计算机科学
COMPUTER SCIENCE

基于上下文门控残差和多尺度注意力的图像重照明网络

王威, 杜响成, 金城

引用本文

王威, 杜响成, 金城. 基于上下文门控残差和多尺度注意力的图像重照明网络[J]. 计算机科学, 2023, 50(9): 168-175.

WANG Wei, DU Xiangcheng, JIN Cheng. Image Relighting Network Based on Context-gated Residuals and Multi-scale Attention [J]. Computer Science, 2023, 50(9): 168-175.

计算机科学 COMPUTER SCIENCE
http://www.jsjkx.com
DOI:10.11896/jsjkx.221000100

基于上下文门控残差和多尺度注意力的图像重照明网络

王 威 杜响成 金 城
复旦大学计算机科学技术学院 上海 200438
(20212010100@fudan.edu.cn)

摘 要 图像重照明普遍应用于图像编辑和数据增强等任务。现有图像重照明方法去除和重建复杂场景下的阴影时,存在阴影形状估计不准确、物体纹理模糊和结构变形等缺陷。针对以上问题,提出了基于上下文门控残差和多尺度注意力的图像重照明网络。上下文门控残差通过聚合局部和全局的空间上下文信息获取像素的长程依赖,保持阴影方向和照明方向的一致性。此外,利用门控机制有效提高网络对纹理和结构的恢复能力。多尺度注意力通过迭代提取和聚合不同尺度的特征,在不损失分辨率的基础上增大感受野,它通过串联通道注意力和空间注意力激活图像中重要的特征,并抑制无关特征的响应。文中还提出了照明梯度损失,它通过有效学习各方向照明梯度,获得了视觉感知效果更好的图像。实验结果表明,与现有的最优方法相比,所提方法在 PSNR 指标和 SSIM 指标上分别提升了 7.47%和 12.37%。

关键词:图像重照明;上下文信息;门控机制;照明梯度;注意力
中图法分类号 TP391

Image Relighting Network Based on Context-gated Residuals and Multi-scale Attention

WANG Wei,DU Xiangcheng and JIN Cheng
School of Computer Science,Fudan University,Shanghai 200438,China

Abstract Image relighting is commonly used in image editing and data augmentation tasks. Existing image relighting methods suffer from estimating accurate shadows and obtaining consistent structures and clear texture when removing and rendering shadows in complex scenes. To address these issues,this paper proposes an image relighting network based on context-gated residuals and multiscale attention. Contextual gating residuals capture the long-range dependencies of pixels by aggregating local and global spatial context information,which maintains the consistency of shadow and lighting direction. Besides,gating mechanisms can effectively improve the network's ability to recover textures and structures. Multiscale attention increases the receptive field without losing resolution by iteratively extracting and aggregating features of different scales. It activates important features by concatenating channel attention and spatial attention,and suppresses the responses of irrelevant features. In this paper,lighting gradient loss is also proposed to obtain satisfactory visual images through efficiently learning the lighting gradients in all directions. Experimental results show that,compared with the current state-of-the-art methods,the proposed method improves PSNR and SSIM by 7.47% and 12.37%,respectively.

Keywords Image relighting,Contextual information,Gating mechanism,Lighting gradient,Attention

1 引言

随着显示设备的发展,人们对图像质量的要求越来越高,合适的照明条件是其必要条件。例如,人们在进行网上教学、网上购物和网上会议时,需要根据需求和场景特点为其赋予合适的照明条件。不合适的照明条件,会使得图像产生异常的纹理、颜色以及冗余的光影。图像重照明以给定照明条件(光源位置和色温)为目标,改变输入图像的照明条件。

目前,图像重照明方法主要包括传统方法和基于深度学习的方法。基于物理的传统图像重照明方法泛化能力弱,局限于肖像、人身、建筑等特定对象。而基于深度学习的图像重照明方法是通用的,只需要训练一个模型,便可以泛化到其他场景进行重照明。然而,大多数基于深度学习的方法需要图像深度信息和多照明方向的视图以供训练。同时,现有方法存在未充分考虑局部和全局空间上下文信息导致阴影形状估计错误,特征表达能力不足导致生成图像中物体结构扭曲、边缘模糊等问题。

根据上述问题,本文提出了融合上下文门控残差和多尺度注意力的网络。具体来说,受到 Retinex 理论的启发,本文将图像重照明分解为场景结构恢复、照明估计和场景重渲染

到稿日期:2022-10-13 返修日期:2023-04-06
基金项目:国家重点研发计划(2019YFB2102800)
This work was supported by the National Key R&D Program of China (2019YFB2102800).
通信作者:金城(jc@fudan.edu.cn)

3 个子任务。在场景结构恢复和照明估计子网中，将目标图像的照明条件（色温、照明方向、照明）转移到输入图像中，而不是在输入图像上直接学习目标照明条件映射。全局信息和局部信息对照明转换操作都非常重要：全局信息有利于增强照明效果和阴影的一致性，局部信息则可以强化细节表现。本文设计了上下文门控残差模块，用于聚合局部和全局空间上下文信息，捕获像素的长程依赖，并利用门控机制抑制信息较少的特征，以产生更精细的图像。

在场景重渲染阶段中，本文提出多尺度注意力模块，通过多个平滑膨胀卷积组级联聚合不同尺度的全局信息，然后使用双重注意力选择利于重照明的关键特征，避免无关信息的干扰。双重注意力由坐标通道注意力[1]和照明空间注意力组成。坐标通道注意力[1]可以捕获跨通道信息，将通道注意力分解为两个并行的特征编码来高效整合空间坐标信息到生成的特征图中，并重新校准通道的权重。照明空间注意力则在空间维度上进行最大池化和平均池化，引导特征在空间维度上具有全局视角。

除此之外，在场景结构恢复和照明估计子网进行阴影去除和重建的过程中，本文结合了自校准卷积和自正则化注意力实现对特征信息的有效标定，强化网络对阴影区域的定位能力。为了避免产生伪影，使用双线性插值加卷积的方法替换反卷积进行上采样，并将多尺度感知模块中的膨胀卷积替换为平滑膨胀卷积。此外，为了生成视觉感知质量更好的图像，设计了照明梯度损失，目的在于通过模糊纹理，使得网络更加关注于各方向上的照明梯度。本文贡献分为以下 4 点：

1）提出了上下文门控残差模块，通过深度卷积和通道注意力机制编码局部和全局空间上下文信息，提升图像的整体协调性。同时，利用门控机制有效提高了网络对纹理和结构的恢复能力。

2）提出了多尺度注意力模块，利用平滑膨胀卷积，在不损失分辨率的基础上增大感受野，并通过通道和空间注意力串联聚焦图像中重要的特征，抑制无关区域的响应。

3）提出了自校准采样模块，通过添加自校准卷积至上采样和下采样模块，实现特征信息的迭代标定，提高了特征的表示能力。

4）提出了照明梯度损失，使得网络能够更加关注各方向上的照明梯度，生成视觉感知质量更好的图像。

2　相关工作

目前的图像重照明主要分为基于物理的图像重照明和基于深度学习的图像重照明两种方法。

2.1　基于物理的图像重照明

基于物理的传统图像重照明方法主要是通过先预测图像的几何形状、材料反射属性和照明条件，然后进行重渲染得到重照明图像的。Zhang 等[2]设计了一个可以同时估计场景中物体的材料反射特性和场景的照明条件的逆渲染框架，从而对空房间中新摆放的物体进行重照明。Duchêne 等[3]将一组相同照明下的多视角户外图像和照明方向作为输入，进而将图像分解为反射层和照明层进行图像重照明。Karsch 等[4]利用几何来改善亮度估计，以进行更好的图像分解重照明。Peers 等[5]和 Reddy 等[6]则通过计算目标场景的照明传输函数来更改输入图像的照明条件。虽然以上方法可以通过物理建模，对输入图像进行重渲染而生成高质量的重照明图像，但是它们大部分是为特定场景而设计的，缺乏泛化性，并且需要复杂的设备或系统来采集或估计明确的照明参数，这些桎梏限制了传统照明的发展。本文的方法是以数据作为导向，模型经数据集训练，即可对单张图像进行重照明。

2.2　基于深度学习的图像重照明

近年来，基于深度学习的图像重照明方法有了显著的进步。这些方法的目的是学习两个照明条件之间的映射，将图像重照明近似看作图像之间的转换问题。UNet[7]网络因能够结合底层信息和高层信息的优势，被广泛应用在低级图像视觉任务中。Wang 等[8]和 Hu 等[9]都针对图像重照明任务包含多个子任务的特点进行改进。Wang 等在 Pix2Pix[10]框架的基础上提出了 DRN[8]。此后，Wang 等[11]在 DRN[8]的基础上做了改进，提出了 MCN[11]。MCN 将下采样特征自标定块和上采样特征自标定块作为编码器和解码器的基本块，强化对图像照明的重标定能力。Hu 等[9]基于 UNet[7]网络，通过两个辅助网络估计引导图像的照明，并向解码器提供照明特征。

Kubiak 等[12]和 Yazdani 等[13]都是基于 Retinex 理论进行图像重照明。前者提出了一种将输入图像分解为内容域和风格域的自监督照明传输方法；后者使用图像内在分解重照明和直接重照明两种策略生成目标图像，再通过权重学习的方式将两者融合。文献[14-15]则考虑了图像的深度信息，设计以 RGB-D 图像作为输入和输出的深度引导重照明网络。Yang 等[14]利用双分叉主干同时提取图像的深度信息和特征，并引入动态膨胀金字塔来融合深度特征。相关方法大多需要图像深度信息或多视图数据集，而本文的方法关注图像本身，减少了对输入的约束，通过上下文门控残差聚合局部和全局空间上下文信息，增强了阴影和照明的空间一致性。同时，本文通过多尺度注意力提升网络对图像纹理和结构的重建能力，并在无需图像深度信息的基础上进一步提升了重照明的效果。

3　基于上下文门控残差和多尺度注意力的图像重照明网络

3.1　网络结构

如图 1 所示，重照明网络由场景结构恢复、照明估计和场景重渲染 3 个部分组成。本文定义 X_α 是在预先设置的照明条件（照明方向、色温、阴影效果）α 下的输入图像，X_β 则是在目标照明条件 β 下的图像。受 Retinex 理论的启发，学习照明的过程分为场景结构恢复操作 $R_\alpha^{-1}(\cdot)$ 和照明估计操作 $R_{\alpha\to\beta}(\cdot)$。前者提取和照明无关的主要场景结构 S，后者将目标照明条件 β 转移到输入图像 X_α 中。

$$S = R_\alpha^{-1}(X_\alpha) \tag{1}$$

$$L = R_{\alpha\to\beta}(X_\alpha) \tag{2}$$

最后，场景重渲染子网得益于上述两个操作提取出的主要场景结构和所估计的阴影，感知全局照明效果。

$$X_\beta = P(S, L) \tag{3}$$

其中，P 代表的是联合渲染的过程。

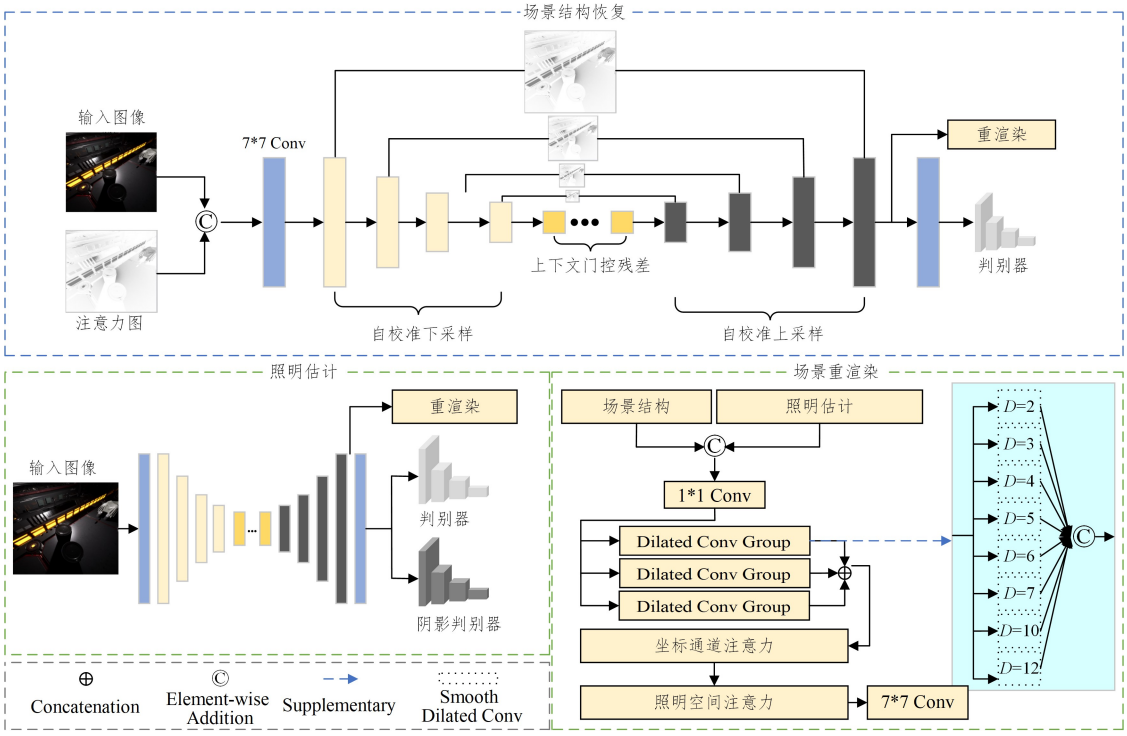

图 1 重照明网络结构图

Fig. 1 Architecture of our image relighting network

3.2 场景结构恢复子网

场景结构恢复子网的主要目标是提取和照明无关的场景结构信息。该子网生成器采用特征自校准下采样编码器-上下文门控残差-特征自校准解码器的结构，判别器采用与 Pix2Pix[10] 相同的四层跨步卷积来递进获取全局表示。给定输入图像 $X_a \in \mathbb{R}^{3\times H\times W}$，由大小为 7×7 的卷积层提取浅层特征表示 $F' \in \mathbb{R}^{C\times H\times W}$，$C$ 表示通道数量，$H\times W$ 代表空间维度。接着，将其通过 4 次特征自校准下采样层提取场景的全局判别特征 $F^{\downarrow 4} \in \mathbb{R}^{16*C\times \frac{H}{16}\times \frac{W}{16}}$。在编码器提取到图像的全局特征 $F^{\downarrow 4}$ 之后，将其输入上下文门控残差块，聚合局部和全局空间上下文信息以丰富特征表示，以更好地进行场景细节结构的恢复。如图 2 所示，特征自校准上采样和下采样采用类似的结构。此外，受到 Enlightengan[16] 的启发，本文归一化 RGB 图像的照明通道 I 至[0,1]，然后通过 $1-\tilde{I}$ 构建自正则化注意力图，并将其调整至合适的大小与各层输入特征图拼接。其中，$\tilde{I}$ 代表的是照明通道 I 的逐元素差值。

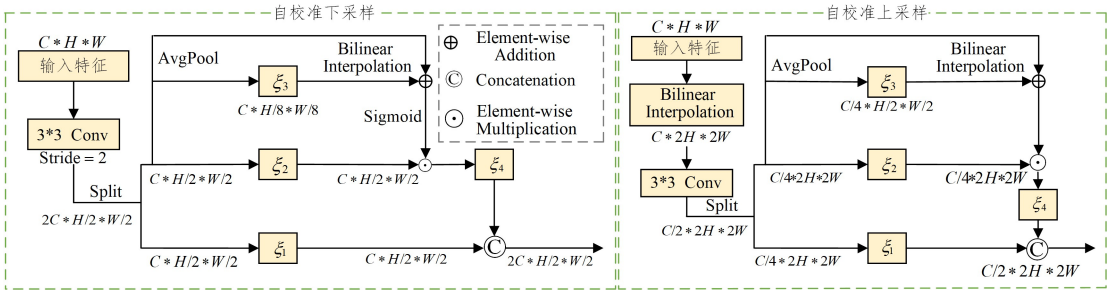

图 2 自校准上采样和下采样模块

Fig. 2 Self-calibrating up-sampling and down-sampling modules

3.2.1 特征自校准采样

引入特征自校准卷积[17] 能够自适应编码每个空间位置周围的上下文信息，增大感受野，利于准确定位阴影区域，从而更好地实施去除和重建阴影的操作。本文设置 $\xi_i(i=1,2,3,4)$ 表示第 i 个大小为 3×3 的卷积操作。第一条路径仅使用 ξ_1 保留特征的原始空间；第二条路径通过 ξ_2,ξ_3,ξ_4 在下采样后的小尺度空间和保持分辨率的原始空间进行卷积特征提取。

$$\widetilde{F_1''}=\xi_1(F''_1) \tag{4}$$

$$W=\sigma(Up(\xi_3(Down(F_2''))+F_2'') \tag{5}$$

$$\widetilde{F_2''}=\xi_4(W*\xi_2(F_2'')) \tag{6}$$

其中，F' 为中间特征，$F_j''(j=1,2)$ 为下采样后通过 1×1 卷积按通道划分成的两个相同的部分，σ 是 Sigmoid 激活函数，Up 和 $Down$ 分别代表双线性插值和平均池化，W 为小尺度空间校准原始空间特征的权重。最后，将校准过的特征和保留原始空间的特征按通道维度相加，生成最终的校准特征 $F''\in\mathbb{R}^{C\times H\times W}$。该过程的表达式如下：

$$F''=concat(\widetilde{F_1''},\widetilde{F_2''}) \tag{7}$$

3.2.2　上下文门控残差

如图 3 所示，上下文门控残差模块使用深度卷积获取来自于空间相邻像素位置的上下文信息以丰富特征。本文定义中间特征 $F_i^{\downarrow 4}\in\mathbb{R}^{16*C\times\frac{H}{16}\times\frac{W}{16}}(i\in[1,9])$，$i$ 表示位于第 i 个残差块。首先，$F_i^{\downarrow 4}$ 通过 1×1 卷积 $C_p(\cdot)$ 在像素层面上进行跨通道上下文聚合；然后，通过 3×3 深度卷积 $C_d(\cdot)$ 在通道层面上进行局部空间上的上下文聚合，得到特征 $F_l\in\mathbb{R}^{32*C\times\frac{H}{16}\times\frac{W}{16}}$；接着，使用门控机制抑制信息较少的特征，促进有效信息向前流动，使得后续层更加关注信息丰富的区域，生成更加清晰的场景结构。整个过程的表达式如下：

$$F_l=C_d(C_p(LN(F_i^{\downarrow 4}))) \tag{8}$$

$$F_l^1,F_l^2=Split(F_l) \tag{9}$$

$$F_c=CA(\Phi(F_l^1)\odot F_l^2)+F_i^{\downarrow 4} \tag{10}$$

其中，LN 代表层归一化；Φ 代表 GELU[18] 非线性激活；$\odot$ 表示对应元素逐个相乘；$Split$ 为分块操作；CA 为 SE[19] 通道注意力，被广泛应用于低级视觉任务。通道注意力提供了两个关键优势：1）聚合全局上下文信息；2）对通道进行权值校准。全局上下文信息有助于准确定位阴影区域，而局部上下文信息有助于更好地进行阴影去除和重建工作。为了进一步强化有用信息的传递，使用两个 1×1 卷积 C_p^1 和 C_p^2 进行线性变换，并在中间插入上述门控机制。

$$F'_c=C_p^2(C_p^1LN(\Phi(F_c^1)\odot F_c^2))+F_c \tag{11}$$

然后，将上下文门控残差与标准残差通过通道压缩模块级联。

$$F_{\text{out}}=Res(Comp[F_c',F_i^{\downarrow 4}]) \tag{12}$$

其中，$Comp$ 代表的是通道压缩的过程。通道压缩通过并行卷积分支进行，一个分支由 1×1 卷积构成，另一个分支由两个 3×3 卷积串联而成，Res 为标准残差块。

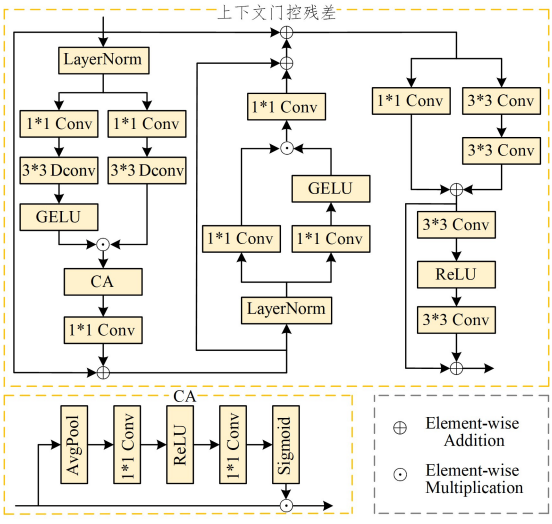

图 3　上下文门控残差模块

Fig. 3　Context-gated residual modules

3.3　照明估计子网

照明估计子网的目标是在输入图像上施加目标照明条件。首先，为了更好地感知全局照明变化，去除带有自正则化注意力的跳跃连接。然后，在使用多尺度判别器的基础上添加结构类似的阴影判别器。阴影判别器首先通过校准值 C 来对低照明区域（黑暗区域、阴影区域）进行校正。校准值表达式如下：

$$C=\min(\alpha,\beta) \tag{13}$$

其中，α 表示所估计的像素亮度，β 为一个超参数，预先定义阴影敏感度阈值，参考先前工作，将其设置为 0.059＝15/255。

3.4　场景重渲染

输入图像在通过场景结构恢复子网和照明估计子网后，得到所估计的场景结构和照明效果。场景重渲染子网的目标是将两者融合，生成具有目标照明条件的图像。首先，将场景结构特征图和照明估计特征图按通道维度进行拼接，并通过 1×1 卷积生成具有固定深度的混合特征图（深度为 8）。膨胀卷积可以在不损失分辨率的情况下扩大感受野并捕获多尺度信息。具体来说，通过 8 个膨胀卷积层并行构成膨胀卷积组来提取多层次信息。然后，通过级联 3 个膨胀卷积组获取更好的信息提取能力。

多尺度感知模块将不同空间感知的特征合并为单个特征图，可以将每个特征通道图看作特定尺度的响应。本文使用了双重注意力，通过坐标通道注意力[1]对不同通道的照明信息进行权值校准，挑选利于重照明的关键特征；同时，通过照明空间注意力提升关键照明区域的特征表达。

3.5　损失函数设计

本文为场景恢复、照明估计和场景重渲染 3 个子网分别设计了相应的损失函数。

3.5.1　场景恢复子网损失函数

场景结构恢复子网的判别器遵循与 Pix2Pix[10] 相同的结构，使用 4 层跨步卷积逐层提取全局表示。对抗损失 $\mathcal{L}_{\text{cGAN}}$ 如下所示：

$$\mathcal{L}_{\text{cGAN}}(G,D)=E_{(\boldsymbol{X},\boldsymbol{Y}_{\text{str}})}[\log D(\boldsymbol{X},\boldsymbol{Y}_{\text{str}})]+E_{\boldsymbol{X}}[\log(1-D(\boldsymbol{X},G(\boldsymbol{X})))] \tag{14}$$

其中，G 和 D 分别代表生成器和判别器，$\boldsymbol{X}$ 和 $\boldsymbol{Y}_{\text{str}}$ 分别代表输入图像和目标无阴影图像。

$$\mathcal{L}_C(G)=E_{(\boldsymbol{X},\boldsymbol{Y}_{\text{str}})}[\sqrt{(\boldsymbol{Y}_{\text{str}}-G(\boldsymbol{X}))^2+\epsilon^2}] \tag{15}$$

$$G^*=\arg\min_G\max_D\mathcal{L}_{\text{cGAN}}\mathcal{L}(G,D)+\lambda\mathcal{L}_C(G) \tag{16}$$

为了使估计图像和目标无阴影图像尽可能接近，使用被近似看作鲁棒的 L_1 损失的 Charbonnier[20] 损失 $\mathcal{L}_C$。Charbonnier[20] 损失有助于恢复全局结构，可以鲁棒地处理异常值。ϵ 是一个非常小的常数，本文设置为 10^{-6}，用于稳定训练。λ 用于平衡对抗损失和 Charbonnier[20] 损失。

3.5.2　照明估计子网损失函数

照明估计子网的损失与场景结构恢复子网有两处不同：1）目标图像为照明图像而不是无阴影图像；2）额外添加了聚焦于阴影区域的对抗训练。具体损失如下：

$$G^*_{\text{light}}=\arg\min_G\max_D\mathcal{L}_{\text{cGAN}}(G,D)+\arg\min_G\max_{D_{\text{shad}}}\mathcal{L}_{\text{cGAN}}(G,D_S)+\lambda\mathcal{L}_C(G) \tag{17}$$

其中，D_S 为额外添加的阴影判别器。

3.5.3　场景重渲染子网损失函数

场景重渲染子网的损失函数主要由 5 部分组成，分别是拉普拉斯损、SSIM 损失、像素级损失、照明梯度损失和感知损失。总的损失函数可表示为：

$$\mathcal{L}_{\text{Re-rendering}}=\mathcal{L}_C+\mathcal{L}_{\text{Light}}+\lambda_1\mathcal{L}_{\text{Lap}}+\lambda_2\mathcal{L}_{\text{SSIM}}+\lambda_3\mathcal{L}_{\text{Per}} \tag{18}$$

其中，$\mathcal{L}_{\text{Lap}}$表示拉普拉斯损失；$\mathcal{L}_{\text{SSIM}}$表示 SSIM 损失；$\mathcal{L}_C$ 表示像素级损失；$\mathcal{L}_{\text{Light}}$表示照明梯度损失；$\mathcal{L}_{\text{Per}}$表示感知损失；$\lambda_1$，$\lambda_2$，$\lambda_3$ 是平衡系数，本文分别设置为 0.1，0.01，0.01。

拉普拉斯损失 $\mathcal{L}_{\text{Lap}}$表示为：

$$\mathcal{L}_{\text{Lap}}=\sum_{ij}(D(\boldsymbol{Y})-D(\widetilde{\boldsymbol{Y}}))_{ij}^2 \tag{19}$$

其中，$D(\cdot)$表示拉普拉斯滤波器和输入图像卷积获得拉普拉斯矩阵的操作；i，j 为第 i 或第 j 个特征图；$\boldsymbol{Y}$，$\widetilde{\boldsymbol{Y}}$ 分别表示目标图像和估计图像。

其次，本文应用了 SSIM 损失，以帮助网络呈现更具有视觉吸引力的图像。SSIM 表示如下：

$$\mathcal{L}_{\text{SSIM}}=1-\text{SSIM}(\boldsymbol{Y},\widetilde{\boldsymbol{Y}}) \tag{20}$$

使用 $\mathcal{L}_C$ 计算逐像素重建误差，具体表达式如下：

$$\mathcal{L}_C=\sqrt{(\boldsymbol{Y},\widetilde{\boldsymbol{Y}})^2+\varepsilon^2} \tag{21}$$

在改变照明方向后，所估计的图像中距离光源越近的物体，照明强度应该越强；而距离光源较远的物体所受的照明强度应该越弱。为了使网络能够更加关注图像照明梯度的重建，首先对估计图像和目标图像进行高斯模糊，通过平滑细节纹理让网络聚焦学习图像中的照明梯度信息。照明梯度损失 $\mathcal{L}_{\text{Light}}$如下：

$$\mathcal{L}_{\text{Light}}=\sqrt{(B(\boldsymbol{Y})-B(\widetilde{\boldsymbol{Y}}))^2+\varepsilon^2} \tag{22}$$

其中，B 表示高斯模糊。感知损失被广泛应用于低级视觉任务。在重照明任务中，感知损失通过预训练深度神经网络提取的特征表示差距来评估估计图像和目标图像之间的视觉差异。感知损失 $\mathcal{L}_{\text{Per}}$定义为：

$$\mathcal{L}_{\text{Per}}=\|VGG_{19}(\boldsymbol{Y})-VGG_{19}(\widetilde{\boldsymbol{Y}})\| \tag{23}$$

其中，$VGG_{19}(\cdot)$代表从 VGG-19 网络中提取特征图的操作。

4 实验结果与分析

4.1 数据集和训练细节

4.1.1 数据集

本文使用由虚幻游戏引擎生成的虚拟数据集 VIDIT[21] 来评价所提方法的性能。VIDIT[21] 数据集提供分辨率为 1024×1024 的图像，包含 390 个不同的虚拟场景，其中 300 个场景用于训练，45 个场景用于验证，45 个场景用于测试。

4.1.2 训练细节

对 3 个子网络（场景结构恢复、照明估计和重渲染）分别进行训练。首先，利用输入图像和无照明图像的所有可能配对来进行场景结构恢复子网的训练；其次，利用输入图像和目标图像配对进行照明估计子网的训练。最后，固定前两个子网，训练场景重渲染网络。训练阶段，图像分辨率由 1024×1024 统一调整为 512×512，使用 Adam[22] 优化器，动量设置为 0.5，学习率为 2×10^{-4}。

4.2 视觉感知主观评价

图 4 展示了利用代表性方法的重照明结果来评估视觉感知质量。DRN[8] 和 WDRN[23] 来自 AIM 2020，WDRN[23] 取得比赛的第一名，它将小波变换应用于 UNet[7] 网络的编码-解码模块，以实现高效的图像照明转换。DRN[8] 则获得最佳的 PSNR。MCN[11] 是基于 DRN[8] 的改进方法。NTIRE 2021 提供图像深度信息，比赛的冠亚军分别为 MBNet[14] 和 OID-DRNet[13]。GridNet[24] 和 RetinexNet[25] 是在去雾和低光照增强领域具有代表性的方法。

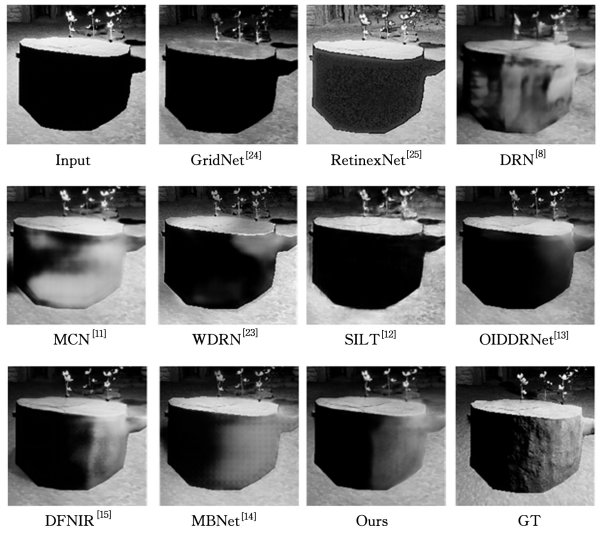

图 4 重照明结果的对比（木桩）

Fig. 4 Comparison of relighting results (wooden piles)

在当图像照明方向发生较大变化时，GridNet[24] 和 RetinexNet[25] 未能进行阴影的去除和重建。如图 4 和图 5 所示，DRN[8]，MCN[11]，WDRN[23]，SILT[12] 和 OIDDRNet[13] 生成的图像存在伪影和结构扭曲等缺陷，且生成阴影与照明方向不一致。同时，虽然 DFNIR[15] 和 MBNet[14] 生成了符合目标照明方向的阴影，但是在阴影边缘与目标照明图像仍然存在差异。从图 5 中可以很明显地看出，本文方法在地面上恢复出的阴影与目标照明有着高度的空间一致性，且能够很好地恢复出纹理复杂的木桶表面阴影。在没有深度图先验情况下，本文通过聚合局部和全局上下文信息，重新校准照明特征，展现出最接近于目标照明条件的全局照明效果，最大化恢复局部纹理，呈现出了更好的视觉效果。

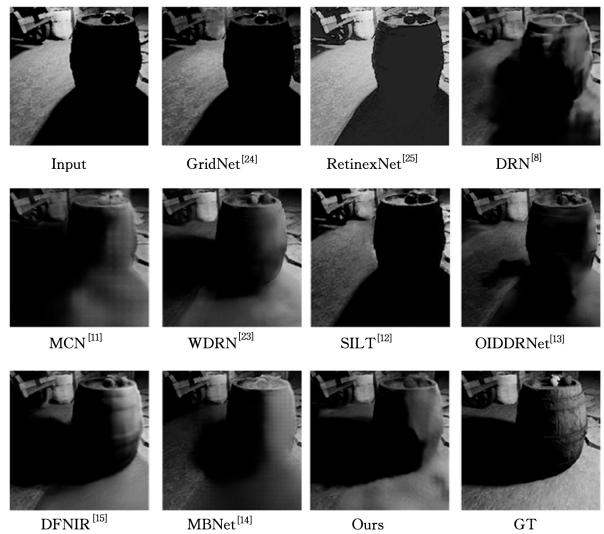

图 5 重照明结果的对比（木桶）

Fig. 5 Comparison of relighting results (barrels)

4.3　图像质量客观评价

选取 PSNR，SSIM[26]，LPIPS[27] 和 MPS[15,28] 4 种评估指标对所提方法与上述 9 种方法进行全面客观的比较，实验结果如表 1 所列。可以看出，文中所提出的方法总体上获得了最高的分数。具体来说，所提方法获得了最高的 PSNR，SSIM[26] 和 MPS[15,28]；同时，其在 LPIPS[27] 指标对比中具有竞争优势。深度信息可以提高网络对场景几何结构的理解，促使其生成符合人类视觉感知的图片。但是拥有深度信息的图像在日常生活中较难获取，这限制了此类方法的适用范围；同时，添加深度信息的操作并不能增强物理材质特性所引起的反光、透射等照明表现。如图 6 所示，OIDDRNet[13]，DFNIR[15] 和 MBNet[14] 均使用了深度图，但只有本文方法很好地展现出了金属地板的反光效果。

表 1　客观指标对比

Table 1　Comparison of objective indicators

Algorithm	*PSNR*(↑)	*SSIM*(↑)	*LPIPS*(↓)	*MPS*(↑)
RetinexNet[25]	12.1490	0.1660	0.5690	0.2990
SRN[29]	15.6720	0.5700	0.4070	0.5820
GridNet[24]	16.6730	0.2810	0.3690	0.4560
DRN[8]	17.5860	0.6090	0.3920	0.6080
WDRN[23]	17.4540	0.6640	0.2770	0.6940
MCN[11]	17.8580	0.6490	0.3800	0.6350
SILT[12]	17.0000	0.6060	0.4490	0.5785
DFNIR[15]	20.6830	0.6630	0.3050	0.6790
OIDDRNet[13]	17.9970	0.6830	0.2780	0.7030
MBNet[14]	18.4740	0.6640	**0.2740**	0.6950
Our Method	**22.2280**	**0.7450**	0.3210	**0.7120**

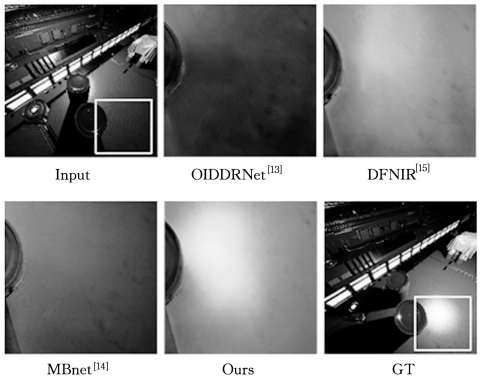

图 6　重照明结果对比（金属地板）

Fig. 6　Comparison of relighting results(metal floor)

4.4　消融实验

为了证明本文方法的有效性，对其中的 3 个模块进行消融实验，如图 7 和表 2 所示。如表 2(b)列所示，添加了上下文门控残差模块后，4 个评价指标均有提升。如图 7(a)列所示，重建过程缺少局部和全局空间上下文信息会导致所估计的阴影区域呈现不规则状，缺少门控机制则会导致被重新照明的无照明区域边缘模糊。图 7(b)添加了上下文门控残差模块后，目标照明下的阴影形状被正确估计，并且具有清晰的边缘。

如表 2(d)列所示，添加多尺度注意力有利于 PSNR 指标的提升。具体来说，如图 7(d)列所示，多尺度感知模块能在不损失分辨率的情况下增大感受野，较好地捕获全局照明信息。

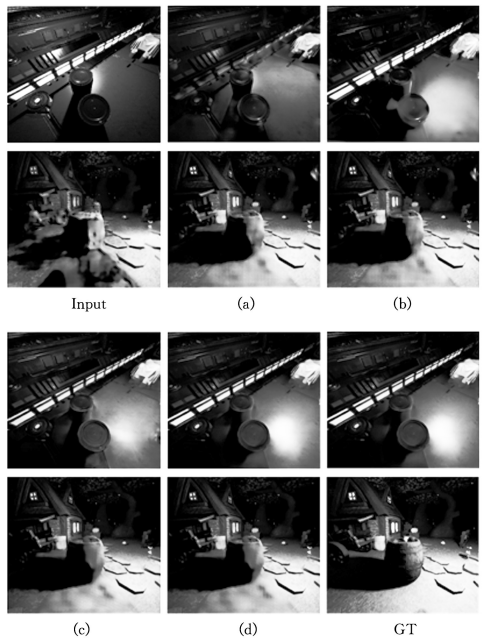

图 7　消融实验结果对比

Fig. 7　Comparison of ablation experimental results

表 2　消融实验的对比结果

Table 2　Comparative results of ablation experiments

Module	(a)	(b)	(c)	(d)
Context-gated residuals	×	√	√	√
Lighting gradient loss	×	×	√	√
Multi-scale attention	×	×	×	√
PSNR(↑)	17.50	20.10	21.70	**22.20**
SSIM(↑)	0.53	0.69	0.73	**0.75**
LPIPS(↓)	0.44	0.37	0.32	**0.32**
MPS(↑)	0.55	0.66	0.70	**0.71**

为了更明显地观察添加模块后所产生的变化，图 8 放大了局部细节进行比较。

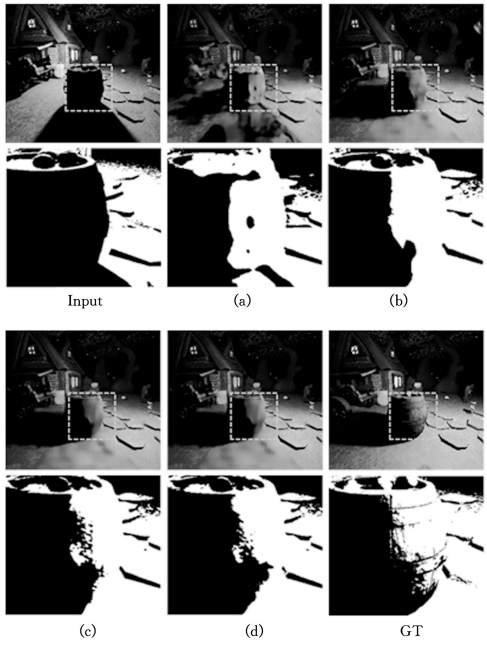

图 8　消融实验图像细节

Fig. 8　Details of ablation experimental images

从图 8(b)列和图 8(c)列木桶上的照明效果可以明显地

看出，添加了照明梯度损失后生成的阴影区域边缘细节更加清晰。本文方法虽然取得了很好的效果，但依然会丢失许多木桶上去阴影的重照明区域的纹理细节，没有将木质纹理很好地恢复出来。其原因是，输入图像的阴影区域包含大量无效像素，难以为纹理重建做出贡献。因此，未来对于相似纹理区域，其阴影部分可通过周围纹理进行填补。从图 7(b)列和图 7(c)列的第一行图像对比中可以明显看出，添加了照明梯度损失所估计的金属地板的反光明显和目标图像更相似。照明梯度损失由于是平滑了细节信息的模糊图像产生，因此能保证网络更加关注于学习不同方向的照明梯度，并生成更符合人类视觉感知的重照明图像。值得注意的是，如图 7 所示，虽然本文方法较好地渲染出了金属地板重照明之后的反光，但是其形状仍不规则，没有聚焦，之后应该在该区域施加惩罚项，以呈现更好的效果。同时，为提升模型鲁棒性，需要增加训练数据集图像中所涵盖的材质，特别是玻璃、塑料、瓷砖等易受照明变化而产生不同的反光和透射效果的材质。

结束语 文中提出了一种基于上下文门控残差和多尺度注意的图像重照明算法。上下文门控残差能够聚合局部和全局空间上下文信息，丰富特征表示能力，并通过门控机制生成更精细的重照明图像。多尺度感知能够扩大感受野，改善全局照明变换效果，同时应用双重注意使得网络关注阴影的去除和重建，保持照明和阴影的空间一致性。此外，本文添加了照明梯度损失，模糊纹理，使得网络更加关注照明的变换，增强了对低频纹理和照明效果差异的辨别能力。实验结果证明，文中所提方法既能完成关注全局照明效果的变换，又能生成符合目标照明形状准确的阴影。值得一提的是，本文算法在未使用深度信息的情况下依然能达到和使用深度信息方法相匹敌的性能。

未来，将侧重依靠周围的像素对重照明区域(阴影区域或黑暗区域)进行纹理修复(参考图 8)；针对金属、玻璃等材质物体的重照明任务进行研究，使重照明的结果更加逼真(参考图 7)。

参 考 文 献

[1] HOU Q, ZHOU D, FENG J. Coordinate attention for efficient mobile network design[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 13713-13722.

[2] ZHANG E, COHEN M F, CURLESS B. Emptying, refurnishing, and relighting indoor spaces[J]. ACM Transactions on Graphics(TOG), 2016, 35(6): 1-14.

[3] DUCHENE S, RIANT C, CHAURASIA G, et al. Multi-view intrinsic images of outdoors scenes with an application to relighting[J]. ACM Transactions on Graphics, 2015, 34(5): 1-16.

[4] KARSCH K, HEDAU V, FORSYTH D, et al. Rendering synthetic objects into legacy photographs[J]. ACM Transactions on Graphics(TOG), 2011, 30(6): 1-12.

[5] PEERS P, MAHAJAN D K, LAMOND B, et al. Compressive light transport sensing[J]. ACM Transactions on Graphics (TOG), 2009, 28(1): 1-18.

[6] REDDY D, RAMAMOORTHI R, CURLESS B. Frequency-space decomposition and acquisition of light transport under spatially varying illumination[C]// European Conference on Computer Vision. Berlin: Springer, 2012: 596-610.

[7] RONNEBERGER O, FISCHER P, BROX T. U-net: Convolutional networks for biomedical image segmentation[C]// International Conference on Medical Image Computing and Computer-assisted Intervention. Cham: Springer, 2015: 234-241.

[8] WANG L W, SIU W C, LIU Z S, et al. Deep relighting networks for image light source manipulation[C]// European Conference on Computer Vision. Cham: Springer, 2020: 550-567.

[9] HU Z, HUANG X, LI Y, et al. SA-AE for any-to-any relighting[C]// European Conference on Computer Vision. Cham: Springer, 2020: 535-549.

[10] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 1125-1134.

[11] WANG Y, LU T, ZHANG Y, et al. Multi-scale self-calibrated network for image light source transfer[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 252-259.

[12] KUBIAK N, MUSTAFA A, PHILLIPSON G, et al. SILT: Self-supervised Lighting Transfer Using Implicit Image Decomposition[J]. arXiv: 2110. 12914, 2021.

[13] YAZDANI A, GUO T, MONGA V. Physically inspired dense fusion networks for relighting[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 497-506.

[14] YANG H H, CHEN W T, LUO H L, et al. Multi-modal bifurcated network for depth guided image relighting[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 260-267.

[15] El HELOU M, ZHOU R, SUSSTRUNK S, et al. NTIRE 2021 depth guided image relighting challenge[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021: 566-577.

[16] JIANG Y, GONG X, LIU D, et al. Enlightengan: Deep light enhancement without paired supervision[J]. IEEE Transactions on Image Processing, 2021, 30: 2340-2349.

[17] LIU J J, HOU Q, CHENG M M, et al. Improving convolutional networks with self-calibrated convolutions[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 10096-10105.

[18] HENDRYCKS D, GIMPEL K. Gaussian error linear units(gelus)[J]. arXiv: 1606. 08415, 2016.

[19] HU J, SHEN L, SUN G. Squeeze-and-excitation networks[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 7132-7141.

[20] LAI W S, HUANG J B, AHUJA N, et al. Fast and accurate

image super-resolution with deep laplacian pyramid networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,41(11):2599-2613.

[21] HELOU M E,ZHOU R,BARTHAS J,et al. VIDIT:Virtual image dataset for illumination transfer[J]. arXiv:2005.05460,2020.

[22] KINGMA D P,BA J. Adam:A method for stochastic optimization[J]. arXiv:1412.6980,2014.

[23] PUTHUSSERY D,PANIKKASSERIL S H,KURIAKOSE M,et al. Wdrn:A wavelet decomposed relightnet for image relighting[C]// European Conference on Computer Vision. Cham:Springer,2020:519-534.

[24] LIU X,MA Y,SHI Z,et al. Griddehazenet:Attention-based multi-scale network for image dehazing[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019:7314-7323.

[25] WEI C,WANG W,YANG W,et al. Deep retinex decomposition for low-light enhancement[J]. arXiv:1808.04560,2018.

[26] WANG Z,BOVIK A C,SHEIKH H R,et al. Image quality assessment:from error visibility to structural similarity[J]. IEEE Transactions on Image Processing,2004,13(4):600-612.

[27] ZHANG R,ISOLA P,EFROS A A,et al. The unreasonable effectiveness of deep features as a perceptual metric[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:586-595.

[28] EI HELOU M,ZHOU R,SUSSTRUNK S,et al. AIM 2020:Scene relighting and illumination estimation challenge[C]// European Conference on Computer Vision. Cham:Springer,2020:499-518.

[29] TAO X,GAO H,SHEN X,et al. Scale-recurrent network for deep image deblurring[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018:8174-8182.

WANG Wei, born in 1995, postgraduate, is a member of China Computer Federation. His main research interests include image processing and visual positioning.

JIN Cheng, born in 1978, Ph.D, professor, Ph.D supervisor, is a member of China Computer Federation. His main research interests include computer vision and multimedia information retrieval.

(责任编辑:柯颖)