Systematic study of lightweight for object detection

Haoyu Liu

Chengdu University of Technology, No. 1, Dongsan Road, Erxianqiao Street, Chenghua District, Chengdu City, Sichuan Province

lhy010621edu@outlook.com

Abstract. Object detection is a critical task in computer vision that has been studied for many years. However, existing object detection models are often computationally expensive and require high-end hardware to achieve real-time performance. To address this issue, many efforts aim to propose lightweight detection deep models to meet the requirements of resource-critical applications. YOLO is such a representative work that relies on a one-stage detection design and a series of optimization strategies. Correspondingly, related object detection works aiming at lightweight requirements based on YOLO are continuously proposed. Nevertheless, there is still a lack of a systematic study on the development of YOLO-based works, making it difficult to understand the lightweight trend for the object detection task. In this paper, we bridge the gap between YOLO's lightweight design and object detection. We comprehensively studied numerous representative works in this field and analyzed their concerns, approaches, performance, and application scenarios in detail. Moreover, we discussed the impacts of these efforts from multiple dimensions and proposed future directions for lightweight detection. Our study and findings are expected to provide valuable suggestions for object detection in low-power and resource-intensive devices.

Keywords: YOLO, Object Detection, Lightweight, Resource-Critical, Time Cost.

1. Introduction

Object Detection Object Detection (OD), abbreviated as OD, plays an indispensable role in various computer vision applications, encompassing fields such as autonomous vehicles, surveillance systems, and robotics. The fundamental objective of Object Detection is to identify target objects within an image and determine both their respective categories and positions. The complexity of this task arises from the diverse appearances, shapes, and orientations of different objects, combined with factors like illumination variations and occlusion during image capture. In recent times, the research community has proposed two main methodologies for Object Detection: two-stage models and one-stage models. The former tend to suffer from intricate model architectures and excessive region proposals, leading to computationally expensive operations. Conversely, one-stage models have garnered more attention due to their simplicity in design and cost optimization. Among these models, YOLO [1] stands as a representative one-stage approach that addresses object detection through regression. By processing the input image only once, YOLO efficiently obtains the positions of target objects, their respective categories, and the corresponding confidence probabilities. This streamlined model design enables YOLO to detect target objects significantly faster than traditional two-stage methods, making it a preferred choice for applications on resource-constrained devices such as mobile phones and IoT devices.

Despite the effectiveness of the original YOLO models, researchers have continuously proposed variants of YOLO tailored to specific tasks or scenarios [2-11], such as YOLO v2-v5 [12-14], YOLO-LITE [8], YOLO tiny [3], and others. However, a comprehensive investigation into the lightweight adaptations of YOLO and its variants is still lacking. This gap hinders a thorough understanding of the development of lightweight approaches for the OD task. Moreover, with the increasing prevalence of OD on mobile and IoT devices, grasping the lightweight OD trend through future research will be essential for enhancing detection efficiency and real-world applications.

To address this research gap, this paper presents a systematic study of lightweight adaptations based on the YOLO framework. Initially, we provide the necessary background knowledge about the OD task and introduce the representative YOLO models. Subsequently, we conduct a comprehensive examination of recent YOLO-based lightweight OD works from multiple research dimensions, including approaches, results, and application scenarios. Furthermore, we delve into a detailed discussion of the existing related efforts. Finally, based on our study, we analyze potential future research directions to further advance the field of lightweight OD.

2. Object detection

Object detection is a classic task in the field of computer vision, encompassing four fundamental problem categories:

(1) Classification problem: determining the category to which an object in the picture (or a certain area) belongs.

(2) Localization problem: identifying where the target object may appear in the image.

(3) Size problem: ascertaining the size of the target object.

(4) Shape problem: recognizing the shape presented by the target object.

To address these problems, two-stage methods were proposed earlier, with representative methods such as R-CNN, SPP-Net, Fast R-CNN, Faster R-CNN, and R-FCN. The fundamental design concept underlying two-stage object detection is as follows: Firstly, the model generates region proposals (RPs) in the original image, which are pre-selection boxes that may potentially contain objects requiring inspection. Subsequently, the objects undergo classification through a CNN or other advanced learning models. The two-stage methods offer advantages in terms of detection precision, owing to their utilization of an anchor mechanism to extract RPs, thereby enhancing detection performance. However, these methods face limitations concerning detection efficiency and training time costs since each stage necessitates learning prediction results based on data inputs derived from the preceding stage. Additionally, the two-stage approaches are also constrained by false positive rates, as an excess of RPs will inevitably lead to redundant detection outcomes. Consequently, achieving real-time performance with these methods requires high-end hardware and renders them unsuitable for deployment on low-power devices, thus limiting their widespread adoption in real-world scenarios.

In stark contrast to the two-stage methods, one-stage methods execute the classification and localization of target objects through direct deep neural network (DNN) based feature extraction. In comparison to their two-stage counterparts, one-stage approaches exhibit superior detection efficiency in both training and detection stages, as they eliminate the time-consuming construction of RPs. One-stage methods generally acquire the ability to learn generalization features of objects, thus mitigating false positives in detection. However, the detection precision of one-stage methods is limited by the open feature extraction networks they utilize. To address this challenge, one-stage methods typically concentrate on constructing deep models. Unlike simple CNNs or their variants, the one-stage network can be divided into different modules to extract various dimensions of image features.

As a representative lightweight one-stage model for object detection, YOLO [1] employs numerous lightweight strategies to achieve enhanced accuracy and efficiency. The details of YOLO and its variants are expounded upon in section 3.3.

3. Study of lightweight object detection

3.1. Global map

To systematically study lightweight works in object detection (OD), we have collected numerous YOLO-based efforts in recent years. Through careful screening, analysis, and summarization, the key information is presented in Table 1, where each work's first author, representative tool, target, approach, result, and scenario are listed.

Table 1. Representative lightweight works based on YOLO.

First Author	Tool[ref]	Lightweight Target	Approach*	Result	Scenario
Rachel Huang	YOLO-LITE	Fast non-GPU object detection applications	1) Based on Tiny-YOLO-V2 2) Image size reducing 3) Remove batch normalization	1) Dataset: PASCAL VOC 2) mAP>30% 3) FPS>20	
Pranav Adarsh	YOLO v3-Tiny	To improve detection speed	1) Fine-grained pooling layer 2) Reduce figure for convolution layer	442% faster than the former variants of YOLO	
Zihan Ni	Light Yolo	To accelerate the refined network	1) Based on YOLO-V2 2) Spatial refinement module 3) Selective channel pruning for discarding channels	1) Accuracy from 96.80% to 98.06% 2) Speed form 40PFS to 125FPS 3) Size form 250M to 4MB.	Handling small-scale gestures in practical applications
Dongyang Zhang	Smart-YOLO	To improve accuracy and speed of real-time object detection with fewer computing resources	1) Reverse bottleneck blocks and deep separable convolutions 2) new Loss function	1) Accuracy is reduced by 21% 2) Up to 4.5x acceleration 3) Model size is only about 1/8 of the original one.	
Jiahuan Jiang	YOLO-V4-lightship	Practical application in maritime safety monitoring and emergency rescue.	1) A multi-channel fusion SAR image processing method 2) Based on the latest YOLO-V4	1) mAP of SSDD is 90.37% 2) Model is Simplified	Ocean monitoring and remote sensing field
ZHOU Long	Lira-YOLO	Siganificantly reduces the network operation complexity.	1) Two YOLO prediction layers with dense connection 2) Residual modules for better convey features. 3) Built a small RD, lightweight range Doppler domain radar image dataset	1) Low network complexity of only 2.980Bflops 2) The parameter quantity is small, only 4.3MB 3) mAP on SSDD reaches 83.21% and 85.46%, respectively	Detection of Ship Targets in Radar Images
Rachel Huang	YOLO-LITE	Create a smaller, faster, and more efficient model increasing the accessibility of real-time object detection to a variety of devices.	1) Trained on the PASCAL VOC dataset and then on the COCO dataset 2) Discard the batch normalization 3) Implementation and operation on non-GPU websites with only 7 layers and 482 million FLOPS	1) 33.81% and 12.26% mAP on PASCAL VOC and COCO, respectively. 2) Speed of 21 FPS on non-GPU and 10 FPS on websites . 2. Running speed is 3.8x faster than MobiletvI.	Running on portable devices such as laptops or mobile phones that lack GPUs

Table 1. (continued)

Yunong Tian	Improved YOLO-V3 (YOLO-V3 dense model)	To detect apples during different growth stages in orchards with fluctuating illumination, and branches and leaves.	1) Collect images of young, bloated, and mature apples 2) Transformation of rotation, color balance and brightness, 3) Blur processing	1) Outperforms the original YOLO-V3 and Faster R-CNN with VGG16 network 2) 0.304s per frame for detection time at a resolution of 3000..	Detecting apples in orchards with fluctuating illumination, complex backgrounds
Rui Huang	YOLO-V3	Fast recognition for electronic components in complex background	1) Proposed a fast recognition method based on deep learning 2) Based on YOLO-V3	1) Accuracy reaches 95.21%, and the speed is 0.0794s 2) Proved the superiority in detection of electronic components.	Electronic components in complex background
Tossaporn Santad	YOLO	Abandoned-baggage detection for security detection in subway stations	1. Proposed a system 2. Designed and developed a GUI for parameter settings in the system	The system operates effectively under constant lighting and camera position conditions	Detection of abandoned luggage
Xiulong Gao	Based on YOLO-V3	Solve high power consumption and poor real-time performance of the autopilot vehicle detection .	1) Visualization of convolutional layers 2) Suitable anchor box 3) Non maximum suppression pruning	Accuracy of detection reaches to 98.7%.	Autonomous driving in automobiles
Guozhan Wang	YOLO-DFD	Solve the problem of dog feces in the living environment.	1) MobileNetV3 serves as backbone network 2) Deep separable convolution 3) Introduce CBAM in neck network 4) SCYLLA union intersection loss	1) mAP reaches 98.66% 2) Parameters are reduced by 82% 3) FPS is increased by 14	Solving the problem of dog excrement in the living environment
Petr Hurtik	Poly-YOLO	A large amount of rewritten labels and inefficient distribution of anchors.	1) Stepped upsampling 2) Hypervolume for aggregate features from lightweight SE-Darknet-53 3) Based on YOLO-v3	1) 60% of its trainable parameters 2) mAP has increased by a relative 40%.	Usually used for sampling on a ladder and Suitable for embedded devices
Wei Li	Based on YOLO-V3	Detect pedestrians in infrared images at night.	Infrared image training and testing based on YOLO-V3.	1) Accurately locate pedestrians at night 2) pedestrians overlap.	Pedestrian detection at night
Khaled Zaatouri	Based on YOLO	Realtime traffic light control based on the traffic flow	1) Extract traffic flow at signalized intersections 2) YOLO-based method to reduce detection time.	Vehicles pass safely within the shortest waiting time	Used to alleviate traffic pressure in real life

3.2. Lightweight of YOLO V1-V5

1) YOLO-V1 [1] represents the initial generation of the YOLO algorithm, primarily devised for target detection. Its key feature lies in enhancing detection speed and reducing false alarm rates. Built upon a CNN framework, YOLO-V1 directly predicts bounding boxes for multiple object categories and locations in a single forward pass. The primary advantage of YOLO-V1 is efficiency, yet it is constrained by scale and resolution, leading to relatively low detection accuracy. Moreover, its ability to handle small objects is comparatively poor.

2) YOLO-V2 [12] constitutes an improvement over YOLO-V1, incorporating a series of techniques such as batch normalization, anchor boxes, convolution expansion, and enhanced activation functions for the convolution layer, all of which further elevate YOLO's detection performance. This results in improved detection accuracy and speed, especially concerning small objects. In comparison to V1, YOLO-V2 features a more complex network structure and an increased number of parameters, leading to lengthier training and inference times.

3) YOLO-V3 [13] introduces more improvements and optimizations, such as multi-scale prediction, feature pyramid network, and upsampling, compared to YOLO-V1 and V2. These optimizations result in YOLO-V3 outperforming YOLO-V2 in terms of detection accuracy, speed, and localization capabilities. YOLO-V3 also supports larger objects and multi-scale detection. However, it is more complex than YOLO-V2 and requires more time and computing resources for training and inference.

4) YOLO-V4 [14] incorporates several technologies, including Scaled-YOLO, CSPDarknet53, CSPResNeXt50, YOLO-V4, and YOLO-V4-neck, which lead to significant improvements in speed and performance. Compared to YOLO-V3, YOLO-V4 shows enhancements in detection accuracy, speed, and robustness. However, YOLO-V4 is more complex, and the resource requirements for training and inference are higher.

5) YOLO-V5 adopts the structure of V4 and improves its portability, making it easier to deploy to new hardware and existing applications by reducing the code base and writing common COCO evaluation scripts. YOLO-V5 also supports various data enhancement methods and model center loss training. Compared to YOLO-V4, YOLO-V5 offers better speed, a smaller model size, easier deployment, and detection accuracy close to YOLO-V4.

3.3. Lightweight of variants

In addition to the YOLO models, there are several variants of YOLO. For instance, Light-YOLO [4] is a new lightweight tool developed based on YOLO-V2 that allows selective channel pruning of discarded channels and spatial refinement of modules. Smart-YOLO [5] is a new lightweight algorithm proposed based on YOLO-V3, utilizing reverse bottleneck blocks and deep separable convolutions. YOLO-LITE [8], constructed on the basis of YOLO-V2, is designed to create a smaller, faster, and more efficient model. YOLO-DFD [16] is an improved YOLO-V4 based lightweight detection algorithm developed to address the problem of dog feces in our living environment. Poly-YOLO [17], based on YOLO-V3, removes rewritten labels and inefficient distribution of anchors. However, the lightweight works for YOLO variants are often designed for non-targeted generalized tasks and are not directly applicable in specific scenarios.

3.4. Specific scenarios

There are also works proposed for specific scenarios. For example, Gao et al. [17] address high power consumption and poor performance in autonomous driving of automobiles by using YOLO-V3 support. Li et al. [18] accurately locate pedestrians at night. Additionally, Zaatouri [19] improves real-time traffic light control, while Tian et al. [9] accurately detect apples in orchards with fluctuating illumination and complex backgrounds. These YOLO-based lightweight works for specific scenarios often lack improvement in model construction.

4. Discussion

4.1. Existing efforts

4.1.1. Advantages. The advantages of the current YOLO-based OD lightweight object detection works are evident in the related publications. Firstly, some of the works aim to improve detection efficiency while maintaining an acceptable level of accuracy. They achieve this goal through improved complex neural network architectures and specific data strategies. This trend can be observed in the evolution of YOLO models. For example, YOLO-V1 predicted the four vertices of the target object using multiple

CNN layers and FC, while YOLO-V2 changed the prediction architecture to focus on anchors and grids and increased the prediction categories significantly. YOLO-V3 introduced more complexities in head categories and detection grids. YOLO-V4 further improved the detection head and changed the loss function to CIoU loss. YOLO-V5 also saw corresponding improvements in the backbone network.

4.1.2. Limitations. Our study reveals that YOLO Lightweight is a viable alternative to existing lightweight object detection models. However, there are limitations within the proposed YOLO-based models or applications. Although many existing efforts aim to balance performance in terms of both accuracy and efficiency, the improvements after YOLO-V4 are still limited. Some methods focus more on improving detection accuracy, but the efficiency for popularization and generalization is not satisfactory. Additionally, with the prevalence of applications on the Internet of Things and mobile terminals, the real-time requirements for detection algorithms are increasing. The current optimization models, including YOLO and its variants, are challenging to be directly used in these real-time scenarios. Special pruning or quantization processing is still required to achieve terminal-related adaptation. Furthermore, although many proposed models claim to achieve both high accuracy and efficiency, they are primarily trained on commonly used datasets like PASCAL VOC and COCO. The effectiveness of these models has not been sufficiently verified under new domains, limiting their practicality.

4.2. Future directions

Based on the analysis of both advantages and limitations of existing lightweight efforts, future works in this direction can be divided into the following three points.

4.2.1. Trade-off between accuracy and efficiency. The trade-off between accuracy and efficiency remains a key focus in lightweight object detection. However, most of the existing works excel in one aspect or show an average performance in both aspects, making it challenging to achieve optimal performance in both areas. Future research should focus on constructing new architectures and algorithms that can strike a better balance between accuracy and efficiency. Advanced techniques in computer vision, such as transformers and data augmentation, should also be integrated into the field. Additionally, breakthroughs in lightweight object detection should be combined with technologies in other related fields, such as hardware device transmission and edge device computing speed, involving cloud computing, edge computing, and distribution system.

4.2.2. Exploration of new application scenarios. Object detection, as a fundamental task in computer vision, has diverse applications. The emergence of YOLO-based lightweight models has extended the scope of potential applications. While some detection scenarios, such as agriculture, traffic, and ships, have been explored, more applications are expected to be discovered. This future direction will make lightweight object detection more valuable and meaningful.

4.2.3. Customized lightweight. As seen in the previous analysis, lightweight object detection tends to be customized and specific. Since achieving high accuracy and efficiency simultaneously is challenging, the task can be separated into several attributes based on user requirements. Detection targets can then be transformed into one or more attributes, such as precision, recall, explanation, generalization, etc. This approach allows for the exploration of more practical functionalities of YOLO-based models for users. This promising future direction aims to solve real object detection problems in various industries while considering resource costs.

5. Conclusion

In this paper, we have made the first effort to understand the development of YOLO-based object detection models from the perspective of lightweight. We comprehensively studied and discussed more than twenty YOLO-based lightweight works. Through this study, we have highlighted the development trend for object detection and the YOLO models. We have also analyzed some interesting findings

discovered during the investigation of lightweight YOLO models. We believe that this study and its findings can uncover the potential philosophy behind model lightweight and further enhance the applicability of object detection in real-world resource-critical scenarios.

References

[1] 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.

[2] Huang R, Pedoeem J, Chen C. YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers[C]//2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018: 2503-2510.

[3] Adarsh P, Rathi P, Kumar M. YOLO v3-Tiny: Object Detection and Recognition using one stage improved model[C]//2020 6th International Conference on Advanced Computing and Communication Systems (ICACCS). IEEE, 2020: 687-694.

[4] Ni Z, Chen J, Sang N, et al. Light YOLO for high-speed gesture recognition[C]//2018 25th IEEE international conference on image processing (ICIP). IEEE, 2018: 3099-3103.

[5] Zhang D, Chen X, Ren Y, et al. Smart-YOLO: A Light-Weight Real-time Object Detection Network[C]//Journal of Physics: Conference Series. IOP Publishing, 2021, 1757(1): 012096.

[6] Jiang J, Fu X, Qin R, et al. High-speed lightweight ship detection algorithm based on YOLO-v4 for three-channels RGB SAR image[J]. Remote Sensing, 2021, 13(10): 1909.

[7] Long Z, Suyuan W E I, Zhongma C U I, et al. Lira-YOLO: a lightweight model for ship detection in radar images[J]. Journal of Systems Engineering and Electronics, 2020, 31(5): 950-956.

[8] Huang R, Pedoeem J, Chen C. YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers[C]//2018 IEEE international conference on big data (big data). IEEE, 2018: 2503-2510.

[9] Tian Y, Yang G, Wang Z, et al. Apple detection during different growth stages in orchards using the improved YOLO-V3 model[J]. Computers and electronics in agriculture, 2019, 157: 417-426.

[10] Huang R, Gu J, Sun X, et al. A rapid recognition method for electronic components based on the improved YOLO-V3 network[J]. Electronics, 2019, 8(8): 825.

[11] Santad T, Silapasupphakornwong P, Choensawat W, et al. Application of YOLO deep learning model for real time abandoned baggage detection[C]//2018 IEEE 7th Global Conference on Consumer Electronics (GCCE). IEEE, 2018: 157-158.

[12] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.

[13] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[14] Bochkovskiy A, Wang C Y, Liao H Y M. Yolov4: Optimal speed and accuracy of object detection[J]. arXiv preprint arXiv:2004.10934, 2020.

[15] Gao X, Ge D, Chen Z. The Research on autopilot system based on lightweight YOLO-V3 target detection algorithm[C]//Journal of Physics: Conference Series. IOP Publishing, 2020, 1486(3): 032028.

[16] Wang G, Feng A, Gu C, et al. YOLO-DFD: A Lightweight Method for Dog Feces Detection Based on Improved YOLO-V4[J]. Journal of Sensors, 2023, 2023.

[17] Hurtik P, Molek V, Hula J, et al. Poly-YOLO: higher speed, more precise detection and instance segmentation for YOLO-V3[J]. Neural Computing and Applications, 2022, 34(10): 8275-8290.

[18] Li W. Infrared image pedestrian detection via YOLO-V3[C]//2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC). IEEE, 2021, 5: 1052-1055.

[19] Zaatouri K, Ezzedine T. A self-adaptive traffic light control system based on YOLO[C]//2018 International Conference on Internet of Things, Embedded Systems and Communications (IINTEC). IEEE, 2018: 16-19.