Efficient Remote Sensing in Agriculture via Active Learning and Opt-HRDNet

Desheng Chen, Shuai Xiao, *Member, IEEE*, Meng Xi, *Member, IEEE*, Ling Dai, and Jiachen Yang, *Senior Member, IEEE*

***Abstract*—As the foundation of human survival, the development of agriculture has always played an important role in social development. The quality of agricultural development determines the speed of social progress. With the development of computer science, using computer technology to solve problems related to agricultural development has become an important research direction in current computer development. In recent years, remote sensing detection has received widespread attention, and the application of remote sensing detection technology in agriculture can provide great convenience for the development of agriculture. Benefit from the development of deep learning, remote sensing detection has made gigantic achievements. However, we have to face some challenges. First, deep learning depends on large scale of data with annotations, which expends inestimable human resources. In addition, as the depth of detection network increases, the amount of parameters explosively extends. In this work, we carry out the research based on the two problems. At the beginning, an active learning method considering classification and localization task is proposed. Our method can choose some few but valuable images. It uses 34.48% amount of training set, up to 97.4% baseline performance, and realize the compression of the scale of dataset, which reduces the trouble of manual labeling. Due to imaging and other factors, there exists many small objects in remote sensing images. So we add the mixed convolution, dilated convolution, and mosaic data augmentation modules into HRDNet network. Experiments on the agriculture dataset indicate that the improved algorithm can obtain about 2% higher than HRDNet. To reduce the number of parameters, we adjust the weighted sum ratio of importance scores dynamically. With the pruning ratio of 80%, the model volume has only 184 MB, degrading 70%. Model compression accelerates the detection speed for seven times on NVIDIA AGX Xavier, with a speed of 6 FPS. Our work will lay a foundation for remote sensing detection.**

***Index Terms*—Agriculture dataset, importance scores, small object detection, Xavier.**

Manuscript received 23 November 2023; revised 25 December 2023 and 28 January 2024; accepted 19 February 2024. Date of publication 23 February 2024; date of current version 8 March 2024. This work was supported in part by the National Natural Science Foundation of China under Grant 62271345 and Grant 62301356, and in part by the Joint Fund of Ministry of Education for Equipment Preresearch under Grant 8091B032254. *(Corresponding author: Shuai Xiao.)*

Desheng Chen, Shuai Xiao, Meng Xi, and Jiachen Yang are with the School of Electrical and Information Engineering, Tianjin University, Tianjin 300072, China (e-mail: chendesheng@tju.edu.cn; xs611@tju.edu.cn; ximeng@tju.edu.cn; yangjiachen@tju.edu.cn).

Ling Dai is with the School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing 210023, China (e-mail: 21210531@njnu.edu.cn).

Digital Object Identifier 10.1109/JSTARS.2024.3369189

I. Introduction

REMOTE sensing detection plays an important role in the field of intelligent perception [1]. In the past, researchers usually used manually designed features for remote sensing detection [2]. However, those small targets in remote sensing images makes it difficult to detect, which seriously influences the development of remote sensing detection.

Since 2012, researchers have paid more attention to deep learning [3], [4], [36]. It has been used in many areas in recent ten years, such as image classification and object detection [5]. At the same time, deep learning has also played a huge role in the development of agriculture [6]. Deep learning has played a significant role in promoting the development of weed and pest identification [7], [8], [37]. Homoplastically, remote sensing detection based on deep learning also makes great progress. Han et al. [9] used genetic algorithm to acquire the suitable nonlinear activation method. The fitness score is calculated according to the performance of classification task on the DOTA dataset. Experiments indicate that the proposed method can better capture the features from remote sensing images and improve the detection performance [9]. In order to improve the performance of small object detection, Liu et al. [10] integrated FPN, soft-NMS, and ROI-align module into the faster RCNN framework. Experiments show that the optimized algorithm can obtain richer information of small objects than the primordial network. Wu et al. [11] utilized attention mechanism to focus on important features of images. The authors also improve the training loss based on saliency loss. Multiscale module is integrated into the network to get richer feature characteristic. Wang et al. [12] proposed a coalescent feature extracting framework, which combines context with target information. Lacking the huge number of remote sensing database for training, the authors borrowed the model pretrained on ImageNet. The experiments indicate that the constructed model is superior to the conventional one. Wang et al. [13] designed a remote sensing detection model, which fuses multiscale information. PFR-DBB algorithm is constructed to filter those superfluous bounding boxes. Detection precision of small object detection has been improved by experiments validation. Huang et al. [14] proposed a remote sensing detection algorithm based on lightweight network, which can fast and accurately extract semantic information of targets in images. The authors also transplant the model to the mobile platform and assess the performance. Xu et al. [15] integrated DenseNet module into the original YOLOv3 backbone to better

Fig. 1. Examples of some classes in agricultural dataset.

capture the feature information. The number of detection layers increase to four in order to adapt small target detection in remote sensing images. Wang et al. [16] used super-resolution reconstruction and generative adversarial network to recover the tiny objects in remote sensing images under the fog weather, which can reach a better detection precision after verification.

On the whole, great achievements have been made in remote sensing detection recently. Nevertheless, there are some problems that we have to solve. On the one hand, remote sensing detection based on deep learning requires large amount of manually labeled data [17], [18]. On the other hand, the detection algorithm contains numerous parameters, which restricts the running speed on the embedded device [19], [20]. Researchers usually use active learning method to choose some few but valuable samples from large-scale database, which can effectively compress the scale of datasets [21], [22]. In addition, model compression methods are proposed to degrade the volume of detection networks and the amount of neuron parameters [23], [24].

In our work, an overall process for remote sensing detection is designed, which is made up of active learning, detection algorithm optimization, and model compression. In detail, our contributions are listed as follows.

1) To degrade the amount of remote sensing dataset, we synthetically consider classification and localization results and construct an active method, which can help select few, but informative samples.
2) The mixed convolution, dilated convolution, and mosaic data augmentation ways are integrated into the original HRDNet to improve the performance for remote sensing detection. Experiments show that the optimized network outperforms incipient one for about 2%.
3) To reduce the number of parameters, we carry out channel pruning method on the improved network. When the pruning ratio is 80%, the model volume can lower 70%. Next, we deploy the compressed model on the resource-limited devices, which can run at 6 FPS on Xavier board.

II. Material and Method

A. Dataset

In our research, we took photos of farmland, established our own agricultural dataset, and conducted experiments on the agricultural dataset. We took 880 agricultural remote sensing photos and obtained 880 remote sensing samples, including 4400 objects. The number of images used for training, validation, and testing is 580, 200, and 100, respectively. In order to expand the number of training sets, this article processed the images in the training set using methods, such as scaling, rotation, and deformation, expanding to 1740 remote sensing samples. Some examples of each class in the dataset are shown in Fig. 1.

B. Overall Framework

Based on the dataset, we can carry out our study for active learning, detection algorithm optimization, and filter pruning. Fig. 2 shows the overall framework of this work. There is no doubt that remote sensing detection relying on large-scale datasets has made great progress in recent years [25], [26]. Meanwhile, we have no choice but to admit that large amounts of images are unaffordable to annotate, which cost lots of human resources [27], [28]. So in this work, we use active learning based on object detection task to choose fewer yet valuable remote sensing samples, which may ease the double of manual annotation.

Following the acquisition of a small-scale dataset, the emphasis of the research shifted toward remote sensing object detection. There are some small objects in images, which increases the difficulty of detection and affects the detection

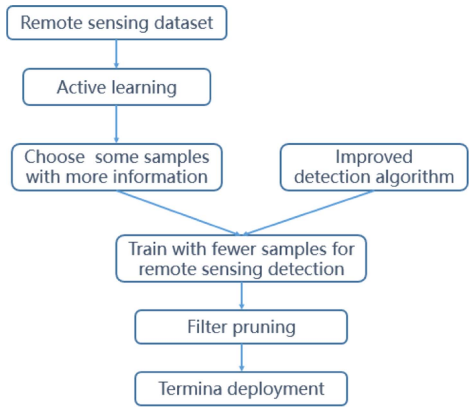

Fig. 2. Overall framework of this work.

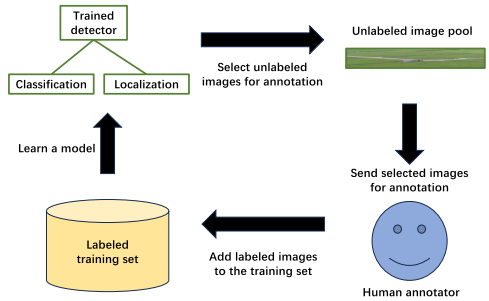

Fig. 3. Process of active learning.

performance. Therefore, we use some tricks to construct the improved detection algorithm. Considering the large number of parameters, filter pruning is carried out to reduce the model size of improved detection algorithm, and realize the deployment on the embedded device.

C. Active Learning Method

Active learning can choose some useful samples from large-scale datasets and then label them. In recent years, research on active learning has developed rapidly. Fig. 3 shows the main process of active learning for object detection task.

Given a remote sensing dataset, we divide it into two parts at first, including a small training set and large query set. The training set is used for detection model training, whereas the query set is used for choosing samples. After obtaining a pretrain detection model, we can set some metrics according to object detection task to measure the value of the samples in query set. Then, some informative images will be chosen, labeled and put into the training set. Obviously the number of images in the training set become larger after every iteration.

In this work, we carry out the research about active learning based on object detection, considering classification uncertainty and localization robustness. According to Fig. 3, a parameterized detection model will be acquired using the training set. Then, we apply the pretrain model on the query set to obtain the outputs of each image. If targets in images can be detected, then, we can get the classification possibility and localization information of the target and obtain the metric scores for choosing samples in active learning. Classification uncertainty can be calculated by

$$U(B) = 1 - P_{\max}(B) \tag{1}$$

According to (1), *U(B)* represents the classification uncertainty results of bounding boxes. $P_{\max}(B)$ is the highest value of classification possibility among all the classes. If $P_{\max}(B)$ is low, it means the detector can not classify the target centainly. We deem that that image can provide more information for classification if labeled. Supposing $P_{\max}(B)$ is on the verge of 1, the detector can classify well so we do not need to choose that image. In short, we want to choose the images in the query set with large values of *U(B)* to improve the classification performance. Besides, localization information of bounding boxes is taken into account. The process of calculating localization robustness can be represented by

$$R\left(B_j\right) = \mathrm{GIOU}\left(B_j, B_j^{\mathrm{gt}}\right). \tag{2}$$

From (2), we calculate the value of GIOU between the detection result and real bounding box. B_j represents the result obtained by the pretrain detection model. B_j^{gt} is the ground truth of this target. $R(B_j)$ can represent the localization robustness of this image. In this part, we want to select the images with low localization robustness, which will help increase the localization exactitude. We consider both classification possibility and localization information. So the total metric can be calculated as

$$S\left(B_j\right) = \left|R\left(B_j\right) + P_{\max}\left(B_0^j\right) - 1\right|. \tag{3}$$

Combining (1) with (2), the final value for choosing samples is shown in (3). $S(B_j)$ represents the total score of each image, B_0^j represents the bounding box associated with the image that is used as a reference in the calculation of $P_{\max}(B_0^j)$. In our active learning method, those images with low value of $S(B_j)$ in the query set can be selected. They contain more useful information and contribute to improving detection performance compared with random choosing.

D. Remote Sensing Detection Algorithm Optimization

Remote sensing images consist of different scales of objects and small targets are no exception. So in this part we pay attention to small object detection. In concrete, we integrate mixed convolution and dilated convolution into HRDNet algorithm [29]. Fig. 4 shows the structure of Opt-HRDNet.

From Fig. 4, Opt-HRDNet consists of image pyramid and feature pyramid. First, three images with different resolutions are fed into the backbone to extract feature maps. $S_0, S_1,$ and S_2 represent three CNNs with different depth. Opt-HRDNet uses a deeper CNN for images with small resolution. In contrast, a fleet CNN is used for large resolution. After obtaining feature maps with different scales, indicated as $G_0, G_1,$ and G_2, Opt-HRDNet merges them according to feature pyramid. Finally, five luxuriant feature maps $F_0', F_1', F_2', F_3',$ and F_4' are used for classification and localization, which can detect multiscale targets. As for YOLOv3 and SSD, the feature information in not rich enough, resulting in low detection performance for rather small targets.

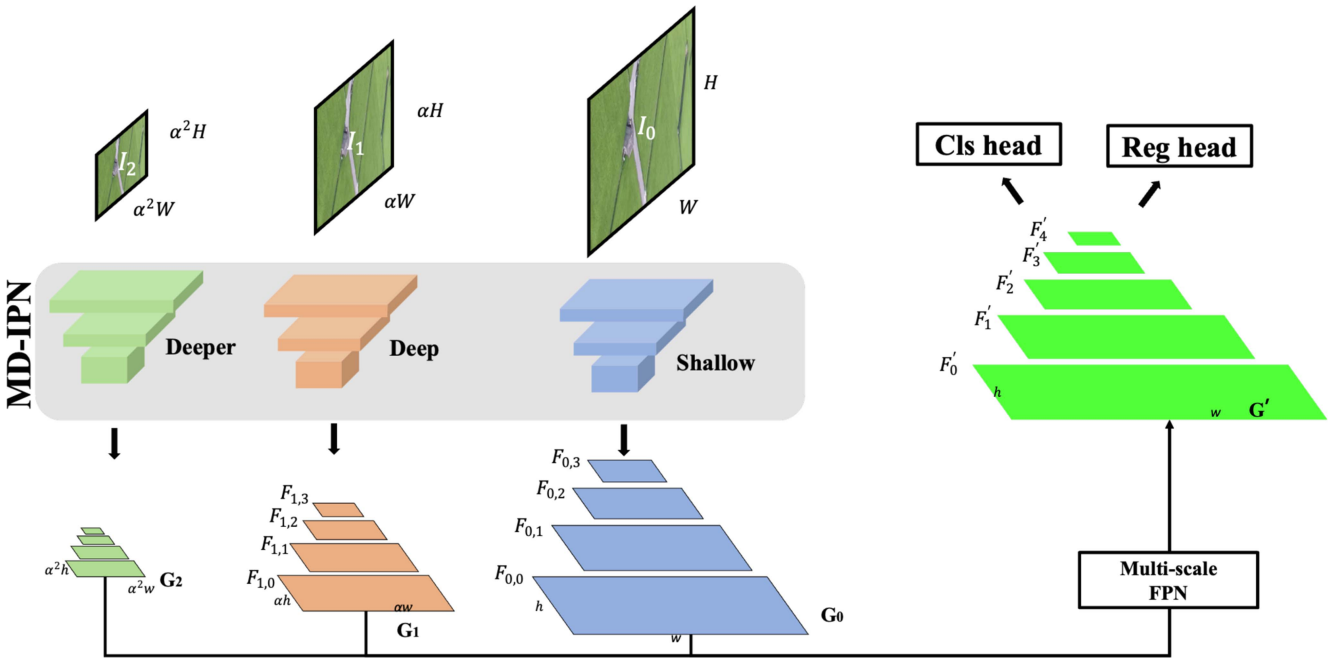

Fig. 4. Architecture of Opt-HRDNet.

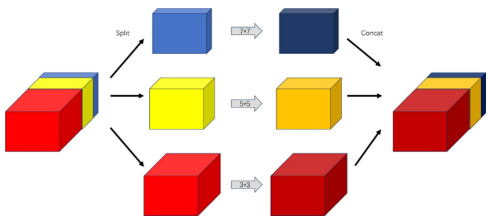

Fig. 5. Architecture of mixed convolution.

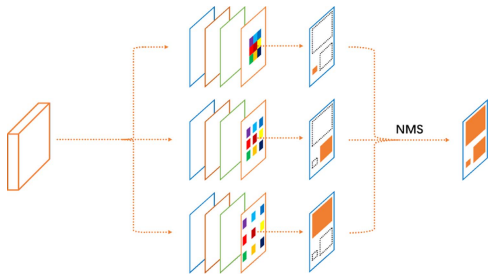

Fig. 6. Architecture of DCNet.

Next, we add some tricks based on HRDNet. To further improve the detection performance for small targets, we add mixed convolution [30] and dilated convolution [31] into HRDNet. The architecture of mixed convolution is shown in Fig. 5.

We use mixed convolution to replace the common convolutional layers in backbone $S_0, S_1,$ and S_2. The mixed convolution consists of three different filters, including 3×3, 5×5, and 7×7. The channels of input feature maps are divided into three parts, which can be calculated with three different sizes of filters, respectively. During calculating, we should guarantee that the width and height of three output feature maps are the same. Meanwhile, the number of channels may be disparate. And, then, we concatenate the three output features along the channel dimension. Filters with different sizes can bring disparate receptive fields. After using mixed convolution, the diversity of feature maps obtained by $S_0, S_1,$ and S_2 can increase remarkably compared with the common convolutional layers. In mixed convolution, convolution kernels of different sizes are applied in the same layer. It can blend features of different receptive fields and improve feature extraction ability.

In front of the final detection layers, we integrate DCNet applied on feature maps $F_0', F_1', F_2', F_3',$ and F_4'. The architecture of DCNet is displayed in Fig. 6. We use DCNet behind each feature map among $F_0', F_1', F_2', F_3',$ and F_4'. From Fig. 6, DCNet contains three dilated convolution layers with dilation ratio values of 1,2,3, which represent different receptive fields.

Then, NMS is used on the three outputs to filter superfluous bounding boxes. The layers with dilation ratios 1,2,3 are responsible for predicting small, middle, and large targets, respectively. Benefited from different dilation ratios, DCNet can increase the richness of feature maps. Dilated convolution can capture multiscale context information, which plays an important role in vision tasks.

In addition, during detection model training phase, we use mosaic data augmentation method borrowed from YOLOv4 [32]. It can mix four different training images, which helps provide context information and increase the robustness of detection algorithm. Finally, we obtain an improved small object detection network with high precision.

E. Channel Pruning

The optimized small object detection network contains huge number of parameters, which requires high-performance servers as support [33]. This is unaffordable for portable embedded

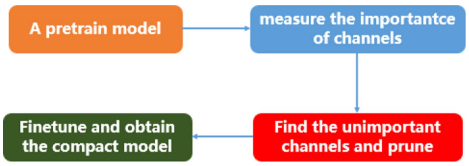

Fig. 7. Process of channel pruning.

devices [34]. Facing this problem, we use filter pruning method to delete those unimportant channels in convolutional layers and realize the reduction of model size and number of parameters. Fig. 7 shows the main process of common channel pruning methods, containing four main steps.

First, a pretrain model with numerous parameters should be trained. Then, we should choose a suitable metric to measure the importance of convolutional channels. According to the values of channel importance score, those channels with small score values will be pruned. Usually channel pruning may lead to the degrading of accuracy, we should finetune the compressed model to compensate accuracy and acquire the compact model, which has fewer parameters than the uncompressed model [35].

In this work, we consider the two parameters in BN layers to measure the importance of convolutional channels. The metric can be calculated by

$$S_l^i = w_1^* \gamma_l^i + w_2^* \beta_l^i \tag{4}$$

where l and i represent the lth layer in the network and ith channel in lth layer, respectively. S is the channel importance score. γ and β correspond to the scale and shift factor in BN layers. $w1$ and $w2$ represent the weighted parameters. We use the weighted sum of scale and shift factor to calculate the score. Given a channel pruning ratio ρ, we can prune those channels with small importance scores. Then, a pruning candidate network will generate. The fitness of this candidate network can be calculated by

$$f = \frac{\alpha}{P/P_{\text{init}} + F/F_{\text{init}}} \tag{5}$$

where f represents the fitness score. P, F, and α represent the number of parameters, FLOPs and detection performance on the testing set. Fitness scores can measure how well a species is adapted to its environment in genetic algorithm. Equation (5) means we want to obtain a compressed model with higher performance and fewer parameters.

In channel pruning phase, we adjust the weighted parameters in (4) based on crossover and variation. Then, many different combinations of weighted parameters will generate. According to these combinations and pruning ratio, we can get the same amount of pruning candidate networks. By calculating fitness scores of these candidate networks with (5), finally an optimal pruning network will be chosen.

After channel pruning, we transplant the compressed model to NVIDIA AGX Xavier platform, which will provide reference for the deployment of remote sensing detection algorithm.

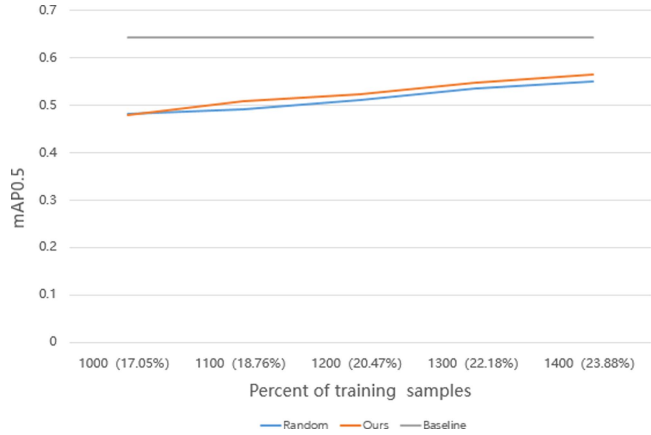

Fig. 8. Active learning results of C = 25 on the agriculture dataset.

Fig. 9. Active learning results of C = 50 on the agriculture dataset.

III. Experiments and Results

A. Experiment Details

For active learning experiments, the software environment includes Python 3.7, PyTorch 1.7, Numpy, PIL, and so on. The number of samples in the agriculture dataset for validation and testing is 200 and 100, respectively. In the initial, we randomly choose 300 samples to make up labeled pool from training set, which contains 1740 remote sensing images in total, in order to train the Opt-HRDNet detection network. The rest of the samples in training set are named as the unlabeled pool. Then, the values of sampling interval C are set as 25/50/75. It means we choose 25/50/75 valuable samples from the unlabeled pool after each active learning iterative process. Next, these chosen samples are labeled and added into the labeled pool, which will be used for the next training process. We consider mAP 0.5 as the detection metric for validation and testing. In our experiments, we repeat the active learning procedure four times.

B. Results for Active Learning

Experiments on the agriculture dataset are carried out so as to validate our active learning method. Figs. 8–10 indicate active learning results of three sampling intervals obtained by different methods. If all the images in the training set are used for training,

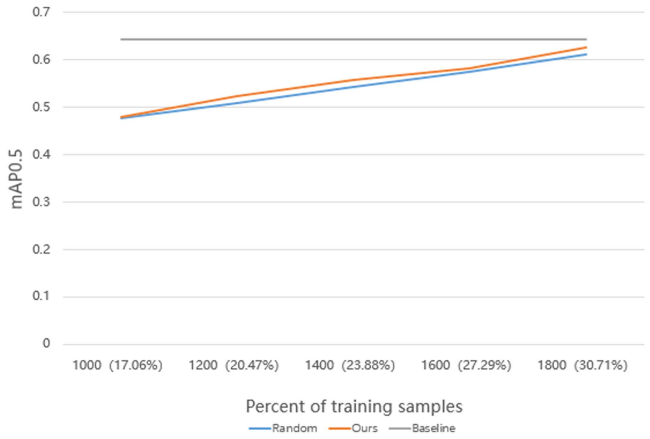

Fig. 10. Active learning results of C = 75 on the agriculture dataset.

we can get the baseline detection metric mAP0.5 of 64.3% on the testing set. The *x*-coordinate and *y*-coordinate express the number of samples in the labeled pool and the detection performance mAP0.5 during active learning process.

There are three curves in the figures. "Random," "Ours," and "Baseline" show the results of random sampling, our method, and the baseline detection performance, respectively. From Fig. 8, our active learning can mAP0.5 metric of 56.5% with 22.99% of the training set, reaching 87.9% of the baseline. Our method is superior to random sampling for 1.4%. When sampling interval C = 50, using 28.74% amount of the training set, random sampling can acquire mAP0.5 of 56.5%, 1.1% lower than our method according to Fig. 9. When C = 75, it means we choose 75 images from unlabeled pool using different methods. By 34.48% of the training set, our method can get mAP0.5 of 62.6%, reaching 97.4% of the baseline metric. Compared with mAP0.5, Random sampling is 1.4% inferior to our method. Equation (3) will help us choose few but beneficial samples, which will help improve detection performance. In contrast, random sampling may choose some unuseful samples. According to the active learning results, our method outperforms random sampling by different sampling intervals on the agriculture dataset.

C. Results for Detection Algorithm Optimization

After active learning, few but valuable samples from training set are obtained. According to Fig. 9, a total of 500 images were utilized for the subsequent experiments, accounting for 28.74% of the entire training set. In the agriculture dataset, there are many small targets existing in the remote sensing images, which affects the detection precision. To further increase the detection performance, we integrate the mixed convolution, dilated convolution, and mosaic data augmentation approach into the original HRDNet. In the experiments, we train the Opt-HRDNet with two Titan Xp GPUs on the Centos system. Table I shows the results of detection algorithm optimization. According to Table I, as for the three metrics, the Opt-HRDNet is 2.1%, 2.5%, and 2.6% higher than original HRDNet, respectively. The precision, recall, and mAP0.5 have increased above 2% compared with the original HRDNet.

TABLE I
RESULTS FOR ALGORITHM OPTIMIZATION

	Precision	Recall	mAP0.5
HRDNet	59.2%	58.4%	57.6%
Opt-HRDNet	61.3%	60.9%	60.2%

TABLE II
RESULTS FOR MODEL COMPRESSION

	Parameters	Model size	mAP0.5	Compression ratio
Opt-HRDNet (baseline)	152M	608MB	60.2%	–
Pruning-50%	67M	268MB	59.8%	55.9%
Pruning-80%	46M	184MB	59.5%	69.7%

The experiment results indicate that Opt-HRDNet integrating the mixed convolution, dilated convolution, and mosaic data augmentation method can help improve the detection performance on the agriculture dataset. There exist different sizes of filters in the mixed convolution. The advantages of different convolution kernel sizes enable better feature extraction under different resolutions. Dilated convolution can arbitrarily enlarge the receptive field without introducing additional parameters. Data augmentation will enrich the feature information of small objects. Thus, the optimized HRRNet outperforms the original network.

D. Results for Channel Pruning

The Opt-HRDNet consists of the enormous amount of model parameters. To realize the decrease of model volume, we use channel pruning to compress the detection model. Table II shows the results of model compression. In channel pruning experiments, we set the pruning ratio of 50% and 80%.

From Table II, the Opt-HRDNet has 152 M parameters and the model volume reaches 608 MB. When the pruning ratio is 50%, the pruned model contains 67 M parameters. The compression ration reaches 55.9% with 0.4% of mAP0.5 drop. When we prune 80% of channels, the model after pruning has only 184 MB, up to a compression ratio about 70%. Finally, we transplant the compressed model to NVIDIA AGX Xavier board. The optimized model without pruning can run no more than 1 FPS. After pruning 80%, the detection speed of compressed model increased a lot, running at 6 FPS. Experiments verify that model compression is beneficial to accelerate the running speed of detection network.

In the future, we will further research how to measure the information when selecting samples from labeled pool. We aim to choose those most informative images. In addition, we want to study how to compress the detection model to the utmost with tolerant performance drop, which can help realize real-time detection on some convenient embedded edge devices. In addition, we will use int8 to replace float32 by model quantization. Also, TensorRT will be utilized to replace PyTorch deepl learning framework and computing resources of CPU and GPU should be made full use of.

IV. CONCLUSION

In this article, we propose a whole framework containing active learning, detection algorithm optimization, and model

compression for remote sensing images. Concretely, we design an active learning method for detection task, which takes both classification and localization results into account. Experiments show that our method can choose more valuable samples than random sampling. To improve the detection performance for small targets, we add the mixed convolution, dilated convolution, and mosaic data augmentation ways into the primordial HRDNet and obtain 2% of mAP0.5 improvement. Then, channel pruning based on dynamic weighted is carried out on the Opt-HRDNet. Under the condition of 80% pruning ratio, we can prune most of model parameters. After compression, the pruned model has only 46 M parameters, reaching a ratio of 69.7% approximately. Ultimately, the pruned model is deployed on the Xavier board, running at 6 FPS, seven times faster than uncompressed model.

REFERENCES

[1] S. Yang and X. Zhao, "Remote sensing image change saliency detection technology," *J. Phys. Conf. Ser.*, vol. 1069, 2018, Art. no. 012110.

[2] J. Yang et al., "Efficient data-driven behavior identification based on vision transformers for human activity understanding," *Neurocomputing*, vol. 530, pp. 104–115, 2023.

[3] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2015, pp. 1–9, doi: 10.1109/CVPR.2015.7298594.

[4] G. Lan, S. Xiao, J. Yang, J. Wen, and M. Xi, "Generative AI-based data completeness augmentation algorithm for data-driven smart healthcare," *IEEE J. Biomed. Health Inform.*, early access, Oct. 30, 2023, doi: 10.1109/JBHI.2023.3327485.

[5] J. Yang, C. Cheng, S. Xiao, G. Lan, and J. Wen, "High fidelity face-swapping with style ConvTransformer and latent space selection," *IEEE Trans. Multimedia*, vol. 26, pp. 3604–3615, 2023, doi: 10.1109/TMM.2023.3313256.

[6] Y. Li, X. Chao, and S. Ercisli, "Disturbed-entropy: A simple data quality assessment approach," *ICT Express*, vol. 8, pp. 309–312, Sep. 2022, doi: 10.1016/j.icte.2022.01.006.

[7] Y. Yang, Y. Li, J. Yang, and J. Wen, "Dissimilarity-based active learning for embedded weed identification," *Turkish J. Agriculture Forestry*, vol. 46, pp. 390–401, 2022, doi: 10.55730/1300-011X.3011.

[8] Y. Li and X. Chao, "Toward sustainability: Trade-off between data quality and quantity in crop pest recognition," *Front. Plant Sci.*, vol. 12, 2021, Art. no. 811241, doi: 10.3389/fpls.2021.811241.

[9] Y. C. Han, J. Wang, and L. Lu, "A typical remote sensing object detection method based on YOLOv3," in *Proc. 4th Int. Conf. Mech. Control Comput. Eng.*, 2019, pp. 520–5203, doi: 10.1109/ICMCCE48743.2019.00121.

[10] R. Liu, Z. Yu, D. Mo, and Y. Cai, "An improved Faster-RCNN algorithm for object detection in remote sensing images," in *Proc. 39th Chin. Control Conf.*, 2020, pp. 7188–7192, doi: 10.23919/CCC50068.2020.9189024.

[11] Q. Wu, X. Yuan, Z. Yao, and Z. Chai, "Attention-based object detection with saliency loss in remote sensing images," *J. Electron. Imag.*, vol. 30, pp. 013 007–013 007, 2021.

[12] E. Wang et al., "Deep fusion feature based object detection method for high resolution optical remote sensing images," *Appl. Sci.*, vol. 9, 2019, Art. no. 1130.

[13] X. Wang, Y. Ban, H. Guo, and L. Hong, "Deep learning model for target detection in remote sensing images fusing multilevel features," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, 2019, pp. 250–253, doi: 10.1109/IGARSS.2019.8898759.

[14] J. Huang, X. Zhang, and H. Xu, "Fast object detection for remote sensing data using a mobile device," in *Proc. IEEE 6th Int. Conf. Big Data Analytics*, 2021, pp. 108–112, doi: 10.1109/ICBDA51983.2021.9402979.

[15] D. Xu and Y. Wu, "Improved YOLO-V3 with DenseNet for multi-scale remote sensing target detection," *Sensors*, vol. 20, 2020, Art. no. 4276.

[16] Y. Wang, G. Sun, and S. Guo, "Target detection method for low-resolution remote sensing image based on ESRGAN and ReDet," *Photonics*, vol. 8, 2021, Art. no. 431.

[17] Y. Senzaki and C. Hamelain, "Active learning for deep neural networks on edge devices," 2021, *arXiv:2106.10836*.

[18] Z. Zhao, Z. Zeng, K. Xu, C. Chen, and C. Guan, "DSAL: Deeply supervised active learning from strong and weak labelers for biomedical image segmentation," *IEEE J. Biomed. Health Inform.*, vol. 25, no. 10, pp. 3744–3751, Oct. 2021, doi: 10.1109/JBHI.2021.3052320.

[19] X. Ding et al., "ResRep: Lossless CNN pruning via decoupling remembering and forgetting," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 4490–4500, doi: 10.1109/ICCV48922.2021.00447.

[20] Z. Huang, W. Shao, X. Wang, L. Lin, and P. Luo, "Rethinking the pruning criteria for convolutional neural network," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2021, pp. 16 305–16 318.

[21] C.-C. Kao, T.-Y. Lee, P. Sen, and M.-Y. Liu, "Localization-aware active learning for object detection," in *Computer VisionACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 26, 2018, Revised Selected Papers, Part VI 14*, 2019, *arXiv:1801.05124*.

[22] S. Huang, T. Wang, H. Xiong, J. Huan, and D. Dou, "Semi-supervised active learning with temporal output discrepancy," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 3427–3436.

[23] Z. Zhan et al., "Achieving on-mobile real-time super-resolution with neural architecture and pruning search," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 4801–4811.

[24] D. Wu and Y. Wang, "Adversarial neuron pruning purifies backdoored deep models," *Advances Neural Inform. Process. Syst.*, vol. 34, 2021, *arXiv:2110.14430*.

[25] J. Choi, I. Elezi, H.-J. Lee, C. Farabet, and J. M. Álvarez, "Active learning for deep object detection via probabilistic modeling," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 10 244–10 253.

[26] Z. Qu, J. Du, Y. Cao, Q. Guan, and P. Zhao, "Deep active learning for remote sensing object detection," 2020, *arXiv:2003.08793*.

[27] A.-K. A.-Tamimi, E. B.-Isaa, and A. A.-Alami, "Active learning for arabic text classification," in *Proc. Int. Conf. Comput. Intell. Knowl. Economy*, 2021, pp. 123–126, doi: 10.1109/ICCIKE51210.2021.9410758.

[28] Z. Yetiştiren, C. Özbey, and H. E. Arkangil, "Different scenarios and query strategies in active learning for document classification," in *Proc. 6th Int. Conf. Comput. Sci. Eng.*, 2021, pp. 332–335.

[29] Z. Liu, G. Gao, L. Sun, and Z. Fang, "HRDNet: High-resolution detection network for small objects," in *Proc. IEEE Int. Conf. Multimedia Expo*, 2021, pp. 1–6, doi: 10.1109/ICME51207.2021.9428241.

[30] M. Tan and Q. V. Le, "MixConv: Mixed depthwise convolutional kernels," 2019, *arXiv:1907.09595*.

[31] Y. Li, Y. Chen, N. Wang, and Z.-X. Zhang, "Scale-aware trident networks for object detection," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2019, pp. 6053–6062, doi: 10.1109/ICCV.2019.00615.

[32] A. Bochkovskiy, C-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," 2020, *arXiv:2004.10934*.

[33] S. Liu et al., "Sparse training via boosting pruning plasticity with neuroregeneration," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2021, pp. 9908–9922.

[34] C. Zhao, Y. Ge, F. Zhu, R. Zhao, H. Li, and M. Salzmann, "Progressive correspondence pruning by consensus learning," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 6444–6453, doi: 10.1109/ICCV48922.2021.00640.

[35] B. Li, B. Wu, J. Su, and G. Wang, "EagleEye: Fast sub-net evaluation for efficient neural network pruning," in *Proc. Eur. Conf. Comput. Vis.*, 2020, pp. 639–654, https://doi.org/10.1007/978-3-030-58536-538

[36] G. Lan et al., "Image aesthetics assessment based on hypernetwork of emotion fusion," *IEEE Trans. Multimedia*, vol. 26, pp. 3640–3650, 2023.

[37] J. Yang, G. Lan, Y. Li, Y. Gong, Z. Zhang, and S. Ercisli, "Data quality assessment and analysis for pest identification in smart agriculture," *Comput. Elect. Eng.*, vol. 103, 2022, Art. no. 108322.

Desheng Chen is currently working toward the Ph.D. degree in communication and information systems with the School of Electrical and Information Engineering, Tianjin University, China.

His research interests include marine exploration, image processing, marine intelligent equipment control, artificial intelligence, and Big Data computing.

Shuai Xiao (Member, IEEE) received the Ph.D. degree in communication and information systems from the School of Electrical and Information Engineering, Tianjin University, Tianjin, China, in 2022.

He is currently an Associate Researcher with Tianjin University. His research interests are image processing, image quality assessment, computer vision, and image forensics.

Meng Xi (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in communication and information engineering from Tianjin University, Tianjin, China, in 2018, 2020, and 2023, respectively.

She is currently a postdoctor with Tianjin University. Her research interests include deep reinforcement learning, AUV control algorithm, and machine learning.

Ling Dai is currently working toward the bachelor's degree in electrical engineering and automation with the School of Electrical and Automation Engineering, Nanjing Normal University, Nanjing, China.

His research interests include artificial intelligence, Big Data computing, control system engineering, and intelligent robot design and manufacturing.

Jiachen Yang (Senior Member, IEEE) received the M.S. and Ph.D. degrees in communication and information engineering from Tianjin University, Tianjin, China, in 2005 and 2009, respectively.

He was a Visiting Scholar with the Department of Computer Science, School of Science, Loughborough University, Loughborough, U.K. He is currently a Professor with Tianjin University. His research interests include ocean information processing, Internet of Things, cloud computing, Big Data analytics, stereo vision research, pattern recognition, and image quality evaluation.