Deep Learning based Automatic Approach using Hybrid Global and Local Activated Features towards Large-scale Multi-class Pest Monitoring

Liu Liu
Institute of Intelligent Machines, and Hefei Institute of Physical Science
Chinese Academy of Sciences
Hefei, China
liuliu66@mail.ustc.edu.cn

Rujing Wang
Institute of Intelligent Machines, and Hefei Institute of Physical Science
Chinese Academy of Sciences
Hefei, China
rjwang@iim.ac.cn

Chengjun Xie
Institute of Intelligent Machines, and Hefei Institute of Physical Science
Chinese Academy of Sciences
Hefei, China
cjxie@iim.ac.cn

Po Yang
Department of Computer Science
Liverpool John Moores University
Liverpool, UK
poyangcn@gmail.com

Sud Sudirman
Department of Computer Science
Liverpool John Moores University
Liverpool, UK
S.Sudirman@ljmu.ac.uk

Fangyuan Wang
Institute of Intelligent Machines, and Hefei Institute of Physical Science
Chinese Academy of Sciences
Hefei, China
wfy710@mail.ustc.edu.cn

Rui Li
Institute of Intelligent Machines, and Hefei Institute of Physical Science
Chinese Academy of Sciences
Hefei, China
lirui@iim.ac.cn

***Abstract*—Monitoring pest in agriculture has been a high-priority issue all over the world. Computer vision techniques are widely utilized in practical crop pest prevention applications due to the rapid development of artificial intelligence technology. However, current deep learning image analytic approaches achieve low accuracy and poor robustness in agriculture pest monitoring task. This paper targets at this challenge by proposing a novel two-stage deep learning based automatic pest monitoring system with hybrid global and local activated feature. In this approach, a Global activated Feature Pyramid Network (GaFPN) is firstly proposed for extracting highly representative features of pests over both depth and spatial position activation levels. Then, an improved Local activated Region Proposal Network (LaRPN) augmenting contextual and attentional information is represented for precisely locating pest objects. Finally, we design a fully connected neural network to estimate the severity of input image under the detected pests. The experimental results on our 88.6K images dataset (with 16 types of common pests) show that our approach outweighs the state-of-the-art methods in industrial circumstances.**

Keywords—Pest Monitoring, Convolutional Neural Network, Global Activated Feature Pyramid Network, Local Activated Region Proposal Network

I. INTRODUCTION

Monitoring pest in agriculture has been a high-priority issue all over the world. The need for better efficiency of inspecting occurrence of pests drives the development of new chemical engineering solutions and innovative pest-monitoring systems, including chemical pesticides [1], image analytic systems [2], automatic adjustable spraying device [3], status estimation of wheat plants [4], remote sensing [5], etc. Computer vision techniques are widely utilized in practical crop pest prevention applications due to the rapid development of artificial intelligence technology. Among these applications, stationary pest trap facilities are the common choice to capture and transform trap images that contain multi-class numerous wild pests in the field [6-9]. Despite that these aforementioned computer vision approaches could enable great success in effective pest monitoring in the wild field, there still remains an open problem due to a challenging fact that many discriminative features of small pest are short of details when hand-crafted features are designed to be selected as pest descriptors. Therefore, towards practical multi-class pest monitoring including localization, classification and severity estimation in the field, it is highly demanding to develop an effective domain specific automatic system.

Recently, the emergence of deep learning techniques has led to significantly promising progress in the field of object detection that requires localization as well as classification [10-12]. Specifically, Convolutional Neural Network (CNN) has exhibited superior capacities in learning invariance in multiple object categories from large amounts of training data. In this context, this paper attempts to study the state-of-the-art deep learning approaches and find out an effective automatic system targeting at solving the challenges of pest monitoring including localization, classification and severity estimation. Our idea is to build a feature pyramid structure on CNN backbone named Global activated Feature Pyramid Network (GaFPN) to extract highly representative features of pests over both depth and spatial position activation levels. Then, an improved Local activated Region Proposal Network (LaRPN) augmenting contextual and attentional information is represented for precisely locating pest objects. Following this motivation, we integrate GaFPN and LaRPN into our two-stage deep learning solution. Finally, we design a fully connected neural network to estimate the severity of input image under the detected pests.

In this paper, we make three major contributions: (1) a novel CNN based pest monitoring system is presented for accurate and effective pest detection. (2) two novel global and local activation branches are introduced to improve the powerfulness of feature extraction and pest localization. (3) we evaluate our system on our 88.6K image dataset on pest localization, classification and severity estimation tasks, which show to outweigh the state-of-the-art deep learning approaches.

978-1-7281-2927-3/19/$31.00 ©2019 IEEE

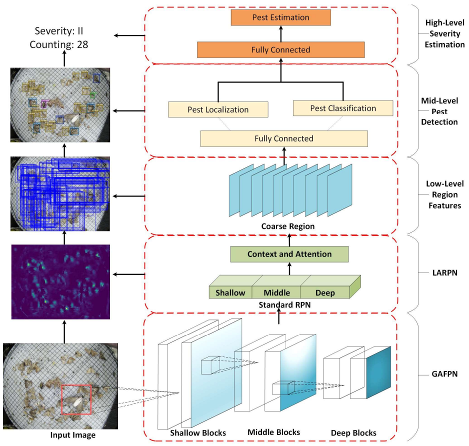

Fig. 1. Workflow of our proposed pest monitoring system

II. System Description

A. Global activated Feature Pyramid Network (GaFPN)

For feature extraction in our pest monitoring task, we build our Global Activated Feature Pyramid Network (GaFPN) shown in Fig. 2 rather than single convolutional network. This property will allow some missing features of tiny pests in pooling layers in one level to be redetected by many pyramid levels. Different from common feature pyramid architecture [12], GaFPN takes full advantages of global information during series of sequential convolution operations. The motivation is that The number of kernels corresponds to be the feature depth and each kernel is learned to extract the specific type of feature such as shape and texture. Thus, one potential way to enhance the quality of features is activating the weights of different kernels (depth of feature maps). On the other hand, there exists a drawback that limited receptive field of convolution operations might result in weaken features of pests in spatial level during training because tiny targets could be confused with nearby context when relatively large kernels are applied. Therefore, it is necessary to apply depth and spatial activation in global level with deep CNN to boost the representational power of pests' feature.

As shown in Fig. 2, feature map from each level is input into our proposed Global Activation Module (GAM) to refine the features, which involves two branches for depth and spatial level activation. In the first part of depth activation branch (the upper), the 3D feature map with shape of $W \times H \times C$ extracted by CNN block is fed into a global pooling layer that takes average in each channel (depth) and generates a lower dimensional (1D) feature ($1 \times 1 \times C$), in which the averaged value represents the global feature of every channel. So the feature vector is learned by a combination of convolutional operations and could be used as depth activation vector. The lower part of GAM in Fig. 4 is spatial activation branch. Similarly, the input 3D feature map ($W \times H \times C$) is fed into a 'global convolution layer' that takes 1×1 size of kernels to reduce the depth of input feature map to 1, which ensures that our spatial factor is learned in spatial level and output a 2D spatial activation vector with shape of $W \times H \times 1$. Then the 2D vector is learned by a series of dilated convolutions [13] to enlarge receptive field. Finally, the output is the broadcast element-wise product of original feature map and two activation vectors.

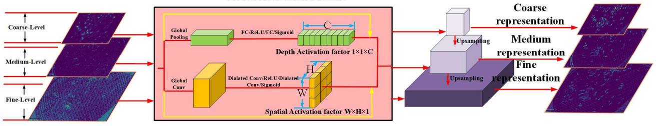

Fig. 2. Structure of Global activated Feature Pyramid Network

B. Local Activated Region Proposal Network (LaRPN)

After extracting powerful enough features, we adopt an improvement of Region Proposal Network (RPN) to enhance the region features in local level by augmenting contextual and attentional information during pest localization phase. So our approach is named Local Activated Region Proposal Network (LaRPN). The first motivation of 'local activation' is that the output region proposals derived from standard RPN might not contain the complete information of target pest, which results in inaccurate pest boxes localization and classification. Secondly, the local spatial positions contribute to the pest regions classification because the key feature for precise region might be the fine-grained characteristics such as colors or shapes of pests' wings. Thus, we introduce contextual and attentional information to precisely locate pest objects

Based on the two motivations, we develop LaRPN to achieve local activation in region proposals, whose structure is shown in Fig. 3. In the first stage, we apply a standard RPN with small scale region templates using sliding window approach to find potential pest locations, in which those templates with Intersection-over-Union (IoU) more than 0.7 to ground truth are extracted as preliminary pest regions. Then the four types of extra contextual regions [14] that are expansion of 1.5 times larger than the preliminary pest regions are augmented and fused into them by Rregion-of-Interest Align [15] operation, in which three different magnifications could cover sufficient contextual information. Thirdly, we develop a self-attention mechanism [16] to activated the spatial positions of these fused regions in local level, in which we add two convolutional layers with 1×1 size kernel to build two extra branches region feature maps. In this way, the relationships among various positions of pests could be extracted by the multiplication of two branches.

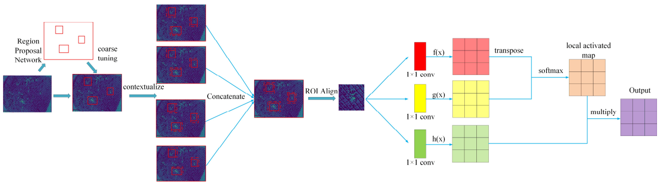

Fig. 3. Structure of Local activated Region Proposal Network

III. Model Optimazation

A. Pest Localization

Localization is a specific problem where the samples are predicted with numerous bounding boxes rather than labels. So differently, it is essential to pay more attention on the spatial accuracy. Thus, we are supposed to employ box regression loss as pest localization training criterion. Referenced by Faster RCNN [11], smooth L1 is selected as the loss function of pest localization task Loss_L:

$$Loss_L = \sum_{i \in (x,y,w,h)} \begin{cases} \tau (t_i - \hat{t}_i)^2, & if \left| t_i - \hat{t}_i \right| \le \tau \\ \left| t_i - \hat{t}_i \right|, & otherwise \end{cases} \qquad (1)$$

Where τ is usually set to 0.5, t_i and $\hat{t}_i$ indicate the coordinates of ground truth and predicted bounding boxes respectively.

B. Pest Classification

After pest localization, we classify each bounding box from LaRPN. Different from binary classification for region proposals in the above, localized boxes are categorized with various types of pest. Therefore, we use multi-class Softmax loss pest classification problem:

$$Loss_C = \sum_{i=1}^{N} -y_i \log(\hat{y}_i) \qquad (2)$$

C. Pest Severity Estimation

In our system, pest severity estimation aims to predict the pest severities of input images, which consist of 5 levels from general to serious corresponding to I-V, in which the 5 severities are labelled by agricultural experts. During high-level semantic estimation task, the input information should be the combined results with those from localization and classification in previous step. In this way, we adopt a variant encoding of one-hot approach [17] that transforms input into a N_{cls}-dimensional vector. In this case, the pest severity estimation is a multi-class classification task so we build several fully connected layers for feature extraction and severity prediction with the Mean Square Error (MSE) loss:

$$Loss_E = -\sum_{i=1}^{N} (y_i - \hat{y}_i)^2 \qquad (3)$$

IV. Experimental Results and Discussion

A. Pest Localization Task

We show pest localization results on our stationary pest image dataset in Table 1, where we compare our method with Faster RCNN [11] and FPN [12] using Inception [18] and ResNet50 [19] as CNN backbones. Observed from Table 1, our approach could significantly outperform another feature pyramid method FPN on two types of CNN backbones, which holds 2.69% and 2.38% Average Precision [20] for localization (AP_L) improvement. Besides, Fig. 4 illustrates the Precision-Recall (PR) Curve for detailed pest localization performance. It can be observed that our method outperforms Faster RCNN by an obvious margin, which could be mainly due to the following two reasons. Firstly, our method with GaFPN applies a pyramid feature extraction architecture and localize pests' regions on multi-level feature maps that could help precisely find pests positions on various scales resulting in pests with different sizes could be searched well, which is evidence from AP_L values of our method in Fig. 6. Secondly, holding global activation factors by our presented GAM for activating the channel and spatial information in global level makes it easier to localize bounding boxes of pests because of much more remarkable features between pests' positions and background.

TABLE 1. Pest Localization Results AP_L

CNN Backbone	Method	AP_L
Inception	Faster RCNN	74.99%
	FPN	76.65%
	Ours	79.34%
ResNet50	Faster RCNN	78.74%
	FPN	80.29%
	Ours	82.67%

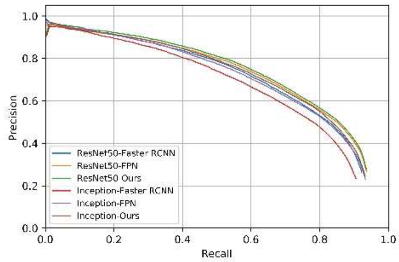

Fig. 4. Precision-Recall curve for pest localization

B. Pest Classification Task

Table 2 presents the pest classification results using different models. Having pest localization information associated with the predicted bounding boxes to pests, our method could achieve more accurate pest recognition performance on 16 pest categories. It is clear that our approach significantly exceeds Faster RCNN in pest classification task on almost all classes using Inception network. Similarly, the homologous advance appears in relatively deep CNN architectures ResNet50, which 3.28% mAP improvement could be obtained. Compared to FPN method, our method still could improve the mAP of pest classification. This gain is largely due to our LARPN's ability to introduce the contextual and local activated information rather than simple fully connected layer for pest recognition, which is helpful to sufficiently learn the features of pests in local level.

TABLE 2. Pest Classification Task Results AP value (%)

Pest ID	**Inception**			**ResNet50**		
	Faster RCNN	**FPN**	**Ours**	**Faster RCNN**	**FPN**	**Ours**
1	51.62	60.24	61.41	57.12	62.13	**64.60**
2	56.26	61.00	63.15	59.70	62.96	**66.01**
3	64.27	67.33	68.22	69.75	70.16	**71.74**
4	80.74	82.10	83.48	83.73	82.82	**84.97**
5	65.65	69.73	71.44	70.17	71.22	**72.07**
6	65.36	68.45	71.61	68.60	68.98	**72.07**
7	63.09	63.30	67.35	68.39	69.46	**71.25**
8	45.31	49.70	51.04	48.57	53.47	**54.50**
9	69.93	71.17	73.36	72.56	72.91	**76.32**
10	75.55	76.27	78.73	79.92	80.58	**80.65**
11	50.71	51.74	54.28	54.45	57.35	**62.36**
12	63.17	66.78	69.06	66.26	69.20	**72.03**
13	77.48	83.31	85.45	84.94	85.18	**85.95**
14	79.43	86.93	**88.21**	87.86	88.03	88.08
15	89.81	89.77	89.82	89.93	89.97	**90.21**
16	69.13	72.51	**75.09**	73.38	74.37	75.05
mean	66.72	70.02	71.98	70.96	72.42	**74.24**

C. Pest Severity Estimation Task

In many conventional machine learning methods, pest severity estimation is considered as a whole image classification task. Differently, our method achieves pest severity estimation based on the input feature vector that fuses the pest localization and classification information from previous tasks as initialization in this task. Table 3 illustrates the pest severity estimation task results, in which we compare our method with other current CNN based image classification methods. As can be seen from the table, our method beats the whole image classification models by a large margin with around 2% accuracy improvement due to the prior pest information.

TABLE 3. Pest severity estimation Task Results Accuracy

CNN Backbone	Method	Accuracy
Inception	Softmax	80.5
	Ours	82.8
ResNet50	Softmax	84.9
	Ours	86.6

Finally, we visualize some pest monitoring images in Fig. 5 that fuse the results from localization, classification and severity estimation tasks. As it can be seen, our method could realize multi-class pest localization and classification under both simple and complicated environments and provide the predicted severity estimation, despite the intractable challenges such as noisy image and tiny objects.

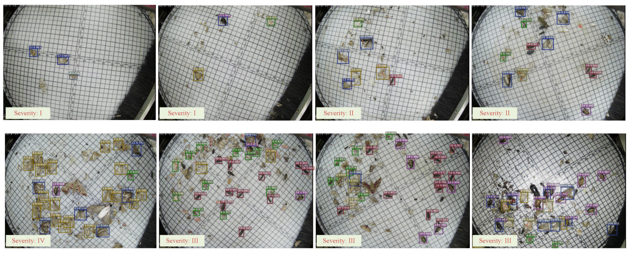

Fig. 5. Examples of pest monitoring results demonstration

V. Conclusion

This paper proposes a novel deep learning approach for automatic pest monitoring in industrial equipment to simultaneously perform three key tasks: localization, classification and severity estimation. Our method successfully realizes efficient and automatic feature extraction with global activated feature pyramid GaFPN structure. Furthermore, we present local activation to enhance position-sensitive features of pest boxes by LaRPN for powerful regions proposal. Under our enriched stationary pest dataset captured by our designed pest monitoring equipment, our method has outperformed the state-of-the-art methods in pest localization, classification and severity estimation tasks. Future work will consider developing more efficient deep learning architecture for real-time pest monitoring.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 31401293, Grant 31671586, and Grant 61773360.

References

[1] B. L. Bures, K. V. Donohue, R. M. Roe and M. A. Bourham, “Nonchemical dielectric barrier discharge treatment as a method of insect control”, IEEE Transactions on Plasma Science, vol. 34, no. 1, pp. 55-62, 2006.

[2] H. J. Liu, S. H. Lee and J. S. Chahl, “A multispectral 3D vision system for invertebrate detection on crops”, IEEE Sensors Journal, vol. 17, no. 22, pp. 7502-7515, 2017.

[3] R. Berenstein, and Y. Edan, “Automatic Adjustable Spraying Device for Site-Specific Agriculture Application,” IEEE Transactions on Automation Science and Engineering, vol. 15, no. 2, pp.641-650, 2018.

[4] Sulistyo, Susanto B., Wai Lok Woo and Satnam Singh Dlay, "Regularized neural networks fusion and genetic algorithm based on-field nitrogen status estimation of wheat plants," IEEE Transactions on Industrial Informatics, vol. 13, no. 1, pp.103-114, 2017.

[5] J. Luo, W. Huang, J. Zhao, J. Zhang, C. Zhao, and R. Ma, “Detecting Aphid Density of Winter Wheat Leaf Using Hyperspectral Measurements,” IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 6, no. 2, pp.690-698, 2013.

[6] W. Ding and G. Taylor, “Automatic moth detection from trap images for pest management,” Computers and Electronics in Agriculture, vol. 123, pp.17-28, 2016.

[7] I. Y. Zayas and P. W. Flinn, “Detection of Insects in Bulkwheat Samples with Machine Vision,” *Transactions of the ASAE*, vol. 41, no. 3, pp. 883, 1998.

[8] J. Cho, J. Choi, M. Qiao, C. W. Ji, H. Y. Kim, K. B. Uhm and T. S. Chon, “Automatic identification of whiteflies, aphids and thrips in greenhouse based on image analysis,” *Red*, vol. 346, no. 246, pp. 244, 2007.

[9] C. Wen, D. E. Guyer and W. Li, “Local feature-based identification and classification for orchard,” insects. *Biosystems engineering*, vol. 104, no. 3, pp. 299-307, 2009.

[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu and A. C. Berg. “SSD: Single shot multibox detector”. In European conference on computer vision, pp.21–37. Springer, 2016.

[11] S. Ren, K. He, R. Girshick and J. Sun, “Faster r-cnn: Towards real-time object detection with region proposal networks,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.

[12] T.Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, “Feature pyramid networks for object detection,” in *IEEE conference on computer vision and pattern recognition*, vol. 1, no. 2, pp. 4, 2017.

[13] P. Wang, P. Chen, Y. Yuan, D. Liu, Z. Huang and X. Hou, “Understanding convolution for semantic segmentation,” in *IEEE Winter Conference on Applications of Computer Vision*, 2018.

[14] P. Hu and D. Ramanan, “Finding Tiny Faces,” in *IEEE conference on computer vision and pattern recognition*, vol. 1, no. 2, pp. 3, 2017.

[15] K. He, et al. "Mask R-CNN," in *2017 IEEE International Conference on Computer Vision (ICCV)*, 2017.

[16] H. Zhang, I. Goodfellow, D. Metaxas and A. Odena. (2018). “Self-Attention Generative Adversarial Networks,” arXiv. [online]. Available: https://arxiv.org/abs/1805.08318.

[17] K. Lee, M. Lam, R. Pedarsani, D. Papailiopoulos and K. Ramchandran, "Speeding up distributed machine learning using codes." *IEEE Transactions on Information Theory*, vol. 64, no. 3, pp.1514-1529, 2018.

[18] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov and A. Rabinovich, “Going deeper with convolutions,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1-9. 2015.

[19] K. He, X. Zhang, S. Ren and J. Sun, “Deep residual learning for image recognition,” in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.

[20] X. Chen, H. Fang, T.Y. Lin, R. Vedantam, S. Gupta, P. Dollár, and C. Zitnick. (2015). “Microsoft COCO captions: Data collection and evaluation server,” arXiv. [online]. Available: https://arxiv.org/abs/1504.00325.