International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 09 Issue 10 ǁ 2021 ǁ PP. 20-26

Deep Convolution Neural Network for Detection of Foreign Matter in Tea

Dr. Amitava Akuli[1], Suprito Dey Sarkar[2], Atri Sain[3], Tamal Dey[1], Abhra Pal[1], Gopinath Bej[1], Sabyasachi Majumdar[1], Dr. Nabarun Bhattacharya[1]
[1]Centre for Development of Advanced Computing, Dept. of Agriculture R&D, Kolkata, India
[2]Dept. of Computer Science & Engineering, Adamas University, Kolkata, India
[3]B P Poddar Institute of Management & Technology, Kolkata, India
Corresponding Author: Dr. Amitava Akuli

Abstract
In this paper, we present an automated tea foreign matter detection system using deep convolution neural network. Tea sample was collected from the tea packaging industry where mixing of foreign matters with tea is a common problem. The manual sorting process for cleaning of foreign particles is laborious and time consuming. The mechanical sorting process is also costly and resource consuming. The images of the tea sample were captured using a digital camera under controlled illumination while passing on a vibratory feeder. We applied two deep neural network algorithms (YOLO v3 and Faster RCNN) to train the tea image dataset collected using the developed system. The performance parameters like AP (average precision), average recall and speed of analysis were considered for comparison of the two algorithms. Using Faster RCNN, we calculated average precision, average recall and average frame rate as 0.42, 0.47 and 4 fps and the values obtained using YOLO v3 were 0.91, 0.67, and 15 fps. The solution may be integrated with the tea sorting system for real-time detection and removal of foreign matter in the tea industry.
Keywords: *tea, foreign matter, object detection, convolutional neural network, Faster RCNN, YOLO v3, precision, recall, frame rate*

Date of Submission: 06-10-2021 Date of acceptance: 20-10-2021

I. INTRODUCTION

Tea is a popular consumed beverage in the world which produces from Camellia Sinensis plant (a sub-tropical evergreen plant native to Asia and is grown worldwide). Tea processing steps include plucking, withering, cut-tear-curling (CTC), oxidizing, drying and packaging. Though there are several types of tea like oolong, white and green; black tea is one of the most widely consumed as due to its bright colour, strong liquor and flavour. Tea cleaning process is necessary to maintain its quality and making sure it's not contaminated with foreign matter during tea packaging. There are different types of foreign matter present in tea, like plastic, paper, rubber pieces, agro-based contaminants, organic fickle matters, ferrous and non-ferrous materials like electrical wire pieces, corrosion materials etc. At present, these objects can be removed by various mechanical processes in the industrial scales like perforated tray, magnetic, electrostatic separation, mechanical destoner etc. But there are still chances that few foreign matter remain in the tea even after performing these processes and can cause serious health problems when we consume the tea. Computer vision backed up by deep learning algorithms have been implemented in agriculture industry for monitoring the quality of crops and soil, identifying weeds, and detecting diseases in crops. The application of deep neural networks like Convolutional Neural Networks (CNN) have been reported performing high accuracy in the field of computer vision detection task. CNN's having deep layers are able to capture more concrete representational information of the image. CNN also performs automatic feature extraction unlike conventional machine learning algorithms. For object detection purpose, there are several algorithms like Faster RCNN [1], YOLO [2], and SSD [3]. Among these, Faster RCNN which is an improvement over Fast RCNN, being the most accurate is not fast enough for real-time usage [2]. Application of Faster R-CNN has been reported with significant detection accuracy for object detection in general, but the performance decreases when the detected objects in the images have problems such as deformation, occlusion etc. [4]. But, YOLO and SSD are both capable for real time video use, since they can give us high frame rates and accurate predictions with faster processing time. Manual sorting process for cleaning of foreign matter in tea industry is laborious, time consuming and prone to human mistakes. To overcome the limitations of the manual sorting process, an automated image analysis solution for object detection using deep learning has been proposed to identify the undesirable bits before packaging, so that the consumer can have an untroubled tea making experience. There are a number of screening processes available in the industry like, a) magnetic separator which separates any kind of ferrous material present, b) electrostatic separator which is applied to separate out materials with large difference in conductivity like plastics, fibers, and tea stalks and c) perforated trays to filter out large impurities. All these methods increase the overall cost of the production and even consume large spaces for installation. The problem arises when the size of the foreign matters and tea particles are same and are not possible to eliminate using aforesaid mechanical processes. We propose an automated image analysis system which comprises of a digital camera, an illumination system to capture the images of black tea samples while passing on a vibratory feeder. This paper compares the performance of two deep neural network algorithms namely Faster R-CNN and YOLO v3 for detection of foreign matters in tea.

II. LITERATURE REVIEW

Nipan et al. [5] proposed a technique to predict the moisture loss (ML) in the soil during withering in the tea manufacturing process. They developed a prototype of an enclosed trough to perform the withering of the tea leaves, and the data collected from it was used to train an artificial neural network and predict the moisture loss. Their result showed a minimum difference between the predicted value and the actual ML value, with a max mean prediction error of 3.6%. Yu-Ting et al. [6] proposed a method to identify the plucking points in tea shoot using Faster RCNN object detector. The model was trained to detect OTTL (one tip with two leaves) regions in the image, and achieved an overall accuracy of 84.91%. Y. Hualin et al. [7] implemented YOLO v3 to detect the plucking points of tender tea shoots. For predicting the category and position of the tea shoot, they used an image pyramid structure to obtain the characteristics map of the tea shoot and built the dataset for the plucking points. The algorithm was trained and achieved 90% under the verification set. Endang et al. [8] in their work proposed a tea clone identification system using an unsupervised learning algorithm derived from Deep Convolutional Generative Adversarial Network (DCGAN). Their proposed encoder DCGAN achieved an accuracy of 91.4% with an optimum loss of 0.32. Gibson et al. [9] proposed a tea fermentation detection method using a Convolutional Neural Network termed TeaNet and compared with other standard machine learning techniques. Their classifier recorded a precision of 0.78-1.00 in fermentation dataset and between 0.65-0.96 for “LabelMe” dataset.

III. MATERIALS AND METHOLOGY

3.1 System description

The block diagram of the machine vision setup is shown in Figure 1. The system comprises of a hopper, a vibratory feeder, vibration control unit, a motor, control panel, vision inspection unit, an actuator and a computer. A digital camera with an illumination arrangement is placed inside the vision inspection cabinet.

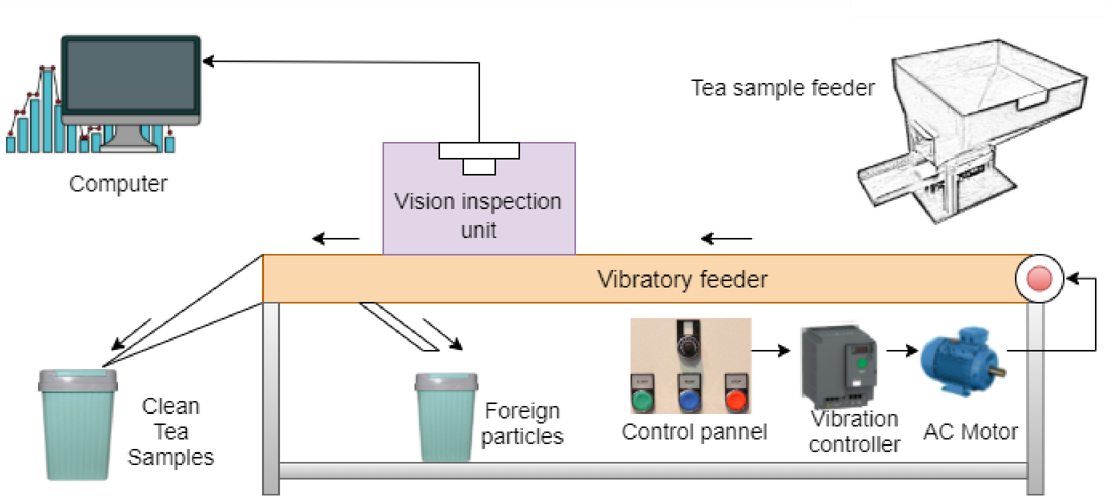

Figure 1. Machine vision setup

The tea sample is allowed to pass on a vibratory feeder. The speed of the movement of the tea particle is controlled by controlling the number of rotations per minute of the motor thorough a speed controller (variable frequency drive). The vibrating frequency of the feeder is set in such a way that the smallest bit of unwanted particles is exposed on the upper layer of the tea sample. A digital camera is mounted on top of the illuminated enclosed chamber on the feeder. The intensity of illumination inside the enclosed chamber is controlled by a light intensity controller and maintained at a desired level between 100 - 120 Lux. The camera is focused on the vibratory feeder through which the tea sample is passed and the video is captured at the frame rate of 30 fps (frames per second) with resolution of 640x480 pixels in RGB mode. The still frames are extracted from the video. Images are taken using the following camera settings: manually adjusted fixed focus, automatic shutter speed adjustment, no zoom, no flash and storage in 24 bit BMP format. The white balance of the camera is set in auto mode. A pneumatic valve is actuated automatically by obtaining a digital signal from a data acquisition card (NI USB 6009) connected to the computer upon detecting foreign matter in the image. The pneumatic valve fitted at the bottom of the system opens a small slit attached to the bottom of the feeder to release a portion of tea sample with foreign matter. Tea sample is fed from a hopper with a controlled flow rate (10 Kg/hour). Two electronic weighing balances are used to measure the foreign matter along with the wastage tea sample and clean tea sample that comes out from the end of the feeding system.

3.2 Data collection

Black tea samples were collected from a local market. The samples were divided into five different batches with 1000 gm of tea in each batch. Different type of foreign matter ranging from big, medium and small sizes were collected from a local tea packaging industry. During experiment, different types of foreign matter were selected randomly and mixed to these batches with a predefined proportion. The foreign matter mixed with tea samples are shown in Figure 2. The tea samples mixed with foreign matter were poured into the hopper. The tea sample was allowed to move on the vibratory feeder. The digital camera captured the video feed with a resolution of 640x480 pixels per frame. Image frames were extracted from the video and each frame was analysed to check the presence of foreign matter. Thus 2000 images were found containing tea sample mixed with foreign matter. Annotations were done for 1500 images for training the algorithms.

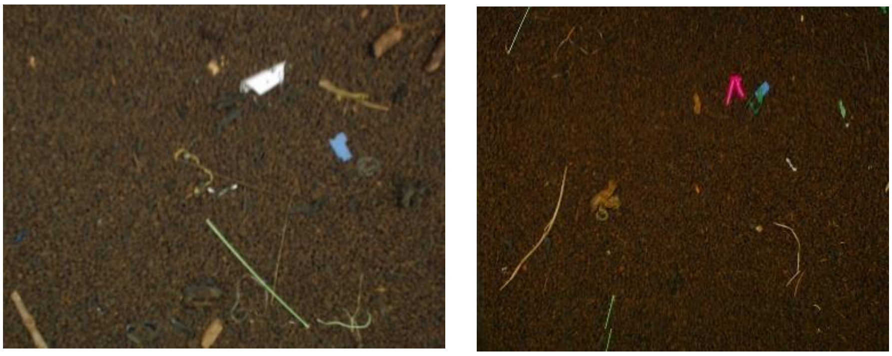

Figure 2. Data captured with different illumination levels

3.3 Object Detection Algorithms

(i) *Faster RCNN*

Faster RCNN [1] algorithm is the improved version of R-CNN and Fast RCNN where Regional Proposal Network (RPM) is used for getting the proposals. The use of RPM not only brings down the region proposal search time from 2 second to 10 milliseconds, but also RPM allows the region proposals to be shared with the later detection stages which causes an overall improvement to the feature representation. First the input image is passed to a base feature extractor, which in this case is pre-trained Inception v2 network. As the input image passes through multiple convolutional and pooling layers of the Inception v2blocks, having 1x1, 3x3, 5x5 convolution filters, we get useful level of information's encoded in the feature map relative to the input image. We then take the feature maps from an intermediated layer of the base network (with the classification layer freeze) and feed it to the Regional Proposal Network. RPM finds whether the object is present in the input feature map and its location.

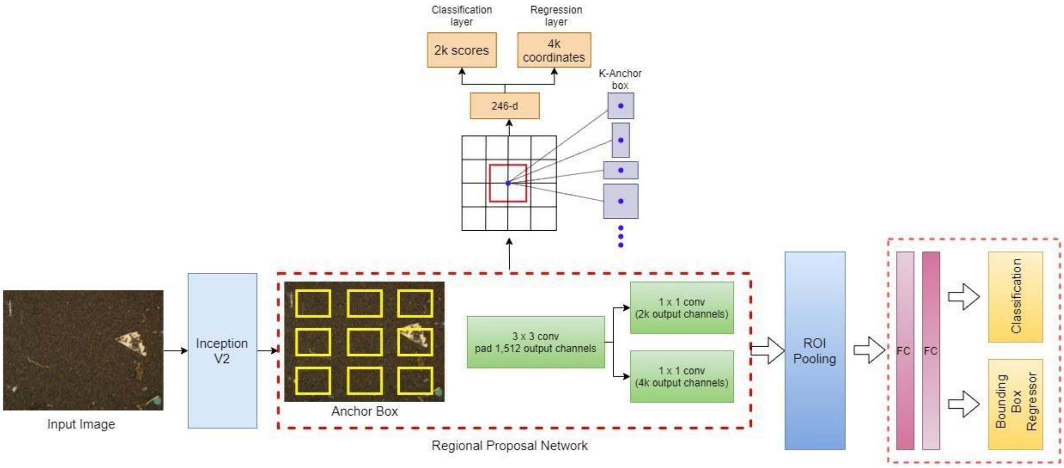

Figure 3. Faster RCNN architecture

This is done using anchor boxes. Anchor boxes are pre-defined bounding boxes, with default boxes of 3 scales 128x128, 256x256, 512x512 and aspect ratios of 1:1, 1:2, 2:1. These boxes are generated throughout the image, and it is the job of RPM to check whether the corresponding anchor contains the object or not, and refine the anchor's coordinates to give bounding box as object proposals or region of interest. To achieve this RPM uses 3x3, 512 convolutional layer on the input feature map and generates 512-d feature map for each location. This is followed by 1x1, 18 for binary classification and 1x1, 36 for bounding box regression. During the training phase, the anchor with IoU (intersection over union) is computed and an overlap of more than 0.7 threshold is considered as foreground class, and those that have an overlap of less than 0.1 IoU is considered as background. Then randomly sampling those anchors to form a mini batch of 256 with maintaining a balance ratio between the foreground and background class. RPM uses these anchor for mini-batch to calculate the binary classification loss and it selects those anchors from the mini-batch which were marked as foreground and uses it to calculate the bounding box regression loss. The output feature map from the RPM is sent to the ROI pooling layer to be resized into 7x7,512 using a 2x2 max-pool kernel. This feature map is sent to the Region-based convolutional neural network (R-CNN) [10] which after flattening is send to two fully connected (FC) layers. The first FC layer is a softmax classification layer for predicting the foreign particle, and the second FC layer is used for predicting the coordinates of the bounding boxes. The architecture of R-CNN is shown in Figure 3.

(ii) YOLO v3

YOLO v3 [11] is a modified variant of YOLO v2 [12] which originally used Darknet-19 [12] as its base network. The backbone of YOLO v3 is darknet-53, which is made up of 53 convolutional layers, followed by batch-normalization and Leaky ReLu activation. For the detection task, the architecture is stacked with an additional 53-layers, giving it a total of 106 layers. YOLO v3 extensively uses 1x1 and 3x3 convolution kernels for feature extraction, followed by residual blocks which helps with the problem of vanishing gradients as the feature maps propagate deeper into the network. No form of pooling is used, instead convolutional layer with a stride of two is used to down sample the feature map; this helps to prevent the lower level features in the image. The architecture of YOLO v3 is explained in Figure 4. YOLO v3 takes the input image and resizes it to 416x416 and feeds it into the Darknet network. After performing multiple convolutions using 1x1 and 3x3, a feature map of 13x13 is obtained which is upscale to 26x26 and 52x52 in layers 82, 94, and 106 respectively. These features maps of different sizes are optimized to detect the smaller particles; the 13x13 layer detects larger particles, the 26x26 layer detects medium size particles, and the 52x52 layer detects smaller particles. Using adaptive feature fusion, the feature maps are rescaled and combined for the final detection. YOLO v3 predicts bounding boxes for each grid cell of the feature map; each of the grid cells, in turn, predicts an object through one of the bounding boxes, if the centre of the object belongs to the receptive field of the cell. For bounding box prediction, pre-defined anchor boxes are used. These anchor boxes have pre-defined aspect ratios, which are determined by running k-means clustering on the entire dataset. In total 9 anchor boxes are used (3 anchor boxes for each of the feature maps). To predict the real height and width of the bounding box, YOLO v3 calculates offsets to the pre-defined anchors. These offsets are called Log-Space Transform. For each anchor boxes, we determine how much the ground truth box overlaps with the anchor box, and pick the one with the highest IoU if the IoU is greater than 50%, then the anchor box detects the object. The network is predicting four bounding box coordinates t_x t_y, t_w, t_z. If the cell is offset from the top left corner of the image by c_x and c_y, then the prediction of the bounding box can be calculated using the equations (i) – (iv).

$$b_x = \sigma(t_x) + c_x \quad \text{(i)}$$
$$b_y = \sigma(t_y) + c_y \quad \text{(ii)}$$
$$b_w = p_w e^{tw} \quad \text{(iii)}$$
$$b_h = p_h e^{th} \quad \text{(iv)}$$

Here, σ denotes sigmoid function, p_w and p_h is the width and height of the bounding box. The error of class probability and position of the object is minimised and the network parameters is updated using back propagation.

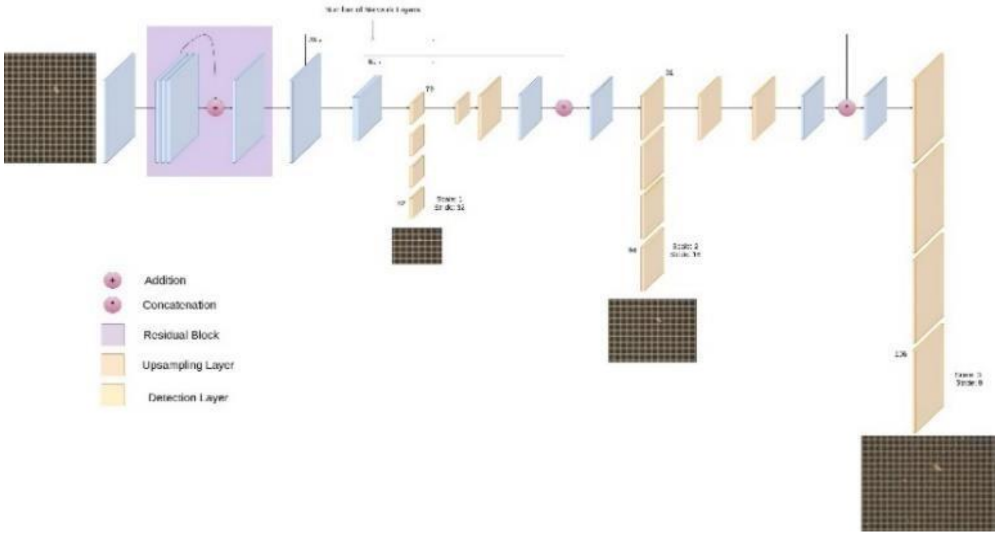

Figure 4. YOLO v3 Network

IV. EXPERIMENTATION AND RESULTS

Both models were trained with 1500 images, and multiple testing took place to ensure the system can detect any type of foreign matter present in the tea. The CNN model requires large amounts of data for training. Because of limited availability of images with foreign matter in tea, we employ transfer learning technique. This technique uses a pre-trained model which was trained previously on large dataset. Faster RCNN algorithm for this work was pre-trained on COCO dataset and we used those pre-trained weights for our custom model detection. Input image of 640*480 size was resized to a min-max dimension of 600,1024 and was fed to the network. During the training phase, anchor boxes of scales: [0.25, 0.5, 1.0, 2.0] and aspect ratios [0.5, 1.0, 2.0] were generated. The non-max suppression-IoU threshold was set to 0.07. We used Stochastic Gradient Decent (SDG) with momentum optimizer, with momentum optimizer value set to 0.9. The initial learning rate was 0.0002 and the model ran for 58,000 steps, for 300 epochs. We considered average precision (AP), recall value and the speed of analysis as our performance criteria. AP represents the mean of positive predictive value which is calculated as the ratio of correct positive predictions to the total predicted positives considering both true positive and false positive results. Recall value represents the proportion of actual positives which was identified correctly by considering both true positive and the false negative results. We got an average precision of 0.42, an average recall value of 0.47, and found the time of analysis per frame as 0.25 sec. The model gave us a classification loss and localization loss of 0.54 and 0.49 respectively. The classification loss and localization loss are shown in Figure 5, where x-axis represents the number of steps and y-axis indicates loss.

Figure 5. Faster RCNN classification and localization loss curve

For YOLO v3, the anchor scale was 1.05, 1.1, 1.2, the optimizer used was Adam, learning rate of 0.001 with 0.9 momentum value. The network was trained for 60,000 steps and converged at a loss value of 0.60 as shown in Figure 6 curve, where x-axis represents iteration number, y-axis represents the loss value and the red line indicates mAP chart. After evaluating, we got AP value of 0.91 and recall of 0.67. YOLO v3 gave us a stable 15 fps and is currently implemented in our system.

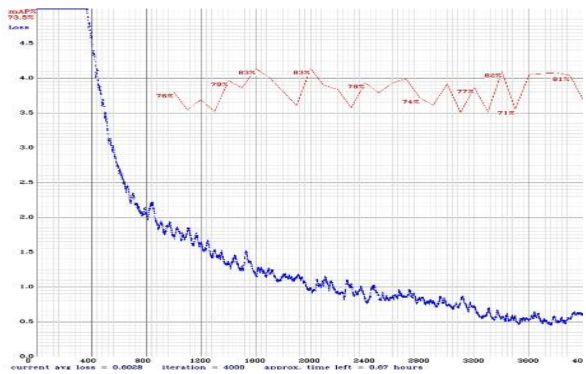

Figure 6. YOLO v3 loss curve

TABLE I. PERFORMANCE COMPARISION

Model	Average Precision	Average Recall	fps
Faster RCNN	0.42	0.47	4
YOLO v3	0.91	0.67	15

Table I. shows the comparison of performance using Faster RCNN and YOLO v3. Faster RCNN model shows recall value as 0.47 and 0.42 as precision value. This implies that it classifies the positive samples correctly, but it has many false positive (i.e. it classifies many negative samples as positive). The higher precision and recall values of YOLO v3 model gives a high confidence score in detecting and classifying the samples. Also, speed is a very important parameter for our system and since we are getting an average speed of over 15 fps with YOLO v3 model, we have implemented this algorithm for our system.

V. CONCLUSION

In this work, we explored the use of two deep CNN object detection models i.e. Faster R-CNN and YOLO v3 in the tasks of detection and classification of foreign matter in tea sample during tea packaging. As our proposed model will be used in detection and classification of foreign matter in real time, we also explored the above models to see the impact of network complexity on overall performance and speed of operation. We make a comparison between R-CNN and YOLO v3 approaches based on predefined performance parameters. Experimental result demonstrates that YOLO v3 model achieves superior performance over Faster R-CNN. Fine tuning the models for our problem yields 0.91 precision, 0.67 recall and analysis speed of 15 fps. From the experimental results, we may conclude that YOLO v3 model shows a better performance on detecting large foreign matters, but it is not robust to detect smaller objects and fails to classify overlapping objects. Further, we propose to improve the performance of our system by experimenting with larger dataset and tuning the parameters of the YOLO v3 model. Our experimental result illustrates the real-time applicability and superiority of YOLO v3 object detection model within this foreign matter detection problem in tea. The algorithm captures those particles and track them until it reaches the opening valve for separation. The outputs of the analysis using Faster RCNN and YOLO v3 are shown in Figure 7. (A) and 7. (B). Future work will consider exploiting implementation of the model in a single board hardware platform integrating with sorting solution.

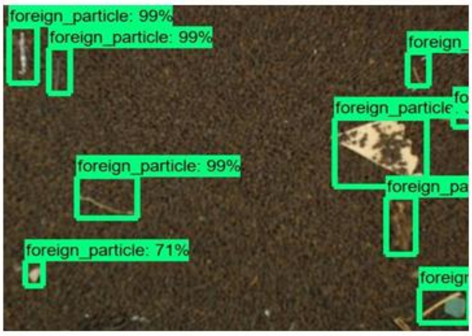

Figure 7. (A) Software output showing the detected foreign matter using Faster RCNN

Figure 7. (B) Software output showing the detected foreign matter using YOLO v3

REFERENCES

[1]. S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[2]. J. Redmon, S. Divvala, R. Girshick and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779-788, doi: 10.1109/CVPR.2016.91.

[3]. Liu W. et al. SSD: Single Shot MultiBox Detector. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9905. Springer, Cham. https://doi.org/10.1007/978-3-319-46448-0_2

[4]. Xiao, Yi and Wang, Xinqing and Zhang, Peng and Meng, Fanjie and Shao, Faming, Object Detection Based on Faster R-CNN Algorithm with Skip Pooling and Fusion of Contextual Information, Sensors 2020, 20, 5490; doi:10.3390/s20195490

[5]. N. Das, K. Kalita, P. K. Boruah and U. Sarma, "Prediction of Moisture Loss in Withering Process of Tea Manufacturing Using Artificial Neural Network," in *IEEE Transactions on Instrumentation and Measurement*, vol. 67, no. 1, pp. 175-184, Jan. 2018, doi: 10.1109/TIM.2017.2754818.

[6]. Yu-Ting Chen, Shih-Fang Chen, Localizing plucking points of tea leaves using deep convolutional neural networks, Computers and Electronics in Agriculture, Vol 171, 2020, 105298, ISSN 0168-1699, https://doi.org/10.1016/j.compag.2020.105298.

[7]. Yang, Hualin & Chen, Long & Chen, Miaoting & Ma, Zhibin & Deng, Fang & Li, Maozhen & Li, Xiangrong, Tender Tea Shoots Recognition and Positioning for Picking Robot Using Improved YOLO-V3 Model. IEEE Access. 7, 2019, 180998-181011, doi. 10.1109/ACCESS.2019.2958614.

[8]. Suryawati, Endang & Zilvan, Vicky & Yuwana, R. & Heryana, Ana & Rohdiana, Dadan & Pardede, Hilman. (2019). Deep Convolutional Adversarial Network-Based Feature Learning for Tea Clones Identifications. 1-5. 10.1109/ICICoS48119.2019.8982416.

[9]. Kimutai, G.; Ngenzi, A.; Said, R.N.; Kiprop, A.; Förster, A. An Optimum Tea Fermentation Detection Model Based on Deep Convolutional Neural Networks. *Data* 2020, *5*, 44. https://doi.org/10.3390/data5020044

[10]. R. Girshick, J. Donahue, T. Darrell and J. Malik, "Region-Based Convolutional Networks for Accurate Object Detection and Segmentation," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 1, pp. 142-158, 1 Jan. 2016, doi: 10.1109/TPAMI.2015.2437384.

[11]. Joseph Redmon, Ali Farhadi, "YOLOv3: An Incremental Improvement", 2018, arXiv:1804.02767

[12]. Joseph Redmon, Ali Farhadi, "YOLO9000: Better, Faster, Stronger", 2016, arXiv:1612.08242