

文章编号: 2096-1618(2022)01-0034-06

# Non-local 注意力机制生成对抗网络的油画修复研究

何 妍, 何 嘉

(成都信息工程大学计算机学院, 四川 成都 610225)

**摘要:**针对部分油画艺术作品存在图像破损的问题,提出一种基于非局部(Non-local)注意力机制生成对抗网络的油画修复方法。首先,在生成器部分,采用扩张卷积和门控卷积替代原网络中的普通卷积层,增强网络的特征提取能力,同时加入 Non-local 注意力机制,提升生成器的修复能力;其次,使用马尔科夫判别器,强化网络的判别效果;最后,在损失函数部分使用感知损失、GAN 损失和 L1 损失,使整个网络的训练更加稳定。网络在开源的 Gal-lerix 油画数据集上进行了验证,实验结果表明:与 Global&Local、ParticalConvGAN 和 Deepfillv2 的修复方法相比,PSNR 和 SSIM 指标均超过 3 种优秀的生成对抗网络,对油画艺术品的修复的效果进行了更好的提升。

**关 键 词:**生成对抗网络;Non-local;油画修复;扩张卷积;门控卷积

**中图分类号:**TP751

**文献标志码:**A

**doi:**10.16836/j.cnki.jcuit.2022.01.006

## 0 引言

油画材料的特殊性、存储方式不当等原因可能导致图像画质出现破损。例如存放环境的温度、湿度的变化可能导致油画掉色,出现破损区域,就会降低油画的价值。通过对油画图像的修复,可以修复其艺术性。然而,油画修复是一个非常耗时的精细工作,随着深度学习图像修复技术<sup>[1]</sup>的出现,可以高效地对油画艺术品进行图像修复。

图像修复(image inpainting)最早由 Bertalmio 等<sup>[2]</sup>在 2000 年提出,是一种基于偏微分方程的图像修复方法。主要思想是将待修复破损区域的周围信息沿着等照度线方向由外向内传播,此方法利用了物理学中热扩散方程的方法。虽然,之前的图像修复方法<sup>[3-5]</sup>在整体结构上能够修复完整的画面,但对图像的修复效果处理不太令人满意。

因此,提出一种基于 Non-local 注意力机制生成对抗网络修复方法,修复破损的油画图像。该方法主要在 3 方面进行了改进:(1)使用门控卷积替换生成网络的普通卷积层,该结构有助于网络进行特征筛选;(2)网络生成器分为第一、第二阶段,通过双阶段设计提升图像细节的修复效果。在第一阶段,使用扩张卷积,以增大图像的感受野,有助于捕捉到更多的图像特征信息;在第二阶段,引入 Non-local 注意力机制,该机制使用的是全局特征信息,让网络感受野的范围进一步扩大,修复图像的效果更好;(3)判别模型采用光谱归一化马尔可夫判别器,用来解决图像边缘不和谐的问题。在损失函数上,采用感知损失、GAN 损失和 L1

重建损失的结合,使训练更加稳定、生成的图像效果也更好。实验证明,文中提出的对不规则的破损油画图像进行修复的方法,与其他修复方法<sup>[6-8]</sup>相比,具有更好的修复效果。

## 1 相关工作

近年来随着深度学习技术<sup>[9]</sup>的快速发展,图像修复成了计算机视觉的热门领域。Satoshi 等<sup>[6]</sup>提出了基于双判别器的图像修复方法 Global&Local,通过双判别器来提升网络对全局信息和局部信息的融合利用,使网络在完整区域中寻找最相似纹理块,进行图像复原。Guilin 等<sup>[7]</sup>在 ECCV 上首次利用部分卷积网络 PartialConvGAN 来实现图像修复。网络主体由生成器和判别器组成,生成器用来修复破损图像,判别器用来判断输入的图片是原始图片还是生成器生成的图片,并将结果反馈给生成器让其进行不断的优化,形成一种博弈状态,使图像修复的效果更真实。最近,Yu 等<sup>[8]</sup>提出 DeepFillv2 网络,它们采用双阶段生成网络结构。首先,在生成器第一阶段,提取图像中的语义信息;其次,在生成器第二阶段将粗生成网络的结果作为输入,使得生成纹理结构更合理。

## 2 网络结构

文中提出的油画图像修复算法的网络结构如图 1 所示,该算法以生成对抗网络为基本的模型架构,由生成器和判别器组成。卷积单元下面数字 48、96、192 代表卷积后输出通道数的大小,4 和 3 分别是输入和输出通道数大小。

收稿日期:2021-07-06

基金项目:四川省科技厅苗子工程资助项目(2019ZJ118);四川省科技厅应用基础重点资助项目(2017JY0011)

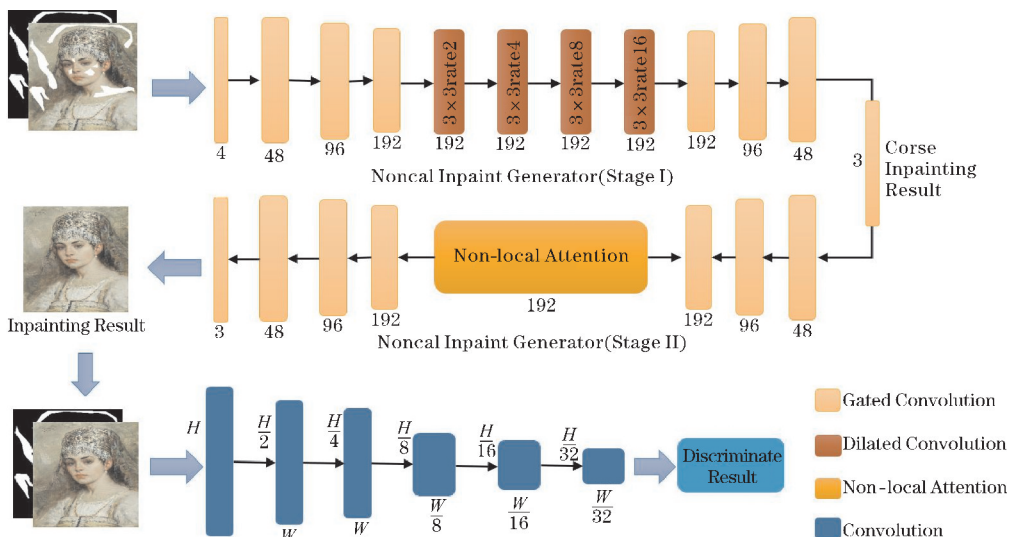


图1 网络结构

## 2.1 生成器

生成器输入是 RGB 破损图和掩膜灰度图组成的四通道图像,输出为修复图像。生成器是由两个编码器-解码器组成,分为第一和第二阶段,先得到粗略的修复结果,然后再进行精细化修复,两者组成完整修复流程。

门控卷积是一种具有特征筛选卷积层,因此,使用门控卷积替换普通卷积。在生成器的第一阶段,引入扩张卷积层<sup>[10]</sup>,在空间维度上增大感受野,让网络获得更多的特征信息,帮助图像修复。在生成器的第二阶段,引入 Non-local<sup>[11]</sup> 注意力机制,利用它建立局部特征信息和全局特征信息的联系,让网络的修复可以依赖于图像的整体结构,提升图像的修复效果。

### 2.1.1 Gated Convolution 门控卷积

门控卷积<sup>[12]</sup>最早出现在 NLP 的任务中,取得了良好的结果,Yu 等<sup>[8]</sup>将其引入图像修复任务。门控卷积的计算分 3 个步骤:首先,将特征图分别输入独立的卷积层,获得经过处理的特征图和通道权重图;然后,使用 Sigmoid 函数激活通道权重图,将值调整到 0~1;再将处理过的特征图和通道权重图进行点乘操作;最后使用 ELU 激活函数进行特征激活,得到最终输出结果。

### 2.1.2 Dilated Convolution 扩张卷积

Dilated convolution 扩张卷积,由 Google 团队<sup>[13]</sup>在 2014 年提出,其思路是在普通卷积核中通过 0 填充的方式,在不增加参数的同时,增大图像的感受野,从而获取更多的特征信息。普通的卷积操作,相当于扩张率为 1 的扩张卷积。

图 2 是扩张率为 2,卷积核大小为 3×3 的扩张卷积<sup>[12]</sup>,它通过 0 填充的方式,使卷积核的感受野大小变成 5×5。相比标准的卷积操作,扩张卷积多了一个参数扩张率,代表扩张的大小。

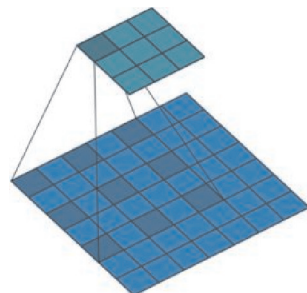


图2 扩张卷积结构

将扩张卷积应用到图像修复中,在生成器第一阶段结构的中间部分加入 4 层扩张卷积,随着层数的加深,扩张卷积层的扩张率不断变大,四层扩张率分别为 2、4、8 和 16。通过叠加扩张率不断变大的扩张卷积层,来实现不断增大图像感受野,增强网络对图像整体特征信息的提取能力和使用能力。

### 2.1.3 Non-local Attention 注意力机制

Non-local<sup>[11]</sup>注意力机制通过计算特征图任意两个位置之间的交互信息,建立特征值远距离依赖,不同于传统卷积只能局限于相邻特征点,它相当于构造一个和特征图尺寸一样大的卷积核,获得更多特征信息。Non-local 注意力机制通过引入全局特征信息,增大图像的感受野,为图像的补全获取更多丰富的特征信息。该机制在深度学习领域得到广泛应用<sup>[11]</sup>;语义分割、目标检测、图像超分辨率、图像去噪和文本检测。因此,文中引入 Non-local 注意力机制,以提升网络的图像修复效果。

Non-local<sup>[11]</sup>如式(1)所示

$$y_i = \frac{1}{C(x)} \sum_j f(x_i, x_j) g(x_j) \quad (1)$$

式中  $x, y$  为输入和输出,  $i$  代表特征值当前位置的响应,  $j$  代表的是全局位置的响应。  $g(x_j) = W_g x_j$ ,  $W_g$  为权重,  $f(x_i, x_j)$  如式(2)所示,  $C(x)$  如式(3)所示。

式中  $f(x_i, x_j)$  是 Non-local 的核心计算所在。其中,

$\theta(x_i) = W_\theta x_i, \varphi(x_j) = W_\varphi x_j, W_\theta$  和  $W_\varphi$  是权重。 $\theta(x_i)$  和  $\varphi(x_j)$  的计算细节如表 1 所示。T 是对矩阵进行转置,  $x_i$  和  $x_j$  分别是当前位置特征值和全局位置的特征值。 $f(x_i, x_j)$  通过矩阵相乘的计算方式, 来建立局部特征信息和全局特征信息的联系。

$$f(x_i, x_j) = e^{\theta(x_i)^T \varphi(x_j)} \tag{2}$$

$$C(x) = \sum_j f(x_i, x_j) \tag{3}$$

式中的  $C(x)$  是归一化因子, 以实现特征的 softmax 归一化, 它使特征值分布趋于中心化, 避免极值的出现(极值会导致全黑或全白的修复结果)。

表 1  $\theta(x_i)$  和  $\varphi(x_j)$  结构表

Name	$\theta(x_i)$	$\varphi(x_j)$
Input Size	[8, 64, 64, 192]	[8, 64, 64, 192]
Conv. Kernel Size	[192, 1×1, 1, 96]	[192, 1×1, 1, 96]
Conv. Output Size	[8, 64, 64, 96]	[8, 64, 64, 96]
Reshape Output Size	[8×64×64, 96]	[96, 8×64×64]

表 2  $g(x_i)$  和 Conv. 1×1×1 结构表

Name	$g(x_j)$	Conv. 1×1×1
Input Size	[8, 64, 64, 192]	[8, 64, 64, 96]
Conv. Kernel Size	[192, 1×1, 1, 96]	[96, 1×1, 1, 192]
Conv. Output Size	[8, 64, 64, 96]	[8, 64, 64, 96]
Reshape Output Size	[8×64×64, 96]	[8, 64, 64, 192]

Non-local 的结构如图 3 所示, 它涉及四次卷积、两次矩阵相乘和一次矩阵拼接计算。其中, B (Batch Size) 代表批处理大小, H (height) 和 W (width) 代表特征图的长和宽, 192 和 96 是特征图的通道数量, 1×1×1 代表卷积核的大小以及卷积步长。⊗ 表示矩阵相乘, ⊕ 表示把两个特征图的特征通道拼接到一起。

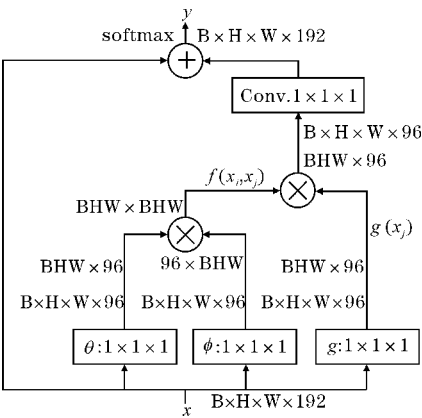


图 3 Non-local attention 结构

表 1 和表 2 是对图 3 中  $\theta(x_i)$ 、 $\varphi(x_j)$ 、 $g(x_i)$ 、Conv. 1×1×1 结构的解释。其中, 第一行是函数的名称; 第二行 Input Size 和第四行 Conv. Output Size 分别是特征图的输入和输出尺寸, 4 个值分别为批大小、高

度、宽度和通道数; 第三行 Conv. Kernel Size 是卷积核尺寸, 4 个值分别为输入通道数、卷积核大小、卷积步长和输出通道数; 第四行 Reshape Output Size 是对卷积输出变形后的特征图大小, 2 个值分别为批大小乘以高度乘以宽度、通道数。

2.2 SN Discriminator 判别器

采用光谱归一化马尔可夫判别器 (SN-Discriminator) 作为模型判别器, 网络通过判别器来进行对抗训练, 进而提升生成器的生成效果。判别器结构如图 4 所示, 图中的 H、W 代表图像的高度和宽度。

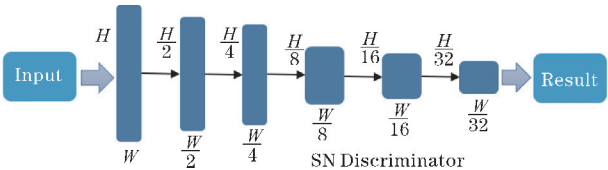


图 4 判别器结构

判别器的输入是修复图片和掩膜灰度图组成的 4 通道图片, 输出是判别结果。判别器的第一层采用的卷积核为 7×7, 大卷积核具有大的感受野, 能获取更多的特征信息; 后五层采用卷积核大小 4×4。表 3 是对 SN-Discriminator 判别器结构的详细解释。表中第一列是卷积层的名字; 第二列和第四列是输入和输出特征图的大小, 4 个参数分别为批大小、高度、宽度、通道数; 第三列是卷积核的大小, 4 个参数分别为输入通道数、卷积核大小、卷积步长、输出通道数。

表 3 SN-Discriminator 判别器结构表

Name	Input Size	Conv. Kernel	Output Size
x1	[8, 256, 256, 4]	[4, 7×7, 1, 64]	[8, 256, 256, 64]
x2	[8, 256, 256, 64]	[64, 4×4, 2, 128]	[8, 128, 128, 128]
x3	[8, 128, 128, 128]	[128, 4×4, 2, 256]	[8, 64, 64, 256]
x4	[8, 64, 64, 256]	[256, 4×4, 2, 256]	[8, 32, 32, 256]
x5	[8, 32, 32, 256]	[256, 4×4, 2, 256]	[8, 16, 16, 256]
x6	[8, 16, 16, 256]	[256, 4×4, 2, 1]	[8, 8, 8, 1]

3 损失函数和优化器

3.1 损失函数

在损失函数上, 本文算法使用 L1 损失 ( $\mathcal{L}_l$ )、感知损失 ( $\mathcal{L}_p$ ) 和 GAN 损失 ( $\mathcal{L}_g$ ), 总损失函数如式 (4) 所示, 超参数  $\alpha=100, \beta=10, \gamma=1$ 。

$$\mathcal{L} = \alpha \mathcal{L}_l + \beta \mathcal{L}_p + \gamma \mathcal{L}_g \tag{4}$$

选取 L1 作为重建损失函数, 如式 (5) 所示

$$\mathcal{L}_l = \frac{1}{HW} \|I - I'\|_1 \tag{5}$$



H 代表图像的高度, W 代表图像的宽度, I 代表原图, I' 代表破损的油画图像。

感知损失为

$$\mathcal{L}_p = \frac{1}{WH} \sum_i \sum_j (I_{i,j} - G(I')_{i,j})^2 \quad (6)$$

$i$  和  $j$  是像素的坐标值, I 是无损原图, I' 是破损图,  $G(I')$  代表的是经过修复后的油画图片。

GAN 损失计算的是均值误差为

$$\mathcal{L}_g = \frac{-\sum_{i=1}^n (I - I')}{n} \quad (7)$$

式中  $n$  是样本总数,  $i$  是样本序号。

### 3.2 Adam 亚当优化器

本文算法使用的是 Adam 优化器<sup>[14]</sup>, 通过一阶求导推导出偏差, 得到目标函数的梯度, 然后使用梯度值优化模型的参数。Adam 优化器参数如下: 学习率为 0.0001, 动量参数 beta1 和 beta2 的大小为 0.5 和 0.999。

## 4 实验数据集

### 4.1 数据收集

使用开源 Gallerix 油画数据集, 从中选取完整的油画图片来验证网络模型的性能, 部分数据集如图 5 所示。使用 3000 张印象主义风格的油画图像作为本文实验数据集, 图像分辨率为 256 像素, 饱和度为 100%, 使用其中 2700 张作为训练集, 300 张作为测试数据集。



图 5 部分数据集展示

### 4.2 破损图像的生成

在模型训练时, 本文算法通过随机生成破损区域的方式进行数据增强, 使用随机角度、长度和宽度的线条去模拟破损区域, 生成的破损掩码区域为占总面积的 10% ~ 25%。每次生成的破损区域不相同, 这丰富了训练数据, 使训练出的网络模型能够适应多种破损形态的油画。通过随机生成破损区域的方式, 进行数据增强, 避免模型的数据过拟合问题, 增强模型的泛化

性能。

## 5 实验及结果

### 5.1 实验环境

实验在一台深度学习服务器上完成, 本文算法通过 Python 语言编程实现, 使用 PyTorch 深度学习框架。服务器的具体配置如下: CPU 型号为英特尔 I7-7700, 操作系统是 Ubuntu 18.04 LTS, 系统的内存大小是 128 GB, 计算显卡使用的是 2 条 11GB 显存的英伟达 GTX 2080Ti 深度学习图形显卡。

### 5.2 实验参数及过程细节

本文算法模型训练 200 个迭代周期, 每个迭代周期的步长为 10000 步, BatchSize 批处理的大小为 8。训练时, 使用的掩模都是非连续随机生成, 占图像面积的 10% ~ 25%。在此范围内, 图像修复效果比较理想, 能较好地完成油画修复的任务。模型训练的前 100 个周期, 损失函数值从 15.01, 逐渐下降到了 6.50。在 100 ~ 200 的周期内, 损失函数值徘徊在 6.0 附近, 并逐渐趋于稳定。损失由最初的 15.01, 下降到最后的 5.62。

### 5.3 评价指标

使用两个图像质量评价指标来定量评估网络的性能, 分别是峰值信噪比 (PSNR) 和结构相似性 (SSIM), 数值越大, 表示失真越小。

$$\text{PSNR} = 10 \times \lg \left( \frac{(2^n - 1)^2}{\text{MSE}} \right) \quad (8)$$

式中的 MSE 是均方误差,  $n$  代表的是每个采样值的比特数。

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (9)$$

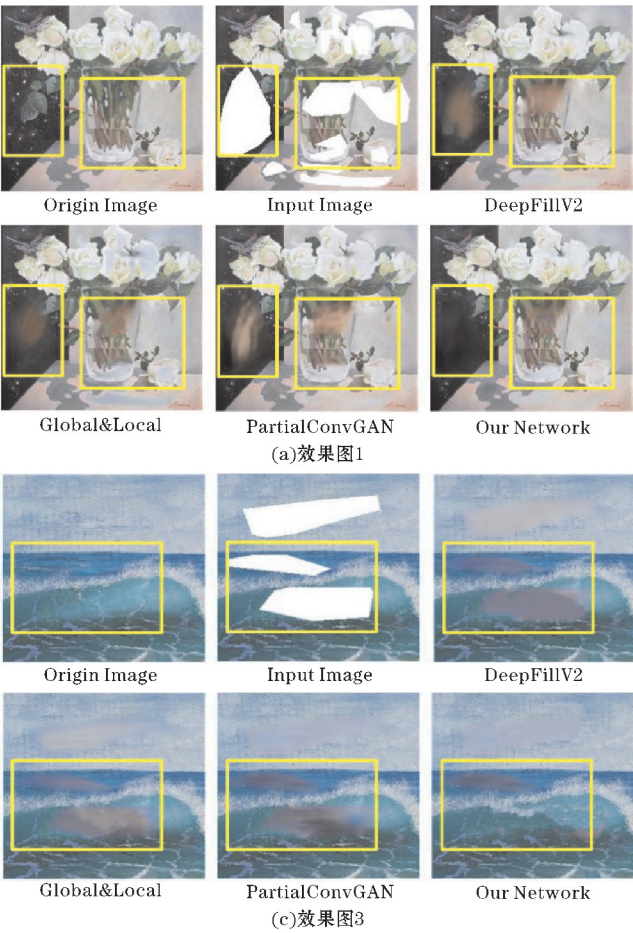
式中  $\mu_x$  和  $\mu_y$  分别代表  $x, y$  的平均值,  $\sigma_x$  和  $\sigma_y$  分别代表  $x$  和  $y$  的标准差,  $\sigma_{xy}$  代表  $x$  和  $y$  的协方差, 而  $C_1, C_2, C_3$  为常数, 避免分母为零导致的计算错误。

### 5.4 实验结果

为验证本文算法图像修复的有效性, 使用 300 张测试图片, 在开源 Gallerix 油画数据集上进行测试。与 3 种经典的图像生成对抗网络进行比较, 3 种网络分别是: 使用双判别器的图像修复网络 Global&Local<sup>[3]</sup>, 使用基于部分卷积的 PartialConv-GAN<sup>[4]</sup>网络, 以及专门针对自由形式掩码修复的 Deepfillv2GAN<sup>[7]</sup>网络, 量化指标如表 4 所示。表 4 的各列依次是模型名称、量化指标以及处理一张图片的平均耗时 (ms)。图 6 是实验的效果图。

表 4 实验结果量化表

模型名称	PSNR/ SSIM	Time/ms
Global&Local <sup>[3]</sup>	21.42/ 0.7877	36
PartialConvGAN <sup>[4]</sup>	22.89 / 0.8192	31
DeepFillv2GAN <sup>[7]</sup>	24.63 / 0.8540	45
本文算法	25.04 / 0.8615	39



实验对比模型一: Global&Local 网络。它是一种使用双判别器的生成对抗网络。虽然它是几种网络中效果较差的,但不可否认它是一种结构简单、性能强大的网络模型,在图像修复领域有着重要的影响。



图 6 实验效果图

实验对比模型二: PartialConvGAN 网络。它的诞生就是用于图像修复,针对图像中非矩形的破损区域进行修复,通过部件卷积实现对图像中有效元素和缺失元素的区别对待。模型训练时,卷积层仅仅对满足有效像素条件的元素进行卷积和归一化操作,然后根据区域权重更新规则实现对权重的更新操作。它的修复效果好于 Global&Local 网络。

实验对比模型三: Deepfillv2 网络。它被用于对人脸图像不规则破损区域进行修复的生成对抗网络,和本文一样采用双编码器和双解码器的生成结构。通过注意力机制来实现对破损区域特征信息的修复,这说明了注意力机制在图像修复中的重要性,在实验效果上,优于 Global&Local 和 PartialConvGAN 网络模型。

本文网络在两项图像修复效果量化指标峰值信噪比和结构相似度均超过其余 3 种网络。同时,通过观察图 6 的实验效果,可以发现本文算法修复的图像纹

理更加清晰、边缘细节上处理很好,明显优于其余 3 种网络。在表 4 的图像处理速度比较中,本文算法处理一张图像需要花 39 ms,与另外 3 种网络相比无明显差异,这也体现了深度学习方法在图像修复中处理速度上的优势。

6 结束语

得益于深度学习图像修复技术的发展,提出一种基于 Non-local<sup>[11]</sup>的非局部注意力机制生成对抗网络的油画修复方法,尽可能地恢复油画图像的价值。该网络受双编码器-解码器生成结构、Non-local 非局部注意力机制和门控卷积扩张卷积结构的启发,在提升破损区域修复效果和色彩的同时,也能对油画破损边缘区域进行修复,较好地提升了油画修复的效果。实验结果表明:本文方法具有对不规则的油画破损区域



进行有效修复的能力。

本文网络的损失函数仍有优化的空间,在对特征的使用上仍然具有改进的潜力,在下一步研究中将考虑加入超图卷积等结构来优化和提升网络。

## 参考文献:

- [1] 郭军胜. 大数据背景下的油画破损区域修复算法设计[J]. 现代电子技术, 2020, 43(19): 31-34.
- [2] Bertalmio M, Sapiro G, Caselles V, et al. Image Inpainting[C]. Proceedings of ACM SIGGRAPH. New Orleans, 2000: 417-424.
- [3] Yu J, Lin Z, Yang J, et al. Generative Image Inpainting with Contextual Attention[C]. IEEE Conference on Computer Vision and Pattern Recognition. United States: IEEE, 2018: 5505-5514.
- [4] 程颖. 基于生成对抗网络图像补全的研究与实现[D]. 成都: 电子科技大学, 2020.
- [5] 赵然. 基于深度学习的图像修复方法综述[J]. 科技风, 2020(18): 130-137.
- [6] Satoshi Iizuka, Edgar Simo-Serra, and Hiroshi Ishikawa. Globally and locally consistent image completion[J]. ACM Transactions on Graphics (TOG), 2017, 36(4): 1-14.
- [7] Guilin Liu, Fitsum A. Reda, Kevin J. Shih, et al. Image Inpainting for Irregular Holes Using Partial Convolutions[C]. European Conference on Computer Vision. Germany: ECCV, 2018: 85-100.
- [8] Yu J, Lin Z, Yang J, et al. Free-Form Image Inpainting with Gated Convolution[C]. IEEE International Conference on Computer Vision. Korea: ICCV, 2019: 4470-4479.
- [9] Goodfellow I J, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]. International Conference on Neural Information Processing Systems. Canada, 2014: 2672-2680.
- [10] F Yu, Koltun V. Multi-Scale Context Aggregation by Dilated Convolutions[OL]. <https://arxiv.org/abs/arXiv:1511.07122>, 2016.
- [11] Wang X, Girshick R, Gupta A, et al. Non-local Neural Networks[C]. IEEE Conference on Computer Vision and Pattern Recognition. United States: IEEE, 2018: 7794-7803.
- [12] Dauphin Y N, Fan A, Auli M, et al. Language Modeling with Gated Convolutional Networks[C]. International Conference on Machine Learning. China, 2017: 933-941.
- [13] Chen L C, Papandreou G, Kokkinos I, et al. Semantic Image Segmentation with Deep Convolutional Nets and Fully Connected CRFs[J]. Computer Science, 2014(4): 357-361.
- [14] Kingma D, Ba J. Adam: A Method for Stochastic Optimization[EB/OL]. <https://arxiv.org/abs/arXiv:1412.6980>, 2014.
- [15] Zhou W, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004: 600-612.

## Research on Oil Painting Restoration of Non-local Attention Mechanism Generative Adversarial Network

HE Yan, HE Jia

(College of Computer Science and Technology, Chengdu University of Information Technology, Chengdu 610225, China)

**Abstract:** Aiming at the problem of image damage in some oil painting artworks, a method of oil painting restoration based on Non-local attention mechanism generative adversarial network is proposed. First, in the part of the generator part, dilated convolution and gated convolution are used to replace the ordinary convolution layer in the original network, to strengthen the judgment effect of the network, and a non-local attention mechanism is added to improve the repair ability restoration effect of the generator; secondly, Markov discrimination is used to strengthen the discriminative effect of the network; finally, the loss function uses perceptual loss, GAN loss, and L1 loss are used in the loss function to make, which makes the training of the entire network more stable. The network of this article is verified on the open source Gallerix oil painting dataset. The experimental results show that the PSNR and SSIM indicators surpass the three excellent generative adversarial networks, and improve the effect of oil painting artwork restoration.

**Keywords:** generative adversarial network; Non-local; oil painting restoration; dilated convolution; gated convolution