

# 基于新生成器结构的图像修复方法<sup>①</sup>



杨 柳, 王 敏, 林 竹

(河海大学 计算机与信息学院, 南京 211100)

通讯作者: 杨 柳, E-mail: 18260061997@163.com

**摘 要:** 针对目前图像修复算法存在的修复效果不连续、缺失大小受限、训练过程不稳定等缺点, 提出了一种基于生成对抗网络的图像修复方法. 利用卷积神经网络, 我们可以真实地修复任意分辨率的图像. 为了实现高分辨率的真实修复效果和对图像特征的充分学习, 我们提出基于 DenseNet 传播源图像的细节和结构得到高分辨率的图像, 实现图像缺失生成; 由于 Iizuka 等人提出的基于双判别器方法中膨胀卷积部分所产生的巨大运算量, 我们提出使用 JPU (Joint Pyramid Upsampling, 联合金字塔上采样) 来加速计算. 在 CelebA 和 ImageNet 上的实验表明, 所提方法能真实地修复大多数的破损图像.

**关键词:** 图像修复; 生成对抗网络; DenseNet; JPU; 双判别器

引用格式: 杨柳, 王敏, 林竹. 基于新生成器结构的图像修复方法. 计算机系统应用, 2020, 29(1): 158-163. <http://www.c-s-a.org.cn/1003-3254/7234.html>

## Image Restoration Method Based on New Generator Structure

YANG Liu, WANG Min, LIN Zhu

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

**Abstract:** Aiming at the shortcomings of current image restoration algorithms, such as discontinuity of repair effect, limitation of missing size, and instability of training process, an image restoration method based on generation antagonistic network is proposed. Using convolutional neural network, we can actually repair images of any resolution. In order to realize the real restoration effect of high resolution and the full learning of image features, we propose to obtain high resolution images based on the details and structure of DenseNet propagating source images, so as to realize the generation of missing images. As Iizuka et al. proposed the large computation amount generated by the expanded convolution part in the two-discriminator method, we propose to use JPU (Joint Pyramid Upsampling) to accelerate the calculation. Experiments in CelebA and ImageNet show that the proposed method can truly repair most broken images.

**Key words:** image restoration; generate antagonism network; DenseNet; JPU; double discriminant unit

图像修复是一种对破损或者存在缺失的图像区域进行补充的技术, 它在缺失区域插入所预测的内容, 为目标移除提供了技术支持. 虽然研究者已经提出了许多图像修复的方法, 例如基于 patch 的图像合成<sup>[1]</sup>, 但是图像修复仍然是一个具有挑战性的研究方向, 它不仅要修复图像纹理, 还需要理解整个场景和被修复对象的语义. 在本文中, 我们考虑使用基于 DenseNet<sup>[2]</sup>结

构的图像缺失生成网络结构来缓解纹理修复中存在的问题, 同时针对双判别器提出使用 JPU<sup>[3]</sup>加速来解决图像纹理和场景连续性的问题.

Pathak 等人提出 Context-Encoder<sup>[4]</sup>基于生成对抗网络训练图像修复的概念, Iizuka 等人提出的文献<sup>[5]</sup>的方法是基于全局和局部一致的图像修复方法, 该方法使用 GAN 对抗性损失训练卷积神经网络<sup>[6]</sup>. 该方法

① 收稿时间: 2019-06-12; 修改时间: 2019-07-08; 采用时间: 2019-07-19; csa 在线出版时间: 2019-12-27

通过双判别器结构对缺失生成器对 Context-Encoder 方法改进, 进行对抗训练, 使得生成的结果更加真实. 该方法尽管可以处理任意大小的任意缺失图像, 但是明显较大的缺失无法修复. 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. 与此同时, 该方法由于膨胀卷积层<sup>[7]</sup>的使用, 使得训练、修复时间很长. 我们提出的方法解决了这两点, 并将进一步提高结果的视觉质量和生成速度.

我们使用 DenseNet 网络架构和卷积神经网络作为网络的基础, 提出了一种新的网络结构, 联合 DenseNet 实现对图像信息的深度利用, 可以实现修复效果的连续性和清晰性, 同时通过 JPU (联合金字塔上采样) 实现训练的加速. 本文的网络结构由 3 个部分组成: 缺失区域生成网络、缺失区域判别器和缺失周边区域鉴别器. 缺失区域生成网络是基于 DenseNet 的, 并使用了提高训练速度的 JPU 替代了传统的扩充卷积层用于生成缺失区域的预测图像. 缺失区域判别器和缺失周边判别器都是用于训练的辅助子网, 这些判别器用于判别图像修复结果是否真实以及与周边上下文是否连续. 我们使用缺失周边部分区域作为判别依据, 联合 JPU 来减少运算的时间. 联合训练双判别器和生成器, 最终使得生成器所生成的效果能“欺骗”判别器. 与现有的方法比较, 尤其是人脸修复, 我们的方法在修复的真实性和与原图的差异性上表现优秀, 在训练时间上也比现有的部分方法速度更快.

综上所述, 本文提出:

(1) 不同于基于全局和局部一致的图像修复, 我们使用充分挖掘图片特征的 DenseNet 结构替代传统的基于 AlexNet 架构的缺失区域生成网络;

(2) 使用 JPU (联合金字塔上采样) 替代传统膨胀卷积层加速计算.

## 1 相关工作

图像修复的传统方法有: 应用变分法对缺失区域进行修复<sup>[8]</sup>, 但是这种方法迭代速度很慢, 对于缺失较大的区域, 修复完成后有明显的模糊. 基于边界的优先级对缺失区域进行顺序填充<sup>[9]</sup>, 是一种基于填充区域轮廓优先级的贪心算法, 它明显的缺点是无法保证整体填充效果连续. 基于 patch 的算法<sup>[1]</sup>通过寻找最相似、最近邻的 patch 来进行修复, 它的优点是修复速度很

快, 但是对于初始化比较敏感, 容易陷入局部极值, 导致修复效果整体不连续. 基于 offset 的算法<sup>[10]</sup>使用图割算法, 利用自然图像中存在的自相似冗余和 patch 块之间的偏移量也存在稀疏的特性计算 offset 来进行图像修复, 该方法减少了边界过度不自然的问题. 但是与我们的方法不同, 它们都不能生成源图像中没有的新对象. 为了生成缺失图像没有的对象, 有人提出使用数据库驱动的方法<sup>[11]</sup>, 通过在数据库中寻找匹配来完成缺失修复, 但是这种方法耗时且修复的对象受数据库的限制.

卷积神经网络 (CNNs) 和对抗神经网络 (GAN)<sup>[6]</sup>也被用于图像补全. Pathak 等人提出上下文编码器 (Context-Encoder, CE)<sup>[4]</sup>, 该算法提出使用 encoder-decoder 作为缺失区域生成器, 将 GAN 作为修复结果判别器. 缺失图像被送入生成器网络进行缺失生成, 修复的结果和真实图像被送入判别器进行对抗判别, 直至判别器无法区分真实图像和修复图像的时候, 训练结束. CE 通过使用均方误差 (MSE) 损失与 GAN 损失相结合的方法, 在  $128 \times 128$  像素图像的中心修复  $64 \times 64$  像素的区域来训练修复网络, 避免了仅使用 MSE 损失时常见的模糊.

基于全局和局部一致的图像修复<sup>[5]</sup>在 CE<sup>[4]</sup>的基础上, 通过引入双判别器结构将基于 CNN 的修复扩展到任意的区域缺失. 缺失图像被送入带有膨胀卷积层的生成网络生成修复结果图像<sup>[12,13]</sup>, 缺失区域的修复结果被送入缺失区域判别器; 整体的修复结果被送入缺失周边判别器, 双判别器结构基本保证图像在缺失局部和缺失周边全局上修复效果一致, 当双判别器无法判别真实图像和修复图像时, 训练结束.

在 encoder 结构中, 得到了低分辨率的特征图, 最终的特征图编码了丰富的语义信息, 却丢失了的图像结构信息, 导致修复效果不清晰. 我们的方法在减少双判别器训练时间的同时, 增强了修复的效果. 通过基于 DenseNet 构建缺失生成网络, 提高修复结果的清晰度; 使用联合上采样的 JPU, 而非传统的膨胀卷积层, 有效减少了训练时间. 同时联合上采样的 JPU 的输入联合多尺度的特征图和 DenseNet 的使用, 实现了有效的特征重用, 保留了更多的空间结构和位置信息, 使得最终的修复效果相比较而言有高分辨率.

## 2 网络结构

我们的方法是基于 DenseNet 训练的图像修复网

络. 图像缺失区域生成使用单个联合 JPU 上采样的生成网络. 为了训练这个网络真实地完成图像, 使用了另外两个子网络, 局部缺失判别网络和非全局的局部缺失周边判别网络. 在训练过程中, 对判别器网络进行训练, 判断图像是否已经完成, 对生成网络进行欺骗. 只有将这 3 种网络训练在一起, 网络才能真实地修复多种图像. 网络结构如图 1 所示.

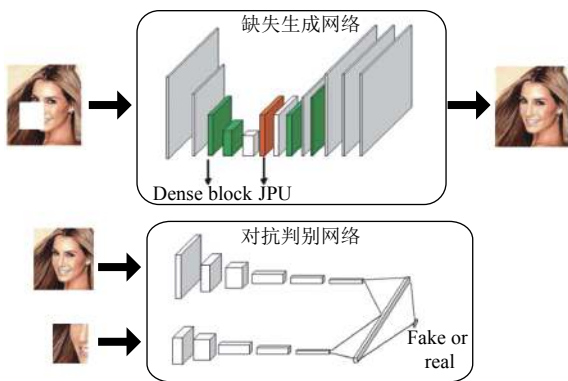


图 1 基于新生成器结构的图像修复网络结构图

## 2.1 DenseNet

随着卷积神经网络的深入, 后面的输出与前面卷积层的特征联系会更少, 如果将前面卷积层的特征图输入到后面的卷积层, 那么训练的网络会更加深入和有效. 原始 encoder-decoder 使用的是基于 AlexNet 结构的, 而我们的方法是基于 DenseNet 结构, 它以前馈的方式将每一层连接到后面的每一层, 加强网络的图像结构、位置等信息重用和特征学习. 传统的卷积网络有  $L$  层, 而 DenseNet 有  $L(L+1)/2$  个直接连接. 对于每一层, 前面所有层的特征图都用作输入, 而它自己的特征图用作后面所有层的输入. DenseNet 有几个优点: 缓解了消失梯度问题, 增强了特征传播, 增强了特征重用, 并大大减少了参数的数量. DenseNet 在大多数情况下都比之前的技术有显著的改进, 同时仅需要更少的计算来实现视觉上更好的修复效果. DenseNet 由 denseblock 组成, 对每个 denseblock:  $X_l = H_l([x_1, x_2, x_3, \dots, x_{l-1}])$ , 第  $l$  层接收前面所有  $l-1$  层的特征图  $x_1, x_2, x_3, \dots, x_{l-1}$  拼接作为输入,  $H_l$  由 3 个操作组成: BN、ReLU、 $3 \times 3$  的卷积. 每个 denseblock 之间的连接被称为过渡层, 过渡层主要由卷积层和池化层构成. 在本文里, 将过渡层和 denseblock 整合为新的 denseblock 结构.

## 2.2 缺失生成网络

缺失生成图像的基本结构信息应该与原图像信息

一致, 在输入特征图和输出特征图之间某些特征可能相似, 利用好这些信息, 可能会生成对比原图更加真实的修复. 根据这个思想, 与其他使用类似于 AlexNet<sup>[14]</sup> 结构的编码-解码结构不同, 我们使用基于 DenseNet 的编码-解码结构. 缺失区域生成网络的输入为有缺失图像, 在缺失之外的区域不做任何更改, 仍为原图像, 输出为修复完成的图像. 缺失区域生成网络为编码-解码结构, 其中编码器部分使用基于 DenseNet 结构对图像进行深度特征压缩提取, 然后使用同样基于 DenseNet 结构的解码器的反卷积结构对图像大小进行恢复, 生成最终的修复图像.

在基于全局和局部一致的图像修复方法<sup>[5]</sup>中使用双判别器的结构中, 使用了 4 个传统膨胀卷积层作为编码-解码部分的连接处理, 来保持图像的感知野. 扩张卷积在获取高分辨率最终特征图中发挥重要作用, 但它也有一些缺陷, 比如增加了计算复杂度和内存占用. 基于这种思想, 除了使用标准的卷积层, 为了减少扩张卷积带来消耗的时间和内存, 我们使用了新的联合上采样模块 JPU<sup>[3]</sup>替代传统的膨胀卷积层. 并且, 膨胀卷积层得到的最终的特征图编码了丰富的语义信息, 却丢失了精细的图像结构信息, 导致目标边界预测不准确, 修复效果不清晰, 而我们使用 JPU 替代扩张卷积层部分能得到高分辨率的特征图. 同时, 在不损失性能的前提下, 将计算复杂度降低了 2 倍左右, 提高了运行速度. 通过实验观察 loss 下降情况, 使用 JPU 连接编码、解码结构, 训练的过程收敛的更快且更加稳定.

JPU 以编码器最后 3 层的特征图作为输入, 与传统的膨胀卷积层的不同之处基于最后的两个卷积层, 传统的使用扩张卷积层是在 encoder 第 4 层后使用一个常规的卷积层, 再放入扩张卷积层; JPU 是用步长为 2 的卷积处理, 然后在放入常规卷积层输出. 整个框架的时间和内存占用大幅减少. 同时由于它能利用跨多级特征图的多尺度特征图, 性能不会出现损失.

生成器部分的损失函数分为两个部分: 对抗性损失和修复距离损失. 对抗性损失:

$$L_{adv} = -E_{Z \sim p_z(Z), (x,y) \sim p_r(x,y)} D(G(z,y), y)$$

对抗性损失是为了生成器尽最大可能“欺骗”判别器; 修复距离损失采用 L2 距离:

$$L_{res} = E_{Z \sim p_z(Z), (x,y) \sim p_r(x,y)} [(M \odot (x - G(z,y)))^2]$$

修复距离损失是为了生成尽量真实的图像, 尽最

大可能使得修复完成的图像与原图像“距离”更小。最后生成器部分的损失函数为:  $L_{adv} + \lambda_1 L_{res}$ , 本文中  $\lambda_1 = 0.0004$ 。

### 2.3 对抗判别网络

缺失区域判别器和缺失周边判别器网络的目标是识别图像是真实的还是修复得到的。该网络基于卷积神经网络, 将图像压缩成小的特征向量。网络的输出由一个全连接层处理得到, 预测修复结果与真实结果的概率值。

我们采用与基于全局和局部一致的图像修复<sup>[5]</sup>类似的双判别器策略, 缺失区域判别器有两个输入: 生成器所生成的缺失区域的图像和真实缺失区域图像; 缺失周边判别器以修复完成的图像和原图像作为输入, 联合两个判别器对抗判别训练。两个判别器的主要区别在于输入图像的大小不同, 局部的输入仅为全局的1/2, 所以缺失区域判别器不需要使用缺失周边判别器的第一层。

最后, 局部判别器得到一个1024维的输出, 表示局部上下文信息; 全局判别器也得到一个1024维的输出, 表示全局上下文信息。两者信息结合, 在使用一个全连接层得到最终的判别概率, 判别图像是修复出来的, 还是真实的。当判别器无法判别图像是修复/真实的时候, 结束训练。整体算法流程图如图2所示。

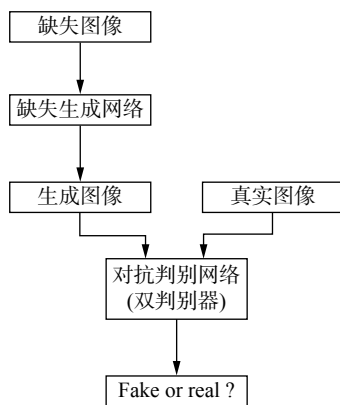


图2 基于新生成器结构的图像修复算法流程图

## 3 实验

本文使用 ImageNet<sup>[15]</sup>和来自香港中文大学的开放数据集 CelebA<sup>[16]</sup>。ImageNet 有大约 1500 万张图片, 有 2.2 万类数据, 我们这里选用其中的 dog 类作为实验数据集; CelebA 是一个大型人脸数据集, 有 10 177 个名人的 202 599 张人脸图像。

### 3.1 各部分的作用

本部分主要分析 DenseNet 架构和 JPU 的作用。

#### 3.1.1 基于 DenseNet 架构

2.1 节从理论上分析了使用基于 DenseNet 架构构建缺失生成网络的优势, 为了进一步从实验上证明该优势, 我们使用了 ImageNet 数据集中的小狗类对使用 AlexNet 作为基本基本架构和使用 DenseNet 作为基本架构基于双判别器方法进行了评估。使用 AlexNet 作为基本架构的修复效果与使用 DenseNet 作为基本架构的修复效果如图 3, 第一列为原始图片, 第二列为缺失图片, 第三列 encoder-decoder 部分使用 AlexNet 结构的 Iizuka 等的方法修复效果, 第四列为使用 DenseNet 作为基本架构的修复效果。可以看出, 使用 DenseNet 作为基本架构的修复效果更加清晰且真实, 对比 Iizuka 等的方法修复效果改进明显。

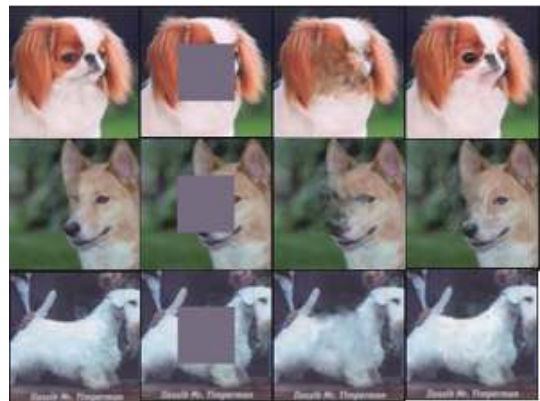


图3 使用 DenseNet 作为基本结构和使用 AlexNet 基本结构的修复效果对比

#### 3.1.2 JPU 加速

第 2.2 节从理论上分析了使用 JPU (联合金字塔上采样) 替代传统膨胀卷积层的优势, 为了进一步从实验上证明该优势, 我们对不带膨胀卷积使用 JPU、带传统膨胀卷积的双判别器方法进行对比, 修复结果如图 4 所示, 我们使用相同的实验设置, 均为中心缺失修复。第一列为原始图像, 第二列为缺失图像, 第三列为使用膨胀卷积层的双判别器网络, 第四列为使用 JPU 替代传统膨胀卷积的方法。由图 4 可以看出, 使用 JPU 替代传统膨胀卷积, 不损失修复效果, JPU 实现了类似传统膨胀卷积的作用; 在轻微或者不牺牲修复效果的情况下, 在同样迭代 50 000 次, batch\_size 设置为 32 的情况下, 使用膨胀卷积层的训练时间为 14.10 h, 使用 JPU

的训练时间为 6.97 h. 在时间上, JPU 大约减少了一倍的训练时间, 减少了内存开销. 同时, loss 曲线的趋势更加稳定.

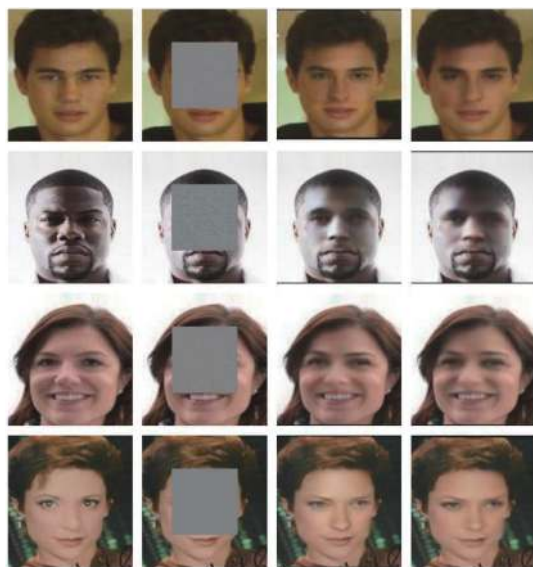


图4 不使用和使用 JPU 加速的修复效果对比

由图4可以看出, 仅仅使用 JPU 只能起到加速训练的作用. 为了解决带传统膨胀卷积的双判别器方法下修复图像模糊的缺点, 我们会在接下来联合使用 JPU 和基于 DenseNet 的架构.

### 3.2 与现有方法的比较

PSNR (Peak Signal to Noise Ratio)<sup>[17]</sup>峰值信噪比是一种全参考的图像质量评价指标. 然而 PSNR 是基于对应像素点间的误差, 即基于误差敏感的图像质量评价. 并未考虑到人眼的视觉特性 (人眼对空间频率较低的对比差异敏感度较高, 人眼对亮度对比差异的敏感度较色度高, 人眼对一个区域的感知结果会受到其周围邻近区域的影响等), 因而经常出现评价结果与人的主观感觉不一致的情况. 所以我们不仅仅使用 PSNR 作为衡量指标, 我们还使用到了 SSIM.

SSIM (Structural SIMilarity)<sup>[18]</sup>结构相似性, 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.

图5 为中心缺失为 32×32 像素大小 (原图大小为 128×128) 的情况下我们的方法与 Iizuka 等的方法修复效果对比. 图6 为中心缺失 54×54 的情况下我们的方法与 Iizuka 等的方法修复效果对比. 图7 为中心缺失

64×64 的情况下我们的方法与 Iizuka 等的方法修复效果对比. 图8 为随机缺失情况下我们的方法与 Iizuka 等的方法修复效果对比. 第一列均为原始图像、第二列均为缺失图像、第三列均为 Iizuka 等的修复效果、第四列均为使用我们的方法 (网络框架如图1) 后的修复效果. 表1 为中心缺失 64×64 情况下修复效果所对应的平均 PSNR 和 SSIM 值对比.



图5 中心缺失为 32×32 的修复效果对比图



图6 中心缺失为 54×54 的修复效果对比图



图7 中心缺失为 64×64 的修复效果对比图

由图5、图6、图7和表1可以看到, 我们的方法可以真实地修复人脸上的缺失区域, 修复结果与原图间的差异性也更小. 在图5 缺失较小的情况下, 我们的方法和 Iizuka 方法修复效果差别不大, 随着缺失的增大, 修复难度的提高, 我们的方法明显优于 Iizuka 的方法. 但在缺失较大的情况下, 如 0.9 比例的缺失, 由于信

息缺损严重,修复效果较差.传统的基于补丁的方法无法完成人脸,因为它要求算法生成不属于该图像一部分的新对象,如眼睛、鼻子和嘴巴.对比 Iizuka 等的方法,我们的方法生成的图像更加清晰且能真实的应用于多种人脸角度.由图 8 可以看到,在缺失大小、位置、形状随机的情况下,本方法的修复效果更真实且清晰.说明本方法优于 Iizuka 等的方法.

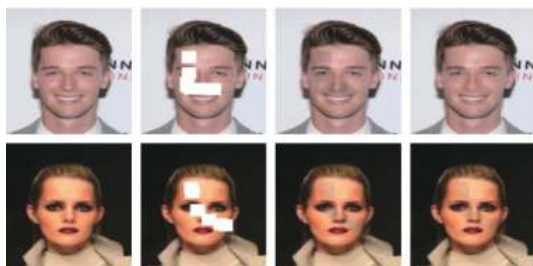


图 8 随机缺失下的修复效果对比图

表 1 我们的方法与 Iizuka 等的方法 PSNR、SSIM 指数对比

方法	PSNR (dB)	SSIM
Iizuka 等的方法	30.18	0.80
我们的方法	36.82	0.91

#### 4 结论

我们提出了一种新的基于双判别器的图像修复方法,实验证明,通过使用 JPU (联合金字塔上采样) 替代膨胀卷积进行计算,在不降低修复效果的前提下,所需要的训练和修复时间大大减少.通过使用 DenseNet 来构造 encoder-decoder 部分,提高修复的真实性,并使得修复效果更加清晰.此外,本文展示了将本方法应用于人脸上的修复效果,与真实人脸差异性较小.

#### 参考文献

- 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.
- Huang G, Liu Z, van der Maaten L, *et al.* Densely connected convolutional networks. *Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu, HI, USA. 2017. 2261–2269.
- Wu HK, Zhang JG, Huang KQ, *et al.* FastFCN: Rethinking dilated convolution in the backbone for semantic segmentation. *arXiv: 1903.11816*, 2019.
- Pathak D, Krähenbühl P, Donahue J, *et al.* Context encoders: Feature learning by inpainting. *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition*. Las Vegas, NV, USA. 2016. 2536–2544.
- Iizuka S, Simo-Serra E, Ishikawa H. Globally and locally consistent image completion. *ACM Transactions on Graphics*, 2017, 36(4): 107.
- Goodfellow IJ, Pouget-Abadie J, Mirza M, *et al.* Generative adversarial nets. *Proceedings of the 27th International Conference on Neural Information Processing Systems*. Montreal, Canada. 2014. 2672–2680.
- Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions. *arXiv: 1511.07122*, 2015.
- 周密, 彭进业, 赵健, 等. 改进的整体变分法在图像修复中的应用. *计算机工程与应用*, 2007, 43(27): 88–90. [doi: 10.3321/j.issn:1002-8331.2007.27.025]
- Criminisi A, Pérez P, Toyama K. Object removal by exemplar-based inpainting. *Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. Madison, WI, USA. 2003. 721–728.
- He KM, Sun J. Image completion approaches using the statistics of similar patches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2014, 36(12): 2423–2435. [doi: 10.1109/TPAMI.2014.2330611]
- Hays J, Efros AA. Scene completion using millions of photographs. *ACM Transactions on Graphics*, 2007, 26(3): 4. [doi: 10.1145/1276377.1276382]
- Agrawal P, Carreira J, Malik J. Learning to see by moving. *Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago, Chile. 2015. 37–45.
- Doersch C, Gupta A, Efros AA. Unsupervised visual representation learning by context prediction. *Proceedings of 2015 IEEE International Conference on Computer Vision*. Santiago, Chile. 2015. 1422–1430.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. *Proceedings of the 25th International Conference on Neural Information Processing Systems*. Lake Tahoe, CA, USA. 2012. 1097–1105.
- Deng J, Dong W, Socher R, *et al.* Imagenet: A large-scale hierarchical image database. *Proceedings of 2009 IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL, USA. 2009. 248–255.
- Liu ZW, Luo P, Wang XG, *et al.* Large-scale celebfaces attributes (CelebA) dataset. 2018.
- 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]
- Horé A, Ziou D. Image quality metrics: PSNR vs. SSIM. *Proceedings of the 2010 20th International Conference on Pattern Recognition*. Istanbul, Turkey. 2010. 2366–2369.