Bridging Non Co-occurrence with Unlabeled In-the-wild Data for Incremental Object Detection

Na Dong, Yongqiang Zhang, Mingli Ding, Gim Hee Lee

National University of Singapore, Harbin Institute of Technology

NeurIPS 2021

NUS National University of Singapore | School of Computing

NEURAL INFORMATION PROCESSING SYSTEMS

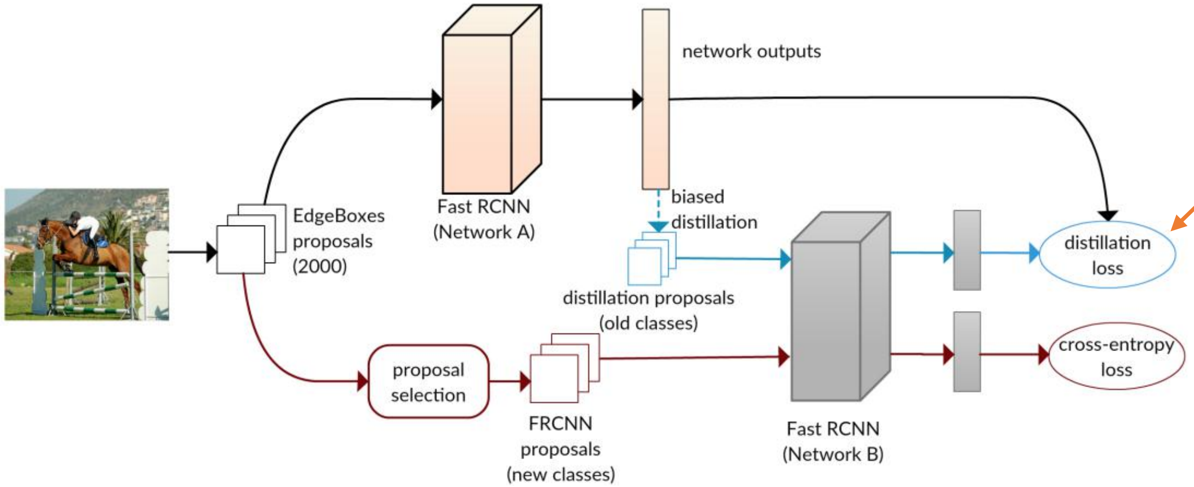
Catastrophic Forgetting

- Deep networks have shown remarkable results in the task of object detection.
- However, their performance suffers critical drops when they are subsequently trained on novel classes without any sample from the base classes originally used to train the model.
- This phenomenon is known as catastrophic forgetting.

Existing Works Requires Co-occurrence

- Existing methods performs well when there is co-occurrence of the unlabeled base classes in the training data of the novel classes.

- This requirement is impractical in many real-world settings since the base classes do not necessarily co-occur with the novel classes.



Most effective when there's co-occurrence!

Shmelkov et al, CVPR 2017

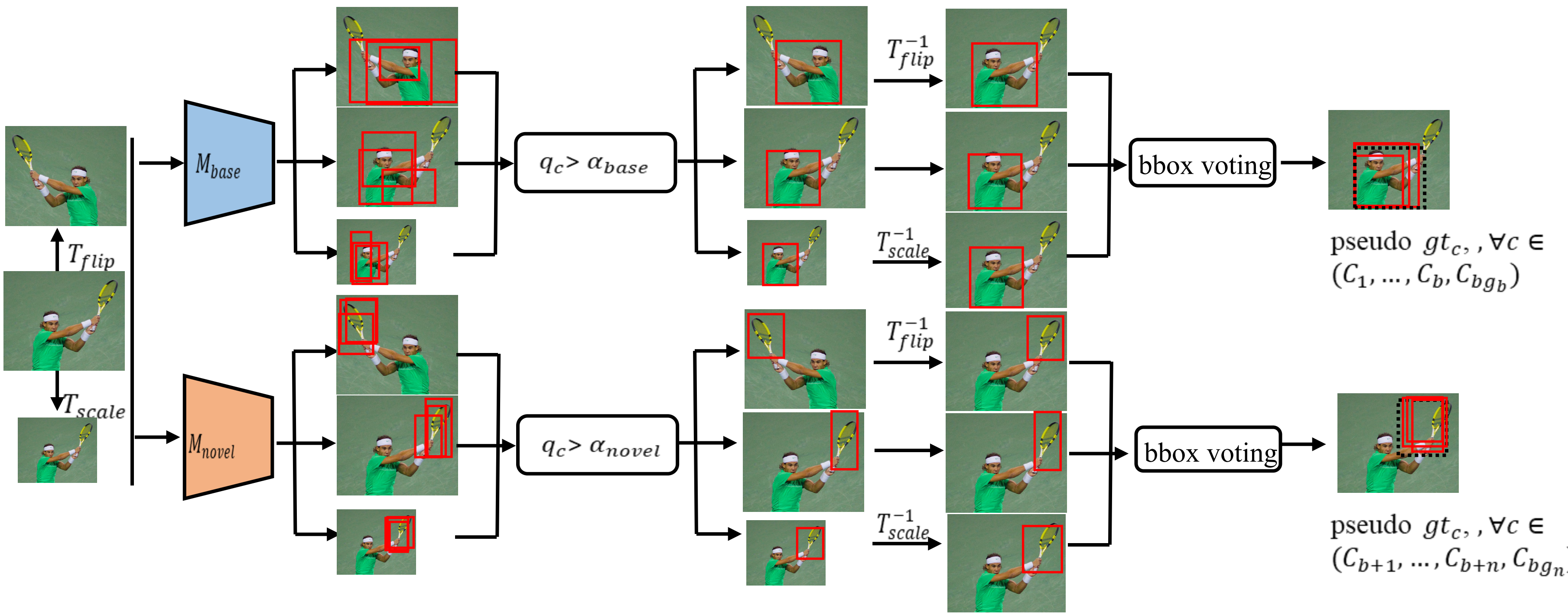
Using In-the-wild Data to Bridge Non Co-occurrence

- We consider a more practical setting of complete absence of co-occurrence of the base and novel classes in the training data.

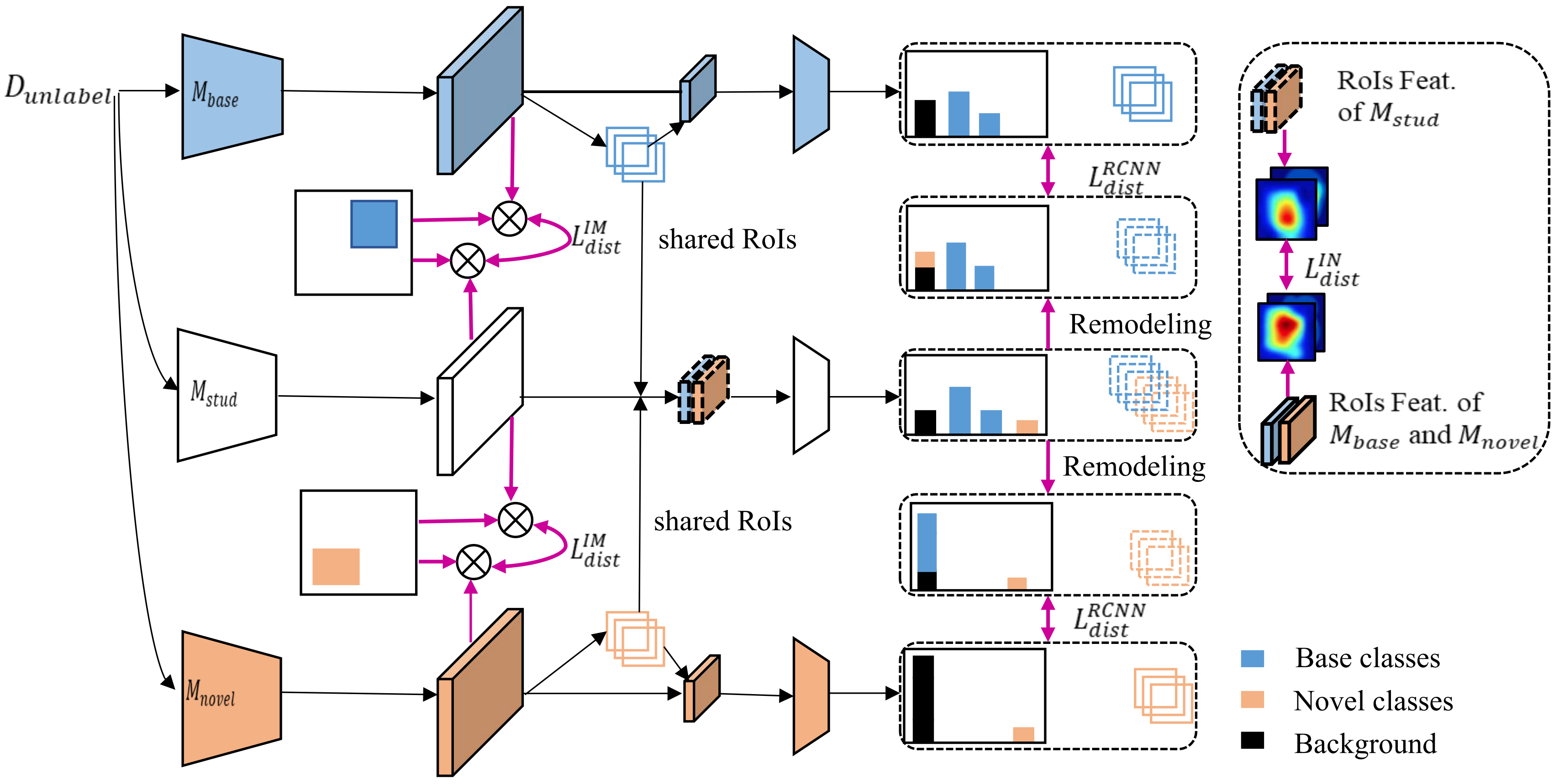
- We propose the use of unlabeled in-the-wild data to bridge the non co-occurrence caused by the missing base classes during the training of additional novel classes.

Our Approach: Blind Sampling Strategy

q_c: Object class probability
α: Threshold

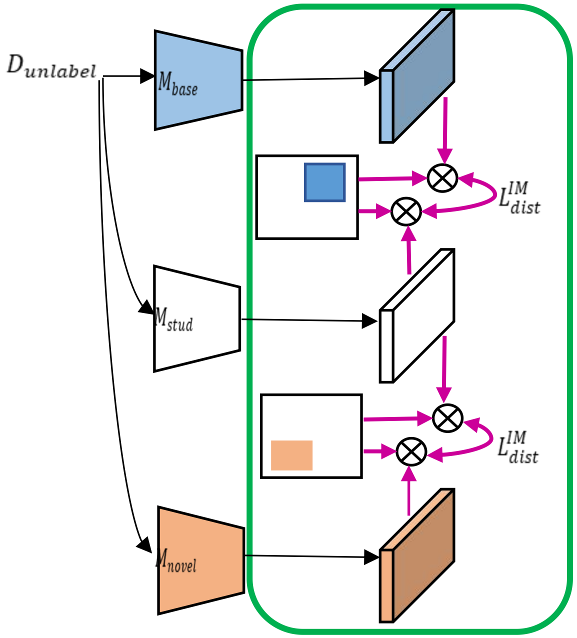


Our Approach: Dual Teacher Distillation



Our Approach: Dual Teacher Distillation

Image-level distillation with ROI Masks



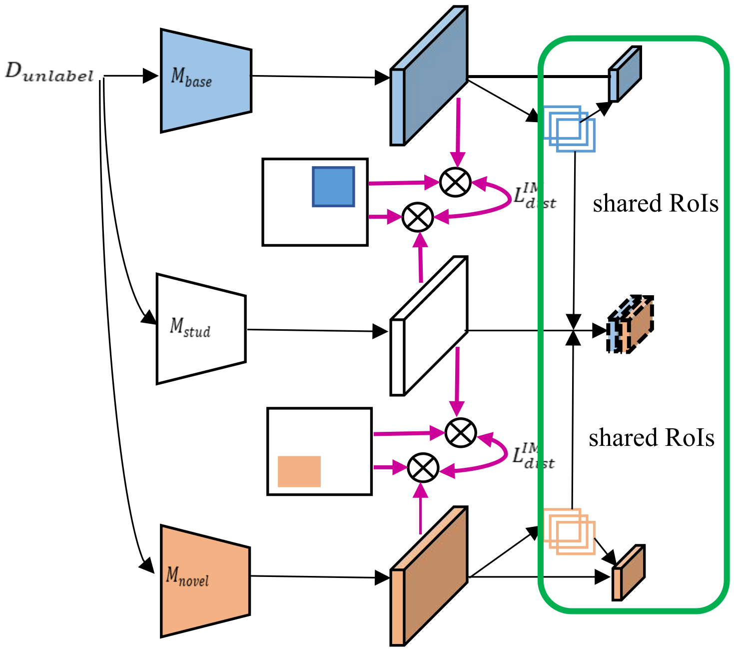
- Base classes
- Novel classes
- Background

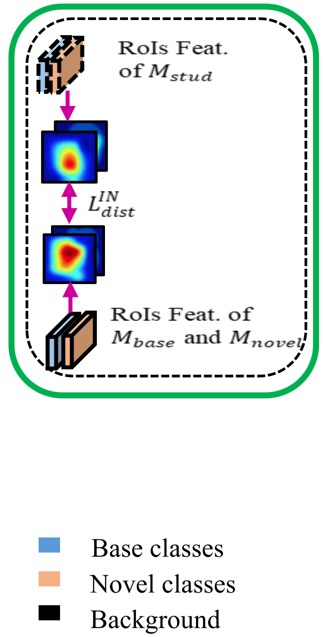
$$\mathcal{L}_{\text{dist}}^{\text{IM}} = \frac{1}{2N^{\text{base}}} \sum_{i=1}^{W} \sum_{j=1}^{H} \sum_{k=1}^{C} Mask_{ij}^{\text{base}} \left\| F_{ijk}^{\text{stud}} - F_{ijk}^{\text{base}} \right\|^2 + \frac{1}{2N^{novel}} \sum_{i=1}^{W} \sum_{j=1}^{H} \sum_{k=1}^{C} Mask_{ij}^{\text{novel}} \left\| F_{ijk}^{\text{stud}} - F_{ijk}^{\text{novel}} \right\|^2,$$

where $N^{\text{base}} = \sum_{i=1}^{W} \sum_{j=1}^{H} Mask_{ij}^{\text{base}}, \; N^{\text{novel}} = \sum_{i=1}^{W} \sum_{j=1}^{H} Mask_{ij}^{\text{novel}}.$

Our Approach: Dual Teacher Distillation

Instance-level distillation with heatmaps



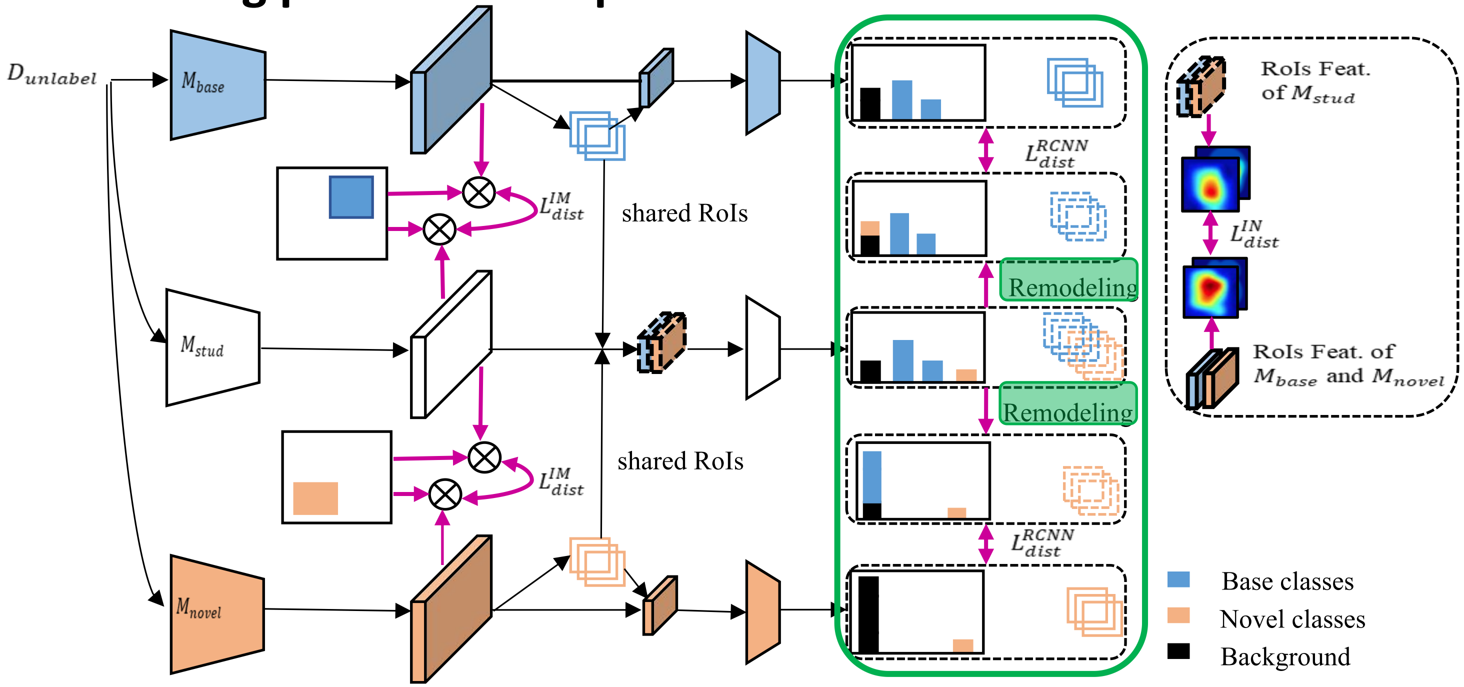


$$\mathcal{L}^{\text{IN}}_{\text{dist}} = \mathcal{L}_{\text{mse}}(\mathcal{H}^{\text{base}} \cup \mathcal{H}^{\text{novel}}, \mathcal{H}^{\text{stud}}) \ ,$$

where $\mathcal{H}^{\text{base}}_{ij} = S(\frac{1}{C}\sum_{k=1}^{C} f^{\text{base}}_{ijk}),\ \mathcal{H}^{\text{novel}}_{ij} = S(\frac{1}{C}\sum_{k=1}^{C} f^{\text{novel}}_{ijk}),\ \mathcal{H}^{\text{stud}}_{ij} = S(\frac{1}{C}\sum_{k=1}^{C} f^{\text{stud}}_{ijk})$

Our Approach: Dual Teacher Distillation

Remodeling prediction outputs



$$\tilde{q}_c^{\text{stud,base}} = \{(q_{\text{bg}}^{\text{stud}} + q_{c_{b+1}}^{\text{stud}} + \cdots + q_{c_{b+n}}^{\text{stud}}), q_{c_1}^{\text{stud}}, q_{c_2}^{\text{stud}}, \ldots, q_{c_b}^{\text{stud}}\}.$$

$$\tilde{q}_c^{\text{stud, novel}} = \{(q_{bg}^{\text{stud}} + q_{c_1}^{\text{stud}} + q_{c_2}^{\text{stud}} + \cdots + q_{c_b}^{\text{stud}}), q_{c_{b+1}}^{\text{stud}}, \ldots, q_{c_{b+n}}^{\text{stud}}\}.$$

$$\begin{aligned}\mathcal{L}_{\text{dist}}^{\text{RCNN}} &= \underbrace{\mathcal{L}_{\text{kl\_div}}(\log(\tilde{q}_c^{\text{stud,base}}), q_c^{\text{base}})}_{} + \lambda\mathcal{L}_{\text{smooth}_{L_1}}(\tilde{r}_c^{\text{stud,base}}, r_c^{\text{base}}) \\ &+ \underbrace{\mathcal{L}_{\text{kl\_div}}(\log(\tilde{q}_c^{\text{stud,novel}}), q_c^{\text{novel}})}_{\text{Class probabilities}} + \underbrace{\lambda\mathcal{L}_{\text{smooth}_{L_1}}(\tilde{r}_c^{\text{stud,novel}}, r_c^{\text{novel}})}_{\text{Bbox parameters}}\end{aligned}$$

Class probabilities

Bbox parameters

Results

- **Left Table:** Results of "19+1" on VOC test set. "1-19"and "20" ("tv") are base and novel classes.
- **Right Table:** Results of "15+5" on VOC test set. "1-15"and "16-20" are the base and novel classes.

Class	Method	mAP(%) (*base* | *novel* | *all*)
1-20 (w/o co-occur)	Ren [24]	73.1 | 55.4 | 72.3
1-19 (w/o co-occur)	Ren [24]	73.4 | – | –
20 (w/o co-occur)	Ren [24]	– | 47.4 | –
(1-19) + (20) (w/o co-occur)	Shmelkov [27]	62.6 | 39.2 | 61.4
	Ours (w category)	**73.3** | **50.7** | **72.2**
	Ours (w/o category)	**71.5** | **46.1** | **70.2**
(1-19) + (20) (w co-occur)	Shmelkov [27]	68.5 | 62.7 | 68.3
	Zhou [40]	70.5 | 53.0 | 69.6
	Ours (w category)	**73.5** | **65.8** | **73.1**

Class	Method	mAP(%) (*base* | *novel* | *all*)
1-20 (w/o co-occur)	Ren [24]	74.4 | 62.7 | 71.7
1-15 (w/o co-occur)	Ren [24]	72.0 | – | –
16-20 (w/o co-occur)	Ren [24]	– | 48.6 | –
(1-15) + (16-20) (w/o co-occur)	Shmelkov [27]	67.2 | 46.1 | 62.0
	Ours (w category)	**70.5** | **49.4** | **65.3**
	Ours (w/o category)	**70.7** | **48.5** | **65.1**
Class (1-15) + (16-20) (w co-occur)	Shmelkov [27]	68.4 | 58.4 | 65.9
	Ours (w category)	**72.7** | **58.4** | **69.1**

w(/o) category: with(out) class overlap in the in-the-wild

NUS National University of Singapore | School of Computing

NEURAL INFORMATION PROCESSING SYSTEMS

Results

- Results of "19+1" on VOC test set. "1-19"and "20" ("tv") are base and novel classes.

Class	Method	mAP(%) (*base* | *novel* | *all*)
1-20 (w/o co-occur)	Ren [24]	73.1 | 55.4 | 72.3
1-19 (w/o co-occur)	Ren [24]	73.4 | – | –
16-20 (w/o co-occur)	Ren [24]	– | 47.4 | –
(1-19) + (20) (w/o co-occur)	Shmelkov [27]	62.6 | 39.2 | 61.4
	Ours	**71.3** | **48.6** | **70.1**
(1-19) + (20) (w co-occur)	Shmelkov [27]	68.5 | 62.7 | 68.3
	Zhou [40]	70.5 | 53.0 | 69.6

Results

- Results of "10+5+5" on VOC tests et. "1-10" are the base classes, and "11-15" and "16-20" are the two groups of sequentially added novel classes.

Class	Method	mAP(%) (*base* | *novel* | *all*)
1-20 (w/o co-occur)	Ren [24]	66.6 | 67.3 | 66.7
1-10 (w/o co-occur)	Ren [24]	57.8 | – | –
11-15 (w/o co-occur)	Ren [24]	– | 62.4 | –
16-20 (w/o co-occur)	Ren [24]	– | 48.6 | –
(1-10)+ (11-15) (w/o co-occur)	Shmelkov [27]	**59.8** | 52.4 | 57.3
	Ours (w category)	57.0 | **62.7** | **58.9**
	Ours (w/o category)	57.3 | **61.7** | **58.8**
(1-10)+ (11-15)+ (16-20) (w/o co-occur)	Shmelkov [27]	**59.0** | 47.3 | 53.1
	Ours (w category)	56.7 | **55.1** | **55.8**
	Ours (w/o category)	56.9 | **53.9** | **55.4**
(1-10)+ (11-15)+ (16-20) (w co-occur)	Zhou [40]	60.3 | 53.1 | 56.7
	Ours (w category)	**68.1** | **64.8** | **66.5**

Results

- Results of “40+40" on COCO minival set. First 40 classes are the old classes, and the next 40 are the added classes.

Class	Method	AP	AP50	AP75	APS	APM	APL
1-80 (w/o co-occur)	Ren [24]	27.7	45.8	29.4	10.8	30.9	42.5
(1-40) + (40-80) (w/o co-occur)	Ours	**22.5**	**40.9**	**23.0**	**8.3**	**25.9**	**34.6**
(1-40) + (40-80) (w co-occur)	Shmelkov [27]	21.3	37.4	-	-	-	-
	Zhou [40]	22.7	36.8	-	-	-	-
	Ours	**23.7**	**42.5**	**24.3**	**8.6**	**26.6**	**37.5**

Results

- Ablation studies for the setting of "19+1" on VOC 2007 test set.

Blind sampling strategy	$\mathcal{L}^{\text{RCNN}} + \mathcal{L}^{\text{RPN}}$	$\mathcal{L}_{\text{dist}}^{\text{RCNN}}$	$\mathcal{L}_{\text{dist}}^{\text{IM}}$	$\mathcal{L}_{\text{dist}}^{\text{IN}}$	mAP(%) $(base \mid novel \mid all)$
✓	✓				64.4 | 35.3 | 62.9
✓	✓	✓			68.0 | 44.9 | 66.9
✓	✓		✓		68.5 | 39.6 | 67.1
✓	✓			✓	67.6 | 39.2 | 66.2
	✓	✓	✓	✓	70.0 | 44.3 | 68.7
✓	✓	✓	✓	✓	**71.5** | **46.1** | **70.2**

NUS National University of Singapore | School of Computing

NEURAL INFORMATION PROCESSING SYSTEMS