BoxTeacher: Exploring High-Quality Pseudo Labels for Weakly Supervised Instance Segmentation

Tianheng Cheng[1], Xinggang Wang[1], Shaoyu Chen[1], Qian Zhang[2], Wenyu Liu[1]

[1]Huazhong University of Science and Technology, [2]Horizon Robotics

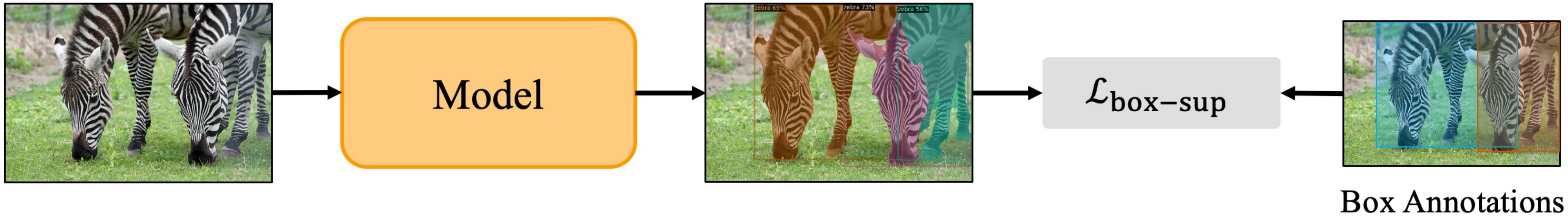
Poster Tag: TUE-AM-299

Paper

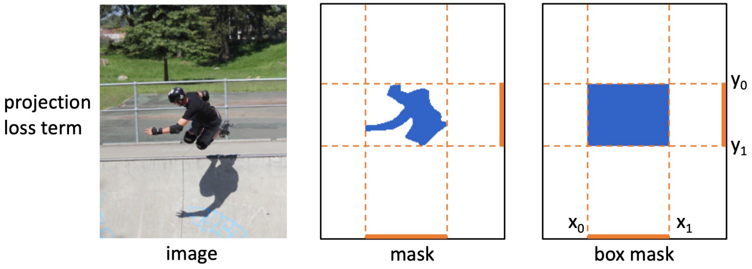
Code

Overview

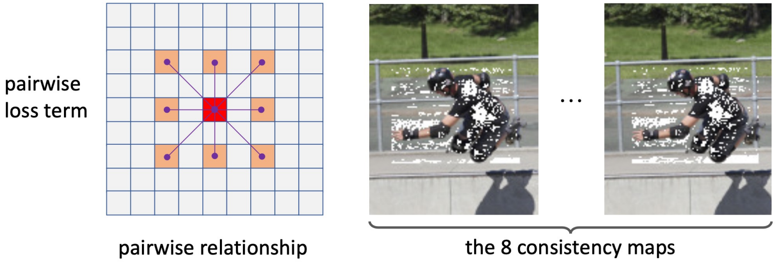
- Box-supervised Instance Segmentation



Using bounding boxes to supervise instance masks: (1) mask projection and (2) pairwise relations



mask projection for global localization

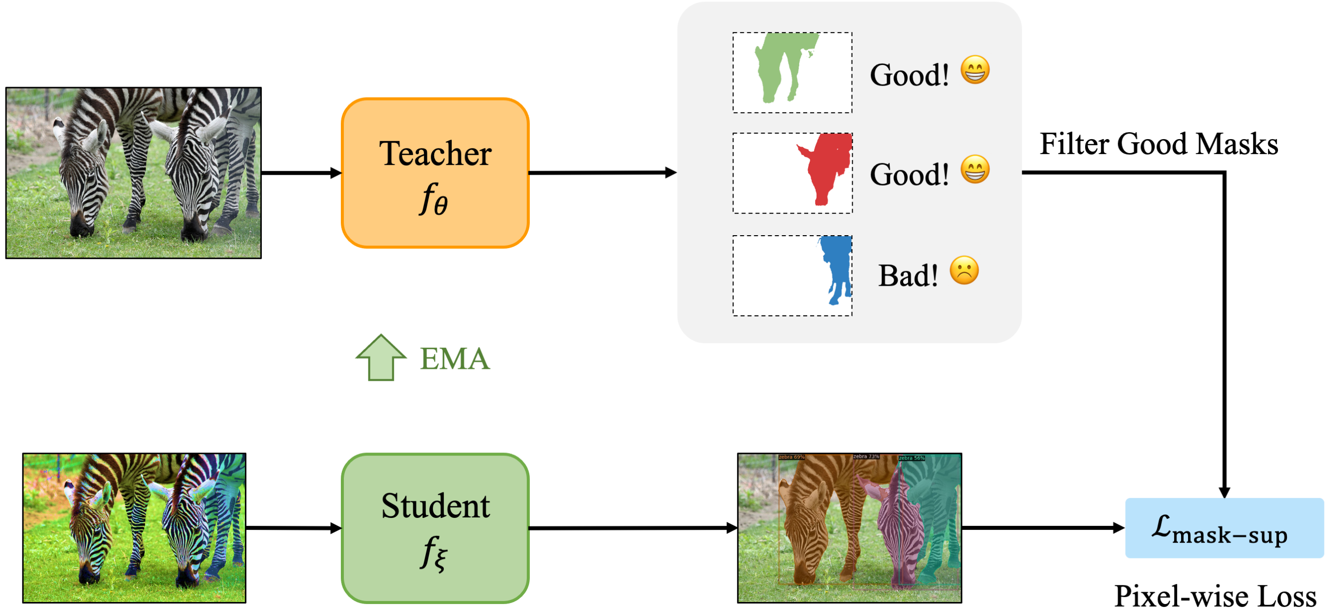


pairwise relations for local boundaries

Tian *et.al.* BoxInst: High-Performance Instance Segmentation with Box Annotations. CVPR 2021

Overview

- BoxTeacher



1. The Teacher generates and filters pseudo masks.
2. The Student learns from pseudo masks and updates the Teacher.

BoxTeacher leverages high-quality pseudo masks and bridges the gap between box-supervised and fully-supervised methods!

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

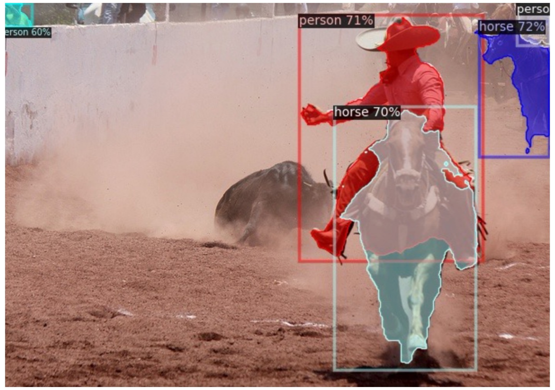
Motivation

Key Observation: box-supervised methods generate high-quality masks!

✓ Accurate localization ✓ Fine boundaries

Prediction

GT



Prediction

GT

Generated by BoxInst (30.7 AP on COCO val)

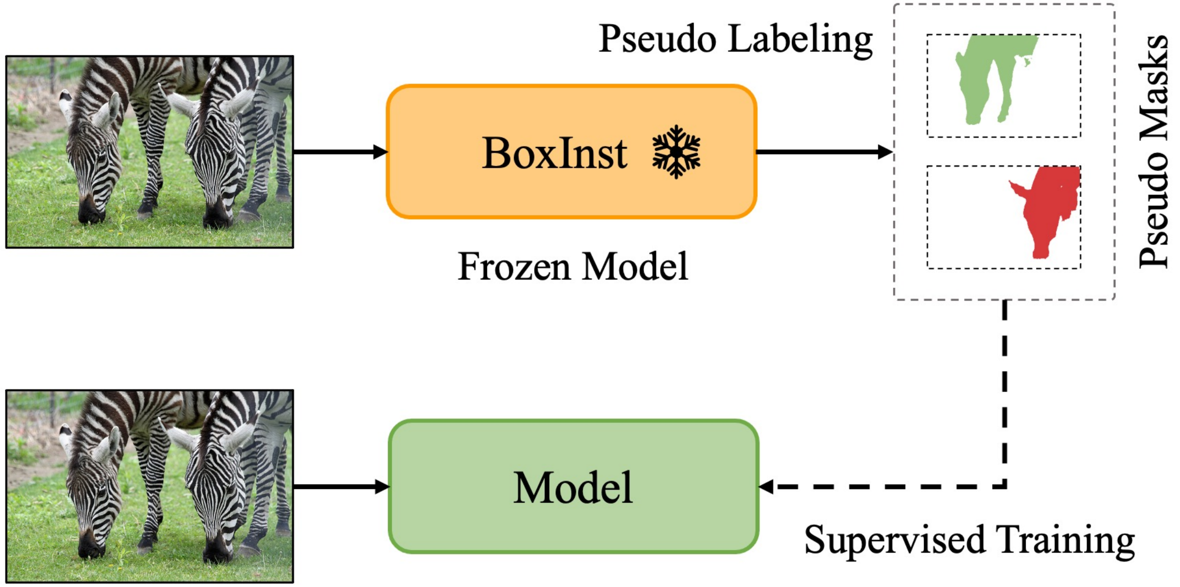
Can we leverage those **high-quality masks** to further improve box-supervised instance segmentation?

Tian *et.al.* BoxInst: High-Performance Instance Segmentation with Box Annotations. CVPR 2021

Method

- ## Naïve Self-Training

Firstly, we adopt a naïve self-training framework…

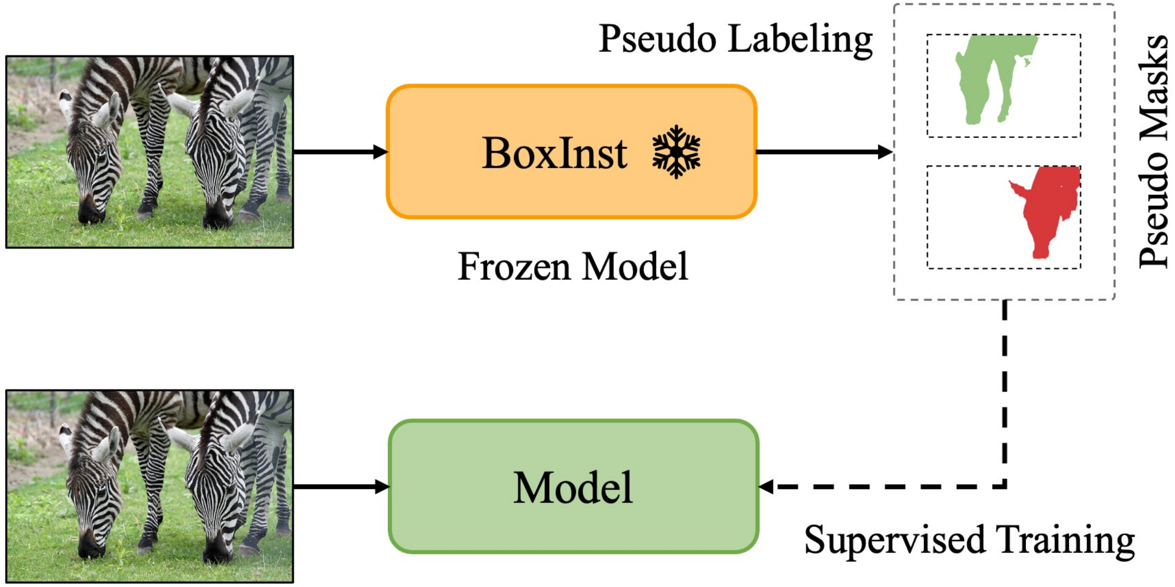


1. Using pre-trained and frozen BoxInst to generate pseudo masks for each image.
2. Training a new instance segmentation model with pseudo labeled samples

Method

▪ Naïve Self-Training

Firstly, we adopt a naïve self-training framework…



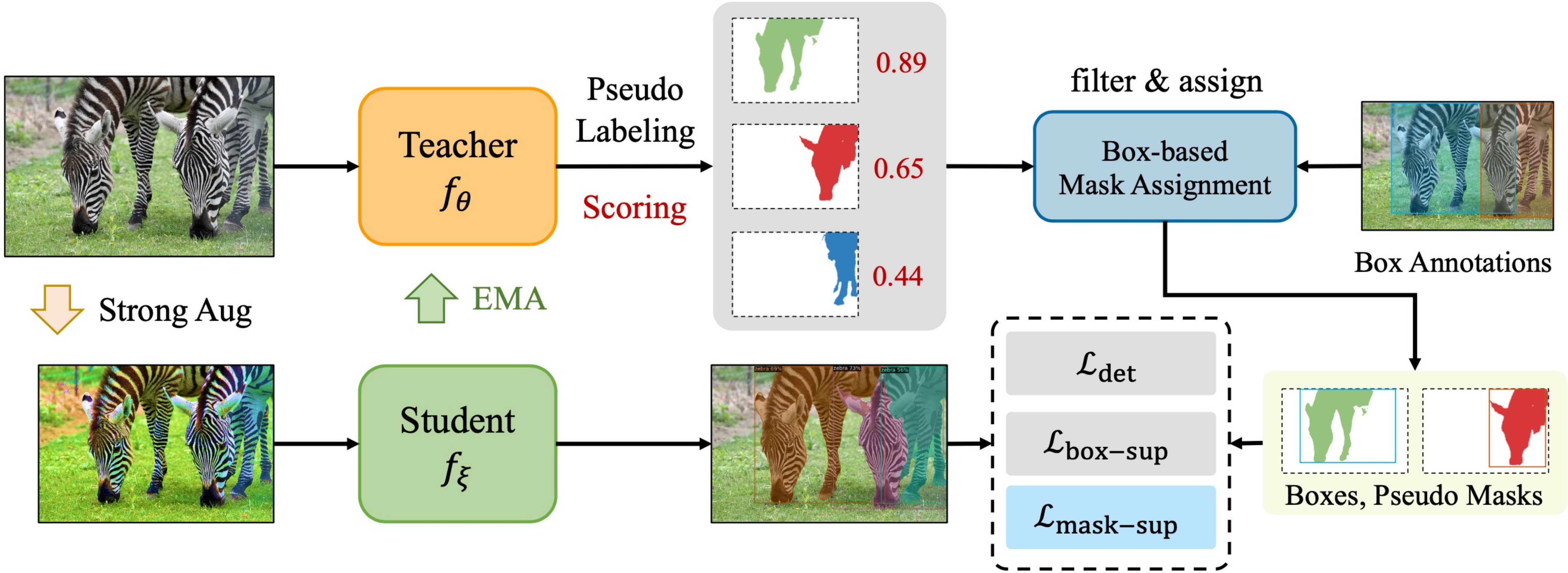
Method	AP	AP_{50}	AP_{75}
BoxInst, 1x	30.7	52.2	31.1
Self-Training, 1x	31.0	53.1	31.6
BoxInst, 3x	31.8	54.0	32.0
Self-Training, 3x	31.3↓	53.8	31.7

Improvements are minor! ☹

Pseudo masks do contain much noise!

BoxTeacher

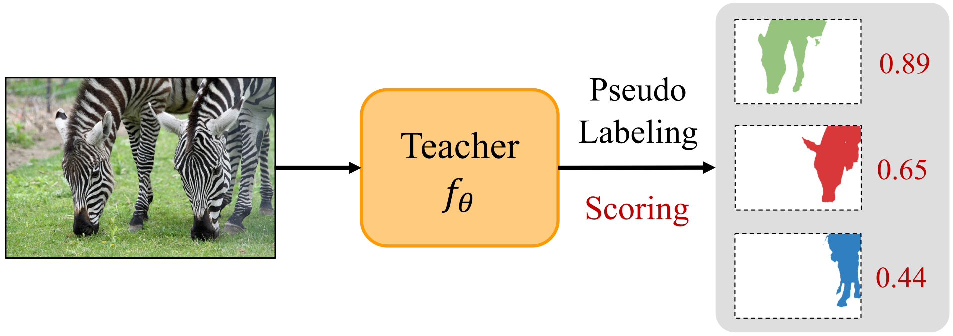
- Overall Architecture



An End-to-End Training Framework, including pseudo labeling and self-training

BoxTeacher

- Pseudo Mask: Generation, Scoring, and Filtering



The Teacher:

✓ generates pseudo masks

✓ estimates **mask confidence scores**

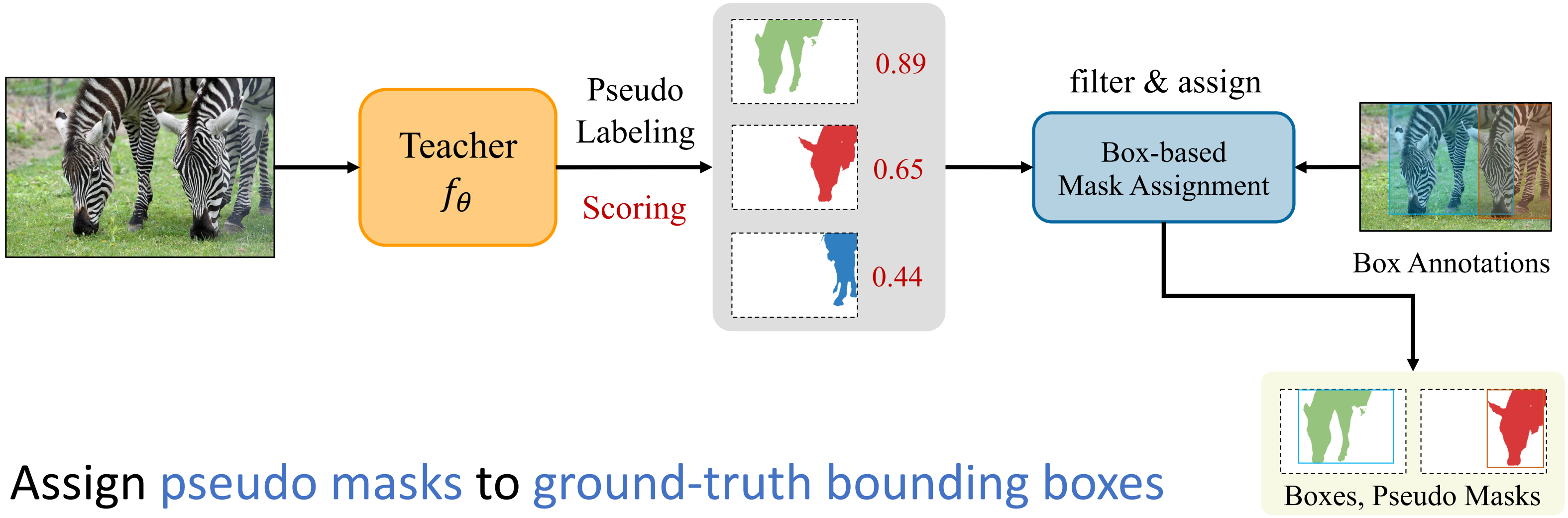
✓ filters low quality masks

Mask-aware Confidence Score: estimate mask quality

$$s_i = \sqrt{c_i \cdot \frac{\sum_{x,y}^{H,W} \mathbb{1}(m_{i,x,y} > \tau_m) \cdot m_{i,x,y} \cdot m^b_{i,x,y}}{\sum_{x,y}^{H,W} \mathbb{1}(m_{i,x,y} > \tau_m) \cdot m^b_{i,x,y}}},$$

BoxTeacher

- Box-based Mask Assignment



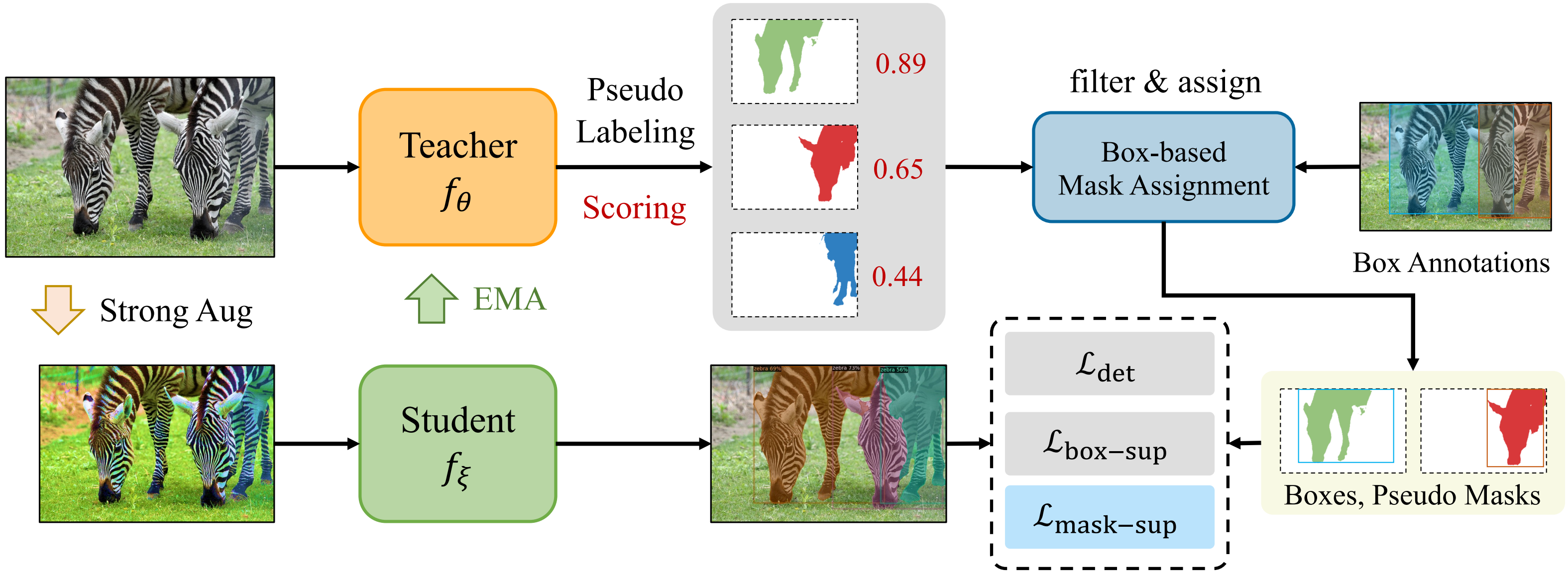
Assign pseudo masks to ground-truth bounding boxes according to:

✓ IoU

✓ Confidence score

BoxTeacher

- Training Student, Updating Teacher



The Student:

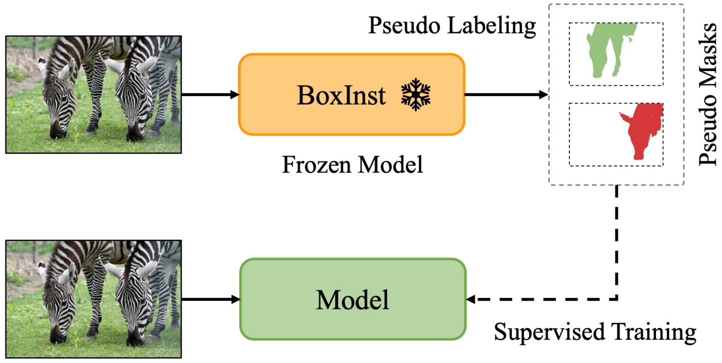
✓ Forwards with **perturbed** images

✓ Computes **detection loss**, **box-supervised loss**, and **mask-supervised loss**

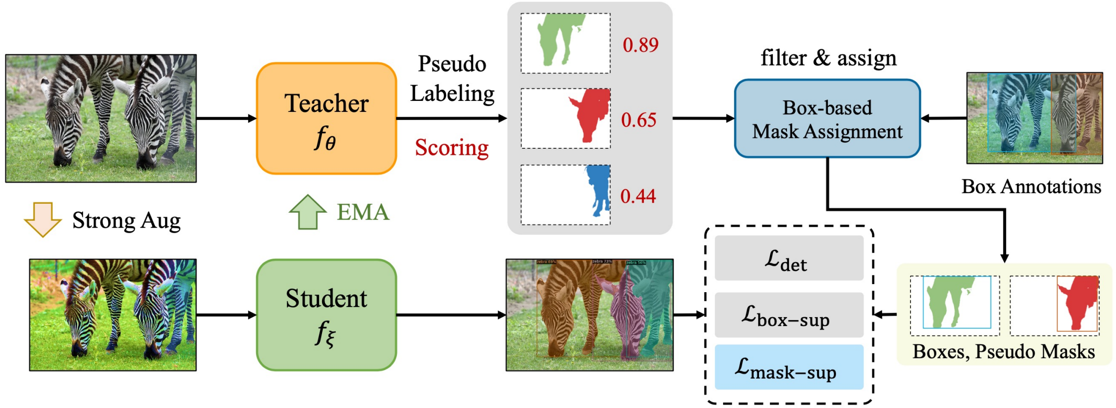
✓ Updates the Teacher via **EMA** (Exponential Moving Average)

BoxTeacher

Compare to Naïve Self-Training



Naïve Self-Training



BoxTeacher

- End-to-end training, simple, efficient
- **Flywheel**: better student leads to better teacher, and better teacher leads to better student
- Better performance

Experimental Results

COCO Instance Segmentation

Method	Backbone	Schedule	AP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Mask-supervised methods.								
Mask R-CNN [23]	R-50-FPN	1×	35.5	57.0	37.8	19.5	37.6	46.0
CondInst [49]	R-50-FPN	1×	35.9	57.0	38.2	19.0	38.6	46.7
CondInst [49]	R-50-FPN	3×	37.7	58.9	40.3	20.4	40.2	48.9
CondInst [49]	R-101-FPN	3×	39.1	60.9	42.0	21.5	41.7	50.9
SOLO [54]	R-101-FPN	6×	37.8	59.5	40.4	16.4	40.6	54.2
SOLOv2 [54]	R-101-FPN	6×	39.7	60.7	42.9	17.3	42.9	57.4
Box-supervised methods.								
BoxInst [51]	R-50-FPN	3×	32.1	55.1	32.4	15.6	34.3	43.5
DiscoBox [31]	R-50-FPN	3×	32.0	53.6	32.6	11.7	33.7	48.4
BoxTeacher†	R-50-FPN	1×	32.9	54.1	34.2	17.4	36.3	43.7
BoxTeacher	R-50-FPN	3×	35.0	56.8	36.7	19.0	38.5	45.9
BBTP [25]	R-101-FPN	1×	21.1	45.5	17.2	11.2	22.0	29.8
BBAM [32]	R-101-FPN	1×	25.7	50.0	23.3	-	-	-
BoxCaseg [53]	R-101-FPN	1×	30.9	54.3	30.8	12.1	32.8	46.3
BoxInst [51]	R-101-FPN	3×	33.2	56.5	33.6	16.2	35.3	45.1
BoxLevelSet [33]	R-101-FPN	3×	33.4	56.8	34.1	15.2	36.8	46.8
BoxLevelSet [33]	R-101-DCN-FPN	3×	35.4	59.1	36.7	16.8	38.5	51.3
DiscoBox [31]	R-101-DCN-FPN	3×	35.8	59.8	36.4	16.9	38.7	52.1
BoxTeacher	R-101-FPN	3×	36.5	59.1	38.4	20.1	40.2	47.9
BoxTeacher	R-101-DCN-FPN	3×	37.6	60.3	39.7	21.0	41.8	49.3
BoxTeacher	Swin-Base-FPN	3×	40.6	65.0	42.5	23.4	44.9	54.2

- ✓ BoxTeacher achieves the State-of-the-Art performance!
- ✓ BoxTeacher bridges the gap between box-supervised and fully-supervised methods, BoxTeacher (36.5) *v.s.* CondInst (39.1)

Experimental Results

- BoxTeacher on Other Datasets

On PASCAL VOC

Method	Backbone	AP	AP_{25}	AP_{50}	AP_{70}	AP_{75}
SDI [28]	VGG-16	-	-	44.8	-	16.3
BoxInst [51]	R-50	34.3	-	59.1	-	34.2
DiscoBox [31]	R-50	-	71.4	59.8	41.7	35.5
BoxLevelSet [33]	R-50	36.3	76.3	64.2	43.9	35.9
BoxTeacher	R-50	38.6	77.6	66.4	46.1	38.7
BBTP [25]	R-101	-	75.0	58.9	30.4	21.6
Arun *et al.* [2]	R-101	-	73.1	57.7	33.5	31.2
BBAM [32]	R-101	-	76.8	63.7	39.5	31.8
BoxInst [51]	R-101	36.4	-	61.4	-	37.0
DiscoBox [31]	R-101	-	72.8	62.2	45.5	37.5
BoxLevelSet [33]	R-101	38.3	77.9	66.3	46.4	38.7
BoxTeacher	R-101	40.3	78.4	67.8	48.0	41.3

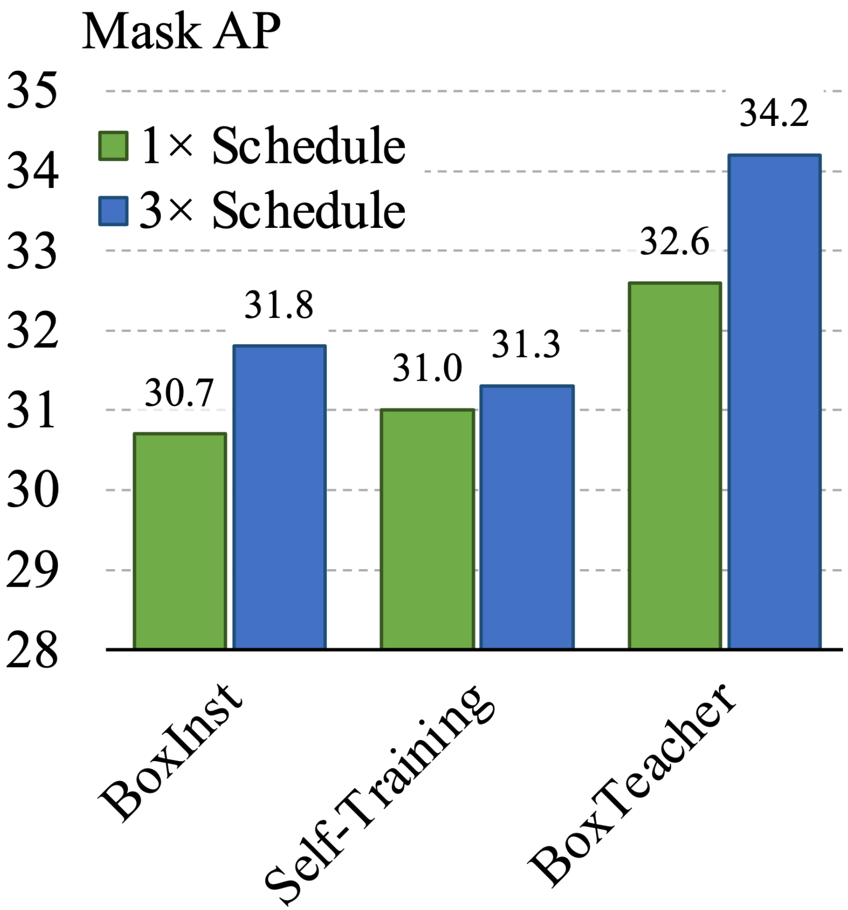
On Cityscapes

Method	Data	AP	AP_{50}
Mask-supervised methods.			
Mask R-CNN [23]	`fine`	31.5	-
CondInst [49]	`fine`	33.0	59.3
CondInst [49]	`fine` + COCO	37.8	63.4
Box-supervised methods.			
BoxInst† [51]	`fine`	19.0	41.8
BoxInst† [51]	`fine` + COCO	24.0	51.0
BoxLevelSet† [33]	`fine`	20.7	43.3
BoxLevelSet† [33]	`fine` + COCO	22.7	46.6
BoxTeacher	`fine`	21.7	47.5
BoxTeacher	`fine` + COCO	26.8	54.2

Experimental Results

- Ablation Experiments

Comparison to Naïve Self-Training



Strong Perturbation for Student

Method	Schd.	weak	strong	AP^b	AP^m
CondInst	1×			39.6	36.2
CondInst	1×	✓		39.6	$35.6^{-0.6}$
CondInst	1×		✓	39.2	$35.3^{-0.9}$
BoxTeacher	1×			39.4	32.6
BoxTeacher	1×	✓		39.1	$32.4^{-0.2}$
BoxTeacher	1×		✓	38.8	$32.2^{-0.4}$
CondInst	3×			41.9	37.5
CondInst	3×		✓	42.0	$37.6^{+0.1}$
BoxTeacher	3×			41.7	34.2
BoxTeacher	3×		✓	41.8	$34.8^{+0.6}$

1. Strong perturbation is useless for fully-supervised training
2. Strong perturbation with longer schedule is beneficial to BoxTeacher

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Experimental Results

- Qualitative Results

BoxTeacher can generate high-quality segmentation masks with fine-grained boundaries!

Horizon Robotics

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

Thanks

Tianheng Cheng

thch@hust.edu.cn

Paper

Code