

# 基于注意力机制的单幅雨滴图像增强

郑顾平 李金华 曹锦纲\*

(华北电力大学控制与计算机工程学院 河北 保定 071003)

**摘要** 在有雨环境下采集到的图片往往会掺杂雨滴或雨条纹,导致图片模糊,而现有去除雨滴实现图像增强算法容易引入更多噪声或不能去除较大雨滴。针对这一问题,提出一种基于注意力机制的单幅雨滴图像增强模型。采用多尺度扩张卷积进行特征提取,引入注意力机制来关注图像中雨滴区域的细节。实验结果表明,该模型可以有效地去除图片上的雨滴,实现雨滴模糊类图像细节的恢复。

**关键词** 扩张卷积 注意力机制 图像增强

**中图分类号** TP391 **文献标志码** A **DOI**:10.3969/j.issn.1000-386x.2020.09.026

## SINGLE RAINDROP IMAGE ENHANCEMENT BASED ON ATTENTION MECHANISM

Zheng Guping Li Jinhua Cao Jingang\*

(School of Control and Computer Engineering, North China Electric Power University, Baoding 071003, Hebei, China)

**Abstract** Pictures collected in a rainy environment are often doped with raindrops or rain stripes, resulting in blurred images. However, the existing image enhancement algorithm for raindrop removal is easy to introduce more noise or fail to remove larger raindrops. Aiming at this problem, this paper proposes a single raindrop image enhancement model based on attention mechanism. It used multi-scale dilated convolution for feature extraction and introduced attention mechanism to pay attention to the details of the raindrop region in the image. The experimental results show that the proposed model can effectively remove the raindrops on the pictures and realize the restoration of the details of the blurred image of raindrops.

**Keywords** Dilated convolution Attention mechanism Image enhancement

## 0 引言

所观测到的雨图像  $O \in \mathbf{R}^{M \times N}$  可以被建模为期望背景层  $B \in \mathbf{R}^{M \times N}$  和雨层  $R \in \mathbf{R}^{M \times N}$  的线性叠加<sup>[1-2]</sup>,用数学公式可以表示为  $O = B + R$ <sup>[3]</sup>。对因雨滴或雨条纹造成的图像模糊,其增强目的是从输入的雨图像  $O$  中分解出无雨背景  $B$ ,其中雨滴层叠加在包含真实场景内容的背景层上,可以将干扰雨滴或雨条纹去除表示为层分解问题,由此可去除图像中的雨滴或雨条纹,增强图像的可见性。

现在已经提出了一些方法来解决图像中雨条纹的去

除,如文献[2]使用双边滤波器将图像分解成低频部分和低频部分,然后通过基于稀疏编码的字典学习将高频部分分解为“雨分量”和“非分量分量”,继而从图像中去除雨水部分,但由于雨水的退化是复杂的,雨水模型不能很好地覆盖真实雨水图像中的一些重要因素,容易造成背景图像过度平滑。文献[1]采用从矩阵到向量结构的低秩模型来捕获时空相关的雨,由于雨和背景纹理图案之间的内在重叠,导致区域过度平滑。文献[4]通过分析每个像素位置处的椭圆核的旋转角度和纵横比来检测雨区域,检测到的雨区域使用非局部均值滤波来去除雨条纹,但该算法在检测不同的密度和大小的雨条纹时不够准确。雨滴虽不如雨条

纹在图像中分布密集,但雨滴在图像中的覆盖面积较大,雨滴图像增强相对困难。图像去除雨滴的研究相对较少,Tanaka等<sup>[5]</sup>通过用由视差参考获得的相应图像区域的纹理替换雨滴区域来从图像中去除雨滴,但容易造成图像加入更多噪声。文献[6]采用运动和强度时间导数来检测雨滴,在检测到雨滴后,通过使用时间强度导数对检测到的部分被雨滴遮挡区域求解混合函数来恢复图像,对于完全遮挡区域,使用视频完成技术恢复,但不适用于单幅雨滴图像。文献[7]提出了一种特殊形式的卷积神经网络来学习如何将雨滴覆盖的图像块映射到干净的图像块上,以实现图像增强,但只适合处理小雨滴类图像。文献[8]将视觉注意力注入生成和辨别网络,更加关注雨滴区域和周围结构,并使用对抗性训练端到端地直接输出处理后的图像。

本文提出了一种基于注意力机制的单幅雨滴图像增强模型,采用多尺度扩张卷积进行特征提取,引入注意力机制来关注图像中雨滴区域的细节,实现端到端的雨滴类模糊图像增强。

## 1 相关技术

### 1.1 扩张卷积

扩张卷积(Dilated Convolutions)又称膨胀卷积或空洞卷积,与其他卷积方式不同的是扩张卷积引入了扩张率参数,即扩张因子(Dilation rate),来表示扩张的大小。

扩张卷积的结构<sup>[9]</sup>如图1所示。设 $F_0, F_1, \dots, F_{n-1}: Z_1 \rightarrow R$ 为离散函数,令 $K_0, K_1, \dots, K_{n-2}: Z_2 \rightarrow R$ 为离散 $3 \times 3$ 滤波器,考虑应用指数增加膨胀的滤波:

$$F_{i+1} = F_{i \times 2} K_i \quad i = 0, 1, \dots, n-2 \quad (1)$$

将 $F_{i+1}$ 中像素 $p$ 的像素感受野区域定义为在 $F_0$ 基础上修改 $F_{i+1}(p)$ 值的像素值,假设 $F_{i+1}$ 中的像素 $p$ 的感受野大小为这些像素的数量,可以将像素在扩张卷积下的感受野大小表示为:

$$F_{i+1} = (2^{i+2} - 1) \times (2^{i+2} - 1) \quad (2)$$

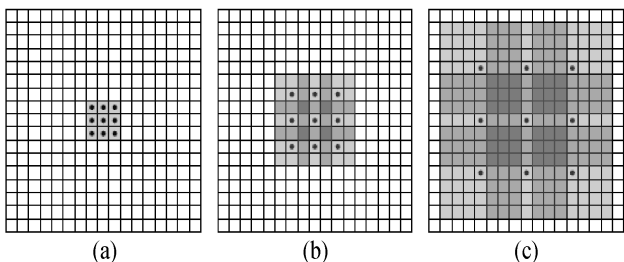


图1 扩张卷积结构图

图1中的点为卷积核,卷积核大小皆为 $3 \times 3$ ,阴影区域为像素的感受野。在图1(a)中,扩张因子的值为1,即 $F_1$ 是由 $F_0$ 进行扩张率为1的扩张卷积产生,可以看到像素的感受野与卷积核为 $3 \times 3$ 的普通卷积感受野相同。图1(b)是建立在图1(a)的基础上,扩张因子的值为2,即 $F_2$ 是由 $F_1$ 进行扩张率为2的扩张卷积产生,其感受野为 $7 \times 7$ ,与卷积核为 $5 \times 5$ 的普通卷积感受野相同。图1(c)是建立在图1(b)的基础上,扩张因子的值为4,即 $F_3$ 是由 $F_2$ 进行扩张率为4的扩张卷积产生,其感受野为 $15 \times 15$ ,与卷积核为 $9 \times 9$ 的普通卷积感受野相同。

由此可以看出,在卷积核的参数不变的情况下,像素的感受野随着扩张因子的增加而成指数增长。扩张卷积可以对像素进行加权和聚合信息,从而能够在不失去分辨率的情况下增强其感受视野,获得更多的特征信息,并可以保证输出的特征映射的大小不发生变化。本文采用不同扩张因子的多尺度扩张卷积进行特征提取。

### 1.2 注意力机制

视觉注意力机制是人类大脑中特有的视觉信号处理机制,注意力机制思想的本质类似人类的视觉注意力,主要意义是对大量的信息进行筛选,选择与当前任务相关的关键性高价值信息。注意力机制思想最早被应用于计算机图像视觉研究,近些年越来越多的研究者使用注意力机制思想与神经网络相结合进行相关问题研究。Mnih等<sup>[10]</sup>在RNN模型上引入了注意力机制来进行图像分类;Bahdanau等<sup>[11]</sup>将注意力机制应用到了自然语言处理方面;Gregor等<sup>[12]</sup>将深度递归神经网络结合了一种新颖的空间注意力机制,模仿人眼的动态,采用顺序变分自动编码框架,实现复杂图像生成。

## 2 模型设计

### 2.1 模型框架

本文提出的基于注意力机制的雨天模糊图像增强模型的整体结构如图2所示。采用不同扩张因子的多尺度扩张卷积增强其感受视野,以获得更多的特征信息;引入注意力机制构造注意力机制模型来关注图像中雨滴区域,以此更好地去除雨滴,实现雨滴模糊类图像细节的恢复。

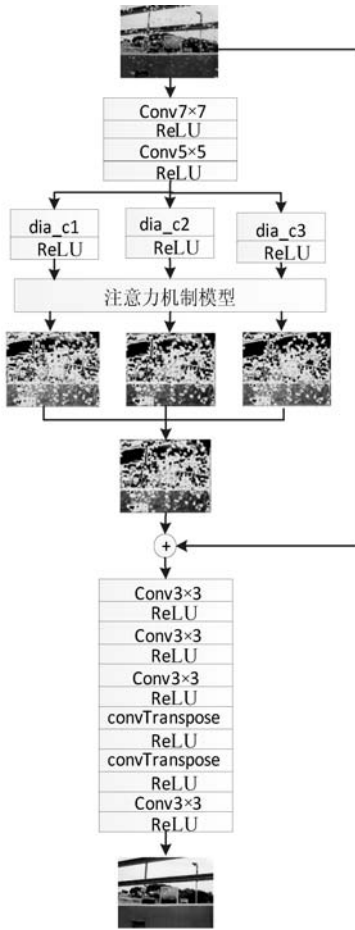


图2 基于注意力机制的雨天模糊图像增强模型

图2中:Conv、dia\_c和convTranspose分别代表卷积层、扩张卷积层和反卷积层;ReLU是激活函数。本文提出的基于注意力机制的雨天模糊图像增强模型包含:

1) 初步提取特征。输入的图像进行两次卷积核尺寸为 $7 \times 7 \times 64$ 、 $5 \times 5 \times 128$ 且步长为1的卷积处理,得到初步提取到的特征。

2) 多尺度扩张卷积提取特征。采用卷积核 $3 \times 3 \times 256$ 的扩张卷积,即dia\_c1、dia\_c2、dia\_c3,扩张因子的值分别为2、4、6,三种扩张卷积处理后得到三种特征。

3) 生成注意力图。将三种特征分别输入注意力机制模型,输出结果为三种注意力图,将三种注意力图与原图像做特征融合。

4) 输出图像。依次进行3次卷积核尺寸为 $3 \times 3 \times 256$ 且步长为1的卷积处理,2次卷积核尺寸为 $4 \times 4 \times 128$ 、 $4 \times 4 \times 64$ 且步长为2的反卷积操作,卷积核尺寸为 $3 \times 3 \times 64$ 且步长为1的卷积处理,最终输出除雨后的图像。

## 2.2 注意力机制模型

本文引入注意力机制帮助对雨滴图像的特征提

取。注意力机制可以使网络关注雨滴区域,能够更好地实现雨滴图像增强,设计的注意力机制模型如图3所示。

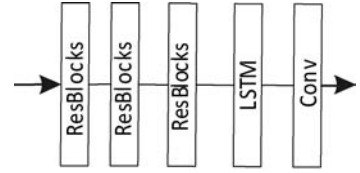


图3 注意力机制模型

可以看到,本文设计的注意力机制模型由三层残差网络(ResBlock)、长短记忆神经网络(LSTM)和卷积层(Conv)组成。

1) 残差网络:由于归一化层会忽略图像特征间的绝对差异,我们所用的残差网络为传统残差网络结构中去掉归一化层,具体结构如图4所示。

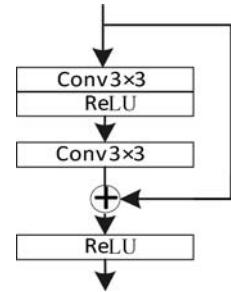
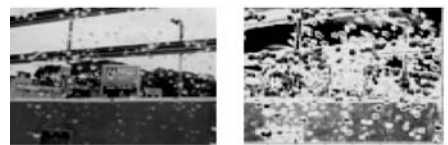


图4 残差网络

2) 长短记忆神经网络(LSTM):选用文献[13]提到的结构。

3) 卷积层:采用卷积核大小为 $3 \times 3$ 、步长为1的普通卷积。

图像经注意力机制模型的输出是注意力图,它是一个0到1的矩阵,也是一个非二元图,其值越大代表它表示的注意力越大。图5为原图和可视化注意力图。



(a) 原图 (b) 注意力图

图5 原图与注意力图

## 2.3 损失函数

本文使用的损失函数由两部分构成:注意力机制模型损失 $L_{att}$ 和感知损失 $L_p$ 。

$$L_{att} = \sum_{t=1}^N \theta^{N-t} L_{MSE}(A_t, M) \quad (3)$$

$$L_p = L_{MSE}(VGG(0), VGG(T)) \quad (4)$$

使用VGG16对网络进行预训练,提取高层的特征进行监督来保证生成图像的质量。式中: $A_t$ 为注意力机制模型在时刻 $t$ 时产生的注意力机制图; $M$ 为二进制

掩码;  $N$ 取4;  $\theta$ 取0.8;  $O$ 为模型处理后输出的去雨滴图像;  $T$ 为雨滴图像对应的原清晰无雨滴图像。

本文模型的总体损失函数为:

$$L = L_{\text{att}} + L_p \quad (5)$$

### 3 实验

#### 3.1 数据样本选取及评价指标

在现有公开的图像训练集中,并没有专门的雨滴类图像训练集,本文通过网络图片库和图像公开数据集等搜集了1 500对图像,每一对图像包含一幅雨滴模糊图和与之对应的无雨滴清晰图像,样例图像如图6所示。本文选取其中1 000对图像作为实验的训练集,剩余500对作为测试集。在进行网络训练之前,本文统一将图片分辨率的大小处理为 $720 \times 480$ 。



图6 样例图像

本文采用客观评价中常用的两种图片质量评价指标:峰值信噪比(PSNR)和结构相似度(SSIM),实验结果数据取平均值。

#### 3.2 模型训练

本文实验模型是建立在Pycharm平台上,采用深度学习TensorFlow框架实现,选用自适应矩估计(ADAM)算法来优化损失函数。模型通过训练数据集不断更新神经网络权重,经过多次训练调优,学习率 $\alpha$ 定为0.000 1,  $\beta_1 = 0$ ,  $\beta_2 = 0.9$ ,在训练过程中,我们将注意力图的值初始化为0.5。

实验的步骤如下:

1) 对数据集图片的分辨率大小进行统一处理,划分出训练集和测试集。

2) 建立基于注意力机制的雨滴模糊图像增强模型,并对网络的参数进行初始化。

3) 利用划分出的训练集对模型进行训练,通过损失函数进行反向传播不断更新模型参数,使计算出的损失最小,以使得模型性能效果最优。

4) 每50对图片为一次训练,设置最大迭代次数为3 000,保存模型参数,使用测试集测试模型。

#### 3.3 多尺度参数设定

本文算法采用具有不同扩张因子的多尺度扩张卷积进行特征值的提取,为研究多尺度参数的设置对实

现去雨效果的影响,在实验中设置了具有不同卷积核 $n$ 的扩张卷积, $n$ 取3、4、5,采用统一的扩张因子即为2、4、6,图7为采用不同的卷积核提取特征后得到的注意力图。图片质量评价结果如表1所示。

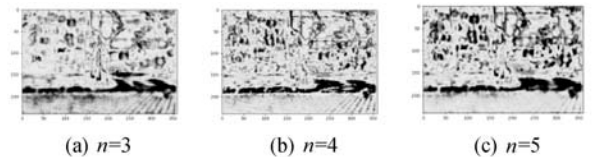


图7 不同尺度下的注意力图

表1 图片质量评价结果

卷积核尺度 $n$	PSNR/dB	SSIM
3	28.37	0.909 1
4	25.61	0.842 5
5	26.41	0.830 3

由图7可以看出,在卷积核大小为3时,得到的可视化注意力图的细节更加细腻,模型能够很好地注意到图像上更多的细节部分。由表1可以看出,在 $n = 3$ 时具有最高的峰值信噪比(PSNR)和结构相似度(SSIM),分别为28.37和0.909 1。因此,采用的扩张卷积的卷积核大小为 $3 \times 3$ 。

#### 3.4 模型性能对比

为说明本文模型的优越性,在3.1节建立的数据集上,与文献[5]及文献[7]模型进行对比实验,图片质量评价结果如表2所示,图8为测试集中3幅图片的实验效果图。

表2 图片质量评价结果

模型	PSNR/dB	SSIM
本文	28.37	0.909 1
文献[5]	26.84	0.851 5
文献[7]	27.41	0.875 3



(a) 雨滴图



(b) 清晰图



(c) 本文模型

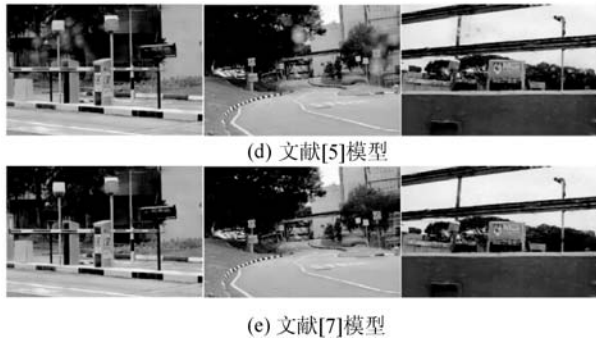


图8 实验结果对比图

由表2可以看出,文献[5]模型的峰值信噪比为26.84,文献[7]模型的峰值信噪比为27.41,而本文模型的峰值信噪比达到了28.37,同时本文提出的模型的结构相似度达到了0.9091高于另外两个模型的结构相似度。由图8可以看出,文献[5]模型不能去除较密集的雨滴,给图像引入了较多的噪声,文献[7]模型处理得到图片的细节不够丰富,出现失真现象,而经本文模型处理后的图片在主观的视觉感受方面取得了较好的效果。因此,本文算法模型无论是从峰值信噪比还是从结构相似度上来看都要优于其他的算法。

## 4 结 语

基于注意力机制思想,本文提出了一种单幅雨滴图像增强模型。首先采用具有不同扩张因子的多尺度扩张卷积对输入含雨滴图片进行特征值提取。其次,引入注意力机制关注图像中雨滴区域的细节,以更好地去除雨滴,恢复图像的细节。最后,实现端到端的雨滴类模糊图像增强。实验结果表明,本文模型可以有效地去除图片上的雨滴,增强雨滴类模糊图像,使图像细节更加丰富,具有较好的图像视觉效果。

## 参 考 文 献

[1] Chen Y L, Hsu C T. A generalized low-rank appearance model for spatio-temporally correlated rain streaks[C]//Proceedings of the IEEE International Conference on Computer Vision, 2013:1968-1975.

[2] Kang L W, Lin C W, Fu Y H. Automatic single image-based rain streaks removal via image decomposition[J]. IEEE Transactions on Image Processing, 2012, 21(4):1742-1755.

[3] Li Y, Tan R T, Guo X, et al. Rain streak removal using layer priors[C]//IEEE Conference on Computer Vision and Pattern Recognition. IEEE Computer Society, 2016:2736-2744.

[4] Kim J H, Lee C, Sim J Y, et al. Single image deraining using an adaptive nonlocal means filter[C]//Proceedings of IEEE International Conference on Image Processing. New York:

IEEE, 2013:914-917.

- [5] Tanaka Y, Yamashita A, Kaneko A. Removal of adherent waterdrops from images acquired with a stereo camera system[J]. IEICE Transactions on Information and Systems, 2006, 89(7):2021-2027.
- [6] You S, Tan R T, Kawakami R, et al. Adherent raindrop modeling, detection and removal in video[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, 38(9):1721-1733.
- [7] Eigen D, Krishnan D, Fergus R. Restoring an image taken through a window covered with dirt or rain[C]//Proceedings of the IEEE International Conference on Computer Vision, 2013.
- [8] Qian R, Tan T, Yang W, et al. Attentive generative adversarial network for raindrop removal from a single image[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.
- [9] Yu F, Koltun V. Multi-scale context aggregation by dilated convolutions[C]//International Conference on Learning Representation, 2016.
- [10] Mnih V, Heess N, Graves A, et al. Recurrent models of visual attention[C]//Advances in Neural Information Processing Systems, 2014:2204-2212.
- [11] Bahdanau D, Cho K, Bengio Y. Neural machine translation by jointly learning to align and translate[EB]. arXiv preprint arXiv:1409.0473, 2014.
- [12] Gregor K, Danihelka I, Graves A, et al. Draw: A recurrent neural network for image generation[EB]. arXiv preprint arXiv:1502.04623, 2015.
- [13] Shi X J, Chen Z, Wang H, et al. Convolutional LSTM network: A machine learning approach for precipitation nowcasting[C]//Advances in neural information processing systems, 2015:802-810.

(上接第117页)

- [16] Bodla N, Singh B, Chellappa R, et al. Soft-NMS—improving object detection with one line of code[C]//2017 IEEE International Conference on Computer Vision (ICCV), 2017.
- [17] Geiger A, Lenz P, Stiller C, et al. Vision meets robotics: the KITTI dataset[J]. The International Journal of Robotics Research, 2013, 32(11):1231-1237.
- [18] Wen L, Du D, Cai Z, et al. UA-DETRAC: a new benchmark and protocol for multi-object detection and tracking[EB]. arXiv:1511.04136, 2015.
- [19] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
- [20] Pepik B. Occlusion patterns for object class detection[C]//IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2013.