



Enhance Then Search: An Augmentation-Search Strategy with Foundation Models for Cross-Domain Few-Shot Object Detection

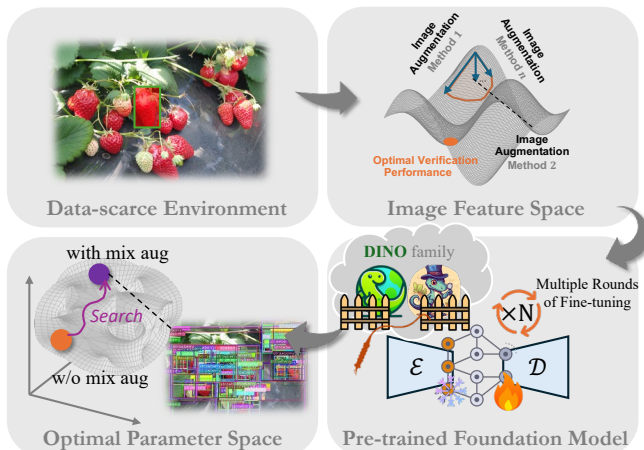
Jiancheng Pan, Yanxing Liu, Xiao He, Long Peng, Jiahao Li, Yuze Sun, Xiaomeng Huang*



What is Enhance Then Search?

Enhance Then Search facilitates efficient few-shot object detection and introduces an approach to solving the **Cross-Domain Few-Shot Object Detection** (CDFSD) problem by efficiently searching for the optimal parameter configuration from the foundation model. Our findings substantially advance the practical deployment of vision-language models in data-scarce environments, offering critical insights into optimizing their cross-domain generalization capabilities without labor-intensive retraining.

Mixed Image Augmentation

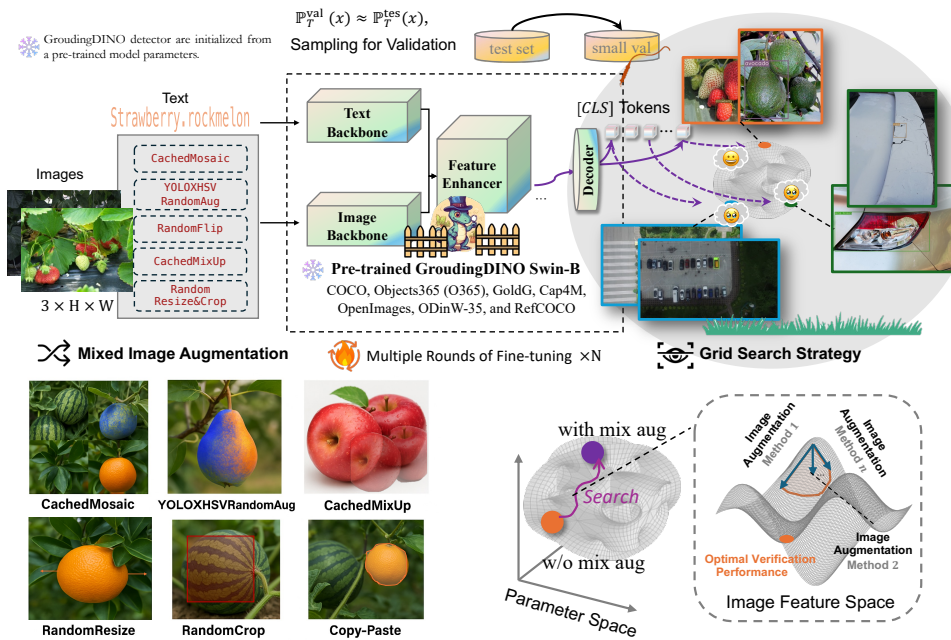


Grid Search Strategy

Building upon foundation model GroundingDINO, we employed several widely used image augmentation methods and established optimization objectives to effectively navigate the expansive domain space in search of optimal subdomains.

Codes and pre-trained models are at <https://github.com/jaychempan/ETS>

Augmentation-Search Strategy for CD-FSOD



Cross-Domain Few-Shot Object Detection

	Method	Setting	ArTaxOr	Clipart1K	DeepFish	DIOR	NEU-DET	UODD	Avg.
1-shot	Detic ViT-L [48]	Close-Source	3.2	15.1	9.0	4.1	3.8	4.2	6.6
	DE-ViT ViT-L [47]	Close-Source	10.5	13.0	19.3	14.7	0.6	2.4	10.1
	CD-ViTO ViT-L [5]	Close-Source	21.0	17.7	20.3	17.8	3.6	3.1	13.9
	GroundingDINO* Swin-B [21]	Open-Source	22.1	56.1	37.5	11.7	10.9	19.2	26.25
	ETS Swin-B (Ours)	Open-Source	28.1	57.5	40.7	12.7	11.7	21.2	28.65
5-shot	Detic ViT-L [48]	Close-Source	8.7	20.2	14.3	12.1	14.1	10.4	13.3
	DE-ViT ViT-L [47]	Close-Source	38.0	38.1	21.2	23.4	7.8	5.0	22.3
	CD-ViTO ViT-L [5]	Close-Source	47.9	41.1	22.3	26.9	11.4	6.8	26.1
	GroundingDINO* Swin-B [21]	Open-Source	61.9	59.1	43.2	27.7	22.2	27.9	40.3
	ETS Swin-B (Ours)	Open-Source	64.5	60.1	44.9	29.3	23.5	28.6	41.8
10-shot	Detic ViT-L [48]	Close-Source	12.0	22.3	17.9	15.4	16.8	14.4	16.5
	DE-ViT ViT-L [47]	Close-Source	49.2	40.8	21.3	25.6	8.8	5.4	25.2
	CD-ViTO ViT-L [5]	Close-Source	60.5	44.3	22.3	30.8	12.8	7.0	29.6
	GroundingDINO* Swin-B [21]	Open-Source	69.8	60.8	41.5	36.4	25.4	27.4	43.6
	ETS Swin-B (Ours)	Open-Source	71.2	61.5	44.1	37.5	26.1	29.8	45.0

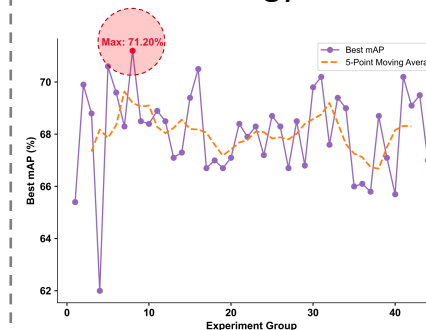
Sampling Rate for Validation

Experimental results under the ArTaxOr dataset (1-shot) with sampling rate.

Sampling Rate	ArTaxOr (1-shot mAP)
0.1	16.7
0.3	16.7
0.5	16.1
0.7	17.0
0.9	16.0
Avg.	16.5

Other Findings

- Search Strategy Feasibility



Search strategy experiments for 10-shot detection results on ArTaxOr dataset.

Visualisation -

Visualisation for 1-shot detection results on DeepFruits dataset.

