DEMNet: Evaluating Solar Differential Emission Measure Inversion with Deep Learning

V. Samuel Pérez-Díaz
Courant Institute
New York University
New York, NY 10011
v.perez.diaz@nyu.edu

Ruoyu Wang
Courant Institute
New York University
New York, NY 10011

Karin Dissauer
NorthWest Research Associates
Boulder, CO 80301
University of Graz, Institute of Physics
Graz, Austria 8010

KD Leka
NorthWest Research Associates
Boulder, CO 80301

David Fouhey
Courant Institute
New York University
New York, NY 10011

Abstract

The Differential Emission Measure (DEM) is a key diagnostic tool for understanding the solar atmosphere; commonly it connects images obtained in the extreme ultraviolet to the thermal structure of the solar corona through mathematical inversion. Traditional DEM inversion methods are based on per-pixel optimization and are very computationally expensive, thus prohibitive for large-scale analysis. We present and analyze **DEMNet**, a deep learning model that emulates a classical solver. DEMNet is far faster, provides uncertainties, and is smoother over multiple dimensions than traditional methods. We analyze DEMNet's predictions as well as several design decisions that are critical to its success.

1 Introduction

Understanding the thermal structure of the solar atmosphere is a key component for fundamental physical questions such as "what heats the solar corona?". NASA's Solar Dynamics Observatory (SDO; [12]) has facilitated pursuit of such questions with extreme ultraviolet (EUV) imaging at high resolution and high cadence from the Atmospheric Imaging Assembly (AIA; [2, 10]). From these observations, one can extract the Differential Emission Measure, which is the distribution of coronal plasma as a function of temperature. Extracting the DEM from observations is a well known inverse problem in solar physics. Classical solvers for DEM inversions include a plethora of techniques, from least squares to MCMC [6, 4, 14, 13, 1, 5, 8, 9].

Conference on Physics and AI at Stanford University (PAI 2026).

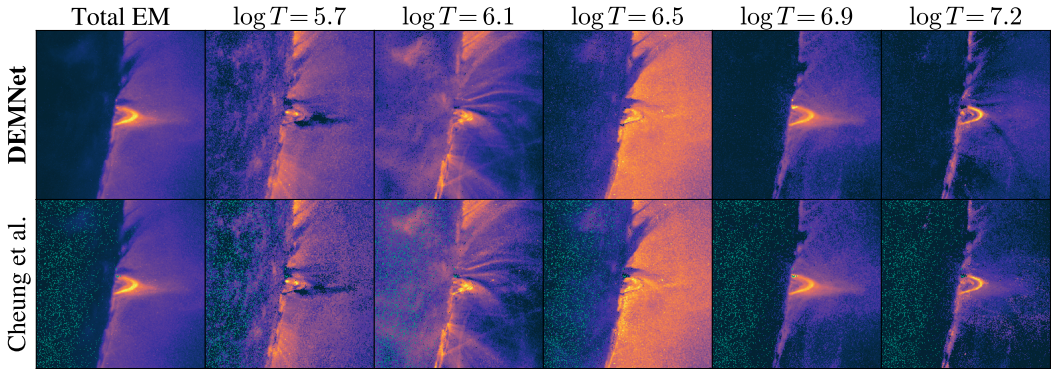

Figure 1: **Comparison of DEMNet with Basis Pursuit [3] method.** We show five log-temperature bins with a column-normalized colormap (Total EM is square-rooted: 2.1–29.9, and $\log T$ bins are cube-root transformed; global min/max after transform: 0.1–6.6). Failure pixels (where the solver fails to find a feasible solution) from the Basis Pursuit method are green. *DEMNet* captures the flaring, active regions and quiet regions while also removing high-frequency noise present in the original Basis Pursuit [3] method. Data: 2017 Sep. 10 15:27:59 UTC. Context: 2017 September 10 X8.2 flare during the impulsive phase, outside the training window (Jan. 2014 – Dec. 2015).

However, traditional methods remain computationally expensive (**≈1.5K px/sec per core**) inadequate for large-scale analysis of the native 4096^2-pixel 12 s AIA data. Thus, deep learning promises an alternative for emulating these techniques, with orders of magnitude faster inference (**≈2.85M px/sec on an L40 GPU**). In the machine learning setting, *DeepEM* [16] is the pioneer for DEM inversions with neural networks. It demonstrated, with a proof of concept, that a neural network can successfully emulate the per-pixel inversion.

In this work, we present **DEMNet**, a deep learning model that maps six AIA EUV images (at 94, 131, 171, 193, 211, 335 Å) to DEMs, trained on a large set of classical inversions. DEMNet builds on *DeepEM*, adding uncertainty quantification with a classification head, Fourier frequency encodings to capture sharp intensity variations, and a forward-model based finetuning stage. The method accurately emulates the training solver (Figure 1) and is temporally consistent (Figure 2). The network learns a consistent mapping of AIA to DEMs, which raises the question: *what and how did the network actually learn?* To answer this, we analyze the learned network.

The DEM Problem. Assuming optically thin plasma, the observed intensity in EUV (or X-ray) channel c is related to the coronal plasma thermal structure by $o_c = \int R_c(T)\,\mathrm{DEM}(T)\,dT$, where $R_c(T)$ is the temperature response function of channel c, and $\mathrm{DEM}(T)$ is the Differential Emission Measure. Discretizing the problem creates a linear system $\mathbf{o} = \mathbf{R}\mathbf{x}$, where $\mathbf{R} \in \mathbb{R}^{6\times 18}$ is the response matrix and $\mathbf{x} \in \mathbb{R}^{18}$ is the DEM vector over 18 temperature bins.

Classical methods. We take [3] (which we refer as Basis Pursuit) as our benchmark classical solver, which has been broadly adopted in the solar physics community. It proposes the inversion based on sparsity, finding the sparsest non-negative solution consistent with the observations to within t noise standard deviations:

$$\min_{\mathbf{x}} ||\mathbf{x}||_1 \quad \text{s.t.} \quad \mathbf{o} - t\boldsymbol{\sigma} \leq \mathbf{R}\mathbf{x} \leq \mathbf{o} + t\boldsymbol{\sigma}, \quad \mathbf{x} \geq \mathbf{0} \tag{1}$$

where $\boldsymbol{\sigma}$ is the estimated uncertainty, and $\mathbf{R}$ is the instrument's response function. We solve this linear program using HiGHS [7].

2 DEMNet

DEMNet is a pixel-wise neural network that maps the 6 EUV filters from AIA observations $\mathbf{o} \in \mathbb{R}^6$ independently per pixel to the DEM vector $\hat{\mathbf{x}} \in \mathbb{R}^{18}$. It is trained on both regression and classification objectives, so uncertainty quantification is available. The regression head produces $\hat{\mathbf{x}}$, direct predictions on the DEM space. The classification head predicts $\hat{\mathbf{c}}$, logits in a binned-space over the DEM values.

Architecture. First, AIA inputs are square-root transformed to make the extreme magnitude range of EUV-AIA intensities manageable for the neural network. However, multi-layer-perceptrons (MLPs) tend to learn low-frequency functions of their inputs [15], which prevents the network from learning the fine, sharp variations in the 6-dimensional EUV input space relevant to DEM inversion. Following [11], we lift each channel to a higher-dimensional Fourier representation: for an AIA channel o_i, we compute sin/cos pairs as $\phi(o_i) = \left[\sin(f_j \pi\, o_i), \cos(f_j \pi\, o_i)\right]_{j=0}^{11}$, with increasing frecuencies $f_j = 2^{(j-2)/2}$.

The encoded input is then passed to a four-layer MLP backbone (widths 128, 256, 256, 256; SiLU activations; $p = 0.2$ dropout). From the final 256-dimensional last layer of the backbone, we connect two heads: (i) a *regression head* (linear projection, no activation) predicting the DEM values directly, and (ii) a *classification head* predicting a distribution of probabilities over the DEM values, in order to get a confidence of the network's predictions; confidence intervals are obtained by computing quantiles over the cumulative distribution function of this predicted distribution. We leave the regression head without a final activation; the motivation for this choice is presented in Section 3.

Data. We generate training DEM targets by running a Basis Pursuit solver [3] on SDO/AIA full disk observations. We create the dataset by selecting samples from January 2014 to December 2015 (1,223 unique timestamps), using a chronological train/val/test split of 917/153/153 timestamps. To keep inversions computationally manageable, AIA images are spatially decimated by $2\times$ to match instrumental resolution (2048^2 pixels). We fix the temperature discretization to the AIA's response function bins, ($\log T$ from 5.5 to 7.2 in steps of 0.1).

To augment the training dataset and teach the network realistic uncertainty, we generated 5 independent realizations per timestamp: one with the original data, and four noisy runs. Each noisy realization is formed by perturbing the original with Gaussian noise. The Gaussian noise is scaled by half the per-pixel photon noise estimate ($\hat{\sigma}_c = \sqrt{o_c}$) [3]. From each full-disk realization, we sample 64 non-overlapping random 256×256 patches, producing approximately 293K training patches for training, and 49k for each val and test splits.

Training. We minimize the joint regression–classification loss

$$\mathcal{L} \;=\; \tfrac{1}{2}\,\mathrm{MSE}(\hat{\mathbf{x}}, \mathbf{x}) \;+\; \tfrac{1}{2}\,\mathrm{CE}(\hat{\mathbf{c}},\, \mathbf{b}(\mathbf{x})) + \lambda \|\mathbf{R}\hat{\mathbf{x}} - \mathbf{o}\|_2^2, \quad (2)$$

where $\hat{\mathbf{x}}$ is the regression prediction, $\mathbf{x}$ the target, $\hat{\mathbf{c}}$ the classification logits, and $\mathbf{b}(\mathbf{x})$ the target bin indices obtained by discretizing $\mathbf{x}$ into 128 sqrt-spaced bins over $[0, 2000]$; this interval was selected by empirically analyzing the distribution of DEM magnitudes over the training set. The last part is a resynthesis term, where $\mathbf{R} \in \mathbb{R}^{6\times 18}$ is the AIA instrument response matrix. This additional loss penalizes DEMs whose predicted AIA intensities diverge from the observations, which is complex to optimize from scratch because there are negative solutions that also minimize the loss. Initially, we set $\lambda = 0$ (no resynthesis). We use AdamW (lr $= 4 \times 10^{-5}$, weight decay 10^{-4}), batch size 64, train for 100 epochs, and select the best checkpoint by validation loss. From this outcome, we set $\lambda = 1$ and re-train for around 40 epochs, which provides a forward-model finetuning for improved resynthesis and more physical DEMs. Note that λ serves as a switch between the two training stages.

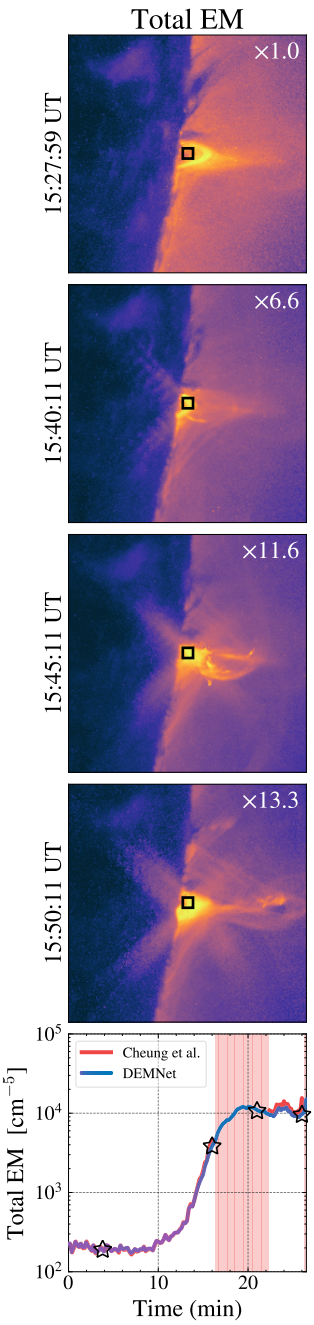

Figure 2: **Temporal snapshots** of DEMNet during the flaring event: 4 times capturing the impulsive phase. (frames are normalized independently and log-transformed, , global min/max: 1.7–9.5; Total EM multipliers in white, relative to the first frame's maximum value). Time series (bottom) from DEMNet and [3] for a pixel (black square); red region are Basis Pursuit failures; stars indicate the times above.

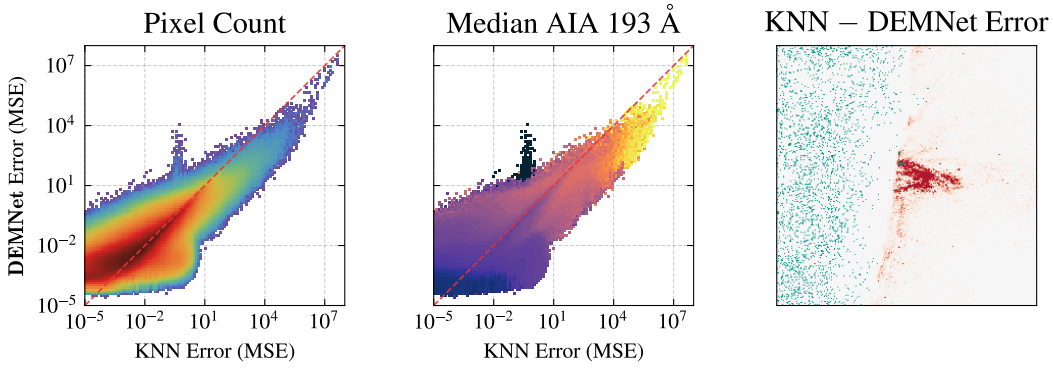

Figure 3: **KNN vs. DEMNet comparison.** (left) Joint density of KNN and DEMNet total-EM errors (1 – 2.3×10^6, log scale); the diagonal pattern for the bulk of the pixels indicates similar performance. (center) Same density, colored by AIA 193 intensity (0.1 – 9.2×10^3 DN s^{-1}, in power norm) showing that differences from the diagonal concentrate primarily at high intensity, brighter regions. (right) Per-pixel MSE difference map same time as in Fig. 1 (-80 – $+80$ MSE units): white is similar error; red indicates DEMNet outperforms KNN; blue indicates the reverse; failures from the Basis Pursuit method is green. The error concentrates in the pre-flare loop, where DEMNet performs better.

3 What and How Does the Network Learn?

Neural network models are often treated as black boxes, which makes physicists cautious of deploying them in scientific data pipelines where errors could carry scientific misinterpretations and mistakes. However, accuracy alone might not be enough to trust a method. We need to understand *why* it works and what structure in the data it exploits.

Performance. On the test set, DEMNet produces a DEM prediction for **every** input, while Basis Pursuit fails on ≈29% of all the pixels. Among the pixels where Basis Pursuit finds a solution, DEMNet achieves a mean relative resynthesis error of **12.2%** (bright regions: 9.8%, quiet regions: 12.7%), compared to **28.3%** (bright: 16.9%, quiet: 31%) for BP. Additionally, DEMNet's predicted distributions are fairly well calibrated: ≈**98%** empirical coverage at the 90% CI for bright pixels. DEMNet runs at **2.85M px/s**, two orders of magnitude faster than classical solvers. DEMNet (without forward-model fine-tuning) achieves a DEM mean relative error of **25.3%** and an MSE of ≈**0.27**, outperforming the 1% (3k samples) model (40.2%, ≈3.7), and a *DeepEM*-equivalent model trained on one full disk (146.6%, ≈4.4). Forward-model finetuning reduces resynthesis MSE from 161.3 to **42.4** at a modest DEM MSE cost (0.27 → 0.40).

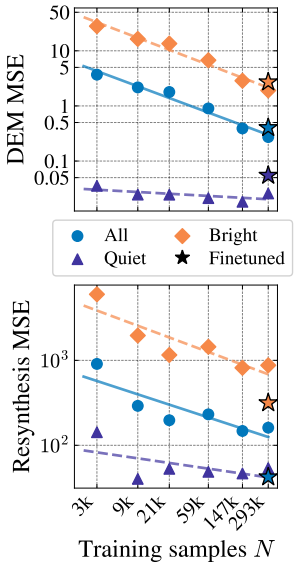

Figure 4: **Scaling laws with finetuning.** DEM MSE (top) and resynthesis MSE (bottom) as a function of training set size, stratified by bright and quiet pixels. The star marks the finetuned full model. Overall performance improves with more data. Finetuning improves resynthesis MSE at a modest cost to DEM MSE.

The Network as a Lookup Table. DEMNet predictions are closely aligned with a K-nearest-neighbor (KNN) on the training set, suggesting the network is acting as a smooth lookup table. We construct a KNN comparison by retrieving the $K = 5$ nearest neighbors in AIA input space over the full dataset using a KD-tree (plus a frequent pixel combination cache), and predict a distance-weighted combination of their DEM inversions. This model is impractically large, but helps analyze our learned MLP. Figure 3 compares the errors of the KNN model and DEMNet. For the most pixels, particularly in the middle region ($10^1 - 1.5 \times 10^2$) of the plot, the density of the errors follow the diagonal closely, suggesting a similar performance. At high intensities, where the most relevant phenomena occur (coronal loops, active regions, all structurally complex and rare), the two methods diverge, where DEMNet

takes advantage of its learned non-linear model. The example error map confirms this: the quiet disk is near-identical (white) but the pre-flare loop is red, where DEMNet outperforms KNN.

Encoding Physical Constraints. In the past *DeepEM* model [16], positivity was enforced with a ReLU at the output layer. We find that this degrades predictions due to the well-known dead neuron problem. Neurons that reach zero in the learning get stuck, and can never update. During training, AdamW updates collapse near-zero values to zero. Conversely, a model without ReLU recovers near-zero values accurately, almost 12% of predictions below 10^{-3}. These near-zero values matter, as they define the tails of the DEM distribution over temperatures. The network might not need the explicit constraint because positivity is already implicit in the training data: classical solvers enforce $\mathbf{x} \geq 0$ by construction. We thus drop the output ReLU and clip negative predictions post hoc at inference.

Scaling Laws. We perform experiments to understand the role of training data size in the overall performance of DEMNet. Figure 4 shows DEM MSE and resynthesis MSE across a log-scaled range of training sizes up to the full dataset. Each model is trained identically. We stratify results into bright pixels (any AIA channel in the top 5% by intensity, $\approx$14% of pixels) and quiet pixels (the rest). Performance broadly improves with more data. The full model achieves the best overall DEM MSE, though its quiet-region error is marginally worse than the model trained with 50% of the full dataset (147k samples). This suggests that the network increasingly focuses on bright pixels (that can have complex temperature structures) at scale. Bright pixels benefit the most from additional data; quiet-sun pixels, as the KNN analysis shows, can be retrieved by a simpler model. Finetuning with the resynthesis loss improves the resynthesis MSE.

Conclusion. We introduced DEMNet, a deep learning framework to produce DEM inversions from a set of AIA observations. The model often behaves similar to a smoothed lookup table over the training data, confirmed by its close alignment with a KNN baseline for the bulk of the pixels. The model stands out when predicting the rarer bright regions, which correspond to the most interesting physical phenomena of the corona. Scale helps broadly and particularly in these important cases. Overall, these findings suggest that the challenge of DEM inversion has a learnable structure, and a method like DEMNet could be incorporated in the data processing pipeline of solar missions in the future.

Acknowledgments and Disclosure of Funding

This work was supported by NASA Heliophysics Innovation and Technology Science (HITS) grant 80NSSC25K7766 (PI: Dr. KD Leka). This work uses data primarily from the Solar Dynamics Observatory (SDO), a NASA mission.

References

[1] Markus J Aschwanden and Paul Boerner. Solar corona loop studies with the atmospheric imaging assembly. i. cross-sectional temperature structure. *The Astrophysical Journal*, 732(2):81, 2011.

[2] Paul Boerner, Christopher Edwards, James Lemen, Adam Rausch, Carolus Schrijver, Richard Shine, Lawrence Shing, Robert Stern, Theodore Tarbell, Alan Title, et al. Initial calibration of the atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo). In *The Solar Dynamics Observatory*, pages 41–66. Springer, 2012.

[3] Mark C. M. Cheung, P. Boerner, C. J. Schrijver, P. Testa, F. Chen, H. Peter, and A. Malanushenko. THERMAL DIAGNOSTICS WITH THE ATMOSPHERIC IMAGING ASSEMBLY ON BOARD THE SOLAR DYNAMICS OBSERVATORY: A VALIDATED METHOD FOR DIFFERENTIAL EMISSION MEASURE INVERSIONS. *The Astrophysical Journal*, 807(2):143, July 2015. Publisher: The American Astronomical Society.

[4] Bart De Pontieu, Juan Martínez-Sykora, Paola Testa, Amy R Winebarger, Adrian Daw, Viggo Hansteen, Mark CM Cheung, and Patrick Antolin. The multi-slit approach to coronal spectroscopy with the multi-slit solar explorer (muse). *The Astrophysical Journal*, 888(1):3, 2020.

[5] C Guennou, F Auchère, E Soubrié, K Bocchialini, S Parenti, and N Barbey. On the accuracy of the differential emission measure diagnostics of solar plasmas. application to sdo/aia. i. isothermal plasmas. *The Astrophysical Journal Supplement Series*, 203(2):25, 2012.

[6] Iain G Hannah and Eduard P Kontar. Differential emission measures from the regularized inversion of hinode and sdo data. *Astronomy & Astrophysics*, 539:A146, 2012.

[7] Q Huangfu and JA Julian Hall. Parallelizing the dual revised simplex method. *Mathematical Programming Computation*, 10(1):119–142, 2018.

[8] Carole Jordan. A discussion on the physics of the solar atmosphere-the structure and energy balance of solar active regions. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 281(1304):391–404, 1976.

[9] Vinay Kashyap and Jeremy J Drake. Markov-chain monte carlo reconstruction of emission measure distributions: application to solar extreme-ultraviolet spectra. *The Astrophysical Journal*, 503(1):450, 1998.

[10] James R Lemen, Alan M Title, David J Akin, Paul F Boerner, Catherine Chou, Jerry F Drake, Dexter W Duncan, Christopher G Edwards, Frank M Friedlaender, Gary F Heyman, et al. The atmospheric imaging assembly (aia) on the solar dynamics observatory (sdo). *Solar Physics*, 275(1):17–40, 2012.

[11] 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.

[12] W Dean Pesnell, B J Thompson, and PC Chamberlin. The solar dynamics observatory (sdo). In *The solar dynamics observatory*, pages 3–15. Springer, 2012.

[13] Joseph Plowman and Amir Caspi. A fast, simple, robust algorithm for coronal temperature reconstruction. *The Astrophysical Journal*, 905(1):17, 2020.

[14] Joseph Plowman, Charles Kankelborg, and Petrus Martens. Fast differential emission measure inversion of solar coronal data. *The Astrophysical Journal*, 771(1):2, 2013.

[15] Nasim Rahaman, Aristide Baratin, Devansh Arpit, Felix Draxler, Min Lin, Fred Hamprecht, Yoshua Bengio, and Aaron Courville. On the spectral bias of neural networks. In *International conference on machine learning*, pages 5301–5310. PMLR, 2019.

[16] Paul J. Wright, Mark C. M. Cheung, Rajat Thomas, Richard Galvez, Alexandre Szenicer, Meng Jin, Andrés Muñoz-Jaramillo, and David Fouhey. Deepem: Demonstrating a deep learning approach to dem inversion, March 2019.