Distilling Object Detectors with Global Knowledge

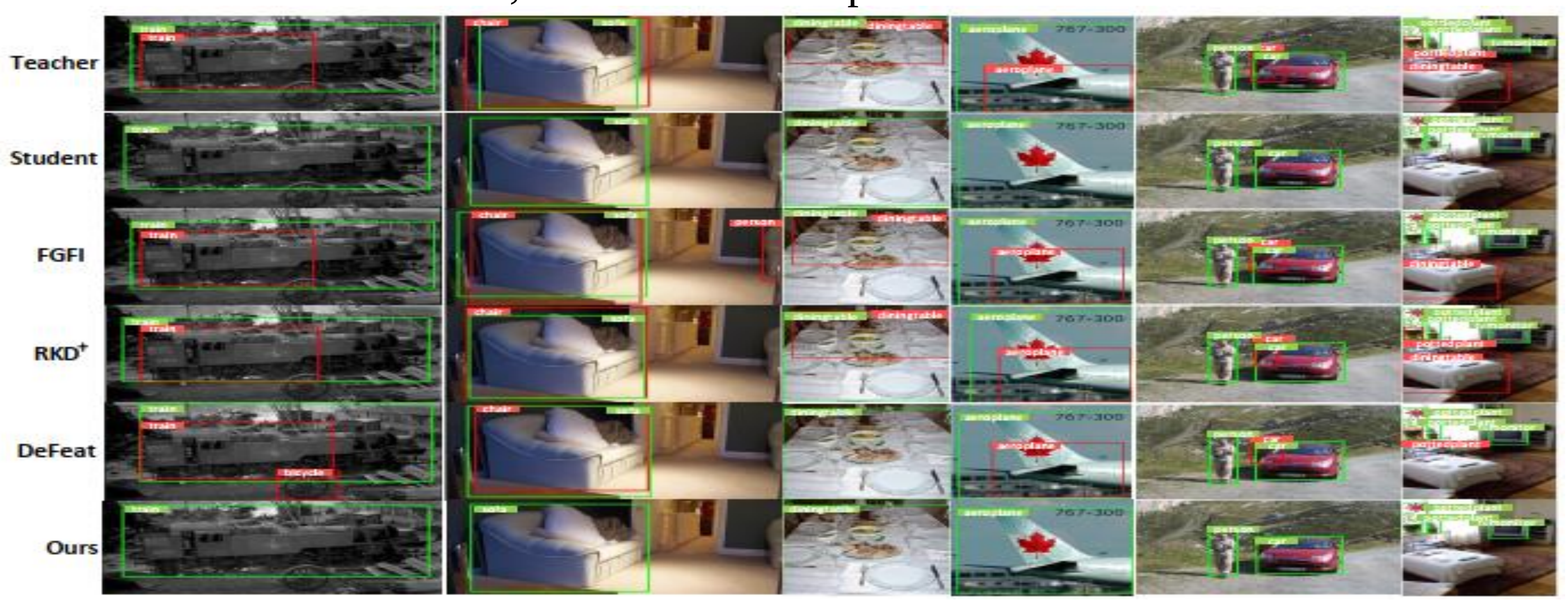
Sanli Tang, Zhongyu Zhang, Zhanzhan Cheng, Jing Lu, Yunlu Xu, Yi Niu, and Fan He

Hikvision Research Institute, Shanghai Jiao Tong University

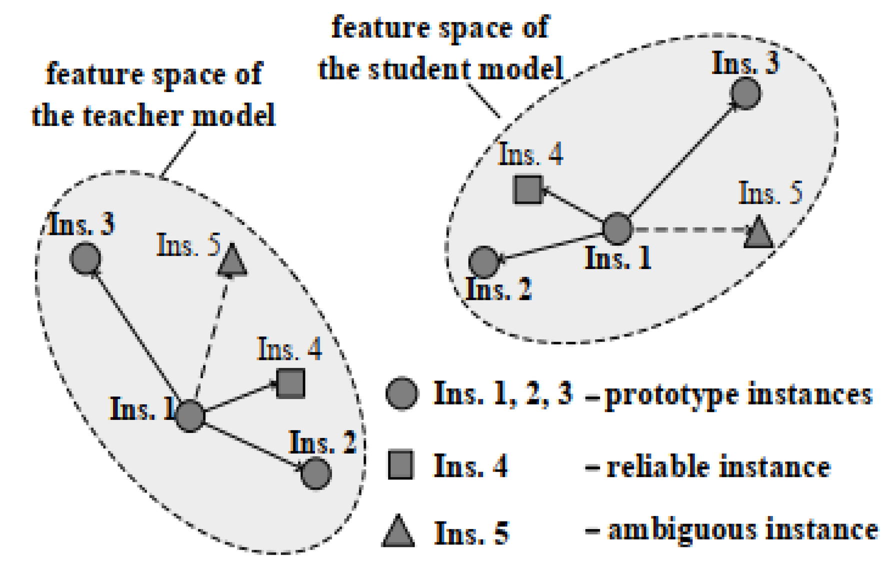
Document Audio Video Analysis & Recognition LAB

Motivation

- local knowledge is of much discrepancy between the teacher and the student in object detection tasks, especially on the ambiguous instances which are blur, truncated, or small.
- The distilling process will suffer from the noisy local knowledge, e.g., the false positives and the localization errors, and lead to sub-optimal.

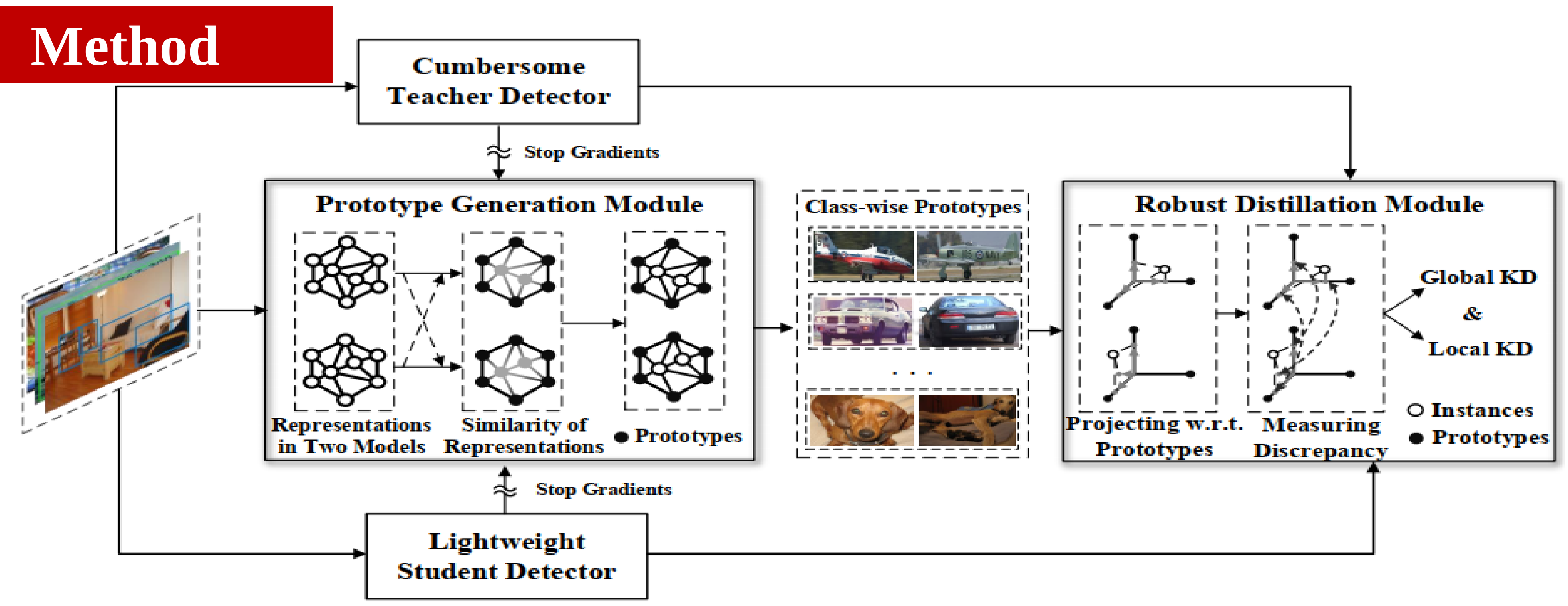


Idea



- Finding a group of common basis vectors in both the feature spaces of the teacher and the student detectors.
- Designing more robust distilling algorithm by measuring the discrepancy of the representations in the two feature spaces

Method



Algorithm 1 Algorithm for selecting prototypes in PGM.

Input:
$\{(\boldsymbol{f}_i^t, \boldsymbol{f}_i^s)\}_{i=1}^N$: features of N instances in *TS-space*;
Parameter:
K: number of prototypes to be selected;
λ: regularization weight;
Output:
$\mathcal{I}$: index set of the prototypes

initialize $n = 0$, the residuals $\boldsymbol{r}_{0,i}^t = \boldsymbol{f}_i^t$, and $\boldsymbol{r}_{0,i}^s = \boldsymbol{f}_i^s \quad \forall i = 1, \cdots, N$;
while $n < K$ **do**
  compute the optimal $w_{n+1,i}^s$ and $w_{n+1,i}^t$ by Eq. [5];
  compute the $\mathcal{L}_{n+1}^k$ with Eq. [4] for each instance by setting $\boldsymbol{g}_{n+1}^s = \boldsymbol{f}_k^s$ and $\boldsymbol{g}_{n+1}^t = \boldsymbol{f}_k^t, \forall k = 1, \cdots, N$;
  append the index k^* into $\mathcal{I}$ where $k^* = \arg\min_k \{\mathcal{L}_{n+1}^k\} \ \forall\ (\boldsymbol{g}_k^s, \boldsymbol{g}_k^t) \in \{(\boldsymbol{f}_i^s, \boldsymbol{f}_i^t)\}_{i=1}^N$; set $\boldsymbol{g}_{n+1}^t = \boldsymbol{f}_{k^*}^t$ and $\boldsymbol{g}_{n+1}^s = \boldsymbol{f}_{k^*}^s$;
  update the residuals $\boldsymbol{r}_{n+1,i}^t$ and $\boldsymbol{r}_{n+1,i}^s$ by Eq. [3];
  set $n = n + 1$;
end while
Return: $\mathcal{I}$

Contribution

(1)A prototype generation module (PGM) is designed to find a group of common basis vectors as the prototypes in the two feature spaces by minimizing reconstruction errors of instances w.r.t. prototypes.
(2) The global knowledge is formed by representing the instances under the prototypes, which shows a smaller gap between the two spaces.
(3) The proposed method achieves new remarkable performance on distilling both single-stage and two-stage detectors on Pascal VOC and COCO benchmarks..

Scan Me for Paper
(**ECCV-2022**)

Ablation

Method	Student	DeFeat [11]		RKD† [34]		Vanilla-KD† [16]	
+prototypes			✓		✓		✓
AP_{50}	81.3	82.0	**82.4**	81.6	**82.0**	81.8	**82.2**

Method	K-Means		DBSCAN [6]		Ambiguous	Ours
Features	Student	Teacher	Student	Teacher	-	Both
AP_{50}	82.3	82.1	82.4	82.2	81.8	**82.9**

Method	Backbone	mAP	AP_{50}	AP_{75}	AP_s	AP_m	AP_l
Cascade R-CNN (teacher)	ResNext101	47.3	66.3	51.7	28.2	51.7	62.7
Faster R-CNN (student)	ResNet50	38.4	59.0	42.0	21.5	42.1	50.3
FBKD [46]	ResNet50	**41.5**	**62.2**	**45.1**	**23.5**	45.0	55.3
Ours	ResNet50	**41.5**	61.9	**45.1**	**23.5**	**45.1**	**55.4**

- Prototypes can be easily combined by existing distilling methods and achieves better performance.
- Prototypes generated by proposed PGM shows better results than the cluster-like methods.
- It Shows competitive performance with much larger teacher.

Performance

Faster-Res101 (teacher)	39.8	60.1	43.3	22.5	43.6	52.8	53.0	32.8	56.9	68.6
Faster-Res50 (student)	38.4	59.0	42.0	21.5	42.1	50.3	52.0	32.6	55.8	66.1
FGFI [43]	39.3	59.8	42.9	22.5	42.3	52.2	52.4	32.2	55.7	67.9
DeFeat [11]	40.3	60.9	**44.0**	23.1	44.1	**53.4**	53.7	33.3	57.7	69.1
FBKD [46]	40.2	60.4	43.6	22.8	43.8	53.2	53.4	32.7	57.1	68.8
GID [4]	40.2	60.8	43.6	**23.6**	43.9	53.0	53.7	33.6	57.7	68.6
Ours	**40.6**	**61.0**	**44.0**	23.4	**44.4**	53.3	**53.8**	**33.9**	**57.9**	**69.2**
Retina-Res101 (teacher)	38.9	58.0	41.5	21.0	42.8	52.4	54.8	33.4	59.3	71.2
Retina-Res50 (student)	37.4	56.7	39.6	20.0	40.7	49.7	53.9	33.1	57.7	70.2
FGFI [43]	38.6	58.7	41.3	21.4	42.5	51.5	54.6	34.7	58.2	70.4
GID [4]	39.1	**59.0**	42.3	**22.8**	43.1	52.3	55.3	**36.7**	59.1	71.1
DeFeat [11]	39.3	58.2	42.1	21.7	42.9	52.9	55.1	33.9	59.6	71.5
FBKD [46]	39.3	58.8	42.0	21.2	43.2	53.0	55.4	34.6	59.7	72.2
FR [5]	39.3	58.8	42.0	21.5	43.3	52.6	-	-	-	-
PFI [23]	39.6	-	-	21.4	44.0	52.5	-	-	-	-
Ours	**39.8**	58.6	**42.6**	21.8	**43.5**	**53.5**	**55.8**	34.1	**60.0**	**72.2**