The University of Manchester
Jodrell Bank Observatory

MANCHESTER 1824

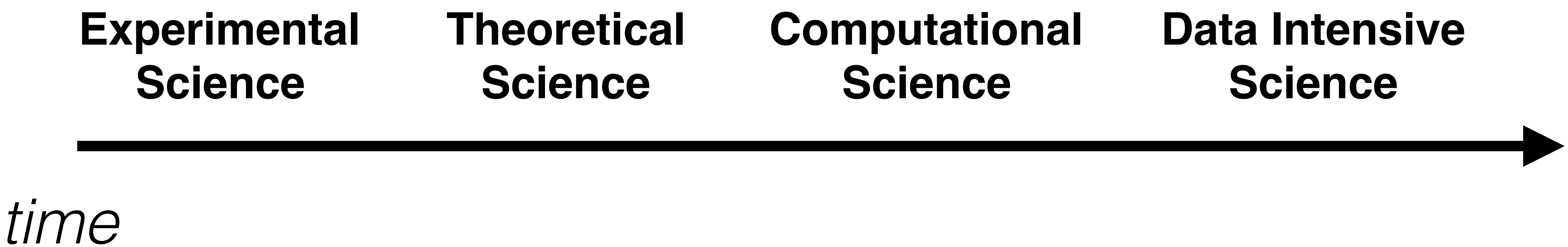
ANNA SCAIFE
JODRELL BANK CENTRE FOR ASTROPHYSICS
@RADASTRAT

DATA-CENTRIC MACHINE LEARNING FOR SKA POST-PROCESSING

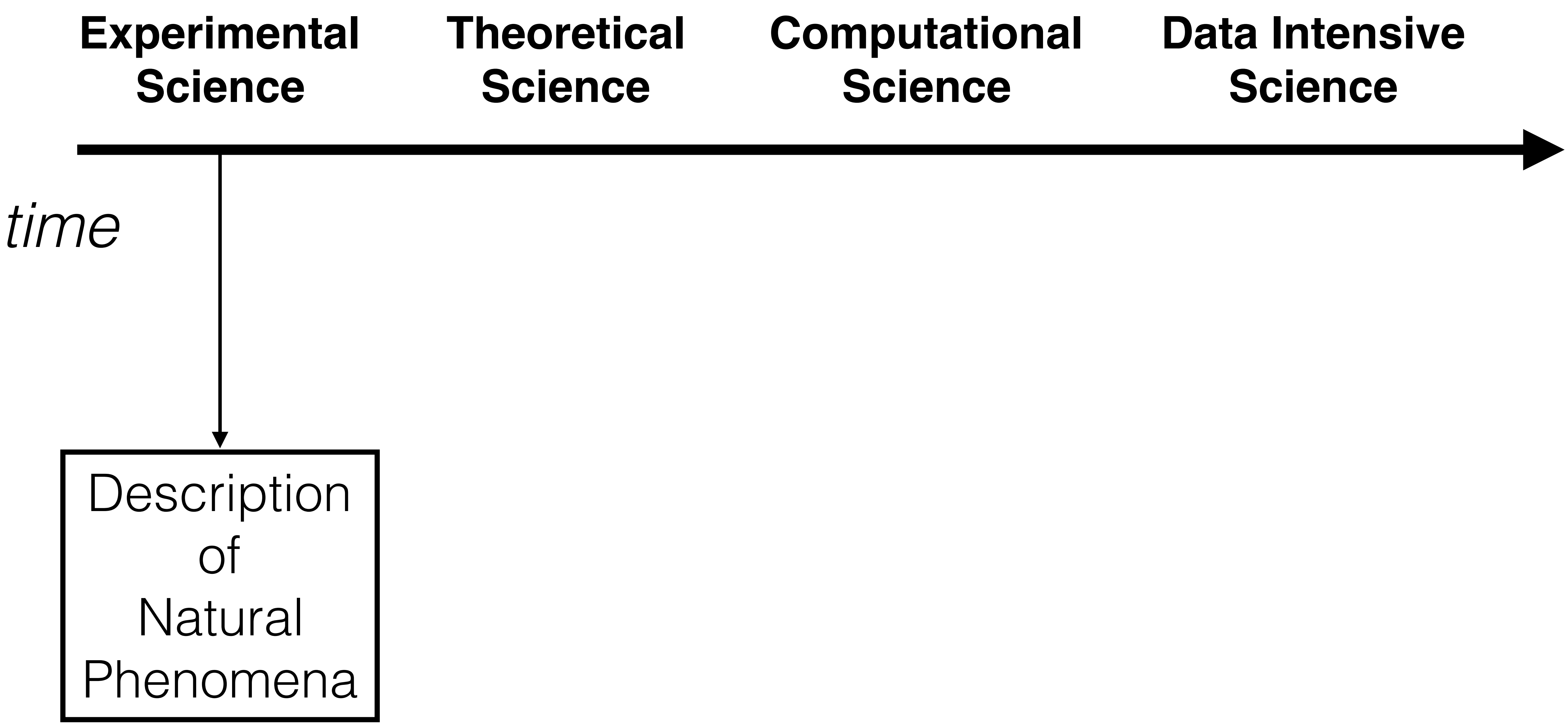
The University of Manchester
Jodrell Bank
Observatory

ANNA SCAIFE
JODRELL BANK CENTRE FOR ASTROPHYSICS
@RADASTRAT

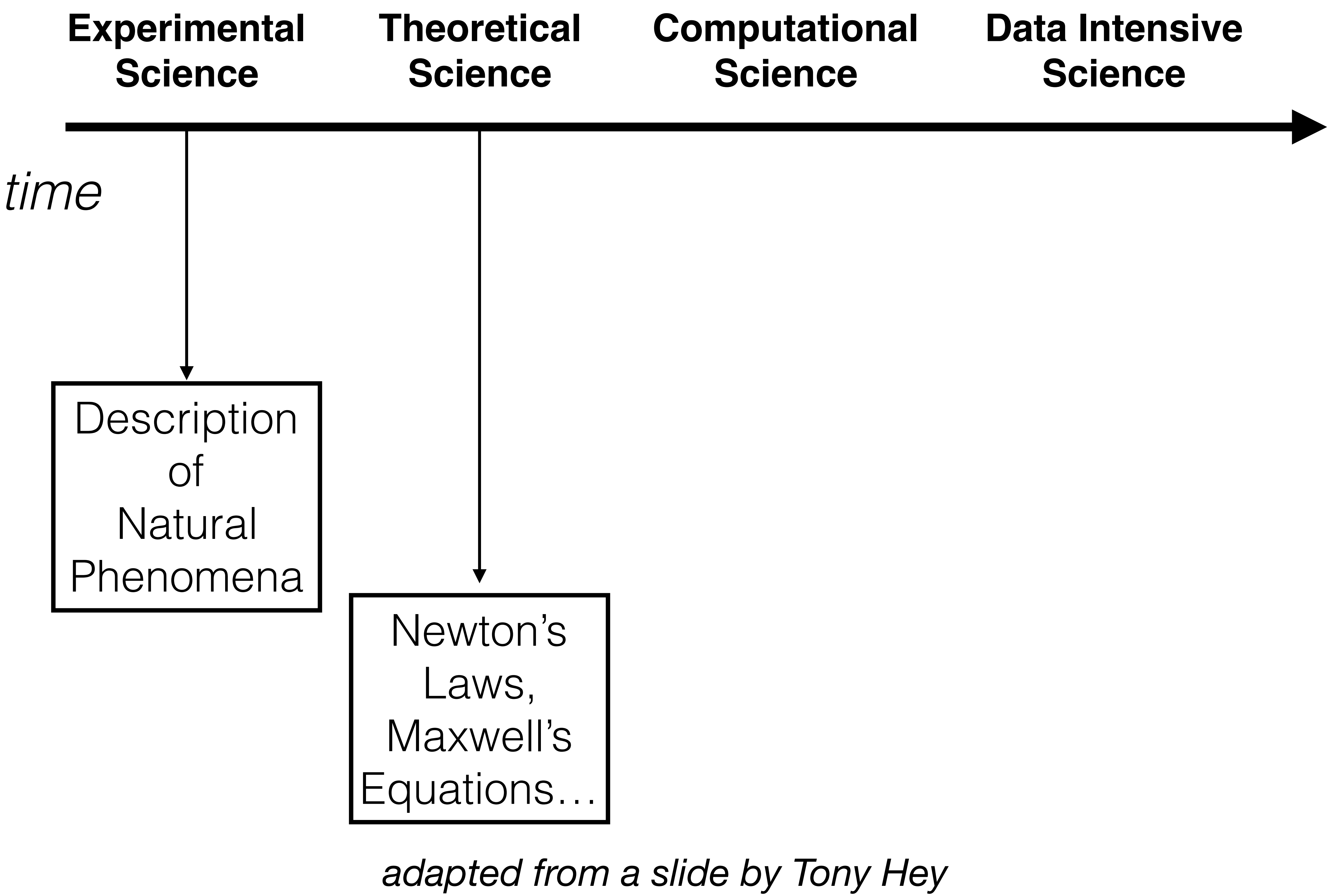
DATA-CENTRIC MACHINE LEARNING FOR SKA POST-PROCESSING

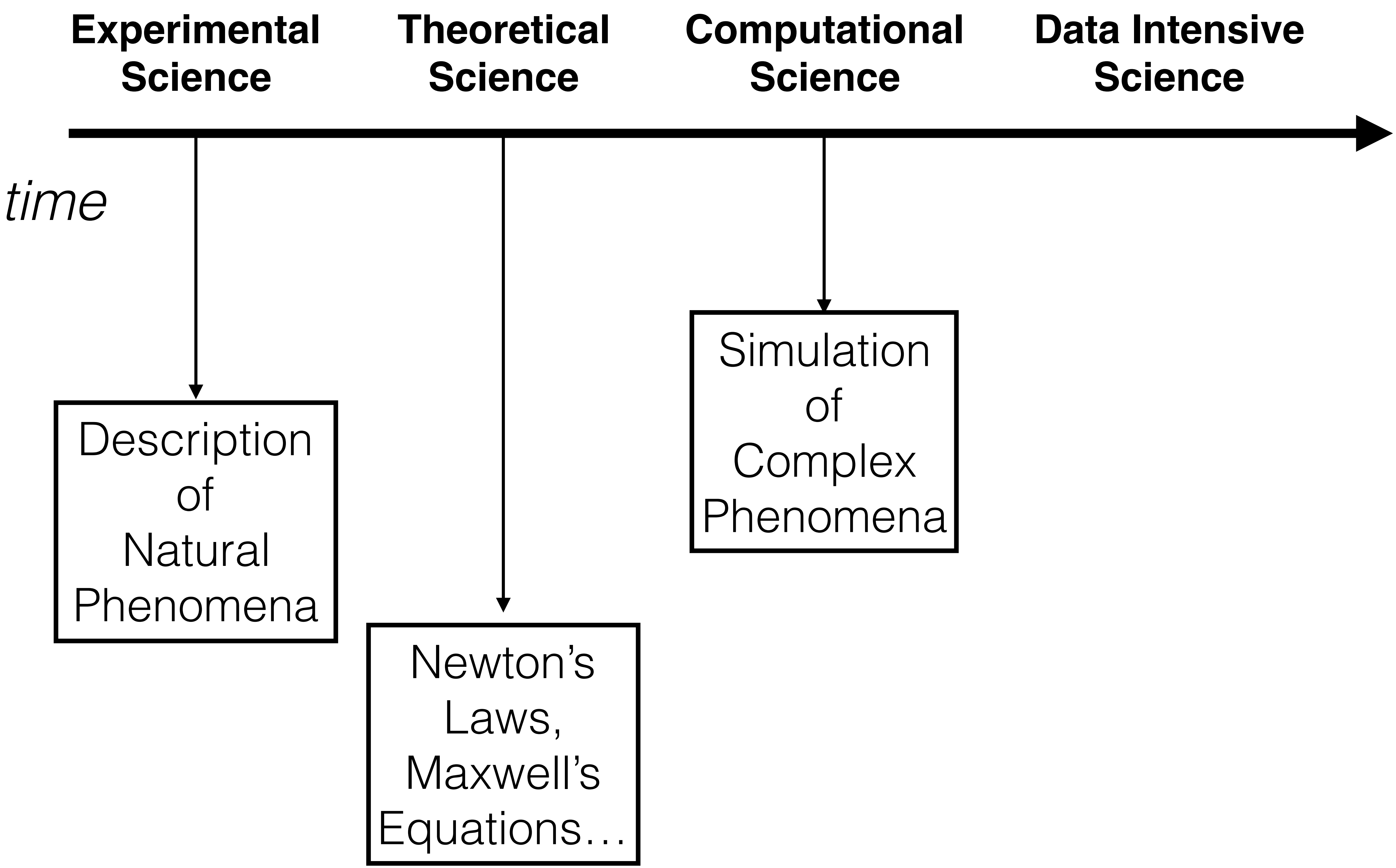


adapted from a slide by Tony Hey

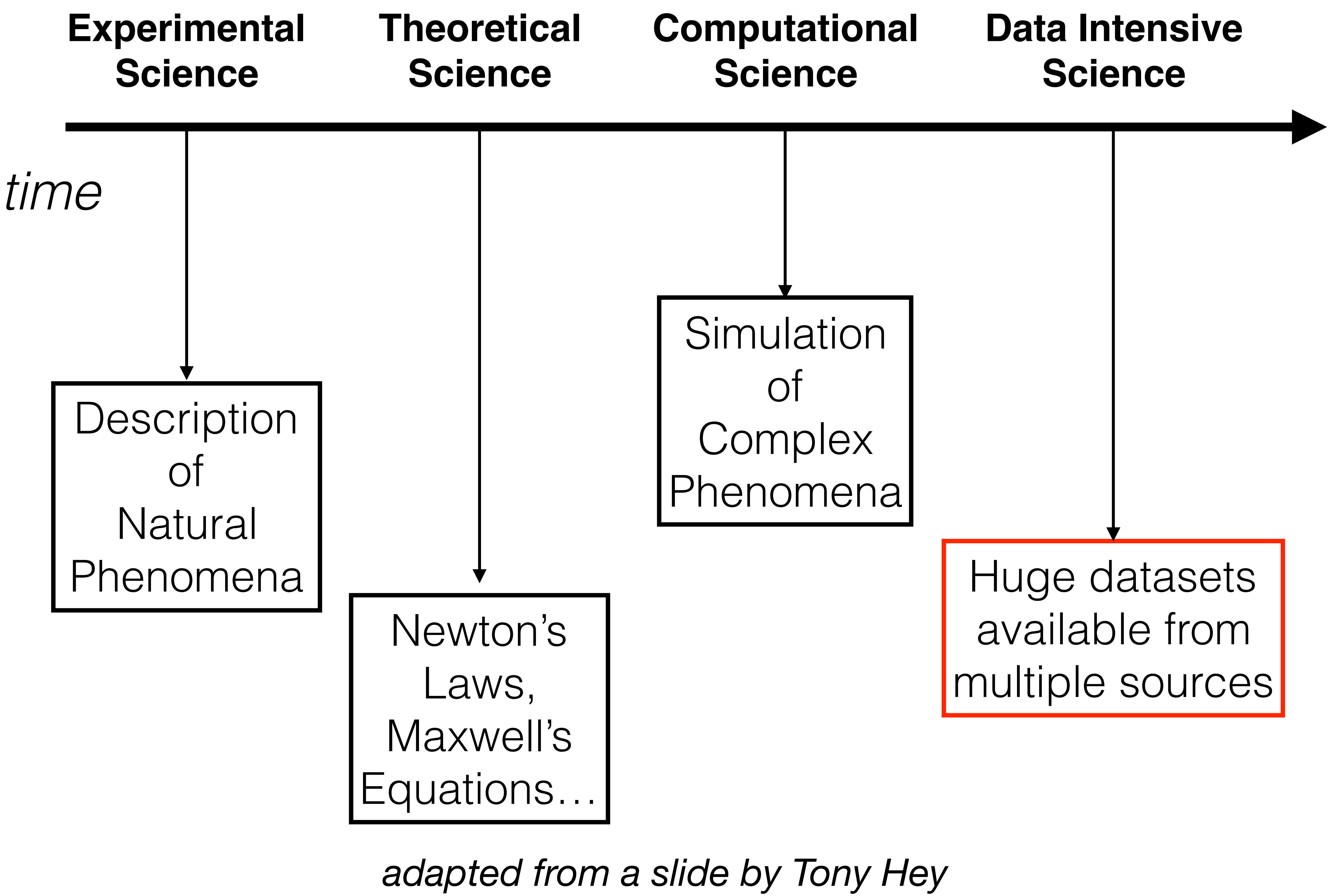


adapted from a slide by Tony Hey


Experimental Science
Theoretical Science
Computational Science
Data Intensive Science
time
Description of Natural Phenomena
Newton's Laws, Maxwell's Equations…
adapted from a slide by Tony Hey



adapted from a slide by Tony Hey


Experimental Science
Theoretical Science
Computational Science
Data Intensive Science
time
Description of Natural Phenomena
Newton's Laws, Maxwell's Equations…
Simulation of Complex Phenomena
Huge datasets available from multiple sources
adapted from a slide by Tony Hey

Too much data is as bad as none at all…

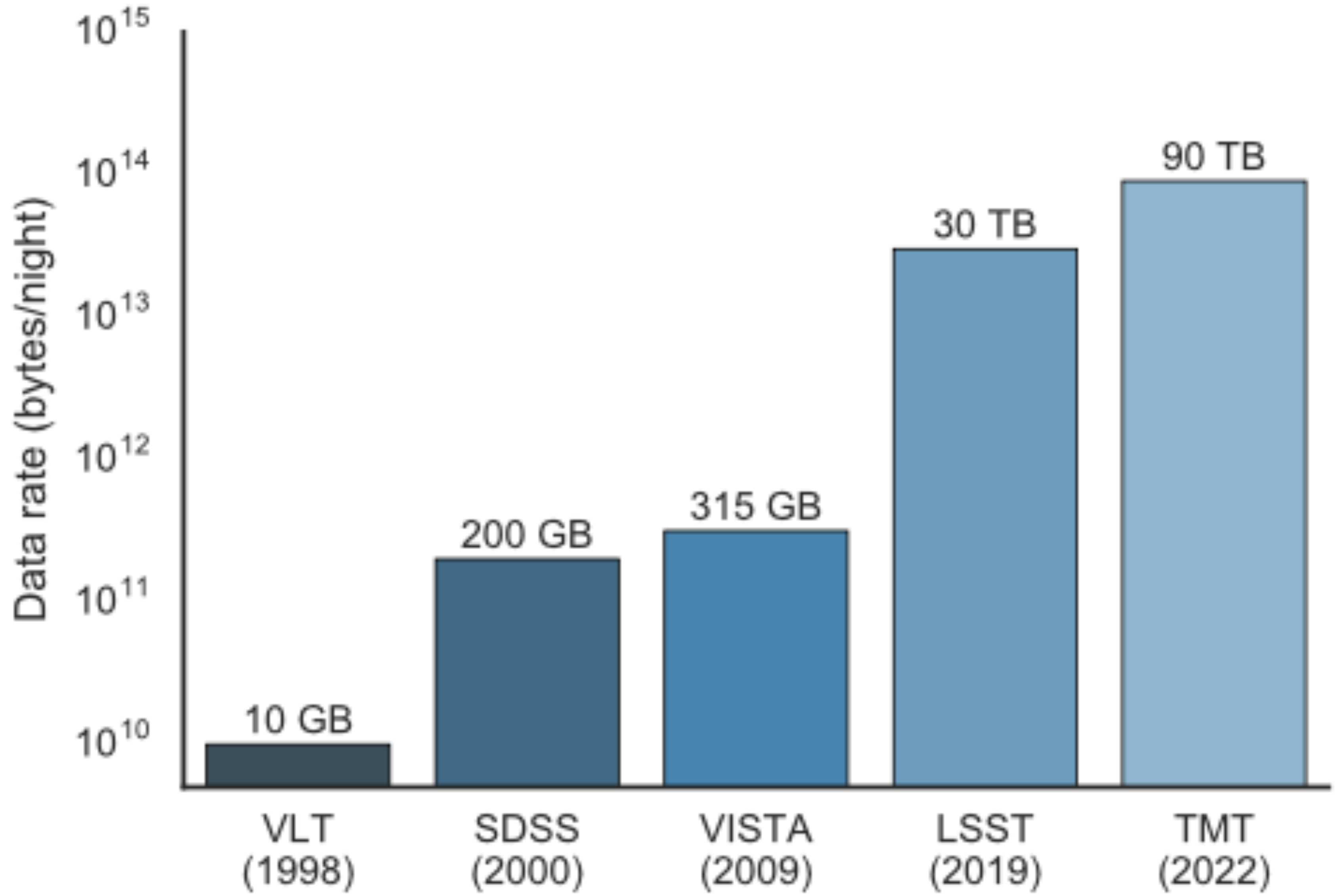
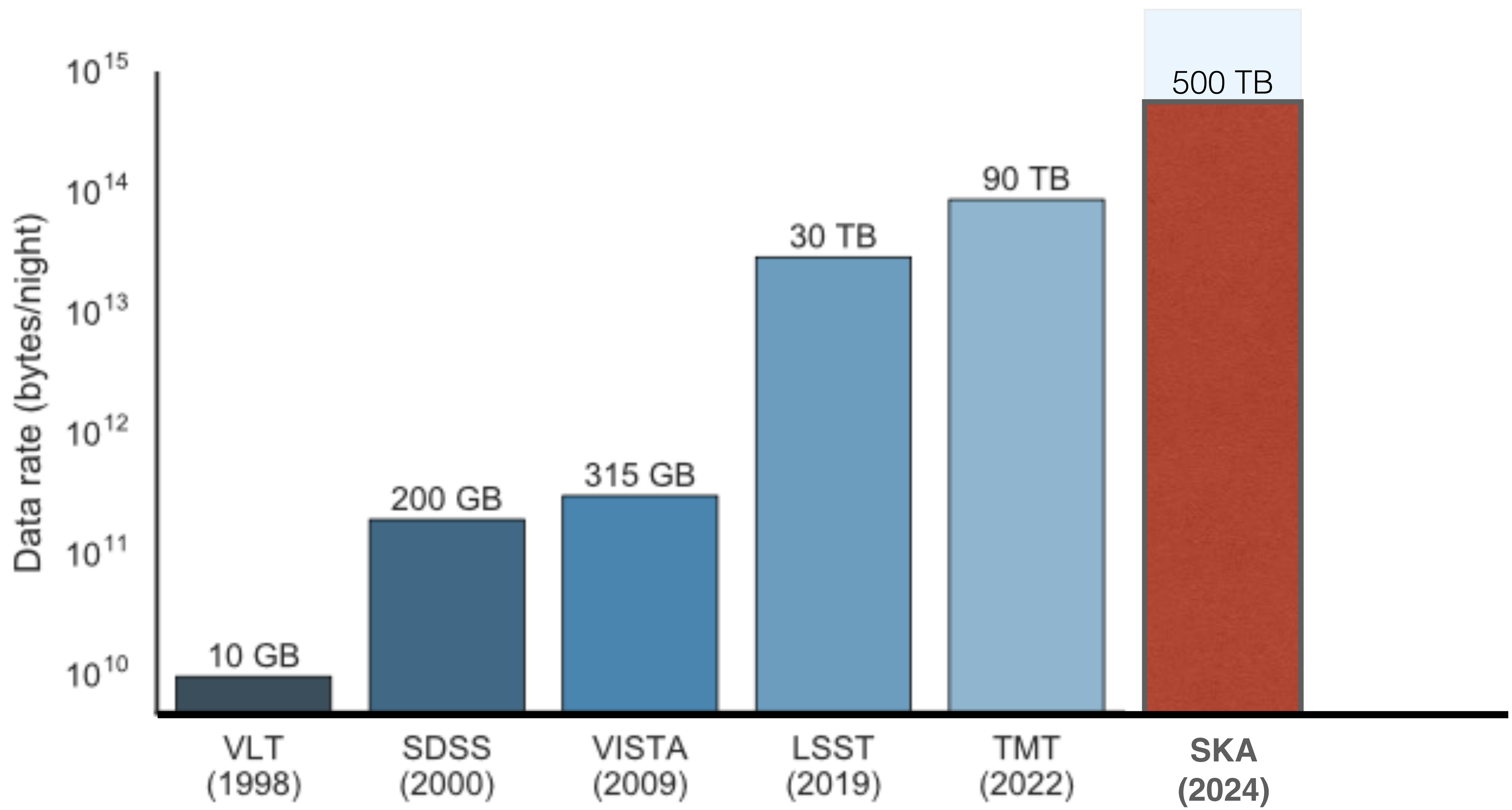


Figure 1: Increasing data volumes of existing and upcoming telescopes: Very Large Telescope (VLT), Sloan Digital Sky Survey (SDSS), Visible and Infrared Telescope for Astronomy (VISTA), Large Synoptic Survey Telescope (LSST) and Thirty Meter Telescope (TMT).

Kremer et al. 2017 arXiv:1704.04650

…and then there's the SKA.



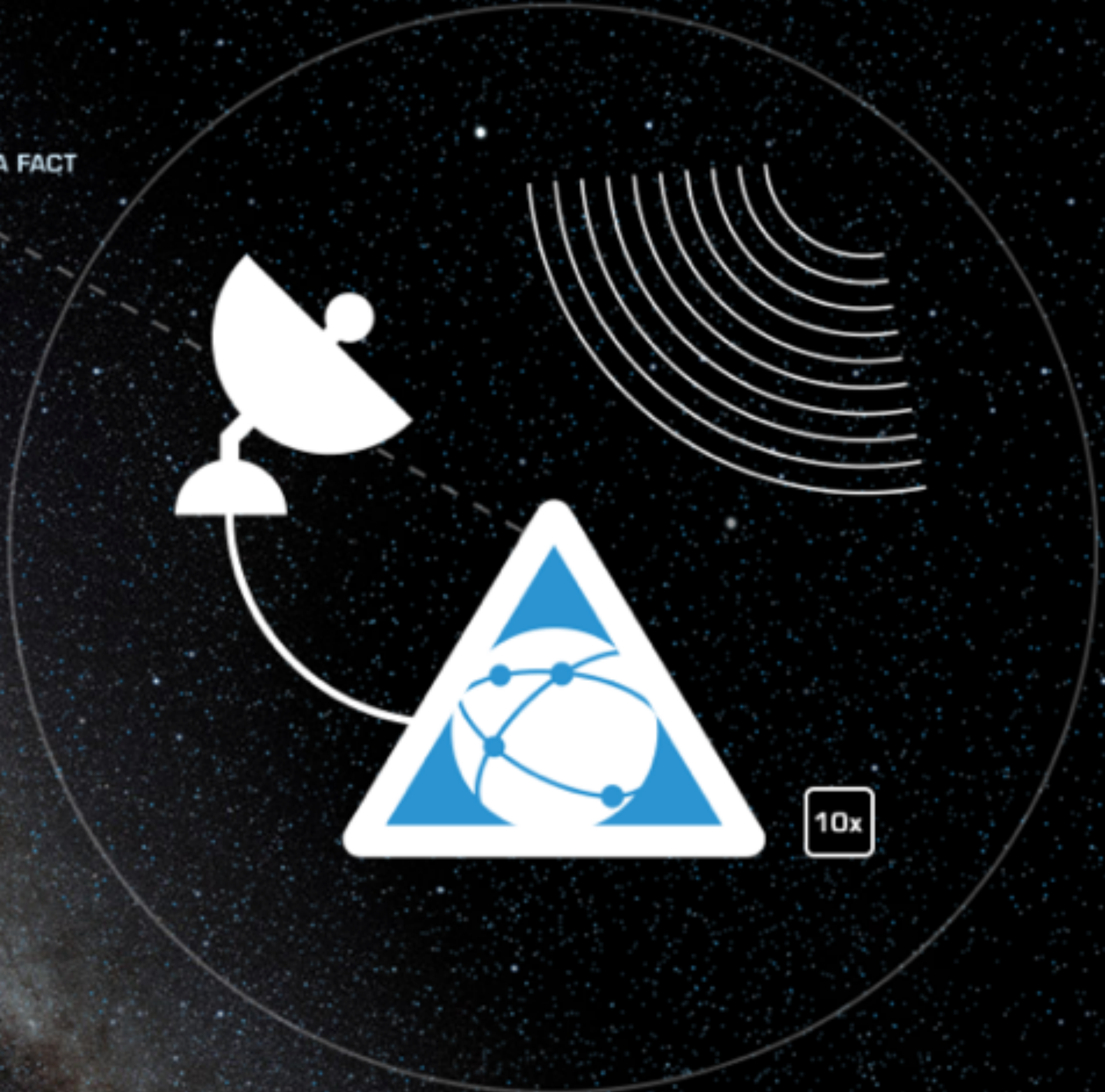
adapted from Kremer et al. 2017 arXiv:1704.04650

AMAZING 1 SKA FACT

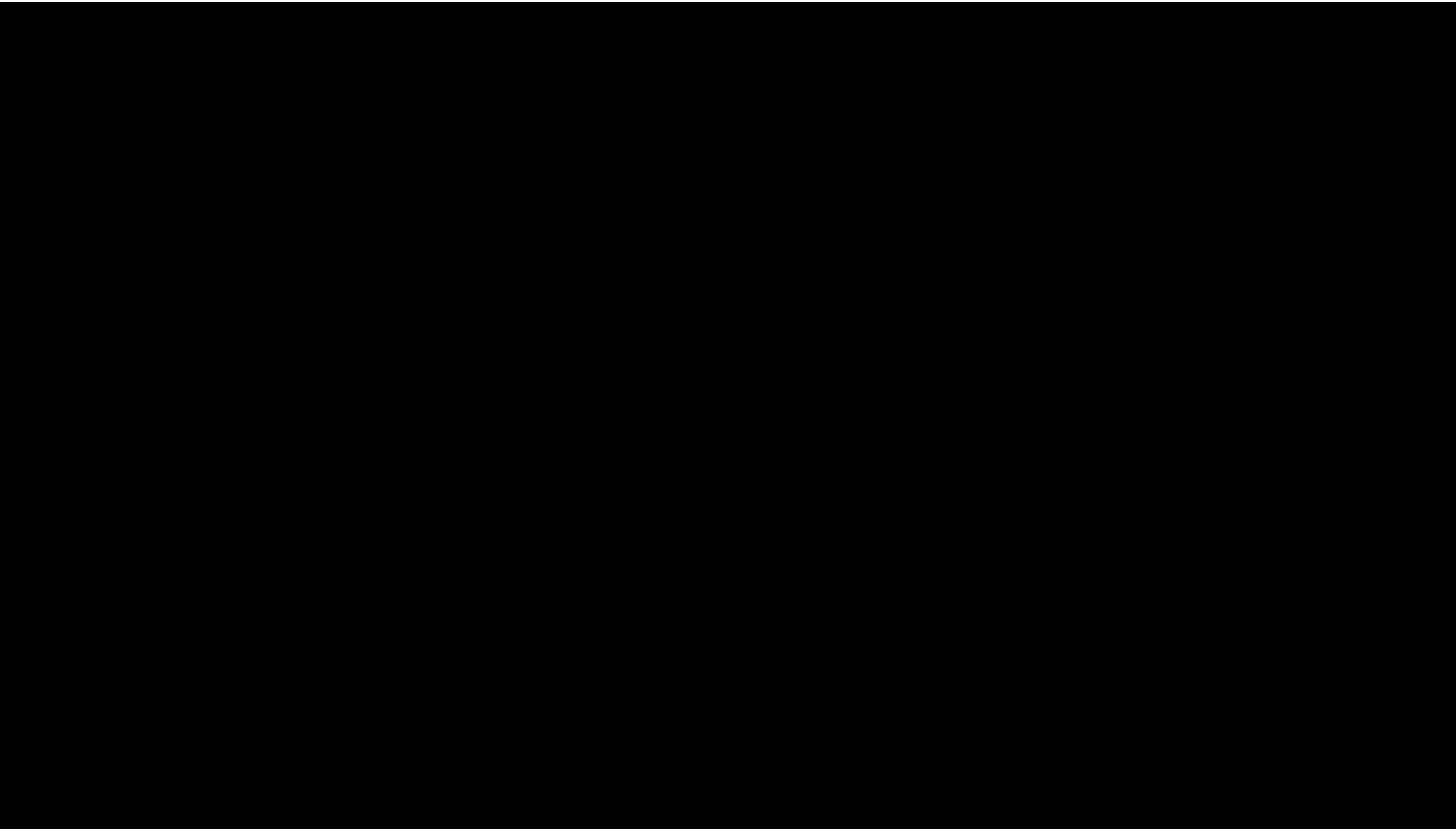
The data collected by the SKA in a single day would take nearly *two million years* to playback on an ipod.

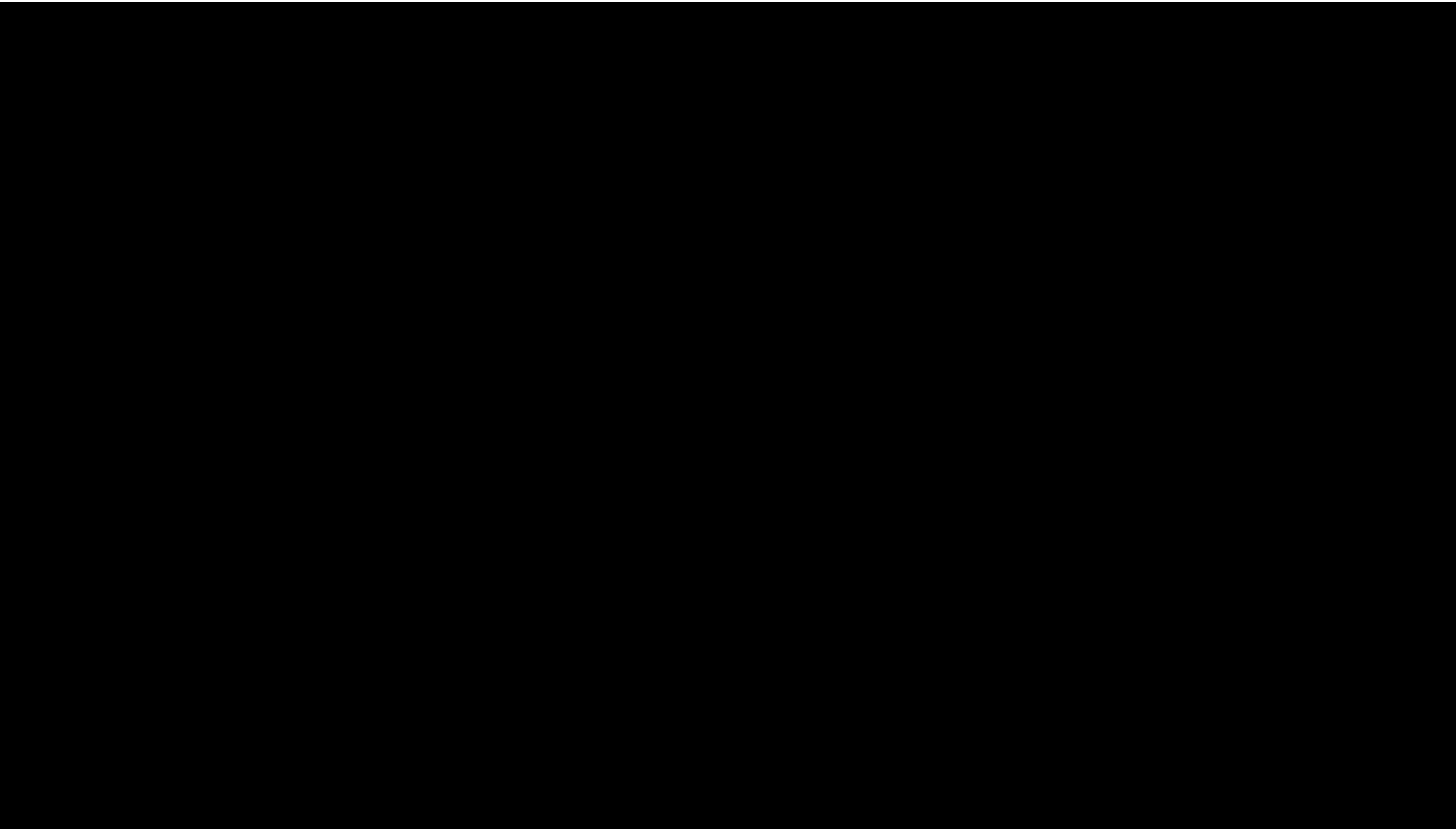
AMAZING **4** SKA FACT

The dishes of the SKA will produce *10 times* the global internet traffic.

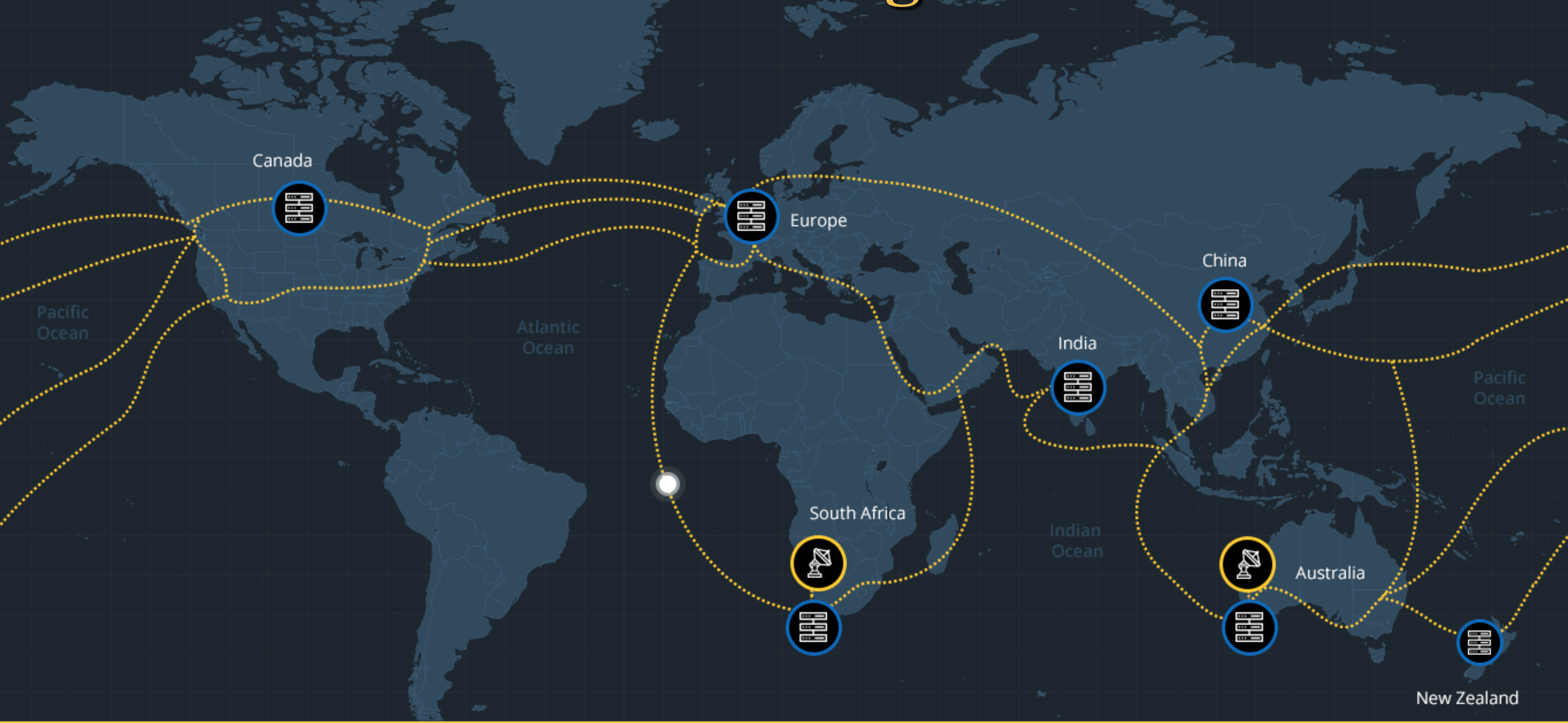


Let’s face it,
we’re not going to be looking
at it all by eye.



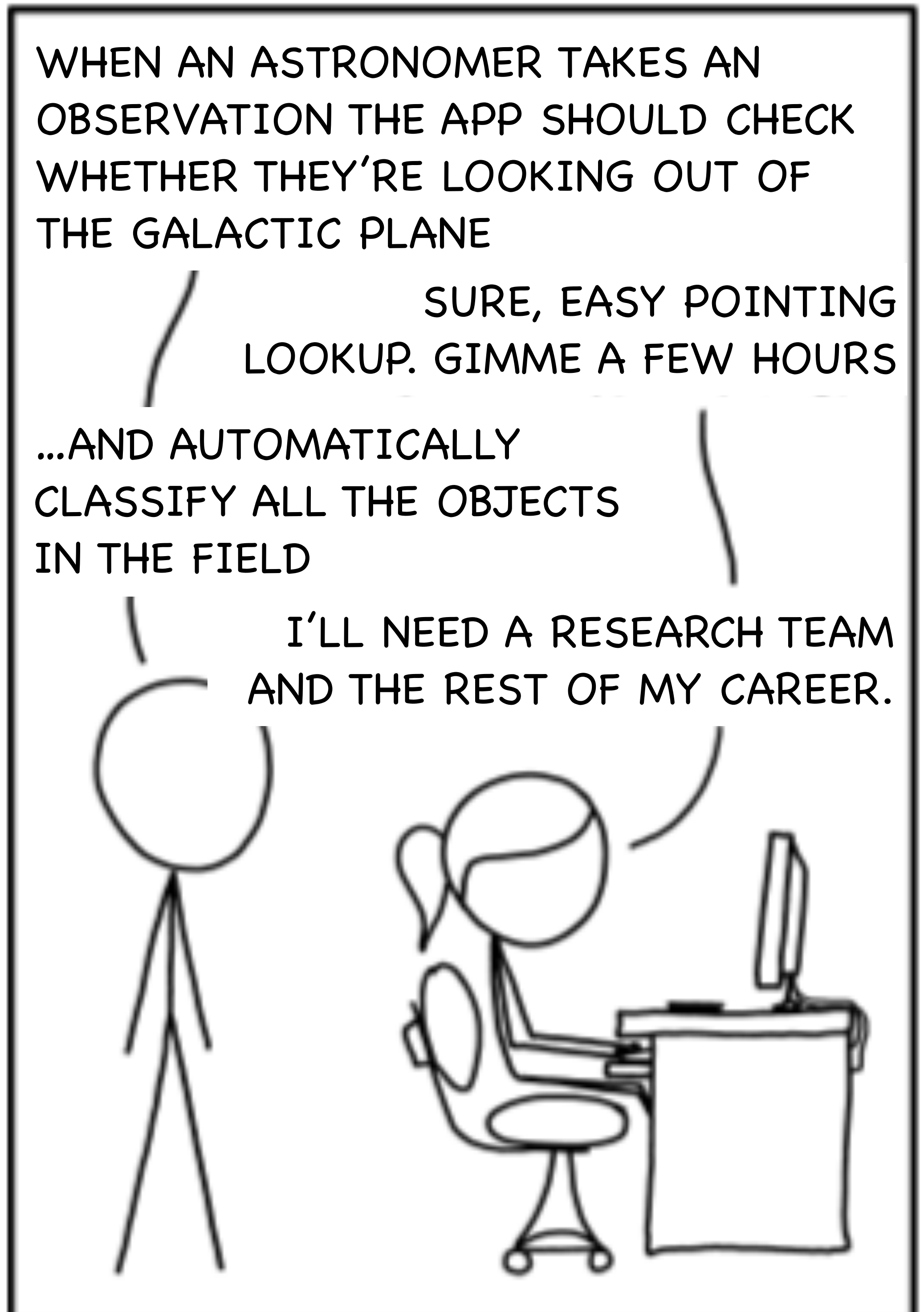


Global Network of SKA Regional Centres



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

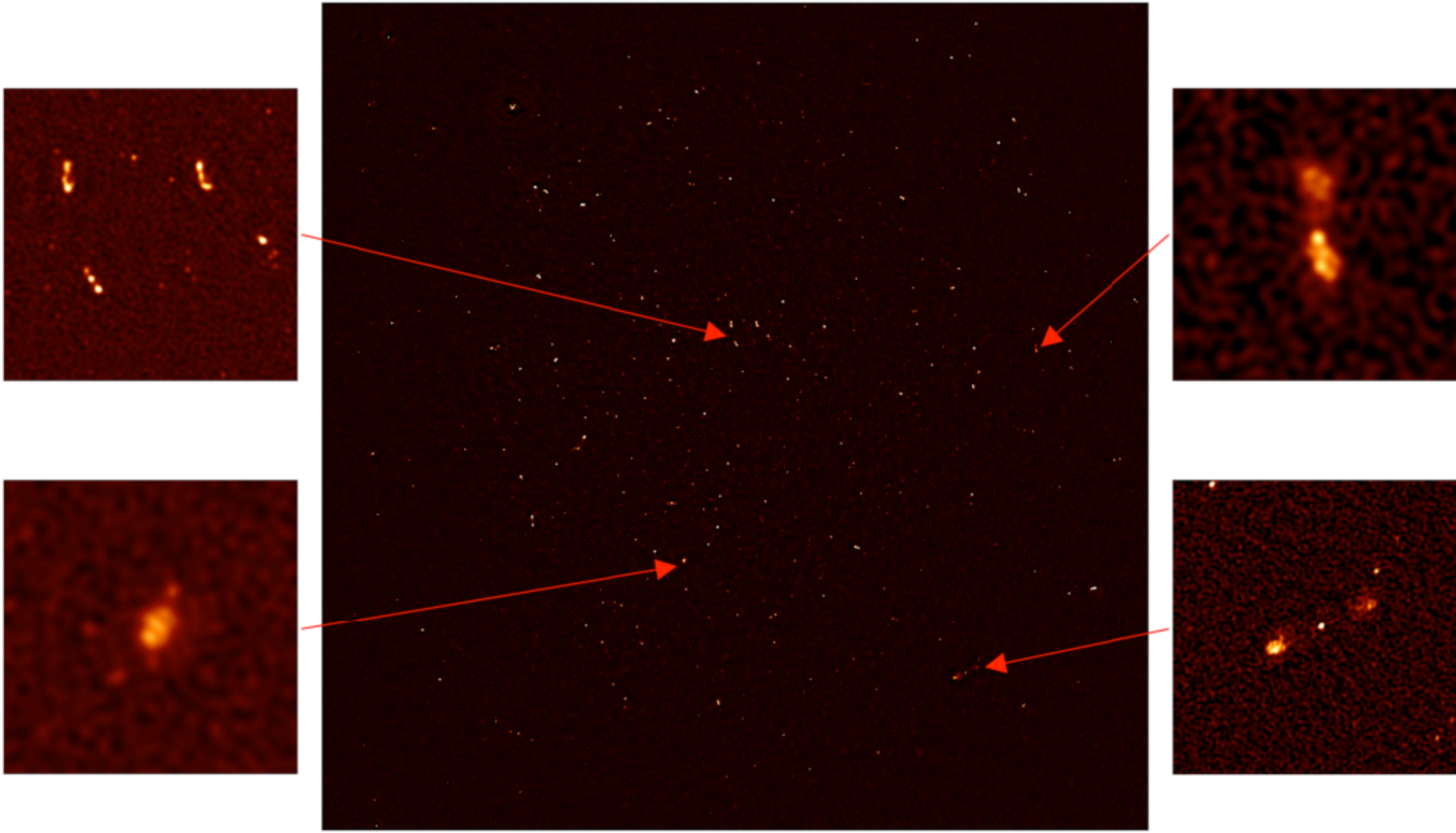
http://xkcd.com/1425/



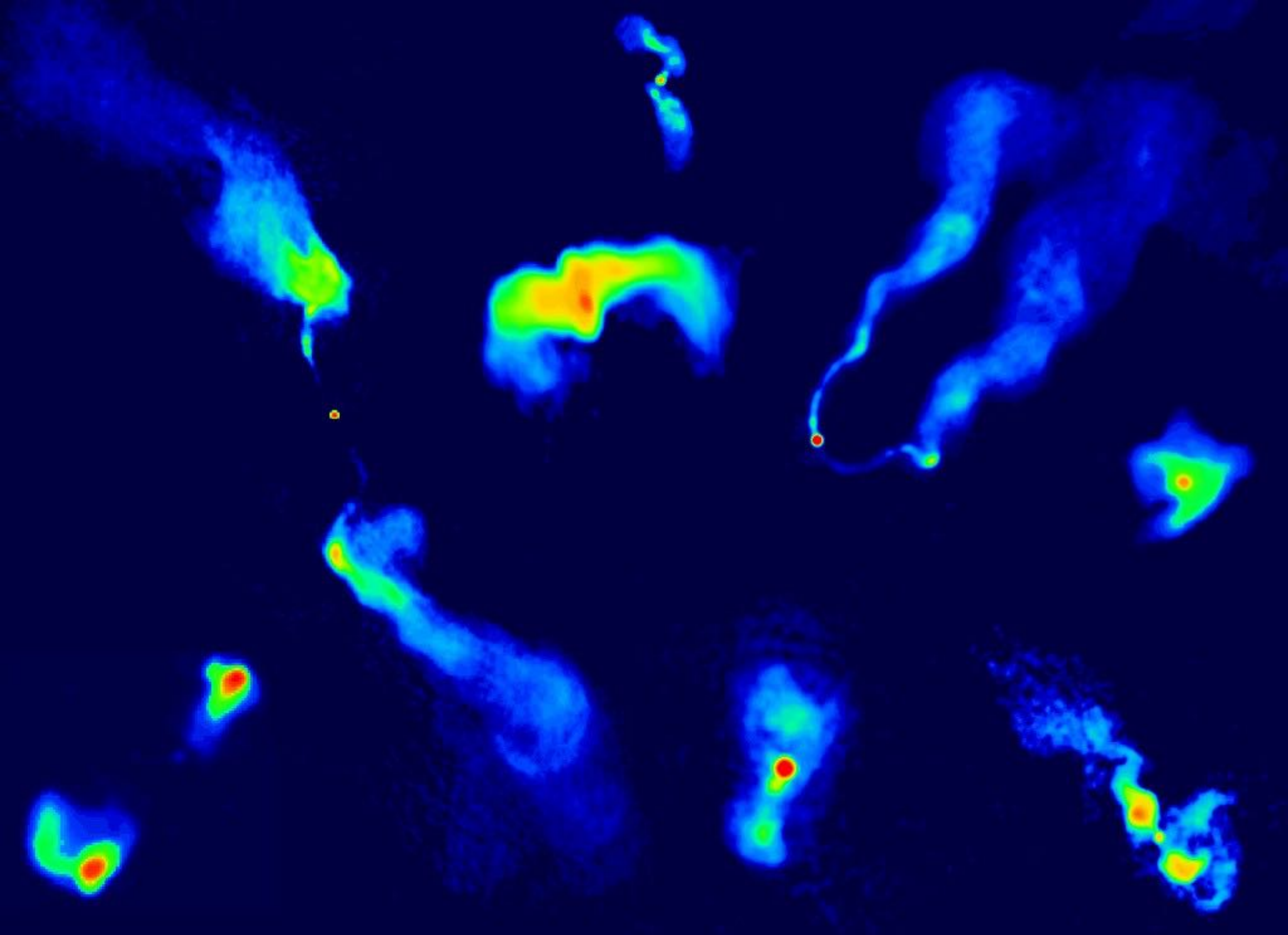
IN ASTRONOMY SUPPORT, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

http://xkcd.com/1425/

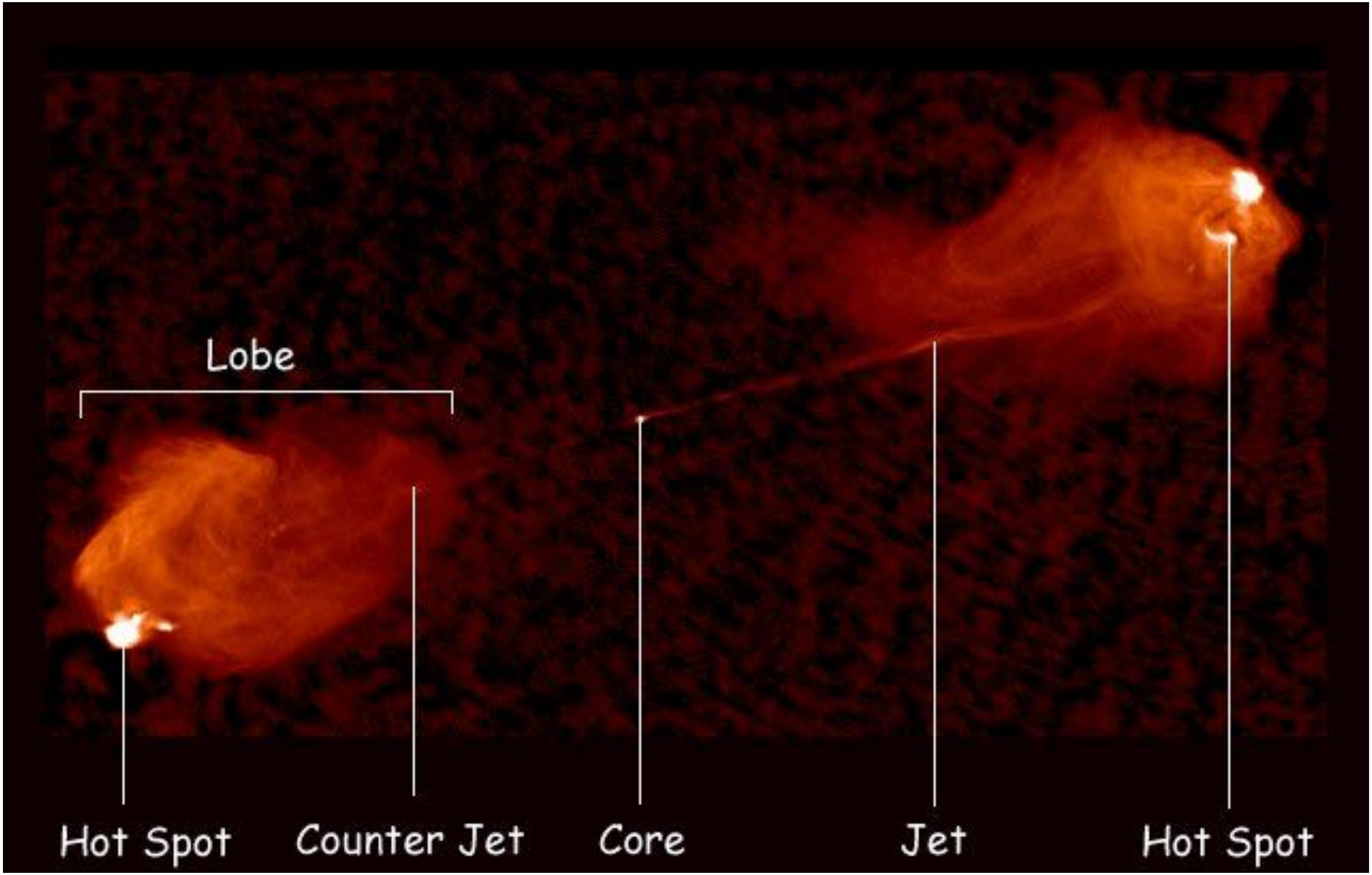

MeerKAT First Light: South African Radio Astronomy Observatory



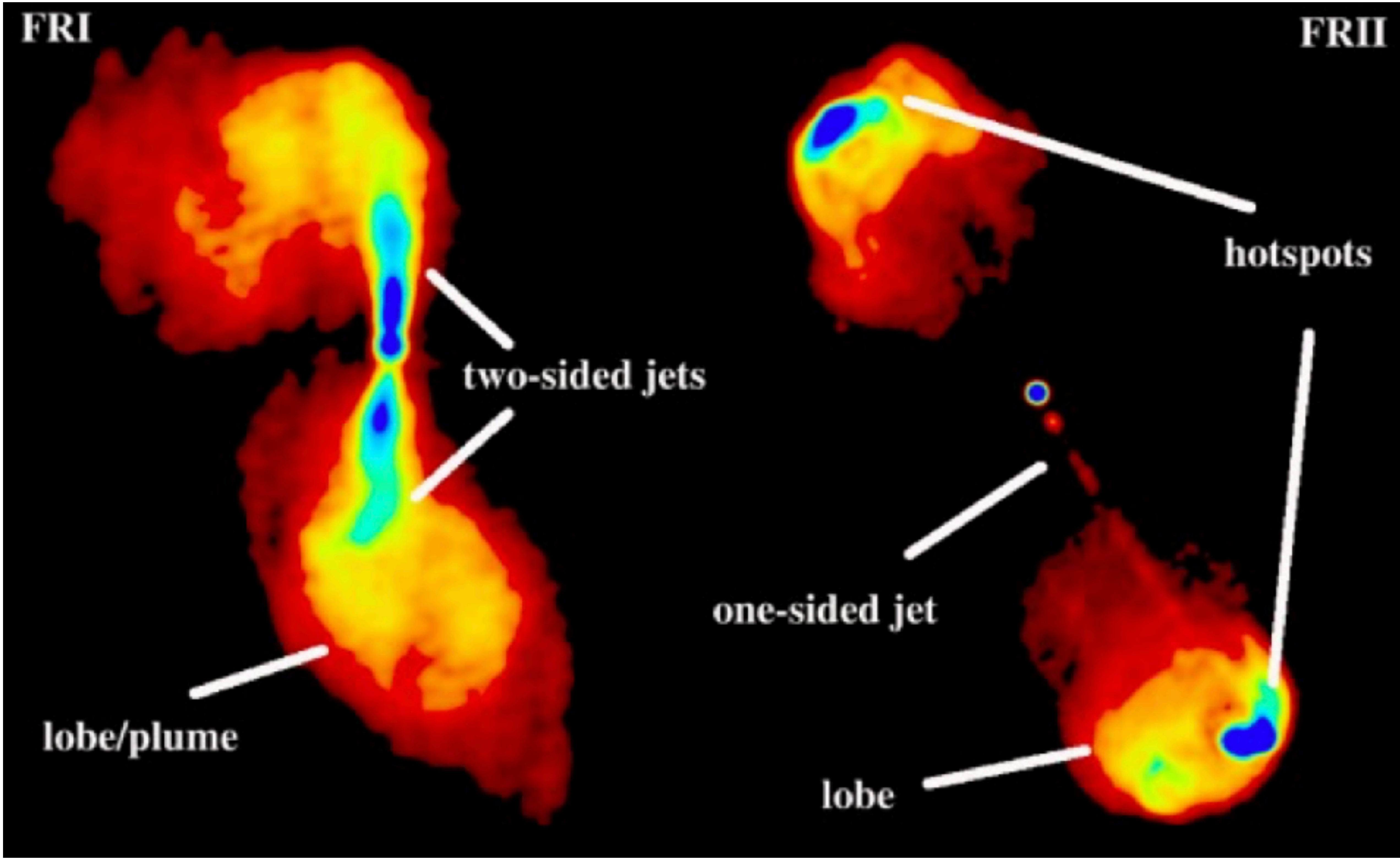
MeerKAT First Light: South African Radio Astronomy Observatory



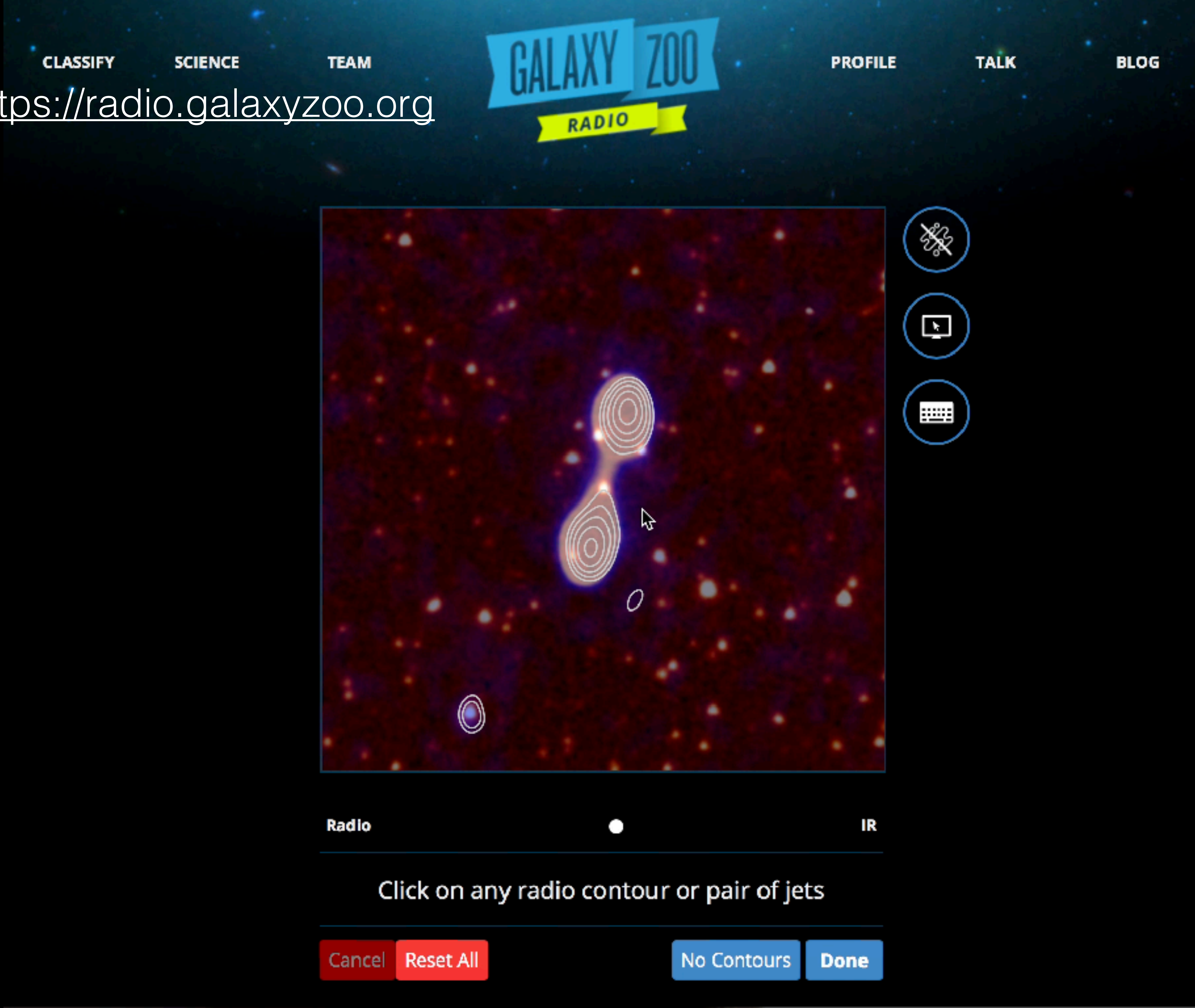
http://www.phys.unm.edu/~gbtaylor/gallery.html



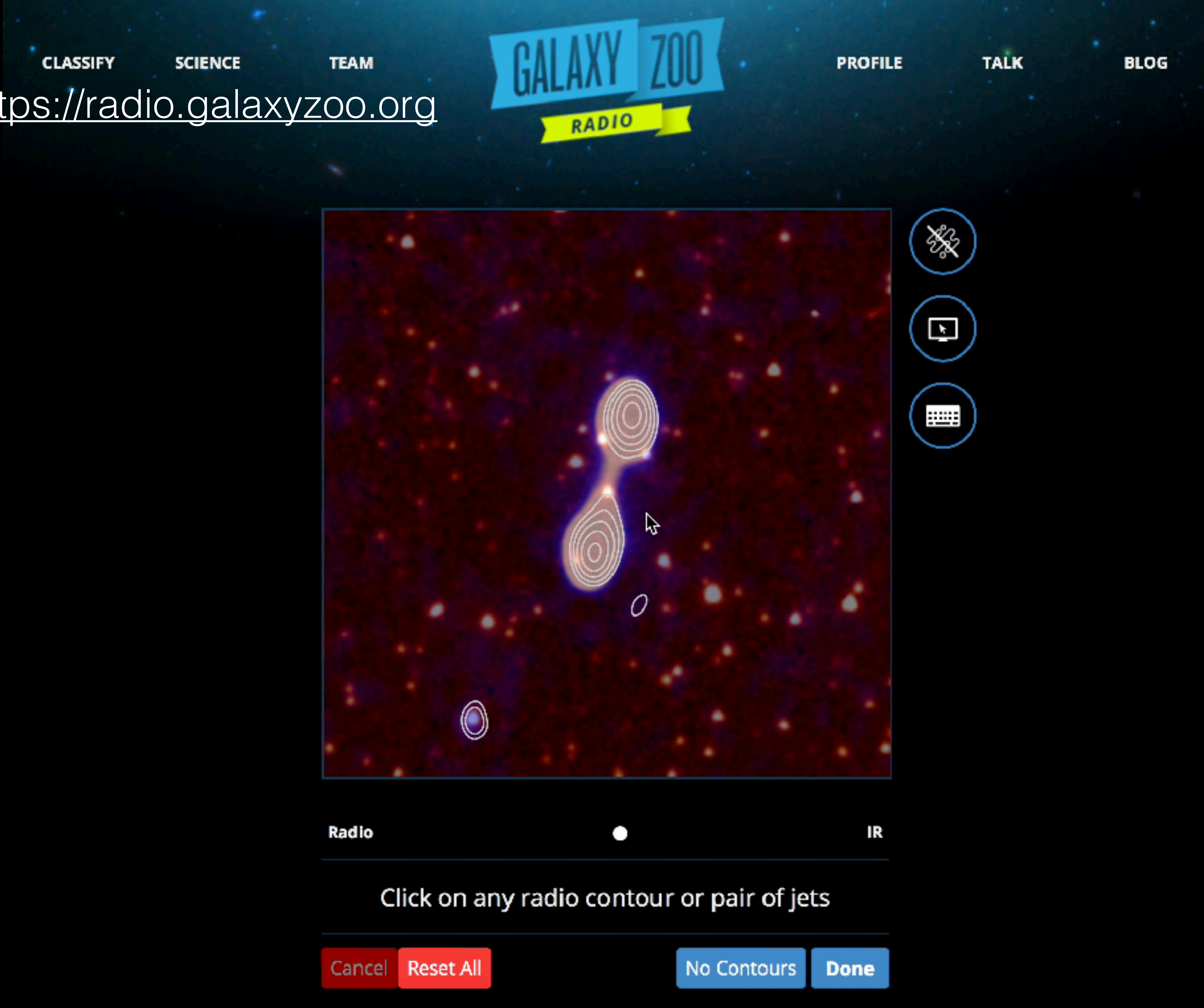
https://www.cv.nrao.edu/~abridle/dragnparts.htm


FRI
FRII
hotspots
two-sided jets
one-sided jet
lobe/plume
lobe

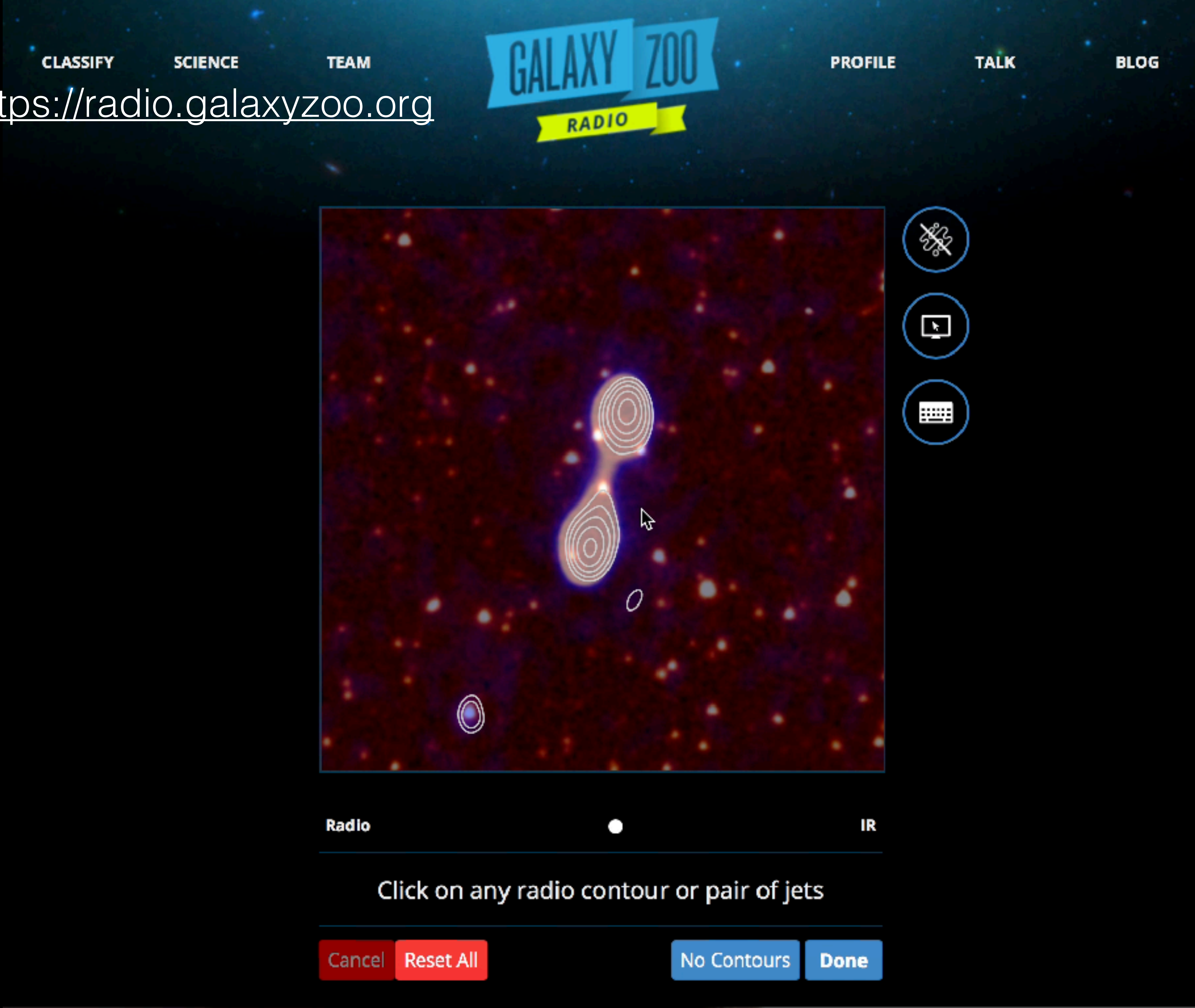
https://radio.galaxyzoo.org



https://radio.galaxyzoo.org



https://radio.galaxyzoo.org

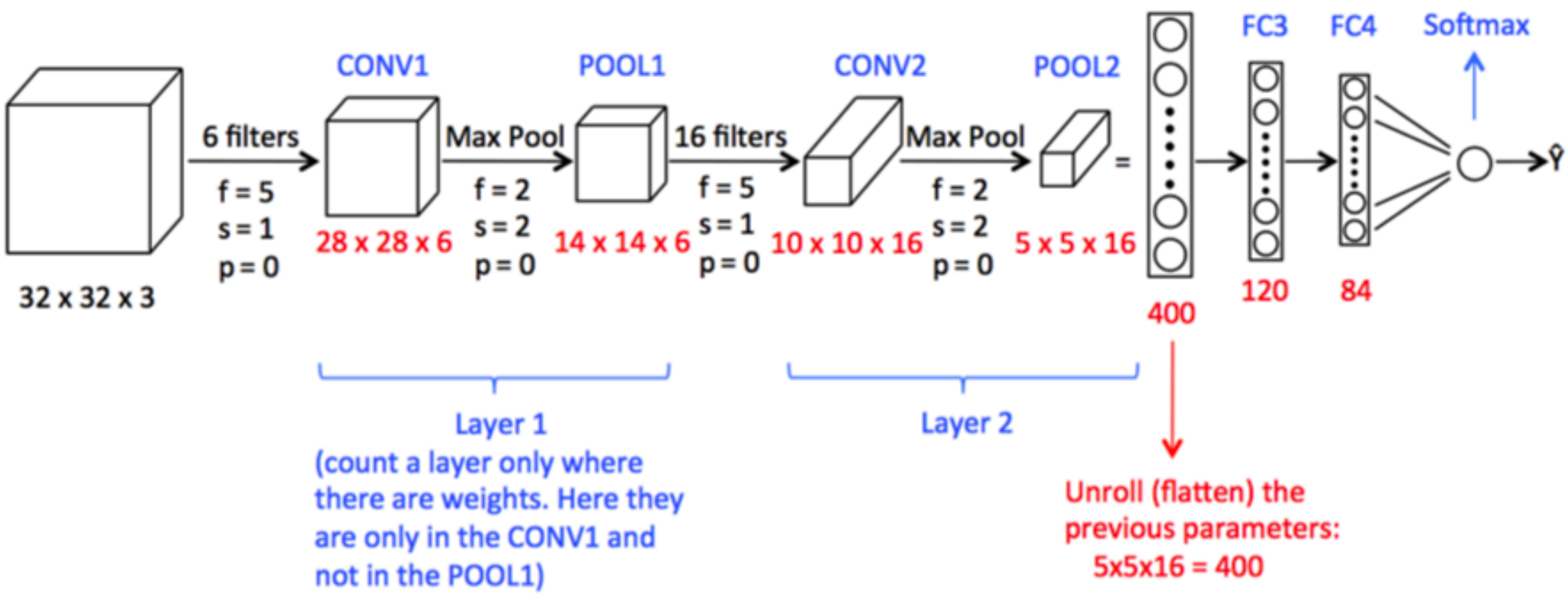


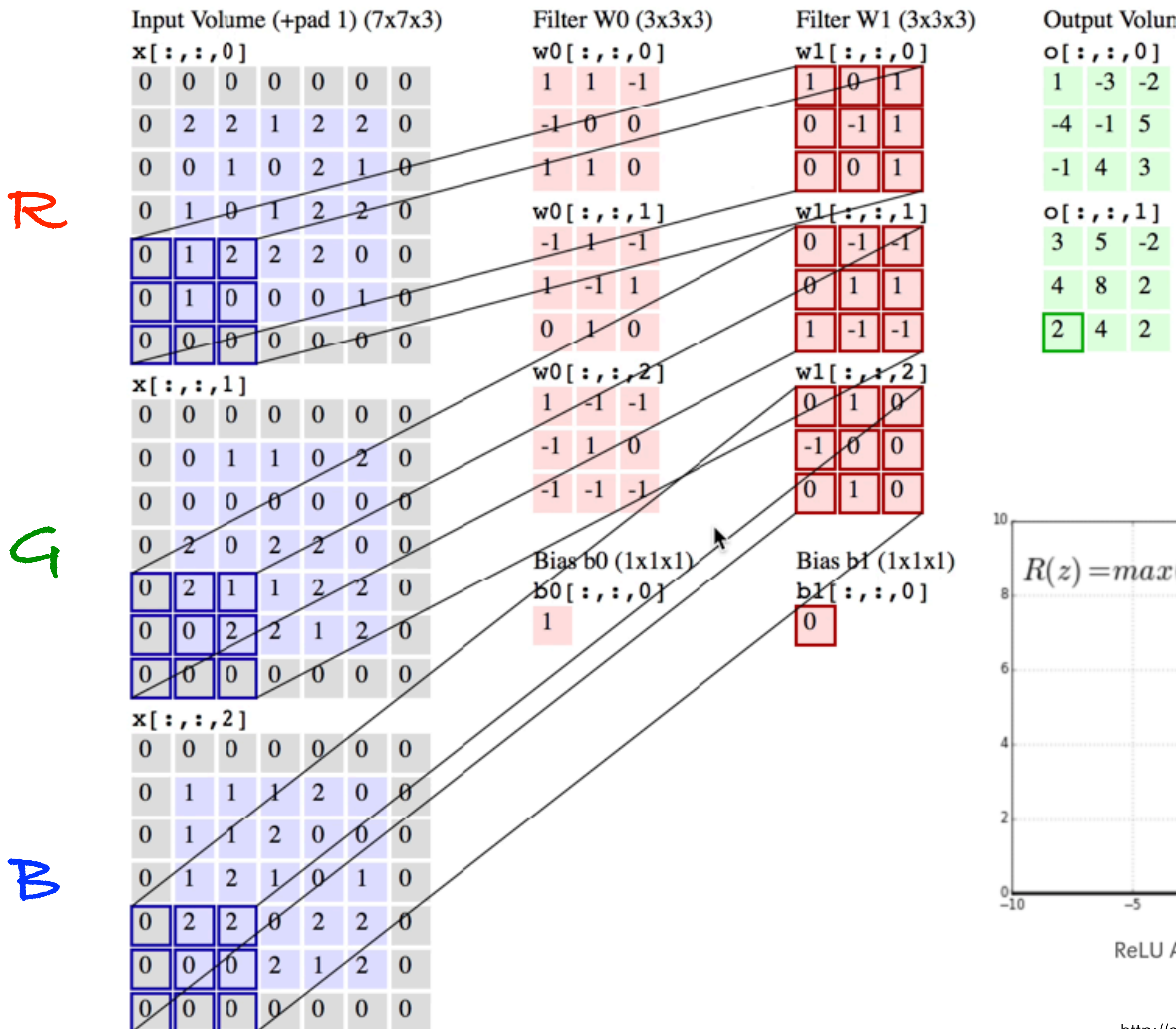
https://medium.com/machine-learning-bites/deeplearning-series-convolutional-neural-networks-a9c2f2ee1524

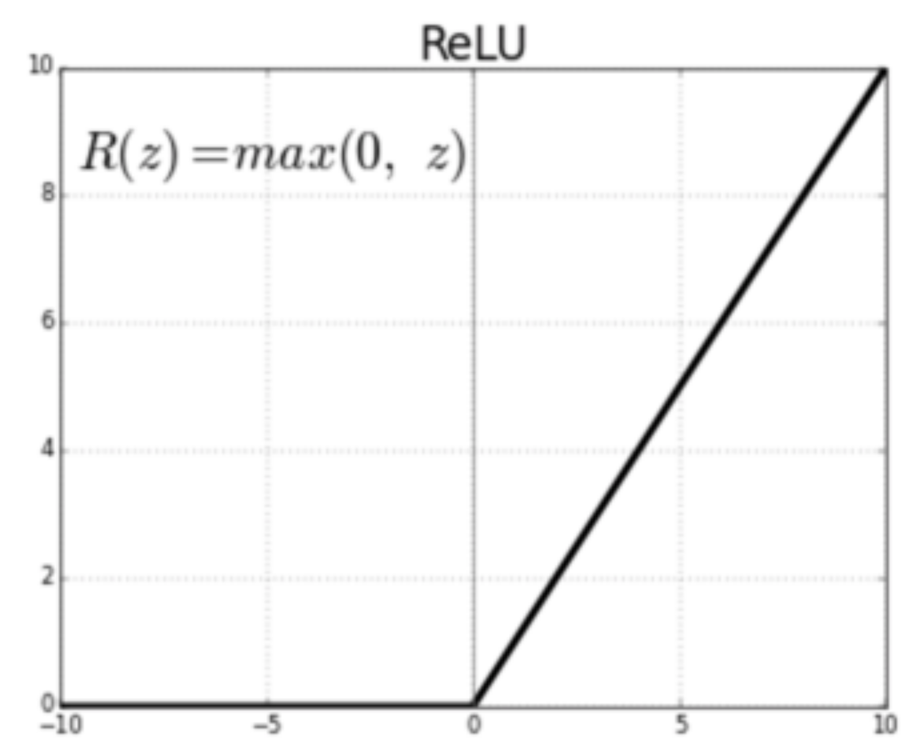
CONVOLUTIONAL NEURAL NETWORKS

3 basic types of layer:

- Convolutional Layer
    - ReLu Layer
- Pooling Layer
- Fully-Connected Layer

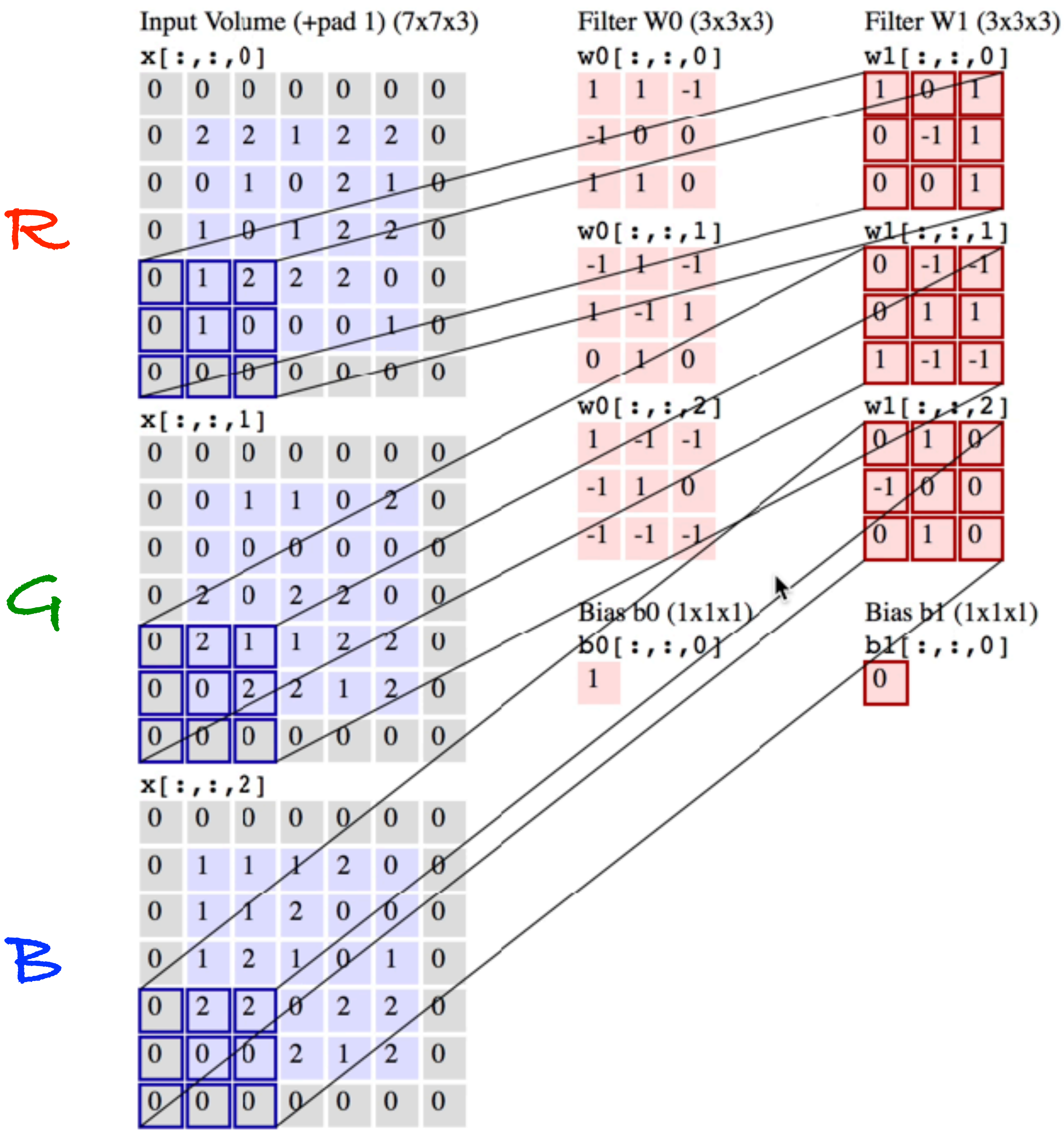


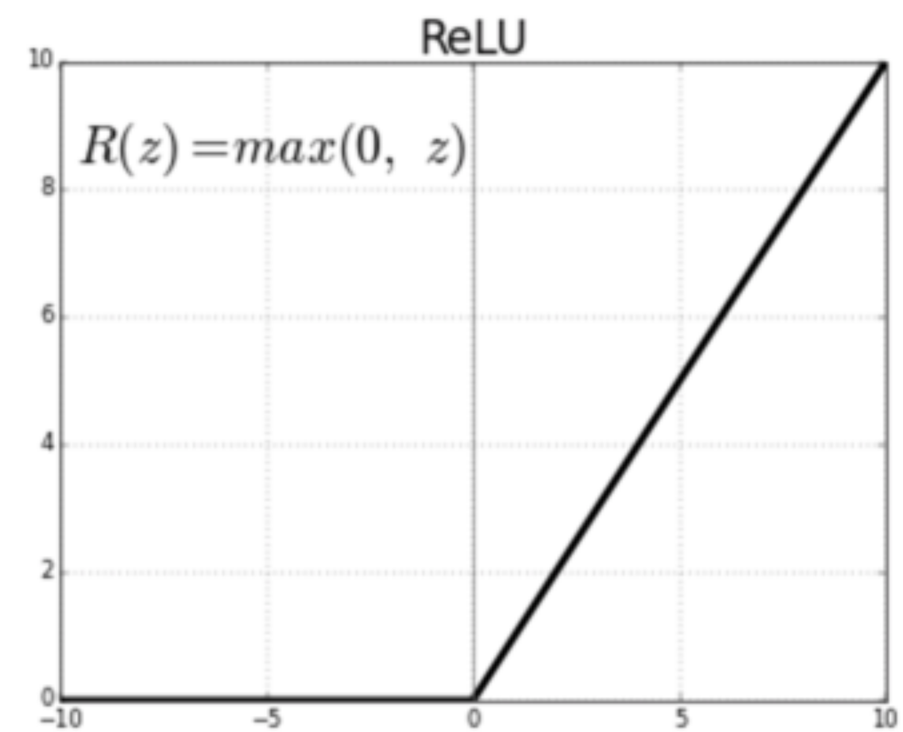




ReLU Activation Function

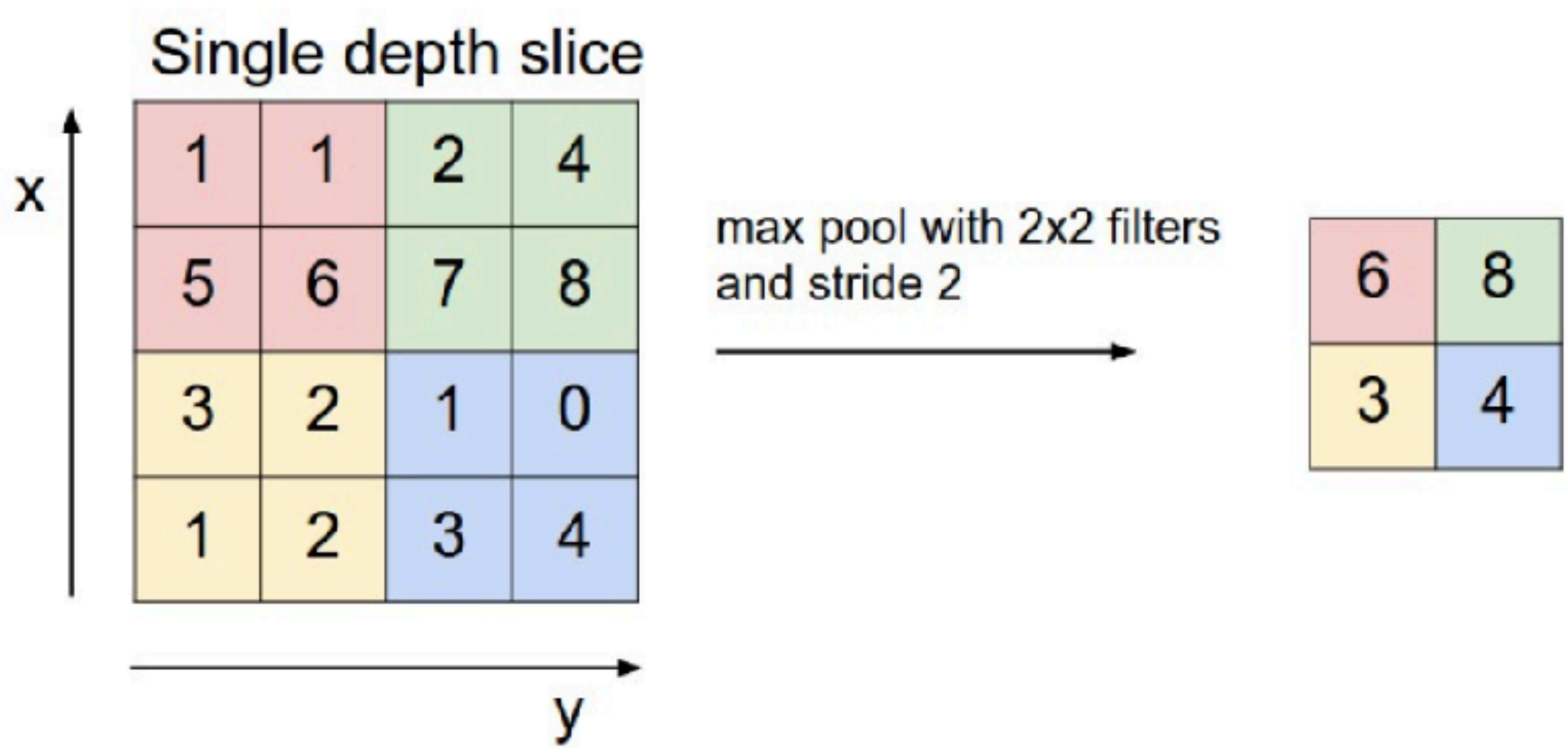
http://cs231n.github.io/convolutional-networks/





ReLU Activation Function

http://cs231n.github.io/convolutional-networks/

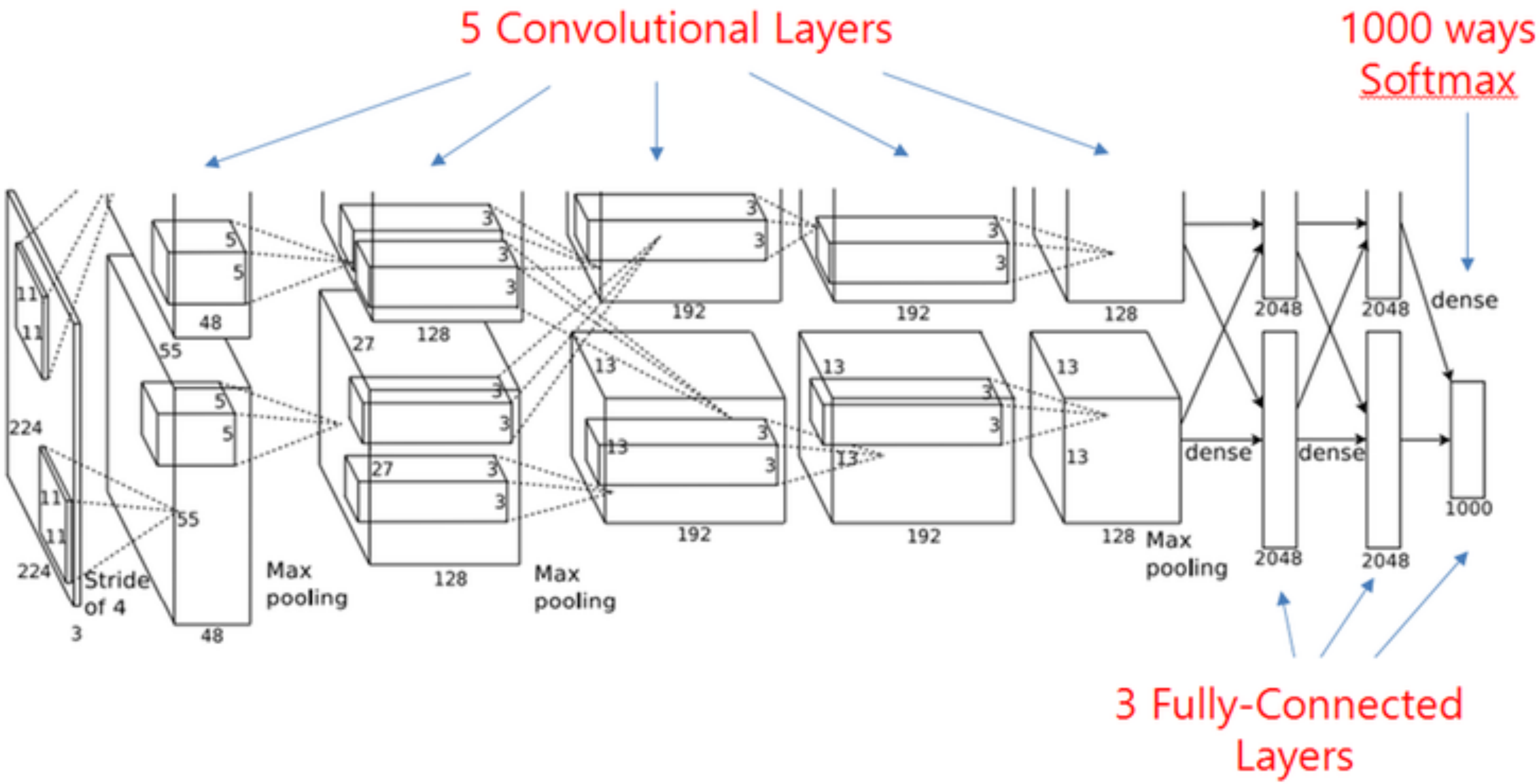


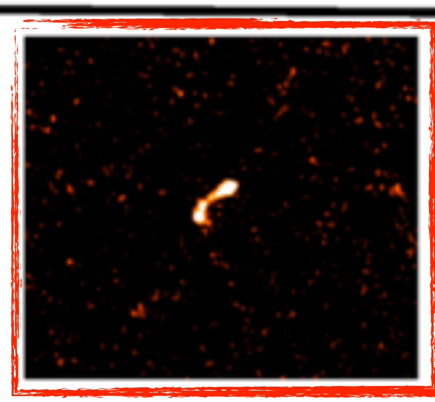
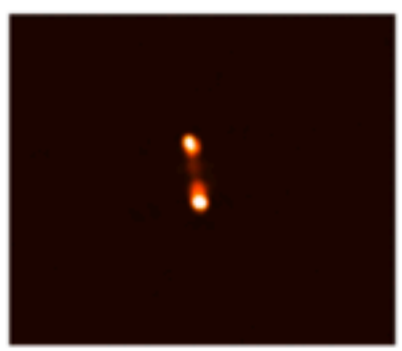
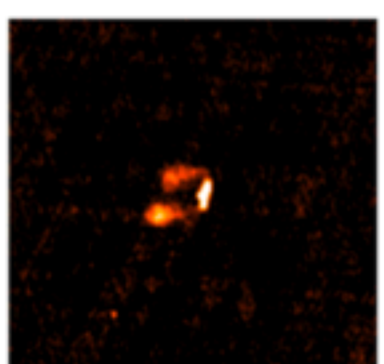
http://cs231n.github.io/convolutional-networks/

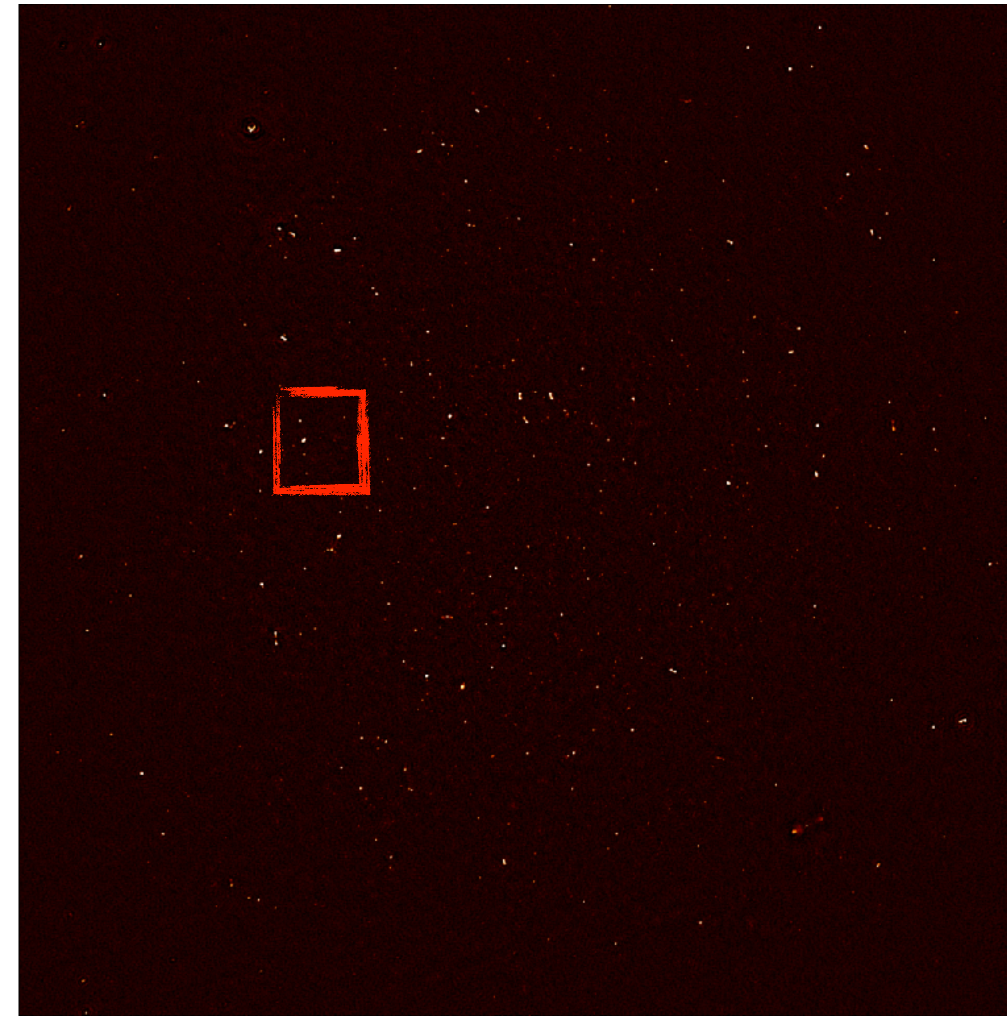
CONVOLUTIONAL NEURAL NETWORKS

- **LeNet**. The first successful applications of Convolutional Networks were developed by Yann LeCun in 1990's.
- **AlexNet**. The first work that popularized Convolutional Networks in Computer Vision. Very similar architecture to LeNet, but deeper, bigger, and with stacked Convolutional Layers.
- **ZFNet**. An improvement on AlexNet by tweaking the architecture hyperparameters, in particular by expanding the size of the middle convolutional layers and making the stride and filter size on the first layer smaller.
- **GoogLeNet**. Includes the Inception Module that dramatically reduced the number of parameters in the network (4M, compared to AlexNet with 60M). Uses Average Pooling instead of Fully Connected layers at the top of the ConvNet, eliminating a large number of low ranked parameters.
- **VGGNet**. Showed that the depth of the network is a critical component for good performance. An extremely homogeneous architecture that only performs 3x3 convolutions and 2x2 pooling from the beginning to the end.
- **ResNet**. Features special skip connections and a heavy use of batch normalization. ResNets are currently the default state of the art Convolutional Neural Network model.

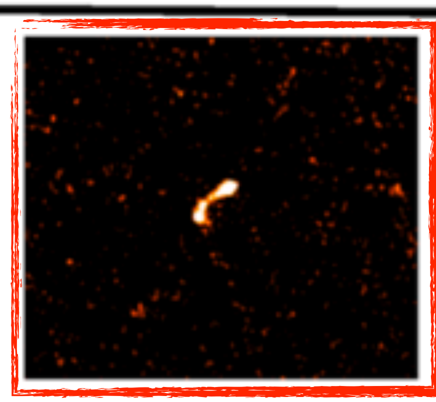
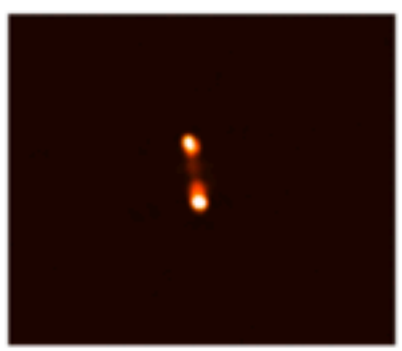
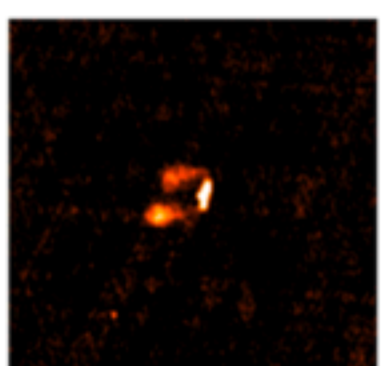
ALEXNET

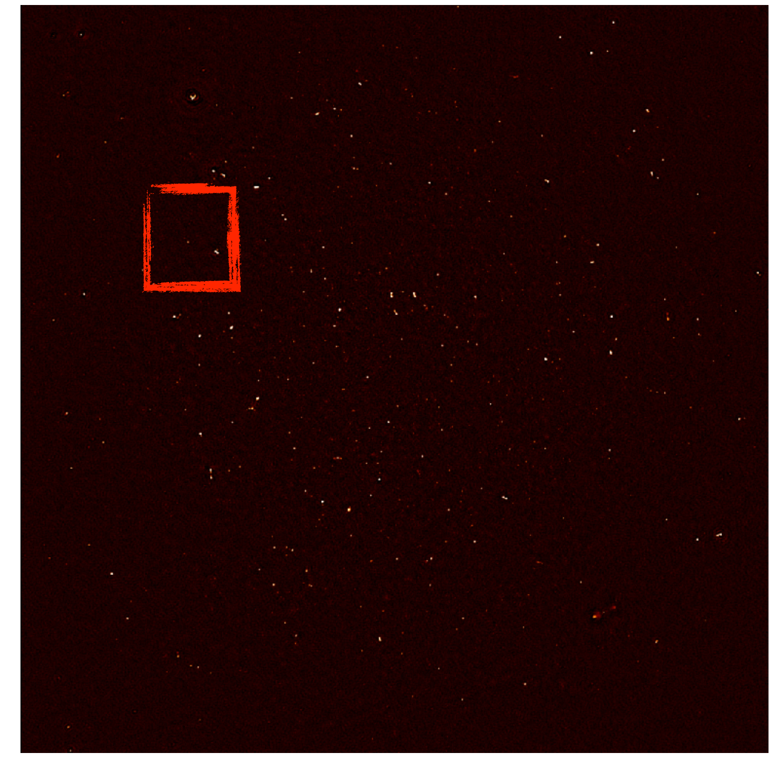

5 Convolutional Layers
1000 ways
Softmax
224
224
3
11
11
11
11
Stride
of 4
5
5
5
5
55
55
48
48
Max
pooling
27
27
3
3
3
3
3
3
128
128
Max
pooling
13
13
192
192
13
13
192
192
13
13
128
128
Max
pooling
dense
dense
dense
2048
2048
2048
2048
1000
3 Fully-Connected
Layers

Name, Coordinates	Source Image	Class True / Predicted
J101937.94+001955.7 10 19 37.94 +00 19 55.7		FRI / FRI
3C 228 09 50 10.77 +14 19 57.3		FRII / FRII
J092645.3+030916 09 47 15.887 +52 51 05.20		BENT / BENT



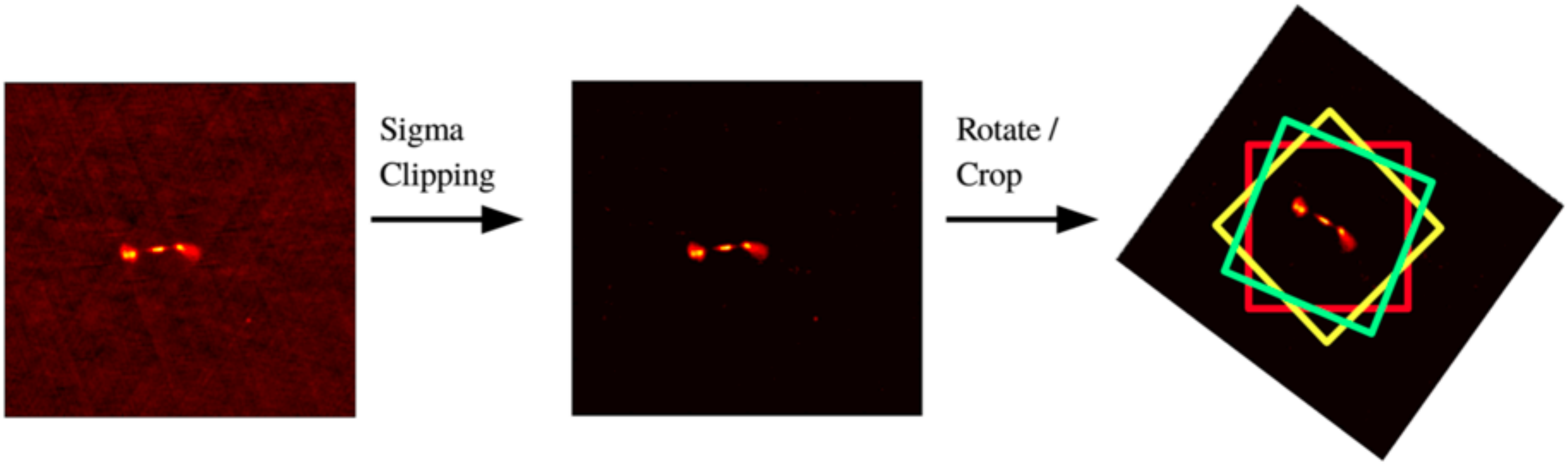
arXiv:1705.03413

Name, Coordinates	Source Image	Class True / Predicted
J101937.94+001955.7 10 19 37.94 +00 19 55.7		FRI / FRI
3C 228 09 50 10.77 +14 19 57.3		FRII / FRII
J092645.3+030916 09 47 15.887 +52 51 05.20		BENT / BENT



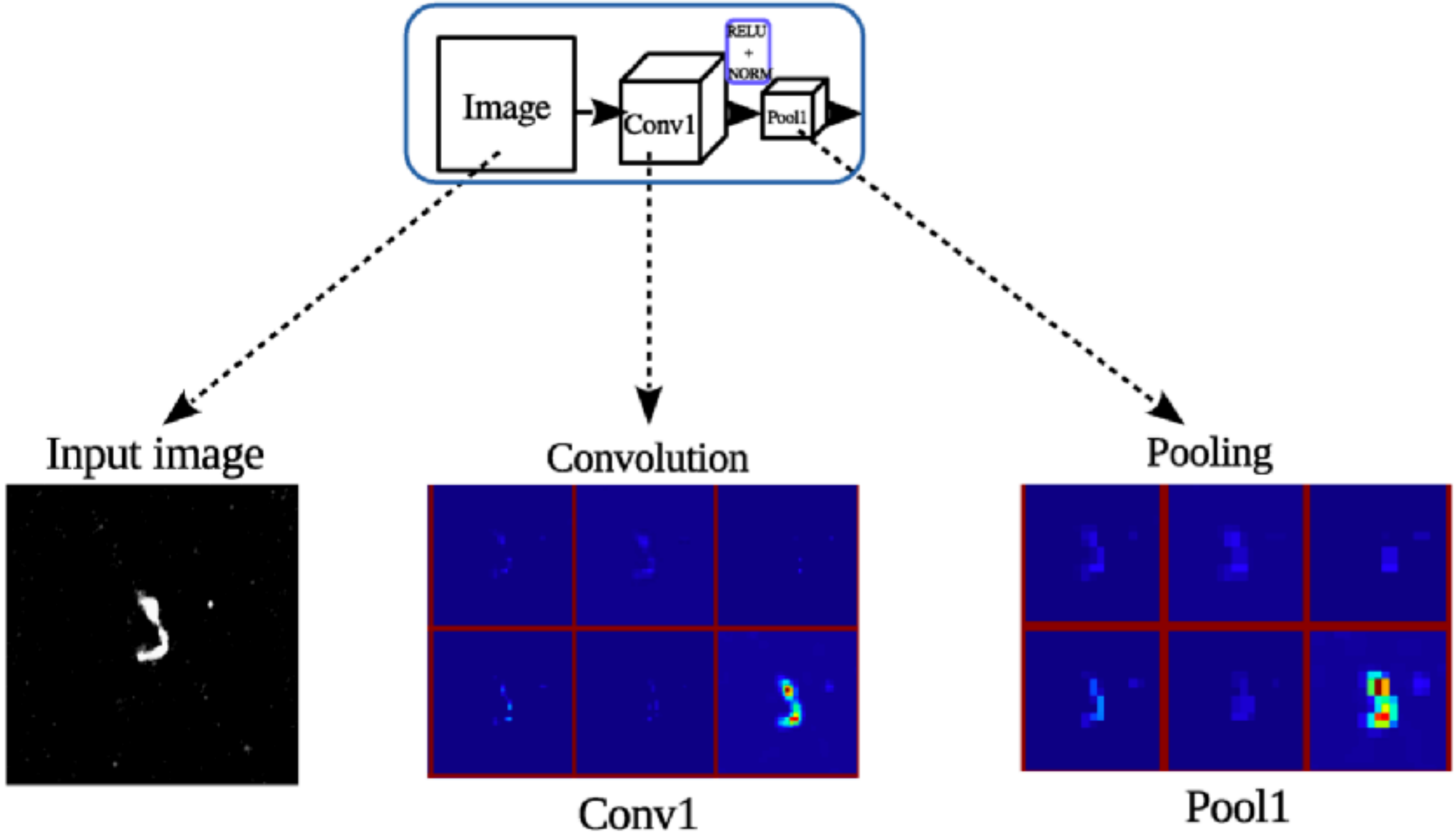
arXiv:1705.03413

RADIO GALAXY CLASSIFICATION

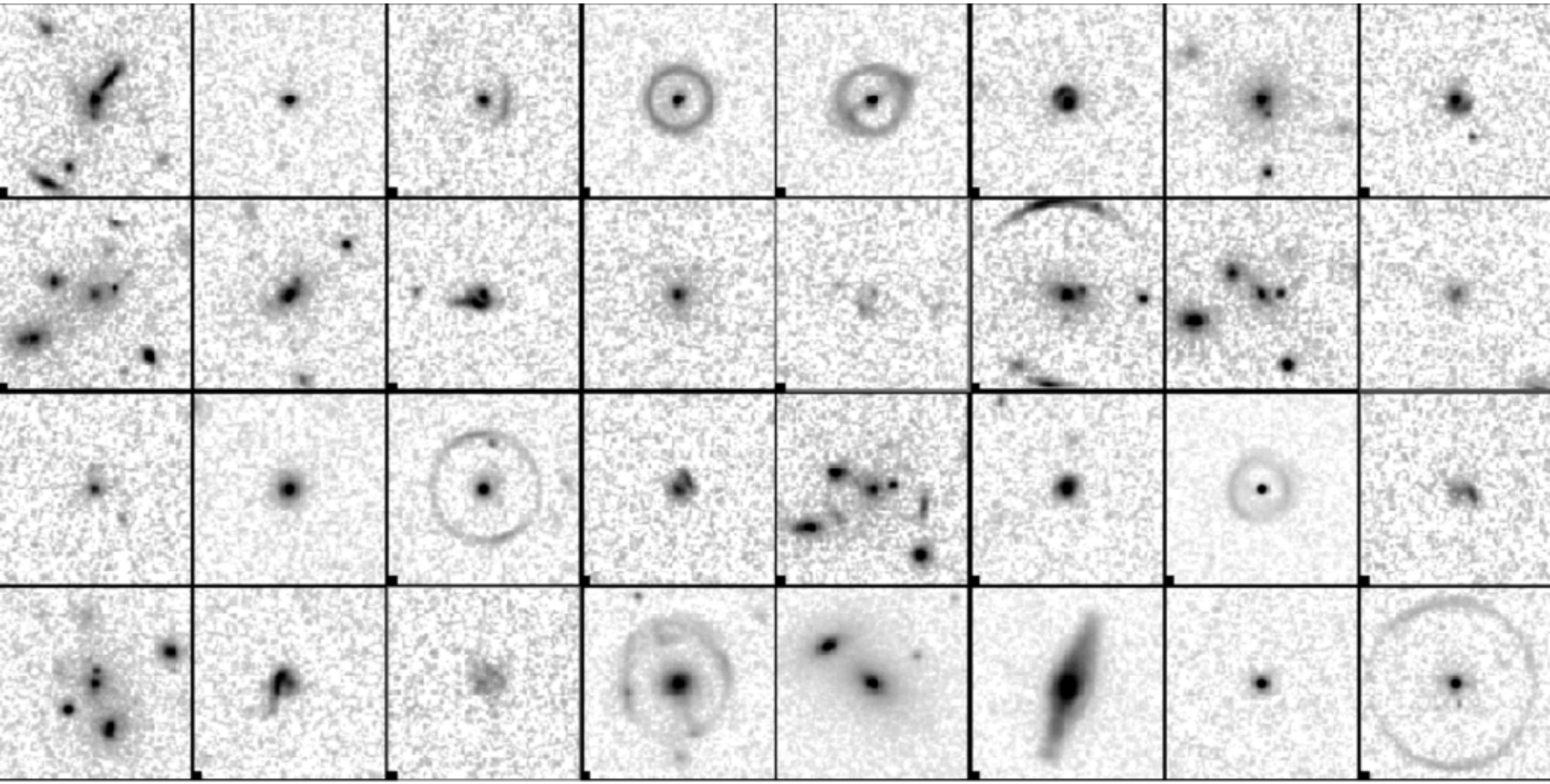


arXiv:1705.03413

RADIO GALAXY CLASSIFICATION



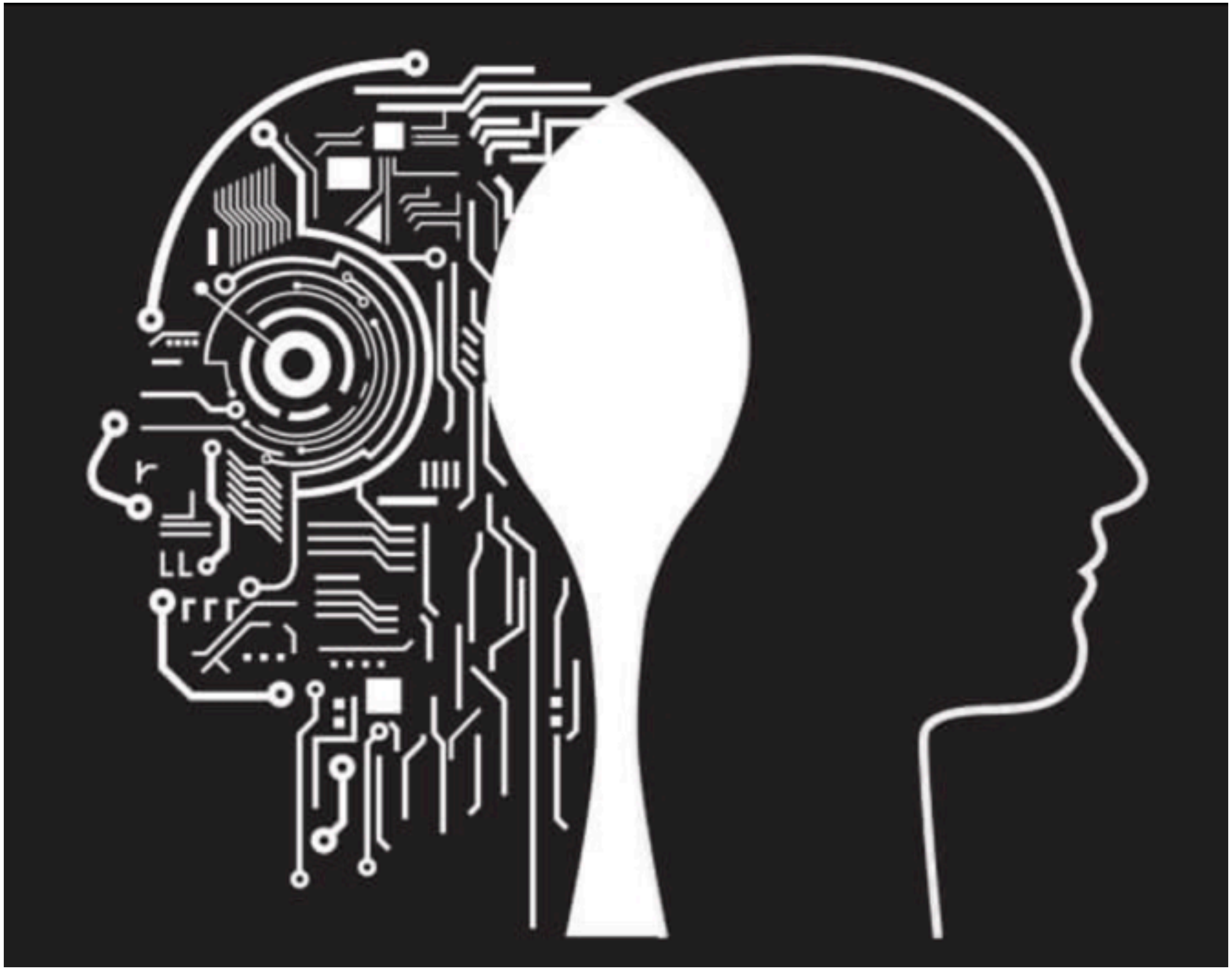
arXiv:1705.03413



arXiv:1802.03609

GRAVITATIONAL LENS FINDING CHALLENGE

100,000 simulated images, 48 hours…



Name	type	AUROC	TPR_0	TPR_{10}	short description
CMU-DeepLens-ResNet-ground3	Ground-Based	0.98	0.09	0.45	CNN
CMU-DeepLens-Resnet-Voting	Ground-Based	0.98	0.02	0.10	CNN
LASTRO EPFL	Ground-Based	0.97	0.07	0.11	CNN
CAS Swinburne Melb	Ground-Based	0.96	0.02	0.08	CNN
AstrOmatic	Ground-Based	0.96	0.00	0.01	CNN
Manchester SVM	Ground-Based	0.93	0.22	0.35	SVM / Gabor
Manchester-NA2	Ground-Based	0.89	0.00	0.01	Human Inspection
ALL-star	Ground-Based	0.84	0.01	0.02	edges/gradiants and Logistic Reg.
CAST	Ground-Based	0.83	0.00	0.00	CNN / SVM
YattaLensLite	Ground-Based	0.82	0.00	0.00	SExtractor
LASTRO EPFL	Space-Based	0.93	0.00	0.08	CNN
CMU-DeepLens-ResNet	Space-Based	0.92	0.22	0.29	CNN
GAMOCLASS	Space-Based	0.92	0.07	0.36	CNN
CMU-DeepLens-Resnet-Voting	Space-Based	0.91	0.00	0.01	CNN
AstrOmatic	Space-Based	0.91	0.00	0.01	CNN
CMU-DeepLens-ResNet-aug	Space-Based	0.91	0.00	0.00	CNN
Kapteyn Resnet	Space-Based	0.82	0.00	0.00	CNN
CAST	Space-Based	0.81	0.07	0.12	CNN
Manchester1	Space-Based	0.81	0.01	0.17	Human Inspection
Manchester SVM	Space-Based	0.81	0.03	0.08	SVM / Gabor
NeuralNet2	Space-Based	0.76	0.00	0.00	CNN / wavelets
YattaLensLite	Space-Based	0.76	0.00	0.00	Arcs / SExtractor
All-now	Space-Based	0.73	0.05	0.07	edges/gradiants and Logistic Reg.
GAHEC IRAP	Space-Based	0.66	0.00	0.01	arc finder

Table 3. The AUROC, TPR_0 and TPR_{10} for the entries in order of AUROC.

arXiv:1802.03609

Name	type	AUROC	TPR_0	TPR_{10}	short description
CMU-DeepLens-ResNet-ground3	Ground-Based	0.98	0.09	0.45	CNN
CMU-DeepLens-Resnet-Voting	Ground-Based	0.98	0.02	0.10	CNN
LASTRO EPFL	Ground-Based	0.97	0.07	0.11	CNN
CAS Swinburne Melb	Ground-Based	0.96	0.02	0.08	CNN
AstrOmatic	Ground-Based	0.96	0.00	0.01	CNN
Manchester SVM	Ground-Based	0.93	0.22	0.35	SVM / Gabor
Manchester-NA2	Ground-Based	0.89	0.00	0.01	Human Inspection
ALL-star	Ground-Based	0.84	0.01	0.02	edges/gradiants and Logistic Reg.
CAST	Ground-Based	0.83	0.00	0.00	CNN / SVM
YattaLensLite	Ground-Based	0.82	0.00	0.00	SExtractor
LASTRO EPFL	Space-Based	0.93	0.00	0.08	CNN
CMU-DeepLens-ResNet	Space-Based	0.92	0.22	0.29	CNN
GAMOCLASS	Space-Based	0.92	0.07	0.36	CNN
CMU-DeepLens-Resnet-Voting	Space-Based	0.91	0.00	0.01	CNN
AstrOmatic	Space-Based	0.91	0.00	0.01	CNN
CMU-DeepLens-ResNet-aug	Space-Based	0.91	0.00	0.00	CNN
Kapteyn Resnet	Space-Based	0.82	0.00	0.00	CNN
CAST	Space-Based	0.81	0.07	0.12	CNN
Manchester1	Space-Based	0.81	0.01	0.17	Human Inspection
Manchester SVM	Space-Based	0.81	0.03	0.08	SVM / Gabor
NeuralNet2	Space-Based	0.76	0.00	0.00	CNN / wavelets
YattaLensLite	Space-Based	0.76	0.00	0.00	Arcs / SExtractor
All-now	Space-Based	0.73	0.05	0.07	edges/gradiants and Logistic Reg.
GAHEC IRAP	Space-Based	0.66	0.00	0.01	arc finder

Table 3. The AUROC, TPR_0 and TPR_{10} for the entries in order of AUROC.

arXiv:1802.03609

Name	type	AUROC	TPR_0	TPR_{10}	short description
CMU-DeepLens-ResNet-ground3	Ground-Based	0.98	0.09	0.45	CNN
CMU-DeepLens-Resnet-Voting	Ground-Based	0.98	0.02	0.10	CNN
LASTRO EPFL	Ground-Based	0.97	0.07	0.11	CNN
CAS Swinburne Melb	Ground-Based	0.96	0.02	0.08	CNN
AstrOmatic	Ground-Based	0.96	0.00	0.01	CNN
Manchester SVM	Ground-Based	0.93	0.22	0.35	SVM / Gabor
Manchester-NA2	Ground-Based	0.89	0.00	0.01	Human Inspection
ALL-star	Ground-Based	0.84	0.01	0.02	edges/gradiants and Logistic Reg.
CAST	Ground-Based	0.83	0.00	0.00	CNN / SVM
YattaLensLite	Ground-Based	0.82	0.00	0.00	SExtractor
LASTRO EPFL	Space-Based	0.93	0.00	0.08	CNN
CMU-DeepLens-ResNet	Space-Based	0.92	0.22	0.29	CNN
GAMOCLASS	Space-Based	0.92	0.07	0.36	CNN
CMU-DeepLens-Resnet-Voting	Space-Based	0.91	0.00	0.01	CNN
AstrOmatic	Space-Based	0.91	0.00	0.01	CNN
CMU-DeepLens-ResNet-aug	Space-Based	0.91	0.00	0.00	CNN
Kapteyn Resnet	Space-Based	0.82	0.00	0.00	CNN
CAST	Space-Based	0.81	0.07	0.12	CNN
Manchester1	Space-Based	0.81	0.01	0.17	Human Inspection
Manchester SVM	Space-Based	0.81	0.03	0.08	SVM / Gabor
NeuralNet2	Space-Based	0.76	0.00	0.00	CNN / wavelets
YattaLensLite	Space-Based	0.76	0.00	0.00	Arcs / SExtractor
All-now	Space-Based	0.73	0.05	0.07	edges/gradiants and Logistic Reg.
GAHEC IRAP	Space-Based	0.66	0.00	0.01	arc finder

Table 3. The AUROC, TPR_0 and TPR_{10} for the entries in order of AUROC.

arXiv:1802.03609

$$\frac{\mathrm{FP}}{\mathrm{TP}} \simeq \frac{\mathrm{FPR}}{\mathrm{TPR}}\left(\frac{\text{number of non-lenses in sample}}{\text{number of lenses in sample}}\right)$$

$\mathbf{TPR_0}$ = the highest TPR reached before a single false positive occurs in the test set of 100,000 cases

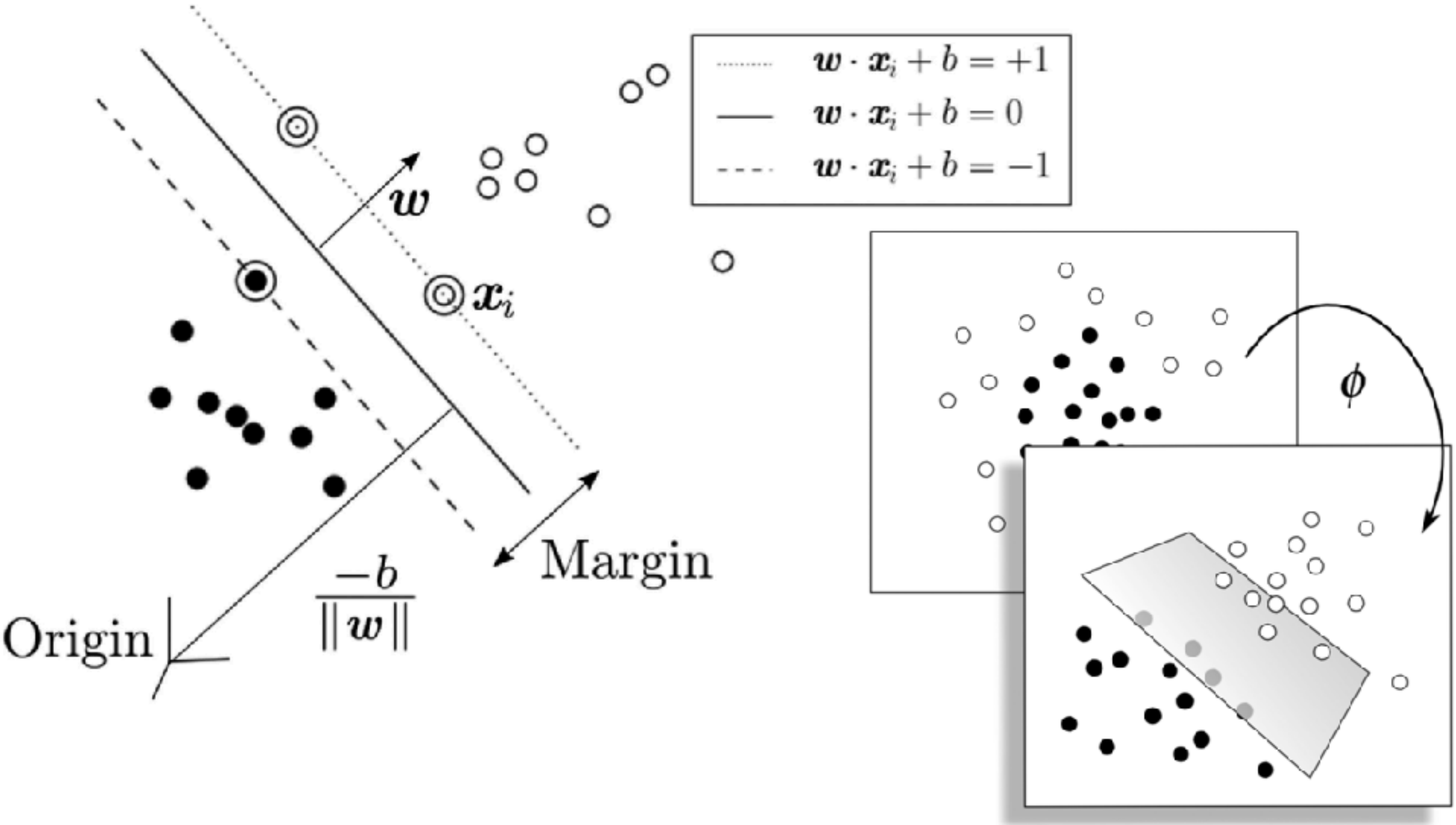
$\mathbf{TPR_{10}}$ = the TPR at the point where fewer than ten false positives are made

arXiv:1802.03609

Name	type	AUROC	TPR_0	TPR_{10}	short description
Manchester SVM	Ground-Based	0.93	0.22	0.35	SVM / Gabor
CMU-DeepLens-ResNet-ground3	Ground-Based	0.98	0.09	0.45	CNN
LASTRO EPFL	Ground-Based	0.97	0.07	0.11	CNN
CMU-DeepLens-Resnet-Voting	Ground-Based	0.98	0.02	0.10	CNN
CAS Swinburne Melb	Ground-Based	0.96	0.02	0.08	CNN
ALL-star	Ground-Based	0.84	0.01	0.02	edges/gradiants and Logistic Reg.
Manchester-NA2	Ground-Based	0.89	0.00	0.01	Human Inspection
YattaLensLite	Ground-Based	0.82	0.00	0.00	SExtractor
CAST	Ground-Based	0.83	0.00	0.00	CNN / SVM
AstrOmatic	Ground-Based	0.96	0.00	0.01	CNN
CMU-DeepLens-ResNet	Space-Based	0.92	0.22	0.29	CNN
GAMOCLASS	Space-Based	0.92	0.07	0.36	CNN
CAST	Space-Based	0.81	0.07	0.12	CNN
All-now	Space-Based	0.73	0.05	0.07	edges/gradiants and Logistic Reg.
Manchester SVM	Space-Based	0.80	0.03	0.07	SVM / Gabor
Manchester1	Space-Based	0.81	0.01	0.17	Human Inspection
LASTRO EPFL	Space-Based	0.93	0.00	0.08	CNN
GAHEC IRAP	Space-Based	0.66	0.00	0.01	arc finder
AstrOmatic	Space-Based	0.91	0.00	0.01	CNN
Kapteyn Resnet	Space-Based	0.82	0.00	0.00	CNN
CMU-DeepLens-ResNet-aug	Space-Based	0.91	0.00	0.00	CNN
CMU-DeepLens-Resnet-Voting	Space-Based	0.91	0.00	0.01	CNN
NeuralNet2	Space-Based	0.76	0.00	0.00	CNN / wavelets
YattaLensLite	Space-Based	0.76	0.00	0.00	Arcs / SExtractor

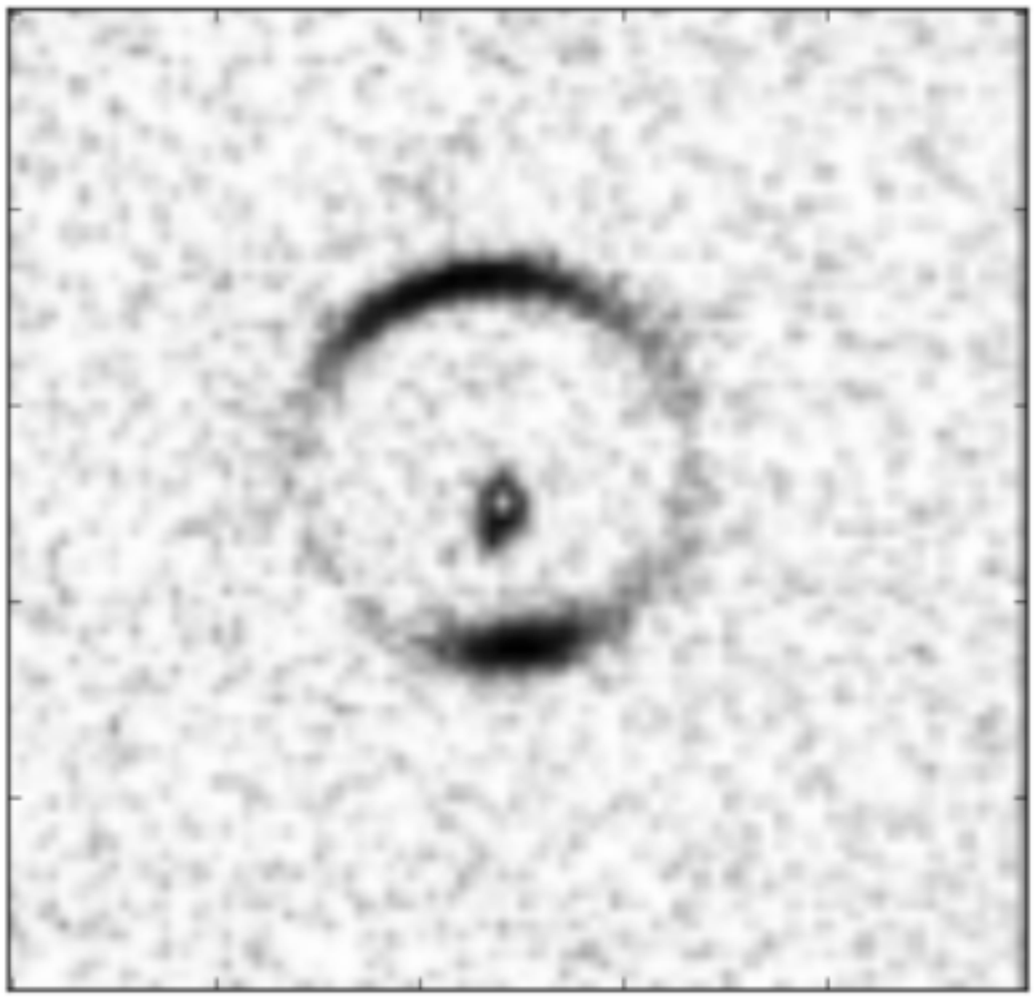
Table 4. The AUROC, TPR_0 and TPR_{10} for the entries in order of TPR_0.

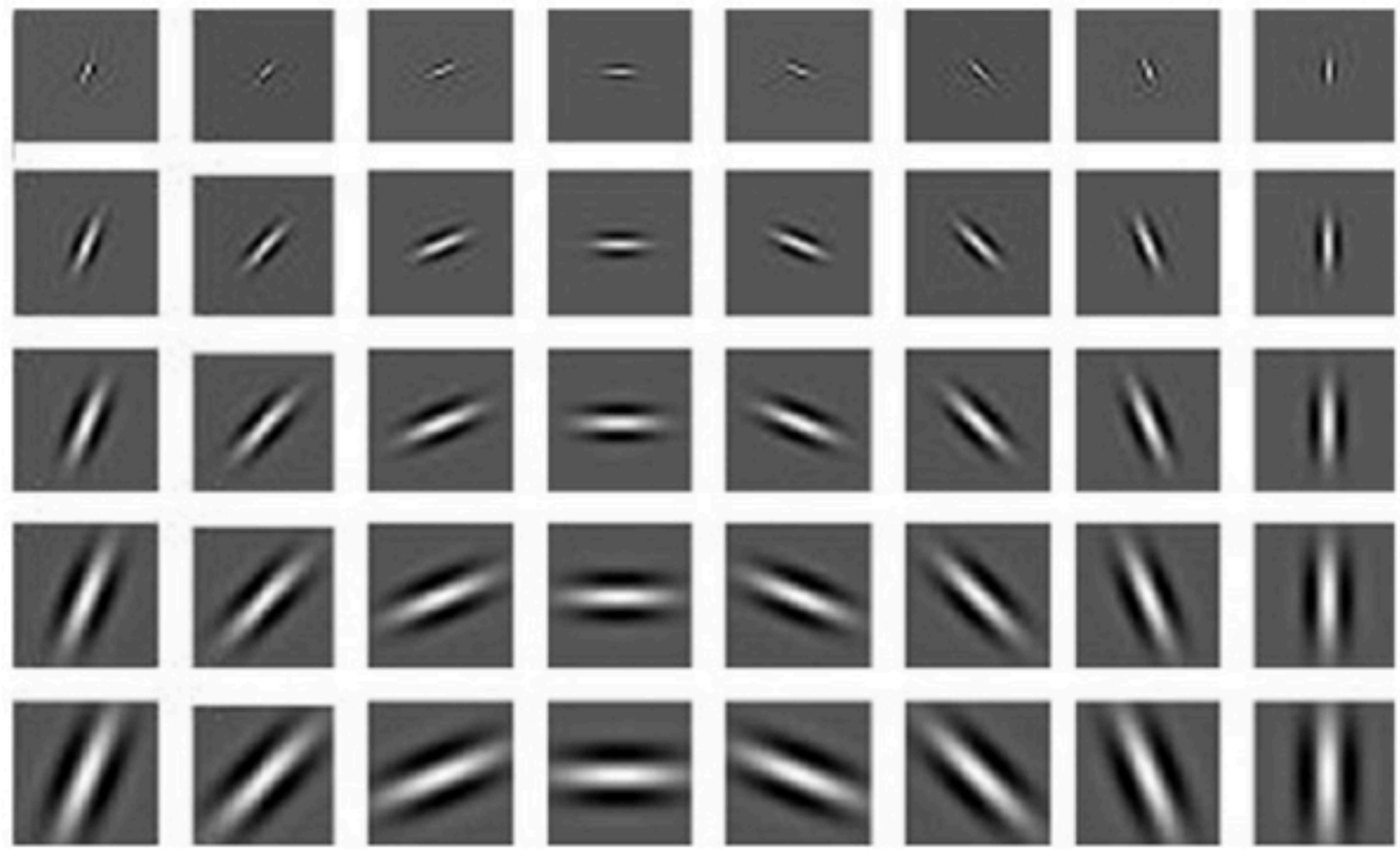
arXiv:1802.03609



arXiv:1802.03609

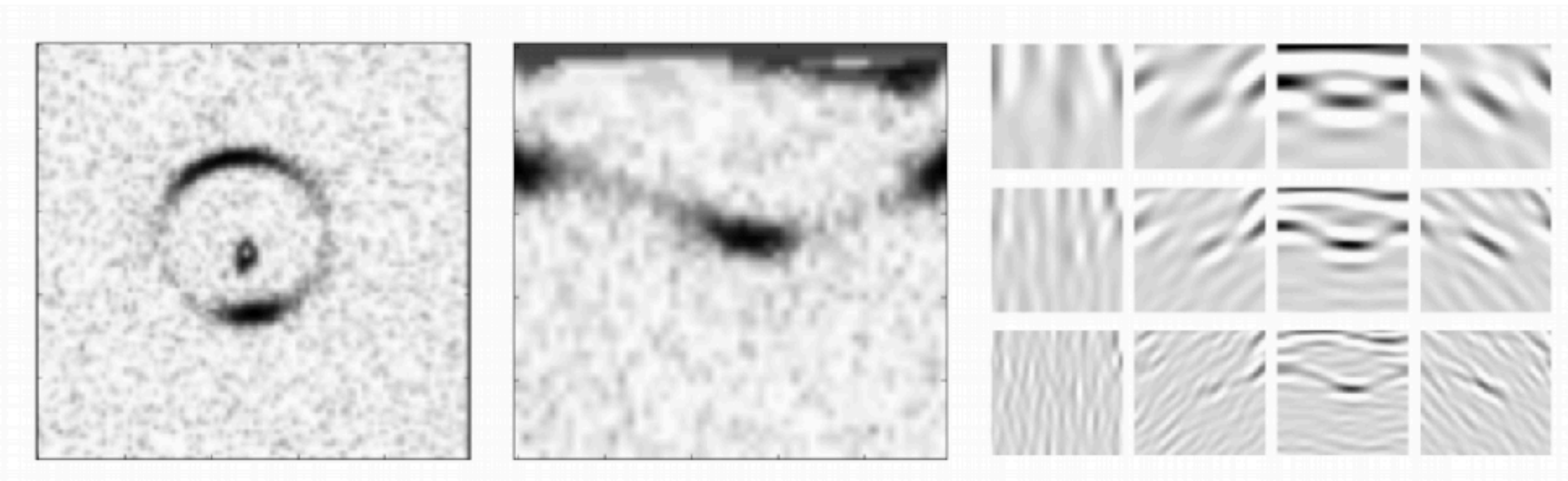
Simulated gravitational lens





Gabor filters: model the simple cells of the mammillian visual cortex (Marcelja 1980)

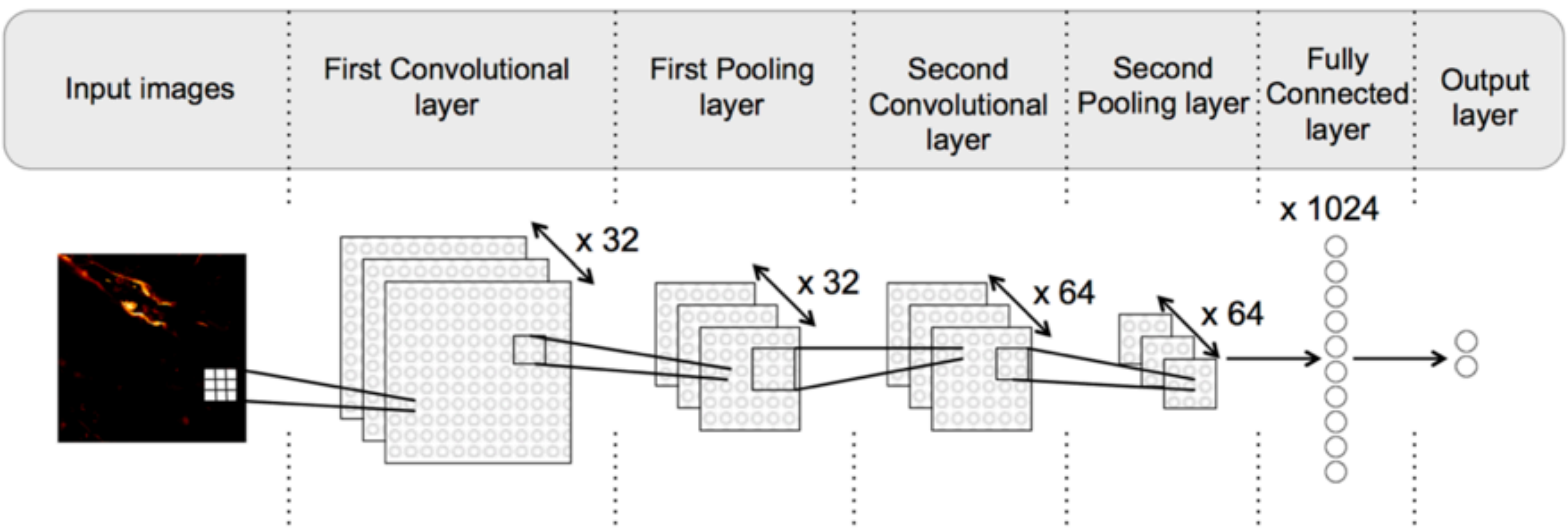
input data → **polar transform** → **Gabor filterbank**



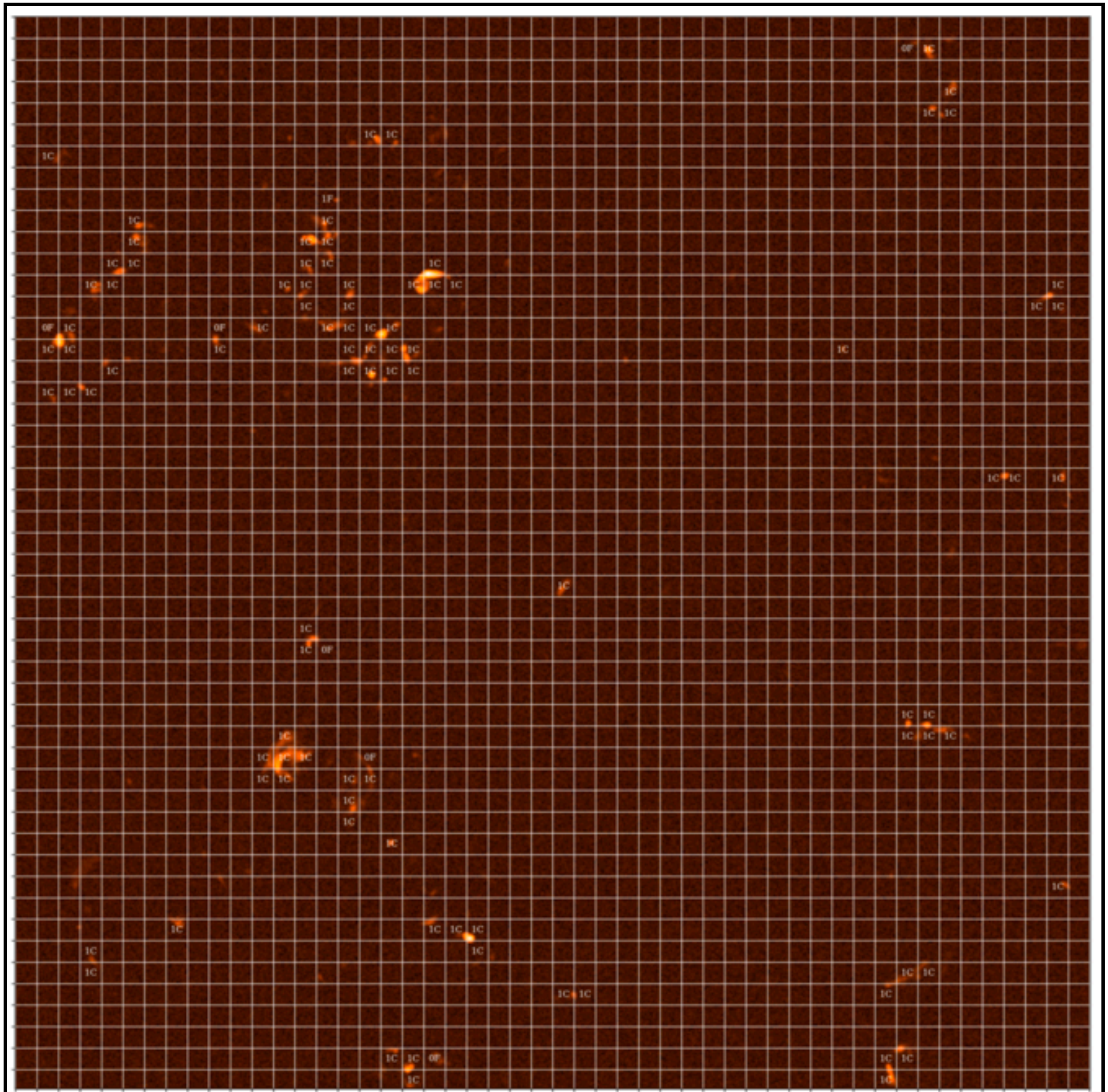
↓

calculate moments (mean, variance, skew, kurtosis, local energy)

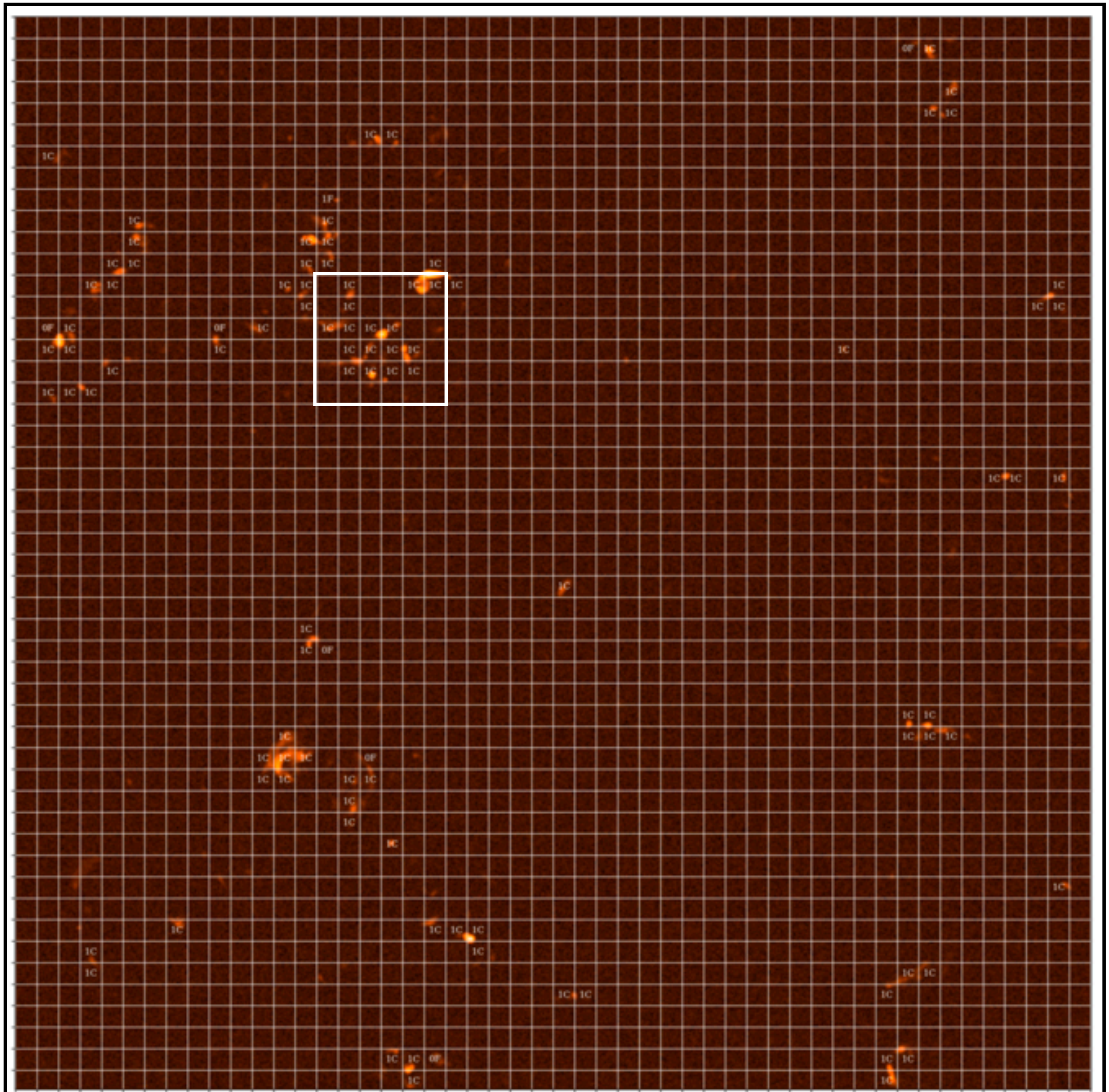
1260 features per example

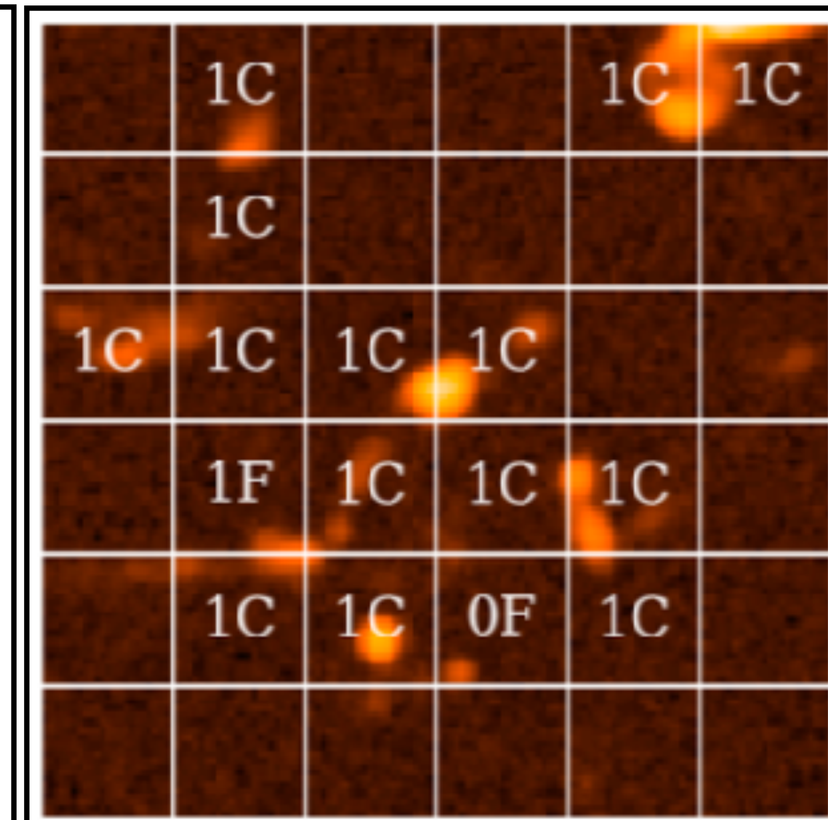


arXiv:1809.03315

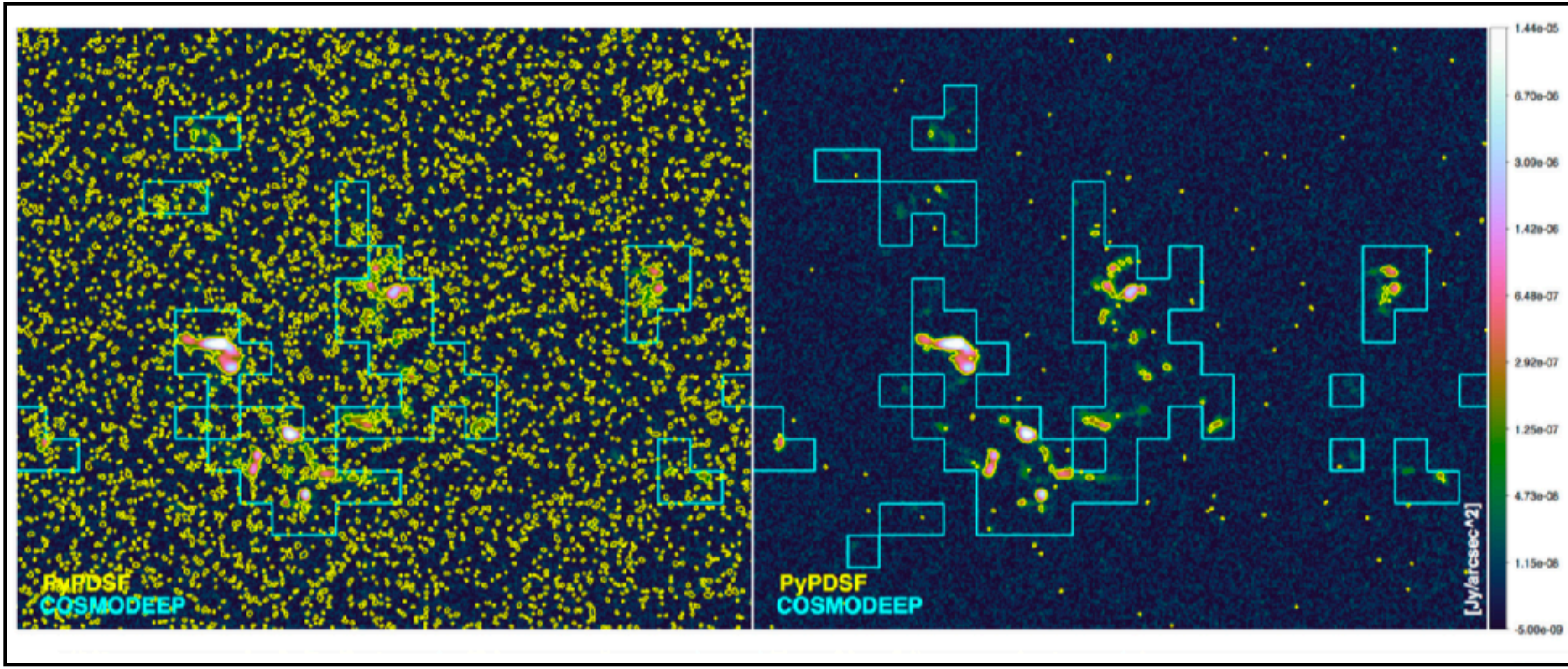


arXiv:1809.03315



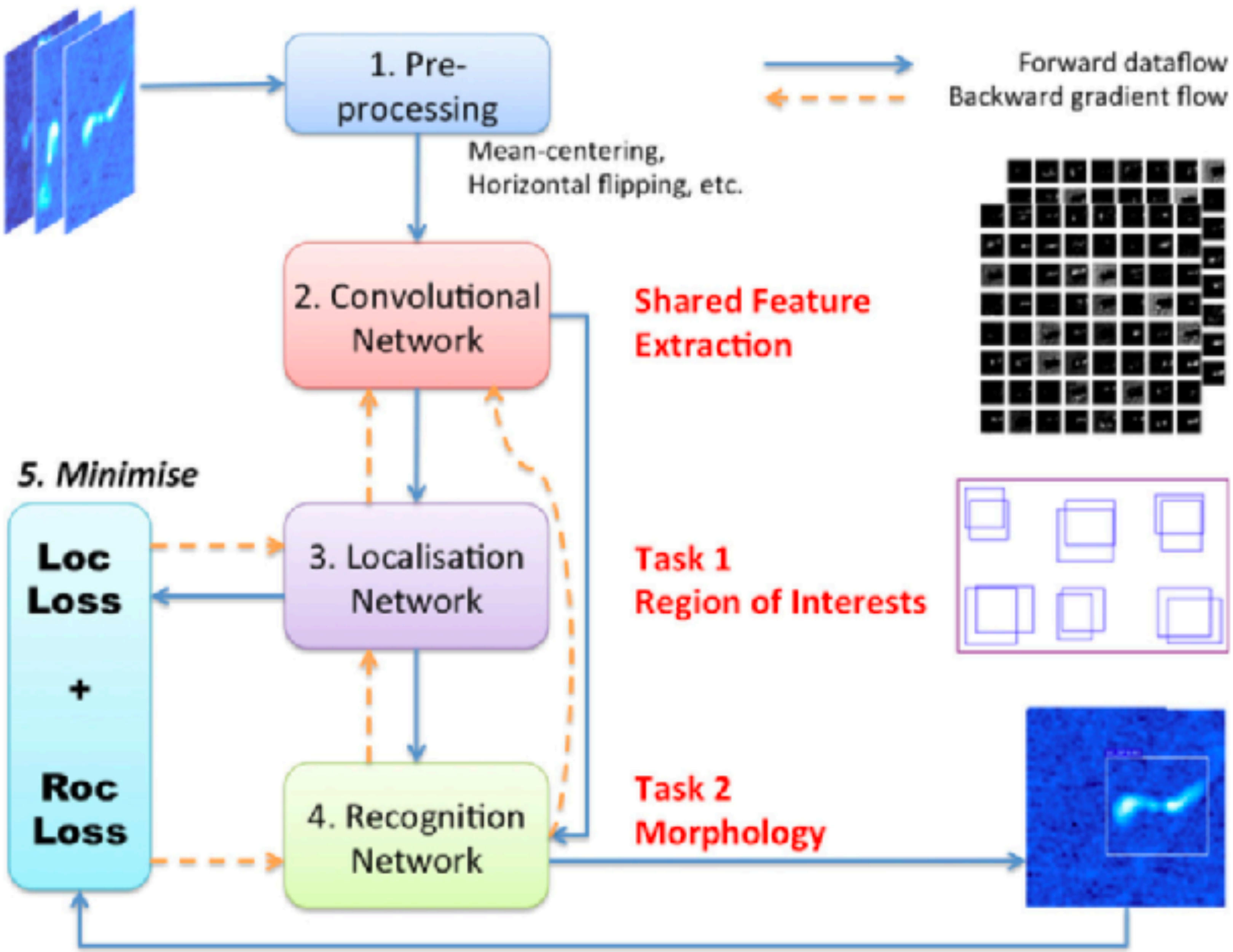


arXiv:1809.03315

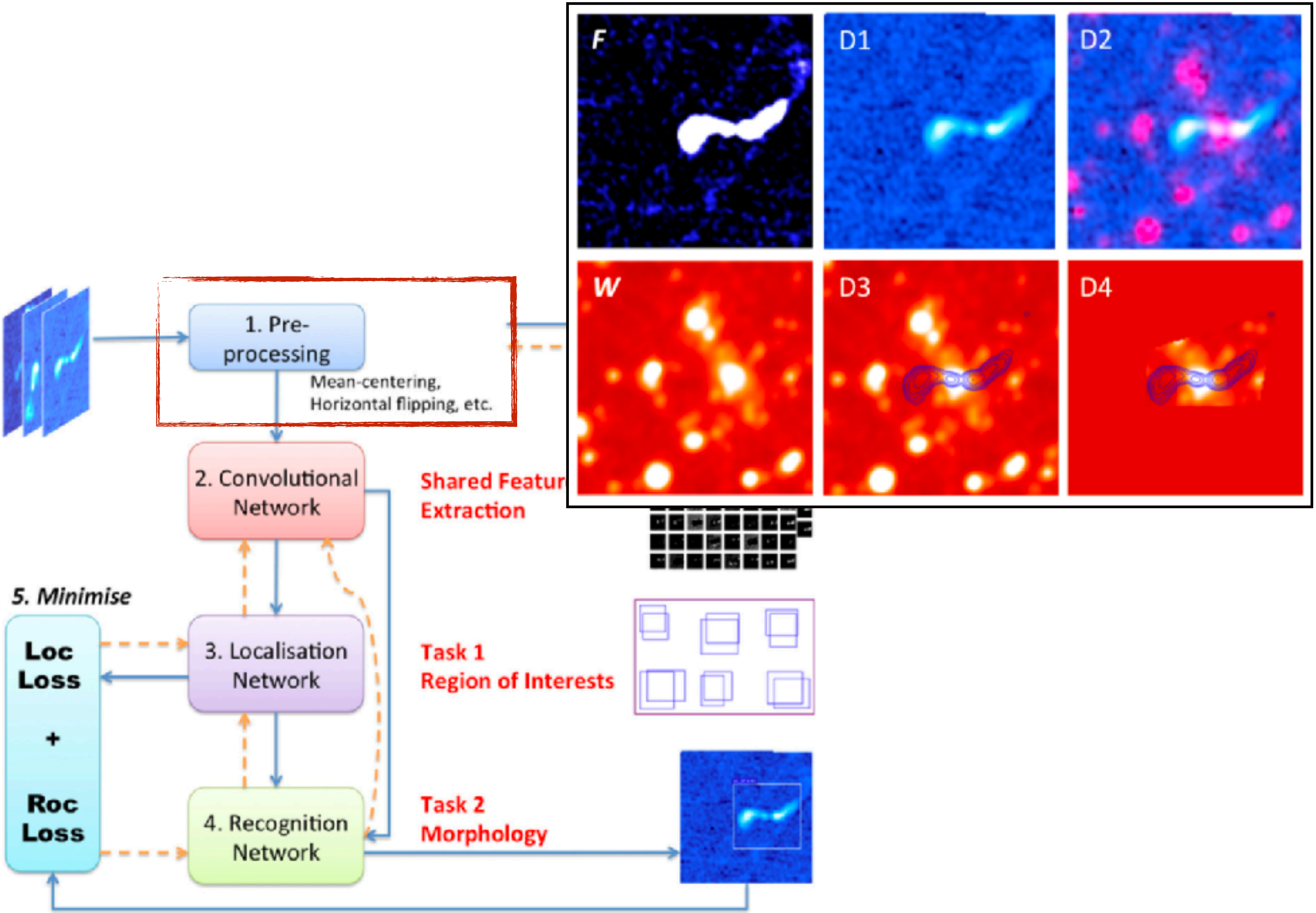


arXiv:1809.03315

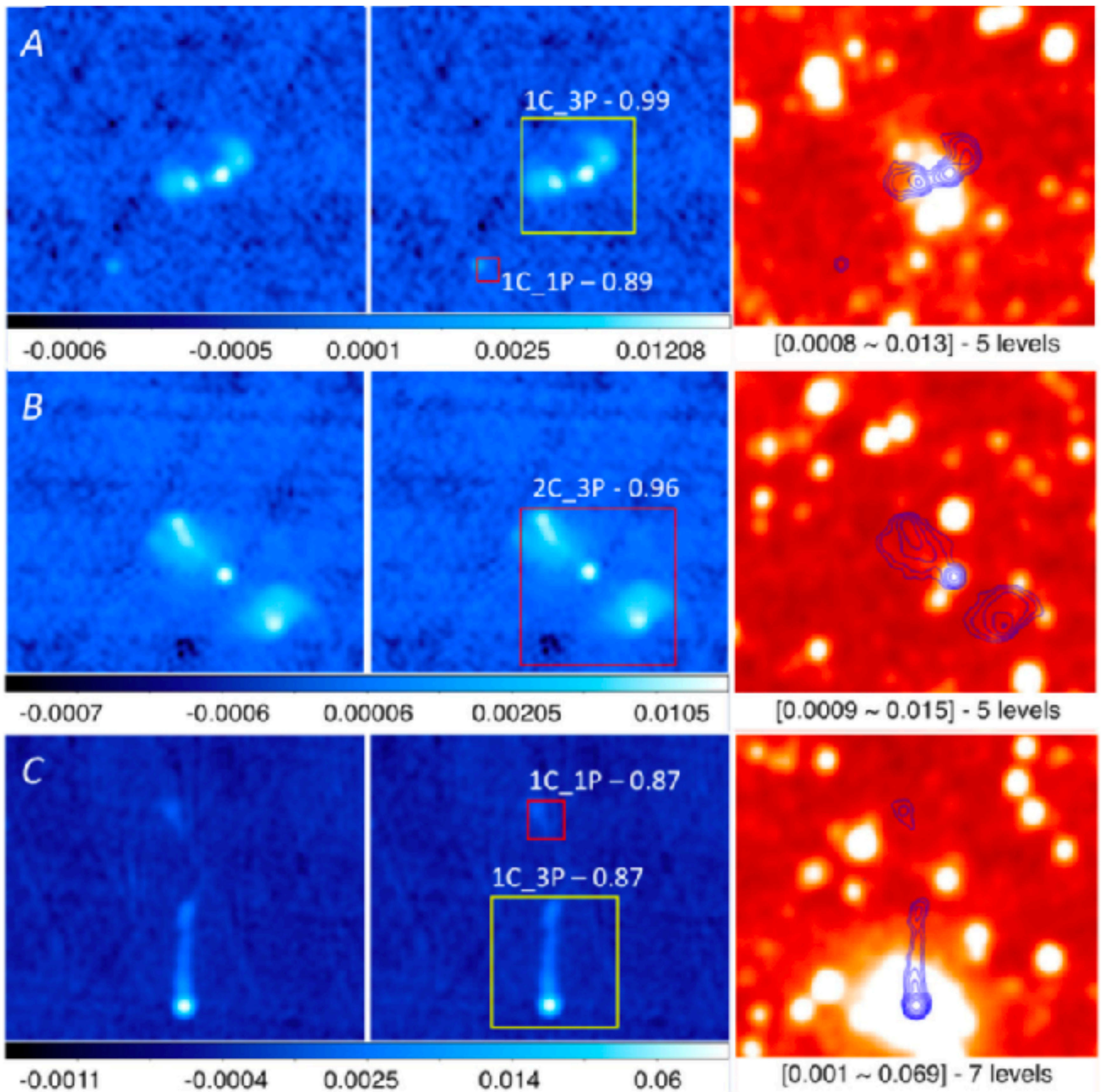
Wu et al. MNRAS **482,** 1211–1230 (2019)



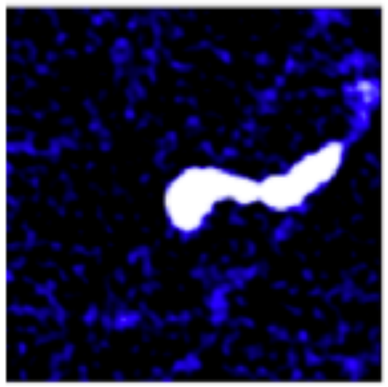
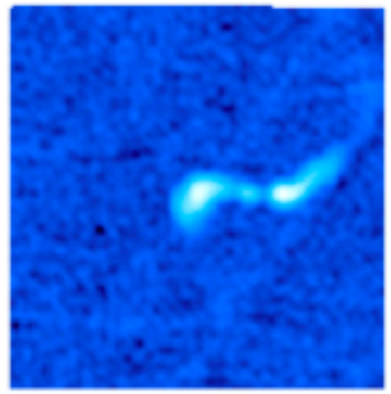
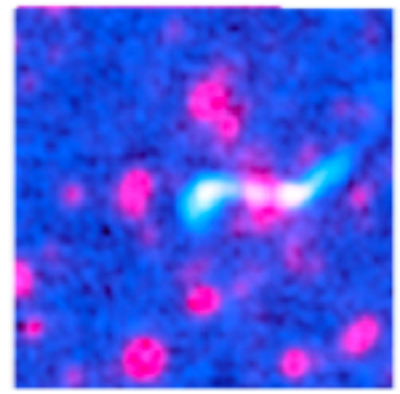
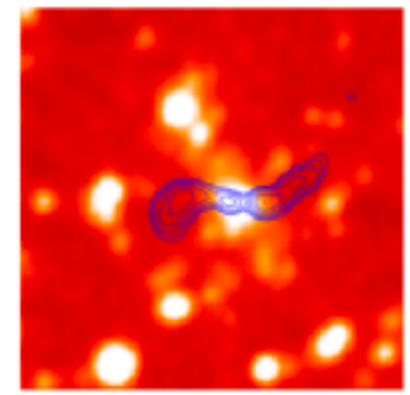
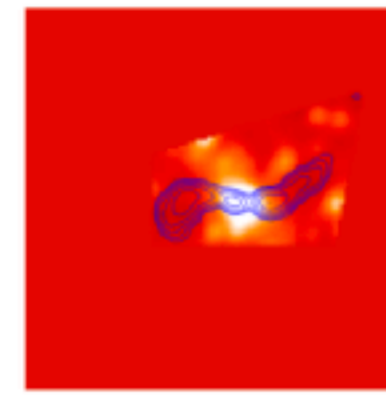
Wu et al. MNRAS **482,** 1211–1230 (2019)



Wu et al. MNRAS **482,** 1211–1230 (2019)



Wu et al. MNRAS **482,** 1211–1230 (2019)

Methods	F	$D1$	$D2$	$D3$	$D4$
					
1C_1P	0.8087	0.8580	0.8242	0.8485	**0.8784**
1C_2P	0.6376	0.6882	0.6843	0.6746	**0.7074**
1C_3P	0.8250	0.8816	0.8561	0.8876	**0.8941**
2C_2P	0.7474	0.7014	0.7231	0.7983	**0.8200**
2C_3P	**0.8087**	0.7099	0.6989	0.8047	0.7916
3C_3P	0.7708	0.8636	0.8561	**0.9424**	0.9269
mean AP	78.5%	78.4%	77.4%	82.6%	**83.6%**

LECTURES & TUTORIALS

https://github.com/as595/NITheP