The Astrophysical Journal, 900:62 (12pp), 2020 September 1 https://doi.org/10.3847/1538-4357/aba95d

A Machine Learning Model to Infer Planet Masses from Gaps Observed in Protoplanetary Disks

Sayantan Auddy and Min-Kai Lin
Institute of Astronomy and Astrophysics, Academia Sinica, Taipei 10617, Taiwan; sauddy@asiaa.sinica.edu.tw
Received 2020 May 6; revised 2020 July 8; accepted 2020 July 23; published 2020 September 1

Abstract

Observations of bright protoplanetary disks often show annular gaps in their dust emission. One interpretation of these gaps is disk–planet interaction. If so, fitting models of planetary gaps to observed protoplanetary disk gaps can reveal the presence of hidden planets. However, future surveys are expected to produce an ever-increasing number of protoplanetary disks with gaps. In this case, performing a customized fitting for each target becomes impractical owing to the complexity of disk–planet interaction. To this end, we introduce Disk Planet Neural Network (DPNNet), an efficient model of planetary gaps by exploiting the power of machine learning. We train a deep neural network with a large number of dusty disk–planet hydrodynamic simulations across a range of planet masses, disk temperatures, disk viscosities, disk surface density profiles, particle Stokes numbers, and dust abundances. The network can then be deployed to extract the planet mass for a given gap morphology. In this work, first in a series, we focus on the basic concepts of our machine learning framework. We demonstrate its utility by applying it to the dust gaps observed in the protoplanetary disk around HL Tau at 10, 30, and 80 au. Our network predicts planet masses of 80 $M_{\oplus}$, 63 $M_{\oplus}$, and 70 $M_{\oplus}$, respectively, which are comparable to those from other studies based on specialized simulations. We discuss the key advantages of our DPNNet in its flexibility to incorporate new physics as well as any number of parameters and predictions, in addition to its potential to ultimately replace hydrodynamical simulations for disk observers and modelers.

Unified Astronomy Thesaurus concepts: Exoplanet detection methods (489); Protoplanetary disks (1300); Exoplanet astronomy (486); Neural networks (1933)

1. Introduction

In the past few decades, exoplanet surveys using multiple techniques have revealed that planets are essentially ubiquitous throughout the galaxies (Cassan et al. 2012; Batalha et al. 2013). Their distribution in sizes and diversity with regard to masses, radii, and composition (e.g., Winn & Fabrycky 2015; Fulton et al. 2017) have put constraints over planet formation theories; for more details, see reviews by Johansen et al. (2014) and Raymond & Morbidelli (2020). However, with the current planet search methods (Fischer et al. 2014), it is often difficult to detect planets around young stars. Planet signatures like spectra and/or light curves from such systems are faint due to enhanced stellar activity in young stars and the presence of protoplanetary disk (hereafter PPDs). Thus, most discovered exoplanets are old ($\sim 10^3$ Myr), with few detections of young planet candidates in systems <10 Myr old (e.g., Yu et al. 2017; Keppler et al. 2018). This limits our ability to constrain the demographics of young planets, particularly during their formation epoch in PPDs. An alternative and/or indirect method to probe the unseen younger population of exoplanets is needed, as that is key to testing planet formation theories.

In this work, we introduce a novel state-of-the-art scheme that uses artificial intelligence (machine learning techniques) to infer properties of young, undetected planets from observed features in PPDs. Our idea is motivated by recent high-resolution observations (e.g., Andrews et al. 2018; Clarke et al. 2018; Huang et al. 2018a; Huang et al. 2018b; Long et al. 2018; Pérez et al. 2018; Liu et al. 2019) revealing complex—possibly planet-induced—features in PPDs with unprecedented precision. Near-infrared imaging with the Atacama Large Millimeter Array (ALMA) has provided a plethora of resolved images of PPDs showing detailed substructures like rings and distinct gaps in systems like HL Tau, as shown in Figure 1 (ALMA Partnership et al. 2015; Yen et al. 2016). Other such systems include HD 169141 (Momose et al. 2015), HD 97048 (van der Plas et al. 2017), and TW Hya (Andrews et al. 2016; Huang et al. 2018c).

The origins of such rings and gaps, although debated extensively, are broadly categorized into those caused due to disk physics and chemistry and those due to disk–planet interaction. Disk-specific mechanism include zonal flows associated with magnetorotational instability turbulence (e.g., Johansen et al. 2009; Simon & Armitage 2014), secular gravitational instability (Youdin 2011; Takahashi & Inutsuka 2014), condensation of molecular species leading to growth and fragmentation of dust around various snowlines (e.g., Zhang et al. 2015; Pinilla et al. 2017), self-induced dust pile-ups (Gonzalez et al. 2015), gap opening by large-scale vortices (Barge et al. 2017), etc.

An alternative explanation, and the one that is of interest to us, is that these gaps are induced by embedded young planets due to their dynamical interactions with the disk (Dipierro et al. 2016; Rosotti et al. 2016; Dong & Fung 2017). Although disk–planet interaction has long been predicted, (Goldreich & Tremaine 1980; Lin & Papaloizou 1993), direct observational evidence has only come recently (Pinte et al. 2019, 2020), as it relies on careful kinematic measurements. On the other hand, annular dust gaps are readily observed global features and thus offer a promising indirect method to detect planets. To this end, a model constraining planet properties from gap features is key to the exploration of the unseen young exoplanet population.

It should be noted that gap and ring formation due to planet–disk interaction and that due to disk processes are not mutually exclusive. For example, snowlines can favor planetesimal formation (Drążkowska & Alibert 2017) by acting as dust traps. In this case, ring formation at the snow line (Morbidelli 2020)

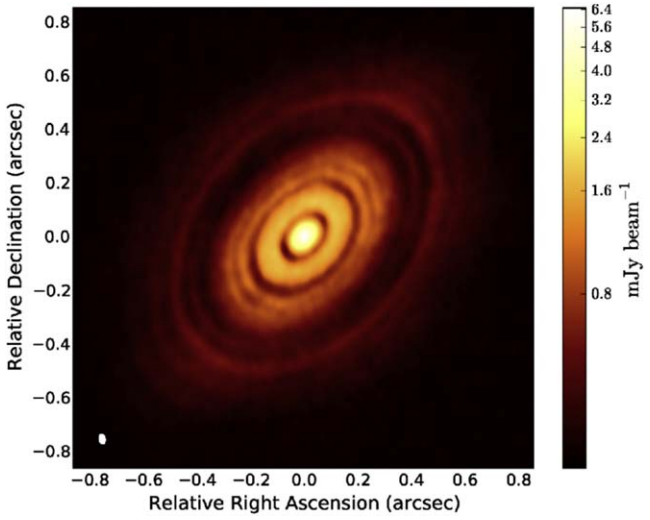

Figure 1. ALMA continuum image of HL Tau (ALMA Partnership et al. 2015). Synthesized beam is shown in the lower left. Adapted from Dipierro et al. (2015).

results in planet formation, which can then induce planet gaps and rings. It is also possible that planet and non-planet gap/ring formation processes occur in the same disk at different radii. However, including additional gap/ring formation processes is beyond the scope of this work. Our model assumes that planet–disk interaction dominates over other mechanisms.

1.1. Identifying Young Planets from Gap Profiles

In this paper, we focus on predicting planet masses from observed, axisymmetric disk gaps in dust emission. For each gap, there are at least two measurable parameters: its depth and width. The depth is the contrast between the undepleted background and the minimum intensity inside the gap, and the width is defined as the radial extent between the inner and outer edge of the gap (e.g., Dong & Fung 2017; Clarke et al. 2018; Long et al. 2018).

Both gap depths and widths have been modeled extensively using analytical and numerically approaches (e.g., Crida et al. 2006; Paardekooper & Papaloizou 2009; Duffell & Macfadyen 2013; Fung et al. 2014; Duffell 2015; Kanagawa et al. 2015a, 2016). However, instrumental limitation (i.e., lack of sensitivity resulting in poor signal-to-noise ratio) often makes it hard to constrain gap depths observationally (Zhang et al. 2018). Thus, gap widths are often considered a better alternative to constrain planet mass (Duffell & Macfadyen 2013).

Recently, Lodato et al. (2019) used a simple empirical relation (hereafter the Lodato model) to infer planet masses from observed gap widths. They assumed the dust gap width $w_{\rm d} = kR_{\rm H}$, where k is the proportionality constant and $R_{\rm H} = (M_{\rm P}/3M_*)^{1/3}R_0$ is the planet's Hill radius, R_0 is its orbital radius, and $M_{\rm P,*}$ are the planet and stellar masses, respectively. This gives

$$M_{\rm P} = \left(\frac{w_{\rm d}}{kR_0}\right)^3 3M_*. \tag{1}$$

Lodato et al. (2019) assumed a value of $k = 5.5$ by averaging results from hydrodynamical simulations of CI Tau and MWC 480. While the simplicity of this model is appealing, it does not account for the fact that gap profiles also depend on disk properties (Crida et al. 2006; Paardekooper & Papaloizou 2009).

A more sophisticated approach was taken by Kanagawa et al. (2016). They ran multiple two-dimensional (2D) hydrodynamic simulations to model gap opening in gaseous disks for relatively massive planets ($0.1M_{\rm J}$–$2M_{\rm J}$) and derived an empirical formula (hereafter the Kanagawa model),

$$M_{\rm P} = 0.0021 \times \left(\frac{w_{\rm g}}{R_0}\right)^2 \left(\frac{h_0}{0.05R_0}\right)^{\frac{3}{2}} \left(\frac{\alpha}{10^{-3}}\right)^{\frac{1}{2}} M_*, \tag{2}$$

to relate planet masses and *gas* gap widths, $w_{\rm g}$. This relation also depends on the disk's local aspect ratio h_0 and dimensionless viscosity parameter α.

A limitation of the Kanagawa model is that it requires gap widths in gas emission, which do not coincide with observed dust gaps unless particles are strongly coupled to the gas (Dipierro et al. 2016). Moreover, low-mass planets can open up dust gaps only, with negligible effect on the gas profile (Dipierro et al. 2016; Rosotti et al. 2016). As will be shown later, using dust gaps instead of gas gaps in Equation (2) leads to an overestimation of planet masses when applying the Kanagawa model.

In order to infer planet masses from dust gaps directly, one typically needs to match observations with customized dusty disk–planet simulations (e.g., Dipierro et al. 2015; Clarke et al. 2018), which introduces at least two more parameters: the strength of dust–gas coupling (or particle size) and dust abundance (or metallicity). Furthermore, as disk models improve, e.g., with realistic thermodynamics (Miranda & Rafikov 2019, 2020), more parameters become necessary to describe planet gaps precisely and place tighter constraints on planet masses. Thus, it will become impractical to perform complex disk–planet simulations to match each observed target in future large surveys of PPDs.

We are therefore motivated to seek a generic, future-proof procedure to model planet gaps. This calls for a method that can encapsulate the multidimensional parameter space associated with disk–planet interaction. In the era of big data, we propose machine learning methods as a solution.

1.2. A Machine Learning Approach

In this article, we implement an efficient and quick-paced method based on state-of-the-art machine learning (ML) techniques. We build an ML model trained with explicit planet–disk simulations, which require only a single definitive run. Our ML model can then estimate planet masses given parameters such as the dust gap width, the disk aspect ratio, disk viscosity, disk surface density profile, metallicity, and particle size.

The advantage of this ML model is threefold. First, our model can be used by observers as a tool to predict planet mass from observed PPDs *without running any simulations*. Second, the model is trained using synthetic data from numerical simulations, hence the predicted mass is more accurate (i.e., closer to the real one) than one obtained using simple empirical relations. Third, our model can easily be extended to accommodate new and improved physics. Doing so simply amounts to generating new data using advanced simulations and adding relevant variables to the ML algorithm.

1.3. Paper Plan

The paper is organized as follows. We first describe the disk–planet systems under consideration in Section 2. We then explain, in some detail, the architecture of our artificial neural network in Section 3 and how the data is preprocessed and the network is trained to predict planet masses from gap profiles. We present the predictions from our ML model in Section 4, where we also compare its performance to previous empirical approaches. In Section 5, we apply it to observed PPDs around HL Tau and AS 209 to infer planet masses. We discuss limitations and future prospects of our basic ML framework in Section 6, and conclude in Section 7. Some technical details for the training step are given in the Appendix.

2. Disk–Planet Interaction

The physical system of interest is a dusty protoplanetary disk with an embedded, gap-opening planet. To model disk–planet interaction using ML, we train the algorithm with a sample of hydrodynamic simulations. Details of the training step are presented in the next section. Here, we describe the physical disk models, simulations, and notations for later reference.

2.1. Disk Model

We consider 2D, razor-thin disk models for simplicity. Cylindrical coordinates (R, ϕ) are centered on the central star of mass M_*. We place a planet of mass M_P on a Keplerian circular orbit at $R = R_0$ and do not consider planet migration. We adopt a rotating frame with the planet and include the planet's indirect potential. We use dimensionless units such that $G = R_0 = M_* = 1$, where G is the gravitational constant. We quote time in units of the planet's orbital period $P_0 = 2\pi/\Omega_{\rm K0}$, where $\Omega_{\rm K} = \sqrt{GM_*/R^3}$ is the Keplerian frequency and subscript 0 denotes evaluation at $R = R_0$.

The disk has an initial gas surface density profile given by:

$$\Sigma_{\rm g}(R) = \Sigma_{\rm g0}\left(\frac{R}{R_0}\right)^{-\sigma}, \tag{3}$$

where $\Sigma_{\rm g0} = 10^{-4}$ (in code units) and σ is the exponent of the density profile. We neglect the disk's gravitational potential as our models have Toomre parameters $Q_0 \gtrsim 80$ (see Section 3.2), which is well above that required for gravitational stability (Toomre 1964).

We consider locally isothermal disks with a fixed sound-speed profile $c_s(R) = h_0 R\Omega_{\rm K}$, where the disk's local aspect ratio h_0 is taken to be a constant, which corresponds to a non-flared disk. The (vertically integrated) gas pressure is then given by $P = c_s^2\Sigma_{\rm g}$. We include viscous forces to mimic possible turbulence in the gas disk using the standard α-prescription (Shakura & Sunyaev 1973), such that the kinematic viscosity $\nu = \alpha c_s^2/\Omega_{\rm K}$. We assume constant α.

We also consider a single population of dust particles modeled as a pressureless fluid (Jacquet et al. 2011), which is valid for Stokes numbers $S_{\rm t} \equiv t_s\Omega_{\rm K} \ll 1$, where t_s is the stopping time characterizing the strength of dust–gas coupling (Weidenschilling 1977). For simplicity, we assume a constant Stokes number, which was found to maintain better numerical equilibrium and behavior near the disk boundaries than a fixed particle-size approach. The dust fluid is initialized with a surface density $\Sigma_{\rm d} = \epsilon\Sigma_{\rm g}$, where ϵ is the constant (global) dust-to-gas ratio. The reference dust surface density is $\Sigma_{\rm d0} = \epsilon\Sigma_{\rm g0}$. The backreaction from the dust drag onto the gas is included.

Table 1
Parameter Space for Hydrodynamic Simulations

Name	Notation	Max.	Min.
Planet mass in Earth masses	$M_{\rm P}/M_\oplus$	120	8
Disk aspect ratio	h_0	0.100	0.025
Disk surface density profile	σ	1.2	0.05
Disk viscosity parameter	α	10^{-2}	10^{-4}
Global dust-to-gas ratio	ϵ	10^{-1}	10^{-2}
Particle Stokes numbers	$S_{\rm t}$	10^{-1}	10^{-3}

The dusty disk is initialized with steady-state drift solutions given by Equations (70)–(73) of Benítez-Llambay et al. (2019), which assumes a constant Stokes number.

2.2. Hydrodynamic Simulations

We simulate the above dusty disk–planet systems with the FARGO3D hydrodynamics code (Benítez-Llambay & Masset 2016). FARGO3D can be run on Graphics Processing Units (GPUs), achieving large speed-ups compared to Central Processing Units (CPUs). For instance, each of our simulation takes approximately ∼12 GPU hr compared to ∼72 hr on a single CPU core. This advantage, together with the GPU clusters at ASIAA and the National Center for High-Performance Computing in Taiwan, enabled us to run a large number of simulations ($\sim 10^3$) necessary to generate the required training data within a reasonable time.

Our computational domain is $R \in [0.4, 2.5]R_0$ and $\phi \in [0, 2\pi]$, and are discretized with 512×512 uniformly spaced cells. Periodic boundaries are applied in ϕ and the radial boundaries are set to their initial (equilibrium) solutions. We also apply wave-killing zones near the disk edges (de Val-Borro 2006) where the disk is relaxed toward its initial state. The planet is introduced from the beginning of the simulation. We use a constant softening length of $r_s = 0.6h_0{\rm R}_0$ for the planet's potential.

2.3. Parameter Space

We mainly focus on low-mass planets because they are more difficult to observe directly, which makes our ML method more useful (Lodato et al. 2019). To this end, we consider planetary masses ranging from $M_P = 8\,M_\oplus$ to $M_P = 120\,M_\oplus$ around a $1\,M_\odot$ central star. We consider a range of values for disk aspect ratios $h_0 \in [0.025, 0.1]$, viscosity parameters $\alpha \in [10^{-4}, 10^{-2}]$, and power-law slope of the surface density profile $\sigma \in [0.05, 1.2]$. The value of σ affects the dimensionless pressure gradient $\eta \equiv (h_0^2/2)d\ln P/d\ln R \propto -h_0^2(1+\sigma)$, which controls the radial drift of dust (Benítez-Llambay et al. 2019). For each disk, we consider one dust species characterized by (fixed) Stokes numbers $S_{\rm t} \in [10^{-3}, 10^{-1}]$ and abundance $\epsilon \in [0.01, 0.1]$. These values of $S_{\rm t}$ roughly translate to a physical radii of dust particles (or pebbles) ranging from ∼1 mm to ∼10 cm, assuming an internal grain density of $1\,{\rm g\,cm^{-3}}$ and gas surface density $100\,{\rm g\,cm^{-2}}$. The parameter space is summarized in Table 1.

Each simulation is allowed to evolve for 3000 orbits or about ∼1 Myr at 45 au around a $1\,M_\odot$ star. The time is adequate for most models to reach a quasi-steady state. For some, the gaps have not fully settled; however, they are still relevant, as they have evolved for an appreciable fraction of the gas disk lifetime.

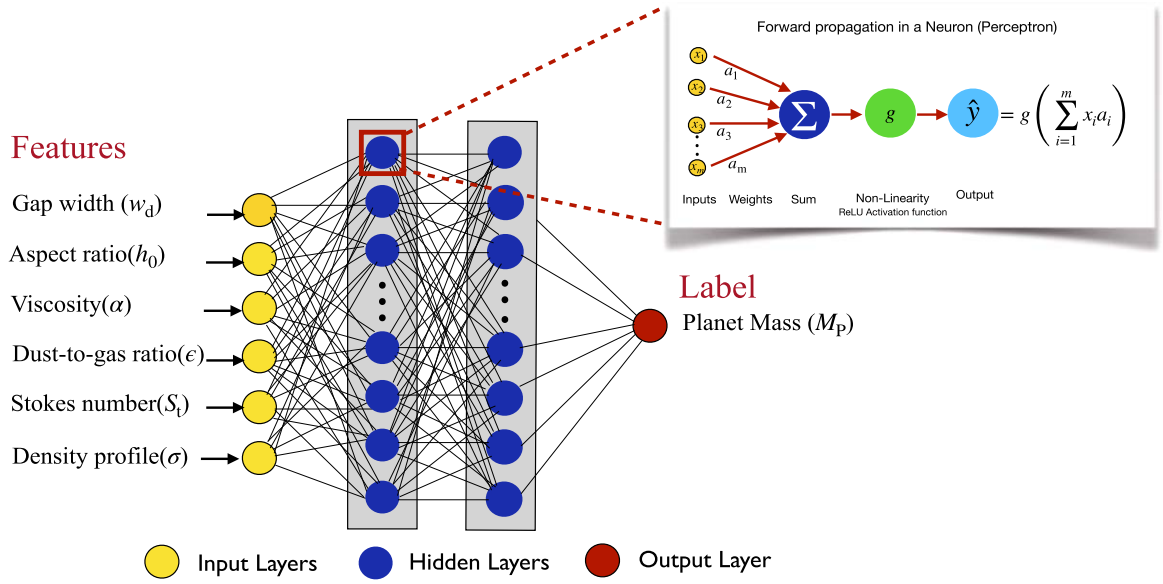

Figure 2. Schematic diagram of the DPNNet using a fully connected multilayer perceptron. First layer (in yellow) takes as input six feature variables. Output layer (in red) gives the planet mass as a target variable. Two hidden layers in the middle (in blue) have 128 units of neurons each. Each neuron performs a simple task via forward propagation, as illustrated in the magnified panel. It finds the linear combination of inputs (weighted sum) and then passes this through a ReLU activation function to generate an output. The network is trained using RMSprop optimizer. More details are given in Section 3.

3. Implementation of Deep Neural Network

The last few years have seen a growth in the use of machine learning, a branch of artificial intelligence, in diverse domains to solve complex real-world problems. In particular Deep Learning (LeCun et al. 2015), a subfield of ML, has been the main driving force behind breakthroughs in image classification (Krizhevsky et al. 2012), speech recognition (Chorowski et al. 2015), text classification (Yang et al. 2016), and more recently, self-driven (autonomous) cars (Grigorescu et al. 2019). Deep Learning allows machines to identify patterns in data using *neural networks* to perform tasks like detection, classification, regression, segmentation, etc. In this work, we are interested in building a neural network and applying it to a *regression* problem (see Goodfellow et al. 2016; Alibert & Venturini 2019) in astronomy. The task of our neural network is to learn from input variables (also called *features*) like gap width, disk aspect ratio, disk viscosity, disk surface density profile, dust abundance, and particle size observed in PPDs to predict a target variable (called *label*) planet mass. In the next section, we discuss the structure and the working principle of the network.

3.1. Setting up the Network

A neural network consists of multiple layers of neurons or perceptrons; for a review, see Schmidhuber (2015). Each neuron performs a simple task via forward propagation, meaning the process of generating output flows in one direction from the input layer to the next layer. A neuron accepts a set of inputs and the corresponding weights. It then passes the linear combination of the inputs through a nonlinear activation function to generate an output. The first layer of the network takes as input the features and the last layer gives (predicts) a target variable that is, ideally, close to the true value. The middle layers are also called the hidden layers, as the states in them are not directly enforced, unlike the input and the output layer. The hidden layers accept the outputs from the previous layer and feed their results into the next layer. A network is said to be *deep* when it has more than one hidden layer.

We design a deep neural network, the Disk Planet Neural Network (DPNNet), to model planetary gaps using a multilayer perceptron. Figure 2 gives a schematic representation of our DPNNet, with two hidden layers, an input layer (yellow) with six feature variables, and an output layer with one target variable. Our network is trained using simulation data to predict the target variable, the planet mass. The training process works iteratively, whereby the weights of all linear combinations of all neurons are adjusted at each step to get the best agreement between the predicted and the actual value. The matching between the actual and the predicted values is quantified by the cost function, the mean square error (MSE). It is defined as

$$C(a) = \frac{1}{N}\sum_{n=1}^{N}(M_{\rm P,predicted,n} - M_{\rm P,simulation,n})^2, \qquad (4)$$

where N is the sample size, $M_{\rm P,simulation,n}$ is the true planet mass used in the simulation, $M_{\rm P,predicted,n}$ is the value predicted by the neural network, and a are the weights. The gradient of the cost function is computed using the chain rule for derivatives, known as backpropagation (Goodfellow et al. 2016). The cost is then minimized iteratively using gradient descent (Ruder 2016). After each iteration, the weights are updated until an optimal value is reached.

In our DPNNet, each hidden layer has 128 units of neurons. They are fully connected, as each unit of a layer is linked to each unit of the previous and the next layer. We use the ReLU activation function (Nwankpa et al. 2018) to introduce nonlinearity for each unit. The network is trained using the RMSprop optimizer (Ruder 2016) with its parameters set to default values. We select a learning rate of 0.0001. The network dimensions are arrived through a series of trial-and-error tests by varying the number of layers and units per layer. A lower validation loss indicates a better model. Thus, we

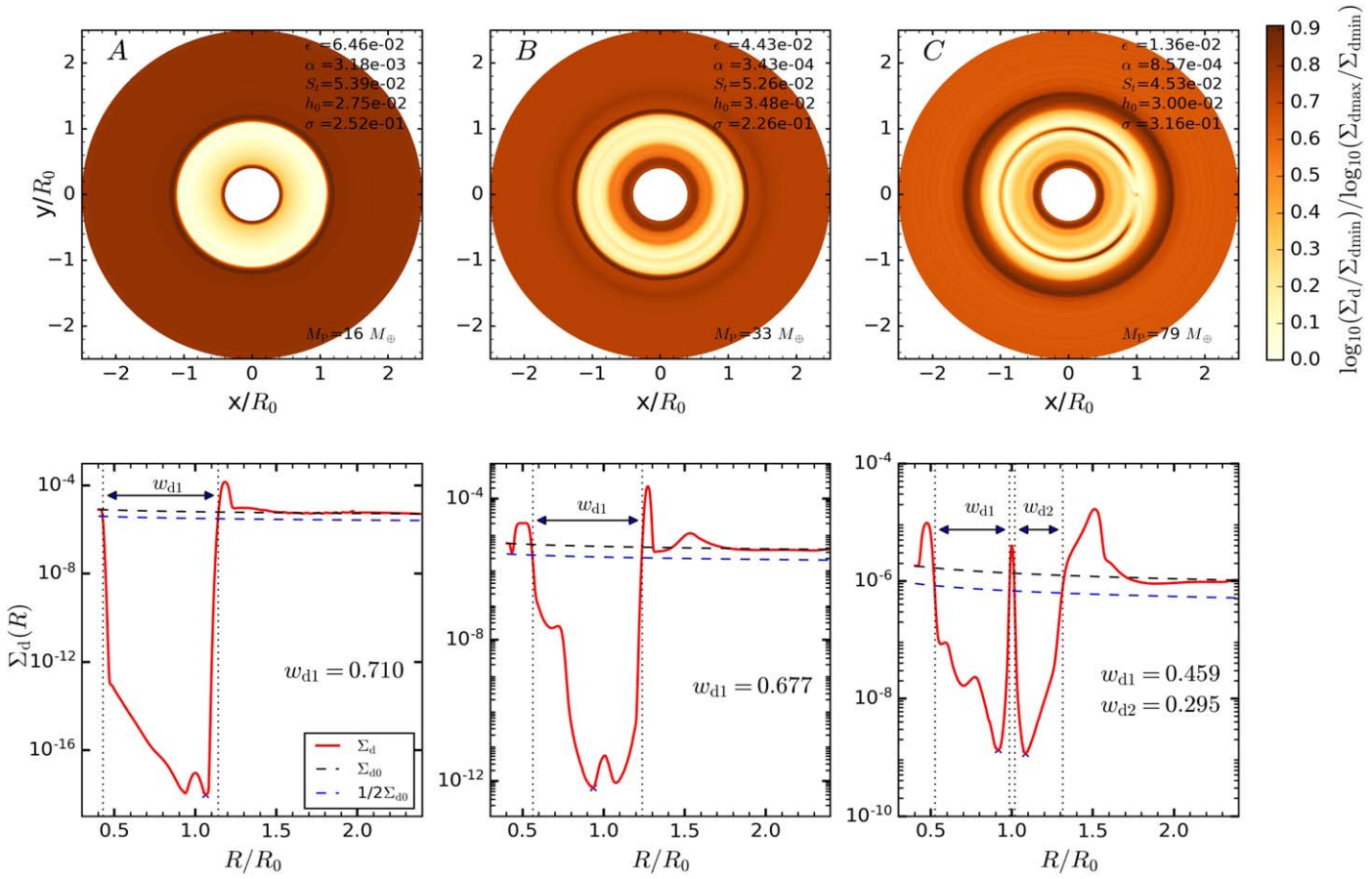

Figure 3. Top panel is the normalized dust surface density distribution after 3×10^3 orbits for different planet masses and different disk initial conditions. Disk parameters for each model are indicated on the top right of each plot. Bottom panel is the radial profile (red line) of the azimuthally averaged surface density for each simulation. Horizontal arrow represents the widths w_{d1}, w_{d2} of the gap(s). Gap width is the distance between the inner and the outer edge of the gap (vertical dotted lines) where $\Sigma_d(R)$ reaches 50% of the initial surface density Σ_{d0} shown in dotted blue line. Cross indicates the minimum surface density Σ_{dmin}.

begin by increasing the size of the layers or adding new layers until we get diminishing returns on the validation loss. We further implement L2 regularization (Ng 2004), with a coefficient of 0.0001, and early stopping to mitigate overfitting.

3.2. Data Acquisition

The parameter space for the disk–planet simulations are sampled using the Latin hypercube sampling (McKay et al. 1979; Iman et al. 1981) method. We generate a uniform random distribution of the parameters, with each value centered within the sampling intervals. We ran a total of 1100 simulations across the range of input parameters described in Section 2.3. Example simulations are shown in Figures 3–4 and discussed in more detail below. We analyze the data from each simulation and measure the width of the gap/gaps opened by the embedded planet in both the dust and gas profiles.

Our simulations show that, for various parameter combinations, a planet can induce multiple but adjacent gaps/rings in the dust (Dong et al. 2017). We thus broadly classify our simulations into two categories: (a) one deep, broad gap as seen in simulations A and B in Figure 3; and (b) two gaps adjacent to a planet (see panel C in Figure 3). To account for case (b), we simply define two dust gap widths for each simulation, w_{d1} and w_{d2}, but set $w_{d2} = 0$ if only a single dust gap is formed. There are are some simulations that open secondary gaps away from the planet (Bae et al. 2017; Miranda & Rafikov 2019). However, we exclude such simulations from our training set as we do not have enough samples to effectively train our models to consider secondary gaps.

Following Kanagawa et al. (2016), we measure the dust gap width as the radial distance between the inner $R_{d,in}$ and outer $R_{d,out}$ edge of the dust gap where the dust surface density Σ_d reaches a predefined threshold fraction of Σ_{d0}, respectively. Thus, the first dust gap width is given by

$$w_{d1} = \frac{R_{d1,out} - R_{d1,in}}{R_0}, \quad (5)$$

where R_0 is the planet's orbital radius, as illustrated in Figure 3. We use similar definitions for the second dust gap width w_{d2} (if it exists) and the gas gap width w_g (see Figure 4). For measuring the dust and gas gap widths, we adopt a threshold fraction of 1/2 of the initial surface density for easy comparison with the Kanagawa model (Kanagawa et al. 2016). One limitation with this definition of w_d is that it fails for shallow gaps (associated with low-mass planets) when the depletion is less than the threshold fraction. However, one can choose any threshold fraction or implement another definition of the gap width. In that case, the DPNNet will be trained using those numbers.

Figure 3 illustrates a subset of simulation outputs from the training set. The top panel shows the dust surface density at the end of 3000 orbits for three independent runs with distinct

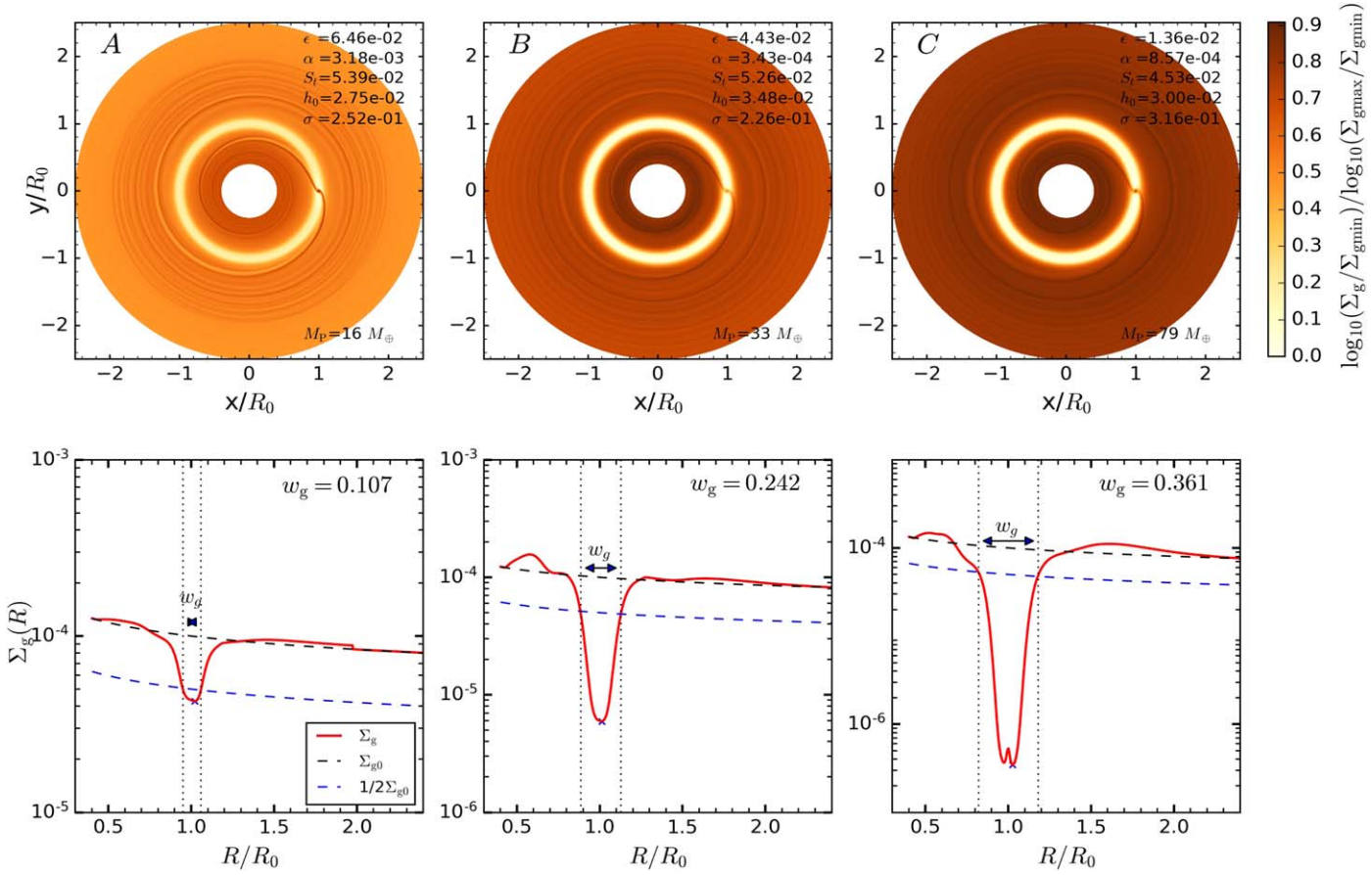

Figure 4. Same as Figure 3, but for gas surface density Σ_g.

planet masses and different disk initial conditions (shown on the upper right corner). Simulations A, B, and C have embedded planets of masses $M_P = 16\,M_\oplus$, $M_P = 33\,M_\oplus$, and $M_P = 79\,M_\oplus$, respectively. The bottom panel is the corresponding radial profile of the azimuthally averaged surface density. The extents of the dust gap widths w_d are shown by the horizontal arrows.

Figure 4 shows the gas gap surface density and one-dimensional radial profile for the same models as in Figure 3. Gas gaps are much smoother and less deep compared to their dust counterparts. The gas gaps mainly fall into category (a), i.e., one gap around the planet, according to the above gap classification scheme. As is evident, gas gap profiles and features are vastly dissimilar, compared to dust gaps.

Figure 3 highlights the fact that the variation in the gap profiles due to planet–disk interaction, which is complex and nonlinear, remains present even in the simple empirical relations of Equations (1)–(2). For instance, even though the embedded planet in simulation B is heavier ($\times 2.1$) compared to simulation A, they both have comparable dust gap width. This is likely due to the differences in the ambient disk initial conditions: for example, simulation B has a higher ($\times 1.2$) aspect ratio and lower viscosity ($\times 0.1$) compared to simulation A. Furthermore, in simulation C, a planet of mass $M_P = 79 M_\oplus$ carves a double dust gap of category (b), possibly due to even lower viscosity, $\alpha = 8.57 \times 10^{-4}$, compared to simulation A and B.

The initial data set consisted of 1100 simulations. The data is preprocessed to eliminate runs that do not open up detectable gaps or have secondary gaps. After the initial screening, we have data from 976 simulations. In addition to the input run parameters (M_P, α, h_0, σ, S_t, ϵ), each simulation is also associated with gap width(s) in both dust (w_{d1}, w_{d2}) and gas (w_g). However, the gas gap width will only be used for comparison with other empirical models (Section 4.2), and is *not* part of our ML model. Table A1 gives the distribution of the parameters of the raw data set and the processed data set after the filtering.

For building our ML model, we assign each simulation a set of feature variables (w_{d1}, w_{d2}, ϵ, α, S_t, h_0, σ) and a target variable (or label), here the planet mass M_P. For each feature variable, the data is normalized using the standard scaling (z-score) by removing the mean and then scaling it by the standard deviation. Note that feature and target variables need not correspond to simulation input parameters and output measurements, respectively. We are free to assign feature and target variables to fit the application, which in our case is to predict planet masses. It would be straightforward to add, remove, or swap variables to predict other quantities.

3.3. DPNNet Training

The DPNNet is implemented using the Google TensorFlow (Abadi et al. 2015), which is an open-source platform for machine learning. The data are randomly split into two blocks: 80% of them used for training and validation, and the remaining 20% for testing. This is crucial because we want to test our model on data that it has not previously seen or been

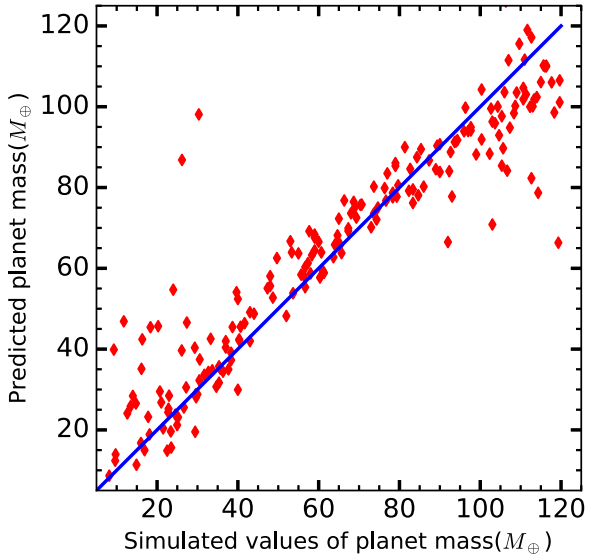

Figure 5. Correlation between the predicted planet mass (in units of $M_\oplus$) using DPNNet and the simulated planet mass used in the numerical simulations.

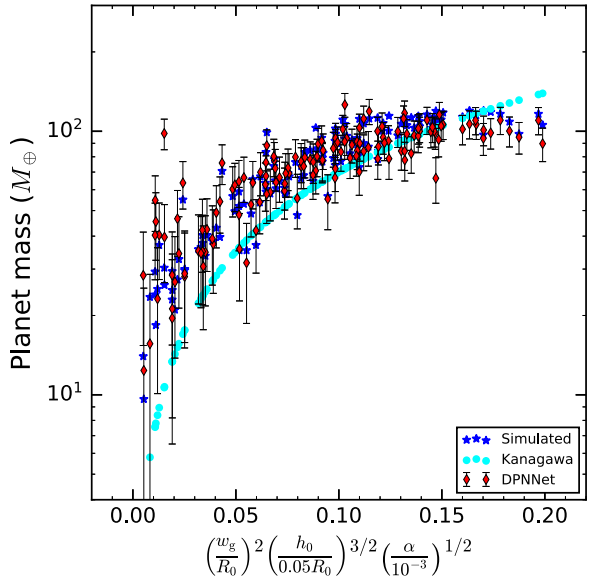

Figure 6. Predicted planet mass from DPNNet using (α, h_0, σ, S_t, ϵ, w_{d1}, w_{d2}) as input and Kanagawa model using (w_g, h_0, α) for input are plotted as a function of gas gap width w_g, aspect ratio h_0, and viscosity α as given in Equation (2). Vertical error bars are prediction uncertainty associated with the network (see Section 4.1). Blue markers are the simulated mass ($M_{P,simulation}$) of the planets.

trained with. The training set is fed into our DPNNet and the weights are adjusted iteratively until they are optimized (see Section 3.1 for details). The validation set is used to check the general performance of DPNNet (e.g., to avoid overfitting). In Figure A1, we plot the training and the validation loss, the MSE, as a function of training epoch for the DPNNet. As is evident, the validation loss gradually decreases and starts to flatten after 600–700 epoch, whereas the training loss is still decreasing, indicating overfitting. Thus, we implement early stopping and stop the iterations around 800 epochs. In the next section, we apply our trained network to predict planet mass.

4. Results

Once the training process is complete, the DPNNet (with optimized weights) is ready to be deployed to predict planet mass from observed disk. We initially test the network's performance by applying it to the test data set. This allows us to quantify the network's overall accuracy in predicting planet mass from unseen data. We also compare our network's prediction with other empirical relations and discuss its advantages.

4.1. DPNNet Prediction

The test data set consists of feature variables along with the true planet mass. The feature variables are normalized with same distribution that the network has been trained on. It is then fed to the trained DPNNet to predict the planet mass. Figure 5 illustrates the correlation between the predicted and the simulated planet mass in Earth-mass ($M_\oplus$) units for all the test samples. The predicted mass (red diamonds) lies along the blue line, indicating a close correlation. We estimate the error ($M_{P,simulation} - M_{P,predicted}$) between the predicted and the simulated mass. It is normally distributed with a mean (μ) and a standard deviation (σ_E) of $0.8\,M_\oplus$ and $12.5\,M_\oplus$, respectively. The model performance is evaluated using standard metrics such as the root mean square error (RMSE) and the mean absolute error (MAE). The DPNNet has an RMSE of $12.5\,M_\oplus$ and an MAE of $7.9\,M_\oplus$ when applied to the test data set.

Furthermore, to lower the bias associated with data sampling, we apply k-Fold cross-validation resampling process using scikit-learn (Pedregosa et al. 2011). Instead of using a single RMSE, we take the average of the RMSE obtained from each of the fivefold resampling process. We get a mean RMSE of $13.0M_\oplus$. This can be regarded as the prediction uncertainty of our network. Given our limited sample size and wide parameter space, the model error is moderate, particularly toward the high-mass end of our parameter space. The network performance can be further improved by using more comprehensive data. However, doing so is beyond the scope of this work.

In the next section, we compare DPNNet with both the Kanagawa (Kanagawa et al. 2016) and the Lodato model (Lodato et al. 2019).

4.2. Model Comparison

We apply the Kanagawa and Lodato models along with DPNNet to the test data set to generate planet masses and compare it with simulated values. This is done as follows. Each run in the test data set is associated with some combination of (M_P, α, h_0, σ, S_t, ϵ, w_{d1}, w_{d2}, w_g). This M_P is the simulated planet mass. The Lodato model (Equation (1)) takes a dust gap width as the sole input to predict a planet mass, in which case we use w_{d1}. Similarly, we only use (w_g, h_0, α) for input in the Kanagawa prediction (Equation (2)). Finally, our DPNNet takes (α, h_0, σ, S_t, ϵ, w_{d1}, w_{d2}) as input to predict a planet mass.

Figure 6 shows the simulated planet mass (blue star) and predicted planet mass using the Kanagawa model[1] (cyan hexagons). The Kanagawa planet masses are comparable to the simulated values with an RMSE of $18.2M_\oplus$. While the Kanagawa model performs reasonably well toward the higher-mass end, one caveat is that it requires gap widths measured in

[1] In applying the Kanagawa model, we select only those simulations that have a measurable gas gap.

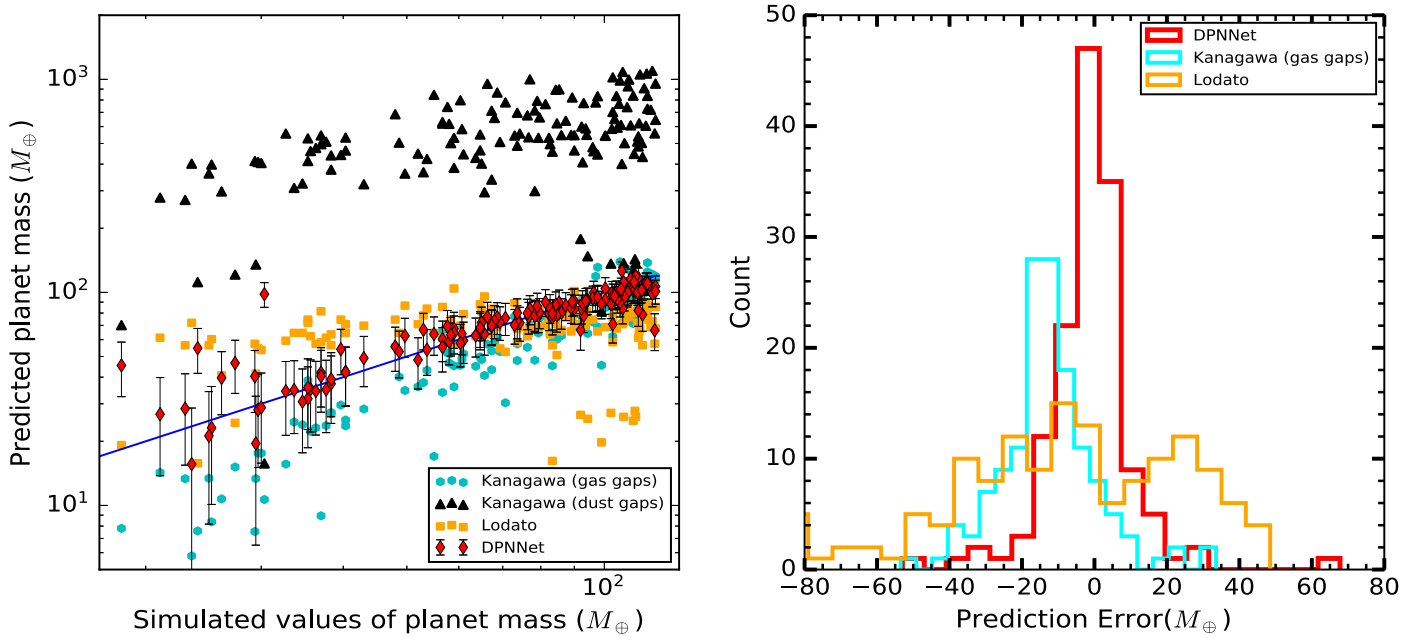

Figure 7. Left: Comparison between the planet mass predicted using different models and the simulated planet mass. DPNNet-predicted mass lies along the blue line, indicating close correlation with the true mass. Black triangles are the masses predicted based on the dust gap width using Kanagawa models, i.e., setting $w_g \rightarrow w_{d1}$ in Equation (2). The Kanagawa models overestimate the planet mass. Orange squares are the masses inferred from Lodato models (Equation (1)). They deviate considerably from the simulated values both in the low- and high-mass ends. The masses inferred (cyan hexagon) from the Kanagawa model using gas gap widths (Equation (2) without modification) are marginally below the blue line near the low-mass end, indicating underprediction. Right: Distribution of the prediction error ($M_{P,simulation} - M_{P,predicted}$) for Kanagawa, Lodato, and DPNNet with ($\mu = -12.7M_\oplus$, $\sigma_E = 13.1M_\oplus$), ($\mu = -7.4M_\oplus$, $\sigma_E = 31.2M_\oplus$), and ($\mu = -1.2M_\oplus$, $\sigma_E = 12.0M_\oplus$), respectively.

gas emission. This may limit its application, as observed gap profiles are mostly from dust emission. One would then first need to estimate gas gap widths, which can introduce large uncertainties (Zhang et al. 2018).

On the other hand, our DPNNet takes as input dust gap widths (a more direct observable) along with other disk features. The DPNNet-predicted planet masses are indicated in red diamonds in Figure 6. The vertical error bars represent the uncertainty ($\pm 13.0M_\oplus$) in the predicted mass. The RMSE for the DPNNet is $12.1M_\oplus$. Our DPNNet predictions have substantially reduced bias and produce better fits compared to the Kanagawa model.

Figure 7 makes a more direct comparison between models. It shows the predicted M_P from dust gap widths and disk parameters as a function of actual planet mass. An exact prediction would lie on the blue line. The black triangles are the predicted masses from the Kanagawa model, but using the dust gap width instead of the gas gap width (setting $w_g \rightarrow w_{d1}$ in Equation (2)). These are an order of magnitude above the blue line, which shows that using dust gap widths in the Kanagawa models leads to significant overestimates in planet mass. The orange squares are the planet masses inferred using the Lodato model. We rescale the proportionality constant k in Equation (1) to adjust for the different way we measure the gap width compared to Lodato et al. (2019). Using the training data set, we fit the measured dust gap width w_{d1} to $k \times (M_P/3M_*)^{1/3}$, where $M_{P,*}$ are the simulated planet and stellar mass, and we obtain $k = 18.4$. The estimated masses deviate considerably from actual simulated values both in the low- and high-mass end. The orange histogram on the right denotes the distribution of the error in the Lodato model. It follows a normal distribution with $\mu = -7.4M_\oplus$ and $\sigma_E = 31.2M_\oplus$.

On the other hand, the Kanagawa model using gas gap widths (cyan hexagons; Equation (2) without modification) typically underestimates planet masses, although it performs better at the high-mass end. The cyan histogram on the right follows a normal distribution with $\mu = -12.7M_\oplus$ and $\sigma_E = 13.1M_\oplus$. The peak of the histogram is considerably shifted from the center, indicating a negative bias in its estimate. In contrast, the mass predicted using the DPNNet lies along the blue line, indicating a much closer correlation than other models. Thus, our DPNNet outperforms both empirical relations when applied to the test data set. The error follows a normal distribution (red histogram) with $\mu = -1.2M_\oplus$ and $\sigma_E = 12.0M_\oplus$.

5. Application to Observations

As a first application of our newly developed DPNNet, for planet gaps, we estimate planet masses responsible for opening the dust gaps observed in the protoplanetary disks around HL Tau and AS 209. Our findings are summarized in Table 2 and discussed below.

5.1. HL Tau

We apply our DPNNet to the HL Tau disk, which shows clear evidence of axisymmetric gaps in dust thermal emission (ALMA Partnership et al. 2015). Several studies using hydrodynamic simulations suggest such rings are due to disk–planet interaction (Dipierro et al. 2015; Dong et al. 2015; Jin et al. 2016). Assuming that each gap is created by a single planet, we use the properties of the observed gap as input features (see Table 2) for our ML model to predict the planet mass responsible for carving out each gap. For instance, the

Table 2
Inferred Planet Masses from Gaps in HL Tau and AS 209

	Observed Features								DPNNet	Other Models					
Name	M_* ($M_\odot$)	$R_{\rm gap}$ (au)	$w_{\rm d1}$	h_0	α	$S_{\rm t}$	ϵ	σ_0	$M_{\rm p0}$ ($M_{\rm J}$) $\pm 0.04 M_{\rm J}$	$M_{\rm p1}$ ($M_{\rm J}$)	$M_{\rm p2}$ ($M_{\rm J}$)	$M_{\rm p3}$ ($M_{\rm J}$)	$M_{\rm p4}$ ($M_{\rm J}$)	$M_{\rm p5}$ ($M_{\rm J}$)	$M_{\rm p6}$ ($M_{\rm J}$)
	1.00	10	0.81	0.05	10^{-3}	0.005	0.01	1.0	0.25	1.40	0.20	0.35	0.20	⋯	⋯
HL Tau	1.00	30	0.23	0.07	10^{-3}	0.005	0.01	1.0	0.20	0.20	0.20	0.17	0.27	⋯	⋯
	1.00	80	0.29	0.10	10^{-3}	0.005	0.01	1.0	0.22	0.50	0.20	0.26	0.55.	⋯	⋯
AS 209	0.83	9	0.42	0.04	10^{-4}	0.016	0.01	1.0	0.13	⋯	⋯	⋯	⋯	0.37	⋯
	0.83	99	0.31	0.08	10^{-4}	0.016	0.02	1.0	0.14	⋯	⋯	⋯	⋯	0.18	0.21

Notes. $M_{\rm P0}$ is the mass inferred using DPNNet, from input features ($w_{\rm d1}$, h_0, α, $S_{\rm t}$, ϵ, σ) given in column 4–9 and setting $w_{\rm d2} = 0$. $M_{\rm P1}$ is the mass predicted by Kanagawa et al. (2016) using the empirical relation. $M_{\rm P2}$, $M_{\rm P3}$, and $M_{\rm P4}$ are the masses obtained using the customized simulation of HL Tau by Dong et al. (2015), Jin et al. (2016), and Dipierro et al. (2015) respectively. $M_{\rm P5}$ is the planet mass in AS 209 inferred by Zhang et al. (2018). $M_{\rm P6}$ is the planet mass found by Fedele et al. (2018), using specialized simulations of AS 209.

disk aspect ratios and gap widths for three identified gaps at $R = 10$, 30, and 80 au in HL Tau are obtained from Kanagawa et al. (2015b, 2016). Those were estimated for an assumed spectral index $\beta = 1.5$, using optical depth and the gas temperature from the brightness temperatures in ALMA Bands 6 and 7. We set the viscosity $\alpha = 10^{-3}$ (Dipierro et al. 2015), a central star mass $M_* = 1\,M_\odot$, and a canonical value of the dust-to-gas ratio $\epsilon = 0.01$ (ALMA Partnership et al. 2015; Dong et al. 2015). The Stokes number is on the order of 10^{-2}–10^{-3} around the three gaps (Dong et al. 2015). For simplicity, we adopt $S_t = 0.005$. Our DPNNet predicts the planet mass in gaps at 10, 30, and 80 au as 80, 63, and 70 $M_\oplus$, with an uncertainty of $\pm 13\,M_\oplus$. These masses correspond to 0.25, 0.20, and 0.22 M_J.

These estimates are in excellent agreement with the planet masses inferred from direct numerical simulations. For instance, Dong et al. (2015) and Jin et al. (2016) generated synthetic images using three-dimensional Monte Carlo Radiative Transfer and hydrodynamical simulations of disk–planet systems to compare with observations. Dong et al. (2015) established that three planets each of mass $0.2 M_{\rm J}$ were responsible for the gaps in HL Tau. Jin et al. (2016) found three embedded planets with masses 0.35, 0.17, and 0.26 $M_{\rm J}$ in the gaps. Furthermore, three-dimensional dusty smoothed particle hydrodynamics calculations by Dipierro et al. (2015) yield comparable masses of 0.20 and 0.27 $M_{\rm J}$ in the two inner gaps, and a larger 0.55 $M_{\rm J}$ mass planet in the outer gap. However, these inferred planet masses (from our DPNNet and specialized simulations), particularly at the inner gap, differ considerably from the Kanagawa model (Kanagawa et al. 2016). They predicted planet masses of 1.40 $M_{\rm J}$, 0.20 $M_{\rm J}$, and 0.5 $M_{\rm J}$ at 10 au, 30 au, and 80 au, respectively, using the dust gap widths. This may be because the Kanagawa model often overestimates planet mass when applied to dust gap width (see Section 4.2).

5.2. AS 209

AS 209 is another interesting system with multiple gaps at 9, 24, 35, 61, 90, 105, and 137 au (Guzmán et al. 2018; Huang et al. 2018a) surrounding a central star of mass $M_* = 0.83\,M_\odot$. Zhang et al. (2018) established, using numerical simulations, that a single planet of mass $\sim 27 M_\oplus$ ($M_{\rm P} = 10^{-4} M_*$) at $R \sim 100$ au in a $\alpha \simeq 10^{-5}$ disk (with radially varying α) can produce all five of the gaps at 24, 35, 62, 90, and 105 au. The disk aspect ratio $h_0 \sim 0.05$–0.06 was constrained using the distance between two dominant gaps at $R = 61$ and 100 au. Zhang et al. (2018) also suggested that the innermost gap at 9 au hosts a second planet. Furthermore, Fedele et al. (2018) used customized 3D hydrodynamical simulations of planet–disk interaction to establish that a single planet of $M_{\rm P} \sim 0.7 M_{\rm Saturn}$ ($\sim 67 M_\oplus$) at $R \sim 103$ can open up the two gaps detected at 62 and 103 au.

However, the DPNNet is not explicitly trained for systems where multiple (>2) gaps are induced by a single planet. Thus, for simplicity, when applying it to AS 209, we only consider the planet-harboring gaps at $R = 9$ and 99 au and ignore the others. The potential effect on the predicted mass due to this approximation is likely to be within the model uncertainty. Advance application considering the influence of these subsidiary gaps will be implemented in future improvements.

As input to our network, we adopt disk parameters and dust gap widths[2] from Zhang et al. (2018). While Zhang et al. (2018) include a range of values for dust abundance, particle size, and disk viscosity, we consider those that are within our parameter space (Table 1). For example, we select $S_t = 30 \times 5.23 \times 10^{-4}$ and $\alpha = 10^{-4}$ for both the gaps. Next, for the innermost gap at $R = 9$ au, the other selected parameters are $w_{\rm d1} = 0.42$, $\epsilon = 0.012$, and $h_0 = 0.04$. For the gap at $R = 99$ au, we have $w_{\rm d1} = 0.31$, $\epsilon = 0.017$ and $h_0 = 0.08$. We set surface density profile $\sigma = 1$ (Fedele et al. 2018) for both the gaps. The feature variables are summarized in Table 2.

The DPNNet predicts planet masses of $39 M_\oplus$ and $45 M_\oplus$ ($\pm 13\,M_\oplus$) in 9 au and 99 au gaps, respectively. These masses correspond to 0.13 and 0.14 $M_{\rm J}$. Our estimates, particularly at $R = 9$ au, are lower compared to Zhang et al. (2018) values, as they find masses of 0.37 $M_{\rm J}$ and 0.18 $M_{\rm J}$ at 9 au and 99 au, respectively, for the same disk parameters. The discrepancy could be partially because Zhang et al. (2018) measured gap widths differently than us and also we did not consider the effects of subsidiary gaps. Furthermore, the disk parameters adopted from them are likely to be uncertain.

Given the simplicity of our setup, the limited parameter range, and the small sample size used in our training set, our DPNNet performs remarkably well, as its predictions are

[2] Note that Zhang et al. (2018) define the gap width differently than us. This is likely to introduce additional uncertainty in the predicted planet mass.

consistent with specialized models and simulations, which were exclusively customized to reproduce selected observations.

6. Discussion

We introduce an ingenious multidisciplinary technique that combines computational astrophysics with deep learning (using neural networks) to characterize unseen exoplanets from observed disks. Deep neural network learns the underlying relationship in a set of data via training. In our case, it maps the connection between the planetary gaps in dust and the complex disk–planet interaction. This enables our DPNNet to directly estimate planet masses from observed dust gap width and disk features.

Previous analytic scaling relations that can infer planet mass from gas surface density are often limited, as PPDs are mostly detected in dust emission. Using these empirical relations requires one to estimate gap profiles in gas from dust continuum flux, which is uncertain and prone to large errors (Zhang et al. 2018). While customized disk–planet simulations for modeling specific PPDs have been successful, this approach may not be efficient in dealing with multiple targets in a large survey. Thus, machine learning, more specifically deep learning, provides an alternative way that is much more accurate and definitive in characterizing unseen exoplanets from observations of PPDs.

The key to a well-trained network is a comprehensive data set that reflects the underlying connection between the input features and the output (target) variables. This necessitates the inclusion of data that incorporate not only a broad parameter space but also capture the diverse disk morphologies due to disk–planet interaction. However, as the first paper in a series, we have only explored a limited range, due to finite resources. This restricts the scope of our network, as discussed below.

6.1. Network Limitations

The network is trained with limited data generated using idealized 2D hydrodynamical simulations. Our limited sample size for the training set affects the network performance, resulting in higher uncertainty in its prediction. For example, our optimized network has an uncertainty of $\pm 13.0 M_{\oplus}$, which corresponds to 11% error even for the most massive planet in our parameter space. The relative uncertainty is much higher at the low-mass end. Runs resulting in complex morphologies such as secondary/tertiary gaps have to be excluded, as we do not have enough samples to successfully train the network to identify them.

Furthermore, we only consider gaps opened by a single planet in a narrow mass range ($8 M_{\oplus} \leqslant M_{\rm P} \leqslant 120 M_{\oplus}$). These allow for smoother gap profiles, which are easier to characterize, compared to massive ones, as those are likely to induce eccentric gaps (Papaloizou 2002; Kley & Dirksen 2006) and have vortices at gap edges (Koller et al. 2003; Li et al. 2005; de Val-Borro et al. 2007). This further prevents the application of the network to disks harboring massive planets.

We only measure the gap widths ($w_{\rm d}$) after 3000 orbits, whence the majority of the simulations have settled into a quasi-steady state. Here, an improved approach would be to measure the gap profiles as a function of time as well, such that our network learns to identify the variations due to time evolution in the observed gap. This would then allow one to relax the assumption that the observed system has reached a steady state, and use the system age as another input parameter.

Our idealized 2D simulations can also be improved. For instance, a fixed particle-size approach in place of a constant Stokes number, adopted here for better numerical behavior, would be a more physical model for dust–gas interaction (Weidenschilling 1977).

Previous work showed that the vertically integrated gap structure in gas is similar between 2D and three-dimensional (3D) simulations (Fung & Chiang 2016). Furthermore, solids tend to settle into a thin dust layer around the disk midplane. Thus, we do not expect the relation between dust gap profiles and planet masses, as identified by our DPNNet based on 2D simulations, to differ significantly from that trained with 3D simulations. Nevertheless, 3D disk models should be considered in the future, as these can also account for disk and planet inclinations, which cannot be captured in 2D. At present, the cost of 3D simulations likely prohibits one from obtaining a sufficiently large training set. However, if this can be overcome, then one only needs to retrain our DPNNet for application to 3D disks.

6.2. Future Prospects

Our DPNNet demonstrates a simple but powerful application of deep learning in constraining the mass of hidden exoplanets. With more realistic simulations and a larger training data set, the network can possibly be used to characterize additional properties like gas surface densities and gap depths from observed PPDs. Improved and new physics can be accommodated by adding more variables to the training set, as far as the neural network is concerned. In the future, with the emergence of more powerful computers, it will be possible to have adequate data to train the network with synthetic images generated from 3D radiative transfer and hydrodynamical simulations, allowing fast and direct comparisons between models and observations. This will potentially make the network versatile enough to entirely replace hydrodynamical simulations for observers and modelers.

7. Conclusions

In this paper, we design a deep neural network, DPNNet, to estimate the mass of a planet from the dust gap it opens in a protoplanetary disk. The important takeaways are summarized as follows:

1. The DPNNet can directly infer planet mass from gap width observed in dust emission and other disk features. It takes as input dust gap widths, gas disk properties (aspect ratio, viscosity, surface density profile), and dust properties (abundance, Stokes numbers) from observations and estimates the embedded planet mass.
2. In contrast to previous empirical models (e.g., Dong et al. 2015; Kanagawa et al. 2016; Lodato et al. 2019), the DPNNet additionally accounts for changes in the predicted planet mass due to variation in dust abundance, Stokes number, and disk surface density profile.
3. The DPNNet is applied to a test data set to evaluate its effectiveness. The DPNNet-predicted mass is closely correlated with the true planet mass. Our current model uncertainty is $13 M_{\oplus}$, but this can be improved in the future with a larger training set.
4. We deploy our DPNNet to infer planet masses from the dust gaps observed in the protoplanetary disks around HL

Tau and AS 209. We find planet masses of 80 $M_\oplus$, 63 $M_\oplus$, and 70 $M_\oplus$ in gaps at 10 au, 30 au, and 80 au, respectively, in the HL Tau disk. Similarly, for the AS 209 disk, we infer embedded planet masses of 39 $M_\oplus$ and 45 $M_\oplus$ at the 9 au and 100 au gaps, respectively. These estimates are in close agreement with results from other models based on specialized disk–planet simulations. The results are summarized in Table 2.

We thank the anonymous referee for constructive comments. This work is supported by the Ministry of Science and Technology of Taiwan (grant number 107-2112-M-001-043-MY3). Numerical simulations were performed on the TIARA cluster at ASIAA, as well as the TWCC cluster at the National Center for High-performance Computing (NCHC). We are grateful to the NCHC for computing time, facilities, and support.

Appendix

Table A1 gives the statistical distribution of the simulation data set before and after the screening. The raw data set corresponds to the data from the entire 1100 simulations. The filtered data set is obtained after the initial cut to remove runs that do not open up detectable gaps or have more that two gaps. While the variables (M_P, α, h_0, σ, S_t, ϵ) are the input parameters for the simulations, the dust gaps width are measured from the dust density profile from the simulation outputs. Figure A1 shows training and the validation loss, the MSE, as a function of training epoch for the DPNNet. As is evident, the validation loss gradually decreases and starts to flatten after 600–700 epoch, whereas the training loss is still decreasing indicating over fitting. We implement early stopping and stop the iterations around 800 epochs.

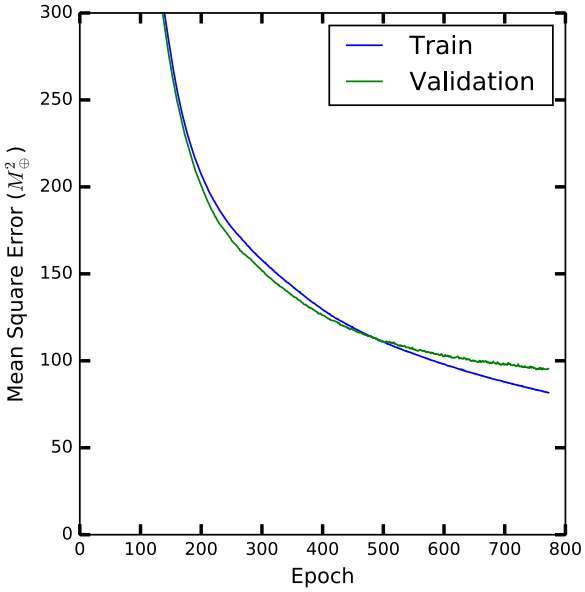

Figure A1. Training and validation loss as a function of training epoch for the DPNNet for computation of the planet mass.

Table A1
Distribution Statistics of Parameters for the Simulation Data

		Raw Data				Filtered Data			
Name	Notation	Max.	Min.	Mean	Std	Max.	Min.	Mean	Std
Planet mass in Earth masses	$M_P/\ M_\oplus$	120.0	8.1	64.2	32.4	120.0	8.2	66.8	31.5
Disk aspect ratio	h_0	0.099	0.025	0.040	0.011	0.091	0.025	0.038	0.009
Disk surface density profile	σ	1.20	0.05	0.54	0.29	1.20	0.05	0.53	0.29
Disk viscosity parameter	α ($\times 10^{-3}$)	10	0.11	5.03	2.86	10	0.13	5.13	2.77
Global dust-to-gas ratio	ϵ	0.10	0.01	0.05	0.03	0.10	0.01	0.05	0.03
Particle Stokes numbers	S_t	0.100	0.001	0.051	0.029	0.10	0.001	0.053	0.028
Dust Gap 1	w_{d1}	0.94	0.00	0.67	0.22	0.94	0.10	0.74	0.11
Dust Gap 2	w_{d2}	0.69	0.00	0.03	0.09	0.47	0.00	0.02	0.07

Note. Raw data ($N = 1100$) correspond to the simulation data set. Filtered data ($N = 976$) are obtained after the initial screening (see Section 3.2 for details).

ORCID iDs

Sayantan Auddy https://orcid.org/0000-0003-3784-8913
Min-Kai Lin https://orcid.org/0000-0002-8597-4386

References

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://www.tensorflow.org/
Alibert, Y., & Venturini, J. 2019, A&A, 626, 21
ALMA Partnership, Brogan, C. L., Pérez, L. M., et al. 2015, ApJL, 808, L3
Andrews, S. M., Huang, J., Pérez, L. M., et al. 2018, ApJL, 869, L41
Andrews, S. M., Wilner, D. J., Zhu, Z., et al. 2016, ApJL, 820, L40
Bae, J., Zhu, Z., & Hartmann, L. 2017, ApJ, 850, 201
Barge, P., Ricci, L., Carilli, C. L., & Previn-Ratnasingam, R. 2017, A&A, 605, A122
Batalha, N. M., Rowe, J. F., Bryson, S. T., et al. 2013, ApJS, 204, 24
Benítez-Llambay, P., Krapp, L., & Pessah, M. E. 2019, ApJS, 241, 25
Benítez-Llambay, P., & Masset, F. 2016, ApJS, 223, 11
Cassan, A., Kubas, D., Beaulieu, J. P., et al. 2012, Natur, 481, 167
Chorowski, J. K., Bahdanau, D., Serdyuk, D., Cho, K., & Bengio, Y. 2015, arXiv:1506.0750, http://papers.nips.cc/paper/5847-attention-based-models-for-speech-recognition.pdf
Clarke, C. J., Tazzari, M., Juhasz, A., et al. 2018, ApJ, 866, 6
Crida, A., Morbidelli, A., & Masset, F. 2006, Icar, 181, 587
de Val-Borro, M., Artymowicz, P., D'Angelo, G., & Peplinski, A. 2007, A&A, 471, 1043
de Val-Borro, M. e. 2006, MNRAS, 370, 529
Dipierro, G., Laibe, G., Price, D. J., & Lodato, G. 2016, MNRAS, 459, 1
Dipierro, G., Price, D., Laibe, G., et al. 2015, MNRAS, 453, 73
Dong, R., & Fung, J. 2017, ApJ, 835, 146
Dong, R., Li, S., Chiang, E., & Li, H. 2017, ApJ, 843, 127
Dong, R., Zhu, Z., & Whitney, B. 2015, ApJ, 809, 93
Drażkowska, J., & Alibert, Y. 2017, A&A, 608, A92
Duffell, P. C. 2015, ApJL, 807, L11
Duffell, P. C., & Macfadyen, A. I. 2013, ApJ, 769, 41
Fedele, D., Tazzari, M., Booth, R., et al. 2018, A&A, 610, 24
Fischer, D. A., Howard, A. W., Laughlin, G. P., et al. 2014, in Protostars and Planets VI, ed. H. Beuther et al. (Tucson, AZ: Univ. Arizona Press), 715
Fulton, B. J., Petigura, E. A., Howard, A. W., et al. 2017, AJ, 154, 109
Fung, J., & Chiang, E. 2016, ApJ, 832, 105
Fung, J., Shi, J.-M., & Chiang, E. 2014, ApJ, 782, 88
Goldreich, P., & Tremaine, S. 1980, ApJ, 241, 425
Gonzalez, J. F., Laibe, G., Maddison, S. T., Pinte, C., & Ménard, F. 2015, MNRAS, 454, L36
Goodfellow, I., Bengio, Y., & Courville, A. 2016, Deep Learning (Cambridge, MA: MIT Press)
Grigorescu, S., Trasnea, B., Cocias, T., & Macesanu, G. 2019, arXiv:1910.07738
Guzmán, V. V., Huang, J., Andrews, S. M., et al. 2018, ApJ, 869, 48
Huang, J., Andrews, S. M., Cleeves, L. I., et al. 2018c, ApJ, 852, 122
Huang, J., Andrews, S. M., Dullemond, C. P., et al. 2018b, ApJL, 869, L42
Huang, J., Andrews, S. M., Pérez, L. M., et al. 2018a, ApJL, 869, L43
Iman, R. L., Helton, J. C., & Campbell, J. E. 1981, JQT, 13, 174
Jacquet, E., Balbus, S., & Latter, H. 2011, MNRAS, 415, 3591
Jin, S., Li, S., Isella, A., Li, H., & Ji, J. 2016, ApJ, 818, 76
Johansen, A., Blum, J., Tanaka, H., et al. 2014, in Protostars and Planets VI, ed. H. Beuther et al. (Tucson, AZ: Univ. Arizona Press), 547
Johansen, A., Youdin, A., & Klahr, H. 2009, ApJ, 697, 1269
Kanagawa, K. D., Muto, T., Tanaka, H., et al. 2015a, ApJL, 806, L15
Kanagawa, K. D., Muto, T., Tanaka, H., et al. 2016, PASJ, 68, 43
Kanagawa, K. D., Tanaka, H., Muto, T., Tanigawa, T., & Takeuchi, T. 2015b, MNRAS, 448, 994
Keppler, M., Benisty, M., Müller, A., et al. 2018, A&A, 617, A44
Kley, W., & Dirksen, G. 2006, A&A, 447, 369
Koller, J., Li, H., & Lin, D. N. C. 2003, ApJ, 596, L91
Krizhevsky, A., Sutskever, I., & Hinton, G. E. 2012, in Advances in Neural Information Processing Systems 25, ed. F. Pereira et al. (New York: Commun. ACM), 84
LeCun, Y., Bengio, Y., & Hinton, G. 2015, Natur, 521, 436
Li, H., Li, S., Koller, J., et al. 2005, ApJ, 624, 1003
Lin, D. N. C., & Papaloizou, J. C. B. 1993, in Protostars and Planets III, ed. E. H. Levy & J. I. Lunine (Tucson, AZ: Univ. Arizona Press), 749
Liu, Y., Dipierro, G., Ragusa, E., et al. 2019, A&A, 622, A75
Lodato, G., Dipierro, G., Zagusa, E., et al. 2019, MNRAS, 486, 453
Long, F., Pinilla, P., Herczeg, G. J., et al. 2018, ApJ, 869, 17
McKay, M. D., Beckman, R. J., & Conover, W. J. 1979, Technometrics, 21, 239
Miranda, R., & Rafikov, R. R. 2019, ApJL, 878, L9
Miranda, R., & Rafikov, R. R. 2020, ApJ, 892, 65
Momose, M., Morita, A., Fukagawa, M., et al. 2015, PASJ, 67, 83
Morbidelli, A. 2020, A&A, 638, A1
Ng, A. Y. 2004, in Proc. Twenty-First International Conference on Machine Learning, Feature Selection, L1 vs. L2 Regularization, and Rotational Invariance (New York: Association for Computing Machinery)
Nwankpa, C., Ijomah, W., Gachagan, A., & Marshall, S. 2018, arXiv:1811.03378
Paardekooper, S.-J., & Papaloizou, J. C. B. 2009, MNRAS, 394, 2297
Papaloizou, J. C. B. 2002, A&A, 388, 615
Pedregosa, F., Varoquaux, G., Gramfort, A., et al. 2011, Journal of Machine Learning Research, 12, 2825
Pérez, L. M., Benisty, M., Andrews, S. M., et al. 2018, ApJL, 869, L50
Pinilla, P., Pohl, A., Stammler, S. M., & Birnstiel, T. 2017, ApJ, 845, 68
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
Raymond, S. N., & Morbidelli, A. 2020, arXiv:2002.05756
Rosotti, G. P., Juhasz, A., Booth, R. A., & Clarke, C. J. 2016, MNRAS, 459, 2790
Ruder, S. 2016, arXiv:1609.04747
Schmidhuber, J. 2015, NN, 61, 85
Shakura, N. I., & Sunyaev, R. A. 1973, A&A, 24, 337
Simon, J. B., & Armitage, P. J. 2014, ApJ, 784, 15
Takahashi, S. Z., & Inutsuka, S.-i. 2014, ApJ, 794, 55
Toomre, A. 1964, ApJ, 139, 1217
van der Plas, G., Wright, C. M., Ménard, F., et al. 2017, A&A, 597, A32
Weidenschilling, S. J. 1977, MNRAS, 180, 57
Winn, J. N., & Fabrycky, D. C. 2015, ARA&A, 53, 409
Yang, Z., Yang, D., Dyer, C., et al. 2016, in Proc. 2016 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Hierarchical Attention Networks for Document Classification (San Diego, CA: Association for Computational Linguistics), 1480
Yen, H.-W., Liu, H. B., Gu, P.-G., et al. 2016, ApJL, 820, L25
Youdin, A. N. 2011, ApJ, 731, 99
Yu, L., Donati, J. F., Hébrard, E. M., et al. 2017, MNRAS, 467, 1342
Zhang, K., Blake, G. A., & Bergin, E. A. 2015, ApJ, 806, 7
Zhang, S., Zhu, Z., Huang, J., et al. 2018, ApJL, 869, L47