Review of Object Detection

Jiwei Ye
School of Intelligent Manufacturing, Hebei Polytechnic University of Industry, Shijiazhuang, China

Abstract: Object detection, as a fundamental and critical problem in the field of computer vision, holds significant theoretical research value and demonstrates extensive industrial application value in numerous practical scenarios. This technology has been applied across various domains, including facial recognition, industrial defect detection, and remote sensing detection. Based on a comprehensive review of relevant domestic and international literature, this paper presents a survey of object detection techniques. It begins by outlining the key developmental stages of object detection technology. Subsequently, it discusses aspects such as datasets, evaluation metrics, and performance analysis, classifies existing object detection algorithms, and analyzes and compares the performance of different algorithms on datasets. Finally, the paper provides an outlook on future research directions in object detection.

Keywords: Object Detection; Convolutional Neural Network; Deep Learning

1. Introduction

Object detection is a key technology in the field of computer vision, which involves identifying and locating one or more target objects from images or videos. It not only requires determining whether a specific object exists in the image, but also determining the specific location of these objects. This is achieved by drawing bounding boxes around the objects. The evolution of object detection technology has gone through a process from traditional methods to deep learning, and currently multiple architectures coexist. After 2012, with the rise of deep convolutional neural networks, object detection technology moved from the traditional method era into the deep learning era, with two-stage detectors and single-stage detectors as the main lines. In recent years, in addition to the continuous evolution of CNN, a diversified development pattern has emerged, with research focusing on anchor-free detectors and detectors based on Transformer. Object detection is a research hotspot in the field of computer vision. Every year, a large number of high-quality papers are published (as shown in Figure 1). Many scholars at home and abroad have conducted systematic research on this and have widely applied it in practice. Currently, it has achieved good results in autonomous driving[1], remote sensing detection[2], and medical image analysis[3].

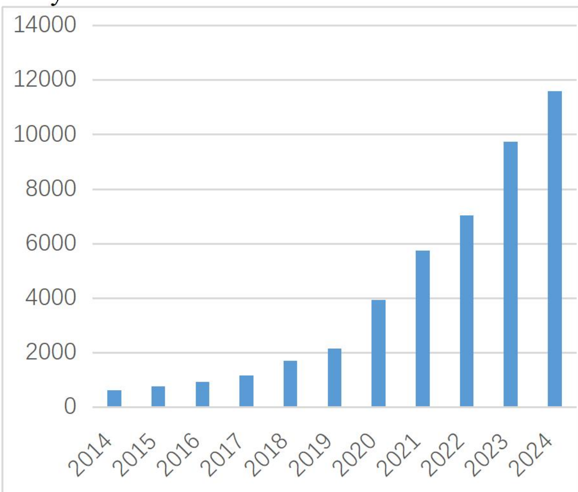

Figure 1. The Number of Papers Related to Object Detection from 2014 to 2024

Data obtained from Google Scholar's advanced search for "object detection" OR "detecting objects"

2. The Key Stages in the Development of Target Detection Technology

2.1 Traditional Object Detection Algorithms

The traditional target detection algorithms consist of three stages: region selection, feature extraction, and classification recognition. The role of region selection is to find candidate regions containing objects. Since the size and position of the target cannot be predicted, traditional methods mainly adopt the exhaustive approach, obtaining a large number of candidate regions by using sliding windows of different scales. The second stage is feature extraction, where methods such as scale-invariant features

[4], Haar features [5], and direction gradient histogram [6] are commonly used to extract feature vectors from the candidate regions. Then, the extracted feature vectors are used as the input of the third stage, and the classifier is used to recognize the features, outputting the recognition result.

The traditional target detection algorithms have the following defects: 1) In the region selection stage, since the size of the sliding window is fixed, if the target has an unconventional shape, the detection performance will drop sharply. Traditional methods mainly adopt the exhaustive approach, thus generating a large number of candidate regions. During the feature extraction and classification recognition stages, the resources of the algorithm will be wasted significantly, reducing the running speed. 2) Feature extraction can only obtain shallow features, which are strongly related to specific tasks. For new tasks, they need to be redesigned and optimized, and the universality is poor, which restricts the detection accuracy. 3) The performance of the classifier is limited by the quality of the input of the first two steps. It highly depends on the researcher's domain knowledge and task design experience, and it is difficult to balance speed and accuracy, making it unable to cope well with variable target detection instances.

2.2 Deep Learning-Based Target Detection Algorithms

During the deep learning era, target detection algorithms mainly follow two main lines: two-stage detectors and single-stage detectors. The representative algorithms of two-stage target detection include R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN. The representative algorithms of single-stage target detection include the YOLO series, SSD series, and RetinaNet.

2.2.1 Two-stage target detection algorithms

(1) R-CNN

R-CNN was proposed by Ross Girshick[7] et al. in 2014. The network structure of R-CNN is shown in Figure 2. In the PASCAL VOC 2012 dataset, the mAP of this algorithm improved from about 35% of the traditional methods to about 62%, showing a significant improvement in performance. First, the candidate regions are obtained using the selective search [8] algorithm, then they are input into a pre-trained CNN model, and a fixed-length feature vector is obtained. Finally, a trained linear support vector machine classifier is used to recognize the feature vector for classification.

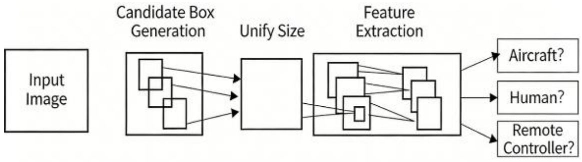

Figure 2. R-CNN Algorithm

The disadvantages of R-CNN:

1) It requires training multiple stages, and each stage is trained independently, resulting in a cumbersome process.

2) Fixed-size transformation of candidate regions will cause image distortion, affecting the accuracy of feature extraction.

3) Each feature vector of the candidate region needs to be written to the disk, occupying a large amount of memory.

(2) Fast R-CNN

Fast R-CNN[9] was proposed by Ross Girshick et al. in 2015. The network structure of Fast R-CNN is shown in Figure 3. This algorithm uses the Region of Interest pooling layer (ROI) to uniformly divide the candidate regions into a fixed number of sub-windows. The algorithm uses a multi-task approach to train the entire network, allowing the two tasks to share the parameters of the convolutional neural network, which can reduce the computational resources required for object detection and improve the detection speed. However, this algorithm still uses selective search in the region selection stage, so it cannot be accelerated using a GPU, and is limited in terms of detection speed and real-time performance.

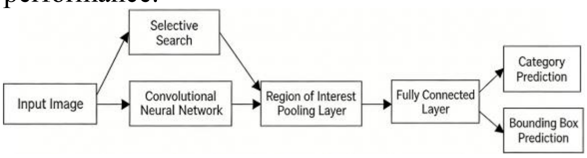

Figure 3. Fast R-CNN Algorithm

(3) Faster R-CNN

Shortly after the proposal of Fast R-CNN, Ren et al. addressed the issue of slow speed in Fast R-CNN due to the use of selective search, and proposed Faster R-CNN[10]. The network structure is shown in Figure 4. This algorithm uses a region prediction network (RPN) instead of the region of interest prediction method, thus enabling acceleration using a GPU, improving the training speed and achieving true end-to-end training. RPN can generate multi-scale anchors, but its detection performance for small targets is not ideal.

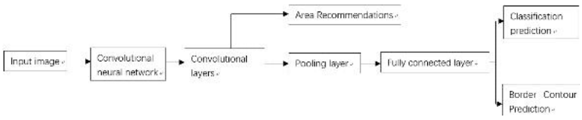

Figure 4. **Faster R-CNN Algorithm**

(4) Mask R-CNN

Mask R-CNN [11] was proposed by He et al. in 2017. Its network structure is shown in Figure 5. This algorithm replaces the ROI Pooling layer in Faster R-CNN with ROI Align. ROI Align uses bilinear interpolation to calculate the floating-point coordinate positions, which can effectively avoid the deviation of feature maps from the original region that occurs during mapping in ROI Pooling, significantly improving the detection accuracy. However, when objects are extremely dense or the target size is very small, the problem of mask loss still occurs. Additionally, due to the addition of an extra branch, the computational cost will increase.

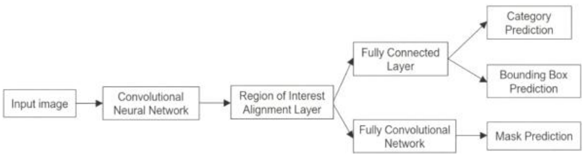

Figure 5. Mask R-CNN Algorithm

2.2.2 Single-stage Object Detection Algorithm

(1) YOLO Series

YOLO was first proposed by Redmon [12] in 2015. It is the earliest single-stage object detection algorithm, redefining object detection as a unified, single regression problem. The network structure is shown in Figure 6. Firstly, the input image is resized to a square size of 448*448 pixels, and then it is input into the convolutional neural network, which is divided into an S*S network, and each grid unit conducts prediction. The network finally outputs a three-dimensional vector. Then, the confidence of each bounding box and the probability of each category are multiplied to obtain the confidence level. Using a threshold, low-probability predictions are filtered out, and through NMS, more accurate results can be obtained.

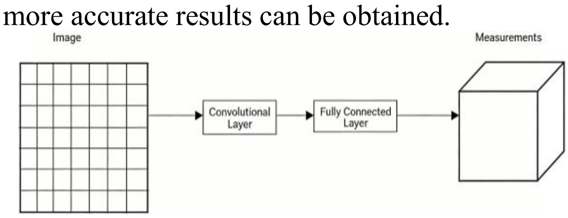

Figure 6. YOLO Algorithm

YOLOv2[13] compared with the previous generation, proposed the Darknet-19 backbone network, using batch high-resolution classifiers and adopting the joint training method, achieving a significant balance between detection speed and accuracy. Its performance surpassed that of the v1 and v2 models, capable of detecting up to 9,000 types of models. Therefore, it was named YOLO9000. However, the anchor boxes of this algorithm still rely on the dataset, and its multi-scale prediction ability is still limited.

YOLOv3 [14] introduced the Darknet-53 backbone network and adopted multi-scale prediction, significantly enhancing the detection ability of small targets. However, the judgment ability for overlapping targets still needs to be improved.

YOLOv4 [15] applied a new backbone network, integrating a large number of training techniques and modules, without increasing the inference cost, significantly improving the accuracy while achieving a substantial improvement in precision. This made YOLOv4 achieve the optimal balance of speed and accuracy in 2020, representing the high maturity of the algorithm in engineering. However, due to its complex architecture and training process, model reproduction and training optimization require high practical experience from researchers.

(2) SSD Series

The SSD[16] algorithm was proposed by Liu et al. in 2016. The network structure is shown in Figure 7. This algorithm combines single-stage detection with multi-scale feature map prediction. It draws on the Anchor idea from Faster R-CNN and sets different scales and aspect ratios for each feature map location to create bounding boxes. This significantly improves the detection effect for small targets. The SDD algorithm achieved the best balance between speed and accuracy at that time, with an mAP of 74.3% on VOC2007 and a frame rate of 58 FPS.

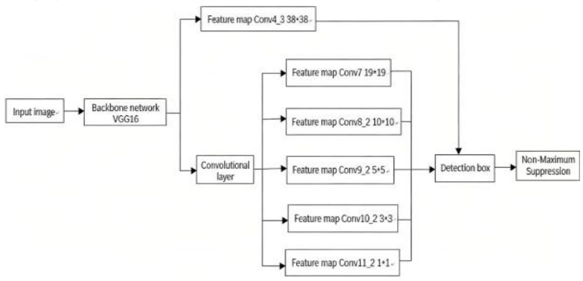

Figure 7. SSD Algorithm

DSSD[17] uses ResNet-101 instead of VGG as the backbone network, and also builds a feature fusion module, enabling deep features to be combined with gradually increasing high-resolution features, significantly improving the detection accuracy for small targets.

However, due to the significant increase in model complexity and computational load, its speed is not as fast as SSD.

FSSD[18] was proposed in 2017 and is a simpler and less computationally expensive feature fusion method compared to DSSD. This algorithm will fuse features of different scales to generate a lightweight feature fusion pyramid, enabling it to trade speed loss for accuracy improvement.

(3) RetinaNet

The structure of RetinaNet[19] is shown in Figure 8. The birth of this algorithm is to solve the problem of how to make a single-stage detector maintain its speed advantage while achieving accuracy even higher than two-stage detectors. This algorithm introduces a focal loss function to enable the model training to focus on difficult-to-classify samples. Using ResNet-101-FPN as the backbone in the COCO dataset, the average precision (AP) reaches 39.1%, and the accuracy is close to that of two-stage detection algorithms.

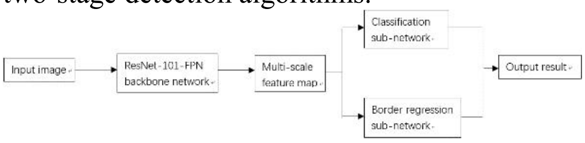

Figure 8. RetinaNet Algorithm

2.3 New Era Object Detection Algorithms

2.3.1 RF-DETR

RF-DETR [20] is a series of real-time object detection models based on the Transformer architecture developed by Roboflow. By combining the latest pre-training techniques with the Transformer framework, it can achieve high detection accuracy and speed. DETR [21] was proposed in 2020, thanks to the self-attention mechanism of the Transformer, which makes the model perform better in handling scenes with occluded objects. DETR matches the predicted boxes with the real boxes using bipartite graph matching loss, avoiding the traditional redundant processing steps and having a more concise structure. However, the original DETR has high computational complexity and slow convergence. To solve this problem, Deformable DETR[22] was proposed, which uses deformable attention modules, focusing on a few key sampling points, which can improve computational efficiency and convergence speed while improving the detection performance for small targets. RF-DETR inherits the design of Deformable DETR and in the COCO benchmark test, RF-DETR is the first real-time detector with an average accuracy over 60%, breaking the performance ceiling of real-time detectors.

2.3.2 RTMDet

RTMDet [23] is an industrial-level real-time detection framework that can perform multiple tasks. It can not only complete the detection of common objects but also extend to fuzzy or rotated object detection. RTMDet is composed of separable convolutions, and this backbone structure can reduce the computational complexity and has compatible capacity between the backbone network and the neck network, achieving a balance between speed and accuracy. On the COCO benchmark, the RTMDet-L model achieved an AP of 52.8% and an inference speed of 300 FPS, surpassing the mainstream industrial detectors of the same period. It has demonstrated strong performance in complex traffic scenarios, industrial defect detection, and small target detection in unmanned aerial vehicle remote sensing.

3. Object Detection Technology: Datasets, Evaluation Metrics, and Performance Analysis

3.1 Datasets

Common datasets used in the object detection field include Pascal VOC, ImageNet, Microsoft COCO, Open Images, etc. From the definition of tasks in Pascal VOC to ImageNet

3.1.1 Pascal VOC

The Pascal VOC [24] dataset was proposed in 2005 and is mainly used for the classification, detection, and segmentation of common objects. This dataset established the evaluation standards and competition platform for the object detection task, defined core evaluation metrics such as mAP, and also gave rise to early network structures such as R-CNN series. Currently, the main ones used are Pascal VOC 2007 and Pascal VOC 2012, where Pascal VOC 2007 contains 11,500 images and 24,640 target objects, and Pascal VOC 2012 contains 17,125 images and 27,450 target objects. This dataset includes four major categories (people, animals, vehicles, and objects) and 20 subcategories.

3.1.2 ImageNet

The ImageNet[25] dataset was jointly established by Stanford University and others in 2009 as a large-scale visual database, mainly used for image classification and detection, hierarchically organized WordNet categories, and contains 1.4

million images, with a detection subset of over 1 million boxes. With massive-scale data, a shift towards the deep learning model has been driven. It provides powerful and universal visual feature extraction capabilities.

3.1.3 Microsoft COCO

The Microsoft COCO[26] dataset was released by Microsoft in 2014 and is mainly used for detecting and segmenting targets in complex scenes with small and dense objects. It contains 118,287 training sets, 5,000 validation sets, and 40,670 test sets. It is currently the most mainstream benchmark, enhancing the model's ability to handle complex scenes.

3.1.4 Open Images

Open Images[27] was launched by Google in 2016 and is used for detecting, classifying, and describing visual relationships. This dataset contains 600 detection categories, 900 images, and 16 million hierarchical labels and relationship annotations. Its protocol is open, which greatly promotes industrial and academic applications, and builds a powerful visual base model.

3.2 Evaluation Metrics

3.2.1 IoU

Intersection over Union (IoU) is a measure of the overlap between the predicted bounding box and the real bounding box. A IoU threshold is usually set to evaluate the prediction results. COCO and other datasets usually use the average value of multiple IoU thresholds to assess the localization accuracy, resulting in more rigorous results.

Calculation method:

$$\mathrm{IoU}=\frac{\mathit{Area\ of\ Overlap}}{\mathit{Area\ of\ Union}} \tag{1}$$

3.2.2 AP/mAP

Average Precision (AP) is a standard for evaluating detection results. Its calculation formula is:

$$\mathrm{A}_{\mathrm{AP}}=\int_0^1 P(t)dt \tag{2}$$

Where $P(t)$ represents the precision at intersection over union threshold t. Mean Average Precision (mAP) is the average of all category APs, which can better complete multi-class target detection tasks. The calculation method is:

$$\mathrm{m}_{\mathrm{mAP}}=\frac{\sum_{n=1}^{N} A_{AP}^{n}}{N} \tag{3}$$

Here, N represents the number of object categories, and A_{AP}^{n} represents the evaluation accuracy rate of measuring the objects.

3.2.3 FPS

FPS is the core metric for measuring the real-time processing speed of the target detection algorithm. The calculation formula is:

$$\mathrm{FPS}=\frac{N}{T_{end}-T_{start}} \tag{4}$$

Here, N represents the number of image frames successfully processed and output within a unit time, and T_{start} refers to the time when the first frame is started to be processed, while T_{end} refers to the time when the last frame is processed and the result is output.

3.3 Performance Analysis

As shown in Table 1, the performance of representative object detection algorithms was compared. Among them, R-CNN, Fast R-CNN, Faster R-CNN, and Mask R-CNN are two-stage object detection algorithms, while YOLO, YOLOv2, YOLOv3, YOLOv4, SSD, DSSD, FSSD, and RetinaNet are single-stage object detection algorithms.

From the data in Table 1, the following conclusions can be drawn: 1) The single-stage detection algorithm uses a more complex backbone network, and its accuracy is higher than that of the two-stage detection algorithm, indicating that the design of a complex backbone network is helpful for improving accuracy. 2) The algorithms based on the more complex Microsoft COCO dataset in the table cannot be directly compared with other algorithms, indicating that the benchmark of COCO is higher and more challenging.

4. Application Areas of Object Detection

4.1 Face Recognition

Face recognition is a very common application area of object detection technology. Face recognition has already deeply integrated into various aspects of people's daily lives, thus attracting a lot of attention. The difficulties encountered in face recognition include the difficulty in extracting features in strong light or low light conditions, as well as the loss of key information due to wearing masks or sunglasses, and the protection of user privacy. To solve these problems, Shaik[28] et al. proposed using autoencoders and GANs to generate training data to improve the accuracy of detection. At the same time, in the aspect of protecting user privacy during the recognition process, Sun[29] et al. adopted homomorphic encryption in the inference stage, using synthetic data instead of real data, which can ensure the normal operation of the recognition system while protecting personal privacy.

4.2 Industrial Product Detection

Industrial product detection is one of the fields with strict requirements for object detection technology. Due to the scarcity of defect samples and the diverse shapes of defects, the complex environment of the production line may affect the extraction of features. To strike a balance between accuracy and real - time performance of the model, Zhang et al. [30]proposed a lightweight module and a multi - branch feature fusion method, which improved the measurement accuracy on the product defect dataset.

Table 1. Performance Evaluation of Object Detection Algorithms

Algorithm	Backbone network	Test dataset	mAP%
R-CNN	AlexNet	Pascal VOC 2007	58.5
Fast R-CNN	VGG-16	Pascal VOC 2007	70.0
Faster R-CNN	ResNet-101	Pascal VOC 2007	73.2
Mask R-CNN	ResNet+FPN	Microsoft COCO	39.8
YOLO	VGG-16	Pascal VOC 2007	63.4
YOLOv2	Darknet-19	Pascal VOC 2007	76.8
YOLOv3	Darknet-53	Pascal VOC 2007	79.6
YOLOv4	CSPDarknet-53	Pascal VOC 2007	87.2
SSD	VGG-16	Pascal VOC 2007	77.2
DSSD	ResNet-101	Pascal VOC 2007	78.6
FSSD	VGG-16	Pascal VOC 2007	78.8
RetinaNet	ResNet-101	Microsoft COCO	34.4

4.3 Remote Sensing Image Detection

The application of remote sensing detection technology is extensive, such as aerospace, ports, military, etc. However, remote sensing image detection faces many problems, such as: the orientations of vehicles are diverse, which is not conducive to positioning; for the target to be detected, the scale differences can be large, which requires a high multi-scale requirement for the model; the detection positions have complex terrain and many obstructions, which can easily cause feature confusion. To solve these problems, Li[31] et al. proposed using deformable convolution, enabling the network to adapt to the shape and scale of the target, and using 3D convolution to fuse multi-scale features, which can improve the detection accuracy for rotating targets.

5. Future Research Directions of Object Detection

The research on object detection is developing towards more efficient, intelligent, and universal directions. The future research directions are summarized as follows:

1) Currently, the algorithms of object detection are increasingly balancing accuracy and speed. In the future, new architectures can be designed to ensure both accuracy and speed while improving the generalization of object detection technology, enabling it to be better applied in daily life.

2) Gradually moving towards global and continuous learning, improving the self-learning ability of the model, making it more intelligent, and enabling it to detect new categories even with only a few samples or textual descriptions, better adapting to environmental changes.

3) Improving the reliability and security of the detection system, making the decision-making of the monitoring system more transparent. Using more robust training methods and defense mechanisms, enhancing the defense mechanism of the model when attacked, and ensuring the safe operation of the monitoring system.

6. Conclusion

Object detection is a fundamental task in computer vision and has become a popular field in computer vision at present. It still has a broad future prospect in the future. This paper arranges the different categories of object detection algorithms in chronological order, summarizes the mechanisms of different algorithms, and focuses on conducting comparative analysis of their performance. On this basis, it also looks forward to the future research directions of object detection technology.

References

[1] Pravallika, M. F. Hashmi and A. Gupta, "Deep Learning Frontiers in 3D Object Detection: A Comprehensive Review for Autonomous Driving," in IEEE Access, vol. 12, pp. 173936-173980, 2024.

[2] Gui, S., Song, S., Qin, R., & Tang, Y. (2024). Remote sensing object detection in the deep learning era—a review. Remote Sensing, 16(2), 327.

[3] Bakhtiyorov, Sanjar, et al. "Real-Time Object Detector for Medical Diagnostics

(RTMDet): A High-Performance Deep Learning Model for Brain Tumor Diagnosis." Bioengineering 12.3 (2025): 274.
[4] Lowe, David G. "Distinctive image features from scale-invariant keypoints." International journal of computer vision 60.2 (2004): 91-110.
[5] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001. Vol. 1. Ieee, 2001.
[6] Dalal, Navneet, and Bill Triggs. "Histograms of oriented gradients for human detection." 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). Vol. 1. Ieee, 2005.
[7] 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.
[8] Uijlings J R R, Van De Sande K E A, Gevers T, et al. Selective search for object recognition[J]. International journal of computer vision, 2013, 104(2): 154-171.
[9] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.
[10] Faster, R. C. N. N. (2015). Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 9199(10.5555), 2969239-2969250.
[11] He K, Gkioxari G, Dollár P, et al. Mask r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2961-2969.
[12] Redmon, Joseph, et al. "You only look once: Unified, real-time object detection."Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[13] Redmon J, Farhadi A. YOLO9000: better, faster, stronger[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 7263-7271.
[14] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.
[15] Bochkovskiy, Alexey, Chien-Yao Wang, and Hong-Yuan Mark Liao. "Yolov4: Optimal speed and accuracy of object detection." arXiv preprint arXiv:2004.10934 (2020).
[16] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Cham: Springer International Publishing, 2016: 21-37.
[17] Fu C Y, Liu W, Ranga A, et al. Dssd: Deconvolutional single shot detector[J]. arXiv preprint arXiv:1701.06659, 2017.
[18] Li Z, Yang L, Zhou F. FSSD: feature fusion single shot multibox detector[J]. arXiv preprint arXiv:1712.00960, 2017.
[19] Lin, Tsung-Yi, et al. "Focal loss for dense object detection." Proceedings of the IEEE international conference on computer vision. 2017.
[20] Robinson, Isaac, et al. "RF-DETR: neural architecture search for real-time detection transformers." arXiv preprint arXiv:2511.09554 (2025).
[21] Carion, Nicolas, et al. "End-to-end object detection with transformers." European conference on computer vision. Cham: Springer International Publishing, 2020.
[22] Zhu, Xizhou, et al. "Deformable detr: Deformable transformers for end-to-end object detection." arXiv preprint arXiv:2010.04159 (2020).
[23] Lyu, Chengqi, et al. "Rtmdet: An empirical study of designing real-time object detectors." arXiv preprint arXiv:2212.07784 (2022).
[24] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (voc) challenge[J]. International journal of computer vision, 2010, 88(2): 303-338.
[25] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.
[26] Lin, TY. et al. (2014). Microsoft COCO: Common Objects in Context. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds) Computer Vision – ECCV 2014. ECCV 2014. Lecture Notes in Computer Science, vol 8693. Springer, Cham. https://doi.org/10.1007/978-3-319-10602-1_48
[27] Kuznetsova, A., Rom, H., Alldrin, N. et al. The Open Images Dataset V4. Int J Comput

Vis 128, 1956–1981 (2020). https://doi.org/10.1007/s11263-020-01316-z
[28] Shaik A N, Villarini B, Argyriou V. A Deep Learning Approach for Facial Attribute Manipulation and Reconstruction in Surveillance and Reconnaissance[J]. arXiv preprint arXiv:2506.06578, 2025.
[29] Sun Z, Liu Z. Ensuring privacy in face recognition: a survey on data generation, inference and storage[J]. Discover Applied Sciences, 2025, 7(5): 441.
[30] Zhang L, Wang Z, Ma Y, et al. Steel surface defect detection algorithm based on improved YOLOv10[J]. Scientific Reports, 2025, 15(1): 32827.
[31] Li Q B, Fan Z C, Zhao X Y. An advanced adaptive detector for oriented objects in remote sensing imagery[J]. Scientific Reports, 2025, 15(1): 33877.
