IGARSS 2023 - 2023 IEEE International Geoscience and Remote Sensing Symposium | 979-8-3503-2010-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IGARSS52108.2023.10282790

CO-DETECTOR: TOWARDS COMPLEX OBJECT DETECTION WITH CROSS-PART FEATURE LEARNING IN REMOTE SENSING

Shuai Yuan[1,4], Juepeng Zheng[2,4], Yanlong Huang[1], Jierui Liu[1], Haohuan Fu[3,4], Ray C. C. Cheung[1]

1 Department of Electric Engineering, City University of Hong Kong, Hong Kong, China;
2 School of Artificial Intelligence, Sun Yet-Sun University, Zhuhai, China;
3 Ministry of Education Key Laboratory for Earth System Modeling, Department of Earth System Science, Tsinghua University, Beijing, China;
4 Tsinghua University (Department of Earth System Science)- Xi'an Institute of Surveying and Mapping Joint Research Center for Next-Generation Smart Mapping, Beijing, China.

ABSTRACT

Object detection in remote sensing imagery builds the essential foundation of aerial and satellite image understanding, being an important role in many common real-world tasks and attracting world-wide attention. In recent years, despite the great progress of common object detection in remote sensing and the proven success of deep learning in this field, yet complex object detection which consists of multiple objects with variable layouts in remote sensing (e.g., coal-fired power plant, airport, sewage treatment plant, etc.) is still challenging for complex composite spatial relationship, non-rigid boundaries, and complicated surrounding textures. These challenges necessitate developing specific complex object detection methods to learn inter-relationship and distinctive and discriminative features in complex objects. To address this problem, in this paper, we propose a method, i.e., CO-Detector, in an end-to-end manner, to achieve various complex composite object detection in remote sensing images with high accuracy and efficiency. The effectiveness of CO-Detector is built on three main parts: (a) First, as surrounding contexts are normally complicated and similar to complex objects, we propose a Tandem Attention Network (TAN), including a channel enhanced network and a spatial enhanced network, with a K-global max/average pooling, to restrain noise disturbance and highlight complex object features and boundaries. (b) Second, we design a Part Region Proposal Network (P-RPN) to learn the inter-relationship between parts in one object, generating part proposals and locating discriminative and distinctive object parts finely. (c) Third, to detect the whole complex object as well as the parts, we propose a Part Detection Network (PDN) to detect the individual parts, and detect the whole object through multi-level fused features. We train our CO-Detector model with three selected categories (i.e., coal-fired power plant, airport, oil storage tank) in three datasets, and conduct comparative experiments to evaluate and verify the performance. The comprehensive experiment results show that our CO-Detector achieves a mAP of 80.23%, outperforming 4.17%-17.83% against other cutting-edge deep learning-based detection methods. The experiment results indicate our CO-Detector has promising performance and potential in various complex object detection in high-resolution remote sensing images, pending to be utilized in real large-scale applications.

Index Terms— Complex composite object detection, inter-relationship, part region proposal, remote sensing image.

1. INTRODUCTION

Complex composite objects provide essential support for our society. For example, coal-fired power plants guarantee the power generation for daily life; airports support convenient transportation. This kind of importance requires accurate monitoring for the sake of better urban management and planning. However, there exist several challenges which still hinder complementation. First, the complex composite objects normally consist of multiple parts. For instance, a coal-fired power plant mainly includes chimneys and condensing towers. The discrete parts own variable layouts and form uncertain inter-relationships between each other, which leads to non-rigid and blurred boundaries of the whole object. Second, these composite objects often locate in industrial areas. Surrounding textures with similar outward appearance interferes the target feature extraction. Besides, these geographic entities weaken the spatial inter-relationships between parts of objects, making it more difficult to detect the whole objects. Third, most of existing object detection methods are designed for single object detection, (e.g., vehicle detection [1,2], building detection [3], tree crown detection [4,5], etc.). Even though CNN-based methods have gained big achievement in many aspects in remote sensing, especially in object detection [6,7], these CNN-based object detection algorithms cannot bridge the gaps between the low-level features and the high-

level understanding caused by distributed inter-relationships between multiple parts.

To address the problem, in this paper, we propose a part-based method, i.e., CO-Detector, in an end-to-end manner, to achieve various complex composite object detection in remote sensing images with high accuracy and efficiency. Three components in CO-Detector, i.e., Tandem Attention Network (TAN), Part Region Proposal Network (P-RPN), and Part Detection Network (PDN) ensure the effectiveness. The evaluation on selected datasets and comparison with other cutting-edge methods prove the potential of our proposed method.

2. USED DATASET

Due to the lack of a specific composite object detection dataset in remote sensing, we collect composite object detection dataset from existing datasets, i.e., DIOR dataset [8], Airbus oil storage tank dataset (AirOil) [9], and FFPP-BUAA60 dataset (FFPP) [10]. The specific dataset collection detail is concluded in Table 1. After data augmentation (i.e., blur, rotation, mirror, etc.), we then divide the dataset into training and testing parts in 7:3 ratio.

Table 1. Statistic of our collected dataset.

Dataset	Category	Resolution	Number
DIOR [8]	Airport	0.1-1m	3,909
AirOil [9]	Oil tank	1m	2,450
FFPP [10]	power plant	1m	892

3. METHODOLOGY

Because a typical composite object consists of various components with non-rigid spatial relationships and blurred boundaries, as well as similar complex textures surroundings, it is desirable to firstly enhance feature representation and cripple the noise disturbance, and then locate discriminative part features to generate part proposals, and finally conduct part and whole classification. As a result, our proposed CO-Detector framework includes (as shown in Fig. 1): a Tandem Attention Network (TAN), a Part Region Proposal Network (P-RPN), and a Part Detection Network (PDN).

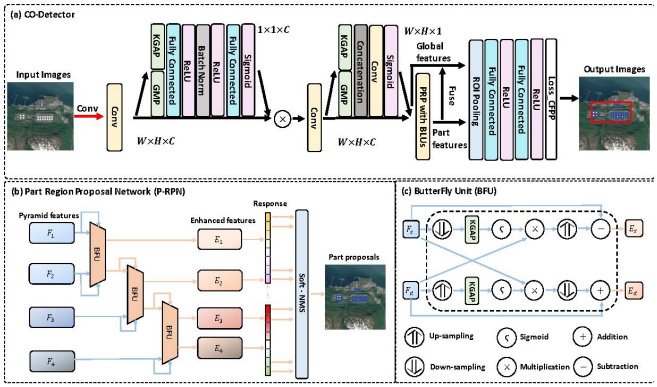

Fig. 1. The architecture of our CO-Detector.

3.1. Tandem Attention Network (TAN)

Composite object owns various components with a non-rigid spatial relationship and blurred boundaries. In addition, the complex background can cause massive obstruction when extracting features and generating RoIs, increasing the false positives and decreasing the accuracy. To address this problem, it is essential to strengthen features of every part and weaken noise disturbance. Here, similar with [11], we propose Tandem Attention Network (TAN), containing both a Channel Enhanced sub-Network (CEN) and a Spatial Enhanced sub-Network (SEN) located beside the backbone in a tandem manner.

CEN has the squeeze-and-excitation block. In the squeeze section, we first propose a K-global average pooling to achieve feature compression on each channel. Compared with the common global average pooling which takes care of the whole image, K-global average pooling concentrates on the K-highest peak response from the feature maps to integrate the whole image across the spatial aspect, which filters other irrelevant noises. Thus, the input feature maps are transformed into K-global average pooling result and max pooling result respectively. In the excitation section, we employ a bottleneck-like construction consisting of two fully connected layers, a batch normalization layer followed by a ReLU activation layer and a Sigmoid activation layer to get the importance of each channel. Then we conduct a channel-wise multiplication to achieve the end-to-end adaptive channel re-calibration.

Tandem with the CEN, SEN lies beside the identity shortcut in the backbone and focuses on the semantic-related parts in features. we use K-global average pooling and replace the fully connected layers with convolution layers. We concatenate the K-global average pooling features and the succedent convolution layer learns the semantic information. With a batch normalization layer and a Sigmoid layer, we could get the spatial importance map, indicating where the network should pay attention. Then after an element-wise multiplication, we can achieve the end-to-end adaptive spatial re-calibration.

3.2. Part Region Proposal Network (P-RPN)

Here we propose a Part Region Proposal Network (P-RPN) to explore the inter-relationship between different scale feature representations containing different parts, without extra convolution computation costs as Fig. 1 shows. we use the pyramid features with four levels (i.e., F_1, F_2, F_3, F_4) as the input feature maps. With the number increasing, the feature transfers from the shallow layer to the deep layer. To extract the inter-relationship between individual parts in shallow layers F_s and deep layers F_d, we embed the ButterFly Unit (BFU) in the P-RPN. Instead of using convolution layers, BFU follows an unsupervised manner.

F_d and F_s are transformed to corresponding multi-level aware features E_d and E_s by up-sampling, down-sampling, and a K-global average pooling along the channel dimension

Table 2. Comparative results on the collected dataset.

Method	Airport	Oil tank	Power plant	mAP
SSD [12]	51.83%	73.27%	62.10%	62.40%
Cascade R-CNN [13]	65.15%	84.82%	73.13%	74.37%
Libra Faster R-CNN [14]	69.37%	83.72%	75.09%	76.06%
CO-Detector (Ours)	**74.89%**	**86.12%**	**79.68%**	**80.23%**

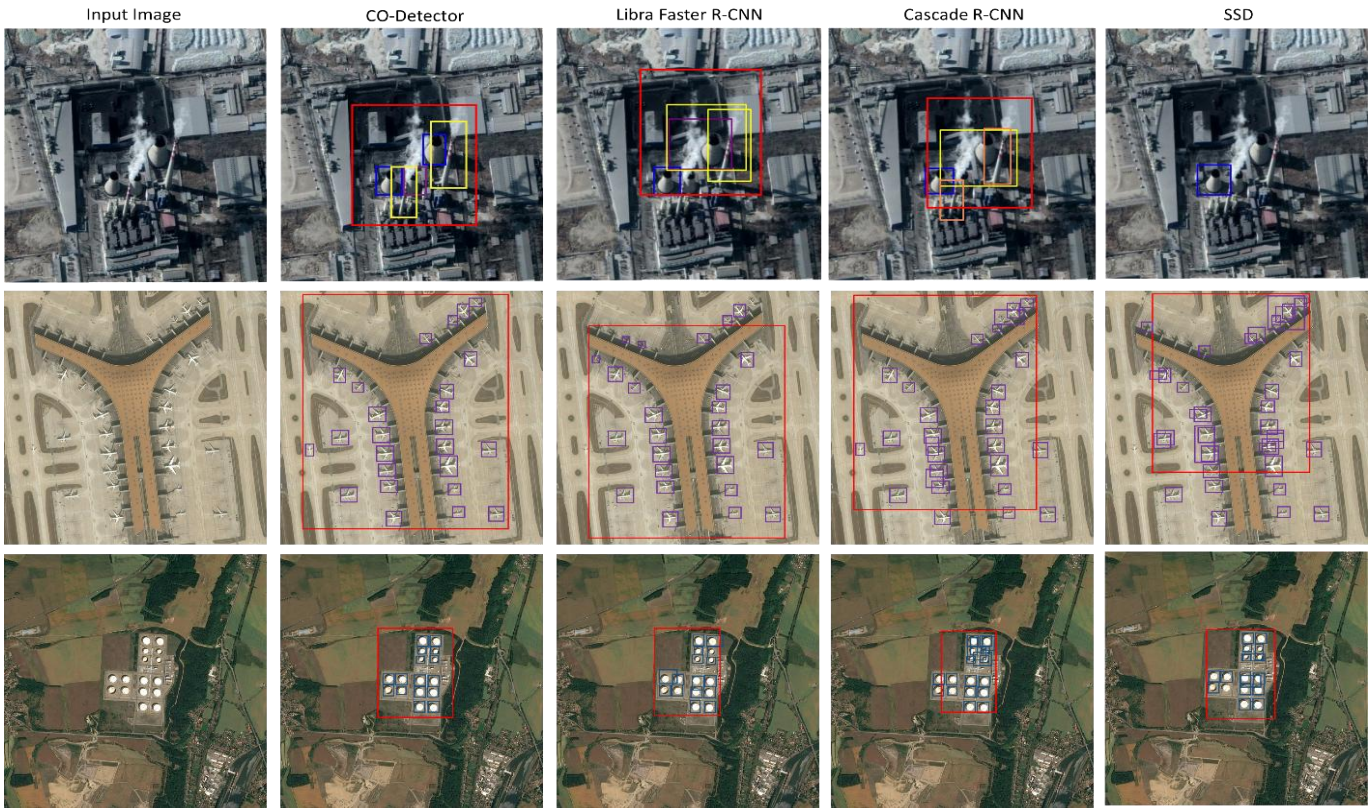

Fig. 2. The visualization results on the collected dataset.

with a Sigmoid activation layer. We note the peak response coordinates in each channel of each multi-regional aware pyramid feature. Then we flatten and concatenate multi-regional aware pyramid features and mark the peak responses. we set anchors to generate part proposals. Then we select the most scored parts with different scales to locate the discriminative parts.

3.3. Part Detection Network (PDN)

We design a Part Detection Network (PDN) to obtain both part and complex classification. PDN takes the part proposals and the global feature as inputs and classifies the parts and the whole object by part and object annotations. By flattening and concatenation of the part proposal features and global image feature together, a more powerful representative feature for whole complex detection is gained.

4. EXPERIMENTAL RESULTS

4.1. Experimental setup

We conduct our experiment on PyTorch deep learning framework and MMDetection framework, with 4 NVIDIA GeForce RTX 2080 Ti GPUs and 30 training epochs. The batch size is set as 4 and the learning rate is 0.005. We use mini-batch stochastic gradient descent (SGD) as the optimizer for classifier training, and set a momentum of 0.9 and a weight decay of 0.0005. Besides, we also use multi-scale training with the long edge set to 2000 and the short edge randomly sampled from [400, 1400].

4.2. Results

We also conduct a comparative study between our MURAN and other state-of-the-art object detection methods, including Cascade R-CNN [12], Libra Faster R-CNN [13] and SSD [14]. We list the comparative results in Table 2. We can see CO-Detector takes the lead in three categories, with 4.17%-17.83% improvements against other deep learning-based object detection methods in respect of mAP. The visualization results are shown in Fig. 2.

5. CONCLUSIONS

In this paper, we propose a CO-Detector, for complex composite object detection, and to learn the inter-relationship between parts in composite objects. CO-Detector consists of three main parts: First, we design a Tandem Attention Network (TAN) to restrain noise disturbance and highlight object features and boundaries. Second, we design a Part Region Proposal Network (P-RPN) to generate part-based proposals and features, locating discriminative object parts. Third, we propose a Part Detection Network (PDN) to classify the individual parts, and detect the whole object through multi-level fused features. We evaluate our model and the comparative experiment results show that CO-Detector reaches the mAP of 80.23% and outperforms 4.17%-17.83% against existing cutting-edge deep learning-based detection methods. In the future, we will step further on transformer architecture in composite object detection to explore the global correlations within complex composite objects.

6. REFERENCES

[1] Ding, Peng, et al. "A light and faster regional convolutional neural network for object detection in optical remote sensing images." ISPRS journal of photogrammetry and remote sensing 141 (2018): 208-218.

[2] Zhang, Tianwen, et al. "Balance learning for ship detection from synthetic aperture radar remote sensing imagery." ISPRS Journal of Photogrammetry and Remote Sensing 182 (2021): 190-207.

[3] Zhang, Lixian, et al. "Making low-resolution satellite images reborn: a deep learning approach for super-resolution building extraction." Remote Sensing 13.15 (2021): 2872.

[4] Zheng, Juepeng, et al. "Growing status observation for oil palm trees using Unmanned Aerial Vehicle (UAV) images." ISPRS Journal of Photogrammetry and Remote Sensing 173 (2021): 95-121.

[5] Zheng, Juepeng, et al. "Surveying coconut trees using high-resolution satellite imagery in remote atolls of the Pacific Ocean." Remote Sensing of Environment 287 (2023): 113485.

[6] Dong, Runmin, et al. "Real-world remote sensing image super-resolution via a practical degradation model and a kernel-aware network." ISPRS Journal of Photogrammetry and Remote Sensing 191 (2022): 155-170.

[7] Cheng, Gong, and Junwei Han. "A survey on object detection in optical remote sensing images." ISPRS journal of photogrammetry and remote sensing 117 (2016): 11-28.

[8] Li, Ke, et al. "Object detection in optical remote sensing images: A survey and a new benchmark." ISPRS journal of photogrammetry and remote sensing 159 (2020): 296-307.

[9] Airbus, “Airbus Oil Storage Detection”, *Kaggle*, 2021. [online]. https://www.kaggle.com/datasets/airbusgeo/airbus-oil-storage-detection-dataset. [2023/5/31].

[10] Zhang, Haopeng, and Qin Deng. "Deep learning based fossil-fuel power plant monitoring in high resolution remote sensing images: A comparative study." Remote Sensing 11.9 (2019): 1117.

[11] Yuan, Shuai, et al. "MUREN: MUltistage Recursive Enhanced Network for Coal-Fired Power Plant Detection." Remote Sensing 15.8 (2023): 2200.

[12] Liu, Wei, et al. "Ssd: Single shot multibox detector." Computer Vision – ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11 – 14, 2016, Proceedings, Part I 14. Springer International Publishing, 2016.

[13] Cai, Zhaowei, and Nuno Vasconcelos. "Cascade r-cnn: Delving into high quality object detection." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[14] Pang, Jiangmiao, et al. "Libra r-cnn: Towards balanced learning for object detection." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.