Journal of Artificial Intelligence Practice (2024)
Clausius Scientific Press, Canada
DOI: 10.23977/jaip.2024.070314
ISSN 2371-8412 Vol. 7 Num. 3

YOLOv1 to YOLOv10: A Comprehensive Review of YOLO Variants and Their Application in Medical Image Detection

Ahealisi Yeerjiang[1,2,a], Zongyu Wang[2,b], Xiangtong Huang[2,c], Jing Zhang[2,d], Qi Chen[2,e], Yucheng Qin[2,f], Jia He[1,2,*]

[1]School of Health Sciences and Engineering, University of Shanghai for Science and Technology, Shanghai, China
[2]Department of Health Statistics, Naval Medical University, Shanghai, China
[a]13943003752@163.com, [b]zongyu79727@126.com, [c]232472769@st.usst.edu.cn, [d]233352431@st.usst.edu.cn, [e]cqchenqi1989@163.com, [f]qinyc10@163.com
[*]Corresponding author: hejia63@yeah.net

Keywords: Deep learning, YOLO, Computer vision

Abstract: The rapid evolution of computer vision has elevated object detection to a central task within the field. In medicine, automated lesion detection has the potential to greatly improve diagnostic efficiency for clinicians. The extraordinary success of deep learning in computer vision has motivated researchers globally to apply these advancements to medical image analysis. Deep learning techniques have demonstrated superior performance in medical image classification, detection, segmentation, registration, and retrieval compared to traditional methods. Among these, the YOLO (You Only Look Once) series of algorithms stands out for their exceptional speed and accuracy, making them a popular choice for medical image detection. This paper presents the underlying principles and structure of the classic YOLO algorithms, reviews their current applications in medical image detection, addresses existing challenges, and explores future directions for the application of YOLO in this domain.

1. Introduction

The rapid advancement of computer technology has elevated object detection, a critical branch of computer vision, to a position of paramount importance. The primary objective of object detection is to accurately identify and locate objects within images. In recent years, deep learning-based object detection technologies have made significant strides, offering improved accuracy, speed, and robustness in complex environments compared to traditional methods. These advancements make deep learning particularly well-suited for medical image detection, as it eliminates the need for complex handcrafted feature design and achieves superior results through end-to-end training. Often, deep learning models outperform conventional detection algorithms. As a result, Computer-Aided Diagnosis (CAD) systems have reached, and in many cases surpassed, the diagnostic accuracy of human experts across various medical fields. CAD systems significantly enhance the sensitivity and specificity of pathological diagnoses, reduce the rate of missed diagnoses, and improve overall diagnostic efficiency. Today, CAD has become a central focus in the field of medical image analysis.

As illustrated in Figure 1, since the introduction of R-CNN [1] in 2014, deep learning-based object detection methods have become the dominant approach, replacing traditional techniques. These methods are typically categorized into two-stage and one-stage approaches based on how candidate regions are generated. Two-stage algorithms, such as R-CNN, SPP-net [2], Fast R-CNN [3], and Faster R-CNN [4], involve an initial phase of region proposal followed by the detection phase. In contrast, one-stage algorithms, like the YOLO series, SSD [5], and RetinaNet [6], bypass the region proposal stage and perform detection directly. Among these, the YOLO series stands out as one of the most successful one-stage object detection algorithms, playing a pivotal role in the evolution of the field. The YOLO series, from YOLOv1 to YOLOv10, is recognized for its representativeness, authority, and broad application, collectively referred to as the classic YOLO algorithms.

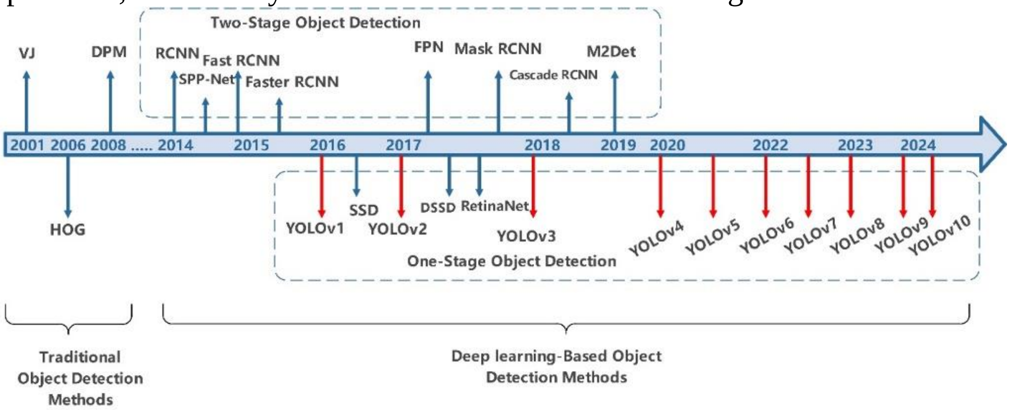

Figure 1: Development history of target detection.

2. Framework of the YOLO Series Algorithms

As illustrated in Figure 2, the YOLO model consists of three critical stages: input, feature extraction, and prediction. In the input stage, images are fed into the model and undergo data augmentation processes to enhance the robustness of the model. Different versions of YOLO vary in terms of the dimensions of input images, allowing for flexibility in adapting to various detection tasks. The feature extraction stage is driven by Convolutional Neural Networks (CNNs), which act as the backbone for extracting essential features from the images. Starting with YOLOv2, a neck layer was introduced, employing techniques such as Feature Pyramid Networks (FPN) to fuse multi-level features, thereby enhancing the network's ability to capture complex patterns. This multi-scale feature fusion is crucial for detecting objects of different sizes with high precision. In the prediction stage, anchor boxes of multiple sizes are generated based on the feature maps, and Non-Maximum Suppression (NMS) is applied to remove redundant bounding boxes, ensuring that the final predictions accurately reflect the detected objects' categories and locations. These architectural innovations and optimizations allow the YOLO model to attain high accuracy in object detection while preserving the rapid processing speeds essential for real-time applications. Additionally, the ability to balance speed and accuracy makes YOLO an appealing choice for various practical applications, from autonomous driving to medical image analysis.

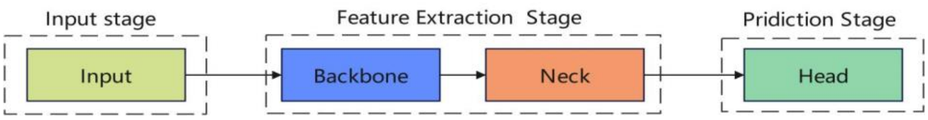

Figure 2: Overall Structure Diagram of the YOLO Algorithm.

3. Development of the YOLO Series Algorithms

YOLOv1 is an end-to-end real-time object detection algorithm introduced by Joseph Redmon et al. in 2016. It revolutionized object detection by transforming the task into a regression problem, processed across the entire image using a 24-layer Convolutional Neural Network (CNN). Unlike traditional two-stage detection methods, YOLOv1 achieves faster detection speeds and higher accuracy by dividing the image into a 7x7 grid. Each grid cell is responsible for detecting objects and predicting bounding boxes, confidence scores, and conditional class probabilities. This approach results in a 7x7x30 prediction matrix for each image.[7]

However, YOLOv1 has some limitations, such as poor prediction accuracy for small objects and imprecise object localization. To address these issues, Redmon et al. proposed the YOLOv2 algorithm in 2017. YOLOv2 introduced a new backbone network, Darknet-19, which includes 19 convolutional layers and 5 pooling layers, enhancing feature extraction capability and detection accuracy. The loss function was optimized in the prediction stage, and anchor boxes were introduced on the feature map to predict object positions in advance. Multi-scale training methods were implemented to accommodate varying image sizes, complemented by the addition of batch normalization layers at the end of each layer to expedite convergence and mitigate overfitting. These improvements significantly enhanced YOLOv2 in terms of detection accuracy, speed, and the number of classes it could classify, making it one of the leading object detection algorithms at the time.[8]

YOLOv3, proposed by Redmon et al. in 2018, features the Darknet-53 backbone network, which includes 53 convolutional layers and combines the advantages of Darknet-19 and ResNet networks. To improve sensitivity to small objects, YOLOv3 introduced a feature pyramid structure, performing upsampling and downsampling through forward and backward propagation, thereby merging deep and shallow features while preserving both semantic and graphic features of the image. Additionally, the detection head was optimized by replacing YOLOv2's single-label Softmax classifier with a multi-label Logistic classifier, enhancing the algorithm's flexibility. The LeakyReLU function was adopted as the activation function, boosting overall feature representation capability. The loss function combined mean square error and cross-entropy loss functions to calculate losses for coordinates, width, height, confidence, and classification information, ultimately summing them to obtain the final loss function.[9]

In 2020, Bochkovskiy et al. proposed YOLOv4, which received recognition from the YOLO series founder Redmon. YOLOv4 introduced several optimizations and innovations based on YOLOv3. In the input stage, Mosaic data augmentation, Cross mini-Batch Normalization (CmBN), and Self-Adversarial Training (SAT) data augmentation methods were introduced. Mosaic, in particular, significantly enhanced the model's ability to detect small objects by randomly scaling and stitching four images together, increasing dataset diversity. In the feature extraction stage, YOLOv4 employed CSPDarknet53 as the backbone network and introduced the Mish activation function and DropBlock regularization to effectively solve gradient information redundancy issues, improving feature extraction efficiency and effectiveness. The neck layer integrated Spatial Pyramid Pooling (SPP) and Path Aggregation Network (PAN) modules, replacing the original FPN module, further enhancing feature fusion capabilities and the model's performance in detecting objects at different scales.[10]

The same year, YOLOv5 was proposed. This method introduced Mosaic data augmentation, adaptive anchor box computation, and adaptive image scaling in the input stage. The CAPDarknet-53 network was used in the feature extraction stage to reduce parameters and computational complexity while maintaining high feature representation capability. The training loss function GIOU-Loss and the DIOU-NMS for prediction box filtering were improved in the prediction layer.

In 2022, Li et al. proposed YOLOv6, which abandoned traditional anchor boxes. In the input stage, images were transformed into uniform sizes through data augmentation. In the feature extraction stage, an efficient reparameterized backbone network, EfficientRep, and Rep-PAN neck were designed based on the RepVGG style. Additionally, decoupling was performed in the detection stage, using a simpler and more efficient Decoupled detection head to reduce the extra latency overhead brought by the decoupling head while maintaining accuracy.[11]

In the same year, researchers led by Alexey introduced YOLOv7, which stood out for its innovation in not employing anchor boxes, a shift that aimed to simplify the object detection process. They enhanced the input stage with advanced data augmentation techniques, including Mixup, Copy-Paste, and Paste-In, which contributed to improved model robustness and generalization by creating more diverse training samples. During the feature extraction phase, the traditional Backbone was redefined. Instead of the conventional layers, YOLOv7 implemented BConv (Block Convolution), E-ELAN (Enhanced Efficient Layer Aggregation Network), and MPConv (Multi-Path Convolution) layers. This change not only optimized the network architecture but also aimed to achieve better feature representation and extraction capabilities. In a significant architectural adjustment, the NECK layer and the HEAD layer were merged into a unified head layer. This newly formulated head layer employed a sophisticated combination of SPPCPC (Spatial Pyramid Pooling Convolutional Pathway), BConv, MPConv, and RepVGG block layers. Together, these components worked in harmony to enhance the model's ability to predict feature maps effectively, thereby improving the overall performance and accuracy of the object detection system. This integration of cutting-edge techniques reflected the ongoing evolution of deep learning architectures in the field of computer vision.[12]

YOLOv8, developed by the same team behind YOLOv5, is an advanced State-of-the-Art (SOTA) model. It introduces new features and improvements based on earlier successful YOLO versions to further enhance performance and flexibility. YOLOv8's backbone architecture is similar to YOLOv5's, but it replaces the C3 module with the C2f module in the feature extraction stage and adjusts channels according to different model scales, significantly improving model performance. In the prediction stage, YOLOv8 adopts a decoupled head structure, separating the classification and detection heads to further optimize the accuracy and efficiency of model outputs.[13]

YOLOv9, developed by the original YOLOv7 team in 2024, has been widely applied in various scenarios. This algorithm introduced the innovative concept of Programmable Gradient Information (PGI), generating reliable gradients through auxiliary reversible branches, maintaining the critical performance of deep features when executing target tasks. The design of PGI avoids the semantic loss that can occur in traditional deep supervision and incurs no additional computational cost. The PGI mechanism allows the selection of appropriate loss functions based on the target task, effectively overcoming some issues in mask modeling. Additionally, the algorithm designed generalized ELAN (GELAN) based on the ELAN, considering the number of parameters, computational complexity, accuracy, and inference speed, enabling users to choose suitable computational blocks according to different inference devices. GELAN achieves higher parameter utilization compared to deep convolution designs based on cutting-edge technology, demonstrating lightweight, fast, and accurate advantages.[14]

In the same year, the YOLOv10 algorithm was proposed by a research team at Tsinghua University.

It adheres to the design principles of the YOLO series, aiming to create a high-performance real-time end-to-end object detector. In the feature extraction stage, large-kernel convolution and partial self-attention are used to improve detection accuracy without significantly increasing computational cost. Spatial-channel decoupled downsampling and rank-guided modules are introduced to reduce computational redundancy and enhance overall model efficiency. In the prediction stage, a consistent dual assignment strategy is introduced, where multiple detection heads are used during training to provide more positive samples and enrich model training. During inference, gradient truncation is used to switch to a one-to-one detection head, eliminating the need for NMS post-processing and reducing inference overhead while maintaining performance.[15]

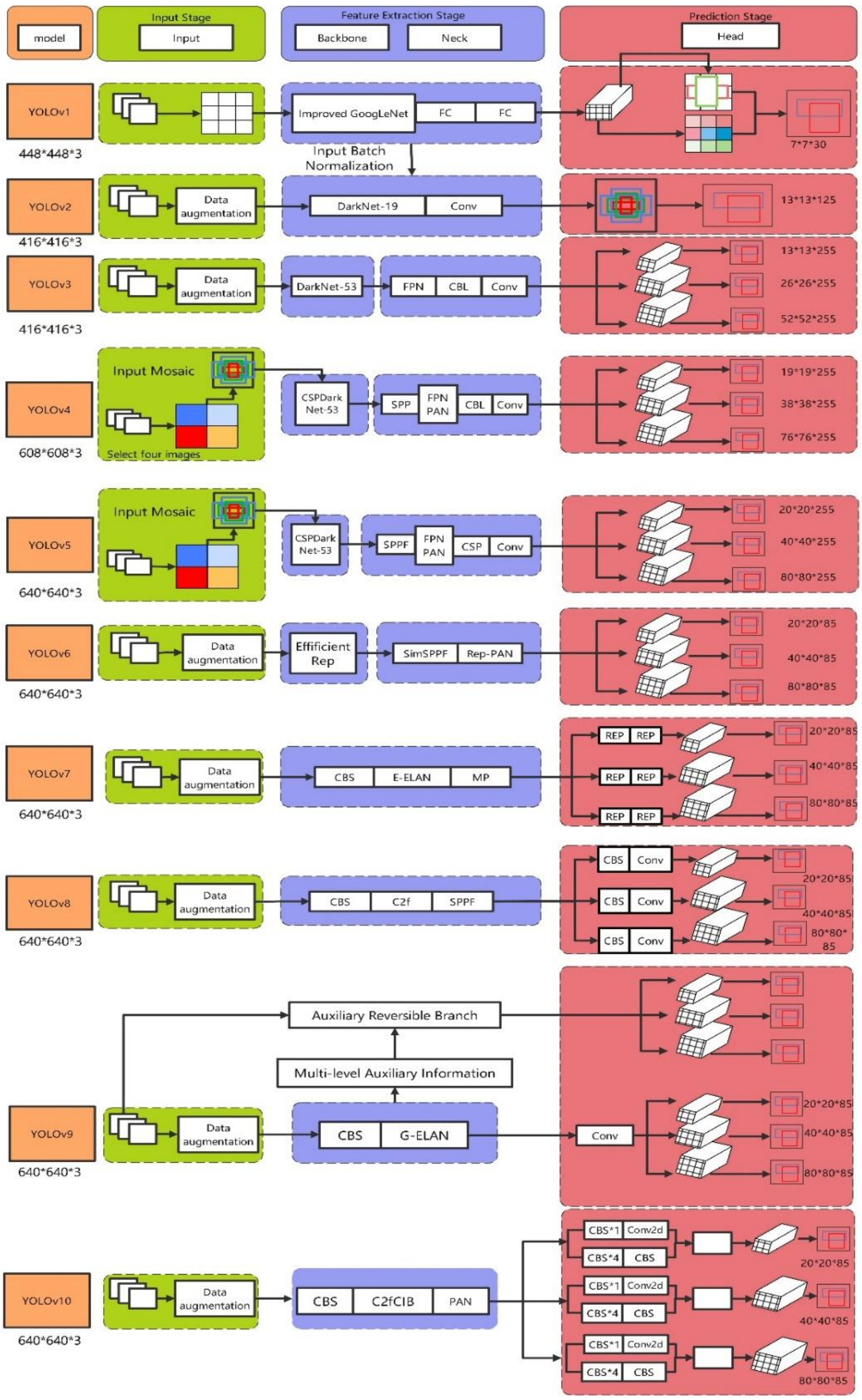

Figure 3: Schematic diagram of YOLOv1~v10 series algorithm network structure.

Figure 3 illustrates the overall architecture of the YOLOv1 to YOLOv10 models, which consist of three primary components: the backbone, neck, and head. Although these models share a similar general structure, distinct techniques are applied within each component to optimize detection performance.

Table 1 outlines the components and performance metrics of the YOLOv1 through YOLOv10 algorithms, highlighting a consistent improvement in performance across successive iterations.

Table 1: Performance comparison of YOLOv1~v10 algorithms

Version	Date	Input	Backbone	Neck	Voc07(AP)	Coco17(AP)
V1	2016	448	GoogLeNet	no	63.4%	
V2	2017	416	DarkNet-19	no	76.8%	
V3	2018	608	DarkNet-53	FPN		33.0%
V4	2020	608	CSPDarkNet-53	SPP		43.5%
V5	2020	608	CSPDarkNet-53	SPP/PAN		50.7% (v5-X)
V6	2022	640	EfficientRep	SPP/Rep-PAN		52.5% (v6-L)
V7	2022	640	CBS+ELAN+MP	SPP/PAN		52.9% (v7-X)
V8	2023	640	CBS+C2f+SPPF	SPP/PAN		53.9% (v8-X)
V9	2024	640	CBS+G-ELAN	PAN		55.6% (v9-E
V10	2024	640	PSA	C2f/PAN		54.4% (v10-X)

4. Applications of YOLO Algorithms in Medical Image Detection

The YOLO series of algorithms represents a groundbreaking real-time object detection system, demonstrating unique advantages in medical image analysis due to its exceptional speed and accuracy. Its real-time processing capabilities are particularly crucial in emergency medical situations, while its high detection accuracy makes it suitable for various types of medical images, including X-rays, CT scans, MRI, and ultrasound images. The flexibility and scalability of YOLO algorithms allow them to adapt to complex medical image analysis tasks, ranging from routine disease screening to intricate pathological analysis. This provides efficient object detection solutions and significantly advances the automation of medical imaging.

4.1. Applications in Tumor Detection

In the field of medical image processing, tumor image detection is a key technology for improving the accuracy and efficiency of pathological diagnoses and is also vital for advancing precision medicine. With the rapid development of deep learning technologies, YOLO algorithms have demonstrated unprecedented potential and effectiveness in tumor image detection, particularly in the detection of breast and brain tumors.

Breast cancer, one of the most common cancers and a leading cause of cancer-related deaths among women, has seen significant advancements through the application of YOLO algorithms. For instance, Prinzi et al. compared YOLOv3, YOLOv5, and YOLOv5-Transformer models using the CBIS-DDSM and INbreast datasets as source datasets and a proprietary dataset as the target dataset. YOLOv5s emerged as the best model with a mean average precision (mAP) of 0.621 on the proprietary dataset.[16] SU et al. employed the YOLO-LOGO model for detecting and segmenting breast cancer, initially using YOLOv5 to detect lesion areas and then using the LOGO training method to segment these areas. This model achieved recall, average precision, F1-score, and IoU of 95.7%, 65.0%, 74.5%, and 64.0%, respectively, on the CBIS-DDSM dataset.[17]

Brain tumors, among the most fatal diseases in adults due to their rapid growth and organ-destructive nature, require accurate detection for correct diagnosis and patient survival. With the continuous updates to YOLO algorithms, research on using YOLO for brain tumor detection has become increasingly prevalent. For example, Kang et al. integrated a Bi-level Attention (BRA), a Generalized feature pyramid networks (GFPN), and a fourth detection head into YOLOv8, developing the novel BGF-YOLO architecture. This model achieved precision, recall, and mAP50 of 91.9%, 92.6%, and 97.4%, respectively, in brain tumor detection.[18] Karacı et al. proposed a three-stage hybrid classification architecture based on YOLO, DenseNet, and Bi-LSTM for classifying gliomas, meningiomas, and pituitary tumors. In this framework, the brain region is first detected through the YOLO algorithm. In the second stage, deep features are extracted from this region via a pre-trained deep learning architecture, and in the final stage, brain tumor classification is performed using the Bi-LSTM network. The YOLO+DenseNet+Bi-LSTM combination showed the best performance with precision, recall, and F1 scores of 99.58%, 99.64%, and 99.61%, respectively.[19]

These studies demonstrate that continuous improvements in YOLO algorithms have significantly enhanced their applications in brain tumor detection and classification. Specifically, incorporating advanced technologies and architectures such as dual-layer attention mechanisms, generalized feature pyramids, and multi-stage hybrid classification frameworks has significantly improved detection accuracy and recall rates. YOLO algorithms have also shown robust capabilities in handling complex backgrounds and detecting small objects. These advancements have enabled YOLO algorithms to achieve outstanding results in real-time detection and precise classification, effectively aiding clinicians in accurately identifying and diagnosing brain tumors.

Table 2: Applications of YOLO series algorithms in tumor detection

Author	Date	Dataset	Disease	Model	Performance Measure
Prinzi[16]	2023	CBIS-DDSM INbreast private dataset	Breast Cancer	YOLOv3 YOLOv5	mAP=0.621
SU[17]	2022	CBIS-DDSM	Breast Cancer	YOLO-LOGO	Recall=95.7 F1-score=74.5 IoU=64.0
Kang [18]	2023	Br35H	Brain Tumor	BGF-YOLO	Precision=91.9 Recall=92.6 mAP=97.4
Karacı [19]	2023	brain tumor dataset	Brain Tumor	YoDenBi-NET	Recall=99.64 Precision=99.58 F1-score=99.61

Table 2: Summary of recent studies utilizing YOLO-based models for tumor detection across various medical imaging datasets. The table provides details on the authors, publication year, datasets used, target diseases, specific YOLO models employed, and the key performance measures reported. The results highlight the effectiveness of these models in achieving high precision, recall, and mAP across different tumor detection applications.

4.2. Applications in Fracture Detection

Fractures are among the most common clinical injuries, and rapid, accurate fracture detection is crucial for effective treatment and rehabilitation. Traditional fracture diagnosis relies on radiologists' manual analysis of X-ray or CT images, which is time-consuming and subject to subjective interpretation. Recently, with the advancement of deep learning technology, the YOLO algorithm has demonstrated excellent performance in object detection and has gradually been introduced into the field of fracture detection. For example, Zhou et al. designed a 3M-YOLOv5 network to detect mandibular fracture locations, achieving mAP, F1-score, recall, and precision of 99.17%, 99.06%, 98.81%, and 99.32%, respectively.[20] Bai et al. used the YOLOv3 algorithm to detect rib fractures, achieving accuracy, recall, and F1-score of 90.5%, 75.5%, and 0.82, respectively.[21] Hržić et al. applied the YOLOv4 model for fracture detection on the GRAZPEDWRI-DX dataset, marking the first instance of a YOLO series model aiding radiologists in more accurately predicting pediatric wrist injuries on X-ray images. Their model achieved precision, recall, F1-score, and accuracy of 95.94%, 95.39%, 95.45%, and 95.39%, respectively.[22] Ju et al. employed the YOLOv8 algorithm to detect fractures in pediatric wrist trauma X-ray images, achieving precision, recall, and mAP of 69.4%, 59.2%, and 63.1%, respectively.[23]

The application prospects of the YOLO algorithm in fracture detection are broad, significantly improving diagnostic efficiency and accuracy. Despite challenges such as data annotation, complexity, and accuracy, the continued development and deeper application of YOLO technology promise an increasingly important role in fracture detection. Through exploration in multimodal fusion, personalized medicine, and integrated AI systems, the YOLO algorithm will further drive the advancement of fracture detection technology, providing stronger support for clinical diagnosis and treatment.

Table 3: Overview of recent studies employing YOLO-based models for fracture detection across various medical imaging datasets. This table details the authors, publication year, datasets used, specific fracture types, YOLO models applied, and the key performance metrics. The results demonstrate the applicability of YOLO models in accurately detecting different types of fractures, with varying degrees of recall, precision, mAP, and F1-scores across different datasets.

Table 3: Applications of YOLO series algorithms in fracture detection

Author	Year	Dataset	Disease	Model	Performance Metrics
Zhou [20]	2023	A tertiary hospital in Ningxia	Mandibular fracture	3M-YOLOv5	Recall=98.81 Precision=99.32 mAP=99.17 F1-score=99.06
Bai [21]	2023	Private Dataset	Rib Fracture	YOLOv3	Recall=75.5 Accuracy=90.5 F1-score=82
Hržić [22]	2022	GRAZPEDWRI-DX	Wrist Fracture	YOLOv4	Recall=95.39 Precision=95.94 Accuracy=95.39 F1-score=95.45
Ju [23]	2023	GRAZPEDWRI-DX	Wrist Fracture	YOLOv8	Recall=59.2 Precision=69.4 mAP=63.1

4.3. Applications in Lesion Area Detection

Identifying lesion areas is crucial in medical diagnosis, aiding doctors in accurately locating and assessing the affected regions. The YOLO algorithm has demonstrated robust capabilities in lesion area identification, particularly in handling various types of medical images, such as X-rays, CT, MRI, ultrasound, and endoscopic images, enabling rapid and accurate detection and annotation.

4.3.1. Lung Nodule Detection

Lung nodules are a common type of pulmonary lesion that can be benign or malignant. Early detection and identification of lung nodules are vital for the early diagnosis and treatment of lung cancer. Traditional lung nodule detection methods primarily rely on radiologists' observation and analysis of CT images, which is time-consuming and prone to subjective factors. As an efficient object detection algorithm, YOLO has shown significant potential in the automatic detection of lung nodules. For instance, Song et al. improved the original YOLOv5 for the detection of small targets like lung nodules, further enhancing detection accuracy and sensitivity, achieving recall and mAP of 91.36% and 86.51%, respectively.[24] Xi et al. used YOLOv3 to detect lung nodules and compared it with SSC, CNN, U-Net, and other models to verify the effectiveness of the YOLOv3 model. The YOLOv3 model achieved an accuracy and recall rate of 92.18% and 97%, respectively.[25]

4.3.2. Diabetic Retinopathy Detection

Diabetic retinopathy (DR) is a common complication of diabetes and one of the leading causes of blindness in adults. Early detection and intervention are crucial to prevent vision loss. Traditional DR detection mainly relies on ophthalmologists' observation and analysis of fundus images, which is time-consuming and prone to subjective factors. As an efficient object detection algorithm, YOLO has shown significant potential in the automatic detection of diabetic retinopathy. For example, Gao et al. used an improved YOLOv4 with embedded SENet to detect microaneurysms, achieving recall, precision, and AP of 89.77%, 87.13%, and 88.92%, respectively.[26] Zhang et al. utilized the YOLOv4-tiny algorithm to detect the optic disc and macular fovea in retinal images, achieving a mAP of 96.11%.[27]

4.3.3. Skin Cancer Detection

Skin cancer is one of the most common malignant tumors, primarily including melanoma, basal cell carcinoma, and squamous cell carcinoma. Early detection and diagnosis are crucial for improving patient survival rates. Traditional skin cancer detection mainly relies on dermatologists' expertise and visual examination, which are subject to a degree of subjectivity and risk of misdiagnosis. As an efficient object detection algorithm, YOLO has shown significant potential in the automatic detection of skin cancer. For example, Ünver et al. first used YOLOv3 for lesion localization and then applied GrabCut for lesion segmentation. The accuracy of YOLOv3 for skin lesion localization was 94.40% on the PH2 dataset and 96% on the ISBI2017 dataset.[28] Nersisson et al. extracted texture features using GLCM and color features using CLCM, combining these texture-color features with YOLO-CNN features to detect skin cancer, achieving accuracy, precision, and recall of 94%, 85%, and 88%, respectively.[29]

Table 4: Applications of YOLO series algorithms in lesion area detection

Author	**Year**	**Dataset**	**Disease**	**Model**	**Performance Metrics**
Song [24]	2023	LUNA16	Lung Nodule	Improved YOLOv5	Recall=91.36 mAP=86.51
Xi [25]	2020	LUNA16	Lung Nodule	YOLOv3	Recall=97 Accuracy=92.18
Gao [26]	2022	Diabetic Retinopathy dataset	Diabetic Retinopathy	Improved YOLOv4	Recall=89.77 Precision=87.1
Zhang [27]	2022	Kaggle diabetic Retinopathy dataset	Diabetic Retinopathy	YOLOv4-tine	mAP=96.11
Ünver [28]	2019	PH2 ISBI2017	Skin Cancer	YOLOv3	Accuracy=94.40 Accuracy=96
Nersisson [29]	2021	ISBI Melanoma dataset	Skin Cancer	Improved YOLOv2	Accuracy=94 Precision=85% Recall=88%

Table 4: Summary of recent studies applying YOLO-based models for lesion area detection in various medical imaging datasets. The table includes details on the authors, year of publication, datasets, types of lesions, models used, and their corresponding performance metrics. These studies demonstrate the effectiveness of improved YOLO models in detecting lung nodules, diabetic retinopathy, and skin cancer, with notable performance in terms of recall, precision, mAP, and accuracy.

5. Conclusion

The YOLO algorithm holds significant promise for medical image detection, offering substantial improvements in the efficiency and accuracy of disease diagnosis. This paper has provided an overview of the YOLO algorithm's structure and evolution, followed by an exploration of its specific applications in tumor detection, fracture detection, and lesion area identification. The analysis of these applications demonstrates YOLO's exceptional performance in efficiently processing large volumes of images, enabling real-time detection, and accurately identifying a wide range of lesion types and sizes when trained on extensive annotated datasets. This capability enhances the accuracy of early diagnoses. In tumor detection, YOLO effectively localizes breast cancer, brain tumors, and thyroid nodules, aiding clinicians in devising more precise treatment strategies. In fracture detection, the algorithm quickly identifies fracture locations and types, minimizing the time and risk associated with manual diagnosis. Additionally, in lesion area identification, YOLO accurately detects conditions such as diabetic retinopathy and skin lesions, supporting clinicians in disease evaluation and treatment planning.

Despite its notable achievements in medical image detection, the YOLO algorithm faces challenges and opportunities for improvement. Future research could focus on enhancing detection accuracy and robustness by integrating YOLO with other image processing technologies and deep learning models, as well as incorporating multimodal data (e.g., imaging and genetic data). The development of personalized detection and diagnostic models that account for individual patient differences and medical histories could lead to more precise medical services tailored to each patient. Embedding the YOLO algorithm into medical devices for real-time patient monitoring and early warning systems could enhance the speed and efficacy of medical responses to abnormalities. Furthermore, establishing high-quality, large-scale medical image datasets to facilitate data sharing and openness, reduce annotation costs, and improve model training will be crucial for future advancements. Integrating the YOLO algorithm with other AI technologies, such as natural language processing and robotic surgery, could pave the way for comprehensive intelligent medical systems, further elevating the quality of healthcare services. Through ongoing technological innovation and expanded applications, the YOLO algorithm is poised to drive the continued development of medical image detection technologies, providing stronger support for clinical diagnosis and treatment, and ultimately improving patient health and quality of life.

Acknowledgements

This work was supported by the Shanghai municipal health commission Special Research Project in Emerging Interdisciplinary Fields [2022JC011].

References

[1] Girshick R, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2014: 580-587.

[2] He K, Zhang X, Ren S, et al. Spatial pyramid pooling in deep convolutional networks for visual recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 37(9): 1904-1916.

[3] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[4] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[J]. Advances in neural information processing systems, 2015, 28.

[5] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016: 21-37.

[6] Lin T Y, Goyal P, Girshick R, et al. Focal loss for dense object detection[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2980-2988.

[7] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.
[8] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
[9] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
[10] Bochkovskiy A. YOLOv4: Optimal Speed and Accuracy of Object Detection[J]. arXiv preprint arXiv:2004.10934, 2020.
[11] Li C, Li L, Jiang H, et al. YOLOv6: A single-stage object detection framework for industrial applications[J]. arXiv preprint arXiv:2209.02976, 2022.
[12] Wang C Y, Bochkovskiy A, Liao H Y M. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2023: 7464-7475.
[13] Varghese R, Sambath M. YOLOv8: A Novel Object Detection Algorithm with Enhanced Performance and Robustness[C]//2024 International Conference on Advances in Data Engineering and Intelligent Computing Systems (ADICS). IEEE, 2024: 1-6.
[14] Wang C Y, Yeh I H, Liao H Y M. Yolov9: Learning what you want to learn using programmable gradient information [J]. arXiv preprint arXiv:2402.13616, 2024.
[15] Wang A, Chen H, Liu L, et al. Yolov10: Real-time end-to-end object detection[J]. arXiv preprint arXiv:2405.14458, 2024.
[16] Prinzi F, Insalaco M, Orlando A, et al. A yolo-based model for breast cancer detection in mammograms[J]. Cognitive Computation, 2024, 16(1): 107-120.
[17] Su Y, Liu Q, Xie W, et al. YOLO-LOGO: A transformer-based YOLO segmentation model for breast mass detection and segmentation in digital mammograms[J]. Computer Methods and Programs in Biomedicine, 2022, 221: 106903.
[18] Kang M, Ting C M, Ting F F, et al. Bgf-yolo: Enhanced yolov8 with multiscale attentional feature fusion for brain tumor detection [J]. arXiv preprint arXiv:2309.12585, 2023.
[19] Karacı A, Akyol K. YoDenBi-NET: YOLO+ DenseNet+ Bi-LSTM-based hybrid deep learning model for brain tumor classification [J]. Neural Computing and Applications, 2023, 35(17): 12583-12598.
[20] Tao Zhou, et al. "Mandibular fracture detection with enhanced feature extraction capabilities in the 3M-YOLOv5 network." Optics and Precision Engineering 31.21 (2023): 3178-3191.
[21] Bai J, Sun J, Cheng X G, et al. Construction and Application of Rib Fracture Diagnosis Model Based on YOLOv3 Algorithm [J]. Fa yi xue za zhi, 2023, 39(4): 343-349.
[22] Hržić F, Tschauner S, Sorantin E, et al. Fracture recognition in paediatric wrist radiographs: An object detection approach [J]. Mathematics, 2022, 10(16): 2939.
[23] Ju R Y, Cai W. Fracture detection in pediatric wrist trauma X-ray images using YOLOv8 algorithm[J]. Scientific Reports, 2023, 13(1): 20077.
[24] SONG Fangfang, SUN Zhaoyong, TIAN Yimin, et al. Improved YOLOv5 Pulmonary Nodule Detection Method[J]. Software Engineering and Applications, 2023, 12: 257.
[25] Xi Xiaoqian, Liu Wei. Auxiliary Diagnosis System for Pulmonary Nodules Based on Object Detection Algorithm[J]. Computer and Modernization, 2020 (11): 1.
[26] Gao W, Shan M, Song N, et al. Detection of microaneurysms in fundus images based on improved YOLOv4 with SENet embedded[J]. Journal of Biomedical Engineering, 2022, 39(4): 713-720.
[27] Wei Z, Hua Z, Yuhong L I U, et al. Research on Optic Disc and Macula Fovea Simultaneous Location and Detection Method on FPGA[J]. Journal of Computer Engineering & Applications, 2022, 58(11).
[28] Ünver H M, Ayan E. Skin lesion segmentation in dermoscopic images with combination of YOLO and grabcut algorithm [J]. Diagnostics, 2019, 9(3): 72.
[29] Nersisson R, Iyer T J, Joseph Raj A N, et al. A dermoscopic skin lesion classification technique using YOLO-CNN and traditional feature model[J]. Arabian Journal for Science and Engineering, 2021, 46(10): 9797-9808.