International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

UAV- Based Object Detection and Tracking: Real-World Applications

***Gadi Jagadeesh*[1]**

[1]***B. tech Student, GMR Institute of Technology, Rajam,532127, India.***

ABSTRACT :

The integration of Unmanned Aerial Vehicles (UAVs) has propelled advancements in object detection, shaping diverse applications across precision agriculture, environmental monitoring, disaster response, and surveillance. The overview of the evolution, challenges, and future directions in UAV-based object detection and tracking.The survey delves into the effectiveness of deep learning approaches for UAV object detection, with a particular focus on applications such as environmental monitoring, precision agriculture, and traffic management. Comparative studies, including mid-to late-season weed detection, underscore the efficacy of models like Faster RCNN and Single Shot Detector. Convolutional Neural Networks (CNNs), including Retina Net, Faster R-CNN , and SSD, exhibit remarkable capabilities in object detection from UAVs. The integration of deep learning extends to online Multi-Object Tracking (MOT) using a Hierarchical Deep High-resolution network (HDHNet). Efficiency and real-time applicability take center stage in UAV applications, with the introduction of an efficient real-time object detection model showcasing superior performance. Addressing uncertainties in outdoor surveillance involves leveraging probabilistic-based motion planning and Partially Observable Markov Decision Processes (POMDP). Ongoing challenges in UAV-based object detection and tracking, such as small object detection and real-time processing, persist. Unified approaches like You Only Look Once (YOLO) redefine real-time object detection, presenting it as a regression task. While excelling in various aspects, challenges in handling small objects in groups necessitate ongoing refinement. Datasets contributions, including the P-DESTRE dataset tailored for long-term pedestrian re-identification research, enrich the field. Comparison of different datasets and algorithms models of the each paper the accuracy values are getting different for different papers .The YOLO V3 + R-CNN model, utilizing the HERIDAL dataset , achieved the highest accuracy of 95.11% among the models analyzed in object detection tracking system.

KEYWORDS :- UAV, drone, computer vision, VisDrone datasets, convolutional neural networks (CNN), Region-Based Conventional neural network (RCCN),Multi-Object Tracking(MOT).

1. INTRODUCTION

In recent years, the integration of unmanned aerial vehicles (UAVs) has propelled advancements in various fields, ranging from precision agriculture and environmental monitoring to disaster response and surveillance. A critical aspect of maximizing the potential of UAVs lies in the development of robust object detection and tracking algorithms. These algorithms play a pivotal role in enabling UAVs to autonomously perceive and interact with their surroundings, making them indispensable in applications such as search and rescue, surveillance, and precision agriculture.The foundation for progress in this domain often rests upon benchmark datasets that facilitate rigorous evaluation of detection and tracking algorithms. The "Detection and tracking meet drones challenge" paper by Zhu et al.[1] introduces the VisDrone datasets, a seminal contribution providing a large-scale collection of drone-captured data. This datasets has since become a cornerstone for researchers, fostering a comprehensive evaluation of visual analysis algorithms tailored specifically for UAV platforms.By encompasses a diverse array of research endeavors, each shedding light on distinct aspects of object detection and tracking for drones. Ajaz.et al[2]develop into deep learning-based approaches for UAV object detection and tracking, exploring applications in environmental monitoring, precision agriculture, and traffic management. Veeranampalayam Sivakumar et al.[3] contribute insights into mid-to late-season weed detection using UAV imagery, evaluating object detection models like Faster RCNN and Single Shot Detector (SSD).Wang.et al[4] pivot the focus to the utilization of Convolutional Neural Networks (CNN) for object detection from UAVs, emphasizing the efficacy of RetinaNet, Faster R-CNN, and SSD. The integration of deep learning into UAV systems is further explored by Huang et al[5], who propose an online Multi-Object Tracking (MOT) approach utilizing a Hierarchical Deep High-resolution network (HDHNet).Efficiency and real-time applicability become paramount considerations in UAV applications. Vaddi et al[6]. introduce an efficient object detection model tailored for real-time UAV applications, demonstrating superior performance using the VisDrone'18 datasets. Sandino et al[7] ,tackle uncertainties in outdoor surveillance by leveraging probabilistic-based motion planning and Partially Observable Markov Decision Processes (POMDP) in their framework.The exploration of UAV-based object detection and tracking extends beyond traditional domains, encompassing applications in disaster response, precision agriculture, and surveillance.The landscape of object detection methodologies has witnessed paradigm shifts with the introduction of unified approaches like You Only Look Once (YOLO), as outlined by Han.et al[8]. YOLO's architecture, framing object detection as a regression task, allows for real-time processing with higher mean average precision, albeit with spatial constraints on bounding box predictions. Despite its successes, challenges arise when dealing with small objects in groups, underscoring the need for ongoing refinement. Notably, the work by Kumar.et al [10] introduces the P-DESTRE dataset, a pivotal contribution to the field, specifically tailored for long-term pedestrian re-identification research. The dataset addresses gaps in existing datasets by providing consistent ID annotations across different days, thereby fostering advancements in pedestrian detection, tracking, and re-identification techniques.Vujasinović.et al[11] delve into the challenges of visual object tracking from UAVs in a 3D environment. While prevailing tracking designs on ground-level benchmarks have been well-established, the integration of 3D information into detection-by-tracking algorithms for UAV onboard visual object tracking presents novel challenges and opportunities.The ubiquity of UAVs extends to disaster response scenarios, as demonstrated by Han.et al[12], who present an Energy-Efficient Tracking-based Action Detector (E^2TAD) that excelled in the UAV-Video Track of the Low-Power Computer Vision Challenge. The versatility of UAVs is further exemplified by Fradi.et al[14], who propose a color-based detection framework for autonomous person detection and tracking. it becomes evident that the challenges in UAV-based object detection and tracking are multifaceted. These challenges encompass small object detection, real-time processing, and the integration of 3D information into tracking algorithms. Additionally, researchers are exploring the fusion of thermal data, machine learning algorithms, and the development of comprehensive datasets to address specific use cases.Advancements in machine learning (ML) techniques have played a pivotal role in addressing complexities within visible light communication (VLC) systems. Saxena.et al[20] provide a comprehensive survey on ML applications in VLC, highlighting algorithms like DBSCAN and OPTICS to reduce design complexity and enhance network performance. While this reference might seem tangential, it underscores the interdisciplinary nature of research, where insights from one field contribute to the development of methodologies in another.

In summary, the comprehensive picture of the evolution, challenges, and future directions of UAV-based object detection and tracking. From dataset contributions, like P-DESTRE, to innovative algorithms like E^2TAD, each reference adds a unique layer to the narrative, collectively contributing to the growth and maturation of this dynamic field. As we delve deeper into the intricacies of each research endeavor, we unravel the intricacies of a field that not only extends the capabilities of UAVs but also pushes the boundaries of what is achievable in computer vision and machine learning.

2. Related work

***Zhu.et al(*2021)** discussed the object detection and tracking datasets and benchmarks, discussing the challenges of collecting large-scale drone-based datasets with manual annotations.It presents the VisDrone dataset, the largest drone captured dataset published to date, which enables extensive evaluation and investigation of visual analysis algorithms for the drone platform.The paper analyzes the current state of the field of large-scale object detection and tracking on drones, highlighting the advancements and challenges faced by different methods and algorithms.The paper proposes future directions for research and development in video analysis on drone platforms, aiming to further advance object detection and tracking algorithms [1].

***Ajaz.et al*(2022)** proposed a comprehensive survey on the research progress and prospects of deep learning (DL)-based unmanned aerial vehicle (UAV) object detection and tracking approaches in various UAV-related tasks such as environmental monitoring, precision agriculture, and traffic management.The methods in this paper have their own experimental environment, experimental data, and even source code, and the computation cost is directly related to the speed, GPU, and backbone model .The DL approaches in object detection and tracking in the remote sensing (RS) field were systematically analyzed according to three UAV topics: single object detection (SOD), video object detection (VID), and multiple object tracking (MOT).The paper provides a DL-based UAV object detection and tracking approaches, which can serve as a foundation for future research and development in this fiel*d.* [2]

***Veeranampalayam Sivakumar et al* (2020)** compares object detection-based convolutional neural network (CNN) models, specifically Faster RCNN and Single Shot Detector (SSD), for mid-to late-season weed detection in soybean fields using low-altitude unmanned aerial vehicle (UAV) imagery. The performance of these models is evaluated in terms of precision, recall, f1 score, and Intersection over Union (IoU).It compares the performance of the object detection models with a patch-based CNN model. It is found that the Faster RCNN model better to the patch-based CNN model in terms of weed detection performance and inference time.Highlights the importance of understanding the potential and identifying the algorithms for on-farm, near real-time weed detection and management.It provides insights into the potential and identification of algorithms for on-farm, near real-time weed detection and management using UAV imagery and CNN models.[3]

***Wang.et al* (2018)** discusses the use of Convolutional Neural Networks (CNN) for object detection from Unmanned Aerial Vehicles (UAVs) and mentions that CNN-based methods have been employed for this purpose in previous studies .The authors mention three representative CNN object detectors used in their study: RetinaNet, Faster R-CNN, and SSD.Highlight that RetinaNet, in particular, achieved state of the art performance in object detection from UAVs, indicating the effectiveness of CNN-based object detectors.The Stanford Drone Dataset (SDD) is used for the experiments, which contains different categories of objects such as pedestrians,bicyclists, cars, etc.The authors suggest that further investigation is needed for deep learning model compression to minimize the computation workload over UAVs .[4]

***Huang et al* (2021)** proposed an online Multi-Object Tracking (MOT) approach in the UAV system that integrates deep high-resolution representation network and data association method in a unified framework .The proposed approach utilizes a Hierarchical Deep High-resolution network (HDHNet) to handle different types and scales of targets and extract effective and comprehensive features during online learning .An adjustable fusion loss function is proposed by combining focal loss and GIoU loss to address class imbalance and hard samples . The tracking process involves applying the detection results to an improved DeepSORT MOT algorithm in each frame, which makes use of target appearance features for matching .Experimental results on the VisDrone2019 MOT benchmark demonstrate that the proposed UAV MOT system achieves the highest accuracy and robustness compared to state-of-the-art methods .The paper also mentions the hardware and software environment used for the experiments, including the CPU, GPUs, and software frameworks .Additionally, the paper describes the fusion loss functions used during the multi-task training process, including the GIoU loss and classification loss. [5]

***Vaddi et al* (2019)** proposed an efficient object detection model for real-time UAV applications, specifically focusing on object detection from the information captured by an on-board camera .The authors use the VisDrone'18 dataset for their studies, which contains different objects such as pedestrians, vehicles, and bicycles .The proposed model utilizes a deep feature pyramid architecture that captures both generic features (such as edge and color) and detailed features specific to the classes in the problem .The model is implemented with both Res Net and Mobile Net as convolutional bases, and it achieves a desirable performance of 30.6 mAP for object detection with an inference time of 14 fps .Comparisons with RetinaNet-ResNet-50 and HAL- RetinaNet show that the proposed model combined with Mobile Net as the back end feature extractor provides the best results in terms of accuracy, speed, and memory efficiency for real-time object detection with drones .The MobileNet backbone in the proposed model is well-suited for embedded system applications, offering a speedup in computing the feature map without sacrificing overall quality.[6]

***Sandino et al* (2022)** focuses on reducing object detection uncertainties in outdoor surveillance using UAVs .It highlights the limited cognition power of UAVs to autonomously interact with the environment and the reliance on human operators for understanding detections .The framework proposed in the paper utilizes a probabilistic-based motion planner and a Partially Observable Markov Decision Process (POMDP) to model the navigation problem in real-time .The motion planner incorporates color and thermal imagery to provide accurate victim localization coordinates and diminish false positive readings from vision-based object detectors.The paper also mentions the off board mode of autonomous navigation, which internally executes the POMDP-based motion planner without a predefined survey plan Preliminary flight tests using thermal imagery were conducted to test the scalability of the framework .[7]

***Han.et al* (2021)** proposedYOLO (You Only Look Once) is a new approach to object detection that frames the problem as a regression task, predicting bounding boxes and class probabilities directly from full images in one evaluation .YOLO's unified architecture allows for end-to-end optimization and real-time processing, with the base model processing images at 45 frames per second and the smaller version, Fast YOLO, processing an impressive 155 frames per second .Compared to other real-time detectors, YOLO has a higher mean average precision (mAP) and is less likely to predict false detections .YOLO imposes spatial constraints on bounding box predictions, limiting the number of nearby objects it can predict and making it less effective for small objects in groups, such as flocks of birds .YOLO's error metric accounts for the importance of deviations in different box sizes by predicting the square root of the bounding box width and height .YOLO unifies the components of object detection into a single neural network, enabling global reasoning about the full image and all objects within it .The YOLO network is pretrained on the Image Net dataset and achieves high accuracy, comparable to GoogLe Net models. [8]

***Zhang.et al* (2022)** discusses the use of discriminative correlation filters (DCFs) in visual tracking approaches for unmanned aerial vehicle (UAV) videos .The authors propose a robust DCF-based tracking framework that incorporates a pretrained rectification network for UAV-based remote sensing .The proposed framework includes a target-specific rectification network that is offline trained to classify the target and background, and a DCF module for fast inference and obtaining potential target locations .The authors also propose a robust and adaptive model update strategy by finetuning both the DCF module and rectification network based on the classification confidence of the estimated result .Experimental results on recent UAV benchmarks demonstrate that the proposed method outperforms other competing algorithms .The paper also mentions the use of deep CNN features from VGG-M and VGG-16 for target representation, and the introduction of a temporal smoothing term for the learned filter to improve robustness against distractors. [9]

***Kumar.et al* (2020)** provides a comprehensive literature review on UAV-based datasets, with a specific focus on datasets related to pedestrian detection, tracking, re-identification, and search The authors highlight the availability of various publicly available datasets for evaluating detection, tracking, and short-term re-identification techniques, but note the lack of datasets suitable for long-term pedestrian re-identification research .The P-DESTRE dataset introduced in the paper fills this gap by providing video sequences captured from UAVs for long-term pedestrian re-identification research, with consistent ID annotations across different days .The paper also compares the performance of state-of-the-art techniques in pedestrian detection, tracking, short-term re-identification, and long-term re-identification using the P-DESTRE dataset, as well as other well-known surveillance datasets .The authors identify the challenging factors and co-variates for UAV-based automated data analysis, such as motion blur, shadows, and occlusions caused by crowded environments, which should be considered in future advancements in this field .[10]

***Vujasinović.et al* (2020)** mentions that there has been significant progress in single visual object tracking in recent years, thanks to the availability of benchmarks.Most existing benchmarks for tracking algorithms focus on ground-level perspectives, resulting in state-of-the-art visual object trackers following the tracking-by-detection paradigm .Three prevailing tracking designs on these benchmarks are discriminative correlation filters, Siamese-based approach, and trackers inspired by correlation filters with small convolutional neural networks.In the field of autonomous driving, there are applications where objects are tracked in a 3D system of reference through a detection-by-tracking paradigm .The authors of the current paper propose an approach that integrates the 3D structure of the observed scene into a detection-by-tracking algorithm for UAV onboard visual object tracking.However, these tracking algorithms are tailored to ground-level perspectives and do not utilize 3D information. [11]

***Han.et al* (2022)** presents a tracking-based solution for video action detection, aiming to accurately and efficiently localize predefined key actions spatially and temporally .The solution consists of three core components: ball-person detection, deep association, and action detection. It also includes shape-texture debiased training and domain-invariant adversarial training to improve detection robustness .The solution won the first place in the UAV-Video Track of the 2021 Low-Power Computer Vision Challenge (LPCVC) .The challenge required contestants to spatio-temporally localize the key action of ball-catching from videos captured by drones, considering accuracy and efficiency .The solution addresses the challenges of lack of training data, robustness, and efficiency in the challenge .The code for the solution is available at a GitHub repository .[12]

Ding.et al **(2021)** proposes a multi-small target detection and tracking method based on improved YOLO and SIFT for drones .The algorithm is evaluated using the VisDrone 2019 MOT benchmark dataset, with leading evaluation indicators MOTA and MOTP being 38.7 and 75.7 respectively .The algorithm meets the complex task requirements of multi-target tracking for UAVs and performs better in secondary evaluation indicators IDF1, ML, and FN .The algorithm framework is implemented based on the Pytorch deep learning framework and the tests are performed on Intel (R) Core (TM) i5-2450M CPU 2.50GHZ, 4GB memory, NVIDIA GeForce GTX 1660 .The experiment includes YOLO v3 network pruning, UAV-YOLO performance evaluation, optimized SIFT feature small target detection, and drone multitarget tracking. [13]

Fradi.et al **(2018)** proposes a color-based detection framework for autonomous person detection and tracking using unmanned aerial vehicles (UAVs) .The approach includes the execution of control commands to switch from detection to automatically following the detected target .The proposed approach is evaluated on videos recorded by drones and demonstrates the effectiveness of accurately following a target in real-time, despite challenges such as lighting changes, speed, and occlusions .The primary objective of the paper is to build an accurate visual representation of the target that can handle changing environments and target appearances .The second objective is to enable following the target by mini-drones, which adds additional challenges to the tracking algorithm, particularly in terms of real-time requirements .The final bounding box of the target is defined as a rectangular shape, with the center position denoted by (Xk, Yk) and the width and height defined by Wk and Hk, respectively. [14]

Wu et al **(2021)** provides a comprehensive survey on the research progress and prospects of deep learning (DL)-based unmanned aerial vehicle (UAV) object detection and tracking approaches .The survey reviews various DL-based methods for UAV-related tasks such as environmental monitoring, precision agriculture, and traffic management .The paper also discusses DL-based solutions for multi-object tracking (MOT) and categorizes them into three mainstream methods: tracking by detection (TBD), single object tracking (SOT)-assisted method, and memory networks .The development of typical methods for MOT, including those specifically designed for UAV data. [15]

Paik and Kim et al **(2022)** mentions the use of SORT and DeepSORT as two-stage object trackers that combine object detection and identity association. SORT uses the Kalman filter and Hungarian algorithm for tracklet prediction and association .DeepSORT improves on SORT by incorporating appearance descriptors for matching detection information and tracks, reducing ID switching issues caused by occlusion .JDE introduces a shared model that combines detection elements and appearance features for re-identification, achieving real-time processing speed .FairMOT builds on JDE's paradigm and overcomes its weaknesses by implementing anchor-free detection, multi-layer feature aggregation, and low-dimensional features to improve re-ID features and scaling performance .[16]

Pan et al **(2019)** mentions the lack of publicly available large-scale drone-based benchmarks or datasets, which hinders the further development in drone-based visual data understanding .The authors refer to the VisDrone2018 dataset, which focuses on core problems in computer vision fields and the challenge workshop, Vision Meets Drone Video Object Detection and Tracking (VisDrone-VDT2018), which proposed plentiful methods for understanding drone-based visual data .The paper also mentions that there are a small number of datasets related to drone platforms, including a dataset for car counting and video sequences for object tracking .Overall, the literature survey in this paper highlights the limited availability of large-scale drone-based benchmarks or datasets, but mentions the VisDrone2018 dataset and other datasets that have contributed to the understanding of drone-based visual data. [17]

Dousai and Lončarić et al **(2022)** proposed Object detection in computer vision has been extensively researched over the past two decades, with several papers achieving exceptional results in detecting objects from ground images .However, detecting objects in aerial images presents additional challenges, such as small object detection and high resolution .Before the evolution of convolutional neural networks, the Viola-Jones (VJ) method proposed in 2001 showed impressive results for face recognition in real-time .Some researchers have explored the use of thermal imaging techniques, specifically thermal infrared (TIR) cameras, to detect humans in aerial images .Burke et al. discussed the limitations and requirements of thermal object detection for effective search and rescue in marine and coastal environments . Doherty et al. presented a simple hardware-based onboard model using thermal and color imagery for human detection.[18]

Jadhav et al **(2020)** proposed Object detection algorithms have achieved significant advances with deep learning techniques in recent years .However, these algorithms are not usually optimal for sequences or images captured by drone-based platforms due to challenges such as view point change, scales, density of object distribution, and occlusion .The paper proposes a model for object detection in drone images using the VisDrone2019 DET dataset .The RetinaNet model is used as the base, with modified anchor scales to handle dense distribution and small object sizes .The paper also introduces Squeeze-and-Excitation (SE) blocks to model channel inter dependencies and adaptively re calibrate feature responses, resulting in improved performance. [19]

Saxena et al **(2023)** provides a detailed review of several machine learning (ML) algorithms to reduce the design complexity of indoor VLC transmission, as well as ML applications in different design aspects to improve system performance .It comprehensively reviews various ML algorithms for reducing the computational complexity in indoor VLC transmission and ML applications in different design requirements to increase network performance .The paper highlights some of the most widely used ML algorithms in the VLC network, including DBSCAN and OPTICS .It also addresses the challenges and future research directions based on machine learning algorithms in VLC.The paper emphasizes the need for further investigation of ML techniques for different real-time VLC application scenarios in the future, particularly in the context of 5G and beyond.[20]

3.METHODOLOGY

3.1 UAV Data Collection:-

Mainly used for the weed detection using the unmanned aerial vehicle (UAV). The Zenmuse X5R camera used is a 16 megapixel camera with 4/3” sensor and 72 degree diagonal field of view. The dimension of the captured images is 4608 × 3456 pixels in three bands—Red, Green, and Blue. To develop an economical solution ,In study focuses on only using RGB imagery. By a 20-m altitude, for the given sensor specifications, the spatial resolution of the output image is 0.5 cm/pixel. DJI Ground Station pro software was used for flight control.

Common weed species at the experimental site were waterhemp (Amaranthus tuberculatus), Palmer Amaranthus (Amaranthus palmeri), common lambs quarters (Chenopodiam album), velvetleaf (Abutilon theophrasti), and foxtail species such as yellow and green foxtails.

3.2 Data Annotation and Processing:-

The Main objective of the study is to develop a weed detection system with on-farm data processing capability.

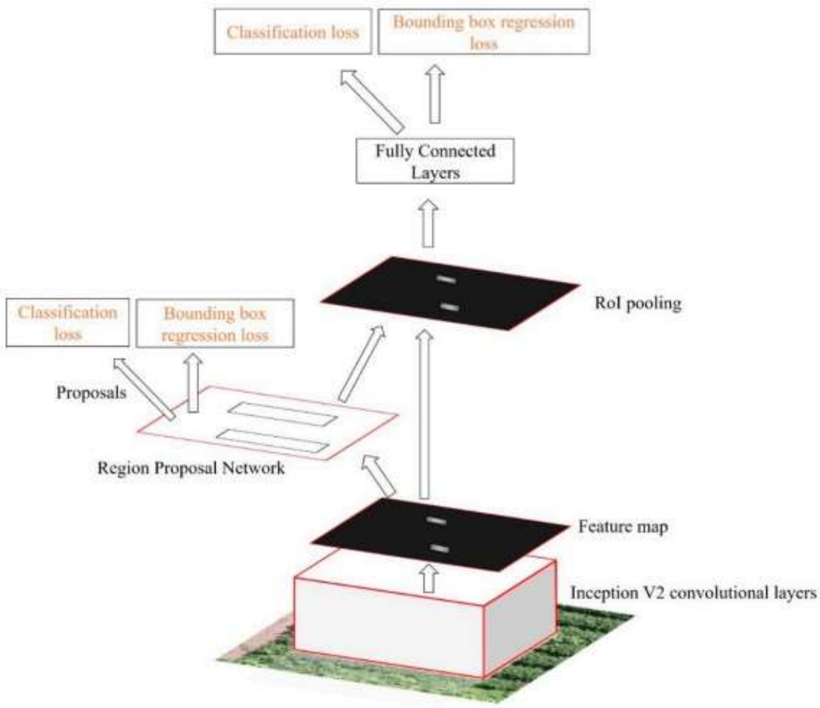


Fig 1 RCNN Architecture

From the fig 1 The Faster Region-based Convolutional neural network (RCNN) model for mid-to late-season weed detection in soybean fields is based on the Faster RCNN architecture, which utilizes the convolutional layers from Inception v2 for feature extraction .The Inception v2 network is known for its translation and scale invariance, achieved by using wider networks with filters of different kernel sizes in each layer .The Faster RCNN architecture includes a region proposal network that defines anchors or fixed boundary boxes at each location, enabling scale-invariant proposals .The region proposal layer uses a convolutional filter to output a confidence score for object and background classes, as well as regression offsets for anchor boxes .The model uses RoI (Region of Interest) pooling to resize bounding box regions of different sizes and aspect ratios to fixed size outputs using max pooling .The Faster RCNN model in this study used input images resized to a fixed size of 1024 x 1024 pixels and employed 4 different scales and 3 different aspect ratios for region proposals, resulting in a total of 12 anchors at each location.

3.3 Faster RCNN:

Region-based Convolutional neural network (RCNN) is a region proposal method-based object detection algorithm that was used for mid-to late-season weed detection in soybean fields using UAV imagery. The Faster RCNN model consists of four sections: the feature extractor, the region proposal network, Region of Interest (RoI) pooling, and classification. It uses the convolutional layers from Inception v2 for feature extraction and fully convolutional Region Proposal Networks (RPN) for proposing better object regions.The performance of the Faster RCNN model was evaluated in terms of precision, recall, f1 score, and mean Intersection over Union (IoU The inference time of the Faster RCNN model for a 1152 x 1152 image was 0.23 seconds.The Faster RCNN model yielded better weed detection performance compared to a patch-based CNN model and had similar inference time.

3.4 Single Shot Detector :

The Single Shot Detector (SSD) model was used for mid-to late-season weed detection in soybean fields using UAV imagery .The performance of the SSD model was evaluated and compared with the Faster RCNN model in terms of precision, recall, f1 score, and mean Intersection over Union (IoU). The optimal confidence threshold of the SSD model was found to be lower than that of the Faster RCNN model, indicating potentially lower generalization performance for mid-to late-season weed detection .The inference time of the SSD model for a 1152 x 1152 image was slightly faster than the Faster RCNN model .Overall, the Faster RCNN model yielded better weed detection performance and had similar inference time compared to the SSD model.

3.5 Evaluation Metrics :

Precision, recall, f1 score, and Intersection over Union (IoU) are the evaluation metrics used in this study.

$$\text{precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

$$IOU = \frac{Area\ of\ Intersection\ of\ two\ boxes}{Area\ of\ Union\ of\ two\ boxes}$$

Here TP refers to True Positive, FP refers to False Positive, and FN refers to False negative. Moreover, mean Average Precision (mAP) is another metric that is commonly used in object detection problems. Hence, IoU here represents the ratio between the intersection of all positive prediction boxes (true positive and false positives in object detection terms) and all ground truth boxes in an image.

CASE STUDY

Enhancing On-Farm Weed Detection in Soybean Fields through UAV-Based Object Detection Models.Investigated the efficacy of two object detection models, Faster RCNN and Single Shot Detector (SSD), for mid- to late-season weed detection in soybean fields using UAV imagery.Both models demonstrated comparable weed detection performance with precision, recall, f1 score, and Intersection over Union (IoU) values around 0.65-0.68.Faster RCNN showed a slight advantage over SSD in terms of generalization performance.Faster RCNN outperformed a patch-based CNN model in both weed detection accuracy and inference speed.

Fig 3

Image Preprocessing:

The image preprocessing step is important because it can help to improve the accuracy of the object detection pipeline. By cleaning up the image and removing noise, we can make it easier for the feature extractor to extract relevant features. We may also want to resize the image to a standard size or normalize the color channels. This can help to improve the performance of the neural networks used in the pipeline.

Feature Extraction:

The feature extraction step is the heart of the object detection pipeline. This is where we extract relevant features from the image that can be used to represent the objects in the image and to distinguish them from each other. Some common features used in object detection include color, texture, shape, and spatial relationships.There are many different ways to extract features from an image. One common approach is to use convolutional neural networks (CNNs). CNNs are a type of neural network that is well-suited for feature extraction from images. CNNs work by learning to extract features from images at different levels of abstraction.

Region Proposal Network (RPN):

The Region Proposal Network (RPN) is a neural network that generates candidate object regions in the image. These regions are called "proposals" and they represent where the RPN thinks objects might be located.The RPN works by sliding a window over the image and generating a proposal for each window location. The RPN then classifies each proposal as an object or not-an-object. The RPN also regresses the bounding box of the object in the proposal.

Fast RCNN:

The Fast RCNN is a neural network that takes the proposals from the RPN and classifies them as objects or not-objects. The Fast RCNN also regresses the bounding boxes of the objects in the proposals.The Fast RCNN works by extracting features from each proposal and then passing those features to a fully connected neural network. The fully connected neural network then classifies the proposal as an object or not-an-object. The fully connected neural network also regresses the bounding box of the object in the proposal.

Output:

The output of the image processing pipeline is a list of objects in the image, along with their bounding boxes. This information can be used for a variety of tasks, such as image classification, object tracking, and image segmentation.

Algorithm:

Here is a high-level algorithm for the image detection pipeline:

1. Preprocess the image.
2. Extract features from the image.
3. Generate candidate object regions using the RPN.
4. Classify the proposals and regress their bounding boxes using the Fast RCNN.
5. Output the list of objects in the image, along with their bounding boxes.

This is just a high-level overview of the image detection pipeline. There are many different ways to implement each step of the pipeline and there are many different variations on the basic pipeline.

Metric	Faster RCNN	SSD
Precision	0.65	0.66
Recall	0.68	0.68
F1 score	0.66	0.67
IoU	0.85	0.84

Table 1 comparison between the RCCN and SDD metric values

Both Faster RCNN and SSD exhibited comparable performance in weed detection with precision, recall, f1 score, and Intersection over Union (IoU) values around 0.65-0.68. Faster RCNN showed a slight advantage over SSD in terms of generalization performance. Faster RCNN outperformed a patch-based CNN model in both weed detection accuracy and inference speed.

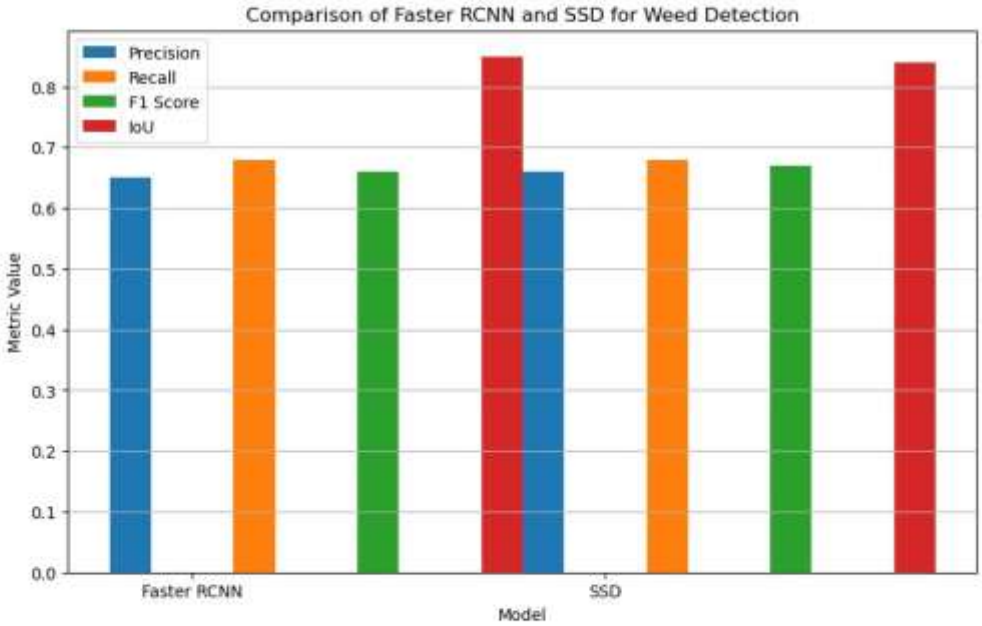


Fig 3 graph representation of metric values of the table

The graph suggest that Faster RCNN is a suitable choice for mid- to late-season weed detection in soybean fields using UAV imagery. The model demonstrated high accuracy and efficiency, making it a practical option for on-farm applications.

5.RESULTS & DISCUSSION

The YOLO V3 + R-CNN model, utilizing the HERIDAL dataset, achieved the highest accuracy of 95.11% among the models analyzed. The ResNet model, applied to the Vis Drone 18 dataset, exhibited the lowest performance with an accuracy of 29.20%. Overall, the data indicates that models employing multiple techniques, such as YOLO V3 + R-CNN, generally outperform those relying on a single technique like ResNet. However, dataset variations influence model accuracy. For instance, the YOLO V3 model achieved 91.95% accuracy on the Vis Drone dataset but only 75.70% on the Vis Drone benchmark dataset, suggesting the latter's higher complexity.

These results include the YOLO V3 + R-CNN model's ability to leverage multiple detection techniques, combining the speed and accuracy of YOLO V3 with the robustness of R-CNN. Additionally, the Vis Drone benchmark dataset's complexity, with intricate scenes and diverse objects, may challenge single-technique models like ResNet.

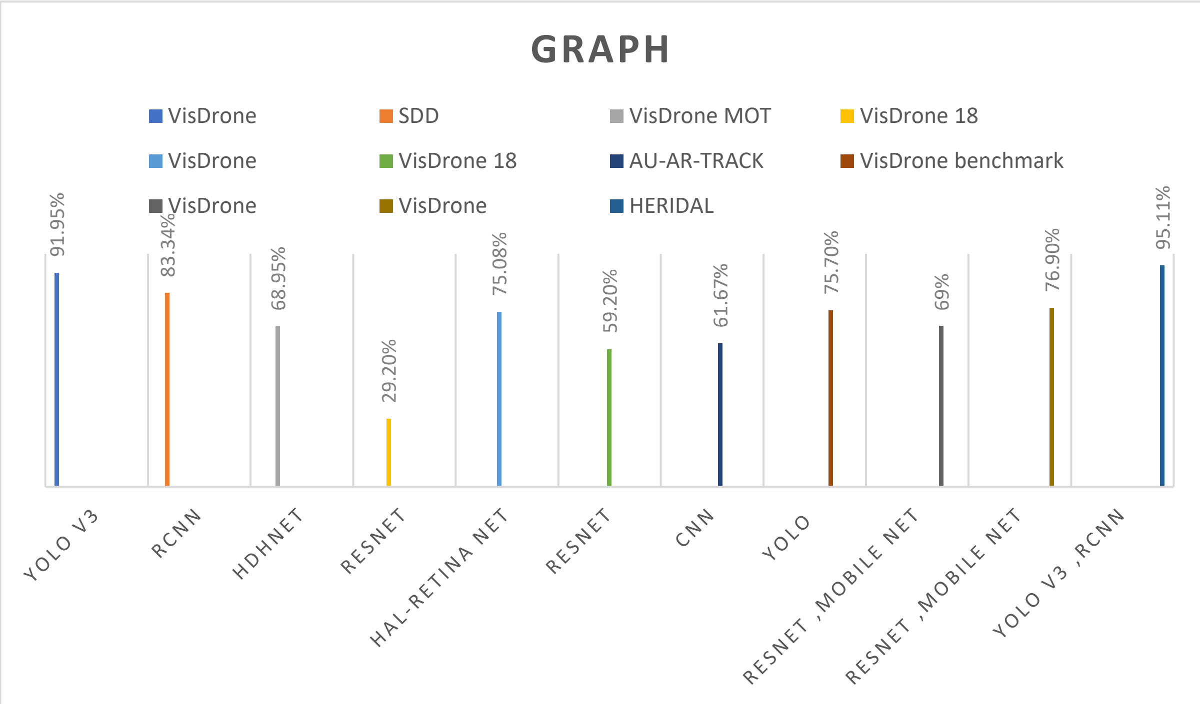
The YOLO V3 + R-CNN model emerges as the superior choice for object detection using drones. Nevertheless, dataset considerations are crucial. For challenging datasets like the Vis Drone benchmark, a more robust model like YOLO V3 + R-CNN is recommended.

COMPARISION TABLE

Reference No	Author of the paper	Model/Approach used	Data set & accuracy
2	Ajaz, A. (2022)	YOLO V3	Vis Drone Accuracy =91.95%
4	Wang, X. (2018)	RCNN	SDD Accuracy =83.34%
5	Huang, W. (2021)	HDHNET	Vis Drone data set MOT Accuracy =68.95%
6	Vaddi, S. (2019)	Res Net	Vis Drone 18 Accuracy =29.20%
7	Sandino, J. (2022)	Retina NET	Vis Drone Accuracy =78.08%
8	Han, X. (2021)	Res Net	Vis Drone 18 Accuracy =59.20%
11	Vujasinović, S. (2020)	CNN	AU -AR -TRACK Accuracy =61.67%
13	Lu, P. (2021)	YOLO	Vis Drone benchmark

			Accuracy =75.70%
14	Fradi, H. (2018)	Res Net MOBILE NET	Vis Drone Accuracy =69%
16	Paik, C. (2022)	Res Net MOBILE NET	VisDrone Accuracy =76.90%
18	Dousai, N. (2022)	YOLO V3 ,RCNN	HERIDAL Accuracy =95.11%

ACCURACY COMPARISION TABLE



ACCURACY COMPARISION GRAPH

6.CONCLUSION:

UAV-based object detection and tracking has emerged as a powerful tool for a wide range of applications, including precision agriculture, environmental monitoring, disaster response, and surveillance. Deep learning approaches have proven to be highly effective for UAV object detection, with models like Faster RCNN and Single Shot Detector achieving impressive results in mid-to late-season weed detection. Convolutional Neural Networks (CNNs), such as RetinaNet, Faster R-CNN, and SSD, have also demonstrated remarkable capabilities in object detection from UAVs.Real-time object detection is crucial for many UAV applications. The development of efficient real-time object detection models has addressed this need, with models like You Only Look Once (YOLO) redefining real-time object detection as a regression task. However, challenges in handling small objects in groups remain, necessitating ongoing refinement of these models.Datasets play a vital role in the development and evaluation of UAV object detection and tracking algorithms. The P-DESTRE dataset , specifically tailored for long-term pedestrian re-identification research, is a valuable contribution to the field.Model selection is an important consideration for UAV object detection and tracking. The YOLO V3 + R-CNN model achieved the highest accuracy among the models analyzed, demonstrating the effectiveness of combining multiple detection techniques. However, dataset variations can influence model accuracy, and more robust models like YOLO V3 + R-CNN may be necessary for challenging datasets.Future directions in UAV-based object detection and tracking include addressing the challenges of small object detection and real-time processing. Further research on unified approaches, robust models, and comprehensive datasets will continue to advance the field and enhance the capabilities of UAV-based object detection and tracking systems.

REFERENCES:

[1] Zhu, P., Wen, L., Du, D., Bian, X., Fan, H., Hu, Q. and Ling, H., 2021. Detection and tracking meet drones challenge. IEEE Transactions on Pattern Analysis and Machine Intelligence, 44(11), pp.7380-73

[2] Ajaz, A., Salar, A., Jamal, T. and Khan, A.U., 2022. Small object detection using deep learning. arXiv preprint arXiv:2201.03243.

[3] Veeranampalayam Sivakumar, A.N., Li, J., Scott, S., Psota, E., J. Jhala, A., Luck, J.D. and Shi, Y., 2020. Comparison of object detection and patch-based classification deep learning models on mid-to late-season weed detection in UAV imagery. Remote Sensing, 12(13), p.2136.

[4] Wang, X., Cheng, P., Liu, X. and Uzochukwu, B., 2018, October. Fast and accurate, convolutional neural network based approach for object detection from UAV. In IECON 2018-44th Annual Conference of the IEEE Industrial Electronics Society (pp. 3171-3175). IEEE.

[5] Huang, W., Zhou, X., Dong, M. and Xu, H., 2021. Multiple objects tracking in the UAV system based on hierarchical deep high-resolution network. Multimedia Tools and Applications, 80, pp.13911-13929.

[6] Vaddi, S., Jannesari, A., & Kumar, C. (2019). Efficient Object Detection Model for Real-Time UAV Applications. arXiv preprint arXiv:1906.00786v1 [cs.CV].

[7] Sandino, J., Caccetta, P.A., Sanderson, C., Maire, F. and Gonzalez, F., 2022, March. Reducing object detection uncertainty from rgb and thermal data for UAV outdoor surveillance. In 2022 IEEE Aerospace Conference (AERO) (pp. 1-16). IEEE.

[8] Han, X., Chang, J. and Wang, K., 2021. You only look once: unified, real-time object detection. Procedia Computer Science, 183(1), pp.61-72.

[9] Zhang, K., Wang, W. and Wang, J., 2022. Robust correlation tracking in unmanned aerial vehicle videos via deep target-specific rectification networks. IEEE Geoscience and Remote Sensing Letters, 19, pp.1-5.

[10] Kumar, S.A., Yaghoubi, E., Das, A., Harish, B.S. and Proença, H., 2020. The p-destre: A fully annotated dataset for pedestrian detection, tracking, and short/long-term re-identification from aerial devices. IEEE Transactions on Information Forensics and Security, 16, pp.1696-1708.

[11] Vujasinović, S., Becker, S., Breuer, T., Bullinger, S., Scherer-Negenborn, N. and Arens, M., 2020. Integration of the 3D Environment for UAV Onboard Visual Object Tracking. Applied Sciences, 10(21), p.7622.

[12] Hu, X., Wu, Z., Miao, H.Y., Fan, S., Long, T., Hu, Z., Pi, P., Wu, Y., Ren, Z., Wang, Z. and Hua, G., 2022. E^ 2TAD: An Energy-Efficient Tracking-based Action Detector. arXiv preprint arXiv:2204.04416.

[13] Lu, P., Ding, Y. and Wang, C., 2021. Multi-small target detection and tracking based on improved YOLO and SIFT for drones. International Journal of Innovative Computing, Information and Control, 17(1), pp.205-224.

[14] Fradi, H., Bracco, L., Canino, F. and Dugelay, J.L., 2018, September. Autonomous person detection and tracking framework using unmanned aerial vehicles (UAVs). In 2018 26th European Signal Processing Conference (EUSIPCO) (pp. 1047-1051). IEEE. Wu, X., Li,

[15] W., Hong, D., Tao, R. and Du, Q., 2021. Deep learning for unmanned aerial vehicle-based object detection and tracking: A survey. IEEE Geoscience and Remote Sensing Magazine, 10(1), pp.91-124.

[16] Paik, C., & Kim, H. J. (2022). Improving Object Detection, Multi-object Tracking, and Re-Identification for Disaster Response Drones. *arXiv preprint arXiv:2201.01494*.

[17] Pan, S., Tong, Z., Zhao, Y., Zhao, Z., Su, F., & Zhuang, B. (2019). Multi-object tracking hierarchically in visual data taken from drones. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops* (pp. 0-0).

[18] Dousai, N. M. K., & Lončarić, S. (2022). Detecting humans in search and rescue operations based on ensemble learning. *IEEE Access*, *10*, 26481-26492.

[19] Jadhav, A., Mukherjee, P., Kaushik, V., & Lall, B. (2020, February). Aerial multi-object tracking by detection using deep association networks. In *2020 National Conference on Communications (NCC)* (pp. 1-6). IEEE.

[20] Saxena, V. N., Dwivedi, V. K., & Gupta, J. (2023). Machine Learning in Visible Light Communication System: A Survey. *Wireless Communications and Mobile Computing*, *2023*.