

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



- 1. Introduction à la détection d'objets few-shot**
- 2. Étude et analyse des performances de la détection d'objets few-shot cross-domain**
- 3. Contributions scientifiques**
- 4. Travaux futurs**
- 5. Formations**

1.1 Principe de la détection d'objets

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) .

1.1 Principe de la détection d'objets

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

1.1 Principe de la détection d'objets

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 Principe de la détection d'objets

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 Principe de la détection d'objets

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



Résultats de la détection

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

1.2 Principe de la détection d'objets few-shot

Détection d'objets n -way k -shot

Étant donné des exemples de support $\{(x_1, a_1), \dots, (x_{nk}, a_{nk})\}$, il s'agit de détecter toutes les occurrences des classes dans \mathcal{C} ($|\mathcal{C}| = n$) dans une image de requête x_q .

En règle générale, k est compris entre 1 et 50.

1.2 Principe de la détection d'objets few-shot

Détection d'objets n -way k -shot

Étant donné des exemples de support $\{(x_1, a_1), \dots, (x_{nk}, a_{nk})\}$, il s'agit de détecter toutes les occurrences des classes dans \mathcal{C} ($|\mathcal{C}| = n$) dans une image de requête x_q . En règle générale, k est compris entre 1 et 50.



Query image

1.2 Principe de la détection d'objets few-shot

Détection d'objets *n-way k-shot*

Étant donné des exemples de support $\{(x_1, a_1), \dots, (x_{nk}, a_{nk})\}$, il s'agit de détecter toutes les occurrences des classes dans \mathcal{C} ($|\mathcal{C}| = n$) dans une image de requête x_q . En règle générale, k est compris entre 1 et 50.



Query image

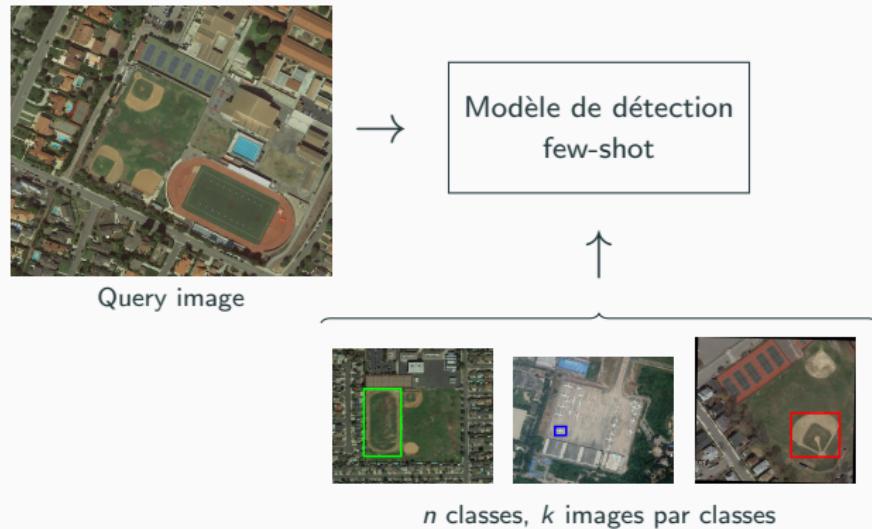


n classes, k images par classes

1.2 Principe de la détection d'objets few-shot

Détection d'objets n -way k -shot

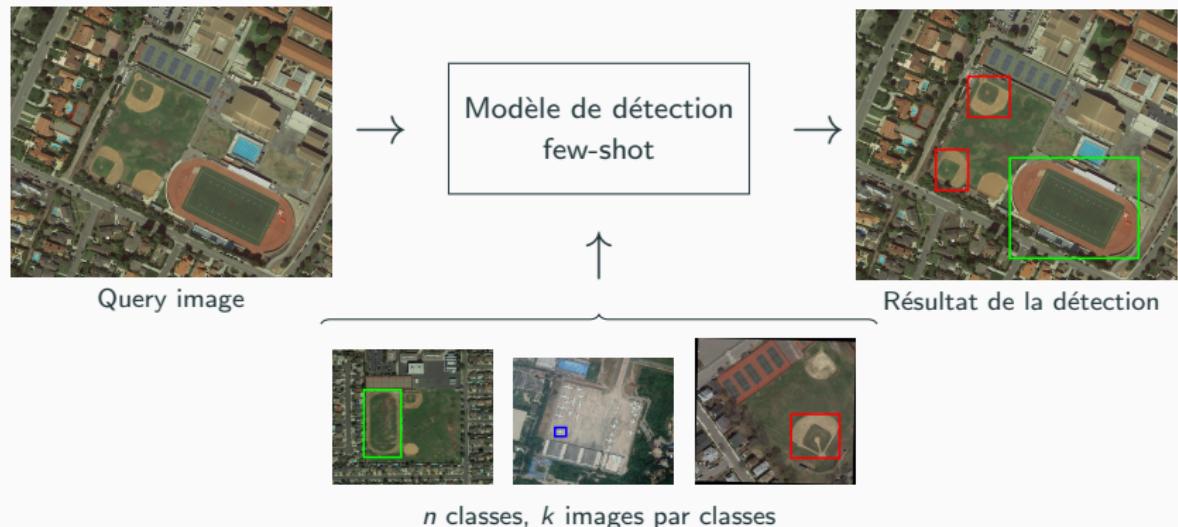
Étant donné des exemples de support $\{(x_1, a_1), \dots, (x_{nk}, a_{nk})\}$, il s'agit de détecter toutes les occurrences des classes dans \mathcal{C} ($|\mathcal{C}| = n$) dans une image de requête x_q . En règle générale, k est compris entre 1 et 50.



1.2 Principe de la détection d'objets few-shot

Détection d'objets n -way k -shot

Étant donné des exemples de support $\{(x_1, a_1), \dots, (x_{nk}, a_{nk})\}$, il s'agit de détecter toutes les occurrences des classes dans \mathcal{C} ($|\mathcal{C}| = n$) dans une image de requête x_q . En règle générale, k est compris entre 1 et 50.



2.1 Principe de la détection d'objets few-shot cross-domain

Détection d'objets few-shot cross-domain

Dans la détection d'objets few-shot cross-domain, on dispose de deux datasets de deux domaines distincts. L'un est composé d'un grand nombre de données annotées (le dataset source) tandis que l'autre est plutôt limité (le dataset cible). L'objectif est de s'aider du dataset source pour réaliser la détection sur le dataset cible. Le modèle doit dans ce cas s'adapter à la fois à de nouvelles classes et à de nouvelles images ce qui complexifie la tâche.



Domaine source



Domaine cible

2.1 Principe de la détection d'objets few-shot cross-domain

Détection d'objets few-shot cross-domain

Dans la détection d'objets few-shot cross-domain, on dispose de deux datasets de deux domaines distincts. L'un est composé d'un grand nombre de données annotées (le dataset source) tandis que l'autre est plutôt limité (le dataset cible). L'objectif est de s'aider du dataset source pour réaliser la détection sur le dataset cible. Le modèle doit dans ce cas s'adapter à la fois à de nouvelles classes et à de nouvelles images ce qui complexifie la tâche.



Dataset source
(grand nombre d'images
annotées)

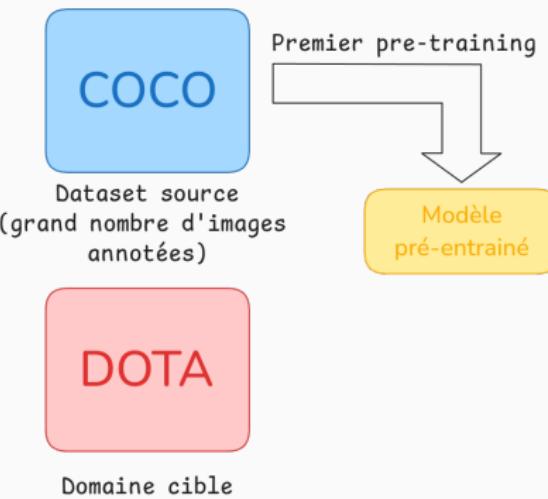


Domaine cible

2.1 Principe de la détection d'objets few-shot cross-domain

Détection d'objets few-shot cross-domain

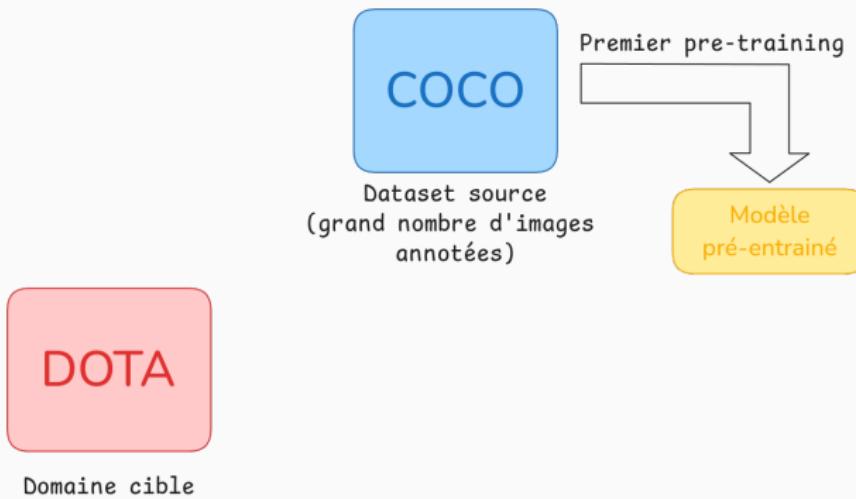
Dans la détection d'objets few-shot cross-domain, on dispose de deux datasets de deux domaines distincts. L'un est composé d'un grand nombre de données annotées (le dataset source) tandis que l'autre est plutôt limité (le dataset cible). L'objectif est de s'aider du dataset source pour réaliser la détection sur le dataset cible. Le modèle doit dans ce cas s'adapter à la fois à de nouvelles classes et à de nouvelles images ce qui complexifie la tâche.



2.1 Principe de la détection d'objets few-shot cross-domain

Détection d'objets few-shot cross-domain

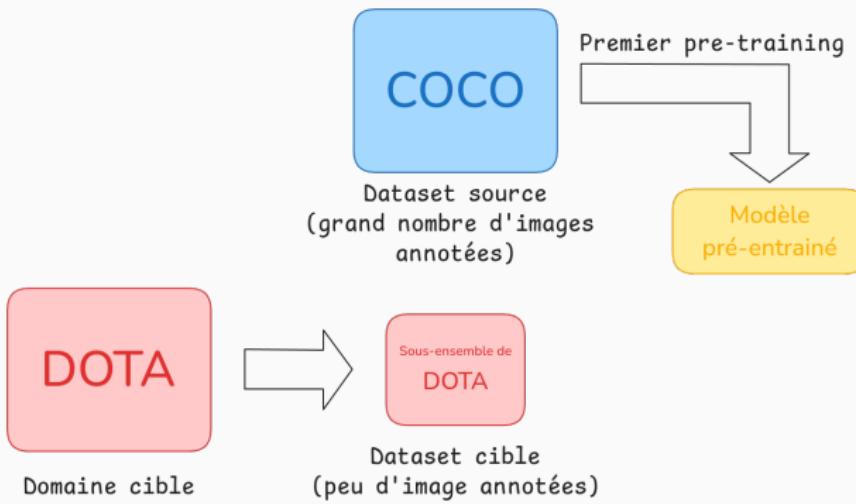
Dans la détection d'objets few-shot cross-domain, on dispose de deux datasets de deux domaines distincts. L'un est composé d'un grand nombre de données annotées (le dataset source) tandis que l'autre est plutôt limité (le dataset cible). L'objectif est de s'aider du dataset source pour réaliser la détection sur le dataset cible. Le modèle doit dans ce cas s'adapter à la fois à de nouvelles classes et à de nouvelles images ce qui complexifie la tâche.



2.1 Principe de la détection d'objets few-shot cross-domain

Détection d'objets few-shot cross-domain

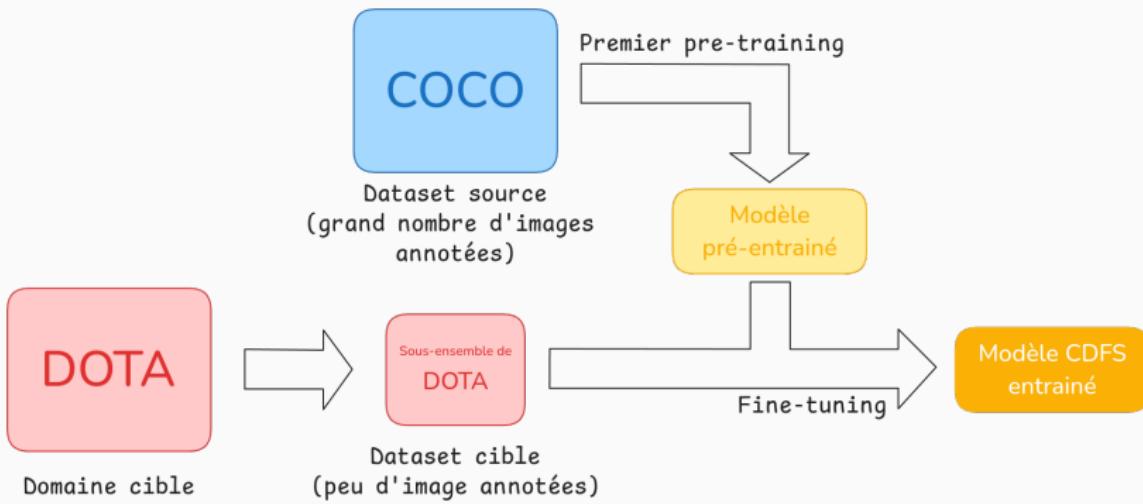
Dans la détection d'objets few-shot cross-domain, on dispose de deux datasets de deux domaines distincts. L'un est composé d'un grand nombre de données annotées (le dataset source) tandis que l'autre est plutôt limité (le dataset cible). L'objectif est de s'aider du dataset source pour réaliser la détection sur le dataset cible. Le modèle doit dans ce cas s'adapter à la fois à de nouvelles classes et à de nouvelles images ce qui complexifie la tâche.



2.1 Principe de la détection d'objets few-shot cross-domain

Détection d'objets few-shot cross-domain

Dans la détection d'objets few-shot cross-domain, on dispose de deux datasets de deux domaines distincts. L'un est composé d'un grand nombre de données annotées (le dataset source) tandis que l'autre est plutôt limité (le dataset cible). L'objectif est de s'aider du dataset source pour réaliser la détection sur le dataset cible. Le modèle doit dans ce cas s'adapter à la fois à de nouvelles classes et à de nouvelles images ce qui complexifie la tâche.



2.2 Principaux travaux réaliser pour répondre aux problèmes de la détection d'objets few-shot cross-domain

Les principaux travaux réalisés au cours de l'année : Principaux travaux réalisés au cours de l'année :

- Développement d'un outil permettant de comparer rapidement et efficacement les performances des modèles d'intelligence artificielle les plus performants.
- Mise en évidence de l'importance du choix d'un dataset source pertinent, avec une estimation de l'écart inter-domaines pour identifier le dataset source le plus adapté.
- Limitation du risque d'overfitting afin d'assurer une meilleure généralisation des modèles et, par conséquent, d'améliorer leurs performances.

2.3 Listes des méthodes de détection d'objets pouvant être utilisé en cross-domain

Name	Backbone
Meta-RCNN (Wu, Sahoo, and Hoi 2020)	ResNet50
TFA w/cos (X. Wang et al. 2020)	ResNet50
FSCE (B. Sun et al. 2021)	ResNet50
DeFRCN (Qiao et al. 2021)	ResNet50
FSDiffusionDet chen2022diffusiondet	ResNet50
Distill-cdfsod (Xiong 2023)	ResNet50
ViTDeT-FT (Y. Li et al. 2022)	ViT-B/14
Detic (Zhou et al. 2022)	ViT-L/14
Detic-FT (Zhou et al. 2022)	ViT-L/14
DE-ViT (Zhang, Y. Wang, and Bouliarias 2024)	ViT-L/14
DE-ViT-FT (Zhang, Y. Wang, and Bouliarias 2024)	ViT-L/14
CD-ViT0 (Fu et al. 2024)	ViT-L/14

Table 1: Liste des méthodes de détection d'objets pouvant être utilisé en cross-domain. Ces méthodes seront toutes tester et comparées dans un benchmark à l'aide d'un outil dévoloppé cette année. Il sera mis à disposition après la publication du benchmark.

2.4 Etude et analyse de l'impact du dataset source sur les performances des modèles de détection d'objets few-shot cross-domain

<i>k</i> shots	DOTA → DIOR	COCO → DIOR	DIOR → DOTA	COCO → DOTA	DOTA → DOTA	DIOR → DIOR
1	20.18	11.10	5.41	4.03	-	-
5	34.43	30.42	25.88	14.45	-	-
10	41.48	38.73	31.99	25.02	60.45	57.51
20	49	48.23	38.77	33.31	-	-
50	54.07	56.97	44.07	43.23	-	-

Table 2: Comparaison des performances de FSDiffusionDet en utilisant différents datasets sources pour la détection d'objet cross-domain. Les performances sont évaluées en mAP (%) rapporté avec un seuil IoU à 0.5.

2.5 Estimation de l'écart entre domaines

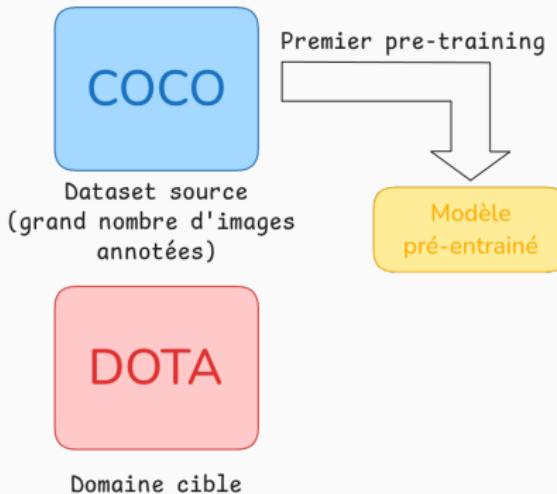


Figure 1: Méthode utilisée pour extraire le vecteur des caractéristiques. L'idée est de remplacer la tête de classification par un module ROI pooling.

2.5 Estimation de l'écart entre domaines

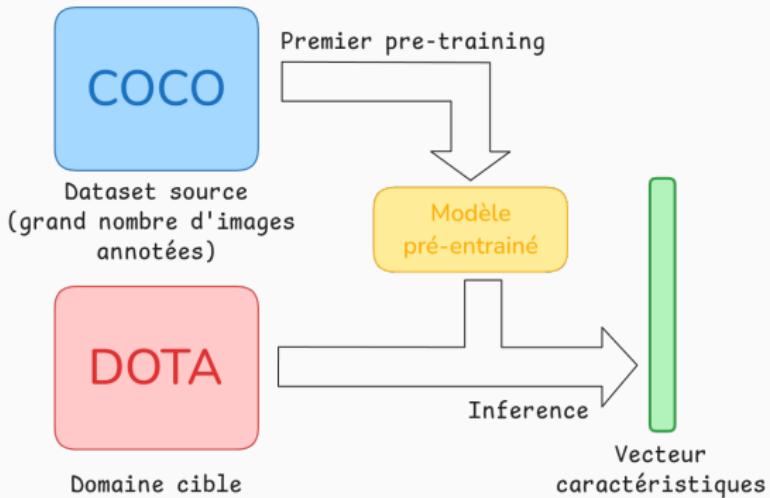


Figure 1: Méthode utilisée pour extraire le vecteur des caractéristiques. L'idée est de remplacer la tête de classification par un module ROI pooling.

2.5 Estimation de l'écart entre domaines

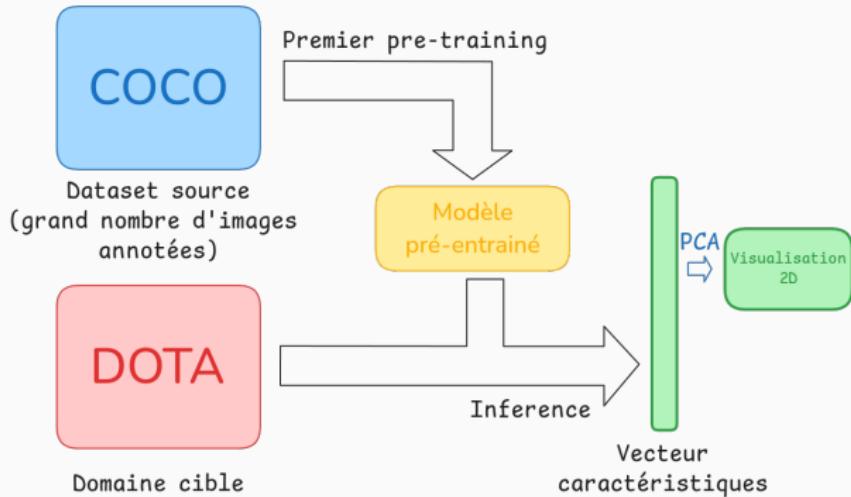


Figure 1: Méthode utilisée pour extraire le vecteur des caractéristiques. L'idée est de remplacer la tête de classification par un module ROI pooling.

2.5 Estimation de l'écart entre domaines

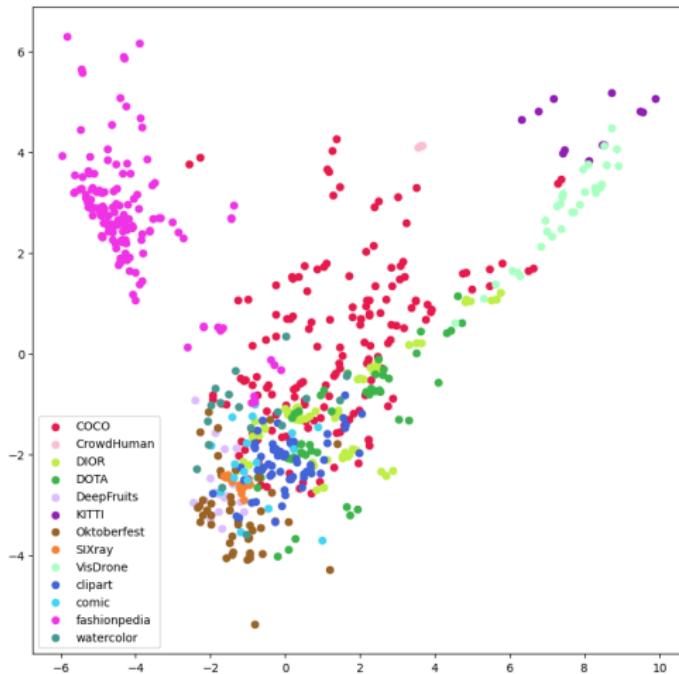
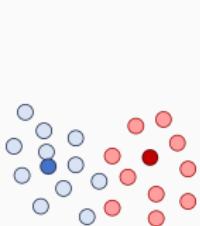
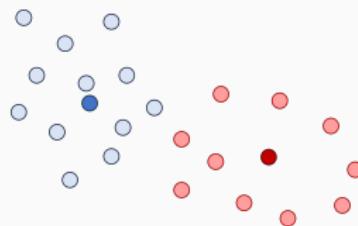


Figure 2: Représentation 2D des projections des barycentres de chaque classe, obtenue après le passage des images de chaque dataset dans un modèle pré-entraîné sur COCO, suivi d'une réduction dimensionnelle par PCA.

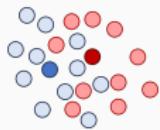
2.5 Estimation de l'écart entre domaines



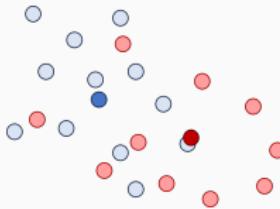
(a) Variance intra-classe faible,
Variance inter-classe forte,



(b) Variance intra-classe forte,
Variance inter-classe forte,



(d) Variance intra-classe faible,
Variance inter-classe faible,



(c) Variance intra-classe forte,
Variance inter-classe faible,

Figure 3: Illustration conceptuelle de la séparabilité des classes en relation avec les variances inter-classes et intra-classes. Les projections dans l'espace latent des images d'une classe sont représentées par des points bleus, tandis que les projections d'une autre classe sont marquées en rouge, avec les barycentres de chaque classe indiqués par des points plus foncés.

2.5 Estimation de l'écart entre domaines

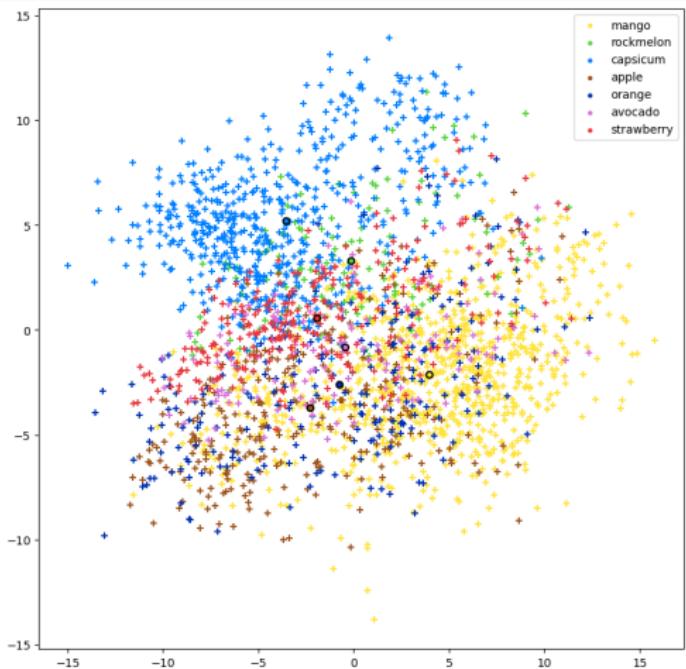


Figure 4: Représentation 2D de la projection des objets de DeepFruits, obtenue après le passage des images de chaque dataset dans un modèle pré-entraîné sur COCO, suivi d'une réduction dimensionnelle par PCA. Les barycentres mis en valeur avec des cercles de couleurs.

2.6 Méthodes PEFT (Parameter Efficient Fine-Tuning) pour limiter l'overfitting

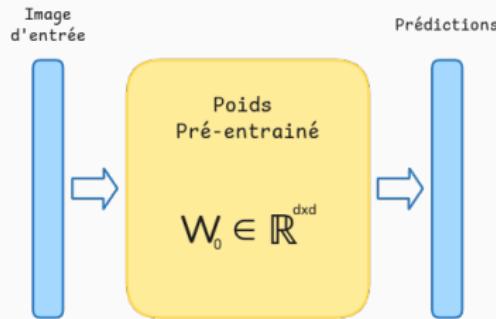


Figure 5: LoRA (Low-Rank Adaptation) est une technique permettant entre autre de limiter le risque d'overfitting pour assurer de meilleures performances.

2.6 Méthodes PEFT (Parameter Efficient Fine-Tuning) pour limiter l'overfitting



Figure 5: LoRA (Low-Rank Adaptation) est une technique permettant entre autre de limiter le risque d'overfitting pour assurer de meilleures performances.

2.6 Méthodes PEFT (Parameter Efficient Fine-Tuning) pour limiter l'overfitting

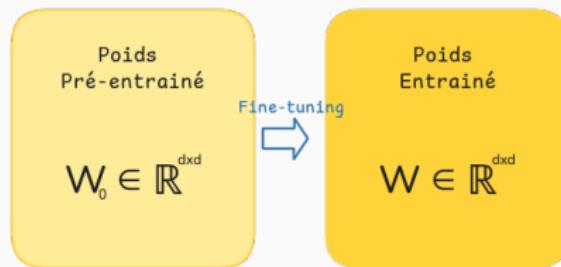


Figure 5: LoRA (Low-Rank Adaptation) est une technique permettant entre autre de limiter le risque d'overfitting pour assurer de meilleures performances.

2.6 Méthodes PEFT (Parameter Efficient Fine-Tuning) pour limiter l'overfitting

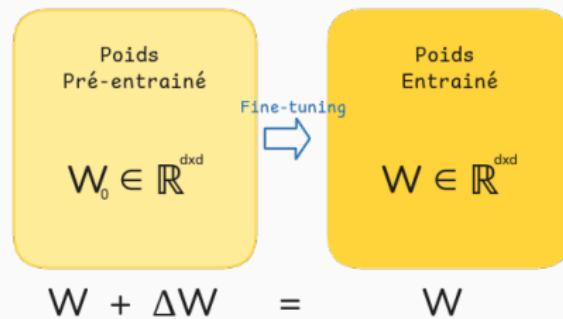


Figure 5: LoRA (Low-Rank Adaptation) est une technique permettant entre autre de limiter le risque d'overfitting pour assurer de meilleures performances.

2.6 Méthodes PEFT (Parameter Efficient Fine-Tuning) pour limiter l'overfitting

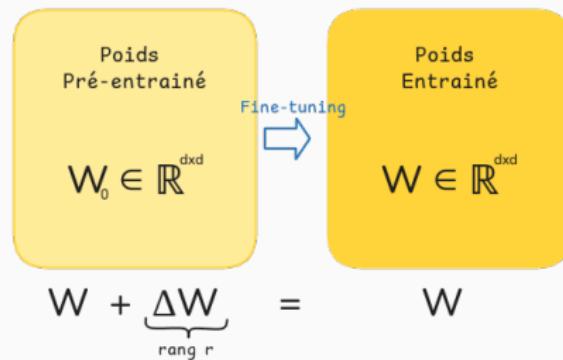


Figure 5: LoRA (Low-Rank Adaptation) est une technique permettant entre autre de limiter le risque d'overfitting pour assurer de meilleures performances.

2.6 Méthodes PEFT (Parameter Efficient Fine-Tuning) pour limiter l'overfitting

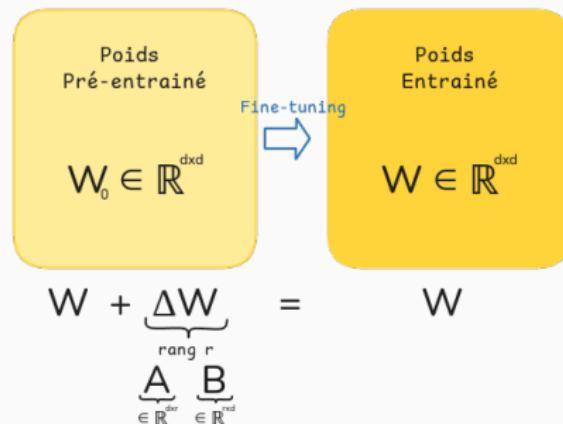


Figure 5: LoRA (Low-Rank Adaptation) est une technique permettant entre autre de limiter le risque d'overfitting pour assurer de meilleures performances.

2.6 Méthodes PEFT (Parameter Efficient Fine-Tuning) pour limiter l'overfitting

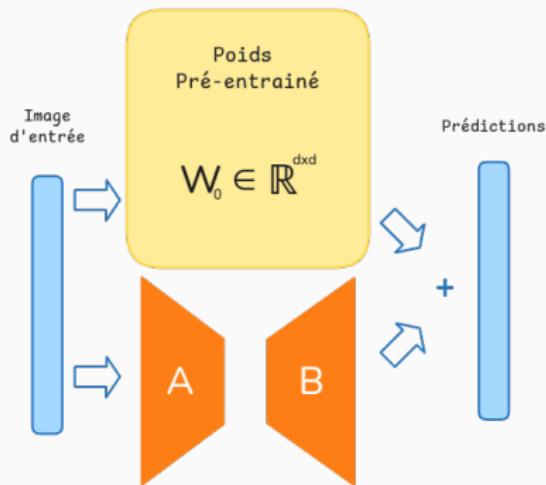


Figure 5: LoRA (Low-Rank Adaptation) est une technique permettant entre autre de limiter le risque d'overfitting pour assurer de meilleures performances.

2.6 Méthodes PEFT (Parameter Efficient Fine-Tuning) pour limiter l'overfitting

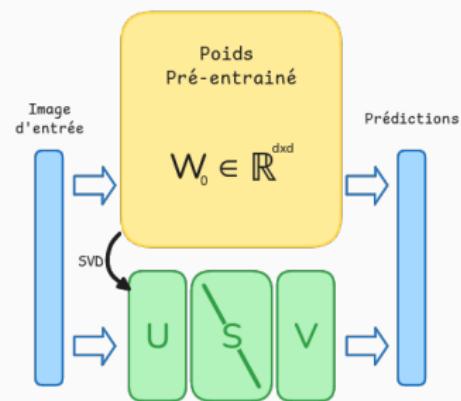


Figure 6: SVF (Y. Sun et al. 2022) se base sur une décomposition en valeurs singulières. Dans ce cas on entraîne seulement les valeurs singulières.

2.7 Premiers résultats avec LoRA

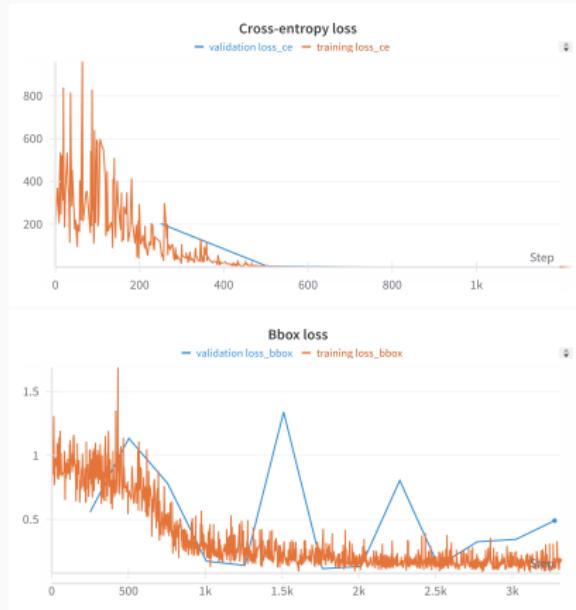


Figure 7: Impact de LoRA sur l'entraînement de FSDiffusionDet sur DIOR en utilisant 10 shots et un rang de 16.

2.7 Premiers résultats avec LoRA

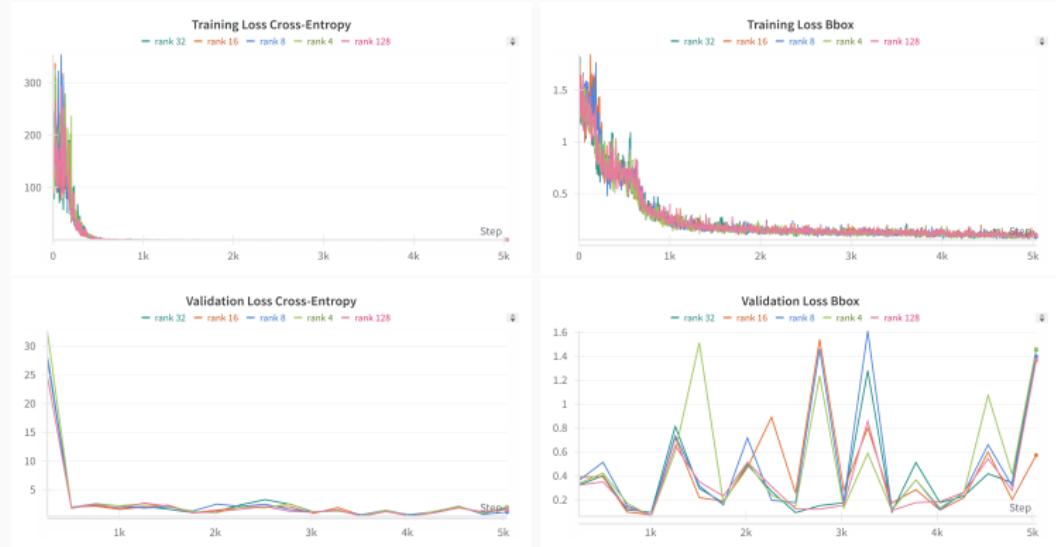


Figure 8: Impact du rang de LoRA sur l'entraînement de FSDiffusionDet sur DIOR en utilisant 10 shots.

3.0 Contributions scientifiques

Articles de conférence acceptés (avant la thèse) :

- Improving Few-Shot and Cross-Domain Object Detection on Aerial Images with a Diffusion-Based Detector – accepté et présenté à **IGARSS** (7 Juil. 2024)
(relecture du papier, participation aux expériences et présentation oral)

Articles de conférence soumis

- Indirect Attention: IA-DETR for one shot Object Detection – soumis à **ICLR** (2024) (participation au code et aux expériences)

4.0 Formations

Formation doctorale suivie :

- Formations professionnalisantes et langues – Anglais de communication niveau faux débutant-intermédiaire **4 ECTS**

Conférences présentées :

- Présentation orale à **IGARSS** (8-12 Juil.) – *Improving Few-Shot and Cross-Domain Object Detection on Aerial Images with a Diffusion-Based Detector* **24 ECTS**

Mobilité :

- Mobilité de 4 mois à l'ETS Montréal (Canada)

5.0 Travaux futurs

- Rédaction d'un article benchmark sur les méthodes cross-domain few-shot incluant les différents fine-tuning et les estimations du domain shift.
- Étude des PEFT pour la regression.
- Étudier l'apport des les VLM (Vision/Language Models) sur le cross-domain few-shot.
- Compléter le gros des formations doctorales internes obligatoires et non obligatoires.

Merci pour votre attention

Des questions ?

✉ hicham.talaoubrid1@edu.univ-paris13.fr

References i

-  Chen, Shoufa et al. (2022). "Diffusiondet: Diffusion model for object detection". In: *arXiv preprint arXiv:2211.09788*.
-  Everingham, Mark et al. (2010). "The pascal visual object classes (voc) challenge". In: *International journal of computer vision* 88.2, pp. 303–338.
-  Fu, Yuqian et al. (2024). *Cross-Domain Few-Shot Object Detection via Enhanced Open-Set Object Detector*. arXiv: 2402.03094 [cs.CV]. URL: <https://arxiv.org/abs/2402.03094>.
-  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.
-  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, Yanghao et al. (2022). *Exploring Plain Vision Transformer Backbones for Object Detection*. arXiv: 2203.16527 [cs.CV]. URL: <https://arxiv.org/abs/2203.16527>.
-  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.
-  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, Yanpeng et al. (2022). "Singular Value Fine-tuning: Few-shot Segmentation requires Few-parameters Fine-tuning". In: arXiv: 2206.06122 [cs.CV].
-  Wang, Xin et al. (July 2020). "Frustratingly Simple Few-Shot Object Detection". In: *International Conference on Machine Learning (ICML)*.
-  Wu, Xiongwei, Doyen Sahoo, and Steven Hoi (2020). "Meta-rcnn: Meta learning for few-shot object detection". In: *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 1679–1687.
-  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.
-  Xiong, Wuti (2023). "CD-FSOD: A Benchmark For Cross-Domain Few-Shot Object Detection". In: *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, pp. 1–5.

References ii

-  Zhang, Xinyu, Yuting Wang, and Abdeslam Boularias (2024). *Detect Everything with Few Examples*. arXiv: 2309.12969 [cs.CV]. URL: <https://arxiv.org/abs/2309.12969>.
-  Zhou, Xingyi et al. (2022). *Detecting Twenty-thousand Classes using Image-level Supervision*. arXiv: 2201.02605 [cs.CV]. URL: <https://arxiv.org/abs/2201.02605>.