Stack-CNN algorithm: A new approach for the detection of space objects

Stack-CNN algorithm: a new approach for the detection of space objects

A. Montanaro[c,2,*], T. Ebisuzaki[b], M. Bertaina[a,1,*]

[a] *Dipartimento di Fisica, Università degli studi di Torino - 10125 Torino, Italy*
[b] *Computational Astrophyiscs Laboratory, RIKEN - 351-0198 Wakoshi, Japan*
[c] *Politecnico di Torino - Department of Electronics and Telecommunications, Italy*

Abstract

We present a new trigger algorithm combining a stacking procedure and a convolutional neural network that could be applied to any space object moving linearly or with a known trajectory in the field of view of a telescope. This includes the detection of high velocity fragmentation debris in orbit. A possible implementation is as trigger system for an orbiting Space Debris remediation system. The algorithm was initially developed as offline system for the Multiwavelength Imaging New Instrument for the Extreme Universe Space Observatory (Mini-EUSO), on the International Space Station. We evaluated the performance of the algorithm on simulated data and compared it with those obtained by means of a more conventional trigger algorithm. Results indicate that this method would allow to recognise signals with $\sim$1% Signal over Background Ratio (SBR) on poissonian random fluctuations with a negligible fake trigger rate. Such promising results lead us to not only consider this technique as an online trigger system, but also as an offline method for searching moving signals and their characteristics (speed and direction). More generally, any kind of telescope (on the ground or in space) such as those used for space debris, meteors monitoring or cosmic ray science, could benefit from this automatized technique. The content of this paper is part of the recent Italian patent proposal submitted by the authors (patent application number: 102021000009845).

Keywords: convolutional neural network; space debris; trigger algorithm

*Corresponding author, work done at Università degli studi di Torino, Italy
[1] marioedoardo.bertaina@unito.it
[2] antonio.montanaro@polito.it

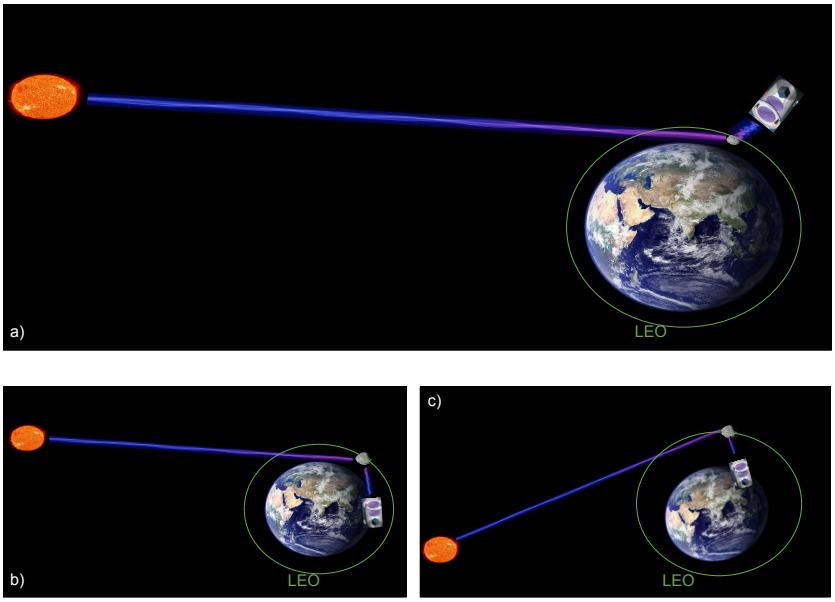


Figure 1: a) Conceptual figure for the SD detection in LEO by a Mini-EUSO-like telescope. It detects the reflected UV light (shown pictorially as a blue-violet light wave) of SD illuminated by the Sun; note that even the telescope itself is illuminated; b) and c) show two other possible configurations that could be more beneficial for SD detection: in b) the detector is positioned along the tangent of its orbit allowing itself to be in penumbra while the LEO SD are illuminated, instead in c) the detector is rotated by 180° and can detect objects in LEO above it. The size of objects is not to scale.

1. Introduction

STACK-CNN is the acronym of STACKing method plus Convolutional Neural Network. It is a completely new detection system that combines two existing algorithms in a specific way. The Stacking Method (SM) was first proposed by Yanagisawa (Yanagisawa, 2003) for Space Debris (SD) detection and independently developed as part of the trigger system for the Extreme Universe Space Observatory JEM-EUSO (Bertaina, 2007).

SM produces a stacked image, that is a sum-image made by overlapping many frames shifted by one or more pixels, depending on the speed and the (opposite) direction that an object (or a particle) can have in the Field of View (FoV) of a telescope. If the SM matches exactly the speed and the

direction of the object, the resulting summed image, or stacked image, is made by a single brighter spot at the initial position of the object. This enhances the Signal over Background Ratio (SBR) of the stacked images with respect to the SBR of the single images by a factor of $\sqrt{n}$ where n is the number of single images.

However, since the specific characteristic of the object is not known a priori, the SM produces a lot of combinations according to all possible configurations. In this scenario, distinguishing the right ones from the wrong ones requires some decision algorithm. If in the past such algorithms exploited the SBR enhancement for recognition of a right combination (leading however to not so high performances), today we can exploit more robust and reliable decision algorithms from computer vision, e.g. Convolutional Neural Networks (CNNs).

CNN is one of the most famous and used Neural Network (NN). It finds application especially in computer vision: image classification, video analysis, anomaly detection, drug discovery and so on. Since its inception, physicists discovered its utility in astronomy, for example for classification of galaxies. Today, whenever there are images or video, like those recorded by telescopes, CNN, and more in general Machine Learning (ML), can give a fundamental contribution to their study. CNNs are ancillaries of the first famous CNN from Le Cun's proposal, LeNet-5 (LeCun, 1998), which was first applied to handwritten digits classification. In the last years, notable improvements concerning new optimizers in the learning phase, e.g. Adadelta (Adadelta, 2012), used in this work, have been introduced, and the automatization of training (advanced backpropagation and available hardware accelerators) on open source platforms have made ML accessible to different scientific communities, including the physics' one. A CNN of this kind is implemented in the STACK-CNN algorithm with suitable adaptations, we refer to subsection 5.2 for more details.

In addition to describing the whole method in detail, we consider one of the many applications of STACK-CNN, i.e. SD detection, keeping in mind a possible implementation on board of new remediation systems.

We analyze SD detection since it has become an urgent problem in recent years and many space agencies are trying to find new tracking systems assembled with instruments to de-orbit or capture as much as possible SD. Among these, a remediation system comprised of a super-wide field-of-view telescope (EUSO) and a novel high-efficiency fibre-based laser system (CAN) can become a feasible solution (Ebisuzaki, 2015).

Such a system could benefit from the STACK-CNN as a trigger system because it is fast, simple to implement and can be mounted on a Field Programmable Gate Unit (FPGA).

As a first step in this direction, the Mini-EUSO detector (Capel, 2018) on board the International Space Station (ISS) is used to make a proof of principle of the detection strategy and possibly tracking new SD. This motivates us to adapt our STACK-CNN for Mini-EUSO data, giving proof of its detection ability and encouraging us to investigate it even as offline data analysis for Mini-EUSO.

Then we compare the method to a standard trigger system developed in the framework of cosmic ray science and adapted for SD, and we show better performances for the STACK-CNN.

This paper is structured in the following way. Section 2 explains what SD are and what risks they can carry. Section 3 describes the observational principle of space debris of the employed system detailing the Mini-EUSO configuration. Section 4 describes the simulation approach developed to test the trigger performance and the conventional method. Section 5 describes the STACK-CNN method. Section 6 presents the results compared to a conventional trigger. A discussion of the results and the conclusions are reported in Section 7.

2. Space Debris

Over the last 60 years, since man began to explore space, several thousand tons of satellites and rockets have been launched. About 7520 are still in orbit according to the last updated informations by ESA (ESA, 2021), of which only 60% are still functioning. The remaining part is catalogued as SD.

Unfortunately these are not the only derelict objects in space, because, during the years, explosions, collisions and other anomalous events (for a total estimated number of 570 events according to (ESA, 2021)) occured in the Earth's orbits, causing chain reactions that produced new space debris of different size and material.

Today the number of debris regularly tracked by Space Surveillance Networks are about 29,240, for a total mass of thousands of tons.

These objects travel at high speeds, of the order of 7 - 9 km/s near Low Earth Orbit (LEO), and can collide with spacecrafts such as the ISS or other manned or unmanned spacecrafts, damaging them and in turn producing new debris.

The largest SD, such as inactive satellites, are easily catalogued and monitored to space safety, while the smallest ones, in the order of millimeter size, although invisible, do not damage operating satellites thanks to specific shields which absorb possible collisions.

A more prominent danger is due to cm-sized space objects. Travelling at high speed, they act as projectiles and can damage solar panels or the main structure of satellites. Moreover, since these fragments are very small and the reflected light is very faint, their detection and monitoring through albedo is rather challenging. Indeed, the exact number is not known, but using statistical models, ESA estimates about 1 million of objects of size between 1 cm and 10 cm. This number will grow over the years as the number of space missions increases.

3. Observation principle of space debris and the Mini-EUSO application

SD themselves do not emit light, but an instrument can detect their reflected light when the SD is illuminated by a laser, by the Sun (see Fig. 1 a)) or by the Moonlight, a phenomenon known as albedo. In such a way, SD appear as tracks crossing the FoV of the focal plane of a detector, enabling the detection and tracking of the object.

In the case of Mini-EUSO the optimal configurations to find this kind of objects are two:

- At sunrise or sunset when the earth is still in umbra while the high atmosphere is already illuminated by the Sun. In this case Mini-EUSO, by looking towards the Earth, can see all the objects in LEO below it, as in Fig. 1 a), provided that the instrument is shielded from direct sun-light;

- With the ISS turned by 90° or 180°, showed in Fig. 1 b) and c) respectively, and the Sun shines from the back to avoid direct sunlight.

Mini-EUSO (Multiwavelength Imaging New Instrument for the Extreme Universe Space Observatory or "UV atmosphere" in the Russian Space Program) is a telescope operating in the UV range (290 - 430 nm) with a square field of view of ∼44° and a ground resolution of ∼6 km (Capel, 2018). It is also equipped by ancillary cameras (NIR and visible), but these are not

needed for the scope of the paper. Only images acquired in UV waveband are considered.

Mini-EUSO was brought to the ISS by the uncrewed Soyuz MS-14, on August 22, 2019 and installed for the first time on the nadir-facing UV transparent window in the Russian Zvezda module of ISS on October 7. Since then, it has been taking data periodically, with installations occurring every couple of weeks on average. The instrument is expected to operate for at least three years. The scientific objectives of Mini-EUSO include, among others, the study of the exposure for the space-based observation of ultra-high-energy cosmic rays, the UV mapping of the Earth, the detection of meteors and space debris, the observation of Transient Luminous Events and bioluminescence, as well as the search for strange quark matter. Examples of the various phenomena observed in the first months of operations can be found in (Bacholle, 2020). The optical system consists of two Fresnel lenses with a diameter of 25 cm. The focal surface, or Photon Detector Module (PDM), consists of 36 MultiAnode Photomultipliers (MAPMTs) tubes, 64 pixels each from Hamamatsu, capable of single photon detection. The read-out is handled by ASICs in frames of 2.5 μs, named as Gate Time Unit (GTU) in the following. Single photon discrimination is 6 ns. Data are then processed by a Zynq based FPGA board which implements a multiple level triggering, allowing the measurement of triggered UV transients for 128 frames at time scales of both 2.5 μs and 320 μs. An untriggered acquisition mode with 40 ms frames performs continuous data taking (Belov, 2018); this is called Integrated Gate Time Unit (IGTU) in the following. Please note that IGTU = $128 \cdot 128 \cdot$GTU. This is the acquisition mode considered for the detection of SD.

4. Simulation studies and standard trigger results

In order to study the performance of Mini-EUSO detector in recognizing the presence of SD in the FoV at sunset or sunrise by means of the STACK-CNN algorithm, we used the EUSO Simulation and Analysis Framework (ESAF) (Berat, 2010). ESAF is a package that provides an end-to-end simulation of the phenomenon from the light source emission, the propagation through the environment, to the simulation of the detector response and its reconstruction algorithms.

The Mini-EUSO configuration is implemented in ESAF and it includes the simulation of a point-like moving light source. Fig. 2 shows an example

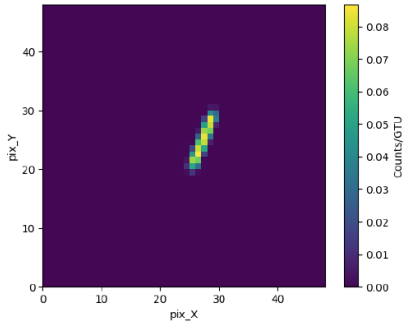


Figure 2: Example of a simulated SD track using ESAF software.

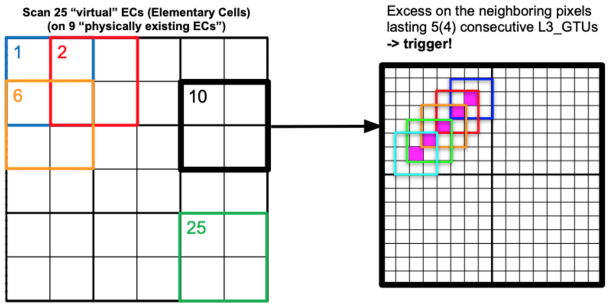


Figure 3: Illustrative sketch of the standard method used as comparison for our method. Figure taken from (Miyamoto, 2019).

of a track generated on the PDM of Mini-EUSO by the simulated source.

The simulations with ESAF allow us to have a benchmark to systematically develop and test different detection strategies, by means of trigger algorithms. For this reason, our proposed method is compared to a standard technique used for Mini-EUSO data (Miyamoto, 2019). In such approach, 25 'virtual' Elementary Cells (ECs) are defined while the trigger scans the entire PDM and looks for an excess in neighboring pixels lasting 5 consecutive IGTUs of 40 ms. One EC consists of 4 MAPMTs. Neighbouring ECs overlap vertically or horizontally by 2 PMTs, or diagonally by 1 PMT. With a threshold on pixel counts of 3σ above the average background calculated every IGTU in the pixel (μ_{pix}), $\mu_{pix} + 3\sigma_{bkg}$, the fake trigger rate is sufficiently low (order of 10^{-5} - 10^{-6} Hz) . A sketch of this method is depicted in Fig. 3. More details can be found in (Miyamoto, 2019).

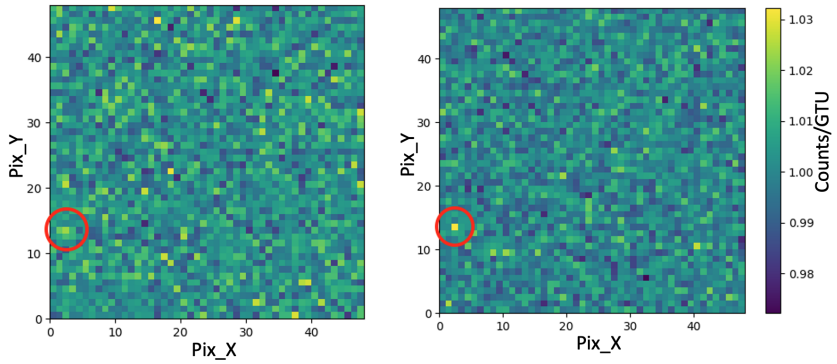


Figure 4: Example of a simulated SD moving from one GTU to another in Poisson background condition. The scale refers to both images.

To estimate the performance of the algorithms, we use the SBR and the Signal over Noise Ratio (SNR) values. For clarification, we define their relation here. The SBR is given by the ratio between the signal integrated over $128 \cdot 128$ GTUs of 2.5 μs and the average of the background integrated over the same time unit. Instead, the SNR has the same numerator of the SBR while its denominator is given by the standard deviation of the background in the same time unit.

The background is modeled by a Poisson distribution with an average of about 1 count/GTU (i.e. one photon count for one GTU of 2.5 μs), in agreement with typical Mini-EUSO findings (Bacholle, 2020).

Under these assumptions, the relation between SBR and SNR integrated over $128 \cdot 128$ GTUs of 2.5 μs is SNR = 128·SBR.

Despite the good performance of the standard trigger techniques that allows to detect SD at SBR levels around 3%, we challenged the possibility to push further the detection, by introducing the STACK-CNN method.

An example of faint SD (below a SBR of 3%) moving in Poisson background can be seen in Fig. 4.

5. The STACK-CNN trigger algorithm

5.1. Stacking Method

In order to explain the method we consider a SD that has a fixed speed $|\overrightarrow{v}|$ and direction θ. For simplicity the debris has only an horizontal velocity

($v_z = 0$) and starts at position (x_0,y_0,h), where $h = 0$ km means at ground level.

At height h, the size of the FoV of one pixel l_p is calculated by rescaling the size of the pixel FoV at ground l_g (∼6 km) by the distance between the height of the detector and of the debris compared to ground level. Naming α the aperture angle of one single pixel:

$$l_p = (400 - h) \cdot tan(\alpha) = \frac{400 - h}{400} \cdot l_g \tag{1}$$

This is useful for rescaling the spatial resolution of a single pixel at the different heights where space objects can be found.

When the acquisition starts, the detector stores the first $n + 1$ images starting from IGTU(t_0) to IGTU(t_n). The STACK-CNN trigger system acquires and processes data continuously. For sake of simplicity, we define in the following as IGTU(t_0) the one in which a SD appears in the FoV of the telescope. We also assume that it lasts until IGTU(t_n), for a total of $n + 1$ IGTUs.

Naming I(t_i) the 48·48 image at IGTU(t_i) and Δt the time difference between two IGTUs, the Stacking Method consists of two iterative steps, a Shift and an Add operation, that iteratively act over all images in temporal order.

The image I(t_1) is shifted by dx along the x-axis and by dy along the y-axis according to the SD motion, but in opposite directions:

$$\begin{aligned} dx &= |\overrightarrow{v}| \cdot \Delta t \cdot cos(-\theta) \\ dy &= |\overrightarrow{v}| \cdot \Delta t \cdot sin(-\theta) \end{aligned} \tag{2}$$

Assuming the SD starts in the center of a pixel, if dx (dy) is smaller than $l_p/2$, then the image is not shifted (meaning that the debris moves within the pixel), otherwise it is shifted by one or two pixels depending on whether dx (dy) is greater than $l_p/2$ or $l_p/2 + l_p$. Once $I(t_1)$ is shifted, it is added to $I(t_0)$ to generate a summed image ΣI_1. The procedure is iteratively applied to all the considered $n + 1$ frames. After the last step, the value of ΣI_n in a pixel (x, y) follows the equation:

$$\Sigma I_n(x, y) = \Sigma_{k=0,n} I(x + k \cdot dx, y + k \cdot dy, t_k) \tag{3}$$

where k indicates by how much an image $I(t_i)$ has to be shifted according to the time delay from t_0 (since $t_k = k \cdot t_0$).

ΣI_n is the stacked image and its SNR is increased by $\sqrt{n+1}$ compared to a single image due to the coherent signal overimposed on random Poisson fluctuations.

It is important to underline that the specific parameters of a SD are not known a priori and the stacking method has to produce all possible combinations; there will be one or more that match fairly well the SD motion while the remaining ones will be wrong and the signal will not be coherent. Once the right combination is found, SM gives speed, direction and the starting pixel position of the SD according to the chosen reference frame. This is an important aspect, because the reference frame fixes a height and a corresponding pixel size that leads to a particular horizontal speed. The number of combinations is a significant aspect too. All possible combinations could lead to a huge amount of images and this number grows with the spatial resolution of the telescope.

These aspects motivate us to explore in the following a Convolutional Neural Network as a decision making algorithm to recognize the right combinations and discard the wrong ones.

5.2. Convolutional Neural Network

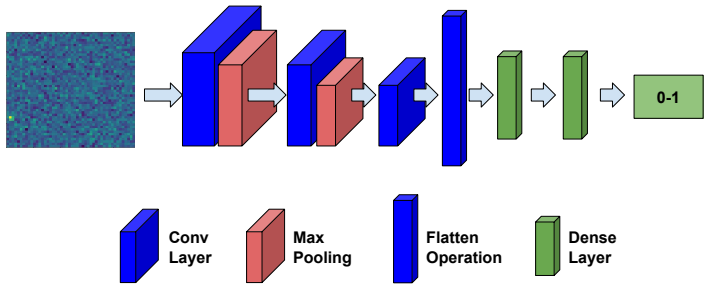
From the revolution of Deep Learning in 2012 (Parloff, 2018), the CNN has received a lot of attention and consequently many researchers made efforts to drive such architecture at performances similar to humans in many computer vision problems and other tasks.

Here we will describe a CNN with basic components, as in its initial versions, since this is the simple model used in the analysis. The motivations for such a choice are provided later on in this section. More details about CNN and more in general about Deep Learning can be found in (Aggarwal, 2018).

An illustration of our architecture is provided in Fig. 5.

The CNN is an Artificial Neural Network composed mainly by three types of blocks each one repeated in a specific order and sequentially stacked. This structure allows to create high-level concepts of the input (represented by the feature maps) with different levels of abstraction according to the depth of each layer. At the end of the network such features are projected into one single unit that, in our binary classification problem, tells us if one stacked image is good or not.

The main blocks composing the CNN are the convolutional, the pooling (or downsample) and the dense (or fully connected) layers.



Input	Conv layer	Max pooling	Conv layer	Max pooling	Conv layer	Flatten	Dense layer	Dense layer	Dense layer
(48,48)	(10,4x4)	(2x2)	(5,4x4)	(2x2)	(1,4x4)	(144)	(144,72)	(144,72)	(72,1)
# param	170	0	805	0	81	0	10440	5256	73

Figure 5: The CNN architecture employed after the Stacking Method. Top image shows an illustrative view of the CNN with its main components; bottom image shows a table where each component is specified with its characteristic and number of parameters. In convolutional layers, (F, $k \times k$) means applying F filters with a kernel size of k, i.e. the window to convolve. The same is true for max pooling where (2×2) is the tile where the maximum is computed. For Dense layers, (m,n) stands for the affine transformation that get the m-dimensional vector and transform it to n-dimensional vector (with n<m). In particular the pooling layer, as the flatten one, has no learning parameters since they change the input without additional parameters. The total number of parameters is 16,825.

The convolutional layer (blue blocks in Fig. 5), as the main block of the network, performs a weighted sum of one pixel and its neighborhood (known in biology as receptive field and with a size specified by the user). Then the same weighted sum is applied to the other pixels until the whole image is processed. Such a shared-weight operation allows to have few parameters in the architecture and makes it equivariant to translations.

A pooling layer (red blocks in Fig. 5) is applied after each convolutional layer. The role of the pooling layer is reducing the dimension of the input feature maps while keeping the most representative values within each group of pixels. There are different types of pooling operations. We will use the max pooling, which takes the maximum over a tile of 2×2 pixels.

Finally, the dense layer (green blocks in Fig. 5), perform an affine transformation, i.e. a matrix multiplication followed by a bias offset and a non linear function. The role of these layers stacked each other is that of bringing the flattened feature maps to the final single output unit by reducing iteratively the dimension of the input while preserving the most useful information for the final classification.

This architecture, implemented in our settings, has to be considered as a *shallow* CNN, since it is not a *deep* CNN in terms of number of stacked layers and number of parameters.

Indeed, our CNN has only 16,825 parameters and 3 Convolutional layers, while a typical CNN involved in computer vision, e.g. GoogLeNet (Szegedy, 2015), consists of 22 layers and 4 millions of parameters.

There are two reasons to choose a simple CNN.

The first one is that we would like to develop a system that could be mounted on board new telescopes through FPGA. It is clear that a huge network with millions of parameters would be difficult to adapt to this system.

The second one is that a CNN with few parameters can learn very well features associated to stacked images that are much simpler than natural images in computer vision, hence deep architectures are not necessary.

While convolutional layers require few parameters, the most computationally expensive part is associated to the dense layers (since each component of the input is involved in the weighted sum).

Although more advanced and powerful CNNs can be computed, it turns out that this very simple architecture is effective for our purpose.

Given a $48 \cdot 48$ stacked image as input, the CNN is trained to provide as output a value between 0 and 1, where 0 means a wrong combination and 1 a correct one. This architecture was found after several attempts, by always

Radius cm	**Pixel Position** pixel (X,Y)	**Speed Range** km/s	**Direction Range** deg	**Height** km
1	(12 - 34 , 12 - 34)	5 - 12	0° - 360°	370

Table 1: Space Debris Parameters simulated with ESAF.

keeping in mind different theoretical aspects: the simplicity of the images that have to be learnt, using few max-pooling to avoid information loss, and few convolutions for elementary shapes in the images. More details can be found in (Montanaro, 2020).

A dataset for CNN study includes three subsets:

- Train: It is the biggest one and has all the images that statistically cover the phase space. Through this, the network updates its weights minimizing a loss function.

- Validation: Usually it is a percentage of the training dataset that is not used for training. Instead, it is used to validate the performance of the network and to verify if some overfitting or loss of generalization is present.

- Test: It includes a lot of images never seen before and it is used for determining the final accuracy and error of the network.

A set of 80 debris simulated with ESAF with the parameters indicated in table 1 is used as a training set. The background level is set to 1 count GTU^{-1}, which is a typical value for the background measured by Mini-EUSO on oceans, due to the UV nightglow and absence of Moon light, see (Bacholle, 2020).

During the Stacking Method, the single images are shifted in the θ direction through steps of 15°, from 0° to 360°, and with a step of 2 km/s for speed starting from 5 km/s until 12 km/s. This leads to 4 combinations of speed and 24 combinations of directions, for a total of 96 combinations. For 80 SD, there are 7680 combinations. A couple of them are shown in Fig. 6.

Before becoming an input for the CNN, the stacked images are normalized in greyscale (i.e. values between 0 and 1) through the following formula:

$$GV = \frac{PV - mV}{MV - mV} \tag{4}$$

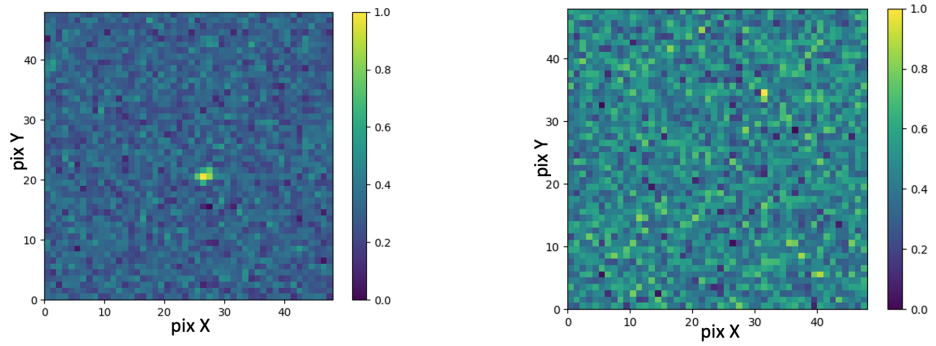


Figure 6: Correct combination coming from SM over 12 frames that matches the debris motion (left). Background Combination coming from SM over 12 frames (right). The scale is normalized between 0 and 1 (greyscale) because this helps the CNN to provide a better classification and training.

where PV indicates the pixel value and mV and MV correspond to the minimum and maximum values among all the pixels in the whole stacked image.

The training dataset consists of about 500 stacked images, one half with right combinations and the other with background ones. The 3% of this set is passed to the validation dataset. The training process, weights updating and model evaluation are done exploiting the high-level API Keras (Gulli, 2017), running on platform TensorFlow. These Python programs are executed on the interactive environment Google Colab notebook (Google Colab, 2015).

The few parameters allow to train CNN efficiently in a short time and without implementation of hardware accelerators.

We select the best CNN to assemble with the SM by training the same architecture with a dataset which includes different SBR: 3%, 1.5% and 0.7%. The network is even trained over different stacked images with 0, 4, 8 and 12 integrated GTUs.

After training, CNN is tested over 30 new SD and 30 new background events. A True Positive Rate (TPR) and a False Positive Rate (FPR) is defined and calculated over the different configurations.

Fig. 7 shows the results. As expected, the TPR grows as the integrated frames increase, until the best case, with 12 integrated frames, where all networks reach a 100% TPR (the three points overlap).

However, the network trained with the SBR of 3% has the highest TPR

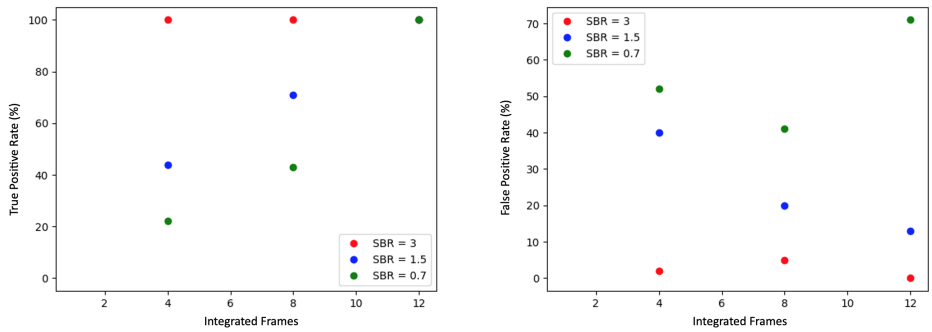


Figure 7: True Positive Rate (left) and False Positive Rate (right) for the CNN trained with different SBR (showed as different colors points) and different number of stacked images (x-axis). Note the following overlap among points in the left plot. Here, there is an overlap between SBR = 3% ,SBR = 1.5% and SBR = 0.7% at the integrated frame 12.

even when the integrated frames are fewer, and this is due to the higher brightness of the signal compared to the two others cases.

Moreover, by looking at the FPR in Fig. 7, we can observe an opposite trend as the number of integrated frames increases, and this is caused by the higher fluctuations that come into play with the stacking method. Anyway, the network trained with the 3% SBR has still a FPR of 0 % while for the other two it is much higher (13% and 70% for 1.5% and 0.7% SBRs, respectively).

This preliminary test shows that the network trained over 0.7% and 1.5% SBRs certainly does not provide reliable results as it is not able to extract clear information from faint debris being confused by noise fluctuations.

If we consider the fact that the final goal of this development is to be associated with a CAN laser that acts as a remediation system, the final solution should have a FPR of the order of 1 per hour to avoid unnecessary shooting.

All these motivations lead us to select the architecture trained with the 3% SBR.

To investigate more deeply the performance of the network with 3% SBR condition, 4.8×10^4 background images were created with Poisson fluctuations around the same average background level as before. This corresponds to a 33 minutes equivalent time for Mini-EUSO. All of them are split in 4000 sequences each one with 12 IGTUs.

SM computes 96 combinations on a single sequence, for a total of 96 $\times$

$4000 = 3.84 \times 10^5$ combinations.

After running the CNN it turns out that even with a threshold of 0.99 for a positive result, i.e. for good combination the CNN output must be greater than 0.99, a FPR of 0.25% is obtained, which corresponds to 1 event every $\sim$3.3 minutes and this is still not acceptable.

Looking more carefully at these fake events, it turns out that they really hold some brighter pixel that deceive the network leading to a wrong prediction. This also means that SM creates, in the space of all possible combinations, some stacked images that are overlays of positive fluctuations; the more combinations are performed, the greater is the risk to get false positives.

To solve this problem, the best solution is to exploit the difference between a SD and a fake background combination, which is that a SD has a steady coherent movement for a long time while the background does not.

When SM finds a right combination, it gives the speed and direction associated with that combination. If SD moves through the focal surface for long time, it stands for more than 12 IGTUs. For this reason, starting from the selected direction it is possible to repeat the stacking procedure once more but this time for many more IGTUs. Moreover, it is possible to produce more correct combinations according to a fine tuning around the selected speed and direction. Such operation enhances the chance to find an optimized combination making more contrast between a real spot and a background fluctuation.

On the contrary, in case of a false positive, a fake event is killed by repeating SM for many IGTUs. After producing these new stacked images, it is the task of CNN to recognize if these are again right combinations. The CNN is the same as in the first level of the algorithm, as it needs to perform the same task, therefore, it is not necessary to re-train a new network.

As a last condition, if CNN has recognized a right combination in both first and second trigger levels, these two selected stacked images must have an overlapping maximum in a neighborhood of at most two pixels.

This system is called STACK-CNN and it is represented in Fig. 8.

It is the last version of STACK-CNN and its robustness is proven over different tests as explained in the next section.

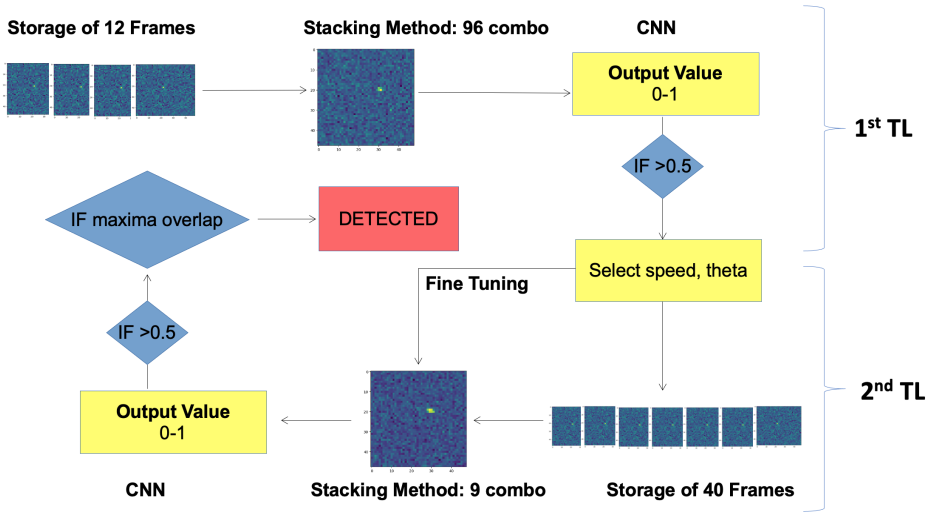


Figure 8: The proposed STACK-CNN trigger system.

6. Results of the STACK-CNN algorithm

A primary test for STACK-CNN is performed over data with pure Poisson background of 1 count · GTU^{-1} for a more reliable evaluation of FPR. 1.08×10^5 images, corresponding to 1 h and 13 min equivalent acquisition time have been generated. This time is organized in sequences of 40 IGTUs for a total of 2700 sequences.

Each sequence is passed through a STACK-CNN that automatically generates 96 combinations in the First Trigger Level (1^{st} TL).

Then the CNN processes all the combinations, giving for each one an output value between 0 and 1, and if one combo has a value greater than 0.5, then the event is passed to the Second Trigger Level (2^{nd} TL) where the whole process is repeated. However, this time the stacking procedure is performed over 40 images according to only a neighborhood of the parameters selected in the 1^{st} TL. In this way, 9 new combinations are produced according to a fine tuning around the selected speed and direction.

Once again, the CNN searches for a good combination; if this is the case, then the last test checks if the two maxima in the stacked images found in the 1^{st} TL and the 2^{nd} TL are overlapped in the same pixel positions (in a neighborhood of two pixels).

Fig. 9 shows two clear examples on how STACK-CNN avoids false positives.

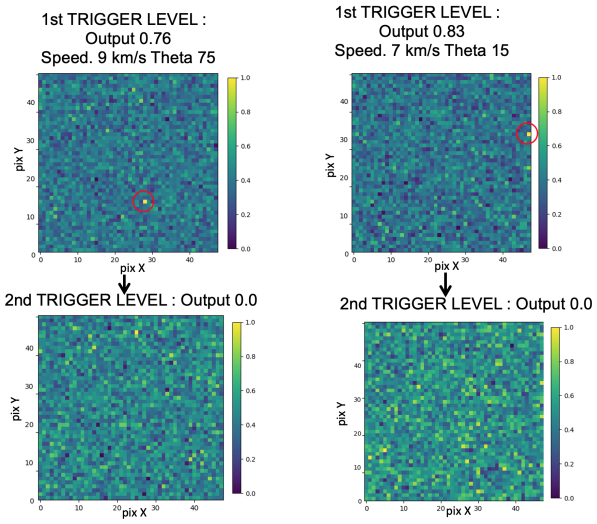


Figure 9: Two examples showing how STACK-CNN manages false positives. If in the top images two spots are visible and correctly found by CNN with the first triggering level, after stacking over 40 GTUs these disappear and CNN classifies them as background.

Though two spots are visible in the two triggered combinations at the first trigger level (top images), when the second stacking method is applied starting from the selected parameters, but with many more images, the resulting stacked images (bottom images) do not have such spots anymore. Therefore, the CNN classifies them correctly as background. The final result is that no background event gives rise to a false positive in 1 hour and 13 minutes of integrated time, making the STACK-CNN a robust and reliable system to the FPR.

As a second test the performance of STACK-CNN is verified on 16 simulated debris with fixed speed of 7.7 km/s, different sizes and distances from the detector. The SD reflectivity is set at 50% in the UV waveband. This is an average value since it depends on debris materials and can span over a large range (from 17% to 92% according to laboratory measurements, see (Miyamoto, 2019) for more details). We compare it to the performance of the standard trigger system.

Fig. 10 shows two examples of stacked images with SD. Compared to the stacked images with only background shown in Fig. 9, after applying the second trigger level, the spots become brighter with respect to the signal in a single image, and they are located in the same portion of the FoV. Therefore,

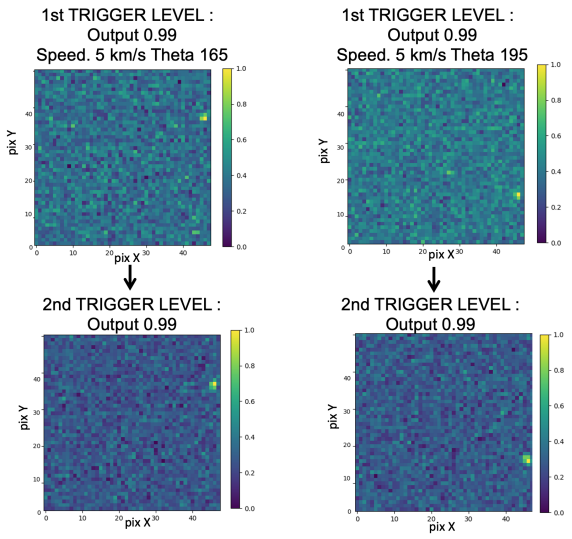


Figure 10: Two examples showing how STACK-CNN manages SD candidates. If in the top images two spots are visible and correctly found by CNN with the first triggering level, after stacking over 40 GTUs these spots are brighter and CNN correctly classifies them as SD.

the SD is correctly classified as a true positive.

The comparison results are shown in Fig. 11.

Red points are the detection limit for the standard method. Instead, blue points show the detection limit for STACK-CNN, both tested over the same dataset produced with ESAF. The improvement of STACK-CNN is clear at all distances, preserving a TPR = 100%.

To go further in the detection limit, we simulate fainter SD of the order of SNR = 1.3 (SBR = 1 %, in some cases the size of the debris is changed to reach this value).

The STACK-CNN can reach maximum distances shown by green points, accepting a TPR = 50%. In terms of SNR this means that the STACK-CNN is able to detect signals up to SNR = 1.3 against SNR = 4 for the standard method. A notable aspect is that even if the CNN is trained over SD with SBR of 3%, the STACK-CNN is able to find fainter debris up to a SBR of 1% thanks to the peculiar combination and optimization of the two methods (Stacking Method and CNN).

Later on, the STACK-CNN has been applied on real data obtained with Mini-EUSO either as background for SD or for the search of meteor events

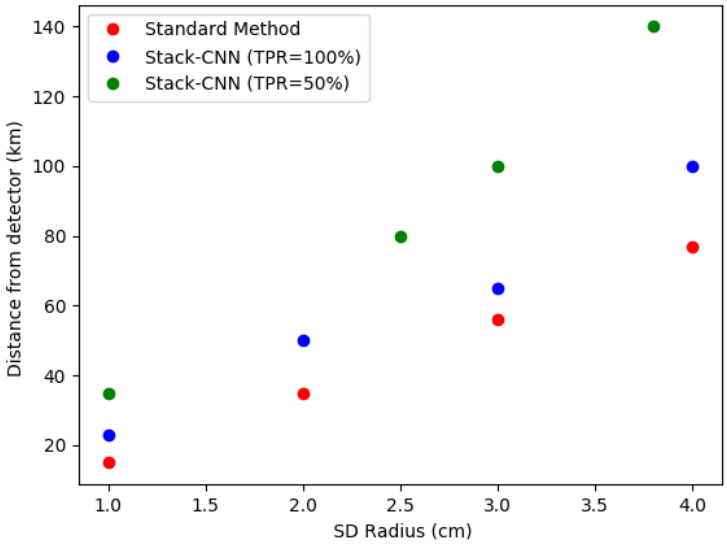


Figure 11: Comparison between the detection limit for the STACK-CNN and the standard method with simulated data.

confirming the superior performance of this method compared to the traditional one. Details of these tests and results are reported in (Montanaro, 2020) and will be subject of a future publication.

All the results are reassuring and even indicating that if the ISS will be turned by 90° or 180° in pitch and roll, Mini-EUSO should be able to detect SD in LEO orbit within acceptable distances and sizes.

7. Discussion and conclusions

A new trigger system combined with a stacking procedure and a *shallow* convolutional neural network has been presented. The STACK-CNN could be applied to any kind of light-sources moving linearly or with a known trajectory in the field of view of a telescope either from ground or from space. Its application on the detection of high velocity fragmentation debris in orbit is shown as a first adaptation. A possible future implementation is on an orbiting debris remediation system comprised of a super-wide field of view telescope like EUSO and a novel high-efficiency fibre-based CAN laser system.

The proposed method has been developed based on an initial proof of concept stage of this system, which is the Mini-EUSO detector on the ISS. By means of a simulation code of space debris we evaluated the performance of the method and compared it with the results obtained by means of a more conventional trigger system. Results indicate that our method allows to recognise signals with $\sim$1% SBR on Poisson fluctuations with a negligible fake trigger rate. This has been done assuming the average background level seen by Mini-EUSO on the ISS and pointing nadir. Most probably the typical background level pointing towards the zenith would be lower as it would not be affected by the airglow, increasing the overall sensitivity.

The next step would be to adapt STACK-CNN for ground or with space based observatories pointing towards the zenith to mimic more realistic conditions. Indeed, it is enough to change some parameters to let the STACK-CNN process images with different spatial and temporal resolutions, and search signals moving with different characteristics (e.g. speed) with respect to SD.

In parallel, the flexibility of this approach allows testing the logic directly on Mini-EUSO data to search for SD, meteors and other point-like sources which share similar behaviour.

Moreover, by simply re-adapting the speed range of STACK-CNN, it could also be applied to cosmic ray science as an offline scanning algorithm. Other practical applications to events not related to physics could be considered too.

The *shallow* CNN involved and the computational speed of the whole STACK-CNN allow the system to be mounted on board of future telescopes (such as for SD removal) inside an FPGA. Indeed CNNs have already been implemented on FPGA, and recent works even show the possibility to embed deeper architectures, see for example (Ghaffari, 2020). Here the authors propose a new framework able to build and run the project on FPGA starting from several popular high-level machine learning libraries, such as Keras (Gulli, 2017) or Pytorch (Paszke, 2017).

Another option for this CNN implementation would be its implementation on a CPU on board the satellite, since it was proven that such network runs easily even on a common laptop.

Concerning the possibility of implementing the algorithm on FPGA, it was already demonstrated that the stacking procedure can be implemented on FPGA as a second level trigger of the JEM-EUSO experiment (Bayer, 2013). Such algorithm is very similar in concept to the one for space debris

presented in this paper. Such FPGA already flew on board of stratospheric balloons with the trigger logic implemented. Details can be found in (Scotti, 2019).

At the end, the last objective will be seeing STACK-CNN that autonomously triggers and classifies events, taking a step forward for the artificial intelligence of space systems.

8. Acknowledgments

The authors would like to thank the Mini-EUSO collaboration for the fruitful discussions, Francesco Fenu for his support about ESAF simulations and Hiroko Miyamoto for the comparison with the standard approach to detect SD in Mini-EUSO. Furthermore, the authors thank Paolo Rumerio for his careful and precious proofreading of the paper. At the end, the authors express their gratitude to the three anonymous referees. Their comments and suggestions were fundamental in improving the clarity and quality of the paper.

T. Yanagisawa *et al*, Detection of Small GEO Debris by Use of the Stacking Method, Transactions of the Japan Society for Aeronautical and Space Sciences, 51(589): 61-70. (2003).

M. Bertaina *et al*, The trigger system of the JEM-EUSO project, Proc. 30th Int. Cosmic Ray Conf. (Merida), (2007).

Y. Lecun *et al*, Gradient-based learning applied to document recognition, in Proceedings of the IEEE, vol. 86, no. 11, pp. 2278-2324, doi: 10.1109/5.726791. (1998).

C. Szegedy *et al*, Going deeper with convolutions. Proceedings of the IEEE conference on computer vision and pattern recognition. (2015).

A. Montanaro, Machine Learning for space images Processing, Master Thesis, University of Torino.
`http://personalpages.to.infn.it/~bertaina/AMontanaro_mthesis_offver.pdf`
(2020).

M. D. Zeiler. ADADELTA: An adaptive learning rate method. 1212. (2012).

T. Ebisuzaki *et al*, Demonstration designs for the remediation of space debris from the International Space Station, Acta Astronautica, 112: 102 - 113. (2015)

F. Capel *et al*, Mini-EUSO: A high resolution detector for the study of terrestrial and cosmic UV emission from the International Space Station, Advances in Space Research, 62: 2954 - 2965. (2018).

`https://www.esa.int/Safety_Security/Space_Debris/Space_debris_by_the_numbers`

S. Bacholle *et al*, Mini-EUSO Mission to Study Earth UV Emissions on board the ISS, The Astrophysical Journal Supplement Series, Vol. 253, pag. 36, https://doi.org/10.3847/1538-4365/abd93d https://arxiv.org/abs/2010.01937. (2021).

A. Belov *et al*, The integration and testing of the Mini-EUSO multi-level trigger system, Advances in Space Research, 62: 2966 - 2976. (2018).

C. Berat *et al*, Full simulation of space-based extensive air showers detectors with ESAF. Astropart. Phys. 33 (4): 221247. (2010).

H. Miyamoto *et al*, Proc. 36th Int. Cosmic Ray Conf. (Madison), PoS(ICRC2019) 253. (2019).

`https://colab.research.google.com/notebooks/intro.ipynb`

A. Ghaffari *et al*, CNN2Gate: An Implementation of Convolutional Neural Networks Inference on FPGAs with Automated Design Space Exploration. Electronics, 9.12: 2200. (2020).

A. Gulli *et al*. Deep learning with Keras. Packt Publishing Ltd, (2017).

A. Paszke *et al*, Automatic differentiation in pytorch. (2017).

C. C. Aggarwal, Neural networks and deep learning. Springer, 10: 978-3. (2018).

R. Parloff, Why Deep Learning Is Suddenly Changing Your Life. Fortune. 2016. Archived from the original on 14 April 2018. Retrieved 13 April (2018).

V. Scotti *et al*, The Data Processor system of EUSO-SPB1. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment, 916: 94-101. (2019).

J. Bayer *et al*, Second level trigger and Cluster Control Board for the JEM-EUSO mission. Proceedings of 33rd International Cosmic Ray Conference, Rio de Janeiro, Brazil. (2013).