



Publication Year	2021
Acceptance in OA	2023-04-18T15:18:33Z
Title	A Deep Learning Method for AGILE-GRID Gamma-Ray Burst Detection
Authors	PARMIGGIANI, NICOLO, BULGARELLI, ANDREA, FIORETTI, VALENTINA, DI PIANO, AMBRA, GIULIANI, ANDREA, Longo, F., VERRECCHIA, Francesco, TAVANI, MARCO, Beneventano, D., Macaluso, A.
Publisher's version (DOI)	10.3847/1538-4357/abfa15
Handle	http://hdl.handle.net/20.500.12386/34085
Journal	THE ASTROPHYSICAL JOURNAL
Volume	914



A Deep Learning Method for AGILE-GRID Gamma-Ray Burst Detection

N. Parmiggiani^{1,2}, A. Bulgarelli¹, V. Fioretti¹, A. Di Piano¹, A. Giuliani³, F. Longo^{4,5,6}, F. Verrecchia^{7,8},
M. Tavani^{9,10,11,12}, D. Beneventano², and A. Macaluso¹³

¹ INAF OAS Bologna, Via P. Gobetti 93/3, 40129 Bologna, Italy; nicolo.parmiggiani@inaf.it

² Università degli Studi di Modena e Reggio Emilia, DIF—Via Pietro Vivarelli 10, I-41125 Modena, Italy

³ INAF-IASF Milano, Via Alfonso Corti 12, I-20133 Milano, Italy

⁴ Dipartimento di Fisica, University of Trieste, via Valerio 2, I-34127 Trieste, Italy

⁵ INFN, sezione di Trieste, via Valerio 2, I-34127 Trieste, Italy

⁶ Institute for Fundamental Physics of the Universe, Via Beirut 2, Trieste, Italy

⁷ ASI Space Science Data Center (SSDC), Via del Politecnico snc, I-00133 Roma, Italy

⁸ INAF-Osservatorio Astronomico di Roma, Via di Frascati 33, I-00078 Monte Porzio Catone, Italy

⁹ INAF-IAPS Roma, via del Fosso del Cavaliere 100, I-00133 Roma, Italy

¹⁰ Dipartimento di Fisica, Università Tor Vergata, via della Ricerca Scientifica 1, I-00133 Roma, Italy

¹¹ INFN Roma Tor Vergata, via della Ricerca Scientifica 1, I-00133 Roma, Italy

¹² Consorzio Interuniversitario Fisica Spaziale (CIFS), villa Gualino—v.le Settimio Severo 63, I-10133 Torino, Italy

¹³ University of Bologna, Department of Computer Science and Engineering (DISI) Viale del Risorgimento 2, I-40136 Bologna, Italy

Received 2019 November 8; revised 2021 April 17; accepted 2021 April 19; published 2021 June 16

Abstract

The follow-up of external science alerts received from gamma-ray burst (GRB) and gravitational wave detectors is one of the AGILE Team's current major activities. The AGILE team developed an automated real-time analysis pipeline to analyze AGILE Gamma-Ray Imaging Detector (GRID) data to detect possible counterparts in the energy range 0.1–10 GeV. This work presents a new approach for detecting GRBs using a convolutional neural network (CNN) to classify the AGILE-GRID intensity maps by improving the GRB detection capability over the Li & Ma method, currently used by the AGILE team. The CNN is trained with large simulated data sets of intensity maps. The AGILE complex observing pattern due to the so-called “spinning mode” is studied to prepare data sets to test and evaluate the CNN. A GRB emission model is defined from the second Fermi-LAT GRB catalog and convoluted with the AGILE observing pattern. Different p -value distributions are calculated, evaluating, using the CNN, millions of background-only maps simulated by varying the background level. The CNN is then used on real data to analyze the AGILE-GRID data archive, searching for GRB detections using the trigger time and position taken from the Swift-BAT, Fermi-GBM, and Fermi-LAT GRB catalogs. From these catalogs, the CNN detects 21 GRBs with a significance of $\geq 3\sigma$, while the Li & Ma method detects only two GRBs. The results shown in this work demonstrate that the CNN is more effective in detecting GRBs than the Li & Ma method in this context and can be implemented into the AGILE-GRID real-time analysis pipeline.

Unified Astronomy Thesaurus concepts: [Gamma-ray astronomy \(628\)](#); [Gamma-ray bursts \(629\)](#); [Convolutional neural networks \(1938\)](#); [Neural networks \(1933\)](#)

1. Introduction

Astrorivelatore Gamma ad Immagini LEggero, Light Imager for Gamma-Ray Astrophysics (AGILE) is a scientific mission of the Italian Space Agency (ASI) that was launched on 2007 April 23 (Tavani et al. 2008, 2009). The AGILE payload detector consists of the Silicon Tracker (ST; Barbiellini et al. 2001; Prest et al. 2003; Bulgarelli et al. 2010; Cattaneo et al. 2011), the SuperAGILE X-ray detector (Feroci et al. 2007), the CsI(Tl) Mini-Calorimeter (MCAL; Labanti et al. 2009), and an AntiCoincidence System (ACS; Perotti et al. 2006). The combination of ST, MCAL, and ACS forms the Gamma-Ray Imaging Detector (GRID). AGILE-GRID is used for observations in the 30 MeV–50 GeV energy range. The Precise Positioning System and the two Star Sensors provide accurate timing, positional, and attitude information. The ST is the core of the AGILE-GRID, and it relies on the process of photon conversion into electron–positron pairs. It consists of 12 trays, the first 10 of which include a tungsten converter followed by a pair of silicon microstrip detectors with strips orthogonal to each other; the last 2 consist only of silicon detectors. The γ -rays are converted in the tungsten (silicon) layers, and a readout electronics acquires and processes the data.

The AGILE team developed an automated pipeline to react to external science alerts received from the Gamma-Ray Coordinates Network (GCN) (Bulgarelli 2019). A science alert is a communication from/to the astrophysical community that a transient phenomenon is occurring in the sky. This automated pipeline can react in a fast way and detect a possible GRB counterpart in AGILE-GRID short-term (< 1000 s) observations. A detection occurs when the pipeline finds a signal with a statistical significance above a defined threshold. The GRB positions and trigger times are known in advance because the pipeline reacts to external science alerts. The analysis is performed using aperture photometry, evaluating the counts detected inside a time window containing the target (T_{on}) and a time window containing only background (T_{off}). The AGILE-GRID instrument has a point-spread function (PSF) $< 10^\circ$ at energy > 50 MeV (Sabatini et al. 2015). The counts are selected from the AGILE-GRID photon list in a radius of 10° from the center of the error localization region reported by the external science alert in order to contain the PSF of the source. The background is evaluated before the trigger time because the true duration of the GRB in the GRID energy range is unknown.

In the current method used by the AGILE-GRID automated pipeline, the significance of a GRB detection is calculated with the Li & Ma formula (Li & Ma 1983) using the counts extracted in the previous steps. The Li & Ma formula is a likelihood ratio method applied to aperture photometry. It is largely used in γ -ray astronomy and by the AGILE Team as a standard analysis for GRB detection. This method, cited from now on as Li&Ma, has two main limits.

The Li&Ma method does not use the shape of the PSF during the analysis, just the event number inside a region defined large enough to include the PSF. Furthermore, the Li&Ma method requires the counts in both the T_{on} and T_{off} time windows not to be too few, with a threshold of 10 usually applied (Li & Ma 1983). The AGILE team, following Li & Ma (1983), defined a threshold of 10 counts for the real-time analysis pipeline, and for this reason, detections with a lower count rate are discarded.

This work proposes a new detection method to overcome these limitations and, in general, improve the AGILE-GRID automated pipeline's capability to detect GRBs during the follow-up of science alerts received from other observatories through the GCN network.

This new method uses a class of deep learning (DL) methods called convolutional neural networks (CNN), described in detail in Section 5.

DL methods (LeCun et al. 2015) are subsets of machine learning (ML) methods. ML methods use automated training algorithms (without human intervention) to learn how to predict the correct output concerning several problems (classification, regression, etc.) without being directly programmed to do this. The training is performed using a training data set that is a subset of the whole population of possible inputs that the model will obtain to predict the output. ML techniques cannot be used directly on the raw data but require a first step of feature engineering (feature extraction) from the raw data. This operation is time-consuming and must be performed by field experts with a complete understanding of the data. Once extracted, the features are used as input for the ML model. DL methods, on the contrary, do not require this feature engineering performed by experts because they can extract features directly from the raw data. DL architectures, called deep neural networks (DNN), are composed of several layers that are able to extract features at different levels of abstraction during the training phase. The number of layers can vary with the problem complexity and the available computing power starting from fewer than 10 layers (Krizhevsky et al. 2012) to more than 100 (He et al. 2016) and up. DNNs have become even more popular in recent years thanks to three main factors: (i) the improvement in computational hardware (e.g., graphical processing unit—GPU) required to train DNNs with millions or billions of parameters, (ii) the availability of huge amounts of data suitable for the training of large DNN models, and (iii) the development of frameworks that can be used to implement these DNN models with standard technologies (e.g., Python).

The CNN developed in this work is used to classify AGILE-GRID intensity maps and detect the presence of GRBs in the field. Intensity maps are counts maps divided by the exposure and therefore report the measurement in $\text{ph cm}^{-2} \text{s}^{-1} \text{sr}^{-1}$ for each pixel. The CNN uses the intensity maps as input and does not require information about the exposure.

The CNN requires a training phase with large simulated data sets of intensity maps representing the average background level and the GRB flux distribution expected in the AGILE-GRID energy range 0.1–10 GeV. The study of the observing pattern is described in Section 3. Section 4 describes the GRB model used to simulate GRBs for CNN training. After performing CNN training, the p -value distribution from only background maps is computed in different observational conditions (Section 6). The CNN is then applied to real data using the GRBs' position and trigger time of the Swift-BAT¹⁴ Fermi-LAT (Ajello et al. 2019) and Fermi-GBM¹⁵ catalogs. Section 7 describes this analysis, and the results show a considerable improvement in the detection capability compared to the Li&Ma method.

The main reasons why the CNN method improves the AGILE-GRID GRBs detection capabilities are:

1. The CNN can be trained on the data of a specific instrument, learning from huge data sets of simulated data, while Li&Ma is a generic method. In fact, the PSF of the AGILE-GRID instrument is used during the CNN training phase to define the size of the kernels used during the convolution process.
2. The CNN is trained with data sets simulated using the background level calculated during real AGILE-GRID observation. In addition, the fluxes of the simulated GRBs are extracted from the Fermi-LAT GRB catalog (Ajello et al. 2019) and scaled to the AGILE energy range. All this knowledge is learned by the CNN, while Li&Ma is applied as is.
3. The CNN does not require a minimum number of events to be applicable. On the contrary, Li&Ma requires at least 10 events in the T_{on} and T_{off} time windows.

This is the first attempt to use a CNN to classify the AGILE-GRID γ -ray sky maps. The results obtained in this work (Section 7) encourage further research in this direction. Future works are planned to use the CNN to classify AGILE-GRID γ -ray sky maps containing more than one source and perform a regression analysis to determine the GRB position and the flux. These kinds of analyses cannot be performed with the Li&Ma method. The method described in this work can also be used to train a CNN network to classify sky maps produced by the next generation of X-ray and γ -ray observatories such as the Cherenkov Telescope Array (Actis et al. 2011; Acharya et al. 2019) or the e-ASTROGAM (De Angelis et al. 2021) and THESEUS (Amati et al. 2021) spacecraft. These observatories will produce more complex sky maps collecting a larger number of events and background information. The CNN can be trained on the observing condition (e.g., the background level) of the specific instrument and can learn detection patterns following the instrument's PSF. These are additional reasons to promote research in this field.

CNNs are used in astrophysics to analyze data in several contexts. In particular, CNNs can be used for image classification problems. As described in Hezaveh et al. (2017), a CNN is used to perform fast and automated gravitational lens analyses. With the next generation of ground and space observatories such as the Vera C. Rubin Observatory

¹⁴ Swift-BAT Gamma-Ray Bursts online catalog: https://swift.gsfc.nasa.gov/archive/grb_table/.

¹⁵ Fermi-GBM Gamma-Ray Bursts catalog: <https://heasarc.gsfc.nasa.gov/W3Browse/fermi/fermigbrst.html>.

(formerly Large Synoptic Survey Telescope—LSST; Thomas et al. 2020), tens of thousands of new lenses are expected to be discovered. A CNN can improve these analyses’ performances and reduce the time required to obtain results in contrast with traditional analysis methods based on the maximum likelihood estimator (MLE). CNN technologies can be used to analyze the Big Data generated by the next generation of observatories, exploiting the GPUs computing power and parallel processing. Several works use CNNs in γ -ray data analysis. In Caron et al. (2018), the CNN approach is used to analyze Fermi-LAT γ -ray maps of the Galactic Center. In Drozdova et al. (2020), the CNN method is used to extract point sources on Fermi-LAT simulated images.

2. Assumptions of This Work

The analyses presented in this paper use parameters inherited from the AGILE real-time analysis pipeline developed for the follow-up of external science alerts. The new method based on DL techniques is compared with the standard method used in this pipeline. The parameters that are not inherited by the AGILE pipeline are defined here to test the CNN model with the common conditions that can be found during the AGILE-GRID observations. This work does not treat rare and complex situations that will be analyzed in future works. Not all parameters are fixed. They may be calculated during the analyses (e.g., the time window used to evaluate the background level). The main assumptions made for this work are:

1. The γ -ray sky maps used to train and evaluate the CNN have a size of 100×100 pixels and a bin size of 0.5° . This map size is defined to be larger than the AGILE-GRID PSF and to include background regions. The bin size is a standard parameter used in the AGILE-GRID data analysis.
2. The γ -ray sky is simulated with a time window of 200 s. This value is selected after the analysis of the AGILE observing pattern described in Section 3.1.
3. The maps are simulated using a representative value for the AGILE-GRID exposure in maps with 200 s time windows, calculated excluding exposure levels under a threshold defined to avoid limit conditions that are not the goal for this work.
4. The energy range considered in this work is 0.1–10 GeV. This energy range is the standard one used by the AGILE Team to perform analysis on AGILE-GRID data and is supported by the AGILE Science Tools’ simulation software.
5. This work is focused on GRBs in the extragalactic region ($|b| > 10^\circ$, where b is the Galactic latitude) to evaluate the new method excluding regions with several background sources and to avoid the diffuse Galactic γ -ray background (Section 3.2).
6. The background levels of the maps classified with the CNN are calculated with an MLE analysis. The AGILE Team defined that the time windows to calculate the background with an MLE analysis must contain a minimum of 10 counts. The time windows are found automatically to have at least 10 counts. For a minimum of 10 counts, the background time windows last from 6 to 32 hr. During extragalactic observations, the AGILE-GRID background is isotropic and mainly dominated by charged particles populating the low-Earth orbit radiation

environment (see Section 3.2 for a full description). This background flux is fairly quiescent, and no significant variations are expected for timescales from several hours to a few days. In fact, when studying the general background fluctuation trends within a year, an average variability of only 30% (1σ) is found.

7. As described in Bulgarelli et al. (2012), the AGILE Team uses analysis regions with a radius of 10° for the AGILE-GRID data centered on the source position to include the PSF of the instrument.
8. The external science alerts considered for this work have a maximum error region of 1° . This scenario covers more than 90% of GRBs presented in the Second Fermi-LAT GRBs catalog and 100% of the GRBs reported in the Swift-BAT GRBs catalog. The science alerts with greater error regions are excluded because this work does not have the goal of finding the source’s position in a blind search.
9. The CNN uses intensity maps (counts maps divided by the exposure maps) as input. This solution makes the CNN exposure independent.

3. Modeling the Observations

The AGILE orbit (quasi-equatorial with an inclination angle of 2.5° and an average altitude of 500 km, 96 minute period) is optimal for low-background γ -ray observations. From July 2007 to October 2009, AGILE observed the γ -ray sky in “pointing mode,” characterized by a quasi-fixed pointing with a slow drift ($\sim 1^\circ/\text{day}$) of the instrument boresight direction following solar panel constraints.

Due to a change in the satellite pointing control system, since November 2009, the AGILE γ -ray observations have been obtained with the instrument operating in “spinning mode” (i.e., the satellite axis sweeps a 360° circle in the sky approximately every 7 minutes). The axis of this circle points toward the Sun, so the whole sky is exposed every six months.

This new mission configuration provides a unique capability to the AGILE satellite to discover transients. The actual spinning configuration of the satellite, together with a large field of view and a sensitivity of typically $F = (1-2) \times 10^{-8} \text{ erg cm}^{-2} \text{ s}^{-1}$ for 100 s time integration, provides a coverage of 80% of the sky, with each sky position covered 200 times per day with 100 s of integration time.

3.1. Parameter Identification

The complex observing pattern of AGILE in “spinning mode” is studied to identify the range of conditions during the observations (average exposure level and background level). These conditions are used to perform the Monte Carlo simulations of the training, validation, and test data sets. Because this work is focused on extragalactic sky regions, a sky region centered at Galactic coordinates $(l, b) = (45, 30)$ is selected to study the exposure and background level. The AGILE-GRID exposure in the center of this sky region is calculated during one spinning revolution. Figure 1 shows the typical exposure pattern for fixed accessible sky regions: the exposure values are calculated in a time window of 500 s divided into 1 s bins and in a radius of 10° . The exposure is highly variable, with a well-defined peak due to the AGILE rotation. The study of the exposure pattern during the AGILE spacecraft spinning is used to determine the time window size

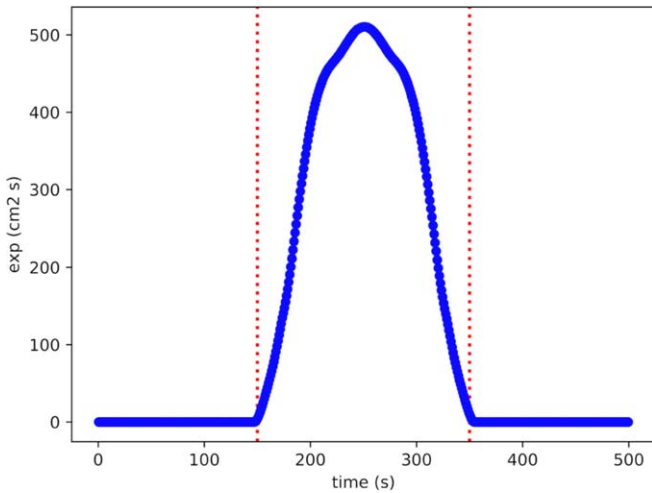


Figure 1. Typical pattern of the AGILE-GRID exposure for a fixed accessible sky region given in values of $[\text{cm}^2 \text{ s}]$ as a function of time during the AGILE spinning mode. The red dotted lines represent the time window of 200 s used for the maps simulation.

used during the maps simulation. A time window of 200 s (shown in Figure 1 with red dotted lines) is selected because it contains an entire spin of AGILE exposure.

The exposure analysis is repeated for a year of observation (2018 January 1, 2019 January 1), with an integration time of 200 s to obtain the exposure distribution. Excluding intervals with no exposure, the mean value obtained is $\sim 20 \times 10^3 \text{ cm}^2 \text{ s}$ (Figure 2). Thus, all exposure levels lower than $20 \times 10^3 \text{ cm}^2 \text{ s}$ are excluded from this work. Then, a new distribution is obtained with a mean value of $\sim 40 \times 10^3 \text{ cm}^2 \text{ s}$. The obtained value is used to simulate the data sets. This procedure aims to focus the training and the evaluation of the CNN on exposure values excluding limit conditions that are not the goal of this work.

3.2. AGILE-GRID Background Estimation

Two background components are taken into account. The diffuse γ -ray background (g_{gal}) is assumed to be produced by the interaction of cosmic rays (CR) with the Galactic interstellar medium, the cosmic microwave background (CMB), and the interstellar radiation field (ISRF). The (quasi) isotropic background (g_{iso}) includes both a contribution from the cosmic extragalactic diffuse emission as well as a component of noise due to residual CR-induced background at the detector level. In the extragalactic regions, the isotropic background dominates the AGILE-GRID data. For this reason, the g_{gal} value is considered equal to zero. More details about the AGILE-GRID background model can be found in Bulgarelli et al. (2019). One year of data (2018 January 1, 2019 January 1) is analyzed using time windows of 6 hr to obtain the distribution of g_{iso} values in an extragalactic position. The time window size is defined to have a mean of 10 counts in a radius of 10° . This counts value is required to perform the statistical analysis of the background level using the MLE method. Figure 3 shows the g_{iso} distribution excluding time windows with fewer than 10 counts in a radius of 10° . The mean of the distribution is $10.4 \times 10^{-5} \text{ cts cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$ and the standard deviation is $3.0 \times 10^{-5} \text{ cts cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$. This distribution is used to simulate

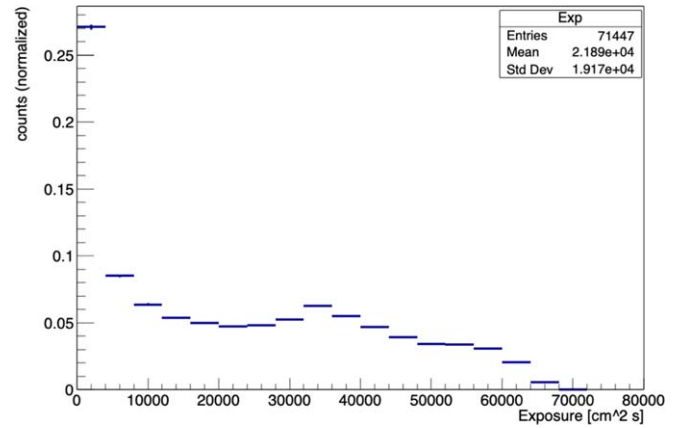


Figure 2. Histogram of exposure values ($\text{cm}^2 \text{ s}$) calculated with 200 s integrations during one year of AGILE-GRID data.

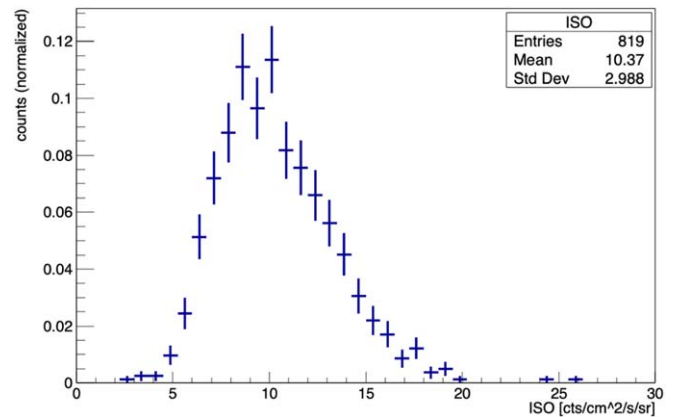


Figure 3. Histogram of g_{iso} values, expressed in $10^{-5} \text{ cts cm}^{-2} \text{ s}^{-1} \text{ sr}^{-1}$, for 6 hours integrations during one year of AGILE-GRID data.

the data sets to train the CNN; more details are given in Section 5.1.

4. GRB Model

AGILE-GRID has detected so far only 11 GRBs, namely GRB 080514B (Giuliani et al. 2008), GRB 090401B (Giuliani et al. 2009; Moretti et al. 2009), GRB 090510 (Giuliani et al. 2010), GRB 100724B (Del Monte et al. 2011), GRB 130327B (Longo et al. 2013), GRB 130427A (Verrecchia et al. 2013), GRB 131108A (Giuliani et al. 2013), GRB 170115B (Verrecchia et al. 2017), GRB 180914B (Verrecchia et al. 2018), GRB 190501A (Lucarelli et al. 2019), and GRB 1905030A (Verrecchia et al. 2019).

These events exhibit some of the main properties of the larger GRB population detected by the Fermi-LAT experiment and discussed, for example, in its First GRB catalog (Ackermann et al. 2013) and confirmed more recently in its Second GRB catalog (Ajello et al. 2019). From here on, the work will refer to the Second Fermi-LAT GRB catalog. In particular, the GRBs' main characteristics at energies greater than 100 MeV, as detected by LAT and used in this study, are the spectral model and its temporal decay. The first one shows a clear flattening of the spectrum to a value of around -2 at late times, independent of other GRB properties, and a typical larger duration concerning lower energies, extending up to 1000 s in the first catalog and up to 10,000 s in the second one. The temporal power-law decay index is clustered around -1 .

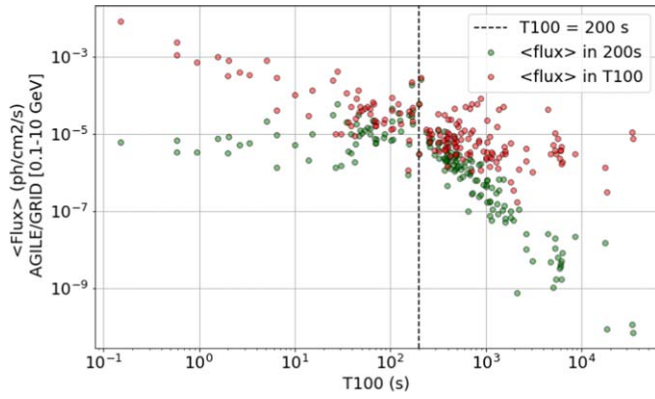


Figure 4. Average photon flux of the second Fermi-LAT GRB catalog population as a function of the Fermi-LAT duration. Red data represent the cataloged flux values within the AGILE-GRID energy range (0.1–10 GeV). In green is the evaluation of the average flux of each event for 200 s emission.

The GRBs detected by the LAT, in its first catalog, were among the brightest detected by the Fermi-GBM, with a fluence generally greater than a few 10^{-6} erg cm $^{-2}$ (see the discussion in the First LAT GRB catalog). In the Second LAT GRB catalog, the fluence limit decreased up to around 10^{-6} erg cm $^{-2}$ for long GRBs.

4.1. Flux Estimation from the LAT Catalog

Under the assumption that the simple power-law model from Ajello et al. (2019) is a good enough estimate of the spectral shape of the Fermi-LAT-detected sources, the $F_{\text{ph}}^{\text{LAT}}$ (ph cm $^{-2}$ s $^{-1}$) within the 0.1–100 GeV energy range observed by Fermi-LAT is scaled to an $F_{\text{ph}}^{\text{GRID}}$ (ph cm $^{-2}$ s $^{-1}$) value within the AGILE-GRID energy range (0.1–10 GeV). The photon fluence was then computed as $f_{\text{ph}} = F_{\text{ph}} \times T_{100}$.¹⁶ The fraction of the integrated photon flux emitted by the source in 200 s exposure time is required to simulate the data sets. For events with T_{100} greater than 200 s, the photon flux emitted within the said exposure time is calculated, assuming the cataloged simple power-law evolution model to weight the loss of later emission. The average photon flux of events with T_{100} less than 200 s is instead mediated over the exposure time to preserve the total photon fluence (Figure 4).

The simulation process is based on the use of a function fitted from GRB data having $F > 6.6 \times 10^{-6}$ ph cm $^{-2}$ s $^{-1}$ to extract random flux values. This exponential function is obtained through the Levenberg–Marquardt method: $y = a \cdot e^{-x/b}$ where $x = \log_{10}(F_{\text{ph}})$, $a = 2.7 \times 10^{-5}$ and $b = 0.523$. In Figure 5 the distribution and fitting function are displayed in the top panel, with residuals shown in the bottom panel.

5. Convolutional Neural Network

A DL approach, described in Section 1, is used to build a GRB detection algorithm. The DL architecture used in this work is the CNN, a class of DNN specifically developed to analyze and classify images (Krizhevsky et al. 2012; Goodfellow et al. 2016). The CNN is trained to classify the intensity maps in the AGILE-GRID energy range to detect GRBs. This CNN has a multiple-layer architecture where each

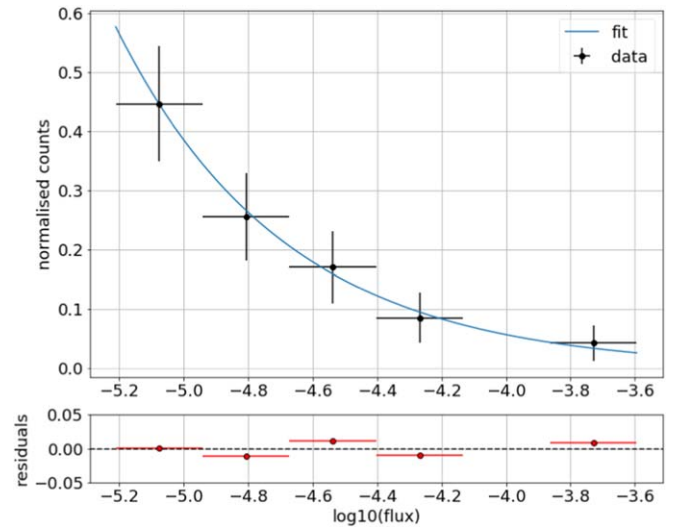


Figure 5. Gamma-ray burst population of the second Fermi-LAT catalog, with photon flux above $6.6 \cdot 10^{-6}$ ph cm $^{-2}$ s $^{-1}$. The distribution fit (blue continuous line) is achieved with an exponential law. Residuals are shown in the bottom panel.

layer can identify specific features inside the image. The supervised learning technique is used to train the CNN. This technique requires the training of the CNN with a labeled data set. These kinds of data sets contain the results of the classification for each element.

As described in Section 1, CNNs are used in several astrophysical contexts, exploiting the features of this technology for data analysis and object classification.

5.1. Data Set Simulation

Three Monte Carlo simulations for the training, validation, and test data sets, each with 40,000 AGILE-GRID intensity sky maps, are performed.

The first step executed to obtain the intensity maps is the counts maps simulation, then intensity maps are obtained from the counts maps. This simulation is performed using BUILD25 of the AGILE Science Tools (Bulgarelli et al. 2019), which includes a sky simulator called AG_multisim. The background event filter called FM3.119 and the instrument response functions (IRFs), called H0025, have been used. The energy range used for these simulations is 0.1–10 GeV. The simulator applies a Poisson-distributed noise to each pixel and produces each resulting counts map exactly as flight data. The simulated maps have an integration time of 200 s and a size of 100×100 pixels with a bin size of 0.5, i.e., $50^\circ \times 50^\circ$. The data sets are labeled to train the CNN with a supervised learning procedure. One of the inputs required for the simulation is the exposure map. The 200 s exposure map (Figure 6) is obtained from AGILE-GRID data, centering the map in the sky region defined before and searching for a time window with an exposure level equal to the mean level found in Section 3.1.

This work is focused on common conditions that can be found during the AGILE observations. The investigations to detect faint GRBs in more complex conditions and with more computing power will be performed in future works.

The γ -ray sky maps used as input for the CNN are intensity sky maps where the counts are divided with the exposure. This operation is used to make the CNN independent from the

¹⁶ The duration of the burst (T_{100}) is defined as the time between the first and last photon detection to be associated with the GRB with probability $p > 0.9$ in the 0.1–10 GeV energy range (Ajello et al. 2019).

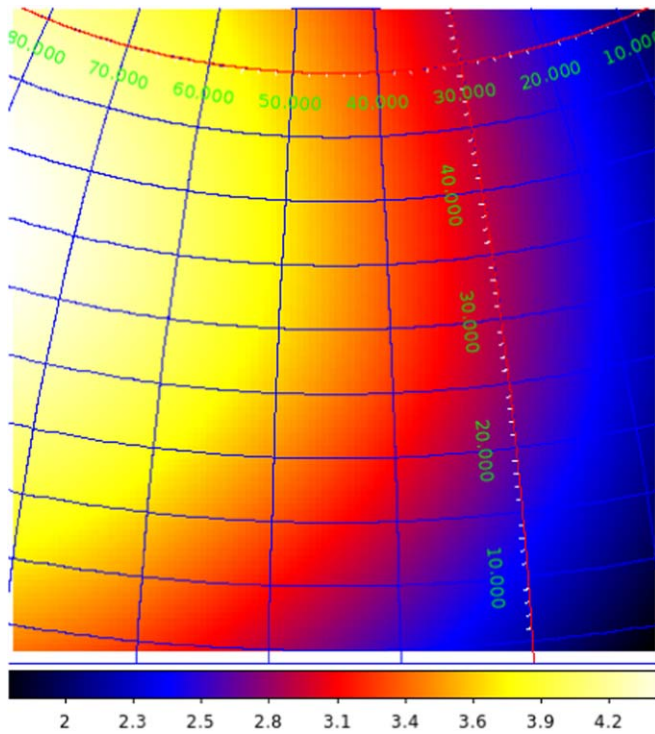


Figure 6. Exposure map used as input for the Monte Carlo simulations, expressed in $\text{cm}^2 \text{ s sr}$. The map is represented in ARC projection and Galactic coordinates, with a bin size = 0.5 .

exposure level. For this reason, the exposure value selected for the simulation is not a critical value for the CNN training.

As described in Section 2, only external alerts with a maximum error localization region smaller than 1° are considered in this work. For this reason, the GRB maps are simulated with a GRB in a random position inside the 1° radius from the maps' center. The external alert error localization region is assumed to be in the center of the simulated maps.

The background level for the simulations is obtained from the isotropic background distribution calculated in Section 3.2; no Galactic diffuse emission is considered.

The fluxes of the sources (simulated GRBs) are randomly generated using the fit function described in Section 4.1. The minimum flux value for the maps simulation is defined to reach a significance of 2σ . The flux and position of the GRBs simulated with this method are thus compatible with real external science alerts. This approach simulates the data sets with the background levels and GRB fluxes obtained from real data, improving the CNN's transfer learning from simulated data sets to real data.

The data sets contain half of the maps with a simulated GRB and other half background only. There are no additional sources simulated into the maps.

Figure 7 shows the counts' distribution of sky maps with a simulated GRB (blue) and background only (red).

5.1.1. Image Preprocessing

Figure 8 shows an example of smoothed intensity maps. The data sets are processed by performing a Gaussian smoothing of the intensity maps with a radius of 6° , assuming it to be twice the value of the AGILE-GRID instrument PSF for the energy range considered in this work.

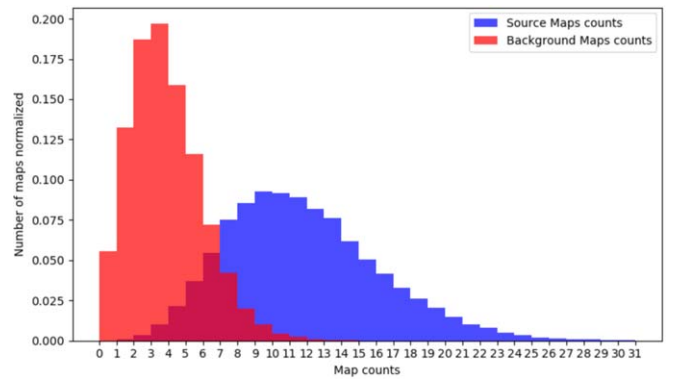


Figure 7. Histogram of the sum of photon counts inside the simulated counts maps. The red histogram is for the background-only maps, the blue histogram for the maps with a GRB.

Two 3D histograms are created by summing up all the intensity maps, pixel by pixel, to verify the counts' spatial distribution in the data set maps. The X and Y axes refer to the map pixels' reference system, and the Z-axis refers to the normalized value of summed pixels from all maps in the data set. Figure 9 shows the histogram obtained from the intensity maps containing a simulated GRB. In this histogram, the peak in the center of the map is due to simulated GRBs. Figure 10 shows the histogram for the background-only maps.

5.2. CNN Architecture

In order to find the best CNN architecture for the GRB detection, more than 700 different parameterizations were tested on a separate validation set. In particular, each CNN architecture tested changed for at least one of the following parameters: epoch number, number of convolutional layers, number of filters in convolutional layers, number of dense layers, batch size, and learning rate. However, as usually happens when working with real-world problems, increasing the neural network's complexity does not always lead to better results. Thus, the final CNN (Figure 11) is chosen according to the best tradeoff between training time and validation performance. This network is composed of 10 layers, and it is implemented using two open-source frameworks, Keras¹⁷ running on top of Tensorflow.¹⁸

The final architecture is defined as follows. The first layer receives an array of maps, each with a size of 100×100 pixels, then a Convolution2D layer with 20 filters is applied. The Convolution2D layer executes the convolution operation on the 2D input map. This layer applies all of the filters defined with a specific kernel size to the input image, producing a Feature Map for each filter. The filters are smaller than the input maps, and they are applied systematically to each overlapping part of the map. Each time a filter is applied on the input maps, it produces a Feature Map pixel. These Feature Maps are then used as input for the next layer of the CNN. In this first Convolution2D layer, the CNN uses filters with a kernel size of 12×12 pixels to identify features within the intensity maps. The size of these filters is defined starting from the PSF of the AGILE-GRID instrument. A 12×12 pixels kernel size with 0.5 side pixels is used to cover an area ($6^\circ \times 6^\circ$) approximately equal to twice the value of the AGILE-GRID PSF. The next

¹⁷ <https://keras.io>

¹⁸ <https://www.tensorflow.org>

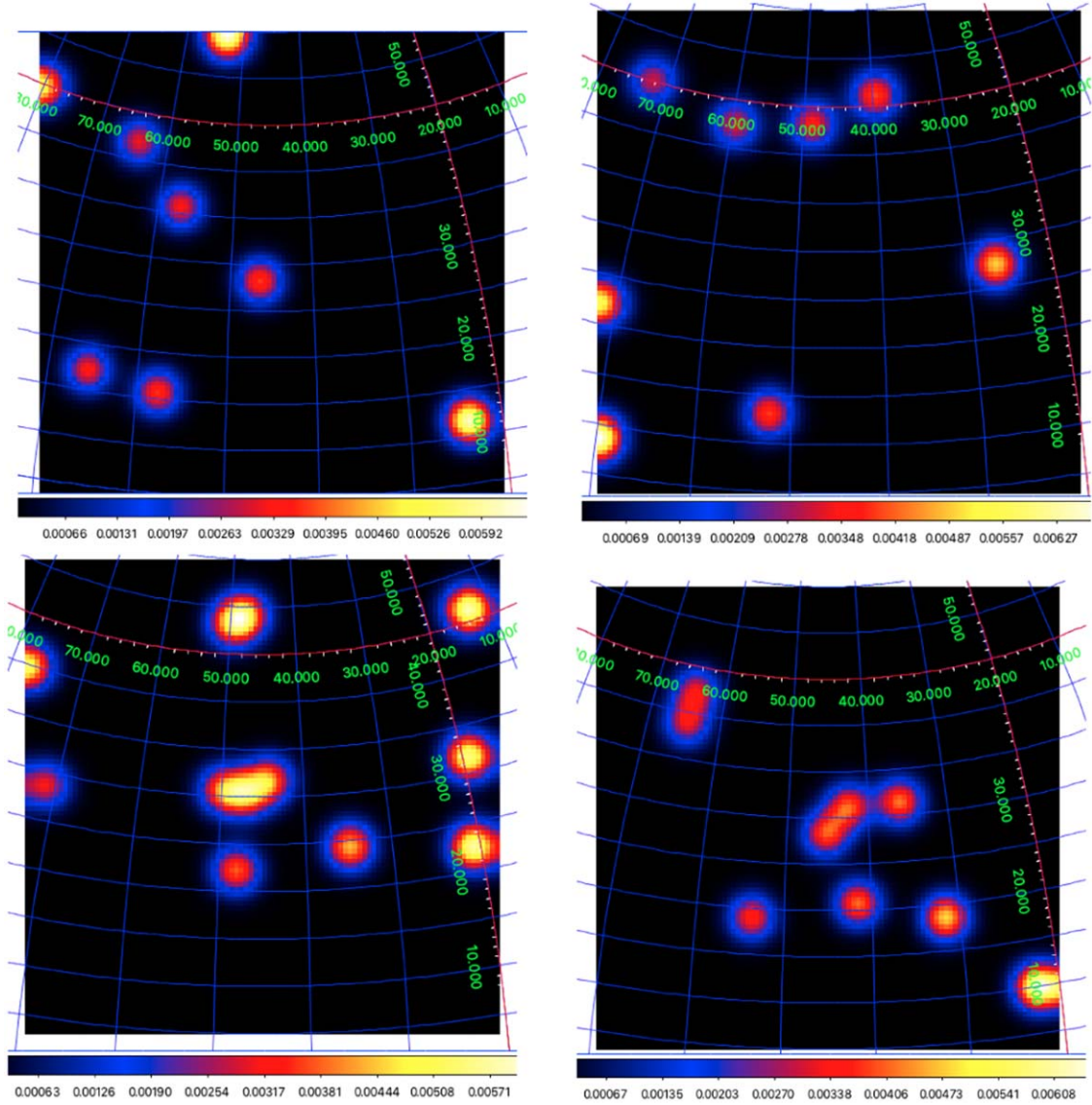


Figure 8. Smoothed intensity maps from the simulated data set used to train the CNN. The top two images are background only. The bottom two images contain a simulated GRB. No other γ -ray sources are present inside the maps. The maps are represented in ARC projection and Galactic coordinates, with a bin size = 0.5° .

layer consists of a MaxPooling2D operation with a kernel size of 2×2 pixels. This layer aims to reduce the size of the image before sending it to subsequent layers. The MaxPooling2D layer executes operations on each Feature Map separately. This layer applies the kernel on each possible overlapping position over the Feature Map and calculates the maximum value for the pixel included in the kernel. This value is reported in the new Feature Map. This layer's kernel size indicates the level of reduction that the layer applies to the Feature Maps. With a kernel size of 2×2 pixels, the Feature Map size is reduced by a factor of 2 and the number of pixels by a factor of 4. This operation speeds up the training of the CNN by reducing the size of the whole model.

Three additional Convolution2D layers and a MaxPooling2D layer are applied to find new features, reducing the image size. At this point, a Dropout layer with the probability of 25% is applied as a regularization technique to prevent overfitting. The Dropout technique consists of setting to zero the output of a percentage of neurons in the CNN. The neurons that are “dropped out” do not contribute to the learning

algorithm. Using the Dropout technique, the CNN assumes a different architecture for each iteration during the training, but all of these architectures share the weights, and the procedure optimizes a single model. The neurons cannot rely on the presence of particular other neurons, reducing the coadaptations of neurons. The model also implements a Dense layer that flattens the 2D map into a single-dimension array of 1000 elements and a Dropout layer with a probability of 50%. The Dense layer is a single-dimension array of neurons fully connected to the next layer of neurons. Finally, a two-neuron Output layer (the last layer of the network) is applied with a Softmax activation function that provides the predicted probabilities of the two classes of intensity maps: background and GRB. The Softmax activation function is applied to the last layer of the CNN to convert the output into a probability distribution; the sum of all outputs is equal to 1.

The CNN's output value, defined as CV, is the probability computed by the Softmax activation function for the two classes. If $CV = 0$, then the map is classified as background only with 100% probability. Otherwise, if $CV = 1$, the map

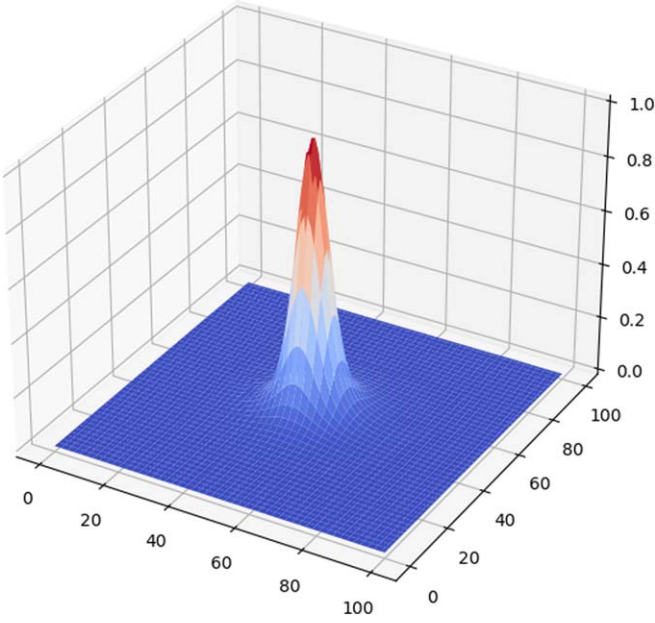


Figure 9. 3D histogram obtained summing all the counts of the smoothed maps of the data set with a GRB. X and Y axes represent the pixels of the maps, while the Z -axis represents the normalized summed counts.

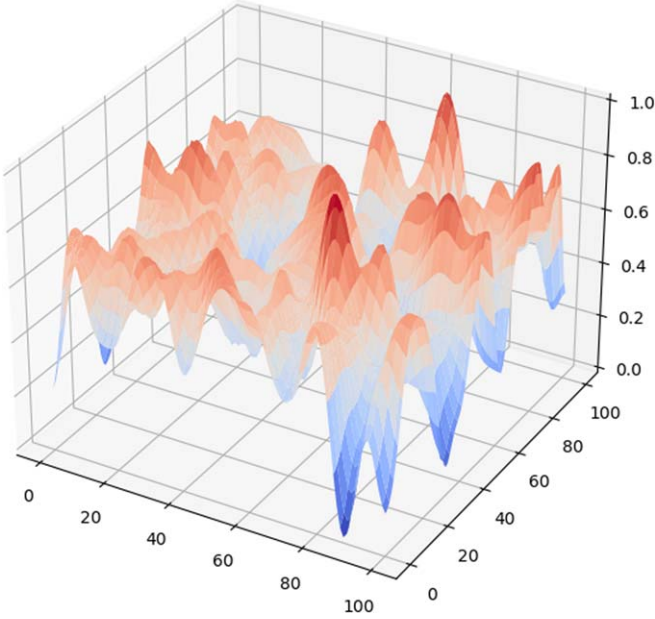


Figure 10. 3D histogram obtained summing the smoothed counts maps of the background-only data set. X and Y axes represent the pixels of the map while the Z -axis represents the normalized summed counts.

contains a GRB signal with 100% probability. Usually, the CVs are numbers between these two opposite situations, and the 0.5 value is the standard threshold between the two classifications.

As a final remark, all of the convolutional layers use the Rectified Linear Unit (ReLU) activation function ($f(x) = \max(0, x)$) that returns as output the input if it is positive; otherwise, it returns zero. This activation function is largely used with the CNN because it improves computing efficiency.

Before starting the CNN training, all the weights of the CNN model (that will be optimized during the training phase) must be randomly initialized with an initialization criterion. In this

work, the initializer used to set weights is a Keras method called the variance scaling initializer. With this method, the weights are initialized with a random number obtained from a uniform distribution $[-\text{limit}, \text{limit}]$ with $\text{limit} = \sqrt{3/n}$ and n equals the average of the numbers of the input and output units. The CNN model also contains biases initialized to zero and used together with weights during the training.

All the experiments are performed using Python 3.6 on an NVIDIA Tesla K80 GPU.

5.3. CNN Training and Testing

Once the set of optimal parameters is obtained, the final training is performed using a batch size of 200 maps, and the CNN model achieves convergence after five epochs (Figure 12). The epoch number defines the number of times that the learning algorithm examined all of the maps inside the training data set. During one epoch, each map in the training data set is used to update the model weights during the learning process. The data set contains thousands of maps. Instead of performing a single training step with the complete data set, each epoch is divided into several iterations. During each iteration, the learning algorithm analyzes a batch of data that is a subsample of the whole data set. The CNN requires a loss function as part of the optimization process to calculate the error for the current state of the model between the predicted output and the expected output. The CNN implements the sparse categorical cross-entropy loss function:

$$CE = -\sum_{i=1}^N t_i \log(f(s)_i), \quad (1)$$

where t_i is the target vector for the i th simulated map, $f(s)_i$ is the prediction of the CNN after the Softmax activation function, and N is the number of maps used for the training. The cross-entropy loss function is used when there are two or more labeled classes. The learning procedure updates the weights to reduce the loss on the next evaluation with an iterative process. The optimization algorithm used in this work to train the CNN is the Adam optimizer applied with a learning rate of 0.001. The Adam optimization algorithm, described in Kingma & Ba (2014), is an extension of the stochastic gradient descent algorithm, and empirical results demonstrate that it works well in many DL applications.

The accuracy and the area under the curve (AUC) are calculated as performance metrics. The accuracy is the percentage of input maps that the CNN classifies correctly with respect to all the maps tested. The AUC is calculated as the area under the receiver operating characteristic (ROC) curve, which is a graphical plot that shows the discrimination ability of a binary classifier when the probability threshold to determine the target class varies. In Fawcett (2006), the ROC curve and the AUC are described in detail. The X -axis of the ROC represents the false-positive rate (FPR), while the Y -axis represents the true-positive rate (TPR):

$$\text{TPR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \text{ and } \text{FPR} = \frac{\text{FP}}{\text{TN} + \text{FP}}, \quad (2)$$

where TP = True Positive, FP = False Positive, TN = True Negative, and FN = False Negative. These values are calculated on the results obtained evaluating the test data set. The

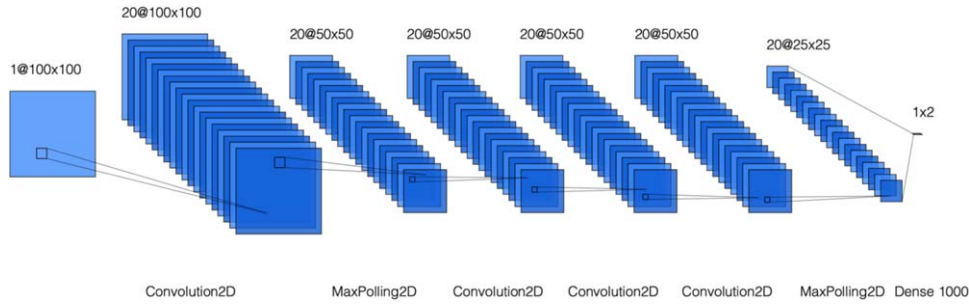


Figure 11. Schema of the CNN architecture created with a graphical tool.¹⁹

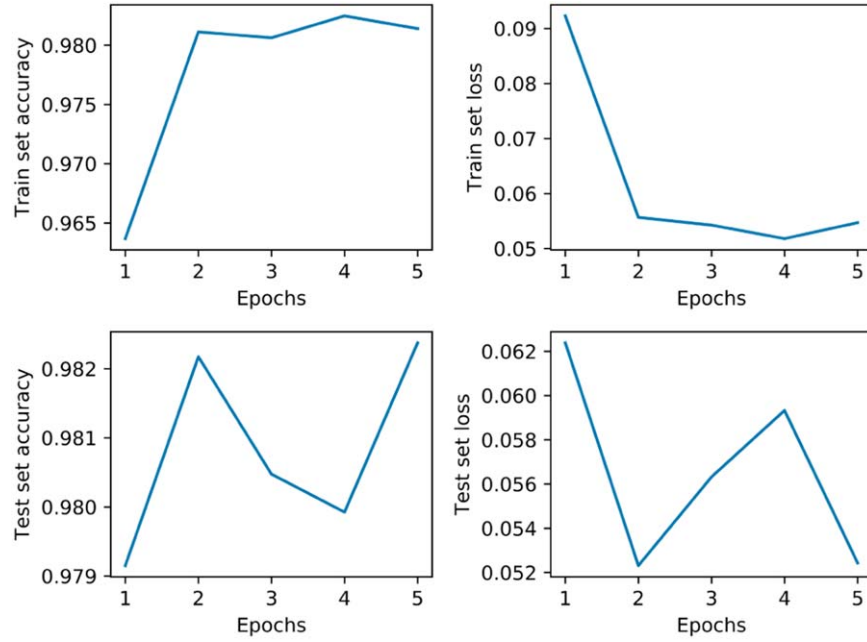


Figure 12. Accuracy and loss values obtained during the five epochs of training of the CNN, with both train and test data sets.

AUC provides an aggregated measure of performance calculated with all the possible classification thresholds represented in the ROC curve. The AUC value ranges between [0, 1], a model with 100% wrong predictions has an AUC = 0, while a model with 100% correct predictions has an AUC = 1. The AUC value should be as close as possible to 1.

Figure 12 shows the accuracy and the loss for training and test sets. As expected, the training accuracy increases with the number of epochs, while the opposite behavior is observed when considering the training loss function. This means that the CNN model gradually learns how to classify the maps correctly. The final CNN has an accuracy of 98.2% on the test data set, which means that the CNN correctly classifies the 98.2% of evaluated maps, and it performs accurately in both classes, GRB and background.

Figure 13 shows the ROC curve of the CNN obtained evaluating the test data set after the training phase. The AUC calculated with the ROC curve is equal to 0.997. This value is very close to 1 and indicates that the model reached a high-performance level.

Notice that, from a technical point of view, similar results when considering different parameterizations and a low

number of epochs to obtain optimal performance indicate that the problem is relatively simple to solve for the CNN.

During the first epoch of the training, the network reaches an accuracy close to the optimal accuracy value, improving it in small quantities in the following epochs. This because the 40,000 γ -ray sky maps that compose the training data set are enough to teach the CNN how to classify maps with and without a GRB. The CNN can learn from this data set all of the information required with a few epochs.

Several different CNN architectures with more layers are tested, but these additional layers do not lead to better results. The additional layers increase the training time, and for this reason, the CNN architecture with the best tradeoff between training time and results is selected.

6. CNN p -value Evaluation

This CNN network is trained to work as part of the automated pipeline for detecting GRBs in AGILE-GRID maps starting from external notices received from other instruments. If the CNN detects a GRB, the AGILE team can communicate this detection to the community.

The CV value provided by the CNN cannot be used directly to determine the significance level of a GRB detection. An evaluation using empty fields to determine the p -value of the

¹⁹ <http://alexlenail.me/NN-SVG/LeNet.html>

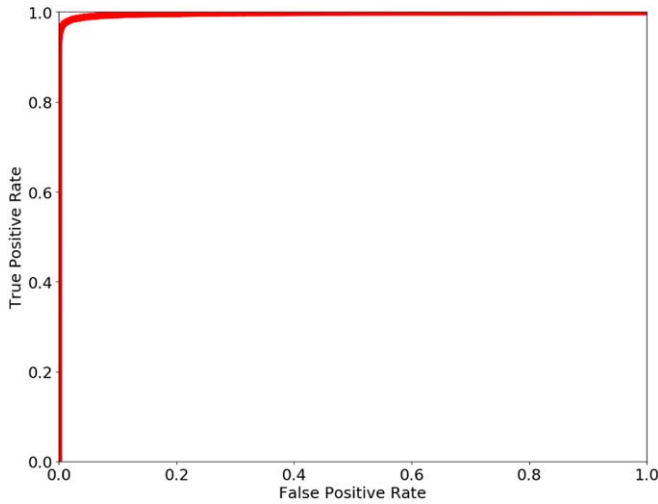


Figure 13. Receiver operating characteristic (ROC) curve. The X-axis of the ROC represents the false-positive rate (FPR), while the Y-axis represents the true-positive rate (TPR).

CNN is performed. The determination of the p -value for an MLE method is described in detail in Bulgarelli et al. (2012). A similar approach is used for the CNN presented here. The p -value distributions are calculated using the CNN output values obtained with background-only maps in different conditions. The main goal of the CNN is to detect GRBs in the context of the AGILE-GRID real-time analysis minimizing the false positives and avoiding the communication of false transient alerts to the community. For this reason, this analysis is focused on background-only maps to obtain the thresholds on CV values used to reject the null hypothesis and classify the map as a GRB map with a certain σ level.

The distribution Φ of the CV values resulting from the CNN analysis procedure on empty simulated fields with a defined level of background is defined to evaluate the p -value. The probability that the result of a trial in an empty field has $CV \geq h$ (that is the complement of the cumulative distribution function) is

$$P(CV \geq h) = \int_h^{+\infty} \Phi(x) dx, \quad (3)$$

which is also called the p -value $p = P(CV \geq h)$ and defines the probability of obtaining that value or larger when the null hypothesis is true.

6.1. p -value Determination for Different Background Conditions

The p -value distribution is strongly affected by the background level. Different p -value distributions are calculated to determine the statistical significance of a CNN detection in different background conditions, allowing this method to be applied to real maps.

Three different background levels have been selected from the background distribution defined in Section 3.2 and reported in Figure 3: the mean level ($g_{\text{iso}} = 10.4 \times 10^{-5}$ cts cm $^{-2}$ s $^{-1}$ sr $^{-1}$) and two 1σ deviations adding or subtracting the standard deviation of 3.0×10^{-5} cts cm $^{-2}$ s $^{-1}$ sr $^{-1}$, obtaining $g_{\text{iso}} = 7.4 \times 10^{-5}$ cts cm $^{-2}$ s $^{-1}$ sr $^{-1}$ and $g_{\text{iso}} = 13.4 \times 10^{-5}$ cts cm $^{-2}$ s $^{-1}$ sr $^{-1}$.

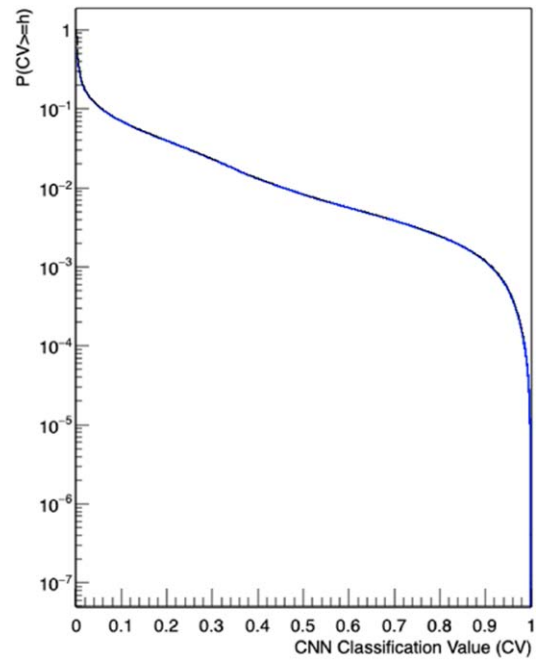


Figure 14. p -value distribution of the CV values, $g_{\text{iso}} = 10.4 \times 10^{-5}$ cts cm $^{-2}$ s $^{-1}$ sr $^{-1}$.

For each background level, a data set of 10 million background-only maps is simulated. The maps are simulated using the parameters and the same observational model used to create the data sets described in Section 5.1. These background-only maps are evaluated using the trained CNN, and the classification results are used to calculate the p -value distribution for each different observing condition. The p -value distribution of CV values obtained with the mean background level is shown in Figure 14. The number of different observing conditions is limited for constraints on computing power and time. More p -value analyses are planned for the future to improve the accuracy of this method.

The simulation software requires an exposure map as input, and a map with the exposure level of $\sim 40 \times 10^3$ cm 2 s, defined in Section 3.1, is selected. The exposure level used to simulate the maps is fixed because the CNN evaluates intensity maps, which are not influenced by different levels of exposure.

Table 1 shows the thresholds of the CNN classification values (CV) reported as $(1-CV)$ in relation to different σ levels for different background conditions. It is possible to note the dependence of the CV thresholds on the background levels. This behavior is expected, given that the detection of a GRB depends strongly on the background conditions. The results reach a maximum significance level of 5σ due to constraints on computing power and time. A fitting function between the three p -value distributions is calculated to estimate the CV thresholds for g_{iso} values different from the three values used to calculate the p -value.

7. AGILE-GRID GRB Search and Results

The GRB catalogs of Swift-BAT, Fermi-LAT, and Fermi-GBM are used to test the trained CNN with real GRBs and real AGILE-GRID data.

Table 1Relation between σ and Threshold on CNN CV Values for Different Observing Conditions

σ	p -value	$g_{\text{iso}} = 7.4$	$g_{\text{iso}} = 10.4$	$g_{\text{iso}} = 13.4$
3	1.35×10^{-3}	2.4×10^{-1}	1.1×10^{-1}	6.5×10^{-2}
3.5	2.32×10^{-4}	5.5×10^{-2}	2.7×10^{-2}	1.4×10^{-2}
4	3.17×10^{-5}	1.5×10^{-2}	4.9×10^{-3}	2.0×10^{-3}
4.5	3.40×10^{-6}	2.2×10^{-3}	8.9×10^{-4}	2.8×10^{-4}
5	2.86×10^{-7}	2.1×10^{-4}	3.5×10^{-5}	2.5×10^{-5}

Note. The table shows thresholds on the CNN values expressed as (1-CV) for different statistical significance levels. These thresholds are calculated in different background conditions defined in Section 6. The g_{iso} are expressed in 10^{-5} cts $\text{cm}^{-2} \text{s}^{-1} \text{sr}^{-1}$.

AGILE-GRID intensity maps are generated for each GRB using the GRB trigger time and the center of the error localization region defined in the catalogs. The integration time and the map size are defined in Section 5.1. This analysis is performed on the consolidated AGILE-GRID data archive. This archive covers a time window that starts on 2010 January 2 and ends on 2019 November 30. A list of 193 GRBs is obtained from these catalogs after applying four filters: (i) the AGILE-GRID map with 200 s of integration starting from the GRB trigger time must have an exposure value greater than the minimum value of $20 \times 10^3 \text{ cm}^2 \text{s}$, fixed in Section 3.1, (ii) the localization error region radius of the GRB must be $< 1^\circ$, (iii) the GRB trigger time must be inside the AGILE-GRID consolidated archive time window, and (iv) the GRBs must be extragalactic, so GRBs with $|b| < 10$ are excluded. From these 193 GRBs the CNN detected 21 GRBs with $\sigma \geq 3$. Not all GRBs detected by other instruments can be detected by the AGILE-GRID due to the different energy ranges or the lower AGILE-GRID sensitivity.

Table 2 shows the list of the detected GRBs, including the statistical significance of the CNN detection, as described in Section 6.1. The background level (g_{iso}), used to determine the right p -value distribution, is evaluated in a time window (T_{off}) preceding the GRB trigger time, starting from 6 hours and expanding it until at least 10 counts are found in a radius of 10° from the GRB position. The g_{iso} value is evaluated using an MLE. The significance is then calculated with g_{iso} and CV values using the fitting function for the p -value threshold defined in Section 6.1.

Table 2 also reports the detection with $\sigma \geq 3$ calculated using the Li&Ma method on the same list of 193 GRBs and with the same On-Off parameters used for the CNN analysis. The Li&Ma method applied to the same maps and with the same parameters can detect only two GRBs.

The N_{on} values are the number of photons inside a 10° area from the center of the GRB's alert error localization region. As already said in Section 1, the Li&Ma method requires at least 10 counts to be applied. The CNN has not this limit and can detect GRBs even when the counts of photons are less than ten. The CNN is more flexible and more suitable in this context of short detection with few photons.

The comparison between the results obtained with the CNN and with the Li&Ma method shows that in this context, the CNN detects more GRB counterparts than the Li&Ma algorithm, improving the detection capability of the AGILE-GRID automated pipeline.

Table 2

List of GRBs Detected with the CNN and Li&Ma Methods

GRB	CNN			Li&Ma		
	T_{on}	N_{on}	σ	T_{on}	N_{on}	σ
100724B	200	9	5			
110530A	200	3	3			
120711A	200	2	3			
121202A	200	2	3.5			
130427A	200	5	5			
130518A	200	2	3.5			
130828A	200	5	5			
131108A	200	11	5	200	11	4.4
141012A	200	3	4			
141028A	200	4	3			
160325A	200	4	4			
160804A	200	2	3.5			
160912A	200	8	4			
170115B	200	4	4			
170127C	200	3	4.5			
170522A	200	5	5			
170710B	200	4	3.5			
180418A	200	2	3.5			
180720B	200	13	5	200	13	4.7
190324B	200	5	3			
190530A	200	6	4.5			

Note. The table shows the comparison between the results obtained with the CNN and the Li&Ma methods searching for GRBs in AGILE-GRID data starting from GRB catalogs of other γ -ray detectors. The T_{on} is expressed in seconds. The N_{on} indicates the number of counts in a radius of 10° from the GRB error localization region's center.

8. Conclusions

This paper describes the method used to train and evaluate a CNN to detect GRBs inside the AGILE-GRID intensity maps. This CNN can be implemented into the AGILE-GRID real-time analysis pipeline to react to external science alerts. The AGILE satellite's complex observing pattern is analyzed and convoluted with a GRB model extracted from the Second Fermi-LAT GRB catalog. After the CNN training, the p -value distribution of the method is calculated in different background conditions. From the p -value distribution, the thresholds on the CNN classification values are defined for each considered observing condition to find the statistical significance of a GRB detection. The CNN is tested with the catalogs of GRBs detected by Swift-BAT, Fermi-LAT, and Fermi-GBM. From these catalogs, a list of GRBs suitable for the AGILE-GRID analysis is extracted. The position and the trigger time presented in these catalogs are used to search for GRBs counterpart in the AGILE-GRID consolidated data archive, analyzing the AGILE-GRID intensity maps with the CNN.

Table 2 shows the GRBs counterpart detected by the CNN from these catalogs having $\sigma \geq 3$. The CNN detects 21 GRBs from a list of 193 observable GRBs. The CNN leads to a 5σ detection of the important GRB 130427A, previously detected by AGILE (Verrecchia et al. 2013) only with an MLE analysis on a very long (12 hr) exposure including the T_0 , while no Li&Ma detection on prompt time interval was obtained. The CNN can detect GRBs in different observing conditions without requiring new training, thanks to the different p -value distributions calculated with different background levels. Not all GRBs detected from other instruments can be detected by AGILE-GRID due to instrument constraints. However, the

results of this work prove that the CNN can improve the detection capability of the AGILE-GRID in the 0.1–10 GeV energy range when an external science alert is received. The same analysis performed with the Li&Ma detects only two GRBs from the same GRB list, confirming the effectiveness of the CNN in the context of the AGILE-GRID real-time analysis.

The AGILE Mission is funded by the Italian Space Agency (ASI) with scientific and programmatic participation by the Italian Institute of Astrophysics (INAF) and the Italian Institute of Nuclear Physics (INFN). Investigation supported by ASI grant I/028/12/6. We thank the ASI management for unfailing support during AGILE operations. We acknowledge the effort of ASI and industry personnel in operating the ASI ground station in Malindi (Kenya), and the data processing done at the ASI/SSDC in Rome: the success of AGILE scientific operations depends on the effectiveness of the data flow from Kenya to SSCD and the data analysis and software management.

ORCID iDs

N. Parmiggiani  <https://orcid.org/0000-0002-4535-5329>
 A. Bulgarelli  <https://orcid.org/0000-0001-6347-0649>
 V. Fioretti  <https://orcid.org/0000-0002-6082-5384>
 A. Di Piano  <https://orcid.org/0000-0002-9894-7491>
 A. Giuliani  <https://orcid.org/0000-0002-4315-1699>
 F. Longo  <https://orcid.org/0000-0003-2501-2270>
 F. Verrecchia  <https://orcid.org/0000-0003-3455-5082>
 M. Tavani  <https://orcid.org/0000-0003-2893-1459>
 D. Beneventano  <https://orcid.org/0000-0001-6616-1753>
 A. Macaluso  <https://orcid.org/0000-0002-1348-250X>

References

Acharya, B. S., Agudo, I., Al Samarai, I., et al. 2019, Science with the Cherenkov Telescope Array: for the Cherenkov Telescope Array Consortium (Singapore: World Scientific Publishing Co)
 Ackermann, M., Ajello, M., Asano, K., et al. 2013, *ApJS*, **209**, 11
 Actis, M., Agnetta, G., Aharonian, F., et al. 2011, *ExA*, **32**, 193
 Ajello, M., Arimoto, M., Axelsson, M., et al. 2019, *ApJ*, **878**, 1

Amati, L., O'Brien, P. T., Götz, D., et al. 2021, *Proc. SPIE*, **395**, 114442J
 Barbiellini, G., Bordignon, G., Fedel, G., et al. 2001, in AIP Conf. Proc., **587**, ed. S. Ritz, N. Gehrels, & C. R. Shrader, **754**
 Bulgarelli, A. 2019, *ExA*, **48**, 199
 Bulgarelli, A., Argan, A., Barbiellini, G., et al. 2010, *NIMA*, **614**, 213
 Bulgarelli, A., Chen, A. W., Tavani, M., et al. 2012, *A&A*, **540**, A79
 Bulgarelli, A., Fioretti, V., Parmiggiani, N., et al. 2019, *A&A*, **627**, A13
 Caron, S., Gómez-Vargas, G. A., Hendriks, L., et al. 2018, *JCAP*, **2018**, 058
 Cattaneo, P. W., Argan, A., Boffelli, F., et al. 2011, *NIMA*, **630**, 251
 De Angelis, A., Tatischeff, V., Argan, A., et al. 2021, *ExA*, in press (arXiv:2102.02460)
 Del Monte, E., Barbiellini, G., Donnarumma, I., et al. 2011, *A&A*, **535**, A120
 Drozdova, M., Broilovskiy, A., Ustyuzhanin, A., & Malyshev, D. 2020, *AN*, **341**, 819
 Fawcett, T. 2006, *PaReL*, **27**, 861
 Feroci, M., Costa, E., Soffitta, P., et al. 2007, *NIM A*, **581**, 728
 Giuliani, A., Cutini, S., Pittori, C., et al. 2009, *GCN*, **9075**
 Giuliani, A., Fuschino, F., Vianello, G., et al. 2010, *ApJL*, **708**, L84
 Giuliani, A., Longo, F., Verrecchia, F., et al. 2013, *GCN*, **15479**
 Giuliani, A., Mereghetti, S., Fornari, F., et al. 2008, *A&A*, **491**, L25
 Goodfellow, I., Bengio, Y., & Courville, A. 2016, *Deep Learning* (Cambridge: MIT Press)
 Hezaveh, Y. D., Perreault Levasseur, L., & Marshall, P. J. 2017, *Natur*, **548**, 555
 He, K., Zhang, X., Ren, S., & Sun, J. 2016, in 2016 IEEE Conf. Computer Vision and Pattern Recognition (CVPR) (Las Vegas, NV), **770**
 Kingma, D. P., & Ba, J. 2014, arXiv:1412.6980
 Krizhevsky, A., Sutskever, I., & Hinton, G. 2012, in NIPS E, **25**, 1097
 Labanti, C., Marisaldi, M., Fuschino, F., et al. 2009, *NIM A*, **598**, 470
 LeCun, Y., Bengio, Y., & Hinton, G. 2015, *Natur*, **521**, 436
 Li, T.-P., & Ma, Y.-Q. 1983, *Apj*, **272**, 317
 Longo, F., Giuliani, A., Marisaldi, M., et al. 2013, *GCN*, **14344**
 Lucarelli, F., Verrecchia, F., Piano, G., et al. 2019, *GCN*, **24361**
 Moretti, E., Longo, F., Barbiellini, G., et al. 2009, *GCN*, **9069**
 Perotti, F., Fiorini, M., Incorvaia, S., Mattaini, E., & Sant Ambrogio, E. 2006, *NIM A*, **556**, 228
 Prest, M., Barbiellini, G., Bordignon, G., et al. 2003, *NIM A*, **501**, 280
 Sabatini, S., Donnarumma, I., Tavani, M., et al. 2015, *ApJ*, **809**, 10
 Tavani, M., Barbiellini, G., Argan, A., et al. 2008, *NIM A*, **588**, 52
 Tavani, M., Barbiellini, G., Argan, A., et al. 2009, *A&A*, **502**, 995
 Thomas, S. J., Barr, J., Callahan, S., et al. 2020, *Proc. SPIE*, **11445**, 114450I
 Verrecchia, F., Lucarelli, F., Pittori, C., et al. 2019, *GCN*, **24683**
 Verrecchia, F., Pittori, C., Giuliani, A., et al. 2013, *GCN*, **14515**
 Verrecchia, F., Ursi, A., Lucarelli, F., Longo, F., & Tavani, M. 2017, *GCN*, **20474**
 Verrecchia, F., Ursi, A., Tavani, M., et al. 2018, *GCN*, **23231**