

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. Introduction**
- 2. Détection d'objet few-shot**
- 3. Challenges de la détection d'objet few-shot sur les images aériennes**
- 4. Méthode PEFT (Parameter Efficient Fine-Tuning)**
- 5. Mon approche**
- 6. Premiers résultats**
- 7. Conclusion**

1.1 Introduction - Principe de la détection d'objet

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 Introduction - Principe de la détection d'objet

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 Introduction - Principe de la détection d'objet

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 Introduction - Principe de la détection d'objet

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) .

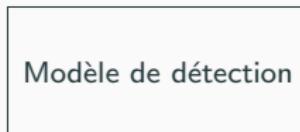


Image d'entrée I

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

1.1 Introduction - Principe de la détection d'objet

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

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

2.1 Détection d'objet few-shot - Principe et définitions

n-way k-shot object detection

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 .

2.1 Détection d'objet few-shot - Principe et définitions

n-way k-shot object detection

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 .



Query image

2.1 Détection d'objet few-shot - Principe et définitions

n-way k-shot object detection

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 .



Query image

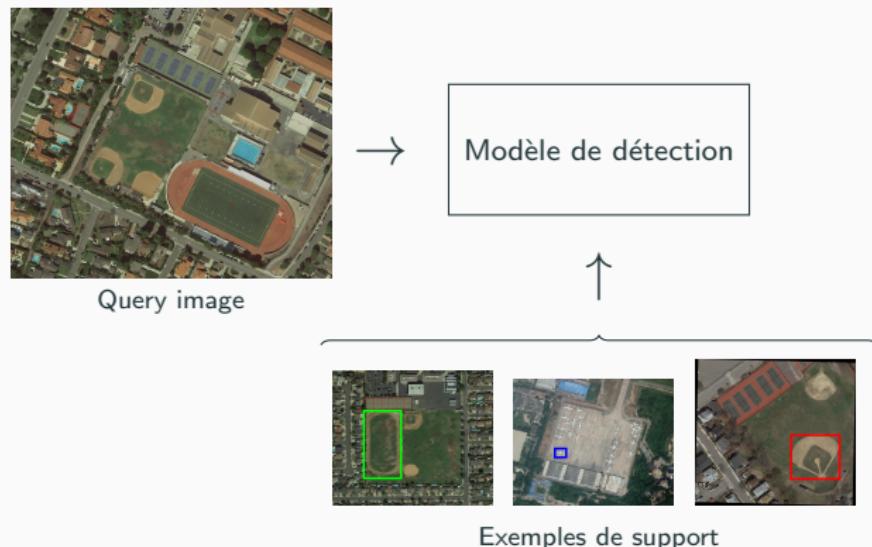


Exemples de support

2.1 Détection d'objet few-shot - Principe et définitions

n-way k-shot object detection

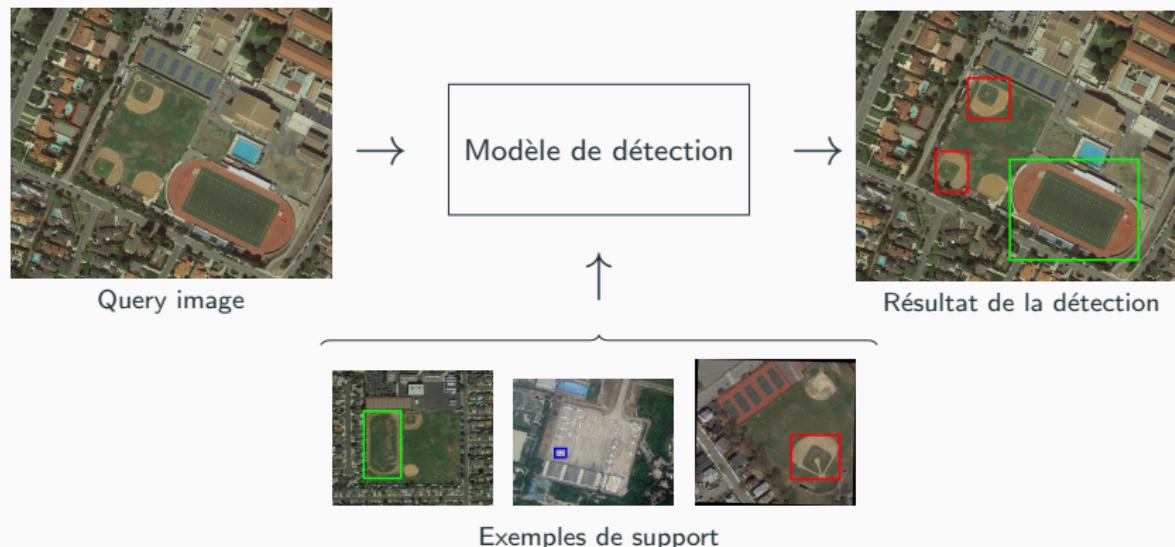
Étant donné des exemples de support $\{(x_1, a_1), \dots, (x_{nk}, a_{nk})\}$, il s'agit de détecter toutes les occurrences des classes dans \mathcal{C} ($|\mathcal{C}| = n$) dans une image de requête x_q .



2.1 Détection d'objet few-shot - Principe et définitions

n-way k-shot object detection

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 .



3.0 Challenges de la détection d'objet few-shot sur les images aériennes

La plupart des méthodes sont évaluées sur des images naturelles : les ensembles de données Pascal VOC et MS COCO. ⇒ cela ne garantit pas de bonnes performances sur les images aériennes.

Les tailles des objets sont extrêmement différentes entre les images aériennes et naturelles.

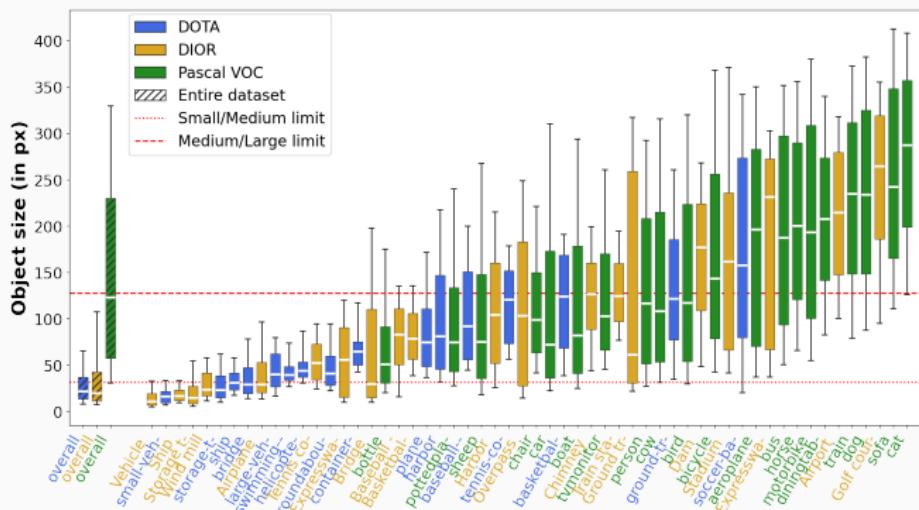


Figure 1: Diagramme en boîte des tailles d'objets dans DOTA (Xia et al. 2018), DIOR (K. Li et al. 2020) et Pascal VOC (Everingham et al. 2010); par classe (**à droite**) et globalement (**à gauche**).

3.0 Challenges de la détection d'objet few-shot sur les images aériennes

Il est impossible de comparer les performances sur différents ensembles de données.

Cependant, il est possible de comparer les performances FSOD par rapport à une référence non few-shot et de comparer cela sur plusieurs ensembles de données.

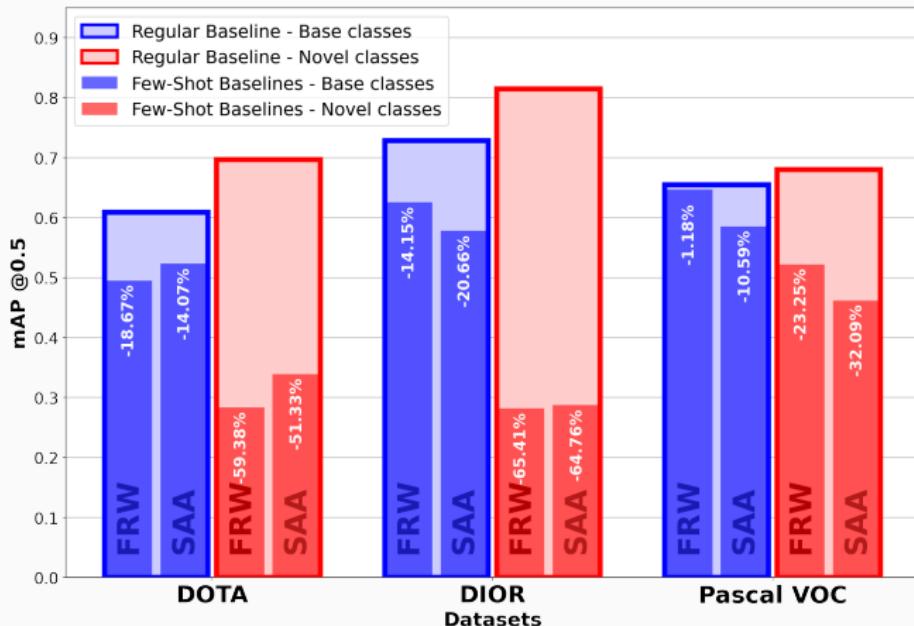


Figure 2: Performances FSOD comparées sur DOTA, DIOR et Pascal VOC. (Le Jeune and Mokraoui 2022)

3.0 Challenges de la détection d'objet few-shot sur les images aériennes

Les grandes différences de taille moyenne entre les classes suggèrent d'analyser les performances par classe. **Corrélation claire entre la taille moyenne des classes et les performances** par rapport à la référence.

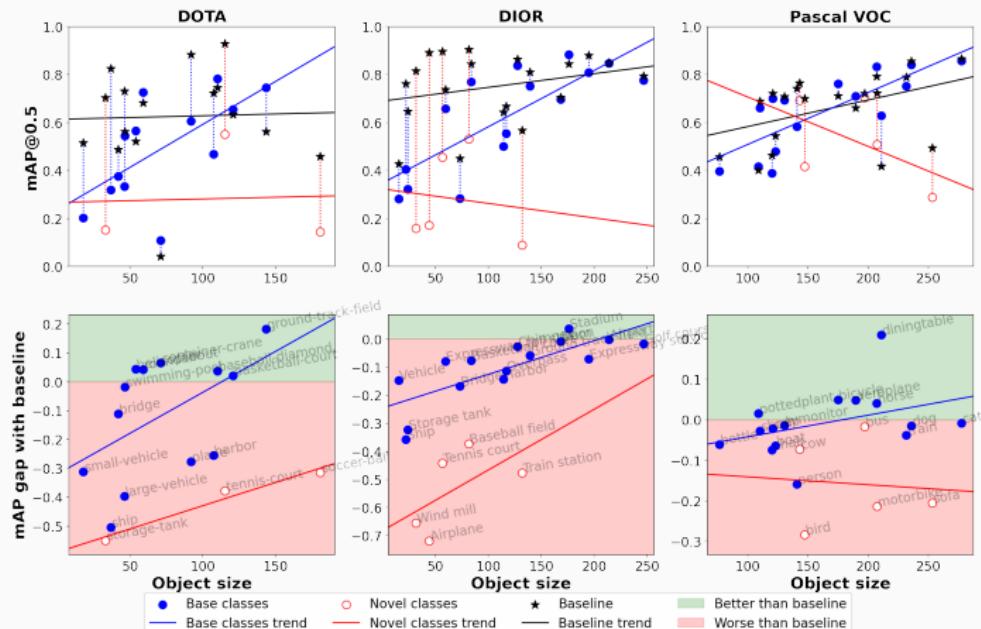
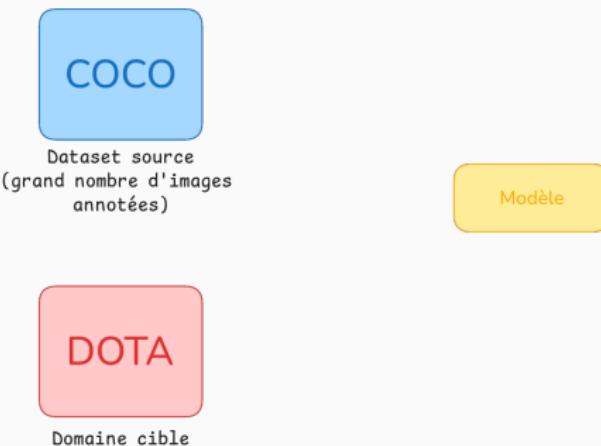


Figure 3: Analyse des performances par classe et comparaison avec la référence non few-shot sur DOTA, DIOR et Pascal VOC. (Le Jeune and Mokraoui 2022)

3.1 Détection d'objet few-shot - Cross-domain

Détection d'objet few-shot cross-domain

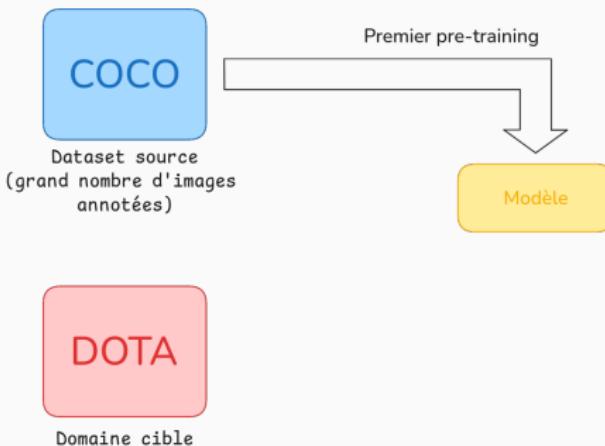
Dans la détection d'objets à faible échantillon inter-domaines (CD-FSOD), deux ensembles de données distincts sont utilisés pendant l'entraînement de base et l'affinage. La CD-FSOD est plus difficile, car le modèle doit non seulement s'adapter à de nouvelles classes, mais aussi à de nouvelles images.



3.1 Détection d'objet few-shot - Cross-domain

Détection d'objet few-shot cross-domain

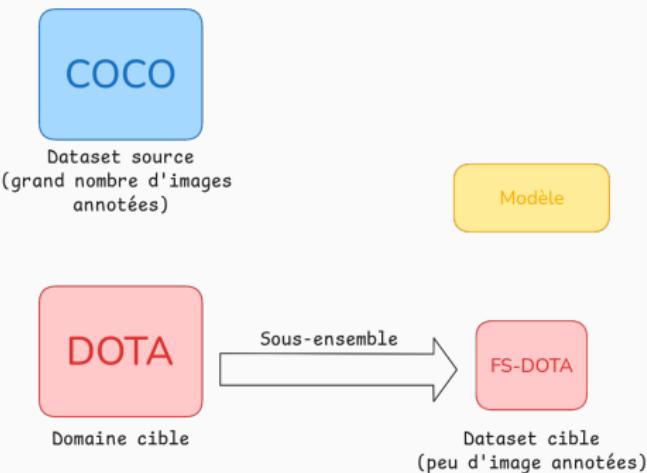
Dans la détection d'objets à faible échantillon inter-domaines (CD-FSOD), deux ensembles de données distincts sont utilisés pendant l'entraînement de base et l'affinage. La CD-FSOD est plus difficile, car le modèle doit non seulement s'adapter à de nouvelles classes, mais aussi à de nouvelles images.



3.1 Détection d'objet few-shot - Cross-domain

Détection d'objet few-shot cross-domain

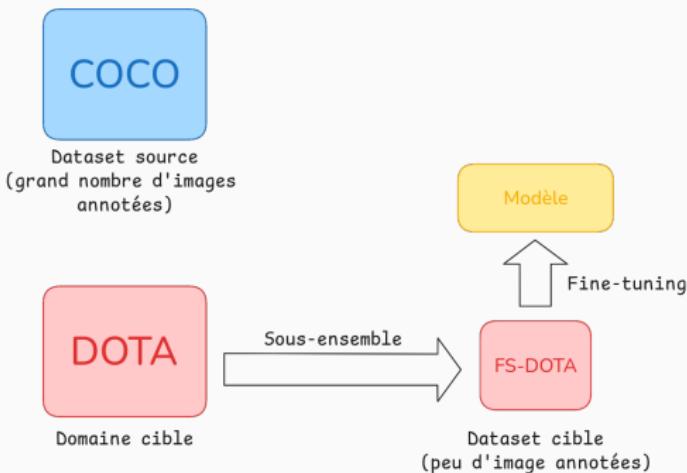
Dans la détection d'objets à faible échantillon inter-domaines (CD-FSOD), deux ensembles de données distincts sont utilisés pendant l'entraînement de base et l'affinage. La CD-FSOD est plus difficile, car le modèle doit non seulement s'adapter à de nouvelles classes, mais aussi à de nouvelles images.



3.1 Détection d'objet few-shot - Cross-domain

Détection d'objet few-shot cross-domain

Dans la détection d'objets à faible échantillon inter-domaines (CD-FSOD), deux ensembles de données distincts sont utilisés pendant l'entraînement de base et l'affinage. La CD-FSOD est plus difficile, car le modèle doit non seulement s'adapter à de nouvelles classes, mais aussi à de nouvelles images.



3.1 Détection d'objet few-shot - 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 Bouali 2024)	ViT-L/14
DE-ViT-FT (Zhang, Y. Wang, and Bouali 2024)	ViT-L/14
CD-VITO (Fu et al. 2024)	ViT-L/14

Table 1: Etats de l'art des méthodes de détection d'objet cross-domain few-shot.

3.1 Détection d'objet few-shot - Cross-domain

	DOTA	DIOR
Few-shot	60.45	57.51
Cross-domain Few-shot (COCO)	25.02	38.73
Cross-domain Few-shot (DOTA)	-	38.44
Cross-domain Few-shot (DIOR)	33.30	-

Table 2: Comparaison des performances de DiffusionDet (Chen et al. 2022) en Few-shot et en Cross-domain Few-shot sur DOTA et DIOR en considérant 3 dataset source (notés entre parenthèse) avec 10 shots.

3.2 Premiers résultats - Importance du dataset source

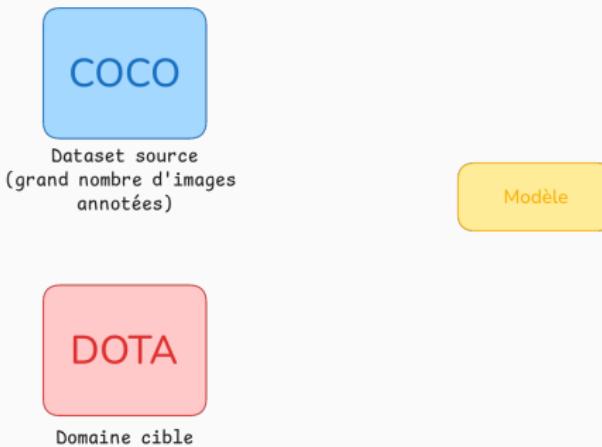
Importance of source dataset in cross-domain

Le choix du dataset source dans la détection d'objets cross-domain est crucial, car il affecte directement la capacité du modèle à se généraliser à de nouveaux domaines cibles non vus.

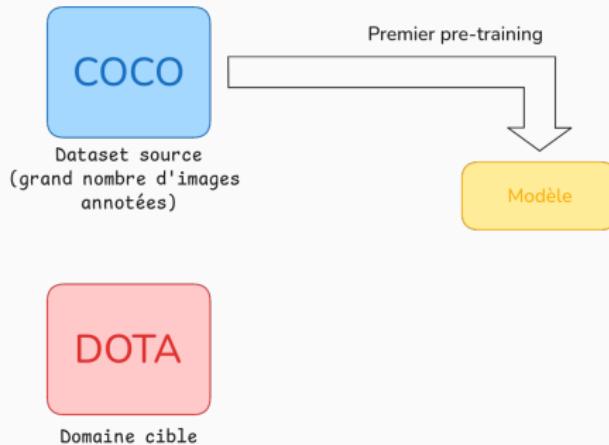
k shots	DOTA → DIOR	COCO → DIOR	DIOR → DOTA	COCO → DOTA
1	20.18	11.10	5.41	4.03
5	34.43	30.42	25.88	14.45
10	41.48	38.73	31.99	25.02
20	49	48.23	38.77	33.31
50	54.07	56.97	44.07	43.23

Table 3: Performance comparison of FSDiffusionDet using different source datasets in a cross-domain setting. Results demonstrate the impact of source dataset diversity and relevance on target domain accuracy.

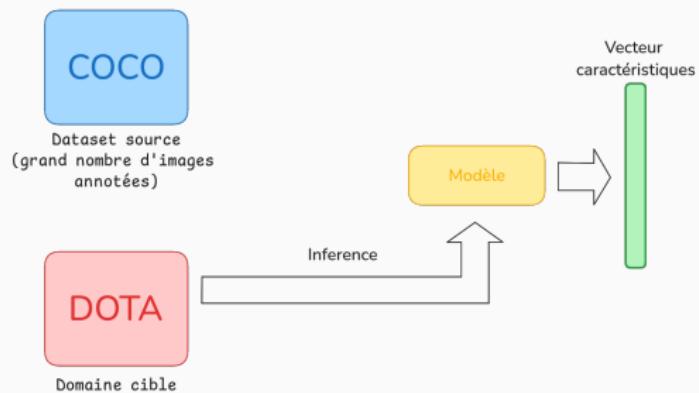
3.3 Premiers résultats - Variances inter-classes et intra-classes



3.3 Premiers résultats - Variances inter-classes et intra-classes



3.3 Premiers résultats - Variances inter-classes et intra-classes



3.3 Premiers résultats - Variances inter-classes et intra-classes

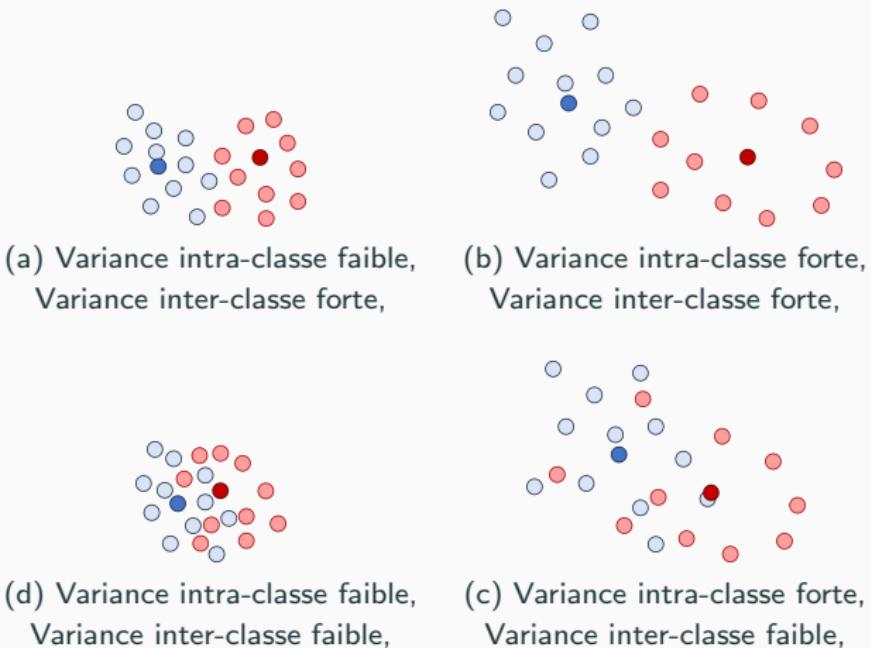
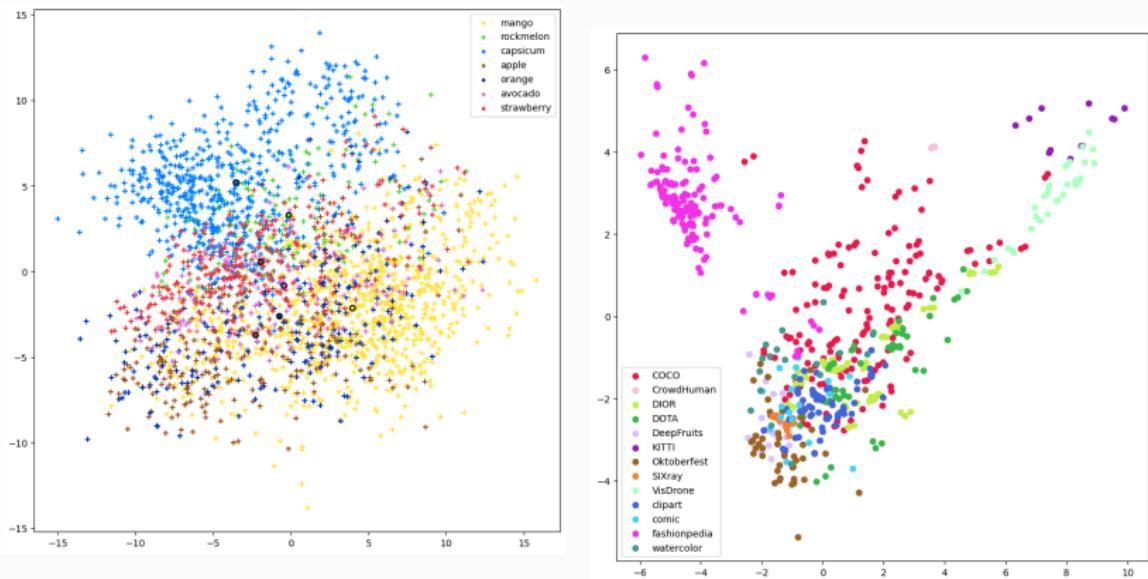


Figure 4: Illustration conceptuelle de la séparabilité des classes en relation avec les variances inter-classes et intra-classes. Les projections dans l'espace latent des images d'une classe sont représentées par des points bleus, tandis que les projections d'une autre classe sont marquées en rouge, avec les barycentres de chaque classe indiqués par des points plus foncés.

3.3 Premiers résultats - Variances inter-classes et intra-classes



(a) Représentation 2D de la projection des objets de DeepFruits après un PCA avec les barycentres mis en valeur avec des cercles de couleurs.

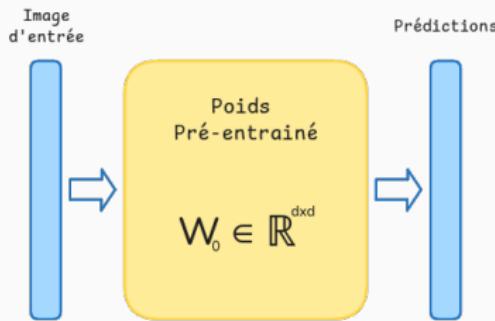
(b) Représentation 2D des projections des barycentres de chaque classe de chaque datasets après un PCA.

Figure 5: Représentation 2D des projections (a) des objets du dataset DeepFruits et (b) des barycentres de chaque classe après un PCA.

4.2 Méthode PEFT - LoRA Low Rank Adaptation

LoRA (Low-Rank Adaptation)

LoRA (Low-Rank Adaptation) est une technique d'optimisation qui ajuste efficacement les modèles de réseaux neuronaux de grande taille en réduisant le nombre de paramètres à ajuster via la factorisation des matrices de poids. En limitant le nombre de paramètres modifiés, elle aide également à réduire le risque d'overfitting tout en maintenant de bonnes performances.



4.2 Méthode PEFT - LoRA Low Rank Adaptation

LoRA (Low-Rank Adaptation)

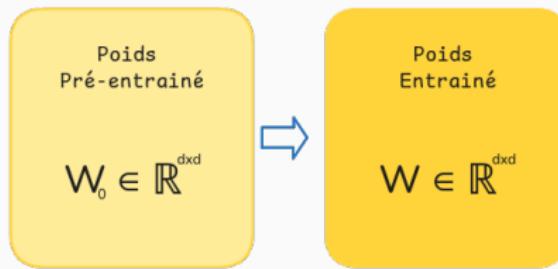
LoRA (Low-Rank Adaptation) est une technique d'optimisation qui ajuste efficacement les modèles de réseaux neuronaux de grande taille en réduisant le nombre de paramètres à ajuster via la factorisation des matrices de poids. En limitant le nombre de paramètres modifiés, elle aide également à réduire le risque d'overfitting tout en maintenant de bonnes performances.



4.2 Méthode PEFT - LoRA Low Rank Adaptation

LoRA (Low-Rank Adaptation)

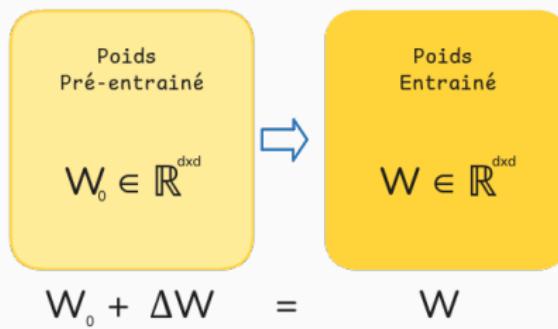
LoRA (Low-Rank Adaptation) est une technique d'optimisation qui ajuste efficacement les modèles de réseaux neuronaux de grande taille en réduisant le nombre de paramètres à ajuster via la factorisation des matrices de poids. En limitant le nombre de paramètres modifiés, elle aide également à réduire le risque d'overfitting tout en maintenant de bonnes performances.



4.2 Méthode PEFT - LoRA Low Rank Adaptation

LoRA (Low-Rank Adaptation)

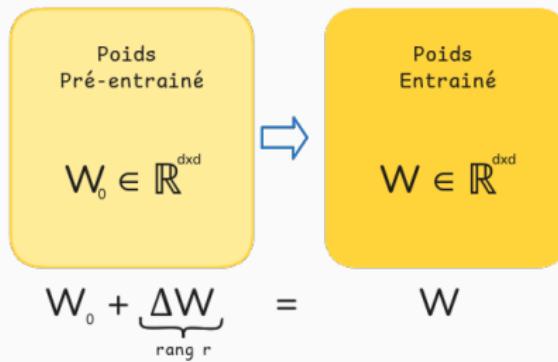
LoRA (Low-Rank Adaptation) est une technique d'optimisation qui ajuste efficacement les modèles de réseaux neuronaux de grande taille en réduisant le nombre de paramètres à ajuster via la factorisation des matrices de poids. En limitant le nombre de paramètres modifiés, elle aide également à réduire le risque d'overfitting tout en maintenant de bonnes performances.



4.2 Méthode PEFT - LoRA Low Rank Adaptation

LoRA (Low-Rank Adaptation)

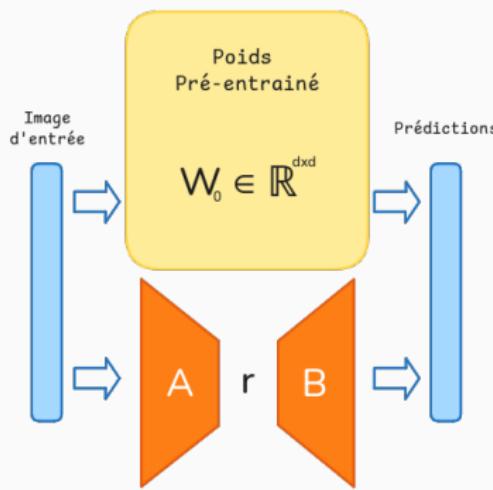
LoRA (Low-Rank Adaptation) est une technique d'optimisation qui ajuste efficacement les modèles de réseaux neuronaux de grande taille en réduisant le nombre de paramètres à ajuster via la factorisation des matrices de poids. En limitant le nombre de paramètres modifiés, elle aide également à réduire le risque d'overfitting tout en maintenant de bonnes performances.



4.2 Méthode PEFT - LoRA Low Rank Adaptation

LoRA (Low-Rank Adaptation)

LoRA (Low-Rank Adaptation) est une technique d'optimisation qui ajuste efficacement les modèles de réseaux neuronaux de grande taille en réduisant le nombre de paramètres à ajuster via la factorisation des matrices de poids. En limitant le nombre de paramètres modifiés, elle aide également à réduire le risque d'overfitting tout en maintenant de bonnes performances.



4.3 Méthode PEFT - SVF Singular Value Fine-tuning

SVF (Singular Value Fine-tuning)

SVF (Singular Value Fine-tuning) est une technique d'ajustement des modèles de réseaux neuronaux qui réduit la complexité en décomposant les matrices de poids en valeurs singulières. En ne modifiant que les valeurs singulières les plus importantes, elle permet d'améliorer l'efficacité tout en limitant l'overfitting, en évitant d'ajuster un trop grand nombre de paramètres.

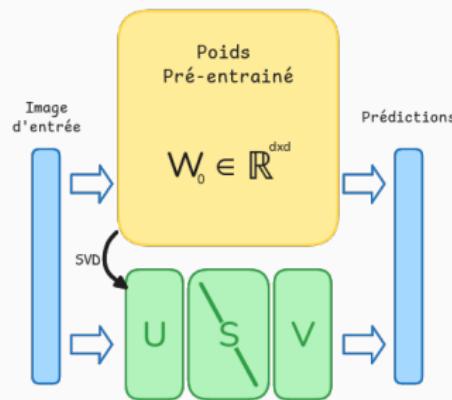


Figure 6: Principe de l'entraînement SVF. Ici on entraîne seulement les valeurs propres (Y. Sun et al. 2022).

4.4 Premiers résultats - Overfitting pour la regression

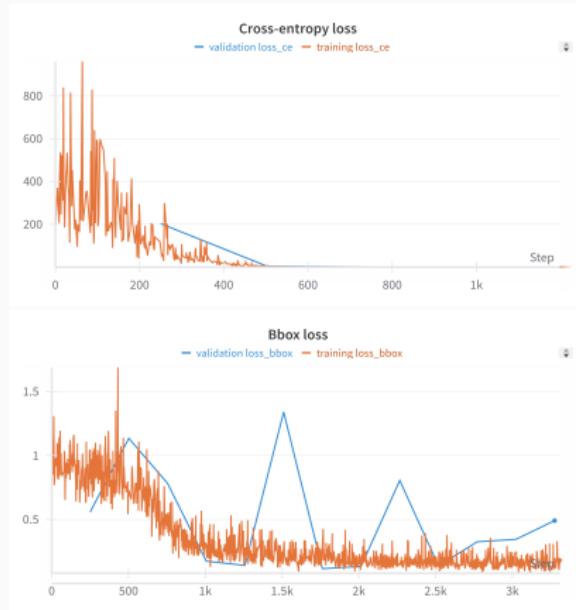


Figure 7: Impact de LoRA sur l'entraînement de FSDiffusionDet sur DIOR en utilisant 10 shots et un rang de 16.

4.4 Premiers résultats - Overfitting pour la regression

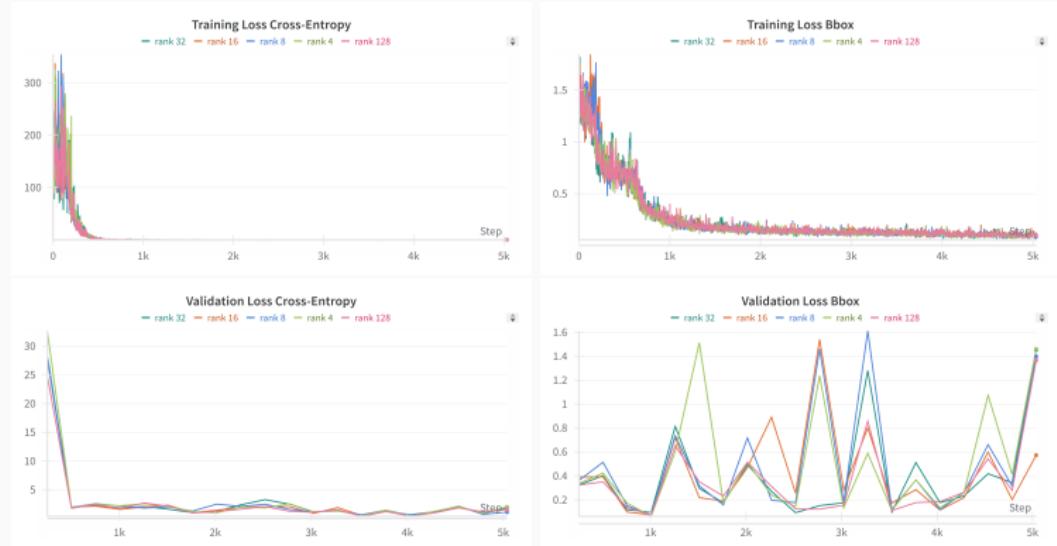


Figure 8: Impact du rang de LoRA sur l'entraînement de FSDiffusionDet sur DIOR en utilisant 10 shots.

6.1 Mon approche - Vu d'ensemble de mes publications

Articles de conférence acceptés (avant la thèse) :

- Straightforward Adaptation of Particle Filter to Fish Eye Images for Top View Pedestrian Tracking – publié à **ICASSP** (14 avr. 2024)
- 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, participation aux expériences et présentation)

Article de journal accepté (avant la thèse) :

- Human tracking in top-view fisheye images: Analysis of familiar similarity measures via HOG and against various color spaces – publié à **Journal of Imaging** (16 avr. 2022).

Articles de conférence soumis

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

6.2 Mon approche - Travaux future

- Expérimentation pour aider au choix du domaine source en cross domain
 - objectif octobre 2024
- Travail sur un framework facilitant l'expérimentation en cross domain
 - objectif octobre 2024
- Utiliser le framework pour réaliser un article benchmark des méthodes cross domain
 - objectif novembre 2024
- Travaux sur des PEFT qui pourrait fonctionner pour la regression
 - objectif début 2025
- Compléter le papier du benchmark pour l'inclure dans un journal
 - objectif début 2025
- Travaux sur les VLM (Vision/Language Models)
 - objectif courant 2025

7.0 Conclusion

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>.