Data science for direct imaging of exoplanets

Machine learning applied to astronomical high-contrast imaging

Carlos Alberto ***Gomez Gonzalez***

EWASS S6, 05/04/2018

UNIVERSITÉ Grenoble Alpes

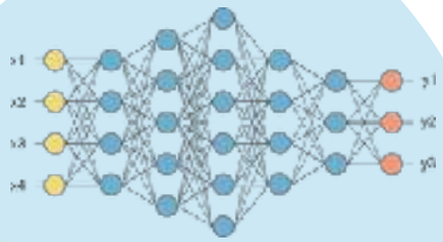
IPAG
Institut de Planétologie et d'Astrophysique de Grenoble

Data Institute
Univ. Grenoble Alpes

Computer science

Machine learning & Stats

Domain knowledge



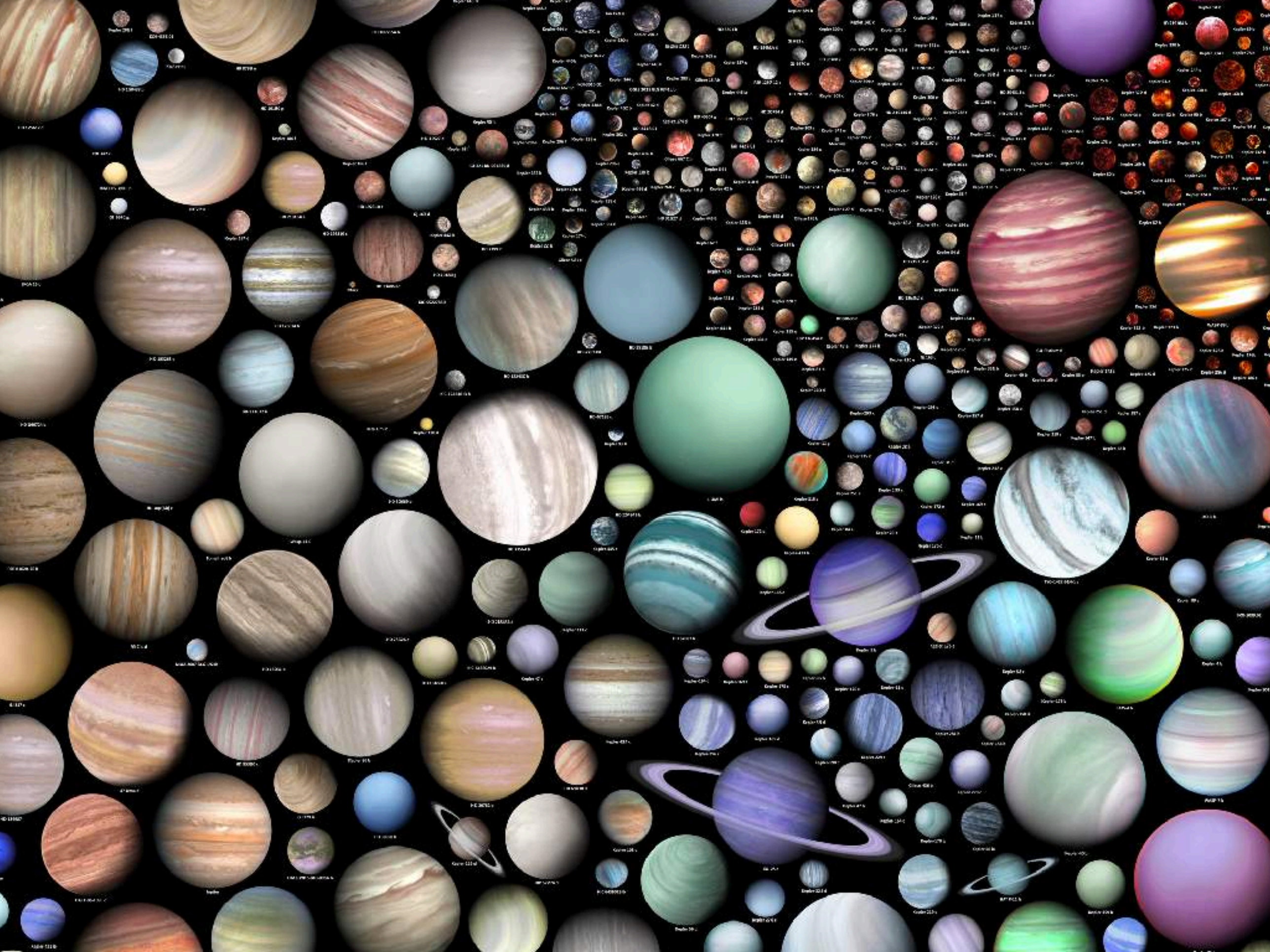
Communication and team skills

(Academic) Data Science

Academic data science

- Interdisciplinary research (natural science with CS, ML, AI fields).
- Integrating cutting-edge AI developments.
- Ensuring the use of robust statistical approaches and well-suited metrics.
- Code + supporting data releasing.
- Data challenges (benchmark datasets).
- Open peer-review.

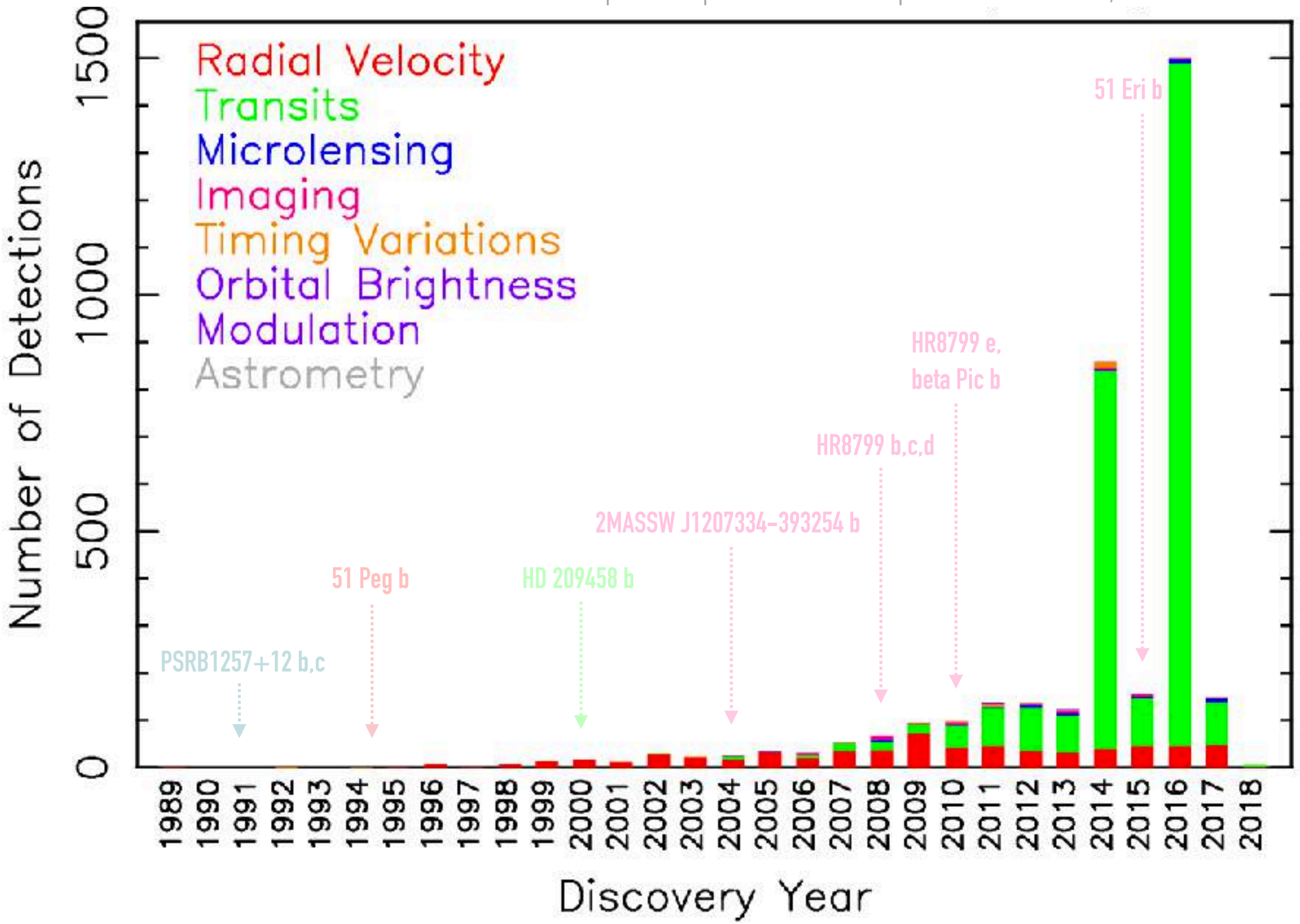
http://jakevdp.github.io/blog/2014/08/22/hacking-academia/



Mostly, we rely on indirect methods for detecting exoplanets

… it's very hard to "see" them

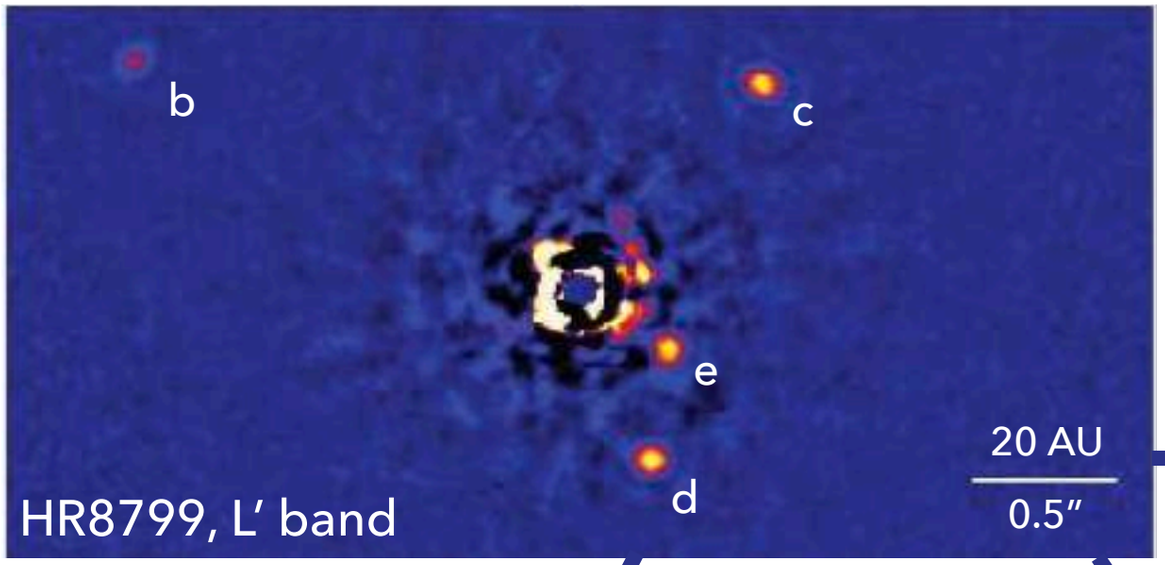
http://exoplanetarchive.ipac.caltech.edu, 25 Jan 2018



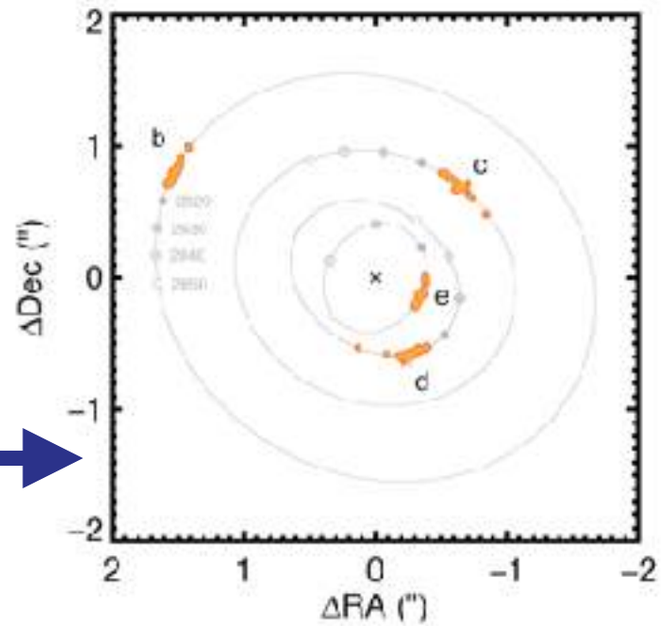
Credit: NASA, http://planetquest.jpl.nasa.gov

Power of direct imaging

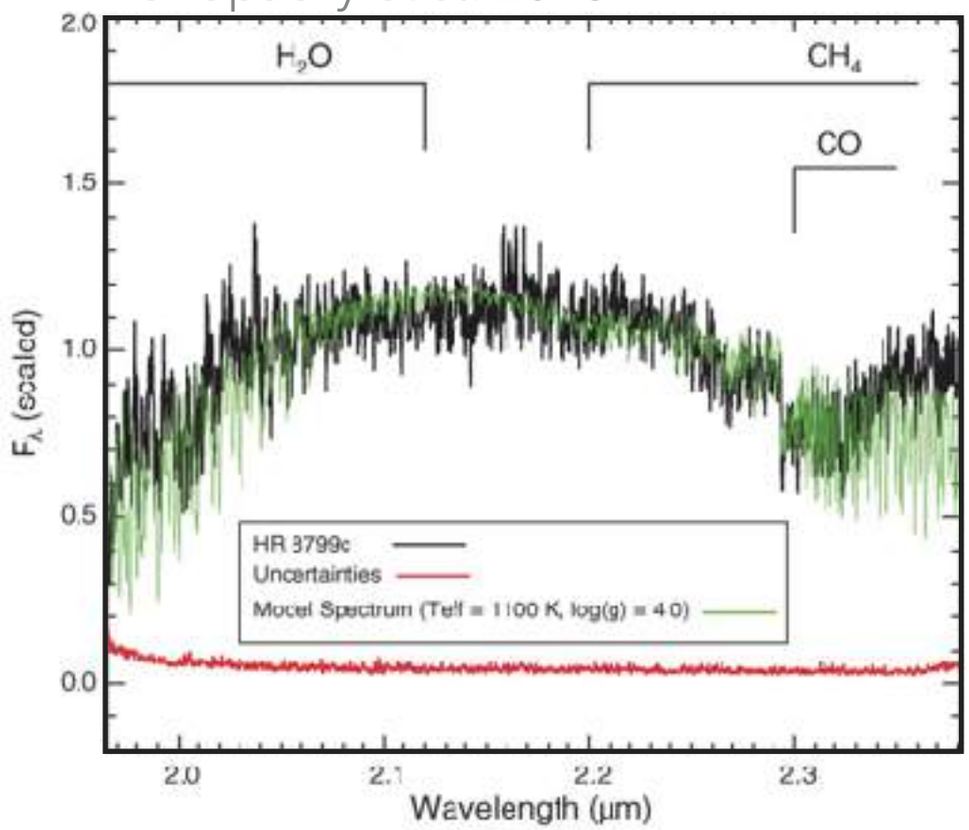
Marois et al. 2010



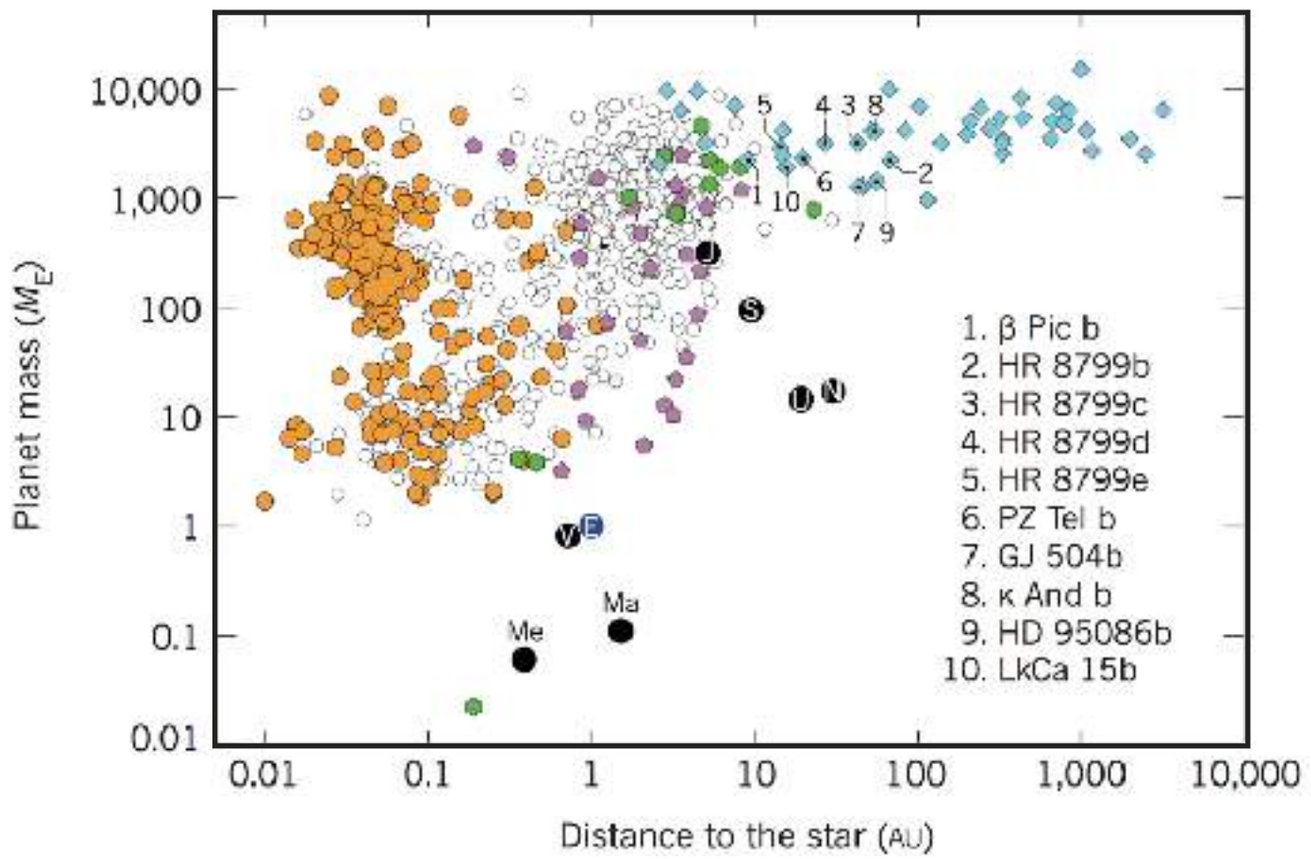
Bowler 2016



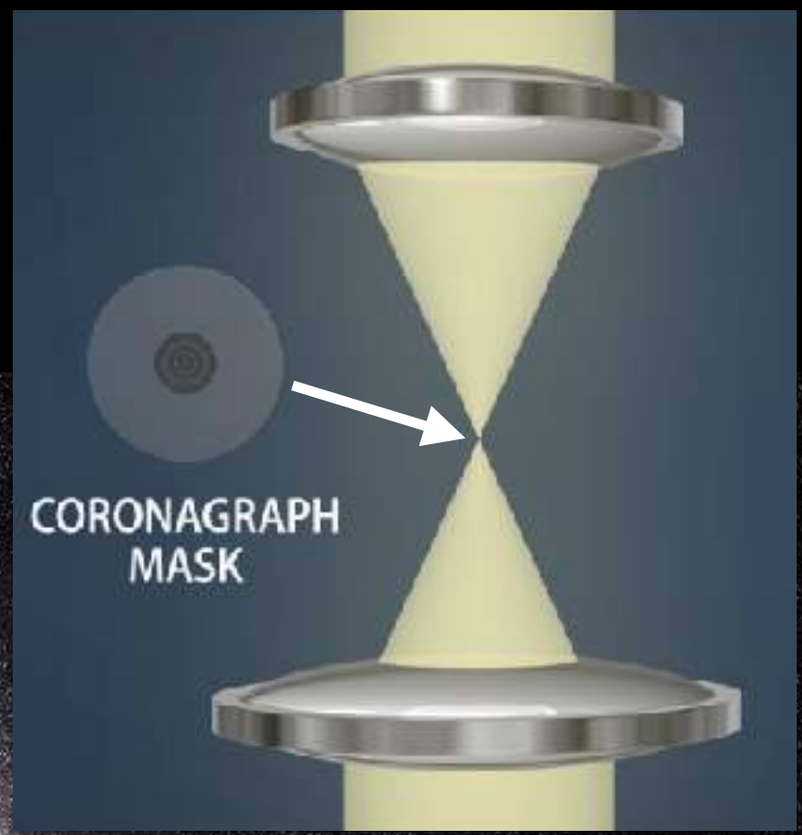
Konopacky et al. 2013



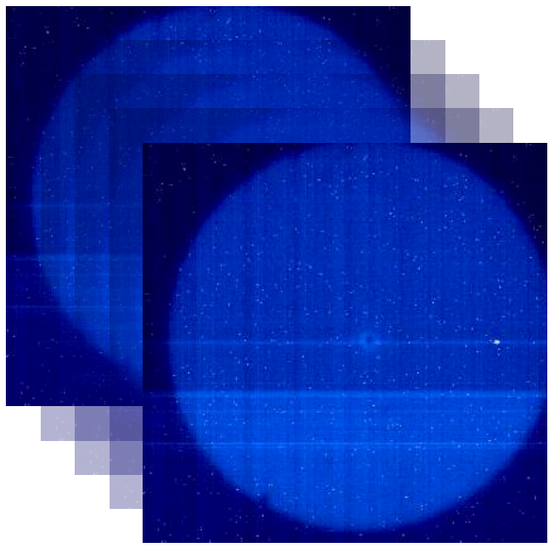
Milli et al. 2016

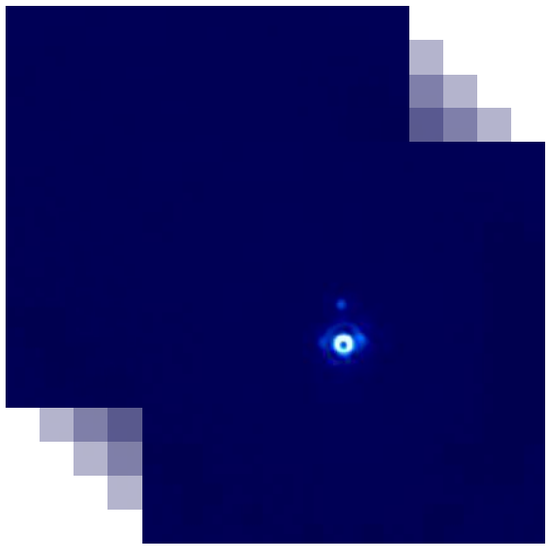


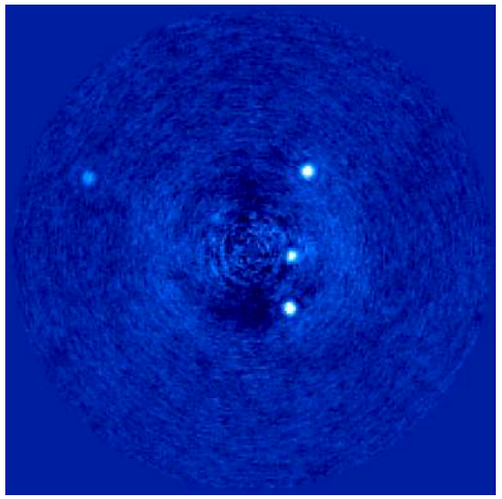
SPHERE, Vigan et al. 2015



Very Large Telescope (VLT), Chile







Raw astronomical images

Basic calibration and "cosmetics"

- Dark/bias subtraction
- Flat fielding
- Sky or thermal background subtraction
- Bad pixel correction

Bad frames removal

Image recentering

Sequence of calibrated images

PSF modeling

- Median
- Pairwise, ANDROMEDA
- LOCI
- PCA, NMF
- LLSG

Model PSF subtraction

De-rotation (ADI) or rescaling (mSDI)

Image combination

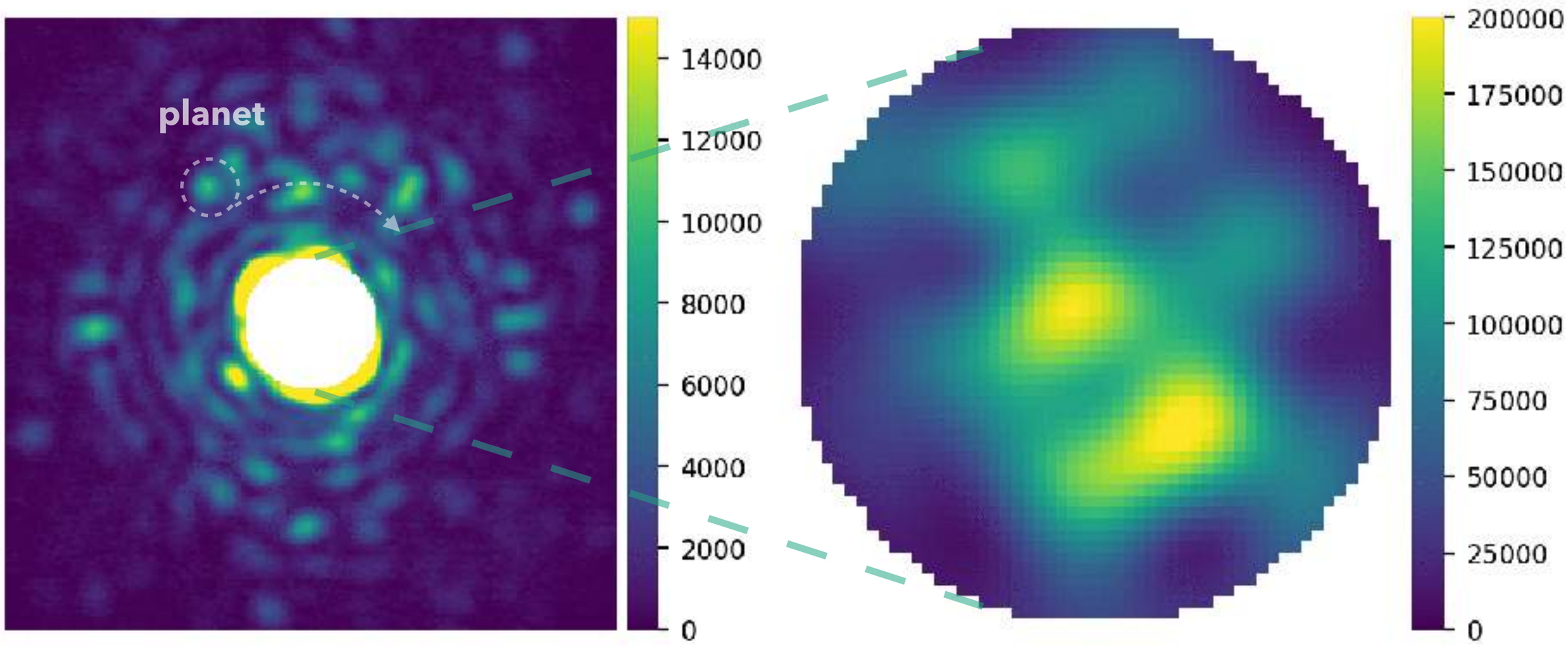
Detection on final residual image

Characterization of detected companions

calib. images

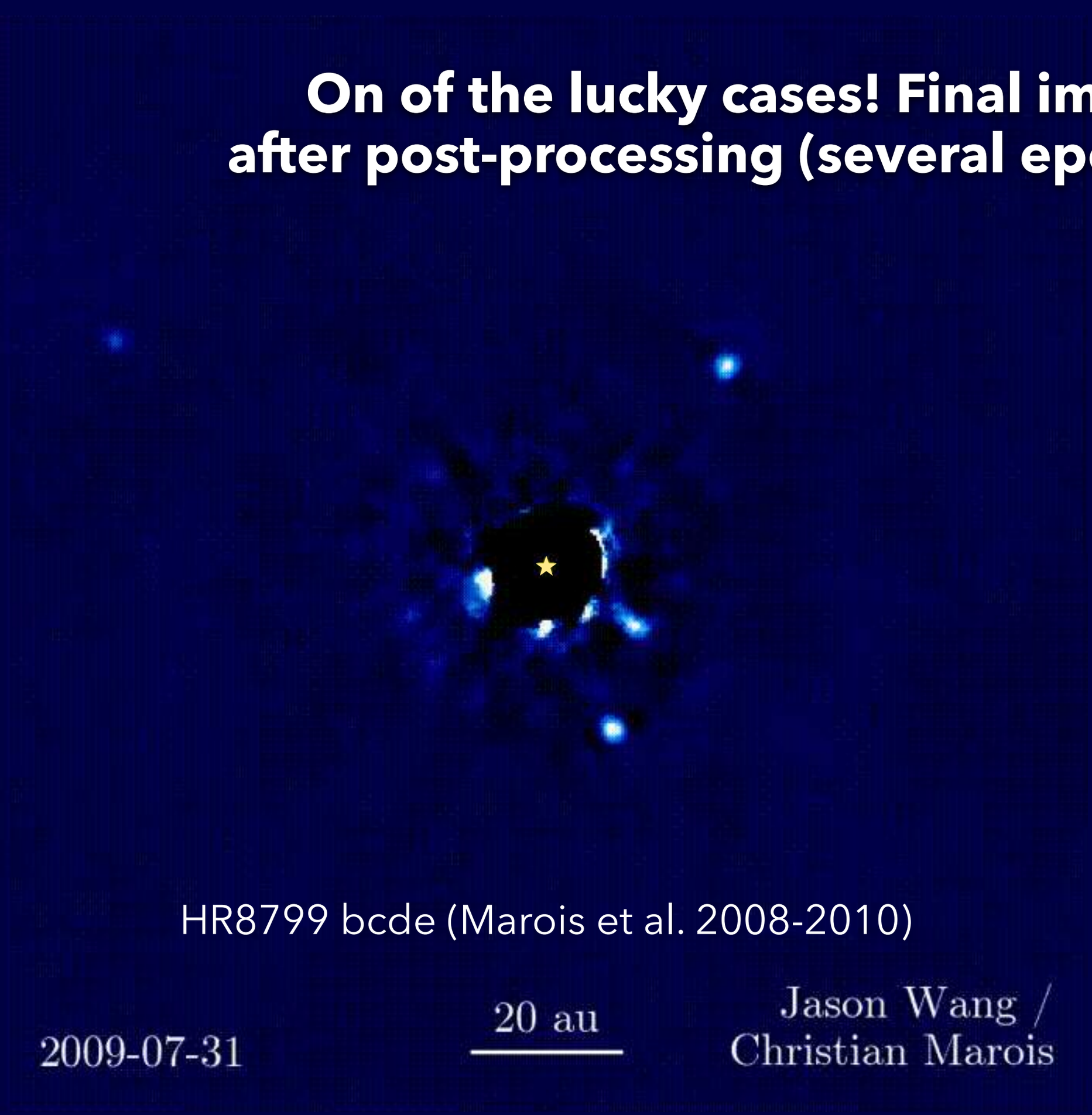
Speckle noise

Angular differential imaging - bright synthetic planet



post-proc.

On of the lucky cases! Final images after post-processing (several epochs)

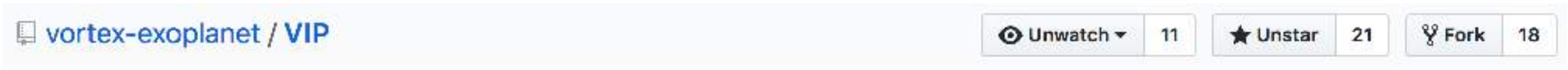


HR8799 bcde (Marois et al. 2008-2010)

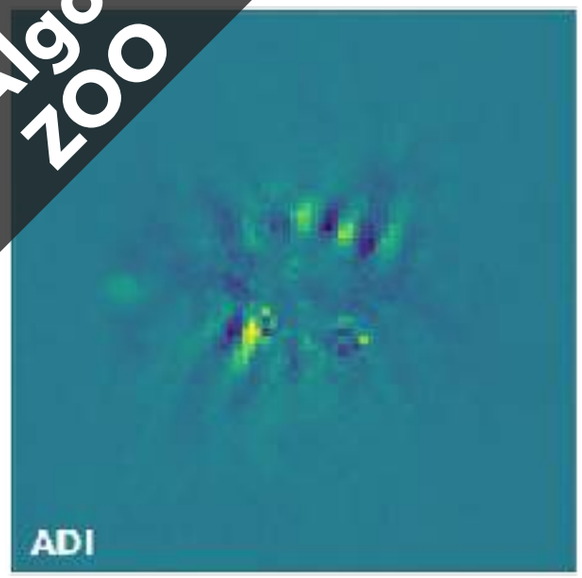
Vortex Image Processing library

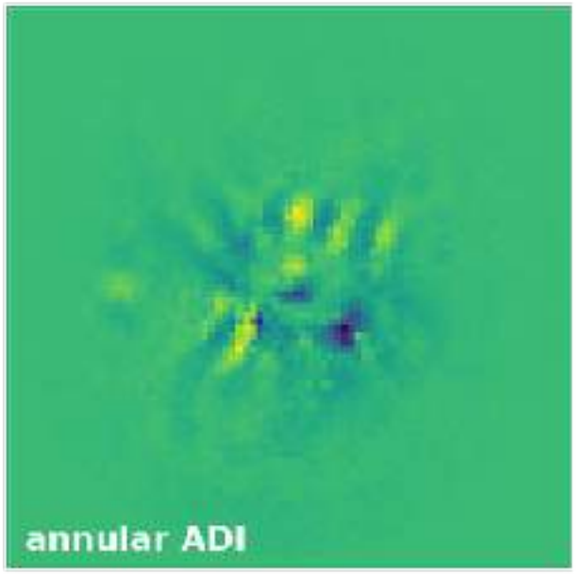
Gomez Gonzalez et al. 2017

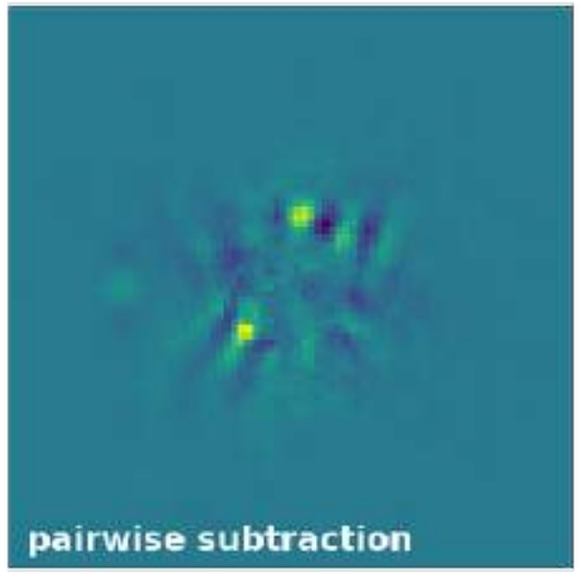
- Available on Pypi
- Documentation (Sphynx): http://vip.readthedocs.io/
- Jupyter tutorial
- https://github.com/vortex-exoplanet/VIP
- Python 2/3 compatibility
- Continuous integration (Travis CI) and automated testing (Pytest)

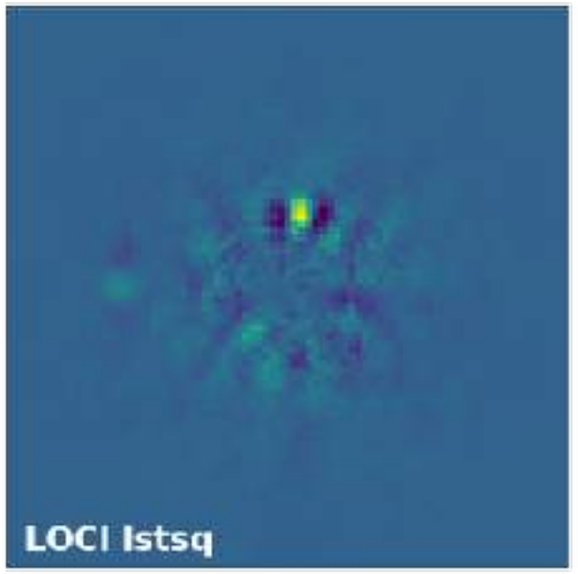


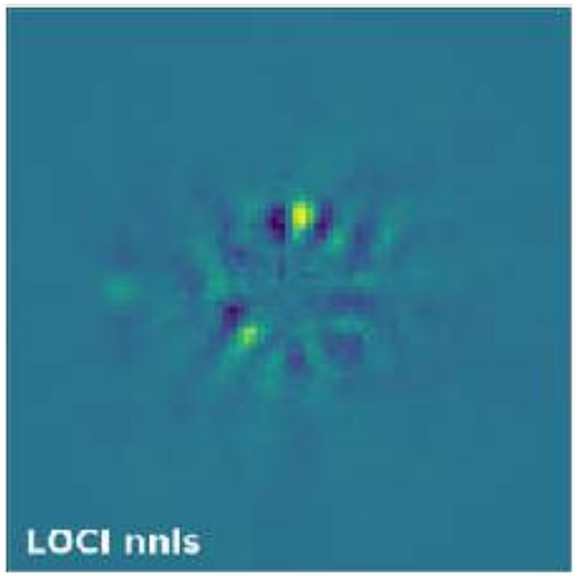
Algo-ZOO

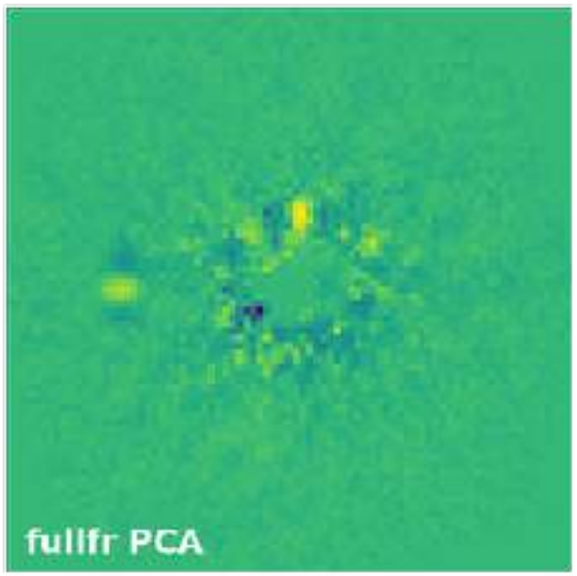

ADI

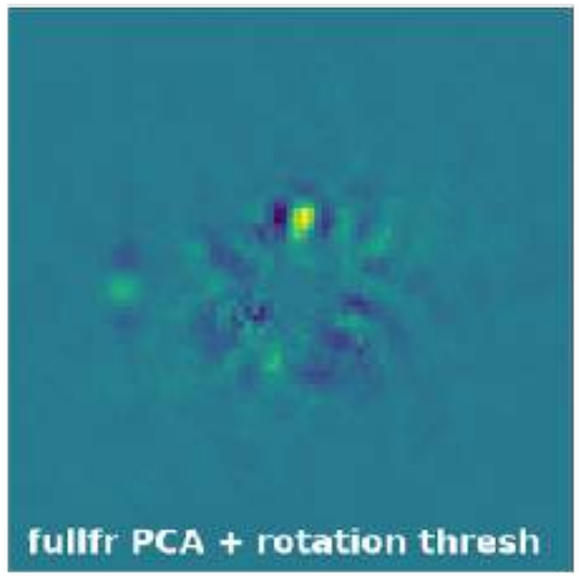

annular ADI

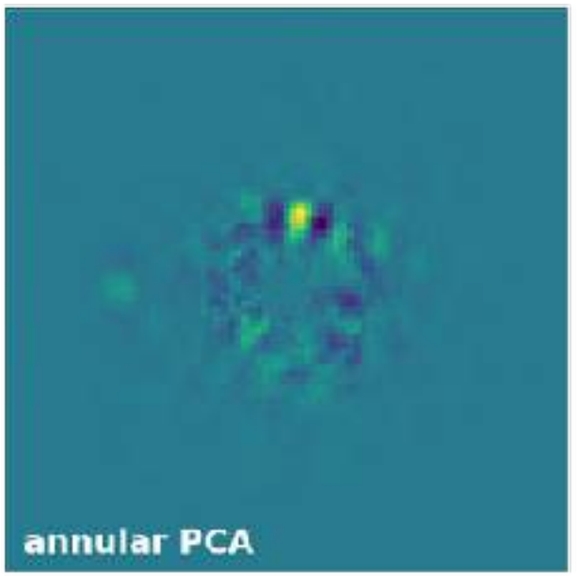

pairwise subtraction

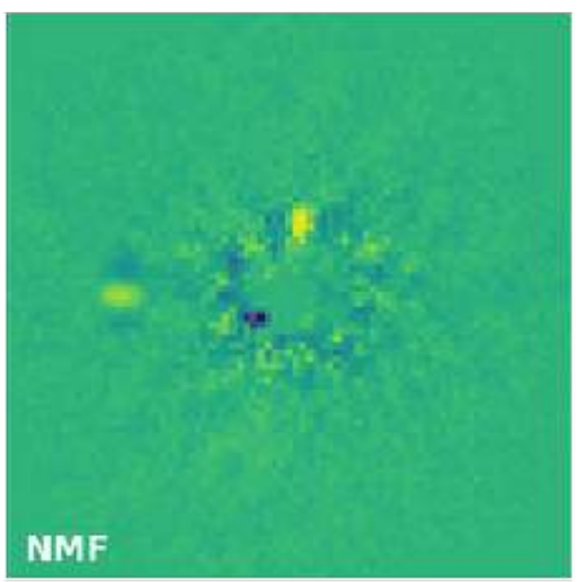

LOCI lstsq


LOCI nnls

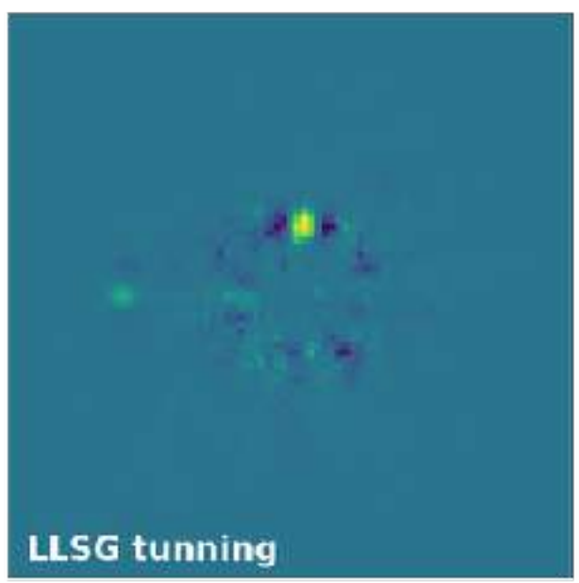

fullfr PCA

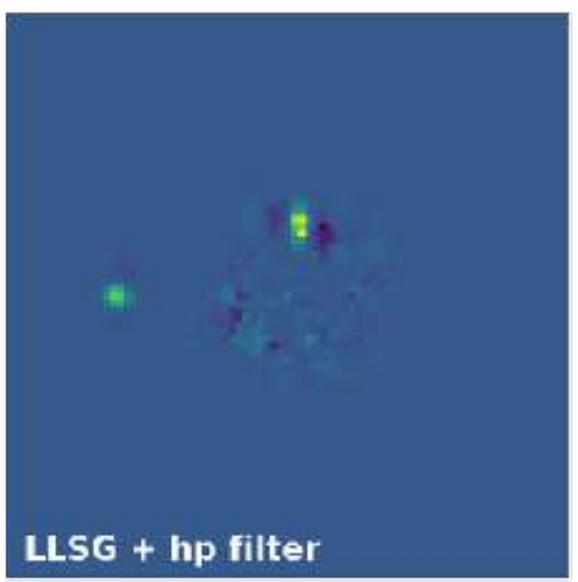

fullfr PCA + rotation thresh


annular PCA


NMF


Gomez Gonzalez et al. 2016
LLSG


LLSG tunning


LLSG + hp filter

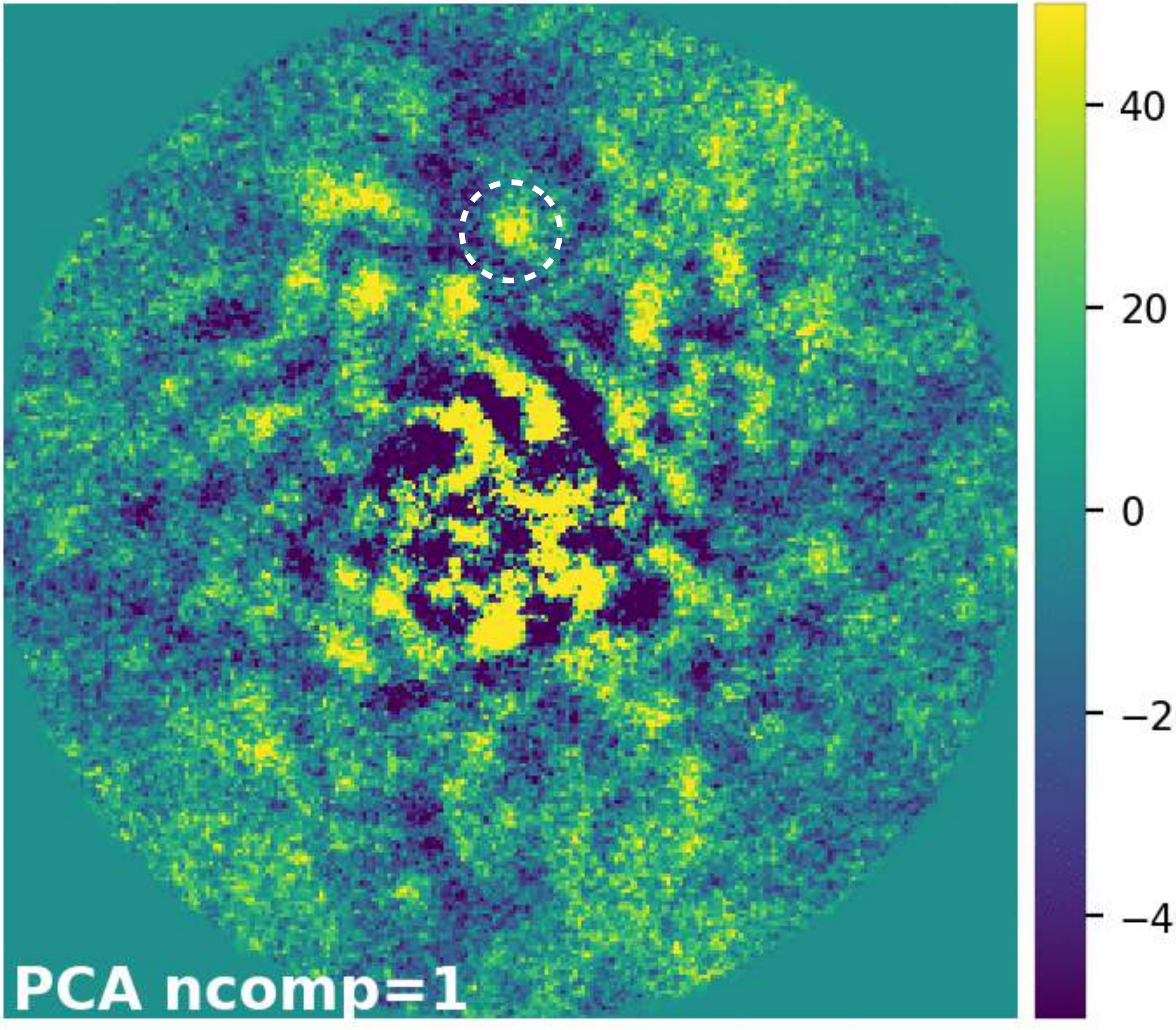
Open source

Open Science & reproducibility

Image sequence

Final residual image

Detection


40
20
0
−20
−40
PCA ncomp=1

“Essentially, all models are wrong, but some are useful.”

George Box

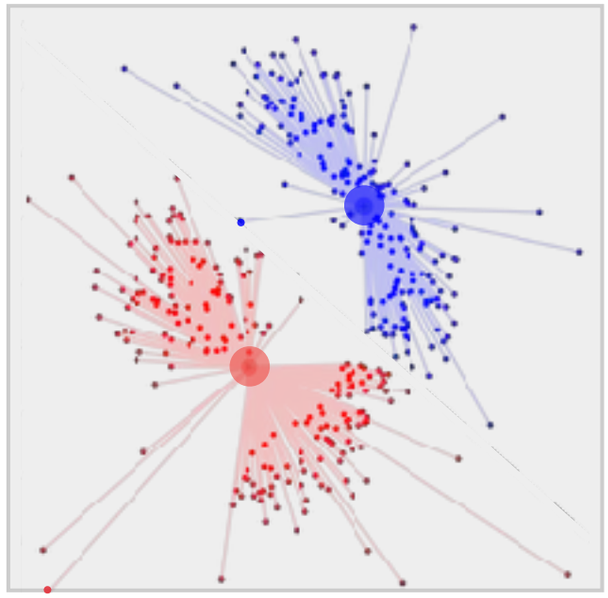
“…if the model is going to be wrong anyway, why not see if you can get the computer to ‘quickly’ learn a model from the data, rather than have a human laboriously derive a model from a lot of thought.”

Peter Norvig

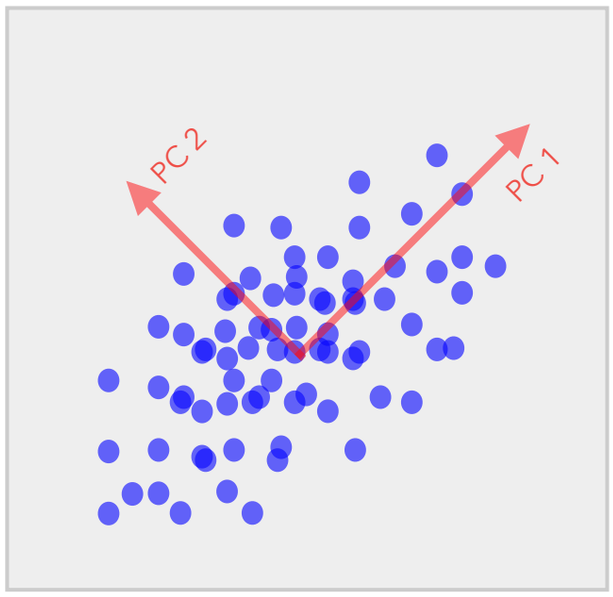
Textbook Machine Learning

Unsupervised

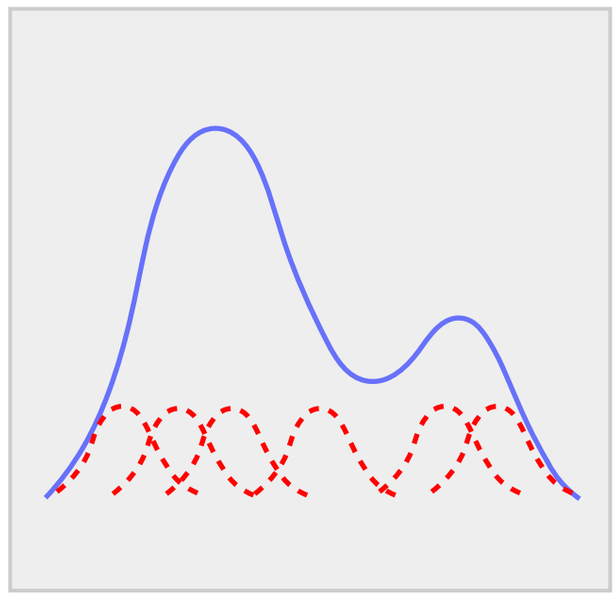
Clustering



Dimensionality reduction

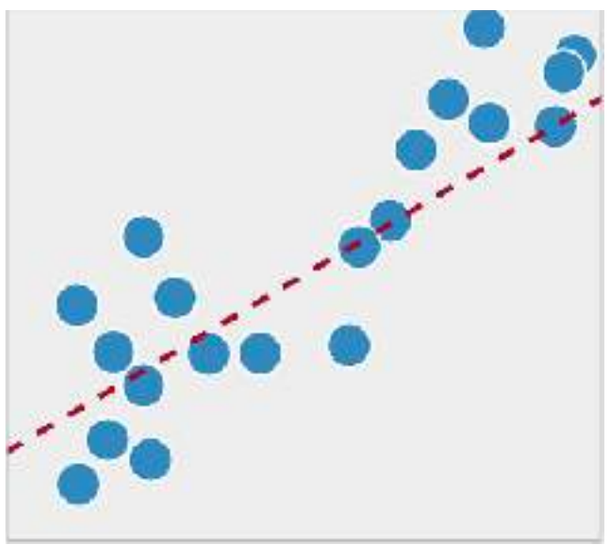


Density estimation

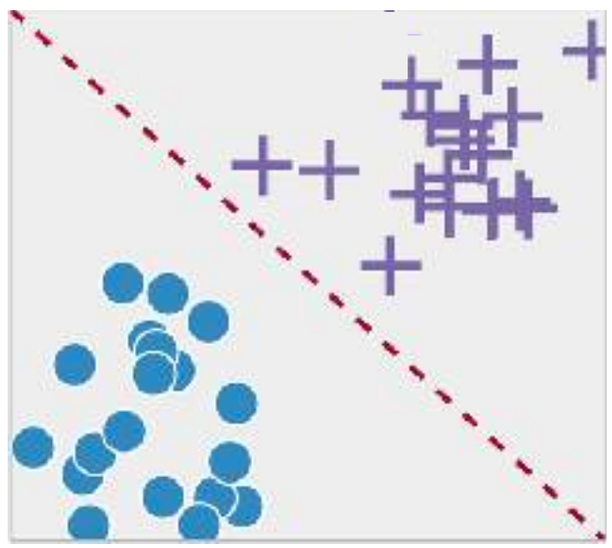


Supervised

Regression



Classification



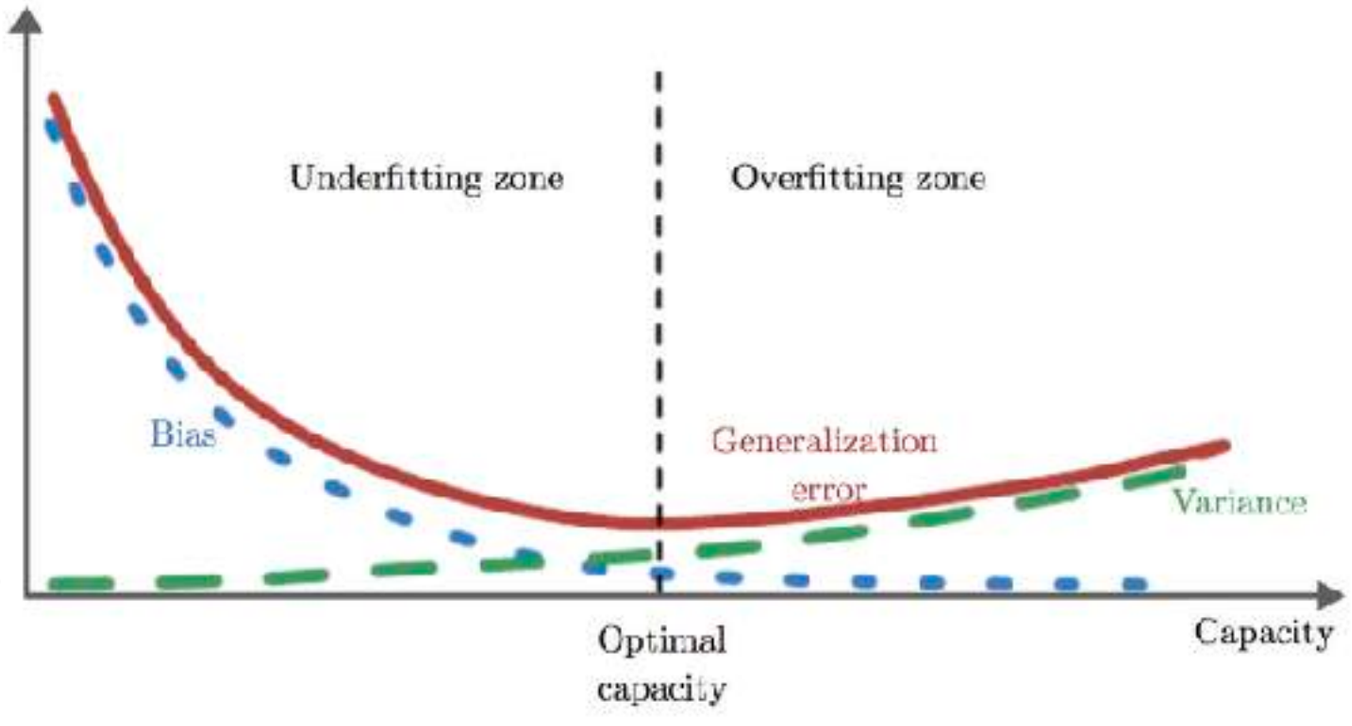
and reinforcement learning

Supervised learning

The goal is to learn a mapping from the input samples to the labels: $f : \mathcal{X} \to \mathcal{Y},$

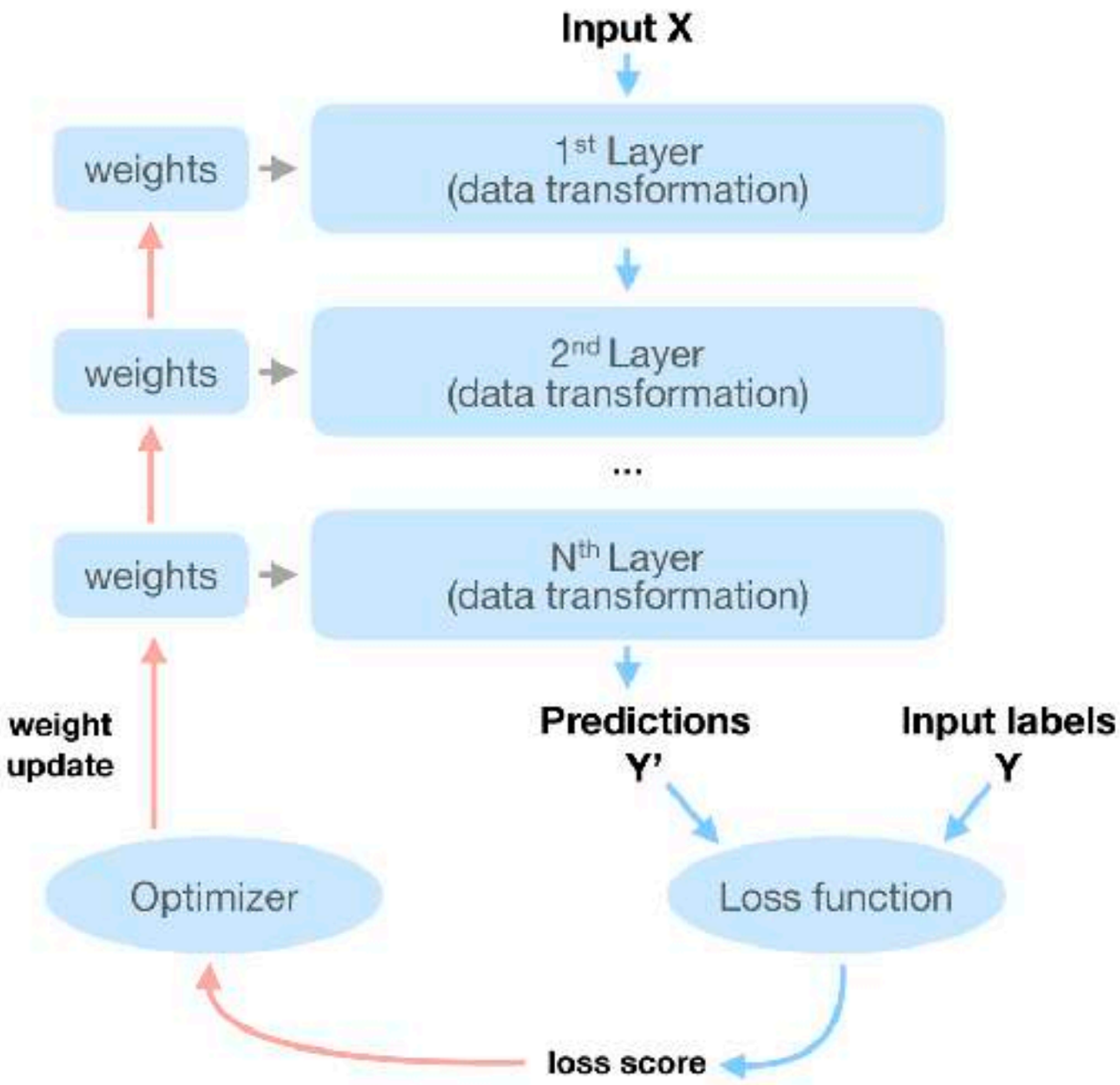
given a labeled dataset $(x_i, y_i)_{i=1,\ldots,n}$:

$$\min_{f \in \mathcal{F}} \frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(y_i, f(x_i)) + \lambda \Omega(f)$$



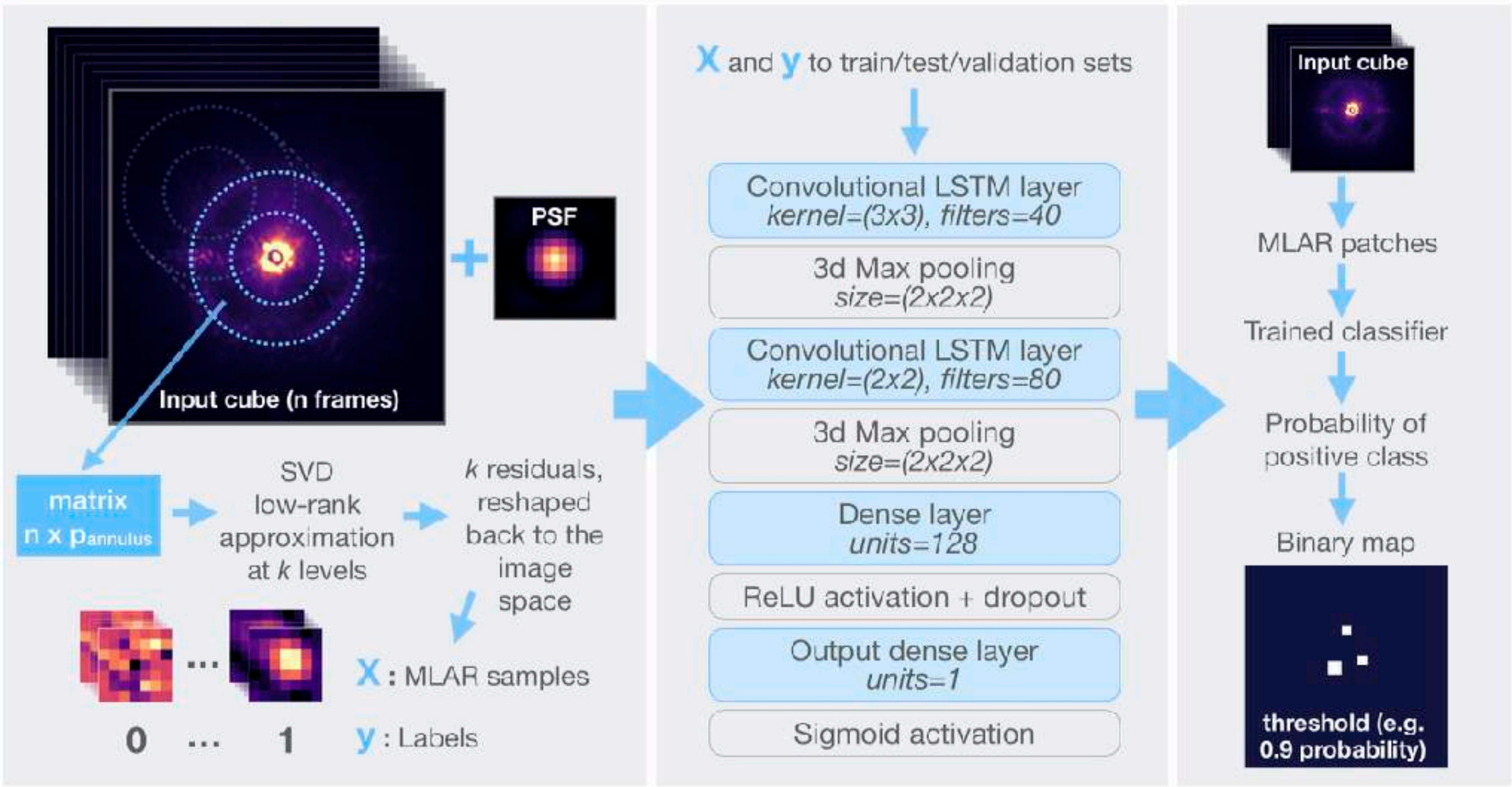
Goodfellow et al. 2016

Deep neural networks



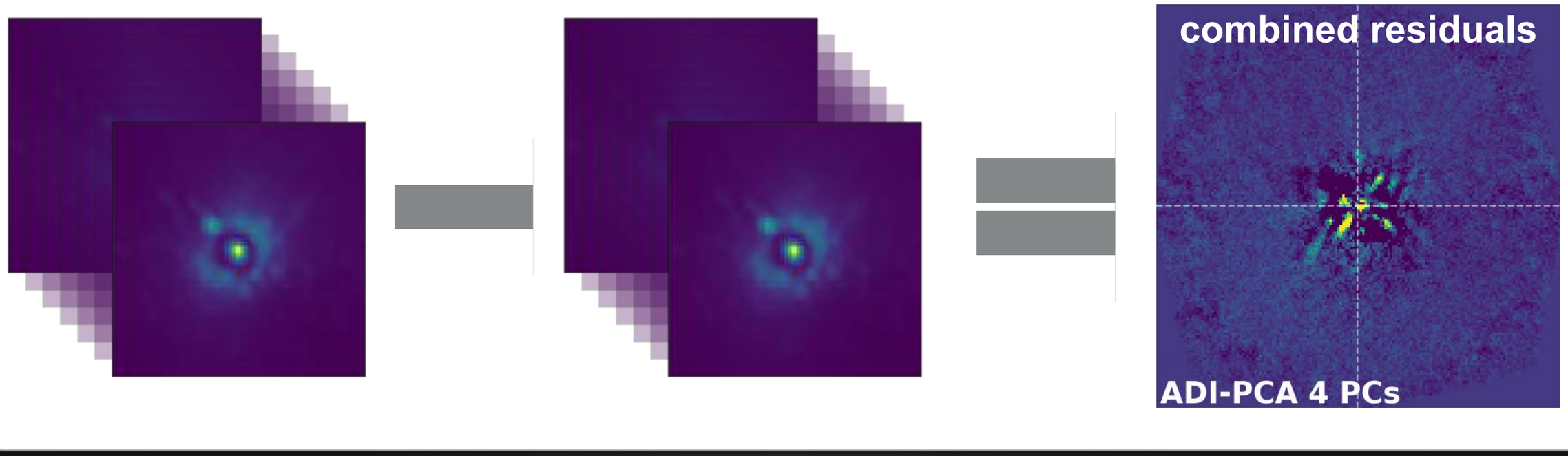
$$f(x) = \sigma_k(A_k\sigma_{k-1}(A_{k-1}...\sigma_2(A_2\sigma_1(A_1x))...))$$

Supervised detection of exoplanets



Gomez Gonzalez et al. 2018

Model PSF subtraction



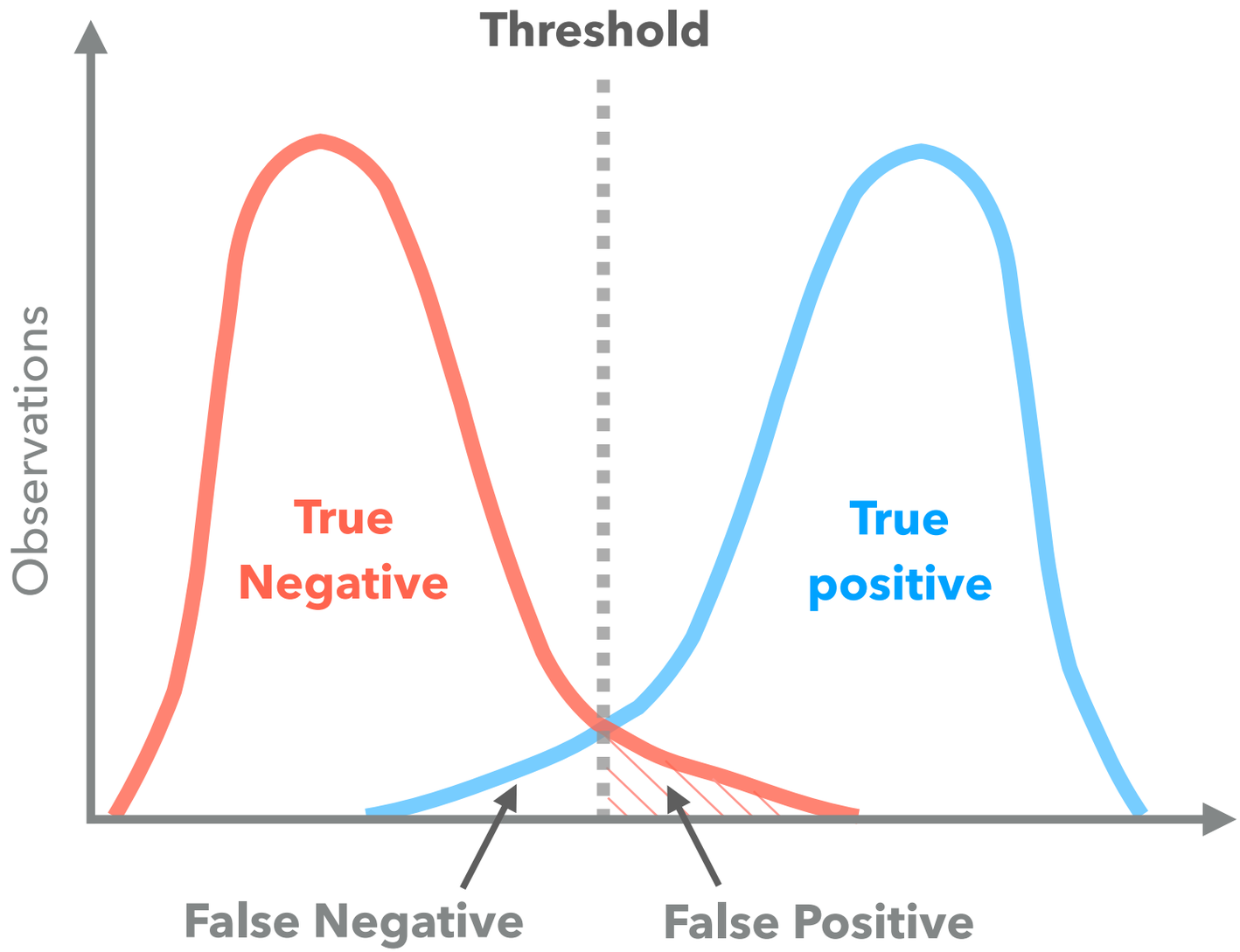
Supervised detection (SODINN)

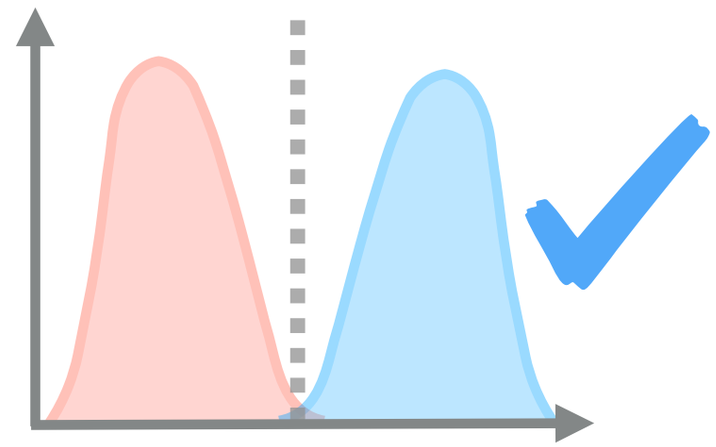
noisy and unlabelled images

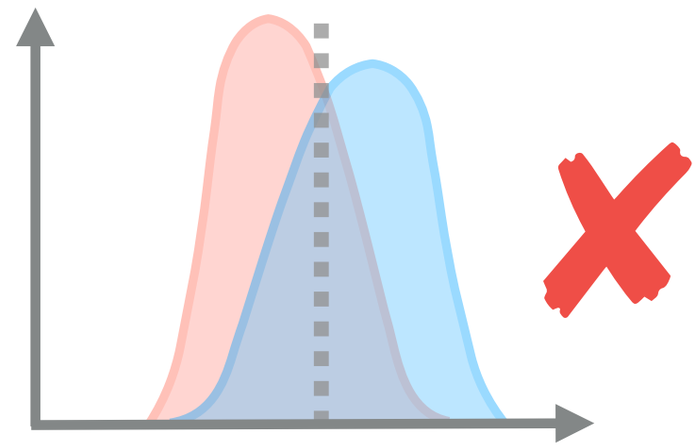
data transformation + adequate (ML) model

higher sensitivity

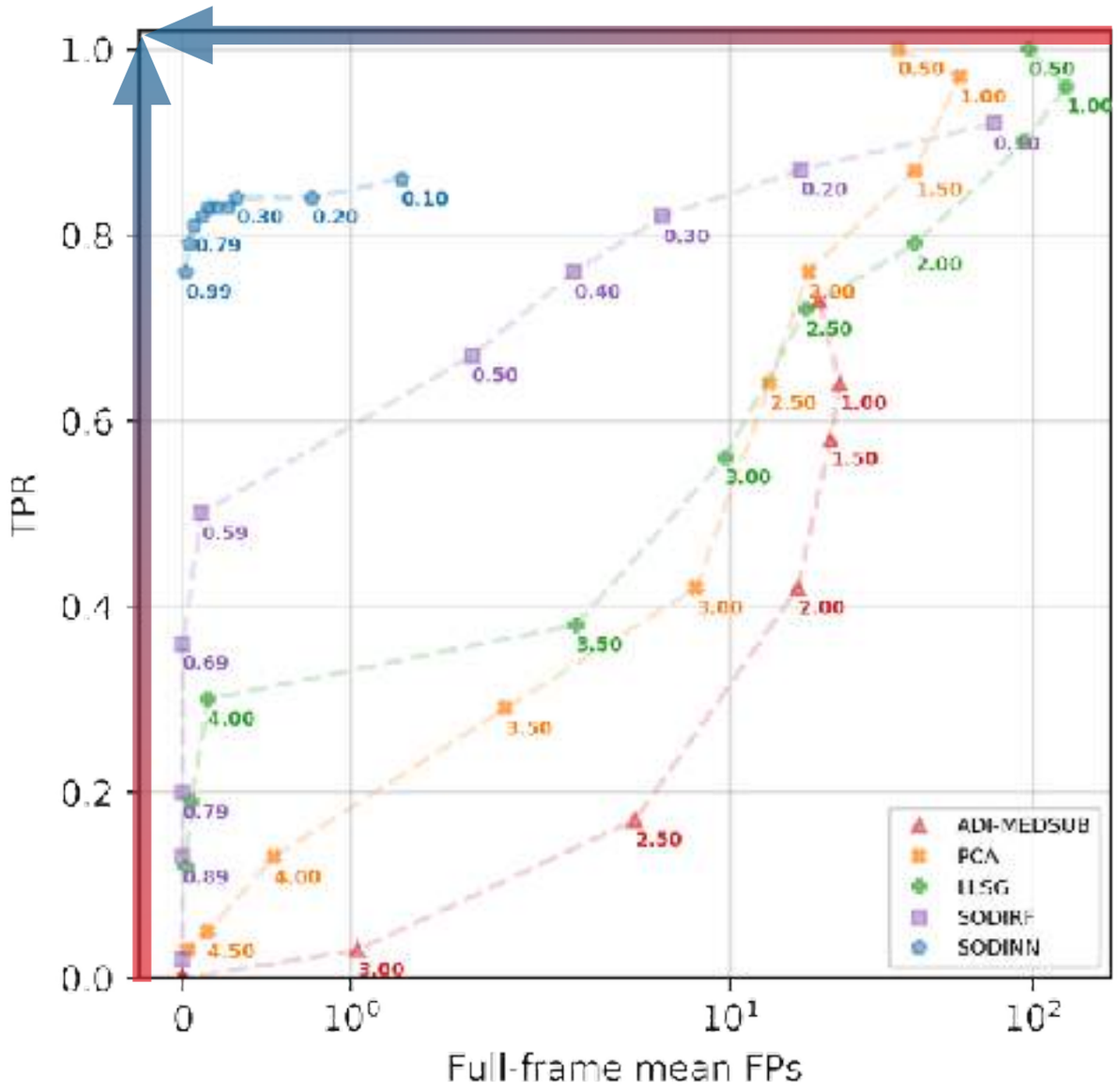
Performance assessment

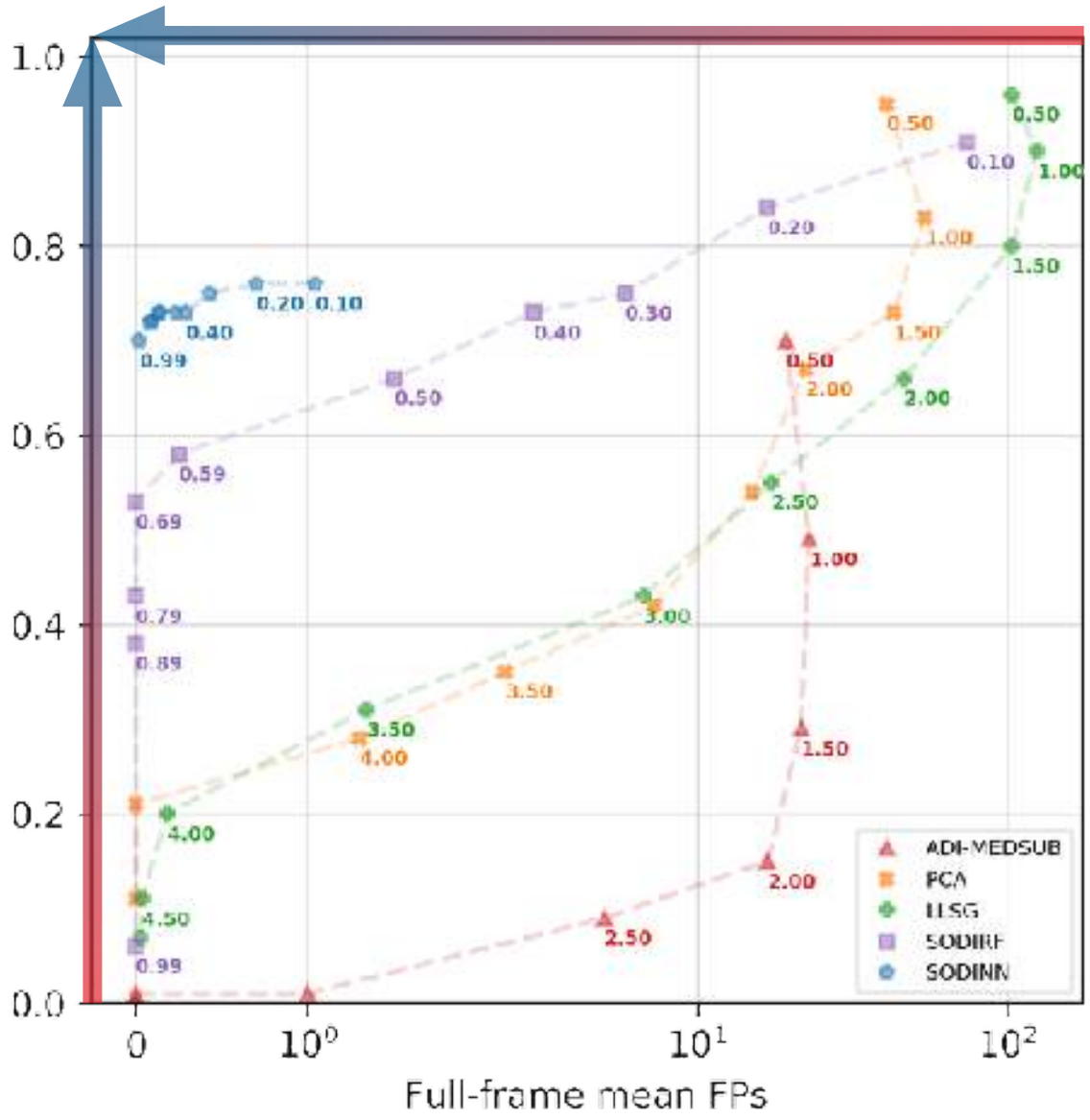






Data-driven performance assessment





Not reinventing the wheel!

VIP

TensorFlow

K

jupyter

NumPy

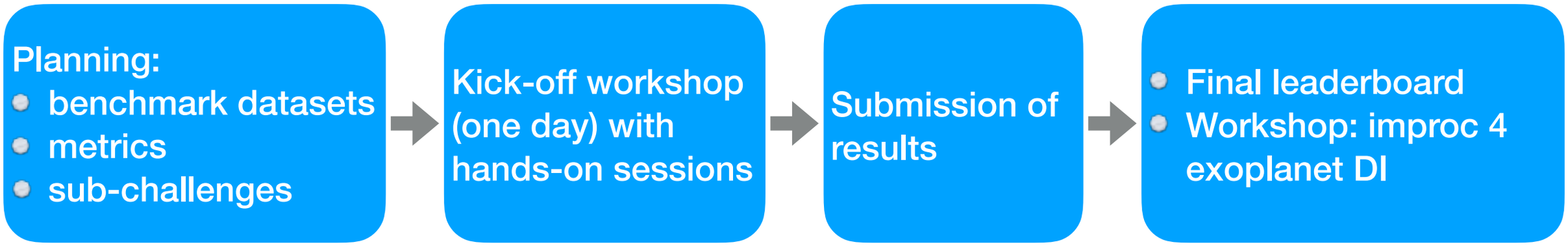
scikit learn

matplotlib

Other ML applications in HCI

- Characterization of companions.
- Spectral retrieval for multi-channel spectral differential imaging.
- Smart use of instrumental telemetry for reference-star differential imaging (survey).
- Common-path wavefront sensing and AO wavefront control.
- Multi-epoch aware model taking into account Keplerian motion.

Exoplanet DI challenge



- International collaboration.
- Whole HCI community (open to DS and ML researchers).
- Using existing platforms/frameworks (Kaggle, RAMP, codalab, crowAI, evalAI).
- Hands-on sessions.
- Workshop on image processing for direct detection of exoplanets.

¡Gracias!

carlos.gomez@univ-grenoble-alpes.fr

carlgogo

carlosalbertogomezgonzalez

https://carlgogo.github.io/