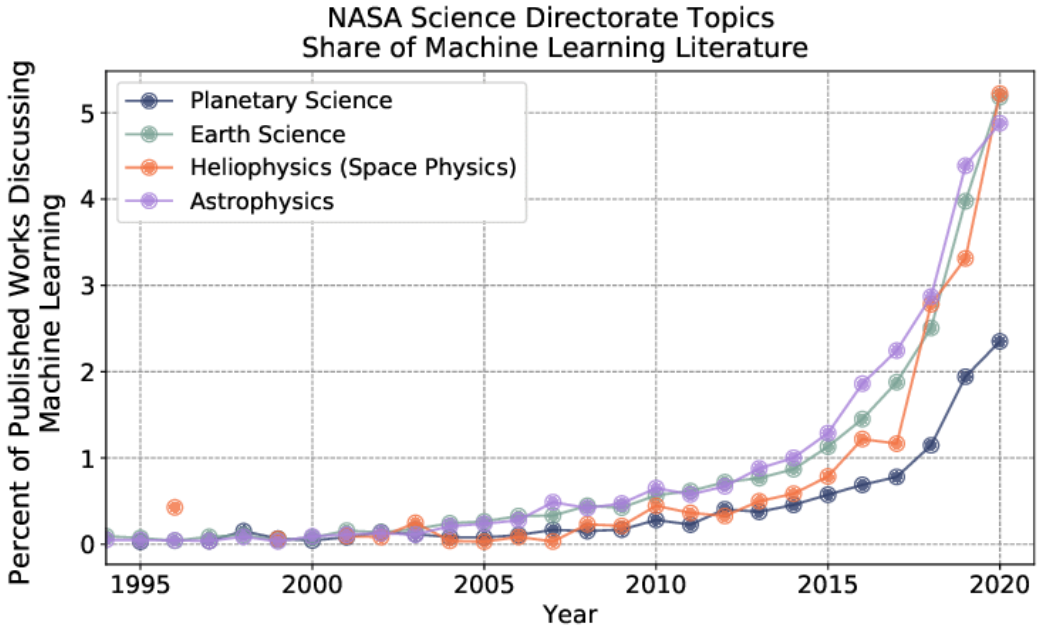
Some thoughts on Scientific ML & some recent applications in astrophysics

Giuseppe Longo
Università di Napoli Federico II
Dep. of Physics, E. Pancini

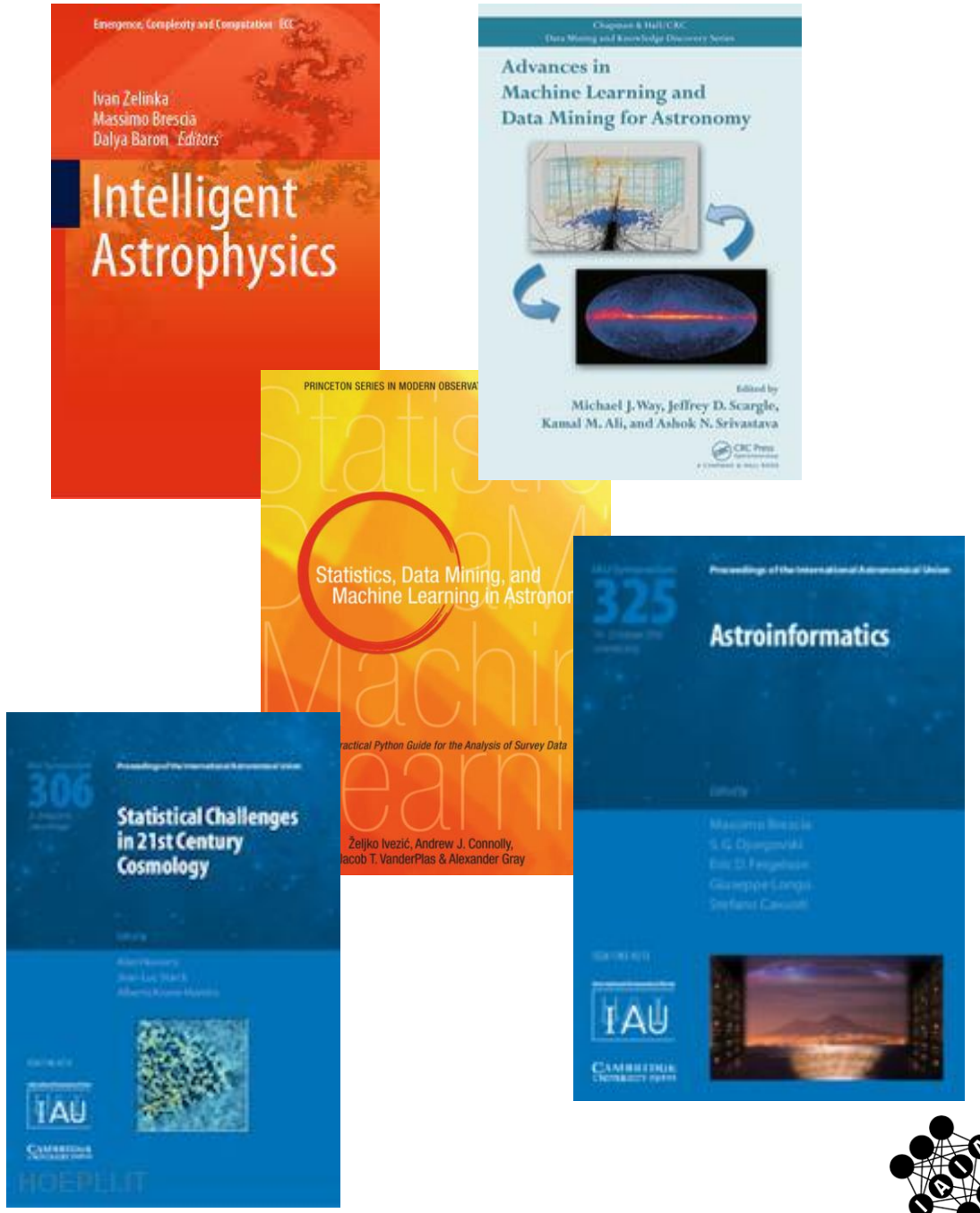
International AstroInformatics Association



A new science… ASTROINFORMATICS

(statistics, machine learning, computer science and domain expertise)

Please forget the AI label … there is nothing more stupid than a ML algorithm.



International AstroInformatics Association

Traditional justification for using ML in Science

- **DATA SIZE**: modern Instruments and detectors produce data flows/streams impossible to handle with traditional methods

Many applications (90%)

- photometric redshifts
- Galaxy classification (traditional approach)
- Identifying transients, etc…
- Etc. …..
- Image segmentation
- Denoising
- Etc. …..

Supervised approaches
Require base of knowledge

Unsupervised approaches
Few templates

International AstroInformatics Association

Traditional justification for using ML in Science

- **DATA SIZE**: modern Instruments and detectors produce data flows/streams impossible to handle with traditional methods

Many applications (90%)

- photometric redshifts
- Galaxy classification (traditional approach)
- Identifying transients, etc…
- Etc. …..
- Image segmentation
- Denoising
- Etc. …..

Supervised approaches
Require base of knowledge

Unsupervised approaches
Few templates

Simplified translation of the process

Data are humongous… «…colleagues and journals have told me good things about ML…»

Professor
(domain expert … old, knows little about ML) has **interesting** but rather **standard** problem…

Hires a **PhD**
(domain expert at large…, knows little about ML but he/she is young, enthusiastic and learns fast)

But… very often suffers from lack of formal understanding of statistical learning and ML ….

Using off the shelf libraries (Github, Kaggle, paperswithcode, ecc.) in three months puts together a code… almost always DL)

Personal Notice
It is Almost impossible for a DL method to perform worse than a human

Code performs equally and sometimed better than humans

Non optimised, quite always in overfitting (too high capacity hence far too complex for the problem…)

Professor
is astonished and happy…

PhD
publishes a paper

arXiv

International AstroInformatics Association

Traditional justification for using ML in Science

- **DATA SIZE**: modern Instruments and detectors produce data flows/streams impossible to handle with traditional methods

Many applications (90%)

- photometric redshifts
- Galaxy classification (traditional approach)
- Identifying transients, etc…
- Etc. …..
- Image segmentation
- Denoising
- Etc. …..

Supervised approaches
Require base of knowledge

Unsupervised approaches
Few templates

Simplified translation of the process

Data are humongous… «…colleagues and journals have told me good things about ML…»

Professor
(domain expert … old, knows little about ML) has **interesting** but rather **standard** problem…

Hires a **PhD**
(domain expert at large…, knows little about ML but he/she is young, enthusiastic and learns fast)

Using off the shelf libraries (Github, Kaggle, paperswithcode, ecc.) in three months puts together a code… almost always DL)

But… very often suffers from lack of formal understanding of statistical learning and ML ….

Personal Notice
It is Almost impossible for a DL method to perform worse than a human

Code performs equally and sometimed better than humans

Non optimised, quite always in overfitting (too high capacity hence far too complex for the problem…)

Professor
is astonished and happy…

PhD
publishes a paper

arXiv

Very much needed and useful

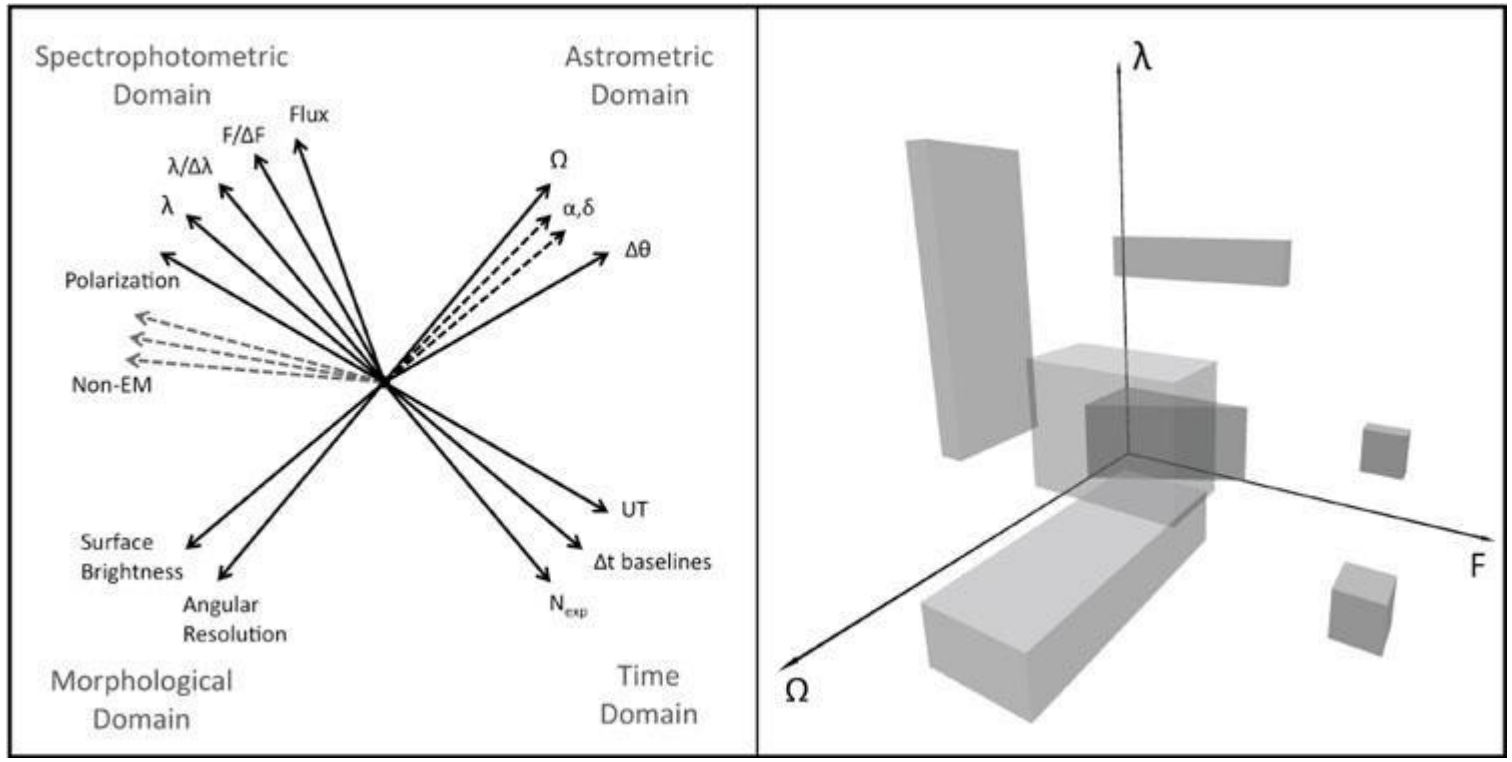
but… standard !

interesting, not very challenging

International AstroInformatics Association

Better (at least more exciting) justification for ML in Science

- **Data complexity**: modern data sets contain complex information (hundreds or thousands of measured parameters) which goes beyond the human brain capability to uncover patterns (trends, data structures, ecc…)



OPS – (Observable/Observed Parameter Space)

Each event (phenomenon, measure, ecc.) **identifies a point in an N-dimensional parameter space** (where N is the number of measured parameters)

Today NN >> 100

Correlations are non trivial structures in this OPS

International AstroInformatics Association

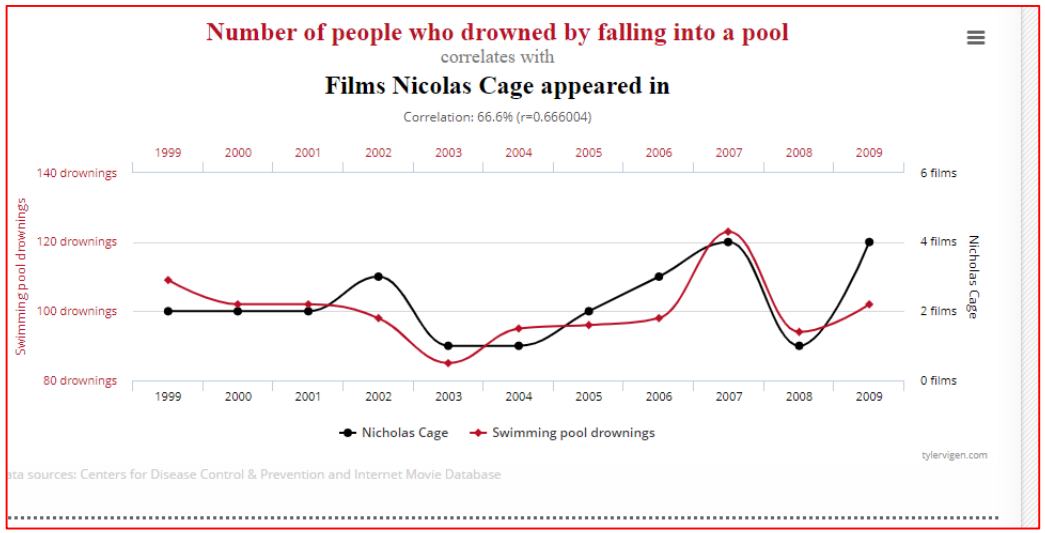
VERY SIMPLIFIED EPISTEMOLOGY

Laws of physics

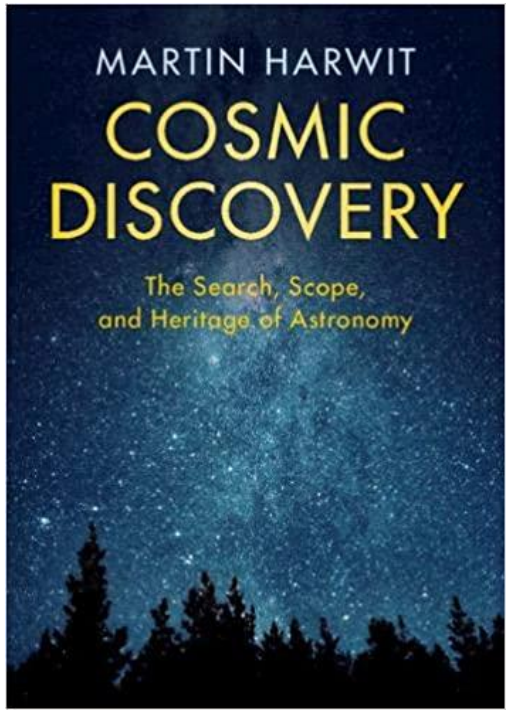
- **Empirical laws** emerge from Identified patterns in the OPS
- Theory is a **compression algorithm** capable to subsume an empirical law in a formula.
- Criterium for a good theory: experimentum crucis (i.e. capability to predict new phenomena)

(in ML, from a statistical point of view this is equivalent at largeto the concept of significance and usefulness of a correlation).

> *Empirical laws are laws that can be confirmed directly by empirical observations. The term "observable" is often used for any phenomenon that can be directly observed, so it can be said that **empirical laws are laws about observable.***
>
> *Rudolph Carnap, 1966*

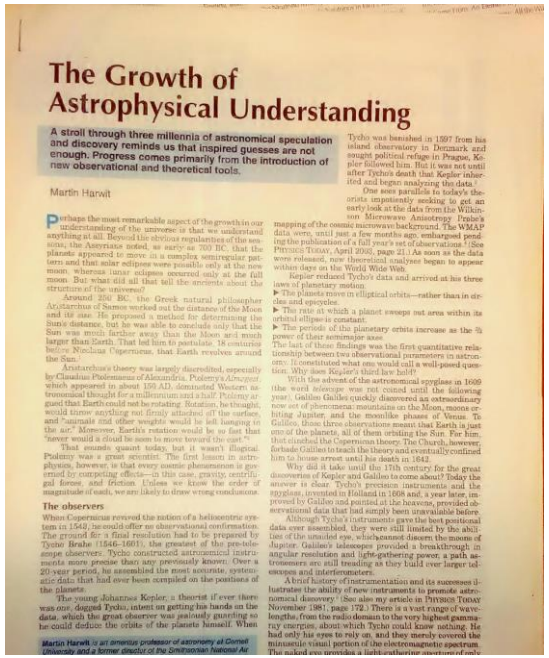


International AstroInformatics Association



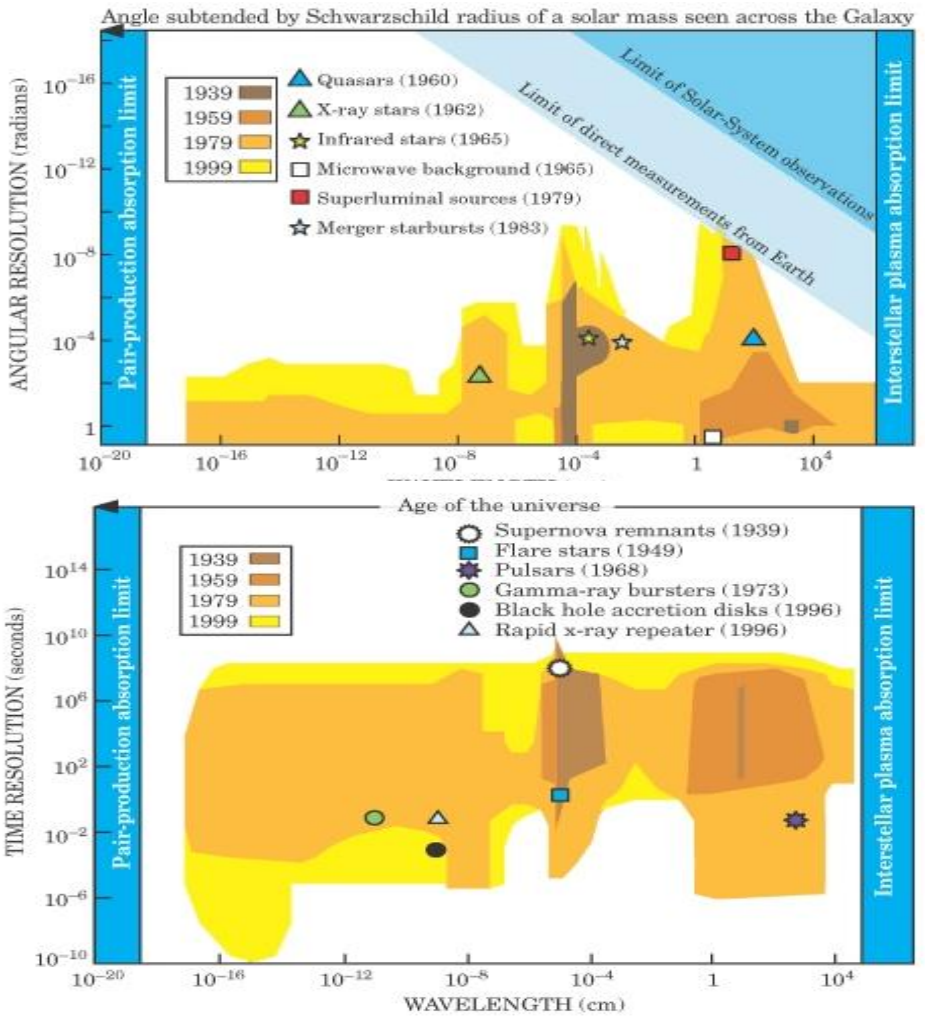
The discovery process is always linked to…

- Enlargements of the parameter space
- better sampling or coverage of the OPS


The Growth of Astrophysical Understanding

A stroll through three millennia of astronomical speculation and discovery reminds us that inspired guesses are not enough. Progress comes primarily from the introduction of new observational and theoretical tools.

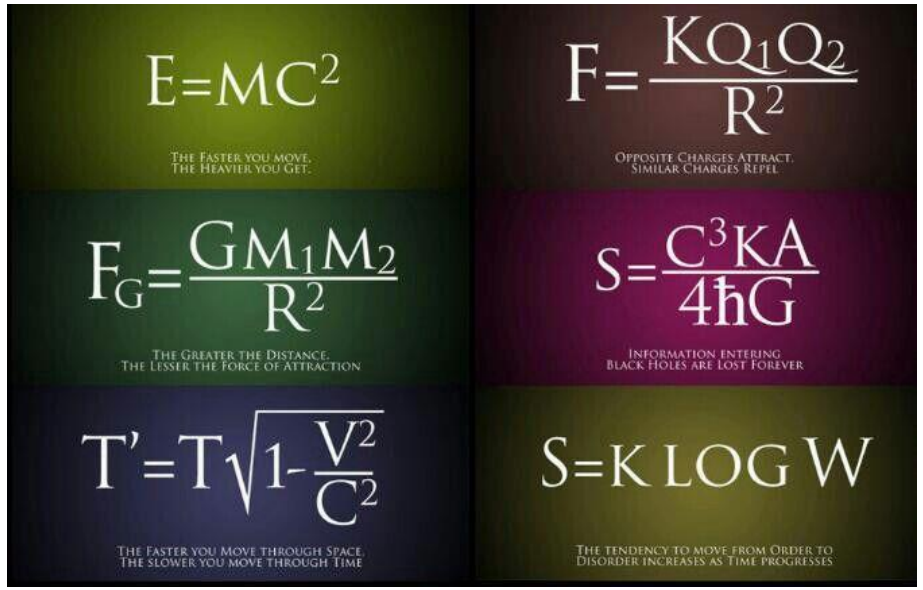
Martin Harwit



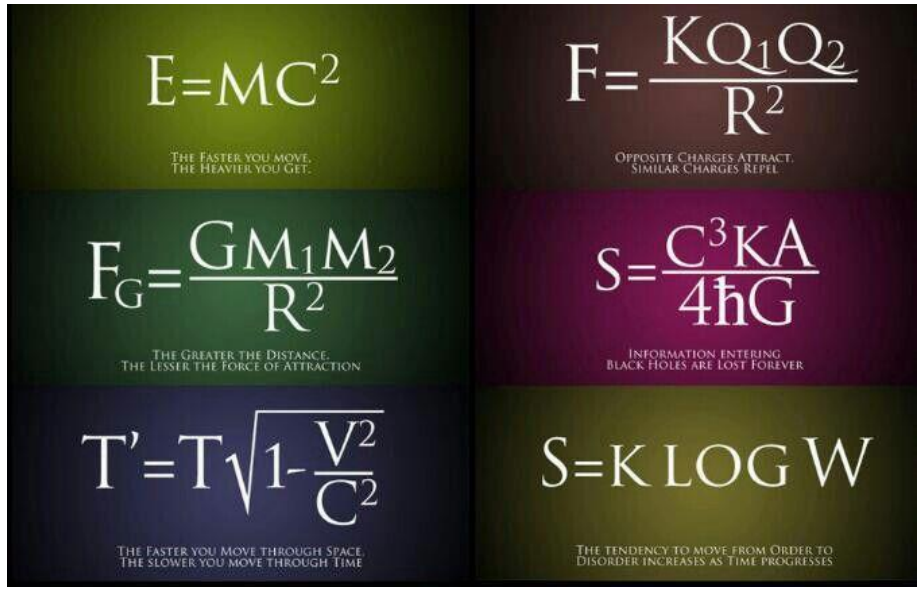
Driven by:

- Technological advances
- Better analytical or numerical tools

International AstroInformatics Association

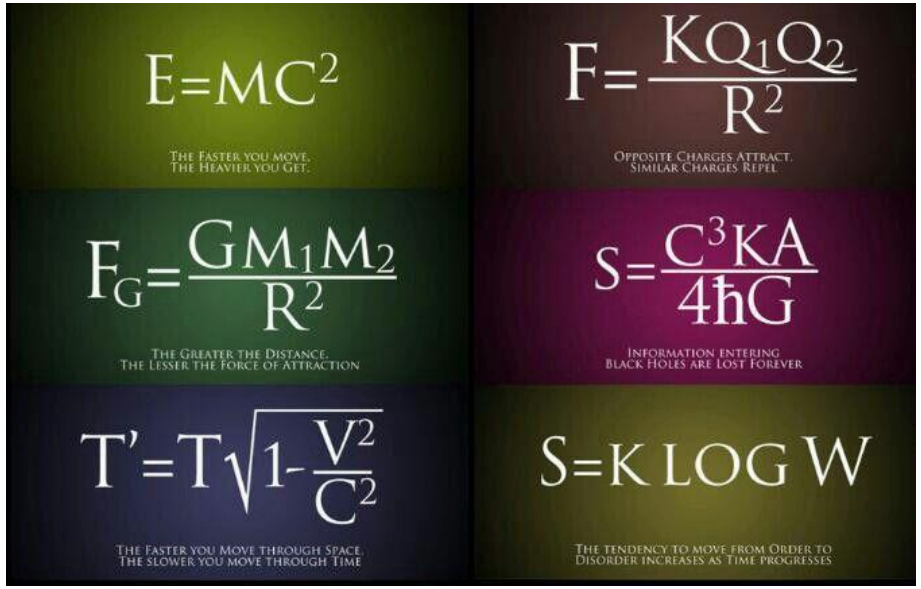


1. We cannot visualize patterns in more than 3-D …. And therefore…
very few (if any) physical (astrophysical, physiological, etc.) laws depend on more that two (max three) independent variables



1. We cannot visualize patterns in more than 3-D …. And therefore… very few (if any) physical (astrophysical, physiological, etc.) laws depend on more that two (max three) independent variables

2. **We do not easily recognize patterns** to which we have never been exposed before…



1. We cannot visualize patterns in more than 3-D …. And therefore… very few physical (astrophysical, ecc.) laws depend on more that two (max three) independent variables

2. **We do not easily recognize patterns** to which we have never been exposed before…

Failure of human brain classifier

Question: do we live in a **simple universe** or, rather, **OUR** description of the universe is biased by the fact that it is «**OUR DESCRIPTION**»?

Why can ML be useful?

Introduction of non linear (sigmoid) neurons transform ANN in almost perfect approximators….

$\boldsymbol{y} = f\left(\mathrm{g}\left(\mathrm{h}(\mathrm{i}\ldots((\mathbf{X})))\right)\right)$ minimize a cost $= \mathcal{L}(\boldsymbol{y}, \hat{\boldsymbol{y}})$ with respect to the weight matrices $\mathbf{W}_i$

Which define the functions f,g,h,i, …

Universal Approximation (Pseudo)**Theorem** (Haykin Pseudo - Theorem)

a neural network with a single layer can approximate any non-linear function to arbitrary accuracy.

Note: while this implies that only a single hidden layer is, in principle, sufficient for any problem, the dimension of this hidden layer may become intractably large for complex problems.

Most deep learning architectures (e.g. ***autoencoders***) stem out of the need to solve this problem (but the substance is the same as for feedforward fully connected networks) together **with the vanishing gradients problem**

The ML paradigm

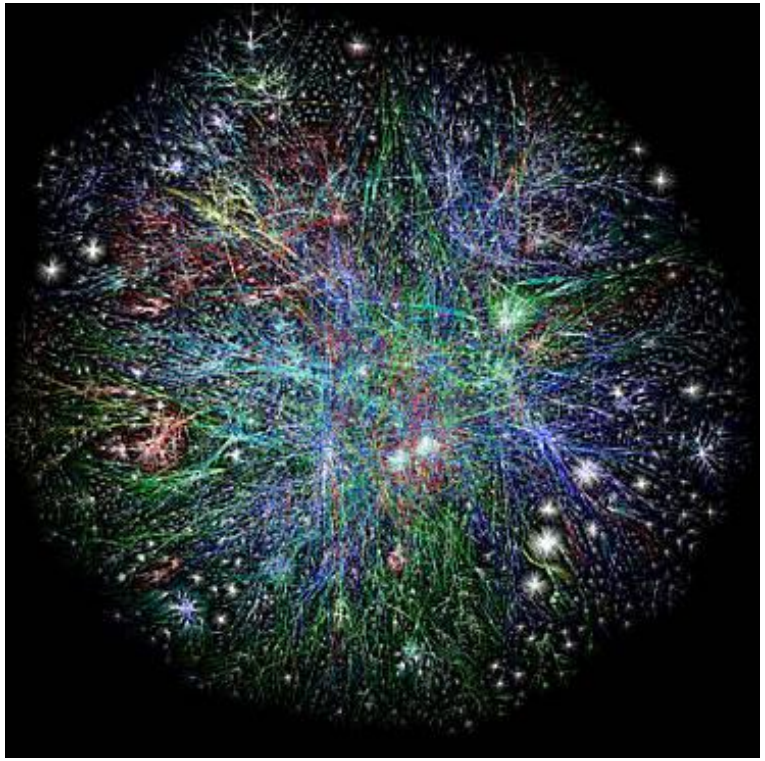
Underlying hypothesis

there exist an underlying unknown (often complex) multi-parametric function (with limited number of degrees (n) of freedom) which maps the input space onto the output space (target).

Find the proper **affine transformation** of the input space which allows to separate the problem

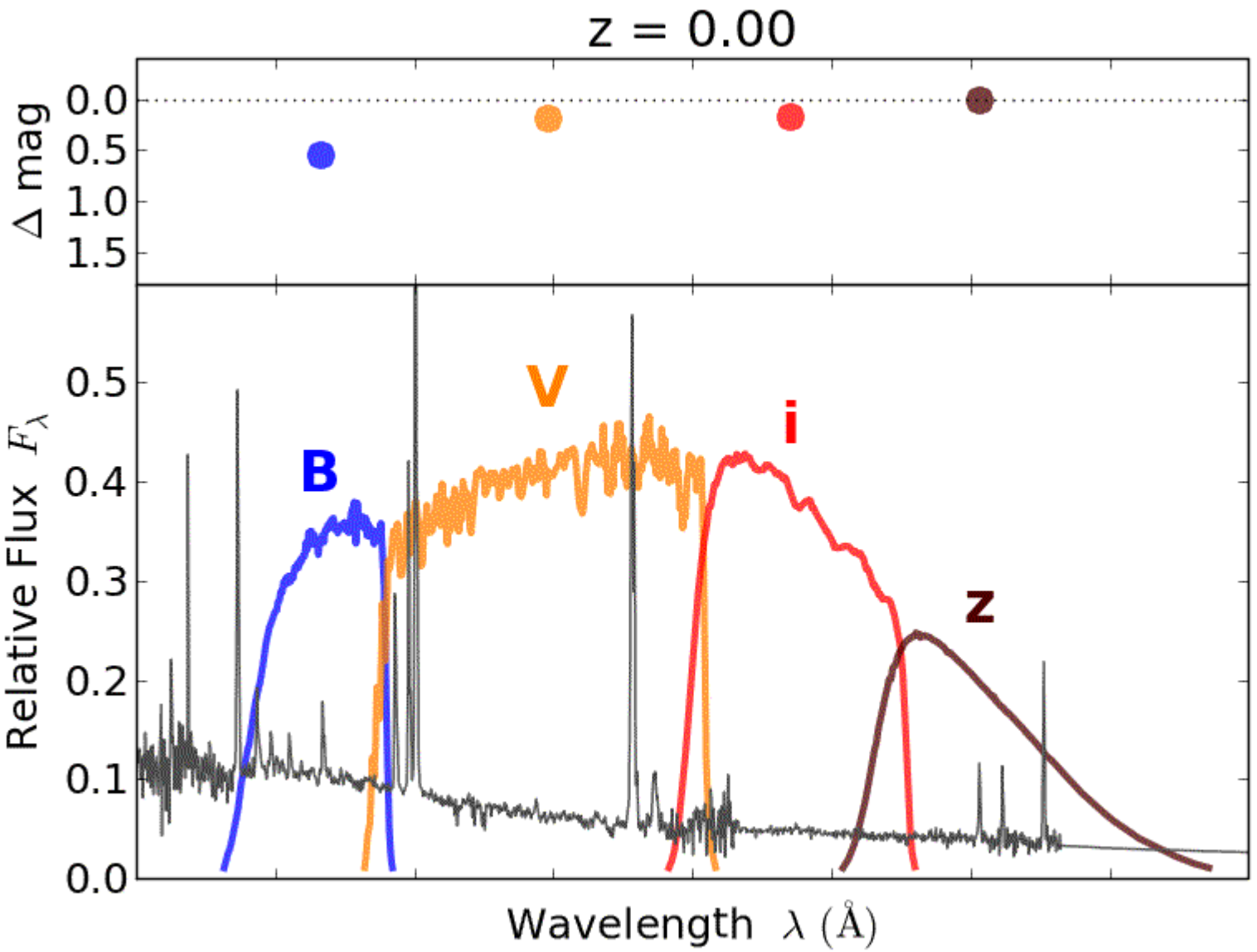
Identify the minimum number (N) of features (complex) which describe the problem

Usually: $N \geq n$



Proper Machine Learning may help understanding complex (n>3) physics (science in general) and to uncover a higher order of complexity.

First Example: Photometric redshifts (supervised problem)



Spectroscopic redshifts

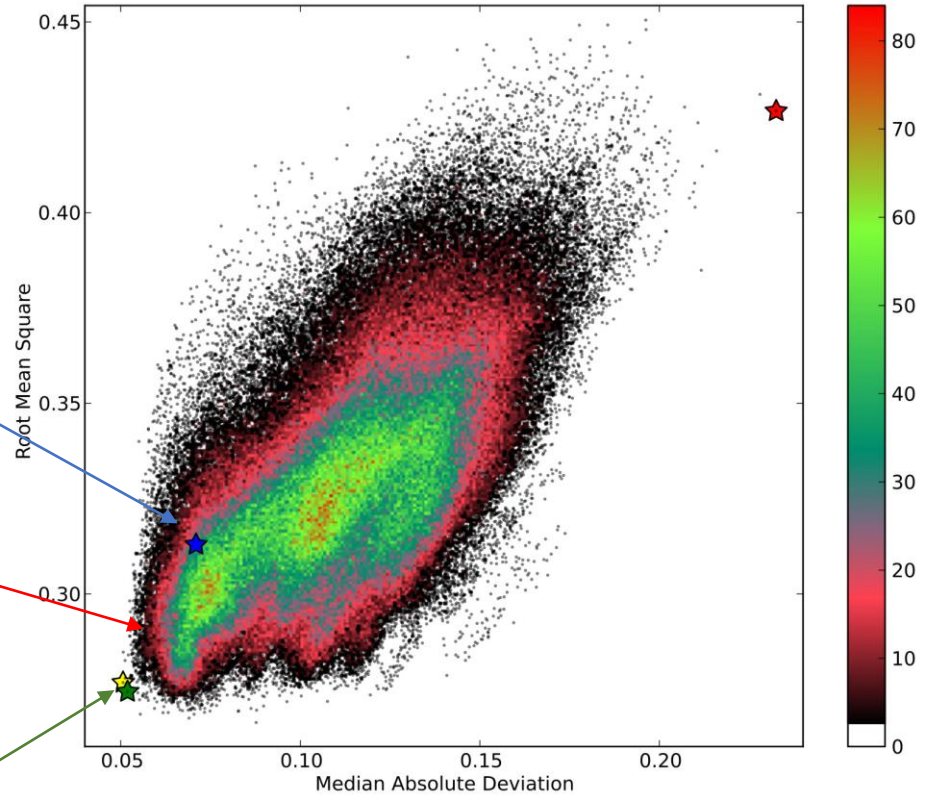
Accurate but troublesome to obtain in large quantities especially for distant galaxies

Photometric redshifts

Less accurate but much much easier to obtain

Laurino et al 2011,

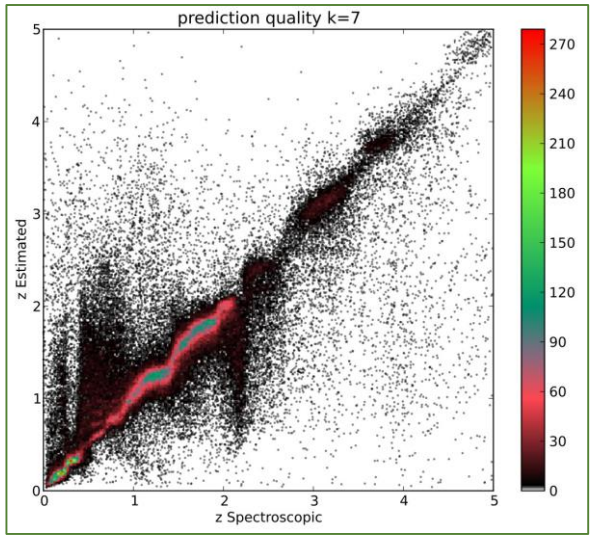
Brescia et al. 2019
ML with traditional approach (i.e. Expert saying which and how many (4) parameters)



QSOs from SDSS
Training set 300.000

A brute force ML approach (K. Polsterer et al., 2015)
All possible combinations of 4 parameters among a selection of 55

$$\frac{n!}{(n-r)!r!} = 341,055 \text{ combinations}$$



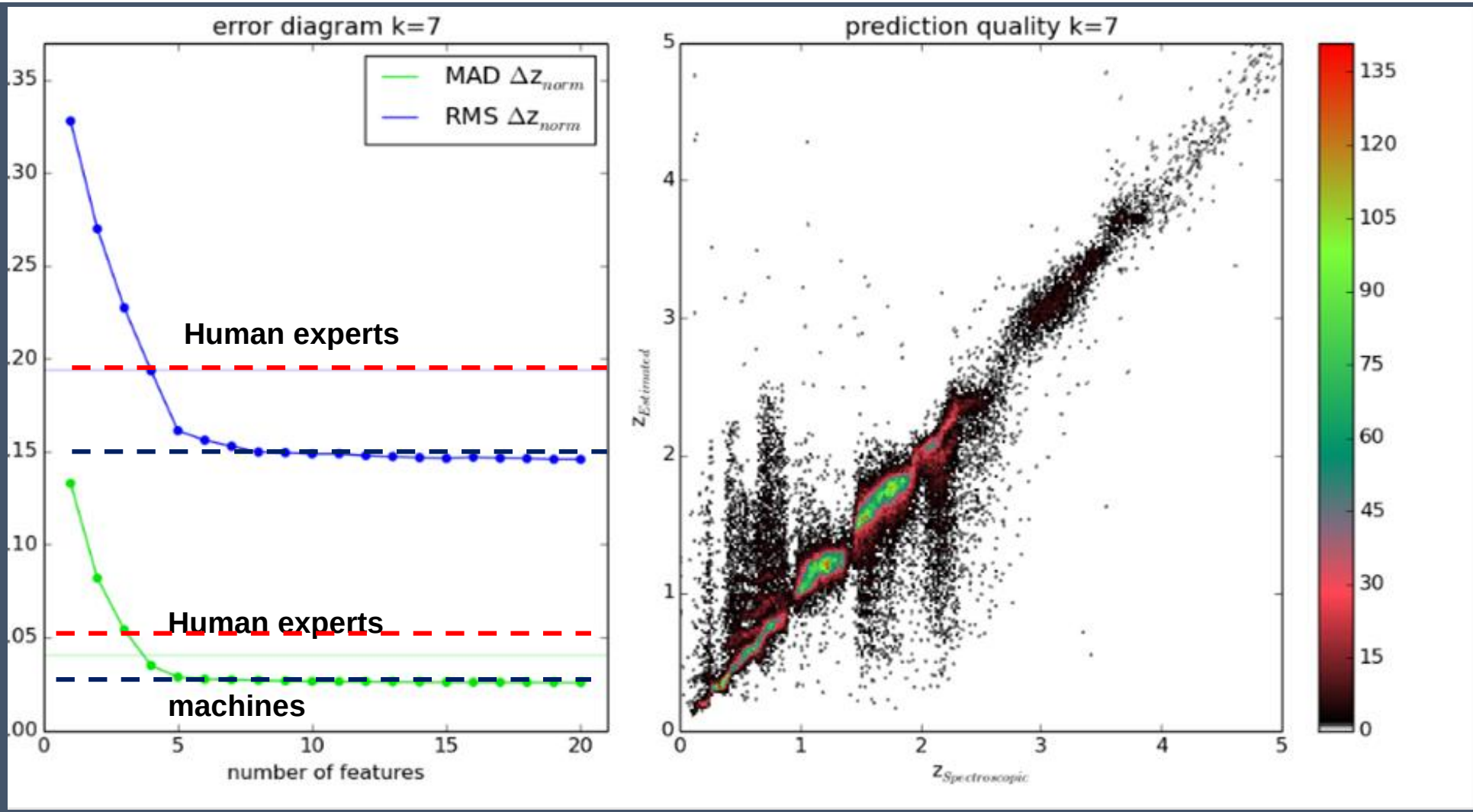
Best combination
$u_{model} - g_{model}$
g_{psf}-r_{model}
z_{psf}-r_{model}
i_{psf}-z_{model}

Which do not make any sense to a domain expert !!!

PSF, Petrosian, Total magnitudes + extinction + errors ….. 585 features…. Let us find the best combination of 10, 11, 12 etc… using FEATURE ADDITION

For just 10 features ….. **1,197,308,441,345,108,200,000** combinations

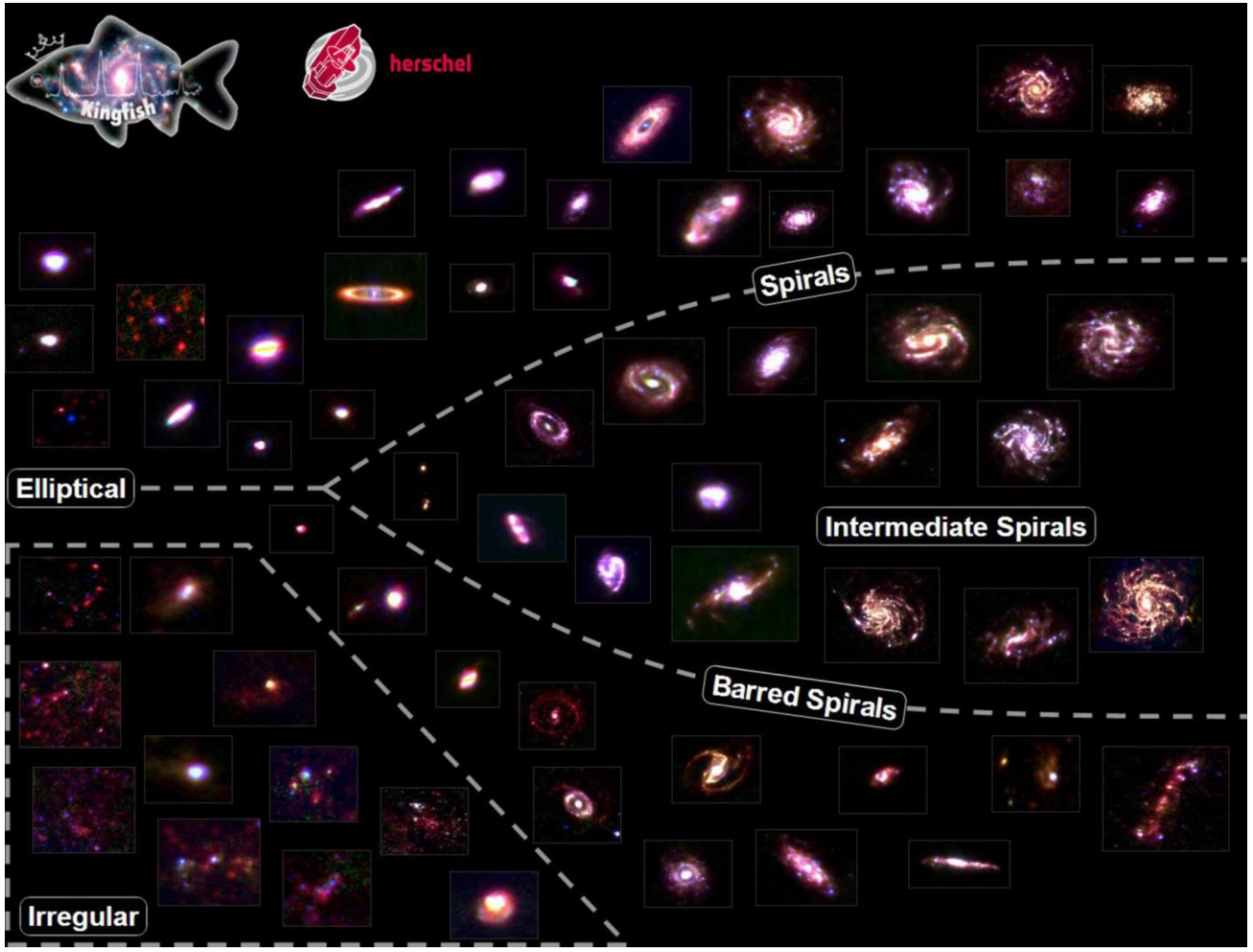
$$u_{psf} - g_{petr}$$
$$dered\left(z_{pdf}\right) - dered\left(i_{petr}\right)$$
$$dered\left(g_{psf}\right) - dered\left(r_{\mathrm{mod}}\right)$$
$$dered\left(r_{psf}\right) - dered\left(z_{\mathrm{mod}}\right)$$
$$\sqrt{\sigma^2_{g_{petr}} - \sigma^2_{r_{\mathrm{mod}el}}}$$
$$dered\left(r_{\mathrm{mod}}\right) - dered\left(i_{\mathrm{mod}}\right)$$
$$i_{psf} - i_{petr}$$
$$dered\left(z_{psf}\right) - dered\left(r_{petr}\right)$$
$$g_{\mathrm{mod}} - g_{petr}$$
$$\sqrt{\sigma^2_{g_{petr}} - \sigma^2_{r_{petr}}}$$



Information reaches a plateau at ca. very complex 10 features (N)

USE ML IN AN INNOVATIVE WAY:

- Look at old problems in an innovative way (it is not matter of using the latest ML model), by helping your TEAM to redefine the problem.



E.G

Galaxy classification.

Galaxy evolution
Galaxy dynamics
Galaxy formation
Cosmological models…..

Biased (KB only at optical NIR wav, resolution hence z, etc….)
Incomplete
Non physical (merger? AGN?, ecc.)

Still most works focus on reproducing it!

arXiv > astro-ph > arXiv:astro-ph/9502078

Search...

Help | Advanced

Astrophysics

[Submitted on 16 Feb 1995]

A COMPARATIVE STUDY OF MORPHOLOGICAL CLASSIFICATIONS OF APM GALAXIES

A. Naim, O. Lahav, R. J. Buta, H. G. Corwin Jr., G. de Vaucouleurs, A. Dressler, J. P. Huchra, S. van den Bergh, S. Raychaudhury, L. Sodre Jr., M. C. Storrie-Lombardi

We investigate the consistency of visual morphological classifications of galaxies by comparing classifications for 831 galaxies from six independent observers. The galaxies were classified on laser print copy images or on computer screen produced from scans with the Automated Plate Measuring (APM) machine. Classifications are compared using the Revised Hubble numerical type index T. We find that individual observers agree with one another with rms combined dispersions of between 1.3 and 2.3 type units, typically about 1.8 units. The dispersions tend to decrease slightly with increasing angular diameter and, in some cases, with increasing axial ratio (b/a). The agreement between independent observers is reasonably good but the scatter is non-negligible. In spite of the scatter the Revised Hubble T system can be used to train an automated galaxy classifier, e.g. an Artificial Neural Network, to handle the large number of galaxy images that are being compiled in the APM and other surveys.

There is an intrinsic error in the KB of 1.8 classification bins !!!!

Almost no one mentions it.

Subproblem:

Technically simple
Conceptually: far reaching

Let us try to find AGN (especially low luminosity AGN since QSO et al are easy)

Reference persons: Lars Doorenbros (CS), Stefano Cavuoti (Astroinf.), Torbaniuk (AGN expert) ACCEPTED

ULISSE: A Tool for One-shot Sky Exploration and its Application to Active Galactic Nuclei Detection

Lars Doorenbos[1], Olena Torbaniuk[2,3], Stefano Cavuoti[4,5], Maurizio Paolillo[2,4,5], Giuseppe Longo[2], Massimo Brescia[4], Raphael Sznitman[1] and Pablo Márquez-Neila[1]

[1] AIMI, ARTORG Center, University of Bern, Murtenstrasse 50, CH-3008 Bern, Switzerland
e-mail: lars.doorenbos@unibe.ch
[2] Department of Physics, University Federico II, Strada Vicinale Cupa Cintia, 21, 80126 Napoli, Italy
[3] Main Astronomical Observatory of National Academy of Sciences, 27 Akademika Zabolotnoho str., 03143 Kyiv, Ukraine
[4] INAF - Astronomical Observatory of Capodimonte, Salita Moiariello 16, I-80131 Napoli, Italy
[5] INFN - Sezione di Napoli, via Cinthia 9, 80126 Napoli, Italy

EfficientNet-b 0 (a type of CNN architecture -Tan & Le 2019), trained for classification on ImageNet as the CNN from which we obtain the features.

Its penultimate layer consists of 1280 channels, leading to a 1280-dimensional feature descriptor for each image (features are extracted from the model, and were derived from natural images rather than astronomical ones. Hence, they are not directly interpretable).

These pretrained features are used by ULISSE to identify objects with similar properties. This is done by performing a similarity search in the feature space.

Nonetheless, we can get an idea of the patterns individual features are looking for, by looking at the images in our dataset which most strongly activate them

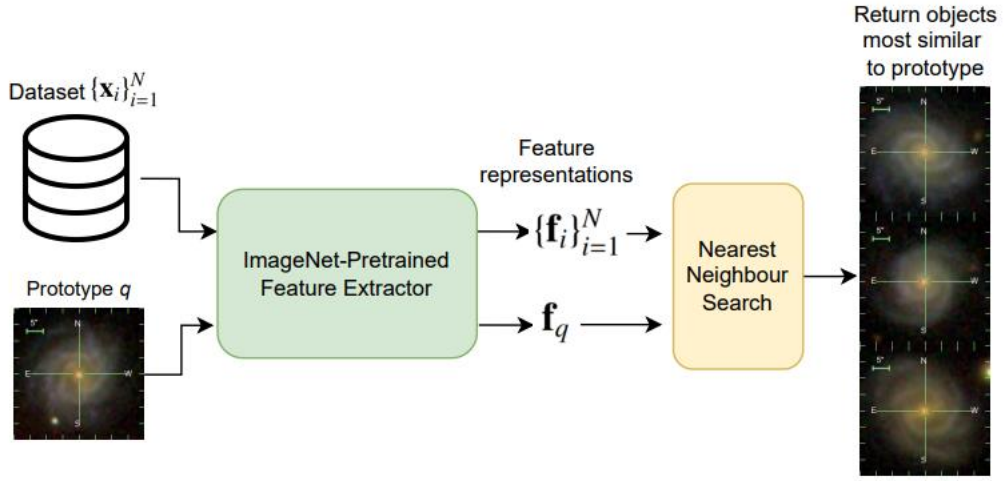


Fig. 2. Overview of ULISSE.

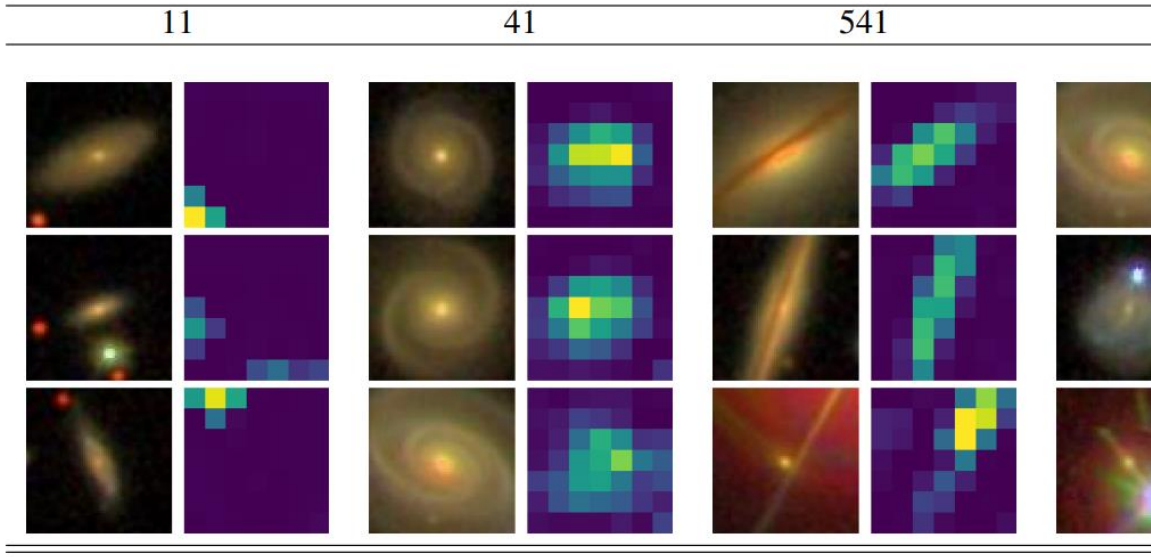


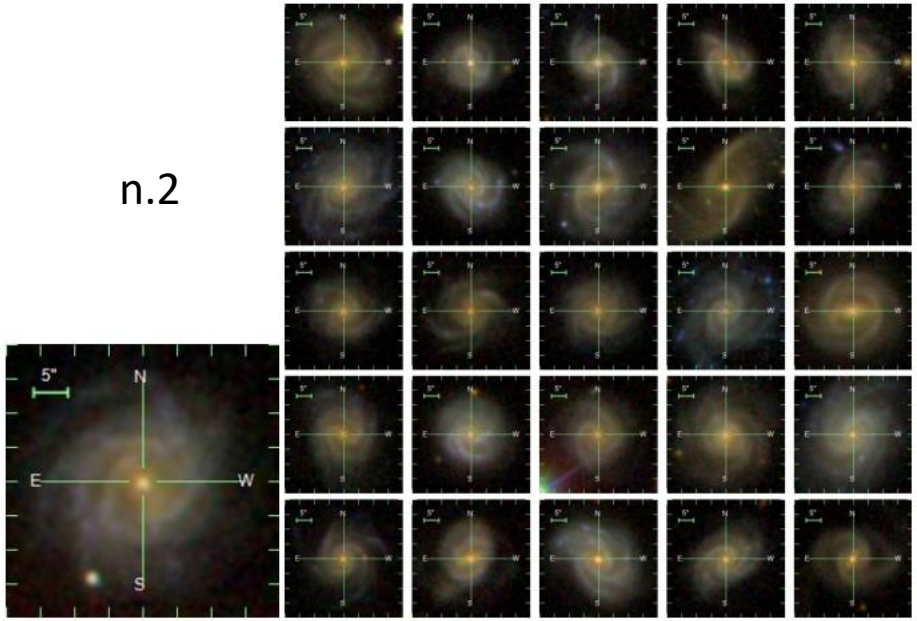
Fig. 1. The three objects in our sample which most strongly activate features 11, 41, 541, 835 together with their feature maps. We provide these for all 1280 features at the http://dame.na

Table 1. The summary of the different datasets studied in this work.

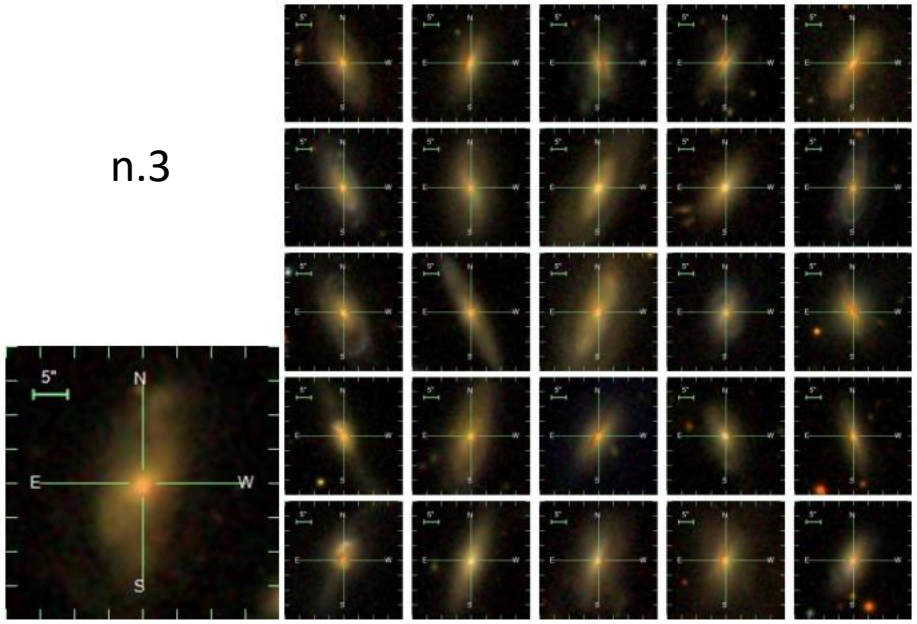
Sample			Fraction		
			Entire	X-ray MOC	Random
Criteria	BPT	AGN	12.0 %	11.8 %	12.0 %
		Composite	5.8 %	5.5 %	5.8 %
		SFG	44.0 %	41.5 %	44.1 %
		Unclassified	38.2 %	41.2 %	38.1 %
	X-ray	AGN	0.2 %	4.0 %	0.2 %
		non-AGN	5.6 %	96.0 %	5.6 %
		Unknown	94.2 %	—	94.2 %
Number of objects			703 422	40 889	99 991

Notes. The fractions represent the percentage of objects in each dataset classified as AGN, SFG or composite according to the optical BPT-diagram or X-ray AGN/non-AGN by X-ray selection criteria (see details in the text). Unknown class indicates the fraction of objects which have not been observed by *XMM-Newton*.

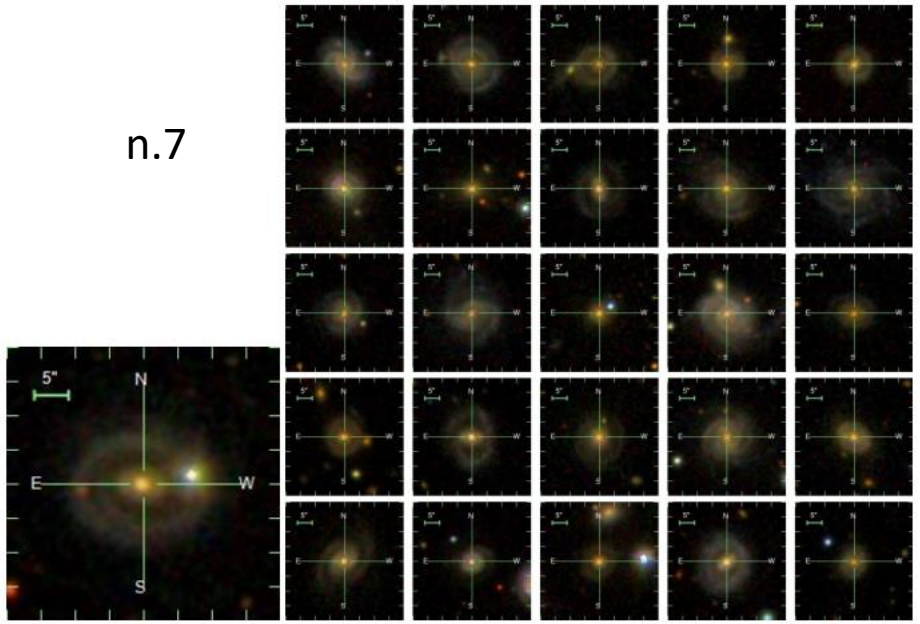
n.2

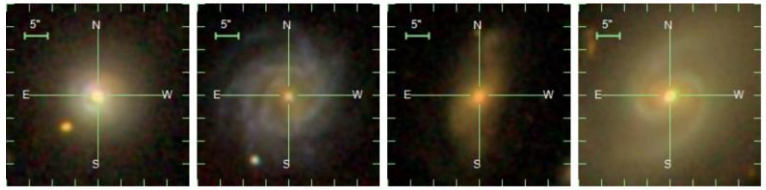


n.3

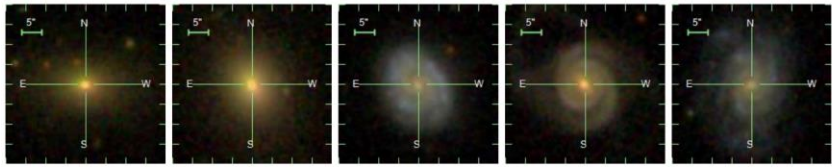


n.7

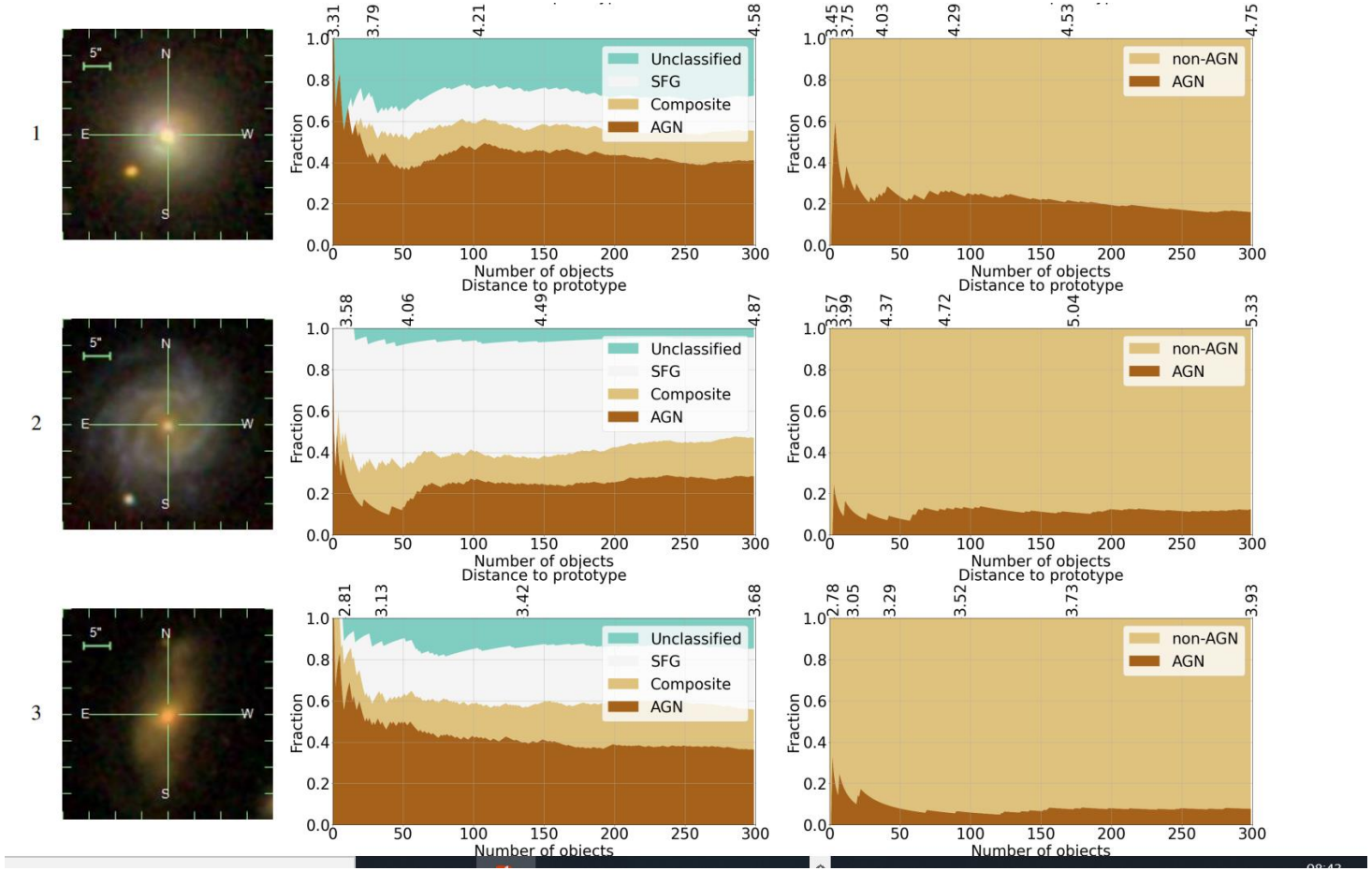


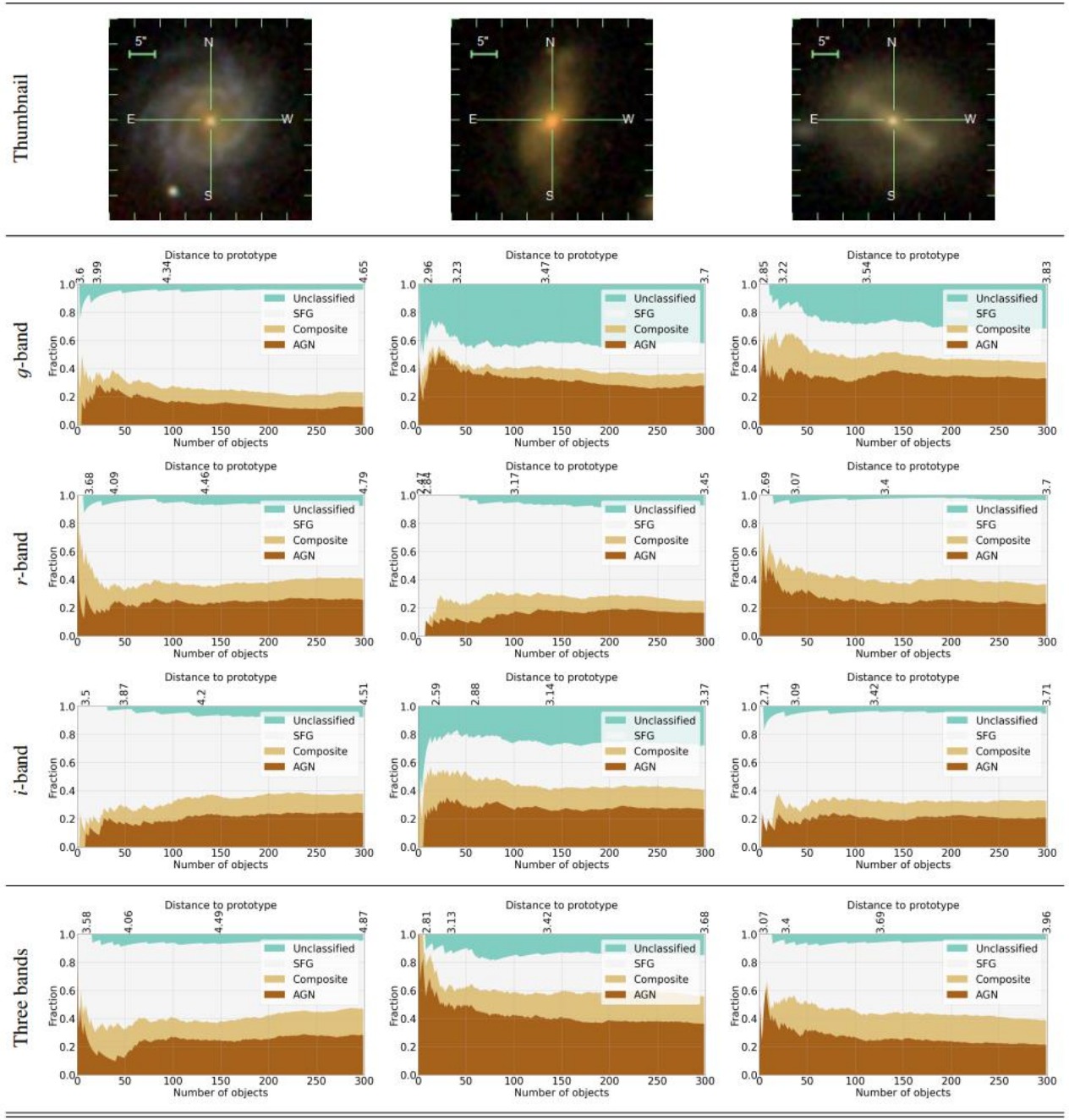
#	1	2	3	4
SDSS	J032525.36-060837.8	J164607.00+422737.4	J153621.30+222913.6	J133548.24+025956.1
RA	51.35569	251.52917	234.08879	203.95103
DEC	−6.14386	42.46041	22.48712	2.99892
Redshift	0.034	0.049	0.089	0.022
Thumbnails				
SDSS class	Galaxy	Galaxy	QSO	Galaxy
SDSS subclass	AGN	Starforming	AGN Broadline	AGN
BPT class	AGN	SFG	AGN	SFG
X-ray class	AGN	AGN	AGN	AGN

AGN prototypes

#	9	10	11	12	13
SDSS	J115928.62+423542.8	J151121.53+072250.6	J134059.80+302058.0	J083114.54+524224.8	J151105.13+053112.7
RA	179.86926	227.83972	205.24919	127.81060	227.77139
DEC	42.59522	7.38073	30.34947	52.70690	5.52020
Redshift	0.114	0.044	0.040	0.064	0.035
Thumbnails					
SDSS class	Galaxy	Galaxy	Galaxy	Galaxy	Galaxy
SDSS subclass	—	—	Starforming	Starforming	Starforming
BPT class	Unclassified	Unclassified	SFG	SFG	SFG
X-ray class	non-AGN	non-AGN	non-AGN	non-AGN	non-AGN

Non - AGN prototypes



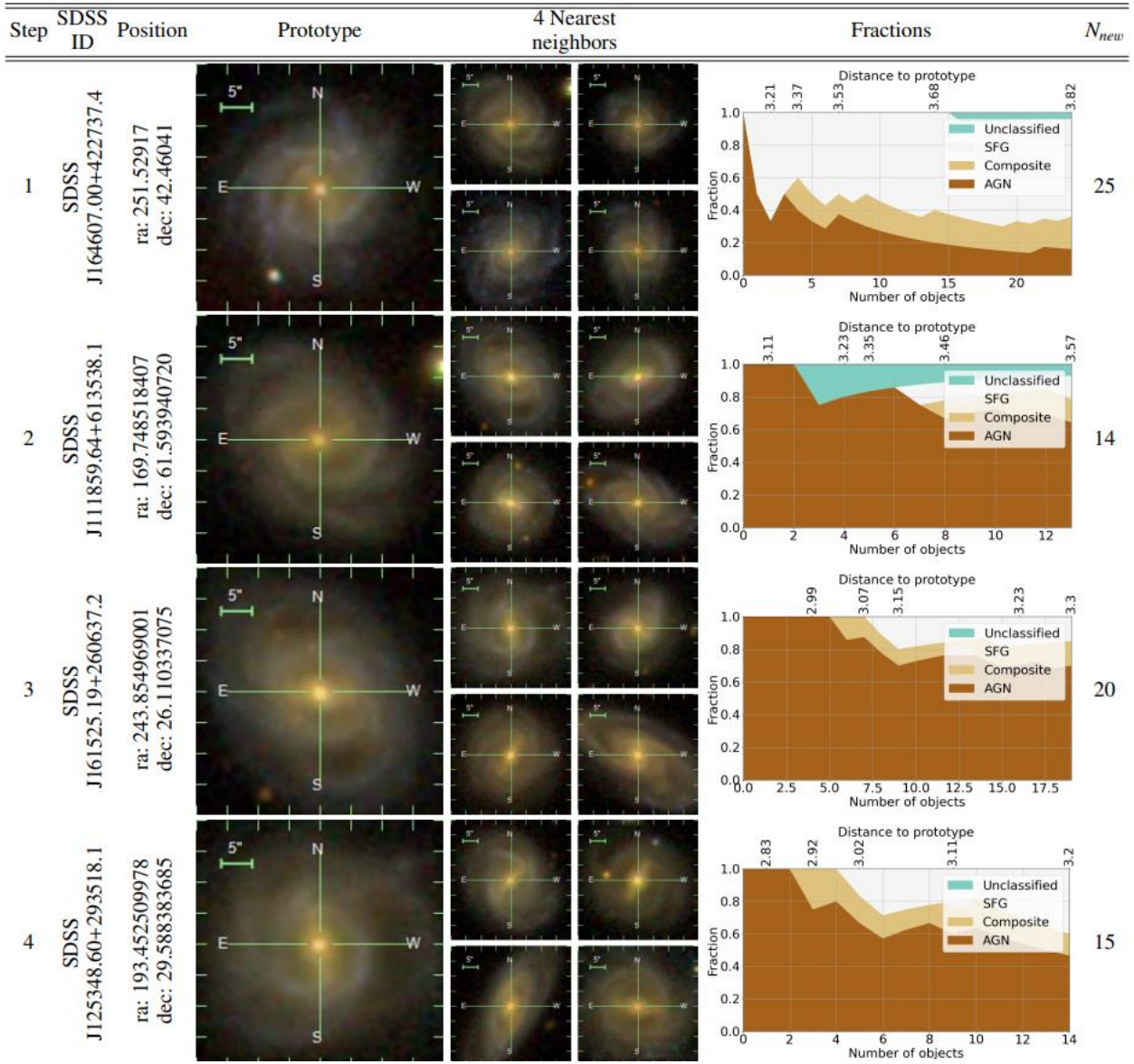

Thumbnail
g-band
r-band
i-band
Three bands
5"
N
E
W
S
Distance to prototype
Unclassified
SFG
Composite
AGN
Fraction
Number of objects

Results from recursive process on single band images

Using this set-up, for the relatively unsuccessful prototype #2.

We obtained a total of 89 objects excluding duplicates in five iterations, of which 38 are AGN (42.7 %, see Table A.5).

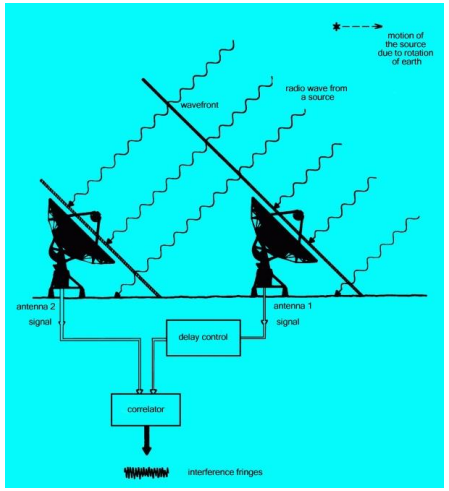
In contrast, the resulting AGN fraction setting n = 89 directly for prototype #2 would only result in 25.8 % AGN.

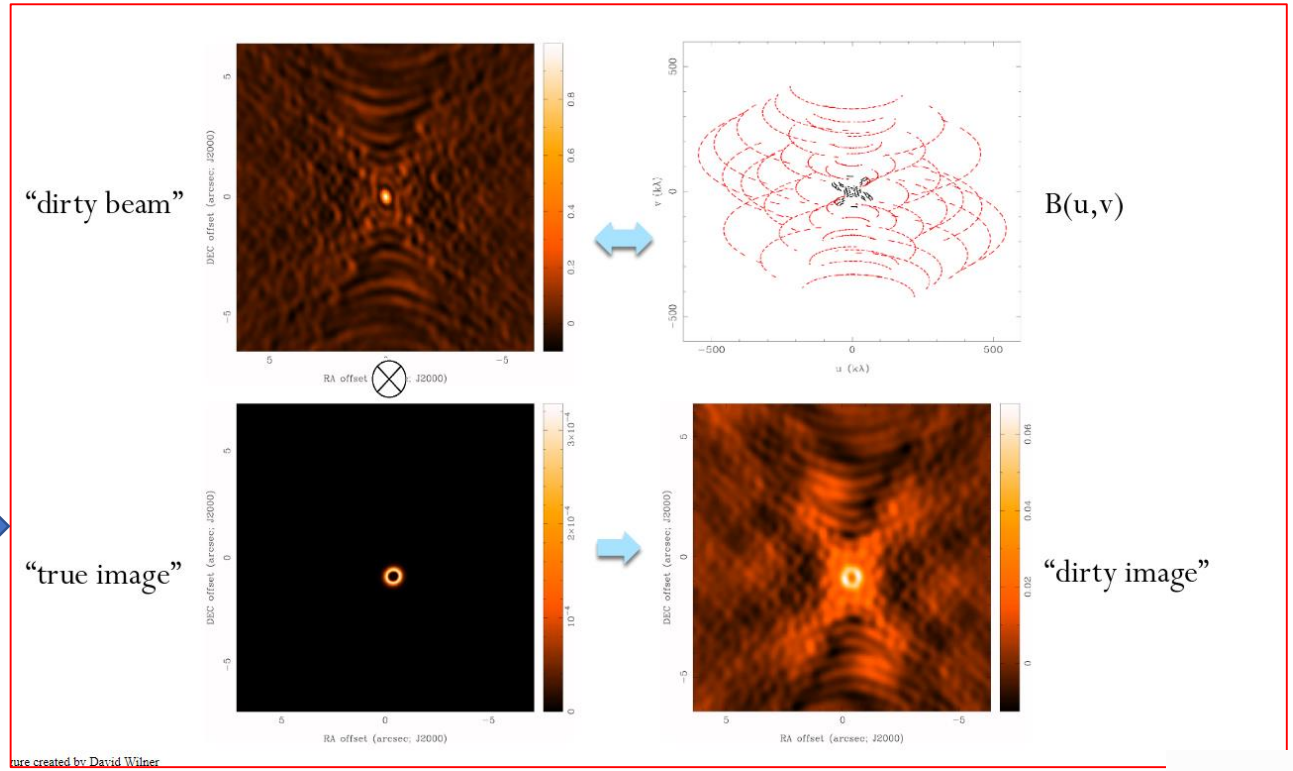


Second example: Finding and characterizing objects in radiointerferometric data cubes

ALMA

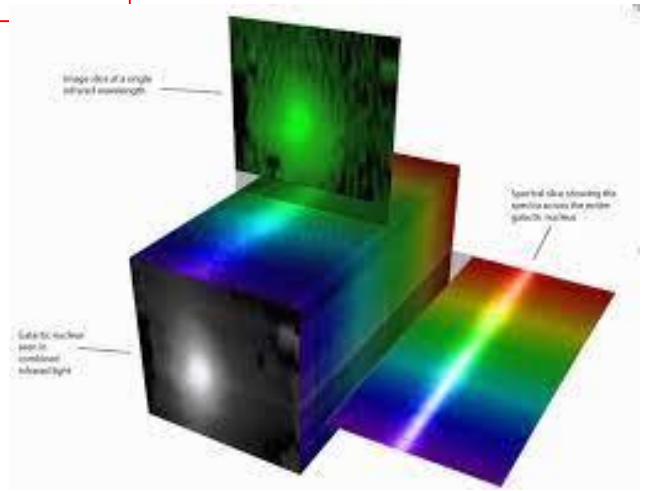
Atacama Large sub Millimeter Array





Data cubes

(RA,dec and λ)



Huge problems:

- Size of the datacube
- Number of data cubes
- Asymmetric beam (lack of u,v coverage)
- Correlated noise structures (artefact of image reconstruction process)
- noise changes with frequency, ecc.), ecc

Old professor (s)

PhD Students:

Michele delli Veneri

(***above…. After finishing the work and three years of insomnia !)***

Lukasz Tychoniec
(simulations)

Reference person: Michele delli Veneri

MNRAS **000**, 1–22 (2022) Preprint 3 September 2022 Compiled using MNRAS LATEX style file v3.0

3D Detection and Characterisation of ALMA Sources through Deep Learning

Michele Delli Veneri,[1,2]★ Łukasz Tychoniec,[3]† Fabrizia Guglielmetti [3], Giuseppe Longo [2], Eric Villard [3]
[1] *INFN Section of Naples, Napoli, via Cintia, 1, Italy, 80126*
[2] *Department of Electrical Engineering and Information Technology, University of Naples "Federico II", Via Claudio, 21, 80125 Naples NA, Italy*
[3] *ESO, Karl-Schwarzschild-Straße 2, 85748 Garching bei München*
[4] *Department of Physics "Ettore Pancini", University of Naples "Federico II" , Via Cintia, 1, Italy, 80126*

Technically very complex
Conceptually: traditional (simple)

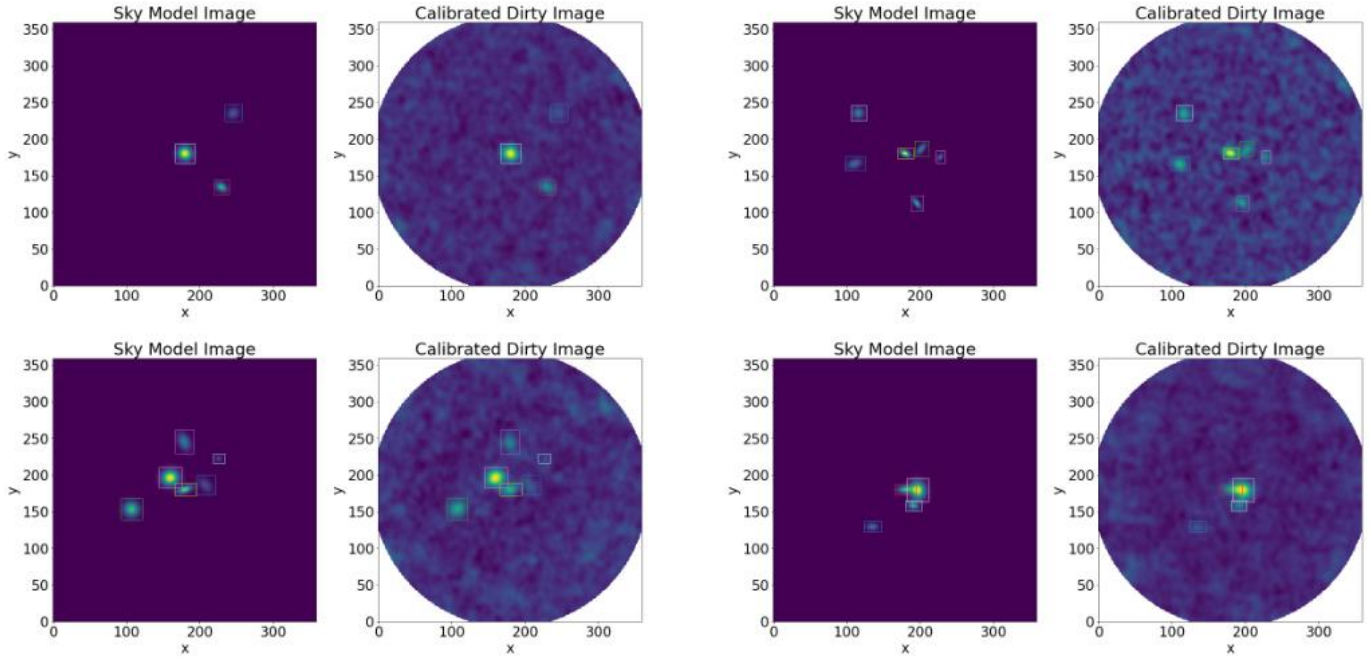
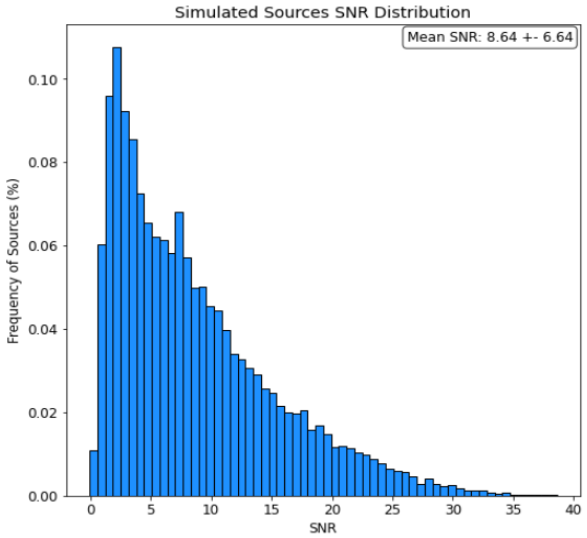


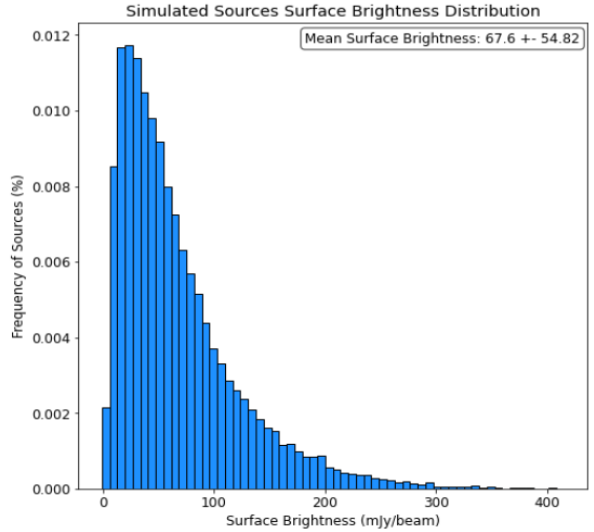
Figure 1. Several examples of frequency stacked dirty/clean cube pairs generated through our simulation code. Sources within the cubes are outlined with colored bounding boxes.

Lack of training data

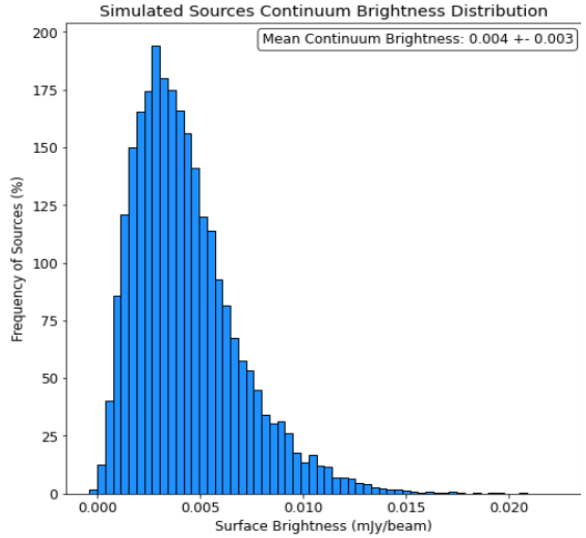
Trained on 20.000 simulated and realistic ALMA data cubes



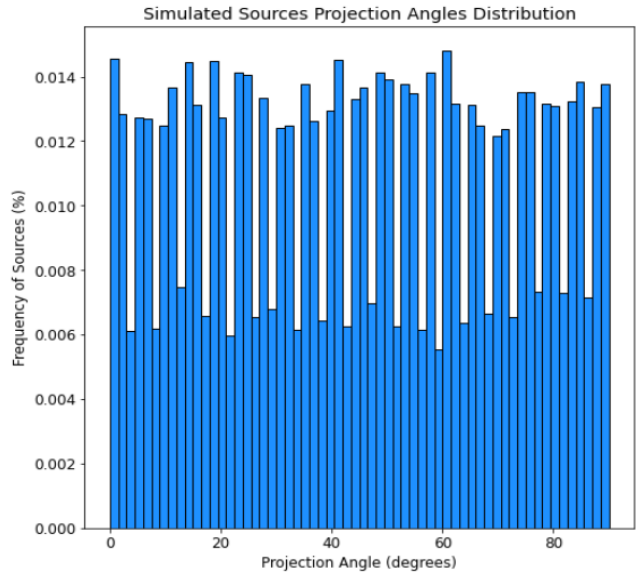
(a) Distribution of the Signal to Noise Ratio of the simulated sources: fraction of simulated sources versus measured SNR (see Eq. 17). The box in the top right corner shows the mean SNR ± its standard deviation.



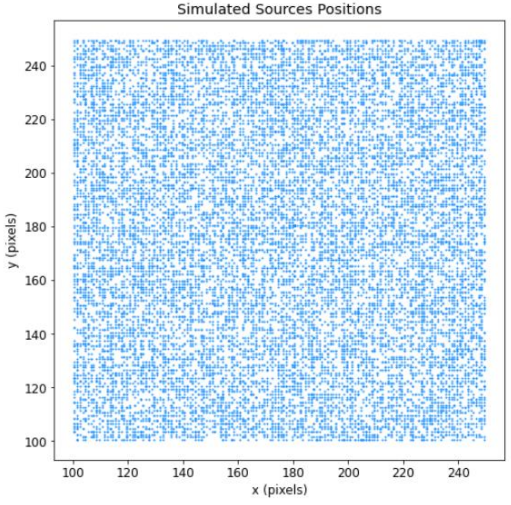
(b) Distribution of the the total brightness of the simulated sources. On the x-axis, the measured total brightness [mJy / beam] is obtained by summing the voxel values in the dirty cubes within sources bounding boxes. On the y-axis, the fraction of simulated sources is provided. The box in the top right corner shows the mean brightness ± its standard deviation.



(c) Distribution of the continuum mean brightness of the simulated sources On the x-axis, the measured mean continuum brightness [mJy / beam] i obtained by selecting all voxels within the x and y limits of the source bounding boxes but outside their boundaries in frequency $[z - fwhm_z, z \cdot fwhm_z]$. On the y-axis, the fraction of simulated sources is shown. The bo



(d) Distribution of the projection angles of the simulated sources: fraction of simulated sources versus projection angles in degrees



(b) Scatter plot showing uniformity in the positions on the xy plane of the simulated sources.

The full pipeline created (not implemented)by delli Veneri et al.

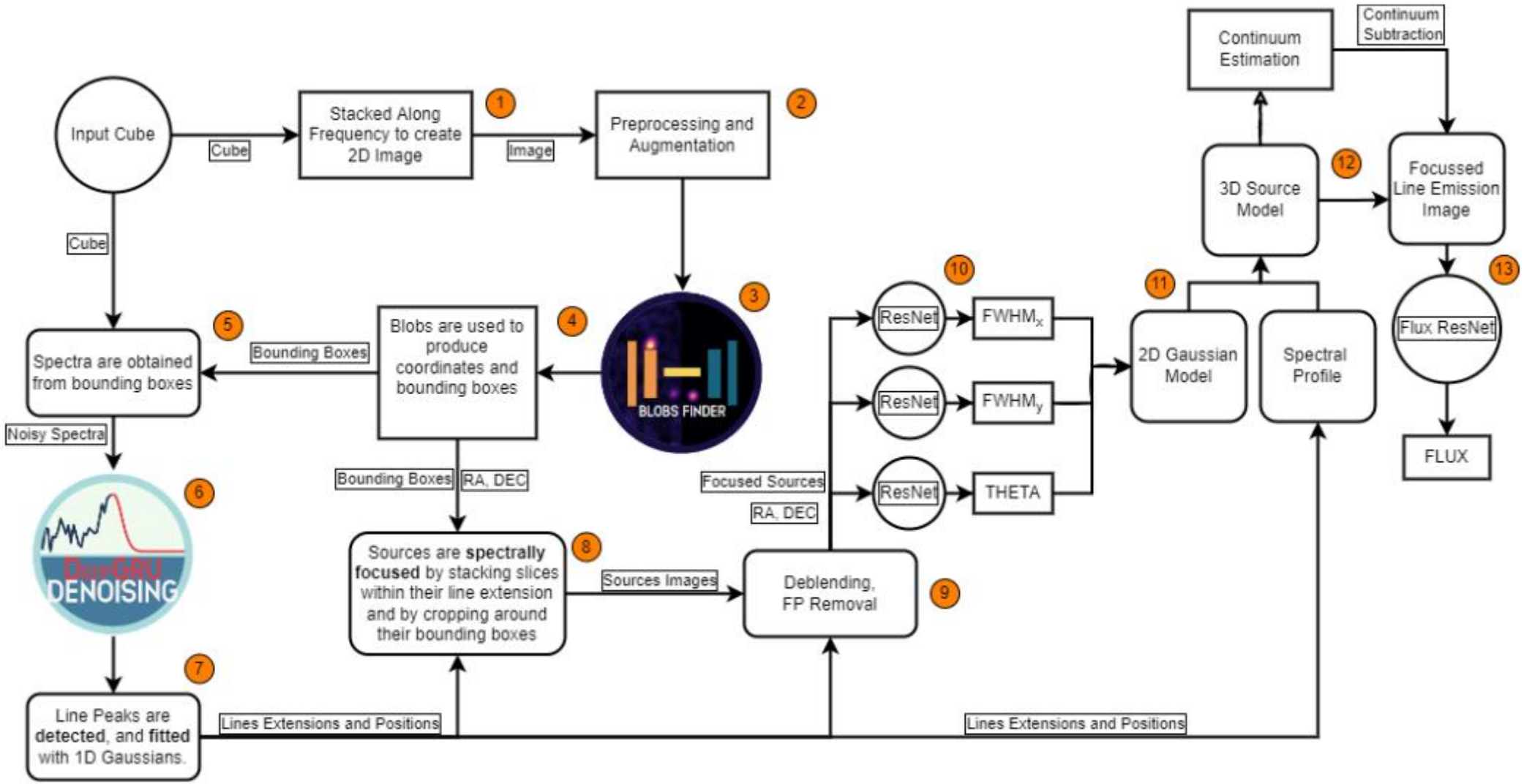


Figure 9. The full pipeline schema. Numbers show the logical flow of the data within the pipeline.

1-st – DL module – «Blobs finder»

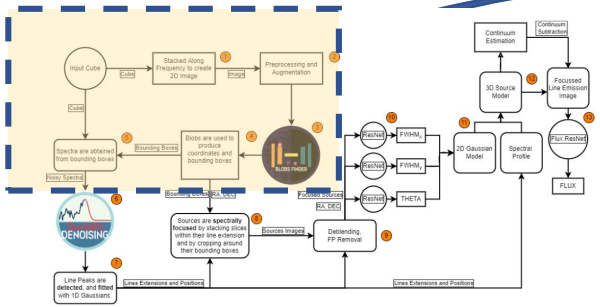
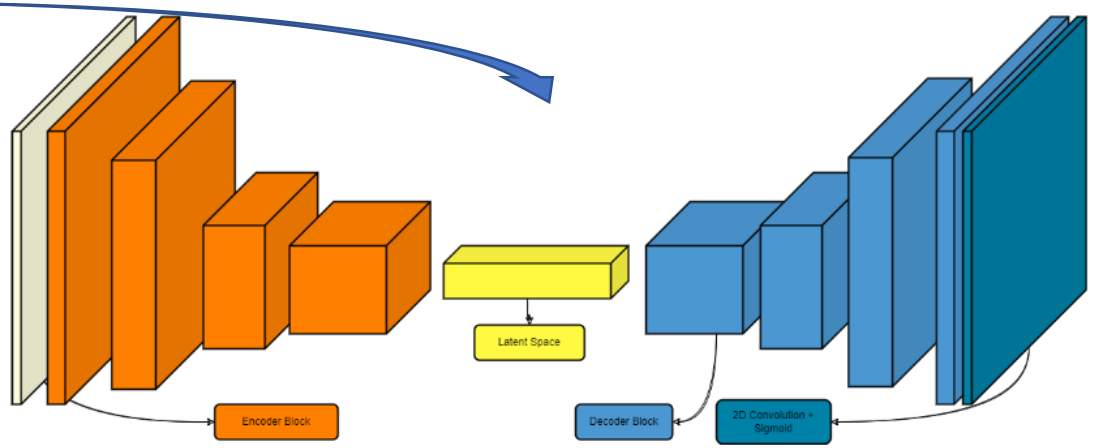
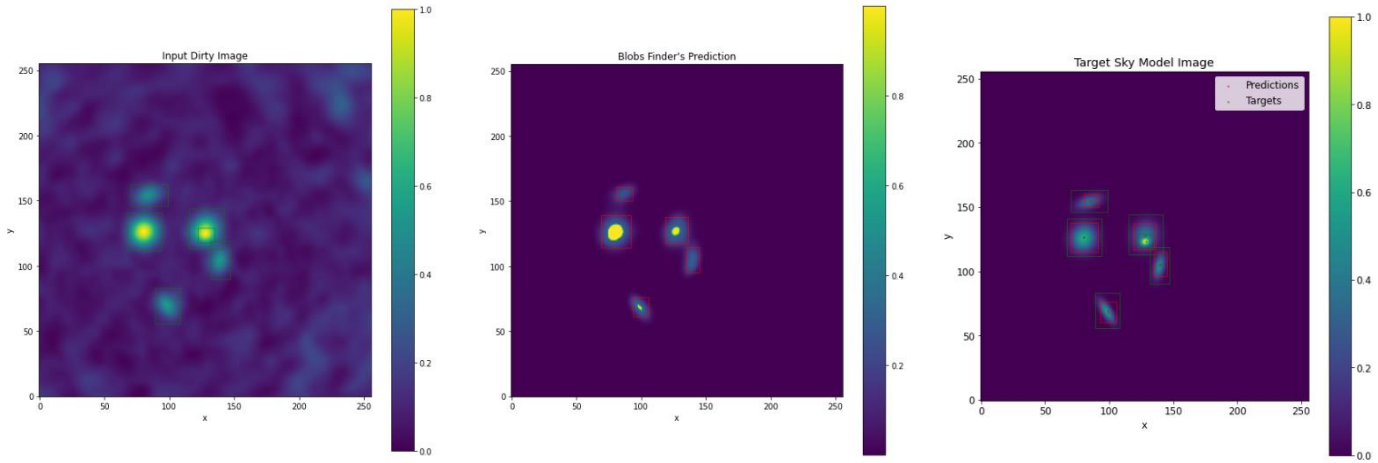


Figure 9. The full pipeline schema. Numbers show the logical flow of the data within the pipeline.



Deep Convolutional autoencoder



Bounding boxes

Encoder:

- 4 CNN blocks (kernel 3, stride 2, leaky RELU)
  2 D normalization batch
- CNN (stride 1, kernel 3, leaky RELU)
  2 dnORMALIZATION BATCH
- Fully connected layer (size 1024)

Decoder

- 4 deconvolutional blocks and a final block (2D bilinear interpolation, stride 2, Leaky ReLU, 2D Batch Normalization layer (upsampling block)
- 2D Transposed Convolution layer (stride 2, kernel size 3, Leaky ReLU, 2D Batch Normalization layer (learnable upsampling block).
- output of the up-sampling block and learnable up-sampling block are tconcatenated and passed to a convolutional block (2D Convolution layer, stride 2, kernel 3, Leaky ReLU, 2D Batch Normalization layer)
- 2D Convolution layer (stride 1, kernel size 3,) final block is a 2D Convolution layer (stride 1, kernel 1) followed by a Sigmoid activation function.

Block Name	Input Size	Output Size
Conv Block 1	$[b, 1, 256, 256]$	$[b, 8, 128, 128]$
Conv Block 2	$[b, 8, 128, 128]$	$[b, 16, 64, 64]$
Conv Block 3	$[b, 16, 64, 64]$	$[b, 32, 32, 32]$
Conv Block 4	$[b, 32, 32, 32]$	$[b, 64, 16, 16]$
FC 1	$[b, 64 \times 16 \times 16]$	$[b, 1024]$
FC 2	$[b, 1024]$	$[b, 64 \times 16 \times 16]$
DeConv Block 1	$[b, 64, 16, 16]$	$[b, 32, 32, 32]$
DeConv Block 2	$[b, 32, 32, 32]$	$[b, 16, 64, 64]$
DeConv Block 3	$[b, 16, 64, 64]$	$[b, 8, 128, 128]$
DeConv Block 4	$[b, 8, 128, 128]$	$[b, 1, 256, 256]$
Final Block	$[b, 1, 256, 256]$	$[b, 1, 256, 256]$

Table 3. Input and Output shapes for each layer of Blobs Finder, where b indicates the batch size, and the horizontal line separates the Encoder from the Deconder network.

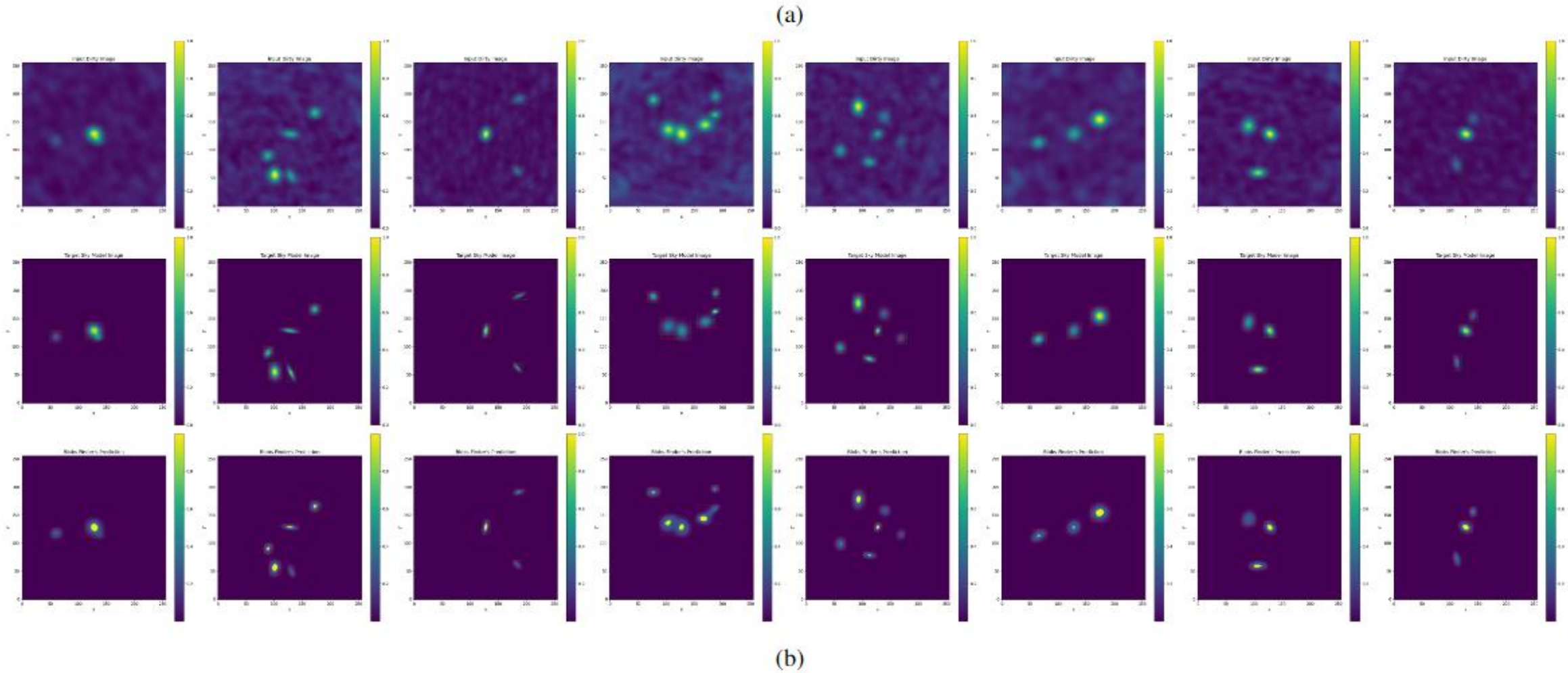


Figure 18. Examples of Blobs Finder predictions on the Test Set. The first row shows input integrated dirty cubes, the middle row the target sky models, and the bottom row, Blobs Finder predicted 2D Source Probability maps. In green are outlined (in the dirty and sky models images) the true bounding boxes, while in red the predicted bounding boxes extracted by thresholding the probability maps.

2-nd – DL module – Spectral focusing (RNN : DEEP GRU – Gated Recurrent Unit)

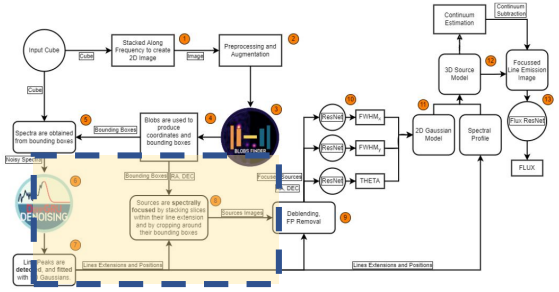


Figure 9. The full pipeline schema. Numbers show the logical flow of the data within the pipeline.

The Deep Gated Recurrent Unit denoises the standardized spectra and outputs 1D probabilistic maps of source emission lines or cleaned spectra.

i) feed to the ResNets the best possible input image of potential candidates

(ii) to deblend the sources

(iii) to remove most false positives.

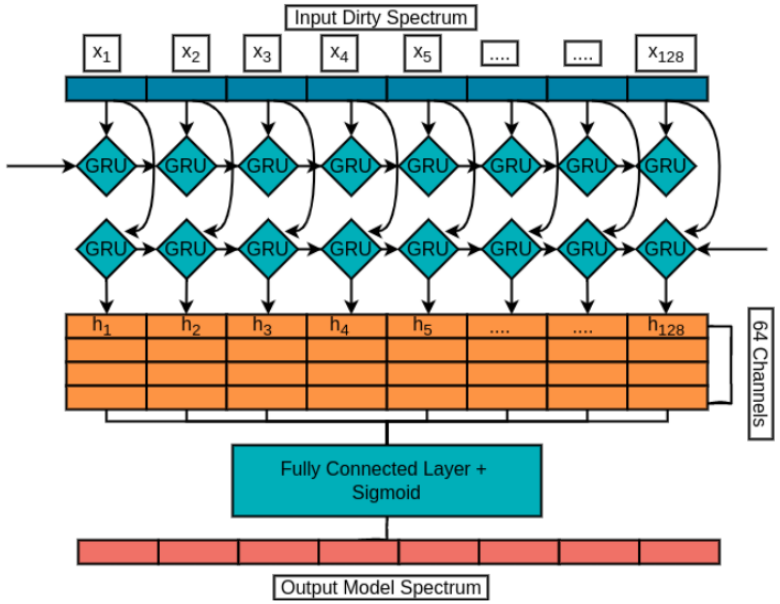


Figure 6. Deep GRU's architecture constituted by two layers of GRUs followed by a Fully connected layer and a Sigmoid activation function.

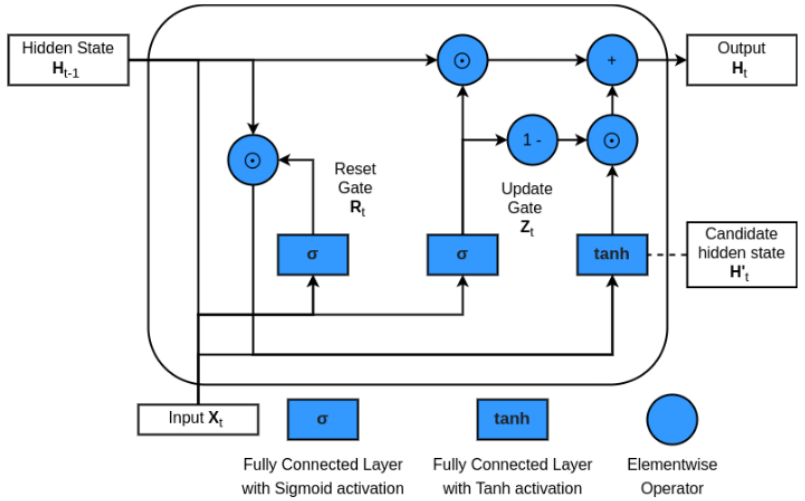


Figure 5. Gated Recurrent Unit architecture showing the flow of data within the network.

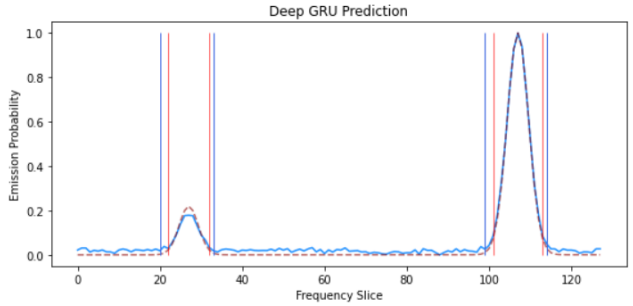


Figure 13. In blue the dirty spectrum extracted from the central source bounding box predicted by Blobs Finder (Fig. 11), in dotted-red the Deep GRU's prediction. Vertical blue bars delimit the true emission ranges, while red bars the predicted emission ranges. Vertical blue and red bars delimit the true and predicted emission ranges, respectively. A secondary fainter source emission peak is detected by Deep GRU and thus the source is flagged for deblending.

Flagged for deblending

3-rd – DL module – Parameters estimation (Battery of RESNETs)

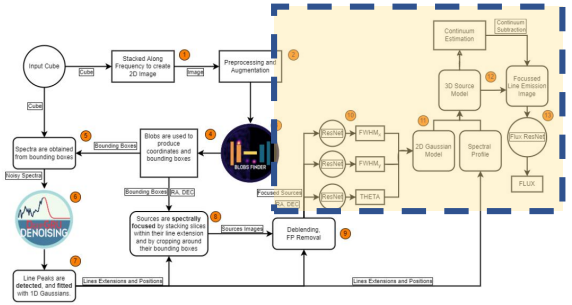


Figure 9. The full pipeline schema. Numbers show the logical flow of the data within the pipeline.

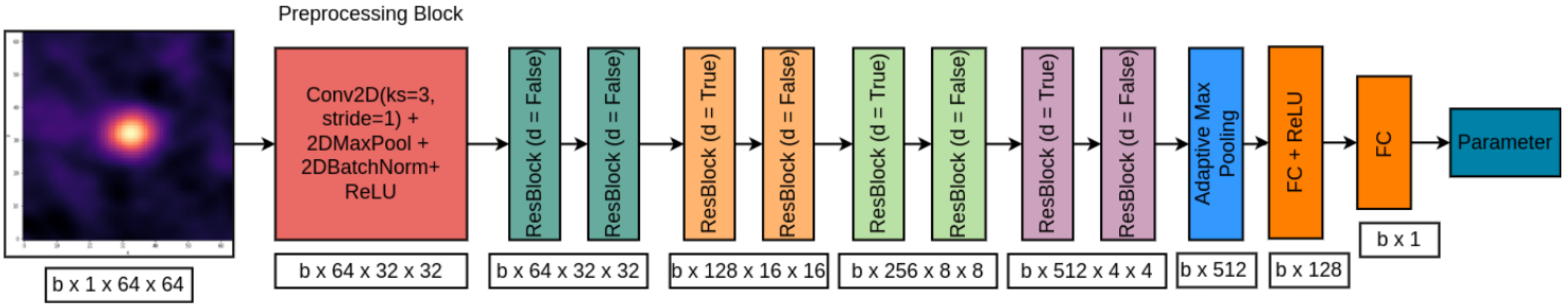
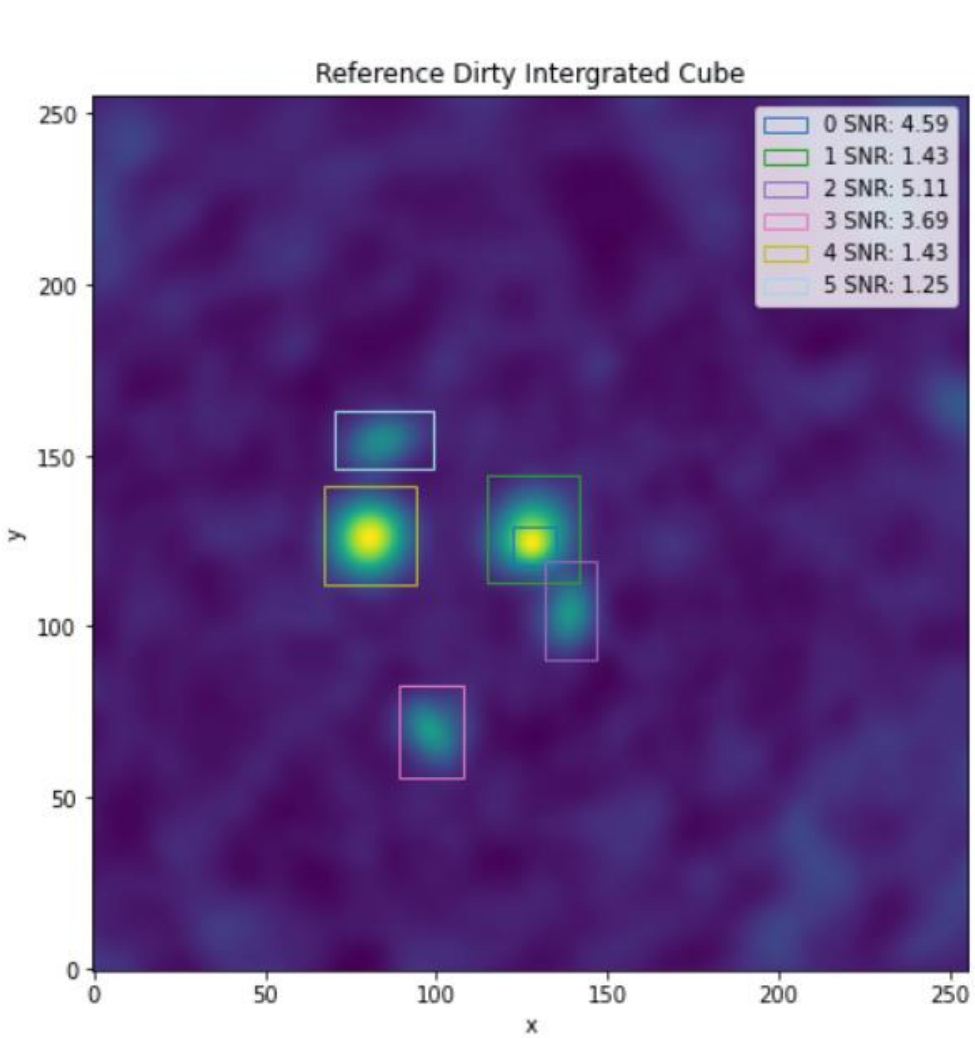
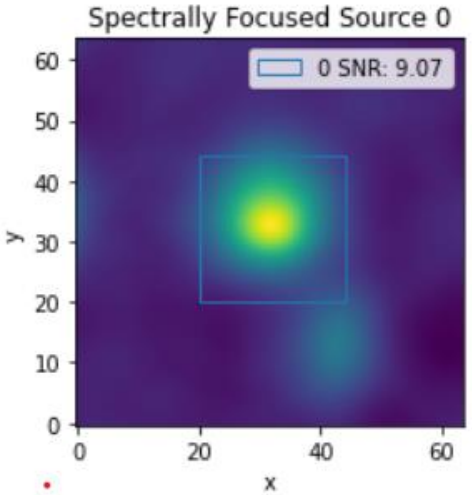
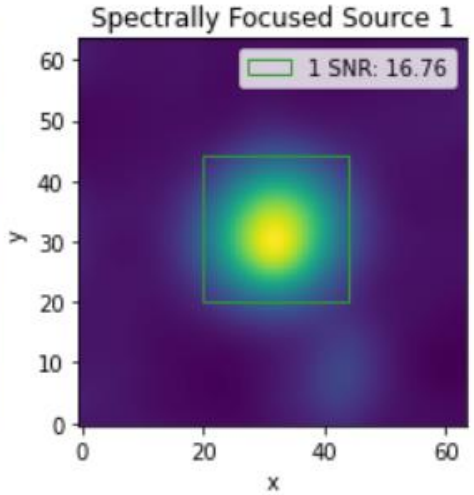
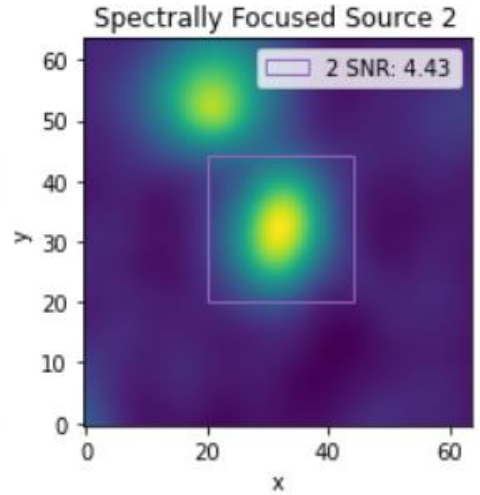


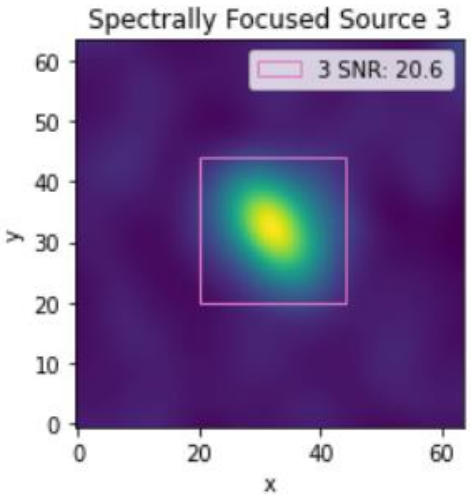
Figure 8. Our implementation of the ResNet architecture. The input image is first preprocessed by a 2D Convolution layer, followed by 2D Max Pooling, 2D Batch Normalization, and a ReLU activation function, and then is forwarded through four blocks of two Residual Blocks (see Fig. 7). The output is then processed via an Adaptive Max Pooling layer and fed to two fully connected layers which map the latent vector of 512 elements to a single scalar (the value of the parameter of interest for the ResNet).

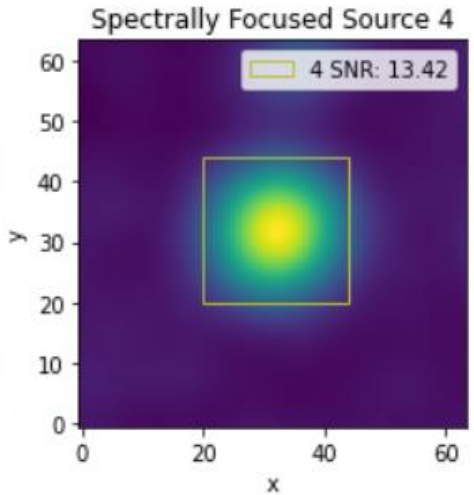

Reference Dirty Intergrated Cube
0 SNR: 4.59
1 SNR: 1.43
2 SNR: 5.11
3 SNR: 3.69
4 SNR: 1.43
5 SNR: 1.25
x
y

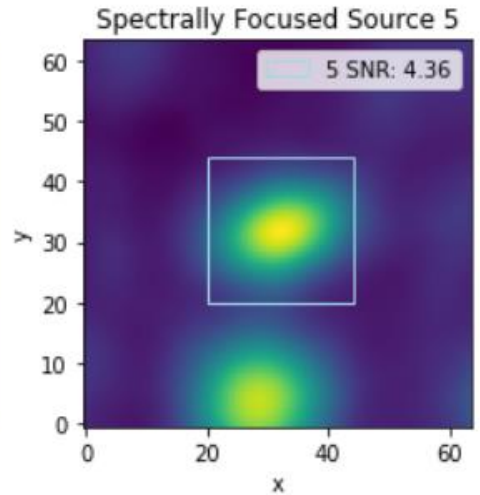

Spectrally Focused Source 0
0 SNR: 9.07
x
y


Spectrally Focused Source 1
1 SNR: 16.76
x
y


Spectrally Focused Source 2
2 SNR: 4.43
x
y


Spectrally Focused Source 3
3 SNR: 20.6
x
y


Spectrally Focused Source 4
4 SNR: 13.42
x
y


Spectrally Focused Source 5
5 SNR: 4.36
x
y

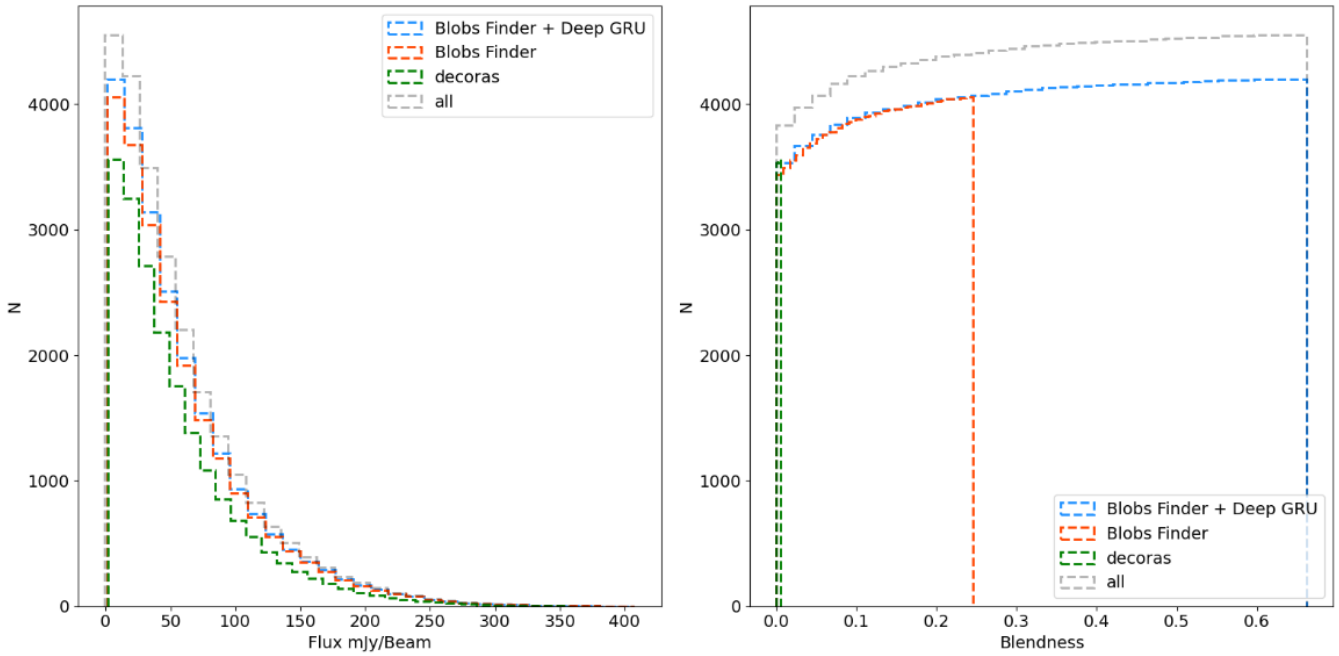


Figure 15. Left: histograms of the detected sources flux densities. Right: cumulative histogram of the detected sources *blendness score* (see the text). In both histograms, we compare our detection pipeline (Blobs Finder + Deep GRU), our implementation of Blobs Finder, DECORAS implementation of Blobs Finder and, we report the histograms for all the test set distribution.

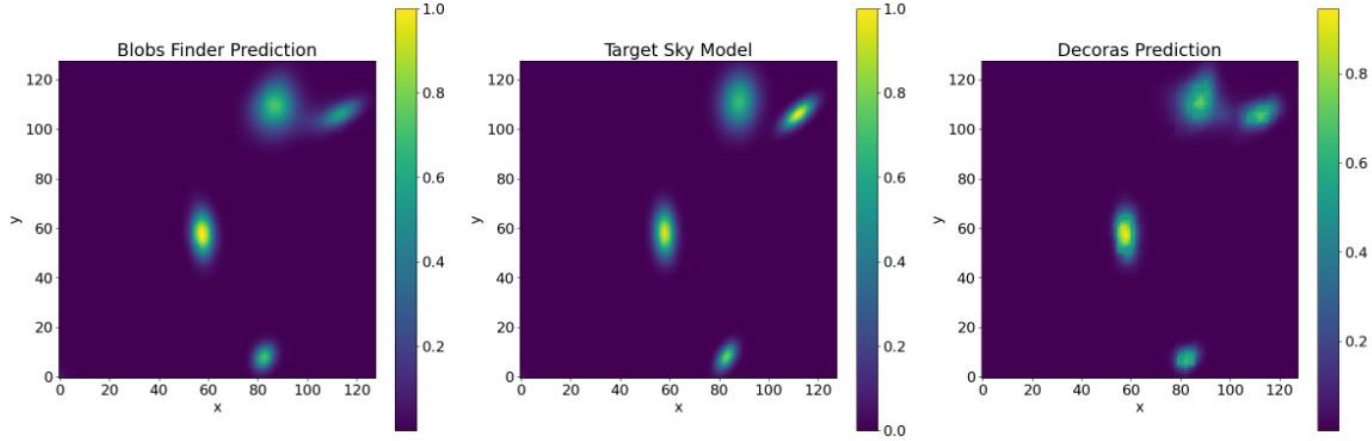


Figure 16. Left: Blobs Finder predicted 2D probabilistic map; Center: true sky model image; Right: DECORAS implementation of Blobs Finder predicted 2D probabilistic map. Predictions and target images have been cropped to 128 by 128 around sources, to better showcase the reconstruction quality.

Algorithm	TP /	FP	FN
Pipeline	4202 (92.3%)	0	354 (7.7%)
blobcat	2779 (61%)	2429	1777 (39%)
Sofia-2	1010 (22%)	4011	3546 (78%)
decoras	3560 (78.2%)	759	996 (21.9%)
Algorithm	**Precision**	**Recall**	**Mean IoU**
Pipeline	1.0	0.923	0.74
blobcat	0.53	0.609	0.61
Sofia-2	0.20	0.22	0.63
decoras	0.82	0.78	0.60

Table 4. Comparison between the sequential proposed source finding pipeline composed by Blobs Finder, DeepGRU and Spectral Focusing, *blobscat*, SOFIA-2 and DECORAS. Columns show true positives (TP), false positives (FP), false negatives (FN), precision, recall and mean intersection over union (Mean IoU) between true bounding boxes and predicted ones. TP and FN are also expressed as fractions over the total number of sources.

Parameter Residual	mean	std
x (pixels)	−0.004	0.73
y (pixels)	−0.005	0.67
FWHMx (pixels)	−0.04	0.46
FWHMy(pixels)	−0.12	0.45
z(slices)	0.0	0.003
Δ_z (slices)	0.0	0.001
pa (degrees)	−0.65	20.28
flux (mJy/beam)	−9.56	20.08

Table 5. This table shows the mean and standard deviation of all the residual distributions between the true target parameters and the predictions made by our pipeline. The x and y positions are computed from Blobs Finder predicted blobs, z positions and extensions Δ_z are computed from Deep GRU predictions, and the remaining parameters are predicted from the four ResNets. Alongside each parameter, we also indicate their unit of measurement.

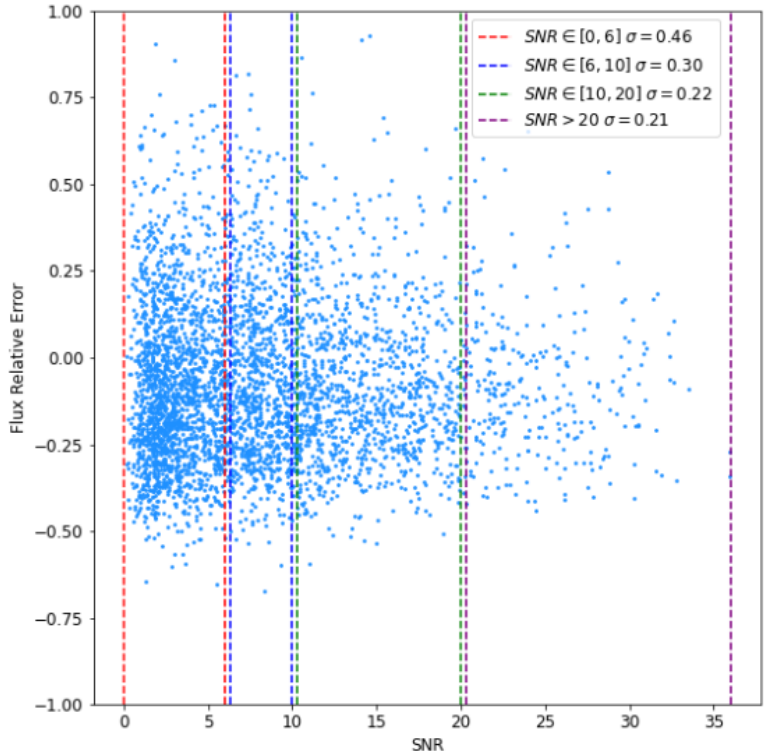
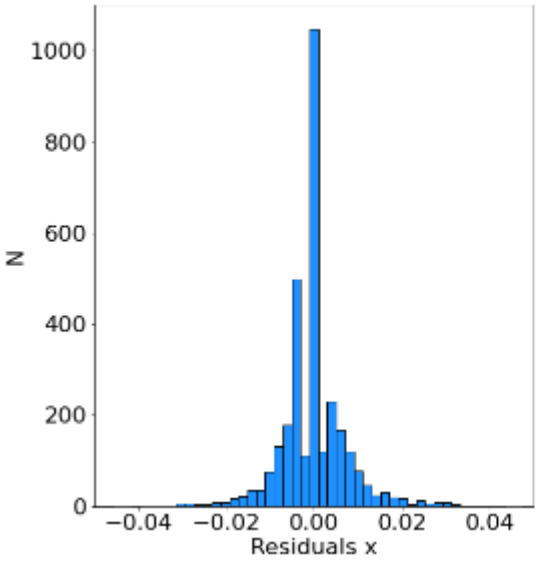
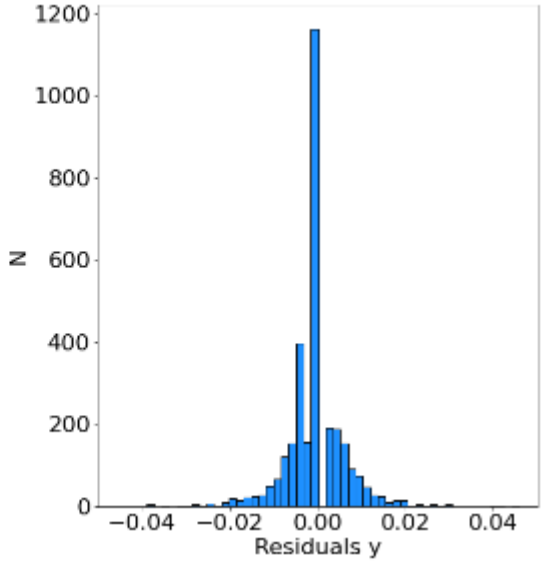
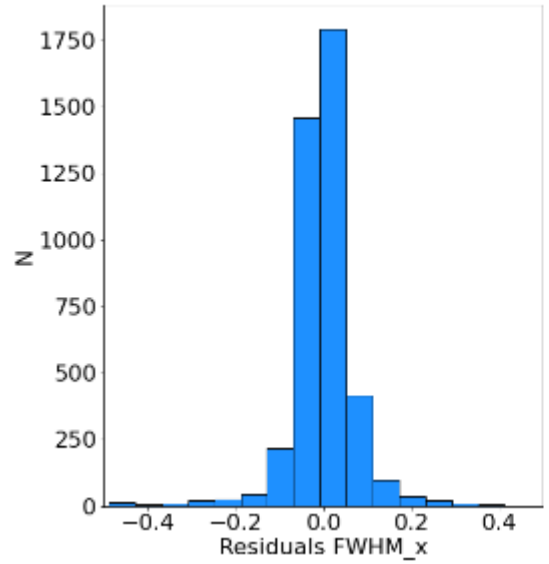
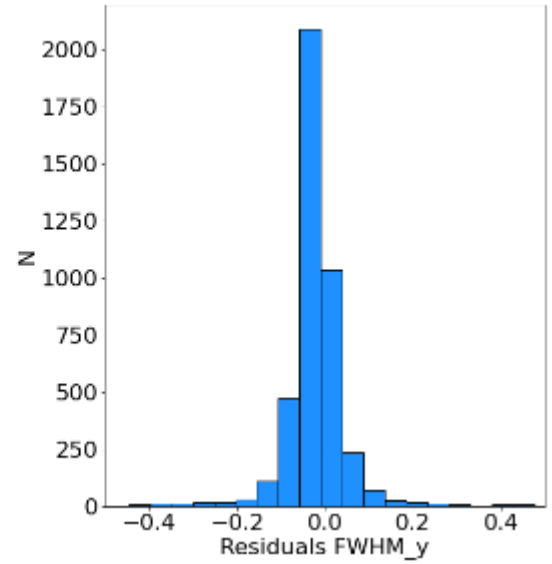


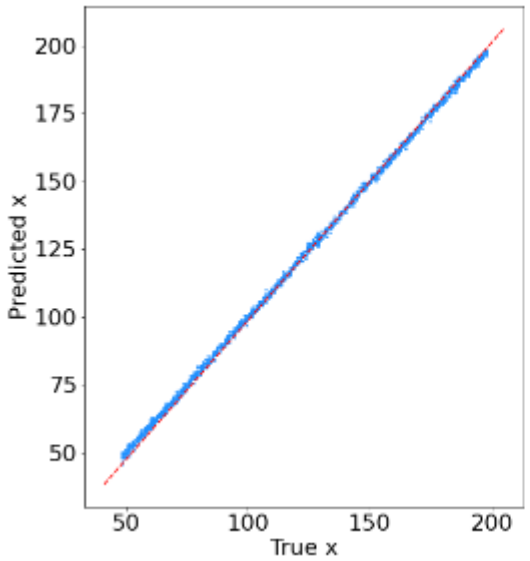
Figure 20. Scatter plot of the sources SNR against their flux densities relative errors. Vertical bars divide the plot in section of SNR. The legend shows, for each SNR interval, the standard deviation of the relative errors.

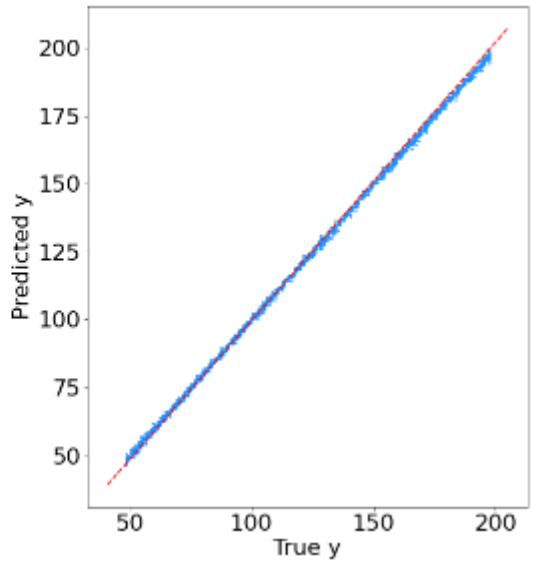

N
Residuals x

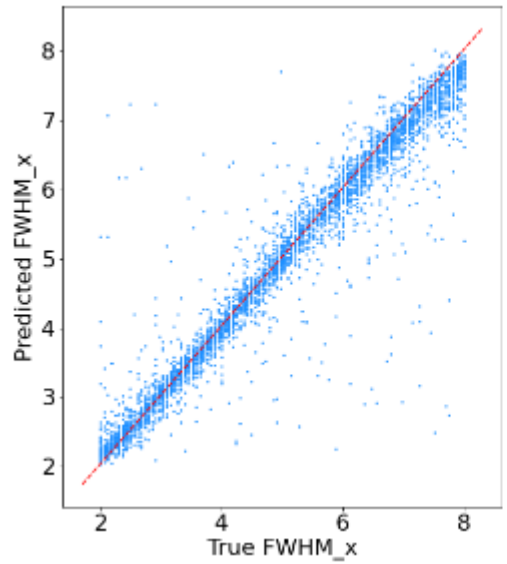

N
Residuals y

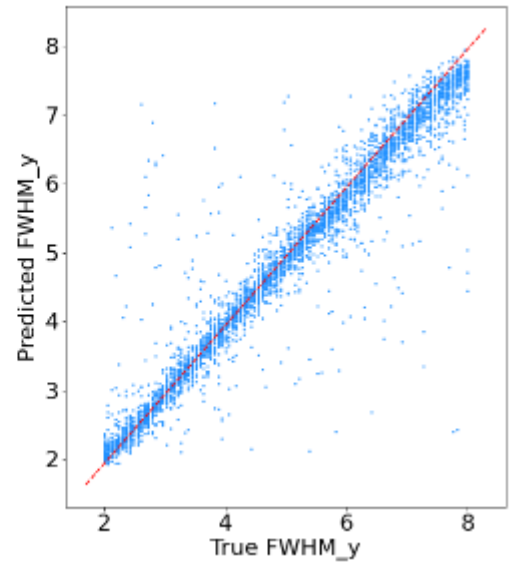

N
Residuals FWHM_x


N
Residuals FWHM_y


Predicted x
True x


Predicted y
True y


Predicted FWHM_x
True FWHM_x


Predicted FWHM_y
True FWHM_y

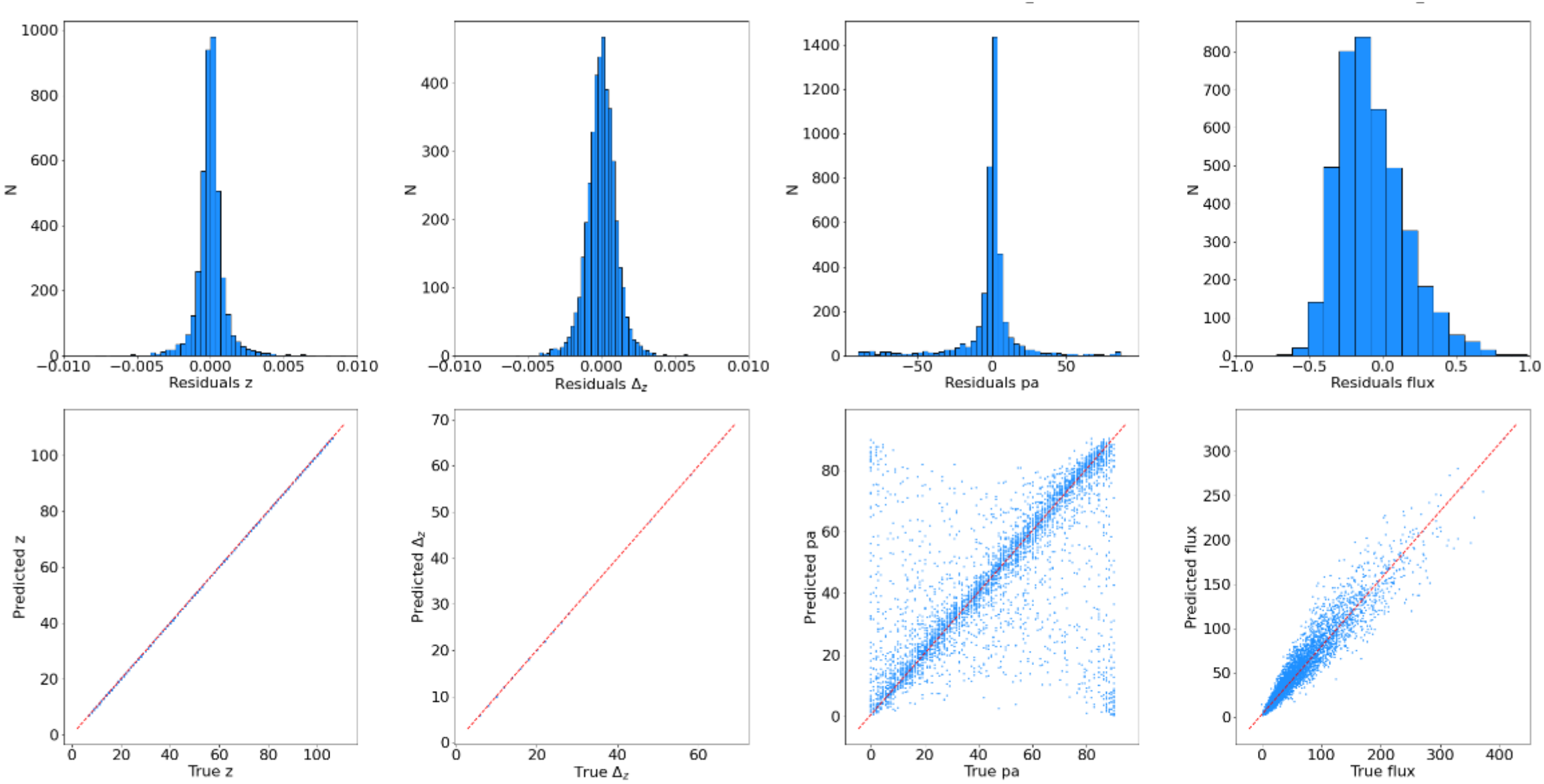


Figure 19. Scatter plots of the true parameters values against the models predictions and the corresponding residuals histograms. The red dotted lines in each scatter plot represent the bisector of the quadrant, i.e. if all instances were perfectly predicted, they would all lie on the red line.

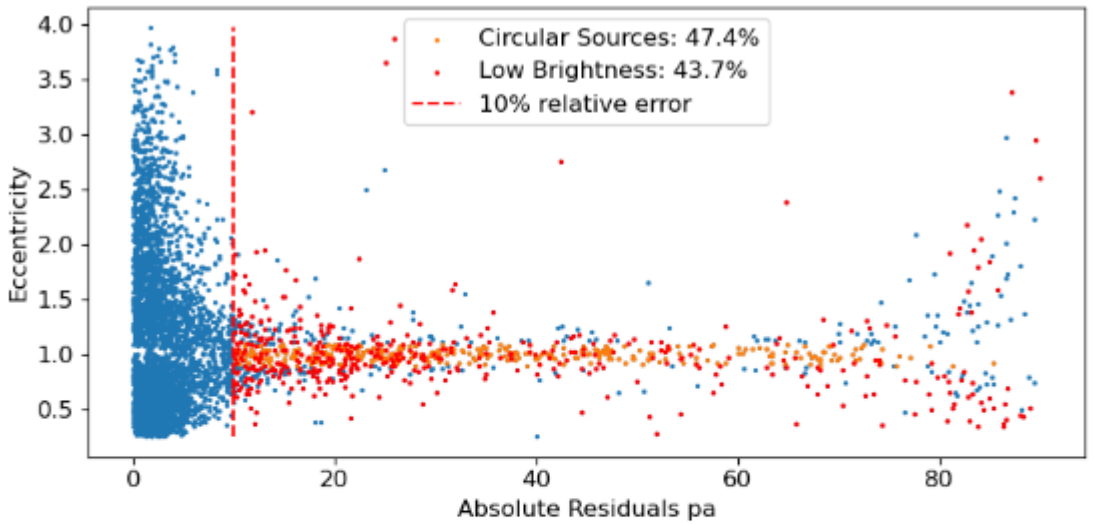


Figure 21. Scatter plot of the sources absolute projection angle residual errors against their eccentricity, defined as the ratio of their FWHMs. The vertical bar delimits the 10% mark for the residual error, while the sources highlighted in orange are circular ($e \simeq 1$) and the ones in red have surface brightness lower then 30 mJy / beam. These sources account for respectively 47.4 and 43.7 of all sources with a relative error higher then 10%.

Conclusions:

- In ML there is a concrete possibility to **expand our knowledge** not only quantitatively but also qualitatively (… achieving a higher level of complexity).
- **Feature selection and feature interpretation** are often ignored but they likely are the most interesting part of the process
- Better use of **simulations** may hugely help in obtaining outstanding results
- **It is NOT a one man job** (every year more than 3000 papers on ML and it is increasing)…. Transition took place around 2010.
- **TODAY**: There is **NO WAY a DOMAIN EXPERT can become also an Expert in ML** . At most you can aim at beccoming a practinioner.
- **CHOOSE ON WHICH SIDE OF THE PROBLEM YOU WANT TO BE** and enjoy it

conclusion

Professor
(domain expert … old, knows little about ML) has **interesting** but rather **standard** problem…

Hires a **PhD**
(domain expert at large…, knows little about ML but he/she is young, enthusiastic and learns fast)

Using off the shelf libraries (Github, Kaggle, paperswithcode, ecc.) in three months puts together a code… almost always DL)

But… very often suffers from lack of formal understanding of statistical learning and ML ….

Code performs equally and sometimed better than humans

Personal Notice
It is Almost impossible for a DL method to perform worse than a human

Non optimised, quite always in overfitting (too high capacity hence far too complex for the problem…)

Professor
is astonished and happy…

PhD
publishes a paper

arXiv