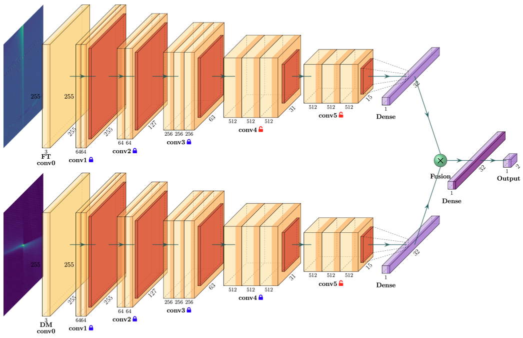
RFI masking recast as a supervised segmentation problem



SUMMARY.
Radio Frequency Interference (RFI) remains one of the main limitations in modern radio astronomy, severely affecting the detection of faint astrophysical signals such as Fast Radio Bursts (FRBs) or pulsars. This project proposes to explore and develop machine learning approaches for automatic RFI detection and masking in high-dimensional radio telescope data.
Students will investigate state-of-the-art deep learning techniques.

OBJECTIVES

- Understand the astrophysical context of transient radio sources: Fast Radio Bursts (FRBs) and Pulsars.
- Gain a solid understanding of Radio Frequency Interference (RFI): its origins, its impact on radio telescope data, and why automated masking is critical at modern facilities such as NenuFAR.
- Frame RFI mitigation as a supervised pixel-wise segmentation problem and understand how deep learning methods can address it.
- Develop hands-on proficiency in building, training, and evaluating segmentation neural networks using modern tools (PyTorch Lightning, MLFlow).
- Critically assess model performance on real observational data and understand the gap between benchmarks and deployment.

PREREQUISITES

☒ S1. Data Sciences
☒ S2. Statistics

THEORY

by Sara El Bouch

This module begins with a concise review of machine learning fundamentals (supervised vs. unsupervised learning, loss functions, gradient descent, overfitting and regularization) before introducing deep learning: convolutional neural networks (CNNs), residual connections, and encoder–decoder architectures. Particular attention is paid to the **segmentation task**, assigning a label to every pixel of an input, as the core methodology for RFI masking.
The astrophysical motivation is woven throughout: students are introduced to the time–frequency representation of radio telescope data (dynamic spectra), the morphology of RFI signals, and the characteristics of genuine astrophysical pulses such as dispersed FRB and pulsar signals. This dual perspective provides the conceptual grounding needed to design and interpret segmentation models responsibly.

APPLICATIONS

by Sara El Bouch

- **Practical work 1: Tooling:** Hands-on introduction to PyTorch Lightning (training loops, callbacks, checkpointing) and MLFlow (experiment tracking, metric logging, model registry). Students train a simple classifier on a toy dataset to build confidence with the workflow.
- **Practical work 2: Segmentation baseline:** Implementation of a U-Net-style encoder–decoder from scratch. Students explore architectural choices (depth, skip connections) and experiment with different loss functions on a labelled RFI dataset.
- **Project: RFI masking on NenuFAR data:** Students apply and refine their segmentation pipeline on real dynamic spectra from the NenuFAR low-frequency array. The project includes data preprocessing (normalisation, augmentation), training on labelled observations, and qualitative and quantitative evaluation of masks. Students are expected to document their experimental choices, analyse failure cases, and propose improvements.

MAIN PROGRESSION STEPS

- Tier 1: **Foundations of ML/AI and radio astronomy context:** Lectures on supervised learning, CNNs, and the physics of FRBs/Pulsars. Introduction to dynamic spectra and RFI phenomenology.
- Tier 2: **Segmentation theory and tooling:** Deep dive into encoder–decoder architectures, loss functions, and evaluation metrics. Practical works 1 and 2.
- Tier 3: **RFI masking project on real data:** End-to-end project on NenuFAR observations, from raw HDF5 data loading to mask generation and critical analysis.

EVALUATION

- Theory grade [30%]
  - Written exam (70%): Covers ML/AI fundamentals, CNN architectures, segmentation losses, and evaluation metrics.
  - Article presentation (30%): Students select and present a research paper on RFI mitigation or related ML for radio astronomy (e.g., FETCH, AOFlagger, or a recent deep learning approach). Graded on clarity, critical analysis, and ability to connect the paper to course material.
- Practice grade [30%]
  - Practical works (30%): Assessed on correctness of implementation, quality of experimental protocol, and written answers to guided questions.
  - Project (70%): Graded on initiative and originality of design choices, rigour of evaluation on

NenuFAR data, quality of analysis (including failure cases), and clarity of the written report.

- `Defense grade [40%]`
  - Oral and slides quality
  - Context
  - Project
  - Answers to questions

BIBLIOGRAPHY & RESOURCES

- FETCH Deep learning classifier for FRB candidates
- Goodfellow, Bengio & Courville, *Deep Learning* (MIT Press), Chapters on CNNs and regularisation.
- Ronneberger et al. (2015), *U-Net: Convolutional Networks for Biomedical Image Segmentation*, MICCAI.
- Offringa et al. (2012), *A morphological algorithm for improved RFI detection* AOFlagger reference.
- NenuFAR documentation and open data portal.
- PyTorch Lightning and MLFlow official documentation.

CONTACT

✉ sara.el-bouch@oca.eu