Locating Hidden Exoplanets in ALMA Data Using Machine Learning

J. P. Terry[1,2] , C. Hall[1,2] , S. Abreau[3,4] , and S. Gleyzer[5]
[1] Department of Physics and Astronomy, The University of Georgia, Athens, GA 30602, USA; jpterry@uga.edu
[2] Center for Simulational Physics, The University of Georgia, Athens, GA 30602, USA
[3] Department of Medicine, Division of Cardiology, University of California, San Francisco, CA 94143, USA
[4] Cardiovascular Research Institute, San Francisco, CA 94158, USA
[5] Department of Physics and Astronomy, The University of Alabama, Tuscaloosa, AL 35487, USA
Received 2022 August 2; revised 2022 November 16; accepted 2022 November 17; published 2022 December 23

Abstract

Exoplanets in protoplanetary disks cause localized deviations from Keplerian velocity in channel maps of molecular line emission. Current methods of characterizing these deviations are time consuming, and there is no unified standard approach. We demonstrate that machine learning can quickly and accurately detect the presence of planets. We train our model on synthetic images generated from simulations and apply it to real observations to identify forming planets in real systems. Machine-learning methods, based on computer vision, are not only capable of correctly identifying the presence of one or more planets, but they can also correctly constrain the location of those planets.

Unified Astronomy Thesaurus concepts: Exoplanet astronomy (486); Exoplanet detection methods (489); Hydrodynamical simulations (767); Protoplanetary disks (1300)

1. Introduction

In recent years, the kinematic analysis of protoplanetary disks has proven important in demonstrating that ring- and gap-like features are indeed due to the presence of unseen planets (Pinte et al. 2018; Teague et al. 2018). The spiral wake from planets causes localized deviations from Keplerian motion known as "kinks." Other mechanisms, such as gravitational instability (Hall et al. 2020), vertical shear instability (Barraza-Alfaro et al. 2021), and magnetorotational instability (Simon et al. 2015), may leave similar kinematic imprints.

Recent kinematic observations of the HD 97048 system revealed a localized velocity deviation signaling the presence of a forming exoplanet (Pinte et al. 2019) too faint to be observed directly at ∼millimeter wavelengths. To estimate the location of the forming exoplanet, several computationally expensive hydrodynamics simulations were run, and synthetic observations were generated and compared "by eye" to real observations to obtain the best fit.

Machine learning offers an alternative route. The use of machine learning in astronomy has exploded in the last decade (Jo & Kim 2019; Alexander et al. 2020; Möller & de Boissière 2020). The study of protoplanetary disks largely relies on images, so machine-learning techniques for computer vision are perfectly suited for application to these kinematics observations. Computer vision is a field of AI that focuses on image and video analyses. From simple transcription of handwritten digits (Cireşan et al. 2011) to locating and classifying objects pixel by pixel (Minaee et al. 2022), computer vision has a wide range of uses. The introduction of advanced architectures, such as residual (He et al. 2016) and inception (Szegedy et al. 2015) blocks, along with qualitatively new algorithms, such as the vision transformer (Vaswani et al.

2017; Dosovitskiy et al. 2020), has led to computer vision models exceeding human experts at times (Zhou et al. 2021).

Given the applicability and strong performance of computer vision techniques, we have applied them to the kinematic analysis of protoplanetary disks.

Previous work has successfully applied deep learning to the problem of inferring properties of planets within protoplanetary disks (Auddy et al. 2021, 2022) using models trained on simulated data, but these efforts do not use kinematic information and instead rely on gaps as the signatures. We demonstrate that a trained network can detect the presence of planets in large data sets much faster than a human, recover all previously identified planets, and possibly identify ones that have been missed in previous analyses. Inferring other properties, such as the number of planets and their masses, may be within the reach of machine-learning techniques as well, which we leave to future work. This paper presents a proof-of-concept application of a relatively simple machine algorithm applied to protoplanetary accretion disks.

The paper is arranged as follows: Section 2 describes the training data and the details of the machine-learning models and their training. Section 3 presents the results, interpretations, suggestions for future work, and limitations. Section 4 gives our conclusions.

2. Methods

Training a model to predict the presence of a planet requires a large amount of data spanning a large parameter space. We run 1000 simulations in this work. The parameter space is sampled using a Latin Hypercube (LHC; McKay et al. 1979), which accounts for previously sampled points to evenly distribute values across the entire parameter space. We sample values from observed ranges inferred from disk surveys (Andrews et al. 2009, 2018; Huang et al. 2018) and widely accepted simulational parameters (Pinte et al. 2018, 2019, 2020).

Table 1 shows the variables sampled and their minimum and maximum values. Approximately one-third of the simulations

<div align="center">

Table 1
Simulation Parameter Ranges for the LHC Sampling and the Stage at Which They Were Applied

</div>

Parameter	Variable	Minimum Value	Maximum Value	Use
Number of planets	N	0	4	PHANTOM, MCFOST
Disk Inner Radius (au)	R_{out}	150	250	PHANTOM, MCFOST
Disk Outer Radius	R_{in}	$0.1R_{out}$	$0.1R_{out}$	PHANTOM, MCFOST
Planetary Mass ($M_\oplus$)	M_{planet}	5	120	PHANTOM, MCFOST
Stellar Mass ($M_\odot$)	$M_\star$	0.5	2.0	PHANTOM, MCFOST
Planet Semimajor Axis (au)	a	$R_{in} + 0.2R_{out}$	$R_{in} + 0.8R_{out}$	PHANTOM
Mass Ratio	$M_{disk} / M_\star$	5×10^{-4}	10^{-2}	PHANTOM
Disk Viscosity	α_{SS}	5×10^{-4}	2.5×10^{-3}	PHANTOM
Surface Density Profile ($\Sigma = \Sigma_0 R^{-p}$)	p	0.1	1.0	PHANTOM
Sound Speed Profile ($c_s = c_{s,0} R^{-q}$)	q	0.1	0.75	PHANTOM
Stellar Age (Myr)	t	0.5	5.0	MCFOST
Inclination (degrees)	i	10	80	MCFOST
Azimuth (degrees)	ϕ	0	360	MCFOST
Distance (pc)	d	100	200	MCFOST
Spectral Resolution (km s^{-1})	δv	0.01	0.1	Synthetic Observations
Spatial Resolution (arcseconds)	$''$	0.025	0.25	Synthetic Observations
rms Noise (%)	δ	0.05	20	Synthetic Observations

have no planets in them to address any class imbalance (i.e., not having an equal number of "at least one planet" class and "no planet" class simulations) while ensuring diversity in the number of planets. After sampling the stellar mass for use in the simulation, stellar temperatures and radii, both of which are used in the radiative transfer calculations, are determined using isochrones calculated by Siess et al. (2000).

2.1. Hydrodynamical Simulations

We run 1000 3D smoothed particle hydrodynamics (SPH) simulations using the PHANTOM (Price et al. 2018) code. System properties, such as stellar mass, disk properties, observational resolution, etc., were LHC-sampled from Table 1. A significant percentage (25%) of these simulations were withheld for testing. Each simulation uses 10^6 SPH particles and evolves for 100 orbits. A given planet's accretion radius is set to one-fourth of its Hill radius. The surface density profile is set to $\Sigma \propto R^{-p}$, and the sound speed is set to $c_s \propto R^{-q}$.

2.2. Velocity Channel Maps

The simulation results are used to create velocity channel maps in only ^{13}CO to keep the size of the parameter space manageable, although ^{12}CO and C^{18}O are also reasonable choices. The channel maps are created using the MCFOST radiative transfer code (Pinte et al. 2006, 2009). For each simulation, multiple points between 10 and 100 orbits are chosen randomly. Since one-third of the simulations have no planets, six snapshots are chosen from each of those simulations while only two snapshots are chosen for simulations with planets. This eliminates the class imbalance. From each snapshot, a velocity channel cube is made using either the ^{13}CO J $2 \to 1$ (220.4 GHz) or $3 \to 2$ (330.6 GHz) transition, which is randomly sampled. These transitions are used because they were also used to identify the planets in HD 163296 and HD 97048 (Pinte et al. 2018, 2019). The final result is a velocity cube, where each velocity channel is an image with 600×600 pixels at a resolution of 1 au pixel^{-1}. This velocity cube is used as the model input. Channel maps are calculated using 10^8 photon packets. The dust-to-gas ratio is 1:100, and dust is composed of carbon and silicates (Draine & Lee 1984).

Stellar radii and temperatures are calculated with isochrones given by Siess et al. (2000) using the LHC-sampled masses and ages. A spherical beam is used.

2.3. Simulated Observations

The velocity channel maps are convolved spatially and spectrally, and noise is added by LHC sampling from Table 1. rms noise is added independently for each pixel, which more faithfully replicates observational effects such as hot pixels. This method results in an average rms of ~10% for each image.
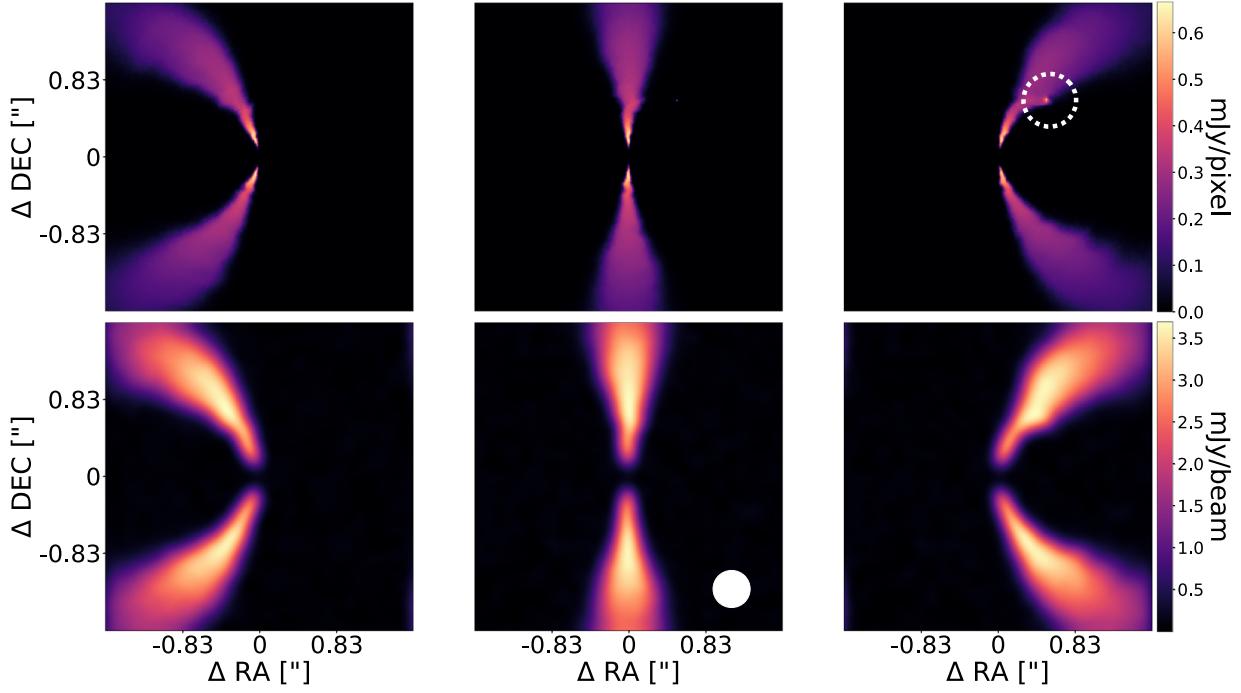
Except for the specific values, this process largely follows that done in Terry et al. (2022). Figures 1 and 2 show the results of convolving selected velocity channels for a disk with and without a planet, respectively. Figure 1 shows that, while the perturbation is strongest in the velocity channel that covers the planet (rightmost column), the perturbation is not necessarily completely localized and can be visible in distant regions of the disk. This is an expected behavior. The deviation is strongest at the location of the planet, but the spiral wake leaves further imprints as it disperses throughout the disk.

By itself, a single velocity channel is poorly suited for our method. Only a few channels display kinks, and channels with kinks can neighbor the exact channel containing the planet. We therefore stack all C velocity channels into a single image of dimension ($H \times W \times C$), where H and W are the height and width of the image (600×600 pixels for these purposes).
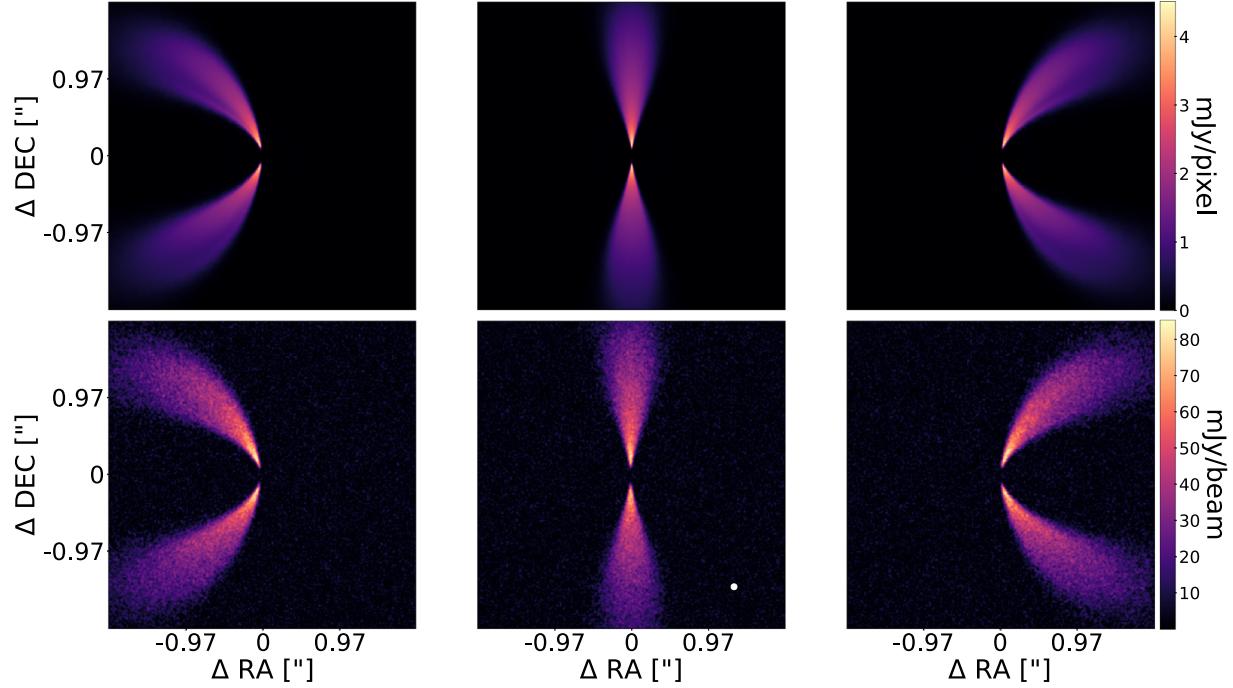
The number of velocity channels in an observation varies but mostly falls in the range ~40–80 channels that cover the disk (Andrews et al. 2018; Öberg et al. 2021). We therefore use $C = 47$, 61, and 75 channels for each observation. The number of channels must be odd to cover the disk symmetrically and still retain the systemic ($\Delta v = 0$) channel. The input velocity of the simulated channels ranges from $|\Delta v| = |v_{systemic} - v| \leqslant 2.5$ km s^{-1} to fully cover the disk.

2.4. Machine-learning Models

The input for all models is a ($600 \times 600 \times C$) image, and their output is a two-component vector with softmax activation (Goodfellow et al. 2016). The softmax function maps the output vector such that the sum of the components equals one. Each component can be interpreted as the predicted probability

Figure 1. Example raw (top row) and convolved, noisy (bottom row) channel maps in a disk with a planet present. The planet (circled in white) is visible as a kink in the right column. The opposite velocity channel is shown in the left column, and the systemic channel is shown in the middle column. The beam size is indicated in the bottom middle image (solid white circle). This disk is one of the smallest and farthest simulated and is observed with some of the worst spatial resolutions, which is why the beam is so large.

Figure 2. Example raw (top row) and convolved, noisy (bottom row) channel maps in a disk without a planet present. The beam size is indicated in the bottom middle image.

that the input belongs to the corresponding class (i.e., "contains at least one planet" class or "does not contain a planet" class). An Adam optimizer (Kingma & Ba 2014) is used to add momentum, which remembers previous gradients, and gradient scaling to a simple gradient descent algorithm. This helps escape local minima and improves the speed and range of convergence. The cross-entropy loss (Hinton et al. 1995)

penalizes confident predictions that are incorrect. Training stops when there is no improvement in the validation loss after eight complete epochs, i.e., eight sweeps through the training data.

Multiple architectures are used to compare performance on two tasks: predicting if the system hosts at least one planet and determining its azimuthal location.

We use two different models based on PyTorch (Paszke et al. 2019) implementations: EfficientNetV2 (Tan & Le 2021) and RegNet (Xu et al. 2022). Neither model uses the default hyperparameters or pretrained weights.

EfficientNetV2 is an updated version of EfficientNet (Tan & Le 2019) that maximizes performance and minimizes the number of trainable parameters and training time. RegNet, on the other hand, seeks to maximize performance by adding variants of convolutional recurrent neural networks to ResNet (He et al. 2016). This can extract greater spatial information at the cost of significantly more trainable parameters.

We perform Bayesian hyperparameter tuning using WANDB (Biewald 2020) to find separate sets of hyperparameters that minimize the validation loss for $C = 47$, 61, and 75. The default parameters for these sweeps are based on default versions "EfficientNetV2 S" and "RegNetY 16GF."

Each architecture is evaluated for accuracy and area under the receiver operating characteristic curve (AUC; Hanley & McNeil 1982). Accuracy in this context is defined as whether a model correctly predicted that a system with planet(s) has at least one planet and similarly for systems with no planets. We choose AUC because it gives a robust measure of the performance of the model across different decision thresholds, i.e., the softmax threshold required to claim the presence of a planet. Substantial confidence is required to claim a planet detection, but setting the decision threshold too high can lead to false negatives. False positives may be more acceptable because further analysis can overrule them, whereas a false negative may lead to no further analysis and result in the planet not being detected. The optimum threshold for this boundary is open for discussion by the community. We show multiple thresholds (50%, 75%, 90%, and 95%) in Figures 3 and 4 to demonstrate how the threshold for claiming the presence of at least one planet affects the results.

While some protoplanetary disks exhibit gaps in the continuum that are highly suggestive of a planet, this does not help one find the azimuthal location. Planet-induced non-Keplerian kinks tend to be strongest near the planet, so our model must determine if and where the kink is in our data in order to give information on the azimuthal location.

One simple approach to this is to look at the activations inside the model itself. The activation strengths highlight regions that the model finds particularly informative for classification. EfficientNetV2 uses SiLU activations (Elfwing et al. 2018), and RegNet uses ReLU (Glorot et al. 2011). We use mean-subtracted activations to achieve this. For a given pixel with an activation value x, the mean-subtracted activation x' is defined as $x' = |x - \langle x \rangle|$, where $\langle x \rangle$ is the mean activation over the entire image. Activations occur most strongly at the location of the strongest deviation from Keplerianity, i.e., the kink.

3. Results and Discussion

3.1. Machine-learning Models

We train six models in total: an EfficientNetV2 and a RegNet, each for 47, 61, and 75 velocity channels. Using a typical decision threshold of 0.5, each model, other than the 47-channel RegNet, has an accuracy of at least 93%. When this decision threshold is increased to 0.95, four models have an accuracy of at least 92%. Five models have an AUC of at least 0.98. This confirms that our models have learned to correctly detect the presence of planets using kinematic signatures.

Table 2 shows the number of parameters as well as relevant performance measures for each final model. All EfficientNetV2 models converge to a similar number of trainable parameters, whereas the number of parameters for the RegNets increases sharply with the number of channels. RegNet is a larger and more complicated architecture, so this is expected. The performance of the RegNet models varies widely compared to the EfficientNetV2 models. The accuracy of EfficientNetV2 models varies by at most 8%, whereas RegNet varies by as much as 31%. By all performance metrics, while the average EfficientNetV2 outperforms the average RegNet, the 61-channel RegNet outperforms all other models.
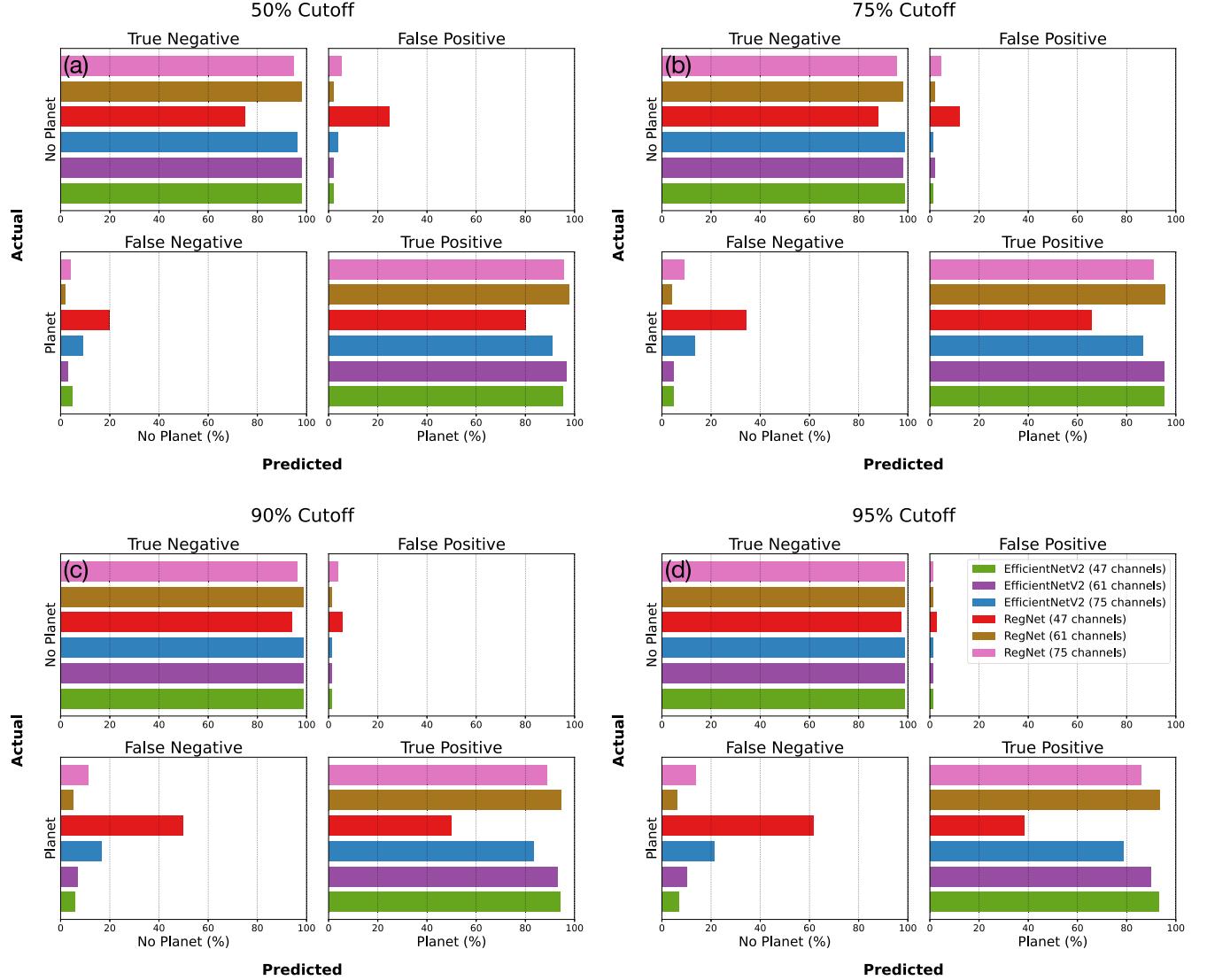
Figure 3 shows the confusion matrix, which describes the distribution of correct and incorrect predictions, for all models applied to the withheld test set. Panel (a) shows the results using a 50% decision threshold, i.e., a planet is predicted if the confidence is over 50%. Panels (b), (c), and (d) show the same for cutoffs of 75%, 90%, and 95%. Figure 4 gives several relevant metrics for model performance and all ROC curves. Claiming a planet is present should come with high confidence, so a 50% decision threshold is not necessarily what should be used in practice. It is encouraging that, for all other models than the 47-channel RegNet, there are no qualitative changes in the results when a higher decision threshold is used; the main effect is simply increasing the number of false negatives and decreasing the number of false positives.

If there are multiple planets in a disk, then multiple kinks may be present. It is therefore prudent to check that the quality of the predictions does not depend on the number of planets in the disk. We check the possibility of this dependence by calculating the accuracy for each model for disks with a given number of planets. Figure 5 gives the results of these calculations. The p-values from Pearson correlation tests shown in Figure 6 demonstrate that only the 47-channel RegNet using a decision threshold of 50% has a statistically significant relationship between accuracy and the number of planets ($p < 0.05$). While the 47-channel RegNet using a decision threshold of 95% does not have a statistically significant relationship, it is clear that the performance is worse when there are planets, i.e., there are many false negatives. This is in line with this model's relatively low AUC and Figure 3. The performance of all other models does not differ significantly with the number of planets.
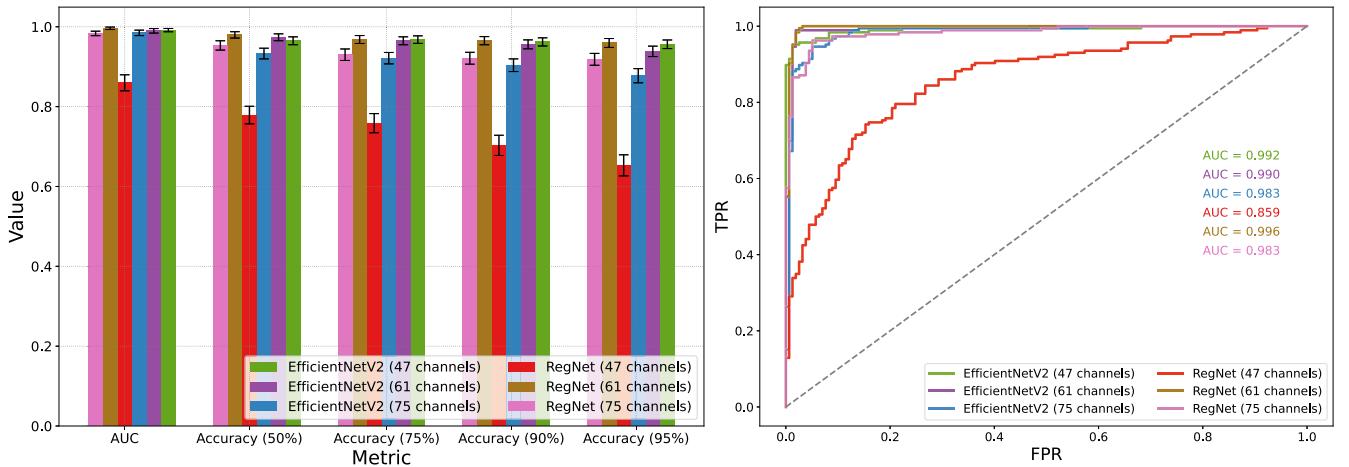
3.2. Simulation Activations

Many disks do not show obvious signs of planets, either in continuum or line emission. Inspecting the activations of the disks helps overcome this. Figure 7 shows the result from a disk that is difficult to classify by eye. The models predict the presence of an exoplanet with high confidence, activating on a subset of channels that are deemed important by the model. This gives information on the azimuthal location, and the radial location may be inferred by either the location with the strongest kink or coincident gaps in continuum images.

Panels (c) and (d) in Figure 7 show the raw and convolved velocity channels, respectively. If one were simply inspecting the velocity channels by eye, it would be trivially easy to miss the kink even in the unconvolved data. The convolved, noisy data show essentially no sign of the planets. Despite lacking any clear signatures that can be observed visually, the models

Figure 3. Confusion matrix from withheld test set for all models. The *x*-axis is the percentage of disks in that class that were predicted accordingly. (a) Using a 50% decision threshold. (b) Using a 75% decision threshold. (c) Using a 90% decision threshold. (d) Using a 95% decision threshold. The top left and bottom right blocks of each panel show the counts of true negatives and true positives, respectively. The top right and bottom left blocks of a panel show the false positives and false negatives, respectively. A perfectly accurate model will have no entries in the top right and bottom left.
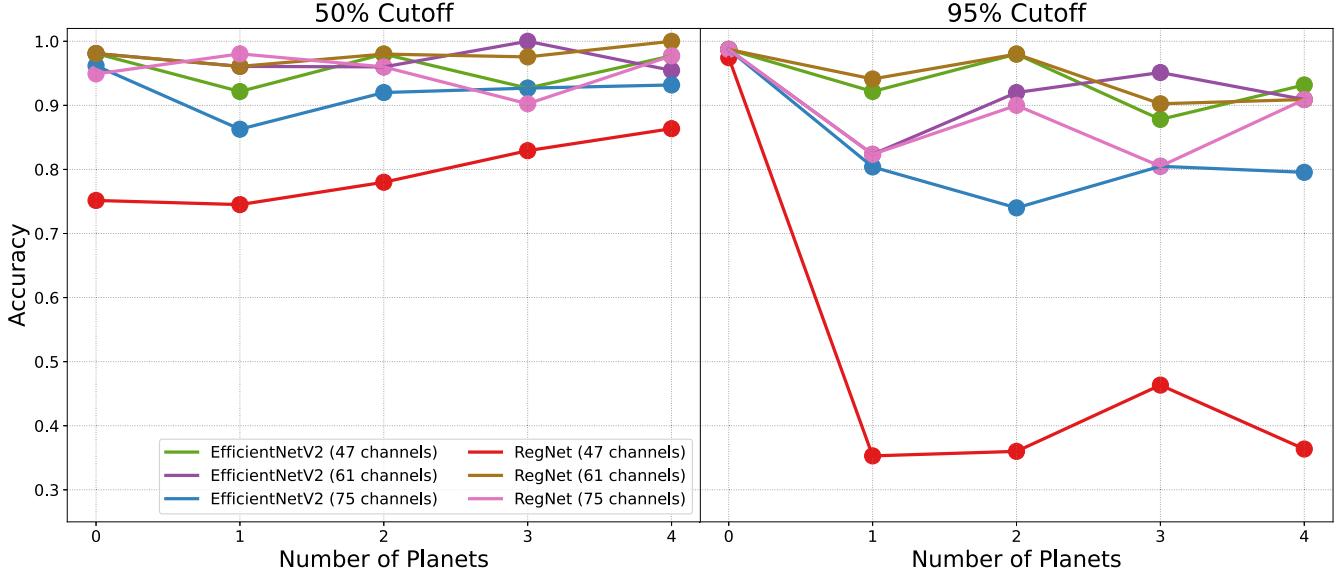
Figure 4. Left: various metrics calculated from the withheld test set for all models. Right: corresponding receiver operating characteristic (ROC) curves. Error bars show 95% confidence intervals that are calculated by bootstrapping each metric 1000 times using random selections of 80% of the test data. The percentages next to the accuracy labels denote the decision threshold.
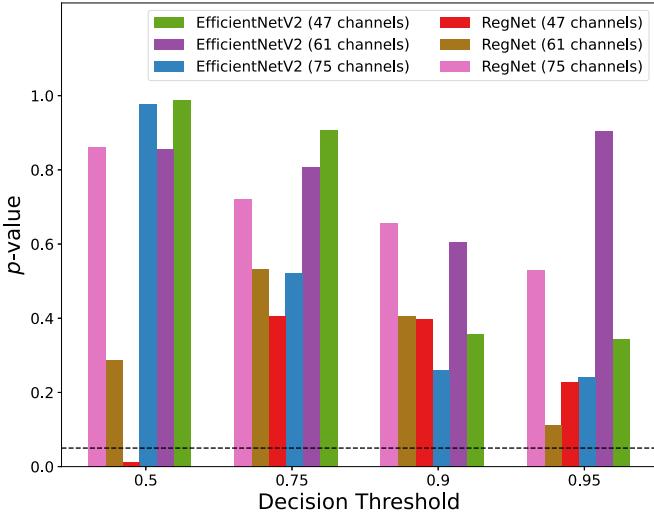
Table 2
Final Model Descriptions and Metrics

Model	Parameters (millions)	Accuracy (50%) (%)	Accuracy (95%) (%)	AUC
EfficientNetV2: 47 Channels	20.2	97 ± 0.5	96 ± 0.5	0.99 ± 0.002
EfficientNetV2: 61 Channels	20.2	97 ± 0.5	94 ± 0.7	0.99 ± 0.003
EfficientNetV2: 75 Channels	20.2	93 ± 0.7	88 ± 0.9	0.98 ± 0.003
RegNet: 47 Channels	51.0	78 ± 1.1	65 ± 1.3	0.86 ± 0.010
RegNet: 61 Channels	62.8	98 ± 0.4	96 ± 0.6	$>0.99 \pm 0.001$
RegNet: 75 Channels	114	95 ± 0.6	92 ± 0.7	0.98 ± 0.003

Note. 95% confidence intervals are calculated by bootstrapping each metric 1000 times using random selections of 80% of the test data. The percentages next to the accuracy labels denote the decision threshold.

Figure 5. Accuracy of all models for a given number of planets. Left: 50% decision threshold. Right: 95% decision threshold.


Figure 6. p-values from testing the correlation between accuracy and the number of planets at different decision thresholds. The dashed black line is $p = 0.05$, which is the standard cutoff for statistical significance.

correctly classify the disks and locate the planets. This demonstrates that our models are able to accurately find planets that the by-eye method may fail to find.
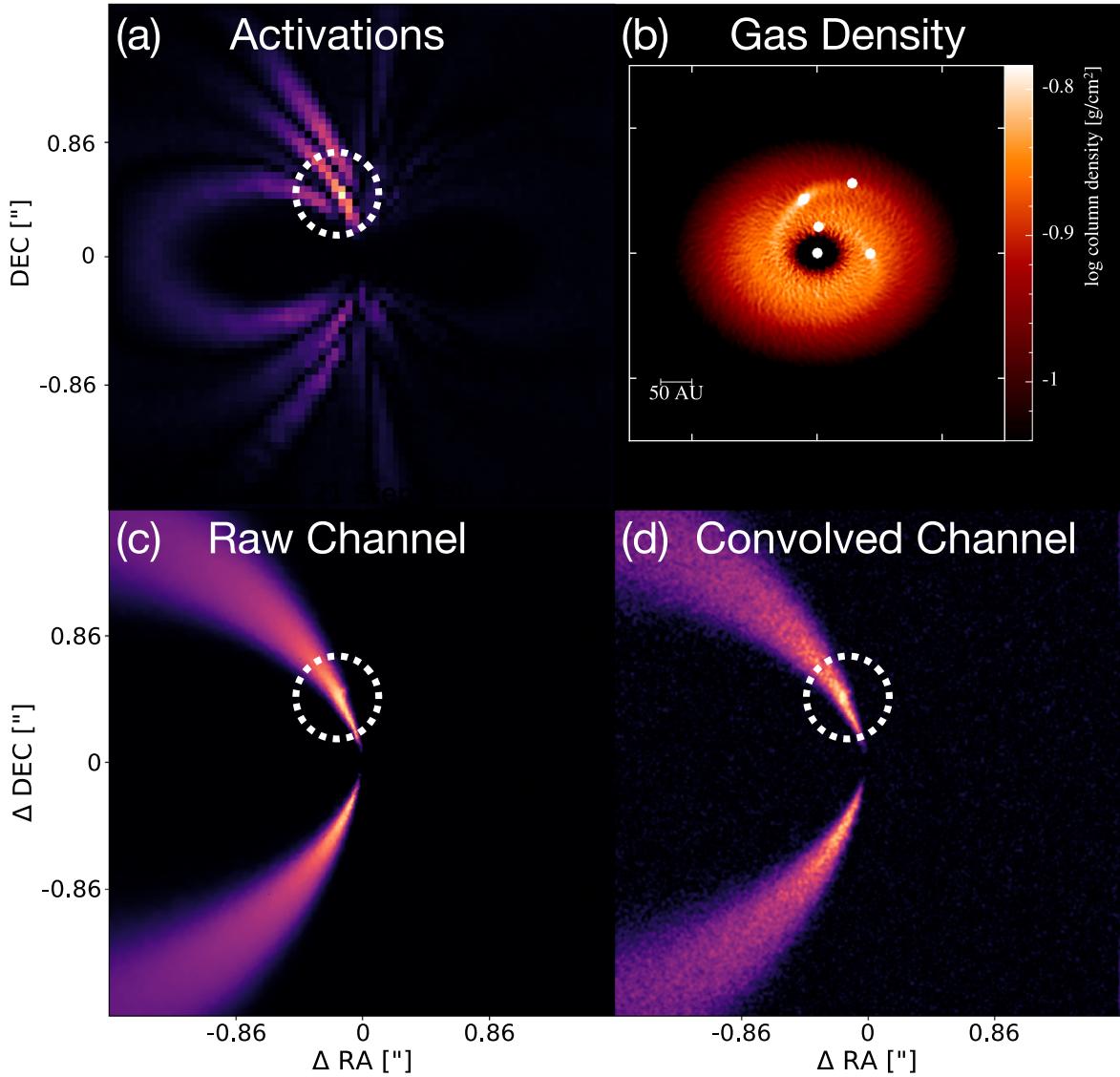
Figure 8 shows a typical activation for a disk with no planet. The entire disk is activated equally, indicating that there is no localized region that exhibits behaviors indicative of a planet.

3.3. Application to Real Observations
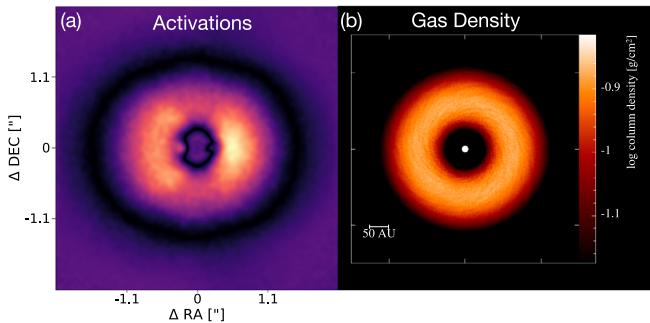
Next, we demonstrate the effectiveness of our model by applying it to real telescope observations. Kinematic detections of planets have been demonstrated for many disks (Pinte et al. 2020), but here we focus on the HD 163296 and HD 97048 disks as a proof of concept (Pinte et al. 2018, 2019). Using the same data as Pinte et al. (2018, 2019), we demonstrate that our models replicate the prediction and estimated location of the forming exoplanet. Figure 9(a) shows the HD 97048 in continuum, and Figure 9(b) shows the velocity channel with the planet's signature circled.

Table 3 shows the softmax values for the observations. For HD 97048, all RegNet models correctly predict a planet with >95% confidence. Two EfficientNetV2 models correctly predict a planet with $\geqslant 85\%$ confidence. Figure 9 shows that the velocity channels that contain the planet are strongly activated. The activation in Figure 9(d) is strong on the planet itself rather than simply the entire velocity channel.

The results for HD 163296 are not as confident, but the two best-performing models predict the presence of at least one planet with over 99% confidence.

Figure 7. (a) Mean-subtracted activations using the 47-channel EfficientNetV2. The innermost planet has $M = 0.28M_J$, the second planet has $M = 0.20M_J$, the third planet has $M = 0.36M_J$, and the outermost planet has $M = 0.02M_J$. (b) Disk column density after an azimuthal rotation of $\phi = 131°$ and inclined by $i = 36°$. (c) Unconvolved velocity channel with the most obvious kink. (d) Corresponding convolved velocity channel. Based on the kink, the approximate location of the planet causing the most significant kink is in the white circle. The planets were correctly predicted with a confidence of $>99\%$.
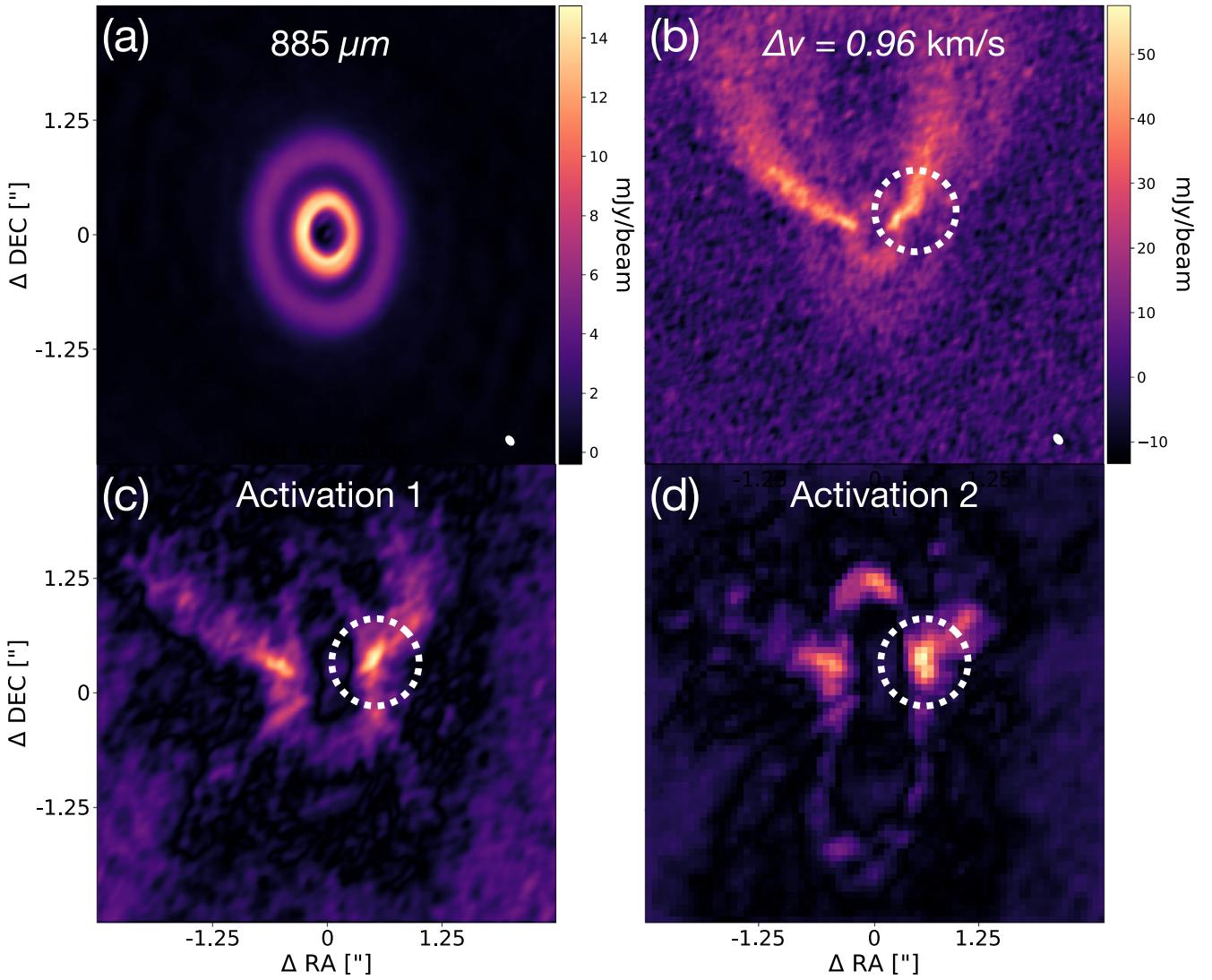
Figure 8. (a) Typical example of mean-subtracted activations for a disk with no planet. (b) Disk column density after an azimuthal rotation of $\phi = 153°$ and inclined by $i = 13°$.

Both classes of the models perform better when there are more input channels, a trend not seen for HD 97048. A possible explanation is the fact that there are about twice as many velocity channels that cover HD 163296 when compared to HD 97048. While HD 97048 has roughly 75 channels covering the disk, HD 163296 has roughly 150. All models applied to HD 163296 must therefore combine significantly more velocity channels, which may decrease their ability to extract information from the images. The activation structure tends to be less informative, perhaps for the similar reasons, so it is omitted. Despite this limitation, the high confidence of some of the models should encourage a human observer to seriously scrutinize the kinematic structure.

3.4. Limitations and Future Work

Our work comes with several limitations. All simulations were run with SPH, so we did not demonstrate that our methods would also be applicable to data from grid-based codes. However, there is typically a qualitative agreement between results from SPH- and grid-based methods (De Val-Borro et al. 2006; Agertz et al. 2007). A good direction for future work would be to train on simulation data from a variety

Figure 9. (a) HD 97048 at 885 μm. (b) $\Delta v = 0.96$ km s^{-1} channel. Bottom: mean-subtracted activations from two different layers for HD 97048 using the 47-channel RegNet. The estimated planet location is given inside the white circle. The beams for the observations are given in the bottom right corner of panels (a) and (b).

Table 3
Prediction Confidences for All Models Tested on HD 97048 and HD 163296

Model	HD 97048 Softmax	HD 163296 Softmax
EffecientNetV2: 47 Channels	85%	37%
EffecientNetV2: 61 Channels	89%	69%
EffecientNetV2: 75 Channels	64%	73%
RegNet: 47 Channels	>99%	74%
RegNet: 61 Channels	96%	>99%
RegNet: 75 Channels	>99%	>99%

of hydrodynamics codes and check that the same results can be obtained.

Any differences in performance of the models between the codes, e.g., if the models consistently performed better on SPH data, would be difficult to interpret, and the cause could be hard to identify. It is possible that the method of domain adaptation (Ben-David et al. 2010), which has already found success in astronomy (Vilalta et al. 2019; Alexander et al. 2021; Ćiprijanović et al. 2022), could be used to encourage the models to overcome any differences between data sets. This is an avenue that is ripe for exploration in future work.

In order to keep the parameter space manageable and demonstrate a proof of concept in this work, we therefore trained on data produced by only one hydrodynamics code (PHANTOM).

The absence of true image segmentation (Minaee et al. 2022) or object detection (Zhao et al. 2019) during training is another limitation that requires addressing in future work. While we present results in the form of activation structures that suggest the location(s) of the planet(s), these activations neither always perfectly coincide with the planets nor capture all planets. User analysis is required in order to pinpoint the radial (and sometimes azimuthal) location(s) of the planets even with the information given by the activation structures. The main utility of the presented models is therefore to give a user an indication as to whether a planet is present and motivate and direct the analysis of the velocity channels. Segmentation and object detection would address these shortcomings by precisely pinpointing the location(s) of the planet(s) without any necessary further analysis.

The assertion of a vertically isothermal disk is a simplifying assumption that is commonly used in simulations of proto-planetary disks (see, e.g., Dipierro et al. 2015; Pinte et al.

2019). It has been demonstrated that the dispersion relation and morphology of the planet's signature can change significantly if the disk is vertically stratified (Ogilvie & Lubow 2002; Juhász & Rosotti 2018).

However, the presence of any disturbance is more important than its specific morphology. There does exist the possibility that vertical stratification dampens the signatures to an extent that the models' performances appreciably decrease, but our results demonstrate that even extremely subtle signals can be found by the models. Regardless, a more realistic treatment of the thermal structure may be useful in future works.

4. Conclusions

Our results show that a machine-learning model trained on synthetic data can determine the presence and location of forming exoplanets in real telescope observations of proto-planetary accretion disks. We apply this model to the HD 97048 and HD 163296 systems. Multiple models predict the presence of a planet with greater than 99% confidence in both systems. The presence and location of the planet in HD 97048 are corroborated by activating strongly on the same location given by Pinte et al. (2019).

Five out of six models have greater than 90% accuracy when using a 50% decision threshold. When using a 95% decision threshold, four out of six models have an accuracy greater than 90%. Five models have an AUC of at least 0.98. While EfficientNetV2 and RegNet perform comparably on the synthetic test data, RegNet outperforms EfficientNetV2 on the real telescope observations.

This work is the first step in a series of works that use machine learning to automate the detection and analysis of objects within protoplanetary disks. Applying full segmentation and object-detection methods will further increase the ability of our models to locate planets and infer properties, such as planetary mass. We leave this work to the future. We envision final models that can be used easily for multiple purposes and be distributed freely throughout the community.

As Atacama Large Millimeter/submillimeter Array (ALMA) continues to deliver larger and larger disk survey data sets, and next-generation telescopes—such as JWST, ngVLA, and the Square Kilometre Array (SKA)—come online, the time/cost of analyzing more and more data will increase. Our proof-of-concept work demonstrates that even when high confidence is demanded, machine learning can automate this task and match and exceed human proficiency.

ORCID iDs

J. P. Terry https://orcid.org/0000-0002-8590-7271
C. Hall https://orcid.org/0000-0002-8138-0425
S. Abreau https://orcid.org/0000-0002-9426-3789
S. Gleyzer https://orcid.org/0000-0002-6222-8102

References

Agertz, O., Moore, B., Stadel, J., et al. 2007, MNRAS, 380, 963
Alexander, S., Gleyzer, S., McDonough, E., Toomey, M. W., & Usai, E. 2020, ApJ, 893, 15
Alexander, S., Gleyzer, S., Reddy, P., Tidball, M., & Toomey, M. W. 2021, arXiv:2112.12121
Andrews, S. M., Huang, J., Pérez, L. M., et al. 2018, ApJL, 869, L41
Andrews, S. M., Wilner, D. J., Hughes, A. M., Qi, C., & Dullemond, C. P. 2009, ApJ, 700, 1502
Auddy, S., Dey, R., Lin, M.-K., Carrera, D., & Simon, J. B. 2022, ApJ, 936, 93
Auddy, S., Dey, R., Lin, M.-K., & Hall, C. 2021, ApJ, 920, 3
Barraza-Alfaro, M., Flock, M., Marino, S., & Pérez, S. 2021, A&A, 653, A113
Ben-David, S., Blitzer, J., Crammer, K., et al. 2010, Mach. Learn., 79, 151
Biewald, L. 2020, Experiment Tracking with Weights and Biases https://www.wandb.com/
Ćiprijanović, A., Kafkes, D., Snyder, G., et al. 2022, MLS&T, 3, 035007
Cireşan, D. C., Meier, U., Masci, J., Gambardella, L. M., & Schmidhuber, J. 2011, arXiv:1102.0183
De Val-Borro, M., Edgar, R. G., Artymowicz, P., et al. 2006, MNRAS, 370, 529
Dipierro, G., Price, D., Laibe, G., et al. 2015, MNRAS Lett., 453, L73
Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. 2020, arXiv:2010.11929
Draine, B. T., & Lee, H. M. 1984, ApJ, 285, 89
Elfwing, S., Uchibe, E., & Doya, K. 2018, NN, 107, 3
Glorot, X., Bordes, A., & Bengio, Y. 2011, in Proc. 14th Int. Conf. Artif. Intell. Stat., 15, (PMLR), 315, https://proceedings.mlr.press/v15/glorot11a.html
Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (Cambridge, MA: MIT Press) http://www.deeplearningbook.org
Hall, C., Dong, R., Teague, R., et al. 2020, ApJ, 904, 148
Hanley, J. A., & McNeil, B. J. 1982, Radiology, 143, 29
He, K., Zhang, X., Ren, S., & Sun, J. 2016, in IEEE Conf. Comput. Vision Pattern Recognit. (CVPR) (Piscataway, NJ: IEEE), 770
Hinton, G. E., Dayan, P., Frey, B. J., & Neal, R. M. 1995, Sci, 268, 1158
Huang, J., Andrews, S. M., Dullemond, C. P., et al. 2018, ApJL, 869, L42
Jo, Y., & Kim, J.-h. 2019, MNRAS, 489, 3565
Juhász, A., & Rosotti, G. P. 2018, MNRAS Lett., 474, L32
Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
McKay, M. D., Beckman, R. J., & Conover, W. J. 1979, Technometrics, 21, 239, http://www.jstor.org/stable/1268522
Minaee, S., Boykov, Y., Porikli, F., et al. 2022, in IEEE Trans. on Pattern Anal. Mach. Intell., 44 (Piscataway, NJ: IEEE), 3523
Möller, A., & de Boissière, T. 2020, MNRAS, 491, 4277
Öberg, K. I., Guzmán, V. V., Walsh, C., et al. 2021, ApJS, 257, 1
Ogilvie, G. I., & Lubow, S. H. 2002, MNRAS, 330, 950
Paszke, A., Gross, S., Massa, F., et al. 2019, in Advances in Neural Information Processing Systems, 32, ed. H. Wallach et al. (Red Hook, NY: Curran Associates, Inc.) https://proceedings.neurips.cc/paper/2019/file/bdbca288fee7f92f2bfa9f7012727740-Paper.pdf
Pinte, C., Harries, T. J., Min, M., et al. 2009, A&A, 498, 967
Pinte, C., Ménard, F., Duchêne, G., & Bastien, P. 2006, A&A, 459, 797
Pinte, C., Price, D. J., Ménard, F., et al. 2018, ApJL, 860, L13
Pinte, C., Price, D. J., Ménard, F., et al. 2020, ApJL, 890, L9
Pinte, C., van der Plas, G., Ménard, F., et al. 2019, NatAs, 3, 1109
Price, D. J. 2007, PASA, 24, 159
Price, D. J., Wurster, J., Tricco, T. S., et al. 2018, PASA, 35, e031
Siess, L., Dufour, E., & Forestini, M. 2000, A&A, 358, 593
Simon, J. B., Hughes, A. M., Flaherty, K. M., Bai, X.-N., & Armitage, P. J. 2015, ApJ, 808, 180
Szegedy, C., Liu, W., Jia, Y., et al. 2015, in IEEE Conf. Comput. Vision Pattern Recognit. (CVPR) (Piscataway, NJ: IEEE), 1, https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html

Tan, M., & Le, Q. V. 2019, in Proc. 36th Int. Conf. Mach. Learn., 97, (PMLR), 6105, https://proceedings.mlr.press/v97/tan19a.html

Tan, M., & Le, Q. V. 2021, in Proc. 38th Int. Conf. Mach. Learn., 139, (PMLR), 10096, https://proceedings.mlr.press/v139/tan21a.html

Teague, R., Bae, J., Bergin, E. A., Birnstiel, T., & Foreman-Mackey, D. 2018, ApJL, 860, L12

Terry, J. P., Hall, C., Longarini, C., et al. 2022, MNRAS, 510, 1671

Vaswani, A., Shazeer, N., Parmar, N., et al. 2017, in Advances in Neural Information Processing Systems, 30 (Red Hook, NY: Curran Associates, Inc.) https://proceedings.neurips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

Vilalta, R., Gupta, K. D., Boumber, D., & Meskhi, M. M. 2019, PASP, 131, 108008

Xu, J., Pan, Y., Pan, X., et al. 2022, in IEEE Trans. on Neural Networks and Learn. Syst. (Piscataway, NJ: IEEE), 1, https://ieeexplore.ieee.org/document/9743274

Zhao, Z.-Q., Zheng, P., Xu, S.-t., & Wu, X. 2019, in IEEE Trans. on Neural Networks and Learn. Syst., 30 (Piscataway, NJ: IEEE), 3212

Zhou, W., Yang, Y., Yu, C., et al. 2021, NatCo, 12, 1259