Computer Modeling in
Engineering & Sciences

Tech Science Press

DOI: 10.32604/cmes.2024.050140

ARTICLE

Highly Differentiated Target Detection under Extremely Low-Light Conditions Based on Improved YOLOX Model

Haijian Shao[1,2,*], Suqin Lei[1], Chenxu Yan[3], Xing Deng[1] and Yunsong Qi[1]

[1]School of Computer, Jiangsu University of Science and Technology, Zhenjiang, 212003, China

[2]Department of Electrical and Computer Engineering, University of Nevada, Las Vegas, NV, 89154, USA

[3]Zhejiang Geely Automobile Research Institute, Ningbo, 315336, China

*Corresponding Author: Haijian Shao. Email: jsj_shj@just.edu.cn

Received: 28 January 2024 Accepted: 12 April 2024 Published: 20 May 2024

ABSTRACT

This paper expounds upon a novel target detection methodology distinguished by its elevated discriminatory efficacy, specifically tailored for environments characterized by markedly low luminance levels. Conventional methodologies struggle with the challenges posed by luminosity fluctuations, especially in settings characterized by diminished radiance, further exacerbated by the utilization of suboptimal imaging instrumentation. The envisioned approach mandates a departure from the conventional YOLOX model, which exhibits inadequacies in mitigating these challenges. To enhance the efficacy of this approach in low-light conditions, the dehazing algorithm undergoes refinement, effecting a discerning regulation of the transmission rate at the pixel level, reducing it to values below 0.5, thereby resulting in an augmentation of image contrast. Subsequently, the coiflet wavelet transform is employed to discern and isolate high-discriminatory attributes by dismantling low-frequency image attributes and extracting high-frequency attributes across divergent axes. The utilization of CycleGAN serves to elevate the features of low-light imagery across an array of stylistic variances. Advanced computational methodologies are then employed to amalgamate and conflate intricate attributes originating from images characterized by distinct stylistic orientations, thereby augmenting the model's erudition potential. Empirical validation conducted on the PASCAL VOC and MS COCO 2017 datasets substantiates pronounced advancements. The refined low-light enhancement algorithm yields a discernible 5.9% augmentation in the target detection evaluation index when compared to the original imagery. Mean Average Precision (mAP) undergoes enhancements of 9.45% and 0.052% in low-light visual renditions relative to conventional YOLOX outcomes. The envisaged approach presents a myriad of advantages over prevailing benchmark methodologies in the realm of target detection within environments marked by an acute scarcity of luminosity.

KEYWORDS

Target detection; extremely low-light; wavelet transformation; highly differentiated features; YOLOX

Nomenclature

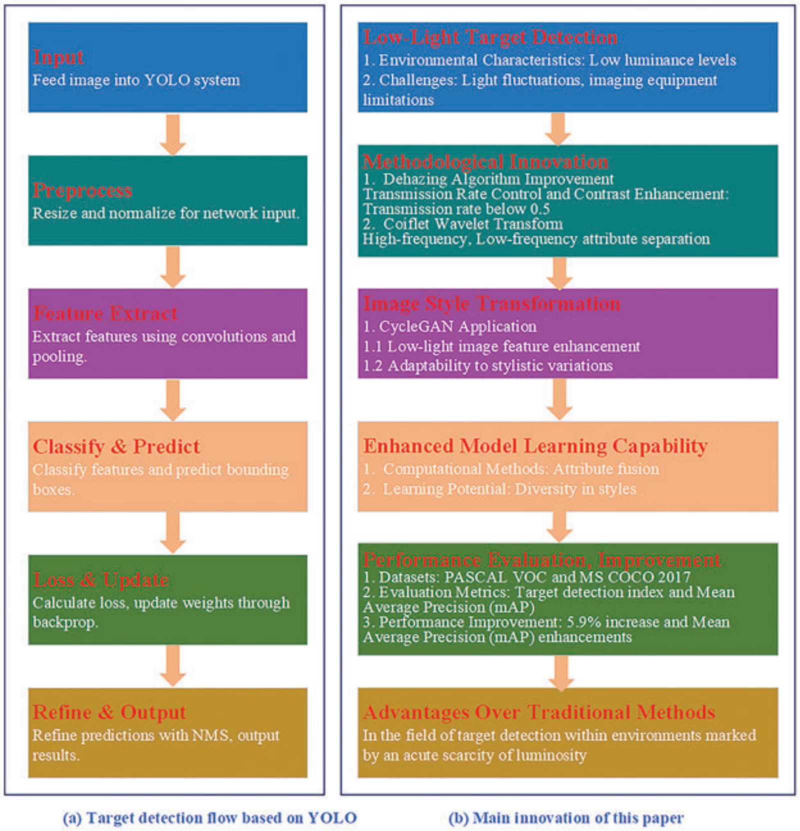
YOLOX	An advancement of the YOLO (You Only Look Once) series
CycleGAN	Cycle-consistent adversarial networks

mAP	Mean average precision
SVM	Support vector machine
CNNs	Convolutional neural networks
GIoU	Generalized intersection over union
DPLNet	Dual-path learning network
MKUO	Multiple kinds of underwater organisms
AGCL	Attribute-guided curriculum learning
AP	Average precision
ASR	Attack success rate
BAWE	Backdoor attack with wavelet embedding
AIOD-YOLO	Aerial image object detection based on YOLO
FHB	Fusarium head blight
HSIs	High-order spatial interactions
LHAB	Hybrid attention block
UAV	Unmanned aerial vehicle
STD-Conv	Spatial-to-depth convolution
Faster R-CNN	Faster region-based convolutional neural network

1 Introduction

Target detection, a pivotal method for identifying and locating multiple distinct targets in images, finds widespread applications in image retrieval [1], image classification [2,3], semantic segmentation [4,5], visual tracking [6], superpixels [7], robot navigation [8], and diverse other fields [9]. The challenge of extracting salient entities from intricate backgrounds is particularly pronounced in natural images, given their diverse contextual complexities. Traditional approaches to target detection, such as the integration of Hog features with Support Vector Machine (SVM) [10], Haar features with Adaboost [11], and the Deformable Parts Model (DPM) algorithm [12,13], predominantly rely on human feature annotation methods. Despite their effectiveness, these methods necessitate manual feature extraction, leading to labor-intensive processes and limited portability [14]. Second-order detection models, exemplified by convolutional neural networks, have emerged as advanced methodologies, surpassing traditional patterns in efficacy and precision. First-order detection models, exemplified by the YOLO [15] target detection model from Facebook's artificial intelligence lab, exhibit noteworthy reductions in detection time. With the utilization of a single training network, the test speed on a single Titan X approaches 45 frames per second, a ninefold improvement over Faster R-CNN (Faster Region-based Convolutional Neural Network).

The aforementioned factors exert a discernible influence on both perceptual subjectivity and the efficacy of target detection. Consequently, this paper endeavors to refine the image processing methodology tailored for environments characterized by an extreme paucity of luminosity, seamlessly integrating it with the YOLOX model. Moreover, it strategically leverages the robust discriminatory attributes intrinsic to Coiflet wavelet feature extraction to enhance detection precision. It is noteworthy, however, that the resultant low-light enhanced image engendered by this algorithm lacks an efficient mechanism for processing light transmittance. This deficiency results in an uneven luminous distribution across the visual rendering, precipitating suboptimal image contrast and manifesting conspicuous blurriness in areas of heightened brightness. These outcomes invariably have repercussions on both the efficacy of target detection and the subjective perceptual experience. The target detection flow based on YOLO, and main innovation of this paper are provided in Fig. 1.


Input
Feed image into YOLO system
Preprocess
Resize and normalize for network input.
Feature Extract
Extract features using convolutions and pooling.
Classify & Predict
Classify features and predict bounding boxes.
Loss & Update
Calculate loss, update weights through backprop.
Refine & Output
Refine predictions with NMS, output results.
Low-Light Target Detection
1. Environmental Characteristics: Low luminance levels
2. Challenges: Light fluctuations, imaging equipment limitations
Methodological Innovation
1. Dehazing Algorithm Improvement
Transmission Rate Control and Contrast Enhancement: Transmission rate below 0.5
2. Coiflet Wavelet Transform
High-frequency, Low-frequency attribute separation
Image Style Transformation
1. CycleGAN Application
1.1 Low-light image feature enhancement
1.2 Adaptability to stylistic variations
Enhanced Model Learning Capability
1. Computational Methods: Attribute fusion
2. Learning Potential: Diversity in styles
Performance Evaluation, Improvement
1. Datasets: PASCAL VOC and MS COCO 2017
2. Evaluation Metrics: Target detection index and Mean Average Precision (mAP)
3. Performance Improvement: 5.9% increase and Mean Average Precision (mAP) enhancements
Advantages Over Traditional Methods
In the field of target detection within environments marked by an acute scarcity of luminosity

(a) Target detection flow based on YOLO (b) Main innovation of this paper

Figure 1: The target detection flow based on YOLO, and main innovation of this paper

The subsequent sections of this paper are delineated as follows: Section 3 expounds upon the high-discrimination feature extraction methodology predicated on the Coiflet wavelet transform, in conjunction with the methodological framework for the discernment of high-discrimination targets within environments characterized by an extreme dearth of luminosity. In Section 4, a comparative analysis of target identification methodologies benchmark approaches is presented, and Section 5 encapsulates the concluding remarks of this paper.

2 Related Works

Object detection methods can be broadly classified into four types, namely: Target Detection Optimization, Feature Extraction and Model Modifications, Application-Specific Enhancements, and Low-Light Image Enhancement Techniques. Each category focuses on distinct aspects of refining and advancing the capabilities of object detection systems to meet the challenges posed by various scenarios and application domains. Object Detection Optimization Workflow with Methods is shown in Fig. 2.

2.1 Target Detection Optimization

Target detection optimization prioritizes efficiency in demanding conditions. Investigate developments in low-light picture-enhancing techniques specifically designed to improve accuracy in detecting targets. Observe the utilization of specialized convolutional neural networks (CNNs) specifically designed for detecting encoded targets, revealing a specialized method for improving the accuracy of target recognition. Participate in this investigation into the domain of target identification, where state-of-the-art tactics redefine effectiveness and precision in many situations.

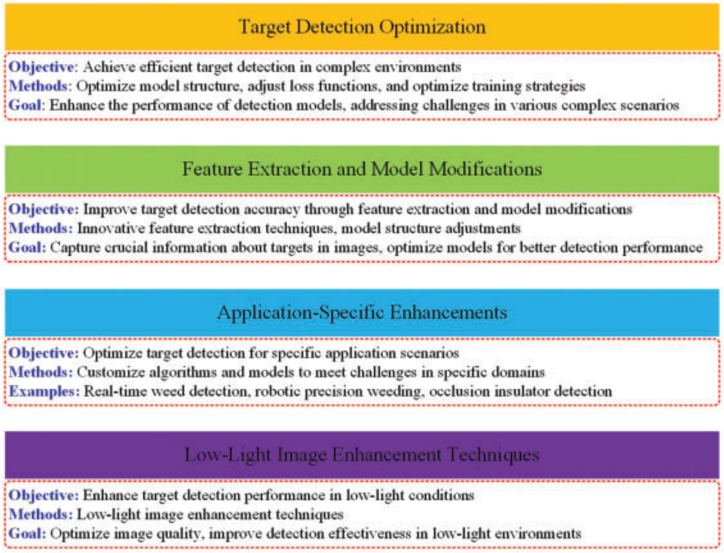

Target Detection Optimization
Objective: Achieve efficient target detection in complex environments
Methods: Optimize model structure, adjust loss functions, and optimize training strategies
Goal: Enhance the performance of detection models, addressing challenges in various complex scenarios
Feature Extraction and Model Modifications
Objective: Improve target detection accuracy through feature extraction and model modifications
Methods: Innovative feature extraction techniques, model structure adjustments
Goal: Capture crucial information about targets in images, optimize models for better detection performance
Application-Specific Enhancements
Objective: Optimize target detection for specific application scenarios
Methods: Customize algorithms and models to meet challenges in specific domains
Examples: Real-time weed detection, robotic precision weeding, occlusion insulator detection
Low-Light Image Enhancement Techniques
Objective: Enhance target detection performance in low-light conditions
Methods: Low-light image enhancement techniques
Goal: Optimize image quality, improve detection effectiveness in low-light environments

Figure 2: Object detection optimization workflow with methods

The YOLOX [16] target detection network, proposed by Joseph Redmon and Ali Farhadi, demonstrated remarkable efficiency by completing the detection of 320 $*$ 320 size images in 22 milliseconds. Following the YOLO (You Only Look Once), Zhu et al. [17] enhanced small-size feature extraction using a deconvolution layer in the residual module. Generalized Intersection over Union (GIoU) replaced IoU, minimizing positional discrepancies. The improved algorithm achieved 94.86% average detection accuracy, a 0.07% higher F1 value, and a 1.16% higher AP value compared to the original. The weeding robot, employing YOLOX, attained a detection rate of 92.45% for maize seedlings and 88.94% for weed recognition at 0.2 m/s. These findings offered crucial insights for real-time weed detection and robotic precision weeding. Leng et al. [18] described Deep-Orga, a lightweight model based on YOLOX that improved organoid detection while requiring some more computing power. They compared the model with classical models on an intestinal organoids dataset, and ablation experiments validated performance improvements. Deep-Orga provided an automated method for organoid morphology evaluation, replacing manual analysis processes. Jing et al. [19] proposed improving fruit detection in YOLOv7 by using the AlphaIoU loss function and optimizing boundary box regression. Experimental results showed significant improvements in accuracy, precision, and recall, enhancing YOLOv7-tiny model performance for fruit detection tasks. Zhan et al. [20] proposed YOLOPX, an anchor-free multi-task learning network for panoptic driving perception. YOLOPX simplified training, enhanced adaptability, and achieved optimal performance. It included a new anchor-free detection head and a lane detection head. On the BDD100K dataset, it showed state-of-the-art performance: 93.7% recall and 83.3% mAP50 for traffic object detection, 93.2% mIoU for drivable area segmentation, and 88.6% accuracy and 27.2% IoU for lane detection. YOLOPX also exhibited faster inference speed than YOLOP, making it a powerful solution for panoptic driving perception. Liu et al. [21] addressed challenges in remote sensing object detection by proposing the SDSDet detector. They introduced a non-reorganized patch-embedding layer and a dual-path learning network (DPLNet) to mitigate spatial artifacts and optimize the learning of intrinsic feature information. Furthermore, an OGF-NEM neighbor-erasing module improved the ability to find small, dense, and multi-scale objects in remote sensing images. SDSDet achieved excellent results on DOTA and MS COCO datasets, with a 42.8% AP on DOTA and 33.3% AP on MS COCO, a model size of 4.87 M, and 95 FPS. Huang et al. [22] filled in gaps in research by creating the Multiple Kinds of Underwater Organisms (MKUO) dataset, which had accurate bounding box annotations for 84 types of underwater organisms. They evaluated existing object detection algorithms on this dataset, establishing a baseline for future reference. They also proposed the Sparse Ghost Module, a lightweight module designed for object detection networks. Substituting the standard convolution with this module significantly reduced network complexity and improved inference speed without obvious detection accuracy loss. Li et al. [23] improved training efficiency for small samples by integrating low-level and high-level features. They enhanced traffic light image learning and dynamic parameter selection, achieving higher average precision on three datasets. The approach demonstrates effectiveness and potential for autonomous driving systems, meeting real-time requirements. He et al. [24] proposed the modulated intensity decoding (MID) method for car body surface defect detection. The method uses encoded fringe patterns that are projected onto the body of the car and captured reflection images to make high-quality surface defective decoded (SDD) images that clearly show defects against the background. Comparison experiments with four typical object detection networks demonstrate the superior effectiveness of MID over conventional approaches using a square-wave-like pattern light source in detecting and classifying car body surface defects.

Ng et al. [25] introduced ICText, the largest dataset for text detection and recognition on integrated circuits, which included labels for character quality attributes. They proposed the Attribute-Guided Curriculum Learning (AGCL) method to leverage labeled attributes for improved object detector performance. AGCL, when applied in a plug-and-play manner to various detectors, achieved higher average precision (AP) on ICText and proved effective on the Pascal VOC dataset, outperforming existing methods without additional computational overhead during inference. Yan et al. [26] suggested AIOD-YOLO, an improved aerial image object detection algorithm based on YOLOv8-s, to address issues associated with high-altitude imaging. They introduced a multibranch contextual information aggregation module, a multilayer feature cascade efficient aggregation network, and an adaptive task alignment label assignment strategy. In tests using the VisDrone dataset, AIOD-YOLO achieved a 7.2% improvement in mAP compared to YOLOv8-s. This demonstrated its proficiency in detecting small objects, managing changes in scale, and handling dense object distribution in aerial images. Zhang et al. [27] developed a method for counting whiteflies and fruit flies in split images, delivering reliable performance. Tailored for mobile devices, this approach showed promise in monitoring pest populations and had potential applications in various small target detection scenarios. Cao et al. [28] tackled challenges in spacecraft datasets by introducing a dataset for key spacecraft component detection and segmentation. The dataset, captured under diverse conditions, facilitated research in spacecraft-related computer vision tasks, providing valuable insights and opportunities for evaluation. Shen et al. [29] presented a method for backdoor attacks on object detection models. They introduced five attack scenarios and assessed their effectiveness using the Attack Success Rate (ASR). Moreover, they proposed a backdoor attack with wavelet embedding (BAWE) to enhance trigger stealth, revealing vulnerabilities in object detection models. Zhu et al. [30] proposed LRSNet, a lightweight remote sensing object detection network for UAVs based on YOLOv5s. It addressed issues such as complex backgrounds, small targets, and hardware limitations by employing MobileNetV3 as the backbone and CM-FPN with effective channel attention modules. LRSNet provided fast and accurate results on a variety of datasets. Ullah et al. [31] introduced an automated method for quantifying crop consistency in the early growth stage, utilizing YOLOv5 for plant counting and a lightweight U-Net for row segmentation. This method achieved high precision in canola plant detection and accurate row segmentation with fewer parameters, offering an efficient alternative for large-field assessment of plant stand count and spacing. Bao et al. [32] proposed a UAV-based method for detecting Fusarium head blight (FHB) in wheat fields. They developed a parallel channel spatial attention module and the PCSA-YOLO detection algorithm to address challenges such as overexposure, different spike orientations, and small lesions. The method outperformed YOLOv5s in precision, recall, and mAP@0.5 when preprocessing, rotation, and scaling were conducted effectively, making it a promising option for identifying FHB in agricultural settings.

Yin et al. [33] introduced HSI-ShipDetectionNet, a lightweight framework designed for accurate small-ship detection in optical remote sensing images. This model excelled at detecting objects and was computationally efficient, thanks to the use of high-order spatial interactions (HSIs) and a hybrid attention block (LHAB). When evaluated on the Kaggle and FAIR1M datasets, HSI-ShipDetectionNet outperformed existing top models. This made it suitable for deployment on resource-constrained platforms, such as those used in maritime surveillance systems. Zhao et al. proposed RA-YOLOX [34], an enhanced version of the YOLOX one-stage object detection model. RA-YOLOX introduced a re-parameterization-aligned decoupled head, which better aligned classification and regression tasks, thus improving the learning of connection information. The novel label assignment scheme of RA-YOLOX focused on high-quality positive samples and effectively filtered out low-quality ones during training. RA-YOLOX offered three lite models that surpassed similar-sized YOLOX models in performance on the MS COCO-2017 validation set. Wang et al. proposed YOLOX_w [35], an advanced algorithm based on YOLOX-X for unmanned aerial vehicle (UAV) object detection. With YOLOX-X as the baseline network, the new algorithm addressed challenges such as complex backgrounds and numerous small objects in UAV images. The algorithm enhanced small object detection by incorporating preprocessing with the SAHI algorithm and data augmentation. It also introduced a shallow feature map into the PAN network, an ultra-lightweight subspace attention module (ULSAM), and optimized the bounding box regression loss function. The experimental results on the VisDrone dataset showed an 8% improvement in detection accuracy compared to YOLOX-X, highlighting its effectiveness and robustness. Zhu et al. introduced YOLOX-ECA [36], a method for detecting damage in conveyor belts. This model was based on YOLOX, CSPDarknet, and the ECA channel attention mechanism. It achieved an accuracy of 95.65% and a speed of 30.50 FPS, effectively addressing training cost, real-time performance, and reliability issues. The inclusion of cross-domain and intra-domain transfer learning techniques further enhanced training performance and robustness, leading to improved detection accuracy compared to existing methods.

2.2 Feature Extraction and Model Modifications

Advancements in feature extraction and model modifications offer improved detection accuracy. Explore the domain of enhancing model size and speed by employing strategic loss functions, revealing the intricate equilibrium between efficiency and accuracy. Experience the potency of efficient fusion and segmentation approaches, revolutionizing the field of detection for enhanced accuracy. Embark on an exploration of feature extraction and model refining, where state-of-the-art breakthroughs redefine the benchmarks for detecting skills.

Luo et al. [37] presented an occlusion insulator detection system that utilized YOLOX to tackle the challenges of locating small insulators with extreme aspect ratios. Their method incorporated an improved SPP (Spatial Pyramid Pooling) module for enhanced semantic information extraction, an AFF-BiFPN (Asymmetric Feature Fusion Bottleneck with BiFPN) module for multi-feature fusion, and an adaptive anchor frame extraction method for improved localization accuracy. The experimental results of this system demonstrated a 90.71% precision and 88.25% recall in identifying defective insulators, showcasing its effectiveness in this niche area. Zhou's model [38] integrated a dynamic label assignment strategy to enhance precision in object detection. The model featured an upgraded CSPNet (Cross Stage Partial Network) and PANet (Path Aggregation Network), which improved multi-scale object detection capabilities. The Rep-CSPNet (Reparameterization CSPNet) ensured fast inference with reparameterization. YOLO-NL (YOLO with Non-Local) achieved a 52.9% mAP on the COCO 2017 dataset, outperforming YOLOX by 2.64%. Remarkably, it reached 98.8% accuracy at 130 FPS for face mask detection in real-life scenarios, using self-built FMD and open-source datasets. Su et al.'s approach [39] focused on enhancing reference average information by introducing MODSNet (Multi-Object Detection System Network) with MODSLayer for capturing rich inter-channel details. By maintaining the original dimensions of the light head, the model ensured lightweight non-dimensionality reduction, which significantly improved both detection accuracy and speed. Experimental results indicated a 27.5%–91.1% accuracy improvement on a crack dataset compared to YOLOX, with a reduction in both parameters and computational complexity. In the context of automated grape harvesting, detecting clusters amidst background clutter and occlusion challenges was crucial. Improvements detailed in [40] included enriching the dataset with random brightness, flip, and mosaic augmentations. The spatial-to-depth convolution (STD-Conv) module enhanced grape feature information by transforming spatial dimensions into depth. A parameter-free attention mechanism (SimAM) was applied to improve YOLOv4, YOLOv5, and YOLOX. The enhanced YOLOX achieved an 88.4% mAP, 87.8% precision, and 79.5% recall, indicating its effectiveness in identifying grape clusters for automated harvesting. In the field of semiconductor defect detection, the proposed FFDG-Cascade system [41] combined a classifier with an object detector for high-speed and accurate results. Zhu et al. enhanced the detection of small defects using shallow-to-deep attentional feature fusion. A comprehensive dataset, which included synthetic samples, helped mitigate overfitting. The integration of synthetic data boosted detector performance by 7.49 mAP. With these improvements, FFDG-Cascade achieved a 61.77% increase in speed, and the average false acceptance and rejection rates decreased by 80.67% and 59.93%, respectively, demonstrating significant advancements in semiconductor defect detection.

Wang et al. [42] introduced an automatic detection approach that utilized a modified version of YOLOX and a segmentation model named DSASNet. DSASNet employs Parallel Twins-SVT self-attention branches to eliminate mode-sensitive features and a mid-fusion module for adaptable feature integration. The incorporation of pyramid-pooling enhances the segmentation capabilities of the model. In the test set, DSASNet outperformed baseline methods with a 93.65% F1-score and 0.881 IoU, demonstrating the accuracy and efficiency of the two-step pavement distress segmentation method. In another study, a one-stage network was proposed to tackle challenges in optical remote sensing object detection, such as effective localization attention, small object compensation, and background separation strategies [43]. Through extensive experiments on public datasets, the model achieved impressive mAPs of 94.2%, 70.7%, and 80.5% on the NWPU VHR-10, DIOR, and DOTA datasets, respectively, showcasing its robustness and adaptability. Du et al. [44] focused on enhancing YOLOv4-tiny for target detection in pastoral environments. To tackle the issue of livestock size variation, they implemented a pyramid network with multiscale feature fusion and introduced a compound multichannel attention mechanism, which significantly improved accuracy. The algorithm, tested on the Jetson AGX embedded platform, achieved an 89.77% detection accuracy and a detection speed of 30 frames per second. In comparison to the standard YOLOv4-tiny, the modified version increased the average detection accuracy by 11.67% while maintaining a similar detection rate.

Zou et al. [45] chosed the high-performance YOLOX-S model for detecting external force objects in transmission line corridors. They enhanced the model to improve multi-object detection and the extraction of irregular features. The addition of a global context block enhanced perception, while a convolutional block attention module improved the recognition of objects with random features. The employment of EIoU facilitated the precise determination of object detection boxes and successful detection of external force targets. Although the model shows promise, there is a recognized need for improvements in smoke recognition, particularly in differentiating between smoke and fog targets.

2.3 Application-Specific Enhancements

Customized improvements for a specific application YOLOX plays a prominent role in transforming the process of detecting and removing weeds with high accuracy in real-time using robots. Examine the utilization of cutting-edge methods to overcome obstacles in occlusion insulator detection, highlighting the flexibility and effectiveness of YOLOX in meeting the unique requirements of the application. Participate in the investigation of customized improvements created to enhance performance in these specific fields.

Ferrante et al. [46] conducted an evaluation of various YOLO-based object detection models, including YOLOv4, Scaled-YOLOv4, YOLOv5, YOLOR, YOLOX, and YOLOv7, for their performance on small datasets related to endangered animals in Brazil. Utilizing the BRA-Dataset and focusing on data augmentation and transfer learning techniques, they found that Scaled-YOLOv4 was particularly effective at reducing false negatives, while YOLOv5 Nano delivered the highest FPS (frames per second) for detection. This research underscores the potential of YOLO-based models in wildlife conservation efforts, where training data may be limited. Jiao et al. [47] introduced a real-time litchi detection method tailored for portable and low-energy edge devices. By employing the YOLOX model, they leveraged a CNN-based single-stage detector to accurately pinpoint litchi fruit locations. Through channel and layer pruning, they compressed the model by 97.1%, yielding a compact 6.9 MB model that maintained a high average precision of 94.9% and an average recall of 97.2%. With an operational speed of 99 FPS, the method outperformed the unprocessed model by a factor of 1.8 in speed, making it an ideal solution for real-time litchi detection in orchards using portable, low-computational harvesting equipment. In the context of industrial production where hardware resources are constrained, Liu et al. [48] developed YOLO_Bolt, a lightweight adaptation of the YOLOv5 model. This model incorporates a ghost bottleneck lightweight convolution to reduce the model's size and an asymptotic feature pyramid network to enhance feature utilization and detection accuracy. By focusing on the loss function and modifying the head structure, they further improved detection precision. The model, which has half the parameters of YOLOv5s, demonstrated an increase in mAP by 2.4% and a 104 FPS improvement in testing on MS COCO 2017 and a custom bolt dataset. On the homemade dataset, the mAP 0.5 saw a 4.2% increase, outperforming YOLOv8s by 1.2%. YOLO_Bolt's enhanced performance provides robust support for workpiece detection in industrial settings. Yu et al. [49] introduced SARGap, a novel full-link automatic pruning approach for SAR (Synthetic Aperture Radar) target detectors, aiming to balance speed and accuracy. SARGap analyzes the network structure, identifies pair-coupled structures, and prunes channels accordingly. An Automatic Pruning Rate Search method (APRS) optimizes pruning rates using a Multiobjective Optimization Loss Function (FPBL). The experiments conducted on large-scale SAR target detection datasets showcased SARGap's superiority in terms of parameter and flop compression with minimal impact on accuracy, making it a versatile tool for deep learning target detectors. Zhang et al. [50] proposed the Object Knowledge Distilled Joint Detection and Tracking Framework (OKD-JDT), which amalgamates the strengths of two-stage and one-stage multiple object tracking (MOT) methods.

Within this framework, the detection network serves as a teacher, guiding feature learning in one-stage methods through knowledge distillation. For the distillation process, they designed adaptive attention learning and employed joint center point distance and Intersection over Union (IoU) for efficient tracklet generation in satellite videos. The experiments conducted using JiLin-1 satellite videos confirmed the effectiveness and state-of-the-art performance of OKD-JDT in delivering accurate and efficient MOT. Lastly, Çınar et al. [51] presented a YOLOv5-based urine analysis system for efficient particle detection in urine sediment examination (USE). The system utilizes artificial intelligence to identify and count various particles, offering automated reporting of components in centrifuged urine samples. YOLOv5m emerged as the most accurate architecture among the evaluated YOLOv5x models, with the highest mAP value of 95.8%. The system, designed to operate on a single-board computer, aims to streamline the process, standardize microscopy, and serve as educational material for laboratory personnel engaged in urinary system disease detection.

2.4 Low-Light Image Enhancement Techniques

The focus is on Low-Light Image Enhancement Techniques, which aim to optimize YOLOX for efficient vehicle detection. Additionally, Crack Detection accuracy is enhanced using MODSNet, and a Two-Step Approach to Pavement Distress Segmentation is introduced to improve efficiency. The journey towards achieving enhanced object detection in difficult low-light circumstances is revealed, providing a glimpse into groundbreaking approaches.

In the Titan X environment, the detection accuracy reaches 57.9%. The challenge of insufficient illumination in images results in suboptimal contrast and fine-grained details, thereby impacting target detection precision. A robust low-light image enhancement solution becomes imperative for restoring informational granularity. For instance, Priyanka et al. [52] employed a principal component analysis framework to enhance low-light images by decomposing luminance-chrominance components. Dong et al. [53] developed a low-light enhancement strategy using a dehazing algorithm, but it lacks effective light transmittance processing, resulting in uneven luminosity distribution, diminished contrast, and noticeable blurriness in brighter regions. Tian et al. [54] addresses challenges in underwater object detection using a lightweight model with image enhancement and multi-attention. MSRCR enhances image quality, YOLOX serves as the baseline, and GhostNet reduces computation. The multi-attention module LCR enhances feature learning and detection accuracy. Experimental results show a mAP of 77.32 with a size of 18.5 MB, 1.25 higher and 46.4 less than the baseline, demonstrating superior detection precision while keeping the model lightweight. Xi et al. [55] propose MPS-YOLO, a multi-scale information fusion network for aerial remote sensing. FPN-P reduces feature loss for similar targets, MRF addresses multi-scale challenges, and ESF enhances detection. Results show a 4.15% accuracy improvement and robustness to difficult targets. Chen et al. [56] comprehensively review object detection, exploring traditional algorithm challenges and analyzing anchor-based, anchor-free, and transformer-based approaches. They detail structures, performance, advantages, and disadvantages.

In summary, the analyzed literature primarily falls into the category of “Target Detection Optimization,” with a focus on enhancing the performance of object detection models. Several studies propose novel approaches and optimizations to address specific challenges in various domains. For instance, Cao et al. [28] address challenges in spacecraft datasets and introduce a dataset for key spacecraft component detection and segmentation, contributing valuable insights to spacecraft-related computer vision tasks. Shen et al. [29] explore backdoor attacks on object detection models, revealing vulnerabilities and emphasizing the importance of model robustness in the face of adversarial scenarios. Furthermore, the introduction of lightweight models, such as LRSNet by Zhu et al. [30], tailored for UAVs, showcases efforts to handle challenges like complex backgrounds and small targets. Ullah et al. [31] propose an automated method for quantifying crop consistency in the early growth stage, combining YOLOv5 for plant counting and a lightweight U-Net for row segmentation. In the domain of agricultural applications, Bao et al. [32] present a UAV-based method for detecting Fusarium head blight (FHB) in wheat fields, employing a parallel channel spatial attention module. Yin et al. [33] introduce HSI-ShipDetectionNet, a lightweight framework for accurate small-ship detection in optical remote sensing images, emphasizing computational efficiency. Notably, several studies extend and optimize the YOLO architecture. Zhao et al. [34] propose RA-YOLOX, an improved version of YOLOX, introducing a re-parameterization-aligned decoupled head for enhanced learning of connection information. Similarly, Wang et al. [35] propose YOLOX_w, an improved YOLOX-X algorithm for unmanned aerial vehicle (UAV) object detection, addressing challenges posed by complex backgrounds and numerous small objects.

Overall, these studies collectively contribute to the advancement of object detection techniques by addressing specific challenges in diverse applications, showcasing the importance of tailored optimizations and novel architectures for improved accuracy and efficiency.

3 Design of High-Discrimination Target Detection Method under Extremely Low-Light Conditions

The paramount phase within the dehazing procedure revolves around the computation of ambient light and transmittance. Fundamentally, transmittance undergoes refinement through the manipulation of $P(x)$ [53], constraining its range to [0,0.5], thereby attenuating the lesser transmittance to a specific threshold. However, this approach disregards the imperative consideration that the extent of reduction should diminish commensurately with the decline of $t(x)$ to $2t^2(x)$. Consequently, the ultimate imaging contrast of the original algorithm experiences degradation, particularly in scenarios where brightness-induced blurring is prevalent. The augmented low-light image enhancement algorithm demonstrably enhances image fidelity to a remarkable degree. In comparison to alternative methodologies, the visual representation derived through the elucidated method in this paper exhibits a conspicuously heightened performance across various metrics, encompassing information entropy, peak signal-to-noise ratio, spectral angle, mean square error, and even average gradient. This augmentation stands as a pivotal assurance for the precision of target detection within environments characterized by an exceedingly low luminance quotient. Moreover, the algorithm significantly amplifies the contrast between brighter regions of the image and the surrounding objects, thereby further accentuating visual perceptibility.

The high-discrimination target identification system tailored for environments characterized by a dearth of luminosity comprises four constituent elements, as shown in Fig. 3. The first section develops a low-light enhancement methodology based on the dehazing technique, which iteratively refines to produce an augmented image with significant contrast differences between bright and dark areas, ultimately improving target identification accuracy. Noteworthy is the employment of a more apt and efficacious constraint framework governing light transmittance. In view of this, it is imperative to recognize that relying solely on the integration of photographs from the dataset into the deep learning network for training purposes leads to a notable loss of both low-frequency and high-frequency information. The second section of this framework dedicates itself to formulating a high-discrimination feature extraction strategy, facilitated by the Coiflet wavelet transform. This technique enables the retention of finer-grained information accessible in each image along the horizontal, vertical, and diagonal axes. The third component introduces a deep learning model predicated on the YOLOX architecture, tailored for training. This model, emphasizing the finer attributes of the image during the training process, culminates in a more precise target detection outcome.

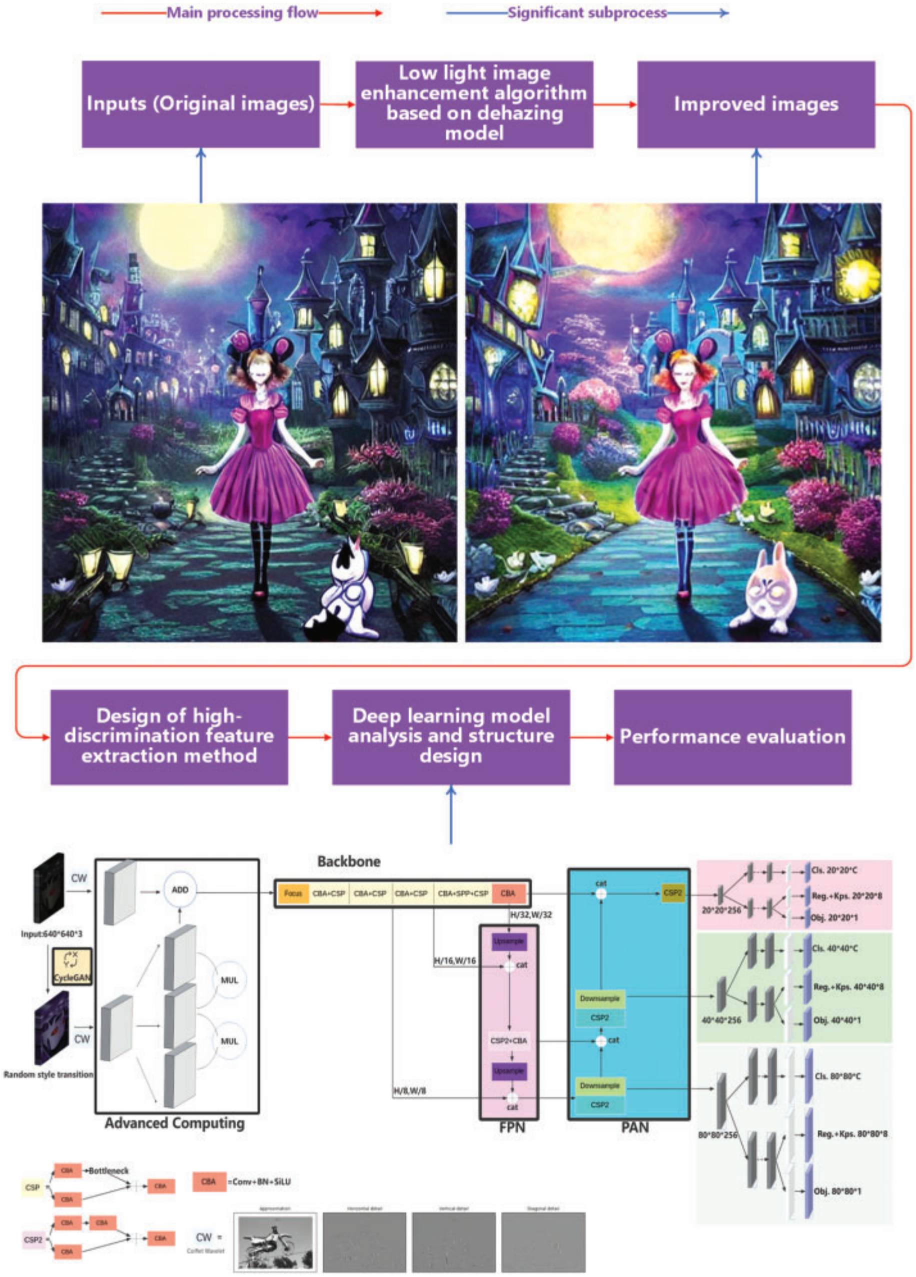


Figure 3: Highly differentiated target detection method process under extremely low-light conditions

The fourth section uses both subjective and objective assessments to give a full grade to both the improved low-light enhancement algorithm and the target identification model described in this paper. The principal contributions of this paper can be summarized as follows: (*i*) The low-light enhancement algorithm, rooted in the dehazing technique, has undergone refinement to exercise finer control over image transmission, thereby amplifying image quality and augmenting precision in target recognition. (*ii*) It is recommended to leverage the Coiflet wavelet for the extraction of highly distinctive features within the image. This strategic choice enables convolutional neural networks (CNNs) to devote greater attention to the image's horizontal, vertical, and diagonal intricacies, consequently heightening the accuracy of target detection. (*iii*) The deep learning model undergoes customization, wherein the integration of CycleGAN [57] serves to broaden the array of image features. The sophisticated computational framework advanced in this paper subsequently harmonizes these diverse image feature types, ultimately utilizing YOLOX for training purposes.

Recognizing the contextual background, it is imperative to note that the input and output images depicted in Fig. 1 serve as illustrative examples. The textual prompt, "Alice in Wonderland," is generated utilizing the Stable Diffusion [58] on Colab. The Text-to-Image latent diffusion model, Stable Diffusion, is a collaborative development involving researchers and engineers from CompVis, Stability AI, and LAION. The model is trained using 512 × 512 LAION-5B subset images, employing a frozen CLIP ViT-L/14 text encoder to condition the model on textual prompts. Noteworthy attributes include the model's lightweight nature, enabling its operation on consumer GPUs, and its 860 M U-Net and 123 M text encoder.

3.1 Low-Light Image Enhancement Algorithm Based on Dehazing Model

The generation of the low-light image involves the application of the atmospheric dehazing model to the inverted representation of the original low-light image, followed by the inversion of the processed output. This intricate process is undertaken to heighten the visual appeal of the image. Aligned with this methodology, the current paper advocates for a low-light image enhancement framework rooted in a dehazing algorithm, resembling the conditions encountered during the capture of photographs in foggy weather. The algorithm commences by subjecting the low-light image to inversion, followed by the application of the atmospheric dehazing model, aimed at restoring clarity. Subsequently, the resulting low-light image is utilized as input before undergoing channel-wise inversion within the 0 to 255 range.

$$R^c(x) = 255 - I^c(x) \tag{1}$$

where c in Eq. (1) represents the three channels of the image in RGB, $I^c(x)$ represents the pixel value of the input image on each channel (I is the set of all pixel values, x is the single-pixel value), $R^c(x)$ represents the pixel of the output image on each channel value (R is the set of all pixel values, x is the value of a single-pixel). Subsequently, the dehazing process is executed according to the model for atmospheric dehazing, as defined by Eq. (2).

$$R(x) = J(x)t(x) + A(1 - t(x)) \tag{2}$$

where $R(x)$ is the brightness of the input image, $J(x)$ is the image obtained after dehazing, and A is the brightness of the original image or scene, that is, the ambient light. $t(x)$ is the transmittance. The existing condition is only $R(x)$, so the ambient light A and the transmittance $t(x)$ must be calculated. To estimate the ambient light, the approach involves traversing all pixels in the image and sorting them in descending order based on the minimum value of each pixel across the three RGB channels. Subsequently, the first 100 pixels are selected, and the ambient light is determined by identifying the set with the highest sum of the three channels among these 100 pixels. The estimation of $t(x)$ is articulated through Eq. (3).

$$t(x) = 1 - w \min_{c\in\{r,g,b\}} \left(\min \left(\frac{R^c(y)}{A^c} \right) \right) \tag{3}$$

By transforming the form of Eq. (2), an equivalent expression, denoted as Eq. (4), can be derived.

$$J(x) = \frac{R(x) - A}{t(x)} + A \tag{4}$$

However, the enhancing impact for low-light images is relatively poor by employing this approach directly. It is considered that the region of interest can be enhanced without effecting the region of non-interest. Therefore, a constraint term $P(x)$ is introduced,

$$P\left(x\right)=\begin{cases}2t\left(x\right),0<t\left(x\right)<0.5\\1,0.5<t\left(x\right)<1\end{cases}\tag{5}$$

Thus, building upon Eqs. (4)–(5), the derived expression is denoted as Eq. (6).

$$J(x)=\frac{R(x)-A}{t(x)P(x)}+A\tag{6}$$

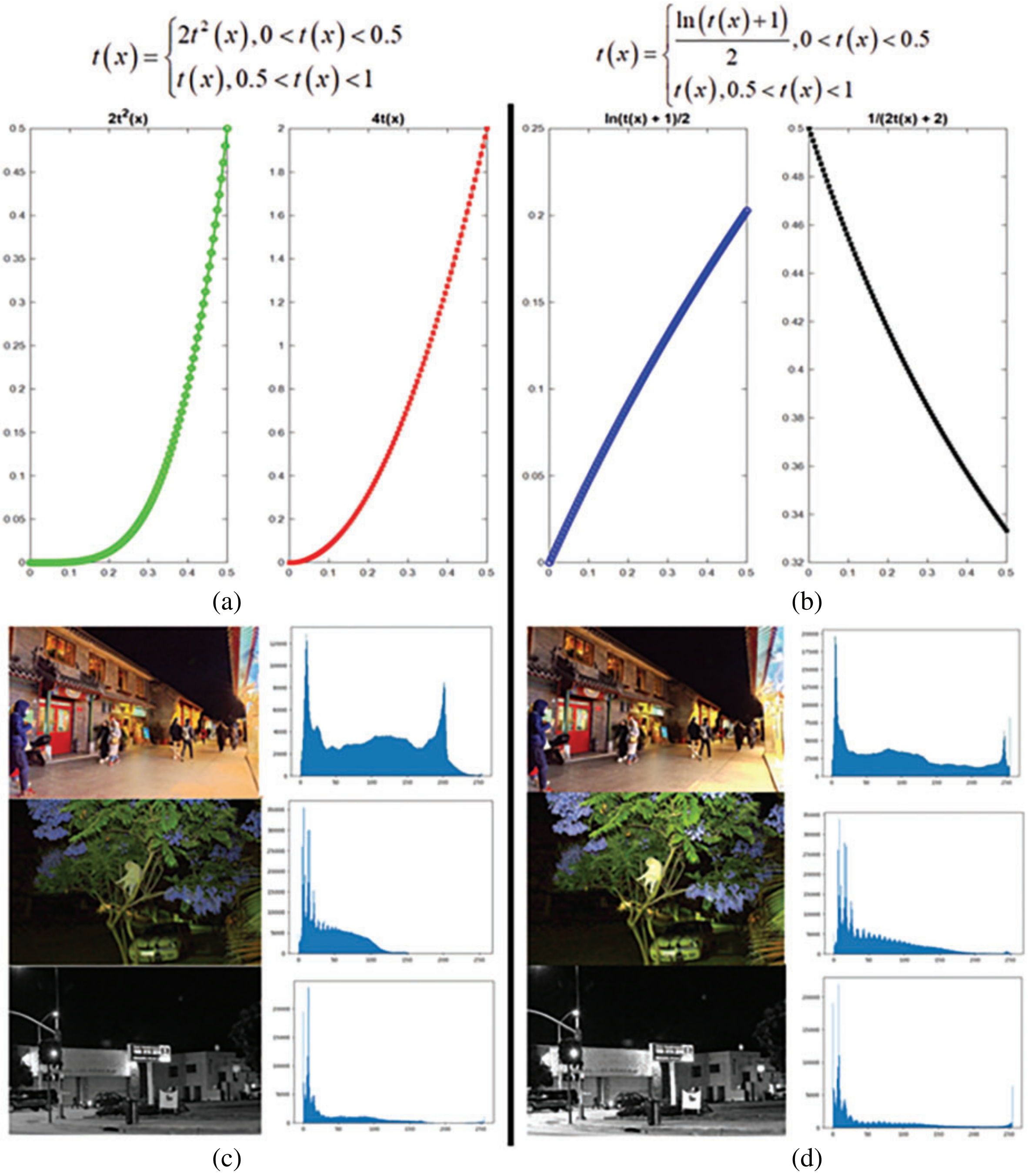
Eq. (6) implies that when $t(x)$ is less than 0.5, the corresponding pixel requires enhancement. Therefore, a small value is assigned to $t(x)$ to decrease $t(x)P(x)$ and increase the RGB intensity of the pixel. Conversely, when $t(x)$ exceeds 0.5, the original value is retained to prevent an excessive increase in the corresponding pixel intensity. In the case of $0 < t(x) < 0.5$, to make the dark place darker, although the constraint term $P(x)$, and reduced $t(x)$ to $2t^2(x)$, reducing the transmittance in the range of 0–0.5 [59], to make the transmittance $t(x)$ in the range of 0–0.5 lower. However, it simply updates the original data to a square value and uses 2 as the coefficient to make the constrained transmittance closer to the original transmittance, to make the transmittance $t(x)$ in the range of 0–0.5 lower.

This approach encounters limitations in making a small value converge further, i.e., as $t(x)$ decreases within the same range, $2t^2(x)$ fails to exhibit increased convergence in each corresponding range of decreasing $t(x)$. In essence, the objective of making smaller values in the range of 0–0.5 even smaller is not effectively achieved. Consequently, the convergence method for $t(x)$ is revised from $2t^2(x)$ to $\ln(t(x)+1)/2$. The nonlinearity of the derivative of $\ln(t(x)+1)/2$ means that, within the same reduction range of $t(x)$, the actual reduction in the value of $\ln(t(x)+1)/2$ is more pronounced. To elucidate the impact of this enhancement, Table 1 compares the changes that occur before and after the enhancement.

Table 1: Fluctuations in value prior to and subsequent enhancements

$t(x)$	$2t^2(x)$	$t_2\left(x\right)$	$t_3\left(x\right)$	$t_4\left(x\right)$
0.5	0.500	0.203	0.180	0.035
0.4	0.320	0.168	0.140	0.037
0.3	0.180	0.131	0.100	0.040
0.2	0.080	0.091	0.060	0.044
0.1	0.020	0.048	0.020	0.048

In Table 1, $t_2\left(x\right)=\ln(t(x)+1)/2$, $t_3\left(x\right)=2(t(x)^2-(t(x)-0.1)^2)$, $t_4\left(x\right)=(\ln(t(x)+1)-\ln(t(x)+0.9))/2$. Analyzing Table 1, it becomes evident that as $t(x)$ transitions from 0.5 to 0.1, the disparity in $2t^2(x)$ gradually diminishes in the process of decreasing from 0.500 to 0.020, exhibiting a drop rate of 0.040. Conversely, the improved $\ln(t(x)+1)/2$ reveals an incremental difference during the descent from 0.203 to 0.048 within the same reduction of $t(x)$, with an augmented magnitude for each decrease. Furthermore, it is noteworthy that when $t(x)$ approximates 0.5, the impact of the coefficient $P(x)$ constraint becomes less conspicuous. This is particularly prominent in the cases of 0.4 and 0.5. Under the constraints of $2t^2(x)$, values such as 0.50 and 0.32 exhibit minimal divergence from the original values. In contrast, the revised constraint not only effectively confines $t(x)$ when it is less than 0.5 but also accentuates the magnitude of restriction as the value diminishes. This results in a final image with enhanced depth and increased realism. In fact, the characteristics of $2t^2(x)$ and $\ln(t(x)+1)/2$ are not only reflected in these special values, the five values in Table 1 are just to illustrate the different constraints of the two strategies. On the image of the function $\ln(t(x) + 1)/2$, the characteristic of this decline is continuous, so all changes in $t(x)$ within the range of 0-0.5 can be constrained, and the visualization effect is shown in Fig. 4.



(a)

(b)

(c)

(d)

Figure 4: The visualization effect: Original and improved $t(x)$ function with derivatives, pre-improved and improved low-light image enhancement

The blue lines in Figs. 4a and 4b depict the functions $2t^2(x)$ and $\ln(t(x)+1)/2$, respectively, while the yellow lines represent their corresponding derivatives. The visual representations and monotonic behavior of these two functions reveal that, despite the contraction of the dependent variable and the initial diminishment of the independent variable in the case of $2t^2(x)$, the magnitude of reduction remains constant. This constancy arises from the linear connection in its derivative, implying that reducing $t(x)$ in a continuous interval does not ensure a more convergent value after applying the $2t^2(x)$ constraint each time, compared to the preceding value in a similar interval. As $t(x)$ falls, the $2t^2(x)$ also falls linearly, which does not have a greater restricting effect on lesser $t(x)$ values. Instead, it moves in the opposite direction. Once more looking at the $\ln(t(x)+1)/2$ function and its derivative image, it can be observed that even if $t(x)$ progressively shrinks from 0.5, $\ln(t(x)+1)/2$ still gradually decreases, with each decline having a higher magnitude. In stark contrast to the original $2t^2(x)$, this constraint technique ensures that the value after the constraint can indeed decrease in tandem with the decrease in $t(x)$. Figs. 4c and 4d illustrate the intuitive effects and accompanying histograms of low-light image processing [53] and the approaches proposed in this paper, respectively. It is evident that the image constrained by the original method exhibits inferior performance in terms of light contrast. It tends to blend with the surrounding brightness, resulting in diminished contrast and clarity, especially in well-lit areas. Conversely, the image processed using the suggested method elicits a more pronounced sensory effect, accentuating the overall contrast of the image. Furthermore, the histogram reveals a greater abundance of data. Following this modification, the entire procedure can be summarized as follows: Step 1) Enter a low-light image. Step 2) Reverse the three channels of the low-light image in the range of 0–255. Step 3) Use atmospheric dehazing model for dehazing; Step 4) Judge the transmittance $t(x)$ obtained in Eq. (3), if $0 < t(x) < 0.5$, $t(x) = ln\left(t\left(x\right)+1\right)/2$, otherwise t(x) keep the original value; Step 5) Reverse the processed image to get a low-light enhanced image.

3.2 Design of High-Discrimination Feature Extraction Method

The Coiflet wavelet exhibits superior performance in terms of orthogonality, biorthogonality, and the maintenance of a good vanishing moment and tight support in both the frequency and time domains. Employing the two-dimensional wavelet transform enables the further subdivision of each image's low-frequency information into high-frequency information at various resolutions. The Coiflet wavelet, characterized by a wide support range of $6N$-1 with proximity to symmetry, serves as a biorthogonal wavelet. The wavelet function features a $2N$ vanishing moment, while the scale function possesses a $2N$-1 vanishing moment. This wavelet's higher compression ratio results in smaller high-frequency coefficients, a flatter filter, and more concentrated image energy after wavelet decomposition. Consequently, maximizing the number of zero wavelet coefficients or minimizing non-zero wavelet coefficients facilitates data compression and noise elimination. This phenomenon is often referred to as the amplitude of the vanishing moment determining the image's vibration level after decomposition. The Coiflet wavelet's effective balance between support length and calculation time is crucial. Longer support lengths require more calculation time and generate more high-amplitude wavelet coefficients, which inversely correlate with the vanishing moment. If the support length is excessively long, boundary issues may arise, while an overly short support length inhibits signal energy concentration due to a diminutive vanishing moment. The concepts of vanishing moment and support length are, therefore, mutually exclusive. The Coiflet wavelet accurately reflects this balance with its favorable vanishing moment and robust support in both the frequency and temporal domains. In terms of orthogonality, biorthogonality, and spectrum utilization rate, the Coiflet wavelet surpasses the standard Gaussian function. The two vector spaces V_j and W_j formed by the scale function of the Coiflet wavelet and the wavelet function are defined as Eqs. (7) and (8), respectively.

$$V_j = s^{i/2}\phi(s^i x - k)|j, k \in Z \tag{7}$$

$$W_j = s^{i/2}\psi(s^i x - k)|j, k \in Z \tag{8}$$

In 2D space, scale-space $V_j(x1, x2)$ and wavelet space $W_j(x1, x2)$ are defined as Eqs. (9) and (10), respectively.

$$V_{j-1}(x1, x2) = V_j(x1, x2) \oplus W_j(x1, x2) \tag{9}$$

$$W_j(x1, x2) = V_{j-1}(x1, x2)/V_j(x1, x2) \tag{10}$$

The scale-space subspaces V_j exhibit a nested relationship, and the wavelet space W_j plays a crucial role in capturing information between the adjacent scale subspaces V_{j-1} and V_j. It functions to capture the information lost as V_{j-1} approaches V_j. In essence, the vector space V_j and the vector space W_j are orthogonal, signifying that W_j can effectively represent the information that cannot be expressed within V_j. Therefore, the combined information from both vector spaces, V_j and W_j, enables a complete representation of the information. For any function $f(x1, x2)$ in the scale-space based on Eqs. (9) and (10), then

$$P_{j-1}f(x1, x2) = P_j f(x1, x2) + D_j f(x1, x2) \tag{11}$$

Since $P_{j-1}f(x1, x2)$ represents the projection of function $f(x1, x2)$ on space $Vj - 1(x1, x2)$, so $P(j - 1)f(x1, x2)$ can be represented by the components $Vj(x1, x2)$ and $W_j(x1, x2)$ in $Vj - 1(x1, x2)$. Since two-dimensional space can be decomposed into one-dimensional space, $Vj - 1(x1, x2)$ can be decomposed as

$$\begin{aligned} V_{j-1}(x1, x2) &= V_{j-1}(x1) \otimes V_{j-1}(x2) \\ &= [V_j(x1) \oplus W_j(x1)] \otimes [V_j(x2) \oplus W_j(x2)] \\ &= [V_j(x1) \otimes V_j(x2)] \oplus [V_j(x1) \otimes W_j(x2)] \oplus [W_j(x1) \otimes V_j(x2)] \oplus [W_j(x1) \otimes W_j(x2)] \end{aligned} \tag{12}$$

$$\phi_{ik1}(x1)\phi_{ik2}(x2) = s^{i/2}\phi(s^i x1 - k1)s^{i/2}\phi(s^i x2 - k2) \tag{13}$$

The orthogonal normalization basis of Eq. (12) is given by Eq. (13). The functions $\phi_{ik1}(x1)$ and $\phi_{ik2}(x2)$ in the aforementioned formula are low-pass scale functions. Therefore, the space $V_j(x1, x2)$ expressed by this formula represents the low-frequency characteristics of the original space. Another component utilized to represent $P_{j-1}f(x1, x2)$ is $W_j(x1, x2)$. In accordance with the corresponding relationship between Eqs. (9) and (10), $W_j(x1, x2)$ is expressed as Eq. (14).

$$W_j(x1, x2) = [V_j(x1) \otimes W_j(x2)] \oplus [W_j(x1) \otimes V_j(x2)] \oplus [W_j(x1) \otimes W_j(x2)] \tag{14}$$

The upper equation can be decomposed into three parts, and each part of the orthogonal basis is defined by Eq. (15).

$$\begin{aligned} &\phi_{ik1}(x1)\psi_{ik2}(x2) = s^{i/2}\phi(s^i x1 - k1)s^{i/2}\psi(s^i x2 - k2) \\ &\psi_{ik1}(x1)\phi_{ik2}(x2) = s^{i/2}\psi(s^i x1 - k1)s^{i/2}\phi(s^i x2 - k2) \\ &\psi_{ik1}(x1)\psi_{ik2}(x2) = s^{i/2}\psi(s^i x1 - k1)s^{i/2}\psi(s^i x2 - k2) \end{aligned} \tag{15}$$

where $\phi_{ik1}(x1)\psi_{ik2}(x2)$, $\psi_{ik1}(x1)\phi_{ik2}(x2)$ and $\psi_{ik1}(x1)\psi_{ik2}(x2)$ represent high-frequency characteristics in horizontal, vertical, and diagonal directions, respectively. Therefore, $P(j - 1)f(x1, x2)$ is represented by components in $V_j(x1, x2)$ and $W_j(x1, x2)$ as

$$\begin{aligned} P_{j-1}f(x1,x2) &= P_j f(x1,x2) + D_j f(x1,x2) \\ &= \sum_{k_1k_2} x^{(j)}_{k_1k_2}\phi_{jk1}(x1)\phi_{jk_2}(x2) + \sum_{k_1k_2} \alpha^{(j)}_{k_1k_2}\phi_{jk1}(x1)\psi_{jk_2}(x2) \\ &\quad + \sum_{k_1k_2} \beta^{(j)}_{k_1k_2}\psi_{jk1}(x1)\phi_{jk_2}(x2) + \sum_{k_1k_2} \gamma^{(j)}_{k_1k_2}\psi_{jk1}(x1)\psi_{jk_2}(x2) \end{aligned} \tag{16}$$

Consequently, $f(x1, x2)$ is decomposed into three directions utilizing both low-frequency and high-frequency features. In the two-dimensional context of target identification, a single sample can be subdivided into four samples, each representing the approximate features of the original image along with high-frequency characteristics in the horizontal, vertical, and diagonal directions. This subdivision is depicted in Fig. 5.

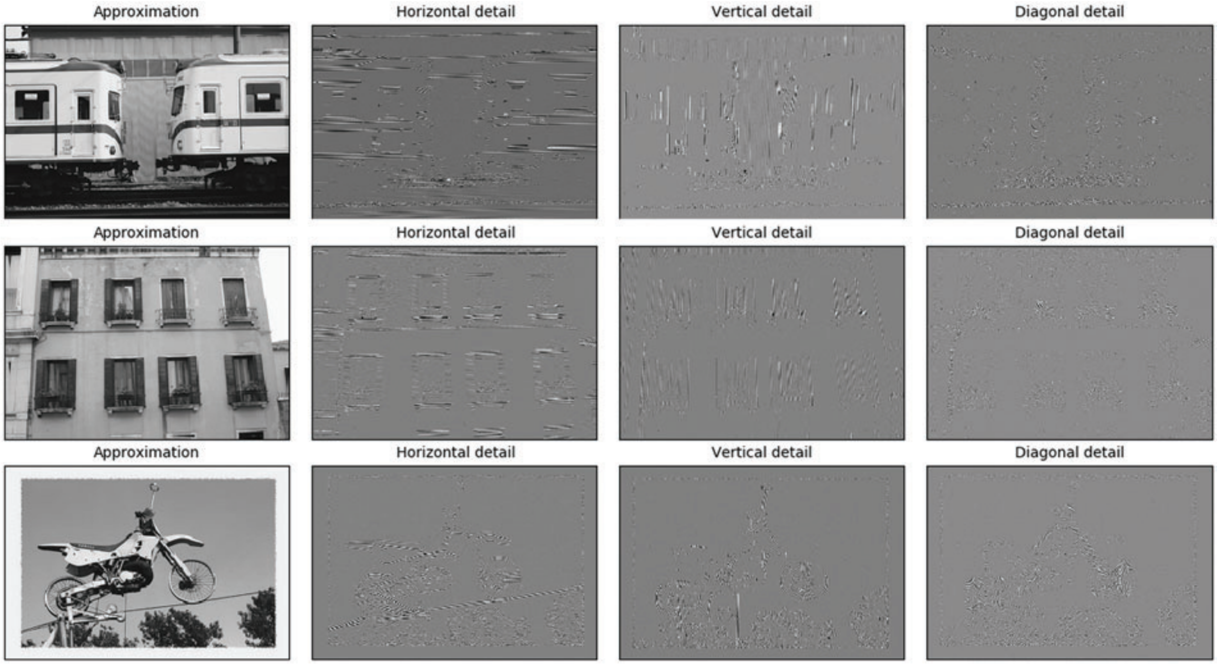


Figure 5: The efficacy of our highly differentiated feature extraction method

3.3 Deep Learning Model Analysis and Structure Design

The proposed approach for target identification in extremely low-light conditions, as illustrated in Fig. 6 is distinguished by the integration of CycleGAN, state-of-the-art computational techniques, and the YOLOX model architecture. The incorporation of the CycleGAN model aims to produce an original image with enhanced generalization capacity for low-light photographs, addressing inherent biases in image presentations across diverse contexts and devices. The generator and discriminator modules of the CycleGAN model, equipped with robust generative and discriminative capabilities, establish a dynamic equilibrium through stochastic image modifications. Following this, high-discrimination feature extraction, facilitated by the Coiflet wavelet transform, is applied to both the processed image and the original input. Advanced computing techniques are then strategically employed to amalgamate nuanced features derived from the two images. As outlined in Eq. (17), the essence of advanced computing involves duplicating a feature layer into three copies, element-wise multiplication of the three feature matrices through subsequent layer multiplication operations, and activation using the rectified linear unit (ReLU) function.

$$F(x) = relu\left(\sum\nolimits_{r=1,2,3} \langle w^r, \otimes_r y\rangle\right) + x \tag{17}$$

where w^r represents the rth feature vector, y is the detailed feature of the style-transformed image, and $\otimes_r y$ signifies the r-order self-outer product of y, providing insights into specific aspect interactions. The detailed details of the original image are then combined with the residual unit.

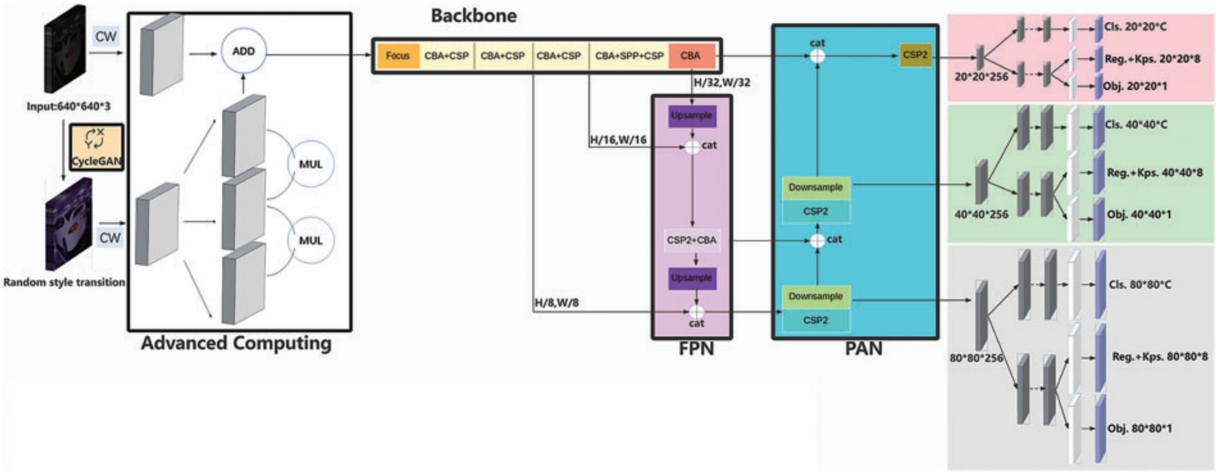


Figure 6: The overall structure of the deep learning model

3.4 *Performance Evaluation*

The enhanced low-light enhancement algorithm proposed in this paper, along with other low-light enhancement algorithms, has been objectively evaluated based on five indicators: information entropy (E), peak signal-to-noise ratio (PSNR), spectral angle (SAM), mean square error (RMSE), and average gradient (G). The calculation methods for these five evaluation indicators are defined by Eqs. (18)–(22), respectively.

$$E = -\sum\nolimits_{i=0,\ldots,255} \sum\nolimits_{j=0,\ldots,255} p_{ij} \log_2 p_{ij} \tag{18}$$

where $P_i j$ represents the proportion of pixels in the image with grayscale values and neighborhood grayscale means in the range of 0–255.

$$PSNR = 10 \times \log\left(\frac{255^2}{\sum_{i=0}^{M-1}\sum_{j=0}^{N-1}\frac{(f(i,j) - g(i,j))^2}{MN}}\right) \tag{19}$$

where M and N represent the rows and columns of the image, respectively, and $f(i,j)$ and $g(i,j)$ represent the pixel values of the i-th row and j-th column before and after the image enhancement, respectively.

$$SAM = \cos^{-1}\frac{d^T x}{(d^T d)^{1/2}(x^T x)^{1/2}} \tag{20}$$

where d is the two-dimensional matrix of the image before enhancement, and x is the two-dimensional matrix of the image after enhancement.

$$RMSE = \left(\frac{1}{MN}\sum_{i=1}^{M\times N}\left(y_i - \hat{y}_i\right)^2\right)^{1/2} \tag{21}$$

where M and N represent the rows and columns of the image, respectively, and y_i and $\hat{y}_i$ represent the pixel values of the i-th row and j-th column before and after the image enhancement, respectively.

$$G = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j}^{N}\left(\frac{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}{MN}\right)^{1/2} \tag{22}$$

where M and N represent the rows and columns of the image, $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ respectively represent the gradient of the image in the horizontal and vertical directions. The assessment of the target detection model mandates a comprehensive evaluation across an extensive array of test datasets. Standard evaluation metrics for neural network models encompass considerations of speed, accuracy, and minimal memory utilization. Within the ambit of the target detection model, a primary criterion for evaluation is the Mean Average Accuracy (mAP), an established metric employed to gauge overall accuracy. The mAP metric computes the average precision across various classes of detected targets, entailing the calculation of *Recall* and *Precision* values for each specific category of target detection. The *Recall* and *Precision*, instrumental in evaluating the performance of the proposed approach, are defined as Eqs. (23) and (24), respectively.

$$Recall = TP/(TP + FP) \tag{23}$$

$$Precision = TP/(TP + FN) \tag{24}$$

where T, F, P and N represent the correct detection, error detection, positive sample and negative sample, respectively. TP represents positive sample detected correctly, FP represents positive sample detected incorrectly, and FN represents the detect false negative samples. The Intersection over Union (IoU) threshold serves as a crucial metric in target detection, quantifying the extent of overlap between the actual bounding box and the predicted bounding box. This threshold is pivotal in determining the validity of a given detection by assessing the ratio of intersection to union. Accurate delineation of a target's precise location within an image is of paramount importance, involving scrutiny of the ground truth position specified in the label and discerning the corresponding visual representation. Subsequently, utilizing the detector, bounding boxes are identified and ranked in descending order of confidence. The bounding box with the highest confidence score is then compared to the ground truth position. If the IoU value exceeds the specified threshold, it is marked as a True Positive (TP), indicating a successful detection. The remaining instances are labeled as False Positives (FP), and the object has not been accurately detected. Based on the *Recall* and *Precision* values, one can construct a series of Precision-Recall curves. The curve has Recall as its abscissa and Precision as its ordinate, with the area under the curve reflecting the accuracy Average Precision *(AP)* of this type. The Mean Average Precision (mAP) represents the average accuracy rate across 20 item categories in the VOC dataset.

$$mAP = \left(\sum_{i=1}^{c} AP\_i \right) / C \tag{25}$$

where C represents the object category of the data set, and AP_i represents the AP of each category. After calculating the AP for all 20 categories, the mAP is obtained by taking the mean (average) of all the individual AP values. It provides an overall measure of the model's performance across all categories in the dataset. mAP is considered an important metric because it considers the performance of the model at multiple confidence thresholds, giving a more comprehensive assessment of its accuracy compared to just a single point on the precision-recall curve.

4 Experiment

4.1 Dataset Used for Experiments

In this experiment, the PASCAL VOC dataset and the MS COCO2017 dataset [60] were employed. The VOC datasets are categorized into vehicle, household, animal, and person classes. The trainval datasets of PASCAL VOC 2007 [61] and PASCAL VOC 2012 [62] are utilized for training and validation, with 90% allocated for training and 10% for validation. The dataset comprises a total of 16,551 photos (5,001 for VOC 2007 and 11,540 for VOC 2012) and 40,058 detection objects (12,608 for VOC 2007 and 27,450 for VOC 2012). The test dataset is the VOC 2007 test collection, containing 4,952 images and 12,032 detection objects, as outlined in Table 2. The MS COCO2017 dataset includes 118,287 training sets, 5,000 validation sets, and 40,670 test sets, covering 80 categories with over 500,000 annotations. With an average of 7.2 targets per image, it stands as the largest and most well-known object detection challenge dataset.

Table 2: Dataset for experimentation

–	Train	Validate	Test
Data set	VOC2007+VOC2012	VOC2007+VOC2012	VOC2007+VOC2012
Images	14896	1655	4952
Objects	36052	4006	12032

4.2 Low-Light Enhancement Algorithm Analysis

This paper subjectively evaluates seven low-light enhancement algorithms based on the Retinex theory, including *SSR*, *MSR*, *MSRCR*, *MSRCP*, *Gimp*, and *FM*. As illustrated in Fig. 7 the six specified models exhibit significant overall distortion, weak contrast, and poor sharpness from the perspective of sensory effects. In contrast, the improved algorithm proposed in this paper offers a better overall sensory experience and more coordinated light-dark contrast compared to previous methods. Five metrics are used in Table 3 to measure how well the improved algorithm works: information entropy (E), peak signal-to-noise ratio (PSNR), spectral angle (SAM), mean square error (RMSE), and average gradient (G). The results show that the proposed algorithm produces an image with increased information entropy and peak signal-to-noise ratio, indicating enhanced information richness and improved image quality with minimized distortion. The spectral angle index is lower for the proposed algorithm, suggesting that the processed image and the original image share more similar spectral characteristics when represented as high-dimensional vectors. This enhances the likelihood that the processed image corresponds to analogous objects in the original image.

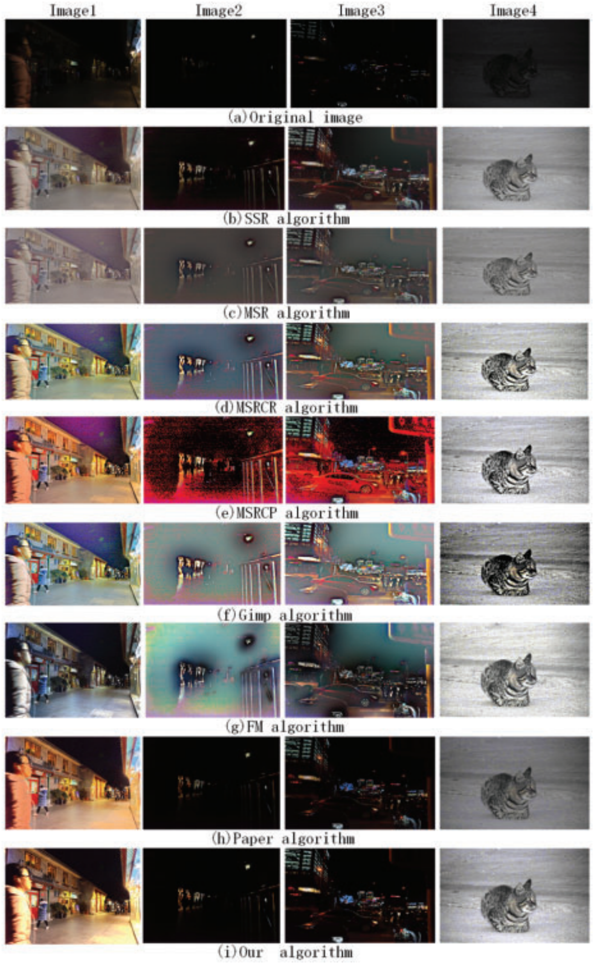


Figure 7: Subjective comparison of several algorithms

Table 3: An objective evaluation of several algorithms

–	E	PSNR	SAM	RMSE	G
SSR	7.3265	6.4193	0.8799	121.03	4.5637
MSR	6.4932	7.0757	0.9767	112.91	4.2225
MSRCR	7.5419	6.5554	0.8993	119.88	10.4752
MSRCP	7.5950	6.4766	0.8945	120.97	11.3002
Gimp	7.6698	6.1437	0.8985	125.70	11.9929
FM	7.7125	7.7785	0.8715	105.13	9.1544
Dong et al. [53]	7.7069	7.5162	0.8260	108.07	6.9699
This paper	7.7424	7.7816	0.7790	104.72	7.2279

The RMSE is a pixel-based metric that quantifies the deviation between the fused image and an ideal reference image. The proposed algorithm results in a lower mean square error, indicating higher image quality due to reduced disparities between the processed and original images. The average gradient represents the mean rate of grayscale transition, reflecting variations in contrast within the microscopic features of the image and serving as an indicator of image sharpness. While some algorithms (MSRCR, MSRCP, Gimp, and FM) exhibit higher index values, implying sharper images, these values do not consistently correlate with improved imaging effects or align with other evaluation criteria. Consequently, these algorithms do not demonstrate superior performance, as they yield images with diminished information richness, heightened distortion, and larger discrepancies compared to the original image.

The algorithm proposed in this research demonstrates its overall superiority across the five indicators in Table 4, resulting in better consistency between the processed and original images, improved image clarity, and reduced distortion. Specifically, the processed image exhibits higher information entropy, enhanced image contrast, and an elevated information preset.

Table 4: The precision of many models in identifying the detected object

–	This paper	YOLOX	YOLOv4	RFBnet	Mobilenet-SSD	Faster R-CNN	M2det
Plane	79.62	74.23	58.51	69.53	39.01	62.90	70.18
Bicycle	81.87	75.78	70.85	79.91	41.79	80.88	79.86
Bird	76.09	55.70	42.82	62.46	33.40	52.60	58.63
Boat	72.88	66.87	39.52	49.86	28.40	50.67	53.16
Bottle	54.22	49.54	32.33	32.38	13.50	29.32	27.00
Bus	86.73	78.52	70.00	75.90	42.81	76.51	73.54
Car	89.43	73.54	66.43	76.14	40.71	74.60	76.87
Cat	74.43	67.65	44.44	71.92	46.76	72.61	64.65
Chair	65.87	55.89	27.18	43.91	22.06	41.73	38.50
Cow	78.67	68.65	21.62	66.87	33.35	55.74	61.84
Table	77.98	70.24	48.89	68.20	44.03	71.90	66.00
Dog	76.56	66.52	48.87	67.08	44.20	65.33	61.95
Horse	89.76	80.54	62.59	77.63	46.73	78.88	77.53
Motor	89.34	78.15	55.62	71.91	42.96	74.58	74.44
Person	85.67	78.62	56.32	66.10	35.02	68.23	65.80
Plant	49.31	40.73	25.19	39.32	19.39	37.90	38.43
Sheep	72.57	69.43	37.15	56.43	31.09	56.12	56.80
Sofa	71.87	64.86	33.64	64.42	43.33	70.30	58.75
Train	86.61	74.98	67.05	79.62	47.12	70.88	80.70
TV	77.83	70.76	46.02	62.16	34.77	62.62	59.37

4.3 Analysis of Image High-Discrimination Feature Extraction Method

In the original YOLOX training process, a significant portion of high-frequency information is susceptible to loss when directly feeding images from the dataset into the deep learning network. To address this, the training regimen incorporates these data with a focus on discerning specific information along the horizontal, vertical, and diagonal axes in each image. This augmentation enhances the precision of target detection. As a result, we categorize the 16,551 images in the training dataset based on their directional information. The partitioning of each original image into distinct components includes smooth approximation segments, horizontal, vertical, and oblique components, as well as a salient portion with diagonal intricacies. The scale function and wavelet function of the Coiflet wavelet achieve this segmentation. Following the extraction of low-frequency and high-frequency features from the training dataset, the amplitude of the resulting feature image undergoes visual scrutiny.

4.4 Performance Analysis of Target Detection Methods

The experimental environment comprises tensorflow-gpu 1.14 and keras 2.1.5. Training is conducted on an NVIDIA 2080Ti 11 G GPU, with 100 batches of the VOC dataset and 250 batches of the COCO dataset. The maximum learning rate and batch size are set at 1e-3 and 8, respectively. Building upon the low illumination enhancement technique, the adverse impact of insufficient light on the target detection task is alleviated. The detection effectiveness of the YOLOX detection model is notably enhanced through the approach of extracting image characteristics via wavelet decomposition.

Fig. 8 depicts a comparison of actual detection between the original YOLOX model and the new target detection model. The detection frame locates the detected object in the image, and the numbers on the detection frame show the target object's proper detection scores. Fig. 9 depicts a polar pie chart illustrating the detection accuracy of each model, employing the enhanced low illumination enhancement method. The chart reflects the model's detection precision for the represented object along the axis. With values ranging from 0 at the center to 100 at the outermost point, the blue coverage area represents the overall accuracy of the model. It is evident from the chart that the detection results using this method surpass those of other models. The analysis was conducted on the COCO dataset, identifying original images, low illumination images, and low illumination enhanced images. The IoU, CLS, and OBJ loss functions in the COCO and VOC datasets are illustrated in Figs. 10 and 11, respectively.

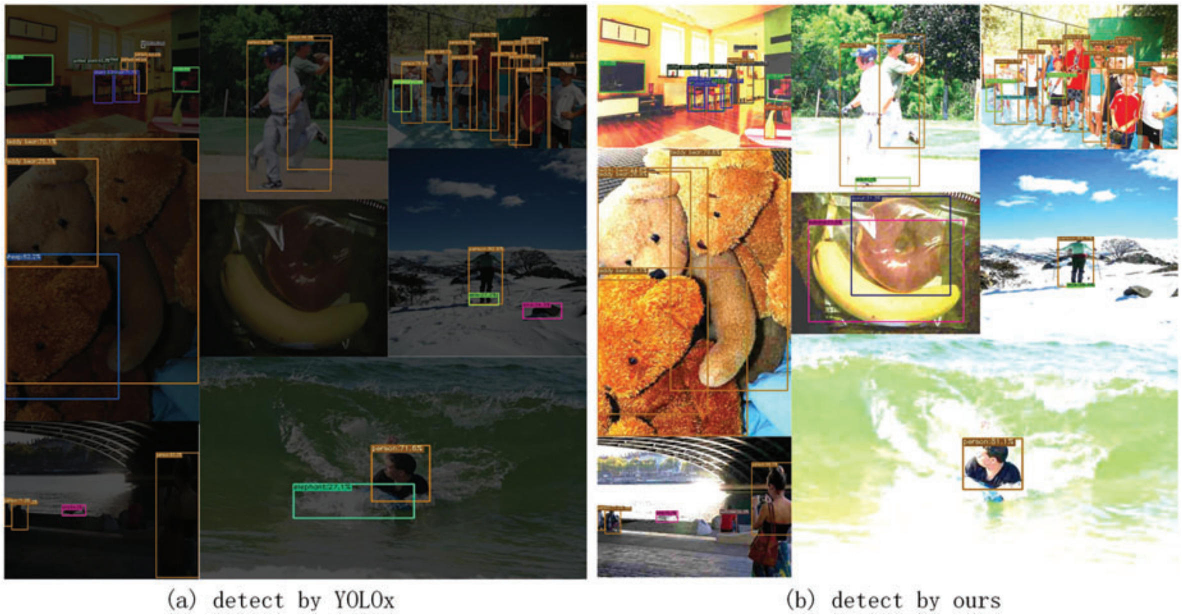


Figure 8: Comparison of target detection effects before and after low-light image enhancement

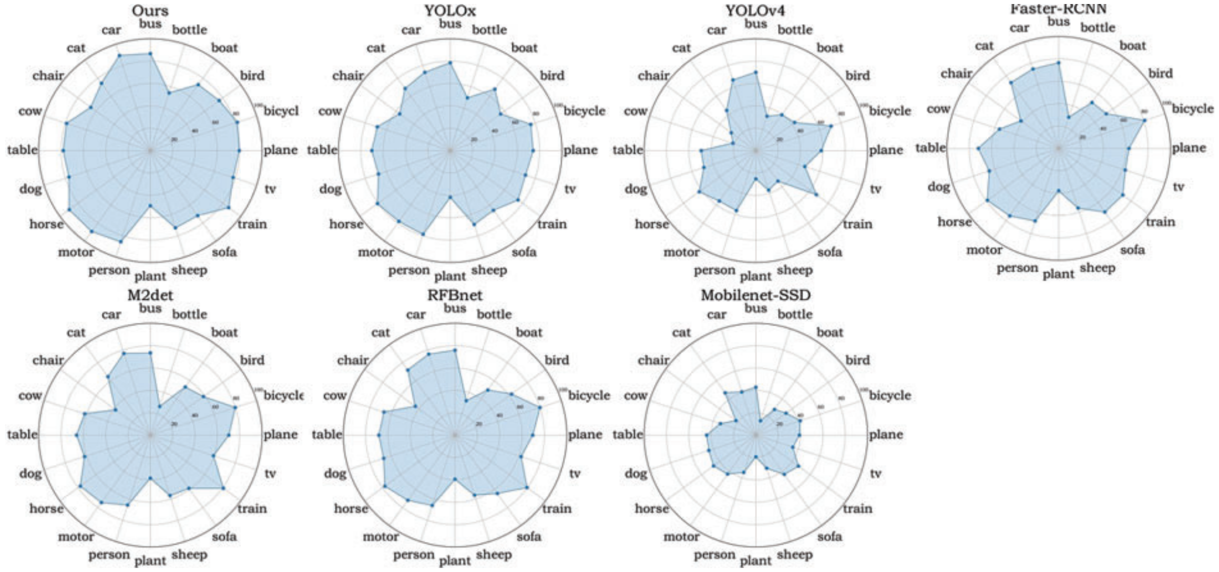


Figure 9: Polar axis pie chart of accuracy comparison between our model and other models on the VOC dataset

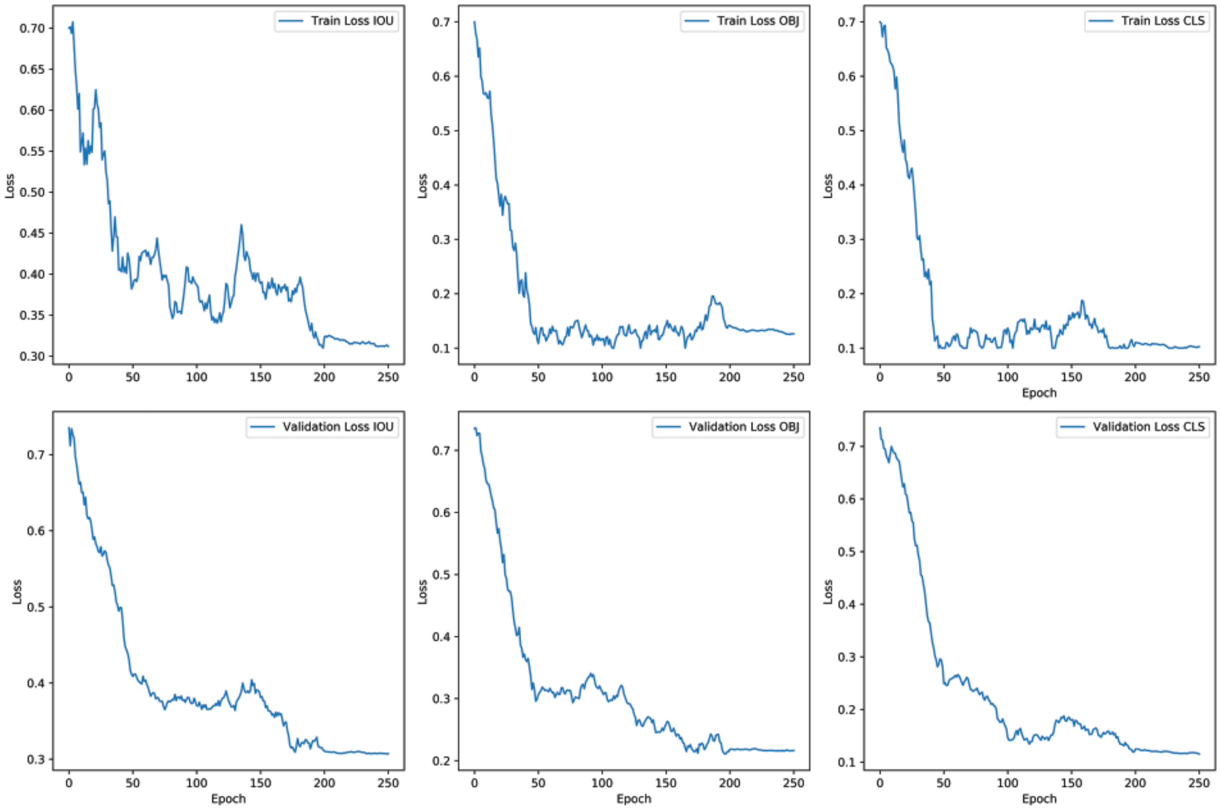


Figure 10: The IoU, CLS, and OBJ loss functions on the COCO dataset

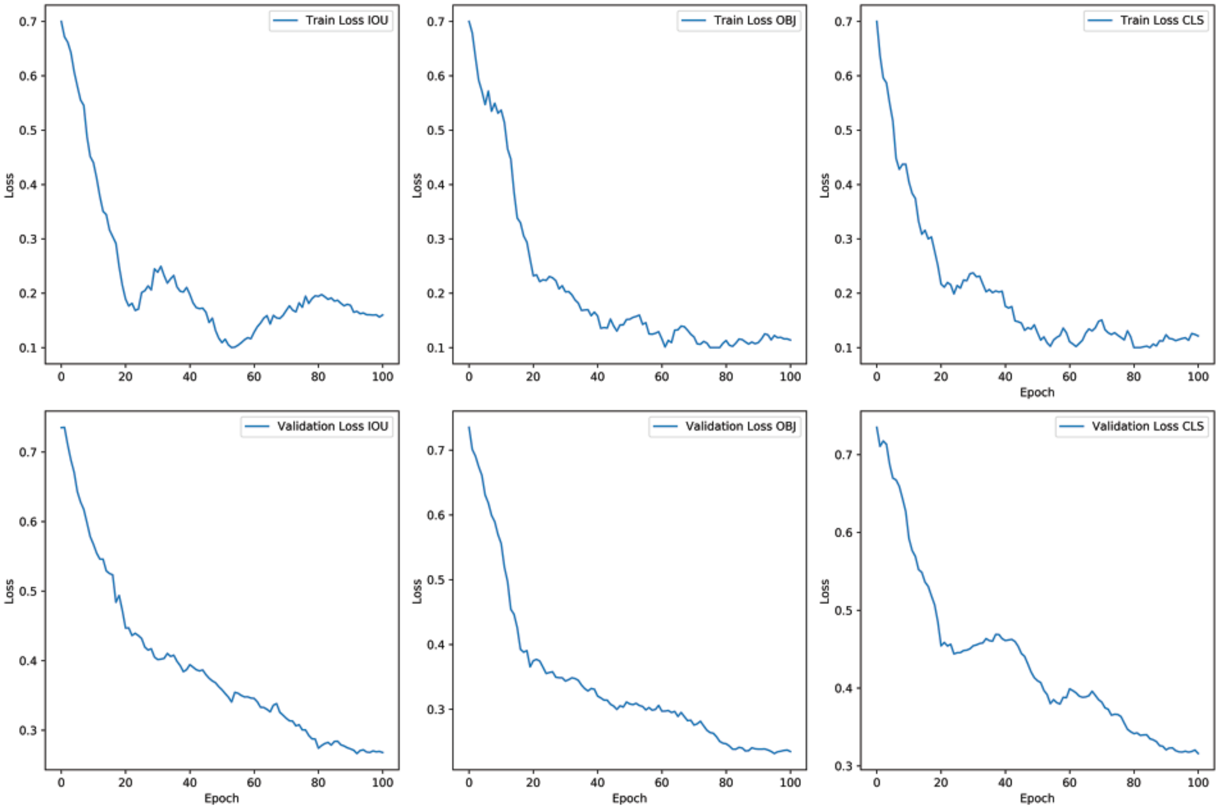


Figure 11: The IoU, CLS, and OBJ loss functions on the VOC dataset

Table 4 displays the AP of the proposed technique, YOLOX, YOLOv4 [63], RFBnet [64], Mobilenet-SSD [65], Faster R-CNN [59], and M2det [66] on the 20 low-light image types in the VOC test dataset. The proposed method exhibits robust color fidelity and improvements in image reproduction. Additionally, the mean Average Precision (mAP) of the target detection models is computed under three scenarios: Low-light environment, low-light image enhancement using the approach described in [53], and low-light image enhancement using our method. The comparison results are presented in Table 5.

Table 5: The accuracy of detecting low-light images using several models with different enhancing approaches

Methods	Proposed	YOLOX	YOLOv4	RFBnet	Mobilenet-SSD	Faster R-CNN	M2det
Dark image	70.96	67.41	54.96	64.06	53.91	65.19	65.09
Dong et al. [53]	72.15	68.47	47.75	64.09	36.52	62.71	62.20
This paper	76.86	73.31	69.83	73.79	63.26	62.75	72.32

The enhancement approach presented in [59] does not guarantee an improvement in target identification precision. In contrast, the proposed target detection model demonstrates a discernible improvement in precision. The data in the table illustrates the model’s detection precision for objects, with the radial axis indicating the precision values. The blue-shaded region, extending from the center (minimum value) to the maximum value of 100, represents the model's overall accuracy. This representation clearly emphasizes that the detection outcomes achieved through this method surpass those of alternative models.

Within the COCO dataset, assessments were conducted using the original image, low-illumination image, and the image post low-illumination enhancement. Tables 6 and 7 display the Average Precision (AP) and Average Recall (AR), respectively. The low-light image detection approach described in this paper provides superior detection effects on targets of various sizes.

Table 6: The analysis of the average precision (AP) of YOLOX in various lighting images

Image	AP@0.50:0.95	AP@0.50	AP@0.75	AP@S	AP@M	AP@L
Original image	0.504	0.690	0.547	0.325	0.561	0.669
Dark image	0.404	0.592	0.426	0.196	0.448	0.597
This paper	0.456	0.643	0.489	0.239	0.508	0.650

Table 7: The augmented reality (AR) of YOLOX in various illumination images

Image	AR@0.50:0.95	AR@0.50	AR@0.75	AR@S	AR@M	AR@L
Original image	0.379	0.614	0.653	0.468	0.712	0.825
Dark image	0.327	0.514	0.549	0.334	0.489	0.687
This paper	0.354	0.566	0.603	0.371	0.505	0.669

5 Conclusion

The presented paradigm for high-discrimination target detection addresses the exigency of localizing targets in conditions of profoundly attenuated luminosity through the strategic integration of advanced methodologies. Notably, this approach incorporates a feature extraction modality based on the wavelet transform, renowned for its discriminatory prowess, along with a sophisticated low-light enhancement protocol utilizing a dehazing algorithm. These modalities are seamlessly engrafted within the YOLOX model framework, resulting in a significant amplification of target detection efficacy. To mitigate the deleterious effects of exceedingly scant luminous flux, the refined low-light enhancement methodology employs an atmospheric dehazing model to ameliorate the inversed representation. The analogy between inversed low-light renditions and fog-occluded images justifiably underpins the deployment of the dehazing algorithm, thereby incrementally refining target identification acuity. Harnessing the adeptness of the Coiflet wavelet in convalescing intricate features across diverse spatial resolutions and high-frequency content, convolutional neural networks (CNNs) demonstrate a penchant for discerning details from multifarious azimuths. This proficiency empowers the model to enhance image fidelity and successfully recover occluded attributes, ultimately refining target recognition even in instances of indeterminacy. To warrant the availability of diverse input feature maps and concomitantly amalgamate nuanced features from a heterogeneous spectrum of image typologies, the proposed approach seamlessly incorporates CycleGAN. This integration augments the model's adaptability for discerning low-light image content, enabling it to accommodate various environmental conditions adeptly. The YOLOX target detection model is subsequently embedded within the approach to further hone detection precision. The symbiosis among high-discrimination feature extraction methodologies, the enhanced low-light enhancement approach, and the YOLOX model results in a resilient and precise stratagem expressly calibrated for the recognition of high-discrimination targets within contexts of exceedingly diminished luminance. Empirical assessments conducted on both the PASCAL VOC dataset and the MS COCO 2017 dataset substantiate the ascendancy of the proposed methodology over antecedent paradigms. These findings underscore its adaptability to real-world exigencies, affirming its stature as an efficacious panacea for target detection within exacting low-light environments.

Acknowledgement: The authors wish to express their appreciation to the reviewers for their helpful suggestions which greatly improved the presentation of this paper.

Funding Statement: This research is supported by National Sciences Foundation of China Grants (No. 61902158).

Author Contributions: Haijian Shao contributed to structural optimization, paper review, code inspection, and provided idea guidance. Suqin Lei and Chenxu Yan performed the programming implementation and drafted the paper. Xing Deng and Yunsong Qi also worked on structural optimization and participated in the paper review. All authors collectively reviewed the manuscript. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data that support the findings of this paper are openly available in [60–62]. The Python code for this paper has been open-sourced and is available on https://github.com/Harmenlv/YOLOX_DarkEnhance.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

1. Dutta A, Akata Z. Semantically tied paired cycle consistency for any-shot sketch-based image retrieval. Int J Comput Vis. 2020;128(10):2684–703. doi:10.1007/s11263-020-01350-x.
2. Wang H, Mo R, Chen Y, Lin W, Xu M, Liu B. Pedestrian and vehicle detection based on pruning YOLOv4 with cloud-edge collaboration. Comput Model Eng Sci. 2023;137(2):2025–47. doi:10.32604/cmes.2023.0269102.
3. Lee JG, Hwang J, Chi S, Seo J. Synthetic image dataset development for vision-based construction equipment detection. J Comput Civ Eng. 2022;36(5):04022020. doi:10.1061/(ASCE)CP.1943-5487.0001035.
4. Zhang Z, Cui Z, Xu C, Yan Y, Sebe N, Yang J. Pattern-affinitive propagation across depth, surface normal and semantic segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 19–20; Long Beach, California, USA. p. 4106–15.
5. Ding H, Jiang X, Shuai B, Liu AQ, Wang G. Semantic correlation promoted shape-variant context for segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 19–20; Long Beach, California, USA. p. 8885–94.
6. Dendorfer P, Osep A, Milan A, Schindler K, Cremers D, Reid I, et al. MOTChallenge: a benchmark for single-camera multiple target tracking. Int J Comput Vis. 2021;129(4):845–81. doi:10.1007/s11263-020-01393-0.

7. Lin D, Ji Y, Lischinski D, Cohen-Or D, Huang H. Multi-scale context intertwining for semantic segmentation. Proceedings of the European Conference on Computer Vision (ECCV), Sep. 2018; Munich, Germany. p. 603–19.
8. Çatal O, Verbelen T, Van de Maele T, Dhoedt B, Safron A. Robot navigation as hierarchical active inference. Neural Netw. 2021;142:192–204. doi:10.1016/j.neunet.2021.05.010.
9. Wang Y, Xiao B, Bouferguene A, Al-Hussein M, Li H. Content-based image retrieval for construction site images: leveraging deep learning–based object detection. J Comput Civ Eng. 2023;37(6):04023035. doi:10.1061/JCCEE5.CPENG-5473.
10. Attarmoghaddam N, Li KF. An area-efficient FPGA implementation of a real-time multi-class classifier for binary images. IEEE Trans Circuits Syst II: Express Briefs. 2022;69(4):2306–10. doi:10.1109/TCSII.2022.3148228.
11. Iepure B, Morales AW. A novel tracking algorithm using thermal and optical cameras fused with mmwave radar sensor data. IEEE Trans Consum Electron. 2021;67(4):372–82. doi:10.1109/TCE.2021.3128825.
12. Huang Z, Li W, Xia XG, Wu X, Cai Z, Tao R. A novel nonlocal-aware pyramid and multiscale multitask refinement detector for object detection in remote sensing images. IEEE Trans Geosci Remote Sens. 2021;60:1–20. doi:10.1109/TGRS.2021.3059450.
13. Mahdi SS, Nauwelaers N, Joris P, Bouritsas G, Gong S, Walsh S, et al. Matching 3D facial shape to demographic properties by geometric metric learning: a part-based approach. IEEE Trans Biom Behav Identity Sci. 2021;4(2):163–72. doi:10.1109/tbiom.2021.3092564.
14. Liu R, Shen J, Wang H, Chen C, Sc C, Asari VK. Enhanced 3D human pose estimation from videos by using attention-based neural network with dilated convolutions. Int J Comput Vis. 2021;129(5):1596–615. doi:10.48550/arXiv.2103.03170.
15. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016 Jun 26–Jul 1; Las Vegas, Nevada, USA. p. 779–88.
16. Ge Z, Liu S, Wang F, Li Z, Sun J. YOLOX: exceeding YOLO series in 2021. arXiv preprint arXiv:210708430 2021.
17. Zhu H, Zhang Y, Mu D, Bai L, Wu X, Zhuang H, et al. Research on improved YOLOX weed detection based on lightweight attention module. Crop Prot. 2024;177:106563. doi:10.1016/j.cropro.2023.106563.
18. Leng B, Jiang H, Wang B, Wang J, Luo G. Deep-Orga: an improved deep learning-based lightweight model for intestinal organoid detection. Comput Biol Med. 2024;169:107847. doi:10.1016/j.compbiomed.2023.107847.
19. Jing Z, Zhu L. Research on YOLOV7-tiny fruit detection algorithm improved by α-IoU. In: International Conference on Algorithm, Imaging Processing, and Machine Vision (AIPMV 2023), 2024; SPIE.
20. Zhan J, Luo Y, Guo C, Wu Y, Meng J, Liu J. YOLOPX: anchor-free multi-task learning network for panoptic driving perception. Pattern Recognit. 2024;148:110152. doi:10.1016/j.patcog.2023.110152.
21. Liu J, Zheng K, Liu X, Xu P, Zhou Y. SDSDet: a real-time object detector for small, dense, multi-scale remote sensing objects. Image Vis Comput. 2024;104898. doi:10.1016/j.imavis.2024.104898.
22. Huang J, Zhang T, Zhao S, Zhang L, Zhou Y. An underwater organism image dataset and a lightweight module designed for object detection networks. ACM Trans Multimed Comput Commun Appl. 2024;20(5):1–23. doi:10.1145/3640465.
23. Li J, Cheang CF, Liu S, Tang S, Li T, Cheng Q. Dynamic-TLD: a traffic light detector based on dynamic strategies. IEEE Sens J. 2024;24(5):6677–86. doi:10.1109/JSEN.2024.3352830.
24. He Y, Wu B, Mao J, Jiang W, Fu J, Hu S. An effective MID-based visual defect detection method for specular car body surface. J Manuf Syst. 2024;72:154–62. doi:10.1016/j.jmsy.2023.11.014.
25. Ng CC, Lin CT, Tan ZQ, Wang X, Kew JL, Chan CS, et al. When IC meets text: towards a rich annotated integrated circuit text dataset. Pattern Recognit. 2024;147:110124. doi:10.1016/j.patcog.2023.110124.

26. Yan P, Liu Y, Lyu L, Xu X, Song B, Wang F. AIOD-YOLO: an algorithm for object detection in low-altitude aerial images. J Electron Imaging. 2024;33(1):013023. doi:10.1117/1.JEI.33.1.013023.
27. Zhang Z, Rong J, Qi Z, Yang Y, Zheng X, Gao J, et al. A multi-species pest recognition and counting method based on a density map in the greenhouse. Comput Electron Agric. 2024;217:108554. doi:10.1016/j.compag.2023.108554.
28. Cao Y, Mu J, Cheng X, Liu F. Spacecraft-DS: a spacecraft dataset for key components detection and segmentation via hardware-in-the-loop capture. IEEE Sens J. 2024;24(4):5347–58. doi:10.1109/JSEN.2023.3347584.
29. Shen M, Huang R. Backdoor attacks with wavelet embedding: revealing and enhancing the insights of vulnerabilities in visual object detection models on transformers within digital twin systems. Adv Eng Inform. 2024;60:102355. doi:10.1016/j.aei.2024.102355.
30. Zhu S, Miao M, Wang Y. LRSNet: a high-efficiency lightweight model for object detection in remote sensing. J Appl Remote Sens. 2024;18(1):016502. doi:10.1117/1.JRS.18.016502.
31. Ullah M, Islam F, Bais A. Quantifying consistency of crop establishment using a lightweight U-Net deep learning architecture and image processing techniques. Comput Electron Agric. 2024;217:108617. doi:10.1016/j.compag.2024.108617.
32. Bao W, Huang C, Hu G, Su B, Yang X. Detection of Fusarium head blight in wheat using UAV remote sensing based on parallel channel space attention. Comput Electron Agric. 2024;217:108630. doi:10.1016/j.compag.2024.108630.
33. Yin Y, Cheng X, Shi F, Liu X, Huo H, Chen S. High-order spatial interactions enhanced lightweight model for optical remote sensing image-based small ship detection. IEEE Trans Geosci Remote Sens. 2024;62:1–16. doi:10.48550/arXiv.2304.03812.
34. Zhao Z, He C, Zhao G, Zhou J, Hao K. RA-YOLOX: re-parameterization align decoupled head and novel label assignment scheme based on YOLOX. Pattern Recognit. 2023;140:109579. doi:10.1016/j.patcog.2023.109579.
35. Wang X, He N, Hong C, Wang Q, Chen M. Improved YOLOX-X based UAV aerial photography object detection algorithm. Image Vis Comput. 2023;135:104697. doi:10.1016/j.imavis.2023.104697.
36. Zhu C, Hong H, Sun H, Wang G, Shen J, Yang Z. Real-time damage detection method for conveyor belts based on improved YOLOX. J Fail Anal Prev. 2023;23(4):1608–20. doi:10.3390/technologies11050114.
37. Luo B, Xiao J, Zhu G, Fang X, Wang J. Occluded insulator detection system based on YOLOX of multi-scale feature fusion. IEEE Trans Power Deliv. 2024;1–12. doi:10.1109/TPWRD.2024.3350162.
38. Zhou Y. A YOLO-NL object detector for real-time detection. Expert Syst Appl. 2024;238:122256. doi:10.1016/j.eswa.2023.122256.
39. Su P, Han H, Liu M, Yang T, Liu S. MOD-YOLO: rethinking the YOLO architecture at the level of feature information and applying it to crack detection. Expert Syst Appl. 2024;237:121346. doi:10.1016/j.eswa.2023.121346.
40. Rong S, Kong X, Gao R, Hu Z, Yang H. Grape cluster detection based on spatial-to-depth convolution and attention mechanism. Syst Sci Control Eng. 2024;12(1):2295949. doi:10.1080/21642583.2023.2295949.
41. Zhu X, Wang S, Su J, Liu F, Zeng L. High-speed and accurate cascade detection method for chip surface defects. IEEE Trans Instrum Meas. 2024;73:1–12. doi:10.3390/app11167657.
42. Wang A, Lang H, Chen Z, Peng Y, Ding S, Lu JJ. The two-step method of pavement pothole and raveling detection and segmentation based on deep learning. IEEE Trans Intell Transp Syst. 2024 doi:10.1109/TITS.2023.3340340.
43. Dong Y, Yang H, Liu S, Gao G, Li C. Optical remote sensing object detection based on background separation and small object compensation strategy. IEEE J Sel Top Appl Earth Obs Remote Sens. 2024 doi:10.1109/JSTARS.2024.3351140.

44. Du X, Qi Y, Zhu J, Li Y, Liu L. Enhanced lightweight deep network for efficient livestock detection in grazing areas. Int J Adv Robot Syst. 2024;21(1):17298806231218865. doi:10.1177/17298806231218865.
45. Zou H, Ye Z, Sun J, Chen J, Yang Q, Chai Y. Research on detection of transmission line corridor external force object containing random feature targets. Front Energy Res. 2024;12:1295830. doi:10.3389/fenrg.2024.1295830.
46. Ferrante GS, Vasconcelos Nakamura LH, Sampaio S, Filho GPR, Meneguette RI. Evaluating YOLO architectures for detecting road killed endangered Brazilian animals. Sci Rep. 2024;14(1):1353. doi:10.1038/s41598-024-52054-y.
47. Jiao Z, Huang K, Wang Q, Zhong Z, Cai Y. Real-time litchi detection in complex orchard environments: a portable, low-energy edge computing approach for enhanced automated harvesting. Artif Intell Agric. 2024;11:13–22. doi:10.1016/j.aiia.2023.12.002.
48. Liu Z, Lv H. YOLO_Bolt: a lightweight network model for bolt detection. Sci Rep. 2024;14(1):656. doi:10.1038/s41598-023-50527-0.
49. Yu J, Chen J, Wan H, Zhou Z, Cao Y, Huang Z, et al. SARGap: a full-link general decoupling automatic pruning algorithm for deep learning-based SAR target detectors. IEEE Trans Geosci Remote Sens. 2024;62:1–18. doi:10.1109/TGRS.2024.3350712.
50. Zhang W, Deng W, Cui Z, Liu J, Jiao L. Object knowledge distillation for joint detection and tracking in satellite videos. IEEE Trans Geosci Remote Sens. 2024;62:1–13. doi:10.3390/rs14102385.
51. Çınar A, Erkuş M, Tuncer T, Ayyıldız H, Tuncer SA. YOLOv5 based detector for eight different urine particles components on single board computer. Int J Imaging Syst Technol. 2024;34(1):e22968. doi:10.1016/j.patcog.2023.110152.
52. Priyanka SA, Wang YK, Huang SY. Low-light image enhancement by principal component analysis. IEEE Access. 2018;7:3082–92. doi:10.1109/ACCESS.2018.2887296.
53. Dong X, Wang G, Pang Y, Li W, Wen J, Meng W, et al. Fast efficient algorithm for enhancement of low lighting video. In: 2011 IEEE Int Conf Multimed Expo; 2011 Jul 15; Barcelona, Spain, IEEE. p. 1–6. doi:10.1109/ICME.2011.6012107.
54. Tian T, Cheng J, Wu D, Li Z. Lightweight underwater object detection based on image enhancement and multi-attention. Multimed Tools Appl. 2024;1–19. doi:10.1007/s11042-023-18008-8.
55. Xi LH, Hou JW, Ma GL, Hei YQ, Li WT. A multi-scale information fusion network based on pixelshuffle integrated with YOLO for aerial remote sensing object detection. IEEE Geosci Remote Sens Lett. 2024;21:1–5. doi:10.1109/LGRS.2024.3353304.
56. Chen W, Luo J, Zhang F, Tian Z. A review of object detection: datasets, performance evaluation, architecture, applications and current trends. Multimed Tools Appl. 2024;1–59. doi:10.1007/s11042-023-17949-4.
57. Almahairi A, Rajeshwar S, Sordoni A, Bachman P, Courville A. Augmented cyclegan: learning many-to-many mappings from unpaired data. In: Proceedings of the 35th International Conference on Machine Learning; 2018 Jul 10–15; Stockholm, Sweden: PMLR. p. 195–204.
58. Rombach R, Blattmann A, Lorenz D, Esser P, Ommer B. High-resolution image synthesis with latent diffusion models. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2022 Jun 19–24; New Orleans, Louisiana, USA. p. 10684–95.
59. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. Adv Neural Inf Process Syst. 2015;28:91–9.
60. Lin TY, Maire M, Belongie S, Hays J, Perona P, Ramanan D, et al. Microsoft coco: common objects in contex. In: Computer Vision–ECCV 2014: 13th Eur Conf, 2014; Zurich, Switzerland: Springer. p. 740–55.
61. Everingham M, van Gool L, Williams CKI, Winn J, Zisserman A. The PASCAL visual object classes challenge 2007 (VOC2007) results; 2007. Available from: http://www.pascal-network.org/challenges/VOC/voc2007/workshop/index.html. [Accessed 2024].

62. Everingham M, van Gool L, Williams CKI, Winn J, Zisserman A. The PASCAL visual object classes challenge 2012 (VOC2012) results; 2012. Available from: http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html. [Accessed 2024].

63. Bochkovskiy A, Wang CY. YOLO HYML. Optimal speed and accuracy of object detection. arXiv preprint arXiv:200410934. 2020. doi:10.48550/arXiv.2004.10934.

64. Liu S, Huang D, et al. Receptive field block net for accurate and fast object detection. In: Proceedings of the European Conference on Computer Vision (ECCV), Sep 2018; Munich, Germany. p. 385–400.

65. Howard AG, Zhu M, Chen B, Kalenichenko D, Wang W, Weyand T, et al. Mobilenets: efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:170404861. 2017. doi:10.48550/arXiv.1704.04861.

66. Zhao Q, Sheng T, Wang Y, Tang Z, Chen Y, Cai L, et al. M2det: a single-shot object detector based on multi-level feature pyramid network. In: Proceedings of the AAAI Conference on Artificial Intelligence, 2019; Honolulu, Hawaii, USA; vol. 33, no. 1, p. 9259–66.