Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16439 - 16450
Received 05 March 2021; Accepted 01 April 2021.

Deep Learning Based Multi Class Wild Pest Identification and Solving Approach Using Cnn

Dr. D. Rajalakshmi[1] V. Monishkumar[2] Segu Balasainarayana[3] Meka Siva Rama Prasad[4]

[1]Assistant Professor, Department of Computer Science and Engineering, R.M.D. Engineering College, Kavaraipettai – 601 206
[2] AWS Cloud Architect, Virtusa Pvt Ltd
[3]B.E. Final Year Student, Department of Computer Science and Engineering, R.M.D. Engineering College, Kavaraipettai – 601 206
[4]B.E. Final Year Student, Department of Computer Science and Engineering, R.M.D. Engineering College, Kavaraipettai – 601 206

ABSTRACT

In several nations, specialised pest and disease control has become a high-priority challenge for the agricultural sector. Image processing has become more automated and cost-effective as a result of its cost-effectiveness. In practical crop protection applications, analytic pest recognition systems are commonly used. in the identification and recognition of multi-class pests on a broad scale.This paper proposes a region-wide end-to-end solution called Deeplearning based multi class wild pest monitoring approach using CNN to address this issue. For feature extraction and enhancement, a novel module channel spatial focus (CSA) is proposed to be fused into the convolutional neural network (CNN) backbone. The second is the region proposal network (RPN), which uses derived feature maps from images to include region proposals as possible pest positions. Our 7-year large-scale pest dataset of 88.6K photographs (16 categories of pests) and 582.1K manually labelled pest items is used to test the method. The experimental findings indicate that the proposed System outperforms state-of-the-art approaches in multi-class pest identification, with a mean average precision (mAP) of 87.43 percent.

KEY-WORDS: Channel-spatial attention(CSA), Convolutional neural network(CNN), multi-class pest detection, position-sensitive score map, region proposal network(RPN).

I. INTRODUCTION

Regarding the growth of crops, one of the important factors affecting crop yield is insect disasters. Since most insect species are extremely similar, insect detection on field crops, such as rice, soybean and other crops, is more challenging than generic object detection. Presently, distinguishing insects in crop fields mainly relies on manual classification, but this is an extremely time-consuming and expensive process. This work proposes a convolution neural network model to solve the problem of multi-classification of crop insects. The model can make full use of the advantages of the neural network to comprehensively extract multifaceted insect features.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16439 - 16450
Received 05 March 2021; Accepted 01 April 2021.

Objective

The main objective of this model is to identify the classification of pest with high accuracy in a cost effective way. Every farmer can use this approach of pest identification to utilize the pesticides correctly. Around 80k image dataset has been fed to the system to get the appropriate results. The System uses the advanced techniques to classify the pests based on the CNN and CSA which produces more accurate results.

Existing System

Manual classification is currently used to identify insects in crop fields, but this is a time-consuming and costly method. Compared with previous classifiers such as k-nearest neighbors and linear discriminate analysis (LDA), support vector machine (SVM) was proposed with Haar-like features to classify insects and obtained a poor performance than the Convolutional Neural Network.

Proposed System

Our System consists of three stages: pest feature extraction, pest regions search and pest prediction. In System, the input image is firstly fed into a CNN backbone to extract feature maps, where CSA module is proposed for feature enhancement. Then we fuse RPN and PSSM for providing pest regions and pest prediction respectively. During the prediction phase, Contextual RoIs are presented as contextual information to improve detection accuracy.

Expected Result

These results area unit outputted by our system supported CNN backbone. The environments of input pictures from top to bottom area unit additional and difficult because it are often seen, our methodology may come through multi-class blighter localization and recognition below each easy and sophisticated environments and supply the anticipated severity estimation. It may be found that, the feature maps in our system diminish the highlights of non-objects and focus additional attention on blighter regions with lighter activation points with our designed design. Therefore, our methodology may perform higher on blighter detection and more and more learn the pests options well.

II. RELATED WORK

There is lots of advanced techniques developed and applied in fashionable agricultural field like leaf diseases identification and bug recognition . Among these works, 2 key steps of ancient pc vision strategies might be summarized: (1) feature extraction that extracts data as feature vectors from pictures. (2) pattern recognition that trains a model to classify classes of input pictures. The relatively early works for gadfly identification was done by World Health Organization achieved classification through RGB multispectral analysis also because the technique projected by that recognized insects through eigen-images extracted by Principle part Analysis (PCA) algorithmic rule. Since then, an excellent deal of achievements emerged within the past few years. Size and color options were conjointly extracted to categorise whiteflies, aphids and thrips . with the exception of size, form and texture options were conjointly chosen for characteristic flower . However, these options were too weak to be insensitive to rotation, scale and translation. To upset it, utilized Scaleinvariant feature rework (SIFT) was utilized with LOSS algorithmic rule to classify insects . Meanwhile,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16439 - 16450
Received 05 March 2021; Accepted 01 April 2021.

compared with previous classifiers like k-nearest neighbours and linear discriminate analysis (LDA) , support vector machine (SVM) was projected with Haar-like options to classify insects and obtained a higher performance than the progressive strategies. In terms of neural network approaches, Artificial Neural Network (ANN) was adopted also as SVM supported their own designed options.

Firstly, a completely unique module Channel-Spatial Attention (CSA) is proposed into CNN primarily based network for feature extraction and enhancement. Secondly, for tormenter localization, we tend to use a region proposal technique combined with PSSM module generating candidate boxes mechanically, within which discourse RoIs (Region of Interest) as discourse data area unit increased. Such this end-to-end technique doesn't need any preprocessing and human intervention and will yield state of- the-art performance.

III. METHODOLOGY

For agriculture pest identification, there exist many open datasets discharged like Butterfly Dataset. However, to our greatest information, few open datasets appropriate for multi-class pest detection task area unit discharged whereas our purpose is to notice totally different sorts of pests at the same time in one image. As a result, we tend to build a dataset for our large-scale multi-class pest detection task. Specifically, the pest image acquisition instrumentality designed for capturing pictures of multi-class pests in our dataset is shown. In this device, the multispectral lightweight lure may emit lightweight for attracting multi-class pests, during which the wavelengths may vary with time in step with habits of pests within the day to make sure differing types of pests can be captured. Then these attracted pests would be surprised by the screen and comprise the pest assortment receptacle on the lowest. At identical time, the HD camera in the receptacle higher than is about to require photos sporadically at fifteen second intervals. when being photographed, the pests would be swept away from the pest assortment receptacle forthwith to avoid accumulation and overlapping. The captured pictures are keep in JPG format at 2592×1944 resolution. Hereafter, each pest in pictures area unit annotated by agricultural specialists with labels and bounding boxes. Finally, 88,670 pictures containing 582,170 pest objects classified in sixteen categories area unit captured.

Architecture of the System

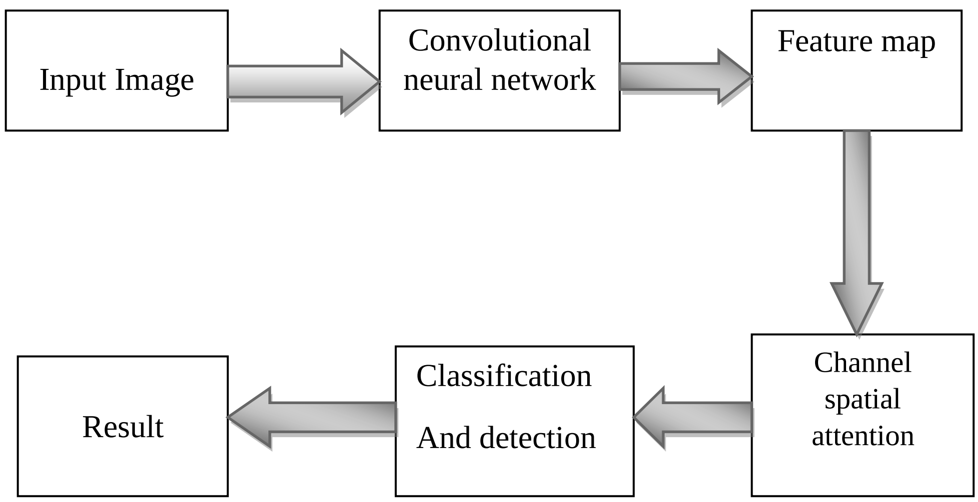


Figure 1. It shows the architecture of the system.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16439 - 16450
Received 05 March 2021; Accepted 01 April 2021.

Convolutional Neural Network

Conventional computer vision employed hand-crafted features to describe the images. Instead, we adopt CNN for automatic feature extraction which is basically composed of 3 parts: convolutional layer, activation function, and pooling layer.

Convolutional Layer

Standard convolutional layer takes a set of filters (also called kernel) as a filterbank to the input and the output feature map in each subsequent layer could be regarded as abstract transformations of image. Take a size of Wl−1×H l−1×C l−1 feature map and a filterbank within C l filters at size of f l × f l×C l−1 in layer l−1 for example, augmenting the other two hyper-parameters padding p l and stride s l , the output feature map in layer l is at size of Wl × H l × C l :

(Wl , H l) = (Wl−1 , H l−1) + 2p l − f l s l + 1 (1)

where b·c denotes floor operation. Note that the number of filters must be equal to that of input feature map. More specifically:

x l j = ϕ(X i∈Mj x l−1 i × f l ij + b l j) (2)

where i and j are indexes of input and output feature maps at range of Wl × H l and Wl−1 × H l−1 respectively. Mj here indicates the receptive field of filter and b l j is bias term.

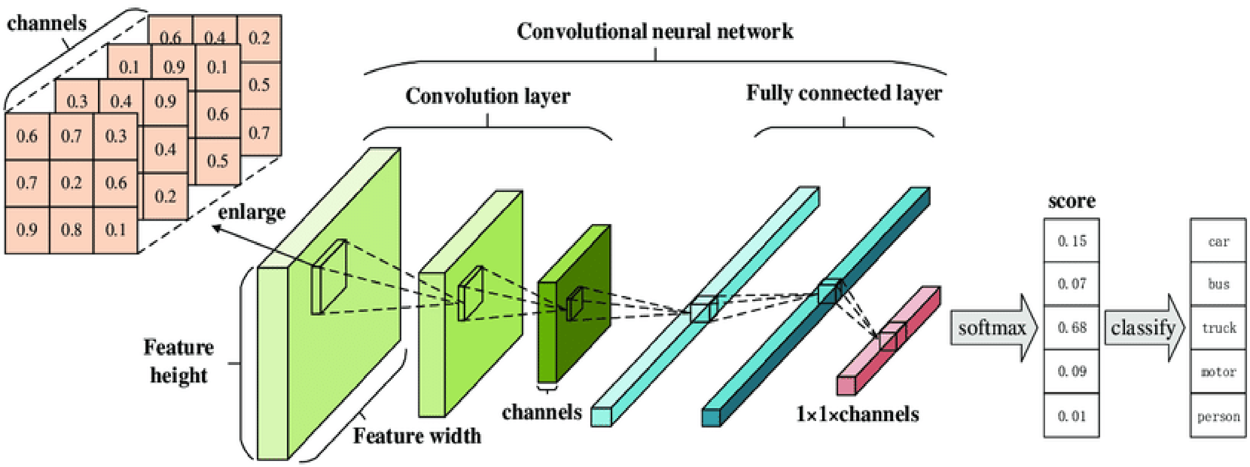
Activation Layer

In the equation (2), ϕ(·) is called activation function applied to achieve element-wise non-linearity in deep learning. Channel-spatial attention architecture. contains many types such as sigmoid, Rectified Linear Units (ReLU). In our method, we utilize ReLU as activation function for faster training because of larger gradient in

(0,∞): ϕ(x) = max(0, x) (3)

POOLING LAYER

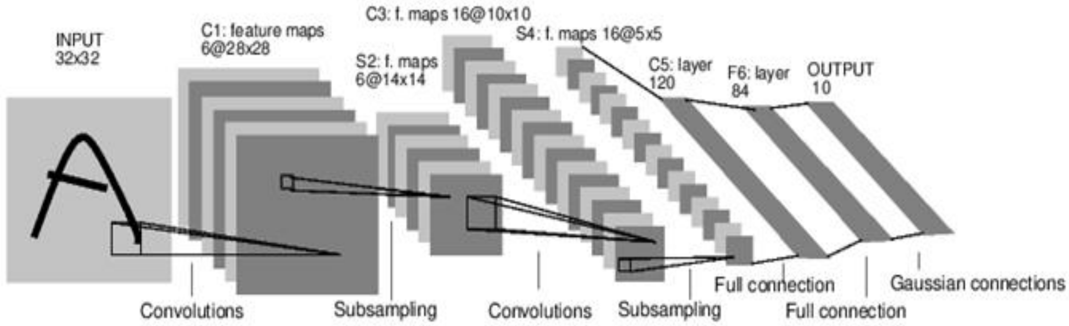
In CNN, pooling layer is usually applied for feature dimension reduction. Besides, spatial translational invariance is another benefit of pooling layer. Among different pooling layer methods, max-pooling layer is selected in PestNet which applies local pooling by preserving maximum of receptive field and discarding other values.



Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16439 - 16450
Received 05 March 2021; Accepted 01 April 2021.

Fully Connected Layer

Now that we are able to notice these high level options, the icing on the cake is attaching a completely connected layer to the top of the network. This layer essentially takes Associate in Nursing input volume (whatever the output is of the conv or ReLU or pool layer preceding it) Associate in Nursingd outputs an N dimensional vector wherever N is that the range of categories that the program needs to choose between. as an example, if you wished a digit classification program, N would be ten since there area unit ten digits. every range during this N dimensional vector represents the chance of an explicit category. as an example, if the ensuing vector for a digit classification program is [0 .1 .1 .75 0 0 0 0 0 .05], then this represents one0th|a tenth} chance that the image may be a 1, a tenth chance that the image may be a a pair of, a seventy fifth chance that the image may be a three, and a five-hitter chance that the image may be a nine (Side note: There area unit different ways in which you'll represent the output, however i'm simply showing the softmax approach). The manner this totally connected layer works is that it's at the output of the previous layer (which as we tend to keep in mind ought to represent the activation maps of high level options) and determines that features most correlate to a selected category. as an example, if the program is predicting that some image may be a dog, it'll have high values within the activation maps that represent high level options sort of a paw or four legs, etc. Similarly, if the program is predicting that some image may be a bird, it'll have high values within the activation maps that represent high level options like wings or a beak, etc. Basically, a FC layer appearance at what high level options most powerfully correlate to explicit|a specific|a selected} category and has particular weights so once you reckon the product between the weights and therefore the previous layer, you get the right chances for the various categories.



A Full Convolutional Neural Network (LeNet)

Training the System

Now, this can be the one facet of neural networks that I advisedly haven't mentioned nevertheless and it's in all probability the foremost necessary half. There is also heaps of queries you had whereas reading. however do the filters within the 1st conv layer grasp to appear for edges and curves? however will the absolutely connected layer grasp what activation maps to appear at? however do the filters in every layer grasp what values to have? The manner the pc is ready to regulate its filter values (or weights) is thru a coaching method known as backpropagation.

Before we tend to get into backpropagation, we must first take a step back and speak about what a neural network needs so as to figure. At the instant we all were born, our minds were fresh. We didn't know what a cat or dog or bird was. during a similar type of way,

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16439 - 16450
Received 05 March 2021; Accepted 01 April 2021.

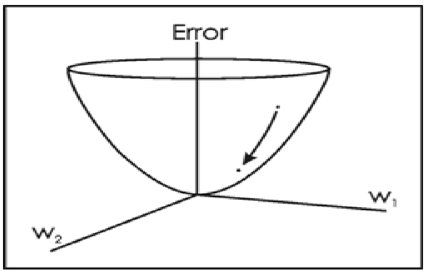
before the CNN starts, the weights or filter values are randomized. The filters don't know to appear for edges and curves. The filters within the higher layers don't know to appear for paws and beaks. As we grew older however, our parents and teachers showed us different pictures and pictures and gave us a corresponding label. this idea of being given a picture and a label is that the training process that CNNs bear. Before getting too into it, let's just say that we've got a training set that has thousands of images of dogs, cats, and birds and every of the pictures incorporates a label of what animal that picture is. Back to backprop.

So backpropagation may be separated into four distinct sections, the passing game, the loss perform, the backward pass, and therefore the weight update. throughout the passing game, you are taking a coaching image that as we tend to keep in mind may be a thirty two x thirty two x three array of numbers and pass it through the total network. On our initial coaching example, since all of the weights or filter values were willy-nilly initialized, the output can in all probability be one thing like [.1 .1 .1 .1 .1 .1 .1 .1 .1 .1], primarily associate degree output that doesn't provide preference to any range specially. The network, with its current weights, isn't able to search for those low level options or therefore isn't able to create any affordable conclusion regarding what the classification may be. This goes to the loss perform a part of backpropagation. keep in mind that what we tend to area unit mistreatment without delay is coaching knowledge. This knowledge has each a picture and a label. Let's say for instance that the primary coaching image inputted was a three. The label for the image would be [0 zero zero one zero zero zero zero zero 0]. A loss perform may be outlined in many alternative ways that however a standard one is MSE (mean square error), that is ½ times (actual -

$$E_{total} = \sum \frac{1}{2}(target - output)^2$$

predicted) square.

Let's say the variable L is same to that value. As you may imagine, the loss will be extraordinarily excessive for the primary couple of schooling images. Now, let's simply think approximately this intuitively. We need to get to some extent in which the expected label (output of the ConvNet) is similar to the schooling label (This approach that our community were given its prediction right).In order to get there, we need to minimize the quantity of loss we have. Visualizing this as simply an optimization trouble in calculus, we need to discover which inputs (weights in our case) maximum directly contributed to the loss(or error) of the community.



One way of visualizing this idea of minimizing the loss is to consider a 3-D graph where the weights of the neural net (there are obviously more than 2 weights, but let's go for simplicity) are the independent variables and the dependent variable is the loss. The task of minimizing the loss involves trying to adjust the weights so that the loss decreases. In visual terms, we want to get to the lowest point in our bowl shaped object. To do this, we have to take a derivative of the loss (visual terms: calculate the slope in every direction) with respect to the weights.

This is the mathematical equal of a dL/dW wherein W are the weights at a specific layer. Now, what we need to do is carry out a backward pass thru the network, that's figuring out which weights contributed maximum to the loss and locating approaches to alter them in order that the loss decreases. Once we compute this derivative, we then visit the closing step

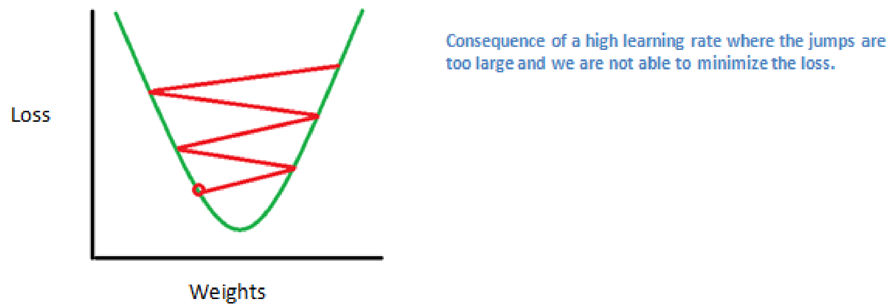
Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16439 - 16450
Received 05 March 2021; Accepted 01 April 2021.

which is the weight replace. This is wherein we take all of the weights of the filters and replace them in order that they extrade withinside the contrary course of the gradient.

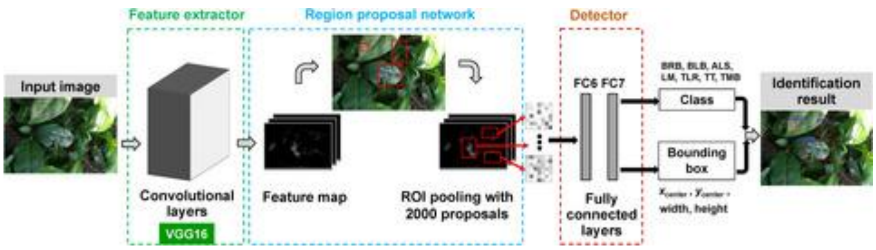
$$w = w_i - \eta \frac{dL}{dW}$$

w = Weight
w_i = Initial Weight
η = Learning Rate

The learning rate may be a parameter that's chosen by the computer programmer. A high learning rate implies that larger steps square measure taken within the weight updates and therefore, it should take less time for the model to converge on AN optimum set of weights.However, a learning rate that's too high might end in jumps that square measure overlarge and not precise enough to succeed in the optimum purpose



The process of pass, loss operate, backward pass, and parameter update is one coaching iteration. The program can repeat this method for a set variety of iterations for every set of coaching pictures (commonly referred to as a batch). Once you end the parameter update on the last coaching, hopefully the network ought to be trained tolerably in order that the weights of the layers area unit tuned properly.



Channel Spatial Attention(CSA)

Inspired by these 2 observations, we tend to propose a unique module Channel-Spatial Attention (CSA) for coefficient channel and spacial data on output from every CNN block to boost the figurative power of feature maps.An intuitive framework of our CSA module within the backbone that consists of 2 components. within the 1st a part of Channel Attention module (the higher part), the 3D feature map with form of W ×H×C extracted by CNN block is input into an additional world pooling layer that takes average pooling from the entire feature maps in every channel to get a lower dimensional (1D) feature vector, within which the averaged worth represents the worldwide feature for every channel. Then, we tend to apply a bunch of convolutional layers with non-linear activation ReLU following. This 1D feature vector is mapped into (0,1) space by adopting Sigmoid perform and therefore the

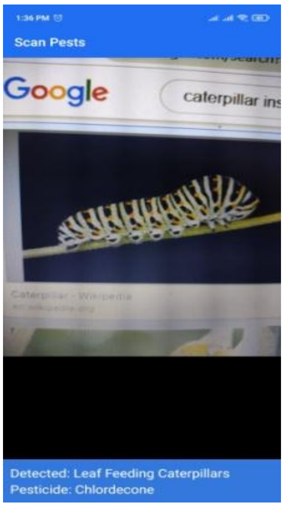
Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16439 - 16450
Received 05 March 2021; Accepted 01 April 2021.

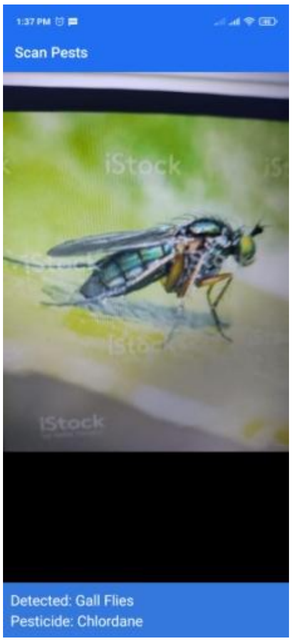
output with form of one × one × C is alleged channel attention issue. Thus, the output of Channel Attention module is that the broadcast element-wise product of the first input 3D feature map and therefore the 1D channel attention issue. during this approach, the input 3D feature map is activated in channel level.

Region Proposal Network (RPN)

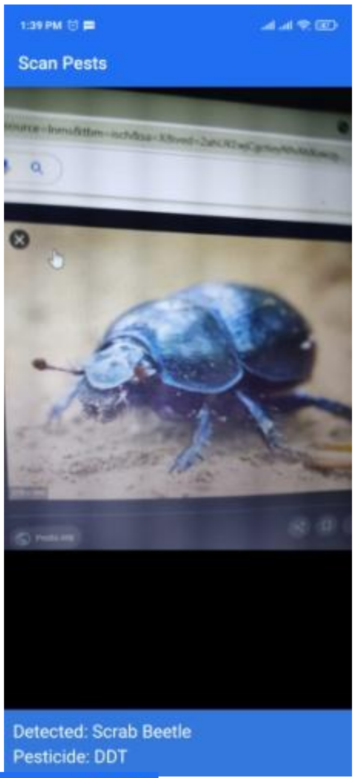
Our System may be a region-based CNN methodology therefore we tend to adopt RPN module to look potential regions of object followed by CNN backbone. As RPN holds its own objectness scores and bounding box regression layer, it may effectively give regions mechanically. In distinction to different relevant strategies e.g. selective search and edge boxes. Region proposal spec. abundant time on thousands of regions, RPN may scale back an oversized range of proposal regions once making certain the standard by introducing numerous anchor boxes for box regression reference.

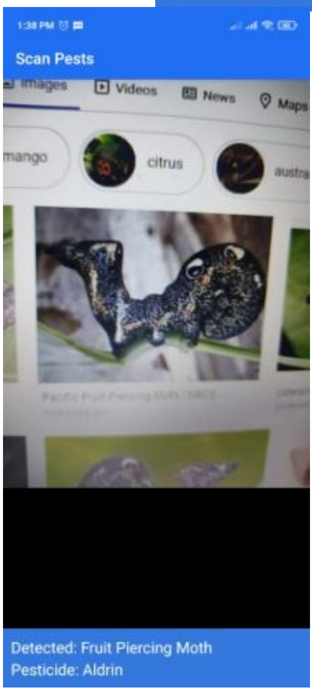
IV. RESULTS





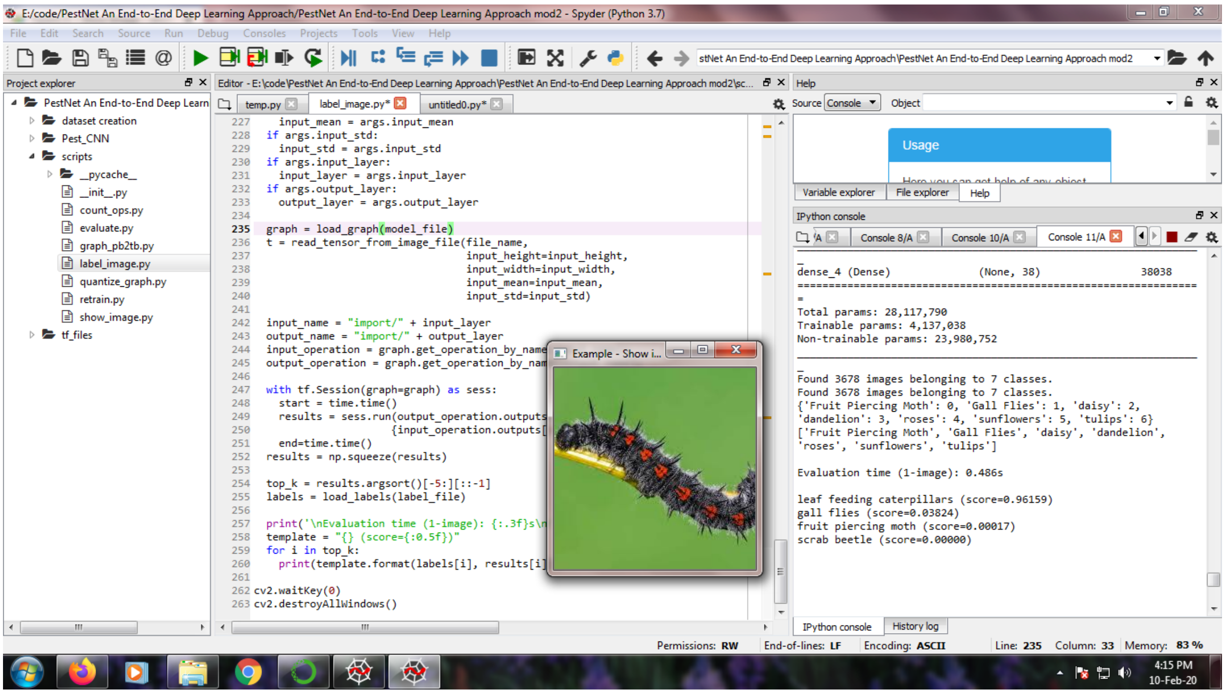
The above images are the results that we obtained from our system. The first image shows the insect identified as leaf feeding caterpillar and the pesticide it suggests as Chlordecone .Likewise the second image shows the insect as Gall flies and the pesticide suggested as Chlordane.





The above images are the results that we obtained from our system. The third image shows the insect identified as Scrab Beetle and the pesticide it suggests as DDT .Likewise the fourth image shows the insect as Fruit Piercing Moth and the pesticide suggested as Aldrin.

Annals of R.S.C.B., ISSN:1583-6258, Vol. 25, Issue 4, 2021, Pages. 16439 - 16450
Received 05 March 2021; Accepted 01 April 2021.



The above image shows the code running in the spyder ide platform which takes the input image and shows result, it is the correct species identification of the insect or pest. It takes the input image and matches with the closely related species and it searches deeper using the CNN methodologies to detect the appropriate pest species identification.

V. CONCLUSION AND FUTURE WORK

This paper projected a unique end-to-end automatic pest detection network , that with success achieves large-scale multi-class tormentor detection. Our system might notice the automated extraction of upper quality options by our projected CSA module that's a unique feature improvement module. moreover, compared to several common object detection strategies, we tend to adopted PSSM rather than FCfor procedure value reduction in pest classification and box regression method. Besides, discourse RoIs was conjointly considered as discourse info in system to any improve detection performance. underneath our enriched dataset MPD2019, The system have achieved the next mAP (87.43%)

among sixteen categories of pests than the progressive strategies. Despite that we develop a novel deep learning based system for pest monitoring task in the field and achieve a successful performance in our dataset, there are several limitations of our method that could be improved in future smart agriculture innovation. Firstly, the unbalanced data structure could be alleviated in the next work. Specifically, due to the difficulty in capturing pests of some rare categories in our pest monitoring equipment, our system tends to identity an unknown pest into the common species, which might improve the risk of inaccurate pest severity warning. Besides, it is necessary to achieve the real-time pest image recognition and detection performance in our system, in which current inference time might be an important factor that limits the advances in agricultural applications. Therefore, future work would target at solving the problem of unbalanced dataset and focus on developing real-time automatic pest monitoring system.

VII. REFERENCES

[1] C. Wu, C. Luo, N. Xiong, W. Zhang, and T. Kim, ``A greedy deep learning method for medical disease analysis," *IEEE Access*, vol. 6, pp. 20021-20030, 2018.

[2] G. Liang, H. Hong, W. Xie, and L. Zheng, ``Combining convolutionalneural network with recursive neural network for blood cell image classification," *IEEE Access*, vol. 6, pp. 36188-36197, 2018.

[3] D. Hu, L.Wang,W. Jiang, S. Zheng, and B. Li, ``A novel image steganography method via deep convolutional generative adversarial networks," *IEEE Access*, vol. 6, pp. 38303_38314, 2018.

[4] R. Girshick. (Apr. 2015). ``Fast R-CNN." [Online]. Available: https:// arxiv.org/abs/1504.08083

[5] J. Dai, Y. Li, K. He, and J. Sun, ``R-FCN: Object detection via region-based fully convolutional networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 379-387.

[6] S. Ren, K. He, R. Girshick, and J. Sun, ``Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137-1149, Jun. 2017.

[7] A. Shrivastava, A. Gupta, and R. Girshick, ``Training region-based object detectors with online hard example mining," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 761-769.

[8] X. Zhang, Y. Qiao, F. Meng, C. Fan, and M. Zhang, ``Identification of maize leaf diseases using improved deep convolutional neural networks," *IEEE Access*, vol. 6, pp. 30370_30377, 2018.

[9] C. Xie *et al.*, ``Multi-level learning features for automatic classification of field crop pests," *Comput. Electron. Agricult.*, vol. 152, pp. 233-241, Sep. 2018.

[10] I. Y. Zayas and P. W. Flinn, ``Detection of insects in bulkwheat samples with machine vision," *Trans. ASAE*, vol. 41, no. 3, p. 883, 1998.

[11] P. J. D. Weeks, M. A. O'Neill, K. J. Gaston, and I. D. Gauld, ``Species identification of wasps using principal component associative memories," *Image Vis. Comput.*, vol. 17, no. 12, pp. 861-866, 1999.

[12] J. Cho *et al.*, ``Automatic identification of whiteflies, aphids and thrips in greenhouse based on image analysis," *Red*, vol. 346, no. 246, p. 244, 2007.

[13] C. Wen, D. E. Guyer, and W. Li, ``Local feature-based identification and classification for orchard insects," *Biosyst. Eng.*, vol. 104, no. 3, pp. 299-307, 2009.

[14] X.-L. Li, S.-G. Huang, M.-Q. Zhou, and G.-H. Geng, ``KNN-spectral regression LDA for insect recognition," in *Proc. 1st Int. Conf. Inf. Sci. Eng. (ICISE)*, Dec. 2009, pp. 1315-1318.

[15] N. Larios, B. Soran, L. G. Shapiro, G. Martinez-Munoz, J. Lin, and T. G. Dietterich, ``Haar random forest features and SVM spatial matching kernel for stonefly species identification," in *Proc. 20th Int. Conf. Pattern Recognit. (ICPR)*, Aug. 2010, pp. 2624-2627.

[16] W. Ding and G. Taylor, ``Automatic moth detection from trap images for pest management," *Comput. Electron. Agricult.*, vol. 123, pp. 17-28, Apr. 2016.

[17] Y. Kaya and L. Kayci, ``Application of artificial neural network for automatic detection of butterfly species using color and texture features," *Vis. Comput.*, vol. 30, no. 1, pp. 71-79, 2014.

[18] K. He, X. Zhang, S. Ren, and J. Sun, ``Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770-778.

[19] R. Girshick, J. Donahue, T. Darrell, and J. Malik, ``Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 580-587.

[20] Y. Shen, H. Zhou, J. Li, F. Jian, and D. S. Jayas, ``Detection of stored-graininsects using deep learning," *Comput. Electron. Agricult.*, vol. 145,pp. 319-325, Feb. 2018.
[21] D. Xia, P. Chen, B.Wang, J. Zhang, and C. Xie, ``Insect detection and classification based on an improved convolutional neural network," *Sensors*,vol. 18, no. 12, p. 4169, 2018.
[22] F. Alvaro, Y. Sook, K. Sang, and P. Dong, ``A robust deep-learningbaseddetector for real-time tomato plant diseases and pests recognition,"*Sensors*, vol. 17, no. 9, p. 2022, 2017.
[23] X. Cheng, Y. Zhang, Y. Chen, Y. Wu, and Y. Yue, ``Pest identificationvia deep residual learning in complex background," *Comput. Electron.Agricult.*, vol. 141, pp. 351-356, Sep. 2017.