

基于扩散模型的天气退化图像修复技术

刘德崇¹ 姚剑敏^{1,2} 胡海龙^{1,2} 陈恩果¹ 严群¹ 程慧敏¹ 徐任政¹

LIU Dechong YAO Jianmin HU Hailong CHEN En'guo YAN qun CHENG Huimin XU Renzheng

摘要

在恶劣天气中拍摄的图片,由于雨雪霾等恶劣天气的影响会对视线有一定的阻挡,导致可见度下降严重,图片的背景失真严重,从而对图片识别、语义分割或者目标监测产生很大的误差。此前,大多数的天气退化图像的修复都是基于深度学习算法或是生成对抗网络为基本架构的,由于去噪扩散模型(DDPM)在计算机视觉上的优势比较大,所以采用以扩散模型为基本架构来进行雨类的天气退化图像的反向采样,经典扩散模型的噪声估计网络是基于 U-Net 的结构,现提出了一种改进的 U-Net 噪声估计网络结构,将通道自适应注意力机制与 U-Net 结合,在图像去雨方面有着较好的恢复表现。

关键词

扩散模型; 通道自适应注意力机制; 天气退化; 图像修复; 图像去雨

doi: 10.3969/j.issn.1672-9528.2024.03.050

0 引言

天气退化图像修复技术的应用具有广泛的需求背景,在像视频监控、自动驾驶等应用场景中,受天气退化影响的图像可能导致对环境信息的理解和识别能力下降,影响相应系统的性能和安全性。通过修复天气退化图像,可以改善环境信息的获取和分析,提高相关应用系统的可靠性和性能。恶劣的天气条件,如雨雪、大风、雾霾等,会导致拍摄的图像质量受损,表现为模糊、噪点、对比度失真等问题。通过天气退化图像修复,可以提高图像的清晰度和对比度,恢复图像的可视质量,提高图像的观赏性和可理解性。

鉴于以上种种需求,就需要对天气图像退化修复技术进行深层次的研究与探索,考虑到生成图像的表现可以更真实地与天气退化现象贴切,所以本次研究采用的是以扩散模型为基本的结构,准确反映了真实的退化现象,解决了仿真与实测样本存在差异性的问题。

图像生成模型具有生成能力、多样性、控制性、学习数据分布和高度表达能力等特点,为本文创造逼真、多样、可控的图像样本提供了有力的工具和方法。比如,变分自编码器(VAE),2013年,Kingma等人提出了变分自编码器(VAE)^[1]。VAE是一种基于概率推断的生成模型,通过学习数据的潜在分布来生成新的样本。VAE通过将输入数据映射到一个潜在空间,并通过优化潜在空间中的概率分布来生

成样本。但是为了平衡两个损失函数,VAE倾向于产生一些模糊的重构图像,影响了实际的表现。生成对抗网络(GAN),2014年,Goodfellow等人提出了生成对抗网络(GAN)^[2],GAN由生成器和判别器组成^[3],它们通过博弈过程相互竞争,生成器试图生成逼真的图像,

而判别器试图区分真实图像和生成图像。但由于生成器倾向于产生相似的图像样本,而忽略了其他可能性。这可能导致生成的样本缺乏多样性、缺乏全面性,无法涵盖整个数据分布。

本次研究基于扩散模型为基础来反向采样生成图片,采用改进的噪声估计网络,将 U-Net 网络与通道注意力机制融合在雨类等天气退化图片中,使其有较好的生成表现力^[4]。

1 相关理论

1.1 扩散模型的介绍

DDPM 是一种生成模型,旨在对图像数据进行建模。DDPM 通过在一系列梯度可逆的扩散步骤后,使用对称网络来对噪声分布进行建模。该模型可以用于生成高质量、高保真度的图像,包括自然图像和艺术风格图像。DDPM 的核心思想是将图像数据视为梯度可逆的变换,通过扩散算法逐步转换噪声信号以生成图像。在这个过程中,采用对称网络结构来建模噪声信号的概率分布,从而使得生成的图像保持高质量和真实感。与传统的生成对抗网络(GAN)相比,DDPM 技术上更加稳定,能够更容易生成高分辨率的图像,并且能够更好地处理图像的细节和结构。

扩散模型的主要过程可分为两个部分。

前向过程^[5](diffusion process):可以理解为将一滴墨

1. 福州大学 福建福州 350108

2. 晋江市博感电子科技有限公司 福建晋江 3662200

[基金项目] 国家重点研发计划(2022YFB3603503),福建省技术攻关重点项目(2023G007)

汁滴入水中并在水中扩散直到整个系统达到平衡，前向过程就是每一步都在图片上加一点噪声，直到整张图片变成一幅纯噪声。 t 时刻的噪化状态和 $t-1$ 时刻的关系用公式描述为：

$$x_t = \sqrt{a_t}x_{t-1} + \sqrt{1-a_t}z_1 \quad a_t = 1 - \beta_t \quad (1)$$

式中： β_t 是噪声调节因子， z 为本文所加入的噪声。

$$x_t = \sqrt{a_t}x_{t-1} + \sqrt{1-a_t}z_2 \quad (2)$$

故可得到 x_t 与 x_0 之间的关系：

$$x_t = \sqrt{a_t}x_0 + \sqrt{1-a_t}z_t \quad (3)$$

因此可得到一个综合的马尔可夫性联合扩散过程的表达式为：

$$q(x_1, x_2, \dots, x_t | x_0) = \prod_{t=1}^T q(x_t | x_{t-1}) \quad (4)$$

噪化状态为：

$$q(x_t | x_0) = N(x_t; \sqrt{a_t}x_0, (1 - a_t)I) \quad (5)$$

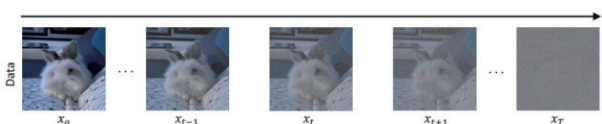


图1 前向过程原理图

逆向过程（Denosing Process）：主要是根据已有的被噪化的图片，学习后估计噪声分布，从而预测出上一步的图片的状态，也就是对图片反向采样去噪的过程，可以将其表示为：

$$p_\theta(x_0, x_1, \dots, x_T) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1} | x_t) \quad (6)$$

从而得到条件分布的表达式：

$$p_\theta(x_{t-1} | x_t) = N(x_{t-1} | \mu_\theta(x_t, t), \Sigma_\theta(x_t, t)) \quad (7)$$

1.2 条件扩散模型

在对图像修复的过程中，所使用的扩散模型与去噪扩散概率模型略有不同，其在前向过程中的任务与经典的模型是一样的。但不同的是，其在反向采样过程中引入图像作为条件，相当于是将低分辨率的图像和噪声的图像进行了一次拼接^[6]。所以反向采样的公式变为：

$$p_\theta(x_0, x_1, \dots, x_T | \hat{x}) = p(x_T) \prod_{t=1}^T p_\theta(x_{t-1} | x_t, \hat{x}) \quad (8)$$

式中： $\hat{x}$ 是引入的作为条件的退化图像。

因此可以得到 $t-1$ 时刻的噪声分布公式：

$$x_{t-1} = \sqrt{a_{t-1}} \left(\frac{x_t - \sqrt{1-a_t} \epsilon_\theta(x_t, \hat{x}, t)}{\sqrt{a_t}} \right) + \sqrt{1-a_{t-1}} \epsilon_\theta(x_t, \hat{x}, t) \quad (9)$$

式中： $\epsilon_\theta(x_t, \hat{x}, t)$ 为当 $\hat{x}$ 为引入的作为条件的退化图像时噪声分布估计。

1.3 噪声估计网络

去噪扩散模型（DDPM）的噪声估计网络部分是基于U-Net的架构，U-Net网络架构典型的特征是一个U型的Encoder—Decoder结构，是一个对称的全卷积神经网络。在前向和反向过程中，U-Net网络部分的作用都是在当前的样

本和时间下对噪声预测。它包括一个向下采样路径（Encoder模块）和一个向上采样路径（Decoder模块）。Encoder模块负责从输入图像中提取高级抽象特征，而Decoder模块则将这些特征映射回原始图像的尺寸。具体来说，U-Net的Encoder部分由一系列卷积操作和池化操作组成，用来逐渐降低特征图的尺寸和通道数。这样可以扩大感受野并提取图像的全局信息。然后，Decoder部分通过上采样操作（通常是反卷积或转置卷积）逐渐恢复特征图的尺寸，并与编码器对应层级的特征进行拼接。这种设计可以同时利用局部和全局信息，使网络能够更好地还原细节和边界信息。确保在图片还原过程中，保留住原有步骤的信息。这种设计可以同时利用局部和全局信息，使网络能够更好地还原细节和边界信息。

在U-Net中，为了更好地控制感受野范围和降低计算量，可以使用跳跃连接（skip connections）。这些连接直接将编码器中的特征图与解码器中对应层级的特征图相连接，使解码器能够直接访问更低层次的特征。这有助于提供更多的上下文信息，以改善预测结果。U-Net结构图如图2所示。

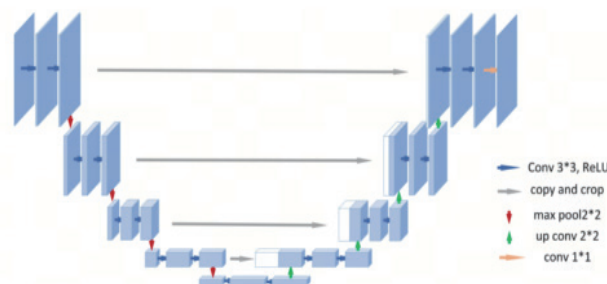


图2 U-Net 结构图

1.4 改进噪声估计网络

传统U-Net结构的噪声估计网络在天气退化图片生成中存在着明显的问题^[7]，当雨线比较密集的时候，生成的恢复图片存在着特征缺失的问题，为了解决这一问题，采用在U-Net网络结构中添加高效通道注意力模块ECA（efficient channel attention）^[8-9]的方法，从而增强U-Net模型对通道信息的关注和任务特定特征的提取能力。

通道注意力机制如图3所示，首先，上一层输入的 X ，送入残差网络，经过一个全局的平均池化（global pooling）将每个通道的特征图都池化成一个值，这个值就代表每个通道的全局特征，也可以将其看成是向量的形式；然后经过一个降维的全连接层（FC）和ReLU之后通过一个升维的全连接层就变成了通道权重向量^[10]；最后通过Sigmoid与原来的特征图 X 相乘得到带权重的特征图的输出。通过对每个通道（channel）的特征图赋予不同的注意力权重，以实现对不同通道特征的加权。这种机制可以使网络在学习和提

取特征时更加关注重要的通道,从而提升网络对于重要特征的捕捉能力^[11]。

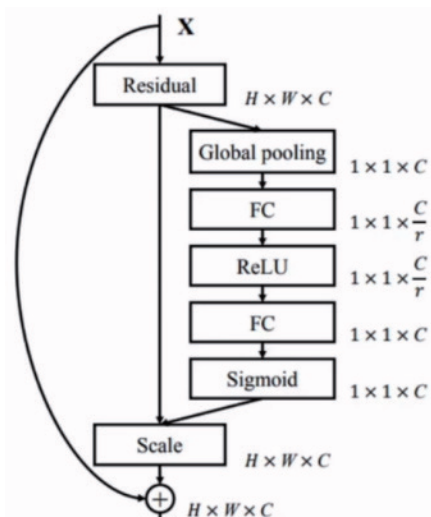


图3 通道注意力模块原理图

相较于其他的通道注意了模块, ECA 比较显著的优势首先是在模型的复杂度方面 ECA 的参数量少, 性能也有明显的提升, 有比较好的竞争力。其次, 通道注意力模块在全局平均池化之后维度是否降低对于其学习非常重要, 避免维度降低导致其对通道注意力模块的预测功能产生副作用, 对于在不同的通道之间特征的捕获效率较低的模块来说, 经过降维在升维的操作之后, 每个特征值与每个特征权重之间的对应关系变得更间接了。因此, ECA 采用了 1 维卷积的不降维—局部跨通道交互的操作。首先特征图经过全局平均池化之后, 经过一个一维的卷积得到一个通道选中的向量, 再经过一个 Sigmoid 得到值为 (0-1) 之间的通道, 选中通道向量后与原来的特征图相乘, 从而得到带通道权重的特征图。高效通道注意力 (ECA) 原理如图 4 所示。

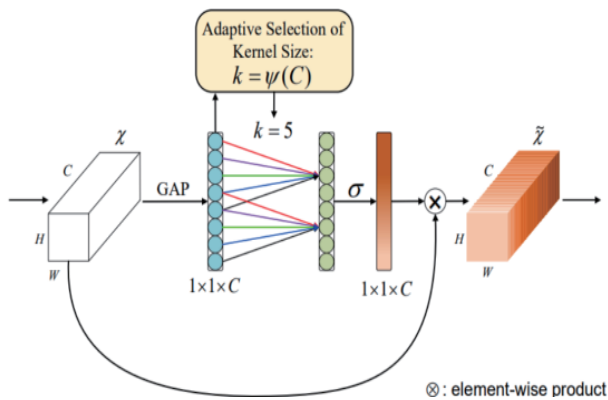


图4 高效通道注意力 (ECA) 原理

本次研究的工作主要是在编码器的模块中结合通道注意力机制 (ECA)^[12], 以增强通道的关注并挖掘重要特征。在编码器的每个模块输出后, 将特征图输入到 ECA 注意

力模块中, 然后将加权后的特征图作为下一个编码器模块的输入。

2 实验结果和分析

2.1 数据集

本次实验所采用的数据集有: 去雨的数据由 Rain100H, Rain800 以及 SPA-Data 组成。Rain800 是一个常用的用于雨天图像去雨算法评估的数据集, 包含有 800 张具有挑战性的雨天图像。Rain100H 数据集是一个广泛用于雨天图像去雨的数据集之一, 包含有 100 张高分辨率的雨天图像。Rain1200 数据集是一个较大规模的用于雨天图像去雨的数据集, 包含有 1200 张图像, 并提供了真实世界和合成的雨天图像, 设置 TEST 的子集作为测试集。

2.2 实验环境

本次实验的硬件环境采用了 NVIDIA Tesla P40 GPU (16 GB), 采用了 adam 训练神经网络模型, 所使用的软件有 PyTorch1.5, VS Code, 服务器为 Ubuntu 完成了模型的训练。所设置的迭代次数为 5000, 学习率为 $2e-4$, 每批次输入 64 对图像。

2.3 实验结果与分析

为了验证此次实验有良好的表现, 实验选择图像质量评价指标 PSNR (peak signal-to-noise ratio) 与 SSIM (structural similarity index) 来对生成的图片进行评估^[13], 并与其他表现良好的天气退化图像算法进行对比。PSNR 指的是峰值信噪比, 峰值信噪比是一种常用于评估图像或视频质量的度量标准。它通过比较原始图像和经过处理的图像之间的均方误差来衡量两者之间的失真程度, 计算公式为:

$$PSNR = 10 \log_{10} \left(\frac{MAX^2}{MSE} \right) \quad (10)$$

MSE 是均方误差 (Mean Squared Error), 用来衡量两幅图像之间的差异。MAX 表示图像像素值的最大可能取值 PSNR 的数值, 通常以分贝 (dB) 为单位, 数值越大表示图像质量的失真越小。SSIM 是结构相似度指标, 是一种用于测量图像或视频质量的度量标准, 它充分考虑了人眼对图像质量的敏感性, 同时也考虑了图像结构的特征, 更适合于评估失真后的图像质量。SSIM 可用于评估图像压缩、去噪、去模糊、超分辨率等各种图像处理和恢复技术。与 PSNR 相比, SSIM 考虑了图像的结构信息, 并更贴近人类感知。SSIM 的取值范围是 [0,1], 其中数值接近 1 表示图像的结构相似度较高, 接近 0 表示图像的结构相似度较低。图 5 为是此次实验天气退化图像所生成的效果图。



(a) 退化图像 (b) Uformer (c) TransWeather (d) Ours

图 5 不同的天气退化算法的对比图片

从图 5 中可以看出，本次实验在雨类的天气退化图像恢复上的效果表现，修复后的图像在视觉上更加自然和清晰，该实验对于退化上的处理更加准确。

对比 PSNR 和 SSIM 在测试集上的评估可以看出本次研究的两项指标对比之前的研究方法都有了小幅度的提升，这可以说明，通过将扩散模型中的 U-Net 网络与通道注意力模块相融合，确实对雨类的天气退化图像有较好的恢复能力，可以高质量地对图片进行恢复处理。

表 1 不同模型结果对比表

网络	PSNR	SSIM
MprNet	29.36	0.901
TransWeather	29.11	0.877
Ours	29.96	0.923

3 结论与展望

针对雨类天气退化图像的修复任务，本文提出了一种基于扩散模型的恢复策略，通过对噪声估计网络结构的改进，即将通道注意力机制与 U-Net 网络中的 Encoder 过程融合，在去雨图像修复上有着良好的表现。

综上所述，基于扩散模型的天气退化图像的修复，在去雨一类上的修复有着很好的表现能力，同时对其他恶劣的天气场景相信也会有不错的还原能力。但是，相对于其他的算法来说，基于扩散模型的天气退化图像的修复在生成速度上相对来说是比较慢的，未来的研究可以把重心放在如何加速扩散模型的采样过程，从而提升整体图像恢复效率上，同时训练雨雪霾等恶劣天气融合在一起的大一统模型。

参考文献：

[1] KINGMA D P, WELLING M. Auto-Encoding Variational Bayes[EB/OL].(2013-11-20)[2023-11-01].<https://arxiv.org/abs/1312.6114>.

[2] GOODFELLOW I, POUGET-ABADIE J, MIRZA M, et al. Generative adversarial networks[J].Communications of the ACM, 2020, 63(11): 139-144.

[3] 李然, 周子溪, 张月芳, 等. 一种结合自注意和多尺度生成对抗网络的图像去雨方法 [J]. 计算机工程与科学, 2021, 43(12): 2216-2222.

[4] 杨爱萍, 王朝臣, 王建, 等. 基于自适应感知金字塔网络

的图像去雨 [J]. 东北大学学报 (自然科学版), 2022,43(4): 470-479.

[5] DHARIWAL P, NICHOL A. Diffusion models beat gans on image synthesis[J].Advances in neural information processing systems, 2021, 34: 8780-8794.

[6] SAHARIA C,HO J,CHAN W, et al.Image super-resolution via iterative refinement[J].IEEE transactions on pattern analysis and machine intelligence, 2022, 45(4): 4713-4726.

[7] SYED W, ADITYA A, SALMAN K,et al. Restormer: efficient transformer for high-resolution image restoration[C]//In IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE,2022:5728–5739.

[8] ZAMIR S W, ARORA A, KHAN S, et al. Restormer: efficient transformer for high-resolution image restoration[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).Piscataway: IEEE, 2022: 5728-5739.

[9] WANG Q, WU B, ZHU P, et al. ECA-Net:efficient channel attention for deep convolutional neural networks[C]// IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway:IEEE, 2020:11531-11539.

[10] 柴国强, 王大为, 芦宾, 等. 基于注意机制的轻量化稠密连接网络单幅图像去雨 [J]. 北京航空航天大学学报, 2022, 48(11): 2186-2192.

[11] ZHANG Y, LI K, LI K,et al. Residual non-local attention networks for image restoration[EB/OL].(2019-05-24)[2023-10-26].<https://arxiv.org/abs/1903.10082>.

[12] SEONGHYEON N, YOUNGBAE H, YASUYUKI M, et al. A holistic approach to cross-channel image noise modeling and its application to image denoising[C]//In IEEE Conference on Computer Vision and Pattern Recognition. Piscataway: IEEE, 2016:683–1691.

[13] WANG Z, CUN X, BAO J, et al. Uformer: a general U-shaped transformer for image restoration[C] // Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway: IEEE, 2022: 17683-17693.

【作者简介】

刘德崇（1998—），男，山东聊城人，硕士，研究方向：深度学习、图像处理等。

姚剑敏（1978—），男，福建莆田人，博士，副研究员，研究方向：人工智能、图像处理、计算机视觉等。

（收稿日期：2023-12-27）