Textile & Leather Review

ISSN 2623-6281 | www.tlr-journal.com | 10.31881/TLR

LPEnHancer: Laplacian Pyramid Enhancement Networks for Object Detection and Beyond Under Low-Light Vision

Fang Zheng

How to cite: Zheng F. LPEnHancer: Laplacian Pyramid Enhancement Networks for Object Detection and Beyond Under Low-Light Vision. Textile & Leather Review. 2026; 9:5444-5467. https://doi.org/10.31881/TLR.2026.5444

How to link: https://doi.org/10.31881/TLR.2026.5444

Published: 27 April 2026

LPEnHancer: Laplacian Pyramid Enhancement Networks for Object Detection and Beyond Under Low-Light Vision

Fang Zheng

School of Cyber Security and Computer Science, Hebei University, Baoding 071000, Hebei, China
20228019013@stumail.hbu.edu.cn

Article
https://doi.org/10.31881/TLR.2026.5444
Published 27 April 2026

ABSTRACT

Low-light conditions pose significant challenges for the extraction of visual features essential for downstream tasks. Previous research has attempted to enhance image representations either by correlating visual quality with machine perception or by designing illumination degradation transformation methods that necessitate pre-training on synthetic datasets. However, we propose that optimizing the enhanced image representation in relation to the specific loss functions of downstream tasks can yield more expressive representations. In this study, we introduce a novel network, LPEnHancer, which consists of two key modules. By utilizing the Laplacian pyramid, the input low-light images are decomposed into multiple scales, where the Detail Mining Module is used to excavate edge/texture features and enhance local features at each scale, and the Low Frequency Enhancement Filter is employed to suppress high-frequency noise. LPEnHancer is a versatile plug-and-play module that can be seamlessly integrated into any low-light vision pipeline. Our extensive experimental results demonstrate that the enhanced representations generated by LPEnHancer significantly and consistently improve performance across various low-light vision tasks, including dark object detection (3.0 mAP improvement on ExDark), dark face detection (2.0 mAP improvement on DARK FACE), nighttime semantic segmentation (0.78 mIoU improvement on ACDC), and dark instance segmentation (1.7 Bbox_mAP and 0.9 Segm_mAP improvements on LIS).

KEYWORDS

low-light object detection, plug-and-play, pyramid enhancement, multi-scale

INTRODUCTION

In recent years, the rapid development of convolutional neural networks (CNNs) and transformers has significantly improved the performance of advanced vision tasks. A large number of target detection models [1-3], semantic segmentation models [6-9], and instance segmentation models [10-12] have been introduced, and

satisfactory results have been achieved on benchmark datasets for different visual tasks. However, most of these models were trained under normal illumination conditions. In practical applications, there are instances of insufficient light due to the limited number of photons and the imperfections of photodetectors, which result in severe noise that “buries” the details of the object, leading to a degradation of image quality and thus limiting the real-world performance of the models.

Currently, numerous neural network-based low-light enhancement methods [13,14]have been proposed to improve the visual perception of low-light images by the human eye. The preprocessing of datasets with these advanced image enhancement techniques has the potential to enhance detection and segmentation performance under low light conditions. However, to build complex nonlinear mappings from low-quality images to their corresponding high-quality versions, many enhancement models require large model sizes, which can be detrimental to real-time detection when applied prior to the detector. Although some light-weight models offer shorter runtimes [15,16], they provide only limited improvements in target detection and semantic segmentation, as they are primarily designed for the human visual system. Another limitation is that many enhancement models are trained using an enhancement loss that measures the distance between the enhanced image and a clean real image. On one hand, clean ground-truth (GT) images may not be available in practical applications. On the other hand, such a loss function treats each pixel equally and does not prioritize structured features that are crucial for target detection and semantic segmentation. Furthermore, training with these two-stage methods can be challenging.

Inspired by these observations and influenced by recent advancements in low-light image enhancement methods (LLIE) and vision-based backbone networks, this paper aims to bridge the gap by exploring an end-to-end trainable enhancement module. This module is designed to jointly optimize the enhancement of downstream task objectives within a single network framework. To achieve this, we propose the Laplacian Pyramid Enhancement Networks (LPEnhancer), a feature enhancer dedicated to low-light images. LPEnHancer reduces noise induced by low-light conditions and learns enriched multiscale hierarchical features, which are beneficial for downstream visual tasks performed in low-light environments.

We combine the LPEnHancer module with RtmDet to build an end-to-end dark target detection, semantic segmentation, and instance segmentation framework that can be trained jointly. In LPEnHancer, we use learnable Gaussian convolution to decompose the image into multiple components with different resolutions for multi-scale feature enhancement while reducing the noise induced by low-light environments. At each

scale, we enhance the layered features using an information mining module, which exists in two branches; the texture information mining branch (TIM) uses pixel differential convolution [17], to capture detailed textures in low-light images, while the local feature enhancement branch (LFE) enhances the components by capturing the dependencies using ordinary depth-separable convolution at multiple scales. In addition, we propose the Low Frequency Enhancement Filter (LEF), which uses a learnable Gaussian filter to capture low-frequency semantic information while suppressing high-frequency noise to capture more feature information. During training, we only use the target detection loss, semantic segmentation loss or instance segmentation loss of the original model, and we verify the effectiveness of our method on Low Light Instance [18], Dark Face [19] and ACDC NightTime[20], and ExDark[21].

The main contributions of this work can be summarized as follows:

1.We propose a novel plug-and-play module, termed LPEnHancer, which enhances low-light features and reduces noise, thereby improving performance under low-light conditions.

2.By integrating LPEnHancer with other visual task models, such as the object detection model RtmDet, we have developed a joint LPEnHancer-RtmDet detection framework.

3.Extensive experiments across four different downstream vision tasks demonstrate that our approach achieves consistent and significant improvements over the baseline, low-light image enhancement (LLIE) methods, and task specific state-of-the-art approaches.

RELATED WORK

Object Detection

Object detection is a cornerstone of the computer vision field and has achieved remarkable success in recent years. It has made key breakthroughs in real-world applications, such as autonomous driving, and is indispensable for object detection in various systems, including autonomous driving perception and surveillance. The current object detection model architecture typically consists of three main components: the backbone network for feature extraction, the neck for feature fusion, and the head for detection. The backbone network is usually a model pre-trained on a large-scale dataset to extract key features, and it typically includes 3 to 5 stages for extracting features at different resolutions. Several CNN-based backbone networks, such as ResNet[22], ConvNeXt [23], and CSPNet [24], have been successful. Additionally, following the great success of Transformers in NLP, many scholars have applied this architecture to computer vision, leading to the proposal of various Transformer-based backbone networks, including Vision Transformer[25], Swin Transformer[26],

and MobileVIT[27]. It is a common practice to integrate feature fusion modules between the backbone and detection networks to combine feature maps of different scales, thereby improving performance [28]. Common feature fusion modules [29] include Feature Pyramid Networks (FPNs), Path Aggregation Networks (PANs), and Bidirectional Feature Pyramid Networks (BIFPNs).

Other visual tasks besides object detection.

In recent years, the field of computer vision has seen significant advancements beyond the traditional domains of face recognition and object detection. Researchers have delved into more complex tasks that require a deeper understanding of the visual content. Among these are semantic segmentation, which involves the pixel-wise classification of images into various categories, and instance segmentation, which further distinguishes individual instances of the same class. These tasks are crucial for applications such as autonomous navigation, medical imaging, and advanced driver-assistance systems (ADAS). Xue et al. [30]have made a notable contribution in this space by designing a contrast learning strategy that not only enhances visual perception but also significantly improves machine performance. Their approach has been rigorously tested and demonstrated impressive results on the Adverse Conditions Dataset for Nighttime Semantic Segmentation with Correspondence Correlation (ACDC) [20]. This dataset presents a unique challenge due to its focus on low-light conditions, which are common in nighttime scenarios and often lead to poor segmentation performance.Moreover, the Low-Light Instance Segmentation (LIS) dataset has been introduced to fill a critical gap in the field.

Low-light Enhancement

The goal of low-light enhancement tasks is to improve human visual perception by recovering image details, correcting color distortions, and providing high-quality images for advanced visual tasks such as target detection. Zhang et al. [31]proposed Kind, a method that can be trained with paired images at different light levels without the need for real ground-truth images. Guo et al. [15,16]proposed Zero-DCE and Zero-DCE++, which transform the low-light enhancement task into an image-specific curve estimation problem. Lv et al. [32] proposed a Multi-branch Low-Light Enhancement Network (MBLLEN), capable of extracting features at different levels and generating the output image through multi-branch fusion. Cui et al. [33] proposed an Illumination Adaptive Translator (IAT), which constructs an end-to-end Transformer through dynamic query learning. However, most low-light enhancement models are complex and significantly impact the real-time performance of detectors. Additionally, using these methods to preprocess entire images in a dataset often

leads to unsatisfactory results. This is because low-light enhancement methods use enhancement loss functions to optimize the network. These loss functions force the network to focus equally on all pixels, ignoring the learning of informative details required for advanced downstream vision tasks (e.g., object detection). Moreover, they can also destroy edge details of objects, leading to unsatisfactory results.

Object Detection in Low-light Condition

Object detection in dark environments poses significant challenges due to the limited number of photons, leading to the loss of edge details in objects. Moreover, images captured in low-light conditions often exhibit substantial noise, which can obscure object features. Consequently, it is difficult for detection networks to accurately infer the entire feature region of an object, thereby extracting sufficient feature information for subsequent prediction tasks. Liu et al. [34] introduced IA-YOLO, a method that enhances detection performance by adaptively processing each image. They developed a differentiable image processing (DIP) module tailored for severe weather conditions and used a compact convolutional neural network (CNN-PP) to optimize the DIP's parameters. Expanding on the work of Liu et al., Karwar et al. [35] proposed GDIP-YOLO, which incorporates a gating mechanism enabling multiple DIPs to function concurrently. Furthermore, Qin et al. [36]presented the Detection Driven Enhanced Network (DENet), designed for target detection under adverse weather conditions. Jade et al. [37]contributed PE-YOLO, leveraging the Sobel operator at various scales to refine object details and enhance model performance. Lastly, Khurram Azeem Hashmi et al.[38] designed a multiscale feature enhancement network, FeatEnHancer, which has demonstrated significant and consistent improvements across several low-light datasets.

PROPOSED APPROACH

The core idea of this thesis is to design a generalized pluggable low-light image enhancement module that adaptively enhances features under low-light image conditions to improve the performance of several downstream vision tasks such as target detection, semantic segmentation, instance segmentation, and face recognition in dark environments. The overall architecture of the LPEnHancer is presented in Figure 1. Our LPEnHancer receives a low-light image as input and then uses pyramid to decompose the image into four components with different scales, which are restored to the original resolution after feature mining and noise reduction for each component to enhance its semantic representation.

Overview

In this paper, a Laplacian image pyramid is employed to decompose the input image into four components with different resolutions, and each component is adaptively enhanced using the proposed Detail Mining Module (DMM) and Low-Frequency Enhancement Filter (LEF). Let the input low-light image be denoted as $I \in R^{640*640*3}$

A learnable Gaussian-like filtering module is adopted for downsampling to generate a set of feature representations with progressively decreasing resolutions:

$$G_1(x) = x \tag{1}$$

$$G_i(x) = LEFfilter(G_{i-1}(x)), \quad i \in \{2,3,4\} \tag{2}$$

where LGFilter denotes a learnable Gaussian-like filtering module with stride, which preserves the number of channels while reducing the spatial resolution of feature maps by half. Specifically, the Gaussian kernel weights are directly learned rather than parameterized by a predefined Gaussian function (e.g., sigma). To maintain Gaussian-like properties, the weights are constrained via softmax normalization to ensure non-negativity and unit-sum. To ensure stable low-pass characteristics, the convolution kernel weights are normalized using a softmax operation after each gradient update during training, enforcing non-negativity and unit-sum constraints.

As the spatial resolution decreases progressively (e.g., from $640 * 640$ to $320 * 320$), the downsampling process inevitably discards high-frequency information, making it inherently non-invertible.

To alleviate this issue, we adopt a Laplacian pyramid decomposition to explicitly preserve high-frequency details during the multi-scale representation process. Each level of the Laplacian pyramid is defined as:

$$L_i(x) = G_i(x) - up(G_{i+1}(x)), i \in \{1,2,3\} \tag{3}$$

$$L_4(x) = G_4(x) \tag{4}$$

where up(·) denotes bilinear upsampling, G_i represents the i -th level of the Gaussian pyramid, and L_i denotes the corresponding Laplacian component. This decomposition separates the image into band-pass

residuals and a low-frequency base component, yielding four components { L_1 , L_2 , L_3 , L_4 } with gradually reduced resolutions.

It is worth noting that the classical Laplacian pyramid is theoretically invertible under the assumption of fixed filters. However, in the proposed framework, the Gaussian kernels are learnable and dynamically updated during training, and thus strict invertibility is not guaranteed. Instead, the proposed model achieves approximate reconstruction by leveraging the preserved residual components together with the learning capacity of the network.

We observe that different levels of the pyramid capture complementary information, ranging from fine-grained details to coarse semantic structures. Therefore, the Detail Mining Module (DMM) is applied to enhance structural details, while the Low-Frequency Enhancement Filter (LEF) is used to suppress noise and strengthen low-frequency semantic representations. These two processes are executed in parallel to improve computational efficiency and increase the frames per second (FPS).

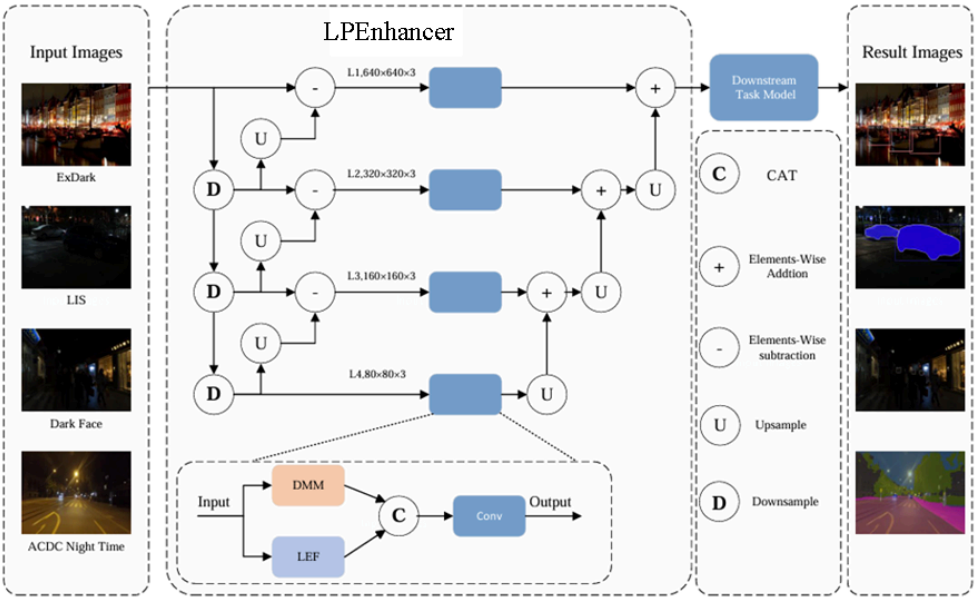

Figure 1. Network architecture of the proposed LPEnHancer

Low-Frequency Enhancement Filter (LEF) and Gaussian Filter Downsampling

Typically, low-light images suffer from insufficient illumination and fluctuations in the number of photons reaching the sensor, resulting in a significant amount of noise, which is predominantly present in the high-frequency components. High-frequency noise can introduce high-frequency interference into the feature

maps of convolutional neural networks (CNNs), thereby reducing the model's ability to extract low-frequency semantic features. To achieve robustness against image noise, the features extracted by the network should be clean and consistent in their response to scene content. To this end, we can incorporate low-pass filters such as Gaussian, mean, and Wiener filters into the network to suppress noise. For instance, methods like MSR[39] and MSRCR[40] have successfully employed Gaussian filtering to mitigate high-frequency noise and decompose images into illumination and reflectance maps. Although these low-pass filters can effectively suppress high-frequency noise without significantly increasing computational load, thereby aiding downstream tasks in achieving better low-light detection/segmentation results, this approach may not be optimal. In each scale component extracted from the Laplacian pyramid, the low-pass component retains the majority of the semantic information present in the image, while the high-frequency components contain texture detail information, which is crucial for subsequent downstream tasks. To more effectively extract low-frequency semantic information and texture detail information during the image reconstruction process and suppress high-frequency noise components, we propose a Low-Frequency Enhancement Filter (LEF). The detailed structure of the LEF is illustrated in Figure 2.

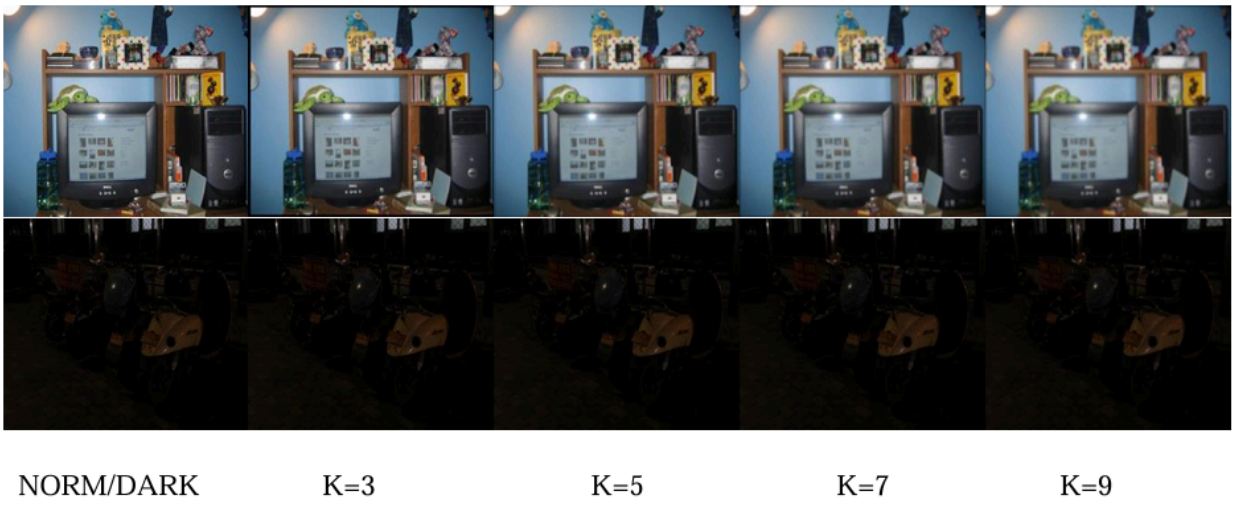

Figure 2. Comparison of the Effects of Different Convolutional Kernel Sizes.

For Gaussian convolution, the size of the convolution kernel dictates the extent of noise reduction capability. As shown in Figure 2, for both low-light and normally illuminated images, the degree of blurring in the image processed by Gaussian convolution increases with the enlargement of the convolution kernel, which also implies that the high frequency components of the image are more severely suppressed. While the application

of these simple low-pass filters is beneficial, they may not fully address the complexity of noise under low-light conditions. Therefore, in the design of the Low-Frequency Enhancement Filter (LEF), to prevent excessive image smoothing and enhance the capability of information extraction, we have decided to employ the aforementioned learnable Gaussian filtering and integrate filters with different kernel sizes. As depicted in Figure 3, inspired by SPP[41] and ASPP[13], our LEF module comprises three branches with convolution kernels of varying sizes and a residual branch. For an input feature, we first perform a standard convolution, followed by learnable Gaussian convolution filtering with kernels of sizes 3×3, 5×5, and 7×7, which differ in the degree of blurring to suppress noise and extract important detail information necessary for downstream tasks. The three outputs are then concatenated with the residual along the channel dimension, and finally, a standard convolution operation is applied to integrate the channel information and adjust the number of channels. In summary, we have introduced the concept of Inception[42], which employs learnable Gaussian convolutions with multiple kernel sizes for the input features. Firstly, different convolutional sizes provide varying receptive fields, enabling feature extraction at different levels. Secondly, the varying sizes of the convolutional kernels imply different capabilities for noise reduction and blurring, allowing the model to aggregate features with different degrees of blur, thereby adaptively selecting the level of noise reduction.

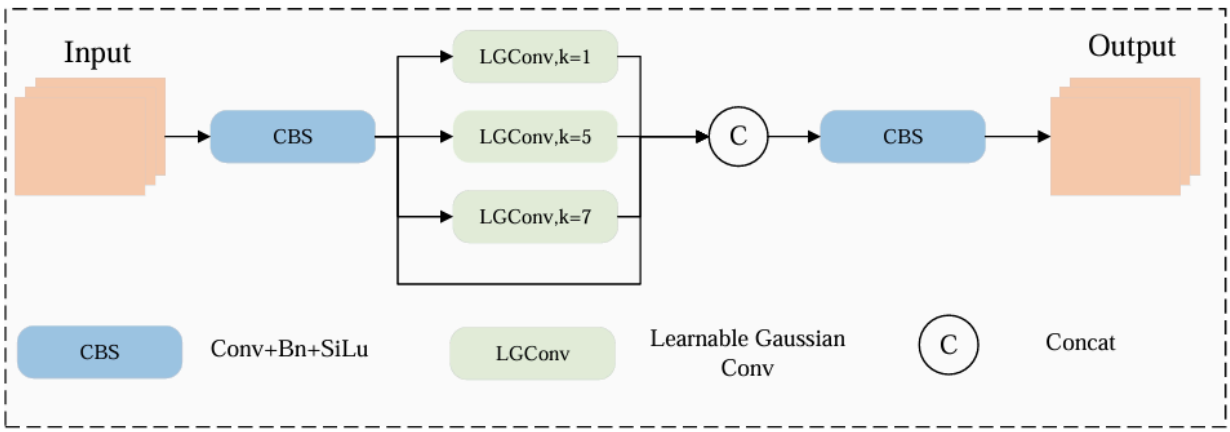

Figure 3. Detailed structure of our proposed Low-Frequency Enhancement Filter (LEF).

Detail Mining Module

We propose a Detail Mining Module (DMM) designed to enhance the high-frequency components within a Laplacian pyramid, focusing on the recovery of fine-grained structural details and texture information. The

module is composed of two parallel branches: a Local Feature Enhancement branch (LFE) and a Texture Information Mining branch (TIM). The overall architecture is illustrated in Figure 4.

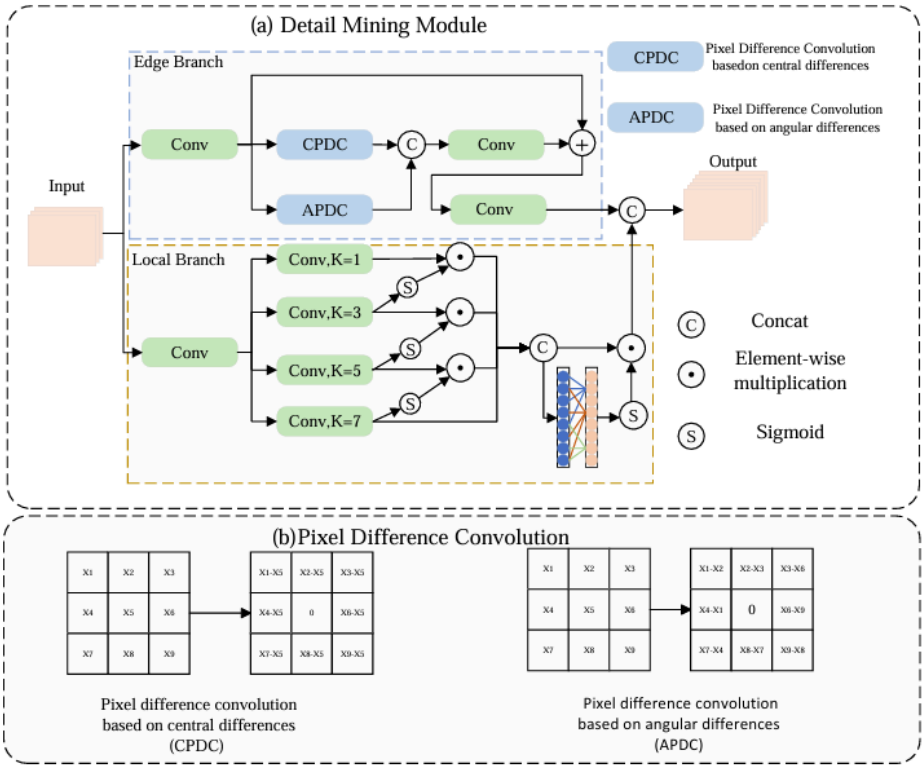

Figure 4: Detailed structure of our proposed Detail Mining Module (DMM).

The LFE branch employs standard convolutions to capture local structural features, while the TIM branch leverages edge-aware operations based on Pixel Difference Convolution (PDC)[17] to further enhance edge and texture characteristics. In contrast, the Low-Frequency Enhancement Filter (LEF) focuses on enhancing low-frequency components for noise suppression and global semantic consistency.

Local Feature Enhancement Branch: To effectively capture local structural information at multiple scales, we design the LFE branch to extract and fuse multi-scale features. Deep neural networks have demonstrated strong representation capability, but increasing depth often leads to higher computational cost and training difficulty. Residual connections (ResNet) alleviate gradient degradation, while DenseNet[43] promotes feature reuse through dense connections.Inspired by these works, our LFE branch aims to extract multi-scale local structural features while modeling interdependencies across different feature channels. Specifically, we first apply a standard convolution with batch normalization to the input features. Then, four convolutional layers with different kernel sizes are used to extract multi-scale features, denoted as $\{x_1, x_3, x_5, x_7\}$.

To facilitate cross-scale interaction, we introduce an attention-guided mechanism where features with larger receptive fields guide those with smaller ones. Concretely, element-wise multiplication is performed between x_1 and Sigmoid(x_3), and similar operations are applied across scales. Finally, the outputs are concatenated along the channel dimension, followed by Efficient Channel Attention (ECA)[44] to emphasize informative channels and suppress less relevant ones.

Texture Information Mining Branch: Edge and texture information are critical visual cues that correspond to high-frequency components in images. In low-light conditions, such information is often severely degraded due to noise and low signal-to-noise ratio, which negatively affects downstream tasks.

To address this issue, we design the TIM branch to explicitly mine and enhance edge and texture features. This branch utilizes Pixel Difference Convolution (PDC), which is particularly effective for capturing local intensity variations.

Specifically, the input feature f_{in} from the Laplacian pyramid is first processed by a standard convolution. Then, two variants of PDC—central difference convolution (CPDC) and angular difference convolution (APDC)—are applied to extract complementary edge-aware features, resulting in intermediate representations f_1 and f_2 . These features are concatenated along the channel dimension and further fused through a convolutional layer. Finally, a residual connection is added to obtain the texture-enhanced output f_{out} .

EXPERIMENTS

Datasets and evaluation metrics

We conducted extensive experiments on the proposed LPEnhancer module to evaluate its performance on several downstream tasks in low-light vision, including generic target detection[21], face detection[19], semantic segmentation [41], and dark semantic segmentation [5].Table 1 summarizes the key statistics of the used datasets. This section first compares the proposed method with a robust baseline, existing low-light image enhancement (LLIE) methods, and state-of-the-art techniques for specific tasks. We then perform an ablation study of important design choices for our LPEnhancer.Below are the details of the dataset and the evaluation metrics.

Exdark. Exclusively dark image dataset is a dedicated dataset for low-light target detection, with a total of 10 lighting environments ranging from very low light to low-light, and a total of 12 object categories. There are 5889 images in the training set, 736 images in the validation set, and 737 images in the test set. We used the

training set for training and evaluated on the test set. In this task, we use mAP and mAP50 to evaluate the accuracy of the model.

LIS. The LIS dataset is a dataset dedicated to low-light instance segmentation, comprising 8 classes and a total of 2230 low-light images, all of which contain pixel-level annotations. Of these images, 1561 are used for training and 669 for evaluation. For the low-light instance segmentation task, we utilize bbox_mAP, bbox_mAP50, segm_mAP, and segm_mAP50 as metrics to evaluate the model's accuracy.

Dark Face. The Dark Face dataset provides 6000 real-world, low-light images of faces captured at night, in school buildings, streets, bridges, overpasses, parks, etc., all labeled with bounding boxes. There is only one category of faces, of which 4000 are used for training and 2000 for evaluating the performance. Since the Dark Face dataset has only one category, in this task we use AP and AP50 to evaluate the accuracy of the model.

ACDC Night Time. The Adverse Conditions Dataset with Correspondences is a semantic segmentation dataset that provides pixel-level annotations for harsh environments, featuring 19 categories, similar to Cityscapes. The dataset categorizes images into four harsh environmental conditions, such as foggy days and nighttime. In this study, we focus exclusively on the nighttime segment, known as ACDC Night Time, which includes 400 low-light images for training and 106 for testing. It should be noted that the dataset is highly imbalanced due to the selection of only a part of it, with some categories having an Intersection over Union (IOU) of 0; however, these categories are still included in the mean Intersection over Union (mIoU) calculation. We utilize mIoU as the metric to assess the model's performance.

Table 1. Dataset Information

Dataset	Task	Class	Train	Val	Test
ExDark	Dark Object Detection	12	5889	736	737
LIS	Instance Segmentation	8	1561	669	#
Dark Face	Face Detection	1	4000	2000	#
ACDC Night Time	Semantic Segmentation	19	400	106	#

Dark Object Detection

SETTING: To perform dark object detection experiments on real-world data, we consider the Exclusively Dark (ExDark) dataset (see Table 1). We report the results using RtmDet[45] as a typical detector. For the experiments, the images were resized to 640 × 640, and we trained using mmdetection[46] with the learning rate set to 0.002, weight decay to 0.0001, batch size of 16, none of the pre-trained weights were used, and 300

epochs were trained using the cosine annealing scheduler. Note that for each target detection framework, the training of our model, the baseline, the LLIE method, and the task-specific state-of-the-art method when we used the same settings. In addition to the high memory requirements of FeatEnhancer and PE-NET which prevent the batch size from being set to 16, we employ the linear scaling rule for the learning rate to adjust the batch size.

We compare our proposed LPEnhancer with several state-of-the-art low-light image enhancement (LLIE) methods, including KIND[31], RAUS [47], EnGAN [48], MBLLEN [32], Zero-DCE[16], and Zero-DCE++[15], as well as SCI[49]. Additionally, we compare our method with state-of-the-art dark object detection methods: FeatEnHancer[38], PENet[37], and DENet[36]. For the LLIE methods, all images are enhanced using their published checkpoints before being passed to the detector. In contrast, for the dark object detection methods, we integrate these models with the backbone network of RtmDet, creating a combined dark object detection method-RtmDet model for training to facilitate direct comparisons.

Table 2. Quantitative Comparison on ExDark. The results obtained on commonly used evaluation metrics are highlighted. Our LPEnhancer brings consistent improvements and achieves new state-of-the-art results with RtmDet.

Methods	mAP	mAP50	Flops/G	Params/M
RtmDet	44.0	70.3	\	\
RAUS	41.7	67.5	\	\
KIND	42.8	67.5	\	\
Zero-DCE	42.9	69.0	\	\
Mbllen	42.5	68.4	\	\
Zero-DCE++	42.7	68.7	\	\
EnGan	43.7	69.6	\	\
SCI	43.1	69.3	\	\
DeNet	46.5	73.3	2.209	0.045
FeatEnhancer	46.6	73.6	44.339	0.14
PeNet	45.7	72.7	11.427	0.092
LPEnhancer(Ours)	47.0	74.0	2.823	0.022

Results: Table 2 presents the results of the LLIE method, dark object detection, and our proposed approach on the EXdark dataset. Notably, our LPEnhancer method consistently and significantly outperforms previous methods, achieving new state-of-the-art mAP50 and mAP scores of 74.0% and 47.0%, respectively. Specifically, compared with the baseline model, our method improves the mAP and mAP50 by 3.0% and 3.7%, respectively. In addition, compared with representative low-light image enhancement (LLIE) based methods, such as Zero-

DCE, our approach achieves further performance gains, with an improvement of approximately 3.1% in mAP. Moreover, when compared with dedicated dark object detection methods, such as FeatEnhancer and DeNet, our method still demonstrates consistent improvements, achieving gains of approximately 0.4%–0.5% in mAP.

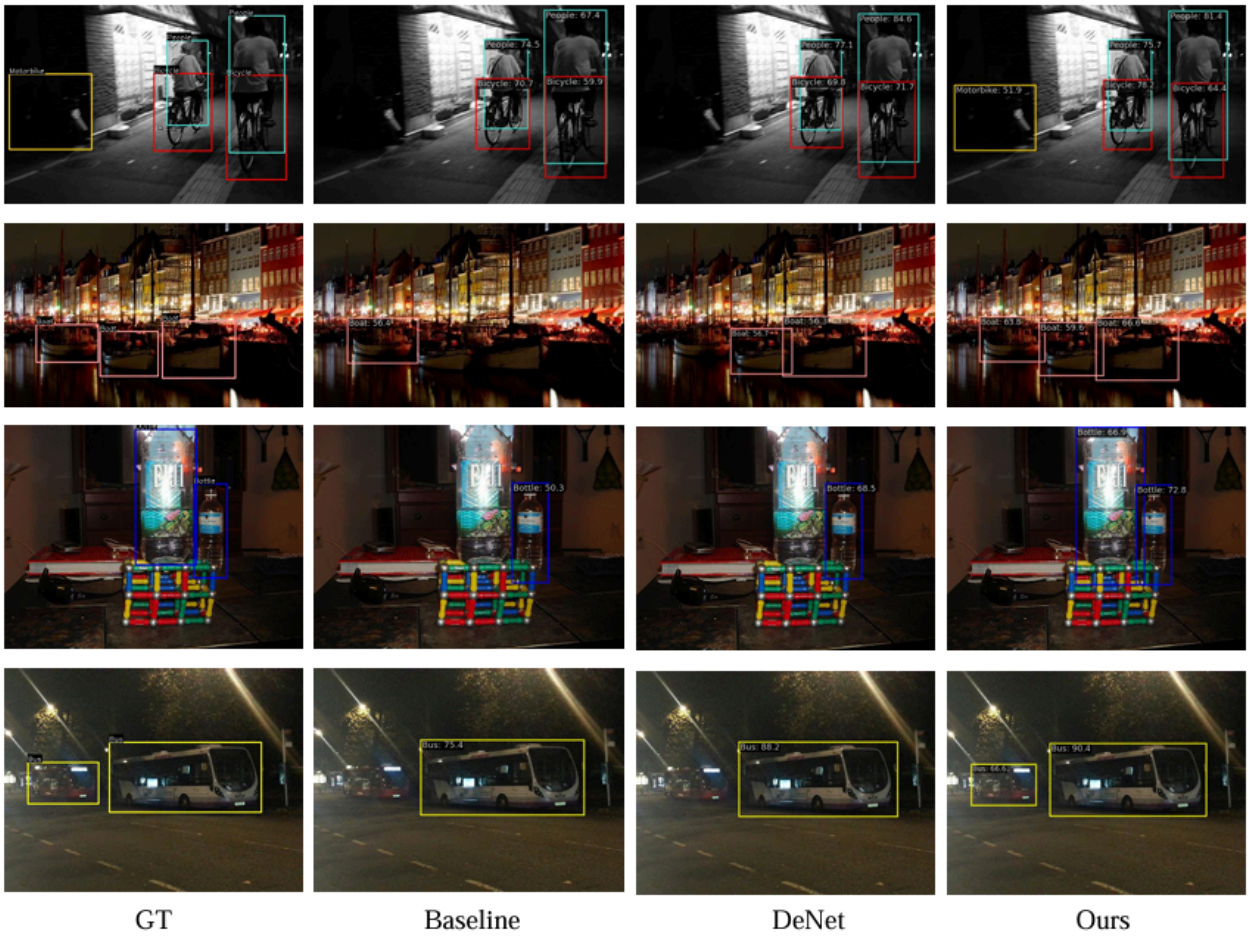

Figure 5: Visual comparison of LPEnhancer with the baseline method and the best competitors on the ExDark dataset. Zoom in for the best view

Furthermore, Figure 5 illustrates four detection examples from our method, the Baseline, and the best-performing dark object detection models. From these examples, it is evident that our method is more accurate than other advanced object-detection models. The enhanced detection capabilities of our LPEnhancer for objects such as motorcycles, buses, and bottles in low-light conditions are apparent, with fewer instances of missed detections. Additionally, for easily detectable targets, our method yields higher confidence scores. These results indicate that our LPEnhancer enhances hierarchical features beneficial for dark object detection, even under poor visual quality conditions, by mitigating the negative impact of noise on detection models

and producing state-of-the-art results. Note that, in terms of parameter count, the LLIE method, which pre-processes the dataset, has a computational and parameter count of zero, while our proposed method and dark object detection account for additional computational and parameter overhead. The inherent parameters of the Baseline are not included in this count. Moreover, we will not list computational and parameter overhead in future experimental tables.

Low light instance segmentation

SETTING: Low-light Instance Segmentation is a dark instance segmentation dataset with pixel-level annotations, and we report results using the instance segmentation version of RtmDet as the segmenter. For the experiments, the images were resized to 640 × 640, and we used mmdetection for training using the ADAW optimizer with the learning rate set to 0.002, weight decay to 0.005, batch size of 16, none of the pre-training weights were used, and 300 epochs were trained using the cosine annealing scheduler. We will report bbox_map, bbox_map50, segm_map, and segm_map50 data from four experiments to measure the performance of each method.

Table 3. Quantitative Comparison on LIS. The results obtained on commonly used evaluation metrics are highlighted. Our LPEnhancer brings consistent improvements and achieves new state-of-the-art results with RTMDet.

Methods	Bbox_mAP/%	Bbox_mAP50/%	Segm_mAP/%	Segm_mAP50/%
RtmDet	45.2	67.7	39.4	62.0
RAUS	44.8	65.9	38.2	60.8
KIND	44.9	66.8	38.5	61.1
ZERO-DCE	45.3	67.2	39.5	61.5
MBLLEN	45.1	67.5	39.2	61.7
ZERO-DCE++	45.2	67.8	39.6	62.3
ENGAN	45.0	67.1	39.3	61.5
SCI	46.1	68.4	40.0	62.6
DENET	46.3	68.5	40.2	63.1
FeatEnhance	46.6	68.7	40.1	63.0
PENET	46.0	67.9	39.6	63.1
MFFormer	47.8	69.1	41.2	64.3
GH-RtmDet	46.9	69.5	39.9	63.3
LPEnHancer(ours)	46.9	69.0	40.5	63.3

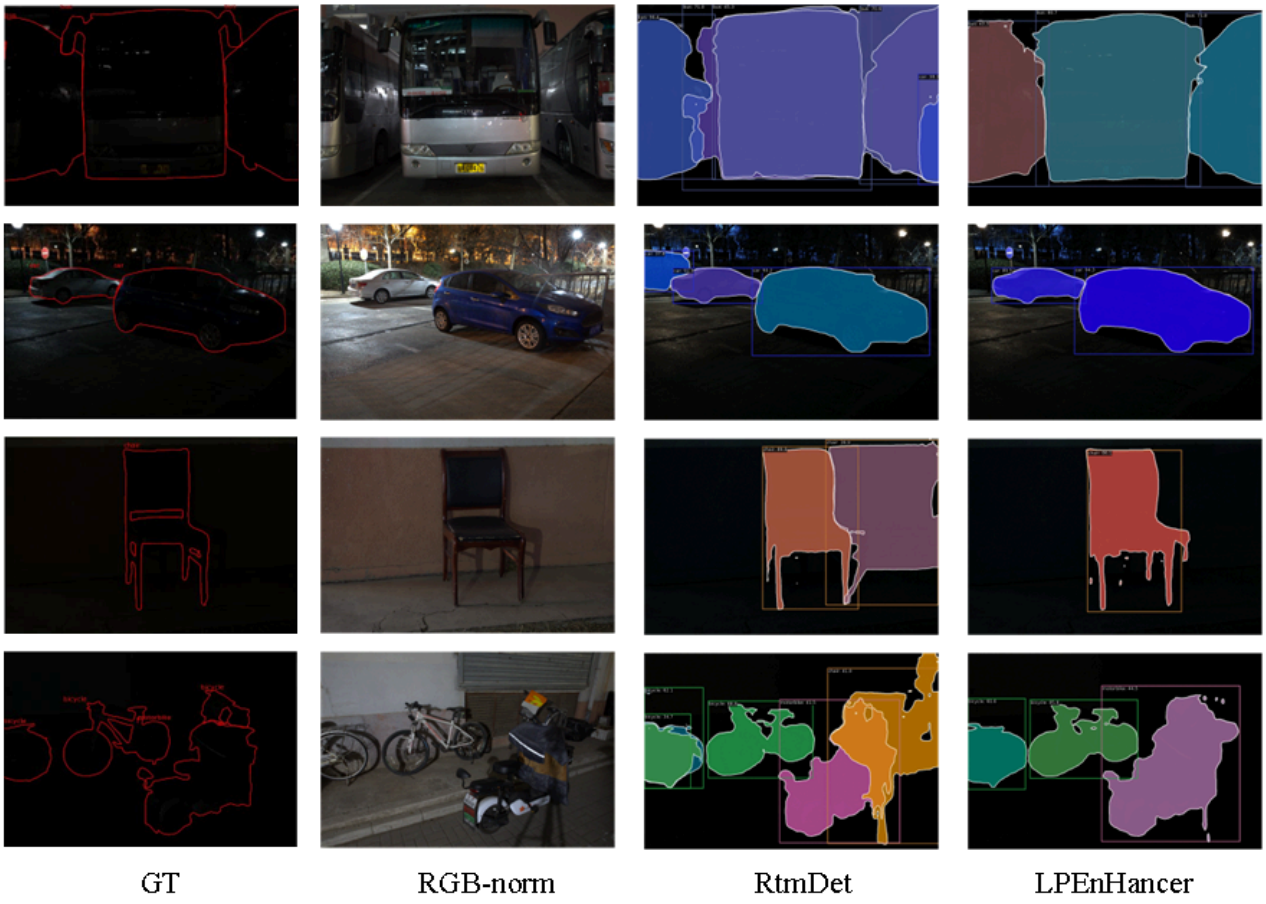

Figure 6: Visual comparison of LPEnhancer with the baseline method ,GT and ours method on the LIS dataset.

Results: Table 3 presents the results of the LLIE method, dark object detection, and our proposed approach on the LIS dataset. In terms of the four key evaluation metrics for instance segmentation, our method achieved rates of 46.9%, 69.0%, 40.5%, and 63.3%, respectively, representing a significant improvement over the baseline. Compared to low-light enhancement methods, DENet, PENet, and FeatEnHancer, our method demonstrated the best competitiveness on the LIS dataset, achieving state-of-the-art results.

Furthermore, Figure 6 presents several visual result examples. Observations reveal that the baseline often mislabels a single object as multiple distinct objects, likely due to the extremely low-light environment where the images were captured. In contrast, our method effectively suppresses noise caused by low illumination and enhances the extraction of edge and texture information, leading to more refined segmentation masks. These results demonstrate that our approach can significantly enhance segmentation quality, even in the face of poor visual quality. The visual examples in Figure 6 illustrate the effectiveness of our method in addressing the challenges of low-light scenarios, where traditional methods frequently struggle to achieve

accurate segmentation. The refinement in mask quality is a direct consequence of the method's enhanced feature extraction capabilities, which are crucial for distinguishing between objects and their backgrounds in conditions of limited visibility.

Face Detection on Dark Face

SETTING: DARK FACE is a challenging face detection dataset released for the UG2 competition. We report results using RtmDet [35] as a typical detector. For the experiments, the images were resized to 640 × 640, and we trained using mmdetection [4] with the learning rate set to 0.002, weight decay to 0.0001, batch size of 16, none of the pre-training weights were used, and 200 epochs were trained using a linear scheduler and a cosine annealing scheduler.

Results: Table 4 delineates the performance of LLIE methods, dark object detection, and our proposed approach on the DarkFace dataset. Notably, certain low-light enhancement methods have achieved minor improvements on the DarkFace dataset. We surmise that these enhancements, by improving the illumination of the minuscule faces within the dataset, have conferred some benefits (for instance, Zero-DCE has contributed a 0.4 AP enhancement). Our method, however, has garnered the most competitive outcomes on the DarkFace dataset, with AP and AP50 attaining 19.8 and 45.8, respectively, and realizing a performance improvement of 2.0 AP and 3.7 mAP over the baseline. We attribute this remarkable enhancement in performance to the fact that the faces within the DarkFace dataset are exceedingly diminutive, and the features extracted by the backbone network are prone to being subsumed by noise. Our method alleviates this issue by effectively suppressing noise.

Table 4. Quantitative Comparison on DarkFace.

Methods	AP/%	AP50/%
RtmDet	17.8	42.1
RAUS	16.5	41.3
KIND	17.6	41.9
ZERO-DCE	18.2	42.4
MBLLEN	17.9	42.2
ZERO-DCE++	18.1	42.1
ENGAN	17.3	41.8
SCI	18.1	42.5
DENET	18.9	44.4
FeatEnhance	19.3	45.2

Methods	AP/%	AP50/%
PENET	19.2	44.9
LPEnHancer(Ours)	19.8	45.8

Night time Semantic Segmentation on ACDC

Settings: We report semantic segmentation results in low-light conditions using nighttime images from the ACDC dataset [41] (see Table 1). DeepLab v2 [6] serves as the baseline for semantic segmentation, trained using mmseg [8] to facilitate direct comparison with contemporary work. We follow the experimental setup outlined in mmseg. The image size is cropped to 512 × 1024 pixels, with a batch size (bs) of 2 and an 80k scheduler. It is noteworthy that although the ACDC dataset is a Cityscapes-formatted dataset with 19 classes, we only utilized nighttime images here, leading to a scarcity of some classes that could not be accurately segmented on the validation set.

Results: We compared our method with several state-of-the-art low-light image enhancement (LLIE) methods. As shown in Table 5, our LPEnhancer achieved an mIoU improvement of 0.78 over the baseline, with a mean Intersection over Union (mIoU) reaching 44.34. Although our method (LPEnhancer) achieved an improvement of 0.78 in mIoU, outperforming low-light enhancement methods and dark object detectors such as DENet, the performance gain brought by LPEnhancer is not significant compared to the previous three experiments. We believe that the possible reason is the pyramid structure of LPEnhancer, which disrupts the inter-pixel connections, leading to difficulty in improving the segmentation accuracy of downstream DeepLab v2.

Table 5. Quantitative Comparison on ACDC Night Time

Methods	mIoU/%
DeepLabv2	43.56
RAUS	22.18
KIND	35.1
ZERO-DCE	27.6
MBLLEN	37.88
ZERO-DCE++	27.35
ENGAN	35.5
SCI	34.5
DENET	42.31
FeatEnhance	43.5
PENET	43.0
LPEnHancer(Ours)	44.44

Ablation Studies

Design of Key Modules: This section presents an ablation study examining the critical design choices incorporated into the Laplacian pyramid enhance networks (LPEnhancer) when integrated into the Real-Time Multi-Task Detection (RtmDet) framework. The individual contributions of the two branches within the Low-Frequency Enhancement Filter (LEF) and the Detail Mining Module (DMM) are delineated in Table 6, with performance metrics reported on the Dark Face dataset. Due to the impact of module integration on the feature map channel dimensions, adjustments were made to the convolutional channel dimensions of the modeled feature locations during the ablation experiments.

The results indicate that the implementation of the Laplacian pyramid and the Low-Frequency Enhancement Filter (LEF) led to improvements of 19.3% and 44.3% in mAP and mAP50, respectively, equating to performance enhancements of 1.5% and 2.1%. Subsequently, the sequential addition of the two branches of the Detail Mining Module (DMM) further increased mAP by 0.2% and 0.3%, with corresponding mAP50 improvements of 0.6% and 1.0%. These results of the ablation study indicate that optimal performance is achieved when all three modules are utilized, suggesting synergy between the modules.

This comprehensive analysis validates the effectiveness of the recommended design improvements and emphasizes the importance of each component in the LPEnhancer to enhance the detection capabilities of the RtmDet in low-light conditions. In addition, the complementary nature of these modules emphasizes the need for a holistic approach to the challenges posed by low-light environments in real-time detection tasks.

Table 6 Ablation analysis on different modules of our method.

LPEnHancer	LEF	Edge	Local	AP	AP50
BaseLine	-	-	-	17.8	42.1
LPEnhancer	√	-	-	19.3	44.3
	√	√	-	19.5	44.9
	√	√	√	19.8	45.8

The Impact of the Number of Levels in the Laplacian Pyramid: To evaluate the impact of the number of levels, N, in the Laplacian pyramid on the final detection/segmentation outcomes, we conducted an ablation study and reported the performance on the Dark Face dataset. Note also that an N value of 1 indicates that the pyramid structure is not utilized.

The results, as shown in Table 7, clearly indicate that an increase in the number of decomposition levels leads to an increase in the computational parameter count. As N increases from 1 to 4, both mean average precision mAP and mAP50 show a monotonic increase with N; however, when N=5, the final outcome shows negligible improvement compared to the result with N=4. Based on the results of this ablation study, and considering both performance and cost, we chose N=4 for the final structure of the LPEnHancer.

Table 7. The effects of the number of levels of Laplacian pyramid.

N	Params/M	Flops/G	AP	AP50
1	0.006	2.089	19.0	44.5
2	0.011	2.647	19.5	44.5
3	0.017	2.843	19.7	45.6
4	0.023	2.917	19.8	45.8
5	0.029	2.980	19.8	45.9

CONCLUSION

This paper proposes LPEnHancer, a novel general-purpose feature enhancement module designed to enrich hierarchical features under low-light conditions and reduce noise, thereby benefiting downstream tasks. We employ the structure of the Laplacian pyramid to decompose low-light images into four components with varying resolutions. At each scale, noise is reduced, and local feature information, as well as texture and edge information, is enhanced to improve performance in downstream tasks. Moreover, our LPEnHancer does not necessitate pre-training on synthetic datasets nor does it rely on loss functions for enhancement. These architectural innovations make LPEnHancer a versatile plug-and-play module. Extensive experiments across four different downstream vision tasks demonstrate that our approach consistently and significantly outperforms baselines, LLIE methods, and task-specific state-of-the-art approaches in terms of both consistency and significance.

Author Contributions

Zheng Fang conceived the study, performed the analysis, and wrote the manuscript.

Conflicts of Interest

The author declares no conflict of interest.

Funding

This research received no external funding.

REFERENCES

[1] Tian Y, Ye Q, Doermann D. Yolov12: Attention-centric real-time object detectors. arXiv preprint arXiv:2502.12524. 2025.

[2] Wang A, Chen H, Liu L, Chen K, Lin Z, Han J, et al. Yolov10: Real-time end-to-end object detection. Advances in neural information processing systems. 2024; 37:107984-108011.

[3] Wang C, He W, Nie Y, Guo J, Liu C, Wang Y, Han K. Gold-yolo: Efficient object detector via gather-and-distribute mechanism. Advances in neural information processing systems. 2023; 36:51094-51112.

[4] Jocher G, Qiu J. Ultralytics yolo11. URL: https://github.com/ultralytics/ultralytics. 2024.

[5] Lei M, Li S, Wu Y, Hu H, Zhou Y, Zheng X, Ding G, Du S, Wu Z, Gao Y. Yolov13: Real-time object detection with hypergraph-enhanced adaptive visual perception. arXiv preprint arXiv:2506.17733. 2025.

[6] Ren T, Liu S, Zeng A, Lin J, Li K, Cao H, Chen J, Huang X, Chen Y, Yan F, et al. Grounded sam: Assembling open-world models for diverse visual tasks. arXiv preprint arXiv:2401.14159. 2024.

[7] Cheng B, Misra I, Schwing A G, Kirillov A, Girdhar R. Masked-attention mask transformer for universal image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022. p. 1290-1299.

[8] Jain J, Li J, Chiu M T, Hassani A, Orlov N, Shi H. Oneformer: One transformer to rule universal image segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023. p. 2989-2998.

[9] Li Z, Yang B, Liu Q, Zhang S, Ma Z, Yin L, Deng L, Sun Y, Liu Y, Bai X. Lira: Inferring segmentation in large multi modal models with local interleaved region assistance. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2025. p. 24056-24067.

[10] Bolya D, Zhou C, Xiao F, Lee Y J. Yolact: Real-time instance segmentation. In: Proceedings of the IEEE/CVF international conference on computer vision. 2019. p. 9157-9166.

[11] Chen H, Sun K, Tian Z, Shen C, Huang Y, Yan Y. Blendmask: Top-down meets bottom-up for instance segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 8573-8581.

[12] Lee Y, Park J. Centermask: Real-time anchor-free instance segmentation. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 13906-13915.

[13] Chen L C, Papandreou G, Kokkinos I, Murphy K, Yuille A L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. IEEE transactions on pattern analysis and machine intelligence. 2017; 40:834-848.

[14] Li Z, Wang Y, Zhang J. Low-light image enhancement with knowledge distillation. Neurocomputing. 2023; 518:332-343.

[15] Li C, Guo C, Loy C C. Learning to enhance low-light image via zero-reference deep curve estimation. IEEE transactions on pattern analysis and machine intelligence. 2021; 44:4225-4238.

[16] Guo C, Li C, Guo J, Loy C C, Hou J, Kwong S, Cong R. Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 1780-1789.

[17] Su Z, Liu W, Yu Z, Hu D, Liao Q, Tian Q, Pietikäinen M, Liu L. Pixel difference networks for efficient edge detection. In: Proceedings of the IEEE/CVF international conference on computer vision. 2021. p. 5117-5127.

[18] Chen L, Fu Y, Wei K, Zheng D, Heide F. Instance segmentation in the dark. International Journal of Computer Vision. 2023; 131:2198-2218.

[19] Yang W, Yuan Y, Ren W, Liu J, Scheirer W J, Wang Z, Zhang T, Zhong Q, Xie D, Pu S, et al. Advancing image understanding in poor visibility environments: A collective benchmark study. IEEE Transactions on Image Processing. 2020; 29:5737-5752.

[20] Sakaridis C, Dai D, Van Gool L. Acdc: The adverse conditions dataset with correspondences for semantic driving scene understanding. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). 2021. p. 10745-10755. doi:10.1109/ICCV48922.2021.01059.

[21] Loh Y P, Chan C S. Getting to know low-light images with the exclusively dark dataset. Computer Vision and Image Understanding. 2019; 178:30-42.

[22] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 770-778.

[23] Liu Z, Mao H, Wu C Y, Feichtenhofer C, Darrell T, Xie S. A convnet for the 2020s. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022. p. 11976-11986.

[24] Wang C Y, Liao H Y M, Wu Y H, Chen P Y, Hsieh J W, Yeh I H. Cspnet: A new backbone that can enhance learning capability of cnn. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops. 2020. p. 390-391.

[25] Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, Dehghani M, Minderer M, Heigold G, Gelly S, Uszkoreit J, Ho J. An image is worth 16x16 words: Transformers for image recognition at scale. In: International Conference on Learning Representations (ICLR). 2021.

[26] Liu Z, Lin Y, Cao Y, Hu H, Wei Y, Zhang Z, Lin S, Guo B. Swin transformer: Hierarchical vision transformer using shifted windows. In: Proceedings of the IEEE/CVF international conference on computer vision. 2021. p. 10012-10022.

[27] Mehta S, Rastegari M. Mobilevit: Light-weight, general-purpose, and mobile-friendly vision transformer. In: International Conference on Learning Representations. 2022.

[28] Kaur J, Singh W. Tools, techniques, datasets and application areas for object detection in an image: a review. Multimedia Tools and Applications. 2022; 81:38297-38351.

[29] Liang W, Xu P, Guo L, Bai H, Zhou Y, Chen F. A survey of 3d object detection. Multimedia Tools and Applications. 2021; 80:29617-29641.

[30] Xue X, He J, Ma L, Wang Y, Fan X, Liu R. Best of both worlds: See and understand clearly in the dark. In: Proceedings of the 30th ACM International Conference on Multimedia. 2022. p. 2154-2162.

[31] Zhang Y, Zhang J, Guo X. Kindling the darkness: A practical low-light image enhancer. In: Proceedings of the 27th ACM international conference on multimedia. 2019. p. 1632-1640.

[32] Lv F, Lu F, Wu J, Lim C. Mbllen: Low-light image/video enhancement using cnns. In: BMVC, Northumbria University. 2018. p. 4.

[33] Cui Z, Li K, Gu L, Su S, Gao P, Jiang Z, Qiao Y, Harada T. You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. In: 33rd British Machine Vision Conference 2022, BMVC 2022, London, UK, November 21-24, 2022, BMVA Press. URL: https://bmvc2022.mpi-inf.mpg.de/0238.pdf. 2022.

[34] Liu W, Ren G, Yu R, Guo S, Zhu J, Zhang L. Image-adaptive yolo for object detection in adverse weather conditions. In: Proceedings of the AAAI conference on artificial intelligence. 2022. p. 1792-1800.

[35] Kalwar S, Patel D, Aanegola A, Konda K R, Garg S, Krishna K M. Gdip: Gated differentiable image processing for object detection in adverse conditions. In: 2023 IEEE International Conference on Robotics and Automation (ICRA), IEEE. 2023. p. 7083-7089.

[36] Qin Q, Chang K, Huang M, Li G. Denet: detection-driven enhancement network for object detection under adverse weather conditions. In: Proceedings of the Asian Conference on Computer Vision. 2022. p. 2813-2829.

[37] Yin X, Yu Z, Fei Z, Lv W, Gao X. Pe-yolo: Pyramid enhancement network for dark object detection. In: International Conference on Artificial Neural Networks, Springer. 2023. p. 163-174.

[38] Hashmi K A, Kallempudi G, Stricker D, Afzal M Z. Featenhancer: Enhancing hierarchical features for object detection and beyond under low-light vision. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. 2023. p. 6725-6735.

[39] Ghosh C, Roy S, Cavalcanti D. Coexistence challenges for heterogeneous cognitive wireless networks in tv white spaces. IEEE Wireless Communications. 2011; 18:22-31.

[40] 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 Transactions on Image processing. 1997; 6:965-976.

[41] He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE transactions on pattern analysis and machine intelligence. 2015; 37:1904-1916.

[42] Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Erhan D, Vanhoucke V, Rabinovich A. Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2015. p. 1-9.

[43] Huang G, Liu Z, Van Der Maaten L, Weinberger K Q. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 4700-4708.

[44] Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. Eca-net: Efficient channel attention for deep convolutional neural networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020. p. 11534-11542.

[45] Lyu C, Zhang W, Huang H, Zhou Y, Wang Y, Liu Y, Zhang S, Chen K. Rtmdet: An empirical study of designing real-time object detectors. arXiv:2212.07784. 2022.

[46] Chen K, Wang J, Pang J, Cao Y, Xiong Y, Li X, Sun S, Feng W, Liu Z, Xu J, et al. Mmdetection: Openmmlab detection toolbox and benchmark. arXiv preprint arXiv:1906.07155. 2019.

[47] Liu R, Ma L, Zhang J, Fan X, Luo Z. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2021. p. 10561-10570.

[48] Jiang Y, Gong X, Liu D, Cheng Y, Fang C, Shen X, Yang J, Zhou P, Wang Z. Enlightengan: Deep light enhancement without paired supervision. IEEE transactions on image processing. 2021; 30:2340-2349.

[49] Ma L, Ma T, Liu R, Fan X, Luo Z. Toward fast, flexible, and robust low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2022. p. 5637-5646.