

1 Improving Gamma-ray Source Searches with Image 2 Processing

3 HAWC Collaboration

4 (a complete list of authors can be found at the end of the proceedings)

5 *E-mail:* rbabu@icecube.wisc.edu, itomura@msu.edu, herzogia@msu.edu,
6 salaza82@msu.edu, nisamehr@msu.edu

7 The High Altitude Water Cherenkov (HAWC) observatory surveys the very-high-energy sky in the energy range from 300 GeV to greater than 100 TeV. Its wide field of view makes it particularly suitable for studying large, extended regions of emission that often contain multiple point and diffuse sources of gamma rays. Existing blind search methods to detect sources in HAWC data use a computationally expensive, iterative source fitting algorithm that can take several days to scan a few degree region in the sky. In this work, we adopt a new approach to speed up the identification of sources using an image processing pipeline. Using image processing filters, and a blob finder algorithm, based on the Determinant of a Gaussian, this pipeline seeds sources accurately up to 300 times faster than the current HAWC source search pipeline. The pipeline's output are then passed onto a global multi-threaded likelihood fitter for accurate source localization. We present the performance of the improved pipeline and discuss prospects for future applications on other astrophysical datasets.

Corresponding authors: Rishi Babu¹, Palmer Wentworth¹, Ian Herzog^{1*}, Dan Salazar¹ Mehr Un Nisa¹

¹ *Department of Physics & Astronomy, Michigan State University, Lansing, MI, USA*

* Presenter

39th International Cosmic Ray Conference (ICRC2025)
15–24 July 2025
Geneva, Switzerland



ICRC 2025
The Astroparticle Physics Conference
Geneva July 15-24, 2025

1. Introduction

The analysis of HAWC data is done using a likelihood-based systematic multi-source search procedure utilizing the threeML software package and the HAL Plugin [1][2]. This approach was inspired by the extended source search described in the Fermi-LAT Extended Source Search Catalog [3]. The pipeline to search for point sources and extended sources within the region of interest (ROI) is described in [4]. This procedure is computationally intensive and often requires multiple days to produce a final model for a region. Often this approach misses fainter sources, which need to be added manually later. This blind search could be complemented by providing a method to seed source locations, which can be assessed and evaluated by likelihood analysis, thereby significantly reducing the computational time and resources spent on finding a model.

In this work, we present the results of an image processing algorithm, using the `scikit-image` [5] python-package, to seed sources from a gamma-ray significance image. Additionally, we show the performance metrics of the image processing pipeline with 60 simulations and compare the speed of the pipeline with conventional likelihood source search method. Finally, we discuss the applications of the image processing algorithm in the context of source seeding in astronomical images, also by extending the application to Fermi-LAT, and potentially with future public LHAASO and SWGO data.

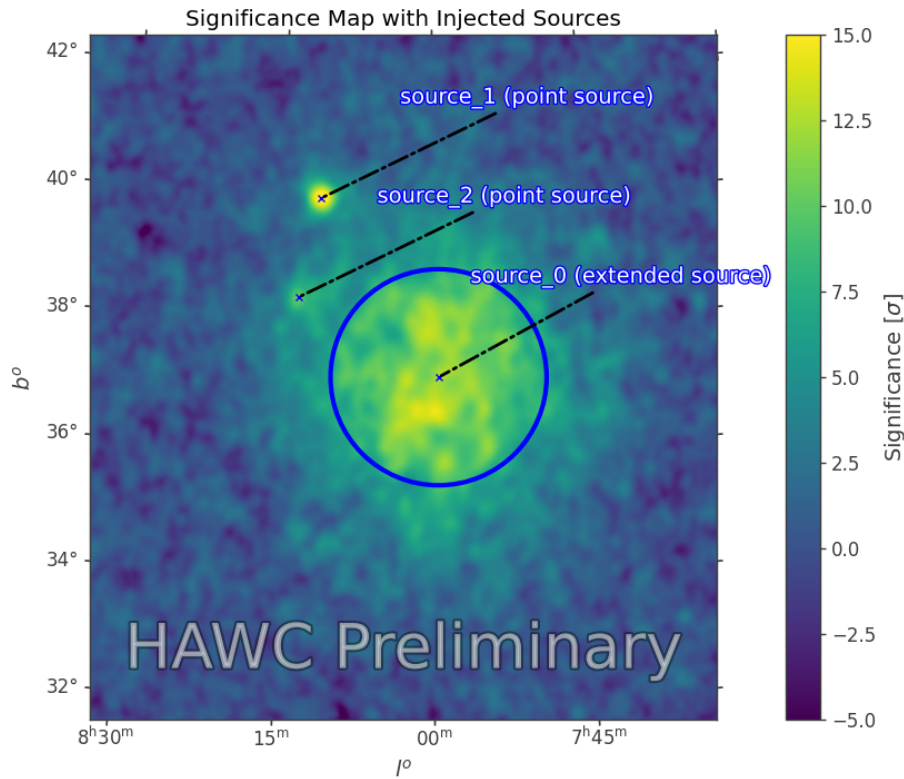


Figure 1: HAWC Significance Sky Map showing the location and morphology of three different simulated sources, marked with their labels and extensions (in galactic coordinates).

2. Source Detection Pipeline

To analyze and seed sources in a region, the first step is to determine the region of interest (ROI) of the source in question. The ROI of the image is chosen to encompass a large region ($\approx 5^0 \times 5^0$) around the potential sources. Once the ROI is found and the image is selected, it is then passed on to the image processing pipeline. The image processing pipeline is described below:

2.1 Image Preprocessing

The significance image is pre-processed to remove significant under-fluctuations. Significance maps show severe under-fluctuations, in regions where there is not enough background statistics and near the edges of the field of view (FOV) of the detector, as negative significance. When the significance is $< \sim -5\sigma$, during the image contrast normalization and blob detection steps (described in the following sections), background fluctuations starts to show up as fake source seeds (noise seeds). To eliminate these noise seeds, the original significance map is clipped to a minimum value of -5σ , the resulting image is shown in Fig.(1). This is done to reduce the noise in the images, while still keeping a large dynamic range for the pixel intensities for further processing. The clipped image is then scanned for peaks with values $\geq 5\sigma$ intensity, which is our detection threshold. If there are peaks found in the map, then the pipeline proceeds to the next step.

2.2 Image Normalization

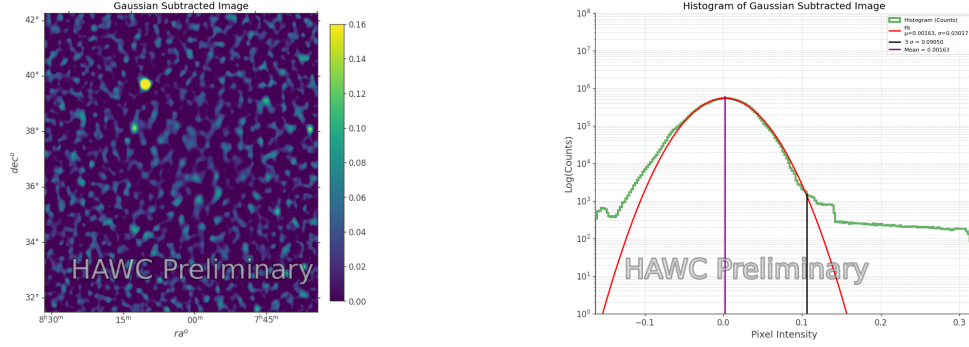
The clipped image shows a large dynamic range which is based on the ROI of the region in analysis, as regions with fainter sources look different from those with very bright sources. To produce consistency in the dynamic range between different regions of the data and to reduce the bias of finding brighter sources, the clipped image is first normalized to values between $[0, 1]$. This is an essential technique for spatial-based image enhancement, as it adjusts the image contrast to bring out the fainter sources in the images.

2.3 Feature Enhancement using Difference of Gaussian method

The normalized image, does not show the presence of any underlying features buried in it. To look at the underlying features in the normalized image, we use a feature enhancement algorithm called the Difference of Gaussians (DoG)[6][7].

The normalized image is convolved with a Gaussian kernel with a smearing radius, equal to the point spread function (PSF) value, averaged over different bins, of the detector corresponding to the declination of the region in analysis. Blurring the normalized image with the Gaussian kernels suppresses high-frequency spatial information such as sharp peaks, sharp edges, fine boundaries.. etc and enhances low-frequency spatial information such as gradual change in intensities. The Gaussian convolved image is then subtracted from the normalized image to produce a "DoG residual image". The resulting image preserves the high frequency components which when examined corresponds to the sources in the images. Point sources and smaller extended sources appear as sharper peaks, whereas large extended sources are removed. From Fig.(2a), it seen that there are 2 main peaks, which correspond to the point sources in the original map of the simulated sources seen in Fig.(1), and the large extended source from the original map is removed. Therefore, the DoG algorithm acts as a spatial bandpass filter, allowing high frequency components to pop up.

A smearing radius of 0.2° is used in the simulation plots shown in the proceeding. The intensity histogram of the DoG residual image is calculated as shown in Fig.(2b). The intensity histogram (green) resembles a zero-centered Gaussian distribution with a skewed tail toward positive values. Therefore, the intensity data is fit with a gaussian function (red), and we calculate the 3σ deviation of the gaussian intensity, which is used for intensity thresholding in the next step.



(a) Residual map after subtracting the 0.2° Gaussian convolved kernel image from the normalized image **(b)** Intensity Histogram of the DoG residual map. The pixel intensity curve are show in green, the gaussian fit to the curve is shown in red. The black line shows the 3σ deviation from the gaussian fit.

Figure 2: Difference of Gaussian (DoG) Algorithm

2.4 Blob Detection and Intensity Thresholding

We employ a blob detection algorithm, which is based on the DoG approach to look for blobs in the images [7]. The blob detection algorithm works by building an array of images, in which the DoG residual image is blurred with increasing smearing radii and the difference between two consecutive images are stored in a 3D array. The algorithm scans each of these images in the 3D array, where it looks for high intensity at the location of each pixel. For each pixel scanned in a single image, it performs a secondary scan of the 8 neighboring pixels. The the pixel is evaluated by comparing its intensity to the intensity of the surrounding pixels and determines if its a local maxima/minima, called as the local extrema. If the pixel is local extrema, then the pixel is compared to its corresponding pixel in all the images from the 3D array, and if the algorithm shows a local extrema in multiple blurred scale images, then it tags the pixel as a blob and the (x,y) pixel position and scale size is returned.

The DoG blob finder returns multiple detected blobs, both real and fake, depending on the local extrema and fluctuations near the boundary of the images. To remove the fake blobs, we apply a boundary cut and intensity thresholding steps.

A boundary cut is applied where blobs detected 1° from the edges are removed. Then, an internal intensity thresholding algorithm rejects the detected blobs, whose local extrema are less than the 3σ deviation from the gaussian fit to the pixel intensity histogram of the DoG residual map. A second thresholding algorithm is applied, where the significance values of the blobs from the input significance map are compared. If the significance value inside the blobs is less than 5 sigma, it is removed from the list of detected blobs.

After filtering the blobs with the intensity thresholds, a source seed list is produced. The seeds in the sources are classified into 2: (a) If the scale of the blob is less than 0.15° , then it is classified as a point-like source (b) If the scale of the blob is greater than 0.15° , then it is classified as an extended-like source. The seed list is passed then passed onto a astropy model file and the sources are evaluated using threeML+HAL Plugin.

3. Performance

We run the pipeline on 60 sets of simulations with a total of 123 sources: (i) 66 point sources (ii) 25: $< 0.5^\circ$ extended sources (iii) 33: $> 0.8^\circ$ extended diffuse sources. We recover 65 point sources, 24 of $< 0.5^\circ$ extended sources and 1 of $> 0.8^\circ$ extended diffuse sources. We see that the algorithm has a 10% false positive rate, most of it coming from detecting very bright $> 0.8^\circ$ extended sources as point-like. The cpu wall time to run the image processing pipeline, over 60 simulations sets was less than 26 hrs, whereas the conventional blind search using likelihood fitting algorithms took over 2 months. The image processing algorithm achieves a $60\times$ speedup compared to the conventional method, corresponding to an approximate performance gain of 98.33%.

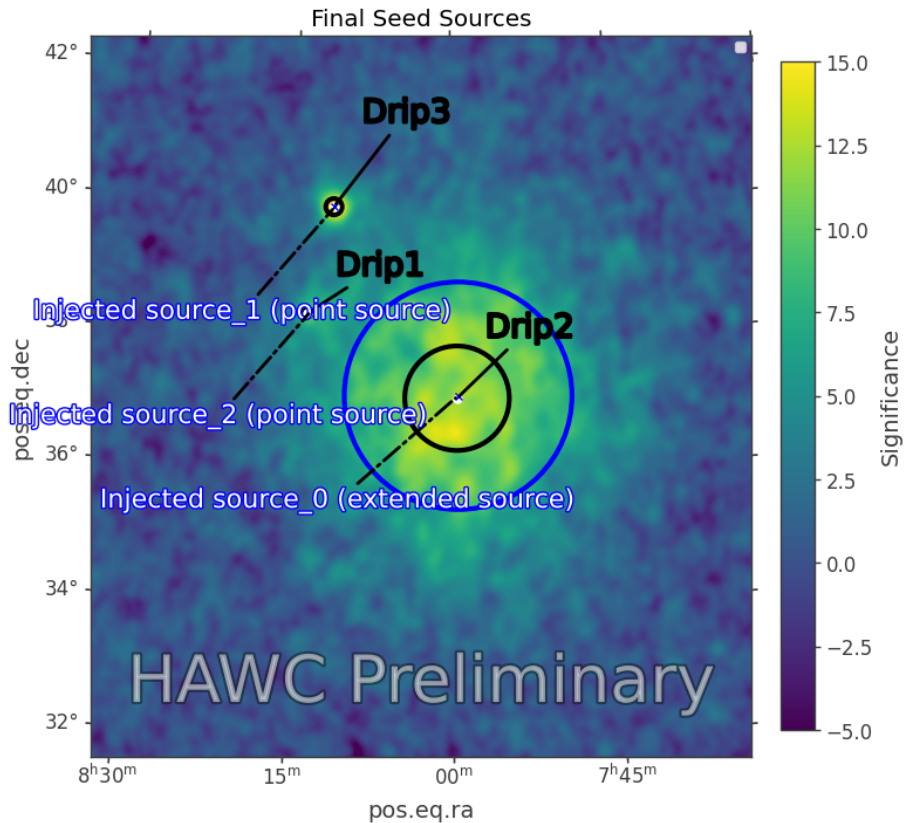


Figure 3: The source seeds detected from the image processing pipeline, along with the simulated source locations.

The output seeds, referred to as Drips, generated by the image processing pipeline on the simulation dataset used in this study are shown in Fig.(3). We see that for the simulation dataset used, the image processing pipeline is fully capable of recovering the simulated sources.

4. Conclusion

We explore the use of modern image processing algorithms for astronomical source detection, demonstrating significantly faster source seeding compared to traditional blind likelihood-based methods. The algorithm presented here achieves >98% accuracy in identifying point sources and >95% accuracy for extended sources, while offering a computational speedup of approximately 98% relative to conventional approaches. The proposed pipeline can be seamlessly extended to source-seeding applications in datasets from instruments like Fermi-LAT, as well as upcoming publicly available data from SWGO and LHAASO, enabling the full utilization of image-based analysis methodologies.

References

- [1] G. Vianello, R.J. Lauer, P. Younk, L. Tibaldo, J.M. Burgess, H. Ayala et al., *The Multi-Mission Maximum Likelihood framework (3ML)*, *arXiv e-prints* (2015) [arXiv:1507.08343](https://arxiv.org/abs/1507.08343) [1507.08343].
- [2] HAWC collaboration, *Characterizing gamma-ray sources with HAL (HAWC Accelerated likelihood) and 3ML*, *PoS ICRC2021* (2021) 828.
- [3] M. Ackermann, M. Ajello, L. Baldini, J. Ballet, G. Barbiellini, D. Bastieri et al., *Search for Extended Sources in the Galactic Plane Using Six Years of Fermi-Large Area Telescope Pass 8 Data above 10 GeV*, **843** (2017) 139 [1702.00476].
- [4] A. Albert, R. Alfaro, C. Alvarez, J.C. Arteaga-Velázquez, D. Avila Rojas, R. Babu et al., *TeV Analysis of a Source-rich Region with the HAWC Observatory: Is HESS J1809-193 a Potential Hadronic PeVatron?*, **972** (2024) 21 [2407.08849].
- [5] S. van der Walt, J.L. Schönberger, J. Nunez-Iglesias, F. Boulogne, J.D. Warner, N. Yager et al., *scikit-image: image processing in Python*, *PeerJ* **2** (2014) e453.
- [6] D. Marr and E. Hildreth, *Theory of edge detection*, *Proceedings of the Royal Society of London. Series B. Biological Sciences* **207** (1980) 187.
- [7] D.G. Lowe, *Distinctive image features from scale-invariant keypoints*, *International journal of computer vision* **60** (2004) 91.

134 **Full Author List: HAWC Collaboration**

135 R. Alfaro¹, C. Alvarez², A. Andrés³, E. Anita-Rangel³, M. Araya⁴, J.C. Arteaga-Velázquez⁵, D. Avila Rojas³, H.A. Ayala Solares⁶,
 136 R. Babu⁷, P. Bangale⁸, E. Belmont-Moreno¹, A. Bernal³, K.S. Caballero-Mora², T. Capistrán⁹, A. Carramiñana¹⁰, F. Carreón³, S.
 137 Casanova¹¹, S. Coutiño de León¹², E. De la Fuente¹³, D. Depaoli¹⁴, P. Desiati¹², N. Di Lalla¹⁵, R. Diaz Hernandez¹⁰, B.L. Dings¹⁶,
 138 M.A. DuVernois¹², J.C. Díaz-Vélez¹², K. Engel¹⁷, T. Ergin⁷, C. Espinoza¹, K. Fang¹², N. Fraija³, S. Fraija³, J.A. García-González¹⁸,
 139 F. Garfías³, N. Ghosh¹⁹, A. Gonzalez Muñoz¹, M.M. González³, J.A. Goodman¹⁷, S. Groetsch¹⁹, J. Gyeong²⁰, J.P. Harding¹⁶, S.
 140 Hernández-Cadena²¹, I. Herzog⁷, D. Huang¹⁷, P. Hütemeyer¹⁹, A. Iriarte³, S. Kaufmann²², D. Kieda²³, K. Leavitt¹⁹, H. León
 141 Vargas¹, J.T. Linnemann⁷, A.L. Longinotti³, G. Luis-Raya²², K. Malone¹⁶, O. Martinez²⁴, J. Martínez-Castro²⁵, H. Martínez-Huerta³⁰,
 142 J.A. Matthews²⁶, P. Miranda-Romagnoli²⁷, P.E. Mirón-Enriquez³, J.A. Montes³, J.A. Morales-Soto⁵, M. Mostafá⁸, M. Najafi¹⁹, L.
 143 Nellen²⁸, M.U. Nisa⁷, N. Omodei¹⁵, E. Ponce²⁴, Y. Pérez Araujo¹, E.G. Pérez-Pérez²², Q. Remy¹⁴, C.D. Rho²⁰, D. Rosa-González¹⁰,
 144 M. Roth¹⁶, H. Salazar²⁴, D. Salazar-Gallegos⁷, A. Sandoval¹, M. Schneider¹, G. Schwefer¹⁴, J. Serna-Franco¹, A.J. Smith¹⁷, Y. Son²⁹,
 145 R.W. Springer²³, O. Tibolla²², K. Tollefson⁷, I. Torres¹⁰, R. Torres-Escobedo²¹, R. Turner¹⁹, E. Varela²⁴, L. Villaseñor²⁴, X. Wang¹⁹,
 146 Z. Wang¹⁷, I.J. Watson²⁹, H. Wu¹², S. Yu⁶, S. Yun-Cárcamo¹⁷, H. Zhou²¹,

147 ¹Instituto de Física, Universidad Nacional Autónoma de México, Ciudad de Mexico, Mexico, ²Universidad Autónoma de Chiapas,
 148 Tuxtla Gutiérrez, Chiapas, México, ³Instituto de Astronomía, Universidad Nacional Autónoma de México, Ciudad de Mexico, Mexico,
 149 ⁴Universidad de Costa Rica, San José 2060, Costa Rica, ⁵Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Mexico,
 150 ⁶Department of Physics, Pennsylvania State University, University Park, PA, USA, ⁷Department of Physics and Astronomy, Michigan
 151 State University, East Lansing, MI, USA, ⁸Temple University, Department of Physics, 1925 N. 12th Street, Philadelphia, PA 19122,
 152 USA, ⁹Universita degli Studi di Torino, I-10125 Torino, Italy, ¹⁰Instituto Nacional de Astrofísica, Óptica y Electrónica, Puebla, Mexico,
 153 ¹¹Institute of Nuclear Physics Polish Academy of Sciences, PL-31342 11, Krakow, Poland, ¹²Dept. of Physics and Wisconsin IceCube
 154 Particle Astrophysics Center, University of Wisconsin—Madison, Madison, WI, USA, ¹³Departamento de Física, Centro Universitario
 155 de Ciencias Exactas Ingenierías, Universidad de Guadalajara, Guadalajara, Mexico, ¹⁴Max-Planck Institute for Nuclear Physics,
 156 69117 Heidelberg, Germany, ¹⁵Department of Physics, Stanford University: Stanford, CA 94305–4060, USA, ¹⁶Los Alamos National
 157 Laboratory, Los Alamos, NM, USA, ¹⁷Department of Physics, University of Maryland, College Park, MD, USA, ¹⁸Tecnológico de
 158 Monterrey, Escuela de Ingeniería y Ciencias, Ave. Eugenio Garza Sada 2501, Monterrey, N.L., Mexico, 64849, ¹⁹Department of
 159 Physics, Michigan Technological University, Houghton, MI, USA, ²⁰Department of Physics, Sungkyunkwan University, Suwon 16419,
 160 South Korea, ²¹Tsung-Dao Lee Institute & School of Physics and Astronomy, Shanghai Jiao Tong University, 800 Dongchuan Rd,
 161 Shanghai, SH 200240, China, ²²Universidad Politécnica de Pachuca, Pachuca, Hgo, Mexico, ²³Department of Physics and Astronomy,
 162 University of Utah, Salt Lake City, UT, USA, ²⁴Facultad de Ciencias Físico Matemáticas, Benemérita Universidad Autónoma de Puebla,
 163 Puebla, Mexico, ²⁵Centro de Investigación en Computación, Instituto Politécnico Nacional, México City, México, ²⁶Dept of Physics and
 164 Astronomy, University of New Mexico, Albuquerque, NM, USA, ²⁷Universidad Autónoma del Estado de Hidalgo, Pachuca, Mexico,
 165 ²⁸Instituto de Ciencias Nucleares, Universidad Nacional Autónoma de Mexico, Ciudad de Mexico, Mexico, ²⁹University of Seoul,
 166 Seoul, Rep. of Korea, ³⁰Departamento de Física y Matemáticas, Universidad de Monterrey, Av. Morones Prieto 4500, 66238, San Pedro
 167 Garza García NL, Mexico