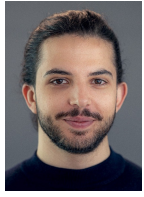
(Deep) Machine Learning for Exoplanet Detection in Direct Imaging at High Contrast

Julien Mairal

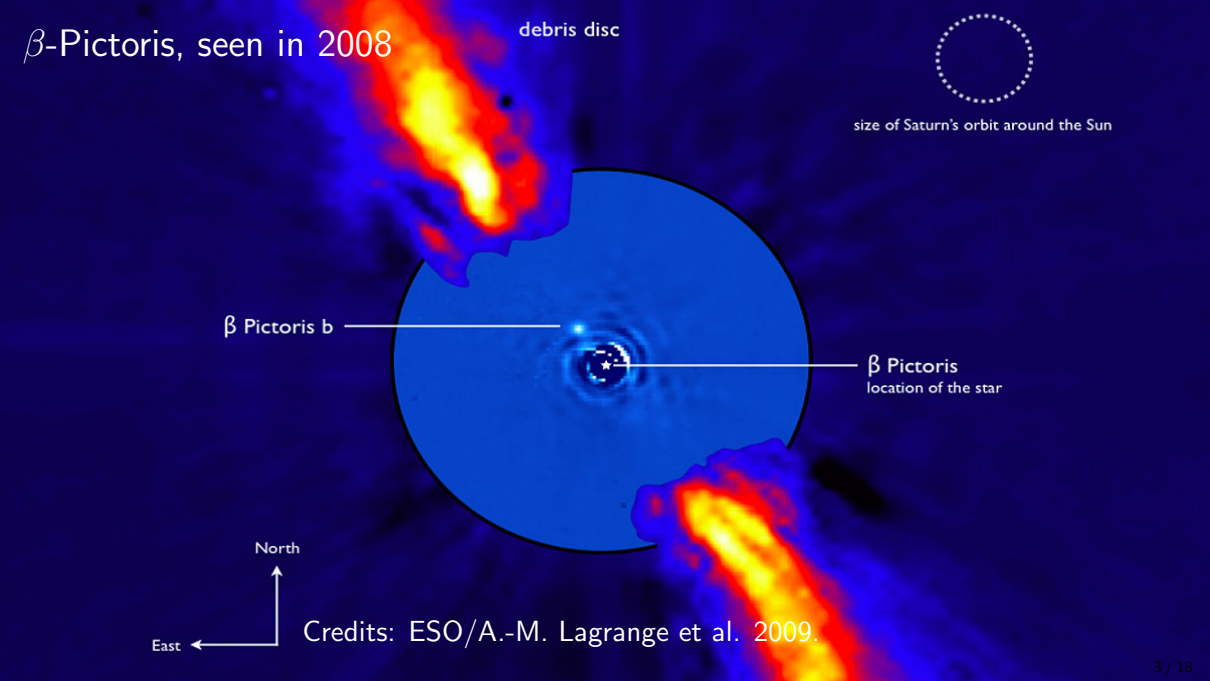
Univ. Grenoble-Alpes, Inria

Collaborators and Publications

- O. Flasseur, T. Bodrito, J. Mairal, J. Ponce, M. Langlois and A.-M. Lagrange. Deep PACO: Combining Statistical Models with Deep Learning for Exoplanet Detection and Characterization in Direct Imaging at High Contrast. *Monthly Notices of the Royal Astronomical Society (MNRAS)*. 2024.
- T. Bodrito, O. Flasseur, J. Mairal, J. Ponce, M. Langlois and A.-M. Lagrange. MODEL&CO: Exoplanet detection in angular differential imaging by learning across multiple observations. *Monthly Notices of the Royal Astronomical Society (MNRAS)*. 2024.


Some figures courtesy of Théo Bodrito and Olivier Flasseur. Collab. PEPR Origins.


β-Pictoris, seen in 2008
debris disc
size of Saturn's orbit around the Sun
β Pictoris b
β Pictoris
location of the star
North
East
Credits: ESO/A.-M. Lagrange et al. 2009.

17 years of observations using VLT/NACO and VLT/SPHERE

Video Credit: Jason Wang and Malachi Noel

A coronograph blocks light emitted by the star.

From visible light to near-infrared: Contrast improves from 10^9 to 10^6
Coronograph: from 10^6 to 10^4
Adaptive optics: from 10^4 to 10^3

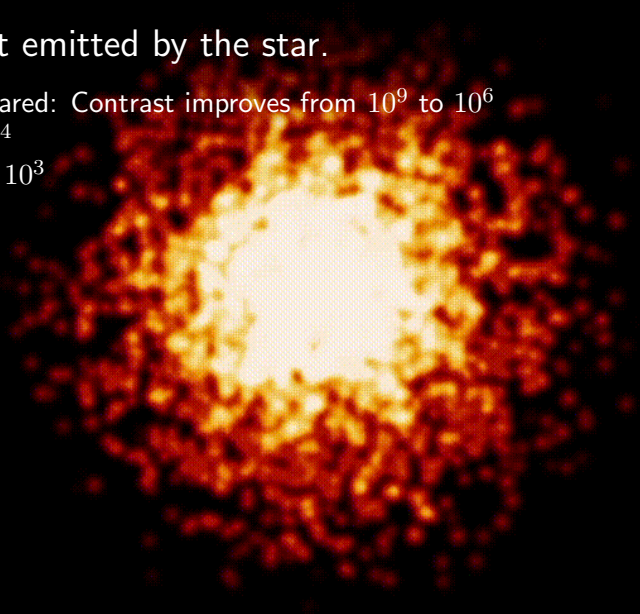
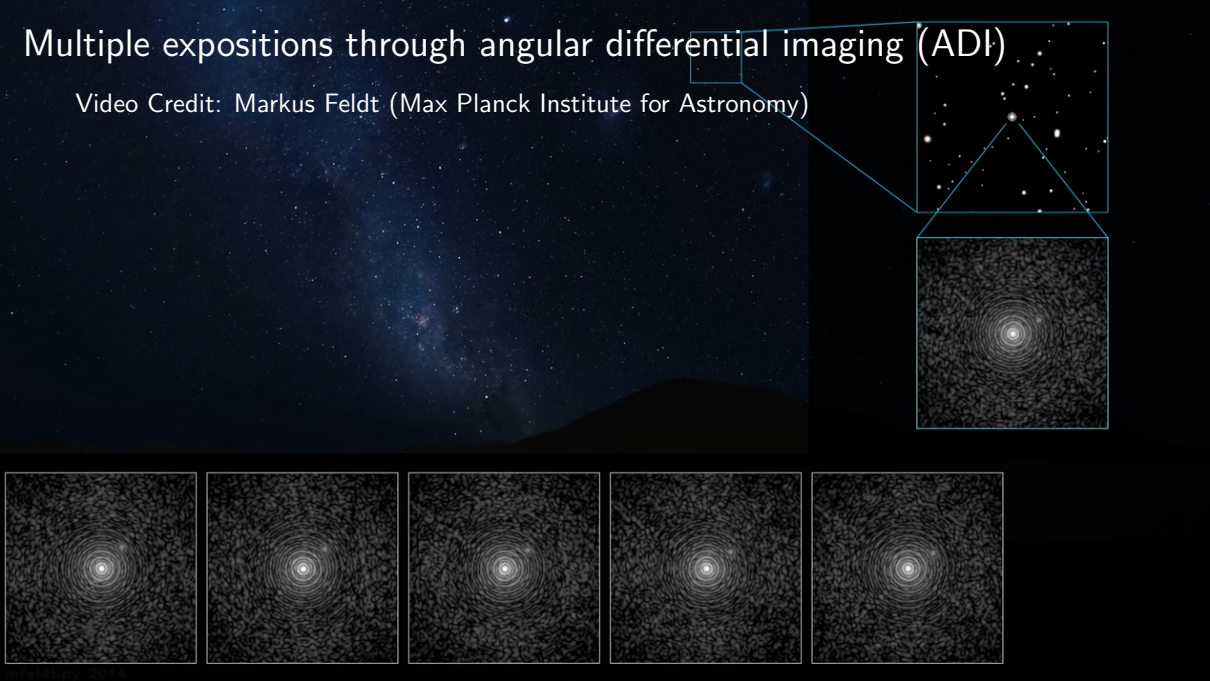


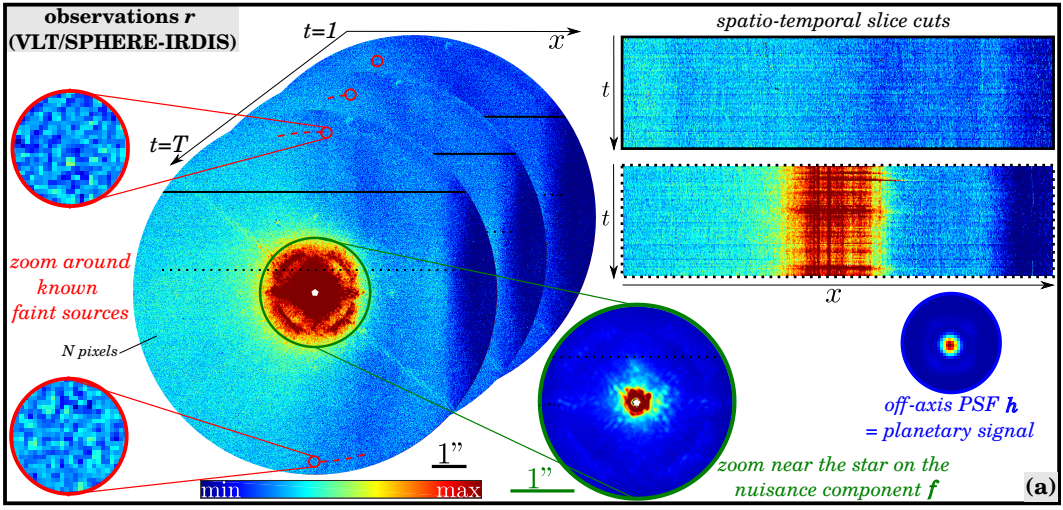
Image Credit: Nasa

Multiple expositions through angular differential imaging (ADI)

Video Credit: Markus Feldt (Max Planck Institute for Astronomy)



Finally: The Data



Speckles are temporally quasi-static but spatially non-stationary.

Challenge for ML: Learning without ground truth data.

Challenge for ML: Learning without ground truth data.

- Learning with **semi-synthetic** data
- Combining an observation-dependent **statistical model** and **deep learning**.

The PACO model [Flasseur et al., 2020]

Observation at time t:

$$I_t = \underbrace{N_t}_{\text{speckle}} + \sum_{i=1}^{k} \underbrace{\alpha_i}_{\text{contrast}} \Big(\underbrace{P}_{\text{off-axis PSF}} \star \underbrace{R_t}_{\substack{\text{rotation} \\ \text{from } t_0 \text{ to } t}} \Big(\underbrace{p_{i,t_0}}_{\substack{\text{position of source } i \\ \text{at time } t_0}} \Big) \Big).$$

- We have a very good estimate of P.
- Rotation is known (due to Earth's rotation).
- Speckle is temporally quasi-static with local spatial correlations: **Patch-based model of speckles as multivariate Gaussian.**
- Estimation of α by **MLE** and detection by **likelihood ratio test**.

The PACO model [Flasseur et al., 2020]

Observation at time t:

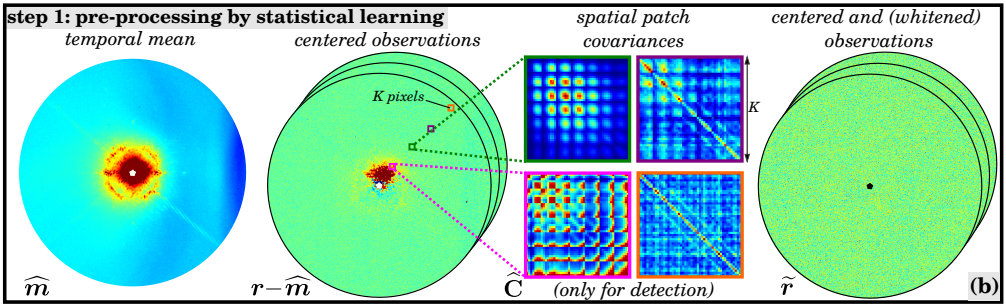
$$I_t = \underbrace{N_t}_{\text{speckle}} + \sum_{i=1}^{k} \underbrace{\alpha_i}_{\text{contrast}} \Big(\underbrace{P}_{\text{off-axis PSF}} \star \underbrace{R_t}_{\substack{\text{rotation} \\ \text{from } t_0 \text{ to } t}} \Big(\underbrace{p_{i,t_0}}_{\substack{\text{position of source } i \\ \text{at time } t_0}} \Big) \Big).$$

- We have a very good estimate of P.
- Rotation is known (due to Earth's rotation).
- Speckle is temporally quasi-static with local spatial correlations: **Patch-based model of speckles as multivariate Gaussian.**
- Estimation of α by **MLE** and detection by **likelihood ratio test**.

This is the **optimization**/model fitting part. Can supervised learning help?

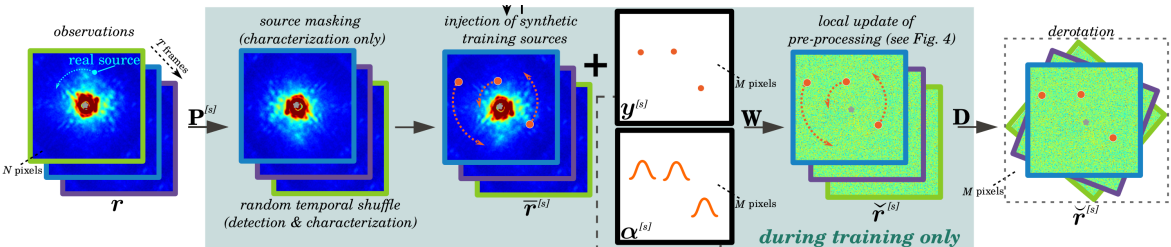
Deep PACO [Flasseur et al., 2024]

1. Temporal **centering** and spatial **whitening** using the PACO model.



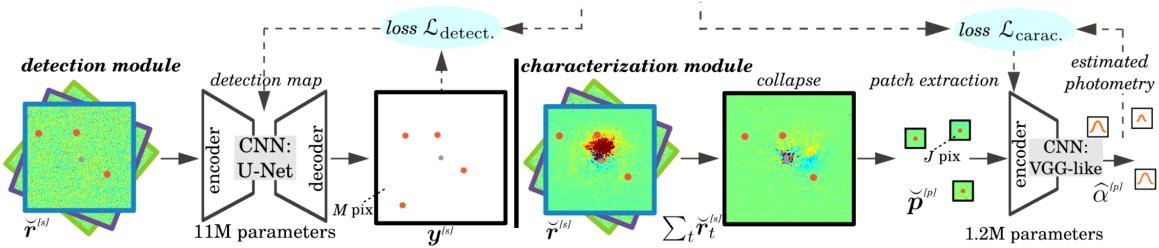
Deep PACO [Flasseur et al., 2024]

1. Temporal **centering** and spatial **whitening** using the PACO model.
2. **Dataset construction:** Synthetic source injection and derotation.

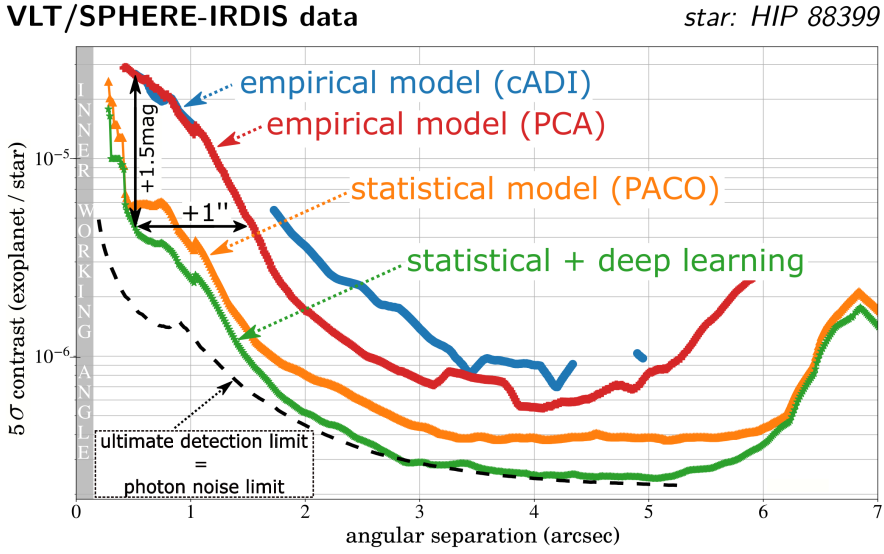


Deep PACO [Flasseur et al., 2024]

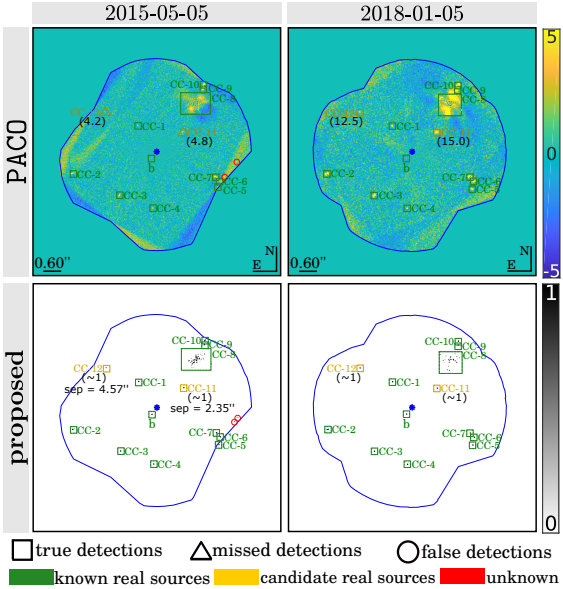
1. Temporal **centering** and spatial **whitening** using the PACO model.
2. **Dataset construction:** Synthetic source injection and derotation.
3. **Supervised learning**:



Results: contrast curves with synthetic sources



Results: observations of real known sources (HD 95086)



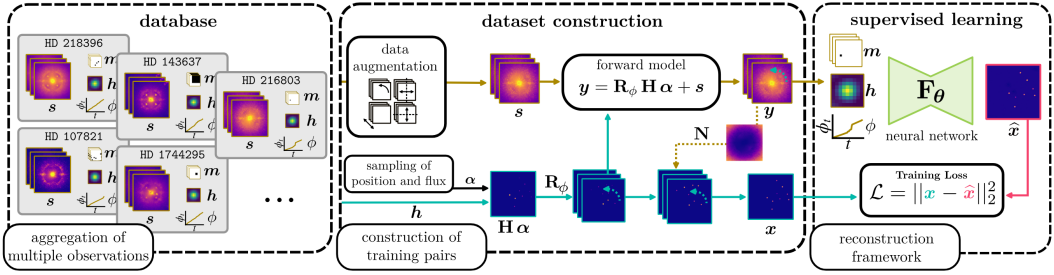
- The star with the largest number of known observed objects.
- Candidate sources are observed on several independent observations.
- Sources may be galaxies, stars, that are very far in the background, or **exoplanets**.

Limitations of deep PACO

Great proof of concept for deep learning but...

- **two-step procedure** with important issues close to the star.
- **observation-dependent model**: impractical for large-scale deployment.
- two independent models for **detection** and **characterization**.
- rigorous calibration but lack of statistical interpretability **(black box)**.

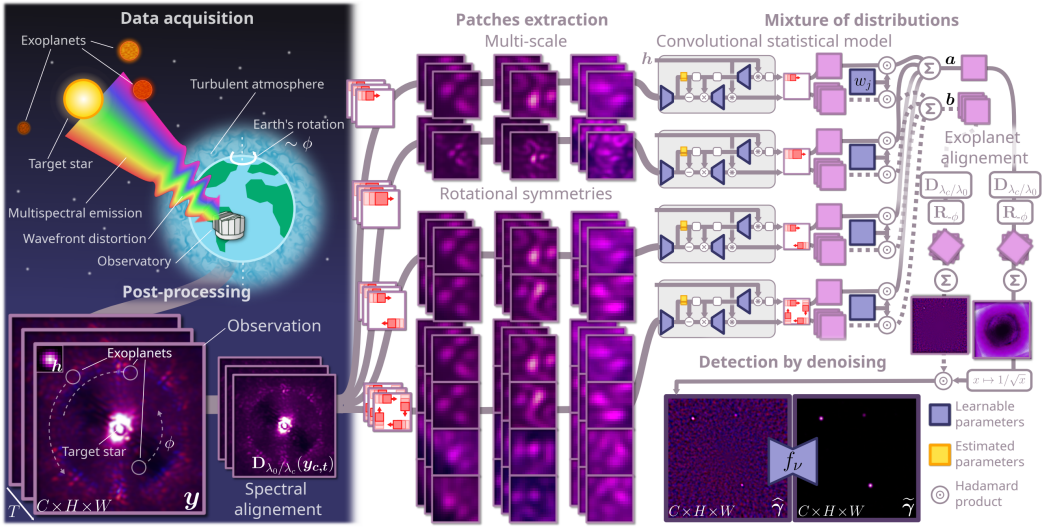
Model&Co [Bodrito et al., 2024]



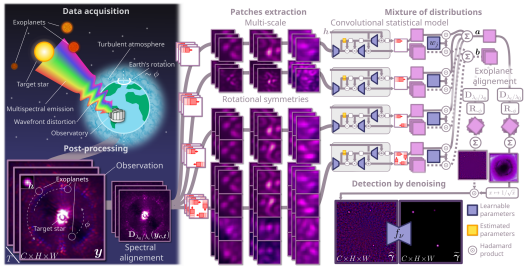
Key points

- One-step procedure integrating the PACO preprocessing within the learning model.
- **Single model** for multiple observations: appropriate for large-scale deployment.
- Smaller model than deep PACO (800K vs. 11M), higher accuracy close to the star.
- Still a **black box**, fails to jointly detect and characterize.

ExoMild [Bodrito et al., 2025] (will be on arXiv on Monday)



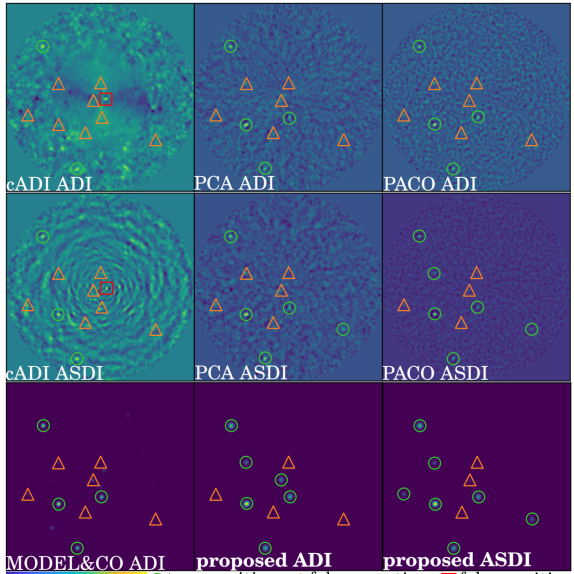
ExoMild [Bodrito et al., 2025] (will be on arXiv on Monday)



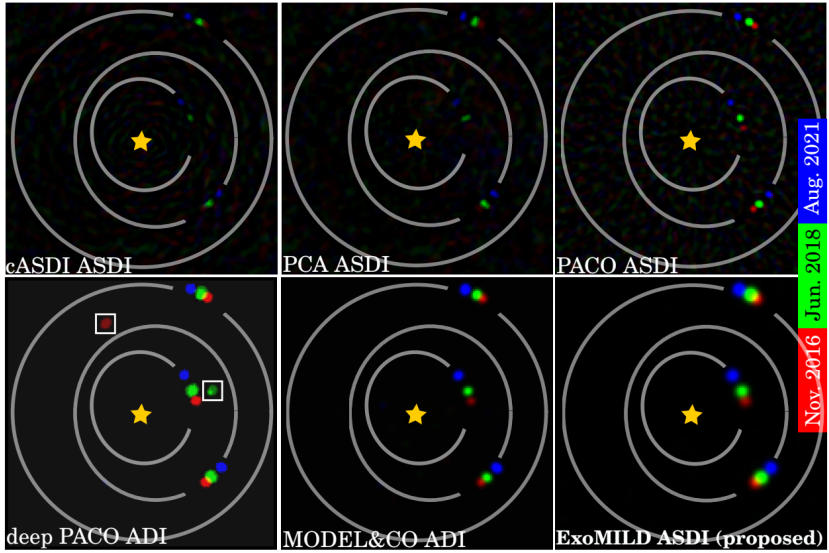
Key points: the big come back of statistical modeling

- A model that integrates pixel correlation across **multiple scales** and **spatial/spectral signal symmetries**.
- Statistical model with a learnable component (keeps the assets of Model&Co).
- Achieves **joint detection and characterization**.

Results on HD 159911, synthetic sources



Results on HR 8799, real sources



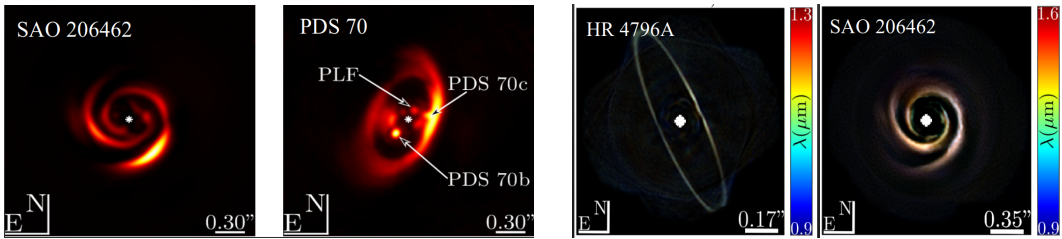
To recap: four years of work following a highly nonlinear path...

	Acc.	Acc. close to the star	statistical interpret.	large-scale (practical)	multispect.	astrometry
PACO	✓	x	✓	✓	✓	✓
Deep PACO	✓✓	x	x	x	✓	x
Model&Co	✓✓✓	✓	x	✓	x	x
ExoMild	✓✓✓	✓	✓	✓	✓	✓

To recap: four years of work following a highly nonlinear path...

	Acc.	Acc. close to the star	statistical interpret.	large-scale (practical)	multispect.	astrometry
PACO	✓	x	✓	✓	✓	✓
Deep PACO	✓✓	x	x	x	✓	x
Model&Co	✓✓✓	✓	x	✓	x	x
ExoMild	✓✓✓	✓	✓	✓	✓	✓

What is next?



If your are into nordic skiing...
feel free to chat with me!

References I

Théo Bodrito, Olivier Flasseur, Julien Mairal, Jean Ponce, Maud Langlois, and Anne-Marie Lagrange. Model&co: Exoplanet detection in angular differential imaging by learning across multiple observations. *Monthly Notices of the Royal Astronomical Society*, 534(2): 1569–1596, 2024.

Théo Bodrito, Olivier Flasseur, Julien Mairal, Jean Ponce, Maud Langlois, and Anne-Marie Lagrange. A new statistical model of star speckles for learning to detect and characterize exoplanets in direct imaging observations. *arXiv*, 2025.

Olivier Flasseur, Loic Denis, Éric Thiébaut, and Maud Langlois. Paco asdi: an algorithm for exoplanet detection and characterization in direct imaging with integral field spectrographs. *Astronomy & Astrophysics*, 637:A9, 2020.

Olivier Flasseur, Théo Bodrito, Julien Mairal, Jean Ponce, Maud Langlois, and Anne-Marie Lagrange. deep paco: Combining statistical models with deep learning for exoplanet detection and characterization in direct imaging at high contrast. *Monthly Notices of the Royal Astronomical Society*, 527(1):1534–1562, 2024.