A citizen-science approach to muon events in imaging atmospheric Cherenkov telescope data: the Muon Hunter

Q. Feng[*][a] **for the VERITAS Collaboration,**[†] **J. Jarvis**[b]

[a]*Physics Department, McGill University, Montreal, QC H3A 2T8, Canada*
E-mail: qi.feng2@mcgill.ca

[b] *ASTERICS, DECS Workpackage, School of Physical Sciences, The Open University, Milton Keynes MK7 6AA, UK*

Event classification is a common task in gamma-ray astrophysics. It can be treated with rapidly-advancing machine learning algorithms, which have the potential to outperform traditional analysis methods. However, a major challenge for machine learning models is extracting reliably labelled training examples from real data. Citizen science offers a promising approach to tackle this challenge.

We present "Muon Hunter", a citizen science project hosted on the Zooniverse platform, where VERITAS data are classified multiple times by individual users in order to select and parameterize muon events, a product from cosmic ray induced showers. We use this dataset to train and validate a convolutional neural-network model to identify muon events for use in monitoring and calibration. The results of this work and our experience of using the Zooniverse are presented.

35th International Cosmic Ray Conference - ICRC2017
12-20 July, 2017
Bexco, Busan, Korea

---

[*]Speaker.

[†]veritas.sao.arizona.edu

1. Introduction

In the past decade, our understanding of the very-high-energy (VHE; $100\,\mathrm{GeV} \lesssim E_\gamma \lesssim 100\,\mathrm{TeV}$) gamma-ray sky has greatly progressed through the use of imaging atmospheric Cherenkov telescopes (IACTs). IACTs image the Cherenkov light of an extensive air shower induced by an incident VHE gamma-ray photon or a cosmic-ray (CR) particle. The air-shower images are then analyzed to reconstruct the information of the incident photons or the CR particles, of which the latter comprise a substantial background in VHE gamma-ray astronomy. The ability to separate gamma rays from CR particles is important, as it is directly related to the sensitivity of the instrument. Such separation is performed based on the morphology of the shower images, and can be formalized as a computer vision classification task.

Classification is a common task in experimental physics. Usually, a few key parameters are chosen, and cuts determined from Monte Carlo simulations are applied to the chosen parameters [e.g. 1]. More recently, the use of machine learning algorithms has become increasingly popular in many branches of physics, including VHE gamma-ray astrophysics to improve event classification [e.g. 2, 3, 4, 5]. One powerful machine learning algorithm, convolutional neural networks (CNN) [6], was used on a small batch of VHE gamma-ray data to detect and characterize muon events [5]. Muons are secondary products in CR-particle air showers that propagate to the ground and emit Cherenkov light, detected as rings or arcs by IACTs. The number of Cherenkov photons from a muon reaching the reflector of a telescope can be estimated from the Cherenkov angle and the traveling direction of the muon, and the distance between the telescope and where the muon impact the ground. Therefore, muon events can be used as a calibration source for the throughput of IACTs [e.g. 7, 8].

For developing supervised machine learning algorithms like the CNN, it is essential to provide correct labels of a large training dataset. This was done using a standard analysis in [5]. In this work, we describe the Muon Hunter[1], a citizen-science project where volunteers label and parameterize muon and non-muon images in VHE gamma-ray data.

2. The VERITAS array and a standard muon analysis

The Very Energetic Radiation Imaging Telescope Array System (VERITAS) is an array of four IACTs located at the Fred Lawrence Whipple Observatory in southern Arizona [30°40'N 110°57'W, 1.3 km a.s.l.; 9]. It is sensitive to gamma rays in the energy range from 85 GeV to >30 TeV with an energy resolution of ∼15% (at 1 TeV). Each of the four telescopes is equipped with a 12-m diameter reflector comprising 345 identical mirror facets, and a 499-pixel photomultiplier tube (PMT) camera covering a field of view of 3.5°. The array has a 68%-containment gamma-ray angular resolution at 1 TeV of ∼0.1°.

Coincident signals from at least two out of the four telescopes are required to trigger an array-wide read-out of the PMT signals, which occurs at a typical rate of ∼400 Hz. Most of these triggers come from CR showers or night sky background noise. For comparison, the brightest steady VHE source, the Crab Nebula, is typically observed by VERITAS at a rate of <15 gamma rays per minute. Muons are produced from charged pion decays in CR hadronic showers and

---

[1]www.muonhunter.org

are background for gamma-ray astrophysics. Single muons reaching ground level usually only trigger one telescope and produce ring images. Therefore the multi-telescope trigger system greatly reduces the muon trigger rate. There are still muon events in the data because the array can be triggered by sub-showers or energetic muons impacting the ground between two telescopes.

We randomly selected 16 observations, each of which is 30 minutes in duration, and analyzed them using one of the standard VERITAS data analysis packages, named *VEGAS* [10]. The procedure for the *VEGAS* analysis of muon events is described as follows [see also 5]:

(1) calculate the brightness-weighted average coordinates of the image (see e.g. Figure 2), and use them as the initial muon-ring center;

(2) calculate the mean ($\bar{r}$) and the variance (σ_r^2) of the distances between all image pixels and the initial centroid, and then use $\bar{r}$ as the initial muon radius;

(3) move the initial centroid by a small step, repeat step (2) and check if the variance σ_r^2 decreases; if so, update the centroid, $\bar{r}$ and the variance σ_r^2;

(4) repeat step (3) to cover a predefined grid around the initial centroid, and return the optimal centroid and radius that minimize σ_r^2;

(5) check if >70% of the pixels fall into a predefined accepted annulus e.g. $(\bar{r}-1.5\sigma_r, \bar{r}+1.5\sigma_r)$; if so, accept this event as a muon event.

A double-pass method based on the above procedure was used in the analysis, and dead or stuck pixels were corrected for. We used the analysis described above as one way to label muon events for the training/test data to build a CNN model, with an additional requirement that $\bar{r} \geqslant 0.5°$. We note that an independent muon analysis using the Hough transform [8] is available but was not used in this work.

3. The Muon Hunter project

The Muon Hunter experiment was developed in collaboration with the ASTERICS Horizon2020 project[2], and the website is hosted by the Zooniverse[3] platform, where researchers in many disciplines can easily publish a citizen-science project through the Zooniverse Project Builder[4]. The key components of a Zooniverse project are data and workflows. As a sequence of specific tasks, the workflows are tailored by the researchers to gain insights on the data. The workflows are streamlined and optimized balancing the experience of the volunteers with the need to obtain accurate results.

About 137,000 VERITAS single-telescope images were served on the Muon Hunter website, most of which are preprocessed using the standard two-level cleaning[10] based on the signal-to-noise ratio of each pixel. A subset of uncleaned images were also uploaded to explore the effect of image cleaning on the results of classification. We received a total of ∼2.1 million classifications, half within the first week after the official launch of the project, from 5,734 volunteers.

---

[2]www.asterics2020.eu/
[3]www.zooniverse.org
[4]www.zooniverse.org/labs

The workflow for identifying the presence of rings in an image is illustrated in Figure 1. We retired an image once a total of 15 volunteers had examined it and finished the workflow. It was possible to draw multiple rings on one image, as sometimes more than one muon was recorded in an image. But this also allowed room for human error, especially when a user was unfamiliar with the workflow. After a user was done with the workflow of a subject image, an option to discuss it in the Talk board was presented. The Talk board allows interactions between experts and volunteers regarding specific images. Collections of interesting images are also established by users, including double muon rings and composite images with a muon ring and a shower. Moreover, to help new users, a short tutorial, a mini course, as well as a detailed About page were available.

Figure 2 illustrates a few typical comparisons between the ring we obtained from the VEGAS analysis and user input. For a clean image with a single muon ring and no other bright components, as illustrated by the left image, the decision and the best ring drawn by VEGAS and volunteers typically agree. For an image with a muon ring and a CR shower as illustrated by the center image, a human is able to isolate and correctly locate the ring, while VEGAS simultaneously tries to fit the shower component and the actual ring with a single ring. This results in an incorrect ring and may affect subsequent analysis. Sometimes, volunteers behave erratically as illustrated by the right image, and provide incorrect input that can be viewed as outliers, which affects the mean and standard deviation of all user input for a given image. Based on the number of votes, we estimate that roughly 0.8% of all votes, and 3% of those votes for the presence of a ring in an image are outliers.
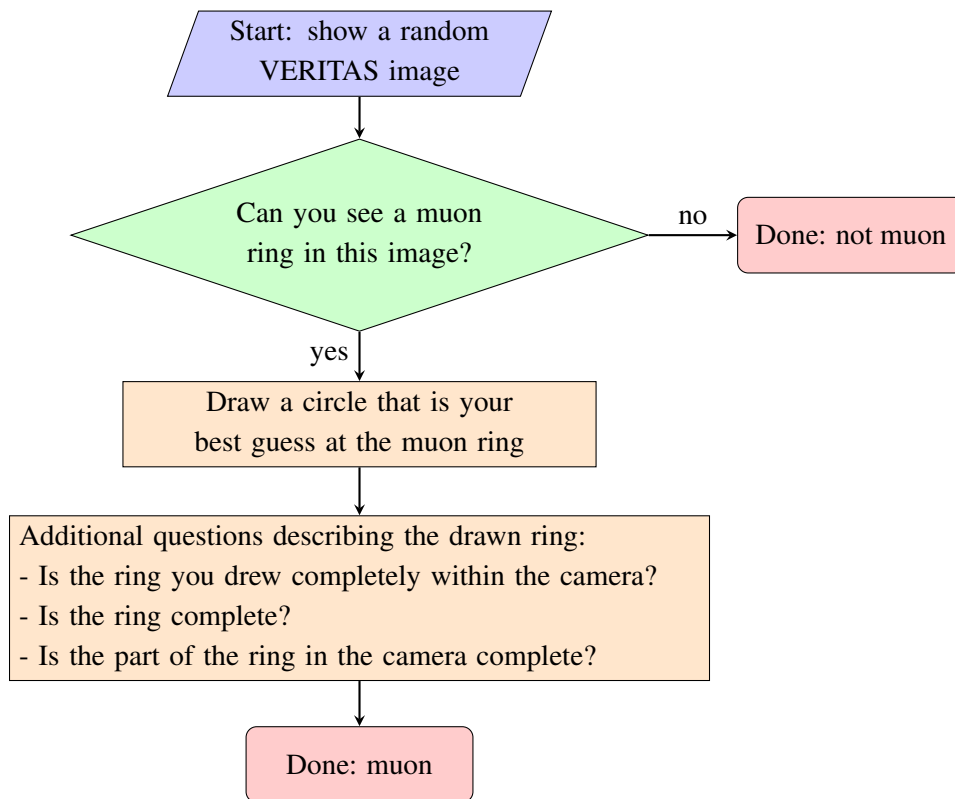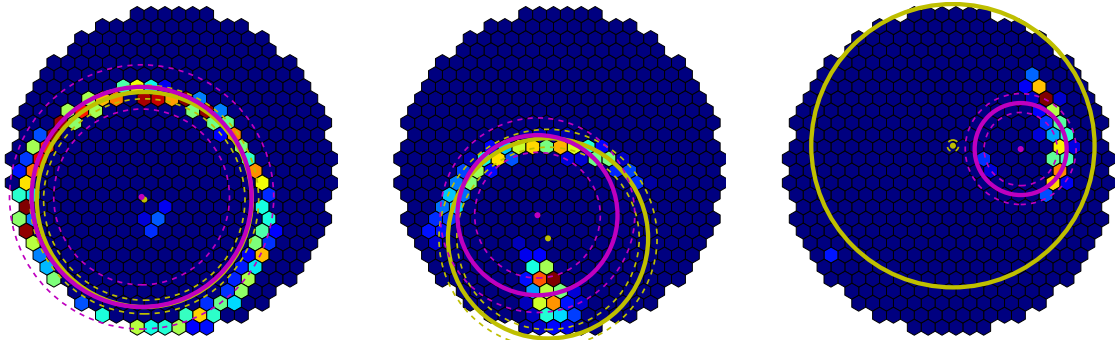

Figure 1: An illustration of the workflow of Muon Hunter.

Figure 2: Examples of VERITAS images with muon rings illustrating the difference between VEGAS analysis and volunteers' opinions. The VEGAS analysis results are shown in magenta, and the Muon Hunter volunteer input is shown in yellow. The circles with solid lines and dashed lines show the mean and the standard deviation of the radius of the ring.

Figure 3 summarizes the input we received from the volunteers of Muon Hunter. The distribution of the number of classifications each user made roughly follows a log-normal distribution (shown as the red dashed curve), with a median of 30 images per user. Assuming the number of classifications from one user is roughly proportional to the time spent, this log-normal distribution is of similar nature to the dwell time of internet users on social media articles [11]. There are 16 volunteers who classified more than 10,000 images, while there are 724 volunteers who only classified one image.

The 15 votes for each image allowed us to estimate the confidence of the users' classifications. The distribution of the fraction of votes for the presence of a ring in each image is shown in the right plot of Figure 3. 85% of the images received unanimous votes from 15 volunteers, 11% of which were muon events and 73% of which were non-muon events. We chose to label all images with 10 or more votes for muons as muon events, and the rest as non-muon events as a second set of training/test data for building a new CNN model.
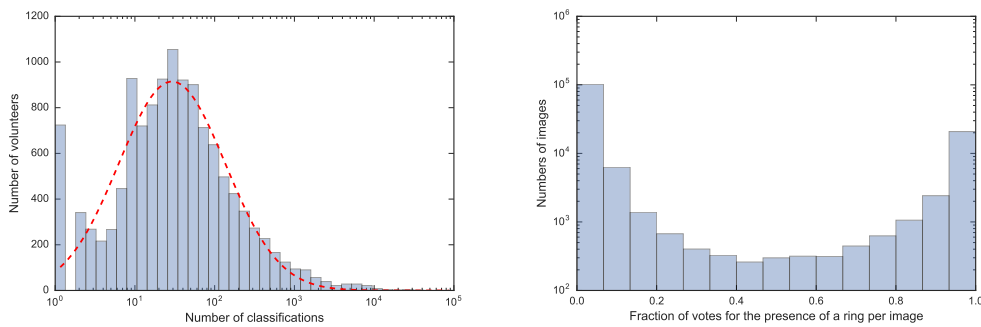
Figure 3: Distributions of the input from Muon Hunter users. The left plot shows the distribution of the number of classifications each user made. The right plot shows the distribution of votes for muons each image received.

4. The CNN classification model: training and evaluation

One purpose of this project is to train a reliable CNN model to classify muon rings. Two sources of labels, provided by the VEGAS analysis and by the Muon Hunter user input, can be used for the training, validation, and testing of the models. A detailed description of the CNN model can be found in [5], and is summarized below.

The preprocessing of VERITAS data includes standard image cleaning followed by image oversampling, which we use to approximately convert an image from its original hexagonal coordinates (due to the geometry of the PMT layout) into square coordinates. This converts a 499-pixel hexagonal image into a 54×54 pixel square image, and stretches the image by roughly 15%.

The oversampled 54×54 pixel images of the muon events and the background events are then used as input features into a CNN model, which is implemented using the *keras* Python deep learning library [12] with the backend *TensorFlow* [13]. The structure of the CNN model is a simplified "VGG"-style model [14], with only three layers of small filters, average pooling and dropout in between filter layers, and a two-layer fully-connected neural network classification model after the convolutional layers.

Treating all images with 10 or more votes for muons as muon events, we were able to train a CNN model with a test accuracy of ∼97%, while the best model using VEGAS labels whose test accuracy was ∼95%. We note that the performance of the previous CNN model [5] trained on a smaller set of data worsened when tested with the new, larger dataset, indicating possible overtraining in the previous model.

5. Summary and outlook

We received a phenomenal response from volunteers to the Muon Hunter project. The input from volunteers helped us gain insight into where the standard analysis is lacking, and train an updated machine learning model using convolutional neural networks.

We are working on isolating images with only one muon ring without other components (e.g. cosmic-ray shower) based on the input from users, and using them as a set of less noisy training examples to further improve the performance of the classification model. The single-muon images only have one relevant radius value, and therefore allow us to train a regression model to predict the radius of the muon ring. More questions will be designed for a future workflow to augment this purpose (e.g. whether a muon image also contains a cosmic ray shower image). Different criteria for converting the user input into a set of labels for the training of the CNN model are also being explored. By introducing a more strict cut on the number of users agreeing on a given image being muon or non-muon, a less ambiguous but smaller data set can be used.

Acknowledgments

References

[1] A. M. Hillas. *Cerenkov light images of EAS produced by primary gamma*. In: *International Cosmic Ray Conference*. Ed. by F. C. Jones. Vol. 3. International Cosmic Ray Conference. Aug. 1985, pp. 445–448.

[2] J. Albert et al. *Implementation of the Random Forest method for the Imaging Atmospheric Cherenkov Telescope MAGIC*. In: *Nuclear Instruments and Methods in Physics Research A* 588 (Apr. 2008), pp. 424–432. DOI: `10.1016/j.nima.2007.11.068`. arXiv: `0709.3719`.

[3] F. Acero et al. *Detection of Gamma Rays from a Starburst Galaxy*. In: *Science* 326 (Nov. 2009), pp. 1080–. DOI: `10.1126/science.1178826`. arXiv: `0909.4651 [astro-ph.HE]`.

[4] M. Krause, E. Pueschel, and G. Maier. *Improved γ/hadron separation for the detection of faint γ-ray sources using boosted decision trees*. In: *Astroparticle Physics* 89 (Mar. 2017), pp. 1–9. DOI: `10.1016/j.astropartphys.2017.01.004`. arXiv: `1701.06928 [astro-ph.IM]`.

[5] Q. Feng, T. T. Y. Lin, and the VERITAS Collaboration. *The analysis of VERITAS muon images using convolutional neural networks*. In: *IAU Symposium*. Vol. 325. IAU Symposium. June 2017, pp. 173–179. DOI: `10.1017/S1743921316012734`. arXiv: `1611.09832 [astro-ph.IM]`.

[6] A. Krizhevsky, Ilya Sutskever, and Geoffrey E. Hinton. *ImageNet Classification with Deep Convolutional Neural Networks*. In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105.

[7] D. Hanna. *Calibration Techniques for VERITAS*. In: *International Cosmic Ray Conference* 3 (2008), pp. 1417–1420. arXiv: `0709.4479`.

[8] J. Tyler and for the VERITAS Collaboration. *Muon Identification with VERITAS using the Hough Transform*. In: *ArXiv e-prints* (July 2013). arXiv: `1307.8361 [astro-ph.IM]`.

[9] J. Holder. *VERITAS: Status and Highlights*. In: *International Cosmic Ray Conference* 12 (2011), p. 137. DOI: `10.7529/ICRC2011/V12/H11`. arXiv: `1111.1225 [astro-ph.HE]`.

[10] P. Cogan. *VEGAS, the VERITAS Gamma-ray Analysis Suite*. In: *International Cosmic Ray Conference*. Vol. 3. International Cosmic Ray Conference. 2008, pp. 1385–1388. arXiv: `0709.4233`.

[11] Peifeng Yin et al. *Silence is also evidence: interpreting dwell time for recommendation from psychological perspective.* In: *KDD*. Ed. by Inderjit S. Dhillon et al. ACM, 2013, pp. 989–997. ISBN: 978-1-4503-2174-7. URL: `http://dblp.uni-trier.de/db/conf/kdd/kdd2013.html#YinLLW13`.

[12] François Chollet et al. *Keras.* `https://github.com/fchollet/keras`. 2015.

[13] M. Abadi et al. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems.* In: *ArXiv e-prints* (Mar. 2016). arXiv: `1603.04467 [cs.DC]`.

[14] K. Simonyan and A. Zisserman. *Very Deep Convolutional Networks for Large-Scale Image Recognition.* In: *ArXiv e-prints* (Sept. 2014). arXiv: `1409.1556 [cs.CV]`.

PoS(ICRC2017)826