Object Detection and Tracking: Camera-based 2D Object Detection (YOLOX)

Lee Xuan Le, Chloe 22410022

Valen Tang Wenwen 22410023

Huang Xiang 12110426

Zhong Jing Chen 12110320

Contents

1 Introduction

1.1 Background and Significance

Object detection is currently a widely researched section of computer vision, with its primary purpose to identify the position of objects in images, before classifying them. Deep convolutional neural networks (CNNs) serve as the backbone for object detection, their efficiency being a great advantage in extracting features closer to real-time execution, compared to non-automatic extraction (Chen et al., 2023).

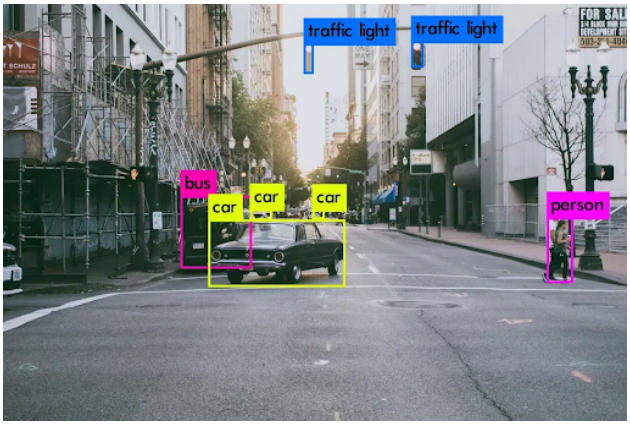

Figure 1: General View of Object Detection and Image Classification (Świeżewski, 2020)

Object detection methods date back as far as to 1986, in which the Histogram of Oriented Objects was introduced, but only gained traction in 2005. Object detection is particularly useful in a multitude of ways, ranging from robotics and self-driving cars to facial recognition and detection for personal and national security. It is of no doubt that object detection has established itself as a powerful tool in today's society of ever-advancing AI, which unfortunately still comes with many potential security and privacy threats. A great example of this would be the emerging issue with Deepfake, in which Deepfake has sadly been abused to generate misleading information or even non-consensual videos, such as Deepfake pornography.

Clearly, object detection has made its mark on the global economy, and it is definitely unwise to ignore the possible issues arising from it. Instead, it is crucial that we understand the uses of object detection, and elevate its advantages to ensure that it creates a positive impact on our society today.

1.2 Motivation

Narrowing object detection down to the finer details, there are many models that have been restlessly trained, improved on and produced. In particular, You Only Look Once Extended (YOLOX) is a high-performance object detection model that belongs to the You Only Look Once (YOLO) family. It is a significant advancement compared to previous YOLO models, greatly exceeding their previous limitations, boosting the model's performance in object detection.

The YOLOX architecture is mainly divided into three main parts, namely the backbone, neck and head. The backbone is responsible for extracting features from the input image, which is made possible with a pre-trained CNN. The neck, which did not exist in initial YOLO versions, combines features from multiple scales in the backbone to improve detection performance. The head will then utilise the extracted features provided by the backbone and neck to carry out predictions and classification (Boesch, 2024).

Out of the many models for object detection, another well-known model is the Region-based Convolutional Neural Networks (R-CNN). Why focus on YOLOX then? It is undeniable that YOLOX has the upper hand in many aspects.

In terms of real-time detection and speed, YOLOX is optimal in real-time object detection, armed with the ability to process images in a single forward pass through the network. Compared to YOLOX, R-CNN mostly uses a multi-step process that is computationally more expensive and slow. This makes YOLOX more suitable for most tasks with real-time processing, such as surveillance systems and self-driving cars.

YOLOX, being built on fresh and more well-developed techniques, also utilises Anchor-free detection and Decoupled Head architecture to produce predictions of enhanced quality and speed. These will be further explained in the later part of the report.

Overall, YOLOX is highly versatile and performs better on a diversity of datasets due to its improved object detection methods, such as data augmentation, higher-quality backbone networks and stronger regularisation techniques. YOLOX is a cutting-edge technology that paves the way for greater worldwide accessibility. It allows for the tracking of multiple objects, even within complex or congested environments. Furthermore, its easy integration into autonomous vehicles will largely improve the mobility for individuals who are unable to drive, such as the elderly or disabled.

Figure 2: Sustainable Development Goals (Eurostat, 2024)

YOLOX holds the potential to amplify the security and efficiency of vehicles, which contributes to the global Sustainable Development Goals (SDGs), particularly reducing road fatalities (SDG 3) and improving city infrastructure (SDG 11).

2 Current Research Status

Tracing back to the origins of YOLOX, YOLO is an algorithm best known for its object detection and real-time processing capabilities.

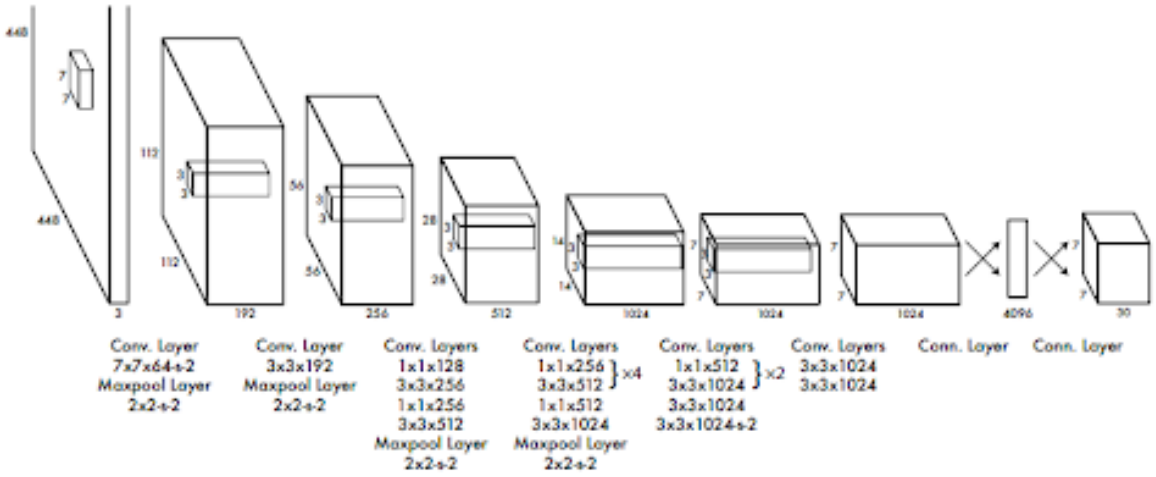

Figure 3: An Architecture of YOLO (Redmon et al., 2016)

The algorithm is implemented by predicting three features. Firstly, grid division, in which YOLO separates the input image into a grid of cells. Secondly, YOLO predicts a number of bounding boxes and their confidence scores. Last but not least, YOLO undergoes a final prediction, in which the probabilities determined in the previous steps are used to predict what the target object is (Boesch, 2024).

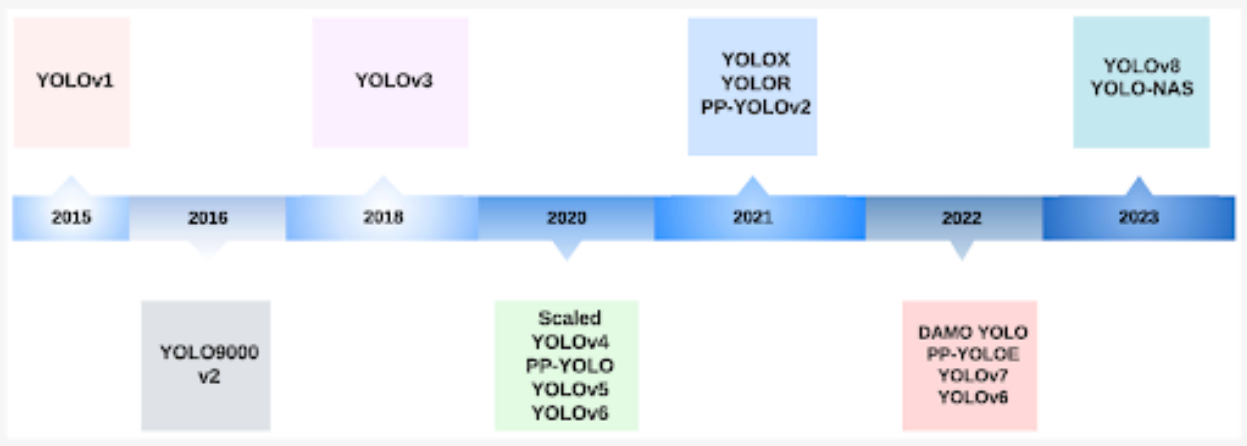

Figure 4: Timeline of YOLO series (Terven, 2023)

Given the increasingly diverse needs globally, several versions have since evolved, each improving upon the gaps of previous models. YOLOX, which was introduced in 2021, was a huge breakthrough as it notably improved YOLO's accuracy with key features, such as an anchor-free design and SimOTA label assignment. Since then, there have been ongoing advancements with the creation of newer versions like YOLOv7 and YOLOv8, which are built upon YOLOX by improving performance for real-time applications (YOLOvx, n.d.).

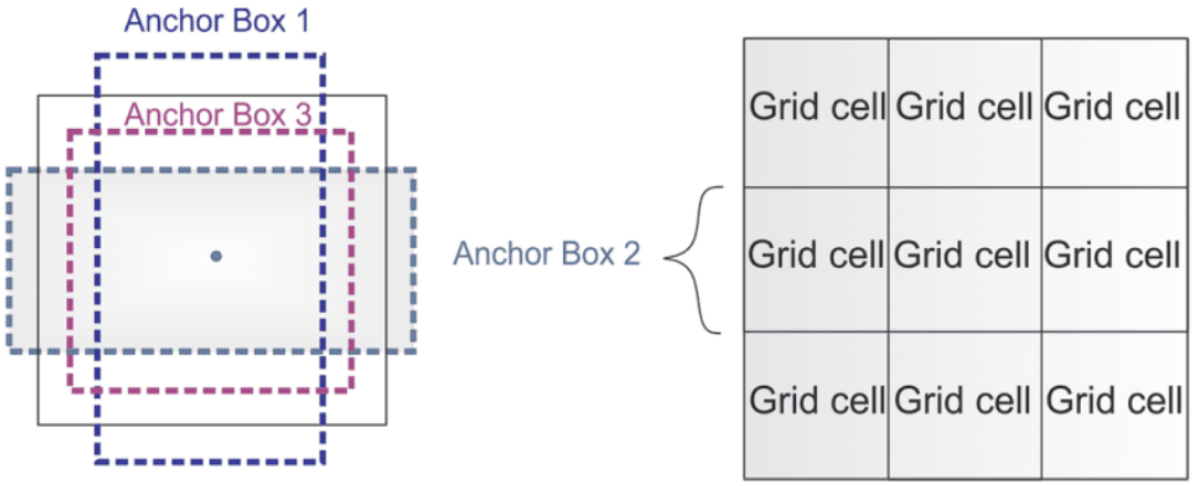

Figure 5: Anchors used in YOLO (Boesch, 2024)

Expanding on the anchor-free design, anchors are predefined bounding boxes, commonly used in the training phase to approximate the actual objects and predict their boundaries (Data-Driven Science, 2023). Unlike previous YOLO versions, YOLOX eliminates the need for these anchors and is capable of directly predicting the bounding boxes. With this feature, it leads to greater flexibility as the model can now better adapt to objects of different shapes and sizes. Furthermore, greater efficiency is also achieved through the reduction in the number of required predictions, enhancing the overall processing speed (Boesch, 2024).

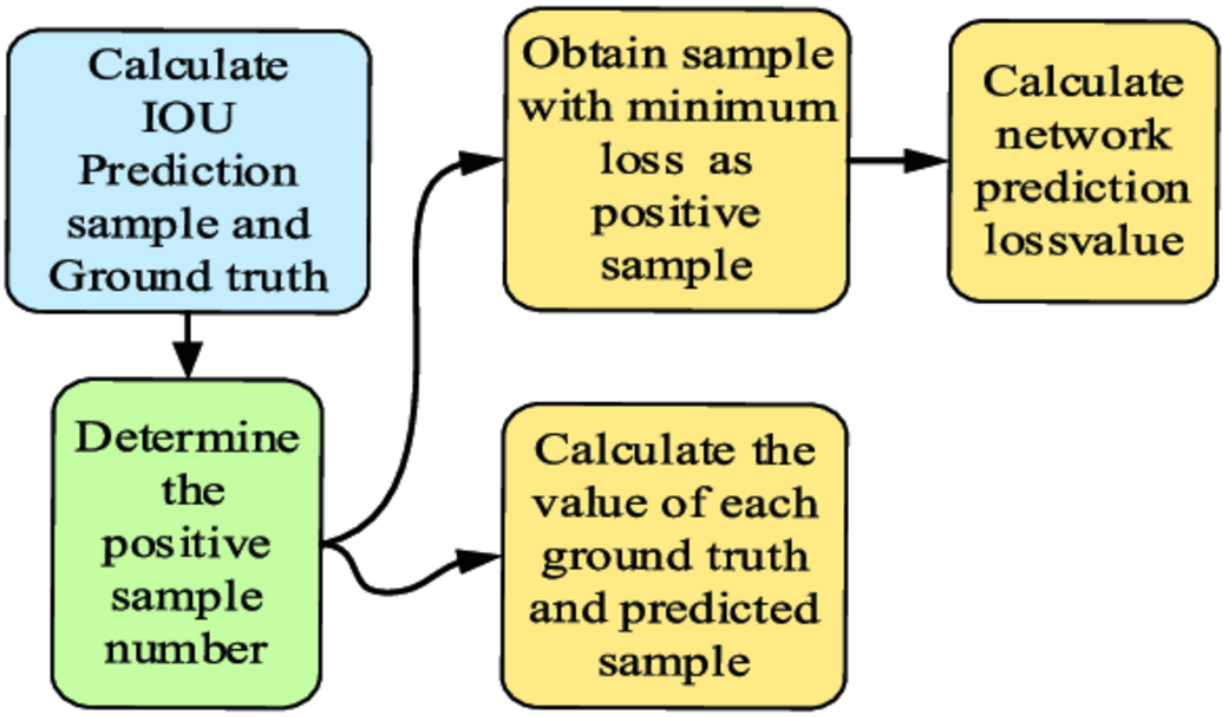

Figure 6: The computational process of SimOTA (Hu et al., 2022)

Elaborating on the SimOTA label assignment, we first introduce OTA, which is known as Optimal Transport Assignment. OTA is a method used in object detection for label assignment with the goal of optimally matching predicted anchors with ground-truth objects. This is achieved by formulating the problem as an Optimal Transport (OT) task and solving it efficiently using the Sinkhorn-Knopp algorithm (Ge et al., 2021b). However, it was

later found that solving the OT problem via the Sinkhorn-Knopp algorithm led to 25% more training time which is computationally expensive and inefficient. Hence, SimOTA, a simplified version, was introduced and subsequently adopted in YOLOX. SimOTA involves selecting the top k predictors with the least associated costs, requiring only a single iteration to approximate the assignments (Mongaras, 2022).

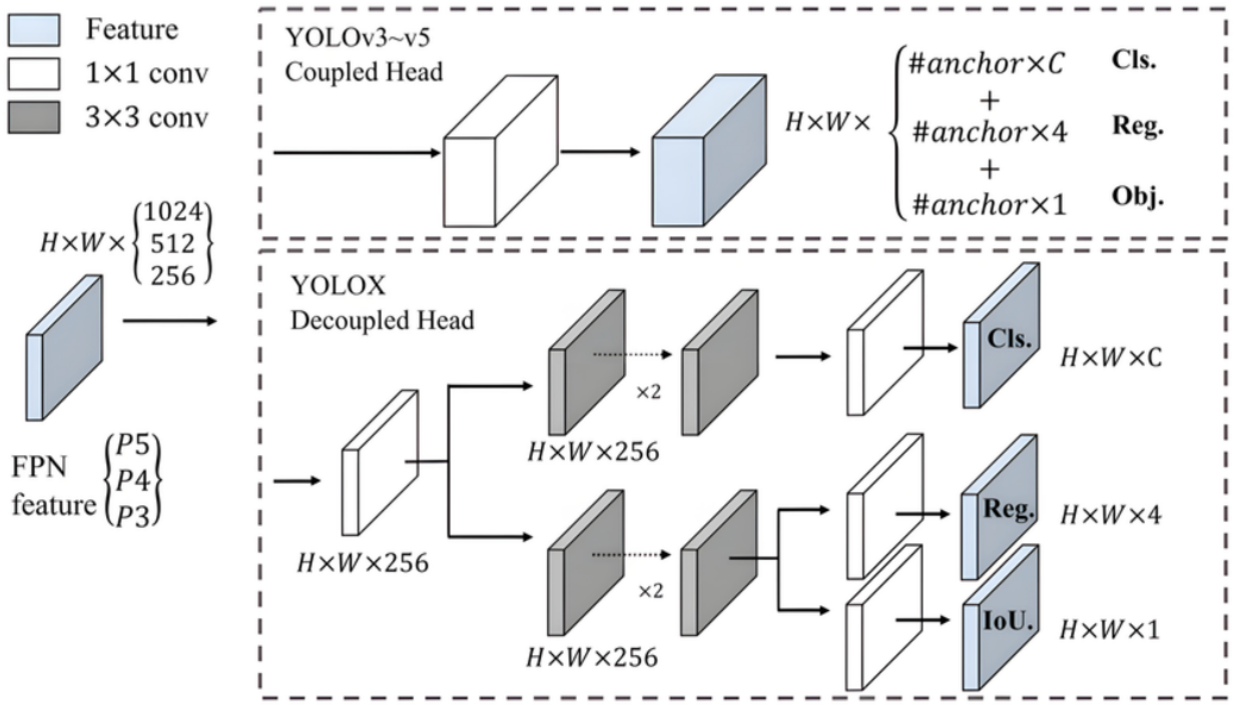

Figure 7: Decoupled Head Structure of YOLOX (Zhang et al., 2023)

The decoupled head feature of YOLOX enables it to effectively target classification and regression tasks, in which it separates the tasks for optimality. This allows it to utilise fewer parameters, decreasing its computational expenses and overhead. (Zhao et al, 2023).

The evolution and usefulness of YOLOX have led to its huge applications in various industries. The most prominent one is the usage of YOLOX to monitor road conditions, detect objects and pedestrians, and enable autonomous vehicles to safely travel on the roads (Saadani, 2024). Many may be aware of automated self-checkout machines at local super-

markets, but how many actually know that YOLOX is what drives these machines behind the scenes? YOLOX has repeatedly proven itself to be useful in the healthcare industry, which employs YOLOX to detect cancer, skin segmentation and pill identification to improve diagnosis accuracy rates (ProX, 2024).

In the agricultural industry, YOLOX is very applicable to identify and categorise crops, pests and diseases to support precision agricultural practices and automate farming operations. In fact, a contemporary YOLOX weed detection model (Zhu et al., 2024) has been developed for meticulous weed identification and control.

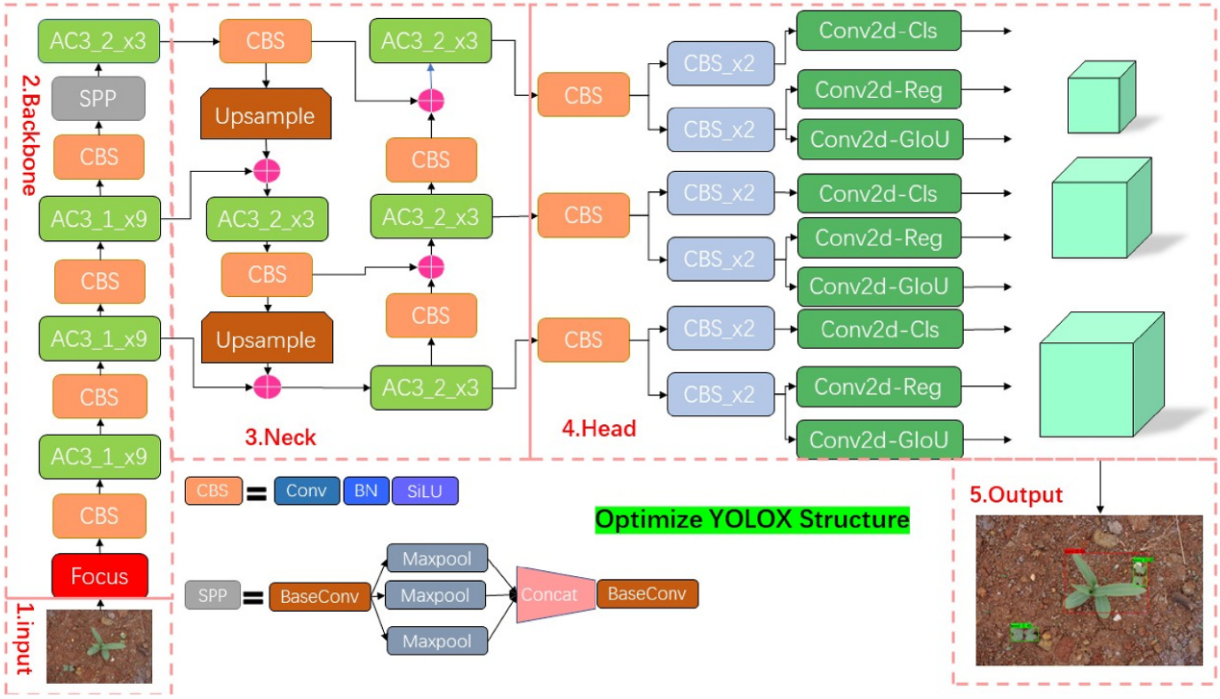

Figure 8: Improved YOLOX Algorithm (Zhu et al., 2024))

The model encompasses a deep network connection lightweight attention mechanism, which is able to quickly and accurately recognise the different types of weeds in maize seedling fields. The lightweight attention module is strongly related to the network of YOLOX-Darknet, which stifles the channel noise effect of leftover computation, increasing the productivity of the detection model. With a deconvolution layer in the residual module and the Generalised

Intersection over Union (GIoU) taking over the Intersection over Union (IoU), the learning capability of the model was greatly boosted.

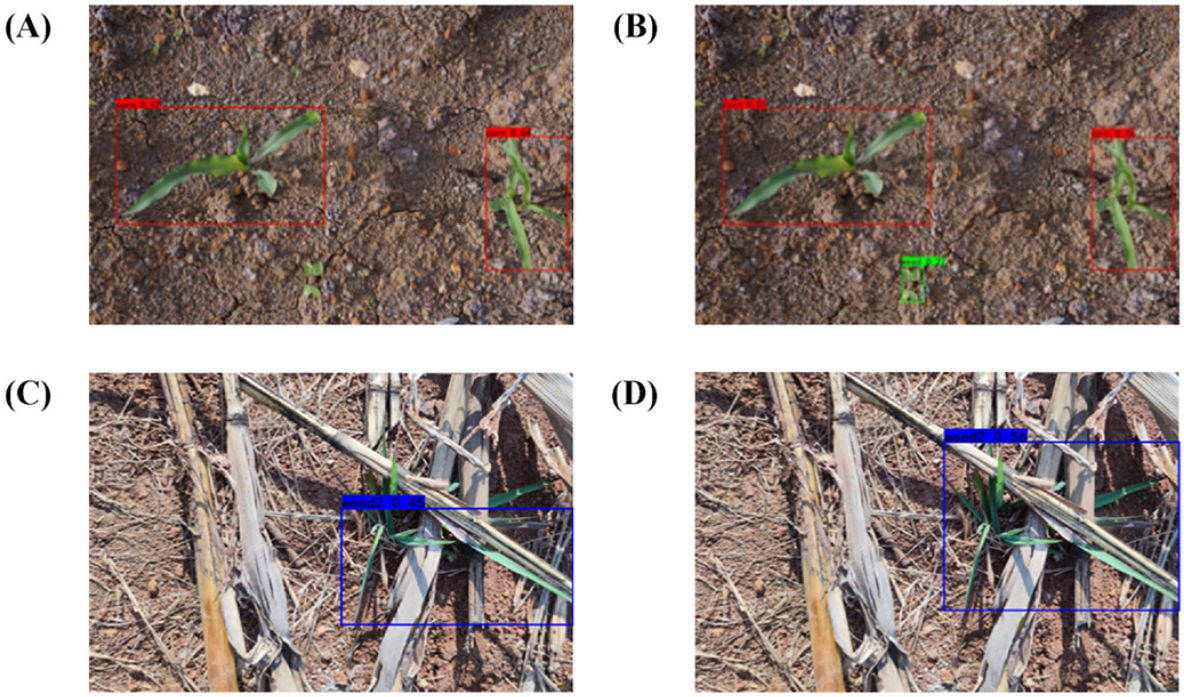

Figure 9: Side-by-side Comparison of Original YOLOX versus Improved YOLOX Detection Results (Zhu et al., 2024)

These features allowed the enhanced algorithm to accomplish a high average detection accuracy of almost 95% in the performance assessment, 0.07% and 1.16% improvement in the F1 value and AP values respectively. The robot with the improved YOLOX algorithm had over 88% for both the detection rate of maize seedlings and weed recognition. This is evident in the figure above, in which the improved YOLOX model was able to accurately detect the number of weeds in images (B) and (C), unlike the original YOLOX model in (A) and (D).

Moreover, despite the numerous benefits that YOLOX has brought about in our daily lives, it has its very own imperfections and associated challenges. Even though YOLOX largely performs well on various datasets, its effectiveness can vary based on the dataset's specific characteristics. Achieving optimal performance on datasets with unique traits often requires

fine-tuning and customisation. YOLOX can detect multiple object classes, but adapting it to new object categories or drastically different scenarios requires appropriate training data, which is likely to be a time-consuming and challenging process. Additionally, detecting very small or overly large objects within the same image remains difficult even though YOLOX has surpassed its predecessors. Addressing this problem may require specific architectural modifications or additional processing techniques (ProX, 2024).

Another possible limitation to various research studies on YOLOX is that YOLOX may not have been compared to an adequate number of alternative detection models. If YOLO remains to be the only one in each study that is studied, this means that the research conclusions and results may miss out on a lot of information and comparisons in terms of efficacy, widespread relevance and productivity of the various available methods for such object detection issues (Gheorghe et al, 2024). Generalising this, YOLOX may therefore be limited in terms of comparisons, especially with the knowledge that YOLOX is a budding tool that could lack research compared to those developed many years ago.

Another point to note is that many of these studies related to YOLO are targeted towards achieving a certain aim – they do not take into account industry standards, economic dynamics and regulatory problems that the real world faces today. This could greatly affect the usage and further advancement of YOLOX in object detection.

To further enhance the benefits or address the loopholes that YOLOX has, considering that no algorithm has been perfected to be able to resolve every issue on Earth, there has been a variety of case studies and research papers that have paired YOLOX with other tools to increase the accuracy of test results. According to a research study (Yan et al., 2023), YOLOX was implemented in a study of algal bioprocessing. This is particularly insightful, which deviates from the common research with YOLOX on autonomous vehicles.

Microalgae, which are microscopic phytoplankton, have the ability to influence carbon dioxide and nutrient levels from flue gases and wastewater, in conjunction with generating im-

portant biofuels, chemicals and fertilisers. To detect and pinpoint mixed algae species and types through manual methods would be highly inefficient and inaccurate – as such, the researchers utilised a YOLOX-s-based microalgae detection method.

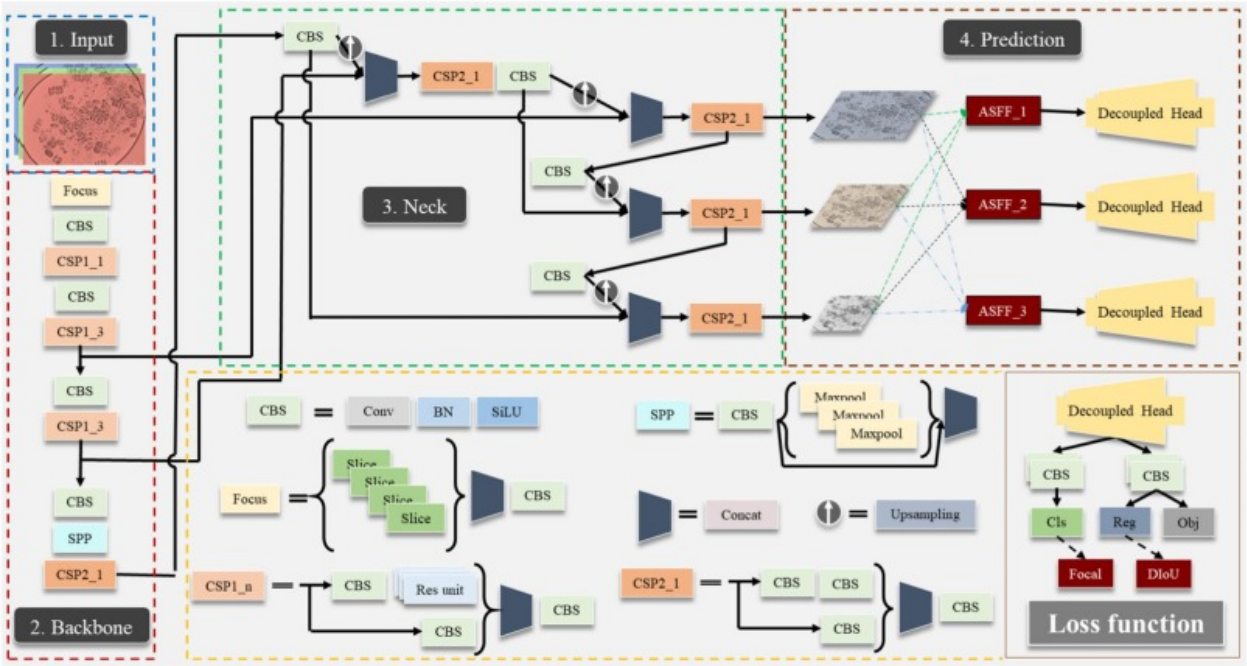

Figure 10: Structure of the Final Improved YOLOX-s (Yan et al., 2023)

To put the complex model into simple terms, as a starting model input, the algorithm was given a multi-level and morphology microalgae image dataset. To further enhance the results of the model, Focal Loss was used for the classification loss to solve the quantity disparity challenge. DIoU Loss was also included for regression loss, which shortened over $\frac{1}{5}$ of the processing time. Along with this, the ASFF module increased the general model achievement by merging features.

$$Precision = \frac{TP}{TP+FP} \tag{9}$$

$$Recall = \frac{TP}{TP+FN} \tag{10}$$

$$F1\_score = \frac{2 \times Precision \times Recall}{Precision+Recall} \tag{11}$$

Figure 11: Formulas used to Calculate Precision, Recall and F1 Score (Yan et al., 2023)

All in all, the Precision and Recall of this enhanced YOLOX-s in microalgae identification attained over 95% and 93% respectively, and the mAP was increased by 3.33%, in comparison with the earlier YOLOX-s.

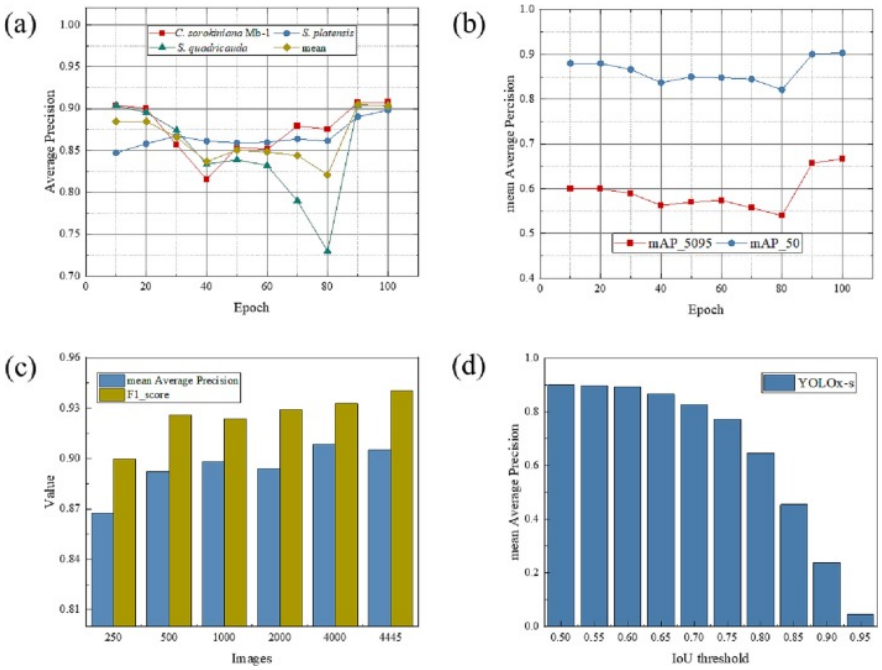

Figure 12: Graphical Results of YOLOX-s (Yan et al., 2023)

Overall, there is a positive trend for the mean average Precision and F1 score for YOLOX. These results suggest that YOLOX is a strong algorithm that is helpful in many aspects, reaching beyond just self-driving cars. Its effects are even more promising and purposeful when paired with other existing tools, which opens up many more research opportunities in terms of the integration and collective use of different materials.

3 Completed Work

SimOTA Equation

We have briefly discussed the SimOTA feature of YOLOX in the previous section, but we now dive deeper into analysing the associated equation.

$$c_{ij} = L_{cls}^{ij} + \lambda L_{reg}^{ij}$$

Figure 13: SimOTA equation

c_{ij}: total cost of matching a ground-truth object (g_i) with a predicted prediction (p_j)

L_{cls}^{ij}: measures the error in predicting the class of the ground-truth object (g_i)

L_{reg}^{ij}: measures the error in predicting the location and size of the ground-truth object (g_i)

λ: balancing coefficient, controls the relative importance of classification (L_{cls}^{ij}) and regression loss (L_{reg}^{ij})

The aim of this equation is to calculate the cost (c_{ij}) and select the top k predictions (p_j) with the lowest costs for each ground-truth object (g_i). By doing so, we can find the best match between predictions (p_j) and ground-truth objects (g_i) to ensure accurate label assignment for training.

Analysis of Baseline Results

We also explored the baseline performance of YOLOX by comparing its results to similar lightweight models.

Models	AP (%)	Parameters	GFLOPs
YOLOv4-Tiny [30]	21.7	6.06 M	6.96
PPYOLO-Tiny	22.7	4.20 M	-
YOLOX-Tiny	**32.8** (+10.1)	5.06 M	6.45
NanoDet[3]	23.5	0.95 M	1.20
YOLOX-Nano	**25.3** (+1.8)	0.91 M	1.08

Figure 14: Baseline results for YOLOX-Tiny and YOLOX-Nano on the COCO validation dataset (Ge et al., 2021a)

YOLOX-Tiny has an AP of 32.8%, which is significantly higher than its counterparts, indicating better accuracy. On the other hand, YOLOX-Nano has an AP of 25.3%, achieving balance between efficiency and accuracy. YOLOX-Tiny has moderately sized parameters (5.06M) while YOLOX-Nano has extremely lightweight parameters (0.91M), indicating lower model complexity. In terms of computational cost, YOLOX-Tiny has GFLOPs of 6.45, which is relatively efficient for its accuracy, while YOLOX-Nano has GFLOPs of 1.08, making it much more efficient but with slightly lower accuracy. This tells us that YOLOX-Tiny is suitable for tasks requiring higher precision, while YOLOX-Nano is ideal for resource-constrained environments. The analysis demonstrates the versatility of YOLOX models for a variety of use cases. These insights provide a solid foundation for future work, including performance evaluations and implementation refinements.

Deployment Results

We also deployed the model successfully. We made use of the COCO 2017 dataset to run a YOLOX-s model.

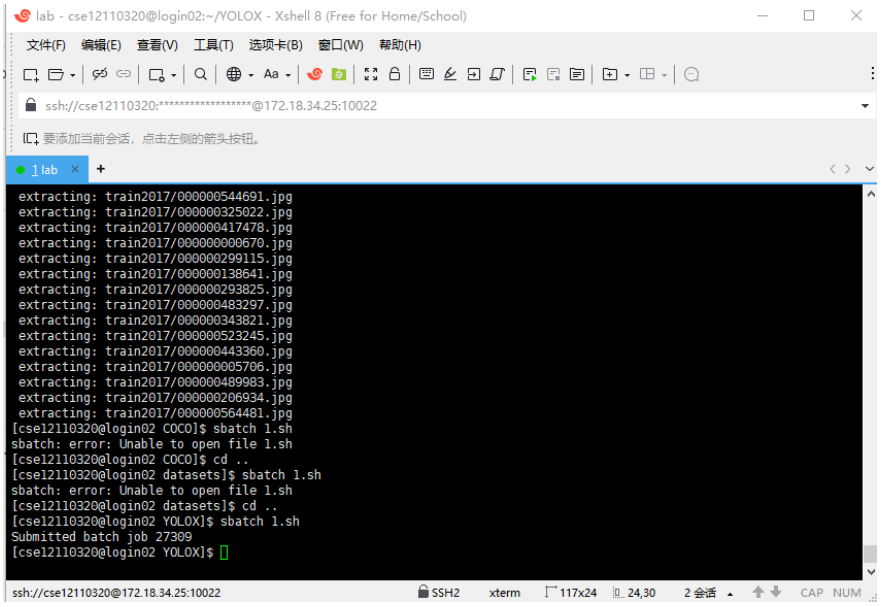

Figure 15: Connect to the server and run the model

Figure 16: Comments to execute

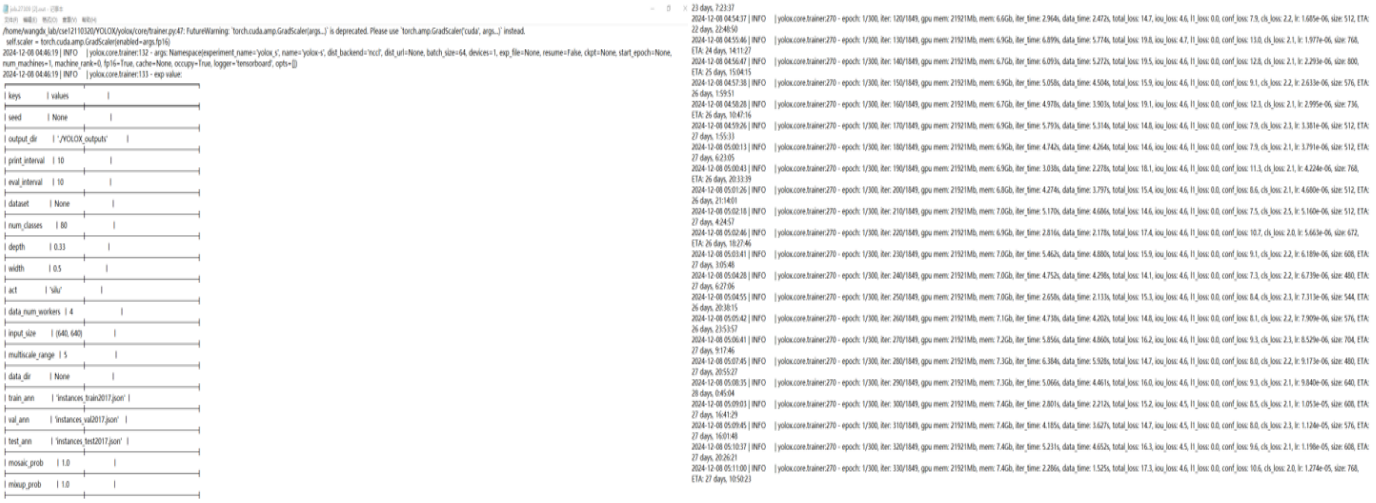

Figure 17: Part of the results

4 Research Plan and Expected Results

Week	Task	Expected Outcomes
12	• Topic selection • Identify the purpose and direction of the project • Conduct secondary research on the current problem	• Conduct a thorough literature review of YOLOX to identify existing gaps and establish a baseline understanding of the problem. • Identify the advantages and limitations of YOLOX in object detection.

Week	**Task**	**Expected Outcomes**
13	• Construct dataset • Train the results • Complete and submit report and ppt	• Gather a diverse dataset for testing, ensuring the dataset includes challenging scenarios like small objects. • Train the YOLOX model on the dataset using existing architecture as a baseline to establish starting performance metrics.
14	• Improve the accuracy of working code • Dataset refinement	• Implement proposed enhancements, such as fine-tuning hyperparameters to improve detection accuracy. • Clean and label datasets, ensuring proper annotations for bounding boxes and object identities.

Week	Task	Expected Outcomes
15	• Collect and analyze results • Visualize possible trends • Further refinement of code	• Evaluate the improved model on validation datasets, analyse performance metrics, and compare with baseline results to measure progress. • Generate visualisations of key trends to identify remaining issues and areas for further refinement. • Enhance code efficiency by reducing unnecessary processing and improving how the system handles tasks in real time.
16	• Summary of key results • Presentation of results	• Summarise the research findings, highlighting the performance gains, improvements made and any unresolved challenges. • Present the final outcomes, including comparisons to the baseline and implications for future work.

5 Potential Challenges and Solutions

Moreover, the research process poses a few potential difficulties.

Firstly, the dataset that will be used may lack diversity. It may not be able to represent a sufficiently wide range of scenarios for the detection algorithm to properly generalise objects within a certain video or image. This is especially significant for different and uncommon-sized objects, and extreme lighting conditions.

To resolve this, we can look to apply developed data augmentation techniques, such as brightness adjustments, rotation, noise addition and random scaling. Libraries such as TensorFlow, PyTorch, and Albumentations provide such robust augmentation options to simulate different conditions to improve the model.

Furthermore, synthetic data could be produced using useful tools, like Unity Perception and Unreal Engine, to generate scenarios that are less or underrepresented in the real dataset, such as congested urban areas or rare object classes.

To further boost the diversity of the datasets, additional real-world datasets can be collected to improve representation, particularly for autonomous driving or aerial surveillance. This is particularly crucial as large databases are able to offer ample training and validation samples to strengthen the soundness of the models and their transferability, execution quality and acceptance testing (Chan et al, 2021).

Secondly, training YOLOX on large training datasets presents significant computational challenges, making it difficult to achieve results within a short timeframe or on modest hardware. This includes limited memory capacity, which can lead to out-of-memory (OOM) errors when processing images or videos of high resolution, and extended training times due to the great number of parameters in YOLOX.

This also highlights the resource inefficiencies that may arise when cloud services are used

without proper scaling, creating unnecessarily high operational costs. Using smaller hardware with limited memory also poses the risk of overfitting as the model may not generalise well to larger datasets.

To tackle this, we can leverage cloud platforms, such as AWS and Google Cloud, to access high-performance computing resources. These platforms provide versatile, on-demand scalability, which will enable us to scale computing power based on our needs (Charter Global, 2024). During the initial testing stage, we can start with a small dataset, before gradually increasing the sample size, until the limit that the computational resources allow. Cloud platforms are also an option, armed with tools to supervise resource usage, helping to optimise training.

Additionally, the implementation of model pruning to reduce unnecessary parameters can also be put in place to lower computational loads (Neo, 2024). Techniques such as weight or neuron pruning could be employed to reduce the size of the model with minimal loss of accuracy. Another effective approach could be to use mixed-precision training, which reduces the computational cost of training by using lower-precision data types for certain operations without sacrificing model performance.

Thirdly, the model may fail to effectively generalise to fresh data, even if it performs well on the training data, particularly with small or unbalanced datasets (Geeks for Geeks, 2024a). This issue mainly arises when the model overfits to the training data, learning patterns specific to the training set that may not be the same as those in the new, unseen data. More frequent classes could be favoured, while less-represented ones are neglected, reducing the model's capability to detect objects in real-world scenarios that differ from training.

Alleviating this issue has several methods. One would be using dropout. Randomly setting a proportion of the weights to zero during training would prevent the model from becoming over-reliant on specific neurons, which encourages more robust learning and reduces overfitting (Srivastava et al., 2014).

Second, weight decay adds a penalty to the loss function (Jain, 2024) based on the magnitude of the model's weight. This discourages excessively or overly large weights, which helps to prevent overfitting, and allows the model to learn more generalised features that are not too specific to the training data.

Third would be label smoothing. This prevents the model from being too confident in its predictions by softening the target labels during training, decreasing the likelihood of overfitting to very specific labels (Geeks for Geeks, 2024b), similar to the aims of weight decay.

The last strategy would be to address class imbalance. The imbalance could make the model more biased towards more frequent classes, which can be mitigated through oversampling underrepresented classes (Chawla et al, 2002) or using weighted loss functions, in which higher weights are assigned to less frequent classes. The model will therefore pay more attention to these classes during training, aiding in solving the issue.

6 References

Boesch, G. (2024, March 11). *YOLOX Explained: Features, Architecture and Applications.* viso.ai. https://viso.ai/computer-vision/yolox/

Chan, H.-P., Samala, R. K., Hadjiiski, L. M., & Zhou, C. (2021, January 1). *Deep Learning in Medical Image Analysis.* National Library of Medicine. https://pmc.ncbi.nlm.nih.gov/articles/PMC7442218/

Charter Global. (2024, November 12). *Choosing the Right Cloud Provider: An In-Depth comparison of AWS, Azure, and Google Cloud.* https://www.charterglobal.com/aws-vs-azure-vs-google-cloud-platforms/

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002, June 1). *SMOTE: Synthetic Minority Over-sampling Technique.* Journal of Artificial Intelligence Research. https://www.jair.org/index.php/jair/article/view/10302

Chen, W., Li, Y., Tian, Z., & Zhang, F. (2023, September). *2D and 3D object detection algorithms from images: A Survey.* Science Direct. https://www.sciencedirect.com/science/article/pii/S2590005623000309

Data-Driven Science. (2023, February 22). *Object Detection with YOLOv5: Understanding Anchor Boxes.* https://datadrivenscience.com/object-detection-with-yolov5-understanding-anchor-boxes/

Eurostat. (2024, April). *Statistics Explained.* https://ec.europa.eu/eurostat/statistics-explained/index.php?title=SDG_-_Introduction

Ge Z., Liu, S., Wang, F., & Li, Z. (2021a, July 18). *YOLOX: Exceeding YOLO Series in 2021.* Semantic Scholar. https://www.semanticscholar.org/paper/YOLOX:-Exceeding-YOLO-Series-in-2021-Ge-Liu/c01b385205e488a731c8c8c11c0c494d426beb03

Ge, Z., Liu, S., Li, Z., Yoshie, O., & Sun, J. (2021b, March 26). *OTA: Optimal Transport Assignment for Object Detection.* arXiv. https://arxiv.org/abs/2103.14259

Geeks for Geeks. (2024a, September 13). *What is Imbalanced Dataset.* https://www.geeksforgeeks.org/what-is-imbalanced-dataset/

Geeks for Geeks. (2024b, September 12). *What is Label Smoothing?* https://www.geeksforgeeks.org/what-is-label-smoothing/

Gheorghe, C., Duguleana, M., Boboc, R. G., & Postelnicu, C. C. (2024, October 31). *Analyzing Real-Time Object Detection with YOLO Algorithm in Automotive Applications: A Review.* Tech Science Press. https://www.techscience.com/CMES/v141n3/58495

Hu, J., Lv, H., Yang, L., & Qiao, P. (2022 November). *The computational processes of simOTA.* ResearchGate. https://www.researchgate.net/figure/The-computational-processes-of-simOTA_fig2_365056716

Jain, S. (2024, December 4). *Regularization in Deep Learning with Python Code.* Analytics Vidhya. https://www.analyticsvidhya.com/blog/2018/04/fundamentals-deep-learning-regularization-techniques/#h-how-does-regularization-help-reduce-overfitting

Mongaras, G. (2022, May 20). *YOLOX Explanation — SimOTA For Dynamic Label Assignment.* Medium. https://gmongaras.medium.com/yolox-explanation-simota-for-dynamic-label-assignment-8fa5ae397f76

Neo, M. (2024, February 29). *A Comprehensive Guide to Neural Network Model Pruning.* Datature. https://www.datature.io/blog/a-comprehensive-guide-to-neural-network-model-pruning

ProX. (2024, June 12). *YOLOX Explained: Features, Architecture and Applications.* https://www.proxpc.com/blogs/yolox-explained-features-architecture-and-applications

Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016, May 9). *You Only Look Once: Unified, Real-Time Object Detection.* https://arxiv.org/pdf/1506.02640

Saadani, T. (2024, January 17). *The Evolution and Applications of YOLO Object Detection: A Comprehensive Exploration.* Medium.

https://medium.com/ubiai-nlp/the-evolution-and-applications-of-yolo-object-detection-a-comprehensive-exploration-9b8dbc0ef2a5

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014, January 1). *Dropout: A Simple Way to Prevent Neural Networks from Overfitting.* Journal of Machine Learning Research. https://www.cs.toronto.edu/~rsalakhu/papers/srivastava14a.pdf

Świeżewski, J. (2020, May 22). *YOLO Algorithm and YOLO Object Detection.* Appsilon. https://www.appsilon.com/post/object-detection-yolo-algorithm

Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023, November 20). *A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS.* MDPI. https://www.mdpi.com/2504-4990/5/4/83

Yan, H., Peng, X., Chen, C., Xia A., Huang, Y., Zhu, X., Zhu X., & Liao, Q. (2023, July). *YOLOx model-based object detection for microalgal bioprocess.* ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S2211926423002114

YOLOvx. (n.d.). *Evolution of YOLO: A Timeline of Versions and Advancements in Object Detection.* https://yolovx.com/evolution-of-yolo-a-timeline-of-versions-and-advancements-in-object-detection/

Zhang, J., Chen, Z., Yan, G., & Wang, Y. (2023 October). *Faster and Lightweight: An Improved YOLOv5 Object Detector for Remote Sensing Images. ResearchGate.* https://www.researchgate.net/publication/374748650_Faster_and_Lightweight_An_Improved_YOLOv5_Object_Detector_for_Remote_Sensing_Images

Zhao, Z., He, C., Zhao G., Zhou, J., & Hao, K. (2023, August). *RA-YOLOX: Re-parameterization align decoupled head and novel label assignment scheme based on YOLOX.* ScienceDirect. https://www.sciencedirect.com/science/article/pii/S0031320323002790

Zhu, H., Zhang, Y., Mu, D., Bai L., Wu, X., Zhuang, H., & Li, H. (2024, March). *Research on improved YOLOx weed detection based on lightweight attention module.* ScienceDirect. https://www.sciencedirect.com/science/article/pii/S026121942300385X