



Università degli Studi di Napoli Federico II
PhD program in
Information Technology and Electrical Engineering

PhD Student: Michele Delli Veneri

Cycle: XXXV

Training and Research Activities Report

Academic year: 2020-21 - PhD Year: Second

Tutor: Prof. Vincenzo Moscato

Co-Tutor: Prof. Giuseppe Longo

Date: October 21, 2021

Training and Research Activities Report

PhD program in Information Technology and Electrical Engineering

PhD student:

Cycle: XXXV

1 Information:

➤ PhD student: Michele Delli Veneri

PhD Cycle: 35th

➤ DR number: DR993895

➤ Date of birth: 06/03/1993

➤ Master Science degree: Physics

University: Federico II

➤ Scholarship type: Funding company – EUSTEMA SPA

➤ Tutor: Prof. Vincenzo Moscato

➤ Co-tutor: Prof. Giuseppe Longo

1 Study and training activities:

Activity	Type ¹	Hours	Credits	Dates	Organizer	Certificate ²
Information Theory	Courses	30	6	01/11/2020 - 01/03/2021	Prof. Anna Maria Tulino	y
Elaborazione dei segnali digitali	Courses	30	6	01/11/2020 - 01/03/2021	Prof. Anna Maria Tulino	y
Picariello Lectures on Data Science Performing an Iso27001 Assessment	Seminar	1	0.2	04/11/2020	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science Connecting the dots: investigating an APT campaign using Splunk	Seminar	1	0.2	11/11/2020	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science Digital Project Management: practices, processes, techniques, tools and scientific approach	Seminar	1	0.2	18/11/2020	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on	Seminar	1.5	0.3	25/11/2020	Prof.	y

Training and Research Activities Report

PhD program in Information Technology and Electrical Engineering

PhD student:

Cycle: XXXV

Data Science #andràtuttobene: Images, Texts, Emojis & Geodata in a Sentiment Analysis Pipeline				0	Giuseppe Longo, Prof. Flora Amato	
Picariello Lectures on Data Science At the Nexus of Big Data, Machine Intelligence, and Human Cognition	Seminar	1	0.2	02/12/202 0	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science Exploiting Deep Learning and Probabilistic Modeling for Behavior Analytics	Seminar	1	0.2	09/12/202 0	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science Data Driven Transformation in WINDTRE through Managers voice	Seminar	2	0.4	16/12/202 0	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science From Photometric Redshifts to Improved Weather Forecast an interdisciplinary view on machine learning	Seminar	1	0.2	13/01/202 1	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science Cybercrime and electronic evidence, The international legal framework for an effective criminal justice response	Seminar	1	0.2	20/01/202 1	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science Artificial Intelligence for	Seminar	1	0.2	21/01/202 1	Prof. Giuseppe Longo,	y

Training and Research Activities Report

PhD program in Information Technology and Electrical Engineering

PhD student:

Cycle: XXXV

notary's sector - a case study					Prof. Flora Amato	
Picariello Lectures on Data Science The era of Industry 4.0: new frontiers in business model innovation	Seminar	1	0.2	03/02/2021	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science Machine Learning: causality lost in translation	Seminar	1.5	0.3	10/02/2021	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science Approaches to Graph Machine Learning	Seminar	1	0.2	17/02/2021	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science Visual Interaction and Communication in Data Science	Seminar	2	0.4	03/03/2021	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science Big Data and Computational Linguistics	Seminar	2	0.4	10/03/2021	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science Sensoria Health	Seminar	1	0.2	17/03/2021	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science The coming revolution of Data driven Discovery	Seminar	1.5	0.3	25/03/2021	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science DoveAndiamoDomani -	Seminar	1.5	0.3	28/04/2021	Prof. Giuseppe Longo,	y

Training and Research Activities Report

PhD program in Information Technology and Electrical Engineering

PhD student:

Cycle: XXXV

Deep Tech					Prof. Flora Amato	
Picariello Lectures on Data Science Artificial Intelligence and 5G combined with holographic technology: a new perspective for remote health monitoring	Seminar	2	0.4	26/05/2021	Prof. Giuseppe Longo, Prof. Flora Amato	y
Picariello Lectures on Data Science Distributional Semantics Methods: How Linguistic features can improve the semantic representation	Seminar	2	0.4	23/06/2021	Prof. Giuseppe Longo, Prof. Flora Amato	y
ALMA I-TRAIN with the European ARC Network , Session I	Course	1	0.2	04/12/2020	European ARC Network	y
ALMA I-TRAIN with the European ARC Network , Session II	Course	1	0.2	15/12/2020	European ARC Network	y
From Cell to Galaxies - "Introductory Remarks - From Cells to Galaxies: Imaging Challenges in Astronomy and Medicine"	Seminar	2	0.4	15/09/2021	Karen Prairrie	y
SAE 2021 - Big4small, Data Science Methodology Transfer: Big to Small, Prof. Edwin Valentijn Keptein Institute for Astrophysics, Univesrity of Groningen, Kai Polsterer University of Haidenberg	Seminar	2	0.4	24/09/2021	Prof. Giuseppe Longo	y
Thriving as a doctoral student in informatics	Seminar	2	0.4	10/09/2021	Prof. Gerardine Fitzpatrick, Dr. Austen Rainer	y
Adventures in Astronomical Time Series Analysis	Seminar	2	0.4	12/09/2021	Jeffrey D. Scargle	y
From Cell to Galaxies -	Seminar	2	0.4	13/10/2021	Karen	y

Training and Research Activities Report

PhD program in Information Technology and Electrical Engineering

PhD student:

Cycle: XXXV

Urvashi Rau (NRAO) "Introduction to Radio Astronomy for Medical Imaging Professionals" and Daniel Sodickson "Introduction to Medical Imaging for Radio Astronomers"				1	Prairiee	
--	--	--	--	---	----------	--

1Courses, Seminar, Doctoral School, Research, Tutorship

2Choose: Y or N

1.1 Study and training activities - credits earned

	Courses	Seminars	Research	Tutorship	Total
Bimonth 1	0.4	1.5	9	0	11.1
Bimonth 2	12	1.5	9	0	22.5
Bimonth 3	0	1.6	9	0	11.1
Bimonth 4	0	0.8	9	0.32	11
Bimonth 5	0	0	4.5	0.32	6.7
Bimonth 6	0	2.0	4.5	0.96	9.9
Total	12.4	7.4	45	1.6	72.3
Expected	30 - 70	10 - 30	80 - 140	0 - 4.8	

2Research activity:

This year, most of my research activity has been dedicated to my Thesis, i.e. the development of a Deep Learning (DL) based architecture (temporarily named UDCNet) for the detection and characterization of sources in Radio data cubes, both of Astrophysical origin (SKA, APERITIF) and of Medical origin (BREST-MRI). Therefore the bulk of time this year has been spent in the actual development of the architecture and all the necessary i/o pipelines for two of the three datasets on which I plan to test my architecture, SKA and BREAST-MRI. In particular, in January 2021, I got involved with a group of researchers all part of COIN (Cosmostatistic Initiative), to the SKA Data Challenge 2, which lasted for six months and ended in July. Nevertheless, we did not win the challenge coming at the 8th place, I learned many lessons regarding the data and its properties, I got the change to test several DL architectures in order to prove the hypothesis on which some of the architectural choices of UDCNet are built upon and proved the possibility of reliably measure sources properties if the detection part of the problem is correctly addressed. At the moment of writing this document, I am co-working on an article summarizing the results of the challenge, and I'm looking forward to a confrontation between the participants to analyze which solutions and practices may be best suited to improve the results which

Training and Research Activities Report

PhD program in Information Technology and Electrical Engineering

PhD student:

Cycle: XXXV

will be extremely beneficial to the scope of my Thesis. Given our participation in the challenge, we will also soon receive access to the complete SKA datacube and its companion dataset which will give me the capability of comparing the performances of UDCNet against the competitors.

The UDCNet (Unet like Detection and Characterization Network) architecture is inspired by state of the art 3D segmentation architectures developed in the context of medical imaging to detect and segment cancerous lesions within MRI volumes and, infact, it shares some design choices with other DL architectures such as 3D Unet, 3D Faster R-CNN, M3D-RPN, Feature Pyramid Network, DenseNet and Fully Convolutional Network for Object Detection and aims to combine features from all the cited architectures and introduce novel ones in order to create an end-to-end detection and characterization framework for radio data. The architecture's final goal is to produce segmentation masks for objects of interest by training several specialized network at the end of a backbone feature extractor and then used the said produced segmentation mask to regress the sources morphological parameters. The specialized architectures for the detection part of the network are as follows: a **regressor** which aims to learn the linear transformation necessary to transform the generated bounding boxes into the ground truth bounding boxes made around ground truth sources, a **classifier** which aims to classify the boxes in several classes of interest plus a background of class and thus reduce the number of false positives and a **pixel classifier** (or **segmenter**) which takes the features belonging to true positive box and aims to classify each pixel of the box, i.e. make a segmentation mask of the source. The produced segmentation masks are used to produced moment-masked versions of the features extracted by the backbone which are then fed to several **regressors** (one for each parameter of interest) that aim to derive the geometrical properties of the sources. The backbone of the architecture is an **U-Net** style network made by **Dense Blocks** which have the capabilities of extracting relevant features at several spatial scales while controlling the number of parameters. This is achieved through bottleneck layers which perform a sort of dimensionality reduction after each convolutional block. The upsampled layers of the U-Net in the decoding path, which should contain relevant features for several spatial scaled, are combined with their correspondent counterparts in the encoding part and fed to the specialized networks. The key idea is that each of these heads should contain both relevant spacial and semantic information about the sources at a given spatial scale. Depending on the subsequent specialized network, the heads are combined or independently feed to the networks ahead. The bottleneck and region proposal networks (classifier and regressor) are trained together since the first epoch of training while the segmenter starts training only if a certain number of epochs or a threshold in the loss function has been reached. The parameter regressors are only trained after the detector has been fully trained. This training strategy must be implemented in order to account for the sparsity of the dataset and the necessities of the loss functions of each network.

Since the 13th of September, I have started my abroad period here in Groningen, working with Prof. Michael Biehl (Computer Scientist), Prof. John McKein (Astrophysicist) and Prof Edwin Valentjin (Astrophysicist) at the University of Groningen. Under their supervision, I'm performing a noise analysis on the SKA 3D cube in order to select on the basis of the RMS of the noise and the local SNR a subsample of "detectable" sources to use as a training set for the DL architecture. Moreover, Prof.

Training and Research Activities Report

PhD program in Information Technology and Electrical Engineering

PhD student:

Cycle: XXXV

McKein will soon make available to me a dataset containing real radio data produced by the WSRT-APERITIF radio telescope here in Groningen which will be the third dataset of which I plan to test UDCNet.

The Breast-MRI dataset has been shared by Prof. Giovanni Mettevier at the Department of Physics of Federico II and, in collaboration with EUSTEMA, I have spent some time developing the preprocessing and i/o pipelines required to apply UDCNet.

These datasets, SKA, APERITIF and BREAST-MRI, will make the three use cases presented in my Thesis to show the performance and capabilities of my UDCNet.

While working on my Thesis, I spent some time working on some side projects such as:

- **Team Science:** in collaboration with COIN, I'm working on measuring the interdisciplinarity of papers found in the Arxiv Public database. My contribution so far has been the development of the pipeline to get the papers through the Arxiv API and the analysis of the subsequent dataset through the use of clustering graph-based machine learning algorithms;
- **HyCASTLE:** on the 21 of November 2020 I presented at the IAU Challenges and Innovations In Computational Astrophysics conference the HyCASTLE algorithm, a hybrid classification algorithm that I developed in collaboration with the late Prof. Picariello. Nevertheless, the official paper was submitted to the Elsevier Knowledge Base System Journal around the same date of the conference, the article is still under revision, but we hope that it will see the light of day soon;
- **Dermatology:** in collaboration with the Department of Medicine of the Federico II, we carried out and published on *American Journal of Clinical Dermatology* the article *Effectiveness and Safety of Long-Term Dupilumab Treatment in Elderly Patients with Atopic Dermatitis: A Multicenter Real-Life Observational Study*, which aims to characterize the effectiveness of the treatment and highlight some baseline markers which can be used to predict the outcome and effectiveness of the treatment after three and six months of Dupilumab therapy;
- **Geology:** in collaboration with the INAF Observatory and the Department of Geology of the Federico II, we published a novel application of Deep Learning to classify terrestrial drainage networks and those found on solar system bodies (*A novel approach to the classification of terrestrial drainage networks based on deep learning and preliminary results on solar system bodies*). The work was published on Nature Sci Rep on the 12th of March 2021. In October I have begun preliminary work for the application of UDCNet to the Hyper-spectral Cubes produced by the PRISMA satellite with the objective of detecting mining deposits and measure their composition and, if undergoing an active mining operation, to assess their impact on surrounding regions
- **Astrophysics:** publication of the work carried out during the first year of the PhD on the TOLIMAN telescope in the book *Intelligent Astrophysics*. Some work has been carried on the development of an anomaly detection algorithm which has proven to be able to improve the performance on the regression of Euclid redshifts by automatically combining several other methods;

Training and Research Activities Report

PhD program in Information Technology and Electrical Engineering

PhD student:

Cycle: XXXV

- **Dentistry:** I have carried out in collaboration with the dentistry firm Guerino Paolantoni a statistical assessment of the effectiveness of a wide range of support therapies for the treatment of Stage IV Parodontitis. An article about our work together is in production.
- **EUSTEMA:** thought the year I have been working with EUSTEMA RND Lab Naples on several topic of common interest such as topic detection and document classification in the legal domain through machine learning. Introductory talks are been made in order to extend the work that I'm carrying for the detection and characterization of cancerous lesions in medical images towards an explainable AI framework.

1 Research products

Effectiveness and Safety of Long-Term Dupilumab Treatment in Elderly Patients with Atopic Dermatitis: A Multicenter Real-Life Observational Study, 22 July 2021, *American Journal of Clinical Dermatology*, doi: [10.1007/s40257-021-00597-5](https://doi.org/10.1007/s40257-021-00597-5);

A novel approach to the classification of terrestrial drainage networks based on deep learning and preliminary results on solar system bodies, Carlo Donadio 1, Massimo Brescia 2, Alessia Riccardo 3, Giuseppe Angora 4, Michele Delli Veneri 5, Giuseppe Riccio 2, 12 March 2021, *Nature Sci Rep* 11, 5875 (2021). <https://doi.org/10.1038/s41598-021-85254-x>;

Periodic Astrometric Signal Recovery Through Convolutional Autoencoders - Intelligent Astrophysics, Part of the *Emergence, Complexity and Computation* book series (ECC, volume 39), <https://doi.org/10.1007/978-3-030-65867-0>, Springer Link

5. Conferences and seminars attended

Training and Research Activities Report

PhD program in Information Technology and Electrical Engineering

PhD student:

Cycle: XXXV

A new source detection and characterization pipeline for radio data-cubes: application to the SKA-DC2 at the EAS 2021 (European Astronomical Society Annual Meeting) Conference 28/06/2021 – 02/07/2021;

Presentation at the IAU Challenges and Innovations In Computational Astrophysics – II, Live Zoom Meeting 18 – 21 November 2020 (<https://cb1-chaica2020.astro.unistra.fr/Welcome.html>). Abstract title: HyCASTLE: a Hybrid Classification System based on Typicality, Labels and Entropy;

The Third National Workshop on the SKA Project - The Italian Route to the SKAO Revolution 4/10/2021 – 8/10/2021 – Virtual, <https://indico.ict.inaf.it/event/1512/timetable/#all.detailed>;

From Cells to Galaxies – Exploring the Synergies between Radio Astronomy and Medical Imaging – September 2021 – March 2022;

6. Periods abroad and/or in international research institutions

Since the 13th of September I have been working at the University of Groningen in the Netherlands under the supervision of Prof. Michael Biehl, Prof. John McKein and Prof. Edwin Valentijn in order to carry forward my Thesis project. In particular, at the time of writing this document, I have spent a bit more than a month and one week here in Groningen and my work has proceeded over two parallel tracks. With Prof. McKein I'm mostly exploring the data, studying the noise distributions and preselecting training SKA sources on the basis of their local SNR, brightness and geometrical parameters. With Prof. Biehl I'm mostly working on the debugging and optimization of the architecture in order to train it on multi-GPU clusters.

Exact dates of my period abroad so far: 13/09/21 - 21/10/21

7. Tutorship

During the year, as authorized by the PhD board, I have helped Prof. Longo with his course of Data Mining MOD.B (U2642) for the Data Science Master Degree by helping Prof. Longo with both frontal lessons and by tutoring students towards the final examinations. Between May and July 2021 I helped out Prof. Longo with 16 hours of direct frontal lesson for a total of 0.64 CFU, while between July and November I have supported several groups of students with their task for the final examinations for a total of 0.96 CFU.

Training and Research Activities Report

PhD program in Information Technology and Electrical Engineering

PhD student:

Cycle: XXXV

8. Plan for year three

During my third year I plan to completely focus on my Thesis. In the upcoming month I will finish debugging the architecture and the data preparation of the SKA data and I hope to get first baselines performances on both the SKA and MRI data by December 2021. Some time will also be spent in laying down some necessary chapters in the thesis. By March 2022 I will have obtained the APERITIF data set from Prof. McKein and I'm hopefully at a good stage in the experiments on the other two datasets. From the initial discussions with Prof. McKein, it seems that minor modifications in the network parameters and i/o pipelines will be needed for the application on the APERITIF data and thus I hope to obtain results there in a relatively short time.

Hoping in the soon publications of the HyCASTLE algorithm, we would like to test the algorithm capabilities on real databases of interest for EUSTEMA and also on the problem of behavior detection on social networks.