

Détection des interférences radio au sein d'observations de pulsars par apprentissage profond

A. Berthereau^{1,2}, L. Guillemot^{1,2}, G. Theureau^{1,2,3}, and I. Cognard^{1,2}

¹LPC2E CNRS-Université d'Orléans, F-45071 Orléans, France

²Station de Radioastronomie de Nançay, Observatoire de Paris, CNRS/INSU, F-18330 Nançay, France

³LUTH, Observatoire de Paris, PSL Research University, CNRS/INSU, Université Paris Diderot, Meudon, France

Keywords: Radioastronomie, Interférences, Deep Learning, CNN, Pulsars.

Abstract

Les pulsars sont des étoiles à neutrons dont l'observation est rendue possible par des faisceaux radio balayant l'espace nous apparaissant sous la forme de signaux pulsés. Ces observations dans les fréquences radio sont cependant sujettes à de nombreuses interférences, appelées RFI pour *Radio Frequency Interferences*, qui sont d'origine humaine et qu'il nous faut traiter afin de pouvoir étudier ces astres. La méthode présentée repose sur l'utilisation d'un réseau profond pour détecter ces interférences dans les spectrogrammes dynamiques des observations radio. En interprétant ces spectrogrammes comme des images dans le plan temps-fréquence, nous utilisons la segmentation d'images pour assigner une classe, RFI ou Non-RFI, à chacun des points de l'image. Les résultats préliminaires montrent que le réseau détecte certaines RFIs tout en étant moins destructif sur le signal de l'étoile que des méthodes de traitement statistique usuelles.

1 Introduction

Les étoiles à neutrons sont des astres compacts issus de l'effondrement puis de l'explosion d'une étoile massive en supernova. Parmi les étoiles à neutrons, les pulsars se caractérisent par la présence d'un faisceau radio issu de chacun des pôles magnétiques balayant l'espace à chaque rotation. Il nous est possible d'observer ces astres sous la forme d'un pulse radio périodique dont la période correspondante à la rotation du pulsar est de l'ordre de la seconde à la milliseconde pour les plus rapides. Les observations radio sont propices à l'apparition d'interférence radio dites RFI pour *Radio Frequency Interferences*, issues des télécommunications et transmissions satellites. Il est nécessaire alors d'identifier ces interférences et de les exclure avant de procéder à l'analyse du signal de l'étoile que l'on étudie. L'identification et la séparation du bruit du signal d'intérêt sont au cœur des problématiques étudiées dans le domaine des réseaux de neurones profonds [1, 2]. Issue de la vision par ordinateur, la méthode de segmentation d'image qui consiste à assigner à chaque pixel une classe, est largement utilisée pour détecter des objets ainsi que les directions de mouvements. En considérant les observations radio comme un ensemble d'images du plan temps-fréquence, l'application des méthodes de segmentation d'images pour la détection d'interférences offre une alternative aux méthodes actuelles de traitement statistiques.

2 Données

Les observations que nous utilisons proviennent du grand radiotélescope de Nançay sur une plage de fréquences de 1.4 GHz à 3.5 GHz, et sous la forme d'un hypercube comprenant les informations en polarisation, fréquence, temps et phase de rotation du pulsar. Chaque observation comprend 128 canaux de fréquence de 4 MHz pour une bande passante de 512 MHz ainsi qu'un nombre variable de sous-intégrations temporelles. La durée des sous-intégrations varie : plus la rotation du pulsar est courte, plus courte sera la sous-intégration augmentant le nombre de ces dernières pour une durée d'observation fixe. De plus, un fichier permettant le masquage des RFIs dans le plan temps-fréquence est généré pour chaque observation sous la forme d'une matrice binaire aux dimensions n_{chan} , n_{sub} , où n_{chan} est le nombre de canaux et n_{sub} le nombre de sous-intégrations. Ce masque est produit par seuillage statistique sur l'observation via la méthode Coast Guard [3].

2.1 Réseau

L'identification des RFIs dans le plan temps-fréquence est une tâche équivalente à la classification de chaque pixel d'une image selon deux classes : RFIs et non-RFIs. Afin de procéder à cette classification, aussi appelée segmentation d'image, nous utilisons un réseau de neurones à convolutions (CNN) permettant l'extraction de caractéristiques discriminantes entre les classes [4]. Le réseau est organisé selon l'architecture U-net [5], avec la partie descendante qui extrait les caractéristiques discriminantes en un vecteur par le biais de convolutions. La partie montante crée un masque à partir de ce vecteur grâce à des convolutions transposées. Ces deux sections du réseaux sont connectées par des concaténations comme présenté sur la figure 1. Les dimensions des masques ainsi créés dépendent uniquement des dimensions des données d'entrée du réseau, ce qui permet au réseau de traiter des observations de taille variable.

L'apprentissage du réseau est supervisé et la fonction de coût optimisée est l'indice de Jaccard moyen représentant le recouvrement entre le masque prédit et le masque vrai, fourni pour l'apprentissage.

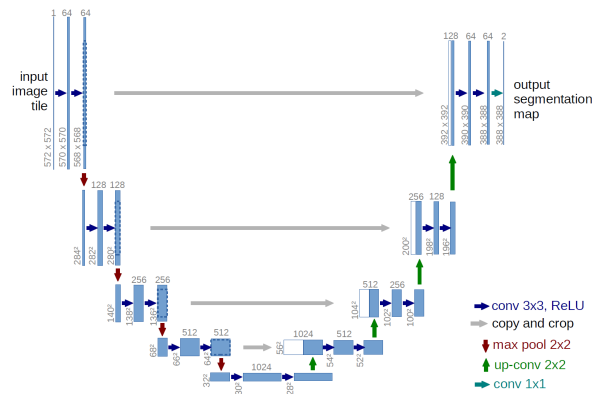


Figure 1: Architecture U-net. Figure extraite de Ronneberger et al. [5].

2.2 Base d'apprentissage

Le réseau est entraîné sur une base de données comprenant les observations radio de Nançay et les masques de ces observations générés par la méthode Coast Guard. Ce masque est considéré comme l'objectif d'apprentissage du réseau. Les observations radio sont réduites en dimensions en déterminant la médiane de l'intensité du signal radio en fonction de la phase rotationnelle du pulsar. L'intégralité de l'information spectrale et temporelle est ainsi conservée dans le spectrogramme, qui est la donnée à traiter pour l'identification des RFIs. La base de données comprend 35 354 observations dont 80% à 1.4 GHz et 66% correspondant à des pulsars millisecondes, dont la période de rotation est inférieure à 10 ms. Les ensembles d'apprentissage, de validation et de test comprennent respectivement 50%, 30% et 20% des observations et masques disponibles.

3 Résultats

Les résultats obtenus montrent une difficulté pour le réseau actuel à identifier certaines RFIs perturbant la ligne de base. L'application du masque prédit par le réseau a un effet moindre sur la ligne de base par rapport aux masques générés par Coast Guard. Néanmoins, les masques prédits sont moins destructifs sur le signal, et les rapports signal à bruit des observations masquées sont proches. Ces résultats préliminaires ont été obtenus lors d'un entraînement restreint à 7% de l'ensemble d'apprentissage.

4 Conclusion

Le réseau à convolutions ici présenté offre une méthode rapide et une alternative à Coast Guard. Actuellement en cours de développement, les résultats préliminaires sont encourageants sur l'identification des RFIs. Malgré une détection difficile de certains types d'interférences, notamment, les interférences à bandes étroites, les masques prédits par le réseau permettent de conserver une part égale voire plus importante du signal du pulsar. L'extension à l'ensemble d'entraînement complet ainsi qu'une optimisation des hyperparamètres du réseau sont des facteurs d'amélioration envisagés.

References

- [1] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Transactions on Medical Imaging*, 2018.
- [2] D. Wang and J. Chen, "Supervised speech separation based on deep learning: An overview," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 2018.
- [3] P. Lazarus, R. Karuppusamy, E. Graikou, R. N. Caballero, D. J. Champion, K. J. Lee, J. P. W. Verbiest, and M. Kramer, "Prospects for high-precision pulsar timing with the new effelsberg psrix backend," *MNRAS*, vol. 458, 2016.
- [4] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, 1998.
- [5] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS, Springer, 2015.