

A machine learning approach for computing solar flare locations in X-rays on-board Solar Orbiter/STIX

Paolo Massa^{*1}, Simon Felix^{1,2}, László István Etesi^{1,2}, Ewan C. M. Dickson³, Hualin Xiao¹, Francesco P. Ramunno^{1,4}, Merve Selcuk-Simsek¹, Brandon Panos¹, André Csillaghy¹, Säm Krucker^{1,5}

¹*Institute for Data Science, University of Applied Sciences and Arts Northwestern Switzerland, Bahnhofstrasse 6, 5210 Windisch, Switzerland*

²*Ateleris GmbH, Neumarkt 1, 5200 Brugg, Switzerland*

³*Institute of Physics, University of Graz, A-8010 Graz, Austria*

⁴*Department of Computer Science, University of Geneva, 1211 Geneva, Switzerland*

⁵*Space Sciences Laboratory, University of California, 7 Gauss Way, 94720 Berkeley, USA*

The Spectrometer/Telescope for Imaging X-rays (STIX) on-board the ESA Solar Orbiter mission retrieves the coordinates of solar flare locations by means of a specific sub-collimator, named the Coarse Flare Locator (CFL). When a solar flare occurs on the Sun, the emitted X-ray radiation casts the shadow of a peculiar “H-shaped” tungsten grid over the CFL X-ray detector. From measurements of the areas of the detector that are illuminated by the X-ray radiation, it is possible to retrieve the (x, y) coordinates of the flare location on the solar disk.

In this paper, we train a neural network on a dataset of real CFL observations to estimate the coordinates of solar flare locations. Further, we apply a post-training quantization technique specifically tailored to the adopted model architecture. This technique allows all computations to be in integer arithmetic at inference time, making the model compatible with the STIX computational requirements. We show that our model outperforms the currently adopted algorithm for estimating the flare locations from CFL data regarding prediction accuracy while requiring fewer parameters. We finally discuss possible future applications of the proposed model on-board STIX.

1 Introduction

The Spectrometer/Telescope for Imaging X-rays (STIX; [1]) is the telescope of the ESA Solar Orbiter mission [2] dedicated to the observation of solar flares. STIX measures the flaring X-ray radiation between 4 and 150 keV by means of Cadmium-Telluride (CdTe) detectors [3], which count the number of incident X-ray photons. Specifically, 30 STIX detectors are mounted behind pairs of tungsten grids, which modulate the incident X-ray radiation and encode infor-

mation on the morphology and location of the X-ray source(s). The units consisting of a detector and a grid pair, named *sub-collimators*, provide indirect information on the flaring X-ray source(s) in terms of a limited number of corresponding 2D Fourier components [4]. Therefore, once the STIX data are downloaded to the ground, it is possible to reconstruct the image of the X-ray source(s) for a specific flaring event by solving a Fourier inversion problem [5].

STIX contains an additional sub-collimator, named the Coarse Flare Locator (CFL), which provides information on the (x, y) coordinates of the flare location on the solar disk with an accuracy of a couple of arcminutes [1]. The CFL sub-collimator consists of a front grid with a peculiar “H-shaped” design, while the rear grid is open. The location of the X-ray shadow cast by the front grid on the detector surface varies depending on the flare’s location on the Sun. Therefore, the (x, y) coordinates of the flare location are uniquely determined from measurements of the illuminated areas of the CFL detector (see Figure 1).

The estimates of the flare location provided by the STIX CFL are extremely important for several reasons. First, they can be sent to other Solar Orbiter instruments for coordinated flare observations. Second, the flare locations estimated on-board are sent to the ground as quick-look data products with reduced latency (within less than a day) [1, 7]. Since, due to telemetry constraints, the STIX science data are available only after several weeks, the flare locations saved in the quick-look data products are important for preliminary analysis and for performing appropriate data requests from the ground [7]. Finally, it is the first time a device like the STIX CFL operates within a space mission. Therefore, the flare locations

^{*}Corresponding author. E-Mail: paolo.massa@fhnw.ch

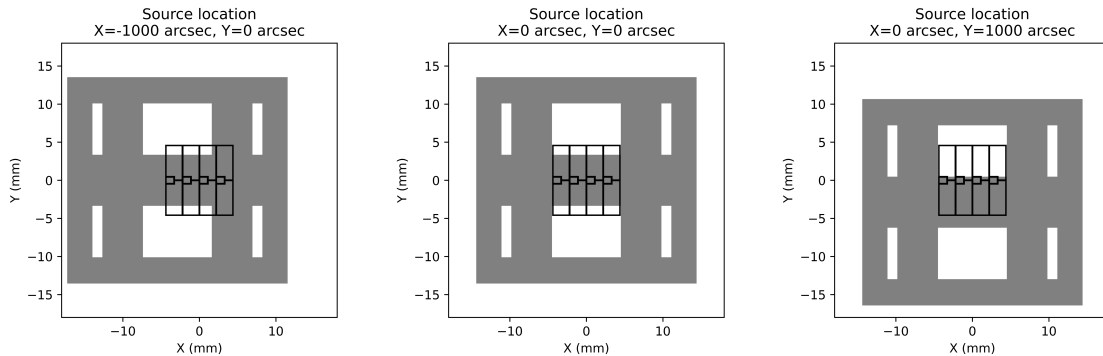


Figure 1: Simulated shadow of the CFL front grid that is cast on the detector surface by three point-like flare sources located in as many different locations on the solar disk. The detector edges are plotted in black, while the grid shadow is overlaid in gray. The simulated flare locations are $(x, y) = (-1000'', 0'')$, $(x, y) = (0'', 0'')$, and $(x, y) = (0'', 1000'')$ from left to right, respectively. The simulation is performed by means of the `pystixsim` library [6].

computed on-board STIX are extremely valuable to demonstrate that this technique works as expected in a deep space environment and, therefore, a similar system could be used in future space weather missions.

In this work, we develop a machine learning approach to predict the coordinates of the flare location from CFL data. We construct a dataset of real CFL measurements and label them with the corresponding location obtained from images of flares reconstructed from STIX data. We train a multi-layer perceptron (MLP) and apply a post-training model quantization technique specifically developed for the adopted MLP architecture. In this way, at inference time, our model applies only integer arithmetic and, therefore, satisfies the computational requirements to be run on-board STIX. We compare the performance of our model with that of the currently used CFL algorithm, showing that the MLP outperforms the CFL algorithm in terms of prediction accuracy.

The paper is organized as follows. Section 2 contains the details of the dataset we constructed, the MLP architecture and training, the CFL algorithm, and the comparison between MLP and the CFL algorithm performances on the test set. Finally, Section 3 contains our discussions on the results and future work.

2 Results

We define an MLP with three hidden layers, each consisting of 100 neurons. The activation function of each neuron is a Rectified Linear Unit (ReLU [8]). Our MLP takes a 9-dimensional input. The first eight entries are the number of counts registered in the large pixels¹

¹A STIX detector is partitioned into 12 units, called *pixels*, which count the number of incident X-ray photons independently (see

of the CFL detector. The ninth entry is the sum of the counts registered by a separate imaging detector (labeled as *7b*; see Table 2 in [1]), which has been chosen since it shares similar calibration properties as the CFL. This ninth value provides an independent estimate of the total flux (i.e., total number of counts per unit of area) that should be recorded in the CFL pixels. For example, let's consider the case when one of the two rows of the CFL pixels is fully covered by the mask shadow, as shown in the right panel in Figure 1. By comparing the number of counts recorded by the illuminated pixels of the CFL with the total number of counts recorded by detector *7b*, the MLP can determine the amount of illuminated area of the CFL pixels and, therefore, estimate the flare location correctly. The MLP's output layer consists of two neurons, which represent the x and y coordinates of the flare location on the solar disk. No activation function is used in the output layer.

2.1 Dataset

We consider 2150 solar flares recorded by STIX between 2021 March 19 and 2023 August 28. In this proof of concept study, we do not consider flaring events with locations outside the full-sensitivity FOV of STIX (see Section 5 in [4]) since they are quite rare (fewer than 10% of the total events) and make the neural network training more challenging. For each event we consider the raw counts measured by the STIX detector pixels in 32 energy channels and in several consecutive time bins, whose number varies from event to event. We sum the counts recorded between 5 and 9 keV, and group the data in bins larger or equal to 8 s. We extract the 9-dimensional vector of each bin to be

Figure 1). In particular, there are 8 large pixels (top and bottom row) and 4 small ones (central row).

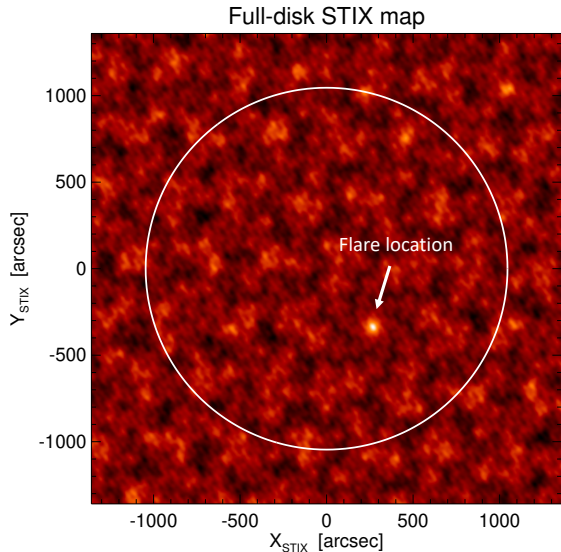


Figure 2: Full-disk image reconstructed from STIX data. The solar limb is plotted as a white circle, and the flare location $(x, y) \simeq (250'', -350'')$ is indicated with a white arrow.

used as input to the MLP. At the same time, the flare’s location, which will be used to train the MLP, is determined for each time bin from measurements provided by the STIX imaging detectors. We recall that STIX does not provide us directly with images of the flaring X-ray radiation. Rather, STIX measures a limited number of 2D Fourier components of the signal; therefore, an image reconstruction process is needed to obtain the flare image from STIX measurements. As shown in Figure 2, we reconstruct a full-disk map from STIX data using a direct Fourier inversion method, named Back Projection [4, 5]. Then, we derive the coordinates of the flare location by identifying the position of the brightest area in the map, i.e., the location at which the X-ray emission is most intense. Note that the STIX image in Figure 2 shows a background pattern around the flare location which does not represent a real flaring emission. This pattern is a ringing effect due to the limited number of Fourier components available for image reconstruction and to the adopted image reconstruction method.

The resulting dataset is divided into training, validation, and test sets as follows. Samples constructed from files recorded between 2021 March 19 and 2022 November 15 (18K data points) are used for training, samples constructed from files recorded between 2022 November 15 and 2023 April 4 (13K data points) are used for validation, and the remaining ones (11K data points) are used for test. We did not use a random split of the dataset to avoid data leakage, where data points from the same flare would end up in both training and test sets. This would result in metrics overestimating the MLP performance.

2.2 MLP implementation and training

We implemented our MLP using the Python Tensorflow library [9]. Two normalization techniques appropriate for the requirements of the input and expected output data are applied prior to training the network. Input arrays are rescaled by their maximum value. Normalizing the input values is permissible, as the intensity of the flare is not relevant when determining its location. Similarly, the labels (i.e., the x and y coordinates of the flare locations) are divided by a value f larger than the maximum dimension of the STIX Field-of-View (FOV). In this case, we set $f = 4000''$. MLP training has been performed with the Adam optimizer [10] and using a Mean Squared Error (MSE) loss function.

2.3 Post-training quantization

Since the STIX data processing unit does not support floating-point arithmetic, we apply a post-training quantization of inputs, weights, and outputs to ensure that all the network computations are performed in 16-bit fixed-point integer arithmetic at inference time. As our specific hardware requirements are not satisfied by the available quantization algorithms, we use the custom approach described in the following. We determined the range of all intermediate results by finding extremal values in a Mixed Integer Linear Programming representation of the network. A mathematical solver [11] proved that all numbers are strictly less than 2^5 , irrespective of the network’s input. Thus, we use 5 bits to represent the integral part, and the remaining 11 bits to represent the fractional part of each number. This way, our proposed model meets the computational requirements to be potentially run on-board the STIX instrument.

To convert the input values to fixed-point form, we use

$$\mathbf{x}_{\text{scaled}} = (\mathbf{x}_{\text{input}} \cdot 2^{11}) // m, \quad (1)$$

where m is the maximum of the 9 input values, and $//$ denotes integer division (applied component-wise where needed). At the end of these operations, the input values are integers between 0 and 2^{11} .

During the forward pass of the network, 16-bit input values are multiplied by 16-bit weights. The results are integers between 0 and 2^{32} , which must be converted back to the 16-bit fixed-point representation by right shifting 11 bits. Finally, we add the bias term and apply the ReLU activation function. In summary, the output vector of a network layer is

$$\mathbf{o} = \text{ReLU}((\mathbf{W} \cdot \mathbf{x} + 2^{10}\vec{\mathbf{1}}) \gg 11 + \mathbf{b}), \quad (2)$$

where \mathbf{W} is the weight matrix, $\vec{\mathbf{1}}$ an all-one vector, \mathbf{b} is the bias array, and \gg denotes the component-

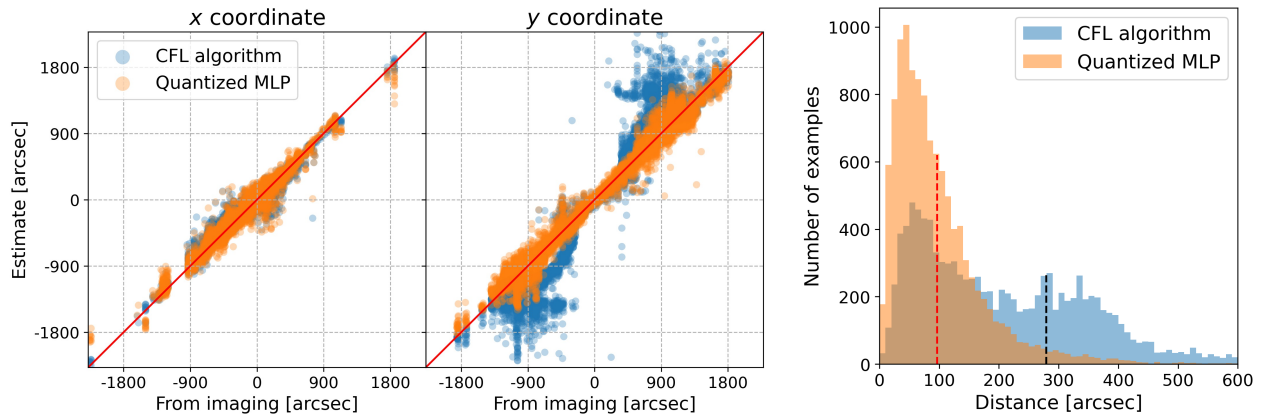


Figure 3: Comparison of the quantized MLP and CFL algorithm performances on the test set examples. Left and middle panels: scatter plots of the x (left) and y (middle) coordinates of the flare locations estimated by the MLP (orange) and by the CFL algorithm (blue) as a function of the coordinates obtained from the reconstructed STIX images. The identity line representing exact estimates is overlaid in red as a reference. Right panel: histograms of the distances between the flare locations estimated by the MLP (orange) and the CFL algorithm (blue) and the locations derived from STIX reconstructed images. The mean values of the quantized MLP distances and of the CFL algorithm ones are plotted with a dashed red and black line, respectively.

wise right shift operator. We note that the same operations are performed in the network’s output layer, sans ReLU activation. Finally, since inputs, weights, and biases of the network are scaled by 2^{11} , and since ReLU is a positively homogeneous function of degree one (i.e., $\text{ReLU}(\alpha \cdot \mathbf{x}) = \alpha \text{ReLU}(\mathbf{x})$ for all $\alpha > 0$), the output of the network is also scaled by 2^{11} . Thus the coordinates \mathbf{c} computed by the network must be scaled back, also taking into account the fact that the output of the network during training has been divided by a value f (see Section 2.2). The array containing the estimated flare location coordinates is then

$$\mathbf{c}_{\text{final}} = (\mathbf{c} \cdot f + 2^{10} \vec{\mathbf{1}}) \gg 11, \quad (3)$$

where \mathbf{c} is the two-dimensional output of the network.

2.4 CFL algorithm

During flaring events, the X-ray radiation is blocked by the tungsten “H-shaped” front grid of the CFL sub-collimator, and a shadow is cast on the corresponding detector (see Figure 1). The current on-board implementation of the CFL algorithm relies on a lookup table containing the values of the area of the eight large CFL pixels that would be illuminated by a point source located at each of 65×65 points on the solar disk, uniformly separated by a distance of $2'$. The CFL algorithm normalizes the eight large pixel measurements by an estimate of the total flux (i.e., total number of counts per unit of area) derived from the imaging sub-collimators, thus obtaining the fraction of illuminated area for each pixel. Then, the method

compares the observed illuminated areas with the entries of the lookup table, and returns the location corresponding to the maximum correlation.

In this proof of concept study, we utilize an on-ground Interactive Data Language (IDL) implementation of the CFL algorithm, and we apply it to our dataset (see Section 2.1). Differently from the on-board implementation, this version applies a final interpolation step between the coordinates corresponding to the largest correlation with the observations². Therefore, the on-ground implementation of the CFL algorithm can improve the $2'$ resolution of the lookup table.

2.5 Results on test set

Figure 3 compares the flare location coordinates obtained by means of the quantized neural network with those obtained by means of the CFL algorithm. The left and middle panels show the scatter plots of the x and y coordinates of the flare locations obtained with the two methods as a function of the coordinates retrieved from the reconstructed STIX images. The locations derived from imaging are the reference values; therefore, exact estimates would result in scatter plots along the identity diagonal, which is plotted in red as a reference. The MLP predictions (orange) of the x coordinate show an agreement with the locations derived from imaging which is comparable with that of the CFL algorithm (blue). However, regarding the y coordinate (middle panel), the MLP is substantially

²The interpolation step is not performed in the on-board implementation of the CFL algorithm to reduce its computational cost.

more accurate than the CFL algorithm. This is particularly true for samples corresponding to flare locations $400'' < |y| < 1200''$. In these locations one of the two rows of the CFL pixels is completely covered by the shadow of the “H-shaped” front grid (see the right panel of Figure 1). To determine the amount of illuminated area in the large pixels, the CFL algorithm normalizes the measurements provided by the CFL pixels using an independent estimate of the total flux derived from the imaging sub-collimators (see Section 2.4). It is likely that the cross-calibration between CFL measurements and those provided by the imaging sub-collimators is not yet optimal, which would explain the bias in the algorithm’s estimate of the flare location.

In the right panel of Figure 3 we plot the histogram of the distances between the flare locations estimated by the methods and those derived from the reconstructed images. The mean distance of the CFL algorithm predictions is $280''$, while the mean distance of the MLP predictions is $97''$, a marked improvement in accuracy.

3 Discussion

In this work, we develop a machine learning method for estimating the (x, y) coordinates of flare locations on the solar disk from STIX CFL data. Our model, coupled with a post-training quantization technique, is more accurate than the algorithm currently adopted on-board STIX. Further, the CFL algorithm is based on a lookup table, which requires $65 \times 65 \times 8 \simeq 34\text{K}$ parameters to be stored in memory, while our proposed MLP utilizes only 21K parameters (i.e., the total number of weights and biases).

The post-training quantization technique specifically developed for the adopted model architecture allows the MLP to use integer arithmetic at inference time, as required by the STIX data processing unit. The quantization technique does not affect the performance of the model. Indeed, the distance between the locations estimated by the quantized MLP and those estimated by the non-quantized model for the test set examples is, on average, $9''$, with a maximum distance of $24''$.

In this work we have considered an MLP with 3 layers and 100 neurons per layer. This choice of the model architecture was done based on some trial and error tests. However, it is possible that even lighter models can be used for addressing the same task with similar accuracy. Before considering using the trained neural network on-board STIX, we will more systematically search for the model architecture which provides the best trade-off between prediction accu-

racy of and low number of parameters by means of well-established techniques [12, 13]. In future work, we will also include flaring events outside the full-sensitivity STIX FOV in the training set.

Although more complex network architectures (e.g., convolutional neural networks) could be considered for this work, we decided to adopt an MLP model since the number of data points (9 elements) and number of output values (2 elements) is very small. Given that the average error of the adopted MLP on the test set examples ($97''$) is already within the design accuracy of CFL sub-collimator ($\sim 2'$), we do not expect more complex architectures to have substantially improved performances compared to the MLP. Further, the MLP model was chosen as it could be easily implemented within the STIX flight software. More complex model architectures would require a greater effort for the implementation on-board STIX, since the flight software can not leverage on standard Python libraries that are usually adopted for defining the models (e.g., Tensorflow [9]). Finally, we demonstrated that the MLP can be made compatible with the STIX hardware requirements by means of an appropriate post-training quantization technique. The same quantization method can not be applied to every model. Therefore, other models could require the implementation of tailored (and potentially complex) quantization methods.

The dataset and the code implemented for this study can be found at https://github.com/pao1omassa/STX_CFL_NN.git.

References

1. Krucker, S. *et al.* The Spectrometer/Telescope for Imaging X-rays (STIX). *A&A* **642**, A15 (2020).
2. Müller, D. *et al.* The Solar Orbiter mission. Science overview. *A&A* **642**, A1 (2020).
3. Meuris, A. *et al.* Caliste-SO X-ray micro-camera for the STIX instrument on-board Solar Orbiter space mission. *Nuclear Instruments and Methods in Physics Research A* **695**, 288–292 (Dec. 2012).
4. Massa, P. *et al.* The STIX Imaging Concept. *Sol. Phys.* **298**, 114 (Oct. 2023).
5. Piana, M., Emslie, A., Massone, A. M. & Dennis, B. R. *Hard X-ray Imaging of Solar Flares* (Springer-Verlag, Berlin, 2022).
6. Xiao, H. *pystixsim: an imaging simulator for the X-ray Imager/Telescope onboard Solar Orbiter STIX version v1.0*. Apr. 2024. <https://doi.org/10.5281/zenodo.10973277>.
7. Xiao, H. *et al.* The data center for the Spectrometer and Telescope for Imaging X-rays (STIX) on board Solar Orbiter. *A&A* **673**, A142. arXiv: 2302.00497 [astro-ph.SR] (May 2023).
8. Bishop, C. M. *Pattern recognition and machine learning* (Springer, 2006).
9. Martín Abadi *et al.* *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems* 2015. <https://www.tensorflow.org/>.

10. Kingma, D. P. & Ba, J. *Adam: A Method for Stochastic Optimization* 2017. arXiv: 1412.6980 [cs.LG]. <https://arxiv.org/abs/1412.6980>.
11. Gurobi Optimization, LLC. *Gurobi Optimizer Reference Manual* 2023. <https://www.gurobi.com>.
12. Hinton, G., Vinyals, O. & Dean, J. *Distilling the Knowledge in a Neural Network* 2015. arXiv: 1503.02531 [stat.ML]. <https://arxiv.org/abs/1503.02531>.
13. White, C. *et al.* *Neural Architecture Search: Insights from 1000 Papers* 2023. arXiv: 2301.08727 [cs.LG]. <https://arxiv.org/abs/2301.08727>.

Acknowledgments

Solar Orbiter is a space mission of international collaboration between ESA and NASA, operated by ESA. The STIX instrument is an international collaboration between Switzerland, Poland, France, Czech Republic, Germany, Austria, Ireland, and Italy.

PM, FPR, MSS, and AC are supported by the Swiss National Science Foundation in the framework of the project Robust Density Models for High Energy Physics and Solar Physics (rodem.ch), CRSII5_193716. HX is supported by the PRODEX experiment agreement No. 4000140488 for STIX. SK is partially supported by the Swiss National Science Foundation Grant 200021L_189180.