BLOOD CELLS CLASSIFICATION USING EMBEDDED MACHINE LEARNING

ZHANG ZIMU

DEPARTMENT OF ELECTRICAL ENGINEERING FACULTY OF ENGINEERING UNIVERSITY OF MALAYA KUALA LUMPUR

2021

BLOOD CELLS CLASSIFICATION USING EMBEDDED MACHINE LEARNING

ZHANG ZIMU

DISSERTATION SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF INDUSTRIAL ELECTRONIC AND CONTROL ENGINEERING

DEPARTMENT OF ELECTRICAL ENGINEERING FACULTY OF ENGINEERING UNIVERSITY OF MALAYA KUALA LUMPUR

2021

UNIVERSITY OF MALAYA ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: ZHANG ZIMU

Matric No: 17220681

Name of Degree: Master Of Industrial Electronic And Control Engineering Title of Project Paper/Research Report/Dissertation/Thesis ("this Work"):

Field of Study:embedded machine learning

I do solemnly and sincerely declare that:

- (1) I am the sole author/writer of this Work;
- (2) This Work is original;
- (3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
- (4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
- (5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
- (6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other action as may be determined by UM.

Candidate's Signature

Date:2021/10/1

Subscribed and solemnly declared before,

Witness's Signature

Date:

Name: Designation:

[BLOOD CELLS CLASSIFICATION USING EMBEDDED MACHINE LEARNING] ABSTRACT

With the development of science and technology, digital image processing has been applied to various fields, especially playing an important role in medicine. This thesis mainly studies the identification of blood cells in complex situations, and proposes a YOLOv3 target detection method. The ResNet network is used to optimize the Darknet-53 feature extraction structure of YOLOv3, and the feature pyramid network is used to obtain the four scale features of the target to fuse the shallow features and deep feature information. Then adjust the influence weight of the loss function according to the size of the detected target, so as to enhance the detection effect of small targets and mutual occluded objects. The experimental results on the data set show that the detection accuracy of the YOLOv3 method can reach 83.74%, and made a graphical interface with Python QT5.

Keywords: YOLOv3, residual network, Darknet-53, cross entropy loss function

ACKNOWLEDGEMENTS

Along the way of completing this research, there's a people who helping me to work this thesis. Firstly, I am grateful to have my supervisor, Dr. Mohamad Sofian bin Abu Talip for all his great effort in guiding and advising me throughout the whole research. He also organized the research guideline for me so that I am able to complete the research in time.

I also would like to gratefully acknowledge the support of lecturers of Department of Electrical Engineering, University of Malaya for handling good of any issues raised by the final year students especially regarding to the final year project.

I thank the Faculty Engineering Electrical Department for providing and arranging some helpful courses and seminar that really useful in the thesis writing especially the End Note program and simulation test.

Most importantly, I would like to gratitude my whole family for keep supporting and encourage me, despite my own limited devotion.

Lastly, I would like to thank and offer my regards to the people who support me in any respect during the completion of the work.

TABLE OF CONTENTS

[BLOOD CELLS CLASSIFICATION USING EMBEDDED MACHINE LEARNING]
Abstractiii
Acknowledgementsiv
Table of Contents
List of Figures
CHAPTER 1: INTRODUCTION1
1.1 RESEARCH BACKGROUND1
1.2 RESEARCH STATUS AND DEVELOPMENT TREND
1.3 RESEARCH CONTENT
1.4 ORGANIZATIONAL STRUCTURE
CHAPTER 2: LITERATURE REVIEW
2.1 CONVOLUTIONAL NEURAL NETWORK ARCHITECTURE
2.1.1 Input layer7
2.1.2 Convolutional layer
2.1.3 Downsampling layer
2.1.4 Activation function
2.1.5 Loss layer function
2.2 FEATURE LEARNING PRINCIPLE OF CONVOLUTIONAL NEURAL
NETWORK
2.3 INTRODUCTION TO YOLOV3
2.4 YOLOV3 NETWORK STRUCTURE
2.5 YOLOV3 DETECTION PROCESS
2.6 ADVANTAGES AND DISADVANTAGES OF YOLOV3 ALGORITHM 20
CHAPTER 3: RESEARCH METHODOLOGY22

3.1 DATA SET SELECTION	
3.2 CALIBRATION OF THE DATA SET	
3.3 MODEL EVALUATION INDEX	
3.4 SUMMARY	
CHAPTER 4: RESULTS AND ANALYSIS	
4.1 YOLO ALGORITHM	
4.2 FEATURE EXTRACTION NETWORK DARKNET-53	
4.3 TARGET DETECTION NETWORK STRUCTURE	
4.4 LOSS FUNCTION	
4.5 RESULTS AND ANALYSIS	
4.6 GRAPHICAL INTERFACE	
CHAPTER 5: CONCLUSION AND OUTLOOK	
References	41

LIST OF FIGURES

Figure 2.1 Schematic diagram of the convolution process	9
Figure 2.2 Comparison of normal connection and sparse connection	11
Figure 2.3 Schematic diagram of weight sharing	11
Figure 2.4 Schematic diagram of the convolution process	16
Figure 2.5 YOLOv3 detection process	20
Figure 3.1 Image of the dataset used in this project	23
Figure 3.2 LabelImg interface	24
Figure 3.3 Interface after marking	24
Figure 4.1 darknet-53	. 29
Figure 4.2 Comparsion of backbones	30
Figure 4.3 yolo v3 network Architecture	32
Figure 4.4 Training data	. 34
Figure 4.5 The precision value and recall	35
Figure 4.6 The F-score value	36
Figure 4.7 The mAP value	37
Figure 4.8 Graphical interface	38

CHAPTER 1: INTRODUCTION

1.1 Research background

Since the advent of computers in the last century, computer technology has developed rapidly. Computer image processing and computer pattern recognition technology are one of them. They are not only used in industrial and agricultural production, aerospace exploration, geological exploration and other fields, but also have been widely used. In various medical diagnosis and biological tests, clinical medical diagnosis is more scientific and standardized. Blood cell morphology examination is of great value for clinical diagnosis, treatment and prognosis judgment. Hematology experts generally call for attention to peripheral blood cell morphology. Detection.

In the research of basic medicine, researchers have used computer technology to research and develop a variety of medical testing equipment, such as high-end equipment such as electrocardiogram and electroencephalogram, to monitor the activities of people's heart, brain, etc., to understand and diagnose The health of the heart and brain, along with the development of computer technology, a variety of medical testing instruments have emerged, and they have shown their skills in clinical medical diagnosis. In various clinical medical diagnostics and biotechnological tests, hematology tests are characterized by a large number and a wide range of aspects. In particular, blood cell counts play an absolute main role in clinical laboratory work: for example, various types of bacteria. Or the diagnosis of viral infections, the diagnosis of various leukemias and anemias, the inspection, screening, and judgment of the side effects of various drugs on human blood cells, the control of the degree of inhibition of the beneficial and harmful cells in the bone marrow tissue by radiotherapy and chemotherapy, etc. Wait and see.

The clinical and pathological routine examinations of many blood diseases or bloodrelated diseases are judged and diagnosed by observing and calculating the volume, surface area, number of blood cells per unit volume, and the morphology of cells and nuclei of blood cells. Disease . In these medical diagnosis and judgments, computer image analysis has become a powerful tool for quantitative analysis and research. It has been used in the identification and classification of blood cells, the count of various cells, and the identification of tumor cells. This greatly reduces the labor intensity of medical testing personnel, and provides medical personnel with quantitative and objective data. However, in order for computer image analysis to obtain better recognition results in clinical medical diagnosis, it must be accurate and efficient. In blood cell image recognition, the key to high accuracy of blood cell image recognition lies in the segmentation and edge of blood cells. Extraction, which involves automatic morphological analysis of blood cells.

The automatic morphological analysis of blood cells is an important development direction of modern biological and clinical medical inspection technology, and it is also an important application direction of computer digital image analysis and pattern recognition technology in the field of clinical medicine. When people are unwell and go to the hospital for diagnosis and treatment, they are used most. The most common is the analysis and testing of various blood. At present, various blood cell analyzers are commonly used in hospitals. These blood cell analyzers can not only analyze and detect multiple parameters such as white blood cells in blood cells, but also The analysis of blood cell data is accurate, easy to operate, and basically fully automated. This greatly liberates the productivity of medical testing personnel, and also greatly facilitates the clinical diagnosis of doctors. Therefore, various blood cell analyzers have become various types of hospitals. One of the most widely used medical testing equipment. Therefore, compared with traditional manual analysis methods, computer image analysis and processing technology not only greatly liberates the productivity of medical examiners and improves the efficiency of medical diagnosis, the analysis and detection of blood cells by computer image technology is more objective and more accurate. I believe that with the continuous development of computer pattern recognition technology and artificial intelligence research, people's requirements for computer image analysis and processing technology will become higher and higher. The requirements are simple, accurate, and qualitative, quantitative, and efficient, making computer image analysis Processing technology has become a powerful tool for blood cell identification and analysis.

However, in recent years, people have been dissatisfied with the accuracy, statistical, quantitative and efficient characteristics of computer image analysis technology. Analyze and process single cells to realize point-to-point research, so as to research and develop new drugs and treatment methods. On the other hand, intelligence has become the development trend of computer medical image processing. Only by accurately segmenting blood cells, computer image analysis can accurately classify and identify blood cells, so as to truly realize the intelligence of computer image analysis in blood cell analysis. change.

1.2 Research status and development trend

For a long time, a lot of medical image information is often observed and counted by the naked eye of medical examiners. At present, the analysis of blood cell images by computer image analysis is still insufficient, and the accurate segmentation of blood cell images cannot be achieved, which affects the classification of blood cell images. Recognition. If the analysis and processing of blood cell images is carried out with the help of computer intelligence and pattern recognition technology, it is possible to provide medical experts with more reliable test data faster, easier, and more accurately, so that medical experts can make better results. Scientific judgment. Therefore, the intelligence of computer image analysis has become the general trend of medical image processing.

Digital image processing technology can eliminate the influence of human subjective factors and obtain objective quantitative data. Therefore, it has become a hot spot in the research and development of medical testing technology and equipment. The emergence of various automatic blood cell recognition software has greatly improved the working conditions of inspectors and reduced labor intensity (Kelleher & Tierney, 2018). At present, the blood cell identifiers used in major medical institutions in China can not only facilitate the classification and statistics of platelets, white blood cells, red blood cells, etc. More than twenty parameters such as packed volume and mononuclear cell ratio. However, it is not enough to just do this. In the current era of advanced medical technology, people have put forward new requirements for the objectivity and accuracy of medical testing. People hope that the results of testing and diagnosis will be more objective and accurate. For this reason, many researchers have devoted themselves to studying the automatic analysis and identification technology of blood cells. Especially in recent years, many automatic cell identification systems have appeared. The main goal is to analyze and identify the original images of blood cells, and then perform blood cells. Segmentation, which can extract single blood cells, and then analyze and process single blood cells, so as to achieve accurate identification and classification of blood cells, greatly reduce the labor of medical staff, and provide a quantitative and objective basis for clinical diagnosis. Therefore, the research and development of automatic cell classification and recognition systems provide more objective and accurate medical information for medical testing and diagnosis, reduce the work intensity of medical staff, and meet people's requirements for medical testing.

1.3 Research Content

This project focuses on the research of blood cell detection based on machine learning, and uses cross-entropy loss function based on the YOLOv3 algorithm to make it more suitable for the field of blood cell detection. The main contents of this project are as follows: Select the YOLOv3 algorithm as the basis to meet the accuracy and speed requirements of blood cell detection. Using the gradient descent algorithm to optimize the process of selecting the target and improve the stability of the algorithm. Improved the loss function of the YOLOv3 algorithm, using the cross-entropy loss function to make the training converge faster, and the coordinate positioning is more accurate. In view of the overlap and occlusion of blood cell detection rate.

After the thesis is completed, blood cells can be successfully identified and a high accuracy rate will be achieved.

1.4 Organizational structure

Chapter One Introduction. This project introduces the research background and significance of blood cell detection, discusses the research status of target detection and blood cell detection at home and abroad, and discusses the main content and organizational structure of this project.

The second chapter is the related theoretical basis of convolutional neural network target detection. Starting from each layer of the convolutional neural network, the detection algorithm based on the convolutional neural network is introduced. Then briefly describe the feature learning principle of volume neural network. Then compare several different types of convolutional neural network algorithms, introduce the YOLOv3 algorithm, introduce the network structure, detection process, advantages and disadvantages of the YOLOv3 algorithm, and lay the foundation for the experiment on YOLOv3 in the following.

The third chapter is the improvement of target detection algorithm based on YOLOv3 algorithm. First introduced the difficulties and challenges of blood cell detection, and then used the cross-entropy loss function for blood cell recognition from the perspective of gradient descent algorithm.

The fourth chapter is the experimental results and analysis of the algorithm. The feasibility of the improved algorithm is verified by AP, actual detection pictures and MR-FPPI curve under different algorithms, and the real-time performance of the improved algorithm is verified by the detection rate and FPS during video detection, and finally the AP is compared with related algorithms.

The fifth chapter Summary and Outlook. Summarized the improvement of this project based on the YOLOv3 algorithm and the effect in blood cell detection, and pointed out the shortcomings of this project and the research direction worth exploring in the future.

CHAPTER 2: LITERATURE REVIEW

This chapter focuses on the convolutional neural network framework, introducing the input layer, convolutional layer, downsampling layer, activation function, and loss function layer of the network. Then introduce the convolutional neural network YOLOv3 and related technologies used in this project, mainly introduce the network structure and detection process of YOLOv3, and analyze the advantages and disadvantages of YOLOv3, paving the way for the use of YOLOv3 in the following. Finally, the data set and algorithm evaluation index of this project are introduced (Zhan, Wang, Ben, Ruan, & Chen, 2019).

2.1 Convolutional neural network architecture

Convolutional neural networks mainly include input layer, convolutional layer, down-sampling layer, activation function, and loss function layer. The division of labor between each layer is clear. The construction of convolutional neural network is to connect different layers in a specific and orderly manner. This section will introduce the convolutional neural network from each layer, and pave the way for the improvement of the network layer later.

2.1.1 Input layer

The input layer is to input the trained or detected pictures into the convolutional neural neural network after preprocessing, usually at the head of the convolutional neural network. If you directly input the picture into the convolutional neural network for training, the following problems will occur: 1. The convergence speed is slow due to the different sizes of the input pictures (Gong, Zhang, Wu, & Ni, 2019). 2. The activation function in the convolutional neural network is limited in the value range, so the target data of the network training needs to be mapped to the value range of the activation

function. If the image is directly input, the value range of the mapping will be too large, making the activation function invalid.

In response to these problems, the input layer is required to preprocess the input image data, such as data standardization, normalization, de-average whitening, etc., so as to integrate the input image data into a certain range, which is convenient for network training and error propagation. The input layer of YOLOv3 processes the image to a size of 640×480.

2.1.2 Convolutional layer

The purpose of the convolutional layer is to perform a convolution operation on the picture input by the input layer to extract picture features. The convolutional layer is usually stacked by multiple convolution kernels. The convolution kernel determines the range of the receptive field of the convolutional layer. The larger the convolution kernel, the larger the receptive field, and the more sensitive the network; the value of the convolution kernel is the network model. This parameter is optimized by training the network model. The number of convolution kernels in the convolution layer, so that one convolution kernel performs convolution operations on a feature picture, and multiple convolution kernels form a set of filters. The number of filters is the number of output feature pictures of the previous convolution set of output feature pictures.

Before the convolution operation is performed on the feature pictures of the l-1th layer, the number of convolution kernels needs to be determined first, so that it is equal to the number of output pictures of the l-1th layer. For a feature x_i^{l-1} in the local receptive field of the l-1th feature picture, use the convolution kernel to convolve it to obtain the convolution value $x_i^{l-1} \times \text{kernel}_{ij}^{l}$, and finally sum all the convolution values

and add the offset variable. The calculation formula for the convolution process is as follows:

$$\mathbf{x}_{j}^{l} = \mathbf{f}(\sum_{i=M_{i}} \mathbf{x}_{i}^{l-1} \times \text{kernel}_{ij}^{l} + \mathbf{B}^{l})$$
(2-1)

Among them, f represents the activation function, which makes the neural network convert the linear regression model into a nonlinear model; x_i^l represents the j-th feature image of the l-th layer after convolution; x_i^{l-1} represents the output of the l-1 convolutional layer. The i-th feature image: kernel represents the i-th convolution kernel in the j-th filter, which is convolved with the i-th feature image output by the l-1 convolution layer: Bl represents the l-th corresponding to the layer is mainly to improve the classification performance of the network: Mj which represents the corresponding feature image area of the convolution kernel in the convolution process (Xiao et al., 2019). When the number of convolution kernels is one, the convolution process is shown in Figure 2.1.



Figure 2.1 Schematic diagram of the convolution process

In the convolution operation shown in Figure 2.1, the input feature image size is 5x5, the convolution kernel size is 3x3, and the convolution step length of the convolution kernel on the feature image is 1, that is, the convolution kernel is on the input feature image once Move one pixel. After the convolution operation, the output feature picture size is 3x3, and the pixel value in the output feature picture is the sum of the product of the convolution kernel and the pixel value of the corresponding feature picture area. The parameters of the convolution kernel are updated by reverse gradient propagation during the training of the convolutional neural network. Different convolution kernels have different feature extraction capabilities and different forms of feature extraction. The feature extraction capabilities of the convolution kernel are optimized by parameter updates to improve The detection capability of the network.

Convolutional neural networks use sparse connections to connect input and output. As can be seen from Figure 2.2, different from traditional neural networks, sparse connections are only for local areas, thereby reducing the network parameters and calculations. At the same time, in order to solve the problem of small receptive fields caused by sparse connections, convolutional neural networks use weight sharing to increase the receptive fields of neurons. Figure 2.3 is a schematic diagram of weight sharing. In the figure, the middle layer neuron connected to the top neuron (y3) is (f2, f3, f4), and the bottom layer neuron connected to the middle layer (f2, 3f, f4) is (x1, x2, x3, x4, x5), which means that the receptive field of the top neuron (y3) at the bottom is (x1, x2, x3, x4, x5). Although the use of sparse connections leads to a reduction in the amount of data between single layers, the use of weight sharing makes the deeper the network layer, the more data the top neuron receives, and while reducing the parameters,

the data circulation is appropriately increased, thereby improving the top layer. The range of the neuron's receptive field.



Figure 2.2 Comparison of normal connection and sparse connection



Figure 2.3 Schematic diagram of weight sharing

Generally speaking, the deeper the number of convolutional network layers, the stronger the feature extraction ability, but when the network layer is too deep, gradient dispersion is likely to occur. In the convolutional neural network, the parameters are propagated through the loss gradient. Assuming that the parameter adjustment function of the first layer of the network is f(1), the loss is $\epsilon(1)$, then the parameter optimization of the first layer needs to calculate the gradient corresponding to the loss of this layer $\delta(1)$:

$$\delta(\mathbf{l}) = (\partial \delta(\mathbf{l})) / (\partial f(\mathbf{l}))) \tag{2-2}$$

This gradient is the gradient of a single layer. According to the chain rule (during the propagation process, the gradient of this layer is equal to the product of the gradients of all layers before this layer), the gradient of the first layer depends on the gradient of the previous layer:

$$\delta(\mathbf{l}) = \partial \delta(\mathbf{l}) / (\partial f(\mathbf{l})) * \delta(\mathbf{l} - \mathbf{l})$$
(2-3)

It can be seen that in the propagation process, if the multi-layer propagation gradient as/f is less than 1, it will cause the gradient in the propagation to gradually disappear, and the underlying parameters can not be updated effectively, resulting in gradient dispersion, making the network unable to achieve the best performance. Therefore, gradient diffusion is an urgent problem in deep convolutional neural networks.

2.1.3 Downsampling layer

The purpose of the downsampling layer is to reduce the parameters of the network layer, compress the free variables in the feature map, further deepen the network features, and improve the robustness of the neural network. In the convolutional neural network, the network needs to extract the typical feature points in the picture. The difference between this feature point and the neighboring feature points is more obvious, and the performance is the most prominent (Chellappa, Roy-Chowdhury, Zhou, & Processing, 2005). At this time, it needs to be extracted through the down-sampling layer. At the same time, the downsampling layer can compress redundant non-key variables in the network and streamline network parameters. The down-sampling layer of YOLOv3 uses a 3×3 convolution kernel. The down-sampling process is basically the same as the convolution process. The difference is that the down-sampling layer sets the step size to 2, that is, a span of two pixels. After the sampling operation, the feature image can be reduced by one time, and the variable compression and extraction can be realized.

2.1.4 Activation function

The purpose of the activation function is to improve the feature expression ability of the network through nonlinear transformation. In the convolutional neural network, the classification of pictures is nonlinear, and the convolutional layer just multiplies and adds the pixel values of the feature pictures and the corresponding parameter values in the convolution kernel. This is a simple linear operation, so it needs to be activated. The function maps the result to a non-linear classification. At present, the commonly used activation functions mainly include Sigmoid activation function, Relu activation function and Leaky Relu activation function.

2.1.5 Loss layer function

The role of the loss function layer is to calculate the network model error through the loss function, adjust the network parameters through the error in the back propagation, and optimize the network model. Simply put, the loss function in the loss function layer is used to indicate the quality of the convolutional neural network. The larger the loss function value, the worse the model detection ability. The loss function layer of YOLOv3 is composed of three types of loss functions: frame positioning loss function, prediction type loss function and confidence loss function. The frame positioning loss function is composed of the center coordinate loss function and the frame width and height loss function.

In YOLOv3, different grid areas are responsible for the prediction frames of the network, and each grid is responsible for predicting 3 prediction frames. For the center coordinates (\hat{x}_i, \hat{y}_i) and border width and height $(\widehat{w}_i, \widehat{h}_i)$ of the i-th prediction box of the j-th grid of the feature image output by the network (Q. Xu et al., 2020), the center coordinate loss function and the border width and height loss function are required to calculate the sum The error between the center coordinates (xi, yi) of the marker frame

and the width and height (wi, hi) of the frame makes the network correct the network parameters during backpropagation to improve the positioning ability of the network. The two loss functions are shown in the formula (Chellappa et al., 2005).

$$Loss1 = \sum_{i=0}^{S^2} \sum_{j=0}^{B} \ell_{ij}^{obj} \left[(\sqrt{x_i} - \sqrt{\hat{x}_i})^2 + (\sqrt{y_i} - \sqrt{\hat{y}_i})^2 \right]$$
(2-4)

$$Loss 2 = \sum_{i=0}^{S^2} \sum_{j=0}^{B} \ell_{ij}^{obj} \left[(\sqrt{w_i} - \sqrt{\hat{w}_i})^2 + (\sqrt{h_i} - \sqrt{\hat{h}_i})^2 \right]$$
(2-5)

Among them, S represents the length and width of the network output feature image, YOLOv3 finally outputs three scale feature images, the image size is the input layer output image, 1/8, 1/16, 1/32, and B represents The number of prediction frames in a network is 3 in YOLOv3, and obj represents the category of the prediction frame output by the network, which is usually represented by a number, and its category.

For the prior prediction box output by the network, it is necessary to calculate the error between its prediction confidence score \hat{p}_i (c) and the corresponding labeled box confidence score p_i (c) (usually 1, which means there is a target in the labeled box), Used to determine whether there is a target in the a priori prediction box, the error is obtained by the confidence loss function, and the loss function formula is as follows.

$$Loss3 = \sum_{i=0}^{S^2} \sum_{j=0}^{B} \ell_{ij}^{obj} \left(p_i(c) - \hat{p}_i(c) \right)^2$$
(2-6)

Finally, the network needs to determine the category error of the prior prediction box, that is, the error between the category probability \hat{C}_i of the prior prediction output by the network and the calibrated category Ci. The error is obtained by the predicted category loss function, and its loss function is shown in the formula:

$$Loss4 = \lambda_{obj} \sum_{i=0}^{S^2} \sum_{j=0}^{B} \ell_{ij}^{obj} (C_i - \hat{C}_i)^2$$
(2-7)

Among them, Ci the value is 0 or 1, Ci is 1 means that the target in the prediction box belongs to the i-th category, and \hat{C}_i is (0,1).

In the end, the overall loss function of YOLOv3 is equal to the sum of each error, expressed as follows: Loss = Loss1 + Loss2+Loss3+ Loss4

2.2 Feature learning principle of convolutional neural network

In general, the convolutional neural network is a hierarchical model.For the input original image data, the convolutional neural network uses a series of operations such as convolutional layer, downsampling layer, and activation function mapping to remove the deep semantic information of the image from This process is called "feedforward calculation". Finally, the final layer of the convolutional neural network transforms its target tasks (classification, positioning, etc.) into real picture data and displays it on the original picture.

It can be seen that the advantages of convolutional neural networks over traditional machine learning algorithms are that they have more layers and deeper feature extraction capabilities. For different input pictures, the convolutional neural network can extract the deeper features of different pictures, that is, it can get the unique features of various objects, so as to judge which category the object belongs to from these features. The feature extraction of convolutional neural network is mainly done through convolution operation, that is, its feature extraction ability is related to the convolution kernel. In a convolution kernel, the parameters in the convolution kernel are arranged in a certain rule. When performing convolution operations with the input picture, if the picture in the receptive field of the convolution kernel has a similar parameter

arrangement to the convolution kernel, then It will show a higher value on the output feature map data, if it is not similar, it will be close to 0. After passing the activation function, the larger value is retained, and the smaller feature value is assigned 0 or negative. In the deep convolutional neural network, higher-order features can be obtained after multiple convolutional layers. Figure 2.6 is a schematic diagram of a simple convolution process. After the input image is subjected to the convolution operation of the convolution kernel in the figure, the edge texture of the image can be extracted, and deeper image features can be extracted through multiple convolution processes.

In the training preparation stage, the parameters of the convolutional network need to be initialized randomly. During training, the network before the input image passes through the loss function will get the a priori prediction frame coordinates, confidence score and classification score. The current training loss value can be calculated through the label file corresponding to the input image, and then the chain derivation rule is used. , You can deduce the gradient descent direction of the loss function with respect to the parameters, so as to adjust the network parameters according to the gradient direction, so that the convolution kernel in the convolutional neural network layer can adapt to the characteristics of the target object to be detected, and achieve more accurate classification and position.



Figure 2.4 Schematic diagram of the convolution process

2.3 Introduction to YOLOv3

Convolutional neural networks have two main tasks to achieve in target detection: first, detect the location of the object; second, identify the category of the object. According to the division of the network structure, the existing convolutional neural networks are divided into two-level convolutional neural networks and end-to-end convolutional neural networks. The two-level convolutional neural network completes the two tasks separately, that is, first obtains a series of prior prediction boxes as samples, and then classifies the samples through the convolutional neural network. The detection accuracy of this algorithm is high, but the model is relatively complex and the network parameters are many. Each module needs independent training. The network model is difficult to optimize, and the detection speed is slow, and it is difficult to detect in real time.

YOLOv3 is a typical end-to-end convolutional neural network, which turns object detection into a regression problem, from inputting a picture to completing the detection in one step. Its advantage is that the network is more streamlined, more robust, and easier to optimize the model. And improvement. The proposal of the YOLOv3 algorithm makes the target detection return from the traditional two-level convolutional neural network to the end-to-end convolutional neural network, and brings a new idea of target detection: combining the positioning and classification tasks, and scanning

the picture can be completed in one time Target detection task.

2.4 YOLOv3 network structure

The advantage of the YOLOv3 algorithm is to realize the detection task with a simple network structure without basically reducing the detection performance. The algorithm first extracts the deep features of the picture according to the traditional convolutional neural network structure, and outputs the target category and predicted

frame in the picture in the final network. YOLOv3 is mainly composed of an input layer, a convolution layer, an activation function, a downsampling layer, an output layer, and a loss function layer.

During training or detection, the image is input into the backbone network after the scale transformation and normalization operation of the input layer. The backbone network of YOLOv3 is the first 52 layers of the darknets53 convolutional network model. The idea of building the network is to use a small convolution kernel to build a deep convolutional network to achieve deep feature extraction of images. The network is built with 3x3 and 1×1 convolution kernels. Both of these convolution kernels are small convolution kernels. Among them, the 3×3 convolution kernel is used to extract image features, and the 1X1 convolution kernel is used to reduce convolution. The number of nuclear channels, thereby reducing the parameters in the network (Zhao, Wang, & Du, 2020).

The backbone network of YOLOv3 has a total of 5 downsampling layers, which are located in the 4th, 9th, 26th, 43rd, and 52nd layers of the network. The size of the input image is reduced by one time after each downsampling layer, and the final network uses the output feature images of the third, fourth, and fifth downsampling layers as the output of the backbone network, and the size of the feature images is 1/8, respectively. 1/16, 1/32,. The main purpose of outputting three feature images of different sizes is to improve the multi-scale detection capability of the network.

After extracting the deep feature information of the picture, the data information is passed to the output layer to obtain the final detection information. The information mainly includes the predicted frame coordinates, the confidence score and the classification score. The output layer outputs $S \times Sx[3x(4+1+class)]$ data information for each size feature image, where SxS represents the final output size of the feature image,

and 3 represents the number of prediction targets in each grid (Qian, Zhou, & Zheng, 2019), 4 represents that the prediction frame coordinate information includes four parameters: center coordinates and length and width, 1 represents the confidence score of the prediction frame, and class represents the total number of categories that the network can predict.

In the training phase, the output parameters of the output layer are passed into the loss function layer, and the label file data are input into the loss function to calculate the loss value of the network, and the network parameters are optimized by backpropagating the loss value to improve the detection performance of the network.

2.5 YOLOv3 detection process

YOLOv3 directly outputs sample candidate frames and classification results in the detection process, which improves the anti-interference ability of the network. At the same time, in order to improve the detection efficiency, YOLOv3 divides the detected pictures into different grids for processing. As shown in Figure 2.8, the algorithm divides the feature image output by the network into a grid of SxS size. If the center of an object is located in this grid, the grid is responsible for detecting the object. Each grid has 3 anchor frames responsible for predicting the target frame, that is, the network can predict at most SXSx3 targets for the entire picture. Finally, the network outputs three types of information for each target, which are the target's coordinate information, the prediction score Pr(O), and the target's probability value $Pr(C_i)$ for each category (Z. Wang et al., 2020).



Figure 2.5 YOLOv3 detection process

The designer only needs to design the functions that he wants to achieve without a logic analyzer, which can be directly connected to the pin signals derived from the chip for observation and analysis.

2.6 Advantages and disadvantages of YOLOv3 algorithm

The advantages of the YOLOv3 algorithm: (1) Since YOLOv3 is an end-to-end convolutional neural network model, the network structure is more streamlined than the two-level convolutional neural network, with fewer network parameters, and easier optimization and improvement. (2) The YOLOv3 algorithm uses more structures to prevent over-fitting, the network generalization ability is stronger, and the model is more robust.

Disadvantages of the YOLOv3 algorithm: (1) In the detection of blood cells in dense scenes, the loss function of YOLOv3 is only for the regression of the prior prediction box and the labeled box, and non-maximum value suppression will judge the occluded blood cell prior prediction box as a false prediction Frames, resulting in the network's

insufficient ability to detect cells that are obscured by each other. (2) When cells are occluded from each other, the features of the occluded cells extracted by YOLOv3 are not the same as the overall cell features, resulting in a lower confidence score corresponding to the prior prediction box, which makes the network judge it as a false prediction box, thus Unable to detect occluded cells. And because of the limited extraction capability of the backbone network of YOLOv3 and the insufficient multi-scale detection capability of the network, it is difficult for the network to detect pedestrians that are too close or too far, which further weakens the network's ability to detect occluded pedestrians.

CHAPTER 3: RESEARCH METHODOLOGY

In target detection based on convolutional neural networks, the training of the network model requires a large number of image data sets. The quality of the data set and the calibration method directly determine the performance of the trained model. In order to obtain high-quality cell data sets, various relevant institutions and schools obtain a large number of blood cell pictures from different scenes and calibrate them into blood cell data sets for model training. Different data sets are trained due to different collection conditions and collection equipment. The effect is also different. Therefore, selecting a suitable pedestrian data set is crucial for the training of the network model. For a trained network model, there are multiple indicators that can judge the quality of the model. The most common ones are accuracy, recall, and loss. Rate, mAP, comprehensive evaluation index F-Score, etc. Different evaluation indexes also have their own choices for the evaluation ability of the model, so which evaluation index is used is very important for the appraisal of model performance. It can be seen that the appropriate data set and evaluation index are very important for the cell detection model. The main work of this section is to introduce the cell data set used in this project. At the same time, in order to enable the data set to be trained with the YOLOv3 algorithm, recalibrate the original pedestrian data set. Finally, we introduce the algorithm evaluation index used in this project, and use the evaluation index to detect the effectiveness of this project on the use of the YOLOv3 algorithm (Xiao et al., 2019).

3.1 Data set selection

So far, the blood cell detection data set is mainly organized by schools and related institutions. The blood cell data set in this project comes from the kaggle data set. Figure 3.1 shows the relevant pictures in the data set. Compared with other pedestrian

datasets, the cells in the images of this dataset are clearer and more numerous, which is convenient for the network to extract the subtle features of the cells. The disadvantage is that the data set has fewer pictures, and there are only 364 training pictures, so it cannot meet the training needs; and there are no cells in the calibration picture in the data set. The uncalibrated edges and occluded blood cells make the network unable to extract the difference during training. Size characteristics of blood cells. Therefore, most model training, especially models with more complex structures, are not trained from scratch, but based on an existing pre-trained model to train their own data set, which can save a lot of training time, thereby Improve training efficiency. In addition, pre-training models are often trained on large public data sets, such as imagenet. So I used the data from the VOC2007 pre-training set here.





3.2 Calibration of the data set

In the existing data set, blood cells are not calibrated, and there are some blood cells that are blocked or incompletely photographed. These will affect the accuracy of the program's recognition of blood cells.



Figure 3.2 LabelImg interface

Then we perform batch annotation, and use the Open Dir button to open the folder of pictures that need to be marked. Use the Change Save Dir button to open the folder where the annotation file is stored. Use the w shortcut key or click create RectBox to start marking, and you need to save it after marking.



Figure 3.3 Interface after marking

3.3 Model evaluation index

There are many evaluation indicators for machine learning models. This project uses mAP, recall, accuracy and comprehensive evaluation indicators (F-Score).

The calculation formula for accuracy is Precision = $\frac{T_p}{T_p+F_p}$, Where Fp represents the number of target blood cells that were falsely detected in the test set.

Accuracy alone cannot fully measure the pros and cons of a classifier: if there are 50 positive and negative samples each, the network will judge all 45 positive samples as negative samples, and all 50 negative samples are judged correctly, and the accuracy at this time is also 100%, because the classifier distinguishes 5 positive samples, and all of them are correct. Therefore, the recall rate needs to be judged together.

Recall rate (Recall) calculation formula is $\text{Recall} = \frac{T_p}{T_p + F_n}$. Among them, Tp represents the number of target blood cells that were correctly detected in the test set, and Fn represents the number of target blood cells that were missed in the test set.

It refers to the proportion of samples predicted to be positive to the true total number of positive samples. If the recall rate is 100%, it means that all positive classes are classified into positive classes by the classifier. If the recall rate is 0%, it means that no positive class is classified as a positive class.

mAP (Mean Average Precision), that is, the average AP value, is the average AP value of multiple verification set individuals, which can be used as an index to evaluate the detection accuracy in target detection. When calculating the AP value of a single category, you need to calculate the P-R curve. P represents the y-axis accuracy rate, and R represents the x-axis recall rate. According to this curve to calculate the enclosed area can get the AP value (Shao et al., 2021).

When calculating the accuracy and recall of each group, you need to know the values of TP, FP, FN, etc., and in the target detection, which is the correct prediction? Two conditions need to be met: 1. The category prediction is correct and the confidence is greater than Specified threshold 2. The IoU of the predicted frame and the real frame output by the grid is also greater than the specified threshold.

Each result predicted by the network is divided into two parts: prediction box and confidence probability. The prediction box will have 4 values, which are the xy coordinates of the upper left corner of the prediction box and the xy coordinates of the lower right corner. If the loU threshold is 0.5, the prediction box is correct, and it can be recorded as TP (True Positive). False Positive), and then subtract TP from the total number of true positive samples to get FN. In this way, a set of accuracy and recall rates are obtained under the condition that the confidence threshold is 0.6. Using different confidence thresholds from 0 to 1, 11 values can be obtained, and then the AP value of a single category can be calculated. Calculate the AP value for each category in the data set, and then add the arithmetic mean to get the mAP (Zhao et al., 2020).

3.4 Summary

This chapter first introduces the functions of each layer of convolutional neural network and the principle of network feature learning, and then introduces a typical end-to-end convolutional neural network algorithm YOLOv3, summarizes the network structure and detection process of YOLOv3, and analyzes the advantages and disadvantages of YOLOv3. Finally, the blood cell detection data set and model evaluation indicators used in the thesis are introduced. In terms of data sets, this chapter analyzes the advantages and disadvantages of the selected blood cell data set, and at the same time modifies the label files in the data set to prepare for model training. In terms

of evaluation indicators, it mainly introduces recall rate, accuracy, F-score and mAP, and analyzes its advantages as evaluation indicators in blood cell testing.

University

CHAPTER 4: RESULTS AND ANALYSIS

4.1 YOLO algorithm

YOLO (You Only Look Once) has received widespread attention since it was proposed. It has gone through three iterations, constantly optimizing itself and absorbing the advantages of other detection models. Different from Faster R-CNN's target detection algorithm based on region prediction (Qian et al., 2019), YOLO uses the idea of regression to directly detect the entire input image to get the target category and location.

The YOLO algorithm realizes end-to-end target detection, and the model divides the input image into SxS grids. If the center of a target is in a certain grid, the grid is responsible for predicting the target. In the detection process, each grid will generate B bounding boxes, and each bounding box must predict the position information (x, y, w, h) and confidence of the bounding box. The confidence is the product of the probability of the target contained in the bounding box (Pr(object)) and the accuracy of the bounding box IOU (Li, Shen, & Li, 2019).

If there is a target in the grid, Pr(object) = 1, and when IoU is 1, it means that the real labeled frame and the predicted frame overlap.

In YOLO v3, the complexity of the model is further improved, and multi-scale fusion is used to predict, and the position and category predictions are made on the multi-scale feature map, thereby improving the accuracy of target detection, and introducing residual network at the same time A better feature extraction network, darknet-53, is proposed (Q. Xu et al., 2020). Based on the above improvements, the YOLO v3 model has better results in accuracy and speed.

4.2 Feature extraction network darknet-53

	Туре	Filters	Size	Output
	Convolutional	32	3 × 3	256×256
	Convolutional	64	$3 \times 3 / 2$	128×128
	Convolutional	32	1×1	
1×	Convolutional	64	3×3	
	Residual			128 × 128
	Convolutional	128	$3 \times 3 / 2$	64×64
1	Convolutional	64	1 × 1	
2×	Convolutional	128	3×3	
100000	Residual			64 × 64
	Convolutional	256	3×3/2	32 × 32
	Convolutional	128	1 × 1	
8×	Convolutional	256	3×3	
	Residual			32 × 32
	Convolutional	512	3×3/2	16 × 16
	Convolutional	256	1 × 1	
8×	Convolutional	512	3×3	
	Residual			16 × 16
	Convolutional	1024	3×3/2	8 × 8
	Convolutional	512	1 × 1	
4×	Convolutional	1024	3 × 3	
	Residual			8×8
	Avgpool		Global	
	Connected		1000	
	Softmax			

Figure 4.1 darknet-53

This new network is better than darknet-19, but still more effective than ResNet-101 or ResNet-152.

Here are some ImageNet results:

Backbone	Top-1	Top-5	Bn Ops	BFLOP/s	FPS
Darknet-19 [15]	74.1	91.8	7.29	1246	171
ResNet-101[5]	77.1	93.7	19.7	1039	53
ResNet-152 [5]	77.6	93.8	29.4	1090	37
Darknet-53	77.2	93.8	18.7	1457	78

Figure 4.2 Comparsion of backbones

Each network uses the same settings for training and testing at 256×256 , with a single cropping accuracy. The running time is measured on a 256×256 Titan X. However, the performance of Darknet-53 is the same as the most advanced classifier, but with fewer floating point operations and faster speed. Darknet-53 is 1.5 times faster than ResNet-101 and 2 times faster than ResNet-152. Darknet-53 also achieves the highest measured floating point operation per second. This means that the network junction makes better use of the GPU, making it detect more efficiently and therefore faster. This is mainly because ResNets only has too many layers, resulting in low efficiency.

4.3 Target detection network structure

YOLO3 further uses 3 feature maps of different scales for object detection. More fine-grained features can be detected.

The results of these three detections are not the same thing. The rough understanding here is that different scales are given to detect objects of different sizes.

The final output of the network has 3 scales, namely 1/32, 1/16, and 1/8;

After the 79th layer, after several convolution operations, a prediction result of 1/32 (13*13) is obtained. The downsampling factor is high. Here the receptive field of the feature map is relatively large, so it is suitable for detecting objects with larger sizes in the image (Xu, Jia, Sun, Liu, & Cui, 2020).

Then this result is concated with the result of the 61st layer through upsampling, and then a 1/16 prediction result is obtained through several convolution operations; it has a medium-scale receptive field and is suitable for detecting medium-scale objects.

After the result of layer 91 is up-sampled, the result of layer 36 is concat. After several convolution operations, the result is 1/8. It has the smallest receptive field and is suitable for detecting small-sized objects (Z.-F. Xu et al., 2020).



Figure 4.3 yolo v3 network Architecture

4.4 Loss function

Here we use two loss functions.

The first is Mean Squared Error (MSE), which is a relatively common loss function, defined as $MSE = \frac{1}{n} \sum_{i}^{n} (\hat{y}_{i} - y_{i})^{2}$.

The second is the BCE loss function. In the case of two classifications, there are only two cases where the model needs to predict the result (Liu, Ren, Wang, & Yuan, 2020). For each category, our predicted probabilities are p and 1 -p. At this time, the expression is:

$$L = \frac{1}{N} \sum_{i} -[y_{i} \times \log(p_{i}) + (1 - y_{i}) \times \log((1 - p_{i}))]$$
(4-1)

 y_i — represents the label of the sampley_i, the positive class is 1 and the negative class is 0.

p_i—— Indicates the probability that the sample p_i is predicted to be a positive class (Liu et al., 2020).

4.5 Results and analysis

Calculate the recall accuracy, ap and f-score every 5 times during the entire training process, and then print them out. And use pyqt5 to design a graphical interface to facilitate blood cell detection.

Metrics	YOLO Layer 0	YOLO Layer 1	YOLO Layer 2
grid_size	12	24	48
loss	2.186347	1.988263	1.550236
x	0.027242	0.017985	0.069879
у	0.065141	0.053314	0.056139
w	0.023274	0.013171	0.008307
h	0.019814	0.026130	0.015680
conf	2.050871	1.877660	1.400229
cls	0.000006	0.000004	0.000001
cls_acc	100.00%	100.00%	100.00%
recall50	0.562500	0.437500	0.500000
recall75	0.437500	0.437500	0.500000
precision	0.900000	0.777778	0.421053
conf_obj	0.479494	0.461717	0.514679
conf_noobj	0.006253	0.004545	0.003870

---- [Epoch 49/50, Batch 103/329] ----

Total loss 5.724845886230469

---- ETA 0:02:30.862347

Figure 4.4 Training data

Precision = number of correct information extracted / number of extracted information, recall rate = number of correct information extracted / number of information in the sample, the two data values are between 0 and 1, the closer the value is 1. The precision or recall rate is higher. As shown in Figure 5.2, both of these data remain at 0.8.



Figure 4.5 The precision value and recall

P and R indicators sometimes have contradictions, so you need to consider them comprehensively. F-Score is the weighted average of Precision and Recall (M. Wang et al., 2020). It can be seen that F1 combines the results of P and R. When F1 is higher, it can indicate that the test method is more effective.



Figure 4.6 The F-score value

When we take different degrees of confidence, we can get different Precision and different Recall. When we get dense enough confidence, we can get a lot of Precision and Recall.

At this time, Precision and Recall can draw a line on the picture, and the area under this line is the AP value of a certain class. In fact, mAP is to average the AP values of all classes. Since our model has only one type, ap and map are the same.

AP measures the quality of the learned model in each category, and mAP measures the quality of the learned model in all categories. After the AP is obtained, the calculation of mAP becomes very simple.



Figure 4.7 The mAP value

4.6 Graphical interface

We can click the Select Image button to select an image in the data set folder, and then click the Cell detection button to output the result image.

At the same time, I also added contrast adjustment and brightness adjustment functions to detect blood cells in different backgrounds to make the detection accuracy more accurate.

Inspection chart	Result graph				
	RBC RBC RBC RBC RBC RBC RBC RBC				
Select Image					
Cell detection					
Edge detection					
Bright a	Bright adjustment				
contrast adjustment					

Figure 4.8 Graphical interface

CHAPTER 5: CONCLUSION AND OUTLOOK

Blood cell target detection has always been one of the hot issues in the medical field, and its wide application in various scenarios has brought great convenience to people's work and life. However, due to the fact that blood cell targets have many small targets, scale scaling, target rotation, background clutter, target dense and diverse shapes, etc., the use of traditional target detection methods or early target detection methods based on deep learning cannot perform well on blood cell targets. Identify and locate. In this project, two factors of detection accuracy and detection speed are comprehensively considered. After comparing several commonly used target detection methods, YOLOv3, which has the best detection speed and detection accuracy, is finally selected as the blood cell target detection algorithm in this project. Finally, this project also designed a graphical interface to improve the processing efficiency in blood cell target detection.

In the first chapter, this project first introduces the research background and significance of blood cell target detection and the current research status of related fields at home and abroad, and then elaborates the main content and chapter arrangement of this project. In the second chapter, this project first briefly introduces the related knowledge of artificial neural network and convolutional neural network, and then focuses on the structure of darknet series network and the principle of YOLO series algorithm. Based on the theories and methods introduced in Chapter 2, this project first analyzes the many small targets in blood cell targets, scale scaling, target rotation, background clutter, dense targets, and diverse shapes. Annotate the data set to improve accuracy. After that, the VOC2007 data set was used for pre-training. Then explained the evaluation criteria of the model. The fourth chapter introduces the network structure of yolov3 and the choice of loss function. The fifth chapter explains

the training process and training results of the entire model, and shows the graphical interface.

Due to the limited time, this project still has its shortcomings. 1. The data set for blood cell detection needs to be expanded. 2. Although this project has been improved on the basis of the original YOLOv3, it has improved the problem of YOLOv3 missed detection, but due to the denser blood cell targets and smaller features, the improved YOLOv3 still has a certain degree of missed detection. In summary, blood cell target detection with high detection accuracy, low missed detection rate, and accurate positioning and identification of the target is an important development direction of cell detection in the future, and there is still a broad space for development.

REFERENCES

- Chellappa, R., Roy-Chowdhury, A. K., Zhou, S. K. J. S. L. o. I., Video, & Processing, M. (2005). Recognition of humans and their activities using video. *1*(1), 1-173.
- Gong, Z., Zhang, Z., Wu, M., & Ni, H. (2019). Distribution Network Load Forecasting Based on Improved BP Neural Network. Paper presented at the 2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM).
- Kelleher, J. D., & Tierney, B. (2018). Data science: MIT Press.
- Li, W., Shen, Z., & Li, P. (2019). *Crack detection of track plate based on YOLO*. Paper presented at the 2019 12th International Symposium on Computational Intelligence and Design (ISCID).
- Liu, Y., Ren, J., Wang, C., & Yuan, X. (2020). Research And Implementation Of Facialnet Based On Convolutional Neural Network. Paper presented at the 2020 5th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS).
- Qian, H., Zhou, X., & Zheng, M. (2019). Detection and Recognition of Abnormal Behavior based on Multi-level Residual Network. Paper presented at the 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC).
- Shao, X., Wei, J., Guo, D., Zheng, R., Nie, X., Wang, G., & Zhao, Y. (2021). *Pedestrian Detection Algorithm based on Improved Faster RCNN*. Paper presented at the 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC).
- Wang, M., Jia, S., Chen, E., Yang, S., Liu, P., & Qi, Z. J. A. I. (2020). A derived least square fast learning network model. *50*(12), 4176-4194.
- Wang, Z., Liu, D., Lei, Y., Niu, X., Wang, S., Shi, L. J. M. T., & Applications. (2020). Small target detection based on bird's visual information processing mechanism. 79, 22083-22105.
- Xiao, D., Shan, F., Li, Z., Le, B. T., Liu, X., & Li, X. J. I. A. (2019). A target detection model based on improved tiny-yolov3 under the environment of mining truck. 7, 123757-123764.
- Xu, Q., Lin, R., Yue, H., Huang, H., Yang, Y., & Yao, Z. (2020). Research on Small Target Detection in Driving Scenarios Based on Improved Yolo Network. *IEEE Access*, 8, 27574-27583. doi:10.1109/access.2020.2966328
- Xu, Z.-F., Jia, R.-S., Sun, H.-M., Liu, Q.-M., & Cui, Z. J. A. I. (2020). Light-YOLOv3: fast method for detecting green mangoes in complex scenes using picking robots. *50*(12), 4670-4687.
- Zhan, J., Wang, J., Ben, Z., Ruan, H., & Chen, S. J. I. A. (2019). Recognition of Angiographic Atherosclerotic Plaque Development Based on Deep Learning. 7, 170807-170819.
- Zhao, X., Wang, X., & Du, Z. (2020). Research on Detection Method for the Leakage of Underwater Pipeline by YOLOv3. Paper presented at the 2020 IEEE International Conference on Mechatronics and Automation (ICMA).
- Wang, Z., Wang, E. & Zhu, Y. Image segmentation evaluation: a survey of methods. Artif Intell Rev 53, 5637–5674 (2020). https://doi.org/10.1007/s10462-020-09830-9
- Liu, Z., Shen, Q., Ma, J. and Dong, Z. (2018), "Research on comment target extracting in Chinese online shopping platform", International Journal of Crowd Science, Vol. 2 No. 3, pp. 247-258. https://doi.org/10.1108/IJCS-09-2018-0019

- Bao,J. & Ye,M.(2017).Head Pose Estimation Based on Robust Convolutional Neural Network. Cybernetics and Information Technologies,16(6) 133-145. https://doi.org/10.1515/cait-2016-0083
- C. Li, Y. Tian, W. Li, J. Tian and F. Zhou, "Low-Contrast Defects Recognition Using Low-Order Residual Network," in IEEE Access, vol. 7, pp. 91193-91201, 2019, doi: 10.1109/ACCESS.2019.2923803.
- T. Niu and Y. Hou, "Density Matrix Based Convolutional Neural Network for Click-Through Rate Prediction," 2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD), 2020, pp. 46-50, doi: 10.1109/ICAIBD49809.2020.9137448.
- R. Zhang, X. Li, L. Zhu, M. Zhong and Y. Gao, "Target detection of banana string and fruit stalk based on YOLOv3 deep learning network," 2021 IEEE 2nd International Conference on Big Data, Artificial Intelligence and Internet of Things Engineering (ICBAIE), 2021, pp. 346-349, doi: 10.1109/ICBAIE52039.2021.9389948.