HAZY REMOTE SENSING IMAGES SEMANTIC SEGMENTATION FOR WEAKLY ANNOTATION BASED ON SALIENCY-AWARE ALIGNMENT STRATEGY

Junda Xu and Libao Zhang

School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China

INTRODUCTION

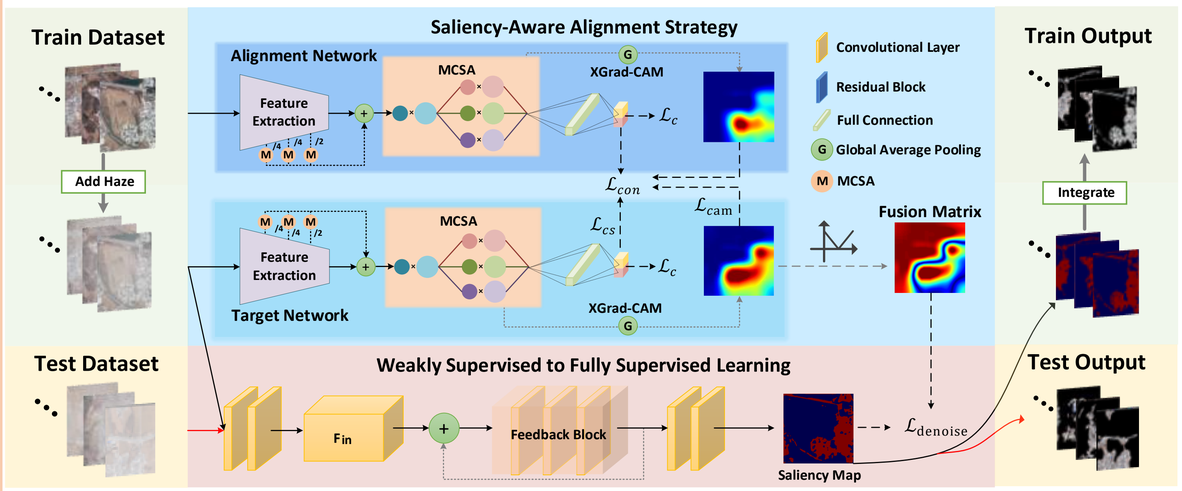

Fig. 1 The framework of the proposed method.

- We propose a new weakly supervised semantic segmentation framework for hazy RSIs under the condition of image-level annotation, comprising alignment network (AN) and target network (TN);
- We propose a network calibration scheme based on a novel saliency-aware alignment strategy, which enables TN to learn the deep feature parameters of AN by incorporating a consistency loss term;
- We design a multi-scale channel-spatial attention (MCSA) module that leverages both spatial and channel attention mechanisms to fuse multi-scale information effectively.

METHODOLOGY

A. Generation of pseudo labels based on Saliency-aware Alignment Strategy

The framework of this method is composed of two classification networks, namely AN and the TN, which share the same structure. The AN is responsible for learning and storing the features extracted from clear images, while the TN focuses on learning features from hazy images, aligning them with those obtained by the AN. The alignment is achieved through the utilization of a consistency loss, wherein we compute the L2 norm on both the category scores and class activation maps of the two networks to quantify the level of alignment between them.

$$\mathcal{L}_{con} = \alpha \mathcal{L}_{cs} + \mathcal{L}_{cam}$$

Subsequently, based on these aligned features, the TN generates pseudo-labels indicating the class in each hazy image.

B. Multi-scale Channel-Spatial Attention Block

We designed a MCSA module to enable the extraction of more comprehensive and diverse features, which consists of a channel attention layer and a multi-scale spatial attention layer in sequence, as shown in Fig. 2. By combining channel and spatial attention, we achieve a more precise feature enhancement. It is worth noting that multi-scale spatial attention is achieved through dilated convolution.

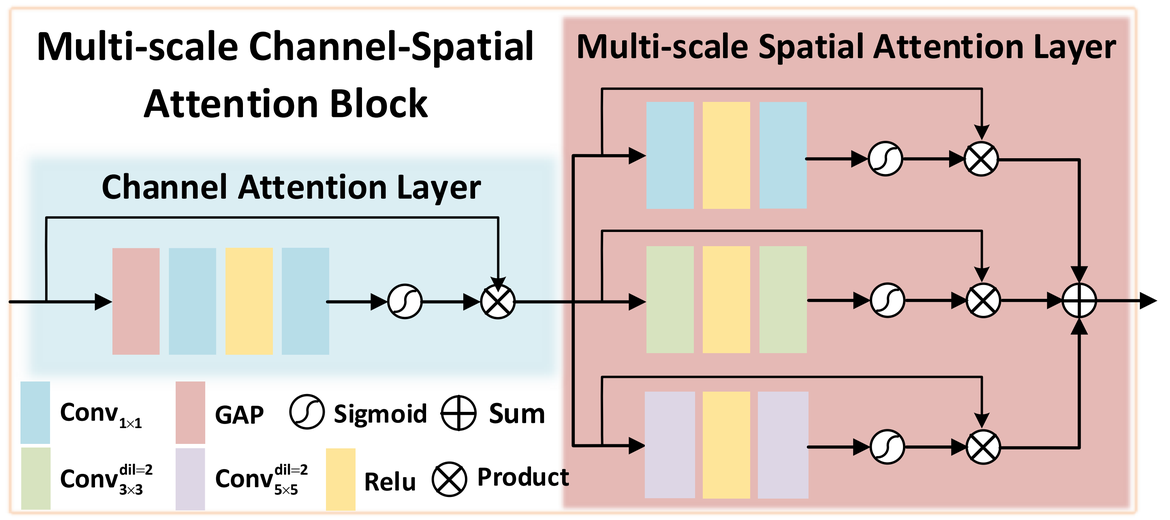

Fig. 2 The architecture of the MCSA.

For a given feature F, it is enhanced and fused by the attention module to get the corresponding channel attention map F_{CA} and spatial attention map F_{SA}, and produce the final feature map F_{MCSA}.

$$F_{CA} = F \otimes A_C, F_{SA}^i = F_{CA} \otimes A_S^i, F_{MCSA} = \sum_i F_{SA}^i$$

RESULTS AND DISCUSSIONS

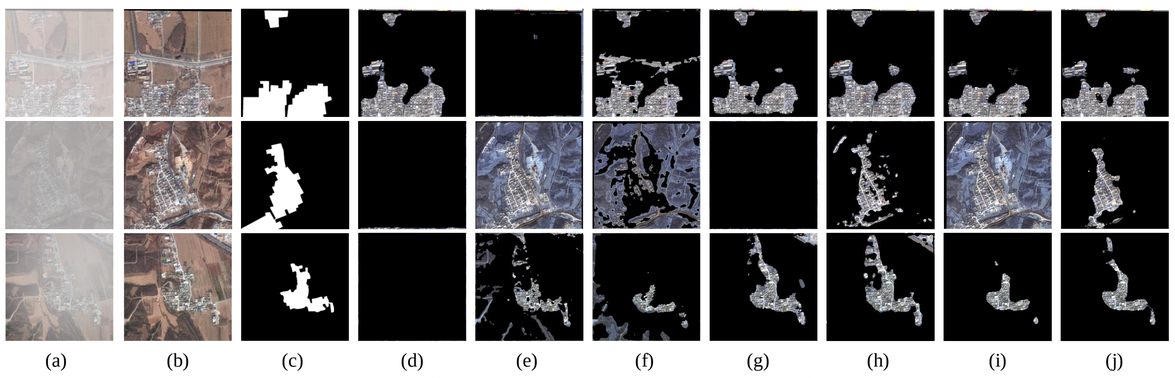

Fig. 3 Visual comparison of segmentation results. (a) Original images. (b) Clear images. (c) Ground truth. (d) PSL. (e) DCP-PSL. (f) AOD-PSL. (g) FFA-PSL. (h) w/o MCSA. (i) w/o CAL. (j) Ours.

Visual comparison shows that our method is superior to the existing methods not only in the recognition of the background but also in the integrity of the object segmentation, especially in the severe haze condition.

Fig. 4 AUC value of the SPOT5 dataset.

Methods	AUC	P	R	F
w/o MCSA	0.865	0.985	0.880	0.928
w/o CAL	0.842	0.977	0.941	**0.958**
Ours	**0.890**	**0.993**	**0.964**	0.922

Table I. AUC, Precision, Recall, F-measure values of ablation experiments.

Quantitative analysis. Fig. 4 shows the excellent performance of our approach under different haze conditions, but slightly worse than FFA-Net under dense conditions on the SPOT5 dataset. The presence of a significant contrast between the foreground and background in the SPOT5 dataset may contribute to this issue, while the task of dehazing does not pose a high level of difficulty.

Ablation experiments. From Fig. 3 and Table I, we designed ablation experiments for two major modules of the method proposed in this paper. The strong effectiveness of both modules in our design is evident.