CHANGE DETECTION NETWORK OF NEARSHORE SHIPS FOR MULTI-TEMPORAL OPTICAL REMOTE SENSING IMAGES

Jingyi Cao, Yanan You, Yuanyong Ning, Wenli Zhou

School of Information and Communication Engineering, Beijing University of Posts and Telecommunications

IGARSS 2020 - 2020 IEEE International Geoscience and Remote Sensing Symposium | 978-1-7281-6374-1/20/$31.00 ©2020 IEEE | DOI: 10.1109/IGARSS39084.2020.9323388

ABSTRACT

Ship change detection is of great significance for maritime safety supervision, wharf vessel management, and vessel life cycle analysis. Nowadays, the ship change information is obtained through the difference between the multiple independent object detection results of multi-temporal images. However, it neglects the temporal correlation of the concerned features, impeding further improvement of the detection accuracy of change detection. Therefore, based on the sequential network, namely convolution LSTM, we built an end-to-end ship change detection (SCD) R-CNN network. The network extracts abstract semantic information reflecting the changed features of ships, and the time correlation between features of the different phases is established. Specifically, the changed features are utilized to guide the judgement of ship change. It is verified from the RS images that the proposed network avoids the misjudgment caused by the errors of object detection in the conventional method. In addition, a higher efficiency is revealed, maintaining the accuracy of the change detection of ship targets.

Index Terms— Change detection, Convolutional neural network, Sequential network, Ship detection

1. INTRODUCTION

Multi-temporal remote sensing (RS) images have been extensively used in desertification surveys, urban sprawl, and other land-based change monitoring tasks. However, little literature focused on its application in marine scenarios, especially ship change detection. The ship change detection task can obtain location and classification results of ship changes in the specified scene, which is indispensable information for berth management, invasion warning, and trajectory analysis.

The development of the deep convolutional neural network (DCNN) improves the analysis efficiency of RS images. Nowadays, the object detection methods based on DCNN reveal enormous potentiality. Faster R-CNN [1], as a relatively mature two-stage detection method, extracts the proposals of targets and performs classification and regression on the proposals. It maintains satisfactory detection accuracy and speed. Therefore, it is available to obtain the changed information through the difference of the independent object detection results.

However, only after several complete object detection processes, a single change result is obtained. Duplicated non-target features are discriminated repeatedly, causing low mission efficiency. Moreover, ignoring the continuity and temporal correlation in the multi-temporal RS images, false detections caused by the accidental factors in a certain image will inevitably affect the final change detection result.

In accordance with the redundancy in object detection, the end-to-end ship change detection (SCD) R-CNN is proposed. It associated the multi-temporal images with the sequential network, realizing feature interaction and extraction of the changed information. Furthermore, under the discriminant of changed features, object detections are conducted in the phases of change.

2. SEQUENTIAL NETWORK

The Long-Term Memory (LSTM) network is inherited from RNN, coping with the long-term dependence problem. In order to establish the correlation of spatial sequence, Convolutional LSTM (ConvLSTM) [2] applied the convolution structure to the traditional LSTM gate control process, effectively extracting the temporal and spatial features of images through convolution kernels with larger receptive field.

The multi-temporal RS images can be regarded as a special sequential data. Therefore, their deep semantic features are attainable to be extracted by the Pseudo-Siamese network [3], and the non-target features can be fused through the ConvLSTM network. After modeling the correlation features, the changed areas are obtained. Theoretically, extracting changed features through the

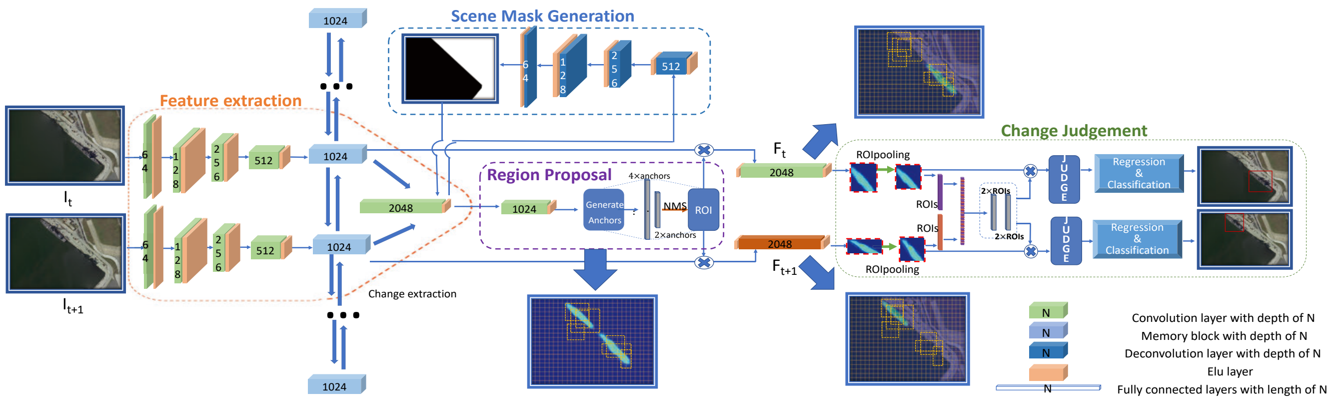

Fig.1 The flowchart of the SCD R-CNN

symmetrical and sequential network can reduce the impact of the redundant information of multi-temporal images on changed ship targets, such as repeating land information. Besides, it will improve the irrelevance of the differential features, and make full use of correlation information between images. Then, in reference to the Region Proposal Network (RPN) in [1], the changed areas are classified and regressed to the proposal regions, delivering changed locations and changed categories.

3. METHODOLOGY

The SCD R-CNN is designed to derive the information of the changed position and the changed types from the multi-temporal RS images, as shown in Fig.1. The modules are described as follows.

3.1. Feature Extraction module

For the consistency of features of each branch, the symmetric Pseudo-Siamese structure is utilized to extraction basic features. Besides, the multi-temporal basic features are associated with memory blocks through the bi-directional convLSTM structure, as shown in Fig. 1. The convLSTM equips with the ability to handle spatial data with sequential relationship, and the bi-directional network [4] balances the features of the previous and next phases, which is adaptive for the area acquisition task of ship change. Finally, the bi-directional outputs are integrated with convolutional layers, acquiring the changed features of ships.

$$
\begin{aligned}
i^{(t)} &= \sigma(W_{xi} * X^{(t)} + W_{hi} * H^{(t-1)} + W_{ci} \odot H^{(t-1)} + b_i) \\
f^{(t)} &= \sigma(W_{xf} * X^{(t)} + W_{hf} * H^{(t-1)} + W_{cf} \odot C^{(t-1)} + b_f) \\
C^{(t)} &= f^{(t)} \odot C^{(t-1)} + i^{(t)} \odot \tanh(W_{xc} * X^{(t)} + W_{hc} * H^{(t-1)} + b_c) \quad (1) \\
o^{(t)} &= \sigma(W_{xo} * X^{(t)} + W_{ho} * H^{(t-1)} + W_{co} \odot C^{(t-1)} + b_o) \\
H^{(t)} &= o^{(t)} \odot \tanh(C^{(t)})
\end{aligned}
$$

The gate control process of convLSTM is shown in Equ. 1. The input gate combines the input $X^{(t)}$ with the state of the previous memory cell $C^{(t-1)}$ to enhance the feature intensity of the changed area. The forgetting gate adjusts the weighted parameters W_{xf} through the state of the previous cell $C^{(t-1)}$. It reduces the response of features with less variation, such as land, by determining the discarded information of the cells. The cell state $C^{(t)}$ is determined by the output of the sum of $f^{(t)}$ and $i^{(t)}$ of the current phase. Then, the output $o^{(t)}$ of rich feature information in the changed target area is obtained.

3.2. Scene Mask Generation module

The scene mask restricts the changed information only detected in the target area (sea), reducing the feature response in the non-target area [5]. Through the deconvolution, the module generates the scene mask with probability distribution between 0 and 1, indicating the possibility of land or sea. Then the updated changed information is obtained by multiplying the generated scene mask with the original ship change feature map.

3.3. Region Proposal module

The slide window method is used to generate multi-scale anchors on the updated changed feature map [1]. The changed probability of each generated anchor is obtained through the fully connected layers. Finally, the regions of interest (ROIs) are obtained through non-maximum suppression (NMS) of the high scores' anchors, representing areas of severe change.

3.4. Change Judgement module

This module is used to realize the quadratic judgment of the changed ROIs and deduce the time phase of change. After obtaining the changed proposal regions, the features on the memory block of each branch are normalized by the ROI pooling. Then, they are analyzed for similarity through the fully connected layers.

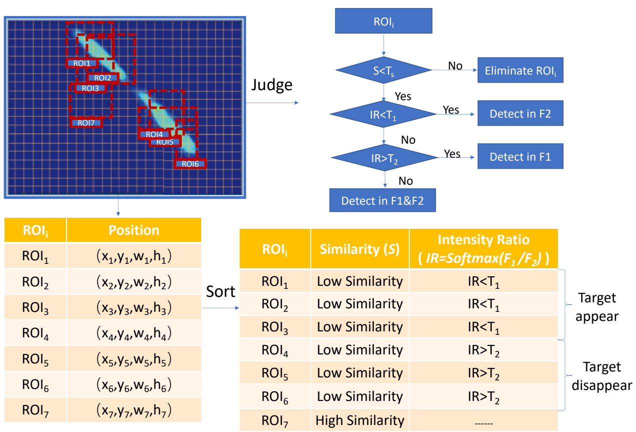

Fig. 2 Algorithm process of change judgement module

Finally, two sets of 1-dimensional tensor with the size of 2*ROIs are generated. The first tensor represents the similarity (S) of the corresponding ROIs, and the second refers to the intensity ratio (IR) of ROIs' features. As shown in Fig.2, the similarity information is sorted by the threshold value Ts, excluding the ROIs without feature change. As for the ROI regions of $S > Ts$, IR is used for analysis. If IR is less than the threshold T_1, object detection is performed on F_2, the target appears. Instead, detection is performed on F_1, when IR exceeds T_2, representing the disappearance of the target. Object detections are simultaneously performed in both ROIs when T_1<IR<T_2, representing target replacement.

3.5. Training

The acquisition of ROIs and scene masks are referenced in [1] and [5], respectively. In the change judgement module, Euclidean distance D of features [6] is the measure index for similarity. The contrastive loss is used to optimize model through augmenting the difference between various features and reducing the difference between the similar features, as Equ. 2,

$$L_{sim}(R_i^{(t)}, R_i^{(t+1)}, y_i) = \frac{1}{2}(1-y_i)D^2 + \frac{1}{2}y_i(max(0,\delta - D))^2 \quad (2)$$

where, y_i represents the label of the ROI_i pairs, $y_i = 1$ means change happened, and $y_i = 0$ means there is no change. D^2 is the penalty terms of changed state, and the punishment $max(0,\delta - D)$ is for the unchanged.

Cross entropy loss is to calculate IR loss, as Equ. 3,

$$L_{IR} = \sum_k \log[IR_k IR_k^* + (1-IR_i)(1-IR_i^*)] \quad (3)$$

IR_k^* is intensity label of ROIs, and IR_k is the predicted intensity. If the changed target is only included in the previous phase, $IR^* = 1$; If target is only included in the next phase, the ship appears, $IR^* = 0$.

4. EXPERIMENTS

Due to the wide shooting interval of the open-source RS images, it is difficult to obtain plenty of training data with annotations. Therefore, it is feasible to fuse the concerned targets in the existed object detection dataset with the specific scene by the image fusion technology to construct a dataset, as shown in Fig.3.

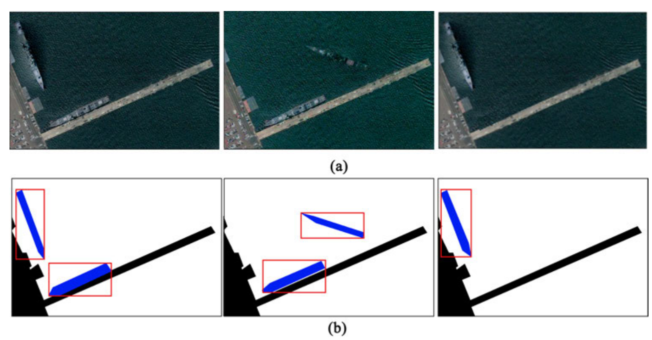

Fig. 3 Samples of dataset for change detection. (a) Fused images; (b) Annotations of target masks and locations.

In order to evaluate the performance, the change detection results of SCD R-CNN are compared with the difference results of the Faster R-CNN. In the difference judgement, the corresponding targets are judged no change when IOU exceeds the threshold of 0.7.

Table I Comparison on change detection performance

	SCD R-CNN	Faster R-CNN
Detection Accuracy	93.37 %	81.23 %
Time Consumption	289 ms	384 ms

From the perspective of detection accuracy, the difference method based on the Faster R-CNN neglects the continuity of the images, processing the non-target information repeatedly. However, as shown in Table I and Fig.4, the SCD R-CNN demonstrates superior accuracy, the impact of the misjudgments caused by false alarms is weakened.

It also can be concluded that, for the Faster R-CNN, the change detection is carried out only in the condition of two complete detection tasks. However, the SCD R-CNN only carries out detection on the generated change feature map, saving the time of twice extraction and classification of ROIs, which results in better performance in time consumption.

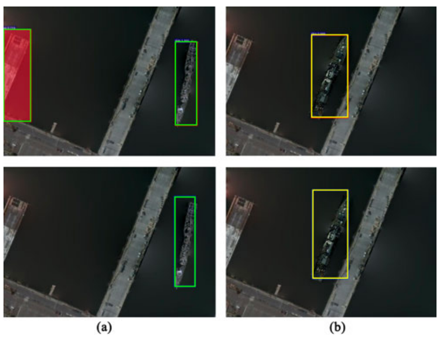

Fig. 4 Samples of the change detection results. The first row comes from the Faster R-CNN, and the second row comes from the SCD R-CNN. (a) The previous phase, (b) the next phase. Thereinto, the green boxes represent the results of the disappearances of ships, and the yellow boxes represent the occurrences of ships. Besides, the boxes covered in red are error detections.

5. CONCLUSION

For extracting changed semantic information on the ship targets of multi-temporal RS images, an end-to-end ship change detection network is proposed in this paper. This network takes multi-temporal images as sequential data, obtaining the basic features through the Pseudo-Siamese network and analyzing change semantic with bi-directional convLSTM. Compared with the method, which calculates the difference between the object detection results, the correlation and dependency on sequential images is considered over the generation of the changed features in the SCD R-CNN. The experiment shows that the SCD R-CNN has better performance at efficiency and detection accuracy, which is capable of the complex offshore scene.

6. REFERENCES

[1] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” *ArXiv150601497 Cs*, Jan. 2016, Accessed: Dec. 08, 2019.

[2] “Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting.” https://arxiv.org/abs/1506.04214.

[3] H. Dong, W. Ma, Y. Wu, M. Gong, and L. Jiao, “Local Descriptor Learning for Change Detection in Synthetic Aperture Radar Images via Convolutional Neural Networks,” *IEEE Access*, vol. 7, pp. 15389–15403, Feb. 2019.

[4] X. Ma and E. Hovy, “End-to-end Sequence Labeling via Bi-directional LSTM-CNNs-CRF,” *ArXiv160301354 Cs Stat*, May 2016.

[5] Y. You, J. Cao, Y. Zhang, F. Liu, and W. Zhou, “Nearshore Ship Detection on High-Resolution Remote Sensing Image via Scene-Mask R-CNN,” *IEEE Access*, vol. 7, pp. 128431–128444, 2019.

[6] Z. Xu, H. Zhang, Y. Wang, X. Chang, and Y. Liang, “L 1/2 regularization,” *Sci. China Inf. Sci.*, vol. 53, no. 6, pp. 1159–1169, Jun. 2010.