Table 4: Ablation experiments quantifying the effectiveness of each component in our SSOD approach using 1% of COCO labels. The first row corresponds to the Soft Teacher baseline and the last row is our SoftER Teacher configuration.

Proposal Similarity Measure	Proposal IoU Regression	$AP_{50:95}$	$AR_{50:95}$
None	✗	22.4	30.8
KL-Divergence	✗	22.8	31.5
Cross-Entropy (Eq. (4))	✗	22.7	31.6
None	✓	22.3	30.8
KL-Divergence	✓	22.9	31.8
Cross-Entropy (Eq. (4))	✓	**23.0**	**32.0**

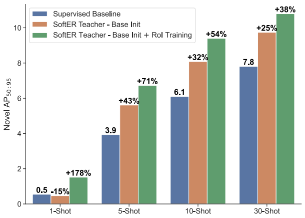

Figure 6: The impact of unlabeled data on semi-supervised few-shot fine-tuning.

Table 5: Ablation experiments evaluated on COCO `val2017` showing the standard procedure of fine-tuning both box classification and regression heads degrades base performance by as much as 21%. Our modified protocol of fine-tuning only the box classifier, while keeping the box regressor fixed, helps retain base detection accuracy with a performance drop of less than 11% for Faster R-CNN and 9% for SoftER Teacher.

Method	Base $AP_{50:95}$	Base $AP_{50:95}$ (60 Classes)				Novel $AP_{50:95}$ (20 Classes)			
		1-Shot	5-Shot	10-Shot	30-Shot	1-Shot	5-Shot	10-Shot	30-Shot
Faster R-CNN (fine-tune `cls+reg`)	39.3	31.2 (↓ 21%)	34.7 (↓ 12%)	34.8 (↓ 11%)	36.7 (↓ 7%)	0.6	3.9	6.0	7.9
Faster R-CNN (fine-tune `cls` only)	39.3	34.9 (↓ 11%)	35.8 (↓ 9%)	35.8 (↓ 9%)	37.1 (↓ 6%)	0.5	3.9	6.1	7.8
SoftER Teacher (fine-tune `cls+reg`)	42.0	33.6 (↓ 20%)	37.8 (↓ 10%)	38.1 (↓ 9%)	39.9 (↓ 5%)	1.5	6.7	9.4	10.8
SoftER Teacher (fine-tune `cls` only)	42.0	38.3 (↓ 9%)	39.1 (↓ 7%)	39.1 (↓ 7%)	40.2 (↓ 4%)	1.5	6.7	9.4	10.8

A Ablation Studies

A.1 SoftER Teacher System Design

Table 4 shows an ablation study on 1% of COCO labels to assess the key elements in our SoftER Teacher approach for SSOD. Compared to the Soft Teacher [55] baseline (first row), the addition of the cross-entropy or KL-divergence measure to enforce proposal consistency leads to a boost in both AP and AR, although the performance difference between the two measures is immaterial. Interestingly, the addition of the IoU regression loss by itself does not produce a performance improvement over the Soft Teacher baseline. However, when we couple IoU regression with the cross-entropy similarity measure, we obtain the best performing configuration (last row). ***SoftER Teacher improves on both precision and recall over the strong Soft Teacher baseline via our proposed Entropy Regression module for proposal learning with complex affine transformations.***

A.2 Semi-Supervised Few-Shot Fine-Tuning with Unlabeled Data

As discussed in Section 3.3, we explore two ways of leveraging unlabeled data to fine-tune the few-shot detector on novel classes: (1) we initialize the few-shot detector with parameters copied from the base *teacher* detector pre-trained with unlabeled data per Eq. (6); and (2) we further train the RoI box classifier and regressor on novel classes using the available few-shot and unlabeled examples while freezing the base backbone, FPN, and RPN components. Figure 6 illustrates semi-supervised base initialization boosts novel AP by as much as 43%, compared to the supervised baseline. In addition to semi-supervised base initialization, training the RoI head on few-shot novel classes with unlabeled images further amplifies the novel AP margin of SoftER Teacher.

A.3 To Freeze or Not to Freeze Box Regressor

The standard two-stage transfer learning procedure [50] fine-tunes the few-shot detector by updating both the RoI box classifier and regressor while keeping everything else frozen. Intuitively, we expect the RPN to produce accurate object regions during base pre-training, especially in the semi-supervised setting where it is further boosted by supplementary unlabeled images. We postulate that only the box classifier needs to be updated during fine-tuning to adapt base representations to novel concepts, and that fine-tuning the regression head is not necessary and may even hurt base performance. Table 5 verifies our intuition that fine-tuning both box classification and regression heads degrades base performance by as much as 21% on COCO `val2017`. By comparison, our modified protocol of fine-tuning only the box classifier helps retain base performance with a drop of less than 11%. Novel performance is unaffected between the two configurations. Our results are corroborated by

Table 6: Generalized FSOD results evaluated on VOC07 `test` over three random partitions. We compare our SoftER Teacher against its Soft Teacher counterpart and strong supervised baselines. We report the mean and 95% confidence interval over 10 random samples for our models. SoftER Teacher with ResNet-50 exceeds the supervised models with ResNet-101 by a large margin across most metrics under consideration.

VOC07 `test` – Split 1	Backbone	Base	Base	Base AP_{50} (15 Classes)			Novel AP_{50} (5 Classes)			Overall AP_{50} (20 Classes)		
Method		AP_{50}	AR_{50}	1-Shot	5-Shot	10-Shot	1-Shot	5-Shot	10-Shot	1-Shot	5-Shot	10-Shot
MPSR [53]	R-101	80.8	–	61.5	69.7	71.6	**42.8**	55.3	61.2	56.8	66.1	69.0
Retentive R-CNN [13]	R-101	80.8	–	80.9	80.8	80.8	42.4	53.7	56.1	71.3	74.0	74.6
TFA [50]	R-101	80.8	–	77.6 ± 0.2	77.4 ± 0.3	77.5 ± 0.2	25.3 ± 2.2	47.9 ± 1.2	52.8 ± 1.0	64.5 ± 0.6	70.1 ± 0.4	71.3 ± 0.3
Faster R-CNN (Our Impl.)	R-50	81.7	88.0	82.0 ± 0.2	82.4 ± 0.1	82.3 ± 0.1	27.9 ± 3.2	52.1 ± 2.1	58.2 ± 1.6	68.5 ± 0.8	74.9 ± 0.5	76.2 ± 0.4
Soft Teacher (Our Impl.)	R-50	85.3	91.2	84.5 ± 0.3	85.2 ± 0.1	85.2 ± 0.1	29.5 ± 4.2	56.2 ± 2.6	62.3 ± 1.8	70.8 ± 1.1	78.0 ± 0.7	79.5 ± 0.5
SoftER Teacher (Ours)	R-50	**85.9**	**92.5**	**84.5 ± 0.4**	**85.5 ± 0.1**	**85.5 ± 0.1**	31.6 ± 3.9	**57.7 ± 2.6**	**63.4 ± 1.7**	**71.3 ± 1.2**	**78.5 ± 0.7**	**80.0 ± 0.4**

VOC07 `test` – Split 2	Backbone	Base	Base	Base AP_{50} (15 Classes)			Novel AP_{50} (5 Classes)			Overall AP_{50} (20 Classes)		
Method		AP_{50}	AR_{50}	1-Shot	5-Shot	10-Shot	1-Shot	5-Shot	10-Shot	1-Shot	5-Shot	10-Shot
MPSR [53]	R-101	81.9	–	60.8	71.2	72.7	**29.8**	**43.2**	**47.0**	53.1	64.2	66.3
Retentive R-CNN [13]	R-101	81.9	–	81.8	81.9	81.9	21.7	37.0	40.3	66.8	70.7	71.5
TFA [50]	R-101	81.9	–	73.8 ± 0.8	76.2 ± 0.4	76.9 ± 0.3	18.3 ± 2.4	34.1 ± 1.4	39.5 ± 1.1	59.9 ± 0.8	65.7 ± 0.5	67.6 ± 0.4
Faster R-CNN (Our Impl.)	R-50	82.9	88.7	83.1 ± 0.1	83.5 ± 0.1	83.3 ± 0.1	18.3 ± 4.3	34.9 ± 1.5	40.6 ± 1.7	66.9 ± 1.1	71.4 ± 0.4	72.6 ± 0.4
Soft Teacher (Our Impl.)	R-50	85.9	91.7	**85.3 ± 0.1**	**85.8 ± 0.1**	**85.7 ± 0.1**	21.3 ± 4.4	39.4 ± 2.0	43.9 ± 1.7	**69.3 ± 1.1**	**74.2 ± 0.6**	75.3 ± 0.4
SoftER Teacher (Ours)	R-50	**86.1**	**92.9**	84.9 ± 0.2	85.6 ± 0.2	85.7 ± 0.2	21.9 ± 4.1	39.6 ± 1.7	45.0 ± 1.9	69.1 ± 1.1	74.1 ± 0.5	**75.5 ± 0.5**

VOC07 `test` – Split 3	Backbone	Base	Base	Base AP_{50} (15 Classes)			Novel AP_{50} (5 Classes)			Overall AP_{50} (20 Classes)		
Method		AP_{50}	AR_{50}	1-Shot	5-Shot	10-Shot	1-Shot	5-Shot	10-Shot	1-Shot	5-Shot	10-Shot
MPSR [53]	R-101	82.0	–	61.6	72.9	73.2	**35.9**	48.9	51.3	55.2	66.9	67.7
Retentive R-CNN [13]	R-101	82.0	–	81.9	82.0	82.1	30.2	**49.7**	50.1	69.0	73.9	74.1
TFA [50]	R-101	82.0	–	78.7 ± 0.2	78.5 ± 0.3	78.6 ± 0.2	17.9 ± 2.0	40.8 ± 1.4	45.6 ± 1.1	63.5 ± 0.6	69.1 ± 0.4	70.3 ± 0.4
Faster R-CNN (Our Impl.)	R-50	82.6	88.0	83.1 ± 0.2	83.6 ± 0.1	83.3 ± 0.1	19.6 ± 1.9	44.1 ± 1.8	51.2 ± 1.3	67.3 ± 0.5	73.7 ± 0.4	75.3 ± 0.3
Soft Teacher (Our Impl.)	R-50	85.6	91.3	**85.2 ± 0.2**	**85.5 ± 0.2**	**85.5 ± 0.1**	21.6 ± 1.6	46.4 ± 2.2	53.1 ± 1.3	**69.2 ± 0.5**	**75.7 ± 0.6**	**77.4 ± 0.3**
SoftER Teacher (Ours)	R-50	**85.7**	**92.5**	84.5 ± 0.2	85.2 ± 0.2	85.3 ± 0.1	22.4 ± 1.6	46.6 ± 2.1	**53.3 ± 1.6**	69.0 ± 0.5	75.6 ± 0.6	77.3 ± 0.5

existing work confirming that the main source of error with FSOD is indeed associated with the box classifier [14, 46]. Recall our goal for FSOD is to maximize novel detection accuracy while minimizing base performance degradation; keeping the box localization parameters fixed during fine-tuning is a simple and straight-forward way to help maintain base class accuracy.

B Additional Quantitative Results

B.1 Generalized Few-Shot Detection on PASCAL VOC

We present the generalized FSOD results on VOC in Table 6, which comprises three random partition splits. We report the ideal supervised base AP from previous work [13, 50] along with our substantially improved semi-supervised base AP to measure the extent of base forgetting. These results further support our observation on the trade-off between novel performance and base forgetting, for which our approach aims to simultaneously optimize. We summarize the following key takeaways.

Base Performance. Our re-implementation of the supervised Faster R-CNN baseline *does not* degrade base performance compared to the TFA benchmark across all three partitions. Base degradation is negligible with SoftER Teacher at less than 1.6%. We attribute this apparent improvement in base performance to our modified procedure of fine-tuning only the RoI box classifier and to our proposed Entropy Regression module enabling SoftER Teacher to achieve superior learning with unlabeled data.

SoftER Teacher *vs.* Supervised Baselines. SoftER Teacher with ResNet-50 surpasses the supervised MPSR, TFA, and Retentive R-CNN models with ResNet-101 by a large margin on the combined overall base + novel AP metric across most experiments under consideration, while being more parameter-efficient. Although MPSR achieves impressive few-shot performance on novel categories, it suffers catastrophic base forgetting by as much as 26%. Retentive R-CNN does not exhibit base class degradation, but generally falls behind on novel class performance.

SoftER Teacher *vs.* Soft Teacher. Both Soft Teacher and SoftER Teacher can harness unlabeled data to boost FSOD. However, we observe that a stronger semi-supervised detector leads to a more effective few-shot detector, with SoftER Teacher slightly edging out Soft Teacher on novel accuracy.

B.2 The Impact of Proposal Quality on Semi-Supervised Few-Shot Detection

We present expansive results on proposal quality and its relationship with semi-supervised few-shot detection in Table 7. Following existing literature [20, 49], we measure proposal quality using the metric AR@p, for $p \in \{100, 300, 1000\}$ proposals, averaged over 10 overlap thresholds between 0.5

Table 7: Proposal quality is highly correlated with semi-supervised few-shot detection. SoftER Teacher produces the best proposal quality AR@p, for $p \in \{100, 300, 1000\}$, among the comparisons, which in turn yields the strongest novel k-shot performances with varying fractions of base labels. All models are equipped with the ResNet-101 backbone. We report the mean and standard deviation over 5 random samples.

Method	% Labeled	AR@100	AR@300	AR@1000	Base $AP_{50:95}$ (60 Classes)			Novel $AP_{50:95}$ (20 Classes)			Overall $AP_{50:95}$ (80 Classes)		
					5-Shot	10-Shot	30-Shot	5-Shot	10-Shot	30-Shot	5-Shot	10-Shot	30-Shot
Faster R-CNN	1	17.3 ± 0.1	22.0 ± 0.2	27.0 ± 0.4	9.8 ± 0.3	10.0 ± 0.4	10.8 ± 0.3	1.9 ± 0.3	2.7 ± 0.1	3.5 ± 0.1	7.8 ± 0.2	8.2 ± 0.3	9.0 ± 0.2
Soft Teacher		27.8 ± 0.8	32.4 ± 0.8	38.1 ± 0.9	19.4 ± 0.7	19.9 ± 0.8	21.2 ± 0.7	5.9 ± 0.8	7.9 ± 0.7	10.1 ± 0.5	16.0 ± 0.6	16.9 ± 0.7	18.4 ± 0.6
SoftER Teacher		28.9 ± 0.7	33.7 ± 0.6	39.4 ± 0.6	19.2 ± 0.6	19.8 ± 0.5	21.1 ± 0.5	6.7 ± 0.3	8.8 ± 0.2	10.8 ± 0.5	16.1 ± 0.5	17.1 ± 0.4	18.5 ± 0.5
Faster R-CNN	5	23.3 ± 0.3	28.7 ± 0.4	34.9 ± 0.5	18.5 ± 0.5	18.9 ± 0.3	20.0 ± 0.5	3.5 ± 0.2	4.6 ± 0.2	5.9 ± 0.3	14.8 ± 0.4	15.3 ± 0.2	16.5 ± 0.4
Soft Teacher		29.8 ± 0.2	35.2 ± 0.2	41.4 ± 0.3	27.5 ± 0.4	27.8 ± 0.5	29.2 ± 0.5	6.7 ± 0.7	8.9 ± 0.4	11.1 ± 0.3	22.3 ± 0.4	23.1 ± 0.3	24.7 ± 0.4
SoftER Teacher		30.5 ± 0.2	35.9 ± 0.2	42.0 ± 0.2	27.5 ± 0.4	27.9 ± 0.4	29.3 ± 0.2	7.9 ± 0.4	10.1 ± 0.5	12.4 ± 0.5	22.6 ± 0.3	23.4 ± 0.3	25.1 ± 0.2
Faster R-CNN	10	25.0 ± 0.2	30.7 ± 0.3	37.5 ± 0.3	22.6 ± 0.4	22.8 ± 0.1	24.2 ± 0.2	3.8 ± 0.5	5.3 ± 0.2	6.8 ± 0.2	17.9 ± 0.3	18.4 ± 0.1	19.9 ± 0.2
Soft Teacher		30.2 ± 0.2	35.9 ± 0.2	42.4 ± 0.2	30.5 ± 0.5	30.7 ± 0.4	32.1 ± 0.3	6.8 ± 0.3	9.0 ± 0.6	11.4 ± 0.3	24.6 ± 0.4	25.3 ± 0.4	26.9 ± 0.3
SoftER Teacher		31.1 ± 0.2	36.7 ± 0.2	43.1 ± 0.3	30.3 ± 0.5	30.6 ± 0.5	32.0 ± 0.4	7.9 ± 1.3	10.4 ± 1.1	12.9 ± 1.0	24.6 ± 0.1	25.6 ± 0.3	27.2 ± 0.3

Table 8: SSOD results on VOC07 `test`. `VOC0712` denotes the combined VOC07+12 `trainval` splits. `COCO-20` is the subset of COCO data having the same 20 classes as VOC. SoftER Teacher outperforms Humble Teacher and Soft Teacher by a convincing margin.

Method	# Labels	Unlabeled	AP_{50}	$AP_{50:95}$	AR_{50}	$AR_{50:95}$
Supervised [47]	VOC07 (5k)	None	76.30	42.60	–	–
Supervised (Our Impl.)	VOC07 (5k)		**79.34**	**49.20**	**85.38**	**57.50**
Supervised [47]	VOC0712 (16k)	None	82.17	54.29	–	–
Supervised (Our Impl.)	VOC0712 (16k)		**84.53**	**57.77**	**89.73**	**65.73**
Humble Teacher [47]	VOC07 (5k)	VOC12	80.94	**53.04**	–	–
Soft Teacher (Our Impl.)			82.37	51.10	88.44	59.49
SoftER Teacher (Ours)			**83.10**	51.26	**89.74**	**60.19**
Humble Teacher [47]	VOC07 (5k)	VOC12 + COCO-20	81.29	54.41	–	–
Soft Teacher (Our Impl.)			82.50	54.47	87.14	62.45
SoftER Teacher (Ours)			**84.09**	**55.34**	**88.90**	**63.58**

Table 9: SSOD results on COCO `val2017`. The † setting refers to self-augmented supervised training without unlabeled data, and ‡ refers to the use of extra `unlabeled2017` images. We report the mean and standard deviation computed over 5 random samples.

COCO `val2017`	Average Precision ($AP_{50:95}$)				
Method	1%	5%	10%	†100%	‡100%
Supervised (Our Impl.)	10.57 ± 0.32	21.33 ± 0.40	26.80 ± 0.26	41.96	41.96
Humble Teacher [47]	16.96 ± 0.38	27.70 ± 0.15	31.61 ± 0.28	–	42.37
Soft Teacher (Our Impl.)	21.38 ± 1.02	30.65 ± 0.19	33.95 ± 0.13	43.51	44.08
SoftER Teacher (Ours)	$\mathbf{21.93 \pm 0.90}$	$\mathbf{31.15 \pm 0.29}$	$\mathbf{34.08 \pm 0.05}$	**43.54**	**44.22**

Method	Average Recall ($AR_{50:95}$)				
	1%	5%	10%	†100%	‡100%
Supervised (Our Impl.)	15.87 ± 0.45	29.07 ± 0.47	36.80 ± 0.46	55.64	55.64
Soft Teacher (Our Impl.)	29.85 ± 0.89	38.68 ± 0.28	43.48 ± 0.25	55.66	56.18
SoftER Teacher (Ours)	$\mathbf{30.90 \pm 0.88}$	$\mathbf{39.60 \pm 0.41}$	$\mathbf{43.90 \pm 0.55}$	**55.68**	**56.22**

and 0.95. Proposal quality AR@p is not to be confused with the detection metric $AR_{50:95}$, which is used to evaluate object coverage computed on a per-category basis and averaged over categories.

B.3 SoftER Teacher Improves Precision and Recall for Semi-Supervised Detection

We present SSOD results for VOC and COCO in Tables 8 and 9, respectively. On both datasets, we re-implement and re-train the supervised and Soft Teacher models for a direct comparison with SoftER Teacher. As part of our re-implementation, we make a conscientious effort to obtain high-quality supervised and Soft Teacher baselines by maximizing their performance output. This is to ensure that any performance boost demonstrated by SoftER Teacher is directly attributed to our entropy regression module for proposal learning with affine transforms.

In Table 8, we compare our best-case supervised baselines to those trained by Humble Teacher [47] and show that ours achieve significantly better detection accuracy. Even in the presence of strong supervised and Soft Teacher baselines, our SoftER Teacher model continues to improve upon its counterparts across almost all AP and AR metrics. Notably, our approach demonstrates superior learning with unlabeled data by narrowing the gap to less than 0.5 AP_{50} between the fully supervised model trained on VOC07+12 (16k labels) and SoftER Teacher trained on VOC07 (5k labels) augmented with unlabeled images from VOC12+COCO-20.

In Table 9, our model consistently outperforms its Soft Teacher counterpart over varying fractions of labeled data, although the impact of proposal learning in SoftER Teacher diminishes as the percentage of labeled data increases. We also experiment with 100% labels, *i.e.*, the entire `train2017` set, in two settings. In the first setting without unlabeled data, referred to as *self-augmented supervised training*, we use the `train2017` set as the source of "unlabeled data" to generate pseudo targets. And in the second setting, we supplement supervised training with `unlabeled2017` images. We observe that even without unlabeled data, SoftER Teacher improves on the supervised baseline by +1.6 AP, suggesting that more representations can still be learned from `train2017` alone. In the setting with additional unlabeled data, our model further boosts accuracy by another +0.7 AP.

Figure 7 illustrates exemplar detections from models trained on 1% of COCO labels, wherein our SoftER Teacher improves on both precision and recall over the comparisons.

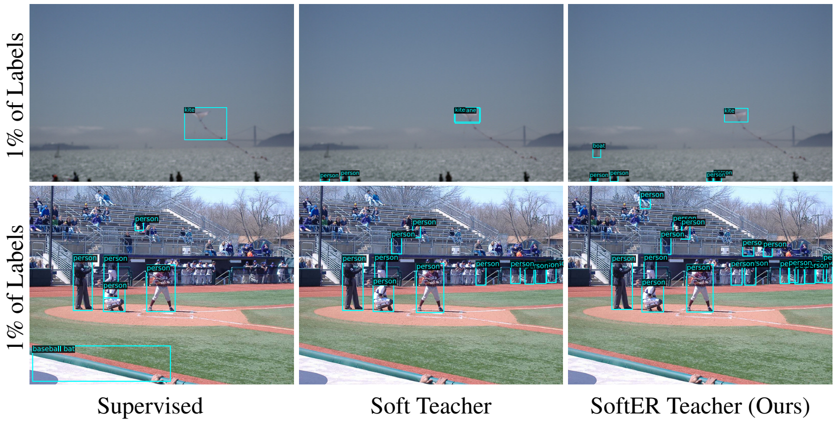

Figure 7: Qualitative detections on COCO `val2017` from models trained on 1% of labels. SoftER Teacher improves on both precision and recall, by recovering more missed objects while making fewer false positive detections, over its corresponding supervised and Soft Teacher counterparts. Best viewed digitally.

Table 10: FSOD results evaluated on COCO `val2017`. We report the mean and 95% confidence interval over 5 random samples for our models. SoftER Teacher with ResNet-50 surpasses TFA with ResNet-101 on both base and novel performances while also uniformly outperforming its Soft Teacher counterpart across all experiments.

COCO `val2017`	Backbone	Base $AP_{50:95}$	Base $AR_{50:95}$	Base $AP_{50:95}$ (60 Classes)				Novel $AP_{50:95}$ (20 Classes)			
Method				1-Shot	5-Shot	10-Shot	30-Shot	1-Shot	5-Shot	10-Shot	30-Shot
TFA w/cos [50]	R-101	39.3	–	31.9 ± 0.7	32.3 ± 0.6	32.4 ± 0.6	34.2 ± 0.4	1.9 ± 0.4	7.0 ± 0.7	9.1 ± 0.5	12.1 ± 0.4
Faster R-CNN (Our Impl.)	R-50	39.3	53.0	34.4 ± 0.6	33.1 ± 0.2	33.2 ± 0.2	35.1 ± 0.3	1.0 ± 0.3	5.1 ± 0.4	7.2 ± 0.4	9.6 ± 0.2
Soft Teacher (Our Impl.)	R-50	41.3	52.8	37.6 ± 0.4	38.0 ± 0.1	37.8 ± 0.3	39.2 ± 0.3	1.7 ± 0.9	6.7 ± 0.4	8.8 ± 0.5	11.2 ± 0.4
SoftER Teacher (Ours)	R-50	42.0	54.4	38.0 ± 0.4	38.4 ± 0.2	38.4 ± 0.2	39.7 ± 0.2	2.4 ± 0.6	8.2 ± 0.3	10.3 ± 0.5	12.9 ± 0.6
SoftER Teacher (Ours)	R-101	**44.4**	**56.1**	**40.7 ± 0.3**	**40.3 ± 0.2**	**40.2 ± 0.3**	**41.4 ± 0.2**	**2.8 ± 0.7**	**8.7 ± 0.6**	**11.0 ± 0.4**	**14.0 ± 0.6**

B.4 Generalized Few-Shot Detection on MS-COCO

We present additional FSOD results on the COCO dataset to include 1-shot detection in Table 10. Here, we observe more supporting evidence to strengthen our empirical finding on the potential relationship between SSOD and FSOD to suggest that a stronger semi-supervised detector leads to a more label-efficient few-shot detector. SoftER Teacher uniformly outperforms Soft Teacher across all metrics and experiments under consideration, most notably on novel class detection.

B.5 SoftER Teacher is Less Prone to Overfitting

We analyze the training behavior of Soft Teacher and SoftER Teacher for semi-supervised detection in Figure 8. For VOC, we train both models on VOC07 `trainval` labels with supplementary unlabeled images from VOC12+COCO-20. We observe from the validation curves that Soft Teacher seems to train faster than SoftER Teacher at the beginning, but has the propensity to overfit more than SoftER Teacher toward the end of training. For COCO, we train on 1% of labels sampled from `train2017` with the remaining 99% as unlabeled data. Similarly, we see from the validation curves that SoftER Teacher continues to improve even when Soft Teacher has reached its performance plateau. We attribute these characteristics to our entropy regression module for proposal learning, which provides SoftER Teacher a degree of robustness against overfitting.

C Implementation Details

C.1 Data Augmentation

We summarize the data augmentation strategy used to train Soft Teacher [55] and SoftER Teacher in Table 11. There are essentially three pipelines or branches of augmentation. The labeled branch uses random resizing and horizontal flipping along with color transformations. The student detector of the unlabeled branch undergoes the full complement of augmentations including strong affine geometric transformations and cutout [10, 57], akin to RandAugment [7], whereas the teacher detector leverages only weak resizing and horizontal flipping. At test time, we resize all instances to the

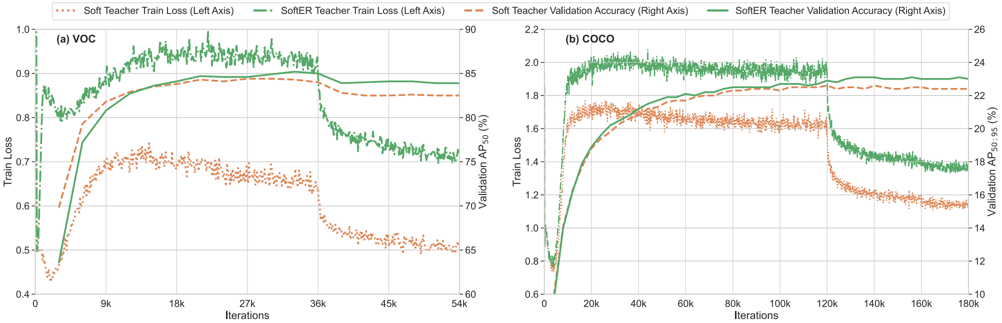

Figure 8: Visualization of training and validation behavior of Soft Teacher and SoftER Teacher on **(a)** VOC07 and **(b)** 1% of COCO labels. **Left:** The validation curve of Soft Teacher tends to overfit more than SoftER Teacher toward the end of training on VOC. **Right:** SoftER Teacher continues to improve even when Soft Teacher has reached its validation performance plateau at the 120k iterations mark.

Table 11: Summary of the data augmentation pipelines used to train Soft Teacher and SoftER Teacher. **Left:** transformations applied to the student trained on labeled data. **Middle:** strong augmentation pipeline applied to the student trained on unlabeled data. **Right:** weak augmentation pipeline applied to the teacher trained on unlabeled data.

Augmentation	Student Labeled Branch	Student Unlabeled Branch (Strong)	Teacher Unlabeled Branch (Weak)
Resize	short edge $\in [400, 1200]$	short edge $\in [400, 1200]$	short edge $\in [400, 1200]$
Flip	$p = 0.5$, horizontal	$p = 0.5$, horizontal	$p = 0.5$, horizontal
Identity	$p = 1/9$	$p = 1/9$	
AutoContrast	$p = 1/9$	$p = 1/9$	
Equalize	$p = 1/9$	$p = 1/9$	
Solarize	$p = 1/9$	$p = 1/9$	
Color	$p = 1/9$	$p = 1/9$	
Contrast	$p = 1/9$	$p = 1/9$	
Brightness	$p = 1/9$	$p = 1/9$	
Sharpness	$p = 1/9$	$p = 1/9$	
Posterize	$p = 1/9$	$p = 1/9$	
Translation		$p = 1/3$, $(x, y) \in (-0.1, 0.1)$	
Shearing		$p = 1/3$, $(x, y) \in (-30^\circ, 30^\circ)$	
Rotation		$p = 1/3$, angle $\in (-30^\circ, 30^\circ)$	
Cutout		$n \in [1, 5]$, size $\in [0.0, 0.2]$	

shorter side of 800 resolution, but otherwise do not perform any test-time augmentation, following standard supervised [40] and semi-supervised [32, 45, 47, 55] protocols.

C.2 Supervised and Semi-Supervised Training

High-Quality Baselines. Following existing literature [32, 45, 47, 55], we evaluate our approach for semi-supervised detection on VOC and COCO 2017 datasets. On both datasets, we re-implement and re-train the supervised Faster R-CNN and Soft Teacher[1] models for a direct comparison with SoftER Teacher. As part of our re-implementation, we make a conscientious effort to obtain the best-case supervised and Soft Teacher baselines by maximizing their performance output. We train the strong supervised baseline by using a longer training schedule (see Tables 12 and 13) and applying diverse color augmentations in addition to random resizing and horizontal flipping (see Table 11). And we re-train Soft Teacher exactly as is according to the authors' source code. This is to ensure that any performance boost demonstrated by SoftER Teacher is directly attributed to our entropy regression module for learning representations from region proposals, and not to changes in model configuration and training protocol.

VOC Evaluation. We experiment with two supervised settings: (1) using VOC07 `trainval` split as labeled data, and (2) utilizing the joint VOC07+12 labeled set as an upper bound for supervised detection performance. We also have two semi-supervised settings: (1) augmenting supervised

[1] We leverage the original authors' source code made publicly available at https://github.com/microsoft/SoftTeacher

Table 12: Supervised and semi-supervised training protocols on PASCAL VOC. `COCO-20` is the subset of COCO data containing objects with the same 20 category names as VOC objects. `Sample Ratio` denotes the blend of (labeled, unlabeled) examples in a mini-batch. All settings are configured for 8× multi-GPU training.

Method	Labeled	Unlabeled	Batch Size	Sample Ratio	lr	lr Step	Iterations
Supervised	VOC07	None	16	(16, 0)	0.02	(12k, 16k)	18k
Supervised	VOC0712		16	(16, 0)	0.02	(36k, 48k)	54k
Soft Teacher	VOC07	VOC12	64	(32, 32)	0.01	(12k, 16k)	18k
SoftER Teacher			64	(32, 32)	0.01	(12k, 16k)	18k
Soft Teacher	VOC07	VOC12+ COCO-20	64	(32, 32)	0.01	(36k, 48k)	54k
SoftER Teacher			64	(32, 32)	0.01	(36k, 48k)	54k
Soft Teacher	VOC0712	COCO-20	64	(32, 32)	0.01	(40k, 52k)	60k
SoftER Teacher			64	(32, 32)	0.01	(40k, 52k)	60k

Table 13: Supervised and semi-supervised training protocols on COCO 2017. The † setting refers to self-augmented supervised training without unlabeled data, and ‡ corresponds to the use of supplementary `unlabeled2017` images. `Sample Ratio` denotes the blend of (labeled, unlabeled) examples in a mini-batch. All settings are configured for 8× multi-GPU training.

% Labeled	Method	Batch Size	Sample Ratio	lr	lr Step	Iterations
1	Supervised	8	(8, 0)	0.01	(120k, 160k)	180k
	Soft Teacher	40	(8, 32)	0.01	(120k, 160k)	180k
	SoftER Teacher	40	(8, 32)	0.01	(120k, 160k)	180k
5	Supervised	8	(8, 0)	0.01	(120k, 160k)	180k
	Soft Teacher	40	(8, 32)	0.01	(120k, 160k)	180k
	SoftER Teacher	40	(8, 32)	0.01	(120k, 160k)	180k
10	Supervised	8	(8, 0)	0.01	(120k, 160k)	180k
	Soft Teacher	40	(8, 32)	0.01	(120k, 160k)	180k
	SoftER Teacher	40	(8, 32)	0.01	(120k, 160k)	180k
†100	Supervised	16	(16, 0)	0.02	(480k, 640k)	720k
	Soft Teacher	64	(32, 32)	0.01	(480k, 640k)	720k
	SoftER Teacher	64	(32, 32)	0.01	(480k, 640k)	720k
‡100	Supervised	16	(16, 0)	0.02	(480k, 640k)	720k
	Soft Teacher	64	(32, 32)	0.01	(480k, 640k)	720k
	SoftER Teacher	64	(32, 32)	0.01	(480k, 640k)	720k

training on VOC07 with VOC12 as unlabeled data, and (2) leveraging the combined VOC12+COCO-20 as unlabeled data. `COCO-20` is the subset of COCO `train2017` having the same 20 category names as VOC objects. Model performance is evaluated on the VOC07 `test` set. Detailed comparative results are given in Table 8.

COCO Evaluation. There are three experimental settings: (1) *Partially labeled*, where we train on $\{1, 5, 10\}$ percent of labels randomly sampled from the `train2017` split while treating the remaining images as unlabeled data. (2) *Fully labeled*, where we leverage the extra 123k images from the `unlabeled2017` set to supplement supervised training on the entire `train2017`. And (3) *Self-augmented supervised training*, where we apply the `train2017` set, discarding all label information, as the source of "unlabeled" data to generate unsupervised pseudo targets. To our knowledge, we are the first to conduct this experiment for semi-supervised detection. For each setting, we also train on the labeled portion alone, without using unlabeled data, to establish the lower-bound supervised baseline. Model performance is evaluated on the `val2017` set. See Table 9 for comparative results.

Top-N Proposals. To learn representations on region proposals, we extract the top 512 proposals, after non-maximum suppression, from each unlabeled image as generated by the student's RPN. Our motivation for selecting the top 512 proposals is to balance the trade-off among accuracy performance, memory requirements, and training duration. Moreover, our choice of $N = 512$ is consistent with $N = 640$ proposals empirically found by Humble Teacher [47] to be an optimal number with regards to detection accuracy.

Training Parameters. We summarize our training protocols on VOC and COCO in Tables 12 and 13 for the supervised, Soft Teacher, and SoftER Teacher models. In general, Soft Teacher and our SoftER Teacher share the same configuration to ensure we can directly measure the impact of proposal learning and its contribution to detection accuracy. All hyper-parameters related to Soft Teacher remain the same, including the EMA momentum, which defaults to 0.999 following common practice in the semi-supervised classification literature [44, 48]. We train our models using vanilla SGD optimization with momentum and weight decay set to 0.9 and 0.0001, respectively. We train on 8× A6000 GPUs each with 48GB of memory. One experiment takes between 12 hours and 10 days to complete, depending on the scope. At test time, we extract the teacher model from the final check-point for evaluation.

C.3 Semi-Supervised Few-Shot Training

In this section, we expound on our protocol for semi-supervised few-shot training on VOC and COCO datasets. We conduct our few-shot experiments on the same VOC and COCO samples provided by the TFA benchmark [50]. The VOC dataset is randomly partitioned into 15 base and 5 novel classes, where there are $k \in \{1, 5, 10\}$ labeled boxes per category sampled from the combined VOC07+12 `trainval` splits. This process is repeated three times to create three partitions. And the COCO dataset is divided into 60 base and 20 novel classes having the same VOC category names with $k \in \{1, 5, 10, 30\}$ shots. We leverage `COCO-20` as the source of external unlabeled data to supplement few-shot training on VOC, and `unlabeled2017` images to augment few-shot experiments on COCO.

Table 14: Protocol for few-shot fine-tuning on VOC and COCO datasets. All settings are configured for $8\times$ multi-GPU training.

# Shot	Parameter	VOC07+12	COCO 2017
1	Batch Size	16	16
	lr	0.001	0.001
	lr Step	9k	14k
	Iterations	10k	16k
	Fine-Tune Layer	`cls+reg`	`cls`
5	Batch Size	16	16
	lr	0.001	0.001
	lr Step	18k	72k
	Iterations	20k	80k
	Fine-Tune Layer	`cls+reg`	`cls`
10	Batch Size	16	16
	lr	0.001	0.001
	lr Step	36k	144k
	Iterations	40k	160k
	Fine-Tune Layer	`cls+reg`	`cls`
30	Batch Size	–	16
	lr	–	0.001
	lr Step	–	216k
	Iterations	–	240k
	Fine-Tune Layer	–	`cls`

Semi-Supervised Base Pre-Training. In the first stage, we train a base detector on base classes, along with the available unlabeled data, according to the formulation described in Section 3.2. For the supervised base detector, we equip Faster R-CNN with the ResNet-101 [18] backbone. For the semi-supervised base detectors, we experiment with Soft Teacher and our proposed SoftER Teacher using the same ResNet-101 backbone. In some experiments, we also employ ResNet-50 to explore parameter-efficient learning with SoftER Teacher. Our motivation for leveraging unlabeled data in the base pre-training step is two-fold: first, we demonstrate the versatility of our approach by not strictly depending on an abundance of base classes. Second, we observe impressive results in the SSOD literature that show unlabeled data can consistently and significantly boost detection performance. Intuitively, any performance gains during semi-supervised base pre-training with unlabeled data should have the potential to boost few-shot detection in the fine-tuning step.

Semi-Supervised Few-Shot Fine-Tuning. In the second stage, we combine the parameters of the (semi-supervised) base detector with those of the novel detector into the overall few-shot detector and fine-tune it on a small balanced training set of k shots per class containing both base and novel examples. Before fine-tuning, we obtain the parameters of the novel detector in two ways. For VOC, we initialize the parameters of the novel classifier and regressor with normally distributed random values, analogous to TFA. For the COCO dataset, we reuse the base model pre-trained in the first stage, but further train the detector head from scrach on novel classes. We optimize the novel detector on both few-shot and unlabeled examples according to the semi-supervised protocols. At the fine-tuning step, we update only the RoI box classifier of the few-shot detector while freezing all other components, including the box regressor. We justify our decision to freeze the RoI box regressor with an ablation study in Appendix A. Table 14 summarizes our few-shot fine-tuning protocol.

D Limitations and Future Work

Although SoftER Teacher demonstrates superior generalized FSOD performance with unlabeled data, there is still much room for improvement. We observe complementary properties of DCFS [14] and Retentive R-CNN [13] which can be combined with SoftER Teacher to further advance FSOD without base degradation. Moreover, it would be inspiring to see how far FSOD can go by integrating unlabeled data with the latest advances in Vision Transformers [4, 11]. Lastly, it would be interesting direction for future work to investigate if our empirical finding connecting SSOD with FSOD can be extended to other SSOD formulations including one-stage detectors, such as the recently introduced Consistent Teacher [51] and Unbiased Teacher v2 [33] detectors.

E Additional Qualitative Results

We present additional visualizations of student and teacher proposals in Figure 9. The student undergoes a wide spectrum of scale, color, and geometric transformations, whereas the teacher receives weakly augmented images as the basis for generating reliable unsupervised pseudo targets to regularize the student's learning trajectory. This multi-stream data augmentation strategy enables the student to tap into abundant region proposals to capture diverse feature representations that would otherwise be lost with aggressive confidence thresholding associated with pseudo-labeling.

Figure 10 illustrates additional qualitative detections from models trained on $\{1, 5, 10\}$ percent of labels sampled from COCO `train2017`. As corroborated by quantitative results, SoftER Teacher improves on both precision and recall over the supervised and Soft Teacher counterparts by recovering

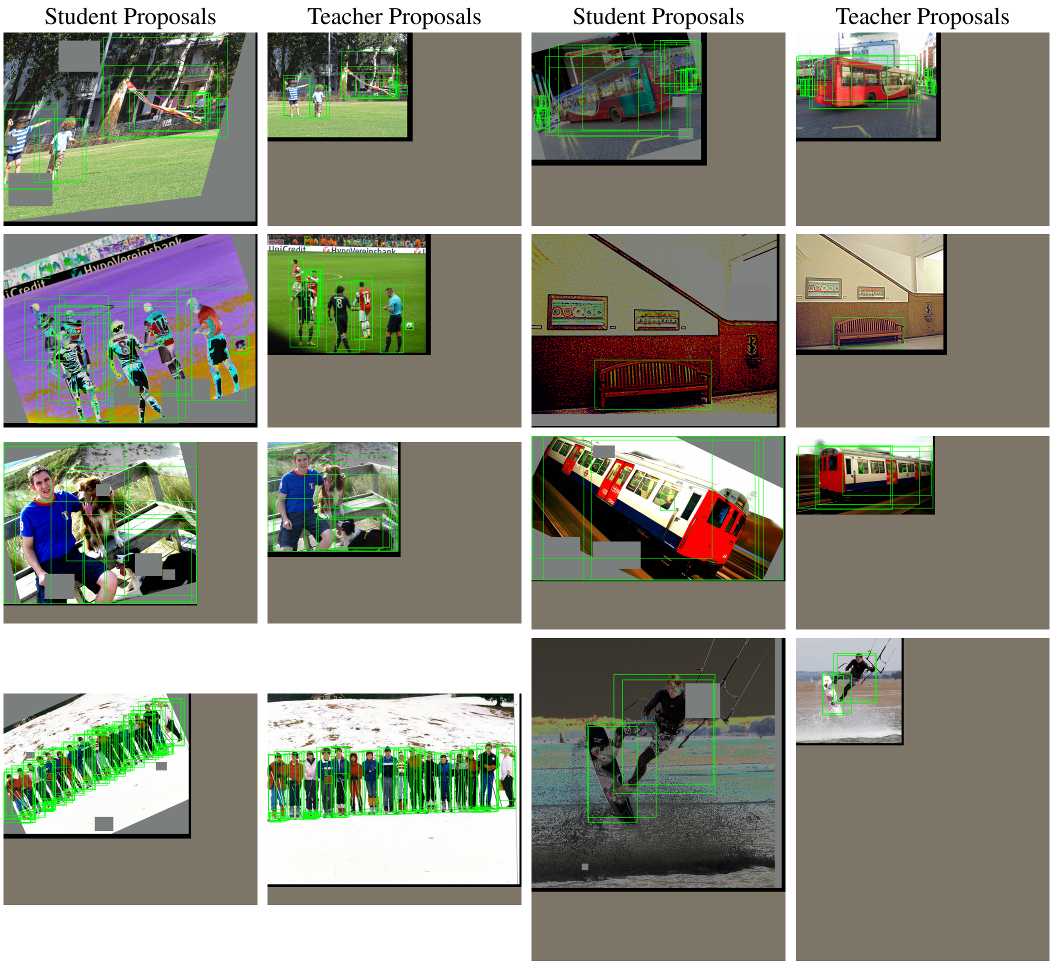

Figure 9: Visualizations of student and teacher proposals with confidence scores greater than 0.99. The student images are subjected to a wide range of complex scale, color, and geometric distortions, whereas the teacher images undergo simple random resizing and horizontal flipping. A pair of student-teacher proposals is aligned between student and teacher images for the purpose of enforcing classification similarity and localization consistency. Best viewed digitally.

more missed objects while making fewer false positive detections. The enhancements over the strong Soft Teacher baseline are especially pronounced in low-label settings and in crowded scenes with small and ambiguous objects, which is the intended benefit specifically designed into SoftER Teacher by way of our entropy regularization module for proposal learning.

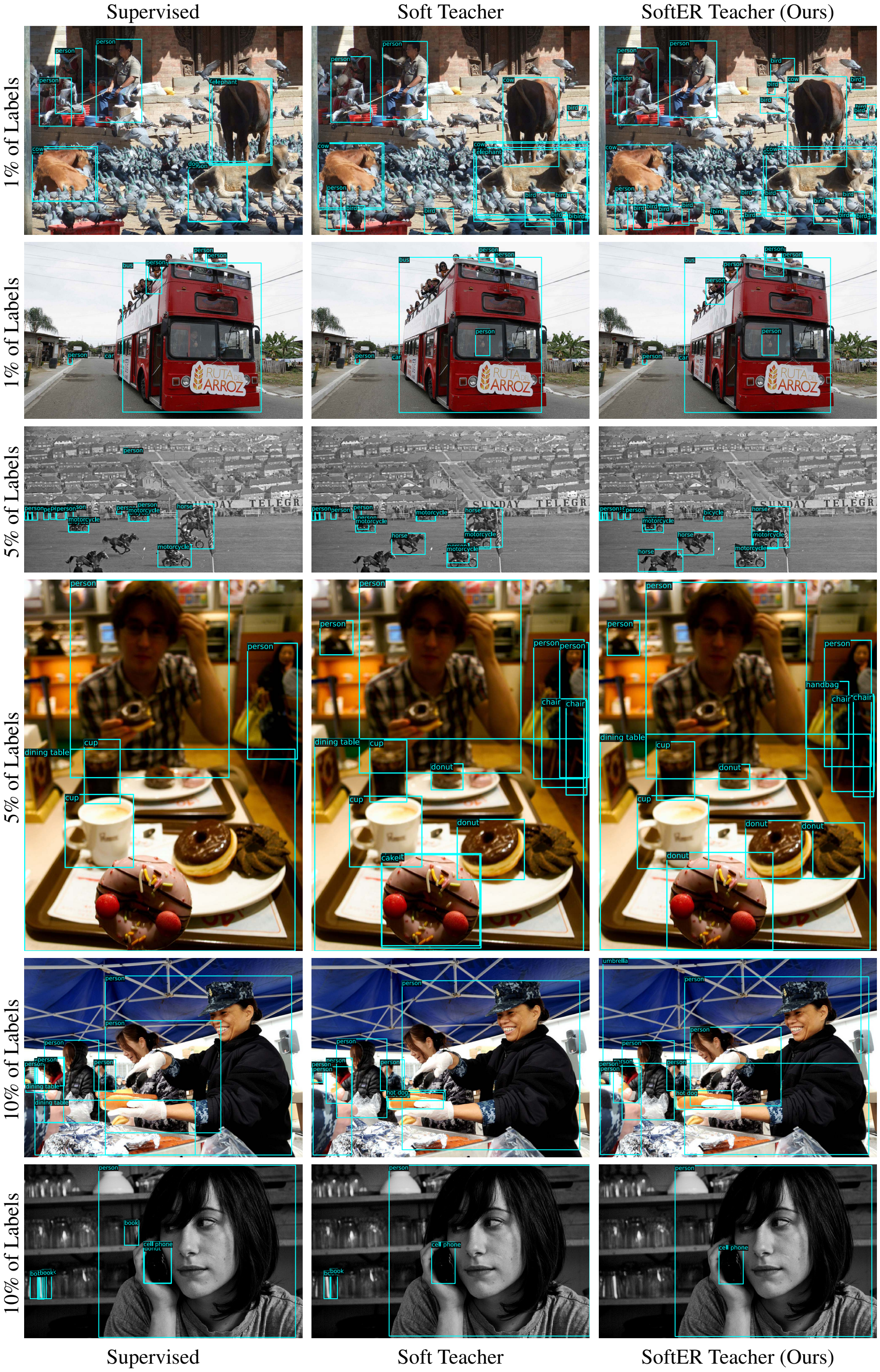

Figure 10: Exemplar detections from models trained on $\{1, 5, 10\}$ percent of labels sampled from COCO `train2017` and visualized on `val2017`. SoftER Teacher captures more object coverage while making fewer false positive detections than its supervised and Soft Teacher counterparts. The enhancements over Soft Teacher are especially pronounced in crowded scenes with small and ambiguous objects. Best viewed digitally.

References

[1] Philip Bachman, Ouais Alsharif, and Doina Precup. Learning with Pseudo-Ensembles. In *NeurIPS*, 2014. 5

[2] Amir Bar, Xin Wang, Vadim Kantorov, Colorado J. Reed, Roei Herzig, Gal Chechik, Anna Rohrbach, Trevor Darrell, and Amir Globerson. DETReg: Unsupervised Pretraining With Region Priors for Object Detection. In *CVPR*, 2022. 1

[3] Yuhang Cao, Jiaqi Wang, Ying Jin, Tong Wu, Kai Chen, Ziwei Liu, and Dahua Lin. Few-Shot Object Detection via Association and Discrimination. In *NeurIPS*, 2021. 3

[4] Nicolas Carion, Francisco Massa, Gabriel Synnaeve, Nicolas Usunier, Alexander Kirillov, and Sergey Zagoruyko. End-to-End Object Detection with Transformers. In *ECCV*, 2020. 19

[5] Kai Chen, Jiaqi Wang, Jiangmiao Pang, Yuhang Cao, Yu Xiong, Xiaoxiao Li, et al. MMDetection: Open MMLab Detection Toolbox and Benchmark. https://arxiv.org/abs/1906.07155, 2019. 7

[6] Xinlei Chen and Kaiming He. Exploring Simple Siamese Representation Learning. In *CVPR*, 2021. 7

[7] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. RandAugment: Practical Automated Data Augmentation with a Reduced Search Space. In *NeurIPS*, 2020. 5, 16

[8] Zhigang Dai, Bolun Cai, Yugeng Lin, and Junying Chen. UPDETR: Unsupervised Pre-Training for Object Detection with Transformers. In *CVPR*, 2021. 1

[9] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR*, pages 248–255, 2009. 7

[10] Terrance DeVries and Graham W. Taylor. Improved Regularization of Convolutional Neural Networks with Cutout. https://arxiv.org/abs/1708.04552, 2017. 16

[11] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. In *ICLR*, 2021. 19

[12] Mark Everingham, Luc Van Gool, Christopher K.I. Williams, John Winn, and Andrew Zisserman. The PASCAL Visual Object Classes (VOC) Challenge. *IJCV*, 88(2):303–338, 2010. 7

[13] Zhibo Fan, Yuchen Ma, Zeming Li, and Jian Sun. Generalized Few-Shot Object Detection without Forgetting. In *CVPR*, 2021. 1, 2, 3, 4, 8, 14, 19

[14] Bin-Bin Gao, Xiaochen Chen, Zhongyi Huang, Congchong Nie, Jun Liu, Jinxiang Lai, Guannan Jiang, Xi Wang, and Chengjie Wang. Decoupling Classifier for Boosting Few-Shot Object Detection and Instance Segmentation. In *NeurIPS*, 2022. 1, 2, 3, 7, 8, 14, 19

[15] Yves Grandvalet and Yoshua Bengio. Semi-Supervised Learning by Entropy Minimization. In *NeurIPS*, 2004. 7

[16] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre H. Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, et al. Bootstrap Your Own Latent: A New Approach to Self-Supervised Learning. In *NeurIPS*, 2020. 7

[17] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *ICCV*, 2017. 4

[18] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *CVPR*, 2016. 4, 19

[19] Geoffrey Hinton, Oriol Vinyals, and Jeffrey Dean. Distilling the Knowledge in a Neural Network. In *NeurIPS Deep Learning and Representation Learning Workshop*, 2015. 2

[20] Jan Hosang, Rodrigo Benenson, Piotr Dollár, and Bernt Schiele. What Makes for Effective Detection Proposals? *IEEE TPAMI*, 38(4):814–830, 2016. 2, 4, 8, 14

[21] Jisoo Jeong, Seungeui Lee, Jeesoo Kim, and Nojun Kwak. Consistency-Based Semi-Supervised Learning for Object Detection. In *NeurIPS*, 2019. 2

[22] Bingyi Kang, Zhuang Liu, Xin Wang, Fisher Yu, Jiashi Feng, and Trevor Darrell. Few-Shot Object Detection via Feature Reweighting. In *ICCV*, 2019. 3, 4

[23] Leonid Karlinsky, Joseph Shtok, Sivan Harary, Eli Schwartz, Amit Aides, Rogerio Feris, Raja Giryes, and Alex M. Bronstein. RepMet: Representative-Based Metric Learning for Classification and Few-Shot Object Detection. In *CVPR*, 2019. 4

[24] Prannay Kaul, Weidi Xie, and Andrew Zisserman. Label, Verify, Correct: A Simple Few Shot Object Detection Method. In *CVPR*, 2022. 1, 2, 3, 4, 8

[25] Siddhesh Khandelwal, Raghav Goyal, and Leonid Sigal. UniT: Unified Knowledge Transfer for Any-shot Object Detection and Segmentation. In *CVPR*, 2021. 3

[26] Samuli Laine and Timo Aila. Temporal Ensembling for Semi-Supervised Learning. In *ICLR*, 2017. 5

[27] Jianan Li, Xiaodan Liang, Yunchao Wei, Tingfa Xu, Jiashi Feng, and Shuicheng Yan. Perceptual Generative Adversarial Networks for Small Object Detection. In *CVPR*, 2017. 7

[28] Yangguang Li, Feng Liang, Lichen Zhao, Yufeng Cui, Wanli Ouyang, Jing Shao, Fengwei Yu, and Junjie Yan. Supervision Exists Everywhere: A Data Efficient Contrastive Language-Image Pre-Training Paradigm. In *ICLR*, 2022. 1

[29] Zeming Li, Chao Peng, Gang Yu, Xiangyu Zhang, Yangdong Deng, and Jian Sun. DetNet: A Backbone Network for Object Detection. In *ECCV*, 2018. 6, 7

[30] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature Pyramid Networks for Object Detection. In *CVPR*, 2017. 4

[31] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In *ECCV*, 2014. 1, 7

[32] Yen-Cheng Liu, Chih-Yao Ma, Zijian He, Chia-Wen Kuo, Kan Chen, Peizhao Zhang, Bichen Wu, Zsolt Kira, and Peter Vajda. Unbiased Teacher for Semi-Supervised Object Detection. In *ICLR*, 2021. 3, 17

[33] Yen-Cheng Liu, Chih-Yao Ma, and Zsolt Kira. Unbiased Teacher v2: Semi-Supervised Object Detection for Anchor-Free and Anchor-Based Detectors. In *CVPR*, 2022. 1, 19

[34] David Lopez-Paz and Marc'Aurelio Ranzato. Gradient Episodic Memory for Continual Learning. In *NeurIPS*, 2017. 2

[35] Matthias Minderer, Alexey Gritsenko, Austin Stone, Maxim Neumann, Dirk Weissenborn, Alexey Dosovitskiy, Aravindh Mahendran, Anurag Arnab, et al. Simple Open-Vocabulary Object Detection with Vision Transformers. In *ECCV*, 2022. 1

[36] Takeru Miyato, Shin-ichi Maeda, Masanori Koyama, and Shin Ishii. Virtual Adversarial Training: A Regularization Method for Supervised and Semi-Supervised Learning. *IEEE TPAMI*, 41:1979–1993, 2017. 5, 7

[37] Avital Oliver, Augustus Odena, Colin Raffel, Ekin D. Cubuk, and Ian J. Goodfellow. Realistic Evaluation of Deep Semi-Supervised Learning Algorithms. In *NeurIPS*, 2018. 7

[38] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *NeurIPS*, pages 8024–8035. Curran Associates, Inc., 2019. 7

[39] Limeng Qiao, Yuxuan Zhao, Zhiyuan Li, Xi Qiu, Jianan Wu, and Chi Zhang. DeFRCN: Decoupled Faster R-CNN for Few-Shot Object Detection. In *ICCV*, 2021. 1, 2, 3, 8

[40] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *NeurIPS*, 2015. 4, 17

[41] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. Generalized Intersection over Union: A Metric and A Loss for Bounding Box Regression. In *CVPR*, 2019. 6, 7

[42] Byungseok Roh, Wuhyun Shin, Ildoo Kim, and Sungwoong Kim. Spatially Consistent Representation Learning. In *CVPR*, 2021. 1

[43] Mehdi Sajjadi, Mehran Javanmardi, and Tolga Tasdizen. Regularization with Stochastic Perturbations for Deep Semi-Supervised Learning. In *NeurIPS*, 2016. 2

[44] Kihyuk Sohn, David Berthelot, Chun-Liang Li, Zizhao Zhang, Nicholas Carlini, Ekin D Cubuk, Alex Kurakin, Han Zhang, and Colin Raffel. Fixmatch: Simplifying Semi-Supervised Learning with Consistency and Confidence. In *NeurIPS*, 2020. 5, 18

[45] Kihyuk Sohn, Zizhao Zhang, Chun-Liang Li, Han Zhang, Chen-Yu Lee, and Tomas Pfister. A Simple Semi-Supervised Learning Framework for Object Detection. https://arxiv.org/abs/2005.04757, 2020. 3, 5, 17

[46] Bo Sun, Banghuai Li, Shengcai Cai, Ye Yuan, and Chi Zhang. FSCE: Few-Shot Object Detection via Contrastive Proposal Encoding. In *CVPR*, 2021. 3, 5, 7, 14

[47] Yihe Tang, Weifeng Chen, Yijun Luo, and Yuting Zhang. Humble Teachers Teach Better Students for Semi-Supervised Object Detection. In *CVPR*, 2021. 2, 7, 15, 17, 18

[48] Antti Tarvainen and Harri Valpola. Mean Teachers are Better Role Models: Weight-Averaged Consistency Targets Improve Semi-Supervised Deep Learning Results. In *NeurIPS*, 2017. 2, 5, 18

[49] Thang Vu, Hyunjun Jang, Trung X. Pham, and Chang D. Yoo. Cascade RPN: Delving into High-Quality Region Proposal Network with Adaptive Convolution. In *NeurIPS*, 2019. 2, 4, 8, 14

[50] Xin Wang, Thomas E. Huang, Trevor Darrell, Joseph E. Gonzalez, and Fisher Yu. Frustratingly Simple Few-Shot Object Detection. In *ICML*, 2020. 3, 4, 7, 8, 13, 14, 16, 18

[51] Xinjiang Wang, Xingyi Yang, Shilong Zhang, Yijiang Li, Litong Feng, Shijie Fang, Chengqi Lyu, Kai Chen, and Wayne Zhang. Consistent-Teacher: Towards Reducing Inconsistent Pseudo-Targets in Semi-Supervised Object Detection. In *CVPR*, 2023. 1, 3, 19

[52] Yu-Xiong Wang, Deva Ramanan, and Martial Hebert. Meta-Learning to Detect Rare Objects. In *ICCV*, 2019. 3

[53] Jiaxi Wu, Songtao Liu, Di Huang, and Yunhong Wang. Multi-Scale Positive Sample Refinement for Few-Shot Object Detection. In *ECCV*, 2020. 8, 14

[54] Wuti Xiong, Yawen Cui, and Li Liu. Semi-Supervised Few-Shot Object Detection with a Teacher-Student Network. In *BMVC*, 2021. 4

[55] Mengde Xu, Zheng Zhang, Han Hu, Jianfeng Wang, Lijuan Wang, Fangyun Wei, Xiang Bai, and Zicheng Liu. End-to-End Semi-Supervised Object Detection with Soft Teacher. In *ICCV*, 2021. 1, 3, 5, 6, 13, 16, 17

[56] Xiaopeng Yan, Ziliang Chen, Anni Xu, Xiaoxi Wang, Xiaodan Liang, and Liang Lin. Meta R-CNN : Towards General Solver for Instance-Level Few-Shot Learning. In *ICCV*, 2019. 3

[57] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random Erasing Data Augmentation. In *AAAI*, 2020. 16