YOLOv11: An Overview of the Key Architectural Enhancements

Rahima Khanam[*] **and Muhammad Hussain**

Department of Computer Science, Huddersfield University, Queensgate, Huddersfield HD1 3DH, UK;
[*]Correspondence: rahima.khanam@hud.ac.uk;

October 24, 2024

ABSTRACT

This study presents an architectural analysis of YOLOv11, the latest iteration in the YOLO (You Only Look Once) series of object detection models. We examine the models architectural innovations, including the introduction of the C3k2 (Cross Stage Partial with kernel size 2) block, SPPF (Spatial Pyramid Pooling - Fast), and C2PSA (Convolutional block with Parallel Spatial Attention) components, which contribute in improving the models performance in several ways such as enhanced feature extraction. The paper explores YOLOv11's expanded capabilities across various computer vision tasks, including object detection, instance segmentation, pose estimation, and oriented object detection (OBB). We review the model's performance improvements in terms of mean Average Precision (mAP) and computational efficiency compared to its predecessors, with a focus on the trade-off between parameter count and accuracy. Additionally, the study discusses YOLOv11's versatility across different model sizes, from nano to extra-large, catering to diverse application needs from edge devices to high-performance computing environments. Our research provides insights into YOLOv11's position within the broader landscape of object detection and its potential impact on real-time computer vision applications.

Keywords Automation; Computer Vision; YOLO; YOLOV11; Object Detection; Real-Time Image processing; YOLO version comparison

1 Introduction

Computer vision, a rapidly advancing field, enables machines to interpret and understand visual data [1]. A crucial aspect of this domain is object detection[2], which involves the precise identification and localization of objects within images or video streams[3]. Recent years have witnessed remarkable progress in algorithmic approaches to address this challenge [4].

A pivotal breakthrough in object detection came with the introduction of the You Only Look Once (YOLO) algorithm by Redmon et al. in 2015 [5]. This innovative approach, as its name suggests, processes the entire image in a single pass to detect objects and their locations. YOLO's methodology diverges from traditional two-stage detection processes by framing object detection as a regression problem [5]. It employs a single convolutional neural network to simultaneously predict bounding boxes and class probabilities across the entire image [6], streamlining the detection pipeline compared to more complex traditional methods.

YOLOv11 is the latest iteration in the YOLO series, building upon the foundation established by YOLOv1. Unveiled at the YOLO Vision 2024 (YV24) conference, YOLOv11 represents a significant leap forward in real-time object detection technology. This new version introduces substantial enhancements in both architecture and training methodologies, pushing the boundaries of accuracy, speed, and efficiency.

YOLOv11's innovative design incorporates advanced feature extraction techniques, allowing for more nuanced detail capture while maintaining a lean parameter count. This results in improved accuracy across a diverse range of computer

vision (CV) tasks, from object detection to classification. Furthermore, YOLOv11 achieves remarkable gains in processing speed, substantially enhancing real-time performance capabilities.

In the following sections, this paper will provide a comprehensive analysis of YOLOv11's architecture, exploring its key components and innovations. We will examine the evolution of YOLO models, leading up to the development of YOLOv11. The study will delve into the model's expanded capabilities across various CV tasks, including object detection, instance segmentation, pose estimation, and oriented object detection. We will also review YOLOv11's performance improvements in terms of accuracy and computational efficiency compared to its predecessors, with a particular focus on its versatility across different model sizes. Finally, we will discuss the potential impact of YOLOv11 on real-time CV applications and its position within the broader landscape of object detection technologies.

2 Evolution of YOLO models

Table 1 illustrates the progression of YOLO models from their inception to the most recent versions. Each iteration has brought significant improvements in object detection capabilities, computational efficiency, and versatility in handling various CV tasks.

Table 1: YOLO: Evolution of models

Release	Year	Tasks	Contributions	Framework
YOLO [5]	2015	Object Detection, Basic Classification	Single-stage object detector	Darknet
YOLOv2 [7]	2016	Object Detection, Improved Classification	Multi-scale training, dimension clustering	Darknet
YOLOv3 [8]	2018	Object Detection, Multi-scale Detection	SPP block, Darknet-53 backbone	Darknet
YOLOv4 [9]	2020	Object Detection, Basic Object Tracking	Mish activation, CSPDarknet-53 backbone	Darknet
YOLOv5 [10]	2020	Object Detection, Basic Instance Segmentation (via custom modifications)	Anchor-free detection, SWISH activation, PANet	PyTorch
YOLOv6 [11]	2022	Object Detection, Instance Segmentation	Self-attention, anchor-free OD	PyTorch
YOLOv7 [12]	2022	Object Detection, Object Tracking, Instance Segmentation	Transformers, E-ELAN reparameterisation	PyTorch
YOLOv8 [13]	2023	Object Detection, Instance Segmentation, Panoptic Segmentation, Keypoint Estimation	GANs, anchor-free detection	PyTorch
YOLOv9 [14]	2024	Object Detection, Instance Segmentation	PGI and GELAN	PyTorch
YOLOv10 [15]	2024	Object Detection	Consistent dual assignments for NMS-free training	PyTorch

This evolution showcases the rapid advancement in object detection technologies, with each version introducing novel features and expanding the range of supported tasks. From the original YOLO's groundbreaking single-stage detection to YOLOv10's NMS-free training, the series has consistently pushed the boundaries of real-time object detection.

The latest iteration, YOLO11, builds upon this legacy with further enhancements in feature extraction, efficiency, and multi-task capabilities. Our subsequent analysis will delve into YOLO11's architectural innovations, including its improved backbone and neck structures, and its performance across various computer vision tasks such as object detection, instance segmentation, and pose estimation.

3 What is YOLOv11?

The evolution of the YOLO algorithm reaches new heights with the introduction of YOLOv11 [16], representing a significant advancement in real-time object detection technology. This latest iteration builds upon the strengths of its predecessors while introducing novel capabilities that expand its utility across diverse CV applications.

YOLOv11 distinguishes itself through its enhanced adaptability, supporting an expanded range of CV tasks beyond traditional object detection. Notable among these are posture estimation and instance segmentation, broadening the

model's applicability in various domains. YOLOv11's design focuses on balancing power and practicality, aiming to address specific challenges across various industries with increased accuracy and efficiency.

This latest model demonstrates the ongoing evolution of real-time object detection technology, pushing the boundaries of what's possible in CV applications. Its versatility and performance improvements position YOLOv11 as a significant advancement in the field, potentially opening new avenues for real-world implementation across diverse sectors.

4 Architectural footprint of Yolov11

The YOLO framework revolutionized object detection by introducing a unified neural network architecture that simultaneously handles both bounding box regression and object classification tasks [17]. This integrated approach marked a significant departure from traditional two-stage detection methods, offering end-to-end training capabilities through its fully differentiable design.

At its core, the YOLO architecture consists of three fundamental components. First, the **backbone** serves as the primary feature extractor, utilizing convolutional neural networks to transform raw image data into multi-scale feature maps. Second, the **neck** component acts as an intermediate processing stage, employing specialized layers to aggregate and enhance feature representations across different scales. Third, the **head** component functions as the prediction mechanism, generating the final outputs for object localization and classification based on the refined feature maps.

Building on this established architecture, YOLO11 extends and enhances the foundation laid by YOLOv8, introducing architectural innovations and parameter optimizations to achieve superior detection performance as illustrated in Figure 1. The following sections detail the key architectural modifications implemented in YOLO11:
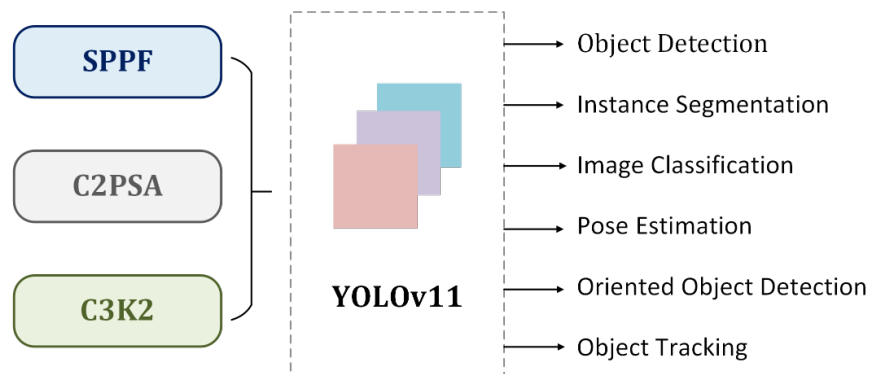

Figure 1: Key architectural modules in YOLO11

4.1 Backbone

The backbone is a crucial component of the YOLO architecture, responsible for extracting features from the input image at multiple scales. This process involves stacking convolutional layers and specialized blocks to generate feature maps at various resolutions.

4.1.1 Convolutional Layers

YOLOv11 maintains a structure similar to its predecessors, utilizing initial convolutional layers to downsample the image. These layers form the foundation of the feature extraction process, gradually reducing spatial dimensions while increasing the number of channels. A significant improvement in YOLO11 is the introduction of the C3k2 block, which replaces the C2f block used in previous versions [18]. The C3k2 block is a more computationally efficient implementation of the Cross Stage Partial (CSP) Bottleneck. It employs two smaller convolutions instead of one large convolution, as seen in YOLOv8 [13]. The "k2" in C3k2 indicates a smaller kernel size, which contributes to faster processing while maintaining performance.

4.1.2 SPPF and C2PSA

YOLO11 retains the Spatial Pyramid Pooling - Fast (SPPF) block from previous versions but introduces a new Cross Stage Partial with Spatial Attention (C2PSA) block after it [18]. The C2PSA block is a notable addition that enhances

spatial attention in the feature maps. This spatial attention mechanism allows the model to focus more effectively on important regions within the image. By pooling features spatially, the C2PSA block enables YOLO11 to concentrate on specific areas of interest, potentially improving detection accuracy for objects of varying sizes and positions.

4.2 Neck

The neck combines features at different scales and transmits them to the head for prediction. This process typically involves upsampling and concatenation of feature maps from different levels, enabling the model to capture multi-scale information effectively.

4.2.1 C3k2 Block

YOLO11 introduces a significant change by replacing the C2f block in the neck with the C3k2 block. The C3k2 block is designed to be faster and more efficient, enhancing the overall performance of the feature aggregation process. After upsampling and concatenation, the neck in YOLO11 incorporates this improved block, resulting in enhanced speed and performance [18].

4.2.2 Attention Mechanism

A notable addition to YOLO11 is its increased focus on spatial attention through the C2PSA module. This attention mechanism enables the model to concentrate on key regions within the image, potentially leading to more accurate detection, especially for smaller or partially occluded objects. The inclusion of C2PSA sets YOLO11 apart from its predecessor, YOLOv8, which lacks this specific attention mechanism [18].

4.3 Head

The head of YOLOv11 is responsible for generating the final predictions in terms of object detection and classification. It processes the feature maps passed from the neck, ultimately outputting bounding boxes and class labels for objects within the image.

4.3.1 C3k2 Block

In the head section, YOLOv11 utilizes multiple C3k2 blocks to efficiently process and refine the feature maps. The C3k2 blocks are placed in several pathways within the head, functioning to process multi-scale features at different depths. The C3k2 block exhibits flexibility depending on the value of the c3k parameter:

- When c3k = False, the C3k2 module behaves similarly to the C2f block, utilizing a standard bottleneck structure.
- When c3k = True, the bottleneck structure is replaced by the C3 module, which allows for deeper and more complex feature extraction.

Key characteristics of the C3k2 block:

- **Faster processing:** The use of two smaller convolutions reduces the computational overhead compared to a single large convolution, leading to quicker feature extraction.
- **Parameter efficiency:** C3k2 is a more compact version of the CSP bottleneck, making the architecture more efficient in terms of the number of trainable parameters.

Another notable addition is the C3k block, which offers enhanced flexibility by allowing customizable kernel sizes. The adaptability of C3k is particularly useful for extracting more detailed features from images, contributing to improved detection accuracy.

4.3.2 CBS Blocks

The head of YOLOv11 includes several CBS (Convolution-BatchNorm-Silu) [19] layers after the C3k2 blocks. These layers further refine the feature maps by:

- Extracting relevant features for accurate object detection.
- Stabilizing and normalizing the data flow through batch normalization.

- Utilizing the Sigmoid Linear Unit (SiLU) activation function for non-linearity, which improves model performance.

CBS blocks serve as foundational components in both feature extraction and the detection process, ensuring that the refined feature maps are passed to the subsequent layers for bounding box and classification predictions.

4.3.3 Final Convolutional Layers and Detect Layer

Each detection branch ends with a set of Conv2D layers, which reduce the features to the required number of outputs for bounding box coordinates and class predictions. The final Detect layer consolidates these predictions, which include:

- Bounding box coordinates for localizing objects in the image.
- Objectness scores that indicate the presence of objects.
- Class scores for determining the class of the detected object.

5 Key Computer Vision Tasks Supported by YOLO11

YOLO11 supports a diverse range of CV tasks, showcasing its versatility and power in various applications. Here's an overview of the key tasks:

1. **Object Detection:** YOLO11 excels in identifying and localizing objects within images or video frames, providing bounding boxes for each detected item [20]. This capability finds applications in surveillance systems, autonomous vehicles, and retail analytics, where precise object identification is crucial [21].

2. **Instance Segmentation:** Going beyond simple detection, YOLO11 can identify and separate individual objects within an image down to the pixel level [20]. This fine-grained segmentation is particularly valuable in medical imaging for precise organ or tumor delineation, and in manufacturing for detailed defect detection [21].

3. **Image Classification:** YOLOv11 is capable of classifying entire images into predetermined categories, making it ideal for applications like product categorization in e-commerce platforms or wildlife monitoring in ecological studies [21].

4. **Pose Estimation:** The model can detect specific key points within images or video frames to track movements or poses. This capability is beneficial for fitness tracking applications, sports performance analysis, and various healthcare applications requiring motion assessment [21].

5. **Oriented Object Detection (OBB):** YOLO11 introduces the ability to detect objects with an orientation angle, allowing for more precise localization of rotated objects. This feature is especially valuable in aerial imagery analysis, robotics, and warehouse automation tasks where object orientation is crucial [21].

6. **Object Tracking:** It identifies and traces the path of objects in a sequence of images or video frames[21]. This real-time tracking capability is essential for applications such as traffic monitoring, sports analysis, and security systems.

Table 2 outlines the YOLOv11 model variants and their corresponding tasks. Each variant is designed for specific use cases, from object detection to pose estimation. Moreover, all variants support core functionalities like inference, validation, training, and export, making YOLOv11 a versatile tool for various CV applications.

6 Advancements and Key Features of YOLOv11

YOLOv11 represents a significant advancement in object detection technology, building upon the foundations laid by its predecessors, YOLOv9 and YOLOv10, which were introduced earlier in 2024. This latest iteration from Ultralytics showcases enhanced architectural designs, more sophisticated feature extraction techniques, and refined training methodologies. The synergy of YOLOv11's rapid processing, high accuracy, and computational efficiency positions it as one of the most formidable models in Ultralytics' portfolio to date [22]. A key strength of YOLOv11 lies in its refined architecture, which facilitates the detection of subtle details even in challenging scenarios. The model's improved feature extraction capabilities allow it to identify and process a broader range of patterns and intricate elements within images. Compared to earlier versions, YOLOv11 introduces several notable enhancements:

Table 2: YOLOv11 Model Variants and Tasks

Model	Variants	Task	Inference	Validation	Training	Export
YOLOv11	yolo11-nano yolo11-small yolo11-medium yolo11-large yolo11-xlarge	Detection	✓	✓	✓	✓
YOLOv11-seg	yolo11-nano-seg yolo11-small-seg yolo11-medium-seg yolo11-large-seg yolo11-xlarge-seg	Instance Segmentation	✓	✓	✓	✓
YOLOv11-pose	yolo11-nano-pose yolo11-small-pose yolo11-medium-pose yolo11-large-pose yolo11-xlarge-pose	Pose/Keypoints	✓	✓	✓	✓
YOLOv11-obb	yolo11-nano-obb yolo11-small-obb yolo1-medium-obb yolo11-large-obb yolo11-xlarge-obb	Oriented Detection	✓	✓	✓	✓
YOLOv11-cls	yolo11-nano-cls yolo11-small-cls yolo11-medium-cls yolo11-large-cls yolo11-xlarge-cls	Classification	✓	✓	✓	✓

1. **Enhanced precision with reduced complexity:** The YOLOv11m variant achieves superior mean Average Precision (mAP) scores on the COCO dataset while utilizing 22% fewer parameters than its YOLOv8m counterpart, demonstrating improved computational efficiency without compromising accuracy [23].

2. **Versatility in CV tasks:** YOLOv11 exhibits proficiency across a diverse array of CV applications, including pose estimation, object recognition, image classification, instance segmentation, and oriented bounding box (OBB) detection [23].

3. **Optimized speed and performance:** Through refined architectural designs and streamlined training pipelines, YOLOv11 achieves faster processing speeds while maintaining a balance between accuracy and computational efficiency [23].

4. **Streamlined parameter count:** The reduction in parameters contributes to faster model performance without significantly impacting the overall accuracy of YOLOv11 [22].

5. **Advanced feature extraction:** YOLOv11 incorporates improvements in both its backbone and neck architectures, resulting in enhanced feature extraction capabilities and, consequently, more precise object detection [23].

6. **Contextual adaptability:** YOLOv11 demonstrates versatility across various deployment scenarios, including cloud platforms, edge devices, and systems optimized for NVIDIA GPUs [23].

YOLOv11 model demonstrates significant advancements in both inference speed and accuracy compared to its predecessors. In the benchmark analysis, YOLOv11 was compared against several of its predecessors including variants such as YOLOv5 [24] through to the more recent variants such as YOLOv10. As presented in Figure 2, YOLOv11 consistently outperforms these models, achieving superior mAP on the COCO dataset while maintaining a faster inference rate [25].

The performance comparison graph depicted in Figure 2 overs several key insights. The YOLOv11 variants (11n, 11s, 11m, and 11x) form a distinct performance frontier, with each model achieving higher COCO mAP^{50-95} scores at their respective latency points. Notably, the YOLOv11x achieves approximately 54.5% mAP^{50-95} at 13ms latency, surpassing all previous YOLO iterations. The intermediate variants, particularly YOLOv11m, demonstrate exceptional efficiency by achieving comparable accuracy to larger models from previous generations while requiring significantly less processing time.

A particularly noteworthy observation is the performance leap in the low-latency regime (2-6ms), where YOLOv11s maintains high accuracy (approximately 47% mAP^{50-95}) while operating at speeds previously associated with much less accurate models. This represents a crucial advancement for real-time applications where both speed and accuracy are critical. The improvement curve of YOLOv11 also shows better scaling characteristics across its model variants, suggesting more efficient utilization of additional computational resources compared to previous generations.
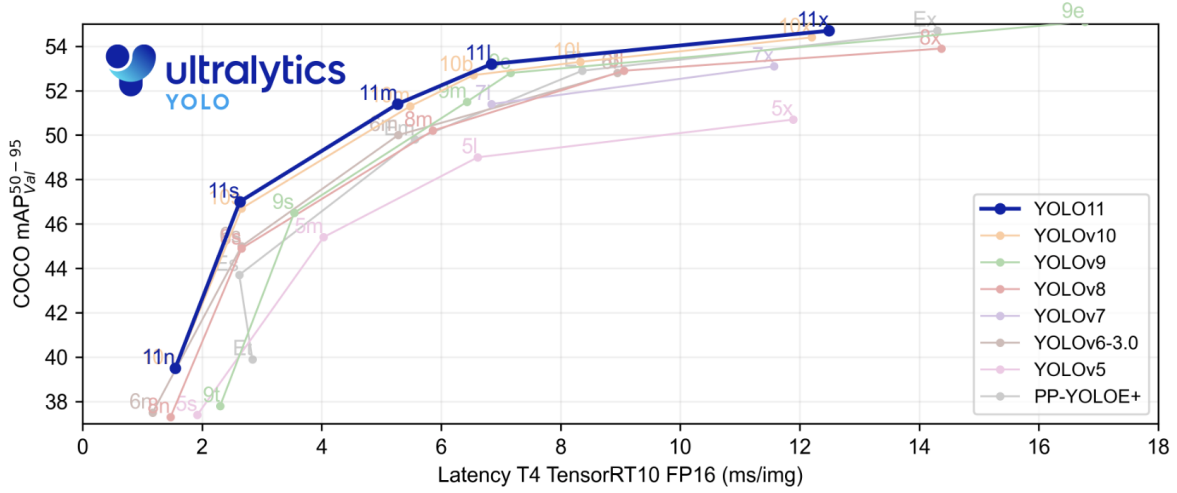
Figure 2: Benchmarking YOLOv11 Against Previous Versions [23]

7 Discussion

YOLO11 marks a significant leap forward in object detection technology, building upon its predecessors while introducing innovative enhancements. This latest iteration demonstrates remarkable versatility and efficiency across various CV tasks.

1. **Efficiency and Scalability:** YOLO11 introduces a range of model sizes, from nano to extra-large, catering to diverse application needs. This scalability allows for deployment in scenarios ranging from resource-constrained edge devices to high-performance computing environments. The nano variant, in particular, showcases impressive speed and efficiency improvements over its predecessor, making it ideal for real-time applications.

2. **Architectural Innovations:** The model incorporates novel architectural elements that enhance its feature extraction and processing capabilities. The incorporation of novel elements such as the C3k2 block, SPPF, and C2PSA contributes to more effective feature extraction and processing. These enhancements allow the model to better analyze and interpret complex visual information, potentially leading to improved detection accuracy across various scenarios.

3. **Multi-Task Proficiency:** YOLO11's versatility extends beyond object detection, encompassing tasks such as instance segmentation, image classification, pose estimation, and oriented object detection. This multi-faceted approach positions YOLO11 as a comprehensive solution for diverse CV challenges.

4. **Enhanced Attention Mechanisms:** A key advancement in YOLO11 is the integration of sophisticated spatial attention mechanisms, particularly the C2PSA component. This feature enables the model to focus more effectively on critical regions within an image, enhancing its ability to detect and analyze objects. The improved attention capability is especially beneficial for identifying complex or partially occluded objects, addressing a common challenge in object detection tasks. This refinement in spatial awareness contributes to YOLO11's overall performance improvements, particularly in challenging visual environments.

5. **Performance Benchmarks:** Comparative analyses reveal YOLO11's superior performance, particularly in its smaller variants. The nano model, despite a slight increase in parameters, demonstrates enhanced inference speed and frames per second (FPS) compared to its predecessor. This improvement suggests that YOLO11 achieves a favorable balance between computational efficiency and detection accuracy.

6. **Implications for Real-World Applications:** The advancements in YOLO11 have significant implications for various industries. Its improved efficiency and multi-task capabilities make it particularly suitable for applications in autonomous vehicles, surveillance systems, and industrial automation. The model's ability to perform well across different scales also opens up new possibilities for deployment in resource-constrained environments without compromising on performance.

8 Conclusion

YOLOv11 represents a significant advancement in the field of CV, offering a compelling combination of enhanced performance and versatility. This latest iteration of the YOLO architecture demonstrates marked improvements in accuracy and processing speed, while simultaneously reducing the number of parameters required. Such optimizations make YOLOv11 particularly well-suited for a wide range of applications, from edge computing to cloud-based analysis.

The model's adaptability across various tasks, including object detection, instance segmentation, and pose estimation, positions it as a valuable tool for diverse industries such as emotion detection [26], healthcare [27] and various other industries [17]. Its seamless integration capabilities and improved efficiency make it an attractive option for businesses seeking to implement or upgrade their CV systems. In summary, YOLOv11's blend of enhanced feature extraction, optimized performance, and broad task support establishes it as a formidable solution for addressing complex visual recognition challenges in both research and practical applications.

References

[1] Milan Sonka, Vaclav Hlavac, and Roger Boyle. *Image processing, analysis and machine vision*. Springer, 2013.

[2] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years: A survey. *Proceedings of the IEEE*, 111(3):257–276, 2023.

[3] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu. Object detection with deep learning: A review. *IEEE transactions on neural networks and learning systems*, 30(11):3212–3232, 2019.

[4] Muhammad Hussain and Rahima Khanam. In-depth review of yolov1 to yolov10 variants for enhanced photovoltaic defect detection. In *Solar*, volume 4, pages 351–386. MDPI, 2024.

[5] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[6] Juan Du. Understanding of object detection based on cnn family and yolo. In *Journal of Physics: Conference Series*, volume 1004, page 012029. IOP Publishing, 2018.

[7] Joseph Redmon and Ali Farhadi. Yolo9000: better, faster, stronger. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7263–7271, 2017.

[8] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[9] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4: Optimal speed and accuracy of object detection. *arXiv preprint arXiv:2004.10934*, 2020.

[10] Roboflow Blog Jacob Solawetz. What is yolov5? a guide for beginners., 2020. Accessed: 21 October 2024.

[11] Chuyi Li, Lulu Li, Hongliang Jiang, Kaiheng Weng, Yifei Geng, Liang Li, Zaidan Ke, Qingyuan Li, Meng Cheng, Weiqiang Nie, et al. Yolov6: A single-stage object detection framework for industrial applications. *arXiv preprint arXiv:2209.02976*, 2022.

[12] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. Yolov7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7464–7475, 2023.

[13] Francesco Jacob Solawetz. What is yolov8? the ultimate guide, 2023. Accessed: 21 October 2024.

[14] Chien-Yao Wang, I-Hau Yeh, and Hong-Yuan Mark Liao. Yolov9: Learning what you want to learn using programmable gradient information. *arXiv preprint arXiv:2402.13616*, 2024.

[15] Ao Wang, Hui Chen, Lihao Liu, Kai Chen, Zijia Lin, Jungong Han, and Guiguang Ding. Yolov10: Real-time end-to-end object detection. *arXiv preprint arXiv:2405.14458*, 2024.

[16] Glenn Jocher and Jing Qiu. Ultralytics yolo11, 2024.

[17] Rahima Khanam, Muhammad Hussain, Richard Hill, and Paul Allen. A comprehensive review of convolutional neural networks for defect detection in industrial applications. *IEEE Access*, 2024.

[18] Satya Mallick. Yolo - learnopencv. https://learnopencv.com/yolo11/, 2024. Accessed: 2024-10-21.

[19] Jingwen Feng, Qiaofeng An, Jiahao Zhang, Shuxun Zhou, Guangwei Du, and Kai Yang. Application of yolov7-tiny in the detection of steel surface defects. In *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, pages 2241–2245. IEEE, 2024.

[20] Ultralytics. Instance segmentation and tracking, 2024. Accessed: 2024-10-21.

[21] Ultralytics Abirami Vina. Ultralytics yolo11 has arrived: Redefine what's possible in ai, 2024. Accessed: 2024-10-21.

[22] Viso.AI Gaudenz Boesch. Yolov11: A new iteration of "you only look once. `https://viso.ai/computer-vision/yolov11/`, 2024. Accessed: 2024-10-21.

[23] Ultralytics. Ultralytics yolov11. `https://docs.ultralytics.com/models/yolo11/s`, 2024. Accessed: 21-Oct-2024.

[24] Rahima Khanam and Muhammad Hussain. What is yolov5: A deep look into the internal features of the popular object detector. *arXiv preprint arXiv:2407.20892*, 2024.

[25] DigitalOcean. What's new in yolov11 transforming object detection once again part 1, 2024. Accessed: 2024-10-21.

[26] Muhammad Hussain and Hussain Al-Aqrabi. Child emotion recognition via custom lightweight cnn architecture. In *Kids Cybersecurity Using Computational Intelligence Techniques*, pages 165–174. Springer, 2023.

[27] Burcu Ataer Aydin, Muhammad Hussain, Richard Hill, and Hussain Al-Aqrabi. Domain modelling for a lightweight convolutional network focused on automated exudate detection in retinal fundus images. In *2023 9th International Conference on Information Technology Trends (ITT)*, pages 145–150. IEEE, 2023.