Knowledge Distillation via Cross Supervising with Attention for Remote Sensing Object Detection

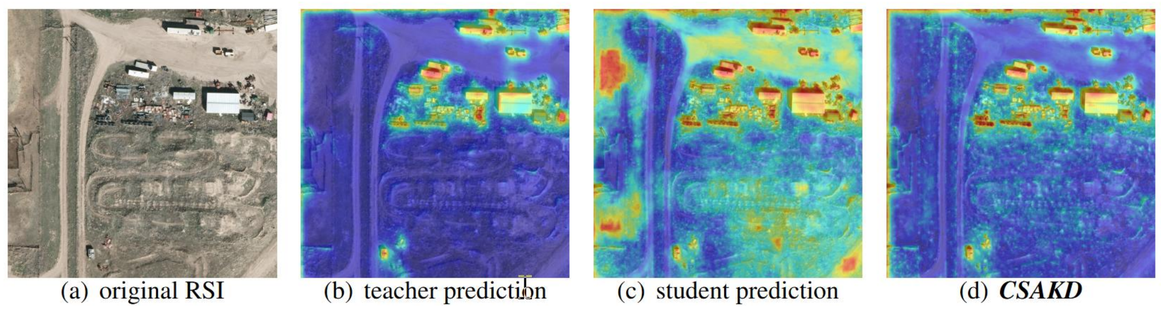
Kefan Zhan An Luo Yunpeng Zeng
Jiaxin Li Yuan Zhang * Kai Hu *

School of Computer Science, Xiangtan University, Hunan, P. R. China

Introduction

Problem

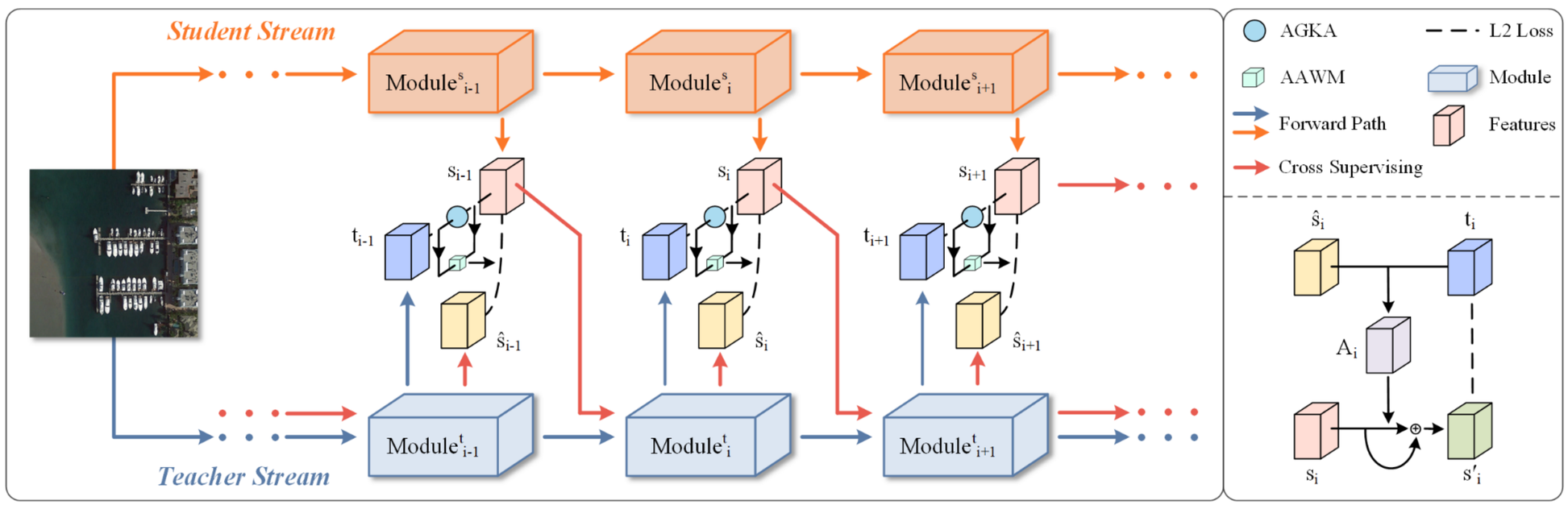
Remote sensing image possesses distinct characteristics such as small objects and complex background, making it extremely difficult for the student model to understand and assimilate the intermediate features of the teacher model. Hence, the effectiveness of feature-based KD methods in RSOD is significantly limited.



(a) original RSI (b) teacher prediction (c) student prediction (d) *CSAKD*

Figure 1: Visualizations of gradient heatmaps from the YOLOv8s. (a) is an original RSI. (b) and (c) are outputs predicted by the teacher and the student, which is trained with OST. (d) is predicted by the student model trained with the proposed *CSAKD*.

Methods



(a) Overview of *CSAKD* (b) AGKA

Figure 2: (a) Overview of the proposed Knowledge Distillation via Cross Supervising with Attention (*CSAKD*). Through Cross Supervising, the student will acquire module-level mentoring, which eliminates the performance discrepancy between the teacher and the student. The Adaptive Attention Weight Modulation (AAWM) is introduced to regulate the intensity of Cross Supervising. (b) Attention-Guided Knowledge Alignment (AGKA), which involves the cross features obtained in Cross Supervising, is capable of enabling the student to focus on the area that the teacher deems crucial.

Total loss

$$\mathcal{L}_{total} = \mathcal{L}_{Det} + \alpha \cdot \sum_{i=0}^{N} \mathcal{L}_{AGKA_i} + \beta \cdot \sum_{i=0}^{N} \mathcal{L}_{Cross_i}$$

Cross Supervising

$$t_i = Module^t_i(t_{i-1}), s_i = Module^s_i(s_{i-1}), \hat{s}_i = \begin{cases} s_0 & i = 0 \\ Module^t_i(s_{i-1}) & i \neq 0 \end{cases}$$

AAWM

$$T_i = GAP(t_i), S_i = GAP(s_i), a_i = \frac{T_i \cdot S_i^T}{\sqrt{d}}$$

$$weight_i = 1 - Sigmoid(a_i), \mathcal{L}_{Cross_i} = weight_i \cdot \|\hat{s}_i - s_i\|_2$$

AGKA

$$q_i = ReLU(t_i), k_i = ReLU(\hat{s}_i), A_i = \frac{Softmax(q_i \cdot k_i^T)}{\sqrt{d_{k_i}}}$$

$$s'_i = s_i + A_i \cdot s_i, \mathcal{L}_{AGKA_i} = \left\|t_i - s'_i\right\|_2$$

Contributions

- We propose a novel self-distillation framework, Cross Supervising with Attention Knowledge Distillation (*CSAKD*). Specifically designed for RSOD, this framework offers module-level mentoring, denoted as Cross Supervising, allowing the student model to acquire knowledge more effectively from the teacher model. In addition, a new Adaptive Attention Weight Modulation (AAWM) is proposed to dynamically adjust the intensity of Cross Supervising.
- We introduce an Attention-Guided Knowledge Alignment (AGKA), which enables the student to concentrate on the area that the teacher deems crucial and have better comprehension of the feature knowledge instead of blindly learning from the intermediate features of the teacher.

* Corresponding authors

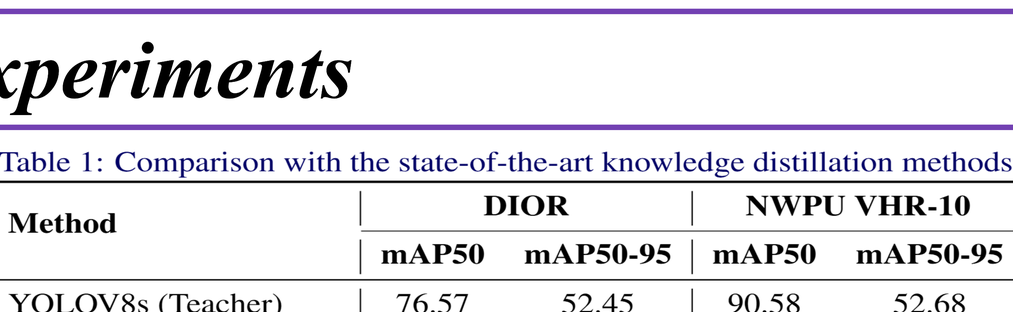
Experiments

Table 1: Comparison with the state-of-the-art knowledge distillation methods

Method	DIOR		NWPU VHR-10	
	mAP50	mAP50-95	mAP50	mAP50-95
YOLOV8s (Teacher)	76.57	52.45	90.58	52.68
YOLOv8s+HQ (Student)	73.84	51.44	88.05	51.53
AFD	75.72	52.30	89.41	52.97
ReviewKD	75.27	51.59	91.00	52.92
OST	75.92	**52.65**	90.25	52.74
CrossKD	75.79	52.42	90.17	53.30
***CSAKD* (Ours)**	**76.27**	52.56	**91.15**	**53.69**

Table 2: Ablation study of key components in our *CSAKD* framework

Key Components		DIOR		NWPU VHR-10	
Cross Supervising	AGKA	mAP50	mAP50-95	mAP50	mAP50-95
×	×	74.21	50.52	87.91	51.02
✓	×	74.28	50.49	88.50	51.08
×	✓	76.14	52.54	88.74	52.26
✓	✓	**76.27**	**52.56**	**91.15**	**53.69**



(a)

(b)

Figure 3: Visualization of detection results on two datasets. (a) DIOR. (b) NWPU VHR-10.