

## 基于内容风格迁移的残缺稀疏文字图像盲修复方法

刘栋斌<sup>1</sup>, 王慧琴<sup>1\*</sup>, 王可<sup>1</sup>, 王展<sup>2</sup>, 甄刚<sup>2</sup><sup>1</sup>西安建筑科技大学信息与控制工程学院, 陕西 西安 710055;<sup>2</sup>陕西省文物保护研究院, 陕西 西安 710075

**摘要** 现有残缺文字图像的修复需要确定Mask区域后对其填补, 如果文字残缺部分剩余信息量过于稀疏, 将无法确定Mask区域。针对该问题, 提出了一种基于内容风格迁移的残缺稀疏文字图像盲修复方法。利用循环生成对抗网络构建修复前后文字图像间的全局关联像素信息, 将残缺文字的图像内容风格特征迁移为完整文字图像从而进行修复; 并在网络中加入自注意力机制对稀疏像素进行全局约束, 解决迁移过程中相隔较远文字稀疏像素之间依赖关系较弱的问题; 同时在自注意力机制中使用最大池化, 提高迁移修复后的文字图像纹理特征; 使用最小二乘损失替换原网络模型中的sigmoid交叉熵损失函数, 提高迁移精度。实验结果表明, 所提方法不借助Mask指导, 能够盲修复稀疏性残缺文字图像中的随机未知缺失区域。

**关键词** 成像系统; 盲修复; 循环生成对抗网络; 自注意力模块; 最小二乘损失

中图分类号 TP391 文献标志码 A

DOI: 10.3788/LOP202259.2411001

## Blind Restoration Method for Incomplete and Sparse Text Images Based on Content Style Transfer

Liu Dongbin<sup>1</sup>, Wang Huiqin<sup>1\*</sup>, Wang Ke<sup>1</sup>, Wang Zhan<sup>2</sup>, Zhen Gang<sup>2</sup><sup>1</sup>School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, Shaanxi, China;<sup>2</sup>Shaanxi Provincial Institute of Cultural Relics Protection, Xi'an 710075, Shaanxi, China

**Abstract** The existing incomplete text image restoration must identify and fill the Mask region. The Mask region cannot be determined if the residual information of the incomplete text part is too sparse. A blind restoration method of incomplete and sparse text images based on content style transfer is proposed to address this issue. Using a cyclic generative adversarial network to construct the global related pixel information between the text images before and after restoration, the image content style features of the incomplete text were transferred to complete text images for restoration. The selfattention mechanism is added to the network to globally restrict the sparse pixels, thereby resolving the issue of weak dependence between the text sparse pixels far away in the migration process. simultaneously, the maximum pooling is used in the selfattention mechanism to enhance the texture features of the text images after migration and restoration. To improve migration accuracy, the least square loss function is used to replace the sigmoid cross-entropy loss function in the original network model. The proposed algorithm can repair random unknown missing regions in sparse incomplete text images without the assistance of Mask, according to experimental results.

**Key words** imaging systems; blind repair; cyclically generated adversarial network; selfattention module; least square loss

## 1 引言

文字是一个民族文化的象征, 记录着古人的智慧与弥足珍贵的历史信息。随着历史的更迭变迁, 各种

被书于石碑、简牍、甲骨、铜器、纸书上的文字, 由于自然侵蚀、人为破坏等原因而残缺不全, 许多重要的文物信息无法得以保留。因此, 文字修复对于文化的发扬及传承都具有很高的价值。传统的修复方法依赖专家

收稿日期: 2021-09-01; 修回日期: 2021-10-08; 录用日期: 2021-10-27

基金项目: 陕西省自然科学基金基础研究计划项目(2021JM-377)、陕西省科技厅科技合作项目(2020KW-012)、陕西省教育厅智库项目(18JT006)、西安市科技局高校人才服务企业项目(GXYD10.1)

通信作者: hqwang@xauat.edu.cn

经验,修复耗时耗力,并且不同专家的修复结果差异较大<sup>[1]</sup>。现代修复方法对文字图像进行数字化采集,使用数字图像处理或基于深度学习的技术对文字进行虚拟修复。张伟等<sup>[2]</sup>对文字图像灰度化后使用 Canny 算子进行文字边缘提取,然后结合人机修复处理,解决了修复人员仅依靠文字先验知识对文字图像进行修复的问题,但如果字迹模糊,Canny 算子边缘检测、修复效果不是很理想。张商珉等<sup>[3]</sup>对侯马盟书中碑文文字颜色进行迁移并细化,去除毛刺均匀膨胀修复,该方法能够处理部分笔画粘连的字体,但对于残断明显的字体,不能达到理想的修复效果。Isola 等<sup>[4]</sup>基于生成对抗网络(GAN)<sup>[5]</sup>提出了一种基于图像内容风格迁移的网络模型 Pix2Pix。Tian 等<sup>[6]</sup>在 Pix2Pix 的基础上构建了 Zi2Zi 文字转换网络模型,将图像迁移应用在汉字字体转换中。任建等<sup>[7]</sup>将基于 Pix2Pix 的网络框架字体转换模型应用到石刻文字虚拟修复中,但在实际修复应用中, Pix2Pix 训练所需要的成对数据集难以获得。陈善雄等<sup>[8]</sup>提出了一种基于双判别器 GAN 的方法,该方法通过训练获得古彝文字生成器,并优化双判别器模型,对古彝文字进行修复。但该方法需要确定修复部分并添加 Mask,如果破损区域残缺像素稀疏,无法确定用于指导破损区域的

Mask 位置,则语义修复效果较差。

残缺稀疏文字图像由于文字结构上的残缺性和像素上的稀疏性,无法确定 Mask 区域来指导残缺区域完成修复。本文提出了一种基于内容风格迁移的残缺稀疏文字图像盲修复方法。使用循环 GAN 对文字图像进行内容风格迁移盲修复,输入为残缺文字图像,输出为修复后的图像,并通过改进的自注意力机制,优化损失函数提高文字修复质量。

## 2 生成对抗网络

传统图像修复需要关联像素间局部信息,依据破损区域的大小,对破损区域进行扩散修复或者块填补<sup>[9]</sup>。盲修复<sup>[10]</sup>通过构建残缺区域与完整图像间全局关联信息来进行修复, GAN 是利用像素间全局关联信息生成图像的创新性方法。

GAN 的思想源于博弈论中的零和博弈,该网络由一个生成器  $G$  和一个判别器  $D$  构成。生成器通过随机噪声  $z$  生成图像  $G(z)$ , 判别器则判断生成图像  $G(z)$  与真实图像  $x$  的差异,并将结果进行反馈。在对抗训练之中,生成器根据判别器的反馈进行修正,生成的图像逐渐接近真实图像,判别器的鉴别能力也逐渐增强,最终达到纳什均衡<sup>[11]</sup>。GAN 的原理如图 1 所示:

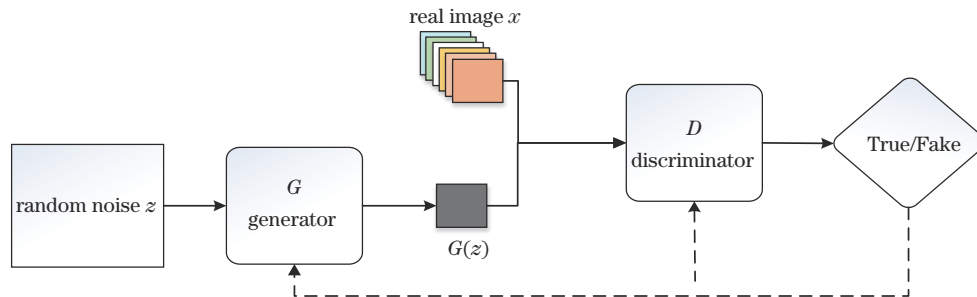


图 1 GAN 原理图

Fig. 1 Schematic of GAN

GAN 的目标函数为

$$\min_G \max_D V(D, G) = E_{x \sim P_{\text{data}}(x)} [\ln D(x)] + E_{z \sim P_z(z)} \left\{ \ln \{1 - D[G(z)]\} \right\}, \quad (1)$$

式中:  $x$  为输入真实图像;  $z$  为随机噪声;  $G(z)$  为生成图像;  $E(\cdot)$  为分布函数的期望值;  $P_{\text{data}}(x)$  为真实图像分布;  $P_z(z)$  为生成图像分布;  $D(x)$  为判别器输出, 是一个介于 0~1 之间的实数, 表示输出图片是真实图片的概率。

对于生成器, 期望生成图像为真, 即最大化  $D[G(z)]$  概率的同时最小化  $\ln \{1 - D[G(z)]\}$ , 目标函数值最小; 对于判别器, 期望生成图像为假, 即最小化  $D[G(z)]$  概率, 目标函数值最大。最终, 训练生成器模型和判别器模型达到纳什均衡, 达到生成图像与真实图像难以区分的目的。

## 3 基于内容风格迁移的残缺稀疏文字图像盲修复

GAN 产生随机噪声与真实样本间全局关联, 但无法进行图像之间内容风格迁移转换。循环 GAN (CycleGAN) 对 GAN 结构进行了突破性创新, 借鉴对偶学习在机器翻译中的具体应用, 使用循环一致性思想实现数据间的循环映射。本实验组在循环的 GAN 基础上, 提出了基于内容风格迁移的残缺稀疏文字图像盲修复算法。该算法在修复前后文字图像间产生全局像素关联, 实现残缺文字图像与完整文字图像之间的内容风格迁移, 将破损文字图像迁移修复完整。

### 3.1 整体设计

所提算法是在循环生成对抗网络的基础上改进而来的, 可不借助 Mask 指导对残缺稀疏文字图像进行全局盲修复。所提算法结构主要由两部分构成, 前向网

络和后向网络之间镜像对称构成环形循环网络,如图 2 所示。前向网络主要包括生成器  $G$ 、 $F$  及判别器  $D_A$ ,后向网络主要包括生成器  $F$ 、 $G$  及判别器  $D_B$ 。在残缺文字图像  $A$  迁移修复为  $B$  的过程中,正向生成器  $G$  和逆向生成器  $F$  进行训练:训练学习正向映射  $G(A \rightarrow B)$ ,使得生成样本的分布  $\text{Fake}_B = G(A)$  尽可能与  $\text{Real}_B$  的分布保持一致即  $\text{Fake}_B \approx \text{Real}_B$ ;训

练学习逆向映射  $F(\text{Fake}_B \rightarrow \text{Predict}_A)$ ,使得  $\text{Predict}_A = F(\text{Fake}_B)$  尽可能与  $A$  的分布保持一致,即  $\text{Predict}_A \approx A$ 。同时引入判别器  $D_A$  来判定生成图像  $\text{Fake}_B$  的所属类别。 $B \rightarrow \text{Fake}_A \rightarrow \text{Predict}_B \approx B$  的内容风格转换过程同理。为增强对残缺文字稀疏像素间的约束,在生成器和判别器中加入改进的自注意力(ISA)对网络进行优化,提高文字迁移修复效果。

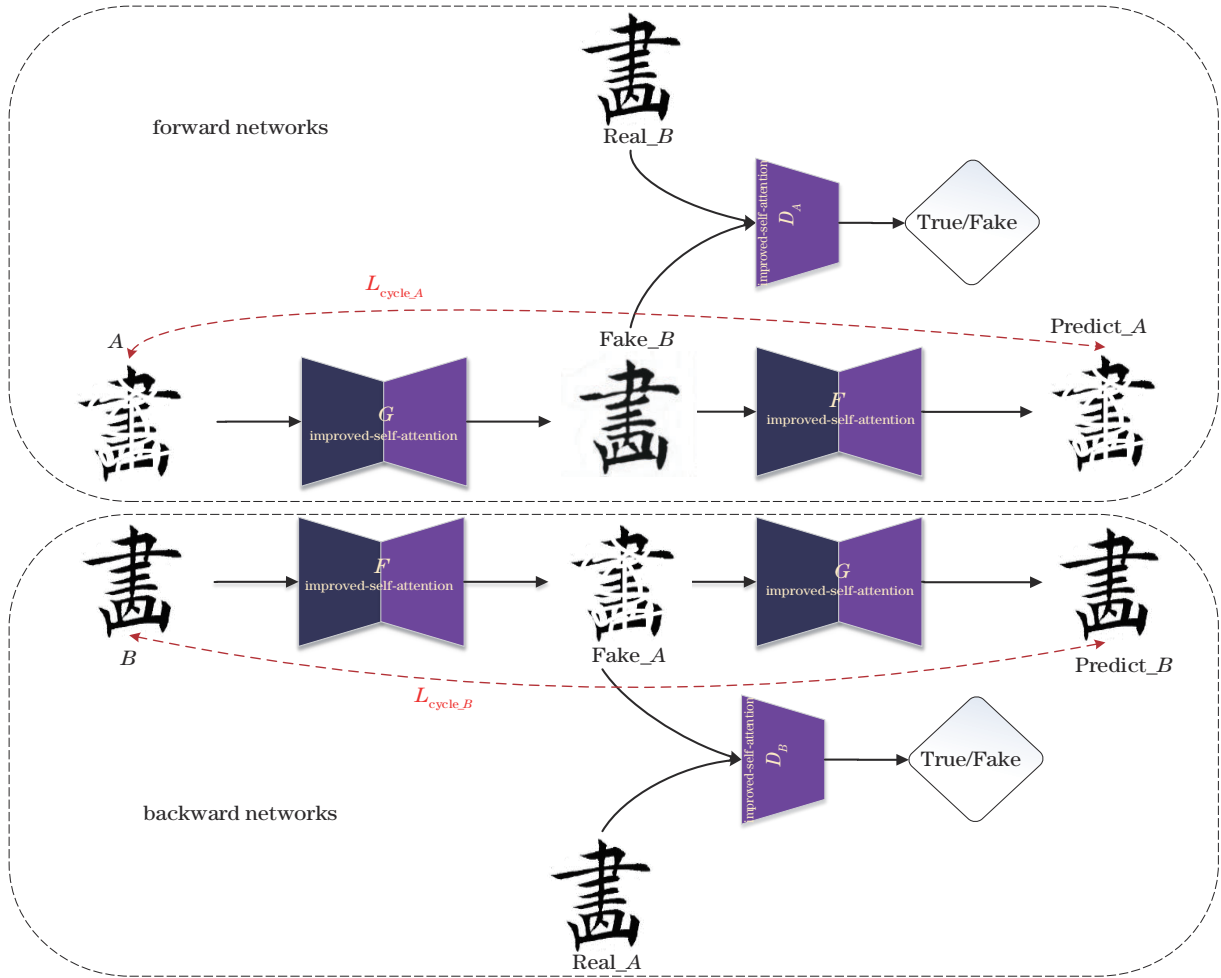


图 2 残缺稀疏文字图像盲修复网络结构图

Fig. 2 Structure diagram of blind restoration network for incomplete sparse text image

### 3.2 改进的自注意力机制

注意力机制源于对人类视觉注意力机制的研究,人在观察东西时注重观察事物感兴趣的区域,人类视觉注意力机制是一种综合内部经验和外部感觉的机制,这种机制使人眼对部分区域的观察更精细。自注意力机制对注意力机制进行了改进,更擅长捕捉事物数据或特征的内部相关性<sup>[12-13]</sup>。

文字特征图提取中卷积层参数误差会造成估计均值偏移,最大值池化可以减小这种均值偏移,提高内容风格迁移修复后的文字图像纹理信息。使用池化深度为 2、步长为 2 的最大值池化对自注意力机制进行改进,如图 3 所示,加入最大值池化后的改进自注意力机制如图 4 所示。

残缺文字图像经过编码,特征图  $\mathbf{x} \in \mathbf{R}^{C \times N}$  分别通过  $1 \times 1$  的卷积被转换到  $f$  和  $g$  特征空间中,计算输入特征图的注意力:

$$\begin{cases} \beta_{j,i} = \frac{\exp(s_{i,j})}{\sum_{i=1}^N \exp(s_{i,j})}, \\ s_{i,j} = f(\mathbf{x}_i)^T g(\mathbf{x}_j) \end{cases} \quad (2)$$

式中:  $f(\mathbf{x}) = \mathbf{W}_f \mathbf{x}$ 、 $g(\mathbf{x}) = \mathbf{W}_g \mathbf{x}$ , 权重学习矩阵  $\mathbf{W}_f \in \mathbf{R}^{\bar{C} \times C}$ 、 $\mathbf{W}_g \in \mathbf{R}^{\bar{C} \times C}$ ,  $C$  为特征图的通道数,  $\bar{C} = C/8$ ;  $N = H \times W$ ,  $H$ 、 $W$  分别为特征图的高和宽;  $\beta_{j,i}$  为图像中第  $i$  个区域对第  $j$  个区域的影响, 2 个位置的特征表示越相似,说明它们之间的相关性越强。

将  $f$  和  $g$  空间计算得到的文字图像注意力图  $\beta$  与

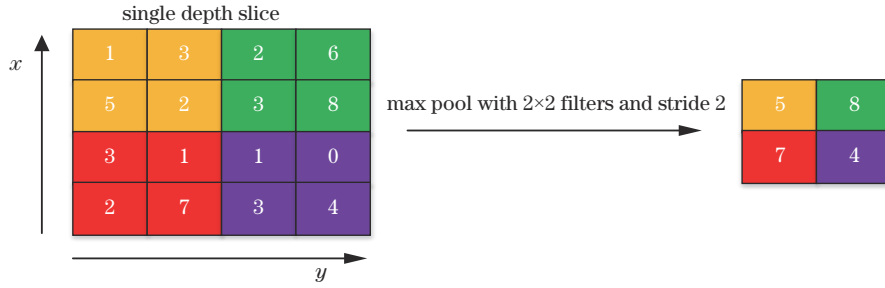


图 3 最大池化图

Fig. 3 Max pooling diagram

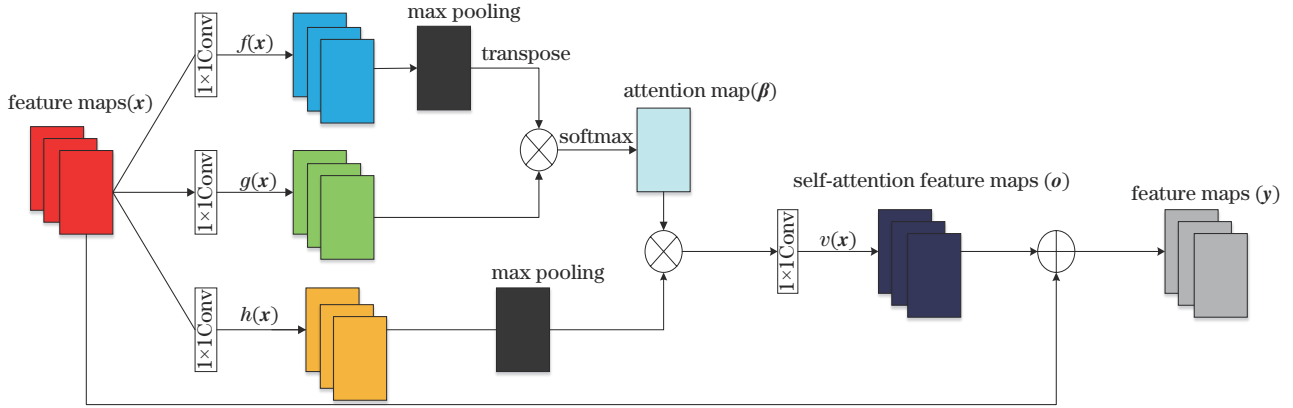


图 4 改进的自注意力机制图

Fig. 4 ISA mechanism diagram

$h(x) = W_h x$  相乘，得到注意力特征图  $o = (o_1, o_2, \dots, o_j, \dots, o_N) \in \mathbf{R}^{c \times N}$ ;

$$\begin{cases} o_j = v \left[ \sum_{i=1}^N \beta_{j,i} h(x_i) \right], \\ v(x_i) = W_v x_i \end{cases} \quad (3)$$

式中： $W_v \in \mathbf{R}^{\bar{c} \times c}$ ； $W_h \in \mathbf{R}^{\bar{c} \times c}$ 。最终将  $o$  层注意力输出  $o_i$  乘以参数  $\gamma$  再和  $x_i$  相加得到最终的注意力特征图  $y_i$ ：

$$y_i = \gamma o_i + x_i, \quad (4)$$

式中： $\gamma$  为可学习参数，初始化为 0，随着训练进行，从

邻域特征依赖逐渐增大为对较远区域依赖的权重。

### 3.3 基于 ISA 的生成器与判别器网络设计

生成器网络和判别器网络决定文字图像生成质量。传统卷积堆叠方式无法在相隔较远的文字图像空间区域像素之间建立长距离依赖关系，本实验组在生成器和判别器中加入改进的自注意力机制对稀疏像素进行全局约束，增强迁移文字内部相关性，提高文字迁移修复质量。生成器网络结构如图 5 所示，主要由编码器 C0~C2、修复模块和解码器 D1~D2 等 3 部分构

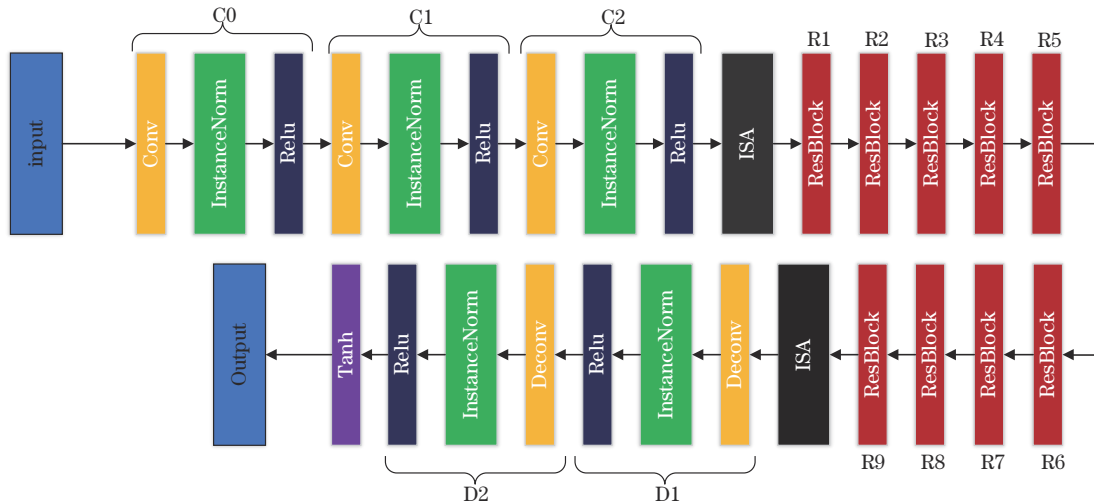


图 5 基于改进自注意力的生成器网络结构

Fig. 5 Generator network structure based on improved self-attention

成。编码器 C0-C2 主要作用为对输入书法文字图像进行特征提取,使用 InstanceNorm 归一化层分别对 3 个卷积层进行归一化后使用 ReLU 激活函数进行激活,即由 3 个 Conv+InstanceNorm+Relu 结构进行堆叠而成。修复模块主要作用为将破损图像  $A$  迁移修复为完整图像  $B$ ,由两个所提 ISA 模块与 9 个 ResNet 中 ResBlock 残差块<sup>[14]</sup>R1~R9 堆叠而成。解码器 D1~D2 的主要作用是还原图像的低级特征,由 2 个 Deconv+InstanceNorm+Relu 结构堆叠而成。最后将 D2 的输出结果再输入 tanh

激活函数中进行非线性激活得到输出。

判别器网络结构如图 6 所示,主要由 Conv+InstanceNorm+Relu 结构堆叠而成。对 H0 层进行卷积特征提取后通过非线性 ReLU 激活函数激活,H1、H2、H3 层在卷积后使用与生成器中相同的 InstanceNorm 归一化后再通过 ReLU 激活函数激活,最后一层经过  $1 \times 1$  卷积将通道数降为 1,并加入改进的自注意力机制来捕捉文字图像像素之间的几何和结构信息。

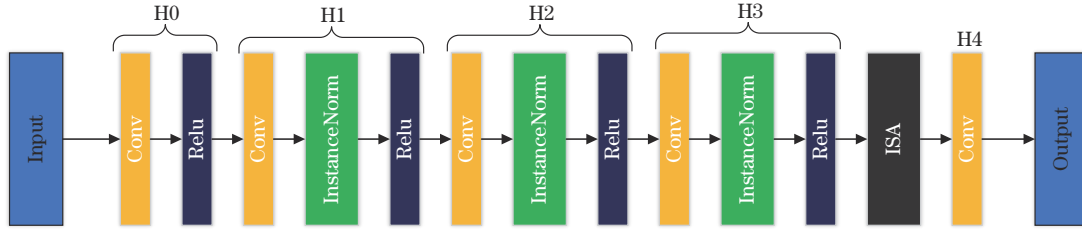


图 6 基于改进自注意力机制的判别器网络结构

Fig. 6 Discriminator network structure based on improved self-attention mechanism

### 3.4 提高迁移精度的损失函数改进

损失函数确定了图像修复网络的权重更新方向,决定着修复前后文字图像之间的差距,对文字图像修复至关重要。因此,为了获得一致性的迁移结果,所提方法采用最小二乘损失、循环一致性损失来共同约束目标函数。

原 CycleGAN 中的 sigmoid 交叉熵损失函数<sup>[15]</sup>无法对被判别器误判的图像进行修正,生成器不优化距离决策边缘较远但被判别为真的图像,影响文字迁移修复的质量。所提方法采用最小二乘损失代替 sigmoid 交叉熵损失函数,文字图像被误判时,惩罚距离图像决策边界较远但被判别为真的生成文字图像,提高网络判别准确度,生成更好的文字修复结果。

生成器  $G$  及判别器  $D_Y$  损失函数的表达式为

$$L_{GAN}(G, D_Y, X, Y) = E_{y \sim p_{\text{ann}}(y)} \left\{ \left[ D_Y(y) - 1 \right]^2 \right\} + E_{x \sim p_{\text{ann}}(x)} \left\{ \left[ D_Y[G(x)] \right]^2 \right\}, \quad (5)$$

式中:  $p_{\text{ann}}(y)$  为完整文字图像的数据分布;  $p_{\text{ann}}(x)$  为残缺文字图像的数据分布。同样,逆向映射的生成器  $F$  及其对应的判别器  $D_X$  损失函数的表达式为

$$L_{GAN}(F, D_X, X, Y) = E_{x \sim p_{\text{ann}}(x)} \left\{ \left[ D_X(x) - 1 \right]^2 \right\} + E_{y \sim p_{\text{ann}}(y)} \left\{ \left[ D_X[G(y)] \right]^2 \right\}. \quad (6)$$

在输入文字图像中选取样本  $(a, b)$ ,  $(a \in A, b \in B)$ , 其中  $A$  是待修复图像域,  $B$  为修复后图像域。  $A$  域中的图像  $a$  都应满足  $a \rightarrow G(x) \rightarrow F[G(x)] = \tilde{a}$ , 同样  $B$  域中图像需要满足  $b \rightarrow G(x) \rightarrow F[G(x)] = \tilde{b}$ , 在  $a \approx \tilde{a}, b \approx \tilde{b}$  的过程中构成循环一致性。循环一致性损失函数的表达式为

$$L_{\text{cycle}}(G, F) = E_{x \sim p_{\text{ann}}(x)} \left\{ \|F[G(x)] - x\|_1 \right\} + E_{y \sim p_{\text{ann}}(y)} \left\{ \|G[F(y)] - y\|_1 \right\}, \quad (7)$$

式中:  $\|\cdot\|_1$  为数据的 1 范数。循环一致性损失保证了  $A$  域中待修复图像修复完成后图像内容的风格不变性。

使用多项损失函数来优化目标函数,使修复图像网络更加稳定,可以获得更好的修复效果:

$$L = L_{GAN}(G, D_Y, X, Y) + L_{GAN}(F, D_X, X, Y) + \varphi L_{\text{cycle}}(G, F), \quad (8)$$

式中:  $\varphi$  为循环一致性损失在总损失  $L$  中所占权重。对于本实验而言,  $\varphi = 10$  时可取得较好的修复效果。

## 4 实验结果及分析

### 4.1 实验数据背景

实验数据采集于天津市蓟州区的独乐寺,由于光照、风化等自然因素和人为破坏等原因,寺中墙壁上字迹已变得残破不堪。观音阁中泥土墙壁上记载的文字表面笔迹残断、墨迹脱落,文字呈现模糊、残断、稀疏性的特点,如图 7 所示。



图 7 天津独乐寺中部分残缺文字图像

Fig. 7 Partial incomplete characters in Tianjin dule Temple

#### 4.2 数据集准备和实验环境

对文字图像的内容风格迁移进行文字修复,从残缺书法图像数据  $A$  迁移修复至完整书法图像数据  $B$ ,使用非匹配数据集  $A$ 、 $B$  进行网络训练。数据集  $A$  为观音阁中楷体书法文字分割处理结果,如图 8 所示。

由于书法文字方面没有标准公共数据集作为参考目标,本实验组根据数据集  $A$  中楷体书法文字特点,搜集自建诸如欧阳询、王羲之、苏轼、黄庭坚等 58 位中国古代书法名家所书的 39459 个楷体书法文字数据集  $B$ ,如图 9 所示。

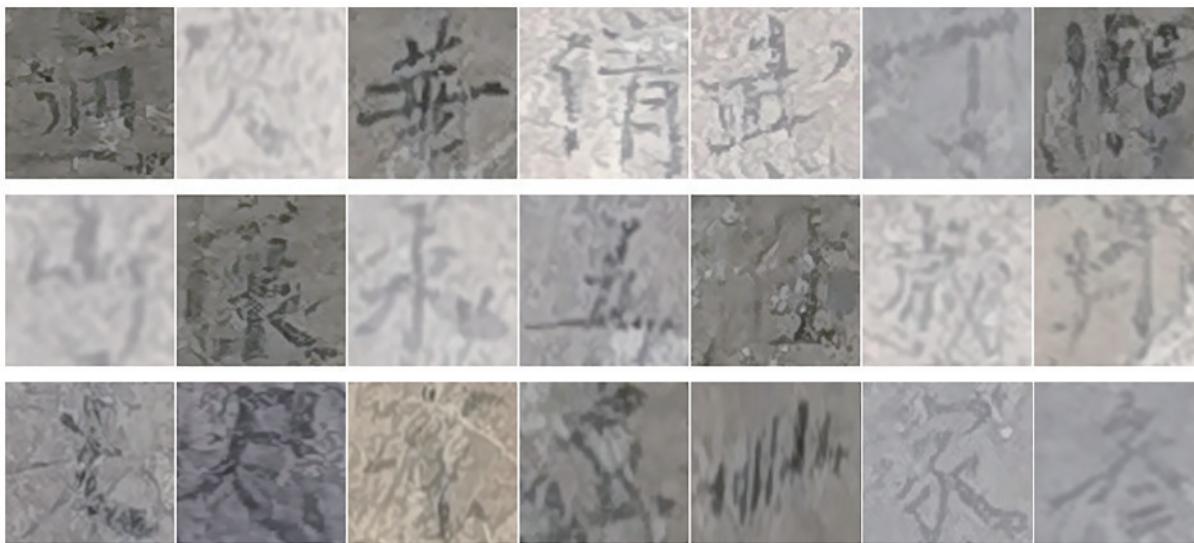


图 8 天津独乐寺残缺文字分割结果图

Fig. 8 Segmentation results of incomplete text in Tianjin dule Temple



图 9 欧阳询、王羲之、苏轼、黄庭坚等人书部分楷体数据

Fig. 9 Some regular script data of Ouyang xun, Wang Xizhi, Su Shi, Huang Tingjian, etc

残缺文字图像分割后采用自适应直方图均衡化 (AHE)<sup>[16]</sup>对文字墨迹与背景之间特征对比度进行增强,突出文字主体。分割增强后的图像比例较小、无法修复,为保证增大图像比例的同时保留图像分辨率,使用超分辨率(SR)重建<sup>[17]</sup>对图像进行4K重建。并对重建后的文字骨架进行提取,减少噪声以提高修复性能。最后,将处理后的数据集  $A$ 、数据集  $B$  文字图像输入修复网络,测试训练模型,得到残缺文字图像修复后结果。

实验环境如下:Windows 10 64 位系统;Intel(R)

Core(TM) i7-8700 CPU @3.20 GHz CPU; 8 GB 内存; NVIDIA GeForce RTX 2070 GPU; 8 GB 显存。网络基于 Python 3.6 TensorFlow 1.8-gpu 实现, batch size 设为 1, 学习率为 0.0002, 采用 Adam 优化器进行优化。

#### 4.3 修复性能评价

一般采用由人眼视觉的主观感受进行的定性评价与客观定量评价相结合的方式评价图像性能。为验证所提盲修复算法的有效性,采用主观定性评价与客观定量评价结合的方式进行评价,客观评价使用峰值信

噪比 (PSNR)、结构相似度 (SSIM) 和均方根误差 (RMSE)。PSNR 是最普遍和使用最广泛的图像客观评价指标, 基于对应像素之间的误差来判定图像质量, 数值越大表示图像质量越好; SSIM 通过对比原图像和修复后图像之间在对比度、结构、亮度等 3 方面的差异来衡量修复前后图像的相似性指标, 数值越大表示修复前后图像越相似; RMSE 也称标准差, 反映了真实

图像和修复图像之间的偏差, RMSE 越小表示原图像与修复后图像越接近。

随机挑选刘炳森、赵孟頫、胡问遂、田蕴章、黄自元、褚遂良、颜真卿、柳公权等 8 位书法名家的“畫”“歲”“會”“美”“福”“倫”“盟”“後”等 8 种破损书法文字图像控制变量地进行客观质量评价来验证所提算法的优势。图 10 为 CycleGAN、CycleGAN(LS+SA) 与所



图 10 实验结果对比。(a) 破损图; (b) CycleGAN; (c) CycleGAN(LS+SA); (d) 所提算法; (e) 原始图像

Fig. 10 Comparison of experimental results. (a) Broken images; (b) CycleGAN; (c) CycleGAN(LS+SA); (d) proposed algorithm; (e) original images

提算法进行盲修复的实验结果对比,其中 CycleGAN (LS+SA)表示在 CycleGAN 中使用最小二乘和自注意力机制的算法。

从图 10(b)可以看出,原始的 CycleGAN 算法能够较大程度完成修复,但在“歲”字及“後”上会出现笔画粘连的情况。从图 10(c)可以看出,对 CycleGAN 进行改进后的 CycleGAN(LS+SA)算法,较大程度解决了

CycleGAN 中的“歲”字及“後”的粘连情况,文字图像修复效果较原始 CycleGAN 有所提高。从图 10(d)可以看出,所提算法修复文字像素清晰、结构连续,修复效果均优于其余算法。

为进一步验证所提算法的有效性,将书法文字图像均调整为  $256 \times 256$ 、3 通道,进行客观评价指标 PSNR、SSIM 和 RMSE 的对比,结果如表 1 所示。

表 1 PSNR、SSIM 和 RMSE 对比结果  
Table 1 PSNR, SSIM, and RMSE comparison results

Evaluation index	Algorithm	Calligraphy image							
		畫	歲	會	美	福	倫	盟	後
PSNR	CycleGAN	12.29	11.98	12.99	14.05	12.14	12.44	13.31	13.32
	CycleGAN(LS+SA)	12.63	<b>13.16</b>	13.29	15.23	12.18	13.21	13.59	13.52
	Proposed algorithm	<b>13.24</b>	12.41	<b>14.04</b>	<b>15.65</b>	<b>12.73</b>	<b>14.27</b>	<b>14.59</b>	<b>14.71</b>
SSIM	CycleGAN	0.7535	0.7374	0.7873	0.8038	0.7252	0.7795	0.8155	0.7854
	CycleGAN(LS+SA)	0.7405	<b>0.7669</b>	0.7820	0.8229	0.7308	0.7933	0.8126	0.7817
	Proposed algorithm	<b>0.7720</b>	0.7392	<b>0.8014</b>	<b>0.8372</b>	<b>0.7457</b>	<b>0.8144</b>	<b>0.8258</b>	<b>0.8143</b>
RMSE	CycleGAN	61.92	64.22	57.14	50.62	62.31	60.87	55.06	55.03
	CycleGAN(LS+SA)	59.56	<b>56.02</b>	55.20	44.14	62.69	55.73	53.31	53.75
	Proposed algorithm	<b>55.58</b>	61.08	<b>50.64</b>	<b>42.05</b>	<b>58.89</b>	<b>49.30</b>	<b>47.53</b>	<b>46.86</b>

表 1 数据表明:原始 CycleGAN 对“畫”“歲”“會”“美”“福”“倫”“盟”“後”等 8 个不同文字进行修复时具有最低的 PSNR、最低的 SSIM 和最高 RMSE,修复效果最差;对 CycleGAN 进行改进后的 CycleGAN(LS+SA)算法在 PSNR、SSIM、RMSE 指标上均优于原始 CycleGAN;而所提算法仅在“歲”字上的客观指标不如 CycleGAN(LS+SA),在其余 7 个文字上修复指标均为

最优。并且从图 10 也可以看出,在“歲”字笔画中“横”的部分,所提算法修复结果明显连续优于 CycleGAN(LS+SA)。即所提算法具有更好的修复效果。

此外,为探究图像破损比例对图像修复性能的影响,随机挑选“動”“伐”“吸”“煩”“符”“館”“資”“益”等 8 个不同破损比例下的文字图像进行修复对比,实验结果如表 2 所示。

表 2 各种破损比例下的 PSNR、SSIM 和 RMSE 对比结果  
Table 2 Comparison results of PSNR, SSIM, and RMSE under various failure ratios

Mask ratio / %	Dataset	Calligraphy image dataset							
		動	伐	吸	煩	符	館	資	益
PSNR	CycleGAN	11.53	12.54	23.73	28.27	32.18	43.39	54.95	62.73
	CycleGAN(LS+SA)	13.97	14.53	13.54	14.43	14.11	11.85	11.56	10.11
	Proposed algorithm	<b>14.40</b>	<b>15.10</b>	<b>14.47</b>	14.54	15.20	12.45	11.74	11.16
SSIM	CycleGAN	14.20	13.75	14.29	<b>14.67</b>	<b>15.63</b>	<b>13.09</b>	<b>12.58</b>	<b>11.81</b>
	CycleGAN	0.7876	0.7968	0.8009	0.7966	0.8395	0.7562	0.7564	0.6988
	CycleGAN(LS+SA)	0.8103	<b>0.8380</b>	<b>0.8302</b>	<b>0.8101</b>	<b>0.8639</b>	0.7697	0.7509	0.7162
RMSE	Proposed algorithm	<b>0.8120</b>	0.7990	0.8254	0.8066	0.8478	<b>0.7921</b>	<b>0.7713</b>	<b>0.7409</b>
	CycleGAN	51.12	47.87	53.67	48.42	50.22	65.14	67.35	79.66
	CycleGAN(LA+SA)	<b>48.58</b>	<b>44.82</b>	<b>48.18</b>	47.83	44.31	60.82	66.02	70.51
Proposed algorithm	49.69	52.39	49.24	<b>47.16</b>	<b>42.19</b>	<b>56.52</b>	<b>59.90</b>	<b>65.49</b>	

从表 2 可以看出,整体而言,对于不同残缺文字图像,随着文字图像破损比例的不断增大,图像修复难度也逐渐增加。3 种修复算法的 PSNR 和 SSIM 逐渐减

小, RMSE 逐渐增大,修复效果下降。不同文字图像破损比例下,所提算法较 CycleGAN 算法而言均有更好的效果。与 CycleGAN(LS+SA)相比,在小破损率的



情况下, CycleGAN (LS+SA) 有着更大 SSIM 与 PSNR, 但随着破损比例的增加, 所提算法在 PSNR、SSIM、RMSE 指标上均是最优的。

#### 4.4 观音阁残缺文字修复评价

为验证所提盲修复算法在残缺稀疏文字修复方面的优势, 随机挑选观音阁中的“熙”“都”“年”“淮”等 4 个残缺稀疏书法文字图像, 使用所提算法与深度神经网络领域 RFR<sup>[18]</sup> 算法进行修复对比实验。

在修复观音阁中文字前, 需要对其进行预处理操作。首先对单字截取后的文字图像进行 AHE 处理, 结果如图 11 所示。从图 11 可以看出, 通过 AHE 算法改变“熙”“都”“年”“淮”图像像素直方图分布, 可以增强图像全局对比度, 有利于分离文字与背景。对原始图像进行 AHE、SR 重建、文字提取预处理后的处理结果如图 12 所示。

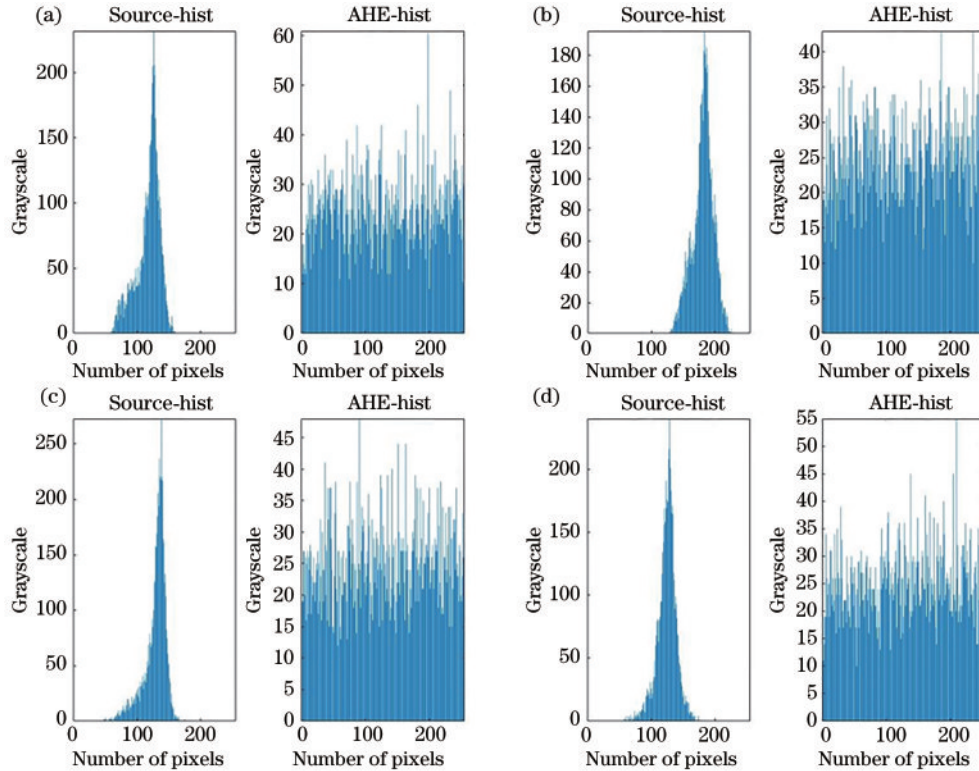


图 11 “熙”“都”“年”“淮”书法图像 AHE。(a)“熙”字书法图像 AHE;(b)“都”字书法图像 AHE;(c)“年”字书法图像 AHE;(d)“淮”字书法图像 AHE

Fig. 11 “Xi” “Du” “Nian” and “Huai” calligraphy image AHE. (a) “Xi” calligraphy image AHE; (b) “Du” calligraphy image AHE; (c) “Nian” calligraphy image AHE; (d) “Huai” calligraphy image AHE

从图 12 可以看出, 原始图像经过 AHE 与 SR 后, 增强了文字与背景之间的对比度, 但含有许多不属于文字的噪声部分。文字提取去除噪声后保留了文字骨架, 结构较完整, 但像素不连续且残缺。采用深度学习图像算法 RFR 与所提算法对图 12(c) 残缺稀疏文字进行修复, 对比结果如图 13 所示:

从图 13 可以看出, 由于人为添加 Mask 的精细程度难以控制, 同时需要对缺失区域 Mask 进行训练, 深度神经网络 RFR 图像修复算法对于残缺稀疏文字图像的修复与文字提取结果相比变化不大, 文字修复效果较差。而所提算法不借助 Mask 指导, 对图 12(c) 残缺文字图像进行全局盲修复, 修复后的结果语义清晰、像素连续、结构清晰, 具有较好的修复结果。

天津独乐寺观音阁中残缺书法文字图像, 没有原始图像作为图像质量评价参考依据。因此, 使用无参考图像修复结果质量评价方法对所提算法修复结果

与 RFR 算法修复结果进行评价。无参考图像质量评价通常以平均意见得分 (MOS)<sup>[19]</sup> 作为感知图像质量评价的主观指标。最常用的是以 5 分制作为 MOS 的给分尺度。以国际无线电咨询委员会制定的 CCIR500-1 号建议书, 将图像分为 5 个等级, 通过人的视觉系统对视觉信号的感知作为图像最终质量, 如表 3 所示。MOS 为平均一组评分尺度中多个评分结果的平均:

$$S_{\text{MOS}} = \frac{1}{N} \sum_{i=1}^N R_i, \quad (9)$$

式中:  $N$  表示受测试人数;  $R$  为受试者给定的评级。

实验评价小组由随机挑选的 4 名普通同学、1 名具有书法文字经验知识学生、随机挑选的 2 名普通老师、2 名具有图像修复经验知识老师及 1 名具有书法文字经验知识的老师组成。对所提盲修复算法与 RFR 算法进行得分测评, 测评结果如表 4 所示:



图 12 图像预处理结果。(a)原始图像;(b)AHE+SR;(c)文字提取

Fig. 12 Image preprocessing results. (a) Original images; (b) AHE+SR; (c) text extraction



图 13 不同修复算法对比结果。(a)人为添加 Mask;(b)RFR算法;(c)所提算法

Fig 13 Comparison results of different repair algorithms. (a) Adding mask artificially; (b) RFR; (c) proposed algorithm

表 3 CCIR500-1 主观评价尺度

Table 3 CCIR500-1 Subjective evaluation scale

Score	Obstruction criterion	Quality standard
5	No deterioration in image quality	Very good
4	The image quality has deteriorated, not hinder viewing	Good
3	The deterioration of image quality slightly hinders viewing	General
2	Obstruction to viewing	Bad
1	Very serious obstruction of viewing	Very bad

由表 4 可以看出,所提算法对残缺稀疏文字“熙”“都”“年”“淮”修复结果的评分,均优于 RFR 算法。即

所提算法对独乐寺中残缺稀疏文字图像的修复可以取得更好的效果。

表 4 主观评价测评结果  
Table 4 Subjective evaluation results

Proposed result	RFR result	Proposed score	RFR score
		4.2	2.6
		3.9	1.7
		4.9	2.5
		3.8	2.0

## 5 结 论

深度神经网络修复需要对图像残缺区域添加训练的 Mask 来完成修复,对于无法精细控制添加 Mask 或全局残缺的图像修复效果较差。针对残缺稀疏文字图像无法人为确定 Mask 区域的问题,采用基于内容风格迁移的盲修复方法,并对修复网络进行优化改进。在自建书法数据集及独乐寺残缺稀疏文字图像上进行了对比,实验结果表明:所提算法能够获得较好的主观文字修复效果,较高的 PSNR 与 SSIM;在稀疏性图像修复中的 MOS 得分也远远高于深度神经网络图像修复的方式,具有较好的修复还原度。未来的工作将继续在修复稳定性及细节上进行探究,使网络达到更好的修复性能。

## 参 考 文 献

- [1] 张娜, 陈露军, 王绪本. 基于水平垂直投影的考古文字修复识别方法[J]. 科技通报, 2014, 30(6): 185-187.  
Zhang N, Chen L J, Wang X B. Method of the archaeological word recognition based on the horizontal and vertical projection[J]. Bulletin of Science and Technology, 2014, 30(6): 185-187.
- [2] 张伟, 王绪本, 金平. Canny 边缘算子在简牍文字修复中的应用[J]. 微计算机信息, 2008, 24(9): 241-242, 250.  
Zhang W, Wang X B, Jin P. Application of Canny operator on repair of Jiandu's letter[J]. Microcomputer Information, 2008, 24(9): 241-242, 250.
- [3] 张商珉. 侯马盟书碑文及色彩虚拟修复技术研究[D]. 太原: 中北大学, 2017: 16-44.

- Zhang S M. Research on inscription and color virtual restoration technology of Houma covenant tablets[D]. Taiyuan: North University of China, 2017: 16-44.
- [4] Isola P, Zhu J Y, Zhou T H, et al. Image-to-image translation with conditional adversarial networks[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition, July 21-26, 2017, Honolulu, HI, USA. New York: IEEE Press, 2017: 5967-5976.
- [5] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems 27: Annual Conference on Neural Information Processing Systems 2014, December 8-13, 2014, Montreal, Quebec, Canada. [S.l.: s.n.], 2014: 2672-2680.
- [6] Tian Y C. Zi2zi: master Chinese calligraphy with conditional adversarial networks[EB/OL]. (2019-08-09) [2021-02-14]. <https://github.com/kaonashi-tyc/>.
- [7] 任健. 面向文物修复的中国书法风格迁移研究与应用[D]. 西安: 西北大学, 2019: 31-45.  
Ren J. Research and application of Chinese calligraphy style transfer for cultural relics restoration[D]. Xi'an: Northwest University, 2019: 31-45.
- [8] 陈善雄, 朱世宇, 熊海灵, 等. 一种双判别器 GAN 的古彝文字符修复方法[J]. 自动化学报, 2022, 48(3): 853-864.  
Chen S X, Zhu S Y, Xiong H L, et al. A restoration method of ancient Yi characters with double discriminator GAN[J]. Acta Automatica Sinica, 2022, 48(3): 853-864.
- [9] Jiang Y L, Liu J, Wang L Y. Double cross algorithm of improved TV image inpainting model[C]//2018 IEEE 3rd Advanced Information Technology, Electronic and Automation Control Conference, October 12-14, 2018, Chongqing, China. New York: IEEE Press, 2018: 1097-1100.
- [10] 苏畅, 付天骄, 张星祥, 等. 基于能量约束的自适应加权图像盲复原算法[J]. 光学学报, 2018, 38(2): 0210001.  
Su C, Fu T J, Zhang X X, et al. Adaptively-weighted blind image restoration algorithm based on energy constraint[J]. Acta Optica Sinica, 2018, 38(2): 0210001.
- [11] Heusel M, Ramsauer H, Unterthiner T, et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium[EB/OL]. (2017-06-26) [2021-02-01]. <https://arxiv.org/abs/1706.08500>.
- [12] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[EB/OL]. (2017-06-12) [2021-02-04]. <https://arxiv.org/abs/1706.03762>.
- [13] 张文秀, 朱振才, 张永合, 等. 基于残差块和注意力机制的细胞图像分割方法[J]. 光学学报, 2020, 40(17): 1710001.  
Zhang W X, Zhu Z C, Zhang Y H, et al. Cell image segmentation method based on residual block and attention mechanism[J]. Acta Optica Sinica, 2020, 40(17): 1710001.
- [14] He K M, Zhang X Y, Ren S Q, et al. Deep residual learning for image recognition[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE

- Press, 2016: 770-778.
- [15] Zhang Z L, Sabuncu M R. Generalized cross entropy loss for training deep neural networks with noisy labels[EB/OL]. (2018-05-20)[2021-02-04]. <https://arxiv.org/abs/1805.07836>.
- [16] 董丽丽, 丁畅, 许文海. 基于直方图均衡化图像增强的两种改进方法[J]. 电子学报, 2018, 46(10): 2367-2375.  
Dong L L, Ding C, Xu W H. Two improved methods based on histogram equalization for image enhancement [J]. Acta Electronica Sinica, 2018, 46(10): 2367-2375.
- [17] Shi W Z, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition, June 27-30, 2016, Las Vegas, NV, USA. New York: IEEE Press, 2016: 1874-1883.
- [18] Li J Y, Wang N, Zhang L F, et al. Recurrent feature reasoning for image inpainting[C]//2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), June 13-19, 2020, Seattle, WA, USA. New York: IEEE Press, 2020: 7757-7765.
- [19] Streijl R C, Winkler S, Hands D S. Mean opinion score (MOS) revisited: methods and applications, limitations and alternatives[J]. Multimedia Systems, 2016, 22(2): 213-227.