ISSN: 3030-3346

METHOD OF USING YOLO MODELS IN IDENTIFYING OBJECTS IN IMAGES

Dauletov Adilbek Yusupbayevich

Alfraganus University,
E-mail: a.dauletov@afu.uz

K E Y W O R D S

YOLO, model, artificial intelligence, deep learning, object detection, analysis.

A B S T R A C T

This article analyzes the capabilities of YOLO (You Only Look Once) models in the field of image recognition and the achievements made with them. The YOLO algorithm stands out among object recognition technologies for its speed and accuracy. Each version of the YOLO models includes its own technological innovations, and each has increased accuracy and efficiency. This article highlights the changes in each version of the YOLO models and their technical capabilities.

Introduction

Object detection is one of the most difficult and at the same time most important tasks in the field of computer vision. This technology is widely used in security systems, autonomous vehicles, medicine and various industrial processes. The efficiency and speed of object detection is directly related to the development of computer vision algorithms, and YOLO models are one of the greatest achievements in this regard[1].

One of the main advantages of YOLO models is the fast and efficient implementation of the object detection process. The YOLO algorithm analyzes the entire image once, as a result of which it becomes possible to instantly detect objects without going through several layers. This is very important in real-time applications, such as security cameras or autonomous vehicles[2].

This article reviews the development of YOLO models, their technical capabilities, and a comparative analysis of different versions. It provides information on the differences between YOLO models, their strengths and weaknesses, and possible areas of application. It also discusses how YOLO models have achieved success in object detection and their impact on current technological solutions.

Literature Review

Today, there are numerous reasons why it is very important for contemporary students to study artificial intelligence (AI) technologies. For this reason, universities are actively incorporating AI-related subjects into their undergraduate and graduate curricula. It is crucial to teach students how these technologies work and how they can be applied to solve various problems [3]. In their article, A.D. Agafina and A.A. Nikitin present the results of testing the functionalities of a text and formula recognition system in Russian and English. They detail the advantages and disadvantages of the developed system, as well as the challenges faced during the development process [4].

Russian researcher Markeev Maksim Valeryevich, in his article "Методика идентификации объектов по уникальным характеристикам на изображениях" ("A Methodology for Identifying Objects by Their Unique Characteristics in Images"), increases the accuracy of the model's performance by observing the movements of whales and dolphins. He creates two different datasets: one for the entire body and another specifically for fins [5]. In the article "Разработка нейросетевого метода в задаче классификации и распознавании изображения" ("Development of a Neural Network Method for Image Classification and Recognition"), M.N.B. Mual, D.V. Kozyrev, G.Zh.K. Uankpo, and E. Nibasumba describe the FastER-RCNN method, which is based on a two-stage neural network. They also present the results of applying the YOLOv3 algorithm at various stages of neural network training [6]. K. Avezov and B. Mominov propose a method for detecting and reporting fires in ship areas using deep learning and computer vision methods [7]. Similarly, to aid in coastal monitoring, Shao and Wang built a Saliency-Aware CNN within the YOLOv2 framework to create a ship detection model. They used an initial version of the SeaShips image dataset [8].

In his article “Сложность распознавания при разработке программного обеспечения для видеомониторинга” (“The Complexity of Recognition in Developing Video Monitoring Software”), Candidate of Technical Sciences A.Yu. Kruchinin examines the issue of selecting the optimal operating mode for video monitoring when using neural network models for recognition. He demonstrates a solution using YOLOv5 models to detect objects separated into two categories, ensuring that results are produced after each frame in real time with minimal delay. Additionally, he analyzes the metrics used for object detection from the perspective of assessing result reliability when no final information about the object is available [9].

Main part

The YOLO family of models is one of the most popular network architectures used for object detection, with each new version introducing significant improvements and new capabilities. While YOLOv1 was known for its speed, it lacked accuracy. YOLOv2 and YOLOv3 improved accuracy in more complex scenes. YOLOv4, leveraging CSP blocks and PANet, enhanced both accuracy and efficiency. YOLOv5 became well-suited for applications requiring lightweight and rapid decision-making. YOLOv6 and YOLOv7 further increased accuracy by deepening the network architecture. The latest version, YOLOv8, enables highly accurate object detection even in complex conditions. The first table below presents an analysis of the YOLO model versions.

Table 1.

Analysis of YOLO model versions

YOLO Version	Year of release	Capabilities	Disadvantages
YOLOv1	2015	One-step detection (fast performance, suitable for real-time use). Using a grid approach, divide the image into grids and detect objects in each slice.	There are problems with detecting small objects, and the fact that only one object can be detected per grid due to grid constraints reduces accuracy.
YOLOv2 (YOLO9000)	2016	Batch normalization to improve accuracy. Wide class recognition (9000+ feature classes). Better detection of small and close-up features using anchor boxes.	It still doesn't have high accuracy, especially on complex images. It requires powerful computing resources to train the model.
YOLOv3	2018	Multi-scale detection (detection of large and small objects at three different scales). Better features with Darknet-53 network. Good balance of accuracy and speed.	More complex and heavier, requiring more computer resources.

YOLOv4	2020	Easy to train and deploy in the PyTorch framework. Different model versions (nano, small, medium, large, xlarge). Fast and efficient training.	There was no scientific paper available when the model was released, which led to criticism in academic circles.
YOLOv5	2020	Easy to train and deploy in the PyTorch framework. Different model versions (nano, small, medium, large, xlarge). Fast and efficient training.	There was no scientific paper available when the model was released, which led to criticism in academic circles.
YOLOv6	2022	New techniques have been used to increase efficiency and accuracy. Accuracy and speed of operation have been improved.	The demand for computer resources has increased.
YOLOv7	2022	Greater detection range for large and small object detection. Detect small and multiple objects with high accuracy.	Requires large computing resources.
YOLOv8	2023	Real-time high-precision detection capability. New structures and optimization methods have been applied.	Requires large computing resources.

YOLO models are used in many areas. Their main areas of application are as follows:

- **Video surveillance and security:** YOLO models are used in surveillance cameras to detect and track people, cars, and other objects. These systems aid in identifying security threats and hazardous situations.
- **Autonomous vehicles:** YOLO is employed to detect objects for cars and other vehicles. For example, it plays an important role in recognizing road signs, pedestrians, and other vehicles.
- **Drones:** YOLO models are applied to help drones detect and track objects. This can be beneficial in agriculture, environmental monitoring, rescue operations, and military fields.
- **Medicine:** YOLO models are used to analyze medical images (e.g., X-rays or ultrasounds) to detect issues like tumors and signs of disease.
- **Agriculture:** By using YOLO to analyze images, it is possible to monitor plant conditions, identify pests, and assess crop yields.
- **Retail and marketing:** YOLO can be utilized to track products in stores, manage warehouse inventory, or analyze customer behavior within the store.
- **Industry and manufacturing:** YOLO models are used for product inspection on production lines, detecting defects, and maintaining quality control.
- **Sports analysis:** Sports analysts use YOLO models to track and analyze players, balls, and other elements during games.

Due to the speed and accuracy of YOLO models, they are widely employed in situations that require real-time monitoring [10].

In this article, we analyze two versions of YOLO models. The first one, YOLOv5, with its high processing speed and ease of learning, is considered ideal for situations requiring rapid decision-making in real time. In terms of speed, it stands out by analyzing objects in a single pass, making it effective for real-time surveillance and video stream analysis. YOLOv5 is written in PyTorch and is well-documented, which allows beginners to quickly master and integrate it [11]. It is well-suited for real-time analysis in fields such as security, autonomous transportation, and robotics. Additionally, it easily integrates with technologies like Python, OpenCV, and TensorRT, enabling optimized inference speeds. For these reasons, YOLOv5 is widely used in scenarios demanding both high speed and accuracy, as shown in Figure 1.

Figure 1. Object Detection with the YOLOv5 Model

This image visually demonstrates how the YOLOv5 object detection model operates. In the image, objects such as people standing next to a car, a backpack, and a suitcase are highlighted with bounding boxes by the model. However, some detections are incorrect or uncertain. For instance, the object near the car's trunk is misclassified as a "train" and has a relatively low confidence score (0.26). This example illustrates both correct and erroneous detections by the model, as well as the concept of confidence levels in object detection.

Table 2.

Accuracy Levels of YOLOv5's Object Detection in the Image

Object	Model	Accuracy
Person 1	YOLOv5	0.51
Person 2	YOLOv5	0.63
Person 3	YOLOv5	0.92
Person 4	YOLOv5	0.90
Backpack	YOLOv5	0.71
Suitcase	YOLOv5	0.59

We will now try to detect the image tested with the YOLOv5 model with the YOLOv8 model. Figure 2 below shows how the YOLOv8 model detects objects in the image.

ISSN: 3030-3346

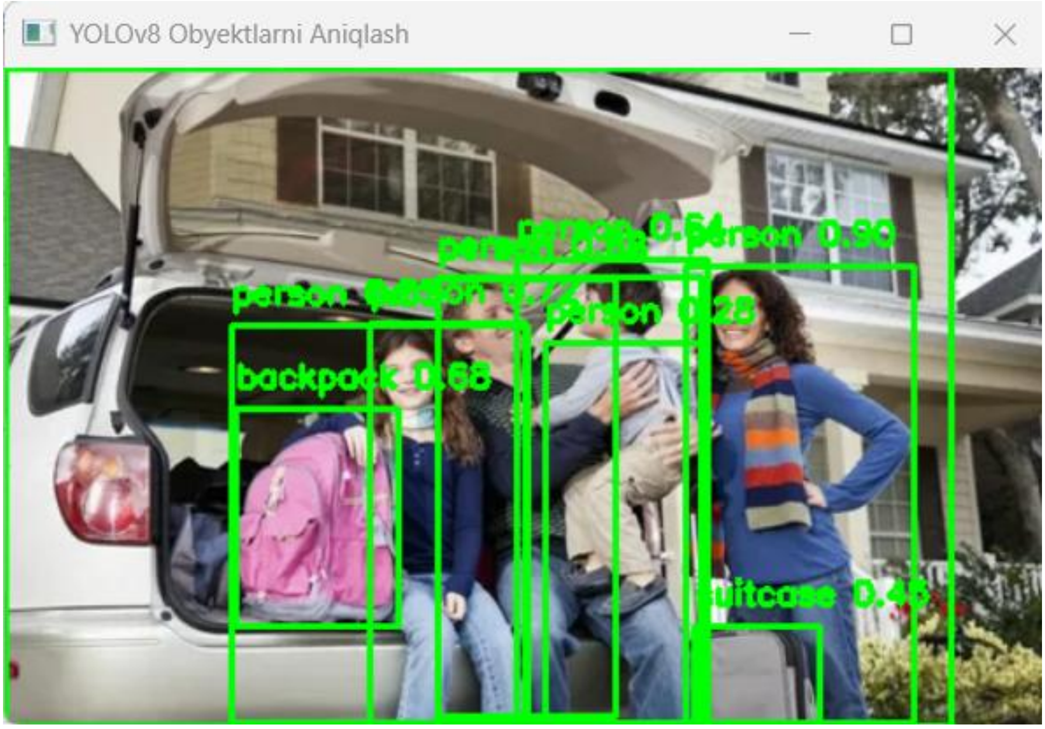

Figure 2. Object detection with the YOLOv5 model

This image illustrates the object detection process of the YOLOv8 model. The model has accurately identified several people standing next to a car, as well as objects like a backpack and a suitcase. Detected objects are enclosed in green bounding boxes, each displaying a confidence level. With its high degree of accuracy, the YOLOv8 model demonstrates its suitability for making reliable real-time decisions. This figure visually represents the model's capability to detect objects in various real-world scenarios.

Table 3.

Accuracy levels of object detection by the YOLOv8 model in the image

Object	Model	Accuracy
Person 1	YOLOv8	0.50
Person 2	YOLOv8	0.99
Person 3	YOLOv8	0.92
Person 4	YOLOv8	0.98
Backpack	YOLOv8	0.68
Suitcase	YOLOv8	0.45

When comparing YOLOv5 and YOLOv8 models, clear differences in accuracy emerge. The YOLOv8 model demonstrates much higher accuracy in detecting people, particularly for the "Person 2" and "Person 4" objects, where accuracy levels are notably improved. Meanwhile, the YOLOv5 model did detect a "Train" object, but with very low accuracy, and YOLOv8 did not detect it at all. Both models recognized the "Backpack" and "Suitcase" objects, though YOLOv5 showed slightly higher accuracy for the suitcase. Overall, YOLOv8 provides consistent, high accuracy for human-related objects, whereas YOLOv5 attempts to identify additional objects but with lower accuracy.

Conclusion

This article presents an analysis of two versions of the YOLO models. YOLO models are widely employed in object detection and have been successfully applied in various sectors, including industry, agriculture, medicine, and security. Their primary advantage is their speed and capability for real-time tracking, contributing to increased efficiency. The evolution of YOLO models from YOLOv1 to YOLOv8 has significantly advanced object detection technology. Which model to choose depends on the situation and requirements: YOLOv5 is recommended for speed and efficiency, while the new YOLOv8 model is preferred for accuracy and stability.

References

1. Redmon J. You only look once: Unified, real-time object detection //Proceedings of the IEEE conference on computer vision and pattern recognition. – 2016.
2. Shafiee M. J. et al. Fast YOLO: A fast you only look once system for real-time embedded object detection in video //arXiv preprint arXiv:1709.05943. – 2017.
3. Вильданов А. Н. Генерация датасетов для учебных задач компьютерного зрения //Инженерный вестник Дона. – 2023. – №. 4 (100). – С. 92-101.
4. Агафьина А. Д., Никитин А. А. Об автоматическом распознавании печатного текста //International Journal of Open Information Technologies. – 2024. – Т. 12. – №. 9. – С. 48-61.
5. Маркеев М. В. Методика идентификации объектов по уникальным характеристикам на изображениях //Universum: технические науки. – 2022. – №. 11-1 (104). – С. 24-27.
6. Муаль М. Н. Б. и др. Разработка нейросетевого метода в задаче классификации и распознавании изображения //Современные информационные технологии и ИТ-образование. – 2021.
7. Avazov K. et al. Fire detection and notification method in ship areas using deep learning and computer vision approaches //Sensors. – 2023. – Т. 23. – №. 16. – С. 7078.
8. Z. Shao, L. Wang, Z. Wang, W. Du and W. Wu, "Saliency-Aware Convolution Neural Network for Ship Detection in Surveillance Video," in IEEE Transactions on Circuits and Systems for Video Technology, vol. 30, no. 3, pp. 781-794, 2020, doi: 10.1109/TCSVT.2019.2897980
9. Кручинин А. Ю. Сложность распознавания при разработке программного обеспечения для видеомониторинга //Программные продукты и системы. – 2023. – Т. 36. – №. 1. – С. 123-129.
10. Нуржанов Ф. Р., Даулетов А. Ю. Математические методы и алгоритмы распознавания изображения лица человека //INTERNATIONAL INNOVATION RESEARCH. – 2017. – С. 123-126.
11. Абрамов Н. С., Фраленко В. П., Хачумов М. В. Обзор методов распознавания образов на основе инвариантов к яркостным и геометрическим преобразованиям //Современные наукоемкие технологии. – 2020. – №. 6-1. – С. 110-117.