3 月工作汇报

Ku Jui

March 2024

Content

1 引言

低光图像增强（Low-Light Image Enhancement, LLIE）是图像处理领域的一个关键任务，旨在提升低光环境下拍摄图像的视觉质量。该领域的最新进展主要由深度学习技术推动，包括多种学习策略、网络架构、损失函数和训练数据集的应用。低光图像增强在视觉监控、自动驾驶和计算摄影等多个领域具有广泛应用。尤其在智能手机摄影领域，由于相机光圈大小、实时处理需求和存储限制，低光环境下的图像拍摄面临着显著挑战。

传统的低光增强方法，如基于直方图均衡化 [1] 和 Retinex 理论 [2–4] 的方法，虽然能够在一定程度上改善图像质量，但存在一些局限性。直方图均衡化能够提高图像的全局对比度，但可能增加背景噪声的对比度并降低有用图像内容的对比度，导致视觉效果不佳。Retinex 算法旨在消除图像照度分量的干扰并还原图像真实色彩，但通常忽略噪声问题，可能导致增强结果中噪声的保留或放大，并且其复杂的优化过程增加了模型的复杂度。

近年来，基于深度学习的低光图像增强技术取得了显著成就，特别是卷积神经网络（CNN）在多个计算机视觉任务中展现出卓越的性能。CNN 通过利用注意力机制 [5,6] 和上下文信息，能够从原始图像中有效提取局部特征 [7–9]。在低光图像中，亮度较低、对比度较弱的区域之间存在一定的关联性和相互作用，如果模型能够捕获全局光照，将有助于恢复图像的整体亮度和对比度 [10, 11][1]。在图像处理领域，尤其是在低光图像恢复和增强方面，保持图像的结构完整性是一个重要的挑战。Fu 等人 [12] 引入了一种加权变分模型，通过边缘感知权重来保持图像的结构完整性，从而在增强过程中保持边缘和细节信息。随后，Wang 等人 [13] 在其提出的 ESRGAN 模型中强调了集成全局和局部特征的重要性，包括保持边缘和纹理细节的完整性，以增强图像的感知质量。最近，Xu [14] 提出了一种基于深度学习的方法，通过分解和增强来恢复低光图像，其中在分解阶段保持了图像的结构信息，包括边缘和纹理细节，这对于后续的增强阶段至关重要。因此，即使在低光条件下，对象的轮廓和边缘仍然是重要的视觉特征，捕获这些长距离的边缘信息对于保持图像的结构完整性至关重要。

在低光条件下捕获和恢复图像中的纹理和模式是图像增强领域的一项重要挑战。Loh 和 Chan [15] 通过对低光照图像的特性进行研究，强调了在弱光条件下保持图像纹理和模式的重要性。进一步地，Jiang 等人 [16] 提出了一种基于生成对抗网络的方法，有效地恢复了低光图像中的细节，包括难以辨识的纹理和模式。同样，Lv 等人 [17] 通过卷积神经网络增强了低光图像和视频，专注于恢复纹理和模式的细节，从而提供更丰富的图像内容。因此，尽管弱光图像中的纹理和模式可能难以辨识，但它们对于理解图像内容仍然非常重要。

理解图像内容的关键在于识别不同对象和场景元素之间的空间关系。Wei 等人 [18] 提出的深度 Retinex 分解方法通过在特征提取过程中考虑长距离依赖关系来进一步提高低光图像的质量。

然而，现有的基于 CNN 的方法在处理局部光照不均匀、颜色信息和细节信息丢失问题时，仍存在过增强或增强不足的挑战，并且增强结果受到感受野大小的限制。因此，开发能够克服这些限制并有效提升低光图像质量的深度学习模型仍然是一个重要的研究方向。

为了解决这些问题，我们提出了一种新颖的方法（方法名待定）。受到 ULite [19] 中深度可分离卷积的启发，我们将 U-Net 网络中的卷积改为轴向深度可分离卷积，以减少模型冗余同时保持增强效果。此外，我们借鉴 Swin-Unet [20] 在 BottleNeck 中使用连续的 Swin Transformer 块以捕获图像长距离特征，相较于传统的 Transformer 块，Swin Transformer 能够减少参数量和模型复杂度。通过以上改进，我们的方法能够有效提升低光图像的视觉质量，同时保持较低的计算复杂度，具有实际应用的潜力。

[1] Retinex 理论的一个关键思想是图像的颜色和亮度感知取决于全局光照条件，因此捕获和调整全局光照信息对于图像增强和恢复至关重要。

本研究提出了三个主要的创新点：(若并行架构未实现或性能不佳，则可尝试仅通过 CNN 分支进行图像的弱光恢复，并相应调整创新点，去除并行架构这一创新点。)

- 首先，提出了一种结合卷积神经网络（CNN）和 Transformer 的并行架构用于弱光图像增强。在该架构中，UNet 网络用于捕获图像的局部特征并进行初步恢复，而 Swin Transformer 块用于捕获图像的长距离特征。最后，通过特征融合块将两者的特征进行融合，以实现更全面的图像增强效果。
- 其次，提出了一个深度语义模块，该模块融合了 Swin Transformer 分支，使 CNN 分支能够有效捕获图像的长距离特征。这种设计增强了 CNN 分支的能力，使其能够更好地理解图像的全局上下文信息。
- 最后，将深度可分离卷积融合进 CNN 分支中，应用于轻量级网络用于提取图像的局部特征。这种设计旨在减少模型的参数量和计算复杂度，同时保持对局部特征的有效提取能力。

2 实验计划

实验的过程中，确保所有的实验在相同的硬件和软件环境下进行，并且为了确保结果的可靠性，可能需要多次运行实验并取平均值。我们主要基于 PyTorch 进行模型的搭建、训练和评估。基于 scikit-image 库计算 PSNR、SSIM 等评价指标。首先构建 U-Net 基本架构模型，然后实现 Swin Transformer 块中的 LocalselfAttention 类，PositionEncoding 类，PositionEmbedding 类。

2.0.1 Dataset

Tab. 1 展示了我们在实验中会使用到的弱光数据集，这些数据集包含真实数据与合成数据。对于每个数据集，我们需要进行如下操作：

- 预处理: 确保所有图像都经过相同的预处理步骤，如尺寸调整、归一化等。
- 分割: 将每个数据集分为训练集、验证集和测试集。

Name	Number	Format	Real/Syn	Video
LOL [18]	500	RGB	Real	
SCIE [21]	4,413	RGB	Real	
VE-LOL-L [22]	2,500	RGB	Real+Syn	

Table 1: Summary of paired training datasets. 'Syn' represents Synthetic.

2.0.2 Train

- 基线模型: 首先，训练基线模型。
- 消融研究: 接着，训练正常 BottleNeck 的模型，以进行消融实验。

2.0.3 Performance Evaluation

对于每个数据集，使用以下指标评估模型性能：

- 峰值信噪比 (PSNR)
- 结构相似性指数 (SSIM)
- 感知图像质量评估 (LPIPS)

2.0.4 Loss Function

$$J_{Huber}(\delta) = \frac{1}{N}\sum_{i=1}^{N}\begin{cases}\frac{1}{2}\|\hat{y}_i - y_i\|_2^2, \|\hat{y}_i - y_i\| < \delta, \\ \delta\left(\|\hat{y}_i - y_i\|_1 - \frac{1}{2}\delta\right), \|\hat{y}_i - y_i\| \geq \delta.\end{cases} \tag{1}$$

$$\ell_{feat}^{\phi,j}(\hat{y}, y) = \frac{1}{C_j H_j W_j}\|\phi_j(\hat{y}) - \phi_j(y)\|_2^2 \tag{2}$$

$$\mathcal{L}^{\text{SSIM}}(P) = 1 - \text{SSIM}(\tilde{p}). \tag{3}$$

一共尝试以下两种损失函数的搭配方式:

- 休伯损失函数和 SSIM 损失函数

$$L_{loss} = \alpha J_{Huber}(\delta) + \beta\mathcal{L}^{\text{SSIM}}(P) \tag{4}$$

- 休伯损失函数，SSIM 损失函数，Perceptual 损失函数 (耗费更多训练时间)

$$L_{loss} = \alpha J_{Huber}(\delta) + \beta\mathcal{L}^{\text{SSIM}}(P) + \gamma\ell_{feat}^{\phi,j}(\hat{y}, y) \tag{5}$$

References

[1] T-L Ji, Malur K Sundareshan, and Hans Roehrig. Adaptive image contrast enhancement based on human visual properties. *IEEE transactions on medical imaging*, 13(4):573–586, 1994.

[2] Edwin H Land. The retinex. In *Ciba Foundation Symposium-Colour Vision: Physiology and Experimental Psychology*, pages 217–227. Wiley Online Library, 1965.

[3] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977.

[4] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. Properties and performance of a center/surround retinex. *IEEE transactions on image processing*, 6(3):451–462, 1997.

[5] 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.

[6] 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.

[7] Anil K Jain and Farshid Farrokhnia. Unsupervised texture segmentation using gabor filters. *Pattern recognition*, 24(12):1167–1186, 1991.

[8] David G Lowe. Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, 60:91–110, 2004.

[9] Timo Ojala, Matti Pietikainen, and Topi Maenpaa. Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Transactions on pattern analysis and machine intelligence*, 24(7):971–987, 2002.

[10] 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.

[11] Shuhang Wang, Jin Zheng, Hai-Miao Hu, and Bo Li. Naturalness preserved enhancement algorithm for non-uniform illumination images. *IEEE transactions on image processing*, 22(9):3538–3548, 2013.

[12] Xueyang Fu, Delu Zeng, Yue Huang, Xiao-Ping Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2782–2790, 2016.

[13] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.

[14] 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.

[15] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42, 2019.

[16] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE transactions on image processing*, 30:2340–2349, 2021.

[17] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, volume 220, page 4, 2018.

[18] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018.

[19] Binh-Duong Dinh, Thanh-Thu Nguyen, Thi-Thao Tran, and Van-Truong Pham. 1m parameters are enough? a lightweight cnn-based model for medical image segmentation. In *2023 Asia Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*, pages 1279–1284. IEEE, 2023.

[20] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-unet: Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.

[21] Jianrui Cai, Shuhang Gu, and Lei Zhang. Learning a deep single image contrast enhancer from multi-exposure images. *IEEE Transactions on Image Processing*, 27(4):2049–2062, 2018.

[22] Haiyang Jiang and Yinqiang Zheng. Learning to see moving objects in the dark. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7324–7333, 2019.

[23] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[24] Xiaoyi Dong, Jianmin Bao, Dongdong Chen, Weiming Zhang, Nenghai Yu, Lu Yuan, Dong Chen, and Baining Guo. Cswin transformer: A general vision transformer backbone with cross-shaped windows. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12124–12134, 2022.

[25] Gedas Bertasius, Jianbo Shi, and Lorenzo Torresani. High-for-low and low-for-high: Efficient boundary detection from deep object features and its applications to high-level vision. In *Proceedings of the IEEE international conference on computer vision*, pages 504–512, 2015.

[26] Saining Xie and Zhuowen Tu. Holistically-nested edge detection. In *Proceedings of the IEEE international conference on computer vision*, pages 1395–1403, 2015.

[27] Yun Liu, Ming-Ming Cheng, Xiaowei Hu, Kai Wang, and Xiang Bai. Richer convolutional features for edge detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3000–3009, 2017.

[28] Ruoxi Deng and Shengjun Liu. Deep structural contour detection. In *Proceedings of the 28th ACM international conference on multimedia*, pages 304–312, 2020.

[29] Zhuo Su, Wenzhe Liu, Zitong Yu, Dewen Hu, Qing Liao, Qi Tian, Matti Pietikäinen, and Li Liu. Pixel difference networks for efficient edge detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 5117–5127, 2021.

[30] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[31] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[32] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.