A deep learning framework for jointly extracting spectra and source-count distributions in astronomy

Florian Wolf[1]   **Florian List**[1]   **Nicholas L. Rodd**[2]   **Oliver Hahn**[1]

[1]University of Vienna, Department of Astrophysics & Department of Mathematics
{florian.wolf, florian.list, oliver.hahn}@univie.ac.at
[2]CERN, Theoretical Physics Department
nick.rodd@cern.ch

Abstract

Astronomical observations typically provide three-dimensional maps, encoding the distribution of the observed flux in (1) the two angles of the celestial sphere and (2) energy/frequency. An important task regarding such maps is to statistically characterize populations of point sources too dim to be individually detected. As the properties of a single dim source will be poorly constrained, instead one commonly studies the population as a whole, inferring a source-count distribution (SCD) that describes the number density of sources as a function of their brightness. Statistical and machine learning methods for recovering SCDs exist; however, they typically entirely neglect spectral information associated with the energy distribution of the flux. We present a deep learning framework able to jointly reconstruct the spectra of different emission components and the SCD of point-source populations. In a proof-of-concept example, we show that our method accurately extracts even complex-shaped spectra and SCDs from simulated maps.

1   Introduction

In many senses, the most fundamental problem in astronomy is to provide a complete catalog of point sources (PSs) within a map of the sky. Individual sources are traditionally characterized down to a detection threshold intrinsic to the instrument which collected the map. For sources below this boundary, characterizing even basic properties such as their location becomes less meaningful; however, properties of the entire population of dim sources can be reliably inferred at a *statistical* level, through, for instance, a source-count distribution (SCD), which reports the distribution of sources as a function of brightness.

Likelihood-based statistical methods have been applied to extract SCDs at various wavelengths—radio, optical, X-ray, gamma-ray, and even in high-energy neutrinos datasets, see e.g. [1, 6, 13, 28, 34, 37, 38, 41, 42, 43]. In principle; these methods are optimal for perfectly modeled data (apart from inter-pixel correlations introduced by the instrument point-spread function (PSF) [13]); however, data mismodeling – present to a certain degree in every analysis – biases the results (e.g. [7, 12, 25, 26, 27].) Recently, deep learning methods have been developed, with a particular focus on gamma-ray astronomy and the data collected by the *Fermi* satellite [15] (e.g. [8, 10, 11, 31]), several of which can determine SCDs from the *Fermi* photon-count maps [5, 30, 36]. Whilst these deep learning methods have their own systematics (see e.g. [30] for an exploration), they also hold the potential of being able to extend beyond what was possible with likelihood-based approaches.

To be explicit, an important limitation of the likelihood approach is that it is by construction blind to any energy or frequency information, studying only the distribution of flux on the celestial sphere. In principle, one can write the likelihood with the energy dimension included, but evaluating this proves to be a significant combinatorial problem that remains unresolved [13]. Deep learning methods,

by contrast, should not be limited in this manner: the input data can be provided binned in energy, allowing for basic quantities like the spectrum (energy distribution) of sub-threshold PS populations to be inferred, without the need for an analytic expression of the likelihood. Despite this promise, existing deep learning approaches have generally ignored energy information.

In this work, we present a first demonstration of a deep learning framework that enables simultaneous inference of spectra and SCDs from astronomical data, thereby exploiting the full information collected by the telescope. We choose to focus on gamma-ray astronomy given the wide class of techniques developed at that wavelength, but we emphasize that our approach could be readily extended to other energies or messengers.

2 Methods

2.1 Modeling

We consider photon-count maps (e.g. of the entire sky) $\boldsymbol{x} = (x_n^e)_{n=1,\ldots,N_{\mathrm{pix}}}^{e=1,\ldots,E} \in \mathbb{N}_0^{N_{\mathrm{pix}} \times E}$, which contain the counts detected in E energy bands in each of the N_{pix} pixels.[1] We model these counts as originating from T emission components, each of which has an individual spatial morphology and spectrum, i.e. $\boldsymbol{x} = \sum_{t=1}^{T} \boldsymbol{x}^t$, where $\boldsymbol{x}^t \in \mathbb{N}_0^{N_{\mathrm{pix}} \times E}$ are the counts from component t.

We distinguish between two types of emission components:

1. **Poissonian** components (e.g. different foregrounds), for which we assume the spatial morphology to be known so that the counts associated with such a component t are simply described by a Poisson draw $\boldsymbol{x}^{t,e} \sim \mathrm{Pois}(A^t \xi^{t,e} \bar{\boldsymbol{x}}^{t,e})$. Here, $\bar{\boldsymbol{x}}^{t,e} \in \mathbb{R}_{\geq 0}^{N_{\mathrm{pix}} \times E}$ models the (known and normalized) spatial morphology of the counts in each pixel and energy bin, A^t is an (unknown) normalization coefficient that characterizes the total brightness of the component, and $\xi^{t,e} \in [0,1]$ with $\sum_{e=1}^{E} \xi^{t,e} = 1$ encodes its (unknown) spectrum.

2. **Point-source** components, for which the situation is somewhat more complicated: here, the underlying spatial distribution p_{PS}^t of the PSs belonging to component t is assumed to be known, but not the individual location of each source. In this proof-of-concept paper, we take the PSs to be uniformly distributed across the sky for simplicity. So for the (unknown) longitude l_s^t and latitude b_s^t of each source s^t, one has $(l_s^t, b_s^t) \sim p_{\mathrm{PS}}^t$ (which is of course the same for all energy bands e). The total number of sources is modeled as a Poissonian random variable $N_{\mathrm{PS}}^t \sim \mathrm{Pois}(\bar{N}_{\mathrm{PS}}^t)$, where the expected number of sources $\bar{N}_{\mathrm{PS}}^t$ associated with component t is unknown. Each source s^t emits a certain number of counts $c_s^{t,e}$ in each energy band e according to a Poisson process $c_s^{t,e} \sim \mathrm{Pois}(\xi^{t,e} \mu_s^t)$, where $\mu_s^t \sim \mathrm{SCD}$ reflects the brightness of that source (integrated over all energy bands), and $\xi^{t,e} \in [0,1]$ with $\sum_{e=1}^{E} \xi^{t,e} = 1$ characterizes the spectrum (which we take to be the same for all sources belonging to component t). The counts $c_s^{t,e}$ from each source s^t are localized around the pixel associated with the sky coordinates (l_s^t, b_s^t), but smeared within a vicinity of this pixel due to the (energy-dependent) instrument PSF. The distribution of the values μ_s^t (i.e., the SCD of component t) and the relative spectrum $\xi^{t,e}$ are both unknown.

Our framework simultaneously targets two objectives: (1) the inference of the **spectra** as described by the spectral shape $\xi^{t,e}$ and an appropriate normalization for each component (both Poissonian and point-like), and (2) the inference of the **SCDs** of all PS components, i.e. the distribution of the source brightnesses μ_s^t.

2.2 Extracting spectra

For recovering spectra from the data, we extend the Gaussian maximum likelihood estimation proposed by Ref. [31] for the inference of count fractions to multiple spectral bands. For simplicity, we neglect the Poisson scatter when estimating spectra and use the actually *realized* counts, rather than the underlying spectral shapes $\xi^{t,e}$ for the labels, as the total number of counts for each component is typically large. Inferring the spectra in terms of counts per energy bin simplifies the task, as this leads to a unified approach for all components (Poissonian and point-like).

---

[1]Herein, we ignore the sky exposure and use the notions of "counts" and "flux" interchangeably.

Since the total number of detected counts in each energy band summed over all components is known from observation, extracting the spectra is tantamount to estimating the *fraction* of counts contributed by each component, separately for each energy bin.

Given a map $\boldsymbol{x}$, we therefore task a neural network (NN) with trainable parameters $\boldsymbol{\theta}$ to provide estimates $\tilde{\boldsymbol{y}}_{\boldsymbol{\theta}} = (\tilde{y}_{\boldsymbol{\theta}}^{t,e}) \in [0,1]^{T \times E}$ for the relative counts of all components t in each energy bin e,

$$\tilde{y}_{\boldsymbol{\theta}}^{t,e} \approx y^{t,e} = \left(\sum_{n=1}^{N_{\text{pix}}} x_n^{t,e} \right) \Big/ \left(\sum_{n=1}^{N_{\text{pix}}} x_n^e \right). \tag{1}$$

We train the NN to estimate posteriors by minimizing the negative Gaussian log-likelihood

$$\mathcal{L}_{\text{spectrum}}(\tilde{\boldsymbol{y}}_{\boldsymbol{\theta}}(\boldsymbol{x}), \boldsymbol{y}) = \sum_{t=1}^{T} \sum_{e=1}^{E} \left(\frac{1}{2(\tilde{\sigma}_{\boldsymbol{\theta}}^{t,e}(\boldsymbol{x}))^2} \left(\tilde{y}_{\boldsymbol{\theta}}^{t,e}(\boldsymbol{x}) - y^{t,e} \right)^2 + \frac{\ln \left[(\tilde{\sigma}_{\boldsymbol{\theta}}^{t,e}(\boldsymbol{x}))^2 \right]}{2} \right). \tag{2}$$

The standard deviation $\tilde{\sigma}_{\boldsymbol{\theta}}^{t,e}(\boldsymbol{x})$ associated with the aleatoric uncertainty of the prediction $\tilde{y}_{\boldsymbol{\theta}}^{t,e}(\boldsymbol{x})$ is estimated by the NN itself, in such a way as to minimize Eq. (2). This loss function assumes the count fractions of the different components in different energy bands to be uncorrelated (which is clearly an approximation, just as the assumption of Gaussianity); however, extending our method to a full covariance structure along the component and/or energy dimension is possible. Another interesting avenue – which we will investigate in future work – is weighting the residual in each energy bin by the total number of counts in that bin, in order to penalize the NN in proportion to its prediction error in terms of total rather than relative counts. Once the NN has been trained, the spectrum of each component t in terms of total counts per energy bin e is readily obtained from $\tilde{\boldsymbol{y}}_{\boldsymbol{\theta}}(\boldsymbol{x})$ by solving for $\sum_{n=1}^{N_{\text{pix}}} x_n^{t,e}$ in Eq. (1).

2.3 Extracting source-count distributions

For extracting the SCDs of the PS components, we adopt the non-parametric approach proposed by Ref. [30] and discretize the SCD of each PS component as a histogram $\boldsymbol{u} = (u^m) \in [0,1]^M$, where M is the total number of (logarithmic) count bins (which is a hyperparameter). As mentioned above, we assume the SCDs to be energy-independent and infer a single SCD for each PS component. Therefore, the only difference in comparison with the framework in that reference is that the input maps $\boldsymbol{x} \in \mathbb{N}^{N_{\text{pix}} \times E}$ on which the NN is trained are energy-dependent, but the final recovered SCDs remain energy-independent, so we will only summarize the key points here.

For computing the SCD value u^m in bin m, we sum up the expected counts contributed by all those PSs whose expected number of counts falls within the m-th count bin; then, we normalize the histogram to sum up to unity, i.e. $\sum_{m=1}^{M} u^m = 1$. This is because the normalization in terms of total counts associated with each component can be obtained from the estimated spectra as $\sum_{e=1}^{E} \sum_{n=1}^{N_{\text{pix}}} x_n^{t,e}$, so estimating relative SCDs is sufficient.

For extracting the SCDs, we use a second NN with parameters $\boldsymbol{\vartheta}$, trained with a loss function rooted in quantile regression [24]. Specifically, we apply the pinball loss [16, 24] to the *cumulative* histogram with $U_m := \sum_{j=1}^{m} u_j$ in each bin, i.e. [29]

$$\mathcal{L}_{\text{SCD}}^{\tau}(\tilde{\boldsymbol{u}}_{\boldsymbol{\vartheta}}(\boldsymbol{x}; \tau), \boldsymbol{u}) = \frac{1}{M} \sum_{m=1}^{M} \left[\left(\tilde{U}_{\boldsymbol{\vartheta}}^m(\boldsymbol{x}; \tau) - U^m \right) \left(\tau - \mathbb{I} \left[\tilde{U}_{\boldsymbol{\vartheta}}^m(\boldsymbol{x}; \tau) < U^m \right] \right) \right], \tag{3}$$

where $\tilde{\boldsymbol{u}}_{\boldsymbol{\vartheta}}(\boldsymbol{x}; \tau)$ is the SCD histogram estimated by the NN at quantile level τ with cumulative counterpart $\tilde{\boldsymbol{U}}_{\boldsymbol{\vartheta}}(\boldsymbol{x}; \tau)$, and $\mathbb{I}[\cdot]$ is the indicator function. The queried quantile level $\tau \in (0, 1)$, which is fed as an additional input to the neural network, is drawn randomly as $\tau \sim U(0, 1)$ during training and can be specified at inference time. As this loss function is minimized by the τ-th quantile level, it enables us to estimate the *distribution* of possible SCDs, i.e. a median estimate ($\tau = 0.5$, in which case the loss function reduces to the ℓ^1-loss applied to the cumulative histograms) plus uncertainties.

2.4 Deep learning model

We use two separate NNs stacked on top of each other, one for extracting the spectra, and another one for inferring the SCDs. The reason for sequentially training two NNs is that, in principle, the spectral

information extracted by the first NN could be exploited by the second NN tasked with the SCD inference; we will investigate this in future work. To accommodate large regions of interest (ROI) for which the sky curvature becomes relevant, we build our NNs onto the `DeepSphere` architecture ([14, 39], see [30, 31, 36] for gamma-ray applications), which models the sky as a graph based on the `Healpix` tessellation of the sphere [20] and provides graph counterparts of convolution and pooling.

3   Results

To demonstrate the effectiveness of our method, we consider photon counts in a $25°$ ROI. The photons originate from an isotropic PS component and a smooth (Poissonian) background component (i.e., $T = 2$). We model the smooth background with the pion decay & bremsstrahlung component of the `Model O` template presented in Ref. [7] (building on Refs. [32, 33]). Although we avoid some of the technicalities present in a realistic analysis (such as the non-uniform sky exposure by gamma-ray telescopes etc.), this experiment comprises the major difficulties one encounters in practice, namely the presence of multiple overlapping emission components (with a background whose morphology varies as a function of energy) and a *Fermi*-like PSF that deteriorates with decreasing energy, described by a combination of King functions.[2]

We model the SCD and the spectra of the isotropic PSs as a mixture of two truncated skew-normal distributions whose location, scale, and skewness parameters are drawn from wide distributions. The motivation for considering a mixture distribution is to highlight the ability of our framework to extract complex spectra and SCDs and, from a physical point of view, to account for the possibility that two distinct (in terms of spectrum and SCD) PS populations with a similar spatial morphology are hidden in the data. For the Poissonian background, we draw the spectrum from a single truncated skew-normal distribution. We discretize the SCDs into $M = 22$ count bins and use $E = 10$ energy bins for the spectra. Our training dataset consists of 940,000 simulated maps at `Healpix` resolution parameter $N_{side} = 256$, and we train our NNs to infer the spectra of both components and the SCD of the PS population. We remark that our SCD estimation framework directly carries over to multiple PS populations with different spatial morphologies (as in Ref. [30, Fig. 8]).

Figure 1 shows the spectra in terms of relative counts, i.e. y in Eq. (1), and SCDs for three maps from the test dataset. We selected these maps in order to showcase some diverse and instructive configurations for the spectra and SCDs, not based on the performance of our NNs. We will present a thorough quantitative analysis and an application of our method to *Fermi* data elsewhere.

The NN accurately recovers the complex shapes of the count fractions which arise from the interplay between the different spectral shapes. The true spectra approximately fall within the 1σ region for the maps shown in the figure, and the performance on other maps is similar.

As for the SCDs, the NN is able to obtain accurate estimates for populations above the one photon line (dashed orange). Below, predictions become less precise: in view of PS emission being degenerate with Poisson emission in the limit of $\ll 1$ expected counts per source, exactly pinpointing the location of the SCD in this regime is no longer possible from a statistical point of view, see e.g. Ref. [30, Fig. 9]. The first and second panel show maps for which the bimodality of the SCD clearly indicates the presence of a faint and a brighter PS subcomponent (described by the mixture of two skew-normal distributions, see above). Our NN produces larger uncertainties for the faint subcomponent (which our NN expresses in terms of the *cumulative* SCDs that we converted to density histograms, indicated by the blue-to-red colored regions), just as expected.

4   Conclusions

We introduced a machine-learning framework for the simultaneous extraction of SCDs and spectra from astronomical data, where we make use of the photon energy information. Including spectral information in the analysis enables more accurate modeling on one hand (as all analysis components such as spatial templates for the different emission components, the instrument PSF, masks for known PSs etc. can be made dependent on the energy), and more informative output on the other hand, as it allows the model to infer the spectra in addition to the SCDs of the PS components. Also, physically

---

[2]https://fermi.gsfc.nasa.gov/ssc/data/analysis/documentation/Cicerone/Cicerone_LAT_IRFs/IRF_PSF.html
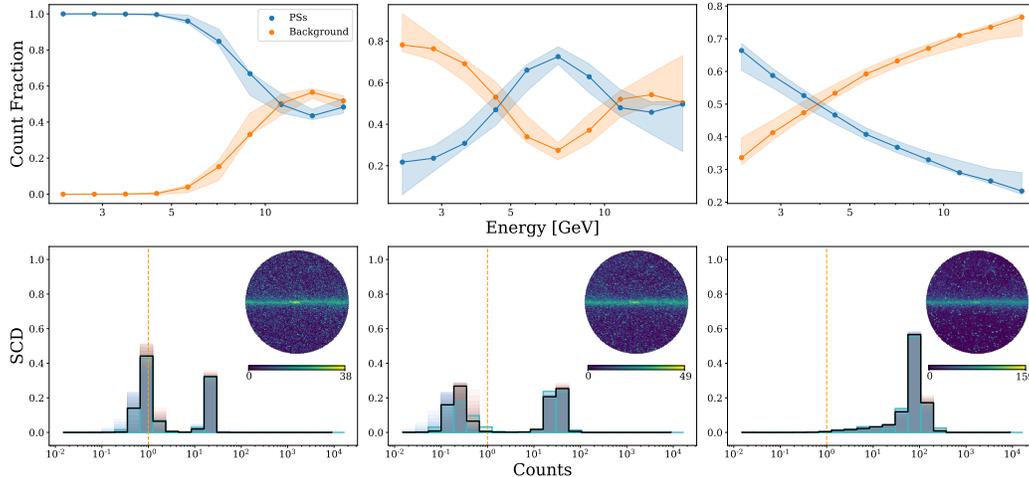
Figure 1: NN predictions for three selected test maps. *Top:* Count fractions for the isotropic PS and the diffuse background component as a function of photon energy. The solid lines represent the true values, whereas the shaded areas cover the 1σ NN predictions. *Bottom:* True vs. reconstructed SCD of the isotropic PS component. Here, the black line is the median prediction (bin-wise in terms of the associated cumulative histogram, see Sec. 2.3), while the faint colored regions indicate the quantiles from 5% to 95%. The blue line shows the true SCD. The orange vertical line corresponds to a single expected photon count per source. In the top right corner, we show the maps (summed over all energy bins) to which the spectra and SCDs belong. The color map is scaled logarithmically. The SCDs are modeled as a mixture of two skew-normal distributions, and their bimodality is clearly visible.

motivated tight priors on the spectra should enable the NNs to obtain tighter constraints as they can learn the correlations between different energy bins; we will carry out a systematic exploration study in this direction in future work. Further, the method could be improved by incorporating the covariance between components or energies in the training.

There are many interesting applications for our method in practice, in particular for *Fermi* data. For instance, there has been an ongoing discussion in the field regarding the composition of the high-latitude sky in the GeV range, specifically in terms of the contributions of star-forming galaxies [40] and blazars (e.g. [4, 35]). Also, the origin of the Galactic Center Excess [18, 22], for which the most popular contenders are a faint population of millisecond pulsars (e.g. [2, 17, 19]) or dark matter annihilation (e.g. [3, 9, 19]), is still unknown. Jointly exploiting spectral properties and SCDs is crucial for further progress on both fronts.

Our method is also ideally suited for applications in X-ray (or neutrino) astronomy, where one often has to deal with complex ROIs and non-isotropic PSFs (e.g. [21]), which poses difficulties for likelihood methods, but is straightforward for simulation-based approaches. There are also clear targets for the future. To begin with, in isolated cases we have observed a marked degradation in the performance of the spectrum reconstruction, and improving performance in those cases will be important before turning to real data. Further, the leading systematic uncertainty for template-based methods is the use of imperfect background models, or equivalently background models that are missing an emission component. Quantifying the performance of our approach in the presence of such imperfections will allow us to determine the systematic uncertainties associated with inferences drawn from actual astrophysical maps.

The computational results presented have been achieved using the Vienna Scientific Cluster (VSC).

References

[1] M. G. Aartsen et al. A Search for Neutrino Point-source Populations in 7 yr of IceCube Data with Neutrino-count Statistics. *The Astrophysical Journal*, 893(2):102, 2020. 10.3847/1538-4357/ab7af9.

[2] K. N. Abazajian. The consistency of Fermi-LAT observations of the galactic center with a millisecond pulsar population in the central stellar cluster. *Journal of Cosmology and Astroparticle Physics*, 2011(03):010, 2011. 10.1088/1475-7516/2011/03/010.

[3] P. Agrawal, B. Batell, P. J. Fox, and R. Harnik. WIMPs at the galactic center. *Journal of Cosmology and Astroparticle Physics*, 2015(05):011, 2015. 10.1088/1475-7516/2011/03/010.

[4] M. Ajello, D. Gasparrini, M. Sánchez-Conde, G. Zaharijas, M. Gustafsson, J. Cohen-Tanugi, C. Dermer, Y. Inoue, D. Hartmann, M. Ackermann, et al. The origin of the extragalactic gamma-ray background and implications for dark matter annihilation. *The Astrophysical Journal Letters*, 800(2):L27, 2015. 10.1088/2041-8205/800/2/L27.

[5] A. Amerio, A. Cuoco, and N. Fornengo. Extracting the gamma-ray source-count distribution below the Fermi-LAT detection limit with deep learning. *Preprint*, 2023. arXiv:2302.01947.

[6] X. Barcons. Confusion Noise and Source Clustering. *The Astrophysical Journal*, 396:460, Sept. 1992.

[7] M. Buschmann, N. L. Rodd, B. R. Safdi, L. J. Chang, S. Mishra-Sharma, M. Lisanti, and O. Macias. Foreground mismodeling and the point source explanation of the Fermi Galactic Center excess. *Physical Review D*, 102(2):23023, 2020. 10.1103/PhysRevD.102.023023.

[8] A. Butter, M. Krämer, S. Manconi, and K. Nippel. Searching for dark matter subhalos in the Fermi-LAT catalog with Bayesian neural networks. *Preprint*, 2023. arXiv:2304.00032.

[9] F. Calore, I. Cholis, and C. Weniger. Background model systematics for the Fermi GeV excess. *Journal of Cosmology and Astroparticle Physics*, 2015(03):038, 2015. 10.1088/1475-7516/2015/03/038.

[10] S. Caron, C. Eckner, L. Hendriks, G. Jóhannesson, R. Ruiz de Austri, and G. Zaharijas. Mind the gap: the discrepancy between simulation and reality drives interpretations of the Galactic Center Excess. *Journal of Cosmology and Astroparticle Physics*, 2023(6), 2023. 10.1088/1475-7516/2023/06/013.

[11] S. Caron, G. A. Gómez-Vargas, L. Hendriks, and R. R. de Austri. Analyzing γ rays of the Galactic Center with deep learning. *Journal of Cosmology and Astroparticle Physics*, 2018(05):058–058, 2018. 10.1088/1475-7516/2018/05/058.

[12] L. J. Chang, S. Mishra-Sharma, M. Lisanti, M. Buschmann, N. L. Rodd, and B. R. Safdi. Characterizing the nature of the unresolved point sources in the Galactic Center: An assessment of systematic uncertainties. *Physical Review D*, 101(2):023014, 2020. 10.1103/PhysRevD.101.023014.

[13] G. H. Collin, N. L. Rodd, T. Erjavec, and K. Perez. A Compound Poisson Generator Approach to Point-source Inference in Astrophysics. *The Astrophysical Journal Supplement Series*, 260(2):29, 2022. 10.3847/1538-4365/ac5cb7.

[14] M. Defferrard, M. Milani, F. Gusset, and N. Perraudin. Deepsphere: a Graph-Based Spherical Cnn. In *8th International Conference on Learning Representations, ICLR 2020*, 2020. arXiv:2012.15000.

[15] Fermi Collaboration. The large area telescope on the fermi gamma-ray space telescope mission. *The Astrophysical Journal*, 697(2):1071–1102, 2009. 10.1088/0004-637X/697/2/1071.

[16] M. Fox and H. Rubin. Admissibility of quantile estimates of a single location parameter. *The Annals of Mathematical Statistics*, pages 1019–1030, 1964. jstor:2238233.

[17] A. Gautam, R. M. Crocker, L. Ferrario, A. J. Ruiter, H. Ploeg, C. Gordon, and O. Macias. Millisecond pulsars from accretion-induced collapse as the origin of the galactic centre gamma-ray excess signal. *Nature Astronomy*, 6(6):703–707, 2022. 10.1038/s41550-022-01658-3.

[18] L. Goodenough and D. Hooper. Possible Evidence For Dark Matter Annihilation In The Inner Milky Way From The Fermi Gamma Ray Space Telescope, 2009. arXiv:0910.2998.

[19] C. Gordon and O. Macías. Dark matter and pulsar model constraints from Galactic Center Fermi-LAT gamma-ray observations. *Physical Review D*, 88:083521, Oct 2013. 10.1103/PhysRevD.88.083521.

[20] K. M. Górski, E. Hivon, A. J. Banday, B. D. Wandelt, F. K. Hansen, M. Reinecke, and M. Bartelmann. HEALPix: A Framework for High-Resolution Discretization and Fast Analysis of Data Distributed on the Sphere. *The Astrophysical Journal*, 622(2):759–771, 2005. 10.1086/427976.

[21] F. A. Harrison, W. W. Craig, F. E. Christensen, C. J. Hailey, W. W. Zhang, S. E. Boggs, D. Stern, W. R. Cook, K. Forster, P. Giommi, et al. The nuclear spectroscopic telescope array (nustar) high-energy x-ray mission. *The Astrophysical Journal*, 770(2):103, 2013. 10.1088/0004-637X/770/2/103.

[22] D. Hooper and L. Goodenough. Dark matter annihilation in the Galactic Center as seen by the Fermi Gamma Ray Space Telescope. *Physics Letters B*, 697(5):412–428, 2011. 10.1016/j.physletb.2011.02.029.

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

[24] R. Koenker and G. Bassett. Regression Quantiles. *Econometrica*, 46(1):33, 1978. 10.2307/1913643.

[25] R. K. Leane and T. R. Slatyer. Revival of the Dark Matter Hypothesis for the Galactic Center Gamma-Ray Excess. *Physical Review Letters*, 123(24):241101, 2019. 10.1103/PhysRevLett.123.241101.

[26] R. K. Leane and T. R. Slatyer. Spurious point source signals in the galactic center excess. *Physical Review Letters*, 125(12):121105, 2020. 10.1103/PhysRevLett.125.121105.

[27] R. K. Leane and T. R. Slatyer. The enigmatic Galactic Center excess: Spurious point sources and signal mismodeling. *Physical Review D*, 102(6):063019, 2020. 10.1103/PhysRevD.102.063019.

[28] S. K. Lee, M. Lisanti, B. R. Safdi, T. R. Slatyer, and W. Xue. Evidence for Unresolved γ-Ray Point Sources in the Inner Galaxy. *Physical Review Letters*, 116(5):051103, 2016. 10.1103/PhysRevLett.116.051103.

[29] F. List. The earth mover's pinball loss: Quantiles for histogram-valued regression. In *International Conference on Machine Learning*, pages 6713–6724. PMLR, 2021. arXiv:2106.02051.

[30] F. List, N. L. Rodd, and G. F. Lewis. Extracting the Galactic Center excess' source-count distribution with neural nets. *Physical Review D*, 104(12):123022, 2021. 10.1103/PhysRevD.104.123022.

[31] F. List, N. L. Rodd, G. F. Lewis, and I. Bhat. Galactic Center Excess in a New Light: Disentangling the γ-Ray Sky with Bayesian Graph Convolutional Neural Networks. *Physical Review Letters*, 125(24):241102, 2020. 10.1103/PhysRevLett.125.241102.

[32] O. Macias, C. Gordon, R. M. Crocker, B. Coleman, D. Paterson, S. Horiuchi, and M. Pohl. Galactic bulge preferred over dark matter for the Galactic centre gamma-ray excess. *Nature Astronomy*, 2:387–392, 2018. 10.1038/s41550-018-0414-3.

[33] O. Macias, S. Horiuchi, M. Kaplinghat, C. Gordon, R. M. Crocker, and D. M. Nataf. Strong evidence that the galactic bulge is shining in gamma rays. *Journal of Cosmology and Astroparticle Physics*, 2019(9):042, 2019. 10.1088/1475-7516/2019/09/042.

[34] D. Malyshev and D. W. Hogg. Statistics of gamma-ray point sources below the Fermi detection limit. *The Astrophysical Journal*, 738(2):181, 2011. 10.1088/0004-637X/738/2/181.

[35] S. Manconi, M. Korsmeier, F. Donato, N. Fornengo, M. Regis, and H. Zechlin. Testing gamma-ray models of blazars in the extragalactic sky. *Physical Review D*, 101:103026, 2020. 10.1103/PhysRevD.101.103026.

[36] S. Mishra-Sharma and K. Cranmer. Neural simulation-based inference approach for characterizing the Galactic Center γ-ray excess. *Physical Review D*, 105(6):063017, 2022. 10.1103/PhysRevD.105.063017.

[37] S. Mishra-Sharma, N. L. Rodd, and B. R. Safdi. NPTFit : A Code Package for Non-Poissonian Template Fitting. *The Astronomical Journal*, 153(6):253, 2017. 10.3847/1538-3881/aa6d5f.

[38] T. Miyaji and R. E. Griffiths. Faint source counts from off-source fluctuation analysis on Chandra observations of the Hubble deep field-north. *The Astrophysical Journal Letters*, 564:L5, 2002. 10.1086/338794.

[39] N. Perraudin, M. Defferrard, T. Kacprzak, and R. Sgier. DeepSphere: Efficient spherical convolutional neural network with HEALPix sampling for cosmological applications. *Astronomy and Computing*, 27:130–146, 2019. 10.1016/j.ascom.2019.03.004.

[40] M. A. Roth, M. R. Krumholz, R. M. Crocker, and S. Celli. The diffuse γ-ray background is dominated by star-forming galaxies. *Nature*, 597(7876):341–344, 2021. 10.1038/s41586-021-03802-x.

[41] P. A. G. Scheuer. A statistical method for analysing observations of faint radio stars. *Proceedings of the Cambridge Philosophical Society*, 53:764–773, Jan. 1957.

[42] H.-S. Zechlin, A. Cuoco, F. Donato, N. Fornengo, and M. Regis. Statistical Measurement of the Gamma-ray Source-count Distribution as a Function of Energy. *The Astrophysical Journal*, 826(2):L31, 2016. 10.3847/2041-8205/826/2/L31.

[43] H.-S. Zechlin, A. Cuoco, F. Donato, N. Fornengo, and A. Vittino. Unveiling the Gamma-Ray Source Count Distribution Below the Fermi Detection Limit With Photon Statistics. *The Astrophysical Journal Supplement Series*, 225(2):18, 2016. 10.3847/0067-0049/225/2/18.

A Details of the neural network architecture and training

We used an Adam optimizer [23] with a learning rate of 5×10^{-4} decaying exponentially at a rate of 0.99985 per iteration. The training dataset consisted of 940,000 maps, and we trained with a batch size of 256 for 80,000 iterations on a single NVIDIA A100 GPUs on the "Vienna Scientific Cluster 5 (VSC5)". The NN architecture can be seen in Tables 1 (spectra) & 2 (SCDs). In total, both NN combined have 6,252,480 parameters from which 6,246,464 are trainable and 6,016 are non-trainable.

Layer	Operations	Output shape	Output n_{side}	Parameters
I	Input map (normalised)	32,892×10	256	-
II	ConvBlock	8,396×32	128	1,696
III	ConvBlock	2,212×64	64	10,432
IV	ConvBlock	606 × 128	32	41,344
V	ConvBlock	164 × 256	16	164,608
VI	ConvBlock	50 × 256	8	328,448
VII	ConvBlock	14 × 256	4	328,448
VIII	ConvBlock	4 × 256	2	328,448
IX	ConvBlock	1 × 256	1	328,448
X	Append $\log_{10}(S_{\text{tot}})$	1 × 257		
XI	ReLU ∘ FC	1 × 2,048		528,384
XII	ReLU ∘ FC	1 × 512		1,049,088
XIII	Reshape ∘ FC	2 × 2 × 10		20,480
XIV	Softmax (means only)	2 × 2 × 10		-

Table 1: Architecture of our NN for estimating the spectra. The ConvBlocks are made of graph convolutions, batch normalisation, ReLU activation and maximum pooling. In layer X we append the logarithmic total counts of the map. The final layer uses a softmax activation for the mean estimates and also returns the estimated uncertainty for each mean value. The output shape corresponds to 2 (means/variances) $\times$ $T(= 2)$ components $\times$ $E(= 10)$ energy bins.

Layer	Operations	Output shape	Output n_{side}	Parameters
I	Input map (normalised)	32,892×10	256	-
II	ConvBlock	8,396×32	128	1,696
III	ConvBlock	2,212×64	64	1,0432
IV	ConvBlock	606 × 128	32	41,344
V	ConvBlock	164 × 256	16	164,608
VI	ConvBlock	50 × 256	8	328,448
VII	ConvBlock	14 × 256	4	328,448
VIII	ConvBlock	4 × 256	2	328,448
IX	ConvBlock	1 × 256	1	328,448
X	Append $\log_{10}(S_{\text{tot}})$	1 × 257		
XI	Append τ	1 × 258		
XII	ReLU ∘ FC	1 × 2,048		530,432
XIII	ReLU ∘ FC	1 × 512		1,049,088
XIV	FC	1 × 22		11,264
XV	Normalized softplus	1 × 22		-

Table 2: Architecture of our NN for estimating the SCDs. Note that in layer X we append the logarithmic total counts of the map, and in layer XI we append the quantile level τ. The final layer uses a normalized softplus activation. The output shape corresponds to a 1 (single PS component) $\times$ $M(= 22)$ count bins.