Détection d'objet sur image aérienne en régime few-shot.

Hicham Talaoubrid

Encadrants :

- Anissa MOKRAOUI (USPN, L2TI, directrice de thèse)
- Ismail BEN AYED (ETS Montréal, LIVIA, co-directeur de thèse)
- Rémi Harvey (COSE, co-encadrant)

Comité de suivi

4 Octobre 2024

Overview of the presentation

1.1 Context & Motivation - Détection d'objet few-shot

Principe de la détection d'objets

Étant donné un ensemble de classes $\mathcal{C}$, trouvez toutes les occurrences d'objets appartenant à une classe $c \in \mathcal{C}$ dans une image I. Chaque objet est représenté par (x_1, y_1, x_2, y_2, c).

Principe de la détection d'objets

Étant donné un ensemble de classes $\mathcal{C}$, trouvez toutes les occurrences d'objets appartenant à une classe $c \in \mathcal{C}$ dans une image I. Chaque objet est représenté par (x_1, y_1, x_2, y_2, c).

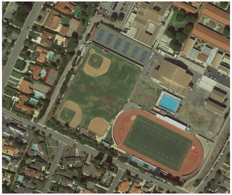


Image d'entrée I

Principe de la détection d'objets

Étant donné un ensemble de classes $\mathcal{C}$, trouvez toutes les occurrences d'objets appartenant à une classe $c \in \mathcal{C}$ dans une image I. Chaque objet est représenté par (x_1, y_1, x_2, y_2, c).

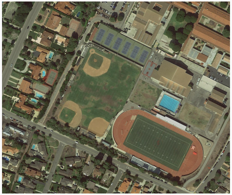


Image d'entrée I

$$\mathcal{C} = \{\text{Terrain de baseball, Piscine, Piste d'athlétisme}\}$$

1.1 Context & Motivation - Détection d'objet few-shot

Principe de la détection d'objets

Étant donné un ensemble de classes $\mathcal{C}$, trouvez toutes les occurrences d'objets appartenant à une classe $c \in \mathcal{C}$ dans une image I. Chaque objet est représenté par (x_1, y_1, x_2, y_2, c).

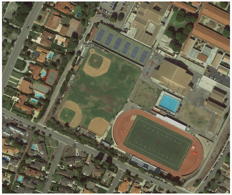


Image d'entrée I

$\longrightarrow$ Modèle de détection

$$\mathcal{C} = \{\text{Terrain de baseball, Piscine, Piste d'athlétisme}\}$$

1.1 Context & Motivation - Détection d'objet few-shot

Principe de la détection d'objets

Étant donné un ensemble de classes $\mathcal{C}$, trouvez toutes les occurrences d'objets appartenant à une classe $c \in \mathcal{C}$ dans une image I. Chaque objet est représenté par (x_1, y_1, x_2, y_2, c).

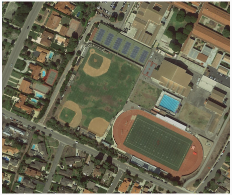
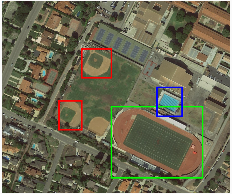


Image d'entrée I

$\longrightarrow$ Modèle de détection $\longrightarrow$



Résultats de la détection

$$\mathcal{C} = \{\text{Terrain de baseball, Piscine, Piste d'athlétisme}\}$$

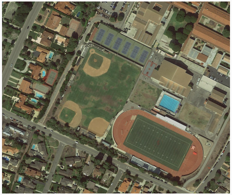
1.1 Context & Motivation - Détection d'objet few-shot

n-way k-shot object detection

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

n-way k-shot object detection

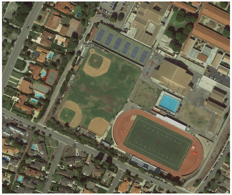
Given support examples $\{(x_1, a_1), \dots, (x_{nk}, a_{nk})\}$ it consists in detecting all occurrences of classes in $\mathcal{C}$ ($|\mathcal{C}| = n$) in a query image x_q.



Query image

n-way k-shot object detection

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

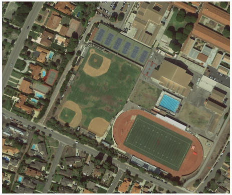


Query image

Support examples

n-way k-shot object detection

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



Query image

$\longrightarrow$ Detection model

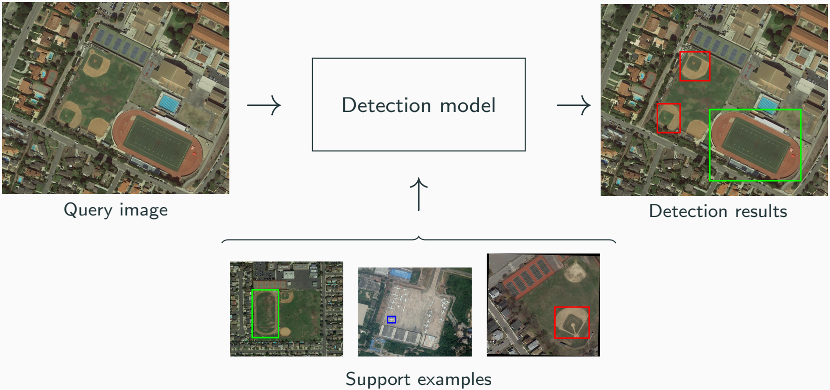
$\uparrow$

Support examples

1.1 Context & Motivation - Détection d'objet few-shot

n-way k-shot object detection

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

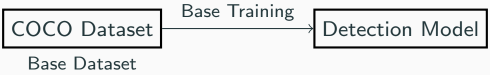


Cross-Domain Few-Shot Object Detection

In Cross-Domain Few-Shot Object Detection (CD-FSOD), two distinct datasets are used during base training and fine-tuning. CD-FSOD is more challenging as the model must not only adapt to new classes but also to new images.

Cross-Domain Few-Shot Object Detection

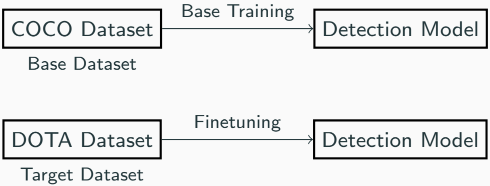
In Cross-Domain Few-Shot Object Detection (CD-FSOD), two distinct datasets are used during base training and fine-tuning. CD-FSOD is more challenging as the model must not only adapt to new classes but also to new images.



1.2 Context & Motivation - Détection d'objet few-shot cross-domain

Cross-Domain Few-Shot Object Detection

In Cross-Domain Few-Shot Object Detection (CD-FSOD), two distinct datasets are used during base training and fine-tuning. CD-FSOD is more challenging as the model must not only adapt to new classes but also to new images.



2.1 Proposed Approaches - Diffusion Based Models

The main idea of DiffusionDet is to apply the diffusion principle to box generation.

Random boxes are first sampled, and a model is trained to refine iteratively the size and position of the boxes so that they localize the objects in the input image.

Specifically, the boxes are iteratively denoised by the model.

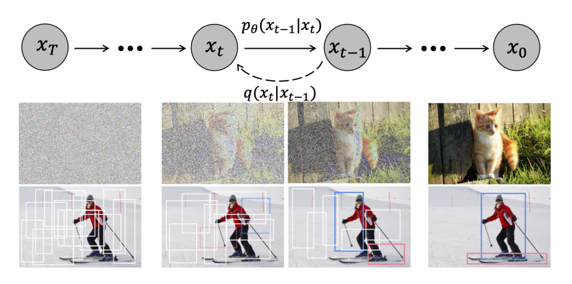


Figure 1: Diffusion model for object detection. In the first line a diffusion model where q is the diffusion process and p_θ is the reverse process. Then a diffusion model for image generation task. And in the last one we propose to formulate object detection as a denoising diffusion process from noisy boxes to object boxes.
taken from DiffusionDet paper Shoufa Chen et al. 2022

The denoising part of DiffusionDet is a lightweight hybrid network, it consists of a self-attention layer (transformer-like) followed by a dynamic layer (called an Instance Interaction layer). The detection head processes object features independently, but the Instance Interaction layer enables interactions between instances. The detection head is applied iteratively to refine the bounding boxes

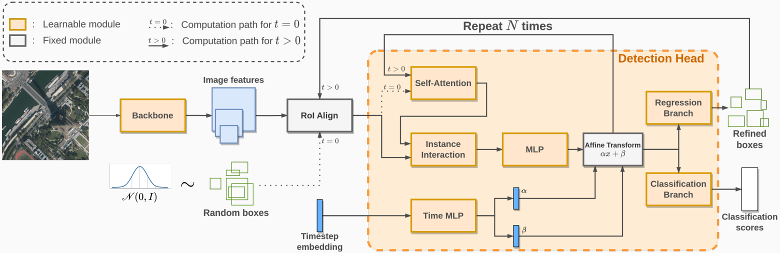


Figure 2: DiffusionDet architecture and detailed detection head design.

2.2 Proposed Approaches - Few-Shot DiffusionDet

Let us introduce some notations.

We denote by $\mathcal{F}_c$ and $\mathcal{F}_r$ the sub-models responsible for classification and regression respectively:

$$\mathcal{F}_c = f_c \circ f, \tag{1}$$

$$\mathcal{F}_r = f_r \circ f, \tag{2}$$

$$\text{with } f = f^{L-1} \circ f^{L-2} \circ \cdots \circ f^{1}. \tag{3}$$

2.2 Proposed Approaches - Few-Shot DiffusionDet

Let us introduce some notations.

We denote by $\mathcal{F}_c$ and $\mathcal{F}_r$ the sub-models responsible for classification and regression respectively:

$$\mathcal{F}_c = f_c \circ f, \tag{1}$$

$$\mathcal{F}_r = f_r \circ f, \tag{2}$$

$$\text{with } f = f^{L-1} \circ f^{L-2} \circ \cdots \circ f^1. \tag{3}$$

To leverage DiffusionDet in the few-shot regime, we propose to adopt a Transfer Learning approach. Once, the model is base-trained, it is prepared for fine-tuning.

A large part of the model is frozen. The crucial part is that we can finely control the amount of frozen layers with a hyperparameter γ.

Denoting frozen layers with a bar ($\overline{f^1}$), the frozen detection model is expressed as

$$f = f^L \circ \cdots \circ f^{\gamma+1} \circ \overline{f^\gamma} \circ \cdots \circ \overline{f^1}. \tag{4}$$

3.1 Results and Analysis - Few-Shot Senarios

Figure 3: Performance ($mAP_{0.5}$) of FSDD (ours), XQSA **lejeune2023xqsa**, FRW Kang et al. 2019, DANA T.-I. Chen et al. 2021 and WSAAN Z. Xiao et al. 2021 on DOTA, DIOR, Pascal VOC and MS COCO against the number of shots. Black dashed lines represent performance achieved with full supervision.

3.1 Results and Analysis - Few-Shot Senarios

Determining how much of the model should be fine-tuned is a complex task in FSOD. Here, we investigate this choice with thorough experiments on four datasets.

Freezing point	Plasticity	DOTA	DIOR	Pascal VOC	COCO
FT whole	100.00 %	60.09	52.17	43.10	17.15
Bias only	35.98 %	**60.45**	55.12	49.90	20.19
BatchNorm only	35.97 %	59.35	55.63	51.96	19.70
Up to stage 1	99.98 %	58.85	53.37	43.81	17.72
Up to stage 4	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

Table 1: Influence of the freezing point on the FS performance on DOTA, DIOR, Pascal VOC, and COCO. mAP is reported with a 0.5 IoU threshold and $k = 10$ shots.

3.2 Results and Analysis - Cross-Domain Senarios

The great results of FSDD on aerial images suggest trying the more complex scenario of CD-FSOD almost untouched in the literature.

-1st Senario: COCO $\rightarrow$ X,

k **Shots**	**DIOR**	**DOTA**	**DeepFruits**	**SIXRay**	**VisDrone**
1	11.10	4.03	38.47	4.80	2.83
5	30.42	14.45	55.58	13.25	5.74
10	38.73	25.02	68.37	21.26	7.50
20	48.23	33.31	73.95	30.06	9.14
50	56.97	43.23	76.65	41.93	11.47

Table 2: Cross-domain performance results on 5 scenarios COCO $\rightarrow$ DIOR / DOTA / DeepFruits (K. Lee et al. 2022) / SIXRay (Miao et al. 2019) / VisDrone (Zhu et al. 2021).

3.2 Results and Analysis - Cross-Domain Senarios

-2^{nd} Senario: DOTA $\rightarrow$ DIOR and DIOR $\rightarrow$ DOTA,

	DOTA $\rightarrow$ DIOR		**DIOR $\rightarrow$ DOTA**	
***k* shots**	**FT head only**	**FT whole**	**FT head only**	**FT whole**
1	**20.18**	9.40	**5.41**	5.09
5	**34.43**	29.57	**25.88**	24.90
10	**41.48**	38.44	31.99	**33.30**
20	**49.00**	45.36	38.77	**41.30**
50	**54.07**	53.51	44.07	**49.22**

Table 3: FSDiffusionDet Cross-Domain (CD-FSOD) results on the scenarios DOTA $\rightarrow$ DIOR and DIOR $\rightarrow$ DOTA. Performance is reported with $mAP_{0.5}$ values.

4.0 Conclusion and Perspectives

To summarize, our contributions are:

- A **novel FSOD model** and a training strategy with greater versatility than existing approaches,
- Extensive experiments to showcase the superiority of **FSDD** over existing **FSOD** methods on **aerial images**, and
- A demonstration of promising results in the challenging **CD-FSOD** scenarios.

Thank you for your attention

Any questions ?

hicham.talaoubrid1@edu.univ-paris13.fr

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, Shoufa et al. (2022). "Diffusiondet: Diffusion model for object detection". In: *arXiv preprint arXiv:2211.09788*.

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.

Everingham, Mark et al. (2010). "The pascal visual object classes (voc) challenge". In: *International journal of computer vision* 88.2, pp. 303–338.

Fan, Qi et al. (2020). "Few-Shot Object Detection with Attention-RPN and Multi-Relation Detector". In: *CVPR*.

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.

Ganea, Dan Andrei, Bas Boom, and Ronald Poppe (2021). "Incremental few-shot instance segmentation". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 1185–1194.

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.

Han, Guangxing 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.

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.

Jeune, Pierre Le et al. (2021). "Experience feedback using Representation Learning for Few-Shot Object Detection on Aerial Images". In: *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, pp. 662–667. DOI: 10.1109/ICMLA52953.2021.00110.

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.

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). "Improving Few-Shot Object Detection through a Performance Analysis on Aerial and Natural Images". In: *30th European Signal Processing Conference (EUSIPCO)*, pp. 513–517. DOI: 10.23919/EUSIPCO55093.2022.9909878.

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*.

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

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

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.

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.

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

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.

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.

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. (2019a). "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.

— (2019b). "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.

References iv

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, 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.

Zhu, Pengfei et al. (2021). "Detection and Tracking Meet Drones Challenge". In: *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1. DOI: 10.1109/TPAMI.2021.3119563.