Continual Detection Transformer for Incremental Object Detection

Poster ID: THU-PM-305

Yaoyao Liu[1] Bernt Schiele[1] Andrea Vedaldi[2] Christian Rupprecht[2]

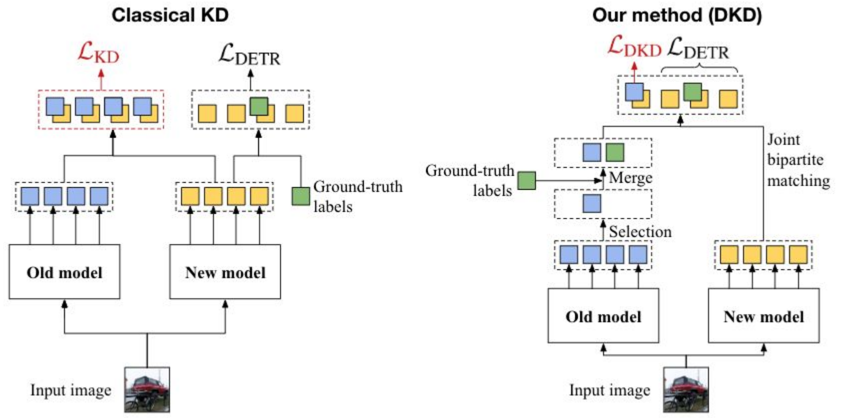
[1]Max Planck Institute for Informatics [2]Visual Geometry Group, University of Oxford

Quick preview: continual detection transformer

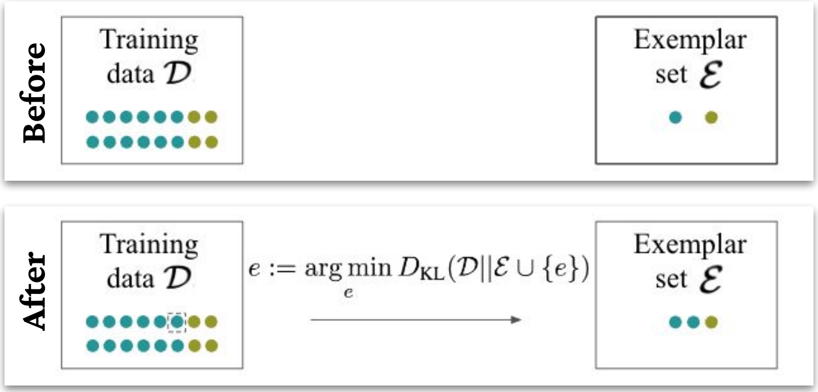
Task: incremental object detection (IOD)

i.e., learning a detector from a **continual data stream** with **limited memory**

- **How to remove the influence of the background?**
  idea: selecting the **most-confident non-background** predictions as **pseudo labels**



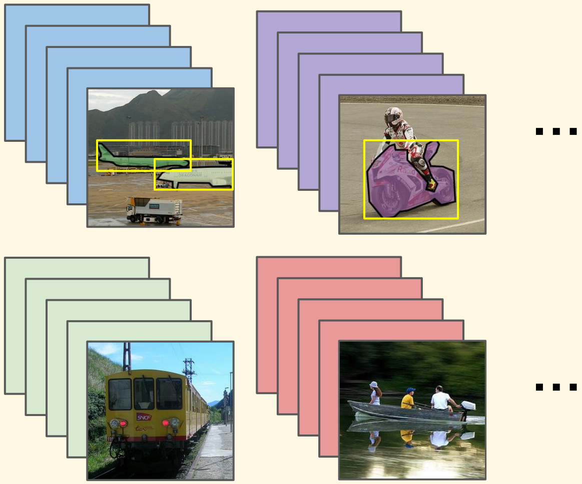
- **How to select an exemplar set following the original distribution?**
  idea: minimizing the **KL divergence** between the **training** set and **exemplar** set



Task: incremental object detection (IOD)

COCO: 80 classes, ~100k images, 10-phase

Phase 1
10k images, **8** classes labeled



Phase 2
10k **new** images, **8 new** classes labeled

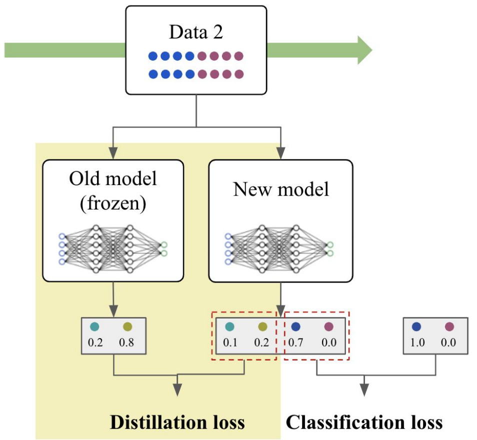
...

Evaluation:
Test for **8** classes

Evaluation:
Test for **16** classes

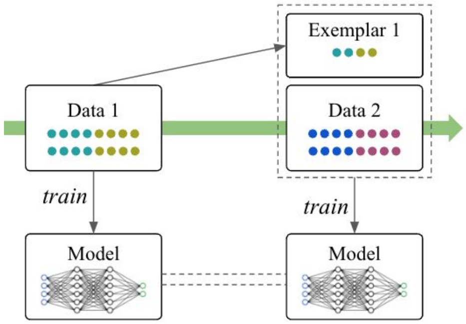
What is the toolbox we have for incremental learning (classification)?

Knowledge distillation (KD)



Basic idea: encourage **the new model's predictions** or **feature maps** to be close to those of **the old model**

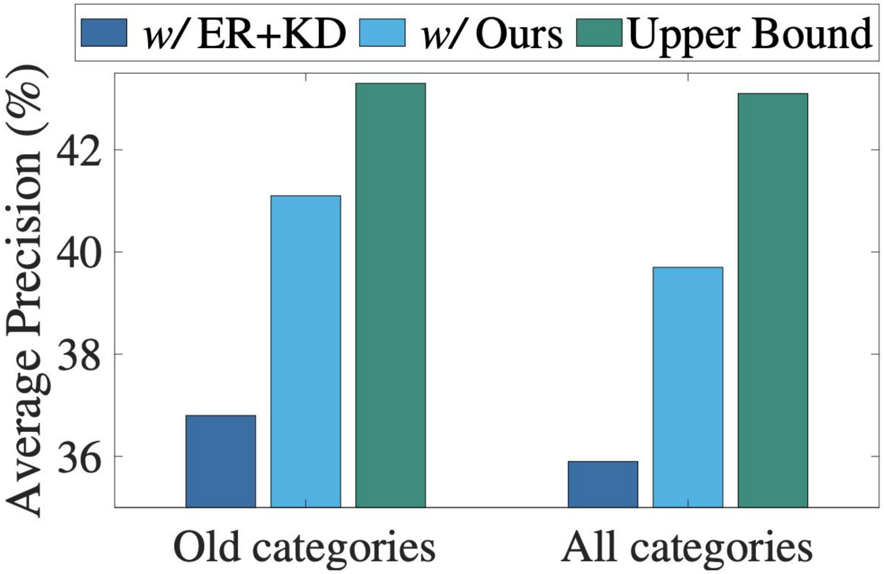
Exemplar replay (ER)



Basic idea: replaying a **small subset** of **the old data** in the following phases

Reference
[1] Rebuffi, Sylvestre-Alvise, et al. "icarl: Incremental classifier and representation learning." CVPR 2017;
[2] Hou, Saihui, et al. "Learning a unified classifier incrementally via rebalancing." CVPR 2019;
[3] Liu, Yaoyao, et al. "Mnemonics training: Multi-class incremental learning without forgetting." CVPR 2020.
[4] Wang, Liyuan, et al. "Memory Replay with Data Compression for Continual Learning." ICLR 2022.

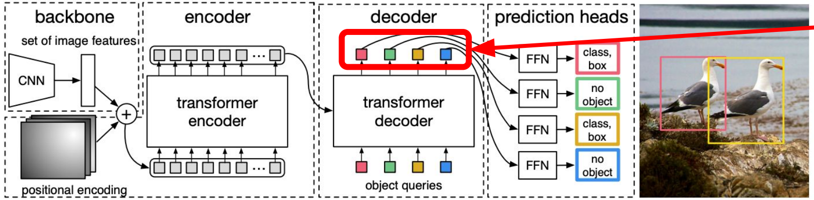
Can we directly apply our toolbox to object detection?



MS COCO, 2-phase, baseline: Deformable DETR

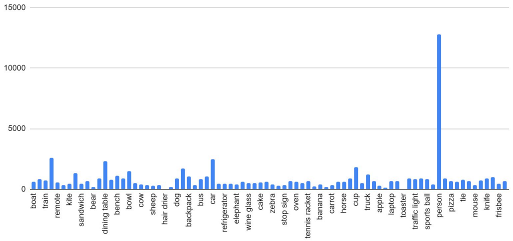
Why KD and ER don’t work well on object detection?

- **Problem 1: the problem of KD:**
  The KD loss is **dominated** by the **background** information



Too many predictions (~300)

- **Problem 2: the problem of ER:**
  The heuristic ER methods **change** the **original** category **distribution** of the training set.



How do we solve these problems?

- **Problem 1: the problem of KD:**
  The KD loss is **dominated** by the **background** information

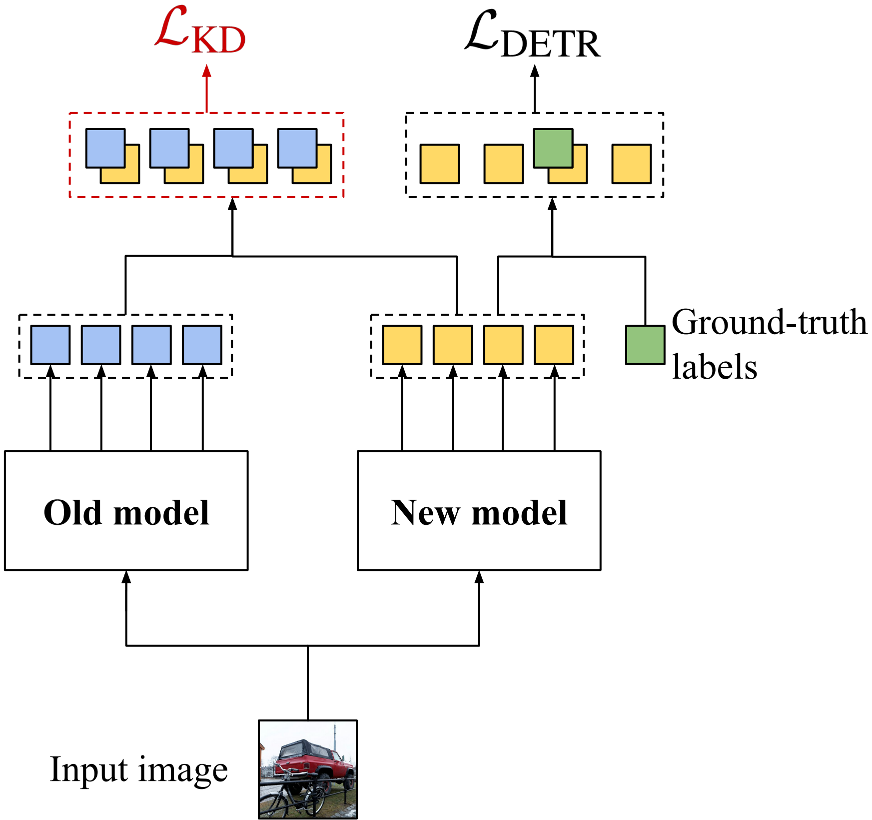
  Our solution: **removing** the **influence** of the **background** information

- **Problem 2: the problem of ER:**
  The heuristic ER methods **change** the **original** category **distribution** of the training set.

  Our solution: selecting an **exemplar set** that **follows the original distribution**

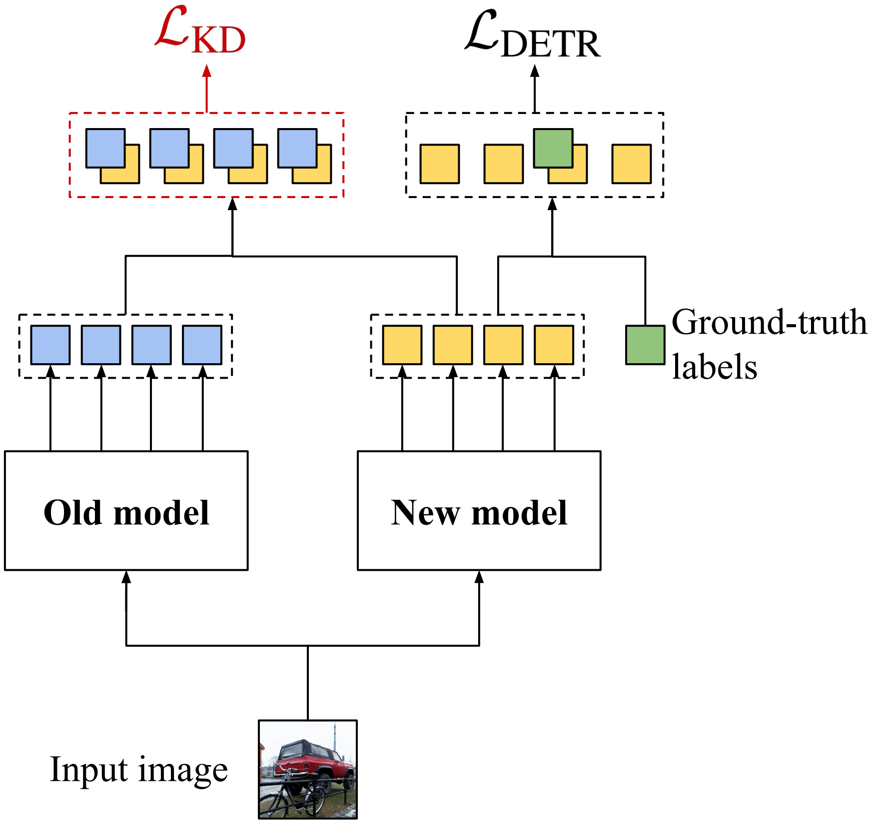
For problem 1: how to remove the influence of the background?

Classical KD



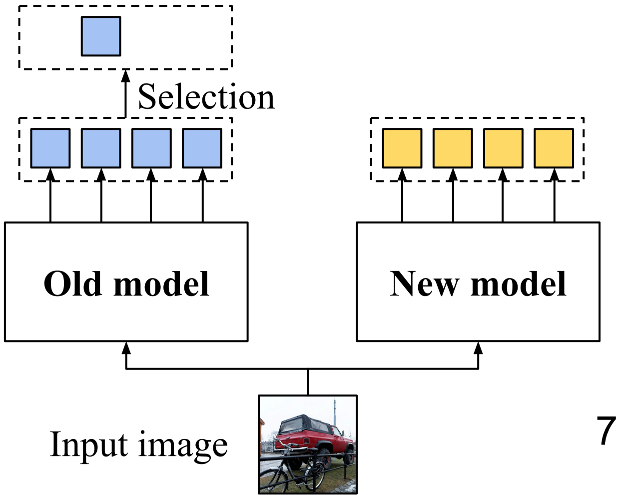
For problem 1: how to remove the influence of the background?

Classical KD



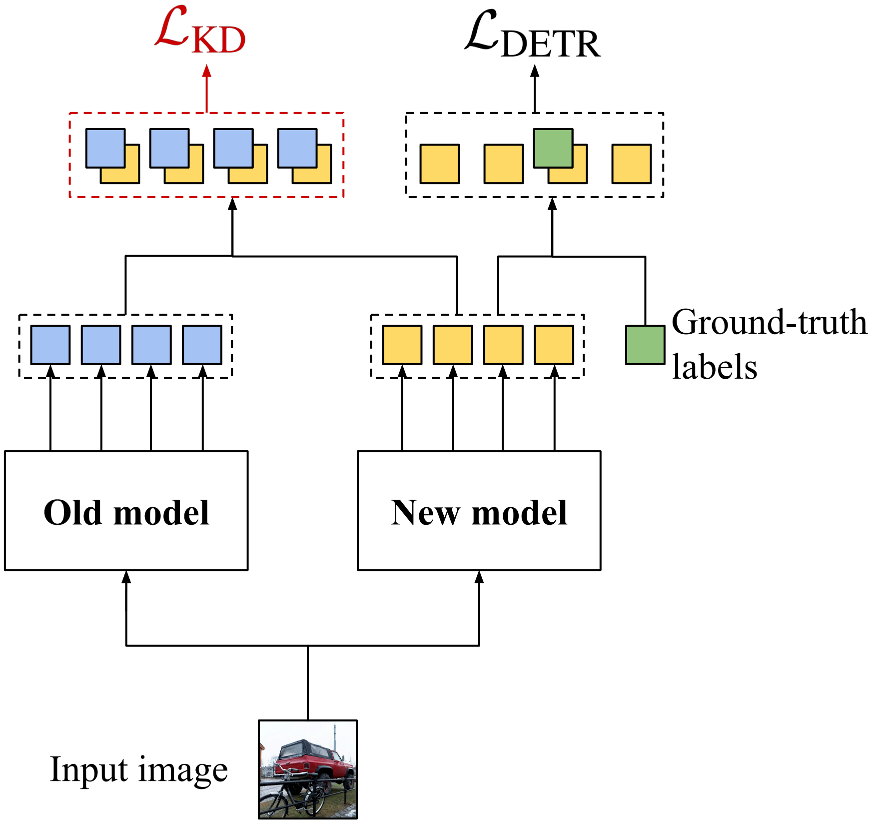
Our method (DKD)

Select the **most confident non-background** predictions

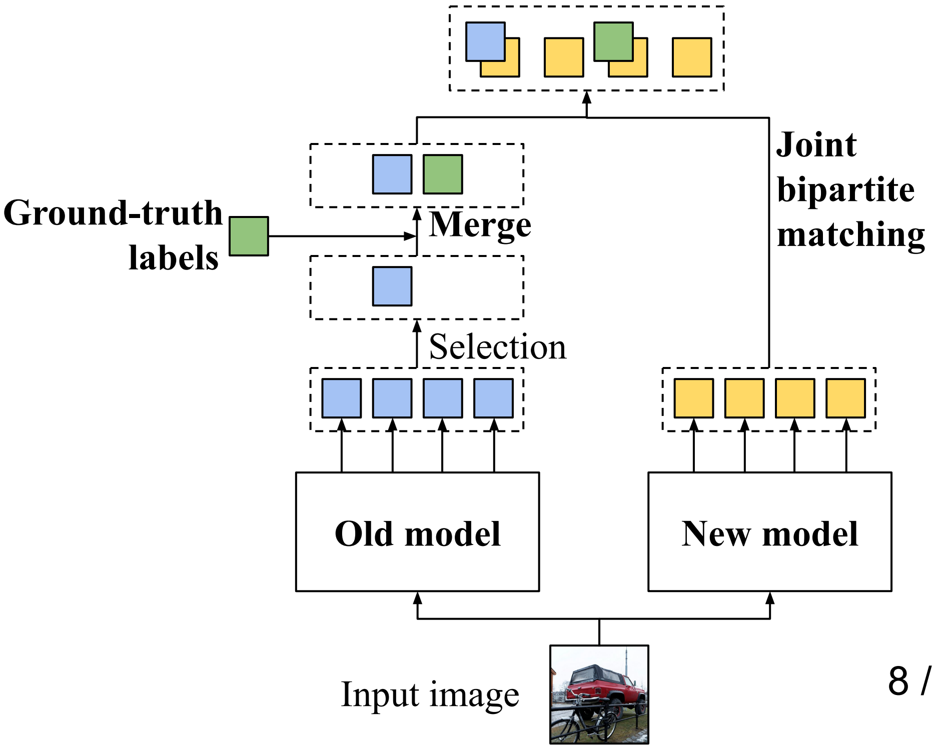


For problem 1: how to remove the influence of the background?

Classical KD

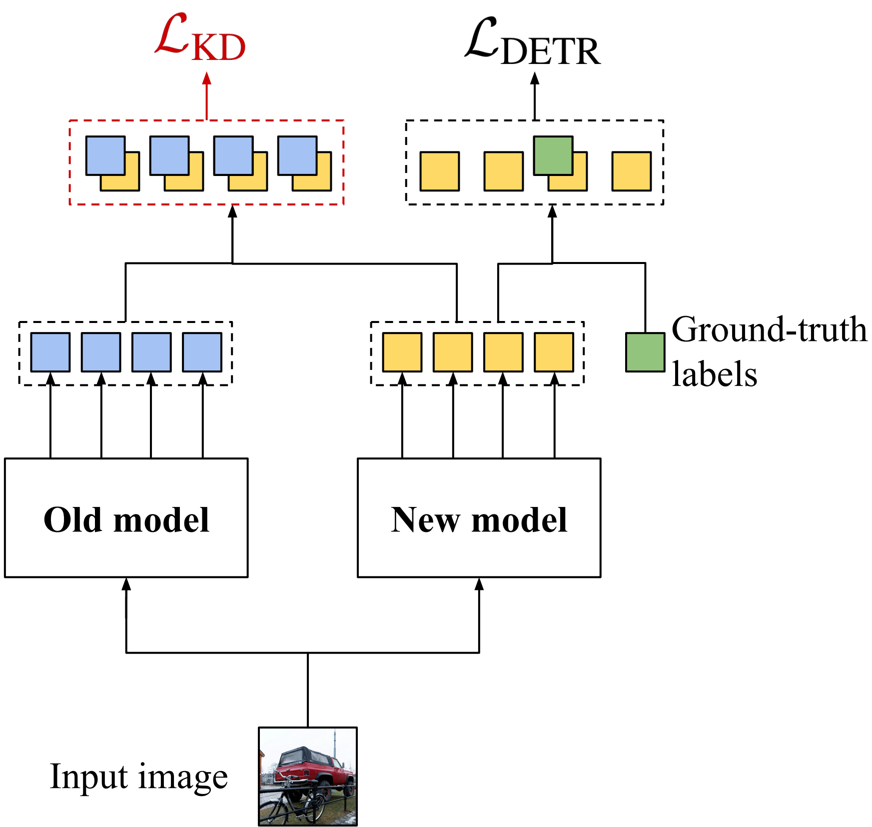


Our method (DKD)

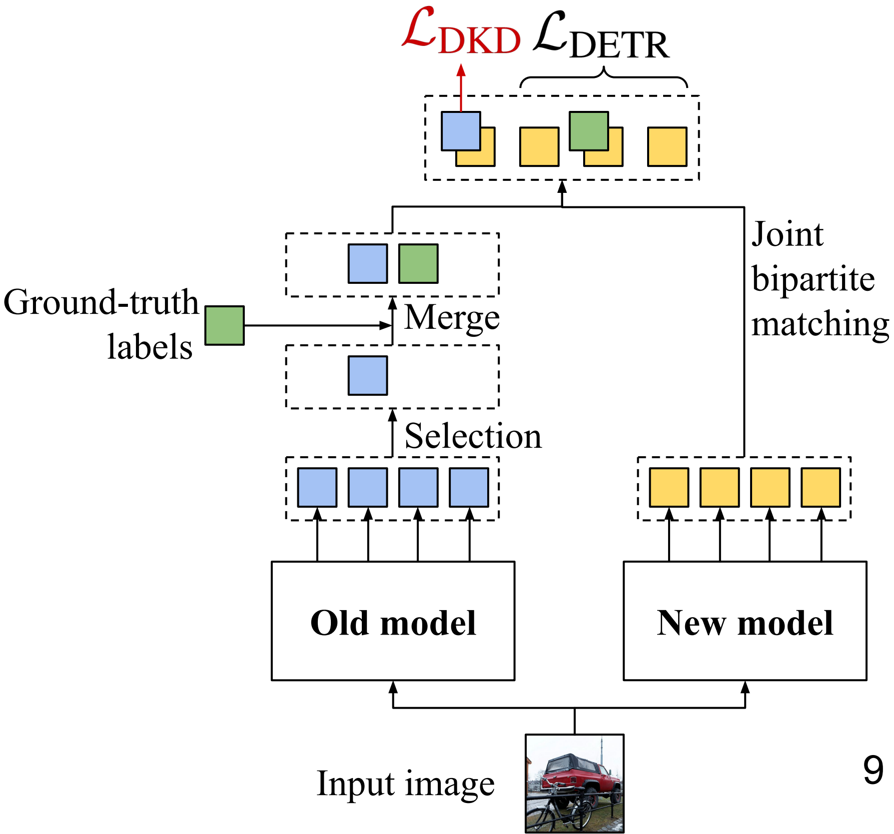


For problem 1: how to remove the influence of the background?

Classical KD



Our method (DKD)



For problem 2: how to select an exemplar set that follows the original distribution?

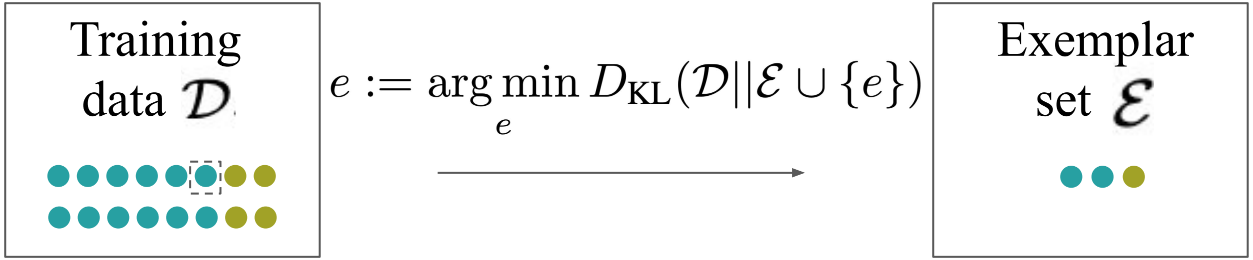
Idea: **minimizing** the **Kullback–Leibler divergence** between the **training** set and **exemplar** set

When we need to select a new exemplar

For problem 2: how to select an exemplar set that follows the original distribution?

Idea: **minimizing** the **Kullback–Leibler divergence** between the **training** set and **exemplar** set

When we need to select a new exemplar



max planck institut informatik

Ablation study: our method vs. classical KD & ER

	Our KD Method		Our ER Method									
Knowledge distillation (KD)	Joint bipartite matching	Pseudo label selection	Exemplar replay (ER)	Distribution preserving calibration	All categories ↑				Old categories ↑			
					AP	AP_S	AP_M	AP_L	AP	AP_S	AP_M	AP_L
					4.2	1.6	4.7	5.8	0.7	0.2	0.8	0.8
✓					24.5	12.4	28.2	35.2	24.0	12.3	27.7	34.4
✓	✓				30.3	19.5	33.0	39.0	33.4	21.8	36.4	43.2
✓	✓	✓			33.9	16.3	37.1	49.2	33.9	16.6	36.8	50.0
✓	✓	✓	✓		37.9	20.8	40.9	50.4	39.0	21.6	41.7	52.3
✓	✓	✓		✓	**40.1**	**23.2**	**43.2**	**52.1**	**41.8**	**24.5**	**44.7**	**54.6**

Observation: our KD & ER achieve better performance than classical KD & ER

Insight: **our design** makes **KD** and **ER perform better** on **object detection** tasks

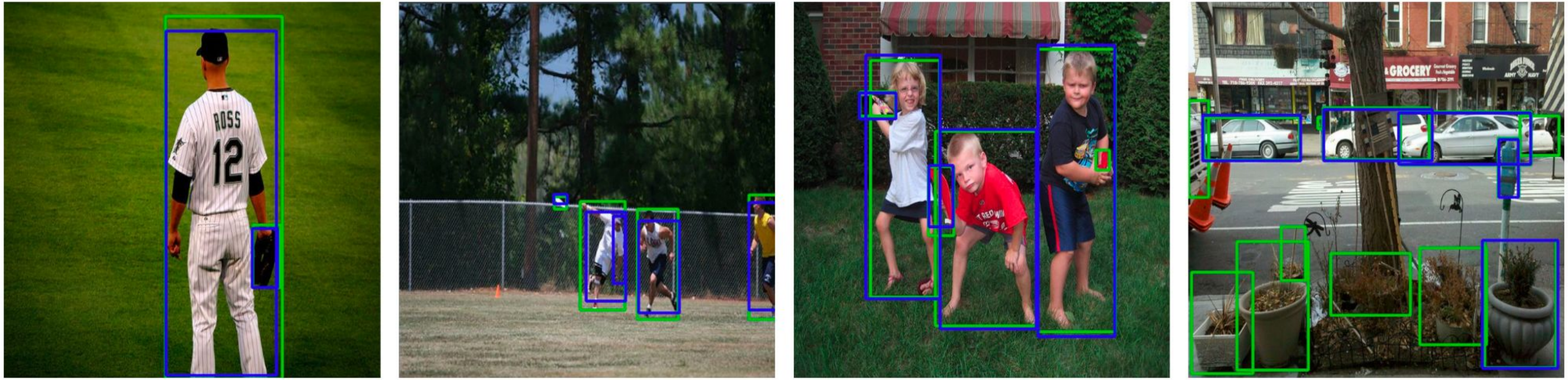
Results: our method + different object detection frameworks

Method	Detection baseline	AP	AP_{50}	AP_{75}
ERD [13]	UP-DETR	$36.2_{\pm 0.3}$	$54.8_{\pm 0.4}$	$39.3_{\pm 0.4}$
CL-DETR (ours)	UP-DETR	$37.6_{\pm 0.2}$	$56.5_{\pm 0.4}$	$39.4_{\pm 0.3}$
LwF [27]	Deformable DETR	$24.5_{\pm 0.3}$	$36.6_{\pm 0.2}$	$26.7_{\pm 0.4}$
iCaRL [42]	Deformable DETR	$35.9_{\pm 0.4}$	$52.5_{\pm 0.3}$	$39.2_{\pm 0.3}$
ERD [13]	Deformable DETR	$36.9_{\pm 0.4}$	$55.7_{\pm 0.4}$	$40.1_{\pm 0.4}$
CL-DETR (ours)	Deformable DETR	$\mathbf{40.1}_{\pm 0.3}$	$\mathbf{57.8}_{\pm 0.4}$	$\mathbf{43.7}_{\pm 0.3}$

Observation: our method can improve different DETR-based architectures

Insight: our method is **generic** to **different DETR-based methods**

Visualizations: our pseudo labels vs. ground-truth labels



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

Visualizations of the old category pseudo (blue) and ground-truth (green) bounding boxes on COCO 2017

Summary: our contributions

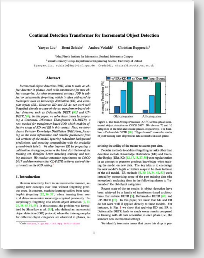
(1) **The DKD loss**
resolving conflicts between distilled knowledge and new evidence
and by ignoring redundant background detections

(2) A **calibration strategy** for **ER**
matching the stored exemplars to the training set distribution

(3) A revised **IOD benchmark protocol**
avoiding observing the same images in different training phases;

(4) **Extensive experiments**
including state-of-the-art results, an in-depth ablation study, and further visualizations.

Thanks!


Continual Detection Transformer for Incremental Object Detection

Yaoyao Liu[1] Bernt Schiele[1] Andrea Vedaldi[2] Christian Rupprecht[2]

Continual Detection Transformer for Incremental Object Detection

Webpage: https://lyy.mpi-inf.mpg.de/CL-DETR/

Code: https://github.com/yaoyao-liu/CL-DETR