Harbin Institute of Technology
Undergraduate Graduation Design (Thesis)

Research on Maize Leaf Counting Based on Computer Vision

Li Daxin
Institute (Department): Computer Science and Technology Major: Computer Science and Technology
Student Number: 1170300825 Supervisor: Liu Xianming Graduation Project
in June 2021 (
Thesis)
Title: Maize Leaf Counting Based on Computer Vision
Computer Science and Technology
Student No. 1170300825 Student
Li Daxin Supervisor Liu Xianming
Date of Defense June 20, 2021

Abstract

Food security is an important issue in every country in the world The most important thing, and as one of the world's three major food and agricultural products, the
breeding and cultivation of corn is an important key link to ensure food security. In recent years, with the advancement and development of science and technology,
plant phenotype research has developed rapidly. In order to get rid of the time-consuming and labor-intensive characteristics of traditional phenotyping research, high-throughput plant phenotyping techniques are gradually emerging. High-throughput plant phenotyping technology can quickly analyze plant
pictures . It can play
an important role in precise breeding and growth monitoring of plants, and has great application prospects.
The rapid development of computing hardware equipment, image processing technology, and deep learning technology has made the combination of artificial intelligence and
agricultural production a general trend. This paper studies the high-throughput phenotyping technology based on computer vision,
designs a corn plant data set in the laboratory environment, and implements the data set by image processing and computer vision.
Phenotypic extraction of leaf number counts in maize plants. In this paper, the computer vision-based corn leaf counting research is carried out from
three .
In this paper, we first model the problem of counting the number of leaves in corn plants, and define the problem of counting corn leaves
as the detection problem of corn leaf tips. For the target detection algorithm of corn leaf tip, due to the large size of the
corn , and the small area of the corn leaf tip label, this paper uses the feature pyramid
method to extract the photo features; considering the amount of calculation, In order to achieve fast training and reasoning work, the algorithm refers to
RetinaNet using a one-stage method, using a fully convolutional network as the head of the regression and classification tasks, and adding a
framework of Focal Loss to achieve detection; because there are generally positive and negative samples in the one-stage target detector The imbalance leads to
the obvious gap with the two-stage target detector. The algorithm uses an adaptive positive and negative sample division method to
divide the positive and negative samples of the anchor points on each layer of the pyramid; for the problem that the leaf tips may overlap, the algorithm uses multiple This method of
feature fusion fully combines context information to realize the guiding
role of blade orientation for blade tip position prediction. Experiments have proved that the method

proposed in this topic for the problem of target detection of corn leaf tips
has a good effect on improving the evaluation index AP.

Keywords
: maize phenotype analysis; maize leaf count; leaf detection; target
detection
y ear s. Amongl ot sofpl ant sandasoneoft hr eemaj orf oodcr opsi nt heworld ld,
t hebr eedi ngandpl ant i ngofmai zear et hekeyl i nkt omai nt ai nf ood
security . Wi t ht hepr ogr essofsci enceandt echnol y pe
t echnol ogyhasdev el opedr api dl y .I nor dert ogetr i doft het i me- consume ngand
l abor - consume ngchar act er i st i csoft r adi ti onalphenot y pi ngr esear ch,high gh- t
hr oughtputpl antphenot y pi ngt echni queshav egr adual l yemer ged.High gh-
t hr oughtpl antphenot y pi ngt echnol ogycanqui ckl yanal y zepl antpi ct ur esand
obt ai nal ar geamountofphenot y pi cdat awi t hr esear n chvash alt uei me.I t
canpl ayani mpor t antr ol ei npr eci sepl antbr eedi ngandgr owt hmoni t or i ng,and
hashugeappl i cat i onpr aspect s.
I nt hi paper ,t hepr obl emofcount i ngt henumberofl eav esofcor npl ant s
i smodel edf ir st ,andt hepr obl emofcount i ngl eav esofcor ni sdef i nedast he
pr obl emofdet ect i ngcor nl eaft i ps.Ai mi ngatt het ar getdet ect i onal gor it hmf or
cor nl eaft ip,duet ot hel ar gesi zeofcor nl eafphot osi nt hel abor at or y
env ir onment ,andt hesmal l ar eamar kedont hecor nl eaft ip,t hi paperusest he
f eat ur epy r ami dmet hodt oext r actt hephot of eat ur es;consi der i ngt heamount
ofcal cul at i on,Toachi ev ef astt r ai ni ngandi nf er encewor k,t heal gor it hmr ef er s t
oRet i naNetusi ngaone- st agemet hod,usi ngaf ul l conv ol ut i onalnet wor kas
t her egr essi onandcl assi fi cat i ont askhead,pl ust heFocalLossf r amewor kt o
achi ev edet ect i on;because set heone- st aget ar getdet ect ori sgener al l ycausedby
t hei mbal anceofposi tiv eandnegat iv esampl es ,t hedi ff er encebet weent he

t wo- st agedet ect orandt het wo- st agedet ect ori sobv i ous.Theal gor it hmuses
anadapt iv eposi tiv eandnegat iv esampl edi vi si onmet hodt odi vi det heposi tive
andnegat iv esampl esf r omt heanchorpoi nt sonachl ay eroft hepy r ami d
f eat ur es;f ort hepr obl emofpossi bl eov er l apoft hel eaft i ps,t heal gor it hmuses
av ar i et yoff eat ur ef usi onmet hodwhi chf ul l ycombi nescont ext uali nf or mat i on t
or eal i zet hegui di ngef f ectofbl adeor i ent at i onont hepr edi ct i onofl eaft ip
posi ti on
. shav epr ov edt hatt hemet hodpr oposed nt hi st opi cf ort he
pr obl emsi nt heobj ectdet ect i onofcor nl eaft i phasagodef f ectont he i
mpr ov ementoft heev al uat i oni indexAP.
Keywor ds:Measur i ngPl y antPhenot pi ngTr ai ts,Cor nLeafCount ,Leaf
Detect i on, Object ect ect i onCatalogue
Abstract
I
Abst r act II
Chapter 1 Introduction 1
1. 1 Subject background and purpose and significance of research 1
1. 2 Internal and external Research Status and Analysis of Direction 2
1. 2. 1 Machine Learning and Deep Learning 2
1. 2. 2 Plant Phenotype Research 2
1. 2. 3 Maize Phenotype Research 3
1. 3 Main Research Content and Main Contributions of this Paper 4
1. 4 Organizational structure of this paper 5

Chapter 2 Object Detection Framework . . . .

. . . .
.
. . . .-6-
2. 1. 1 There are anchor detection methods. . . . .-6-
2. 1. 2 Anchor-free detection methods . . . .-6-
2.2 Maize Leaf Tip Phenotype Dataset . . . . -7-
2. 2.1 Data Acquisition. . . . .-7-
2. 2. 2 Blade tip marking. . . . .-8-
When labeling datasets, target boxes of different sizes are used for labeling. The size of the target frame
is mainly controlled at about 30x. The target frame area distribution of the dataset after labeling is shown in , where the abscissa represents the area of the bounding box, and the ordinate represents the statistical number of the dataset.
The two pictures in the first row are marked when the leaves are occluded or overlapped. In this case, if there is no
obvious tip texture, it will not be marked, and if the shape of the tip can be distinguished, it will be marked.
The two pictures in the second row are examples of malnourished leaves. For this kind of leaves, because
the leaf tip and the texture of the leaf cannot be distinguished, they are not labeled to avoid mixing noise in the data set. . . . .-8-
2.3 Object Detection Framework. . . . . - 10-
2.3.1 Skeleton network. . . . . - 11-
2.3.2 Task Head Network. . . . .
-12-2 .4 Loss function. . . . .-13-
2.4.1 Classification loss function. . . . . -13 -
2.4.2 Regression loss function. . . . . -14- 2.5 Evaluation indicators
. . . . -15-
2. 6 Chapter Summary. . . . .-15-

Chapter 3 Leaf tip detection method. . . . .-16-
3.1 Network structure. . . . .-16- 3.2 Adaptive positive and negative sample division. . . . .
. . . .-16- 3.3 Context information fusion. . . . .
. . . .-18- 3. 4 Centrality branches. . . . .
. . . . -19- 3.5 Experiments. . . . . . .
. . . .

. . . . . -20- 3
. 5.1 Comparison of testing standards. . . . . -21- 3. 5. 2 Display of test results
. . . . -22 -3. 6
Summary of this chapter. . . . .
. . . .
-25-
Conclusion . . . . -26-
References . . . . -27-
original Disclaimer. . . . . -30
- Acknowledgments. . . . . -31 -

Chapter 1 Introduction

1.1 Topic background and research Purpose and Significance

Facing the challenges of resource scarcity, climate change, and global population growth, crop yields and quality will need to be increased in a sustainable . Genetic improvement through breeding is the best . With the rapid development of functional genomics, more and more crop genomes have been sequenced and dozens of genes affecting key agronomic traits have been identified. However, due to the lack of crop phenotypic data, current genome sequence information has not been fully exploited to understand the complex characteristics of polygenes. Efficient, automated and accurate technologies and platforms that capture phenotypic data that can be correlated with genomic information for important as genotyping. Therefore, high-throughput phenotyping has become a major bottleneck restricting crop breeding . Plant phenomics has been defined as the high-throughput and accurate acquisition and analysis of multidimensional phenotypes of crop growth stages at the organism level, including cell, tissue, organ, individual plant plot and field levels. Crop phenotypic expression involves complex interactions between genotypes and environmental factors, including climate, soil factors, abiotic/biotic factors and crop management practices. With the rapid development of novel sensors, imaging technologies, and analytical methods , many basic platforms for phenotypic analysis have been developed [1] . Computer vision for extracting useful information from plant images and videos is rapidly becoming an essential technique in plant phenomenology .

The phenology approach to plant science aims to determine the relationship between genetic diversity and phenotypic traits of plant species by using non-invasive and high-throughput measurements of quantitative parameters reflect a plant's lifetime traits and physiology. state. Recent advances in DNA sequencing technology allow us to rapidly obtain population-scale maps of genomic variation. Combining high-throughput analysis platforms for DNA sequencing and plant phenotypic analysis , to use genetic maps to map genotypes to phenotypes, to explore complex quantitative traits in plants (such as growth, environmental stress tolerance, Genetic factors for disease resistance and yield) offer opportunities. Genetic approaches, including quantitative trait locus (QTL) analysis and genome-wide association studies (GWASs).

Furthermore, models of the relationships between genotype/phenotype maps of individuals in breeding populations can be used to compute genomically estimated breeding values to select optimal parents for new crosses selected for genomic selection in crop breeding. Therefore, high-throughput phenotyping assisted by computer vision with the aid of various sensors and image analysis algorithms will play a role in improving crop yields in the context of population statistics and climate change . Crucial role.

The so-called food is the most important thing for the people, and food security is the top priority of our country's national security. Therefore, among many plants , the research value of maize, one of the three major food crops, is extremely valuable. According to the announcement of the National Bureau of Statistics on grain production data in 2020 , corn production in 2020 will account for 38.9% of China's total grain production. At the same time , in addition to being used as food (food or feed), corn also has huge economic value. Corn is an important raw material for fuel, medicine, chemical industry and other industries. At the same time, corn can be processed, corn starch, edible alcohol, etc. Both products are products of corn reprocessing. In addition, maize is a typical gramineous plant with a wide range of gramineous plants (wheat, rice, sorghum and other food crops are basically gramineous plants), and the method of studying and observing the phenotype of maize can be easily extended to other crops. In the study of undergraduate plants. Phenotyping maize can be used for precision breeding on the one hand and growth monitoring of maize plants on the other hand. By comparing the phenotype data with the known data, the genotype and phenotype can be correlated and analyzed, so as to screen the functional genes that are highly correlated with excellent traits, and then breed new varieties that meet specific needs, so as to obtain optimal High-quality and high-yielding crop varieties; through the dynamic monitoring of crop phenotype changes , planting management strategies can be formulated in real time and personalized , and if there are diseases and insect pests, they can also be predicted and discovered in time, making agricultural operations more accurate and efficient. Harbin Institute of Technology Undergraduate Graduation Project (Thesis) 3 1. 2 Research status and analysis in this direction at home and abroad 1. 2. 1 Machine learning and deep learning Machine learning is a data analysis method that can automatically identify internal patterns. Data can be tables , images, text and other data. The process of machine learning entails building a model on an initial data set (called the training data set) , and then using an independent data set (called the test set) to verify the model's performance on data that was not used training. Deep learning has a strong abstract expression ability. Deep neural networks belong to a class of general fitting methods, which means that no matter what function we want to learn, it can be fitted with a deep learning network [2] . Deep learning models automatically perform feature extraction on input data without using any manual feature input. Object detection, as one of the research fields of computer vision, can analyze the concept . As such, the purpose of object detection is to locate objects in a given image and determine which category each object belongs to. Traditional object detection methods first extract feature descriptors, such as HOG and SIFT. Then, based on the extracted feature descriptors, they train classifiers such as support vector machines (SVM) and AdaBoost to distinguish the target object from all other categories. Recently, deep learning based object detection methods have been proposed . These methods, Si ngl e- ShotDetection (SSD) [3] , YouOnlyLookOnce (YOLO) [4] , and Faster-R-CNN [5] use neural networks to automatically extract the necessary image features, and traditional They greatly improve their accuracy compared to object detection methods. However , these methods also consume data and are computationally intensive to train. 1.2.2 Plant Phenotyping Research There is no shortage of use cases in agriculture when it comes to applying machine learning, image processing, and deep learning to object detection . Traditional image processing-based methods, commonly referred to as image segmentation, have been applied to mango, apple, tomato, and grape for detection and counting in images [6] . Despite the success of these methods, they usually require a large number of high-resolution images with minimal image noise, cannot handle large , and each image can only recognize Produce a crop. Ok et al. used a machine learning approach [7] to demonstrate that the random forest (RF) algorithm and maximum likelihood classification were indeed applicable to the successful classification of wheat, rice, corn, sugar beets, tomatoes, and peppers in fields using satellite imagery . Furthermore, Zawbaa et al. [8] designed an experiment to automatically classify images of apples, strawberries, and oranges using RF and the k-nearest neighbor model [9] . Their research further demonstrates the success of machine learning capabilities in agriculture. Furthermore, Guo et al. [10] applied quadratic SVM to accurately detect and count sorghum heads from unmanned aerial vehicle (drone) images. While these examples show the power of modern machine learning in object detection, especially in agriculture, they are not without flaws, namely limitations of traditional machine learning methods that do not generalize well to objects with Objects with different image resolutions, different image scaling (distance from camera to object) and different object orientations (object angle). Due to the ability of deep learning, the ability to recognize multiple objects in an image, and the lack of object-oriented requirements, there is a large literature on the application of deep learning in the field of agriculture recently. In 2019, Ghosal et al. used their RetinaNet-based method to detect and count sorghum heads from drone images [11] . This deep learning approach significantly outperforms previous work on sorghum detection and counting by Guo et al. With the advent of transfer learning , it is possible to pre-train the model on such datasets and transfer the information of the model to detect similar objects without requiring a long training time [12] . 1. 2. 3 Research on maize phenotype In recent years, a lot of work has emerged in the research of image processing technology for maize phenotype. The phenotypic characteristics of maize plants and maize itself include seeds, leaves, ears, etc. In phenotyping studies on maize seeds, single maize seed images are mainly used as input. Paul sen Harbin Institute of Technology Undergraduate Graduation Project (Thesis) 5 et al [13] used traditional image processing methods to determine the length, width and projected area of corn kernels. Ni et al. [15] used deep learning methods to classify corn seeds. Khaki et al. [14] used a convolutional sliding window method to realize the counting of corn kernels on corn ears. The proposed method can successfully detect corn kernels with a low detection error, and can also be used in Kernels detected on a batch of ears of corn positioned at different angles . The positioning of corn ears is the key to solve the problem of automatic corn threshing. Kurt ulmus et al. [19] used SVM to locate ears of corn in the field. Wang et al. [20] obtained images under different viewing angles by rotating the ear of corn , and used binocular stereo vision to calculate the point cloud of the ear of corn to perform 3D reconstruction of the ear of corn. Lu et al. [21] used pixel-level annotations to locate corn ears in the field and regress the number of corn ears at the same time. There is a wide range of research on the properties of maize leaves, and there are different research emphases for different problems. Aiming at the problem of plant center location, Ji a et al. [16] used the top-view RGB image to locate the plant center by detecting, tracking and expanding the main vein line of the leaf. For the corn leaf counting problem, Miao et al. [17] used deep learning to solve the leaf counting problem, and used simulated images to assist in training the leaf counting method , making the counting accuracy higher. In addition, leaves are an important organ to identify whether there is drought or disease in maize. Ahmad et al. [18] used the RGB features and HIS features of maize leaves to describe maize leaves caused by water and nitrogen and normal physiological growth of plants. The stress level caused by the color change , and the classification features were extracted based on the HIS features for identification. Harbin Institute of Technology Undergraduate Graduation Design (Thesis) 6 1.3 The main research content and main contribution of this paper The main research content of this paper is the counting of corn leaf tips based on the target detection algorithm . The input of the algorithm is the RGB corn image in the laboratory environment, and the output is the target box of the leaf tip . as the picture shows. The main contributions of this paper are as follows: 1) A certain scale of data was collected and manually screened to form a data set in a laboratory environment, and a labeling scheme that is practical and convenient to implement the target detection algorithm was formulated. Corn Leaf Tip Detection Dataset for Object Detection Algorithms. 2) Model the corn leaf detection problem as an object detection problem for the corn leaf tip. At the same time, a small target detection framework is proposed to solve the problem that the marked leaf tip area in high-resolution images is too small, which effectively improves the effect of leaf tip detection. 1.4 Organizational structure of this paper The first chapter is the introduction. Firstly, it introduces the research background of this topic, and the research significance in the two fields of agricultural production and computer vision algorithm. Then it introduces the research status of related research content, including the development trend of computer vision and plant phenomics related to this topic, the research status of maize phenotype analysis, and the analysis method for the trait of maize leaf tip status quo. Finally, the main content and contribution of this topic are described . The second chapter mainly introduces the object detection framework. Firstly, the data collection method is introduced, and then the labeling strategy is formulated around the problem of corn leaf tip occlusion, and the statistical results of the labeled data set are displayed . Afterwards, related work in the field of target detection is introduced. The field of target detection is mainly divided into anchor-based target detectors and anchor-free detectors. Among them, anchor-based detectors are divided into one-stage and two-stage two types. There are two types of anchor-free detectors: keypoint-based and centerpoint-based detectors. Then, this paper introduces the network structure used, including the skeleton network and the task head network, and then explains the design of the classification loss function and regression loss function respectively. Finally, the evaluation metrics . The third chapter designs and selects the network structure and optimization method for the blade tip detection method. In order to solve the problem of unbalanced positive and negative samples, the algorithm uses an adaptive positive and negative sample division method to divide the positive and negative samples on the anchor points on each layer of the pyramid; in view of the problem that leaf tips may overlap, the algorithm uses the pyramid pooling method combined with The context information enriches the high-level semantic representation; the centrality branch is introduced . At the end of this chapter, the experimental design is introduced, and the experimental results are analyzed. The conclusion part is a summary of the full text, and at the same time, the future research direction of this topic is expounded. Harbin Institute of Technology Undergraduate Graduation Design (Thesis) -8- Chapter 2 Target Detection Framework This chapter mainly explains the target detection framework. Firstly, the relevant work . The current target detection methods are mainly divided into two types: anchor points and anchor-free methods. Among them, the anchor point method is divided into two stages and one stage; the anchor point method is divided into two types: There are two types based on center point and key point. To implement the object detection framework, a leaf tip detection dataset needs to be designed. Then introduce the one-stage target detection framework used in this topic . The one-stage detector is functionally divided into two parts: feature extraction and leaf tip detection. The network structure includes a skeleton network for feature extraction and classification and regression for downstream detection tasks. network. At the end, the classification loss function and regression loss function used are introduced respectively. 2.1 Object detection related work 2.1.1 Anchor point detection method Two-stage method: The emergence of FasterR-CNN [5] established the dominance . The network structure of FasterR-CNN includes a region of interest extraction network and a prediction network R- CNN. First, the region of interest is used to screen out a large number of anchor points, and the positive and negative samples are divided, and then the task head network for regression and classification. Since then, many works have been improved based on this framework, including modifying the network structure, applying context and attention mechanisms, feature fusion and enhancement, etc. Recently, two-stage anchor-based methods still have leading results on common object detection standards. One-stage method: With the emergence of SSD [3] , the one-stage anchor-based method has also attracted the attention of the industry. This is because compared with the two-stage complex and cumbersome network framework, the one-stage method has a lightweight and simple network framework. Network framework, so it has obvious advantages in computational complexity. SSD introduces multi-scale network layers in a convolutional network , and lays anchor points on multi-scale network layers to directly perform classification tasks and regression tasks. After Harbin Institute of Technology undergraduate graduation project (thesis) -9- , there are also many SSD-based work emerging, such as multi-scale feature context information fusion, design and use new loss functions, refinement and matching of anchor points, etc. Now the one-stage anchor-based method can achieve very close performance to the one-stage two-stage anchor-based method with faster inference speed. 2. 1. 2 The anchor-free detection method is based on the key point method: the key points used in this method are obtained through a predefined or pre-learned method, and then a prediction frame is generated to detect objects. Coner Net [29] detects the bounding box of the object through the upper left and lower right key points , and Center Net [30] expands the key point pairs of Coner Net into triple points. RepPoints [31] represent objects with a series of sample points, learn how to enclose the spatial content and point out semantically important local regions. Center-based method: This anchor-free method regards the center point of the object as the foreground to define the positive sample, and then predicts the distance from the positive sample to the four boundaries of the target box. YOLO [4] divides the picture into SxS grids, and the grid containing the center point is responsible for detecting objects. DenseBox [32] defines a filled circular area as a positive sample. FCOS [33] considers all locations inside the object bounding box as positive samples, along with four distances and a novel centroid scoring method to detect objects. 2. 2 Maize leaf tip phenotype data set The target detection method based on deep learning relies on the data set to mark the detection target. Considering that there is currently no public data set for corn leaf tip, this topic first conducts corn leaf tip Production of phenotype datasets . This chapter describes the data collection of the dataset and the leaf tip labeling work respectively. Harbin Institute of Technology Undergraduate Graduation Project (Thesis) -10- 2. 2.1 Data collection Use the maize image data under laboratory conditions provided by the official website of Huazhong Agricultural University [26] , choose the Mai ze_ Ji anbi ngYan&WanNengYang data using RGB images set. This data set is an image of maize plants collected under laboratory conditions using a high-throughput plant phenotyping platform system. The data set labels . For a plant, the data set collects plant pictures in 16 growth stages, and the plant pictures included include both small plants and large plants. Crawl all the pictures displayed in this data set. The uniform is 1028x, and the larger the picture size, the higher the resolution. The total data size obtained through crawling is 31232 (1952* 16) pictures. Among all the 1952 plant data, some plants need to be selected for labeling . The plant selection criteria are as follows: 1) The pictures of the 16 growth cycles of the plant have a complete plant body, that is, there will be no incomplete corn plants. 2) Choose plants with fewer weeds in the pots to avoid introducing unnecessary noise in the dataset. Finally, a total of 400 images of 25 plants were selected for labeling, of which 15 were used as training sets and 10 were used as test sets. The results are evaluated on the test set as the model training results. The statistical data of the data set and the labeling results are shown in Table 2-1. It can be seen from the figure that there are about 7 target frame labels . Harbin Institute of Technology Undergraduate Graduation Project (Thesis) -11- Table 2- 1 Maize leaf tip labeling data set statistics plant number collection number image number labeling target frame number crawling total data 1952 16 31232 - training set labeling 15 16 240 1417 Test set labeling 10 16 160 923 2. 2. 2 Leaf tip labeling When labeling the dataset, target boxes of different sizes are used for labeling. The size of the target frame is mainly controlled at about 30x. The distribution of the target frame area of the dataset after labeling is shown in Figure 2-1, where the abscissa represents the area of the bounding box, and the ordinate represents the statistical number of the dataset. The two pictures in the first row are marked when the leaves are occluded or overlapped. , for this case, if there is no obvious tip texture, it will not be marked , and if the shape of the tip can be distinguished, it will be marked. The two pictures in the second row are examples . For this kind of leaves, because the leaf tip and the texture of the leaf cannot be distinguished, they are not labeled to avoid mixing noise in the data set. It can be seen from the figure that the area of the label frame is concentrated around 500-800. Compared with the image size of 1028x, the area of the target frame is very small. Therefore, a challenge in the research of leaf tip target detection is the problem of small target object detection. Harbin Institute of Technology Undergraduate Graduation Design (Thesis) -12- Figure 2-1 The actual labeling process of the target frame area distribution of the self-made dataset follows the following principles: 1) Mark the tip of each leaf as much as possible 2) For the case where the leaf tip is blocked, only mark the leaf tip where the texture of the leaf tip can be clearly seen. 3) For the slender leaves on the 2D image, the annotation should be as large as possible so that the leaves have the basic shape of the leaf tip. For some stunted leaves, the shape is slender and the color is yellowish brown. For such leaves, if they are too slender and difficult to distinguish, considering that this situation is not very common and the data set is small, it will not be marked. An example of labeling is shown in Figure 2-2, and examples of specific labeling details and descriptions are shown in Figure 2-3: Harbin Institute of Technology Undergraduate Graduation Project (Thesis) -13- Figure 2-2 Example of target box labeling for undergraduate graduation project of Harbin Institute of Technology (Paper) -14- Figure 2- 3 Specific examples of labeling details 2. 3 Target detection framework Based on the consideration of transferring follow-up work to embedded development equipment, the designed network needs to be as light as possible to achieve a balance between real-time performance and accuracy , consider the structure of a joint feature fusion network with a simple task head using a skeleton network. The specific structure is shown in Figure 2-2 below: Figure 2-2 One-stage detector network [28] framework Harbin Institute of Technology Undergraduate Graduation Design (Thesis) -15- The feature extraction module uses the form of a skeleton network plus a feature pyramid, and the subsequent detection work Applying the one-stage framework, connect the task sub-network directly after each layer of features for regression and classification. The following chapters introduce the network structure in detail. 2.3.1 Skeleton Network The skeleton network plays the role of extracting high-level semantic features from the input image. The input of the skeleton network is a picture of a uniform . Due to the need for random gradient descent to update the network weights, the input of the skeleton network is often organized into a small batch of tensors of the same size, and its shape can be represented as (bat ch_ si ze, num_ channel s, H, W), where bat chsi ze is the batch size, numchannel s is the number of image channels , H and W represent the uniform height and width after cropping. The output of the skeleton network is a multi-dimensional tensor obtained after the convolutional neural network operation, and its shape is (bat ch_ si ze,num_ f eat ur e_ channel s,H ' ,W '), where numf eat ur echannel s is After passing through ' and W ' represent the downsampled size obtained through the convolutional neural network . In the experiment, unless otherwise specified, all skeleton networks use Resnet S as the skeleton network. Feature Pyramid Networks (FPN [27]) can use the information output by each layer of the network to generate a multi-layer feature representation. On the basis of the traditional convolutional neural network, FPN realizes the enhancement of feature expression ability. Through the method of feature fusion, it can better represent semantic information of each level, even the low-level semantic information can also take into account the high-level semantic information. As shown in Figure 2-3, the feature pyramid network mainly includes: the bottom-up path, that is, the skeleton network generates features of different dimensions; the bottom-up path is to fuse high-level semantic information with low-level semantic information to enhance Low-level semantic information; side connections reduce the dimensionality of the features extracted by the skeleton network to facilitate feature fusion on the top-down pathway. Harbin Institute of Technology Undergraduate Graduation Design (Thesis) -16- Figure 2-3 Feature Pyramid Network [27] Structure bottom-up path: This path is completed by the skeleton network, which is a common CNN feature that is down-sampled layer by layer from bottom to top , The process of halving the feature size and doubling the number of channels. The lower layer saves low-level speech information such as edges, etc.; the higher layer reflects high-level semantic information such as object outlines, and even categories. Top-down path: The feature output feature map of the upper layer has a smaller size, but preserves higher-level, generalized semantic information . Such high-level information plays a key role . Therefore, when we process each layer of information, we will refer to the high-level information of the previous layer as its input. When implementing it, we need to use upsampling of bilinear interpolation to enlarge the upper layer features to the same size, and then use the result of horizontal connection one by one. Elements are added. The side connection connects two channels. The side connection is realized by 1×1 convolution. The 1×1 convolution can reduce the number of channels without changing the feature size . The number of channels is the same as that of the right channel. 1×1 Convolution can achieve filtering in the channel dimension. In addition, in target detection, before performing downstream tasks, the feature map obtained by each layer needs to be subjected to a 3x convolution. Harbin Institute of Technology Undergraduate Graduation Project (Thesis) -17- 2. 3. 2 The task head network lays anchor points on the multi-layer features generated by the skeleton network, and divides positive and negative . Generally speaking, if the IOU between an anchor point and an actual target box is greater than 0.5, the anchor point is considered a positive sample, otherwise it is considered a negative sample. The anchor point is used to regress and predict an actual target box, and for an anchor point, it can only be used to detect one actual target box at most. After dividing the positive and negative samples, the detected targets are selected for all anchor points according to IOU . Afterwards, downstream regression and classification tasks are performed on the anchor points. There are two types of task heads in the target detection task, namely the classification task head and the target frame regression task head. The task head is connected to the feature output of the feature fusion network, that is to say, each layer of features corresponds to two task heads. Both task heads are implemented using fully convolutional networks. The classification task head is responsible for predicting the category of the anchor point according to the feature map. The network structure is four 3x x i nchannel s convolutional layers, followed by a 3×3×KA convolutional layer, where A represents each position The number of anchor points, K represents the number of predicted categories, the output shape is (batch_ size,KA,Hf, Wf), and Hf, Wf represent the height and width of the feature map. The regression task head and the classification task head are responsible for fitting the anchor point to the actual target frame according to the feature map in parallel. The network structure is four 3×3×i n_ channel s convolutional layers, followed by a 3× 3×3×(4A) The convolutional layer of , 4 represents the amount of transformation to return to, and the output shape is (batch ch_ si ze, 4A, Hf, Wf). It should be noted that although the classification and regression network structures are similar, the parameters are not shared. 2.4 Loss function Harbin Institute of Technology undergraduate graduation project (thesis) -18- The loss function is mainly divided into two parts, namely the classification loss function and the regression loss function. The classification loss function is used as a measure of the prediction accuracy of the target box category, and the regression The loss function is used as a measure of the prediction accuracy of the target box, and the centrality loss function is a measure of the centrality prediction accuracy. The total loss function takes the three into consideration, as shown in the following formula (2- 1): L=L cl s+ (2−1) where λ can be regarded as a Lagrangian multiplier, whose function is to adjust the classification loss and regression Contribution . 2.4.1 Classification loss function In the process of target detection, since a large number of anchor points will be laid, and the actual number of target detection frames is far smaller than the number of , this leads to a lot of positive samples when dividing The anchor points are classified as negative samples, which is the so-called imbalance between positive and negative samples . For the two-stage target detector, the RPN network screens the anchor points to obtain regions of interest (ROIs), which balances the number of positive and negative samples to a certain extent, and the samples processed by the downstream task head are relatively balanced. For the one-stage target detector, the downstream task head needs to directly face the unbalanced positive and negative samples. The imbalance between positive and negative samples makes the accuracy of one-stage object detectors not as good as that of two-stage object detectors. FocalLoss [28] is a loss function on classification tasks, which alleviates the above problems and narrows the gap between one-stage detectors and two-stage detectors. FocalLoss is modified on the basis of cross entropy , the form is as follows: FL(pt)=− (1−) l og( ) (2−2) where pt is defined as follows: =1 p t=1− *h* , (2 −3) Harbin Institute of Technology Undergraduate Project (Thesis) -19- Use 1−pt trade-off, reduce the weight for samples that are easy to classify, that is, samples with high probability, and increase the weight for samples , so that the final model can be better Handle hard examples. 2.4.2 Regression loss function The target of regression is modeled as four indicators, as shown in (2- 4): tx=(−) / , =(−) / *h* , tw=l og(/) , *h*=l og(*h*/ *h*) , / , , , (2−4) where x, y , w, h represent the center coordinates, length and width of the target frame, x, xa, x ⋆ represent the predicted target frame, Anchors and actual target boxes (y , w, h similarly).

This section shows the visualization results of leaf tip target detection, and the displayed pictures are all the results generated by the ATSS PPM method.

Figure 3-3 and Figure 3-4 respectively show the detection results of large plants and small plants when the leaves are clearly visible and unobstructed,

where frame is marked in green, and the prediction score is drawn in red at the position of the prediction frame , the actual target box is

marked . Because of the drawing order, the green color of the prediction frame will be blocked by the actual target frame, but it can be judged whether the detector detects the leaf according to

whether there is a
prediction . It can be seen that whether it is a mature corn plant or a young
corn plant, the leaf tip detection algorithm can identify it very well. For the detection of the leaf tip, it is difficult to achieve absolute
accuracy intuitively. The deviations shown in the figure are all in the In addition, for each prediction frame, its prediction score is
1.0, because there is only one target category in this target detection task, and it is easy for the classification network to
converge .
Although the mAP of the detection algorithm can reach 89.6%, there are still some cases where the algorithm of this topic cannot handle it well
, resulting in misrecognition.
The display and analysis of misrecognition results is shown in Figure 3-5. The color of the picture has changed from the original picture due to the regularization of the picture . Among them, in the first three pictures, the weeds in the subjective view of the flower pot were misidentified, which is mainly because the
texture of the leaf tip of the weed is similar to that of the corn leaf tip. The first picture in the second row misidentified the position of the blade tip, and this type of error
is acceptable in practical applications. The second picture in the second row misrecognized the turning point of the blade, and the picture in the third chapter of the second row did not recognize the
blurred
, covered
, and small areas leaf tip. Figure 3-3 The results of leaf tip inspection - 3. 6 Summary of this chapter This chapter mainly introduces the target detection method for leaf tips, and firstly describes . Then, around the overlap of leaves in the actual problem, we optimized along the idea of joint context information for feature fusion , and added a pyramid pooling layer after the last layer of features of the skeleton network to fully integrate the context information of the last layer of features. The centrality branch is added to alleviate the problem that the detection results deviate from the actual target box. Finally, an experiment . The experiment compared the existing anchor-based, anchor-free, one-stage, and two-stage classic algorithms with the leaf tip detection network designed in this topic, and visualized the detection results. . From the detection index, the leaf tip detection network proposed in this project can achieve higher average accuracy than the two-stage algorithm, and the mAP reaches 89.6%. Combined with the visualization results, the recognition method has a good recognition effect on the leaves of corn plants in the laboratory environment. Although there are certain misunderstandings, they are within the tolerable range. Harbin Institute of Technology Undergraduate Graduation Design (Thesis) -35- Conclusion This paper focuses on the problem of counting corn leaves based on computer vision. This paper starts from three aspects: problem modeling, algorithm research and experimental design. 1. Modeling the computer vision-based corn leaf counting problem The computer vision-based corn leaf counting problem is modeled as a small target detection problem for corn leaf tips. The maize leaf tip annotation data set was constructed , and the phenotypic characteristics of the maize leaf tip were marked by using reasonable annotation indicators that meet the target detection standards. 2. The target detection algorithm for the corn leaf tip. Due to the large size of corn leaf photos in the laboratory environment, and the small area of corn leaf tip labeling, this paper uses the feature pyramid method to extract multi-layer image features; considering the amount of calculation, in order to achieve fast training and inference Work, the algorithm refers to the RetinaNet framework using a one-stage method , using the full convolutional network as the head of the regression, classification, and centrality tasks, and the loss function uses GI OU, Focal Loss, and cross-entropy; in order to achieve better detection results, The algorithm uses an adaptive positive and negative sample division method to divide the positive and negative samples of the anchor points on each layer of features; in view of the possible overlap of leaf tips, the algorithm uses two feature fusion methods to fully combine context information to make predictions. 3. Experimental Design Experiments compare the existing anchor-based, anchor-free, one-stage, two-stage classical algorithms with the leaf tip , and visualize the monitoring results. It can be seen that the leaf tip

detection network can achieve higher average accuracy than the classical algorithm, and the mAP can reach 89.6%. Combined with the visualization results, the recognition method has a good recognition effect on the leaves of corn plants in the laboratory environment . Although there are certain misunderstandings, they are all within the acceptable range. 4. The next research direction is to build a high-throughput maize plant phenotypic analysis system. The system takes the RGB image of the maize plant as input, and can realize other phenotypic characteristics of Harbin Institute of Technology (thesis) -36- maize. Rapid and automatic extraction of leaf semantic segmentation, internode distance, plant height, stem thickness, etc. For the data set , we only marked a small part of the data. If the data set is expanded, the detection and generalization effects should be further improved. Judging from the results of the algorithm implementation, the slender shape of the leaf tip cannot be detected well, and further research is needed. For the detection of corn leaves , it is necessary to build a new data set and train the model. Consider using the method of knowledge distillation to reduce the size of the model in order to achieve low power consumption and real-time characteristics on the embedded platform. At the same time, considering that there are still missed detections in the target detection results , simple manual corrections can be performed on the analysis results generated by the system. References
[1]Song,P. ,Wang,J. ,Guo,X. ,Yang,W. ,&Zhao,C.(2021).High- t hr oughtput phenot y pi ng:B eaki ngt hr ought hebot tl enecki nf ut ur ecr opbr eedi ng.TheCr op Journal ,xxxx.ht t ps: // doi . or g . 1016/ j . cj . 2021. 03. 015 [2]Goodf el l ow,I . Bengi o,Y. ;Cour ville,A.Deepl ear ni ng;MI Tpress,2016. [3]Li u,W. ;Anguel ov ,D. ;Er han,D. ;Szegedy,C. ;Reed, S. ;Fu,CY ;Ber g,ACSsd: Si ngl eshotmul ti boxdet ect or .Eur opeanconf er enceoncomput erv i si on.Springer , 2016,pp.21–37. [4]Redmon,J. ;Di vv al a,S. ;Gi r sh ck,R. ;Far hadi ,A.Youonl yl ookonce:Uni fi ed, r eal ti meobj ectdet ect i on.Pr oceedi ngsoft heI EEEconf er enceoncomput erv i si on andpat t er nr ecogni ti on,2016,pp.779–788. [5]Gi r shi ck,R.Fastr - cnn.Pr oceedi ngsoft heI EEEi nt er nat i onalconf er enceon comput erv i si on,2015,pp. [7] _ Ok,AO ;Akar ,O. ;Gungor ,O.Ev al uat i onofr andomf or estmet hodf oragr i cul t ur al cr opcl assi fi cat i on.Eur opeanJournal of Remote eSensi ng ,45,421–432. [8]Zawbaa,HM ;Hazman,M. ;Abbass,M. ;Hassani en,AEAut omat i cf r ui t cl assi fi cat i onusi ngr andomf or estal gor it hm.201414t hI nt er nat i onalConf er ence Harbin Industry Undergraduate Graduation Design (Thesis) -37- onHy br i dI nt el li gentSy st ems.I EEE,2014,pp.164–168. [9]Cover ,T. ;Har t ,P.Near estnei ghborpat t er ncl assi fi cat i on.I EEEt r ansact i onson i nf or mat i ont heor y ,13,21–27. [10] Guo,W. ;Zheng,B. ;Pot gi et er ,AB ;Di ot ,J. ;Wat anabe,K. ;Noshi ta,K. ;Jor dan, DR ;Wang,X. ;Wat son,J. ;Ni nomi ya,S. ;Chapman,SCAer i alI mager yAnal y si s – Quantifyi ngAppear anceandNumberofSor ghumHeadsf orAppl i cat i onsi n Breedi ng and Agronomy .Front i er sin Pl antSci ence L,9,1544. doi : 10. 3389/ f pl s. 2018. 01544. [11] Ghosal,S. ;Zheng,B. ;Chapman,SC ;Pot gi et er ,AB ;Jor dan,DR ;Wang,X. ;Singh,AK ;Singh,A . ;Ni nomi ya,S. ;ot her s.Aweakl ysuper vi seddeep l ear ni ngf r amewor kf orsor ghumheaddet ect i onandcount i ng.Pl antPhenomi cs 2019,2019,1525874. [12] aTan,C. ;Sun, F. ;Kong,T. ;Zhang,W. ;Yang,C. ;Li u,C.Asur v eyondeep t r ansf erl ear ni ng.I nt er nat i onalconf er enceonar tifi ci alneur alnet work ks.Spr i nger , 2018,pp.270–279. [13] PAULSENMR,WI GGERW,LI TCHFI ELDJB,etal .Computeri mageanal y sesf or det ect i onofmai zeandsoy beanker nelqual it yf act or s[J]. t ur al Engi neer i ngResearch,1989,43: 93- 101. [14] Khaki ,S. ,Pham,H. ,Han,Y. ,Kuhl ,A. ,Kent ,W. ,&Wang,L.(2020) . Conv ol ut i onalneur alnet work ksf ori mage- basedcor nker neldet ect i onand count i ng.Sensor s(Swi t zer l and) ,20(9) ,1–14.ht t ps: / / doi . or g . 3390/ s [15] NIC,WANGD,VI NSONR,etal .Aut omat i ci nspect i onmachi nef ormai ze ker nel sbasedondeepconv ol ut i onalneur alnet work[J] .Bi osy st emsengi neer i ng , 2019, 178: 131- 144. [16] JI AJ,KRUTZG.Locat i onoft hemai zepl antwi t hmachi nev i si on[J].Jour nalof agri cul t ur alengi neer i ngr esear ch,1992,52 : 169- 181. [17] MI AOC,HOBANTP,PAGESA,etal .Si mul at edpl anti magesi mpr ov emai ze Harbin Institute of Technology Undergraduate Graduation Project (Thesis) -38- l eafcount i ngaccur acy [J].bi oRxi v ,2019: 706994. [18] AHMADIS,REI DJF.Ev al uat i onofcol ourr epr esent at i onsf ormai zei images[J] .Jour nalofAgr i cul t ur alEngi neer i ngResearch ch,1996,63(3) : 185- 195. [19] KURTULMUŞF,KAVDI RI .Detect i ngcor nt assel susi ngcomput erv i si onand support tv ect ormachi nes[J] .Expert Sy st ems with Appl i cat ions,2014, 41(16) : 73907397. [20] WANGC,GUOX,WUS,etal .Thr eedi mensi onalr econst r uct i onofmai zeear basedoncom-put erv i si on. [J] .Nongy eJi xi eXuebaoQTr ansact i onsoft he Chi neseSoci et yf orAgr i cul t ur al Machi ner y ,2014,45(9): 274- 253. [21] LUH,CAOZ,XI AOY,etal .Tassel net :count i ngmai zet assel si nt hewi l

dv ia l ocalcount sr egr essi onnet work[J].Pl antmet hods,2017,13(1): 79. [22] ZHANGX,LIX,ZHANGB,etal .Aut omat edr obustcr op- r owdet ect i oni nmai ze f i el dsbasedonposi ti oncl ust er i ngal gor it hm andshor t estpat hmet hod[J] . I mageanal y si sof anat omi calt r ai t si nst al kt r ansect i onsofmai zeandot hergr asses[J] .Pl ant Met hods,2015,11(1): 26. [24] BAWEJAHS,PARHART,MI RBODO, etal .St al knet :Adeepl ear ni ngpi pel i ne f orhigh gh-t hr oughtputmeasure mentofpl antst al kcountandst al kwi dt h[C] / / Fi el d and Ser vi ceRobot i cs. Springer ,2018:271- 284. [25] KHANZ,RAHI MI - EI CHIV,HAEFELES,etal .Est i mat i onofv eget at io oni ndi ces f orhigh gh-t hr oughtputphenot y pi ngofwheatusi ngaer i ali magi ng[J].Pl antmet hods , 2018,14(1): 20. [26] Mai zeJi anbi ngYan&WannengYang. ht tp: // pl ant phenomi cs. hzau. edu . l ´ ar ,RossB.Gi r sh ck,Kai mi ngHe,Bhar at hHar i har an, andSer geJ.Bel ongi e.Feat ur epy r ami dnet wor ksf orobj ectdet ect i on.I nCVPR, undergraduate of Harbin Institute of Technology Graduation Project (Thesis) -39- 2017. Harbin Institute of Technology Undergraduate Graduation Project (Thesis) -40- [28] Tsung- YiLin,Pr iy aGoy al ,RossB.Gi r sh ck,Kai mi ngHe,andPi ot r Dol l ´ ar .Focall ossf ordenseobj ectdet ect i on.I nI CCV,2017. [29] HeiLawandJi aDeng.Corner net :Det ect i ngobj ect saspai r ed key poi nt sI nECCV,2018. [30] Kai wenDuan, SongBai ,Li ngxiXi e,HonggangQi ,Qi ng-mi ngHuang, andQiTi an.Cent er net :Key poi ntt ri pl et sf orobj ectdet ect i on.I nI CCV, 2019. [31] ZeYang,ShaohuiLi u,HanHu, Li weiWang,andStephenLi n. Reppoi nt s:Poi ntsetr epr esent at i onf orobj ectdet ect i on.I nI CCV,2019. [32] Li chaoHuang,YiYang,Yaf engDeng,andYi nanYu.Dense-box: Uni fyi ngl andmar kl ocal i zat i onwi t hendt oendobj ectdet ect i on.CoRR, 2015. [33] ZhiTi an,ChunhuaShen,HaoChen,andTongHe.FCOS:f ul ly conv ol ut i onalonest ageobj ectdet ect i on.I nI CCV,2019. [34] Rezat of ighi ,SHetal .“ Gener al i zedI nt er sect i onOv erUnion:A Met ri candaLossf orBoundi ngBoxReg essi on. ”2019I EEECVF Conf er enceonComput erVi si onandPat er nRecogni on (CVPR) (2019):658- 666. [35] Zhang,S. ,Chi ,C. ,Yao,Y. ,Lei ,Z. ,&Li ,SZ(2020) .Bri dgi ngt hegap bet weenanchor - basedandanchor - fr eedet ect i onv i aadapt iv et r ai ni ng sample eel ect i on . [36] Zhao ,H. ,Shi ,J. ,Qi ,X. ,Wang,X. ,&Ji a,J.(2017) .Py r ami dscene par si ngnet work k.Pr oceedi ngs-30t hI EEEConf er enceonComput er Vi si onandPat t er nRecogniti on,CVPR ,2017- Janua,6230–6239. Harbin Institute of Technology Undergraduate Graduation Project (Thesis) Harbin Institute of Technology Undergraduate Graduation Project (Thesis) Original Statement I solemnly declare: In Harbin Industry During the university's bachelor's degree, the graduation design (thesis) "Research on Counting Maize Leaf Based on Computer Vision" is the result of my independent research work under the guidance of my supervisor. Individuals and groups that have made important contributions to the research work of this article have been clearly indicated in the article , and other unmarked parts do not include research results that have been published or written by others, and there is no purchase or writing by others , plagiarism and falsification of data and other fraudulent acts. I am willing to take legal responsibility for this statement. Author's signature: Date: June -30, 2020- Harbin Institute of Technology Undergraduate Graduation Project (Thesis) Acknowledgments Four-year undergraduate, time flies. College life should also come to an end . 2020 will be unforgettable. At the beginning of the year, the COVID-19 epidemic was severe, and I couldn’t go back to school. I had to stay at home and prepare for the rematch of the software competition while taking online classes. Prepare for postgraduate entrance examination. Fortunately, although the postgraduate entrance examination was a long journey, I managed keep clear of the clouds and mist to see the moonlight. I went ashore after the postgraduate entrance examination, and my residence qualification in Harbin Institute of Technology was considered to be continued. First of all, I would like to thank the country and the heroes who fought against the epidemic. Under the leadership of the party, we quickly controlled the epidemic and realized the resumption of work, production and school. It is very important for me to return to school in time to prepare for the postgraduate entrance examination, which is extremely inefficient as long as I study at home. I would also like to thank the government's student loan program, which helped my family relieve a lot of financial pressure. I would like to thank my parents sincerely. They gave me unconditional and reserved support and trust. It is the support of my parents made me devote myself to the postgraduate entrance examination without hesitation. Along the way, the kindness of parents never dare to forget . I wish my parents good health. Thanks to Mr. Liu Xianming for his guidance on my final thesis. During the postgraduate entrance examination, Mr. Liu gave me the greatest understanding and tolerance, and did not urge the progress of my graduation project, so that I had enough review time to prepare for the postgraduate entrance examination. Everything is ready, I only owe Dongfeng, I only hope that I can publish more papers under the guidance of Mr. Liu during the postgraduate period. Thanks to Wang Panpan from the Shenzhen Institute of Genomics, Academy of Agricultural Sciences. She spent a lot of time and did a lot of labeling work in building the laboratory corn plant dataset. At the same time, it was her urging me to continue to advance the graduation project after the preliminary examination of the postgraduate entrance examination. Thanks to my roommates for their tolerance and companionship over the past three years. Thank you Luo Ge for being able to work together . The friendship is unforgettable. I also wish Brother Hua smooth work, and Brother Sun and Brother Biao to make great achievements in the new environment . As for Brother Luo and Brother Zhang, the graduate students are also good roommates. Wish us all a bright future. Thanks to everyone who has helped me along the way. On the way to school, elementary school, junior high school, high school, and college time are all behind us. Although it is the end of a small stage , the future with infinite possibilities is slowly unfolding. What's more, today's crisis coexists, and we should grasp the great changes unseen . We should forge ahead and have a bright future. Share with you all.