

密度与颜色信息引导的文本图像阴影消除算法^①



柏 畅, 张 玲

(武汉科技大学 计算机科学与技术学院, 武汉 430065)

通信作者: 张 玲, E-mail: zhling@wust.edu.cn

摘 要: 现有文本图像阴影消除方法已取得了一定的进步, 但是这些方法主要关注图像本身和文本背景颜色信息, 忽视了真实场景中文本图像通常伴随密度不一致的阴影, 因此这些方法可能存在图像局部细节丢失、内容不协调等问题. 针对这个问题, 本文提出了一种基于密度和颜色信息引导的文本图像阴影消除算法. 首先, 设计了特征提取模块, 以提取输入图像的阴影密度和背景颜色信息. 随后, 在每一级模块输入之前, 利用自适应模块调整初始的阴影密度特征, 并结合背景颜色信息指导后续的阴影消除过程. 为了更好地提取图像的全局与局部特征, 本文提出了密度和颜色引导的 Transformer 模块和阴影敏感的局部特征提取模块, 并将两者结合作为主要阴影消除模块. 实验结果表明, 相比现阶段的文本图像阴影消除方法, 所提出的网络模型在性能上更加优越.

关键词: 文本图像阴影消除; 深度学习; 阴影密度指导; 背景颜色指导; 自适应调整

引用格式: 柏畅,张玲.密度与颜色信息引导的文本图像阴影消除算法.计算机系统应用,2025,34(7):152-162. <http://www.c-s-a.org.cn/1003-3254/9884.html>

Document Image Shadow Removal Algorithm Guided by Density and Color Information

BAI Chang, ZHANG Ling

(School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan 430065, China)

Abstract: Existing methods for shadow removal in document images have made certain progress. However, these approaches primarily focus on the image itself and the background color information, neglecting the fact that in real-world scenarios, document images often contain shadows with varying densities. Consequently, these methods may result in issues such as loss of local details and content inconsistencies. To address these challenges, a document image shadow removal algorithm guided by density and color information is proposed. A feature extraction module is first designed to extract shadow density and background color information from the input image. Before each module receives its input, an adaptive module adjusts the initial shadow density features, which are then combined with the background color information to guide the subsequent shadow removal process. To better extract both global and local features, a density- and color-guided Transformer module, along with a shadow-sensitive local feature extraction module, is proposed. The two modules are combined to form the primary shadow removal module. Experimental results demonstrate that, compared to current methods, the proposed network model outperforms existing text image shadow removal techniques.

Key words: document image shadow removal; deep learning; shadow density guidance; background color guidance; adaptive adjustment

① 基金项目: 湖北省自然科学基金 (2023AFB615)

收稿时间: 2024-11-13; 修改时间: 2024-12-09, 2025-01-15; 采用时间: 2025-01-21; csa 在线出版时间: 2025-05-27

CNKI 网络首发时间: 2025-05-28

在日常生活中,文本随处可见,如教科书、传单、杂志和漫画等。这些文本通常需要保存为电子文件,以便进行数字存档或在线消息传输。随着移动设备的广泛流行,人们更喜欢使用各类便捷移动设备来数字化文本副本。然而,拍摄过程中,光源常被拍摄设备或拍摄者的手臂等遮挡,导致图像上出现阴影^[1,2],严重影响视觉质量和信息可读性,甚至妨碍后续自动化处理任务,如目标检测^[3,4]和图像分割^[5]等。在这些场景中,清晰的文本图像对于准确的目标检测至关重要。因此,文本图像阴影消除成为各类视觉应用中的一项必要图像处理任务。

尽管自然图像阴影消除已经取得了显著进展^[6-23],但是这些方法在文本图像上的表现通常不佳,原因在于它们与自然图像的特征截然不同。例如,自然图像强调背景内容(无阴影图像),而文本图像则侧重于文本内容。如果不考虑文本图像的特定属性,传统的自然图像处理方法在应用于文本图像时通常会产生不正确的结果,而基于深度学习的方法同样也会受到影响。

目前的文本图像阴影消除方法深入研究了文本图像的特定特征,然而在处理复杂光照条件下的图像时,往往会导致阴影残留或内容不一致的问题。特别是在多光源的照明环境中,文本上的阴影表现出多样化和高度不均匀的特性。具体而言,文本中的阴影可以大致分为高密度的硬阴影和低密度的软阴影。硬阴影通常会导致下层表面的纹理和颜色信息大量丢失,而软阴影则由于照明强度不均匀,其边界与非阴影区域逐渐融合。不同阴影类型带来了多样化的处理挑战。

为了解决这个问题,本文提出了一种基于密度和颜色信息引导的文本图像阴影消除网络。该网络能够自适应地处理不同类型的阴影。并通过对阴影密度进行自适应调整,再结合背景颜色信息对图像进行补光,以实现更精准的阴影去除效果。首先,本文构建了一个特征提取模块,用于提取输入文本图像的初始阴影密度信息和背景颜色信息。然后,在输入到每一级模块之前,本文创新地提出了自适应密度调节模块,它可对初始阴影密度信息进行调整,使得网络能够获取动态的阴影密度信息,让阴影消除过程更加精准。此外,本文还设计了密度和颜色引导的 Transformer 模块和阴影敏感的局部特征提取模块,并将它们结合使用作为模型的主要阴影消除模块。通过自适应调整阴影密度信息和背景颜色信息,网络能够持续更新和优化图像的全局和局部特征,从而进一步提升阴影消除的效果。实

验证明,相比现阶段的文本图像阴影消除方法,本文提出的网络模型在性能上更加优越。综上所述,本文的贡献可以总结为以下几点。

(1) 设计了一个结合阴影密度和背景颜色引导的文本图像阴影消除网络,并构建了一个特征提取模块,用于更准确地提取输入文本图中的阴影密度和背景颜色信息。

(2) 引入了一个自适应密度调整模块,通过在不同阶段对初始阴影密度信息进行调整,实现对后续模块的精确密度信息指导,从而提升阴影消除效果的质量。

(3) 提出了密度和颜色引导的 Transformer 模块和阴影敏感的局部特征提取模块,并将它们结合作为网络的基本组件。Transformer 和 CNN 的结合使网络能够处理长距离依赖关系的同时,关注图像的局部特征,进一步提高阴影消除结果的细节保留能力和内容一致性。

1 相关工作

1.1 自然图像的阴影消除算法

传统的自然图像阴影消除方法通常关注阴影的不同物理特性^[10,11]。Finlayson 等人^[11]基于梯度一致性重建了阴影消除的图像。然而,由于光照变化,这些方法容易引入明显的阴影边界伪影。Shor 等人^[12]定义了阴影区域和非阴影区域之间的仿射关系。Xiao 等人^[13,14]通过将光照从非阴影区域传递到阴影区域来消除阴影。然而,这些方法依赖于参考的非阴影区域,当参考区域不理想时,往往会导致光照不一致。

许多基于学习的方法已经被提出用于自然图像的阴影消除。例如,DeshadowNet^[15]提取了多上下文特征来预测阴影层以消除阴影。Wang 等人^[8]使用堆叠的条件 GAN 进行联合的阴影检测和消除。Zhang 等人^[17]探索了使用 GAN 的残差和光照进行阴影消除。ARGAN^[18]利用注意力递归生成对抗网络来进行阴影检测和消除。Liu 等人^[19]通过生成阴影通过弱监督训练进行阴影消除。最近,ShadowFormer^[9]通过在嵌入的特征空间中上下上下文信息从非阴影区域传递到阴影区域。

虽然这些方法对自然图像有效,但由于自然图像和文本图像的特性差异,它们在文本图像阴影消除任务中表现不佳。

1.2 文本图像的阴影消除算法

大多数传统的文本阴影消除算法使用启发式算法来挖掘文本图像的特定特征。Bako 等人^[1]通过估计阴

影消除了阴影,但在强阴影下该方法在边界上仍然留有痕迹. Oliveira 等人^[20]使用自然邻域插值法来估计阴影图. Jung 等人^[21]采用水填充法通过将输入图像转换为地形表面来校正文本图像的光照.该方法在弱或中等水平的阴影下表现良好,但在处理重阴影场景时容易产生颜色退化的结果.

最近, Lin 等人^[22]提出了通过估计恒定背景来消除文本图像阴影的 BEDSR-Net. 这是第 1 个专门为文本图像阴影消除设计的深度网络, 利用了文本图像的特定属性. Zhang 等人^[23]进一步提出颜色感知背景提取网络 (CBENet), 提取空间变化的背景以提升阴影消除效果. 然而, 这些方法在强调利用文本背景颜色信息的同时忽视了阴影密度信息对于文本阴影消除任务的影响, 这在处理高对比度图像时可能导致色彩失真. 而 Li 等人^[24]提出了频率感知阴影消除网络 (FSENet), 结合拉普拉斯金字塔的多频分解技术, 将高分辨率图像分解为高频与低频分量, 其中低频部分主要负责光照和颜色校正, 高频部分则用于细节与边缘恢复. 然而, FSENet 在消除阴影的过程中忽视了文本的特定属性也没有使用阴影密度信息选择合适的高低频特征进行融合, 导致了其低频模块无法完全抑制颜色失真, 高频模块则可能引入噪声, 使得最终消除结果出现较明显的褪色伪影和残留阴影.

2 文本图像阴影消除算法

本节首先介绍提出模型的总体架构. 接着详细讲解模型中的各个模块, 包括阴影密度信息提取模块、阴影密度信息自适应调整模块、密度和颜色引导的 Transformer 模块以及阴影敏感的局部特征提取模块. 最后介绍模型所使用的损失函数.

2.1 网络总架构

本文提出的基于密度和颜色信息指导的文本图像阴影消除网络如图 1 所示. 该网络由特征提取模块、自适应密度调整模块和阴影消除模块 3 部分组成, 其中自适应密度调整模块和阴影消除模块共同组成主干网络, 采用了 3 级水平编码器和解码器结构. 这种结构设计使得网络能够逐级提取和处理图像的多尺度特征, 从而更有效地应对不均匀的阴影强度和复杂背景颜色. 首先, 特征提取模块根据输入图像 I_{shadow} 提取对应的阴影密度特征 S_1 、 S_2 、 S_3 和背景颜色特征 C_1 、 C_2 、 C_3 . 由于在网络的不同编码和解码部分, 图像处于不同的状态, 因此在网络的每一层前使用自适应密度调整模块对阴影密度特征进行调整, 并利用调整后的特征对后续模块进行指导. 这使得网络在每个阶段都能获取有关阴影密度的信息, 有助于在特征提取过程中更准确地捕捉阴影的特性, 从而提高特征提取质量, 避免过度补光导致的图像失真.

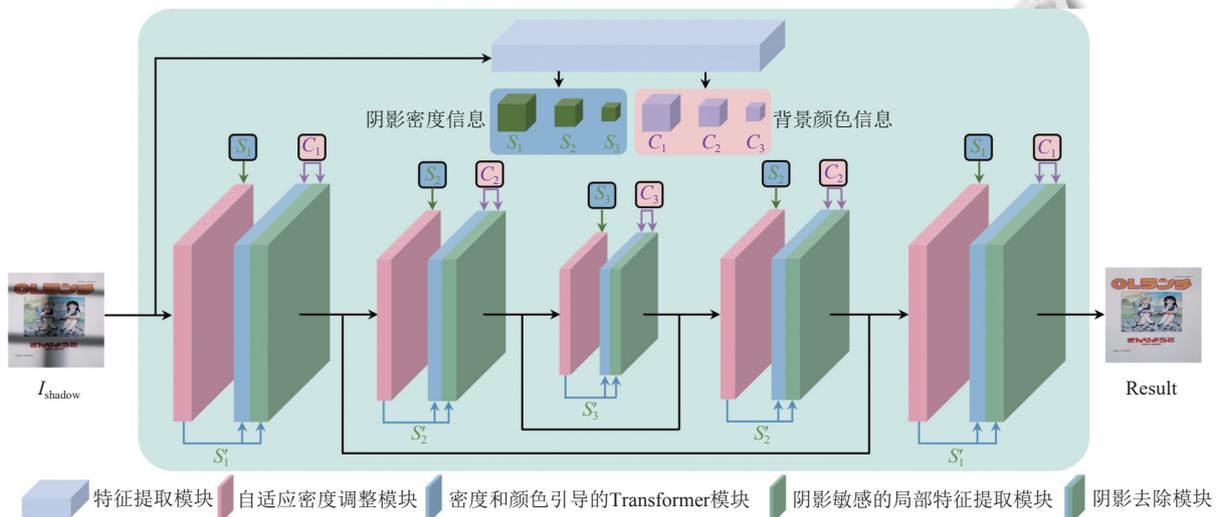


图 1 网络结构图

然而, 传统的卷积神经网络 (CNN) 在处理全局特征时存在局限性, 特别是在捕捉图像中远距离像素之间的依赖关系时. 而 Transformer 擅长捕捉全局特征, 能够有效处理图像中长距离的依赖关系, 这对于文本

阴影消除任务中的全局一致性非常重要. 但相比于 CNN 网络, Transformer 缺乏对局部特征提取的先验知识. 这意味着它不能像 CNN 那样天然地捕捉到图像中的局部特征和空间结构. 因此, 本文将 Transformer 模块和

CNN 模块相结合,使得网络可以捕捉图像中远距离像素之间的依赖关系的同时,也可以关注到图像当中的局部特征,从而进一步提升阴影消除效果。

2.2 特征提取模块

在文本阴影消除任务中,阴影区域的颜色补偿直接依赖于该区域的背景颜色,同时阴影密度的变化对图像的清晰度和色彩保真度产生显著影响。然而,现有的基于深度学习的文本阴影消除方法通常未能有效结合阴影密度与背景颜色,导致消除结果中频繁出现颜色失真和细节退化等伪影现象。为解决这一问题,本文构建了一个能够估计出阴影区域密度分布及其对应背景颜色信息的特征提取模块。通过这种方式,阴影消除网络能够基于阴影密度自适应调整策略,并结合背景颜色的先验信息进行颜色补偿,以实现像素级的精确处理。该机制确保了在不同阴影强度下,模型能够动态

优化消除效果,并有效恢复颜色一致性,从而提升整体阴影消除的准确性和鲁棒性。

CBENet^[23]已经证实了其在提取空间变化背景图像方面的能力,能够精确地描绘文本背景颜色。然而,它生成的背景图像是模糊的,导致难以准确区分不同区域的边界(如图2(a)所示)。为了解决这一问题,特征提取模块引入了 Sobel 边缘检测算法^[25]来捕捉图像的结构信息(如图2(b)所示)。考虑到文本图像的平面成像特性,不涉及深度信息,其纹理边缘信息较为简单且主要集中在水平和垂直方向,Sobel 算子能够有效捕捉这些特征。同时,Sobel 算子计算简单、效率高,且对噪声具有较好的抗性,适合高效、准确地处理文本图像阴影消除任务。另外,通过保持预训练好的 CBENet 模型参数不变(即冻结网络的参数),获取背景颜色信息,并将这些颜色信息与先前获得的结构信息拼接在一起得到 F_C 。

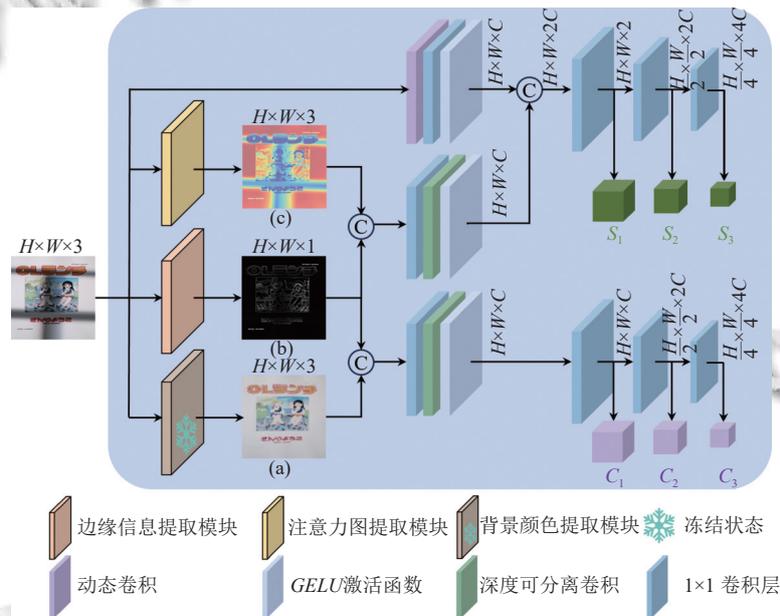


图2 特征提取模块

另外,受 Grad-CAM^[26]方法的启发,利用 CBENet 学习的空间分布信息,通过其最后一个卷积层的梯度,生成阴影注意力图(如图2(c)所示)。

由 Grad-CAM 方法生成的阴影注意力图能够有效捕捉阴影区域的密度分布,其中阴影越深的区域在注意力图中,对应的蓝色像素值越高,同时该阴影注意力图可以有效区分阴影图像中无阴影背景(红色)和其他(蓝色)像素的位置,通过这种方法,本文模型可以避免使用阴影遮罩,减少因阴影遮罩引入的错误信息,并

节省准备遮罩的精力。然而,该注意力图存在一定局限性,即文本颜色过深的内容在注意力图中也可能呈现出蓝色像素,被误认为阴影。因此为了更好地区分深色文本内容和阴影,特征提取模块也将 Sobel 算法提取出的纹理边缘信息和阴影注意力图拼接在一起得到 F_S 。然后,对 F_C 、 F_S 使用卷积核大小为 5、空洞率为 3 的深度可分离卷积进行特征提取。深度可分离卷积相较于普通卷积具有更大的感受野,在减少卷积神经网络中的计算量和参数数量的同时,保持较高的性能。接着,

使用 1×1 的卷积对其进行特征变换, 增强其特征的表达能力, 最后通过一个激活函数得到对应的特征 B_1 、 M_1 . 该过程计算公式如下:

$$\begin{cases} F_C = \text{Concat}(\text{Sobel}(I_{\text{shadow}}), \text{Color}(I_{\text{shadow}})) \\ F_S = \text{Concat}(\text{Sobel}(I_{\text{shadow}}), \text{Map}(I_{\text{shadow}})) \\ M_1, B_1 = \text{GELU}(\text{PWConv}(\text{DWConv}(F_C, F_S))) \end{cases} \quad (1)$$

其中, Sobel 为边缘检测函数, Color 为背景颜色估计函数, Map 为阴影注意力图计算函数. DWConv 为深度可分离卷积层, PWConv 为 1×1 卷积层, Concat 为拼接操作.

同时, 该模块使用动态可变形卷积对阴影图片进行特征提取, 同样通过一个 1×1 的卷积核和 GELU 激活函数得到特征 M_2 , 该过程公式如下:

$$M_2 = \text{GELU}(\text{PWConv}(\text{Deform}(I_{\text{shadow}}))) \quad (2)$$

其中, Deform 为动态可变形卷积.

最后, 将 M_1 和 M_2 拼接在一起输入到一个 1×1 卷积层, 对其进行融合和降维, 得到阴影密度特征 S_1 . 随后, 再依次使用两个 1×1 的卷积对其进行下采样, 得到

S_2 和 S_3 . 同时, B_1 也进行同样的操作, 得到背景颜色特征 C_1 、 C_2 、 C_3 .

在文本阴影消除任务中, 结合阴影密度信息和背景颜色信息指导图像阴影消除具有多方面的优势. 首先, 阴影密度信息可使网络精准识别阴影的分布和强度, 从而精确消除阴影而不误删图像中的其他细节. 其次, 背景颜色信息可为阴影区域提供准确的颜色补偿, 避免色彩失真, 保持图像的外观一致性和整体质量. 此外, 这些信息提供了图像的局部与全局特征, 网络通过融合这些特征, 能够更深入理解图像的结构与内容, 从而提高阴影消除效果.

2.3 自适应密度调整模块

图像在网络的不同阶段会经历不同的处理操作, 其特征表示会不断变化. 因此, 需要为其提供更加精准的阴影密度特征, 以指导后续模块的处理. 为此, 本文提出了自适应密度调整模块. 如图 3 所示. 该模块的设计目的是在网络的每个阶段动态调整阴影密度信息, 从而为后续密度和颜色引导的 Transformer 模块和阴影敏感的局部特征提取模块提供更准确的信息指导.

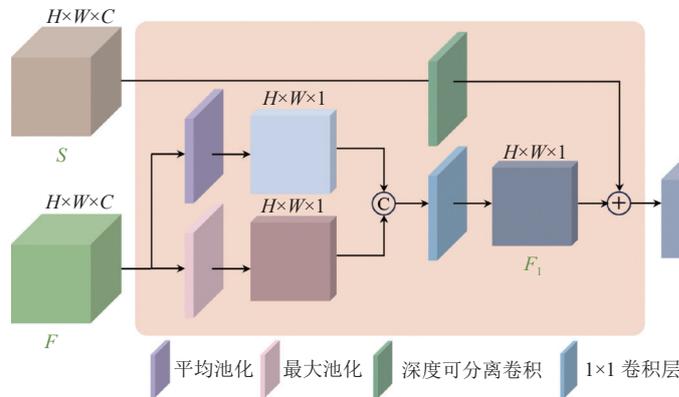


图3 自适应密度调整模块

具体来说, 输入特征 F 首先通过平均池化和最大池化操作, 得到对应的池化结果. 随后, 将池化后的结果拼接在一起, 并通过一个 1×1 卷积进行融合和调整, 得到变化系数 A . 利用变化系数 A 对阴影密度特征进行调整, 得到特征 F_1 . 该过程表示如下:

$$\begin{cases} A = \text{PWConv}(\text{Concat}(F_{\text{avg}}, F_{\text{max}})) \\ F_1 = A \odot S \end{cases} \quad (3)$$

其中, F_{avg} 表示平均池化结果, F_{max} 表示最大池化结果, \odot 表示元素相乘.

该操作可以帮助网络集中关注输入特征中的重要

区域, 并根据这些区域对初始阴影密度特征 S 进行自适应调整. 随后, 将输入的阴影密度特征 S 通过一个卷积核大小为 5、空洞率为 3 的深度可分离卷积层进行处理得到特征 F_2 . 最后, 将上述调整后的特征 F_1 和 F_2 相加, 得到最终调整后的阴影密度特征 S' . 该过程表示为:

$$S' = F_1 + \text{DWConv}(S) \quad (4)$$

通过自适应调整模块, 网络能够更有效地聚焦于图像中的阴影区域, 并在各阶段动态提取和精确校正阴影密度信息. 该动态调整机制显著提升了对不同密度阴影的消除精度, 有效抑制了过度亮化带来的图像

失真, 确保网络在各阶段均能获取准确的阴影密度信息, 从而在决策时实现更为精确的阴影消除效果。

2.4 密度和颜色引导的 Transformer 模块

在真实场景中, 图像中的阴影密度往往是不一致的. 这种不一致性给 Transformer 模块提取全局信息带来了很大的挑战, 特别是对于硬阴影. 不准确的全局信息可能导致色彩失真和对比度不平衡, 从而导致图像

外观不一致.

基于此, 本文提出了密度和颜色引导的 Transformer 模块, 如图 4 所示. 该模块主要由多头自注意力模块和前馈网络组成, 旨在解决阴影消除任务中全局特征提取问题. 具体来说, 该模块的作用是利用阴影密度信息来调整特征表示, 使得网络在处理图像时能够更好地关注重要区域并准确建立长距离依赖关系.

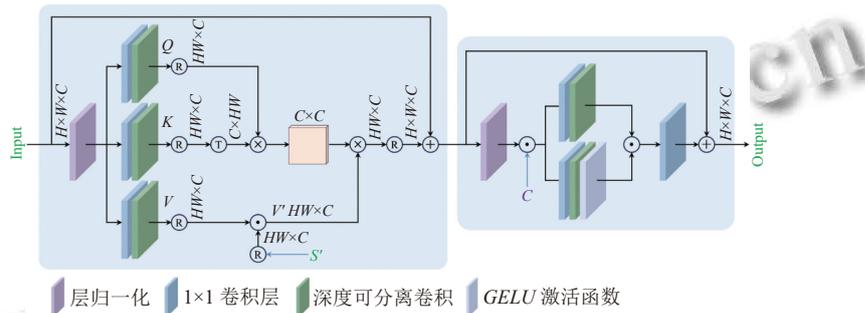


图 4 密度和颜色引导的 Transformer 模块

在多头自注意力模块当中, 首先对输入特征 F 经过层归一化操作, 然后分别通过 3 个 1×1 卷积和深度可分离卷积, 分别得到对应的 Q 、 K 和 V , 分别代表查询、键和值. 考虑到不同区域会有不同密度的阴影, 硬阴影区域通常会丢失更多的文本信息, 更加难以恢复. 而软阴影区域可以提供上下文信息, 帮助修复硬阴影区域. 而调整后的阴影密度特征表示每个阶段不同区域的阴影密度, 通过利用这些信息来调整自注意力机制中的权重分配. 使模型能够更加关注到硬阴影区域, 并从无阴影和软阴影区域学习有用的上下文信息. 因此利用调整后的阴影密度特征 S' 对特征 V 进行调整得到调整后的特征 V' . 该过程表示如下:

$$V' = V \odot \text{Reshape}(S') \quad (5)$$

其中, Reshape 表示调整形状.

然后, 将 Q 与转置后的 K 相乘, 得到全局注意力, 再将全局注意力与调整后的特征 V' 相乘, 并经过 Reshape 操作得到与输入特征 F 相同的形状, 最后与输入特征相加, 得到输出结果. 该过程表示如下:

$$F = (Q \times K^T) \odot V' + F \quad (6)$$

其中, \times 表示矩阵相乘, T 表示转置操作.

在前馈网络中, 将中间结果输入到前馈网络中, 首先进行层归一化, 然后背景颜色特征对其进行引导. 接着, 引导后的特征分别输入到对应的分支结构中, 得到

的结果相乘, 再经过一个 1×1 卷积层后与输入特征相加, 最终得到输出特征.

密度和颜色引导的 Transformer 模块通过自注意力机制和前馈网络, 结合阴影密度调整特征表示, 准确捕捉长距离依赖关系, 利用背景颜色信息更好地恢复全局一致性特征, 从而提高文本图像阴影消除的效果.

2.5 阴影敏感的局部特征提取模块

在真实场景中, 不均匀的阴影强度和复杂的背景颜色给 CNN 模块在提取局部信息时带来了显著挑战, 尤其是在不同颜色背景且阴影强度还差异很大的情况. 不准确的局部信息可能导致阴影消除结果中出现不真实的细节或伪影现象, 严重影响图像的视觉质量和真实性.

为了解决上述问题, 本文提出了阴影敏感的局部特征提取模块. 如图 5 所示. 该模块旨在解决阴影消除任务中局部特征提取的问题. 该模块利用背景颜色信息和阴影密度信息来指导图像阴影消除过程, 使得网络能够更精准地补偿图像中阴影部分的颜色, 从而提升阴影消除的效果和质量.

首先, 输入特征 F 经过批量归一化操作, 然后利用背景颜色特征 C 对其进行引导, 随后输入到一个 1×1 卷积层进行初步特征变换, 接着通过一个 5×5 卷积层和一个深度可分离卷积层进行特征提取得到特征 F' . 5×5 卷积层具有较大的感受野, 有助于捕捉更大范围内

的上下文信息. 然后, 将提取到的特征 \bar{F} 与归一化后的特征拼接在一起, 经过一个 1×1 卷积层、激活函数和另一个 1×1 卷积层进行特征融合和降维 (以下将其简化描述为 MLP) 得到输出结果. 激活函数能够引入非线性, 提升模型的表达能力. 这个过程的目的是将拼接后

的特征进一步融合, 减少冗余信息, 提高特征的紧凑性和有效性. 该过程表示如下:

$$\begin{cases} \bar{F} = DWConv(Conv(PWConv(BatchNorm(F) \odot C))) \\ F = MLP(Concat(\bar{F}, BatchNorm(F))) \end{cases} \quad (7)$$

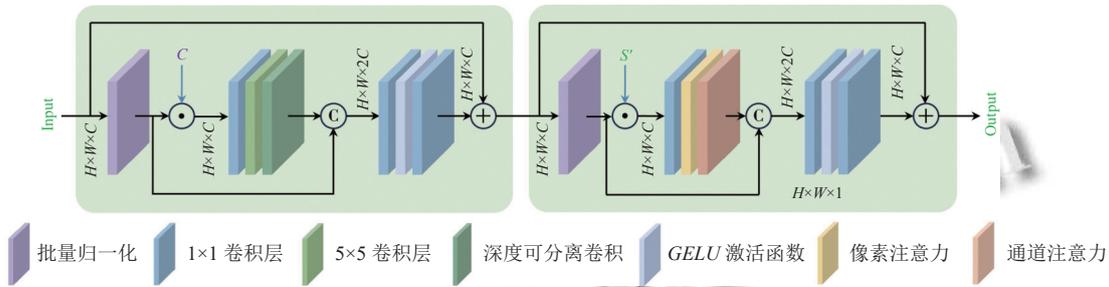


图5 阴影敏感的局部特征提取模块

随后该结果被输入到下一个模块, 再次进行层归一化操作, 以确保特征处理过程的一致性和稳定性. 利用调整后的阴影密度特征 S' 和其进行点积操作, 并输入到一个 1×1 卷积层进行初步特征调整. 接着, 使用像素注意力^[27]和通道注意力^[28]对调整后的特征进行提取得到中间特征 \bar{F} . 像素注意力机制通过计算每个像素的重要性权重, 使模型关注图像中的重要细节, 增强关键区域的清晰度, 从而更有效地消除阴影并恢复图像细节. 通道注意力机制通过调整特征通道的权重, 优化特征提取和信息融合的效率, 使网络更好地利用特征通道中的重要信息. 通过这两种注意力机制的协同作用, 模型能够更加灵活和精准地处理特征, 提高阴影消除效果. 最后将中间特征 \bar{F} 与归一化后的特征拼接后输入到一个 MLP 当中得到最终的输出特征. 该过程表示为:

$$\bar{F} = CA(PA(PWConv(BatchNorm(F) \odot S'))) \quad (8)$$

其中, CA 表示通道注意力, PA 表示像素注意力.

阴影敏感的局部特征提取模块通过结合卷积层、深度可分离卷积层、像素注意力和通道注意力机制, 能够更有效提取和融合特征, 从而显著提升文本图像阴影消除的质量.

2.6 损失函数

该网络的损失函数 L_{total} 由两项构成, 分别是外观一致性损失 $L_{appearance}$ 和结构一致性损失 $L_{structure}$. 其定义如下:

$$L_{total} = \alpha L_{appearance} + \beta L_{structure} \quad (9)$$

其中, α 和 β 为对应损失的权重值, 分别取值为 1 和 0.05.

- 外观一致性损失: 将生成图像和目标图像的每个像素间的差值的绝对值相加, 其定义如下:

$$L_{appearance} = \|I_{gt} - I_{shadow}\|_1 \quad (10)$$

其中, I_{shadow} 表示最终阴影消除的图像, I_{gt} 表示真实的无阴影图像.

- 结构一致性损失: 用来评估输出图像与真实图像的结构损失. 其计算公式为:

$$L_{structure} = \|VGG(I_{gt}) - VGG(I_{shadow})\|_2 \quad (11)$$

其中, $VGG(\cdot)$ 是预先训练好的 VGG19^[29] 模型的特征提取器.

综合这两项损失, 网络能够更有效地进行训练, 从而提升文本图像阴影消除的效果.

3 实验结果与分析

3.1 实验准备

- 实验设置: 基于 PyTorch 框架搭建和实现, 并在单块 NVIDIA GeForce RTX 3090 GPU 上进行训练. 实验过程中使用 AdamW 优化器^[30]对模型进行训练和反向优化. 训练图像为 256×256 , 并采用随机裁剪、随机翻转和随机旋转进行数据增强操作. 初始学习率设置为 0.0002, 并使用 CosineAnnealingLR 学习率调整策略对其进行调整, 批量大小设置为 8.

- 数据集: 为了公平比较, 所有方法均在 RDD^[23] 和 SD7K^[24] 数据集上进行训练和测试. RDD 数据集是为文本阴影消除任务设计的真实世界数据集, 包含了

纸张、书籍、宣传册等常见的阴影和无阴影文本。该数据集共收集了 4916 对图像, 其中 4371 对用于训练, 545 对用于测试。而 SD7K 数据集专注于高清文本阴影消除, 提供了更为丰富和挑战性的测试环境。该数据集由 7620 个样本组成, 其中 6860 对用于训练, 760 对作为测试集。SD7K 数据集的显著特点在于使用了超过 30 种不同的遮挡物类型以及 350 多份文本材料, 而且每个阴影图像中的阴影面积占总面积的比例大约为 41.38%, 标准差约为 13.58%。这样的设置增加了数据集的多样性, 能够有效模拟真实场景下的文本阴影消除任务。

● 评估指标: 为了评估论文中方法的性能, 本文采用了峰值信噪比 (PSNR)^[31]、结构相似性 (SSIM)^[32] 两个常用的图像质量评价指标。PSNR 用于衡量文本图像去除阴影后图像与真实无阴影图像间的差异, PSNR 值越高表示图像失真越小, 质量越高。SSIM 用于衡量图像的结构相似性, 更好地反映了人类视觉系统的感知, SSIM 值越接近 1 表示图像质量越高。

3.2 对比实验

表 1 展示了在相同硬件条件下, 本文方法在 RDD 和 SD7K 数据集上的表现均优于其他方法, 尤其在阴影条件复杂、文本图像色彩丰富的 SD7K 数据集上, 本文方法显著优于其他方法。表 1 中列出了 2 种自然图像阴影消除方法 (ST-CGAN^[8] 和 ShadowFormer^[9]) 和 5 种文本图像阴影消除方法 (文献[21]、BEDSR-Net^[22]、ShadocNet^[33]、FSENet^[24] 和 BGShadowNet^[23])。在文本颜色变化较少的 RDD 数据集上, ST-CGAN^[8] 和 ShadowFormer^[9] 表现尚可, 但在 SD7K 数据集上表现不佳。这主要是因为自然图像阴影消除方法在处理复杂背景颜色的文本图像时存在局限性。而在文本图像阴影消除方法中, 文献[21]倾向于冲淡颜色, 其泛化性较差。BEDSR-Net^[22]、ShadocNet^[33] 和 BGShadowNet^[23] 在 SD7K 数据集上的表现不佳, 这可能是因为它们对阴影密度的感知不敏感, 难以处理复杂的阴影情况。而 FSENet^[24] 在 RDD 数据集上的效果一般, 主要原因在于其设计更适合处理高分辨率图像, 如 SD7K 数据集。综上所述, 本文方法在不同特性的文本图像中均能稳定提供优异结果, 展现出比现有方法更强的稳健性。

图 6 提供了一些视觉阴影去除结果, 进一步展示了本文方法的优越性。可以看到, 文献[21]方法由于忽视了文本复杂的背景颜色, 导致了显著的颜色失真和

光照不均。ShadowFormer^[9] 和 ShadocNet^[33] 未能充分利用图像中的局部信息和背景颜色信息, 因此结果中出现了伪影和光照失真。基于恒定背景假设的 BEDSR-Net^[22] 在处理过程中未考虑阴影密度信息, 导致其在阴影边界处有时会产生伪影。FSENet^[24] 忽略了图像的全局信息和阴影密度, 导致阴影消除结果外观不一致。BGShadowNet^[23] 同样未考虑文本中的阴影密度信息, 导致光照失真问题。相比之下, 本文所提方法能够有效恢复阴影区域的光照, 避免伪影的产生, 结果与真实图像更加接近。

表 1 本文所提出的方法与其他方法在各个数据集上的对比

方法	RDD		SD7K	
	PSNR (dB)	SSIM	PSNR (dB)	SSIM
文献[21]	14.364	0.861	20.863	0.881
ST-CGAN ^[8]	34.328	0.974	23.407	0.892
ShadowFormer ^[9]	36.337	0.978	23.810	0.903
BEDSR-Net ^[22]	34.928	0.973	24.930	0.928
ShadocNet ^[33]	31.870	0.961	22.672	0.897
FSENet ^[24]	32.443	0.985	28.691	0.971
BGShadowNet ^[23]	37.585	0.983	29.143	0.979
本文方法	37.725	0.990	30.089	0.987

3.3 消融实验

为了验证本文所提出的网络中各个模块的有效性, 进行了一系列消融实验, 针对所提文本阴影消除模型设置了不同结构上的变化, 并对变化后的模型进行了实验。以探索各模块的贡献。具体包括以下 5 个变化后的模型。

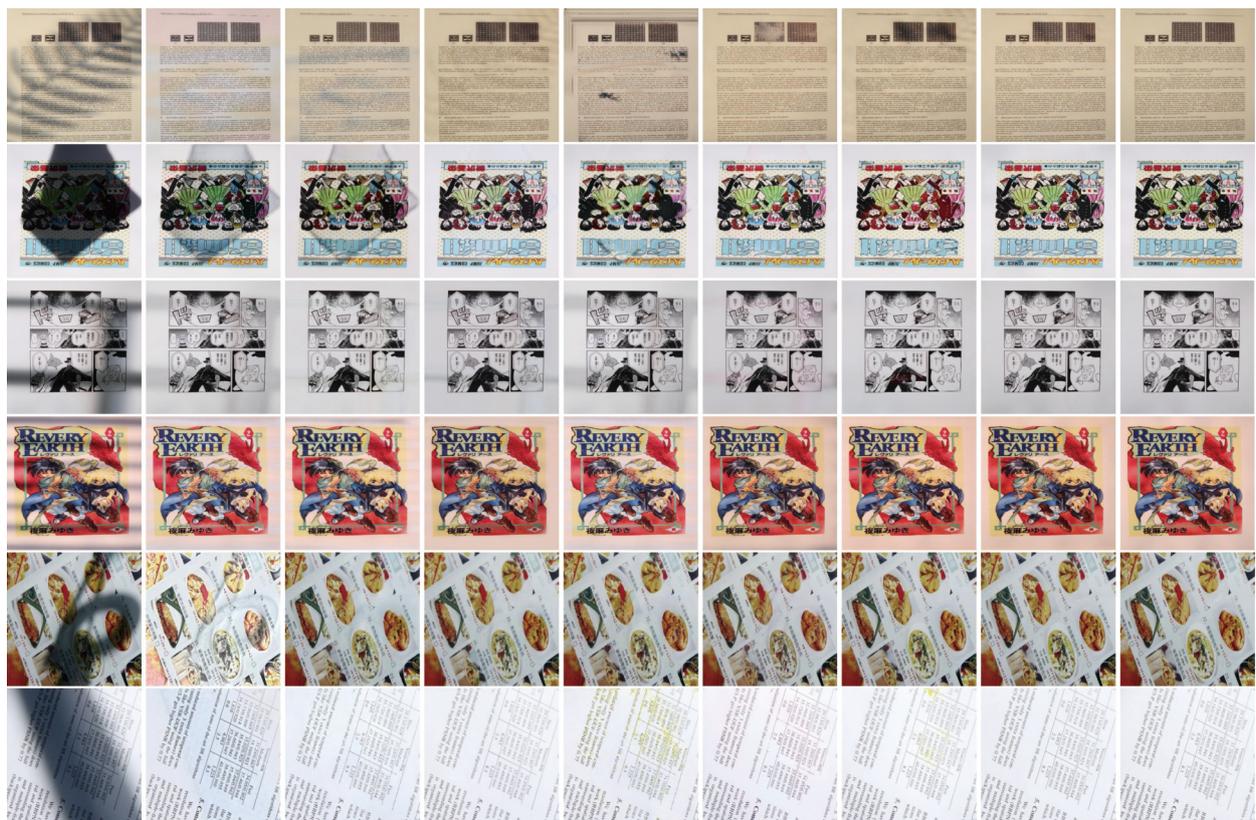
变体 1: 移除背景颜色提取, 只使用阴影密度提取指导 Transformer 和阴影敏感的局部特征提取模块, 评估背景颜色信息对文本阴影消除效果的影响。

变体 2: 移除阴影密度提取, 改用输入特征替代初始阴影密度特征, 并将其输入至自适应密度调整模块, 只使用背景颜色提取指导 Transformer 和阴影敏感的局部特征提取模块, 评估阴影密度信息对文本阴影消除效果的影响。

变体 3: 移除自适应密度调整模块对阴影密度信息进行调整, 仅使用初步提取到的阴影密度信息对 Transformer 和阴影敏感的局部特征提取模块进行指导。

变体 4: 移除阴影敏感的局部特征提取模块, 仅使用密度和颜色引导的 Transformer 模块。

变体 5: 移除密度和颜色引导的 Transformer 模块, 仅使用阴影敏感的局部特征提取模块。



(a) 阴影图片 (b) 文献[21] (c) ShadowFormer (d) BEDSR-Net (e) ShadocNet (f) FSENet (g) BGShadowNet (h) 本文方法 (i) 真实无阴影图像

图6 在RDD和SD7K数据集上的视觉比较图

为保证实验公平性, 每组消融实验均在SD7K数据集上进行, 并在对应测试集上完成了评估. 结果如表2所示, 证明各个模块均有效提升了网络性能.

表2 消融实验结果

方法	PSNR (dB)	SSIM
变体1	29.187	0.984
变体2	29.013	0.981
变体3	28.815	0.975
变体4	28.872	0.979
变体5	28.541	0.973
本文方法	30.089	0.987

移除背景颜色信息(变体1)后, PSNR数值下降0.902 dB, 表明背景颜色信息对于文本阴影消除效果至关重要; 去除阴影密度提取模块(变体2)导致PSNR下降1.076 dB, 说明阴影密度信息有助于防止图像过度补光而引起失真; 不使用自适应密度调整模块(变体3)时, PSNR数值下降1.274 dB, 证明在不同阶段提供精准阴影密度信息是必要的; 去除阴影敏感局部特征提取模块(变体4)导致PSNR下降1.217 dB, 体现出局部特征提取对提升效果的重要性; 移除密度与颜色

引导的Transformer模块(变体5)时, PSNR下降1.548 dB, 表明全局信息对于保持图像整体一致性具有关键作用. 整体来看, 本文方法中的各模块均对文本图像阴影消除任务做出了有效贡献.

图7展示了各变体的可视化结果: 变体1缺少背景颜色信息引导, 硬阴影区域出现颜色失真或过度补光; 变体2因缺少阴影密度信息, 轻度阴影区域边缘出现伪影; 变体3移除自适应密度调整模块, 导致过度亮化引发图像失真; 变体4缺少局部特征提取模块, 结果细节表现不真实; 变体5去除Transformer模块引导, 导致图像整体外观不一致. 相比之下, 本文方法综合各模块后, 生成了更加自然、真实的阴影消除结果.

4 总结

本文提出了一种基于密度和颜色信息指导的文本图像阴影消除网络, 通过阴影密度和背景颜色信息的引导, 显著提高了文本图像阴影消除的精度和效果. 该网络首先通过特征提取模块提取阴影区域的密度分布及其对应的背景颜色信息. 接着, 使用自适应密度调整模块对提取的阴影密度进行调节, 确保阴影区域的密

度信息更加准确和可靠. 在此基础上, 本文提出了密度和颜色引导的 Transformer 模块和阴影敏感的局部特征提取模块, 并将它们结合, 作为模型的基本组件. 通过密度和颜色信息的引导, Transformer 模块能够有效处理长距离依赖关系, 确保阴影消除后的图像在全局上保持一致性; 而阴影敏感的局部特征提取模块则能够提取局部特征, 确保阴影区域的局部细节得到准确处理, 避免颜色失真和细节退化等伪影现象. 实验证明

本文所提出的方法在文本图像阴影消除方面的可行性和有效性. 然而本文方法存在一定局限性, 当面对具有复杂光照条件或极端阴影情况的文本图像时, 阴影消除效果可能欠佳, 出现阴影残留或过度消除现象. 同时, 模型在计算资源消耗上相对较高. 针对这些局限, 未来将研究引入更多场景先验知识, 提升模型对复杂情况的适应性. 同时探索不同方法来进一步减少所需参数, 使其能够在智能设备上使用.



图7 消融实验的可视化结果

参考文献

- 1 Bako S, Darabi S, Shechtman E, *et al.* Removing shadows from images of documents. Proceedings of the 13th Asian Conference on Computer Vision. Taipei: Springer International Publishing, 2017. 173–183.
- 2 Gatos B, Pratikakis I, Perantonis SJ. Adaptive degraded document image binarization. Pattern Recognition, 2006, 39(3): 317–327.
- 3 皇甫俊逸, 孟乔, 孟令辰, 等. 基于 GhostNet 与注意力机制的 YOLOv5 交通目标检测. 计算机系统应用, 2023, 32(4): 149–160. [doi: 10.15888/j.cnki.csa.009048]
- 4 张凯悦, 张鸿. 基于注意力机制网络的航运监控图像识别模型. 计算机应用, 2021, 41(10): 3010–3016. [doi: 10.11772/j.issn.1001-9081.2020121899]
- 5 杨康叶, 鄂春学. 基于 RGB 模型颜色相似性的彩色图像分割. 计算机系统应用, 2013, 22(3): 128–131, 160. [doi: 10.3969/j.issn.1003-3254.2013.03.030]
- 6 张华南, 冯全, 杨梅, 等. 基于 YCbCr 颜色空间的叶片阴影检测与去除. 计算机系统应用, 2015, 24(11): 262–265. [doi: 10.3969/j.issn.1003-3254.2015.11.045]
- 7 郝灿, 朱信忠, 赵建民, 等. 基于改进型 LBP 特征的运动阴影去除算法. 计算机系统应用, 2010, 19(5): 80–83. [doi: 10.3969/j.issn.1003-3254.2010.05.019]
- 8 Wang JF, Li X, Yang J. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City: IEEE, 2018. 1788–1797.
- 9 Guo LQ, Huang SY, Liu D, *et al.* ShadowFormer: Global

- context helps image shadow removal. arXiv:2302.01650, 2023.
- 10 Arbel E, Hel-Or H. Shadow removal using intensity surfaces and texture anchor points. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2011, 33(6): 1202–1216. [doi: [10.1109/TPAMI.2010.157](https://doi.org/10.1109/TPAMI.2010.157)]
 - 11 Finlayson GD, Hordley SD, Lu C, *et al.* On the removal of shadows from images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006, 28(1): 59–68. [doi: [10.1109/TPAMI.2006.18](https://doi.org/10.1109/TPAMI.2006.18)]
 - 12 Shor Y, Lischinski D. The shadow meets the mask: Pyramid-based shadow removal. *Computer Graphics Forum*, 2008, 27(2): 577–586.
 - 13 Xiao CX, She RY, Xiao DL, *et al.* Fast shadow removal using adaptive multi-scale illumination transfer. *Computer Graphics Forum*, 2013, 32(8): 207–218.
 - 14 Xiao CX, Xiao DL, Zhang L, *et al.* Efficient shadow removal using subregion matching illumination transfer. *Computer Graphics Forum*, 2013, 32(7): 421–430. [doi: [10.1111/cgf.12250](https://doi.org/10.1111/cgf.12250)]
 - 15 Qu LQ, Tian JD, He SF, *et al.* DeshadowNet: A multi-context embedding deep network for shadow removal. *Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition*. Honolulu: IEEE, 2017. 2308–2316.
 - 16 Le H, Samaras D. Shadow removal via shadow image decomposition. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 8577–8586.
 - 17 Zhang L, Long CJ, Zhang XL, *et al.* RIS-GAN: Explore residual and illumination with generative adversarial networks for shadow removal. *Proceedings of the 2020 AAAI Conference on Artificial Intelligence*, 2020, 34(7): 12829–12836. [doi: [10.1609/aaai.v34i07.6979](https://doi.org/10.1609/aaai.v34i07.6979)]
 - 18 Ding B, Long CJ, Zhang L, *et al.* ARGAN: Attentive recurrent generative adversarial network for shadow detection and removal. *Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision*. Seoul: IEEE, 2019. 10212–10221.
 - 19 Liu ZH, Yin H, Wu XY, *et al.* From shadow generation to shadow removal. *Proceedings of the 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Nashville: IEEE, 2021. 4925–4934.
 - 20 Oliveira DM, Lins RD, de França Pereira e Silva G. Shading removal of illustrated documents. *Proceedings of the 10th International Conference on Image Analysis and Recognition*. Aveiro: Springer, 2013. 308–317.
 - 21 Jung S, Hasan MA, Kim C. Water-filling: An efficient algorithm for digitized document shadow removal. *Proceedings of the 14th Asian Conference on Computer Vision*. Perth: Springer, 2019. 398–414.
 - 22 Lin YH, Chen WC, Chuang YY. BEDSR-Net: A deep shadow removal network from a single document image. *Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Seattle: IEEE, 2020. 12902–12911.
 - 23 Zhang L, He YH, Zhang Q, *et al.* Document image shadow removal guided by color-aware background. *Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. Vancouver: IEEE, 2023. 1818–1827.
 - 24 Li ZN, Chen XH, Pun CM, *et al.* High-resolution document shadow removal via a large-scale real-world dataset and a frequency-aware shadow erasing net. *Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV)*. Paris: IEEE, 2023. 12415–12424.
 - 25 郑欢欢, 冯治东, 李瑞华. 基于 Sobel 算子的道路图像边缘检测算法. *榆林学院学报*, 2023, 33(2): 60–63. [doi: [10.16752/j.cnki.jylu.2023.02.014](https://doi.org/10.16752/j.cnki.jylu.2023.02.014)]
 - 26 Selvaraju RR, Cogswell M, Das A, *et al.* Grad-CAM: Visual explanations from deep networks via gradient-based localization. *Proceedings of the 2017 IEEE International Conference on Computer Vision*. Venice: IEEE, 2017. 618–626.
 - 27 Thakur RK, Maji SK. Multi scale pixel attention and feature extraction based neural network for image denoising. *Pattern Recognition*, 2023, 141: 109603. [doi: [10.1016/j.patcog.2023.109603](https://doi.org/10.1016/j.patcog.2023.109603)]
 - 28 Yang WQ, Yuan Y, Zhang DH, *et al.* An effective image classification method for plant diseases with improved channel attention mechanism aECANet based on deep learning. *Symmetry*, 2024, 16(4): 451. [doi: [10.3390/sym16040451](https://doi.org/10.3390/sym16040451)]
 - 29 Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv:1409.1556, 2014.
 - 30 Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv:1412.6980, 2014.
 - 31 Huynh-Thu Q, Ghanbari M. Scope of validity of PSNR in image/video quality assessment. *Electronics Letters*, 2008, 44(13): 800–801. [doi: [10.1049/el:20080522](https://doi.org/10.1049/el:20080522)]
 - 32 Wang Z, Bovik AC, Sheikh HR, *et al.* Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 2004, 13(4): 600–612. [doi: [10.1109/TIP.2003.819861](https://doi.org/10.1109/TIP.2003.819861)]
 - 33 Chen XH, Cun XD, Pun CM, *et al.* ShadocNet: Learning spatial-aware tokens in Transformer for document shadow removal. *Proceedings of the 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. Rhodes Island: IEEE, 2023. 1–5.

(校对责编:王欣欣)