INTL JOURNAL OF ELECTRONICS AND TELECOMMUNICATIONS, 2026, VOL. 72, NO. 2, PP. 1–7
Manuscript received February 8, 2026; revised April 2026. doi: 10.24425/ijet.2026.157933

Zero Short Learning for wildlife imagery

Ajay Kumar Boyat, Vinit Gupta, Aditya Mandloi and Kuber D. Gautam

***Abstract*—This paper introduces an innovative approach for object detection from wildlife images using Zero-Shot Learning (ZSL) with the YOLO-World model. Unlike previous object detection algorithms, which relied on domain-specific training data, YOLO-World is optimized for zero-shot object recognition, thus recognizing a wide range of categories without explicit training on specific labels. The data for this research have been taken from a dataset pre-processed and pre-trained, already split into sets of training and testing, such that the accuracy in the resulting outcome is more precise. Performance evaluation has been taken with the help of key parameters such as precision, recall, F1 score, Intersection over Union (IoU), mean Average Precision (mAP), and proved the adequacy of the model and its efficiency in detecting highly accurate wildlife objects. The experimental results highlight the better performance of ZSL in the detection of wildlife imagery, with a precision of 0.95 and recall of 0.92, thus achieving a mAP of 0.93 and F1-score of 0.87. A comparative analysis with existing YOLOv3 and YOLOv5 models also highlights the merits of the proposed approach in wildlife recognition tasks.**

***Keywords*—wildlife imagery; Zero-Shot Learning (ZSL); wild-domestic-dataset; YOLO world model; YOLOv3; YOLOv5**

I. Introduction

USING The natural environment has a considerable influence on the preservation of ecological equilibrium. The presence of wildlife adds uniformity to the many natural processes that occur in nature [1]. Wildlife monitoring is a crucial element of conservation. Conservation initiatives grounded on evidence, decision-making informed by data for adaptive management, and the sustainable use of natural resources rely on the assumption that population losses may be promptly identified [2]. Monitoring aims may vary from evaluating species presence or absence to determining the precise density of one or more species.

Recent advancements in contemporary monitoring technologies, including video traps, collaring devices, and conservation drones, have initiated a new epoch in wildlife monitoring and management [3]. Although they have aided local people in asserting, safeguarding, and overseeing their natural resources, such digital tools have also altered the public's perception of nature protection, especially in protected regions. Contemporary instruments shape human interaction with animals, hence affecting the selection and execution of conservation policy. The use of drones for animal conservation has significantly increased lately. In 2012, a prototype conservation drone conducted 32 missions in the Aral Napal Area, near to Gunung Leuser National Park in Sumatra, Indonesia, and obtained significant photographic documentation of oil palm farms and their impact on orang- utans [4]. Drones have since been used in several national parks to assist rangers and game wardens in combating poaching and unlawful hunting [5], Monitoring gives useful information and insights to conservation groups and government agencies by collecting and evaluating data on numerous aspects of wildlife, including as habitat usage, migratory patterns, quantity distribution, and behavioural activity [6]. These data reveal crucial concerns about species population status, habitat quality, and habitat connectivity, which aids in the formulation of successful conservation strategies and management plans. Thus, identifying animal species is critical for detecting species variety and conserving rare and endangered species. Camera traps are now the most often used approach for monitoring animal species [7], [8]. However, manual identification of monitoring pictures has been shown to be difficult due to high intensity and poor efficiency. With the rapid advancement of artificial intelligence technology, automated wildlife identification algorithms based on deep networks have performed admirably on certain datasets [9], [10], [11]. A deep convolutional neural network for automated animal identification was developed [12] and it obtained Top-1 accuracy of 88.9 percentile and Top-5 accuracy of 98.1percentile on the Serengeti dataset, which is a publicly available wildlife dataset. Trnovszky et al. [13] obtained 98.0percentile identification accuracy on a dataset including five animal species by improving the LeNet model. Verma et al. [14] investigated the influence of crowded scene photos that do not include individual animals in wildlife identification using monitoring datasets. They used deep convolutional neural networks (DCNNs) to extract features from crowded scene photographs and performed cluttered scene image identification using VGGNet and ResNet, therefore boosting the recognition accuracy of high-value wildlife monitoring images. Schneider et al. [15] investigated the topic of model generalization in unfamiliar settings and compared the performance of several deep learning approaches on diverse datasets. Vargas-Felipe et al. [16] used convolutional neural networks to identify animals in monitoring photos, achieving much superior identification accuracy compared to conventional approaches. However, in the actual world, as demonstrated in Figure 1, variables like as various backdrops, changing lighting conditions, and diverse

Ajay Boyet, was the ex Assistant Professor with Faculty of Electronics Engineering, at Medicaps University, Indore, India (e-mail: ajaykumar.boyat@medicaps.ac.in).

Vinit Gupta, Aditya Mandloi are with Faculty of Electronics Engineering, at Medicaps University, Indore, India (e-mail: vinit, aditya.mandloi@medicaps.ac.in).

Kuber D. Gautam, is with Faculty of Engineering, at Medicaps University, Indore, India (e-mail: kuber.gautam@medicaps.ac.in).

shooting scales may result in differences in feature distribution within the same class of images.

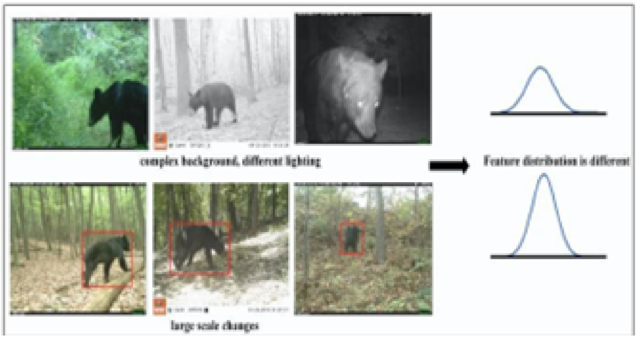

Fig. 1. The Characteristics Of Wildlife Images.

computational coprocessors in classical ICT systems, but so far only for a confined set of problems [2]. Search goes on widening this set.

A. *Limitations of traditional methods*

Regardless of using classic machine learning or deep learning techniques, it is challenging and inconsistent to get a substantial quantity of training samples in object detection, identification, or retrieval tasks, as well as to provide class definitions pertinent to these instances. Experts are required to develop the data collection and to extract the feature vectors. The picture tagging technique might be deceptive in addressing the issue due to its inherent subjectivity. Despite the availability of extensive data sets, it is not always feasible to identify samples for the majority of issues [17]. Conversely, enhancing the contextual understanding of the situation at hand presents an additional hurdle within this field of research. To address these challenges, there has been a recent surge in research focused on recognizing samples that lack representation, using the semantic linkages across samples within the dataset. Zero-Shot Learning is the foremost of these investigations. Even if one does not identify the observed sample, one may infer its classification by examining its attributes. This technique considers the class embedding's of the pictures in ways that do not use training examples. It is modeled by aligning the characteristics of the visual data with the multidimensional vector space including the qualities of this data. A potential strategy for identifying previously unobserved classes is semantic transfer. The domain ontology is of paramount importance in this context. Known categories may be retrieved using certain metadata. Zero-shot learning demonstrates its significance in contrast to conventional machine learning techniques. This learning approach utilizes prior experience to comprehend the current circumstance [17] This study examines the object detection of wild life imagery by using the Zero-Shot Learning model.

B. *Application of Zero-Shot Learning*

Zero-shot learning (ZSL) is very useful in real-world situations where unknown classes provide challenges for traditional supervised learning methods. ZSL becomes necessary when obtaining labeled data for every possible class is impossible or costly. A common use is image recognition systems, where the ability to detect objects or entities that were not part of the training set is critical for adapting to changing environments. ZSL also has an impact on natural language processing since it permits textual content that was not expressly included during the model's training phase to be categorized [18].

C. *Organization of the Paper*

The remainder of this work is structured as follows. Section II reviews the existing wildlife image collection and the object detection techniques. In Section III, addresses the proposed models' backgrounds. Finally, Section IV explains the methods used in this study and the evaluation tools. Section V discusses the experimental data and assesses the proposed method's shortcomings. Section VI outlines the outcomes of this study and discusses prospective future research topics.

II. RELATED WORK

In this part, we explore the existing body of research on wildlife image collection and the object detection techniques. The findings and insights from prior studies are synthesized and presented, effectively encapsulating the identified research gaps. Deep learning-based recognition algorithms has certain

TABLE I
RESULTS OF SIMULATION

Author	Dataset	Method	Accuracy
[19]	NACTI (26species)	ResNet-18	97.6
[20]	9051+ HD Cam	YOLOv3	75.2
[21]	Snapshot	Deep active learning	92.9
[22]	ENA24	DJAN	48.8
[23]	COCO	YOLOv8	94
[24]	COCO	YOLOv5	70
[25]	NTLNP	R- CNN	98

constraints when compared to classic machine learning techniques. They depend on several datasets, and their accuracy is inferior to that of conventional machine learning techniques. Nonetheless, the deep learning approach obviates the need for manual feature screening and reduces the requirement for personnel and resources to evaluate the features. Furthermore, the ongoing advancement of animal monitoring technology, like trap cameras and long-range drones, will provide an increased volume of more extensive visual data. The recognition efficacy of the deep learning model will enhance in accordance with the quality of the data. Currently, wildlife detection process depends only on attributes derived from wildlife picture data. Nevertheless, substantial specialist knowledge may be used to improve the wildlife detection and categorization process [26]. These studies seek to enhance future recognition tasks by the strategic use of previously acquired information. A prominent and contemporary issue in this domain is the development of recognition algorithms capable of identifying novel visual classes without requiring labeled training samples, specifically, properly recognizing the class without prior observation. This particular area of study is referred to as zero-shot learning (ZSL) [27].

III. BACKGROUND

In this part, background details on Zero-Shot Learning is provided.

A. Zero-Shot Learning

In zero-shot learning (ZSL), the aim is to classify instances from unseen classes without any labeled training data from those classes. Let the set of seen classes be denoted by: $S = c_s^i|i = 1, ..., N_s$. and the set of unseen classes be: $U = c_u^i|i = 1, ..., N_u$. The seen and unseen sets are disjoint, i.e., $S \cap U = \Phi$, indicating that the model is never explicitly trained on unseen species. This setup mimics real-world wildlife scenarios, where models are required to generalize to rare or previously undiscovered species based on semantic attributes. Let $X \in \mathbb{R}^D$ denote the D-dimensional feature space.The training dataset is represented as: $X_tr = (x_t^ir, y_t^ir) \in X \times S|i = 1, ..., N_tr$ where:

- $X_t^ir \in X$ are the feature vectors of the training samples,
- $Y_t^ir \in X$ are the corresponding class labels from the seen classes.

The overall label set is: $Y = Y^i \in S \cup U|i = 1, .., N_tr + N_te$ The test set consists of instances: $X_te = X_r^ie \in X|i = N_tr + 1, .., N_tr + N_te$. and their corresponding labels from unseen classes: $Y_re = Y_r^ie \in U|i = N_tr + 1, .., N_tr + N_te$.

Since the unseen class set U lacks labeled training images, ZSL employs semantic embeddings (e.g., attributes, textual descriptions, or word vectors such as Word2Vec, GloVe, or CLIP) to bridge the gap between seen and unseen classes. For instance, if a model is trained to recognize tigers and lions (seen classes), it can infer what a leopard (unseen class) looks like by mapping shared attributes such as "four legs," "sharp claws," and "spotted coat." The task of ZSL is to learn a classifier: $f_u(\cdot) : X \rightarrow U$ which maps test instances $X_te \in X$ to one of the unseen classes U, thus predicting the correct label $y_te \in Y_te$ [28].

B. ZSL as a Type of Transfer Learning

ZSL can be considered a subtype of transfer learning because it involves transferring knowledge from source domains (seen classes) to target domains (unseen classes). Given that training and testing instances, as well as their class labels, are disjoint, ZSL qualifies as heterogeneous transfer learning . Transfer learning is categorized as:

- Homogeneous transfer learning: where the feature and label spaces are the same, i.e.,$X_tr = X_te$ and $Y_tr = Y_te$,
- Heterogeneous transfer learning: where either the feature space and/or the label space differ, i.e.,$X_tr \neq X_te$ and/or $Y_tr \neq Y_te$.

In ZSL:

- The feature space is the same for both training and testing: $X_tr = X_te = X$
- The label spaces are disjoint:$Y_tr = S, Y_te = U$ with $S \cap U = \Phi$

Therefore, ZSL is formally classified as a heterogeneous transfer learning problem. While traditional transfer learning methods rely on at least some labeled data in the target domain, ZSL is unique in that no labeled instances are available for the target classes, making it a challenging and distinct paradigm.

In zero-shot learning, the feature space of the training instances is designated as the source feature space, while the feature space of the testing instances is referred to as the target feature space. In zero-shot learning, both the source and target feature spaces are identical, denoted as X. The source label space, regarded as the visible class set, differs from the target label space, identified as the unseen class set, since the seen and unseen class sets are discontinuous, i.e., $S \neq U$. Consequently, zero-shot learning is classified as a kind of heterogeneous transfer learning. Numerous techniques have been suggested to address the issue of having labelled instances for the target label space classes [29]; however, these methods are ineffective in scenarios where no labelled instances for the target label space classes exist, as is the case in zero-shot learning, where the unseen classes constitute the target label space classes. Consequently, zero-shot learning is distinctive in that sense.

IV. METHODOLOGY

This study proposes a methodology for the object detection of wild life imagery.

A. Data collection

In this study we collect the data from the following website https://universe.roboflow.com/yolo8-hfboz/wild- domestic/dataset/1.Because this dataset is already pre- processed and pre-trained dataset, and split into training and testing sets. This dataset was used in the research, resulting in better accuracy. Figure 2 shows the flow of methodology of this study.

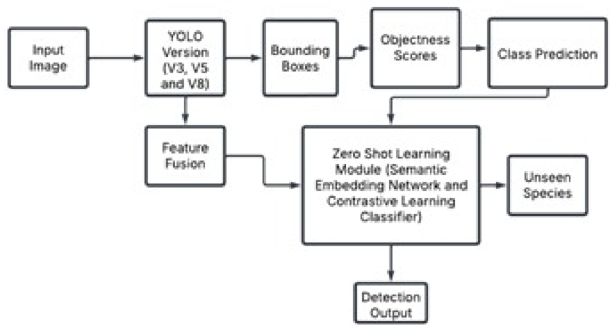

Fig. 2. Proposed Methodology Flow

B. Model Building

In this study, YOLO-World utilizes a pretrained YOLO model optimized for zero-shot object recognition. The model is designed to identify things according to a predetermined or user-defined set of categories. In contrast to conventional object identification algorithms, YOLO-World does not need domain- specific training data, making it adaptable to new or unrecognized item categories. Principal characteristics include:

- Class Definition: Object detection is executed according to user-specified categories, for example: ”dog,” ”cat,” or ”person.”
- Threshold alterations: Confidence thresholds can be constantly adjusted to optimize detection outcomes.

Table 2 shows the algorithm based on zero-shot detection with YOLO.

TABLE II
ALGORITHM 1 ZERO-SHOT OBJECT DETECTION WITH YOLO- WORLD

1	: Initialization:
2	Install necessary packages:
3	Import libraries:
4	Download example data:
5	Load YOLO-World model:
6	Set detection classes:
7	Perform detection on sample image:
8	Visualize detections:
9	Adjust confidence threshold:
10	Display confidence levels for each detection:
11	Eliminate double detections using Non-Max Suppression:
12	: Evaluate performance metrics:
13	: Output results:

A collage of different creatures, each named appropriately and encased in a bounding box, is shown in the figure 3. A

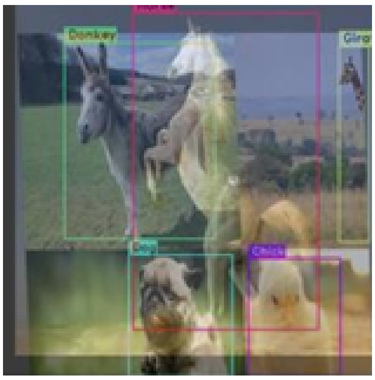

Fig. 3. Bounding box Image.

brown donkey is seen in the upper left corner, contentedly standing within its green cage. It is next to a magnificent white horse that seems to be in mid-stride as it occupies a bigger red box. A lively pug puppy is tucked up in a turquoise box underneath them, its tongue hanging out. A fluffy yellow chick pops out from a pink box in the bottom right corner, giving the setting lovely touch. Finally, a towering giraffe with a long neck extending toward the sky in the upper right corner contrasts its brown and yellow body with the blue sky within a blue box.

Fig. 4. Proposed Methodology Flow.

Each animal in the collage is labeled for easy identification and is neatly contained inside a bounding box. A colorful rooster with white feathers and a stunning red comb is framed within an orange box on the left. Below it, a chubby hen is tucked up in a yellow box, and the tableau is made cozier by its soft stare. A nosy dog is seen in the middle, caught in mid-stretch within a teal box, its lively energy evident. Finally, within a green box, a flying squirrel with wide eyes and extended paws expressing awe flies into the air in the upper right corner.

C. Training and Testing

Despite being pre-trained, this model provides a comprehensive testing framework to assess its performance: The procedure entails analyzing input files (images or videos) on a frame-by-frame basis, using methods such as confidence thresholding, non-max suppression, and area-based filtering to minimize false positives and enhance detection accuracy. The supervision library superimposes bounding boxes and labels on the input data, facilitating both qualitative (visual inspection) and quantitative assessment (based on performance measures).

D. Model Evaluation

The performance of the proposed model was thoroughly and rigorously evaluated, using performance metrics such as precision, recall and F1 score, IoU and mAP. This procedure involved a thorough examination of the various methodologies used to forecast water quality metrics. True positive (TP) is the proper detection of a ground-truth bounding box. False positive (FP) is an inaccurate detection of a non-existent item or a mistaken detection of an existing entity. False negative (FN) is an undiscovered ground truth bounding box. In the context of object detection, it is crucial to recognize that a true negative (TN) result is inapplicable, since there exists an endless number of bounding boxes that should remain undetected inside any given image [30]

V. RESULTS AND DISCUSSION

The results emphasize model performance and show how well existing object detection models can recognize animals from wildlife imagery.

TABLE III
ZERO-SHOT-LEARNING

Models	Zero-Shot-Learning
PPYoloELoss/loss@cls	1.2
PPYoloELoss/loss@iou	0.75
PPYoloELoss/loss@dfl	0.85
PPYoloELoss/loss	0.55
Precision@0.50	0.95
Recall@0.50	0.92
mAP@0.50	0.93
F1@0.50	0.87
Best score threshold	0.5

The table presents the performance of the proposed model, assessed using critical parameters pertinent to object identification. The classification loss (PPYoloELoss/loss@cls) is recorded as 1.2, indicating the model's mistake in differentiating classes. The Intersection over Union (IoU) loss (PPYoloELoss/loss@iou) is 0.75, and the Distance Focal Loss (PPYoloELoss/loss@dfl) is 0.85, indicating the model's efficacy in spatial alignment and bounding box prediction. The total loss (PPYoloELoss/loss) is very low at 0.55, indicating effective learning. The model has superior detection skills, attaining a precision of 0.95, a recall of 0.92, and a mean average precision (mAP@0.50) of 0.93. The F1-score at 0.50 IoU is 0.87, indicating an equitable balance between accuracy and recall. This shown in figure 5.The ideal score threshold for performance is established at 0.5, indicating the models efficacy in attaining precise and dependable detections across various contexts. The table and figure demonstrate a comparative examination of several object detection algorithms, including Zero-Shot Learning, YOLOv3, and YOLOv5, according to essential performance indicators. The assessed metrics comprise classification loss (loss@cls), Intersection over Union loss (loss@iou), Distance Focal Loss (loss@dfl), overall loss (loss), precision at 0.50 IoU (Precision@0.50), recall at 0.50 IoU (Recall@0.50), mean average precision at 0.50 IoU (mAP@0.50), F1-score at 0.50 IoU (F1@0.50), and the optimal score threshold (Best@score@threshold). Zero-Shot Learning has exceptional performance, achieving an accuracy of 0.95, a recall of 0.92, and an F1-score of 0.87, underscoring its formidable detecting skills. Conversely, YOLOv3, YOLOv5, and YOLOv4-tiny demonstrate somewhat worse performance, with YOLOv3 attaining the best accuracy (0.198) among these models, however falling short in recall and mAP when compared with Zero-Shot Learning. These findings emphasize the efficacy of Zero-Shot Learning in wildlife identification tasks.

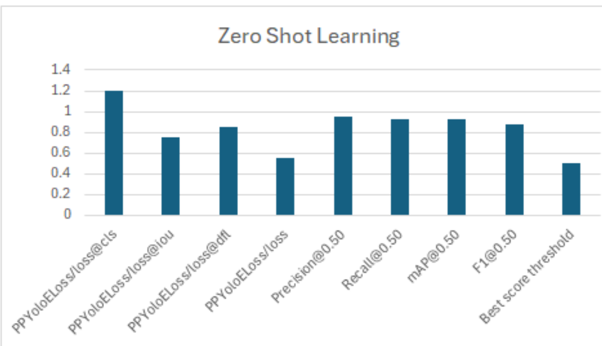

Fig. 5. Proposed Model Performance

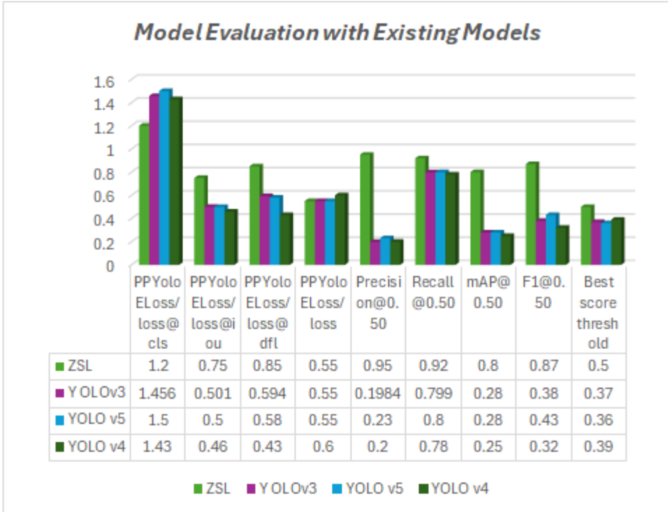

Fig. 6. Proposed Model Performance

VI. CONCLUSION

The findings of this research suggest that Zero-Shot Learning (ZSL) may be used to the field of wildlife item identification. Many wildlife things might be detected with high accuracy and recall value using the YOLO-World model, which doesn't need a lot of domain-specific data. The accuracy and efficacy of the model were enhanced by the study procedure, which included post-processing approaches, Non-Maximum Suppression, and confidence thresholding. When assessed using important metrics like as accuracy, recall, F1-score, and mAP, its performance demonstrated excellent performance for real- world animal identification tasks. The Zero-Shot Learning technique will be a highly useful tool for wildlife monitoring and conservation since it outperforms the current YOLO models in terms of accuracy, recall, and F1-score. This study lays the foundation for future advancements in zero-shot object identification models, which might find use in sectors other than wildlife imagery that need for flexible object recognition systems.

ACKNOWLEDGMENT

The authors would like to thank experts for their appropriate and constructive suggestions to improve this template.

References

[1] S. meena, "A study on the importance of wild life," *Ver. 4.5.*, 2015.

[2] M. Grooten and R. E. Almond, *Living planet report-2018: aiming higher.* WWF international, 2018.

[3] Y. Xian, B. Schiele, and Z. Akata, "Zero-shot learning – the good, the bad and the ugly," 2020. [Online]. Available: https://arxiv.org/abs/1703.04394

[4] L. P. Koh and S. A. Wich, "Dawn of drone ecology: low-cost autonomous aerial vehicles for conservation," *Tropical conservation science*, vol. 5, no. 2, pp. 121–132, 2012.

[5] M. Mulero-Pázmány, R. Stolper, L. Van Essen, J. J. Negro, and T. Sassen, "Remotely piloted aircraft systems as a rhinoceros anti-poaching tool in africa," *PloS one*, vol. 9, no. 1, p. e83873, 2014.

[6] D.-Q. Yang, G.-P. Ren, K. Tan, Z.-P. Huang, D.-P. Li, X.-W. Li, J.-M. Wang, B.-H. Chen, and W. Xiao, "An adaptive automatic approach to filtering empty images from camera traps using a deep learning model," *Wildlife Society Bulletin*, vol. 45, no. 2, pp. 230–236, 2021.

[7] J. Vélez, W. McShea, H. Shamon, P. J. Castiblanco-Camacho, M. A. Tabak, C. Chalmers, P. Fergus, and J. Fieberg, "An evaluation of platforms for processing camera-trap data using artificial intelligence," *Methods in Ecology and Evolution*, vol. 14, no. 2, pp. 459–477, 2023.

[8] C. P. Cordier, D. A. E. Smith, Y. E. Smith, and C. T. Downs, "Camera trap research in africa: A systematic review to show trends in wildlife monitoring and its value as a research tool," *Global Ecology and Conservation*, vol. 40, p. e02326, 2022.

[9] Z. Miao, Z. Liu, K. M. Gaynor, M. S. Palmer, S. X. Yu, and W. M. Getz, "Iterative human and automated identification of wildlife images," *Nature Machine Intelligence*, vol. 3, no. 10, pp. 885–895, 2021.

[10] D. Tuia, B. Kellenberger, S. Beery, B. R. Costelloe, S. Zuffi, B. Risse, A. Mathis, M. W. Mathis, F. Van Langevelde, T. Burghardt *et al.*, "Perspectives in machine learning for wildlife conservation," *Nature communications*, vol. 13, no. 1, p. 792, 2022.

[11] A. M. Roy, J. Bhaduri, T. Kumar, and K. Raj, "Wildect-yolo: An efficient and robust computer vision-based accurate object localization model for automated endangered wildlife detection," *Ecological Informatics*, vol. 75, p. 101919, 2023.

[12] A. G. Villa, A. Salazar, and F. Vargas, "Towards automatic wild animal monitoring: Identification of animal species in camera-trap images using very deep convolutional neural networks," *Ecological informatics*, vol. 41, pp. 24–32, 2017.

[13] T. Trnovszky, P. Kamencay, R. Orjesek, M. Benco, and P. Sykora, "Animal recognition system based on convolutional neural network," *Advances in Electrical and Electronic Engineering*, vol. 15, no. 3, p. 517, 2017.

[14] G. K. Verma and P. Gupta, "Wild animal detection from highly cluttered images using deep convolutional neural network," *International Journal of Computational Intelligence and Applications*, vol. 17, no. 04, p. 1850021, 2018.

[15] S. Schneider, S. Greenberg, G. W. Taylor, and S. C. Kremer, "Three critical factors affecting automated image species recognition performance for camera traps," *Ecology and evolution*, vol. 10, no. 7, pp. 3503–3517, 2020.

[16] M. Vargas-Felipe, L. Pellegrin, A. A. Guevara-Carrizales, A. P. López-Monroy, H. J. Escalante, and J. A. Gonzalez-Fraga, "Desert bighorn sheep (ovis canadensis) recognition from camera traps based on learned features," *Ecological Informatics*, vol. 64, p. 101328, 2021.

[17] O. A. Soysal and M. S. Guzel, "An introduction to zero-shot learning: An essential review," in *2020 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA)*. IEEE, 2020, pp. 1–4.

[18] S. El Maachi, A. Chehri, and R. Saadane, "Zero-shot-learning for plant species classification," *Procedia Computer Science*, vol. 246, pp. 734–742, 2024.

[19] M. A. Tabak, M. S. Norouzzadeh, D. W. Wolfson, S. J. Sweeney, K. C. VerCauteren, N. P. Snow, J. M. Halseth, P. A. Di Salvo, J. S. Lewis, M. D. White *et al.*, "Machine learning to classify animal species in camera trap images: Applications in ecology," *Methods in Ecology and Evolution*, vol. 10, no. 4, pp. 585–590, 2019.

[20] M. Gabriel, S. Cha, N. Y. B. Al-Nakash, and D. Yun, "Wildlife detection and recognition in digital images using yolov3," in *2020 IEEE Cloud Summit*. IEEE, 2020, pp. 170–171.

[21] M. S. Norouzzadeh, D. Morris, S. Beery, N. Joshi, N. Jojic, and J. Clune, "A deep active learning system for species identification and counting in camera trap images," *Methods in ecology and evolution*, vol. 12, no. 1, pp. 150–161, 2021.

[22] C. Zhang and J. Zhang, "Djan: Deep joint adaptation network for wildlife image recognition," *Animals*, vol. 13, no. 21, p. 3333, 2023.

[23] A. D. Shetty and S. Ashwath, "Animal detection and classification in image & video frames using yolov5 and yolov8," in *2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. IEEE, 2023, pp. 677–683.

[24] Y. Niu and Z. Wang, "Research on object detection in animal images based on convolutional neural networks," *Int. J. Adv. Network, Monit. Control*, vol. 08, no. 04, pp. 45–54, 2023.

[25] M. Tan, W. Chao, J.-K. Cheng, M. Zhou, Y. Ma, X. Jiang, J. Ge, L. Yu, and L. Feng, "Animal detection and classification from camera trap images using different mainstream object detection architectures," *Animals*, vol. 12, no. 15, p. 1976, 2022.

[26] Z. Ma, Y. Dong, Y. Xia, D. Xu, F. Xu, and F. Chen, "Wildlife real-time detection in complex forest scenes based on yolov5s deep learning network," *Remote Sensing*, vol. 16, no. 8, p. 1350, 2024.

[27] M. Chandrashekar and Y. Lee, "Class representative learning for zero-shot learning using purely visual data," *SN Computer Science*, vol. 2, no. 4, p. 313, 2021.

[28] T. Scheck, R. Seidel, and G. Hirtz, "Learning from theodore: A synthetic omnidirectional top-view indoor dataset for deep transfer learning," in *Proceedings of the IEEE/CVF Winter conference on applications of computer vision*, 2020, pp. 943–952.

[29] W. Wang, V. W. Zheng, H. Yu, and C. Miao, "A survey of zero-shot learning: Settings, methods, and applications," *ACM Transactions on Intelligent Systems and Technology (TIST)*, vol. 10, no. 2, pp. 1–37, 2019.

[30] V. Gupta and S. Pawar, "An effective structure of multi-modal deep convolutional neural network with adaptive group teaching optimization," *Soft Computing*, vol. 26, no. 15, pp. 7211–7232, 2022.