Artificial Intelligence for Science and Operations in Astronomy

An ESA/ESO Workshop
Working together in support of science

SCIOPS 2022

16 – 20 May 2022
ESO, Garching, Germany

Abstract Booklet

All times are in CEST (UTC +2)

★ **Invited Talks** (IS): 30+10 min

Contributed Talks: 15 + 5 min

Poster session: 3 + 2 min

R Remote

Monday 16 May

Chairperson: Michael Sterzik

13:45 - 14:00	Conference Welcome
14:00 - 14:40	**Pierre-Philippe Mathieu (IS)** The Rise of Artificial Intelligence for Earth Observation (AI4EO)
14:40 - 15:00	Marcos López-Caniego ESA Virtual Assistant in ESASky: enabling archival data exploration via natural language processing
15:00 - 15:15	BREAK
15:15 - 15:55	**Emille Ishida (IS)** Fink: incorporating expert knowledge in machine learning applications for large scale sky surveys
15:55 - 16:15	Jeroen Audenaert An all-sky stellar variability machine learning classification framework for TESS and PLATO
16:15 - 16:30	BREAK
16:30 - 16:50	John F. Suárez Pérez [R] Assessing the quality of massive spectroscopic surveys with unsupervised machine learning

The Rise of Artificial Intelligence for Earth Observation (AI4EO)

Pierre-Philippe Mathieu & the Φ-lab team
ESA, Frascati, Italy

The world of Earth Observation (EO) is rapidly changing as a result of exponential advances in sensor and digital technologies. The speed of change has no historical precedent. Recent decades have witnessed extraordinary developments in ICT, including the Internet, cloud computing and storage, which have all led to radically new ways to collect, distribute and analyse data about our planet. This digital revolution is also accompanied by a sensing revolution that provides an unprecedented amount of data on the state of our planet and its changes.

Europe leads this sensing revolution in space through the Copernicus initiative and the corresponding development of a family of Sentinel missions. This has enabled the global monitoring of our planet across the whole electromagnetic spectrum on an operational and sustained basis. In addition, a new trend, referred to as “New Space”, is now rapidly emerging through the increasing commoditization and commercialization of space.

These new global data sets from space lead to a far more comprehensive picture of our planet. This picture is now even more refined via data from billions of smart and inter-connected sensors referred to as the Internet of Things. Such streams of dynamic data on our planet offer new possibilities for scientists to advance our understanding of how the ocean, atmosphere, land and cryosphere operate and interact as part on an integrated Earth System. It also represents new opportunities for entrepreneurs to turn big data into new types of information services.

However, the emergence of big data creates new opportunities but also new challenges for scientists, business, data and software providers to make sense of the vast and diverse amount of data by capitalizing on powerful techniques such as Artificial Intelligence (AI). Until recently AI was mainly a restricted field occupied by experts and scientists, but today it is routinely used in everyday life without us even noticing it, in applications ranging from recommendation engines, language services, face recognition and autonomous vehicles.

The application of AI to EO data is just at its infancy, remaining mainly concentrated on computer vision applications with Very High-Resolution satellite imagery, while there are certainly many areas of Earth Science and big data mining / fusion, which could increasingly benefit from AI, leading to entire new types of value chain, scientific knowledge and innovative EO services.

This talk will present some of the ESA research / application activities and partnerships in the AI4EO field, inviting you to stimulate new ideas and collaboration to make the most of the big data and AI revolutions.

ESA Virtual Assistant in ESASky: enabling archival data exploration via natural language processing

Marcos López-Caniego
Aurora Techology for ESA/ESAC, Villanueva de la Cañada, Spain

ESASky (sky.esa.int) is a powerful visualisation tool developed by the Data Science and Archives Division at ESAC that can be used to discover multi-mission, multi-wavelength and multi-messenger astronomical data anywhere on the sky in a very intuitive way. Since its inception ESASky has been at the forefront of archival data exploration and an increasing number of astronomers are using ESASky to do their science. In parallel, a wide variety of AI technologies have become available in recent years that will enable new ways to interact with astronomical data centres. In this context, we would like to show how the ESA Virtual Assistant (EVA) has been integrated in ESASky and how it allows users to explore its data contents using natural language processing. In practice what this means is that users can interact with EVA using simple sentences that trigger the commanding of ESASky via its API. At the moment the virtual assistant has been programmed with a sufficient number of interactions to carry out most of the tasks that a typical user would do in the ESASky web interface with the mouse. However, exposing the virtual assistant to a larger audience will allow us to train the neural network running behind the virtual assistant to understand more complex sentences and commands, opening new avenues for more advanced exploration of ESASky.

Fink: incorporating expert knowledge in machine learning applications for large scale sky surveys

Emille Ishida
CNRS, Clermont, France

Next generation experiments such as the Vera Rubin Observatory Legacy Survey of Space and Time (LSST) will provide an unprecedented volume of time-domain data opening a new era of big data in astronomy. To fully harness the power of these surveys, we require analysis methods capable of dealing with large data volumes that can identify promising transients within minutes for follow-up coordination.
In this talk I will present Fink, a broker developed to face these challenges. Fink is based on high-end technology and designed for fast and efficient analysis of big data streams. It has been chosen as one of the official LSST brokers and will receive the full data stream. I will highlight the state-of-the-art machine learning techniques used to generate early classification scores for a variety of time-domain phenomena including kilonovae and supernovae, as well as for artifacts, like satellites glitches. Such methods include Deep Learning advances and Active Learning approaches to coherently incorporate available information, delivering increasingly more accurate added values throughout the duration of the survey.
I will also highlight the potential for discovery that can be unveiled by adding domain knowledge as part of the learning process. This will require an adaptation of learning system as well as as an update on how domain knowledge experts interact with the data taking process.

An all-sky stellar variability machine learning classification framework for TESS and PLATO

Jeroen Audenaert
KU Leuven, Leuven, Belgium

The TESS Data for Asteroseismology (T'DA) working group within TESS Asteroseismic Science Consortium (TASC) is responsible for processing the tens of millions of stars observed by NASA's Transiting Exoplanet Survey Satellite (TESS). In order to process this vast amount of data, we developed a machine learning framework to automatically classify the observed stars according to their stellar variability type. The framework aggregates the predictions from multiple distinct machine learning classifiers and combines their individual predictions into a global optimal classification by means of a metaclassifier. While this machine learning methodology and its resulting classifications are already of prime importance on their own for TESS science and future space missions, in this contribution we will specifically explore how the classifications can be coupled with spectroscopic information obtained from ground-based telescopes. This way, we will create a tool that is ideal for selecting the optimal targets for space missions such as ESA's upcoming PLATO (PLAnetary Transits and Oscillations of stars) mission, irrespective of whether it is a core or complementary science programme.

R

Assessing the quality of massive spectroscopic surveys with unsupervised machine learning

John F. Suárez Pérez
Universidad de Los Andes, Bogotá, Colombia

Massive spectroscopic surveys targeting tens of millions of stars and galaxies are starting to dominate the observational landscape in the 2020 decade. For instance, a night of observation with the Dark Energy Spectroscopic Instrument (DESI) can measure on the order of 100k spectra, each spectrum sampled over 2k wavelength points, approximately. Assessing the quality of such a massive data flow requires new approaches to complement visual inspection by humans. In this work, we explore the Uniform Manifold Approximation and Projection (UMAP) as a technique to assess the data quality of DESI. We use UMAP to project DESI nightly data into a 2-dimensional space. Sometimes in this space, we are able to find a small number of outliers. After visual inspection of those outliers, we find that they correspond to instrument fluctuations that can be then fully diagnosed by inspecting the raw data, allowing the development of an appropriate solution through data re-processing. These results pave the way for to use machine learning techniques to automatically monitor the health of massive spectroscopic surveys.

Tuesday 17 May

Chairperson: Paula Sánchez Sáez

09:00 - 09:20	Michele Delli Veneri Data Cleaning and Detection and Characterization of Sources in ALMA data through Deep Learning
09:20 - 09:40	Juan Gil Log Analysis as an Operational Tool at Paranal Observatory
09:40 - 10:00	Michael Johnson VAMPIRA - Automated Provenance Generation for Astronomical Pipelines
10:00 - 10:15	BREAK
10:15 - 10:35	Stefan Schuldt Machine learning investigations for LSST: Strong lens mass modeling and photo-z estimation
10:35 - 11:15	**Elena Cuoco (IS)** *R* Artificial Intelligence application to Gravitational wave transient signals
11:15 - 11:30	BREAK
11:30 - 11:50	Caroline Heneka Networks Learning the Universe: From 3D (hydrogen tomography) to 1D (classification of spectra)
11:50 - 12:10	Sandor Kruk Using deep learning and crowdsourcing to survey asteroid trails in ESA's Hubble data archive
12:10 - 14:00	LUNCH

Chairperson: Vicente Navarro

14:00 - 14:40	**Camille Avestruz (IS)** AI Enabled Data Exploitation for Astrophysics and Cosmology
14:40 - 15:00	Daniel Muthukrishna Real-Time Detection of Anomalies in Large-scale transient surveys
15:00 - 15:15	BREAK
15:15 - 15:55	**Maggie Lieu (IS)** *R* Pushing the limits of Astronomy using AI
15:55 - 16:15	Carter Rhea A New Paradigm in X-ray Spectral Analysis
16:15 - 16:30	BREAK
16:30 - 17:00	POSTER SESSION 1

Data Cleaning and Detection and Characterization of Sources in ALMA data through Deep Learning

Michele Delli Veneri
University of Naples Federico II, Italy

The Atacama Large Millimeter/submillimeter Array (ALMA) is an aperture synthesis telescope consisting of 66 high-precision antennas. Currently ALMA is providing 1TB of scientific data daily. Within the next decade, at least one order of magnitude of increased in data rate is foreseen. In terms of imaging products, ALMA will produce single field and mosaic cubes of at least two orders of magnitude larger than the current cube size. Since the number of observed spectral lines at once will be duplicated, advanced algorithms are needed to provide shorter processing time (ALMA Pipeline Team, 2021) while handling larger images. For such reason we have developed a Deep Learning Pipeline which starts from the raw data cubes and outputs model cubes on which source detection is performed. The pipeline consists of two Convolutional Autoencoders respectively tasked with learning to denoise the 3D continuum emission and, subsequently, HI emission after continuum subtraction, 3D Faster RCNN architecture which learns how to detect and classify (from noise spikes and artifacts) sources within the cube, and a set of specialized ResNets that take as input the bounding boxes produced by the Faster RCNN and regress the sources morphological parameters. To test the performance of the pipeline and characterize the robustness against increasing levels of noise and sources morphological complexity, several sets of increasing complex simulated observations are produced through the CASA software.

Log Analysis as an Operational Tool at Paranal Observatory

Juan Gil
ESO, Santiago, Chile

The Paranal Observatory is the host of VLT/VLTI, a world-class telescope that contains systems with different maturity levels. In the middle term Paranal will operate ELT and CTA. We are exploring new techniques to deal with this increasing complexity. A good part of system behavior is well captured by software logs, and several techniques brought from academic literature and as industry practices have been implemented as tools for system monitoring, error profiling, anomaly detection, and process discovery among others. Also, several prototypes have been made using ML and natural language processing techniques with the aid of Paranal DATALAB, a Docker-based data warehouse that provides years of historical data. This work shows the result of our experience and identifies key challenges for the adoption of log analysis techniques in the operation of other astronomical facilities.

VAMPIRA - Automated Provenance Generation for Astronomical Pipelines

Michael Johnson
Max Planck Institute for Radio Astronomy, Bonn, Germany

We present VAMPIRA, an automated provenance generation tool designed for data-intensive astronomical pipelines. VAMPIRA will provide users with a record of the processes, data, infrastructure, and users involved within a pipeline as well as their relations. By empowering astronomers with this information, they will be able to make informed decisions on the trustworthiness of data products or pipelines. This provenance information can be used to solve the so-called 'black box problem' prevalent within AI research. The intricate and complex architecture of modern artificial intelligence (AI) can obfuscate the objective of their operation from human comprehension. This lack of transparency may then raise concerns over the reliability and trustworthiness of AI implementations. The importance of provenance within the inherently complex processing of AI is exacerbated by the up to exabyte scale datasets expected from the next generation of astronomical survey telescopes.

Machine learning investigations for LSST: Strong lens mass modeling and photo-z estimation

Stefan Schuldt
MPA/TUM, Garching/Munich, Germany

Strong lensing analyses and photometric redshifts are both integral for cosmological studies with the Rubin Observatory Legacy Survey of Space and Time (LSST). With convolutional neural networks, we can obtain for both significant gain in the performance and speed. In my talk, I will present the new achievements of the HOLISMOKES collaboration in modeling galaxy-scale lenses using a residual neural network that was trained, validated and tested on very realistic mocks. This network enables us to predict the mass model parameters values with uncertainties, allowing us to analyze the huge amount of lenses soon to be discovered by LSST and predicting in advance the next appearing image and corresponding time delays in case of a lensed transient such as a supernova. I will further present our dedicated model comparison of 32 real lenses, once obtained with the network and once with traditional techniques, for which we developed an automation procedure to minimize the user input time drastically. Furthermore, I will briefly highlight our newly developed method NetZ to predict the photo-z based only on the pure galaxy images. Both networks are able to estimate the parameter values in fractions of a second on a single CPU while the lens modeling with non-automated traditional techniques takes typically weeks to month. With both networks and also our automated traditional pipeline, we are ready to process the huge amount of images obtained with LSST in the near future.

R

Artificial Intelligence application to Gravitational wave transient signals

Elena Cuoco
European Gravitational Observatory, Pisa, Italy

In recent years, Machine and Deep learning techniques approaches have been introduced and tested for solving problems in astrophysics. In Gravitational Wave (GW) science many teams in the LIGO-Virgo collaboration have experimented, on simulated data or on real data of LIGO and Virgo interferometers, the power and capabilities of machine learning algorithms both for the detector noise and gravitational wave astrophysical signal characterisation. I will show report on the application of Artificial Intelligence for the detection and classification of transient signals due to noise disturbances or to GW signals from Core Collapse Supernovae or Compact Binary Coalescence signals either in simulated data or in real data.

Networks Learning the Universe: From 3D (hydrogen tomography) to 1D (classification of spectra)

Caroline Heneka
Hamburg Observatory, Germany

With ongoing and future experiments, we are set to enter a more data-driven era in astronomy and astrophysics. To optimally learn the Universe from low to high redshift I advocate besides new observational techniques for the application of modern machine learning. For example in 3D, tomography of line intensity such as the 21cm line of hydrogen can both teach about properties of sources and gaseous media between. In this talk I first showcase the use of networks that are tailored for tomography to directly infer dark matter and astrophysical properties. I compare network models and highlight how a comparably simple 3D architecture (3D-21cmPIE-Net) that mirrors the data structure performs best. I present well-interpretable gradient-based saliency and discuss robustness against foregrounds and systematics via transfer learning. I complement these findings with a discussion of lower redshift results for the recent SKA Data Challenge, where hydrogen sources were to be detected and characterised in a large (TB) 3D cube. I will highlight my team'92s lessons-learned; our networks performed especially well when asked to characterise flux and size of sources bright in 21cm. Finally, moving from 3D to 1D, for the classification infrastructure group of the new ESO workhorse 4MOST (4-metre Multi-Object Spectroscopic Telescope), I propose an object classification layer to efficiently group the ~40,000 spectra per night (40 million in total) the instrument will collect.

Using deep learning and crowdsourcing to survey asteroid trails in ESA's Hubble data archive

Sandor Kruk
Max Planck Institute for Extraterrestrial Physics, Garching, Germany

The Hubble Space Telescope (HST) archives hide many unexpected treasures, such as trails of asteroids, showing a characteristic curvature due to the parallax induced by the orbital motion of the spacecraft. We have explored two decades of HST data for serendipitously observed asteroid trails with a deep learning algorithm on Google Cloud, trained on classifications from the Hubble Asteroid Hunter (www.asteroidhunter.org) citizen science project. The project was set up as a collaboration between the ESAC Science Data Centre, Zooniverse, and engineers at Google as a proof of concept to valorize the rich data in the ESA archives. I will present the first results from the project, finding 1700 asteroid trails in the HST archives (Kruk et al., 2022). The majority of the trails we found are faint (typically > 22 mag) and correspond to previously unidentified asteroids. Identifying the asteroids in HST images allows us to refine their ephemerides, as well as study their orbital distribution. I will argue that a combination of AI and crowdsourcing is an efficient way of exploring increasingly large datasets by taking full advantage of the intuition of the human brain and the processing power of machines. This comprehensive survey of asteroids required only ten hours of wall clock time to train the algorithm and scan the entire HST archival dataset, illustrating the immense scientific potential of these techniques and the benefits of exploiting the vast amounts of data in the archive.

★

AI Enabled Data Exploitation for Astrophysics and Cosmology

Camille Avestruz
University of Michigan, Ann Arbor, USA

We can characterize AI-enabled science as topics with specific problems solved or informed by techniques that learn directly from data. For astrophysics and cosmology, there are a handful of cases where AI approaches are clearly helpful. But, how much can we draw conclusions from or propagate findings to downstream analysis? In this talk, I will lay out examples of how AI can benefit science efforts in astrophysics and cosmology. This includes the use of relatively interpretable models to gain or confirm physical intuition for the purposes of building better-informed physical models, as well as the use of AI to improve the performance of a relatively rote task. I will end with commentary on a case example in which AI needs to be implemented with care to allow for scientifically valid and interpretable downstream analysis, particularly in the context of error propagation considerations and training set dependence.

Real-Time Detection of Anomalies in Large-scale transient surveys

Daniel Muthukrishna
Massachusetts Institute of Technology, Cambridge, USA

Time-domain astronomy has reached an incredible new era where unprecedented amounts of data are becoming available with new surveys such as LSST expected to observe over 10 million transient alerts every night. In this talk, I'll discuss the issue that with such large data volumes, the astronomical community will struggle to identify rare and interesting anomalous transients that have previously been found serendipitously. I'll present two methods of automatically detecting anomalous transient light curves in real-time. The first modelling approach is a probabilistic neural network built using Temporal Convolutional Networks (TCNs) and the second is an interpretable Bayesian parametric model of a transient. We demonstrate our methods' ability to provide anomaly scores as a function of time on light curves from the Zwicky Transient Facility. We show that the flexibility of neural networks, the attribute that makes them such a powerful tool for many regression tasks, is what makes them less suitable for anomaly detection when compared with our parametric model. The parametric model is able to identify anomalies with respect to common supernova classes with high precision and recall for most rare classes such as kilonovae and tidal disruption events. Our ability to identify anomalies improves over the lifetime of the light curves. Our

framework, used in conjunction with transient classifiers, will enable fast and prioritised follow-up of unusual transients from new surveys.

★ R

Pushing the limits of Astronomy using AI

Maggie Lieu
University of Nottingham, United Kingdom

Larger, deeper, multi-wavelength – these are the key words that astronomers yearn for on their journey to producing cutting edge research. Astronomy is big data. This, combined with the growth of machine learning, has radically transformed the way that we work with scientific data. It has expanded our capabilities of making new discoveries in both archival and new astronomical datasets. It has allowed us to do science more efficiently and more robustly. Machine learning is not magic but how far can we go using it? Here, I talk about some of the incredible things machine learning has already achieved in astronomy and what's next.

A New Paradigm in X-ray Spectral Analysis

Carter Rhea
University of Montreal, Quebec, Canada

X-ray spectral analysis is a powerful tool available to astronomers to study differing astrophysical phenomena from X-ray binaries, galactic black holes, and the intracluster medium. A new Bayesian paradigm is emerging in the field of X-ray spectral analysis. However, continued concerns over the choice of priors dominate the conversation. With our new machine learning methodology employing Mixture Density Networks (MDN), we use posterior target distributions calculated by an MDN as the priors for a full Bayesian inference approach to X-ray spectroscopy. Additionally, we discuss the potential of deconvolving observed X-ray spectra from the instrumental response using a Recurrent Inference Machine (RIM). Our findings indicate that using a RIM to deconvolve the spectrum and then passing the deconvolved spectrum to well-tuned MDN results in inaccurate estimates of the temperature and metallicity values which are critical in the study of galaxy clusters, plasma physics, and feedback astrophysics. In this talk, we will also discuss the implications for use cases and demonstrate the power of this exciting new methodology in our exploration of galaxy clusters.

Poster Session 1

1. **Miguel Doctor Yuste (Telespazio for ESA, Villanueva de la Cañada, Spain)**

EVA (ESA Virtual Assistant): The conversational AI platform for Space

Machine Learning is a sub-field of the computer science and the Artificial Intelligence (AI) that provides systems with the ability to act and learn without being explicitly programmed. One of the fields where machine learning techniques have been applied with great results is the Natural Language Processing (NLP). Namely, conversational AI is a branch of NLP focused on giving computers the ability to understand and interact with text and spoken words in much the same way human beings can. ESA Virtual Assistant (EVA) is a WEB platform capable of defining, building, training and deploying customized virtual assistants to power ESA services with conversational AI skills. These virtual assistants can be defined as software agents that converse through a chat interface, so they are able to have a real conversation with the user in order to provide him/her with some kind of valuable action. In this presentation we aim to give an overview of EVA platform placing emphasis on the main features and the advantages over other alternatives. Moreover, we will discuss about integration capabilities, highlighting some real world examples like ESASky, Cosmos portal and Gaia-ADQL. Finally, we will offer a quick hands-on demo to illustrate all concepts and techniques explained during the presentation.

2. **Ali-Dib Mohamad (CAP3/NYU, Abu Dhabi, UAE)** R

Identification of craters and boulders with deep learning

We will present a single machine learning model capable of detecting with very high accuracy craters on both rocky and icy objects, and boulders on the surface of 67P/C-G. We will furthermore discuss the diverse scientific results this model has provided, from a boulders map for 67P, to a new catalogue of Charon's craters.

3. **Fiorenzo Stoppa (Radboud University, Nijmegen, The Netherlands)**

ASID-L. Fast Source Localization in Optical Images

With the ever-increasing survey speed of optical wide-field telescopes and the importance of discovering transients when they are still young, rapid and reliable source localization is paramount. AutoSourceID-Light (ASID-L) is an innovative framework that uses computer vision techniques that can naturally deal with large amounts of data and rapidly localize sources in optical images. The ASID-L algorithm is based on a U-Net (Ronneberger et al. 2015) network and enhanced with a Laplacian of Gaussian filter (Chen et al. 1987) enabling outstanding performances in the localization of sources. Application on images of the MeerLICHT telescope demonstrates the great speed and localization power of ASID-L not only in low and mid crowded fields, but particularly in areas with more than 150 sources per square arcminute.

4. **Christophe Morisset (Instituto de Astronomia - UNAM, Ensenada, Mexico)** R

Machine learning methods applied to interstellar medium studies

I will present the preliminirary results of using Machine Learning methods (ANN, SVM, XGBoost) in research related to Interstellar Medium: determination of ionization correction factors (ICFs), determination of abundances, modeling galaxies emission.

5. Connor Bottrell (Kavli IPMU, Kashiwa, Japan) R

Unraveling galaxy merger histories with deep learning

Mergers between galaxies can be drivers of morphological transformation and various physical phenomena, including star-formation, black-hole accretion, and chemical redistribution. These effects are seen clearly among galaxies that are currently interacting (pairs) - which can be selected with high purity spectroscopically with correctable completeness. Galaxies in the merger remnant phase (post-mergers) exhibit some of the strongest changes, but are more elusive because identification must rely on the remnant properties alone. I will present results from my recent paper combining images and stellar kinematics to identify merger remnants using deep learning (arXiv:2201.03579). I show that kinematics are not the smoking-gun for improving remnant classification purity and that high posterior purity remains a significant challenge for remnant identification in the local Universe. However, an alternative approach which treats all galaxies as merger remnants and reframes the problem as an image-based deep regression yields exciting results.

6. Francesco Guarneri (INAF - Istituto Nazionale di Astrofisica, Trieste, Italy) R

QUBRICS: machine learning for searching bright, high-redshift quasars

The size and complexity of current astronomical datasets has grown to the point of making human analysis, in many cases, impractical or even impossible. With the advent of future observatories (e.g. LSST) and the generation of an unprecedented amount of data, automatic tools to extract information have become mandatory. A practical example of this situation is the search for bright, high-z QSOs in wide area surveys. These targets, among the rarest sources in the sky, enable a wide range of scientific applications both in cosmology and fundamental physics, but their number in the southern sky is relatively scarce. To fill the gap, we developed QUBRICS (QUasars as BRIght beacons for Cosmology in the Southern hemisphere) that is based on machine learning techniques applied to current and future photometric databases. Since 2019 over 450 new, bright ($i < 18$ and/or $Y<18.5$) and high-redshift ($z > 2.5$) QSOs have been identified using different, complementary methods (e.g. CCA, PRF, XGB). This talk will describe the QSO selection algorithms, their performances and the current state of QUBRICS, highlighting peculiarities, lessons learned, future prospects and scientific advancements enabled by QSOs discovered by QUBRICS. As an example, the gain produced by QUBRICS to carry out the Sandage Redshift-Drift Test at the ELT will be shown.

Wednesday 18 May

Chairperson: Michelle Lochner

09:00 - 09:20	Maxime Quesnel A physics-based deep learning approach for focal-plane wavefront sensing
09:20 - 09:40	Bartomeu Pou Mulet A Non-Linear Control Method with Reinforcement Learning for Adaptive Optics with Pyramid Sensors
09:40 - 10:00	Jalo Nousiainen Advances in model-based reinforcement learning for adaptive optics control
10:00 - 10:15	BREAK
10:15 - 10:55	**Wolfgang Kerzendorf (IS)** AI for proposal handling and selection
10:55 - 11:15	Siouar Bensaid Segmentation of the Galactic ISM Filaments using Deep Learning and enriched HiGAL catalogue
11:15 - 11:30	BREAK
11:30 - 12:00	POSTER SESSION 2

A physics-based deep learning approach for focal-plane wavefront sensing

Maxime Quesnel
University of Liège, Liège, Belgium

High-contrast imaging instruments are today primarily limited by non-common path aberrations appearing between the scientific and wavefront sensing arms. These aberrations can produce quasi-static speckles in the science images that are difficult to distinguish from exoplanet signatures. With the help of recent advances in deep learning, we have developed in previous works a method that implements convolutional neural networks (CNN) to estimate pupil-plane phase aberrations from point spread functions (PSF). Here we take it a step further by incorporating into the deep learning architecture the optical propagation occurring inside the instrument. The motivation behind this is to give a physical meaning to the models and to improve their robustness to various conditions. We explore how a variational autoencoder architecture that contains a differentiable optical simulator as the decoder can be used for that task. Because this unsupervised learning approach reconstructs the PSFs, it is not required to know the true wavefront aberrations in order to train the models, which is particularly promising for on-sky applications. We investigate different configurations of such a physics-based deep learning method and compare their performance to a standard CNN approach.

A Non-Linear Control Method with Reinforcement Learning for Adaptive Optics with Pyramid Sensors

Bartomeu Pou Mulet
Barcelona Supercomputing Center (BSC), Spain

Extreme Adaptive Optics (AO) systems are designed to provide high resolution and high contrast observing capabilities on the largest ground-based telescopes through exquisite phase reconstruction accuracy. In that context, the pyramid wavefront sensor (P-WFS) has shown promise to deliver the means to provide such accuracy due to its high sensitivity. However, traditional methods cannot leverage the highly non-linear P-WFS measurements to their full potential. We present a predictive control method based on Reinforcement Learning (RL) for AO control with a P-WFS. The proposed approach is data-driven, has no assumptions about the system's evolution, and is non-linear due to the usage of neural networks. First, we discuss the challenges of using an RL control method with a P-WFS and propose solutions. Then, we show that our method outperforms an optimized integrator controller. Finally, we discuss its possible path for an actual implementation.

Advances in model-based reinforcement learning for adaptive optics control

Jalo Nousiainen
Lappeenranta-Lahti University of Technology, Lappeenranta, Finland

One of the prime science cases of the next generations of high contrast imaging instruments on ground-based telescopes is the direct imaging of potentially habitable exoplanets. In order to reach the high contrast needed, the instruments are equipped an extreme adaptive optics (XAO) systems. The XAO systems will control thousands of actuators from kilohertz to several kilohertz at a framerate. Most of the habitable exoplanets are located at small angular separation from the host star, where the XAO control might be the limiting factor in the contrast. An up-and-coming field of research aimed at improving AO control methods is the application of fully data-driven control methods such as Reinforcement learning (RL). RL is an active branch of machine learning that aims to learn a control task via interaction with the environment and hence can be seen as an automated approach for AO control. In particular, model-based reinforcement learning (MBRL) has been shown to cope with both temporal and misregistration errors. Moreover, MBRL is shown to learn continuously on timescales of some seconds and automatically adjust to changing conditions. This work shows advances in the practical implementation of MBRL in a lab setup and simulation environments.

AI for proposal handling and selection

Wolfgang Kerzendorf
Michigan State University, East Lansing, USA

Many aspects of modern peer review have not changed from its inception in the 18th century despite drastic changes in the scientific community. Specifically, contrarily to the early days of peer review, it has become a significant challenge to identify experts that can effectively review the more and more specialized fields of science. The problem is exacerbated by the ever-rising number of researchers (having grown by 15% between 2014 and 2018 according to a UNESCO report) also seen through the staggering increase of publications and proposals (doubling every 14 years in astronomy). Some say that peer review has not adequately innovated as technology has advanced and the dissemination of publications has surged, creating a space for stagnant and biased reviews. We have developed and deployed a novel form of peer review in which the proposers become reviewers themselves known as distributed peer review. We enhanced this process using natural language processing and AI technologies to find an optimal match between the pool of proposals and the reviewers. In this talk, I will present this potential solution that was trialed at ESO and is about to go into full use. I will discuss potential other applications of AI in the field of peer review. I will close with an outlook of current and future experiments in peer review.

Segmentation of the Galactic ISM Filaments using Deep Learning and enriched HiGAL catalogue

Siouar Bensaid
Astrophysics Laboratory of Marseille, France

Understanding star formation is key to understand galaxy evolution. Observations collected in the Herschel Infrared Galactic plane Survey (Hi-GAL) revealed the ubiquitous presence of filamentary structures in the Galactic Plane. Filaments host star formation. Therefore, it is very interesting to study them and analyze the environment effect on their life cycle. A first step in understanding filaments consists in detecting them in the Galactic Plane. Detection algorithms based on standard image processing techniques present several limitations. Taking into consideration the drastic progress in Artificial Intelligence (AI), along with the data abundance about filaments, we propose to explore filament segmentation using Deep Learning (DL) framework. In this paper, we use state-of-the-art image segmentation in DL, U-Net, to segment filaments in all the Galactic Plane, in H2 column density images of the Galactic Plane obtained with Herschel Hi-GAL data. The used ground truth consists in the Hi-GAL filament catalogue provided by Schisano et al., enriched with filaments detected by the getSF algorithm. Obtained results reveal more structures than provided in the catalogue, which makes AI-based methods very promising for such application.

Poster Session 2

7. Cristiano Sabiu (University of Seoul, South Korea) R

Machine Learning the nature of Dark Matter with 21cm tomography

We present forecasts on the detectability of Ultra-light axion-like particles (ULAP) from future 21cm radio observations around the epoch of reionization (EoR). We show that the axion as the dominant dark matter component has a significant impact on the reionization history due to the suppression of small scale density perturbations in the early universe. This behavior depends strongly on the mass of the axion particle. Using numerical simulations of the brightness temperature field of neutral hydrogen over a large redshift range, we construct a suite of training data. This data is used to train a convolutional neural network that can build a connection between the spatial structures of the brightness temperature field and the input axion mass directly. We construct mock observations of the future Square Kilometer Array survey, SKA1-Low, and find that even in the presence of realistic noise and resolution constraints, the network is still able to predict the input axion mass.

8. Roland Szakacs (ESO, Garching Germany)

Detecting MgII Absorbers in Quasar Spectra: A Machine Learning Approach

Large scale surveys such as SDSS, DESI and upcoming 4MOST surveys provide an unprecedented amount of quasar spectra to explore. Given the vast number of spectra, it is unfeasible to explore them individually to infer systems of interest. Machine learning has become an increasingly important tool in astrophysics to approach such problems. One of these upcoming surveys is the 4MOST community survey 4HI-Q. The survey will obtain ~ 1 million R=20 000 quasar spectra to search for HI and metal absorption line systems in quasar sightlines at low to high redshifts. Among the targeted metal absorbers is MgII (2796, 2803). Using TNG50 from the Illustris project, we create realistic MgII absorption spectra and inject them into quasar mock spectra including Lyman alpha forests. In preparation for the 4HI-Q survey, we use this dataset to train a neural network to efficiently and accurately detect MgII absorption lines within quasar sightlines. We demonstrate that this method is a viable and accurate alternative to traditional methods for the detection of systems of interest in quasar spectra for current and future large-scale surveys.

9. Magda Arnaboldi (ESO, Garching Germany)

Enhanced discoverability and download of the ESO science data content via programmatic access

The ESO science archive is now equipped with new powerful interfaces which allow the astronomical community to discover and access its content, both interactively and programmatically. The latter is best suited to enable AI investigations on the full content of the ESO archive. The ESO programmatic access layer allows the discovery of science reduced data, scientific catalogues, raw data, calibration reference files, measurements and profiling of various atmospheric parameters of the La Silla Paranal observatory, and also of the ALMA science data. These data are all exposed via the same standard interfaces allowing an increased level of discoverability of the entire archive content, and is therefore a way to maximize the scientific return of the observatory.

10. Maja Jabłońska (Astronomical Observatory of University of Warsaw, Poland) R

Differentiable stellar disk integration

Inversion of stellar abundance maps based on time series of spectra is a demanding task. The Doppler Imaging (DI) method is successful, but since it is based on regularized chi-square minimization, it is computationally expensive. The first step to make the progess is to improve the method of generating spectrum for a stellar surface abundances and rotational velocities map. We present a python package that uses Jax (Autograd and XLA package) and is therefore state-of-the-art in terms of computational efficiency. Autograd lets us obtain the gradients of the operations, paving the way for machine learning models which would be able to solve the inverse problem - reconstruct stellar surface maps from the spectral time series. Moreover, we are using general radial velocity maps, which can be easily extended to modelany velocity fields, e.g., those of pulsating stars, allowing for fast spectral modelling for stars with various features.

11. Suhyun Shin (Seoul National University, South Korea) R

The survey for low-luminosity quasars at z~5 using artificial neural network and Bayesian statistics

Low-luminosity quasars ($M_{1450} \lesssim -25$ mag) at $z \sim 5$ are key ingredients to construct a reliable faint end of the quasar luminosity function. In our work, we tried to make a novel quasar selection consisting of an artificial neural network and Bayesian statistics to find the faintest quasars reaching $M_{1450} \sim -22$ mag, which is ~ 1 mag deeper than previous surveys. Using quasar SED models and a source catalog from the deep layer of the Hyper Suprime-Cam Subaru Strategic Program (HSC-SSP), we prepared training samples for quasar at z=4.5-5.5 and non-quasar classes, respectively. An artificial neural network determines whether an object is a quasar or not using multi-dimensional decision boundaries in color spaces. Bayesian information criterion indicates promising candidates by comparing quasar and star best-fit SED models. The combination of artificial neural network and Bayesian statistics utilizes all the flux measurements, enabling us to maximize the chance for finding high-redshift quasars as well as minimize the probability of containing contamination sources in our final candidates. As a result, we could build the quasar luminosity function with the most reliable faint-end slope of $-1.6_{-0.19}^{+0.21}$, which is a mid-value of previously reported slopes from –2.0 to –1.2.

12. MD Redyan Ahmed (Indian Institute of Science Education and Research, Kolkata, India)

Investigating the efficiency of the SPIIR pipeline in O3 offline Gravitational Wave search

Gravitational waves are 'ripples' in space-time caused by some of the most violent and energetic processes in the Universe. Detecting and analyzing the information carried by gravitational waves is allowing us to observe the Universe in a way never before possible, providing astronomers and other scientists with their first glimpses of literally unseeable wonders. SPIIR is one of the fastest of the five online pipelines.The SPIIR pipeline is distinguished from other CBC search pipelines in several aspects as it adopts the summed parallel infinite impulse response (SPIIR) method for matched filtering. This method is expected to be more efficient computationally than the traditional Fourier method when a filtering delay of less than 10s is intended. The other main difference is that the SPIIR pipeline selects candidates based on the maximum network likelihood ratio (MNLR) principle which is referred to as the coherent method. I use SPIIR and investigate its efficiency during the third LIGO-Virgo observing run (O3). I do this via investigating the distribution of False alarm rates of Gravitational waves via machine learning algorithms. I try to classify false alarms from real gravitational waves based on the analysis of data from GW detectors. My work will support SPIIR group's efforts to produce an 'offline' catalog of SPIIR GW.

Thursday 19 May

Chairperson: Camille Avestruz

09:00 - 09:40	**Michelle Lochner (IS)** Enabling New Discoveries with Machine Learning
09:40 - 10:00	Paula Sanchez Saez Searching for different AGN populations in massive datasets with Machine Learning
10:00 - 10:15	BREAK
10:15 - 10:55	**Vicente Navarro (IS)** ESA Data Exploitation Platforms: Accelerating Space and Navigation Science
10:55 - 11:15	Vanessa Moss *R* The quest for an autonomous ASKAP: automating the next-generation of survey telescopes
11:15 - 11:30	BREAK
11:30 - 11:50	Sam Sweere Deep Learning-Based Super-Resolution and De-Noising for XMM-Newton EPIC-pn
11:50 - 12:10	Laura Manduchi *R* Can Neural Networks be used to understand X-ray spectra?
12:10 - 14:00	LUNCH

Chairperson: Thomas Klein

14:00 - 14:20	Amelia Yu *R* Previously Undiscovered Exoplanets Detected with Deep Learning
14:20 - 15:00	**María Francisca Yañez Castillo (IS)** We are facing a double transformation: Digital and Sustainable
15:00 - 15:15	BREAK
15:15 - 15:55	**Natalie Behara (IS)** Moving Beyond Traditional KPIs: Data-Driven Performance Monitoring in Operations
15:55 - 16:15	Alessandro Terreri Neural Networks and PCA coefficients to identify and correct aberrations in Adaptive Optics
16:15 - 16:30	BREAK
16:30 - 16:50	Martino Romaniello Letting the data speak - The experience with running Deep Learning on the ESO Science Archive
16:50 - 17:10	Ignacio Toledo The ALMA Data Science Initiative: Building a Data-driven Organization to Improve Operations

Enabling New Discoveries with Machine Learning

Michelle Lochner
University of the Western Cape, Cape Town, South Africa

The next generation of telescopes such as the SKA and the Vera C. Rubin Observatory will produce enormous data sets, far too large for traditional analysis techniques. Machine learning has proven invaluable in handling massive data volumes and automating many tasks traditionally done by human scientists. Telescope scheduling can also be improved by machine learning, but scheduling choices, human or otherwise, can have significant impact on the resultant scientific results from large surveys. In this talk, I will discuss recent research on the observing strategy of the Rubin Observatory and the role machine learning can play in scheduling. I will also explore the use of machine learning for automating the discovery and follow-up of interesting astronomical phenomena. I will highlight how, for both scheduling and anomaly detection, the human-machine interface will play a critical role in maximising scientific discovery with automated tools.

Searching for different AGN populations in massive datasets with Machine Learning

Paula Sanchez Saez
ESO, Garching, Germany

Brightness variations of active galactic nuclei (AGNs) offer key insights into their physical emission mechanisms and related phenomena. These variations also provide us an alternative way to identify AGN candidates that could be missed by more traditional selection techniques. In this talk I will first introduce the ALeRCE light curve classifier, that uses a hierarchical imbalanced Random Forest and variability features computed from ZTF light curves, to classify each source into more than 15 subclasses, including three classes of AGNs (host-dominated, core-dominated, and Blazar). Then, I will present an anomaly detection (AD) technique designed to identify AGN light curves with anomalous behaviours in massive datasets, like the ZTF data releases. The main aim of this technique is to identify changing-state AGNs (CSAGNs) at different stages of the transition, but it can also be used for more general purposes, such as cleaning massive datasets for AGN variability analyses.

ESA Data Exploitation Platforms: Accelerating Space and Navigation Science

Vicente Navarro
ESA, Madrid, Spain

At the European Space Astronomy Centre (ESAC) near Madrid, the ESAC Science Data Centre (ESDC) hosts ESA archives for Astronomy, Planetary and Heliophysics Space Science. Furthermore, the GNSS Science Support Centre (GSSC), with special attention to Galileo and EGNOS, consolidates an ESA archive for scientific exploitation of Global Navigation Satellite Systems (GNSS). The sheer amount of data derived from current development plans for Space Science and Navigation Systems, constitutes a unique opportunity for research. Nevertheless, this deluge of data triggers a depart from standard processes and architectures supporting archives exploitation. More than ever, science innovation requires to leverage multiple data sources, sophisticated computing algorithms and cross-domain collaboration to achieve its goals. Hence, motivated by this increasingly complex landscape, 'ESA Datalabs' and 'GSSC Now' platforms expand Space and Navigation Science archives respectively, with a brand-new palette of collaboration, exploration, and analysis capabilities. Designed from the ground-up with a focus on end-user productivity, these two platforms accelerate science unleashing the full potential of ESA archives. Transparently to the user, the two platforms implement advanced network and computing capabilities to leave behind the discovery and download science era. 'ESA Datalabs' and 'GSSC Now' showcase the new era characterized by archives tight coupling with exploitation tools as a service.

R

The quest for an autonomous ASKAP: automating the next-generation of survey telescopes

Vanessa Moss
CSIRO, Sydney, Australia

The Australian Square Kilometre Array Pathfinder is an array of 36 x 12m dishes, and is the newest radio telescope owned and operated by CSIRO, Australia's national science agency. As a telescope designed to carry out large all-sky surveys, it necessitated a significant divergence from the existing astronomer-operated ATNF telescopes. This presented the observatory with a unique opportunity to redefine what science operations look like from the foundations up, rather than inheriting existing structures and systems. Since the start of Pilot Surveys Phase I in mid 2019, we have aimed to increase automation of ASKAP as much as possible, with the end goal of carrying out science operations in a way which maximises the autonomy of the system and minimises the reliance on human-centric decision making. Alongside a number of changes to the way the telescope is operated and monitored, a key ongoing development has been SAURON (Scheduling Autonomously Under Reactive Observational Constraints), which has been primarily responsible for science observation scheduling since late 2020. I will present an overview of our progress so far towards full survey operations starting in 2022, especially with respect to the critical role of automation, autonomy and AI in paving the path to the next generation of complex telescope systems.

Deep Learning-Based Super-Resolution and De-Noising for XMM-Newton EPIC-pn

Sam Sweere
ESAC, Madrid, Spain

The field of AI image enhancement has been rapidly evolving over the last few years and is able to produce impressive results on non-astronomical images. Here we present the first application of Machine Learning based super-resolution (SR) and de-noising (DN) to enhance X-ray images from European Space Agency's XMM-Newton telescope. We propose XMM-SuperRes and XMM-DeNoise - deep learning-based models that can generate enhanced SR and DN images from real observations. The models are trained on realistic XMM-Newton simulations such that XMM-SuperRes will output images with two times smaller point-spread function and with improved noise characteristics. The XMM-DeNoise model is trained to produce images with 2.5x the input exposure time from 20 to 50 ks. When tested on real images, DN improves the image quality by 8.2%, as quantified by the global peak-signal-to-noise ratio. These enhanced images allow identification of features that are otherwise hard or impossible to perceive in the original or in filtered/smoothed images with traditional methods. We demonstrate the feasibility of using our deep learning models to enhance XMM-Newton X-ray images to increase their scientific value in a way that could benefit the legacy of the XMM-Newton archive.

R

Can Neural Networks be used to understand X-ray spectra?

Laura Manduchi
ETH, Zürich, Switzerland

X-ray spectra are usually analysed by fitting the data to a known physical emission model. This is usually done manually with the help of spectral-fitting programs such as XSPEC, which aim to find the global minimum of a known function in a multi-dimensional parameter space using the method of grid-search. Such methods turn inefficient for an increasing number of spectra. Machine learning models, on the other hand, have recently achieved tremendous success with remarkable predictive capabilities. They may present an alternative solution in analysing X-ray spectra that could overcome the limits of the existing techniques. In this work, we propose Deep Spectra, a neural network composed of a Fully Convolutional Autoencoder, to denoise and disentangle the X-ray source to different components, and Convolutional Neural Networks to predict the parameters of each spectra component. We train the network using a dataset of simulated spectra from the Active Galactive Nuclei using XSPEC. On simulated data, the proposed approach shows a neat increase in performance compared to standard spectra fitting routines. We aim at extending the proposed approach to real-world data from the XMM-Newton Catalogue, however, several challenges need to be addressed and they represent interesting future research directions.

R

Previously Undiscovered Exoplanets Detected with Deep Learning

Amelia Yu
Henry M. Gunn High School, PAUSD, Palo Alto, USA

Using deep learning with the Adam optimization algorithm in this research, I detected 11 previously undiscovered exoplanets in the Kepler data. Although some of the exoplanet transit signals were evident, others were not as strong and need further evaluation. By using my own code, open source libraries, and deep learning packages TensorFlow, I developed a Python program for exoplanet search. The program first normalizes the transit light curves, trains the deep learning model, folds the light curves to intensify the transit signals, then uses the model to search for exoplanet transits in the Kepler light curves. Among the newly detected exoplanets, 9 of them are ultra-short period (USP) exoplanets with orbital periods shorter than a day, and the 2 others are short period exoplanets with periods between 1 to 10 days. Because the Kepler mission lasted for nine years and observed each star for a selected period of time, there are much more Kepler Objects of Interest (KOI) with shorter periods than those with long periods in the NASA database. This may be a reason why the orbital periods of the detected exoplanets in this study are shorter than 10 days. Meanwhile, the detection of these new exoplanets, especially the USP exoplanets, can shed light on their kind and expand our views on their planetary systems. Finally, these findings show that artificial intelligence such as deep learning can be an effective technological tool to detect objects of interest in astronomy big data.

We are facing a double transformation: Digital and Sustainable

María Francisca Yañez Castillo
Microsoft, Chile

Successful organizations from different sectors are connecting their sustainability and digital efforts into a single strategy as part of their core activities.
However, the intersection between Sustainability and Digital Transformation remains as a Blue Ocean with plenty of opportunities and challenges.
Blurring industry boundaries is one of the consequences of this double transformation.
In this new scenario, native digital-sustainable organizations and people are arising. Actually, “the natives” seem to enjoy and dance with the new digital and sustainable rules. Contrary, most organizations and people are struggling with transformations. Digital has dominated the agenda in the last decades, and more recent sustainability has exploded with new demands and challenges. So, can the organizations embrace this double transformation and interact with the natives?
Is this transformational scenario valid only for business or somehow is it inundated other sectors? What is happening with astronomy and other disciplines?
Astronomy is also facing this mix with native and transforming not only in terms of technology, but also in terms of people. In this presentation we will see how social challenges are being faced with astronomy methodologies and new technologies, and at the same time we will see how other industries have answers to astronomy challenges through technology.

Moving Beyond Traditional KPIs: Data-Driven Performance Monitoring in Operations

Natalie Behara
ESO, Santiago, Chile

There exists a fundamental link between data quality, system performance and environmental conditions. An end-to-end monitoring approach allows for better system understanding and opens the door for more efficient root cause investigations when anomalies occur. To prepare for future operations, Paranal established a dedicated data & system analysis framework, investigating different operational scenarios, testing new techniques and technologies. With industry partners we are exploring the best data infrastructure suited for multi-site, multi-instrument operations to achieve reliable and robust data access for operations. By increasing system understanding we are paving the way to fully integrated operations.

Neural Networks and PCA coefficients to identify and correct aberrations in Adaptive Optics

Alessandro Terreri
INAF, Rome, Italy

Static and quasi-static aberrations represent a great limit for high contrast imaging in large telescopes. Among them the most important ones are all the aberrations not corrected by Adaptive Optics (AO) system, called Non-Common Path Aberrations (NCPA). An estimate of the NCPA can be obtained by a trial-and-error approach or by more sophisticated techniques of focal plane wavefront sensing. In all cases, a fast procedure is desirable to limit the telescope downtime and to repeat, if needed, the correction procedure to cope with the temporal variation of the NCPA.In this work, through simulated images, I will describe the application of a supervised NN for the mitigation of NCPA in high contrast imaging at visible wavelengths and I will apply this method to fast imagers such as SHARK-VIS, the forthcoming visible band high-contrast imager for LBT. Preliminary results show a measurement accuracy of the NCPA of 2 nm RMS for each sensed Zernike mode in turbulence-free conditions, and 5 nm RMS per mode in presence of a residual turbulence corresponding to a WFE=42.5 nm RMS, a typical value during LBT AO system calibration. This measurement accuracy is sufficient to guarantee that, after correction, NCPA residuals in the system are negligible compared to the typical WFE >100 nm RMS of the best AO systems at large telescopes. Our simulations show this method is robust even in the presence of turbulence-induced aberrations that are not labelled in the training phase of the NN.

Letting the data speak - The experience with running Deep Learning on the ESO Science Archive

Martino Romaniello
ESO, Garching, Germany

The ESO Science Archive offers millions of data files to its science community. It is, then, crucial to its success that it also provides advanced query capabilities to guide the researchers to identify which of those data are of interest for their specific use. The broad hypothesis behind the work presented here is that letting the abundant real astrophysical data speak for itself, with minimal supervision and no labels, can reveal interesting patterns and distil information in ways that may facilitate data discovery. The need for supervision to be minimal is a must, given the sheer size and diversity of the archive contents. To this end, we have applied Deep Learning techniques to large datasets in the ESO Science Archive, most notably the entire data history of the HARPS high-resolution spectrograph. In our first round of experiments, we have obtained promising results in two directions. On the one side, the networks have learned to infer physical parameters of the celestial sources, such as radial velocity and effective temperature, just by being exposed to a large number of stellar spectra, without being explicitly asked to do so. On the other side, networks with the same basic architecture have shown an inclination to separate stellar and telluric spectral components, again without being instructed to do so. The lessons learned from the exercise and prospects will be presented.

The ALMA Data Science Initiative: Building a Data-driven Organization to Improve Operations

Ignacio Toledo
Joint ALMA Observatory, Santiago, Chile

Building fully-functional and successful observatories like ALMA requires the use of cutting-edge technologies and state-of-the-art engineering development. However, just 15 years ago most of the ground based observatories worked like laboratories, and our field is still learning how to apply other disciplines, like operations research, system engineering and optimization, to find the way to operate these "big data factories" in the most efficient and effective way possible. One important aspect is the ability to make decisions based on data. We need to answer questions like 'how are we doing', 'why are things going well or bad', 'what is the reason for the success or for the failure', 'how can we improve the success rates or prevent failures', etc. in order to make informed decisions. Since ALMA stores a huge amount of data everyday, and not just science data, it was easy to realize that we had a gold-mine waiting to be exploited. Four years ago, ALMA started a collaboration with the industry to provisionalize a platform and infrastructure that would allow us to use this data and our skills to answer these questions and make better decisions. This allowed us to understand what were our weaknesses, which are our strengths and what challenges we must still solve. The purpose of this talk is to share with the audience what we have learned, what have we achieved and what is our road ahead, in the process of applying data science to improve operations.

Friday 20 May

Chairperson: Christophe Arviset

09:00 - 09:40	**François Lanusse (IS)** Going Beyond Common Deep Learning Limitations with Deep Probabilistic Modeling
09:40 - 10:00	Jason Spyromilio Why AI is wrong for apparatus calibration
10:00 - 10:15	BREAK
10:15 - 10:35	Claudia Comito R The missing link between massive data and AI: parallel computing with Heat
10:35 - 10:55	Nick Cox EXPLORE - Innovative Scientific Data Exploration and Exploitation Applications for Space Sciences
10:55 - 11:15	Srividya Subramanian Astrophysics with INODE
11:15 - 11:30	BREAK
11:30 - 12:10	**Michelle Ntampaka (IS)** R ML for Astronomy: Cautionary Tales for the Community
12:10 - 12:30	Final comments

★

Going Beyond Common Deep Learning Limitations with Deep Probabilistic Modeling

François Lanusse
CEA Saclay, Gif-sur-Yvette, France

With an upcoming generation of wide-field cosmological surveys, cosmologists are facing new and outstanding challenges at all levels of scientific analysis, from pixel-level data reduction to cosmological inference. As powerful as Deep Learning (DL) has proven to be in recent years, in most cases a DL approach alone proves to be insufficient to meet these challenges, and is typically plagued by issues including robustness to covariate shifts, interpretability, and proper uncertainty quantification, impeding their exploitation in scientific analysis.
In this talk, I will instead advocate for a unified approach merging the robustness and interpretability of physical models, the proper uncertainty quantification provided by a Bayesian framework, and the inference methodologies and computational frameworks brought about by the Deep Learning revolution.
In particular, we will see how deep generative models can be embedded within principled physical Bayesian modeling to solve a number of high-dimensional astronomical ill-posed inverse problems, from deblending galaxy images to inferring the large scale matter distribution in the Universe. As another example, I will present efforts to combine deep learning with cosmological simulators as a way to ensure physically plausible fast simulations, which can then be embedded in larger inference schemes.

Why AI is wrong for apparatus calibration

Jason Spyromilio
ESO, Garching, Germany

The use of AI for operations is being promoted by various actors: "The data shall lead you to the truth". The analysis of multiple components of data sets to seek correlations is very powerful and can provide views into the system that humans would find difficult or even impossible to imagine. In the presentation it will be argued that usage of this methodology in a massively non-linear system is likely to result in incorrect operation of the system, producing correlated systematic errors that will be imprinted in the data. It is argued that while the usage of AI for analysis of a system is important, its inversion for the operation of the apparatus is inappropriate.

R

The missing link between massive data and AI: parallel computing with Heat

Claudia Comito
Jülich Supercomputing Centre, Jülich, Germany

When it comes to enhancing exploitation of massive data, machine learning and AI methods are very much at the forefront of our awareness. Much less so is the need for, and complexity of, applying these techniques efficiently across memory-distributed data volumes. Heat [1, 2] is an open-source Python library for high-performance data analytics, machine learning, and deep learning. It provides highly optimized algorithms and data structures for tensor computations using CPUs, GPUs and distributed cluster systems. Heat's Numpy-like API makes writing scalable, GPU-accelerated applications straightforward - at the same time, parallelism implemented under the hood via MPI provides a significant improvement in efficiency and performance with respect to, e.g., Dask. Born out of a large-scale collaboration in applied sciences, Heat also acts a platform for collaboration and knowledge transfer within data-intensive science. In this presentation, I will show you the inner workings of the library, tell you about our collaborations with the astrophysics and space science community (massively parallel signal-processing capabilities for the SKA-MPG telescope among others) and hopefully gain from you some insight into how to best support data-intensive astro operations going forward.

References: [1] Gotz, M., Debus, C., Coquelin, et al.: "HeAT - a Distributed and GPU-accelerated Tensor Framework for Data Analytics"; [2] *https://github.com/helmholtz-analytics/heat*

EXPLORE - Innovative Scientific Data Exploration and Exploitation Applications for Space Sciences

Nick Cox
ACRI-ST Toulouse France

We present the EXPLORE Horizon 2020 Research & Innovation project and its first outcomes. EXPLORE's main objective is to develop and deploy a suite of scientific data applications (SDAs) to foster exploitation of scientific data from astrophysics and planetary space missions. Six SDAs will produce new scientific data products and offer new scientific services. Four of the applications will leverage data primarily from Gaia, developing tools to help understand the evolution of our galaxy, the 3D distribution of interstellar matter, and support the discovery, classification, and characterisation of stars. The other two applications will integrate data from a range of lunar missions to focus on characterisation of the Moon's surface and potential human landing sites. The SDAs will utilise state-of-the-art Artificial Intelligence and visual analytics to enhance science return and discovery. In parallel, we built a small virtual platform (explore-platform.eu) that is used to develop, validate, and test each SDA. Later it could be offered as a service. The SDAs will be deployed, under open-source licences, on different cloud science platforms (e.g., ESA Datalabs and ESCAPE Science Analysis Platform) to stimulate uptake and sustainability. This open science project will help democratise space science exploration and exploitation. This project has received funding from the European Union's Horizon 2020 research and innovation pro-gramme under grant agreement No 101004214.

Astrophysics with INODE

Srividya Subramanian
Max-Planck-Institut für extraterrestrische Physik, Garching, Germany

In this presentation I introduce “INODE”, an end-to-end data exploration system, an unified comprehensive platform that provides extensive access to open datasets. INODE offers sustainable services in (a) data modeling and linking, (b) integrated query processing using natural language, (c) guidance, and (d) data exploration through visualization, thus facilitating the user in discovering new insights. I will demonstrate our unique system with a prime focus on Astrophysics usecase.

R

ML for Astronomy: Cautionary Tales for the Community

Michelle Ntampaka
Space Telescope Science Institute, Baltimore, USA

Astronomy is entering an era of data-driven science, due in part to modern machine learning techniques that enable powerful new data analysis methods. As we shift to a more data-driven scientific approach, we must consider how to build models that can be trusted.
In this talk, I will show examples of ML successes and failures from other fields, and will discuss the implications to astronomy, including steps we can take to build and train models that our peers can trust.