DOUBLE REVERSE REGULARIZATION NETWORK BASED ON SELF-KNOWLEDGE DISTILLATION FOR SAR OBJECT CLASSIFICATION

Bo Xu*, Hao Zheng*, Zhigang Hu, Liu Yang§, Meiguang Zheng, Xianting Feng, Wei Lin

School of Computer Science and Engineering, Central South University, China

{xubocsu, zhenghao, zghu, yangliu, zhengmeiguang, xiantingfeng, linwei}@csu.edu.cn

Abstract

In current synthetic aperture radar (SAR) object classification, one of the major challenges is the severe overfitting issue due to the limited dataset (few-shot) and noisy data. Considering the advantages of knowledge distillation as a learned label smoothing regularization, this paper proposes a novel Double Reverse Regularization Network based on Self-Knowledge Distillation (DRRNet-SKD). Specifically, through exploring the effect of distillation weight on the process of distillation, we are inspired to adopt the double reverse thought to implement an effective regularization network by combining offline and online distillation in a complementary way. Then, the Adaptive Weight Assignment (AWA) module is designed to adaptively assign two reverse-changing weights based on the network performance, allowing the student network to better benefit from both teachers.

Motivation

Results of DLB and Tf-KD on OpenSARShip:

BackBone	Method	KD Weight	**Accuracy(%)**
VGG-11	Baseline	Fixed : 0	73.07±2.38
	DLB	Fixed: 0.3	76.69±1.56(+3.62)
		Fixed: 0.7	77.82±1.13(+4.75)
		Unfixed:α_t	**78.14±0.37(+5.07)**
	Tf-KD	Fixed: 0.3	74.24±1.19(+1.17)
		Fixed: 0.7	74.67±1.72(+1.60)
		Unfixed:α'_t	**75.15±0.24(+2.08)**

$$\alpha_t = \frac{t}{T}, \alpha'_t = 1 - \frac{t}{T} \tag{1}$$

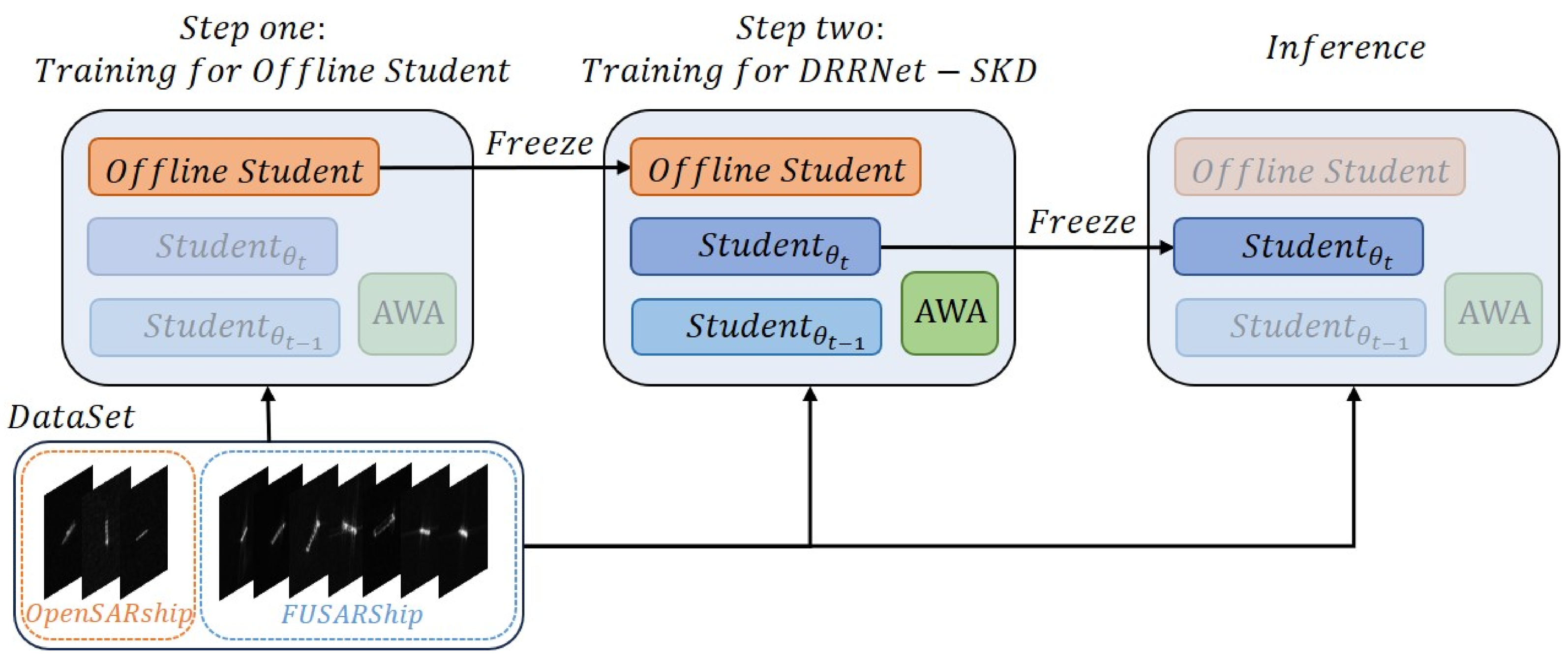
where T is the total number of epoch. We think that the weight of offline distillation should decrease and the weight of self-distillation should increase as student performance grows.

Acknowledgements

This work is supported by the Natural Science Foundation of China (Grant No: 62172442).

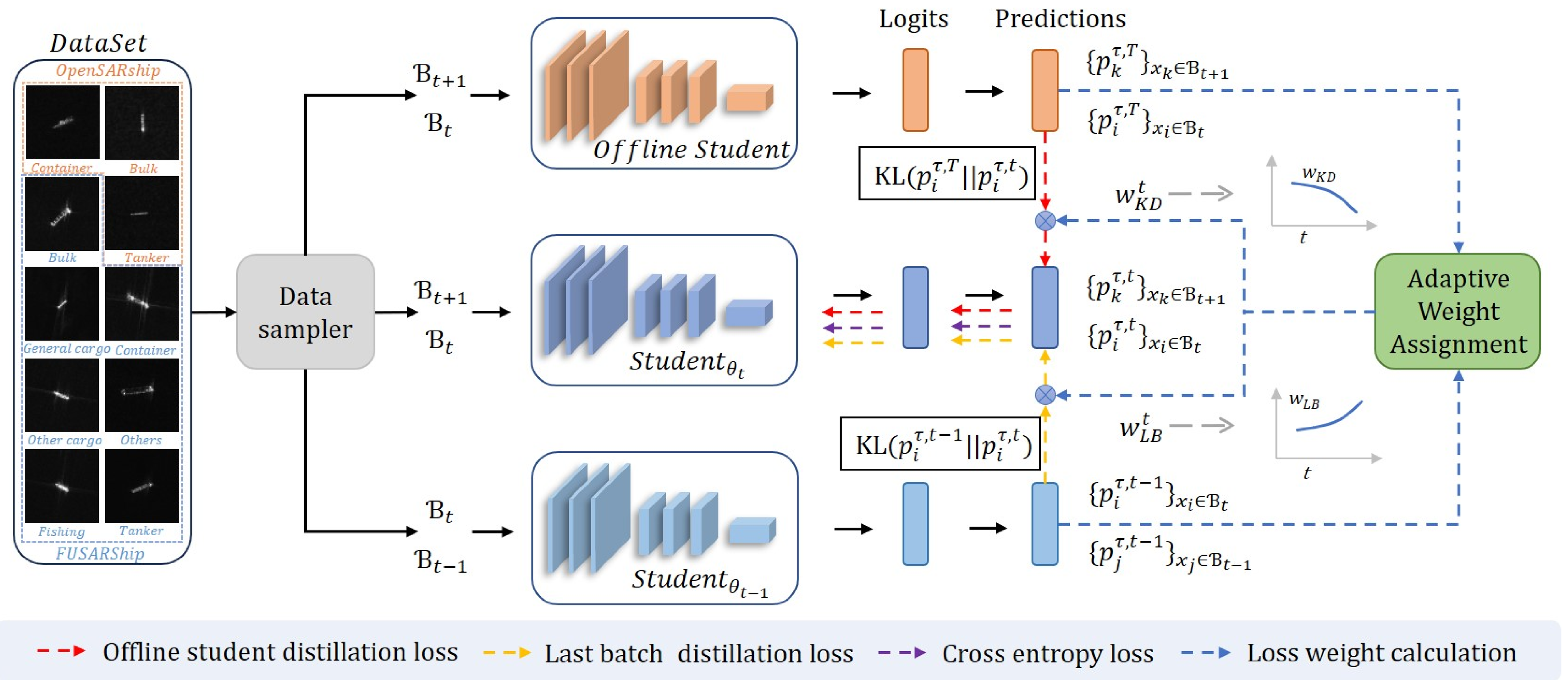
Introduction

Our DRRNet-SKD has a two-stage training procedure. The first stage is training the pre-trained model via DLB, whose parameters will be frozen as offline student in the next step. The second phase is training a new student model from scratch through DRRNet-SKD. Then the well-trained student model can perform inference independently.



Methodology

The $\mathcal{B}_t$, θ_t, $p^{\tau,t}$ represent a mini-batch of data samples, trainable parameters and the soften outputs indexed in the t^{th} iteration. The offline student and last batch student impose double reverse regularization on the current batch student training with the assistance of adaptive weight assignment.



$$\mathcal{L}_{ON} = -\sum_{c=1}^{C} y^c \log\left(\varphi\left(\boldsymbol{p}_{\theta_{t-1}}/\alpha_\tau\right)\right), \quad \mathcal{L}_{OF} = -\sum_{c=1}^{C} y^c \log\left(\varphi\left(\boldsymbol{p}^T/\alpha_\tau\right)\right). \tag{2}$$

$$w_{LB}^t = exp(\mathcal{L}_{OF} - \mathcal{L}_{ON}), \quad w_{KD}^t = \alpha - w_{LB}^t, \quad \alpha \geq w_{LB}^t. \tag{3}$$

Eventually, the overall loss function of our DRRNet-SKD is summarized as:

$$\mathcal{L} = \mathcal{L}_{CE} + w_{LB} \cdot \mathcal{L}_{LB} + w_{KD} \cdot \mathcal{L}_{KD}. \tag{4}$$

Results

Dataset	Method	AlexNet	VGG-11	VGG-16	ResNet-18	DenseNet-121
OpenSARShip	Baseline	68.88±2.05	73.07±2.38	72.34±2.51	72.15±1.25	75.69±1.60
	LSR	70.19±2.79(+1.31)	76.94±0.86(+3.87)	74.34±1.61(+2.00)	73.20±1.29(+1.05)	78.03±0.72(+2.34)
	Tf-KD	69.35±1.22(+0.47)	74.67±1.72(+1.60)	75.04±1.73(+2.70)	72.75±0.97(+0.60)	75.15±1.99(-0.54)
	DLB	70.15±1.47(+1.27)	77.82±1.13(+4.75)	74.63±0.63(+2.29)	74.46±1.08(+2.31)	76.62±1.26(+0.93)
	Ours	**70.67±0.49(+1.79)**	**80.03±0.87(+6.96)**	**78.15±1.72(+5.81)**	**75.10±0.91(+2.95)**	**78.48±0.72(+2.79)**
	Method	AlexNet	VGG-16	VGG-19	ResNet-18	ResNet-50
FUSAR-Ship	Baseline	79.36±0.34	83.51±0.77	82.51±0.70	79.48±0.39	80.29±0.46
	LSR	80.31±0.36(+0.95)	85.41±0.70(+1.90)	84.34±0.40(+1.83)	80.67±0.30(+1.19)	81.40±0.46(+1.11)
	Tf-KD	79.60±0.76(+0.24)	84.35±0.80(+0.84)	83.24±0.88(+0.73)	80.88±0.47(+1.40)	81.91±0.35(+1.62)
	DLB	80.24±0.58(+0.88)	84.53±0.36(+1.02)	84.79±0.35(+2.28)	80.71±0.41(+1.23)	81.17±0.41(+0.88)
	Ours	**81.02±0.49(+1.66)**	**86.07±0.54(+2.56)**	**86.97±0.40(+4.46)**	**81.58±0.44(+2.10)**	**82.24±0.33(+1.95)**