

Apprentissage Actif Pour La Classification d'Images en Few-Shot

Aymane ABDALI^{1,2} Bartosz BOGUSLAWSKI¹ Lucas DRUMETZ² Vincent GRIPON²

¹Schneider Electric

²IMT Atlantique, UMR CNRS 6285, Lab-STICC, F-29238 Brest, France

Résumé – Nous introduisons le problème de la classification active en Few-Shot (Active Few-Shot Classification, AFSC) où l'objectif est de classifier un petit ensemble de données initialement non étiquetées, compte tenu d'un budget d'étiquetage très limité. Ce problème peut être considéré comme un paradigme rival de la classification Few-Shot transductive classique (Transductive Few-Shot Classification, TFSC), car ces deux approches sont applicables dans des conditions similaires. Nous proposons d'abord une méthodologie qui combine l'inférence statistique et une stratégie d'apprentissage actif à deux niveaux originale qui s'insère bien dans ce cadre. Nous adaptons ensuite plusieurs benchmarks de référence en vision dans le cadre TFSC. Nos expériences montrent que les avantages potentiels de l'AFSC peuvent être substantiels, avec des gains en précision moyenne pondérée allant jusqu'à 10% par rapport aux méthodes TFSC de pointe pour le même budget d'étiquetage. Nous estimons que ce nouveau paradigme pourrait conduire à de nouveaux développements et standards dans les paramètres d'apprentissage à faible disponibilité de données.

Abstract – We introduce the problem of Active Few-Shot Classification (AFSC) where the objective is to classify a small, initially unlabeled, dataset given a very restrained labeling budget. This problem can be seen as a rival paradigm to classical Transductive Few-Shot Classification (TFSC), as both these approaches are applicable in similar conditions. We first propose a methodology that combines statistical inference, and an original two-tier active learning strategy that fits well into this framework. We then adapt several standard vision benchmarks from the field of TFSC. Our experiments show the potential benefits of AFSC can be substantial, with gains in average weighted accuracy of up to 10% compared to state-of-the-art TFSC methods for the same labeling budget. We believe this new paradigm could lead to new developments and standards in data-scarce learning settings.

1 Introduction

Nous considérons le problème de la classification d'un ensemble d'échantillons de données sans ensemble d'entraînement donné, mais avec un budget d'étiquetage –très– limité. Un tel problème est susceptible de se poser dans des contextes où les échantillons de données sont disponibles mais leur étiquetage est coûteux. Ces contextes sont souvent rencontrés dans des applications industrielles.

Lorsque le nombre d'échantillons de données est réduit, le domaine de la classification Transductive Few-Shot (Transductive Few-Shot Classification, TFSC) comprend de nombreuses solutions qui peuvent atteindre une grande précision. Le principe consiste en l'utilisation d'un grand ensemble de données génériques pour entraîner des extracteurs de caractéristiques efficaces. Dans le contexte de référence en TFSC, les échantillons étiquetés (labels) de la tâche Few-Shot sont généralement considérés comme étant tirés uniformément au hasard.

Le domaine de l'apprentissage actif a proposé de nombreuses solutions. Cependant, l'adaptation de l'apprentissage actif au régime de faible données, en particulier pour la vision, est peu explorée et notre motivation principale est de considérer l'apprentissage actif non pas comme un cadre d'apprentissage difficile, mais comme un outil à utiliser lors de la construction d'un ensemble de données étiquetées et de prouver que cela peut avoir un impact considérable sur les résultats.

Dans cet article, nous introduisons donc une formulation du problème de classification active en Few-Shot mentionné ci-dessus (Active Few-Shot Classification, AFSC). Nous proposons une méthodologie combinant l'inférence statistique

avec des heuristiques, puis introduisons des benchmarks de référence en adaptant ceux du domaine de la TFSC.

2 Travaux Connexes

2.1 Apprentissage avec Peu de Données (Few-Shot)

Le problème d'apprentissage en Few-Shot learning attire de plus en plus d'attention au cours des dernières années. Différentes branches sont explorées notamment le meta-learning [17] et l'apprentissage de métriques [26]. Ces méthodes nécessitent l'utilisation d'un modèle pré-entraîné, et les avancées dans ce domaine ont rendu disponibles une myriade d'extracteurs de caractéristiques efficacement entraînés avec diverses techniques [12, 2]. La classification dans le cadre inductif s'appuie principalement sur des méthodes simples [2], tandis que des méthodes plus diverses ont été utilisées dans le cadre transductif permettant des résultats plus prometteurs [12, 13, 4]. Les auteurs de [22] ont également remis en question les hypothèses de l'équilibre des classes dans ce cadre et son importance dans la récente ligne de travaux. Notre travail s'inscrit dans cette reconsidération de la configuration transductive vers une approche plus agnostique à la distribution de classes et de labels.

Notre travail est également lié aux techniques d'auto-supervision [29], car l'aspect Cold-start, c'est-à-dire partir de zéro exemples, est crucial pour nous et nous utilisons des méthodes non supervisées tant que nous n'avons pas d'exemples étiquetés disponibles.

2.2 Apprentissage Actif

Une partie majeure de ce document tourne autour de l'augmentation significative des scores que nous pouvons observer si nous sélectionnons les bons exemples à annoter, ainsi que de l'exploration des heuristiques que nous avons envisagées pour effectuer cette sélection. Une grande partie de la littérature actuelle sur l'apprentissage actif n'est pas adaptée aux scénarios Few-Shot car la plupart des techniques utilisées qui ont des bases théoriques solides dépendent fortement des modèles utilisés. Ces techniques cherchent généralement à réduire la variance du modèle ou à baser la sélection sur l'incertitude de ce dernier [7, 18]. Ces techniques deviennent difficiles dans un contexte "Cold-start" [10, 28] où nous partons de zéro exemples étiquetés. Avoir trop peu d'échantillons et/ou un budget d'étiquetage rend cela encore plus difficile car nous ne pouvons pas nous permettre de subir une phase initiale d'étiquetage non pertinent basé sur un modèle instable. Cependant, certains documents récents ont montré que dans le traitement du langage naturel, les modèles pré-entraînés peuvent être utilisés pour surmonter ce défi [15]. Dans notre configuration Few-Shot, nous utilisons également les extracteurs de caractéristiques pour obtenir des distributions de données plus proches des gaussiennes multimodales. Cela nous permet de mélanger des méthodes d'échantillonnage de diversité et des méthodes d'échantillonnage basées sur l'incertitude [27, 23] pour obtenir un pipeline d'étiquetage efficace.

3 Problématique

Nous supposons que nous disposons d'un extracteur de caractéristiques et d'un petit ensemble de données composé de N échantillons appartenant à K nouvelles classes. Ces classes sont disjointes de celles de l'ensemble de données utilisé pour l'entraînement de l'extracteur de caractéristiques. La tâche de classification active en Few-Shot est de prédire la classe de chaque échantillon de \mathcal{N} après avoir révélé l'annotation d'au plus $\ell \leq N$ échantillons. Les performances sont mesurées avec une précision pondérée sur tous les échantillons, y compris ceux étiquetés.

Pour simplifier, nous appelons ce problème une tâche à K classes, ℓ labels, N échantillons. La différence majeure avec le cadre TFSC réside dans le fait que les échantillons étiquetés sont choisis plutôt que tirés au hasard. Cela donne un nombre variable de labels par classe, y compris zéro pour certaines d'entre elles dans le pire des cas.

3.1 Méthodologie

La méthode proposée repose sur plusieurs étapes. La première consiste à entraîner un extracteur de caractéristiques pour transformer les images brutes en vecteurs de caractéristiques de haute dimension. Nous effectuons plusieurs étapes de pré-traitement sur les vecteurs de caractéristiques obtenus. Ensuite, nous déduisons une distribution de probabilité sur tous les échantillons en utilisant une méthodologie d'espérance-maximisation (EM). Enfin, nous proposons une procédure d'apprentissage actif séquentiel que nous détaillons dans la suite.

3.2 Extraction de caractéristiques et pré-traitement

Dans cet article, nous utilisons des modèles prêts à l'emploi pour générer les vecteurs de caractéristiques, que nous normalisons ensuite (moyenne zéro et norme unitaire). Enfin, nous les lissions en représentant les données sous forme de graphe et en propageant les caractéristiques des sommets à leurs voisins les plus proches, comme décrit dans [11, 24].

3.3 Inférence Statistique

Soit \mathbf{z}_i le i -ème vecteur de caractéristiques pré-traité correspondant au i -ème échantillon dans \mathcal{N} . Nous considérons un modèle simple où, pour chaque classe k , la distribution de $\mathbf{z}/y = k$ suit une loi normale multivariée avec une moyenne μ_k et une covariance Σ_k . Si nous considérons une distribution isotrope, Σ_k devient $\sigma_k^2 \mathbf{I}$ où $\sigma_k \in \mathbb{R}^+$ et \mathbf{I} est la matrice identité. Nous utilisons un algorithme standard de soft- K -means au vu de son efficacité pour l'apprentissage transductif [2].

3.4 Sélection Active Proposée

Après l'inférence statistique, nous voulons sélectionner les échantillons les plus informatifs pour l'étiquetage. Les solutions classiques reposent sur des critères [3] qui mesurent la confiance comme la marge entre les probabilités ou les distances aux centroïdes des clusters.

Au lieu d'utiliser ces approches classiques, nous proposons un nouveau critère basé sur le rapport des densités logarithmiques que nous appelons **Log-probs Soft K-means Sampling (LSS)**. En d'autres termes, nous considérons que l'échantillon qui a le plus bas **rapport de logarithme de probabilité (lpr)** d'appartenance à son cluster k plutôt qu'à d'autres clusters est le plus confiant.

Le logarithme de la densité de $\mathbf{z}/y=k$ est

$$\log p(\mathbf{z}/y=k) = \log \left(\frac{1}{\sqrt{2\pi}\sigma_k} \right) - \frac{1}{2\sigma_k} \|\mathbf{z} - \mu_k\|^2$$

Nous définissons le lpr comme le rapport des logarithmes de densités sur les différentes distributions de caractéristiques des labels :

$$lpr_k(\mathbf{z}) = \frac{\log p(\mathbf{z}/y=k)}{\sum_{i=1}^K \log p(\mathbf{z}/y=i)}$$

Pour $a_k = \log \left(\frac{1}{\sqrt{2\pi}\sigma_k} \right)$ et $b_k = 1/2\sigma_k$ Le lpr d'un échantillon \mathbf{z} appartenant au cluster k est :

$$lpr_k(\mathbf{z}) = \frac{a_k - b_k \|\mathbf{z} - \mu_k\|^2}{\sum_{i=1}^K (a_i - b_i \|\mathbf{z} - \mu_i\|^2)}$$

Nous estimons la moyenne $\hat{\mu}_k$ et l'écart-type $\hat{\sigma}_k$ de la distribution des caractéristiques de classe en utilisant le clustering.

En utilisant les logarithmes des densités au lieu des densités elles-mêmes, les calculs sont plus robustes aux estimations. Sous l'hypothèse d'une déviation standard constante

pour toutes les classes. $\sigma = 1/\sqrt{2\pi}$ Le ratio devient :

$$lpr_k(\mathbf{z}) = \frac{\|\mathbf{z} - \mu_k\|^2}{\sum_{i=1}^K \|\mathbf{z} - \mu_i\|^2}.$$

Nous utilisons une pipeline séquentielle naïve pour la sélection active. Cette pipeline comprend plusieurs "rounds" où l'on sélectionne K labels i.e un label suivant chacun des K clusters (c'est-à-dire chaque lpr_k). Le premier round consiste à sélectionner les échantillons les plus confiants tandis que dans les rounds suivants, nous étiquetons les échantillons les moins confiants. Cette approche combine l'échantillonnage de diversité, en échantillonnant des exemples de chaque cluster de manière à assurer la représentativité de la classe, et l'échantillonnage d'incertitude, en sélectionnant des exemples de faible confiance dans la plupart des rounds.

4 Expériences

4.1 Datasets

Nous adaptons deux benchmarks standard à AFSC : mini-ImageNet et tiered-ImageNet. Pour des raisons de concision et étant donné que ces benchmarks sont couramment utilisés dans le domaine de TFSC, nous omettons leur description détaillée ici. Veuillez vous référer à la version longue [1] pour plus de détails.

4.2 Protocole de Benchmarks et Comparaisons

Nos paramètres par défaut sont des ensembles d'échantillons de taille $75 + \ell$ avec $\ell = 5, 25$ pour des tâches 5-classes. Comme décrit dans [22], les classes des ensembles d'échantillons sont distribuées de manière aléatoire selon la distribution de Dirichlet avec des paramètres $\alpha \mathbf{1}$, où $\mathbf{1}$ est le vecteur unité constant avec K coordonnées et $\alpha = 2$. Nous évaluons les méthodes en calculant le score moyen sur 10 000 tâches [4]. Nous avons choisi d'utiliser la précision pondérée pour mieux prendre en compte le déséquilibre entre classes.

Pour nos expériences, nous utilisons deux extracteurs de caractéristiques différents. Le premier est un ResNet-18 [4, 22] et le second un ResNet-12 [2]. Les deux permettent généralement d'atteindre des performances proches de l'état de l'art dans les benchmarks TFSC.

4.3 Résultats et Discussions

4.3.1 Comparaison entre AFSC et TFSC

Dans la première expérience, nous avons cherché à quantifier l'intérêt de l'utilisation de la classification active par rapport à l'approche habituelle de TFSC. Dans le tableau 1, nous avons rapporté les résultats qui comprennent des méthodes utilisées en transductif ainsi que les résultats de notre méthode d'inférence statistique dans la configuration standard TFSC que nous appelons "labels égaux par classe". Nous comparons ensuite les performances lorsque nous sélectionnons des labels au hasard et lorsque nous utilisons la méthode active proposée. Comme nous avons utilisé les résultats de [22], nous rapportons ici, exceptionnellement, la précision moyenne non pondérée au lieu de la pondérée. Nous avons également mis à

TABLE 1 : Comparaison de la précision entre les méthodes TFSC de pointe et la méthode AFSC proposée sur mini-ImageNet et tiered-ImageNet. Toutes les valeurs sont en pourcentage.

| | | mini-ImageNet tiered-Imagenet | | | |
|----------------------------------------|--------------------|---------------------------------|-------------|-------------|-------------|
| Configuration des labels | méthode | 5-labels | 25-labels | 50-labels | 100-labels |
| Labels égaux par classe (aléatoire) | PT-MAP [12] | 62.6 63.6 | 75.3 77.5 | 81.3 83.1 | 70.4 88.6 |
| | LaplacianShot [30] | 67.6 74.0 | 86.2 89.3 | 90.5 92.8 | 94.0 95.3 |
| | BD-CSPN [13] | 69.0 75.8 | 85.1 88.7 | 89.7 92.0 | 93.4 94.8 |
| | TIM [4] | 69.3 75.8 | 84.5 88.1 | 89.4 91.2 | 93.2 94.6 |
| | α -TIM [22] | 69.4 76.0 | 86.9 89.9 | 91.6 93.6 | 94.8 96.1 |
| | soft K-means | 70.7 74.7 | 83.4 84.8 | 87.7 88.8 | 91.8 92.6 |
| Sélection aléatoire de labels | soft K-means | 64.3 66.9 | 84.7 85.7 | 89.6 90.0 | 93.3 93.5 |
| Sélection active de labels | LSS (ours) | 73.5 76.1 | 87.7 87.5 | 93.8 92.2 | 96.8 95.8 |

jour les valeurs de [22] pour tenir compte de notre mesure de précision où les exemples étiquetés sont pris en compte. Nous remarquons une nette augmentation de précision lorsque nous optons pour une sélection active d'étiquettes plutôt qu'aléatoire. Dans le tableau 1, nous pouvons clairement voir que, dans des conditions similaires, notre échantillonnage actif permet d'obtenir de meilleurs scores par rapport à une sélection de labels uniforme par classe. Nous observons également que, à mesure que le nombre de labels augmente, l'écart entre les scores de sélection aléatoire et active se réduit. Cela est en partie dû à la manière de calculer la précision dans notre formulation du problème, car nous incluons les labels dans le calcul de la précision pondérée. Un autre effet intéressant dans le tableau 1 est que la configuration de sélection aléatoire de labels finit par dépasser la configuration de labels égaux par classe à mesure que le nombre de labels augmente, ce qui met en évidence que le choix de labels uniformes à partir d'une distribution déséquilibrée n'est pas nécessairement optimal.

4.3.2 Effets de la Sélection Active

Nous approfondissons l'effet du critère actif dans le Tableau 2. Pour les deux ensembles de données, nous comparons les performances des critères d'échantillonnage à une sélection aléatoire, ainsi qu'à un oracle dans le cas de 5 labels. Nous calculons l'oracle en exécutant une méthode Nearest Class Mean sur toutes les combinaisons possibles de 5 étiquettes parmi les 80 échantillons totaux. Pour cet oracle, nous ne calculons que 1000 exécutions aléatoires pour mini-ImageNet et 100 pour tiered-ImageNet. Cela met en évidence le potentiel de performance qui peut être atteint. De plus, nous obtenons déjà des scores comparables à la configuration standard de l'état de l'art avec des heuristiques simples. Nous pouvons constater que la marge et la K-medoid fonctionnent bien au-delà de la sélection aléatoire et que la méthode LSS proposée dépasse légèrement ces critères.

5 Conclusion

Nous avons introduit une nouvelle formulation pour le problème de la classification active en Few-Shot (AFSC). Ce problème peut être vu comme un nouveau paradigme rival de la classification Few-Shot transductive (TFSC). Nous avons proposé une méthodologie simple qui repose sur deux étapes principales : une méthode d'inférence statistique et une étape de sélection active dans laquelle nous sélectionnons des échantillons à étiqueter (labels) en fonction de la méthode Log-probs

TABLE 2 : Comparaison des critères de sélection active sur les jeux de données considérés. Nous ajoutons une stratégie oracle qui consiste en la sélection qui donne les meilleures performances. Toutes les valeurs sont en pourcentage.

| labels ℓ | stratégie active | mini-ImageNet | tiered-ImageNet |
|---------------|------------------|-----------------------------------|-----------------------------------|
| 5 labels | aléatoire | 62.2 \pm 0.31 | 61.2 \pm 0.30 |
| | oracle | 95.2 \pm 0.21 | 97.4 \pm 0.3 |
| | marge | 69.0 \pm 0.30 | 69.0 \pm 0.26 |
| | K-medoid | 75.6 \pm 0.28 | 75.2 \pm 0.27 |
| | LSS | 76.1 \pm 0.26 | 75.7 \pm 0.27 |

Soft K-Means Sampling. Nous avons adapté des benchmarks du domaine de la TFSC et nous avons prouvé que le cadre proposé peut atteindre de nouveaux niveaux de précision par rapport à la TFSC. Nous nous attendons à ce que l’AFSC soit applicable à un large éventail d’applications du monde réel et, en tant que tel, définisse un nouveau cadre prometteur pour les contributions futures.

6 Remerciements

Ce travail a été réalisé dans le cadre d’un programme de doctorat CIFRE soutenu par Schneider Electric et ANRT (Association Nationale Recherche Technologie).

Références

- [1] Aymane ABDALI, Vincent GRIPON, Lucas DRUMETZ et Bartosz BOGUSLAWSKI : Active few-shot classification : a new paradigm for data-scarce learning settings, 2022.
- [2] Yassir BENDOU, Yuqing HU, Raphael LAFARGUE, Giulia LIOI, Bastien PASDELOUP, Stéphane PATEUX et Vincent GRIPON : Easy : Ensemble augmented-shot y-shaped learning : State-of-the-art few-shot classification with simple ingredients, 2022.
- [3] Rinu BONEY et Alexander ILIN : Semi-supervised few-shot learning with prototypical networks. *CoRR abs/1711.10856*, 2017.
- [4] Malik BOUDIAF, Imtiaz ZIKO, Jérôme RONY, José DOLZ, Pablo PIAN-TANIDA et Ismail BEN AYED : Information maximization for few-shot learning. *Advances in Neural Information Processing Systems*, 33:2445–2457, 2020.
- [5] Da CHEN, Yuefeng CHEN, Yuhong LI, Feng MAO, Yuan HE et Hui XUE : Self-supervised learning for few-shot image classification. *In ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1745–1749. IEEE, 2021.
- [6] Ting CHEN, Simon KORNBLITH, Mohammad NOROUZI et Geoffrey HINTON : A simple framework for contrastive learning of visual representations. *In International conference on machine learning*, pages 1597–1607. PMLR, 2020.
- [7] David A COHN, Zoubin GHAMRANI et Michael I JORDAN : Active learning with statistical models. *Journal of artificial intelligence research*, 4:129–145, 1996.
- [8] Yingjie GU, Zhong JIN et Steve C CHIU : Active learning combining uncertainty and diversity for multi-class image classification. *IET Computer Vision*, 9(3):400–407, 2015.
- [9] Ruiting HOU, Hong CHANG, Bingpeng MA, Shiguang SHAN et Xilin CHEN : Cross attention network for few-shot classification. *Advances in Neural Information Processing Systems*, 32, 2019.
- [10] Neil HOULSBY, José Miguel HERNÁNDEZ-LOBATO et Zoubin GHAMRANI : Cold-start active learning with robust ordinal matrix factorization. *In International Conference on Machine Learning*, pages 766–774. PMLR, 2014.
- [11] Yuqing HU, Vincent GRIPON et Stéphane PATEUX : Graph-based interpolation of feature vectors for accurate few-shot classification. *In 2020 25th International Conference on Pattern Recognition (ICPR)*, pages 8164–8171. IEEE, 2021.
- [12] Yuqing HU, Vincent GRIPON et Stéphane PATEUX : Leveraging the feature distribution in transfer-based few-shot learning. *In International Conference on Artificial Neural Networks*, pages 487–499. Springer, 2021.
- [13] Jinlu LIU, Liang SONG et Yongqiang QIN : Prototype rectification for few-shot learning. *In European Conference on Computer Vision*, pages 741–756. Springer, 2020.
- [14] Yanbin LIU, Juho LEE, Minseop PARK, Saehoon KIM et Yi YANG : Transductive propagation network for few-shot learning. 2018.
- [15] Thomas MÜLLER, Guillermo PÉREZ-TORRÓ, Angelo BASILE et Marc FRANCO-SALVADOR : Active few-shot learning with fasl. *arXiv preprint arXiv :2204.09347*, 2022.
- [16] Hieu T NGUYEN et Arnold SMEULDERS : Active learning using pre-clustering. *In Proceedings of the twenty-first international conference on Machine learning*, page 79, 2004.
- [17] Andrei A RUSU, Dushyant RAO, Jakub SYGNOWSKI, Oriol VINYALS, Razvan PASCANU, Simon OSINDERO et Raia HADSELL : Meta-learning with latent embedding optimization. *arXiv preprint arXiv :1807.05960*, 2018.
- [18] Burr SETTLES : Active learning literature survey. 2009.
- [19] David I SHUMAN, Sunil K NARANG, Pascal FROSSARD, Antonio ORTEGA et Pierre VANDERGHEYNST : The emerging field of signal processing on graphs : Extending high-dimensional data analysis to networks and other irregular domains. *IEEE signal processing magazine*, 30(3):83–98, 2013.
- [20] Christian SIMON, Piotr KONIUSZ, Richard NOCK et Mehrtash HARANDI : Adaptive subspaces for few-shot learning. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4136–4145, 2020.
- [21] Jake SNELL, Kevin SWERSKY et Richard ZEMEL : Prototypical networks for few-shot learning. *Advances in neural information processing systems*, 30, 2017.
- [22] Olivier VEILLEUX, Malik BOUDIAF, Pablo PIAN-TANIDA et Ismail BEN AYED : Realistic evaluation of transductive few-shot learning. *Advances in Neural Information Processing Systems*, 34, 2021.
- [23] Gaoang WANG, Jenq-Neng HWANG, Craig ROSE et Farron WALLACE : Uncertainty sampling based active learning with diversity constraint by sparse selection. *In 2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, pages 1–6. IEEE, 2017.
- [24] Yan WANG, Wei-Lun CHAO, Kilian Q WEINBERGER et Laurens van der MAATEN : Simpleshot : Revisiting nearest-neighbor classification for few-shot learning. *arXiv preprint arXiv :1911.04623*, 2019.
- [25] Felix WU, Amauri SOUZA, Tianyi ZHANG, Christopher FIFTY, Tao YU et Kilian WEINBERGER : Simplifying graph convolutional networks. *In International conference on machine learning*, pages 6861–6871. PMLR, 2019.
- [26] Liu YANG et Rong JIN : Distance metric learning : A comprehensive survey. *Michigan State University*, 2(2):4, 2006.
- [27] Yi YANG, Zhigang MA, Feiping NIE, Xiaojun CHANG et Alexander G HAUPTMANN : Multi-class active learning by uncertainty sampling with diversity maximization. *International Journal of Computer Vision*, 113(2):113–127, 2015.
- [28] Michelle YUAN, Hsuan-Tien LIN et Jordan BOYD-GRABER : Cold-start active learning through self-supervised language modeling. *arXiv preprint arXiv :2010.09535*, 2020.
- [29] Xiangyu YUE, Zangwei ZHENG, Shanghang ZHANG, Yang GAO, Trevor DARRELL, Kurt KEUTZER et Alberto Sangiovanni VINCENTELLI : Prototypical cross-domain self-supervised learning for few-shot unsupervised domain adaptation. *In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13834–13844, 2021.
- [30] Imtiaz ZIKO, Jose DOLZ, Eric GRANGER et Ismail Ben AYED : Laplacian regularized few-shot learning. *In International Conference on Machine Learning*, pages 11660–11670. PMLR, 2020.