

文章编号: 1672-8785(2024)07-0029-06

基于空间特征融合的双路径 图像去噪网络

祖雅婷 李梦琪 张艺萌 王 赫

(1. 中国电子科技集团公司第十一研究所, 北京 100015;

2. 陆军装备部驻北京地区军事代表局驻北京地区第二军事代表室, 北京 100015)

摘 要: 深度卷积神经网络(Convolutional Neural Network, CNN)在图像去噪领域受到广泛关注。然而, 随着网络深度的增加, 大多数深度 CNN 会出现性能饱和、学习能力下降等问题。提出了一种结合局部和全局特征的双路径去噪网络, 将两个不同结构的网络组合后构成一个双路径模型, 增加网络的宽度, 从而获得更多不同的特征。通过长路径连接融合全局和局部特征, 增强层间相关性。注意力机制利用当前阶段引导前一阶段的输入, 获得更多的特征。实验结果表明, 我们提出的网络模型在 Set12 和 BSD68 两个数据集中的峰值信噪比(Peak Signal-to-Noise Ratio, PSNR)分别达到了 32.95 dB 和 31.74dB。同时, 主观视觉效果(如图像的边缘等细节)恢复得更好、更清晰。

关键词: 图像去噪; 深度学习; 注意力机制; 卷积神经网络; 双路径

中图分类号: TP391.41 **文献标志码:** A **DOI:** 10.3969/j.issn.1672-8785.2024.07.004

Dual-Path Image Denoising Network Based on Spatial Feature Fusion

ZU Ya-ting, LI Meng-qi, ZHANG Yi-meng, WANG He

(1. The 11th Research Institute of China Electronics Technology Group Corporation, Beijing 100015, China;

2. The Second Military Representative Office of the Military Representative Bureau of the

Army Armaments Department in Beijing, Beijing 100015, China)

Abstract: Deep convolutional neural networks (CNNs) have attracted much attention in the field of image denoising. However, with the increase in network depth, most deep CNNs have problems such as performance saturation and learning decline. In this paper, a dual-path denoising network combining local and global features is proposed. Two networks with different structures are combined to form a dual path model, and the width of the network is increased to obtain more different features. The global and local features are integrated through long path connections to enhance interlayer correlation. The attention mechanism uses the current stage to guide the input of the previous stage to obtain more features. The experimental results show that the PSNR values of the proposed network model reach 32.95 dB and 31.74 dB in Set12 and BSD68 datasets, re-

收稿日期: 2023-05-04

作者简介: 祖雅婷(1998-), 女, 辽宁沈阳人, 硕士, 主要研究方向为图像处理与计算机视觉。

E-mail: zuyating@naver.com

spectively. At the same time, the subjective visual effects such as image edges and other details are recovered better and clearer.

Key words: image denoising; deep learning; attention mechanism; convolutional neural network; dual path

0 引言

在实际应用中, 数字图像在传输过程中经常受到视频设备和外界环境的干扰。由于这些原因, 在获取数字图像的过程中, 图像的质量会下降。噪声是表示图像质量恶化的因素之一。含有噪声的图像会干扰识别或获取图像中的必要信息。图像去噪算法主要可分为基于传统和深度学习的两种研究方法。

Buades A 等人提出了非局部均值(Non-local Mean, NLM)算法^[1]。NLM 滤波取图像中所有像素的平均值, 并根据这些像素与目标像素的相似程度进行加权处理。传统去噪方法中最具代表性的是 BM3D 算法^[2]。它采用 NL-means 算法^[1], 找到与给定图像模块相似的图像模块, 并将其分组到一个 3D 块中, 然后利用协同改变增强图像的稀疏性, 从而进行图像去噪。虽然传统去噪方法也取得了不错的图像去噪效果, 但手动设置参数、时间耗费较高等缺点也逐渐显露。

近年来, 随着深度学习技术的发展, 神经网络的应用克服了传统方法带来的弊端。神经网络以其出色的特征表示能力和灵活的模块化架构, 在图像去噪领域取得了显著的效果。Jain V 等人^[3]首次使用 CNN 进行图像去噪。Zhang K 等人^[4]也提出了一种去噪卷积神经网络(DnCNN)。DnCNN 算法主张随着网络的深入, 模型的性能逐渐提高。采用残差学习(Residual Learning, RL)^[5]和批量归一化(Batch Normalization, BN)^[6]进行快速训练, 提高去噪性能。根据文献^[4], 随着网络层数的加深, 图像中会提取出更多的特征, 同时也证明了该模型具有较高的性能。但是, 当网络深度超过一定深度时, 性能会随之下降。当浅层学习能力的降低限制了深层的学习能力时, 则会出现

性能饱和的问题。为了改善这个问题, 强化学习也不能很好地应用于更深层的网络。

本文基于以上问题提出了一种带有特征增强的双路径 CNN 图像去噪算法。具体来说, 将两个不同结构的网络结合增加了网络的宽度, 从而可以挖掘出图像中更多维度的特征。标准卷积和空洞卷积的结合使用可从给定的噪声图像中提取更多有用的细节特征。这样既提高了去噪性能, 又提高了去噪效率。利用长路径融合全局和局部特征, 提高模型浅层与深层之间的相关性。接下来, 我们使用注意力机制来利用后一阶段为网络的前一层提供补充信息。最后, 使用 RL 构造干净的图像。此外, 本文假设噪声为加性高斯白噪声(AWGN), 噪声级分别为 15、25 和 50。AWGN 能较好地模拟复杂分量的真实噪声。

1 相关研究

1.1 残差学习

虽然增加网络深度可以提高图像去噪的性能, 但通过简单地对网络进行堆叠来增加网络的深度, 则会产生梯度消失、爆炸或衰减到很小等问题。因此, He K 等人^[5]提出的 RL 被用于解决网络深化带来的性能退化问题。基本框架如图 1 所示。

残差模块由两层组成, X 为残差单元的输入值, $H(\bullet)$ 代表残差单元的输出值。当 $F(\bullet) = 0$ 时, 构成恒等映射。残差值 $F(\bullet)$ 的表达式为

$$F(X) = H(X) - X \quad (1)$$

$$H(X) = F(X) + X \quad (2)$$

$H(\bullet)$ 可以通过具有 shortcut 的前馈神经网络来实现。在这种情况下, shortcut 只是执行恒等映射, 它们的输出被添加到堆叠层的输出中。优化残差映射比优化原始的、未引用的映

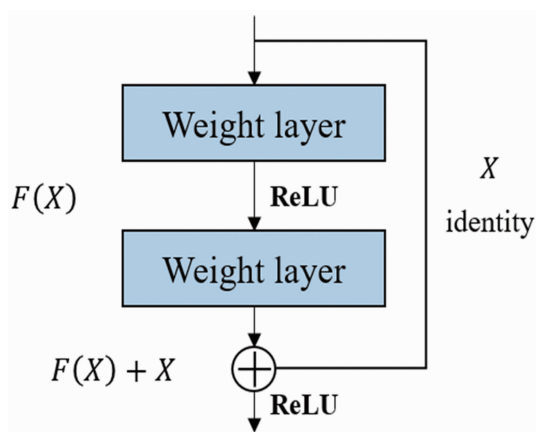


图 1 残差模块

射更加容易。强化学习的目的是通过学习噪声图像与原始图像之间的残差来代替学习它们的共同信息。通过在浅层网络中添加一个跳跃连接,可以在不降低更深深度网络能力的情况下构建网络。

1.2 空洞卷积

在图像去噪中,上下文是恢复受损像素的重要因素。扩大感受野是 CNN 捕获更多特征信息的常用方法。空洞卷积是标准卷积的一种特殊形式,可在不做池化损失信息的情况下加大感受野,使得内部数据结构得到保留。它具有无需添加额外参数即可扩展感受野的优点。图 2 所示为标准卷积和空洞卷积的基本示例。

对于尺寸为 5×5 的普通图像,采用图 2(a)所示的 3×3 标准卷积和图 2(b)所示的扩展因子为 2 的空洞卷积来执行卷积操作,得到的图形块大小分别为 3×3 和 1×1 。对于扩展因子为 2 的空洞卷积,实际上只有 9 个红点参与卷积运算,其余零点权重为 0。虽然卷积核的大小只有 3×3 ,但是这个卷积的感受野与 5×5 卷积核一致。膨胀率决定了内核中值之间的间隔。通过扩展卷积核,可以与普通卷积操作相同的计算成本作用于更广泛的感受域。

2 提出的方法

下面介绍基于神经网络的图像去噪网络。在模型架构的设计上,我们将两个子网络融合为一个网络结构。众所周知,层次较深的网络

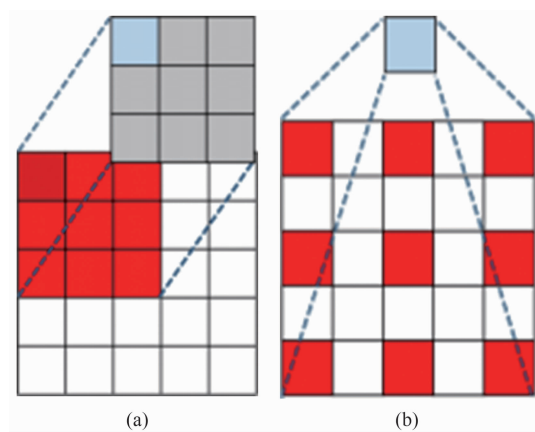


图 2 (a)标准卷积; (b)空洞卷积

会导致梯度消失或爆炸。一般来说,不同的网络架构可以提取不同的特征信息。这些提取的特征在图像去噪任务是互补的。通过增大网络宽度来提高性能是实现去噪任务的有效途径。

因此,我们提出了一个基于双路径的图像去噪网络(见图 3)。该网络包括两个不同的子网络:全局特征增强子网络和局部融合捕捉噪声子网络。两个子网络组合的输出大小为 $64 \times 3 \times 3 \times 2 \times 64$ 。具体来说,该层的输入和输出为 64,卷积核大小为 3×3 。组合结果以 Conv+BN+ReLU 三层计算。最后,将输出结果与残差学习相结合,得到去噪图像。其中,Conv 表示卷积层,BN 表示批处理归一化层,ReLU 表示激活函数层。首先,我们提出的全局特征增强子网络有 17 层。第一层为 Conv+ReLU,最后一层为 Conv。Conv+BN+ReLU 层设置在第 11 层和第 16 层。在网络的其余部分设置扩展系数为 2 的 Conv+BN+ReLU 层。在全局特征增强子网络中,标准卷积和空洞卷积的联合使用有助于我们在更广的图像范围内获得更多的特征来提高去噪性能。在两个子网络对输入的噪声图像进行处理后,使用 Concat 来实现连接功能,并将两个子网络中获取到的细节融合,从而提高计算效率。局部融合噪声捕获子网络层数为 18 层。网络的第 2、5、9、12 层扩展为 Conv+BN+ReLU,16~18 层为 Conv,其余层设置为 Conv+BN+ReLU。在该子网

络中,我们使用串联操作将输入的噪声图像与第 16 层的输出图像相融合,从而增强去噪模型的表示能力。因此,最终输出大小为 $64 \times 3 \times 3 \times 2C$, 其中 C 表示输入图像的通道数。此外,使用 Tanh 函数将得到的特征转换为非线性。由于复杂的背景很容易隐藏图像或视频中的信息,使训练变得更加困难,因此我们使用网络中的注意力机制来优化去噪模型。

注意力机制利用当前阶段引导前一阶段学习有噪声的信息,有利于处理未知噪声图像。第 17 层是一个大小为 1×1 的卷积核,用于将得到的特征压缩成一个向量作为前一阶段调整的权重。利用得到的权重乘以第 16 层的输出来提取更突出的噪声特征。最后,

RL 被用来重建干净的图像。

3 实验与结果

3.1 实验设置

对于高斯图像去噪,我们使用 DIV2K 数据集集中的 800 幅图像作为训练集来训练模型,并选取 Set12 和 BSD68 测试数据集。Patch 大小为 96×96 , 批尺寸 (Batch) 大小为 16。整体网络训练的迭代次数 (Epoch) 设置为 60。初始学习率设置为 0.0001, 迭代 30 次后的学习率设置为 0.00001。Adam 优化器被用来优化学习率。通过 PyTorch 和 Python 3.9 训练和测试所提出的去噪网络。所有的实验都是在 Intel 酷睿 i5-3570 CPU、12G RAM 和 NVIDIA GeForce GTX 3080 Ti GPU 上进行。

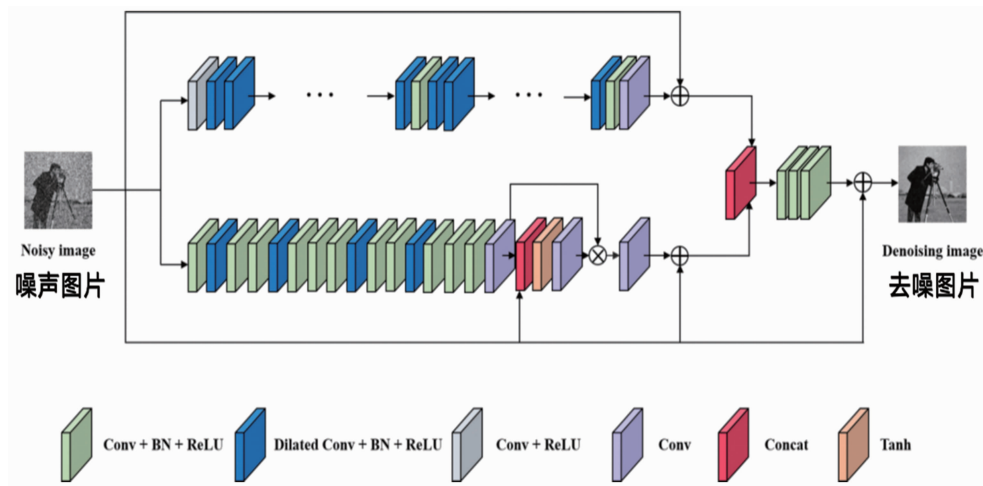


图 3 提出的网络结构

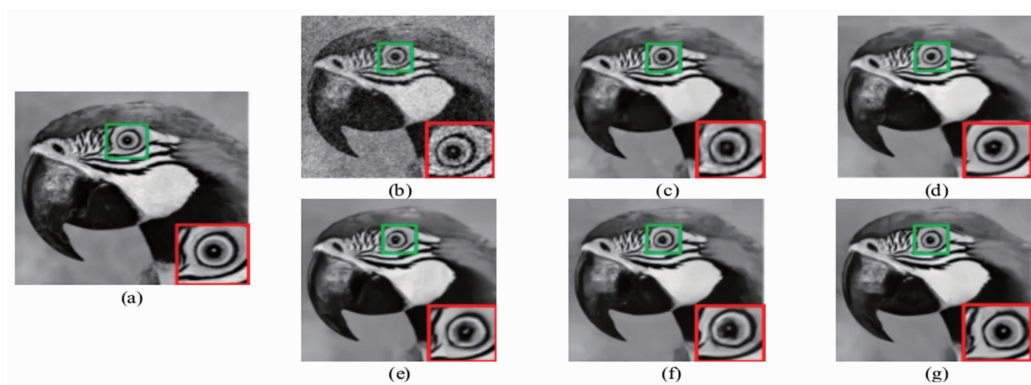


图 4 Set12 中的鸚鵡图片在不同网络结构下的去噪结果: (a)原图; (b)噪声图片($\sigma=50$); (c) BM3D; (d) DnCNN; (e) FFDNet; (f) ECNDNet; (g)提出的网络

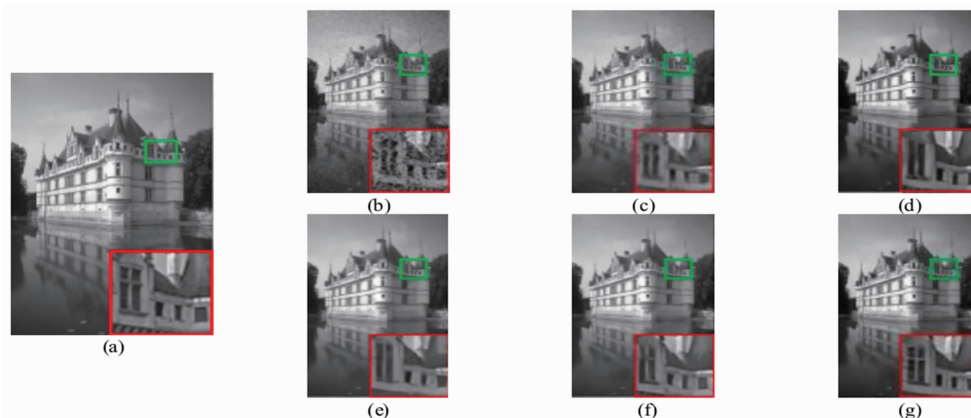


图 5 BSD68 中的城堡图片在不同网络结构下的去噪结果: (a)原图; (b)噪声图片($\sigma=50$); (c) BM3D; (d) DnCNN; (e) FFDNet; (f) ECNDNet; (g)提出的网络

表 1 Set12 和 BSD68 在噪声等级为 15、25 和 50 时的平均 PSNR 值

去噪算法	Set12			BSD68		
	噪声等级 σ			噪声等级 σ		
	15	25	50	15	25	50
噪声	24.60	20.17	14.15	24.61	20.17	14.15
BM3D	32.40	30.00	26.74	31.10	28.19	25.56
DnCNN	32.82	30.38	27.12	31.72	29.21	26.23
FFDNet	32.73	30.42	27.29	31.63	29.19	26.29
ECNDNet	32.81	30.39	27.15	31.71	29.22	26.23
所提网络	32.95	30.57	27.43	31.74	29.26	26.32

3.2 损失函数

PSNR 用于客观评价所训练的网络模型的图像去噪效果。在训练过程中,使用均方误差 (Mean Squared Error, MSE)作为网络的损失函数,计算生成图像和原始干净图像的 MSE。MSE 的计算公式为

$$MSE = L(\theta) = \frac{1}{2N} \sum_{i=1}^N \|R(y_i) - (y_i - x_i)\|^2 \quad (3)$$

表 1 列出了在 Set12 和 BSD68 数据集上分别采用四种去噪方法以及我们提出的去噪算法的平均 PSNR 值。可以看出,我们提出的网络在噪声水平分别为 15、25 和 50 的情况下优于 BM3D、DnCNN、FFDNet 和 ECNDNet。不同噪声水平下的 PSNR 最好和次好结果分别以粗体和斜体突出表示。当 Set12 数据集的噪声水平为 15 时,我们提出的网络的平均 PSNR 比 DnCNN 高 0.13 dB。当噪声水平为 25 时,我们提出的网络的平均 PSNR 比 FFDNet 高出

0.15 dB。当噪声水平为 50 时,我们提出的网络的平均 PSNR 比 FFDNet 高 0.14 dB。可以看出,对于数据集 BSD68,当噪声水平为 15、25 和 50 时,采用该方法分别比 DnCNN、ECNDNet 提高 0.02 dB、0.04 dB 和 0.03 dB。

如图 4 所示,我们选择 Set12 中鸚鵡的原始图像,用绿色矩形和红色矩形分别标记鸚鵡的整个眼睛和放大后眼睛的局部细节。当噪声水平为 50 时,与其他算法的结果图像中产生的边缘模糊和伪影问题相比,本文算法得出的图像边缘更清晰、更真实。图 4 中鸚鵡的眼睛轮廓变得更清晰,瞳孔变得更平滑。这表明我们的网络在恢复图像中某些物体的边缘和细节方面都有显著的改进。如图 5 所示,我们选择了 BSD68 的图像。使用绿色和红色矩形分别标记整个城堡的窗户部分和放大的窗户局部细节。当噪声水平为 25 时,其他算法的窗口边缘有轻微模糊、边缘恢复不佳等问题,并且丢失了部分窗缝的细节。我们的网络结果更清晰

地显示了窗口的边缘,并恢复了窗口接缝的细节。纹理结构的恢复程度高于对比算法恢复的图像质量。事实表明,我们提出的网络在人眼主观视觉评价中也取得了优异的结果。

4 结束语

本文提出了一种结合局部和全局特征的双路径 CNN 去噪算法。该模型结合了两种不同的网络架构,通过学习图像中更多不同的特征来增强图像去噪性能。此外,通过长路径将网络中的全局和局部相结合,提高了各层在噪声图像上的相关性。注意力机制的使用还为前一阶段网络层提供了补充信息。RL 用于从噪声图像中分离噪声,并在网络中获得潜在的干净图像。

实验数据表明,该算法的目标 PSNR 值较以往算法有较好的提高;主观视觉效果较好,对边缘轮廓和部分形状丢失细节的恢复质量较高,呈现出更清晰的图像恢复结果,体现了去噪的灵活性。虽然采用本文算法测出的结果优于现有算法,但该算法仍然在一些构图复杂的图像中效果不明显。后续将会研究一些特定细节的恢复,并继续在图像去噪方面开展相关工作。

参考文献

- [1] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising [C]. San Diego: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 2005.
- [2] Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering [J]. *IEEE Transactions on Image Processing*, 2007, **16**(8): 2080–2095.
- [3] Jain V, Seung S. Natural image denoising with convolutional networks [J]. *Advances in Neural Information Processing Systems*, 2008, **21**: 769–776.
- [4] Zhang K, Zuo W, Chen Y, et al. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising [J]. *IEEE Transactions on Image Processing*, 2017, **26**(7): 3142–3155.
- [5] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition [C]. Las Vegas: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016.
- [6] Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift [C]. Lille: 32nd International Conference on Machine Learning (ICML), 2015.
- [7] Tian C, Xu Y, Zuo W. Image denoising using deep CNN with batch renormalization [J]. *Neural Networks*, 2020, **121**: 461–473.
- [8] Tian C, Xu Y, Fei L, et al. Enhanced CNN for image denoising [J]. *CAAI Transactions on Intelligence Technology*, 2019, **4**(1): 17–23.
- [9] Zhang K, Zuo W, Zhang L. FFDNet: Toward a fast and flexible solution for CNN-based image denoising [J]. *IEEE Transactions on Image Processing*, 2018, **27**(9): 4608–4622.
- [10] Tian C, Xu Y, Li Z, et al. Attention-guided CNN for image denoising [J]. *Neural Networks*, 2020, **124**: 117–129.