

局部-全局关系耦合的低照度图像增强

王克琪^{1,3}, 钱宇华^{1,2*}, 梁吉业², 刘畅¹, 黄琴¹, 陈路¹, 贾洁茹¹

1. 山西大学大数据科学与产业研究院, 太原 030006
2. 山西大学计算智能与中文信息处理教育部重点实验室, 太原 030006
3. 盘古深度智能信息技术有限公司, 太原 030006

* 通信作者. E-mail: jinchengqyh@sxu.edu.cn

国家自然科学基金重点项目 (批准号: 62136005), 国家重点研发计划 (批准号: 2020AAA0106100, 2018YFB1004300)、山西省重点研发计划 (批准号: 201903D421003)、山西省高等学校科技创新项目 (2020L0016)、国家自然科学基金 (批准号: 61673249, 62003200, 62106133) 和山西省海外归国人员研究项目 (批准号: 2017023, 2016004) 资助

摘要 卷积神经网络目前在人工智能多个领域均取得了不同程度的进展. 卷积计算是基于参数共享的滑动机制, 这导致卷积神经网络更多地关注特征信息的局部关系, 对全局关系的建模能力有限. 局部关系和全局关系对特征的表达均具有重要的作用. 为此, 本文聚焦于如何对特征信息的局部-全局关系进行构建并有效耦合, 从而挖掘更加丰富的特征信息, 进而提高特征的判别性. 本文提出了局部-全局关系耦合模块, 该模块是由特征提取、基于深度卷积 (Depth-wise Convolution, DWConv) 的局部关系构建分支、基于多头自注意力机制 (Mutli-head Self-attention, MHSA) 的全局关系构建分支和基于点向卷积 (Point-wise Convolution, PWConv) 的关系耦合四部分组成. 基于此模块, 本文构建了编解码结构的局部-全局关系耦合神经网络, 该网络可以对特征信息的局部-全局关系进行建模, 增强特征信息的表征能力, 进而提升模型的性能. 为验证所提算法的有效性, 本文在低照度图像增强任务上, 使用基准数据集与其它算法进行了实验对比. 实验结果表明, 本文所提出的方法取得了较好的图像增强结果, 优于当前先进的图像增强方法. 最后, 本文通过消融实验和扩展实验从多个角度进一步验证了有效耦合局部-全局关系的重要性和可扩展性.

关键词 低照度图像增强, 局部-全局关系, 计算机视觉, 卷积神经网络, 深度学习

1 引言

深度学习作为当今人工智能领域的研究热点已被广泛应用于计算机视觉^[1~6]、自然语言处理^[7,8] 和行为识别^[9] 等任务. 卷积神经网络作为深度学习中的一项重要工作, 已经成为现代计算机视觉任务中不可或缺的部分. 卷积操作是利用基于参数共享的滑动机制对输入的内容进行计算, 这种方式对特征信息的全局关系建模能力较弱. 如何对特征信息的局部-全局关系有效建模是一项重要的研究内容, 本文尝试从以下两个角度回答这个问题:

- ① 如何构建特征信息的局部关系和全局关系?
- ② 如何对已构建的局部关系和全局关系进行有效的耦合?

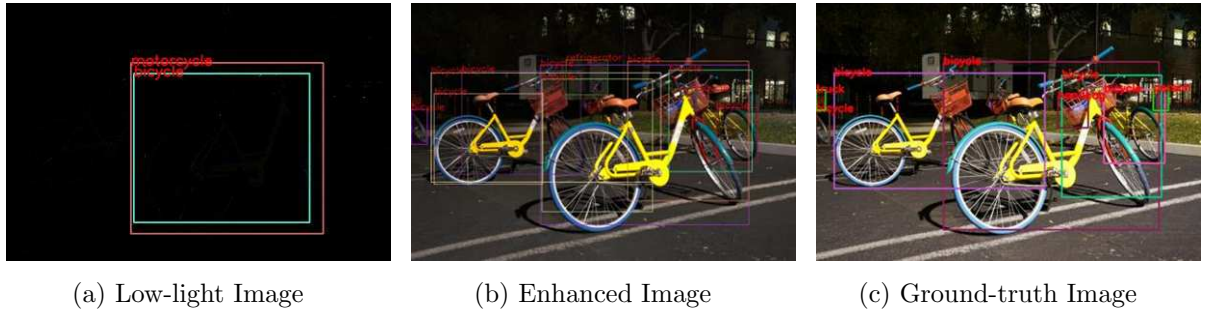


图 1 (网络版彩图) 图像质量对目标检测算法的影响

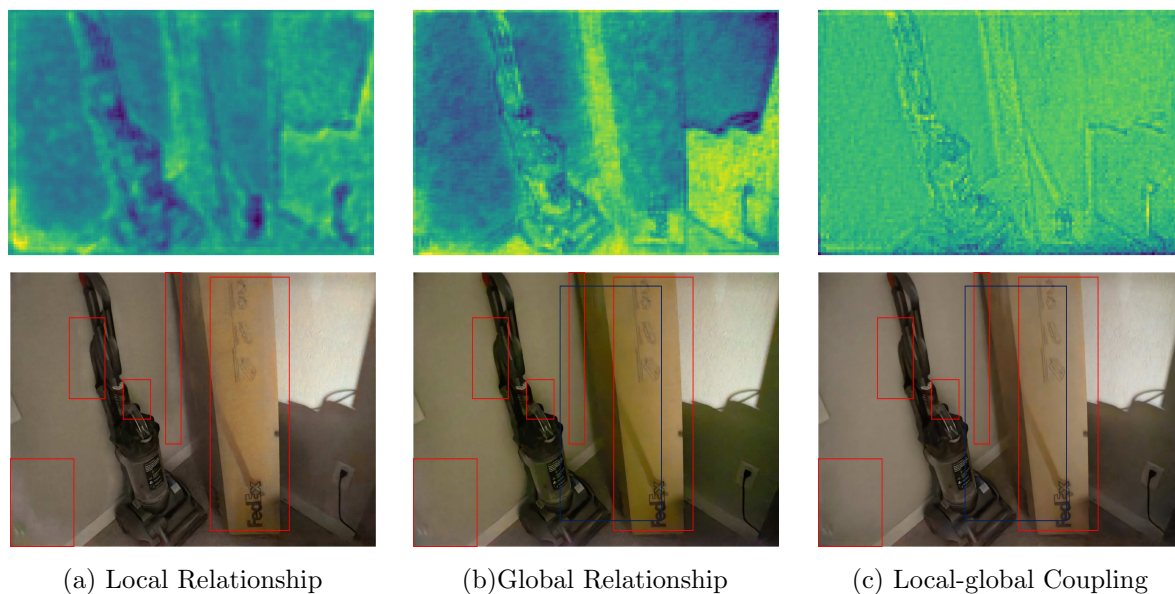
Figure 1 (Color online) The influence of image quality on object detection algorithm

针对以上两个问题, 本文提出了局部-全局关系耦合模块 (Local-Global Relationship Coupling Block, LGRCB), 该模块可以同时特征信息的局部关系和全局关系进行建模, 并能有效耦合两部分关系. 基于此模块, 本文构建了局部-全局关系耦合神经网络 (Local-Global Relationship Coupling Neural Network, LGRCN). 为了验证所提算法的有效性, 本文在低照度图像增强任务上进行了深入的研究.

在夜间或光照条件较差的情况下, 由于成像设备捕获的光子数量有限, 成像后的图像普遍存在信噪比低、能见度较差、以及细节丢失严重等问题. 此类图像或视频直接影响了人类的视觉感知能力, 对公共安全、交通管理等应用领域的监控造成了极大的不便, 且影响了其它基于视觉信息的计算机视觉算法的性能, 如目标检测、语义分割、人体姿态估计等. 图 1 展示了不同质量的图像对目标检测算法的影响, 图 1(a) 是成像设备在低照度环境下捕获的图像, 图 1(b) 是使用本文算法增强后的图像, 图 1(c) 是使用相机的长曝光模式 (曝光时间为10 秒) 捕获的图像, 从图中可以看到图像质量对目标检测算法的性能影响较大 (本文使用的目标检测算法是 Fast R-CNN^[1]).

由于成像过程是一项不可逆操作, 图像增强算法是一种有效提高图像质量的手段. 现阶段已有大量的研究人员对低照度图像增强算法开展了一系列研究, 并取得了大量的研究成果, 如基于直方图均衡的方法^[10,11]、基于Retinex理论的方法^[15,16]、基于卷积神经网络的方法^[2,30,31]等. 在现实场景中, 前两种方法具有较大的局限性, 难以适用于复杂多变的图像内容和光照环境, 而基于卷积神经网络的低照度图像增强方法具有较强的适应性和鲁棒性. 由于卷积操作是基于参数共享的滑窗机制, 其仅考虑了邻域范围内的特征信息, 这限制了其直接构建特征信息全局关系的能力, 因此增强图像仍存在一定程度的质量问题 (噪声、伪影等). 图 2(a) 是只考虑局部关系的特征图和增强图, 从图示可以看到, 仅考虑特征间的局部关系会造成属于同一区域的低频信息的特征表示不一致, 造成增强图像出现色彩过渡不均衡、伪影等问题 (图 2(a) 中红色框标注). 构建特征间的全局关系^[21~23], 可以使得特征中的纹理结构, 边缘细节等信息被充分利用, 从而更好地理解场景内容, 恢复图像质量. 自注意力机制是一种有效构建特征间全局关系的方法^[26~28], 可以直接对不同位置的特征信息进行全局关系建模. 图 2(b) 是使用全局关系构建的特征图和增强图, 从图中可以看到, 增加全局关系可以增强特征的代表能力, 但是由于其忽略了特征间的局部关系, 从而造成增强图像色彩偏差较大 (图 2(b) 蓝色框标注).

由此可以看到, 在深度神经网络中仅考虑局部关系或全局关系的特征建模方法都具有一定局限性, 本文提出的局部-全局关系耦合模块 (Local-Global Relationship Coupling Block, LGRCB) 可以同时特征信息的局部关系和全局关系进行建模, 并能有效地对两部分关系进行耦合, 从而增强特征代表能力, 提升增强图像质量. 从图 2(c) 中可以看到, 增加局部-全局关系耦合的模型刻画的特征更加清晰, 增强图像的视觉质量有明显提高, 对偏色、色彩过渡不均匀、伪影等问题有较好的改善. LGRCB由



(a) Local Relationship

(b) Global Relationship

(c) Local-global Coupling

图 2 (网络版彩图) 局部-全局关系对增强图像质量的影响

Figure 2 (Color online) The effectiveness of the local-global relationship on enhanced image quality

特征提取、局部关系构建分支、全局关系构建分支以及特征耦合四部分组成。在特征提取部分, 本文采用通道分割操作提取输入的特征, 并将提取到的两部分特征分别输入到局部关系分支和全局关系分支。为保证该模块的轻量化, 在局部关系的建立上本文使用深度卷积 (Depth-wise Convolution, DWConv)。在全局关系的建立上本文使用多头自注意力机制 (Mutli-head Self-attention, MHSA), 使得每一个特征可以直接和其它特征建立全局关系。在耦合策略的选择上, 本文使用点向卷积 (Point-wise Convolution, PWConv) 对建模后的局部特征和全局特征进行耦合。

本文利用多个公开数据集在低照度图像增强任务上进行了大量实验。实验结果表明, 本文方法在定量分析的结果上优于其它方法, 在定性分析的结果中显著提升增强图像的视觉质量, 充分证明了本文方法的有效性。同时, 本文通过消融实验从多个角度分析了有效耦合局部-全局关系的重要性以及本文使用的全局特征提取方法的合理性。最后, 为了验证局部-全局关系耦合模块的可扩展性, 本文在多个任务上进行了扩展实验。实验结果表明, 通过在原方法上添加局部-全局关系耦合模块, 可以为原方法性能带来不同程度的性能提升。

2 相关工作

本节对一些相关工作进行简单回顾。在 2.1 小节回顾了低照度图像增强相关的研究, 在 2.2 小节回顾了局部关系与全局关系相关的工作。

2.1 低照度图像增强

光照条件是影响相机成像质量的关键因素。在低照度或夜间环境下, 由于成像设备捕获的光子数量有限, 因此难以清晰成像。图像增强算法是提高低照度图像的视觉质量的有效的手段。在图像增强算法中, 直方图均衡算法^[10,11]是一类经典且有效的算法, 该方法使用图像直方图对对比度进行调整, 从而达到增强图像视觉质量的效果。受人类视觉感知理论启发, Retinex 理论^[12~14]认为图像可以被分

解为光照图像和反射图像两部分. 文献 [16,17] 提出了基于该理论的低照度图像增强方法. 文献 [18] 结合 Retinex 理论和深度学习将输入内容分解为光照和反射两部分分别进行增强. 为了对网络进行有效的训练, 作者收集了具有配对图像的低照度图像数据集 (Low-light Paired Dataset, LOL). 传统的低照度图像增强算法处理的图像大部分为 RGB 图像, RGB 图像是 RAW 格式数据经过相机成像算法 (Image Signal Processor, ISP) 处理后获得. 在成像算法处理过程中会存在信息损失的问题, RGB 图像通常仅包含 8-bit 的信息量, 而 RAW 格式数据却拥有 12-bit 或者更高比特的信息量, 因此 RAW 格式数据比 RGB 图像存储着更加丰富的信息, 更加适用于低照度图像增强算法. 针对此问题, 文献 [2] 做了大量的研究, 作者收集了 SID (See-in-the-dark) 数据集, 该数据集包含有低照度环境下的 RAW 格式数据以及对应的长曝光模式下的清晰图像. 为建模适用于 RAW 格式数据的增强算法, 作者提出了基于深度学习算法的低照度图像增强方法, 该方法以 RAW 格式数据作为输入, 输出 RGB 格式的图像.

2.2 局部关系与全局关系

在卷积神经网络中, 大量的研究通过探索与构建更加合理的网络结构以提高卷积神经网络对特征的代表能力, 提升模型性能. 卷积操作通过参数共享的滑窗机制对像素局部邻域内的关系进行建模, 导致它对像素的全局关系的建模能力较弱. 然而, 全局关系对构建更有效的特征表示是至关重要的. 文献 [19] 指出了对数据间建立关联关系的重要性. 文献 [20] 在聚类集成算法上分析了局部与全局关系的重要性. 在图像处理中, 全局关系同样对算法性能的提升具有重要作用. 在经典的图像去噪算法 NL-means^[21] 和 BM3D^[22] 中都指出了全局关系建模的重要性. 自注意力机制是一种有效的全局特征的构建方式. 基于自注意力机制的 Transformer^[24] 和 Non-local Neural Network^[25] 分别在自然语言处理领域和计算机视觉领域取得了意义重大的进展. 近年来, 越来越多的研究人员致力于探索 Transformer 在计算机视觉任务上的可行性^[26~28], 并在多个视觉任务上取得了一定的突破. 然而, 目前的多数工作是在计算机视觉的中高级任务上开展的研究, 针对低级视觉任务的研究十分稀少. 再者, 目前较多的工作关注于探索更有效的全局关系建模方法或构建更鲁棒的卷积神经网络结构, 对有效耦合局部-全局关系的研究较少.

3 局部-全局关系耦合的低照度图像增强方法

本节对局部-全局关系耦合的低照度图像增强算法进行详细介绍. 在 3.1 小节介绍本文提出的算法的整体流程; 在 3.2 小节对本文提出的局部-全局关系耦合模块进行阐述; 在 3.3 小节对局部-全局关系耦合模块的实例化方法进行介绍.

3.1 低照度图像增强流程

本节对低照度图像增强流程进行详细阐述, 具体的低照度图像增强算法流程如图 3 所示. 其中红色箭头流是 RGB 图像的处理流程, 黑色箭头流是 RAW 数据的处理流程. 针对 RAW 格式数据的增强, 本文将其分为数据预处理、图像增强处理及像素恢复处理三部分进行逐级处理^[2]. 在实施过程中, 首先对数据进行预处理操作, 具体的操作流程如下:

① 由于成像设备的传感器自身存在暗电流, 这导致成像设备在没有捕捉到光子的情况下也会产生电压值, 因此需要对成像设备捕获的 RAW 数据进行黑电平矫正, 即减去暗电流产生的噪声, 计算

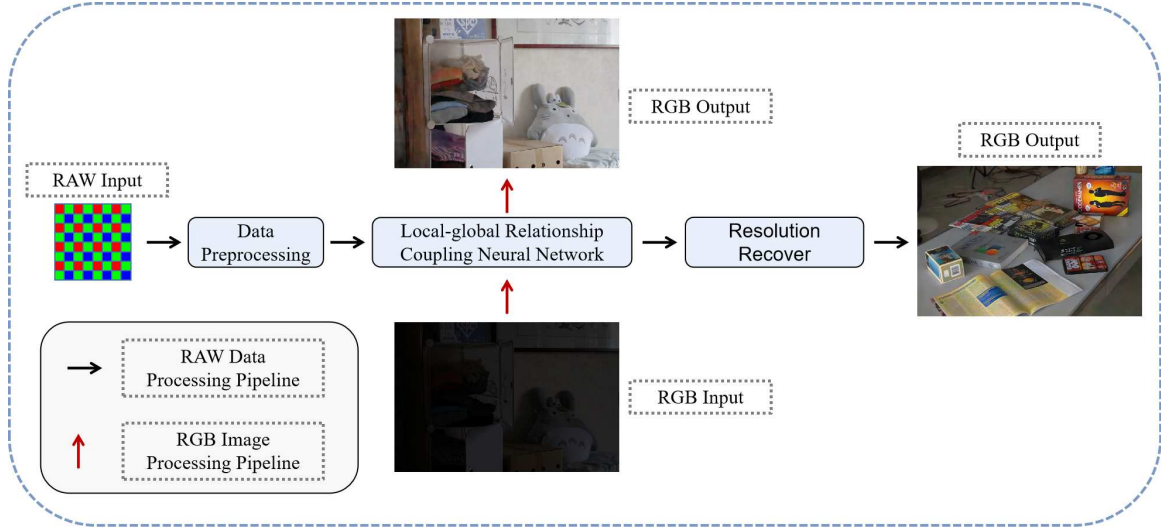


图 3 (网络版彩图) 低照度图像增强流程图

Figure 3 (Color online) The pipeline of low-light image enhancement

过程如式 (1) 所示:

$$RAW' = RAW - OB. \quad (1)$$

其中, RAW 代表相机传感器光电转换后的数据, OB (Optical Black) 代表黑电平的数值大小, RAW' 代表黑电平矫正后的数据;

② 为降低卷积神经网络的资源消耗, 对 RAW' 进行分辨率缩放操作. 针对 Sony 相机的数据, 按 Bayer 阵列中像素 RGG B 的排列方式, 通过像素变换操作, 将 RAW' 分辨率缩减为原分辨率的 $1/2$, 通道数由 1 通道扩展至 4 通道; 针对 Fuji 相机的数据, 将 RAW' 分辨率缩减为原分辨率的 $1/3$, 通道数由 1 通道扩展至 9 通道;

③ 使用缩放系数 Ratio 对输入数据进行放缩处理, 式 (2) 给出了数据的放缩处理方式, 式 (3) 给出了 Ratio 的计算方式:

$$RAW'' = RAW' * Ratio, \quad (2)$$

$$Ratio = \frac{E_{gt}}{E_{raw}}. \quad (3)$$

其中, RAW'' 是经过处理后的数据, E_{gt} 是参考图像的曝光时间, E_{raw} 是输入的 RAW 数据的曝光时间. 将预处理后的数据输入局部-全局关系耦合的卷积神经网络对图像进行增强, 最后使用 sub-pixel^[34] 操作将网络的输出恢复为三通道的 RGB 图像.

如图 4 (LGRCN 部分) 所示是本文构建的局部-全局关系耦合神经网络 (Local-Global Relationship Coupling Neural Network, LGRCN), 其结构类似于 U-net 网络, 由编码器和解码器两部分组成. 在编码器部分, 首先将输入的图像编码到一个较低的维度 F_{low} , 随后将 F_{low} 输入到局部-全局关系耦合模块. 最后, 将耦合后的特征 F_{cou} 输入到解码器将图像恢复至原始分辨率. 本文使用的编码器和解码器均是由四个卷积块构成, 不同的是编码器由两个卷积层和一个池化下采样层构成, 解码器由两个卷积层和一个反卷积上采样层构成.

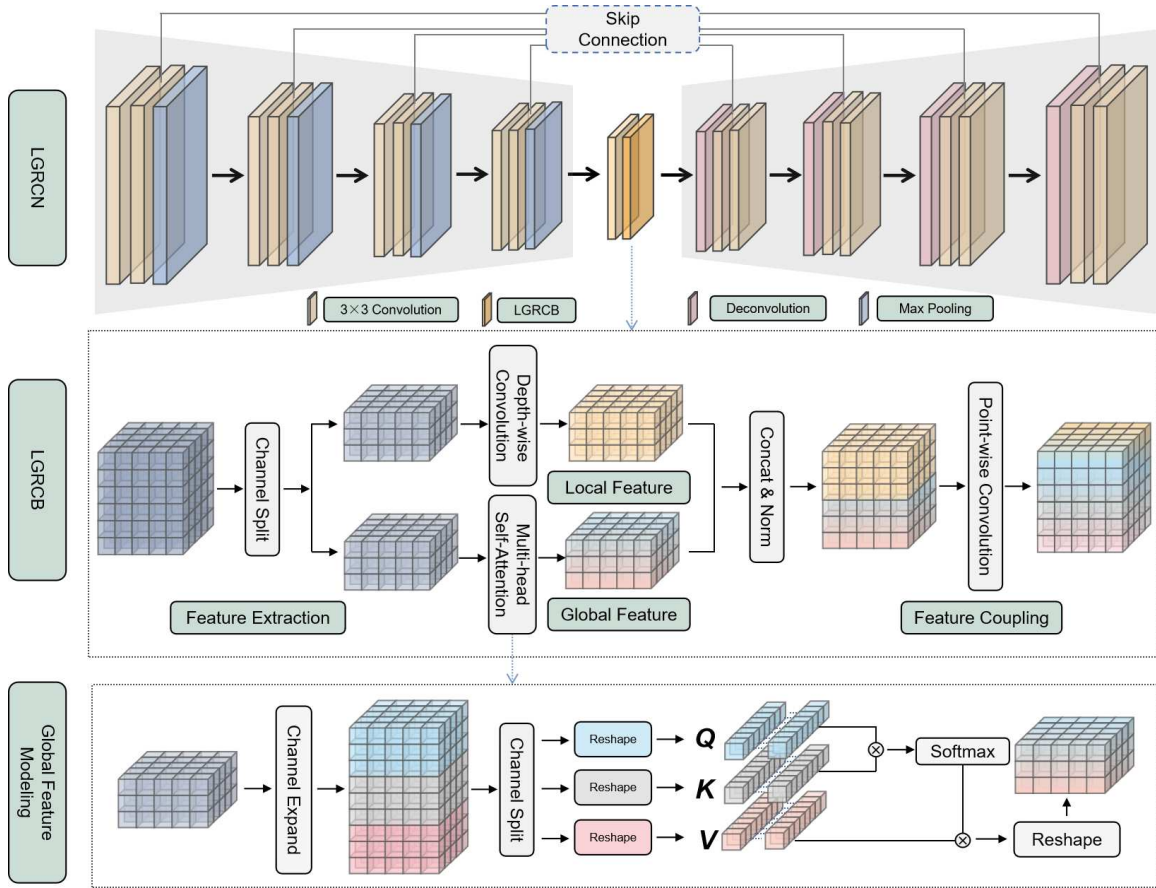


图 4 (网络版彩图) 局部-全局关系耦合神经网络

Figure 4 (Color online) Local-Global relationship coupling neural network

3.2 局部-全局关系耦合模块

卷积神经网络是基于权重共享的滑动窗口操作对输入的特征信息进行计算, 这种计算方式仅能构建窗口中心像素与邻接像素间的关系, 难以直接刻画不同位置的像素关系, 特别是对位置距离较远的像素. 然而, 特征信息的局部关系和全局关系对卷积神经网络构建有效的特征表示均有重要影响. 为增强卷积神经网络对全局关系的建模能力, 本文提出了局部-全局关系耦合模块. 该模块由特征提取、局部关系建模分支、全局关系建模分支以及特征耦合四个子模块构成. 特征提取子模块对输入的特征信息进行特征提取, 并将提取的特征分别输入到局部关系建模分支和全局关系建模分支. 局部关系建模分支和全局关系建模分支分别对特征信息的局部关系和全局关系进行建模, 并通过特征耦合模块对构建的局部特征和全局特征进行有效耦合. 该模块可以有效增加特征的代表能力, 进而提升网络性能.

局部-全局关系耦合模块. 接下来本文对局部-全局关系耦合模块进行详细的阐述. 设输入到局部-全局关系耦合模块的特征信息为 $\mathbf{F}_{\text{low}} \in \mathbb{R}^{C \times H \times W}$, 首先将其输入到特征提取子模块得到新的特征 $\{\mathbf{F}_{\text{local}}, \mathbf{F}_{\text{global}}\}$, 再分别将 $\mathbf{F}_{\text{local}}$ 输入局部关系建模分支, $\mathbf{F}_{\text{global}}$ 输入全局关系建模分支. 最后, 将两个分支构建的特征通过特征耦合模块进行有效耦合, 得到耦合后的特征 $\mathbf{F}_{\text{cou}} \in \mathbb{R}^{C \times H \times W}$. 局部-全

局关系耦合模块定义如式 (4) 所示:

$$\mathbf{F}_{\text{cou}} = f(l(\mathbf{F}_{\text{local}}), g(\mathbf{F}_{\text{global}})). \quad (4)$$

其中, $l(\cdot)$ 表示局部关系建模过程, $g(\cdot)$ 表示全局关系建模过程, $f(\cdot, \cdot)$ 表示局部-全局关系耦合过程.

局部关系建模分支. 在局部关系建模分支, 特征信息内的特征点仅与滤波器窗口内的特征信息构建局部关系, 对于局部关系的计算如式 (5) 所示:

$$l(\mathbf{F}_{\text{local}}(i, j)) = \sum_{m=-h/2}^{h/2} \sum_{n=-w/2}^{w/2} \mathbf{F}_{\text{local}}(i+m, j+n) \mathbf{A}(m, n). \quad (5)$$

其中, (i, j) 和 (m, n) 分别表示特征信息的位置索引和滤波器的位置索引, (h, w) 表示滤波器窗口的大小, \mathbf{A} 表示滤波器的权重矩阵.

全局关系建模分支. 在全局关系建模分支, 特征信息内的每一个特征点与特征信息的所有特征点构建全局关系, 对于全局关系的计算方式如式 (6) 所示:

$$g_i(\mathbf{F}_{\text{global}}) = \sum_{\forall j} \frac{\exp\{\theta_i(\mathbf{F}_{\text{global}})(\phi_j(\mathbf{F}_{\text{global}}))^T\}}{\sum_{i=0}^{N-1} \exp\{\theta_i(\mathbf{F}_{\text{global}})(\phi_j(\mathbf{F}_{\text{global}}))^T\}} \varphi_j(\mathbf{F}_{\text{global}}). \quad (6)$$

其中, $g_i(\mathbf{F}_{\text{global}})$ 表示 i 位置的特征与其它可能的 j 位置的特征构建关系, $\theta(\cdot)$, $\phi(\cdot)$ 和 $\varphi(\cdot)$ 表示特征嵌入过程, N 表示特征点的数量.

特征耦合模块. 在特征耦合模块, 首先将局部关系建模分支和全局关系建模分支构建的特征在通道维度进行拼接, 随后对拼接后的特征进行归一化操作, 并记为 $\mathbf{F}_{\text{cat}} \in \mathbb{R}^{C \times H \times W}$, 最后对 \mathbf{F}_{cat} 进行特征耦合得到 $\mathbf{F}_{\text{cou}} \in \mathbb{R}^{C \times H \times W}$.

3.3 局部-全局关系耦合模块实例化

图 4 (LGRCB 部分) 是本文在低照度图像增强任务中对局部-全局关系耦合模块的实例化方式. 在特征提取子模块的实例化方式上, 本文采用通道分割操作对其实例化. 设输入到局部-全局关系耦合模块的特征为 $\mathbf{F}_{\text{low}} \in \mathbb{R}^{C \times H \times W}$, 随后通过通道分割操作得到 $\{\mathbf{F}_{\text{local}}, \mathbf{F}_{\text{global}}\} \in \mathbb{R}^{\frac{C}{2} \times H \times W}$. 在局部关系建模分支本文采用深度卷积 (Depth-wise Convolution, DWConv), 这样可以在保证性能的同时降低该模块的参数量. 深度卷积是一种高效且轻量的卷积运算方式, 相比于传统的卷积层运算, 其缺少了构建特征信息通道关系的能力, 然而这种关系在全局关系构建分支已被隐式构建. 在全局关系建模分支, 本文借鉴了文献 [24] 中多头自注意力机制的设计并对其进行了轻微的修改. 如图 4 (Global Feature Modeling 部分) 所示, 本文首先使用点向卷积 (Point-wise Convolution, PWConv) 对输入的特征 $\mathbf{F}_{\text{global}} \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ 进行通道扩张操作 (Channel Expand) 得到 $\mathbf{F}_{\text{global}} \in \mathbb{R}^{\frac{3C}{2} \times H \times W}$, 随后将 $\mathbf{F}_{\text{global}} \in \mathbb{R}^{\frac{3C}{2} \times H \times W}$ 通过通道分割操作划分为 $\mathbf{Q} \in \mathbb{R}^{\frac{C}{2} \times H \times W}$, $\mathbf{K} \in \mathbb{R}^{\frac{C}{2} \times H \times W}$, $\mathbf{V} \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ 三部分, 再使用矩阵变换操作将其变换为 $\mathbf{Q} \in \mathbb{R}^{M \times N \times \frac{C}{2M}}$, $\mathbf{K} \in \mathbb{R}^{M \times N \times \frac{C}{2M}}$, $\mathbf{V} \in \mathbb{R}^{M \times N \times \frac{C}{2M}}$, 其中 M 代表多头注意力机制中头的数量, 本文中 $M=8$, $N=H \times W$. 随后通过自注意力机制 (Self-Attention) 计算特征的全局关系, 最后通过矩阵变换将计算的结果 $\mathbf{F}_{\text{global}} \in \mathbb{R}^{M \times N \times \frac{C}{2M}}$ 变换为 $\mathbf{F}_{\text{global}} \in \mathbb{R}^{\frac{C}{2} \times H \times W}$ 得到全局关系特征. 自注意力机制的计算方式如式 7 所示:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^T}{\sqrt{d_k}}\right)\mathbf{V}. \quad (7)$$



图 5 (网络版彩图) SID Sony 数据集样例展示
Figure 5 (Color online) Examples of SID Sony Dataset



图 6 (网络版彩图) LOL 数据集样例展示
Figure 6 (Color online) Examples of LOL Dataset

在对特征的局部关系和全局关系进行建模后, 本文使用特征耦合模块对两部分特征进行有效耦合. 在特征耦合模块本文首先对局部关系建模分支和全局关系建模分支建模的特征在通道维进行特征拼接得到 $\mathbf{F}_{\text{cat}} \in \mathbb{R}^{C \times H \times W}$, 随后对其进行特征归一化操作, 并使用点向卷积 (Point-wise Convolution, PWConv) 对归一化后的特征进行特征融合, 最终得到耦合后的特征 $\mathbf{F}_{\text{cou}} \in \mathbb{R}^{C \times H \times W}$.

4 实验与结果

本节对本文提出的局部-全局关系耦合神经网络进行了大量的实验分析, 并在两个低照度图像增强任务常用的公开数据集 (LOL^[17]和 SID Sony^[2]) 上进行了测试, 通过大量的对比实验、消融实验和扩展实验验证了本文所提算法的有效性.

4.1 数据集

本文使用 LOL (Low light Paired Dataset)^[17] 数据集和 SID^[2] 数据集评测本文提出的算法. LOL 数据集包含 500 组真实场景下拍摄的图像, 每一组图像包含有配对的图像, 也就是正常照度图像和低照度图像, 其中有 485 组训练数据, 15 组测试数据. SID 数据集由 SID Sony 数据集 (SIDS) 和 SID-Fuji 数据集 (SIDF) 两部分组成. SIDS 数据集是由 Sony $\alpha 7s$ II 摄像机拍摄收集. 该数据集同时包含室内和室外场景图像, 在室内场景下采集图像的光照强度值在 0.03 lux 到 0.3 lux 之间, 在室外场景下采集图像的光照强度值在 0.2 lux 到 5 lux 之间. SIDS 数据集共包含 231 张长曝光图像和 2697 张不同曝光时长的短曝光图像, 其中每一张长曝光图像对应多张不同曝光时长的短曝光图像, 图像的分辨率为 4240×2832 . 在最初发布的 SIDS 数据集测试集中存在输入图像和标签图像未配准的问题, 作者在修正版本中删除了未配准的图像, 为方便书写本文将修正前和修正后的数据集分别命名为 SIDS-A 和 SIDS-B. 为了进行更客观的实验分析, 本文在两个版本数据集均进行测试. SIDF 数据集是由 Fujifilm X-T2 摄像机拍摄收集. 该数据集共包含 193 张长曝光图像和 2397 张不同曝光时长的短曝光图像, 其中每一张长曝光图像对应多张不同曝光时长的短曝光图像, 图像的分辨率为 6000×4000 . SID 数据集和 LOL 数据集的最大区别在于, SID 数据集使用的是相机传感器捕获的 RAW 数据, 而 LOL 数据集使用的是 RGB 图像. 为便于展示真实场景下数据集的详细信息, 图 5 列举了 SIDS 数据集的样例图像, 图 6 列举了 LOL 数据集的样例图像, 这两组图像的左图为正常照度图像, 右图为低照度图像.

4.2 训练细节与参数设置

接下来对本文实验过程中的训练细节与参数设置进行详细介绍. 在使用 LOL 数据集的实验中, 本文使用低照度图像作为输入, 使用正常照度的图像作为标签信息, 共进行 1500 次迭代训练, 在训练

的初始阶段将学习率设置为 10^{-4} , 进行 1000 次迭代后将学习率降为 10^{-5} , 训练过程中未进行数据增强操作. 在使用 SID 数据集的实验中, 本文使用短曝光 RAW 数据作为输入, 对应的长曝光 RGB 图像作为标签信息, 训练过程中使用随机裁剪, 随机反转和随机旋转等数据增强方法扩充样本量, 增加训练样本的多样性, 其中随机裁剪的大小为 512×512 , 共进行 4000 次迭代训练, 训练初始阶段将学习率设置为 10^{-4} , 进行 2000 次迭代后将学习率设置为 10^{-5} . 本文使用 L_1 损失作为本文的损失函数, 使用 Adam 优化器优化网络参数, 优化器的参数设置为默认参数.

4.3 对比的方法

为合理验证本文算法的有效性, 本文与现有算法分别从定性和定量两方面进行了对比分析. LOL 数据集使用的对比算法包括直方图均衡算法 (HE)、多曝光融合算法 (EFF) [32]、RetinexNet [17]、ZeroDCE [30]、Kind [31]. SID 数据集使用的对比算法包括直方图均衡算法 (HE)、多曝光融合算法 (EFF) [32]、DRCN [33]、LTS [2]、SubPixelCNN (SPCNN) [34]、EDSR [35]、SCGA-Net [36]. 在上述选择的对比方法中, 包含有基于经典的图像处理方法的图像增强方法 [32]、基于有监督训练方式的低照度图像增强方法 [2, 17, 31, 36]、基于无监督训练方式的低照度图像增强方法 [30]、及一些在图像复原领域常用的深度学习方法 [33~35].

4.4 度量标准

本文使用三个常用的图像质量评价指标对本文算法进行定量分析, 其分别是峰值信噪比 (Peak Signal-to-Noise, PSNR)、结构相似性 (Structural Similarity Indexm, SSIM) 以及基于深度学习的有参考图像评价指标 [37] (Deep Image Structure and Texture Similarity, DISTS).

PSNR:

$$\text{PSNR} = 10 \times \log_{10}\left(\frac{\text{MAX}_I^2}{\text{MSE}}\right), \quad (8)$$

$$\text{MSE} = \frac{1}{m \times n} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [\mathbf{I}(i, j) - \mathbf{K}(i, j)]^2, \quad (9)$$

$$\text{MAX}_I = 2^B - 1. \quad (10)$$

其中, \mathbf{I} 和 \mathbf{K} 分别表示标签图像和增强后的图像, B 代表图像的 bit 数, m 和 n 代表图像的长和宽.

SSIM:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \quad (11)$$

其中, μ_x^2 和 μ_y^2 为 x 和 y 的均值, σ_x^2 和 σ_y^2 分别表示 x 和 y 的方差, σ_{xy} 为 x 和 y 的协方差, c_1 和 c_2 是两个常量, 分别等于 $(k_1L)^2$ 和 $(k_2L)^2$, k_1 和 k_2 是两个超参数, L 为像素值的动态范围的最大值 $2^B - 1$ (对于 8-bit 图像该值为 255).

DISTS: 在低照度图像增强任务中, 由于成像设备和光照条件的限制, 真实场景拍摄的标签图像往往含有噪声. 如图 7 所示, 由于标签图像拍摄时的 ISO 较高 (12800), 因此可以看到其含有较大的噪声. 在低照度图像增强问题中, 去除增强图像的噪声也是一项重要的内容, 因此使用 PSNR 和 SSIM 这类未考虑标签含有噪声情况的评价指标可能会对定量分析的客观性带来一定的影响. 为了更加客观对该任务进行定量分析, 本文进一步使用了 DISTS [37] 做为评价指标, 该指标可以在标签图像和增强图像存在纹理变化, 噪声等情况下仍保持较为合理的评价.



图 7 (网络版彩图) 标签图像含有噪声图像
 Figure 7 (Color online) The ground truth image with noise

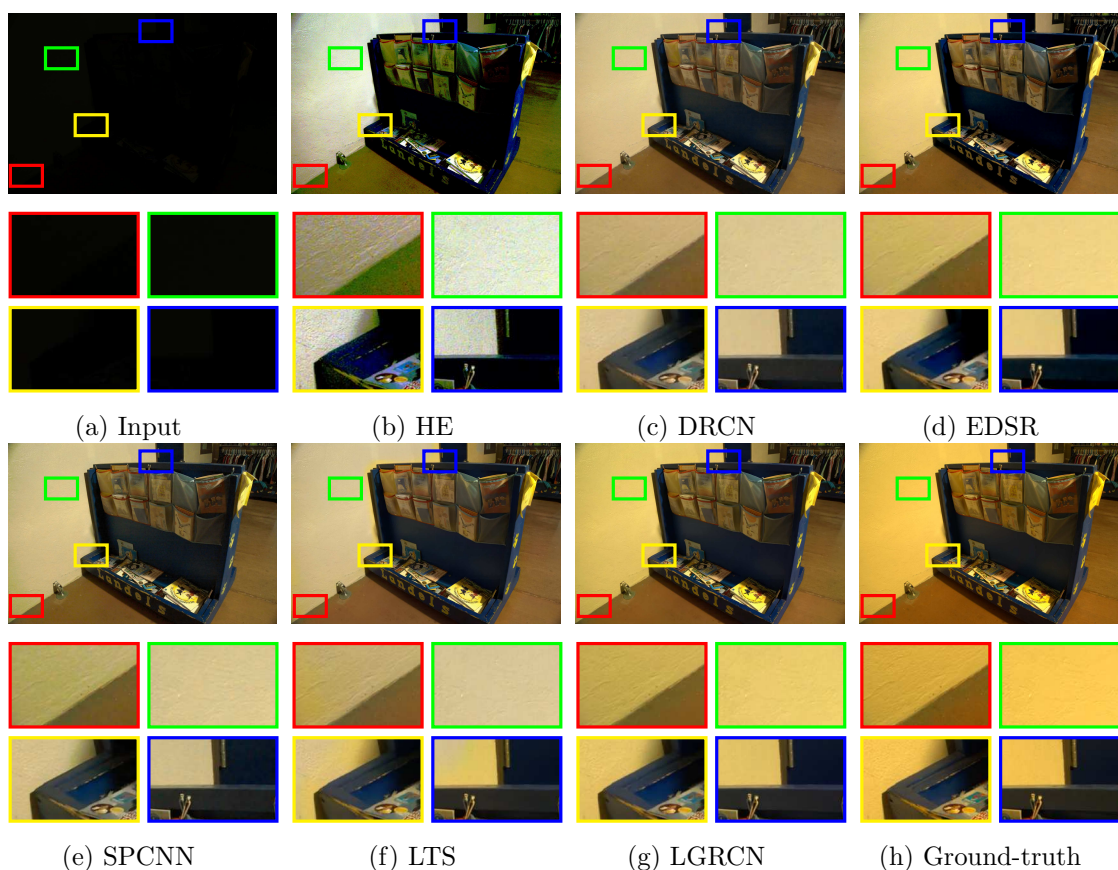


图 8 (网络版彩图) SIDS 数据集定性分析
 Figure 8 (Color online) Qualitative analysis on SIDS Sony Dataset

4.5 实验结果分析

为验证本文所提算法的合理性和有效性, 本文算法 (LGRCN) 和现有的一些图像增强算法分别在 SIDS-A 数据集 (RAW 数据)、SIDS-B 数据集 (RAW 数据)、SIDF 数据集 (RAW 数据) 和 LOL 数

据集 (RGB图像) 上进行了定量分析和定性分析. 表 1 展示了本文算法在 SIDS-A 数据集上与其它算法定量比较的结果, 表 2 展示了本文算法在 SIDS-B 数据集上与其它算法定量比较的结果, 表 3 展示了本文算法在 SIDF 数据集上与其它算法定量比较的结果, 表 4 展示了本文算法在 LOL 数据集上与其它算法定量比较的结果, 最优结果加粗表示. 在 SID Sony 数据集中, 输入的低照度图像曝光时间分为 0.04 秒、0.1 秒和 0.03 秒三种设置, 对应的长曝光标签图像的曝光时间为 10 秒. 为了充分验证本文算法的有效性, 本文分别对三种不同曝光时间的图像 ($\times 100$ 、 $\times 250$ 、 $\times 300$) 以及全部图像 (All) 进行了实验对比. 通过对实验结果分析可知, 本文算法在多个数据集上均取得了最优结果.

由表 1、表 2、和表 3 可知, 基于深度学习的图像增强方法与直方图均衡算法 (HE)、多曝光融合算法 (EFF) 等传统的图像增强方法相比, PSNR、SSIM 和 DISTs 的值均具有较大地提升. 由图 8 可知, 传统算法虽可以直接处理极低照度图像, 提升图像的亮度, 但其不能较好地抑制图像噪声, 恢复图像色彩信息, 因此对极低照度图像进行增强时, 图像质量的改善有限. 相比于其它基于卷积神经网络的图像增强方法 (DRCN、EDSR、SubPixelCNN、LTS 和 SCGA-Net), 本文算法同时考虑了图像特征的局部关系和全局关系, 具有更好的特征表达能力, 因此实验结果明显优于其它五种算法. 图 8(a) 为低照度的输入图像, 图 8(b) 为使用 HE 算法获得的增强图像, 图 8(c) 为使用 DRCN 算法获得的增强图像, 图 8(d) 为使用 EDSR 算法获得的增强图像, 图 8(e) 为使用 SPCNN 算法获得的增强图像, 图 8(f) 为使用 LTS 算法获得的增强图像, 图 8(g) 使用本文算法获得的增强图像, 图 8(h) 为标签图像. 由于图 8(a) 的成像环境光照条件较差且成像设备的曝光时间较短, 为了抑制图像的噪声, 成像设备设置的 ISO 值较低, 因此成像设备捕获的图像较黑. 从图 8 标注的四个方框中可以看到, 本文算法的增强图像和标签图像更加接近, 且较好地修正了色彩过渡不自然的问题, 由此可以验证局部-全局关系耦合的有效性. 特别的, 图 8(f) (LTS 算法) 与图 8(g) (LGRCN 算法) 相比, 虽使用的都是 U-net 结构, 但前者使用的全卷积的网络结构, 而本文 (LGRCN) 加入了局部-全局关系耦合. 从图 8(f) 可以看到, 在使用全卷积网络的结果中, 由于仅考虑了特征信息间的局部关系, 因此有明显色彩不连续现象, 特别是在墙壁部分其色彩过渡具有明显的断层感. 从图 8(g) 可以看到, 当使用了局部-全局关系耦合模块后, 上述问题有很好的改善. 如图 9 所示, 对于一些困难样本, 也就是由于拍摄时的光照强度很低或者曝光时间太短造成样本质量较差. 对于这样的样本在考虑了局部-全局关系耦合后的增强图像有较为显著的质量提升, 由此也进一步证明了局部-全局关系耦合的有效性和重要性.

由表 4 实验结果可以看到, 本文方法 (LGRCN) 与 HE、EFF^[32]、RetinexNet^[17]、Zero-DCE^[30] 以及 Kind^[31] 方法相比, 均具有较大的提升. 图 10 展示了本文算法与其他算法在 LOL 数据集上的定性对比结果, 其中图 10(a) 为低照度的输入图像, 图 10(b) 为 HE 算法处理后的增强图像, 图 10(c) 为 EFF 算法处理后的增强图像, 图 10(d) 为 RetinexNet 算法处理后的增强图像, 图 10(e) 为 Kind 算法处理后的增强图像, 图 10(f) 为 Zero-DCE 算法处理后的增强图像, 图 10(g) 为本文算法 (LGRCN) 处理后的增强图像, 图 10(h) 代表标签图像. 从图 10(g) 可以看到, 本文算法的增强结果色彩更加准确, 对噪声的抑制更加明显, 且图像细节更加清晰. 本文通过与其它算法在 SID 数据集和 LOL 数据集上的实验对比可知, 本文算法性能更优, 由此验证了本文算法的有效性和鲁棒性.

4.6 消融实验

为了进一步探索局部-全局关系耦合的重要性, 本文在 SID Sony 数据集对消融模型进行训练, 并在 SIDS-A 进行了测试. 同时从模型性能 (PSNR、SSIM、DISTs)、模型参数量 (Param)、模型计算量 (FLOPs) 以及推理时间 (Runtime) 四个方面对不同的消融方法进行了分析. 在计算模型的计算量以及推理时间时, 本文使用了 1024×1024 分辨率的 RAW 数据进行测试, 在推理时间的计算上为消



图 9 (网络版彩图) 局部-全局关系对增强图像质量的影响

Figure 9 (Color online) The effectness of local-global relationship on enhanced image quality

表 1 本文算法与其它算法在 SIDS-A 数据集上的实验结果对比

Table 1 Comparison results on SIDS-A Dataset

SIDS-A	× 100			× 250			× 300			All		
	PSNR↑	SSIM↑	DISTS↓	PSNR↑	SSIM↑	DISTS↓	PSNR↑	SSIM↑	DISTS↓	PSNR↑	SSIM↑	DISTS↓
HE	12.268	0.073	0.283	11.526	0.463	0.407	11.853	0.504	0.426	11.880	0.056	0.375
EFF [32]	14.633	0.288	0.323	13.587	0.216	0.429	13.587	0.210	0.484	13.917	0.236	0.417
DRCN [33]	28.035	0.789	0.154	26.430	0.763	0.222	26.993	0.735	0.243	27.142	0.760	0.208
SPCNN [34]	27.497	0.831	0.150	25.040	0.696	0.257	24.953	0.657	0.275	25.697	0.700	0.240
EDSR [35]	25.830	0.704	0.141	24.389	0.678	0.198	24.615	0.653	0.220	24.923	0.677	0.189
LTS [2]	29.597	0.810	0.091	28.004	0.790	0.138	28.489	0.765	0.158	28.683	0.787	0.130
LGRCN	30.742	0.814	0.081	29.056	0.795	0.126	28.854	0.767	0.150	29.506	0.790	0.121

除设备自身带来影响, 同时在相同的环境下重复了 10000 实验, 并计算其均值. 如表 5 所示, 消融实验一 (w/ Conv) 使用传统的卷积层替换本文提出的局部-全局耦合模块. 消融实验二 (w/ MHSA) 使用多头自注意力机制替换本文提出的局部-全局耦合模块. 消融实验三 (w/ DWConv) 使用深度可分离卷积替换本文提出的局部-全局耦合模块. 消融实验四 (w/ LGRCB) 使用本文提出的局部-全局耦合模块. 通过实验结果可以看到, 使用深度可分离卷积替换本文提出的局部-全局耦合模块的方法可以在参数量、计算量、运行时间上取得较优的结果, 然而在定量分析的结果中表现却较差, 相比于使用 LGRCB 的方法, 其 PSNR 值下降了 1.075 dB. 使用 LGRCB 的方法, 其在定量分析的结果中取得最好的结果, 同时在参数量、计算量、运行时间上也达到了较好的平衡. 通过以上的四组消融实验, 充分证明了局部-全局关系耦合的必要性.

为了对本文使用的全局关系构建方法的合理性进行更深入的研究, 接下来使用其它方法替换本文使用的方法, 这些方法包括 Transformer (TF) [24] (仅使用了单层编码器)、Non-local Attention [25] (N-

表 2 本文算法与其它算法在 SIDS-B 数据集上的实验结果对比

Table 2 Comparison results on SIDS-B Dataset

SIDS-B	× 100			× 250			× 300			All		
	PSNR↑	SSIM↑	DISTS↓	PSNR↑	SSIM↑	DISTS↓	PSNR↑	SSIM↑	DISTS↓	PSNR↑	SSIM↑	DISTS↓
HE	12.268	0.073	0.283	11.526	0.046	0.407	11.853	0.050	0.426	11.880	0.056	0.375
EFF ^[32]	14.633	0.288	0.323	13.587	0.216	0.429	13.597	0.210	0.484	13.916	0.236	0.417
DRCN ^[33]	28.509	0.790	0.159	26.801	0.762	0.229	27.020	0.732	0.244	27.410	0.760	0.213
SPCNN ^[34]	27.669	0.756	0.185	25.335	0.695	0.267	24.948	0.654	0.277	25.901	0.698	0.246
EDSR ^[35]	26.316	0.712	0.141	24.785	0.687	0.201	24.625	0.649	0.220	25.193	0.680	0.190
LTS ^[2]	29.944	0.807	0.096	28.238	0.787	0.145	28.357	0.761	0.160	29.031	0.786	0.134
SCGA-Net ^[36]	-	-	-	-	-	-	-	-	-	29.32	-	-
LGRCN	31.462	0.815	0.084	29.654	0.796	0.132	28.950	0.766	0.152	29.937	0.790	0.125

表 3 本文算法与其它算法在 SIDF 数据集上的实验结果对比

Table 3 Comparison results on SIDF Dataset

SIDF	× 100			× 250			× 300			All		
	PSNR↑	SSIM↑	DISTS↓	PSNR↑	SSIM↑	DISTS↓	PSNR↑	SSIM↑	DISTS↓	PSNR↑	SSIM↑	DISTS↓
DRCN ^[33]	26.798	0.741	0.168	25.147	0.671	0.257	24.105	0.671	0.257	25.580	0.694	0.226
SPCNN ^[34]	26.597	0.716	0.190	23.880	0.614	0.305	22.330	0.576	0.332	24.636	0.648	0.263
EDSR ^[35]	26.311	0.710	0.142	24.468	0.637	0.234	23.632	0.612	0.263	25.041	0.662	0.202
LTS ^[2]	27.895	0.757	0.117	25.893	0.684	0.195	24.710	0.654	0.221	27.374	0.709	0.168
SCGA-Net ^[36]	-	-	-	-	-	-	-	-	-	27.63	-	-
LGRCN	29.306	0.762	0.010	27.335	0.693	0.167	26.182	0.667	0.200	27.876	0.716	0.147

表 4 本文算法与其它算法在 LOL 数据集上的实验结果对比

Table 4 Comparison results on LOL Dataset

LOL	Metric	HE	EFF ^[32]	RetinexNet ^[17]	Zero-DCE ^[30]	Kind ^[31]	LGRCN
	PSNR (dB)↑	14.541	14.478	16.774	16.733	20.862	22.993
	SSIM↑	0.377	0.521	0.420	0.520	0.832	0.846
	DISTS↓	0.200	0.187	0.230	0.187	0.112	0.092

表 5 本文方法消融实验的实验结果

Table 5 Ablation study on SID Sony Dataset

SIDS-B	Metric	w/ Conv	w/ MHSA	w/ DWConv	w/ LGRCB
	PSNR (dB)↑	29.031	29.451	28.862	29.937
	SSIM ↑	0.786	0.786	0.784	0.790
	DISTS↓	0.134	0.134	0.136	0.125
	Params (M)↓	7.761	6.189	5.406	5.865
	FLOPs (G)↓	54.978	53.368	52.567	53.036
	Runtime (ms)↓	8.986	9.324	8.803	9.250

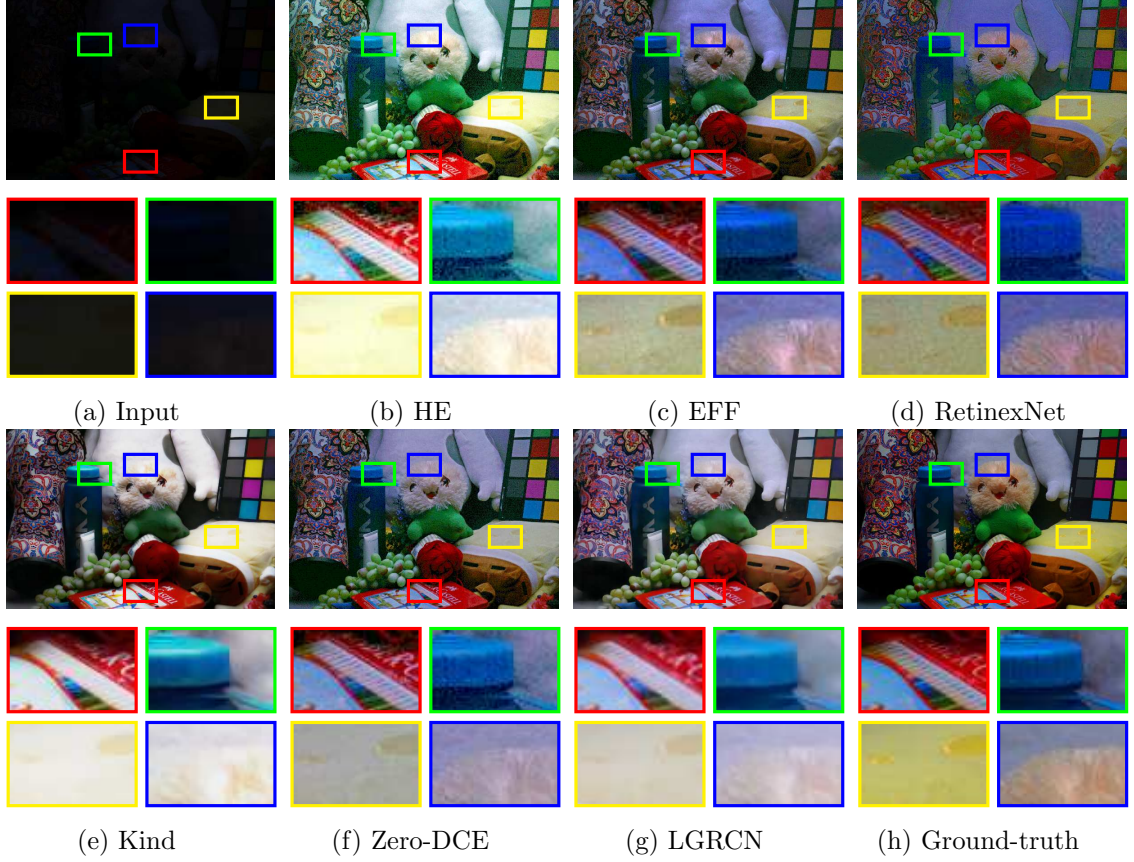


图 10 (网络版彩图) LOL 数据集定性分析

Figure 10 (Color online) Qualitative analysis on LOL Dataset

LA)、Global Self-Attention^[38] (GSA)、Criss-Cross Attention^[39] (CCA). 如表 6 所示, 其中 MHSA 代表本文使用的多头自注意力机制, $A \rightarrow B$ 代表使用 B 替换 A. 本文使用的全局关系构建方法在模型性能, 模型复杂度和推理时间上取得了较好的平衡.

表 6 使用不同全局关系建立方法的实验对比

Table 6 Comparison results using different methods of establishing global relationships

SIDS-B	Metric	MHSA	MHSA \rightarrow TF	MHSA \rightarrow NLA	MHSA \rightarrow GSA	MHSA \rightarrow CCA
	PSNR (dB) \uparrow	29.937	29.798	29.787	29.847	29.687
	SSIM \uparrow	0.790	0.787	0.787	0.787	0.786
	DISTS \downarrow	0.125	0.130	0.130	0.130	0.130
	Params (M) \downarrow	5.865	8.296	5.799	5.930	5.749
	FLOPs (G) \downarrow	53.036	55.518	52.969	53.103	52.918
	Runtime (ms) \downarrow	9.250	10.920	9.046	9.227	9.484

如表 7 所示, 本文对局部-全局关系耦合模块 (LGRCB) 的添加位置对局部-全局关系耦合神经网络 (LGRCN) 性能的影响做了进一步的研究, 其中 n^{th} CB 代表在第 n 个卷积块的位置替换该层卷积层. 从表中可以看到局部-全局关系耦合模块在 LGRCN 的 4th CB 时, LGRCN 获得了最优的 S-

SIM 值, 在 LGRCN 的 7th CB 时, 获得了最优的 PSNR、DISTS 值, 但是此时模型的参数量较高分别是 7.288 M 和 7.643 M. 而在 5th CB 时, 虽然此时 LGRCN 在 PSNR、SSIM 和 DISTS 上的值比 4th CB 和 7th CB 略微低一些, 但是此时 LGRCN 的参数量却比这两个位置低很多. 因此, 本文的默认方案是在该位置使用 LGRCB.

表 7 LGRCB 在 LGRCN 不同位置的实验结果

Table 7 Experimental results using different methods of establishing global relationships

SIDS - B	Metric	1 st CB	2 nd CB	3 rd CB	4 th CB	5 th CB	6 th CB	7 th CB	8 th CB	9 th CB
	PSNR (dB) ↑	29.051	29.690	29.848	29.891	29.937	29.894	30.069	29.885	15.693
	SSIM ↑	0.782	0.787	0.788	0.791	0.790	0.790	0.789	0.785	0.379
	DISTS ↓	0.139	0.130	0.126	0.124	0.125	0.125	0.121	0.121	0.510
	Params (M) ↓	7.754	7.732	7.643	7.288	5.865	7.288	7.643	7.732	7.754
	FLOPs (G) ↓	52.890	52.863	52.858	52.888	53.036	52.888	52.858	52.863	52.890
	Runtime (ms) ↓	10.563	9.569	9.177	9.268	9.250	9.289	9.170	9.632	10.601

4.7 扩展实验

为了进一步验证局部-全局关系耦合模块的可扩展性, 本文分别在图像超分辨率任务和图像去雨任务对本文所提方法进一步的进行了验证. 对于图像超分辨率任务, 本文采用 IMDN^[40] 作为实验的基准网络, 通过在该网络上集成局部-全局关系耦合模块对本文方法进行验证, 并将修改后的方法记为 IMDN-LGRC. 按照 IMDN^[40] 文中的实验设计, 本文在 DIV2K 数据集对原模型和修改后的模型进行了训练, 并在 Set5 数据集进行了测试. 对于图像去雨任务, 本文采用 PRN^[41] 作为实验的基准网络, 通过在该网络上集成局部-全局关系耦合模块对本文方法进行验证, 并将修改后的方法记为 PRN-LGRC. 按照 PRN^[41] 文中的实验设计, 本文在 Rain100H 数据集的训练集对两个模型进行重新训练, 并在 Rain100H 数据集的测试集进行了测试. 如表 8 所示, 通过添加局部-全局关系耦合模块可在不同程度上提升基准网络的性能, 并且不会明显增加基准网络的参数量. 进一步的, 本文在去雨任务上进行了定性分析. 如图 11 所示, 在对 PRN^[41] 增加局部-全局关系耦合模块后可以有效提升去雨图像的视觉质量, 特别是图中标框的区域, 可以看到 PRN^[41] 的去雨图像仍存在一定程度的雨痕, 然而在增加了局部-全局关系耦合模块后可以很好的修复这些问题.

表 8 本文方法在其它任务的实验结果

Table 8 The results of our method on other tasks

Task	Dataset	Scale	Method	PSNR (dB) ↑	SSIM ↑	Params (M) ↓
SR	Set5	×2	IMDN ^[40]	37.981	0.960	0.694
			IMDN-LGRC	38.083	0.961	0.740
		×3	IMDN ^[40]	34.355	0.927	0.703
			IMDN-LGRC	34.418	0.927	0.749
		×4	IMDN ^[40]	32.097	0.893	0.715
			IMDN-LGRC	32.281	0.896	0.761
Derain	Rain100H	-	PRN ^[41]	28.083	0.884	0.095
		-	PRN-LGRC	28.347	0.886	0.107

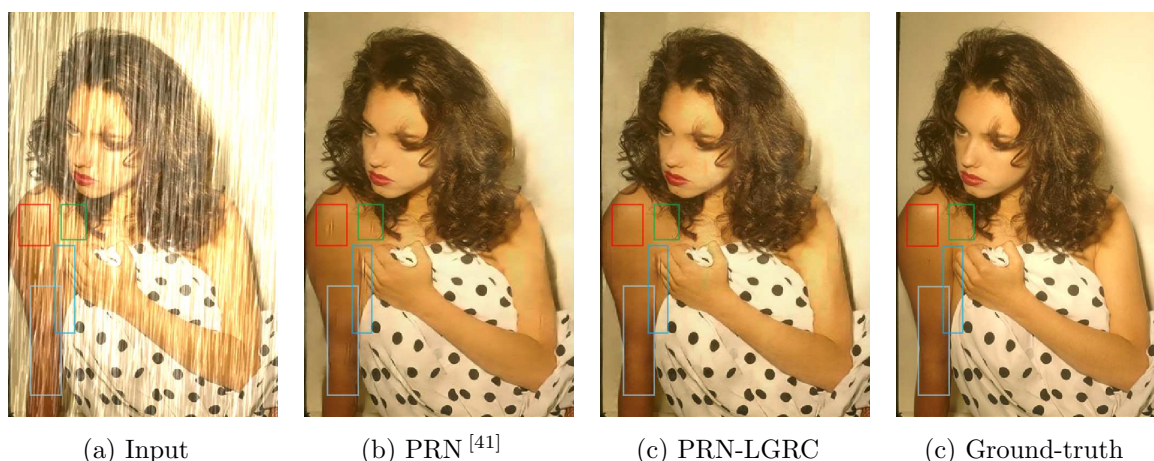


图 11 (网络版彩图) RPN 和 RPN-LGRC 图像去雨的结果
Figure 11 (Color online) Image Deraining results by RPN and RPN-LGRC

5 总结

本文针对现阶段卷积神经网络存在不足之处, 基于局部-全局关系耦合的思想, 提出了新型的低照度图像增强算法, 该方法分别利用深度卷积和多头自注意力机制获取特征信息的局部关系和全局关系, 并通过特征耦合模块对具有局部关系的特征和具有全局关系的特征进行有效耦合. 本文方法弥补了传统卷积神经网络以及多头自注意力机制获取特征信息的局限性, 实现图像中不同关系的特征信息互补, 提升特征信息的丰富程度, 进而增强卷积神经网络的性能. 本文通过大量的对比实验验证了所提算法的有效性, 并通过消融实验验证了耦合局部关系和全局关系的必要性以及本文全局关系构建方式的合理性.

特别地, 本文所提出的局部-全局关系耦合模块是一种即插即用的结构, 可以适用于多种网络结构以及任务中. 在扩展实验中, 本文在图像超分辨率任务和图像去雨任务对本文方法的通用性进行了验证, 通过在基准网络上简单的添加局部-全局关系耦合模块, 可以为基准网络带来不同程度性能的提升. 在未来的工作中, 我们会继续对局部-全局关系耦合方法进行更深入的研究, 在更多的任务上探索局部关系与全局关系耦合的可行性.

参考文献

- 1 Ren S Q, He K M, Girshick R B, et al. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*, 2017, 39: 1137 - 1149
- 2 Chen C, Chen Q, Xu J, et al. Learning to see in the dark. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 3291-3300
- 3 Gao P, Jiang Z, You H, et al. Dynamic fusion with intra- and inter-modality attention flow for visual question answering. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019. 6632-6641
- 4 Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks. In: *Proceedings of Advances in Neural Information Processing Systems*, 2012. 1097-1105
- 5 He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 770-778
- 6 Zhang X B, Lin Z Y, Tian W X, et al. Mask-wearing recognition in the wild. *Sci Sin Inform*, 2020, 50:1110 - 1120 [张修宝, 林子原, 田万鑫, 等. 全天候自然场景下的人脸佩戴口罩识别技术. *中国科学: 信息科学*, 2020, 50: 1110 - 1120]

- 7 Alaparthi S, Mishra M. Bidirectional Encoder Representations from Transformers (BERT): A sentiment analysis odyssey. 2020. ArXiv: 2007.01127
- 8 Li X C, Zhan D C. A semantic relation preserved word embedding reuse method. *Sci Sin Inform*, 2020,50: 813 - 823 [李新春, 詹德川. 一种保持语义关系的词向量复用方法. *中国科学: 信息科学*, 2020, 50: 813 - 823]
- 9 Li D, Zhang W S. Attentive pooling for group activity recognition. *Sci Sin Inform*, 2021, 51: 399 - 412 [李定, 张文生. 面向群体行为识别的注意力池化机制. *中国科学: 信息科学*, 2021, 51: 399 - 412]
- 10 Kim Y T. Contrast enhancement using brightness preserving bi-histogram equalization. *IEEE Trans Consum Electr*, 1997, 43: 1-8
- 11 Pisano E D, Zong S, Hemminger B M, et al. Contrast limited adaptive histogram equalization image processing to improve the detection of simulated spiculations in dense mammograms. *J Digit Imaging*, 1998, 11: 193-200
- 12 Land E H. The retinex theory of color vision. *Sci Am*, 1977, 237: 108-129
- 13 Land E H. An alternative technique for the computation of the designator in the retinex theory of color vision. In: *Proceedings of the National Academy of Sciences of the United States of America*, 1986. 3078-3080
- 14 Jobson D J, Rahman Z U, Woodell G A. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Trans on Image Process*, 1997, 7: 965-976
- 15 Fu X Y, Zeng D L, Huang Y, et al. A weighted variational model for simultaneous reflectance and illumination estimation. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 2782-2790
- 16 Park S, Yu S, Moon B, et al. Low-light image enhancement using variational optimization-based retinex model. *IEEE Trans Consum Electr*, 2017, 63: 178-184
- 17 Wei C, Wang W J, Yang W H, et al. Deep retinex decomposition for low-light enhancement. 2018. ArXiv:1808.04560
- 18 Jiang G X, Wang W J, Qian Y H, et al. A Unified Sample Selection Framework for Output Noise Filtering: An Error-Bound Perspective. *J Mach Learn Res*, 2021, 22: 1-66
- 19 Cheng H H, Qian Y H, Hu Z G, et al. Association mining method based on neighborhood perspective. *Sci Sin Inform*, 2020, 50: 824 - 844 [成红红, 钱宇华, 胡治国, 等. 基于邻域视角的关联关系挖掘方法. *中国科学: 信息科学*, 2020, 50: 824 - 844]
- 20 Li F J, Qian Y H, Wang J T, et al. Clustering ensemble based on sample's stability. *Artif Intell*, 2019, 273: 37-55
- 21 Buades A, Coll B, Morel J. A non-local algorithm for image denoising. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2005. 60-65
- 22 Dabov K, Foi A, Katkovnik V, et al. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Trans on Image Process*, 2007, 16: 2080-2095
- 23 Glasner D, Bagon S, Irani M. Super-resolution from a single image. In: *Proceedings of IEEE Conference on Computer Vision*, 2009. 349-356
- 24 Vaswani A, Shazeer N, Parmar N, et al. Attention is All you Need. In: *Proceedings of Advances in Neural Information Processing Systems*, 2017. 6000-6010
- 25 Wang X, Girshick R, Gupta A, et al. Non-local neural networks. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 7794-7803.
- 26 Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale. 2020. ArXiv: 2010.11929
- 27 Zhang D, Zhang H W, Tang J H, et al. Feature pyramid transformer. In: *Proceedings of European Conference on Computer Vision*, 2020. 323-339
- 28 Zhu X Z, Wu W J, Lu L W, et al. Deformable DETR: Deformable transformers for end-to-end object detection. 2020. ArXiv: 2010.04159
- 29 Ba J L, Jamie R K, Geoffrey E H. Layer normalization. 2016. ArXiv: 1607.06450
- 30 Guo C L, Li C Y, Guo J C, et al. Zero-reference deep curve estimation for low-light image enhancement. In: *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2020. 1777-1786
- 31 Zhang Y H, Zhang J W, Guo X J. Kindling the darkness: A practical low-light image enhancer. In: *Proceedings of the 27th ACM International Conference on Multimedia*, 2019. 1632-1640
- 32 Ying Z Q, Li G, Ren Y R, et al. A new image contrast enhancement algorithm using exposure fusion framework. In: *Proceedings of the 17th International Conference on Computer Analysis of Images and Patterns*, 2017. 36-46
- 33 Kim J, Lee J K, Lee K M. Deeply-recursive convolutional network for image super-resolution. In: *Proceedings of*

- IEEE Conference on Computer Vision and Pattern Recognition, 2016. 1637-1645
- 34 Shi W Z, Caballero J, Huszar F, et al. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2016. 1874-1883
- 35 Lim B, Son S Y, Kim H, et al. Enhanced deep residual networks for single image super-resolution. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2017. 1132-1140
- 36 Ren D D, Li J B, Han M, et al. SCGA-Net: Skip Connections Global Attention Network for Image Restoration. Computer Graphics Forum, 2020, 39: 507-518
- 37 Ding K Y, Ma K D, Wang S Q, et al. Image Quality Assessment: Unifying Structure and Texture Similarity. IEEE Trans Pattern Anal Mach Intell, 2020
- 38 Shen Z R, Bello I, Vemulapalli R, et al. Global Self-Attention Networks for Image Recognition. 2020. ArXiv: 2010.03019
- 39 Huang Z L, Wang X G, Wei Y C, et al. CCNet: Criss-Cross Attention for Semantic Segmentation. IEEE Trans Pattern Anal Mach Intell, 2020
- 40 Hui Z, Gao X, Yang Y, et al. Lightweight image super-resolution with information multi-distillation network. In: Proceedings of the 27th ACM International Conference on Multimedia, 2019. 2024-2032
- 41 Ren D W, Zuo W M, Hu Q H, et al. Progressive image deraining networks: A better and simpler baseline. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2019. 3932-3941

Local-Global Coupling Relationship based Low-light Image Enhancement

Keqi WANG^{1,3}, Yuhua QIAN^{1,2*}, Jiye LIANG², Chang LIU¹, Qin HUANG¹, Lu CHEN¹ & Jieru JIA¹

1. *Institute of Big Data Science and Industry, Shanxi University, Taiyuan 03006, China;*

2. *Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, Shanxi University, Taiyuan 030006, China;*

3. *Pangole Deep Intelligence, Taiyuan 030006, China*

* Corresponding author. E-mail: jinchengqyh@sxu.edu.cn

Abstract Convolutional neural networks have made significant progress in various fields of artificial intelligence. However, the convolution operation which is based on shared parameter sliding window mechanism only focuses on modeling the local relationship and is insufficient to establish global relationship. Nonetheless, the local and global relationships are both crucial for feature representation. Therefore, this paper concentrates on how to efficiently construct and couple the local and global relationships to mine more abundant feature information and enhance the discriminability of features. To this end, a local-global coupling module is designed, which is composed of four basic components: feature extraction, local relationship branch based on Depth-wise Convolution (DWConv), global relationship branch based on Multi-head self-attention (MHSA), and local-global relationship coupling based on Point-wise convolution (PWConv). Grounded on this, a local-global relationship coupling with encoder-decoder structure neural network is proposed, which can efficaciously model the local-global coupling relationship, strengthen the discriminative ability of feature information, and improve the performance of the model. Extensive experiments on low-light image enhancement datasets demonstrate the effectiveness of the proposed method, which significantly outperforms the state-of-the-art counterparts. Furthermore, the importance and effectiveness of local-global relationship coupling is analyzed through ablation and extended experiments.

Keywords low-light image enhancement, local-global relationship, computer vision, convolution neural network, deep learning



Keqi WANG was born in 1992. Master. His main research interests include deep learning, computer vision and low-light image enhancement.



Yuhua QIAN was born in 1976. He received his M.S. and Ph.D. degrees in computers with applications at Shanxi University in 2005 and 2011, respectively. He is a professor at the Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education, China. He is actively pursuing research in pattern recognition, computer vision, machine learning, granular computing, and artificial intelligence.



Jiye LIANG was born in 1962. He received the M.S. and Ph.D. degrees from Xi'an Jiaotong University, Xi'an, China, in 1990 and 2001, respectively. He is a professor at School of Computer and Information Technology and Key Laboratory of Computational Intelligence and Chinese Information Processing of Ministry of Education at Shanxi University. His current research interests include computational intelligence, granular computing, data mining, and

knowledge discovery.



Chang LIU was born in 1996. She is a postgraduate student at the institute of Computer and Information Technology, Shanxi University. Her main research interests include image enhancement and deep learning.