Event Selection and Background Rejection in Time Projection Chambers Using Convolutional Neural Networks and a Specific Application to the AdEPT Gamma-ray Polarimeter Mission

Richard L. Garnett[a], Soo Hyun Byun[a,b], Andrei R. Hanu[a,b], Stanley D. Hunter[c]

[a]*Radiation Sciences Graduate Program, McMaster University, Tandem Accelerator Building, 1280 Main St. W, Hamilton, ON, L8S 4K1, Canada*
[b]*Department of Physics and Astronomy, McMaster University, Hamilton, ON, L8S 4K1, Canada*
[c]*NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA*

Abstract

The Advanced Energetic Pair Telescope (AdEPT) gamma-ray polarimeter uses a Time Projection Chamber (TPC) for measuring pair production events and is expected to generate a raw instrument data rate four orders of magnitude greater than is transmittable with typical satellite data communications. `GammaNet`, a Convolutional Neural Network (CNN), proposes to solve this problem by performing event classification on-board for pair production and background events, reducing the data rate to a level that can be accommodated by typical satellite communication systems. In order to train `GammaNet`, a set of 1.1x10^6 pair production events and 10^6 background events were simulated for AdEPT using the `Geant4` Monte Carlo code. An additional set of 10^3 pair production and 10^5 background events were simulated to test `GammaNet`'s capability for background discrimination. With optimization, `GammaNet` has achieved the proposed background rejection requirements for Galactic Cosmic Ray (GCR) proton events. Given the best case assumption for downlink speeds, signal sensitivity for pair production ranged between 1.1±0.5% to 69±2% for 5 and 250 MeV incident gamma rays. This range became 0.1±0.1% to 17±2% for the worst case scenario of downlink speeds. The application of a feature visualization algorithm to `GammaNet` demonstrated decreased response to electronic noise and events exiting or entering the frame and increased response to parallel tracks that are close in proximity. `GammaNet` has been successfully implemented and shows promising results.

Keywords: Pair production; Neural network; Machine vision; Radiation; Event classification; Event discriminator

1. Introduction

Recent advances in machine learning and computer vision have led to astonishing improvements in image classification performance [1–3], where algorithms estimate the likelihood that an input image belongs to a set of labels that describe features contained within the image. Current state of the art algorithms perform with around 4% top-5 error [1] [2]. These results were demonstrated on test sets of images from the ImageNet Large Scale Visual Recognition Competition (ILSVRC) [4] which contain images belonging to 10^3 different classes.

The application of machine learning to event classification in radiation detection is a natural progression of the field given that radiation detectors produce highly structured signals. These signals are often dependent on the nature of interacting radiation, and the type of interaction undergone. High energy physics projects such as the Large Hadron Collider have utilized machine learning applications for event classification [5, 6]. There has also been implementations of machine vision for image classification in radiation imaging detectors using Convolutional Neural Networks (CNNs) to classify neutrino interactions at Fermilab and the Ash River Laboratory [7].

The CNN application explored in this work has been developed for the event classification of images generated from a large (8 m^3) Time Projection Chamber (TPC) being designed for the Advanced Energetic Pair Telescope (AdEPT) [8], a mission to measure medium-energy gamma-ray polarimetry. The design details of AdEPT are discussed in detail in [8], and briefly summarized in Section 1.1.

1.1. The AdEPT Instrument

Astrophysical gamma rays are a means to probe the most extreme non-thermal processes in the Universe and their study provides valuable insight into the fundamental physics and structure of the most powerful particle accelerators. Most studies of astrophysical gamma rays have been in the ~20 MeV to 300 GeV energy range, using measurements from the AGILE [9] and Fermi [10] space telescopes. However, neither instrument was optimized for polarization sensitivity or observations in the medium energy (~0.1–200 MeV) band, where many astrophysical objects exhibit unique behavior. The medium energy gamma-ray band has so far proven difficult to study due to competing photon interactions, namely Compton scatter and

Email addresses: `garnetri@mcmaster.ca` (Richard L. Garnett), `hanua@mcmaster.ca` (Andrei R. Hanu)

[1] Where the top-5 error is determined by the fraction of test images for which the correct label is not among the five labels considered most probable by the algorithm.

pair production. Each of these interactions generate different signatures, and the manner in which polarization information is gathered consequently requires differing algorithms and instrumentation [8, 11, 12]. The optimization of a detector for both pair production and Compton scatter interactions on-board a satellite is prohibitive. The challenge is further exacerbated by the Galactic Cosmic Ray (GCR) background, which is an extragalactic source of charged atomic nuclei at extremely high kinetic energy. The GCR background cannot be effectively shielded for on satellites given their high kinetic energy, which can extend to several TeV per nucleon. In addition, the fluence of GCR particles exceeds the astrophysical gamma-ray flux by approximately four orders of magnitude.

Next-generation telescopes are being developed with the goal of characterizing the complete signature of gamma rays including their direction, energy, arrival time, and polarization. The most promising space missions (AdEPT [8], HARPO [13], and SMILE-I/II [14, 15]) proposed to explore the gamma-ray sky in the medium energy range are based on low-density gaseous TPC technologies that enable precise, three-dimensional tracking of particle interactions.

The AdEPT mission, for which `GammaNet` is being developed as a proposed event classifier, is one such medium energy gamma-ray polarimeter. The science data for AdEPT will consist of pair production interactions, with a background composed primarily of GCR and Compton scatter interactions. Compton scatter, though a photon interaction of interest for characterizing the medium energy gamma-ray spectra, is considered background for the AdEPT mission. Compton scatter is considered background because the AdEPT instrument is not designed to measure polarization for this interaction. The AdEPT TPCs takes advantage of the Micro-Well Detector (MWD) technology augmented with the negative ion drift technique [16] to construct an instrument with the largest volume that can be accommodated in the rocket fairings currently available to MIDEX missions, 8 m^3. The active gas volume of the TPC is bounded on the top and bottom faces by an array of MWDs defining the 400 μm X- and Y-coordinate spatial resolution of the TPC [8]. The uniform electric field in the active volume provides a constant ionization charge drift velocity. Measurement of a relative arrival time of the signals on the detector strips provides the third, Z-coordinate. The use of the negative ion drift technique in the AdEPT TPC design [8] effectively reduces electron drift diffusion in the gas, making possible drift distances up to 1 m. With the applied electric field, ionization charge can traverse the Z dimension of the detector within a maximum of 50 ms.

The use of the negative ion drift technique precludes the use of an anti-coincidence system, as used in HARPO [13], resulting in large raw data rates. This requires an alternative on-board processing approach for discrimination of GCR tracks and gamma-ray interactions. The 8 m^3 version of AdEPT is estimated to produce an uncompressed data rate of $\sim$16 Gbps, far too large for current satellite communication. Currently the Fermi Large Area Telescope mission [10, 17, 18] achieves an average science data downlink of 1.5 Mbps, while planned communications methods aim to achieve an average 50 Mbps downlink [19]. The range of possible average downlinks leaves two to four orders of magnitude difference between the raw data rate and communications data rate for the AdEPT mission. Our proposed solution is to use computer vision feature recognition algorithms running on-board the spacecraft in real time to discriminate gamma-ray interactions of interest from the abundant GCR background. The desired outcome for the algorithm is to perform event classification within 50 ms with a background rejection rate between 99.99% and 99.69%, which would reduce the raw data rate to one which can be accommodated by satellite downlink. The hardware to be used for AdEPT has not yet been chosen, though commercial solutions are available that offer enough computing power for `GammaNet`. One such solution is Innoflight's Compact Flight Computer 500, which is radiation tolerant up to 30 krad, and is space rated. Additionally, National Aeronautics and Space Administration (NASA) is investigating the suitability of System on a Chip (SoC) solutions available from NVIDIA [20].

In this paper we explore `GammaNet`, a CNN trained on simulated images from a high resolution gaseous TPC, and its performance in classifying gamma-ray events on images contaminated with a GCR background. To evaluate the performance of `GammaNet`, we performed a Receiver Operating Characteristic (ROC) analysis [21] to assess how the background rejection threshold influences the specificity and sensitivity of the classifier. Specificity is determined as the rate at which negative events are correctly classified as negative. Sensitivity is determined as the rate at which positive events are correctly classified as positive. The result of our ROC study demonstrated that `GammaNet` can reliably achieve the proposed background rejection rates of between 99.99% and 99.69%. At these rates of background rejection, `GammaNet` correctly classifies between 10±1% and 52±2% of pair production images over the energy range of interest. The simulation used for generating training and testing data, as well as the architecture and training protocol for GammaNet, are described thoroughly in Sections 2 and 3. An analysis of `GammaNet's` performance and failures is presented in Section 4. Observations are presented in Section 5 for the features utilized by `GammaNet` classifying the simulation images of AdEPT.

2. Monte Carlo Simulations of AdEPT

The AdEPT raw TPC data consists of two orthogonal projections, XZ and YZ, of the tracks in the active gas volume. The simulation of the response and readout of the AdEPT instrument was carried out using the GEANT4 Monte Carlo toolkit [22, 23]. The application, which is named `G4AdEPTSim` [24], simulates the passage of GCR protons and gamma rays through an active volume filled with 1.5 atmospheres of Ar and CS_2 at a temperature of 293 degrees K with a sub-scale size of 25x25x25 cm^3, and full-scale size of 8 m^3. The use of the sub-scale volume was to determine what level of downscaling was viable for use in `GammaNet`, and subsequently the full-scale volume was used. Downscaling of the simulation results is necessary because the time to train and run classification for any CNN is strongly correlated to the image size passed to it. The full size AdEPT TPC will produce images of 5000 x 5000 pixels, which would be prohibitively slow in terms of both training and time to classification during operation.

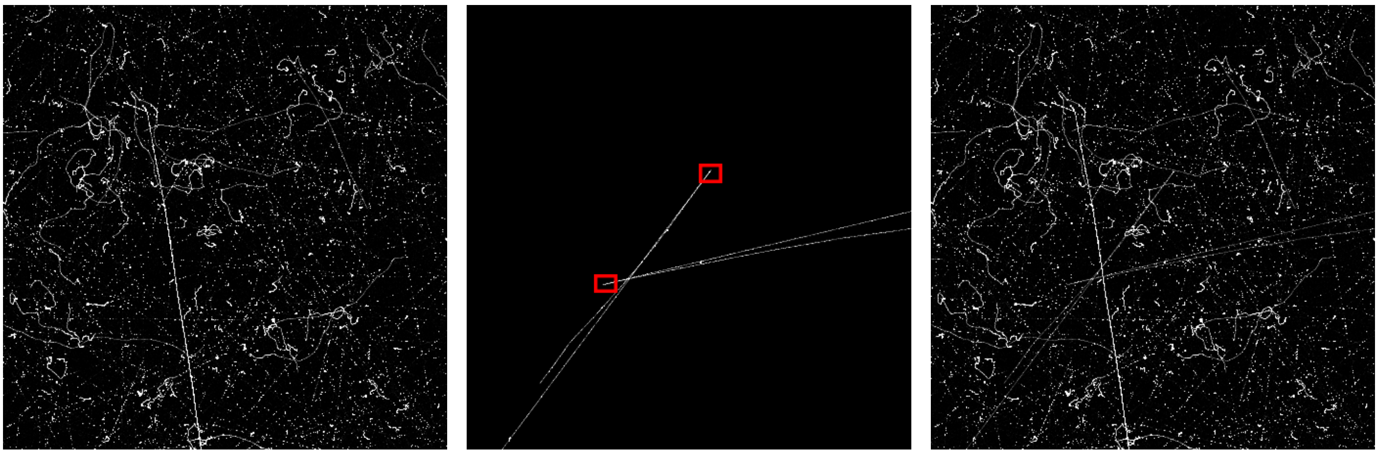
The physics included in the simulation account for the different types of interactions between source particles and the Ar gas. These include hadronic physics for the interaction of GCR protons, electromagnetic physics for the interaction of gamma rays and electrons, and photo-absorption ionization model to accurately model the primary ionization and energy loss of relativistic charged particles in low density media. `G4AdEPTSim` produces the ideal response of AdEPT, reporting the number of ionization electrons, their X-, Y-, and Z-coordinates, and the energy deposited in the active volume by a single incident particle.

AdEPT is proposed for launch into a low-Earth orbit with a 550 km altitude and a 28 degree inclination. The background environment in such an orbit is well-known and consists predominantly of cosmic diffuse radiation, atmospheric gamma-ray emissions, reactions induced by albedo neutrons, and background produced by satellite materials activated by fast protons and alpha particles [25–29]. In the 0.1 to 200 MeV energy range, the instrument background is dominated by charged particles in the Van Allen belt impinging on the spacecraft, cosmic diffuse radiation, and atmospheric gamma-ray emissions.

`G4AdEPTSim` models the simulated events using a spherical volume source of radius 22 cm for the sub-scale version, and 1.73 m for the full-scale version, which is concentric with the active volume. The arrival direction of the simulated particles is isotropically distributed on a sphere, producing a uniform distribution of the source particles within the sphere. For this work, the background component consisted of only GCR protons with the energy spectrum from the Space Environment Information System for the expected AdEPT orbital conditions. GCR protons were selected as the background because they comprise the majority of the GCR fluence. Astrophysical sources of gamma rays simulated with mono-energetic energies ranging from 5–250 MeV were generated using the same source geometry as background.

Each simulation run of the full-scale AdEPT instrument included 375 incident GCR protons or two incident gamma rays, for background and signal respectively. The sub-scale simulation runs consisted of five incident GCR protons or two incident gamma rays to account for the reduced surface area relative to the full-scale instrument. The number of incident particles were chosen in each case to fit the expected number of primary tracks, given the AdEPT instrument parameters [8], within the 50 ms collection window. There are two incident gamma rays for both simulations because the anticipated pair production rate in the full size simulation is less than one, although there is still the probability of two pair production events occurring within one collection window. The source geometry allowed for the possibility of particles to miss the active volume, but results were only recorded if at least one particle interacted with the active volume. The source geometry used allows for a varying number of tracks to be recorded from each simulation run, although the number of simulated particles was constant between runs.

Per simulation run the number of ionization electrons in 400 x 400 x 400 μm^3 voxels was recorded, corresponding to the nominal resolution of the AdEPT instrument. The number of ionization electrons in each voxel is then projected onto the XZ and YZ planes to generate images. To emulate the response of the AdEPT detector, electronic noise was added to the signal output for each set of images. The addition of electronic noise was performed by adding a randomly generated number of electrons, from a normal distribution with standard deviation of two and a mean of zero, to each pixel of an image. In addition to electronic noise, background events were added to every gamma-ray image in the form of GCR protons. To do the background event addition, GCR proton images were generated with electronic noise and gamma-ray images without. Each gamma-



(a) (b) (c)

Fig. 1. XZ projection of the sensitive volume of the AdEPT simulation. a) GCR background image containing several proton tracks with added electronic noise. b) gamma-ray image, containing two pair production events with the vertices outlined in red for illustrative purposes. c) Combination image that would be used for training and testing `GammaNet`. These simulation images have had their contrast adjusted for better viewing in this paper.

ray image then had a unique GCR image added to it. Gamma-ray images were generated without addition of electronic noise to ensure GCR images and the composite gamma-ray images would have a constant amount of electronic noise. Figure 1 shows an example of the process used for generating the pair production data set, where an image containing two pair production events is added to a background GCR image with two tracks.

Correctly labeled image sets were generated from these simulations for both training and testing of `GammaNet`. The training image sets contained 1.1x10^6 pair production images and 10^6 background GCR proton images. The testing image sets contained 1.5x10^3 pair production images, 1.5x10^3 Compton scatter images, and 10^6 background GCR images. The Compton scatter images were included in testing, but not training, as an additional source of background. `GammaNet` was found to be less sensitive to the Compton scatter images than pair production. The initial intuition when applying a CNN to this classification problem was that the CNN would be able to pick up on the discerning characteristic of pair production events compared to GCR proton tracks. The key signature of pair production being the vertex created by the electron-positron pair at the interaction site shown in Figure 1b. These pair production signatures are further discussed in the results, Sections 4 and 5.

3. GammaNet

`GammaNet` was inspired by the successes of a CNN designed for classification of neutrino interaction events in the NOvA experiment at Fermilab [30]. The CNN showed an increased performance compared to the state-of-the-art algorithms currently deployed for classification of neutrino events at Fermilab. Specifically, there was a relative increase of 40% sensitivity for electron neutrino signals, going from 35% to 49% [7]. However, the implementation for `GammaNet` is more generic than that used for the NOvA detector in that `GammaNet` only classifies to two classes, as opposed to the 13 classes used for the NOvA experiment. `GammaNet` produces a probability that the input image from AdEPT contains a pair production event, which, above a certain threshold, will result in a positive signal, and below will produce a negative signal. This simplicity allows for faster classification with an unsophisticated architecture.

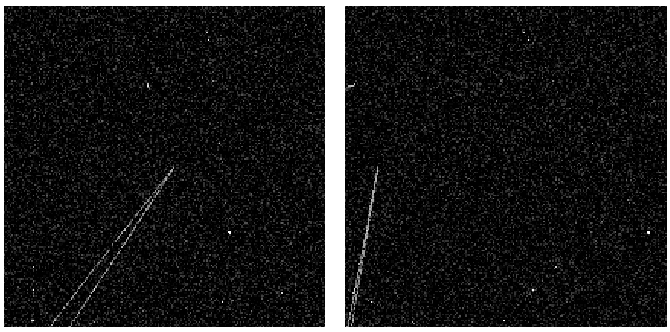
3.1. GammaNet Architecture

The XZ and YZ projection images that the AdEPT TPC produce are used as the input of two identical instances of `GammaNet` for classification. The classifications of the two projections are then compared using a boolean operation, where if either projection produced a positive signal, the event was determined to be positive. Comparing both projections helps reduce errors associated with the positron and electron tracks overlapping in a projection, which would appear as a singular track. Having the two orthogonal projections ensures that this overlap is avoided in at least one of the images provided to `GammaNet`, avoiding misclassification of pair production events. An example of this issue is shown in Figure 2, where in the first projection, the two tracks from the pair production are well separated, and the alternative projection shows them overlapping to an extent. The architecture of `GammaNet` is presented in Section 7.1.

`GammaNet`'s architecture is an adaptation of `GoogLeNet` [31] with modifications needed for reduced time to classification and the stringent background rejection requirements of AdEPT. To reduce time to classification, the overall network size was truncated by utilizing only one inception module, where an inception module is a network in network design created by Google [31]. Table 1 lists the results from `GammaNet` when operating with a threshold of 0.5 for classification of background and pair production when differing the number of inception modules. From these results it is shown that the highest GCR background rejection rate was achieved with a single inception module. It was required to implement mixed precision math in `GammaNet` in order to attain the rejection rate necessary for AdEPT. Mixed precision math was implemented by using double precision in the inner product layer and the softmax layer shown in Figure 8 m) and l).

Number of Inception Modules	Pair Production Sensitivity (%)	Background GCR Rejection Rate (%)
1	**93.17**	**96.30**
2	94.28	93.91
3	93.47	96.10

Table 1. Tabulated results of `GammaNet` pair production sensitivity and background rejection rate for differing numbers of inception modules. Pair Production sensitivity reported as highest of the 5–250 MeV energy sets whereas background rejection rate was calculated from only one set.



(a) (b)

Fig. 2. XZ and YZ projections of the same event generated in the sub-scale simulation, with a downsampling rate of 3. In a), the XZ projection, a well separated pair production track is shown in the lower half of the image. In b), the YZ projection, an overlapping pair production track is shown in the lower left of the image.

3.2. Training

The training process for `GammaNet` involves passing a simulation image through it, after which the parameters of each layer in the network are updated based on the negative gradient of that output with respect to each parameter. Training is continued until the network converges on a steady state of accuracy with respect to a testing data set that is separate from the training data. The training procedure is governed by a handful of parameters, called hyperparameters, used by `NVCaffe` to determine how training is carried out [32]. The hyperparameters used for training `GammaNet` can be seen in Section 7.2. The results of training are shown in Figure 3, and demonstrates that `GammaNet` converges upon a solution quickly while training. Training was continued for 5 million iterations for each version of `GammaNet`, and to 2 million iterations for the `VGG16` [33] architecture. The training graphs of subsequent networks were omitted for the sake of brevity, though each network reached similar results to Figure 3.

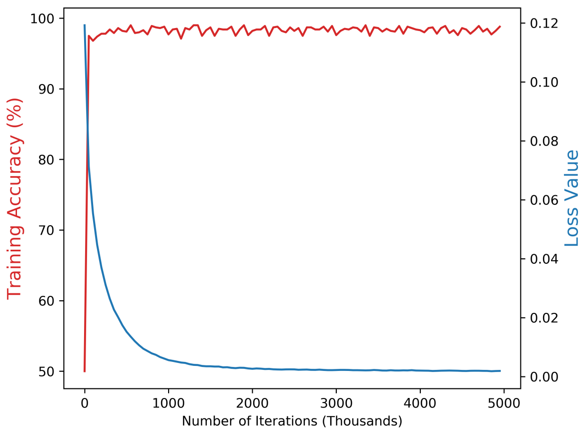
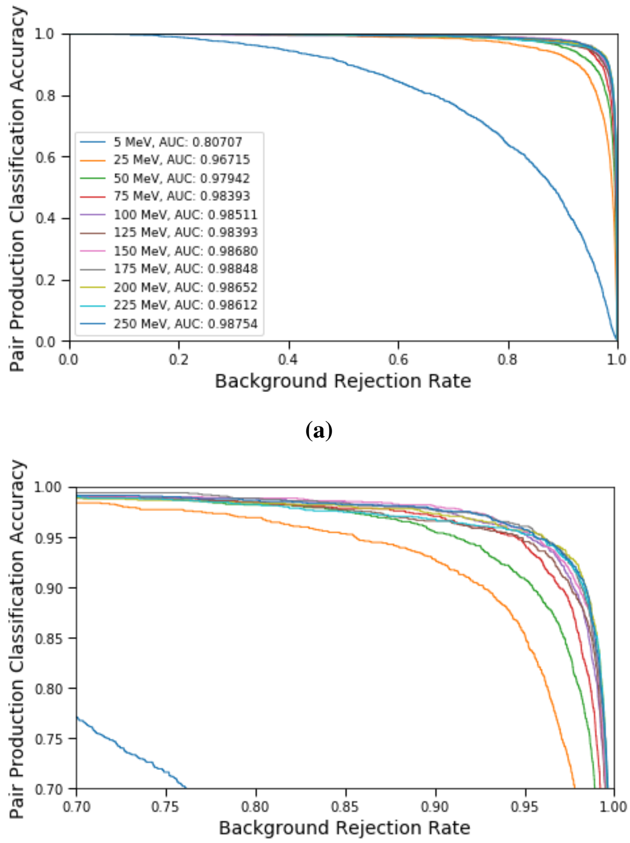


Fig. 3. Graph of the training results for `GammaNet` with 1 inception module. This training data was generated with the sub-scale simulation, with a downsampling rate of 3x. The left axis contains the accuracy of `GammaNet` on the validation data set, and the right axis contains the loss value averaged over every 50k training iterations.

4. Results

The final layer of `GammaNet`, Figure 8 m, outputs the probability that a given input image contains a gamma-ray pair production signal or is purely background. To analyze the performance of `GammaNet` as a binary classifier a ROC analysis [21], which determines a classifier's specificity and sensitivity at different threshold values, was conducted. In the ROC algorithm, the list of classification outputs produced by `GammaNet` for the image set is sorted by decreasing value of probability for the pair production event class. The threshold value is then iterated through the list of pair production class probabilities. For each iteration the classification probability for pair production produced by `GammaNet` in response to the input image is compared to the threshold. If the classification probability for pair production is lower than the threshold, the image is classified as background. If the classification probability for pair production is above threshold, the image is classified as pair production. Utilizing the threshold for classification allows for an event classified as pair production to be either a true positive or false positive event. The number of true positive and false positive events is then tallied and normalized to the number of images in the set to generate the true and false positive rates for each threshold value.

The ROC plot provides a graphic representation of the classifier's response to threshold levels for true positive and false positive rates. The ROC curve generated for `GammaNet` is shown in Figure 4. Area Under the ROC Curve (AUC) in Figure 4 ranges from 0.807 to 0.988 depending on incident gamma-ray energy, demonstrating the general level of performance of `GammaNet` as a binary classifier. The individual points on Figure 4 show the sensitivity to pair production of `GammaNet` at a given background rejection rate, which can be used to determine what threshold to run `GammaNet` at to satisfy the requirements of AdEPT for background rejection rates.



(a)

(b)

Fig. 4. a) The ROC curves generated using the described algorithm for each pair production data set as classified by `GammaNet`, using 11x downsampled images. The AUC is provided in the legend for each incident gamma-ray energy, with an area of 1 being a perfect classifier, and 0.5 being random selection. b) A subsection of Figure 4a is presented to display the nuanced features of the plot.

Using the ROC analysis, it is possible to investigate the performance of `GammaNet` with respect to the rate of downsampling used on the simulation data. To perform downsampling the output of the simulation had the spatial extent for each dimension of a voxel increased by a given multiple, N, wherein all values contained within the original voxels were summed into the new voxel. Downsampling results in the projection images being reduced by a factor of N^2, which significantly reduces the time taken to train and perform classification with `GammaNet`. The impacts of downsampling on signal sensitivity are displayed in Table 3, where downsampling rates between 1–11 were investigated using the sub-scale simulation. When downsampling by 1 the voxel size is maintained at 400 x 400 x 400 μm^3, and when downsampling by 11 the voxel size is reduced to 4.4 x 4.4 x 4.4 mm^3.

From the results in Table 3, it is shown that any amount of downsampling outperforms the alternative of no downsampling, with a decrease in signal sensitivity for increasing downsampling rates. This increase in sensitivity for any amount of downsampling is expected to be due to the original images containing discontinuities in the ionization tracks from the pair production events, this is reduced or entirely removed when downsampling the image. This gain of sensitivity is then diminished with greater degrees of downsampling as the higher rates of downsampling reduce the ability to distinguish both arms of the pair production tracks. Given the large image size generated by the full-scale simulation, a downsampling value of 11 was used for the remainder of the work when utilizing the full-scale simulation. Training of `GammaNet` on the full-scale simulation data at a downsampling rate of 11 took 30 days of compute time, proving investigating `GammaNet's` performance on lower downsampling rates with the full-scale simulation data to be prohibitively time consuming.

Background Rejection Rate (%)	`GammaNet` Pair Production Sensitivity (%)	VGG16 Pair Production Sensitivity (%)
99.990±0.002	65±2	28±2
99.97±0.003	73±2	38±2
99.94±0.005	78±2	46±2
99.87±0.007	84±2	57±2
99.81±0.009	87±2	61±2
99.75±0.01	89±2	64±2
99.69±0.01	90±1	66±2

Table 2. Pair Production sensitivity for `GammaNet` and VGG16 at varying background rejection rates corresponding to anticipated downlink speeds. Performance comparison results were generated using the sub-scale simulation data, with a downsampling rate of 3.

A cursory investigation between the performance of `GammaNet` relative to other neural network architectures was performed. In this investigation another neural network architecture was chosen, VGG16 [33], given it outperformed GoogLeNet in the ILSVRC. VGG16 was trained in the exact same manner as `GammaNet` and the performance of the two networks were compared. The performance comparison between `GammaNet` and VGG16 was carried out with a downsampling rate of 3, and with data produced from the sub-scale simulation. Table 2 provides the results from each network when classifying the sub-scale simulation data, with `GammaNet` shown to largely outperform VGG16 over the entire range of background rejection rates investigated. This result is not what would be anticipated given that VGG16 outperformed GoogLeNet in the ILSVRC competition, but the task of event classification for AdEPT utilizes more sparse images. These results imply that `GammaNet` is more well suited for classifying background images than is VGG16, which ultimately is the primary task of `GammaNet` for AdEPT.

The performance of `GammaNet` when classifying Compton scatter events was of interest as well given that it is the main gamma-ray interaction contributing to background in the AdEPT instrument, and the similarity in track structure compared to pair production. The rate of misclassification for Compton scatter events as pair production events provides information about the features that `GammaNet` uses for classifying the input. The main differentiation between the pair production and Compton scatter tracks is the presence of only a singular track for Compton scatter and the absence of the vertex from pair production. Support for the importance of these features for classification is shown in Figure 5. As the incident gamma-ray energy increases, so too does the signal sensitivity for pair production. The increase in signal sensitivity is due to the increased energy of the positron-electron pair producing more linear tracks, closer in proximity, and with more distinct vertices. This is supported by the negligible increase in signal sensitivity for Compton scatter events.

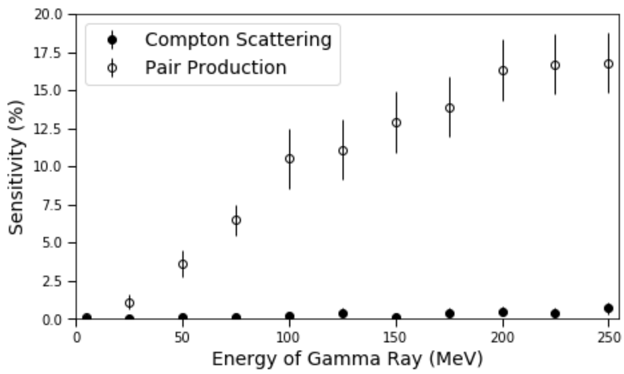


Fig. 5. Plot of the sensitivity for Compton scatter and pair production image sets as the energy of the incident gamma ray varies, using a downsampling rate of 11 on the full-scale simulation. These sensitivities were calculated using a threshold value that generated a 99.990±0.002% background rejection rate. Errors were calculated using binomial statistics with a 95% confidence interval.

Due to the large raw data rate and the limits of satellite communications, it is required to achieve a background rejection rate of 99.99% to 99.69% in order for the data to be transmitted. To achieve this background rejection rate, the threshold for a pair production event classification has to be set quite

		Downsampling Rate					
		1	3	5	7	9	11
Data Rate Limit (Mbps avg.)	Background Rejection Rate (%)	Signal Sensitivity (%)					
1.5	99.99±0.002	43±2	65±2	50±2	41±2	40±2	28±2
5	99.97±0.003	47±3	73±2	65±2	54±2	49±2	45±2
10	99.94±0.005	54±3	78±2	73±2	62±2	56±2	56±2
20	99.87±0.007	64±2	84±2	81±2	74±2	67±2	66±2
30	99.81±0.009	68±2	87±2	84±2	78±2	72±2	71±2
40	99.75±0.01	71±2	89±2	86±2	81±2	75±2	74±2
50	99.69±0.01	74±2	90±1	87±2	83±2	78±2	77±2

Table 3. Pair production sensitivity of `GammaNet`, for sub-scale simulation images, given the desired background rejection rate with differing factors of downsampling. The data rate limits are sampled between the proposed minimum and maximum as described in Section 1. The background rejection rates listed are calculated by using the ratio of the raw data rate and the data rate limit, assuming the signal is approximately entirely background. Each data set was generated from the sub-scale simulation, using the given downsampling rate. `GammaNet` was then trained and tested on those data sets. The reported pair production signal sensitivities are the average sensitivity for the energies simulated. Error was calculated using binomial statistics with a 95% confidence interval.

high, which results in a number of pair production events being misclassified as background events. Table 4 shows the average rate at which `GammaNet` classifies pair production and Compton scatter events as a positive event, given different background rejection rates. These results were generated using the full-scale simulation with a downsampling rate of 11. The classification accuracies were averaged over the energies simulated for pair production and Compton scatter. It is shown in Table 4 and Figure 5 that at the proposed 99.99% background rejection rate, we obtain a pair production sensitivity between 0.1±0.1% and 17±2%, depending on incident photon energy, with an average of 10±1%. For the best case scenario of 99.69% background rejection, the signal sensitivity increases to a range of 1.1±0.5% to 69±2%, again depending on incident photon energy, with an average of 52±2%. In both cases, the sensitivity to Compton scatter is quite small, which is beneficial for the mission due to Compton scatter representing background for the AdEPT mission. The relatively low sensitivity to pair production events at low energy will reduce the effectiveness of the instrument, but this impact can be mitigated during mission design by implementing image compression, where these calculations were done assuming no compression.

In this study, the test set of GCR protons contained 10^6 events, with twice as many images. Operating at 99.99% background rejection resulted in 100 GCR proton events being classified as positive, considered false positives events. False positives occur when at least one of two projections is classified as positive by `GammaNet`. Figure 6a-j shows 10 of the GCR proton events that resulted in false positive classifications. Figure 6k-t shows 10 pair production events that resulted in `GammaNet` producing the lowest response for pair production classification. The projection shown for the GCR proton events are the projection resulting in a positive classification, and the projections shown for the pair production events are the most representative of the characteristics resulting in false negative classification. In the false positive images, Figure 6a-j, extended delta tracks are observed with at least one point of track crossing. This observation demonstrates that `GammaNet` responds to extended contiguous tracks, and track crossings, as signals of pair production events. In addition, Figure 6a contains a pair production event occurring from a GCR proton track which results in a false positive classification, showing `GammaNet` responds significantly to the vertex of a pair production event. In the false negative images, Figure 6k-t, three features can be observed in the pair production images: short track length in Figure 6k-q, overlapping of the two tracks making it appear as a singular track in Figure 6i-o, and deep inelastic scattering events in Figure 6r-t.

Background Rejection Rate (%)	Pair Production Sensitivity (%)	Compton Scatter Sensitivity (%)
99.990±0.002	10±1	0.3±0.3
99.97±0.003	16±2	0.4±0.3
99.94±0.005	26±2	0.7±0.4
99.87±0.007	37±2	1.3±0.6
99.81±0.009	44±2	1.7±0.6
99.75±0.01	47±2	1.9±0.7
99.69±0.01	52±2	2.2±0.7

Table 4. Pair production and Compton scatter sensitivity at varying background rejection rates corresponding to anticipated downlink speeds. The GCR proton background rejection rate was calculated for one set of background images. Each data point for Compton scatter and pair production sensitivity were generated by averaging the sensitivity over all simulated gamma-ray energies. All data here were generated using the full-scale simulation with a downsampling rate of 11.

5. GammaNet Visualization

As the use of CNNs becomes more prevalent in research, it is of increasing interest how the CNN performs the classification and what features of the input it uses to do so. Grad-CAM[34] is a recent algorithm developed to answer these questions by

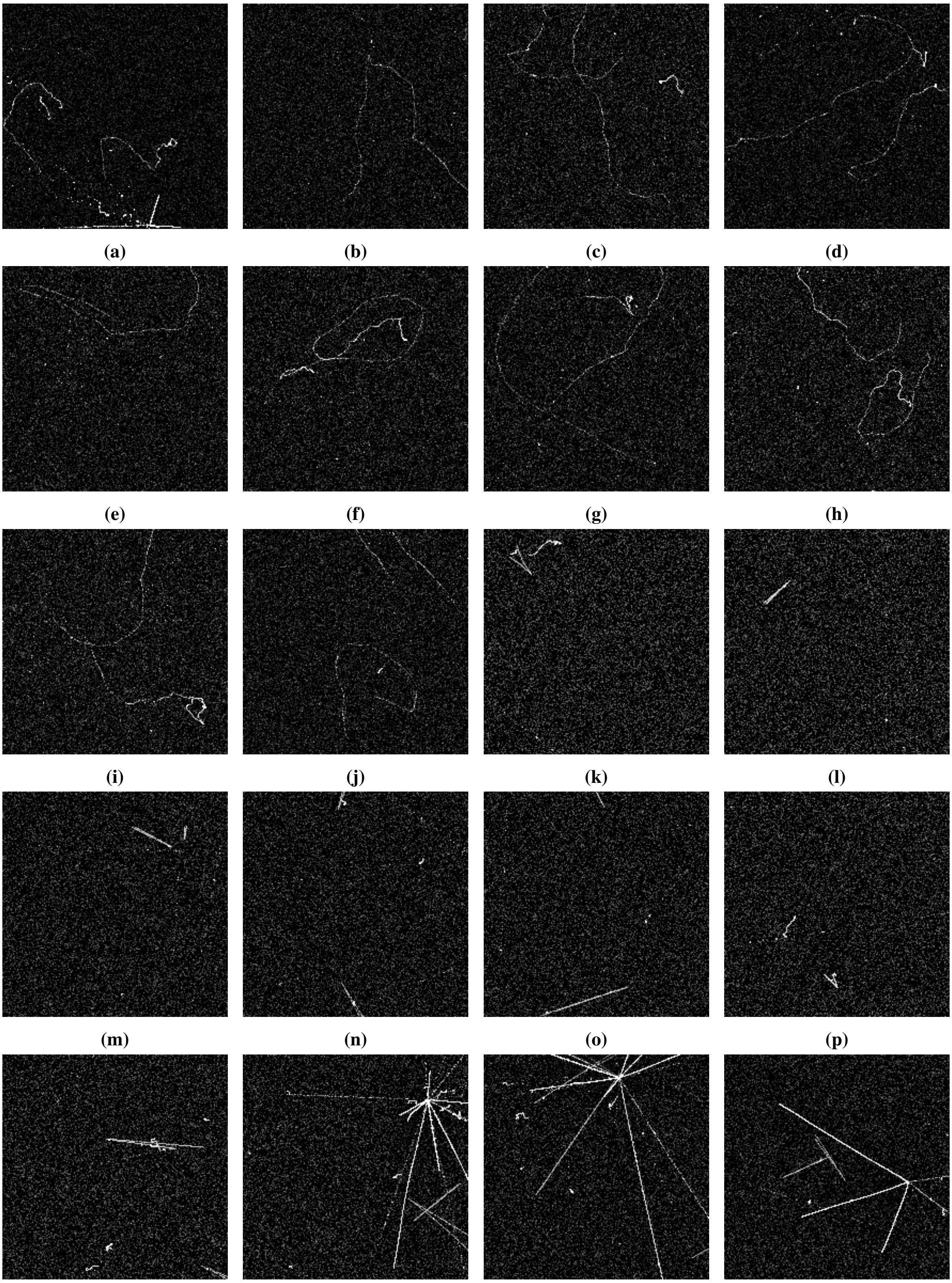


Fig. 6. a)-j) Projection images of simulated GCR proton events that resulted in false positive classifications. Only the projection image resulting in a false positive is shown, the alternate projection is not included because no event produced a false positive in both projections. k)-t) Projection images of pair production events that produced the lowest response in `GammaNet` for the pair production event class. The projection shown is most representative of the cause for false negative classification.

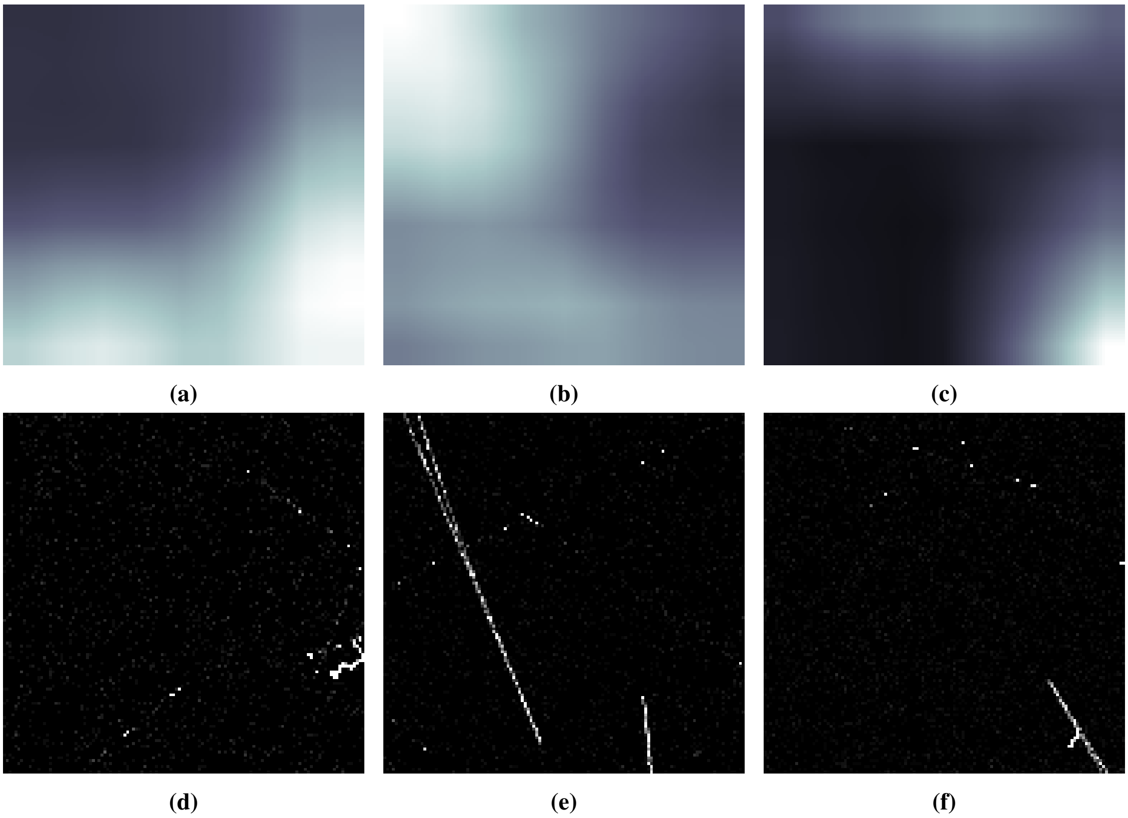


Fig. 7. a-c) Images generated by the Grad-CAM algorithm that demonstrate the features that GammaNet utilizes for classifying images as background or signal. d-f) The simulation images used to generate the respective Grad-CAM images, with d) and f) being signal events and e) being a background event.

providing an activation map for input images that shows the regions the CNN used most within the image during classification. Figure 7 shows the Grad-CAM images generated for GammaNet with one background image, Figure 7d, and two signal images, Figures 7e and 7f. These images were generated using the sub-scale simulation of AdEPT because the lower track density provides interpretable results. Figure 7a shows that for the background class, GammaNet utilized sparsely ionizing tracks and delta rays present in Figure 7d, resulting correctly in a background classification. Figure 7b demonstrates that for the signal class, GammaNet utilizes the separate, nearly parallel, tracks of the pair production event preferentially over the overlapped pair production event at the bottom of Figure 7e, resulting in an accurate positive classification. Lastly, Figure 7c results in a false background classification of the pair production image, Figure 7f, with GammaNet using the sparsely ionized GCR tracks and the delta generated from the pair production track.

6. Conclusion

The event classification requirements of the AdEPT mission dictate a background rejection rate between 99.99% and 99.69% which must be achieved within a 50 ms time window determined by the instrument collection rate. `GammaNet`, using mixed precision enabled by `NVCaffe`, was able to achieve a background rejection rate of 99.990±0.004%. These results were achieved using the full-scale simulation, classifying on images downsampled at 11x. The time for inference was found to be on average 6.8 ms utilizing a NVIDIA GTX 1080 Graphics Processing Unit (GPU), which has 8.2 TFLOPS of single precision compute performance. This implies that, as is, `GammaNet` would require 1.1 TFLOPS of single precision compute available to it from the on-board flight computer. The AdEPT mission is still in the development stages, and thus the flight computer has not been chosen. Commercially available flight computers are capable of meeting this demand. Additionally NASA is investigating the use of commercial SoC solutions that possess greater than 1 TFLOPS performance [20]. In its current iteration, `GammaNet` is not prohibitively compute intensive for use as an on-board event classifier.

These background rejection requirements have limited the sensitivity to pair production images to a range of 0.1±0.1% to 17±2% for 99.99% background rejection and 1.1±0.5% to 69±2% for 99.69% background rejection, for incident photon energies from 5–250 MeV. The low sensitivity lowers the effectiveness of the AdEPT instrument, however these values were generated using conservative estimates. These results show that `GammaNet` achieves the desired background rejections of AdEPT, making it a serious consideration for use on-board the satellite for event classification.

These performance estimates include no image compression, and downlink bandwidth afforded by current and near future satellite communication [10, 17–19]. No image compression was used as a conservative assumption due to the data handling system for the AdEPT satellite not yet being decided. Simple lossless compression afforded by the PNG format produces compression ratios nearing 2 for the simulation images used in this study. As more systems aboard the AdEPT satellite are designed and implemented, more precise determination of the operational parameters of `GammaNet` can be achieved. Reductions in the raw data rate will allow `GammaNet` to operate at a lower background rejection rate, affording increased pair production sensitivity.

Grad-CAM was implemented for `GammaNet` in order to discern the features that `GammaNet` uses during classification of the simulation images. The results from this application support the supposition that `GammaNet` utilizes features that are characteristic of background for the respective classification, such as lower ionization density relative to the pair production tracks and the presence of delta rays. For the positive class of pair production the network responds strongly to semi-parallel tracks that are close in proximity, indicative of energetic pair production events.

7. Appendix

7.1. GammaNet Architecture

The architecture used for `GammaNet` is shown schematically in Figure 8 and is comprised of convolution, Rectified Linear Unit (ReLU) [35], maximum or average pooling, Local Response Normalization (LRN) [36, 37], dropout [38], concatenation, inner product, and softmax [36, 39] operations. All of these operations come preprogrammed in `NVCaffe`, a platform for developing and programming CNNs [32], which was used for the development of `GammaNet`.

7.2. GammaNet Hyperparameters

The hyperparameters used in training `GammaNet` are as follows:

test_iter: 1000
test_interval: 50000
base_lr: 0.0001
display: 1000
max_iter: 10000000
lr_policy: "step"
gamma: 0.96
momentum: 0.9
weight_decay: 0.0002
stepsize: 320000
snapshot: 49000
snapshot_prefix: "/path/to/your_prefered_directory"
solver_mode: GPU
net: "/path/to/your_network.prototxt"
test_initialization: false
average_loss: 40
iter_size: 1

Acknowledgements

This work has received support from the National Sciences and Engineering Research Council of Canada through research grants to S.H.B. Support for A.R.H was given by an appointment to the NASA Postdoctoral Program at the Goddard Space Flight Center administered by Oak Ridge Associated Universities through a contract with NASA. We gratefully acknowledge the support of NVIDIA Corporation with the donation of a GPU which was used to accelerate computations for `GammaNet` in this research.

References

[1] S. Hoo-Chang, H. R. Roth, M. Gao, L. Lu, Z. Xu, I. Nogues, J. Yao, D. Mollura, R. M. Summers, Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning, IEEE transactions on medical imaging 35 (5) (2016) 1285.

[2] C. Szegedy, S. Ioffe, V. Vanhoucke, A. A. Alemi, Inception-v4, inception-resnet and the impact of residual connections on learning, in: Thirty-first AAAI conference on artificial intelligence, 2017.

[3] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, K. Keutzer, Squeezenet: Alexnet-level accuracy with 50x fewer parameters and 0.5 mb model size, arXiv preprint arXiv:1602.07360.

[4] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, International Journal of Computer Vision (IJCV) 115 (3) (2015) 211–252.

[5] P. Baldi, P. Sadowski, D. Whiteson, Searching for exotic particles in high-energy physics with deep learning, Nature communications 5 (2014) 4308.

[6] P. Baldi, K. Bauer, C. Eng, P. Sadowski, D. Whiteson, Jet substructure classification in high-energy physics with deep neural networks, Physical Review D 93 (9) (2016) 094034.

[7] A. Aurisano, A. Radovic, D. Rocco, A. Himmel, M. Messier, E. Niner, G. Pawloski, F. Psihas, A. Sousa, P. Vahle, A convolutional neural network neutrino event classifier, Journal of Instrumentation 11 (09) (2016) P09001.

[8] S. D. Hunter, P. F. Bloser, G. O. Depaola, M. P. Dion, G. A. DeNolfo, A. Hanu, M. Iparraguirre, J. Legere, F. Longo, M. L. McConnell, et al., A pair production telescope for medium-energy gamma-ray polarimetry, Astroparticle Physics 59 (2014) 18–28.

[9] M. Tavani, G. Barbiellini, A. Argan, F. Boffelli, A. Bulgarelli, P. Caraveo, P. Cattaneo, A. Chen, V. Cocco, E. Costa, et al., The AGILE mission, Astronomy & Astrophysics 502 (3) (2009) 995–1013.

[10] W. Atwood, A. A. Abdo, M. Ackermann, W. Althouse, B. Anderson, M. Axelsson, L. Baldini, J. Ballet, D. Band, G. Barbiellini, et al., The large area telescope on the Fermi gamma-ray space telescope mission, The Astrophysical Journal 697 (2) (2009) 1071.

[11] F. Lei, A. Dean, G. Hills, Compton polarimetry in gamma-ray astronomy, Space Science Reviews 82 (3-4) (1997) 309–388.

[12] M. Forot, P. Laurent, I. Grenier, C. Gouiffès, F. Lebrun, Polarization of the crab pulsar and nebula as observed by the integral/ibis telescope, The Astrophysical Journal Letters 688 (1) (2008) L29.

[13] D. Bernard, P. Bruel, M. Frotin, Y. Geerebaert, B. Giebels, P. Gros, D. Horan, M. Louzir, P. Poilleux, I. Semeniouk, et al., HARPO: a TPC as a gamma-ray telescope and polarimeter, in: SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics, 2014, pp. 91441M–91441M.

[14] A. Takada, H. Kubo, H. Nishimura, K. Ueno, K. Hattori, S. Kabuki, S. Kurosawa, K. Miuchi, E. Mizuta, T. Nagayoshi, et al., Observation of diffuse cosmic and atmospheric gamma rays at balloon altitudes with an electron-tracking Compton camera, The Astrophysical Journal 733 (1) (2011) 13.

[15] K. Ueno, T. Mizumoto, K. Hattori, N. Higashi, S. Iwaki, S. Kabuki, Y. Kishimoto, S. Komura, H. Kubo, S. Kurosawa, et al., Development of the balloon-borne sub-MeV gamma-ray Compton camera using an

electron-tracking gaseous TPC and a scintillation camera, Journal of Instrumentation 7 (01) (2012) C01088.

[16] C. Martoff, R. Ayad, M. Katz-Hyman, G. Bonvicini, A. Schreiner, Negative ion drift and diffusion in a TPC near 1 bar, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment 555 (1-2) (2005) 55–58.

[17] C. Meegan, G. Lichti, P. Bhat, E. Bissaldi, M. S. Briggs, V. Connaughton, R. Diehl, G. Fishman, J. Greiner, A. S. Hoover, et al., The Fermi gamma-ray burst monitor, The Astrophysical Journal 702 (1) (2009) 791.

[18] R. A. Cameron, Fermi large area telescope operations: progress over 4 years, in: SPIE Astronomical Telescopes+ Instrumentation, International Society for Optics and Photonics, 2012, pp. 84481J–84481J.

[19] B. Robinson, D. Boroson, C. Schieler, F. Khatri, O. Guldner, S. Constantine, T. Shih, J. Burnside, B. Bilyeu, F. Hakimi, et al., TeraByte InfraRed Delivery (TBIRD): a demonstration of large-volume direct-to-Earth data transfer from low-Earth orbit, in: Free-Space Laser Communication and Atmospheric Propagation XXX, Vol. 10524, International Society for Optics and Photonics, 2018, p. 105240V.

[20] W. Powell, M. Campola, T. Sheets, A. Davidson, S. Welsh, Commercial off-the-shelf gpu qualification for space applications.

[21] T. Fawcett, An introduction to ROC analysis, Pattern recognition letters 27 (8) (2006) 861–874.

[22] S. Agostinelli, J. Allison, K. a. Amako, J. Apostolakis, H. Araujo, P. Arce, M. Asai, D. Axen, S. Banerjee, G. . Barrand, et al., GEANT4 a simulation toolkit, Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment 506 (3) (2003) 250–303.

[23] J. Allison, K. Amako, J. Apostolakis, H. Araujo, P. A. Dubois, M. Asai, G. Barrand, R. Capra, S. Chauvie, R. Chytracek, et al., Geant4 developments and applications, IEEE Transactions on nuclear science 53 (1) (2006) 270–278.

[24] A. R. Hanu, G4AdEPTSim, `https://github.com/AndreiHanu/G4AdEPTSim/releases` (2018).

[25] G. D. Badhwar, The radiation environment in low-earth orbit, Radiation research 148 (5s) (1997) S3–S10.

[26] E. R. Benton, E. Benton, Space radiation dosimetry in low-earth orbit and beyond, Nuclear Instruments and Methods in Physics Research Section B: Beam Interactions with Materials and Atoms 184 (1-2) (2001) 255–294.

[27] D. Zhou, D. OSullivan, E. Semones, W. Heinrich, Radiation field of cosmic rays measured in low earth orbit by cr-39 detectors, Advances in Space Research 37 (9) (2006) 1764–1769.

[28] G. Weidenspointner, M. Varendorff, S. Kappadath, K. Bennett, H. Bloemen, R. Diehl, W. Hermsen, G. Lichti, J. Ryan, V. Schönfelder, The cosmic diffuse gamma-ray background measured with comptel, in: AIP Conference Proceedings, Vol. 510, American Institute of Physics, 2000, pp. 467–470.

[29] R. C. Henry, Diffuse background radiation, The Astrophysical Journal Letters 516 (2) (1999) L49.

[30] D. Ayres, N. Collaboration, et al., NOvA proposal to build a 30 kiloton off-axis detector to study neutrino oscillations in the Fermilab NuMI beamline, arXiv preprint hep-ex/0503053.

[31] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 1–9.

[32] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, T. Darrell, Caffe: Convolutional architecture for fast feature embedding, in: Proceedings of the 22nd ACM international conference on Multimedia, 2014, pp. 675–678.

[33] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv preprint arXiv:1409.1556.

[34] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, D. Batra, Grad-CAM: Visual explanations from deep networks via gradient-based localization, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 618–626.

[35] V. Nair, G. E. Hinton, Rectified linear units improve restricted boltzmann machines, in: Proceedings of the 27th international conference on machine learning (ICML-10), 2010, pp. 807–814.

[36] A. Krizhevsky, I. Sutskever, G. E. Hinton, Imagenet classification with deep convolutional neural networks, in: Advances in neural information processing systems, 2012, pp. 1097–1105.

[37] G. E. Hinton, N. Srivastava, A. Krizhevsky, I. Sutskever, R. R. Salakhutdinov, Improving neural networks by preventing co-adaptation of feature detectors, arXiv preprint arXiv:1207.0580.

[38] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov, Dropout: A simple way to prevent neural networks from overfitting, The Journal of Machine Learning Research 15 (1) (2014) 1929–1958.

[39] S. Lawrence, C. L. Giles, A. C. Tsoi, A. D. Back, Face recognition: A convolutional neural-network approach, IEEE transactions on neural networks 8 (1) (1997) 98–113.

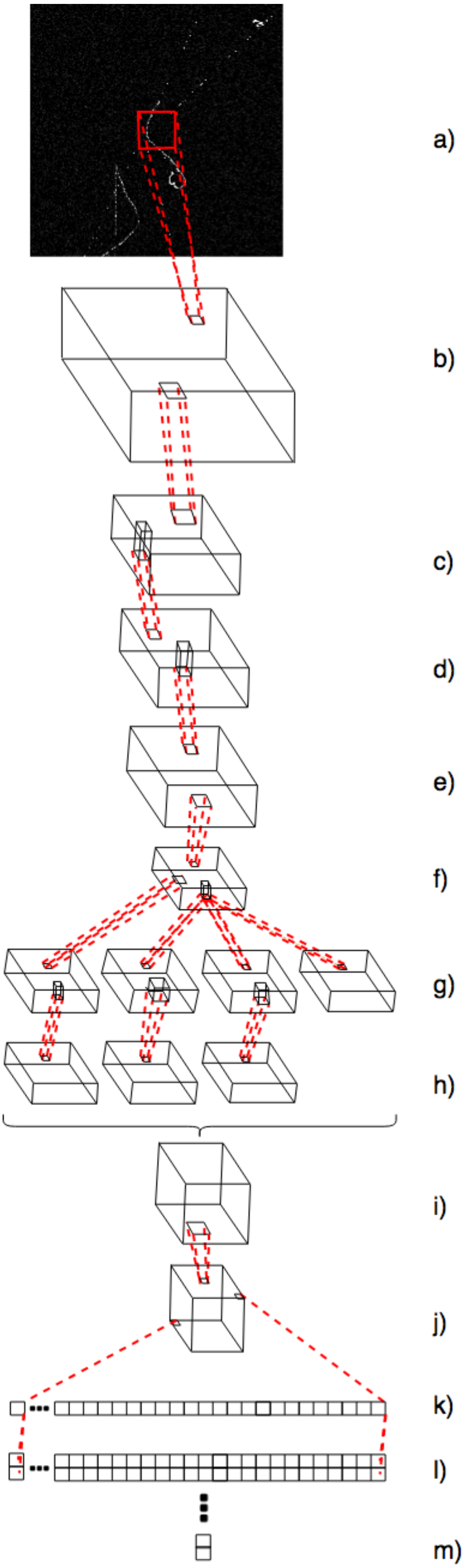


Fig. 8. Diagram depicting the architecture and layers used for `GammaNet`. All functions depicted in this diagram are from the preprogrammed operations included in the `NVCaffe` library.

a) input to the network of an AdEPT simulation image.

b) first convolution layer made of a 7x7 convolution with a stride of 2, where stride is the spacing between the center of successive convolutions performed on the previous layer. The convolution is followed by a ReLU operation, where all negative values are made to be 0.

c) 3x3 max pooling layer with a stride of 2, where max pooling takes a subset of the previous layer and outputs the maximum value from that subset. The 3x3 max pooling is followed by a LRN operation, where the values of the max pooling output are normalized along the depth of the output.

d) 1x1 convolution with stride of 1 followed by a ReLU.

e) 3x3 convolution with a stride of 1 followed by a ReLU and LRN operation.

f) 3x3 max pooling layer with a stride of 2.

g) inception module used by `GoogLeNet` [31], part 1, from top to bottom is: 3 1x1 convolutions of stride 1 and a 3x3 max pooling with a stride of 1.

h) inception module used by `GoogLeNet` [31], part 2, from top to bottom is a 3x3 convolution of stride 1, a 5x5 convolution of stride 1, and a 1x1 convolution of stride 1.

i) concatenation along the depth of the previous 3 operations in h), where the separate outputs are combined into one 3 dimensional matrix.

j) 7x7 average pooling with a stride of 1, where average pooling takes a subset of the previous layer and provides the average value for an output. The average pooling is followed by a dropout operation, where randomly some values in the output are set to 0 with a programmed probability.

k) the flattening of j) into a vector.

l) inner product between vector k) and the parameters of l) where there is a set of parameters for each class contained in the output, with two classes in the case of `GammaNet`. The parameters of l) were stored in double precision and the inner product calculated using double precision.

m) 2 values output by the softmax operation, which takes the output of the inner product layer as an input for the softmax function. The softmax function provides the probability that the original input image belongs to each class of the network, pair production or background for `GammaNet`. The softmax function was calculated using double precision and its results were also produced with double precision.