Mechatronics and Automation Technology
J. Xu (Ed.)
© 2022 The authors and IOS Press.
This article is published online with Open Access by IOS Press and distributed under the terms of the Creative Commons Attribution Non-Commercial License 4.0 (CC BY-NC 4.0).
doi:10.3233/ATDE221146

Image Recognition Technology and Development of YOLO Algorithm

Chenbo WU, Qiang LUO[1] and Derong SHOU
Chongqing Three Gorges University, China

Abstract. With the rapid development of image recognition technology, it has its applications in medical, security and other fields, but this technology has many areas to be improved, such as the accuracy of recognition, real-time and other issues, which have always been a hot research topic in this field .This article involves the development process of image recognition (traditional image processing, machine learning, deep learning), and focuses on the advantages and disadvantages of the popular YOLO algorithm in image recognition. Finally, it describes the technical problems faced by the current target detection technology and the corresponding solutions.

Keywords. Target detection, deep learning, YOLO algorithm.

1. Introduction

Nowadays, with the rapid development of target detection technology, many people study its algorithm function and optimize its recognition accuracy, real-time performance and stability from different angles. Understanding its development history is conducive to better learning this technology, this paper introduces its development history, and the YOLO algorithm, so that we can have a more comprehensive understanding of this technology, get inspiration from it, and learn better.

2. Traditional Image Processing

Target detection technology is a hot field of scientific research today, and its development has experienced a stage from traditional digital image processing to machine learning and then to deep learning[1]. Each stage has its own unique detection method and typical algorithm, its development process is shown in figure 1.

[1] Corresponding Author, Qiang Luo, Chong qing Three Gorges University, Chongqing, China; E-mail :qluosxxy@bjtu.edu.cn

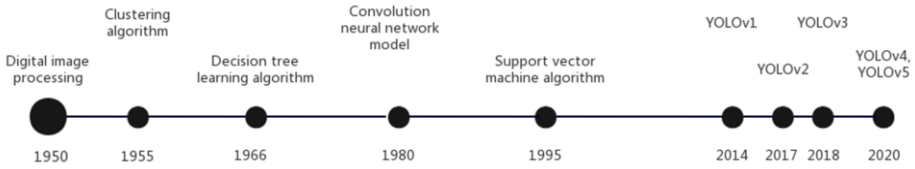

Figure 1. Development History of typical algorithms for Target Detection Technology

In the 1960s, the traditional digital image processing technology was born as a discipline. In the early days, people only used it for simple image processing in order to facilitate study the image. After that, this technology became familiar to people, and it began to infiltrate other fields, such as medical X-ray imaging. After research and development, the traditional image processing constructed a digital human vision system, which is called image understanding or computer vision, and was applied to the geographic information system. Later, the development of wavelet theory led to the development of image processing technology, so this traditional technology began to develop rapidly[2]. During the development of traditional image technology at that time, many problems were encountered, among which the most important problems are the hardware operation and storage capacity of the computer, the software information processing technology and how to make the expression of the computer convenient for human observation. At that time, the computer hardware was difficult to deal with large image information, and the software development technology was not mature, which made this technology less functional and high cost. Digital image processing mainly realizes its function in the following ways and forms: digitizing the image to facilitate image follow-up operation by sampling and quantifying[3]; Through coding, decoding and other methods to achieve image compression to reduce image storage space without affecting human observation as much as possible; computer uses mathematical calculation to achieve image enhancement and restoration to make it more efficient in recognition.; image segmentation of special areas for next image processing and analysis; description of image volume and outline; image recognition function which is enlightening for subsequent machine learning.

At present, there is an attempt to apply the image recognition function of digital image processing technology to agricultural picking robot. Its image recognition function is mainly based on the difference of color, shape and texture. For some fruits with bright colors and high recognition, fruit recognition can be carried out based on color features, such as double Otsu segmentation algorithm, which effectively reduces the recognition time, but its disadvantage is that it requires high illumination, and its recognition efficiency is greatly reduced in the case of night or uneven illumination. Shape-based image recognition, typical boundary feature method, geometric parameter method and so on, illumination has little influence on this recognition method, but when there is branch occlusion, the recognition efficiency of this method will be reduced; texture-based recognition method can avoid the above two common interference, but texture recognition method needs to be within a threshold range to be effective, beyond this range will not be able to identify. As a result, there are a variety of comprehensive methods of recognition, in order to deal with different fruits, different picking environments, and improve the recognition accuracy[4].The comparison of processing technologies is shown in table 1.

Table 1. Comparison of traditional digital image processing techniques

Processing method	Advantage	Limit	Suitable goal
Based on color	Can efficiently distinguish brightly colored recognition objects.	Not suitable for environments where the light is weak or uneven.	Citrus, apple
Based on shape	The requirement of lighting condition is not high, and the target contour can be extracted.	It is not suitable for identifying objects with serious branch and leaf occlusion.	Eggplant, tomato
Based on texture	It can effectively distinguish between recognition object and background.	Weak light or serious occlusion will affect the recognition efficiency.	Dragon fruit, balsam pear
Based on color, shape and texture	Make up for the limitations of a single treatment method	Different convergence priorities need to be set for different objects and environments	Oranges, mangoes

In the future, digital image processing will develop in the direction of high speed, high resolution, intelligence and standardization, realizing the reconstruction of two-dimensional image into three-dimensional image, development chip used in robot vision, optimization algorithm and so on[5].

3. Machine Learning

With the development of artificial intelligence driven by the level of science and technology, machine learning has stepped onto the stage of history . Machine learning is to achieve some functions by making computers learn knowledge like human beings, summarize laws, and constantly improve and perfect the mode of operation of computers[6]. The development of machine learning can be traced back to the 1960s, when the research methods only focused on the goal results, not the process of realization. By modifying the system parameters and the execution instructions of the machine, let the computer find the process method to achieve the goal by itself. This way is the machine learning in the period of enlightenment, but the effect is not satisfactory. Around the 1970s, computers were allowed to imitate human learning, human thinking and logical thinking, induction and summary, so as to complete the learning tasks. at this time, it was considered that attention should be paid to the realization process, and then around the 1980s, people put forward a variety of different and multifaceted learning methods, synthesizing the characteristics of different functions, so as to make the computer more accurately close to the learning goal and combine with reality. As a result, the effect of machine learning has been significantly improved, and machine learning has officially risen[7]. Today, the way of learning is ever-changing, he has been divided into several categories , The main types of machine learning are shown in figure 2.

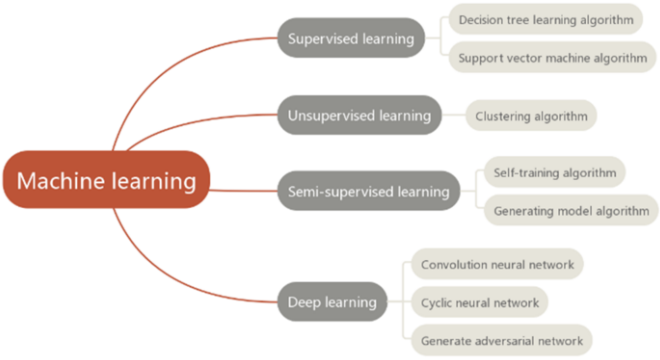

Figure 2. Major machine learning classifications

Supervised learning, giving the computer a set of marked training data to make the computer infer the functions it should achieve, the common are decision tree learning algorithm, support vector machine algorithm; unsupervised learning, which gives a set of unmarked training data, inputs the data directly to the computer, and lets it summarize and form a model to achieve the training purpose. This method is generally used in areas where manual tagging is difficult or expensive, and the typical algorithm is clustering algorithm; semi-supervised learning, which is a combination of supervised learning and unsupervised learning, uses a set of training data (most of which are unlabeled and only a few are labeled) for pattern recognition. It has been paid attention to by people because of its high efficiency and high accuracy, such as self-training algorithm and model generation algorithm[8]; deep learning, which is a category of pattern analysis methods, through multi-layer processing, combining "low-level" feature representation to transform it into "high-level" feature representation, to achieve a simple model to complete complex tasks. It makes full use of feature learning and transforms features from the original space to a new space through layer-by-layer feature transformation, which increases the accuracy of classification and prediction[9].Convolution neural network model and deep trust network model are typical deep learning models.

4. Deep Learning

Some companies have developed and used certain functions on the basis of deep learning, such as computer vision, speech recognition and so on. The development of deep learning plays a very important role in the progress of image recognition technology. Deep learning can be divided into two types of target detection:2-stage target detection and 1-stage target detection. The typical 2-stage target detection algorithm is RCNN series algorithm, the two steps of inputting images and generating target suggestion area and sending classifier classification are completed by different networks, his detection accuracy is high but slow; 1-stage target detection typical algorithm is YOLO series algorithm, its input image and output detection frame and classification label are completed by the same network, compared with 2-stage, it has lower detection accuracy but faster speed. Combined with the actual situation, the latter is also more used in industry.

5. Overview of YOLO Algorithm

In 1-stage target detection, YOLO algorithm has great advantages compared with other algorithms, and has been used in production and life. It has received great attention and research, and now it has evolved from yolov1 algorithm to yolov5 algorithm. The image to be processed by Yolov1 algorithm is divided into several rectangular regions averagely. If the center position of an object to be tested falls on this rectangular region, the rectangular region is responsible for predicting the object to be tested. Each region predicts B bounding box (including the center position, height, width and confidence of the predicted object), and predicts the type of object. Confidence indicates the accuracy of predicting object information, which is expressed by formula (1):

$$C=\Pr(obj)*IOU_{truth}^{pred} \tag{1}$$

If the object to be tested falls in a rectangular region, the first term of the formula takes 1, otherwise it takes 0, and the second term IOU is the union of the intersection ratio of the predicted bounding box and the real object. By multiplying the information predicted by the rectangular region with the information predicted by bounding box, the accuracy of the predicted object belonging to a certain class is obtained, and then a threshold is designed to clear the result below the threshold, and the rest is the target detection result. The loss function of Yolov1 is expressed by formula (2):

$$\begin{aligned} loss=&\lambda_{coord}\sum_{i=0}^{s^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[(x_i-\bar{x}_i)^2+(y_i-\bar{y}_i)^2\right]+\lambda_{coord}\sum_{i=0}^{s^2}\sum_{j=0}^{B}1_{ij}^{obj}\left[(\sqrt{w_i}-\right.\\ &\left.\sqrt{\overline{w_i}})^2+(\sqrt{h_i}-\sqrt{\overline{h_i}})^2\right]+\sum_{i=0}^{s^2}\sum_{j=0}^{B}1_{ij}^{obj}(C_i-\overline{C_i})^2+\lambda_{noobj}\sum_{i=0}^{s^2}\sum_{j=0}^{B}1_{ij}^{noobj}(C_i-\\ &\overline{C_i})^2+\sum_{i=0}^{s^2}1_{i}^{obj}\sum_{c\in classes}(p_i(c)-\overline{p_i}(c))^2 \end{aligned} \tag{2}$$

Among them 1_{ij}^{obj} Judge whether there is an object in the rectangular area, there is an object that is equal to 1, and there is no object equal to 0.1_{ij}^{noobj} and 1_{ij}^{obj} have the opposite rule of value.The first term of the formula is the loss function of the central coordinates, and the second term is the square root mean square error of height and width, which makes both large and small sizes equally sensitive to detection. the latter three items are to judge whether the rectangular area contains the object to be tested and the type of the object to be tested.λ_{coord}=5， λ_{noobj}=0.5,This value increases the stability of the model.Yolov1 algorithm has strong versatility and can detect most kinds of objects, but each rectangular region can only detect one object, which is not conducive to detect the dense model of target objects, and the error rate of target location is also high.

Yolov2 is a big improvement over yolov1 and extends to detect 9000 different objects, also known as yolo9000.The improvement is:(1)Batch Normalization. The batch normalization operation of the samples entering the neural network can solve the problem of gradient disappearance and gradient explosion in back propagation, and better convergence speed and effect can be obtained.(2)High resolution classifier, is trained with 224*224 images, fine-tuned with 448*448 high-resolution samples, and then trained to reduce the impact of sudden changes in resolution.(3)Convolution with anchor boxes, which draws lessons from the method of Faster R-CNN to predict the offset value and confidence of Anchor Box, does not predict coordinate values directly,

which improves the learning efficiency of neural network and greatly improves recall.(4)Dimension clusters, K-mean clustering analysis is carried out on the frames marked in the training set to find the prior frames that better match the sample size.(5)Direct location prediction adjusts the prediction formula to determine the center of the prediction frame in a specific rectangular area, which makes the network more stable. (6)Fine-Grained Features, and retains the details in the feature graph, so that the detail features of smaller objects can be retained.(7)Multi-Scale Training, so that yolov2 can be compatible with different sizes of pictures for recognition. These improvements make yolo algorithm more competitive than other image recognition algorithms.

Yolov3 is more complex than the previous model, and the detection speed and accuracy can be balanced by changing the structure of the model. Its main improvement has two points: (1)The improvement of the network structure refers to the practice of ResNet, modifies the last layer and the penultimate layer of ResNet, optimizes the network output, and makes its detection speed become very fast.(2)Multi-scale training makes the algorithm more advantageous to small object detection. On the basis of yolov3, yolov4 optimizes and improves various parameters of yolov3, constantly adjusts various data parameters, constantly tries data combinations of different parameters, finds the best matching parameters, and improves its performance in all aspects. The advantage of Yolov5 is that when all aspects of performance are not inferior to yolov4, its model construction is very small, nearly 90 percent smaller . The summary of algorithm upgrade and optimization is shown in figure 3.

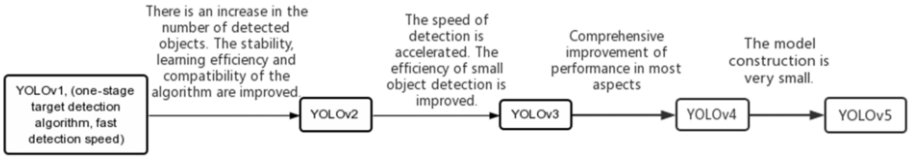

Figure 3. Upgrade part of YOLO algorithm

6. Problems and Prospect of Target Detection Technology

Nowadays, with the rapid development of image recognition technology, in addition to YOLO algorithm, there are also R-CNN, SPP-NET, Fast R-CNN and other algorithms, they have their own advantages and disadvantages, according to different occasions, choose the appropriate algorithm to use, the current deep learning wants to make a further development breakthrough, mainly from the following aspects: (1) improve the theoretical model, deep learning in China for a short time. Many theoretical models are not complete enough[10].(2) optimize the parameters of the balance model, which requires researchers to constantly experiment, summarize and innovate, and improve the performance of the algorithm. (3) optimize the training data set, which affects the efficiency of deep learning, and improves the performance of the algorithm through optimization. (4) improve the robustness of the system and reduce the risk of system collapse in use. (5) optimize the reduction of accuracy caused by the transformation of viewing angle, different proportion of object size, transformation of light and shadow, complex environment, occlusion and so on.

Where there is a problem, there is a direction to solve the problem, these problems

show that the target detection technology still has a lot of room for growth, its potential remains to be developed, and this technology will be better used in various fields in the future.

Acknowledgments

This work was supported by the Artificial Intelligence + Intelligent Agricultural subject Group Open Fund Project(Study on the Prediction of the recognition and location of Ancient Red Citrus in Natural Environment)(ZNNYKFB201907)

References

[1] Zhang Q, Zhang R M and Chen B. A Survey of Image recognition Technology based on Deep Learning [J], Journal of Hebei Academy of Sciences. 2019, 36(3): 28-36.
[2] Ji SL, et,. A Summary of Research on robustness of Deep Learning Model [J], Computer journals. 2022, 45(1): 190-206.
[3] Chen H Q, Wan Y L and Wang G G. Research Progress of Digital Image processing Technology [J], Industrial control computer. 2013, 26(1): 72-74.
[4] Zheng T X, Jiang M Z and Feng M C. A Survey of Target recognition and Localization methods for picking Robot based on Vision. Journal of instrumentation, 2021, 41(9): 28-51.
[5] Yang X J, Zhang F D and Hu C B, A Survey of Machine Learning[J], Scientific and technological economic market. 2021(10): 3.
[6] Yu K, Jia L, Chen Y Q, etc.Yesterday, today and tomorrow of deep learning[C], Computer research and development. CNKI; WanFang, 2013: 6.
[7] Meng Z L, Li T L. Summary and Prospect of the Development of Machine Learning Technology, Integrated circuit application. 2020. 37(10): 56-57.
[8] Chen K and Zhu Y. A Survey of Machine Learning and its related algorithms, Statistics and Information Forum, 2007, 22(5): 105-112.
[9] Hu Y, etc. Summary and discussion on Deep Learning [J], Journal of Intelligent Systems. 2019, 14(1): 1-19.
[10] Chen B Q, Liu H L and Meng F B. The present situation and Development Direction of Digital Image processing Technology[J], Journal of Jishou University (Natural Science Edition), 2009, 30(01): 63-70.