

基于多级解码网络的图像修复

刘微容¹, 米彦春¹, 杨帆¹, 张彦¹, 郭宏林², 刘仲民¹

(1. 兰州理工大学电气工程与信息工程学院, 甘肃兰州 730050;
2. 大型电气传动系统与装备技术国家重点实验室, 甘肃天水 741000)

摘 要: 当前主流的图像修复方法重点依赖于自动编解码网络, 此类方法试图利用编码阶段压缩后的信息在解码阶段恢复出原始图像. 然而自编码网络在压缩过程中必然存在信息丢失, 仅利用压缩后的信息难以得到细节丰富的修复结果, 主要表现为模糊和修复区域周围明显的边缘响应. 本文针对图像信息利用不完备的问题, 提出多级解码网络(Multi-Stage Decoding Network, MSDN), 由多个解码器对编码阶段各层特征进行解码并聚合, 增大对编码器不同尺度特征的利用率, 进而得到更能反映缺损区域内容的特征映射. 在国际公认数据集上组织的对比实验结果表明, MSDN修复的图像视觉效果有一定提升.

关键词: 图像修复; 编解码器; 多级解码网络(MSDN)

中图分类号: TP391

文献标识码: A

文章编号: 0372-2112(2022)03-0625-12

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20201451

Generative Image Inpainting with Multi-Stage Decoding Network

LIU Wei-rong¹, MI Yan-chun¹, YANG Fan¹, ZHANG Yan¹, GUO Hong-lin², LIU Zhong-min¹

(1. College of Electrical and Information Engineering, Lanzhou University of Technology, Lanzhou, Gansu 730050, China;
2. State Key Laboratory of Large Electric Drive System and Equipment Technology, Tianshui, Gansu 741000, China)

Abstract: Current image inpainting methods mainly rely on automatic encoding and decoding networks. These methods try to use the information compressed in the encoding stage to restore an original image in the decoding stage. While, it is difficult to reconstruct detailed inpainting results by using only compressed information. Due to the loss of information during compression, there are visual artifacts in the results, such as blurring and obvious edge response around the reconstructed area. Aimed at the problem of incomplete utilization of image information, this manuscript proposed a multi-stage decoding network (MSDN). The MSDN decodes and aggregates features of each layer in the encoder by multiple decoders successively, which can increase utilization of features from different layers in the encoding stage and obtain better feature maps to reflect the defected area. The experiment results, which are conducted on internationally recognized datasets, show that visual effects of images generated by MSDN have been improved.

Key words: image inpainting; automatic encoding and decoding networks; multi-stage decoding network

1 引言

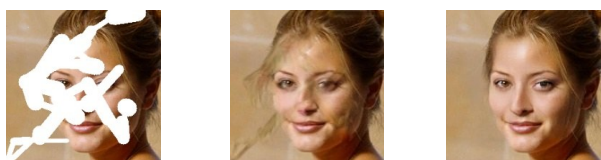
图像修复是根据破损图像的已知信息, 利用计算机技术推断出缺损区域像素的计算机视觉任务, 也称为图像补全. 图像修复的总体目标是重建语义上合理、视觉上逼真的完整图像, 其在遮挡区域去除、特定对象移除和珍贵历史资料修复等方面均有重要的应用价值^[1].

在深度学习方法提出之前, 图像修复理论研究主要集中在基于扩散的方法^[2, 3]和基于块匹配的方法^[4, 7]. 基

于扩散的方法利用距离场机制将相邻像素的外观信息传播到目标区域, 这种方法仅依赖于缺损区域周围的像素值, 因此只能修复小型破损; 对于大面积受损图像, 会产生伪影及中心区域模糊的现象. 基于块匹配的图像修复方法, 假设修复区域内容可以从周围已知区域找到, 以迭代的方式从未缺损区域搜索相关块复制到缺损区域. 但是, 搜索过程往往伴随大量的计算消耗, 匹配速度缓慢. 为减少运行时间, 提高内存效率, Barnes 等人^[4]提出了随机化搜索方法 PatchMatch, 可有

效加快搜索速度、提升修复质量. PatchMatch方法凭借其优越的背景修复能力被应用于许多应用程序中,例如经典的Photoshop商业软件. 基于块匹配的方法虽然在背景修复和重复性结构修复任务中表现出优越性能,但因其未借助高级语义信息来引导搜索过程,故难以应用至人脸等高度模式化图像的修复任务^[8,10].

随着卷积神经网络(Convolutional Neural Networks, CNN)^[11]和生成对抗网络(Generative Adversarial Networks, GAN)^[12]等深度学习方法的快速发展,基于CNN和GAN的编解码网络^[11, 13, 16]从大规模数据集中学习图像的语义特征,捕获缺损区域并利用已有特征重建完整图像,实现了语义层面的合理修复. 但是,基于编解码的图像修复方法在压缩编码过程中不可避免地存在信息丢失现象,单一的解码网络难以从压缩后的少量信息中准确重建出期望的结果,从而导致修复结果出现模糊或边缘响应等视觉伪影,严重影响视觉观测效果(如图1所示). 因此,严重的信息丢失问题已成为制约编解码图像修复方法性能提升的技术瓶颈.



(a)掩码图像 (b)出现视觉伪影图像 (c)本文方法修复结果
图1 是否充分利用编码部分信息修复结果对比

现有编解码方法^[11, 13, 16]对缺损区域的修复仅仅依赖于小尺度下的高级语义特征,忽略了在图像编码过程中可以产生丰富不同尺度特征的事实,而大尺度下存在的边缘信息在图像纹理和细节的重建过程中可以起到至关重要的作用. 因此,针对信息利用不完备的问题,本文对编码部分不同尺度特征施加多级解码,并将其结果与主解码器聚合以指导下一级特征的重建. 此外,进一步利用注意力机制来增强对关键区域的关注度,以产生更真实的细节信息.

本文提出的多级解码网络(Multi-Stage Decoding Network, MSDN)基于不同尺度特征增强的机理,可有效解决图像修复过程出现的模糊、边缘响应严重等影响视觉效果的问题.

本文的主要贡献如下:

(1)提出了多级解码网络MSDN,设计了主、副解码器,充分利用编码部分不同尺度下的高低级特征,共同指导合成主解码网络的各尺度特征;

(2)提出了并行连接机制,将注意力转移网络以并联的方式引入主解码网络. 注意力转移网络从上一层特征中匹配复制到相似块后,所得特征与上一层特征相结合,共同作用填充下一层特征缺损区域.

2 相关工作

2.1 图像修复

近几年,基于深度学习的图像修复方法逐渐被众多科研工作者所关注,其修复目标是:在已知待修复图像 y_m 的条件下,基于深度网络具有的映射关系 $f(\cdot)$,重建完整的图像 $\hat{y}$,如式(1)所示:

$$\hat{y} = f(y_m; \beta) \quad (1)$$

式中, β 指网络参数.

图像修复是一个不适定问题^[17],即修复结果并不连续地依赖于已知输入,一张缺损图像可能对应多个修复结果. 为了降低不适定性,需要在网络训练时对待修复图像施加一定先验,利用先验知识来重建缺失内容. 期望预训练网络重建的完整图像 $\hat{y}$ 尽可能与原始图像 y 相似,即损失函数最小,从而得到如式(2)所示优化目标函数:

$$\beta^* = \arg \min_{\beta} \sum_{k=1}^K \lambda_k \ell_k(\hat{y}, y) \quad (2)$$

式中, β^* 指网络修复性能最佳时的参数值, $\ell_k(\cdot)$ 指损失函数, λ_k 为各项损失函数的权衡参数.

2.2 基于卷积自编码网络的图像修复

基于深度学习的图像修复方法大多采用主流的编解码网络架构,其机理为对称神经网络压缩重构待修复图像^[18],如图2所示. 编码器将一张破损的图像 y_m 映射到低维的特征空间,再通过解码器重建出完整图像 $\hat{y}$,实现了从缺损图像到完整图像的端到端映射^[19].

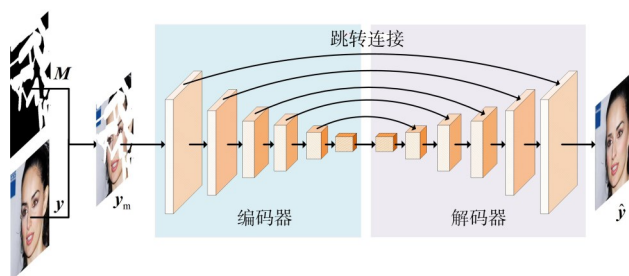


图2 带有跳跃连接的编解码网络

上下文解码(Context Encoder, CE)^[13]网络是首个采用自动编解码结构的图像修复方法,其通过卷积神经网络来填充 64×64 的中心缺损,并使用 L_2 像素级重建损失和对抗性损失训练网络模型,取得了语义上可信的修复结果,然而CE^[13]在生成精细纹理方面效果并不理想. 之后,Iizuka等人^[14]提出使用全局和局部判别器来生成全局图像和局部图像,并提出用扩张卷积替换CE中使用的通道级全连接层,有效提高了修复质量,但是,该方法需要添加后处理步骤以加强边界颜色的一致性. Yu等人^[15]提出上下文注意力(Context Attention, CA)机制,并将其应用在细粒度网络,以搜索匹配

与粗粒度网络预测结果相似度最高的背景块. 然而, 注意力机制和细粒度网络的串联形式可能会使前一级匹配到的错误信息向后传递, 难以保证邻域信息一致性^[20]. Liu 等人^[21]将部分卷积引入图像修复任务, 用部分卷积层更新掩码, 重新对卷积计算后的特征映射进行归一化, 确保卷积滤波器将注意力集中在已知区域的有效信息, 以处理不规则缺损区域. Wang 等人^[22]提出一种生成式多列神经网络 (Generative Multi-Column Convolutional Neural Network, GMCNN) 架构, 将三个编解码器相并联, 在训练阶段采用隐式多样化马尔可夫随机场正则化方案, 提出用置信值驱动的重建损失, 根据空间位置施加不同约束, 该方法在矩形掩码和不规则掩码下均表现出较好的修复性能.

以上所述方法均用到编码-解码的思想, 但是多个编解码器并联或串联的方式并不能从根本上解决信息丢失后重建的问题, 模糊和明显的边缘响应依旧是制约图像修复效果的关键因素.

3 所提方法

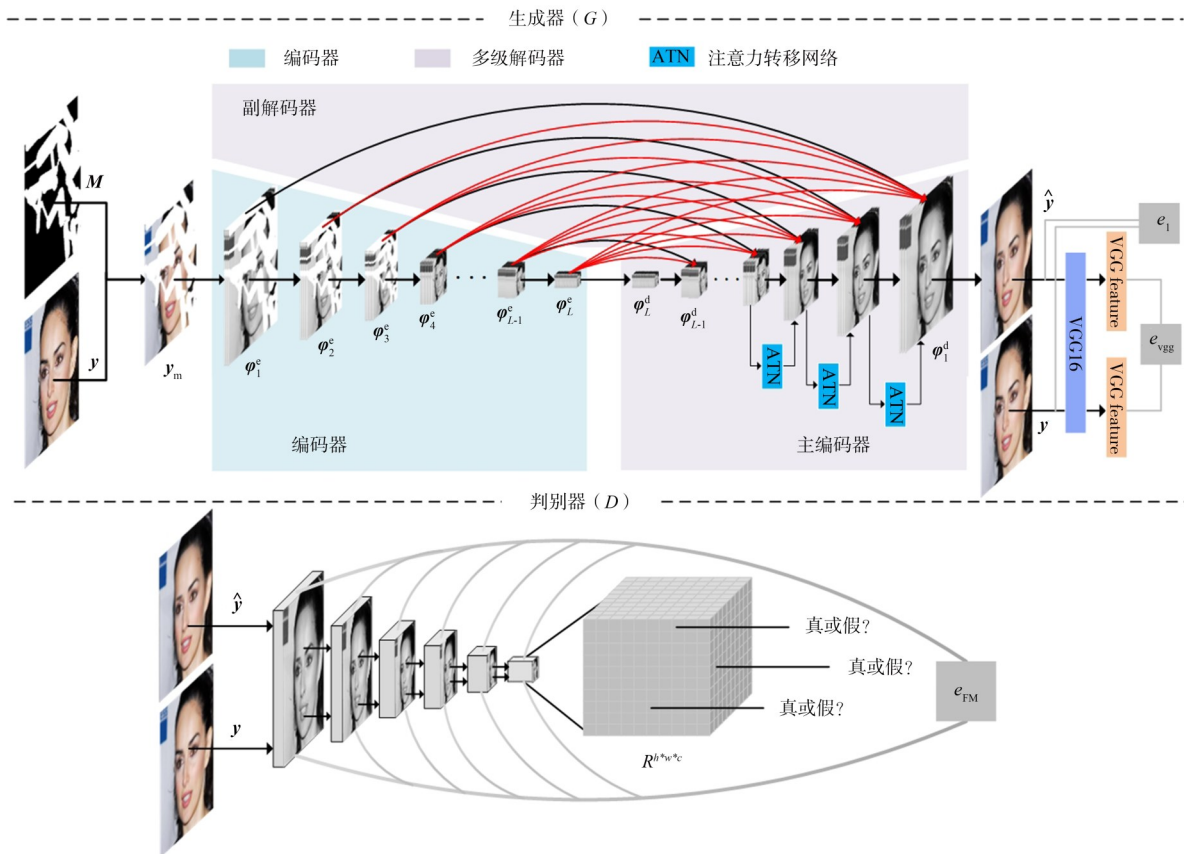
基于编解码网络的深度学习修复方法, 提取自编码阶段的不同尺度特征必然包含由浅层到深层、由

低级到高级的特征信息, 但是单一的解码器无法实现对编码阶段多尺度特征信息的综合利用. 因此本文提出多级解码网络, 其在传统编解码网络中引入多个副解码器, 将主流方法中被忽略的编码阶段信息迁移应用至主解码器中, 实现对不同尺度特征信息的完整利用. 此外, 本文以并联的方式引入注意力转移网络 (Attention Transfer Network, ATN), 通过加权的方式消除一部分错误, 保证向后传递信息的正确性, 以克服传统串联方式在搜索未缺损区域过程中匹配到错误信息时, 不可避免地将错误直接向后传播的问题.

3.1 MSDN 网络架构

如图 3 所示, 多级解码网络 MSDN 由不可或缺的生成器 G 和判别器 D 共同构成, 其中生成器 G 将缺损图像 y_m 映射为完整图像 $\hat{y}$, 判别器 D 用来判别完整图像 $\hat{y}$ 的真假. 基于经典的生成对抗原理, G 和 D 两个网络通过连续的“零和博弈”, 最终达到纳什平衡, 得到最佳的网络修复结果.

生成器 G 包含四个网络模块, 分别为编码器、主解码器、副解码器和注意力转移网络. 编码器和主解码器组成基本的编解码网络架构. 副解码器提取编码阶段



生成器 G 中, 红色曲线表示一个或多个反卷积操作, 称之为副解码器, 黑色曲线表示跳转连接.

图 3 多级解码网络的生成器和判别器体系架构.

不同尺度下所包含的语义及细节层面特征,并与主编码器对应尺度的特征相聚合.此外,考虑到接近网络末端的特征映射已恢复出大部分缺损内容,包含更多细节信息,因此将ATN并行连接到主编码器的后三层,进一步将相似块从未缺损区域匹配复制到缺损区域.

判别器 D 是一个 K 层的全卷积网络,即谱归一化马尔科夫判别器(Spectral Normalized Markovian Discriminator, SN-PatchGAN)^[23]. D 映射输出一个形状为 R^{h*w*c} 的三维特征,其中 h, w, c 分别表示高度、宽度和数量,相当于有 $h*w*c$ 个小判别器同时工作,判别结果的可靠性增加.

3.2 自动编解码网络

本文提出的自动编解码网络同样包括编码器和解码器两部分,但是不同于常见的编码、解码架构,此处将解码器划分为主解码器和副解码器,并且称之为多级解码器.

3.2.1 编码器

如图3中生成器 G 所示,编码器是一个 L 层的全卷积网络,通过连续的卷积操作对输入图像 y_m 进行逐级空间压缩,提取到特征映射 $\varphi_1^e, \varphi_2^e, \dots, \varphi_L^e$.在压缩编码过

程中,编码器不仅能逐步提取到不同尺度的上下文信息,还可依据缺损区域周围的信息初步填充缺损内容.从 φ_1^e 到 φ_L^e 依次提取到包含更多语义信息的高级特征,但同时也丢失了大量细节信息.

3.2.2 多级解码器

多级解码器主要包括一个主解码器和多个副解码器,均为反卷积网络的叠加.主解码器网络层数和编码器保持一致,也为 L ,主要作用是逐层恢复出各尺度对应的特征映射.对于一个 L 层的主解码器,各层的特征映射与编码部分相对应,分别表示为 $\varphi_L^d, \varphi_{L-1}^d, \dots, \varphi_2^d, \varphi_1^d$.主解码器第一层特征映射 φ_L^d 的形成仅依赖于编码器中最后一层 φ_L^e :

$$\varphi_L^d = \varphi_L^e \quad (3)$$

对于主解码器中间第 $j(4 \leq j < L)$ 层的特征,其聚合过程如图4所示.首先由副解码器将编码部分不同尺度的特征 $\varphi_{i+1}^e, \dots, \varphi_L^e$ 在当前尺度下解码至与目标 φ_j^d 大小相同,将所解码的特征 $f_d(\varphi_{i+1}^e), \dots, f_d(\varphi_L^e)$ 跟主解码器解码的特征 $f_d(\varphi_{j+1}^d)$ 相聚合,综合考虑主副解码器得到最终聚合后的目标 φ_j^d ,公式化如式(4).

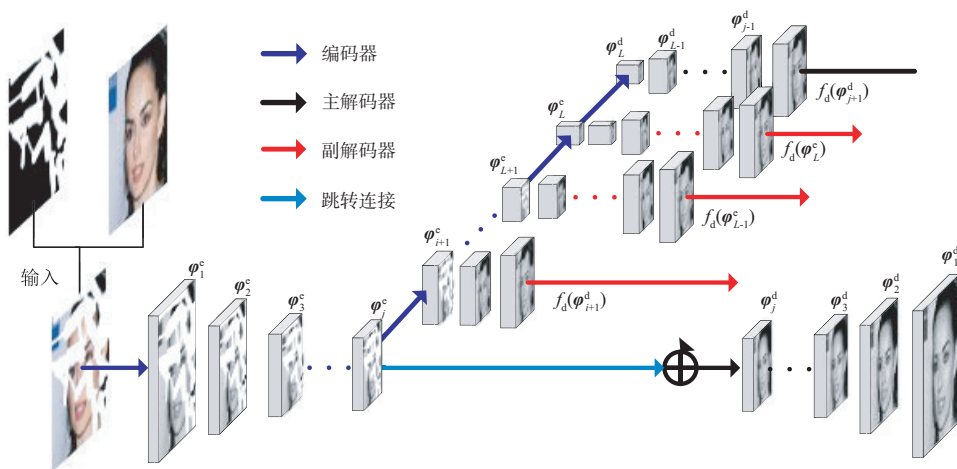


图4 不同尺度特征汇聚示意图,其中黑色箭头表示主解码过程,红色箭头表示副解码过程

$$\begin{aligned} \varphi_j^d &= \varphi_i^e \oplus f_d(\varphi_{i+1}^e) \cdots \oplus f_d(\varphi_L^e) \oplus f_d(\varphi_{j+1}^d) \\ &= \varphi_i^e \oplus f_d(\varphi_{i+1}^e, \dots, \varphi_L^e, \varphi_{j+1}^d) \end{aligned} \quad (4)$$

其中, $i \in [4, L]$, 呈递增趋势, 表示编码器中间层索引; $j \in [L, 4]$, 呈递减趋势, 为主解码器中间层索引. $\oplus$ 表示通道维叠加操作, $f_d(\cdot)$ 表示对不同尺度特征进行的反卷积操作, 即解码过程.

主解码最后三层的形成过程如图3所示, $\varphi_j^d (j=3, 2, 1)$ 在式(4)所示聚合方式的基础上, 附加融合了ATN所得新的特征表示.

$$\varphi_j^d = \varphi_j^e \oplus f_d(\varphi_L^e, \dots, \varphi_{j+1}^e, \varphi_{j+1}^d) \oplus f_a(\varphi_{j+1}^d) \quad (5)$$

其中, $f_a(\cdot)$ 表示注意力模块ATN对应的操作.

3.3 并行连接的ATN

已有方法引入注意力机制时,都是将其串联在网络层间,如CA^[15]和Pen-Net^[24],然而串联方式存在固有的弊端.如图5(a)所示,若红色区域表示匹配到的错误信息,串联方式会导致网络将错误信息逐层向后传递,进而产生错误的修复结果.因此,本文以并联方式将注意力机制引入MSDN中.如图5(b)所示,并联模式的ATN既能充分利用注意力转移网络强大的匹配复制能力,又能保证所传递信息的准确性.

注意力转移网络^[24, 25]:注意力转移网络ATN如图6所示,通过匹配计算缺损区域内外块之间的相关性,可将未缺损区域的内容加权复制到缺损区域.

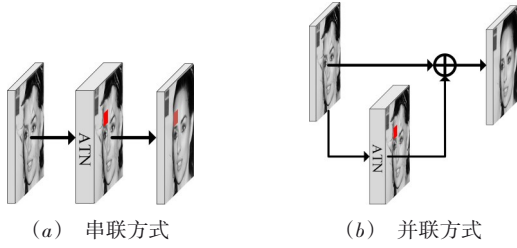


图5 ATN的连接形式

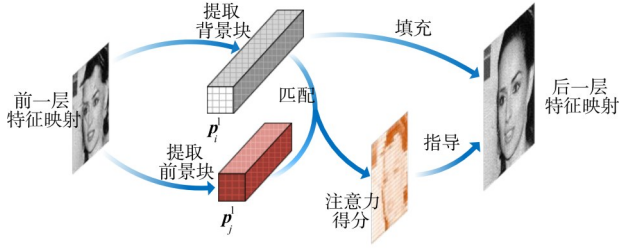


图6 注意力转移网络(ATN)

通常,缺损区域内外的相关性用余弦相似性 s'_{ij} 来衡量:

$$s'_{ij} = \left\langle \frac{\mathbf{p}_i^l}{\|\mathbf{p}_i^l\|_2}, \frac{\mathbf{p}_j^l}{\|\mathbf{p}_j^l\|_2} \right\rangle \quad (6)$$

其中, $\mathbf{p}_i^l$ 是提取自第 l 层未缺损区域的第 i 个块, $\mathbf{p}_j^l$ 是第 l 层缺损区域的第 j 个块. 求出第 l 层缺损区域内外块之间的余弦相似性 s'_{ij} 后,再用Softmax函数进行处理,得到 l 层每个块的注意力得分 $\mathbf{C}_{j,i}$, $\mathbf{C}_{j,i}$ 越大,表示对应块之间的相似性越高:

$$\mathbf{C}_{j,i} = \frac{\exp(s'_{ij})}{\sum_{i=1}^N \exp(s'_{ij})} \quad (7)$$

获得 l 层的注意力得分 $\mathbf{C}_{j,i}$ 后,便可用注意力得分加权的上下文指导填充 $l+1$ 层特征图中的漏洞:

$$\mathbf{p}_j^{l+1} = \sum_{i=1}^N \mathbf{C}_{j,i} \mathbf{p}_i^l \quad (8)$$

其中, $\mathbf{p}_j^{l+1}$ 是提取自第 $l+1$ 层缺损区域的第 j 个前景块,缺损区域总共被划分为 N 个小块.

3.4 损失函数

损失函数作为训练过程中非常关键的约束条件,主要目的是最小化原始图像和重建图像之间的差异. 为了完成训练,首先组织训练样本,给定 n 个真实样本集合 $Y = \{\mathbf{y}_i\}_{i=1}^n, \mathbf{y}_i \sim p_{\text{data}}$,通过生成网络 G 生成的样本集合为 $\hat{Y} = \{\hat{\mathbf{y}}_i\}_{i=1}^n, \hat{\mathbf{y}}_i \sim p_G$,依据 Y 和 $\hat{Y}$ 之间不同类型的差异来定义损失函数. 本文所提MSDN中,损失函数包括三部分,如图3中灰色框所示,分别为生成对抗损失 ℓ_{GAN} 、重建损失 ℓ_{rec} 和特征匹配损失 ℓ_{FM} ,各项损失函数采用加权相加的形式:

$$\ell_{\text{total}} = \ell_{\text{GAN}} + \lambda_1 \ell_{\text{rec}} + \lambda_2 \ell_{\text{FM}} \quad (9)$$

其中 λ_1, λ_2 分别为重建损失、特征匹配损失的权衡参数.

3.4.1 生成对抗损失

GAN网络在生成器 G 和判别器 D 相互博弈中进行迭代优化,生成对抗损失 ℓ_{GAN} 反映出生成网络 G 生成的图像与真实图像之间的相似程度. 对抗损失惩罚并促使生成网络 G 生成细节更为丰富的图像,可以表示为:

$$\ell_{\text{GAN}} = \mathbb{E}_{\mathbf{y}_i \sim p_{\text{data}}} [\log D(\mathbf{y}_i)] + \mathbb{E}_{\hat{\mathbf{y}}_i \sim p_G} [\log(1 - D(\hat{\mathbf{y}}_i))] \quad (10)$$

其中, p_{data} 和 p_G 分别表示真实数据分布和生成数据分布, $\mathbb{E}$ 为数学期望.

3.4.2 重建损失

为了保证图像重建前后的一致性,本文引入重建损失 ℓ_{rec} ,主要包括像素级损失 ℓ_1 和VGG损失 ℓ_{vgg} . ℓ_1 损失通过计算像素值的相似性保证像素级别的精确重建, ℓ_{vgg} 约束语义信息实现高频信息的精确表达:

$$\ell_{\text{rec}} = \ell_1 + \ell_{\text{vgg}} = \frac{1}{n} \sum_{i=1}^n (\|\mathbf{y}_i - \hat{\mathbf{y}}_i\|_1 + \|VGG_l(\mathbf{y}_i) - VGG_l(\hat{\mathbf{y}}_i)\|_1) \quad (11)$$

其中VGG是一个预训练CNN网络, $VGG_l(\cdot)$ 则为VGG16网络每个最大池化层之后第 l 个卷积层得到的特征映射.

3.4.3 特征匹配损失

为能有效解决生成对抗网络训练过程中出现的梯度消失问题,本文引入特征匹配损失 ℓ_{FM} . 特征匹配损失函数要求生成图像和真实图像在判别网络中的特征中心靠近,通过比较判别器中间层的激活映射,迫使生成器生成与真实图像相似的特征表示,从而稳定训练过程^[9, 26, 27]. 特征匹配损失 ℓ_{FM} 公式化表述如下:

$$\ell_{\text{FM}} = \frac{1}{n} \sum_{i=1}^n \sum_{k=1}^K \frac{1}{N_k} \|D_k(\mathbf{y}_i) - D_k(\hat{\mathbf{y}}_i)\|_1 \quad (12)$$

其中 K 为判别器的卷积层数, D_k 为判别器 D 第 k 层的激活结果, N_k 表示判别器 D 第 k 个卷积层激活后的元素个数.

4 实验

4.1 实验设置

4.1.1 实验平台及实验参数

本文所提方法MSDN在训练及测试中,使用的硬件平台为Intel(R) Core(TM) i7-8700 CPU (3.2 GHz)和单个的NVIDIA TITAN Xp GPU (12 GB),软件平台为TensorFlow 1.10.0.

在训练过程中,使用Adam优化器,初始学习率设置为 1×10^{-4} ,后期再将其调整为 1×10^{-5} ,用于对模型进行微调. 一阶动量和二阶动量分别被设为 $\beta_1 = 0.5, \beta_2 = 0.9$. 对于损失函数的平衡参数,经多次调参后确定为

$\lambda_1=1.2, \lambda_2=0.01$. 网络设计时, 编码器和主解码器的卷积层均为 $L=6$; 判别器中全卷积层数 K 也为 6.

4.1.2 实验数据集及对比方法

在上述软硬件平台及参数设置下, 分别在人脸数据集 Celeba-HQ^[28]、立体墙面数据集 Facade^[29]、场景图像数据集 Places2^[30] 以及自行组织的壁画数据集 Mural 上进行训练及测试. 分批将各数据集下的图像送入网络, 每批次送 6 张, 训练和测试所用图片大小均为 256×256 . 训练集及测试集的样本量划分如表 1 所示.

表 1 训练集、测试集划分

数据集	训练集	测试集	总数
Celeba-HQ ^[28]	25000	500	30000
Facade ^[29]	556	50	606
Places2 ^[30]	8020000	6628	8026628
Mural	1000	92	1092

本文采用以下四种主流的图像修复方法跟所提出的 MSDN 方法进行比较:

PatchMatch^[4]: 一个典型的基于块的方法, 从周围已知环境中寻找相似块复制到缺损区域, 被应用于常见的图像处理软件中.

CA^[15]: 一种分为粗细精度两阶段训练的模型, 采用两个编解码器串联的形式, 在细精度网络中使用上下文注意力机制.

GMCNN^[22]: 一种生成式多列神经网络架构, 采用三个编解码器并联的形式.

Pen-Net^[24]: 一种基于 U-Net^[31] 架构的金字塔式图像修复方法, 采用“由深到浅, 多次补全”的策略修复受损图像.

4.1.3 评价指标

为了客观评价先进方法 CA^[15], GMCNN^[22] 以及本文提出的 MSDN 方法的修复效果, 在相同的实验条件下, 采用以下几种客观指标来评价重建质量:

(1) 峰值信噪比 (Peak Signal-to-Noise Ratio, PSNR)^[32] 通过图像之间最大可能像素值 Z 和均方误差 MSE 来定义:

$$\begin{aligned} \text{MSE} &= \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ \text{PSNR} &= 10 \log_{10} \left(\frac{Z^2}{\text{MSE}} \right) \end{aligned} \quad (13)$$

其中, Z 取值 255, PSNR 的值通常在 20 到 40 之间, 越高表示重建图像质量越好.

(2) 结构相似度 (Structural SIMilarity, SSIM)^[33] 是在比较图像亮度、对比度特性的基础上, 衡量图像之间的结构相似度.

$$\text{SSIM}(y_i, \hat{y}_i) = \frac{(2\mu_{y_i} \mu_{\hat{y}_i} + C_1)(\sigma_{y_i \hat{y}_i} + C_2)}{(\mu_{y_i}^2 + \mu_{\hat{y}_i}^2 + C_1)(\sigma_{y_i}^2 + \sigma_{\hat{y}_i}^2 + C_2)} \quad (14)$$

其中, μ 和 σ 分别表示图像像素强度的均值和方差; $C_1 = (k_1 Z)^2$, $C_2 = (k_2 Z)^2$ 为保持稳定的常数, $k_1 \ll 1$ 且 $k_2 \ll 1$, Z 是可能的最大像素值, 通常情况下取值为 255.

(3) 弗雷切特起始距离 (Frechet Inception Distance, FID)^[34] 是计算真实图像与修复图像特征向量之间距离的性能指标. FID 分数越低, 生成图像质量越好, 与原图相似性越高.

(4) L_1 损失^[35] 通过计算重建图像与原始图像绝对差值的总和, 来评判两张图像在像素层面的相似性.

$$L_1_loss = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (15)$$

4.2 结果分析

该小节针对不同修复方法所得修复结果, 从主观、客观以及用户研究三方面来对比分析各方法的优劣. 由于 MSDN 在多级解码过程中实现了高级语义特征指导下对低级纹理细节特征的重建, 因此得到了较为满意的评价结果.

4.2.1 定性评价

4.2.1.1 不规则缺损下修复结果分析

以下分别在 Celeba-HQ、Facade、Places2 三个数据集上讨论各算法的修复性能.

在人脸数据集 Celeba-HQ^[25] 上, 使用不同算法得到的修复结果如图 7 所示. 相对于 PatchMatch^[4] 修复方法, MSDN 生成的图像语义上更加合理, 避免了 PatchMatch 因匹配错误而产生错误结果的情况. 与 CA^[15] 生成的结果相比, MSDN 基本消除了修复区域模糊和扭曲的现象, 生成结果更加平滑, 实现了从破损区域到未破损区域的完美过渡. GMCNN^[22] 在待修复区域面积较小的情况下, 修复性能优越, 但对于较大面积的缺损, 其表现出水波纹状的视觉模糊, 影响图像整体观测效果.

在立体墙面数据集 Facade^[29] 上的修复效果对比如图 8 所示. 块匹配方法 PatchMatch^[4] 在某些内容重复性较强的图像修复任务中表现出很好的性能, 如图 8(b) 上图, 但由于缺乏对图像整体语义的理解, 会出现信息匹配错误的情况, 如图 8(b) 下图所示. CA^[15] 和 GMCNN^[22] 在小面积缺损时均能重建出完整的缺失内容, 待修复区域面积较大时, 仍然会出现修复错误和模糊的情况, 如图 8(c)、(d) 所示. 本文所提方法 MSDN 整体性能较为稳定, 不会因待修复图像结构、缺损面积大小而影响到修复效果.

不同算法在自然场景图像数据集 Place2^[30] 上的修复效果如图 9 所示, 对于背景重复性自然场景图像, PatchMatch^[4] 和 MSDN 修复效果基本持平, 如图 9(b)、(e) 的上图. 但对于内容复杂的缺损图像而言, Patch-

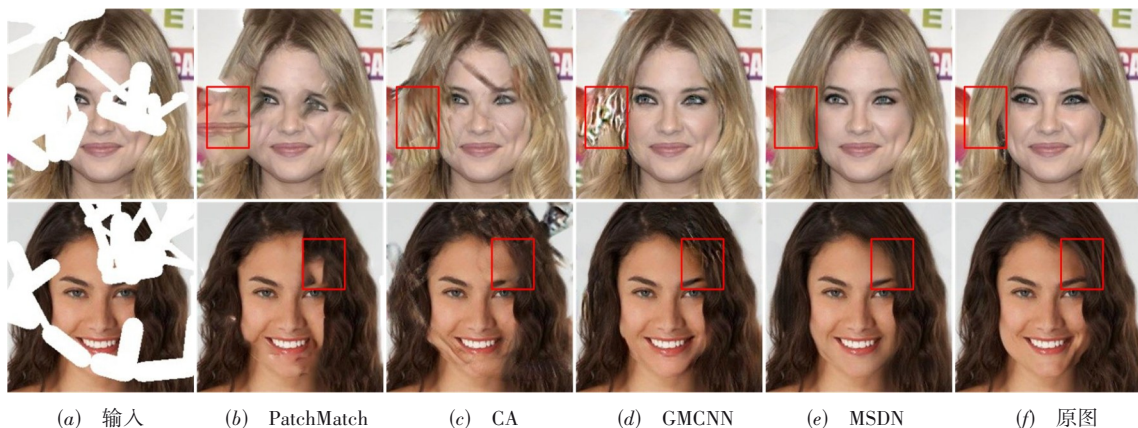


图7 不同方法在 CelebA-HQ 数据集上的重建效果图

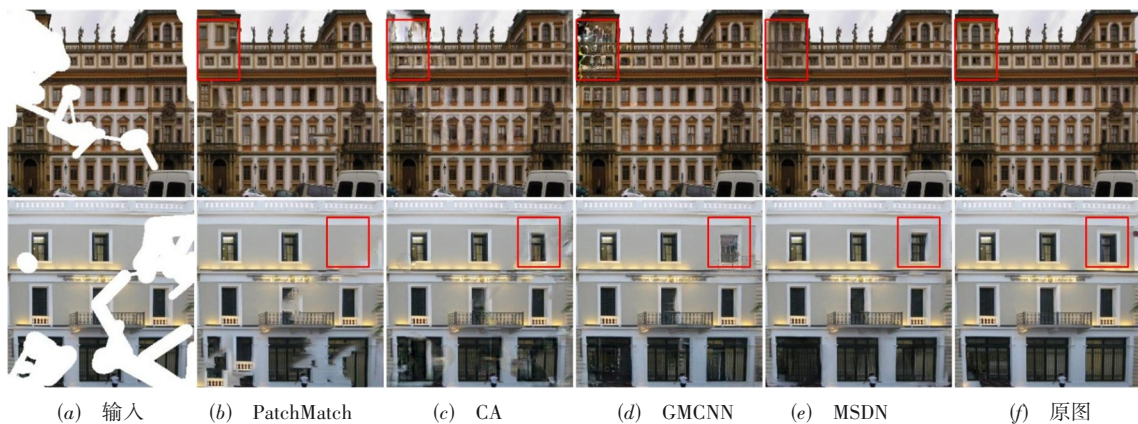


图8 不同方法在 Facade 数据集上的重建效果图,放大观察效果更佳

Match 修复性能骤减,如图9(b)下图所示,而MSDN依然可以修复出连续性较强的缺失内容,如图9(e)下图所示. CA^[15]修复区域内外一致性差的问题在 Places2 数据集中表现的依旧很明显,严重影响到图像的观测效果. 同样, GMCNN^[22]修复大面积缺损时出现的水波纹状模糊现象在此并未得以改善,如图9(c)所示.

除了在公认数据集 CelebA-HQ^[28]、Facade^[29]、Places2^[30]上进行的性能验证外,我们还将所提方法 MSDN 扩展到壁画图像修复任务中. 在自行组织的壁画数据集 Mural 上观测其重建效果,视觉展示如图10所示. 从图中可知, MSDN 在修复壁画图像时,基本实现了对缺损区域细节的重现,得到了视觉上完整、语义上合理的修复结果.

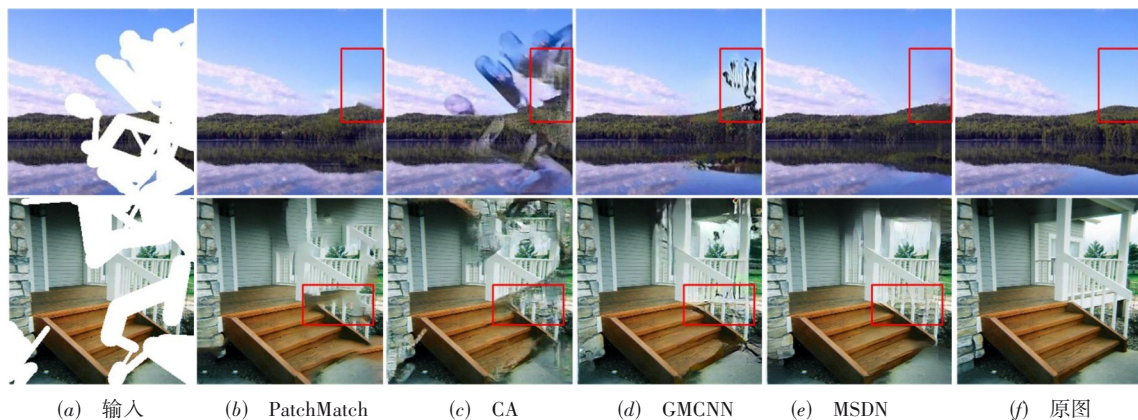
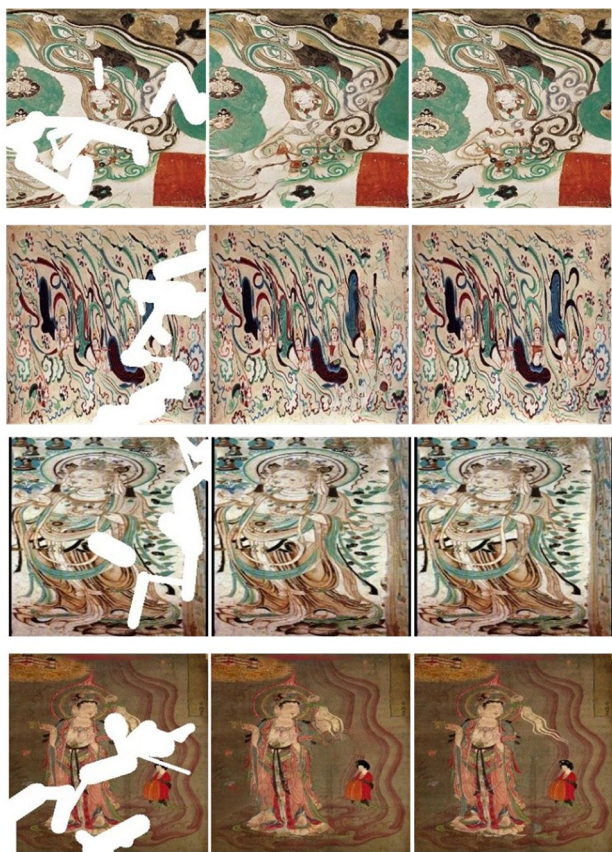


图9 不同方法在 Places2 数据集上的重建效果图



(a) 输入 (b) MSDN (c) 原图

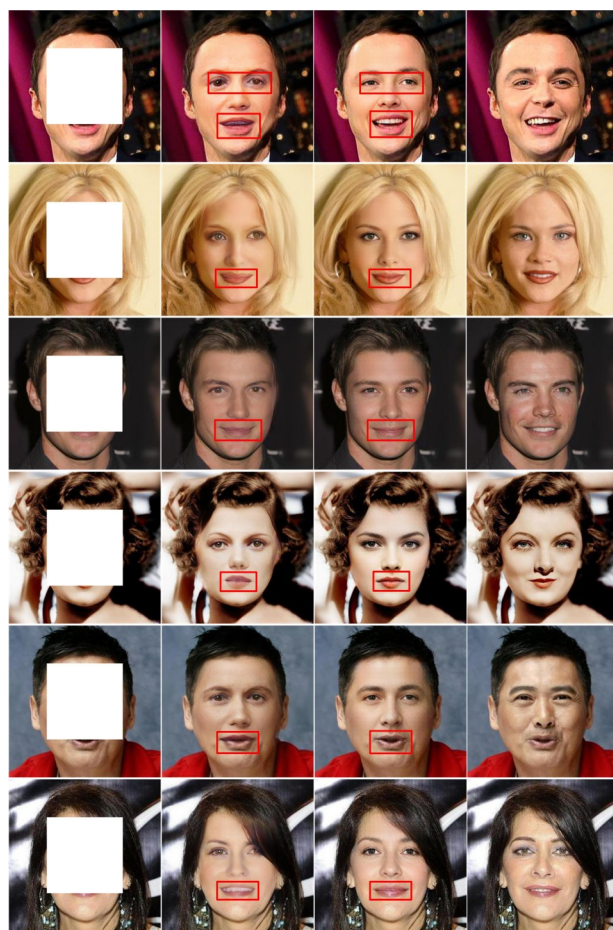
图10 MSDN在Mural数据集上的重建效果图

4.2.1.2 矩形缺损时修复结果分析

为了进一步验证本文所提方法 MSDN 在矩形缺损下的修复效果,在人脸数据集 Celeba-HQ^[28]上重新训练网络并进行相应测试,并与2019年 CVPR 中 Zeng 等人所提方法 Pen-Net^[24]进行矩形掩码下的修复效果对比.如图11所示,图11(a)~(d)依次表示待修复图像、Pen-Net^[24]修复结果、MSDN 修复结果和原图.由图可知, Pen-Net^[24]修复出的人脸图像趋于模糊,修复区域边缘出现色差、过度不连续等问题,主要表现在嘴唇部分.而 MSDN 修复的人脸图像五官清晰,色彩一致性较好,仅通过肉眼已难以分辨出原图还是修复后的图像.

4.2.2 定量评价

对于测试集中的每一张图片,我们设置了大小不同的缺损区域,即不同比例的掩码面积,并使用三种不同的图像修复方法得到对应的修复结果.为了量化模型性能,表2列出了用不同方法修复破损图像的客观评价分值,分别使用标准度量指标 PSNR, SSIM, FID, L_1_loss 来计算修复后的图像与原图之间的相似性及差异.总体而言,本文提出的 MSDN 在四个度量指标下均



(a) 输入 (b) Pen-Net (c) MSDN (d) 原图

图11 Pen-Net(b)及MSDN(c)在Celeba-HQ数据集上的重建效果图

优于 CA^[15]和 GMCNN^[22].该结果说明 MSDN 相对于主流的编解码网络,修复效果有明显提升,尤其是在人脸图像的修复中,这是因为 MSDN 加强了对结构化信息的关注度.

4.2.3 用户研究

对图像修复效果的评价一般以主观为主,因此本文进行了用户研究以量化主观评价.我们从每个测试集随机选取100张图片,在不同范围的掩码下获得三种方法(CA^[15], GMCNN^[22], MSDN)的修复结果.按照掩码大小将900张图片分三批展示给实验小组中的24个成员(有10个成员从事图像方面的研究,14个不曾接触过),每次展示三张由不同方法得到的顺序混乱的图像,在不限时的前提下让评分员对每张图片打分,分值范围0~10.

最后分批求出24个评分员对不同方法所得图像打分的均值,并列出图12所示的分值统计图.由图可知,在不同比例的缺损面积下,本文提出的 MSDN 所得结果用户认可度高.

表2 各种方法的定量评价结果(不规则掩码).

	掩码面积	CA	GMCNN	MSDN
		CelebA-HQ/Facade/Places2	CelebA-HQ/Facade/Places2	CelebA-HQ/Facade/Places2
PSNR [↑]	10~20%	25.585/24.151/22.492	29.729/27.431/27.318	31.076/27.887/27.559
	30~40%	22.948/23.151/18.502	26.651/24.934/22.725	28.375/25.362/23.316
	50~60%	21.337/21.536/17.553	24.785/23.318/21.417	26.545/23.819/22.202
	均值	23.29/22.946/19.516	27.055/25.227/23.82	28.665/25.689/24.359
SSIM [↑]	10~20%	0.9239/0.9230/0.9025	0.9486/0.9398/0.9336	0.9578/0.9443/0.9330
	30~40%	0.8786/0.8762/0.7896	0.9152/0.9020/0.8510	0.9322/0.9087/0.8528
	50~60%	0.8406/0.8328/0.7477	0.8875/0.8669/0.8170	0.9110/0.8767/0.8216
	均值	0.8810/0.8773/0.8132	0.9171/0.9029/0.8672	0.9336/0.9099/0.8691
FID [↓]	10~20%	34.036/16.880/30.179	14.594/11.055/15.841	10.057/11.137/15.582
	30~40%	53.223/27.261/50.728	27.849/17.323/27.026	16.974/17.366/24.963
	50~60%	73.466/37.111/66.768	37.526/22.782/38.028	21.841/22.313/34.569
	均值	53.575/27.084/49.225	26.656/17.053/26.965	16.290/16.938/25.038
L_1 Loss [↓]	10~20%	0.0406/0.0399/0.0594	0.0228/0.0291/0.0307	0.0196/0.0280/0.0299
	30~40%	0.0669/0.0647/0.1313	0.0382/0.0475/0.0711	0.0319/0.0459/0.0671
	50~60%	0.0901/0.0883/0.1589	0.0522/0.0645/0.0889	0.0434/0.0623/0.0823
	均值	0.0658/0.0634/0.1165	0.0377/0.047/0.0635	0.0316/0.0454/0.0576

↑表示越大越好, ↓表示越小越好. 每组实验最好的评价结果已在表中用粗体标出.

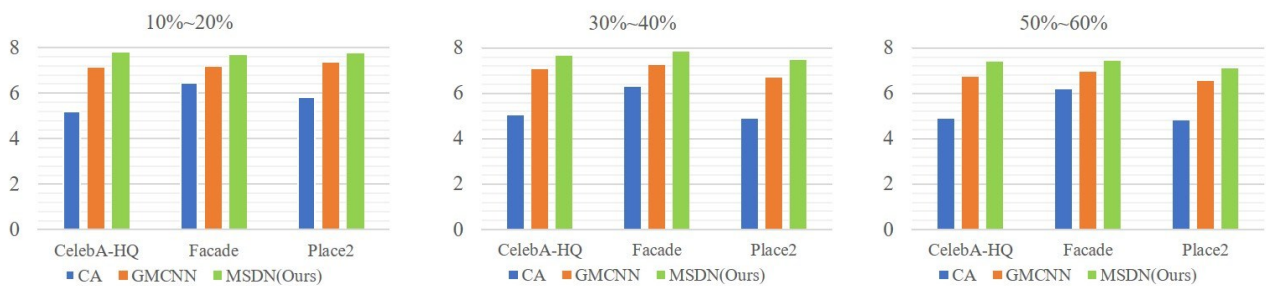


图12 用户研究结果统计图

4.3 消融对比实验

为了进一步说明并行连接 ATN 的有效性, 分别组织实验验证去除 ATN 以及串行连接 ATN 时的修复效果, 并与本文中的并行连接方式所得结果进行对比, 如图 13 所示. 其中, 图 13(a) 为输入的待修复图像, 图 13(b)~(d) 分别表示无 ATN 时的修复结果及串行、并行连接 ATN 后的修复结果, e 列为原始图像. 由图 13(b) 可知, 无 ATN 时, 修复结果虽然具备较好的内外一致性,

但在缺损严重的区域会出现模糊现象, 如图 13(b) 上图的修复结果, 在修复内容较为复杂时, 甚至会表现出修复结果错误、扭曲的现象, 如图 13(b) 下图的修复结果. 串行连接 ATN 后, 网络整体性能有一定提升, 但模糊和结果错误的问题并未得到彻底解决, 如图 13(c) 所示. ATN 的并联方式通过抵消一部分匹配到的错误信息, 有效解决了无 ATN 及串联 ATN 时出现的修复区域模糊、匹配信息错误的问题, 如图 13(d) 所示.

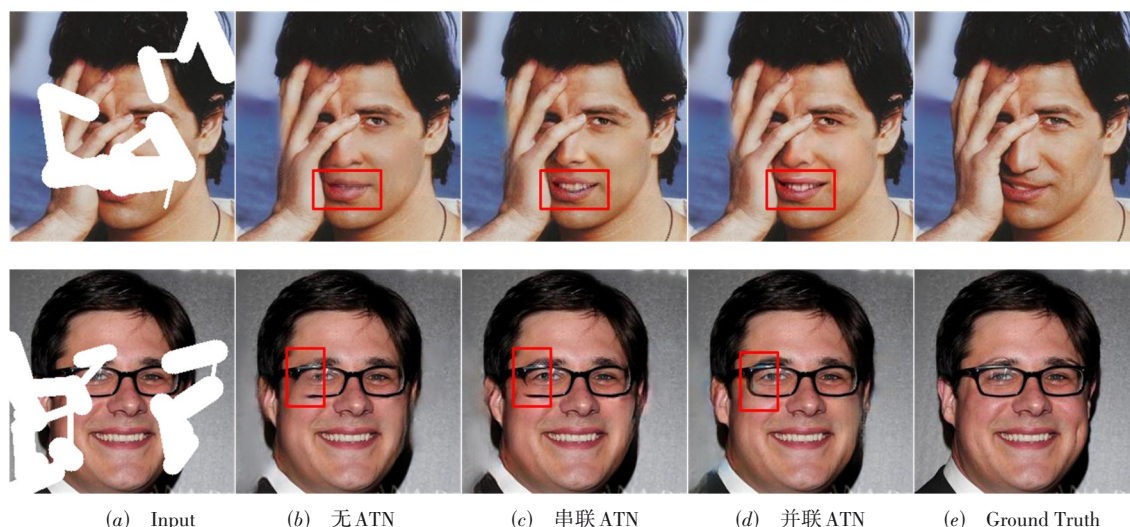


图 13 注意力转移网络消融对比实验效果图

5 结论

本文提出了一种多级解码网络 MSDN, 由副解码器对编码阶段各尺度特征进行解码, 获得不同尺度的特征表示, 并将其聚合至主解码器, 共同指导下一级特征的重建。此外, 本文以并联的方式将注意力机制引入主解码器, 重建出更精确、视觉效果更佳的预测结果。通过大量实验证明, MSDN 可有效生成细节丰富、边缘过渡平滑、视觉上逼真的完整图像。

参考文献

- [1] ELHARROUSS O, ALMAADEED N, et al. Image inpainting: a review[J]. *Neural Processing Letters*, 2020, 51(2): 2007-2028.
- [2] Coloma Ballester, BERTALMIO M, CASELLES V, et al. Filling-in by joint interpolation of vector fields and gray levels[J]. *IEEE Transactions on Image Processing*, 2001, 10(2): 1200-1211.
- [3] ASHIKHMIN M. Synthesizing natural textures[C]//Proceedings of the 2001 symposium on Interactive 3D graphics - SI3D 01. New York: ACM Press, 2001: 217-226.
- [4] CONNELLY B, ELI S, ADAM F, et al. Patchmatch: a randomized correspondence algorithm for structural image editing[J]. *ACM Transactions on Graphics*, 2009, 28(3): 1-11.
- [5] Ruzic Tijana, Pizurica Aleksandra. Context-aware patch-based image inpainting using markov random field modeling[J]. *IEEE Trans Image Process*, 2015, 24(1): 444-456.
- [6] 李志丹, 和红杰, 尹忠科, 等. 基于 curvelet 方向特征的样本块图像修复算法[J]. *电子学报*, 2016, 44(1): 150-154.
LI Zhi-dan, HE Hong-jie, YIN Zhong-ke, et al, Exemplar based image inpainting algorithm using direction features of curvelet transform[J]. *Acta Electronica Sinica*, 2016, 44(1): 150-154. (in Chinese)
- [7] 吴晓军, 李功清. 基于样本和线性结构信息的大范围图像修复算法[J]. *电子学报*, 2012, 40(8): 1509-1514.
WU Xiao-jun, LI Gong-qing. Large scale image inpainting based on exemplar and structure information[J]. *Acta Electronica Sinica*, 2012, 40(8): 1509-1514. (in Chinese)
- [8] 强振平, 何丽波, 陈旭, 等. 深度学习图像修复方法综述[J]. *中国图象图形学报*, 2019, 24: 447-463.
- [9] 曹承瑞. 基于密集判别与注意力特征传播的细粒度图像修复[D]. 兰州: 兰州理工大学, 2020.
- [10] 王贵杭. 基于几何信息的图像修复[D]. 合肥: 中国科学技术大学, 2015.
- [11] YANG C, LU X, LIN Z, et al. High-resolution image inpainting using multi-scale neural patch synthesis[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2017: 4076-4084.
- [12] Goodfellow Ian J, Jean Pouget-Abadie, Mehdi Mirza, et al. Generative adversarial nets[C]// Proceedings of the 27th International Conference on Neural Information Processing Systems-Volume 2. Cambridge, MA, US: MIT Press, 2014: 2672-2680
- [13] PATHAK D, KRÄHENBÜHL P, DONAHUE J, et al. Context encoders: feature learning by inpainting[C]//2016 IEEE Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2016: 2536-2544.
- [14] IIZUKA S, SIMO-SERRA E, ISHIKAWA H. Globally and locally consistent image completion[J]. *ACM Transactions on Graphics*, 2017, 36(4): 1-14.

- [15] YU J H, LIN Z, YANG J M, et al. Generative image inpainting with contextual attention[C]//2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2018: 5505-5514.
- [16] 罗会兰, 敖阳, 袁璞. 一种生成对抗网络用于图像修复的方法[J]. 电子学报, 2020, 48(10): 1891-1898.
LUO H L, AO Y, YUAN P. Image inpainting using generative adversarial networks[J]. Acta Electronica Sinica, 2020, 48(10): 1891-1898. (in Chinese)
- [17] 王斌, 胡辽林, 曹京京, 等. 基于小波域稀疏最优的图像修复方法[J]. 电子学报, 2016, 44(3): 600-606.
WANG B, HU L L, CAO J J, et al. Image restoration based on sparse-optimal strategy in wavelet domain[J]. Acta Electronica Sinica, 2016, 44(3): 600-606. (in Chinese)
- [18] 熊锐, 张雷洪, 蒋周杰, 等. 基于编码-解码对称神经网络的高分辨率图像重构机理研究[J]. 光学仪器, 2019, 41(4): 36-41.
XIONG R, ZHANG L H, JIANG Z J, et al. Research on high-resolution image reconstruction mechanism based on coding-decoding symmetric neural network[J]. Optical Instruments, 2019, 41(4): 36-41. (in Chinese)
- [19] 魏域林. 层间特征融合与多注意力的图像修复算法研究[D]; 兰州理工大学, 2020.
- [20] 李志丹, 程吉祥, 刘家伟. 基于结构偏移映射统计和多方向特征的MRF图像修复算法[J]. 电子学报, 2020, 48(5): 985-989.
LI Z D, CHENG J X, LIU J W. MRF image inpainting algorithm based on structure offsets statistics and multi-direction features[J]. Acta Electronica Sinica, 2020, 48(5): 985-989. (in Chinese)
- [21] LIU G L, REDA F A, SHIH K J, et al. Image inpainting for irregular holes using partial convolutions[M]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 89-105.
- [22] WANG Yi, TAO Xin, QI Xiao-juan, et al. Image inpainting via generative multi-column convolutional neural networks[C]//Proceedings of the 32nd International Conference on Neural Information Processing Systems. Red Hook, NY: Curran Associates Inc, 2018: 329-338.
- [23] YU J H, LIN Z, YANG J M, et al. Free-form image inpainting with gated convolution[C]//2019 IEEE/CVF International Conference on Computer Vision (ICCV). Piscataway, NJ: IEEE, 2019: 4470-4479.
- [24] ZENG Y H, FU J L, CHAO H Y, et al. Learning pyramid-context encoder network for high-quality image inpainting [C]//2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ: IEEE, 2019: 1486-1494.
- [25] 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). Piscataway, NJ: IEEE, 2020: 7757-7765.
- [26] 鲍建敏. 基于生成对抗网络的图像合成[D]. 合肥: 中国科学技术大学, 2019.
BAO J M. Image Synthesis Based on Generative Adversarial Networks[D]. Hefei: University of Science and Technology of China, 2019. (in Chinese)
- [27] Kamyar Nazeri, Eric Ng, Tony Joseph, et al. Edgeconnect: generative image inpainting with adversarial edge learning[EB/OL]. (2019-01-11). <https://arxiv.org/abs/1901.00212v3>.
- [28] KARRAS T, AILA T, LAINE S, et al. Progressive growing of GANs for improved quality, stability, and variation [EB/OL]. (2018-02-26). <https://arxiv.org/abs/1710.10196>.
- [29] Radim Tyleček, Radim Šára. Spatial pattern templates for recognition of objects with regular structure[C]//German Conference on Pattern Recognition. Berlin: Springer, 2013: 364-374.
- [30] Zhou Bo-lei, Agata Lapedriza, Aditya Khosla, et al. Places: a 10 million image database for scene recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 40(6): 1452-1464.
- [31] Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-Net: convolutional networks for biomedical image segmentation[C]// Medical Image Computing and Computer-Assisted Intervention-MICCAI 2015, Part III. Berlin: Springer, 2015: 234-241.
- [32] HORÉ A, ZIOU D. Image quality metrics: PSNR vs. SSIM[C]//2010 20th International Conference on Pattern Recognition. Piscataway, NJ: IEEE, 2010: 2366-2369.
- [33] WANG Zhou. Image quality assessment : from error visibility to structural similarity[J]. IEEE Transactions on Image Processing, 2004, 13(4): 600-612.
- [34] HEUSEL M, RAMSAUER H, UNTERTHINER T, et al. GANs trained by a two time-scale update rule converge to a local Nash equilibrium[C]//NIPS 17: Proceedings of the 31st International Conference on Neural Information Processing Systems. Red Hook, NY, US: Curran Associates Inc, 2017: 6629-6640.

- [35] GAO Xiao-li, FANG Yi-xin. A note on the generalized degrees of freedom under the L1 loss function[J]. Journal of Statistical Planning & Inference, 2011, 141(2): 677-686.

作者简介



刘微容 男, 1976年12月出生于湖南祁阳, 兰州理工大学教授, 主要研究方向为机器视觉与人工智能、复杂系统先进控制理论与应用等.
E-mail: liu_weirong@163.com



米彦春 女, 1994年6月出生于甘肃定西, 兰州理工大学硕士研究生, 主要研究方向为机器视觉与人工智能.



杨帆 男, 1996年7月出生于湖北宜昌, 兰州理工大学硕士研究生, 主要研究方向为机器视觉与人工智能.



张彦 女, 1992年6月出生于甘肃白银, 兰州理工大学硕士研究生, 主要研究方向为机器视觉与人工智能.



郭宏林 男, 1977年4月出生于甘肃天水, 天水电气传动研究所集团有限公司高级工程师, 主要从事电气传动系统、高精度电源及数字控制系统、智能装备等领域研究.



刘仲民 男, 1978年5月出生于甘肃白银, 兰州理工大学副教授, 主要研究方向为机器视觉、模式识别与图像处理.