1.1 Title

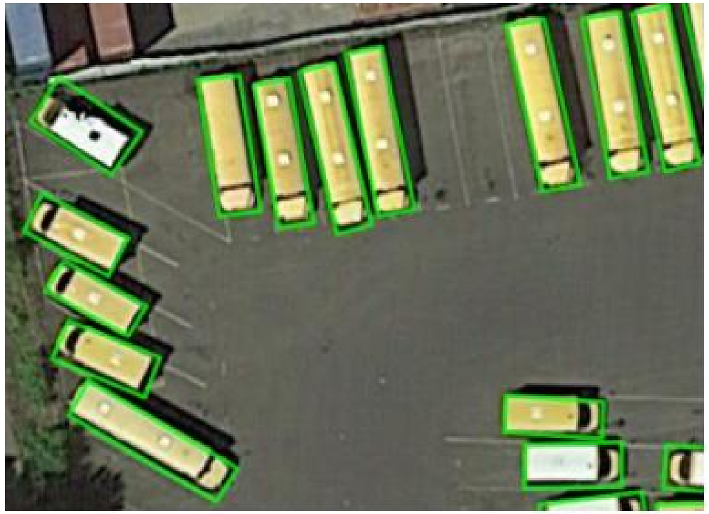
SOOD: Towards Semi-Supervised Oriented Object Detection

Wei Hua[*1], Dingkang Liang[*1], Jingyu Li[1], Xiaolong Liu[1], Zhikang Zou[2], Xiaoqing Ye[2], Xiang Bai[†1]

[1]Huazhong University of Science and Technology, {whua_hust, dkliang, xbai}@hust.edu.cn
[2] Baidu Inc., China

1.2 Background

（1） 研究场景：俯瞰图；
物体小、密集、边界框存在方向性



(a) Arbitrary rotating objects

(b) Small and dense objects

1.2 Background

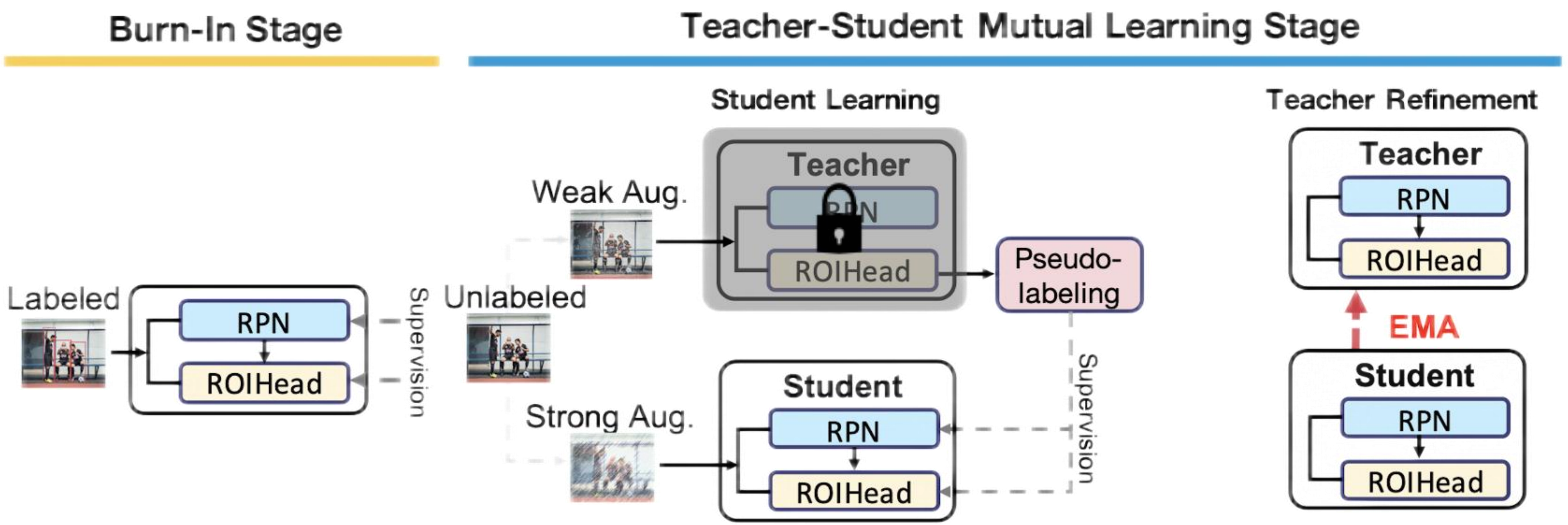
（2）Oriented Object Detection

与传统目标检测器不同，方向目标检测器使用带有方向性的边界框进行物体预测。

Oriented

1.2 Background

（3）半监督目标检测框架：教师-学生模型



$$\theta_t \leftarrow \alpha\theta_t + (1-\alpha)\theta_s$$

1.2 Background

（4）Optimal transport 最优传输

当给定**两个沙盘**时（每盘沙子可以**代表一个概率分布**），可以通过很多方式将一个沙盘传输到另一个沙盘。

基于**传输单个沙粒的局部花费**，每一种传输方法均对应一个全局花费。最优传输的目的就是**寻找总体花费最少的传输方案。**

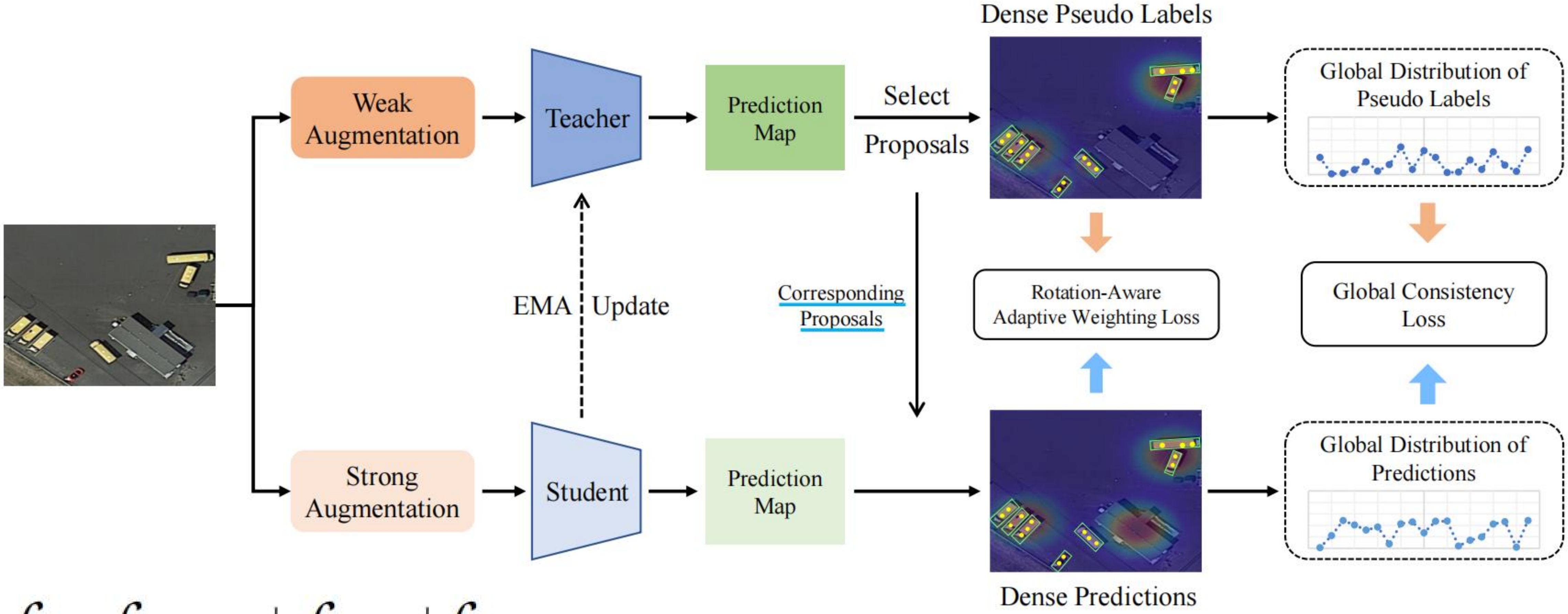
a b c

1.3 Contribution

· Proposes the **first** Semi-supervised Oriented Object Detection method

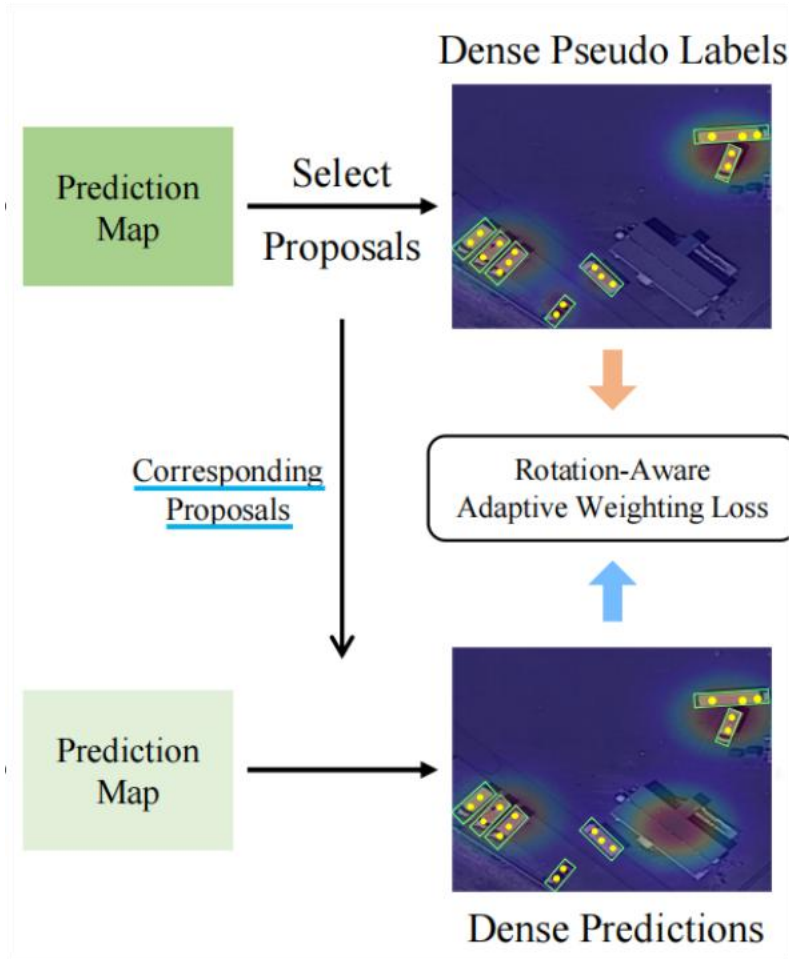
· Two simple yet effective **losses that enforce the instance-level and set-level consistency** between the students and the teacher’s predictions.

2.1 Pipeline



$$\mathcal{L} = \underbrace{\mathcal{L}_{RAW} + \mathcal{L}_{GC}}_{\mathcal{L}_u} + \mathcal{L}_s.$$

2.2 Rotation-aware Adaptive Weighting Loss



The orientation difference can be used to dynamically adjust the unsupervised loss.

$$\mathcal{L}_{RAW} = \sum_{i}^{N_p} \omega_i^{rot} \mathcal{L}_{u}^{i}$$

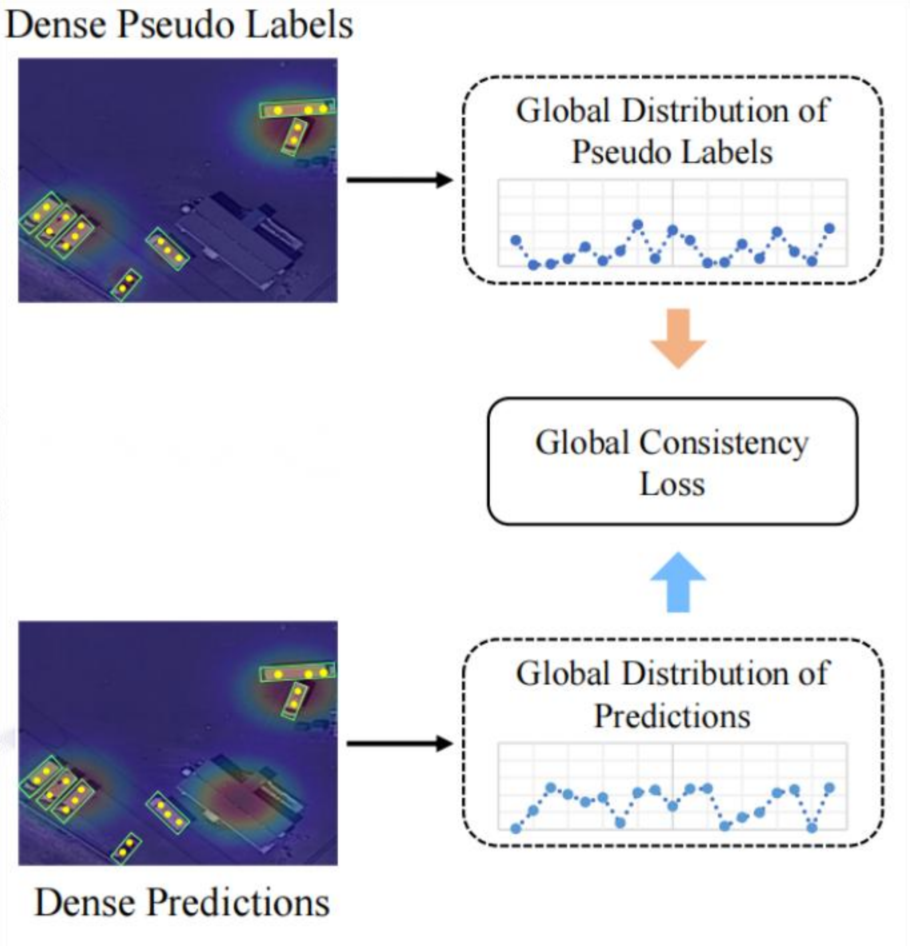
Np: pseudo label number

$$\omega_i^{rot} = 1 + \sigma_i,$$

α：show the importance of adjusting orientation.

$$\sigma_i = \alpha \frac{|r_i^t - r_i^s|}{\pi}, r_i^t, r_i^s \in [-\frac{\pi}{2}, \frac{\pi}{2})$$

2.2 Global Consistency Loss



使用最优传输方法

（1）概率分布：

$$\mathbf{d}_i^t = e^{\mathbf{s}_{i,c(i)}^t}$$

$$\mathbf{d}_i^s = e^{\mathbf{s}_{i,c(i)}^s}$$

s:classification scores

c(i):index of category

$$C_{i,j} = C_{i,j}^{dist} + C_{i,j}^{score},$$

（2）局部花费：

$$C_{i,j}^{dist} = \frac{\|\mathbf{z}_i^t - \mathbf{z}_j^s\|_2^2}{\max_{1<=a,b<=N_p} \|\mathbf{z}_a^t - \mathbf{z}_b^s\|_2^2},$$

$$C_{i,j}^{score} = \frac{\|\mathbf{s}_{i,c(i)}^t - \mathbf{s}_{j,c(j)}^s\|_1}{\max_{1<=a,b<=N_p} \|\mathbf{s}_{a,c(a)}^t - \mathbf{s}_{b,c(b)}^s\|_1},$$

z :2D coordinates of the sample

（3）全局花费：

$$\mathcal{L}_{GC}(\mathbf{d}^t, \mathbf{d}^s) = \left\langle \boldsymbol{\lambda}^*, \frac{\mathbf{d}^t}{\|\mathbf{d}^t\|_1} \right\rangle + \left\langle \boldsymbol{\mu}^*, \frac{\mathbf{d}^s}{\|\mathbf{d}^s\|_1} \right\rangle$$

3.1 Experiment

1. 数据集：DOTA-v1.5, which contains 2806 large aerial images and 402,089 annotated oriented objects。
2. detector： FCOS
3. backbone：ResNet-50 with FPN

Setting	Method	Publication	10%	20%	30%
Supervised	Faster R-CNN* [31]	NeurIPS 2016	43.43	51.32	53.14
	FCOS† [38]	ICCV 2019	42.78	50.11	54.79
Semi-supervised	Unbiased Teacher* [25]	ICLR 2021	44.51	52.80	53.33
	Soft Teacher* [45]	ICCV 2021	48.46	54.89	57.83
	Dense Teacher† [51]	ECCV 2022	46.90	53.93	57.86
	SOOD† (**ours**)	-	**48.63**	**55.58**	**59.23**

Method	Publication	mAP
Unbiased Teacher* [25]	ICLR 2021	$66.12 \xrightarrow{-1.27} 64.85$
Soft Teacher* [45]	ICCV 2021	$66.12 \xrightarrow{+0.28} 66.40$
Dense Teacher† [51]	ECCV 2022	$65.46 \xrightarrow{+0.92} 66.38$
SOOD† (**ours**)	-	$65.46 \xrightarrow{+2.24} \mathbf{67.70}$

3.1 Experiment

Setting	RAW	GC	mAP		
			10%	20%	30%
I	-	-	47.24	54.07	57.74
II	✓	-	47.82	55.21	58.93
III	-	✓	47.71	54.72	58.70
IV	✓	✓	**48.63**	**55.58**	**59.23**

(1)

Setting	Distance	Score	mAP
I	-	-	47.82
II	-	✓	48.10
III	✓	-	47.94
IV	✓	✓	**48.63**

(2)

Setting	α	mAP
I	1	47.77
II	10	47.87
III	50	**48.63**
IV	100	47.95

(3)

$$C_{i,j} = C_{i,j}^{dist} + C_{i,j}^{score},$$

$$C_{i,j}^{dist} = \frac{\|\mathbf{z}_i^t - \mathbf{z}_j^s\|_2^2}{\max_{1<=a,b<=N_p} \|\mathbf{z}_a^t - \mathbf{z}_b^s\|_2^2},$$

$$C_{i,j}^{score} = \frac{\|\mathbf{s}_{i,c(i)}^t - \mathbf{s}_{j,c(j)}^s\|_1}{\max_{1<=a,b<=N_p} \|\mathbf{s}_{a,c(a)}^t - \mathbf{s}_{b,c(b)}^s\|_1}.$$

$$\sigma_i = \alpha \frac{|r_i^t - r_i^s|}{\pi}, r_i^t, r_i^s \in [-\frac{\pi}{2}, \frac{\pi}{2})$$

3.1 Experiment

真实框 supervised Dense Teacher SOOD