PAHTEI
REFEREED & REVIEWED JOURNAL

ISSN: 1609-1620

VOLUME 27 ISSUE 06 2025

Pp. 142-147 DOI: https://doi.org/10.32010/IXNR6858

COMPARATIVE ANALYSIS OF REAL-TIME COMPUTER VISION SYSTEMS

Elviz Ismayilov[1], Hasanrza Hasanli[2]
[1,2] Azerbaijan State Oil and Industry University
[1] Department of General and Applied Mathematics
[2] Department of Computer Engineering
[1] Docent, Director of Digital Development and Innovations Center, elviz.ismayilov@gmail.com
[2] Master student, hasanrza.888@gmail.com
ORCID: [1]0000-0002-3152-059X; [2]0009-0000-8462-9866

ABSTRACT

Real-time object detection poses a significant problem in modern computer vision, especially in areas that include autonomous driving, intelligent surveillance systems, robotics, and innovative manufacturing processes. This work provides a comparative study of two highly-used and high-performing models in the object detection field: the newly released You Only Look Once version 12 (YOLOv12), known for its high speed in processing and ease of use, and Faster R-CNN that has a ResNet-50-FPN backbone and is known as a two-stage detection model that is highly valued for its accuracy and performance in the process of feature extraction. The main objective of this work is the evaluation of both models in terms of performance in real-time applications with considerations of variables including inference speed, computational complexity, the number of parameters, and the general efficiency of each model.

To enable the evaluation, both models were tested under the same experimental conditions using a test image benchmark and run in a GPU-based Google Colab environment. The models were compared in terms of average inference time in seconds, frames per second (FPS), total parameter count in millions, and floating point operations (FLOPs). The outcome of the experiment showed that despite the architectural optimizations of the YOLOv12 for enhanced real-time performance, the Faster R-CNN model surprisingly out-performed in terms of FPS and showed a lower inference time in the given configuration. However, in contrast, the YOLOv12 showed considerably higher model complexity that may make it more suited for generalized performance in more complex or diverse deployment situations.

The results indicate that Faster R-CNN suits applications that value research and high accuracy and tolerate a slight increase in inference times. In contrast, YOLOv12 exhibits better versatility for edge computing platforms and real-time processing use cases due to its modular architecture, low weight, and hardware acceleration compatibility. This side-by-side analysis reveals valuable information about the trade-offs between precision, speed, and computation requirements and aids in more effective decision-making for real-time computer vision application model selection.

Keywords: Real-time object detection, YOLOv12, Faster R-CNN, ResNet-50-FPN, inference speed, frames per second (FPS), model complexity, computational cost, deep learning, edge computing, two-stage detector, single-stage detector, computer vision, performance comparison.

Introduction

Computer vision has evolved from research in the academic level to practical, applied usage over the past few years at a tremendous rate. Real-time object detection, or the ability of a machine to recognize and locate objects from images or video streams with minimal delay, is one of the most valuable elements of vision systems nowadays. Real-time object detection is now a key condition for a wide range of applications, including autonomous vehicles, security and surveillance, industrial automation, and robotics.

Since real-time operation is needed, the issue of using the correct detection model arises. There are models optimized for performance with capabilities for handling a huge volume of visual data within a limited time and those optimized for accuracy and stability of detection. The choice is based on specific application needs such as tolerance for delay, constraints on hardware as well as required accuracy [7].

Two of the many options present at the time of writing have distinguished themselves due to their performance and popularity: YOLOv12 and Faster R-CNN using a ResNet-50-FPN backbone. YOLOv12 is the latest from the highly reputed YOLO series of models that has always had the standing of being efficient and fast [1, 2, 8]. Faster R-CNN is an embodiment of a different ideology - a two-stage detection network that is generally more accurate at the cost of increased computational complexity [3, 4, 5]. It places those two models under the microscope and pits them against each other under live conditions. Rather than examining accuracy on its own, the comparison considers a spectrum of factors influencing live useability, from inference time to frames per second, to number of parameters and general complexity of a model. Side by side and split out, the goal is to provide constructive, actionable insight that is able to inform on what object detection models to use when deploying for live purposes.

The core aim of the current study is to compare the relative performance of two widely applied object detection models in use nowadays—YOLOv12 and Faster R-CNN based on a ResNet-50-FPN backbone—against real-life usage contexts. These models represent different design principles: YOLOv12, being the latest in the YOLO family of models, has been trained for speed and efficiency in design and is best applied in use scenarios that require prompt response times [1, 2, 10]. On the contrary, Faster R-CNN has superior detection capabilities and richer feature extraction properties but is normally more resource-demanding [3, 4].

Contrasting from the conventional approach of rating models based on accuracy in several benchmarks prevalent in other studies, this research has a more inclusive approach that also considers practical performance in real scenarios. The models are subjected to same testing conditions and are compared in regards to inference time, frames per second (FPS), number of parameters, and floating-point operations (FLOPs). These particular factors have been chosen due to representation of trade-offs for developers and engineers in deciding suitability of models for deployment into production environments, especially on edge hardware or resource-restricted platforms [6, 9].

The aim is not one of expressing absolute superiority of one model over all others but one of making clear in which contexts each model achieves outstanding performance and of each model's inherent limitations. By exploring both results and analytical insights, this study attempts to enable wiser and better-informed choice for professionals operating within computer vision in academic research, industrial use, or in applied fields like autonomous guidance and perceptual surveillance.

Methods

To compare the effectiveness of YOLOv12 against Faster R-CNN based on ResNet-50-FPN models in real-time scenarios, a systematic experimentation framework was applied for both models. The evaluation of performance included a number of key metrics: average inference time in seconds, frames per second (FPS), total model parameters in units of millions of parameters, and computational complexity described in floating-point operations (FLOPs). These particular metrics were selected since they capture most of the key factors related to the deployment of object detection models in time-constrained practical scenarios, for instance, in autonomous systems and edge-computing hardware [6, 7, 9].

Models were created using open-source implementations. The YOLOv12 implementation was based on its official framework and documentation provided by Ultralytics [8, 10]. YOLOv12 architecture features an improvement over its predecessors and includes an efficient backbone network, improved attention mechanisms, and hardware deployment-optimal optimizations for deployment on resource-restricted hardware [1, 2]. The model was run in a Google Colab environment using the Ultralytics Python API under GPU acceleration.

The Torchvision model zoo's standard ResNet-50-FPN backbone was used to construct faster R-CNN [9]. The detection mechanism of this model is two-staged: a Region Proposal Network (RPN) is used to generate region suggestions first, and a second step is used to categorize and refine these ideas [3, 4]. Multi-scale feature extraction, which is particularly useful for detecting objects of different sizes, is made possible by the Feature Pyramid Network (FPN), which improves performance [4].

The same dataset – COCO128, a condensed subset of the COCO dataset intended for rapid prototyping – was used to run both models. This maintained appropriate computational needs while enabling a fair and uniform comparison between models. To remove variance brought forth by variations in hardware or data, each model processed the same input image under the same circumstances.

FPS was calculated using reciprocal inference time, and all timing values were averaged over several runs to minimize noise. Model introspection tools from PyTorch and Ultralytics were used to get parameter counts and FLOPs.

The models' capabilities in real-time scenarios are directly and fairly compared thanks to this experimental design, which also sheds light on the models' advantages and disadvantages in various deployment scenarios.

Conclusion

One goal of the current research was to determine the effectiveness of YOLOv12 compared to Faster R-CNN using ResNet-50-FPN, two of the leading frameworks in real-time object detection. Even though both frameworks are widely applied in the field of computer vision, they solve the issue through essentially different approaches. YOLOv12 is a single-stage detector and proves highly efficient in scenarios in which speed is given top priority. On the contrary, Faster R-CNN's two-stage approach tends to deliver higher accuracy but comes with a higher computational cost.

To understand the implications in practice of these differences, both models were tested in similar conditions: a GPU-accelerated environment using Google Colab and an optimized benchmark dataset (COCO128). A systematic comparison was done in order to quantify and compare some key metrics – mean inference time, frames per second (FPS), total parameters, and floating point operations per second (FLOPs) – In an overall evaluation of performance shown in Table 1.

Table 1. Performance Comparison Between YOLOv12 and Faster R-CNN

Model	Avg Inference Time (s)	FPS	Parameters (M)	FLOPs (G)
YOLOv12	0.276	3.62	85	175
Faster R-CNN (ResNet-50-FPN)	0.136	7.37	42	120

This difference is also clearly illustrated in Figure 1, which highlights the FPS values of both models. The results shown do not conclusively indicate Faster R-CNN as being the best choice in all scenarios. The performance may have been due either to the simplicity of the test data or certain optimizations built into the base model implementation. On the other hand, YOLOv12's highly increased parameter count and increased number of FLOPs suggest that its network has a larger capacity in processing more challenging scenarios, particularly when run on highly capable hardware platforms including edge devices or accelerators, e.g., TensorRT [1, 2, 10].

In most scenarios, Faster R-CNN seems specifically beneficial in scenarios in which accuracy is more crucial than speed of processing, for instance in medical images or document processing, especially in research institutions or server-based platforms [3, 4, 9]. YOLOv12 is best suited for edge computing, robotics, or mobile platforms in which speed of inference and deployment is a top requirement [8, 10].

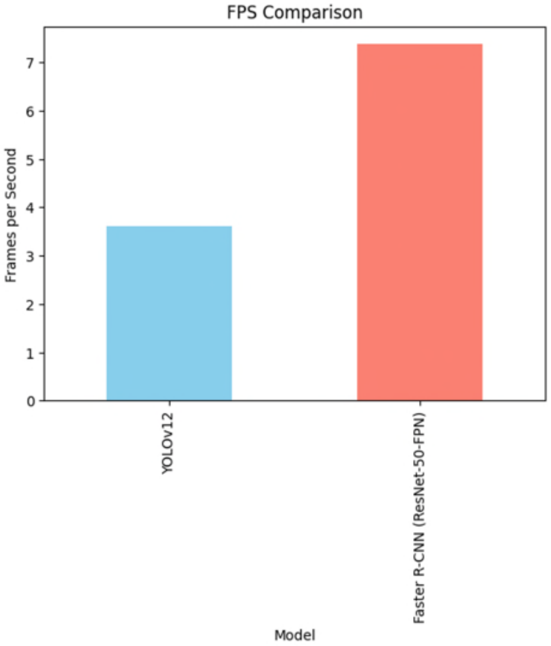

Figure 1. FPS Comparison Between YOLOv12 and Faster R-CNN

Overall, a one-size-fits-all solution for all circumstances does not really exist. All models have unique strengths and understanding these trade-offs is crucial for making informed choices. Our hope is that comparative and empirical information provided throughout this report will encourage more intentional model selection in real-time vision projects.

REFERENCES

1. Tian, Y., Ye, Q., & Doermann, D. (2025). YOLOv12: Attention-Centric Real-Time Object Detectors. *arXiv preprint arXiv:2502.12524*. https://arxiv.org/abs/2502.12524
2. Alif, M. A. R., & Hussain, M. (2025). YOLOv12: A Breakdown of the Key Architectural Features. *arXiv preprint arXiv:2502.14740*. https://arxiv.org/abs/2502.14740arXiv
3. Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *Advances in Neural Information Processing Systems*, 28. https://arxiv.org/abs/1506.01497
4. Lin, T.-Y., Dollár, P., Girshick, R., He, K., Hariharan, B., & Belongie, S. (2016). Feature Pyramid Networks for Object Detection. *arXiv preprint arXiv:1612.03144*. https://arxiv.org/abs/1612.03144arXiv
5. Girshick, R. (2015). Fast R-CNN. *arXiv preprint arXiv:1504.08083*. https://arxiv.org/abs/1504.08083arXiv+2arXiv+2MDPI+2
6. Tahir, H., Khan, M. S., & Tariq, M. O. (2021). Performance Analysis and Comparison of Faster R-CNN, Mask R-CNN and ResNet50 for the Detection and Counting of Vehicles. *2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*. https://www.researchgate.net/publication/350927615ResearchGate
7. Jonathan Hui. (2018). Object Detection: Speed and Accuracy Comparison (Faster R-CNN, R-FCN, SSD, FPN, RetinaNet and YOLOv3). *Medium*. https://jonathan-hui.medium.com/object-

detection-speed-and-accuracy-comparison-faster-r-cnn-r-fcn-ssd-and-yolo-5425656ae359Medium
8. Ultralytics. YOLOv12 Documentation https://docs.ultralytics.com/models/yolo12
9. PyTorch. fasterrcnn_resnet50_fpn — Torchvision Main Documentation. https://pytorch.org/vision/main/models/generated/torchvision.models.detection.fasterrcnn_resnet50_fpn.htmlMedium+1en.wikipedia.org+1
10. Roboflow. YOLOv12: State-of-the-Art Object Detection Model. https://yolov12.com

REAL VAXT REJIMINDƏ KOMPÜTER GÖRMƏ SISTEMLƏRININ MÜQAYISƏLI TƏHLILI

Elviz İsmayılov[1], Həsənrza Həsənli[2]
[1,2]Azərbaycan Dövlət Neft və Sənaye Universiteti
[1]"Ümumi və Tətbiqi Riyaziyyat" kafedrası
[2]"Komputer Mühəndisliyi" kafedrası
[1]Dosent, Rəqəmsal İnkişaf və İnnovasiyalar Mərkəzinin direktoru, elviz.ismayilov@gmail.com
[2]Magistr tələbəsi, hasanrza.888@gmail.com

XÜLASƏ

Real vaxtda obyektin aşkarlanması müasir kompüter görmə sahəsində, xüsusilə avtonom nəqliyyat, ağıllı müşahidə sistemləri, robototexnika və innovativ istehsal prosesləri kimi sahələrdə ciddi problemlərdən biri kimi çıxış edir. Bu işdə obyekt aşkarlanması sahəsində geniş istifadə olunan və yüksək nəticələr göstərən iki modelin müqayisəli təhlili aparılır: yüksək emal sürəti və istifadədə sadəliyi ilə tanınan You Only Look Once (YOLO) modelinin son versiyası olan YOLOv12 və xüsusilə dəqiqlik və təsvir xüsusiyyətlərinin çıxarılmasında yüksək performansı ilə seçilən ResNet-50-FPN əsaslı iki mərhələli aşkarlama modeli olan Faster R-CNN. Bu işin əsas məqsədi hər iki modelin real vaxt tətbiqlərindəki performansının qiymətləndirilməsidir və burada inferensiya sürəti, hesablama mürəkkəbliyi, parametr sayı və ümumi effektivlik kimi dəyişənlər nəzərə alınır. Qiymətləndirməni təmin etmək üçün hər iki model eyni eksperimental şərtlər altında test şəkilləri ilə sınaqdan keçirilmiş və GPU əsaslı Google Colab mühitində işə salınmışdır. Modellər orta inferensiya vaxtı (saniyə ilə), saniyədə kadr sayı (FPS), milyonlarla parametr sayı və üzən nöqtəli əməliyyatlar (FLOPs) baxımından müqayisə edilmişdir. Eksperimentin nəticələri göstərdi ki, YOLOv12-nin real vaxt performansını artırmaq üçün edilmiş arxitektur optimallaşdırmalara baxmayaraq, Faster R-CNN FPS baxımından daha yaxşı nəticə göstərmiş və verilmiş konfiqurasiyada daha aşağı inferensiya vaxtı təqdim etmişdir. Digər tərəfdən, YOLOv12-nin model mürəkkəbliyi xeyli yüksək olmuş və bu onu daha mürəkkəb və müxtəlif tətbiq mühitlərində ümumiləşdirilmiş performans üçün daha uyğun etmişdir. Nəticələr göstərir ki, Faster R-CNN tədqiqat yönümlü və yüksək dəqiqlik tələb olunan, lakin azacıq yüksək inferensiya vaxtına dözümlü tətbiqlər üçün uyğundur. Əvəzində, YOLOv12 özünün modulyar arxitekturası, yüngül çəkili dizaynı və hardware sürətləndirici uyğunluğu ilə kənar hesablama platformaları və real vaxtda emal tətbiqləri üçün daha əlverişlidir. Bu yanaşı müqayisə dəqiqlik, sürət və hesablama tələbləri arasında mövcud balansı aydın şəkildə göstərir və real vaxt kompüter görmə tətbiqləri üçün model seçimini daha effektiv şəkildə aparmağa imkan verir.

Açar sözlər: real vaxtda obyektin aşkarlanması, YOLOv12, Faster R-CNN, ResNet-50-FPN, inferensiya sürəti, saniyədə kadr sayı (FPS), model mürəkkəbliyi, hesablama xərci, dərin öyrənmə, kənar hesablama, iki mərhələli detektor, bir mərhələli detektor, kompüter görməsi, performans müqayisəsi.

СРАВНИТЕЛЬНЫЙ АНАЛИЗ СИСТЕМ КОМПЬЮТЕРНОГО ЗРЕНИЯ В РЕАЛЬНОМ ВРЕМЕНИ

Эльвиз Исмаилов[1]**, Гасанрза Гасанли**[2]

[1,2] Азербайджанский Государственный Университет Нефти и Промышленности
[1,2] кафедра «Общая и прикладная математика»
[1] Доцент, директор Центра цифровых разработок и инноваций elviz.ismayilov@gmail.com
[2] Магистрант, hasanrza.888@gmail.com

РЕЗЮМЕ

Обнаружение объектов в реальном времени является одной из серьезных проблем современного компьютерного зрения, особенно в таких областях, как беспилотный транспорт, интеллектуальные системы наблюдения, робототехника и инновационные производственные процессы. В этом исследовании сравниваются две широко используемые и весьма успешные модели обнаружения объектов: YOLOv12, последняя версия модели You Only Look Once (YOLO), известная своей высокой скоростью обработки и простотой использования, и Faster R-CNN, двухэтапная модель обнаружения на основе ResNet-50-FPN, которая особенно примечательна своей точностью и высокой производительностью при извлечении характеристик изображения. Основная цель данной работы — оценить производительность обеих моделей в приложениях реального времени, где учитываются такие переменные, как скорость вывода, вычислительная сложность, количество параметров и общая эффективность. Для проверки обе модели были протестированы с тестовыми изображениями в одинаковых экспериментальных условиях и запущены в среде Google Colab на базе графического процессора. Результаты показывают, что Faster R-CNN подходит для научно-исследовательских и высокоточных приложений, которые требуют, но допускают несколько большее время вывода. Вместо этого YOLOv12 больше подходит для платформ периферийных вычислений и приложений обработки в реальном времени благодаря своей модульной архитектуре, легкой конструкции и совместимости с аппаратными ускорителями. Это наглядное сравнение наглядно иллюстрирует существующий баланс между точностью, скоростью и вычислительными требованиями, что позволяет более эффективно выбирать модели для приложений компьютерного зрения в реальном времени.

Ключевые слова: Обнаружение объектов в реальном времени, YOLOv12, Faster R-CNN, ResNet-50-FPN, скорость инференса, количество кадров в секунду (FPS), сложность модели, вычислительные затраты, глубокое обучение, периферийные вычисления, двухэтапный детектор, одноэтапный детектор, компьютерное зрение, сравнение производительности.

Publishing history: 30.06.2025
Article received: 22.05.2025
Article accepted: 10.06.2025