

本科生论文设计方向

Ku Jui

October 2023

目录

I Pre-Knowledge	1
1 背景介绍	1
2 方向	2
2.1 方向一	3
2.2 方向二	4
2.3 方向三	4
3 创新想法	5
3.1 创新想法的支撑知识点	6

Part I

Pre-Knowledge

1 背景介绍

弱光图像增强 (Low-light image enhancement, LLIE) 是图像处理中的一个重要任务, 其目标是提升在低光环境下拍摄的图像的感知质量。这个领域的近期进展主要由深度学习方法主导, 包括不同的学习策略、网络架构、损失函数和训练数据。

低光图像增强在不同领域享有广泛的应用, 包括视觉监控、自动驾驶和计算摄影。特别是, 智能手机摄影已经变得无处不在和突出。受限于相机光圈的大小、实时处理的要求以及存储器的约束, 在昏暗环境中用智能手机的相机拍摄照片尤其具有挑战性。

用于低光增强的传统方法包括基于直方图均衡的方法和基于 Retinex 模型的方法。然而, 这些方法存在一些局限性, 例如在 Retinex 模型中通常忽略噪声, 因此在增强结果中保留或放大噪声; 找到有效的先验或正则化是具有挑战性的, 不准确的先验或正则化可能导致增强结果中的伪像和颜色偏差; 由于其复杂的优化过程, 运行时间相对较长。

近年来基于深度学习的 LLIE 取得了引人注目的成功。基于深度学习的解决方案比传统方法具有更好的准确性、鲁棒性和速度, 因此越来越受到关注。然而, 现有的低光照图像增强技术聚焦于构建数据驱动的深

度网络, 通常其网络模型复杂, 导致计算效率低、推理速度慢, 并且由于对于训练数据分布的依赖性导致其在未知场景下的性能缺乏保障。

为了解决这些问题并推动该领域的发展, 研究者们提出了一些新颖的方法和工具。例如, 他们提出了一个大尺度低光图像与视频数据集, 并开发了一个包含多种主流 LLIE 方法的在线平台。这些工具可以帮助研究者和开发者更好地理解和改进现有技术, 并为未来研究提供宝贵资源。

2 方向

根据目前的文献研究, 在 LLIE 任务中, 一种方法是采用边缘检测思想来引导弱光图像增强¹, 其策略是将边缘图与初步恢复的图像进行串联, 然后通过单一的增强网络来实现最终的图像恢复 (见Fig. 1)。恢复结果的质量往往会受到最后的增强网络设计和初步恢复图像所包含特征与真实图像特征之间的相似度的影响。

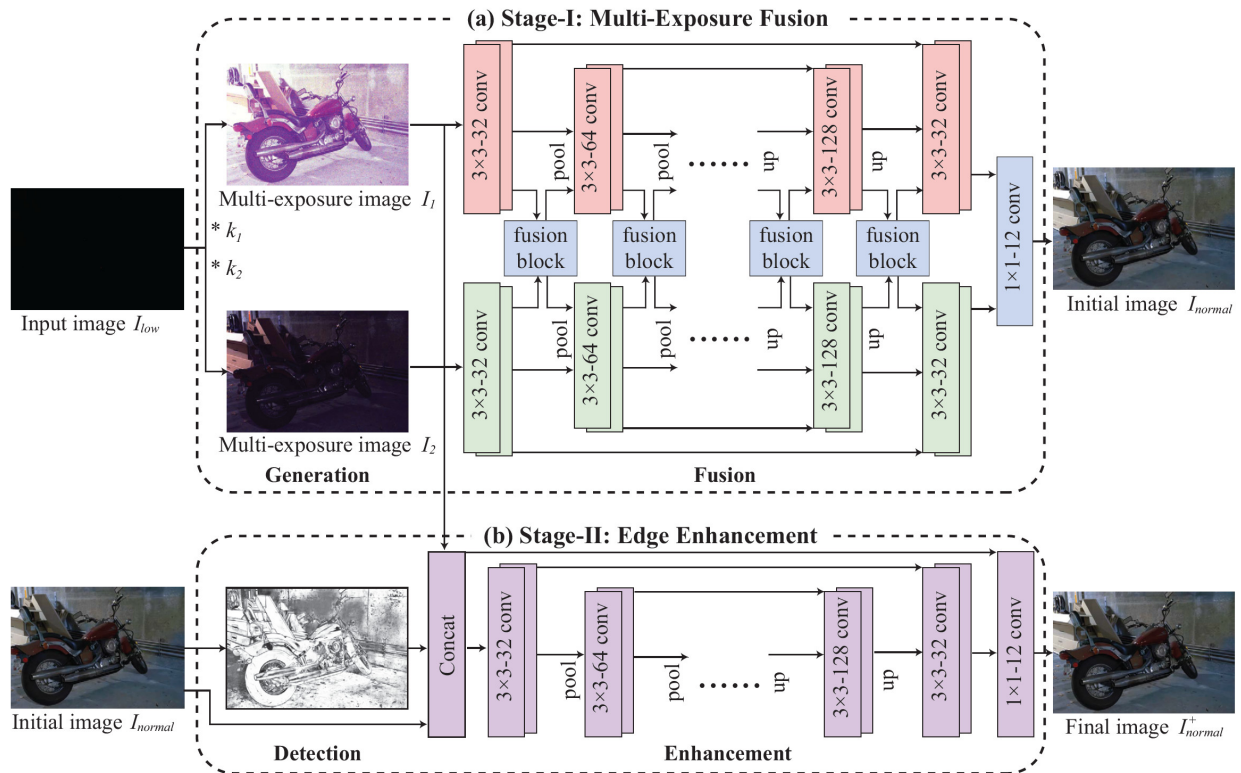


Figure 1: 该 LLIE 结构源自 [1], 如其 Multi-Exposure Fusion 部分采用多曝光融合结构, 与由 Initial image 生成的边缘图进行 Concat, 后续通过一个 U-Net 网络进一步恢复图像。

根据目前的文献调研, 当前的此类方法的特点只集中在通过边缘图指导图像恢复, 且恢复出来的图像还是有明显噪点。

¹为什么要采用图像的边缘图片来指导弱光图像恢复? 首先, 图像边缘图仅需要很少的数据量, 其剔除了不相关的信息, 保留了图像重要的结构属性。其次, 低光图像通常具有很高的对比度, 现有的方法可能无法在极暗或极亮的区域恢复图像边缘细节。同时, 当物体边缘不清晰时, 像素级损失往往会模糊边缘, 破坏图像细节。而加入边缘先验可以降低优化外观重构时的不适定程度。最后, 人类视觉系统对边缘高度敏感, 保留结构信息对图像重建任务的性能至关重要。定义边缘可以通过学习区分黑暗区域的不同物体来指导增强过程, 而不仅仅是识别低光区域。通过保留图像中的结构属性, 这使得物体之间的可见性更好。

2.1 方向一

图像初步恢复的方法目前主要包括以下几种：

- (1) 利用多曝光技术²获得具有不同曝光度的图像，然后结合恢复网络进行初步恢复³；
- (2) 我们提出的 CNN 和 Transformer 并行架构（见Fig. 2）。
- (3) 直接通过神经网络进行弱恢复。

但由于噪声和伪影的存在，以及色彩与真实图像之间的差异，后者结果通常并不理想。

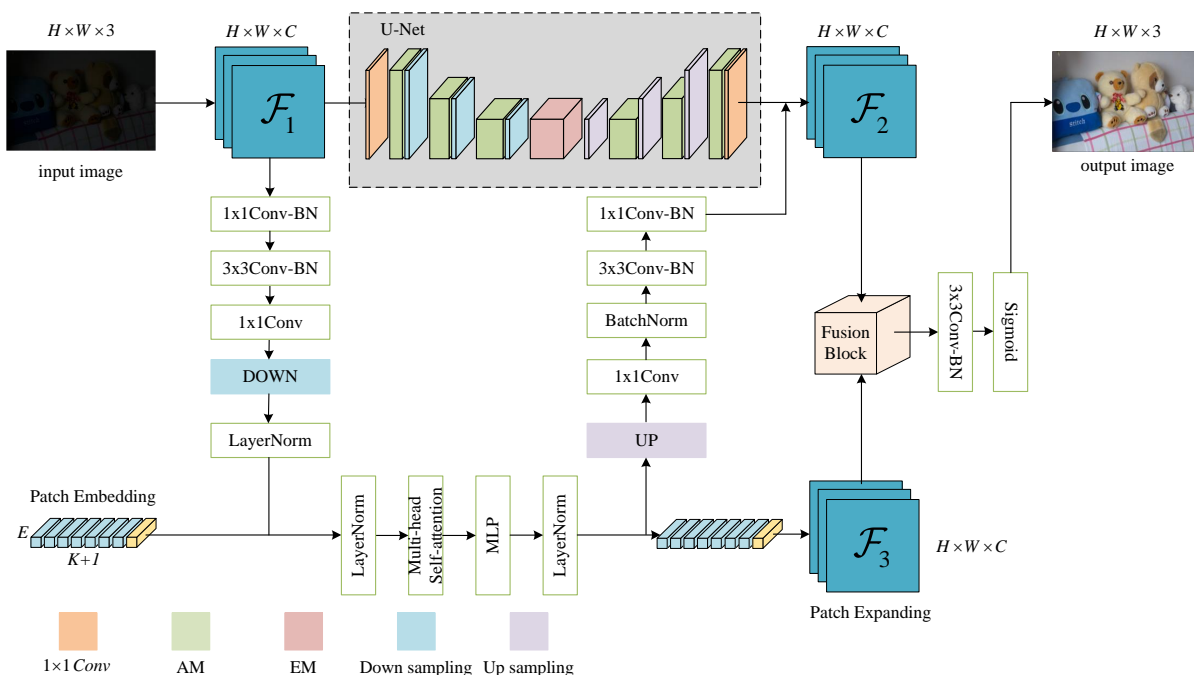


Figure 2: Low-Lighting 图像恢复网络。结构受 Conformer [2]（见 Fig. 4）启发，将原来结构中 CNN 分支的 ResNet 结构修改为 U-Net 结构。其采用一个 U-Net 和 ViT 的并行架构，通过 U-Net 结构得到一个弱恢复的弱特征图 \mathcal{F}_2 ，通过 ViT 融合的特征可以初步增强弱特征图 \mathcal{F}_2 ，ViT 的输出经过 Patch Expanding 的特征图 \mathcal{F}_3 经过与 U-Net 输出的特征图 \mathcal{F}_2 融合之后得到一个初步恢复的图片 output image，用以后续参与图片的进一步恢复。

要求构建上述两种方法中的一种模型（或采用第三种方法，但结果要求需要优于 SOTA），训练和测试数据采用公开数据集 **LOL-v2** 和 **Imagenet** 数据集，进行消融实验（去除 ARMB 或者整体去除 ViT），可以将其与 Uformer 架构进行对比。

²多曝光技术已被实验证明可以有效的解决 LLIE 任务中普遍存在的色彩失真问题。

³常见的架构如基于 GAN 的多曝光图像融合架构，基于端到端深度学习框架下多曝光图像融合算法，基于稀疏表示和可平移复方向金字塔变换的多曝光融合框架。

2.2 方向二

此外，获取边缘图的方法也各不相同，在 LLIE 任务中，常见方法采用**边缘检测网络**来生成边缘图。Canny 和 Sobel 算子通常被用作对比实验，也有一些研究采用辅助边缘网络以提高边缘图生成的准确性。这些方法在弱光图像恢复研究中都具有一定的学术和实际应用价值。目前在 LLIE 领域中获取边缘图分化成了两个不同的方向，一种是采用初步恢复图像获取边缘图（如Fig. 1），另一种是直接采用弱光图像获取边缘图。

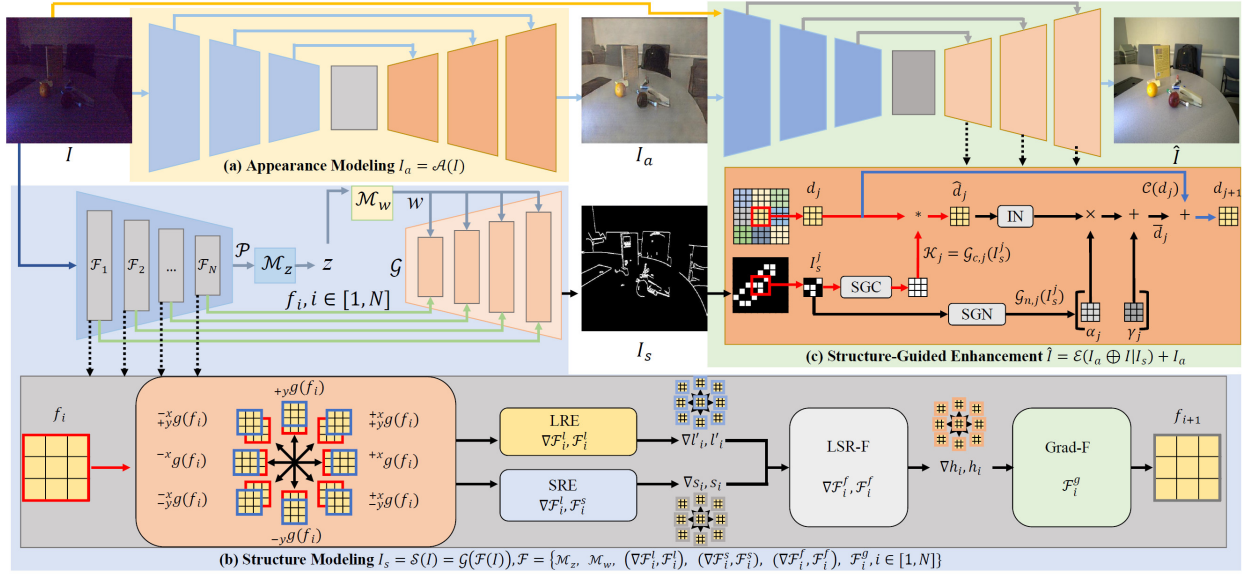


Figure 3: 该结构的边缘图直接从 I 中通过 StyleGAN 得到边缘图 I_s ，而非从 I_α 中通过边缘网络获取边缘图。

图像处理任务中边缘检测的各种方法：

- (1) 手工设计各种滤波器来生成边缘图。
- (2) 根据人类设计的特征使用数据驱动模型来预测边缘（随机决策森林来学习边缘斑块）
- (3) 深度学习方法从原始数据中学习复杂的特征表示端到端边缘检测模型。

要求通过深度学习方法构建一种边缘网络，若是上述提到的第一个方向，则该网络是无监督的，若是上述提到的第二个方向，则该网络是有监督的，可以使用 Canny 或 Sobel 算子作为先验。

2.3 方向三

在增强网络的设计方面，主要采用了两种主要架构：StyleGAN 和 U-Net。这些网络通常通过串联操作或将边缘图像的分块作为卷积核的方式来引导图像恢复过程。关键在于边缘图如何为初步恢复图片提供先验增强。

可以使用直接使用方向一的架构，如果选择方向三，边缘图可以通过 Canny 算子或 Sobel 算子获取，整个模型的输入为弱光图像和边缘图像。此外，需要对模型进行消融实验。

3 创新想法

在目前的研究中，我们探索了不同的边缘图生成方法，并引入了基于 VGG 网络的方法，同时采用 Canny 边缘检测作为参考对比方法。我们的主要目标是设计一种综合恢复策略，结合了以下几个关键元素：原始图像的 Y/Cb/Cr 通道信息、边缘图以及初步恢复图。

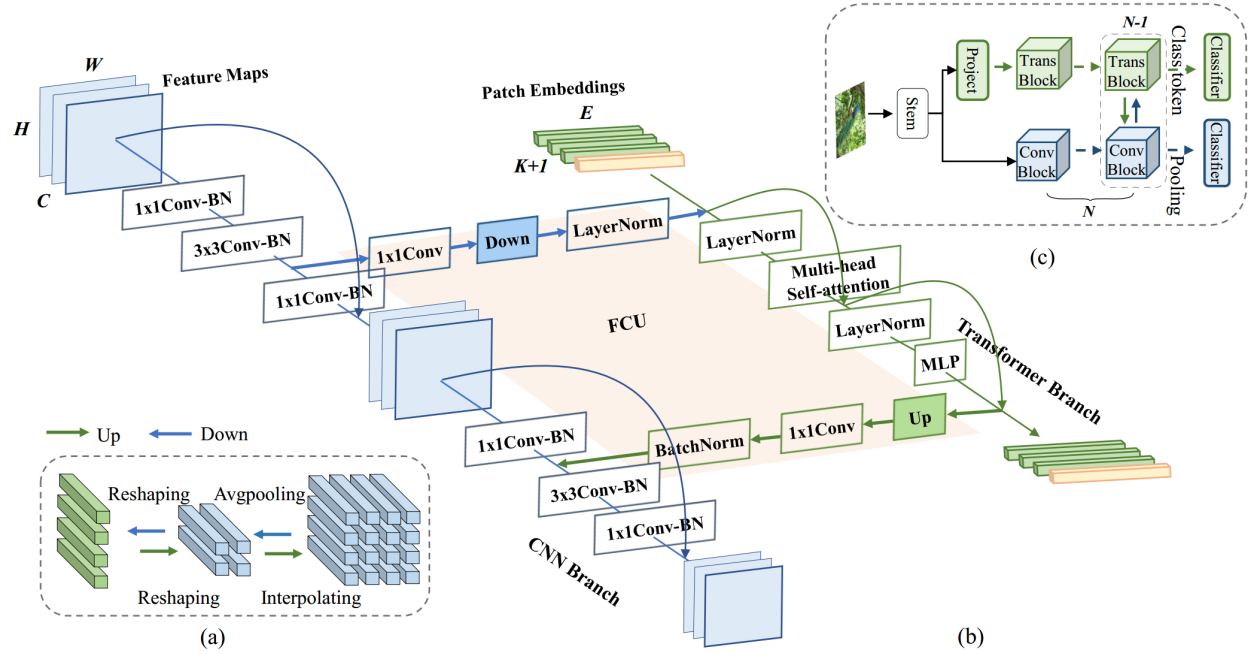


Figure 4: Network architecture of the proposed Conformer. (a) Up-sampling and down-sampling for spatial alignment of feature maps and patch embeddings. (b) Implementation details of the CNN block, the transformer block, and the Feature Coupling Unit (FCU). (c) Thumbnail of Conformer.

在初步恢复图的生成方面，我们受到了 Conformer [2] 结构的启发，在卷积神经网络 (CNN) 中，卷积操作擅长提取局部特征，但难以捕获全局表示。Conformer 是一种混合网络结构，包含了并行的 CNN 分支和 Transformer 分支，其通过特征耦合模块融合局部与全局特征，目的在于不损失图像细节的同时捕捉图像全局信息。我们借鉴 Conformer 并发结构 (见 Fig. 4)，将其中 CNN 分支的 ResNet 结构修改为 U-Net 结构，我们的设计采用了采用了一种复合型卷积神经网络 CNN 与 Transformer 网络的混合模型 (见 Fig. ??)。这一模型旨在从弱光图像中提取长距离和短距离特征，以更全面地捕捉图像的信息。通过特定的特征融合策略，我们期望能够得到初步恢复的图像。

至于增强网络的设计，我们目前尚未确定具体的方法。然而，我们的研究侧重于构建一个整体恢复框架，该框架利用边缘图、初步恢复图和通道信息的融合，以提高弱光图像的质量。我们将继续探索合适的增强网络结构，以进一步提升恢复结果的性能。这一研究旨在为弱光图像恢复领域提供新的理论和方法，以改善图像质量并拓展应用领域。

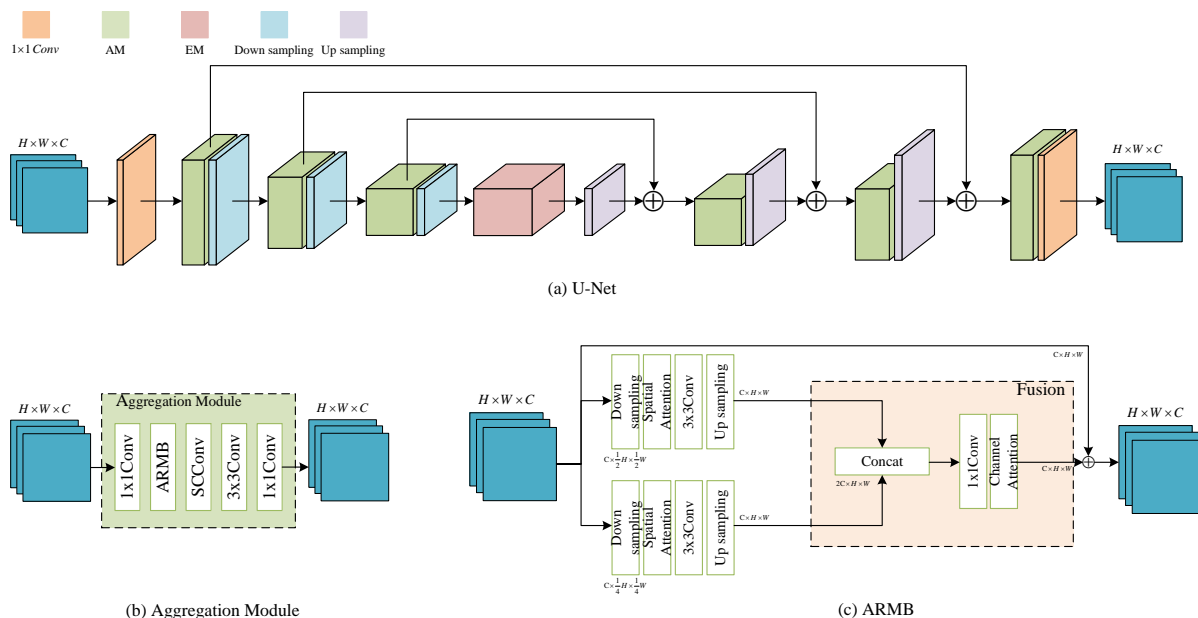


Figure 5: U-Net 图像初步恢复网络及其所属的模块。

3.1 创新想法调研支撑

Attention in CV

在计算机视觉领域，研究人员使用了不同类型的注意力机制，目的是使模型聚焦于图像本身。这大致分为两种类型：一种是基于通道的注意力机制，即 Squeeze-and-Excite。另一个是基于空间的注意力机制 [3]。然而，研究发现这些类型有三个局限性：

- (1) 卷积捕捉远距离特征的能力较差；
- (2) 注意力机制关注所有输入；当分辨率较大时，计算成本增加；
- (3) 不同时关注通道和空间维度关注信息。这就造成了信息的浪费和注意力的减弱。

针对第一个限制 [4] 将所有卷积替换为独立的自注意层和空间意识的独立自注意层，替换后的模型为纯注意模型。纯注意模型增强了网络对远距离特征连接的建模能力。在 ImageNet 分类任务和 COCO 对象检测任务中，该模型优于使用卷积的基线模型。对于第三个限制，Woo 等人 [3] 提出了一种卷积块注意模型 CBAM，这是一种轻量级的通用注意模型。该模型从通道维度和空间维度两方面关注重要信息。由于 CBAM 从多个维度提取重要信息，该模型被广泛应用于目标检测、目标分类等任务。实验表明，CBAM 模块可以显著提高网络的表达能力。同时，CBAM 可以搭配 SCConv [5] 结构，SCConv 是一种名为可以即插

即用的卷积模块，目的是减少卷积神经网络中特征之间的空间和通道冗余，从而压缩 CNN 模型并提高其性能。

U-Net for LLIE

U-Net 网络由卷积层、下采样、上采样和跳过连接操作组成，其编码器和解码器具有对称结构。长期以来，U-Net 网络及其改进结构由于其强大的特征学习和特征重构能力，在图像语义分割领域取得了巨大的成功。U-Net 网络在弱光图像增强方面显示出出色的效果。Chen 等 [6]创新性地提出了一种基于 U-Net 框架的弱光图像增强网络。该网络的核心架构是多尺度上下文聚合网络 (CAN) 和 U-Net 网络。之后的工作 [6,7]也将 U-Net 架构应用到自己的网络中，其网络包含两个子网。即图像约简子网和感知损失子网，其中图像约简子网具有与 [6]相同的 U-Net 架构。虽然 [6,7]在使用U-Net架构增强图像方面取得了一定的效果，但 [8]认为 [6,7]网络忽略了全局信息，导致增强图像中的颜色不自然和物体伪影问题。为了改善以往网络的缺陷，利用 U-Net 架构，[8]在U-Net的编码器和解码器之间增加了全局信息感知模块(global information awareness module, GIA)，GIA 可以将全局信息整合到整个网络中，改善增强图像中的颜色不一致和伪影。经过对 [6–8]的仔细研究，基于 U-Net 网络的弱光图像增强方法取得了很好的效果。然而，该网络仍然存在两个局限性：

- (a) 跳过连接只融合相同尺度的特征，导致编码器和解码器之间存在较大的语义差距；
- (2) 全局上下文信息的连接相对稀缺。

这两个限制容易导致增强图像的细节丢失、对比度差和颜色信息不准确。Zhou等 [9,10]改进了 U-Net 架构生成 U-Net++，在 U-Net++ 中将跳跃连接重新设计为一系列嵌套的密集跳跃连接，实现了多个特征的融合，加强了不同层次的特征信息接触。

CNN for LLIE

CNN 在许多计算机视觉任务中表现出了令人印象深刻的结果。CNN 利用注意力机制 [11,12]和上下文信息，从原始图像中生成注意感知并提取多尺度特征 [13,14]。基于 CNN 的低光图像增强方法不断发展。例如，基于 CNN 的自适应弱光图像增强框架 [15]极大地增强了图像对比度、颜色和细节信息。然而，现有的基于 CNN 的方法大多侧重于图像亮度、纹理和颜色的恢复 [16]。由于局部光照不均匀，颜色信息和细节信息丢失严重，容易出现过增强或增强不足的问题。

References

- [1] Minfeng Zhu, Pingbo Pan, Wei Chen, and Yi Yang. Eemefn: Low-light image enhancement via edge-enhanced multi-exposure fusion network. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13106–13113, 2020.
- [2] Zhiliang Peng, Wei Huang, Shanzhi Gu, Lingxi Xie, Yaowei Wang, Jianbin Jiao, and Qixiang Ye. Conformer: Local features coupling global representations for visual recognition. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 367–376, 2021.
- [3] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

- [4] Prajit Ramachandran, Niki Parmar, Ashish Vaswani, Irwan Bello, Anselm Levskaya, and Jon Shlens. Stand-alone self-attention in vision models. *Advances in neural information processing systems*, 32, 2019.
- [5] Jiafeng Li, Ying Wen, and Lianghua He. Scconv: Spatial and channel reconstruction convolution for feature redundancy. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6153–6162, 2023.
- [6] Chen Chen, Qifeng Chen, Jia Xu, and Vladlen Koltun. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3291–3300, 2018.
- [7] Syed Waqas Zamir, Aditya Arora, Salman Khan, Fahad Shahbaz Khan, and Ling Shao. Learning digital camera pipeline for extreme low-light imaging. *Neurocomputing*, 452:37–47, 2021.
- [8] Zibo Meng, Runsheng Xu, and Chiu Man Ho. Gia-net: Global information aware network for low-light imaging. In *European Conference on Computer Vision*, pages 327–342. Springer, 2020.
- [9] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018.
- [10] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging*, 39(6):1856–1867, 2019.
- [11] Chenglin Yang, Siyuan Qiao, Adam Kortylewski, and Alan Yuille. Locally enhanced self-attention: Combining self-attention and convolution as local and context terms. *arXiv preprint arXiv:2107.05637*, 2021.
- [12] Cheng Zhang, Qingsen Yan, Yu Zhu, Xianjun Li, Jinqiu Sun, and Yanning Zhang. Attention-based network for low-light image enhancement. In *2020 IEEE international conference on multimedia and expo (ICME)*, pages 1–6. IEEE, 2020.
- [13] Juncheng Li, Faming Fang, Kangfu Mei, and Guixu Zhang. Multi-scale residual network for image super-resolution. In *Proceedings of the European conference on computer vision (ECCV)*, pages 517–532, 2018.
- [14] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. Learning enriched features for real image restoration and enhancement. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXV 16*, pages 492–511. Springer, 2020.
- [15] Xiaoxiao Li, Xiaopeng Guo, Liye Mei, Mingyu Shang, Jie Gao, Maojing Shu, and Xiang Wang. Visual perception model for rapid and adaptive low-light image enhancement. *arXiv preprint arXiv:2005.07343*, 2020.

-
- [16] Ke Xu, Xin Yang, Baocai Yin, and Rynson WH Lau. Learning to restore low-light images via decomposition-and-enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2281–2290, 2020.