



煤炭科学技术

煤炭科学研究院 COAL SCIENCE AND TECHNOLOGY

基于Transformer和自适应特征融合的矿井低照度图像亮度提升和细节增强方法

田子建 吴佳奇 张文琪 陈伟 周涛 杨伟 王帅

引用本文:

田子建, 吴佳奇, 张文琪, 等. 基于Transformer和自适应特征融合的矿井低照度图像亮度提升和细节增强方法[J]. 煤炭科学技术, 2024, 52(1): 297–310.

TIAN Zijian, WU Jiaqi, ZHANG Wenqi. An illuminance improvement and details enhancement method on coal mine low-light images based on Transformer and adaptive feature fusion[J]. Coal Science and Technology, 2024, 52(1): 297–310.

在线阅读 View online: <https://doi.org/10.13199/j.cnki.cst.2023-0112>

您可能感兴趣的其他文章

Articles you may be interested in

基于深度卷积特征的露天矿卡车装载状况识别技术研究

Study on recognition technology of truck loading condition in open-pit mine based on deep convolutional features

煤炭科学技术. 2021, 49(10): 167–176 <http://www.mtkxjs.com.cn/article/id/c08f5f3b-8582-470a-a82f-3b8ecb636a36>

基于深度学习的煤矸石识别技术与实现

Coal and gangue recognition technology and implementation based on deep learning

煤炭科学技术. 2021, 49(12): 202–208 <http://www.mtkxjs.com.cn/article/id/5f975297-7e00-4ebe-aa8d-a4f4808d034b>

基于CS架构的煤矿井下图像处理算法研究

Research on processing algorithm of image in underground coal mine based on CS framework

煤炭科学技术. 2018(2) <http://www.mtkxjs.com.cn/article/id/13f7458c-3d99-4146-a3a9-1c6d0679e488>

基于深度学习的多产品煤料图像分类方法研究

Study on multi-product coal image classification method based on deep learning

煤炭科学技术. 2021, 49(9): 117–123 <http://www.mtkxjs.com.cn/article/id/3ceaf1e6-7be5-4bd9-8238-18b18fe8b9c1>

多尺度分区统一化LBP算子井下人员人脸识别方法

Underground personnel face recognition method based on multi-scale unified LBP operator

煤炭科学技术. 2019(12) <http://www.mtkxjs.com.cn/article/id/e019c116-2fd6-4967-ba0b-643a58030811>

基于改进Otsu的煤体CT图像阈值分割算法的研究

Research on threshold segmentation algorithm of coal CT images based on improved Otsu

煤炭科学技术. 2021, 49(1): 264–271 <https://doi.org/10.13199/j.cnki.cst.2021.01.023>



关注微信公众号, 获得更多资讯信息



移动扫码阅读

田子建, 吴佳奇, 张文琪, 等. 基于 Transformer 和自适应特征融合的矿井低照度图像亮度提升和细节增强方法[J]. 煤炭科学技术, 2024, 52(1): 297–310.

TIAN Zijian, WU Jiaqi, ZHANG Wenqi, *et al.* An illuminance improvement and details enhancement method on coal mine low-light images based on Transformer and adaptive feature fusion[J]. Coal Science and Technology, 2024, 52(1): 297–310.

基于 Transformer 和自适应特征融合的矿井低照度图像亮度提升和细节增强方法

田子建¹, 吴佳奇¹, 张文琪¹, 陈伟^{1,2}, 周涛³, 杨伟¹, 王帅^{1,4}

(1. 中国矿业大学(北京)机电与信息工程学院, 北京 100083; 2. 中国矿业大学计算机科学与技术学院, 江苏徐州 221116; 3. 中国石油大学(华东)石油工程学院, 山东青岛 266580; 4. 国家矿山安全监察局内蒙古局, 内蒙古呼和浩特 010010)

摘要: 高质量矿井影像为矿山安全生产提供保障, 也有利于提高后续图像分析技术的性能。矿井影像受低照度环境的影响, 易出现亮度低, 照度不均, 颜色失真, 细节信息丢失严重等问题。针对上述问题, 提出一种基于 Transformer 和自适应特征融合的矿井低照度图像亮度提升和细节增强方法。基于生成对抗思想搭建生成对抗式主体模型框架, 使用目标图像域而非单一参考图像驱动判别器监督生成器的训练, 实现对低照度图像的充分增强; 基于特征表示学习理论搭建特征编码器, 将图像解耦为亮度分量和反射分量, 避免图像增强过程中亮度与颜色特征相互影响从而导致颜色失真问题; 设计 CEM-Transformer Encoder 通过捕获全局上下文关系和提取局部区域特征, 能够充分提升整体图像亮度并消除局部区域照度不均; 在反射分量增强过程中, 使用结合 CEM-Cross-Transformer Encoder 的跳跃连接将低级特征与深层网络处特征进行自适应融合, 能够有效避免细节特征丢失, 并在编码网络中添加 ECA-Net, 提高浅层网络的特征提取效率。制作矿井低照度图像数据集为矿井低照度图像增强任务提供数据资源。试验显示, 在矿井低照度图像数据集和公共数据集中, 与 5 种先进的低照度图像增强算法相比, 该算法增强图像的质量指标 PSNR、SSIM、VIF 平均提高了 16.564%, 10.998%, 16.226% 和 14.438%, 10.888%, 14.948%, 证明该算法能够有效提升整体图像亮度, 消除照度不均, 避免颜色失真和细节丢失, 实现矿井低照度图像增强。

关键词: 图像增强; 图像识别; 生成对抗网络; 特征解耦; Transformer

中图分类号: TP391

文献标志码: A

文章编号: 0253-2336(2024)01-0297-14

An illuminance improvement and details enhancement method on coal mine low-light images based on Transformer and adaptive feature fusion

TIAN Zijian¹, WU Jiaqi¹, ZHANG Wenqi¹, CHEN Wei^{1,2}, ZHOU Tao³, YANG Wei¹, WANG Shuai^{1,4}

(1. School of Mechanical Electronic & Information Engineering, China University of Mining and Technology- Beijing, Beijing 100083, China; 2. School of Computer Science & Technology, China University of Mining and Technology, Xuzhou 221116, China; 3. School of Petroleum Engineering, China University of Petroleum (East China), Qingdao 266580, China; 4. Inner Mongolia Bureau of the National Mine Safety Administration, Inner Mongolia Autonomous Region, Hohhot 010010, China)

Abstract: High quality mine images can provide guarantee for mine safety production, and improve the performance of subsequent image analysis technologies. Affected by low illuminance environment, mine images suffer low brightness, uneven brightness, color distortion, and serious loss of details. Aiming at the above problems, an illuminance improvement and details enhancement method on coal mine low-light images based on Transformer and adaptive feature fusion was proposed to enhance the brightness and detail of mine low illuminance

收稿日期: 2023-02-07

责任编辑: 周子博

DOI: 10.13199/j.cnki.cst.2023-0112

基金项目: 国家自然科学基金资助项目(52074305, 52274160); 国家自然科学基金委员会-山西省人民政府煤基低碳联合基金资助项目(U1510115)

作者简介: 田子建(1964—), 男, 湖南望城人, 教授, 博士生导师, 博士。E-mail: Tianzj0726@126.com

通讯作者: 陈伟(1978—), 男, 江苏徐州人, 教授, 博士生导师, 博士。E-mail: chenwdavior@163.com

images. Based on the idea of generative confrontation, a framework of generative adversary agent model was built, and the target image domain was used instead of a single reference image to drive discriminator to supervise the training of generator, so as to achieve adaptive enhancement of low illuminance images; The feature encoder was built based on the feature representation learning theory to decouple the image into illuminance component and reflection component, the method can avoid the interaction between illuminance and color features during image enhancement to solve the color distortion; the CEM-Transformer Encoder was designed to enhance the brightness component, the method can improve the overall image brightness and eliminate the local area brightness unevenness, by capturing the global context and extracting the local area features; In the process of reflection component enhancement, the skip connection combined with CEM-Cross-Transformer Encoder was used to adaptively fuse low-level features with features at the deep CNN layers, which can effectively avoid the loss of detailed features, and ECA-Net was added to the encode network to improve the feature extraction efficiency of the shallow CNN layers. The low illuminance mine image dataset was produced to provide data resources for the low illuminance mine image enhancement task. The experiments show that, compared with five advanced low illuminance image enhancement algorithms, the quality indicators PSNR, SSIM and VIF of the images enhanced by the algorithm are improved by 16.564%, 10.998%, 16.226% and 14.438%, 10.888% and 14.948% on average on the low illuminance mine image dataset and the public dataset. And the algorithm also perform well in subjective visual evaluation. The above results prove that the algorithm can effectively improve the overall image brightness and eliminate the uneven brightness, thus to achieve mine low illuminance image enhancement.

Key words: image enhancement; image recognition; generative adversarial network; feature disentanglement; Transformer

0 引言

近年来,煤矿开采方法向智能、环保、安全方向发展^[1],矿井监视系统的重要性日益提高,不仅为矿井安全生产提供保障,也推动了矿山智能化发展。煤矿井下工作环境具有照明条件差、电磁噪声密集、粉尘及雾气密度大等不利因素,致使采集的画面亮度低,光照不均,颜色饱和度低和细节特征模糊,严重干扰人工观测,并且不清晰的监视图像将影响目标追踪^[2]、目标识别^[3]、图像语义分割^[4]、目标检测^[5-6]等图像分析技术的性能。为了得到亮度高、光照均匀、颜色均衡且细节信息完整的矿井影像,需要对矿井低照度图像进行有效增强处理^[7]。

传统低照度图像增强方法主要包括直方图均衡化、灰度变换法和 Retinex 算法。直方图均衡化通过扩大像素灰度级的动态范围,使像素灰度值在灰度直方图上均匀分布,从而提高图像亮度,例如 BBHE^[8]、DSIHE^[9]等。灰度变换法使用特定的灰度线性变换或非线性变换函数对像素点进行处理^[10],通过调节像素值的动态范围来提高图像的对比度。Retinex 算法假设图像由光照分量和反射分量组成,通过对两分量进行特定处理以提升图像的视觉效果,包括单尺度 Retinex 算法 (SSR)^[11]、多尺度 Retinex 算法 (MSR)^[12]、颜色修复 Retinex 算法 (MSRCR)^[13]等。近年来,深度学习技术发展迅速,大量基于深度学习的图像增强算法被提出,2017 年 Lore 基于卷积神经网络率先提出低照度增强网络模型 LLNet^[14]; 2018

年 Wei C 将 Retinex 理论模型与卷积神经网络相融合设计出 RetinexNet^[15]。近年来以 GAN^[16](Generative Adversarial Networks)为基础而设计的无监督模型大量涌现,如 StarGAN^[17]、InfoGAN^[18]及 CycleGAN^[19]、EnlightenGAN^[20]等,其在图像风格转换、图像翻译、低质量图像增强等领域展现了出色的性能。

然而上述方法仍具有较多不足之处,传统方法需基于专家经验进行手动调参且针对不同场景需更改模型参数,模型设计过程低效;现有的传统算法无法有效解决亮度不均现象。基于 CNN 的深度学习算法也存在较多缺陷,首先,卷积核操作具有空间局限性,无法建立像素的远距离依赖,导致其只能处理局部区域特征,无法充分考虑图像的上下文信息,该缺陷使其难以解决各局部区域的光照不均问题;其次,在特征提取过程中,为获取较大感受野,需堆叠大量卷积层,造成大量计算损耗,并导致细节、纹理等低级特征和结构、轮廓等中级特征在较深卷积层处大量丢失,图像信息的损失将严重影响人眼视觉及各类图像应用技术对图像内容的正确感知;上述方法并未有效解决颜色失真问题,极低照度图像需要大幅提升图像亮度来改善图像质量,由于图像中亮度和颜色特征具有耦合关系,亮度的大幅变化会造成图像颜色激变,从而引起颜色失真,使得图像关键区域呈现错误的信息,影响后续的图像分析工作。

针对上述问题,提出一种基于 Transformer 和自适应特征融合的矿井低照度图像亮度提升和细节增

强方法。基于生成对抗思想搭建生成对抗式主体网络框架,使低照度图像特征不趋向于参考图像而是参考图像的特征域,避免由于部分参考图像质量较差而干扰模型增强性能,从而实现低照度图像的充分增强;受特征表示学习理论启发,基于文献 [21] 的特征提取网络,搭建特征编码器将图像解耦为包含细节、结构、颜色的反射分量和包含亮度特征的亮度分量,从而解决颜色特征随亮度改变而造成的颜色失真问题;针对反射分量,设计一种结合交叉注意力机制的跳跃连接,实现低级特征和高维特征的自适应融合来解决深层网络处特征丢失问题;针对亮度分量,将原始 Transformer^[23] 与深度可分离卷积^[24] (DS-Conv, Depthwise Separable Convolution Block) 相结合,设计了 CEM-Transformer Encoder,通过自注意力机制建立像素间长期依赖关系,实现图像中各局部区域的有效交互,并使用深度可分离卷积层充分提取局部特征,从而在实现全局亮度提升的同时,消除光照不均现象。试验结果显示文中算法在各类试验中表现出色,证明文中算法能够有效改善矿井低照度图像质量,有利于提升矿井监视系统及下游图像分析技术的性能。

1 基于 Transformer 和多尺度特征融合的矿井低照度图像亮度提升和细节增强方法

1.1 整体模型结构

受生成对抗网络启发,设计生成对抗主体网络框架,使生成器各模块与判别器进行对抗博弈不断优化性能,最终学习到最优参数;基于特征表示学习搭建两类编码器 (Reflection component Encoder, Illuminance component Encoder),并将低照度图像的 f_L 与参考图像的 f_R 进行对齐,从而将图像解耦为反射分量 (f_R) 和亮度分量 (f_I),有利于避免颜色特征受亮度变化的影响而发生颜色失真;设计全局-局部增强网络 (GLENet, Global-Local enhancement Network) 提升 f_I 的全局亮度水平,并使局部区域亮度交互性地自适应增强,从而在 F_I 中消除照度不均;设计多尺度特征融合网络 (MFNet, Multi-scale features fusion Network) 提高对 f_R 的低级特征提取效率,并将低级特征与高维特征进行自适应融合,从而实现 F_R 中细节特征的有效增强,避免细节特征的大量丢失;最终由解码器 (Decoder) 将 F_R 和 F_I 解码为增强图像,实现低照度图像增强。如图 1 所示。

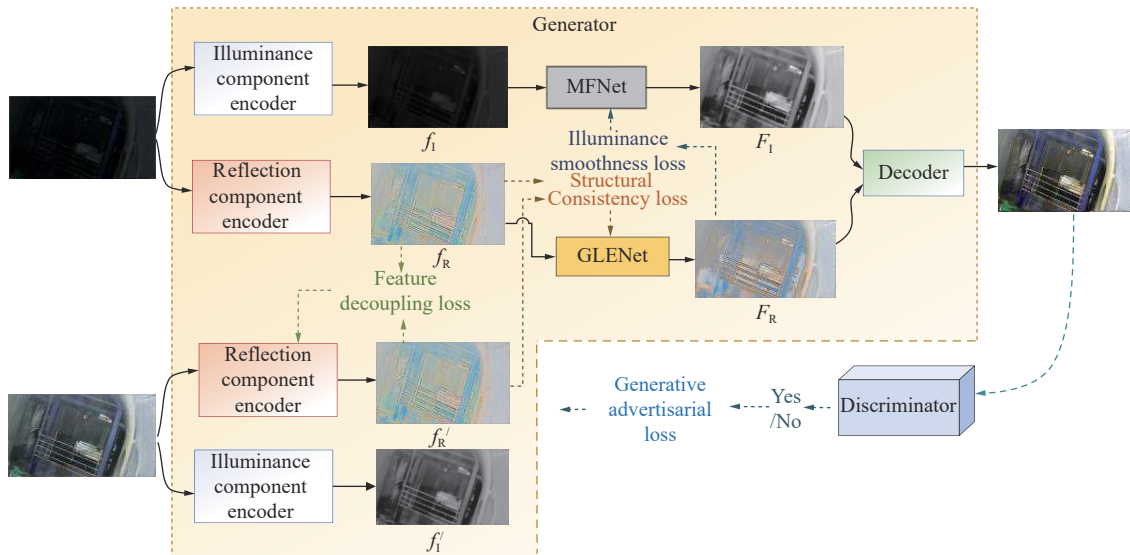


图 1 整体模型结构

Fig.1 Structure of overall model

1.2 生成对抗模型

文中使用该 GAN 思想搭建生成对抗主体网络,特征编码器、双通道并行增强网络及解码器作为文中生成对抗模型的生成器,通过增强低照度图像的亮度分量和反射分量来提升图像质量;设计全局-局部判别器作为文中模型的判别器,通过对增强图像

及其局部区域的特征进行真实性鉴别,从而在训练阶段促进生成器各模块优化。不同于 RetinexNet^[15] 以单一参考图像的光照分量和反射分量来监督原始图像的增强,文中使用全体参考图像组成的目标图像域来监督图像增强,能避免由于部分参考图像本身质量低而影响模型的图像增强能力。

相较于传统判别器,文中使用全局-局部判别器鉴别增强图像的真实性,其中,全局判别器用于鉴别整体图像的整体风格特征,局部判别器能够鉴别增强图像的随机图像块是否符合目标图像域特征并鉴别局部区域细节恢复情况及亮度分布的合理性,从而促使生成器在改善整体图像质量的同时,有效避免细节损失及光照不均现象。并且文中判别器在批归一化 (Batch Normalization) 的基础上,增添谱归一化 (Spectral Normalization)使局部判别器具有 Lipschitz 连续性,能够缓解判别器性能变化的剧烈程度,从而改善模型训练过程中,由于判别器性能提升过快导致生成器发生梯度消失现象,有利于提升整体模型的鲁棒性^[25]。文中判别器的详细网络结构如图 2 所示。

全局判别器的卷积核尺寸均为 3*3,步长均为 2,输出特征图通道数分别为 {16, 32, 64, 128, 256, 1},局部判别器的卷积核同上,输出特征图通道数分别为 {16, 32, 64, 1}。

1.3 特征编码器

特征表示学习^[26-27]旨在根据图像中各类物理特

征的特性,使用潜在的特征分布表示各类物理特征,便于后续进行隐空间特征解耦。为实现亮度分量与反射分量的有效解耦,文中根据反射分量、亮度分量的特点搭建 2 类编码器,并受文献 [15] 中 Decom-Net 的启发,将低照度图像与目标图像的反射分量进行对齐,从而将亮度分量从原图中分离。解码器将增强后的两分量融合为最终增强图像。详细网络结构如图 3 所示。

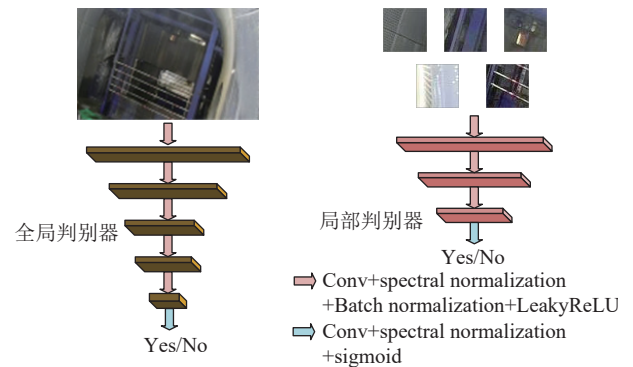


图 2 全局-局部判别器网络结构

Fig.2 Structure of global-local discriminator

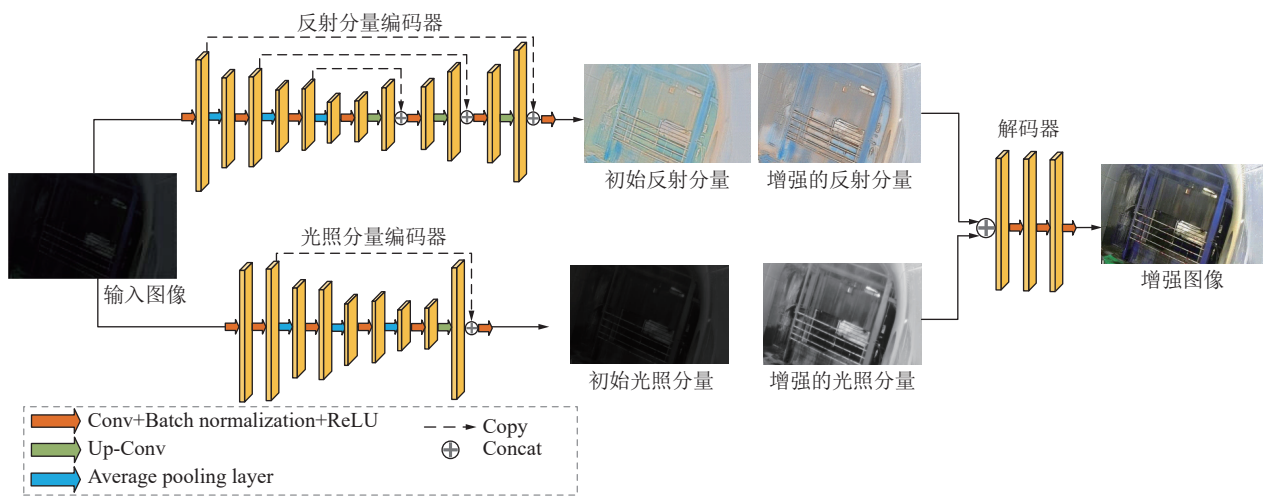


图 3 特征编码器和解码器网络结构

Fig.3 Structure of feature encoders and decoder

文中亮度分量包含图像的亮度特征,文献 [21] 使用全局池化层来提取整体风格特征,而亮度分量属于低级图像特征不适用全局池化层提取,所以文中使用浅层卷积层提取图像的亮度特征,并且为了对各局部区域的亮度分量进行平滑度调节,文中加深网络深度获取抽象级更高的亮度分布,通过将二者进行信道级拼接,从而得到包含亮度分布的亮度分量。各特征图通道数为 {3, 8, 16, 32, 64, 1}。

反射分量编码器由细节、纹理、颜色等低级特征

和结构、语义等高维特征组成,反射分量编码器的结构受文献 [21] 中内容编码器的启发,使用 U-Net 提取多尺度特征;并且在网络中加入跳跃连接,避免低级特征丢失;使用转置卷积进行特征图上采样,相较于普通插值操作,转置卷积产生的特征冗余小,特征映射误差小^[28]。相比于文献 [21] 的编码器,文中结构能够保留更多细节信息和颜色特征,并且网络层较深,特征提取充分。反射分量编码器中卷积层的卷积核尺寸均为 3*3,步长均为 2,各特征图通道数

分别为{3, 16, 32, 64, 32, 16, 1}。

解码器先使用 Concat 层拼接两分量, 再使用 3 层 Conv 对拼接特征进行自适应特征调节, 使两分量的融合效果更自然, 各 Conv 通道数为{16, 32, 1}。

1.4 双通道并行增强模块

传统 Retinex 方法将图像分解为光照分量和反射分量后, 通过去除光照分量并保留反映物体本质属性的反射分量来增强图像。该方法在处理常规低照度图像时表现出色, 但由于煤矿低照度图像的亮度极低、照度不均, 且反射分量中颜色特征暗淡, 细节模糊, 导致上述方法对其增强效果有限。文中根据两分量的特点设计双通道并行增强模块, 包含多尺度特征融合网络 (MFNet, Multiscale features Fusion Network) 和全局-局部增强网络 (GLENet, Global and Local features Enhancement Network), 在增强亮度分量的同时, 对反射分量进行充分增强, 从而有效提升图像质量。

1.4.1 全局-局部增强网络

照度不均均是低照度图像增强领域的重点问题, 现有的图像亮度提升方法解决该问题的效果不佳, 表现在只使用一种标准提升全局图像亮度, 导致局部区域的阴影和过曝^[29-30], 或者能够对不同曝光度区域进行增强^[31-32], 但亮度提升有限且无法实现各区域之间交互性地自适应增强, 导致增强图像的棋盘效应严重。

Transformer^[23] 最初在 NLP 领域被提出, 近年来, 被引入计算机视觉领域来解决图像任务, 在图像分类、目标检测和图像分割任务中均有较好的性能表现^[33]。不同于卷积操作受限于局部运算只能依靠堆叠卷积层扩大感受野获取全局上下文关系, Transformer 通过将二维特征图转化为一维序列, 再计算序列间相关性, 能够快速捕获像素间远距离依赖关系, 利用该特性实现图像增强过程中各局部区域的有效交互, 从而避免增强图像的照度不均。但原始 Transformer 仍存在缺陷: 无法充分学习局部区域细节特征, 该缺陷易影响图像细微处亮度提升效果。因此在 Transformer 的基础上融合卷积层能够高效提取局部区域特征的优点, 设计了 CEM-Transformer Encoder, CEM(CNN Extraction Module) 能够在 transformer 模型获取全局上下文信息的同时, 提高模型对局部区域特征的捕获能力, 从而能够充分提升图像全局亮度并消除照度不均。并且为降低 Transformer 的高计算复杂度, 使用计算成本较小的深度可分离卷积层搭建 CEM。

设计的全局-局部增强网络由 3 个 CTE-Down sampling-Block 和 3 个 CTE-Up sampling-Block 组成, 并使用跳跃连接进行低维特征复用, 详细网络结构如图 4 所示。其中, CTE-Down sampling-Block 包含特征嵌入模块、CEM-Transformer Encoder、2*2 Stride 卷积下采样层, CTE-Up sampling-Block 使用转置卷积层进行上采样。

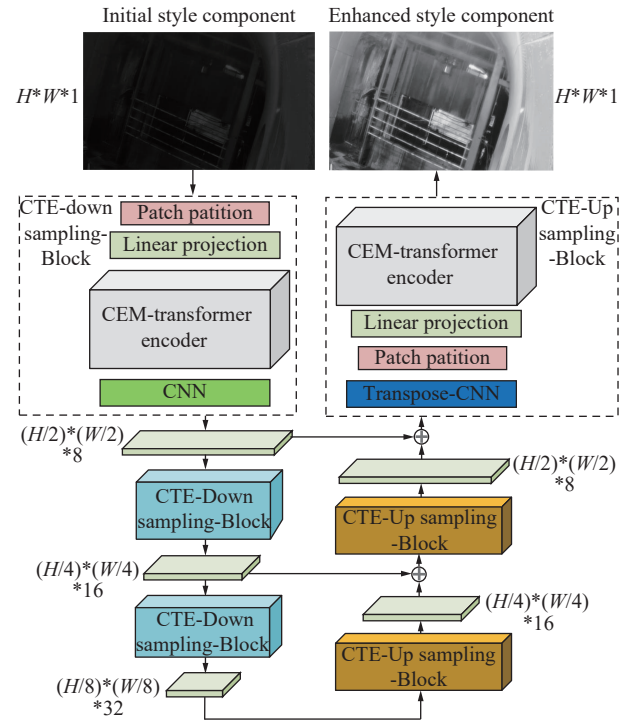


图 4 全局-局部增强网络结构

Fig.4 Structure of GLENet

对二维图像特征完成 Patch Embedding 编码后, 将生成序列输入 CEM-Transformer Encoder, 如图 5 所示, 首先使用 Layer normalization 层对输入序列进行归一化处理, 提高模型训练的鲁棒性, 再输入多头自注意力模块获取各序列间的相关性, 从而捕获图像中各局部区域之间的长期依赖。自注意力机制将输入向量通过特征变换矩阵 W_q 、 W_k 、 W_v 分别投影为查询向量 Q 、键向量 K 和值向量 V , 且 Q 、 K 、 V 向量的维度相同; 再将三者输入自注意力机制进行计算来捕获像素间远距离依赖关系:

$$\text{Attention}(Q, K, V) = A \cdot V = \text{Softmax}(Q \cdot K^T / \sqrt{d}) \cdot V \quad (1)$$

式中, A 为注意力矩阵; $Q \cdot K^T$ 用于计算 Q 中指定元素与 K 中每个元素的相似程度, 即相似度权重; Softmax 为数值归一化函数; $1/\sqrt{d}$ 为适应 Softmax 的数值缩放操作; $A \cdot V$ 为注意力矩阵与值向量的每个元素进行加权聚合。

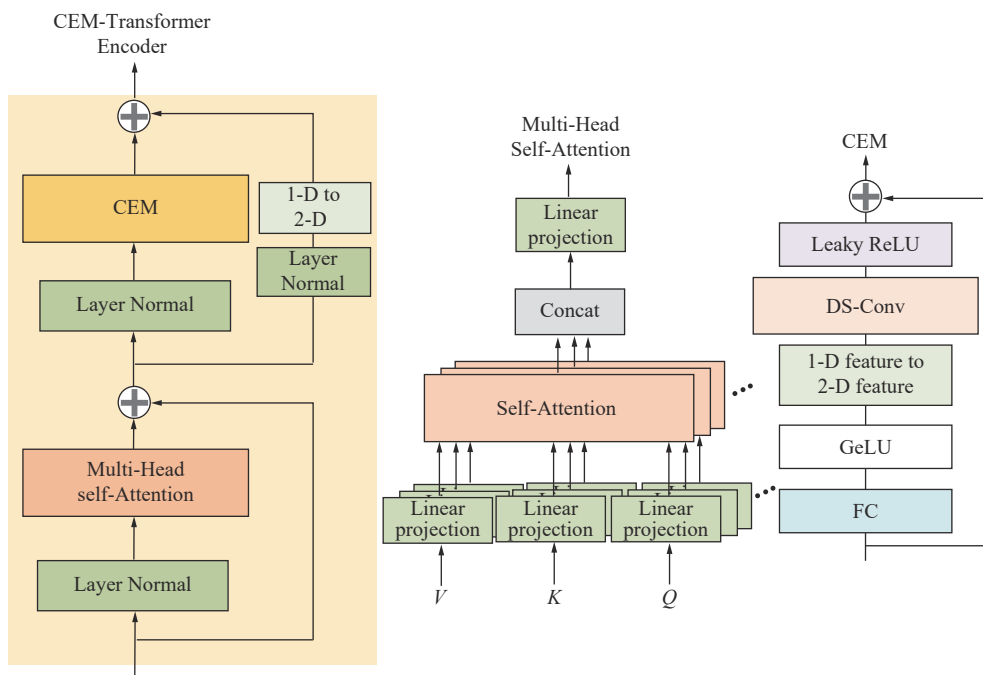


图5 CEM-Transformer Encoder 结构

Fig.5 Structure of CEM-Transformer Encoder

多头自注意力机制^[34]通过设置多个自注意力头,能够充分考虑各序列之间相关程度,有利于建立鲁棒性更强的依赖关系。多头自注意力机制在得到查询向量 Q 、键向量 K 和值向量 V 后,将 Q 、 K 、 V 根据注意力头个数进行平均划分,并各自完成 $\text{Attention}(Q, K, V)$ 计算,最后将所有注意力头结果融合:

$$\text{MSA}(x) = \mathbf{M}_{\text{MSA}}[\text{SA}_1(x); \text{SA}_2(x); \dots; \text{SA}_m(x)] \quad (2)$$

式中, SA 为单一自注意力头; m 为注意力头数量; \mathbf{M}_{MSA} 为融合矩阵。

为加强 Transformer 模型对局部区域特征的提取,设计了卷积提取模块改进原始 Transformer Encoder, CEM 网络结构如图 5 所示,相比于原始网络,DS-Conv 的卷积特性能够使模型增强对局部区域特征的感知能力,有利于全局-局部增强网络优化图像局部范围曝光度,并且 DS-Conv 的低计算复杂度特性能够降低整体模型的计算成本。CEM 中,首先使用全连接层对序列元素进行升维,并将一维向量映射为二维空间特征,再使用 DS-Conv 提取特征图的局部区域信息,GeLU 及 Leaky ReLU 均为特征变换后的激活函数,跳跃连接结构用于增加模型训练的稳定性。

1.4.2 多尺度特征融合网络

多尺度特征融合网络使用 U-net 网络作为主体框架进行多尺度特征提取,如图 6 所示。相比于亮度分量,反射分量包含的特征较丰富,为加强解码器对低级特征的提取效率,在解码器的浅层网络处添

加高效通道注意力模块 ECA-Net,用于增强对细节、纹理特征的关注;并且为避免深层编码解码结构造成的大量低级特征丢失,在 U-Net 网络中添加跳跃连接,从而在解码过程中补充丢失的低级特征,并且在跳跃连接中融合 CEM-Cross-Transformer Encoder,能够筛选不重要的特征,强调模型所需的特征,该设计能够有效避免深度网络层处低级特征丢失,从而在增强高级语义特征的同时,保留大量细节、颜色等低级特征。

CEM-Cross-Transformer Encoder 将 CEM-Transformer Encoder 中 Multi-Head Self-Attention 替换为 Multi-Head Cross-Attention,如图 6 所示,通过交叉注意力机制能够得出两输入之间各局部区域的相似性,见式(3)。多尺度特征融合网络使用融合 CEM-Cross-Transformer Encoder 的跳跃连接,将解码器提取的高维特征与跳跃连接引入的低维特征,按照各尺度特征之间的特征相似度进行自适应融合,从而加强高维特征所关注的低维特征,过滤非关注特征,实现对低级特征的自适应复用,有利于进一步提升模型对细节特征的增强效果。

$$\text{Cross-Attention}(Q_x, K_y, V_y) = \text{Softmax}(Q_x \cdot K_y^T / \sqrt{d}) \cdot V_y \quad (3)$$

式中, Q_x 、 K_y 和 V_y 分别为输入 x 、 y 通过特征变换矩阵映射得到的查询向量、键向量和值向量。

ECA-Net^[22] 是一种针对深度卷积神经网络的轻量级信道注意力模块,网络结构如图 6 所示。

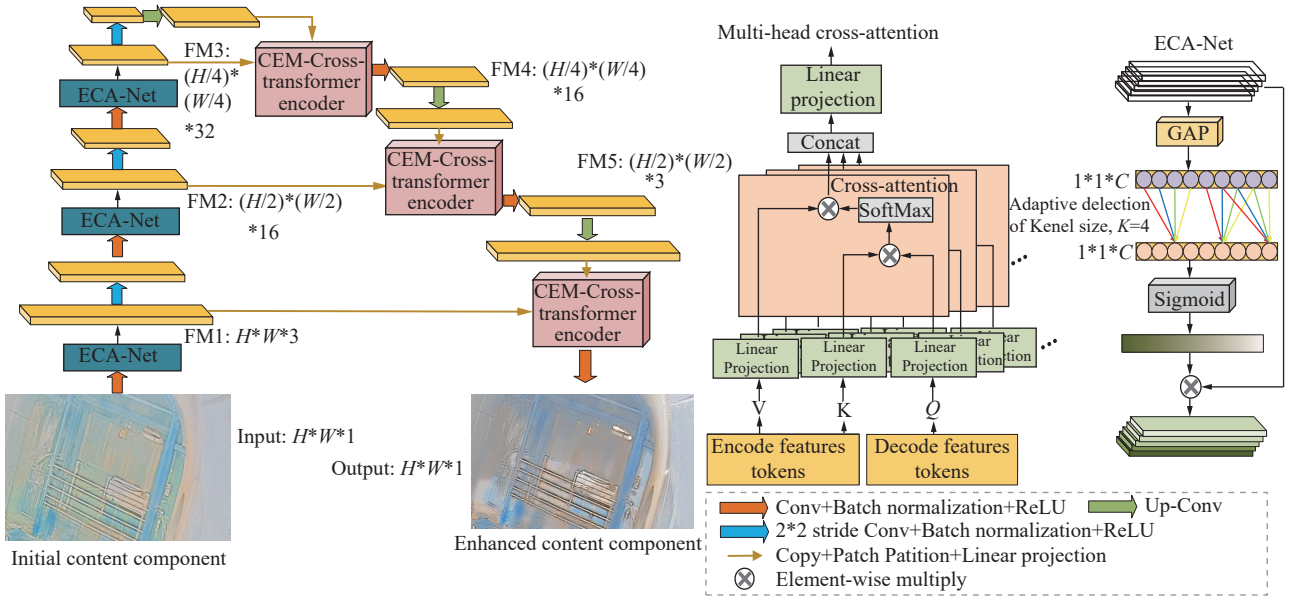


图 6 多尺度特征融合网络结构

Fig.6 Structure of MFNet

首先对输入特征图进行信道级全局平均池化 (GAP, Global Average Pooling) 从而得到各通道特征图的全局信息, 见式(4); 再根据特征向量的通道维数确定 K 值, 再使用一维卷积和激活函数进行特征权重计算, 见式(5); 最后将各通道权重与原始特征图的特征通道进行对应相乘, 从而加强对重要特征的关注度, 见式(6)。

$$x_{\text{GAP}} = \sum_{x=1}^H \sum_{y=1}^W x_{\text{Input}}(x, y) / H * W \quad (4)$$

式中, x_{Input} 为单通道输入特征图; $H*W$ 为特征图尺寸; x_{GAP} 为全局池化结果。

$$W = \text{Sigmoid}(\text{CNN}_{1D-k}(x_{\text{GAP}})) \quad (5)$$

式中, W 为包含各通道特征权重的权重向量; Sigmoid 为激活函数; CNN_{1D-k} 为一维卷积; k 为卷积核大小, x_{GAP} 为一维特征向量。

$$X_{\text{out}} = X_{\text{Input}} + W X_{\text{Input}} \quad (6)$$

式中, X_{Input} 为输入特征图; X_{out} 为经 ECA-Net 进行信道级重要性划分的特征图。

1.5 损失函数

本文模型的损失函数包括生成对抗损失和图像增强损失, 生成对抗损失 L_{GAN} 包括 L_{GANI} 、 L_{GANg} , 分别对应局部、全局判别器损失函数; 图像增强损失函数包含特征解耦损失 (L_{FD} , Feature Decoupling Loss) 结构一致性损失 (L_{SC} , Structural Consistency Loss)、照度平滑性损失 (L_{IS} , Illuminance Smoothness Loss):

$$L_{\text{total}} = L_{\text{GANI}} + L_{\text{GANg}} + L_{\text{SC}} + L_{\text{IS}} + L_{\text{FD}} \quad (7)$$

本文判别器网络使用全局-局部判别器结构, 全局判别器通过鉴别图像整体特征风格, 能够增强全局图像的亮度、颜色特征, 局部判别器作用于生成图像的随机图像块, 能够避免细节信息丢失和光照不均现象。两个判别器均使用相对判别损失^[35], 能够使生成器快速收敛, 避免由于生成器体积过大影响训练效率, 增强训练稳定性:

$$L_{\text{GANI}} = L_{\text{GANg}} = -E_{x_r \sim P_{\text{real}}, x \sim P_{\text{data}}} [\log \sigma(D(x_r) - D(G(x)))] - E_{x_r \sim P_{\text{real}}, x \sim P_{\text{data}}} [\log \sigma(D(G(x)) - D(x_r))] \quad (8)$$

其中, x_r 为真实图像; x 为输入图像; E 为数据期望值; P_{data} 与 P_{real} 为输入图像与真实图像的数据分布; G 为生成器; D 为判别器; σ 为激活函数 Sigmoid。

特征解耦损失 L_{FD} 根据 Retinex 理论通过在图像解耦过程中, 将低照度图像与其参考图像的反射分量保持一致, 从而实现原始图像中反射分量与亮度分量的分离:

$$L_{\text{FD}} = |R_L - R_N| + |R_L I_N - S_N| + |R_N I_L - S_L| \quad (9)$$

式中, S_L 、 I_L 、 R_L 分别为低照度图像及其亮度分量、反射分量; S_N 、 I_N 、 R_N 分别为正常照度图像的参考图像及其亮度分量、反射分量。

结构一致性损失 L_{SC} 的设计受文献^[36]中 Perceptual Losses 的启发, 使用特征提取网络 VGG-19 提取增强前后的反射分量的轮廓、结构特征和语义信息, 再将两者趋向一致, 从而保持原始图像在增强前后的空间结构一致性:

$$L_{\text{SC}} = |f(R') - f(R)| + \frac{1}{N} \sum_{k=1}^N |f(R'_k) - f(R_k)| \quad (10)$$

式中, R 和 R' 分别为增强前后反射分量; R_k 为局部区域反射分量; 由于 VGG-19 的第 4 层 conv3-256 输出的特征图兼顾结构轮廓特征及语义信息, 因此 L_{sc} 使用该卷积层输出特征图进行差异性比较, $f(\cdot)$ 表示输出特征图; N 为滑动框平移次数, 本文设置滑动框尺寸为 2×2 , 则 $N = (\text{width}/2) \times (\text{hight}/2)$, width 与 hight 表示特征图的宽和高。

亮度平滑性损失 L_{ls} 根据反射分量的图像信息对边缘轮廓、细节区域的亮度分量作自适应平滑度调节。在局部细节区域, 增大对亮度分量的平滑强度, 从而使原图中对应区域的亮度值更均匀; 在边缘轮廓区域施加低强度亮度分量平滑处理, 有利于保持原始图像中物体的边缘轮廓。

$$L_{ls} = \text{Exp}(-(\nabla R_k)^2) \nabla I_k \quad k \in N \quad (11)$$

$$\nabla R = \sqrt{(\nabla R_x)^2 + (\nabla R_y)^2} \quad (12)$$

$$\nabla I = \sqrt{(\nabla I_x)^2 + (\nabla I_y)^2} \quad (13)$$

其中, N 为滑动框平移次数, 滑动框与式(9)相同; ∇R 为该区域反射分量的梯度值; ∇R_x 与 ∇R_y 分别为水平和垂直梯度; I 为该区域的亮度分量。

2 试验分析

2.1 试验准备

2.1.1 试验数据集

文中矿井低照度图像数据集采集自鄂尔多斯马泰壕矿井低光/正常光监视影像, 该数据集共有 1 495 组图像对, 其中 1 095 组为原始图像对, 400 组图像对是通过翻转、随机裁剪等数据增强方法生成。随机选取 1 300 组图像对作为训练集, 195 组图像对为测试集。此外, 选取公共低照度图像数据集 LOL^[15]、SCIE^[14]、SID^[37] 对本文模型进行性能测试。其中, SCIE、SID 中配对图像的目标图像选取长曝光参考图像; SID 图像格式由 RAW 格式转换为 JPG 格式的 RGB 图像。

为满足无监督网络模型的训练要求, 文章将矿井低照度图像数据集的训练集中低照度图像和正常照度图像平均分为两组, 再交叉组合, 从而形成两组非配对图像子数据集, 在训练过程中将两个子数据集顺序输入模型。公共低照度图像数据集的非配对版本制作方法同上。

2.1.2 仿真试验环境

试验设备为 Windows7+NVIDIA GTX 960 GPU 的计算机。使用 Adam 优化器, 学习率为 10^{-4} , 矩估

计指数衰减率分别 β_1 、 β_2 为 0.9、0.999。BatchSize 设置为 8, 训练轮数为 240 个 epoch。

采用全参考图像质量评估指标^[38]: 峰值信噪比 (Peak-Signal to Noise Ratio, PSNR)、结构相似度 (Structure Similarity, SSIM)、视觉信息保真度 (Visual Information Fidelity, VIF) 对增强图像进行评估。三类指标均能感知对比图像之间的照度、颜色、细节、结构特征的变化, 但 PSNR 主要反映颜色特征的增强效果, 颜色失真程度; SSIM 能够评估增强图像的亮度特征, 评估整体图像亮度提升水平, 亮度分布合理性及局部区域是否存在阴影和过曝; VIF 评估增强图像中细节、纹理信息保留度。三类指标通过均值±方差的数据形式显示评估结果, 数值越高, 表示增强图像质量越好。

2.2 消融试验

2.2.1 网络结构消融试验

为验证本文模型中特征编码器结构、生成对抗网络框架、CEM-Transformer Encoder 和 ECA-Net、CEM-Cross-Transformer Encoder 的有效性, 本文通过设计特征编码器、CEM-Transformer Encoder 的对比网络结构, 删减判别器和 ECA-Net、CEM-Cross-Transformer Encoder, 对各模块进行有效性评估, 其中, 特征编码器的对比网络结构使用 RetinexNet^[15] 的 DecomNet, 或者去除特征编码器直接使用 MFNet 和 GLENet 依次增强原始图像; CEM-Transformer Encoder 的对比网络结构分别为深度可分离卷积块 (DS-Conv+Batch Normalization+Leaky ReLU), 或 ViT 的 Transformer Encoder。试验中, 除待评估模块外, 其他模块的损失函数及参数设置不变, 测试结果见表 1、表 2, 最优结果被标记为粗体, A w B 表示 A 模型使用 B 结构, A w/o B 表示 A 模型去除 B 结构, 消融试验展示如图 7、图 8 所示。

由表 1、表 2 可知, 对原始模型进行改动均会造成全部图像质量指标的下降。当去除判别器时, 增强图像的所有质量指标均大幅下降, 说明判别器对生成器具有显著的监督作用, 生成对抗网络结构能实现模型的无监督训练。当使用对比网络结构替换 CEM-Transformer Encoder 时, SSIM 下降幅度较大, 证明 CEM-Transformer Encoder 通过在亮度分量中捕获全局性依赖关系, 并提取局部区域低级特征, 能够有效提升整体图像亮度, 避免局部区域的光照不均现象。当去除 ECA-Net 时, VIF 及 PSNR 下降明显, 证明 ECA-Net 能够加强浅层网络对低级特征的提取效率。当去除 CEM-Cross-Transformer Encoder

表 1 基于矿井低照度图像数据集的网络结构消融试验结果

Table 1 Ablation experimental results of network structures on coal mine low-light image dataset

待测对象	模型结构	评估指标 (均值 ± 方差)		
		VIF	PSNR	SSIM
生成对抗网络框架	本文模型	3.821±0.055	24.37±0.45	0.749±0.016
	本文模型 w/o 判别器	2.845±0.193	17.65±1.86	0.562±0.094
CEM-Transformer Encoder	本文模型 w 深度可分离卷积块	3.687±0.083	23.83±0.67	0.652±0.029
	本文模型 w Transformer Encoder	3.556±0.105	23.25±0.61	0.682±0.035
ECA-Net	本文模型 w/o ECA-Net	3.461±0.072	22.19±0.94	0.715±0.019
CEM-Cross-Transformer Encoder	本文模型 w/o CEM-Cross-Transformer Encoder	3.519±0.058	22.76±0.47	0.693±0.047
特征编码器	本文模型 w DecomNet	3.762±0.069	23.13±0.32	0.705±0.013
	本文模型 w/o 特征编码器	3.141±0.279	19.52±1.09	0.613±0.041

表 2 基于 LOL 的网络结构消融试验结果

Table 2 Ablation experimental results of network structures on LOL

待测对象	模型结构	评估指标 (均值 ± 方差)		
		VIF	PSNR	SSIM
生成对抗网络框架	本文模型	3.673±0.032	23.61±0.29	0.715±0.011
	本文模型 w/o 判别器	2.752±0.118	15.63±1.02	0.551±0.048
CEM-Transformer Encoder	本文模型 w 深度可分离卷积块	3.466±0.039	22.18±0.21	0.634±0.013
	本文模型 w Transformer Encoder	3.419±0.053	21.74±0.41	0.672±0.014
ECA-Net	本文模型 w/o ECA-Net	3.537±0.038	21.86±0.18	0.681±0.011
CEM-Cross-Transformer Encoder	本文模型 w/o CEM-Cross-Transformer Encoder	3.475±0.094	22.03±0.47	0.656±0.035
特征编码器	本文模型 w DecomNet	3.488±0.034	21.79±0.32	0.687±0.009
	本文模型 w/o 特征编码器	3.231±0.087	18.47±0.63	0.601±0.029

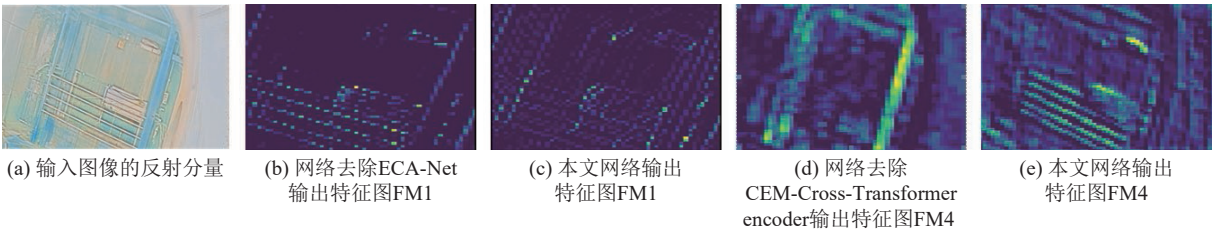


图 7 多尺度特征融合网络结构消融试验结果

Fig.7 Ablation experiment results of MFNet structure

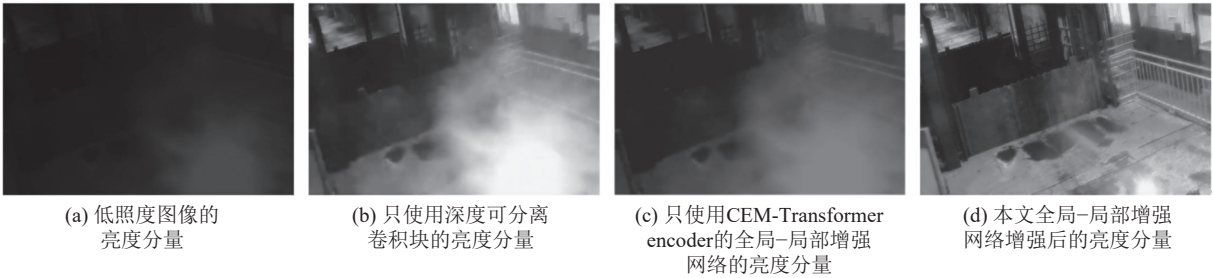


图 8 全局-局部增强网络结构消融试验结果

Fig.8 Ablation experiment results of GLENet structure

时, VIF 及 PSNR 大幅下降, 证明 ECA-Net 通过抑制非重要特征, 强调重要特征, 能够在较深网络层处, 实现低级特征的合理复用, 有利于增强图像的细节、颜色特征。当去除特征编码器时, 3 类评价指标均显

著下降,且PSNR下降幅度较大,证明特征解耦能有效解决颜色失真,并且多通道增强模式通过针对特定分量进行增强能提升图像增强效果;当使用DecomNet进行特征解耦时,全体评价指标均下降,证明根据亮度分量和反射分量特性搭建的特征编码器的解耦效果更好。

相比于图7b—7e中语义特征更显著,并且图7e包含更多的低维特征,细节特征明显,证明CEM-Cross-Transformer Encoder能够有效实现多尺度特征的自适应融合。图8b的亮度提升明显,但全局亮度提升的同时,局部光晕同样得到增强,图8c的亮度较低且光晕未消除,相比之下,图8d的亮度提升明显,局部光晕缩小的同时,相邻区域的亮度得到自适应增强,证明全局—局部增强网络结构能够提升全局图像亮度,且对各局部区域的亮度实现自适应增强,有效避免棋盘效应。

2.2.2 损失函数消融试验

设计消融试验评估各类损失函数的有效性,试验结果见表3、表4,最优值标记为粗体形式。

表3 基于矿井低照度图像数据集的损失函数消融试验结果
Table 3 Ablation experimental results of loss functions on the coal mine low-light image dataset

损失函数	评估指标(均值±方差)		
	VIF	PSNR	SSIM
Total loss	3.821±0.055	24.37±0.45	0.749±0.016
Total loss w/o L_{sc}	3.246±0.052	22.27±0.84	0.659±0.023
Total loss w/o L_{is}	3.579±0.097	23.76±0.53	0.561±0.087
Total loss w/o L_{fd}	3.424±0.084	22.63±0.47	0.638±0.051
Total loss w/o $L_{GANI}+L_{GANg}$	2.752±0.118	15.63±1.02	0.551±0.048

表4 基于LOL的损失函数消融试验结果
Table 4 Ablation Experimental results of loss functions on LOL

损失函数	评估指标(均值±方差)		
	VIF	PSNR	SSIM
Total loss	3.673±0.032	23.61±0.29	0.715±0.011
Total loss w/o L_{sc}	3.272±0.081	22.87±0.23	0.624±0.017
Total loss w/o L_{is}	3.317±0.053	22.45±0.17	0.609±0.052
Total loss w/o L_{fd}	3.409±0.042	22.16±0.38	0.617±0.044
Total loss w/o $L_{GANI}+L_{GANg}$	2.752±0.118	15.63±1.02	0.551±0.048

由表3、表4可知,当去除 L_{sc} 时,增强图像的VIF大幅下降,证明 L_{sc} 能有效保留大量图像细节信息。当去除 L_{is} 时,增强结果SSIM有显著下降,VIF

数值轻微下降,说明 L_{is} 能合理增强图像的亮度分布,使结构轮廓特征更明显,局部细节区域亮度表现得更自然。当去除 L_{fd} 时,增强图像的VIF、PSNR、SSIM均下降,证明 L_{fd} 能实现亮度分量和反射分量的有效分解,有利于后续对两分量进行特定增强。模型训练过程中不使用 L_{GANI} 、 L_{GANg} ,将导致判别器失效,此时所有图像质量评估指标均大幅下降,说明若不使用生成对抗主体框架,模型将无法实现无监督训练。

2.3 对比试验

使用5种先进低照度图像增强算法RetinexNet^[15]、MSRCR^[13]、ExCNet^[39]、MBLLEN^[40]、EnlightenGAN^[20]作为对比算法与文中算法进行图像增强对比试验。再对试验结果进行主观评价和基于全参考图像质量指标的客观评估,以测试本文算法的低照度图像增强性能。

2.3.1 对比试验的主观评价

使用矿井低照度图像数据集及公共低照度图像数据集LOL^[15]、SCIE^[14]、SID^[37]对本文模型性能进行主观评价。试验结果如图9所示,其中第1行为低照度图像,前3个场景从矿井低照度图像数据集中选取,其余3个场景分别从3种公共数据集中随机选取,第2行是经RetinexNet增强后的结果,第3行到第7行依次分别是经MSRCR、ExCNet、MBLLEN、EnlightenGAN和本文算法增强的结果。使用HSV-3D直方图来直观展示某场景的增强结果,如图10所示,其中,Hue表示色调,Saturation表示颜色饱和度,Value表示明度。

由图9可知,RetinexNet能够大幅提升图像亮度,显著提高对比度,但在改善亮度的同时,易造成不同程度的颜色失真,例如,场景1中红色电子指示灯颜色失真,易导致误判;场景2中,局部区域照度不均,光源周围区域偏黄色;场景3中输送带滚筒区域颜色失真,易造成对运载物体的误判;场景4、5、6中包含大量颜色鲜艳的物体,当发生颜色失真时,图像的颜色饱和度过高,致使整体偏红色,影响图像的视觉效果。MSRCR能够提升图像的全局对比度,但是亮度提升能力有限,导致图像信息大量丢失,并且由于场景1、2、3环境较暗,使细节损失尤为严重,例如场景1中墙壁区域不清晰,场景3中输送带左侧区域部分细节不可见,且输送带滚筒区域细节丢失严重。经ExCNet增强的图像,亮度、对比度提升有限,图像整体偏灰,见场景1、2、3,且场景2中光源区域有过曝现象;并且对于包含鲜艳颜色物体的图像,如场景

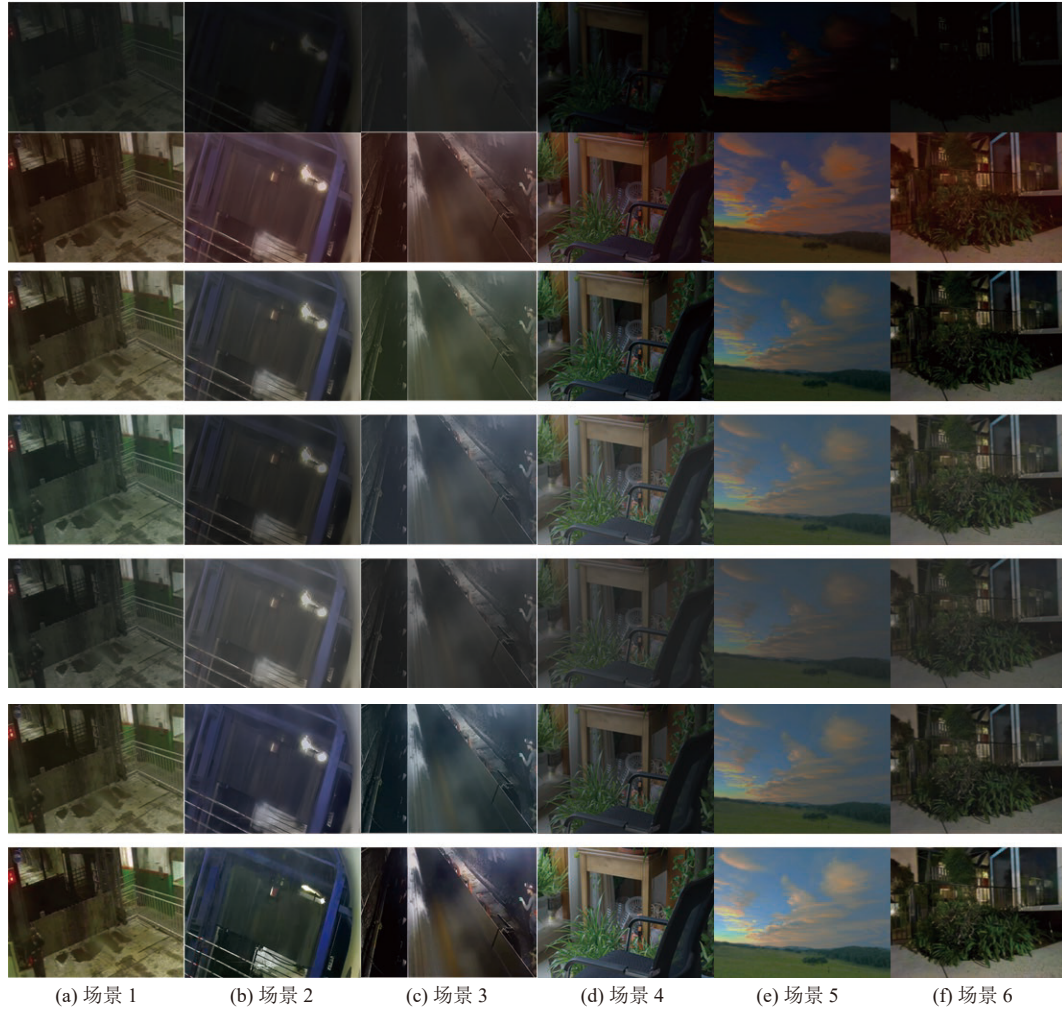


图 9 不同算法在各类数据集上增强结果

Fig.9 Experiment results of different algorithms on all datasets

4、5、6, 亮度过低导致物体颜色暗淡, 视觉效果较差。经 MBLLN 增强的图像, 亮度提高不明显, 对于从矿井环境中采集的极低照度图像, MBLLN 无法有效提升其对比度, 导致阴影、光晕现象严重, 各类目标颜色偏黑色, 影响后续对物体状态的判断。EnlightenGAN 能有效提升图像的亮度和对比度, 但会导致颜色不均衡, 并且对于极低照度图像, 增强后颜色失真更严重, 如场景 3, 增强结果整体呈蓝色, 且场景 2 中光源区域光晕较大。相比之下, 本文算法增强后的图像, 亮度提升明显, 无光照不均, 例如场景 2 的强光源区域光晕显著缩小、细节特征明显; 且原始极低照度图像在增强后无颜色失真现象, 图像信息保留度高, 视觉效果较好, 例如, 场景 6 中绿植在亮度提升后仍保持色彩均衡, 场景 3 中输送带滚筒区域色彩均衡, 细节清晰可见, 图像信息无明显丢失。在图 10 中, 图像依次为 RetinexNet^[15]、MSRCR^[13]、ExCNet^[39]、MBLLN^[40]、EnlightenGAN^[20] 的增强图像, 及增强图像对应的 HSV-3D 直方图, 由图 10 可

见, 相比于其它算法, 本文算法增强图像在 H、S、V 三通道中分布于更高灰度级, 且各通道的动态范围更大, 进一步直观地展现了本文算法的出色性能。

2.3.2 对比试验的客观评价

在矿井低照度图像数据集及 LOL 数据集中分别进行对比试验, 使用 PSNR、SSIM 和 VIF 对增强图像进行质量评估, 并测试各模型的推理耗时, 以均值±方差的形式展示评估结果, 见表 5、表 6, 最优结果被标记为粗体。

表 5、表 6 分别为各类算法在矿井低照度图像数据集、公共数据集中对比试验结果。算法推理速度方面, EnlightenGAN 的推理耗时最少, ExCNet 由于图像迭代优化过程较复杂所以耗时较长, 本文方法相较于多数对比算法具有较好的推理速度, 后续将进一步提升算法的推理效率, 以更好地实现实时图像增强。图像质量指标方面, RetinexNet 的 PSNR 和 VIF 值较低, 说明其增强图像颜色失真严重; MBLLN 的 SSIM 和 VIF 值过低, 证明其亮度增强

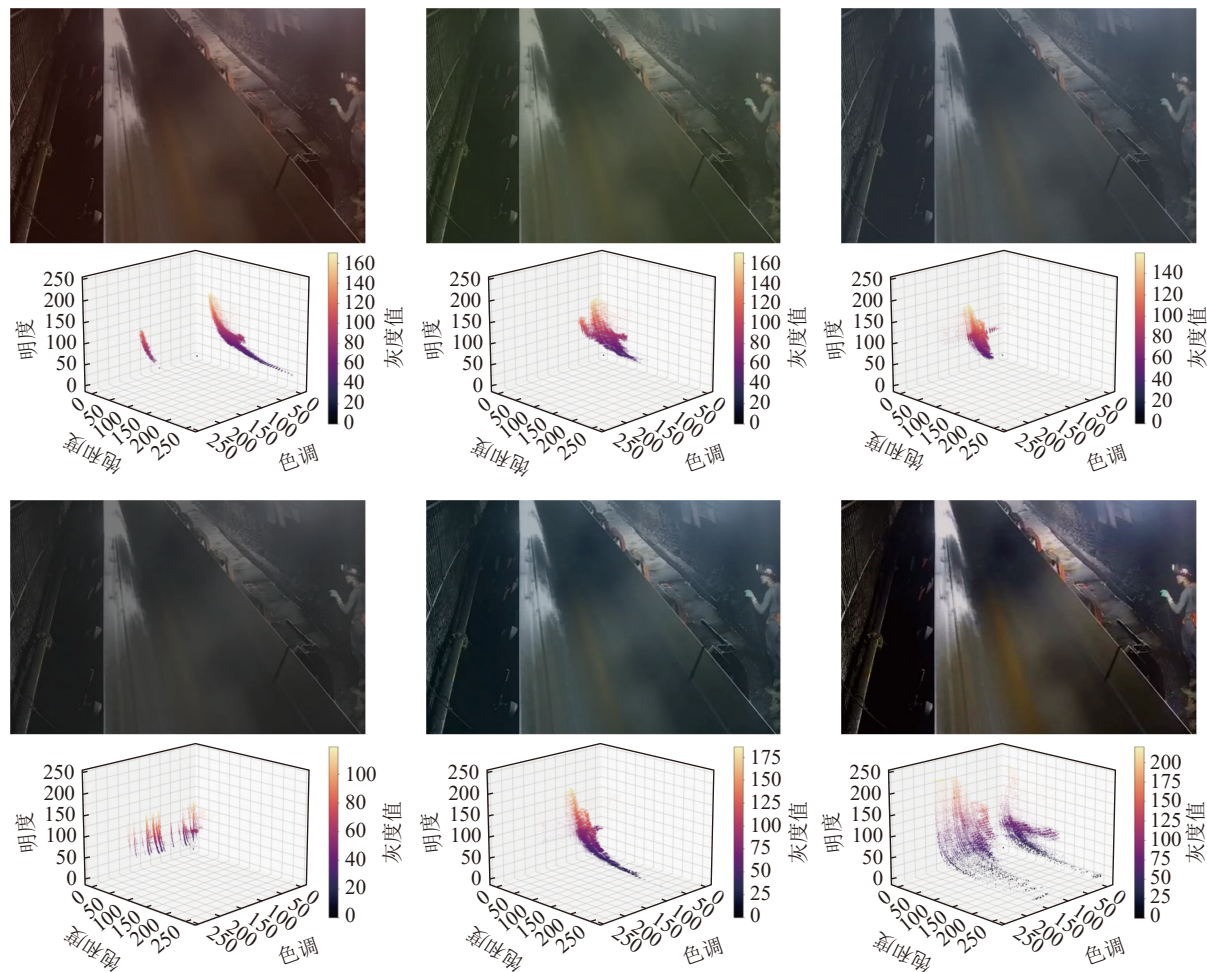


图 10 典型增强图像的 HSV 直方图结果

Fig.10 HSV histogram results of typical enhanced images

表 5 不同算法在矿井低照度图像数据集上客观评估

Table 5 Objective evaluation of different algorithms on the coal mine low-light image dataset

模型	评估指标 (均值 ± 方差)			运行时间
	PSNR	SSIM	VIF	
RetinexNet ^[15]	19.96±1.01	0.682±0.059	3.181±0.311	0.227±0.011
MSRCR ^[13]	21.87±2.14	0.658±0.072	3.413±0.421	10.094±0.573
ExCNet ^[39]	20.51±1.95	0.692±0.097	3.353±0.181	24.638±1.942
MBLEN ^[40]	20.92±2.53	0.623±0.121	3.020±0.219	15.334±1.165
EnlightenGAN ^[20]	21.38±2.19	0.728±0.052	3.519±0.328	0.011±0.001
本文模型	24.37±0.45	0.749±0.016	3.821±0.055	1.213±0.106

表 6 不同算法在 LOL 上客观评估

Table 6 Objective evaluation of different algorithms on LOL

模型	评估指标 (均值 ± 方差)			运行时间
	PSNR	SSIM	VIF	
RetinexNet ^[15]	19.33±1.63	0.653±0.039	3.015±0.133	0.168±0.008
MSRCR ^[13]	20.85±1.87	0.615±0.043	3.332±0.340	9.131±0.514
ExCNet ^[39]	20.34±1.77	0.632±0.107	3.265±0.229	24.638±1.942
MBLEN ^[40]	21.17±2.12	0.628±0.094	2.995±0.294	13.792±0.962
EnlightenGAN ^[20]	21.62±2.08	0.703±0.036	3.415±0.315	0.015±0.001
本文模型	23.61±0.29	0.715±0.011	3.673±0.032	1.155±0.078

能力较差,无法有效恢复阴影中的图像内容,易造成增强图像的结构特征与图像细节严重损失;本文算法在全体评价指标中表现较好,在矿井低照度图像数据集中相较于对比算法,本文算法增强图像的 PSNR 值、SSIM 值和 VIF 值平均提高了 16.564%, 10.998%, 16.226%, 在公共数据集中, PSNR 值、SSIM 值和 VIF 值平均提高了 14.438%, 10.888%, 14.948%, 证明本文算法能有效提升低照度图像的亮度,消除照度不均,增强颜色特征,避免颜色失真,保留大量细节信息,实现矿井低照度图像增强。

3 结 论

1) 基于生成对抗思想设计的生成对抗式主体网络架构使用目标图像域而非单一参考图像监督模型训练,有利于低照度图像的充分增强,避免了由于少数参考图像质量较低而影响模型的图像增强性能。

2) 基于特征表示学习理论搭建特征编码器将图像解耦为反射分量和亮度分量,有效解决了图像增强过程中亮度与颜色特征相互影响导致的颜色失真问题。

3) 双通道并行增强网络根据亮度分量和反射分量的特点采用特定方法对二者进行有效增强。设计 CEM-Transformer Encoder 通过在亮度分量中捕获长程依赖和局部上下文关系,能够在充分提升整体图像亮度的同时,解决照度不均问题,消除局部区域的阴影和过曝;在多尺度特征融合网络中使用 ECA-Net 能够提升浅层网络的特征提取效率,使用融合 CEM-Cross-Transformer Encoder 的跳跃连接能够实现多尺度特征的自适应融合,有效避免细节特征丢失。

4) 文中算法在自制的矿井低照度图像数据集和公共低照度图像数据集中均有出色表现,在两类数据集中,相比于 5 种先进的低照度图像增强算法,文中算法增强图像的质量指标 PSNR、SSIM、VIF 平均提高了 16.564%, 10.998%, 16.226% 和 14.438%, 10.888%, 14.948%, 证明文中算法能够有效提升矿井低照度图像的亮度,消除照度不均,增强颜色特征,避免细节信息丢失,实现低照度图像质量增强。后续将着重进行轻量化模型的研究,尝试将算法部署到硬件设备进行实时图像增强,并且思考如何将图像增强与目标检测、图像分割等任务相结合,从而使图像增强模块在提升图像质量的同时,能够有效辅助高级视觉任务。

参考文献(References):

[1] 王 国 法, 王 虹, 任 怀 伟, 等. 智慧煤矿 2025 情景目标和发展路

径[J]. 煤炭学报, 2018, 43(2): 295–305.

WANG Guofa, WANG Hong, REN Huaiwei, *et al.* 2025 scenarios and development path of intelligent coal mine[J]. Journal of China Coal Society, 2018, 43(2): 295–305.

[2] 陈 伟, 任 鹏, 田子建, 等. 基于注意力机制的无监督矿井人员跟踪[J]. 煤炭学报, 2021, 46(S1): 601–608.

CHEN Wei, REN Peng, TIAN Zijian, *et al.* Unsupervised mine personnel tracking based on attention mechanism[J]. Journal of China Coal Society, 2021, 46(S1): 601–608.

[3] 韩江洪, 卫 星, 陆 阳, 等. 煤矿井下机车无人驾驶系统关键技术[J]. 煤炭学报, 2020, 45(6): 2104–2115.

HAN Jianghong, WEI Xing, LU Yang, *et al.* Driverless technology of underground locomotive in coal mine[J]. Journal of China Coal Society, 2020, 45(6): 2104–2115.

[4] 杨 潇, 陈 伟, 任 鹏, 等. 基于域适应的煤矿环境监控图像语义分割[J]. 煤炭学报, 2021, 46(10): 3386–3396.

YANG Xiao, CHEN Wei, REN Peng, *et al.* Coal mine monitoring image semantic segmentation based on domain adaptation[J]. Journal of China Coal Society, 2021, 46(10): 3386–3396.

[5] 郝 帅, 张 旭, 马 旭, 等. 基于 CBAM-YOLOv5 的煤矿输送带异物检测[J/OL]. 煤炭学报: 1–11[2024-04]. DOI: [10.13225/j.cnki.jccs.2021.1644](https://doi.org/10.13225/j.cnki.jccs.2021.1644).

HAO Shuai, ZHANG Xu, MA Xu, *et al.* Foreign object detection in coal mine conveyor belt based on CBAM-YOLOv5[J/OL]. Journal of China Coal Society: 1–11[2022-04-04]. DOI: [10.13225/j.cnki.jccs.2021.1644](https://doi.org/10.13225/j.cnki.jccs.2021.1644).

[6] 单鹏飞, 孙浩强, 来兴平, 等. 基于改进 Faster R-CNN 的综放煤矸混合放出状态识别方法[J]. 煤炭学报, 2022, 47(3): 1382–1394.

SHAN Pengfei, SUN Haoqiang, LAI Xingping, *et al.* Identification method on mixed and release state of coal-gangue masses of fully mechanized caving based on improved Faster R-CNN[J]. Journal of China Coal Society, 2022, 47(3): 1382–1394.

[7] 程德强, 钱建生, 郭星歌, 等. 煤矿安全生产视频 AI 识别关键技术研究综述[J]. 煤炭科学技术, 2023, 51(2): 349–365.

CHENG Deqiang, QIAN Jiansheng, GUO Xingge, *et al.* Review on key technologies of AI recognition for videos in coal mine[J]. Coal Science and Technology, 2023, 51(2): 349–365.

[8] KIM Yeong-Taeg. Contrast enhancement using brightness preserving bi-histogram equalization[J]. *IEEE transactions on Consumer Electronics*, 1997, 43(1): 1–8.

[9] WANG Yu, CHEN Qian, ZHANG Baomin. Image enhancement based on equal area dualistic sub-image histogram equalization method[J]. *IEEE transactions on Consumer Electronics*, 1999, 45(1): 68–75.

[10] WANG Jianwei. An enhancement algorithm for low-illumination color image with preserving edge[J]. *Computer Technology Development*, 2018, 28(1): 116–120.

[11] LIU Lizhu, WANG Haiying. Image enhancement using a nonlinear method with an improved Single-Scale Retinex algorithm[C]//2011 International Conference on Electronics, Communications and Control (ICECC), 2011: 2086–2089.

[12] RAHMAN Ziaur, JOBSON Daniel J, WOODSELL Glenn A. Multi-scale retinex for color image enhancement[C]//Proceedings of 3rd IEEE International Conference on Image Processing, 1996, 3: 1003–1006.

- [13] JIANG Xingfang, WANG Ge, SHEN Weimin. A method of color image enhancement using color advanced retinex[J]. *Journal of Optoelectronics. Laser*, 2008, 19(10): 1402–1404.
- [14] LORE Kin Gwn, AKINTAYO Adedotun, SARKAR Soumik. LL-Net: a deep autoencoder approach to natural low-light image enhancement[J]. *Pattern Recognition*, 2017, 61: 650–662.
- [15] WEI Chen, WANG Wenjing, YANG Wenhan, *et al.* Deep Retinex Decomposition for Low-Light Enhancement[J]. *arXiv preprint arXiv: 1808.04560*, 2018.
- [16] GOODFELLOW Ian, POUGET-ABADIE Jean, MIRZA Mehdi, *et al.* Generative adversarial nets[J]. *Advances in Neural Information Processing Systems*, 2014, 27.
- [17] CHOI Yunje, CHOI Minje, KIM Munyoung, *et al.* Stargan: Unified generative adversarial networks for multi-domain image-to-image translation [C]//In Proceedings of the IEEE conference on computer vision and pattern recognition, 2018: 8789–8797.
- [18] CHEN Xi, DUAN Yan, HOUTHOOFT Rein, *et al.* InfoGAN: interpretable representation learning by info-ration maximizing generative adversarial nets[C]//Conference and Workshop on Neural Information Processing Sys-tems. Barcelona: MIT Press, 2016, 2180–2188.
- [19] ZHU Junyan, PARK Taesung, ISOLA Phillip, *et al.* Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision, 2017: 2223–2232.
- [20] JIANG Yifan, GONG Xinyu, LIU Ding, *et al.* Enlightengan: Deep light enhancement without paired supervision[J]. *IEEE transactions on image processing*, 2021, 30: 2340–2349.
- [21] LIU Siyuan, THUNG Kim-Han, QU Liangqiong, *et al.* Learning MRI artefact removal with unpaired data[J]. *Nature Machine Intelligence*, 2021, 3(1): 60–67.
- [22] WANG Qilong, WU Banggu, ZHU Pengfei, *et al.* ECA-Net: Efficient channel attention for deep convolutional neural networks[C]. *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020: 11534–11542.
- [23] VASWANI Ashish, SHAZEER Noam, PARMAR Niki, *et al.* Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30.
- [24] FRAN C. Deep learning with depth wise separable convolutions[C]//IEEE conference on computer vision and pattern recognition (CVPR), 2017: 1251–1258.
- [25] MIYATO Takeru, KATAOKA Toshiki, KOYAMA Masanori, *et al.* Spectral normalization for generative adversarial networks[J]. *arXiv preprint arXiv: 1802.05957*, 2018.
- [26] CHEN Xi, DUAN Yan, HOUTHOOFT Rein, *et al.* Infogan: Interpretable representation learning by information maximizing generative adversarial nets[J]. *Advances in Neural Information Processing Systems*, 2016, 29.
- [27] HIGGINS Irina, MATTHEY Loic, PAL Arka, *et al.* beta-vae: Learning basic visual concepts with a constrained variational framework[C]//International Conference on Learning Representations, 2016.
- [28] RADFORD Alec, METZ Luke, CHINTALA Soumith. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. *arXiv preprint arXiv: 1511.06434*, 2015.
- [29] 张雅荔, 李文元, 李昌禄, 等. 融合注意力引导的多尺度低照度图像增强方法 [J/OL]. *西安电子科技大学学报*: 1–9[2022-11-30]. <http://kns.cnki.net/kcms/detail/61.1076.TN.20221129.0832.004.html>.
ZHANG Yali, LI Wenyuan, LI Changlu, *et al.* Method for enhancement of the multi-scale low-light image by combining an attention guidance[J/OL]. *Journal Of XiDian University*: 1–9[2022-11-30]. <http://kns.cnki.net/kcms/detail/61.1076.TN.20221129.0832.004.html>.
- [30] 陈 勇, 陈 东, 刘焕淋, 等. 基于深度卷积神经网络的无参考低照度图像增强 [J]. *电子与信息学报*, 2022, 44(6): 2166–2174.
CHEN Yong, CHEN Dong, LIU Huanlin, *et al.* Unreferenced Low-lighting Image Enhancement Based on[J]. *Journal of Electronics & Information Technology Deep Convolutional Neural Network*, 2022, 44(6): 2166–2174.
- [31] 李 烁, 王 慧, 耿则勋, 等. 双范数混合约束的遥感影像亮度不均变分校正[J]. *测绘学报*, 2018, 47(12): 1621–1629.
LI Shuo, WANG HUI, GENG Zexun, *et al.* Variational uneven illumination correction with double-norm hybrid constraints for remote sensing imagery[J]. *Acta Geodaetica et Cartographica Sinica*, 2018, 47(12): 1621–1629.
- [32] 王冬云, 唐 楚, 鄂世举, 等. 基于导向滤波 Retinex 和自适应 Canny 的图像边缘检测 [J]. *光学精密工程*, 2021, 29(2): 443–451.
WANG Dongyun, TANG Chu, E Shiju, *et al.* Image edge detection based on guided filter Retinex and adaptive Canny[J]. *Optics and Precision Engineering*, 2021, 29(2): 443–451.
- [33] KHAN Salman, NASEER Muzammal, HAYAT Munawar, *et al.* Transformers in vision: a survey[J]. *ACM Computing Surveys (CSUR)*, 2022, 54(10): 1–41.
- [34] DOSOVITSKIY Alexey, BEYER Lucas, KOLESNIKOV Alexander, *et al.* An image is worth 16x16 words: Transformers for image recognition at scale[J]. *arXiv preprint arXiv: 2010.11929*, 2020.
- [35] JOLICOEUR-MARTINEAU Alexia. The relativistic discriminator: a key element missing from standard GAN[J]. *arXiv preprint arXiv: 1807.00734*, 2018.
- [36] JOHNSON Justin, ALAHI Alexandre, FEI-FEI Li. Perceptual losses for real-time style transfer and super-resolution[C]//European conference on computer vision. Springer, Cham, 2016: 694–711.
- [37] CHEN Chen, CHEN Qifeng, XU Jia, *et al.* Learning to see in the dark[C]//In IEEE Conference on Computer Vision and Pattern Recognition, 2018: 3291–3300.
- [38] WANG Zhou, SIMONCELLI Eero P, BOVIK Alan C. Multiscale structural similarity for image quality assessment[C]//The Thirty Seventh Asilomar Conference on Signals, Systems & Computers, Pacific Grove, CA, USA, 2003: 1398–1402.
- [39] ZHANG Lin, ZHANG Lijun, LIU Xiao, *et al.* Zero-shot restoration of back-lit images using deep internal learning[C]//Proceedings of the 27th ACM international conference on multimedia, 2019: 1623–1631.
- [40] LYU Feifan, LU Feng, WU Jianhua, *et al.* MBLEN: low-light image/video enhancement using CNNs[J]. *BMVC*, 2018: 4.