GRENZE Scientific Society

Grenze International Journal of Engineering and Technology, January Issue

Automatic Crop and Weed Detection using Deep Learning: A Comparative Analysis of YOLO Models

Dr. Premasudha B G[1], Murali K L[2] and Akshay M J[3]
[1-2]Dept. of MCA, Siddaganga Institute of Technology, Tumkur, India
Email: bgpremasudha@sit.ac.in, si23mc028@sit.ac.in
[3]Dept. of ISE, Jawaharlal Nehru New College of Engineering, Shivamogga, India
Email: akshaymj@jnnce.ac.in

***Abstract*— Managing weeds efficiently is essential to securing optimal crop production, particularly in the face of rising herbicide resistance and the push toward environmentally sustainable farming. This study evaluates the performance of advanced deep learning-based object detection models for identifying crops and weeds in realistic field conditions. Over 1,400 RGB field images were gathered via drone platforms, covering multiple crop growth phases and challenging visual scenarios, including variable lighting and partial plant coverage. Four models—YOLO-NAS, YOLOv11, YOLOv12, and RF-DETR—were trained and tested under the same experimental setup. The models were assessed using standard object detection metrics such as mean Average Precision at 0.5 (mAP@0.5), precision, and recall. Among all tested models, YOLOv12 achieved the highest performance, recording 89.0% mAP@0.5, 89.4% precision, and 81.9% recall, indicating strong suitability for real-time detection tasks. YOLO-NAS followed closely, offering a well-balanced combination of accuracy and detection consistency, enhanced by its Neural Architecture Search driven structure. RF-DETR achieved the top precision score but lagged in recall. YOLOv11 delivered comparable accuracy alongside rapid inference times. Overall, YOLOv12 emerged as the most promising choice for scalable, real-time deployment using drone imagery. The findings emphasize the benefits of combining drone acquired imagery with automated AI detection tools to enhance precision farming and support sustainable land use practices.**

***Index Terms*— Weed and Crop Detection, Deep Learning, drone-based imagery, YOLO models, Precision Agriculture, Real-time detection.**

I. INTRODUCTION

Weeds are one of the primary constraints in modern agriculture, competing aggressively with crops for essential resources such as sunlight, water, and soil nutrients. Their presence leads to significant reductions in yield and increased production costs due to additional labor and chemical inputs. Traditional weed management techniques, including manual weed identifying, mechanical cultivation, and uniform herbicide spraying, though widely practiced, are labor-intensive, time-consuming, and harmful to the environment due to unnecessary spray of pesticides. To address these limitations, precision agriculture approaches have been introduced, which enable site-specific weed control and can reduce herbicide usage by more than 50% while preserving soil and crop health [9].

Grenze ID: 01.GIJET.12.1.501

Earlier approaches relied on handcrafted features such as color, shape, and texture to distinguish weeds from crops, but these techniques often failed under varying lighting conditions or when crops and weeds exhibited similar morphology [15]. The rise of deep learning (DL) and computer vision (CV) has transformed weed detection, offering highly accurate, real-time solutions that outperform traditional image processing approaches. Modern object detection frameworks such as Faster R-CNN and YOLO (You Only Look Once) have made automated weed identification more robust and scalable for agricultural use [17,12].

Among these frameworks the YOLO family has gained particular attention due to its single-shot architecture, which carries out object detection and position estimation together in a single processing cycle. This design removes the need for multiple processing stages found in region-based detectors, enabling faster inference without sacrificing accuracy. Such efficiency is especially valuable in real-time field applications like drone-assisted crop monitoring where rapid decision-making is essential. Recent YOLO variants, including YOLO-NAS [8], incorporate innovations such as anchor-free detection and neural architecture search (NAS), while YOLOv11 and YOLOv12 [3] improve detection capabilities for small or overlapping weeds. Additionally, transformer-based architectures like RF-DETR [2] utilize self-attention mechanisms to enhance detection accuracy in dense and complex vegetation. Comparative studies have shown that while RF-DETR achieves the highest precision, YOLOv12 delivers superior overall results, with higher mAP and recall, making it a strong candidate for real time weed detection.

The YOLO models utilized in this study are adopted with a unified three-stage architecture tailored for crop and weed detection: a backbone that extracts hierarchical visual features from field images, a neck that fuses multi-scale features to enhance detection of small and overlapping weeds among crops, and a head that generates class predictions (crop vs. weed) along with bounding box coordinates.

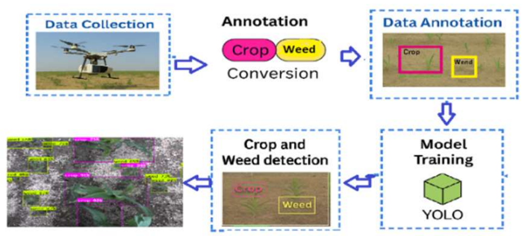

Fig. 1: Architecture for weed and crop detection by YOLO object detectors.

The proposed system for crop and weed detection follows a streamlined workflow involving data collection, annotation preparation, model training, and evaluation. Field images are captured using drones under diverse lighting and growth conditions to reflect real-world scenarios, and the dataset used includes some images that were already annotated. These annotations follow the YOLO labeling format, assigning class labels (e.g., "0" for crop, "1" for weed) along with bounding box coordinates and class identifiers. As illustrated in Fig.1, the annotated dataset displays bounding boxes over the crops and weeds, each labeled with their respective class, making it visually easier to distinguish between objects during training and evaluation. This visual labeling helps in both training and validating object detection models by providing clear ground truth references. The dataset undergoes preprocessing steps such as resizing, normalization, and augmentation to enhance robustness, after which it is divided into training, validation, and testing subsets. Four object detection models YOLO-NAS, YOLOv11, YOLOv12, and RF-DETR are trained with fine-tuned hyperparameters. Performance is assessed using metrics like mAP@0.5, precision, and recall to determine the most accurate and efficient model for real-time deployment in smart farming. This end-to-end pipeline supports precision agriculture by enabling reliable, automated crop-weed identification.

This introduction provides the foundation for the sections that follow in this paper Section II presents a Related Work on existing deep learning models for crop and weed detection; Section III. Materials and Methods details the architectures of YOLOv11, YOLOv12, YOLO-NAS, and RF-DETR, Section IV discusses comparative results and analysis; and Section V concludes with insights and potential future directions for deploying these models in precision agriculture.

II. Literature Review

Research in automated weed detection has expanded rapidly with the advancement of deep learning methods, enabling real time detection and classification in agricultural environments. A comparative study using UAV imagery demonstrated that deep learning approaches significantly outperform traditional handcrafted feature methods, especially when leveraging multi scale feature fusion and high-resolution aerial data, though detecting small or overlapping weeds remains challenging [11]. Additionally, in contexts with class imbalance where weeds are underrepresented compared to crops or backgrounds the application of techniques such as data augmentation, weighted loss functions, and transfer learning has been shown to substantially improve model performance [14].

Studies have evaluated YOLO object detectors in turfgrass scenarios, comparing YOLOv5 through YOLOv8, where YOLOv8 achieved the highest mAP@0.5 of 0.9795 under varied lighting conditions, but all weeds were treated as a single class without crop–weed coexistence consideration [9]. Fine-grained object detection approaches, such as FGOD-YOLOv8, integrated multi-scale dilated attention and cross-scale feature fusion to improve small object detection in cluttered fields [1].

Custom datasets have been developed in [10] for specific crop types, such as sunflower and sugar beet, are tested with YOLO, EfficientDet, and DETR models, where YOLO consistently outperformed others in precision and speed but faced limitations in detecting small weeds and lacked mixed crop weed evaluations. YOLOv5-based detection of weeds, crops, and diseased leaves achieved high mAP but classified crops and weeds separately, failing to address coexistence patterns [12].

Reviews on AI applications in agriculture have emphasized the dominance of YOLO and Faster R-CNN [4,15] for weed detection, identifying barriers such as hardware costs, dataset imbalance, and lack of region-specific datasets Lightweight and optimized deep learning models have been proposed to improve efficiency for field deployment, though their applicability in Indian conditions remains underexplored [5,13].

Improved YOLOv5 architectures have demonstrated strong accuracy (mAP > 90%) in real-time weed recognition under complex field conditions, yet most focus solely on weed detection [6]. UAV-based YOLOv7 and YOLOv8 models have shown high accuracy but grouped all weeds into a single class without crop discrimination [11].

Modified YOLOv5 models have been applied to specific crops, such as sesame, achieving high detection accuracy but again treating weeds as a single class [7]. YOLO-NAS models, developed through neural architecture search, have been applied for real-time crop–weed detection, achieving a balance between accuracy and recall that supports precision agriculture [8].

YOLOv12 and RF-DETR have been compared for real-time weed or similar object detection, with RF-DETR achieving a notably higher mAP@50 of 0.9464 in single-class detection and 0.8298 in multi-class detection, while YOLOv12 demonstrated stronger performance in mAP@50:95, notably reaching 0.7620 for single-class and 0.6622 for multi-class scenarios. This indicates that RF-DETR excels at precise localization, whereas YOLOv12 [2] provides more balanced classification, is superior in overall metric evaluation, and shows higher mAP and recall in real-time weed detection.

Comparative evaluations of YOLOv11 and YOLOv12 architectures in agricultural detection tasks have shown both achieving high accuracy, with YOLOv12 slightly outperforming YOLOv11 in certain metrics [3].

Applications of deep learning in weed detection have also included comparative studies using UAV imagery and multiple detection models, highlighting UAV-based monitoring as an efficient approach for large-scale weed mapping [11]. Lightweight deep learning models have been applied for multi-object detection and classification of crops and weeds, showing the potential of resource-efficient deployment [13]. Class imbalance issues in weed datasets have been studied, with methods proposed to improve recognition in real-world field imagery [14].

Earlier works faced challenges in accurately detecting both crops and weeds, as they relied on handcrafted features such as color and texture for classification, which struggled under variable field conditions [16]. The introduction of YOLO [17] as a unified real-time object detector marked a significant shift toward deep learning based approaches for agricultural detection tasks.

Most existing studies focus on detecting weeds and crops separately, and only YOLO NAS addresses detecting them together in the same field. This limits their use in real world conditions where both coexist. Our proposed system evaluates YOLO V12 achieving a strong mAP score with the other models like YOLOV11, YOLO-NAS, RFDETR to overcome this gap by enabling simultaneous detection of crops and weeds with high accuracy, even in challenging conditions such as overlapping plants, varying sizes, and changing lighting.

III. MATERIALS AND METHODS

A. Data Collection and Dataset Preparation

The dataset for this project was sourced from the Roboflow platform, containing over 1,400+ high resolution RGB images already annotated with crop and weed classes. These images were captured using handheld cameras and drone-mounted imaging systems in open field conditions, covering various lighting scenarios, plant growth stages, and soil textures to reflect real-world agricultural diversity. The dataset includes both crops and multiple weed species coexisting in the same field, with realistic cases of overlapping vegetation and partial occlusions. For consistency across all models, images were standardized to 640×640 pixels and pixel intensities normalized to improve training convergence. To further increase dataset diversity and simulate real-field variations, augmentation techniques such as random flips, rotations up to 15°, brightness and hue shifts, and mosaic augmentation were applied. To ensure class balance and capture different field scenarios, the labeled dataset was partitioned into 80% for training, 10% for validation, and 10% for testing.

B. Algorithms and Training Configuration

This study compared four advanced object detection architectures YOLOv11, YOLOv12, YOLO-NAS, and RF-DETR through systematic evaluation. The selection spans two architectural paradigms: convolution-driven detectors (YOLO series) and transformer-driven designs (RF-DETR), YOLO-NAS, developed through Neural Architecture Search, achieves a balanced trade-off between speed and accuracy by optimizing its design automatically. RF-DETR applies self-attention to capture contextual relationships in dense vegetation, albeit with greater computational overhead. YOLOv11's compact design is optimized for real-time use on edge devices, whereas YOLOv12 integrates anchor-free detection with advanced feature fusion to enhance detection of small or overlapping weeds. Given these enhancements, YOLOv12 was chosen as the baseline for detailed comparisons.

All models were trained in PyTorch, with preprocessing and augmentation handled via the Roboflow platform. Experiments ran on a high-end workstation featuring an NVIDIA RTX 4090 GPU (24 GB VRAM), Intel Core i9 CPU, 64 GB RAM, and Ubuntu 22.04 LTS. Models were initialized with pretrained COCO weights, applying transfer learning to speed up convergence and enhance performance on the agricultural dataset Training for each model spanned over 100 epochs with a batch size of 16, using SGD optimizer (momentum 0.937, weight decay 0.0005). A cosine annealing scheduler decayed the learning rate from 0.001 to promote stable convergence.

The loss function combined CIoU loss for bounding box regression with binary cross-entropy for classification and objectness prediction. Real-time monitoring via Roboflow's dashboard tracked metrics like precision, recall, and mAP, aiding hyperparameter tuning. To ensure a fair comparison, the same configuration was applied to all models input images were resized to 640×640 pixels, pretrained COCO weights were used for initialization, and identical loss functions and optimization settings were applied. This consistent setup enabled a robust evaluation of YOLOv12 against YOLOv11, YOLO-NAS, and RF-DETR in subsequent analyses.

C. Evaluation Metrics

Model performance was evaluated using standard object detection metrics commonly applied in agricultural computer vision. The main measure, mean Average Precision at a 0.5 threshold (mAP@0.5), reflected overall detection accuracy by summarizing the precision–recall relationship for crop and weed classes. Precision measured the share of correctly identified weeds among all weed predictions, showing how well the model avoids false positives. Recall indicated the percentage of actual weeds successfully detected, highlighting the system's ability to minimize false negatives. Additionally, inference speed, expressed in frames per second (FPS), was monitored to evaluate suitability for real-time UAV-based field inspections. Collectively, these metrics offered a comprehensive basis for comparing detection accuracy and practical deployment potential of the tested models

IV. RESULTS AND DISCUSSION

YOLO is a family of real-time object detection models that perform both classification and localization in a single forward pass of a neural network. Unlike traditional region-based detectors, which process images in multiple stages, YOLO processes an image by splitting it into a grid, where each cell outputs bounding boxes, object class probabilities, and associated confidence values. This single-shot approach enables YOLO to achieve exceptional speed without sacrificing accuracy, making it particularly suited for applications like drone-based agricultural monitoring. In weed detection, YOLO models can analyze field images to identify crops and weeds simultaneously, drawing bounding boxes to indicate object locations and assigning confidence scores to show detection certainty. Depending on the variant, YOLO can use either anchor-based or anchor-free detection, enabling it to detect objects of different sizes, including small or partially hidden weeds. Its efficiency, adaptability, and compatibility with lightweight edge devices have made it a favored choice for precision farming. In this study, four advanced object detection models YOLOv11, YOLOv12, YOLO-NAS, and RF-

DETR were trained and tested on a custom-labeled dataset of over 1,400 RGB field images containing mixed crop and weed scenarios. The dataset, captured under diverse lighting conditions, varying plant growth stages, and partial occlusions, was annotated into two categories crop and weed. Images were split into training (80%) and validation/testing (20%) sets. Model performance was assessed using standard evaluation metrics, including precision, recall, and mean Average Precision (mAP).

YOLOv12, used as the benchmark, demonstrates the most balanced performance, achieving an mAP@0.5 of 89.0%, precision of 89.4%, and recall of 81.9%. As shown in Fig.2, its mAP metrics graph displays a steep improvement curve close to 0.88, indicating high sensitivity with minimal false positives. Its anchor-free architecture and feature fusion layers contribute to superior detection of small and overlapping weeds, which is critical under real field conditions.

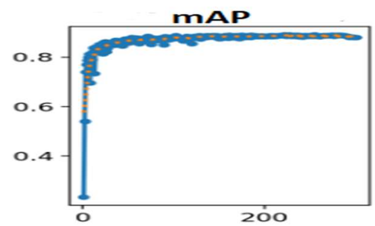

Fig.2: Performance metricsof YOLOV12 model.

Compared with YOLOv12, YOLO-NAS achieves a similar mAP@0.5 (88.7%) but leads in recall (83.5%), signifying its strength in detecting nearly all weed instances. As depicted in Fig.3, its mAP metrics graph indicates slightly more false positives than YOLOv12, aligning with its lower precision (88.2%). This trade-off positions YOLO-NAS as preferable in scenarios where completeness of detection is prioritized over absolute precision, such as minimizing missed weeds during selective spraying

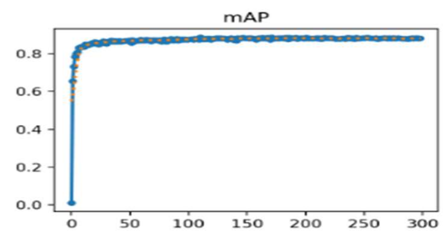

Fig.3: Performance metrics of YOLOV-NAS model.

YOLOv11 achieves mAP@0.5 of 88.6%, matching YOLOv12's recall (81.9%) with slightly lower precision (88.7%). As illustrated in Fig.4, the mAP metrics graph displays moderate separation between true and false positives, confirming competitive yet slightly inferior performance in distinguishing weeds from crops compared to YOLOv12. However, its lightweight design ensures the fastest inference speed, making it ideal for embedded platforms or UAVs with limited computational resources.

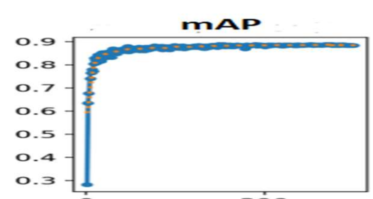

Fig.4: Performance metricsof YOLOV12 model.

RF-DETR, employing a transformer-based architecture, delivers the highest precision (89.4%) but the lowest recall (77%) and mAP@0.5 (85.7%) among the models. As presented in Fig.5, the mAP metrics graph illustrates strong confidence in its detections but limited coverage, missing a larger proportion of weeds. While beneficial in dense vegetation where precise localization is critical, its higher computational cost and slower inference restrict real-time applicability compared to YOLOv12.

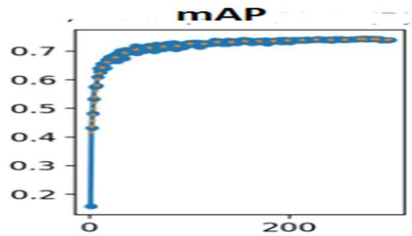

Fig.5: Performance metricsof RF-DETR model.

A. Quantitative Performance Evaluation

The experimental results summarized in Table 1 indicate that all four models YOLOv12, YOLO-NAS, YOLOv11, and RF-DETR delivered high detection accuracy for crop and weed identification. mAP@0.5 values ranged from 85.7% to 89.0%, with YOLOv12 achieving the highest score, closely followed by YOLO-NAS and YOLOv11. Precision remained strong across all models, with YOLOv12 and RF-DETR both reaching 89.4%, while YOLO-NAS achieved the highest recall at 83.5%, highlighting its strength in capturing true positives. Overall, YOLOv12 emerged as the top-performing model, delivering the highest overall accuracy while maintaining strong precision and recall, making it the most reliable choice for this task.

TABLE I: PERFORMANCE EVOLUTION OF TRAINED MODELS

Model	Precision (%)	Recall (%)	mAP@0.5
YOLOv12	89.4	81.9	89.0
YOLO-NAS	88.2	83.5	88.7
RF-DETR	89.4	77.0	85.7
YOLOv11	88.7	81.9	88.6

V. CONCLUSION

In this study, we compared four advanced object detection models YOLO-NAS, YOLOv11, YOLOv12, and RF-DETR to solve the important problem of detecting crops and weeds at the same time in real farming conditions. We used a custom-labeled dataset of over 1,400+ RGB images taken in different lighting and field conditions to train and test the models. Their performance was measured using standard metrics like mAP@0.5, precision, and recall. Among them, YOLOv12 gave the most balanced results, with the highest mAP (89.0%) and precision (89.4%), making it ideal for real use in smart farming. YOLO-NAS had the highest recall (83.5%), meaning it was best at finding as many objects as possible, which is useful when missing a detection is costly. YOLOv11, though lighter and faster, still performed well across all metrics, making it a good choice for mobile devices or drones. RF-DETR, which uses a transformer-based design, worked well in busy or complex scenes but had slightly lower recall than the YOLOV12 model. Unlike earlier studies that only detected weeds or crops separately or used simpler datasets, this work shows a solution that can detect both together. This is important for real-world farming needs like precision spraying or drone monitoring. These results can help in selecting the right model for precision agriculture and guide future work, such as implementing these models in real-time systems, expanding the dataset, or adapting them for different regions to detect both crops and weeds.

REFERENCES

[1] Z. Li, Y. Yan, M. Wei, et al., “FGOD-YOLOv8: Fine-grained object detection applied to crops and weeds,” IEEE Signal Processing Letters, vol. 32, pp. 791–796, 2025. doi: 10.1109/LSP.2024.3519272.

[2] A. Allmendinger, A. O. Saltık, G. G. Peteinatos, A. Stein, and R. Gerhards, “Performance analysis of YOLOv12 and RF-DETR detectors for real-time weed identification in complex orchard settings,” arXiv preprint, arXiv, Jan. 2025. [Online] Available: https://arxiv.org/abs/2501.17387

[3] D. Ribeiro, D. Tavares, E. Tiradentes, F. Santos, and D. Rodriguez, “Assessment of YOLOv11 and YOLOv12 architectures for automated classification and detection of immature Macauba fruits,” Agriculture, vol. 15, no. 15, art. 1571, 2025. doi: 10.3390/agriculture15151571.

[4] A. Bhat, A. Shukla, M. M. Ali, J. S. Sandhu, and K. L. Sahrawat, “Deep learning models for enhancing food security through AI in agriculture,” Discover Artificial Intelligence, vol. 2, no. 1, pp. 1–11, 2024. doi: 10.1016/j.discia.2024.100042.

[5] B. Liu, C. Zhang, Y. Fang, and D. Li, "Crop–weed recognition in precision farming using deep learning integrated with metaheuristic optimization," Intelligent Systems with Applications, vol. 20, pp. 200–292, 2024. doi: 10.1016/j.iswa.2024.200292.
[6] H. Zhang, J. Sun, W. Zhang, et al., "Enhanced YOLOv5-based approach for real-time crop and weed detection in challenging field environments," Precision Agriculture, vol. 25, pp. 219–236, 2024. doi: 10.1007/s11119-023-10073-1.
[7] S. Sonawane and N. N. Patil, "Weed identification in sesame crops using a modified YOLOv5 deep learning model," Indian Journal of Weed Science, vol. 56, pp. 194–199, 2024. doi: 10.5958/0974-8164.2024.00031.X.
[8] S. Patel, D. Patel, A. Patel, J. Nanavati, U. Patel, A. Faldu, and A. Shingala, "YOLO-NAS-based real-time detection of crops and weeds: A neural architecture search-optimized solution," International Journal of Engineering Trends and Technology (IJETT), vol. 72, no. 10, pp. 35–45, Oct. 2024. doi: 10.14445/22315381/IJETT-V72I10P104.
[9] M. Sportelli, O. E. Apolo-Apolo, M. Fontanelli, et al., "YOLO object detector evaluation for weed recognition in diverse turfgrass conditions," Applied Sciences, vol. 13, no. 14, pp. 1–16, 2023. doi: 10.3390/app13148502
[10] D. A. Abuhani, M. ElMohandes, M. Haj Hussain, et al., "Detection of crops and weeds in sunflower and sugar beet fields using SSD-based methods," in Proc. IEEE Int. Conf. Omni-layer Intelligent Systems (COINS), 2023. doi: 10.1109/COINS57856.2023.10189257.
[11] T. B. Shahi, et al., "Comparative study of UAV-based weed detection using deep learning," Drones, vol. 7, no. 10, pp. 6–24, 2023. doi: 10.3390/drones7100624.
[12] P. K. Doddamani and R. G. P., "YOLOv5 algorithm for detecting crops and weeds," in Proc. IEEE MysuruCon, pp. 1–6, 2022. doi: 10.1109/MysuruCon.2022.00000.
[13] X. Zhang, D. Zhang, Y. Xu, et al., "Lightweight deep learning model for multi-object detection of crops and weeds," Frontiers in Plant Science, vol. 13, pp. 109–165, 2022. doi: 10.3389/fpls.2022.1091655.
[14] A. S. M. Hasan, F. Sohel, D. Diepeveen, H. Laga, and M. G. K. Jones, "Addressing class imbalance in weed recognition using deep learning," arXiv preprint, 2021. [Online] Available: https://arxiv.org/abs/2112.07819
[15] K. Hu, et al., "A review of deep learning techniques for in-crop weed identification," arXiv preprint, 2021. [Online] Available: https://arxiv.org/abs/2103.14872
[16] H. Mekhalfa an d F. Yacef, "Crop/weed classification using supervised learning with color and texture-based features," arXiv preprint, 2021. [Online] Available: https://arxiv.org/abs/2106.10581
[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "YOLO: A unified, real-time object detection framework," arXiv preprint, arXiv:1506.02640, 2016. [Online] Available: https://arxiv.org/abs/1506.02640