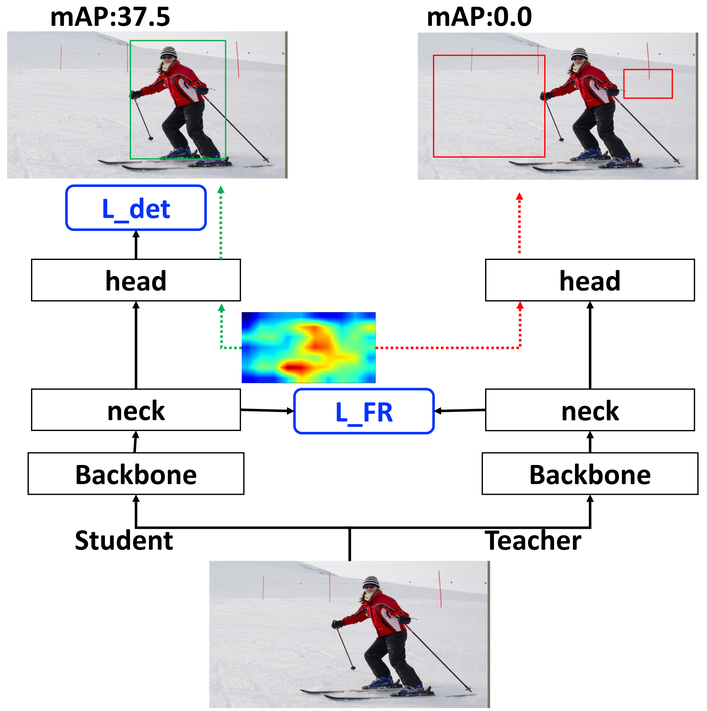
Dist2: Distribution-Guided Distillation for Object Detection

Tianchu Guo, Pengyu Li, Wei Liu, Bin Luo, Biao Wang
Alibaba Group, DAMO Academy

BMVC 2022

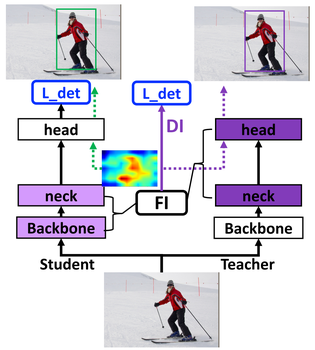
Motivations

- Ideal imitation
  - Student feature map = Teacher feature map
  - Student feature map + teacher head = correct detection result
- current state
  - The detection results are completely wrong fail if we forward the neck feature map of the student to the head of the teacher even after the feature-based and relation-based losses converge to a small value



Method

Overview



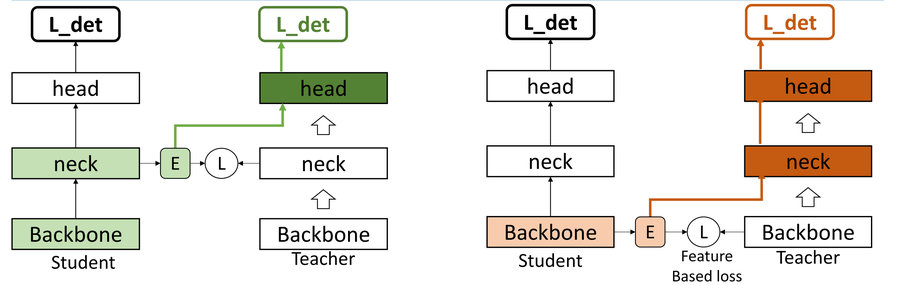
- directly forward the student's feature map to the teacher to make the final predictions.
- Optimized the prediction of the teacher.

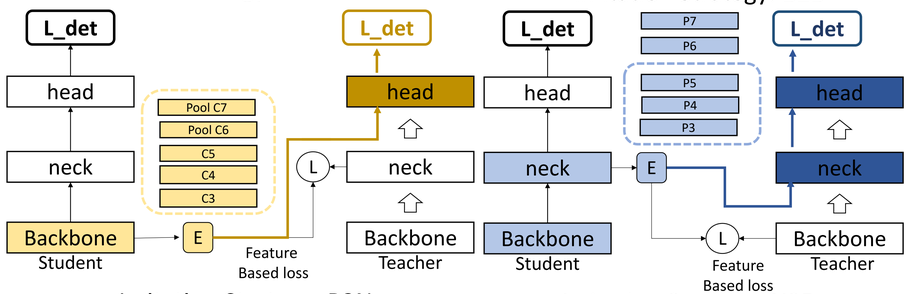
$$L_{det} = L_{cls}(P, P^{gt}) + L_{reg}(t, t^{gt})$$

$$L_{\text{feat}}(\text{feat}^{\text{S}}, \text{feat}^{\text{T}}) = \frac{1}{\text{K}} \left\| \text{f}_{\text{trans}}(\text{feat}^{\text{S}}) - \text{feat}^{\text{T}} \right\|_2^2$$

$$L = L_{det}(P, t) + \lambda_{feat} \sum_{is} L_{feat}(feat_{is}^{S}, feat_{is}^{T}) + \lambda_{DI} \sum_{is} DI^{is}$$

DI and FI Architecture Details





Results

Results on MS COCO 2017 dataset.

	Method	Imit.Strat.	mAP
FCOS	ResNeXt101(T)	-	42.7
	MobileV2C128(S)	-	30.4
	+Hint	N2N	$35.1^{+1.1}$
	+pa	N2N	$31.8^{+1.4}$
	+NonLocal	N2N	$33.3^{+1.9}$
	Ours:+DI	**N2N**	**$34.7^{+4.3}$**
	Ours:+Dist2	**All2All**	**$36.4^{+6.0}$**

	Method	Imit.Strat.	mAP
Retina	ResNeXt101(T)	-	41.1
	MobileV2C128(S)	-	31.0
	+Hint	N2N	$31.6^{+0.6}$
	+pa	N2N	$31.4^{+0.4}$
	+NonLocal	N2N	$31.9^{+0.9}$
	Ours:+DI	**N2N**	**$32.4^{+1.4}$**
	Ours:+Dist2	**All2All**	**$32.5^{+1.5}$**

Effect of each imitation strategy in FI.

FCOS ResNeXt101(42.7)-ResNet50C128(33.4)				
Imit.Strat.Num.	1	2	3	4
mAP	38.3	38.3	38.0	37.9

Effect of using more imitation strategies in FI.

FCOS ResNeXt101(42.7)-ResNet50C128(33.4)				
Imit.Strat.	N2N	B2B	B2N	N2B
+DI	$38.3^{+4.9}$	$38.3^{+4.9}$	$38.0^{+4.6}$	$37.9^{+4.5}$