Autoencoding Labeled Interpolator, Inferring Parameters From Image, And Image From Parameters

Ali SaraerToosi[1,2] **Avery Broderick**[1,2]
[1]Perimeter Institute [2]University of Waterloo
31 Caroline St N 200 University Ave W
{asaraertoosi, abroderick}@perimeterinstitute.ca
{asaraert, abroderick}@uwaterloo.ca

Abstract

The Event Horizon Telescope (EHT) provides an avenue to study black hole accretion flows on event-horizon scales. Traditionally, fitting a semi-analytical model to EHT observations requires the construction of synthetic images, which is computationally expensive. This study presents an image generating tool in the form of a generative machine learning model, which extends the capabilities of a variational autoencoder. This tool can rapidly and continuously interpolate between a training set of images and can retrieve the defining parameters of those images. Trained on a curated set of synthetic black hole images, our tool showcases success in both interpolating and generating images, and retrieving the physical parameters. By reducing the computational cost of generating an image, this tool facilitates parameter estimation and model validation for observations of black hole systems.

1 Introduction

The Event Horizon Telescope (EHT) is a very long baseline interferometer [Collaboration, 2019a,b] created to observe black holes on event-horizon scales, making it possible to study the astrophysical and gravitational processes in the vicinity of black holes [Broderick et al., 2009, Broderick et al., 2011, Broderick et al., 2014, 2016]. So far, Sagittarius A* [Collaboration, 2022] and Messier 87* [EHT, 2019] have been imaged, revealing a lot of information about the structure and dynamics of the accretion disks around these black holes. To gain more information, however, utilizing forward models is essential as the Physical models are the most interpretable.

EHT data is fit within Bayesian parameter estimation frameworks[Broderick and EHT, 2020]. This is done by sampling a model likelihood such that the data provides direct posteriors on the model parameters. Here, to demonstrate our methodology, we concentrate on a particular physical model called RIAF [Broderick & Loeb, 2006, Pu and Broderick, 2018]; however, our method can be used for any well-behaved physical model. RIAF is a simple semi-analytical model of the accretion flow around a black hole that approximates the dynamics of the accretion disk with a thick static disk. However, generating images with normal methods requires simulations and is computationally expensive [Broderick and EHT, 2020].

In this paper, we create a rapid and reliable non-linear interpolation tool based on the principles of generative models in machine learning. This tool, named Autoencoding Labeled Interpolator Network (ALINet), is a new type of variational autoencoder (VAE) [Kingma and Welling, 2014] and due the nature of its structure, it is physically interpretable. ALINet, therefore, is a tool that can retrieve physical parameters from an image. We additionally trained an inverse network that does the job of ALINet in reverse; it finds the image when physical parameters are given.

Machine Learning and the Physical Sciences Workshop, NeurIPS 2023.

In section 2, we discuss in more detail how a traditional VAE works and then we describe our own method. As an example, we train ALINet on RIAF images, spanning 5 physical parameters: black hole spin (a), inclination angle(i), disk thickness (H/R), non-thermal electron density (n_{nth}), and sub-Keplerian fraction (κ). In section 3, we evaluate how well the network can recover these physical parameters and we show that the inverse network can faithfully recover the physical parameters when it is paired with ALINet.

2 Methods

2.1 Variational Autoencoder

A traditional VAE is comprised of an encoder and a decoder. The encoder takes images as input and encodes them down to a few probability distributions. Samples from these distributions are called latent variables. Furthermore, the space in which the latent variables live is called the latent space (for a comprehensive introduction to VAEs, see for example, [Kingma and Welling, 2014]) The goal of the VAE is to learn how to encode a high-dimensional image to a low-dimensional representation and learn to reconstruct the image from the latent representation of an image while keeping the latent space distributions close to the prior distributions. The VAE accomplishes these goals by minimizing a loss function which is called the Evidence Lower Bound (ELBO) and is shown in Equation 1 (Neal and Hinton [1998]).

$$\text{ELBO}_{VAE}(\phi,\,\theta) = -\underbrace{E_{z\sim q_\phi(z\,|\,D)}[log\langle p_\theta(D\,|\,z)\rangle]}_{\text{reconstruction error}} + \underbrace{D_{KL}[q_\phi(z\,|\,D)\;||\;p(z)]}_{\text{regularization}}, \tag{1}$$

where ϕ and θ represent the parameters (weights and biases) of the encoder and the decoder, respectively. D_{KL} is the Kullback–Leibler divergence, which is the measure of how close a probability distribution is to another [Shlens, 2014]. The term containing the KL-divergence is responsible to keep the latent distributions close to the prior. D is representative of the available data or images, z denotes the latent space variables or latent parameters, $p(z)$ is the prior, $p_\theta(D|z)$ is the likelihood of D occurring assuming the parameters z, $q_\phi(z|D)$ is the “surrogate distribution" which the VAE should learn and make as close to the posterior $p(z|D)$.

The problem with VAEs is that the distributions learned by the autoencoder in the latent space are not readily interpretable. [Shavlik, 1992, Mahendran and Vedaldi, 2015, Ghahramani, 2015, Ravid Shwartz-Ziv, 2017]. However, since the difference in the images is caused by the physical parameters, there should exist a mapping from the latent parameters to the physical parameters.

2.2 Autoencoding Labeled Interpolator Network (ALINet)

In this structure, which we name Autoencoding Labeled Interpolator (or ALINet for short), we stitch the labels (physical parameters) to the VAE using a new neural network branch connected to the latent parameters on the decoder side, whose sole job is to find the mapping from the latent parameters to the physical labels. The loss function has therefore three components this time, as shown in Equation 2.

$$\text{ELBO}_{ALI}(\phi,\theta) = \text{ELBO}_{VAE}(\phi,\,\theta) + \alpha \times \underbrace{\sum_i (y_i - \hat{y}_i)^2}_{\text{SSE}} \tag{2}$$

$\text{ELBO}_{VAE}(\phi,\,\theta)$ is the loss function of the traditional VAE described in Equation 1. However, note that in our model, ϕ still represents the learnable parameters in the encoder, but θ represents all the learnable parameters in both branches of ALINet decoder. y_i and $\hat{y}_i$ are the truth and network output for the physical parameters, and α is the hyperparameter that tunes the strength of the extra term in the loss function, i.e. how much emphasis is put on predicting the exact physical parameters. Since the dimensions of the images are significantly bigger than the number of parameters, reconstruction loss tends to be considerably larger than the parameter SSE loss. α is important to make the sum squared error (SSE) loss comparable to the reconstruction loss.

2.3 Inverse Network

The architecture described in subsection 2.2 can retrieve physical parameters given an image. To generate an image given physical parameters, we train an inverse neural network that learns the mapping from physical parameters to the latent distributions. These latent distributions then will be fed to the ALINet decoder to create images. The loss function for this network, therefore, is as follows:

$$\text{Loss} = \alpha \times \sum_i (\hat{\mu} - \mu)^2 + \sum_i (\hat{\text{logvar}} - \text{logvar})^2 \tag{3}$$

Where μ and logvar are the means and natural logarithm of variances of the latent distributions representing the physical parameters. Variables with hats represent the predictions of the network. α is a multiplier to adjust the emphasis of the loss function.

3 Experiments/Results

3.1 ALINet: Parameters From Image

We used 160000 RIAF black hole (BH) images for training, 20000 for validation. We chose these images in such a way that they cover the entire parameter space (look at Table 1 for parameter ranges). We choose the latent space to be 5-dimensional. We also use max-min normalization to bring the data to the range $[0, 1]$. The range of each parameter can be found in table 1. In equation 2 we choose $\alpha = 40000$. We train the model for 45 epochs and chose a batch size of 64. We use learning rates of 10^{-3}, 10^{-4}, and 10^{-5} each for 15 epochs. In all of our setups, we used the Adam optimizer for training. [Kingma and Ba, 2017] Training on the Beluga supercomputer [Baldwin, 2012] takes 1 day to complete. We use 40 CPU cores and 1 GPU for training.

After training, the model is tested with 20,000 testing images. In second row of Figure 1, we plot the output when ALINet is used to reconstruct 5 randomly chosen images from the test data set. For all the test images, the difference between the predicted value of black hole image parameters and the true values given to the simulation can be found in Figure 2. These errors are important and should be added as systematic errors to any further analyses.

3.2 Inverse Network: Image From Parameters

We use the same dataset as subsection 3.1. We first train the network for 2 epochs with learning rate of 10^{-4} and then for another 2 epochs with learning rate of 10^{-5}. In Equation 3 we chose $\alpha = 10$. The error of the physical parameters extracted when inverse network (InvNet) + ALINet decoder is used is shown in Figure 2. These figures and Table 1 show that the $1 - \sigma$ errors for all the parameters for both ALINet and inverse model are less than or equal to 2.5%. Furthermore, the third row of Figure 1, we generate images using InvNet + ALINet. These images are generated by using the parameter values of the first row of this figure as input to InvNet + ALINet. The second branch of ALINet also outputs the physical parameter corresponding to each image, which match the input within a small error margin.

Table 1: Parameter ranges for the RIAF black hole model

a	$\cos i$	H/R	n_{nth}	κ
[-0.98, 0.98]	[-0.99, 0.99]	[0.05 , 2]	[0 , 0.05]	[0.01 , 1]

4 Conclusions

In this work we improve on a traditional variational autoencoder by adding a second branch to the decoder which finds the underlying physical parameters that describe an image. Moreover, we train an inverse network which finds an image corresponding to input parameters. We show that these models extract the images and parameters with small errors. ALINet and inverse network can be utilized

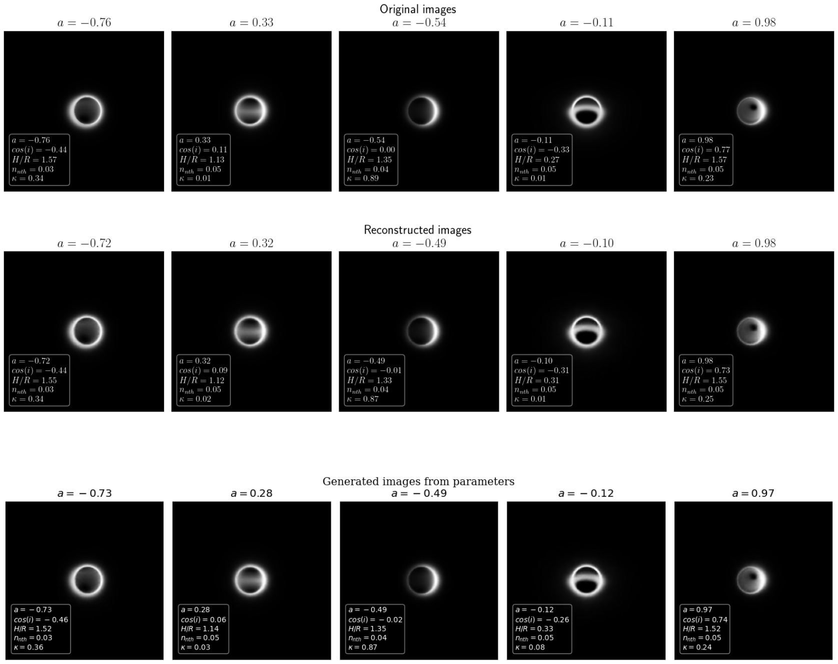

Figure 1: 5 random sample images and their reconstructed outputs from the black hole testing dataset. The number above each image corresponds to the true value of the spin for that image in the first row and the reconstructed spin value from the second branch of the ALINet decoder for the second row. In the third row, the images generated by inputting the truth parameter values in ALINet+InvNet, and the predicted parameters from the second branch of ALINet decoder are shown. All the relevant physical parameters for each image are displayed in a box in the image.

as interpolation tools in fitting astrophysical models to EHT black hole data to highly constrain the allowed range of values for the physical parameters in each model.

Even though here we have used a RIAF model as an example, all the methodology is general to any well-behaved function and can be used for a variety of applications, such as other black hole models like GRMHDs [Narayan et al., 2012] or in multi-messenger astronomy, e.g. for detecting features in gravitational wave signals in binary black hole systems [LIGO, 2016].

Acknowledgements

We thank Boris Georgiev and Sebastian Wetzel for helpful comments. This work was supported in part by Perimeter Institute for Theoretical Physics. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade. A.E.B. receives additional financial support from the Natural Sciences and Engineering Research Council of Canada through a Discovery Grant. This research was enabled in part by support provided by Calcul Québec (www.calculquebec.ca) and the Digital Research Alliance of Canada (alliancecan.ca).

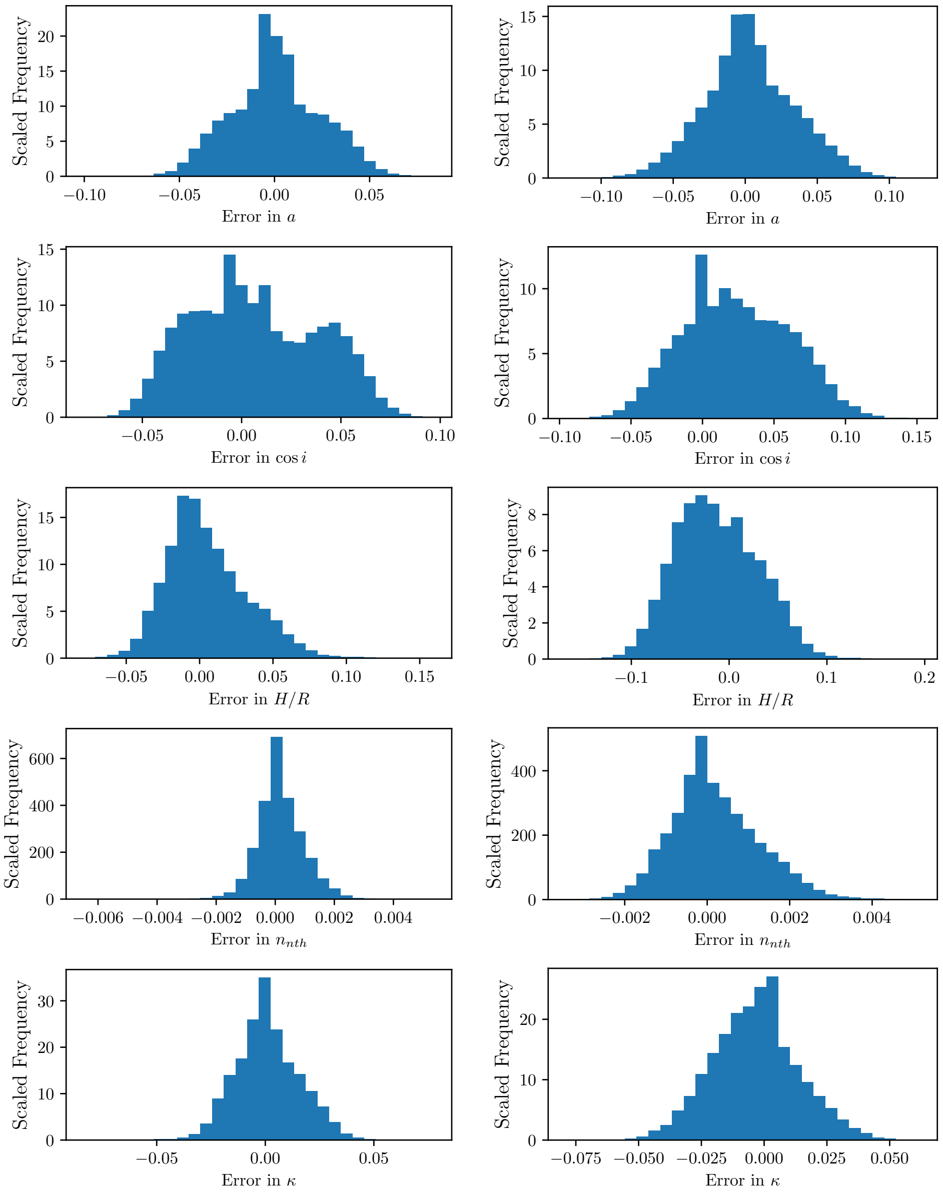

Figure 2: The error of the predicted values by **left:** ALINet model for $20,000$ test images and **right:** inverse network for $20,000$ test parameters. The error is calculated by subtracting the predicted value for each parameter from the truth value of that parameter, i.e. if X is the parameter in question, we do $X_{\text{original}} - X_{\text{reconstructed}}$ to obtain the plots.

References

S. Baldwin. Compute canada: Advancing computational research. *Journal of Physics: Conference Series*, 2012.

A. E. Broderick and T. EHT. Themis: A parameter estimation framework for the event horizon telescope. *The Astrophysical Journal*, 2020.

A. E. Broderick, V. L. Fish, S. S. Doeleman, and A. Loeb. Estimating the parameters of sagittarius a*'s accretion flow via millimeter vlbi. *The Astrophysical Journal*, 697(1):45, apr 2009. doi: 10.1088/0004-637X/697/1/45. URL `https://dx.doi.org/10.1088/0004-637X/697/1/45`.

A. E. Broderick, T. Johannsen, A. Loeb, and D. Psaltis. Testing the no-hair theorem with event horizon telescope observations of sagittarius a*. *The Astrophysical Journal*, 784(1):7, feb 2014. doi: 10.1088/0004-637X/784/1/7. URL `https://dx.doi.org/10.1088/0004-637X/784/1/7`.

A. E. Broderick, V. L. Fish, M. D. Johnson, K. Rosenfeld, C. Wang, S. S. Doeleman, K. Akiyama, T. Johannsen, and A. L. Roy. Modeling seven years of event horizon telescope observations with radiatively inefficient accretion flow models. *The Astrophysical Journal*, 820(2):137, mar 2016. doi: 10.3847/0004-637X/820/2/137. URL `https://dx.doi.org/10.3847/0004-637X/820/2/137`.

Broderick & Loeb. Imaging optically-thin hotspots near the black hole horizon of Sgr A* at radio and near-infrared wavelengths. *Monthly Notices of the Royal Astronomical Society*, 367(3):905–916, 04 2006. ISSN 0035-8711. doi: 10.1111/j.1365-2966.2006.10152.x. URL `https://doi.org/10.1111/j.1365-2966.2006.10152.x`.

Broderick et al. Evidence for Low Black Hole Spin and Physically Motivated Accretion Models from Millimeter-VLBI Observations of Sagittarius A*. *ApJ*, 735:110, Jul 2011. doi: 10.1088/0004-637X/735/2/110.

E. Collaboration. Studying Black Holes on Horizon Scales with VLBI Ground Arrays. *arXiv*, 51:256, Sept. 2019a. doi: 10.48550/arXiv.1909.01411.

E. H. T. Collaboration. First sagittarius a* event horizon telescope results. i. the shadow of the supermassive black hole in the center of the milky way. *The Astrophysical Journal Letters*, 930(2):L12, may 2022. doi: 10.3847/2041-8213/ac6674. URL `https://dx.doi.org/10.3847/2041-8213/ac6674`.

T. E. H. T. Collaboration. First m87 event horizon telescope results. ii. array and instrumentation. *The Astrophysical Journal Letters*, 875(1):L2, apr 2019b. doi: 10.3847/2041-8213/ab0c96. URL `https://dx.doi.org/10.3847/2041-8213/ab0c96`.

EHT. First m87 event horizon telescope results. i. the shadow of the supermassive black hole. *ApJ*, 2019.

Z. Ghahramani. Probabilistic machine learning and artificial intelligence. *Nature*, 2015.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *ArXiv e-prints*, 2017.

D. P. Kingma and M. Welling. Auto-encoding variational bayes. *ICLR*, 2014.

LIGO. Observation of gravitational waves from a binary black hole merger. *Phys. Rev. Lett.*, 2016.

A. Mahendran and A. Vedaldi. Understanding deep image representations by inverting them. *CVPR*, 2015.

R. Narayan, A. Są dowski, R. F. Penna, and A. K. Kulkarni. GRMHD simulations of magnetized advection-dominated accretion on a non-spinning black hole: role of outflows. *Monthly Notices of the Royal Astronomical Society*, 2012.

R. M. Neal and G. E. Hinton. A view of the em algorithm that justifies incremental, sparse, and other variants. *ArXiv e-prints*, 1998.

H.-Y. Pu and A. E. Broderick. Probing the innermost accretion flow geometry of sgr a* with event horizon telescope. *The Astrophysical Journal*, 863(2):148, aug 2018. doi: 10.3847/1538-4357/aad086. URL `https://dx.doi.org/10.3847/1538-4357/aad086`.

N. T. Ravid Shwartz-Ziv. Opening the black box of deep neural networks via information. *ArXiv e-prints*, 2017.

G. T. . J. W. Shavlik. Interpretation of artificial neural networks: Mapping knowledge-based neural networks into rules. *ArXiv e-prints*, 1992.

J. Shlens. Notes on kullback-leibler divergence and likelihood theory. *ArXiv e-prints*, 2014.