Distilling Image Classifiers in Object Detectors

Shuxuan Guo, Jose M. Alvarez, Mathieu Salzmann
CVLab, EPFL & NVIDIA

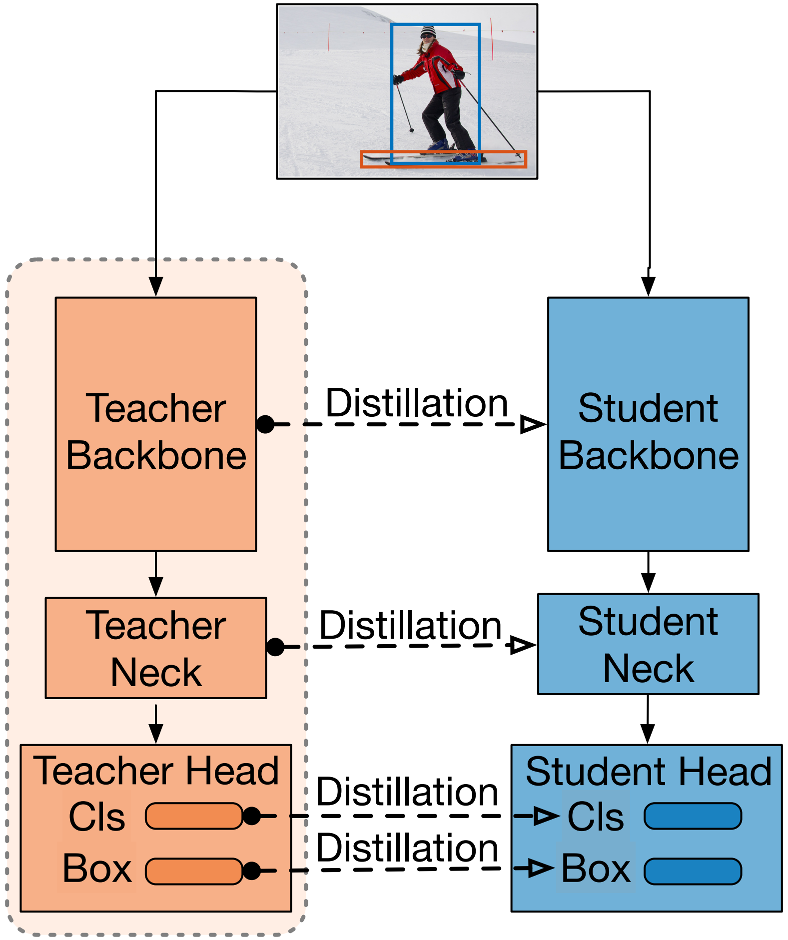
NeurIPS 2021

Knowledge Distillation in Object Detection

- Compact object detectors
  - One-stage methods
    SSD, YOLO …
  - Two-stage methods
    RCNN family with lightweight backbones

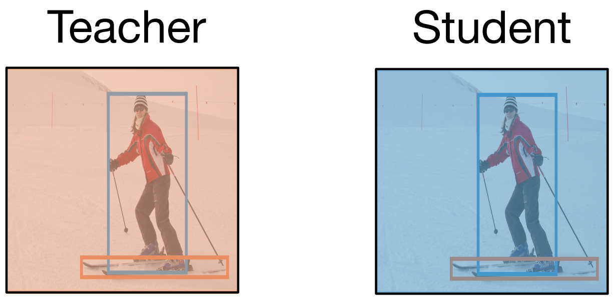
Knowledge Distillation in Object Detection

- Detector-to-detector knowledge distillation



Knowledge Distillation in Object Detection

- Detector-to-detector knowledge distillation

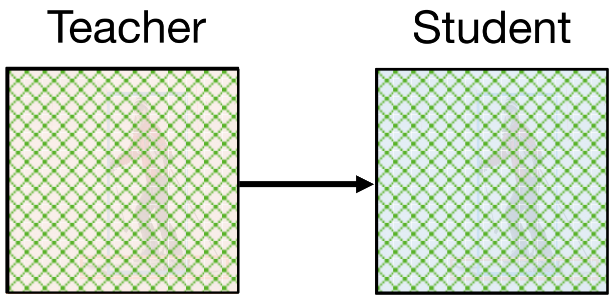


(a) global feature adaptation

Chen et al. NeurIPS2017

Knowledge Distillation in Object Detection

- Detector-to-detector knowledge distillation

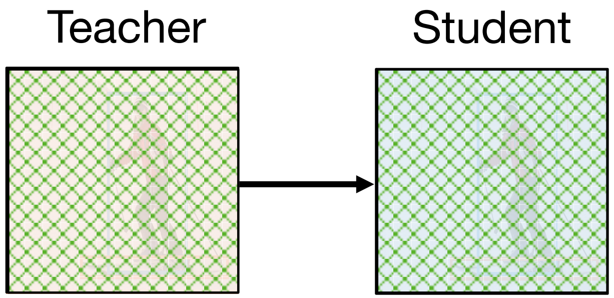


(a) global feature adaptation

Chen et al. NeurIPS2017

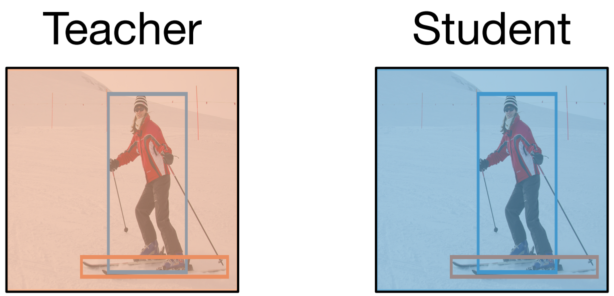
Knowledge Distillation in Object Detection

- Detector-to-detector knowledge distillation



(a) global feature adaptation

Chen et al. NeurIPS2017

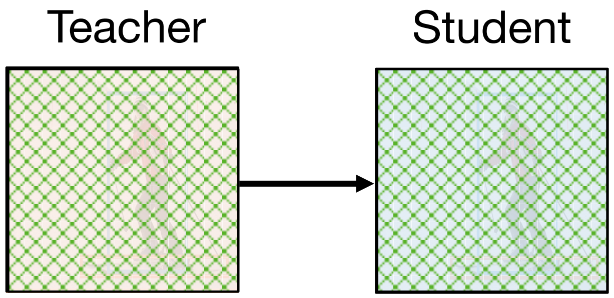


(b) positive feature imitation

Wang et al. CVPR2019

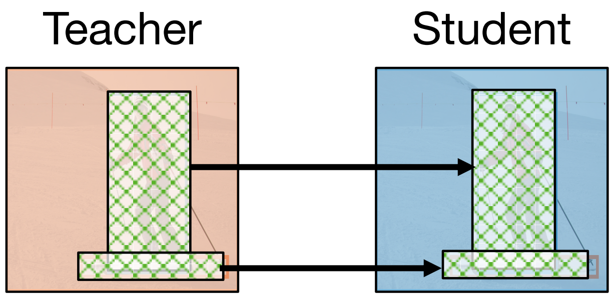
Knowledge Distillation in Object Detection

- Detector-to-detector knowledge distillation



(a) global feature adaptation

Chen et al. NeurIPS2017

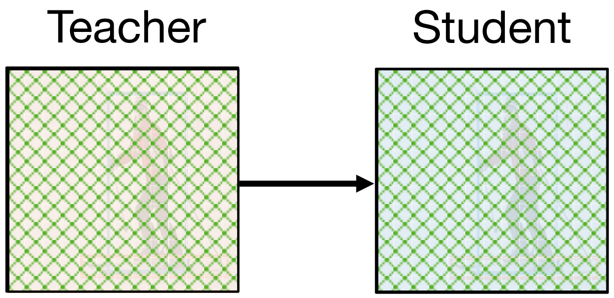


(b) positive feature imitation

Wang et al. CVPR2019

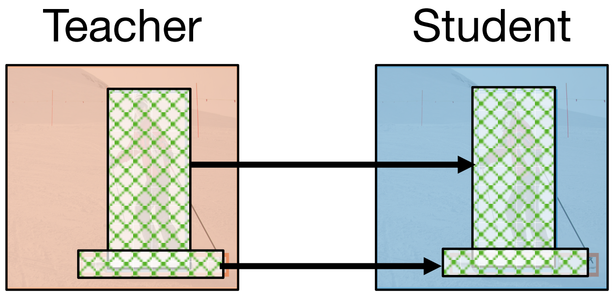
Knowledge Distillation in Object Detection

- Detector-to-detector knowledge distillation



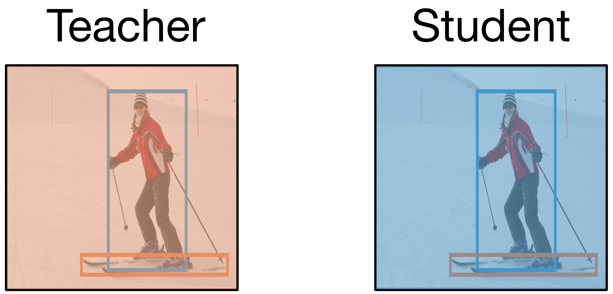
(a) global feature adaptation

Chen et al. NeurIPS2017



(b) positive feature imitation

Wang et al. CVPR2019

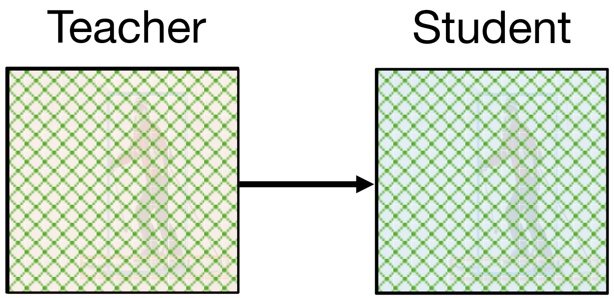


(c) attention-based feature

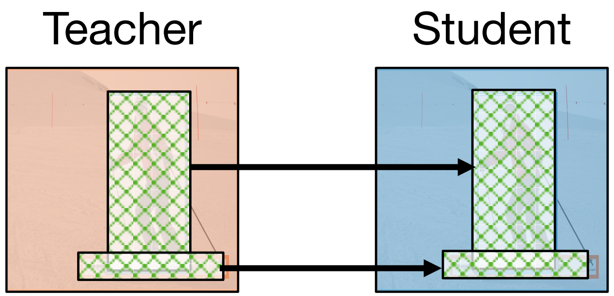
Zhang et al. ICLR2021

Knowledge Distillation in Object Detection

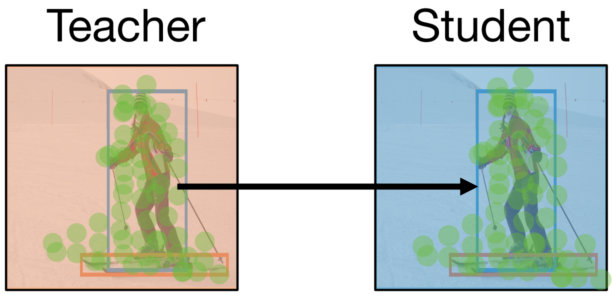
- Detector-to-detector knowledge distillation



(a) global feature adaptation
Chen et al. NeurIPS2017



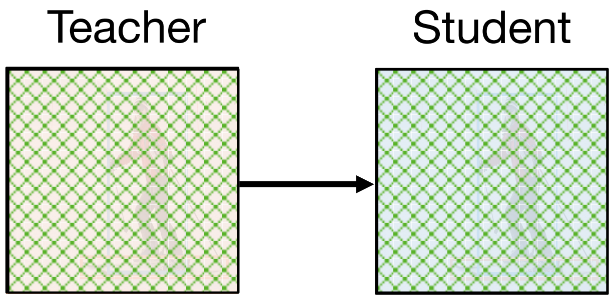
(b) positive feature imitation
Wang et al. CVPR2019



(c) attention-based feature
Zhang et al. ICLR2021

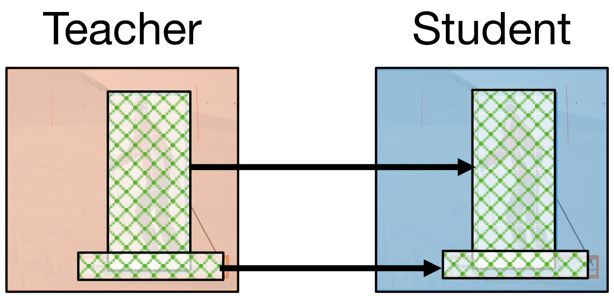
Knowledge Distillation in Object Detection

- Detector-to-detector knowledge distillation



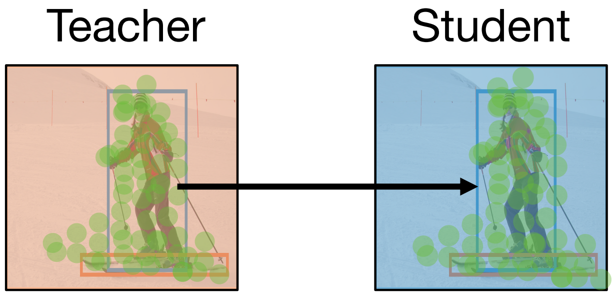
(a) global feature adaptation

Chen et al. NeurIPS2017



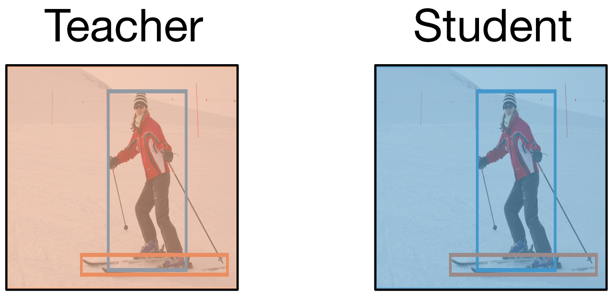
(b) positive feature imitation

Wang et al. CVPR2019



(c) attention-based feature

Zhang et al. ICLR2021

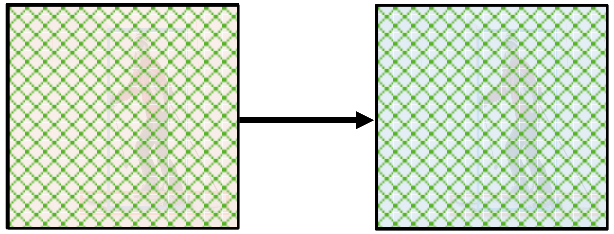


(d) decouple pos and neg feature

Guo et al. CVPR2021

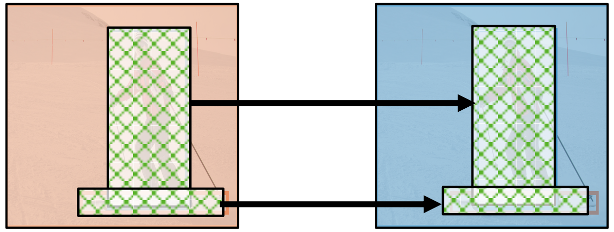
Knowledge Distillation in Object Detection

- Detector-to-detector knowledge distillation



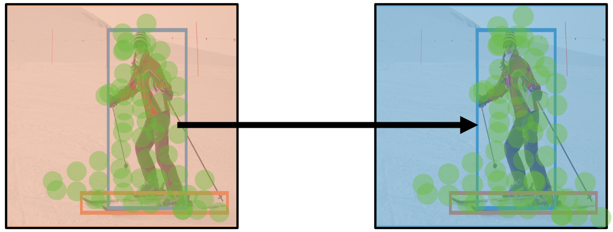
(a) global feature adaptation

Chen et al. NeurIPS2017



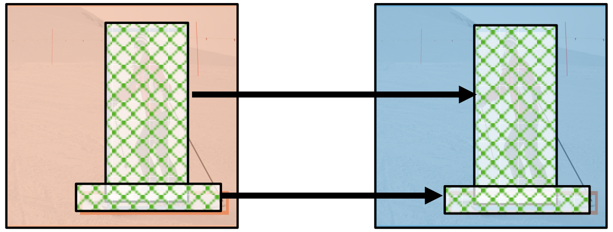
(b) positive feature imitation

Wang et al. CVPR2019



(c) attention-based feature

Zhang et al. ICLR2021

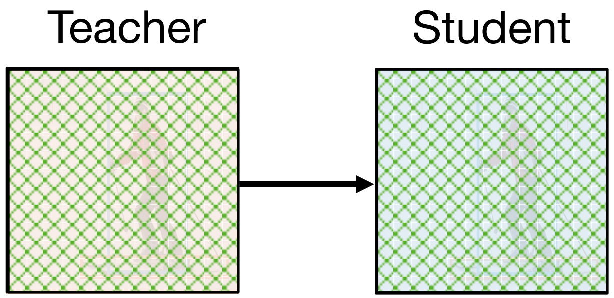


(d) decouple pos and neg feature

Guo et al. CVPR2021

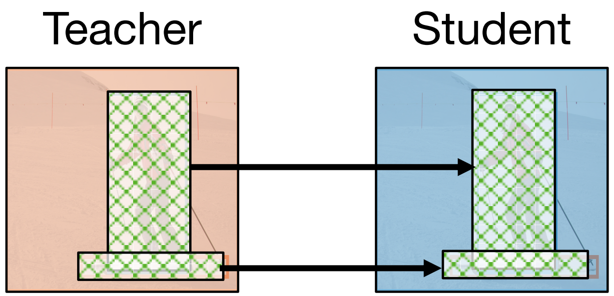
Knowledge Distillation in Object Detection

- Detector-to-detector knowledge distillation



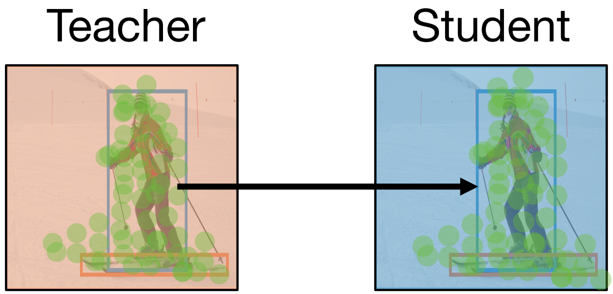
(a) global feature adaptation

Chen et al. NeurIPS2017



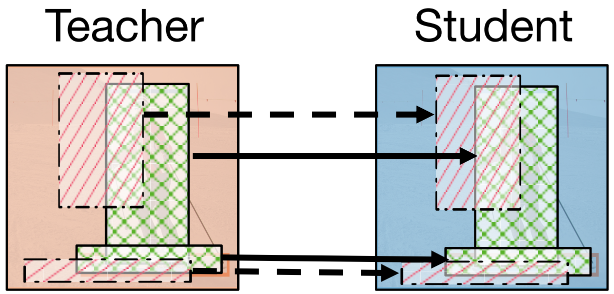
(b) positive feature imitation

Wang et al. CVPR2019



(c) attention-based feature

Zhang et al. ICLR2021

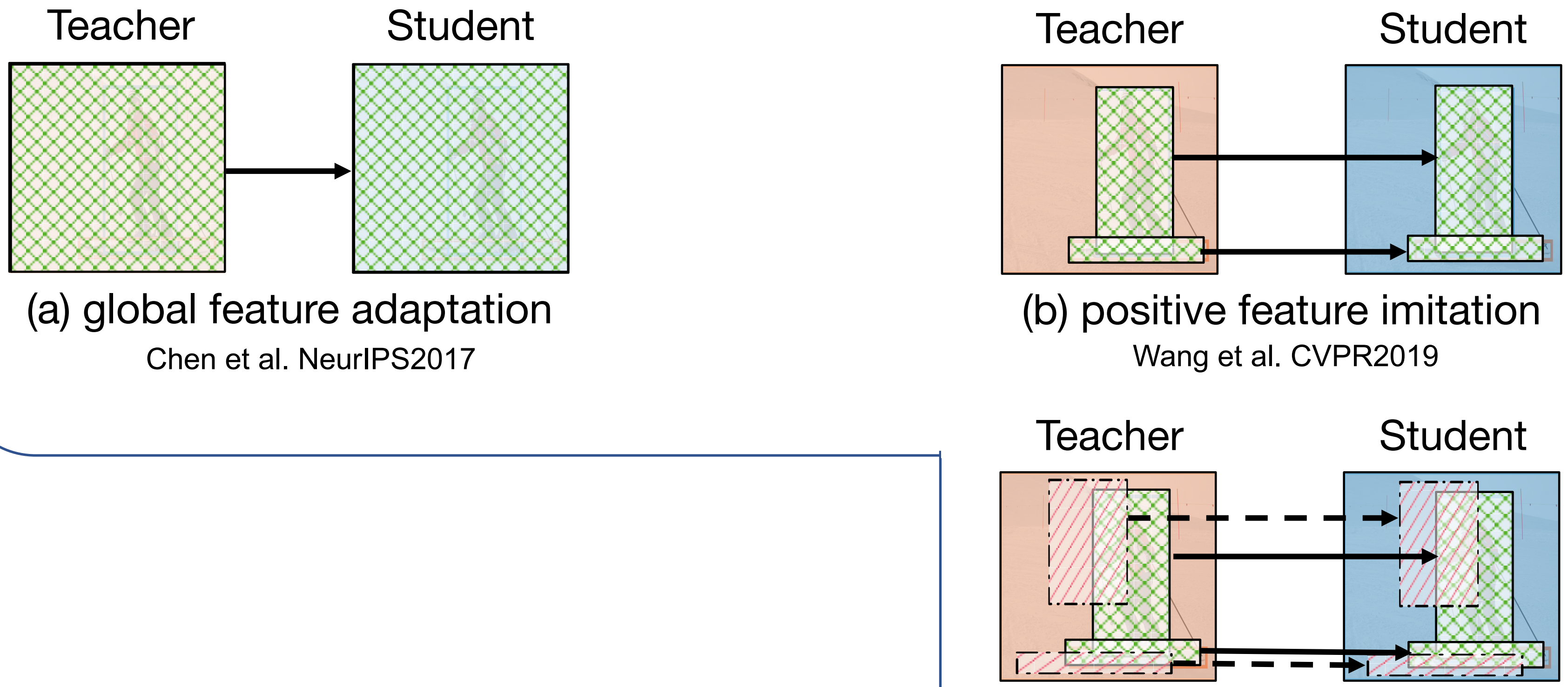


(d) decouple pos and neg feature

Guo et al. CVPR2021

Knowledge Distillation in Object Detection

❖ Detector-to-detector knowledge distillation



(a) global feature adaptation
Chen et al. NeurIPS2017

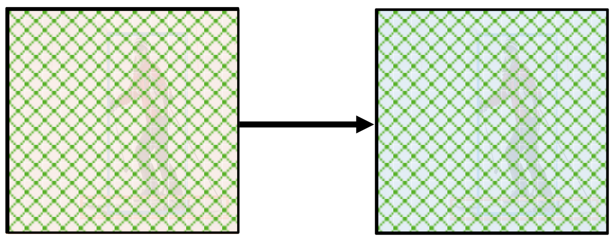
(b) positive feature imitation
Wang et al. CVPR2019

(d) decouple pos and neg feature
Guo et al. CVPR2021

❑ Require same kind of detection framework

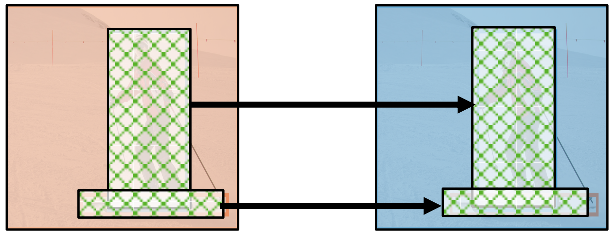
Knowledge Distillation in Object Detection

- Detector-to-detector knowledge distillation



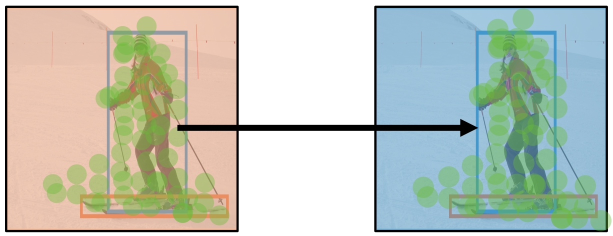
(a) global feature adaptation

Chen et al. NeurIPS2017



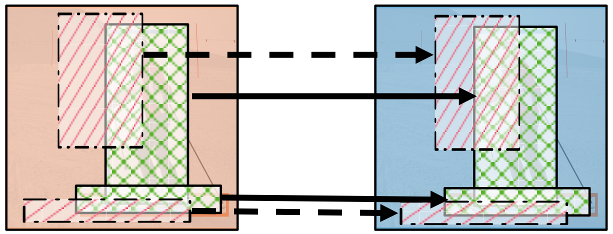
(b) positive feature imitation

Wang et al. CVPR2019



(c) attention-based feature

Zhang et al. ICLR2021



(d) decouple pos and neg feature

Guo et al. CVPR2021

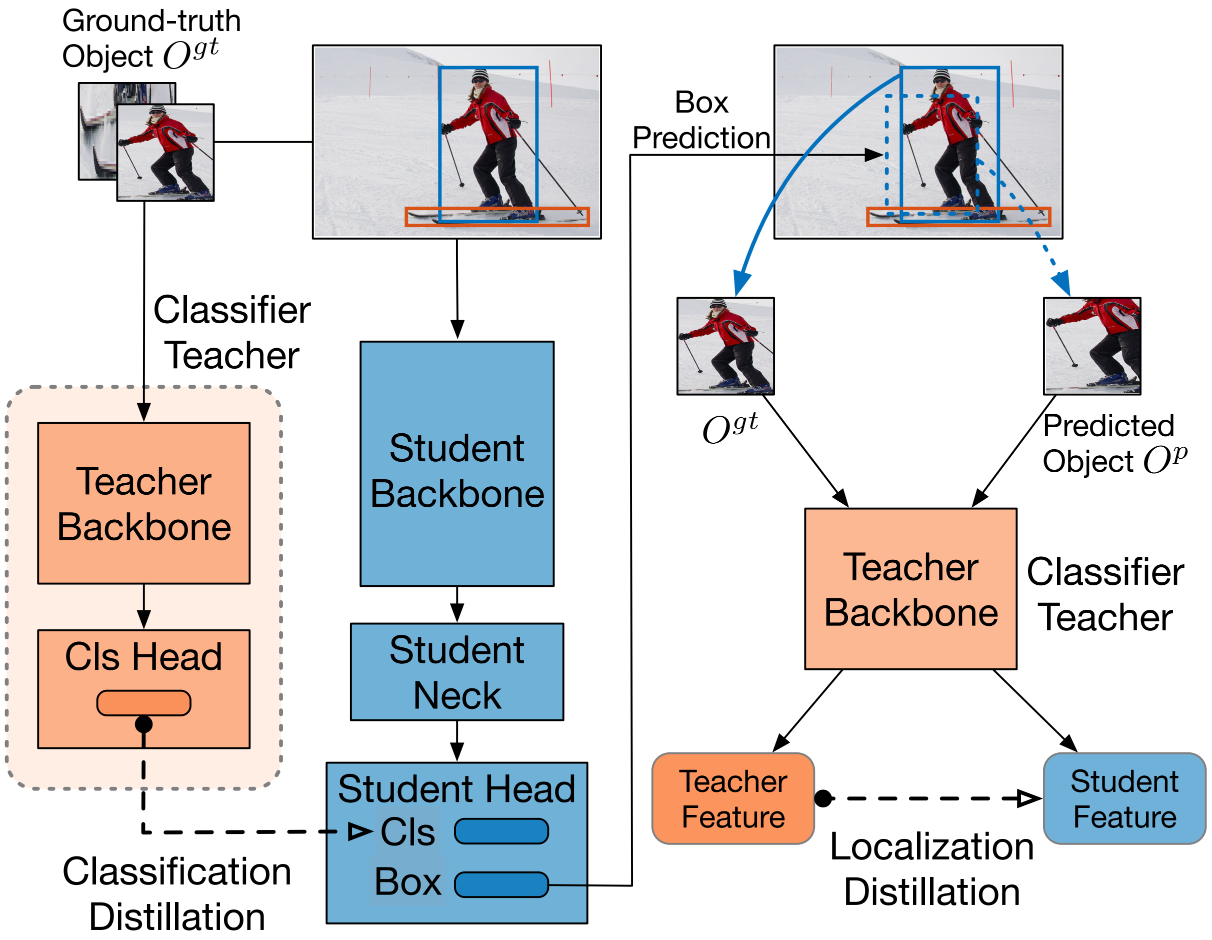
- Feature only, across architectures
- Require same kind of detection framework

Motivation

- Inferior performance of the detection classification head
  - Foreground-background classes imbalance

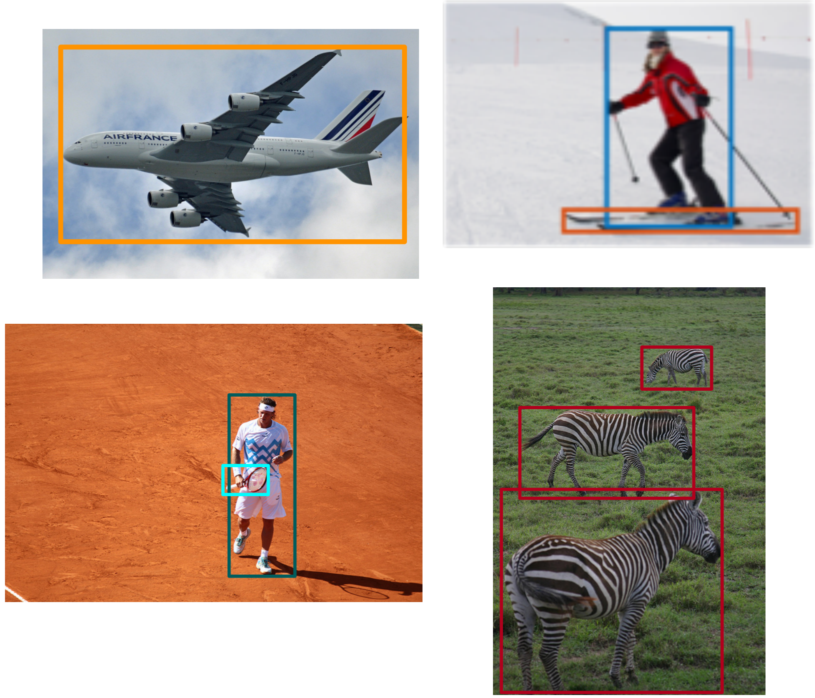
- Localization error is one of the key errors for the compact detection models

Our Method: Classifier Teacher to Detector Student



Our Method: Classifier Teachers

- Dataset



$$\mathcal{D}_{det} = \{class\_labels, bboxes\}$$

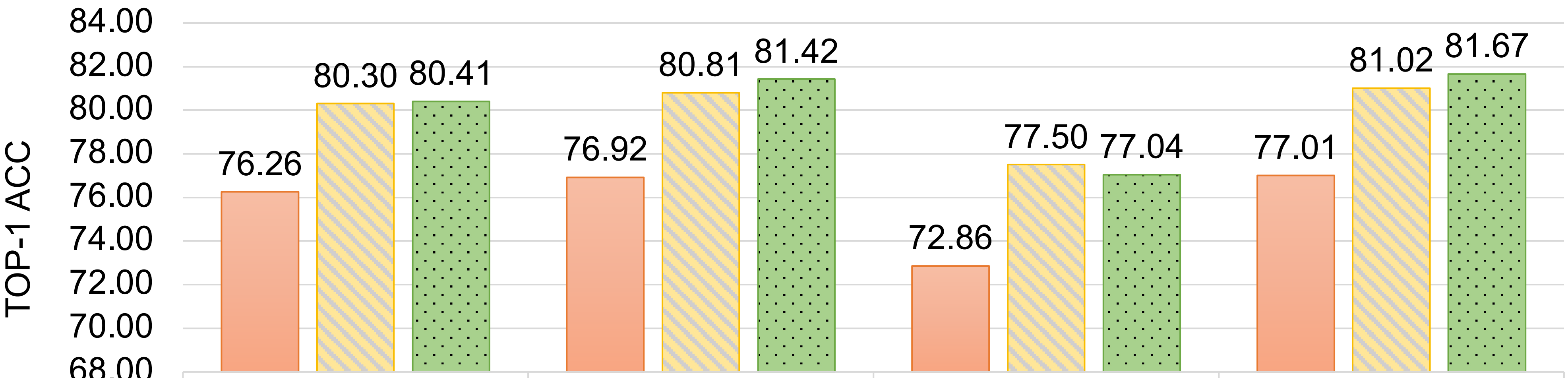
for each image

$$\mathcal{D}_{cls} = \{class\_label\}$$

for each object from all images

Our Method: Classifier Teachers

❖ Training



	SSD + CEL	general + CEL	general + FL	general + CEL + FL
56×56	76.26	76.92	72.86	77.01
112×112	80.30	80.81	77.50	81.02
224×224	80.41	81.42	77.04	81.67

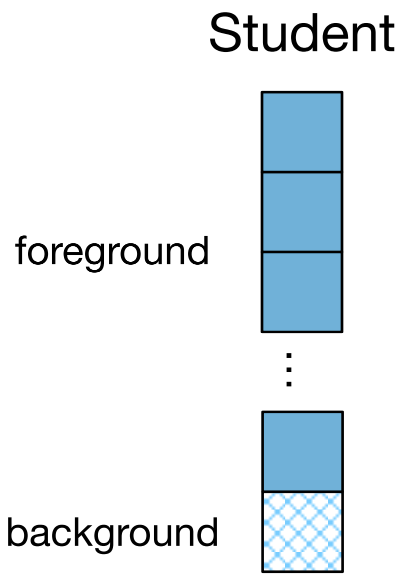
* Data augmentation: SSD -- data augmentation for SSD; general -- data augmentation for Faster RCNN and RetinaNet.
** Training loss: CEL -- cross-entropy loss; FL -- focal loss.

- ❑ The same classification teacher is used for all two-stage Faster RCNNs and one-stage RetinaNets in our classifier-to-detector distillation method.

Our Method: Knowledge Distillation for Classification

- Categorical cross-entropy loss

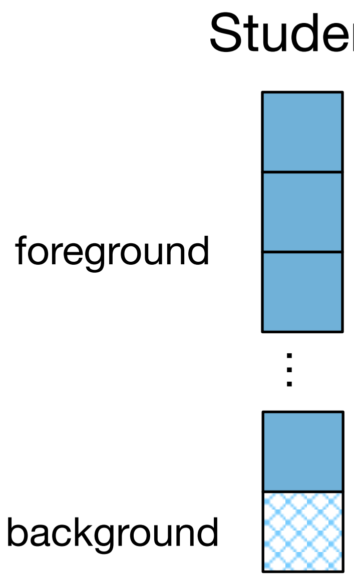
softmax probability



Our Method: Knowledge Distillation for Classification

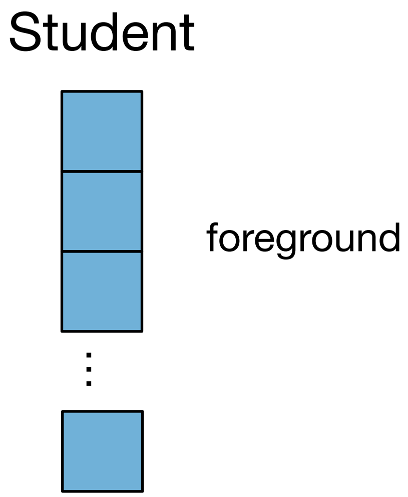
- Categorical cross-entropy loss

softmax probability



- Binary cross-entropy loss

sigmoid probability



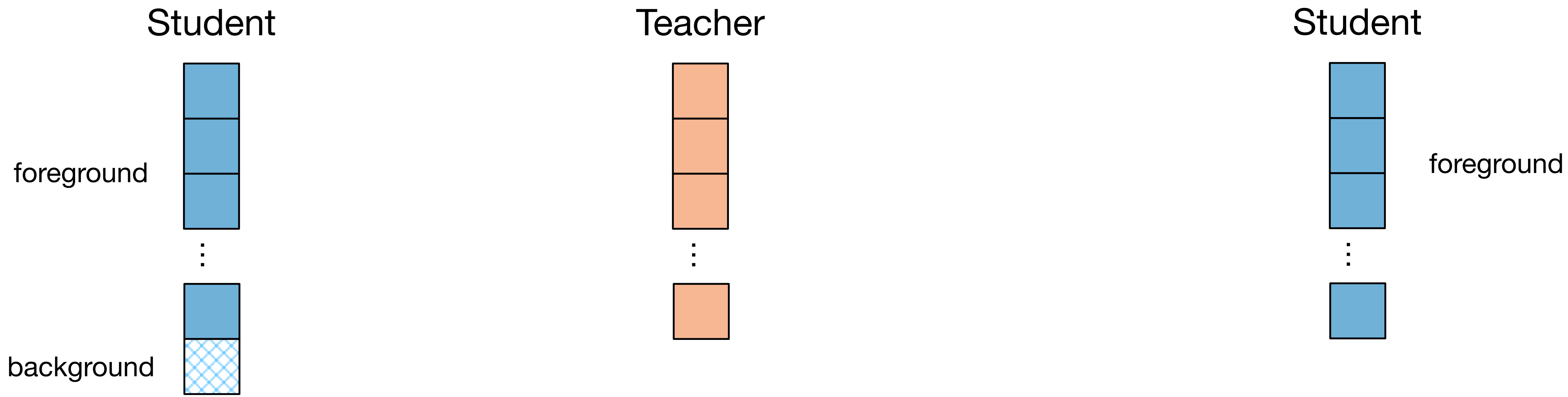
Our Method: Knowledge Distillation for Classification

- Categorical cross-entropy loss

  softmax probability

- Binary cross-entropy loss

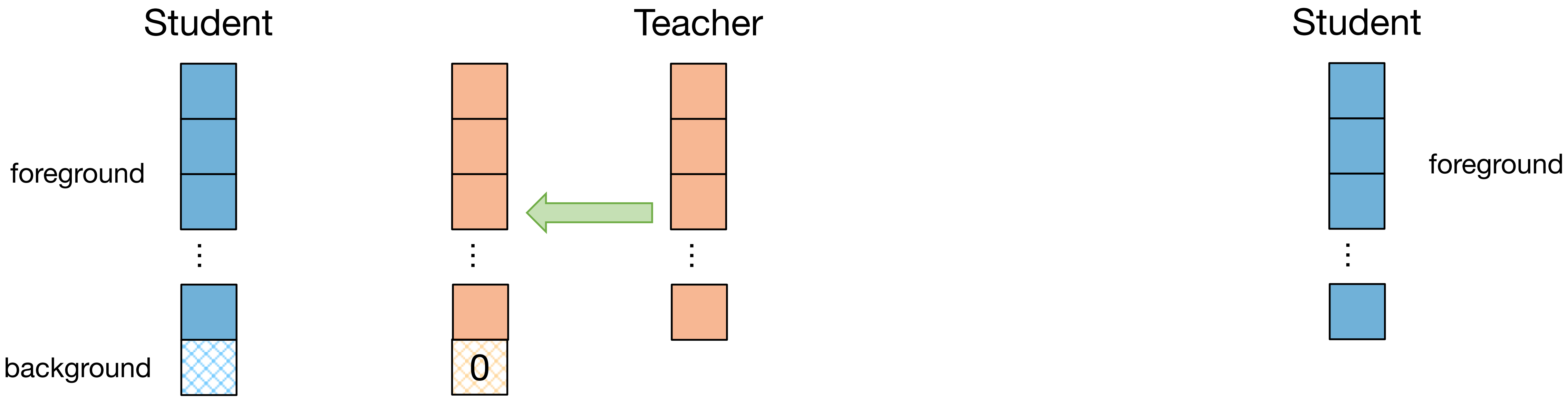
  sigmoid probability



Our Method: Knowledge Distillation for Classification

- Categorical cross-entropy loss
  softmax probability

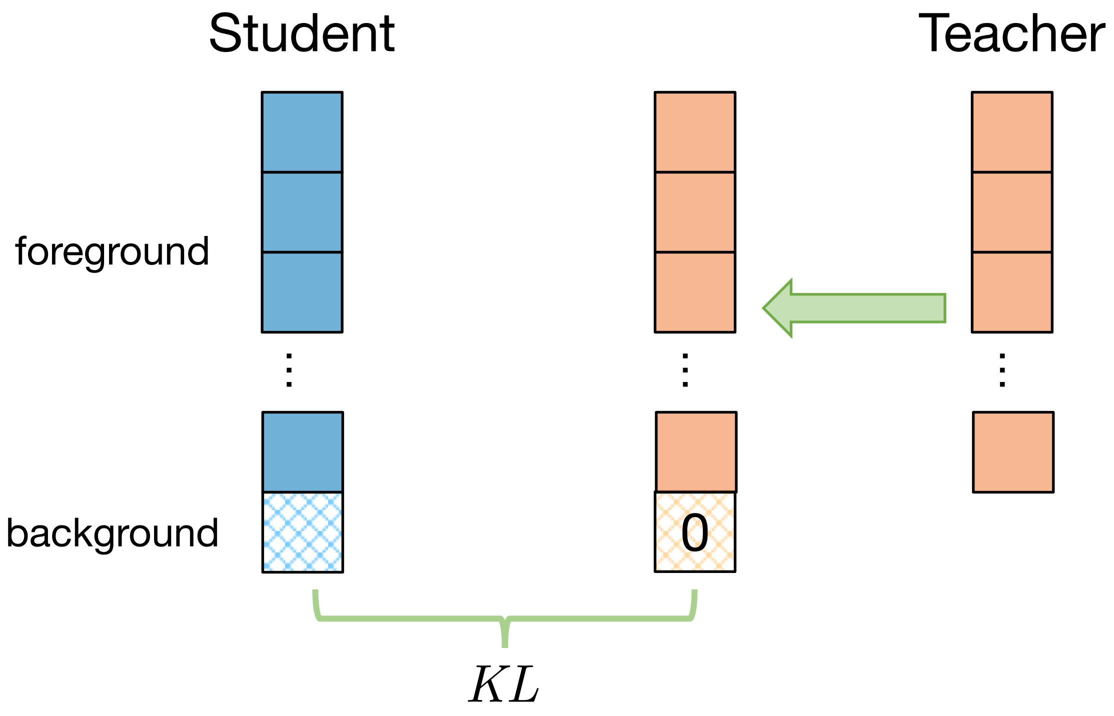
- Binary cross-entropy loss
  sigmoid probability



Our Method: Knowledge Distillation for Classification

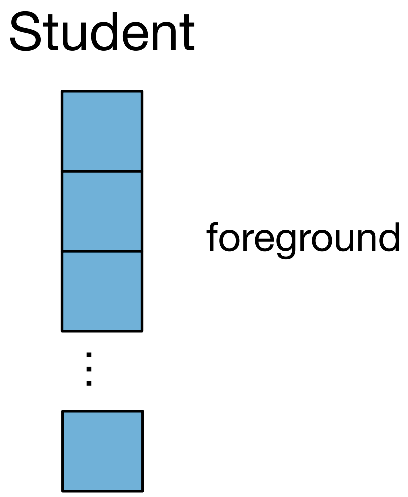
- Categorical cross-entropy loss

softmax probability



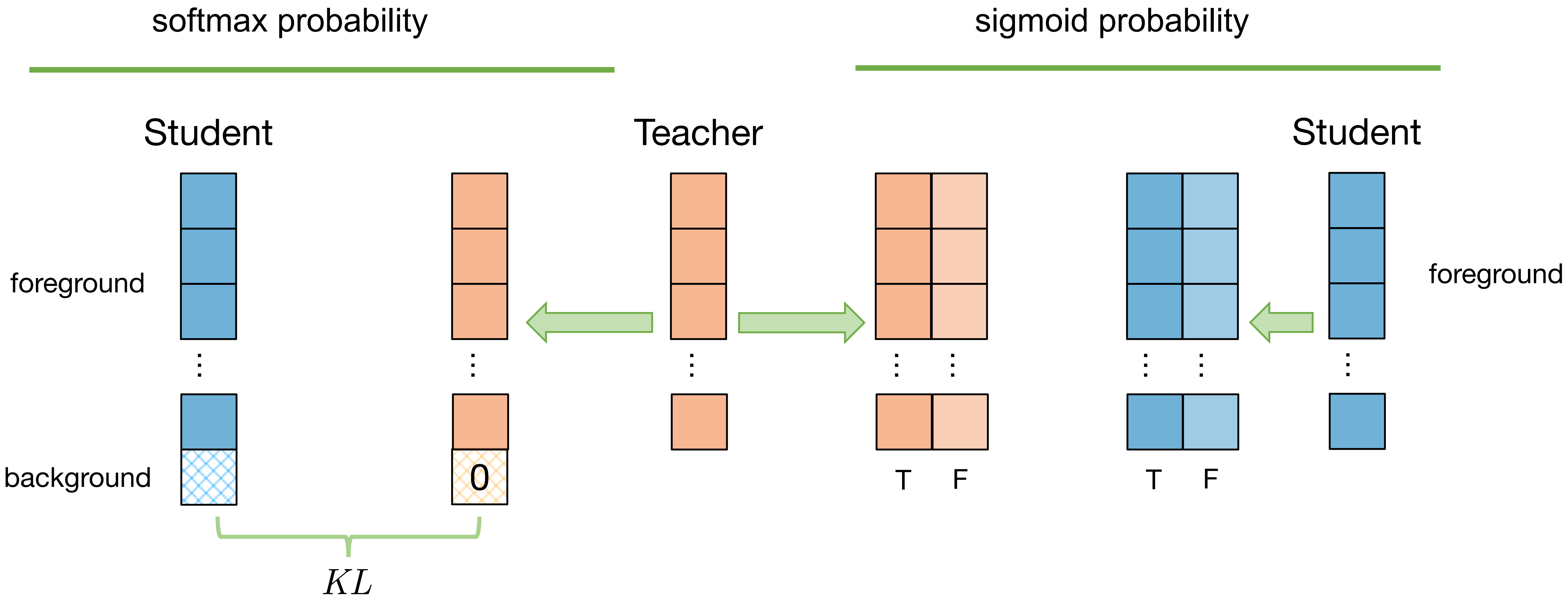
- Binary cross-entropy loss

sigmoid probability



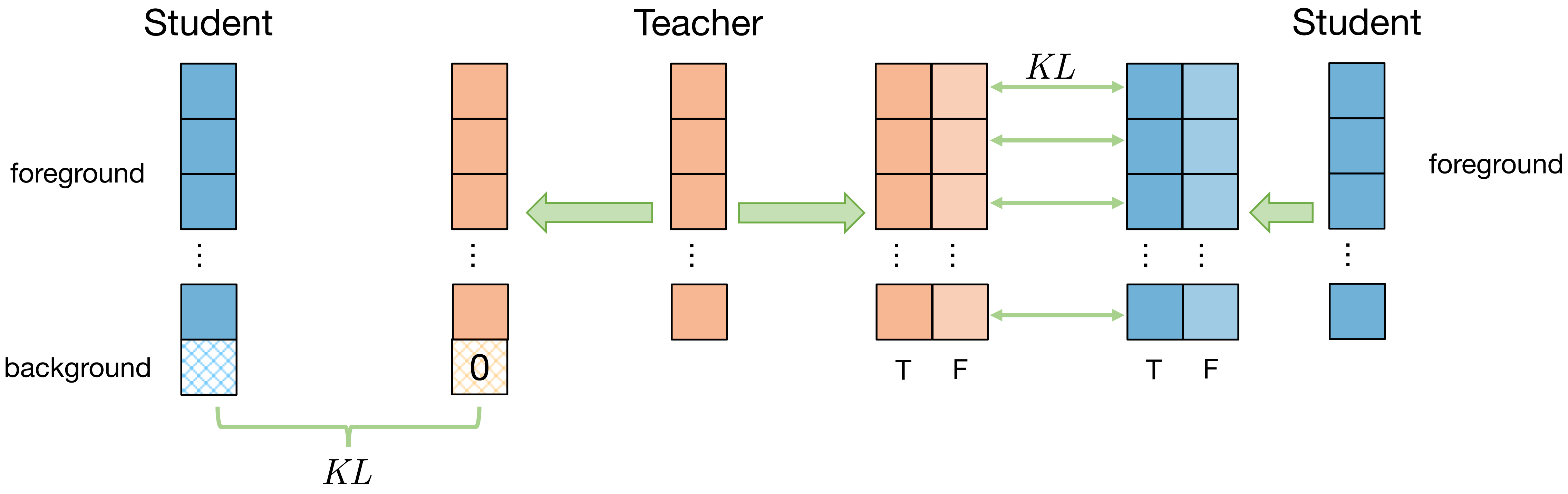
Our Method: Knowledge Distillation for Classification

- Categorical cross-entropy loss
- Binary cross-entropy loss

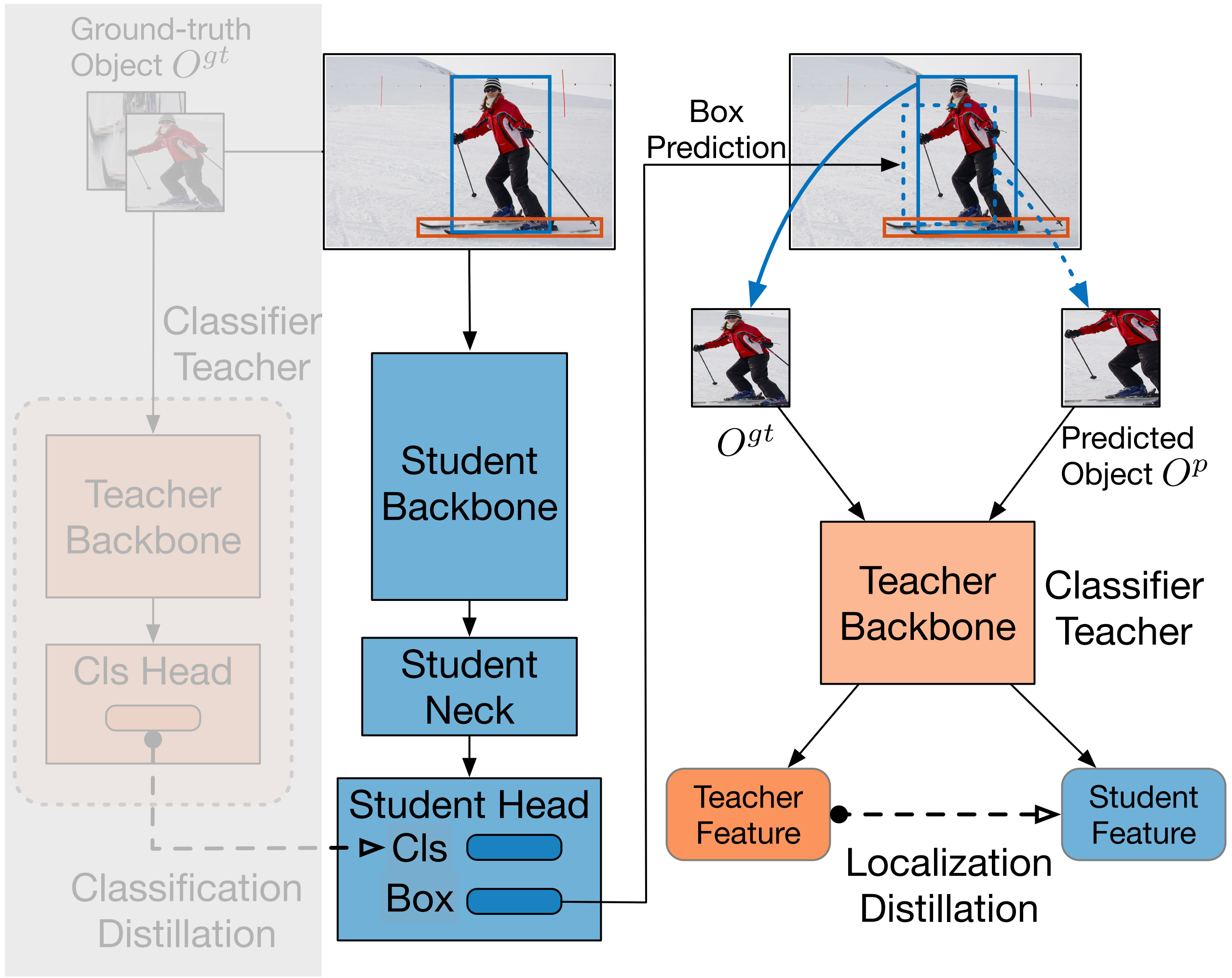


Our Method: Knowledge Distillation for Classification

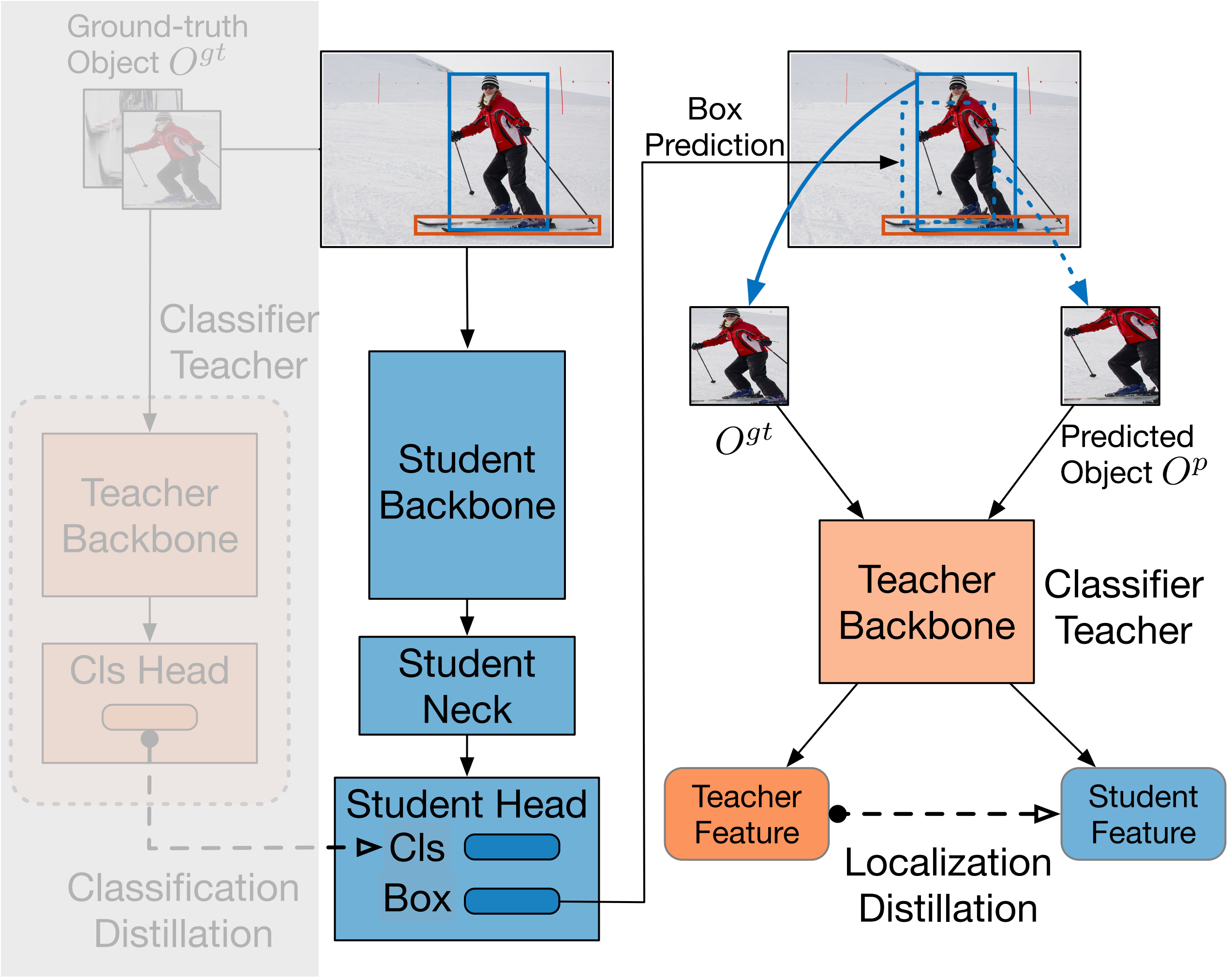
- Categorical cross-entropy loss
  softmax probability
- Binary cross-entropy loss
  sigmoid probability



Our Method: Knowledge Distillation for Localization



Our Method: Knowledge Distillation for Localization



- Spatial transformer
  - Given a bounding box

$$B_k = (x_1, y_1, x_2, y_2)$$

  - Compute the transformer matrix

$$A_k = \begin{bmatrix} (x_2 - x_1)/w & 0 & -1 + (x_1 + x_2)/w \\ 0 & (y_2 - y_1)/h & -1 + (y_1 + y_2)/h \end{bmatrix}$$

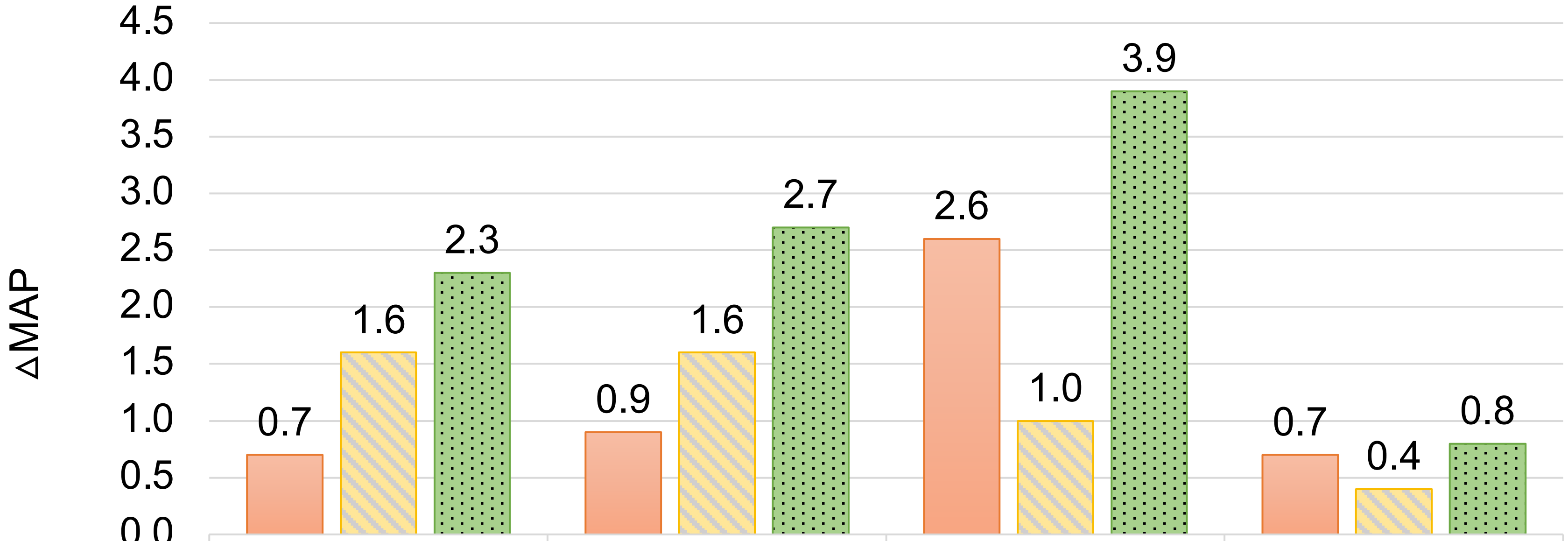
  - Get the object region

$$O_k^p = f_{ST}(A_k, I, s)$$

I -- input image

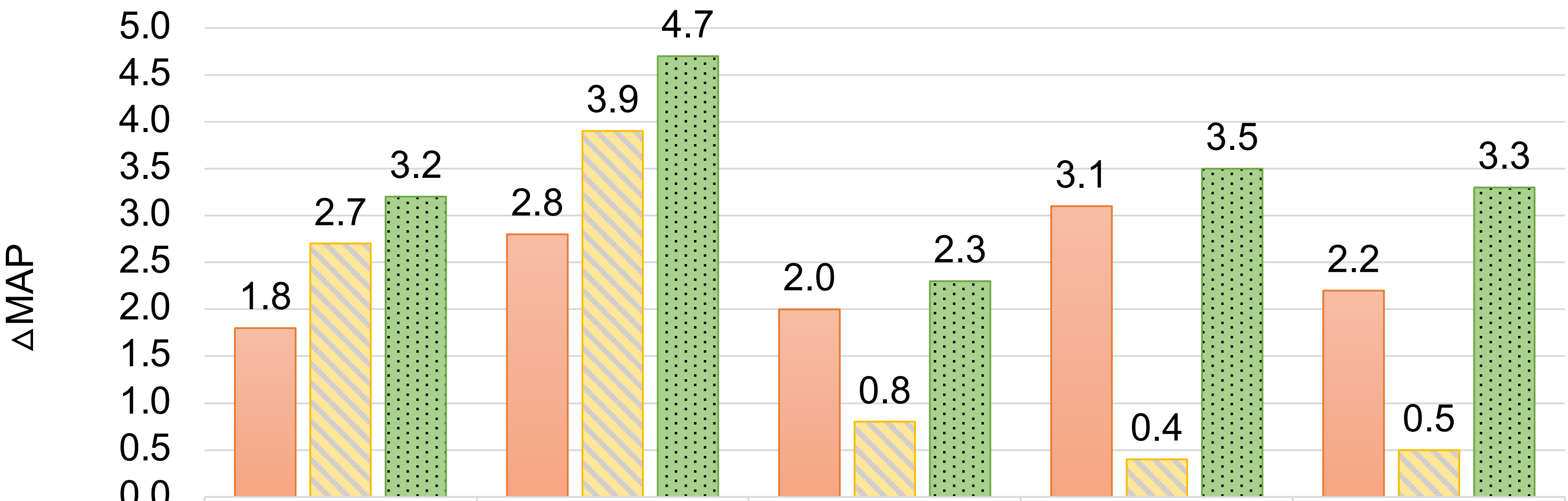
s -- grid sampling size

Results on Compact Detectors



	SSD300	SSD512	Faster RCNN-QR50	Faster RCNN-MV2
KD_cls	0.7	0.9	2.6	0.7
KD_loc	1.6	1.6	1.0	0.4
KD_cls + KD_loc	2.3	2.7	3.9	0.8

Comparing to State-of-the-art Distillation



	SSD512	Faster RCNN-QR50	Faster RCNN-MV2	Faster RCNN-R50	RetinaNet-R50
FKD	1.8	2.8	2.0	3.1	2.2
Ours	2.7	3.9	0.8	0.4	0.5
FKD+Ours	3.2	4.7	2.3	3.5	3.3

* FKD: Zhang et al. ICLR2021. Improve object detection with feature-based knowledge distillation: Towards accurate and efficient detectors.

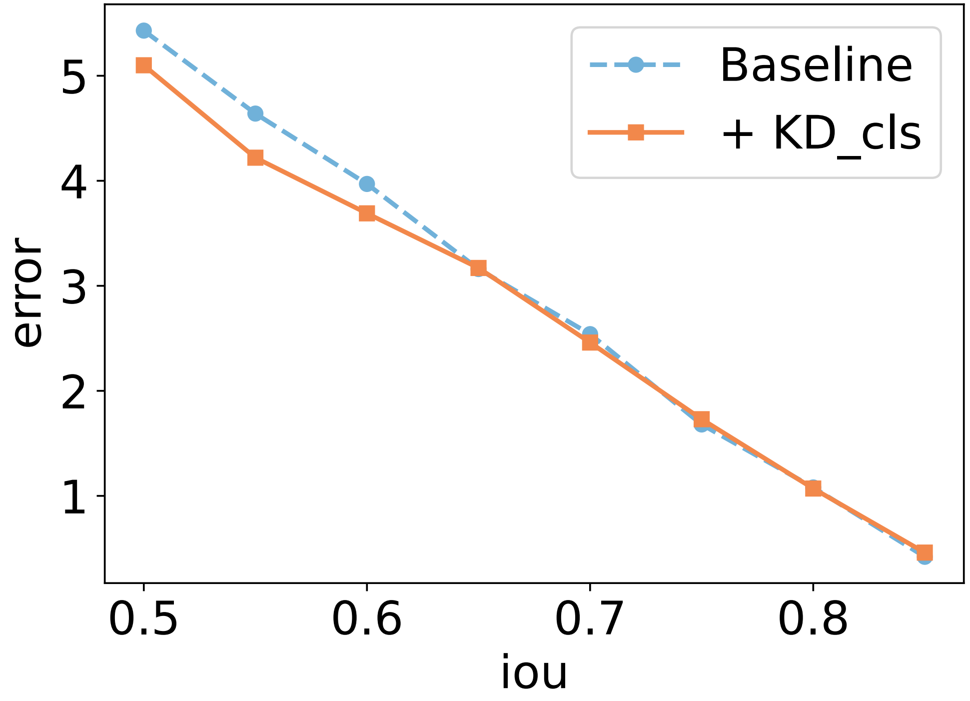
Comparing to State-of-the-art Distillation

	SSD512	Faster RCNN-QR50	Faster RCNN-MV2	Faster RCNN-R50	RetinaNet-R50
FKD	1.8	2.8	2.0	3.1	2.2
Ours	2.7	3.9	0.8	0.4	0.5
FKD+Ours	3.2	4.7	2.3	3.5	3.3

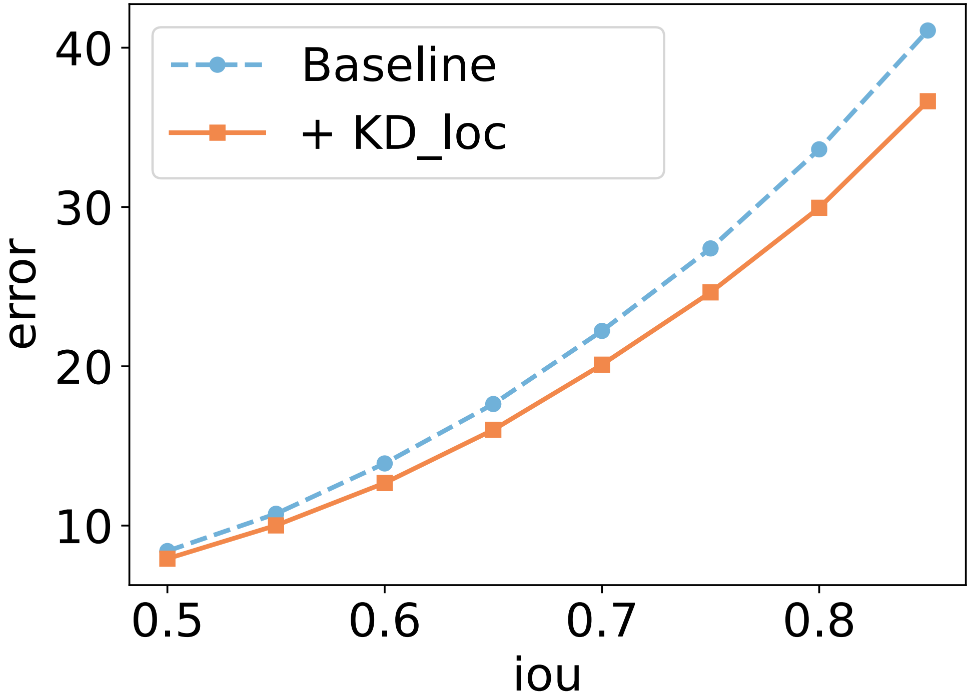
* FKD: Zhang et al. ICLR2021. Improve object detection with feature-based knowledge distillation: Towards accurate and efficient detectors.

Analysis

- Detection error analysis



(a) Classification error



(b) Localization error

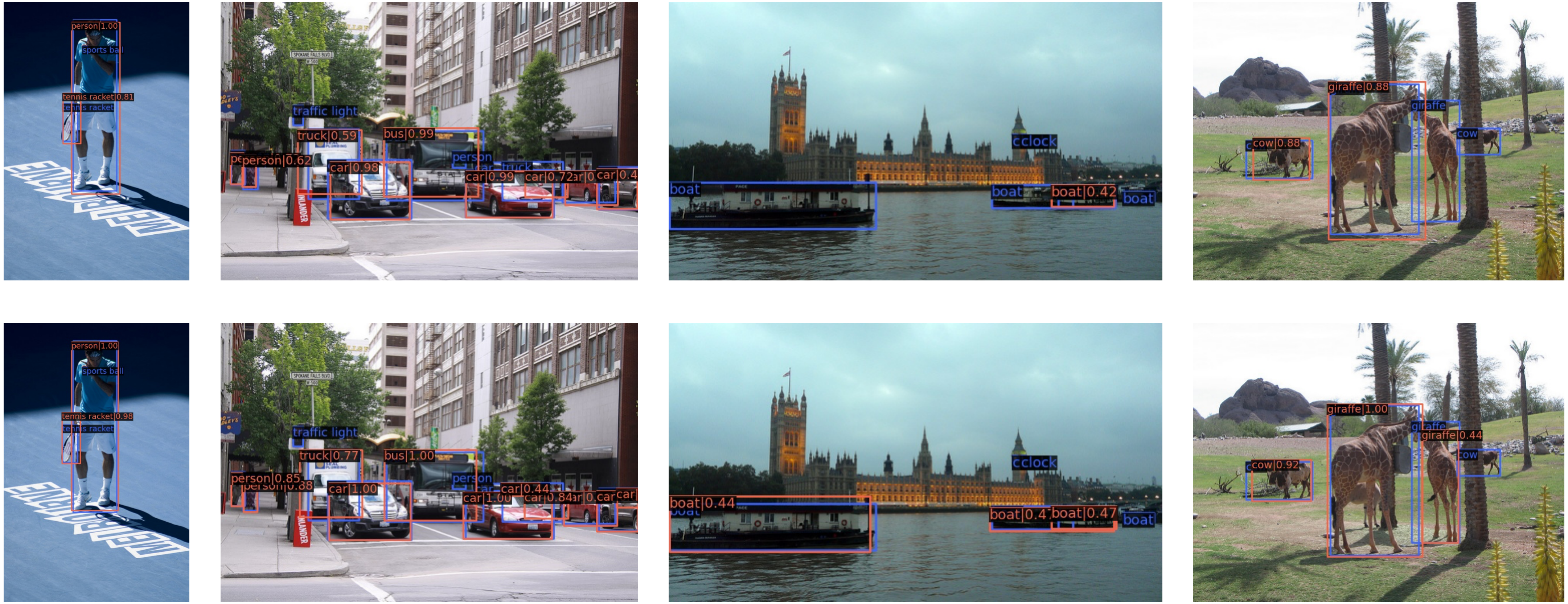
- Complementary nature of classification distillation and localization distillation.

Analysis

- ❖ Qualitative analysis

Baseline

Ours



(a) (b) (c) (d)

- ❖ More precise bounding box predictions
- ❖ Higher classification confidences

EPFL NVIDIA

Summary

- Our classifier-to-detector distillation improves both the classification accuracy and the localization ability of the student.

- Our classifier-to-detector distillation achieves better performance than detector-to-detector distillation.

- Our work opens the door to a new approach to distillation beyond object detection: Knowledge should be transferred not only across architectures, but also across tasks.

Code is available @ github.com/NVlabs/DICOD/
Please check our paper for more details.

SCAN ME

EPFL NVIDIA