Improving Few-Shot Object Detection through a Performance Analysis on Aerial and Natural Images

Pierre Le Jeune
L2TI, Université Sorbonne Paris Nord
COSE

Anissa Mokraoui
L2TI, Université Sorbonne Paris Nord

30th European Signal Processing Conference
Belgrade – Sepbtember 2, 2022

Overview of the presentation

1. Few-Shot Object Detection

2. Difference Between Natural and Aerial Images

3. FSOD Performance Analysis

4. Bridging the Performance Gap

1.1 Few-Shot Object Detection - Definition

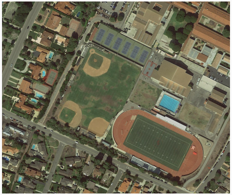
n-way k-shot object detection

Given support examples $\{(\boldsymbol{x_1}, \boldsymbol{a_1}), \ldots, (\boldsymbol{x_{nk}}, \boldsymbol{a_{nk}})\}$ it consists in detecting all occurrences of classes in $\mathcal{C}$ ($|\mathcal{C}| = \boldsymbol{n}$) in a query image $\boldsymbol{x_q}$.

1.1 Few-Shot Object Detection - Definition

n-way k-shot object detection

Given support examples $\{(\boldsymbol{x_1}, \boldsymbol{a_1}), \ldots, (\boldsymbol{x_{nk}}, \boldsymbol{a_{nk}})\}$ it consists in detecting all occurrences of classes in $\mathcal{C}$ ($|\mathcal{C}| = \boldsymbol{n}$) in a query image $\boldsymbol{x_q}$.

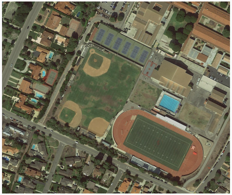


Query image

1.1 Few-Shot Object Detection - Definition

n-way k-shot object detection

Given support examples $\{(\boldsymbol{x}_1, \boldsymbol{a}_1), \ldots, (\boldsymbol{x}_{nk}, \boldsymbol{a}_{nk})\}$ it consists in detecting all occurrences of classes in $\mathcal{C}$ ($|\mathcal{C}| = \boldsymbol{n}$) in a query image $\boldsymbol{x}_q$.

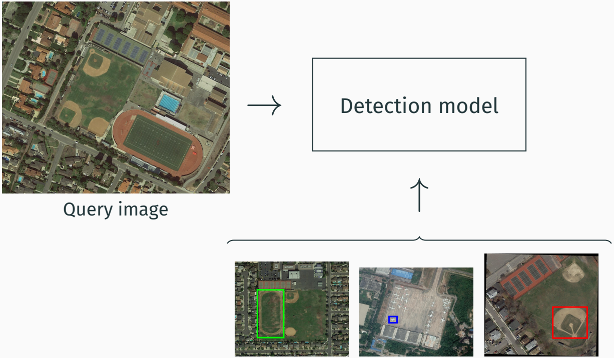


Query image

Support examples

n-way k-shot object detection

Given support examples $\{(\boldsymbol{x}_1, \boldsymbol{a}_1), \ldots, (\boldsymbol{x}_{nk}, \boldsymbol{a}_{nk})\}$ it consists in detecting all occurrences of classes in $\mathcal{C}$ ($|\mathcal{C}| = \boldsymbol{n}$) in a query image $\boldsymbol{x_q}$.

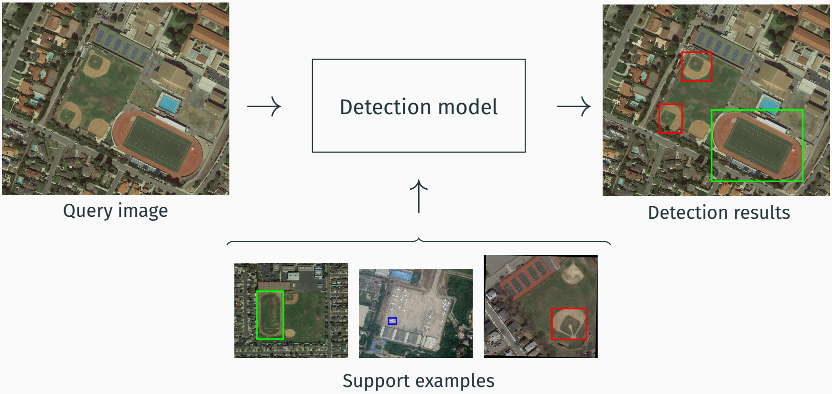


Support examples

1.1 Few-Shot Object Detection - Definition

n-way k-shot object detection

Given support examples $\{(\boldsymbol{x}_1, \boldsymbol{a}_1), \ldots, (\boldsymbol{x}_{nk}, \boldsymbol{a}_{nk})\}$ it consists in detecting all occurrences of classes in $\mathcal{C}$ ($|\mathcal{C}| = \boldsymbol{n}$) in a query image $\boldsymbol{x}_{\boldsymbol{q}}$.



1.2 Few-Shot Object Detection - General Principle

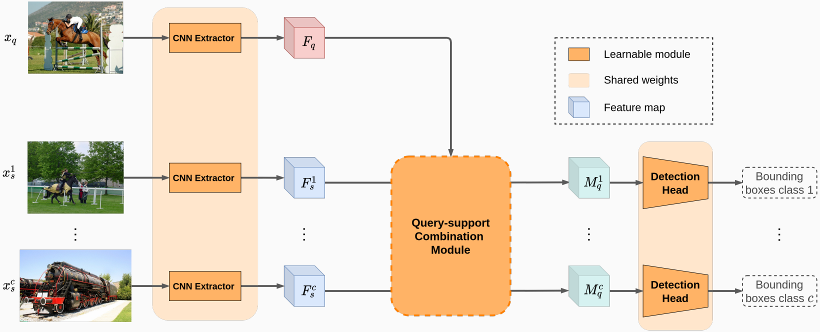


Figure 1: Attention-based Few-Shot Object Detection principle.

2. Difference between Natural and Aerial Images

Most methods are evaluated on natural images: Pascal VOC and MS COCO datasets.
⇒ **No guarantee good performance on aerial images.**

	# classes	# instances	Size (in pixels)			Object occupancy
			Mean	Std	Std/Mean	
DOTA	16	190k	33	37	1.12	0.13
DIOR	20	190k	42	58	1.38	0.17
Pascal VOC	20	50k	153	113	0.74	0.40

Table 1: Object size statistics (in pixel) for DOTA (Xia et al. 2018), DIOR (K. Li et al. 2020) and Pascal VOC (Everingham et al. 2010) datasets.

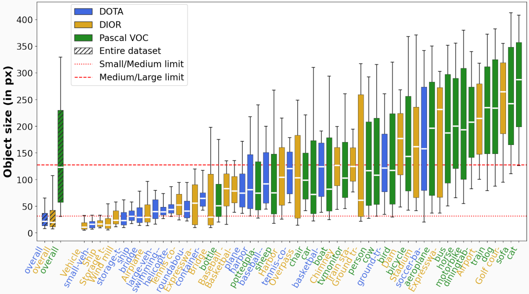


Figure 2: Boxplots of objects' size in DOTA, DIOR and Pascal VOC; per class **(right)** and overall **(left)**.

Relative FSOD performance against non few-shot baseline to make cross datasets comparison.

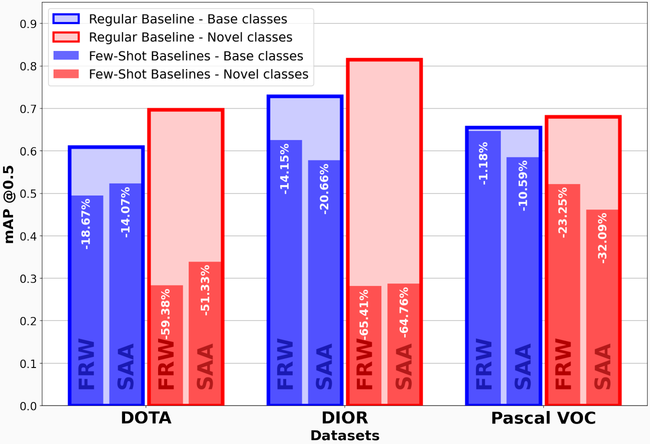


Figure 3: FSOD performance compared on DOTA, DIOR and Pascal VOC for FRW (Kang et al. 2019) and SAA (Xiao et al. 2021).

Clear correlation between average class size and few-shot performance.

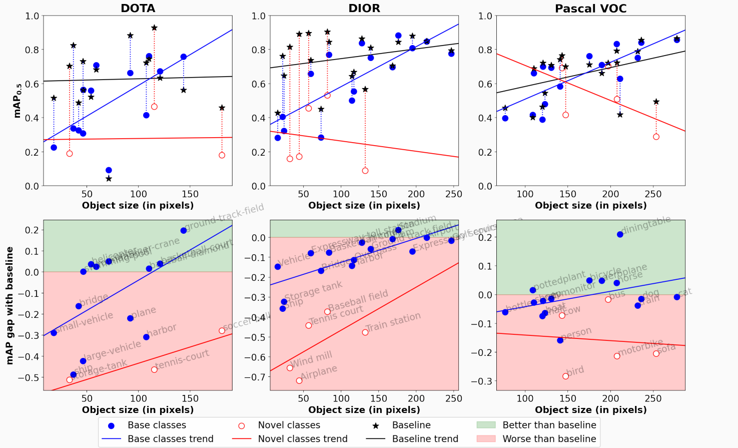


Figure 4: Per class performance analysis and comparison with non few-shot baseline on DOTA, DIOR and Pascal VOC separately.

Clear correlation between average class size and few-shot performance.

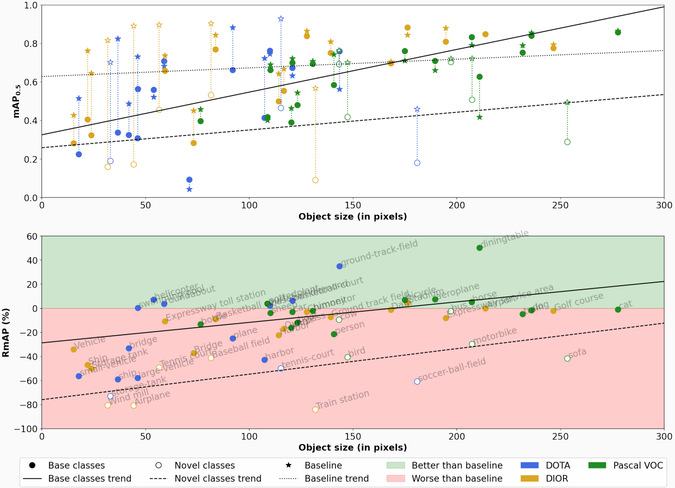


Figure 5: Per class performance analysis and comparison with non few-shot baseline on DOTA, DIOR and Pascal VOC.

4.1 Bridging the Performance Gap - Cropping Strategies

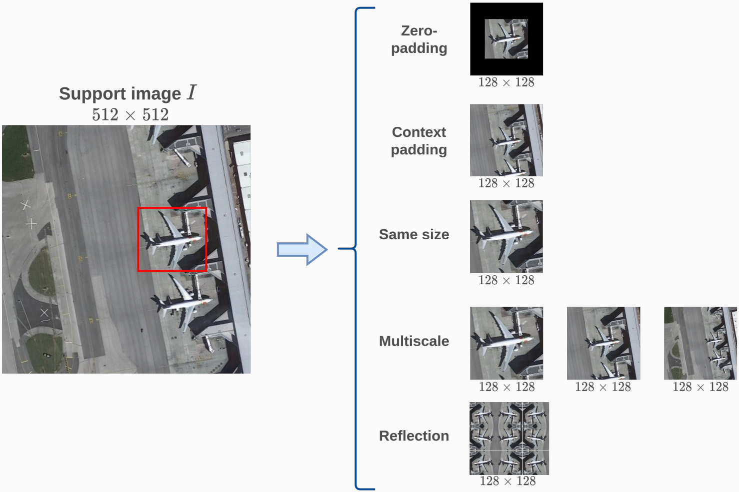


Figure 6: Proposed and evaluated support extraction strategies.

4.1 Bridging the Performance Gap - Cropping Strategies

Cropping strategy is important in FSOD: to condition the detection on a class, it must be easy to extract class features from the support images.

	Base classes				Novel classes			
	Mean	Small	Medium	Large	**Mean**	Small	Medium	Large
Zero-padding (default)	0.237	**0.099**	0.261	0.254	0.132	0.034	0.132	0.178
Context padding	0.243	0.074	0.281	0.240	0.136	0.034	0.115	0.245
Same size	0.238	0.085	0.271	0.241	**0.153**	0.030	**0.168**	**0.300**
Multi-scale	0.231	0.088	0.260	**0.272**	0.145	0.039	0.131	0.255
Reflection	**0.247**	0.086	**0.282**	0.253	0.128	**0.048**	0.139	0.246
Mixed	**0.247**	0.079	0.281	0.247	0.142	0.030	0.124	0.285

Table 2: Comparison of support extraction strategies on base and novel classes with DOTA dataset and FRW method with 10 shots. The performance is measured as in Lin et al. 2014, i.e. mAP is computed with multiple IoU thresholds ranging from 0.5 to 0.95 and separately on objects of different sizes small, medium, and large.

4.2 Bridging the Performance Gap - Augmentation Tecniques

Regular augmentation techniques are not designed for object detection:

- **Random cut-out** can completely mask objects out, making the detection impossible.
- **Random crop-resize** can crop parts of the images without any object.

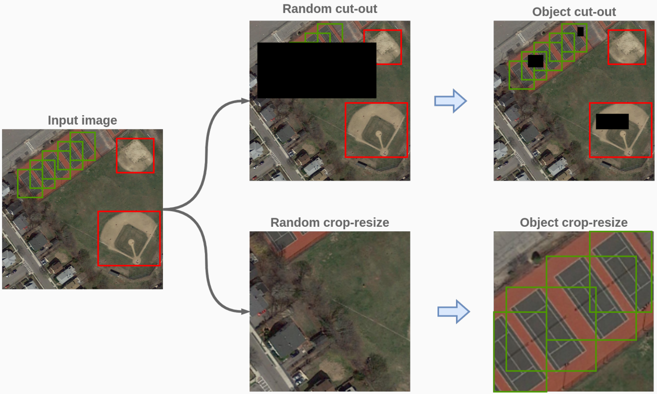


Figure 7: Object-level augmentation techniques.

Augmentation is key to provide better **robustness** in FSOD, more than with regular object detection because it improves the representation both for support and query images.

# Shots		Baseline	+ Flip	+ Color	+ Cutout	+ Crop
1	**Base**	**0.488**	0.458	0.460	0.472	0.457
	Novel	0.062	0.052	0.069	0.064	**0.100**
3	**Base**	**0.511**	0.475	0.470	0.461	0.452
	Novel	0.144	0.186	0.186	0.197	**0.220**
5	**Base**	**0.527**	0.494	0.501	0.503	0.487
	Novel	0.193	0.237	0.251	0.250	**0.259**
10	**Base**	**0.538**	0.508	0.508	0.504	0.503
	Novel	0.286	0.312	0.281	0.341	**0.359**

Table 3: Cumulative study of the proposed augmentation techniques on DOTA with the FRW method. mAP with a 0.5 IoU threshold is reported for different number of shots.

4.3 Bridging the Performance Gap - Overall results

The choice of the cropping strategy and the augmentation techniques yield results superior to the **state-of-the-art on DOTA and DIOR**.

Consistent improvements are also achieved on DOTA and Pascal VOC over the baseline. The results hold for multiple FSOD methods.

	DOTA				DIOR					Pascal VOC			
	FRW		SAA		FRW			SAA		FRW		SAA	
	Baseline	Ours	Baseline	Ours	Baseline	Ours	X. Li et al., 2021	Baseline	Ours	Baseline	Ours	Baseline	Ours
Base classes	**0.495**	0.485	**0.523**	0.467	**0.625**	0.615	0.540	0.578	**0.618**	**0.647**	0.610	**0.585**	0.531
Novel classes	0.283	**0.371**	0.339	**0.351**	0.282	**0.356**	0.320	0.287	**0.334**	0.522	**0.549**	0.462	**0.488**

Table 4: $mAP_{0.5}$ with 10 shots on three different datasets DOTA, DIOR and Pascal VOC. For each dataset, and each method the table compares the performance with our improvements (augmentations and *same-size* extraction) against the baseline.

5. Key Takeaways

1. **Object size varies greatly between aerial and natural images.**

5. Key Takeaways

1. **Object size varies greatly between aerial and natural images.**

2. **Small objects are particularly difficult for FSOD.**

5. Key Takeaways

1. **Object size varies greatly between aerial and natural images.**
2. **Small objects are particularly difficult for FSOD.**
3. **Augmentation and cropping strategy are crucial.**

5. Key Takeaways

1. **Object size varies greatly between aerial and natural images.**
2. **Small objects are particularly difficult for FSOD.**
3. **Augmentation and cropping strategy are crucial.**
4. **Better attention mechanisms are required.**

References i

- Everingham, Mark et al. (2010). “The pascal visual object classes (voc) challenge”. In: *International journal of computer vision* 88.2, pp. 303–338.
- Kang, Bingyi et al. (2019). “Few-shot object detection via feature reweighting”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8420–8429.
- Li, Ke et al. (2020). “Object detection in optical remote sensing images: A survey and a new benchmark”. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 159, pp. 296–307.
- Li, Xiang, Jingyu Deng, and Yi Fang (2021). “Few-Shot Object Detection on Remote Sensing Images”. In: *IEEE Transactions on Geoscience and Remote Sensing*, pp. 1–14. DOI: `10.1109/TGRS.2021.3051383`.
- Lin, Tsung-Yi et al. (2014). “Microsoft coco: Common objects in context”. In: *European conference on computer vision*. Springer, pp. 740–755.
- Xia, Gui-Song et al. (2018). “DOTA: A large-scale dataset for object detection in aerial images”. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3974–3983.
- Xiao, Zixuan et al. (2021). “Few-Shot Object Detection With Self-Adaptive Attention Network for Remote Sensing Images”. In: *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* 14, pp. 4854–4865.

Thank you for your attention

Any questions ?

pierre.lejeune@edu.univ-paris13.fr

https://pierlj.github.io