PPDONet: Deep Operator Networks for Fast Prediction of Steady-State Solutions in Disk-Planet Systems

Shunyuan Mao[1] **Ruobing Dong**[1] **Lu Lu**[2] **Kwang Moo Yi**[3] **Sifan Wang**[4] **Paris Perdikaris**[5]

Abstract

We have created a tool called the Protoplanetary Disk Operator Network (PPDONet) that quickly predicts disk-planet interactions in protoplanetary disks. Our tool uses Deep Operator Networks (DeepONets), a type of neural network that learns non-linear operators to accurately represent both deterministic and stochastic differential equations. PPDONet maps three key parameters in a disk-planet system – the Shakura & Sunyaev viscosity α, the disk aspect ratio h_0, and the planet-star mass ratio q – to the steady-state solutions for disk surface density, radial velocity, and azimuthal velocity. We've validated the accuracy of PPDONet's solutions with an extensive array of tests. Our tool can calculate the result of a disk-planet interaction for a given system in under a second using a standard laptop. PPDONet is publicly accessible for use.

1. Introduction

Planets form in protoplanetary disks surrounding newborn stars. These disks are made of gas and dust, and can extend to hundreds of AU. Gradually, under the influence of gravitational and aerodynamic forces, the material in these disks starts to clump together, forming larger entities that ultimately develop into planets. During formation, planets gravitationally interact with their host disks, generating large-scale structures such as gaps, spiral arms, and dust clumps. This gravitational interplay, which can be modelled using fluid dynamics, is termed disk-planet interaction (Kley & Nelson, 2012). Simulating the evolution of protoplanetary disks and disk-planet interactions is vital for understanding various aspects of planetary formation, including the accretion and dispersion of disk materials (Tabone et al., 2022), the interpretation of observed disk structures (Dong et al., 2015; Dong & Fung, 2017; Liu et al., 2018), the determination of planetary properties (Fung et al., 2014; Fung & Dong, 2015; Zhang et al., 2018), and the study of planets' orbital evolution (Paardekooper et al., 2022).

For decades, numerical methodologies have been established to simulate the evolution of disks and disk-planet interactions (Paardekooper et al., 2022). Nonetheless, these methods are typically computing-intensive. For instance, a simulation of a 2D disk with a resolution of 270 (r) by 810 (θ) for $2,000$ orbits necessitates 20 GPU hours (Fung et al., 2014). Consequently, simulating numerous disk-planet systems or probing the parameter space of their interactions can be resource-intensive. Furthermore, given the rapidly increasing number of observed disks (Benisty et al., 2022; Bae et al., 2022), replicating all these observations through simulations is becoming increasingly impractical.

There exist two fundamental reasons underpinning the high computational cost of extensive disk simulations. Primarily, every numerical simulation is executed independently. That is, despite possible similarities in the outcomes of simulations with slightly varied configurations, these outcomes aren't reused, leading to a computational cost that increases linearly with the number of simulations performed. Secondly, inherent computational complexity arises due to the nature of these simulations. Achieving accurate outcomes in studies of disk evolution and disk-planet interactions necessitates a significant progression through numerical time increments, whose duration is governed by the Courant–Friedrichs–Lewy condition. Typically, A disk-planet system might undergo evolution over thousands of orbits to attain a steady state, correlating to millions of these time units.

In this study, we apply Machine Learning (ML) to overcome both issues, allowing us to "learn" from similar simulations and eliminate the requirement for time-stepping execution.

---

*Equal contribution [1]Department of Physics and Astronomy, University of Victoria, Victoria, BC V8P 5C2, Canada [2]Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA [3]Department of Computer Science, University of British Columbia, Vancouver, BC V6T 1Z4, Canada [4]Graduate Group in Applied Mathematics and Computational Science, University of Pennsylvania, Philadelphia, PA 19104, USA [5]Department of Mechanical Engineering and Applied Mechanics, University of Pennsylvania, Philadelphia, PA 19104, USA. Correspondence to: Shunyuan Mao <symao@uvic.ca>, Ruobing Dong <rbdong@uvic.ca>.

We propose a novel tool, built upon DeepONets, capable of immediately predicting the steady-state disk structure, encompassing both surface density and velocities, in a disk-planet system. We train DeepONets (§3) using hydrodynamic simulations derived from numerical solvers, and test the networks by reproducing empirical relationships observed in disk-planet interactions (§4).

2. Related Work

Recently, ML techniques have been applied to model disk-planet systems, although not for the same problem as we are interested in—to predict the surface density and velocity fields of a disk-planet system. Auddy & Lin (2020) used a deep neural network to predict planet mass based on 1D surface density profiles of gaps. Further expanding on this, Auddy et al. (2021) employed a convolutional neural network (CNN) to deduce planet mass from the 2D dust density distribution. Similarly, Zhang et al. (2022) implemented a comparable neural network to estimate planet mass from synthetic observations. However, the focus of these techniques is on deducing scalar outputs from disk morphologies rather than predicting the distribution of physical quantities.

DeepONets, introduced in Lu et al. (2021), offer a novel framework that substantially cuts down the computational expenses of simulating physical systems. Specifically, they are adept at learning non-linear operators which encapsulate deterministic and stochastic differential equations. These networks are equipped to predict solutions to differential equations characterized by parametric boundary conditions, initial conditions, or forcing terms. Consequently, DeepONets exhibit enhanced generalization and rapid convergence relative to the quantity of training data, as demonstrated in Lu et al. (2021).

3. Implementation

3.1. DeepONets

Our study leverages the DeepONet architecture (Lu et al., 2021), which we summarize briefly. DeepONet employs a neural network to approximate a mathematical operator, denoted as G. In the context of disk-planet systems, this operator G transforms a vector of scalar parameters, $\vec{p}$, into a distribution function, $G(\vec{p})$, which can be evaluated at any given location. Specifically, our networks incorporate scalar parameters $\vec{p}$ such as the Shakura & Sunyaev viscosity α, disk aspect ratio h_0, and the planet-star mass ratio q. The resulting function $G(\vec{p})(r, \theta)$ yields the surface density or velocity at a specific point (r, θ). Consequently, the DeepONet structure includes two types of inputs: scalar parameters, $\vec{p}$, and coordinates (r, θ), with a single output that represents surface density, radial velocity, or azimuthal velocity.
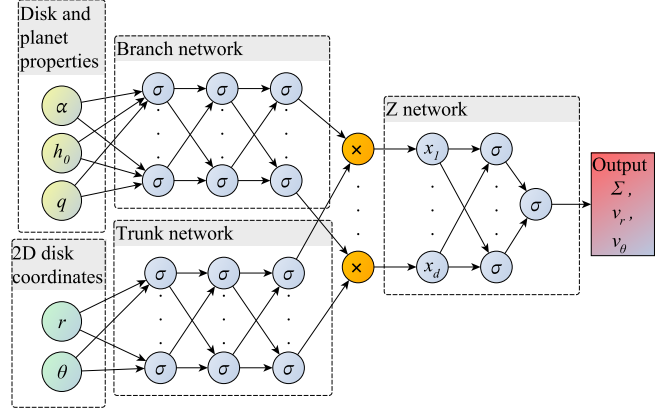

Figure 1. PPDONet architecture used in this work. A fully connected subnetwork, called "branch network", encodes scalar parameters, the disk and planet properties, while another subnetwork, called "trunk network", encodes the coordinates. We produce the inputs x_i, $i = 1, ..., d$, $d = 50$, for the third network ("Z network") by element-wise multiplication of the outputs from the two subnetworks. The Z network is a single-hidden-layer network whose output represents surface density or velocity. The σ in blue nodes are neurons in layers.

Figure 1 delineates the PPDONet architecture, with scalar parameters $\vec{p}$ being processed by a "branch network" subnetwork, which is fully connected with four hidden layers each hosting 100 neurons. Simultaneously, the coordinate inputs are managed by the "trunk network," another fully-connected subnetwork with five layers each having 256 neurons. The inputs x_i, $i = 1, ..., d$, $d = 50$, for the third network (the "Z network"), shown in Figure 1, are derived through element-wise multiplication of the outputs of the branch and trunk networks. The Z network, equipped with a single hidden layer, generates the final output representing surface density or velocity. For fitting surface density and azimuthal velocity, we use the self-scalable Tanh activation function (Stan) (Gnanasambandam et al., 2022) due to its fast convergence and minimal error rates. However, in the case of learning radial velocity, the Tanh activation function exhibits better performance than Stan.

3.2. Hydrodynamic simulations for training, validating, and testing neural networks

We generate 768 hydrodynamic simulations of disk-planet systems using the FARGO3D code (Masset, 2000; Benítez-Llambay & Masset, 2016), varying α, h_0, and q within the parameter space outlined in Table 1. We then partition the simulations into training, validation and testing datasets. For more comprehensive details, please refer to Appendix A.

Table 1. Disk and planet parameter space

PARAMETER	MINIMUM	MAXIMUM
α	3×10^{-4}	0.1
h_0	0.05	0.1
q	5×10^{-5}	2×10^{-3}

3.3. Network training

We employ three distinct neural networks to approximate the steady-state solution for surface density Σ, radial velocity v_r, and azimuthal velocity v_θ. Considering the broad range spanned by some inputs, we transform them into logarithmic scales. We then normalize both the scalar parameters and coordinate inputs. To enhance the learning process, we introduce two modifications: 1) we convert surface density to a logarithmic scale, and 2) we remove the background from the radial and azimuthal velocities. As a result, our loss function for surface density takes the form of

$$L_\Sigma = \frac{1}{N} \sum_{i=1}^{N} \left[\log(\Sigma_i^{\mathrm{pred}}) - \log(\Sigma_i^{\mathrm{truth}}) \right]^2, \quad (1)$$

while the loss functions for velocities are expressed as

$$L_{v_r} = \frac{1}{N} \sum_{i=1}^{N} \left[v_{r\,i}^{\mathrm{pred}} - v_{r\,i}^{\mathrm{truth}} \right]^2 \quad (2)$$

and

$$L_{v_\theta} = \frac{1}{N} \sum_{i=1}^{N} \left[v_{\theta\,i}^{\mathrm{pred}} - v_{\theta\,i}^{\mathrm{truth}} \right]^2. \quad (3)$$

The indices i go through all the grid points in all the simulations in a training batch [1]. We treat FARGO3D outputs as the "ground truth," denoted by the superscript $^{\mathrm{truth}}$, while the superscript $^{\mathrm{pred}}$ indicates neural network predictions. The hyperparameters employed in our networks and training routines can be found in Table 2 (Appendix B).

4. Tests

We present a series of tests. In §4.1, we compare the 2D maps of predicted surface density and velocity with the actual data generated by FARGO3D. Subsequently, in §4.2, we investigate the effectiveness of our model on a collection of disks by recreating the gap depth relationship documented in earlier research. More tests are in Appendix C.1, C.2, and C.3.

4.1. 2D maps of surface density and velocities

Figure 2 illustrates a typical case ($\alpha = 5.2 \times 10^{-4}$, $h_0 = 0.053$, and $q = 1.6 \times 10^{-3}$) emphasizing the gaps. The left column displays the FARGO3D generated ground truth, the middle column presents the PPDONet predictions, and the two right columns highlight the differences or ratios between them. To simplify the comparison in the perturbed radial and azimuthal velocity, we deduct their initial values from the corresponding panels. We compute the ratio between the predicted and actual surface density to emphasize the gap region. For velocities, which can possess both positive and negative values, we exhibit the absolute errors.

A visual inspection of the comparison between the model predictions and actual data (panels a, c, e vs b, d, f) reveals negligible discrepancies in the location and contrast of both the density waves and gaps, despite the sharp morphology of these perturbations compared to the background.

4.2. Empirical dependence of the gap depth on disk and planet properties

Gap Depth (Σ_{gap}), the average surface density inside a gap, depends on α, h_0, and q. Fung et al. (2014) simulated 21 disk-planet systems and established an empirical correlation for planet masses less than five Jupiter masses:

$$\Sigma_{\mathrm{gap}}/\Sigma_0 = 0.14 \left(\frac{q}{0.001} \right)^{-2.16} \left(\frac{\alpha}{0.01} \right)^{1.41} \left(\frac{h_0}{0.05} \right)^{6.61}. \quad (4)$$

Using PPDONet, we generate 500 disks on a laptop within a ten-minute span and determine the gap depth for these disks. Adopting the same methodology as Fung et al. (2014), we average the surface density within the annulus $|r - r_{\mathrm{p}}| < 2\max(R_{\mathrm{H}}, h)$, excluding the area $|\phi - \phi_{\mathrm{p}}| < 2\max(R_{\mathrm{H}}, h)/r_{\mathrm{p}}$. We fit the data with a power-law similar to Fung et al. (2014)

$$\begin{aligned} \Sigma_{\mathrm{gap}}/\Sigma_0 &= 0.13 \pm 0.02 \left(\frac{q}{0.001} \right)^{-2.50 \pm 0.01} \\ &\times \left(\frac{\alpha}{0.01} \right)^{1.421 \pm 0.007} \left(\frac{h_0}{0.05} \right)^{7.26 \pm 0.03} \end{aligned} \quad (5)$$

and measure the standard errors of the estimate, namely standard deviation of the error terms, using IBM SPSS Statistics. We get an adjusted R^2 equaling 0.995, indicating an excellent fit. Figure 3 illustrates the data, the prediction from Equation (4), and our fit. Due to the construction of the horizontal axis, we represent the prediction of Equation (4) with dots rather than a curve. Although our fit's power-law indices vary slightly from Fung et al. (2014)'s, both show the same trend: more massive planets create deeper gaps in

---

[1]To compare with FARGO3D, the grid points for surface density are cell-centered, while velocities are face-centered (see Benítez-Llambay & Masset, 2016, §2.3)
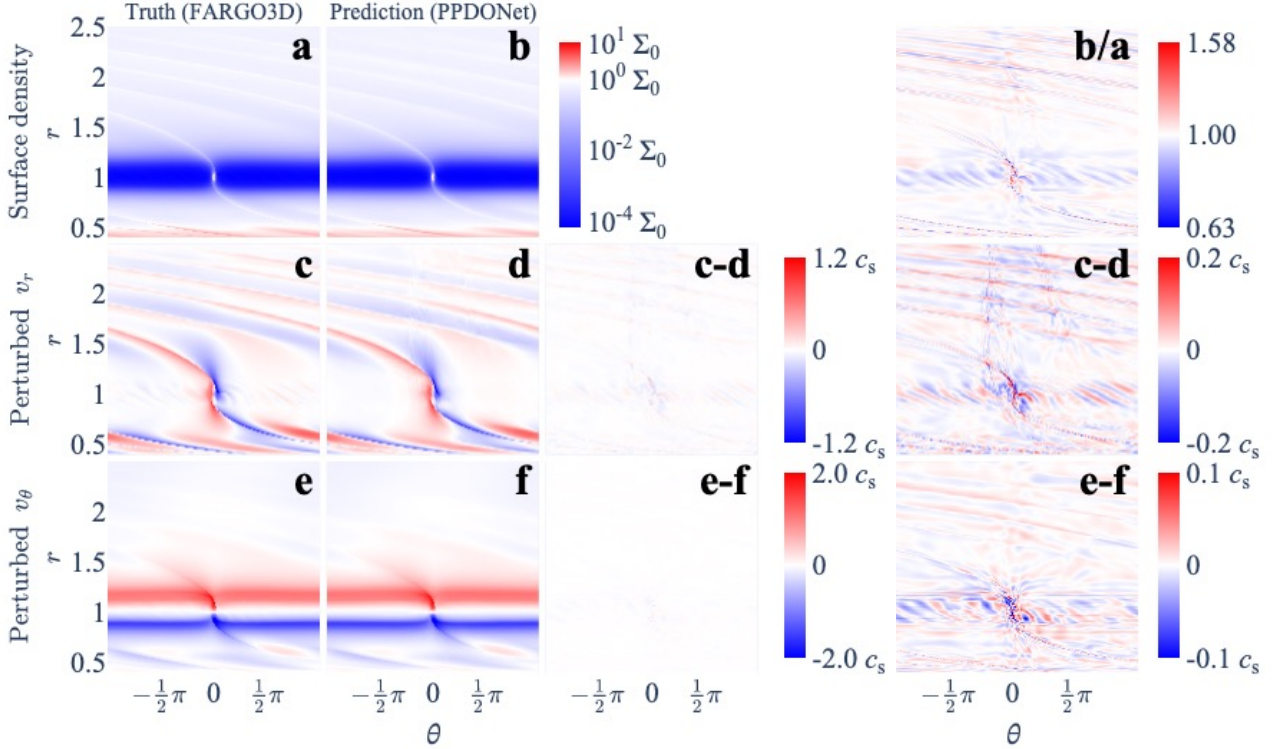
Figure 2. Two-dimensional comparisons of predicted surface density and velocity distribution in polar coordinates (radius r and azimuth θ). $(\alpha, h_0, q) = \left(5.2 \times 10^{-4}, 0.053, 1.6 \times 10^{-3}\right)$. Left (a, c, and e): Ground truth generated from FARGO3D simulations. Middle (b, d, and f): Neural network predictions. For the two left columns, we subtract the initial value from the two velocities to highlight the perturbations. Two right columns: Differences or ratios. Surface density differences are measured by the ratio $\Sigma^{\mathrm{pred}}/\Sigma^{\mathrm{truth}}$, so that the gap region is highlighted. Velocity differences are shown in absolute errors.
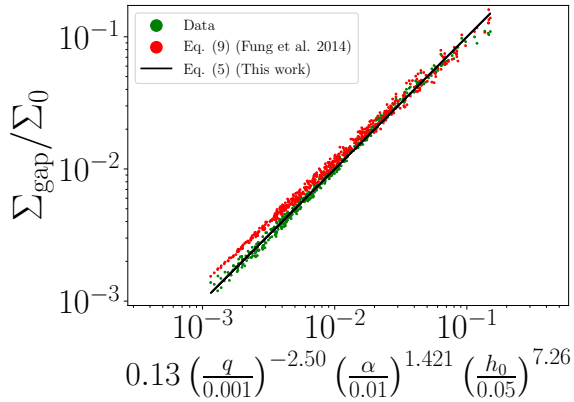

Figure 3. The Empirical relationship of the gap depth Σ_{gap} as a function of α, h_0, and q. We plot the measured gap depth from PPDONet-simulated disks (green) and the best-fitting relationship (black). We compare them with the relationship from Fung et al. (2014) (red).

disks with lower viscosity and aspect ratio. We highlight that with hundreds of disk models, we can not only derive

a fitting function but also ascertain the *uncertainties* in the fitted parameters, a challenging feat with only tens of disk models produced by numerical solvers.

5. SUMMARY

We introduce PPDONet, a machine learning tool, which employs Deep Operator Networks (DeepONets; Lu et al. 2021) to accurately predict the steady-state solutions of disk-planet interactions in parameterized protoplanetary disks. PPDONet's training utilizes 448 disk models derived from FARGO3D simulations. This trained tool maps three scalar parameters – Shakura & Sunyaev viscosity α, disk aspect ratio h_0, and planet-star mass ratio q – to the steady-state solutions of disk surface density, radial velocity, and azimuthal velocity. As of now, PPDONet stands as the sole public tool capable of solving this forward problem. It predicts the structures of 500 disk-planet systems within mere minutes on a standard laptop, drastically outpacing traditional numerical simulations.

To evaluate the performance of PPDONet, we compare the 2D maps of predicted surface density and velocity with those

produced by FARGO3D (§4.1). The two are consistent with each other with little visible difference. We use PPDONet to generate a sample of disk-planet interaction containing 500 disks and revisit the empirical relationship between gap depth and α, h_0, and q (§4.2). We also reproduce more empirical relationships in appendix sections C.2 and C.3. Overall, we recover previously found correlations. In addition, thanks to the one to two orders of magnitude larger samples that PPDONet is able to quickly produce compared with those previously produced using conventional numerical solvers, we are able to constrain the uncertainties in the fitting parameters, a nearly impossible task in the past due to small sample sizes.

References

Auddy, S. and Lin, M.-K. A machine learning model to infer planet masses from gaps observed in protoplanetary disks. *The Astrophysical Journal*, 900(1):62, 2020.

Auddy, S., Dey, R., Lin, M.-K., and Hall, C. Dpnnet-2.0. i. finding hidden planets from simulated images of protoplanetary disk gaps. *The Astrophysical Journal*, 920(1):3, 2021.

Bae, J., Teague, R., Andrews, S. M., Benisty, M., Facchini, S., Galloway-Sprietsma, M., Loomis, R. A., Aikawa, Y., Alarcón, F., Bergin, E., et al. Molecules with alma at planet-forming scales (maps): A circumplanetary disk candidate in molecular-line emission in the as 209 disk. *The Astrophysical Journal Letters*, 934(2):L20, 2022.

Benisty, M., Dominik, C., Follette, K., Garufi, A., Ginski, C., Hashimoto, J., Keppler, M., Kley, W., and Monnier, J. Optical and near-infrared view of planet-forming disks and protoplanets. *arXiv preprint arXiv:2203.09991*, 2022.

Benítez-Llambay, P. and Masset, F. S. Fargo3d: a new gpu-oriented mhd code. *The Astrophysical Journal Supplement Series*, 223(1):11, 2016.

Dong, R. and Fung, J. What is the mass of a gap-opening planet? *The Astrophysical Journal*, 835(2):146, 2017.

Dong, R., Zhu, Z., and Whitney, B. Observational signatures of planets in protoplanetary disks. i. gaps opened by single and multiple young planets in disks. *The Astrophysical Journal*, 809(1):93, 2015.

Fung, J. and Dong, R. Inferring planet mass from spiral structures in protoplanetary disks. *The Astrophysical Journal Letters*, 815(2):L21, 2015.

Fung, J., Shi, J.-M., and Chiang, E. How empty are disk gaps opened by giant planets? *The Astrophysical Journal*, 782(2):88, 2014.

Glorot, X. and Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256. JMLR Workshop and Conference Proceedings, 2010.

Gnanasambandam, R., Shen, B., Chung, J., Yue, X., et al. Self-scalable tanh (stan): Faster convergence and better generalization in physics-informed neural networks. *arXiv preprint arXiv:2204.12589*, 2022.

Kanagawa, K. D., Muto, T., Tanaka, H., Tanigawa, T., Takeuchi, T., Tsukagoshi, T., and Momose, M. Mass constraint for a planet in a protoplanetary disk from the gap width. *Publications of the Astronomical Society of Japan*, 68(3):43, 2016.

Kingma, D. P. and Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Kley, W. and Nelson, R. Planet-disk interaction and orbital evolution. *Annual Review of Astronomy and Astrophysics*, 50:211–249, 2012.

Liu, S.-F., Jin, S., Li, S., Isella, A., and Li, H. New constraints on turbulence and embedded planet mass in the hd 163296 disk from planet–disk hydrodynamic simulations. *The Astrophysical Journal*, 857(2):87, 2018.

Lu, L., Jin, P., Pang, G., Zhang, Z., and Karniadakis, G. E. Learning nonlinear operators via deeponet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, 3(3):218–229, 2021.

Masset, F. Fargo: A fast eulerian transport algorithm for differentially rotating disks. *Astronomy and Astrophysics Supplement Series*, 141(1):165–173, 2000.

Miranda, R. and Rafikov, R. R. Gaps and rings in protoplanetary disks with realistic thermodynamics: The critical role of in-plane radiation transport. *The Astrophysical Journal*, 904(2):121, 2020.

Paardekooper, S.-J., Dong, R., Duffell, P., Fung, J., Masset, F. S., Ogilvie, G., and Tanaka, H. Planet-disk interactions. *arXiv preprint arXiv:2203.09595*, 2022.

Sobol', I. M. On the distribution of points in a cube and the approximate evaluation of integrals. *Zhurnal Vychislitel'noi Matematiki i Matematicheskoi Fiziki*, 7(4): 784–802, 1967.

Tabone, B., Rosotti, G. P., Lodato, G., Armitage, P. J., Cridland, A. J., and van Dishoeck, E. F. Mhd disc winds can reproduce fast disc dispersal and the correlation between accretion rate and disc mass in lupus. *Monthly Notices of the Royal Astronomical Society: Letters*, 512 (1):L74–L79, 2022.

Wu, C., Zhu, M., Tan, Q., Kartha, Y., and Lu, L. A comprehensive study of non-adaptive and residual-based adaptive sampling for physics-informed neural networks. *Computer Methods in Applied Mechanics and Engineering*, 403:115671, 2023.

Yun, H. G., Kim, W.-T., Bae, J., and Han, C. Properties of density and velocity gaps induced by a planet in a protoplanetary disk. *The Astrophysical Journal*, 884(2): 142, 2019.

Zhang, S., Zhu, Z., Huang, J., Guzmán, V. V., Andrews, S. M., Birnstiel, T., Dullemond, C. P., Carpenter, J. M., Isella, A., Pérez, L. M., et al. The disk substructures at high angular resolution project (dsharp). vii. the planet–disk interactions interpretation. *The Astrophysical Journal Letters*, 869(2):L47, 2018.

Zhang, S., Zhu, Z., and Kang, M. Pgnets: planet mass prediction using convolutional neural networks for radio continuum observations of protoplanetary discs. *Monthly Notices of the Royal Astronomical Society*, 510(3):4473–4484, 2022.

A. Hydrodynamic Simulations

We focus on large-scale structures, such as gaps and spiral arms, induced by a single planet on a fixed circular orbit in gaseous disks. The disk's initial profiles are described by the equations:

$$\Sigma \;=\; \Sigma_0 \left(r/r_{\mathrm{p}}\right)^{-1/2}, \tag{6a}$$

$$v_r \;=\; -\frac{3}{2}\alpha h_0{}^2 \sqrt{\frac{GM_\odot \left(1+q\right)}{r}}, \tag{6b}$$

$$v_\theta \;=\; \sqrt{1-\frac{3}{2}h_0{}^2}\sqrt{\frac{GM_\odot \left(1+q\right)}{r}}, \tag{6c}$$

where Σ is the surface density, v_r is the radial velocity, v_θ is the azimuthal velocity, α is the Shakura & Sunyaev viscosity, and h_0 is the disk aspect ratio. Both α and h_0 are kept constant throughout the disk. The models are developed for a fixed initial radial surface density profile and a radially constant disk aspect ratio. The planet is at $r = r_{\mathrm{p}}$ with mass $m_{\mathrm{p}} = qM_\odot$. The initial surface density at r_{p} is $\Sigma_0 = 1$. This setup ensures our model corresponds to planet-free steady-state accretion disks (Fung et al., 2014). The boundary conditions are fixed and determined by initial values, which ensure a constant mass inflow (Fung et al., 2014).

The outcomes of disk-planet interaction in our simulations are determined by three parameters: α, h_0, and q. Massive planets open deep gaps, while large viscosity and aspect ratio hinder the opening of deep gaps. With $q > 2 \times 10^{-3}$, or $\alpha < 3 \times 10^{-4}$, a disk may develop vortices or other asymmetric and time-varying structures (Fung et al., 2014). In this work, we only focus on disks capable of reaching a steady state with parameters bounded by Table 1.

To collect data for training neural networks, choose the best ML models, and test their performance on unseen parameters, we generate 768 FARGO3D (Masset, 2000; Benítez-Llambay & Masset, 2016) hydrodynamic simulations with $r \times \theta = 381 \times 1143$ resolution. We sample α, h_0, and q from Sobol sequences (Sobol', 1967), quasi-random low-discrepancy sequences effective in generating inputs for machine learning tasks (Wu et al., 2023). We divide the 768 FARGO3D simulations into three groups with 448, 64, and 256 cases for training, validation, and testing, respectively. In training, we compare neural network predictions with simulations to formalize loss functions. In testing, we measure errors on unseen simulations to assess the generalization of our neural networks. All simulations are run for $0.314\tau_\nu$ to reach (quasi) steady state, where τ_ν is the disk viscous timescale

$$\tau_\nu \approx \frac{r_{\mathrm{p}}^2}{\nu} = \frac{1}{\alpha c_{\mathrm{s}} h} = \frac{1}{\alpha h_0{}^2}. \tag{7}$$

B. Hyperparameters of PPDONet

The hyperparameters used in our networks and training processes are listed in Table 2.

B.1. Training Dataset Size and Test Error

The predictive accuracy of the network is contingent on the sizes of the training datasets (Lu et al., 2021, Figure 4), and can be problem dependent. To investigate the data size dependency in our problem, we train twelve networks, each targeting surface density, radial velocity, and azimuthal velocity, using training datasets of sizes 64, 128, 256, and 448, respectively. The networks are then tested on a dataset comprising 256 unseen FARGO3D simulations (see § 3.2).

As the number of FARGO3D simulations in the training data increases, the test errors initially decrease before eventually plateauing due to factors such as limited network capacity (Figure 4). To provide a clearer visualization of the comparison, we plot the two-dimensional difference between the FARGO3D simulations and the network predictions in Figure 5. The representative case and difference calculation methodology employed in this figure correspond to those used in Figure 2.

C. Tests

C.1. 1D surface density profile

Gaps are one of the most important disk structures whose profiles are closely related to and can be used to constrain the properties of gap-opening planets (Kanagawa et al., 2016). We carry out quantitative comparisons between PPDONet

Table 2. Machine learning hyperparameters

| HYPER-PARAMETER	DEFAULT VALUE
TRAINING DATA SIZE	448
VALIDATION DATA SIZE	64
TESTING DATA SIZE	256
BATCH SIZE	32
NUM. STEPS FOR EACH BATCH	3
INITIALIZATION	GLOROT NORMAL[A]
LEARNING RATE	0.0005
LEARNING RATE DECAY RATE	0.9
LEARNING RATE TRANSITION STEPS	2000
OPTIMIZER	ADAM[B]
TOTAL NUMBER OF ITERATIONS	10^5
BRANCH NETWORK LAYER SIZE	100, 100, 100, 100, 50
TRUNK NETWORK LAYER SIZE	256, 256, 256, 256, 256, 50
Z NETWORK LAYER SIZE	100, 1

[A] GLOROT & BENGIO (2010)
[B] KINGMA & BA (2014)
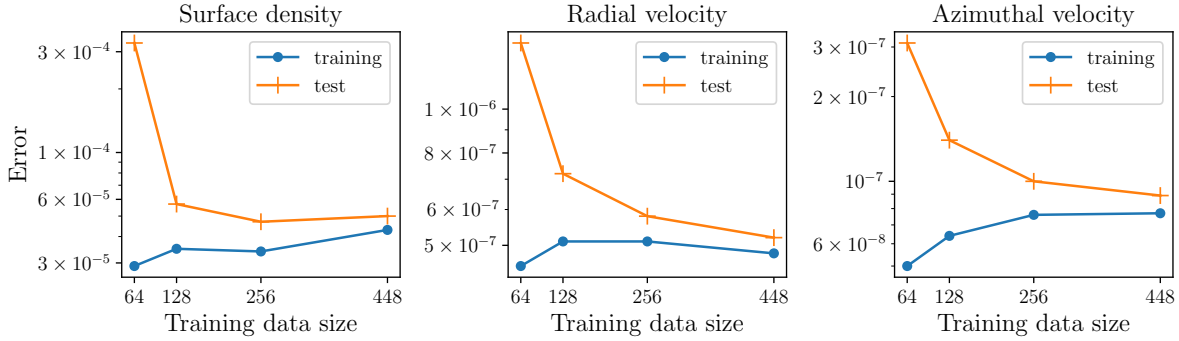
Figure 4. Comparison of test errors for neural networks trained using training dataset with different sizes. The three columns represent networks trained for surface density (left), radial velocity (middle), and azimuthal velocity (right), respectively. We calculate the losses using the method described in §3.3.

predictions and the ground truth by analyzing the azimuthally averaged 1D surface density radial profile of gaps. To measure 1D profiles, we mask regions contaminated by the planet and azimuthally average the surface density. One representative example ($\alpha = 5.2 \times 10^{-4}$, $h_0 = 0.053$, and $q = 1.6 \times 10^{-3}$) is shown in Figure 6. The gap profiles from FARGO3D and PPDONet prediction overlap with no noticeable difference, and the ratio between the two deviates from unity by $\sim 1\%$, indicating an excellent agreement.

C.2. Empirical dependence of the gap width on disk and planet properties

Gap width (Δ_{gap}) relates to disk and planet properties and is an observable frequently used to constrain planet masses in observations (Kanagawa et al., 2016; Zhang et al., 2018). We measure the gap width in our sample and examine the empirical relationship in Kanagawa et al. (2016). We adopt a similar method in measuring Δ_{gap} as Kanagawa et al. (2016), i.e., the radial separation between the inner and outer gap edges, with the edges identified at $0.5\,\Sigma_0$. We fit the data by a power law:

$$
\begin{aligned}
\Delta_{\mathrm{gap}}/R_{\mathrm{p}} \;=\; & 0.47 \pm 0.04 \left(\frac{q}{0.001}\right)^{0.442 \pm 0.003} \\
\times \;& \left(\frac{\alpha}{0.01}\right)^{-0.201 \pm 0.002} \left(\frac{h_0}{0.05}\right)^{-0.94 \pm 0.02}.
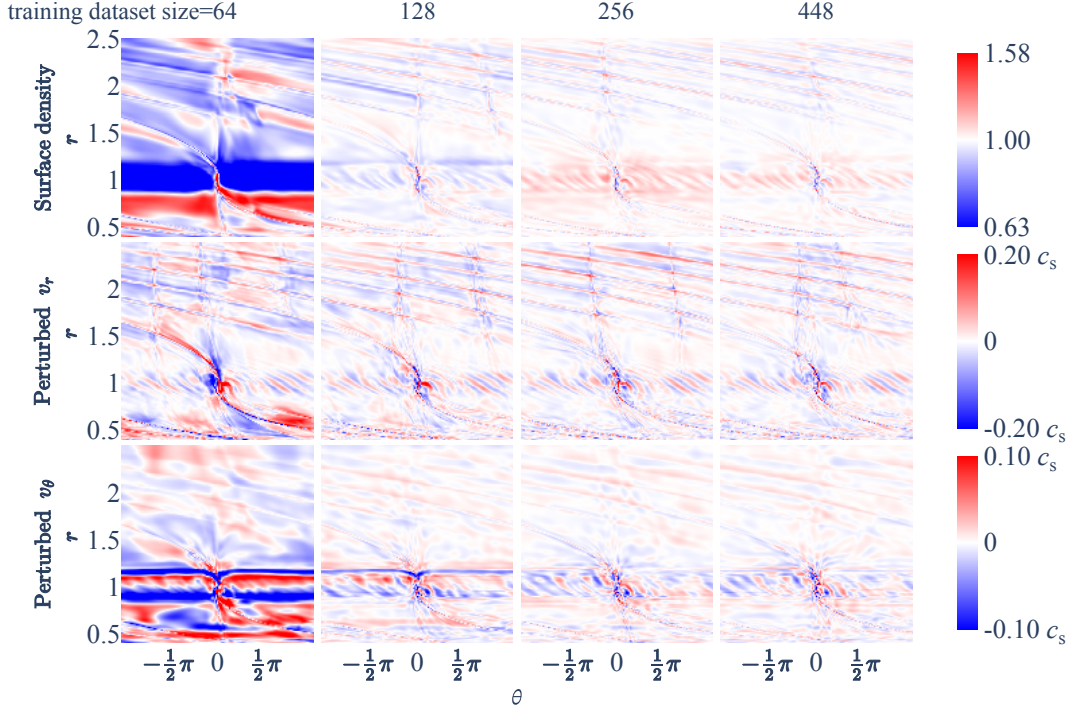\end{aligned}
\tag{8}
$$

8

Figure 5. Comparison of prediction errors for neural networks trained with different dataset sizes. The four columns represent networks trained on dataset sizes of 64 (leftmost column), 128, 256, and 488 (rightmost column) samples, respectively. The representative case and difference calculation methodology correspond to those utilized in Figure 2.

In comparison, Equation (4) in Kanagawa et al. (2016) gives

$$\Delta_{\mathrm{gap}}/R_{\mathrm{p}} = 0.39 \left(\frac{q}{0.001}\right)^{0.5} \left(\frac{\alpha}{0.01}\right)^{-0.25} \left(\frac{h_0}{0.05}\right)^{-0.75}. \tag{9}$$

The two are shown in Figure 7, a reproduction of Figure 3 in Kanagawa et al. (2016), with $19\times$ more data points (500 in our case *vs* 26 in Kanagawa et al. (2016)). Again, the more than one order of magnitude bigger sample size enables us to obtain uncertainties in the fitted parameters.

C.3. Empirical dependence of the azimuthal velocity perturbation on disk and planet properties

We also examine the azimuthally averaged normalized perturbed rotational velocity $v_\theta(r, \theta)$: $\delta v_\theta(r) \equiv \langle (v_\theta(r, \theta) - v_{\theta,0}(r, \theta)) / v_{\theta,0}(r, \theta) \rangle$, where $v_{\theta,0}(r, \theta)$ is the initial value. We then define the difference between the maximum and minimum $\delta v_\theta(r)$ within a radius range of $r \in (0.5r_{\mathrm{p}}, 1.5r_{\mathrm{p}})$ as $\Delta \delta v_\theta \equiv \max[\delta v_\theta(r)] - \min[\delta v_\theta(r)]$ [2] Yun et al. (see 2019, Figure 1d). The measurements are conducted on two datasets each containing 500 disks in different parameter spaces. First, we generate disks from PPDONet with the same range of α as Yun et al. (2019): $3 \times 10^{-4} < \alpha < 3 \times 10^{-3}$. We measure the perturbed velocity amplitudes in our sample (green dots) following Yun et al. (2019) and obtain the best fit (black line):

$$\Delta \delta v_\theta = h_0 \frac{(0.010 \pm 0.001) K^{1.09 \pm 0.03}}{1 + (0.048 \pm 0.003) K^{0.82 \pm 0.02}}, \tag{10}$$

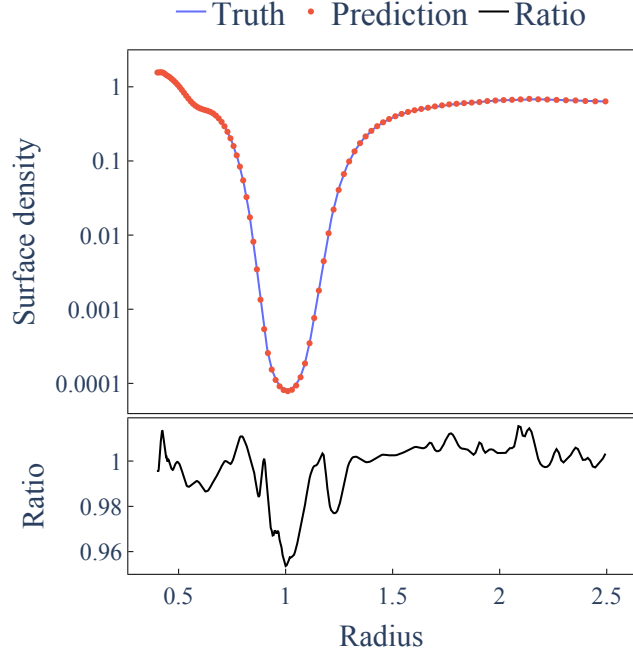[2]The $\Delta \delta v_\theta$ in our work is defined as δ_V in Yun et al. (2019).

Figure 6. Gap profile comparison for one example case with $\alpha = 5.2 \times 10^{-4}$, $h_0 = 0.053$, and $q = 1.6 \times 10^{-3}$. Blue: Ground truth from FARGO3D. Red: PPDONet prediction. Bottom panel: The ratio between the two.
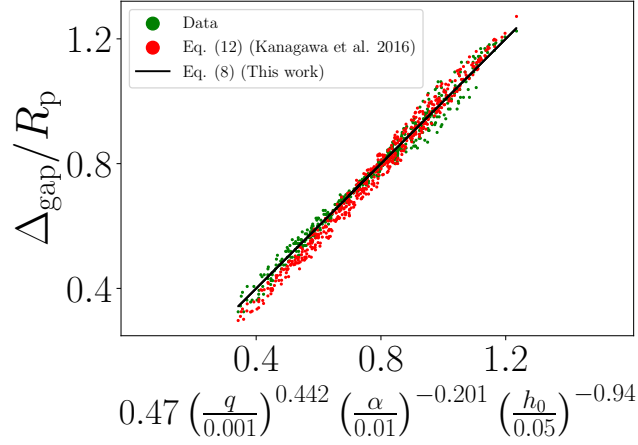

Figure 7. Empirical relationships for the gap width Δ_{gap} against α, h_0, and q. We plot the gap width measured from predicted disks (green), our best-fitting relationship (black), and the relationship from previous work (Kanagawa et al., 2016) (red).

where $K \equiv q^2 h_0^{-5} \alpha^{-1}$. This is similar to the one obtained by (red dots) Yun et al. (2019)

$$\Delta \delta v_\theta = h_0 \frac{0.007 K^{1.38}}{1 + 0.06 K^{1.03}}. \tag{11}$$

The data and the two fits are shown in Figure 8. Next, we generate disks from PPDONet with larger α, $3 \times 10^{-3} < \alpha < 0.1$, and measure $\Delta \delta v_\theta$ (grey dots). We find neither relationship is suitable for accurately characterizing disks with high viscosity. This is as expected, because when $\alpha \gtrsim 0.01$ viscous damping of density waves strongly affects the evolution of waves and their profiles (Miranda & Rafikov, 2020).
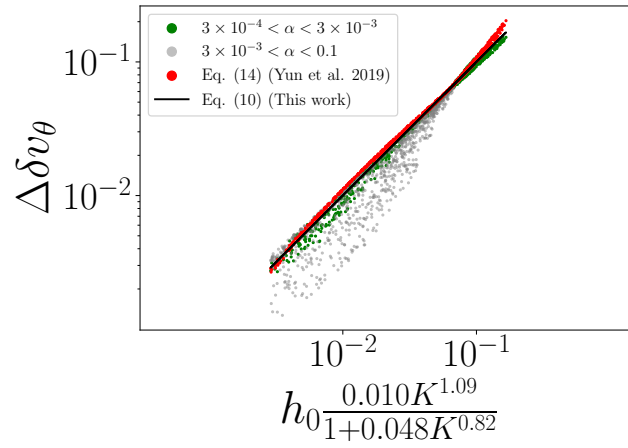
Figure 8. The empirical relationship of amplitudes of perturbed rotational velocity $\Delta\delta v_\theta$ against the dimensionless parameter $K \equiv q^2 h_0^{-5}\alpha^{-1}$. The green and grey dots are the measurements from PPDONet-predicted disks with $3 \times 10^{-4} < \alpha < 3 \times 10^{-3}$ and $3 \times 10^{-3} < \alpha < 0.1$, respectively. The black line is the relationship of this work, fitted from the green dots, and the red sample points are evaluated with the relationship from Yun et al. (2019).