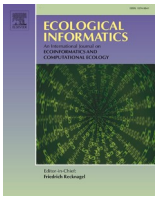
Ecological Informatics 82 (2024) 102705

Contents lists available at ScienceDirect

Ecological Informatics

journal homepage: www.elsevier.com/locate/ecolinf



Adaptive image processing embedding to make the ecological tasks of deep learning more robust on camera traps images

Zihe Yang [a,b,c], Ye Tian [a,b,c], Junguo Zhang [a,b,c,*]

[a] School of Technology, Beijing Forestry University, Beijing 100083, China
[b] Key Lab of State Forestry and Grassland Administration on Forestry Equipment and Automation, Beijing 100083, China
[c] Research Center for Biodiversity Intelligent Monitoring, Beijing Forestry University, Beijing 100083, China

ARTICLE INFO

Keywords:
Adaptive image processing
Camera traps
Deep learning
Ecological tasks

ABSTRACT

Camera traps serve as a valuable tool for wildlife monitoring, generating a vast collection of images for ecologists to conduct ecological investigations, such as species identification and population estimation. However, the sheer volume of images poses a challenge, and the integration of deep learning into automated ecological investigation tasks remains complex, particularly when dealing with low-quality images in long-term monitoring programs. Existing approaches often struggle to strike a balance between image enhancement and deep learning for ecological tasks, thereby overlooking crucial information contained within low-quality images. This research introduces a pioneering adaptive image processing module (AIP) that seamlessly incorporates image processing into camera trap ecological tasks, elevating the performance of wildlife monitoring activities. Specifically, a differentiable image processing (DIP) module is presented to enhance low-quality images, with its parameters predicted by a Non-local based parameter predictor (NLPP). Additionally, an end-to-end approach based on hybrid data containing both original and synthetic data is proposed, encompassing adaptive image processing methods and downstream tasks for camera traps, adaptable to various scenarios. This approach effectively reduces the manual labor and time required for professional image processing. When applied to real-world camera trap images and synthetic image datasets, our method achieves an accuracy of 92.26% and 86.65% in classifying wildlife, respectively, demonstrating its robustness. By outperforming alternative methods under harsh conditions, the application of the adaptive image processing module instills greater confidence in deep learning applications within complex environments.

1. Introduction

Long-term tracking observations are vital for wildlife research, requiring sufficient and reliable data. In recent years, camera traps have emerged as a popular, non-intrusive, and cost-effective method for collecting wildlife observation data, playing a significant role in ecological conservation efforts (Delisle et al., 2023). These cameras facilitate the modeling of population size, distribution, and environmental interactions, enabling the estimation of animal abundance, prediction of animal movements (Li et al., 2022), assessment of species richness, and comprehension of animal behavior (Nazir and Kaleem, 2021). Despite the widespread use of camera trap networks (Niedballa et al., 2016), which can amass billions of images (Thau et al., 2019), the conversion of raw images into actionable information typically relies on manual labeling of each image (Schaus et al., 2020; Swanson et al., 2015). The labor-intensive process of manual labeling poses a significant challenge in camera trap surveys, limiting their application in large-scale research endeavors. Hence, the implementation of automated methods becomes necessary to reduce manual costs and facilitate the widespread utilization of camera traps.

Traditional machine learning techniques have been extensively employed for automatic image recognition and have demonstrated their significance in the field (Petso et al., 2022). In wildlife recognition, these methods involve manually applying filters to preprocess images and extract features. Subsequently, these features are utilized to train and evaluate a classifier. However, the reliance on manual filters renders these methods susceptible to external interference, making them effective only in controlled environments with simple backgrounds, such as laboratories (Midori and Alejandro, 2020), while proving challenging to implement in complex environments. Moreover, such approaches often

* Corresponding author at: School of Technology, Beijing Forestry University, Beijing 100083, China.
E-mail address: zhangjunguo@bjfu.edu.cn (J. Zhang).

https://doi.org/10.1016/j.ecoinf.2024.102705
Received 2 June 2023; Received in revised form 27 May 2024; Accepted 28 June 2024
Available online 29 June 2024

necessitate manual annotation of objects of interest, impeding their scalability in large-scale wildlife datasets (Patrick et al., 2019).

Deep learning, on the other hand, replaces manual filters with feature extractors trained on extensive datasets, mitigating the influence of external noise. Leveraging this advantage, deep learning has achieved remarkable advancements in various domains, including machine translation (Vaswani et al., 2017), speech recognition (Dong et al., 2018), and computer vision (He et al., 2016). Deep Convolutional Neural Networks (DCNN) represent a classic deep learning model extensively utilized for image processing (Krizhevsky et al., 2017). Extensive research has demonstrated that DCNN can effectively extract ecological information from camera trap images, including species labels, counts, and behaviors (Borowiec et al., 2022; Frank and Volker, 2021; Manuel et al., 2021). While deep learning has proven successful in extracting information from camera trap images, the inherent complexity of the natural environment continues to present significant challenges (Alexander et al., 2017). For example:

(1) Exposure shift is a common issue encountered in camera trap images captured at various locations, leading to significant variations in exposure levels. These variations can be observed in diverse environments such as woodlands, grasslands, and mountainous areas, among others.
(2) Illumination variation is a significant factor that affects camera trap observations under different conditions, particularly during periods of intense sunlight.
(3) Appearance change is due to different light supplementation strategies of the camera trap in nighttime environments and may simultaneously lead to changes in animal identification features.
(4) Image blurring may result in masking of image features. Focusing defects and lens dust are direct causes of blurred image formation.

The natural environment presents various visual challenges for camera trap images, resulting in low-quality images that lack information. Even human annotators find it difficult to interpret such images. Some researchers have attempted to enhance the performance of deep learning models on low-quality images by incorporating attention mechanisms (Enlin et al., 2023; Xianchong et al., 2023). While these methods have shown promising results, they often struggle to uncover hidden information within images captured in harsh environments. In contrast, other researchers have proposed unsupervised domain adaptation (UDA) as a potential solution (Chen et al., 2018; Hnewa and Radha, 2021). However, UDA's effectiveness in low-light object detection is limited due to significant disparities between normal and weak lighting conditions. To address these challenges, image processing aims to enhance images by accentuating overall or local features, amplifying differences between object features, and suppressing irrelevant details (Qi et al., 2022). In this study, we will apply the Differential Image Processing (DIP) module consisting of five mapping functions to provide a comprehensive solution for the diverse array of low-quality images encountered. Fig. 1 illustrates examples of images before and after processing, highlighting the potential for recovering additional information when images are appropriately enhanced under adverse conditions.

Undoubtedly, image processing proves effective in handling low-quality images, yet it represents only one aspect of the challenge. The difficulty lies in determining the appropriate parameter, which can hinder the overall performance and scalability of the process. Traditional image processing algorithms typically require parameter settings based on prior experience, such as the histogram equalization algorithm (Lin et al., 2015), partial differential equation algorithm, and Retinex algorithm (Mading et al., 2018). However, enhancing image quality does not necessarily guarantee improved detection performance. To achieve adaptive image enhancement, some researchers have employed small Convolutional Neural Networks (CNNs) to flexibly predict parameter (Hu et al., 2018; Zeng et al., 2020). Nevertheless, CNNs excel at extracting local features but have limitations in capturing global feature representations, which are crucial for predicting the threshold of image processing algorithms (Wang et al., 2018). In this study, we propose a compact network incorporating a non-local layer capable of extracting long-distance features to predict parameters for the image processing module. Additionally, we introduce a joint training approach for image adaptive enhancement modules and deep learning models to ensure that the image processing results contribute to improved performance of the deep learning model. The proposed method demonstrates superior performance across a diverse range of environments. it enables expedited transitions from field surveys to reporting, even in challenging circumstances, thereby facilitating efficient data collection and analysis on a larger scale.

The contributions of this study can be summarized as follows:

- A DIP module is proposed to tackle the challenges presented by complex mixed low-quality environments.
- A non-local layer is incorporated into the parameter predictor to extract long-distance features and enhance the prediction accuracy.

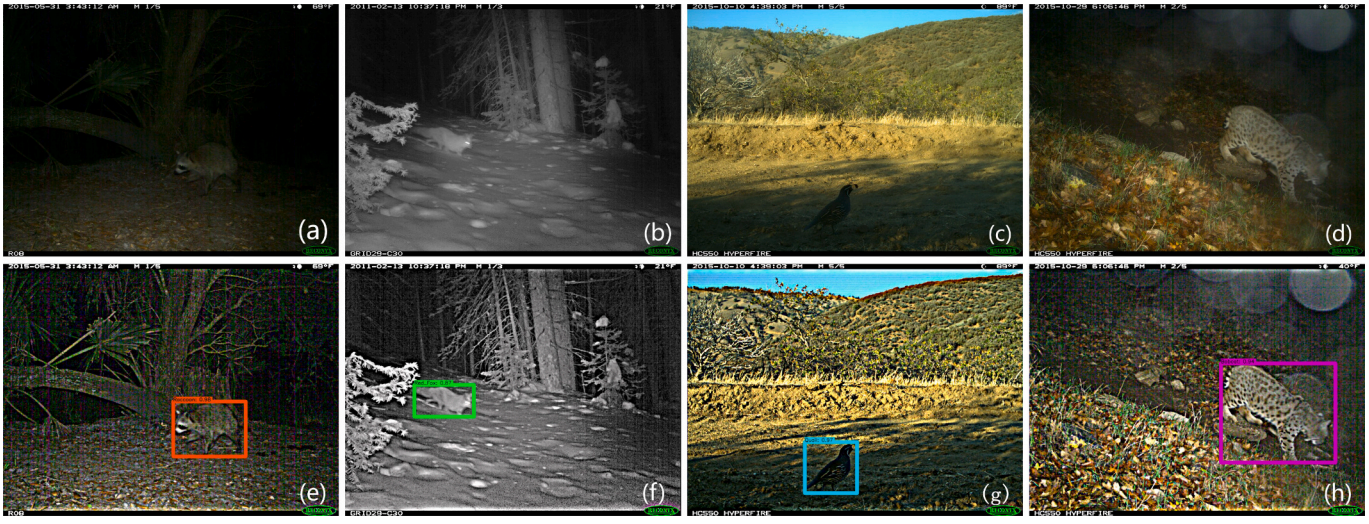


Fig. 1. Example of object detection under adverse conditions. Figures (a)-(d) show samples of low-quality images, while Figures (e) - (h) show the effect of the adaptive image processing module (AIP).

- A joint training method is proposed to ensure that the image processing effect aligns with enhancing the performance of the deep learning model.
- The proposed method can be seamlessly integrated into pre-existing deep learning models, enhancing their resilience and performance in low-quality environments.

1.1. Deep learning

The term "deep learning" refers to the utilization of neural networks, a statistical model for data representation in solving complex problems (LeCun et al., 2015; Christopher and Hugh, 2024). These neural networks are constructed through training, involving a large set of inputs and labeled outputs. Comprising layered nonlinear transformations, a neural network contains numerous adjustable parameters. Training a neural network requires multiple iterations, during which the network may produce incorrect results. The discrepancy between the current and expected outputs is calculated as loss values. Optimization algorithms such as Stochastic gradient descent (SGD) (Herbert and Sutton, 1951), Adaptive Gradient (AdaGrad) (Duchi et al., 2011), Root Mean Square Prop (RMSProp) (Tieleman and Hinton, 2012), and Adam (Kingma and Ba, 2014) are then employed to assess each parameter's contribution to the loss values and adjust them to minimize losses. This iterative optimization process is performed for each image in the dataset, gradually refining the parameters to minimize losses (Norouzzadeh et al., 2021).

For computer vision tasks in deep learning, CNNs are commonly employed (Krizhevsky et al., 2017). Convolutional layers form the fundamental structure of CNNs, enabling the learning of feature maps that capture spatial patterns in an image (Christopher and Hugh, 2024; LeCun et al., 2015). Another crucial component in CNNs is the max-pooling layer, which reduces computational complexity and enhances robustness by partitioning the feature maps into regions and selecting the maximum values within each region (Christopher and Hugh, 2024; LeCun et al., 2015). Several network architectures have gained standardization due to their significant performance milestones, serving as the basis for numerous related studies. Notable examples of such architectures include AlexNet (Krizhevsky et al., 2017), VGG19 (Simonyan and Zisserman, 2015), GoogLeNet/InceptionNet (Christian and Liu, 2015), and ResNet (He et al., 2016), among others.

1.2. Image classification

Image classification involves the application of computer algorithms to quantitatively analyze images and assign them to predefined classes, replacing the need for human visual interpretation. In the context of camera trap images, image classification algorithms are extensively employed, providing a probability estimation of the input image belonging to different species (Mohammad et al., 2018; Tabak et al., 2018). While training an image classification model with image-level labels is relatively straightforward, this model structure presents several recognized limitations:

(1) Typically, supervised image classification models can only predict the most probable label for a given input image, which restricts their ability to handle images containing multiple species simultaneously.
(2) Due to their inherent predictive nature, image classification algorithms are primarily suited for addressing classification tasks in downstream ecological applications. When applied to non-classification problems such as counting, their performance may be suboptimal (Mohammad et al., 2018).
(3) During the training phase, supervised image classification models are adept at discerning the most salient regions within an image that correspond to the designated labels. Nonetheless, there is a tendency for these models to concentrate on the background elements, which may not be pertinent to the subjects of interest, such as animals under scrutiny (Miao et al., 2019). As a result, their transferability is limited when faced with images from new locations where the background differs (Tabak et al., 2018).

1.3. Object detection

Object detection is a fundamental task in computer vision for accurately locating and recognizing instances of predefined object categories in an image. It is particularly well-suited for camera trap imagery, which often contains multiple species categories. There are two main approaches: one-stage and two-stage methods (Zhao et al., 2019).

The two-stage object detection approach involves generating a large number of region proposals (around 1 k–2 k) from the image, which are potential regions of interest. These proposals are then used to predict the classification and localization of the targets, resulting in object detection (Girshick, 2015; Girshick et al., 2014; Ren et al., 2015). However, the complex nature and time-consuming steps of the two-stage method make it challenging to meet real-time computational requirements.

On the other hand, the one-stage object detection method predicts the position and class of objects in the input image simultaneously, eliminating the need for an explicit proposal stage. Notable examples of one-stage methods include various improved versions of YOLO (You Only Look Once) (Bochkovskiy et al., 2020; Redmon et al., 2016; Redmon and Farhadi, 2017, 2018) and SSD (Single Shot Detection) (Wei et al., 2016) models. Typically, one-stage methods offer faster inference speeds compared to two-stage methods. In this study, the widely used one-stage detector YOLOv3 (Redmon and Farhadi, 2018) was selected as the baseline detector for ecological tasks and to enhance its performance in processing low-quality camera trap images.

2. Materials and methods

The following subsections detail the steps of our workflow when studying wildlife images from camera traps.

2.1. Datasets

This study utilized camera trap images obtained from the North American Camera Trap Images (NACTI) dataset from 2010 to 2016 (Michael et al., 2018). The dataset comprised trips conducted across five locations in the United States (California, Colorado, Florida, South Carolina, and Texas) and one location in Canada (Saskatchewan). These sites encompassed diverse ecosystems, including coniferous forests, deciduous forests, wetlands, and grasslands. However, variations in lighting conditions and differences in ecosystems resulted in challenges such as color loss, texture degradation, and reduced details of the captured wildlife. Furthermore, the presence of dust on the camera lens introduced blurriness to the images. Refer to Fig. 2 for visual illustration.

To overcome the labeling errors in the extended NACTI dataset and the computational resource constraints associated with processing a large number of images (3.7 million), several measures were taken in this study. First, species not relevant to the ecological survey, such as categories like livestock and vehicles, were removed from the dataset. In addition, fuzzy categories, such as birds, were also eliminated. Subsequently, a subset of the dataset, which included 11,856 images covering 13 different wildlife species, was randomly selected for further analysis. The details of species is shown in Table 1. To train the model, a random sample comprising 70% of the images for each species was selected, while the remaining 30% were designated for testing. This allocation resulted in 8302 images for training and 3554 images for testing. The test datasets encompassed all the ecological environments previously mentioned. Refer to Table 2 for further details.

Fig. 2. Examples of ecosystems and low-quality images. (a) Conifer Forests. (b) Deciduous Forests. (c) Grassland. (d) Wetlands. (e) Complex Light. (f) White Light. (g) Infra-red. (h) Lens Dust. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of the datasets. The images are split into an 7:3 ratio for the training and test sets.

Species	Training Images	Test Images	Species	Training Images	Test Images
Cougar	1542	661	Jackrabbit	404	173
Black bear	1435	615	Quail	261	112
Bobcat	1360	583	Red squirrel	201	86
Wild boar	1070	459	Red fox	170	73
Coyote	1024	439	Opossum	61	26
Raccoon	833	357	Marten	50	21
Wild turkey	567	243			

Table 2
Summary of the ecological environment in test datasets. Number and frequency of each species in all environmental types.

Species	Conifer forests	Deciduous forests	Grassland	Wetlands
Cougar	2/0.31%	563/88.86%	72/11.34%	0/0.00%
Black bear	100/17.12%	466/79.79%	18/3.08%	0/0.00%
Bobcat	3/0.52%	449/77.28%	105/18.07%	24/4.13%
Wild boar	0/0.00%	238/56.94%	77/18.42%	103/24.64%
Coyote	156/40.73%	146/38.12%	29/7.57%	52/13.58%
Raccoon	0/0.00%	11/3.37%	40/12.27%	275/84.36%
Wild turkey	18/10.06%	50/27.93%	9/5.03%	102/56.98%
Jackrabbit	0/0.00%	91/52.60%	82/47.40%	0/0.00%
Quail	0/0.00%	63/87.50%	9/12.50%	0/0.00%
Red squirrel	86/100%	0/0.00%	0/0.00%	0/0.00%
Red fox	73/100%	0/0.00%	0/0.00%	0/0.00%
Opossum	0/0.00%	0/0.00%	8/38.10%	13/61.90%
Marten	21/100%	0/0.00%	0/0.00%	0/0.00%
Total	459	2077	449	569

2.2. Proposed method

In this section, we describe the AIP module and downstream ecological tasks introduced in this study, as illustrated in Fig. 3. For the AIP module, we detail a non-local-based parameter predictor (NLPP) and a DIP module included. For the downstream task, we provide an end-to-end joint training scheme, ensuring that the adaptive image processing module learns to enhance image quality in a task-driven manner by effectively handling the downstream losses.

2.2.1. DIP module

The proposed DIP module is designed to simulate the techniques of professional photographers, adjusting lighting, color, and sharpening. These three steps contain 5 image mapping functions listed in Table 3, in which $P_i = (r_i, g_i, b_i)$and $P_o = (r_o, g_o, b_o)$ are the input and output RGB value.

The process of light adjustment involves gamma and contrast adjustments, aiming to modify the image brightness and unveil hidden information. The gamma mapping function can be achieved through straightforward multiplication and power transformation operations. Similarly, the contrast mapping function establishes a linear interpolation between the original image and a grayscale image with constant intensity, as illustrated in Table 3. The enhanced function $En(P_i)$ is defined as follows:

$$Lum(P_i) = 0.27r_i + 0.67g_i + 0.06b_i, \tag{1}$$

$$En(P_i) = P_i \times \frac{\frac{1}{2}(1 - cos(\pi \times Lum(P_i)))}{Lum(P_i)} \tag{2}$$

where $P_i = (r_i, g_i, b_i)$ are the input RGB value.

The color adjustment process encompasses tone and white balance (WB) adjustments, aimed at correcting low-quality lighting conditions in the natural environment and mitigating the impact of domain offset. White balance operates by applying a pixel-by-pixel mapping to the images. Tone adjustment, on the other hand, is achieved by constructing piecewise linear curve functions that maintain monotonicity. Lparameters used to represent this curve are denoted as $\{t_0, t_1, t_2, \cdots, t_{L-1}\}$ and the specific points on the tone curve are indicated as $(k/L, T_k/T_L)$where $T_k = \sum_{l=0}^{k-1} t_l$ varies. To enhance the input image, differentiable $\{t_0, t_1, t_2, \cdots, t_{L-1}\}$ parameters are predicted as follows:

$$P_0 = \frac{1}{T_L}\sum_{j=0}^{L-1} \text{clip}(L \cdot P_i - j, 0, 1)t_k, \tag{3}$$

where, the value of L is usually 8, which is sufficient to fit a typical color curve.

The sharpening adjustment applies sharpening mask technology to enhance image details. The process can be described in Table 3, where

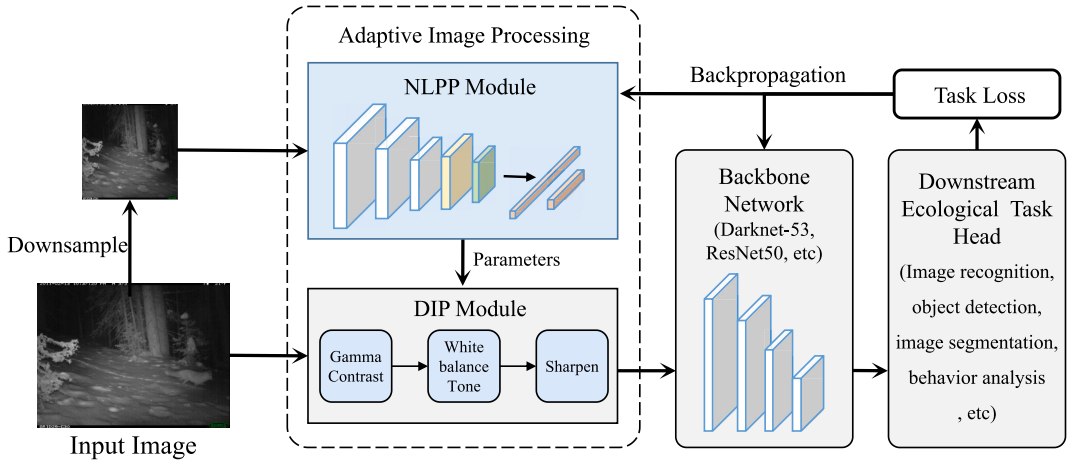


Fig. 3. Workflow diagram for integrating adaptive image processing into various deep neural network-based ecological tasks.

Table 3
The mapping functions of DIP module, including the enhanced function $En(\cdot)$ and the Gaussian mapping $Gau(\cdot)$.

Filter	Parameters and Defined	Mapping Function
Gamma	G:gamma value	$P_o = P_i^G$
Contrast	α:contrast value	$P_o = \alpha \cdot En(P_i) + (1-\alpha) \cdot P_i$
White balance	W_r, W_g, W_b:factors	$P_o = (W_r r_i, W_g g_i, W_b b_i)$
Tone	t_i:tone parameter	$P_o = (L_{t_r}(r_i), L_{t_g}(g_i), L_{t_b}(b_i))$
Sharpen	λ:positive scaling factor	$P_o = P_i + \lambda \cdot (P_i - Gau(P_i))$

$Gau(\cdot)$ denotes a Gaussian mapping, and λ is a differentiable parameter. By adjusting the parameter λ, the sharpening process achieves improved image processing results.

2.2.2. NLPP module

In the image signal processing pipeline, image enhancement typically involves the utilization of several fine-tunable mappings, with their hyperparameters adjusted by engineers based on their experience (Mosleh et al., 2020). However, this process of parameter tuning for different scenarios is time-consuming. To overcome this limitation, a recommendation is made to employ a compact non-local network instead of relying solely on experienced engineers to estimate hyperparameters. This approach offers a faster and more effective solution. When predicting parameters for DIP, the NLPP primarily focuses on capturing the large-scale environmental conditions depicted in the image rather than emphasizing high-definition details. Therefore, downsampling the input images into the size of 256 × 256 using bilinear interpolation is sufficient for completing the parameter estimation process. The NLPP network, as shown in Table 4, comprises seven layers and encompasses a mere 150 K parameters. To extract environmental features effectively, a non-local layer (Wang et al., 2018) is incorporated within the NLPP network to capture long-term dependencies in the images. The NLPP module is trained to generate more suitable hyperparameters for the DIP module.

Table 4
Structure of the NLPP network model.

Input Size	Type	Kernel Size	Stride
$256^2 \times 3$	Convolutional Layer	3 × 3	2
$128^2 \times 16$	Convolutional Layer	3 × 3	2
$64^2 \times 32$	Convolutional Layer	3 × 3	2
$32^2 \times 32$	Non-local Layer	–	–
$32^2 \times 32$	Max Pooling Layer	4 × 4	4
$4^2 \times 32$	Fully Connected Layer	–	–
$4^2 \times 64$	Fully Connected Layer	–	–

2.2.3. Downstream ecological tasks

In this paper, OD is chosen as a downstream task, which is combined with an AIP module for a wide range of ecological studies, including animal recognition and counting. The YOLOv3 one-stage detector is chosen as the wildlife object detection network due to its widespread recognition and successful implementation in various practical applications (Redmon and Farhadi, 2018). YOLOv3 enhances the accuracy of detecting small objects by predicting multi-scale feature maps, making it particularly effective for camera trap images. To integrate the proposed adaptive image processing module, it is seamlessly inserted into the YOLOv3 method, as depicted in Fig. 4.

3. Experiments

The effectiveness of the proposed method was evaluated in both real-world and synthetic scenes.

3.1. Implementation details

The training protocol outlined by (Redmon and Farhadi, 2018) was followed during the training process. The YOLOv3 model served as the baseline for all experiments. To expand the training dataset, various data augmentation techniques were employed, including random adjustments to image size, image flipping, cropping, and conversion. The AIP-OD model was trained using the Adam optimizer (Kingma and Ba, 2014). The training process consisted of 150 epochs with a batch size of 4. The learning rate initially started at $10^{-4}$and gradually decreased to over the course of training. The experiments were conducted using Tensorflow 1.14 and executed on a GTX2080Ti GPU.

3.2. Evaluating indicator

The evaluation of models commonly utilizes Average Precision (AP) as a mainstream performance metric. AP represents the area under the precision-recall curve, denoted as P (precision) and R (recall) at different thresholds.

The mean Average Precision (mAP) is derived by calculating the AP for each wildlife species. The mAP is defined as follows:

$$mAP = \frac{\sum_{i=1}^{C} AP_i}{C} \tag{4}$$

Here, C represents the number of wildlife species. A higher mAP value of the object detector indicates better performance and vice versa.

To provide a more precise assessment of the model's performance under specific conditions, this research incorporates accuracy (Acc) as

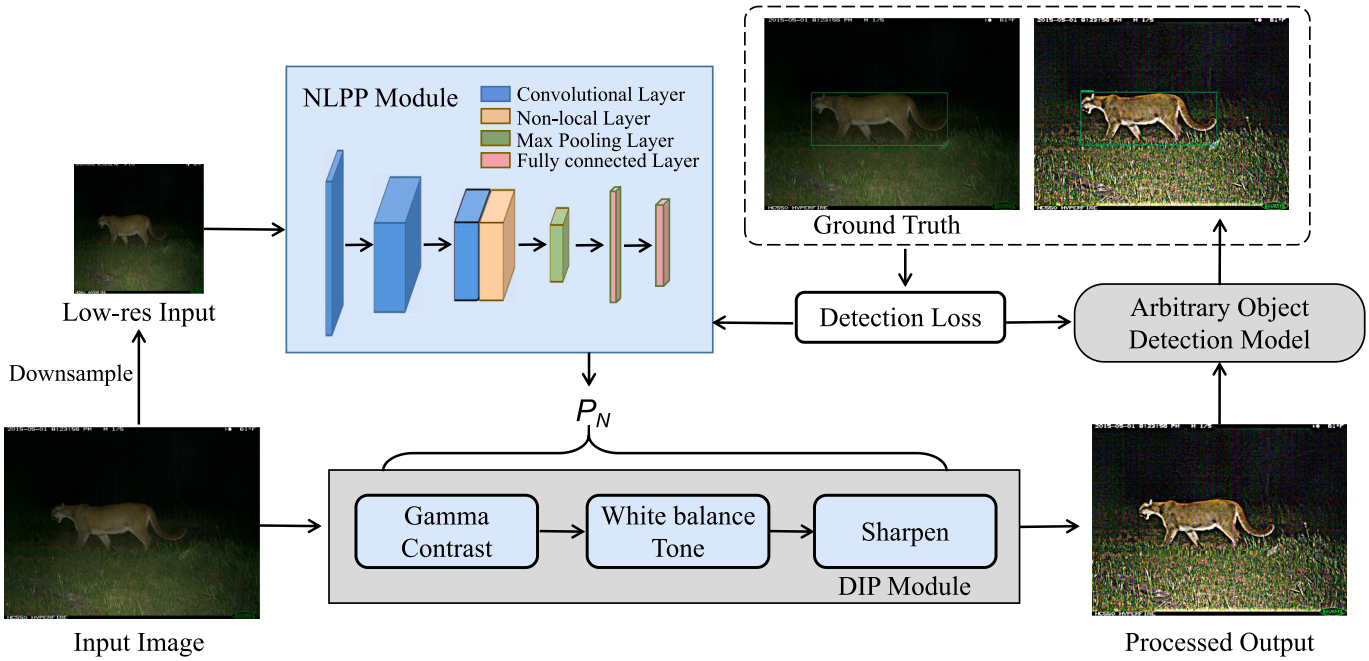


Fig. 4. A sketch of wildlife object detection seamlessly inserted an adaptive image processing module(AIP-OD).

an additional evaluation metric in certain cases. The accuracy rate is defined as follows:

$$Acc = \frac{TP}{TP + FP + FN} \tag{5}$$

where *TP*, *FP* and *FN* represents the number of true positives, false positives, and false negatives, respectively.

3.3. Construct synthetic images datasets

Low-quality datasets for object detection under real-world conditions are often limited in availability, and the data size is typically insufficient to train a stable CNN-based detector. To address this challenge, the researchers developed a synthetic image dataset (SI-Dataset) using light transformation and lens blur techniques to replicate low-quality conditions. Given the computational complexity involved, the generation of low-quality images was conducted prior to the training phase. Specifically, effects such as lack of focus and lens dust were simulated using an atmospheric scattering model, as outlined below:

$$I(x) = J(x)e^{-\beta}d(x) + A\left(1 - e^{-\beta}d(x)\right) \tag{6}$$

The blurred image, denoted as $I(x)$, is generated using an atmospheric scattering model, while the scene radiance (real-world image) is represented by $J(x)$. The global atmospheric light is denoted by A, and the lens blur effect is simulated by adjusting the value of β. The scene depth, $d(x)$, is also considered in the generation process. Each image has a 50% chance of being added with the lens blur effect, as visually illustrated in Fig. 5.

Furthermore, the light transformation step simulates exposure variations observed in different ecological regions. This is achieved through the synthesis of images under adverse lighting conditions using a transformation function, denoted as $f(x) = x^{\gamma}$. Here, x represents the intensity of the input pixels, and the value of γ is randomly sampled from a uniform distribution within the range $[0.3, 0.7] \cup [1.5, 5.5]$, as depicted in Fig. 5. The real-world datasets exhibit variations in γ, with 10 different values considered. Consequently, the SI-Dataset is expanded to 10 times the original quantity, resulting in a total of 118,560 images. During testing, all real-world test images are randomly substituted with their corresponding counterparts from the SI-Dataset.

Fig. 5. Examples of the synthetic images datasets. (a) $\gamma = 0.3$. (b)$\gamma = 0.4$. (c)$\gamma = 0.5$and blur. (d)$\gamma = 0.6$and blur. (e)$\gamma = 0.7$and blur. (f)$\gamma = 1.5$and blur. (g)$\gamma = 2.5$and blur. (h)$\gamma = 3.5$. (i)$\gamma = 4.5$. (j)$\gamma = 5.5$.

3.4. A hybrid data training scheme

A hybrid data training scheme containing both original and synthetic data was considered to address the scarcity of low-quality images in real-world environments. During training, each image was subjected to a 2/3 probability of being randomly replaced with the corresponding image from the SI-Dataset. This approach ensured that the network received a diverse range of training examples that encompassed both real-world and synthetic data. By incorporating hybrid data, all modules within the AIP-OD model underwent end-to-end loss detection training, allowing for the optimization of image enhancement specifically for the task at hand. In this training scheme, the NLPP modules were weakly supervised by the detection loss in the absence of ground truth images. During testing, the SI-Test-Dataset was constructed by replacing the real-world dataset with synthetic data. This dataset substitution ensured that the AIP-OD model was evaluated on a comprehensive range of image content, enabling it to adaptively process images based on their specific characteristics. The combination of the hybrid data training mode and the SI-Test-Dataset contributed to the overall effectiveness of the AIP-OD model in handling images with varying quality and environmental conditions.

3.5. Experimental results

To assess the efficacy of AIP-OD, a comparative analysis was conducted with baseline YOLOv3, ZeroDCE (Guo et al., 2020), and IA-YOLO (Liu et al., 2022) using both real-world datasets and the SI-Test-Dataset. In addition, we also tested the performance of AIP on the latest YOLOv8s, which has one-fifth the parameters of the YOLOv3. The mAP results are presented in Table 5. This evaluation aimed to demonstrate the superior performance of AIP-OD in handling challenging environmental conditions and low-quality images.

In subsequent experiments, the baseline model in AIP-OD was YOLOv3. Fig. 6 illustrates the detection accuracy of AIP-OD predictions for individual species, encompassing both the real-world dataset and the SI-Test-Dataset. Among the species examined, the top three with the highest accuracy were jackrabbit, bobcat, and cougar. Conversely, the species with the lowest accuracy were red fox, opossum, and marten. Notably, there were substantial variations in accuracy across different species when testing with both normal and synthetic data. Opossum and marten exhibited a decrease in accuracy of over 14% compared to other species.

Table 6 demonstrates the performance of our model across diverse ecological environments. In this scenario, the utilization of mAP as an evaluation indicator is impractical given the variability in species composition across different ecological settings. Instead, the evaluation metric employed is Acc, which serves as a reliable measure of the influence of the environment on model performance.

3.6. Ablation experiment

Adverse-light conditions were synthesized using the transformation $f(x) = x^{\gamma}$, where the value of $\gamma$was varied to represent different intensities of synthetic lighting. To assess the robustness of our method across a range of synthetic strengths, six distinct strengths were identified based on the difference between γ and 1. A strength of 0 corresponds to a value of 1, indicating that the test image originates from the real-world dataset. For strengths 1 to 5, half of the light in the test image was enhanced while the other half was weakened. For instance, a strength of 1 implies that the light-enhanced image possesses a value of 0.7, whereas the light-weakened image has a γ value of 1.5. The performance of our model at different synthetic strengths is depicted in Fig. 7. It should be noted that AIP-OD(Real-world) and YOLOv3(Real-world) were exclusively trained on real-world datasets, without incorporating hybrid data.

Table 5
Performance Comparison of real-world datasets and SI-Test-Dataset.

Methods	mAP (%)	
	Real-world datasets	SI-Test-Dataset
YOLOv3	88.61	83.40
ZeroDCE (Guo et al., 2020)	–	57.70
IA-YOLO (Liu et al., 2022)	90.91	86.57
AIP-OD(YOLOv3)(our)	92.26	86.65
YOLOv8s	78.60	68.83
AIP-OD(YOLOv8s)(our)	80.07	71.10

To verify the robustness of our method under blurred images, SI-Test-Dataset without blur is built, called SI-Test-Dataset (no blur). The differences in accuracy between the two datasets serve as indicators of the models' robustness. Table 7 provides an overview of the performance of the four models on both datasets, highlighting the variations in their performance.

In the real-world scenarios, the model demonstrates robust adaptability in low-quality conditions. Fig. 8 (a)-(b) and (c)-(d) illustrate the performance of the model when different methods of fill light are employed, namely white light and infrared. Similarly, Fig. 8 (e)-(f) and (g)-(h) show the model's performance under complex lighting conditions and the presence of lens dust.

The validation of the SI-Dataset using the model for images with varying synthetic strengths and blurring interference resulted in favorable detection performance, as depicted in Fig. 9. It is worth noting that images representing the same ecological environment can be simulated with diverse synthetic strengths to emulate different light conditions. As seen in the Fig. 9, the probability of snow is greater in coniferous forest environments, while the probability of green vegetation is greater in wetland environments. The backgrounds of deciduous forest environments can be made more complex by falling branches and leaves, compared to the simpler grassland environments.

4. Discussion

Camera traps are widely utilized as a popular, non-invasive, and cost-effective method for gathering observational data on wildlife. However, it is inevitable that low-quality images are captured by these cameras in the intricate and unpredictable environments of the wild. In the domains of agriculture and fisheries, recent research has predominantly focused on enhancing the performance of deep learning approaches for low-quality images through the incorporation of attention mechanisms (Enlin et al., 2023; Xianchong et al., 2023; Zan et al., 2022). However, most of these studies have relied on manually collected images, which inherently exhibit environmental selectivity. In contrast, camera trap images are passively captured in any environment where wild animals may appear. Consequently, camera trap images face a more complex and challenging environment, which can obscure crucial features of the wildlife. Researchers have begun to notice the challenges in camera trap images due to complex environments and have tested different base models for their performance in various ecological tasks (Qi et al., 2024; Sajid and Mohammad, 2024). In this study, a novel adaptive image processing module is proposed, which seamlessly integrates into various deep neural network-based ecological tasks in camera trap projects. The aim is to recover potentially obscured information about the objects present in the images.

To enable adaptive image enhancement, a small CNN is employed to flexibly predict hyperparameters for image enhancement. Previous work by Guo et al. (2020) utilized unsupervised training to fine-tune the CNN-based parameter predictor known as ZeroDCE. However, the results demonstrated that it did not enhance the accuracy of the baseline YOLOv3 model in complex environments. This can be attributed to the fact that unsupervised image quality improvement, such as ZeroDCE (Guo et al., 2020), may not provide an absolute advantage in terms of detection performance. In contrast, our proposed method, along with IA-

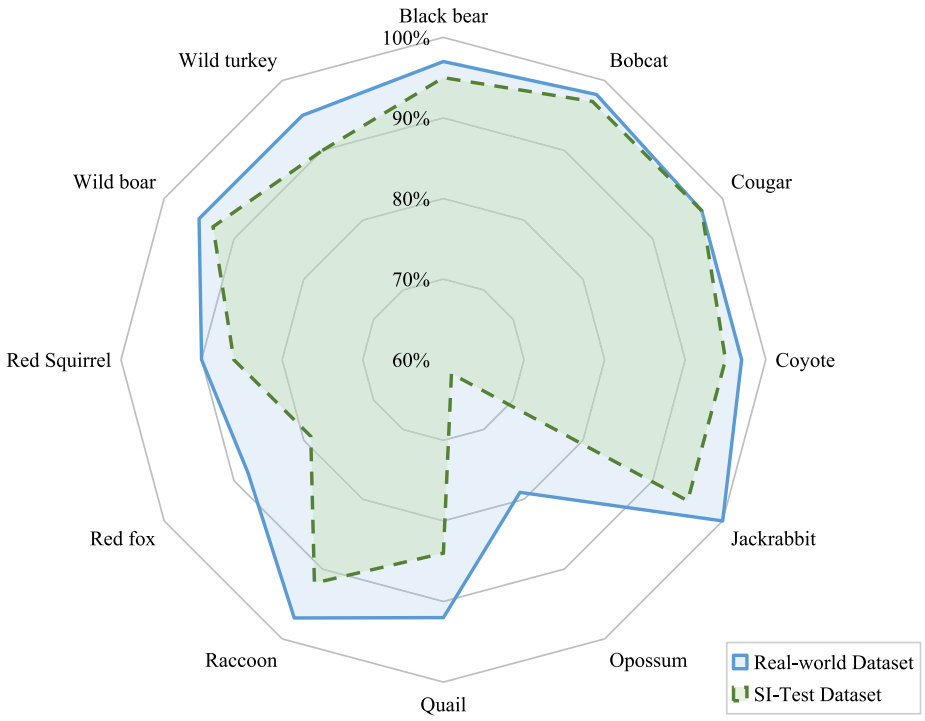


Fig. 6. Detection accuracy on real-world dataset and SI-Test Dataset.

Table 6
Statistics for the real-world datasets and SI-Test-Dataset in different ecological environments.

Ecological Environment Type	Acc(%)	
	real-world datasets	SI-Test-Dataset
Conifer forests	83.03	77.74
Deciduous forests	94.30	90.08
Grassland	93.77	90.16
Wetlands	92.44	86.51

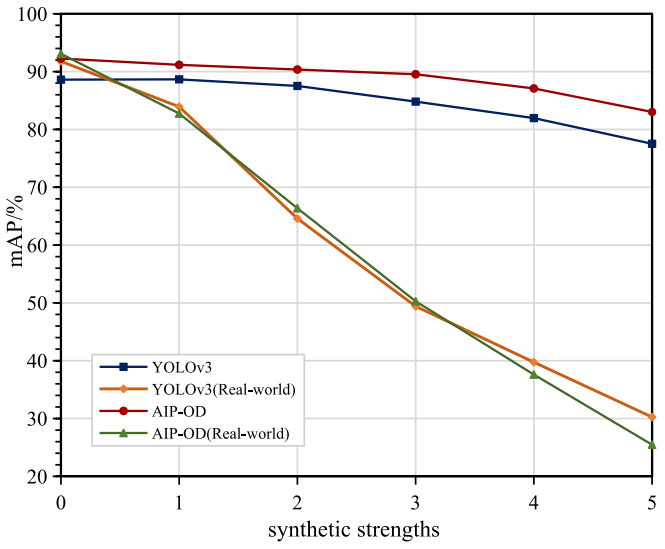


Fig. 7. Performance results of our method at various synthetic strengths.

YOLO (Liu et al., 2022), utilizes joint training to fine-tune the parameter predictor. The superior results of both methods compared to the baseline YOLOv3 model underscore the necessity of joint training. Furthermore,

Table 7
Performance on SI-Test-Dataset and SI-Test-Dataset (no blur).

Methods	mAP (%)	
	SI-Test-Dataset	SI-Test-Dataset(no blur)
YOLOv3(Real-world)	54.23	56.03
YOLOv3	83.40	82.47
AIP-OD(Real-world)	53.45	57.99
AIP-OD	86.65	86.82

our method exhibits improved accuracy by 1.35% and 0.08% in comparison to IA-YOLO on both datasets, respectively. This serves as evidence that the non-local layers incorporated into the CNN effectively extract long-distance features that facilitate parameter prediction. In additional baseline experiments, the model with added AIP module improved by 1.47% and 2.27% compared to the YOLOv8s baseline model on both datasets, respectively. This proves that our proposed AIP module can be flexibly applied to different baseline models and improve its performance in low-quality images.

The long tail effect is a common phenomenon observed in wildlife datasets, wherein species with a larger amount of data tend to exhibit higher accuracy, as depicted in Fig. 6, for example, cougar and bobcat. However, even with a smaller sample size, identification of rabbits still demonstrate high accuracy rates. This can be attributed to the fact that rabbits typically occupy a single environment and are predominantly active during the night. Consequently, the lower accuracy observed in tail categories is primarily influenced by two factors: the complexity of the environment and the limited availability of sufficient data. Existing studies on the long tail effect primarily focus on addressing the issue of data deficiency (Fagner et al., 2023), often overlooking the improvement of model robustness in complex environments. The proposed adaptive image processing module in this article presents an alternative perspective and offers a potential solution to tackle the challenges associated with the long tail effect.

In real-world scenarios, the accuracy rate for conifer forest environments was 83.03%, while for other environments it ranged between 92% and 94%. The coniferous forest environment exhibits a wide range

Fig. 8. Detection results of our method in real-world images. (a)-(b) White light. (c)-(d) Infra-red. (e)-(f) Complex light. (g)-(h) Lens Blur. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

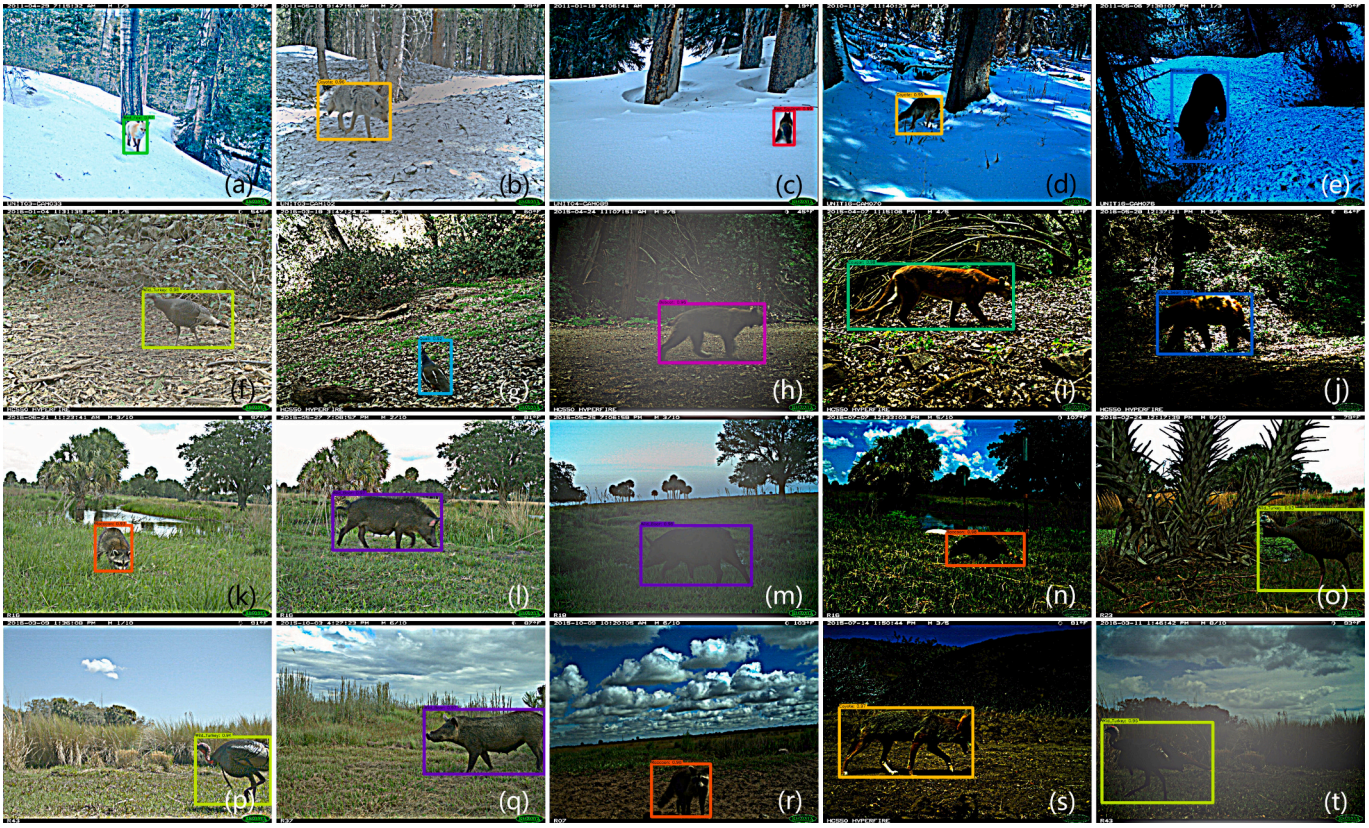


Fig. 9. Detection results of our method for various ecological environments in the SI-Dataset. Figures (a)-(e) show samples of Conifer Forests. Figures (f)-(j) show samples of Deciduous Forests. Figures (k)-(o) show samples of Wetlands. Figures (p)-(t) show samples of Grassland.

of seasonal variations, including snowfall, intense sunlight, and more. Moreover, it is the only area where infrared light is utilized. The performance of our models across different ecological environments demonstrates the influence of their complex conditions. In the comparison using the SI-Test-Dataset, the wetland area showed more variability with a decrease of 5.93% in performance. This can be attributed to the unbalanced nature of the data in wetland areas, where certain rare species are more susceptible to degradation when subjected to environmental changes.

In scenarios with Strength 0 (real-world datasets), AIP-OD (Real-world) demonstrates the highest mAP, yet it exhibits limited robustness when faced with low-quality images. By training with hybrid data, the model's robustness is improved, resulting in superior performance compared to the baseline YOLOv3 across all synthetic strengths. The experimental results substantiate the reciprocal enhancement between the adaptive image processing method and hybrid data training. The optimal outcomes are achieved when both adaptive image processing and hybrid data training are employed in tandem.

The robustness of the model to image blur can be assessed using both the SI-Test-Dataset and the SI-Test-Dataset (no blur), which solely differ in the presence or absence of lens blur. In comparison to alternative methods, our approach exhibits minimal sensitivity to blur, with a marginal difference of only 0.17%. The AIP-OD(Real-world) model trained only on real-world datasets is worse in performance compared to YOLOv3. The AIP module enables the model to be more adapted to the training dataset. However, there are fewer low quality images in the real world compared to the synthetic set. Thus the scalability of the AIP-OD (Real-world) model is limited to the case where there are fewer low quality images. This outcome substantiates the robustness of our method when handling blurred images and underscores the importance of utilizing both the hybrid data and image processing modules.

In real-world scenarios, the model successfully retrieves significant information about the original blurred objects and objects under varying lighting conditions. While the adaptive DIP module may occasionally introduce minor visual perceptual noise, it substantially enhances local image gradients by leveraging image semantics. Consequently, the model demonstrates remarkable adaptability to the complex environments encountered in the wild.

The SI-Dataset is designed to create a comprehensive and diverse test set by incorporating various lighting changes and blurring effects in different ecological environments. Thorough testing on the SI-Dataset serves as a crucial validation for assessing the robustness of our model across different environmental conditions. By incorporating the AIP module and training it with mixed data, the ecological task model exhibits a high degree of resilience when confronted with significant environmental variations. This dataset serves as an excellent foundation for evaluating the generalization capabilities of subsequent research.

5. Conclusions

This manuscript introduces an AIP designed to enhance the quality of camera trap images. This innovative approach can be effortlessly incorporated into all pre-existing models, thereby improving their performance with the addition of a minimal number of parameters. A NLPP addresses the issue of threshold determination in image processing. The NLPP module utilizes a non-local layer to capture long-distance features to improve prediction accuracy. Furthermore, a joint training approach is proposed, which simultaneously trains the parameter predictor and the deep learning model to ensure that the image processing effect contributes to the overall performance improvement of the deep learning model. Experimental results demonstrate the seamless integration of our method into existing deep learning models, enhancing their robustness in low-quality environments. When applied to real-world camera trap images and synthetic image datasets, our adaptive image processing method achieves impressive mAP values of 92.26% and 86.65%, respectively, in object detection tasks.

Notably, our approach focuses on enhancing overall performance by incorporating adaptive image processing modules external to the deep learning model, distinguishing it from previous studies that primarily concentrate on improving the model itself. This alternative perspective offers valuable insights into addressing numerous ecological research challenges based on camera traps. In addition, our approach can be generalized to the field of image processing for other scenes or types of complex environments, providing ideas for solving the problem of low-quality image recognition and processing in complex environments with uncertain lighting and blurred lenses. However, it is important to acknowledge that our study focused solely on investigating the effects of lighting and blur on camera trap images. Therefore, future research will delve into exploring the impact of more extreme environmental conditions, such as strong winds, heavy rain, blizzards, and others, on the images captured by camera traps.

Funding

This research was funded by the Forestry Achievement Promotion Program of the National Forestry and Grassland Administration (Grant number [2019]04) and Beijing Municipal Natural Science Foundation (Grant number 6192019).

Institutional review board statement

Not applicable.

Informed consent statement

Not applicable.

CRediT authorship contribution statement

Zihe Yang: Conceptualization, Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – original draft. **Ye Tian:** Conceptualization, Formal analysis, Methodology, Validation, Writing – review & editing. **Junguo Zhang:** Funding acquisition, Project administration, Resources, Supervision, Validation.

Declaration of competing interest

The authors declare no conflict of interest.

Data availability

Dataset NACTI is available at http://lila.science/datasets/nacti.

References

Alexander, G.V., Augusto, S., Francisco, V., 2017. Towards automatic wild animal monitoring: identification of animal species in camera-trap images using very deep convolutional neural networks. Eco. Inform. 41, 24–32. https://doi.org/10.1016/j.ecoinf.2017.07.004.

Bochkovskiy, A., Wang, C., Liao, H., 2020. YOLOv4:optimal speed and accuracy of object detection. arXiv preprint. https://doi.org/10.48550/arXiv.2004.10934.

Borowiec, M.L., Dikow, R.B., Frandsen, P.B., McKeeken, A., Valentini, G., White, A.E., 2022. Deep learning as a tool for ecology and evolution. Methods Ecol. Evol. 13, 1640–1660. https://doi.org/10.1111/2041-210X.13901.

Chen, Y., Li, W., Sakaridis, C., Dai, D., Van Gool, L., 2018. Domain adaptive faster r-cnn for object detection in the wild. In: Proceedings of IEEE/CVF Conference Computer Vision Pattern Recognition, Salt Lake City, USA. 18-22 June, pp. 3339–3348. https://doi.org/10.1109/cvpr.2018.00352.

Christian, S., Liu, W., 2015. Going deeper with convolutions. In: Proceedings of IEEE/CVF Conference Computer Vision Pattern Recognition, Boston, USA. 7-12 June. https://doi.org/10.48550/arXiv.1409.4842.

Christopher, M.B., Hugh, B., 2024. Deep Learning: Foundations and Concepts. Springer.

Delisle, Z.J., Henrich, M., Palencia, P., Swihart, R.K., 2023. Reducing bias in density estimates for unmarked populations that exhibit reactive behaviour towards camera traps. Methods Ecol. Evol. 14, 3100–3111. https://doi.org/10.1111/2041-210X.14247.

Dong, L., Shuang, X., Bo, X., 2018. Speech-transformer: A no-recurrence sequence-to-sequence model for speech recognition. In: In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Calgary, Canada. 15–20 April, pp. 5884–5888. https://doi.org/10.1109/ICASSP.2018.8462506.

Duchi, J., Hazan, E., Singer, Y., 2011. Adaptive subgradient methods for online learning and stochastic optimization. J. Mach. Learn. Res. 12, 2121–2159. https://doi.org/10.5555/1953048.2021068.

Enlin, L., Liwei, W., Qiuju, X., Rui, G., Zhongbin, S., Yonggang, L., 2023. A novel deep learning method for maize disease identification based on small sample-size and complex background datasets. Eco. Inform. 41, 102011 https://doi.org/10.1016/j.ecoinf.2023.102011.

Fagner, C., Eulanda, M.S., Juan, G.C., 2023. Bag of tricks for long-tail visual recognition of animal species in camera-trap images. Eco. Inform. 76, 102060 https://doi.org/10.1016/j.ecoinf.2023.102060.

Frank, S., Volker, S., 2021. Identification of animals and recognition of their actions in wildlife videos using deep learning techniques. Eco. Inform. 61, 101215 https://doi.org/10.1016/j.ecoinf.2021.101215.

Girshick, R., 2015. Fast r-cnn. In: Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile. 7-13 December https://doi.org/1440-1448.10.1109/ICCV.2015.169.

Girshick, R., Donahue, J., Darrell, T., Malik, J., 2014. Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE

Conference on Computer Vision and Pattern Recognition. Columbus, USA. 23-28 June. https://doi.org/10.1109/CVPR.2014.81.
Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R., 2020. Zero-reference deep curve estimation for low-light image enhancement. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, USA. 13–19 June, pp. 1780–1789. https://doi.org/10.1109/CVPR42600.2020.00185.
He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 26 June - 1 July, pp. 770–778. https://doi.org/10.1109/CVPR.2016.90.
Herbert, R., Sutton, M., 1951. A stochastic approximation method. Ann. Math. Stat. 400-407 https://doi.org/10.1109/TSMC.1971.4308316.
Hnewa, M., Radha, H., 2021. Multiscale domain adaptive Yolo for cross-domain object detection. In: IEEE International Conference on Image Processing. Taipei, China. 21-26 September. https://doi.org/10.1109/icip42928.2021.9506039.
Hu, Y., He, H., Xu, C., Wang, B., Lin, S., 2018. Exposure: a white-box photo post-processing framework. ACM Trans. Graph. 37 (2), 26. https://doi.org/10.1145/3181974.
Kingma, D., Ba, J., 2014. Adam: A Method for Stochastic Optimization. *arXiv*. https://doi.org/10.48550/arXiv.1412.6980.
Krizhevsky, A., Sutskever, I., Hinton, G., 2017. ImageNet classification with deep convolutional neural networks. Commun. ACM 60 (6), 84–90. https://doi.org/10.1145/3065386.
LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444. https://doi.org/10.1038/nature14539.
Li, X.H., Tian, H.D., Piao, Z.J., Wang, G.M., Xiao, Z.S., Sun, Y.H., Gao, E.H., Holyoak, M., 2022. Cameratrapr: an r package for estimating animal density using camera trapping data. Eco. Inform. 69, 101597 https://doi.org/10.1016/j.ecoinf.2022.101597.
Lin, L., Ronggang, W., Wenmin, W., Wen, G., 2015. A low-light image enhancement method for both denoising and contrast enlarging. In: Proceedings of the IEEE International Conference on Image Processing. Quebec Cit, Canada, 27–30 September, pp. 3730–3734. https://doi.org/10.1109/ICIP42928.2021.9506039.
Liu, W., Ren, G., Yu, R., Guo, S., Zhu, J., Zhang, L., 2022. Image-adaptive YOLO for object detection in adverse weather conditions. In: The AAAI Conference on Artificial Intelligence, Online, 22 February - 1 March, pp. 1792–1800. https://doi.org/10.48550/arXiv.2112.08088.
Mading, L., Jiaying, L., Wenhan, Y., Xiaoyan, S., Zongming, G., 2018. Structure-revealing low-light image enhancement via robust Retinex model. IEEE Trans. Image Process. 6 (27), 2828–2841. https://doi.org/10.1109/TIP.2018.2810539.
Manuel, V., Luis, P., Aldo, A.G.A., Pastor, L., Hugo, J.E., Jose, A.G., 2021. Desert bighorn sheep (Ovis canadensis) recognition from camera traps based on learned features. Eco. Inform. 64, 101328 https://doi.org/10.1016/j.ecoinf.2021.101328.
Miao, Z., Gaynor, K.M., Wang, J., Liu, Z., Muellerklein, O., Norouzzadeh, M.S., 2019. Insights and approaches using deep learning to classify wildlife. Sci. Rep. 9 (1), 8137. https://doi.org/10.1038/s41598-019-44565-w.
Michael, A., Mohammad, S., David, W., Steven, J., 2018. Machine learning to classify animal species in camera trap images: applications in ecology. Methods Ecol. Evol. 10 (4), 585–590. https://doi.org/10.1111/2041-210X.13120.
Midori, T., Alejandro, I.L., 2020. Image-based insect species and gender classification by trained supervised machine learning algorithms. Eco. Inform. 60, 101135 https://doi.org/10.1016/j.ecoinf.2020.101135.
Mohammad, S.N., Anh, N., Margaret, K., Alexandra, S., Meredith, S.P., Craig, P., Jeff, C., 2018. Automatically identifying, counting, and describing wild animals in camera-trap images with deep learning. Proc. Natl. Acad. Sci. 115 (25), 5716–5725. https://doi.org/10.1073/pnas.1719367115.
Mosleh, A., Sharma, A., Onzon, E., Mannan, F., Heide, F., 2020. Hardware-in-the-loop end-to-end optimization of camera image processing pipelines. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, USA, 13-19, June, 2020, pp. 7529–7538. https://doi.org/10.1109/CVPR42600.2020.00755.
Nazir, S., Kaleem, M., 2021. Advances in image acquisition and processing technologies transforming animal ecological studies. Eco. Inform. 61, 101212 https://doi.org/10.1016/j.ecoinf.2021.101212.
Niedballa, J., Sollmann, R., Courtiol, A., Wilting, A., Jansen, P., 2016. CamtrapR: an R package for efficient camera trap data management. Methods Ecol. Evol. 7 (12), 1457–1462. https://doi.org/10.1111/2041-210X.12600.
Norouzzadeh, M.S., Morris, D., Beery, S., Joshi, N., Jojic, N., Clune, J., 2021. A deep active learning system for species identification and counting in camera trap images. Methods Ecol. Evol. 12, 150–161. https://doi.org/10.1111/2041-210X.13504.
Patrick, B., Bill, C., Ashish, B., Vivek, G., Derek, E.L., 2019. An automated program to find animals and crop photographs for individual recognition. Eco. Inform. 50, 191–196. https://doi.org/10.1016/j.ecoinf.2019.02.003.
Petso, T., Jamisola, R.S., Mpoeleng, D., 2022. Review on methods used for wildlife species and individual identification. Eur. J. Wildl. Res. 68, 3. https://doi.org/10.1007/s10344-021-01549-4.
Qi, Y., Yang, Z., Sun, W., Lou, M., Lian, J., Zhao, W., Deng, X., Ma, Y., 2022. A comprehensive overview of image enhancement techniques. Arch. Computat. Methods Eng. 29, 583–607. https://doi.org/10.1007/s11831-021-09587-6.
Qi, S., Yu, G., Xi, G., Xinhui, G., Yufeng, C., Hongfang, W., Jianping, G., Tianming, W., Lei, B., 2024. Benchmarking wild bird detection in complex forest scenes. Eco. Inform. 80, 102466 https://doi.org/10.1016/j.ecoinf.2024.102466.
Redmon, J., Farhadi, A., 2017. YOLO9000: better, faster, stronger. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, USA. 21-26 July, pp. 7263–7271. https://doi.org/10.1109/CVPR.2017.690.
Redmon, J., Farhadi, A., 2018. Yolov3: an incremental improvement. arXiv preprint. https://doi.org/10.48550/arXiv.1804.02767.
Redmon, J., Divvala, S., Girshick, R., Farhadi, A., 2016. You only look once: unified, real-time object detection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA. 26 June-1 July, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91.
Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: towards real-time object detection with region proposal networks. IEEE Trans. Pattern Anal. Mach. Intell. 39 (6), 1137–1149. https://doi.org/10.1109/TPAMI.2016.2577031.
Sajid, N., Mohammad, K., 2024. Object classification and visualization with edge artificial intelligence for a customized camera trap platform. Eco. Inform. 79, 102453 https://doi.org/10.1016/j.ecoinf.2023.102453.
Schaus, J., Uzal, A., Gentle, L.K., Baker, P.J., Bearman-Brown, L., Bullion, S., Gazzard, A., Lockwood, H., North, A., Reader, T., Scott, D.M., Sutherland, C.S., Yarnell, R.W., 2020. Application of the random encounter model in citizen science projects to monitor animal densities. Remote Sens. Ecol. 6, 514–528. https://doi.org/10.1002/rse2.153.
Simonyan, K., Zisserman, A., 2015. Very deep convolutional networks for large-scale image recognition. In: International Conference on Learning Representations. San Diego, USA. 7-9 May. https://doi.org/10.48550/arXiv.1409.1556.
Swanson, A., Kosmala, M., Lintott, C., Simpson, R., Smith, A., Packer, C., 2015. Snapshot Serengeti, high-frequency annotated camera trap images of 40 mammalian species in an African savanna. Sci. Data 2, 150026. https://doi.org/10.1038/sdata.2015.26.
Tabak, M.A., Norouzzadeh, M.S., Wolfson, D.W., Sweeney, S.J., Miller, R.S., 2018. Machine learning to classify animal species in camera trap images: applications in ecology. Methods Ecol. Evol. 10 (4), 585–590. https://doi.org/10.1111/2041-210X.13120.
Thau, D., Ahumada, J.A., Birch, T., Fegraus, E., Mcshea, W.J., 2019. Artificial intelligence's role in global camera trap data management and analytics via wildlife insights. Biodiv. Inform. Sci. Stand. 3, e38233 https://doi.org/10.3897/biss.3.38233.
Tieleman, T., Hinton, G., 2012. Lecture 6.5-rmsprop: divide the gradient by a running average of its recent magnitude. COURSERA: Neural Networks Machine Learn. Tech. Rep. 4 (2), 26–31.
Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I., 2017. Attention is all you need. In: Neural Information Processing Systems. Long Beach, USA. 4-9 December, pp. 6000–6010. https://doi.org/10.48550/arXiv.1706.03762.
Wang, X., Ross, G., Abhinav, G., He, K., 2018. Non-local neural networks. arXiv preprint. https://doi.org/10.48550/arXiv.1711.07971.
Wei, L., Dragomir, A., Dumitru, E., Christian, S., Scott, R., Cheng-Yang, F., Alexander, B. C., 2016. SSD: single shot multiBox detector. In: European Conference on Computer Vision. Amsterdam, Netherlands. 10-16 October 2016, p. 9905. https://doi.org/10.1007/978-3-319-46448-0_2.
Xianchong, X., Yong, L., Lu, L., Peng, Y., Jianyi, Z., 2023. MAD-YOLO: a quantitative detection algorithm for dense small-scale marine benthos. Eco. Inform. 75, 102022 https://doi.org/10.1016/j.ecoinf.2023.102022.
Zan, W., Yiming, L., Xuanli, W., Dezhang, M., Lixiu, N., Guiqin, A., Xuanhui, W., 2022. An improved faster R-CNN model for multi-object tomato maturity detection in complex scenarios. Eco. Inform. 72, 101886 https://doi.org/10.1016/j.ecoinf.2022.101886.
Zeng, H., Cai, J., Li, L., Cao, Z., Zhang, L., 2020. Learning image-adaptive 3D lookup tables for high performance photo enhancement in real-time. IEEE Trans. Pattern Anal. Mach. Intell. 44 (4), 2058–2073. https://doi.org/10.1109/TPAMI.2020.3026740.
Zhao, Z., Zheng, P., Xu, S., Wu, X., 2019. Object detection with deep learning: a review. IEEE Trans. Neural Networks Learn. Syst. 30 (11), 3212–3232. https://doi.org/10.1109/TNNLS.2018.2876865.