

These 2021

Title **Uncertainties for large scale deep learning-based image reconstruction**

Laboratory: IRFU/DAP/CosmoStat, CEA Saclay

Supervisors: Jean-Luc Starck & Jalal Fadili

Contact: jstarck@cea.fr ☎ : 01 69 08 57 64 <http://jstarck.cosmostat.org>

Jalal.Fadili@greyc.ensicaen.fr ☎ : 02 31 45 29 22 <https://fadili.users.greyc.fr>

Keywords: deep learning, spherical data, radio-interferometry

Subject:

Deep learning (DL) has changed the way of solving inverse problems. Many scientific challenges remain that must be met for its deployment in astronomical imagery: i) taking into account the physical training model, ii) estimating the uncertainties on reconstructed images, iii) generalization, and iv) the volume of data for scaling up. To quantify the uncertainties, we have introduced a probabilistic DL approach (Remy et al., 2020), which makes it possible to derive the a posteriori distribution of the solution. This requires however to use expensive simulation techniques (MCMC) which does not allow its use in ambitious projects like Euclid or SKA. Several challenges will be addressed in this thesis:

- Develop a new DL method to quantify uncertainties, while enjoying theoretical guarantees of coverage. We will rely on conformal quantile regression, a new method derived from theoretical statistics (Romano et al., 2019).

- Generalization: We recently proposed a new architecture of neural networks (the learnets, Ramsi et al., 2020), which has the advantage of including certain properties of the wavelet transform such as exact reconstruction. This type of architecture should provide a solution to the generalization problem.

- The scaling on data of dimension 3 or 4. It will then be a question of extending the results obtained in order to be able to efficiently handle this type of data.

The last challenge of this thesis will be to set up these new tools to solve problems in two large international projects, for dark matter maps with Euclid and SKA.

Résumé:

Incertitudes pour la reconstruction d'images à grande échelle basée sur l'apprentissage profond

Au delà de la classification d'images, l'apprentissage profond est en passe de changer la manière d'aborder la résolution des problèmes inverses en imagerie. Ainsi, de magnifiques résultats ont pu être obtenus au CEA en reconstruction d'image MRI (Ramzi et al, 2020) et en astrophysique (Sureau et al, 2020). Toutefois, de nombreux défis scientifiques subsistent qu'il est nécessaire de relever avant d'envisager un déploiement de tels outils en imagerie astronomique. Il s'agit en particulier de:

- la prise en compte du modèle de formation physique;

- la robustesse et la quantification des incertitudes pour munir l'image reconstruite d'intervalles de confiance;

- la généralisation;

- la taille des jeux de données d'entraînement, et le volume des données pour le passage à l'échelle.

A titre d'exemple, les images obtenues avec l'instrument Square Kilometer Array (SKA) auront une taille de 50000 x 50000 pixels. Les données peuvent également être en dimension 3 ou 4 (temps et fréquence en plus des deux dimensions spatiales).

Pour quantifier les incertitudes, nous avons récemment introduit une approche probabiliste (Remy et al., 2020), qui permet de dériver la distribution à posteriori de l'estimation donnée par apprentissage profond. Cette approche nécessite toutefois l'utilisation de techniques très coûteuses de simulation Monte-Carlo type

MCMC, ce qui ne permet pas un déploiement raisonnable de ce genre d'approches dans des projets ambitieux comme Euclid ou SKA.

Les travaux de cette thèse s'articuleront autour de plusieurs directions répondant ainsi aux défis plus haut:

- La première direction concerne l'élaboration de nouvelles méthodes pour quantifier les incertitudes et estimer les barres d'erreur, tout en jouissant de garanties théoriques de couverture, et éviter l'écueil d'être conservatif. On s'appuiera en particulier sur la régression conforme de quantiles, une nouvelle méthode issue des statistiques théoriques très récemment introduite dans (Romano et al., 20).
- La seconde direction concerne le problème de généralisation. Nous avons récemment proposé une nouvelle architecture de réseaux de neurones, que nous appelons learnet (Ramsi et al., 2020), qui a l'avantage d'inclure certaines propriétés de la transformée en ondelettes comme la reconstruction exacte. Nous pensons que ce type d'architecture peut apporter une solution au problème de généralisation, mais cela reste à étudier et démontrer.
- La dernière direction concerne le passage à l'échelle sur des données de dimension 3 ou 4. Il s'agira alors d'étendre les résultats obtenus pour pouvoir manipuler efficacement ce type de données.

Le second challenge de cette thèse sera de mettre en place les nouveaux outils pour résoudre des problèmes dans deux grands projets internationaux. Tout d'abord dans le cadre de la mission spatiale Euclid, dont le lancement est en prévu en 2022 et qui permettra de reconstruire des cartes de matière noire sur 150000 degrés carrés. Le CEA est responsable de la production de ces cartes et il sera essentiel d'obtenir des images de très grande qualité avec des barres d'erreur. Le second projet, SKA, représente un extraordinaire Big Data challenge. Un aspect intéressant de la radio-astronomie avec SKA est sa complémentarité avec Euclid. En effet il a été démontré que la puissance du weak lensing pour contraindre les paramètres cosmologiques en radio avec SKA et en optique avec Euclid était très similaire, mais que la combinaison des deux relevés du ciel permettait de s'affranchir de manière extrêmement robuste des systématiques. Le CEA peut jouer un rôle important dans cette synergie, d'une part par son rôle clé dans Euclid, et d'autre part en développant les outils nécessaires au weak lensing pour SKA.

L'environnement scientifique:

La thèse se déroulera au sein du groupe de recherche pluridisciplinaire, CosmoStat, au Département d'Astrophysique du CEA sous la direction de Jean-Luc Starck et de Jalal Fadili. L'équipe a une très forte expertise dans différents domaines (problèmes inverses, weak lensing, machine learning etc.). L'étudiant thésard disposera donc d'un environnement idéal pour mener à bien ses travaux. A l'issue de sa thèse, l'étudiant maîtrisera des outils sophistiqués, le savoir faire pour résoudre des problèmes inverses avec des techniques d'apprentissage profond. Il pourra alors soit continuer dans le domaine académique, soit exploiter ses compétences dans le domaine industrie.

References

1. H. Garsden, J. Girard, J.-L. Starck, S. Corbel, C. Tasse et al, "LOFAR Sparse Image Reconstruction", *Astronomy and Astrophysics*, 575, A90, 2015.
2. M. Jiang, J. Bobin and J.-L. Starck, "Joint Multichannel Deconvolution and Blind Source Separation", *SIAM Journal on Imaging Sciences*, 10, 4, pp. 1997-2021, 2017.
3. F. Nammour, M. A. Schmitz, F. M. Ngolè Mboula, J.-L. Starck and J. N. Girard, "Galaxy Image Restoration with Shape Constraint", submitted, 2020.
4. Z. Ramzi, P. Ciuciu and J.-L. Starck. "[Benchmarking MRI reconstruction neural networks on large public datasets](#)", *Applied Sciences*, 10, 1816, 2020. doi:10.3390/app10051816
5. N. Jeffrey, F. Lanusse, O. Lahav, J.-L. Starck, "[Learning dark matter map reconstructions from DES SV weak lensing data](#)", *Monthly Notices of the Royal Astronomical Society*, 492, 4, p.5023-5029, 2020. doi: [10.1093/mnras/staa127](#)
6. F. Sureau, A. Lechat and J.-L. Starck, "[Deep Learning for space-variant deconvolution in galaxy surveys](#)", *Astronomy and Astrophysics*, 641, A67, 2020.