Proceedings of CONF-SEML 2026 Symposium: Computational Analysis and Modeling in Complex Intelligent Systems
DOI: 10.54254/2755-2721/2026.GU33444

An Investigation of Mainstream YOLO Object Detection Models and Applications

Yiping Xu

Beijing-Dublin International College, Beijing University of Technology, Beijing, China
yiping.xu1@ucdconnect.ie

Abstract. With the rapid progress of object detection algorithms, the You Only Look Once (YOLO) model, differing from traditional two stage detector, relies on single-stage detection architecture to achieve high accuracy while ensuring fast inference speed. This paper systematically introduces the network architectures and the applications of the four mainstream YOLO models included YOLOv5, YOLOv8, YOLO11 and YOLO26. This article aims to form a review to help relevant learners establish an overall awareness and preliminary understanding of the YOLO series models. Meanwhile, in the paper, the author compares and analyzes the core architecture and performance of YOLO11 and YOLO26, which is current newer version. The result indicates that compared with YOLO11, YOLO26 possessed a better performance with regard to detection accuracy, inference speed, end-to-end inference without NMS processing, and lightweight deployment. However, since YOLO26 was released recently, its practical applications and related research in various fields are still relatively limited. Furthermore, in the article, the author summarizes the utilization of YOLO models in the domain of autonomous driving, industrial defect detection and security system, where real-time detection is greatly necessary, showing YOLO models have practical advantages and a promising future. Future researches will surround with the comprehensive cross-domain deployment of YOLO26 to promote its implementation in more real-time detection scenarios.

Keywords: Machine learning, deep learning, YOLO

1. Introduction

Computer Vision (CV), a key component of artificial intelligence, serves as the core perception entry that endows artificial intelligence with the ability to see. Object detection, a fundamental approach to realizing CV, aims to identify and local all objects appearing within a visual scene. This technology combines image classification with object localization by enclosing objects within bounding boxes and assigning corresponding labels [1]. Recently, object detection has achieved a remarkable progress. Within the various object detection methods, YOLO model outperforms through its excellent comprehensive performance, completing detection in a single pass, which has gained outstanding influence and has been widely utilized in various scenarios [1]. From 2015, when YOLOv1 was published for the first time, to 2026, when the newest version of YOLO26 is released, the YOLO family has undergone continuous evolution. Notably, YOLO26 achieves an

© 2026 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.GU33444

architectural breakthrough by completely eliminating the Non-maximum Suppression (NMS) post-processing step, improving inference speed on CPUs by up to 43%, severing as the first mainstream detection approach to realize true end-to-end inference [2].

With the advancement of the YOLO series, multidimensional research has emerged in this field. Some studies comprehensively review the developmental trajectory of YOLO models or focus on introducing specific versions offering evaluation and comparison across core architecture, innovations, performance, and efficiency. Other research examines the different applications of YOLO, including autonomous driving, smart agriculture, and medical image. These researches have proved the wide attention and application of YOLO model for the past eleven years.

However, due to the rapid iteration of YOLO versions and the recent release of YOLO26 in early 2026, research on the newest version remains in an initial stage, with a limited number of relevant studies currently available. Furthermore, existing research tends to focus either on the technical evolution of YOLO or on its application within a single field, lacking different version comparisons and guidance for model selection across different application scenarios. These limitations highlight the necessity of a comprehensive introduction to various YOLO versions, accompanied by cross-version and cross-scenario selection evaluations.

Therefore, the content of this study is as follows: 1) Section 2 outlines the mainstream YOLO models versions released on the official YOLO website, including YOLOv5, YOLOv8, YOLOv9, YOLO11, and YOLO26, with focusing on YOLO11 and YOLO26. 2) Section 3 compares YOLO11 and YOLO26 in terms of core architecture and performance. 3) Section 4 analyzes the application of YOLO models across various domains and provide cross-version and cross-scenario selection evaluations for practical deployment. 4) Section 5 makes a conclusion of the whole study.

2. Introduction of YOLO variants

2.1. The background of YOLO

Deep learning (DL), belonging to the field of machine learning (ML), emerged in the early 21st century and rapidly rose in 2006, by virtue of abundant datasets and high-performance computational resources [3]. In the domain of deep learning, convolutional neural networks (CNN) are one of the most broadly deployed deep network models in recent years, capable of automatically extracting key features from images and visual data without manual supervision [4]. Under the background where deep learning and computer vision boosted rapidly, diverse algorithms for object detection grounded in CNN have continued to evolve and innovate, with typical representatives including R-CNN, the YOLO series, and SSD. In accordance with implementation, the algorithms mentioned above can be classified into two categories based on their implementation: the using two-stage detectors and the one-stage detectors. Typical examples of two-stage detection mechanisms include region-based convolutional neural networks (R-CNN), Fast R-CNN, and Faster R-CNN, whose working principle is to generate candidate region proposals firstly and classify finely [5]. Although two-stage detection algorithms offer high accuracy by focusing only on regions with high confidence scores, they suffer from complex architectures and long inference times [1,3]. In contrast, one-stage detectors without the region proposal stage, features simpler architectures and extremely high inference speeds, with typical models including the YOLO series and SSD [1].

DOI: 10.54254/2755-2721/2026.GU33444

2.2. YOLOv5

In 2020, Glenn Jocher who was the CEO of Ultralytics and his team released YOLOv5 [6] on GitHub two months after the release of YOLOv4. Compared with YOLOv4, YOLOv5 reduced network parameters and speed up inference while improving the accuracy [7]. Its innovation lies in the first-time application of PyTorch as the underlying framework, which is different from all previous versions built on the Darknet framework [1,7]. YOLOv5 retains multiple architectural components of YOLOv4 and uses CSPDarknet-53 as the backbone network. It enhances gradient flow and reduces computational redundancy through combining Darknet-53 and CSP [1]. YOLOv5 employs PANet as the neck structure, and transmits the processed feature information to the prediction module by building a feature pyramid and integrating internal levels of visual features [6]. The head of YOLOv5 is used to receive the output of the neck and generate predictions of object categories and bounding boxes [6]. The most significant innovation of YOLOv5 is that it not only uses the PyTorch framework for development for the first time instead of utilizing the Darknet framework but also develops YOLOv5 in python rather than in C language. In addition, YOLOv5 also refined the classic three-stage architecture consisting of backbone network (Backbone), neck network (Neck) and head network (Head), which was initially established in YOLOv4. The majority of subsequent mainstream versions, such as YOLOv6, YOLOv7, YOLOv8 and YOLO11, optimized and improved in some aspects, such as the module design, detection head structure and training strategy, based on this overall architecture paradigm. Therefore, YOLOv5 is an important milestone in the development of the YOLO series, and establishes the foundation for the basic architecture of the modern YOLO model.

2.3. YOLOv8

In January 2023,YOLOv8 was released on Ultralytics, which had achieved a top performance in accuracy and speed and especially improved in the tiny object detection [2,8]. The researchers devised the C2f module with some improvements in the CSPLayer belong to CSPDarknet-53 and employed C2f module as the backbone network of YOLOv8 [9]. Meanwhile, YOLOv8 allowed EfficientDet being replacement of the backbone network of YOLOv8 [1]. Similarly, based on inheriting PANet adopted by YOLOv5, YOLOv8 combines with SPPF accomplishing multi-scale feature aggregation [1]. In addition, YOLOv8 has innovations in the head structure, where adopted a decoupled head, and detection accuracy and efficiency are both improved [9,2]. YOLOv8 is another major iteration based on the success of YOLOv5. Compared with another YOLO models, it shows more balanced performance.

2.4. YOLO11

The earliest YOLO11 was released on Ultralytics in September 2024 with outstanding accuracy, speed and efficiency [2]. In the backbone network, YOLO11 adopts C3k2 Block superseding C2f module applied in YOLOv8. C3k2 block has achieved the balance between performance and speed through utilizing two small convolutions [10]. Additionally, on the foundation of maintaining the Spatial Pyramid Pooling - Fast (SPPF) block, the backbone network of YOLO11 incorporates a new Cross Stage Partial with Spatial Attention (C2PSA) block, which improves spatial focusing ability [1]. YOLO11 continues to employ the same neck network with previous YOLO model in response to produce the final result of object detection. In summary, although YOLO11 is able to achieve high

DOI: 10.54254/2755-2721/2026.GU33444

accuracy with a minimal number of parameters, it has a worse stability than YOLOv8. The users enable to choose YOLO11 or YOLOv8 according to different needs.

2.5. YOLO26

YOLO26 was introduced firstly in September 2025 and was officially open-sourced by Ultralytics in January 2026. YOLO26 is the latest YOLO model, which eliminates Non Maximum Suppression (NMS) realizing end-to-end model and builds for edge and low-power devices [2]. The architecture design of YOLO26 is extremely similar only with fewer optimization in the Backbone and Neck, and the majority of improvements focus on the neck network and training paradigm [11]. YOLO26 discards the Distribution Focal Loss (DFL) block and employs box regression, which reduces the dependence of NMS realizing a higher accuracy with a little number of parameters [11]. From architecture, YOLO26 does not achieve major breakthrough, but there are new expansion functions which have greatly improved the practicality and scene adaptability of the model, making up for the conservative nature of the architecture. Moreover, the most of innovations of YOLO26 concentrate on training paradigm, which will be introduced in detail in Section 3.

3. Performance comparison and discussion

3.1. Core architecture of YOLO11 and YOLO26

The section only elaborates the differences between YOLO11 and YOLO26:

YOLO11 depends on Non-Maximum Suppression (NMS) while YOLO26 eliminates the usage of NMS. Inspired by the approach used in YOLOv10, YOLO26 employs a native one-to-one prediction head, realizing end-to-end NMS-Free inference and resolving the problem of latency [12].

YOLO11 adopts the DFL mechanism in its detection head, whereas YOLO26 discards DFL and implements box regression by explicitly predicting coordinates [11]. In comparison with YOLO11, YOLO26 is unnecessary to rely on complex softmax operations, which is more suitable to be adopted to low energy consumption edge devices. Additionally, YOLO26 only returns a single value no longer depending on distribution-based output [11]. Therefore, YOLO26 speeds up to 43% faster CPU inference and enables to output higher-confidence detection results using fewer parameters, which constitutes a significant technological breakthrough.

Building upon the SGD optimizer used in YOLO11, YOLO26 adopts a new optimizer called MuSGD, which incorporates Muon into the optimization process. MuSGD delivers more stable convergence during training and is particularly practical for small models [12].

In the last C3k2 block before the detect block, compared with YOLO11, YOLO26 carries out some optimizations through altering the repeat parameter from n to 1, improving the accuracy of YOLO26 [11].

While YOLO11 relies on a combination of static loss weights and Task-Aligned Learning (TAL), YOLO26 advances this by adopting progressive loss balancing (ProgLoss) alongside Small-Target-Aware Label Assignment (STAL). In comparison, YOLO26 achieves a significant progression in small object detection, while ensuring faster convergence and lower training memory usage [11].

DOI: 10.54254/2755-2721/2026.GU33444

3.2. Performance comparison of YOLO11 and YOLO26

Table 1. Performance comparison of YOLO26 and YOLO11 on the COCO dataset

ModelVariant	Resolution (pixels)	mAP@0.5:0.95 (val)	End-to-End mAP@0.5:0.95 (val)	CPU Inference Latency (ONNX, ms)	SpeedT4 TensorRT10 (ms)	params (Million)	Computational Complexity (FLOPs, Billion)
YOLO26n	640	40.9	40.1	38.9±0.7	1.7±0.0	2.4	5.4
YOLO26s	640	48.6	47.8	87.2±0.9	2.5±0.0	9.5	20.7
YOLO26m	640	53.1	52.5	220.0±1.4	4.7±0.1	20.4	68.2
YOLO26l	640	55.0	54.4	286.2 ± 2.0	6.2±0.2	24.8	86.4
YOLO26x	640	57.5	56.9	525.8±4.0	11.8±0.2	55.7	193.9
YOLO11n	640	39.5	39.5	56.1±0.8	1.5±0.0	2.6	6.5
YOLO11s	640	47.0	47.0	90.0±1.2	2.5±0.0	9.4	21.5
YOLO11m	640	51.5	51.5	183.2±2.0	4.7±0.1	20.1	68.0
YOLO11l	640	53.4	53.4	238.6±1.4	6.2±0.1	25.3	86.9
YOLO11x	640	54.7	54.7	462.8±6.7	11.3±0.2	56.9	194.9

Both YOLO11 and YOLO26 released five models: nano (n), small (s), medium (m), large (l) and xlarge (x), which can meet different hardware conditions. This section focuses on comparing the YOLO models used to implement detection tasks.

In YOLO26, the COCO mAP(50–95) increases with the growth of model scales, from YOLO26n at 40.9% to YOLO26x at 57.7% shown in Table 1, which shows a outstanding performance. Meanwhile, across CPU (ONNX) and GPU (TensorRT), different YOLO26 models maintain predictable and low inference latency, which is respectively lower than transformer-based alternatives [13]. Additionally, the five model variants also present good performance in parameters and FLOPs, which reflects that YOLO26 has the foundation to be widely used, and also indicates the original intention of YOLO26 to be developed specifically for edge and low-power devices.

In YOLO11, the trend of the COCO mAP(50–95) is consistent with YOLO26, among which the COCO mAP can reach 54.7%. In particular, the CPU latency of YOLO11n is 56.1 ms, while YOLO11x rises to 120.8 ms. The disproportionate 2.15x growth stems from the cumulative cost of DFL and NMS operations, which scales with the number of detections and become more burdensome as the model capacity increases.

Therefore, under the premise of the same hardware conditions, the performance of YOLO26 is better than YOLO11 across all metrics, although the difference is marginal in some aspects, YOLOn26's CPU latency is reduced by 43% compared to YOLO11n in terms of CPU latency, which is a substantial breakthrough. In addition, in terms of accuracy and the number of parameters, YOLO26 achieved a higher accuracy with minimal parameters. For instance, YOLO26x has

DOI: 10.54254/2755-2721/2026.GU33444

marginally fewer parameters than YOLO11x, yet delivers a 2.8 percentage point improvement in accuracy. All of these major breakthroughs are mainly attributable to end-to-end NMS-free design.

In conclusion, YOLO26 consistently outperforms YOLO11 in accuracy and inference efficiency at all comparable scales. YOLO26 successfully created a more accurate and faster model dedicated to edge and low-power devices by innovating the detection head and training paradigm, making YOLO26 a significant step in the domain of object detection.

4. YOLO models applications

The YOLO model uses a single-stage detector to combine target classification and positioning into one step. Its simple architecture and fast inference speed are suitable for scenarios that require real-time performance [1]. The deployment of YOLO models expands as the models continue to develop, having a wide range of applications, particularly in the field of autonomous driving, remote sensing and smart agriculture. This section mainly focuses on the scenarios requiring real-time performance.

4.1. Autonomous driving

In the field of autonomous driving, YOLO models serve as a key technology improving the development of autonomous driving, which are able to detect the objectives of vehicles, pedestrians and traffic signs timely [14]. The deployments of YOLO models in autonomous driving improved, with the upgrade of different versions. The YOLO models such as YOLOv3, YOLOv5, YOLOv8 and some improved YOLO variants are deployed in autonomous driving [14]. At present, YOLO11 is the latest version applied in autonomous driving possessing various advantages. For instance, YOLO11 utilizes Cross-stage Partial Spatial Attention(C2PSA) as its attention mechanism significantly enhancing the accuracy of detecting smaller, occluded, or rapidly moving vehicles [15]. In addition, in vehicle detection, the accuracy and speed of YOLO11 have an evident strengths. Specially, its mAP@0.5 can reach 0.743 and inference speed can achieve 290 frames per second (FPS) [15]. Although YOLO11 is a robust solution in autonomous driving, the limitations include being difficult to handle rare objects, and being affected by highly variable conditions [15]. Meanwhile, there are searches mention a new model based on improved YOLO11 algorithm, whose mAP@0.5 can reach 0.853 [16]. Finally, the YOLO26 proposed in 2025 can also be considered for applying in autonomous driving to achieve higher detection accuracy.

4.2. Industrial detection

Traditional inspection methods and traditional object detection and location method highly depend on manual operation, which probably causes subjectivity, inefficiency, and inconsistency and also may unable to demonstrate good robustness and generalization capabilities [17,18]. Additionally, in the context of increasingly fierce global market competition, the automation level and real-time analysis capabilities have become key factors to maintain the core competitiveness of an enterprise or factory [17]. YOLO models, by virtue of the ability of detecting efficiently and timely, provide efficient and flexible solutions for industrial defect detection [19]. In the industrial defect detection, YOLO models are mainly applied to test product quality, identify and locate items, and detect the working environment [18]. With the continuous updating of YOLO model versions, different versions also present different performance in industrial detection scenarios. For instance, the defect detection accuracy of YOLOv10 and YOLO11 has been significantly increased and presents a better performance in dealing with complex scenario. This is due to the innovations in Transformer

DOI: 10.54254/2755-2721/2026.GU33444

framework, adaptive resolution scaling and mechanisms of self-attention computation [17]. However, it is still challenging to deploy YOLO models in the industrial defect detection. Regarding fabric defect inspection, when facing various types of defects, complex fabric textures, and changing illumination conditions, it is desired to enhance the detection performance of the model [17].

4.3. Security system

In the context when public safety is crucial, the core objective pursued by the intelligent security system is equipped with the capability to accurately and rapidly detect and evaluate threats [20]. With high inference speed and the ability to detect in real-time sense, YOLO models are able to achieve automated real-time monitoring of public places, which effectively prevents security incidents and respond to emergencies [21]. In the field of intelligent security system, real-time surveillance is one of the most vital application areas of YOLO models and the primary scenario refers to weapon detection, crowd analysis and the detection of specific criminal activities [1]. The latest YOLO version deployed in the security system possesses better detection accuracy and performance facing complex scenarios such as occlusion, low lighting, and crowded scenes [21]. However ,future efforts in YOLO models applied in security system should include the combination of YOLO and advanced temporal models, continue to enhance robustness under challenging conditions and improving generalization to reduce reliance on labeled datasets [21].

5. Conclusion

In conclusion, this research mainly studies the YOLO model in the field of deep learning, concentrating on the evolution of mainstream YOLO models and applications. With the iteration of YOLO technology, the performance of YOLO models continues to enhance through the architectural improvement. YOLO26, as the latest variant, while the accuracy comprehensively increases, it achieves end-to-end NMS-free inference by improving CPU inference speed and realizes the model more lighter, which is more suitable for edge and low-power devices. However, because YOLO26 has been around for a short time, it has few cross-field applications. Future research will further expand the cross-domain application of the YOLO26 model in typical scenarios such as Unmanned Aerial Vehicle(UAV) inspections, remote sensing, and intelligent surveillance to give full play to its advantages of efficient, lightweight, and end-to-end reasoning.

References

[1] L. T. Ramos and A. D. Sappa, "A Decade of You Only Look Once (YOLO) for Object Detection: A Review, " in IEEE Access, vol. 13, pp. 192747-192794, 2025.
[2] Ultralytics. (2026). Ultralytics YOLO docs. https: //docs.ultralytics.com/
[3] Diwan, T., Anirudh, G., & Tembhurne, J. V. (2023). Object detection using YOLO: challenges, architectural successors, datasets and applications. multimedia Tools and Applications, 82(6), 9243-9275.
[4] Alzubaidi, L., Zhang, J., Humaidi, A. J., Al-Dujaili, A., Duan, Y., Al-Shamma, O., ... & Farhan, L. (2021). Review of deep learning: concepts, CNN architectures, challenges, applications, future directions. Journal of big Data, 8(1), 53.
[5] Xie, X., Cheng, G., Wang, J., Yao, X., & Han, J. (2021). Oriented R-CNN for object detection. In Proceedings of the IEEE/CVF international conference on computer vision (pp. 3520-3529).
[6] Sirisha, U., Praveen, S. P., Srinivasu, P. N., Barsocchi, P., & Bhoi, A. K. (2023). Statistical analysis of design aspects of various YOLO-based deep learning models for object detection. International Journal of Computational Intelligence Systems, 16(1), 126.
[7] Olorunshola, O. E., Irhebhude, M. E., & Evwiekpaefe, A. E. (2023). A comparative study of YOLOv5 and YOLOv7 object detection algorithms. Journal of Computing and Social Informatics, 2(1), 1-12.

DOI: 10.54254/2755-2721/2026.GU33444

[8] Varghese, R., & Sambath, M. (2024). Yolov8: A novel object detection algorithm with enhanced performance and robustness. In 2024 International conference on advances in data engineering and intelligent computing systems (ADICS) (pp. 1-6). IEEE.
[9] Terven, J., Córdova-Esparza, D.-M., & Romero-González, J.-A. (2023). A Comprehensive Review of YOLO Architectures in Computer Vision: From YOLOv1 to YOLOv8 and YOLO-NAS. Machine Learning and Knowledge Extraction, 5(4), 1680-1716.
[10] Khanam, R., & Hussain, M. (2024). Yolov11: An overview of the key architectural enhancements. arXiv preprint arXiv: 2410.17725.
[11] Hidayatullah, P., & Tubagus, R. (2026). YOLO26: A Comprehensive Architecture Overview and Key Improvements. arXiv preprint arXiv: 2602.14582.
[12] Satya Mallick (2024). Yolo - learnopencv. https: //learnopencv.com/yolo11/, 2024. Accessed: 2024-10-21.
[13] Sapkota, R., Cheppally, R. H., Sharda, A., & Karkee, M. (2025). YOLO26: key architectural enhancements and performance benchmarking for real-time object detection. arXiv preprint arXiv: 2509.25164.
[14] Wei, J., As'arry, A., Rezali, K. A. M., Yusoff, M. Z. M., Ma, H., & Zhang, K. (2025). A review of yolo algorithm and its applications in autonomous driving object detection. IEEE Access.
[15] Alif, M. A. R. (2024). Yolov11 for vehicle detection: Advancements, performance, and applications in intelligent transportation systems. arXiv preprint arXiv: 2410.22898.
[16] Liang, E., Wei, D., Li, F., Lv, H., & Li, S. (2025). Object detection model of vehicle-road cooperative autonomous driving based on improved YOLO11 algorithm. Scientific Reports, 15(1), 32348.
[17] Mao, M., & Hong, M. (2025). YOLO object detection for real-time fabric defect inspection in the textile industry: A review of YOLOv1 to YOLOv11. Sensors, 25(7), 2270.
[18] Kang, S., Hu, Z., Liu, L., Zhang, K., & Cao, Z. (2025). Object detection YOLO algorithms and their industrial applications: Overview and comparative analysis. Electronics, 14(6), 1104.
[19] Hussain, M. (2023). YOLO-v1 to YOLO-v8, the rise of YOLO and its complementary nature toward digital manufacturing and industrial defect detection. Machines, 11(7), 677.
[20] Sanjalawe, Y., & Alqudah, H. (2024). Integrating enhanced security protocols with moving object detection: A YOLO-based approach for real-time surveillance. In 2024 2nd International Conference on Cyber Resilience (ICCR) (pp. 1-6). IEEE.
[21] Widodo, Y. B., Sibuea, S., & Agustino, R. (2026). YOLO in Suspicious Human Activity Recognition for Intelligent Environmental Security Systems: A Review. Jurnal Teknologi Informatika dan Komputer, 12(1), 101-119.