Using Variational Autoencoding to Infer the Masses of Exoplanets Embedded in the Disks of Gas and Dust Orbiting Young Stars

Sayed Shafaat Mahmud
Department of Physics and Astronomy
Colgate University
13 Oak Drive, Hamilton, NY 13346
smahmud@colgate.edu

Sayantan Auddy
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Dr, Pasadena, CA 91109
sayantan.auddy@jpl.nasa.gov

Ramit Dey
Department of Physics and Astronomy
University of Western Ontario
1151 Richmond St, London, ON N6A 3K7, Canada
rdey@uwo.ca

Neal Turner
Jet Propulsion Laboratory
California Institute of Technology
4800 Oak Grove Dr, Pasadena, CA 91109
neal.turner@jpl.nasa.gov

Jeffrey Bary
Department of Physics and Astronomy
Colgate University
13 Oak Drive, Hamilton, NY 13346
jbary@colgate.edu

Abstract

Unseen young planets can be characterized by analyzing emission from dust grains in protoplanetary disks. The masses of embedded planets are inferred through numerical simulations, empirical relations, or deep learning models. In this study, we employ Variational Autoencoders (VAEs) to infer planetary parameters from simulated protoplanetary disk images while quantifying the uncertainties in the predictions. Our approach provides a robust framework for parameter estimation and uncertainty quantification, which will enhance the reliability of characterizing the nature of embedded protoplanets from disk observations.

1 Introduction

The concentric rings and gaps observed in images of protoplanetary disks (PPDs) through dust emission [1, 2, 12] are frequently interpreted as signs of an embedded planet. However, due to the limitations of current planet search techniques [10, 16], directly detecting planets within these disks remains challenging. Characterizing these large-scale, planet-induced morphological changes in protoplanetary disks (PPDs) offers a unique opportunity to investigate young, unseen planets during their formation. This is typically done by comparing these features with theoretical models of planetary gaps generated through customized simulations [20, 6, 19, 8, 11, 7, 14, 13] or empirical relations [15, 17]. However, with the growing number of observed disks and the complexity of disk-planet simulations, running hundreds of customized simulations for each system is inefficient and impractical as a method for inferring the presence, location, and masses of embedded planets. Various deep-learning models [3, 4] have enabled the detection and characterization of exoplanets

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

from observed data with great accuracy. However, a key challenge with traditional deep learning models is the difficulty in quantifying the uncertainties associated with their predictions.

Given an observed or simulated set of PPD images, a probabilistic deep learning model (such as VAEs) can learn to model the underlying distribution of parameters that shape the morphology of these PPDs. In this paper, we implement a VAE-based model and utilize the learned parameter distribution (projected in the lower dimensional latent space of the VAE) to infer planet mass from PPD images. Since the latent space is given in terms of a probabilistic distribution, we can quantify the uncertainties associated with the predictions.

2 Methodology

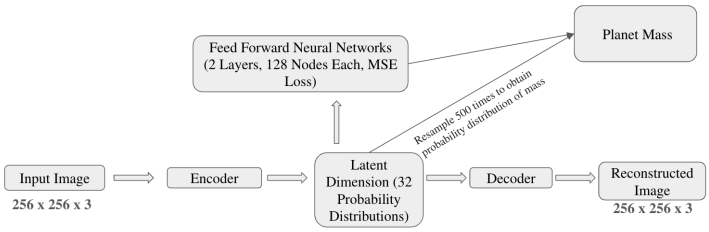

Figure 1: Schematic of network architecture. The model takes PPD images as input and predicts the planet's mass. It additionally outputs the reconstructed image.

A VAE model consists of an encoder and a decoder. The encoder maps the high-dimensional input image x into a latent space z. The goal of the decoder is to reconstruct the original input x from z. As the model is trained/optimized to perform this image reconstruction via a lower-dimensional latent space, the key features (or distribution of parameters) present in the input set get embedded in the latent space.

For our model, the VAE approximates the posterior distribution $p(\alpha|x)$, where α represents planet mass, and x is the disk image. To train the model we consider a loss function that consists of two parts: the evidence lower bound (ELBO) and the mean squared error (MSE) between the predicted planet mass α_{pred} and the true planet mass α_{true}:

$$\mathcal{L}_{\text{total}} = \mathbb{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - \text{KL}[q_\phi(z|x)||p(z)] + \text{MSE}(\alpha_{\text{pred}}, \alpha_{\text{true}}) \tag{1}$$

Here, $q_\phi(z|x)$ is the encoder's approximation of the posterior distribution, and $p(z)$ is a standard normal prior on the latent variables. The reparameterization trick allows the model to learn via backpropagation by expressing z as a function of a mean μ, standard deviation σ, and random noise ϵ, where $z = \mu + \sigma \cdot \epsilon$. The training process iteratively updates the model weights by minimizing the total loss function over batches of disk images, optimizing for both image reconstruction and planet mass prediction.

2.1 VAE-based image reconstruction and parameter inference

The VAE-based model that we implemented is designed to both reconstruct the image and infer the masses of the embedded planets. The encoder consists of three convolutional layers, progressively reducing the spatial dimensions of the image while extracting the necessary features. After passing through the encoder, the output is flattened and passed through fully connected layers to generate the latent space variables (mean μ and log variance $\log \sigma^2$).

To reconstruct the input image, the latent vector z is passed to the decoder, which consists of a fully connected layer followed by transposed convolutional layers. The output is a reconstructed image, matching the original input dimensions, with a sigmoid activation to constrain pixel values between 0 and 1. In parallel, the latent vector z is passed through a separate feedforward network to predict the mass of the planet. This network comprises two fully connected layers and outputs a scalar value representing the predicted planet mass. The VAE itself operates in an unsupervised fashion, learning

a latent probabilistic distribution of disk morphologies. Notably, no prior information about the true planet masses is passed to the VAE, maintaining its unsupervised nature. This approach contrasts with fully supervised models like Convolutional Neural Networks (CNNs), where the true labels directly guide the parameter extraction process.

2.2 Dataset

The training dataset is generated using the hydrodynamic simulation code FARGO3D [5]. We model a simple 2D disk, around a solar mass star, set at a temperature of 10^4 K, with embedded planets of masses $M_\oplus$ located at radius R_p. FARGO3D simulates the morphological changes in dust within protoplanetary disks caused by the presence of embedded single or multiple gap-opening planets of different masses. Each hydrodynamic output is then post-processed with radiative transfer calculations using RADMC3D [9] to create synthetic images that closely resemble the observed images. We train the model on a dataset consisting of 3000 images randomly sampled from 45000 synthetic images.

3 Results

The model is trained for 150 epochs using T4 GPUs on Google Colab. We use Adam optimizer and a learning rate of $1e-3$. We adopt a batch size of 32 images having a resolution of 256×256. Once the VAE-based model is trained, it can reconstruct new images for a given input and predicts the corresponding planet mass based on the disk planet morphology. In this section, we quantify the reconstruction and predict planet mass along with uncertainties.

3.1 Planet Mass Predictions

The trained VAE predicts the mass of the embedded planets from PPD images. Figure 3 compares the predicted masses with the true masses. The scatter plot shows the relationship between the true planet masses and the mean predicted masses across the test set. The error bars represent the standard deviation of the predicted masses, calculated by sampling 500 predictions for each test image to quantify the model's uncertainties.

Figure 3 captures the correlation between the predicted and the true planet mass. As is evident, most predicted values fall close to the dashed line, indicating good agreement between the true and predicted masses. This is further demonstrated by an R^2 value is 0.86. [1]

The error bars indicate the uncertainty in the mass predictions, which increases marginally for larger masses. This suggests that the model's predictions are more precise for smaller masses and tend to exhibit greater variance as the mass increases. However, the overall trend shows that the VAE provides reliable predictions, with uncertainty appropriately accounted for in the error estimates.

The histogram in Figure 3 captures the distribution of the predicted planet mass of a sample image selected randomly from the test dataset. The uncertainty associated with the prediction can be estimated from the variance of the predicted masses. For this test case, we get a $\sigma = 55.92 M_\oplus$ and a mean mass of $\mu = 910.17 M_\oplus$.

3.2 Image Reconstruction Performance

Figure 2 shows two examples of reconstructed protoplanetary disk images from the test dataset, along with the respective ground truth images. The reconstructed images closely resemble the true input images, indicating that the VAE was able to capture the underlying structure of the protoplanetary disks. Additionally, to quantify the reconstruction performance the Structural Similarity Index (SSIM) values for each pair of true and reconstructed images are given. The first pair of images achieve an SSIM score of 0.991, while the second pair has an SSIM of 0.992 out of a maximum value of 1, indicating that the reconstructions are highly accurate.

[1] **The outliers in Figure 3 correspond to certain systems with extremely small dust grains where we don't see an explicit ring formation. In future studies, we will incorporate such systems in the training dataset more systematically, so that the network can infer these cases with more confidence.**

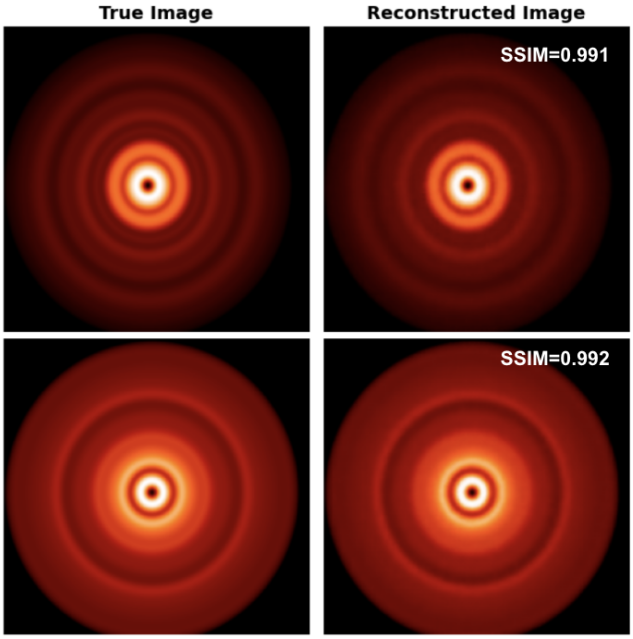

Figure 2: True and reconstructed protoplanetary disk images from the test set. The images on the left show the true disk images, while the images on the right display their corresponding reconstructions. SSIM values quantify the similarity between true and reconstructed images, with the first image pair having an SSIM of 0.991 and the second image pair an SSIM of 0.992.

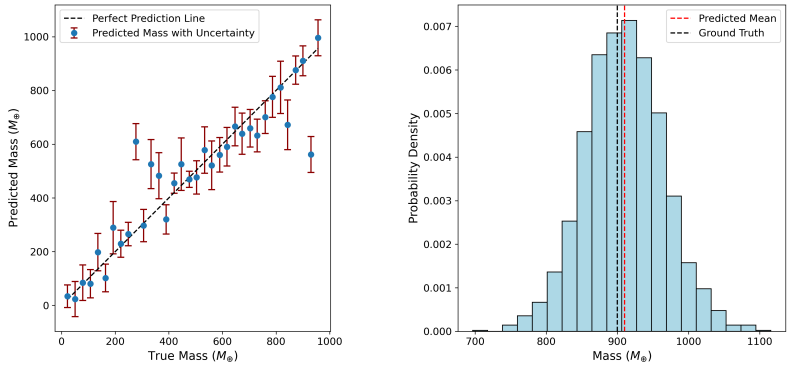

Figure 3: Left: Scatter plot comparing the true planet masses with the mean predicted masses, with error bars representing the standard deviation of the predictions (500 samples per test image). The dashed line indicates the perfect prediction line. The R^2 score for this relationship is 0.86. Right: Distribution of the predicted planet mass of a sample image selected randomly from the test dataset. The mean values along with the standard deviations of the predicted planet masses are $\mu = 910.17 M_\oplus$ and $\sigma = 55.92 M_\oplus$. The true value of the planet mass is $M_{True} = 900 M_\oplus$

4 Limitations and future scope

One of the main limitations of this model is that it is trained on a simulated image dataset. Even though this dataset mimics real-world observations, it doesn't take into consideration any instrumental or environmental noise. As a step forward we would update the training dataset by using an ALMA simulator (e.g. CASA [18]) to refine the synthetic simulations even further.

For generating the dataset we use a pre-determined sampling grid for the parameter set. Out-of-sample real observed data (or for ones belonging to the tail of the parameter distribution) will inherit a very high prediction variance/uncertainty. To address this we will refine the parameter space, used for generating the training dataset, with feedback from real observations.

A trained network's effectiveness is largely dependent on the quality of its training dataset, which should ideally cover all possible scenarios. However, designing simulations that achieve this is not only difficult but also costly in terms of data generation. In this paper, we rely solely on simulated images and focus on a limited parameter space due to constraints in computational resources.

5 Conclusion

We implemented a VAE-based model to predict the masses of embedded exoplanets from images of protoplanetary disks. By leveraging the probabilistic nature of the VAE's latent space, we were able to quantify the uncertainties associated with these predictions, providing a robust framework for parameter inference. To our knowledge, this is the first application of a probabilistic generative model (VAE architecture) to infer planet masses from protoplanetary disk images.

The use of VAE enables us to infer planet masses probabilistically, offering a unique advantage over conventional CNNs, which produce deterministic predictions. By incorporating a probabilistic latent space, we achieve not only accurate predictions but also an assessment of the uncertainties associated with these predictions. This probabilistic approach is especially valuable in astrophysical contexts, where inherent uncertainties in observations are significant. Interestingly, our experiments show that decoupling the feedforward neural network from the VAE during training improves accuracy, possibly by enabling the VAE to focus solely on reconstructing latent distributions, while the feedforward network independently optimizes the mass prediction task.

Although the model performs well on simulation data, its application to real observed data requires further work. To bridge the gap between simulated and real-world datasets, we plan to incorporate observational noise from ALMA data into our training pipeline. This enhancement aims to ensure the model can generalize effectively to observations and provide reliable predictions.

Our approach represents a significant step toward leveraging probabilistic generative models for astrophysical parameter inference. Future work extends this framework to multi-planetary systems or incorporate additional physical parameters to capture more complex disk morphologies and planetary interactions. This ongoing development has the potential to advance our understanding of planet formation and characterization through innovative machine learning techniques.

References

[1] ALMA Partnership et al. "The 2014 ALMA Long Baseline Campaign: First Results from High Angular Resolution Observations toward the HL Tau Region". In: 808.1, L3 (July 2015), p. L3. DOI: 10.1088/2041-8205/808/1/L3. arXiv: 1503.02649 [astro-ph.SR].

[2] Sean M. Andrews et al. "Ringed Substructure and a Gap at 1 au in the Nearest Protoplanetary Disk". In: 820.2, L40 (Apr. 2016), p. L40. DOI: 10.3847/2041-8205/820/2/L40. arXiv: 1603.09352 [astro-ph.EP].

[3] Sayantan Auddy and Min-Kai Lin. "A Machine Learning Model to Infer Planet Masses from Gaps Observed in Protoplanetary Disks". In: 900.1, 62 (Sept. 2020), p. 62. DOI: 10.3847/1538-4357/aba95d. arXiv: 2007.13779 [astro-ph.EP].

[4] Sayantan Auddy et al. "DPNNet-2.0. I. Finding Hidden Planets from Simulated Images of Protoplanetary Disk Gaps". In: *The Astrophysical Journal* 920.1 (Oct. 2021), p. 3. DOI: 10.3847/1538-4357/ac1518. URL: https://doi.org/10.3847/1538-4357/ac1518.

[5] Pablo Benítez-Llambay and Frédéric S. Masset. "FARGO3D: A New GPU-oriented MHD Code". In: 223.1, 11 (Mar. 2016), p. 11. DOI: 10.3847/0067-0049/223/1/11. arXiv: 1602.02359 [astro-ph.IM].

[6] A. Crida, A. Morbidelli, and F. Masset. "On the width and shape of gaps in protoplanetary disks". In: *Icarus* 181 (Apr. 2006), pp. 587–604. DOI: 10.1016/j.icarus.2005.10.007. eprint: arXiv:astro-ph/0511082.

[7] Paul C Duffell. "A SIMPLE ANALYTICAL MODEL FOR GAPS IN PROTOPLANETARY DISKS". In: *The Astrophysical Journal Letters* 807 (2015), p. L11. DOI: 10.1088/2041-8205/807/1/L11.

[8] Paul C Duffell and Andrew I Macfadyen. "GAP OPENING BY EXTREMELY LOW-MASS PLANETS IN A VISCOUS DISK". In: *The Astrophysical Journal* 769.6pp (2013), p. 41. DOI: 10.1088/0004-637X/769/1/41.

[9] C. P. Dullemond et al. "RADMC-3D: A multi-purpose radiative transfer tool". In: ascl:1202.015 (Feb. 2012).

[10] D. A. Fischer et al. "Exoplanet Detection Techniques". In: *Protostars and Planets VI*. Ed. by Henrik Beuther et al. Jan. 2014, p. 715. DOI: 10.2458/azu_uapress_9780816531240-ch031. arXiv: 1505.06869 [astro-ph.EP].

[11] Jeffrey Fung, Ji-Ming Shi, and Eugene Chiang. "HOW EMPTY ARE DISK GAPS OPENED BY GIANT PLANETS?" In: *The Astrophysical Journal* 782.11pp (2014), p. 88. DOI: 10.1088/0004-637X/782/2/88.

[12] Jane Huang et al. "CO and Dust Properties in the TW Hya Disk from High-resolution ALMA Observations". In: 852.2, 122 (Jan. 2018), p. 122. DOI: 10.3847/1538-4357/aaa1e7. arXiv: 1801.03948 [astro-ph.SR].

[13] John D. Ilee et al. "Observing protoplanetary discs with the Square Kilometre Array - I. Characterizing pebble substructure caused by forming planets". In: 498.4 (Nov. 2020), pp. 5116–5127. DOI: 10.1093/mnras/staa2699. arXiv: 2009.00562 [astro-ph.SR].

[14] Kazuhiro D Kanagawa et al. "MASS ESTIMATES OF A GIANT PLANET IN A PROTOPLANETARY DISK FROM THE GAP STRUCTURES". In: *The Astrophysical Journal Letters* 806 (2015), p. L15. DOI: 10.1088/2041-8205/806/1/L15.

[15] Kazuhiro D. Kanagawa et al. "Mass constraint for a planet in a protoplanetary disk from the gap width". In: *Publications of the Astronomical Society of Japan* 68.3 (2016), pp. 1–7. ISSN: 2053051X. DOI: 10.1093/pasj/psw037.

[16] Chien-Hsiu Lee. "Exoplanets: Past, Present, and Future". In: *Galaxies* 6.2 (Apr. 2018), p. 51. DOI: 10.3390/galaxies6020051. arXiv: 1804.08907 [astro-ph.EP].

[17] Giuseppe Lodato et al. *The newborn planet population emerging from ring-like structures in discs*. Tech. rep. 2019. URL: www.exoplanet.eu,.

[18] J. P. McMullin et al. "CASA Architecture and Applications". In: *Astronomical Data Analysis Software and Systems XVI*. Ed. by R. A. Shaw, F. Hill, and D. J. Bell. Vol. 376. Astronomical Society of the Pacific Conference Series. Oct. 2007, p. 127.

[19] S.-J. Paardekooper and J. C. B. Papaloizou. "On the width and shape of the corotation region for low-mass planets". In: 394 (Apr. 2009), pp. 2297–2309. DOI: 10.1111/j.1365-2966.2009.14512.x. arXiv: 0901.2263.

[20] Shangjia Zhang et al. "The Disk Substructures at High Angular Resolution Project (DSHARP). VII. The Planet-Disk Interactions Interpretation". In: *The Astrophysical Journal Letters* 869 (2018), p. L47. DOI: 10.3847/2041-8213/aaf744. URL: https://doi.org/10.3847/2041-8213/aaf744.