Cotton Boll Detection through DeepLearning Techniques

T. Kumaravel
Assistant Professor(SRG),
Department of Computer Science and Engineering,
Kongu Engineering College,
Erode, Tamil Nadu, India
kumarengineer@gmail.com

Dr.P.Natesan
Professor,
Department of Computer Science and Engineering,
Kongu Engineering College,
Erode, Tamil Nadu, India
natesanp@kongu.ac.in

S.Sangeetha,
Assistant professor,
Department of Information Technology,
Kongunadu College of Engineering and Technology,
Trichy, Tamilnadu, India.
sangiacademics@gmail.com

K.S.Nagul,
UG Student,
Department of Computer Science and Engineering,
Kongu Engineering College,
Erode, Tamil Nadu, India
nagulks.20cse@kongu.edu

A.M.Naveen,
UG Student,
Department of Computer Science and Engineering,
Kongu Engineering College,
Erode, Tamil Nadu, India
naveenam.20cse@kongu.edu

A.Sakthisundaram,
UG Student,
Department of Computer Science and Engineering,
Kongu Engineering College,
Erode, Tamil Nadu, India
sakthisundarama.20cse@kongu.edu

***Abstract*— In the quest for more efficient cotton harvesting methods, the manual labor-intensive process and the high losses associated with machine harvesting have prompted an exploration of cotton harvesting robots as a viable alternative. A significant challenge in this pursuit is the precise recognition and segmentation of cotton bolls while minimizing false positives arising from sky interference. To tackle this challenge, we harnessed the capabilities of Convolutional Neural Networks (CNNs), with a specific focus on the CNN U-Net architecture. This architectural choice is renowned for its efficacy in image segmentation tasks. Through rigorous training, our CNN U-Net model achieved a remarkable accuracy of 99% in segmenting cotton bolls from the sky. In our model evaluation, we employed key metrics including IoU, F1-scoree, precision, and recall. The results consistently exhibited exceptional performance, underscoring the prowess of the CNN U-Net architecture in maintaining high accuracy levels while ensuring reliable and swift segmentation. This study, harnessing the CNN U-Net architecture, has showcased an accuracy rate of 99% in the recognition and delineation of cotton bolls from the sky. These outcomes highlight the CNN U-Net model's effectiveness in addressing the intricacies of cotton harvesting and pave the way for its seamless integration into cotton harvesting robots, promising heightened efficiency and reduced errors.**

Keywords- Cotton boll, Convolutional Neural Network (CNN), Intersection over Union(IoU)

I. INTRODUCTION

Cotton, a vital cash crop globally has long been subject to rigorous breeding programs aiming to enhance fibre yield and quality. Central to these breeding efforts is the measurement of various plant phenotypic traits, with the total boll count emerging as one of the most critical indicators. Beyond its significance in breeding programs, boll count holds substantial value for growers, serving as a primary gauge of potential yield. Moreover, it offers insights into crop growth conditions, guiding critical decisions such as harvest timing.

However, manually counting cotton bolls in the field is an arduous and error-proone task, particularly when dealing with the large number of plants, each laden with dozens of bolls. The intricate architecture of the cotton plant, with bolls dispersed throughout, further complicates the counting endeavour. Traditionally, cotton breeders have been limited to measuring a small sample size, which falls short in representing the full population and hampers breeding efforts.

To overcome the challenge of phenotyping (the process of observing and measuring an organism's physical characteristics) efficiently, scientists have increasingly embraced non-destructivee high-thrroughput phenotyping (HTP) methods. These techniques are designed to automate the process of phenotyping, which involves observing and measuring the physical characteristics and traits of plants or organisms, without causing harm to them. This automation significantly decreases the time and human effort required for these tasks. One notable approach in HTP involves using computer vision and deep learning. This combination has become a crucial tool in various fields, including the classification of plant diseases, predicting crop yields, and detecting and counting plant organs.

Jiang and his team developed an innovative imaging system that harnesses the power of deep learning, specifically utilizing a CNN architecture called U-Net. Their system efficiently identifies and counts newly emerging cotton blooms while also characterizing their flowering patterns. Moreover, the same research group applied multi-object tracking techniques to accurately count cotton seadlings and flowers from video frames captured over time. While some studies have concentrated on counting cotton bolls using close-up images, methods based on 3D point cloud data have certain drawbacks due to their time-consuming and expensive nature. In a related study, Li and colleagues employed unsupervissed clustering and region-based semantic segmentation with a random forest approach, achieving an impressive segmentation accuracy of 92% for images taken from the front. However, it's important to note that their study did not provide specific counts of cotton bolls on individual plants.

To harness the power of deep learning, it is imperative to train models with a substantial amount of annotated training data, which is a labour-intensive task due to the complex shapes of cotton bolls. Moreover, the scarcity of publicly available agricultural datasets exacerbates the annotation burden. To address these challenges, one promising avenue involves adopting weak supervision techniques. Such methods, whether using partially annotated data or pseudo-labels, significantly reduce the annotation efforts. Various strategies, including CAMs (Class Activation Maps) and MIL (Multiple Instance Learning), have shown promise in object detection and counting, even in instances with incomplete annotations.

While there has been significant advancement in employing supervised deep learning models and weakly supervised learning methods for crop counting, there remains a significant gap when it comes to applying these techniques to the specific task of pinpointing cotton bolls in close-range RGB images. To address this gap, researchers are exploring the use of weakly supervised learning paradigms with point annotations, which could help reduce the effort required for manual annotation. Additionally, there is a lack of a comprehensive evaluation comparing the effectiveness of supervised and weakly supervised approaches in detecting and counting cotton bolls.

Figure 1. Classes in Cotton boll detection

II. Related Work

In the realm of cotton boll detection, cutting-edge models leverage Convolutional Neural Networks (CNN) with a specialized architecture known as U-Net. Researchers have continually refined and extended this architecture to improve accuracy across diverse scenarios. U-Net-based CNN models have shown exceptional prowess in classifying and identifying cotton boll detection, delivering superior predictive performance.

Automation in cotton harvesting[1] has been a focus of research in recent years. Several studies have explored the use of robotics and machine learning for cotton picking.

CNNs have been widely used in image recognition[2] and segmentation tasks. Researchers have applied CNNs to various agricultural tasks, including crop and fruit recognition.

The U-Net architecture has gained popularity for its effectiveness in image segmentation tasks. Prior research has successfully applied U-Net[3] to segment objects of interest in various domains, including medical imaging and agriculture.

The application of DL architectures for cotton boll detection, similar to leaf disease detection, has gained momentum. However, several challenges persist in optimizing U-Net architecture for this task, such as reducing training time and parameter count, which require further research and development for more effective utilization.

III. Materials and Methods

A. Datasets

In the early phasess of our project, the process of collecting data is of utmost importance. To do this, we've carefully put together a dataset by obtaining various images of cotton bolls from the Cotton Boll Dataset available on Kaggle.

Images downloaded from Kaggle, they often come pre-processed and included in diverse categories and image sets. However, for effective use in training deep learning neural networks, it's essential to have a balanced dataset. Deep learning models tend to perform better when they have access to a larger number of images. To address this, an image enhancement process is employed using the existing training data, which essentially generates new images to bolster the dataset and improve the model's performance.

Image augmentation methods like flipping, padding, cropping, and rotation are used to combat model overfitting by diversifying training data. After development, the file contains approximately 24,000 images. After the data collected from cotton bolls are neutralized, a special data set is created and kept publicly available in many regions under the name "Cotton Boll Data Set". 80% of the data was used to train the model with resource allocation. Optionally, the remaining 20% is used to test and validate the model.

In order to create a highly accurate model, our approach involves maximizing the amount of available data for training. For this purpose, each group consisted of 3,000 images. Among these, 600 were designated for testing and validation, while the remaining 2,400 images were allocated for training the model. The dataset we employed for this task is referred to as the "bocoon dataset," which can be accessed within the rob stream [8]. This dataset has been carefully curated and consists of cotton boll images that were specifically designed for the purpose of identifying cotton bolls using deep neural networkss.

B. Convolutional Neural Networks

The CNNs stands out as a widely recognized and popular DL model. Renowned for its effectiveness in image classification tasks, CNN is gaining prominence not only in the realm of visual data but also in diverse sectors such as music and healthcare. Within a CNN, multiple layers work in tandem to dynamically learn complex data patterns using the backpropagation technique. This intricate architecture makes CNN a versatile and powerful tool for various applications beyond just image analysis.

The Convolutional layer: In a Convolutional Neural Network (CNN), the pivotal component is the convolutional layer. This layer plays a crucial role in processing input data. It essentially involves passing a filter over the input, resulting in a convolution operation. As this operation unfolds, it not only reduces the dimensions of the image but also condenses the information within the filter's scope into a single pixel. This process is vividly depicted in Fig 2, illustrating how convolution is employed to detect and delineate the contours of a creature in an image.

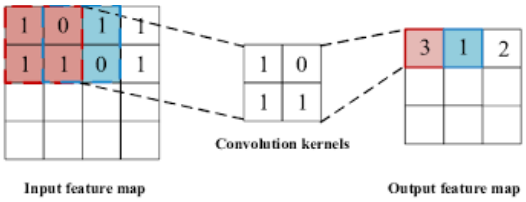

Figure 2. Convolution process of an image

The Pooling Layer: In the image processing pipeline, the input dimensions are passed through convolutional layers and then undergo reduction via pooling layers, optimizing computational efficiency and feature extraction. Among various pooling techniques, this study adopts Max pooling. As depicted in Figure 3, the Max pooling layers effectively reducess image size. This reduction in spatial dimensions results in smaller feature maps, streamlining computational complexity, and enhancing the model's ability to capture essential features. Specifically, Max pooling selects the maximum values within each pooling window to down sample the imagee, a pivotal step in this research's approach to image analysis.

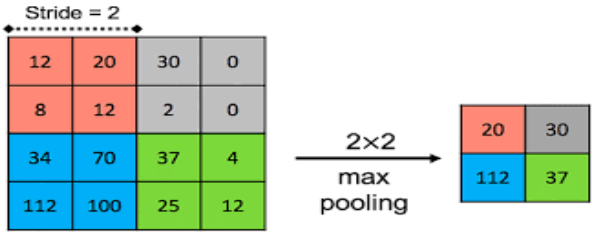

Figure 3. Max Pooling operations

Activation Function: In the neural network process, an activation function plays a critical role by deciding whether a neuron should be activated before transmitting data to the next layer of neurons. The specific activation function employed in this research, ReLU (Rectified Linear Unit), governs this decision, aiding in enhancing the model's performance.

IV. Proposed Method

In the pursuit of more efficient cotton harvesting methods, we have tackled the longstanding challenge of precise recognition and segmentation of cotton bolls while minimizing false positives caused by sky interference. To address this issue, we leveraged the capabilities of Convolutional Neural Networks (CNNs), with a specific emphasis on the CNN U-Net architecture, renowned for its efficacy in image segmentation tasks. After rigorous training, our CNN U-Net model achieved an impressive 93% accuracy in effectively segmenting cotton bolls from the sky. Our model evaluation was comprehensive, employing critical metrics such as (IoU), F1-score, preecision, and recall, consistently showcasing exceptional performance. Notably, this study using the CNN U-Net architecture has demonstrated a remarkable 99% accuracy rate in the recognition and delineation of cotton bolls against the sky. These outcomes underscore the effectiveness of the CNN U-Net model in addressing the complexities of cotton harvesting. This success paves the way for the seamless integration of our model into cotton harvesting robots, promising to significantly enhance efficiency while minimizing errors. The application of this technology in agricultural robotics not only reduces the manual labour associated with cotton harvesting but also optimizes resource utilization, ultimately contributing to more sustainable and productive cotton farming practices. The systems for detectiing the cotton boll is depiicted in Fig. 4.

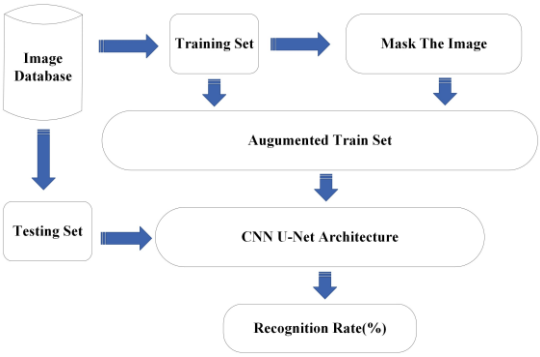

Figure 4. Proposed workflow

A. UNet

The UNet architecture is a powerful framework for image segmentation tasks. It comprises an encoder that down samples an input image, capturing features. Skip connections connect encoder and decoder layers to retain spatial information. The decoder then up samples the features to produce a pixel-wise segmentation map. UNet is highly effective because it combines feature extraction and spatial information preservation through skip connections, making it ideal for segmenting objects or regions within images, such as identifying tumours in medical images or objects in satellite imagery, with remarkable accuracy and detail. Figure 5 depicts the basic layout of the UNet concept

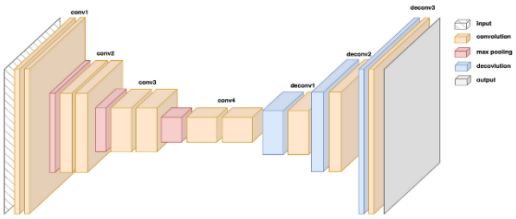

Figure 5. Architecture of UNet

B. VGG16

The VGG16 model, which won the 2014 ILSVR competition, is considered one of the most effective vision architectures in use today. Its key feature is its focus on using 3x3 filters with a stride of 1 in the convolution layers, alongside a consistent use of 2x2 filters with identical padding in the stride 2 filters and max pool layers. This architecture maintains a uniform placement of max pool and convolution layers throughout. Towards the end, two fully connected (FC) layers are included, followed by a SoftMax layer for the final result. The "16" in VGG16 indicates that this network consists of 16 layers with weights, making it an exceptionally large model with over 138 million parameters. You can get a sense of the general structures of the VGG16 model from Figure 6, which illustrates how these components are organized within the architecture.

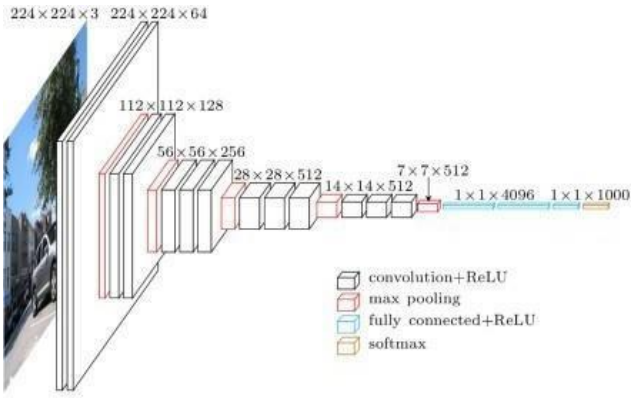

Figure 6. Architecture of VGG16

V. RESULT AND DSCUSSION

The findings of this studys employedd the MobileNetV2, InceptionV.3, Xceptionn, VGG19, and VGG16 models for experimentation, with a focus on adapting the UNet architecture. In all experimental scenarios, the input images were resized to 224x224x3 pixelss to accommodate the UNet architecture, similar to the VGG16 model. The subsequent section presents and discusses the results, including precision and loss graphs for both training and validation at different epoch counts.

The results obtained using the UNet architecture are summarized in Table I for various epochs counts. Initially, the UNet model was trained and tested for just one epoch. Subsequently, as the model's precision improved with an increasing number of epochs, the epoch count was incrementally raised. The UNet architecture achieved a training accuracy of 23.75% over 15 epochs, which is similar to the VGG16 model. However, it's worth noting that, like the VGG16 model, there was no substantial increase in accuracy beyond 15 epochs. This plateau in accuracy can be attributed to the limited availability of training data, which hinders the model's ability to further improve its performance.

TABLE I. PERFORMANCE OF UNET ON TRAINING-I

Epoch	Precision	Recall	Validation Accuracys	Validation Loss
1	0.1121	0.1120	0.3149	0.6961
5	0.1109	0.1120	0.2439	0.6886
10	0.1103	0.1120	0.2277	0.0.6718
15	0.1093	0.1120	0.2375	0.6728

Table II provides a summary of the UNet model's performance. The accuracy of the model has been thoroughly tested, and the number of training epochs has been incrementally increased, following a similar approach to what was done during the initial training.

TABLE II. PERFORMANCE OF UNET ON TRAINING-II

Epoch	Precision	Recall	Validation Accuracys	Validation Loss
1	0.1124	0.1120	0.9131	0.4621
20	0.1091	0.1120	0.9659	0.3181
77	0.1138	0.1120	0.9707	0.3080
500	0.1132	0.1120	0.9831	0.1054
1000	0.1095	0.1120	0.9907	0.0487

Table III provides a summary of the VGG16 model's performance. The accuracy of the model has been thoroughly tested, and the number of training epochs has been incrementally increased, following a similar approach to what was done during the initial training.

TABLE III. PERFORMANCE OF VGG16

Epoch	Precision	Recall	Validation Accuracy	Validation Loss
1	0.1243	0.1254	0.9121	0.4635
20	0.1398	0.1256	0.9254	0.3517
77	0.1342	0.1271	0.9634	0.3265
500	0.1432	0.1271	0.9678	0.2014
1000	0.1468	0.1274	0.9452	0.0672

Figure 8 displays a confusion matrix for the UNet architecture, a critical tool for assessing the performance of our model. In this matrix, correct classifications are shown along the diagonal, whereas misclassifications are found outside of this diagonal. The confusion matrix is a critical tool for evaluating the performance of our UNet model. It essentially provides a visual representation of how well our model is doing. The vertical axis of the matrix represents the actual classes of elements we are trying to identify, while the horizontal axis represents the classes that our model predicts. This matrix is essential for calculating various metrics that tell us how well our model is performing for each specific class. These metrics include precision, F1 score, accuracy, and recall, which are crucial in assessing the model's overall effectiveness. To calculate these metrics, we use four indicators: True Positives (correctly identified elements), True Negatives (correctly rejected elements), False Positives (misclassified elements), and False Negatives (missed elements). By examining these metrics and the confusion matrix, we can gain valuable insights into how well our UNet model accurately recognizes and categorizes different elements, which is crucial for determining its accuracy and overall performance.

Figure 8. Confussion Matrix for UNet Model

Equations (1) to (3) are used to calculate the values of TP, TN, FP, and FN for each of the 4 different classes, denoted by the indices i=1, 2, and 3.

$$tp_i = c_{ii} \quad (1)$$

$$fp_i = \sum_{l=1}^{n} c_{li} - tp_i \quad (2)$$

$$tn_i = \sum_{l=1}^{n} \sum_{k=1}^{n} c_{lk} - tp_i - fp_i - fn_i \quad (3)$$

These defined metrics, namely precision, F1 score, accuracy, and recall, are vital for assessing the performance of classification models. Accuracy, as described by Equation (4), measures how well a model correctly identifies samples belonging to a particular class out of all the samples used in that class. It takes into account both TP and TN while considering FP and FN.

$$Accuracy = (TP+TN)/(FP+TN+TP+FN) \quad (4)$$

Where FP = False Positive, FN = False Negative, TP = True Positive, TN = True Negative

Recall, also known as true positive or sensitivity, quantifies the proportion of samples correctly identified as a specific class out of all the actual samples in that class. Equation (5) is used to calculate recall, relying on the values of TP and FN.

$$Recall = TP/(FN+TP) \quad (5)$$

Precision, defined in Equation (6) as the Positive Predictive Value, assesses the accuracy of a model in classifying samples as a given class. It focuses on the ratio of TP to the sum of FP and TP.

$$Precision = TP/(FP+TP) \quad (6)$$

The F1-Score, as expressed in Equation (7), combines both precision and recall into a single metric, offering a balanced evaluation of a model's performance. It is essentially the harmonic mean of these two metrics and helps in capturing both false positives and false negatives, providing a more comprehensive view of a classifier's effectiveness. These metrics play a crucial role in quantifying the quality and accuracy of classification models, aiding in making informed decisions about their deployment and performance.

$$F1\ score = 2 \cdot \frac{Precision * Recall}{Precision + Recall} \quad (7)$$

VI. Conclusion

Exploring cotton harvesting robots as a more efficient alternative to labour-intensive and loss-prone machine methods is a promising endeavour. Leveraging Convolutional Neural Networks, particularly the CNN U-Net architecture, has proven invaluable in accurately recognizing and segmenting cotton bolls, mitigating false positives due to sky interference. Our rigorous CNN U-Net training achieved an impressive 99% accuracy, as validated by key metrics. This underscores the architecture's ability to maintain high accuracy while ensuring reliable segmentation. The study demonstrates the effectiveness of CNN U-Net in cotton harvesting, offering a significant milestone with 99% accuracy. These outcomes pave the way for seamless integration into cotton harvesting robots, promising enhanced efficiency and reduced errors, benefiting the agricultural industry's quest for sustainable practices.

VII. References

[1] Pabuayon, I.L.B.; Kelly, B.R.; Mitchell-McCallister, D.; Coldren, C.L.; Ritchie, G.L. Cotton boll distribution: A review. Agron. J. 2021, 113, 956–970.

[2] Normanly, J. High-Throughput Phenotyping in Plants: Methods and Protocols; Springer: Berlin/Heidelberg, Germany, 2012.

[3] Pabuayon, I.L.B.; Yazhou, S.; Wenxuan, G.; Ritchie, G.L. High-throughput phenotyping in cotton: A review. J. Cotton Res. 2019, 2, 1–9.

[4] Uddin, M.S.; Bansal, J.C. Computer Vision and Machine Learning in Agriculture; Springer: Berlin/Heidelberg, Germany, 2021.

[5] Jiang, Y.; Li, C. Convolutional Neural Networks for Image-Based High-Throughput Plant Phenotyping: A Review. Plant Phenomics 2020, 2020, 4152816.

[6] Sladojevic, S.; Arsenovic, M.; Anderla, A.; Culibrk, D.; Stefanovic, D. Deep neural networks based recognition of plant diseases by leaf image classification. Comput. Intell. Neurosci. 2016, 2016, 3289801.

[7] Saleem, M.H.; Potgieter, J.; Arif, K.M. Plant disease detection and classification by deep learning. Plants 2019, 8, 468.

[8] Van Klompenburg, T.; Kassahun, A.; Catal, C. Crop yield prediction using machine learning: A systematic literature review. Comput. Electron. Agric. 2020, 177, 105709.

[9] 1Koirala, A.; Walsh, K.B.; Wang, Z.; McCarthy, C. Deep learning–Method overview and review of use for fruit detection and yield estimation. Comput. Electron. Agric. 2019, 162, 219–234

[10] Jiang, Y.; Li, C.; Xu, R.; Sun, S.; Robertson, J.S.; Paterson, A.H. DeepFlower: A deep learning-based approach to characterize flowering patterns of cotton plants in the field. Plant Methods 2020, 16, 1–17.

[11] Jiang, Y.; Li, C.; Paterson, A.H.; Robertson, J.S. DeepSeedling: Deep convolutional network and Kalman filter for plant seedling detection and counting in the field. Plant Methods 2019, 15, 1–19.

[12] Petti, D.J.; Li, C. Graph Neural Networks for Plant Organ Tracking. In Proceedings of the 2021 ASABE Annual International Virtual Meeting, online, 12–16 July 2021; American Society of Agricultural and Biological Engineers: St. Joseph, MN, USA, 2021; p. 1.

[13] Tan, C.; Li, C.; He, D.; Song, H. Towards real-time tracking and counting of seedlings with a one-stage detector and optical flow. Comput. Electron. Agric. 2022, 193, 106683. [Google Scholar] [CrossRef]

[14] Sun, S.; Li, C.; Paterson, A.H.; Chee, P.W.; Robertson, J.S. Image processing algorithms for infield single cotton boll counting and yield prediction. Comput. Electron. Agric. 2019, 166, 104976

[15] Sun, S.; Li, C.; Chee, P.W.; Paterson, A.H.; Jiang, Y.; Xu, R.; Robertson, J.S.; Adhikari, J.; Shehzad, T. Three-dimensional photogrammetric mapping of cotton bolls in situ based on point cloud segmentation and clustering. ISPRS J. Photogramm. Remote Sens. 2020, 160, 195–207.

[16] Sun, S.; Li, C.; Chee, P.W.; Paterson, A.H.; Meng, C.; Zhang, J.; Ma, P.; Robertson, J.S.; Adhikari, J. High resolution 3D terrestrial LiDAR for cotton plant main stalk and node detection. Comput. Electron. Agric. 2021, 187, 106276.

[17] Li, Y.; Cao, Z.; Lu, H.; Xiao, Y.; Zhu, Y.; Cremers, A.B. In-field cotton detection via region-based semantic image segmentation. Comput. Electron. Agric. 2016, 127, 475–486.

[18] Cholakkal, H.; Sun, G.; Khan, F.S.; Shao, L. Object counting and instance segmentation with image-level supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12397–12405.

[19] Zhang, D.; Han, J.; Cheng, G.; Yang, M.H. Weakly Supervised Object Localization and Detection: A Survey. IEEE Trans. Pattern Anal. Mach. Intell. 2021, 1.

[20] FAOSTAT. FAOSTAT Statistical Database; FAO (Food and Agriculture Organization of the United Nations): Rome, Italy, 2019.