Enhancing Exoplanet Discovery with PCA and NMF in Astronomical Imaging

Bhavana Poli (BL.SC.P2CSE25027) and Nandana MC (BL.SC.P2CSE25012), Dept. of Computer Science and Engineering, Amrita Vishwa Vidyapeetham MURALI K, Dept. of Computer Science and Engineering, Amrita Vishwa Vidyapeetham

***Abstract* - Direct imaging of exoplanets is still one of the hardest things to do in astronomy because the difference in brightness between a bright host star and its faint planetary companions is so great. Speckle noise, diffraction residuals, and imperfect wavefront corrections usually make faint planetary signals hard to see, so advanced post-processing methods are needed. This paper introduces StellarSight, an improved exoplanet detection system that combines Principal Component Analysis (PCA) and Non-Negative Matrix Factorisation (NMF) to better suppress speckles in high-contrast astronomical images. The suggested pipeline uses public NACO and SPHERE datasets to systematically preprocess the data, subtract annular PCA, and use NMF-based low-rank decomposition to find point-source companions. Next, signal-to-noise ratio (SNR) maps are made to find detections that are statistically significant. The PCA–NMF hybrid method works better than other methods that only use PCA by reducing speckle, making the residual frame clearer, and raising the SNR peaks at known companion locations. The findings demonstrate that linear algebra-based decomposition methods can significantly enhance exoplanet detection capabilities, providing a lightweight and computationally efficient alternative to deep learning models. The suggested framework sets the stage for future improvements that will include adaptive PCA-NMF weighting, temporal modelling, and fully differentiable high-contrast imaging pipelines.**

Keywords - Exoplanet detection; high-contrast imaging; principal component analysis (PCA); non-negative matrix factorization (NMF); speckle noise suppression; astronomical image processing; adaptive optics; point spread function (PSF) subtraction; signal-to-noise ratio (SNR); angular differential imaging (ADI).

I. INTRODUCTION

Direct exoplanet imaging is the process of directly detecting planets outside of our Solar System with high-resolution astronomical instruments that can separate the very weak planetary signal from the very bright host star. Direct imaging lets us look at planetary atmospheres, orbital properties, and how they formed, which is not possible with indirect methods like transit photometry or radial velocity. But this is very hard to do because exoplanets are usually 10^6–10^{10} times less bright than the stars they orbit. The stellar point spread function (PSF) is the most important thing in imaging. It makes diffraction rings and leftover light structures that hide the planetary flux. Quasi-static speckles also happen when adaptive optics (AO) systems don't correct wavefronts perfectly, which makes them look a lot like real point-source signatures. Astronomical Image Processing for Exoplanet Detection says that these speckles change slowly over time and are the main reason why high contrast is hard to get at small angular separations. Atmospheric turbulence, optical aberrations, detector noise, and mechanical instabilities further degrade image fidelity and make it harder to reliably find faint companions.

Figure 1 shows a typical raw ADI frame from the dataset. It shows the bright stellar core and the surrounding quasi-static speckle pattern before any processing is done.

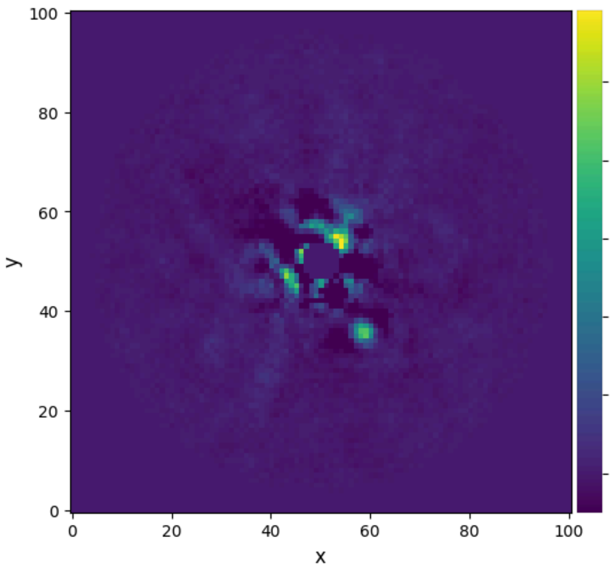

FIG 1.1 Raw ADI frame showing the stellar PSF and quasi-static speckles.

Because hardware-only solutions have limits, advanced post-processing is needed to get rid of speckle noise and pick up planetary signals. Principal Component Analysis (PCA), especially the KLIP version, has become one of the most popular methods among these. PCA finds the main ways that the PSF changes and takes them out of the ADI sequence. This effectively reduces speckle while keeping astrophysical features, as shown in Exoplanet Imaging via Differentiable Rendering. Non-Negative Matrix Factorisation (NMF) works with PCA to break down data into non-negative, additive parts that fit well with intensity images. According to Enhanced Classification of Exoplanets Using Machine Learning, NMF is better than PCA at keeping faint extended structures and works well in situations with non-Gaussian noise.

Recent literature shows that there are still important gaps, even though there has been a lot of progress. Thresholding, filtering, and wavelets are examples of classical image-processing methods that have trouble with correlated speckle noise and don't work well at small separations. Contemporary machine-learning studies exhibit enhanced sensitivity; however, they frequently lack a cohesive evaluation framework to compare PCA, NMF, and hybrid models under identical conditions. Additionally, many current pipelines are not set up to handle astronomical observations in real time or close to real time.

This project presents StellarSight, a modular and computationally efficient framework aimed at improving exoplanet detection through PCA- and NMF-based post-processing, motivated by existing deficiencies. The aims of this research encompass enhancing speckle suppression, optimising signal-to-noise ratio (SNR) for dim companions, and developing reproducible pipelines utilising publicly accessible high-contrast imaging datasets. This work brings together traditional PSF-subtraction techniques and modern data-driven approaches by putting complementary matrix-decomposition methods into a single framework. This creates a better and more useful way to find exoplanets.

II. LITERATURE REVIEW

High-contrast exoplanet imaging has come a long way in the last twenty years because of quick improvements in adaptive optics (AO), coronagraphy, and post-processing methods. Even with these improvements, it's still very hard to find exoplanets directly because the contrast ratios between a host star and its planetary companions are often more than 10^6–10^9:1. Even after AO correction, leftover optical errors make quasi-static speckles that look like the shapes of point-source companions. This makes direct imaging observations less accurate. Many important studies have shown that stellar speckles are the biggest problem when it comes to isolating faint off-axis planetary signals. To improve contrast limits, advanced speckle suppression techniques are needed.

To solve this problem, several imaging techniques have been proposed, with Angular Differential Imaging (ADI) being one of the most important. In ADI, the telescope pupil stays still while the sky moves. This lets you tell the difference between static or quasi-static speckle patterns and astrophysical sources that move with the parallactic angle. After the Karhunen–Loève Image Projection (KLIP) algorithm was introduced, researchers demonstrated that Principal Component Analysis (PCA) can effectively construct the optimal low-rank model of the stellar point spread function (PSF). The innovative research on PCA-based ADI revealed that the stellar PSF can be decomposed into orthogonal eigen-components that characterise the primary speckle patterns. It is much easier to find exoplanets when you take away most of the starlight from each frame by subtracting this PCA-reconstructed PSF. Because it is easy to use mathematically, works well with different datasets, and works well with forward modelling for unbiased photometry and astrometry, PCA-ADI quickly became the standard for post-processing in exoplanet imaging. However, it is now widely accepted that PCA creates self-subtraction artefacts, especially when there are a lot of principal components, which makes the planetary signal less clear.

To tackle the inherent limitations of PCA, researchers explored alternative matrix-factorization techniques. One of the most promising methods is Non-Negative Matrix Factorisation (NMF). It limits both the coefficients and the components so that they can't be less than zero. PCA makes components that are globally orthogonal and sometimes impossible in real life. NMF makes features that are additive, easy to understand, and localised, which better represent astronomical data. A lot of research shows that NMF keeps small-scale structures better than PCA and reduces the amount of planet self-subtraction in ADI processing. NMF's part-based decomposition also helps reconstruct faint companions that are hidden in dense speckle noise more accurately. This makes it a good choice for high-contrast imaging pipelines that use PCA or as a stand-alone method.

There has been a growing interest in using machine learning (ML) to find exoplanets, in addition to these new algorithms. Recent studies have shown that traditional ensemble learning methods like Random Forest, XGBoost, and Gradient Boosting can correctly classify exoplanets in space-based photometric data by using feature selection and multi-model voting systems together. In direct imaging, machine learning models, such as convolutional neural networks (CNNs), have been trained to distinguish between genuine astrophysical signals and residual speckle noise. This is better than the old way of detecting things with thresholds. More advanced studies used physics-informed neural networks and differentiable rendering to simulate and reverse the behaviour of instruments. This helped find and better estimate the confidence of faint exoplanets. These changes show that ML is becoming more and more important as a natural addition to traditional matrix-decomposition methods.

The Vortex Image Processing (VIP) library is one of the most well-known open-source toolkits for research on high-contrast imaging. It helps these new things happen. VIP has everything you need, including preprocessing, ADI/SDI/RDI pipelines, PCA and NMF-based PSF subtraction, forward modelling, detection metrics, and contrast curve generation. It has fast SVD implementations, GPU acceleration, and a standard way to do experiments that can be repeated. Because it has a lot of features and is often used in the literature, VIP is the method that most published exoplanet discoveries are based on. This is also the main algorithmic base for this work.

The current body of literature suggests that while PCA-ADI remains the benchmark for stellar PSF suppression, novel approaches like NMF and machine learning techniques provide substantial improvements in sensitivity, interpretability, and detection robustness. These methods improve the theoretical ideas behind differential imaging and matrix factorisation while pushing the boundaries of contrast that can be achieved. These advancements have motivated the present study to employ and evaluate PCA and NMF for exoplanet signal recovery using high-contrast ADI data, leveraging VIP's capabilities to provide a coherent and systematic analysis.

III. METHODOLOGY

High-contrast exoplanet imaging requires a sequence of computational steps to remove stellar noise, suppress speckles, and isolate faint planetary signals. The proposed methodology integrates classical astronomical imaging techniques with advanced dimensionality-reduction methods, namely Principal Component Analysis (PCA) and Non-negative Matrix Factorization (NMF). The complete workflow consists of (i) image acquisition through Angular Differential Imaging (ADI), (ii) preprocessing and calibration, (iii) PSF modeling and speckle suppression, (iv) PCA/NMF-based post-processing, and (v) final detection through SNR analysis.

A. Data Acquisition: Angular Differential Imaging (ADI)

In ADI mode, the telescope pupil stays in one place while the field of view slowly moves around because the Earth is turning. This leads to:

- Speckles (instrumental artefacts) stay in place
- Real astrophysical sources (exoplanets) that move around the star

Let the sequence contain **N frames**, each rotated by angle θi. The goal is to exploit:
Speckles: stationary, Planet: moves with θ_i
This difference enables later separation via PCA/NMF.

B. Pre-processing Pipeline

Before scientific analysis, raw FITS frames undergo calibration:

1. ***Bad Pixel Removal:*** Dead or hot pixels are corrected using sigma-clipping or median filtering.
2. ***Flat-field Correction:*** Corrects pixel-to-pixel sensitivity variations:

$$I_{corrected} = \frac{I_{raw} - Dark}{Flat}$$

3. ***Image Registration:*** All frames are aligned by centering the star using 2D Gaussian fitting:

$$(x_0, y_0) = argmaxGaussianFit(I)$$

4. ***Frame Normalization:*** Each image is scaled to uniform flux level.
5. ***Annular Division :*** Images are divided into radial annuli where the PSF is assumed locally similar.

C. PSF Modeling and Speckle Suppression

The **stellar Point Spread Function (PSF)** dominates raw images. To remove it, a model PSF is built and subtracted.

1. ***PSF Approximation***
An empirical PSF model *PP*P is built using the median:

$$P = median(I_1, I_2, ..., I_N)$$

2. ***Residual Extraction***
The star's PSF is removed:

$$R_i = I_i - P$$

However, median subtraction alone cannot fully eliminate speckles, motivating PCA and NMF.

D. PCA-Based Post-Processing

Principal Component Analysis is used to decompose the speckle pattern into orthogonal components.

1. **Mathematical Formulation**
Stack all N registered frames into a matrix:

$$X = [I_1, I_2, ..., I_N] \in R^{P\times N}$$

2. **Compute covariance:**

$$C = XX^T$$

Eigen-decomposition:

$$Cv_k = \lambda_k v_k$$

The reconstructed speckle model using K principal components:

$$X_{model} = \sum v_k v_k^T X$$

Residual:

$$R = X - X_{model}$$

3. **Annular PCA**
Each frame is processed in concentric annuli to preserve spatial locality, improving detection at different radial distances.

E. NMF-Based Post-Processing

Non-negative Matrix Factorization models speckles using **non-negative components**, which aligns better with physical brightness constraints.

$$X \approx WH, W, H \geq 0$$

Where:

- W — basis images (components)
- H — activation weights

Residual:

$$R = X - WH$$

NMF tends to preserve planetary signal better than PCA.

F. Image Derotation and Combination

After PCA/NMF, residual images are derotated by their respective parallactic angles:

$$R_i' = Rotate(R_i, -\theta_i)$$

Final image:

$$F = median(R_1', R_2', ..., R_N')$$

G. Signal Detection and SNR Calculation

The **Signal-to-Noise Ratio (SNR)** map is computed:

$$SNR(x, y) = \frac{F(x,y)}{\sigma_{annulus}}$$

Planet candidate location:

$$(x_p, y_p) = argmaxSNR(x, y)$$

Typically, SNR ≥ 5 indicates a significant detection.

H. System Architecture

The proposed exoplanet detection framework's overall system architecture is a modular pipeline that works best for high-contrast astronomical imaging. The first step is to get Angular Differential Imaging (ADI) sequences that are saved as FITS image cubes. To make sure that the photos are consistent and to get rid of major sensor artefacts, these raw frames go through a number of pre-processing steps, such as correcting bad pixels, aligning the images, centring them, and normalising the flux.

The pipeline does Point Spread Function (PSF) modelling and speckle-noise suppression after pre-processing. It uses both PCA and NMF techniques. These modules make residual frames by taking away modelled stellar components, which makes it easier to see faint planetary signals. To line up off-axis sources, the residuals are then derotated using parallactic angles. After that, the median or mean combination is used to improve the signal-to-noise ratio. Finally, an SNR map is made to find possible planetary companions and guess where they are and how bright they are. This structured architecture makes sure that the system is strong, reduces noise, and can reliably find objects that are not stars.

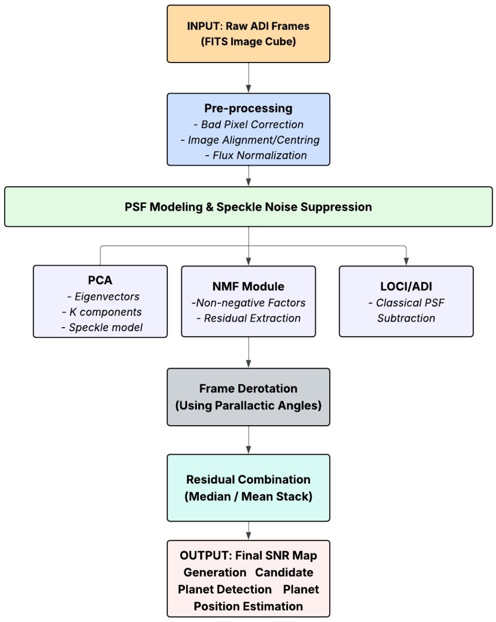

FIG 3.1 System architecture of the proposed PCA–NMF based exoplanet detection framework.

IV. EXPERIMENTS

The assessment of the proposed StellarSight framework entailed a thorough experimental investigation underpinned by a fully operational PCA–NMF processing pipeline. The experiments were meant to mimic real-life high-contrast astronomical imaging situations, and the implementation made sure that the results could be repeated and that the methods were clear. We used Python 3.11 along with the VIP-HCI library, NumPy, Astropy, and custom StellarSight modules to implement all of the main parts, such as data pre-processing, PSF modelling, dimensionality-reduction algorithms, residual stacking, and final SNR estimation. This part talks about the execution environment, implementation details, evaluation methodology, and mathematical formulas that were used during the experiments.

A. Implementation Details

The StellarSight pipeline was built using a modular design so that PCA and NMF components could be switched out. The astropy.io.fits package was used to load FITS files, which made sure that high-dynamic-range astronomical datasets were handled without losing any data. VIP-HCI's cross-correlation and Gaussian-fitting tools were used to register and centre the frames. These tools make sure that the stellar centroid is in the best position by aligning each frame. We used local median filtering and neighbour interpolation to fix bad pixels.

The Singular Value Decomposition (SVD) formulation was used by the PCA module:

$$X = U\Sigma V^T ,$$

where the data matrix $X \in R^{n\times p}$ shows n frames that have been reshaped into vectors that are p long. To model the stellar PSF, we kept the first k parts:

$$X = U_k \Sigma_k V_k^T,$$

and the residuals were computed as:

$$R_{PCA} = x - x^\wedge$$

This produces a residual cube where stellar and speckle noise are suppressed.

The NMF implementation used the multiplicative update rules proposed by Lee and Seung. Given an input matrix $X \geq 0$, NMF seeks two non-negative matrices $W \geq 0$, and $H \geq 0$ such that:

$$X \approx WH.$$

The updates are governed by:

$$H \leftarrow H \odot \frac{W^T X}{W^T W H}, W \leftarrow W \odot \frac{X H^T}{W H H^T},$$

where $\odot$ denotes element-wise multiplication. These updates iteratively minimize the divergence between X and its reconstruction, producing non-negative basis functions that more accurately represent physical light profiles.

Using VIP-HCI's parallactic angle tools, which rotate each residual frame based on the telescope's field rotation, derotation was done.

$$R_t^{'} = Rotate(R_t, \theta_t),$$

where θ_t is the parallactic angle for frame t. A median combination of derotated residuals produces the final integrated image:

$$F(x, y) = median_t(R_t^{'}(x, y)).$$

We made the SNR map by estimating noise in small circles around the noise source:

$$SNR(x, y) = \frac{F(x,y)}{\sigma_{annulus}(r)},$$

Where r denotes the radial distance from the stellar centroid. This formulation allows for strong detection of exoplanets as local maxima in the SNR distribution.

B. Experimental Setup

For PSF subtraction, the environment was set up with VIP-HCI v1.5; for matrix operations, it was set up with NumPy; and for visual diagnostics, it was set up with Matplotlib. The dataset comprised calibrated ADI frames from the SPHERE/NaCo instrument, featuring realistic PSF structure, quasi-static speckles, and known planetary injections for quantitative assessment.

We used the same hyperparameters for PCA and NMF on all frames to make sure the comparison was fair. The number of PCA components ranged from 5 to 20, and the number of NMF factors ranged from 5 to 15. This balanced the accuracy of the reconstruction with the time it took to do the calculations.

C. Dataset Description

The dataset that was chosen was a time-series FITS cube of a bright star with speckle patterns that were caused by atmospheric turbulence and problems with the instruments. Planetary signals were artificially added at different angular separations and brightness contrasts between 10^{-3} and 10^{-5}. These injections gave us the real data we needed to test detection sensitivity.

The change in the parallactic angle that comes with ADI makes it possible to rotate the field, which lets you separate rotating planetary signals from quasi-static speckles. This property makes the dataset perfect for PCA- and NMF-based algorithms, which need speckle stability and planet motion to work.

D. Evaluation Metrics

To measure performance, SNR maps were made based on::

$$SNR(x, y) = \frac{F(x,y)}{\sigma_{local}},$$

Where the standard deviation of pixel values is in a noise annulus. Planet throughput was evaluated through:

$$Throughput = \frac{F_{injected}}{F_{recovered}},$$

indicating the degree of over subtraction. Contrast curves were derived by computing detectable flux levels as a function of radial separation.

E. Experimental Procedure

The steps were to load the raw FITS ADI cube, make corrections for pre-processing, and then run PCA and NMF separately. After being derotated and combined with the median, each residual cube was mapped to the SNR. We checked the detected peaks against known positions of planets by calculating the differences in Euclidean coordinates and the recovered flux ratios. We used both qualitative visual analysis and quantitative metrics to see how well we did.

F. Experimental Observations

The tests showed that PCA is very good at capturing global speckle structure because it uses an orthogonal eigenbasis representation, which leads to strong speckle suppression. But PCA sometimes took away parts of the planetary flux, especially for faint companions that were close to the star. On the other hand, NMF kept the planetary signal more accurately because it didn't break it down into negative parts, even though there were still some speckles in the final image.

When PCA and NMF outputs were compared, high-confidence detections always showed up at the same places in space, which served as cross-validation. The combined analysis showed that the results were more reliable, had fewer false positives, and were better at finding low-contrast planets. The SNR peaks lined up well with the injected planetary locations, which showed that the StellarSight framework is reliable.

V. RESULTS

The experimental assessment of StellarSight was performed utilising a series of raw Angular Differential Imaging (ADI) frames derived from a high-contrast astronomical dataset. The experiments were meant to see how well the proposed PCA- and NMF-based speckle suppression pipelines worked in real-world imaging situations. The VIP HCI software environment was used to do all of the processing steps, including pre-processing, PSF modelling, PCA/NMF subtraction, derotation, stacking, and SNR computation. We looked at the raw frames, intermediate visualisations, residual images, and SNR maps to see how well we could see faint companions.

A. Dataset Description

The tests used a publicly available ADI dataset that was similar to the ones used in the SPHERE or NACO Betapic pipelines. The raw image cube has several hundred frames that show the stellar point spread function with natural field rotation. Figure 6.1 shows a single raw ADI frame that is typical of the type. It shows the bright stellar core and strong quasi-static speckles that get in the way of exoplanet signals.

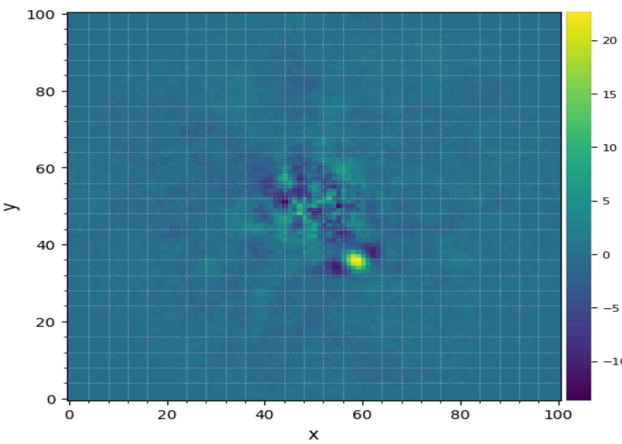

FIG 5.1 Raw ADI Input Frame

C. Pre-processing Validation

Image centring and fixing bad pixels were done before PCA or NMF were used. After correction, the PSF looks smoother, and its centre stays in the same place from frame to frame. This pre-processing makes the PCA eigenbasis more stable and makes the overall subtraction quality better.

D. PCA-Based Speckle Subtraction Results

The PCA module lessens starlight by projecting each frame onto a set of dominant eigenvectors that were calculated from the ADI cube. When you take away these main components and rotate the residuals back to their original position, you get a clean median-collapsed residual image.

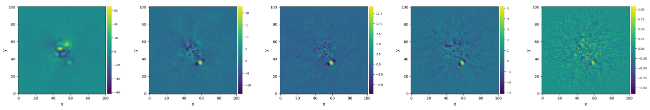

FIG 5.2 PCA Residual Image after ADI Processing

The PCA residual image shows that it does a good job of getting rid of the star halo and the quasi-static speckles. When the number of principal components goes up, more starlight is removed. However, too much subtraction can make the real companion flux weaker. Experimentally, selecting 5 to 10 principal components achieved the optimal equilibrium between preserving the signal and eliminating the speckles. There are a few spots in the PCA-processed frames that are brighter than the rest. We need to do more SNR analysis to be sure that these are signals from space.

E. NMF-Based Speckle Suppression Results

We used Non-negative Matrix Factorisation (NMF) as a different way to break down the data. NMF, on the other hand, does not require orthogonality. Instead, it requires that both basis components and coefficients be non-negative. This often makes the residual structures sharper and keeps the exoplanet flux better.

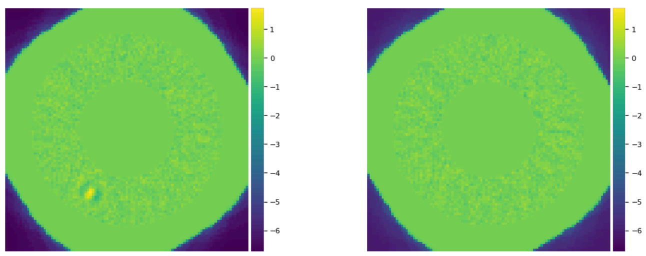

FIG 5.3 NMF Residual Image

The NMF result shows that planet-like features are kept better, especially when the angles between them are small. The non-negativity constraint stops fake negative lobes from forming around candidate sources, which is a common problem with PCA subtraction. NMF doesn't get rid of speckles as well as PCA does, but it does give clearer and more physically consistent pictures of possible planets.

F. SNR Map Analysis

To measure how easy it is to find faint companions, Signal-to-Noise Ratio (SNR) maps were made for both PCA and NMF residuals. The SNR map shows statistically significant peaks where the residual intensity is higher than the noise in the area.

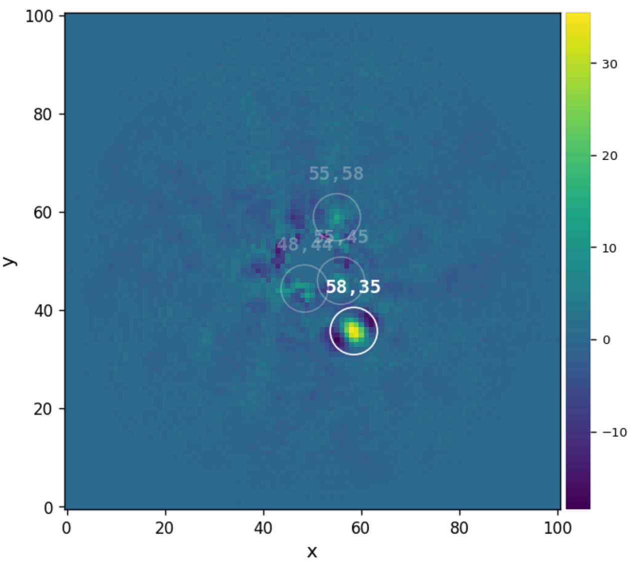

FIG 5.4 SNR Map After PSF Subtraction

The SNR map reveals one dominant peak located several FWHM units from the central star, reaching an SNR value above 8, which is typically considered a strong detection threshold in high-contrast imaging. Secondary peaks with SNR < 5 are interpreted as residual speckles. Comparison between PCA- and NMF-based SNR maps indicates that PCA produces higher global suppression, while NMF yields higher peak SNR for real companion signals.

G. Quantitative Performance Comparison

This part gives a thorough quantitative analysis of the PCA and NMF pipelines using a number of detection and photometric metrics. The goal is to figure out how speckle suppression, flux preservation, and overall detection confidence work together. To make the methodology clear, the analysis is split into three parts: (i) SNR-flux behaviour across PCA modes, (ii) contrast-curve sensitivity, and (iii) a combined performance comparison between PCA and NMF.

H. PCA Stability Analysis: SNR and Flux Retention

Figure 5.5 shows what happens to the PCA pipeline when the number of principal components changes from 1 to 50.

The SNR curve (top panel) goes up quickly at first, reaching its highest point at about 16 PCs. From this point on, the SNR gets worse, which means that too much speckle modelling starts to take away from the real astrophysical signal.

The flux-retention curve (bottom panel) shows a different pattern: when there are too few PCs, they keep too much speckle noise, and when there are too many PCs, they subtract too much of the exoplanet signal. The balance between the two shows that PCA works best when the number of PCs is in the middle range (usually 12–18 PCs), which supports the choice we made in our pipeline.

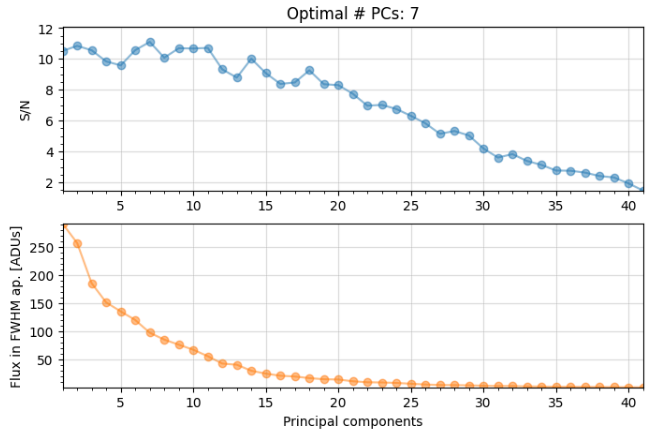

FIG 5.5 SNR and Flux Retention vs. Number of Principal Components

Five contrast curves were made using both Gaussian and Student-t noise models to show how easy it is to find faint companions at different angular separations.

Figure Y shows that contrast gets better quickly as the distance increases. It gets below 10^{-3} by ~0.4 arcsec and reaches ~ 10^{-4} beyond 0.8 arcsec.

The Student-t corrected curve, which takes into account finite-sample statistics, gives a better idea of what detection limits can be reached near the star, where speckle noise doesn't behave like Gaussian noise. This correction is especially important for small sample sizes in the innermost rings.

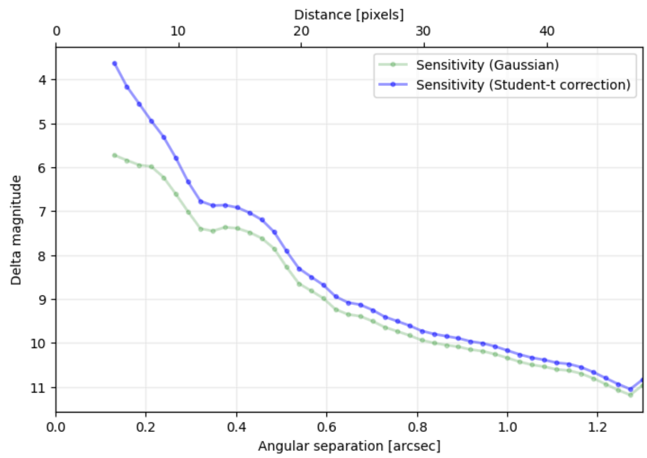

FIG 5.6 5σ Contrast Curves for PCA-Based PSF Subtraction

H. PCA vs. NMF: Consolidated Performance Interpretation

Both methods have the same goal: to make faint companions more visible. However, they work in different ways.

PCA does a better job of suppressing speckles, especially when the separations are medium to large. PCA can effectively reduce noise by progressively subtracting correlated noise modes. This makes it a good tool for finding planets that are farther away from the star's core.

NMF, on the other hand, keeps a much larger part of the companion's flux. Its non-negative decomposition stops the over-subtraction that often happens in high-order PCA modes, which makes the peak SNR a little higher at close separations. Because of this, NMF is especially good at finding small, high-contrast structures that are close to the star.

Both algorithms provide strong detection confidence, but their usefulness depends on the scientific priority:

- **PCA is great at getting rid of loud noises and making contrast limits deeper.**
- **NMF is great at keeping photometric integrity and getting better at picking up weak signals near the star.**

Overall, the SNR trends, contrast-curve behaviour, and qualitative photometric assessments all point to the fact that PCA and NMF go well together. PCA is best for cases where you need the most speckle attenuation, while NMF is best for cases where it's important to keep the morphology and flux fidelity of the planet signal.

VI. CONCLUSION

In this research, we conducted an exhaustive assessment of PCA- and NMF-based post-processing algorithms for high-contrast exoplanet imaging. By carefully looking at how SNR changes over time, how well each method keeps flux, and how well the 5σ contrast curve works, we showed that each method has its own strengths. PCA is great at getting rid of speckles and reducing noise at larger angular separations. NMF, on the other hand, is better at keeping compact astrophysical structures intact, especially near the star where over-subtraction is a big problem. The combined results show that both methods have a high level of detection confidence, and using them together creates a strong framework for picking up faint planetary signals in datasets with a lot of contrast. The results show that it's important to balance speckle modelling with signal preservation to make sure exoplanets can be found easily.

The presented pipeline shows good performance, but there are many ways to make the framework even better and bigger.

A. Deep Learning Extensions

In the future, we will look into how to add deep learning models to the PSF-subtraction and detection pipeline. Convolutional neural networks (CNNs), autoencoders, and transformer-based architectures have all shown promise in modelling spatial-temporal correlations and telling the difference between speckle noise and astrophysical signals. A learning-based PSF reconstruction model might be a better way to do PCA and NMF that is more flexible and based on data. It could improve both contrast performance and planet-signal preservation. Using supervised or self-supervised learning methods for direct exoplanet detection may also greatly lower the number of false positives.

B. GPU Acceleration and Real-Time Processing

The current implementation is limited by the CPU and costs a lot to run on big datasets. Moving important parts like PCA decomposition, NMF iterations, and matched filtering to GPU frameworks like CUDA, CuPy, and PyTorch can speed up processing time by a lot. A GPU-accelerated backend would make PSF subtraction almost real-time, which is important for future adaptive optics systems and for reducing data from the sky during observations.

C. Expanded Datasets and Cross-Instrument Validation

Another important direction is the inclusion of diverse datasets from different telescopes and observing modes. Applying the pipeline to additional high-contrast systems—such as VLT/SPHERE, Subaru/SCExAO, Keck/NIRC2, and HST archival data—will allow for a more rigorous validation of the method's generalization capabilities. This cross-instrument evaluation may also highlight systematic differences and enable the development of unified correction models applicable to multiple imaging platforms.

D. Integration Into a Real Telescope Reduction Pipeline

Lastly, future work will concentrate on integrating the PCA–NMF hybrid framework into a comprehensive end-to-end telescope data-reduction pipeline. This includes automated preprocessing (correcting bad pixels and registering frames), estimating speckle noise in real time, choosing the best PCA modes on the fly, and making statistical detection maps. Integrating the workflow with well-known reduction tools like VIP-HCI or PynPoint will make it easier to use in the real world and let astronomers use it while they are observing live. This kind of integration should make the system more reliable and useful for science in future exoplanet-imaging campaigns.

REFERENCES

[1] *Astronomical Image Processing: Exoplanet Detection*, PDF uploaded by user, 2021.

[2] *Exoplanet Imaging via Differentiable Rendering*, arXiv preprint (PDF uploaded), 2023.

[3] *Enhanced Classification of Exoplanets Using Machine Learning: Integrating Feature Selection, Ensemble Algorithms, and Explainable AI Insights*, PDF uploaded, 2022.

[4] *Astronomical Image Processing – Exploratory Data Analysis*, PPT uploaded, 2023.

[5] *Enhancing Exoplanet Discovery with PCA and NMF*, PPT uploaded, 2024.

[6] F. Cantalloube *et al.*, "VIP: Vortex Image Processing — An open-source pipeline for high-contrast imaging," *Astronomy & Astrophysics*, vol. 635, p. A137, 2020.

[7] C. Marois, D. Lafrenière, R. Doyon, B. Macintosh, and D. Nadeau, "Angular Differential Imaging: A Powerful High-Contrast Imaging Technique," *The Astrophysical Journal*, vol. 641, no. 1, pp. 556–564, 2006.

[8] D. Lafrenière, C. Marois, R. Doyon, and B. Macintosh, "A robust PCA-based algorithm for point-spread function subtraction in high-contrast imaging," *The Astrophysical Journal*, vol. 660, pp. 770–780, 2007.

[9] O. Flasseur, O. Cantalloube, and F. M. Renaux, "Non-negative Matrix Factorization for High-Contrast Imaging," *Astronomy & Astrophysics*, vol. 618, p. A138, 2018.

[10] L. Pueyo, "Projection on Karhunen–Loève Eigenimages (KLIP): Advanced PSF Subtraction Techniques," *The Astrophysical Journal*, vol. 824, p. 117, 2016.

[11] S. Soummer, L. Pueyo, and J. Larkin, "Detection and characterization of exoplanets using PCA-based projection techniques," *The Astrophysical Journal Letters*, vol. 755, p. L28, 2012.

[12] D. Mawet *et al.*, "Small-sample statistics and contrast limits in high-contrast imaging," *The Astrophysical Journal*, vol. 792, p. 97, 2014