Stage Master 2 ML x Astro

Title: Investigating Diffusion Models for Astronomical Image Deconvolution - boosting the synergy between Euclid and LSST

Names and affiliations of the principal and co-investigators (the team)

Alexandre Boucaud, APC, France (PI) <aboucaud@apc.in2p3.fr>
Marc Huertas-Company, IAC, Spain <marc.huertas.company@gmail.com>

Motivation:

The European Space Agency's Euclid mission, with its space-based observations, provides high spatial resolution images but has limited wavelength coverage. Conversely, Vera C. Rubin Observatory's Legacy Survey of Space and Time (LSST), a ground-based survey, offers extensive wavelength coverage but faces challenges in spatial resolution due to atmospheric distortions. Unifying these two datasets via deconvolution methods can lead to unparalleled high-resolution, multi-wavelength observations of the universe, addressing a variety of astrophysical and cosmological questions. The proposed project aims to develop and investigate the efficacy of diffusion models for astronomical image deconvolution hence maximizing the synergistic datasets of Euclid and LSST.

Image deconvolution is the process of reversing the optical distortion that takes place during image capture. In astronomical imaging, this is especially crucial given the presence of atmospheric turbulence, instrumental noise, and other sources of degradation that can compromise the clarity and quality of the observations. Traditional algorithms have long been used for deconvolution, but with the rise of AI, particularly deep learning, there's been a paradigm shift in the way these problems are approached. However, even if some works exist as a proof-of-concept with very promising results using Generative Adversarial Networks for example, the deployment for scientific analysis still does not exist, partly because of stability problems (hallucination) and lack of flexibility.

Diffusion models have recently gained traction in the machine learning community as a powerful generative model. They model the data generation process as a diffusion process, essentially a Markov Chain transitioning from a noisy version of the data to the clean data over several timesteps. Compared to other generative models such as GANs (Generative Adversarial Networks), VAEs (Variational Autoencoders), and normalizing flows, diffusion models have several key advantages which are fundamental for our purpose: Sample Quality, Stability of Training, Diversity of Generated Samples, Flexibility, Robustness to Noise and easy Conditional Generation among others.

The goal of this project is therefore to develop and test diffusion models for astronomical image deconvolution, with the idea of maximizing the synergies between Euclid's high spatial resolution and LSST extensive wavelength coverage. Given the unavailability of both actual Euclid and LSST data, the project will utilize high-resolution, denoised images from the Illustris TNG hydrodynamical simulations as prior information, and mock observations from the Hyper Suprime Cam (HSC) survey as the input images.

Proposed work and implementation:

1. Develop and adapt diffusion models for astronomical image deconvolution, ensuring compatibility with the unique features of astronomical data.

2. Use high-resolution, denoised images from the Illustris TNG simulations as a prior and mock observations from the HSC survey as input images, simulating the synergies between Euclid and LSST.

3. Evaluate the performance of the proposed method through rigorous validation and comparison with traditional deconvolution techniques.

Expected Outcomes:

1. A tailored diffusion model capable of effectively handling astronomical image data, paving the way for its application to real Euclid and LSST datasets.

2. Improved images that fully exploit the combination of high spatial resolution and extensive wavelength coverage, simulating the potential synergies between Euclid and LSST.

3. A comprehensive assessment of the proposed method's performance, establishing its viability for future applications to real datasets from Euclid and LSST.

Timeline:

During the 6-month internship, the student will spend **3 or 4 months at the IAC, in Spain** (Canarias).

- Month 1-2: Development of the diffusion model framework and its adaptation to astronomical data (APC).
- Month 3-4: Application of the model to simulated data, using Illustris TNG images as priors and HSC survey mock observations as input (IAC).
- Month 5: Optimization of the methodology and in-depth analysis of the results (IAC).
- Month 6: Finalization of the project, including documentation and preparation of results for potential presentation (APC).

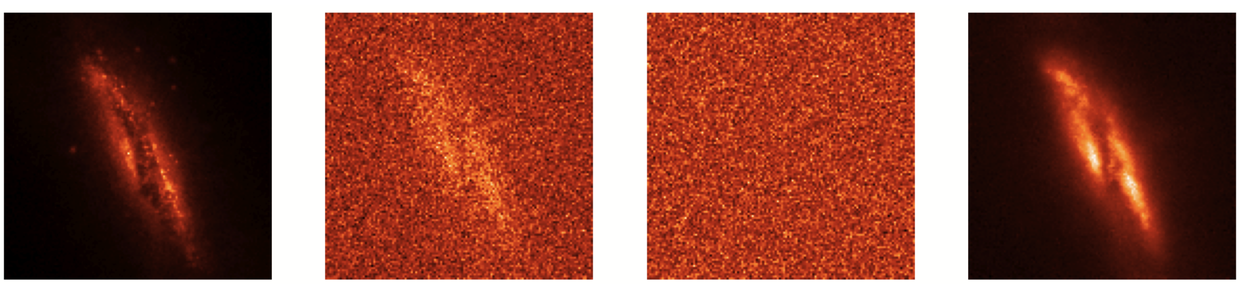

Figure 1. Proof-of-concept example. The rightmost panel shows a simulated image from the IllustrisTNG model without noise. The leftmost panel is the generated image by a simple diffusion model. The middle panels show 2 steps in the denoising process.

Summary:

This internship project represents a unique opportunity to explore the potential of diffusion models in addressing the challenges posed by the integration of diverse astronomical datasets. By simulating the synergies between Euclid and LSST using high-quality hydrodynamical simulations and mock observations, we aim to lay the groundwork for future applications of this methodology to real data from these two groundbreaking missions.