



HAL
open science

Combining statistical learning with deep learning for improved exoplanet detection and characterization

Olivier Flasseur

► **To cite this version:**

Olivier Flasseur. Combining statistical learning with deep learning for improved exoplanet detection and characterization. Adaptive Optics for Extremely Large Telescopes 7th Edition, ONERA, Jun 2023, Avignon, France. 10.13009/AO4ELT7-2023-016 . hal-04414840

HAL Id: hal-04414840

<https://hal.science/hal-04414840v1>

Submitted on 24 Jan 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Combining statistical learning with deep learning for improved exoplanet detection and characterization

Olivier Flasseur^a, Théo Bodrito^b, Julien Mairal^c, Jean Ponce^{b,d}, Maud Langlois^a,
and Anne-Marie Lagrange^{e,f}

^aCentre de Recherche Astrophysique de Lyon, CNRS, Univ. de Lyon,
Univ. Claude Bernard Lyon 1, ENS de Lyon, France

^bDépartement d'Informatique de l'École Normale Supérieure
(ENS-PSL, CNRS, Inria), France

^cUniv. Grenoble Alpes, Inria, CNRS, Grenoble INP, LJK,
38000 Grenoble, France

^dCourant Institute of Mathematical Sciences,
Center for Data Science, New York Univ., USA

^eLaboratoire d'Études Spatiales et d'Instrumentation en Astrophysique,
Observatoire de Paris, Univ. PSL, Sorbonne Univ., Univ. Paris Diderot, France

^fUniv. Grenoble Alpes, Institut de Planétologie
et d'Astrophysique de Grenoble, France

ABSTRACT

In direct imaging at high contrast, the bright glare produced by the host star makes the detection and the characterization of sub-stellar companions particularly challenging. In spite of the use of an extreme adaptive optics system combined with a coronagraphic mask to strongly attenuate the starlight contamination, dedicated post-processing methods combining several images recorded with the pupil tracking mode of the telescope are needed to reach the required contrast.

Among the large variety of post-processing methods of the field, the PACO algorithm capturing locally the spatial correlations of the data with a multi-variate Gaussian model shown better detection sensitivity than standard methods of the field (e.g., cADI, PCA, TLOCI). However, there is a room for improvement to increase the detection sensitivity due to the approximate fidelity of the statistical model embedded in PACO with respect to the observations.

Further author information: Send correspondence to Olivier Flasseur.
E-mail: olivier.flasseur@univ-lyon1.fr, Telephone: +33 (0)4 78 86 83 70.

In that context, we recently proposed to combine the statistics-based model of PACO with a deep learning approach in a three-step algorithm. First, the data are centered and whitened locally using the PACO framework to improve the stationarity and the contrast in a preprocessing step. Second, a convolutional neural network (CNN) is trained in a supervised fashion to detect the signature of synthetic sources in the preprocessed science data. Finally, the trained network is applied to the preprocessed observations and delivers a detection map. A second network is trained to infer locally the photometry of detected sources. Both deep models are trained from scratch with a custom data augmentation strategy allowing to generate a large training set from a single spatio-temporo-spectral dataset. This strategy can be applied to process jointly the images of observations conducted with angular, and eventually spectral, differential imaging (A(S)DI). In this proceeding, we present in a unified framework the key ingredients of the deep PACO algorithm both for ADI and ASDI.

We apply our method on several datasets from the the IRDIS imager of the VLT/SPHERE instrument. Our method reaches, in average, a better trade-off between precision and recall than the comparative algorithms.

Keywords: High angular resolution – techniques: image processing – methods: numerical – methods: statistical – methods: data analysis.

1. INTRODUCTION

The future thirty meters class telescopes (e.g., ELT, GMT, TMT) will enable exploring much deeper the inner environment of nearby solar-type stars than existing facilities. This goal raises several challenges from a data science point of view, including: (i) approaching the ultimate performance of the instruments by an optimal extraction of the signals of the sought objects, (ii) capturing a highly spatially structured nuisance component subject to strong temporal fluctuations, and (iii) building a model of the nuisance component from several datasets to bypass the limits of ADI at very short angular separations. In that context, data-driven approaches combining statistical modeling with deep learning could be highly valuable to deal with the complexity of high-contrast observations.

In previous works, we have developed the PACO algorithm [8, 5, 7, 11, 10] dedicated to the post-processing of A(S)DI observations [14, 16] for exoplanet detection and characterization by direct imaging at high contrast. PACO captures locally the spatial correlations of the nuisance component (i.e., speckles plus other sources of noise) with a scaled mixture of multi-variate Gaussian models. Its parameters are estimated in a data-driven fashion at the scale of a patch of a few tens of pixels. PACO delivers reliable detection confidences with an improved sensitivity with respect to the classical processing methods of the field (e.g., cADI [14, 12], PCA [18, 1], TLOCI [15]). However, it remains room for improvement, especially at short angular separations.

Very recently, we proposed a new algorithm, dubbed deep PACO [9, 6], improving the detection sensitivity of PACO. The proposed method combines the statistics-based model of PACO with a deep learning model in a three steps procedure. First, the data are centered and whitened locally using the PACO framework to improve the stationarity and the contrast in a pre-processing step. Second, a convolutional neural network is trained from scratch, in a supervised fashion, to detect the signature of synthetic sources in the pre-processed science data. Finally, the trained network is applied to the pre-processed observations and delivers a detection map. Photometry of detected sources can be estimated by a second deep neural network. Both models are trained from scratch with a custom data augmentation strategy allowing to generate large training sets from a single spatio-temporo-spectral dataset.

We review the main ingredients of the deep PACO algorithm in Sect. 2. Section 3 presents comparative results obtained on VLT/SPHERE data. Finally, Sect. 4 draws our main conclusions and briefly discusses some future research directions.

2. THE DEEP PACO ALGORITHM

2.1 Direct Model of the Observations

A stack $\mathbf{r} \in \mathbb{R}^{N \times T \times L}$ of high-contrast images is composed of N -pixels frames, recorded at times $t \in \llbracket 1; T \rrbracket$ and in spectral channels $\ell \in \llbracket 1; L \rrbracket$. The contribution $\mathbf{r}_\ell \in \mathbb{R}^{N \times T}$ writes:

$$\mathbf{r}_\ell = \mathbf{f}_\ell + \sum_{p=1}^P \alpha_{p,\ell} \mathbf{h}_\ell(\phi_p), \quad (1)$$

where $\mathbf{f}_\ell \in \mathbb{R}^{N \times T}$ and $\mathbf{h}_\ell \in \mathbb{R}^{N \times T}$ are respectively the contribution of the nuisance component (i.e., speckles and other additive sources of noise) and of any point-like source taking the form of the off-axis point-spread function (PSF), at spectral channel ℓ . The contribution of a source $p \in \llbracket 1; P \rrbracket$ is centered at location $\mathcal{F}_{t,\ell}(\phi_p)$ in the t -th image of the ℓ -th channel, where ϕ_p is its initial location on an image at a time t_{ref} and spectral channel λ_{ref} of reference. $\mathcal{F}_{t,\ell}$ is a deterministic function accounting for the apparent motion of the sought sources induced by ASDI.

2.2 Statistical Modeling of the Nuisance Component

As in [10], we model the fluctuations of the nuisance component \mathbf{f} by a statistical model whose parameters are estimated locally (in this work, at a scale of non-overlapping square patches of K pixels). The locality of the model allows to account for the non-stationarity of the nuisance. We note \mathbb{P} the set of pixel locations on which the parameters of the statistical model should be evaluated. The distribution of a patch of nuisance $\mathbf{f}_{n,t,\ell}$ centered at pixel $n \in \mathbb{P}$, at time t , and channel ℓ is modeled by a scaled multi-variate Gaussian: $\mathcal{N}(\mathbf{m}_{n,\ell}, \sigma_{n,t,\ell}^2 \mathbf{C}_n)$. The sample estimators $\{\widehat{\mathbf{m}}_{n,\ell}, \widehat{\sigma}_{n,t,\ell}^2, \widehat{\mathbf{S}}_n\}$ of $\{\mathbf{m}_{n,\ell}, \sigma_{n,t,\ell}^2, \mathbf{C}_n\}$ are obtained from the TL patches $\mathbf{r}_n \in \mathbb{R}^{K \times T \times L}$ through the maximum-likelihood:

$$\begin{cases} \widehat{\mathbf{m}}_{n,\ell} = \left(\sum_t \widehat{\sigma}_{n,t,\ell}^{-2} \right)^{-1} \sum_t \widehat{\sigma}_{n,t,\ell}^{-2} \mathbf{r}_{n,t,\ell}, \\ \widehat{\mathbf{S}}_n = \frac{1}{TL} \sum_{t,\ell} \widehat{\sigma}_{n,t,\ell}^{-2} (\mathbf{r}_{n,t,\ell} - \widehat{\mathbf{m}}_{n,\ell})(\mathbf{r}_{n,t,\ell} - \widehat{\mathbf{m}}_{n,\ell})^\top, \\ \widehat{\sigma}_{n,t,\ell}^2 = K^{-1} (\mathbf{r}_{n,t,\ell} - \widehat{\mathbf{m}}_{n,\ell})^\top \widehat{\mathbf{S}}_n^{-1} (\mathbf{r}_{n,t,\ell} - \widehat{\mathbf{m}}_{n,\ell}), \end{cases} \quad (2)$$

The number of samples involved in the computation of $\widehat{\mathbf{S}}_n$ being lower than the number K of pixels in a patch, the sample covariance $\widehat{\mathbf{S}}_n$ is very noisy. As in [10], we regularize it by *shrinkage* [13, 4] to form $\widehat{\mathbf{C}}_n$ as follows:

$$\widehat{\mathbf{C}}_n = (1 - \widehat{\rho}_n) \widehat{\mathbf{S}}_n + \widehat{\rho}_n \widehat{\mathbf{F}}_n, \quad (3)$$

where $\widehat{\mathbf{F}}_n$ is a diagonal matrix with only the sample variances. The weight $\widehat{\rho}_n$ sets a bias-variance trade-off and it is estimated in data-driven fashion:

$$\widehat{\rho}_n = \frac{\text{tr}(\widehat{\mathbf{S}}_n^2) + \text{tr}^2(\widehat{\mathbf{S}}_n) - 2 \sum_{k=1}^K [\widehat{\mathbf{S}}_n]_{kk}^2}{(Q + 1) \left(\text{tr}(\widehat{\mathbf{S}}_n^2) - \sum_{k=1}^K [\widehat{\mathbf{S}}_n]_{kk}^2 \right)}, \quad (4)$$

with $Q = \left(\sum_{t,\ell} \widehat{\sigma}_{n,t,\ell}^{-2} \right)^2 / \left(\sum_{t,\ell} \widehat{\sigma}_{n,t,\ell}^{-4} \right)$ the equivalent number of patches involved in the computation of $\widehat{\mathbf{C}}_n$ in the presence of the scaling factors $\{\widehat{\sigma}_{n,t,\ell}^2\}_{t=1:T, \ell=1:L}$. Given the statistics of the nuisance, the dataset \mathbf{r} is preprocessed by centering and whitening to form $\widetilde{\mathbf{r}} \in \mathbb{R}^{N \times T \times L}$ with attenuated spatial structures:

$$\widetilde{\mathbf{r}}_{n,t,\ell} = \mathbf{W}_{n,t,\ell} \mathbf{r}_{n,t,\ell} = \widehat{\sigma}_{n,t,\ell}^{-1} \widehat{\mathbf{L}}_n^\top (\mathbf{r}_{n,t,\ell} - \widehat{\mathbf{m}}_{n,\ell}), \quad (5)$$

such that $\widehat{\mathbf{L}}_n \widehat{\mathbf{L}}_n^\top = \widehat{\mathbf{C}}_n^{-1}$.

2.3 Detection by Supervised Deep Learning

We aim to produce a detection map $\hat{\mathbf{y}}$ in $[0; 1]^M$, where each pixel-value represents the pseudo-probability that a source is centered at that location at time t_{ref} . In order to perform supervised training of our model, we generate S pairs $\{\check{\mathbf{r}}^{[s]}; \mathbf{y}^{[s]}\}_{s=1:S}$ of synthetic samples/ground truths, resulting from the massive injection of point-like sources. The trained model is data-dependent, i.e. it differs for each A(S)DI dataset due to the high variability of the nuisance component from one observation to the other. We thus resort to data-augmentation to generate a large training basis from a single dataset: we apply a random permutation of the T images of each spectral channel for each training sample s . Synthetic sources are injected inside the temporally permuted data following the direct model (1) to form the intermediate datasets $\{\tilde{\mathbf{r}} \in \mathbb{R}^{N \times T \times L}\}_{s=1:S}$:

$$\tilde{\mathbf{r}}_\ell^{[s]} = \mathbf{P}^{[s]} \mathbf{f}_\ell + \sum_{p=1}^{P^{[s]}} \alpha_{p,\ell}^{[s]} \mathbf{h}_\ell(\phi_p^{[s]}), \quad (6)$$

with \mathbf{P} the operator performing a random permutation of the temporal frames. For each synthetic source p and if $L > 2$, we select randomly a synthetic SED $\alpha_p \in \mathbb{R}^L$ from a custom library of 10,000 sub-stellar spectra generated with ExoREM [3]. Before injection of synthetic sources, the initial dataset \mathbf{r} is pre-processed to form $\tilde{\mathbf{r}}$. After injection of synthetic sources, the set $\mathbb{S}^{[s]}$ of locations impacted by the signal of the fiducial sources is determined. Outside $\mathbb{S}^{[s]}$, the pre-processed images are obtained from the temporal permutation of $\tilde{\mathbf{r}}$. Inside $\mathbb{S}^{[s]}$, the statistics of the nuisance and the pre-processed images are updated given the contamination of the $P^{[s]}$ injected sources to form $\{\check{\mathbf{r}} \in \mathbb{R}^{N \times T \times L}\}_{s=1:S}$ such that:

$$\check{\mathbf{r}}_{n,t,\ell}^{[s]} = \begin{cases} \mathbf{W}_{n,t,\ell} \tilde{\mathbf{r}}_{n,t,\ell}^{[s]}, & \text{for } n \in \mathbb{S}^{[s]} \cap \mathbb{P}, \\ \mathbf{P}^{[s]} \tilde{\mathbf{r}}_{n,t,\ell}, & \text{for } n \in \mathbb{P} - \mathbb{S}^{[s]} \cap \mathbb{P}. \end{cases} \quad (7)$$

Finally, the apparent motion of the injected sources are compensated to co-align their signals within $\{\check{\mathbf{r}}^{[s]} \in \mathbb{R}^{N \times T \times L}\}_{s=1:S}$:

$$\check{\mathbf{r}}_{t,\ell}^{[s]} = \mathbf{D}_{t,\ell} \check{\mathbf{r}}_{t,\ell}^{[s]}, \quad (8)$$

with $\mathbf{D}_{t,\ell}$ the operator performing a rotation of the image at time t and spectral channel ℓ by the opposite of the parallactic angle at time t .

At training time, the network parameters are optimized by minimizing the Dice2 loss [19]:

$$\mathcal{L}^{[s]} = 1 - \frac{\sum_m \mathbf{y}_m^{[s]} \hat{\mathbf{y}}_m^{[s]} + \epsilon}{\sum_m \mathbf{y}_m^{[s]} + \hat{\mathbf{y}}_m^{[s]} + \epsilon} - \frac{\sum_m (1 - \mathbf{y}_m^{[s]}) (1 - \hat{\mathbf{y}}_m^{[s]} + \epsilon)}{\sum_m 2 - \mathbf{y}_m^{[s]} - \hat{\mathbf{y}}_m^{[s]} + \epsilon}, \quad (9)$$

where $\{\mathbf{y}^{[s]}; \hat{\mathbf{y}}^{[s]}\}$ is a set of ground truth and predicted detection maps, and ϵ is a small stability parameter. At validation time, we compute from a predicted detection map $\hat{\mathbf{y}}^{[s]} \in [0; 1]^M$ thresholded at $\tau \in [0; 1]$, the true positive rate (TPR) and the false discovery rate (FDR). From these two quantities, receiver operating curves (ROCs) are built and the area under ROCs (AUC) is derived as an overall measurement of the trade-off between precision and recall.

Concerning the architecture, we use a U-Net with a ResNet18 as encoder backbone*. The network weights are optimized from scratch with AMSGrad. The number of epochs, the number of samples per epoch, the batch size, the weight decay, and the learning rate are discussed in details in [9, 6].

3. RESULTS ON VLT/SPHERE DATA

For our experiments, we have selected three datasets from the VLT/SPHERE-IRDIS imager [2]. The performance of deep PACO are compared with the cADI, PCA, and PACO algorithms. The details about the observing

*https://github.com/qubvel/segmentation_models.pytorch

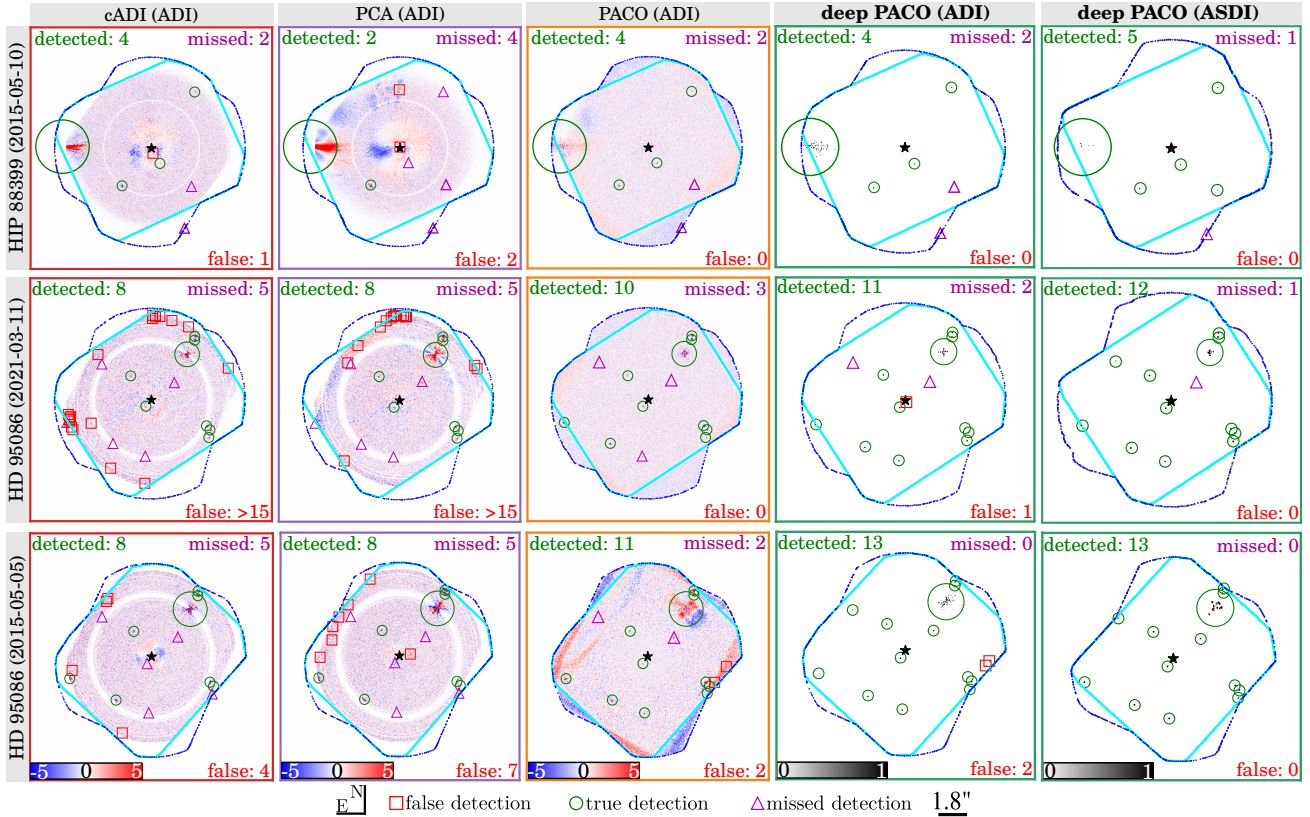


Figure 1. Detection maps obtained with cADI, PCA, PACO, and deep PACO on three datasets (HIP 88399 –2015-05-10–, HD 95086 –2021-03-11–, HD 95086 –2015-05-05–) from the VLT/SPHERE-IRDIS instrument.

parameters, the pre-reduction of the recorded datasets, and the setting of the comparative algorithms are discussed in [9, 6].

Figure 1 shows the detection maps obtained with the comparative algorithms. The detected peaks are classified as true, missed, or false detections based on a thresholding of the output map at $\tau = 5$ for the algorithms producing a signal-to-noise map (i.e., cADI, PCA, PACO), and at $\tau = 0.5$ for deep PACO producing a pseudo-probability map. For the four considered algorithms, the results are obtained by processing the first spectral channel, i.e. exploiting solely ADI. For the proposed approach, we compare ADI and ASDI by also considering a joint processing of the two spectral channels of the dual band observations. These results illustrate the ability of deep PACO to reach, in average, a better detection sensitivity and, at the same time, to avoid some spurious false alarms produced by the comparative algorithms. The joint spectral processing allows to further improve the trade-off between precision and recall.

4. CONCLUSION

As a proof of concept, we applied our method on several datasets from the VLT/SPHERE-IRDIS instrument. We compared the deep PACO method against some state-of-the-art algorithms of the field, including PACO. With ADI, our method reaches in average a better trade-off between precision and recall than the comparative algorithms. A joint processing of spatio-temporo-spectral observations obtained with ASDI allows to further improve the detection performance.

On the methodological side, we are currently investigating ways to improve the deep learning stage of the proposed algorithm by jointly modeling the typical distribution of the nuisance component from multiple datasets. This methodology would strongly differ from reference differential imaging (RDI, [17]) in the sense that a

highly non-linear model would be learned from the archived observations and would exploit several prior domain knowledge. On the application side, we will consider to adapt and to apply the deep PACO algorithm on simulated ELT high-contrast data (e.g., from ELT/HARMONI) to ground its benefits in the context of the future giant telescopes.

ACKNOWLEDGMENTS

This work is supported by the French National Programs (PNP and PNPS) and by the Action Spécifique Haute Résolution Angulaire (ASHRA) of CNRS/INSU co-funded by CNES. This project is also supported in part by the European Research Council (ERC) under the European Union’s Horizon 2020 research and innovation programme (COBREX; grant agreement n° 885593). The work of TB and JP was supported in part by the Inria/NYU collaboration, the Louis Vuitton/ENS chair on artificial intelligence and the French government under management of Agence Nationale de la Recherche as part of the *Investissements d’avenir* program, reference ANR19-P3IA0001 (PRAIRIE 3IA Institute). The work of JM was supported in part by the ERC grant number 714381 (SOLARIS project) and by ANR 3IA MIAI@Grenoble-Alpes (ANR-19-P3IA-0003).

The raw data used in this article are freely available on the ESO archive facility at http://archive.eso.org/eso/eso_archive_main.html. They were pre-reduced with the SPHERE Data Centre, jointly operated by OSUG/IPAG (Grenoble), PYTHEAS/LAM/CESAM (Marseille), OCA/Lagrange (Nice), Observatoire de Paris/LESIA (Paris), and Observatoire de Lyon/CRAL (Lyon, France). This work was also granted access to the HPC resources of IDRIS under the allocation 2022-AD011013643 made by GENCI.

References

- [1] Adam Amara and Sascha P Quanz. “PYNPOINT: an image processing package for finding exoplanets”. In: *Monthly Notices of the Royal Astronomical Society* 427.2 (2012), pp. 948–955.
- [2] J-L Beuzit et al. “SPHERE: the exoplanet imager for the Very Large Telescope”. In: *Astronomy & Astrophysics* 631 (2019), A155.
- [3] Benjamin Charnay et al. “A self-consistent cloud model for brown dwarfs and young giant exoplanets: comparison with photometric and spectroscopic observations”. In: *The Astrophysical Journal* 854.2 (2018), p. 172.
- [4] Yilun Chen et al. “Shrinkage algorithms for MMSE covariance estimation”. In: *IEEE Transactions on Signal Processing* 58.10 (2010), pp. 5016–5029.
- [5] Olivier Flasseur et al. “An unsupervised patch-based approach for exoplanet detection by direct imaging”. In: *IEEE International Conference on Image Processing*. 2018, pp. 2735–2739.
- [6] Olivier Flasseur et al. “deep PACO: Combining statistical models with deep learning for exoplanet detection and characterization in direct imaging at high contrast”. In: *under minor review in Monthly Notices of the Royal Astronomical Society* (2023).
- [7] Olivier Flasseur et al. “Exoplanet detection in angular and spectral differential imaging: local learning of background correlations for improved detections”. In: *SPIE Astronomical Telescopes + Instrumentation*. Vol. 10703. International Society for Optics and Photonics. 2018, 107032R.
- [8] Olivier Flasseur et al. “Exoplanet detection in angular differential imaging by statistical learning of the nonstationary patch covariances - The PACO algorithm”. In: *Astronomy & Astrophysics* 618 (2018), A138.
- [9] Olivier Flasseur et al. “Exoplanet detection in angular differential imaging: combining a statistics-based learning with a deep-based learning for improved detections”. In: *Adaptive Optics Systems VIII*. Vol. 12185. SPIE. 2022, pp. 1154–1167.
- [10] Olivier Flasseur et al. “PACO ASDI: an algorithm for exoplanet detection and characterization in direct imaging with integral field spectrographs”. In: *Astronomy & Astrophysics* 637 (2020), A9.
- [11] Olivier Flasseur et al. “Robustness to bad frames in angular differential imaging: a local weighting approach”. In: *Astronomy & Astrophysics* 634 (2020), A2.

- [12] A-M Lagrange et al. “A probable giant planet imaged in the β Pictoris disk: VLT/NaCo deep L'-band imaging”. In: *Astronomy & Astrophysics* 493.2 (2009), pp. L21–L25.
- [13] Olivier Ledoit and Michael Wolf. “A well-conditioned estimator for large-dimensional covariance matrices”. In: *Journal of Multivariate Analysis* 88.2 (2004), pp. 365–411.
- [14] Christian Marois et al. “Angular Differential Imaging: A Powerful High-Contrast Imaging Technique”. In: *The Astrophysical Journal* 641.1 (2006), p. 556.
- [15] Christian Marois et al. “TLOCI: A fully loaded speckle killing machine”. In: *Proceedings of the International Astronomical Union* 8.S299 (2013), pp. 48–49.
- [16] René Racine et al. “Speckle noise and the detection of faint companions”. In: *Publications of the Astronomical Society of the Pacific* 111.759 (1999), p. 587.
- [17] Garreth Ruane et al. “Reference star differential imaging of close-in companions and circumstellar disks with the, NIRC2 vortex coronagraph at the W.M. Keck observatory”. In: *The Astronomical Journal* 157.3 (2019), p. 118.
- [18] Rémi Soummer, Laurent Pueyo, and James Larkin. “Detection and characterization of exoplanets and disks using projections on Karhunen-Loève eigenimages”. In: *The Astrophysical Journal Letters* 755.2 (2012), p. L28.
- [19] Lu Wang et al. “An improved Dice loss for pneumothorax segmentation by mining the information of negative areas”. In: *IEEE Access* 8 (2020), pp. 167939–167949.