



A Machine Learning Enhanced Approach for Automated Sunquake Detection in Acoustic Emission Maps

Vanessa Mercea¹ · Alin Razvan Paraschiv² · Daniela Adriana Lacatus² · Anca Marginean¹ · Diana Besliu-Ionescu³

Received: 29 August 2022 / Accepted: 7 November 2022 / Published online: 9 January 2023
© The Author(s) 2022

Abstract

Sunquakes are seismic emissions visible on the solar surface, associated with some solar flares. Although discovered in 1998, they have only recently become a more commonly detected phenomenon. Despite the availability of several manual detection guidelines, to our knowledge, the astrophysical data produced for sunquakes is new to the field of machine learning. Detecting sunquakes is a daunting task for human operators, and this work aims to ease and, if possible, to improve their detection. Thus, we introduce a dataset constructed from acoustic egression-power maps of solar active regions obtained for Solar Cycles 23 and 24 using the holography method. We then present a pedagogical approach to the application of machine-learning representation methods for sunquake detection using autoencoders, contrastive learning, object detection and recurrent techniques, which we enhance by introducing several custom, domain-specific data augmentation transformations. We address the main challenges of the automated sunquake-detection task, namely the very high noise patterns in and outside the active region shadow and the extreme class imbalance given by the limited number of frames that present sunquake signatures. With our trained models, we find temporal and spatial locations of peculiar acoustic emission and qualitatively associate them to eruptive and high energy emission. While noting that these models are still in a prototype stage, and there is much room for improvement in metrics and bias levels, we hypothesize that their agreement on example use cases has the potential to enable detection of weak solar acoustic manifestations.

Keywords Active regions · Flares · Waves · Helioseismology

✉ V. Mercea
mercea.vanessa@gmail.com

¹ Technical University of Cluj-Napoca, Cluj-Napoca, 400114, Romania

² High Altitude Observatory, National Center for Atmospheric Research, PO box 3000, Boulder, CO 80307-3000, USA

³ Astronomical Institute of the Romanian Academy, Str. Cutitul de Argint, Nr. 5, Bucharest, 040557, Romania

1. Introduction

Local helioseismology has been the primary tool used to study occasional emission of seismic transients, more commonly known as sunquakes, coming from the solar surface or from sources submerged in the solar interior that are sometimes released by solar flares. Kosovichev and Zharkova (1998) first discovered sunquakes as expanding rings in the Dopplergram data using a factor of four image-enhancement technique. The technique revealed the expanding rings of an almost circular-shaped surface ripple after the 9 July 1996 X2.6 solar flare. Donea, Braun, and Lindsey (1999) obtained egression-power maps using the method developed by Lindsey and Braun (2000) for this particular sunquake event. Since then, sunquakes have become a more commonly detected phenomenon (Donea, 2011; Kosovichev, 2011; Besliu-Ionescu, Donea, and Cally, 2017; Sharykin and Kosovichev, 2020). Typically, sunquakes are associated with intense reconnection events in the solar atmosphere resulting in strong solar flares, of X- or M- spectral class, although Sharykin, Kosovichev, and Zimovets (2015) found signatures related to a weaker C.7 class event.

There are several methods to detect a seismic emission. From a chronological point of view these are: time–distance analysis (Kosovichev and Zharkova, 1998), seismic holography (Lindsey and Braun, 2000), and seismic ripples detection (movie method) (Kosovichev and Zharkova, 1998; Kosovichev, 2011; Sharykin and Kosovichev, 2020). Each method has its advantages and disadvantages; some of them show the source, some the wave propagation, and usually not all methods show signatures for the same flares (Sharykin and Kosovichev, 2020).

Sunquake selection criteria require a continuous detection of acoustic signal of at least a couple of minutes and a reasonable source signal-to-noise ratio. These criteria are further extended in Section 2.1. More elaborate and precise selection criteria have been proposed in the literature (e.g. Chen and Zhao, 2021), where a sunquake selection rule can be the conjuncture between impulsive flaring events and downward background oscillatory velocity, occurring at the same location. In practice, an intricate analysis needs to be done on a flare-by-flare basis on the entire dataset. We thus could not utilize such criteria in this work.

Despite the variety of available methods for sunquake detection, an attempt to automate the detection process in order to reduce human effort has yet to emerge. To facilitate this, we construct a machine learning (ML) ready dataset covering SC23 and SC24. To our knowledge, these specialized astrophysical data products were not previously processed via ML techniques. This work takes the first steps in this direction and describes the components of several ML-enhanced detection models, as a result of an extensive array of experiments centered on representation learning techniques. See Bengio, Courville, and Vincent (2013) for a review of these techniques. We discuss findings from our experiments and identify challenging aspects.

This work is structured as follows: Section 2 describes the methods used for data ingestion, pre-processing, holography analysis, and preparation for ML algorithm ingestion. The ML methodology is described in detail in Section 3, in preparation for Section 4, where particularities and limitations of two detection models are presented along with an analysis of current results on both known and tentative newly detected sunquake events. Lastly, Section 5 summarizes the discussion of models limitations and main results, and suggests avenues of interest for future development. Furthermore, in Appendices A–D we expand on additional methods and more standard methodological approaches that have been explored while pursuing the task at hand that have not been included in the final proposed solution.

2. Data Ingestion and Pre-Processing

2.1. The Helioseismic Holography Method

In this work, we have used the helioseismic-holography analysis of Lindsey and Braun (2000) to process raw data and identify the seismic emission during solar flares from a selected list of sunquake events. The method is applied to photospheric Dopplergram maps such as those provided by the *Michelson Doppler Imager* (MDI: Scherrer et al., 1995) onboard the *Solar and Heliospheric Observatory* (SOHO: Domingo, Fleck, and Poland, 1995) and the *Helioseismic and Magnetic Imager* (HMI: Schou et al., 2012) onboard the *Solar Dynamics Observatory* (SDO: Pesnell, Thompson, and Chamberlin, 2012).

In general terms, the helioseismic-holography technique is used to image acoustic sources on and beneath the Sun's photosphere. It reconstructs phase-coherent p -mode acoustic waves that are observed at the photosphere into the solar interior to render stigmatic images of the subsurface sources that have perturbed this surface.

The solar interior refracts downward-going waves back toward the Sun's surface due to the temperature gradient below the photosphere leading to increasing sound speed towards the interior. Helioseismic holography images a selected area, in a way that is "broadly analogous to how the eye treats electromagnetic radiation at the surface of the cornea, wave-mechanically refocusing radiation from submerged sources to render stigmatic images" (Lindsey and Braun, 2000). In order to obtain these images, holography uses a pupil defined as an annulus with radius 15–45 Mm, to image the focus, situated a considerable distance from the pupil, and computes the "ingression" [H_-] and "egression" [H_+]. The ingression and the egression are obtained from the wave-field at the surface [Ψ] through theoretical Green's functions (Lindsey and Braun, 2000).

The egression power $P(\mathbf{r}, t) = |H_+(\mathbf{r}, t)|^2$ is extensively used in detecting or studying acoustic sources and absorbers (Ionescu, 2010). This equation is used to calculate the egression power for each pixel in the image. Therefore, using this technique, we can create maps of egression power around ARs in order to detect the seismic sources.

Following Ionescu (2010), we employ a temporal selection criteria in which a eight-minute duration of the egression power signature is required for a positive sunquake identification due to an artifact of the truncation of the helioseismic spectrum by a 2 mHz wide pass-band. The egression power signatures that result are temporally smeared to a minimum effective duration of order

$$\Delta t = \frac{1}{\Delta \nu} = \frac{1}{2 \text{ mHz}} \approx 500 \text{ seconds.} \quad (1)$$

Other selection criteria using this technique concern the position of the source relative to the active region (AR), its intensity in terms of a 3σ pixel enhancement, and a clear increase–decrease type signal, or even more complex criteria as suggested by Chen and Zhao (2021). For this ML-oriented work, where absolute intensity measurements tend to lose significance, the most important applicable constraint is the temporal-selection criteria.

The band centered at 6 mHz was chosen for this work because it showed a much higher signal-to-noise ratio compared to other bands (Donea, Braun, and Lindsey, 1999), making it easier to detect the source (egression power in absolute terms ≈ 4.1 times that of the 6 mHz mean quiet Sun, Donea, Braun, and Lindsey, 1999).

2.2. Sunquake Identification and Dataset Creation

For this work, we produced a sunquake dataset spanning two solar cycles, SC23 and SC24, using the holography method described above. To identify potential sunquake candidates, we have used the curated lists of Besliu-Ionescu et al. (2012) for SC23, totaling 15 MDI observed events, along with the positive holography identifications from SC24 of Sharykin and Kosovichev (2020) totaling 80 events. In the case of SC24, multiple sunquakes have been identified in the same AR, but at different times during its disk crossing. For the complete list and parameters of sunquakes that were positively identified and marked “+” by Sharykin and Kosovichev (2020) see their Table 1. To create our dataset, we used temporal cubes of 5–7 mHz holography-acoustic-egression maps binned to a resolution of 256×256 pixels that lead to approximately 1.5” in physical resolution leading to $\approx 200 \times 200$ acoustic maps. This binning is required in order to get adequate signal-to-noise ratios necessary for a positive sunquake detection. For all positive events in both SC23 and SC24, we have created acoustic datacubes spanning three-hour intervals, as required by the holography methodology, with observations starting approximately 1–1.5 hours before the sunquake-inducing flare, and following for an additional 1.5–2 hours. The data cadence of MDI (SC23) is 60 seconds, while the data cadence of HMI (SC24) is 45 seconds. The before and after temporal limits are flexible in order to not bias an ML analysis by generating negative and positive frames at the same temporal location in each individual cube. This setup generates severely imbalanced datasets, meaning that only a small number of the observed frames (3–4%) contain expected sunquake signal. Small datagaps of less than eight frames ($< 360\text{--}480$ seconds $<$ minimum effective duration) have been deemed safe to be interpolated in the cubes.

The very high noise patterns outside ARs along with oscillatory moving patterns that result from the holographic analysis (see Figure 1) make a very challenging dataset. Because of this, we have not considered the number of sunquakes in the two input lists to be definitive. Our curated positive events for our ML-model training represent only the clearly identifiable and distinguishable sunquakes that were inspected and annotated by us based on the sunquake-detection criteria discussed in Section 2.1. Adding to this, the holography analysis on solar ARs is generally considered a convoluted process that requires manual input. We have bypassed these limitations by creating a Sunpy Version 4.0.x (The SunPy Community et al., 2020; Mumford et al., 2022) enabled batch script where we have queried and downloaded magnetogram and dopplergram data from the MDI and HMI repositories corresponding to the temporal slots around our selected sunquakes. We then computed the central heliographic Stonyhurst to Carrington position of the AR at the time of the sunquakes by retrieving and interpolating data from the SolarMonitor service (www.solarmonitor.org). The positions were automatically ingested into the holography method along with a set of other fixed parameters, enabling us to process the sunquake lists and produce temporal cubes of acoustic-egression power corresponding to each event.

The final acoustic datasets from both SC23 and SC24 that are employed by the methodologies described below are available in the online Kaggle SunquakeNet repository (Mercea et al., 2022). The repository also includes the region selections resulting from manual identifications of sunquakes from Besliu-Ionescu et al. (2012) and Sharykin and Kosovichev (2020). Individual-frame datasets ready for ML ingestion in JPEG format separated into positive vs. negative sunquake detections spanning all utilized sunquake datacubes are also included.

2.3. ML Dataset Creation

Before diving into the ML-driven methodology, we will first describe the processes that data undergo in preparation for model ingestion.

To create an ML-suited dataset based on the egression-power maps, frames are extracted from the volumes into gray-scale 224×224 pixels PNG files and a MinMax normalization is applied to each datacube. The obtained images are labeled as follows. An image is considered to be positive, i.e. to contain a sunquake, if the corresponding frame in the egression power map cube shows the presence of a sunquake, and negative otherwise. To identify whether a frame is positive, the starting time present in the egression-power-map file header is used to describe a point in time and, respectively, a corresponding frame range for one sunquake. Based on this equivalence, the sunquake begin and end times are correlated to frame indices. Figure 1 depicts six randomly extracted negative and positive samples from the ML dataset.

One reason for dividing the volumes into 2D images lies in the small quantity of event samples at hand, with only 15 volumes containing sunquakes recorded in SC23 and 38 volumes containing sunquakes in SC24 annotated at the time of this research. An additional 17 datacubes with no events are available from SC23, but are not used in the proposed models due to already increased class imbalance. Moreover, 43 additional cubes from SC24 have been downloaded and processed, but they present weaker and less-clear signatures, caused by the noise threshold inside the ARs achieved via holography. In these cases, the manual sunquake identifications were less certain, where in a few cases we are unsure if we recovered any helioseismic signal at all. As shown by Sharykin and Kosovichev (2020), different events were sensitive to different methods, where here we only employed holography. Thus, the events associated with the 43 additional cubes were dropped from the analysis as our main scope revolves around the correct identification of clear seismic signatures. Although this increases the imbalance, not including these events in the training and validations of current preliminary models, will ensure that we are not inducing incorrect information when testing different ML approaches. In a subsequent project, including these less certain events will be precisely the focus for fine-tuning the models, after identifying a suitable threshold for applying holography.

By dividing the volumes, the sequence structure of the data is not disregarded, as a significant part of the experiments, including the models in Section 4, introduce a form of sequence modeling. A second reason behind this division is the wider variety of deep-learning methods available for 2D-image datasets as opposed to movielike datasets.

From the sunquake catalog presented in Table 1, 53 of the listed events are used for model training and validation, marked with “+”. An additional four datacubes without annotated sunquake signatures are used for testing and analysis, marked with “-”. These latter sets were used to analyze the results for the proposed approaches, with the goal of identifying emissions that are too weak to be detected using conventional methods.

The level of imbalance in the obtained ML dataset is a challenging aspect. By combining both SC23 and SC24 data, the quantity of positive images in the dataset totals 845 ($205 + 640$), while the negative count totals 13,055 ($3891 + 9164$). Some “blank” all-black buffer frames in datacubes are of course excluded from the counts and analysis.

3. Machine Learning Approaches for Sunquakes

To derive a detection model capable of capturing sunquake signatures, several experiments are performed using ML techniques of increasing complexity. These can be divided into

Table 1 Sunquake (SQ) events between 09 July 1996 and 15 January 2005 from SC23 and between 15 February 2011 and 07 September 2017 from SC24. Events marked with “+” are used in training and validating the proposed representation learning models. Events marked with “-” are part of an additional dataset, and not sunquake annotated, and are used for post-learning testing.

Event	SQ Start Frame Index	SQ End Frame Index	Labeled
09 July 1996 09:01	69	87	+
06 June 2000 14:57	47	66	+
24 November 2000 04:54	62	76	+
06 April 2001 19:13	21	36	+
10 April 2001 05:01	47	56	+
24 September 2001 09:35	38	48	+
15 July 2002 19:52	60	74	+
23 July 2002 00:27	30	43	+
21 August 2002 05:24	43	58	+
23 October 2003 08:45	34	42	+
28 October 2003 11:00	170	178	+
29 October 2003 20:35	138	149	+
16 July 2004 13:48	56	69	+
13 August 2004 18:07	15	24	+
15 January 2005 00:33	40	54	+
15 February 2011 01:44	92	104	+
18 February 2011 09:55	107	117	+
18 February 2011 12:59	103	111	+
07 September 2011 22:32	82	105	+
24 September 2011 20:29	90	99	+
25 September 2011 08:46	-	-	-
26 September 2011 05:06	129	139	+
30 December 2011 03:03	-	-	-
05 March 2012 19:27	89	109	+
06 March 2012 04:01	95	110	+
06 March 2012 07:52	82	94	+
06 March 2012 12:23	106	122	+
09 March 2012 03:22	88	103	+
08 May 2012 13:02	-	-	-
09 May 2012 21:01	93	114	+
10 May 2012 04:01	-	-	-
04 July 2012 09:47	112	130	+
04 July 2012 14:35	88	103	+
05 July 2012 03:25	92	106	+
05 July 2012 11:39	89	104	+
05 July 2012 20:09	87	106	+
06 July 2012 01:37	82	96	+
06 July 2012 13:26	84	100	+
24 October 2013 10:30	89	98	+
06 November 2013 13:39	86	103	+

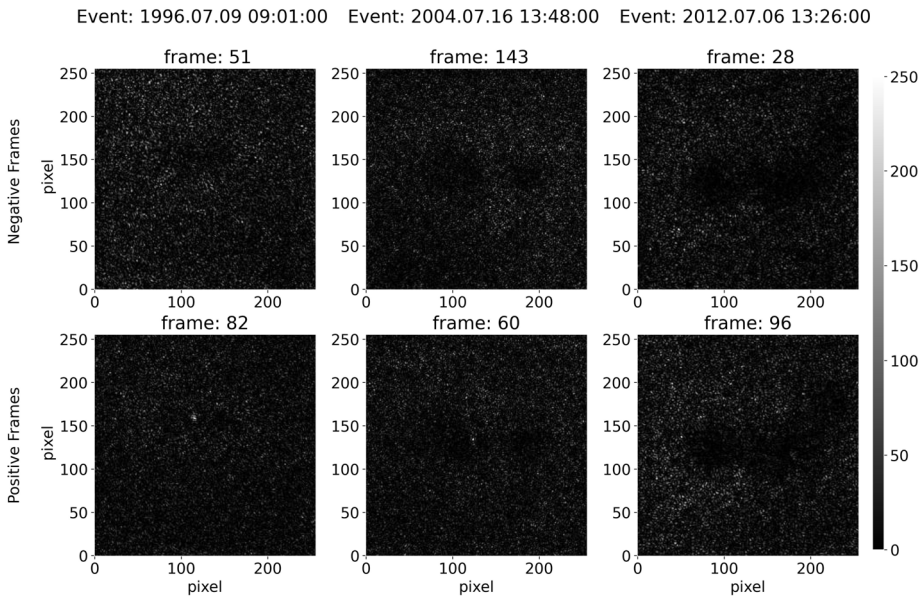


Figure 1 A compilation of six eight-bit scaled intensity data samples from the ML-prepared dataset. *Row titles* indicate the label and *column titles* indicate event identifiers and frame indices.

Table 1 (Continued.)

Event	SQ Start Frame Index	SQ End Frame Index	Labeled
07 November 2013 03:34	84	105	+
08 November 2013 04:20	85	104	+
10 November 2013 05:08	87	103	+
02 February 2014 06:24	SQ1: 90, SQ2: 105	SQ1: 108, SQ2: 114	+
07 February 2014 10:25	86	120	+
09 November 2014 15:24	88	102	+
03 January 2015 09:40	89	106	+
10 March 2015 23:46	95	109	+
11 March 2015 16:11	90	105	+
22 August 2015 21:19	89	97	+
28 September 2015 14:53	85	104	+
30 September 2015 13:18	85	97	+
04 September 2017 20:28	86	102	+
05 September 2017 01:03	86	100	+
06 September 2017 08:57	91	111	+
07 September 2017 10:11	86	101	+
07 September 2017 14:20	99	115	+

two phases: Methods described in Appendix A and Section 3.4 are trained on SC23 data; Methods in Section 3.1 are trained using the combined SC23 and SC24 datasets, based on the process presented above (Section 2.1). The decision to combine the datasets is based

on preliminary findings, which indicated that, given the low-data regime, the amount of AR morphologies that an ML model is exposed to needs to be increased such that the model is able to shift the focus from learning representations of ARs to learning sunquake signatures.

The main ML metrics that will be used to judge model performance are: Precision ($\frac{TP}{TP+FP}$), Recall ($\frac{TP}{TP+FN}$), F1-Score ($\frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$), and Accuracy ($\frac{TP+TN}{TP+FP+TN+FN}$). The TP and FN labels denote true and false positives, and TN , FN the true and false negatives. GT will be used to denote Ground Truth, or real label.

Because the data have domain-specific particularities, learning useful representations from the input has been a primary research goal. Initial experiments are centered on less complex methodologies, starting with small scale convolutional neural networks (CNNs), which prove to be insufficient. Transfer learning from ImageNet (Deng et al., 2009) with common CNN architectures also failed to converge to fully reliable performance metrics. One possible explanation is that ImageNet contains natural images that are quite different from our data, and consequently, the features captured by the pre-trained networks are not sufficiently relevant for our task characterised also by a limited and imbalanced dataset. For this reason, we decided to first move towards representation learning in the form of autoencoder approaches, which also rendered less-satisfactory performance. These are described in more detail in Appendix A.

In this article, we focus on contrastive learning (CL) methods and results, covering supervised (that rely on annotations) and unsupervised (that do not use annotations) objectives, recurrence techniques, and observations on relevant data augmentations. The next subsections describe these experiments and methods, and they highlight the identified challenges and limitations.

3.1. Contrastive Learning Approach

Experiments performed on the holography data indicated that the main challenges in sunquake classification include class imbalance, low-data regime problems, and the inability of autoencoder-based approaches to capture the relevant sunquake features in the latent distribution. The latent distribution represents the reduced parameter vector space produced by an ML model. As a result, the second phase of the experiments is focused on a more recent computer-vision methodology, namely the CL. Cao et al. (2021) argue that, as opposed to autoencoder methods, the goal of CL is not that of finding exact distributions for data samples, but rather about discriminating between different samples.

CL initially emerged as a self-supervised method for learning from visual representations and showed relevant improvements over previous state-of-the-art models, matching supervised network performances (Chen et al., 2020). The goal of CL is intuitive: train a network to generate close latent-space representations for pairs of data points that are similar to one another and distinctive representations for dissimilar pairs.

Recently, a supervised approach to CL has been proposed by Khosla et al. (2020), where clusters of points belonging to the same class are pulled together in the embedding space, while simultaneously pushing apart clusters of samples from different classes. Figure 2 offers an overview of the differences between self-supervised and supervised CL approaches.

The literature presents a number of contrastive-loss functions: max margin loss (Liu et al., 2016), triplet loss (Weinberger, Blitzer, and Saul, 2005), multi-class N-pair loss (Sohn, 2016), and SimCLR loss (Chen et al., 2020), but the main idea behind this method of learning has not changed drastically over the years.

Appendix B.1 describes experiments where a fully self-supervised approach is pursued with the goal of avoiding the problems arising from class imbalance. On average, results of

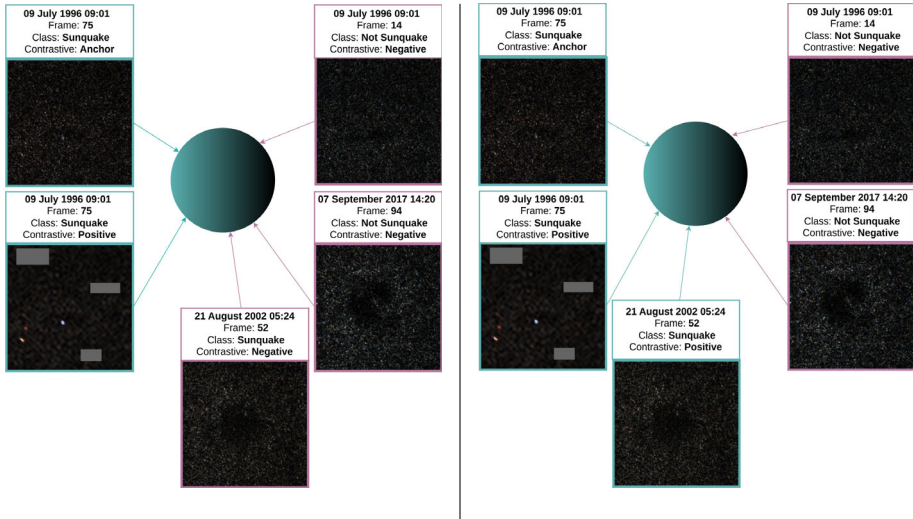


Figure 2 Self-supervised (*left*) and supervised contrastive learning (*right*) Visualization for sunquake image datasets. The self-supervised contrastive loss contrasts the anchor and an augmented version of it (positive) against all other samples (negative), regardless of class. The supervised contrastive loss contrasts the anchor, an augmented version of it, and all the data samples of the same class (positive) against all other samples (negative).

this approach are similar to those of the autoencoder methods. We note that this approach is not able to capture many relevant features. Appendix B.2 describes experiments with a supervised-contrastive model, where we find that some relevant and sometimes distinct features are captured. Metric improvements over the self-supervised model are encouraging, but the class-imbalance factor still posed a great impact on the classification results.

To tackle this, we combine the self-supervised and supervised contrastive models in a unitary pipeline using a two-step approach. Self-supervised and supervised contrastive loss functions are shown in Equations 4–5 and are described in detail in Appendices B.1 and B.2, respectively. For the self-supervised contrastive training, positives are upsampled with geometric transforms. For the supervised part, class weighting is introduced in the contrastive loss, consistent with the equations of Zhong et al. (2022) in order to tackle the imbalance effect. Under this loss, where the tailed-class samples are assigned larger weights, the model achieves the highest precision. The updated weighted supervised contrastive loss is:

$$l_i = \sum_{i \in I} \frac{-w_{y_i}}{|P(i)|} \sum_{p \in P(i)} \log \frac{e^{\frac{\text{sim}(z_i z_p)}{\tau}}}{\sum_{k=1}^{2N} \text{mask}_{[k \neq i]} e^{\frac{\text{sim}(z_i z_k)}{\tau}}}, \tag{2}$$

$$w_{y_i} = \frac{1 - \beta}{1 - \beta^{N_{y_i}}}, \tag{3}$$

where $\beta \in [0, 1)$ is a hyper-parameter and $\frac{1}{w_{y_i}}$ is the effective number of samples for class y_i . In our work, a value of 0.9999 is used for β to halve the effective number of negative samples. During training, Adam (Kingma and Ba, 2014) and stochastic gradient descent (SGD) are used as optimizers. The prior is quickly diverging for the majority of experiments. SGD with momentum, weight decay, and warm-up is found to generalize better.

ResNet-18, Resnet-50, and DenseNet-121 (Huang et al., 2017) backbones are considered case by case, in an attempt to improve results by increasing model complexity.

We note that a set of augmentations additionally play an important role in convergence. These are discussed below. Two relevant models extracted from the above experiments and their results are analyzed in depth in Section 4.

3.2. Data Augmentation and Sampling

To lower the impact of the class imbalance, we experiment with different transforms for augmenting our input data, including: center and random crops, sharpness, color jitter, posterize, invert, auto contrast, solarization, Gaussian blur, normalization, vertical/horizontal flips, and random rotations of 90°, 180°, 270°. In Appendix C we detail the main transforms that are applicable, following the categories of transformations described by Yang et al. (2022). We comment on the applicability of each of these augmentation approaches, as not all of these usually standard transformations are useful for our particular task, some are detrimental to the learning process. We iterate below only custom transforms that are applied to our sunquake data in addition to the more generic approaches illustrated in Appendix C:

- i) *Custom Time-Based Mixing*: When experimenting with various time preserving techniques, we introduce an augmentation that combines subsequent gray-scale frames into a single “3D-like” sample. Channels are represented by consecutive frames. This technique enables the contrastive model to better focus on similarities between samples that are close to one another in both time and space. We remind the reader that sunquakes occur in successive frame series.
- ii) *Custom Solarized Low Pass Filter*: This method is a custom implementation of a color-space transform (Appendix C.5). Given that the process of obtaining seismograms (maps of distances traveled by the wave front) includes a last step of applying a Fourier transform with respect to the azimuthal angle (Kosovichev and Zharkova, 1998), we believe that an augmentation based on such a transform may be useful in enhancing the sunquake details. We apply a low-pass filter followed by a solarize transform. The threshold for both transforms is set to 50. We report that this mix of transforms enhances high-frequency signals, so that sunquake features are amplified. This enhancement is shown in Figure 3, where a more intense sunquake spot is observed on the augmented images. One downside to this transform is that it also enhances high-frequency areas that are not necessarily sunquakes. A more evident example for this is shown in the third image in Figure 3. However, when using this transform with a 50% probability, the contrastive loss no longer stagnates during training, as it does with other transforms typically used in CL. While this statement applies to our data, we cannot draw general conclusions about its applicability. For example, the use of this custom transform with the CIFAR-10 dataset revealed a significant deterioration in the classification results (of $\approx 20\%$ in F1-Score). Because we use this augmentation to better infer domain knowledge to the CL approaches, we experimented with its limited applicability as a method to explain the CL predictions in Section 4.3.
- iii) *Customized Random Erasing*: Random erasing is detailed in Appendix C.6. Here, we develop and apply a custom random erasing implementation in order to provide a higher degree of label-preservation for our data. For this, we propose a method for decreasing the probability of occluding a sunquake: instead of adding standard large-sized rectangle(s) of erased areas, we augment with using multiple small erasing rectangles, of sizes covering up to 4% of the original input image, with an up to 50% probability of application. The smaller sized rectangles have a high probability of being applied, while the large

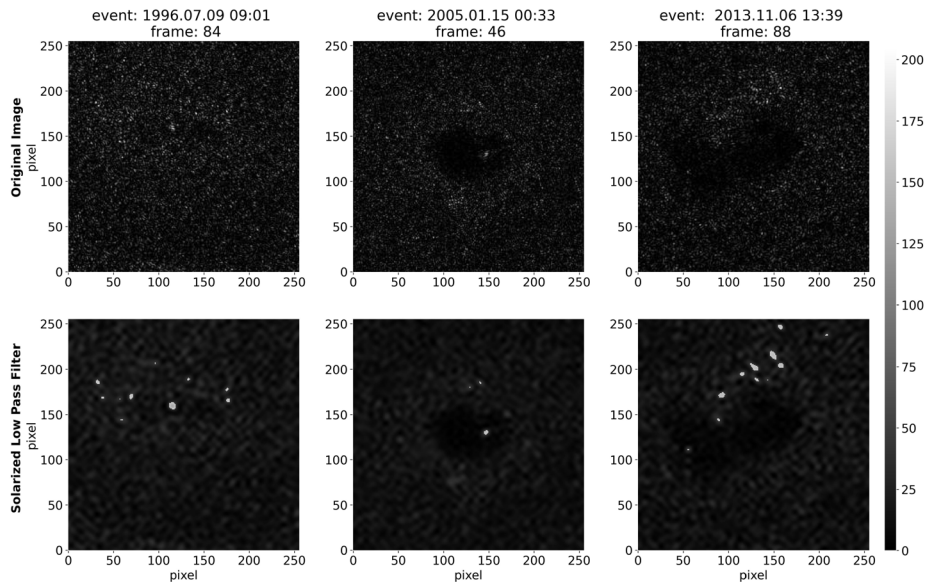


Figure 3 Solarized low-pass filter applied to a selection of eight-bit intensity scaled frames from the SC23 and SC24 datasets. *Frame indices and events are mentioned in the column titles.*

sized ones have a significantly lower probability. For instance, 50% of the images have an erasing rectangle covering 1% of the image. 5% of the images have an erasing rectangle covering 4% of the image. Between zero and eight erasing regions may be applied to a singular image (see the *Contrastive: Positive* image in Figure 2).

3.3. Recurrent Techniques

During the dataset-preparation process described in Section 2.3, the sequence component of the input data is lost. We hypothesize that the main reason for the modest results of single-frame approaches such as those of autoencoder-based models (Appendix A) lies in the inability of the models to capture sunquake signatures from only single data samples. Therefore, separate methods need to be introduced to maintain the time-series structure of the data, as sunquake signatures are only valid if visible at successive times and frames.

To reintroduce the sequence information, the output from the representation model is taken and then combined with previous and next sample embeddings using a sliding window of various temporal sizes. To interpret the clustering of the encoded data points, dimensionality reduction techniques are applied, specifically principal component analysis (PCA: Maćkiewicz and Ratajczak, 1993) and uniform manifold approximation and projection (UMAP: McInnes et al., 2018). The sliding-window methodology is enhanced with the inclusion of the UMAP components for each selected window. This approach, when applied to contrastive representations, yielded weak results (Macro Avg. Metrics: Precision 0.50, Recall 0.51, F1-Score 0.48, Accuracy 0.74), compared to its successor.

These more standard approaches still prove insufficient for contrastive representations. Losing the sequence information during image-based learning makes it difficult for the model to correlate consecutive samples of the same sunquake. Not all single frames that should be positively identified are recovered, as confirmed by the predictions. As an answer to this, we propose the custom time based mixing transform described in Section 3.2.

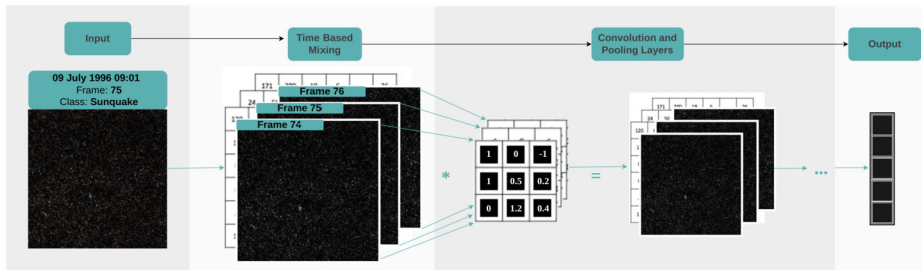


Figure 4 Simplified view of how convolutional filters are applied to the input throughout a CNN-based network. Starting *leftmost*, each input frame is rebuilt using the custom time based mixing transform using a three window input. Then, convolutional and pooling layers are applied to the transformed image. The *rightmost* output data product is flattened to the target latent dimension.

For each model-input gray-scale sample i , a three-dimensional image is created, where the first dimension, that typically represents the channels, stores the sample at index $i - 1$, i , and $i + 1$. In this way, the convolution and pooling filters applied to the sample throughout the models' network capture the evolution of the sample over a time period of three steps. A visualization of this process is presented in Figure 4. The models described in Section 4 make use of data constructed in this manner. This attempt of incorporating the time component significantly improves results (see Table 2).

3.4. Object Detection

Because we currently do not employ additional methods for the explainability of the contrastive models described in Section 3.1, the location of predicted acoustic-emission sources and sunquakes is difficult to infer. Thus, an additional faster region-based CNN (R-CNN) (Ren et al., 2017) based object detection (OD) model is introduced to facilitate interpretation of the model outputs.

The OD model predicts regions as box coordinates around sunquakes in single-frame samples. We note that this model is not related or linked to the CL models described above, and its sole purpose is that of facilitating potential locations for sunquakes in frames where the CL models predict features.

As the OD model is trained only on region-annotated positive samples, it predicts sunquake regions for the majority of samples at runtime. The role of predicting the time of occurrence of sunquakes falls to the contrastive model. By utilizing the predictions yielded by the CL model, positive frames (or positions in time) are identified. Then, the OD model is used for extracting potential regions (or locations) for these predicted frames, to reduce the manual effort of testing the model on unlabeled data.

4. Results and Interpretation

This section presents our most notable results and findings. Following the experiments described in Section 3.1, we train two CL sunquake-detection models with different particularities as described below. The outputs of these models are combined with OD model outputs to derive both temporal and location information of sunquakes for additional datasets.

Due to the large size of the models and the long training time, cross validation is not performed, on account of computational requirements. A locked random state is used at run

time for reproducibility. We note that it is possible for the results relying on this form of data preparation to suffer changes on alternative data shuffles.

4.1. CL Models

In this section we provide the architecture and parameters of two CL models and analyze their predictions. Significant implementation features are shared by both models.

For both CL approaches, the combined SC23 and SC24 data are split during loading in an 80/20% manner into training and validation sets.

We apply the custom time based mixing transform to the entire dataset, followed by the custom low pass and solarize transforms with a threshold of 50 and with a 50% probability. Then, we randomly apply one single positional transform out of horizontal or vertical flip, 90°, 180°, or 270° random rotations, with a probability of 50%. Finally, we apply the random erase transform as described above, generating up to eight gray erasing rectangles of varying proportions.

Before moving on to describing the models, we must emphasize the problem of external bias induction. This occurs when, willingly or not, the model is presented with additional information from the exterior that facilitates training but may impact the reliability of results. From our experiments, we identified two biases that impose a level of risk.

Firstly, when loading the data, if the entire dataset is shuffled before performing the split, different samples associated with the same event may appear in both training and test data, inferring external information to the model regarding previously seen ARs.

Secondly, when upsampling with geometric transformations, even though they are also applied at runtime to all samples, a transformation bias is induced to the model, making it more inclined to predict sunquakes for geometrically transformed samples. When performing inference, no geometric transformations should be applied to the input data so as not to affect the reliability of the predictions.

The impact of such biases is not fully clear. To explain this, we prepared the model in Section 4.1.1, which presents none of the above biases, to offer a clean overview of baseline capabilities. Sections 4.1.2 and 4.3 will describe an in-depth prediction analysis resulting from an impacted model. The analysis is performed on the additional datasets, denoted with “-” in Table 1.

4.1.1. High Precision and Accuracy CL Model with No Known Biases

To mitigate the shuffle bias described earlier, we group the input data by event and only perform shuffling after dividing the groups between training and validation data. This assures that both split sets are self-contained in terms of included ARs. To better illustrate, if we assume that the N th frame of an event is present in the training set, then so are all other frames belonging to that same event, and none are present in the validation set.

We train a self-supervised ResNet-18 ImageNet-pre-trained CL model for 500 epochs on the SC23 and SC24 datasets (using positive upsampling), followed by a weighted multi-head supervised contrastive model for another 100 epochs (without upsampling). For the supervised model, we used an encoding size of 512, a projection dimension of 128, and a τ -parameter of 0.07. The first head includes a linear layer, a ReLU activation function, and an output-encoding layer, while the second head is used for performing and monitoring classification during training and consists of a linear layer.

We then take the resulting encodings for the training and test sets and pass them through an extensive list of classifiers. Building upon the results shown in Table 2, we further analyze

Table 2 Macro average performances of different classifiers over CL-produced embeddings, trained with and without SMOTE augmentation, on the SC23 and SC24 test data (2622 negative and 186 positive samples). The bold entries represent the highest scores for each metric.

Classifier	K-NN (Bagging)		SVC (Poly)		SVC (RBF)		Logistic Regression		SGD	
	None	SMOTE	None	SMOTE	None	SMOTE	None	SMOTE	None	SMOTE
Precision	0.63	0.65	0.66	0.84	0.49	0.59	0.64	0.64	0.54	0.62
Recall	0.54	0.54	0.54	0.54	0.50	0.54	0.54	0.54	0.54	0.54
F1-Score	0.55	0.55	0.55	0.55	0.49	0.55	0.55	0.55	0.54	0.55
Accuracy	0.93	0.93	0.93	0.94	0.93	0.92	0.93	0.93	0.89	0.93
Metric Avg.	0.662	0.667	0.67	0.715	0.605	0.65	0.665	0.665	0.627	0.66

Table 3 TP, FP, FN sum for SMOTE augmented SVC (poly) predictions and positive GT count for each event in the SC23 and SC24 test data (2622 negative and 186 positive samples). In the case of events where $TP \equiv 2$, the model predicts the exact transition frames into and out from the SQ.

Event Date	Counts				Event Date	Counts			
	TP	FP	FN	GT		TP	FP	FN	GT
09 July 1996 09:01	0	0	19	19	04 July 2012 09:47	2	0	17	19
06 April 2001 19:13	0	0	16	16	06 July 2012 13:26	2	0	15	17
24 September 2001 09:35	0	0	11	11	08 November 2013 04:20	2	0	18	20
23 July 2002 00:27	0	6	14	14	11 March 2015 16:11	2	0	14	16
05 March 2012 19:27	2	0	19	21	28 September 2015 14:53	2	0	18	20
06 March 2012 07:52	2	0	11	13					

the embeddings with respect to the predictions provided by the polygonal kernel SVC. When loading data to this model, we separate events entirely between training and test data, and we infer the contrastive encoding on raw images only, avoiding both the shuffle and the transformation biases.

To further tackle imbalance, besides encapsulating weights in the contrastive loss according to Equation 2, we use SMOTE augmentation (Lemaître, Nogueira, and Aridas, 2017) on the training data. We synthetically up-sample positives with a sampling strategy of 0.2, and down-sample negatives with a sampling strategy of 0.75. By this, first the positive samples are increased so that they measure 20% of the total count of the negative samples, and the negative samples are reduced until the positive count is equal to 75% of the negative count, leading to 2753 negative samples and 2065 positive samples used for training the classifiers.

Performance of this model before and after applying SMOTE augmentation are provided in Table 2. We see improvements of up to 20% in precision, and a 1% accuracy increase in SVC (poly). Logistic regression seems to not be affected by this augmentation, which may be explained by how the dense minority-class is distributed closely to the sparse majority-class, as depicted in Figure 5.

Although Table 2 shows a high precision score, recall is small with only 20 predicted sunquakes samples, out of which 6 are false positives. The test set contains 11 sunquake events, listed in Table 3. As the table shows, 7 events are recovered by this model.

To further analyze the particularities of the correct predictions and identify the reason behind false positives, several characteristics of the embeddings are analyzed, beginning with the UMAP clustering, cosine and euclidean distances, and cosine similarities between

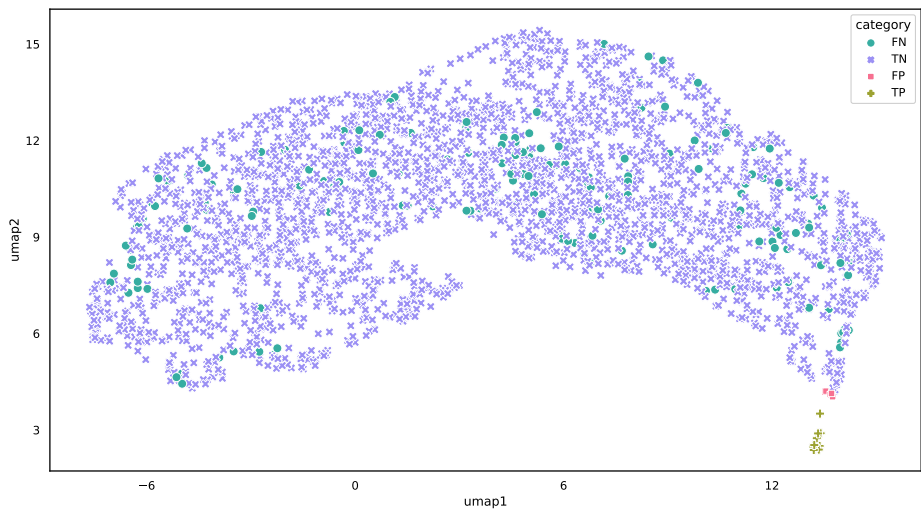


Figure 5 The predictions of the high precision and accuracy model, on the test set (20% of the SC23 and SC24 data), colored by their prediction correctness and clustered by UMAP components.

consecutive frames. Using euclidean distance provides almost identical results. Findings indicate that there is a considerable difference between the embeddings associated to data in SC23 and SC24, indicating either that models may recognize the different measuring instruments.

In Figure 5, a UMAP plot is presented, depicting on the two axes synthetically extracted features based on the contrastive embeddings. The correctly predicted samples are visibly clustered at the tip of the other points, in the lower right-hand corner. FP are grouped together very close to TP, which is expected considering the use of a polygonal kernel during classification.

At a first glance, an evident issue with the embedding clustering lies in the distribution of FN samples, which are randomly distributed alongside the TN.

Figure 6 shows the measured cosine distances between consecutive frames, as a non-outlying plot of a successfully recovered sunquake in the 06 July 2012 13:26 event from the test set. We observe that the predicted frames corresponding to the recovered events are the leftmost and rightmost margins of the sunquake. A maximal cosine distance between the embedding vectors outputted by the CL component of our model appear between frames 83–84 and 100–101. The colors indicate the model's predictions with respect to each frame's embedding vector. These highly distanced frames are exactly the human-identified transitions into and out of a sunquake. This high cosine distance is considered to be the main reason behind the model's positive prediction for these transition frames. We note that all of the recovered sunquakes in SC24 maintain roughly the same characteristics.

We attempt to justify this behavior by pinpointing that for this experiment, when training the CL ML model, each input data sample is augmented with the custom time-based mixing transform. We hypothesize that because of this, the embedding is capable of capturing a gradient in the intensity of the sunquake region between the channels of an individual sample. Further increasing the number of used channels might improve this result.

To understand why the immediately nearby frames located right before and after a sunquake are not also predicted as positives in spite of the visibly large cosine distance that they

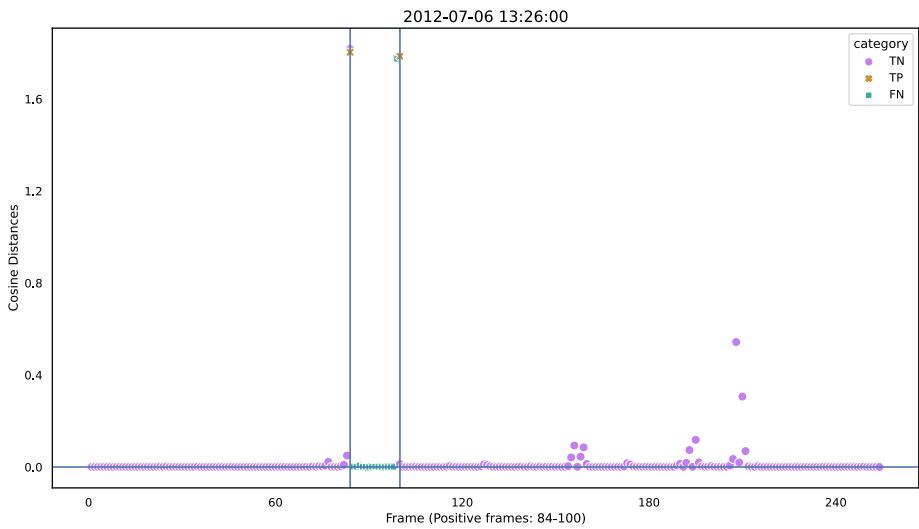


Figure 6 Cosine distances computed between consecutive frames' embedding for the 06 July 2012 13:26 event in the SC23 and SC24 test data, colored by prediction correctness. The x-axis is represented by the frame index for this event, and the y-axis denotes the Cosine Distance value.

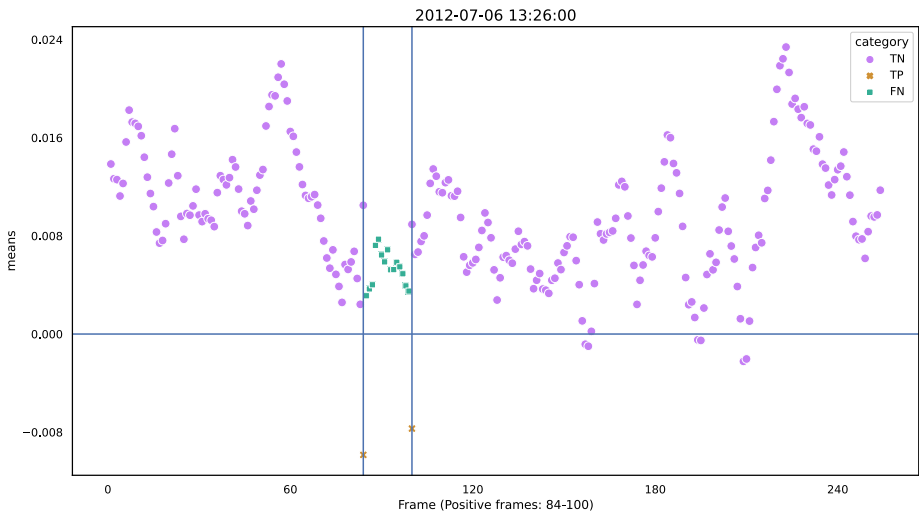


Figure 7 Means for each frame embedding vector for the 06 July 2012 13:26 event in the SC23 and SC24 test data, colored by prediction correctness. The x-axis is represented by the frame index for this event, and the y-axis denotes the embedding mean value.

also present, we look into means and medians, sample level standard deviation, and embedding vectors difference from the mean. All these characteristics indicate a typical behavior for immediate pre- and post-quake frames as compared to other event frames, and a more evident discrepancy for sun quake marginal frames.

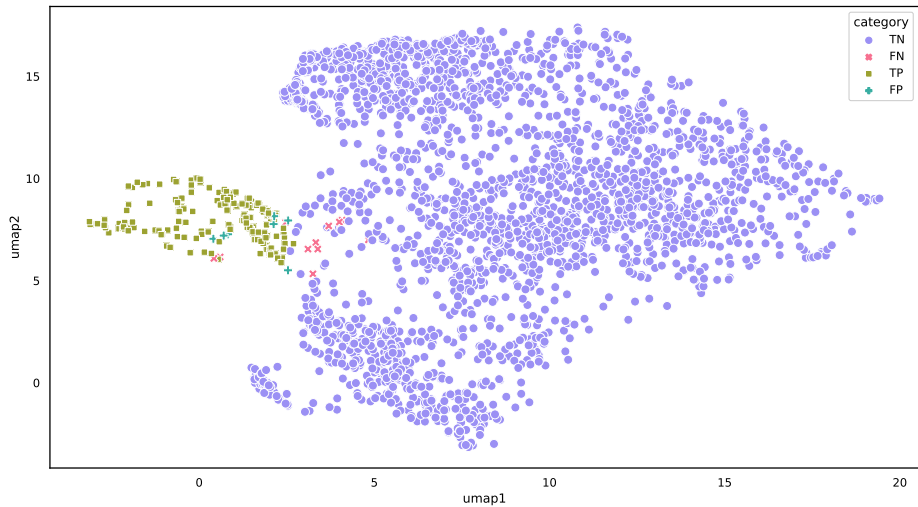


Figure 8 The predictions of the high metrics with shuffle bias model, on the test set (20% of the SC23 and SC24 data), colored by their correctness and clustered by UMAP components.

An example of this behavior is depicted in Figure 7. We can easily see that the TP embeddings have a much lower mean than the adjacent ones. This behavior is also occurring in the other characteristic plots, but to a less evident extent.

There are four events for which the model is not able to identify correct sunquake frames, as seen in Table 3. However, the model does show common behavior when marking positives with respect to the cosine distances between consecutive frames. For example, in the 23 July 2002 00:27 event, the model predicts a shift in the gradient intensity at frames 170–173 and 233–235, but this is not due to any known sunquake. This can result either from an abnormal noise pattern or from events that were not visually identified by human observers.

Lastly, we apply this model to the data associated to the additional datasets marked with “-” in Table 1. For these four unlabeled events, the predictions are as follows:

For 08 May 2012 13:02, one sunquake is predicted, around frames 180–188. This will be analyzed in more detail below. For 30 December 2011 03:03, a significant number of frames are predicted as sunquakes, which, given the high precision of this model, indicates that the dataset might be too different from those previously seen. Visually, the AR does not seem that clear with respect to the noisy quiet-Sun area around it. One sub-interval of positive predictions is analyzed in depth below. For 10 May 2012 04:11 and 25 September 2011 08:46 no sunquakes are detected. Each of the additional sets should have observation of one sunspot, as retrieved from our literature source (Sharykin and Kosovichev, 2020, Table 1).

4.1.2. High Metrics CL Model with Shuffle Bias

We train a supervised CL model with a DenseNet-121 ImageNet-pre-trained backbone, batch normalization, dropout, global average pooling, τ -parameter of 0.1, and an encoding and projection of size 20 on the SC23 and SC24 dataset (using positive upsampling) for ≈ 50 epochs. This model encapsulates the shuffling bias, in that it is trained on the fully shuffled SC23 and SC24 dataset, at sample level. The encoding produced by this model is used to perform the classification. For this, similar to the previous model, an SVC classifier with a polygonal kernel is chosen.

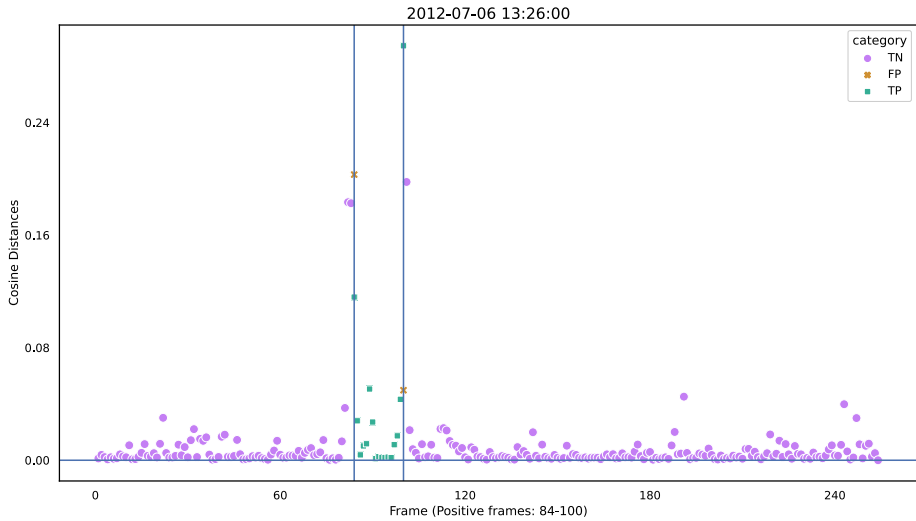


Figure 9 Cosine distances computed between consecutive frames' embeddings vectors for the 06 July 2012 13:26 event in the SC23 and SC24 test data, colored by prediction correctness.

This model comes as a significant improvement to its predecessor in terms of metrics, which are presented in Table 4. Despite the boost, this model introduces a level of uncertainty due to the present bias related to data shuffling. To train the model, we modify the data loader so that data are no longer grouped by their corresponding event date before the split. Hence, after shuffling, the training and validation set may contain frames belonging to the same initial cube, such as the N th frame belonging to event 06 April 2001 19:13 residing in the training set, and the $N + K$ th frame residing in the test set. This modification facilitates the model's ability to produce meaningful embeddings for test data, as the same AR might be present in the training-set samples.

Figure 8 shows clustering of the embeddings produced by the model for the test set. The distribution is quite sparse, but there is a clear separation between the positive and negative class. Falsely predicted samples are tightly coupled in between both clusters, thus making it difficult for the model to clearly classify them.

In spite of the shuffle bias, this model presents plots of cosine distances between consecutive frame embeddings, as that in Figure 9, maintain the characteristics of those in the high-precision model described earlier in Section 4.1.1.

Common behavior displayed between the two different models indicates that aspects such as embedding means and cosine distances are relevant during the process of learning the given task. It also validates the assumption that by using the custom time-based mixing augmentation described in Section 3.2, transitions to and from a sunquake are adequately captured by CL models.

This particular model presents no weighting at the CL level, indicating that models are much more easily trained on sample-level shuffled data. Even though this model appears to lack imbalance impediments, we provide a comparison between classification with and without SMOTE augmentation in Table 4. An improvement in F1-Score up to 30% is noted for K-NN and SVC (RBF), but logistic regression seems to suffer from the augmentation, with a 6% decrease in precision, as analogous to the no performance gains observed for logistic

Table 4 Macro average performances of different classifiers over CL produced embeddings, trained with and without SMOTE augmentation, on the SC23 and SC24 test data (2622 negative and 186 positive samples). The bold entries represent the highest scores for each metric.

Classifier	K-NN (Bagging)		SVC (Poly)		SVC (RBF)		Logistic Regression		SGD	
	None	SMOTE	None	SMOTE	None	SMOTE	None	SMOTE	None	SMOTE
Precision	0.97	0.80	0.97	0.98	0.99	0.94	0.97	0.91	0.91	0.89
Recall	0.54	0.96	0.96	0.97	0.86	0.99	0.98	0.99	0.94	0.93
F1-Score	0.55	0.86	0.97	0.97	0.91	0.96	0.98	0.95	0.94	0.93
Accuracy	0.94	0.95	0.99	0.99	0.98	0.99	0.99	0.99	0.98	0.98
Metrics Avg.	0.750	0.892	0.947	0.977	0.935	0.97	0.98	0.96	0.942	0.932

regressors in Section 4.1.1. The SVC (poly) which is the top average scoring SMOTE-based classifier gains a small 1% improvement in precision and recall.

4.2. OD Model

This section provides the validation results of the OD model for both SC23 and SC24 test data. The model was trained on 80% of the positive events in SC23 using a Faster R-CNN architecture and a ResNet-50 backbone, pre-trained on ImageNet. This selection is due to the limitations of MDI. The instrument was sensitive enough to only capture stronger sunquake events, leading to a selection bias in which most detected events have good signal-to-noise ratio in the egression power maps. For this localization task, this choice has the advantage of producing a clean qualitative dataset, but also the disadvantage of incompleteness.

The most commonly used validation metric in OD is the intersection over union (IoU), which quantifies the degree of overlap between the predicted region box and the GT. Table 5 shows four types of metrics. The IoU metric results i) on our data appear underwhelming. This is because the manually annotated regions vary in size between events, oftentimes including padding. For this reason, we analyze the detections using additional metrics that better capture the desired outcome of this model. We look at: correct signature coverage ii), the overlap of averages over different minimum sunquake duration iii), and the percentages of predicted boxes inside the GT iv). Additional information and visualizations on decreased IoU values in the case of correct detections for the test data in SC24 are provided in Appendix D.

For SC24, the predicted IoU and the GT boxes overlap by $\approx 21.5\%$. An average coverage of correctly localized sunquake signatures in 44.2% of the total positive frames for singular events. This means that although sunquake locations are recovered, not all consecutive frames corresponding to one event are successful in capturing the signal. By manually reviewing predictions, we found that the model tends to perform better for non-marginal sunquake frames, supported by the fact that a significant part of the training data contains stronger examples. Moreover, we test introducing the required minimum duration of predicting the sunquake signature in the same spot. We find that while increasing the minimum duration time, the average percentage of identified signatures for singular events decreases down to 39.3%.

We assume the 44% and 62.6% event-level average of correctly marked signature regions for SC24 and SC23, respectively, to be sufficient for our current goal of enhancing the CL model predictions with a probable location component. Importantly, we note that the currently described OD model should not be considered adequate as a standalone detection tool.

Table 5 Event-level OD model performances for capturing sunquake signature boxes for the positive samples in SC23 and SC24 test sets.

	Metric	SC23 Data	SC24 Data
i)	IoU avg.	29.5%	21.5%
ii)	correct-signature coverage	62.6%	44.2%
iii)	min. duration 1 frame: correct-signature detections avg.	100%	84.8%
iii)	min. duration 4 frames: correct-signature detections avg.	75%	63.6%
iii)	min. duration 6 frames: correct-signature detections avg.	75%	45.4%
iii)	min. duration 8 frames: correct-signature detections avg.	75%	39.3%
iv)	frame-level predicted boxes included in GT	30%	20%

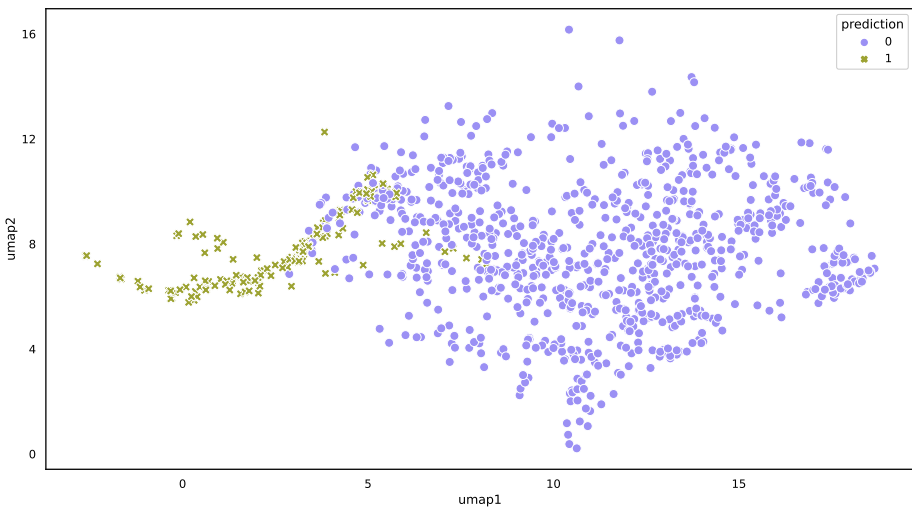


Figure 10 The predictions of the high metrics with shuffle bias model, on SC24 additional dataset, colored by their prediction value and clustered by UMAP components. 0 represents a negative prediction and 1 represents a sunquake prediction.

4.3. An Analysis of Sunquake Detections on Additional Datasets

As both models’ UMAP representations in Figures 5 and 8 present a number of FP detections that are clustered very close to the TP detections, we ask the following: Are these CL trained models sensitive enough to detect additional sunquakes that might be hard or nearly impossible to discern by human observers? We devise an experiment where we analyze the outputs of the trained CL models on the additional SC24 AR datasets. These data were not used in any training or validation phases, and are marked with “-” in Table 1.

For the aforementioned “-” events in Table 1, the CL model predicts a total of 93 positive frames and 923 negatives. The OD model described in Section 3.4 is then applied to the positively predicted frames to extract potential regions. Figure 10 shows the clustering of embeddings, colored by prediction. Although slightly different in distribution, the UMAP is quite consistent in interpretation to the test set clustering shown in Figures 5 and 8 discussed above.

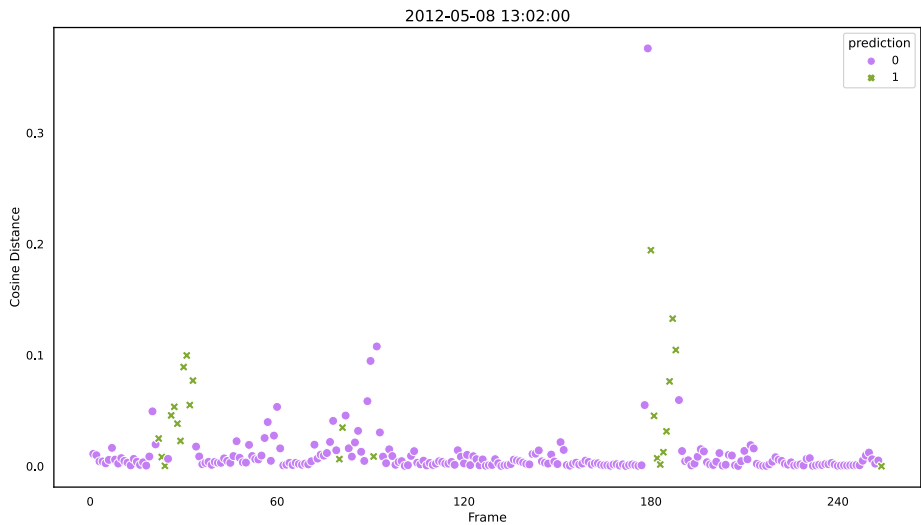


Figure 11 Cosine distances computed between consecutive frames' embedding vectors for the 08 May 2012 13:02 dataset of SC24, colored by their prediction value. 0 represents a negative prediction and 1 represents a sunquake prediction.

The predictions are as follows: For the two datasets of 10 May 2012 04:11 and 25 September 2011 08:46 frame 199 is marked, and frames 87–89 are marked as sunquakes, respectively. Per our identification and selection criteria, one and respectively three frames are insufficient to justify a sunquake signature. Analysis of the higher-atmosphere data showed no candidate eruptions. With respect to the 08 May 2012 13:02 dataset, two sunquakes are predicted at frames 22–33 and 180–188 in different locations. A cosine-distances plot is shown for this event's embeddings in Figure 11. Similar to the test events, the medium-high and high values of this characteristic for sunquake margins are maintained for both detections, respectively. The first prediction proved to be a false positive. The second sunquake is predicted by both CL models, and it is further scrutinized below. The 30 December 2011 03:03 dataset contains six sunquake identifications. Five cases either proved inconsistent with the temporal-detection criteria, or the OD did not successively converge on the same location for the entire CL duration. The last detected event [15–22] is given further consideration below.

We aimed to enhance the level of explainability of the contrastive model, in the absence of other implemented methods, by additionally utilizing our most impactful augmentation, the solarized low pass filter custom transform, alongside the OD approach.

4.3.1. 08 May 2012 13:02 Prediction Analysis

In Figure 12 we observe the predicted 08 May 2012 13:02 sunquake at frames 180–188 using both detection approaches. Both models described in Sections 4.1.1 and 4.1.2 predict a sunquake occurrence in this temporal interval. Starting from this, a manual review of the respective regions is performed to identify sunquake presence and evaluate the behavior of the algorithms used. The OD identification shows a feature at position [80, 135] (Figure 12 rows 3 and 4). The detection marked with a purple box maintains a fixed position starting from Frame 181, close to the center of the AR complex. The solarized low pass filter presents

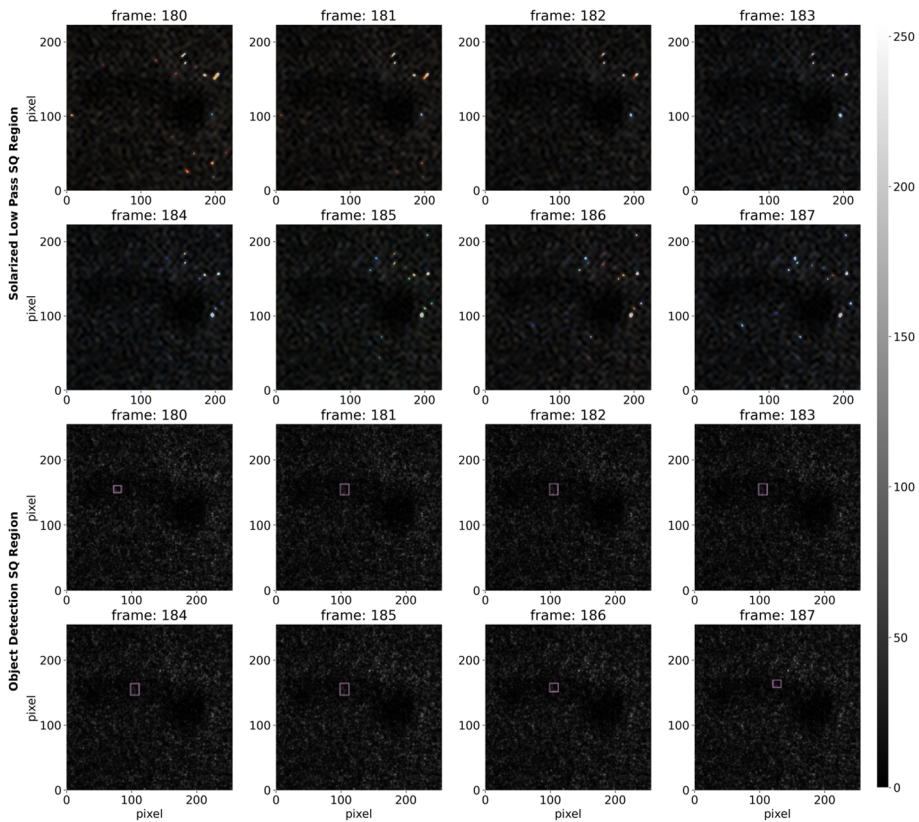


Figure 12 Position of identified acoustic signatures based on solarized low pass filter and OD regions for dataset 08 May 2012 13:02, at positively predicted frames [180–188]. Eight-bit intensity scaling is used.

a spot of gradually increasing intensity on the right-side region at position $\approx [200, 100]$ (Figure 12 rows 1 and 2). The other features are too short-lived to be classified as sunquakes. These are at two completely different locations, where each appears consistent temporally with the CL model detection.

To disentangle this aspect, we use higher atmosphere observations from SDO's *Atmospheric Imaging Assembly* (AIA: Lemen et al., 2012) and from the *Reuven Ramaty High-Energy Solar Spectroscopic Imager* (RHESSI: Lin et al., 2002) to probe the eruptive and high-energy signatures that are usually associated to sunquake activity. The AIA data calibrated to Level-1 data are obtained from the JSOC (jsoc.stanford.edu) around the predicted acoustic-source time intervals. The RHESSI high-energy source location is computed using the CLEAN algorithm (Hurford et al., 2002) by integrating the signal in the 6–12 eV range measured around times of the maximum X-ray emission for each event.

We explore the emission of the solar atmosphere during the times indicated by the OD kernel detection. For this flaring event, the most significant signatures are observed in the AIA 304 Å emission originating from the solar chromosphere and transition region, as presented in Figure 13. The OD detected position (cyan square) appears to overlap the foot-points of a mostly chromospheric flaring event. The weak RHESSI X-ray 6–12 eV source (green contours) is found to match the location of the AIA flaring. We note that the flaring

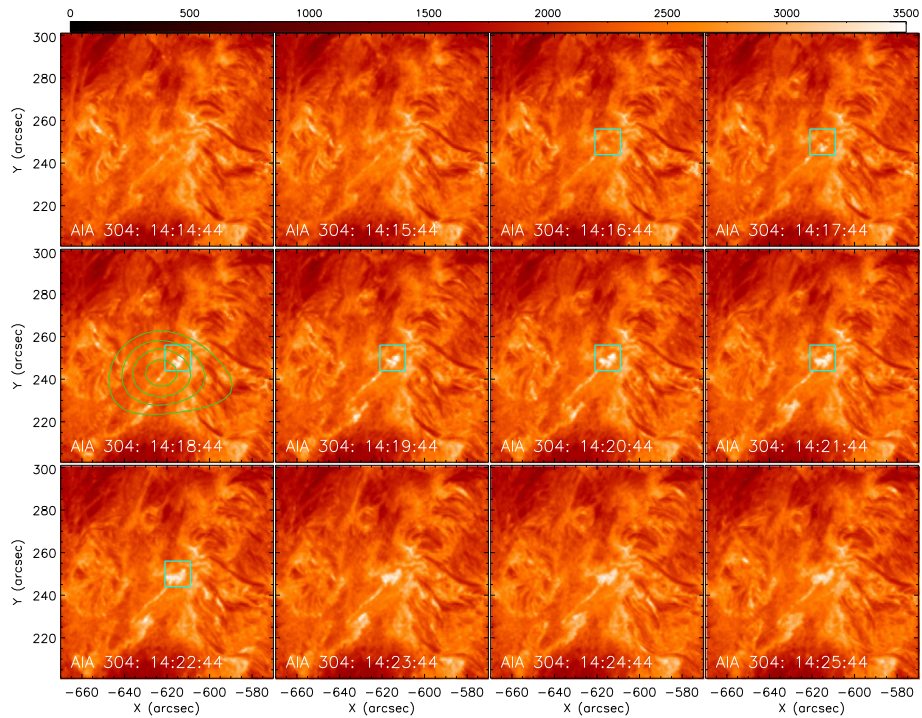


Figure 13 Flaring activity as seen in the AIA 304 Å channel related to the 08 May 2012 13:02 dataset. The observation time for each AIA frame is included at the bottom of each frame. For the temporal interval when the OD kernel is identified, its position is marked by a cyan box in the corresponding frames. The RHESSI high-energy X-ray 6–12 eV signature location is shown as the green contours for the time of the maximum flaring.

appears to be visible for more frames beyond the OD-marked interval. This is consistent with the fact that the signal in the egression-power map is generated by the impulsive initiation of the flaring, while afterwards the chromosphere continues to radiate the generated energy. The second location inferred from the solarized low pass filter is discarded, as we could not find any clear eruptive or high-energy manifestation that can be associated with this location and time.

4.3.2. 30 December 2011 03:03 Prediction Analysis

For our second example dataset, 30 December 2011 03:03, the solarized low pass filter is shown in Figure 14, while Figure 15 shows the hot AIA 94 Å channel in which most of the emission is recorded for this particular flare. The locations of the OD kernel (cyan) and the RHESSI source (green) are also included. This event occurs very close to the solar limb, so projection effects are non-negligible in both the AIA emission and in the data used for acoustic-signature identification. The egression-power maps are de-projected to remove the solar rotation, while the AIA and RHESSI data are significantly influenced by projection effects. In addition, the egression-power maps are mapping the solar photosphere, while the AIA 94 Å channel is mapping the very high corona. Thus, we explain the small mismatch of about 10'' – 15'' between the source in the egression-power maps and the AIA

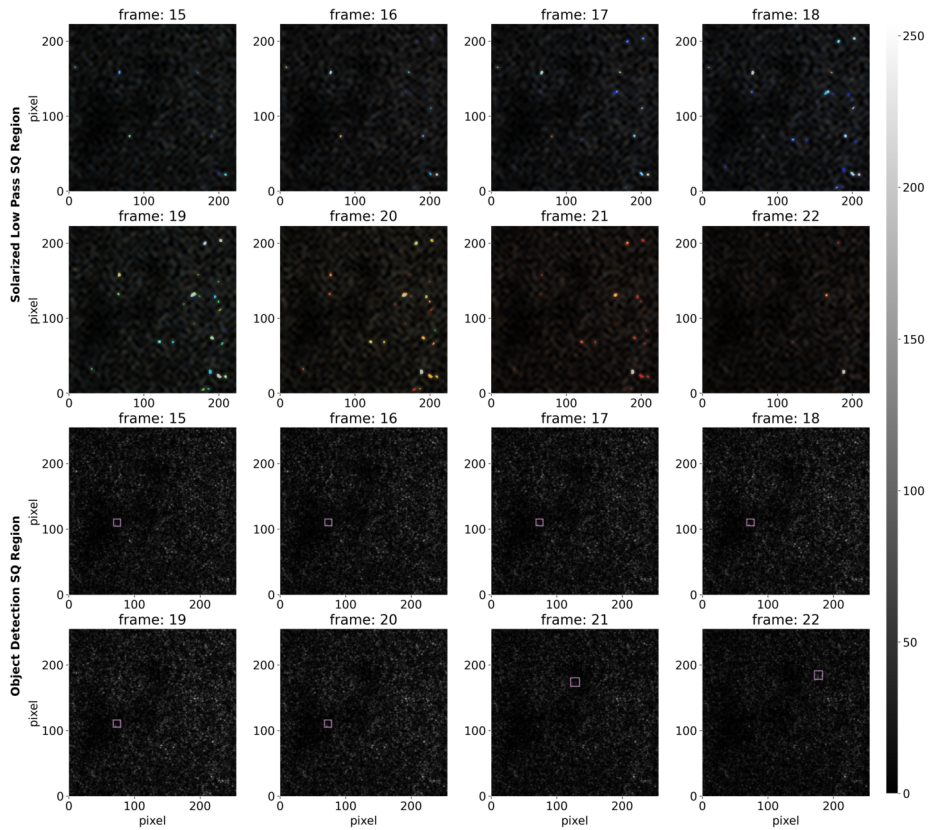


Figure 14 Position of candidate sunquake signatures based on solarized low pass filter and OD regions for dataset 30 December 2011 03:03, at positively predicted frames [15, 23]. Eight-bit intensity scaling is used.

and RHESSI data as a product of superposing all these effects. In this case, the solarized low pass filter shown in Figure 14 did not capture our small acoustic region of interest, and it has not identified other stronger acoustic sources with sufficient lifetime for consideration. We note that the insufficient frames where the OD has consistently detected the flaring location (≈ 300 seconds), makes this event not fully compatible with a sunquake identification. This aspect, when coupled with the small location mismatch and the high projection effects of the observation, makes this association less strong than in the case of the 08 May 2012 13:02 dataset, but still relevant, at least with respect to qualitative and prospective application criteria.

4.3.3. Discussion

In both examples, we observe the AIA and RHESSI data to show eruptions accompanied by high-energy X-ray emission with class $\approx C1.0$ at the approximate location of the OD source during the same temporal intervals of the CL predicted sunquake intervals in both CL models. We tentatively hypothesize, by conjecture, that these source locations might be desiderated weak acoustic emission signatures that are produced by less powerful eruptions, even weaker than the source discussed by Sharykin, Kosovichev, and Zimovets (2015).

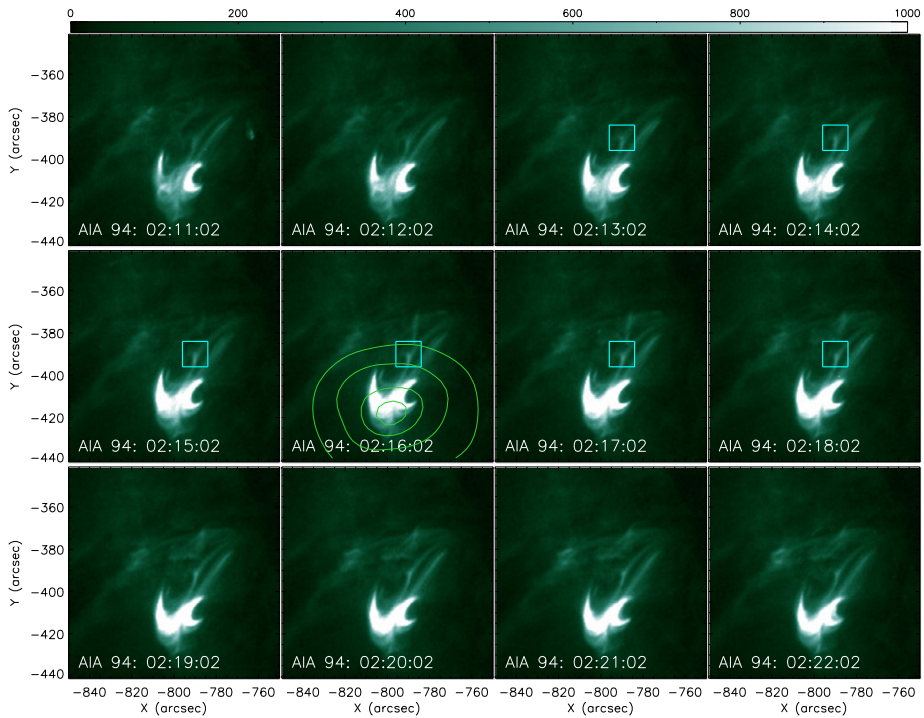


Figure 15 Flaring coronal loops as seen in AIA 94 Å channel related to the 30 December 2011 event, with the position of identified OD kernel marked by the cyan box and the RHESSI high-energy 6–12 eV signature location shown as the green contours.

Figure 16 presents a more detailed analysis of the acoustic emission accompanying the AIA and RHESSI flares. The a) and b) panels show that both events are visually identifiable in egression power maps. The total emitted power over the three-hour background in ARs (P/P_{avg}) exceeded 7 and 14 in individual locations for the 31 December 2011 and 08 May 2012 events, respectively. The c) and d) panels show the P/P_{avg} integrated over different kernel sizes over the three hours of observation of each AR. The more compact kernels (purple) maintain detection levels of above $4\sigma_{\text{ar}}$ in both events discussed above. We note that detection limits would decrease even further to a $\approx 3\sigma_{\text{qs}}$ level if evaluating the temporal median signal in regions outside the less noisy AR shadows.

Following the discussion of Chen and Zhao (2021), we assess that these detections represent very weak sources. In addition, we note that in general a modest 4σ -like detection does not completely exclude us from the risk of spurious signal. To further constrain the detection, we additionally track the kernels over longer temporal intervals of the order of one day for each event (panels e) and f)). Both discussed detections remain the most dominant features in the tracked regions, with detection levels reaching a more optimistic $5\sigma_{\text{ar}}$ in this extended temporal noise statistics. These concerns are also somewhat alleviated by the temporal consistency, and the sets of clustered pixels manifesting similarly. Lastly, when following the temporal-evolution curves in panels c) and d) during the peak acoustic emission, we find that the 2011 event maintains a consistent $> 3\sigma_{\text{ar}}$ detection for ≈ 495 consecutive seconds, on the edge of the detection threshold set by Equation 1. The 2012 event falls marginally under the established threshold, where a $> 3\sigma_{\text{ar}}$ detection is maintained for ≈ 450 seconds. Thus,

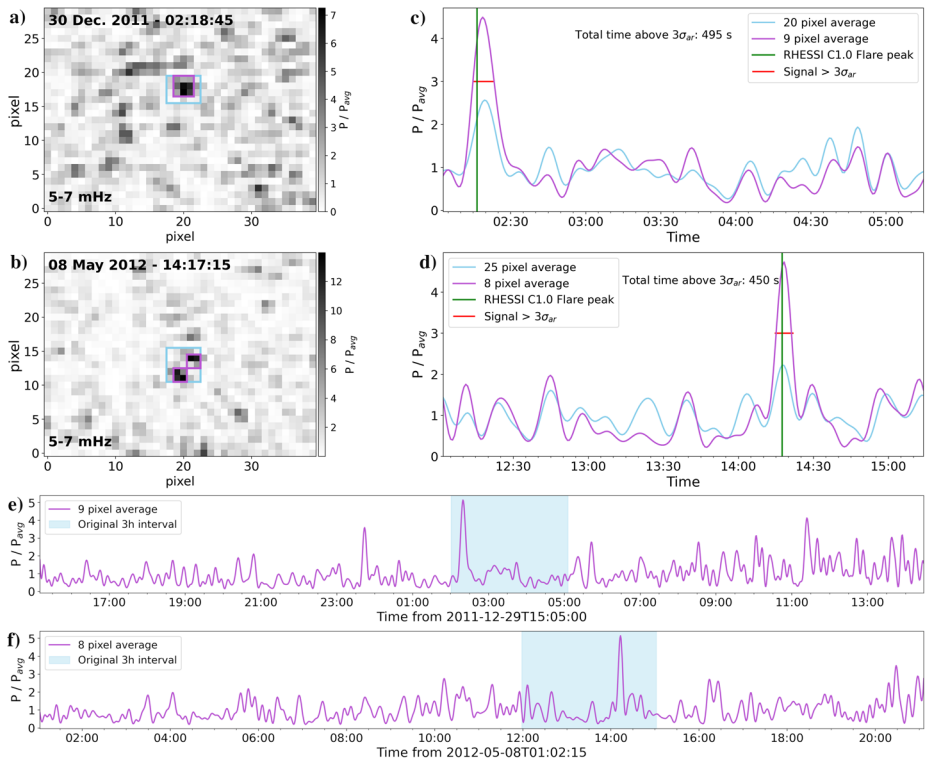


Figure 16 Panels **a)** and **b)**: Background normalized acoustic-emission power maps in the 4–7 mHz band at the maximum of emission for the two detected weak sources. Panels **c)** and **d)**: Evolution of the measured power with respect to background over two highlighted regions over our standard 3 hour dataset interval, one strictly including the acoustic kernels (*purple*), the other also containing some background contribution (*light blue*). The *green vertical line* corresponds to the time of maximum RHESSI emission, and the *horizontal red line* marks the interval when the retrieved signal is over the 3σ level. Panels **e)** and **f)**: The highlighted *purple regions* are tracked for longer temporal intervals to better constrain the background signal. The original 3 hour intervals are highlighted in *blue*.

although a set of stacking correlations between reasonable acoustic power, co-temporal AIA flaring, along with qualitatively overlapping RHESSI sources present a consistent set of evidence, we stress that these two aforementioned events would be hard to classify as bona-fide sunquakes when following a traditional holography-based helioseismology analysis, as a number of methodological and statistical criteria (detection limit, temporal and spatial correlations, kernel size, temporal length, etc.) can hold up to only a qualitative level. These limitations can originate from any combination of instrumental, ML model, interpretation, or statistics effects.

We also remind the reader that our findings represent only a footing, upon which we encourage further exploration, as it is based on only two out of four analyzed datasets that contained one apparently positive example each. Preparing and ingesting a large dataset of acoustic emission of ARs that do not formally contain known sunquakes is required alongside more thorough detection-limit statistics in order to confirm this result. Such a dataset should not be included in the training/validation and used only for highlighting temporal and spatial location of potentially unknown acoustic events. When using only HMI, such

a type of study can also help with establishing a lower flare energy detection limit when concerning ML applicability.

Although the solarized low pass filter improves detection metrics, as has been demonstrated in Section 3.2 and exemplified in Figure 3, we find it is prone to introducing artifact regions into outputs when used in a stand-alone manner for explainability, as it is agnostic of our imposed criteria. For example, due to current holography limitations, the acoustic sources usually need to be inside the ARs to be identifiable. We thus deduce that although using this custom augmentation is tremendously beneficial to the overall training and validation of solar acoustic-emission datasets, careful consideration should be put into its explainability applications and sources that it might locate, as they have a significant chance to be FP. As a corollary, we note that this augmentation will increase the FP metrics of a trained model. This aspect again illustrates the complexity in interpreting these observations in a criterion that is agnostic to ML.

False detections such as the second location at frames 180–188 in event 08 May 2012 13:02 indicate that our current models are not very robust to spurious correlations; for instance strong shifts in the gradient intensity in the ARs shadow, which are essentially noise, are predicted as sunquakes. To alleviate this, techniques that instruct the model to distinguish between false and correct sunquake signatures need to be employed. As a first step towards this, we plan to label our future datasets into multiple classes, such as: strong SQ signature, weak SQ signature, AR shadow intensity shift (not SQ), etc. These labels can then be adopted in the CL methods to encourage producing sunquake representations that are dissimilar to their unfounded lookalikes.

We reiterate that both identified acoustic-emission sources, although promising as potential applications, do not present the actual position of sunquake identifications from the perspective of the CL models, which only provide the temporal component. In the future, we will add the significant effort required to augment these trained models with a suite of CL-specific explainability features, which will be able to extract positions directly from the model output.

5. Conclusion

In this work, we presented a pedagogical approach on the application of ML methodologies to acoustic emission image data of solar ARs. We constructed a curated dataset of major sunquakes during SC23 and SC24, using the holography method. An extensive list of representation-learning-centered ML experiments are performed on this dataset, and two end-to-end models are analyzed from a solar-physics phenomenological standpoint. Figure 17 offers a step-by-step visualization of the process behind the proposed methodological setup. The following summary reiterates our most relevant findings and plans for future improvements.

- **Dataset construction:** We created a comprehensive dataset of acoustic-emission power maps for ARs that contain sunquake signatures. These ML ready datasets are available in the linked repository. We emphasize that characteristics of this dataset include a large imbalance factor of ≈ 1 to 15–25 positive to negative ratio per dataset, very high noise patterns outside ARs, and the presence of artifacts of oscillatory moving features that result from the holographic analysis.
- **Impact of custom augmentations:** three domain-specific transforms were developed and introduced. They significantly improve our CL-based models' performance on our difficult dataset: customized random erase, solarized low pass filter, time based mix.

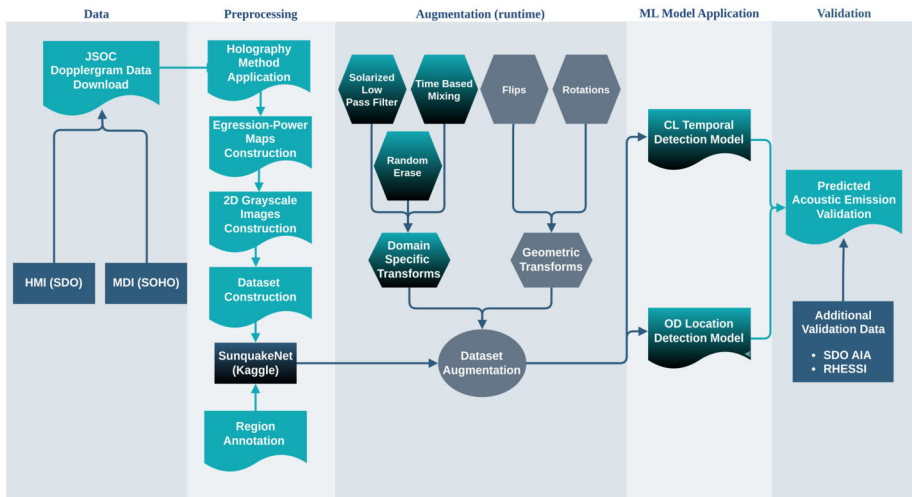


Figure 17 Process diagram of proposed solution. The *gray-labeled* augmentations represent standard ML approaches, while the others depict custom transforms that were found to improve the sunquake prediction model.

- Dissimilar embeddings for sunquake start and end times: Both discussed CL models produce embeddings that have a large cosine distance towards their vicinity in the case of sunquake-transition frames, meaning that they are adept at predicting sunquake start and end frames. This is facilitated by the custom time-based-mixing transform.
- Temporal and spatial components: Our CL-trained models tend to predict a high number of apparently FP detections, compelling us to enhance temporal CL predictions with OD-provided locations, to provide ML-explainability. We found correlations between weaker acoustic-emission sources and solar eruptions accompanied by high-energy X-ray emission with spectral class $> C1.0$ in AIA and RHESSI data. Two such examples are discussed in depth, leading us to qualitatively hypothesize that our approach might be usable to detect weaker acoustic manifestations than previously possible. We stress that these two tentative detections could not be sufficiently validated using helioseismology. A statistically meaningful analysis will be pursued in the future for a quantitative confirmation.
- Autoencoder approach limitations: No autoencoder-based approach, regardless of complexity, proved usable for our dataset. On a dataset-by-dataset basis, basic clusterization of sunquake-positive signals could be obtained for multiple sunquake datacubes.
- Custom augmentations as means of explainability: Although the custom solarized low pass filter augmentation significantly improves both CL models, we find it is not suitable for use as a stand-alone tool for ML-output explainability.
- Unmet minimum sunquake-duration criteria: The model described in Section 4.1.2 is also sensitive in predicting shorter duration signals, be they acoustic in nature or not. This is explained by configuring the time based mixing transform only on the series of three frames. In the future, we will experiment with increasing this window. This was not currently possible due to unfeasible computational costs.
- Unaddressed spurious correlations: False detections such as strong shifts in the gradient intensity in the AR's shadow, which are essentially noise, indicate that the models do not yet distinguish only true sunquake signatures. We believe a more fine-grained separation in signatures is needed to encourage the models to produce sunquake representations that

are dissimilar to their unfounded lookalikes. The additional representations might also prove to be of significant interest.

- **Impact of noise:** Noise has long been the biggest enemy of astronomers and data scientists. While the ML techniques and custom transforms that we introduced facilitate learning, they are not reliable for studying the effects of noise in their current state, as they all have a chance of obscuring the sunquake information. Moreover, conventional noise-reduction methods such as mean or Gaussian filters are also not reliable because noise patterns in regression-power maps are not distinctive enough from sunquake signatures.

As noise continues to impose several challenges, additional pre-learning steps for performing noise reduction are paramount in improving the results and learning complexity. For this, we plan to repurpose the autoencoder model such that the respective model learns AR shapes, instead of sunquake signatures. This will also allow autoencoder training on additional datasets that do not necessarily contain sunquakes. Following this, a pre-learning step can possibly be applied to the input image to blacken out the area outside the AR shadow, obtained from the model's reconstruction, so that the latent space needed for the main CL training may be largely reduced.

Appendix A: AutoEncoder Approaches

As expected, MLPs, simple CNNs, and transfer learning from ImageNet, which were first applied, failed to achieve good performance due to the complexity of the data, the highly imbalanced property and the transferred-feature relevance.

For this reason, before attempting to apply CL and OD techniques described in this article, we decided to use representation learning in the form of unsupervised, or self-supervised, autoencoder approaches. For these, a ResNet-6 (He et al., 2016) backbone is chosen, mainly because a more complex backbone would easily overfit given the small dataset size of SC23.

In self-supervised learning, original annotations are disregarded, but task-specific synthetic annotations are obtained from the data such that the learning is still a supervised one.

Autoencoders are neural networks that reduce the dimensionality of the data and learn to reconstruct the data from the obtained encoding to a representation that is as similar to the original data as possible.

This category of neural networks was first introduced by Hinton and Salakhutdinov (2006) as an improvement to dimensionality reduction techniques such as principal component analysis (Maćkiewicz and Ratajczak, 1993). Autoencoders were later proposed as generative models (Kingma and Welling, 2014, 2019), and further improved by Zhang et al. (2019) by reducing training impediments given by inconsistencies between the data space and the latent space. We decided to choose this approach as it deals well with noisy data, noise being the main impediment for our holography datasets.

A.1 Reduced AutoEncoder

First, a plain autoencoder model with convolutional layers is implemented, with a latent space of size varying from 32 to 4096, with mean squared error loss. The reconstructions for data samples are manually validated to confirm sunquake information is preserved. Due to varying morphological structure between events, reconstructions obtained from latent sizes lower than 1024 do not recover the originals, while larger latent sizes fail in boosting the classification performance, even though they support good reconstructions.

A.2 Variational AutoEncoder

If the autoencoder objective is extended to include and simultaneously optimize both reconstruction and latent-space distribution losses, the latent state for an observation is closer to others but deviates when necessary to describe distinctive features in the input (Kingma and Welling, 2019). This enables balanced latent-state representations of the input data, and provides artificially generated reconstructions that follow the learned distributions. Following Kingma and Welling (2014), we modify our model into a variational autoencoder (VAE), by updating the encoder to also output a distribution in the latent space with explicitly modeled variance.

Although VAE theoretically improves classifications through the learned representation, the experiments performed on our dataset still show weak reconstructions, depicting almost entirely black images. The classification results obtained from the latent representation of the input data also demonstrate that sunquake highlights are not captured in the latent space. During classification, experiments are performed using different loss functions, but even imbalance-specific losses, such as focal loss (Lin et al., 2017), are still incapable of providing reliable results.

Additional different variants of VAE are utilized to find a more suitable design that would i) provide reconstructions focused on the sunquake rather than the AR of the input sample, and ii) is able to capture small variations around or inside the AR where sunquakes usually occur. We take note of Cai, Gao, and Ji (2019), who found that the typically used squared L_2 -reconstruction loss function results in problems such as blurry images, and that it is also known to be sensitive to large noise.

Chen, Chen, and Zhang (2019) replace VAE's L_2 -objective with the log–hyperbolic cosine (log–cosh) loss, which behaves as L_2 at small values and as L_1 at large values. Compared to L_2 , according to Chen, Chen, and Zhang (2019), the log–cosh loss improves the reconstruction without damaging the latent-space optimization, thus automatically keeping a balance between the reconstruction and the generation.

The images in our dataset are noisy, especially in areas outside ARs. Moreover, the sunquake signature represents only a small variation in the data for a short duration of time. Because of this, we believe that an objective that is less sensitive to noise could improve the model's focus.

We find that the log–cosh loss benefits the training process and improves test accuracy by $\approx 25\%$ when compared to the reduced autoencoder (see Table 6). Despite the improvement, sunquakes are still unidentifiable in the reconstructions with the naked eye. The reconstructions are blurry representations of the AR, with little to no detail, as depicted in Figure 18.

A.3 Analysis of the Latent Features Quality

To perform classification on top of the encoded data, various classifiers are trained over the encoding provided by the log–cosh VAE: K-nearest neighbors (KNN) with and without bagging, linear classifier, support vector classifier (SVC) (Cortes and Vapnik, 1995), logistic regression, fine-tuned unfrozen layers of the VAE, stochastic gradient descent, and random forest (Ho, 1995). The results are not satisfactory, but are the first to provide a truly meaningful input. A report of the best results achieved using variational autoencoders as backbones for our models is displayed in Table 6. Although still unreliable for classification, the reconstructions may prove potentially useful for tasks related to masking the AR with the purpose of reducing noise present in the dataset.

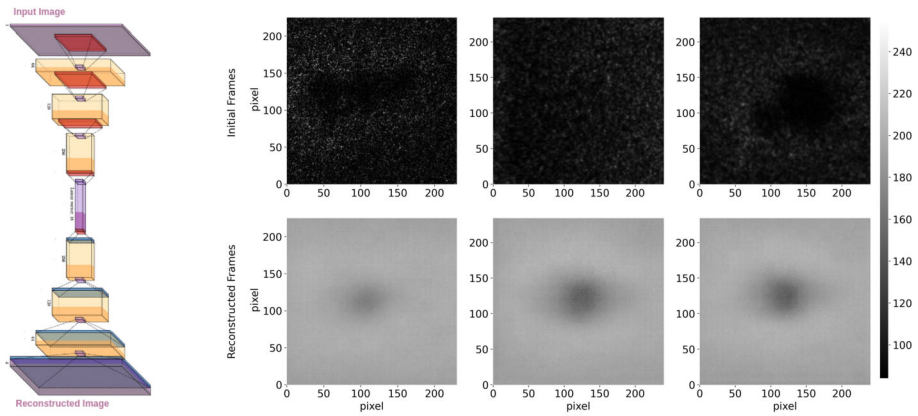


Figure 18 Autoencoder architecture (Xu et al., 2020) (left), Variational autoencoder reconstructions of three randomly chosen eight-bit scaled intensity data samples from the ML prepared dataset (right).

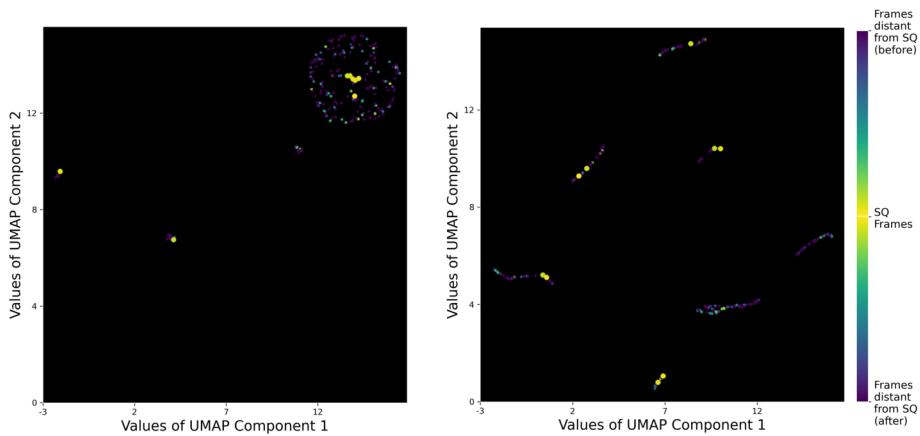


Figure 19 UMAP components constructed from autoencoder-produced embeddings, for all samples in a singular event using 250 (left) and 15 (right) neighbors. Sample count is 256. The colorbar goes symmetrically from yellow for frames corresponding to the sunquake, to purple for frames on either side of the sunquake.

Figure 19 shows UMAPs built from the latent encoding of sunquake samples. On the left panel, we observe that clusters are formed in concentric circles, having the sunquake observations at the center. On the right panel, where fewer neighbors are selected, the clustering is lost. We infer that the observation's background noise and the morphological structure of the AR play a significant role in the autoencoder output, as background counts in the AR are systematically lower, and morphologies are not consistent between the sunquake events. We found these effects to be significant enough to hinder a proper cluster-like classification of the sunquakes by using an autoencoder-based solution with reduced complexity.

It is important to underline that the clustering mentioned above is analyzed per sunquake event. The same analysis performed on the encoding extracted from the entire dataset shows no evident clusters. This supports our hypothesis that the encoding obtained using the mentioned unsupervised techniques is incapable of clustering data points from sunquakes with very different ARs in a two-dimensional setting. However, the few positive results com-

Table 6 Macro average performances of different classifiers over autoencoder-produced embeddings, trained with SMOTE augmentation, on the SC23 test data (171 negative and 26 positive samples). The bold entries represent the highest scores for each metric.

Classifier AE Type	K-NN (Bagging)		SVC (Poly)		Logistic Regression		SGD	
	AE	Log-Cosh VAE	AE	Log-Cosh AE	AE	Log-Cosh VAE	AE	Log-Cosh VAE
Precision	0.50	0.57	0.51	0.76	0.50	0.48	0.50	0.54
Recall	0.46	0.56	0.52	0.59	0.54	0.48	0.50	0.55
F1-Score	0.39	0.56	0.05	0.61	0.39	0.48	0.39	0.54
Accuracy	0.62	0.82	0.05	0.88	0.59	0.78	0.61	0.74
Metric Avg.	0.49	0.62	0.28	0.71	0.5	0.55	0.5	0.59

binized with an increased precision (see Table 6) when classifying the data show that some of the relevant aspects in the data are captured by the encoder, especially considering that no recurrent techniques are used in this approach.

We present the less-satisfactory performance of the autoencoder-based methods discussed above, in Table 6.

Appendix B: Self-Supervised and Supervised Contrastive Learning Approaches

This section describes the experiments performed using self-supervised and supervised contrastive learning approaches that led to establishing the two-step learning approach described in Section 3.1.

B.1 Self-Supervised Contrastive Learning

The framework proposed by Chen et al. (2020) is followed, where, for all the images in each data batch, two augmentations are drawn out so that a double-sized batch is obtained. Specific details on what augmentations were used in our experiments are presented in Section 3.2. Batch data are then passed through the CL model, where SimCLR loss (Chen et al., 2020) is used. The loss is defined as:

$$l_{i,j} = -\log \frac{e^{\frac{\text{sim}(z_i z_j)}{\tau}}}{\sum_{k=1}^{2N} \text{mask}_{[k \neq i]} e^{\frac{\text{sim}(z_i z_k)}{\tau}}}. \quad (4)$$

In the equation above, (i, j) represents a positive pair of samples, mask is an indicator function evaluating to 1 if $k \neq i$, z represents the embedding obtained from passing the input through the model, sim represents cosine similarity. τ is a parameter, called the temperature in ML terminology, used to regularize the contribution of representations (Chen et al., 2020). This does not correspond to a physical temperature.

At low temperatures, smaller distances are favored, which leads to better capturing of minor differences, but making training more difficult since numerical instability increases (Khosla et al., 2020). In the experiments performed on our data, a temperature factor lower than $\tau = 0.07$ is not usable due to this instability, and generally values of $\tau \in [0.07, 0.1]$ work best.

We experiment with dimension size for the encoding (latent representation resulting from the backbone) and the projection (latent representation resulting from the contrastive head) varying from 20 and up to 2048 features.

Because the discussed approach is self-supervised, labels are not considered. Classification is performed using a weighted cross-entropy loss after the contrastive training is finalized. We experiment with freezing network layers in the backbone network during classification. This process prevents any updates to the weights of the frozen layers of the network and assures that changes inferring from classification are not so drastic as to erase relevant information learned by the network during CL.

When classifying the reduced representations, we experiment with using both the encoding and the projection. On average, results of this approach are similar to those of the autoencoder methods. Seeing that this type of approach is not able to capture sunquake signatures in the latent representation, we shift to the recently emerging supervised CL (Khosla et al., 2020).

B.2 Supervised Contrastive Learning

The implementation of a supervised-contrastive model as compared to the self-supervised one differs only in the loss function. The supervised-contrastive loss function (Khosla et al., 2020), written in the same manner as Equation 5 is defined for each anchor i as:

$$l_i = \sum_{i \in I} \frac{-1}{|P(i)|} \sum_{p \in P(i)} \log \frac{e^{\frac{\text{sim}(z_i z_p)}{\tau}}}{\sum_{k=1}^{2N} \text{mask}_{[k \neq i]} e^{\frac{\text{sim}(z_i z_k)}{\tau}}}. \quad (5)$$

The meaning of most parameters is maintained as in Equation 4. P represents the set of indices of the positives, which are all of the samples that share the same class as the anchor and their augmentations.

To verify the validity of the implementation of our models, an experiment is performed on a commonly used small-scale ML dataset, CIFAR-10 (Krizhevsky, 2009). The model setup is kept in place, and for augmentations both CIFAR specific augmentations and ours are used. Findings of this experiment show that the model is only capable of convergence after ≈ 100 epochs (top-1 classification accuracy > 90.0). Up to this point, the model is stuck in a local minimum. We deduce that several more epochs are needed to improve convergence on our data, especially because the data structure has a higher complexity, a larger imbalance factor, is noisier, and is larger in size.

Increasing the number of epochs for the supervised model shows a similar behavior in terms of local minima, but the class imbalance factor still posed a great impact on the classification results. A two-step approach combining self-supervised with supervised CL models was found to be robust in capturing sunquake features, as shown in Section 3.1.

Appendix C: Data Augmentation and Sampling Approaches

This section describes all the data-augmentation methods briefly mentioned in Section 3.2, following the categories proposed by (Yang et al., 2022).

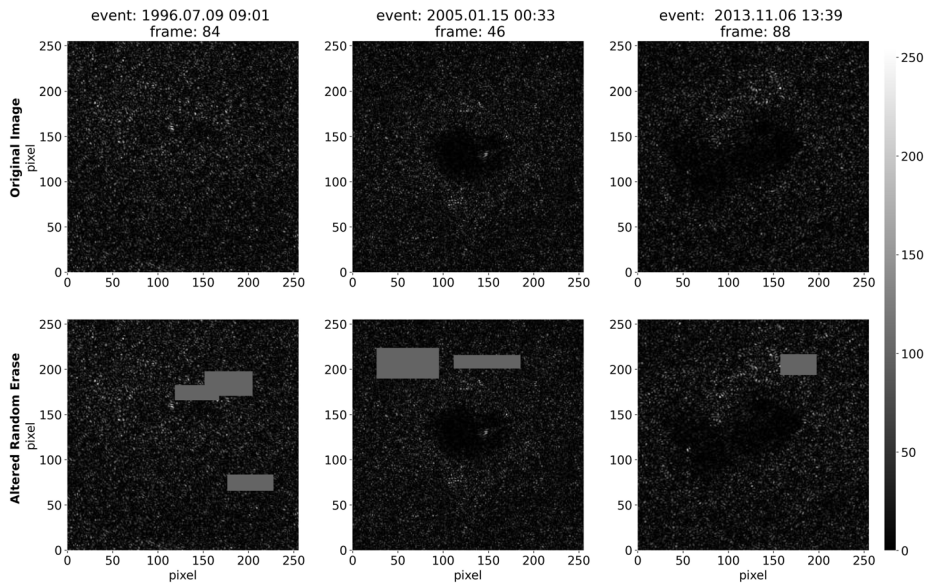


Figure 20 Altered random erase applied to a selection of eight-bit intensity scaled frames from the SC23 and SC24 dataset. Frame indices and events are mentioned in the column titles.

C.1 Geometry

Flips: We apply horizontal and vertical flips as a method of upsampling by adding flipped copies of the positive samples to the dataset. We also apply flip operations at runtime, to ensure that the transforms do not infer external information to the model regarding positive samples.

Crops: The random-crop transform is typically used in most CL approaches. We found it unsuitable for our data since the relevant area for the sunquake signature is very small, and there is a chance of cropping out that specific area. Yang et al. (2022) also argue that this might not be a label-preserving transformation. Adding this augmentation to our data at runtime prevents convergence of supervised methods.

Rotations: We apply right rotation on axes of 90° , 180° , and 270° . Because our data and labels are invariant to rotations, this transformation is usable. As with flip operations, rotations are used as a means to upsample our dataset with the purpose of reducing class imbalance. To make sure that the model does not learn to associate geometric transforms with a positive class, we apply at least the same number of rotated frames to the negative samples at runtime.

Translations: Because the data are not translation invariant, we avoid this category of transforms to preserve the labels.

C.2 Kernel Filters

Gaussian Blur: We experiment with Gaussian blur and argue that it is not always label preserving, hindering the sunquake information for some of the positive samples. We found it to not provide notable improvements in the learning process, and thus we discard this transformation.

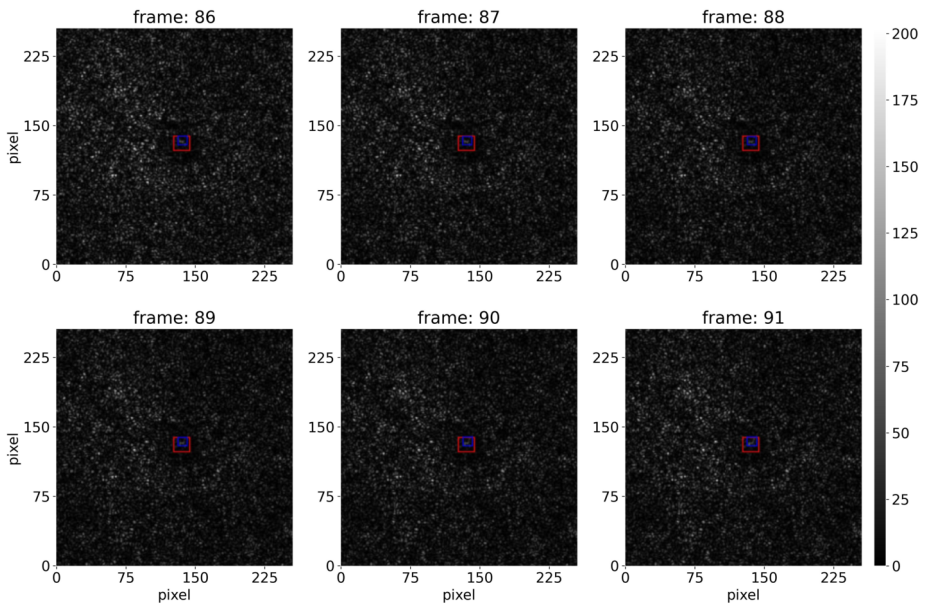


Figure 21 Position of identified OD kernel marked by the *blue box* vs. GT kernels marked by the *red box* for six eight-bit scaled intensity data samples from the event 07 September 2011 22:32 in SC24. IoU varies between 30% and 35% for the given frames.

C.3 Mixing Images

A custom *Custom Time-Based Mixing* augmentation is described in Section 3.2.

C.4 Auto Augment Policies

ImageNet and CIFAR Auto Augment Policies: During the experiment with the CIFAR dataset described in Section 3.1 we apply the auto augment policies generated for imagenet and CIFAR on both the CIFAR-10 dataset and ours, and while the imagenet policy works well on CIFAR-10, when applied to our dataset the results were found to be poor.

C.5 Color Space

As our data are gray-scale, we do not use many color-altering transforms (i.e. color jitter). We experiment with contrast, posterize, brightness, and solarization, and manually review the augmented data to analyze the impact on the sunquake preservation. With the exception below, color-space transforms are ineffective.

We applied a color space-like transform, *Custom Solarized Low Pass Filter*, as described in Section 3.2.

C.6 Occlusion

Random Erasing: This type of augmentation, developed by Zhong et al. (2020) and inspired by the mechanisms of dropout regularization (Yang et al., 2022), is typically used for scenarios where the detection mechanism tends to fail due to occlusions. It forces the model

to learn more relevant features about an image, while preventing overfitting. In typical CL tasks, occlusion is achieved via the random-crop transform. In this case, we aim to shift the model's focus from the AR to the sunquake. In Section 3.2 we describe customizing this type of augmentation to force the network to pay attention to all areas inside the image. We do this by occluding random areas of both the noisy regions and the AR (see Figure 20).

Appendix D: IoU Metric Limitations in Sunquake Detection

Figure 21 highlights the OD prediction boxes and the GT boxes for an event in SC24, where the discrepancy between a correct detection and a low IoU discussed in Section 3.4 is exemplified. For this event, despite a valid signature identification, the average OD IoU for all frames is small (22.3%), and individual IoUs for the frames depicted in the figure are as follows: 34.1%, 32.7%, 30.8%, 33%, 34.2%, and 36%.

List of Acronyms

AIA	<i>Atmospheric Imaging Assembly</i>
AR	Active Region
CL	Contrastive Learning
CNN	Convolutional Neural Networks
FN	False Negative
FP	False Positive
GT	Ground Truth
HMI	<i>Helioseismic and Magnetic Imager</i>
IoU	Intersection over Union
MDI	<i>Solar and Heliospheric Observatory</i>
ML	Machine Learning
MLP	Multilayer Perceptron
OD	Object Detection
PCA	Principal Component Analysis
RBF	Radial Basis Function
R-CNN	Region-Based Convolutional Neural Networks
RHESSI	<i>Ramaty High Energy Solar Spectroscopic Image</i>
SC	Solar Cycle
SDO	<i>Solar Dynamics Observatory</i>
SGD	Stochastic Gradient Descent
SMOTE	Synthetic Minority Over-Sampling Technique
SOHO	<i>Solar and Heliospheric Observatory</i>
SQ	Sunquake
SVC	Support Vector Classifier
TN	True Negative
TP	True Positive
UMAP	Uniform Manifold Approximation and Projection
VAE	Variational AutoEncoder

Acknowledgments The authors appreciate the very pertinent comments from the reviewer, which were crucial towards improving this work. The authors thank Benoit Tremblay for the initial pre-submission review of this work. We acknowledge and appreciate the conversations and generous guidance provided by Alina Donea and Charles Lindsey towards pursuing this project. The authors also thank the Technical University

of Cluj-Napoca for providing access to the DGX system, which facilitated the execution of the presented experiments. Raw data and calibration instructions are obtained courtesy of NASA SDO/HMI and SOHO/MDI science teams. SOHO is a mission of international cooperation between ESA and NASA. The authors welcome and appreciate the open data policy of the SDO, SOHO, and RHESI missions. This research used version 4.0.X The SunPy Community et al. (2020) of the SunPy open source software package. This work has made use of NASA's Astrophysics Data System (ADS).

Author contributions V. Mercea and A. Marginean worked on the ML experiments and ML analysis of results, A.R. Paraschiv, D.A. Lacatus and D. Besliu-Ionescu worked on applying the solar acoustic holography methods and on the analysis of results from a heliophysics standpoint. All authors contributed to writing the Abstract and Sections 1 and 5. V. Mercea wrote Sections 2.3, 3, 4.1, 4.2. A.R. Paraschiv wrote Section 2.2. V. Mercea, A.R. Paraschiv, D.A. Lacatus wrote Section 4.3. D. Besliu-Ionescu wrote Section 2.1, D.A. Lacatus prepared Figures 13, 15, 16, and V. Mercea prepared all other figures. All authors contributed to the Appendix. All authors analyzed and discussed the results. All authors reviewed the manuscript.

Funding A.R. Paraschiv and D.A. Lacatus are funded by the High Altitude Observatory of the National Center for Atmospheric Research, facilities that are sponsored by the National Science Foundation under cooperative agreement No. 1852977. D. Besliu-Ionescu was supported by a grant of the Romanian Ministry of Research, Innovation and Digitalization, CNCS/CCCDI-UEFISCDI, project number PN-III-P2-2.1-SOL-2021-2-0192, within PNCDI III.

Data Availability The datasets generated and/or analyzed during the current study are available in our kaggle sunquakeNet repository: DOI.

Code Availability The most current models can be provided by us upon reasonable request.

Declarations

Competing interests The authors declare no competing interests.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bengio, Y., Courville, A., Vincent, P.: 2013, Representation learning: a review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1798.
- Besliu-Ionescu, D., Donea, A., Cally, P.: 2017, Current state of seismic emission associated with solar flares. *Sun Geosph.* **12**, 59. [ADS](#).
- Besliu-Ionescu, D., Donea, A., Cally, P., Lindsey, C.: 2012, Web-based comprehensive data archive of seismically active solar flares. In: Mariş, G., Demetrescu, C. (eds.) *Advances in Solar and Solar-Terrestrial Physics*, Research Signpost, Kerala, 31. [ADS](#).
- Cai, L., Gao, H., Ji, S.: 2019, Multi-stage variational auto-encoders for coarse-to-fine image generation. In: Berger-Wolf, T., Chawla, N. (eds.) *Proc. 2019 SIAM Internat. Conf. Data Mining (SDM)*, SIAM, Philadelphia, 630. 978-1-61197-567-3. [DOI](#).
- Cao, Z., Li, X., Feng, Y., Chen, S., Xia, C., Zhao, L.: 2021, ContrastNet: unsupervised feature learning by autoencoder and prototypical contrastive learning for hyperspectral imagery classification. *Neurocomputing* **460**, 71. [DOI](#).
- Chen, P., Chen, G., Zhang, S.: 2019, Log Hyperbolic Cosine Loss Improves Variational Auto-Encoder. openreview.net/forum?id=rkgjlvC9Ym.

- Chen, R., Zhao, J.: 2021, A possible selection rule for flares causing sunquakes. *Astrophys. J.* **908**, 182. DOI. ADS.
- Chen, T., Kornblith, S., Norouzi, M., Hinton, G.: 2020, A simple framework for contrastive learning of visual representations. In: Daumé, H., Singh, A. (eds.) *Proc. of the 37th Internat. Conf. on Machine Learn., Proc. of Machine Learning Research* **119**, 1597. PMLR, Virtual. proceedings.mlr.press/v119/chen20j.html.
- Cortes, C., Vapnik, V.: 1995, Support-vector networks. *Mach. Learn.* **20**, 273.
- Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., Fei-Fei, L.: 2009, ImageNet: a large-scale hierarchical image database. In: Essa, I., Kang, S.B., Pollefeys, M. (eds.) *2009 IEEE Conf. on Comp. Vision Patt. Recog.*, IEEE Press, New York, 248. DOI.
- Domingo, V., Fleck, B., Poland, A.I.: 1995, The SOHO mission: an overview. *Solar Phys.* **162**, 1. DOI. ADS.
- Donea, A.: 2011, Seismic transients from flares in solar cycle 23. *Space Sci. Rev.* **158**, 451. DOI. ADS.
- Donea, A.-C., Braun, D.C., Lindsey, C.: 1999, Seismic images of a solar flare. *Astrophys. J. Lett.* **513**, L143. DOI. ADS.
- He, K., Zhang, X., Ren, S., Sun, J.: 2016, Deep residual learning for image recognition. In: Agapito, L., Berg, T., Kosecka, J., Zelnik-Manor, L. (eds.) *2016 IEEE Conf. Comp. Vision Patt. Recog. (CVPR)*, 770.
- Hinton, G.E., Salakhutdinov, R.R.: 2006, Reducing the dimensionality of data with neural networks. *Science* **313**, 504. DOI.
- Ho, T.K.: 1995, Random decision forests. In: Kasturi, R., Lorette, G., Yamamoto, K. (eds.) *Proc. 3rd Internat. Conf. on Doc. Analys. and Recog.* **1**, IEEE Press, New York, 278.
- Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: 2017, Densely connected convolutional networks. In: Rehg, J., Liu, Y., Wu, Y., Taylor, C. (eds.) *2017 IEEE Conf. Comp. Vision Patt. Recog. (CVPR)*, IEEE Press, New York, 2261. DOI.
- Hurfurd, G.J., Schmahl, E.J., Schwartz, R.A., Conway, A.J., Aschwanden, M.J., Csillaghy, A., Dennis, B.R., Johns-Krull, C., Krucker, S., Lin, R.P., McTiernan, J., Metcalf, T.R., Sato, J., Smith, D.M.: 2002, The RHESSI imaging concept. *Solar Phys.* **210**, 61. DOI. ADS.
- Ionescu, D.: 2010, Seismic emissions from solar flares. Dissertation, Faculty of Science, Monash University. DOI.
- Khosla, P., Teterwak, P., Wang, C., Sarna, A., Tian, Y., Isola, P., Maschinot, A., Liu, C., Krishnan, D.: 2020, Supervised contrastive learning. In: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M.F., Lin, H. (eds.) *Adv. Neural Inform. Proc. Sys.* **33**, Curran, New York, 18661. proceedings.neurips.cc/paper/2020/file/d89a66c7c80a29b1bdbab0f2a1a94af8-Paper.pdf.
- Kingma, D.P., Ba, J.: 2014, Adam: a method for stochastic optimization. [arXiv](https://arxiv.org/abs/1412.0441).
- Kingma, D.P., Welling, M.: 2014, Auto-encoding variational bayes. In: Bengio, Y., LeCun, Y. (eds.) *2nd Internat. Conf. Learning Represent., ICLR 2014*, Conference Track Proceedings, 2014. DOI.
- Kingma, D.P., Welling, M.: 2019, An introduction to variational autoencoders. *Found. Trends Mach. Learn.* **12**, 307.
- Kosovichev, A.G.: 2011, Helioseismic response to the X2.2 solar flare of 2011 February 15. *Astrophys. J. Lett.* **734**, L15. DOI. ADS.
- Kosovichev, A.G., Zharkova, V.V.: 1998, X-ray flare sparks quake inside the Sun. *Nature* **393**, 317. ADS.
- Krizhevsky, A.: 2009, Learning multiple layers of features from tiny images. Technical report, University of Toronto.
- Lemaître, G., Nogueira, F., Aridas, C.K.: 2017, Imbalanced-learn: a Python toolbox to tackle the curse of imbalanced datasets in machine learning. *J. Mach. Learn. Res.* **18**. jmlr.org/papers/v18/16-365.html.
- Lemen, J.R., Title, A.M., Akin, D.J., Boerner, P.F., Chou, C., Drake, J.F., Duncan, D.W., Edwards, C.G., Friedlaender, F.M., Heyman, G.F., Hurlburt, N.E., Katz, N.L., Kushner, G.D., Levay, M., Lindgren, R.W., Mathur, D.P., McFeaters, E.L., Mitchell, S., Rehse, R.A., Schrijver, C.J., Springer, L.A., Stern, R.A., Tarbell, T.D., Wuelser, J.-P., Wolfson, C.J., Yanari, C., Bookbinder, J.A., Chemists, P.N., Caldwell, D., Deluca, E.E., Gates, R., Golub, L., Park, S., Podgorski, W.A., Bush, R.I., Scherrer, P.H., Gummis, M.A., Smith, P., Auker, G., Jerram, P., Pool, P., Soufli, R., Windt, D.L., Beardsley, S., Clapp, M., Lang, J., Waltham, N.: 2012, The Atmospheric Imaging Assembly (AIA) on the Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 17. DOI. ADS.
- Lin, R.P., Dennis, B.R., Hurfurd, G.J., Smith, D.M., Zehnder, A., Harvey, P.R., Curtis, D.W., Pankow, D., Turin, P., Bester, M., Csillaghy, A., Lewis, M., Madden, N., van Beek, H.F., Appleby, M., Raudorf, T., McTiernan, J., Ramaty, R., Schmahl, E., Schwartz, R., Krucker, S., Abiad, R., Quinn, T., Berg, P., Hashii, M., Sterling, R., Jackson, R., Pratt, R., Campbell, R.D., Malone, D., Landis, D., Barrington-Leigh, C.P., Slassi-Sennou, S., Cork, C., Clark, D., Amato, D., Orwig, L., Boyle, R., Banks, I.S., Shirey, K., Tolbert, A.K., Zarro, D., Snow, F., Thomsen, K., Henneck, R., McHedlishvili, A., Ming, P., Fivian, M., Jordan, J., Wanner, R., Crubb, J., Preble, J., Matrangola, M., Benz, A., Hudson, H., Canfield, R.C., Holman, G.D., Crannell, C., Kosugi, T., Emslie, A.G., Vilmer, N., Brown, J.C., Johns-Krull, C., Aschwanden, M., Metcalf, T., Conway, A.: 2002, The Reuven Ramaty High-Energy Solar Spectroscopic Imager (RHESSI). *Solar Phys.* **210**, 3. DOI. ADS.

- Lin, T.-Y., Goyal, P., Girshick, R., He, K., Dollár, P.: 2017, Focal loss for dense object detection. In: Cucchiara, R., Matsushita, Y., Sebe, N., Soatto, S. (eds.) *2017 IEEE Internat. Conf. Comp. Vis. (ICCV)*, IEEE Press, New York, 2999. DOI.
- Lindsey, C., Braun, D.C.: 2000, Basic principles of solar acoustic holography – (invited review). *Solar Phys.* **192**, 261. ADS.
- Liu, W., Wen, Y., Yu, Z., Yang, M.: 2016, Large-margin softmax loss for convolutional neural networks. [arXiv](#).
- Maćkiewicz, A., Ratajczak, W.: 1993, Principal components analysis (PCA). *Comput. Geosci.* **19**, 303. DOI.
- McInnes, L., Healy, J., Saul, N., Großberger, L.: 2018, UMAP: uniform manifold approximation and projection. *J. Open Sour. Softw.* **3**, 861. DOI.
- Mercea, V., Paraschiv, A.R., Lacatus, D.A., Marginean, A., D., B.: 2022, *SunquakeNet*, Kaggle online dataset. Egression power map images for sunquake events in Solar Cycle 23 & 24. DOI.
- Mumford, S.J., Freij, N., Stansby, D., Christe, S., Ireland, J., Mayer, F., Shih, A.Y., Hughitt, V.K., Ryan, D.F., Liedtke, S., Hayes, L., Pérez-Suárez, D., I., V.K., Chakraborty, P., Inglis, A., Barnes, W., Pattinaik, P., Sipócz, B., Sharma, R., Leonard, A., Hewett, R., Hamilton, A., Manhas, A., MacBride, C., Panda, A., Earnshaw, M., Choudhary, N., Kumar, A., Singh, R., Chanda, P., Haque, M.A., Kirk, M.S., Konge, S., Mueller, M., Srivastava, R., Jain, Y., Bennett, S., Baruah, A., Arbolante, Q., Charlton, M., Maloney, S., Mishra, S., Paul, J.A., Chorley, N., Chouhan, A., Himanshu, Zivadinovic, L., Modi, S., Verma, A., Mason, J.P., Sharma, Y., Naman9639, Bobra, M.G., Manley, L., Rozo, J.I.C., Ivashkevich, K., Chatterjee, A., von Forstner, J.F., Stern, K.A., Bazán, J., Jain, S., Evans, J., Ghosh, S., Malocha, M., Visscher, R.D., Stańczak, D., Singh, R.R., SophieLemos, Verma, S., Airmansmith97, Buddhika, D., Alam, A., Pathak, H., Sharma, S., Agrawal, A., Rideout, J.R., Park, J., Bates, M., Mishra, P., Gieseler, J., Shukla, D., Taylor, G., Dacie, S., Dubey, S., Jacob, Cetusic, G., Reiter, G., Sharma, D., Inchaurreandieta, M., Goel, D., Bray, E.M., Meszaros, T., Sidhu, S., Russell, W., Surve, R., Parkhi, U., Zahniy, S., Eigenbrot, A., Robitaille, T., Pandey, A., Price-Whelan, A., J. A., Chicrala, A., Ankit, Guennou, C., D'Avella, D., Williams, D., Verma, D., Ballew, J., Murphy, N., Lodha, P., Bose, A., Augspurger, T., Krishan, Y., honey, neerajkulk, Ranjan, K., Hill, A., Keşkek, D., Altunian, N., Bhoje, A., Singaravelan, K., Kothari, Y., Molina, C., Agrawal, K., mridulpandey, Nomiya, Y., Streicher, O., Wiedemann, B.M., Mampaey, B., Agarwal, S., Gomillion, R., Gaba, A.S., Letts, J., Habib, I., Dover, F.M., Tollerud, E., Arias, E., Briseno, D.G., Bard, C., Srikanth, S., Stone, B., Jain, S., Kustov, A., Smith, A., Sinha, A., Tang, A., Kannoja, S., Mehrotra, A., Yadav, T., Paul, T., Wilkinson, T.D., Caswell, T.A., Braccia, T., yasintoda, Pereira, T.M.D., Gates, T., platipo, Dang, T.K., W. A., Bankar, V., Kaszynski, A., Wilson, A., Bahuleyan, A., Stevens, A.L., B. A., Shahdadpuri, N., Dedhia, M., Mendero, M., Cheung, M., Mangaonkar, M., Schoentgen, M., Lyes, M.M., Agrawal, Y., resakra, Ghosh, K., Hiware, K., Gyenge, N.G., Chaudhari, K., Krishna, K., Buitrago-Casas, J.C., Qing, J., Mekala, R.R., Wimbish, J., Calixto, J., Das, R., Mishra, R., Sharma, R., Babuschkin, I., Mathur, H., Kumar, G., Verstringe, F., Attie, R., Murray, S.A.: 2022, *SunPy*, Zenodo. DOI.
- Pesnell, W.D., Thompson, B.J., Chamberlin, P.C.: 2012, The Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 3. DOI. ADS.
- Ren, S., He, K., Girshick, R., Sun, J.: 2017, Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **39**, 1137. DOI.
- Scherrer, P.H., Bogart, R.S., Bush, R.I., Hoeksema, J.T., Kosovichev, A.G., Schou, J., Rosenberg, W., Springer, L., Tarbell, T.D., Title, A., Wolfson, C.J., Zayer, I., MDI Engineering Team: 1995, The solar oscillations investigation – Michelson Doppler Imager. *Solar Phys.* **162**, 129. DOI. ADS.
- Schou, J., Scherrer, P.H., Bush, R.I., Wachter, R., Couvidat, S., Rabello-Soares, M.C., Bogart, R.S., Hoeksema, J.T., Liu, Y., Duvall, T.L., Akin, D.J., Allard, B.A., Miles, J.W., Rairden, R., Shine, R.A., Tarbell, T.D., Title, A.M., Wolfson, C.J., Elmores, D.F., Norton, A.A., Tomczyk, S.: 2012, Design and ground calibration of the Helioseismic and Magnetic Imager (HMI) instrument on the Solar Dynamics Observatory (SDO). *Solar Phys.* **275**, 229. DOI. ADS.
- Sharykin, I.N., Kosovichev, A.G.: 2020, Sunquakes of Solar Cycle 24. *Astrophys. J.* **895**, 76. DOI. ADS.
- Sharykin, I.N., Kosovichev, A.G., Zimovets, I.V.: 2015, Energy release and initiation of a sunquake in a C-class flare. *Astrophys. J.* **807**, 102. DOI. ADS.
- Sohn, K.: 2016, Improved deep metric learning with multi-class N-pair loss objective. In: Lee, D., Sugiyama, M., Luxburg, U., Guyon, I., Garnett, R. (eds.) *Adv. in Neural Inform. Proc. Sys.* **29**, Curran, New York, 1849. [proceedings.neurips.cc/paper/2016/file/6b180037abbebea991d8b1232f8a8ca9-Paper.pdf](#).
- The SunPy Community, Barnes, W.T., Bobra, M.G., Christe, S.D., Freij, N., Hayes, L.A., Ireland, J., Mumford, S., Perez-Suarez, D., Ryan, D.F., Shih, A.Y., Chanda, P., Glogowski, K., Hewett, R., Hughitt, V.K., Hill, A., Hiware, K., Inglis, A., Kirk, M.S.F., Konge, S., Mason, J.P., Maloney, S.A., Murray, S.A., Panda, A., Park, J., Pereira, T.M.D., Reardon, K., Savage, S., Sipócz, B.M., Stansby, D., Jain, Y., Taylor, G., Yadav, T., Rajul, Dang, T.K.: 2020, The SunPy project: open source development and status of the version 1.0 core package. *Astrophys. J.* **890**, 68. DOI.

- Weinberger, K.Q., Blitzer, J., Saul, L.: 2005, Distance metric learning for large margin nearest neighbor classification. In: Weiss, Y., Schölkopf, B., Platt, J. (eds.) *Adv. Neural Inform. Proc. Sys.* **18**, 1473, MIT Press, Massachusetts. [DOI](#).
- Xu, F., Cervone, G., Franch, G., Salvador, M.: 2020, Multiple geometry atmospheric correction for image spectroscopy using deep learning. *J. Appl. Remote Sens.* **14**, 1. [DOI](#).
- Yang, S., Xiao, W., Zhang, M., Guo, S., Zhao, J., Shen, F.: 2022, Image data augmentation for deep learning: a survey. [DOI](#).
- Zhang, Z., Zhang, R., Li, Z., Bengio, Y., Paull, L.: 2019, Perceptual generative autoencoders. *CoRR*, [arXiv](#).
- Zhong, Z., Zheng, L., Kang, G., Li, S., Yang, Y.: 2020, Random erasing data augmentation. In: Vincent Conitzer, F.S. (ed.) *The Thirty-Fourth AAAI Conf. Artificial Intelligence*, AAAI Press, New York, 13001.
- Zhong, Z., Cui, J., Lo, E., Li, Z., Sun, J., Jia, J.: 2022, Rebalanced Siamese contrastive mining for long-tailed recognition. [arXiv](#).

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.