A Poisson-process AutoDecoder for Astrophysical, Time-variable, X-ray Sources

Yanke Song
Department of Statistics
Harvard University
ysong@g.harvard.edu

Ashley Villar
Center for Astrophysics | Harvard
& Smithsonian
The NSF AI Institute for Artificial Intelligence and Fundamental Interactions
ashleyvillar@cfa.harvard.edu

Rafael Martínez-Galarza
Center for Astrophysics | Harvard
& Smithsonian
juan.martinez.galarza@cfa.harvard.edu

Abstract

X-ray observing facilities such as the Chandra X-ray Observatory and the eROSITA all sky survey have detected millions of astronomical sources associated with high-energy phenomena. The arrival of photons as a function of time follows a Poisson process and can vary by orders-of-magnitude, presenting obstacles for downstream tasks such as source classification, physical property derivation and anomaly detection. Previous work has either failed to directly capture the Poisson nature of the data or only focuses on Poisson rate function reconstruction. In this work, we present Poisson Process AutoDecoder (PPAD). PPAD is a neural field decoder that maps fixed-length latent features to continuous Poisson rate functions across energy band and time via unsupervised learning. It reconstructs the rate function and yields a representation at the same time. We demonstrate the efficacy of PPAD reconstruction, regression, classification and anomaly detection experiments using the Chandra Source Catalog.

1 Introduction

Massive volumes of X-ray data are being produced at unprecedented rates thanks to ongoing X-ray surveys and missions, such as the Chandra X-ray Observatory [5], the XMM-Newton[23] telescope, and the eROSITA survey [12], which together contain approximately 2 million individual X-ray sources in the sky (and several million individual detections). Automated and data-driven learning of the individual X-ray photons associated with these sources is necessary, as it enables various downstream applications at massive scale. However, X-ray source vary by orders-of-magnitude in terms of X-ray photons detected, as well as in the distribution of photon energies and relevant timescales. Many sources are well-within the Poisson limit–with telescopes receiving just a few photons per exposure per source.

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

In X-ray astronomy, machine learning classification has primarily relied on the use of hand-crafted features using both supervised and unsupervised approaches [11, 17, 7, 24, 2]. Manual feature engineering requires strong domain knowledge and intensive computation, and may not select the most representative features. For time-variable sources, other lines of work use the more primitive light curve data (binned photon arrival times describing source variability) and apply deep neural networks to encode light curve segments into low dimensional embeddings. These embeddings are either learned in an unsupervised manner [14, 15, 19] or trained end-to-end with some supervised objectives [1, 4, 20, 26]. However, such methods still require light curve reconstruction, for which the naive binning strategy ignores the intrinsic stochastic nature of photon arrivals, thereby creating artifacts in reconstructed light curves which are especially severe for low intensity sources. The more robust likelihood-based Gregory-Loredo algorithm [8] incorporates the Poisson stochasticity, yet still requires binning with < 20 bins for efficient model selection, thereby limiting the capacity. Moreover, the GL algorithm does not naturally allow downstream learning.

In this work, we introduce the Poisson Process Auto-Decoder (PPAD), a pipeline that embeds raw event files (i.e., photon arrival times and tagged energies) to latent representations in an unsupervised manner. PPAD makes the following three key contributions: 1) It employs a neural field for light curve reconstruction, offering infinite resolution and bypassing the binning in previous approaches. 2) It uses a Poisson likelihood-based approach that respects the intrinsic stochasticity of X-ray sources. 3) Via an auto-decoder, it learns fixed-length latent representations of variable-length event files, offering great flexibility for downstream tasks. We showcase the remarkable performance of PPAD via the accuracy of reconstructed light curves and the adaptability of learned latent representations on regression, classification and anomaly detection.

2 Method

Modeling Photon Arrivals as Poisson Processes: X-ray sources are detected / stored as "event files" – a collection of photon arrival times, each tagged with the energy of the photon. The arrival times are commonly modelled as an inhomogenerous Poisson process [3] . For an inhomogeneous Poisson process with rate $r(t)$ (effectively the light curve intensity), an established result [18] precisely characterizes the likelihood of a list of events $\{t_i\}_{i=1}^n$ during an observation interval $[0, T)$:

$$\text{likelihood}(t_1, ..., t_n; r) = \left(\prod_{i=1}^{n} r(t_i)\right) \exp\left(-\int_0^T r(t)dt\right). \tag{1}$$

Here the integral can be approximated via discretization. Now given a list of events $\{t_i\}_{i=1}^n$ on $[0, T)$, we would like to find the rate function $r(t)$ that maximizes the likelihood, or minimize the negative log-likelihood, of this event file. However, this is an ill-posed problem, because Dirac rates at arrival times would lead to an unbounded likelihood. We therefore choose to add a smoothness constraint in the form of a total variation (TV) penalty. Specifically, for the set of discretization points $0 = \tau_1 \leq ... \leq \tau_N = T$, the total variation of the rate function r on these points is defined as:

$$\text{TV}(r; \tau_1, ..., \tau_N) = \frac{1}{N-1} \sum_{i=1}^{N-1} |r(\tau_i) - r(\tau_{i+1})| \tag{2}$$

We compute the TV on both the set of arrival times as well as an additional set of uniformly discretized points on $[0, T)$ to ensure that the penalty adequately samples at both low count and high count regions. The combined reconstruction loss is the sum of the negative log likelihood and the TV penalty, weighted by the hyperparameter λ_{TV}.

Neural light curve reconstruction: Given the loss defined in Section 2, we would like to find the rate function that minimizes the loss $\mathcal{L}(r)$. We adopt the now common practice (e.g. [13]) to parametrize r with a one-dimensional neural field for its weak dependence on assumptions and its resolution-free property. Specifically, a neural network (with weights ϕ) takes time t as an input and output $r_\phi(t)$ such that $r_\phi(t) \approx r(t)$. We then use standard gradient descent algorithms (e.g., Adam [10]) to minimize the loss defined in Section 1 by tuning ϕ. Upon convergence, r_ϕ yields the reconstructed light curve of the given event file.

To enhance the network's ability to learn high frequency signals, we adopt another common practice (e.g. [13]) to apply the following sinusoidal Positional Encoding (PE) to the input t:

$$\gamma(t) = [\bar{t}, \sin(2^0\pi\bar{t}), \cos(2^0\pi\bar{t}), ..., \sin(2^{L-1}\pi\bar{t}), \cos(2^{L-1}\pi\bar{t})]. \tag{3}$$

Here $\bar{t} = t/T$. The encoding $\gamma(t)$ maps t to a $(2L+1)$-dimensional vector $\gamma(t)$ with features in different frequencies, which is then fed into the neural network to produce the output $r_\phi(\gamma(t))$.

Learning latent representation via auto-decoding: To learn the latent representations of a catalog of events, we use an auto-decoder based approach (see e.g. [16, 21]). This is because the common auto-encoder approach is *not* well-suited for our purpose as the event files are variable in length and highly stochastic, preventing autoregressive type encoders to achieving good convergence. Instead, we randomly initialize a latent variable $\boldsymbol{z}$ for each event file, which is fed together with the encoding $\gamma(t)$ through the neural network r_ϕ to produce the reconstructed light curve. For a set of event files $\{t_{ji}\}, 1 \le j \le m, 1 \le i \le n_j$ coming from m sources, we reconstruct m light curves $r^{(j)}(t) \approx r_\phi(t; \boldsymbol{z}^{(j)})$ with the same global weights, ϕ, and different latent variables $\boldsymbol{z}^{(j)}, 1 \le j \le m$. The set of latents are optimized together with the neural network weights during training. We regularize the latent space by adding $\|\boldsymbol{z}^{(j)}\|_2^2$ to the loss; this ensures a compact manifold in the latent space and helps with convergence.

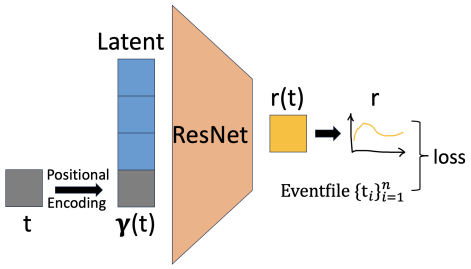

Figure 1: Illustration of PPAD. Latent vectors are concatenated to positionally encoded time t and fed to the shared decoder. The output $r(t)$ for $t \in [0, T)$ forms the light curve, which are used to compute the loss function in 4 together with the event files. Gradients are back-propagated to both the ResNet and the latents.

Putting it together: Poisson Process Auto-Decoder: The full PPAD architecture is displayed in Fig. 1 The loss function of PPAD contains three parts: likelihood, total variation penalty, and latent norm penalty. Furthermore, we bring back the energy marking, and generalize the neural field to output a rate value for each energy bin. Summarizing these components, our final loss function has the following formulation:

$$\begin{aligned}
\mathcal{L}_{\text{total}}(\phi; \{\boldsymbol{z}_j\}_{j=1}^M) &= \sum_{j=1}^{M} \left(\sum_{k=1}^{K} \left(\mathcal{L}_{\text{neg-loglikelihood}}^{(j,k)} + \lambda_{\text{TV}} \mathcal{L}_{\text{TV}}^{(j,k)} \right) + \lambda_{\text{latent}} \mathcal{L}_{\text{latent}}^{(j)} \right), \\
\mathcal{L}_{\text{neg-loglikelihood}}^{(j,k)} &= -\sum_{i=1}^{n_{j,k}} \log r_\phi^{(k)}(\gamma(t_{i,k}); \boldsymbol{z}^{(j)}) + \int_0^T r_\phi^{(k)}(\gamma(t); \boldsymbol{z}^{(j)}) dt, \\
\mathcal{L}_{\text{TV}}^{(j,k)} &= \left[\frac{1}{N-1} \sum_{i=1}^{N-1} |r_\phi^{(k)}(\gamma(\tau_i); \boldsymbol{z}^{(j)}) - r_\phi^{(k)}(\gamma(\tau_{i+1}); \boldsymbol{z}^{(j)})| \right. \\
&\left. + \frac{1}{n-1} \sum_{i=1}^{n-1} |r_\phi^{(k)}(\gamma(t_i); \boldsymbol{z}^{(j)}) - r_\phi^{(k)}(\gamma(t_i); \boldsymbol{z}^{(j)})| \right], \\
\mathcal{L}_{\text{latent}}^{(j,k)} &= \|\boldsymbol{z}^{(j)}\|_2^2,
\end{aligned} \tag{4}$$

where $j = 1, ..., M$ refers to event files, $k = 1, ..., K$ refers to energy bins, $t_{i,k}, i = 1, ..., n_{j,k}$ refers to photon arrivals in the k-th energy bin, $\tau_i, i = 1, ..., N$ refers to evenly discretized points, and γ is the positional encoding defined in Eqn. 3. During training, ϕ and $\{\boldsymbol{z}_j\}_{j=1}^M$ are optimized together. The neural network ϕ is a ResNet [9] which takes a $(d_{\text{latent}} + 2L + 1)$-dimensional input (concatenation of the latent vector and the positional time encoding) and outputs a K dimensional vector representing the rate function at K energy bins. At test/inference time for a new event file, ϕ is

frozen and only a new latent $\boldsymbol{z}$ is optimized:

$$\textbf{Training}: \hat{\phi}, \{\hat{\boldsymbol{z}}^{(j)}\}_{j=1}^{M} := \underset{\phi;\{\boldsymbol{z}_j\}_{j=1}^{M}}{\arg\min}\ \mathcal{L}_{\text{total}}(\phi; \{\boldsymbol{z}^{(j)}\}_{j=1}^{M}). \tag{5}$$

$$\textbf{Inference}: \hat{\boldsymbol{z}} := \underset{\boldsymbol{z}}{\arg\min}\ \mathcal{L}_{\text{total}}(\hat{\phi}; \boldsymbol{z}). \tag{6}$$

3 Experiments

Data and training We collect 109 656 event files from the Chandra Source Catalog [6], truncate all event files into 8-hour segments, bin energies into soft (0.5-1.2 keV), medium (1.2-2.0 keV), and hard (2.0-7.0 keV) categories. Additional implementation details can be found in Appendix A.

Light curve reconstruction Fig. 2 displays light curve reconstruction results on four sources. The PPAD reconstructs a wide range of light curve patterns, including periodic signals, flares, dips and non-varying sources. The reconstruction quality remains high for the integrated ("bolometric") light curve as well as for specific energy bands. Also note the smoothing effect of reconstructed light curves, which filters out stochasticities in event files.

Learning source properties To demonstrate that the learned latents contain rich information about corresponding X-ray sources, we explored relations between various source properties and the learned latents. In Fig. 3, we observe clear continuous trend of hardness ratios (hard-medium, medium-soft, hard-soft, respectively) in PCA space, as well as clear clustering of variable sources (medium variability probability, broad variability probability, and broad variability index) in tSNE space. Moreover, we performed prediction tasks on various source properties and collected results in Table 1. We see that using vanilla Random Forests without tuning, one can obtain remarkable accuracy on hardness ratio prediction, variable source classification as well as source type prediction. Specifically, with much less tuning effort, our source type classification accuracies are comparable to [17] and slightly lower than [25], which uses additional optical and infrared information.

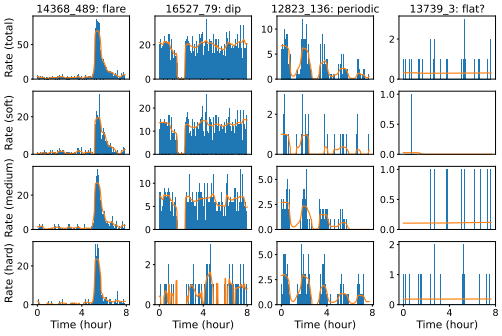

Figure 2: Binned event files vs light curves reconstructed by PPAD. Rate from top row to bottom row: total, high, medium, low. Event files are binned every 5 minutes, and rates are normalized the same way.

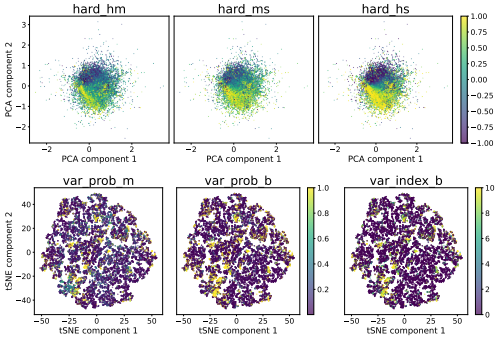

Figure 3: Top: top 2 principal components of latent features colored by hardness ratios. Bottom: tSNE components of latent features colored by variability probabilities/indices.

Anomaly detection To mimic astronomical sources discovery, we performed anomaly detection on learned latents. With a target source in mind, we search for nearest neighbors of the target latents in the embedding space. As Fig. 4 shows, 5 of the 7 neighbors of the targeted flare also exhibit some flare-like behavior, demonstrating the potential of PPAD in discovering interesting unlabeled sources.

Regression Traget	MSE	R^2
hard_ms	0.02	0.87
hard_hm	0.01	0.88
hard_hs	0.02	0.93
Classification Target	**Accuracy**	**F1 Score**
var_index_b > 0.5?	0.92	0.63
source type	0.62	0.25
YSO vs AGN	0.75	0.70

Table 1: Regression/classification performance using learned latent features. All models use a random forest with 100 trees and default hyper-parameters Train-test split is $0.8 - 0.2$ without validation set. SMOTE is applied in classification case to resolve class imbalance.

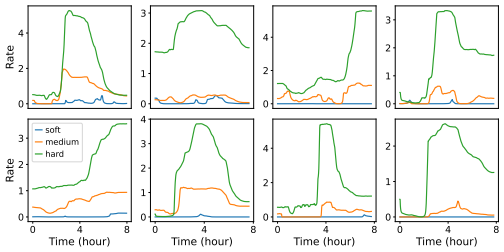

Figure 4: Targeted anomaly (upper left) and 7 found sources who are closest in the latent space. Most of the found sources exhibit flare-like behavior in one or more of the energy bands, as the targeted anomaly source does.

4 Conclusion

This work proposes PPAD, a novel framework for end-to-end unsupervised X-ray sources learning. PPAD simultaneously reconstructs light curves and learns latent representations that are useful for various downstream tasks. Extensive experiments validates the superior performance of PPAD. A notable limitation of PPAD is its relatively costly test time optimization compared Auto-encoders. Another limitation is that one need to tune the network capability in order to strike a good balance between light curve reconstruction quality and downstream task performance. Future work aims to address these limitations as well as apply PPAD to the full Chandra source catalog, releasing classifications and anomaly scores for all sources.

References

[1] Ignacio Becker, Karim Pichara, Márcio Catelan, Pavlos Protopapas, Carlos Aguirre, and Fatemeh Nikzat. Scalable end-to-end recurrent neural network for variable star classification. *Monthly Notices of the Royal Astronomical Society*, 493(2):2981–2995, 2020.

[2] Shubham Bhardwaj, Maria G Dainotti, Sachin Venkatesh, Aditya Narendra, Anish Kalsi, Enrico Rinaldi, and Agnieszka Pollo. Grb optical and x-ray plateau properties classifier using unsupervised machine learning. *Monthly Notices of the Royal Astronomical Society*, 525(4):5204–5223, 2023.

[3] Webster Cash. Parameter estimation in astronomy through application of the likelihood ratio. *Astrophysical Journal, Part 1, vol. 228, Mar. 15, 1979, p. 939-947.*, 228:939–947, 1979.

[4] Tom Charnock and Adam Moss. Deep recurrent neural networks for supernovae classification. *The Astrophysical Journal Letters*, 837(2):L28, 2017.

[5] Ian N. Evans, Janet D. Evans, J. Rafael Martínez-Galarza, Joseph B. Miller, Francis A. Primini, Mojegan Azadi, Douglas J. Burke, Francesca M. Civano, Raffaele D'Abrusco, Giuseppina Fabbiano, Dale E. Graessle, John D. Grier, John C. Houck, Jennifer Lauer, Michael L. McCollough, Michael A. Nowak, David A. Plummer, Arnold H. Rots, Aneta Siemiginowska, and Michael S. Tibbetts. The Chandra Source Catalog Release 2 Series. *arXiv e-prints*, page arXiv:2407.10799, July 2024.

[6] Ian N Evans, Francis A Primini, Kenny J Glotfelty, Craig S Anderson, Nina R Bonaventura, Judy C Chen, John E Davis, Stephen M Doe, Janet D Evans, Giuseppina Fabbiano, et al. The chandra source catalog. *The Astrophysical Journal Supplement Series*, 189(1):37, 2010.

[7] Daniel Giles and Lucianne Walkowicz. Systematic serendipity: A test of unsupervised machine learning as a method for anomaly detection. *Monthly Notices of the Royal Astronomical Society*, 484(1):834–849, 2019.

[8] PC Gregory and Thomas J Loredo. A new method for the detection of a periodic signal of unknown shape and period. *Astrophysical Journal, Part 1 (ISSN 0004-637X), vol. 398, no. 1, Oct. 10, 1992, p. 146-168., Research supported by NSERC*, 398:146–168, 1992.

[9] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[10] Diederik P Kingma. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[11] Miloš Kovačević, Mario Pasquato, Martino Marelli, Andrea De Luca, Ruben Salvaterra, and A Belfiore. Exploring x-ray variability with unsupervised machine learning-i. self-organizing maps applied to xmm-newton data. *Astronomy & Astrophysics*, 659:A66, 2022.

[12] A. Merloni, G. Lamer, T. Liu, M. E. Ramos-Ceja, H. Brunner, E. Bulbul, K. Dennerl, V. Doroshenko, M. J. Freyberg, S. Friedrich, E. Gatuzz, A. Georgakakis, F. Haberl, Z. Igo, I. Kreykenbohm, A. Liu, C. Maitra, A. Malyali, M. G. F. Mayer, K. Nandra, P. Predehl, J. Robrade, M. Salvato, J. S. Sanders, I. Stewart, D. Tubín-Arenas, P. Weber, J. Wilms, R. Arcodia, E. Artis, J. Aschersleben, A. Avakyan, C. Aydar, Y. E. Bahar, F. Balzer, W. Becker, K. Berger, T. Boller, W. Bornemann, M. Brüggen, M. Brusa, J. Buchner, V. Burwitz, F. Camilloni, N. Clerc, J. Comparat, D. Coutinho, S. Czesla, S. M. Dannhauer, L. Dauner, T. Dauser, J. Dietl, K. Dolag, T. Dwelly, K. Egg, E. Ehl, S. Freund, P. Friedrich, R. Gaida, C. Garrel, V. Ghirardini, A. Gokus, G. Grünwald, S. Grandis, I. Grotova, D. Gruen, A. Gueguen, S. Hämmerich, N. Hamaus, G. Hasinger, K. Haubner, D. Homan, J. Ider Chitham, W. M. Joseph, A. Joyce, O. König, D. M. Kaltenbrunner, A. Khokhriakova, W. Kink, C. Kirsch, M. Kluge, J. Knies, S. Krippendorf, M. Krumpe, J. Kurpas, P. Li, Z. Liu, N. Locatelli, M. Lorenz, S. Müller, E. Magaudda, C. Mannes, H. McCall, N. Meidinger, M. Michailidis, K. Migkas, D. Muñoz-Giraldo, B. Musiimenta, N. T. Nguyen-Dang, Q. Ni, A. Olechowska, N. Ota, F. Pacaud, T. Pasini, E. Perinati, A. M. Pires, C. Pommranz, G. Ponti, K. Poppenhaeger, G. Pühlhofer, A. Rau, M. Reh, T. H. Reiprich, W. Roster, S. Saeedi, A. Santangelo, M. Sasaki, J. Schmitt, P. C. Schneider, T. Schrabback, N. Schuster, A. Schwope, R. Seppi, M. M. Serim, S. Shreeram, E. Sokolova-Lapa, H. Starck, B. Stelzer, J. Stierhof, V. Suleimanov, C. Tenzer, I. Traulsen, J. Trümper, K. Tsuge, T. Urrutia, A. Veronica, S. G. H. Waddell, R. Willer, J. Wolf, M. C. H. Yeung, A. Zainab, F. Zangrandi, X. Zhang, Y. Zhang, and X. Zheng. The SRG/eROSITA all-sky survey. First X-ray catalogues and data release of the western Galactic hemisphere. , 682:A34, February 2024.

[13] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[14] Brett Naul, Joshua S Bloom, Fernando Pérez, and Stéfan Van Der Walt. A recurrent neural network for classification of unevenly sampled variable stars. *Nature Astronomy*, 2(2):151–155, 2018.

[15] Jakub K Orwat-Kapola, Antony J Bird, Adam B Hill, Diego Altamirano, and Daniela Huppenkothen. Light-curve fingerprints: an automated approach to the extraction of x-ray variability patterns with feature aggregation–an example application to grs 1915+ 105. *Monthly Notices of the Royal Astronomical Society*, 509(1):1269–1290, 2022.

[16] Jeong Joon Park, Peter Florence, Julian Straub, Richard Newcombe, and Steven Lovegrove. Deepsdf: Learning continuous signed distance functions for shape representation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 165–174, 2019.

[17] Víctor Samuel Pérez-Díaz, Juan Rafael Martínez-Galarza, Alexander Caicedo, and Raffaele D'Abrusco. Unsupervised machine learning for the classification of astrophysical x-ray sources. *Monthly Notices of the Royal Astronomical Society*, 528(3):4852–4871, 2024.

[18] Jakob Gulddahl Rasmussen. Lecture notes: Temporal point processes and the conditional intensity function. *arXiv preprint arXiv:1806.00221*, 2018.

[19] Benjamin J Ricketts, James F Steiner, Cecilia Garraffo, Ronald A Remillard, and Daniela Huppenkothen. Mapping the x-ray variability of grs 1915+ 105 with machine learning. *Monthly Notices of the Royal Astronomical Society*, 523(2):1946–1966, 2023.

[20] Christopher J Shallue and Andrew Vanderburg. Identifying exoplanets with deep learning: A five-planet resonant chain around kepler-80 and an eighth planet around kepler-90. *The Astronomical Journal*, 155(2):94, 2018.

[21] Vincent Sitzmann, Michael Zollhöfer, and Gordon Wetzstein. Scene representation networks: Continuous 3d-structure-aware neural scene representations. *Advances in Neural Information Processing Systems*, 32, 2019.

[22] Yanke Song, V Ashley Villar, and Juan Rafael Martínez-Galarza. A poisson-process autodecoder for astrophysical, time-variable, x-ray sources. In *Machine Learning and the Physical Sciences Workshop @ NeurIPS 2024*, 2024.

[23] N. A. Webb, M. Coriat, I. Traulsen, J. Ballet, C. Motch, F. J. Carrera, F. Koliopanos, J. Authier, I. de la Calle, M. T. Ceballos, E. Colomo, D. Chuard, M. Freyberg, T. Garcia, M. Kolehmainen, G. Lamer, D. Lin, P. Maggi, L. Michel, C. G. Page, M. J. Page, J. V. Perea-Calderon, F. X. Pineau, P. Rodriguez, S. R. Rosen, M. Santos Lleo, R. D. Saxton, A. Schwope, L. Tomás, M. G. Watson, and A. Zakardjian. The XMM-Newton serendipitous survey. IX. The fourth XMM-Newton serendipitous source catalogue. , 641:A136, September 2020.

[24] Sara Webb, Michelle Lochner, Daniel Muthukrishna, Jeff Cooke, Chris Flynn, Ashish Mahabal, Simon Goode, Igor Andreoni, Tyler Pritchard, and Timothy MC Abbott. Unsupervised machine learning for transient discovery in deeper, wider, faster light curves. *Monthly Notices of the Royal Astronomical Society*, 498(3):3077–3094, 2020.

[25] Hui Yang, Jeremy Hare, Oleg Kargaltsev, Igor Volkov, Steven Chen, and Blagoy Rangelov. Classifying unidentified x-ray sources in the chandra source catalog using a multiwavelength machine-learning approach. *The Astrophysical Journal*, 941(2):104, 2022.

[26] Yanxia Zhang, Yongheng Zhao, and Xue-Bing Wu. Classification of 4xmm-dr9 sources by machine learning. *Monthly Notices of the Royal Astronomical Society*, 503(4):5263–5273, 2021.

A Implementation Details

We provide implementation details for our experiments in Section 3.

Data Among the collected $109\,656$ event files, most has low signal-to-noise ratio. Therefore, we first created a higher quality dataset by removing each event file with probability $1/(1+\exp(900^{0.99}cdotl^{0.01}-900))$, where l is the length (number of arrivals) in the event file. The filtering effectively removes mostly low-count event files, and resulted in a high quality dataset of $14\,891$ event files. Further, to prevent the model from learning the peculiarity at $t=0$, we randomly shifted each event file by $t \sim \text{Unif}[t_1, t_2]$ where t_1, t_2 are first two arrival times.

Network architecture The ResNet takes a $(d_{\text{latent}}+2L+1)$-dimensional input with $d_{\text{latent}}=8$ and $L=12$. It maps the input to a 512-dimensional hidden vector via a fully connected input layer.. The hidden vector is then passed through 5 fully connected ResNet blocks, maintaining dimensionality. Lastly, a fully connected output layer maps the hidden vector to the output of dimension 3, representing light curve value at $K=3$ energy bins. ReLU activation is used in between consecutive layers.

Loss function For hyperparameters in Eqn. 4, we used $\lambda_{\text{TV}}=10, \lambda_{\text{latent}}=1$. The time interval $[0,T)$ with $T=8$ hours is divided into 2048 evenly spaced bins when we calculate the integral from $\mathcal{L}_{\text{neg-loglikelihood}}$ and a part of $\mathcal{L}_{\text{TV}}$.

Training The training has 3 stages. For stage 1, we train both the network and the latent using the filtered high quality dataset for 1200 epochs. For stage 2, we switch to the full dataset, but freeze the network and only train the latents for 200 epochs to provide a good initialization. For stage 3, we again train the latents and networks together for 300 epochs. We use the Adam optimizer [10] with default hyperparameters. The learning rate for the latents is 1e-3 for Stages 1&2 and 1e-4 for Stage 3. The learning rate for network weights is always $1/10$ of that for the latents. We use a batch size of 64. The whole training takes approximately 5 days on a single Nvidia V100 GPU.

Experiments All regression/classification experiments use a Random Forest with 100 trees and default hyperparameters, with a 80%-20% train-test split. Standard deviation of numbers are provided by running the experiments 5 times each, using random states 2020-2024.

[22]

B Acknowledgement

This research has made use of data obtained from the Chandra Source Catalog provided by the Chandra X-ray Center (CXC).