Object Detection in Aerial Images in Scarce Data Regime

Détection d'objets dans des images aériennes en cas de faible supervision

Pierre LE JEUNE

L2TI (UR 3043)
Université Sorbonne Paris Nord
COSE

Soutenance de thèse **Mardi 3 Octobre 2023**

Membres du Jury

Céline HUDELOT	Professeur — *CentraleSupélec*	**Rapportrice**
Paul HONEINE	Professeur — *Université Rouen Normandie*	**Rapporteur**
Emmanuel DELLANDREA	MCF HDR — *École Centrale de Lyon*	**Examinateur**
Ismail BEN AYED	Professeur — *ETS Montréal*	**Examinateur**
Fangchen FENG	MCF — *Université Sorbonne Paris Nord*	**Examinateur**
Hervé GUIOT	Président — *COSE*	**Invité**
Anissa MOKRAOUI	Professeur — *Université Sorbonne Paris Nord*	**Directrice**

Overview of the Presentation

1

Introduction and Context

Context of this PhD: CIFRE between COSE and L2TI.

- SMB from aeronautic and defense sector.
- Tier-1 supplier for French state.

COSE develops, among other products, aerial surveillance systems.

Laboratoire de Traitement et Transport de l'Information (L2TI – UR 3043).

Organized in two teams:

- Multimedia team
- Network team

Context of this PhD: CIFRE between COSE and L2TI.

- SMB from aeronautic and defense sector.
- Tier-1 supplier for French state.

COSE develops, among other products, aerial surveillance systems.

Laboratoire de Traitement et Transport de l'Information (L2TI – UR 3043).

Organized in two teams:

- Multimedia team
- Network team

Laboratoire Commun IRISER

Intelligence, ReconnaIssance et SurveillancE Réactive

ANR-21-LCV3-0004

This PhD falls within the scope of the CAMELEON project: COSE's next-gen airborne surveillance system.

GEOspatial INTelligence (GEOINT)

Georeferenced pieces of information about human activity on earth. Includes coordinates, date, and metadata.

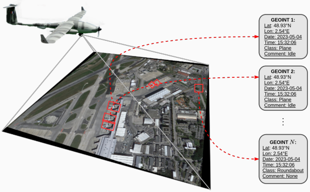


Figure 1: Creation of GEOINT in CAMELEON

Figure 2: Global Scanner System, to be replaced by CAMELEON.

PhD Objective: Automate the creation of GEOINT

- Localization of objects of interest in the images
- Recognition and classification of the objects.

1.1 Industrial Context and Objectives

This PhD falls within the scope of the CAMELEON project: COSE's next-gen airborne surveillance system.

GEOspatial INTelligence (GEOINT)

Georeferenced pieces of information about human activity on earth. Includes coordinates, date, and metadata.

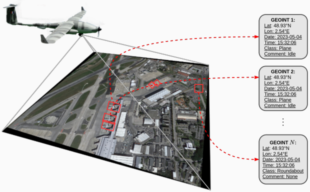


Figure 1: Creation of GEOINT in CAMELEON

Figure 2: Global Scanner System, to be replaced by CAMELEON.

PhD Objective: Automate the creation of GEOINT

- Localization of objects of interest in the images
- Recognition and classification of the objects.

Constraints:

- Images are extremely large.
- On-board resources are limited.
- Training sets are not available.

This PhD falls within the scope of the CAMELEON project: COSE's next-gen airborne surveillance system.

GEOspatial INTelligence (GEOINT)

Georeferenced pieces of information about human activity on earth. Includes coordinates, date, and metadata.

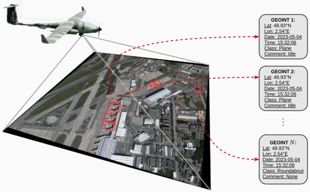


Figure 1: Creation of GEOINT in CAMELEON

Figure 2: Global Scanner System, to be replaced by CAMELEON.

PhD Objective: Automate the creation of GEOINT

- Localization of objects of interest in the images
- Recognition and classification of the objects.

Constraints:

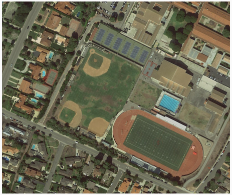
- Images are extremely large.
- On-board resources are limited.
- **Training sets are not available.**

Regular Object Detection

- Given a set of classes $\mathcal{C}$, find all occurrences of objects belonging to any class $\boldsymbol{c} \in \mathcal{C}$ in an image $\boldsymbol{I}$. Each object $\boldsymbol{i}$ is represented as a bounding box $\boldsymbol{b_i} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{c})$.
- Large annotated dataset available.

Regular Object Detection

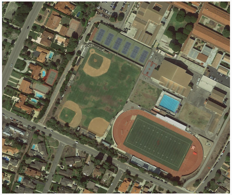
- Given a set of classes $\mathcal{C}$, find all occurrences of objects belonging to any class $\boldsymbol{c} \in \mathcal{C}$ in an image $\boldsymbol{I}$. Each object $\boldsymbol{i}$ is represented as a bounding box $\boldsymbol{b_i} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{c})$.
- Large annotated dataset available.



Input image $\boldsymbol{I}$

Regular Object Detection

- Given a set of classes $\mathcal{C}$, find all occurrences of objects belonging to any class $\boldsymbol{c} \in \mathcal{C}$ in an image $\boldsymbol{I}$. Each object $\boldsymbol{i}$ is represented as a bounding box $\boldsymbol{b}_i = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{c})$.
- Large annotated dataset available.



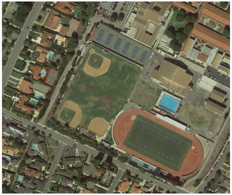
Input image $\boldsymbol{I}$

$$\mathcal{C} = \{\text{Baseball-diamond, Swimming-pool, Ground-track-field}\}$$

1.2 Regular Object Detection Definition

Regular Object Detection

- Given a set of classes $\mathcal{C}$, find all occurrences of objects belonging to any class $\boldsymbol{c} \in \mathcal{C}$ in an image $\boldsymbol{I}$. Each object $\boldsymbol{i}$ is represented as a bounding box $\boldsymbol{b_i} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{c})$.
- Large annotated dataset available.



Input image $\boldsymbol{I}$

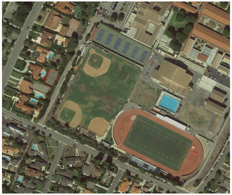
$\longrightarrow$ Detection model

$\mathcal{C} = \{$Baseball-diamond, Swimming-pool, Ground-track-field$\}$

1.2 Regular Object Detection Definition

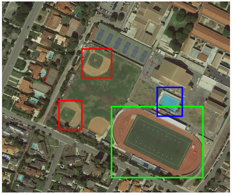
Regular Object Detection

- Given a set of classes $\mathcal{C}$, find all occurrences of objects belonging to any class $\boldsymbol{c} \in \mathcal{C}$ in an image $\boldsymbol{I}$. Each object $\boldsymbol{i}$ is represented as a bounding box $\boldsymbol{b_i} = (\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{w}, \boldsymbol{h}, \boldsymbol{c})$.
- Large annotated dataset available.



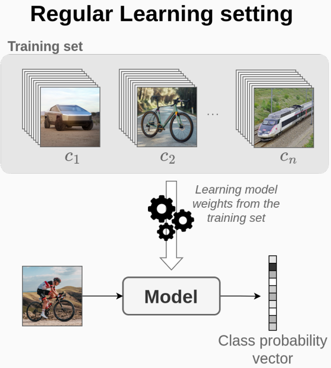
Input image ***I***

$\longrightarrow$ Detection model $\longrightarrow$

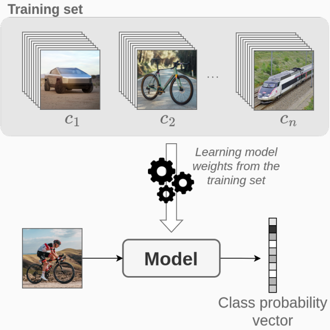


Detection results

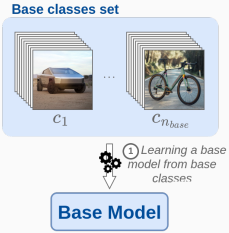
$\mathcal{C} = \{$Baseball-diamond, Swimming-pool, Ground-track-field$\}$


Regular Learning setting
Training set
c_1
c_2
...
c_n
Learning model weights from the training set
Model
Class probability vector

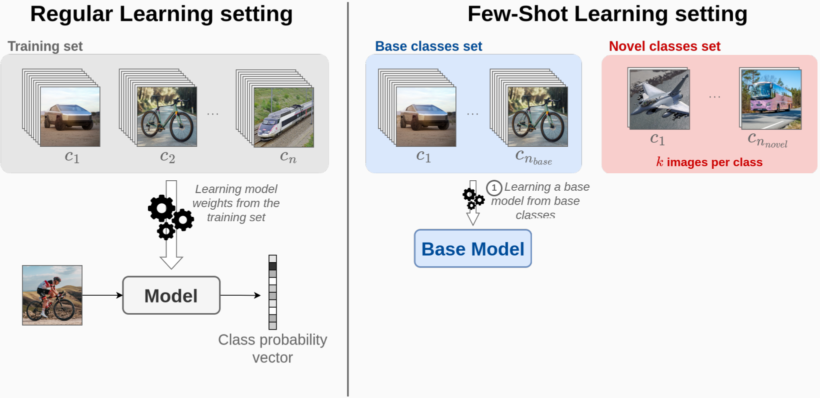
Regular Learning setting

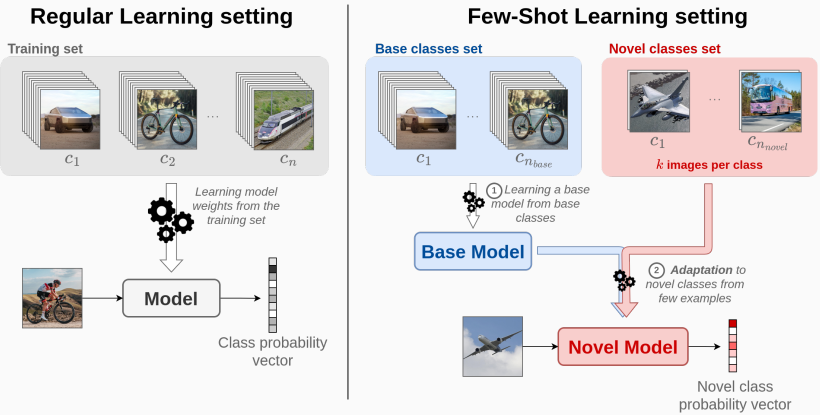


Few-Shot Learning setting



1.3 Few-Shot Learning Principle




Regular Learning setting
Training set
c_1
c_2
...
c_n
Learning model weights from the training set
Model
Class probability vector
Few-Shot Learning setting
Base classes set
c_1
...
$c_{n_{base}}$
Novel classes set
c_1
...
$c_{n_{novel}}$
k images per class
1 Learning a base model from base classes
Base Model
2 Adaptation to novel classes from few examples
Novel Model
Novel class probability vector

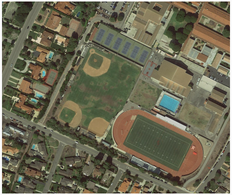
1.4 Few-Shot Object Detection Definition

N-way *K*-shot object detection

- Given support examples $\{(\boldsymbol{I_1}, \boldsymbol{b_1}), \dots, (\boldsymbol{I_{NK}}, \boldsymbol{b_{NK}})\}$ it consists in detecting all occurrences of classes in $\mathcal{C}$ ($|\mathcal{C}| = \boldsymbol{N}$) in a query image $\boldsymbol{I_q}$.
- Classes divided in two sets: **base classes** for which plenty of annotations are available, and **novel classes** for which only $\boldsymbol{K}$ annotations are available per class.

N-way *K*-shot object detection

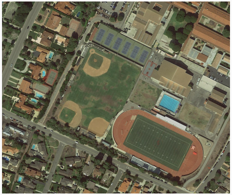
- Given support examples $\{(\boldsymbol{I_1}, \boldsymbol{b_1}), \ldots, (\boldsymbol{I_{NK}}, \boldsymbol{b_{NK}})\}$ it consists in detecting all occurrences of classes in $\mathcal{C}$ ($|\mathcal{C}| = \boldsymbol{N}$) in a query image $\boldsymbol{I_q}$.
- Classes divided in two sets: **base classes** for which plenty of annotations are available, and **novel classes** for which only $\boldsymbol{K}$ annotations are available per class.



Query image $\boldsymbol{I_q}$

N-way *K*-shot object detection

- Given support examples $\{(I_1, b_1), \dots, (I_{NK}, b_{NK})\}$ it consists in detecting all occurrences of classes in $\mathcal{C}$ ($|\mathcal{C}| = N$) in a query image I_q.
- Classes divided in two sets: **base classes** for which plenty of annotations are available, and **novel classes** for which only K annotations are available per class.



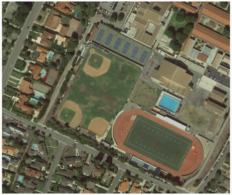
Query image I_q

Support examples (3-ways 1-shot)

1.4 Few-Shot Object Detection Definition

N-way *K*-shot object detection

- Given support examples $\{(I_1, b_1), \ldots, (I_{NK}, b_{NK})\}$ it consists in detecting all occurrences of classes in $\mathcal{C}$ ($|\mathcal{C}| = N$) in a query image I_q.
- Classes divided in two sets: **base classes** for which plenty of annotations are available, and **novel classes** for which only K annotations are available per class.



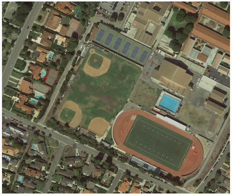
Query image I_q

$\longrightarrow$ Detection model $\uparrow$

Support examples (3-ways 1-shot)

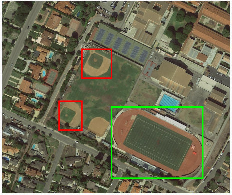
N-way *K*-shot object detection

- Given support examples $\{(\boldsymbol{I}_1, \boldsymbol{b}_1), \ldots, (\boldsymbol{I}_{NK}, \boldsymbol{b}_{NK})\}$ it consists in detecting all occurrences of classes in $\mathcal{C}$ ($|\mathcal{C}| = \boldsymbol{N}$) in a query image $\boldsymbol{I}_q$.
- Classes divided in two sets: **base classes** for which plenty of annotations are available, and **novel classes** for which only $\boldsymbol{K}$ annotations are available per class.



Query image $\boldsymbol{I}_q$

→ Detection model →



Detection results

Support examples (3-ways 1-shot)

1.5 Summary of the Contributions

1. **Investigation of FSOD challenge with aerial Images:**
   - Small objects are more numerous in aerial images, poor examples in the few-shot regime.
   - A first approach with metric learning: Prototypical Faster R-CNN.

1.5 Summary of the Contributions

1 Investigation of FSOD challenge with aerial Images:

- Small objects are more numerous in aerial images, poor examples in the few-shot regime.
- A first approach with metric learning: Prototypical Faster R-CNN.

2 Attention-based FSOD can successfully deal with small objects:

- Attention Alignment Fusion Framework to fairly compare Attention-based FSOD methods.
- Cross-Scale Query-Support Alignment for small object Detection.

1. **Investigation of FSOD challenge with aerial Images:**
   - Small objects are more numerous in aerial images, poor examples in the few-shot regime.
   - A first approach with metric learning: Prototypical Faster R-CNN.
2. **Attention-based FSOD can successfully deal with small objects:**
   - Attention Alignment Fusion Framework to fairly compare Attention-based FSOD methods.
   - Cross-Scale Query-Support Alignment for small object Detection.
3. **Few-Shot DiffusionDet a Fine-tuning approach:**
   - FSDiffusionDet outperforms all methods on aerial images.
   - Promising results in the challenging Cross-Domain scenarios.

1.5 Summary of the Contributions

1. **Investigation of FSOD challenge with aerial Images:**
   - Small objects are more numerous in aerial images, poor examples in the few-shot regime.
   - A first approach with metric learning: Prototypical Faster R-CNN.
2. **Attention-based FSOD can successfully deal with small objects:**
   - Attention Alignment Fusion Framework to fairly compare Attention-based FSOD methods.
   - Cross-Scale Query-Support Alignment for small object Detection.
3. **Few-Shot DiffusionDet a Fine-tuning approach:**
   - FSDiffusionDet outperforms all methods on aerial images.
   - Promising results in the challenging Cross-Domain scenarios.
4. **Carefully designed loss functions can improve small object detection:**
   - Scale-adaptive Intersection over Union (SIoU) improves small object detection.
   - SIoU loss allows precise control of the training balance between small and large objects.
   - SIoU aligns better with human perception, improves model evaluation.

1.5 Summary of the Contributions

1. **Investigation of FSOD challenge with aerial Images:**
   - Small objects are more numerous in aerial images, poor examples in the few-shot regime.
   - A first approach with metric learning: Prototypical Faster R-CNN.
2. **Attention-based FSOD can successfully deal with small objects:**
   - Attention Alignment Fusion Framework to fairly compare Attention-based FSOD methods.
   - Cross-Scale Query-Support Alignment for small object Detection.
3. **Few-Shot DiffusionDet a Fine-tuning approach:**
   - FSDiffusionDet outperforms all methods on aerial images.
   - Promising results in the challenging Cross-Domain scenarios.
4. **Carefully designed loss functions can improve small object detection:**
   - Scale-adaptive Intersection over Union (SIoU) improves small object detection.
   - SIoU loss allows precise control of the training balance between small and large objects.
   - SIoU aligns better with human perception, improves model evaluation.

2

Challenges of Few-Shot Object Detection in Aerial Images

2.1 Small Objects are Difficult to Detect

First contribution: Investigation of the performance gap between Aerial and Natural images in FSOD

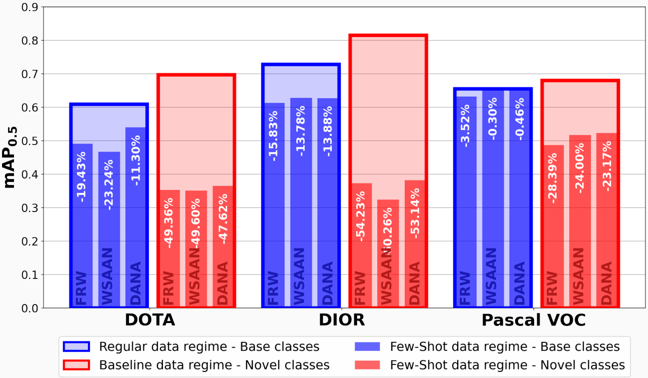
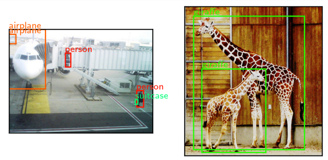


Figure 3: Few-Shot Detection performance compared on three distinct datasets.

2.1 Small Objects are Difficult to Detect

Aerial images vs. Natural images

(a) DOTA



(b) COCO

Figure 4: Examples images from two datasets.

Key specificities of aerial images:

- More and smaller objects
- Arbitrary rotations
- Densely packed objects

- Small objects are poor examples for the models and miscondition the detection.
- Few-Shot performance increases with object size, faster than in regular OD.

2.2 More Complex Scenarios in Real Applications

FSOD setting assumes a few assumptions $\rightarrow$ unrealistic for real application

FSOD setting assumes a few assumptions $\rightarrow$ unrealistic for real application

Generalized Few-Shot Object Detection (G-FSOD):

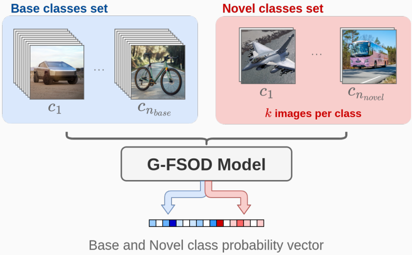


Figure 5: Generalized Few-Shot Object Detection (G-FSOD) principle.

FSOD setting assumes a few assumptions → unrealistic for real application

Generalized Few-Shot Object Detection (G-FSOD):

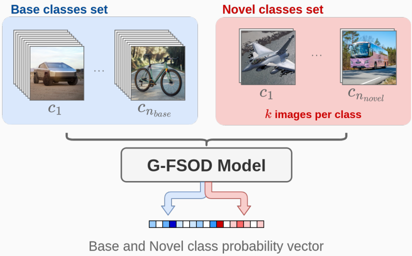


Figure 5: Generalized Few-Shot Object Detection (G-FSOD) principle.

Cross-Domain Few-Shot Object Detection (CD-FSOD):

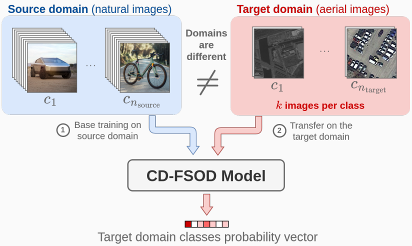


Figure 6: Cross-Domain Few-Shot Object Detection (CD-FSOD) principle.

FSOD setting assumes a few assumptions $\rightarrow$ unrealistic for real application

CD-FSOD is of practical use for COSE:

- While real case images will be aerial images, general aspects can change drastically (weather conditions, seasons, altitude, etc.).
- Classes are unknown before a mission, but their number is limited.
- Can rely on extremely large datasets for base training.

Increased difficulty: the model must adapt both to novel classes and new kinds of images.

Cross-Domain Few-Shot Object Detection (CD-FSOD):

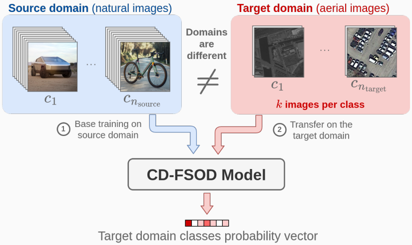


Figure 5: Cross-Domain Few-Shot Object Detection (CD-FSOD) principle.

Key takeaways

- Increased difficulty of detecting small objects in the Few-Shot Regime.
- FSOD performance increases with object size, but the trend is stronger than in regular settings.
- G-FSOD and CD-FSOD are more realistic but more challenging scenarios than Few-Shot.
- CD-FSOD is of particular interest to COSE as it matches its applications.

Key takeaways

- Increased difficulty of detecting small objects in the Few-Shot Regime.
- FSOD performance increases with object size, but the trend is stronger than in regular settings.
- G-FSOD and CD-FSOD are more realistic but more challenging scenarios than Few-Shot.
- CD-FSOD is of particular interest to COSE as it matches its applications.

Main research orientations of this Thesis:

2.2 More Complex Scenarios in Real Applications

Key takeaways

- Increased difficulty of detecting small objects in the Few-Shot Regime.
- FSOD performance increases with object size, but the trend is stronger than in regular settings.
- G-FSOD and CD-FSOD are more realistic but more challenging scenarios than Few-Shot.
- CD-FSOD is of particular interest to COSE as it matches its applications.

Main research orientations of this Thesis:

- Improving the few-shot detection of small objects (section 3).

Key takeaways

- Increased difficulty of detecting small objects in the Few-Shot Regime.
- FSOD performance increases with object size, but the trend is stronger than in regular settings.
- G-FSOD and CD-FSOD are more realistic but more challenging scenarios than Few-Shot.
- CD-FSOD is of particular interest to COSE as it matches its applications.

Main research orientations of this Thesis:

- Improving the few-shot detection of small objects (section 3).
- Designing methods for Cross-Domain scenarios (section 4).

3

Improving Small Object Detection in the Few-Shot Regime

Overview of the FSOD Literature

Approach	Abbreviation	Venue	Date	Detection Framework	Multiscale	Datasets	Aerial / Natural Images
Attention-based	FRW [Kang et al., 2019]	ICCV	2019	YOLO	No	Pascal / COCO	Natural
	OSOD-CACE [Hsieh et al., 2019]	NEURIPS	2019	Faster RCNN	Yes	Pascal / COCO	Natural
	Meta R-CNN [Yan et al., 2019]	ICCV	2019	Faster RCNN	No	Pascal / COCO	Natural
	FSOD-RSI [X. Li et al., 2021]	TGRS	2020	YOLO	Yes	DIOR / NWPU VHR	Aerial
	ARPN [Q. Fan et al., 2020]	CVPR	2020	Faster RCNN	Yes	COCO	Natural
	VEOW [Y. Xiao et al., 2020]	ECCV	2020	Faster RCNN	Yes	Pascal / COCO	Natural
	KT [Kim et al., 2020]	SMC	2020	Faster RCNN	Yes	Pascal	Natural
	OSOD-WFT [X. Li, L. Zhang, et al., 2020]	Preprint	2020	FCOS	Yes	Pascal / COCO / ImageNet Loc	Natural
	ONCE [J.-M. Perez-Rua et al., 2020]‡	CVPR	2020	Center Net	No	Pascal / COCO / Deepfashion	Natural
	WSAAN [Z. Xiao et al., 2021]	TAEORS	2021	Faster RCNN	Yes	RSOD / NWPU VHR	Aerial
	FSOD-FPDI [Yuxuan Gao et al., 2021]	MDPI	2021	FCOS	Yes	DOTA / NWPU VHR	Aerial
	Meta-FRCNN [G. Han, S. Huang, et al., 2022]	AAAI	2022	Faster RCNN	Yes	Pascal / COCO	Natural
	Meta-DETR [G. Zhang, Luo, et al., 2022]	TPAMI	2021	DETR	No	Pascal / COCO	Natural
	DRL [W. Liu, H. Li, et al., 2021]	Preprint	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	DANA [T.-I. Chen et al., 2021]	TM	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	SP [H. Xu et al., 2021]	Access	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	JCACR [Chu et al., 2021]	ICIP	2021	YOLO	Yes	Pascal / COCO	Natural
	TI-FSOD [A. Li et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	SAM [X. Huang et al., 2021]	MDPI	2021	Faster RCNN	No	NWPU VHR-10 / DIOR	Aerial
	FSOD-FCT [G. Han, Ma, et al., 2022]	CVPR	2022	Faster RCNN	No	Pascal / COCO	Natural
	SAR-DRM [Shiqi Chen et al., 2022]	TGRS	2022	Faster RCNN	No	FUSAR-GEN	Aerial
	FSOD-PSI [Ouyang et al., 2022]	JDT	2022	YOLO	Yes	Pascal / COCO	Natural
	SAFT [Y. Zhao et al., 2022]	CVPR	2022	FCOS	Yes	Pascal / COCO	Natural
	APSP [H. Lee et al., 2022]	WACV	2022	Faster RCNN	No	Pascal / COCO	Natural
	KFSOD [S. Zhang et al., 2022]	CVPR	2022	Faster RCNN	Yes	Pascal / COCO	Natural
	FSODS [Zhou et al., 2022]	TGRS	2022	YOLO	Yes	SMCDD-FS	Aerial
	TIN-FSOD [N. Liu et al., 2023]	Arxiv	2023	Faster RCNN	Yes	NWPU VHR/ DIOR / HRRSD	Aerial
	FSOD-ICF [Jiang et al., 2023]	WACV	2023	Faster RCNN	Yes	Pascal / COCO	Natural
Attention / Metric Learning	PNPDet [G. Zhang, Cui, et al., 2021]	WACV	2021	Center Net	No	Pascal / COCO	Natural
	UPE [A. Wu, Y. Han, et al., 2021]	ICCV	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	GenDet [Liyang Liu et al., 2021]	NNLS	2021	FCOS	Yes	Pascal / COCO	Natural
Metric learning	RepMet [Karlinsky et al., 2019]	CVPR	2018	Faster RCNN	Yes	Pascal / ImageNet Loc	Natural
	RN-FSOD [Yang et al., 2020]	NEURIPS	2020	Faster RCNN	Yes	Pascal / ImageNet Loc	Natural
	MDODD [X. Zhao et al., 2021]†	ICCV	2021	Faster RCNN	No	Pascal / COCO	Natural
	FSCE [B. Sun et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	GD-FSOD [A. Wu, S. Zhao, et al., 2021]	NEURIPS	2021	Faster RCNN	Yes	Pascal / COCO	Natural
Fine-tuning Strategy	LSTD [H. Chen et al., 2018]	AAAI	2018	Faster RCNN	Yes	Pascal / COCO / ImageNet Loc	Natural
	MSPSR [Jiaxi Wu et al., 2020]	ECCV	2020	Faster RCNN	Yes	Pascal / COCO	Natural
	TFA [X. Wang et al., 2020]	ICML	2020	Faster RCNN	Yes	Pascal / COCO / LVIS	Natural
	WOFG [Z. Fan et al., 2021]†	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	Hallu-FSOD [W. Zhang et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	DHP [Wolf et al., 2021]	ICCVW	2021	Faster RCNN	Yes	iSAID / NWPU VHR	Aerial
	LVC [Kaul et al., 2022]	CVPR	2021	Faster RCNN	No	Pascal / COCO	Natural
	FSCN [Y. Li et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	FADI [Cao et al., 2021]	NEURIPS	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	DeFRCN [Qiao et al., 2021]	ICCV	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	SIMPL [Y. Xu et al., 2021]	TAEORS	2022	YOLO	No	xView (plane only)	Aerial
	DETReg [Bar et al., 2022]	CVPR	2022	Deformable DETR	Yes	COCO	Natural
	CFA [Guirguis et al., 2022]†	CVPRW	2022	Faster RCNN	No	Pascal / COCO	Natural
	CIR [Y. Wang et al., 2022]	TGRS	2022	Faster RCNN	Yes	NWPU VHR-10 / DIOR	Aerial
	NIMPE [W. Liu, C. Wang, et al., 2022]	ICASSP	2022	Faster RCNN	Yes	COCO	Natural
	HDA [She et al., 2022]	IROS	2022	Faster RCNN	Yes	COCO	Natural
	MDB [S. Wu et al., 2022]	LNCS	2022	Faster RCNN	No	Pascal / COCO	Natural
	DCB [B.-B. Gao et al., 2022]†	NEURIPS	2022	Faster RCNN	Yes	Pascal / COCO	Natural
	CPP-FSOD [Lin et al., 2023]	Preprint	2023	Faster RCNN	Yes	Pascal / COCO	Natural
	I-DETR [Dong et al., 2022]‡	AAAI	2023	Deformable DETR	No	Pascal / COCO	Natural
Meta-Learning	MetaDet [Y.-X. Wang et al., 2019]	ICCV	2019	Faster RCNN	No	Pascal / COCO	Natural
	Sylph [Yin et al., 2022]‡	CVPR	2022	Faster RCNN	No	COCO / LVIS	Natural
Cross-Domain	OA-FSUI2IT [L. Zhao et al., 2022]	AAAI	2022	Faster RCNN	Yes	Multiple datasets	Natural
	Acro FOD [Yipeng Gao et al., 2022]	ECCV	2022	YOLO	Yes	Multiple datasets	Natural
	CD-CutMix [Nakamura et al., 2022]	ACCV	2022	Faster RCNN	No	Multiple datasets	Natural
	CD-FSOD [Xiong et al., 2022]	Preprint	2022	Faster RCNN	Yes	Multiple datasets	Aerial
	CD-MDB [K. Lee et al., 2022]	ECCV	2022	Faster RCNN	Yes	Multiple datasets	Aerial

Overview of the FSOD Literature – Fine-tuning vs. Attention-based approaches

Approach	Abbreviation	Venue	Date	Detection Framework	Multiscale	Datasets	Aerial / Natural Images
Attention-based	FRW [Kang et al., 2019]	ICCV	2019	YOLO	No	Pascal / COCO	Natural
	OSOD-CACE [Hsieh et al., 2019]	NEURIPS	2019	Faster RCNN	Yes	Pascal / COCO	Natural
	Meta R-CNN [Yan et al., 2019]	ICCV	2019	Faster RCNN	No	Pascal / COCO	Natural
	FSOD-RSI [X. Li et al., 2021]	TGRS	2020	YOLO	Yes	DIOR / NWPU VHR	Aerial
	ARPN [Q. Fan et al., 2020]	CVPR	2020	Faster RCNN	Yes	COCO	Natural
	VEOW [Y. Xiao et al., 2020]	ECCV	2020	Faster RCNN	Yes	Pascal / COCO	Natural
	KT [Kim et al., 2020]	SMC	2020	Faster RCNN	Yes	Pascal	Natural
	OSOD-WFT [X. Li, L. Zhang, et al., 2020]	Preprint	2020	FCOS	Yes	Pascal / COCO / ImageNet Loc	Natural
	ONCE [J.-M. Perez-Rua et al., 2020]‡	CVPR	2020	Center Net	No	Pascal / COCO / Deepfashion	Natural
	WSAAN [Z. Xiao et al., 2021]	TAEORS	2021	Faster RCNN	Yes	RSOD / NWPU VHR	Aerial
	FSOD-FPDI [Yuxuan Gao et al., 2021]	MDPI	2021	FCOS	Yes	DOTA / NWPU VHR	Aerial
	Meta-FRCNN [G. Han, S. Huang, et al., 2022]	AAAI	2022	Faster RCNN	Yes	Pascal / COCO	Natural
	Meta-DETR [G. Zhang, Luo, et al., 2022]	TPAMI	2021	DETR	No	Pascal / COCO	Natural
	DRL [W. Liu, H. Li, et al., 2021]	Preprint	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	DANA [T.-I. Chen et al., 2021]	TM	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	SP [H. Xu et al., 2021]	Access	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	JCACR [Chu et al., 2021]	ICIP	2021	YOLO	Yes	Pascal / COCO	Natural
	TI-FSOD [A. Li et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	SAM [X. Huang et al., 2021]	MDPI	2021	Faster RCNN	No	NWPU VHR-10 / DIOR	Aerial
	FSOD-FCT [G. Han, Ma, et al., 2022]	CVPR	2022	Faster RCNN	No	Pascal / COCO	Natural
	SAR-DRM [Shiqi Chen et al., 2022]	TGRS	2022	Faster RCNN	No	FUSAR-GEN	Aerial
	FSOD-PSI [Ouyang et al., 2022]	JDT	2022	YOLO	Yes	Pascal / COCO	Natural
	SAFT [Y. Zhao et al., 2022]	CVPR	2022	FCOS	Yes	Pascal / COCO	Natural
	APSP [H. Lee et al., 2022]	WACV	2022	Faster RCNN	No	Pascal / COCO	Natural
	KFSOD [S. Zhang et al., 2022]	CVPR	2022	Faster RCNN	Yes	Pascal / COCO	Natural
	FSODS [Zhou et al., 2022]	TGRS	2022	YOLO	Yes	SMCDD-FS	Aerial
	TIN-FSOD [N. Liu et al., 2023]	Arxiv	2023	Faster RCNN	Yes	NWPU VHR/ DIOR / HRRSD	Aerial
	FSOD-ICF [Jiang et al., 2023]	WACV	2023	Faster RCNN	Yes	Pascal / COCO	Natural
Attention / Metric Learning	PNPDet [G. Zhang, Cui, et al., 2021]	WACV	2021	Center Net	No	Pascal / COCO	Natural
	UPE [A. Wu, Y. Han, et al., 2021]	ICCV	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	GenDet [Liyang Liu et al., 2021]	NNLS	2021	FCOS	Yes	Pascal / COCO	Natural
Metric learning	RepMet [Karlinsky et al., 2019]	CVPR	2018	Faster RCNN	Yes	Pascal / ImageNet Loc	Natural
	RN-FSOD [Yang et al., 2020]	NEURIPS	2020	Faster RCNN	Yes	Pascal / ImageNet Loc	Natural
	MDODD [X. Zhao et al., 2021]†	ICCV	2021	Faster RCNN	No	Pascal / COCO	Natural
	FSCE [B. Sun et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	GD-FSOD [A. Wu, S. Zhao, et al., 2021]	NEURIPS	2021	Faster RCNN	Yes	Pascal / COCO	Natural
Fine-tuning Strategy	LSTD [H. Chen et al., 2018]	AAAI	2018	Faster RCNN	Yes	Pascal / COCO / ImageNet Loc	Natural
	MSPSR [Jiaxi Wu et al., 2020]	ECCV	2020	Faster RCNN	Yes	Pascal / COCO	Natural
	TFA [X. Wang et al., 2020]	ICML	2020	Faster RCNN	Yes	Pascal / COCO / LVIS	Natural
	WOFG [Z. Fan et al., 2021]†	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	Hallu-FSOD [W. Zhang et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	DHP [Wolf et al., 2021]	ICCVW	2021	Faster RCNN	Yes	iSAID / NWPU VHR	Aerial
	LVC [Kaul et al., 2022]	CVPR	2021	Faster RCNN	No	Pascal / COCO	Natural
	FSCN [Y. Li et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	FADI [Cao et al., 2021]	NEURIPS	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	DeFRCN [Qiao et al., 2021]	ICCV	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	SIMPL [Y. Xu et al., 2021]	TAEORS	2022	YOLO	No	xView (plane only)	Aerial
	DETReg [Bar et al., 2022]	CVPR	2022	Deformable DETR	Yes	COCO	Natural
	CFA [Guirguis et al., 2022]†	CVPRW	2022	Faster RCNN	No	Pascal / COCO	Natural
	CIR [Y. Wang et al., 2022]	TGRS	2022	Faster RCNN	Yes	NWPU VHR-10 / DIOR	Aerial
	NIMPE [W. Liu, C. Wang, et al., 2022]	ICASSP	2022	Faster RCNN	Yes	COCO	Natural
	HDA [She et al., 2022]	IROS	2022	Faster RCNN	Yes	COCO	Natural
	MDB [S. Wu et al., 2022]	LNCS	2022	Faster RCNN	No	Pascal / COCO	Natural
	DCB [B.-B. Gao et al., 2022]†	NEURIPS	2022	Faster RCNN	Yes	Pascal / COCO	Natural
	CPP-FSOD [Lin et al., 2023]	Preprint	2023	Faster RCNN	Yes	Pascal / COCO	Natural
	I-DETR [Dong et al., 2022]‡	AAAI	2023	Deformable DETR	No	Pascal / COCO	Natural
Meta-Learning	MetaDet [Y.-X. Wang et al., 2019]	ICCV	2019	Faster RCNN	No	Pascal / COCO	Natural
	Sylph [Yin et al., 2022]‡	CVPR	2022	Faster RCNN	No	COCO / LVIS	Natural
Cross-Domain	OA-FSUI2IT [L. Zhao et al., 2022]	AAAI	2022	Faster RCNN	Yes	Multiple datasets	Natural
	Acro FOD [Yipeng Gao et al., 2022]	ECCV	2022	YOLO	Yes	Multiple datasets	Natural
	CD-CutMix [Nakamura et al., 2022]	ACCV	2022	Faster RCNN	No	Multiple datasets	Natural
	CD-FSOD [Xiong et al., 2022]	Preprint	2022	Faster RCNN	Yes	Multiple datasets	Aerial
	CD-MDB [K. Lee et al., 2022]	ECCV	2022	Faster RCNN	Yes	Multiple datasets	Aerial

Overview of the FSOD Literature – Application on Aerial Images

Approach	Abbreviation	Venue	Date	Detection Framework	Multiscale	Datasets	Aerial / Natural Images
Attention-based	FRW [Kang et al., 2019]	ICCV	2019	YOLO	No	Pascal / COCO	Natural
	OSOD-CACE [Hsieh et al., 2019]	NEURIPS	2019	Faster RCNN	Yes	Pascal / COCO	Natural
	Meta R-CNN [Yan et al., 2019]	ICCV	2019	Faster RCNN	No	Pascal / COCO	Natural
	FSOD-RSI [X. Li et al., 2021]	TGRS	2020	YOLO	Yes	DIOR / NWPU VHR	Aerial
	ARPN [Q. Fan et al., 2020]	CVPR	2020	Faster RCNN	Yes	COCO	Natural
	VEOW [Y. Xiao et al., 2020]	ECCV	2020	Faster RCNN	Yes	Pascal / COCO	Natural
	KT [Kim et al., 2020]	SMC	2020	Faster RCNN	Yes	Pascal	Natural
	OSOD-WFT [X. Li, L. Zhang, et al., 2020]	Preprint	2020	FCOS	Yes	Pascal / COCO / ImageNet Loc	Natural
	ONCE [J.-M. Perez-Rua et al., 2020]‡	CVPR	2020	Center Net	No	Pascal / COCO / Deepfashion	Natural
	WSAAN [Z. Xiao et al., 2021]	TAEORS	2021	Faster RCNN	Yes	RSOD / NWPU VHR	Aerial
	FSOD-FPDI [Yuxuan Gao et al., 2021]	MDPI	2021	FCOS	Yes	DOTA / NWPU VHR	Aerial
	Meta-FRCNN [G. Han, S. Huang, et al., 2022]	AAAI	2022	Faster RCNN	Yes	Pascal / COCO	Natural
	Meta-DETR [G. Zhang, Luo, et al., 2022]	TPAMI	2021	DETR	No	Pascal / COCO	Natural
	DRL [W. Liu, H. Li, et al., 2021]	Preprint	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	DANA [T.-I. Chen et al., 2021]	TM	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	SP [H. Xu et al., 2021]	Access	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	JCACR [Chu et al., 2021]	ICIP	2021	YOLO	Yes	Pascal / COCO	Natural
	TI-FSOD [A. Li et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	SAM [X. Huang et al., 2021]	MDPI	2021	Faster RCNN	No	NWPU VHR-10 / DIOR	Aerial
	FSOD-FCT [G. Han, Ma, et al., 2022]	CVPR	2022	Faster RCNN	No	Pascal / COCO	Natural
	SAR-DRM [Shiqi Chen et al., 2022]	TGRS	2022	Faster RCNN	No	FUSAR-GEN	Aerial
	FSOD-PSI [Ouyang et al., 2022]	JDT	2022	YOLO	Yes	Pascal / COCO	Natural
	SAFT [Y. Zhao et al., 2022]	CVPR	2022	FCOS	Yes	Pascal / COCO	Natural
	APSP [H. Lee et al., 2022]	WACV	2022	Faster RCNN	No	Pascal / COCO	Natural
	KFSOD [S. Zhang et al., 2022]	CVPR	2022	Faster RCNN	Yes	Pascal / COCO	Natural
	FSODS [Zhou et al., 2022]	TGRS	2022	YOLO	Yes	SMCDD-FS	Aerial
	TIN-FSOD [N. Liu et al., 2023]	Arxiv	2023	Faster RCNN	Yes	NWPU VHR/ DIOR / HRRSD	Aerial
	FSOD-ICF [Jiang et al., 2023]	WACV	2023	Faster RCNN	Yes	Pascal / COCO	Natural
Attention / Metric Learning	PNPDet [G. Zhang, Cui, et al., 2021]	WACV	2021	Center Net	No	Pascal / COCO	Natural
	UPE [A. Wu, Y. Han, et al., 2021]	ICCV	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	GenDet [Liyang Liu et al., 2021]	NNLS	2021	FCOS	Yes	Pascal / COCO	Natural
Metric learning	RepMet [Karlinsky et al., 2019]	CVPR	2018	Faster RCNN	Yes	Pascal / ImageNet Loc	Natural
	RN-FSOD [Yang et al., 2020]	NEURIPS	2020	Faster RCNN	Yes	Pascal / ImageNet Loc	Natural
	MDODD [X. Zhao et al., 2021]†	ICCV	2021	Faster RCNN	No	Pascal / COCO	Natural
	FSCE [B. Sun et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	GD-FSOD [A. Wu, S. Zhao, et al., 2021]	NEURIPS	2021	Faster RCNN	Yes	Pascal / COCO	Natural
Fine-tuning Strategy	LSTD [H. Chen et al., 2018]	AAAI	2018	Faster RCNN	Yes	Pascal / COCO / ImageNet Loc	Natural
	MSPSR [Jiaxi Wu et al., 2020]	ECCV	2020	Faster RCNN	Yes	Pascal / COCO	Natural
	TFA [X. Wang et al., 2020]	ICML	2020	Faster RCNN	Yes	Pascal / COCO / LVIS	Natural
	WOFG [Z. Fan et al., 2021]†	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	Hallu-FSOD [W. Zhang et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	DHP [Wolf et al., 2021]	ICCVW	2021	Faster RCNN	Yes	iSAID / NWPU VHR	Aerial
	LVC [Kaul et al., 2022]	CVPR	2021	Faster RCNN	No	Pascal / COCO	Natural
	FSCN [Y. Li et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	FADI [Cao et al., 2021]	NEURIPS	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	DeFRCN [Qiao et al., 2021]	ICCV	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	SIMPL [Y. Xu et al., 2021]	TAEORS	2022	YOLO	No	xView (plane only)	Aerial
	DETReg [Bar et al., 2022]	CVPR	2022	Deformable DETR	Yes	COCO	Natural
	CFA [Guirguis et al., 2022]†	CVPRW	2022	Faster RCNN	No	Pascal / COCO	Natural
	CIR [Y. Wang et al., 2022]	TGRS	2022	Faster RCNN	Yes	NWPU VHR-10 / DIOR	Aerial
	NIMPE [W. Liu, C. Wang, et al., 2022]	ICASSP	2022	Faster RCNN	Yes	COCO	Natural
	HDA [She et al., 2022]	IROS	2022	Faster RCNN	Yes	COCO	Natural
	MDB [S. Wu et al., 2022]	LNCS	2022	Faster RCNN	No	Pascal / COCO	Natural
	DCB [B.-B. Gao et al., 2022]†	NEURIPS	2022	Faster RCNN	Yes	Pascal / COCO	Natural
	CPP-FSOD [Lin et al., 2023]	Preprint	2023	Faster RCNN	Yes	Pascal / COCO	Natural
	I-DETR [Dong et al., 2022]‡	AAAI	2023	Deformable DETR	No	Pascal / COCO	Natural
Meta-Learning	MetaDet [Y.-X. Wang et al., 2019]	ICCV	2019	Faster RCNN	No	Pascal / COCO	Natural
	Sylph [Yin et al., 2022]‡	CVPR	2022	Faster RCNN	No	COCO / LVIS	Natural
Cross-Domain	OA-FSUI2IT [L. Zhao et al., 2022]	AAAI	2022	Faster RCNN	Yes	Multiple datasets	Natural
	Acro FOD [Yipeng Gao et al., 2022]	ECCV	2022	YOLO	Yes	Multiple datasets	Natural
	CD-CutMix [Nakamura et al., 2022]	ACCV	2022	Faster RCNN	No	Multiple datasets	Natural
	CD-FSOD [Xiong et al., 2022]	Preprint	2022	Faster RCNN	Yes	Multiple datasets	Aerial
	CD-MDB [K. Lee et al., 2022]	ECCV	2022	Faster RCNN	Yes	Multiple datasets	Aerial

Overview of the FSOD Literature – Cross-Domain FSOD

Approach	Abbreviation	Venue	Date	Detection Framework	Multiscale	Datasets	Aerial / Natural Images
Attention-based	FRW [Kang et al., 2019]	ICCV	2019	YOLO	No	Pascal / COCO	Natural
	OSOD-CACE [Hsieh et al., 2019]	NEURIPS	2019	Faster RCNN	Yes	Pascal / COCO	Natural
	Meta R-CNN [Yan et al., 2019]	ICCV	2019	Faster RCNN	No	Pascal / COCO	Natural
	FSOD-RSI [X. Li et al., 2021]	TGRS	2020	YOLO	Yes	DIOR / NWPU VHR	Aerial
	ARPN [Q. Fan et al., 2020]	CVPR	2020	Faster RCNN	Yes	COCO	Natural
	VEOW [Y. Xiao et al., 2020]	ECCV	2020	Faster RCNN	Yes	Pascal / COCO	Natural
	KT [Kim et al., 2020]	SMC	2020	Faster RCNN	Yes	Pascal	Natural
	OSOD-WFT [X. Li, L. Zhang, et al., 2020]	Preprint	2020	FCOS	Yes	Pascal / COCO / ImageNet Loc	Natural
	ONCE [J.-M. Perez-Rua et al., 2020]‡	CVPR	2020	Center Net	No	Pascal / COCO / Deepfashion	Natural
	WSAAN [Z. Xiao et al., 2021]	TAEORS	2021	Faster RCNN	Yes	RSOD / NWPU VHR	Aerial
	FSOD-FPDI [Yuxuan Gao et al., 2021]	MDPI	2021	FCOS	Yes	DOTA / NWPU VHR	Aerial
	Meta-FRCNN [G. Han, S. Huang, et al., 2022]	AAAI	2022	Faster RCNN	Yes	Pascal / COCO	Natural
	Meta-DETR [G. Zhang, Luo, et al., 2022]	TPAMI	2021	DETR	No	Pascal / COCO	Natural
	DRL [W. Liu, H. Li, et al., 2021]	Preprint	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	DANA [T.-I. Chen et al., 2021]	TM	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	SP [H. Xu et al., 2021]	Access	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	JCACR [Chu et al., 2021]	ICIP	2021	YOLO	Yes	Pascal / COCO	Natural
	TI-FSOD [A. Li et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	SAM [X. Huang et al., 2021]	MDPI	2021	Faster RCNN	No	NWPU VHR-10 / DIOR	Aerial
	FSOD-FCT [G. Han, Ma, et al., 2022]	CVPR	2022	Faster RCNN	No	Pascal / COCO	Natural
	SAR-DRM [Shiqi Chen et al., 2022]	TGRS	2022	Faster RCNN	No	FUSAR-GEN	Aerial
	FSOD-PSI [Ouyang et al., 2022]	JDT	2022	YOLO	Yes	Pascal / COCO	Natural
	SAFT [Y. Zhao et al., 2022]	CVPR	2022	FCOS	Yes	Pascal / COCO	Natural
	APSP [H. Lee et al., 2022]	WACV	2022	Faster RCNN	No	Pascal / COCO	Natural
	KFSOD [S. Zhang et al., 2022]	CVPR	2022	Faster RCNN	Yes	Pascal / COCO	Natural
	FSODS [Zhou et al., 2022]	TGRS	2022	YOLO	Yes	SMCDD-FS	Aerial
	TIN-FSOD [N. Liu et al., 2023]	Arxiv	2023	Faster RCNN	Yes	NWPU VHR/ DIOR / HRRSD	Aerial
	FSOD-ICF [Jiang et al., 2023]	WACV	2023	Faster RCNN	Yes	Pascal / COCO	Natural
Attention / Metric Learning	PNPDet [G. Zhang, Cui, et al., 2021]	WACV	2021	Center Net	No	Pascal / COCO	Natural
	UPE [A. Wu, Y. Han, et al., 2021]	ICCV	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	GenDet [Liyang Liu et al., 2021]	NNLS	2021	FCOS	Yes	Pascal / COCO	Natural
Metric learning	RepMet [Karlinsky et al., 2019]	CVPR	2018	Faster RCNN	Yes	Pascal / ImageNet Loc	Natural
	RN-FSOD [Yang et al., 2020]	NEURIPS	2020	Faster RCNN	Yes	Pascal / ImageNet Loc	Natural
	MDODD [X. Zhao et al., 2021]†	ICCV	2021	Faster RCNN	No	Pascal / COCO	Natural
	FSCE [B. Sun et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	GD-FSOD [A. Wu, S. Zhao, et al., 2021]	NEURIPS	2021	Faster RCNN	Yes	Pascal / COCO	Natural
Fine-tuning Strategy	LSTD [H. Chen et al., 2018]	AAAI	2018	Faster RCNN	Yes	Pascal / COCO / ImageNet Loc	Natural
	MSPSR [Jiaxi Wu et al., 2020]	ECCV	2020	Faster RCNN	Yes	Pascal / COCO	Natural
	TFA [X. Wang et al., 2020]	ICML	2020	Faster RCNN	Yes	Pascal / COCO / LVIS	Natural
	WOFG [Z. Fan et al., 2021]†	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	Hallu-FSOD [W. Zhang et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	DHP [Wolf et al., 2021]	ICCVW	2021	Faster RCNN	Yes	iSAID / NWPU VHR	Aerial
	LVC [Kaul et al., 2022]	CVPR	2021	Faster RCNN	No	Pascal / COCO	Natural
	FSCN [Y. Li et al., 2021]	CVPR	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	FADI [Cao et al., 2021]	NEURIPS	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	DeFRCN [Qiao et al., 2021]	ICCV	2021	Faster RCNN	Yes	Pascal / COCO	Natural
	SIMPL [Y. Xu et al., 2021]	TAEORS	2022	YOLO	No	xView (plane only)	Aerial
	DETReg [Bar et al., 2022]	CVPR	2022	Deformable DETR	Yes	COCO	Natural
	CFA [Guirguis et al., 2022]†	CVPRW	2022	Faster RCNN	No	Pascal / COCO	Natural
	CIR [Y. Wang et al., 2022]	TGRS	2022	Faster RCNN	Yes	NWPU VHR-10 / DIOR	Aerial
	NIMPE [W. Liu, C. Wang, et al., 2022]	ICASSP	2022	Faster RCNN	Yes	COCO	Natural
	HDA [She et al., 2022]	IROS	2022	Faster RCNN	Yes	COCO	Natural
	MDB [S. Wu et al., 2022]	LNCS	2022	Faster RCNN	No	Pascal / COCO	Natural
	DCB [B.-B. Gao et al., 2022]†	NEURIPS	2022	Faster RCNN	Yes	Pascal / COCO	Natural
	CPP-FSOD [Lin et al., 2023]	Preprint	2023	Faster RCNN	Yes	Pascal / COCO	Natural
	I-DETR [Dong et al., 2022]‡	AAAI	2023	Deformable DETR	No	Pascal / COCO	Natural
Meta-Learning	MetaDet [Y.-X. Wang et al., 2019]	ICCV	2019	Faster RCNN	No	Pascal / COCO	Natural
	Sylph [Yin et al., 2022]‡	CVPR	2022	Faster RCNN	No	COCO / LVIS	Natural
Cross-Domain	OA-FSUI2IT [L. Zhao et al., 2022]	AAAI	2022	Faster RCNN	Yes	Multiple datasets	Natural
	Acro FOD [Yipeng Gao et al., 2022]	ECCV	2022	YOLO	Yes	Multiple datasets	Natural
	CD-CutMix [Nakamura et al., 2022]	ACCV	2022	Faster RCNN	No	Multiple datasets	Natural
	CD-FSOD [Xiong et al., 2022]	Preprint	2022	Faster RCNN	Yes	Multiple datasets	Aerial
	CD-MDB [K. Lee et al., 2022]	ECCV	2022	Faster RCNN	Yes	Multiple datasets	Aerial

Attention-base Few-Shot Object Detection principle:

Adapt features from the query image on the fly during inference from a few annotated support examples. Built from three main components:

- **Backbone:** extracts features from the images.
- **Query-Support Combination Module:** combines query and support features.
- **Detection Head:** performs object detection in a class-agnostic manner.

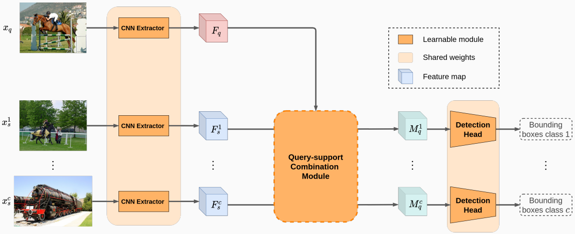


Figure 6: Attention-based FSOD principle.

Attention-base Few-Shot Object Detection principle:

Adapt features from the query image on the fly during inference from a few annotated support examples. Built from three main components:

- **Backbone:** extracts features from the images.
- **Query-Support Combination Module:** combines query and support features.
- **Detection Head:** performs object detection in a class-agnostic manner.

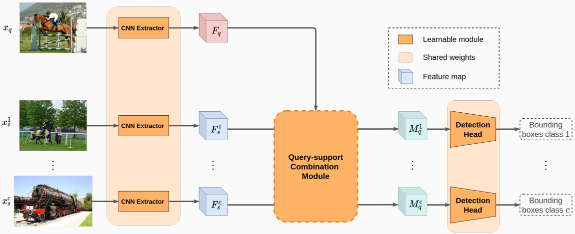


Figure 6: Attention-based FSOD principle.

Great variety of Query-Support Combination Modules

► Introduction of a modular framework, called **Alignment Attention Fusion (AAF) Framework**, to ease comparison and re-implementation [Le Jeune et al., 2022].

3.1 Cross-Scale Query-Support Alignment (XQSA)

Objective: propose a better Query-Support combination block to improve small object detection.

3.1 Cross-Scale Query-Support Alignment (XQSA)

Objective: propose a better Query-Support combination block to improve small object detection.

XQSA's motivation

- Combines query and support features from different scales together.
- Allows matching query and support objects from different sizes.
- *properties not available in the literature.*

3.1 Cross-Scale Query-Support Alignment (XQSA)

Objective: propose a better Query-Support combination block to improve small object detection.

XQSA's motivation

- Combines query and support features from different scales together.
- Allows matching query and support objects from different sizes.
- ▶ *properties not available in the literature.*

Following our AAF framework, the Query-Support combination block is split into three components:

- **Self Attention**: filters query and support features independently.
- **Spatial Alignment**: locally compares features from query and support.
- **Feature Fusion**: aggregates relevant information for detection.

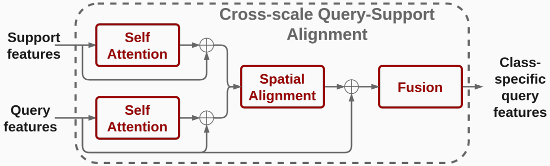


Figure 7: Overall structure of the Cross-Scale Query-Support Alignment block.

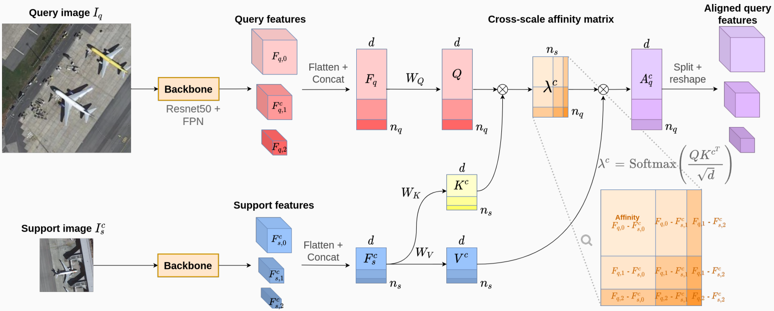


Figure 8: Illustration of the Spatial Alignment block in XQSA.

Concatenated multiscale features

Per level features

$$Q = F_q W_Q = [F_{q,0}, F_{q,1}, F_{q,2}]\, W_Q, \quad (1)$$

$$K^c = F^c_s W_K = [F^c_{s,0}, F^c_{s,1}, F^c_{s,2}]\, W_K, \quad (2)$$

$$V^c = F^c_s W_V = [F^c_{s,0}, F^c_{s,1}, F^c_{s,2}]\, W_V, \quad (3)$$

Learned projection matrices

$$\lambda^c = \mathrm{Softmax}\left(\frac{QK^{c^T}}{\sqrt{d}}\right), \quad (4)$$

$$A^c_q = \lambda^c V^c. \quad (5)$$

Comparison with two existing methods:

- Feature Reweighting (FRW) [Kang et al., 2019].
- Dual AwareNess Attention (DANA) [T.-I. Chen et al., 2021].
- Cross-Scale Query-Support Alignment (XQSA) [Le Jeune et al., 2023].

▶ Two aerial datasets DOTA and DIOR, and two natural datasets Pascal VOC and MS COCO.

		DOTA				DIOR				Pascal VOC				MS COCO			
		All	S	M	L	All	S	M	L	All	S	M	L	All	S	M	L
Base Classes	FRW	49.04	25.48	59.17	63.37	62.20	8.21	48.66	80.67	63.21	15.67	47.94	**81.73**	29.03	13.08	35.87	48.00
	DANA	**53.98**	**37.00**	**62.27**	**70.32**	**62.71**	**10.92**	**49.34**	**83.17**	**65.17**	**18.14**	**50.58**	80.11	**38.14**	**23.30**	**51.85**	**56.38**
	XQSA	51.11	26.10	59.41	64.30	59.88	10.64	45.69	82.34	62.13	15.60	48.64	75.94	31.56	16.13	40.13	49.83
Novel Classes	FRW	37.29	13.99	34.11	59.31	36.29	2.48	33.74	59.38	48.72	16.44	26.71	**68.27**	24.09	11.53	22.45	**38.69**
	DANA	36.38	14.33	40.00	**64.64**	38.18	3.21	34.91	**60.99**	52.26	10.05	24.67	67.23	24.75	12.01	**29.40**	37.95
	XQSA	**41.00**	**17.84**	**44.57**	54.46	**41.51**	**4.12**	**40.69**	58.21	**53.94**	**19.46**	**34.86**	66.14	**25.03**	**12.57**	26.05	38.55

Table 1: Performance comparison between XQSA, FRW, and DANA. $\text{mAP}_{0.5}$ values are reported separately for base (top) and novel (bottom) classes on DOTA, DIOR, Pascal VOC, and MS COCO with $K = 10$ shots. mAP values are reported for All, Small ($\sqrt{wh} < 32$), Medium ($32 \leq \sqrt{wh} < 96$) and Large ($\sqrt{wh} \geq 96$) objects.

Comparison with two existing methods:

- Feature Reweighting (FRW) [Kang et al., 2019].
- Dual AwareNess Attention (DANA) [T.-I. Chen et al., 2021].
- Cross-Scale Query-Support Alignment (XQSA) [Le Jeune et al., 2023].

► Two aerial datasets DOTA and DIOR, and two natural datasets Pascal VOC and MS COCO.

		DOTA				DIOR				Pascal VOC				MS COCO			
		All	S	M	L	All	S	M	L	All	S	M	L	All	S	M	L
Base Classes	FRW	49.04	25.48	59.17	63.37	62.20	8.21	48.66	80.67	63.21	15.67	47.94	**81.73**	29.03	13.08	35.87	48.00
	DANA	**53.98**	**37.00**	**62.27**	**70.32**	**62.71**	**10.92**	**49.34**	**83.17**	**65.17**	**18.14**	**50.58**	80.11	**38.14**	**23.30**	**51.85**	**56.38**
	XQSA	51.11	26.10	59.41	64.30	59.88	10.64	45.69	82.34	62.13	15.60	48.64	75.94	31.56	16.13	40.13	49.83
Novel Classes	FRW	37.29	13.99	34.11	59.31	36.29	2.48	33.74	59.38	48.72	16.44	26.71	**68.27**	24.09	11.53	22.45	**38.69**
	DANA	36.38	14.33	40.00	**64.64**	38.18	3.21	34.91	**60.99**	52.26	10.05	24.67	67.23	24.75	12.01	**29.40**	37.95
	XQSA	**41.00**	**17.84**	**44.57**	54.46	**41.51**	**4.12**	**40.69**	58.21	**53.94**	**19.46**	**34.86**	66.14	**25.03**	**12.57**	26.05	38.55

Table 1: Performance comparison between XQSA, FRW, and DANA. $\text{mAP}_{0.5}$ values are reported separately for base (top) and novel (bottom) classes on DOTA, DIOR, Pascal VOC, and MS COCO with $K = 10$ shots. mAP values are reported for All, Small ($\sqrt{wh} < 32$), Medium ($32 \leq \sqrt{wh} < 96$) and Large ($\sqrt{wh} \geq 96$) objects.

Performance analysis:

3.1 Cross-Scale Query-Support Alignment (XQSA)

Comparison with two existing methods:

- Feature Reweighting (FRW) [Kang et al., 2019].
- Dual AwareNess Attention (DANA) [T.-I. Chen et al., 2021].
- Cross-Scale Query-Support Alignment (XQSA) [Le Jeune et al., 2023].

► Two aerial datasets DOTA and DIOR, and two natural datasets Pascal VOC and MS COCO.

		DOTA				DIOR				Pascal VOC				MS COCO			
		All	S	M	L	All	S	M	L	All	S	M	L	All	S	M	L
Base Classes	FRW	49.04	25.48	59.17	63.37	62.20	8.21	48.66	80.67	63.21	15.67	47.94	**81.73**	29.03	13.08	35.87	48.00
	DANA	**53.98**	**37.00**	**62.27**	**70.32**	**62.71**	**10.92**	**49.34**	**83.17**	**65.17**	**18.14**	**50.58**	80.11	**38.14**	**23.30**	**51.85**	**56.38**
	XQSA	51.11	26.10	59.41	64.30	59.88	10.64	45.69	82.34	62.13	15.60	48.64	75.94	31.56	16.13	40.13	49.83
Novel Classes	FRW	37.29	13.99	34.11	59.31	36.29	2.48	33.74	59.38	48.72	16.44	26.71	**68.27**	24.09	11.53	22.45	**38.69**
	DANA	36.38	14.33	40.00	**64.64**	38.18	3.21	34.91	**60.99**	52.26	10.05	24.67	67.23	24.75	12.01	**29.40**	37.95
	XQSA	**41.00**	**17.84**	**44.57**	54.46	**41.51**	**4.12**	**40.69**	58.21	**53.94**	**19.46**	**34.86**	66.14	**25.03**	**12.57**	26.05	38.55

Table 1: Performance comparison between XQSA, FRW, and DANA. $\text{mAP}_{0.5}$ values are reported separately for base (top) and novel (bottom) classes on DOTA, DIOR, Pascal VOC, and MS COCO with $K = 10$ shots. mAP values are reported for All, Small ($\sqrt{wh} < 32$), Medium ($32 \leq \sqrt{wh} < 96$) and Large ($\sqrt{wh} \geq 96$) objects.

Performance analysis:

► XQSA largely improves the detection of small objects both for natural and aerial images in the few-shot regime.

Comparison with two existing methods:

- Feature Reweighting (FRW) [Kang et al., 2019].
- Dual AwareNess Attention (DANA) [T.-I. Chen et al., 2021].
- Cross-Scale Query-Support Alignment (XQSA) [Le Jeune et al., 2023].

► Two aerial datasets DOTA and DIOR, and two natural datasets Pascal VOC and MS COCO.

		DOTA				DIOR				Pascal VOC				MS COCO			
		All	S	M	L	All	S	M	L	All	S	M	L	All	S	M	L
Base Classes	FRW	49.04	25.48	59.17	63.37	62.20	8.21	48.66	80.67	63.21	15.67	47.94	**81.73**	29.03	13.08	35.87	48.00
	DANA	**53.98**	**37.00**	**62.27**	**70.32**	**62.71**	**10.92**	**49.34**	**83.17**	**65.17**	**18.14**	**50.58**	80.11	**38.14**	**23.30**	**51.85**	**56.38**
	XQSA	51.11	26.10	59.41	64.30	59.88	10.64	45.69	82.34	62.13	15.60	48.64	75.94	31.56	16.13	40.13	49.83
Novel Classes	FRW	37.29	13.99	34.11	59.31	36.29	2.48	33.74	59.38	48.72	16.44	26.71	**68.27**	24.09	11.53	22.45	**38.69**
	DANA	36.38	14.33	40.00	**64.64**	38.18	3.21	34.91	**60.99**	52.26	10.05	24.67	67.23	24.75	12.01	**29.40**	37.95
	XQSA	**41.00**	**17.84**	**44.57**	54.46	**41.51**	**4.12**	**40.69**	58.21	**53.94**	**19.46**	**34.86**	66.14	**25.03**	**12.57**	26.05	38.55

Table 1: Performance comparison between XQSA, FRW, and DANA. $\mathrm{mAP}_{0.5}$ values are reported separately for base (top) and novel (bottom) classes on DOTA, DIOR, Pascal VOC, and MS COCO with $K = 10$ shots. mAP values are reported for All, Small ($\sqrt{wh} < 32$), Medium ($32 \leq \sqrt{wh} < 96$) and Large ($\sqrt{wh} \geq 96$) objects.

Performance analysis:

► XQSA largely improves the detection of small objects both for natural and aerial images in the few-shot regime.

► Improvements at the cost of slight performance drop on larger objects.

Comparison with two existing methods:

- Feature Reweighting (FRW) [Kang et al., 2019].
- Dual AwareNess Attention (DANA) [T.-I. Chen et al., 2021].
- Cross-Scale Query-Support Alignment (XQSA) [Le Jeune et al., 2023].

► Two aerial datasets DOTA and DIOR, and two natural datasets Pascal VOC and MS COCO.

		DOTA				DIOR				Pascal VOC				MS COCO			
		All	S	M	L	All	S	M	L	All	S	M	L	All	S	M	L
Base Classes	FRW	49.04	25.48	59.17	63.37	62.20	8.21	48.66	80.67	63.21	15.67	47.94	**81.73**	29.03	13.08	35.87	48.00
	DANA	**53.98**	**37.00**	**62.27**	**70.32**	**62.71**	**10.92**	**49.34**	**83.17**	**65.17**	**18.14**	**50.58**	80.11	**38.14**	**23.30**	**51.85**	**56.38**
	XQSA	51.11	26.10	59.41	64.30	59.88	10.64	45.69	82.34	62.13	15.60	48.64	75.94	31.56	16.13	40.13	49.83
Novel Classes	FRW	37.29	13.99	34.11	59.31	36.29	2.48	33.74	59.38	48.72	16.44	26.71	**68.27**	24.09	11.53	22.45	**38.69**
	DANA	36.38	14.33	40.00	**64.64**	38.18	3.21	34.91	**60.99**	52.26	10.05	24.67	67.23	24.75	12.01	**29.40**	37.95
	XQSA	**41.00**	**17.84**	**44.57**	54.46	**41.51**	**4.12**	**40.69**	58.21	**53.94**	**19.46**	**34.86**	66.14	**25.03**	**12.57**	26.05	38.55

Table 1: Performance comparison between XQSA, FRW, and DANA. $\text{mAP}_{0.5}$ values are reported separately for base (top) and novel (bottom) classes on DOTA, DIOR, Pascal VOC, and MS COCO with $K = 10$ shots. mAP values are reported for All, Small ($\sqrt{wh} < 32$), Medium ($32 \leq \sqrt{wh} < 96$) and Large ($\sqrt{wh} \geq 96$) objects.

Performance analysis:

- ► XQSA largely improves the detection of small objects both for natural and aerial images in the few-shot regime.
- ► Improvements at the cost of slight performance drop on larger objects.
- ► Large overall improvements on aerial datasets.

3.2 A Fine-Tuning Approach with Few-Shot Diffusion Detector

DiffusionDet translates the detection task into a denoising problem:

1. Generate random boxes.
2. Iteratively denoise the boxes to localize objects.
3. Classify objects inside the resulting boxes.

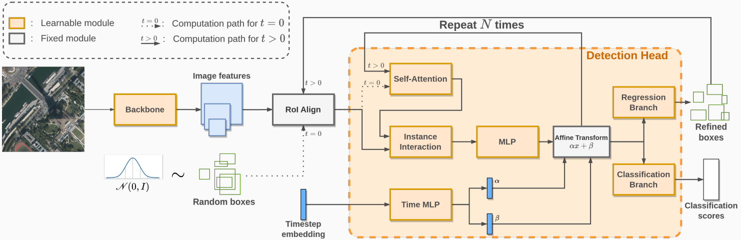


Figure 9: DiffusionDet principle.

3.2 A Fine-Tuning Approach with Few-Shot Diffusion Detector

DiffusionDet translates the detection task into a denoising problem:

1. Generate random boxes.
2. Iteratively denoise the boxes to localize objects.
3. Classify objects inside the resulting boxes.

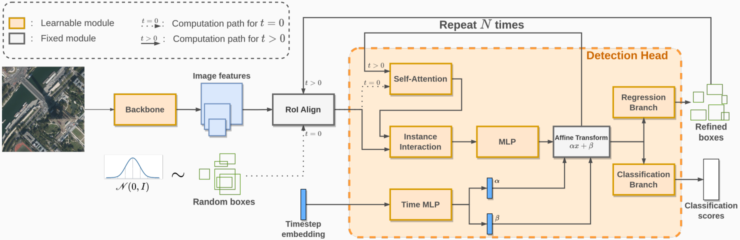


Figure 9: DiffusionDet principle.

Key properties of DiffusionDet:

- High performance on small objects in the regular setting.
- No prior on box generation (*e.g.* anchors boxes).
- Ability to increase the number of detections without retraining.

3.2 A Fine-Tuning Approach with Few-Shot Diffusion Detector

Few-Shot Diffusion Detector (FSDD): A Fine-Tuning strategy for DiffusionDet

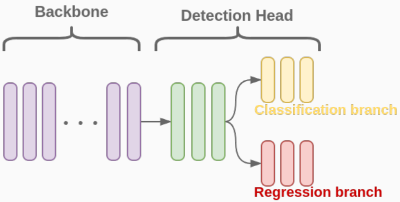


Figure 10: Per-layer representation of the detection model. Grey layers are frozen.

Few-Shot Diffusion Detector (FSDD): A Fine-Tuning strategy for DiffusionDet

1. Train DiffusionDet in a regular manner on base classes.

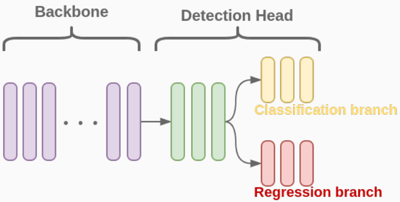


Figure 10: Per-layer representation of the detection model. Grey layers are frozen.

3.2 A Fine-Tuning Approach with Few-Shot Diffusion Detector

Few-Shot Diffusion Detector (FSDD): A Fine-Tuning strategy for DiffusionDet

1. Train DiffusionDet in a regular manner on base classes.
2. Re-initialize the last classification according to the number of novel classes.

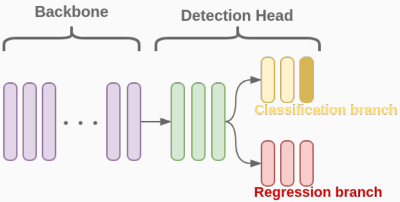


Figure 10: Per-layer representation of the detection model. Grey layers are frozen.

Few-Shot Diffusion Detector (FSDD): A Fine-Tuning strategy for DiffusionDet

1. Train DiffusionDet in a regular manner on base classes.
2. Re-initialize the last classification according to the number of novel classes.
3. Partly freeze the model.

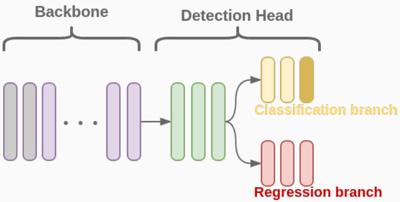


Figure 10: Per-layer representation of the detection model. Grey layers are frozen.

Few-Shot Diffusion Detector (FSDD): A Fine-Tuning strategy for DiffusionDet

1. Train DiffusionDet in a regular manner on base classes.
2. Re-initialize the last classification according to the number of novel classes.
3. Partly freeze the model.

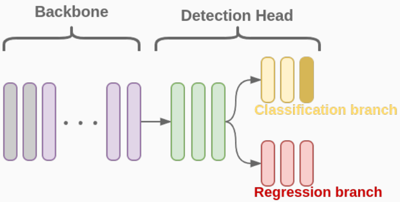


Figure 10: Per-layer representation of the detection model. Grey layers are frozen.

Freezing sweet spot:

- Backbone frozen up to stage ***i***.

Few-Shot Diffusion Detector (FSDD): A Fine-Tuning strategy for DiffusionDet

1. Train DiffusionDet in a regular manner on base classes.
2. Re-initialize the last classification according to the number of novel classes.
3. Partly freeze the model.

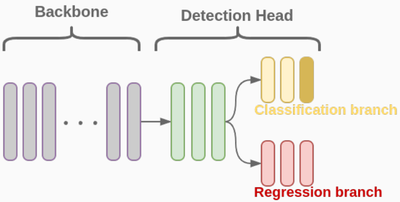


Figure 10: Per-layer representation of the detection model. Grey layers are frozen.

Freezing sweet spot:

- Backbone frozen up to stage i.
- Backbone fully frozen.

Few-Shot Diffusion Detector (FSDD): A Fine-Tuning strategy for DiffusionDet

1. Train DiffusionDet in a regular manner on base classes.
2. Re-initialize the last classification according to the number of novel classes.
3. Partly freeze the model.

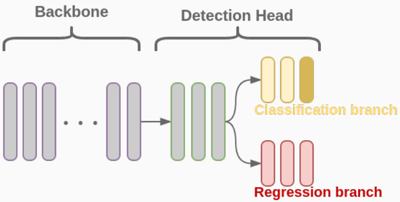


Figure 10: Per-layer representation of the detection model. Grey layers are frozen.

Freezing sweet spot:

- Backbone frozen up to stage *i*.
- Backbone fully frozen.
- Classification and regression branches only.

Few-Shot Diffusion Detector (FSDD): A Fine-Tuning strategy for DiffusionDet

1. Train DiffusionDet in a regular manner on base classes.
2. Re-initialize the last classification according to the number of novel classes.
3. Partly freeze the model.

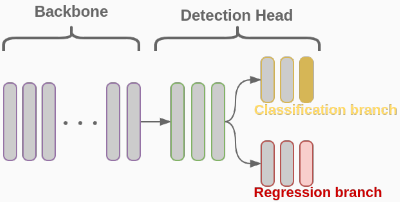


Figure 10: Per-layer representation of the detection model. Grey layers are frozen.

Freezing sweet spot:

- Backbone frozen up to stage *i*.
- Backbone fully frozen.
- Classification and regression branches only.
- Last layers only.

Few-Shot Diffusion Detector (FSDD): A Fine-Tuning strategy for DiffusionDet

1. Train DiffusionDet in a regular manner on base classes.
2. Re-initialize the last classification according to the number of novel classes.
3. Partly freeze the model.

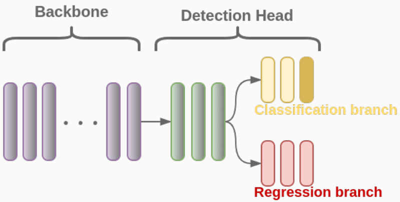


Figure 10: Per-layer representation of the detection model. Grey layers are frozen.

Freezing sweet spot:

- Backbone frozen up to stage *i*.
- Backbone fully frozen.
- Classification and regression branches only.
- Last layers only.
- Bias or norm parameters in the whole model.

3.2 A Fine-Tuning Approach with Few-Shot Diffusion Detector

Few-Shot Diffusion Detector (FSDD): A Fine-Tuning strategy for DiffusionDet

1. Train DiffusionDet in a regular manner on base classes.
2. Re-initialize the last classification according to the number of novel classes.
3. Partly freeze the model.

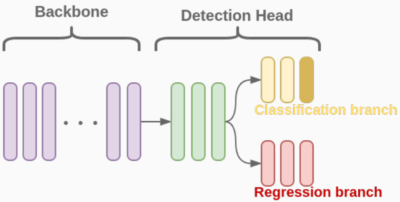


Figure 10: Per-layer representation of the detection model. Grey layers are frozen.

Freezing sweet spot:

- Backbone frozen up to stage *i*.
- Backbone fully frozen.
- Classification and regression branches only.
- Last layers only.
- Bias or norm parameters in the whole model.
- Fine-tune everything.

Few-Shot Diffusion Detector (FSDD): A Fine-Tuning strategy for DiffusionDet

1. Train DiffusionDet in a regular manner on base classes.
2. Re-initialize the last classification according to the number of novel classes.
3. Partly freeze the model.
4. Re-initialize optimizer and learning rate scheduler.

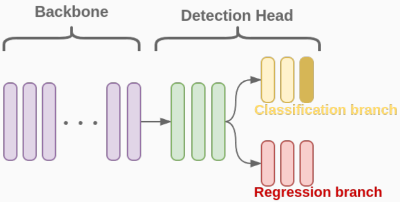


Figure 10: Per-layer representation of the detection model. Grey layers are frozen.

Freezing sweet spot:

- Backbone frozen up to stage *i*.
- Backbone fully frozen.
- Classification and regression branches only.
- Last layers only.
- Bias or norm parameters in the whole model.
- Fine-tune everything.

3.2 A Fine-Tuning Approach with Few-Shot Diffusion Detector

Few-Shot Diffusion Detector (FSDD): A Fine-Tuning strategy for DiffusionDet

1. Train DiffusionDet in a regular manner on base classes.
2. Re-initialize the last classification according to the number of novel classes.
3. Partly freeze the model.
4. Re-initialize optimizer and learning rate scheduler.
5. Train the model with the ***NK*** images available for novel classes.

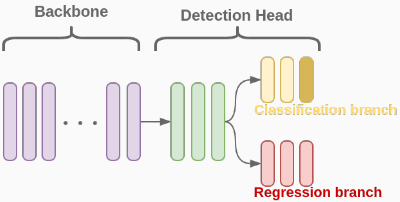


Figure 10: Per-layer representation of the detection model. Grey layers are frozen.

Freezing sweet spot:

- Backbone frozen up to stage *i*.
- Backbone fully frozen.
- Classification and regression branches only.
- Last layers only.
- Bias or norm parameters in the whole model.
- Fine-tune everything.

3.2 A Fine-Tuning Approach with Few-Shot Diffusion Detector

Experimental comparison with existing methods on DOTA, DIOR, Pascal VOC and MS COCO:

- Feature Reweighting (FRW) [Kang et al., 2019].
- Dual AwareNess Attention (DANA) [T.-I. Chen et al., 2021].
- Self-Adaptive Attention (SAA) [Z. Xiao et al., 2021]
- Prototypical Faster R-CNN (PFRCNN) [Le Jeune, Mokraoui, et al., 2021]
- Cross-Scale Query-Support Alignment (XQSA) [Le Jeune et al., 2023]

Experimental comparison with existing methods on DOTA, DIOR, Pascal VOC and MS COCO:

- Feature Reweighting (FRW) [Kang et al., 2019].
- Dual AwareNess Attention (DANA) [T.-I. Chen et al., 2021].
- Self-Adaptive Attention (SAA) [Z. Xiao et al., 2021]
- Prototypical Faster R-CNN (PFRCNN) [Le Jeune, Mokraoui, et al., 2021]
- Cross-Scale Query-Support Alignment (XQSA) [Le Jeune et al., 2023]

	DOTA				DIOR				Pascal VOC				MS COCO			
	All	S	M	L	All	S	M	L	All	S	M	L	All	S	M	L
FRW	35.29	13.99	34.11	59.31	37.29	2.48	33.74	59.38	48.72	16.44	26.71	68.27	24.09	11.53	22.45	**38.69**
DANA	36.50	14.32	40.28	**64.65**	38.18	3.21	34.91	60.99	52.26	10.05	24.67	67.23	24.75	12.01	29.40	37.95
SAA	35.12	-	-	-	32.38	-	-	-	51.70	-	-	-	21.42	-	-	-
PFRCNN	11.55	-	-	-	9.16	-	-	-	-	-	-	-	-	–	-	-
XQSA	41.00	17.84	44.57	54.46	41.51	4.12	40.69	58.21	53.94	**19.46**	**34.86**	66.14	**25.03**	**12.57**	**26.05**	38.55
FSDiffusionDet	**57.93**	**45.99**	**61.33**	53.25	**55.80**	**14.66**	**54.14**	**72.82**	**55.80**	15.05	30.20	**69.64**	24.03	5.17	19.23	38.62

Table 2: Novel classes performance of FSDiffusionDet on DOTA, DIOR, Pascal VOC and MS COCO datasets. The models employed to produce this figure have been finetuned with $K = 10$ shots.

3.2 A Fine-Tuning Approach with Few-Shot Diffusion Detector

Experimental comparison with existing methods on DOTA, DIOR, Pascal VOC and MS COCO:

- Feature Reweighting (FRW) [Kang et al., 2019].
- Dual AwareNess Attention (DANA) [T.-I. Chen et al., 2021].
- Self-Adaptive Attention (SAA) [Z. Xiao et al., 2021]
- Prototypical Faster R-CNN (PFRCNN) [Le Jeune, Mokraoui, et al., 2021]
- Cross-Scale Query-Support Alignment (XQSA) [Le Jeune et al., 2023]

	DOTA				DIOR				Pascal VOC				MS COCO			
	All	S	M	L	All	S	M	L	All	S	M	L	All	S	M	L
FRW	35.29	13.99	34.11	59.31	37.29	2.48	33.74	59.38	48.72	16.44	26.71	68.27	24.09	11.53	22.45	**38.69**
DANA	36.50	14.32	40.28	**64.65**	38.18	3.21	34.91	60.99	52.26	10.05	24.67	67.23	24.75	12.01	29.40	37.95
SAA	35.12	-	-	-	32.38	-	-	-	51.70	-	-	-	21.42	-	-	-
PFRCNN	11.55	-	-	-	9.16	-	-	-	-	-	-	-	-	–	-	-
XQSA	41.00	17.84	44.57	54.46	41.51	4.12	40.69	58.21	53.94	**19.46**	**34.86**	66.14	**25.03**	**12.57**	**26.05**	38.55
FSDiffusionDet	**57.93**	**45.99**	**61.33**	53.25	**55.80**	**14.66**	**54.14**	**72.82**	**55.80**	15.05	30.20	**69.64**	24.03	5.17	19.23	38.62

Table 2: Novel classes performance of FSDiffusionDet on DOTA, DIOR, Pascal VOC and MS COCO datasets. The models employed to produce this figure have been finetuned with $K = 10$ shots.

▶ Impressive overall performance on aerial images.

3.2 A Fine-Tuning Approach with Few-Shot Diffusion Detector

Experimental comparison with existing methods on DOTA, DIOR, Pascal VOC and MS COCO:

- Feature Reweighting (FRW) [Kang et al., 2019].
- Dual AwareNess Attention (DANA) [T.-I. Chen et al., 2021].
- Self-Adaptive Attention (SAA) [Z. Xiao et al., 2021]
- Prototypical Faster R-CNN (PFRCNN) [Le Jeune, Mokraoui, et al., 2021]
- Cross-Scale Query-Support Alignment (XQSA) [Le Jeune et al., 2023]

	DOTA				DIOR				Pascal VOC				MS COCO			
	All	S	M	L	All	S	M	L	All	S	M	L	All	S	M	L
FRW	35.29	13.99	34.11	59.31	37.29	2.48	33.74	59.38	48.72	16.44	26.71	68.27	24.09	11.53	22.45	**38.69**
DANA	36.50	14.32	40.28	**64.65**	38.18	3.21	34.91	60.99	52.26	10.05	24.67	67.23	24.75	12.01	29.40	37.95
SAA	35.12	-	-	-	32.38	-	-	-	51.70	-	-	-	21.42	-	-	-
PFRCNN	11.55	-	-	-	9.16	-	-	-	-	-	-	-	-	–	-	-
XQSA	41.00	17.84	44.57	54.46	41.51	4.12	40.69	58.21	53.94	**19.46**	**34.86**	66.14	**25.03**	**12.57**	**26.05**	38.55
FSDiffusionDet	**57.93**	**45.99**	**61.33**	53.25	**55.80**	**14.66**	**54.14**	**72.82**	**55.80**	15.05	30.20	**69.64**	24.03	5.17	19.23	38.62

Table 2: Novel classes performance of FSDiffusionDet on DOTA, DIOR, Pascal VOC and MS COCO datasets. The models employed to produce this figure have been finetuned with $K = 10$ shots.

- ▶ Impressive overall performance on aerial images.
- ▶ Large improvement on small object on aerial images.

Experimental comparison with existing methods on DOTA, DIOR, Pascal VOC and MS COCO:

- Feature Reweighting (FRW) [Kang et al., 2019].
- Dual AwareNess Attention (DANA) [T.-I. Chen et al., 2021].
- Self-Adaptive Attention (SAA) [Z. Xiao et al., 2021]
- Prototypical Faster R-CNN (PFRCNN) [Le Jeune, Mokraoui, et al., 2021]
- Cross-Scale Query-Support Alignment (XQSA) [Le Jeune et al., 2023]

	DOTA				DIOR				Pascal VOC				MS COCO			
	All	S	M	L	All	S	M	L	All	S	M	L	All	S	M	L
FRW	35.29	13.99	34.11	59.31	37.29	2.48	33.74	59.38	48.72	16.44	26.71	68.27	24.09	11.53	22.45	**38.69**
DANA	36.50	14.32	40.28	**64.65**	38.18	3.21	34.91	60.99	52.26	10.05	24.67	67.23	24.75	12.01	29.40	37.95
SAA	35.12	-	-	-	32.38	-	-	-	51.70	-	-	-	21.42	-	-	-
PFRCNN	11.55	-	-	-	9.16	-	-	-	-	-	-	-	-	–	-	-
XQSA	41.00	17.84	44.57	54.46	41.51	4.12	40.69	58.21	53.94	**19.46**	**34.86**	66.14	**25.03**	**12.57**	**26.05**	38.55
FSDiffusionDet	**57.93**	**45.99**	**61.33**	53.25	**55.80**	**14.66**	**54.14**	**72.82**	**55.80**	15.05	30.20	**69.64**	24.03	5.17	19.23	38.62

Table 2: Novel classes performance of FSDiffusionDet on DOTA, DIOR, Pascal VOC and MS COCO datasets. The models employed to produce this figure have been finetuned with $K = 10$ shots.

- ▶ Impressive overall performance on aerial images.
- ▶ Large improvement on small object on aerial images.
- ▶ Base classes performance is much higher.

	DOTA		DIOR		Pascal VOC		MS COCO	
Method	Base	Novel	Base	Novel	Base	Novel	Base	Novel
FRW	49.04	35.29	61.30	37.29	63.21	48.72	29.03	24.09
DANA	53.99	36.50	62.71	38.18	65.17	52.26	38.14	24.75
SAA	46.72	35.12	62.79	32.38	65.27	51.70	40.87	21.42
PFRCNN	36.32	11.55	42.37	9.16	-	-	-	-
XQSA	51.11	41.00	59.88	41.51	62.13	53.94	31.56	**25.03**
FSDiffusionDet	**69.58**	**57.93**	**81.71**	**55.80**	**74.63**	**55.80**	**51.91**	24.03

Table 3: FSDiffusionDet baseline compared with other FSOD methods. mAP is reported with a 0.5 IoU threshold and $K = 10$ shots.

Definition and properties of Intersection over Union

IoU is a **box similarity criterion**.

Key component of all detection frameworks: leveraged as loss function, for example selection, NMS, and model evaluation.

Definition and properties of Intersection over Union

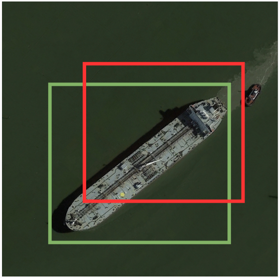
IoU is a **box similarity criterion**.

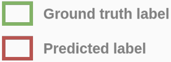
Key component of all detection frameworks: leveraged as loss function, for example selection, NMS, and model evaluation.

Definition and properties of Intersection over Union

IoU is a **box similarity criterion**.

Key component of all detection frameworks: leveraged as loss function, for example selection, NMS, and model evaluation.





Definition and properties of Intersection over Union

IoU is a **box similarity criterion**.

Key component of all detection frameworks: leveraged as loss function, for example selection, NMS, and model evaluation.

Definition and properties of Intersection over Union

IoU is a **box similarity criterion**.

Key component of all detection frameworks: leveraged as loss function, for example selection, NMS, and model evaluation.

Definition and properties of Intersection over Union

IoU is a **box similarity criterion**.

Key component of all detection frameworks: leveraged as loss function, for example selection, NMS, and model evaluation.

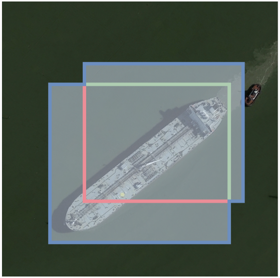
- IoU is scale-invariant

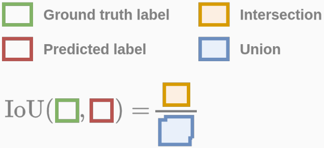
Definition and properties of Intersection over Union

IoU is a **box similarity criterion**.

Key component of all detection frameworks: leveraged as loss function, for example selection, NMS, and model evaluation.

- IoU is scale-invariant
- Scale-invariance is problematic for small objects as detectors do not have this property.



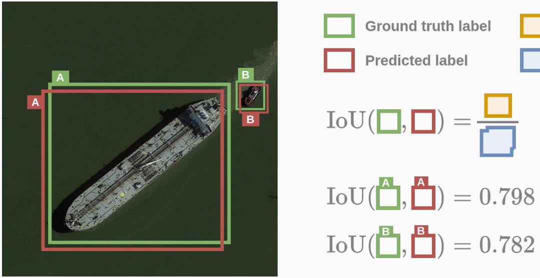


Definition and properties of Intersection over Union

IoU is a **box similarity criterion**.

Key component of all detection frameworks: leveraged as loss function, for example selection, NMS, and model evaluation.

- IoU is scale-invariant
- Scale-invariance is problematic for small objects as detectors do not have this property.
- Small prediction shift can have a large influence on the IoU with ground truth: **problematic for training and evaluation**.

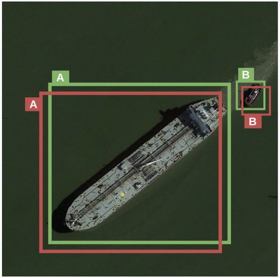


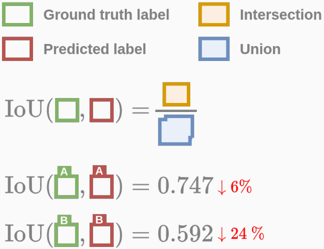
Definition and properties of Intersection over Union

IoU is a **box similarity criterion**.

Key component of all detection frameworks: leveraged as loss function, for example selection, NMS, and model evaluation.

- IoU is scale-invariant
- Scale-invariance is problematic for small objects as detectors do not have this property.
- Small prediction shift can have a large influence on the IoU with ground truth: **problematic for training and evaluation.**





Scale-Adaptive Intersection over Union (SIoU) [Le Jeune et al., 2023]

$$\mathrm{SIoU}(b_1, b_2) = \mathrm{IoU}(b_1, b_2)^p \qquad \text{with } p = 1 - \gamma e^{-\frac{\sqrt{a}}{\kappa}} \tag{6}$$

where a is the mean area of the two boxes b_1 and b_2 ($a = \frac{w_1 h_1 + w_2 h_2}{2}$).
$\gamma \in [-\infty, 1]$ and $\kappa \in \mathbb{R}_+^*$ are hyper-parameters to control SIoU's behavior.

$b_1 = (x_1, y_1, w_1, h_1)$ and $b_2 = (x_2, y_2, w_2, h_2)$.

Scale-Adaptive Intersection over Union (SIoU) [Le Jeune et al., 2023]

$$\mathrm{SIoU}(b_1, b_2) = \mathrm{IoU}(b_1, b_2)^p \qquad \text{with } p = 1 - \gamma e^{-\frac{\sqrt{a}}{\kappa}} \tag{6}$$

where a is the mean area of the two boxes b_1 and b_2 ($a = \frac{w_1 h_1 + w_2 h_2}{2}$).
$\gamma \in [-\infty, 1]$ and $\kappa \in \mathbb{R}_+^*$ are hyper-parameters to control SIoU's behavior.

$b_1 = (x_1, y_1, w_1, h_1)$ and $b_2 = (x_2, y_2, w_2, h_2)$.

Scale invariance of IoU is relaxed in a controllable manner.

- γ controls the direction of the relaxation: criterion values are either **boosted** or **decreased**.

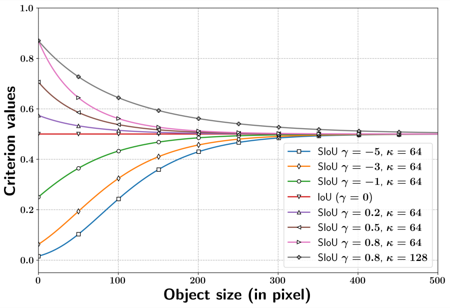


Figure 11: Illustration of the scale invariance relaxation of SIoU.

Scale-Adaptive Intersection over Union (SIoU) [Le Jeune et al., 2023]

$$\text{SIoU}(\boldsymbol{b_1}, \boldsymbol{b_2}) = \text{IoU}(\boldsymbol{b_1}, \boldsymbol{b_2})^{\boldsymbol{p}} \qquad \text{with } \boldsymbol{p} = \boldsymbol{1} - \gamma \boldsymbol{e}^{-\frac{\sqrt{\boldsymbol{a}}}{\kappa}} \tag{6}$$

where $\boldsymbol{a}$ is the mean area of the two boxes $\boldsymbol{b_1}$ and $\boldsymbol{b_2}$ ($\boldsymbol{a} = \frac{w_1h_1+w_2h_2}{2}$).
$\gamma \in [-\infty, \boldsymbol{1}]$ and $\kappa \in \mathbb{R}_+^*$ are hyper-parameters to control SIoU's behavior.

$\boldsymbol{b_1} = (x_1, y_1, w_1, h_1)$ and $\boldsymbol{b_2} = (x_2, y_2, w_2, h_2)$.

Scale invariance of IoU is relaxed in a controllable manner.

- γ controls the direction of the relaxation: criterion values are either **boosted** or **decreased**.

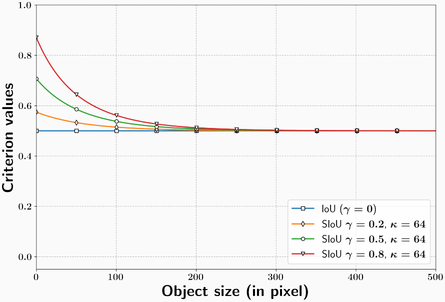


Figure 11: Illustration of the scale invariance relaxation of SIoU.

Scale-Adaptive Intersection over Union (SIoU) [Le Jeune et al., 2023]

$$\mathrm{SIoU}(b_1, b_2) = \mathrm{IoU}(b_1, b_2)^p \qquad \text{with } p = 1 - \gamma e^{-\frac{\sqrt{a}}{\kappa}} \tag{6}$$

where a is the mean area of the two boxes b_1 and b_2 ($a = \frac{w_1 h_1 + w_2 h_2}{2}$).
$\gamma \in [-\infty, 1]$ and $\kappa \in \mathbb{R}_+^*$ are hyper-parameters to control SIoU's behavior.

$b_1 = (x_1, y_1, w_1, h_1)$ and $b_2 = (x_2, y_2, w_2, h_2)$.

Scale invariance of IoU is relaxed in a controllable manner.

- γ controls the direction of the relaxation: criterion values are either **boosted** or **decreased**.

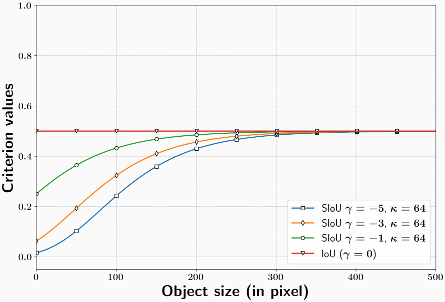


Figure 11: Illustration of the scale invariance relaxation of SIoU.

Scale-Adaptive Intersection over Union (SIoU) [Le Jeune et al., 2023]

$$\mathrm{SIoU}(b_1, b_2) = \mathrm{IoU}(b_1, b_2)^p \qquad \text{with } p = 1 - \gamma e^{-\frac{\sqrt{a}}{\kappa}} \tag{6}$$

where a is the mean area of the two boxes b_1 and b_2 ($a = \frac{w_1 h_1 + w_2 h_2}{2}$).
$\gamma \in [-\infty, 1]$ and $\kappa \in \mathbb{R}_+^*$ are hyper-parameters to control SIoU's behavior.

$b_1 = (x_1, y_1, w_1, h_1)$ and $b_2 = (x_2, y_2, w_2, h_2)$.

Scale invariance of IoU is relaxed in a controllable manner.

- γ controls the direction of the relaxation: criterion values are either **boosted** or **decreased**.
- κ controls the speed at which IoU's behavior is recovered.

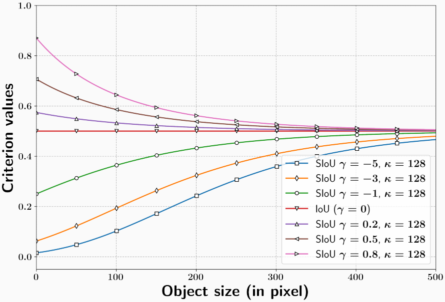


Figure 11: Illustration of the scale invariance relaxation of SIoU.

Scale-Adaptive Intersection over Union (SIoU) [Le Jeune et al., 2023]

$$\mathrm{SIoU}(b_1, b_2) = \mathrm{IoU}(b_1, b_2)^p \qquad \text{with } p = 1 - \gamma e^{-\frac{\sqrt{a}}{\kappa}} \tag{6}$$

where a is the mean area of the two boxes b_1 and b_2 ($a = \frac{w_1 h_1 + w_2 h_2}{2}$).
$\gamma \in [-\infty, 1]$ and $\kappa \in \mathbb{R}_+^*$ are hyper-parameters to control SIoU's behavior.

$b_1 = (x_1, y_1, w_1, h_1)$ and $b_2 = (x_2, y_2, w_2, h_2)$.

Scale invariance of IoU is relaxed in a controllable manner.

- γ controls the direction of the relaxation: criterion values are either **boosted** or **decreased**.
- κ controls the speed at which IoU's behavior is recovered.

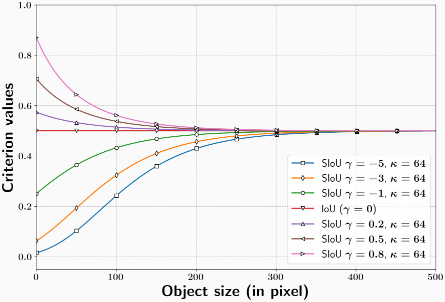


Figure 11: Illustration of the scale invariance relaxation of SIoU.

IoU and SIoU as loss functions.

IoU Loss:

$$\mathcal{L}_{\mathrm{IoU}}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i) = \mathbf{1} - \mathrm{IoU}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i), \qquad (7)$$

$$\mathrm{SIoU}(\boldsymbol{b_1}, \boldsymbol{b_2}) = \mathrm{IoU}(\boldsymbol{b_1}, \boldsymbol{b_2})^{\boldsymbol{p}},$$
$$\text{with } \boldsymbol{p} = \mathbf{1} - \gamma \boldsymbol{e}^{-\frac{\sqrt{\boldsymbol{a}}}{\kappa}}.$$

IoU and SIoU as loss functions.

IoU Loss:

$$\mathcal{L}_{\mathrm{IoU}}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i) = \mathbf{1} - \mathrm{IoU}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i), \tag{7}$$

SIoU Loss:

$$\mathcal{L}_{\mathrm{SIoU}}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i) = \mathbf{1} - \mathrm{SIoU}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i). \tag{8}$$

$$\mathrm{SIoU}(b_1, b_2) = \mathrm{IoU}(b_1, b_2)^p, \quad \text{with } p = 1 - \gamma e^{-\frac{\sqrt{a}}{\kappa}}.$$

IoU and SIoU as loss functions.

IoU Loss:

$$\mathcal{L}_{\mathrm{IoU}}(\hat{b}_i, b_i) = 1 - \mathrm{IoU}(\hat{b}_i, b_i), \tag{7}$$

SIoU Loss:

$$\mathcal{L}_{\mathrm{SIoU}}(\hat{b}_i, b_i) = 1 - \mathrm{SIoU}(\hat{b}_i, b_i). \tag{8}$$

Both loss functions can be extended to output negative values when boxes do not overlap, following Generalized IoU [Rezatofighi et al., 2019].

$$\boxed{\begin{aligned}\mathrm{SIoU}(\boldsymbol{b}_1, \boldsymbol{b}_2) &= \mathrm{IoU}(\boldsymbol{b}_1, \boldsymbol{b}_2)^{\boldsymbol{p}}, \\ \text{with } \boldsymbol{p} &= \mathbf{1} - \gamma \boldsymbol{e}^{-\frac{\sqrt{\boldsymbol{a}}}{\kappa}}.\end{aligned}}$$

IoU and SIoU as loss functions.

IoU Loss:

$$\mathcal{L}_{\mathrm{IoU}}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i) = \mathbf{1} - \mathrm{IoU}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i), \tag{7}$$

SIoU Loss:

$$\mathcal{L}_{\mathrm{SIoU}}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i) = \mathbf{1} - \mathrm{SIoU}(\hat{\boldsymbol{b}}_i, \boldsymbol{b}_i). \tag{8}$$

Both loss functions can be extended to output negative values when boxes do not overlap, following Generalized IoU [Rezatofighi et al., 2019].

SIoU loss can control the training balance between small and large objects.

3.3 Scale-Adaptive IoU for Training Few-Shot Detection Models

$$\mathrm{SIoU}(b_1, b_2) = \mathrm{IoU}(b_1, b_2)^p,$$
$$\text{with } p = 1 - \gamma e^{-\frac{\sqrt{a}}{\kappa}}.$$

IoU and SIoU as loss functions.

IoU Loss:

$$\mathcal{L}_{\mathrm{IoU}}(\hat{b}_i, b_i) = 1 - \mathrm{IoU}(\hat{b}_i, b_i), \quad (7)$$

SIoU Loss:

$$\mathcal{L}_{\mathrm{SIoU}}(\hat{b}_i, b_i) = 1 - \mathrm{SIoU}(\hat{b}_i, b_i). \quad (8)$$

Both loss functions can be extended to output negative values when boxes do not overlap, following Generalized IoU [Rezatofighi et al., 2019].

SIoU loss can control the training balance between small and large objects.

- With $\gamma < 0$, $\mathrm{SIoU}(\hat{b}_i, b_i) \leq \mathrm{IoU}(\hat{b}_i, b_i)$, hence $\mathcal{L}_{\mathrm{SIoU}}(\hat{b}_i, b_i) \geq \mathcal{L}_{\mathrm{IoU}}(\hat{b}_i, b_i)$.

3.3 Scale-Adaptive IoU for Training Few-Shot Detection Models

$$\mathrm{SIoU}(b_1, b_2) = \mathrm{IoU}(b_1, b_2)^p, \quad \text{with } p = 1 - \gamma e^{-\frac{\sqrt{a}}{\kappa}}.$$

IoU and SIoU as loss functions.

IoU Loss:

$$\mathcal{L}_{\mathrm{IoU}}(\hat{b}_i, b_i) = 1 - \mathrm{IoU}(\hat{b}_i, b_i), \qquad (7)$$

SIoU Loss:

$$\mathcal{L}_{\mathrm{SIoU}}(\hat{b}_i, b_i) = 1 - \mathrm{SIoU}(\hat{b}_i, b_i). \qquad (8)$$

Both loss functions can be extended to output negative values when boxes do not overlap, following Generalized IoU [Rezatofighi et al., 2019].

SIoU loss can control the training balance between small and large objects.

- With $\gamma < 0$, $\mathrm{SIoU}(\hat{b}_i, b_i) \leq \mathrm{IoU}(\hat{b}_i, b_i)$, hence $\mathcal{L}_{\mathrm{SIoU}}(\hat{b}_i, b_i) \geq \mathcal{L}_{\mathrm{IoU}}(\hat{b}_i, b_i)$.
- Small objects have more influence on the overall loss.

3.3 Scale-Adaptive IoU for Training Few-Shot Detection Models

$$\mathrm{SIoU}(b_1, b_2) = \mathrm{IoU}(b_1, b_2)^p, \quad \text{with } p = 1 - \gamma e^{-\frac{\sqrt{a}}{\kappa}}.$$

IoU and SIoU as loss functions.

IoU Loss:

$$\mathcal{L}_{\mathrm{IoU}}(\hat{b}_i, b_i) = 1 - \mathrm{IoU}(\hat{b}_i, b_i), \tag{7}$$

SIoU Loss:

$$\mathcal{L}_{\mathrm{SIoU}}(\hat{b}_i, b_i) = 1 - \mathrm{SIoU}(\hat{b}_i, b_i). \tag{8}$$

Both loss functions can be extended to output negative values when boxes do not overlap, following Generalized IoU [Rezatofighi et al., 2019].

SIoU loss can control the training balance between small and large objects.

- With $\gamma < 0$, $\mathrm{SIoU}(\hat{b}_i, b_i) \leq \mathrm{IoU}(\hat{b}_i, b_i)$, hence $\mathcal{L}_{\mathrm{SIoU}}(\hat{b}_i, b_i) \geq \mathcal{L}_{\mathrm{IoU}}(\hat{b}_i, b_i)$.
- Small objects have more influence on the overall loss.
- Training is then biased to improve the localization of small targets.

$$\text{SIoU}(b_1, b_2) = \text{IoU}(b_1, b_2)^p, \quad \text{with } p = 1 - \gamma e^{-\frac{\sqrt{a}}{\kappa}}.$$

IoU and SIoU as loss functions.

IoU Loss:

$$\mathcal{L}_{\text{IoU}}(\hat{b}_i, b_i) = 1 - \text{IoU}(\hat{b}_i, b_i), \tag{7}$$

SIoU Loss:

$$\mathcal{L}_{\text{SIoU}}(\hat{b}_i, b_i) = 1 - \text{SIoU}(\hat{b}_i, b_i). \tag{8}$$

Both loss functions can be extended to output negative values when boxes do not overlap, following Generalized IoU [Rezatofighi et al., 2019].

SIoU loss can control the training balance between small and large objects.

- With $\gamma < 0$, $\text{SIoU}(\hat{b}_i, b_i) \leq \text{IoU}(\hat{b}_i, b_i)$, hence $\mathcal{L}_{\text{SIoU}}(\hat{b}_i, b_i) \geq \mathcal{L}_{\text{IoU}}(\hat{b}_i, b_i)$.
- Small objects have more influence on the overall loss.
- Training is then biased to improve the localization of small targets.

With $\gamma > 0$, SIoU becomes more suitable than IoU for model evaluation as it aligns better with human perception (shown with a user study).

Comparison with existing criteria on DOTA:

- IoU and GIoU [Rezatofighi et al., 2019]
- α-IoU [He et al., 2021]
- Normalized Wasserstein Distance (NWD) [C. Xu et al., 2022]
- Scale-Adaptive Intersection over Union (SIoU) [Le Jeune et al., 2023]

Loss	Base classes All	Base classes S	Base classes M	Base classes L	Novel Classes All	Novel Classes S	Novel Classes M	Novel Classes L
IoU	50.67	**25.83**	57.49	68.24	32.41	10.06	47.87	67.09
α-IoU	46.72	13.24	55.21	**69.94**	33.95	12.58	46.58	**74.50**
SIoU	**53.62**	24.07	**61.91**	67.34	**39.05**	**16.59**	**54.42**	**74.49**
NWD	50.79	19.19	58.90	67.90	41.65	28.26	50.16	65.06
GIoU	52.41	**26.94**	61.17	63.00	41.03	24.01	**52.13**	69.78
GSIoU	**52.91**	22.14	**61.19**	**66.02**	**45.88**	**34.83**	51.26	**70.78**

Table 4: Few-shot performance comparison between several criteria: IoU, α-IoU, SIoU, NWD, GIoU, and GSIoU trained on DOTA. mAP is reported with a 0.5 IoU threshold for small (S), medium (M), large (L), and all objects. $K = 10$ shots.

	XQSA	Base classes All	Base classes S	Base classes M	Base classes L	Novel Classes All	Novel Classes S	Novel Classes M	Novel Classes L
DOTA	w/ GIoU	52.41	**26.94**	61.17	63.00	41.03	24.01	**52.13**	69.78
	w/ GSIoU	**52.91**	22.14	**61.19**	**66.02**	**45.88**	**34.83**	51.26	**70.78**
DIOR	w/ GIoU	58.90	10.38	40.76	80.44	47.93	9.85	47.61	68.40
	w/ GSIoU	**60.29**	**11.28**	**43.24**	**81.63**	**52.85**	**13.78**	**53.73**	**71.22**
Pascal	w/ GIoU	51.09	**13.93**	**40.26**	62.01	48.42	18.44	36.06	59.99
	w/ GSIoU	**54.47**	13.88	40.13	**66.82**	**55.16**	**22.94**	**36.24**	**67.40**
COCO	w/ GIoU	19.15	**8.72**	22.50	30.59	26.25	11.96	23.95	38.60
	w/ GSIoU	**19.57**	8.41	**23.02**	**31.07**	**27.11**	**12.81**	**26.02**	**39.20**

Table 5: Few-shot performance on four datasets: DOTA, DIOR, Pascal VOC and COCO. GIoU and GSIoU losses are compared. mAP is reported with a 0.5 IoU threshold and for all object sizes. $K = 10$ shots.

Comparison with existing criteria on DOTA:

- IoU and GIoU [Rezatofighi et al., 2019]
- α-IoU [He et al., 2021]
- Normalized Wasserstein Distance (NWD) [C. Xu et al., 2022]
- Scale-Adaptive Intersection over Union (SIoU) [Le Jeune et al., 2023]

	Base classes				Novel Classes			
Loss	All	S	M	L	All	S	M	L
IoU	50.67	25.83	57.49	68.24	32.41	10.06	47.87	67.09
α-IoU	46.72	13.24	55.21	69.94	33.95	12.58	46.58	74.50
SIoU	53.62	24.07	61.91	67.34	39.05	16.59	54.42	74.49
NWD	50.79	19.19	58.90	67.90	41.65	28.26	50.16	65.06
GIoU	52.41	26.94	61.17	63.00	41.03	24.01	52.13	69.78
GSIoU	52.91	22.14	61.19	66.02	45.88	34.83	51.26	70.78

Table 4: Few-shot performance comparison between several criteria: IoU, α-IoU, SIoU, NWD, GIoU, and GSIoU trained on DOTA. mAP is reported with a 0.5 IoU threshold for small (S), medium (M), large (L), and all objects. $K = 10$ shots.

		Base classes				Novel Classes			
	XQSA	All	S	M	L	All	S	M	L
DOTA	w/ GIoU	52.41	26.94	61.17	63.00	41.03	24.01	52.13	69.78
	w/ GSIoU	52.91	22.14	61.19	66.02	45.88	34.83	51.26	70.78
DIOR	w/ GIoU	58.90	10.38	40.76	80.44	47.93	9.85	47.61	68.40
	w/ GSIoU	60.29	11.28	43.24	81.63	52.85	13.78	53.73	71.22
Pascal	w/ GIoU	51.09	13.93	40.26	62.01	48.42	18.44	36.06	59.99
	w/ GSIoU	54.47	13.88	40.13	66.82	55.16	22.94	36.24	67.40
COCO	w/ GIoU	19.15	8.72	22.50	30.59	26.25	11.96	23.95	38.60
	w/ GSIoU	19.57	8.41	23.02	31.07	27.11	12.81	26.02	39.20

Table 5: Few-shot performance on four datasets: DOTA, DIOR, Pascal VOC and COCO. GIoU and GSIoU losses are compared. mAP is reported with a 0.5 IoU threshold and for all object sizes. $K = 10$ shots.

► SIoU and GSIoU losses dominate other critera.

Comparison with existing criteria on DOTA:

- IoU and GIoU [Rezatofighi et al., 2019]
- α-IoU [He et al., 2021]
- Normalized Wasserstein Distance (NWD) [C. Xu et al., 2022]
- Scale-Adaptive Intersection over Union (SIoU) [Le Jeune et al., 2023]

	Base classes				Novel Classes			
Loss	All	S	M	L	All	S	M	L
IoU	50.67	**25.83**	57.49	68.24	32.41	10.06	47.87	67.09
α-IoU	46.72	13.24	55.21	**69.94**	33.95	12.58	46.58	**74.50**
SIoU	**53.62**	24.07	**61.91**	67.34	**39.05**	**16.59**	**54.42**	74.49
NWD	50.79	19.19	58.90	67.90	41.65	28.26	50.16	65.06
GIoU	52.41	**26.94**	61.17	63.00	41.03	24.01	**52.13**	69.78
GSIoU	**52.91**	22.14	**61.19**	**66.02**	**45.88**	**34.83**	51.26	**70.78**

Table 4: Few-shot performance comparison between several criteria: IoU, α-IoU, SIoU, NWD, GIoU, and GSIoU trained on DOTA. mAP is reported with a 0.5 IoU threshold for small (S), medium (M), large (L), and all objects. $K = 10$ shots.

		Base classes				Novel Classes			
	XQSA	All	S	M	L	All	S	M	L
DOTA	w/ GIoU	52.41	**26.94**	61.17	63.00	41.03	24.01	**52.13**	69.78
	w/ GSIoU	**52.91**	22.14	**61.19**	**66.02**	**45.88**	**34.83**	51.26	**70.78**
DIOR	w/ GIoU	58.90	10.38	40.76	80.44	47.93	9.85	47.61	68.40
	w/ GSIoU	**60.29**	**11.28**	**43.24**	**81.63**	**52.85**	**13.78**	**53.73**	**71.22**
Pascal	w/ GIoU	51.09	**13.93**	**40.26**	62.01	48.42	18.44	36.06	59.99
	w/ GSIoU	**54.47**	13.88	40.13	**66.82**	**55.16**	**22.94**	**36.24**	**67.40**
COCO	w/ GIoU	19.15	**8.72**	22.50	30.59	26.25	11.96	23.95	38.60
	w/ GSIoU	**19.57**	8.41	**23.02**	**31.07**	**27.11**	**12.81**	**26.02**	**39.20**

Table 5: Few-shot performance on four datasets: DOTA, DIOR, Pascal VOC and COCO. GIoU and GSIoU losses are compared. mAP is reported with a 0.5 IoU threshold and for all object sizes. $K = 10$ shots.

- SIoU and GSIoU losses dominate other critera.
- SIoU and GSIoU brings large improvements for small object detection.

3.3 Scale-Adaptive IoU for Training Few-Shot Detection Models

Comparison with existing criteria on DOTA:

- IoU and GIoU [Rezatofighi et al., 2019]
- α-IoU [He et al., 2021]
- Normalized Wasserstein Distance (NWD) [C. Xu et al., 2022]
- Scale-Adaptive Intersection over Union (SIoU) [Le Jeune et al., 2023]

	Base classes				Novel Classes			
Loss	**All**	**S**	**M**	**L**	**All**	**S**	**M**	**L**
IoU	50.67	**25.83**	57.49	68.24	32.41	10.06	47.87	67.09
α-IoU	46.72	13.24	55.21	**69.94**	33.95	12.58	46.58	**74.50**
SIoU	**53.62**	24.07	**61.91**	67.34	**39.05**	**16.59**	**54.42**	**74.49**
NWD	50.79	19.19	58.90	67.90	41.65	28.26	50.16	65.06
GIoU	52.41	**26.94**	61.17	63.00	41.03	24.01	**52.13**	69.78
GSIoU	**52.91**	22.14	**61.19**	**66.02**	**45.88**	**34.83**	51.26	**70.78**

Table 4: Few-shot performance comparison between several criteria: IoU, α-IoU, SIoU, NWD, GIoU, and GSIoU trained on DOTA. mAP is reported with a 0.5 IoU threshold for small (S), medium (M), large (L), and all objects. $K = 10$ shots.

		Base classes				Novel Classes			
	XQSA	**All**	**S**	**M**	**L**	**All**	**S**	**M**	**L**
DOTA	w/ GIoU	52.41	**26.94**	61.17	63.00	41.03	24.01	**52.13**	69.78
	w/ GSIoU	**52.91**	22.14	**61.19**	**66.02**	**45.88**	**34.83**	51.26	**70.78**
DIOR	w/ GIoU	58.90	10.38	40.76	80.44	47.93	9.85	47.61	68.40
	w/ GSIoU	**60.29**	**11.28**	**43.24**	**81.63**	**52.85**	**13.78**	**53.73**	**71.22**
Pascal	w/ GIoU	51.09	**13.93**	**40.26**	62.01	48.42	18.44	36.06	59.99
	w/ GSIoU	**54.47**	13.88	40.13	**66.82**	**55.16**	**22.94**	**36.24**	**67.40**
COCO	w/ GIoU	19.15	**8.72**	22.50	30.59	26.25	11.96	23.95	38.60
	w/ GSIoU	**19.57**	8.41	**23.02**	**31.07**	**27.11**	**12.81**	**26.02**	**39.20**

Table 5: Few-shot performance on four datasets: DOTA, DIOR, Pascal VOC and COCO. GIoU and GSIoU losses are compared. mAP is reported with a 0.5 IoU threshold and for all object sizes. $K = 10$ shots.

- SIoU and GSIoU losses dominate other critera.
- SIoU and GSIoU brings large improvements for small object detection.
- For aerial images, it induces large overall detection performance gains.

3.4 Key Takeaways

Three contributions for small object detection in the few-shot regime

Three contributions for small object detection in the few-shot regime

1. Cross-Scale Query-Support Alignment (XQSA), an attention mechanism for small FSOD.
   - 👍 XQSA largely improves the detection performance of small objects in the few-shot regime.
   - 👍 Very helpful for aerial images.
   - 👎 Improvements on small targets at the cost of larger objects, more polyvalent attention mechanisms should be developed.

Three contributions for small object detection in the few-shot regime

1. Cross-Scale Query-Support Alignment (XQSA), an attention mechanism for small FSOD.
   - 👍 XQSA largely improves the detection performance of small objects in the few-shot regime.
   - 👍 Very helpful for aerial images.
   - 👎 Improvements on small targets at the cost of larger objects, more polyvalent attention mechanisms should be developed.

2. Few-Shot DiffusiontDet a fine-tuning-based approach for small FSOD.
   - 👍 Substantial improvements on FSOD for aerial images with large gains on small objects with learnable box prior.
   - 👍 Much easier to train and scales better with the number of shots.
   - 👎 Find a way to predict how much freezing will be optimal for a dataset.

3.4 Key Takeaways

Three contributions for small object detection in the few-shot regime

1. Cross-Scale Query-Support Alignment (XQSA), an attention mechanism for small FSOD.
   - 👍 XQSA largely improves the detection performance of small objects in the few-shot regime.
   - 👍 Very helpful for aerial images.
   - 👎 Improvements on small targets at the cost of larger objects, more polyvalent attention mechanisms should be developed.

2. Few-Shot DiffusiontDet a fine-tuning-based approach for small FSOD.
   - 👍 Substantial improvements on FSOD for aerial images with large gains on small objects with learnable box prior.
   - 👍 Much easier to train and scales better with the number of shots.
   - 👎 Find a way to predict how much freezing will be optimal for a dataset.

3. Scaled-Adaptative Intersection over Union (SIoU) is a controllable relaxation of IoU.
   - 👍 Largely improves small object detection in the few-shot regime by shifting the training balance between small and large objects.
   - 👍 Better aligned with human perception and well-suited for model evaluation.
   - 👎 Limited gains in regular settings and with DiffusionDet.
   - 👎 Requires the tuning of γ and κ.

4

Addressing more Complex Scenarios

Promising performance of FSDiffusionDet allows envisioning Cross-Domain applications

Differences with Few-Shot Object Detection:

- Use two separate datasets between base training and fine-tuning.

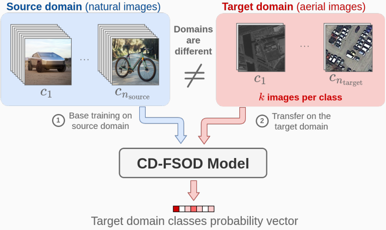


Figure 12: Cross-Domain Few-Shot Object Detection (CD-FSOD).

Promising performance of FSDiffusionDet allows envisioning Cross-Domain applications

Differences with Few-Shot Object Detection:

- Use two separate datasets between base training and fine-tuning.
- Base classes are all classes of the source dataset.

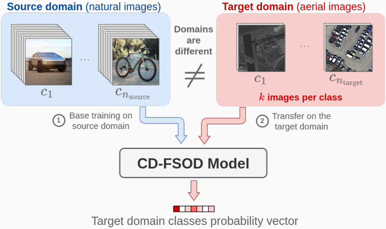


Figure 12: Cross-Domain Few-Shot Object Detection (CD-FSOD).

Promising performance of FSDiffusionDet allows envisioning Cross-Domain applications

Differences with Few-Shot Object Detection:

- Use two separate datasets between base training and fine-tuning.
- Base classes are all classes of the source dataset.
- Novel classes are all classes of the target dataset.

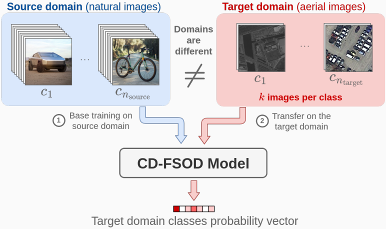


Figure 12: Cross-Domain Few-Shot Object Detection (CD-FSOD).

Promising performance of FSDiffusionDet allows envisioning Cross-Domain applications

Differences with Few-Shot Object Detection:

- Use two separate datasets between base training and fine-tuning.
- Base classes are all classes of the source dataset.
- Novel classes are all classes of the target dataset.
- Target dataset only has ***K*** images for each class.

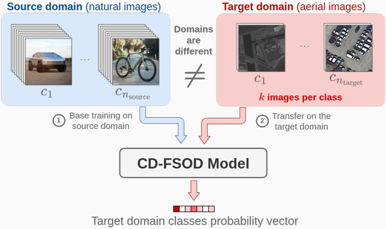


Figure 12: Cross-Domain Few-Shot Object Detection (CD-FSOD).

Promising performance of FSDiffusionDet allows envisioning Cross-Domain applications

Differences with Few-Shot Object Detection:

- Use two separate datasets between base training and fine-tuning.
- Base classes are all classes of the source dataset.
- Novel classes are all classes of the target dataset.
- Target dataset only has K images for each class.

Only interested in the detection performance on the novel classes (*i.e.* the target classes).

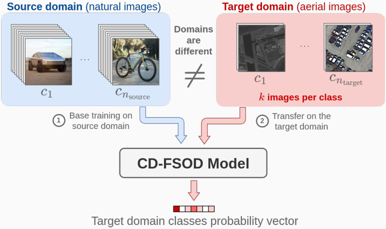


Figure 12: Cross-Domain Few-Shot Object Detection (CD-FSOD).

COCO → Anything scenarios

Base training on COCO and fine-tuning on another dataset, experiment with **DOTA** [Xia et al., 2018], **DIOR** [K. Li et al., 2020], **DeepFruit**, **SIXRay** [Miao et al., 2019], **CipArt** [Inoue et al., 2018], **VisDrone** [Y. Sun et al., 2022].

K Shots	DIOR	DOTA	DeepFruits	SIXRay	ClipArt	VisDrone
1	11.10 ± 0.32	4.03 ± 0.26	38.47 ± 1.42	4.80 ± 0.87	2.09± 0.19	2.83± 0.17
5	30.42 ± 0.69	14.45 ± 0.43	55.58 ± 1.36	13.25 ±1.14	5.26 ± 0.15	5.74± 0.22
10	38.73 ± 0.65	25.02 ± 0.65	68.37 ± 2.01	21.26 ±1.33	5.69 ± 0.10	7.50± 0.10
20	48.23 ± 0.33	33.31 ± 0.46	73.95 ± 0.53	30.06 ±1.09	6.10 ± 0.22	9.14± 0.35
50	56.97 ± 0.60	43.23 ± 0.68	76.65 ± 0.78	41.93 ±1.02	6.44 ± 0.16	11.47± 0.27

Table 6: Cross-domain performance results on 6 scenarios COCO → DIOR / DOTA / DeepFruits / SIXRay / ClipArt / VisDrone. The average $\text{mAP}_{0.5}$ is reported with a 95% confidence interval.

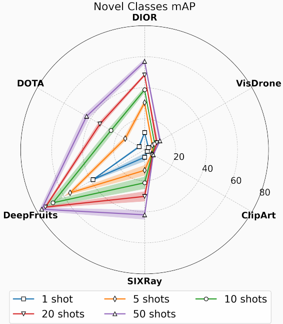


Figure 13: Performance of FSDiffusionDet on multiple COCO ⟶ X scenarios.

COCO $\longrightarrow$ Anything scenarios

Base training on COCO and fine-tuning on another dataset, experiment with **DOTA** [Xia et al., 2018], **DIOR** [K. Li et al., 2020], **DeepFruit**, **SIXRay** [Miao et al., 2019], **CipArt** [Inoue et al., 2018], **VisDrone** [Y. Sun et al., 2022].

K Shots	DIOR	DOTA	DeepFruits	SIXRay	ClipArt	VisDrone
1	11.10 ± 0.32	4.03 ± 0.26	38.47 ± 1.42	4.80 ± 0.87	2.09± 0.19	2.83± 0.17
5	30.42 ± 0.69	14.45 ± 0.43	55.58 ± 1.36	13.25 ±1.14	5.26 ± 0.15	5.74± 0.22
10	38.73 ± 0.65	25.02 ± 0.65	68.37 ± 2.01	21.26 ±1.33	5.69 ± 0.10	7.50± 0.10
20	48.23 ± 0.33	33.31 ± 0.46	73.95 ± 0.53	30.06 ±1.09	6.10 ± 0.22	9.14± 0.35
50	56.97 ± 0.60	43.23 ± 0.68	76.65 ± 0.78	41.93 ±1.02	6.44 ± 0.16	11.47± 0.27

Table 6: Cross-domain performance results on 6 scenarios COCO $\longrightarrow$ DIOR / DOTA / DeepFruits / SIXRay / ClipArt / VisDrone. The average $\text{mAP}_{0.5}$ is reported with a 95% confidence interval.

- As in classical Few-Shot setting, performance improves with K. Promising performance with a reasonable amount of annotations.

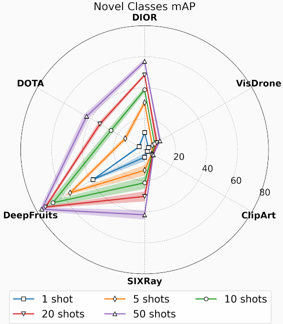


Figure 13: Performance of FSDiffusionDet on multiple COCO $\longrightarrow$ X scenarios.

COCO $\rightarrow$ Anything scenarios

Base training on COCO and fine-tuning on another dataset, experiment with **DOTA** [Xia et al., 2018], **DIOR** [K. Li et al., 2020], **DeepFruit**, **SIXRay** [Miao et al., 2019], **CipArt** [Inoue et al., 2018], **VisDrone** [Y. Sun et al., 2022].

K Shots	DIOR	DOTA	DeepFruits	SIXRay	ClipArt	VisDrone
1	11.10 $\pm$ 0.32	4.03 $\pm$ 0.26	38.47 $\pm$ 1.42	4.80 $\pm$ 0.87	2.09$\pm$ 0.19	2.83$\pm$ 0.17
5	30.42 $\pm$ 0.69	14.45 $\pm$ 0.43	55.58 $\pm$ 1.36	13.25 $\pm$1.14	5.26 $\pm$ 0.15	5.74$\pm$ 0.22
10	38.73 $\pm$ 0.65	25.02 $\pm$ 0.65	68.37 $\pm$ 2.01	21.26 $\pm$1.33	5.69 $\pm$ 0.10	7.50$\pm$ 0.10
20	48.23 $\pm$ 0.33	33.31 $\pm$ 0.46	73.95 $\pm$ 0.53	30.06 $\pm$1.09	6.10 $\pm$ 0.22	9.14$\pm$ 0.35
50	56.97 $\pm$ 0.60	43.23 $\pm$ 0.68	76.65 $\pm$ 0.78	41.93 $\pm$1.02	6.44 $\pm$ 0.16	11.47$\pm$ 0.27

Table 6: Cross-domain performance results on 6 scenarios COCO $\longrightarrow$ DIOR / DOTA / DeepFruits / SIXRay / ClipArt / VisDrone. The average $\text{mAP}_{0.5}$ is reported with a 95% confidence interval.

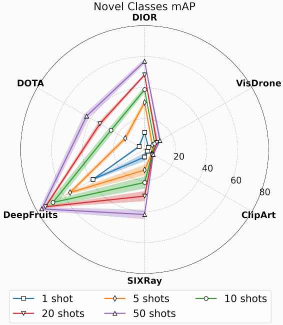


Figure 13: Performance of FSDiffusionDet on multiple COCO $\longrightarrow$ X scenarios.

- As in classical Few-Shot setting, performance improves with K. Promising performance with a reasonable amount of annotations.
- With $K = 10$, reduced performance for DOTA and DIOR, more difficult task (more classes)

COCO $\longrightarrow$ Anything scenarios

Base training on COCO and fine-tuning on another dataset, experiment with **DOTA** [Xia et al., 2018], **DIOR** [K. Li et al., 2020], **DeepFruit**, **SIXRay** [Miao et al., 2019], **CipArt** [Inoue et al., 2018], **VisDrone** [Y. Sun et al., 2022].

K Shots	DIOR	DOTA	DeepFruits	SIXRay	ClipArt	VisDrone
1	11.10 ± 0.32	4.03 ± 0.26	38.47 ± 1.42	4.80 ± 0.87	2.09± 0.19	2.83± 0.17
5	30.42 ± 0.69	14.45 ± 0.43	55.58 ± 1.36	13.25 ±1.14	5.26 ± 0.15	5.74± 0.22
10	38.73 ± 0.65	25.02 ± 0.65	68.37 ± 2.01	21.26 ±1.33	5.69 ± 0.10	7.50± 0.10
20	48.23 ± 0.33	33.31 ± 0.46	73.95 ± 0.53	30.06 ±1.09	6.10 ± 0.22	9.14± 0.35
50	56.97 ± 0.60	43.23 ± 0.68	76.65 ± 0.78	41.93 ±1.02	6.44 ± 0.16	11.47± 0.27

Table 6: Cross-domain performance results on 6 scenarios COCO $\longrightarrow$ DIOR / DOTA / DeepFruits / SIXRay / ClipArt / VisDrone. The average $\text{mAP}_{0.5}$ is reported with a 95% confidence interval.

- As in classical Few-Shot setting, performance improves with K. Promising performance with a reasonable amount of annotations.
- With $K = 10$, reduced performance for DOTA and DIOR, more difficult task (more classes)
- Difficulties with some datasets, probably because of poor annotation quality.

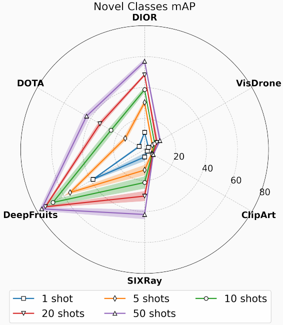


Figure 13: Performance of FSDiffusionDet on multiple COCO $\longrightarrow$ X scenarios.

Aerial Cross-Domain scenarios

Base training on DOTA and fine-tune on DIOR, and vice-versa.

DIOR $\longrightarrow$ DOTA								
	Backbone frozen				Fully fine-tuned			
***K* shots**	**All**	**S**	**M**	**L**	**All**	**S**	**M**	**L**
1	**5.41**	2.72	6.28	4.51	5.09	3.08	6.72	4.07
5	**25.88**	16.99	31.47	22.50	24.90	15.85	29.67	22.27
10	31.99	17.64	36.90	31.23	**33.30**	15.97	37.13	32.45
20	38.77	21.68	46.49	34.79	**41.30**	21.97	45.90	41.08
50	44.07	29.22	52.66	41.00	**49.22**	29.41	55.94	52.82

Table 7: FSDiffusionDet Cross-domain results on the scenario DIOR $\longrightarrow$ DOTA.

DOTA $\longrightarrow$ DIOR								
	Backbone frozen				Fully fine-tuned			
***K* shots**	**All**	**S**	**M**	**L**	**All**	**S**	**M**	**L**
1	**20.18**	5.53	16.96	23.43	9.40	3.86	9.15	8.95
5	**34.43**	9.99	31.12	47.03	29.57	8.70	25.80	35.76
10	**41.48**	12.85	36.62	53.85	38.44	10.50	32.58	47.27
20	**49.00**	16.39	40.23	62.79	45.36	15.29	36.51	55.05
50	**54.07**	18.70	43.83	67.58	53.51	19.49	41.27	63.04

Table 8: FSDiffusionDet Cross-domain results on the scenario DOTA $\longrightarrow$ DIOR.

Aerial Cross-Domain scenarios

Base training on DOTA and fine-tune on DIOR, and vice-versa.

	DIOR $\longrightarrow$ DOTA							
	Backbone frozen				Fully fine-tuned			
K shots	All	S	M	L	All	S	M	L
1	**5.41**	2.72	6.28	4.51	5.09	3.08	6.72	4.07
5	**25.88**	16.99	31.47	22.50	24.90	15.85	29.67	22.27
10	31.99	17.64	36.90	31.23	**33.30**	15.97	37.13	32.45
20	38.77	21.68	46.49	34.79	**41.30**	21.97	45.90	41.08
50	44.07	29.22	52.66	41.00	**49.22**	29.41	55.94	52.82

Table 7: FSDiffusionDet Cross-domain results on the scenario DIOR $\longrightarrow$ DOTA.

	DOTA $\longrightarrow$ DIOR							
	Backbone frozen				Fully fine-tuned			
K shots	All	S	M	L	All	S	M	L
1	**20.18**	5.53	16.96	23.43	9.40	3.86	9.15	8.95
5	**34.43**	9.99	31.12	47.03	29.57	8.70	25.80	35.76
10	**41.48**	12.85	36.62	53.85	38.44	10.50	32.58	47.27
20	**49.00**	16.39	40.23	62.79	45.36	15.29	36.51	55.05
50	**54.07**	18.70	43.83	67.58	53.51	19.49	41.27	63.04

Table 8: FSDiffusionDet Cross-domain results on the scenario DOTA $\longrightarrow$ DIOR.

- Higher performance than with COCO as source, **promising for COSE's applications.**

Aerial Cross-Domain scenarios

Base training on DOTA and fine-tune on DIOR, and vice-versa.

DIOR → DOTA								
	Backbone frozen				Fully fine-tuned			
K shots	All	S	M	L	All	S	M	L
1	**5.41**	2.72	6.28	4.51	5.09	3.08	6.72	4.07
5	**25.88**	16.99	31.47	22.50	24.90	15.85	29.67	22.27
10	31.99	17.64	36.90	31.23	**33.30**	15.97	37.13	32.45
20	38.77	21.68	46.49	34.79	**41.30**	21.97	45.90	41.08
50	44.07	29.22	52.66	41.00	**49.22**	29.41	55.94	52.82

Table 7: FSDiffusionDet Cross-domain results on the scenario DIOR → DOTA.

DOTA → DIOR								
	Backbone frozen				Fully fine-tuned			
K shots	All	S	M	L	All	S	M	L
1	**20.18**	5.53	16.96	23.43	9.40	3.86	9.15	8.95
5	**34.43**	9.99	31.12	47.03	29.57	8.70	25.80	35.76
10	**41.48**	12.85	36.62	53.85	38.44	10.50	32.58	47.27
20	**49.00**	16.39	40.23	62.79	45.36	15.29	36.51	55.05
50	**54.07**	18.70	43.83	67.58	53.51	19.49	41.27	63.04

Table 8: FSDiffusionDet Cross-domain results on the scenario DOTA → DIOR.

- Higher performance than with COCO as source, **promising for COSE's applications.**
- Different freezing sweet spot for DOTA → DIOR and DIOR → DOTA.

Aerial Cross-Domain scenarios

Base training on DOTA and fine-tune on DIOR, and vice-versa.

	DIOR $\longrightarrow$ DOTA							
	Backbone frozen				Fully fine-tuned			
K shots	All	S	M	L	All	S	M	L
1	**5.41**	2.72	6.28	4.51	5.09	3.08	6.72	4.07
5	**25.88**	16.99	31.47	22.50	24.90	15.85	29.67	22.27
10	31.99	17.64	36.90	31.23	**33.30**	15.97	37.13	32.45
20	38.77	21.68	46.49	34.79	**41.30**	21.97	45.90	41.08
50	44.07	29.22	52.66	41.00	**49.22**	29.41	55.94	52.82

Table 7: FSDiffusionDet Cross-domain results on the scenario DIOR $\longrightarrow$ DOTA.

	DOTA $\longrightarrow$ DIOR							
	Backbone frozen				Fully fine-tuned			
K shots	All	S	M	L	All	S	M	L
1	**20.18**	5.53	16.96	23.43	9.40	3.86	9.15	8.95
5	**34.43**	9.99	31.12	47.03	29.57	8.70	25.80	35.76
10	**41.48**	12.85	36.62	53.85	38.44	10.50	32.58	47.27
20	**49.00**	16.39	40.23	62.79	45.36	15.29	36.51	55.05
50	**54.07**	18.70	43.83	67.58	53.51	19.49	41.27	63.04

Table 8: FSDiffusionDet Cross-domain results on the scenario DOTA $\longrightarrow$ DIOR.

- Higher performance than with COCO as source, **promising for COSE's applications.**
- Different freezing sweet spot for DOTA $\longrightarrow$ DIOR and DIOR $\longrightarrow$ DOTA.
- Need for fine-tuning sweet spot estimation tools, *e.g.* a dataset/domain distance measure:
  - Intuition: compatible domains require less plasticity and fine-tuning.
  - Take domain shift into account.
  - Relationship between base/source and novel/target classes (intra and inter-class variance).
  - Work in progress...

5

Conclusion and Perspectives

Summary of this presentation

Summary of this presentation

- **Investigation of FSOD challenge with aerial Images:**
  - Small objects are more numerous in aerial images, poor examples in few-shot.

Summary of this presentation

- **Investigation of FSOD challenge with aerial Images:**
  - Small objects are more numerous in aerial images, poor examples in few-shot.
- **Improvements of small object detection with Attention-based and Fine-tuning approaches:**
  - XQSA significantly improves small object detection.
  - FSDiffusionDet outperforms all methods on aerial images and promising results in the Cross-Domain scenarios.

5. Conclusion and Perspectives

Summary of this presentation

- **Investigation of FSOD challenge with aerial Images:**
  - Small objects are more numerous in aerial images, poor examples in few-shot.
- **Improvements of small object detection with Attention-based and Fine-tuning approaches:**
  - XQSA significantly improves small object detection.
  - FSDiffusionDet outperforms all methods on aerial images and promising results in the Cross-Domain scenarios.

	XQSA		**FSDiffusionDet**
👍	Adaptability to new classes	👍	Simple and quicker training
👍	Very low-shot performance	👍	Shot scalability
👎	Complex training scheme	👎	Fine-tuning mandatory
👎	Inference speed	👍	Cross-domain performance

Summary of this presentation

- **Investigation of FSOD challenge with aerial Images:**
  - Small objects are more numerous in aerial images, poor examples in few-shot.

- **Improvements of small object detection with Attention-based and Fine-tuning approaches:**
  - XQSA significantly improves small object detection.
  - FSDiffusionDet outperforms all methods on aerial images and promising results in the Cross-Domain scenarios.

	XQSA		FSDiffusionDet
👍	Adaptability to new classes	👍	Simple and quicker training
👍	Very low-shot performance	👍	Shot scalability
👎	Complex training scheme	👎	Fine-tuning mandatory
👎	Inference speed	👍	Cross-domain performance

- **Carefully designed loss function can improve small object detection:**
  - SIoU loss allows precise control of the training balance between small and large objects.
  - SIoU increase detection performance and helps for model evaluation.

5. Conclusion and Perspectives

Research perpectives

Research perpectives

- **Attention-based approaches**

Research perpectives

- **Attention-based approaches**
  - Design more versatile attention mechanisms for small and large objects.
  - Reduce memory footprint of attention modules to speed up training and improve scalability.
  - Develop fine-tuning-free approaches (possible but poor results so far).

Research perpectives

- **Attention-based approaches**
  - Design more versatile attention mechanisms for small and large objects.
  - Reduce memory footprint of attention modules to speed up training and improve scalability.
  - Develop fine-tuning-free approaches (possible but poor results so far).

- **Fine-tuning approaches**

Research perpectives

- **Attention-based approaches**
  - Design more versatile attention mechanisms for small and large objects.
  - Reduce memory footprint of attention modules to speed up training and improve scalability.
  - Develop fine-tuning-free approaches (possible but poor results so far).

- **Fine-tuning approaches**
  - Develop predictive tools to find optimal plasticity (including theoretical groundings):
    - Find a measure of difficulty for a given scenario, ideally few-shot compatible.
    - Understand the need for plasticity and how it should be distributed within the models.
  - Find a way to adapt the models without fine-tuning:
    - Propose transductive approaches for object detection.
    - Require solving regression in a transductive manner (*e.g.* Transductive Ridge Regression [Cortes et al., 2006]).

Research perpectives

- **Attention-based approaches**
  - Design more versatile attention mechanisms for small and large objects.
  - Reduce memory footprint of attention modules to speed up training and improve scalability.
  - Develop fine-tuning-free approaches (possible but poor results so far).

- **Fine-tuning approaches**
  - Develop predictive tools to find optimal plasticity (including theoretical groundings):
    - Find a measure of difficulty for a given scenario, ideally few-shot compatible.
    - Understand the need for plasticity and how it should be distributed within the models.

  - Find a way to adapt the models without fine-tuning:
    - Propose transductive approaches for object detection.
    - Require solving regression in a transductive manner (*e.g.* Transductive Ridge Regression [Cortes et al., 2006]).

- **Scaled-adaptive Intersection over Union**

Research perpectives

- **Attention-based approaches**
  - Design more versatile attention mechanisms for small and large objects.
  - Reduce memory footprint of attention modules to speed up training and improve scalability.
  - Develop fine-tuning-free approaches (possible but poor results so far).

- **Fine-tuning approaches**
  - Develop predictive tools to find optimal plasticity (including theoretical groundings):
    - Find a measure of difficulty for a given scenario, ideally few-shot compatible.
    - Understand the need for plasticity and how it should be distributed within the models.
  - Find a way to adapt the models without fine-tuning:
    - Propose transductive approaches for object detection.
    - Require solving regression in a transductive manner (*e.g.* Transductive Ridge Regression [Cortes et al., 2006]).

- **Scaled-adaptive Intersection over Union**
  - Understand in depth the training balance between small and large objects.
  - Investigate its influence on other components of the detection frameworks (*e.g.* NMS and example selection).

Research perpectives

- **Attention-based approaches**
  - Design more versatile attention mechanisms for small and large objects.
  - Reduce memory footprint of attention modules to speed up training and improve scalability.
  - **Develop fine-tuning-free approaches (possible but poor results so far).**
- **Fine-tuning approaches**
  - Develop predictive tools to find optimal plasticity (including theoretical groundings):
    - Find a measure of difficulty for a given scenario, ideally few-shot compatible.
    - Understand the need for plasticity and how it should be distributed within the models.
  - **Find a way to adapt the models without fine-tuning:**
    - Propose transductive approaches for object detection.
    - Require solving regression in a transductive manner (*e.g.* Transductive Ridge Regression [Cortes et al., 2006]).
- **Scaled-adaptive Intersection over Union**
  - Understand in depth the training balance between small and large objects.
  - Investigate its influence on other components of the detection frameworks (*e.g.* NMS and example selection).

Industrial perspectives

Industrial perspectives

- **Focus on CD-FSOD as closest from COSE Applications.**

Industrial perspectives

- **Focus on CD-FSOD as closest from COSE Applications.**
- **Optimization and deployment of Few-Shot detection models:**
  - ► Few-Shot Detection originally planned as a *ground processing*, *i.e.* with much looser constraints.
  - ► Few-Shot models are much more challenging to optimize/quantize, existing tools do not support custom layers.
  - ► Quantization and model size reduction reduce performance, not suitable for Few-Shot models.

Industrial perspectives

- **Focus on CD-FSOD as closest from COSE Applications.**
- **Optimization and deployment of Few-Shot detection models:**
    - Few-Shot Detection originally planned as a *ground processing*, *i.e.* with much looser constraints.
    - Few-Shot models are much more challenging to optimize/quantize, existing tools do not support custom layers.
    - Quantization and model size reduction reduce performance, not suitable for Few-Shot models.
- **Automate fine-tuning and deployment pipeline for regular and Few-Shot Detection models.**

Industrial perspectives

- **Focus on CD-FSOD as closest from COSE Applications.**

- **Optimization and deployment of Few-Shot detection models:**
  - ► Few-Shot Detection originally planned as a *ground processing*, *i.e.* with much looser constraints.
  - ► Few-Shot models are much more challenging to optimize/quantize, existing tools do not support custom layers.
  - ► Quantization and model size reduction reduce performance, not suitable for Few-Shot models.

- **Automate fine-tuning and deployment pipeline for regular and Few-Shot Detection models.**

First flights in the following months: plenty of opportunity for further testing and refinements!

Thank you for your attention

Special thanks to:

- *The members of the jury*
- *COSE and all my colleagues*
- *Everyone who supported me during the past 3 years*

Any Questions ?

Appendix Table of Content

A. Influence of Object Size on Few-Shot Performance

**Detection of small objects is much more difficult in the few-shot regime.
Performance increases with object size.**

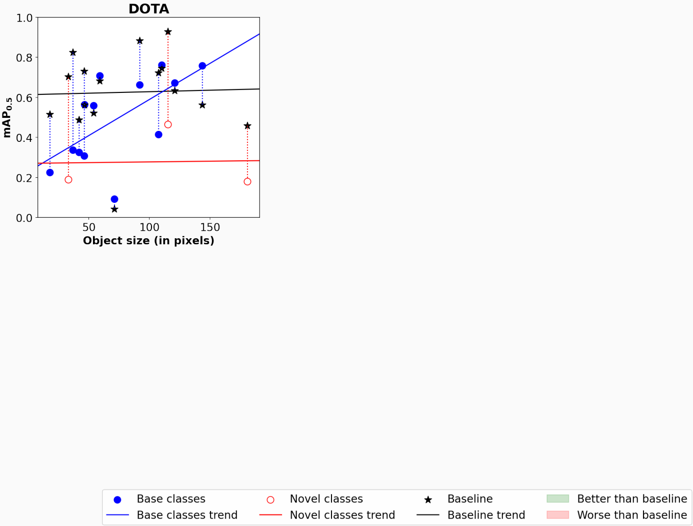


Figure 14: Per-class detection performance against object size, split by dataset.

A. Influence of Object Size on Few-Shot Performance

Detection of small objects is much more difficult in the few-shot regime.
Performance increases with object size.

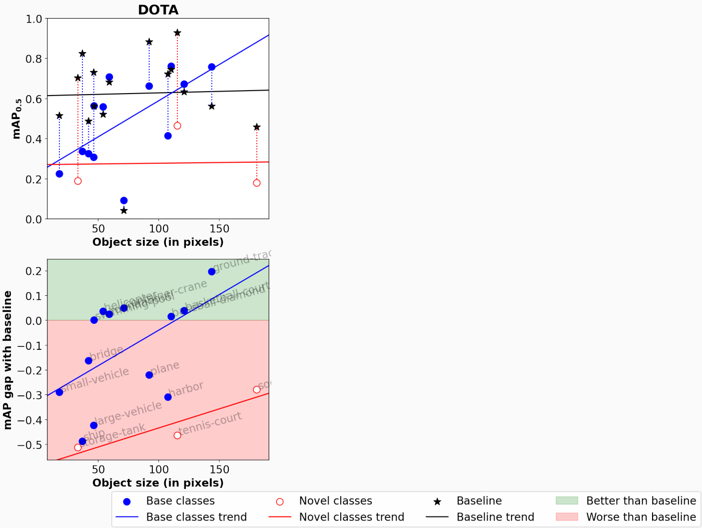


Figure 14: Per-class detection performance against object size, split by dataset.

A. Influence of Object Size on Few-Shot Performance

Detection of small objects is much more difficult in the few-shot regime. Performance increases with object size.

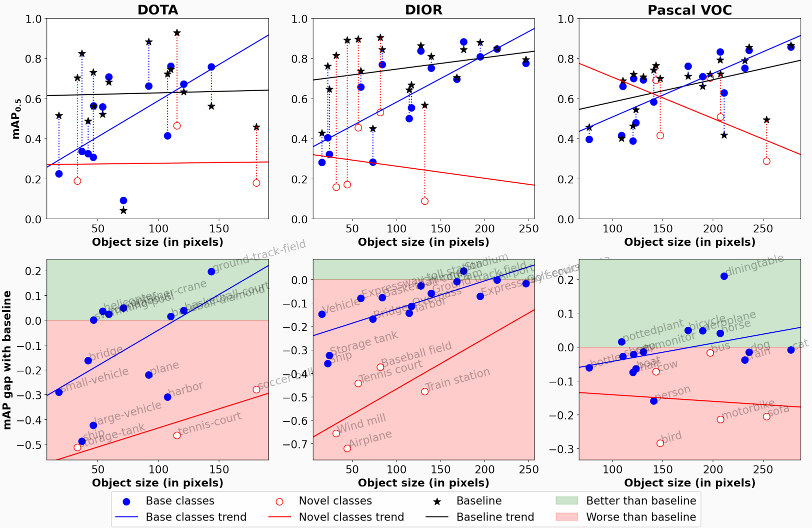


Figure 14: Per-class detection performance against object size, split by dataset.

A. Influence of Object Size on Few-Shot Performance

Detection of small objects is much more difficult in the few-shot regime.
Performance increases with object size.

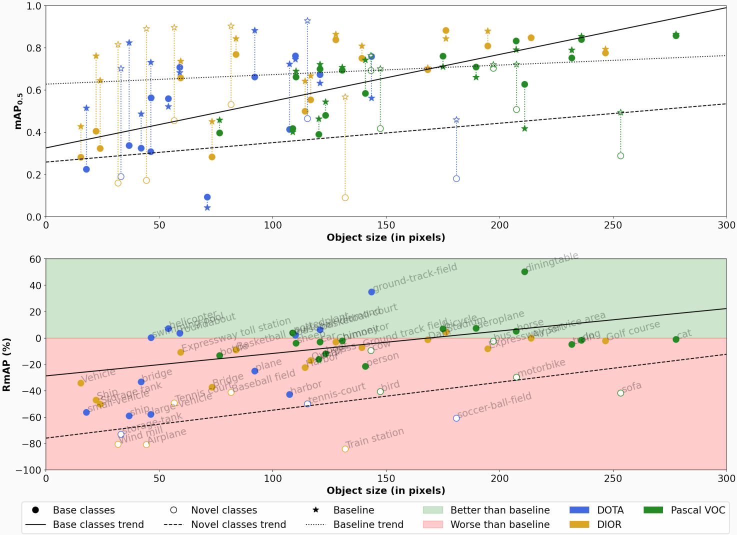


Figure 14: Per-class detection performance against object size.

User study conducted on 74 participants, more than 3000 individual answers.

Factor Analysis on human ratings:

- Only object size has a significant influence on human ratings.
- SIoU compensates the rating shift with object size.

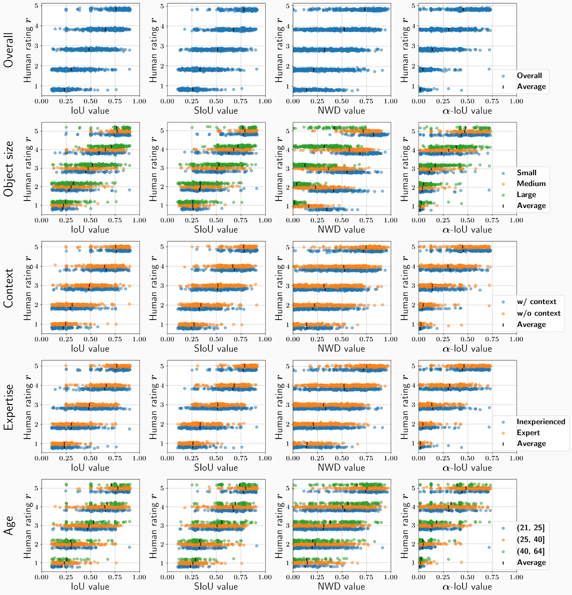


Figure 15: Rating against IoU, SIoU ($\gamma = 0.2$, $\kappa = 64$), NWD and α-IoU ($\alpha = 3$) values, overall and for different groupings of the variables of interest (object size, presence of contextual

Criteria empirical distributions split per human rating and object size.

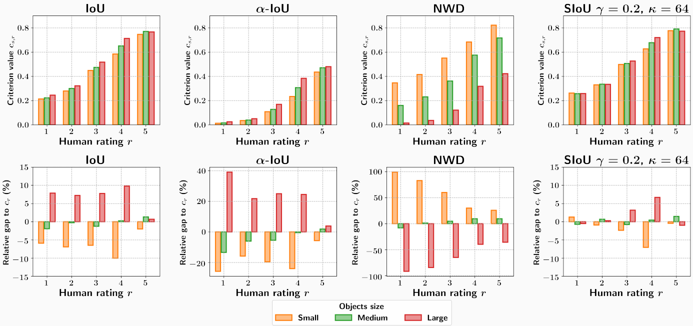
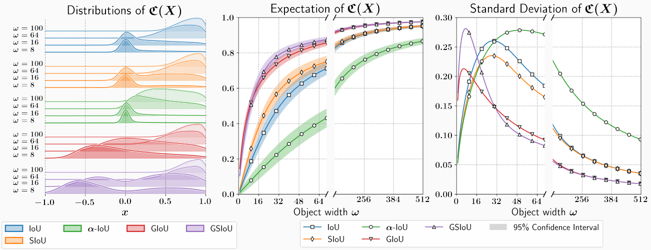


Figure 16: Criteria' scores for different object sizes and human ratings $r \in \{1, 2, 3, 4, 5\}$

$$c_{s,r} = \frac{\mathfrak{C}_{s,r} - \sum_{s} \mathfrak{C}_{s,r}}{\sum_{s} \mathfrak{C}_{s,r}}, \tag{9}$$

with $\mathfrak{C}_{s,r}$ is the average criterion value ($\mathfrak{C} \in \{\text{IoU}, \text{SIoU}, \alpha\text{-IoU}, \text{NWD}\}$) for an object size s and a rating r.

C. Criteria Distributions Analysis



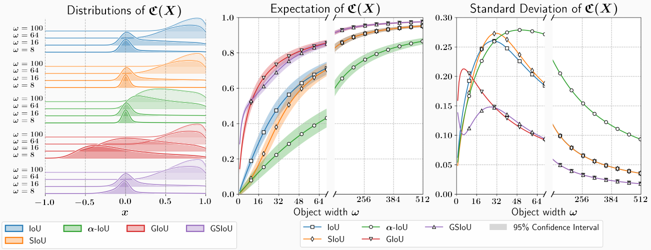


Figure 17: Criteria's distribution comparison **(top)** $\gamma = 0.5$ and $\kappa = 16$ for SIoU and GSIoU, **(bottom)** $\gamma = -4$ and $\kappa = 16$.

D. Additional Results for SIoU

Influence of γ and κ on the performance.

	Base classes				Novel Classes			
γ	**All**	**S**	**M**	**L**	**All**	**S**	**M**	**L**
0.5	47.09	21.29	54.67	65.48	30.50	8.83	44.97	65.89
0.25	45.94	21.60	54.39	63.40	30.96	12.53	42.37	64.14
0	52.41	26.94	61.17	63.00	41.03	24.01	52.13	69.78
-0.5	52.80	**27.16**	61.19	64.61	41.06	25.20	50.18	**72.04**
-1	53.03	23.20	61.53	66.68	42.77	27.55	**52.01**	70.76
-2	**54.06**	23.68	**62.69**	66.62	43.67	30.04	51.69	69.66
-3	52.91	22.14	61.19	66.02	**45.88**	**34.83**	51.26	70.78
-4	53.59	22.50	62.48	66.18	42.43	27.56	51.79	68.70
-9	53.11	20.98	62.13	**67.00**	42.63	30.53	48.89	68.62

Table 9: Evolution of the few-shot performance (XQSA with GSIoU loss) on DOTA for various values of γ ($\kappa = 16$ is fixed). mAP is reported with a 0.5 IoU threshold and for all object sizes.

	Base classes				Novel Classes			
κ	**All**	**S**	**M**	**L**	**All**	**S**	**M**	**L**
4	51.65	21.50	59.76	65.85	42.98	30.33	48.57	**73.41**
8	52.70	21.96	61.49	66.43	**44.16**	**31.35**	50.70	71.99
16	**54.06**	**23.68**	62.69	66.62	43.67	30.04	51.69	69.66
32	53.88	22.33	**63.00**	**67.35**	37.36	23.65	44.60	66.29
64	52.82	21.79	61.46	66.77	43.68	29.43	**52.47**	69.46
128	53.42	21.73	62.90	66.75	41.32	26.85	49.40	70.38

Table 10: Evolution of the few-shot performance (XQSA with GSIoU loss) for various values of κ ($\gamma = -2$ is fixed).

D. Additional Results for SIoU

SIoU is beneficial with multiple attention-based FSOD methods.

	XQSA	Base classes All	S	M	L	Novel Classes All	S	M	L
FRW	w/ GIoU	**34.60**	**16.15**	**48.61**	**59.00**	32.00	15.29	**44.50**	54.77
	w/ GSIoU	30.36	11.94	44.30	54.87	**32.94**	**16.69**	42.87	**62.64**
DANA	w/ GIoU	48.09	27.34	**66.06**	**68.00**	**44.49**	**30.10**	52.24	74.40
	w/ GSIoU	**50.10**	**32.19**	65.46	67.77	41.40	21.07	**54.80**	**75.23**
XQSA	w/ GIoU	**45.30**	**26.94**	61.17	63.00	41.03	24.01	**52.13**	69.78
	w/ GSIoU	43.42	22.14	**61.19**	**66.02**	**45.88**	**34.83**	51.26	**70.78**

Table 11: Performance comparison with three different FSOD methods: Feature Reweighting Kang et al. 2019 (FRW), Dual Awareness Attention T.-I. Chen et al. 2021 (DANA) and our Cross-scale Query-Support Alignment (XQSA), trained with GIoU and GSIoU. mAP is reported with a 0.5 **IoU threshold** for small (S), medium (M), large (L) and all objects.

SIoU's influence on Regular Detection.

FCOS	DOTA All	S	M	L	DIOR All	S	M	L
w/ GIoU	34.9	17.4	36.6	43.3	48.1	10.1	40.3	63.2
w/ GSIoU	**36.8**	**17.5**	**40.4**	**45.2**	**49.2**	**11.0**	**41.2**	**66.1**

Table 12: Regular Object Detection performance on DOTA and DIOR datasets with GIoU and GSIoU ($\gamma = -3$ and $\kappa = 16$) losses. mAP is computed with several IoU thresholds (0.5 to 0.95) as it is commonly done in regular detection.

D. Additional Results for SIoU

Evaluation with SIoU as the evaluation threshold.

Loss	Base classes All	S	M	L	Novel Classes All	S	M	L
IoU	55.81	35.03	62.57	70.05	39.10	18.58	53.93	68.83
α-IoU	53.05	20.60	61.05	**72.41**	41.93	20.99	55.74	76.79
SIoU	**59.77**	36.38	67.29	70.06	49.51	31.06	62.53	**77.24**
NWD	58.80	34.16	66.81	70.05	53.66	42.02	62.53	68.92
GIoU	59.27	**44.07**	**66.91**	65.46	49.02	35.10	57.58	74.30
GSIoU	59.32	35.32	66.29	69.03	**57.70**	**46.77**	**65.56**	73.67

Table 13: Few-shot performance comparison between several criteria: IoU, α-IoU, SIoU, NWD, GIoU and GSIoU trained on DOTA. mAP is reported with a 0.5 **SIoU threshold** for small (S), medium (M), large (L), and all objects.

	XQSA	Base classes All	S	M	L	Novel Classes All	S	M	L
DOTA	w/ GIoU	59.27	**44.07**	66.91	65.46	49.02	35.10	57.58	**74.30**
	w/ GSIoU	**59.32**	35.32	**66.29**	**69.03**	**57.70**	**46.77**	**65.56**	73.67
DIOR	w/ GIoU	62.06	17.49	45.55	82.22	53.81	23.79	53.46	71.63
	w/ GSIoU	**63.81**	**17.77**	**49.62**	**82.53**	**58.79**	**25.60**	**59.28**	**73.78**
Pascal	w/ GIoU	55.51	26.10	**46.82**	64.31	52.43	28.97	40.73	62.58
	w/ GSIoU	**58.74**	**27.47**	46.56	**68.93**	**58.92**	**31.36**	**41.65**	**69.71**
COCO	w/ GIoU	21.46	12.77	24.79	31.86	29.21	17.36	27.62	40.05
	w/ GSIoU	**21.97**	**12.80**	**25.72**	**32.35**	**29.94**	**18.87**	**29.93**	**40.47**

Table 14: Few-shot performance on three datasets: DOTA, DIOR, Pascal VOC and COCO. GIoU and GSIoU losses are compared. mAP is reported with a 0.5 **SIoU threshold** and for various object sizes.

Finding the fine-tuning sweetspot

Freezing point	Plasticity rate	DOTA	DIOR	Pascal VOC	MS COCO
FT whole	100.00 %	60.09	52.17	43.10	17.15
Up to stage 1	99.98 %	58.85	53.37	43.81	17.72
Up to stage 2	99.47 %	57.41	53.21	41.23	17.73
Up to stage 3	96.57 %	59.88	54.36	47.57	19.49
Up to stage 4	79.66 %	56.13	**57.51**	**53.72**	**21.88**
FT head only	35.97 %	51.82	55.70	51.72	19.96
FT last layer only	0.03 %	0.05	0.11	0.53	0.01
Bias only	35.98 %	**60.45**	55.12	49.90	20.19
BatchNorm only	35.97 %	59.35	55.63	51.96	19.70

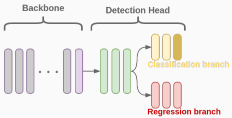
Table 15: Influence of the amount of plasticity on the FS performance on DOTA, DIOR, Pascal VOC and MS COCO. mAP is reported with a 0.5 IoU threshold, $K = 10$ shots.

E. Additional Results for Few-Shot Diffusion Detector

Finding the fine-tuning sweetspot

Freezing point	Plasticity rate	DOTA	DIOR	Pascal VOC	MS COCO
FT whole	100.00 %	60.09	52.17	43.10	17.15
Up to stage 1	99.98 %	58.85	53.37	43.81	17.72
Up to stage 2	99.47 %	57.41	53.21	41.23	17.73
Up to stage 3	96.57 %	59.88	54.36	47.57	19.49
Up to stage 4	79.66 %	56.13	**57.51**	**53.72**	**21.88**
FT head only	35.97 %	51.82	55.70	51.72	19.96
FT last layer only	0.03 %	0.05	0.11	0.53	0.01
Bias only	35.98 %	**60.45**	55.12	49.90	20.19
BatchNorm only	35.97 %	59.35	55.63	51.96	19.70

Table 15: Influence of the amount of plasticity on the FS performance on DOTA, DIOR, Pascal VOC and MS COCO. mAP is reported with a 0.5 IoU threshold, $K = 10$ shots.



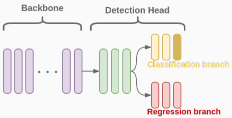
- Freezing the 4 first stages of the backbone is a sensible compromise.

E. Additional Results for Few-Shot Diffusion Detector

Finding the fine-tuning sweetspot

Freezing point	Plasticity rate	DOTA	DIOR	Pascal VOC	MS COCO
FT whole	100.00 %	60.09	52.17	43.10	17.15
Up to stage 1	99.98 %	58.85	53.37	43.81	17.72
Up to stage 2	99.47 %	57.41	53.21	41.23	17.73
Up to stage 3	96.57 %	59.88	54.36	47.57	19.49
Up to stage 4	79.66 %	56.13	**57.51**	**53.72**	**21.88**
FT head only	35.97 %	51.82	55.70	51.72	19.96
FT last layer only	0.03 %	0.05	0.11	0.53	0.01
Bias only	35.98 %	**60.45**	55.12	49.90	20.19
BatchNorm only	35.97 %	59.35	55.63	51.96	19.70

Table 15: Influence of the amount of plasticity on the FS performance on DOTA, DIOR, Pascal VOC and MS COCO. mAP is reported with a 0.5 IoU threshold, $K = 10$ shots.



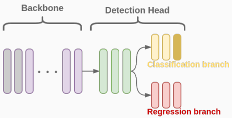
- Freezing the 4 first stages of the backbone is a sensible compromise.
- For DIOR, Pascal VOC and COCO, performance increases as plasticity decreases, up to stage 4.

E. Additional Results for Few-Shot Diffusion Detector

Finding the fine-tuning sweetspot

Freezing point	Plasticity rate	DOTA	DIOR	Pascal VOC	MS COCO
FT whole	100.00 %	60.09	52.17	43.10	17.15
Up to stage 1	99.98 %	58.85	53.37	43.81	17.72
Up to stage 2	99.47 %	57.41	53.21	41.23	17.73
Up to stage 3	96.57 %	59.88	54.36	47.57	19.49
Up to stage 4	79.66 %	56.13	**57.51**	**53.72**	**21.88**
FT head only	35.97 %	51.82	55.70	51.72	19.96
FT last layer only	0.03 %	0.05	0.11	0.53	0.01
Bias only	35.98 %	**60.45**	55.12	49.90	20.19
BatchNorm only	35.97 %	59.35	55.63	51.96	19.70

Table 15: Influence of the amount of plasticity on the FS performance on DOTA, DIOR, Pascal VOC and MS COCO. mAP is reported with a 0.5 IoU threshold, $K = 10$ shots.



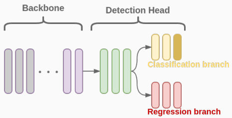
- Freezing the 4 first stages of the backbone is a sensible compromise.
- For DIOR, Pascal VOC and COCO, performance increases as plasticity decreases, up to stage 4.

E. Additional Results for Few-Shot Diffusion Detector

Finding the fine-tuning sweetspot

Freezing point	Plasticity rate	DOTA	DIOR	Pascal VOC	MS COCO
FT whole	100.00 %	60.09	52.17	43.10	17.15
Up to stage 1	99.98 %	58.85	53.37	43.81	17.72
Up to stage 2	99.47 %	57.41	53.21	41.23	17.73
Up to stage 3	96.57 %	59.88	54.36	47.57	19.49
Up to stage 4	79.66 %	56.13	**57.51**	**53.72**	**21.88**
FT head only	35.97 %	51.82	55.70	51.72	19.96
FT last layer only	0.03 %	0.05	0.11	0.53	0.01
Bias only	35.98 %	**60.45**	55.12	49.90	20.19
BatchNorm only	35.97 %	59.35	55.63	51.96	19.70

Table 15: Influence of the amount of plasticity on the FS performance on DOTA, DIOR, Pascal VOC and MS COCO. mAP is reported with a 0.5 IoU threshold, $K = 10$ shots.



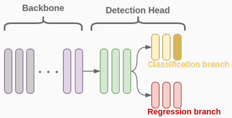
- Freezing the 4 first stages of the backbone is a sensible compromise.
- For DIOR, Pascal VOC and COCO, performance increases as plasticity decreases, up to stage 4.

E. Additional Results for Few-Shot Diffusion Detector

Finding the fine-tuning sweetspot

Freezing point	Plasticity rate	DOTA	DIOR	Pascal VOC	MS COCO
FT whole	100.00 %	60.09	52.17	43.10	17.15
Up to stage 1	99.98 %	58.85	53.37	43.81	17.72
Up to stage 2	99.47 %	57.41	53.21	41.23	17.73
Up to stage 3	96.57 %	59.88	54.36	47.57	19.49
Up to stage 4	79.66 %	56.13	**57.51**	**53.72**	**21.88**
FT head only	35.97 %	51.82	55.70	51.72	19.96
FT last layer only	0.03 %	0.05	0.11	0.53	0.01
Bias only	35.98 %	**60.45**	55.12	49.90	20.19
BatchNorm only	35.97 %	59.35	55.63	51.96	19.70

Table 15: Influence of the amount of plasticity on the FS performance on DOTA, DIOR, Pascal VOC and MS COCO. mAP is reported with a 0.5 IoU threshold, $K = 10$ shots.



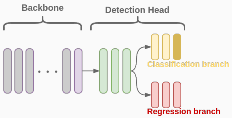
- Freezing the 4 first stages of the backbone is a sensible compromise.
- For DIOR, Pascal VOC and COCO, performance increases as plasticity decreases, up to stage 4.

E. Additional Results for Few-Shot Diffusion Detector

Finding the fine-tuning sweetspot

Freezing point	Plasticity rate	DOTA	DIOR	Pascal VOC	MS COCO
FT whole	100.00 %	60.09	52.17	43.10	17.15
Up to stage 1	99.98 %	58.85	53.37	43.81	17.72
Up to stage 2	99.47 %	57.41	53.21	41.23	17.73
Up to stage 3	96.57 %	59.88	54.36	47.57	19.49
Up to stage 4	79.66 %	56.13	**57.51**	**53.72**	**21.88**
FT head only	35.97 %	51.82	55.70	51.72	19.96
FT last layer only	0.03 %	0.05	0.11	0.53	0.01
Bias only	35.98 %	**60.45**	55.12	49.90	20.19
BatchNorm only	35.97 %	59.35	55.63	51.96	19.70

Table 15: Influence of the amount of plasticity on the FS performance on DOTA, DIOR, Pascal VOC and MS COCO. mAP is reported with a 0.5 IoU threshold, $K = 10$ shots.



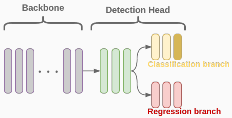
- Freezing the 4 first stages of the backbone is a sensible compromise.
- For DIOR, Pascal VOC and COCO, performance increases as plasticity decreases, up to stage 4.

E. Additional Results for Few-Shot Diffusion Detector

Finding the fine-tuning sweetspot

Freezing point	Plasticity rate	DOTA	DIOR	Pascal VOC	MS COCO
FT whole	100.00 %	60.09	52.17	43.10	17.15
Up to stage 1	99.98 %	58.85	53.37	43.81	17.72
Up to stage 2	99.47 %	57.41	53.21	41.23	17.73
Up to stage 3	96.57 %	59.88	54.36	47.57	19.49
Up to stage 4	79.66 %	56.13	**57.51**	**53.72**	**21.88**
FT head only	35.97 %	51.82	55.70	51.72	19.96
FT last layer only	0.03 %	0.05	0.11	0.53	0.01
Bias only	35.98 %	**60.45**	55.12	49.90	20.19
BatchNorm only	35.97 %	59.35	55.63	51.96	19.70

Table 15: Influence of the amount of plasticity on the FS performance on DOTA, DIOR, Pascal VOC and MS COCO. mAP is reported with a 0.5 IoU threshold, $K = 10$ shots.



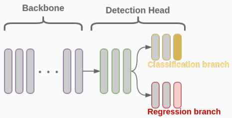
- Freezing the 4 first stages of the backbone is a sensible compromise.
- For DIOR, Pascal VOC and COCO, performance increases as plasticity decreases, up to stage 4.

E. Additional Results for Few-Shot Diffusion Detector

Finding the fine-tuning sweetspot

Freezing point	Plasticity rate	DOTA	DIOR	Pascal VOC	MS COCO
FT whole	100.00 %	60.09	52.17	43.10	17.15
Up to stage 1	99.98 %	58.85	53.37	43.81	17.72
Up to stage 2	99.47 %	57.41	53.21	41.23	17.73
Up to stage 3	96.57 %	59.88	54.36	47.57	19.49
Up to stage 4	79.66 %	56.13	**57.51**	**53.72**	**21.88**
FT head only	35.97 %	51.82	55.70	51.72	19.96
FT last layer only	0.03 %	0.05	0.11	0.53	0.01
Bias only	35.98 %	**60.45**	55.12	49.90	20.19
BatchNorm only	35.97 %	59.35	55.63	51.96	19.70

Table 15: Influence of the amount of plasticity on the FS performance on DOTA, DIOR, Pascal VOC and MS COCO. mAP is reported with a 0.5 IoU threshold, $K = 10$ shots.



- Freezing the 4 first stages of the backbone is a sensible compromise.
- For DIOR, Pascal VOC and COCO, performance increases as plasticity decreases, up to stage 4.

E. Additional Results for Few-Shot Diffusion Detector

Finding the fine-tuning sweetspot

Freezing point	Plasticity rate	DOTA	DIOR	Pascal VOC	MS COCO
FT whole	100.00 %	60.09	52.17	43.10	17.15
Up to stage 1	99.98 %	58.85	53.37	43.81	17.72
Up to stage 2	99.47 %	57.41	53.21	41.23	17.73
Up to stage 3	96.57 %	59.88	54.36	47.57	19.49
Up to stage 4	79.66 %	56.13	**57.51**	**53.72**	**21.88**
FT head only	35.97 %	51.82	55.70	51.72	19.96
FT last layer only	0.03 %	0.05	0.11	0.53	0.01
Bias only	35.98 %	**60.45**	55.12	49.90	20.19
BatchNorm only	35.97 %	59.35	55.63	51.96	19.70

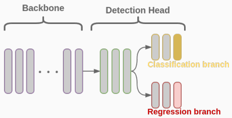


Table 15: Influence of the amount of plasticity on the FS performance on DOTA, DIOR, Pascal VOC and MS COCO. mAP is reported with a 0.5 IoU threshold, $K = 10$ shots.

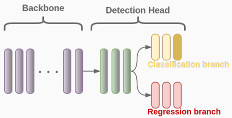
- Freezing the 4 first stages of the backbone is a sensible compromise.
- For DIOR, Pascal VOC and COCO, performance increases as plasticity decreases, up to stage 4.
- For DOTA, the trend is reversed.

E. Additional Results for Few-Shot Diffusion Detector

Finding the fine-tuning sweetspot

Freezing point	Plasticity rate	DOTA	DIOR	Pascal VOC	MS COCO
FT whole	100.00 %	60.09	52.17	43.10	17.15
Up to stage 1	99.98 %	58.85	53.37	43.81	17.72
Up to stage 2	99.47 %	57.41	53.21	41.23	17.73
Up to stage 3	96.57 %	59.88	54.36	47.57	19.49
Up to stage 4	79.66 %	56.13	**57.51**	**53.72**	**21.88**
FT head only	35.97 %	51.82	55.70	51.72	19.96
FT last layer only	0.03 %	0.05	0.11	0.53	0.01
Bias only	35.98 %	**60.45**	55.12	49.90	20.19
BatchNorm only	35.97 %	59.35	55.63	51.96	19.70

Table 15: Influence of the amount of plasticity on the FS performance on DOTA, DIOR, Pascal VOC and MS COCO. mAP is reported with a 0.5 IoU threshold, $K = 10$ shots.



- Freezing the 4 first stages of the backbone is a sensible compromise.
- For DIOR, Pascal VOC and COCO, performance increases as plasticity decreases, up to stage 4.
- For DOTA, the trend is reversed.
- Fine-tuning some parameters in every layer produces impressive results with reduced plasticity.

E. Additional Results for Few-Shot Diffusion Detector

FSDiffusionDet scales much better than other techniques with the number of shots K.

However, lower performance in very low shot settings $K < 5$.

K	DOTA	DIOR	Pascal VOC	MS COCO
1	4.19	27.17	22.24	7.43
2	9.83	40.31	31.98	12.45
3	27.61	43.54	29.52	15.75
5	39.00	46.92	38.08	19.33
10	52.05	54.32	52.64	24.99
20	62.79	60.24	59.26	28.76
30	67.32	65.28	64.19	31.19
50	71.91	71.21	67.81	34.64
100	72.27	77.05	71.31	38.77

Table 16: Influence of the number of shots on the few-shot object detection performance of FSDiffusionDet on DOTA, DIOR, Pascal VOC and MS COCO. Performance is reported with $mAP_{0.5}$.

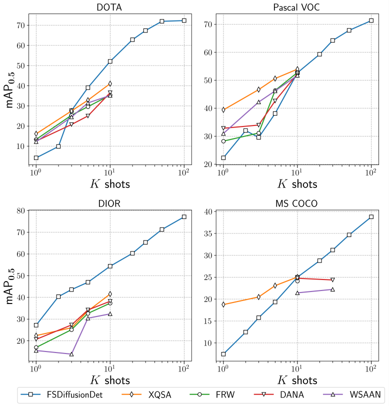


Figure 18: Performance of FSDiffusionDet, XQSA, FRW, DANA and WSAAN on DOTA, DIOR, Pascal VOC and MS COCO against the number of shots. Performance is reported with $mAP_{0.5}$.

E. Additional Results for Few-Shot Diffusion Detector

Influence of the pre-training of the backbone

Backbones	DOTA	DIOR	Pascal VOC	MS COCO
Scratch	7.28	8.72	13.72	0.38
ImageNet	**52.05**	54.32	52.64	**24.99**
DINO	46.84	**55.88**	**54.58**	23.94
CLIP	40.36	51.61	49.81	19.83

Table 17: Study of the influence of the backbone pre-training. $\text{mAP}_{0.5}$ is provided only for base classes, the blue and red colors to distinguish between base and novel classes are no longer required.

Influence of the number of proposals

# of Proposals	DOTA	DIOR	Pascal VOC	MS COCO
200	41.57	52.92	**52.86**	**23.24**
250	47.97	47.62	52.28	22.61
300	**55.76**	51.77	51.81	22.46
350	52.27	50.41	50.63	22.13
400	46.49	49.98	50.55	20.04
450	53.11	53.07	51.06	20.48
500	52.03	**55.31**	51.44	20.25

Table 18: Analysis of FSDiffusionDet performance ($\text{mAP}_{0.5}$) against the number of proposals on DOTA, DIOR, Pascal VOC and MS COCO datasets.

F. Prototypical Faster R-CNN

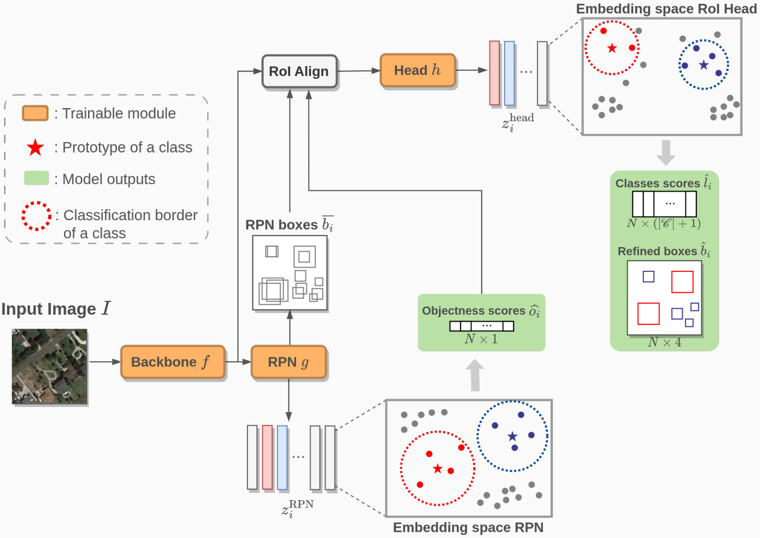


Figure 19: Illustration of the architecture of Prototypical Faster R-CNN.

G. Qualitative Resultswith Prototypical Faster R-CNN

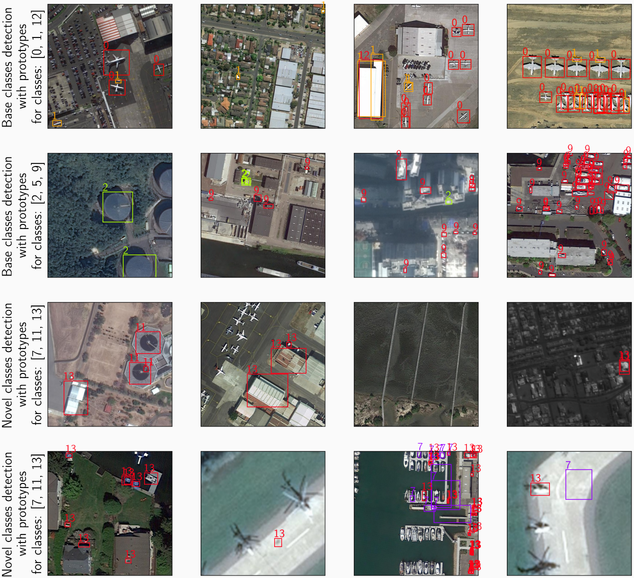


Figure 20: Qualitative results with Prototypical Faster R-CNN.

G. Qualitative Resultswith Attention-based methods

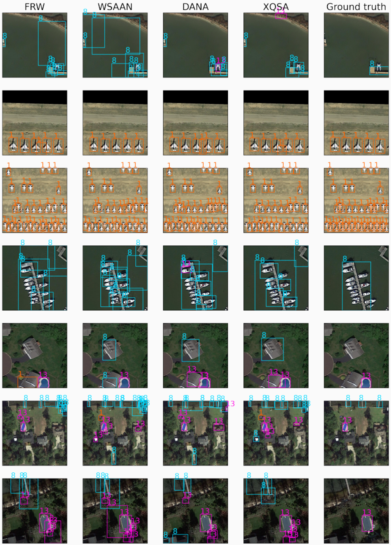


Figure 21: Qualitative results with Attention-based methods on base classes.

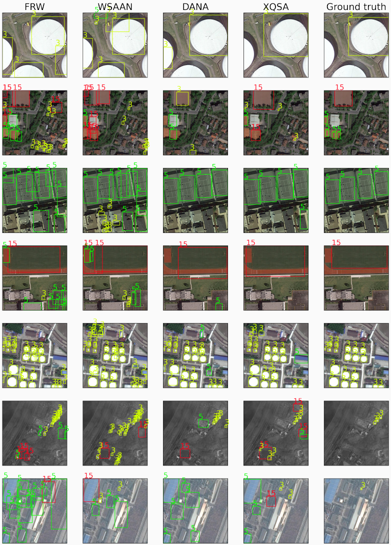


Figure 22: Qualitative results with Attention-based methods on novel classes.

G. Qualitative Resultswith FSDiffusionDet

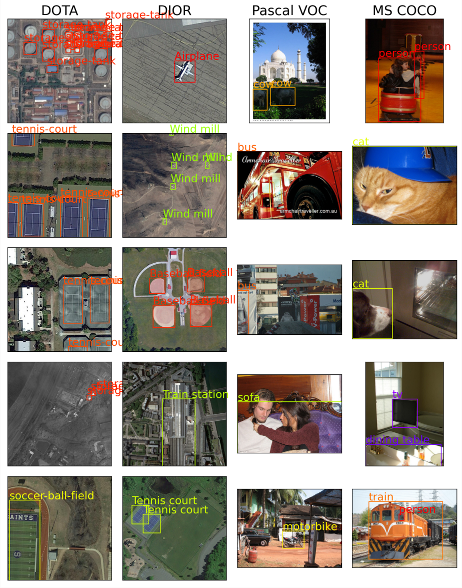


Figure 23: Qualitative results with Attention-based methods on novel classes.

References i

Bar, Amir et al. (2022). “Detreg: Unsupervised pretraining with region priors for object detection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14605–14615.

Cao, Yuhang et al. (2021). “Few-Shot Object Detection via Association and DIscrimination”. In: *Advances in Neural Information Processing Systems* 34, pp. 16570–16581.

Chen, Hao et al. (2018). “Lstd: A low-shot transfer detector for object detection”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 32. 1.

Chen, Shiqi et al. (2022). “Few Shot Object Detection for SAR Images via Feature Enhancement and Dynamic Relationship Modeling”. In: *Remote Sensing* 14.15, p. 3669.

Chen, Tung-I et al. (2021). “Dual-Awareness Attention for Few-Shot Object Detection”. In: *IEEE Transactions on Multimedia*, pp. 1–1. DOI: `10.1109/TMM.2021.3125195`.

Chu, Jinghui et al. (2021). “Joint Co-Attention And Co-Reconstruction Representation Learning For One-Shot Object Detection”. In: *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 2229–2233. DOI: `10.1109/ICIP42928.2021.9506387`.

Cortes, Corinna and Mehryar Mohri (2006). “On transductive regression”. In: *Advances in neural information processing systems* 19.

Dong, Na et al. (2022). “Incremental-detr: Incremental few-shot object detection via self-supervised learning”. In: *arXiv preprint arXiv:2205.04042*.

Fan, Qi et al. (2020). “Few-shot object detection with attention-RPN and multi-relation detector”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4013–4022.

Fan, Zhibo et al. (2021). “Generalized Few-Shot Object Detection without Forgetting”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4527–4536.

Gao, Bin-Bin et al. (2022). “Decoupling Classifier for Boosting Few-shot Object Detection and Instance Segmentation”. In: *Advances in Neural Information Processing Systems*.

Gao, Yipeng et al. (2022). “AcroFOD: An Adaptive Method for Cross-Domain Few-Shot Object Detection”. In: *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIII*. Springer, pp. 673–690.

Gao, Yuxuan et al. (2021). “A Fast and Accurate Few-Shot Detector for Objects with Fewer Pixels in Drone Image”. In: *Electronics* 10.7, p. 783.

Guirguis, Karim et al. (2022). “CFA: Constraint-based Finetuning Approach for Generalized Few-Shot Object Detection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 4039–4049.

Han, Guangxing, Shiyuan Huang, et al. (2022). “Meta faster r-cnn: Towards accurate few-shot object detection with attentive feature alignment”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. 1, pp. 780–789.

Han, Guangxing, Jiawei Ma, et al. (2022). “Few-shot object detection with fully cross-transformer”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5321–5330.

He, Jiabo et al. (2021). “Alpha-IoU: A Family of Power Intersection over Union Losses for Bounding Box Regression”. In: *Advances in Neural Information Processing Systems* 34, pp. 20230–20242.

Hsieh, Ting-I et al. (2019). “One-Shot Object Detection with Co-Attention and Co-Excitation”. In: *Advances in Neural Information Processing Systems 32*.

Huang, Xu et al. (2021). “Few-Shot Object Detection on Remote Sensing Images via Shared Attention Module and Balanced Fine-Tuning Strategy”. In: *Remote Sensing* 13.19, p. 3816.

Inoue, Naoto et al. (2018). “Cross-domain weakly-supervised object detection through progressive domain adaptation”. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 5001–5009.

Jiang, Xinyu et al. (2023). “Few-Shot Object Detection via Improved Classification Features”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 5386–5395.

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.

Karlinsky, Leonid et al. (2019). “Repmet: Representative-based metric learning for classification and few-shot object detection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 5197–5206.

Kaul, Prannay, Weidi Xie, and Andrew Zisserman (2022). “Label, verify, correct: A simple few shot object detection method”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14237–14247.

Kim, Geonuk, Hong-Gyu Jung, and Seong-Whan Lee (2020). “Few-Shot Object Detection via Knowledge Transfer”. In: *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pp. 3564–3569.

Le Jeune, Pierre and Anissa Mokraoui (2022). "A Comparative Attention Framework for Better Few-Shot Object Detection on Aerial Images". In: *arXiv preprint arXiv:2210.13923*.

Le Jeune, Pierre and Anissa Mokraoui (2023a). "Cross-Scale Query-Support Alignment Approach for Small Object Detection in the Few-Shot Regime". In: *2023 IEEE International Conference on Image Processing (ICIP)*.

Le Jeune, Pierre and Anissa Mokraoui (2023b). "Rethinking Intersection Over Union for Small Object Detection in Few-Shot Regime". In: *arXiv preprint arXiv:2307.09562*.

Le Jeune, Pierre, Anissa Mokraoui, et al. (2021). "Experience feedback using Representation Learning for Few-Shot Object Detection on Aerial Images". In: pp. 662–667. DOI: 10.1109/ICMLA52953.2021.00110.

Lee, Hojun, Myunggi Lee, and Nojun Kwak (2022). "Few-shot object detection by attending to per-sample-prototype". In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 2445–2454.

Lee, Kibok et al. (2022). "Rethinking Few-Shot Object Detection on a Multi-Domain Benchmark". In: *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XX*. Springer, pp. 366–382.

Li, Aoxue and Zhenguo Li (2021). "Transformation invariant few-shot object detection". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3094–3102.

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.

Li, Xiang, Lin Zhang, et al. (2020). "One-shot object detection without fine-tuning". In: *arXiv preprint arXiv:2005.03819*.

Li, Yiting et al. (2021). "Few-shot object detection via classification refinement and distractor retreatment". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 15395–15403.

Lin, Shaobo et al. (2023). "An Effective Crop-Paste Pipeline for Few-shot Object Detection". In: *arXiv preprint arXiv:2302.14452*.

Liu, Liyang et al. (2021). "Gendet: Meta learning to generate detectors from few shots". In: *IEEE Transactions on Neural Networks and Learning Systems*.

Liu, Nanqing et al. (2023). "Transformation-Invariant Network for Few-Shot Object Detection in Remote Sensing Images". In: *arXiv preprint arXiv:2303.06817*.

Liu, Weijie, Haohe Li, et al. (2021). “Dynamic Relevance Learning for Few-Shot Object Detection”. In: *arXiv preprint arXiv:2108.02235*.

Liu, Weijie, Chong Wang, et al. (2022). “Novel instance mining with pseudo-margin evaluation for few-shot object detection”. In: *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 2250–2254.

Miao, Caijing et al. (2019). “Sixray: A large-scale security inspection x-ray benchmark for prohibited item discovery in overlapping images”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2119–2128.

Nakamura, Yuzuru et al. (2022). “Few-shot Adaptive Object Detection with Cross-Domain CutMix”. In: *Proceedings of the Asian Conference on Computer Vision*, pp. 1350–1367.

Ouyang, Yan et al. (2022). “Few-shot object detection based on positive-sample improvement”. In: *Defence Technology*.

Perez-Rua, Juan-Manuel et al. (2020). “Incremental few-shot object detection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13846–13855.

Qiao, Limeng et al. (2021). “Defrcn: Decoupled faster r-cnn for few-shot object detection”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 8681–8690.

Rezatofighi, Hamid et al. (2019). “Generalized intersection over union: A metric and a loss for bounding box regression”. In: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 658–666.

She, Yihang et al. (2022). “Fast Hierarchical Learning for Few-Shot Object Detection”. In: *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, pp. 1993–2000.

Sun, Bo et al. (2021). “FSCE: Few-shot object detection via contrastive proposal encoding”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7352–7362.

Sun, Yiming et al. (2022). “Drone-Based RGB-Infrared Cross-Modality Vehicle Detection Via Uncertainty-Aware Learning”. In: *IEEE Transactions on Circuits and Systems for Video Technology* 32.10, pp. 6700–6713. DOI: 10.1109/TCSVT.2022.3168279.

Wang, Xin et al. (July 2020). “Frustratingly Simple Few-Shot Object Detection”. In: *International Conference on Machine Learning (ICML)*.

Wang, Yu-Xiong, Deva Ramanan, and Martial Hebert (2019). “Meta-learning to detect rare objects”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9925–9934.

Wang, Yan et al. (2022). "Context Information Refinement for Few-Shot Object Detection in Remote Sensing Images". In: *Remote Sensing* 14.14, p. 3255.

Wolf, Stefan et al. (2021). "Double Head Predictor based Few-Shot Object Detection for Aerial Imagery". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 721–731.

Wu, Aming, Yahong Han, et al. (2021). "Universal-Prototype Enhancing for Few-Shot Object Detection". In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9567–9576.

Wu, Aming, Suqi Zhao, et al. (2021). "Generalized and discriminative few-shot object detection via SVD-dictionary enhancement". In: *Advances in Neural Information Processing Systems* 34, pp. 6353–6364.

Wu, Jiaxi et al. (2020). "Multi-scale positive sample refinement for few-shot object detection". In: *European Conference on Computer Vision*. Springer, pp. 456–472.

Wu, Shuang et al. (2022). "Multi-faceted Distillation of Base-Novel Commonality for Few-Shot Object Detection". In: *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part IX*. Springer, pp. 578–594.

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, Yang and Renaud Marlet (2020). "Few-shot object detection and viewpoint estimation for objects in the wild". In: *European Conference on Computer Vision*. Springer, pp. 192–210.

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.

Xiong, Wuti and Li Liu (2022). "CD-FSOD: A Benchmark for Cross-domain Few-shot Object Detection". In: *arXiv preprint arXiv:2210.05311*.

Xu, Chang et al. (2022). "Detecting Tiny Objects in Aerial Images: A Normalized Wasserstein Distance and A New Benchmark". In: *ISPRS Journal of Photogrammetry and Remote Sensing (ISPRS JPRS)*.

Xu, Honghui et al. (2021). "Few-Shot Object Detection via Sample Processing". In: *IEEE Access* 9, pp. 29207–29221.

Xu, Yang et al. (2021). *SIMPL: Generating Synthetic Overhead Imagery to Address Zero-shot and Few-Shot Detection Problems*. DOI: `10.48550/ARXIV.2106.15681`. URL: `https://arxiv.org/abs/2106.15681`.

- Yan, Xiaopeng et al. (2019). “Meta r-cnn: Towards general solver for instance-level low-shot learning”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 9577–9586.
- Yang, Yukuan et al. (2020). “Restoring negative information in few-shot object detection”. In: *Advances in neural information processing systems* 33, pp. 3521–3532.
- Yin, Li, Juan M Perez-Rua, and Kevin J Liang (2022). “Sylph: A hypernetwork framework for incremental few-shot object detection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9035–9045.
- Zhang, Gongjie, Kaiwen Cui, et al. (2021). “PNPDet: Efficient few-shot detection without forgetting via plug-and-play sub-networks”. In: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3823–3832.
- Zhang, Gongjie, Zhipeng Luo, et al. (2022). “Meta-DETR: Image-Level Few-Shot Detection with Inter-Class Correlation Exploitation”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*. DOI: 10.1109/TPAMI.2022.3195735.
- Zhang, Shan et al. (2022). “Kernelized few-shot object detection with efficient integral aggregation”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 19207–19216.
- Zhang, Weilin and Yu-Xiong Wang (2021). “Hallucination improves few-shot object detection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 13008–13017.
- Zhao, Lifan, Yunlong Meng, and Lin Xu (2022). “OA-FSUI2IT: A Novel Few-Shot Cross Domain Object Detection Framework with Object-Aware Few-Shot Unsupervised Image-to-Image Translation”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 36. 3, pp. 3426–3435.
- Zhao, Xiangyun, Xu Zou, and Ying Wu (2021). “Morphable detector for object detection on demand”. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4771–4780.
- Zhao, Yizhou, Xun Guo, and Yan Lu (2022). “Semantic-aligned fusion transformer for one-shot object detection”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7601–7611.
- Zhou, Zheng et al. (2022). “FSODS: A Lightweight Metalearning Method for Few-Shot Object Detection on SAR Images”. In: *IEEE Transactions on Geoscience and Remote Sensing* 60, pp. 1–17.