VIVA-Tech International Journal for Research and Innovation
ISSN(Online): 2581-7280

Volume 1, Issue 7 (2024)
Article No. COMP_20
PP 1-11

VIVA-TECH INTERNATIONAL JOURNAL FOR RESEARCH AND INNOVATION

ANNUAL RESEARCH JOURNAL
ISSN(ONLINE): 2581-7280

COP : TARGET RECKS USING YOLOv8

Abhishek Mandavkar[1], Dishant Save[2], Yash Patil[3], Prof. Janhavi Sangoi[4]

[1] *(Department of Computer Engineering, Mumbai University, Maharashtra, India)*
[2] *(Department of Computer Engineering, Mumbai University, Maharashtra, India)*
[3] *(Department of Computer Engineering, Mumbai University, Maharashtra, India)*
[4] *(Professor, Department of Computer Engineering, Mumbai University, Maharashtra, India)*

Abstract: *With the increasing need for effective wildlife monitoring and conservation efforts, computer vision technologies have emerged as powerful tools for automating animal detection in diverse environments. This paper introduces an innovative framework for the detection of Indian exclusive animals—species found exclusively in India—employing the YOLOv8 (You Only Look Once) object detection model. The proposed system is reinforced by a meticulously annotated dataset created through the Computer Vision Annotation Tool (CVAT), focusing specifically on the distinctive fauna inhabiting the Indian subcontinent. The YOLOv8 model, renowned for its speed and accuracy, is employed to detect animals in images and video frames. The YOLOv8 model is tailored to detect and classify indigenous animal species, ensuring its adaptability to the unique ecological contexts of India. By harnessing the real-time capabilities of YOLOv8, the system enables efficient and timely monitoring of exclusive wildlife populations, addressing the urgent need for accurate and scalable solutions in conservation efforts. The CVAT annotated dataset encapsulates a diverse array of Indian endemic species, encompassing various habitats and environmental conditions. The manual annotation process ensures precision in delineating bounding boxes around animals, contributing significantly to the enhancement of the model's detection accuracy for region-specific fauna. Addressing challenges such as diverse animal poses, complex backgrounds, and varying lighting conditions, our framework demonstrates its adaptability to the specific conditions prevalent in India. This work contributes to the growing body of research in wildlife conservation and monitoring, providing a scalable and accurate solution for automated animal detection. The proposed framework stands as a valuable tool for researchers, conservationists, and wildlife managers dedicated to safeguarding the unique biodiversity of India and its integral role in global ecological balance.*

Keywords- *Android, Annotations, CVAT, Detection, Endemic species, YOLOv8.*

I. INTRODUCTION

From self-driving cars and surveillance systems to augmented reality and healthcare, the ability to precisely identify and locate objects within an image or video stream has revolutionized various industries. Object detection, the process of identifying and locating objects within images or video frames, is a pivotal technology with applications ranging from autonomous vehicles to surveillance systems and beyond. In the realm of numerous object detection techniques, YOLO (You Only Look Once) distinguishes itself as an innovative method recognized for its remarkable speed and precision. This extensive manual explores the captivating domain of custom object detection through the utilization of YOLO. Introduced initially in 2015 and enhanced in subsequent iterations, YOLO boasts real-time object detection capabilities. It is particularly renowned for its ability to simultaneously locate and classify objects within an image with remarkable speed. However, while the standard YOLO model is impressive, customizing it to detect specific objects tailored to your unique requirements is where its true potential shines.

Establishing custom object detection with YOLO opens up a multitude of possibilities. Whether it's recognizing specialized industrial components, unique species in ecological research, or custom products in the retail sector, this project demonstrates the versatility and potential of YOLO to adapt and excel in diverse contexts. This endeavor has delved into the intricacies of data preparation, model training, and fine-tuning to empower YOLO to identify custom objects. The journey of establishing custom object detection using YOLO is one marked by technical challenges, innovative solutions, and a commitment to pushing the boundaries of computer vision.In this report, we will share the methodologies, insights, and outcomes of our journey in

Volume 1, Issue 7 (2024)
Article No. COMP_20
PP 1-11

establishing custom object detection using YOLO. The success we've achieved highlights the impactful potential of this technology, offering solutions across diverse industries and serving as evidence of the constantly advancing realm of artificial intelligence and computer vision.

This project has encompassed every facet of custom object detection, from the meticulous curation and preparation of datasets to the fine-tuning of YOLO's neural networks.

The outcomes of this endeavour not only provide practical solutions but also serve as a testament to the boundless possibilities of modern artificial intelligence. This endeavour is poised to not just encourage additional investigation but also enable individuals to initiate their own ventures in the dynamic realm of computer vision..

II. REVIEW OF LITERATURE SURVEY

Nithin Kumar ,Nagarathna and Francesco Flammini [1], A zoology subfield called entomology covers nonentity- related exploration. We discovered that a further thorough disquisition is needed to identify the species position of insects due to the vast number of dangerous nonentity populations. Entomology exploration is pivotal because it opens new avenues and benefits for chemistry, drug, engineering, and the medicinal inspired by insects and nature. Insects rob and annihilate a third of the world's crops, performing in the loss of multitudinous products, and businesses suffer losses. Quick and accurate identification of insects is essential to avoid fiscal losses and progress the study of entomology. Researchers find motivation in insects when creating robotics, sensors, mechanical designs, aerodynamics, and smart systems. These factors make scholarly exploration on nonentity discovery pivotal for demonstrating biodiversity. Relating the order position an nonentity belongs to is vital in determining insects. Knowing the order position is necessary to separate the type. Nonentity orders dating back to 2002 were discovered.

Sergejs Kodors, Edgars Rubauskis, Marks Sondors, lmars Apeinans, unars Lacis, Imants Zarembo [2], Pears are the third most economically important fruit crop encyclopaedically reaching25.7 million tons in 2021. Although it isn't the most important fruit crop in Latvia, it forms a veritably important niche product with high added value, and the area of pear growing is about 200. Timely and accurate vaticinator of fruit yield is also of great profitable significance to optimally plan post-harvest conditioning, storehouse installations and deals. Developing similar yield soothsaying systems has been going on for a long time for colourful fruit factory species. Still, the disadvantage of all these soothsaying systems is the need for high- quality, large- scale data because the delicacy of the developed model and the correspondence of the read and real gathered yield depend on it.

Tausif Diwan1 & G. Anirudh2 & Jitendra V. Tembhurne1 [3], offers a thorough examination of the You Only Look Once (YOLO) object detection framework. The paper covers various aspects related to YOLO, including challenges in object detection, architectural advancements, datasets commonly used for evaluation, and practical applications. It addresses the inherent challenges in object detection tasks, such as handling objects of varying sizes, dealing with class imbalance, and addressing the need for large annotated datasets. The paper explores the evolution and improvements made in YOLO's architecture and its successors. These datasets are essential for benchmarking and validating the effectiveness of detection algorithms. The paper likely explores real-world applications where YOLO and related models are employed for object detection. The authors likely employed key components of the deep learning tech stack. Overall, the paper offers valuable insights into the challenges, advancements, datasets, and real-world applications associated with YOLO-based object detection, making it a valuable resource for researchers and practitioners in the field of computer vision.

Sita M. Yadav, Sandeep M. Chaware [4], it explores video object detection techniques, including both traditional computer vision methods and deep learning approaches. The paper likely discusses various methods and algorithms used for detecting objects within videos. It may compare the effectiveness of traditional computer vision techniques, which include methods like feature extraction and tracking, with modern deep learning methods, such as convolutional neural networks (CNNs) and their variants, for video object detection tasks. Overall, this paper likely provides insights into the evolving landscape of video object detection, highlighting the advantages and limitations of both traditional and deep learning-based approaches in the context of this critical computer vision task. It may also discuss applications and challenges associated with video object detection.

Zhuo Bian, And LiangliangWang. [5] Frame difference is a quick diversity grounded segmentation approach for object discovery, unfortunately, it gets trapped in over-segmented when the pixels of interest over time lap each other. This paper presents a rather fast visual object discovery approach able of approaching the position of moving object under heavy background noise or big imbrication caused by negative similarity. Object discovery

Volume 1, Issue 7 (2024)
Article No. COMP_20
PP 1-11

is a introductory but critical content confederated nearly with image segmentation, object shadowing and recognition in computer vision, all of which have considerable implicit demand in. the field of videotape surveillance, mortal- computer commerce, virtual reality, robotics, intelligent transportation system and others. It's no wonder that a large number of attempts have been made to exploit what features can be used to represent the interest area from noninterest corridor which plays the crucial part, despite central challenges from script complexity, scale variation, occlusion, illumination changing and others.

Mohammed Boukabous, Mostafa Azizi [6], this journal presents an innovative approach to crime prediction through the integration of object detection and deep learning techniques. The research is conducted at the Mathematics, Signal and Image Processing, and Computing Research Laboratory (MATSI) at the Superior School of Technology (ESTO), Mohammed First University in Oujda, Morocco. In this study, the authors leverage the power of computer vision and deep learning to develop a system capable of predicting crimes based on image and video data. This research holds the potential to significantly contribute to the field of law enforcement and public safety by automating the process of crime prediction through advanced technology. The paper's findings and methodologies could prove valuable for both researchers and practitioners working on enhancing crime prevention strategies.

Thomas Tsao and Doris Y. Tsao[7] The primate visual system undergoes a complex and not fully understood process of perceptual association, wherein it transforms retinal images into a stable perception of distinct objects. Traditionally, this process involves addressing two challenging problems: segmentation, which deals with grouping visual pixels into separate objects within a single image, and shadowing, which tackles linking objects across images despite changes in appearance. This paper explores the computational basis of the ability to consistently track objects and demonstrates that this problem can potentially be solved without the need for supervision or literacy. Moving beyond image-based approaches to segmentation and shadowing, the study introduces a figure-grounded approach that views vision as an inverse plates problem. The research reveals that inferring 3D structures from images is entirely constrained when the input comprises a sequence of images depicting a scene with either moving bystanders or objects.

Mingqi Gao, Feng Zheng, James J. Q. Yu, Caifeng Shan, Guiguang Ding ,Jungong Han [8] Videotape object segmentation, considered a fundamental challenge in the field of video understanding, focuses on delineating objects of interest across a provided sequence of videotape. Lately, with the advancements of deep literacy ways, deep neural networks have shown outstanding performance advancements in numerous computer vision operations, object segmentation in videotape has received extensive support and thorough exploration. Latterly, we summarise the datasets for training and testing a videotape object segmentation algorithm, as well as common challenges and evaluation criteria. Next, former workshop are grouped and reviewed grounded on how they prize and use spatial and temporal features, where their infrastructures, benefactions and the differences among each other are developed. This composition is anticipated to serve as a tutorial and source of reference for learners intended to snappily grasp the current progress in this exploration area and interpreters interested in applying the videotape object segmentation styles to their problems.

Jyoti Kini, Fahad Shahbaz Khan, Salman Khan, Mubarak Shah [9], this article presents an innovative approach to self-supervised video object segmentation, addressing the growing need for automated video object segmentation. This is a critical element in computer vision, serving an essential function in applications like video editing and autonomous navigation. In this paper, the authors propose a novel method that leverages self-supervised learning techniques to achieve accurate and robust video object segmentation. This innovative approach allows the model to learn from unlabeled video data, reducing the need for extensive manual annotations, which is a critical advantage in practical applications. The proposed method is extensively evaluated on various benchmark datasets, demonstrating impressive results in terms of video object segmentation accuracy and generalization across diverse scenarios. The authors also highlight the potential real-world applications of their approach, emphasizing its relevance in fields such as video editing, robotics, and surveillance.

Yadang Chen, Duolin Wang, Zhiguo Chen, Zhi-Xin Yang, and Enhua Wu [10], videotape object segmentation, which aims to draw a detailed object mask on videotape frames, is extensively applicable to colourful fields similar as autopilots, videotape editing, and videotape conflation. Propagation- grounded styles use the target's temporal consonance, and calculate on the mask from former frames. To clarify, Mask Track merges the segmentation mask from the preceding frame with the current frame, creating the mask for the current frame. Still, these styles suffer from occlusion problems and error drift. Matching- grounded styles uses the first frame of a given videotape as a reference frame and descry the segmented object singly in each frame. These styles are more robust and reduce the impact of occlusion, but don't take full advantage of spatiotemporal information.

Volume 1, Issue 7 (2024)
Article No. COMP_20
PP 1-11

Consequently, the performance and delicacy of some mongrel algorithms are bettered on the former two classes.

Yue Wu HKUST, Rongrong Gao HKUST, Jaesik Park POSTECH, Qifeng Chen HKUST [11], Can an AI system predict a photo or create a plastic video based on a single visual observation? With an accurate videotape vaticinator model, an intelligent agent can plan its stir according to the prognosticated videotape. Unborn video generation ways can also be used to synthesize a long videotape by constantly extending the future of the videotape. Utmost being styles attack the videotape vaticinator task by generating unborn videotape frames one by one in an unsupervised fashion. These approaches synthesize unborn frames at the pixel position without unequivocal modelling of the movements or semantics of the scene. Therefore, it's difficult for the model to grasp the conception of object boundaries to produce different movements for different objects.

Haidi Zhu, Haoran Wei, Baoqing Li, Xiaobing Yuan and Nasser Kehtarnavaz [12], Videotape object discovery involves detecting objects using videotape data as compared to conventional object discovery using static images. Two operations that have played a major part in the growth of videotape object discovery are independent driving and videotape surveillance. In 2015, videotape object discovery came a new task of the Image Net Large Scale Visual Recognition Challenge (ILSVRC2015). In general, object discovery approaches can be grouped into two major orders one- stage sensors and two- stage sensors. One- stage sensors are frequently more computationally effective than two- stage sensors. Still, two- stage sensors are shown to produce advanced rigor compared to one- stage sensors. Still, using object discovery on each image frame doesn't take into consideration the following attributes in videotape data Since there live both spatial and temporal correlations between image frames, there are point birth redundancies between conterminous frames. Detecting objects from poor quality frames leads to low rigor. Videotape object discovery approaches attempt to address the above challenges. Some approaches make use of the spatial-temporal information to ameliorate delicacy, similar as fusing features on different situations.

Weiguo YI and BO Wang [13], In the realm of underwater operations, difficulties arise due to the complex underwater environment, varying lighting conditions, and the presence of small targets. The identification of underwater targets is crucial for locating and acknowledging objects, a task commonly accomplished using technologies like Remotely Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs). However, persistent issues such as accuracy in detecting small targets, false positives, and missed detections require attention. To address these challenges, the research proposes an enhanced algorithm for the bio-identification of small underwater targets, named UWSC-YOLOv7, which is based on the YOLOv7 detection framework. Modifications to the network include channel feature information augmentation, multiscale check detection, and adjustments to the loss function, resulting in improved performance metrics such as mean Average Precision (mAP). The UWSC-YOLOv7 network demonstrates promise for enhancing underwater small target identification, supporting applications ranging from underwater biological monitoring to deep-sea research and shallow-sea farming. The evolution of target detection methods, particularly those based on deep learning, continues to show improvement over the years, with algorithms like CNN, SSD, and YOLO gaining prominence. The ongoing difficulty revolves around identifying small targets, worsened by issues like poor illumination and obstacles causing occlusion. Recent progress, incorporating contextual attention mechanisms and specific area criteria, seeks to tackle these issues. Despite advancements, improving networks for detecting small targets continues to be a key focus for future research endeavours.

U. Sirisha1, S. Phani Praveen2, Parvathaneni Naga Srinivasu2, Paolo Barsocchi5, Akash Kumar Bhoi [14], Computer vision is a dynamic field striving to equip machines with the ability to comprehend complex visual data. Object detection stands out as a significant challenge within this domain, involving the identification and localization of objects in images or videos. Despite initial hurdles, deep learning emerged as a prominent subfield of machine learning and AI in the early 2000s, thanks to advancements in artificial neural networks and support vector machines. The accessibility of large datasets and powerful computing resources since 2006 has notably boosted the adoption of deep learning techniques. Object detection encompasses various applications, including stock value prediction, speech recognition, and intrusion detection. It relies on a combination of algorithms and sophisticated learning structures to precisely recognize and classify items across diverse data formats like images, videos, and audio. Machine learning models, which include support vector machines, decision trees, and random forests, automate the identification and location of objects through techniques like feature extraction and classification. In contrast, deep learning models such as CNNs, R-CNNs, SSDs, and YOLO models use multiple layers of processing units to efficiently extract intricate features for accurate object detection. The training process for deep learning models in object detection involves utilizing extensive labeled datasets, where each object is annotated with its class label and bounding-box coordinates. These models are pivotal in various practical applications, including autonomous driving, surveillance, and robotics. Object

Volume 1, Issue 7 (2024)
Article No. COMP_20
PP 1-11

detection tasks encompass classification, localization, and segmentation. The objective is to comprehensively understand objects in images or videos by simultaneously performing all three tasks. Particularly, segmentation ensures precise boundary delineation, proving essential in scenarios requiring detailed object analysis, such as medical imaging or satellite imagery analysis. Object detection continues to progress, driven by advancements in both traditional machine learning methods and deep learning architectures.

Hangyue Zhao,Hongpu Zhang,Yanyun Zhao [15], The author aims to address the challenges faced in object detection tasks in maritime scenarios, particularly in Search and Rescue (SAR) missions and autonomous ship navigation. They highlight the limitations of existing object detection models in detecting small objects in maritime drone capture scenarios. To overcome these limitations, the author proposes an improved version of the YOLOv7 model called YOLOv7-sea, incorporating the SimAM attention module and additional components for detecting small objects. The paper discusses various strategies employed to enhance detection performance, such as Test Time Augmentation (TTA) and Weighted Box Fusion (WBF). Experimental results demonstrate the effectiveness of the proposed approach, achieving significant improvements in object detection accuracy on maritime UAV imagery. The study contributes to advancing computer vision methods in marine and freshwater fields, with potential applications in SAR missions and marine surveillance tasks.

III. ANALYSIS TABLE

Table 1 Analysis Table

Title	Summary	Advantages	Open Challenges
YOLO-Based Light-Weight Deep Learning Models for Insect Detection System with Field Adaption [1]	The most inconceivable diversity, cornucopia, spread, and rigidity in biology are set up in insects. The foundation of nonentity study and pest operation is nonentity recognition. Still, utmost of the current nonentity recognition exploration depends on a small number of nonentity taxonomic experts. We can use computers to separate insects directly rather of professionals because of the quick advancement of computer technology.	The deep feature extraction function makes it perform well in image, audio, and text data. Easy to update data through back propagation. Different architectures are suitable for different problems. The hidden layer reduces the dependence of algorithm on feature engineering.	One of the main challenges of deep learning is the need for large amounts of data and computational resources. Neural networks learn from data by adjusting their parameters to minimize a loss function, which measures how well they fit the data.
Rapid Prototyping of Pear Detection Neural Network with YOLO Architecture in Photographs [2]	Estimating fruit yield and predicting outcomes are crucial steps in making informed decisions in agribusiness, aiming to enhance efficiency in both fruit cultivation and sales. The prediction of yield relies on concrete data obtained through regular yield assessments.	One of the main advantages of YOLO v7 is its speed. It can process images at a rate of 155 frames per second, much faster than other state-of-the-art object detection algorithms. Even the original baseline YOLO model was capable of processing at a maximum rate of 45 frames per second.	YOLO might encounter difficulties with compact or intersecting items because it has limitations in predicting a set number of boxes per cell. It may also miss some objects that do not fit well into the grid, or produce false positives for background regions.
Object detection using YOLO:	The paper addresses key challenges in object detection,	The paper likely provides a	Some open challenges in this context include

challenges, architectural successors, datasets and applications [3]	such as handling objects of different sizes, dealing with class imbalances, and the need for extensive annotated datasets. It explores the evolution of YOLO's architecture and how subsequent versions have enhanced object detection capabilities.	comprehensive overview of object detection using the YOLO (You Only Look Once) framework. It may cover the challenges faced in object detection tasks, the evolution of YOLO architecture, available datasets, and real-world applications.	addressing object detection difficulties like handling diverse object sizes and class imbalance, as well as the need for large annotated datasets.
Video Object Detection through Traditional and Deep Learning Methods [4]	The paper explores video object detection techniques, covering traditional computer vision methods and deep learning approaches. The paper provides insights into the evolving landscape of video object detection, highlighting the pros and cons of both approaches and discussing practical applications and challenges in this critical field of computer vision.	It delivers a comprehensive overview of video object detection techniques, encompassing both traditional computer vision methods and modern deep learning approaches. This breadth of coverage enables readers to develop a holistic understanding of the field and make informed decisions about methodology selection.	Video object detection presents several challenges in achieving real-time performance, maintaining object tracking consistency amid occlusions and motion, handling scale and viewpoint variations, and detecting multiple objects simultaneously.
Detecting Moving Object via Projection of Forward Backward Frame Difference [5]	- Introduce the importance of moving object detection in computer vision and video analysis. Explain the specific approach of "Projection of Forward-Backward Frame Difference" and its significance.	Provide an overview of significant research papers, projects, or case studies that have contributed to the development and understanding of this specific technique. Cite and summarize these influential sources.	Discuss the challenges and limitations associated with the "Projection of Forward-Backward Frame Difference" method, such as handling complex scenes or dynamic lighting conditions.
Image and video-based crime prediction using object detection and deep learning [6]	By harnessing computer vision and state-of-the-art object detection methods, the authors aim to automatically predict crimes from image and video data. Researchers and practitioners in the field can benefit from the findings and methodologies presented in this paper.	Firstly, it represents an innovative and advanced approach to crime prediction, harnessing the capabilities of computer vision and deep learning algorithms. Secondly, the research conducted at MATSI in Morocco demonstrates the potential to automate the crime prediction process using image and video data, which can	Ensuring the accuracy and reliability of object detection in complex real-world environments, where lighting, weather, and object occlusions can affect results, remains a significant hurdle. Second, handling privacy and ethical concerns related to the use of image and video data in crime prediction systems is essential. Third, the scalability and real-time

Volume 1, Issue 7 (2024)
Article No. COMP_20
PP 1-11

		significantly improve the efficiency and accuracy of law enforcement efforts.	processing capabilities of such systems must be addressed to make them practical for law enforcement agencies.
A topological solution to object segmentation and tracking. [7]	The World consists of elements such as objects, the ground, and the sky. To visually perceive these objects, one must address two fundamental challenges: dividing visual input into distinct units and keeping track of the identities of these units despite alterations in appearance caused by object distortion, shifting perspectives, and dynamic occlusion.	Topological solutions for object segmentation and tracking offer several advantages, which make them a promising approach in the field of computer vision and image analysis.	Address the current challenges in topological object segmentation and tracking. Suggest potential future directions for research in this field, such as combining TDA with deep learning techniques or developing real-time topological solutions.
Deep learning for video object segmentation: a review. [8]	- Introduce the significance of video object segmentation and its applications in computer vision, robotics, and autonomous systems. Explain the importance of deep learning techniques in advancing video object segmentation. Provide an overview of the foundational concepts of deep learning, including neural networks and convolutional neural networks (CNNs).	Deep learning for video object segmentation offers several advantages, making it a prominent approach in computer vision and video analysis. They can learn complex features and temporal dependencies in video sequences, leading to precise segmentation results.	Describe commonly used datasets for video object segmentation, such as DAVIS, SegTrack, and YouTube-Objects. Highlight benchmark challenges that have driven the development of deep learning algorithms in this domain.
Self-Supervised Video Object Segmentation via Cutout Prediction and Tagging [9]	This journal paper introduces a promising self-supervised approach for video object segmentation, offering a valuable contribution to the field of computer vision. The combination of cutout prediction, tagging, and a novel loss function enhances the model's ability to segment objects accurately in videos, opening up opportunities for more efficient and automated video analysis in various applications.	By integrating cutout prediction and tagging mechanisms, the method enhances object boundary understanding and tracking performance, making it robust across diverse video scenarios. These innovations hold the potential to significantly improve video analysis applications, including video editing, robotics, and surveillance, by automating and enhancing object segmentation tasks.	Achieving a balance between segmentation accuracy and computational efficiency is essential, particularly in robotics and autonomous systems demanding low-latency processing. Scaling the approach to handle extensive video datasets and deploying it in resource-constrained environments poses challenges in data management, model optimization, and hardware constraints.

Volume 1, Issue 7 (2024)
Article No. COMP_20
PP 1-11

Global video object segmentation with spatial constraint Module [10]	This article presents a semi-supervised video object segmentation network that is both lightweight and efficient, utilizing the spatiotemporal memory framework. The method integrates a Global Environment (GC) module to attain real-time segmentation with superior performance.	CNN extracts features automatically. Single feature extraction by CNN. The object detection rate is high. Adding feature pyramid model, and good for small object detection. Based on Google Net, fast in speed.	Overall, our solution is efficient and compatible, and we hope it will set a strong baseline for other real-time video object segmentation solutions in the future.
Future Video Synthesis with Object Motion Prediction [11]	We present an approach to prognosticate unborn videotape frames given a sequence of nonstop videotape frames in the history. With this procedure, our system exhibits much lower tearing or deformation artefact compared to other approaches.	They can be used in interior and exterior passage areas. While they are particularly useful in spaces with many people, they can also be used in certain areas of homes or residential areas such as entryways or gardens.	For future frame prediction. Our method produces future Frames by firstly decomposing possible moving objects into Currently-moving or static objects.
A Review of Video Object Detection: Datasets, Metrics and Methods [12]	Though there are well-established techniques for identifying objects in static images, employing them on video data frame by frame faces two challenges. These challenges encompass restricted computational efficiency arising from redundancy among image frames or the failure to exploit the temporal and spatial correlation of features across frames. Additionally, there is a difficulty in addressing real-world conditions like motion blur and occlusion.	It allows us to be more efficient and to focus on some other task while the machine works on its own. Object Detection is a key task of Computer Vision. Thanks to it, the machine is able to locate, identify and classify an object in an image.	Challenges still remain for further improving the accuracy and speed of the video object detection methods. This section presents the major challenges and possible future trends as related to video object detection.
Research on Underwater Small Target Detection Algorithm Based on Improved YOLOv7 [13]	An improved underwater small target detection technique utilizing YOLOv7 is proposed to address issues of high miss detection rate and poor underwater scene recognition. Through simulation tests, it outperforms other conventional target detection techniques by enhancing accuracy and reducing false detections.	Increased Accuracy: By merging SENet attention mechanism, enhancing FPN network topology, and incorporating EIoU loss function, the technique improves detection accuracy. Reduced Model Complexity: The technique focuses on crucial feature information of small	Real-world Application: The proposed technique can be applied in real underwater operations, supporting human activities like biological monitoring and fishing. Specialized Uses: Beyond underwater operations, the network's capabilities can be extended to deep-sea research and shallow-sea

		targets, reducing model complexity while maintaining accuracy. Faster Detection: Utilizing YOLOv7 network as the basic network enables high detection speed alongside accuracy.	farming, broadening its potential applications.
Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection [14]	This research explores the realm of object detection within computer vision, with a specific emphasis on the progress made in deep neural networks over the last ten years. The investigation involves a comparison between two main categories of object detectors, an examination of the significance of detection accuracy and inference speed, and an exploration of different YOLO variations along with their corresponding performance metrics.	Enhanced Effectiveness: The detection accuracy of deep neural networks, especially YOLO and its variations, has been notably improved, surpassing that of certain two-stage detectors in specific situations. Optimized Efficiency: YOLO detectors prioritize rapid inference, making them well-suited for applications that require real-time processing, such as autonomous driving and surveillance.	Further Enhancements: Continuous advancements in deep learning techniques and architectures will likely lead to even better object detection models with improved accuracy and efficiency. Diverse Applications: Object detection has broad applications across various domains, including autonomous systems, surveillance, robotics, medical imaging, and satellite imagery analysis, indicating a growing need for robust and versatility of the detection algorithms.
YOLOv7-sea: Object Detection of Maritime UAV Images based on Improved YOLOv7 [15]	This article primarily discusses the object detection task within this context, highlighting the challenges faced in maritime scenarios. To address these challenges, an improved version of the YOLOv7 model, named YOLOv7-sea, is proposed. This model incorporates the SimAM attention module to enhance feature extraction and introduces additional components for detecting small objects. Various strategies are employed to enhance detection performance, including Test	Enhanced YOLOv7-sea Model Marine Computer Vision SAR and maritime surveillance, aiming to detect humans, boats, and various objects in open water	Existing models may not directly suit maritime drone capture scenarios due to small object detection difficulties and sea surface interference, indicating the need for specialized approaches. While datasets like Sea Drones See provide valuable training and evaluation data, challenges related to data collection, data annotation, and domain adaptation persist, requiring ongoing efforts to address dataset biases and limitations for

	Time Augmentation (TTA) and Weighted Box Fusion (WBF).		improved model performance.

IV. CONCLUSION

The YOLOv8-CVAT framework, encapsulated within an Android application, stands as a beacon of technological innovation with successful implementation of YOLOv8-CVAT within an Android application propelled the boundaries of wildlife monitoring but has also laid the foundation for a transformative paradigm in conservation efforts [8]. Initially tailored for Indian exclusive species, reveals its potential for cross-domain application by seamlessly integrating with various other datasets. This work sparks further collaborations, adaptations, and innovations, leading to a future where advanced technology becomes an integral ally in the ceaseless effort to protect and preserve the diverse and unique wildlife that graces our planet.

REFERENCES

[1] Kumar, N.; Nagarathna; Flammini, F. YOLO-Based Light-Weight Deep Learning Models for Insect Detection System with Field Adaption. Agriculture 2023, 13, 741. https://doi.org/10.3390/ agriculture13030741

[2] Sergejs Kodors, Marks Sondors, Gunārs Lācis, Edgars Rubauskis, Ilmārs Apeināns, Imants Zarembo, Environment. Technology. Resources. Rezekne, Latvia Proceedings of the 14th International Scientific and Practical Conference. Volume 1, 81-85, 2023

[3] Tausif Diwan1 & G. Anirudh2 & Jitendra V. Tembhurne1, "Object detection using YOLO: challenges, architectural successors, datasets and applications", Department of Computer Science & Engineering, Indian Institute of Information Technology, Nagpur, India and Department of Data science and analytics, Central University of Rajasthan, Jaipur, Rajasthan, India, 2022

[4] Sita M. Yadav, Sandeep M. Chaware, "Video Object Detection through Traditional and Deep Learning Methods", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958 (Online), Volume-9 Issue-4, April, 2020

[5] Zhuo Bian1 and LiangliangWang2, "Detecting Moving Object via Projection of Forward-Backward Frame Difference", International Journal of Future Generation Communication and Networking Vol. 9, No. 5 (2016), pp. 143-152 http://dx.doi.org/10.14257/ijfgcn.2016.9.5.14

[6] Mohammed Boukabous, Mostafa Azizi, "Image and video-based crime prediction using object detection and deep learning", Bulletin of Electrical Engineering and Informatics Vol. 12, No. 3, June 2023, pp. 1630~1638

[7] Thomas Tsaoa and Doris Y. Tsao, "A topological solution to object segmentation and tracking", National Academy of Sciences elected in, 2020

[8] Mingqi Gao1,2 · Feng Zheng2 · James J. Q. Yu2 · Caifeng Shan3 · Guiguang Ding4 · Jungong Han, "Deep learning for video object segmentation: a review", Artifcial Intelligence Review (2023) 56:457–531, April 2022

[9] Jyoti Kini, Fahad Shahbaz Khan, Salman Khan, Mubarak Shah, "Self-Supervised Video Object Segmentation via Cutout Prediction and Tagging", MBZ University of Artificial Intelligence UAE, 22 April 2022

[10] Yadang Chen1 , Duolin Wang1 (), Zhiguo Chen1 , Zhi-Xin Yang2 , and Enhua Wu, "Global video object segmentation with spatial constraint module", Computational Visual Media https://doi.org/10.1007/s41095-022-0282-8, 2022

[11] Yue Wu, Rongrong Gao, Jaesik Park, Qifeng Chen, "Future Video Synthesis with Object Motion Prediction", arXiv:2004.00542v2 [cs.CV] 15 Apr 2020

[12] Haidi Zhu 1,2, Haoran Wei 3 , Baoqing Li 1,* , Xiaobing Yuan 1 and Nasser Kehtarnavaz, "A Review of Video Object Detection: Datasets, Metrics and Methods", Appl. Sci. 2020, 10, 7834

[13] Weiguo YI and BO Wang, "Research on Underwater Small Target Detection Algorithm Based on Improved YOLOv7", date of publication 28 June 2023, Digital Object Identifier 10.1109/ACCESS.2023.3290903

[14] U. Sirisha1 • S. Phani Praveen2 • Parvathaneni Naga Srinivasu2 • Paolo Barsocchi5 • Akash Kumar Bhoi, "Statistical Analysis of Design Aspects of Various YOLO-Based Deep Learning Models for Object Detection", International Journal of Computational Intelligence Systems (2023) 16:126

Volume 1, Issue 7 (2024)
Article No. COMP_20
PP 1-11

[15] Hangyue Zhao,Hongpu Zhang,Yanyun Zhao, "YOLOv7-sea: Object Detection of Maritime UAV Images based on Improved YOLOv7", 2023 IEEE/CVF Winter Conference on Applications of Computer Vision Workshops (WACVW), Waikoloa, HI, USA, 2023, pp. 233-238, doi: 10.1109/WACVW58289.2023.00029