模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

# Current Status and Development Trend of Semi-supervised Objective Detection Research

欣子豪

2024.3.11

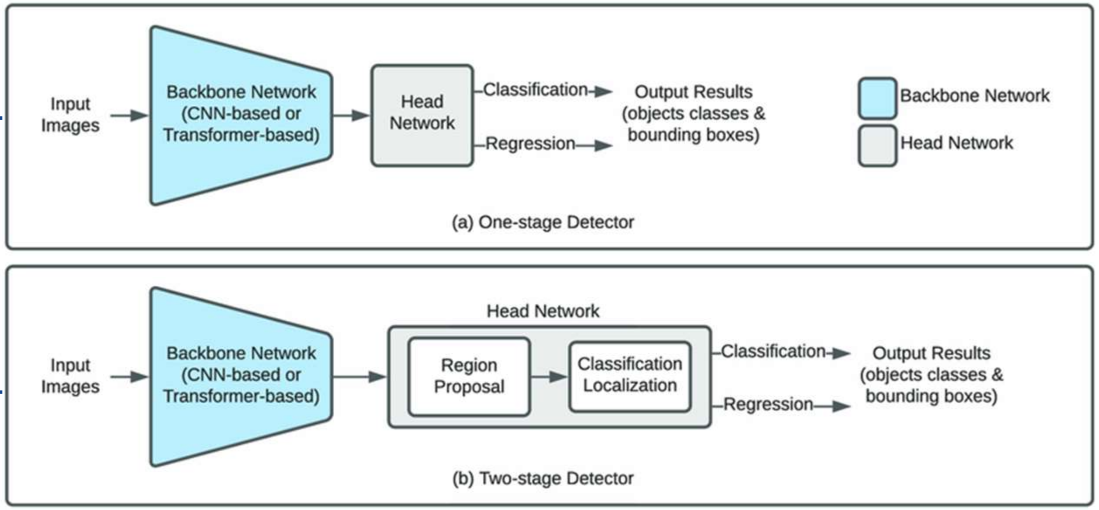
# Background



SSOD

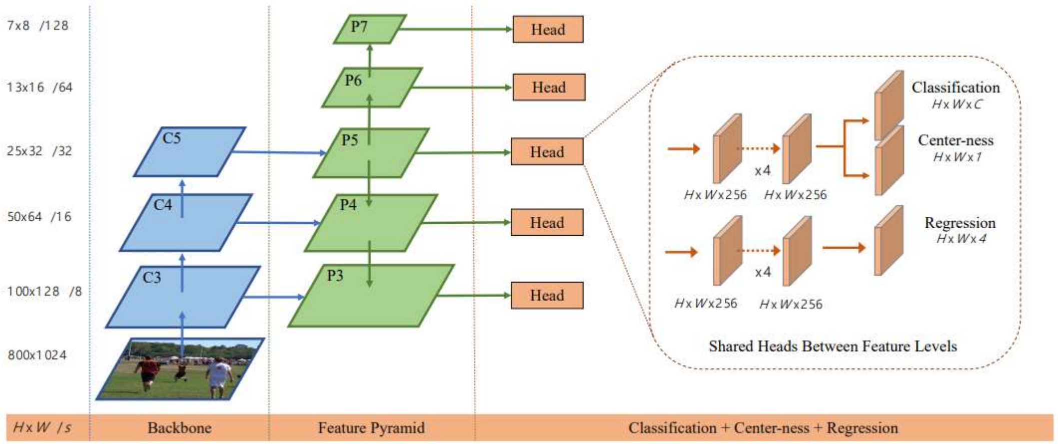
## Object Detection

① **anchor-based**




② **anchor-free**

**(FCOS)**



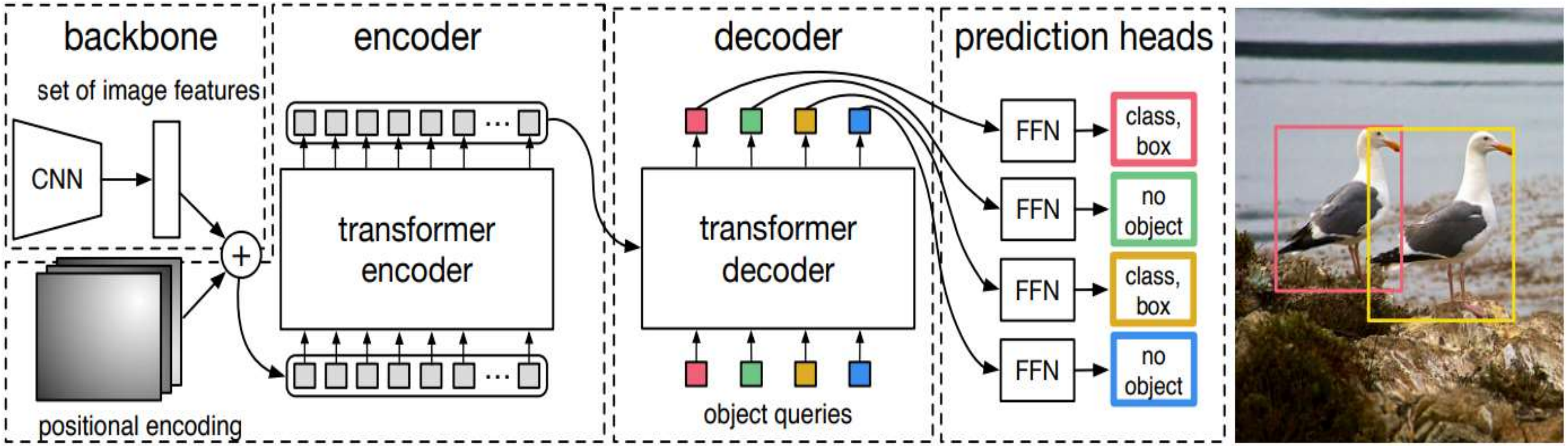
# Background

① anchor-based
- one-stage
- two-stage

② anchor-free

(FCOS)

③ DETR-based

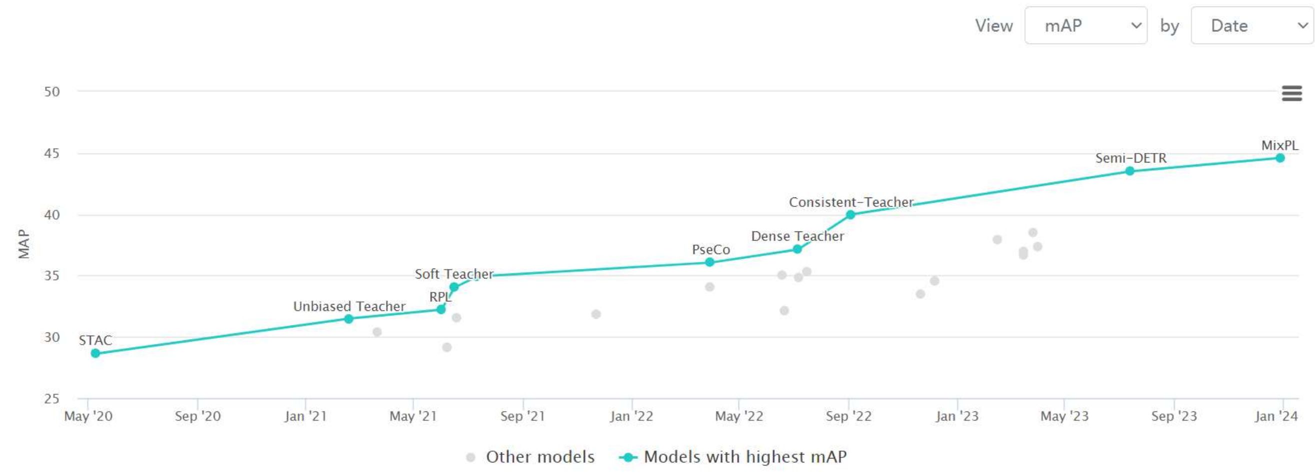


# Background



## Semi-Supervised Object Detection on COCO 10% labeled data

Leaderboard Dataset




| Rank | Model | | mAP | detector | Year |
|---|---|---|---|---|---|
| 1 | **MixPL** | DETR | 44.6 | DINO-Res50 | 2023 |
| 2 | **Semi-DETR** | DETR | 43.5 | DINO-Res50 | 2023 |
| 3 | **Consistent-Teacher** | One-stage | 40.0 | RetinaNet-Res50 | 2022 |
| 4 | **ARSL** | Anchor-free | 38.5 | FCOS-Res50 | 2023 |
| 5 | **Efficient Teacher** | One-stage | 37.9 | YOLOv5-L | 2023 |
| 6 | **Revisiting Class Imbalance** | Two-stage | 37.4 | FasterRCNN-Res50 | 2023 |
| 7 | **Dense Teacher** | Anchor-free | 37.13 | FCOS-Res50 | 2022 |
| 8 | **MixTeacher-FCOS** | Anchor-free | 36.95 | FCOS-Res50 | 2023 |
| 9 | **MixTeacher-FRCNN** | Two-stage | 36.72 | FRCNN-Res50 | 2023 |
| 10 | **PseCo** | Two-stage | 36.06 | FasterRCNN-Res50 | 2022 |

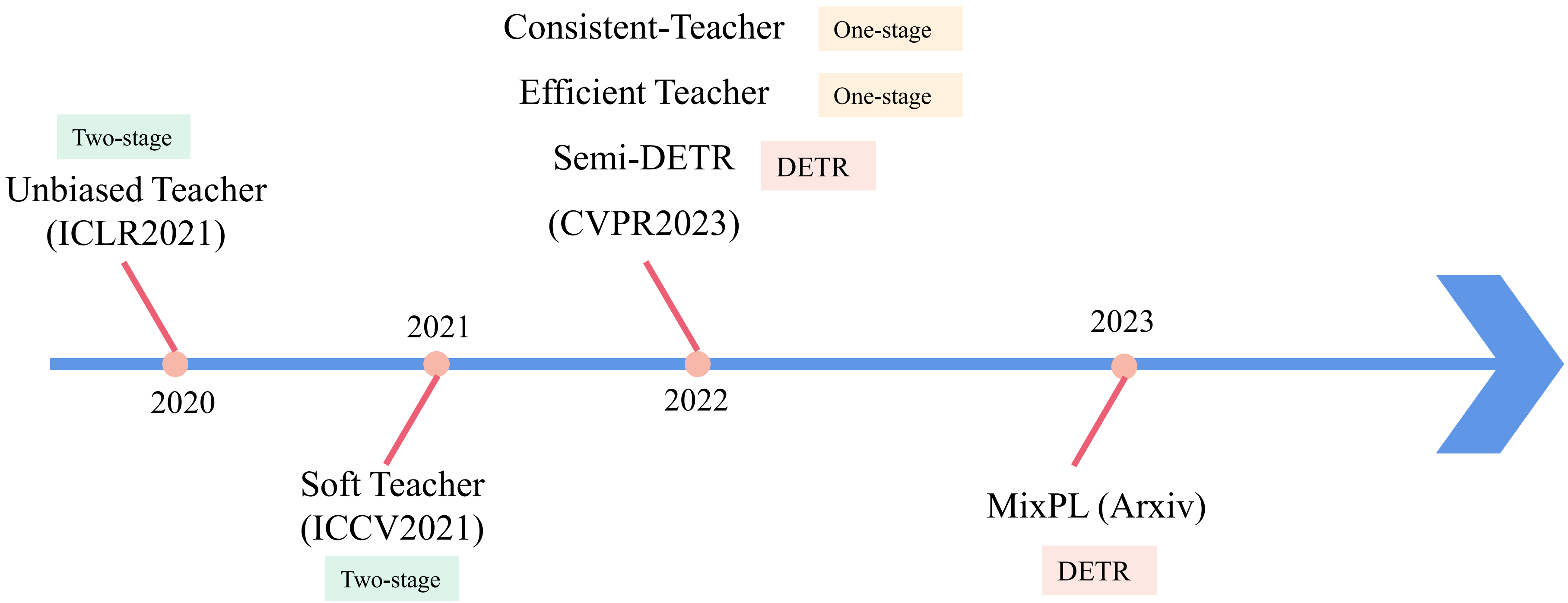
# Background

ParNeC 模式识别与神经计算研究组 PAttern Recognition and NEural Computing

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 12 | **Unbiased Teacher v2** | Anchor-free | 35.08±0.02 | FCOS-Res50 | Unbiased Teacher v2: Semi-supervised Object Detection for Anchor-free and Anchor-based Detectors | | | 2022 |
| 13 | **Adaptive Class-Rebalancing** | Two-stage | 34.92±0.22 | | Semi-Supervised Object Detection with Adaptive Class-Rebalancing Self-Training | | | 2021 |
| 14 | **VC** | Two-stage | 34.82 | FasterRCNN-Res50 | Semi-supervised Object Detection via Virtual Category Learning | | | 2022 |
| 15 | **ASTOD** | Two-stage | 34.58 | | Adaptive Self-Training for Object Detection | | | 2022 |
| 16 | **Omni-DETR** | DETR | 34.1 | | Omni-DETR: Omni-Supervised Object Detection with Transformers | | | 2022 |
| 17 | **Soft Teacher** | Two-stage | 34.04 | FasterRCNN-Res50 | End-to-End Semi-Supervised Object Detection with Soft Teacher | | | 2021 |
| 18 | **SSOD with OCL and RUPL** | Two-stage | 33.53 | | Semi-Supervised Object Detection with Object-wise Contrastive Learning and Regression Uncertainty | | | 2022 |
| 19 | **RPL** | Two-stage | 32.23±0.14 | | Rethinking Pseudo Labels for Semi-Supervised Object Detection | | | 2021 |
| 20 | **Il-net** (resnet-50) | Two-stage | 32.166 | | Improving Localization for Semi-Supervised Object Detection | | | 2022 |

# Background

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing



模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

# Unbiased Teacher For Semi-Supervised Object Detection

**Yen-Cheng Liu**[1,2]*, **Chih-Yao Ma**[2], **Zijian He**[2], **Chia-Wen Kuo**[1], **Kan Chen**[2], **Peizhao Zhang**[2], **Bichen Wu**[2], **Zsolt Kira**[1], **Peter Vajda**[2]
[1]Georgia Tech, [2]Facebook Inc.
{ycliu,cwkuo,zkira}@gatech.edu,
{cyma,zijian,kanchen18,stzpz,wbc,vajdap}@fb.com

Two-stage

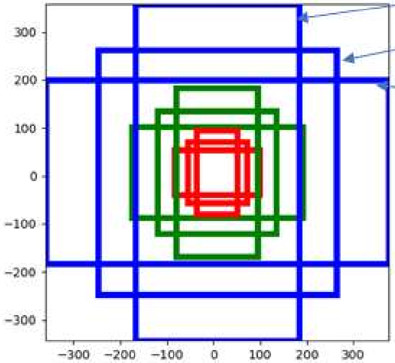
ICLR 2021

# Background

**Generate Anchors**

Given:

- Set of aspect ratios (0.5, 1, 2)
- Stride length (downscaling performed by resnet head: 16)
- Anchor Scales (8, 16, 32)




Aspect Ratio: 0.5

Aspect Ratio: 1

Aspect Ratio: 2

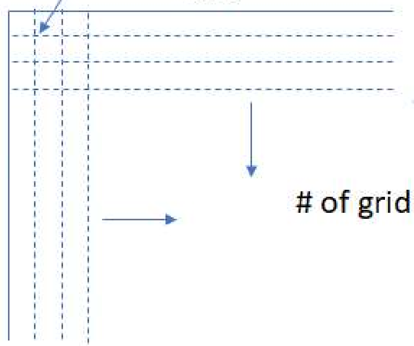
# of anchors: 9

Add anchors to every grid location

16,16

800

600

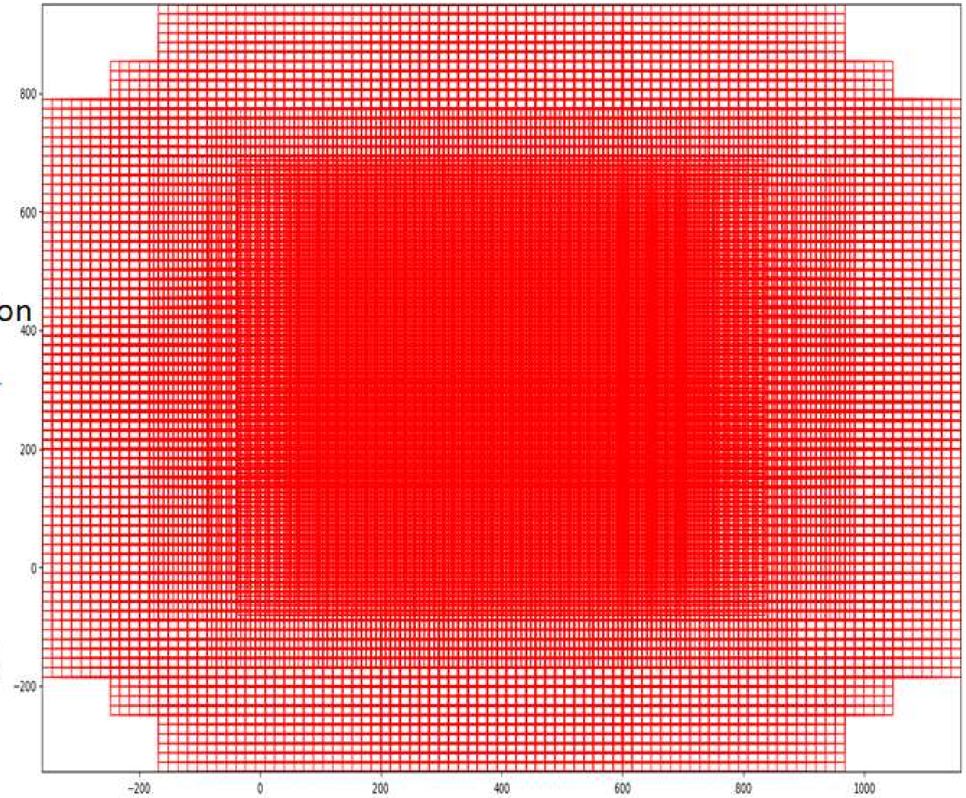


Create uniformly spaced grid with spacing = stride length

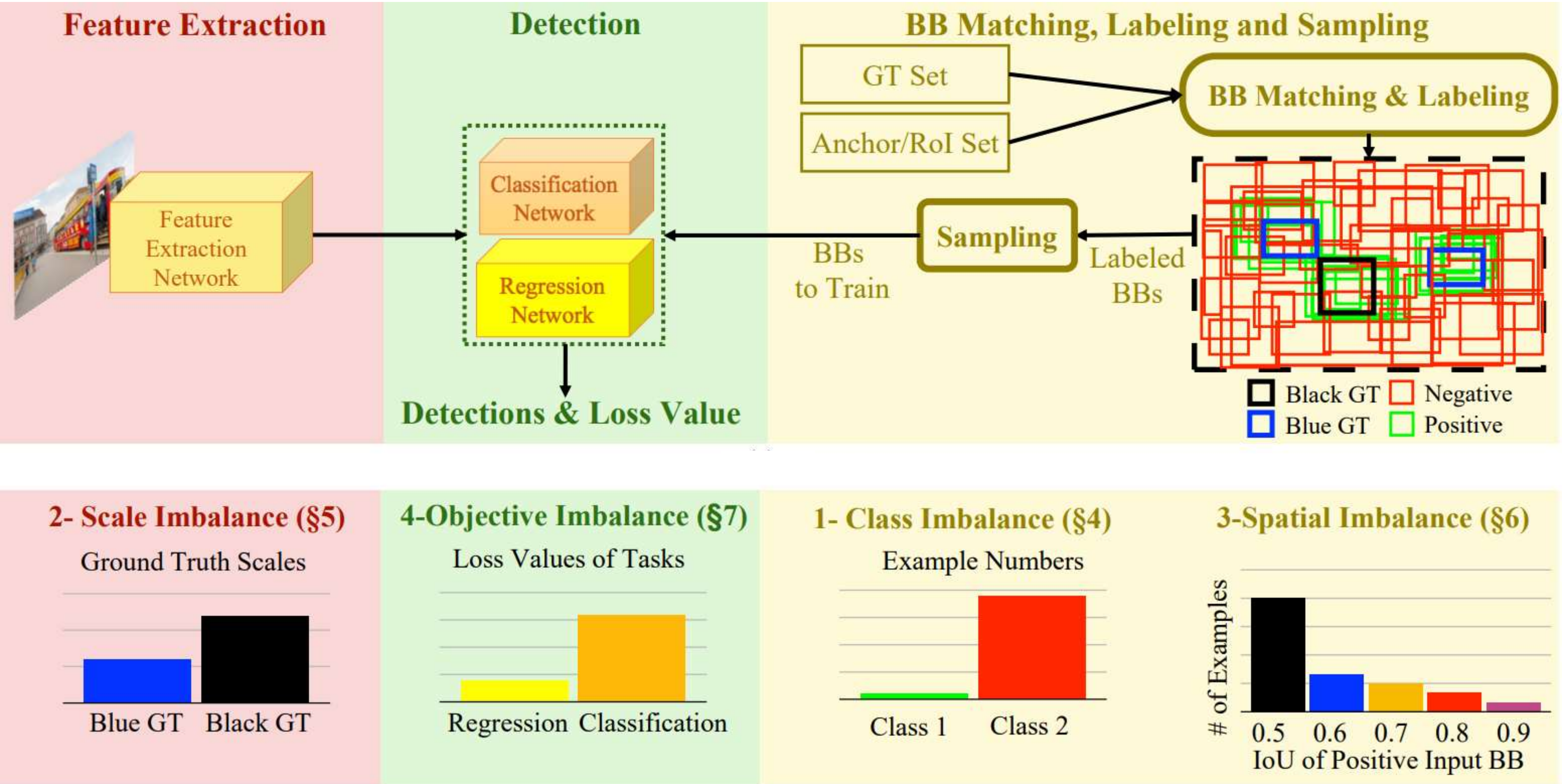
# of grid locations: $\frac{800}{16} * \frac{600}{16} = 1900$

Total number of anchors: 1900*9 = 17100

Some boxes lie outside the image boundary

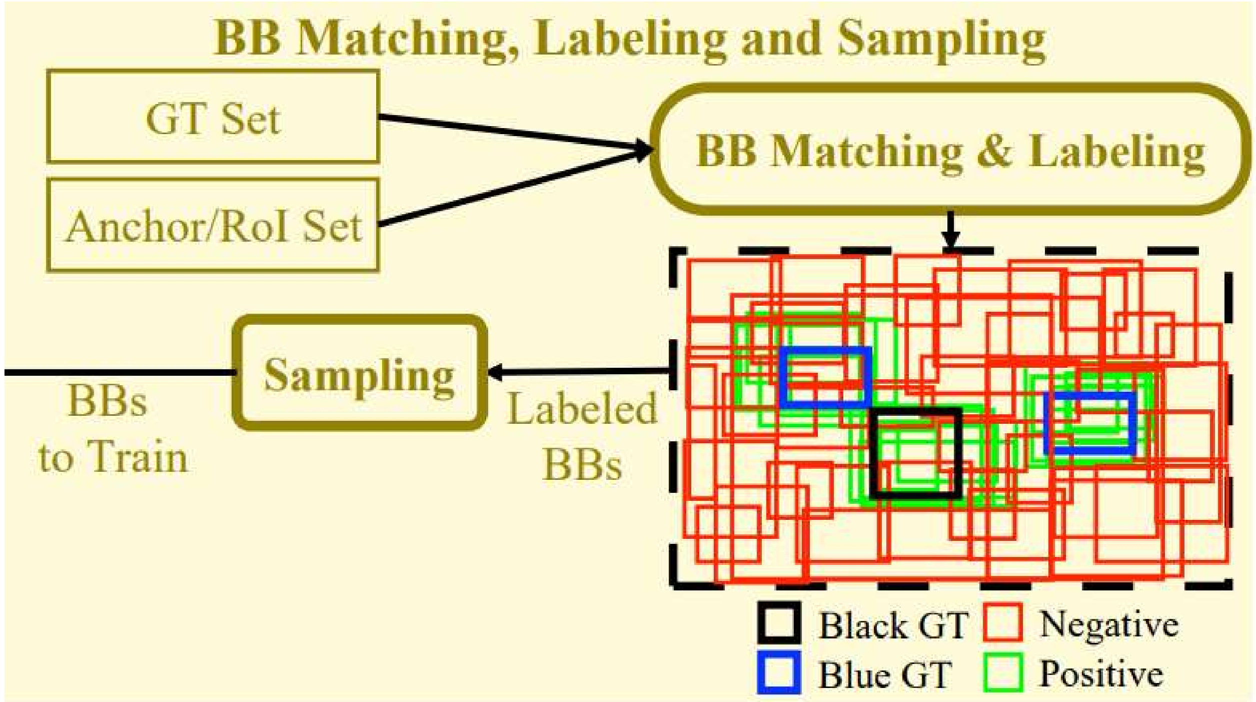


# Background



Feature Extraction
Feature Extraction Network
Detection
Classification Network
Regression Network
Detections & Loss Value
BB Matching, Labeling and Sampling
GT Set
Anchor/RoI Set
BB Matching & Labeling
Sampling
BBs to Train
Labeled BBs
Black GT
Negative
Blue GT
Positive
2- Scale Imbalance (§5)
Ground Truth Scales
Blue GT
Black GT
4-Objective Imbalance (§7)
Loss Values of Tasks
Regression
Classification
1- Class Imbalance (§4)
Example Numbers
Class 1
Class 2
3-Spatial Imbalance (§6)
# of Examples
0.5
0.6
0.7
0.8
0.9
IoU of Positive Input BB

# Background

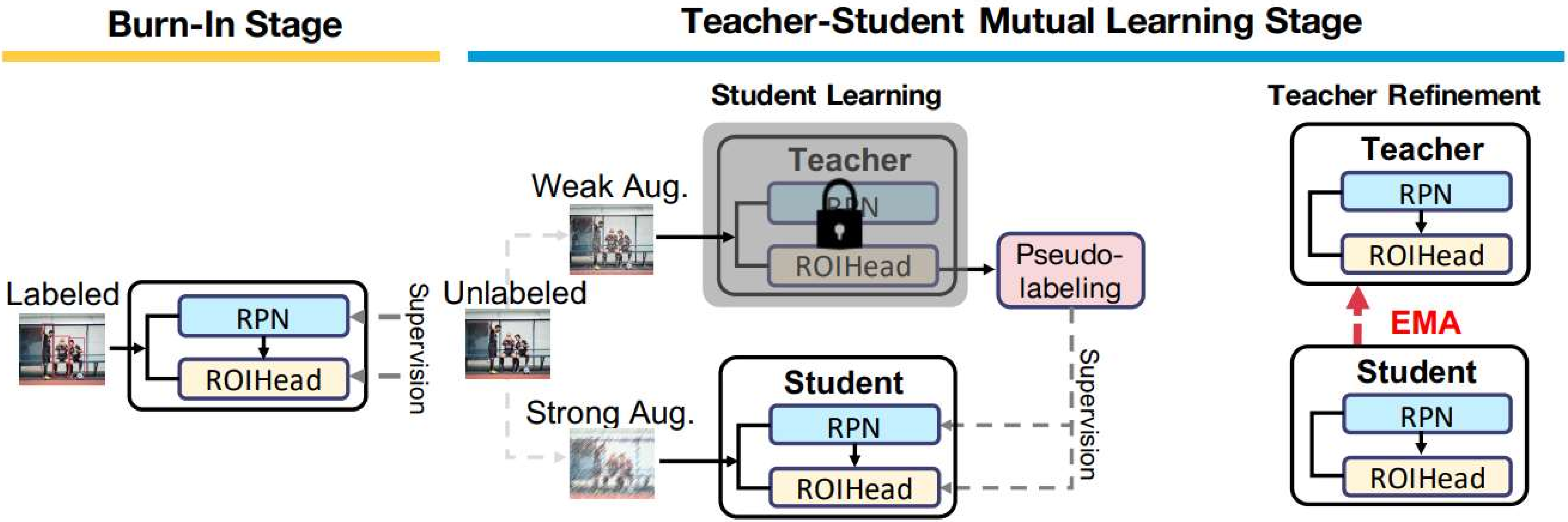



Pseudo-labeling Methods:

- Imbalance between background and foreground
- Imbalance between classes

# Unbiased Teacher





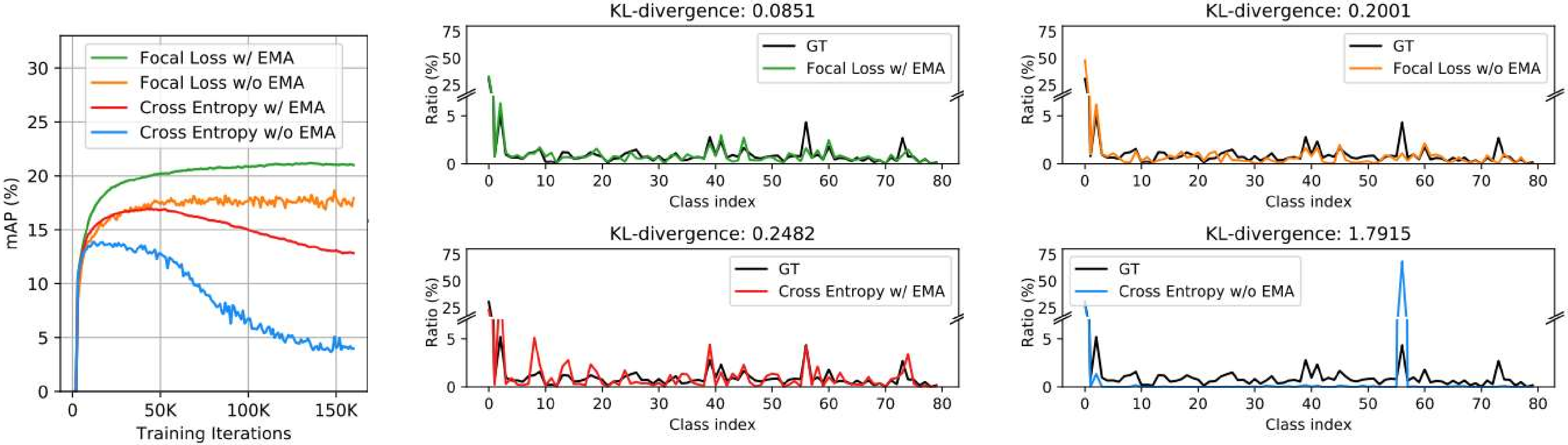

Burn-in: $\mathcal{L}_{sup} = \sum_i \mathcal{L}_{cls}^{rpn}(\boldsymbol{x}_i^s, \boldsymbol{y}_i^s) + \mathcal{L}_{reg}^{rpn}(\boldsymbol{x}_i^s, \boldsymbol{y}_i^s) + \mathcal{L}_{cls}^{roi}(\boldsymbol{x}_i^s, \boldsymbol{y}_i^s) + \boxed{\mathcal{L}_{reg}^{roi}(\boldsymbol{x}_i^s, \boldsymbol{y}_i^s)}$

Mutual Learning: $\mathcal{L}_{unsup} = \sum_i \mathcal{L}_{cls}^{rpn}(\boldsymbol{x}_i^u, \hat{\boldsymbol{y}}_i^u) + \boxed{\mathcal{L}_{cls}^{roi}(\boldsymbol{x}_i^u, \hat{\boldsymbol{y}}_i^u)}$ $\qquad \theta_s \leftarrow \theta_s + \gamma \frac{\partial(\mathcal{L}_{sup} + \boldsymbol{\lambda}_u \mathcal{L}_{unsup})}{\partial \theta_s}$

# Unbiased Teacher



**KL-divergence between the ground-truth labels distribution and the pseudo-label distribution.**




Focal loss:

$$\mathrm{FL}(p_t) = -\alpha_t(1-p_t)^\gamma \log(p_t)$$

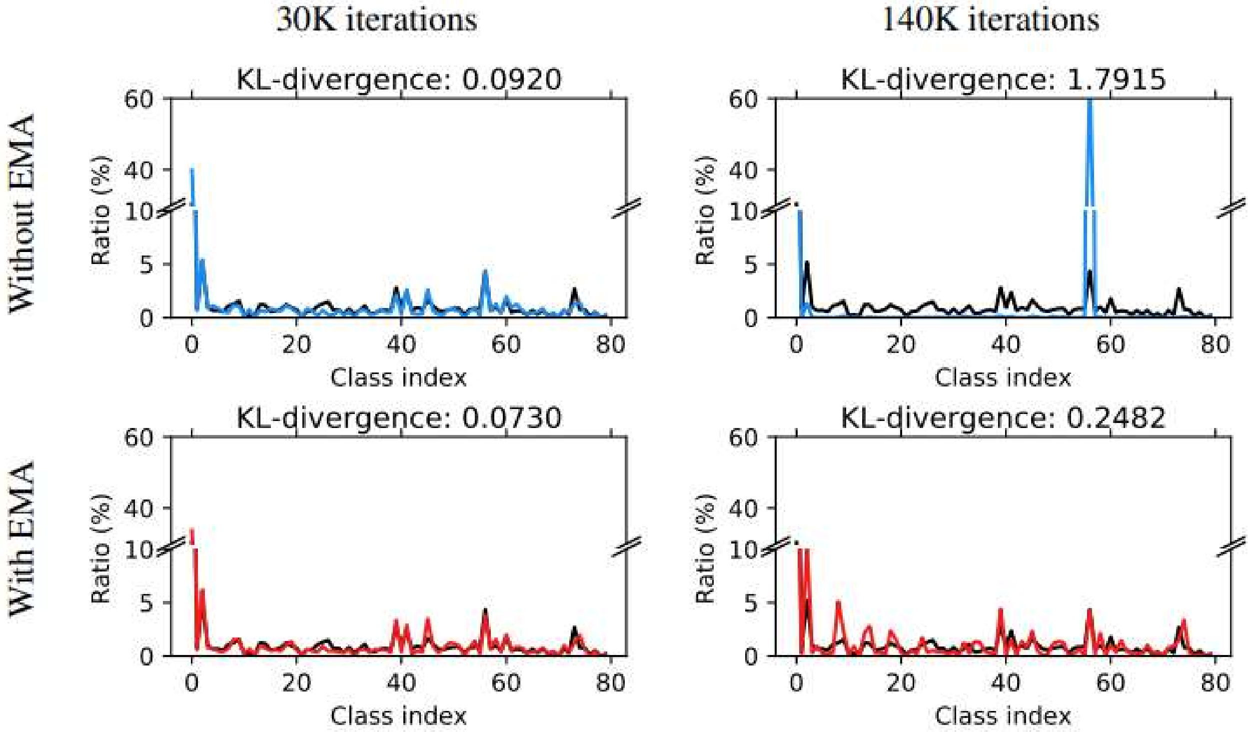
① Mitigating Easy Samples: small weights to easy samples.

② Balancing Class Distribution: balance the weights of positive and negative samples by adjusting gamma.

# Unbiased Teacher



**EMA on Imbalanced Pseudo-labeling**




$$\theta_t^i = \hat{\theta} - \gamma \sum_{k=1}^{i-1} (1 - \alpha^{-k+(i-1)}) \frac{\partial(\mathcal{L}_{sup} + \boldsymbol{\lambda}_u \mathcal{L}_{unsup})}{\partial \theta_s^k}$$

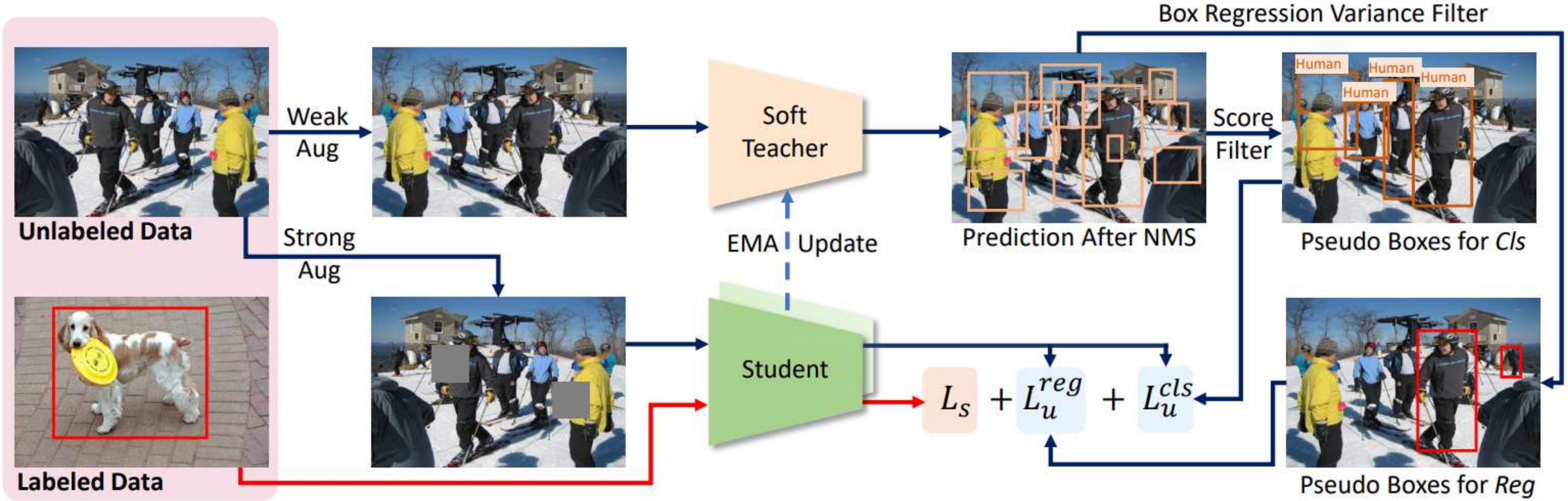
# End-to-End Semi-Supervised Object Detection with Soft Teacher

Mengde Xu[1†*] Zheng Zhang[1,2*‡] Han Hu[2‡] Jianfeng Wang[2] Lijuan Wang[2] Fangyun Wei[2]
Xiang Bai[1] Zicheng Liu[2]
[1]Huazhong University of Science and Technology
{mdxu,xbai}@hust.edu.cn
[2]Microsoft
{zhez,hanhu,jianfw,lijuanw,fawe,zliu}@microsoft.com

Two-stage

ICCV 2021

# Soft Teacher




$$\mathcal{L}_s = \frac{1}{N_l}\sum_{i=1}^{N_l}\left(\mathcal{L}_{\text{cls}}(I_l^i) + \mathcal{L}_{\text{reg}}(I_l^i)\right)$$

$$\mathcal{L}_u = \frac{1}{N_u}\sum_{i=1}^{N_u}\left(\mathcal{L}_{\text{cls}}(I_u^i) + \mathcal{L}_{\text{reg}}(I_u^i)\right)$$

$$\mathcal{L}_u^{\text{cls}} = \frac{1}{N_b^{\text{fg}}}\sum_{i=1}^{N_b^{\text{fg}}} l_{\text{cls}}(b_i^{\text{fg}}, \mathcal{G}_{\text{cls}}) + \sum_{j=1}^{N_b^{\text{bg}}} w_j l_{\text{cls}}(b_j^{\text{bg}}, \mathcal{G}_{\text{cls}})$$

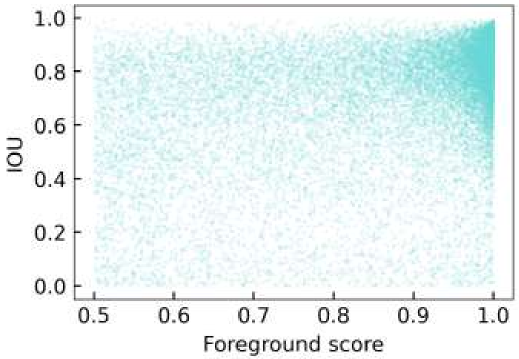
pseudo boxes

$$w_j = \frac{r_j}{\sum_{k=1}^{N_b^{\text{bg}}} r_k}$$

reliability score

(the background score produced by the teacher model)

# Soft Teacher



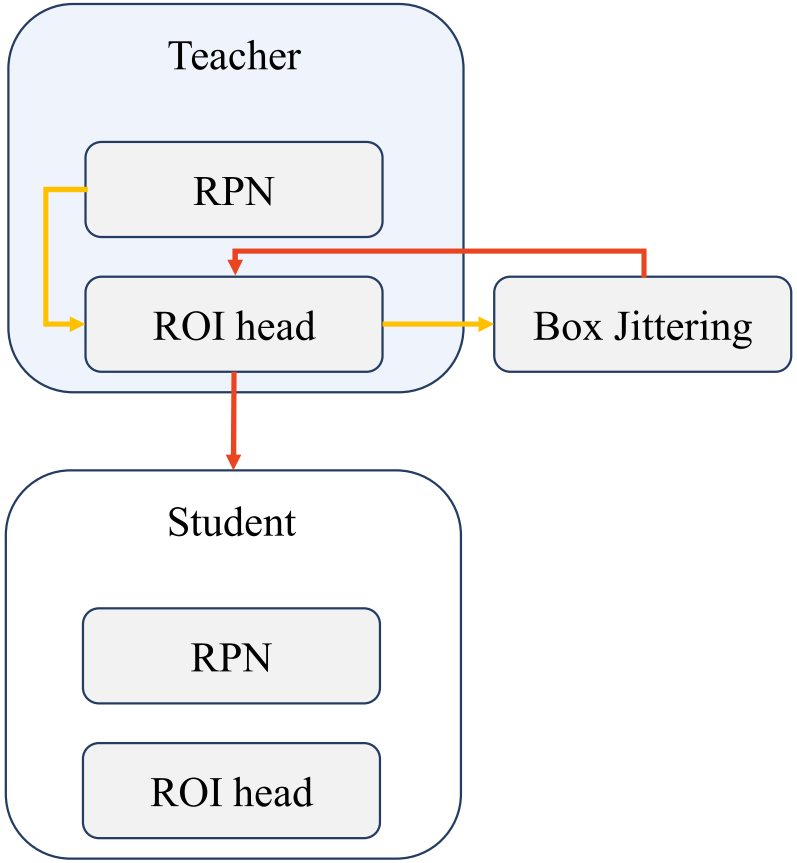

Not strong positive correlation

$$\hat{b}_i = \text{refine}(\text{jitter}(b_i))$$

$$\bar{\sigma_i} = \frac{1}{4}\sum_{k=1}^{4}\frac{\sigma_k}{0.5(h(b_i)+w(b_i))}$$

$$\mathcal{L}_u^{\text{reg}} = \frac{1}{N_b^{\text{fg}}}\sum_{i=1}^{N_b^{\text{fg}}} l_{\text{reg}}(b_i^{\text{fg}}, \mathcal{G}_{\text{reg}})$$

$$\mathcal{L}_u = \frac{1}{N_u}\sum_{i=1}^{N_u}(\mathcal{L}_u^{\text{cls}}(I_u^i, \mathcal{G}_{\text{cls}}^i) + \mathcal{L}_u^{\text{reg}}(I_u^i, \mathcal{G}_{\text{reg}}^i))$$




Estimating the localization reliability of a candidate pseudo box by measuring the consistency of its regression prediction

模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

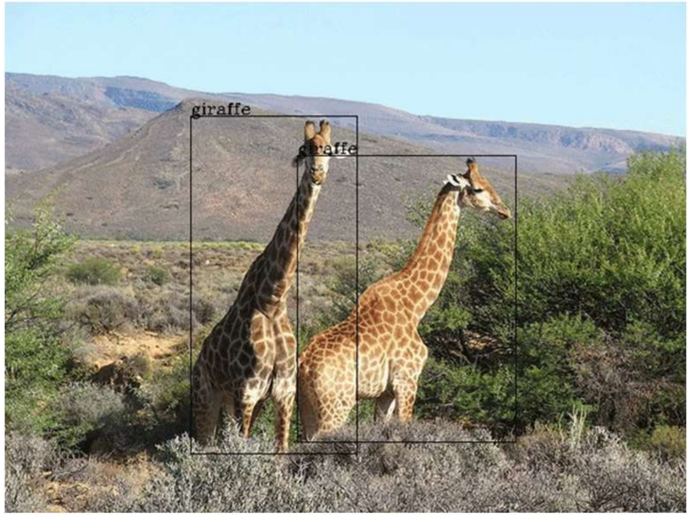
# Consistent-Teacher: Towards Reducing Inconsistent Pseudo-targets in Semi-supervised Object Detection


Xinjiang Wang[1*] Xingyi Yang[3*†] Shilong Zhang[2], Yijiang Li[1‡]
Litong Feng[1] Shijie Fang[4‡] Chengqi Lyu[2] Kai Chen[2] Wayne Zhang[1]
[1]SenseTime Research [2]Shanghai AI Laboratory [3]National University of Singapore [4]Peking University
wangxinjiang@sensetime.com, xyang@u.nus.edu

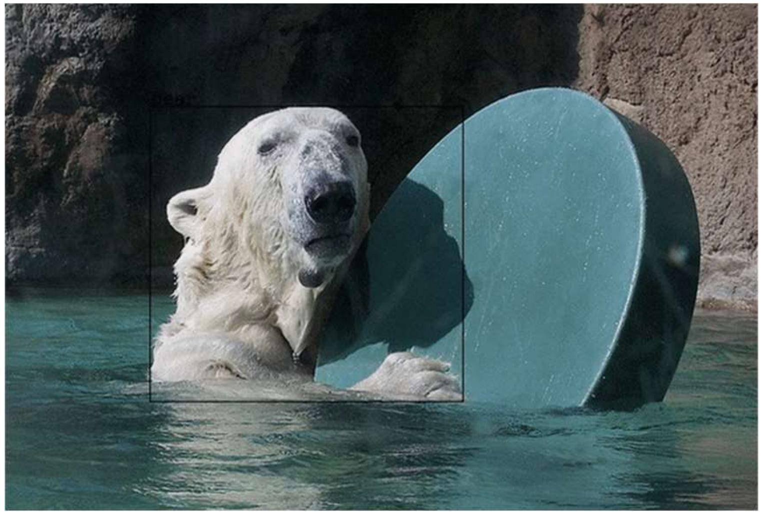

One-stage

CVPR 2023

# Consistent Teacher



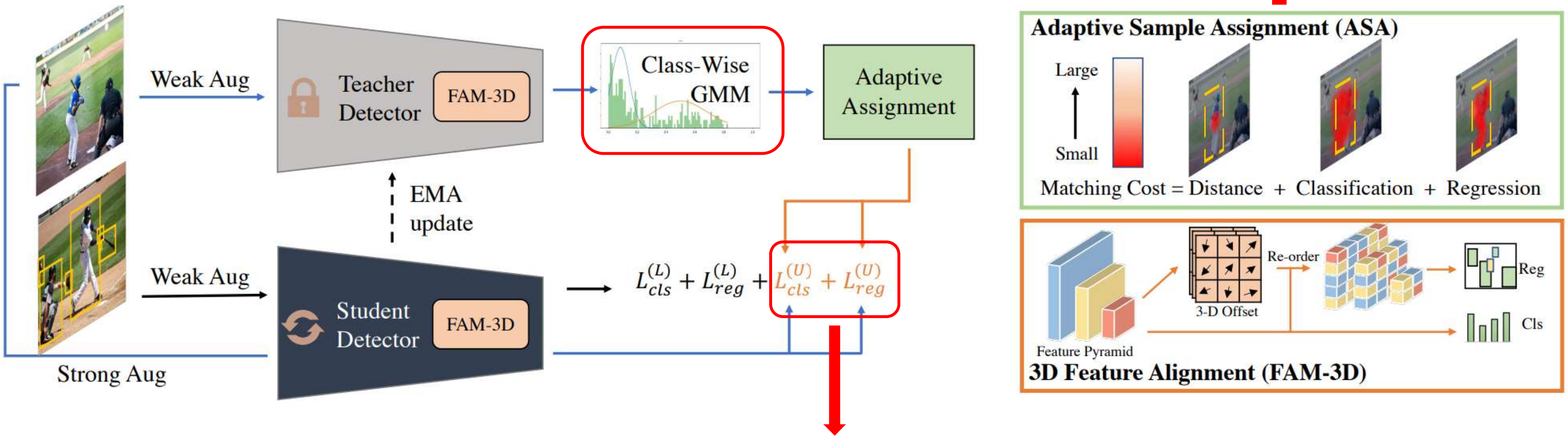




# Consistent Teacher

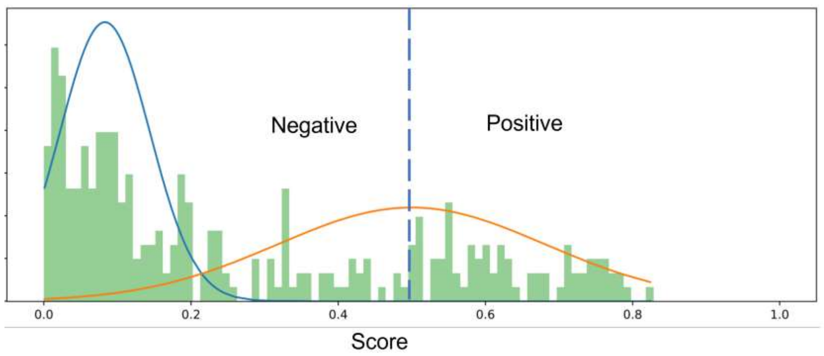


$$C_{nl} = \mathcal{L}_{cls}(p_n, y_l) + \lambda_{reg}\mathcal{L}_{reg}(p_n, y_l) + \lambda_{dist}C_{dist}$$




$$\mathcal{L} = \frac{1}{N}\sum_i \Big[\mathcal{L}_{cls}\big(f_s(T(\mathbf{x}_i^l)), \mathbf{y}_i^l\big) + \mathcal{L}_{reg}\big(f_s(T(\mathbf{x}_i^l)), \mathbf{y}_i^l\big)\Big]$$
$$+ \lambda_u \frac{1}{M}\sum_j \Big[\mathcal{L}_{cls}\big(f_s(T'(\mathbf{x}_j^u)), \hat{\mathbf{y}}_j^u\big) + \mathcal{L}_{reg}\big(f_s(T'(\mathbf{x}_j^u)), \hat{\mathbf{y}}_j^u\big)\Big]$$

# Consistent Teacher




unlabeled data have two modalities: positive and negative

$$\mathcal{P}(s^c) = w_n^c \mathcal{N}(s^c | \mu_n^c, (\sigma_n^c)^2) + w_p^c \mathcal{N}(s^c | \mu_p^c, (\sigma_p^c)^2)$$

Gaussian distribution

weight

EM algorithm

$$\mathcal{P}(pos | s^c, \mu_p^c, (\sigma_p^c)^2)$$

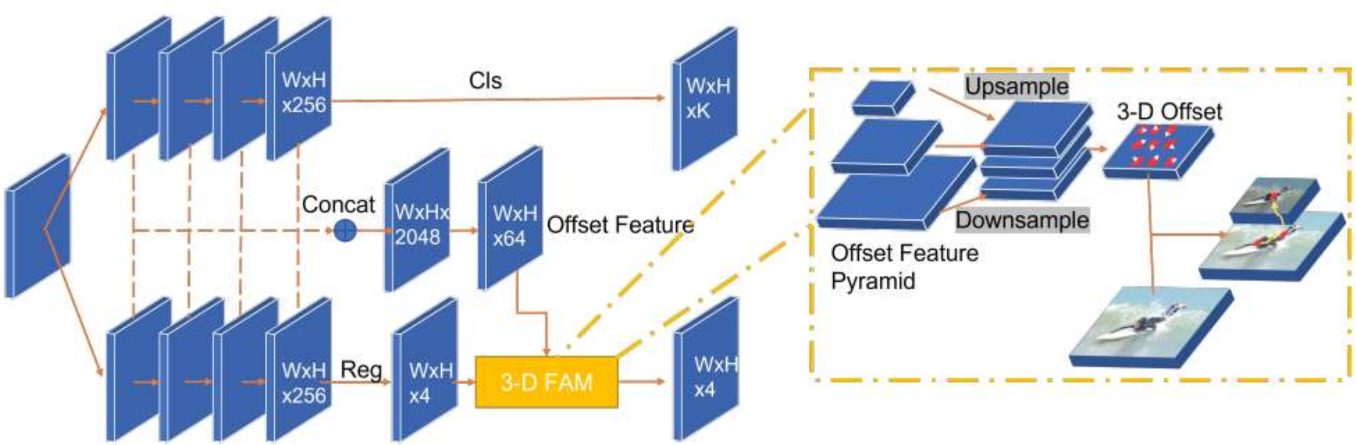
the probability that detection should be set as the pseudo-target for the student

$$\tau^c = \operatorname*{argmax}_{s^c} \mathcal{P}(pos | s^c, \mu_p^c, (\sigma_p^c)^2)$$

# Consistent Teacher



3D Feature Alignment




Adaptive Sample Assignment

$$\hat{c} = \underset{c}{\operatorname{argmin}} \mathcal{L}(f_t(\mathbf{x}^u), c)$$

$$\begin{aligned} C_{ij} =& \lambda_{cls} C_{cls} + \lambda_{reg} C_{reg} + \lambda_{dist} C_{dist} \\ \text{where } C_{cls} &= L_{cls}(\texttt{Pred}(p_i)_{cls}, b_i) \\ C_{reg} &= L_{reg}(\texttt{Pred}(p_i)_{reg}, b_i) \\ C_{dist}(i,j) &= 10^{\|\mathbf{d}(p_j, b_i)\|_2} \end{aligned}$$

模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

# Semi-DETR: Semi-Supervised Object Detection with Detection Transformers

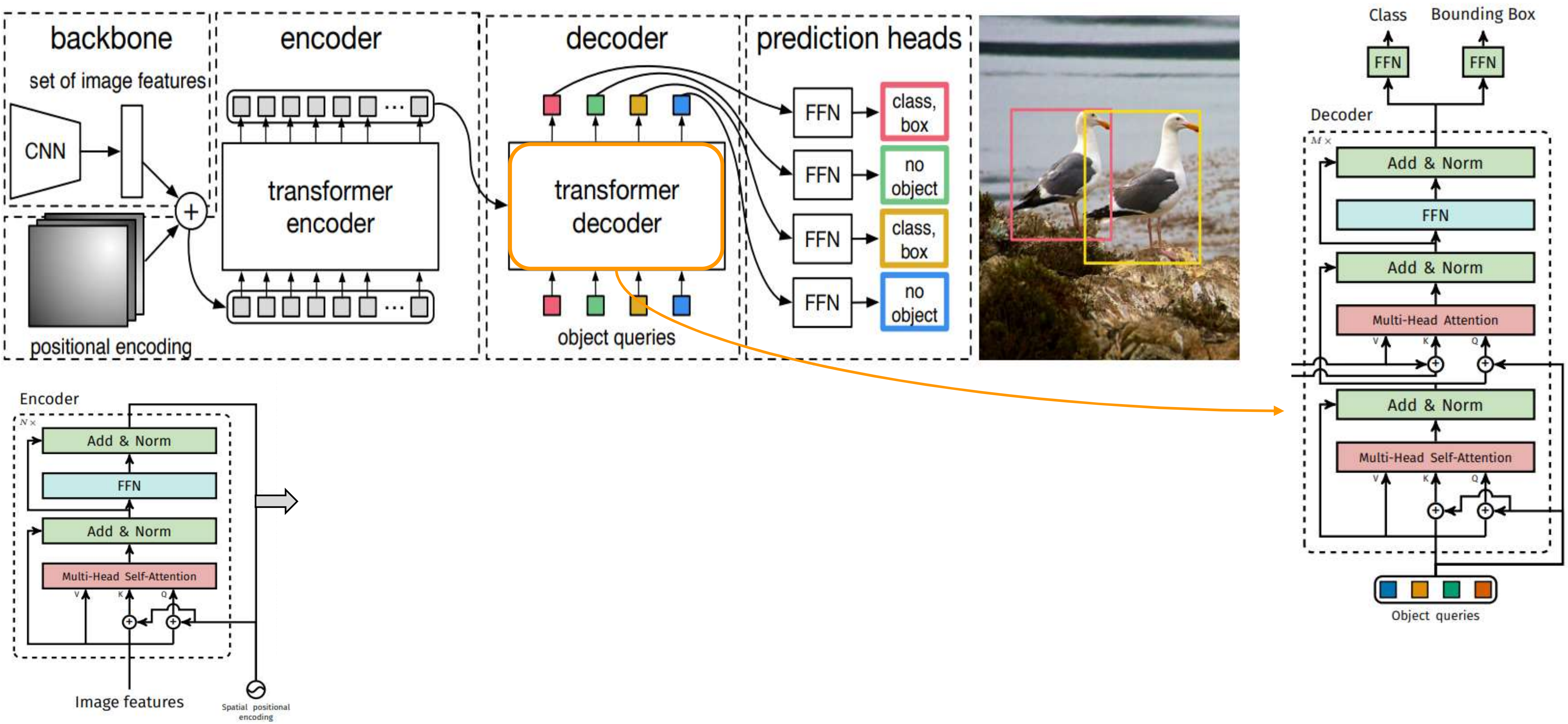
Xinjiang Wang[1*] Xingyi Yang[3*†] Shilong Zhang[2], Yijiang Li[1‡]
Litong Feng[1] Shijie Fang[4‡] Chengqi Lyu[2] Kai Chen[2] Wayne Zhang[1]
[1]SenseTime Research [2]Shanghai AI Laboratory [3]National University of Singapore [4]Peking University
wangxinjiang@sensetime.com, xyang@u.nus.edu

DETR

CVPR 2023

# DETR

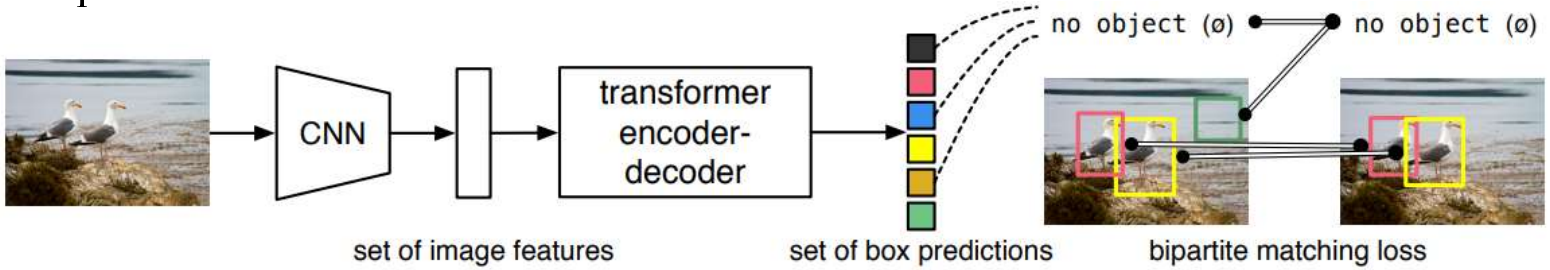





Carion N, Massa F, Synnaeve G, et al. End-to-end object detection with transformers[C]//European conference on computer vision. Cham: Springer International Publishing, 2020: 213-229.

# DETR

Object detection set prediction loss




Search for a permutation of N elements with the lowest cost:

$$\hat{\sigma} = \underset{\sigma \in \mathfrak{S}_N}{\arg\min} \sum_{i}^{N} \mathcal{L}_{\text{match}}(y_i, \hat{y}_{\sigma(i)})$$

A pair-wise matching cost between gt and prediction:

$$\mathcal{L}_{\text{match}}(y_i, \hat{y}_{\sigma(i)}) = -\mathbb{1}_{\{c_i \neq \varnothing\}} \hat{p}_{\sigma(i)}(c_i) + \mathbb{1}_{\{c_i \neq \varnothing\}} \mathcal{L}_{\text{box}}(b_i, \hat{b}_{\sigma(i)})$$

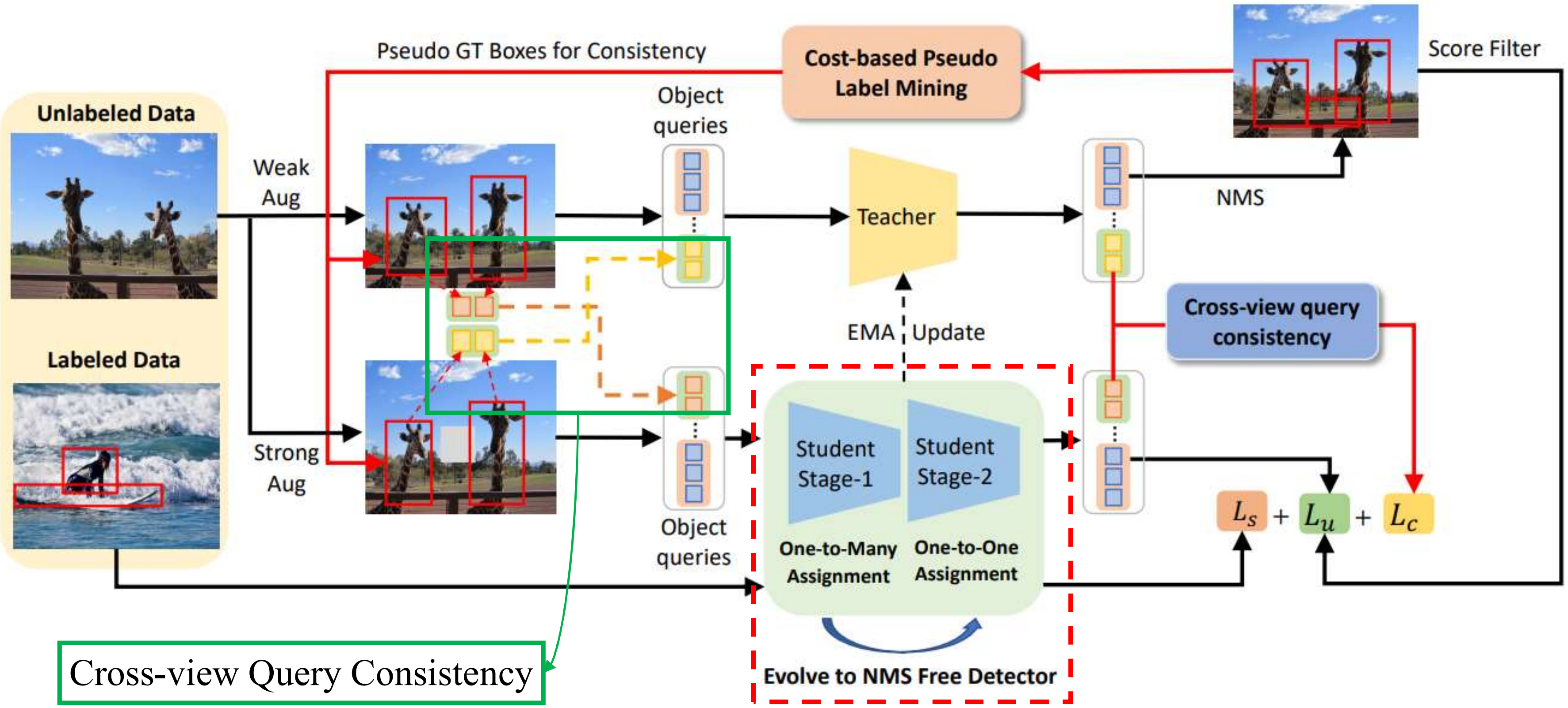
Find one-to-one matching for direct set prediction without duplicates (Hungarian loss):

$$\mathcal{L}_{\text{Hungarian}}(y, \hat{y}) = \sum_{i=1}^{N} \left[ -\log \hat{p}_{\hat{\sigma}(i)}(c_i) + \mathbb{1}_{\{c_i \neq \varnothing\}} \mathcal{L}_{\text{box}}(b_i, \hat{b}_{\hat{\sigma}}(i)) \right]$$

# Semi-DETR



$$C_{ij} = \lambda_1 C_{Cls}(p_i, \hat{p_j}) + \lambda_2 C_{GIoU}(b_i, \hat{b_j}) + \lambda_3 C_{L_1}(b_i, \hat{b_j})$$



# Semi-DETR



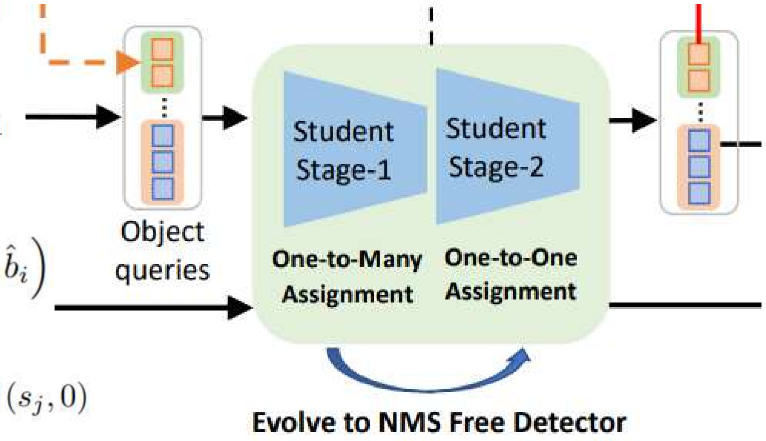
One-to-many assignment:

$$\hat{\sigma}_{o2m} = \left\{ \underset{\boldsymbol{\sigma_i} \in C_N^M}{\arg\min} \sum_{j=1}^{M} \mathcal{C}_{\text{match}} \left( \hat{y}_i^t, \hat{y}^s_{\boldsymbol{\sigma_i(j)}} \right) \right\}_{i=1}^{|\hat{y}^t|}$$

$$\mathcal{L}_{reg}^{o2m} = \sum_{i=1}^{N_{\text{pos}}} \hat{m}_i \mathcal{L}_{GIoU} \left( b_i, \hat{b}_i \right) + \sum_{i=1}^{N_{\text{pos}}} \hat{m}_i \mathcal{L}_{L_1} \left( b_i, \hat{b}_i \right)$$

$$\mathcal{L}_{cls}^{o2m} = \sum_{i=1}^{N_{\text{pos}}} |\hat{m}_i - s_i|^{\gamma} BCE\left(s_i, \hat{m}_i\right) + \sum_{j=1}^{N_{\text{neg}}} s_j^{\gamma} BCE\left(s_j, 0\right)$$

$$\mathcal{L}^{o2m} = \mathcal{L}_{cls}^{o2m} + \mathcal{L}_{reg}^{o2m}$$




One-to-one assignment:

$$\hat{\sigma}_{o2o} = \underset{\sigma \in \xi_N}{\arg\min} \sum_{i=1}^{N} \mathcal{C}_{\text{match}} \left( \hat{y}_i^t, \hat{y}^s_{\sigma(i)} \right)$$

# Semi-DETR

cross-view query embeddings:

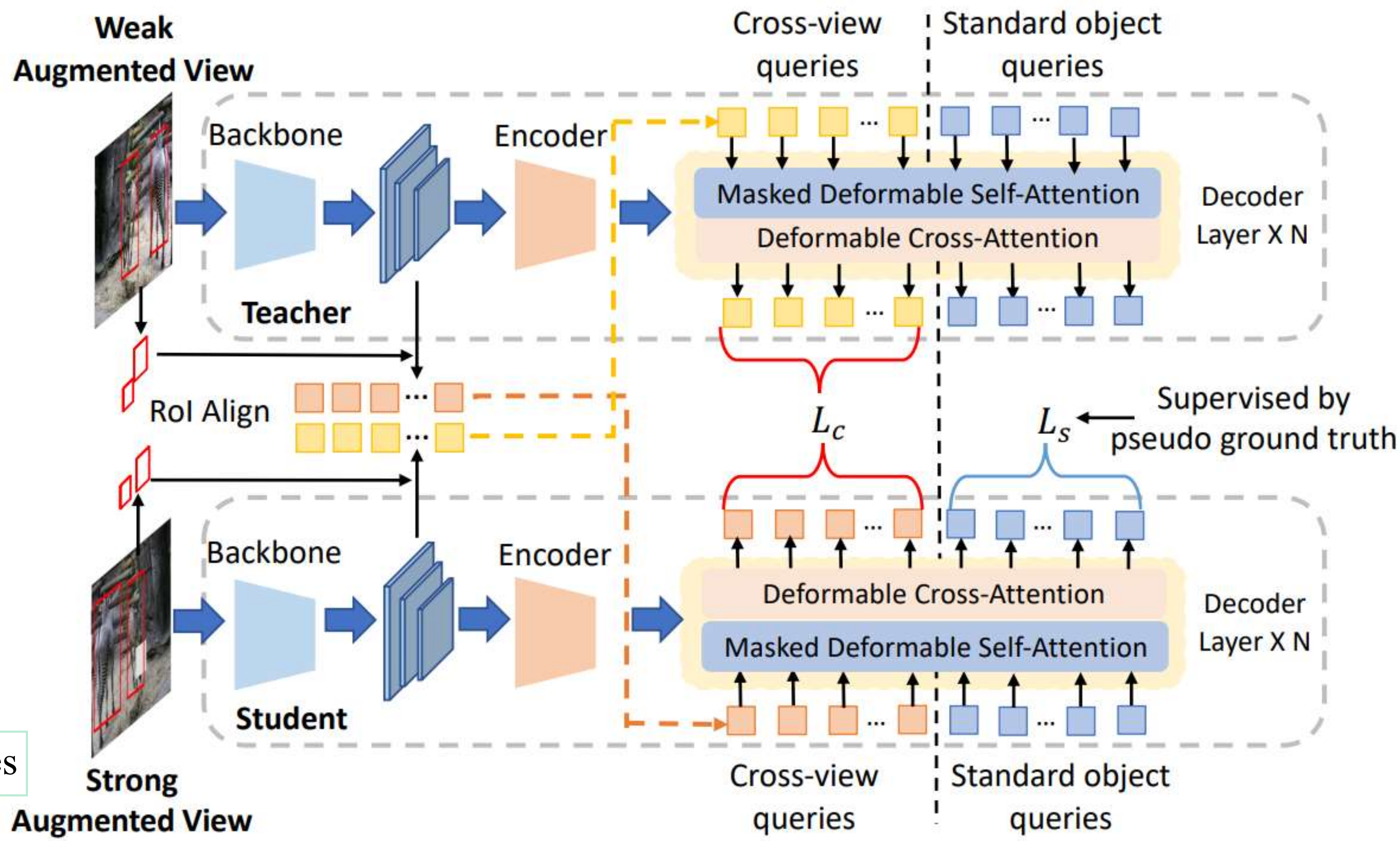
$$c_t = \mathrm{MLP}(\mathrm{RoIAlign}(F_t, b))$$
$$c_s = \mathrm{MLP}(\mathrm{RoIAlign}(F_s, b))$$

decoded features of standard (cross-view) queries

original object queries

$$\hat{o}_t, o_t = \mathrm{Decoder}_t([c_s, q_t], E_t|A)$$
$$\hat{o}_s, o_s = \mathrm{Decoder}_s([c_t, q_s], E_s|A)$$

encoded image features




consistency loss: $\mathcal{L}_c = \mathrm{MSE}(\hat{o}_s, \mathrm{detach}(\hat{o}_t))$

模式分析与机器智能
工业和信息化部重点实验室
MIIT Key Laboratory of
Pattern Analysis & Machine Intelligence

ParNeC 模式识别与神经计算研究组
PAttern Recognition and NEural Computing

# Mixed Pseudo Labels for Semi-Supervised Object Detection

Zeming Chen[1*] Wenwei Zhang[2,4] Xinjiang Wang[3] Kai Chen[4] Zhi Wang[1]

[1]Shenzhen International Graduate School, Tsinghua University

[2]S-Lab, Nanyang Technological University [3]SenseTime Research [4]Shanghai AI Laboratory

czm20@mails.tsinghua.edu.cn wenwei001@ntu.edu.sg wangxinjiang@sensetime.com

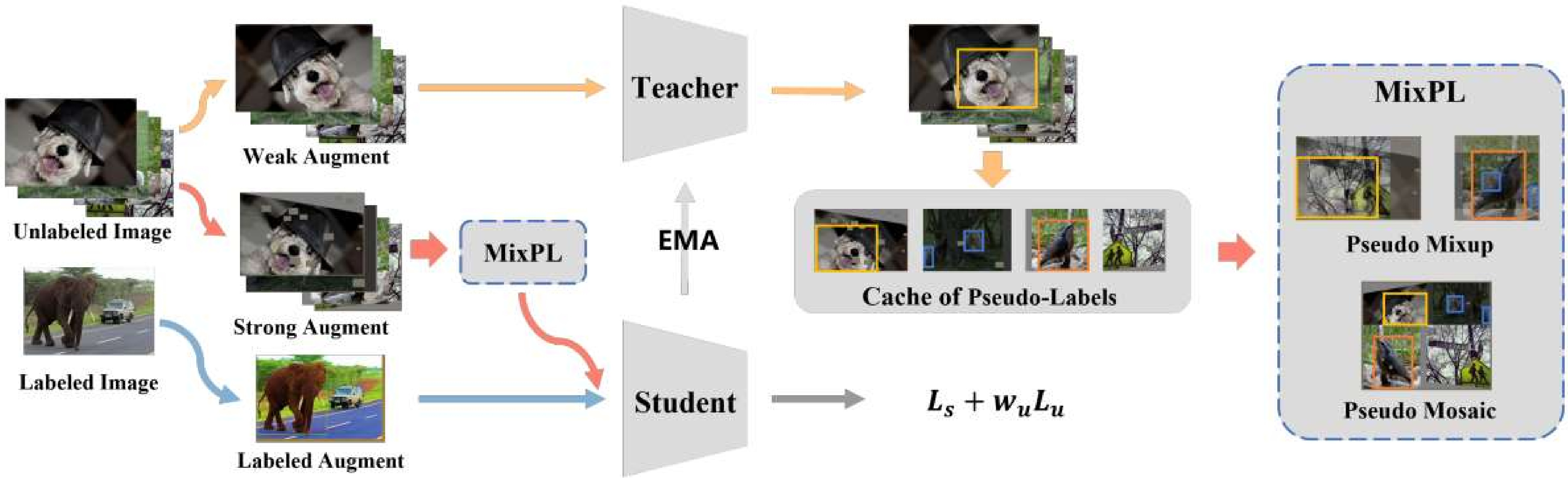
chenkai@pjlab.org.cn wangzhi@sz.tsinghua.edu.cn

DETR

Arxiv 2023.12.12

# MixPL




Unlabeled Image
Weak Augment
Strong Augment
Labeled Image
Labeled Augment
MixPL
Teacher
EMA
Student
Cache of Pseudo-Labels
$L_s + w_u L_u$
MixPL
Pseudo Mixup
Pseudo Mosaic

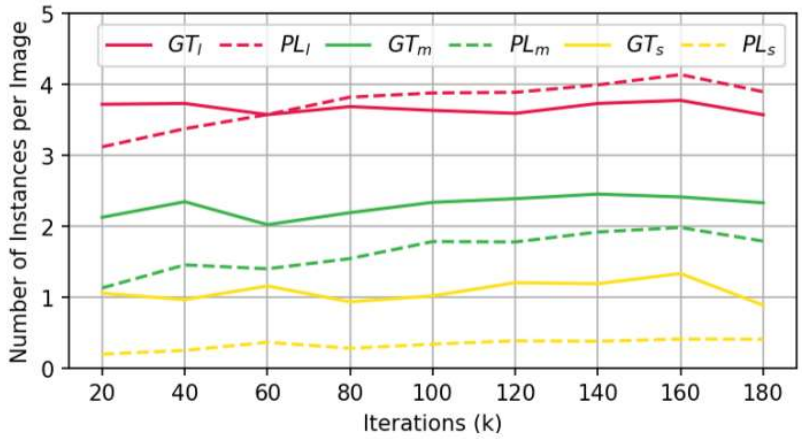
# MixPL

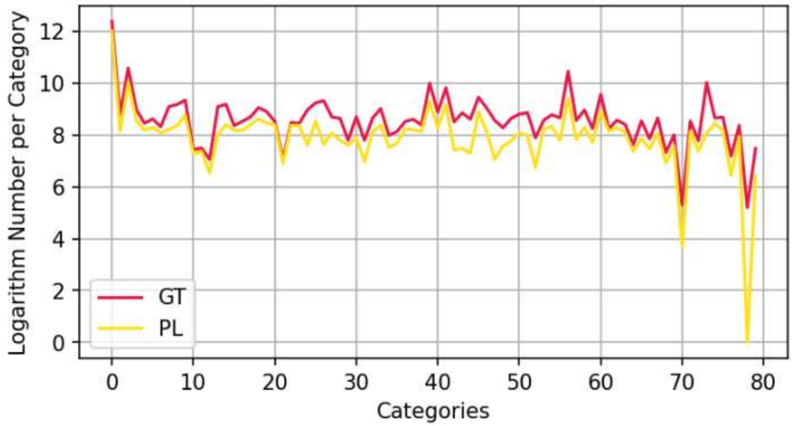
Why mosaic?

Imbalance : Scal

Large obj -> Medium obj
Medium obj -> Small obj

Labeled Resampling:
Oversampling of tail categories

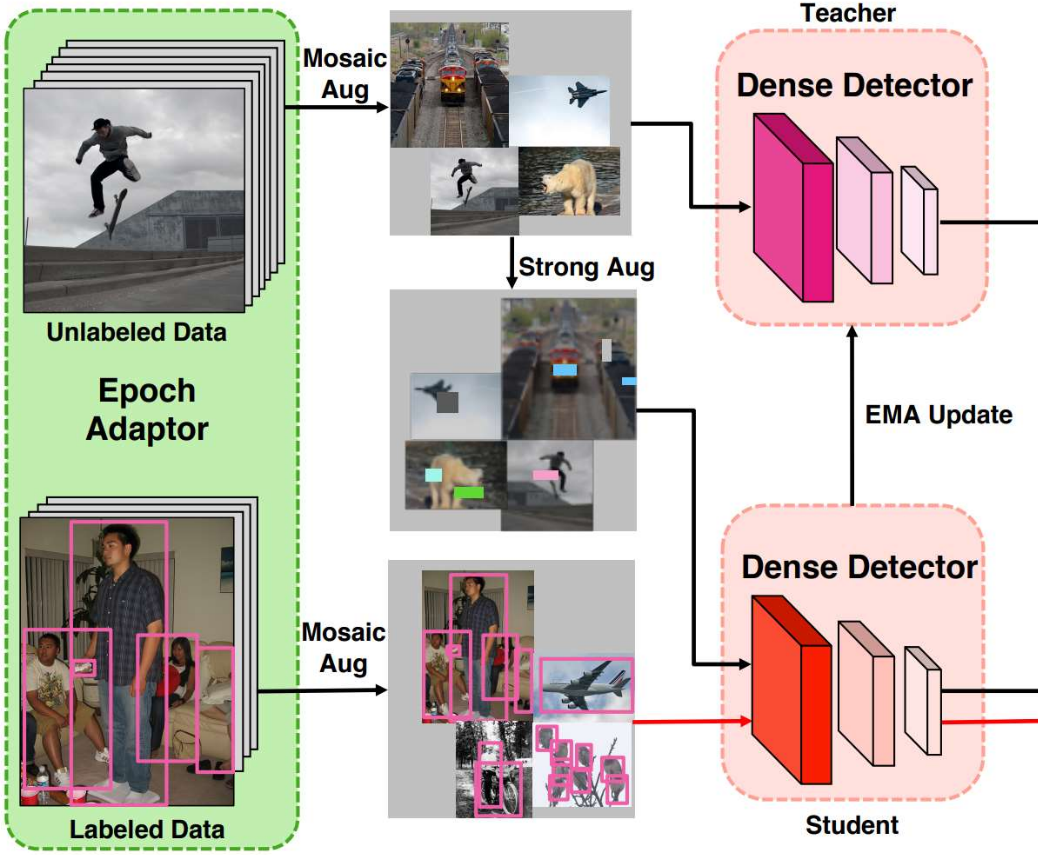


# Efficient Teacher (CVPR 2023)

ParNeC 模式识别与神经计算研究组 PAttern Recognition and NEural Computing

| Method | Resolution | Mosaic | Param. | FLOPs | $AP_{50:95}(\%)$ |
|---|---|---|---|---|---|
| Faster R-CNN [24] | [1333,800] | | 39.8M | 202.31G | 40.3 |
| FCOS [31] | [1333,800] | | 32.02M | 200.59G | 38.5 |
| YOLOv5 $w/o$ | [640,640] | | 46.56M | 109.59G | 41.2 |
| YOLOv5 [14] | [640,640] | ✓ | 46.56M | 109.59G | 49.0 |
| YOLOv7 [33] | [640,640] | ✓ | 37.62M | 106.59G | 51.5 |
| RetinaNet [19] | [1333,800] | | 37.74M | 239.32G | 39.5 |
| Dense Detector | [640,640] | ✓ | 42.13M | 169.61G | 44.86 |

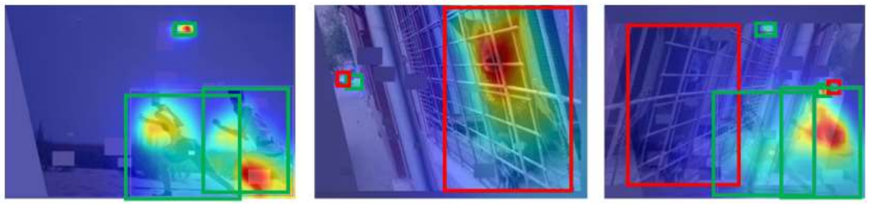


# MixPL

Why mixup?

| Detector | Filter | AP | $AP_{50}$ | $AP_{75}$ | $AP_s$ | $AP_m$ | $AP_l$ |
|---|---|---|---|---|---|---|---|
| Faster R-CNN | × | 34.7 | 54.6 | 37.6 | 19.5 | 37.1 | 46.1 |
| | ✓ | 34.7 | 54.7 | 37.4 | 19.3 | 37.6 | 45.8 |
| RetinaNet | × | 17.9 | 29.0 | 18.8 | 8.1 | 21.7 | 29.6 |
| | ✓ | 34.1 | 52.7 | 36.0 | 18.1 | 36.8 | 46.2 |

Effectiveness of filtering *empty images*.



(a) (b) (c)

Figure 4. Grad-CAM of different augmented images with pseudo-labels for Faster R-CNN on COCO 10%. (c) is the Pseudo Mixup image of (a) and (b). The green boxes indicate correct pseudo-labels and the red boxes indicate missed objects. Pseudo Mixup is effective in reducing the gradient response of missed objects and enhancing the gradient response of correct pseudo-labels.

# MixPL



| Method | 10% COCO |
|---|---|
| Labeled Only | 23.86 |
| CSD | 22.46 |
| STAC | 28.64 |
| Instant Teaching | 30.40 |
| Humble teacher | 31.61 |
| Unbiased Teacher | 31.50 |
| Soft Teacher | 34.04 |
| ACRST | 34.92 |
| PseCo | 36.06 |
| Labeled Only | 24.10 |
| Unbiased Teacher v2 | 32.61 |
| DSL | 36.22 |
| Dense Teacher | 37.13 |
| S4OD | 32.90 |
| Mean-Teacher | 35.50 |
| Consistent-Teacher | **40.00** |

| Detectors | Supervised | DetMeanTeacher | MixPL |
|---|---|---|---|
| RetinaNet | 27.2 | 34.1 (+6.9) | 36.0 (+1.9) |
| CenterNet | 27.7 | 34.9 (+7.2) | 36.7 (+1.8) |
| FCOS | 26.9 | 35.9 (+9.0) | 37.3 (+1.4) |
| ATSS | 28.2 | 35.5 (+7.3) | 37.0 (+1.5) |
| TOOD | 29.5 | 38.2 (+8.7) | 38.9 (+0.7) |
| Faster R-CNN | 26.6 | 34.7 (+8.1) | 37.2 (+2.5) |
| Cascade R-CNN | 28.0 | 37.3 (+9.3) | 40.0 (+2.7) |
| Sparse R-CNN | 29.3 | 36.8 (+7.5) | 38.2 (+1.4) |
| Deformable DETR | 31.3 | 39.3 (+8.0) | 40.5 (+1.2) |
| DAB DETR | 27.5 | 33.7 (+6.2) | 36.2 (+2.5) |
| DINO | 35.7 | 43.2 (+7.5) | 44.4 (+1.2) |

Table 9. Improvements on various detectors on COCO 10%.

# Conclusion

1. Unbiased Teacher constructed the basic framework of SSOD based on teacher-student.
2. Soft-teacher, building on Unbiased Teacher, established the paradigm of two-stage SSOD.
3. Consistent Teacher and Efficient Teacher were the first to apply one-stage methods in SSOD, addressing the challenge of obtaining high-quality pseudo-labels directly from dense predictions.
4. Semi-DETR was the first to incorporate DETR-based approaches into SSOD, with DETR-based frameworks being well-suited for solving SSOD problems.

**Foreground-background imbalance**: Focal loss, Mosaic aug, Mixup aug …

**Pseudo Label Inconsistency**:
EMA(Unbiased teacher), FAM3D(consistent teacher), Box Jettering(soft teacher), Dense Detector(efficient teacher), Cross-view query consistency method(Semi-DETR) …

**Assignment**: adaptive anchor assignment(consistent teacher) …

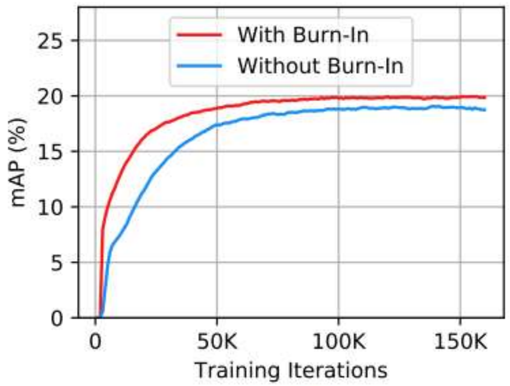
# Conclusion



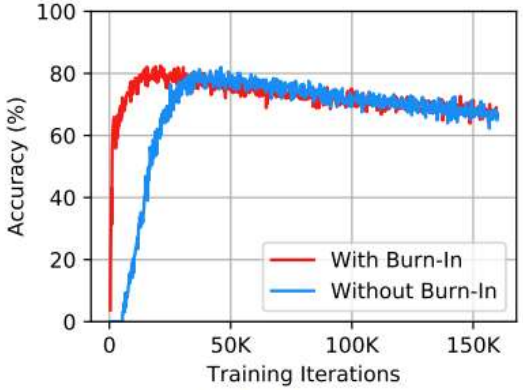
| | | | | | | |
|---|---|---|---|---|---|---|
| **Performance**: | DETR | > | Two-stage | > | One-stage | |
| **Training Cost**: | DETR | > > | Two-stage | > | One-stage | |
| **Speed**: | One-stage | > | Two-stage | > | DETR | |
| **Flexibility**: | Two-stage | > | DETR | ≈ | One-stage | |

# Conclusion

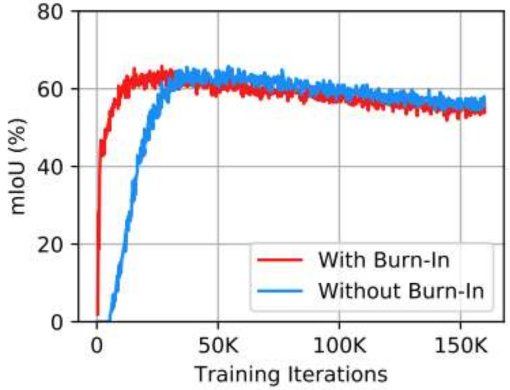
Burn-in stage



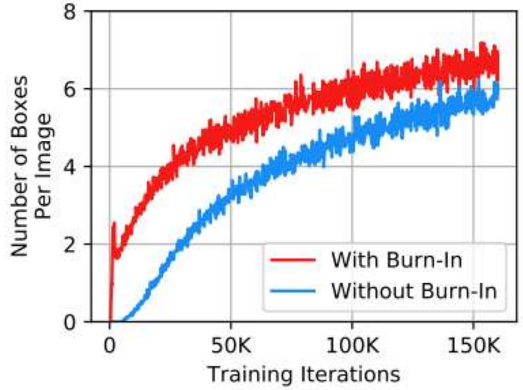
(a) mAP



(b) Box Accuracy



(c) mIoU



(d) Number of Pseudo-Boxes

Unbiased Teacher (two stage)