

SUJET DE STAGE: **Etude de la robustesse des méthodes semi-supervisées pour la classification, application pour la détection de pigments dans les œuvres d'art**

MOTS CLÉS: apprentissage profond, IA frugale, apprentissage semi-supervisé, imagerie éléments chimiques

LABORATOIRES D'ACCUEIL - Laboratoire des signaux et systèmes (L2S), Laboratoire SATIE, Univ. Paris-Saclay, Gif-sur-Yvette, France.

DESCRIPTION DU SUJET DE RECHERCHE ET DU TRAVAIL ATTENDU :

Contexte et objectifs - Ces dernières années, les algorithmes d'apprentissage profond ont pris une place prépondérante dans de nombreux domaines scientifiques grâce à leurs performances remarquables dans des tâches variées d'apprentissage automatique et de traitement de l'information. Cependant, ces algorithmes nécessitent l'accès à de très grandes bases de données labélisées, ce qui limite leur utilisation dans des contextes moins grand public et complique la confiance que l'on peut avoir en ces méthodes. En ce qui nous concerne, nous nous intéresserons dans ce stage au problème de la classification des pigments dans les œuvres d'art par fluorescence des rayons X, ou XRF. En effet, cette modalité d'imagerie a l'avantage de fournir une cartographie des éléments chimiques non organiques, notamment le chrome (Cr), le plomb (Pb), le fer (Fe), et bien d'autres [1]. De plus, la distribution spatiale de la quantité d'un élément chimique renseigne sur l'appartenance de ce dernier aux couches du support, à un pigment ou au liant de la couche image. La question qui se pose alors est comment retrouver ces pigments dans les œuvres d'art existantes et comment utiliser cette information pour en mesurer le vieillissement. Pour cela les techniques utilisées sont basées sur de la détection. On va chercher dans le spectre des images les caractéristiques permettant de détecter les pigments utilisés. La détection peut être considérée comme un problème d'apprentissage automatique semi-supervisé, l'idée étant de segmenter l'image de départ pour y retrouver les spectres des pigments utilisés. Suivant les couleurs ou les artistes, on peut aussi devoir chercher des composantes peu ou mal représentées. Ainsi on s'intéresse actuellement à des techniques adaptées à de petites bases de données, en particulier le *few shot learning* [4] ou les méthodes semi-supervisées.

L'objectif de ce stage est d'étudier les méthodes d'apprentissage automatique semi-supervisées adaptées à la détection de pigments dans les œuvres d'art à l'aide de la technologie XRF. Ces données sont très bruitées avec un bruit s'apparentant à du speckle difficile à filtrer. Nous proposons de nous appuyer sur des approches semi-supervisées où seulement un petit nombre de pixels non ambigus sont étiquetés en amont. L'étiquetage des autres pixels des images du dataset d'entraînement se fait alors en affectant des pseudo-labels estimés par prédiction. Ces approches peuvent être non robustes si on ne prend pas en compte l'incertitude des pseudo-labels pour chaque échantillon individuellement. De premiers travaux sur le sujet ont été menés dans [2]. Dans le stage, nous étudierons alors l'intérêt d'approches à base de *credal sets* qui représentent l'incertitude par des ensembles convexes de distribution et qui sont mathématiquement très solides, en particulier les prédicteurs de Venn [3].

TRAVAIL PROPOSÉ :

Le premier objectif de ce stage est d'étudier les méthodes de classification et de segmentation basées sur les prédicteurs de Venn dans le cadre donné par l'imagerie XRF. En particulier il devra étudier la robustesse des méthodes de segmentation par rapport aux données bruitées et aux erreurs de labélisation. Ce travail sera ensuite adapté au problème de classification, en particulier en développant des méthodes multiclassées basées sur les prédicteurs de Venn-Abers [3].

Le ou la candidat.e devra posséder de solides connaissances en apprentissage par réseaux de neurones mais également en traitement d'image.

References

- [1] C Colantonio, L Clivet, E Laval, Y Coquinot, C Maury, M Melis, and C Boust. Integration of multispectral imaging, xrf mapping and raman analysis for noninvasive study of illustrated manuscripts: the case study of fifteenth century "humay meets the princess humayun" persian masterpiece from louvre museum. *The European Physical Journal Plus*, 136(9):958, 2021.

- [2] Sofiane Daimellah, Sylvie Le Hégarat-Masclé, and Clotilde Boust. Domain-informed and neural-optimized belief assignments: A framework applied to cultural heritage. *International Journal of Approximate Reasoning*, 187:109534, 2025.
- [3] Côme Rodriguez, Vitor Martin Bordini, Sebastien Destercke, and Benjamin Quost. Self learning using venn-abers predictors. In *Conformal and Probabilistic Prediction with Applications*, pages 234–250, 2023.
- [4] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples: A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3):1–34, 2020.

CADRE :

- Le stagiaire recevra une gratification de stage standard, d'environ 600 € par mois.
- Sous réserve de financement, ce stage pourra se poursuivre en thèse.

EENCADRANTES :

Aurélia Fraysse est maître de conférences à l'université Paris-Saclay depuis 2008. Elle a obtenu son doctorat en 2005 en mathématiques appliquées à l'UPEC et a ensuite effectué un post doctorat à Telecom Paris. Elle est titulaire de l'habilitation à diriger des recherches depuis 2017. Elle est membre de l'équipe Problèmes Inverses, pôle Signaux et Statistiques, du Laboratoire des Signaux et Systèmes (CNRS/CentraleSupélec/Univ. Paris-Saclay) dont elle a été responsable de 2013 à 2020. Ses travaux de recherche portent sur les développements méthodologiques nécessaires à l'amélioration de la reconstruction dans le cadre de problèmes inverses mal posés.

Clotilde Boust est chercheuse au Centre de Recherche et de Restauration des musées de France, Ministère de la Culture. Elle est responsable depuis 2013 du groupe imagerie du département recherche, qui effectue les images spectrales, 3D et par rayons X pour l'analyse des oeuvres des musées. Docteure en optronique de l'Université Paris 6 en 2005, elle a été Maître de Conférence à l'Université de Nantes puis Versailles Saint Quentin en sciences et technologies de la communication de 2006 à 2013. Ses travaux de recherche portent sur l'optimisation de l'imagerie et des traitements croisés pour l'analyses et la conservation des oeuvres d'arts.

Sylvie Le Hégarat-Masclé est professeure de l'Université Paris-Saclay, chercheuse au laboratoire SATIE dont elle est aussi directrice adjointe. Docteur (1996) en Signal et Image de TélécomParisTech et HDR (2005) de l'Université de Versailles Saint-Quentin en 2005, ses travaux concernent le raisonnement et l'apprentissage statistique sous incertitude appliqués à la reconnaissance de formes, la vision par ordinateur et la fusion de données. Elle est membre élue du comité de pôle A de l'ED STIC (depuis 2019).

CANDIDATURES :

Pour candidater, merci de fournir un CV, une lettre de motivation ainsi que l'ensemble des relevés de notes à votre disposition.