计算机系统应用 ISSN 1003-3254, CODEN CSAOBN
Computer Systems & Applications,2021,30(1):122−128 [doi: 10.15888/j.cnki.csa.007742]
©中国科学院软件研究所版权所有.
E-mail: csa@iscas.ac.cn
http://www.c-s-a.org.cn
Tel: +86-10-62661041

基于新编码器和相似度约束的图像修复①

林　竹, 王　敏

(河海大学 计算机与信息学院, 南京 211100)
通讯作者: 林　竹, E-mail: 18860402972@163.com

摘　要: 现有的图像修复方法存在受损区域修复痕迹明显、语义不连续、不清晰等问题, 针对这些问题本文提出了一种基于新型编码器并结合上下文感知损失的图像修复方法. 本文方法采用生成对抗网络作为基本网络架构, 为了能够充分学习图像特征得到更清晰的修复结果, 引入了 SE-ResNet 提取图像的有效特征; 同时提出联合上下文感知损失训练生成网络以约束局部特征的相似性, 使得修复图像更加接近原图且更加真实自然. 本文在多个公共数据集上进行实验, 证明了本文所提方法能够更好地对破损图像进行修复.

关键词: 生成对抗网络; 图像修复; 残差网络; 上下文感知损失

引用格式: 林竹,王敏.基于新编码器和相似度约束的图像修复.计算机系统应用,2021,30(1):122–128. http://www.c-s-a.org.cn/1003-3254/7742.html

Image Inpainting Based on New Encoder and Similarity Constraint

LIN Zhu, WANG Min

(College of Computer and Information, Hohai University, Nanjing 211100, China)

Abstract: The existing image repair methods have some problems such as obvious trace, semantic discontinuity, unclear, etc. To solve these problems, this study proposes an image repair method based on a new encoder and context-aware loss. In this paper, the generative adversarial network is adopted as the basic network architecture. In order to fully learn the image features and get clearer repair results, SE-ResNet is introduced to extract the effective features of the image. At the same time, the joint context-aware loss training generating network is proposed to constrain the similarity of local features, so that the repaired image is closer to the original and more real and natural. Experiments on multiple public datasets in this paper prove that the proposed method can repair the damaged images better.

Key words: generative adversarial network; image inpainting; residual network; contextual loss

图像修复是在保证图像的清晰度和它的语义连贯性的前提下, 利用图像中已有的信息, 修补图像中的缺失信息或者移除图像中的特定信息的图像处理技术. 此技术的核心挑战在于为缺失的区域合成视觉逼真和语义上合理的像素, 以便与现有的像素保持一致. 图像修复具有重要的现实意义, 尤其在艺术作品的保护, 修复老照片以及基于图像的渲染和计算机摄影方面有很多应用.

传统的图像修复方法多采用图像级特征来处理修复问题. 块匹配 (patch-match)[1] 方法提出寻找最佳匹配的补丁块来重建缺失区域. 此方法在本质上可以提供真实的纹理信息, 但是他们仅仅是利用给定上下文的低级特征, 缺乏预测上下文高层次特征的能力.

为了更好地提取语义层面的信息, 也得益于神经网络和深度学习的不断发展, 更多的基于深度学习的方法被提出. 上下文编码器 (context-encoder)[2] 是语义修复的一个开创性方法, 给定一个缺失区域的掩码, 训练一个神经网络对上下文信息进行编码并对受损内容

① 收稿时间: 2020-05-23; 修改时间: 2020-06-16; 采用时间: 2020-06-30; csa 在线出版时间: 2020-12-31

© 中国科学院软件研究所 http://www.c-s-a.org.cn

进行预测, 同时利用对抗损失来使得生成的图像尽可能真实, 能够较好地修复缺失区域. 但是此方法仍然缺乏一些修复细节并且存在明显的修复痕迹.

为了能够修复更高分辨率的图像, Iizuka 等人提出添加膨胀卷积层来增加感受野并且联合局部和全局判别器提升修复的一致性[3]. 但是此方法无法修复缺失较大的图像, 尽管视觉上修复纹理较连续, 但是存在某些高度结构化的图片修复后存在模糊和不真实的情况, 有较明显的修复痕迹.

为了解决这些问题, 该文提出了基于 SE-ResNet[4,5] 并联合上下文感知损失的方法. 网络结构由 4 个部分组成: 生成网络, 上下文感知损失 (contextual loss)[6] 网络以及全局上下文判别网络, 局部上下文判别网络. 生成网络是一个基于 SE-ResNet 的全卷积神经网络用生成完整的图像, 全局和局部上下文判别网络都是作为训练时的附属网络, 并且也是添加了 SE-ResNet 残差块的网络, 用来判别生成的图像是否连续且真实. 上下文感知损失网络是一个基于 VGG16[7] 的特征提取网络, 用来辅助从相似性上修复图像.

综上所述, 本文提出:

(1) 在生成网络和全局上下文以及局部上下文判别网络部分添加了基于 SE-ResNet 的残差块更好的提取特征.

(2) 增加了上下文感知损失网络以辅助约束局部高频特征的相似性来修复图像.

1 相关内容

传统的图像修复方法是利用待修补区域周边来传播纹理信息, 这些方法只能修复缺失较小的图像而且存在明显的修复痕迹. 基于补丁块的方法通过搜索最佳的匹配块来优化修复效果, 但是这些方法只能提供已有的纹理信息, 不能从语义结构上理解图像, 也不能对物体结构做出合理的预测.

随着深度生成式模型的快速发展, 尤其是生成对抗网络 (GAN)[8], GAN 在生成高分辨率图像方面有着可观的表现. 但是训练一个生成高清图像并且稳定的 GAN 网络是相当困难的. 近期也提出了一些稳定训练的方法, 例如 DC-GAN, WGAN-GP, Wasserstein GAN (WGAN), LSGAN 和 Progressive GANs[9]. 这些方法为稳定训练提供了实用的技巧, 能够产生更加真实的高清图像.

由 Pathak 等人[2] 提出的 Context-encoder 基于生成对抗网络 (GAN), 扩大了可修复缺失范围. 其中的生成网络是通过判别器辅助训练, 判别器是用来判别生成器生成的图像是否为真.

由 Iizuka 等人[3] 提出的基于全局和局部一致的图像修复在 Context-encoder 的基础上引入了全局和局部判别器, 全局判别器判别生成的图像整体的连贯性, 同时局部判别器专注于生成图片的小区域以保证局部一致性. Iizuka 等人还在生成网络中加入了膨胀卷积层代替 Context-encoder 当中的全连接层, 这些操作都是为了扩大网络的感受野.

Yu 等人[10] 提出一种端到端的图像修复模型, 通过采用一种堆叠型的生成网络确保与周边颜色以及纹理的连贯性, 同时引入了注意力模块从距离较远的区域提取近似待修复区域的特征.

Liu 等人[11] 提出通过在卷积过程中更新掩膜并使用更新的掩膜值归一化卷积核的权重值, 保证卷积核能够专注于有效的像素值.

Yu 等人[12] 通过引入门控卷积, 学习一种特征通道的动态选择机制, 以提高色彩的一致性, 同时提出一种高效的判别器 SN-PatchGAN 用于辅助修复随机缺失的图像.

在生成网络的编码部分提取图像的特征得到低分辨率的特征图, 特征图中编码了丰富的语义信息, 然而在下采样过程当中丢失了图像的结构和语义信息, 使得修复效果不清晰. 该文通过在编码部分添加基于 SE-ResNet 的残差块来更好地提升网络特征利用以得到清晰的图像; 同时在判别器部分也添加基于 SE-ResNet 的残差块, 目的是加深判别器网络深度, 提升网络特征利用, 使得判别器更好地判别生成图像与真实图像. 为了在语义上理解并修复图像, 该文增加了上下文感知损失网络约束局部特征的相似性, 然后使用此损失来训练生成网络.

2 网络结构

2.1 SE-ResNet

近年来卷积神经网络在众多领域上都取得了巨大的突破. 卷积神经网络由一系列卷积层、非线性层和下采样层构成, 这样能够从全局感受野上去捕获图像的特征来进行图像的描述. SENet (Squeeze-and-Excitation Networks) 采用一种全新的特征重标定策略, 主要通过显示地建模特征通道之间的相互依赖关系. 具体来说

© 中国科学院软件研究所 http://www.c-s-a.org.cn

就是通过学习的方式来自动获取到每个特征通道的重要程度, 然后依照这个重要程度去提升有用的特征并抑制对当前任务用处不大的特征, 这点能为图像修复生成网络的编码器以及判别器部分提取有效特征带来巨大的作用.

假设网络的输入是 x (通道数为 $c1$), 通过一系列的卷积等操作过后得到通道为 $c2$ 的特征. 那么 SENet 所做的就是使用新的方式来得到通道为 $c2$ 的特征. 第一个操作就是 Squeeze, 即特征压缩, 主要采用平均池化操作来进行特征压缩, 在每一个通道上得到一个一维的实数, 这个实数表示在特征通道上相应的全局分布. 紧接着是 Excitation, 即激活, 这个操作类似于循环神经网络中的门机制. 通过一个全连接层得到特征每个通道的权重, 以此来显式的建模通道间的相关性. 最后通过 Reweight 操作将上一层输出的权重通过乘法加权到 x 的每个通道上, 因为上一层输出的权重代表着每个通道的重要程度, 这时就完成了对原始特征的重新标定. ResNet 已经广泛被用做提取特征, 同时能够减少参数. 在本文中我们将 SENet 嵌入到 ResNet 残差块中作为一个 SE-ResNet 模块添加到生成网络和判别网络中.

2.2 生成网络

生成网络是一个基于全卷积的网络. 生成网络的输入是一张三通道的图像, 输出是一张相同大小的修复后图像. 生成网络采用了编解码的结构, 在进一步处理图像之前先降低分辨率, 从而减少内存使用和计算时间. 降低分辨率需要使用下采样, 下采样的过程很容易丢失图像的结构和语义信息, 使得修复效果不清晰, 因此仅仅对图像做了两次下采样, 并且添加了基于 SE-ResNet 的残差网络块 (见 3.1 节) 代替传统的下采样方式. 当特征图降为原来的 1/4 时, 使用 4 个膨胀卷积层作为编码-解码部分的连接处理[13], 来保持图像的感知野. 膨胀卷积在获取高分辨率最终特征图中发挥重要作用. 生成网络的结构如图 1 所示.

生成网络的损失函数分为 3 个部分: 内容损失, 对抗性损失和上下文感知损失.

对抗性损失是为了生成器尽最大可能“欺骗判别器”, 对抗性损失定义如下其中 G 代表生成网络, x 代表完整图片, M为图像掩膜, 生成网络的输入为破损图像; D 代表判别网络, 输入为生成的修复图片:

$$L_{\text{adv}} = -E_{x\sim p_r(x)}D(G(M\odot x)) \tag{1}$$

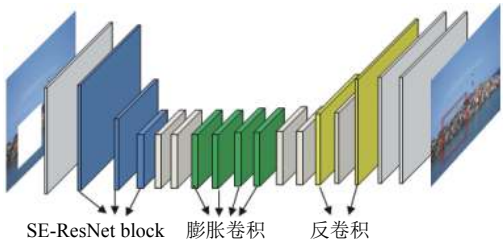

图 1 生成网络结构图

内容损失采用 L2 距离, 内容损失是为了生成尽量真实的图像, 尽最大可能使得待修复完成的图像与原图像“距离”更小. 假设M为图像掩膜, 其与输入图像具有同样大小的尺寸, 其中值为 0 的部分代表破损像素点, 值为 1 的部分代表未破损像素点, 则定义内容损失为如下的形式:

$$L_{\text{res}} = -E_{x\sim p_r(x)}[\|M\odot(x-G(M\odot x))\|_2] \tag{2}$$

其中, 符号$\odot$代表两个矩阵的对应元素相乘.

上下文感知损失 (见 3.4 节) 采用 VGG16 特征提取器, 得到生成图像和原始图像的特征图, 通过特征图计算相似度作为损失来对生成网络进行训练, 上下文感知损失是为了让生成网络具有能够约束局部特征的相似性的能力, 使图像修复得更加逼真, 与原图更加相似. 其形式如下, 其中Φ表示 VGG16 网络, CX 为计算相似度的函数:

$$L_{CX} = -\log[CX(\Phi(x),\Phi(G(M\odot x)))] \tag{3}$$

最后生成器的部分的损失函数如下, 其中λ_1 和 λ_2分别表示两个损失的系数:

$$L_{\text{res}} + \lambda_1 L_{\text{adv}} + \lambda_2 L_{CX} \tag{4}$$

2.3 判别网络

判别网络是由全局上下文判别网络和局部上下文判别网络组成, 分别对应整体图像和局部图像, 它们的作用就是判别生成图像是真还是假. 为了让判别器具有更好的判别能力, 该文在两个判别网络中都添加了基于 SE-ResNet 的残差网络块, 使其能够提升特征利用. 全局和局部的网络会将图像压缩成小的特征向量, 然后连接起来作为最终结果去判断图像真假. 判别网络如图 2 所示, 其中全局和局部判别网络的倒数第 2 至倒数第 4 层是 SE-ResNet 残差块.

全局上下文判别器将整个输入的图裁剪到 160×160 大小并作为网络输入, 它包括 3 个普通卷积层和 3 个 SE-ResNet 残差块以及一个全连接层, 输出为一个 1024 维的向量, 表示全局上下文信息. 其中前 4 层

© 中国科学院软件研究所 http://www.c-s-a.org.cn

普通卷积层和残差块均采用步幅 2 来减少图像大小并双倍增加输出的通道数 (第一层卷积输出通道数为 64), 最后两层通道数不变.

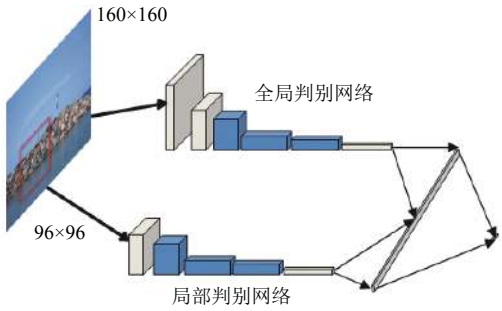

图 2 判别网络结构图

局部上下文判别器与全局类似, 局部网络的输入为 96×96 大小的图像, 此图像包括了缺失区域, 但不全是缺失区域, 还有一部分是未缺失图像. 由于输入只是全局网络输入的缺失部分, 该文去掉全局网络中的第一层作为局部网络. 输出同样为一个 1024 维的向量, 表示局部上下文信息.

最后全局和局部上下文判别网络的输出被拼接成一个 2048 维的向量, 然后送入一个全连接层并输出一个实数, 最后经过一个 Sigmoid 函数将其映射到 [0, 1] 范围代表图像真假的概率.

判别器的损失如下:

$$L_{\text{dis}} = -E_{x\sim p_r}[\log(D(x)) + \log(1 - D(G(x)))] \tag{5}$$

2.4 上下文感知损失

对于生成网络, 我们希望它具有能够约束局部特征的相似性, 使图像修复得更加逼真. 从近期的图像修复所使用的损失来看, 普遍使用的是像素级的损失即要求修复后的图像与完整图像对齐[10–12], 这些方法都不利于生成一个完整而又清晰的修复结果. 因此该文提出了联合上下文感知损失网络训练生成网络, 将图像视为特征的集合, 忽略特征的空间位置, 不要求图像完全对齐, 允许局部形变, 并根据特征之间的相似性度量图像之间的相似性. 让生成的图像和原始图像经过 VGG16 特征提取器, 得到图像的特征图, 通过特征图计算相似度作为损失来对生成网络进行训练.

上下文感知损失网络接收两张 160×160 大小的图像, 经过已经训练好的 VGG16 后输出感知损失. 假设 x 为输入图像, 则表示生成网络, 表示 VGG16 网络, 表示计算相似度的函数, 那么上下文感知网络的损失可以表示为:

$$L_{CX} = -\log[CX(\Phi(x), \Phi(G(M \odot x)))] \tag{6}$$

更进一步的, 对于两个输入的图像 x, y 经过 VGG16 的特征提取后的 x_i, y_j, 其中 CX 函数计算两张图像的相似性函数如下, 对于每个特征 y_j, 找到与它最相似的特征 x_i, y_j, 然后对所有 y_j 求和相应的特征相似值:

$$CX(x, y) = CX(X, Y) = \frac{1}{N}\sum_j \max_i CX_{ij} \tag{7}$$

其中,

$$CX_{ij} = w_{ij} \Big/ \sum_k w_{ik} \tag{8}$$

其中, w_{ij} 表示特征 x_i 与 y_j 的相似性, 通过下式计算得到.

$$w_{ij} = \exp\left(\frac{1 - d_{\text{similar}}}{h}\right) \tag{9}$$

上式通过求幂从距离转换到相似性, 距离由下式计算得到.

$$d_{\text{similar}} = \frac{d_{ij}}{\min_k d_{ik} + \varepsilon} \tag{10}$$

对 d_{ij} 归一化, 其中 d_{ij} 为 x_i 与 y_j 的余弦距离. 上述 d_{ij} 计算公式为:

$$d_{ij} = \left(1 - \frac{(x_i - \mu_y)\cdot(y_j - \mu_y)}{\|x_i - \mu_y\|_2 \|y_j - \mu_y\|_2}\right) \tag{11}$$

其中,

$$\mu_y = \frac{1}{N}\sum_j y_j \tag{12}$$

在训练过程中, 通过不断的减小此损失 (L_{CX}) 来优化生成网络, 生成网络因此而具有约束局部特征的相似性的功能.

3 实验

3.1 数据集

本文使用来自香港中文大学的开放数据集 CelebA[14] 和 LFW[15]. CelebA 是一个大型的人脸属性数据集, 包含 1 万多个名人身份的 20 多万张图片. LFW 数据集是一个无约束自然场景人脸识别数据集, 该数据集由 13 000 多张全世界知名人士互联网自然场景环境人脸图片组成.

3.2 训练过程

在实验中, 将 CelebA 的 12 万张图片作为训练集,

© 中国科学院软件研究所 http://www.c-s-a.org.cn

剩下的图片作为测试集, 将式 (4) 中λ_1设置为 0.0004, λ_2设置为 0.004, 将 batch size 设置为 12, 图片被裁剪成 160×160 像素大小然后被送进网络训练. 首先单独对生成网络训练 90000 次, 然后对判别网络训练 10000 次, 最后联合生成网络, 判别网络训练 400 000 次. 对于 LFW 数据集我们将其中的 1 万多张图片作为训练集剩下的图片作为测试集, 将式 (4) 中λ_1设置为 0.0004, λ_2设置为 0.004, 将 batch size 设置为 16, 图片被裁剪成 160×160 像素大小, 训练步骤与上述相同.

3.3 SE-ResNet 的效果分析

在 3.1 节该文从理论上分析了基于 SE-ResNet 残差块的生成网络和判别网络的优势, 为了进一步从实验上证明, 该文使用了 CelebA 数据集对添加 SE-ResNet 残差块进行了定性分析. 如图 3. 第 1 列为原始图片, 第 2 列为缺失图片, 第 3 列为不使用 SE-ResNet 残差块的方法修复后图片, 第 4 列为使用 SE-ResNet 残差块的修复后图片. 可以看出, SE-ResNet 的使用对修复效果的影响显著, 这也验证了 SE-ResNet 残差块的有效性.

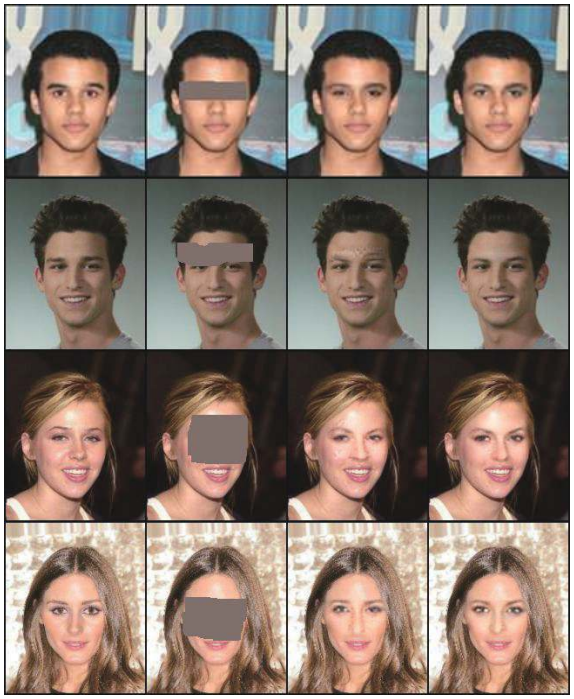

图 3 添加 SE-ResNet 残差块与否对比图

3.4 上下文感知损失的效果分析

在 3.4 节该文从理论上分析了联合上下文感知损失网络训练生成网络的有效性, 为了验证上下文感知损失网络在修复效果上的重要性, 分别在训练时采用上下文感知损失网络和不采用上下文感知损失网络对网络进行训练, 结果如图 4 所示. 第 1 列为原始图片, 第 2 列为缺失图片, 第 3 列为不采用上下文感知损失网络修复后图片, 第 4 列为采用上下文感知损失网络修复后图片. 从图 4 的第 4 行可以看出, 对缺失嘴部的修复中, 不采用上下文感知损失网络修复后的图片存在较为严重的修复痕迹, 该文的方法能够有效的减少修复痕迹, 与原始图像相似度更高.

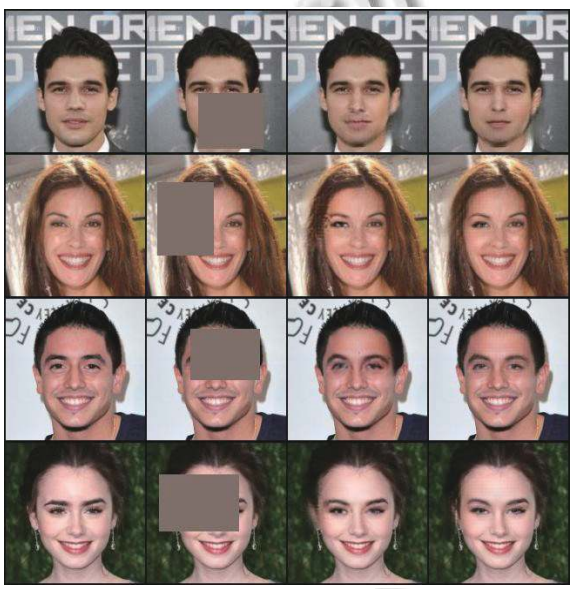

图 4 采用上下文感知损失与否对比图

3.5 与现有方法的比较

为了证明该文方法的优越性, 接下来将从各个方面与文献 [3] 的方法进行对比分析.

该文使用了两种评价标准来评测修复效果. 峰值信噪比 (Peak Signal to Noise Ratio, PSNR)[16] 是一种全参考的图像质量评价指标. 结构相似性 (Structural SIMilarity, SSIM)[17], 是一种全参考的图像质量评价指标, 它分别从亮度、对比度、结构 3 方面度量图像相似性. SSIM 取值范围 [0, 1], 值越大, 表示图像失真越小.

从 CelebA 的测试数据集中随机的选取 128 张图片, 分别计算中心缺失为 1/4 和中心缺失为 1/3 的原图, 用文献 [3] 方法修复的图像以及本文的方法修复的图像的 PSNR 以及 SSIM 值, 然后进行对比. 从表 1 可以看出, 本文的方法优于文献 [3] 的方法, 更加接近原图. 另外, 图 5 定性分析了对比效果. 前两行为中心缺失 1/2 的修复效果对比图, 后两行为中心缺失 1/4 的修复

© 中国科学院软件研究所 http://www.c-s-a.org.cn

效果对比图, 从左到右依次为原图, 缺失图片, 文献 [3] 的方法修复后图片, 本文方法修复后图片. 可以看出文献 [3] 的修复结果存在较多的瑕点, 该文的修复效果更加清晰.

表 1 与文献 [3] 方法的 PSNR 和 SSIM 指标对比

方法	PSNR (dB)	SSIM
文献[3]	30.18	0.80
本文方法	36.90	0.92

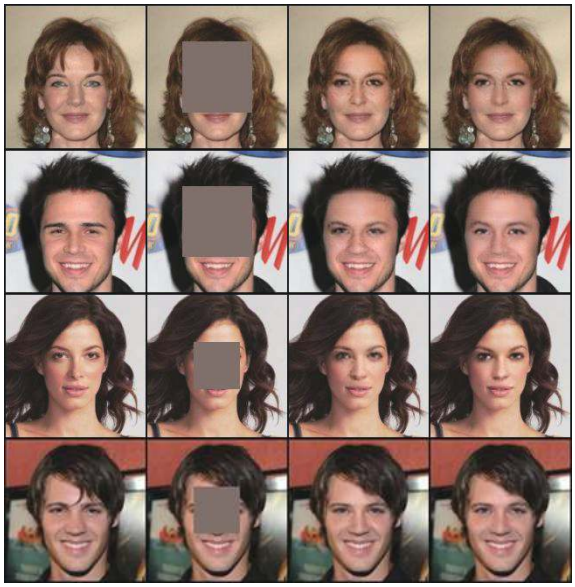

图 5 与文献 [3] 方法的中心缺失修复效果对比图

另外图 6 展示了在随机缺失的情况下文献 [3] 方法与本文方法的对比, 可以看出在图像的高频信息部分缺失时 (如眼睛、鼻子、嘴巴等), 文献 [3] 方法修复细节较差, 且存在较多的伪影, 而本文的方法不存在, 修复效果更好.

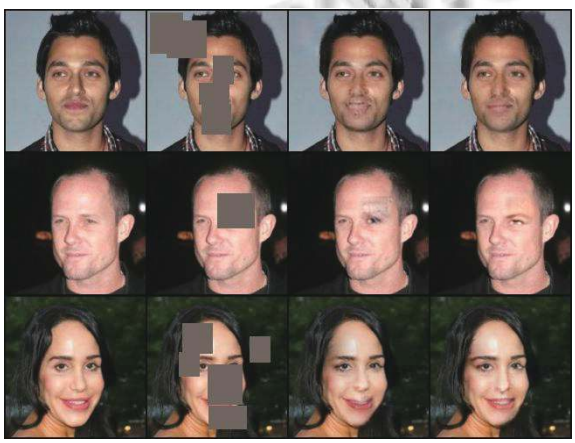

图 6 与文献 [3] 方法的随机缺失修复效果对比图

图像边缘检测图中标识了数字图像中亮度变化明显的点, 反映了图像中的重要的结构属性特征, 因此本文从修复前后边缘检测图的对比来判断图像的修复效果. 图 7 中第 1 列为原图, 第 2 列为原图的边缘检测图, 第 3 列为修复后图像的边缘检测图. 从图中可以看出, 修复后的边缘检测图与原图比较接近, 表明本文方法能从结构上理解并修复图像.

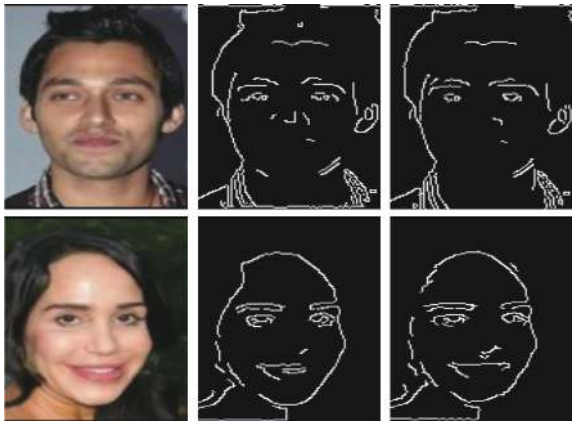

图 7 边缘检测图对比

4 结束语

本文提出了一种基于 SE-ResNet 的生成对抗网络联合上下文感知损失的方法来进行图像修复工作, 通过在生成网络和判别网络部分加入 SE-ResNet 模块, 提升网络特征利用, 使得修复图像更清晰. 通过联合 contextual loss 约束局部特征的相似性, 使得修复图像更加逼真. 多个实验证明, 该文的方法在图像修复上具有重要的作用. 然而当原始图像的分辨率较大或者缺失范围越大, 该方法修复结果还是会存在明显的模糊, 并且训练时间更长, 因此如何解决这个问题有待于进一步研究. 基于深度学习的语义分割技术已经较为成熟, 将语义分割技术与图像修复相结合是一项非常有意义的工作, 因此下一步的研究重点就是如何将语义分割应用于图像修复.

参考文献

1 Darabi S, Shechtman E, Barnes C, *et al*. Image melding: Combining inconsistent images using patch-based synthesis. ACM Transactions on Graphics, 2012, 31(4): 82.

2 Pathak D, Krahenbühl P, Donahue J, *et al*. Context encoders: Feature learning by inpainting. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition.

© 中国科学院软件研究所 http://www.c-s-a.org.cn

Las Vegas, NV, USA. 2016. 2536–2544.

3 Iizuka S, Simo-Serra E, Ishikawa H, *et al*. Globally and locally consistent image completion. ACM Transactions on Graphics, 2017, 36(4): 107.

4 Hu J, Shen L, Sun G. Squeeze-and-excitation networks. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 7132–7141.

5 He KM, Zhang XY, Ren SQ, *et al*. Deep residual learning for image recognition. Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, NV, USA. 2016. 770–778.

6 Mechrez R, Talmi I, Zelnik-Manor L, *et al*. The contextual loss for image transformation with non-aligned data. Proceedings of the 15th European Conference on Computer Vision. Munich, Germany. 2018. 800–815.

7 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, 2014.

8 Goodfellow I, Pouget-Abadie J, Mirza M, *et al*. Generative adversarial nets. Proceedings of the 27th International Conference on Neural Information Processing Systems. Cambridge, MA, USA. 2014. 2672–2680.

9 Song YH, Yang C, Shen YJ, *et al*. SPG-Net: Segmentation prediction and guidance network for image inpainting. arXiv: 1805.03356, 2018.

10 Yu JH, Lin Z, Yang JM, *et al*. Generative image inpainting with contextual attention. Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA. 2018. 5505–5514.

11 Liu HY, Jiang B, Xiao Y, *et al*. Coherent semantic attention for image inpainting. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Republic of Korea. 2019. 4169–4178.

12 Yu JH, Lin Z, Yang JM, *et al*. Free-form image inpainting with gated convolution. Proceedings of 2019 IEEE/CVF International Conference on Computer Vision. Seoul, Republic of Korea. 2019. 4471–4480.

13 Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. arXiv: 1511.07122, 2015.

14 Liu ZW, Luo P, Wang XG, *et al*. Large-scale celebfaces attributes (CelebA) dataset. http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html, 2018.

15 Huang GB, Ramesh M, Berg T, *et al*. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Amherst: University of Massachusetts, 2007.

16 佟雨兵, 张其善, 祁云平. 基于 PSNR 与 SSIM 联合的图像质量评价模型. 中国图象图形学报, 2006, 11(12): 1758–1763. [doi: 10.11834/jig.2006012307]

17 Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM. Proceedings of the 20th International Conference on Pattern Recognition. Istanbul, Turkey. 2010. 2366–2369.

© 中国科学院软件研究所 http://www.c-s-a.org.cn