征 Z_c , Squeeze 过程可以表示为:

$$Z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W X_c(i, j) \quad (1)$$

通过上述过程得到新特征图 F_{ex} , 则此 Excitation 过程可以表示为:

$$S = F_{ex}(Z, W) = \sigma(g(Z, W)) \quad (2)$$

其中, $g(\cdot)$ 指 $eRelu$ 激活函数, $\sigma(\cdot)$ 指 $Sigmoid$ 激活函数。 S 与右侧多尺度卷积结合得到关系 R_7 , 此过程可以表示为:

$$R_7 = S \otimes (((cont(R_1, R_2, R_3)cont(R_4, R_5, R_6)))) \quad (3)$$

其中, “ \otimes ” 是张量积, $cont(\cdot)$ 为合并连接。融合的特征图 R_7 进行 $Scale$ 运算, 特征权重进行重新标定, 最终完成了通道资源的分配, 通过跳跃连接将原始图和 R_7 进行连接, 则图像 X_1 可以表示为:

$$X_1 = X \oplus Sc(R_7) \quad (4)$$

其中, Sc 是图像的比例缩放运算, 即 $Scale$, “ \oplus ” 是通道连接。至此, 通过对通道上像素点的权重重新标定和不同尺度得到特征结合, 完成这一阶段图像修复得到 X_1 。原始的 SE block 使用 $ReLU$ 作为激活函数, 但是当输入值为负值时, 会导致神经元不再学习, 且训练速度较慢, 因此, 在改进的结构中使用 $eRelu$ 作为激活函数。改进 SE block 的模型图, 如图 2 所示。

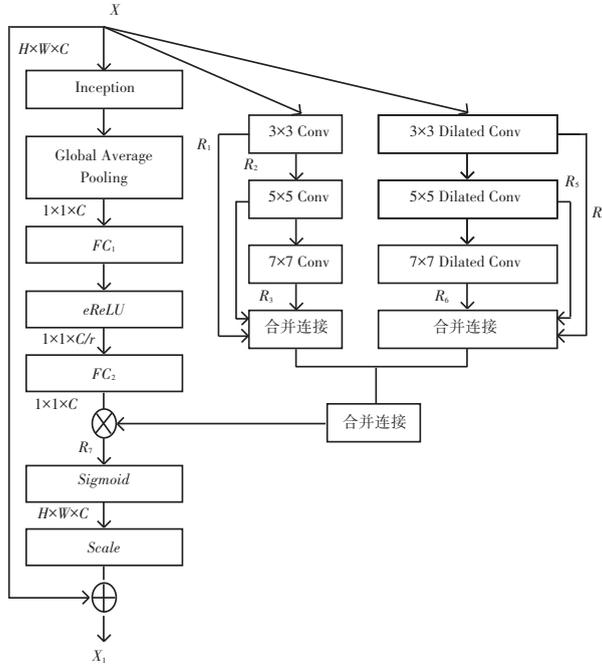


图 2 改进的 SE block 模型图

Fig. 2 Improved SE block model

1.3 改进损失函数

为了使修复后图像与原图像在风格上统一, 纹理细节更清晰, 在原本损失函数的基础上加入指导损失函数和风格损失函数。图像修复的基本损失函数有重建损失函数和对抗损失函数。这里的重建损失函数表示为:

$$L_{l1} = \|\Phi(I; W) - I^{gt}\|_1 \quad (5)$$

其中, Φ 是特征向量; I 是网络的第 i 层; W 是需要学习的模型参数; I^{gt} 是真实图像。

对抗损失函数表示为:

$$L_{adv} = \min_w \max_D E_{I^{gt} \sim p_{data}(I^{gt})} [\log D(I^{gt})] + E_{I \sim p_{miss}(I^{gt})} [\log(1 - D(\Phi(I; W)))] \quad (6)$$

其中, G 是生成器; D 是鉴别器; I^{gt} 指真实图像; Φ 是特征向量; I 是网络的第 i 层; W 是需要学习

的模型参数; $p_{data}(I^{gt})$ 是真实图像的分布; $p_{miss}(I^{gt})$ 是破损图像的分布。由于引入了转移连接层, 则加入指导损失函数, 指导损失函数^[8]表示为:

$$L_g = \sum_{y \in \Omega} \|(\Phi_{L-l}(I))_y - (\Phi_l(I^{gt}))_y\|_2^2 \quad (7)$$

其中, y 是缺失区域; Ω 是图像全部区域; Φ_l 是 l 层特征图; Φ_{L-l} 是 $L-l$ 层特征图; I 是破损图片; I^{gt} 是真实图像。进行图像训练时需要寻找某类图片的风格, 便于图像修复。

进一步地, 研究推得风格损失函数可写为:

$$L_{style} = E_j [\|G_j^\Phi(I^{out}) - G_j^\Phi(I^{gt})\|_1] \quad (8)$$

其中, G_j^Φ 是 $C_j \times C_j$ Gram 构造的矩阵图; I^{gt} 是真实图像; I^{out} 是修复后图像。

最终将多种损失函数结合起来作为整体损失函数, 定义为:

$$L_{total} = \lambda_{rec} L_{l1} + \lambda_{adv} L_{adv} + \lambda_g L_g + \lambda_{sty} L_{style} \quad (9)$$

其中, $l1$ 为重构损失; L_{adv} 为对抗损失; L_g 为指导损失; L_{style} 为风格损失; λ_{rec} , λ_{adv} , λ_g , λ_{sty} 分别为各损失函数的参数; 参数初始化时, $\lambda_{rec} = 1$, $\lambda_{adv} = 0.002$, $\lambda_g = 0.01$, $\lambda_{sty} = 10$ 。

2 实验结果与分析

2.1 数据集及环境配置

本文采用了2种国际标准数据集,即 Celeb A 人脸数据集和 Places2 场景数据集。其中, Celeb A 的每张人脸数据都进行了特征标注, Places2 数据集中含有400多个场景,可以满足对于数据多样性的需求。本文的运行设备为:中央处理器是 Intel 10th i7,显卡是 GPU GeForce GTX 1650Ti。文中的运行环境使用 Pytorch+Tensflow1.4 框架,搭配 Python 的多个库进行实验验证与分析。

数据的预处理过程:任意尺寸的图像输入,经过预处理,将图像裁剪成分辨率为 256×256 的统一图像大小。将输入图像分别与掩码和随机掩码进行结合,模拟破损图像。

2.2 评价指标

图像修复处理的过程中,评价图像修复质量有2个指标:峰值信噪比^[21] (Peak Signal to Noise Ratio, $PSNR$) 和结构相似性^[22] (Structural Similarity, $SSIM$)。其中, $PSNR$ 通常用于描述各种形式的信号的质量属性, $PSNR$ 值越大,说明图像信号质量越高。 $PSNR$ 可由如下公式计算求得:

$$PSNR = 10 \log \left(\frac{i_{max}^2}{MSE} \right) = 20 \log \left(\frac{i_{max}}{\sqrt{MSE}} \right) \quad (10)$$

其中, i_{max} 指图像点颜色的最大值, MSE 为均方差。设有2幅 $m \times n$ 的单色图像 I, K , 这里对 MSE 的数学定义可以表示为:

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [I(i, j) - K(i, j)]^2 \quad (11)$$

其中, $I(i, j)$ 是真实图像的像素点, $K(i, j)$ 是通过修复得到图像的像素点。

结构相似性 ($SSIM$) 是一种衡量2幅图像相似度的指标,主要通过亮度、对比度和结构三方面来度量图像之间的相似性。结构相似性的范围为 $[0, 1]$, 当2幅图像一模一样时, $SSIM$ 的值为1, 即当 $SSIM$ 值越大, 图像相似性越高。研究推得 $SSIM$ 的数学计算公式为:

$$SSIM(x, y) = \frac{(2\mu_x \mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (12)$$

其中, x, y 分别为真实图像和重建图像; μ_x 为真实图像的均值; μ_y 为重建图像的均值; σ_x, σ_y 分别为图像 x, y 的标准差; σ_x^2, σ_y^2 分别为图像 x, y 的方差; σ_{xy} 为 x, y 的协方差; c_1, c_2 为常数, 为了避免分母为零的情况, c_1, c_2 通常取值为 $0.01^2, 0.03^2$ 。

2.3 实验结果

采用2种不同的数据集与中心掩码模拟受损图像,通过对原始图像 GT、破损图像、CE 修复图像、Shift-Net 修复图像和本文提出方法的修复图像的有效性进行对比,挑选有代表性的实验结果加以对比说明,实验结果如图3所示。

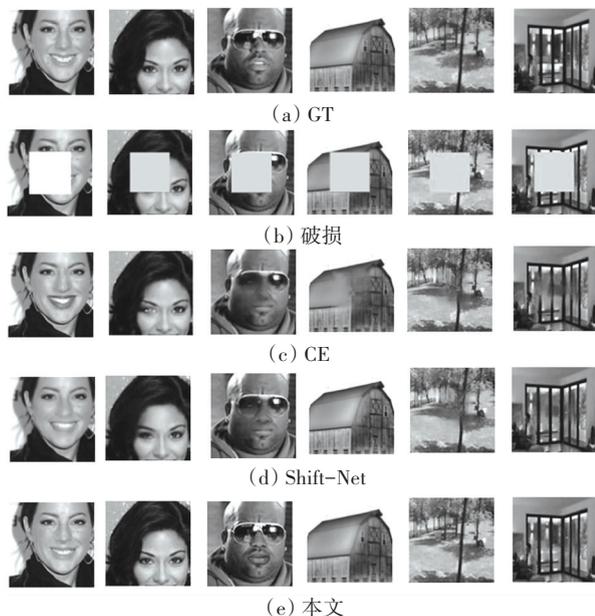


图3 实验结果图

Fig. 3 Experimental results

图3(c)中 CE 使用上下文自动编码器的生成对抗网络,对于大面积缺失图像,修复结果容易出现伪影,不能产生复杂的纹理结构,对于图像在有些地方有严重失真现象,如图3(b)所示的第一幅图像,左、右2只眼睛瞳孔颜色不一致。观察图3(c)第一幅图像发现,出现了明显的伪影、以及鼻子的失真。观察图3(b)的第四幅图像可知,有明显的伪影以及缺少细节纹理。图3(d)中,Shift-Net 采用了转移连接层,通过跳跃连接,使特征图具有良好的全局性,修复图片看起来更加真实,但是缺少对于细节处理、过于简单,局部有小范围的伪影,如图3(d)所示的第一幅图像,鼻子处有小部分伪影。另见图3(c)的第四幅图像可知,整体颜色一致,但是缺少一些内容语义,处理简单,修复部分缺少与右侧对称的窗户。本文添加了转移连接层和改进的通道注意力,对于

特征权重重新分配,得到图像在语义和纹理细节上优于其他2种方法。表1则为在2种不同数据集中,采用中心掩码的图像修复,利用评价指标对结果进行定量分析。由表1可知,本文在像素与结构上优于其他方法。

表1 采用中心掩码不同算法的比较

Tab. 1 Comparison of different algorithms using center mask

模型	PSNR/ db		SSIM	
	Cele A	Places2	Cele A	Places2
CE	24.16	24.73	0.84	0.87
Shift-Net	26.51	26.84	0.88	0.91
本文	27.84	28.16	0.89	0.93

3 结束语

本文以生成对抗网络为基础框架,对模型进行改进,生成器采用具有对称性的U-Net,在此基础上对于目前图像修复中存在语义不连贯、纹理不清晰的问题,引入改进的通道注意力,采用了多尺度和更大感受野与SE-block结合的方法,进行通道特征调整,重定特征权重;其次添加转移连接层,借助于U-Net的跳跃连接确定破损区域的最相似点进行修复;最后在损失函数上增加了指导损失函数和风格损失函数,加强已知区域与破损区域之间的约束关系。通过实验结果可知,对于语义不连贯、纹理不清晰的破损图像,修复取得了较好的效果。

参考文献

[1] 马世斌,杨文芳,张焜. SPOT6 卫星图像处理关键技术研究[J]. 国土资源遥感,2015,27(03):30-35.

[2] 刘李漫,谭龙雨. 数字图像处理中的医学图像处理应用[J]. 教育现代化,2019,6(85):214-216.

[3] 宋孝忠,张群. 煤岩显微组分组图像自动化识别系统与关键技术[J]. 煤炭学报,2019,44(10):3085-3097.

[4] 高慧琳. 基于卷积神经网络的军事图像分类[J]. 计算机应用研究,2017,34(11):3518-3520.

[5] 段乃侠. 基于多媒体图像技术的招贴信息自动分类方法研究[J]. 自动化与仪器仪表,2019(07):18-21,26.

[6] SCHNIEDER B. Image inpainting[J]. SIGGRAPH, 2000, 7:417.

[7] BERTALMIO M, SAPIRO G, CASELLES V, et al. Image inpainting[C]// Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Technique. [S.l.]: ACM, 2000:417-424.

[8] BENGIO Y. Learning deep architectures for AI[J]. Foundations and Trends in machine Learning, 2009, 2(1):1-217.

[9] HOCHREITRE S, SCHMIDHUBER J. Long short-term memory

[J]. Neural Computation, 1997,9(8):1735-1780.

[10] LECUN Y, BENGIO Y, HINTON G. Deep learning[J]. Nature, 2015, 521(7553):436-444.

[11] BERTALMIO M, BERTOZZI A L, SAPIRO G. Navier-stokes, fluid dynamics, and image and video inpainting [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Kauai, HI; IEEE, 2001:355-362.

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

[13] PATHAK D, KRHENBUHL P, DONAHUE J, et al. Context encoders: Feature learning by inpainting [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA;IEEE, 2016:2536-2544.

[14] SHELHAMER E, LONG J, DARRELL T. Fully convolutional networks for semantic segmentation [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017,39(4):640-651.

[15] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial nets [C] //Proceedings of the 27th International Conference on Neural Information Processing Systems. Montreal, Canada;Springer,2014, 2, 2672-2680.

[16] RONNEBERGER O, FISCHER P, BROX T. U-Net: Convolutional networks for biomedical image segmentation [C]// Proceedings of the Medical Image Computing and Computer-Assisted Intervention. Heidelberg;Springer,2015: 234-241.

[17] YAN Zhaoyi, LI Xiaoming, LI Mu, et al. Shift-net: Image inpainting via deep feature rearrangement [M]// FERRARI V, HEBERT M, SMINCHISESCU C, et al. Computer Vision - ECCV 2018. ECCV 2018. Lecture Notes in Computer Science. Cham;Springer, 2018, 11218: 3-19.

[18] HU Jie, SHEN Li, ALBANIE S, et al. Squeeze-and-excitation networks[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020,42(8):2011-2023.

[19] JADERBERG M, SIMONYAN K, ZISSERMAN A, et al. Spatial transformer networks [C]//Proceedings of the Advance in Neural Information Processing Systems. Cambridge: MIT Press, 2015: 2017-2025.

[20] ZHANG Yulun, LI Kunpeng, LI Kai, et al. Image super-resolution using very deep residual channel attention networks [C] //Proceedings of the 2018 European Conference on Computer Vision, LNCS 11211. Berlin: Springer. 2018: 294-310.

[21] 佟雨冰,张其善,祁云平. 基于PSNR与SSIM联合的图像质量评价模型[J]. 中国图像图形学报,2006(12):1758-1763.

[22] HORÉ A, ZIOU D. Image quality metrics: PSNR vs. SSIM [C]//2010 20th International Conference on Pattern Recognition. Istanbul, Turkey;IEEE,2010:2366-2369.

[23] SIMONYAN K, ZISSERMAN A. Very deep convolutional networks for large-scale image recognition [C]// 3rd International Conference on Learning Representations, ICLR 2015. San Diego: dblp, 2015: 1-14.

[24] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with Deep Convolutional Generative Adversarial Networks [J]. arXiv preprint arXiv:1511.06434,2015.