LightStar-Net: A Pseudo-Raw Space Enhancement for Efficient Low-Light Object Detection

Xin Feng[1], Jie Wang[2], Siping Wang[3]
Chongqing University of Technology, Chongqing, China
xfeng@cqut.edu.cn

Jiehui Zhang*
Xi'an University of Science and Technology, Xi'an, China
jiehui_zhang@hotmail.com

Abstract

In low-light conditions, detectors trained on normal-light data often experience significant performance degradation. To address this issue, low-light image enhancement methods are commonly employed to improve detection performance. However, existing human vision-oriented enhancement techniques have shown limited effectiveness, while most machine vision-oriented methods rely on standard RGB image processing or RAW space conversion, often neglecting the preservation of key object features and incurring high computational costs. To overcome these limitations, we propose an efficient low-light object detection method based on Pseudo-RAW space enhancement—LightStar-Net. This method combines a Pseudo-RAW space Enhancement module (PRE) with a lightweight network, enhancing detection capabilities for machine vision in low-light environments. Using inverse mapping to convert RGB images into Pseudo-RAW feature space, the model dynamically adjusts image enhancement parameters to optimize detection performance. On benchmark datasets such as ExDark and DARK FACE, LightStar-Net achieves outstanding accuracy and inference speed. With a simple structure requiring only 3K parameters, it significantly improves detector performance in low-light environments.

***Keywords:** Low-Light Detection, Pseudo-RAW, Machine Vision, LightStar-Net*

1. Introduction

Object detection is a fundamental task in computer vision, aiming to recognize and locate objects within an image. Despite significant advancements in object detection algorithms [6, 9, 15, 22, 23, 25], real-world applications in low-light environments continue to face various challenges. Recently, many works have been proposed to enhance low-light image visual perception for downstream tasks such as object detection, semantic segmentation, and depth estimation.

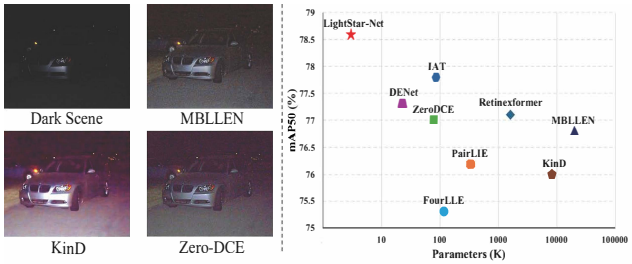

Figure 1. The left part: the results of enhancing night-time images using three different methods: MBLLEN [18], Kind [36], and Zero-DCE [10]. The right part: the results of our method compared to other methods on the ExDark [16] dataset.

Low-light object detection typically employs two main frameworks: a two-stage framework oriented towards human vision and a single-stage framework oriented towards machine vision. In the human vision-oriented framework, the enhancement network [10, 18, 36] and the detector operate as independent modules. The enhancement network is pre-trained on paired low-light and standard-light image datasets [2, 29], transforming images into well-lit versions before training the detector. However, while this processing improves image brightness, it can reduce color saturation and increase noise (as shown in Figure 1 left part), negatively impacting detection performance.

In contrast, the machine vision-oriented framework focuses on enhancing image attributes critical for detection tasks, such as contrast and target clarity, by end-to-end connecting the enhancement network [4, 5, 21, 34] and the detector, optimizing them together. This approach directly

*Corresponding author

boosts detection performance, reduces false positives, and enhances the algorithm's generalization ability under varying lighting conditions.

However, machine vision enhancement methods predominantly rely on standard RGB image processing, which can overlook the preservation and enhancement of critical target features during processing. Recently, some studies have proposed inverse mapping networks that restore RGB images to corresponding RAW space through processes such as decolorization, inverse transformation, and denoising, allowing the retrieval of more feature information [13, 35]. Nonetheless, in low-light environments, RAW images may not fully utilize this information due to fixed processing workflows, potentially leading to detail loss or increased noise in certain cases.

Additionally, some enhancement methods overlook inference speed, causing the detection model to wait for enhancement during application. Typically, most enhancement methods [33, 37] employ a down-sampling followed by an up-sampling approach to enhance images, such as the Laplacian pyramid structure to boost low-frequency information and restore high-frequency details. However, this makes the enhancement model overly complex.

To address these issues, this paper proposes an efficient low-light object detection network enhanced in the **Pseudo-RAW** space—LightStar-Net. This network is based on the concept of Pseudo-RAW space enhancement and dynamically generates optimized parameters according to the features and scenes of input images. This mechanism enables adaptive adjustment of image processing under various lighting conditions and environments, optimizing enhancement effects while avoiding detail loss caused by fixed workflows in traditional RAW image processing methods.

Specifically, LightStar-Net consists of two main components: the Pseudo-RAW space Enhancement module (PRE) and a Lightweight Enhancement Network. In the PRE, we introduce an inverse mapping network (IMRGB) and a pseudo-image signal processing (DOISP) enhancement network. The IMRGB module inversely maps RGB images to the Pseudo-RAW image feature space, enabling in-depth analysis of image features. Subsequently, the DOISP network directly operates in this Pseudo-RAW image feature space to generate adaptive optimization parameters. This design effectively avoids reduced contrast between objects and the background caused by excessive enhancement, thereby minimizing interference with the subsequent detection process.

Meanwhile, the lightweight enhancement network accelerates inference speed by reducing model parameters and computational complexity. It employs fewer convolutional layers or smaller convolutional kernels and utilizes structures like depthwise separable convolutions to significantly lower computational loads. This enables rapid inference in resource-constrained environments, making it particularly well-suited for low-light image processing tasks.

Building on this foundation, we introduce an auxiliary training strategy to expedite the extraction and learning of latent features from the PRE. To achieve this, we designed the feature stimulation network (FS-Net), which effectively integrates deep features from the PRE, optimizing the lightweight network's training process and reducing computational burdens. This collaborative mechanism ensures the lightweight enhancement network can quickly adapt to low-light environments, improving inference efficiency while maintaining high detection accuracy and responsiveness in practical applications. During the inference stage, the PRE module and FS-Net are omitted. In this way, LightStar-Net not only addresses detail loss but also significantly enhances the overall performance and practicality of low-light object detection.

We combined the proposed LightStar-Net with the classic detector YOLOv3 [22] to create an end-to-end efficient low-light object detection algorithm framework, named LightStar-YOLO. Extensive experiments were conducted on two low-light detection datasets, EXDark [16] and DarkFace [32]. The results demonstrate that LightStar-YOLO achieves state-of-the-art performance in low-light detection tasks. Notably, our model has only **3K** parameters (as shown in Figure 1 right part), significantly fewer than previous state-of-the-art low-light detection models. Additionally, the average inference speed is **0.0004s** per image, which is ten times faster than current leading methods [4, 5, 21]. Our contributions can be summarized as follows:

- We propose a novel enhancement network, LightStar-Net, for efficient low-light object detection. This network dynamically generates optimized parameters and adaptively adjusts based on the features and scenes of input images, ensuring optimal image processing results under varying lighting conditions. By avoiding detail loss caused by fixed workflows in traditional techniques when processing RAW images, this approach significantly enhances overall object detection performance.
- To accelerate the inference speed of LightStar-Net, we design FS-Net as an auxiliary training strategy that effectively integrates deep features from the PRE. This collaborative mechanism optimizes the lightweight enhancement network's training process while reducing computational burdens, enabling rapid adaptation to low-light environments. During the inference stage, omitting the PRE and FS-Net further improves inference efficiency, enhancing both the performance and practicality of low-light object detection.
- Extensive experimental results on two low-light detection datasets demonstrate that LightStar-YOLO out-

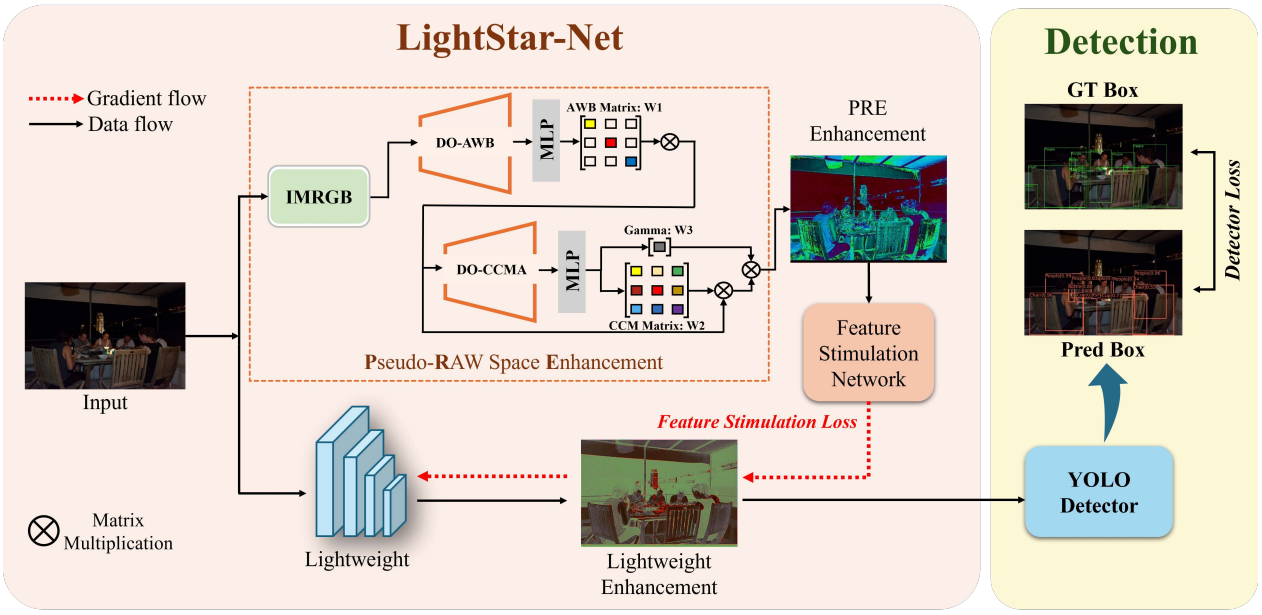

Figure 2. **The overall structure of LightStar-YOLO.** Includes a Pseudo-RAW space enhancement module (PRE), a lightweight machine vision enhancement network, a feature stimulation network (FS-Net), and the YOLO detector. During the training phase, FS-Net is used to extract image features enhanced by PRE and the lightweight network. In the inference phase, the PRE module and FS-Net are discarded.

performs current leading methods. Notably, LightStar-Net has only **3K** parameters and achieves a processing time of just **0.0004s** per image.

2. Related work

Object Detection. Mainstream object detectors are generally categorized into two types: single-stage and two-stage detectors. Single-stage detectors, such as SSD [15], YOLO [22], and FCOS [25], directly predict object bounding boxes and class labels in one step. In contrast, two-stage detectors, like RCNN [9], Faster R-CNN [23], and R-FCN [6], first generate candidate regions and then classify these regions while performing bounding box regression for refinement. The rapid advancement of object detection technologies has been largely driven by the availability of large-scale datasets, such as COCO [14] and Open Images [12], which provide extensive annotated examples for training and evaluation.

Low-Light Image Enhancement. The progress of deep learning has spurred significant advancements in low-light image enhancement techniques. Lore et al. [17] pioneered the use of deep autoencoders for dark-light image enhancement networks, establishing this field. Guo et al. [10] introduced Zero-DCE, which estimates light enhancement curves for no-reference images. Wei et al. [29] utilized the Retinex decomposition model for simultaneous light enhancement and denoising, further improving image quality. Cai et al. [1] incorporated a Transformer structure to model both reflectance and illumination damage, achieving remarkable results. While these methods aim to restore low-light images to well-lit scenes, the enhanced images often suffer from reduced color saturation and noise due to increased brightness.

The superior imaging quality of RAW images has prompted researchers to explore their use in low-light enhancement. Chen et al. [2] developed the first low-light enhancement network based on RAW images, demonstrating significant improvements in noise suppression and color saturation over traditional RGB methods. Xing et al. [31] introduced an inverse mapping network that restores RGB images to RAW space, enabling feature extraction. Zamir et al. [35] proposed the CycleISP network, which cyclically maps RGB images to RAW, adds noise, and then converts them back to RGB, facilitating the synthesis of realistic noise datasets for denoising RAW images. Although RAW images retain more details in low-light environments, they also introduce challenges such as increased noise, color distortion, and higher computational costs, which can affect model adaptability.

Low-light Object Detection. In the field of low-light object detection, a common approach is to enhance images before detection to produce brighter images [8, 19, 30], known as the human vision-oriented two-stage enhancement detection framework. Another method combines image enhancement with detection in a machine vision-oriented single-stage approach. Related research [11, 24] proposes image restoration training pipelines to improve detection robustness. Cui et al. [4] introduced the Illumination Adaptive Transformer (IAT), which dynamically adjusts image brightness by converting sRGB im-

ages to RAW-RGB space using inverse mapping. Du et al. [7] presented DAI-Net, which enhances low-light object detection through day-night domain adaptation, integrating Retinex theory and incorporating an interchange-redecomposition-coherence procedure to improve image decomposition. These methods primarily rely on RGB images and face challenges in effectively capturing complex lighting variations, often resulting in suboptimal performance in specific scenarios.

3. Method

In this section, we first introduce the PRE, the core component of LightStar-Net, providing a detailed description of its structure and composition. Next, we discuss the overall architecture of the efficient low-light detection framework, LightStar-YOLO, and explain the concept of auxiliary training, focusing on how FS-Net enhances the lightweight enhancement network's ability to extract and learn potential features from the PRE during pre-training.

3.1. Pseudo-RAW Space Enhancement

The PRE is responsible for enhancing discriminative features and contrast in dark regions for machine vision while adjusting the image color balance to suit machine vision requirements. As shown in Figure 2, the PRE comprises the Inverse Mapping Network (IMRGB) and the Pseudo-image Signal Processing (DOISP) Enhancement Network. The IMRGB module maps RGB image space to a feature space referred to as the Pseudo-RAW image feature space, which does not contain actual RAW data. This feature space is not designed to restore the RGB image to its corresponding RAW image but rather provides a simulated imaging approach optimized for machine vision during end-to-end training.

The DOISP module processes the Pseudo-RAW image feature space further. Unlike the encoder-decoder structures typically used in two-stage enhancement detection algorithms, the DOISP module directly acts on the Pseudo-RAW image feature space to perform feature optimization, similar to Image Signal Processing (ISP). This approach avoids the loss of image detail and texture caused by upsampling and downsampling in encoder-decoder structures. It also mitigates the problem of reduced contrast between the target and background caused by excessive enhancement, which can interfere with subsequent detection. Additionally, the DOISP module employs an adaptive parameter optimization approach. The enhanced features are ultimately transmitted to the detector. By leveraging the strengths of ISP-based enhancement methods, the DOISP module adaptively optimizes the Pseudo-RAW image feature space under the constraints of the object detection loss function, enabling end-to-end low-light object detection.

Inverse Mapping Network. As shown in Figure 3,

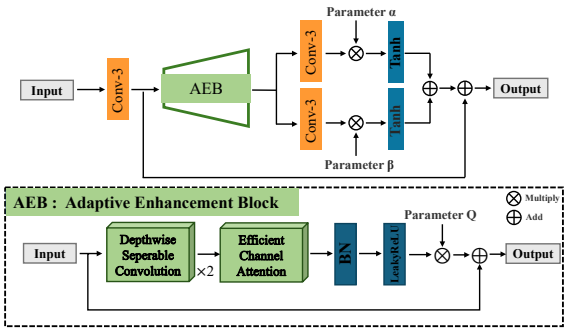

Figure 3. IMRGB Network Structure

the IMRGB module uses stacked Adaptive Enhancement Blocks (AEB) to process low-light images and extract mapping information. Initially, the module generates two independent feature maps and adjusts their channel numbers using convolution operations and a Tanh activation function.

$$[K_1, K_2] = f_{\text{AEB}}(X) \tag{1}$$

Here, K_1 and K_2 represent two independent feature maps, and X denotes the RGB image matrix under low-light conditions, and $\mathbf{X}_i$ signifies the Pseudo-RAW image feature space. The parameters α and β are adjustable and learned during training.

The AEB module primarily consists of depthwise separable convolution and the Efficient Channel Attention (ECA) mechanism [28]. Depthwise separable convolution effectively extracts image features while reducing both parameters and computational costs. The ECA mechanism uses one-dimensional convolution to adaptively adjust channel weights, enabling the extraction of key structural information from low-light images.

The ECA attention mechanism begins by compressing each channel of the feature map from a two-dimensional feature to a single value through a global average pooling layer (GAP), forming the basis for subsequent steps:

$$z_c = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \mathbf{F}_c(i, j) \tag{2}$$

where F_c represents the feature map of the c_{th} channel.

Next, the kernel size of the adaptive one-dimensional convolution is determined and applied to the feature map to generate the weight vector for each channel:

$$\mathbf{w} = \sigma(\text{Conv1D}(\mathbf{z})) \tag{3}$$

where Conv1D represents the one-dimensional convolution operation, $\mathbf{z}$ represents the global average pooling values of all channels, and σ represents the Sigmoid function.

Then, the output of the one-dimensional convolution is converted into the attention weights of the channels using

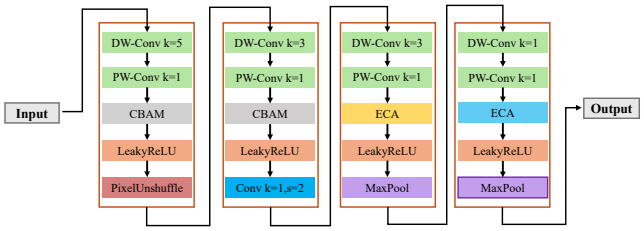

Figure 4. DO-AWB Network Structure

the Sigmoid function:

$$\mathbf{w} = \sigma(\mathbf{w}) \tag{4}$$

Finally, these learned channel weights are applied to each channel of the original feature map to achieve adaptive feature attention:

$$\mathbf{F}'_c = \mathbf{w}_c \cdot \mathbf{F}_c \tag{5}$$

where $\mathbf{F}'_c$ represents the c_{th} channel feature map after processing by the ECA attention mechanism, and $\mathbf{w}_c$ represents the attention weight of the c_{th} channel.

Pseudo-image Signal Processing. RGB images are converted into corresponding Pseudo-RAW image feature space through the IMRGB module. In the real RAW imaging and ISP algorithm enhancement mechanism, RAW images undergo ISP processing, including image compression, white balance correction, black level correction, color correction, sharpening, and other steps, which finally convert the images to RGB images. Based on this enhancement idea, we designed a pseudo-ISP enhancement network called DOISP, which directly estimates various ISP correction parameters from Pseudo-RAW images for enhancing the Pseudo-RAW image feature space.

In DOISP, we propose two enhancement sub-modules: DO-AWB (Detection-Optimized Auto White Balance, As shown in Figure 4) and DO-CCMA (Detection-Optimized Color Correction Matrix and Gamma) to simulate the enhancement process in ISP.

In ISP processing, the purpose of auto white balance processing is to compensate for color deviations caused by the color temperature environment and inherent color channel gain deviations of the shooting instrument by changing the gain of color channels in the image, thus allowing the obtained image to correctly reflect the true colors of objects. Inspired by AWB, DO-AWB uses neural networks to dynamically simulate the gain values of the three color channels:

$$W_{\mathbf{1}} = f_{\mathrm{MLP}}(f_{DO-AWB}(X_i) = \begin{pmatrix} t_1 & 0 & 0 \\ 0 & t_2 & 0 \\ 0 & 0 & t_3 \end{pmatrix} \tag{6}$$

$$X_e = W_{\mathbf{1}} \otimes X_i \tag{7}$$

In Eq. 6, X_i represents the simulated RAW image, $t_i, i \in (1, 3)$ represents the gain value of each color channel optimized, and $W_{\mathbf{1}}$ represents the generated optimization parameters. Eq. 7 represents the simulated RAW image enhanced by the generated optimization parameters, where $\otimes$ represents matrix multiplication, and X_e represents the image enhanced by DO-AWB.

The DO-CCMA module simulates the process of the color correction matrix and gamma adjustment by learning the color correction matrix and gamma values through a multi-layer perceptron (MLP):

$$W_{\mathbf{2}} = f_{\mathrm{MLP}_1}(f_{DO-CCMA}(X_e)) = \begin{pmatrix} t_1 & t_2 & t_3 \\ t_4 & t_5 & t_6 \\ t_5 & t_8 & t_9 \end{pmatrix} \tag{8}$$

$$W_{\mathbf{3}} = f_{\mathrm{MLP}_2}(f_{DO-CCMA}(X_e)) = (t_{10}) \tag{9}$$

$$X_t = (W_{\mathbf{2}} \otimes X_e)^{W_{\mathbf{3}}} \tag{10}$$

where Eq. 8 $W_{\mathbf{2}}$ represents the generated color correction matrix parameter values, $t_i, i \in (1, 9)$ represents the color correction values for each channel. In DO-CCMA, the Gamma coefficient $W_{\mathbf{3}}$ is responsible for adjusting the global brightness of the image under machine vision. The entire optimization process of DO-CCMA is shown in Eq 10, and X_e represents the final enhanced image.

3.2. Overall Structure of LightStar-YOLO

As shown in Figure 2, the efficient low-light detection algorithm framework for auxiliary training consists of a PRE module, a lightweight enhancement network, the YOLO [22] detector, and the feature stimulation network (FS-Net). This detection algorithm framework is mainly divided into two stages: the training stage and the inference stage.

In the training stage, we first combine the lightweight machine vision enhancement network with the YOLO detector to form the end-to-end joint detection framework LightStar-YOLO. Next, we freeze the pre-trained weights of PRE and use FS-Net to extract the image features enhanced by PRE and the lightweight enhancement network. By further enhancing the learning capability of the lightweight network through the feature excitation loss function, we aim to better capture the potential features of PRE, thus improving the overall detection performance of the network. In the inference stage, to maintain efficient inference speed, we discard the PRE and feature stimulation

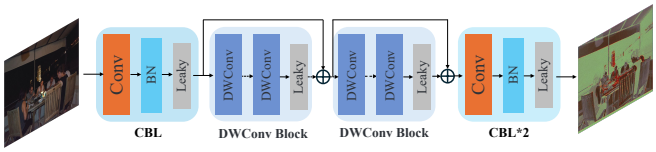

Figure 5. Lightweight Network Structure

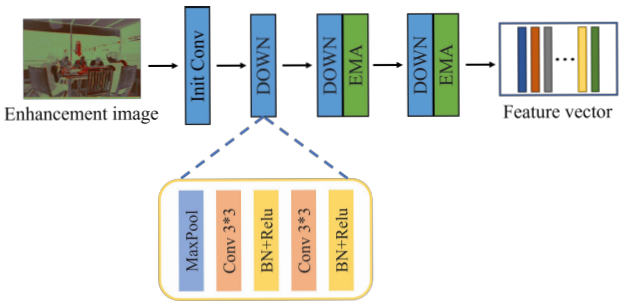

Figure 6. Feature Stimulation Network Structure

modules, retaining only the detection network LightStar-YOLO, composed of the lightweight network and YOLO. This design reduces network complexity while still ensuring the efficiency and accuracy of the detection task. Through this structural design, the entire detection algorithm framework achieves superior performance and rapid response in practical applications.

Lightweight Enhancement Network. Figure 5 illustrates that the design of the lightweight enhancement network mainly consists of CBL convolution blocks and depthwise separable convolution blocks, which draws on the design principles of residual networks. In visual enhancement tasks, downsampling feature maps can lead to a loss of spatial resolution, which diminishes the model's ability to capture fine structural information in images, making it prone to losing detail during the subsequent image reconstruction phase. To preserve as many original input features as possible throughout the enhancement process, the size of the feature maps is kept consistent with the input size, allowing for more accurate depiction and enhancement of details in the image, thus improving image quality and processing results.

Feature Stimulation Network. Based on the idea of auxiliary training, we propose a feature simulation network (FS-Net, as shown in Figure 6). The primary goal of this network is to motivate the lightweight enhancement network to better extract and learn the potential features from the pre-trained PRE network, thereby addressing the issue of poor machine vision enhancement results due to the simple structure and weak feature extraction capability of the lightweight network when processing low-light images. The input data for FS-Net comes from the results of image enhancement processed by the PRE module and the lightweight network. In the construction of FS-Net, we primarily rely on a series of stacked Down modules. Whenever an enhanced image passes through a Down module, the size of the feature map is pooled to half the size of the previous feature map. To strengthen the excitation effect, FS-Net introduces an EMA [20] attention mechanism at the ends of the upper and lower parallel branches.

3.3. Loss Function

Tung et al. [26] suggested that in knowledge distillation, preserving knowledge by computing feature map similarity enables the teacher and student networks to produce similar activations for the same samples, thereby improving the distillation effect. Based on this idea, we propose the Channel Similarity Matrix Loss Function. This loss function measures the difference in high-level feature representations by comparing the inner products of corresponding channels in two feature maps, thereby enhancing LightStar-Net's performance in machine vision imaging. The formula for defining the Channel Similarity Matrix Loss Function is as follows:

$$D_c = \sum_{i}^{w} \sum_{j}^{h} A_{b,c,i,j} \cdot B_{b,c,i,j} \tag{11}$$

$$L_{CS} = \frac{1}{N_c} \sum_{c}^{N_c} \left(\frac{\boldsymbol{D_{i,j}}}{\boldsymbol{W^2}} - \frac{\hat{\boldsymbol{D}}_{\boldsymbol{ij}}}{\boldsymbol{W^2}} \right)^2 \tag{12}$$

Where D_c denotes the inner product of corresponding feature map tensors A and B for each channel C. b represents the batch size, N_c represents the number of channels, and w and h denote the width and height of the feature maps, respectively. i and j represent indices for width and height, respectively. W^2 denotes the square of the width. The overall definition of the Feature Stimulation Loss Function is expressed as follows:

$$L_k = L_{\text{smooth}_{L_1}} + L_{CS} \tag{13}$$

In the overall algorithm framework for training the machine vision enhancement module, we define the total loss function as the sum of the detector loss function and the feature excitation loss function. The specific formula for definition is as follows:

$$L_k = \lambda_1 L_{\text{smooth}_{L_1}} + \lambda_2 L_{CS} + \lambda_3 L_{\text{dec}} \tag{14}$$

where λ_1, λ_2, and λ_3 are balancing coefficients. Based on empirical observations, we set $\lambda_1 = 0.5$, $\lambda_2 = 0.3$, $\lambda_3 = 1.0$.

4. Experiments

4.1. Training Details

We implemented our work using the open-source object detection toolbox MMDetection [3]. During the training process, we employed data augmentation strategies such as random cropping and random flipping to enhance the training of the algorithm framework for machine vision enhancement. To ensure uniform data size, we resized images to

Method	Bicycle	Boat	Bottle	Bus	Car	Cat	Chair	Cup	Dog	Motorbike	People	Table	mAP50(%)↑
YOLO [22] Baseline	80.7	73.9	77.5	92.1	83.1	66.5	71.5	78.6	76.4	76.7	81.5	57.0	76.3
MBLLEN [18] -YOLO	82.2	76.7	76.5	92.5	83.1	72.4	71.5	77.3	78.5	74.5	80.8	55.6	76.8
KinD [36] -YOLO	80.8	77.2	74.7	92.0	84.5	67.2	70.7	78.9	77.7	74.7	80.0	54.0	76.0
ZeroDCE [10] -YOLO	79.1	79.1	76.5	91.5	85.2	68.8	71.7	76.7	78.9	77.1	82.0	57.2	77.0
Retinexformer [1] -YOLO	82.0	80.5	80.9	91.3	83.1	70.8	70.3	76.9	75.8	75.4	80.8	57.8	77.1
PairLIE [8] -YOLO	82.5	76.7	76.4	91.6	82.9	71.0	70.0	76.9	78.9	72.6	79.9	55.3	76.2
FourLLE [27] -YOLO	81.8	78.1	74.2	91.3	82.6	67.7	69.1	73.0	76.1	74.8	79.5	55.3	75.3
MAET [5]	83.1	78.5	75.6	92.9	83.1	73.4	71.3	79.0	79.8	77.2	81.1	57.0	77.7
IAT-YOLO [4]	79.8	76.9	78.6	92.5	83.8	73.6	72.4	78.6	79.0	79.0	81.1	57.7	77.8
DENet [21]	80.4	79.7	77.9	91.2	82.7	72.8	69.9	80.1	77.2	76.7	82.0	57.2	77.3
PE-YOLO [34]	84.7	79.2	79.3	92.5	83.9	71.5	71.7	79.7	79.7	77.3	81.8	55.3	78.0
DAI-Net [7]	83.8	75.8	75.1	94.2	84.1	74.9	73.1	79.2	82.2	76.4	80.7	59.8	78.3
LightStar-YOLO (ours)	83.6	80.5	77.0	91.4	84.8	73.9	73.9	80.6	78.1	78.6	82.1	58.4	78.6

Table 1. Comparison of detection accuracy of different methods on ExDark. Red indicates the best result, and blue indicates the second best result.

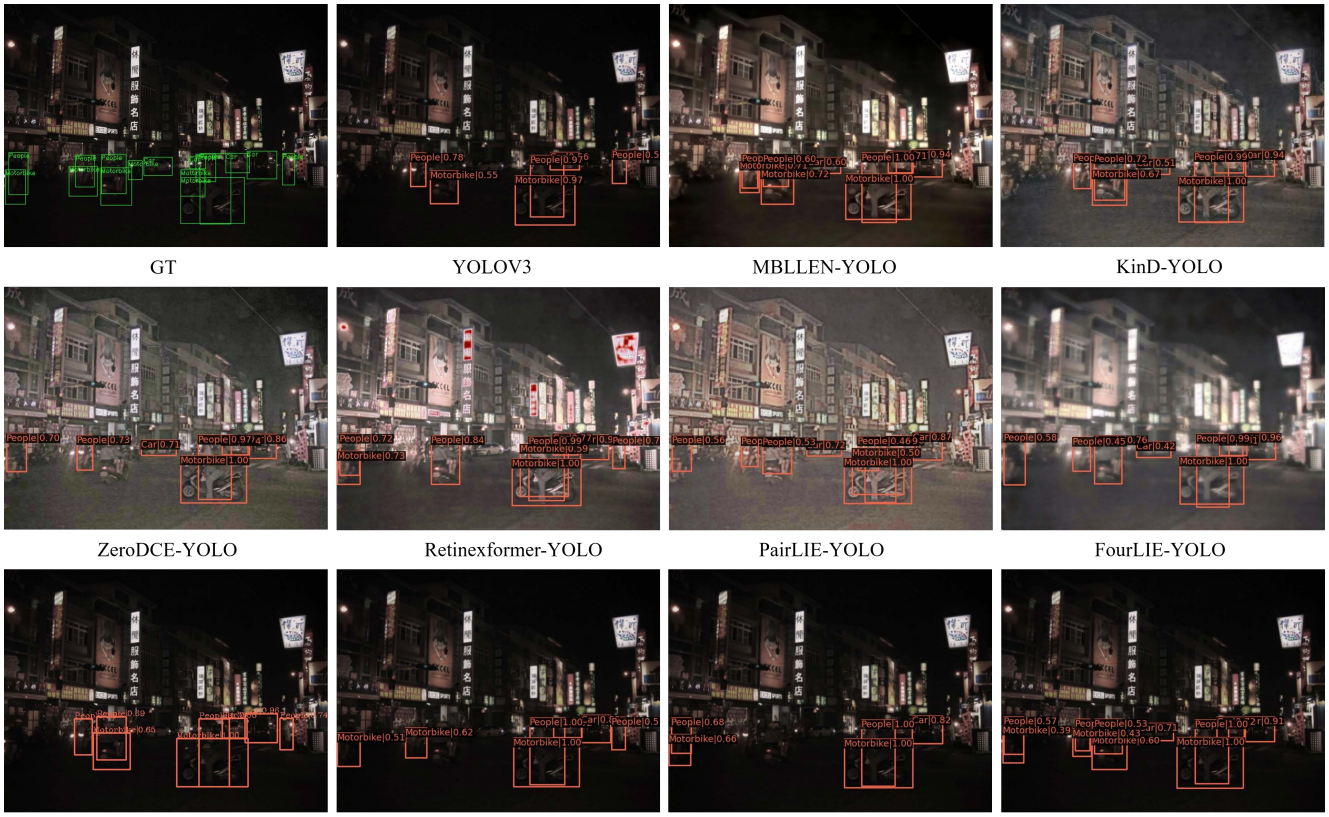

Figure 7. Visualization of detection results in nighttime street environments using different methods

608 × 608 pixels before applying augmentation. For optimization strategies, we utilized stochastic gradient descent (SGD) as the optimizer to fine-tune the enhanced detection model. The initial learning rate was set to 0.001, and we employed a step-based learning rate decay mechanism. The entire training process lasted for 25 epochs. To ensure stable training and smooth convergence, we used a learning rate warm-up strategy at the beginning of training. All experiments were conducted on hardware equipped with an Intel Core i7-12700F processor, 128GB of memory, and an NVIDIA RTX 3090 graphics card.

4.2. Object Detection in Darkness

ExDark [16] and DARK FACE [32] are two well-known low-light detection datasets. We independently trained the model using these two datasets and validated its performance with their respective test sets, and then compared it with several leading low-light scene enhancement detec-

tion models. Among them, MBLLEN [18], KinD [36], ZeroDCE [10], FourLLE [27], PairLIE [8], and Retinexformer [1] emphasize low-light image enhancement methods, where images are preprocessed before applying the YOLO detector. On the other hand, MAET [5], IAT [4], DENet [21], PE-YOLO [34] and DAI-Net [7] are single-stage low-light detection models aimed at machine vision. In the evaluation, we primarily focused on the mean Average Precision (mAP) at an IOU threshold of 0.5 and performed a visual analysis of the results on the ExDark [16] dataset.

The detection accuracy performance on the ExDark [16] dataset is shown in Table 1. The analysis results indicate that: 1) In low-light image enhancement models based on human visual recovery, simply adding the YOLO detector did not significantly improve detection performance in low-light scenes. In fact, this approach may even lead to a decline in the original method's detection performance. For example, the detection performance of KinD-YOLO and PairLIE-YOLO slightly decreased compared to the YOLO baseline model. ZeroDCE-YOLO's mAP50 (%) accuracy improved by only 0.7% over the YOLO baseline model, showing no significant enhancement in detection performance. 2) In integrated enhancement methods and domain generalization, the YOLO detector's detection performance in low-light scenes showed improvement. For example, PE-YOLO's mAP50 (%) accuracy improved by 1.7% over the YOLO baseline model, demonstrating some competitiveness. Our proposed LightStar-Net does not require pretraining on other dark datasets. LightStar-YOLO's mAP50 (%) accuracy improved by 2.3% over the YOLO baseline model. Compared to the state-of-the-art DAI-Net, the mAP50 (%) accuracy metric improved by 0.3%.

Figure 7 shows the detection performance of LightStar-YOLO in a nighttime street scene, where it outperforms DENet-YOLO and IAT-YOLO in detection results. Table 1 and Figure 7 show that the FS-Net Feature Stimulation function enables LightStar-YOLO to learn latent features from the PRE module, enhancing its machine vision capability in low-light environments. Furthermore, it outperforms other models used for low-light scene enhancement detection. These results demonstrate the effectiveness of the proposed detection algorithm framework, particularly in improving detection accuracy through the auxiliary training strategy based on the Pseudo-RAW image enhancement method.

Similar comparative experiments were conducted on the DARK FACE [32] dataset to verify the generalization capability and detection accuracy of the detection algorithm that combines Pseudo-RAW space Enhancement methods with auxiliary training strategies across different datasets. As shown in Table 2, the detection accuracy of LightStar-YOLO is 1.5% higher than that of the current state-of-the-art algorithm, with a mAP50 (%) detection accuracy of 58.5%. Additionally, LightStar-YOLO's detection accuracy also surpasses traditional two-stage enhancement detection framework methods, such as MBLLEN-YOLO, Retinexformer-YOLO, and FourLIE-YOLO. This further validates the effectiveness of the detection algorithm that combines Pseudo-RAW image enhancement methods with auxiliary training strategies in maintaining detection accuracy and demonstrates its strong generalization capability across various datasets.

Method	Face	mAP50(%)↑
YOLO [22] Baseline	54.0	54.0
MBLLEN [18] -YOLO	51.6	51.6
KinD [36] -YOLO	51.6	51.6
ZeroDCE [10] -YOLO	54.2	54.2
Retinexformer [1] -YOLO	57.8	57.8
PairLIE [8] -YOLO	56.8	56.8
FourLLE [27] -YOLO	51.2	51.2
MAET [5]	55.8	55.8
IAT-YOLO [4]	53.1	53.1
DENet [21]	51.2	51.2
PE-YOLO [34]	51.1	51.1
DAI-Net [7]	57.0	57.0
LightStar-YOLO (ours)	58.5	58.5

Table 2. Comparison of detection accuracy of different methods on DarkFace. Red indicates the best result, and blue indicates the second best result.

4.3. Inference Efficiency Analysis

Table 3 lists the performance of different models in terms of the number of parameters, runtime, and frames per second (FPS). The MAET [5] method is excluded because it does not require additional parameters and computations. Since all methods are based on the unmodified YOLO model, we only compare the image enhancement network parts of each model to analyze and compare their inference efficiency. The image size used in this experiment is 256×256. Notably, LightStar-Net performs best in terms of model parameters and FLOPs (floating-point operations), with the shortest inference time and the highest FPS. This result demonstrates the significant advantage of our proposed detection algorithm that combines Pseudo-RAW image enhancement methods with auxiliary training strategies in inference speed. Combining the data from Table 1 and Table 3, it can be observed that LightStar-Net exhibits excellent inference speed while maintaining detection accuracy, with an inference speed that is three times that of DENet [21]. Therefore, LightStar-Net has a notable efficiency advantage without significantly sacrificing machine vision enhancement capability. This lightweight and fast model is suitable for applications requiring strict real-time performance, such as smart driving, intelligent surveil-

Method	FLOPs(G)↓	Parameters(K)↓	Times(s)↓	FPS↑
MBLLEN [18]	19.95	20470	1.98121	0.50
KinD [36]	356.72	8160	0.03802	26.30
ZeroDCE [10]	2.53	80	0.00201	497.51
Retinexformer [1]	17.01	1605	0.01301	76.83
PairLIE [8]	22.34	341	0.00370	270.15
FourLLE [27]	2.54	119	0.00475	210.49
IAT [4]	1.44	90	0.00332	300.64
DENet [21]	0.31	24	0.00123	811.93
LightStar-Net (ours)	0.17	3	0.00037	2702.70

Table 3. Comparison of inference efficiency of different models. Red indicates the best result, and blue indicates the second best result.

lance, and edge computing devices with limited memory resources.

4.4. Ablation Study

To evaluate the impact of each component on overall performance, a series of ablation experiments were conducted on the ExDark [16] dataset. These experiments involved removing or modifying specific parts of the model one by one to observe the effects on performance.

IMRGB and DOISP. The PRE consists of the IMRGB and DOISP modules. We conducted ablation experiments to evaluate their impact. Table 4 shows that adding only the IMRGB module increased the mAP50 (%) detection accuracy of LightStar-YOLO to 78.1%, an improvement of 1.8% over the base YOLO detector, indicating its positive impact. The accuracy achieved by adding only the DOISP module was 77.5%. Although DOISP improved performance, its impact was relatively weaker. However, when both IMRGB and DOISP modules were used together, the detection accuracy increased to 78.6%. This demonstrates their combined effect on model performance, effectively optimizing the detection results. The IMRGB module primarily adapts low-light RGB images to a Pseudo-RAW feature space. Even without the optimization of DOISP, the detector's loss function can still optimize the Pseudo-RAW feature space, enhancing the alignment of images with its perspective, thus improving detection performance to some extent. The DOISP module further enhances the Pseudo-RAW feature space for machine vision. Directly enhancing RGB images provides limited improvement to the detector. Therefore, PRE first maps the RGB image to the machine vision feature space and then enhances the features to achieve optimal performance.

DO-CCMA and DO-AWB. The DOISP module consists of two enhancement sub-modules: DO-CCMA and DO-AWB. To investigate the effects of these two modules on enhancing the feature space of Pseudo-RAW images, we conducted a series of ablation experiments. The comparison results in Table 5 show that LightStar-YOLO with the DO-AWB enhancement sub-module improved detection accuracy by 0.2%compared to the version using only

Method	IMRGB	DOISP	mAP50(%)↑
YOLO [22]	×	×	76.3
LightStar-YOLO	✓	×	78.1
LightStar-YOLO	×	✓	77.5
LightStar-YOLO	✓	✓	78.6

Table 4. Comparison of ablation experiments for IMRGB module and DOISP module.

	IMRGB	DOISP		mAP50(%)↑
		DO-AWB	DO-CCMA	
LightStar-YOLO	✓	×	×	78.1
LightStar-YOLO	✓	✓	×	78.3
LightStar-YOLO	✓	×	✓	78.5
LightStar-YOLO	✓	✓	✓	78.6

Table 5. Comparison of lesion experiments using DO-CCMA and DO-AWB modules.

the IMRGB module. Similarly, LightStar-YOLO with the DO-CCMA enhancement sub-module improved detection accuracy by 0.4% compared to the version using only the IMRGB module. This indicates that both DO-CCMA and DO-AWB have a positive impact on enhancing the feature space of Pseudo-RAW images. The DO-CCMA module introduces a Color Correction Matrix (CCM) capable of correcting color parameters and includes gamma parameters. Through gamma brightness adjustment, this module effectively enhances the global brightness of the image in machine vision. Therefore, compared to the DO-AWB module, the DO-CCMA module has a more significant effect on enhancing the feature space of Pseudo-RAW images.

Feature Stimulation Loss. Similarly, we validated the effectiveness of the Channel Similarity Matrix Loss. As shown in Table 6, we first added the $smooth_{L_1}$ loss function independently. The results indicated that this feature excitation module significantly enhanced overall performance, increasing mAP50 (%) detection accuracy by 1.6% percentage points compared to the YOLO baseline, thereby demonstrating the effectiveness of the $smooth_{L_1}$ loss function as a feature excitation mechanism. This loss function combines the characteristics of both L_1 and L_2, allowing LightStar-Net to converge quickly while remaining robust during optimization. Next, we independently added the Channel Similarity Matrix loss function, which showed that LightStar-YOLO's detection accuracy improved by 1.3% percentage points compared to the YOLO baseline. Although its contribution to LightStar-Net was lower, when combined with the $smooth_{L_1}$ loss function, LightStar-YOLO's detection accuracy reached 78.6%, achieving the best results. This indicates a synergistic effect between the two loss functions. The Channel Similarity Matrix loss function, unlike $smooth_{L_1}$, measures differences between two feature maps. This enables the network to capture low-light visual information more accurately, enhancing machine vi-

Method	L_{CS}	$smooth_{L_1}$	mAP50(%)↑
YOLO	×	×	76.3
LightStar-YOLO	✓	×	77.6
LightStar-YOLO	×	✓	77.9
LightStar-YOLO	✓	✓	78.6

Table 6. Comparison of ablation experiments for feature stimulation loss functions.

Method	EMA	mAP50(%)↑
LightStar-YOLO	×	78.1
LightStar-YOLO	✓	78.6

Table 7. EMA attention ablation experiment comparison.

sion performance in complex environments.

Additionally, we conducted modular ablation experiments on FS-Net using the ExDark [16] dataset to validate the effectiveness of the EMA [20] attention mechanism. As shown in Table 7, the EMA attention mechanism effectively aligns the high-level features of the two enhancement networks. After adding this mechanism, the detection accuracy of LightStar-YOLO increased by 0.5%, indicating that the EMA attention mechanism effectively guides FS-Net in extracting key feature information. Through a cross-space learning strategy, the EMA attention mechanism successfully integrates the feature maps output by the two parallel sub-networks, enhancing the attention network's ability to extract features.

5. Conclusion

In this paper, we propose an efficient low-light object detection network enhanced in the Pseudo-RAW space, named LightStar-Net. To enhance the efficiency of machine vision-oriented low-light object detection, we introduce FS-Net with an auxiliary training strategy, enabling the lightweight network in LightStar-Net to rapidly extract deep latent features from the PRE module. LightStar-Net is embedded within an end-to-end object detection framework and optimized for performance. Extensive experiments on low-light object and face detection tasks demonstrate that the proposed LightStar-YOLO surpasses existing state-of-the-art methods in both detection accuracy and model complexity.

6. Acknowledgement

This work was supported by the Natural Science Foundation of Chongqing (Grant No. CSTB2022NSCQ-MSX0493), National Natural Science Foundation of China (Grant No. 62102309)

References

[1] Y. Cai, H. Bian, J. Lin, H. Wang, R. Timofte, and Y. Zhang. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12504–12513, 2023. 3, 7, 8, 9

[2] C. Chen, Q. Chen, J. Xu, and V. Koltun. Learning to see in the dark. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3291–3300, 2018. 1, 3

[3] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, et al. Mmdetection: Open mmlab detection toolbox and benchmark. *arXiv preprint arXiv:1906.07155*, 2019. 6

[4] Z. Cui, K. Li, L. Gu, S. Su, P. Gao, Z. Jiang, Y. Qiao, and T. Harada. You only need 90k parameters to adapt light: a light weight transformer for image enhancement and exposure correction. *arXiv preprint arXiv:2205.14871*, 2022. 1, 2, 3, 7, 8, 9

[5] Z. Cui, G.-J. Qi, L. Gu, S. You, Z. Zhang, and T. Harada. Multitask aet with orthogonal tangent regularity for dark object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 2553–2562, 2021. 1, 2, 7, 8

[6] J. Dai, Y. Li, K. He, and J. Sun. R-fcn: Object detection via region-based fully convolutional networks. *Advances in neural information processing systems*, 29, 2016. 1, 3

[7] Z. Du, M. Shi, and J. Deng. Boosting object detection with zero-shot day-night domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12666–12676, 2024. 4, 7, 8

[8] Z. Fu, Y. Yang, X. Tu, Y. Huang, X. Ding, and K.-K. Ma. Learning a simple low-light image enhancer from paired low-light instances. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 22252–22261, 2023. 3, 7, 8, 9

[9] R. Girshick, J. Donahue, T. Darrell, and J. Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014. 1, 3

[10] C. Guo, C. Li, J. Guo, C. C. Loy, J. Hou, S. Kwong, and R. Cong. Zero-reference deep curve estimation for low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 1780–1789, 2020. 1, 3, 7, 8, 9

[11] K. A. Hashmi, G. Kallempudi, D. Stricker, and M. Z. Afzal. Featenhancer: Enhancing hierarchical features for object detection and beyond under low-light vision. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6725–6735, 2023. 3

[12] A. Kuznetsova, H. Rom, N. Alldrin, J. Uijlings, I. Krasin, J. Pont-Tuset, S. Kamali, S. Popov, M. Malloci, A. Kolesnikov, et al. The open images dataset v4: Unified image classification, object detection, and visual relationship detection at scale. *International journal of computer vision*, 128(7):1956–1981, 2020. 3

[13] Z. Li, S. Yi, and Z. Ma. Rendering nighttime image via cascaded color and brightness compensation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 897–905, 2022. 2

[14] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick. Microsoft coco: Common objects in context. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014. 3

[15] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg. Ssd: Single shot multibox detector. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer, 2016. 1, 3

[16] Y. P. Loh and C. S. Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42, 2019. 1, 2, 7, 8, 9, 10

[17] K. G. Lore, A. Akintayo, and S. Sarkar. Llnet: A deep autoencoder approach to natural low-light image enhancement. *Pattern Recognition*, 61:650–662, 2017. 3

[18] F. Lv, F. Lu, J. Wu, and C. Lim. Mbllen: Low-light image/video enhancement using cnns. In *BMVC*, volume 220, page 4. Northumbria University, 2018. 1, 7, 8, 9

[19] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo. Toward fast, flexible, and robust low-light image enhancement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 5637–5646, 2022. 3

[20] D. Ouyang, S. He, G. Zhang, M. Luo, H. Guo, J. Zhan, and Z. Huang. Efficient multi-scale attention module with cross-spatial learning. In *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2023. 6, 10

[21] Q. Qin, K. Chang, M. Huang, and G. Li. Denet: detection-driven enhancement network for object detection under adverse weather conditions. In *Proceedings of the Asian Conference on Computer Vision*, pages 2813–2829, 2022. 1, 2, 7, 8, 9

[22] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 1, 2, 3, 5, 7, 8, 9

[23] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015. 1, 3

[24] S. Sun, W. Ren, T. Wang, and X. Cao. Rethinking image restoration for object detection. *Advances in Neural Information Processing Systems*, 35:4461–4474, 2022. 3

[25] Z. Tian, C. Shen, H. Chen, and T. He. Fcos: Fully convolutional one-stage object detection. arxiv 2019. *arXiv preprint arXiv:1904.01355*, 1904. 1, 3

[26] F. Tung and G. Mori. Similarity-preserving knowledge distillation. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 1365–1374, 2019. 6

[27] C. Wang, H. Wu, and Z. Jin. Fourllie: Boosting low-light image enhancement by fourier frequency information. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 7459–7469, 2023. 7, 8, 9

[28] Q. Wang, B. Wu, P. Zhu, P. Li, W. Zuo, and Q. Hu. Eca-net: Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11534–11542, 2020. 4

[29] C. Wei, W. Wang, W. Yang, and J. Liu. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560*, 2018. 1, 3

[30] Y. Wu, C. Pan, G. Wang, Y. Yang, J. Wei, C. Li, and H. T. Shen. Learning semantic-aware knowledge guidance for low-light image enhancement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1662–1671, 2023. 3

[31] Y. Xing, Z. Qian, and Q. Chen. Invertible image signal processing. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6287–6296, 2021. 3

[32] W. Yang, Y. Yuan, W. Ren, J. Liu, W. J. Scheirer, Z. Wang, T. Zhang, Q. Zhong, D. Xie, S. Pu, et al. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing*, 29:5737–5752, 2020. 2, 7, 8

[33] X. Yi, H. Xu, H. Zhang, L. Tang, and J. Ma. Diff-retinex: Rethinking low-light image enhancement with a generative diffusion model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 12302–12311, 2023. 2

[34] X. Yin, Z. Yu, Z. Fei, W. Lv, and X. Gao. Pe-yolo: Pyramid enhancement network for dark object detection. In *International Conference on Artificial Neural Networks*, pages 163–174. Springer, 2023. 1, 7, 8

[35] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, M.-H. Yang, and L. Shao. Cycleisp: Real image restoration via improved data synthesis. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2696–2705, 2020. 2, 3

[36] Y. Zhang, J. Zhang, and X. Guo. Kindling the darkness: A practical low-light image enhancer. In *Proceedings of the 27th ACM international conference on multimedia*, pages 1632–1640, 2019. 1, 7, 8, 9

[37] D. Zhou, Z. Yang, and Y. Yang. Pyramid diffusion models for low-light image enhancement. *arXiv preprint arXiv:2305.10028*, 2023. 2