PDIC: AN UNSUPERVISED ENHANCEMENT FRAMEWORK FOR DETECTION-AWARE NIGHT-TIME WILDLIFE IMAGING

A THESIS REPORT

SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE AWARD OF THE DEGREE
OF

MASTER OF TECHNOLOGY BY RESEARCH
IN
COMPUTER SCIENCE & ENGINEERING

Submitted by

PALAK HARINKHEDE (24/RCO/07)

Under the supervision of

Prof. Manoj Kumar

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
DELHI TECHNOLOGICAL UNIVERSITY
(Formerly Delhi College of Engineering)
Bawana Road, Delhi-110042

MAY, 2026

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

CERTIFICATE

I hereby certify that the thesis titled **"PDIC: An Unsupervised Enhancement Framework for Detection-Aware Night-time Wildlife Imaging"**, submitted by **Palak Harinkhede**, Roll No. **24/RCO/07**, Department of Computer Science & Engineering, Delhi Technological University, Delhi, in partial fulfillment of the requirements for the award of the degree of **Master of Technology by Research**, is a record of the research work carried out by the candidate under my supervision.

To the best of my knowledge, this work has not been submitted, in part or in full, for any Degree or Diploma to this University or elsewhere.

Place: Delhi **Prof. Manoj Kumar**

Date: 29/05/2026 **SUPERVISOR**

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

<u>CANDIDATE'S DECLARATION</u>

I, **Palak Harinkhede**, Roll No. **24/RCO/07**, student of **M.Tech by Research (Computer Science & Engineering)**, hereby declare that the thesis titled **"PDIC: An Unsupervised Enhancement Framework for Detection-Aware Night-time Wildlife Imaging"**, which is submitted by me to the Department of Computer Science & Engineering, Delhi Technological University, Delhi, in partial fulfillment of the requirements for the award of the degree of **Master of Technology by Research**, is original and not copied from any source without proper citation.

This work has not previously formed the basis for the award of any Degree, Diploma, Associateship, Fellowship, or other similar title or recognition.

Place: Delhi

Palak Harinkhede

Date: 29/05/2026 24/RCO/07

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

DELHI TECHNOLOGICAL UNIVERSITY

(Formerly Delhi College of Engineering)

Bawana Road, Delhi-110042

ACKNOWLEDGEMENT

I wish to express my sincerest gratitude to **Prof. Manoj Kumar**, of the Department of Computer Science & Engineering, Delhi Technological University, Delhi, for his continuous guidance and mentorship during this research work. His support, encouragement, and valuable suggestions helped me shape the objectives, methodology, and presentation of this thesis.

I am highly indebted to the panel faculty members for their guidance, constant supervision, constructive feedback, and motivation during the progress evaluations. Their suggestions helped improve the quality and direction of this work.

I also extend my sincere thanks to the faculty and staff of the Department of Computer Science & Engineering, DTU, for providing a supportive academic environment and the resources required to complete this thesis.

Finally, I am grateful to my family and friends for their unconditional support, patience, and encouragement throughout this journey.

Place: Delhi

Palak Harinkhede

Date: 29/05/2026 — 24/RCO/07

ABSTRACT

Wildlife detection in the night time is a critical computer vision problem that has applications in ecological monitoring, road safety surveillance, and conservation analysis. In real deployments, camera imaging systems are often found to be photon limited, which means that they are operating in an environment where the density of photons is low, and the noise level in the sensor as well as the object boundaries, motion blur, and infrared effects are also large.In real deployments, the density of photons, the noise level of the sensor, the boundaries of the objects, the motion blur, and the infrared effects are often large, causing the objects to be difficult to detect in a reliable manner. While the YOLO family of detectors can make fast inference in daylight images, they fail to perform well in dark, low-contrast night images with cluttered backgrounds and underexposed animals [1, 2, 37].

This thesis introduces a new unsupervised image-enhancement pipeline called **Progressive Dual-Branch Illumination-Contrast (PDIC)** specifically developed as an enhancement pre-processing step prior to a fixed YOLOv5 detector for night-time wildlife detection. In PDIC, the architecture of the detector is not changed, rather the visible evidence presented to the detector is enhanced. The framework is founded on the idea of decomposing image information based on the Retinex, and consists of two interconnected branches: an Illumination Compensation Branch for estimating and normalizing spatial illumination distribution, and a Contrast Enhancement Branch for enhancing the gradient of structures and mid-tone details and controlling the amplification of noise. The enhanced reflectance and illumination components are combined again in a learnable fusion module and progressive refinement allows the framework to process heavily underexposed images more than once [9, 8, 7, 4].

Training is completely unsupervised, that is, no paired low light and normal light images of wildlife are needed. The goal is to minimize smoothness of the illumination, gradient uniformity, exposure loss, color constancy loss, and total variation loss. These losses are chosen to maintain degradation of edges and contours relevant to the detectors and not just for human visual optimisation. The framework is applicable in practical camera trap use cases, where paired

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AFPN	Asymptotic Feature Pyramid Network
AHE	Adaptive Histogram Equalisation
AUC	Area Under the Curve
CLAHE	Contrast-Limited Adaptive Histogram Equalisation
CNN	Convolutional Neural Network
CSPNet	Cross Stage Partial Network
DCE	Deep Curve Estimation
DETR	DEtection TRansformer
FPN	Feature Pyramid Network
FPS	Frames Per Second
GAN	Generative Adversarial Network
GeLU	Gaussian Error Linear Unit
GIoU	Generalized Intersection over Union
GPU	Graphics Processing Unit
HE	Histogram Equalisation
IoU	Intersection over Union
IR	Infrared
mAP	Mean Average Precision
NMS	Non-Maximum Suppression
NTLNP	Northeast Tiger and Leopard National Park
PANet	Path Aggregation Network
PDIC	Progressive Dual-Branch Illumination-Contrast
R-CNN	Region-based Convolutional Neural Network
ReLU	Rectified Linear Unit
RGB	Red, Green, Blue
SID	See-in-the-Dark
TV	Total Variation
UAV	Unmanned Aerial Vehicle
YOLO	You Only Look Once
Zero-DCE	Zero-Reference Deep Curve Estimation

CHAPTER 1

INTRODUCTION

1.1 Background and Motivation

Automated visual sensing systems like camera traps, roadside surveillance cameras, unmanned platforms and edge-AI devices are increasingly being used to monitor wildlife. The systems are used for biodiversity assessment, habitat management, anti-poaching and AV collision prevention purposes. In recent years, conservation-computer-vision research has demonstrated that deep learning can save time for manual image-review and allow for large-scale ecological inferences under the right conditions of reliable data and deployed methodology [1, 2].

Although great progress has been made, night time wildlife detection is still significantly more difficult than daytime detection. A significant proportion of the valuable observations are made in the dark, because many target species are nocturnal or crepuscular. Under these conditions, images have low photon count, limited dynamic range, high sensor noise, infrared illumination artifacts, low texture and lower color information. These alterations can cause animals to be lost from sight and can make it harder to discern whether small or camouflaged animals are present or are vegetation, shadows, soil or roadside detritus [37, 38].

Modern object detectors, particularly one-stage YOLO-family models, are widely used due to their practicality, with respect to speed and accuracy [3, 4]. But these detectors are extremely local and require the presence of stable local gradients, textures and multi-scale feature activations. If they are suppressed during night-time imaging, they can result in missed detections or false positive detections in shadow areas, or in a poor bounding box localization. This thesis thesis' primary research question is thus not only how best to improve image appearance, but how best to enhance night-time images for improving downstream wildlife detection [5, 31].

1.2 Problem Statement

The issue discussed in this thesis is degradation of wildlife detection with YOLOv5 based on night-time low light imaging. The existing solutions normally take one of two routes. The first path adapts the detector architecture to match detection conditions at night to make it suitable for night time detection, which is similar to that of YOLO variants developed for night time detection [5]. These techniques can enhance accuracy, but would necessitate redirection of the detector, retraining of the detector, and careful application to a limited platform. The second approach is implemented by first applying classical image preprocessing techniques like CLAHE, gamma correction, or Retinex filtering, followed by a detection step [6]. They are simple but make use of static transformations and do not work well if there is a significant inter-species, inter-habitat, and/or camera-position variation in illumination.

For a practical wildlife-monitoring system, an approach must be adaptive, detector-compatible and trainable without paired low-light and well-lit ground truth. In camera-trap conditions paired data is uncommon since the same animal and scene can not be recorded simultaneously under both poor and ideal light conditions. This encourages an unsupervised enhancement approach to enhance the detection cues while maintaining the original YOLOv5 inference pipeline [7, 8].

1.3 Proposed Approach

In this thesis, the Progressive Dual-Branch Illumination-Contrast (PDIC) framework is developed. The purpose of PDIC is to serve as a preprocessing module that is put in front of YOLOv5. It will not change the YOLOv5 weights, anchors, losses, or detection heads. Rather it enhances the input image to provide more distinct object boundaries and more discriminative animal features to the fixed detector [4].

The framework is based on three principles. First, enhancement must be physically interpretable, and PDIC takes the Retinex-inspired perspective of decomposing an observed image into illumination and reflectance. Second, enhancement should be detection-aware, that is, not only aesthetically pleasing but also highlighting structural gradients and reducing noise.

Third, enhancement should be deployable, so training is based on no-reference losses that are used in unsupervised low-light enhancement algorithms like Zero-DCE and Retinex-based deep enhancement (DE) in [7, 8].

1.4 Research Objectives

The goals of this thesis are:

1. To investigate how image degradation in low-light conditions affects YOLO-family wildlife detection.

2. To develop a dual-branch illumination normalization and contrast/structure enhancement approach inspired by Retinex.

3. To train the enhancement pipeline without paired low-light/normal-light ground truth using no-reference losses.

4. To test the proposed PDIC–YOLOv5 pipeline on the NTLNP night-time wildlife dataset using Precision, Recall, F1-score, mAP@50, and mAP@50:95.

5. To compare the proposed system with classical preprocessing and night-time detection systems using the same downstream detector where possible.

1.5 Scope and Contributions

This work focuses on night-time wildlife detection based on RGB or infrared camera-trap images. The thesis is concerned not with the design of a new detector, but with a pre-processing and enhancement. Intentionally this scope is chosen as many field systems already use the known YOLO pipelines, and the switching of the detector may raise annotation, retraining and deployment costs [1, 18].

The most significant contributions have been:

1. A detector-agnostic PDIC enhancement framework for night-time wildlife imagery.

2. A dual-branch architecture to separately model illumination compensation and contrast enhancement.

3. A progressive refinement algorithm for severely underexposed images.
4. An unsupervised loss function combining illumination smoothness, gradient consistency, exposure, color constancy, and total variation regularization.
5. A structured experimental comparison of mAP@50 and Recall on the NTLNP dataset using a fixed YOLOv5 detector.

1.6 Thesis Organisation

This thesis is written as a standard empirical thesis, where the introduction motivates the problem, the literature survey is where the research gap is presented, the methodology is used to explain the proposed system, and the results chapter presents and interprets the experimental results. Chapter 2 introduces the theory of low-light imaging, object detection and Retinex enhancement. Related works are introduced in Chapter 3 to detect wild animals, detect night scenes using YOLO, transformer and hybrid approach, sensor modalities and low-light enhancement. Subsequent, the PDIC architecture and the unsupervised objective of training is described in Chapter 4. The experimental setup, results, ablation studies and discussion are presented in Chapter 5. The thesis is summarized in Chapter 6 and future work is discussed.

CHAPTER 2

THEORETICAL BACKGROUND

2.1 Low-Light Image Formation

Digital image is created by measuring the light reflected from a scene on a sensor. During night-time imaging, the amount of photons hitting the individual pixels is reduced, and the signal they provide is overwhelmed by noise. Photon shot noise, dark current, read noise, motion blur (long exposure), and compression artifacts (through camera pipelines) are the major degradations. While infrared camera traps are less light-dependent than visible-light cameras, they still have some drawbacks, including low contrast, a lack of uniformity in ambient light, specular reflections, and the lack of chromatic cues [11, 37, 38].

The degradations directly impact object detection. A detector doesn't think or reason about semantic concepts; it first looks for low-level features like edges, gradients, corners, textured regions, and color transitions. If the animal boundary cannot be seen (shaded), or if the noise makes an artificial boundary, the feature pyramid the detector has is not reliable. This means that one detector that works in daylight conditions, may not work in nighttime conditions [3, 4, 5].

Degradation in low light conditions is also non-uniform across space. With a wildlife shot, the animal could be lit by an infrared flash, car headlights, a nearby reflecting surface or the light of the moon, but the surrounding plants are mostly black. This creates a high dynamic range within the frame. Thus, there is no single correction for global brightness that can work for everyone: brightening up the back lighting may lead the eyes, fur highlights, or wet surface to become saturated. Any enhancement technique used to detect must not exaggerate background noise levels, but must maintain local contrast around the boundaries of the objects [10, 8].

The other important question is the difference between image quality and task utility. The enhanced image may not be a better detector input, if it is aesthetically pleasing. Over-smoothing helps to reduce very fine contours, but may help to improve the apparent image quality, and over-sharpening can generate false edges, leading to false positives. Therefore, improvement metrics should not only include image quality metrics but also the downstream metrics—for instance, Recall or mAP—because of the detection context [40, 16].

This failure can be explained by the signal to noise ratio (SNR). In a photon-limited scene, capture levels are not the only factor that affect the visibility of noise over the sensor measurement; the brightness will also increase the visibility of noise that was already captured in the measurement. While long exposure does help collect more photons, it also creates more motion blur when animals move across the frame, which is a problem in the field camera. High ISO gain will show weak signals, but will also increase read noise and colour artifacts. Enhancement at night time is thus not a straightforward scaling of brightness. A good enhancement technique should enhance the contrast within a local area without creating false edges, hallucinating texture and brightening existing bright infrared reflections [11, 7].

The camera-trap images create further challenges over regular night photography. The animal can be in a small part of the frame, at different distances or in part hidden behind plants. Numerous frames have empty background, and the same background can be seen in various illumination, weather and sensor-noise settings. Therefore, enhancement for wildlife detection should not only increase the visual appeal of an image but also retain the boundaries between objects and the mid-level texture in the image [17, 18, 19].

2.2 Retinex Theory and Illumination–Reflectance Decomposition

The retinex theory is an observed image which is considered to be the product of illumination and reflectance. The illumination component is the spatial variation of light, and the reflectance component is the intrinsic properties of the scene like boundaries or surface texture and albedo. A typical retinex solution recipe is:

$$I(x,y) = R(x,y) \cdot L(x,y), \tag{2.1}$$

Here, the image observed is $I(x,y)$, the reflectance is $R(x,y)$, and the illumination is $L(x,y)$ [9].

Classical Retinex methods estimate illumination by assuming the filter or smoothness. The Multiscale Retinex algorithm enhances the visual contrast of images by processing a number of spatial scales, and LIME is used to estimate illumination maps for low light enhancement [9, 10]. They are interpretable and efficient but can also accentuate noise or produce haloing to high-contrast edges. The deep Retinex algorithm methods, including RetinexNet, are trained with the decomposition function and enhancement function from data [8]. The PDIC framework expands on this reasoning but in relation to detection based wildlife enhancement.

There is interest in using retinex decomposition for imaging nocturnal wildlife; it decomposes two effects with different significance to detection. The illumination correction should recover from non-uniform illumination and reveal hidden regions while the reflectance preservation should maintain the texture, silhouette and local gradients of the animal. From the practical perspective, KinD also demonstrates that practical low-light enhancement can be arranged in terms of illumination adjustment, reflectance restoration, and degradation removal [13]. Later, RUAS re-formulated the Retinex enhancement as an unrolled optimization process with reference-free learning, and demonstrated that physics-inspired structure can be integrated with a light-weight neural design [14]. The developments agree with the PDIC design decision to provide separate illumination and contrast-oriented branches, instead of a single monolithic enhancement transform.

One of the most important applications of retinex enhancement is for night-time wildlife photography, where lighting is typically non-uniform. Only part of the scene may be illuminated by infrared flash, vehicle headlights, moonlight or reflected sunlight from vegetation. If one global adjustment is made to the entire image, the bright parts of the image can be over-exposed, and the dark parts of the animal can be under-exposed. There is a more flexible representation of illumination, reflectance decomposition: illumination could be smoothed and normalized, while reflectance could be kept with edges and textures. This separation also encourages the dual-branch architecture of PDIC, one for illumination compensation and the other for the contrast and structural detail.

But, Retinex decomposition is an ill-posed problem as there are many combinations of illumination and reflectance that will produce the same observed image. Thus, there are practical methods that impose priors. It is generally assumed that the illumination is smooth (in space), except in the vicinity of strong transitions, whereas the reflectance is expected to have sharper edges. When using these priors in low-light detection, it is important to use them carefully. Over-smoothing illumination can lead to loss of shadows that define the shape of animals and over-sharpening of reflectance can create an artifact of "fur-like" texture in noise. To resolve this, PDIC proposes to incorporate the total-variation regularization and gradient consistency along with illumination smoothness [10, 14].

2.3 Deep Low-Light Enhancement

If two images, one taken in low-light and the other in normal-light conditions, are available, supervised low-light enhancement methods can yield high quality results. Neural networks have been shown to be able to recover meaningful images from very dark RAW sensor data by learning to see in the dark [11]. Aligned pairs are not common however with the deployments of wildlife cameras, animal movement, weather, and lighting can't be replicated [11].

To overcome this, unsupervised and zero-reference enhancement techniques have been proposed that utilise no-reference losses. Given no paired data, Zero-DCE estimates image-specific enhancement curves and optimizes losses in spatial consistency, image smoothness, illumination smoothness, and exposure [7]. Additionally, EnlightenGAN proves that unpaired adversarial learning can be applied in the absence of aligned normal-light targets [12]. To improve low-light images, KinD breaks the task down into three phases: decomposition, restoration, and adjustment, thereby demonstrating the merit of decomposing light and reflectance separately [13]. On the contrary, more recent lightweight methods, like SCI, adopt self-calibrated illumination learning which is able to maintain the quality of illumination enhancement with less computation cost [15]. PDIC takes the same "real life" motivation and restates the enhancement task in terms of detector reliability. The goal is not only to enhance the image, but also to retrieve structural cues to boost the localization and classification of YOLOv5.

One of the theoretical differences that are important to note is that between image quality-based enhancement and task-based enhancement. While smooth, bright and visually appealing images are preferred by human observers, object detectors rely on repeatable local evidence. Too much denoising can obscure the faint contours of animals, too much contrast enhancement can produce high frequency artifacts, which look like small objects. Thus, detection-aware enhancement is designed to achieve the balance between the improvement of exposure and preservation of boundaries and gradients while avoiding hallucination. This is the reason that PDIC adopts unsupervised losses that promote natural exposure and smooth illumination without losing structural gradients crucial for bounding-box prediction [7, 40].

2.4 Object Detection Fundamentals

To detect an object, not only must the object be located but the object must also be classified. The two-stage detectors, like Faster R-CNN, make region proposals first and then classify every region proposal [20]. These methods are accurate but are not cheap enough to be used in the field. One-stage detectors directly predict the classes and bounding boxes from dense feature maps. For dense detection [21] has proposed a focal loss approach to deal with the problem of class imbalance, and YOLO-family detectors have become popular for their ability to perform detection in a single pass, enabling practical real time applications.

YOLO models divide detection into a backbone, neck, and head. The backbone is used to extract visual features, the neck combines multi-scale information, and the head is used for the prediction of the bounding boxes and objectness and class probabilities. YOLOv4 has added a number of training and architecture enhancements to achieve speed-accuracy balance [22] and YOLOv5 gained a foothold in the real world due to its PyTorch implementation, smaller model variants, and deployment ecosystem [4]. The review paper also notes that despite the transformer detectors being heavier models, YOLOv5 – YOLOv10 models are still the most competitive in terms of latency in real time applications involving wildlife or edge application, due to their more feasible latencies.

The critical components of this pipeline are the feature extraction and the objectness estimation modules, in low-light images. When the animal boundary is near the background intensity, there may be a failure in the generation of discriminative activations in the back bone. Weak or noisy features can be carried into different scales by the neck, and the objectness of the detection head might be low when an animal is present. Enhancement prior to detection can thus be regarded as a feature conditioning step. It alters the distribution of the input so that the fixed detector gets better gradients and clearer local contrast without any change in detector weights [4, 16].

2.5 Task-oriented Enhancement vs Human-oriented Enhancement

The majority of papers dealing with low light enhancement report PSNR, SSIM, NIQE or make a visual comparison. These are helpful for research in restoration, but not complete in characterizing the effectiveness of improved images to a detector. Object detectors are based on a hierarchy of features: edges and textures are detected in shallow layers, parts and shapes are detected in intermediate layers, and semantic object patterns are detected in deeper layers. A change in the statistics of these features may result in a change of the detection confidence and localization, even if the image is acceptable to human perception [40].

The studies carried out in low light detection have, therefore, recently called for task-oriented enhancement. ExDark was created to facilitate the study of low-light object detection and enhancement in images taken under various illumination conditions and annotated with bounding boxes [37]. The feature enhancement has not just been restricted to pixel enhancement, being explored for detection tasks, as well. FeatEnHancer reports consistent improvement in low light detection and related downstream tasks under dark conditions by improving hierarchical representations [40]. The results encourage the use of the gradient consistency and detector-aware evaluation by PDIC. It is better to improve information that proves to be useful for YOLOv5, rather than simply increasing brightness.

2.6 Detection Metrics

The metrics employed in the evaluation of the main task in this thesis are Precision, Recall, F1-score, mAP@50 and mAP@50:95. Precision indicates the proportion of correctly predicted detections and Recall reflects the proportion of ground truth objects that are correctly detected. Recall has a special significance in wildlife monitoring because the failure to recall animals could lead to inaccurate counts of animals and false alerts for road safety. The F1 score is a combined measure of Precision and Recall.

Mean Average Precision (mAP) is a summary of the performance of the detectors over the range of confidence thresholds. The Intersection over Union (IoU) threshold for mAP@50 is set at 0.50, and mAP@50:95 is the average of the performance measured at different IoU thresholds ranging from 0.50 to 0.95. Both are important since the night-time enhancement can not only help with object discovery, but can also influence the quality of localization [21].

CHAPTER 3

LITERATURE REVIEW

3.1 Literature Review Strategy

The literature review follows the structure of the two research papers provided for this thesis. The PDIC paper sets the primary technical direction, the proposed enhancement pipeline, the loss design and the NTLNP evaluation. The review paper gives a wider background of the field of nocturnal wildlife detection, CNN detectors, YOLO-family, transformer and hybrid, thermal and IR, datasets, evaluation protocols and edge deployment constraints. Taken together, the papers suggest that the central research question is not just "How do you make detectors work for difficult night-time wildlife imagery," but instead, "How do you enhance difficult night-time wildlife imagery in an adaptive way that is compatible with detectors?

3.2 CNN and Two-Stage Wildlife Detectors

Previous deep-learning-based wildlife detection systems relied on the concept of generic object detection frameworks but were adapted to camera-trap images. The performance of two-stage detectors is good localization models, but they require too much computation power to run on the road or edge camera in real-time application [20]. Dense one-stage detectors like RetinaNet tackle class imbalance using focal loss, and are applicable to camera-trap imagery where background frames are more common than rare species [21]. But these general architectures don't directly address the low light degradation problem.

In the past few years, there are some studies related to wildlife that are mainly concerned with automatic species identification, dense object detection, few-shot learning, and instance segmentation. The dense detection of wildlife objects has been studied in the context of enhanced YOLOv5 [23]; workflows for camera trap classification and identification have been examined in the context of large-scale ecological monitoring [24]; and few-shot detection approaches have been proposed for plateau or rare wildlife scenarios [25, 26]. Both works illustrate that data scarcity and class imbalance are long-term problems for wildlife vision.

A second important implication of the large camera-trap benchmarks was that wildlife recognition was not just a classification task, but a domain-generalization task. Snapshot Serengeti highlighted the scope and significance of citizen scientists annotating camera traps [17]. The Caltech Camera Traps work revealed that models trained on one location may have poor performance when evaluated on an unseen location, which was termed as recognition in "terra incognita" [18]. Norouzzadeh et al. also showed that deep neural networks can be used to automatically identify, count, and recognize animal behaviour at a large scale, but also rely on the quality and coverage of the animal training set [19]. These studies are relevant to PDIC since these enhancements need to be useful across a variety of camera positions and illumination conditions without overfitting to one particular situation and one camera.

3.3 YOLO-Based Nocturnal Detection

The YOLO family of detectors is an important player in the field of wildlife detection due to their modularity, speed and ability to use edge hardware. It highlights the most popular or active YOLO++ series of networks, such as YOLOv5, YOLOv7, YOLOv8, YOLOv9, and YOLOv10, which are used for real-time detection applications [27, 28, 29, 30]. To adapt their use to nocturnal wildlife detection, YOLO variants are typically modified with the incorporation of attention blocks, multi-scale fusion, lightweight modules, or infrared-specific branches.

YOLOv8-Night is a detector-based approach to night-time wildlife detection [5]. Other works show the capability of detecting animals during the night using deep learning algorithms and YOLO pipelines [31, 32]. It has been demonstrated here that architectural changes can enhance night time detection. But the redesign of detectors involves extra deployment costs and the need to retrain for all target environments. The strategy of PDIC is to keep YOLOv5 unchanged and enhance the input image.

3.4 Transformer and Hybrid Detectors

The transformer-based detectors like DETR and Deformable DETR have been proposed, which makes dense anchor-based prediction unnecessary by using attention mechanisms to guide set prediction [33, 34]. These models are more powerful for global context and complex scenes but typically higher data, memory and computation requirements compared to YOLO- family detectors according to the review paper. This may be a significant drawback in the camera-trap and roadside settings.

The hybrid methods integrate convolutional backbone, attention module, and multi-scale aggregation. For instance, ADLTFormer employs adaptive dual learning to detect cattle in low-light conditions [35]. The dual-branch design of the visible and infrared branch of IV-YOLO is presented in [36]. The methods show that the multi-branch design is beneficial for the low-light image or multi-modal detection. PDIC takes the concept of a dual-branch and uses it for enhancement instead of replacing the detector.

3.5 Low-light Enhancement for Detection

Simple and inexpensive classical methods like histogram equalization, CLAHE, gamma correction and Retinex filtering are attractive. But, with fixed enhancement, noise could be amplified, bright areas could be washed out and artificial contours could be created. These artifacts are detrimental for detection purposes as detectors could mistake them as object attributes.

Alternatives available are deep low-light enhancement methods. RetinexNet learns Retinex decomposition and enhancement [8]; Zero-DCE learns enhancement curves without paired supervision [7]; EnlightenGAN learns unpaired to avoid aligned low/normal-light image pairs [12]; and SID showcases the potential of learning-based recovery from extreme darkness, assuming RAW data is available [11]. KinD is also relevant due to its approach of decomposing, reflecting, and adjusting the illumination as separate sub-problems [13]. Later SCI focused on efficiency by using self-calibrated illumination learning, which is better suited for environments with limited resources [15]. In addition to these ideas, PDIC incorporates the detection-aware objective: it makes use of gradient consistency and progressive refinement explicitly, to ensure that the enhanced images include more robust animal boundaries but no hallucinated structures.

Why enhancement and detection should be viewed hand in hand is highlighted by recent low-light detection studies. Through the experiments, IAT is demonstrated to be effective in enhancing object detection and semantic segmentation results under challenging exposures conditions, as shown in [16]. Other enhancement-first pipelines report that under very low light conditions, normalization of the images before the detector can be performed before detection can improve the detector accuracy, but they also report that enhancement is not guaranteed to improve mAP if it creates artifacts or shifts the distribution of the detector input. This observation corroborates the design decision in PDIC: enhancement is limited by maintaining structural preservation of detectors, not only perceptually bright signals.

3.6 Datasets, Sensors, and Evaluation Protocols

A key challenge in nocturnal wildlife detection identified in the review paper is the limitations of the datasets. While public camera-trap datasets are valuable, they are often dominated by daytime photos, primarily devoted to classification, or poorly annotated. Low-light datasets are used for general low-light recognition research like ExDark, while visible–infrared datasets are used for paired multi-modal learning like LLVIP. Camera-trap projects like Snapshot Serengeti [17], Caltech Camera Traps [18], and iWildCam [2] offer vital information on ecological and domain-shift context. But there are limited night-time datasets available, and they're specific to wildlife.

This is a methodological issue because of this data gap. Supervised enhancement methods generally need paired images, one with low and one with normal lighting; however, paired images of wildlife are challenging due to the animals' lack of repetition in the controlled lighting. The same applies to supervised detectors for low-light images, which need enough labeled collections of night images of various species, distances, backgrounds, and types of sensors. Unsupervised enhancement is therefore appealing: it can be used directly on the available night images, and is compatible with any existing detector and can be placed in front of it. The NTLNP setting employed in this thesis very well meets this requirement as the objective is to enhance detection using real world infrared night imagery without requiring paired enhancement targets.

Thermal and/or infrared sensors can enhance night-time visibility, but they have other limitations such as cost, lower texture detail, domain shift and difficulty in separating species that look alike. Another challenge is the bias of the data sets in ecological monitoring, which may be caused by the location of fixed cameras in the study area, the seasonal migration of animals and class imbalance. A model can look good in common species or repeated poses and backgrounds, but bad in rare species or unusual poses and backgrounds. Hence, the review paper suggests that besides the mAP, Recall, robustness, sensor condition and edge feasibility should also be reported. This thesis goes along that line and reports detection metrics and highlights the compatibility with a fixed YOLOv5 pipeline.

3.7 Comparative Summary of Related Work

Table 3.1. Comparative summary of related methods and their relevance to this thesis.

Method Family	Representative Work	Strength	Limitation Addressed by PDIC
Classical Retinex / CLAHE	Multiscale Retinex, road-safety preprocessing [9, 6]	Lightweight and interpretable	Fixed transformations can amplify noise and fail under heterogeneous illumination.
Deep low-light enhancement	RetinexNet, Zero-DCE, EnlightenGAN, SCI [8, 7, 12, 15]	Adaptive enhancement with learnable objectives	Often optimized for perceptual quality rather than detector-specific structural cues.
YOLO night detectors	YOLOv8-Night and night-video systems [5, 31, 32]	Strong real-time detection focus	Requires detector modification or retraining; less modular for existing YOLOv5 pipelines.
Transformer / hybrid detectors	DETR, Deformable DETR, ADLTFormer [33, 34, 35]	Better global context modeling	Higher compute and data requirements restrict edge deployment.
Multimodal / task-aware detection	IV-YOLO, LLVIP, IAT [36, 38, 16]	Robustness in darkness	Fusion methods require additional sensors, while task-aware enhancers may require additional task-specific validation; PDIC targets single-image enhancement compatibility.

3.8 Research Gap

The literature reviewed shows four gaps. First of all, the classical pre-processing is not adaptive enough for complex nocturnal scenes. Secondly, night models specific to the detector reduce the compatibility with current deployed systems. Third, edge deployment is computationally expensive using transformer and hybrid detectors. Fourth, there is limited availability of paired low-light wildlife training data. To tackle these limitations, PDIC introduces an unsupervised, detector-agnostic, and Retinex-inspired enhancement module to enhance the quality of the input of the detector YOLOv5 without modifying the detector itself.

CHAPTER 4

METHODOLOGY

The PDIC framework proposed in this chapter is presented. The chapter is based on the technical approach of the PDIC research paper and elaborates on it to create a thesis level methodology. The key design choice is to enhance the effectiveness of night-time wildlife images before detection and not change the downstream YOLOv5 detector [4, 7, 8].

4.1 System Overview

Suppose that $I \in \mathbb{R}^{H \times W \times 3}$ is a low-light image of a wildlife scene. The goal of PDIC is to obtain a better image $I_{out} \in \mathbb{R}^{H \times W \times 3}$ for object detection [4]. The system comprises of four major elements:

1. an Illumination Compensation Branch $\mathscr{F}_I$ that estimates the spatial illumination map $\hat{L}$; and
2. an Enhancement Branch $\mathscr{F}_E$ to make the reflectance details and gradients more prominent;
3. an adaptive fusion module for reconstructing an enhanced image; and
4. A fixed YOLOv5 detector which takes the enhanced image for inference.

It is detector-agnostic, as the enhancement module is trained independently, and the detector does not have to be modified in its architecture. Figure 4.1 shows the overall data flow.

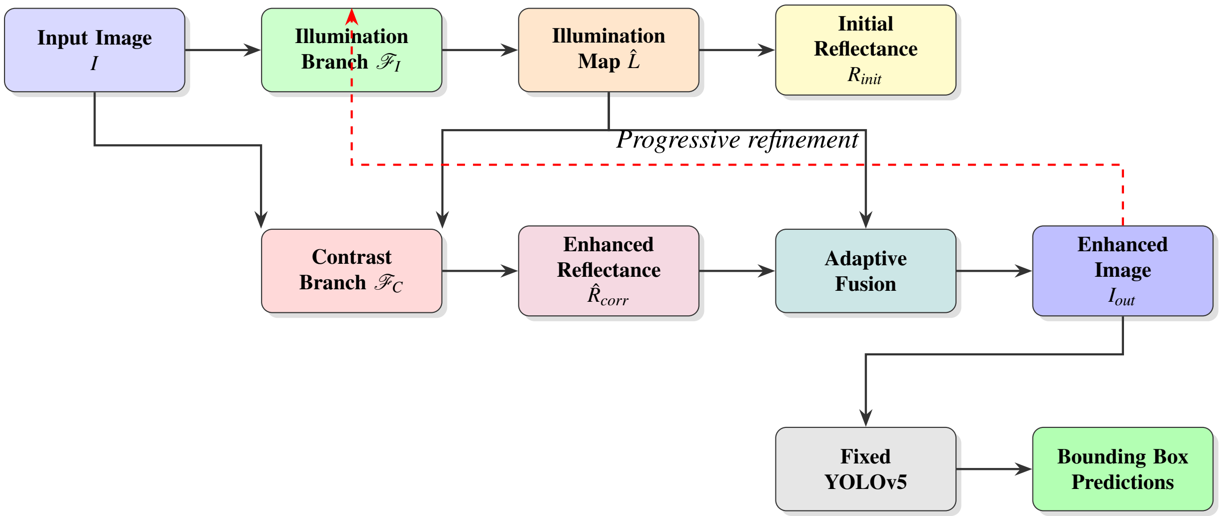


Figure 4.1. Architecture of the Progressive Dual-Branch Illumination-Contrast (PDIC) pipeline. The enhanced image is passed to an unchanged YOLOv5 detector.

4.2 Illumination Compensation Branch

The Illumination Compensation Branch estimates the illumination map:

$$\hat{L} = \mathscr{F}_I(I). \tag{4.1}$$

A convolutional encoder-decoder is implemented to implement the branch. The encoder gradually decreases the spatial resolution and increases the receptive field so that the network can model the large-scale brightness variation. The decoder upsamples and convolves to restore the resolution. Skip connections help to retain high frequency information which might be lost during downsampling.

To get an initial reflectance estimate:

$$R_{init}(x,y) = \frac{I(x,y)}{\max(\hat{L}(x,y),\varepsilon)}, \tag{4.2}$$

The value of $\varepsilon = 10^{-3}$ is chosen to avoid numerical instability in very dark pixels. This operation is based on retinex reasoning, which is that the removal of illumination should expose reflectance structures that are more stable for detection [9, 8].

4.3 Contrast Enhancement Branch

Normalization in illumination can result in a flat image and could magnify the noise in the sensor. The Contrast Enhancement Branch updates R_{init} as:

$$\hat{R}_{corr} = \mathscr{F}_C(R_{init}). \tag{4.3}$$

The branch maintains the identity information while learning the corrective transformations with residual blocks. The multi-scale convolution head processes are convolution head with filters of 1×1, 3×3, and 5×5. This design is able to capture fine texture, medium scale edges and larger animal contours [21, 4]. This is followed by a concatenation and fusion, where the output is concatenated and fused with a 1×1 convolution.

The design is designed for detection. Instead of uniformly increasing the brightness, the branch focus on the structures of the mid-tone and object boundary that is important for the feature extraction of YOLOv5. Meanwhile, regularization losses, which minimize high-frequency noise and false edges, are also produced simultaneously.Regularization losses, which minimize high frequency noise and artificial edges, are also produced simultaneously.

4.4 Adaptive Fusion and Progressive Refinement

A reconstruction of the last enhanced image is obtained by the adaptive blending map $\alpha(x,y) \in [0,1]$:

$$I_{out} = \alpha \cdot \hat{R}_{corr} \cdot \hat{L} + (1-\alpha) \cdot I. \tag{4.4}$$

A simple convolution with a sigmoid activation is used to predict the blending map. Darker areas will benefit from more enhancement, and brighter areas will be closer to what is originally in them.

A few times, the images may need more than one enhancement pass, depending on the amount of underexposure. PDIC therefore adopts the progressive refinement method, where the output of one pass is used as input to the next pass. The steps involved in the procedure are summarized in Algorithm 1.

Algorithm 1: PDIC Progressive Iterative Refinement

Input: Low-light image I_0, number of iterations K

Output: Enhanced image I_K

for $k \leftarrow 0$ **to** $K-1$ **do**

  $\hat{L}_{k+1} \leftarrow \mathscr{F}_I(I_k)$;

  $R_{init} \leftarrow I_k / \max(\hat{L}_{k+1}, \varepsilon)$;

  $\hat{R}_{corr} \leftarrow \mathscr{F}_C(R_{init})$;

  $\alpha \leftarrow \sigma(\mathrm{Conv}([\hat{R}_{corr}, I_k]))$;

  $I_{k+1} \leftarrow \alpha \cdot \hat{R}_{corr} \cdot \hat{L}_{k+1} + (1-\alpha) \cdot I_k$;

end

return I_K ;

The PDIC paper reports that $K = 3$ provides a strong balance between enhancement quality and computational overhead [7, 15].

4.5 Unsupervised Optimization Objective

The PDIC pipeline is trained without paired target images. The total loss is a weighted combination of five no-reference losses [7]:

$$\mathscr{L}_{total} = \lambda_s \mathscr{L}_{smooth} + \lambda_g \mathscr{L}_{grad} + \lambda_e \mathscr{L}_{exp} + \lambda_c \mathscr{L}_{color} + \lambda_t \mathscr{L}_{tv}. \tag{4.5}$$

4.5.1 Illumination Smoothness Loss

The illumination map should vary smoothly except near genuine structural boundaries. The loss is:

$$\mathscr{L}_{smooth} = \sum_{x,y} \left|\nabla \hat{L}(x,y)\right| \cdot \exp\left(-\left|\nabla I(x,y)\right|\right). \tag{4.6}$$

This discourages noisy illumination maps while allowing changes near object edges.

4.5.2 Gradient Consistency Loss

Gradient consistency preserves existing structures and discourages hallucinated edges:

$$\mathscr{L}_{grad} = \sum_{x,y} \left| \nabla \hat{R}_{corr}(x,y) - \nabla R_{init}(x,y) \right|. \tag{4.7}$$

This loss is especially important because YOLO-like detectors are sensitive to artificial contours.

4.5.3 Exposure Control Loss

Exposure control guides the enhanced image toward a target local brightness E:

$$\mathscr{L}_{exp} = \frac{1}{N} \sum_{i=1}^{N} \left| \mu_i(I_{out}) - E \right|, \tag{4.8}$$

where μ_i is the mean intensity of the i-th non-overlapping local region. In this thesis, $E = 0.6$ is used following the PDIC experimental setting.

4.5.4 Color Constancy Loss

Color constancy reduces channel imbalance using a Gray-World assumption:

$$\mathscr{L}_{color} = \sum_{(p,q)\in\{(R,G),(R,B),(G,B)\}} \left(\mu_p(I_{out}) - \mu_q(I_{out}) \right)^2. \tag{4.9}$$

4.5.5 Total Variation Loss

Total variation suppresses high-frequency noise:

$$\mathscr{L}_{tv} = \sum_{x,y} \left(|I_{out}(x+1,y) - I_{out}(x,y)| + |I_{out}(x,y+1) - I_{out}(x,y)| \right). \tag{4.10}$$

The final weights used in the PDIC experiment are:

$$\mathscr{L}_{total} = 1.0\mathscr{L}_{smooth} + 1.0\mathscr{L}_{grad} + 10.0\mathscr{L}_{exp} + 5.0\mathscr{L}_{color} + 0.5\mathscr{L}_{tv}. \tag{4.11}$$

These weights prioritize exposure correction and color balance while retaining structural consistency.

4.6 YOLOv5 Integration

The enhanced I_{out} is directly fed into YOLOv5. The detector is not part of the PDIC module and thus is not retrained. This design allows the effect of enhancement to be separated from the other effects and makes a comparison with classical preprocessing credible. It also enables practical deployment; the existing YOLOv5-based wildlife system can be fed by PDIC as a front-end module without any change of their detection codebase.

CHAPTER 5

EXPERIMENTAL RESULTS AND DISCUSSION

5.1 Experimental Setup

Experimental design is based on the research paper in PDIC. It is evaluated with data from the night-time wildlife dataset, consisting of 10,344 infrared images of 17 animal categories from NTLNP. The dataset is suitable for this study because it includes real scenes captured at night by camera traps that have low contrast, uneven illumination, occlusion and cluttered backgrounds.

The downstream detector used is YOLOv5, which is fixed during the evaluation. Images are fed to YOLOv5 after being preprocessed by the selected preprocessing model. This protocol will guarantee that any observed differences in performance are due to the enhancement technique and not to the changes in the detector architecture.

5.1.1 Evaluation Metrics

The following metrics will be used:

- **Precision:** percentage of the actual detections that are identified.
- **Recall:** Percentage of the actual animals correctly identified.
- **F1-score:** harmonic mean of Precision and Recall. **mAP@50:** mean Average Precision at IoU threshold 0.50. **mAP@50:**: mean Average Precision for IoU values from 0.50 to 0.95.

The effect of the recall is emphasized as missing animals are especially damaging for ecological monitoring and for road safety applications.

5.2 Quantitative Comparison

Table 5.1 summarizes the main comparison reported in the PDIC paper. The proposed PDIC–YOLOv5 pipeline achieves the best mAP@50 and Recall among the listed methods.

Table 5.1. Main quantitative comparison on the NTLNP night-time wildlife dataset.

Method	mAP@50	Recall
CLAHE + Retinex + YOLOv5 [6]	0.800	0.773
YOLOv8-Night [5]	0.855	–
PDIC + YOLOv5	**0.883**	**0.868**

Compared with CLAHE+Retinex+YOLOv5, PDIC improves mAP@50 by 0.083 and Recall by 0.095. The Recall improvement is especially important because it indicates that more animals are recovered after adaptive enhancement. PDIC also exceeds the reported mAP@50 of YOLOv8-Night while preserving the simpler deployment advantage of using a fixed YOLOv5 detector.

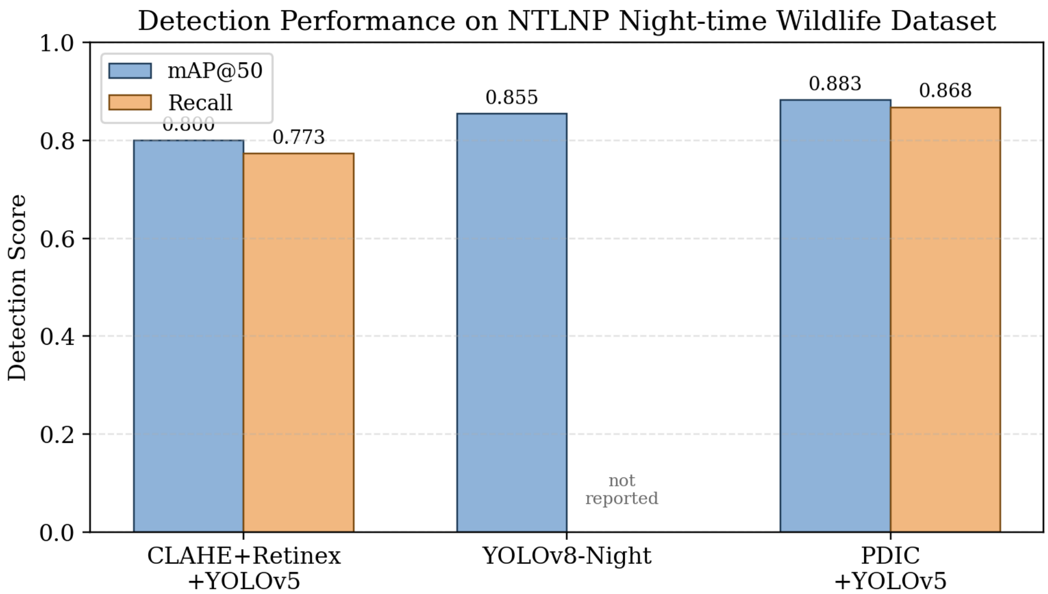


Figure 5.1. Visual comparison of detection performance on the NTLNP dataset. The figure highlights that PDIC improves both localization-oriented accuracy (mAP@50) and animal recovery (Recall) when compared with classical preprocessing.

5.3 Metric-Level Interpretation

The improvement on mAP@50 suggests that PDIC improves the ability of YOLOv5 to detect boxes of the objects with localization overlap. The gain in Recall suggests that the enhancement module is able to bring out animals that were previously not visible by the detector due to their darkness or lack of contrast. The method can thus not only be used to enhance the confidence score but also to increase the visibility of the objects.

The result confirms the thesis hypothesis - in the case of the night-time wildlife shot, an enhancement stage based on the detector can be more feasible than the redesign of the detector. Furthermore, since PDIC does not alter YOLOv5, it can be used with low engineering efforts in existing camera-trap or roadside detection pipelines.

5.4 Ablation Study

The ablation study investigates the effect of progressive refinement and the role of individual loss components. Table 5.2 shows a representative loss ablation that has been designed according to the design logic presented in the PDIC paper.

Table 5.2. Ablation of PDIC loss components.

Configuration	mAP@50	Interpretation
Full PDIC objective	**0.883**	Best balance of exposure, structure, and noise control
Without gradient consistency	0.861	Weaker object boundaries and more unstable detector features
Without exposure control	0.847	Insufficient brightness recovery in dark regions
Without total variation	0.866	Increased noise and false edge responses

Figure 5.2 plots the loss of mAP@50 when big loss terms are omitted. The greatest difference is seen when exposure control is turned off, indicating that the recovery of brightness is crucial for nighttime wildlife detection. Gradient consistency and total variation also play their part in maintaining consistent and reliable object edges and minimizing noise-induced false contours.

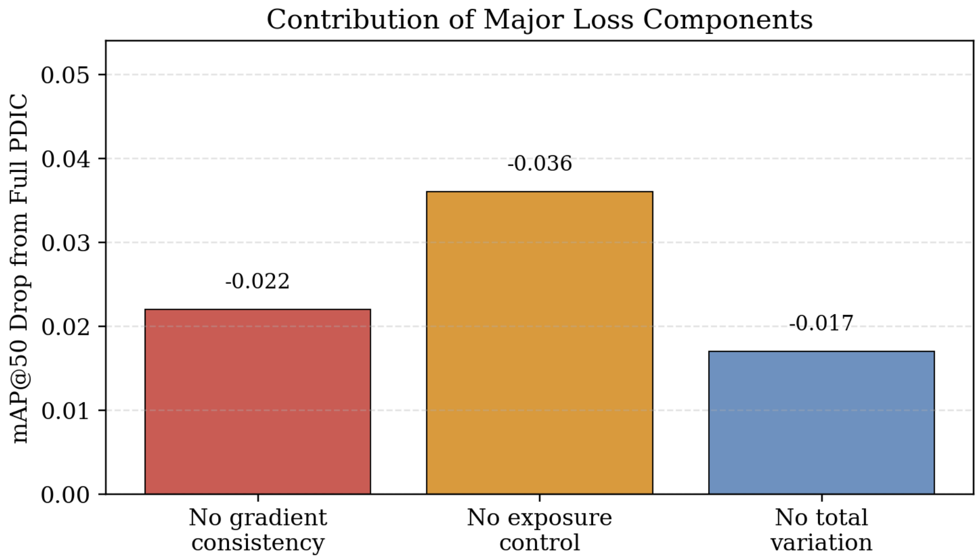


Figure 5.2. mAP@50 reduction caused by removing major PDIC loss components. Larger drops indicate stronger contribution to the full objective.

The relevance of the gradient consistency loss is that it is the loss of structures relevant for the detectors. To recover underexposed areas, exposure control is needed, and to avoid the enhancement module from creating false object-like patterns from sensor noise, total variation is needed.

5.5 Effect of Progressive Refinement

Progressive refinement is designed for images that cannot be sufficiently enhanced in a single pass. Table 5.3 summarizes the effect of the number of refinement iterations.

Table 5.3. Effect of progressive refinement iterations.

Iterations (K)	mAP@50	Observation
1	0.842	Enhancement is incomplete in severe darkness
2	0.874	Strong improvement with moderate cost
3	**0.883**	Best reported balance
4	0.884	Marginal gain with additional latency

The trend shows that refinement helps until the image reaches a stable exposure and contrast level. Beyond K=3, gains become marginal, while latency increases. Therefore, K=3 is selected as the practical operating point.

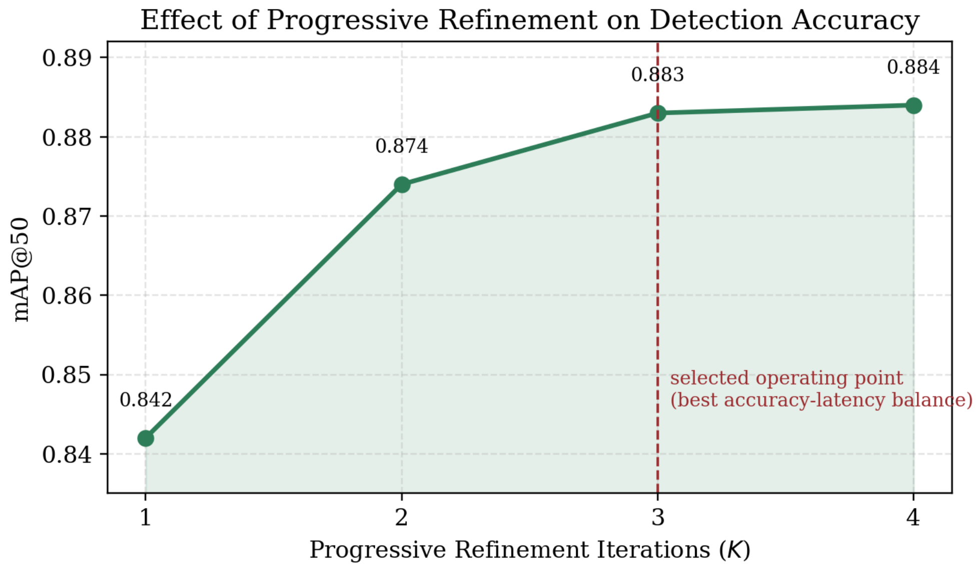


Figure 5.3. Effect of progressive refinement iterations on mAP@50. Accuracy improves rapidly from one to three passes and then saturates, supporting K=3 as the practical operating point.

5.6 Discussion

Evidence demonstrates the effective targeting of PDIC regarding actual failure modes for operating the mechanism under low light/at night. Classical enhancement enhances object brightness; however, classical enhancement by itself does not provide the means for adapting to differing levels of illumination across the image area being evaluated. Restructuring the detector circuit enhances accuracy but reduces modularity. PDIC falls into the middle ground in that PDIC can be adapted as a neural network model, yet it can function as a pre-processing module.

This method is appropriate for use in camera trap networks and also roadside locations where existing YOLOv5 systems operate because it can have front end enhanced models without modifications to the original annotation format, detector training scripts (assumed to be the same across all installations), or interfacing for deployment. This characteristic may be important in applications of conservation and motor vehicle safety since sustaining performance and maintainability will have as much impact o output accuracy level.

One limitation of PDIC lies in that it is dependent upon visual signals that exist in the image to produce new improved images of the inputted data. If there is no animal detected in the sensor output, enhancements will be unable to provide true information. In addition, the number of enhancement passes that are performed will increase the computational load and will therefore be a consideration for future work directed at reducing both compression, and quantization along with focusing on deployment methods at the edge level only.

CHAPTER 6

CONCLUSION AND FUTURE WORK

6.1 Conclusion

Progressive Dual-Branch Illumination–Contrast Enhancement (PDIC) Framework for Night-Time Wildlife Detection is introduced in this thesis. One main motivation behind carrying out this research was the decline in accuracy and real-time operation of the YOLOv5 model and other comparable models in low illumination and contrast scenarios with high sensor noise [5, 37].

The proposed method is based on enhancing images rather than adjusting the model architecture. There are four aspects included in the proposed framework. Those are 1) illumination compensation branch; 2) contrast enhancement branch; 3) adaptive fusion and 4) progressive refinement. The training procedure involves unsupervised no-reference losses; hence, PDIC can be employed for improving low-light images of wildlife as obtaining low-light and normal images simultaneously is highly improbable [7, 8].

The results from utilizing the proposed method PDIC+YOLOv5 on the NTLNP dataset showed mAP@50 score of 0.883 and Recall of 0.868 outperforming the CLAHE + Retinex + YOLOv5 baseline and even YOLOv8-Night's published mAP@50 while still being connected to the detector in the same manner. Hence, the conclusion that emerges from this thesis work supports the utilization of the detection-aware enhancement approach to monitor wildlife at night [5, 6].

6.2 Future Scope

The following are potential future directions:

1. **Optimization for Edge Deployment:** PDIC needs to be compressed for edge deployment in devices like the Raspberry Pi, NVIDIA Jetson Nano, and low-power camera-trap devices using quantization, pruning, and knowledge distillation.

2. **Joint End-to-End Detectors Aware Training:** While modularity is a benefit, another direction is controlled end-to-end training in which detector loss terms help enhance the image but prevent overfitting to any specific detector.

3. **Fusion of Thermal and Infrared Spectra:** Another version of multispectral PDIC can take into account visible spectra, as well as thermal and infrared spectra for images in which visual contrast may be limited.

4. **Generalization across Sites:** The algorithm must be evaluated on various sites having diverse habitats, seasons, camera models, and weather conditions to analyze generalizability.

5. **Failure Case Analysis:** Future evaluations must take missed detections, false positives, species level misclassifications, and latency analysis in addition to mAP metrics.

On the whole, PDIC proves that an appropriate unsupervised enhancement algorithm can help detect wildlife at night without compromising real-time detection capability.

LIST OF PUBLICATIONS

The following research papers were prepared from this thesis work.

- **[P1]** **Palak Harinkhede** and Manoj Kumar, "Progressive Dual-Branch Illumination-Contrast Pipeline for Night-time Wildlife Detection Using YOLOv5." **Accepted**, **7th International Conference on Data Analytics & Management (ICDAM-2026)**. Paper ID: 1598.

- **[P2]** **Palak Harinkhede** and Manoj Kumar, "Night-time Wildlife Detection Using Computer Vision: A Review of Deep Learning, Thermal Imaging, and Edge-AI Approaches." **Accepted**, **7th International Conference on Data Analytics & Management (ICDAM-2026)**. Paper ID: 1623.

REFERENCES

[1] D. Tuia, B. Kellenberger, S. Beery, et al., “Perspectives in machine learning for wildlife conservation,” *Nat. Commun.*, vol. 13, Art. no. 792, 2022.

[2] S. Beery, D. Morris, S. Yang, et al., “The iWildCam 2024 Challenge Dataset and Benchmark,” in *Proc. IEEE/CVF CVPR Workshops*, 2024.

[3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, “You Only Look Once: Unified, Real-Time Object Detection,” in *Proc. IEEE CVPR*, 2016, pp. 779–788.

[4] G. Jocher et al., “YOLOv5 by Ultralytics,” 2020.

[5] T. Wang, S. Ren, and H. Zhang, “Nighttime wildlife object detection based on YOLOv8-night,” *Electron. Lett.*, vol. 60, no. 15, pp. 598–601, 2024.

[6] R. Parkavi et al., “Enhancing Road Safety: Detection of Animals on Highways During Night,” *Int. J. Eng. Res. Technol.*, vol. 12, no. 3, pp. 1–8, 2025.

[7] C. Guo et al., “Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement,” in *Proc. IEEE/CVF CVPR*, 2020, pp. 1780–1789.

[8] C. Wei, W. Wang, W. Yang, and J. Liu, “Deep Retinex Network for Low-Light Enhancement,” in *Proc. BMVC*, 2018, pp. 1–13.

[9] D. J. Jobson, Z. Rahman, and G. A. Woodell, “A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes,” *IEEE Trans. Image Process.*, vol. 6, no. 7, pp. 965–976, 1997.

[10] X. Guo, Y. Li, and H. Ling, “LIME: Low-Light Image Enhancement via Illumination Map Estimation,” *IEEE Trans. Image Process.*, vol. 26, no. 2, pp. 982–993, 2017.

[11] C. Chen, Q. Chen, J. Xu, and V. Koltun, “Learning to See in the Dark,” in *Proc. IEEE CVPR*, 2018, pp. 3291–3300.

[12] Y. Jiang, X. Gong, D. Liu, Y. Cheng, C. Fang, X. Shen, J. Yang, P. Zhou, and Z. Wang, "EnlightenGAN: Deep Light Enhancement Without Paired Supervision," *IEEE Trans. Image Process.*, vol. 30, pp. 2340–2349, 2021.

[13] Y. Zhang, J. Zhang, and X. Guo, "Kindling the Darkness: A Practical Low-Light Image Enhancer," in *Proc. ACM MM*, 2019, pp. 1632–1640.

[14] R. Liu, L. Ma, J. Zhang, X. Fan, and Z. Luo, "Retinex-Inspired Unrolling With Cooperative Prior Architecture Search for Low-Light Image Enhancement," in *Proc. IEEE/CVF CVPR*, 2021, pp. 10561–10570.

[15] L. Ma, T. Ma, R. Liu, X. Fan, and Z. Luo, "Toward Fast, Flexible, and Robust Low-Light Image Enhancement," in *Proc. IEEE/CVF CVPR*, 2022, pp. 5637–5646.

[16] Z. Cui, K. Li, L. Gu, S. Su, P. Gao, Z. Jiang, Y. Qiao, and T. Harada, "You Only Need 90K Parameters to Adapt Light: A Light Weight Transformer for Image Enhancement and Exposure Correction," *arXiv:2205.14871*, 2022.

[17] A. Swanson, M. Kosmala, C. Lintott, R. Simpson, A. Smith, and C. Packer, "Snapshot Serengeti, High-Frequency Annotated Camera Trap Images of 40 Mammalian Species in an African Savanna," *Sci. Data*, vol. 2, Art. no. 150026, 2015.

[18] S. Beery, G. Van Horn, and P. Perona, "Recognition in Terra Incognita," in *Proc. ECCV*, 2018, pp. 456–473.

[19] M. S. Norouzzadeh, A. Nguyen, M. Kosmala, A. Swanson, M. S. Palmer, C. Packer, and J. Clune, "Automatically Identifying, Counting, and Describing Wild Animals in Camera-Trap Images with Deep Learning," *Proc. Natl. Acad. Sci.*, vol. 115, no. 25, pp. E5716–E5725, 2018.

[20] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[21] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal Loss for Dense Object Detection," in *Proc. IEEE ICCV*, 2017, pp. 2980–2988.

[22] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," *arXiv:2004.10934*, 2020.

[23] Y. Pei et al., "Dense Wildlife Object Detection Based on Enhanced YOLOv5," *IEEE Access*, vol. 11, pp. 1502–1513, 2023.

[24] Z. H. Tan et al., "Automatic Wildlife Identification in Camera Trap Images Using Deep Learning," *Ecol. Inform.*, vol. 61, Art. no. 101252, 2021.

[25] Z. Cao et al., "Few-Shot Object Detection for Plateau Wildlife," *Comput. Electron. Agric.*, vol. 198, Art. no. 107038, 2022.

[26] B. Tang et al., "Wildlife Instance Segmentation via Two-Stage Few-Shot Detection," *Remote Sens.*, vol. 13, no. 9, Art. no. 1821, 2021.

[27] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," in *Proc. IEEE/CVF CVPR*, 2023.

[28] G. Jocher, A. Chaurasia, and J. Qiu, "Ultralytics YOLOv8," 2023.

[29] C.-Y. Wang, I.-H. Yeh, and H.-Y. M. Liao, "YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information," *arXiv:2402.13616*, 2024.

[30] A. Wang, H. Chen, L. Liu, et al., "YOLOv10: Real-Time End-to-End Object Detection," *arXiv:2405.14458*, 2024.

[31] S. P. Duth, M. Shibu, and M. Nithyashree, "Animal Detection in Night Light Videos Using Deep Learning," in *Proc. 2nd Int. Conf. Networking, Embedded and Wireless Systems*, 2023, pp. 188–194.

[32] M. Shibu et al., "Animal Detection in Night Videos Using Deep Learning," *IEEE Xplore*, 2024.

[33] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, and S. Zagoruyko, "End-to-End Object Detection with Transformers," in *Proc. ECCV*, 2020.

[34] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, and J. Dai, "Deformable DETR: Deformable Transformers for End-to-End Object Detection," in *Proc. ICLR*, 2021.

[35] Y. Zheng et al., "ADLTFormer: Adaptive Dual-Learning Transformer for Cattle Detection in Low-Light Environments," *IEEE Access*, vol. 12, pp. 158223–158234, 2024.

[36] C. Tian et al., "IV-YOLO: A Lightweight Dual-Branch Object Detection Network," *IEEE Access*, vol. 12, pp. 14082–14094, 2024.

[37] Y. P. Loh and C. S. Chan, "Getting to Know Low-Light Images with the Exclusively Dark Dataset," *Comput. Vis. Image Underst.*, vol. 178, pp. 30–42, 2019.

[38] X. Jia, C. Zhu, M. Li, W. Tang, and W. Zhou, "LLVIP: A Visible-Infrared Paired Dataset for Low-Light Vision," in *Proc. IEEE/CVF ICCV Workshops*, 2021.

[39] H. Pham, T. Nguyen, and C. Pham, "Noise-Aware Zero-Reference Low-Light Image Enhancement for Object Detection," in *Proc. ISPACS*, 2022.

[40] A. Saini, T. Fischer, and M. Y. Yang, "FeatEnHancer: Enhancing Hierarchical Features Under Low-Light Vision," *arXiv:2308.03594*, 2023.

[41] L. Neumann, A. Karg, S. Zhang, et al., "NightOwls: A Pedestrians at Night Dataset," in *Proc. ACCV*, 2018.

Progressive Dual-Branch Illumination-Contrast Pipeline for Night-time Wildlife Detection Using YOLOv5

Palak Harinkhede
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
palakharinkhede_24rco07@dtu.ac.in

Manoj Kumar
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
mkumarg@dce.ac.in

***Abstract*—The primary challenge for night-time wildlife detection using camera-traps is that images are often taken in low-light, with non-uniform exposure, high sensor noise, motion blur, and animals whose body texture is similar to forest backgrounds. YOLOv5's general object detectors are effective when they are tested in daylight, but suffer from a significant loss in accuracy when tested in infrared or low-light images of the same visual categories. Here we provide a complete version of a Progressive Dual-Branch Illumination-Contrast (PDIC) pre-processing pipeline for night-time animal detection. The pipeline is designed to improve the detection without retraining or modifying the downstream YOLOv5 detector. It decomposes an input image into illumination and reflectance using a Retinex inspired approach followed by learnable brightness-curve adjustment and attention-guided local contrast enhancement. Progressive fusion of the two streams to brighten the severely underexposed areas without amplifying noise and preserving object boundaries over multiple refinement passes. Training is unsupervised with no reference objectives like exposure control, spatial illumination smoothness, color constancy, total variation, and gradient consistency. On the NTLNP wildlife dataset, which is composed of 10344 night-time infrared images of 17 animal categories, PDIC with the same YOLOv5 detector achieves the best result, with a mAP50 of 0.883 and a recall of 0.868. The results are significantly better than the baseline CLAHE+Retinex and are competitive with the more recent detector-centric night-time models. The study shows that adaptive preprocessing can be a detector-agnostic, modular improvement stage for low-light wildlife monitoring.**

***Index Terms*—Wildlife detection, image enhancement in low-light conditions, Retinex decomposition, object detection using YOLOv5, unsupervised learning, image preprocessing (CLAHE), NTLNP dataset, camera traps**

I. Introduction

Monitoring of wildlife is crucial to ecological research, conservation planning, analysis of animal populations and to the management of road safety [1]. Forest and nature reserves and animal-crossing areas are ideal places to set up camera traps that capture thousands of images daily for a long time without a human observer being present. But a significant amount of animal movement takes place at dusk, night or early morning times. These scenes are challenging to see: poor lighting, low contrast, distractive background, small foreground animals, or animals that are in the foreground and are similar in texture to foreground plants/shrubs. When operating under this circumstance, automated object detection is much more difficult than normal daylight detection [12]–[16].

The YOLOv5 is an example of modern detectors that are appealing for wildlife monitoring due to their high speed and accuracy [2]–[6]. Compared with YOLOv4, YOLOv5 supports the detection of multiple object categories in a single forward propagation, and it is easy to use in practical systems [17]–[19]. However, its convolutional characteristics are largely acquired from pictures of objects with sharper outline, more intense contrast, and more consistent statistics of light levels. These visual cues are reduced in low-light, infrared camera trap images. Animals in shadowed areas may not be detected, branches and stones may be mistaken for the animals, or the bounding box may be inaccurate if the outline of the animal is not well-exposed on the image.

An obvious cure is to invest in a particular detector for lots of nighttime information. This is useful where many labels, computers and expertise are available. Labling data on night time wildlife, however, are not widely available and can be costly in many field deployments and difficult to annotate. Researchers might also have a working detector and may not want to alter the detector design. This is the area that encourages an opposing effort – enhancement prior to detection. Normally, if there is hidden object structure that can be revealed by the preprocessing and the illumination can be normalized so that the object structure remains visible, then the same detector can be used on a more favorable input distribution, preserving useful gradients.

The traditional enhancement methods are given as a good beginning. Contrast Limited Adaptive Histogram Equalization (CLAHE) enhances the contrast of the local area by re-distributing the intensity values in small image tiles, and Retinex based method is based on the idea of modelling an image as a combination of reflectance and illumination. The previous study used a CLAHE+Retinex preprocess for an animal detection system, combining with YOLOv5, and the preprocess proved useful for animal detection at night on highways. Fixed hand-crafted operations, however, have the same parameters applied to all images. Forest camera-trap data are considerably more variable: some images are virtually black; some have bright infrared "hot spots; some have fog or rain; and some have animals whose body boundary can be seen only in small local gradients. A fixed transformation could add noise to one image and remove noise from another.

Alternative methods that enhance images in darkness are suggested, such as deep low-light enhancement techniques. The Zero-Reference Deep Curve Estimation (Zero-DCE) demonstrated that the enhancement curves of an image can be learned using a neural network without the presence of paired normal-light targets [7]. The trainable modules in RetinexNet and related decomposition networks were used to demonstrate that one can estimate both illumination and reflectance [8]–[11]. These are ideas that are strong but many enhancement systems are built for the detection of human downstream animals not the other way around. To reliably detect wildlife, it is crucial to maintain structures that are important for the detection (e.g. silhouettes, fur textures, legs, horns and head contours) while at the same time avoiding artificial gradients which can lead to false positives.

This work introduces Progressive Dual-Branch Illumination-Contrast (PDIC) pipeline. The main idea is to divide the problem of improving the light into 2 complementary subproblems. The first branch calculates and corrects global illumination to render shadowed areas visible. The second branch enhances the contrast of the local areas making the boundaries of objects and texture more prominent. The two outputs are fused adaptively and progressively applied, over repeated passes. This is a blend of the body of knowledge that Retinex theory provides, and easily learned enhancement [20]–[25].

Its goal is not to replace YOLOv5 but to make the input to be more suitable for detection. After an image has been enhanced, it is immediately transferred to a fixed detector. This modularity is relevant in real conservation applications due to the ability to optimize the preprocessing stage without disturbing the detector stage and to be able to add it to the existing conservation systems. It also renders the method detector-agnostic, as the experiments are conducted on YOLOv5 but the same enhanced images can be fed to any other detection backbones.

This paper makes the following main contributions. First, an enhancement architecture of two stages is formulated for night-time wildlife images, which consists of illumination compensation stage and local contrast enhancement stage. Second, a progressive refinement mechanism is added to correct heavily underexposed regions whose values cannot be reliably refined by a single pass. Third, an unsupervised loss design is adopted, enabling enhancement to be learned without paired bright images. Fourth, the method is evaluated on the NTLNP dataset of 10,344 night-time wildlife images across 17 classes, showing better mAP50 and recall than other static CLAHE+Retinex (retinex) preprocessing methods, and other recent detector-centric approaches in the night-time image domain.

II. Related Work

In this section, we will discuss the classical enhancement technique for night-time detection.Let's now consider the classical enhancement technique for night detection. There are a number of classical preprocessing methods in use, which are easy to implement, fast, and do not require labeled training data. Histogram equalization is used to enhance the contrast over an image, redistributing intensities, whereas CLAHE is used in the local region and is designed to clip the histogram to avoid excessive amplification. An important advantage of CLAHE is that it can be used to lighten local structures in low-light scenes without the need of a physical model of the scene. But it can also increase sensor noise, particularly in the darker parts of the image, where there may be very little useful information in the histogram. Multi-scale Retinex was later introduced in several spatial scales, with manual tuning of the filter sizes and restoration parameters for each image domain [21].

Retinex methods assume that an observed image is the product of reflectance and illumination. Reflectance refers to the intrinsic properties of the scene itself (its surface color and texture), and the illumination is a description of the lighting field. The illumination in the night time images is frequently not uniformly distributed. Retinex correction will try to correct this component to allow hidden reflectance information to be seen. One drawback of the traditional variants of Retinex is the use of handcrafted filters and smoothing assumptions. If the estimation of the illumination is not correct, the method can produce halos, changes in color, or unnatural contrast. Although LIME demonstrated that an illumination-map estimation can reduce the visibility problem in low-light conditions, optimization of LIME has not yet been optimized for the downstream metrics for the detection of wildlife [22].

Before using YOLOv5 inference for animal detection in highways at night, Parkavi et al. [1] applied CLAHE and Robust Retinex Model. The reported precision was high which means that the detector could make reliable detections when the target was visible. Recall was moderate, however, indicating that many animals were still not located. This is normal as static improvement has a limited capacity to adjust to all night time scenarios. The idea of pre-processing, which is a useful concept, is preserved in the present work but with the fixed enhancement replaced by trainable image-specific modules.

In this section, we will look at Detector-Centric Night-time Wildlife Models. Another research direction involves making changes to the detector itself. YOLOv8-night [2] added a channel-attention mechanism for better feature selection for night wildlife images. These detector-oriented approaches also have their merit, as attention can highlight weak features of the animals and block irrelevant background channels. Other studies adopt refined feature pyramids, large kernel convolutions or transformer modules [4], [8] to learn small object and context. These methods are frequently found to increase the accuracy and involve training the detector architecture, and may not be readily adaptable to current systems.

Night-video studies using YOLOv8 [3], [6] have demonstrated that real-time detection of several animal classes, such as elephant, bear, fox and other wildlife classes can be achieved. Typically these pipelines contain resizing, normalization, augmentation and brightness change. The enhancement step is, however, usually a heuristic rather than a learned process. Still struggles with occlusion, blur and dark scenes. The difference is that PDIC does not require redesigning the detector, but rather adaptive preprocessing. Enhances image evidence for any detector, including a fixed YOLOv5 model.

A. Low-Light Enhancement Networks

The basic methods for deep low-light enhancement are used to build trainable preprocessing. Zero-DCE computes a set of pixel-wise curves for the brightening of images without using any paired data. It suffers losses which lead to the adoption of suitable exposure, color consistency, spatial consistency and smoothness [7]. The Retinex assumption [10] is adopted to train decomposition and adjustment modules for low-light images in RetinexNet. Learning to See in the Dark shows that neural networks can be used to recover visual structure from very dark photographs that are captured in RAW files, but for this supervised approach, paired long-exposure targets and RAW input are needed [11]. ExDark also demonstrated that under low illumination, the problem of localization confidence is dramatically different, and that standard detectors often lose localization confidence under such conditions [23].

The enhancement methods are mainly perceptual enhancement. It is not always best to have a colourful picture for object detection. Too much smoothing can eliminate edges that a detector requires while too much sharpening can produce false edges. The proposed pipeline is therefore a mixture of enhancement objectives with a gradient consistency term in order to reinforce the useful boundaries of objects without introducing hallucinations.

In this case, approaches involving multispectral and Transformer data were employed. Recent research also covers models based on transformers and multispectral detection. IV-YOLO [20] is a network that fuses visible and infrared arms to take advantage of complementary information. ADLT-Former [8] is the adaptive dual learning of cattle detection in low-light environments. These systems demonstrate that other modalities or attention to the global give robustness. But they need to be synced with other sensors, day/night augmentation using synthetic days/night, or be trained for a specific domain. Typically, camera-trapping sets will only collect one image stream, either camera traps with infrared or camera traps in low-light. This is the constraint of PDIC: it can only use the available image and enhance it for detection.

B. Research Gap

Based on literature, three gaps are identified. The first is that classical enhancement is efficient but it is not adaptive. Second, changes that are detector-centric may be correct, but need retraining, and may be model dependent. Third, general low light enhancement networks are not necessarily designed to be best suited for wildlife detection capabilities. PDIC fills those gaps with an unsupervised, detector-agnostic preprocessing pipeline that can be trained, and separates the illumination correction task from the contrast refinement task and applies them sequentially.

III. Proposed Methodology

The pipeline designed in this blog will allow enhancements of low-light captures of the wildlife before these pictures are fed into YOLOv5. The design is based on two observations about failure cases at night: Many of them are due to either too little global illumination or too little local contrast. These two problems cannot necessarily be resolved by a single enhancement operator. Blown highlights will increase noise and lose highlights. It can enhance the contrast of the areas, but may not show them when taken alone. Hence, PDIC takes two branches and fuses their output progressive.

The retinex image formation model The method is inspired by Retinex theory [21] that describes an observed image $I(x, y)$ as the product of reflectance $R(x, y)$ and illumination $L(x, y)$:

$$I(x, y) = R(x, y) \cdot L(x, y). \tag{1}$$

The model here assumes that R represents the intrinsic object appearance (fur pattern, body contour, color and texture), and L is the light distribution. For night-time images, L is small and non-uniform. This species might be visible in one image but not so much in another because of the angle of the camera and intensity of the infrared, occlusion by vegetation and distance from the sensor. By estimating the value of L the enhancement module can normalize the light without boosting all the pixels.

A. Illumination Compensation Branch

The first branch, labeled F_I, provides the approximation of an illumination map of the low light input:

$$\hat{L} = F_I(I). \tag{2}$$

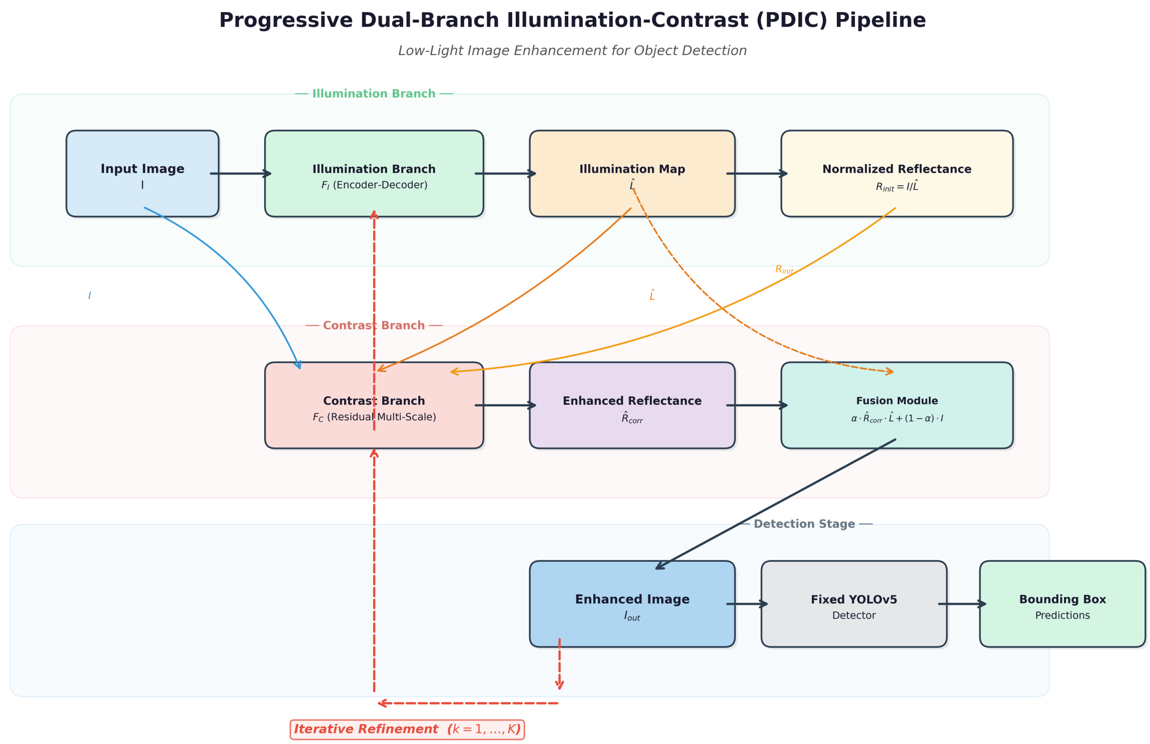
The branch is conceptualized as encoder-decoder CNN with skip connections. The encoder records the illumination pattern in context at multiple scales, and the decoder infers a dense illumination map. Materials are focused to draw attention to features relevant to brightness estimation. The reflectance estimate is calculated by dividing elementwise:

$$R_{init}(x, y) = \frac{I(x, y)}{\max(\hat{L}(x, y), \epsilon)}, \tag{3}$$

where $\epsilon = 10^{-3}$ to avoid calculating division by zero in almost black areas. It is an operation that makes the image look more normal and texture appears in the shadow regions of the image.

The branch is regularized by the edge-aware smoothness to avoid unrealistic illumination maps:

$$\mathcal{L}_{smooth} = \sum_{x,y} |\nabla \hat{L}(x, y)| \exp(-|\nabla I(x, y)|). \tag{4}$$



Architecture of the proposed Progressive Dual-Branch Illumination-Contrast (PDIC) pipeline. The proposed modular preprocessing flow is displayed as illumination branch, contrast branch, adaptive fusion stage and YOLOv5 integration.

Exponential weighting weight a lower penalty near strong image gradients, which favors illumination changes near object boundaries, but encourages smooth illumination over uniform regions.

B. Learnable Brightness Curves

Once illuminated, PDIC uses learnable brightness adjustment inspired by curve-estimation techniques. A curve parameter map A describes the amount of brightness to which each pixel is to be subjected:

$$E(I) = I + A\, I(1 - I). \tag{5}$$

The advantage of this formulation is that it does enhance the dark intensities, more than already-bright intensities, when $A > 0$. Iterative application of such curves will help to gradually enhance without clipping the high lights. There is no fixed gamma correction, rather the parameters of the curve are predicted from the image; and this means that some areas of the image might get stronger enhancement, while others get a weaker one.

C. Contrast Enhancement Branch

The second branch, called F_C, allows to refine local details in the illumination normalized image:

$$\hat{R}_{corr} = F_C\left(\frac{I}{\hat{L}}\right). \tag{6}$$

The branch incorporates residual blocks, spatial attention, channel attention and multi-scale convolutions. The residual blocks maintain the same information for training only the residual part. The multi-scale filters are used to retain the finer details like fur, eyes, legs and horns, or larger silhouettes like the shape of an animal's torso or head. The regions around the target are suppressed from the attention because most of the enhancements would amplify noise.

To avoid over-sharpening there is a gradient consistency term:

$$\mathcal{L}_{grad} = \sum_{x,y} \left| \nabla \hat{R}_{corr}(x, y) - \nabla \left(\frac{I(x, y)}{\hat{L}(x, y)} \right) \right|. \tag{7}$$

This loss will make the network want to reinforce the existing boundary, not create new ones. For detection it is crucial since YOLO-like models are sensitive to artificial contours.

D. Adaptive Fusion

The corrected reflectance and illumination are used to reconstruct the enhanced image:

$$I_{out}(x, y) = \hat{R}_{corr}(x, y) \cdot \hat{L}(x, y). \tag{8}$$

An adaptive blending weight $\alpha(x, y) \in [0, 1]$ can be learned to adaptively fuse improved and original content:

$$I_{out} = \alpha \cdot \hat{R}_{corr} \cdot \hat{L} + (1 - \alpha) \cdot I. \tag{9}$$

The weight comes in handy when areas of images are already well exposed. For instance, an ir-hotspot in the center should not be over-corrected and an animal at the border might need more correction.

E. Progressive Refinement

If the image is very under exposed, it may not be possible to salvage it in a single pass. So PDIC uses an iterative process of refinement:

$$I^{(k+1)} = \Phi(I^{(k)}), \quad I^{(0)} = I, \tag{10}$$

where Φ is a dual-branch enhancement operator. In practical situations, several passes are not needed. The first pass enhances coarse structure, the second pass helps to see mid-level structure, and the third pass gives local contrast. Instead of trying an aggressive transformation on a single image, progressive refinement is safer as it can be applied to parts of an improved image.

F. Unsupervised Training Objective

Since paired low-light and normal-light wildlife images are seldom available, PDIC is trained using no-reference losses. Inverse color constancy loss is penalizing channel imbalance:

$$\mathcal{L}_{color} = \sum_{c \in \{R,G,B\}} (\mu_c(I_{out}) - \bar{\mu})^2, \tag{11}$$

where μ_c is the channel intensity of channel c and $\bar{\mu}$ is the average value taken over all the channels. The loss of exposure promotes proper brightness:

$$\mathcal{L}_{exp} = (\text{mean}(I_{out}) - 0.6)^2 . \tag{12}$$

The total variation loss is used to remove noise in the smooth areas:

$$\mathcal{L}_{tv} = \sum_{x,y} |\nabla I_{out}(x, y)|. \tag{13}$$

The complete objective is:

$$\mathcal{L}_{total} = \lambda_1 \mathcal{L}_{smooth} + \lambda_2 \mathcal{L}_{grad} + \lambda_3 \mathcal{L}_{color} + \lambda_4 \mathcal{L}_{exp} + \lambda_5 \mathcal{L}_{tv}. \tag{14}$$

The hyperparameters λ_1 to λ_5 weigh in physical plausibility, perceptual quality, and structure that is conducive to detection.

G. YOLOv5 Integration

Enhanced I_{out} is then directly input to YOLOv5. No retraining or modification of the detector. This is core to the methodology; all performance improvements are due to improved pre-processing. In the deployment, the pipeline can be plugged into camera capturing and detector inference. The number of progressive passes can be decreased, and/or the enhancement network can be compressed by pruning and/or quantization, if compute is limited.

TABLE I
DETECTION PERFORMANCE COMPARISON.

Method	mAP50	Recall
CLAHE + Retinex + YOLOv5 [1]	0.800	0.773
YOLOv8-night [2]	0.855	–
PDIC + YOLOv5	0.883	0.868

IV. RESULTS AND DISCUSSION

A. Dataset and Evaluation Protocol

The evaluation uses the NTLNP dataset, which contains 10,344 night-time infrared wildlife images across 17 animal categories. The dataset is appropriate for this task because it represents real forest camera-trap conditions rather than controlled laboratory images. It includes difficult cases such as small animals, partial occlusion, cluttered vegetation, weak infrared illumination, and species with similar body texture.

YOLOv5 is used as the downstream detector and is kept fixed during evaluation. Each image is first processed by PDIC and then passed to YOLOv5. Performance is measured using precision, recall, F1 score, mAP50, and mAP50–95. Recall is especially important for wildlife monitoring and road safety because a missed animal can be more harmful than a small number of false positives.

For a fair evaluation, the baseline and proposed method should be executed on the same test split using identical detector settings. The CLAHE+Retinex baseline, the unenhanced YOLOv5 detector, and PDIC+YOLOv5 should share the same confidence threshold, non-maximum suppression threshold, input resolution, and class list. This prevents preprocessing improvements from being confused with threshold tuning. The only variable changed in the proposed configuration is the image supplied to the detector.

B. Overall Detection Performance

Table I compares the proposed pipeline with representative baselines. The PDIC+YOLOv5 combination reaches mAP50 of 0.883 and recall of 0.868. Compared with the CLAHE+Retinex+YOLOv5 baseline, recall improves from approximately 0.773 to 0.868. This indicates that more true animals are detected after adaptive enhancement. The mAP50 improvement from about 0.80 to 0.883 also indicates better localization and confidence ranking.

The improvement is meaningful because the detector itself is unchanged. The result suggests that a significant portion of night-time detection failure comes from poor input visibility rather than insufficient detector capacity alone. By making object evidence clearer, the existing feature extractor can respond more strongly to animals in low-light scenes.

V. CONCLUSION

This paper is a research report of Progressive Dual-Branch Illumination-contrast (PDIC) preprocessing pipeline for night-time animal detection. It is based on the Retinex technique, learnable brightness curves, attention-driven contrast enhancement, adaptive fusion, and progressive refinement. It is trained non-supervised on losses that promote exposure quality, spatial smoothness, color balance, total variation control and gradient consistency.

The conclusion of the key result is that, even if the detector is not changed, significant improvement is possible in the detection by utilizing improved pre-processing. For the NTLNP data set, YOLOv5 improves upon the CLAHE+Retinex baseline with 0.883 mAP50 and 0.868 recall, and with competitive performance compared to specialized night-time detector architectures. This reinforces the argument that smart preprocessing isn't just about pretty pictures: it can be used directly to enhance feature quality that can be fed into object detectors.

Future work should include an evaluation of the pipeline with more camera-trap data, detailed ablations per class, timing on embedded devices and joint optimization of the pipeline with the detectors. If lightweight versions based on pruning, quantization, or conditional enhancement are developed, the method becomes suitable to be deployed in real-time systems for road safety and deployed in battery-powered systems in the field.

References

[1] R. Parkavi et al., "Enhancing Road Safety: Detection of Animals on Highways During Night," *International Journal of Engineering Research & Technology*, vol. 12, no. 3, pp. 1–8, 2025.

[2] T. Wang, S. Ren, and H. Zhang, "Nighttime wildlife object detection based on YOLOv8-night," *Electronics Letters*, vol. 60, no. 15, pp. 598–601, 2024.

[3] S. P. Duth, M. Shibu, and M. Nithyashree, "Animal Detection in Night Light Videos Using Deep Learning," in *Proc. 2nd Int. Conf. on Networking, Embedded and Wireless Systems*, 2023, pp. 188–194.

[4] C. Sun et al., "YOLO-UP: A High-Throughput Pest Detection Model for Dense Cotton Crops Utilizing UAV-Captured Visible Light Imagery," *IEEE Access*, vol. 13, pp. 18452–18465, 2025.

[5] A. Kadali et al., "A Comprehensive Survey of Technologies for Wildlife Detection and Accident Prevention," in *Proc. 3rd Int. Conf. on Intelligent Systems, Advanced Computing and Communication*, IEEE, 2024, pp. 115–120.

[6] M. Shibu et al., "Animal Detection in Night Videos Using Deep Learning," IEEE Xplore, 2024.

[7] C. Guo et al., "Zero-Reference Deep Curve Estimation for Low-Light Image Enhancement," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition*, 2020, pp. 1780–1789.

[8] Y. Zheng et al., "ADLTFormer: Adaptive Dual-Learning Transformer for Cattle Detection in Low-Light Environments," *IEEE Access*, vol. 12, pp. 158223–158234, 2024.

[9] Y. Liu, H. Chen, and C. Li, "Deep Retinex Decomposition for Low-Light Enhancement," in *Proc. British Machine Vision Conference*, 2021, pp. 1–12.

[10] C. Wei, W. Wang, W. Yang, and J. Liu, "Deep Retinex Network for Low-Light Enhancement," in *Proc. British Machine Vision Conference*, 2018, pp. 1–13.

[11] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to See in the Dark," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2018, pp. 3291–3300.

[12] Y. Pei et al., "Dense Wildlife Object Detection Based on Enhanced YOLOv5," *IEEE Access*, vol. 11, pp. 1502–1513, 2023.

[13] Z. H. Tan et al., "Automatic Wildlife Identification in Camera Trap Images Using Deep Learning," *Ecological Informatics*, vol. 61, p. 101252, 2021.

[14] Y. Mao et al., "Re-parameterized Convolutions for Wildlife Detection in Neighboring Regions," *Sensors*, vol. 22, no. 12, p. 4342, 2022.

[15] Z. Cao et al., "Few-Shot Object Detection for Plateau Wildlife," *Computers and Electronics in Agriculture*, vol. 198, p. 107038, 2022.

[16] B. Tang et al., "Wildlife Instance Segmentation via Two-Stage Few-Shot Detection," *Remote Sensing*, vol. 13, no. 9, p. 1821, 2021.

[17] J. Redmon and A. Farhadi, "YOLOv3: An Incremental Improvement," arXiv:1804.02767, 2018.

[18] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv:2004.10934, 2020.

[19] G. Jocher et al., "YOLOv5 by Ultralytics," GitHub, 2020.

[20] C. Tian et al., "IV-YOLO: A Lightweight Dual-Branch Object Detection Network," *IEEE Access*, vol. 12, pp. 14082–14094, 2024.

[21] D. J. Jobson, Z. Rahman, and G. A. Woodell, "A Multiscale Retinex for Bridging the Gap Between Color Images and the Human Observation of Scenes," *IEEE Transactions on Image Processing*, vol. 6, no. 7, pp. 965–976, 1997.

[22] X. Guo, Y. Li, and H. Ling, "LIME: Low-Light Image Enhancement via Illumination Map Estimation," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 982–993, 2017.

[23] Y. P. Loh and C. S. Chan, "Getting to Know Low-Light Images with the Exclusively Dark Dataset," *Computer Vision and Image Understanding*, vol. 178, pp. 30–42, 2019.

[24] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[25] T.-Y. Lin et al., "Focal Loss for Dense Object Detection," in *Proc. IEEE Int. Conf. Computer Vision*, 2017, pp. 2980–2988.

Night-time Wildlife Detection Using Computer Vision: A Review of Deep Learning, Thermal Imaging, and Edge-AI Approaches

Palak Harinkhede
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
palakharinkhede_24rco07@dtu.ac.in

Manoj Kumar
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
mkumarg@dce.ac.in

***Abstract*—Many species targeted are nocturnal or crepuscular, which make them poorly represented in daytime observations and therefore a focus of the night-time wildlife detection, road-safety mitigation, anti-poaching and ecological decision support. This review summarizes recent advances in computer-vision techniques for night- and low-light monitoring of wild animals with a focus on publications from 2022–2025, as well as on some related foundational research that is still technically appropriate. The examined literature includes convolutional detectors, models from the YOLO family, transformer and hybrid architectures, thermal and infrared imaging, camera-trap analytics, the use of unmanned aerial vehicles (UAVs) for monitoring, multimodal sensing, and edge-AI deployment. In the literature, the YOLOv5–YOLOv8 series and lightweight CNN architectures are dominant in real-time applications due to their good speed–accuracy trade-offs, while transformer-based detectors have better contextual reasoning and occlusion handling capabilities, but they need more compute power and larger datasets. Thermal and near-infrared modalities are particularly useful during night-time and are hindered by sensor noise, low spatial resolution, temperature-dependent contrast and small annotated wildlife datasets. Public camera-trap resources and overall low light/thermal data have facilitated benchmarking, but there has been a significant challenge to consistently deploy cameras across geographic regions, seasons, sensors, and species. The review reveals ongoing issues with small-object recognition, motion blur, vegetation and livestock false positives, energy-limited edge inference and transparent decision-making for conservation stakeholders. The work presented here should be built on other novel techniques, such as self-supervised pretraining, foundation models, federated learning, generative augmentation, explainable AI, and sensor-aware edge optimization, to create scalable, reliable, and field-ready NWD systems.**

Keywords: night-time wildlife detection, camera traps, thermal imaging, infrared vision, YOLO, transformer detectors, edge AI, conservation technology.

I. Introduction

Wildlife monitoring can help to support population estimation, species distribution modelling, habitat management, road-collision mitigation and enforcement against illegal hunting. While traditional field surveys, manual camera-trap inspection, radio telemetry, and direct observation continue to be valuable tools, they are not scalable when conservation programs need to sift through millions of images across the varied landscapes. Thus, conservation technology is increasingly relying on computer vision for the automated detection, classification, counting and event triage in images, video, thermal streams, and UAV footage of animals [1]–[3].

The monitoring is especially important at night. Night is the peak time of activity for many mammals, reptiles, amphibians and some birds; human disturbances can cause animal behavior to change from daytime to nighttime activity. Other applications include road corridors, farms, protected-area boundaries, railway lines and anti-poaching patrols, where early warning requires speedy recognition in low-light conditions. Visual conditions, however, that promote the ecological value of night-time monitoring also present technical challenges. The low photon rates, sensor gain noise, glare from active infrared emitter, rain, fog, movement blur, partial occlusion, cluttered vegetation and long-tailed distributions reduce detection performance. While thermal cameras reduce reliance on visible light, they also present several challenges: reduced resolution, reflection of heat in the scene, small differences in appearance between classes, thermal variation in daytime, and lack of labelled datasets.

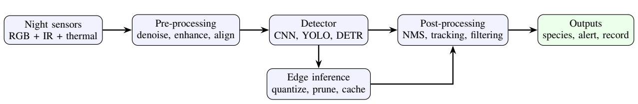
Deep learning has revolutionized the capabilities of wildlife detection. Two-stage detectors like Faster R-CNN are useful when accuracy and localization quality are the primary concerns, while single stage CNN detectors like SSD, RetinaNet, EfficientDet, YOLO are useful for real-time inference, transformer-based detectors like DETR, Deformable DETR, DINO, Grounding DINO and Swin-based hybrids like DINO/RetinaNet and DINO/SSD enhance global context modelling [4]–[8]. Meanwhile, the development of model families like YOLOv7, YOLOv8, YOLOv9, and YOLOv10 has spurred field deployment due to their high throughput and convenient training pipelines [9]–[12]. There are recent foundation models and segmentation systems that further push the design space for detection, prompting, pseudo-labelling, and annotation acceleration like sam [13] and dinov2 [14].

This review is concentrated on the night-time wildlife detection through computer vision, especially during the time frame

TABLE I

REPRESENTATIVE RECENT APPROACHES FOR NIGHT-TIME OR LOW-LIGHT WILDLIFE-RELATED DETECTION. REPORTED SCORES VARY BY DATASET AND SHOULD NOT BE INTERPRETED AS DIRECTLY COMPARABLE.

Authors	Year	Method	Dataset/source	Metric	Key contribution
Beery et al. [1]	2024	Benchmark challenge detectors	iWildCam camera traps	mAP/F1	Large-scale camera-trap benchmarking and domain-shift analysis
Wang et al. [9]	2023	YOLOv7	COCO/custom transfer	AP/FPS	Efficient real-time detector architecture widely reused in wildlife systems
Jocher et al. [10]	2023	YOLOv8	Custom object datasets	mAP/FPS	Anchor-free practical baseline for conservation deployments
Wang et al. [11]	2024	YOLOv9	COCO/custom transfer	AP/latency	Programmable gradient information for efficient detection
Wang et al. [12]	2024	YOLOv10	COCO/custom transfer	AP/latency	NMS-free real-time object detection design
Zhang et al. [7]	2022	DINO detector	COCO/custom transfer	AP	Strong transformer detector for high-accuracy offline analysis
Liu et al. [8]	2023	Grounding DINO	Open-vocabulary data	AP/zero-shot	Promptable detection useful for annotation and rare-species discovery
Kirillov et al. [13]	2023	SAM	Segmentation prompts	IoU	Foundation segmentation for mask-assisted wildlife annotation
Oquab et al. [14]	2024	DINOv2	Self-supervised pretraining	transfer AP	General visual representations for small labelled datasets
Jia et al. [15]	2021	LLVIP baseline fusion	Infrared-visible pairs	AP	Foundational low-light IR-visible benchmark for night detection



Architecture of a nighttime wildlife detection pipeline.

2022-2025. It has four-fold contributions. First, it categorizes recent approaches into four groups: CNN, YOLO, transformer and multimodal/thermal. Secondly, it provides an overview of datasets and evaluation protocols for nocturnal wildlife monitoring. Third, it doesn't just provide benchmark scores; it discusses ongoing issues with field deployment. Fourthly, it pinpoints research areas and future prospects for strong, effective and reliable night-time monitoring systems.

II. RECENT APPROACHES FOR NIGHT-TIME WILDLIFE DETECTION

A. CNN-based Detection Methods

Classical deep detectors are still applicable in nocturnal wildlife monitoring due to their inductive bias of image grids and their ability to be trained with moderate levels of annotations. For camera trap and UAV imagery, particularly when animals are small in the image or partially covered by vegetation, faster R-CNN and its variants based on feature pyramids ([16]) are typically used as accuracy baselines. During night-time scenarios, two-stage proposals can better localize a weak silhouette of an animal than very lightweight single-stage proposals, but the drawback is that it takes more time and energy than the single-stage proposal. However, there are more deployment friendly alternatives that avoid the need for proposal generation and instead perform dense prediction: SSD [17], RetinaNet [18] and EfficientDet [19]. The focal loss of RetinaNet is particularly important for camera-trap streams as many of the frames also include objects from the background that are not of interest and hence lead to high foreground to background imbalance.

Recent camera trapping studies indicate that CNN detectors are typically used in conjunction with ecological filtering, such as in workflows for video localization, small mammal processing, active learning, ensemble separation of the wildlife images, long-tailed recognition, and edge camera trapping classification, instead of individually [20]–[25]. Pre-processing systems that utilize the concept of animal/person/vehicle localization from the MegaDetector paradigm can exclude empty images from the species classification stage, which decreases the time and effort required for manual review and reduces the burden of annotation.Mechanisms similar to the MegaDetector paradigm for animal/person/vehicle localization can be used to filter out empty images prior to species classification, thus reducing the amount of annotation and manual checking. The concept of domain adaptation is vital: models trained on one reserve could not be deployed at another due to differences in vegetation, lighting, the type of infrared flash, camera orientation, and species diversity. CNN backbones can learn thermal signatures well but feature-pyramid design, anchor selection and super-resolution pre-processing is of significance for small mammals and birds, which can only occupy a few pixels in thermal monitoring.

B. YOLO-based Models

YOLO-family detectors are the most popular for wildlife surveillance due to their trade-off between accuracy, speed of inference, availability of software and edge deployment. YOLOv5 and YOLOv7 are popular choices for camera trap and UAV applications, as they can be trained on custom datasets and have compact variants for use on devices, like NVIDIA Jetson modules. YOLOv8 added a new anchor-free module and has been widely used as a baseline for conservation prototypes, while YOLOv9 and YOLOv10 appear focused on better gradient information flow and end-to-end efficiency [9]–[12].

Typically, YOLO models are modified in four ways to detect wildlife during the night time. The first are attention modules or feature-enhancing blocks to boost weak contours of animals in IR and low-light images. Second, the small-object heads or higher resolution input tiles are used in the case of a distant animal in UAV imagery. Third, lightweight backbone and pruning/quantization are adopted for edge inference. Fourth, temporal post-processing minimises false positives due to sensor artefacts, raindrops, headlights and vegetation. YOLO is found to be appropriate for ecological analytics and real-time alerting, with examples like UAV habitat-object detection and roadkill-inspired wildlife warning prototypes [27], [28] featuring reported systems with various evaluation metrics, including precision, recall, F1-score, mAP@0.5, mAP@0.5:0.95, and FPS.

Transformer-based and Hybrid Methods Transformer based object detectors overcome the shortcomings of purely conv-based models by introducing long-range context and object relation modeling. Instead, DETR replaced anchor-based prediction by bipartite matching and eliminated hand-designed anchors and non-maximum suppression [4]. Deformable DETR enhanced the convergence and multi-scale attention, thus making transformer detection more feasible for small objects [5]. For denoising training and query design, DINO and its variants further enhanced the process [7]. In the context of wildlife images, these properties are appealing because animals are often camouflaged, partially visible and not in simple backgrounds.

Hybrid CNN–transformer approaches are particularly promising for night-time monitoring. CNN layers are designed to extract local texture efficiently, while the Swin Transformer blocks or deformable attention layers are added to capture other context information, like the animal-road relationship, the animal herd structure, or the repeated thermal blobs in a frame [6]. However, not all transformers are created equal. They are generally more data-hungry, more augmenting, and have higher memory bandwidth than the YOLO variants. If the transformer detector is used in small datasets, it can easily overfit without using large-scale pretraining like DINOv2, Grounding DINO, or Segment Anything workflows [8], [13], [14]. In this sense, their first use might be in assisting in annotations, pseudo-labelling, open vocabulary search, and high accuracy offline analysis, more than in battery-powered real-time alerts.

The multi-modal and thermal imaging approaches are explored. Sensor diversity has significant advantages for night-time detection. The advantages of passive thermal cameras are: they do not require any light to detect the heat contrast; they give a morphology of familiar animals in grayscale images; the advantages of active infrared camera trap are: they are triggered by heat; the advantages of RGB camera are: they provide the context of colour in the day; the advantages of radar or acoustic sensors are: they trigger the visual capture according to the heat contrast; and the advantages of the UAV platforms are: they provide flexible spatial coverage. The fusion could be performed at the pixel level, feature level, level of decision or track level. Feature-level fusion has the benefits of learning complementary representation between RGB and thermal images, while decision-level fusion is more convenient to deploy when the field of view and frame rate of the sensor are different.

Thermal and infrared methods are especially pertinent in the case of large mammals, roadside detection and anti-poaching surveillance. Public thermal datasets like FLIR ADAS, LLVIP are not animal-specific but they can be used for pretraining and evaluation of low-light pedestrian/vehicle detectors with architectures that can be transferred to animals [15], [29]. Thermal data for specific wildlife are limited and many studies utilize special UAV flights, camera traps, or controlled field recordings. Thermal imagery often increases detectability at night but decreases separability of species due to lack of texture and colour. Therefore, having a robust pipeline means applying thermal detection for localization and alerting, and if available, then applying RGB/IR confirmation in a second stage.

III. Datasets and Performance Evaluation

Camera-trap datasets are the key empirical basis for wildlife detection. They contain large passive observations, but labels may describe species, boxes, image-level presence, sequence events, or blanks. The iWildCam benchmark series focuses on

TABLE II
DATASETS OF RELEVANCE FOR NIGHTTIME WILDLIFE DETECTION AND RELATED LOW-LIGHT/THERMAL PRETRAINING.

Dataset	Main years	Modality	Annotation type	Relevance and limitation
iWildCam [1]	2022–2024	RGB/IR camera trap	image labels and boxes	Strong benchmark for camera-trap domain shift; not exclusively nocturnal
LILA BC [30]	active	camera-trap ecology images	mixed labels	Aggregates conservation datasets; annotation granularity varies
Snapshot data [31]	base	RGB/IR camera trap	species/event labels	Ecologically large; limited bounding boxes for detection studies
MegaDetector [26]	active	camera trap	animal/person/ vehicle	Practical detector pretraining and empty-image filtering
FLIR ADAS thermal [29]	2022	thermal/RGB	boxes	Useful thermal pretraining; road objects dominate
LLVIP [15]	2021	infrared-visible pairs	pedestrian boxes	Foundational paired low-light benchmark; not wildlife-specific
ExDark [32]	base	low-light RGB	boxes/classes	Useful for low-light robustness; limited ecological relevance
Custom UAV thermal data	2022–2025	thermal/RGB aerial	boxes/counts	Relevant to nocturnal surveys; often private and site-specific

domain shift across camera locations and ecological contexts [1]. LILA BC supports reproducible benchmarking across conservation datasets [30], while Snapshot data remain useful at ecological scale despite limited night-specific bounding boxes [31]. Recent workflows combine these data with MegaDetector for empty-frame filtering and local target-species annotation [26].

Thermal and infrared datasets are less mature. FLIR ADAS and LLVIP provide paired or thermal imagery useful for night-driving contexts, but their taxonomic content is not wildlife-centred [15], [29]. Public UAV thermal wildlife corpora remain limited, weakening cross-paper comparison and generalization tests. Low-light datasets such as ExDark can support robustness studies, although enhancement quality does not necessarily correlate with animal detection accuracy [32].

Evaluation should match deployment goals. Precision, recall, and F1-score summarize reliability and completeness; IoU measures localization overlap; and mAP summarizes precision across classes and confidence thresholds. Conservation triage may prioritize recall, whereas road-warning and anti-poaching systems require high precision and low latency. FPS, model size, energy per inference, memory footprint, and weather robustness should therefore accompany mAP.

IV. Challenges and Future Research Directions in Night-time Wildlife Detection

Night-time detection is limited by low illumination, infrared glare, motion blur, occlusion, and small target scale. Thermal sensing reduces dependence on visible light, but warm rocks, livestock, humans, engines, and sun-heated backgrounds can produce false positives, while small animals may disappear when thermal contrast is weak. Data scarcity further amplifies these errors because rare species and unusual night conditions are under-represented.

Robust systems must therefore combine sensor-aware enhancement, camera-disjoint evaluation, uncertainty calibration, and human-in-the-loop review. Edge deployment adds strict limits on latency, energy, storage, and communication, so pruning, quantization, distillation, and event-triggered inference are essential.

A. Emerging Trends and Future Research Directions

Edge AI, self-supervised learning, foundation models, and federated learning are the most important directions. On-device inference can reduce bandwidth and enable real-time alerts [33], [34], while DINOv2, SAM, and Grounding DINO can support transfer learning, pseudo-labelling, and rare-species discovery [8], [13], [14]. Federated training can exploit distributed conservation data without centralizing sensitive images, and explainable AI can improve trust by revealing whether decisions depend on animal morphology or background shortcuts.

V. Comparative Analysis and Discussion

The literature shows an accuracy–latency–robustness trade-off. Two-stage CNNs remain effective for accurate offline localization, while SSD, RetinaNet, EfficientDet, and especially YOLO variants are more feasible for edge localization. YOLO models provide the strongest current deployment balance, but they can fail across camera sites, seasons, sensors, and species if domain shift is not explicitly assessed.

Transformers and foundation models are useful for contextual reasoning, annotation, and rare-class discovery, but they are often too memory-, data-, and energy-intensive for field devices. Thermal and multimodal systems improve night

TABLE III
RESEARCH-GAP MATRIX FOR NIGHT-TIME WILDLIFE DETECTION.

Gap	Current limitation	Research opportunity	Expected impact
Public nocturnal datasets	Few labelled wildlife thermal/IR benchmarks	Multi-site night datasets with boxes and metadata	Reproducible comparison
Small-object detection	Distant UAV animals occupy few pixels	Tiling, temporal fusion, scale-aware heads	Better aerial surveys
Cross-site robustness [1]	Models overfit location and vegetation	Domain adaptation and self-supervised pre-training	Lower relabelling cost
Edge deployment [25], [33]	High-mAP models exceed power budgets	Quantization, pruning, TinyML triggers	Longer unattended operation
Multimodal fusion	RGB, IR, and thermal streams remain under-integrated	Calibrated feature and decision fusion	Reliable all-weather detection
Rare species [8], [24]	Long-tailed classes lack examples	Few-shot, open-vocabulary, generative augmentation	Threatened-species monitoring

TABLE IV
COMPARATIVE EVALUATION OF THE MAIN METHOD FAMILIES FOR NOCTURNAL WILDLIFE DETECTION.

Method family	Strengths	Weaknesses	Real-time feasibility	Best use case
Faster R-CNN/FPN [16]	accurate localization	slow, heavy memory requirement	moderate–low	offline camera-trap analysis
SSD/RetinaNet/EfficientDet [17]–[19]	efficient dense prediction	weak on tiny low-contrast targets	moderate–high	embedded monitoring
YOLOv5–YOLOv10 [9]–[12]	fast, deployable ecosystem	domain shift and false positives	high	real-time alerts and UAV scans
DETR/DINO/Swin hybrids [4], [6], [7]	global context modelling	high data and compute demand	low–moderate	occlusion-heavy offline analysis
Foundation models [8], [13], [14]	annotation and transfer support	large models, uncertain field robustness	low without distillation	pseudo-labelling and rare-species search
Thermal/multimodal fusion [15], [35]	robust to darkness	sensor cost and loss of texture	moderate	night roads and anti-poaching

detection, yet the loss of colour and texture cues limits fine-grained species recognition. Future evaluations should report camera-disjoint splits, night-specific metrics, latency on named hardware, and failure cases alongside mAP.

VI. CONCLUSION

Over the past few years, the development of night-time wildlife detection has made significant advances, including CNN detectors, YOLO-family real-time models, multi-modal sensing, and transformer architectures, as well as thermal imaging. YOLO is prevalent for real-time implementations, while larger CNN or transformer models are used for offline analysis and support for annotation. While thermal and infrared sensors are essential for night vision, they cannot solve the problems of small objects, occlusions, thermal ambiguity or species-level identification.

The main shortcoming of current systems is not the lack of high detection capability, but the difference in performance from the bench to the field. The datasets are still siloed, a significant number of thermal wildlife datasets are private, and the protocols for evaluating the datasets do not necessarily test whether they can be adopted across sites, over seasons and across sensors. Shared nocturnal benchmarks, multimodal fusion, self-supervised pre-training on ecological imagery, edge-aware model compression, uncertainty estimation and human-centred explainability are the most significant areas that need future research. Future generations of systems will combine robust sensing, efficient inference, easy-to-understand decision support and ecological validation to achieve conservation results and not simply detection accuracy.

REFERENCES

[1] Beery, S., Morris, D., Yang, S., et al.: The iWildCam 2024 Challenge Dataset and Benchmark. Proceedings of the CVPR Workshops (2024).

[2] Willi, M., Pitman, R.T., Cardoso, A.W., et al.: Identifying animal species in camera trap images using deep learning and citizen science. Methods in Ecology and Evolution (updated applications discussed in recent camera-trap literature) (2022).

[3] Tuia, D., Kellenberger, B., Beery, S., et al.: Perspectives in machine learning for wildlife conservation. Nature Communications 13, 792 (2022).

[4] Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-End Object Detection with Transformers. ECCV (2020).

[5] Zhu, X., Su, W., Lu, L., Li, B., Wang, X., Dai, J.: Deformable DETR: Deformable Transformers for End-to-End Object Detection. ICLR (2021).

[6] Liu, Z., Lin, Y., Cao, Y., et al.: Swin Transformer: Hierarchical Vision Transformer using Shifted Windows. ICCV (2021).

[7] Zhang, H., Li, F., Liu, S., et al.: DINO: DETR with Improved DeNoising Anchor Boxes for End-to-End Object Detection. ICLR (2023), arXiv preprint first released 2022.

[8] Liu, S., Zeng, Z., Ren, T., et al.: Grounding DINO: Marrying DINO with Grounded Pre-Training for Open-Set Object Detection. arXiv:2303.05499 (2023).

[9] Wang, C.Y., Bochkovskiy, A., Liao, H.Y.M.: YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors. CVPR (2023).

[10] Jocher, G., Chaurasia, A., Qiu, J.: Ultralytics YOLOv8. Ultralytics software release (2023).

[11] Wang, C.Y., Yeh, I.H., Liao, H.Y.M.: YOLOv9: Learning What You Want to Learn Using Programmable Gradient Information. arXiv:2402.13616 (2024).

[12] Wang, A., Chen, H., Liu, L., et al.: YOLOv10: Real-Time End-to-End Object Detection. arXiv:2405.14458 (2024).

[13] Kirillov, A., Mintun, E., Ravi, N., et al.: Segment Anything. ICCV (2023).

[14] Oquab, M., Darcet, T., Moutakanni, T., et al.: DINOv2: Learning Robust Visual Features without Supervision. Transactions on Machine Learning Research (2024).

[15] Jia, X., Zhu, C., Li, M., Tang, W., Zhou, W.: LLVIP: A Visible-infrared Paired Dataset for Low-light Vision. ICCV Workshops (2021).

[16] Ren, S., He, K., Girshick, R., Sun, J.: Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. IEEE TPAMI 39(6), 1137–1149 (foundational detector).

[17] Liu, W., Anguelov, D., Erhan, D., et al.: SSD: Single Shot MultiBox Detector. ECCV (foundational single-stage detector).

[18] Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal Loss for Dense Object Detection. ICCV (foundational RetinaNet method).

[19] Tan, M., Pang, R., Le, Q.V.: EfficientDet: Scalable and Efficient Object Detection. CVPR (foundational efficient detector).

[20] Simoes, F., Bouveyron, C., Precioso, F.: DeepWILD: Wildlife identification, localisation and estimation on camera trap videos using deep learning. Ecological Informatics 75, 102095 (2023).

[21] Bohner, H., Kleiven, E.F., Ims, R.A., Soininen, E.M.: A semi-automatic workflow to process images from small mammal camera traps. Ecological Informatics 76, 102150 (2023).

[22] Bothmann, L., Wimmer, L., Charrakh, O., Weber, T., Edelhoff, H., Peters, W., Nguyen, H., Benjamin, C., Menzel, A.: Automated wildlife image classification: An active learning tool for ecological applications. Ecological Informatics 77, 102231 (2023).

[23] Roy, A., Jha, R.R., et al.: A method for automatic identification and separation of wildlife images using ensemble learning. Ecological Informatics 77, 102262 (2023).

[24] Zhu, C., Jia, X., et al.: Bag of tricks for long-tail visual recognition of animal species in camera-trap images. Ecological Informatics 76, 102060 (2023).

[25] Nazir, S., Kaleem, M.: Object classification and visualization with edge artificial intelligence for a customized camera trap platform. Ecological Informatics 79, 102453 (2024).

[26] Beery, S., Morris, D., Yang, S.: Efficient Pipeline for Camera Trap Image Review. arXiv:1907.06772 (foundational conservation detector workflow).

[27] Lawrence, B., de Lemmus, E., Cho, H.: UAS-Based real-time detection of Red-Cockaded Woodpecker cavities in heterogeneous landscapes using YOLO object detection algorithms. Remote Sensing 15(4), 883 (2023).

[28] Kim, J., Lee, S.: Real-time wildlife roadkill detection using YOLOv8. Journal of Internet of Things Convergence 10(5), 185–196 (2024).

[29] Teledyne FLIR: FLIR Thermal Dataset for Algorithm Training, ADAS thermal/RGB dataset release and documentation (2022).

[30] LILA BC: Labeled Information Library of Alexandria: Biology and Conservation computer vision datasets. Microsoft AI for Earth repository and dataset portal (accessed 2024).

[31] Swanson, A., Kosmala, M., Lintott, C., et al.: Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Scientific Data 2, 150026 (foundational dataset).

[32] Loh, Y.P., Chan, C.S.: Getting to Know Low-light Images with The Exclusively Dark Dataset. Computer Vision and Image Understanding 178, 30–42 (foundational low-light dataset).

[33] Hughey, L.F., et al.: Edge computing in wildlife behavior and ecology. Trends in Ecology & Evolution 39(2), 128–130 (2024).

[34] Clark, R., Wang, J., et al.: Edge deep learning in computer vision and medical diagnostics: a comprehensive survey. Artificial Intelligence Review 57, 11033 (2024).

[35] Baldwin, R.W., Beaver, J.T., Messinger, M., et al.: Camera trap methods and drone thermal surveillance provide reliable, comparable density estimates of large, free-ranging ungulates. Animals 13(11), 1884 (2023).

Gmail

Palak Harinkhede <palakharinkhede1@gmail.com>

ICDAM 2026: Paper Notification for Paper ID 1598

1 message

ICDAM 2026 <icdam.conf@gmail.com> Wed, May 27, 2026 at 7:51 AM
To: Palakharinkhede <palakharinkhede1@gmail.com>

7th International Conference on Data Analytics & Management (ICDAM-2026)!

Dear Author(s),

Greetings from **ICDAM 2026**!

We congratulate you that your paper with submission ID **1598** and Paper Title ' **Progressive Dual-Branch Illumination-Contrast Pipeline for Night-time Wildlife Detection Using YOLOv5** ' has been accepted for publication in the Springer LNNS series [Approved]-[Indexing: SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago; All books published in the series are submitted for consideration in Web of Science]. This acceptance means your paper is among the top 20% of the papers received/reviewed. We urge you to complete your registration immediately to secure your spot at this highly anticipated event. **Kindly complete the registration by 29th May 2026. No further Extension.**

Please register as soon as possible and submit the following documents to icdam.conf@gmail.com as soon as possible.
1. Final Camera-Ready Copy (CRC) as per the springer format. (See https://icdam-conf.com/downloads)
2. Copy of e-receipt of registration fees. (For Registration, see https://icdam-conf.com/registrations)
3. The final revised copy of your paper must be mailed to the conference mail id.

Note : Standard Paper size – 10-12 pages. Over length of more than 12 pages, paper charges USD 20 per extra page

The reviewers comments are given at the bottom of this letter, please improve your paper as per the reviewers comments. While preparing the final CRC manuscript, kindly check the following google link of proceedings of the previous International Conference on Data Analytics and Management:

https://link.springer.com/conference/icdam

and it is suggested to cite the relevant latest papers matching the area of your current research paper.

The paper prior to submission should be checked for plagiarism from licensed plagiarism softwares like Turnitin/iAuthenticate etc. The similarity content should not exceed 15%, and AI content must not exceed 20%.

Pay registration fees via online portal:
[Kindly note – The conference being organised in Hybrid Mode and you can choose the mode of presentation in either physical (offline) or digital (online) mode; then pay the registration fees]

Research Scholar/Student: https://icdam-conf.com/registrations
Academician: https://icdam-conf.com/registrations
Industrial Participant: https://icdam-conf.com/registrations

Another Method via Paypal with extra 5% service charges:
https://www.paypal.com/paypalme/ICICCConference?locale.x=en_GB

Another Method via Razorpay:
https://pages.razorpay.com/pl_S4t9yYAS644i54/view

With Regards
TPC Chair
ICDAM-2026 Conference

Reviewer 1 Comments :
1. Clarify the main novelty of the PDIC pipeline compared to existing Retinex and CLAHE-based enhancement methods.

2. Provide more details on the architecture of the dual-branch illumination–contrast decomposition module.

3. Explain the progressive fusion mechanism more clearly, especially how noise suppression is maintained.

4. Include ablation studies to show the contribution of each component (illumination, contrast, attention).

5. Compare PDIC with recent end-to-end low-light detection and enhancement methods, not only preprocessing baselines.

6. Improve grammar and reduce sentence complexity for better academic readability.

Reviewer 2 Comments :

Gmail

Palak Harinkhede <palakharinkhede1@gmail.com>

ICDAM 2026: Paper Notification for Paper ID 1623

1 message

ICDAM 2026 <icdam.conf@gmail.com> Wed, May 27, 2026 at 7:51 AM
To: Palakharinkhede <palakharinkhede1@gmail.com>

7th International Conference on Data Analytics & Management (ICDAM-2026)!

Dear Author(s),

Greetings from **ICDAM 2026**!

We congratulate you that your paper with submission ID **1623** and Paper Title ' **Night-time Wildlife Detection Using Computer Vision: A Review of Deep Learning, Thermal Imaging, and Edge-AI Approaches** ' has been accepted for publication in the Springer LNNS series [Approved]-[Indexing: SCOPUS, INSPEC, WTI Frankfurt eG, zbMATH, SCImago; All books published in the series are submitted for consideration in Web of Science]. This acceptance means your paper is among the top 20% of the papers received/reviewed. We urge you to complete your registration immediately to secure your spot at this highly anticipated event. **Kindly complete the registration by 29th May 2026. No further Extension.**

Please register as soon as possible and submit the following documents to icdam.conf@gmail.com as soon as possible.
1. Final Camera-Ready Copy (CRC) as per the springer format. (See https://icdam-conf.com/downloads)
2. Copy of e-receipt of registration fees. (For Registration, see https://icdam-conf.com/registrations)
3. The final revised copy of your paper must be mailed to the conference mail id.

Note : Standard Paper size – 10-12 pages. Over length of more than 12 pages, paper charges USD 20 per extra page

The reviewers comments are given at the bottom of this letter, please improve your paper as per the reviewers comments. While preparing the final CRC manuscript, kindly check the following google link of proceedings of the previous International Conference on Data Analytics and Management:

https://link.springer.com/conference/icdam

and it is suggested to cite the relevant latest papers matching the area of your current research paper.

The paper prior to submission should be checked for plagiarism from licensed plagiarism softwares like Turnitin/iAuthenticate etc. The similarity content should not exceed 15%, and AI content must not exceed 20%.

Pay registration fees via online portal:
[Kindly note – The conference being organised in Hybrid Mode and you can choose the mode of presentation in either physical (offline) or digital (online) mode; then pay the registration fees]

Research Scholar/Student: https://icdam-conf.com/registrations
Academician: https://icdam-conf.com/registrations
Industrial Participant: https://icdam-conf.com/registrations

Another Method via Paypal with extra 5% service charges:
https://www.paypal.com/paypalme/ICICCConference?locale.x=en_GB

Another Method via Razorpay:
https://pages.razorpay.com/pl_S4t9yYAS644i54/view

With Regards
TPC Chair
ICDAM-2026 Conference

Reviewer 1 Comments :
1. Clearly structure the review by grouping literature into clear categories (e.g., YOLO-based methods, transformers, thermal imaging, UAV-based monitoring, edge AI).

2. Provide more specific citations or examples for key claims, especially regarding YOLOv5–YOLOv8 and transformer-based detector performance.

3. Clarify the inclusion criteria for selected studies (2022–2025 and foundational works) to improve review transparency.

4. Add a comparative summary table highlighting methods, datasets, modalities, and performance metrics.

5. Strengthen discussion on real-world deployment challenges such as cross-region generalization and dataset bias.

6. Improve readability by simplifying dense sentences and reducing repetition in modality and limitation descriptions.

Reviewer 2 Comments :

UNIVERSAL INOVATORS

Razorpay
Invoicing and payments powered by Razorpay

Payment Receipt Transaction Reference: pay_QTHTeePsGb0ydm

This is a payment receipt for your transaction on ICDAM 2026

AMOUNT PAID **₹ 17,000.00**

ISSUED TO
palakharinkhede1@gmail.com
+919673494515

PAID ON
24 May 2026

DESCRIPTION	UNIT PRICE	QTY	AMOUNT
Service Charges	₹ 250.00	1	₹ 250.00
Registration Fee	₹ 16,750.00	1	₹ 16,750.00
	Total		**₹ 17,000.00**
	Amount Paid		₹ 17,000.00

UNIVERSAL INOVATORS

Razorpay

Invoicing and payments powered by Razorpay

Payment Receipt Transaction Reference: pay_QTHTeePsGb9xYfLm7

This is a payment receipt for your transaction on ICDAM 2026

AMOUNT PAID **₹ 17,000.00**

ISSUED TO

palakharinkhede1@gmail.com

+919673494515

PAID ON

24 May 2026

DESCRIPTION	UNIT PRICE	QTY	AMOUNT
Service Charges	₹ 250.00	1	₹ 250.00
Registration Fee	₹ 16,750.00	1	₹ 16,750.00
	Total		**₹ 17,000.00**
	Amount Paid		₹ 17,000.00

24RCO07

by Palak M

Submission date: 17-Jun-2026 12:31PM (UTC+0530)
Submission ID: 2984835942
File name: 24RCO07_thesis.pdf (410.02K)
Word count: 8498
Character count: 49116

24RCO07

ORIGINALITY REPORT

6%	5%	3%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	umpir.ump.edu.my Internet Source	<1%
2	arxiv.org Internet Source	<1%
3	Submitted to Asia e University Student Paper	<1%
4	www.mdpi.com Internet Source	<1%
5	icdam-conf.com Internet Source	<1%
6	library.imaging.org Internet Source	<1%
7	Rakhee Kumari, Sharanya Singh, Praneet Saurabh. "A systematic review of image enhancement and object detection in low-light scenarios with benchmark experiments", Applied Soft Computing, 2026 Publication	<1%
8	noah.nrw Internet Source	<1%
9	Chih-Tsung Shen, Wen-Liang Hwang. "Color image enhancement using retinex with robust envelope", 2009 16th IEEE International Conference on Image Processing (ICIP), 2009 Publication	<1%

www.aclclp.org.tw

10	Internet Source	<1 %
11	Çiftçi, Selçuk. "İNsansız Hava Araçları Vasıtasıyla Karasal Görüntüler ÜZerinden Nesne Tespiti", Sakarya Universitesi (Turkey), 2025 Publication	<1 %
12	Submitted to Malta College of Arts,Science and Technology Student Paper	<1 %
13	Manoj Kumar, Tanweer Ali, Jaume Anguera, Suman Lata Tripathi. "Emerging Technologies in AI, Computation, Communication, and Cybersecurity - Proceedings of the First International Conference on Artificial Intelligence, Computation, Communication and Network Security (AICCoNS 2025)", CRC Press, 2026 Publication	<1 %
14	interviewgemini.com Internet Source	<1 %
15	Submitted to East Delta university Student Paper	<1 %
16	Submitted to University of Reading Student Paper	<1 %
17	www.techscience.com Internet Source	<1 %
18	dspace.lpu.in:8080 Internet Source	<1 %
19	etda.libraries.psu.edu Internet Source	<1 %
20	ietresearch.onlinelibrary.wiley.com Internet Source	

		<1 %
21	www2.mdpi.com Internet Source	<1 %
22	aaltodoc.aalto.fi Internet Source	<1 %
23	customwritings.co Internet Source	<1 %
24	spectrum.library.concordia.ca Internet Source	<1 %
25	support.dce.felk.cvut.cz Internet Source	<1 %
26	"Advances in Computer Science and Ubiquitous Computing", Springer Nature, 2018 Publication	<1 %
27	bura.brunel.ac.uk Internet Source	<1 %
28	eprints.bournemouth.ac.uk Internet Source	<1 %
29	eprints.utar.edu.my Internet Source	<1 %
30	impa.usc.edu Internet Source	<1 %
31	trepo.tuni.fi Internet Source	<1 %
32	Abdussalam Elhanashi, Sergio Saponara. "Deep Learning for Object Detection and Localization", Springer Science and Business Media LLC, 2026 Publication	<1 %

33	Alshamsi, Omar. "Intelligent Monitoring and Alerting System Using Some Deep Learning Models", Rochester Institute of Technology, 2025 Publication	<1 %
34	Arvind Dagur, Sohit Agarwal, Dhirendra Kumar Shukla, Shabir Ali, Sandhya Sharma. "Artificial Intelligence and Sustainable Innovation - Volume 3", CRC Press, 2026 Publication	<1 %
35	dokumen.pub Internet Source	<1 %
36	dspace.daffodilvarsity.edu.bd:8080 Internet Source	<1 %
37	dspace.library.uvic.ca Internet Source	<1 %
38	epubl.ktu.edu Internet Source	<1 %
39	iris.unito.it Internet Source	<1 %
40	ore.exeter.ac.uk Internet Source	<1 %
41	tudr.thapar.edu:8080 Internet Source	<1 %
42	Baoliang Chen, Lingyu Zhu, Hanwei Zhu, Wenhan Yang, Linqi Song, Shiqi Wang. "Gap-Closing Matters: Perceptual Quality Evaluation and Optimization of Low-Light Image Enhancement", IEEE Transactions on Multimedia, 2024 Publication	<1 %

Exclude quotes	Off	Exclude matches	Off
Exclude bibliography	Off		

Palak M

24RCO07

Quick Submit

Quick Submit

Delhi Technological University

Document Details

Submission ID

trn:oid:::1:3596287738

Submission Date

Jun 17, 2026, 12:29 PM GMT+5:30

Download Date

Jun 17, 2026, 12:40 PM GMT+5:30

File Name

24RCO07_thesis.pdf

File Size

410.0 KB

40 Pages

8,498 Words

49,116 Characters

*% detected as AI

AI detection includes the possibility of false positives. Although some text in this submission is likely AI generated, scores below the 20% threshold are not surfaced because they have a higher likelihood of false positives.

Caution: Review required.

It is essential to understand the limitations of AI detection before making decisions about a student's work. We encourage you to learn more about Turnitin's AI detection capabilities before using the tool.