Preprocessing of Astronomical Images from the NEOWISE Survey for Near-Earth Asteroid Detection

Rachel Meyer and José Manjarrés, Ph.D.

School of STEM, Olivet Nazarene University
Bourbonnais, IL

rrmeyer@olivet.edu

jemanjarres@olvet.edu

I. Introduction

Asteroid detection has become increasingly important to focus on and improve, as near-Earth asteroids (NEAs) have proven dangerous even at small sizes and velocities. The images collected from telescopes observing NEAs and potentially hazardous asteroids (PHAs) require a significant amount of human interaction and processing time [1], [2]. Most larger asteroids have been detected, but some are still yet to be discovered, specifically dimmer and slower-moving asteroids. Survey telescopes take pictures of the same part of the night sky multiple times per night to observe the transients moving over time [1], [3], [4], [5].

A. NEOWISE Survey

The Near-Earth Object Widefield Infrared Survey Explorer (NEOWISE) was created to find NEAs, but much of the sky is pictured in two infrared wavelength bands (3.4 and 4.6 μm) and includes data on multiple types of sources. The two different bands will capture different features based on the wavelengths given off from each source. So, allowing for pictures to be taken in two different wavelengths allows for more features to be observed on the sources. Each image taken through NEOWISE has a specific scan id as well as a frame number, allowing them to have unique names. The map in Figure 1 shows the sky areas where the NEOWISE survey has collected images during the past 7 years of operation. Some of the data collected include the location in astronomical units of right ascension and declination, absolute magnitude, visual magnitude, and the difference in the infrared bands. Absolute magnitude is the brightness of a celestial object as it would be seen from a set distance, whereas visual magnitude is the brightness of the object as viewed from Earth, where each object is placed at a different distance.

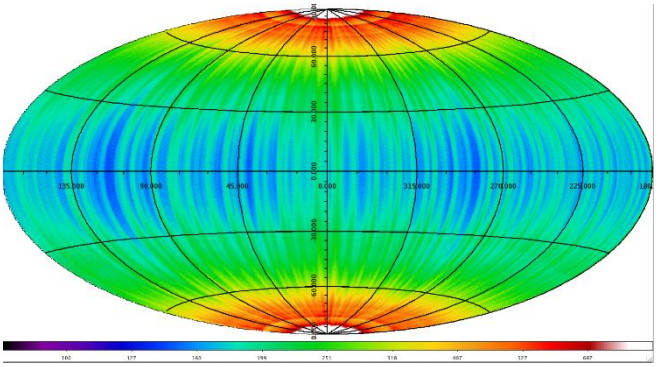


Figure 1. NEOWISE Sky Coverage. This depicts the sky with astronomical coordinates of right ascension and declination. Right ascension is on the x-axis in degrees from 0 to 360°, declination is on the y-axis in degrees from -90 to 90°. The projection map shows the Single-exposure depth-of-coverage accumulated by the NEOWISE survey.

B. Image Processing

The standard asteroid detection technique is to recognize streaks. This method performs poorly on detecting slower-moving asteroids that may only appear as a couple of pixels [3]. Recent studies have been investigating the implementation of Machine Learning to deal with deciphering the images collected from these telescopes [6], [7] and finding these more irregular asteroids. The raw data is often processed in diverse ways to fit better with the research's objective and to be efficiently used with Machine Learning. For example, Duev et al. [2] decided to convert the images into grayscale because the color was not necessary to detect bright pixels; it simplified the data and the processing. Also, in multiple works such as [8], [9], [10], and [3], the known sources are subtracted from images such as stars, planets, or satellites. The differenced images are then used for the research, focusing solely on the transients or unknowns. Another example of preparing the data is seen in [3], where Rabeendran and Denneau use random horizontal and vertical image flipping rotations to avoid bias in the results. To efficiently use the machine learning models, the images need to go through a pre-processing

step, including different changes such as background removal, thresholding, or known object removal.

C. Research Question

The problem addressed in this research is what pre-processing steps need to be taken to better detect dimmer and slower-moving asteroids in the NEOWISE image survey? To explore this issue, we implement different techniques with the collected images as well as the source data.

II. Methods

A. Data Collection

We collect data for this research from the IRSA database, specifically from the NEOWISE survey. To retrieve the images and other source data, we use MySQL programming to perform the database inquiries. The images are single-frame images from a range of –50 to 50 degrees in declination and the full range of right ascension. We collect data on 550,264 sources which includes asteroids, comets, planets, and planetary satellites. Each source data include the right ascension, declination, absolute magnitude, visual magnitude, the magnitude in bands 1 and 2, the difference between the band values, the identification for the image, and the x and y coordinate of the source on that image, which is calculated from the right ascension and declination. To convert the astronomical coordinates into pixels we first convert the right ascension and declination values that are originally in degrees, into arcseconds (1 degree is equal to 3600 arcseconds). We then convert to pixels because the image pixels collected from NEOWISE are reported to have 2.75 arcseconds per pixel. To save the images and data as workable files we save the images as NumPy arrays and the data as separate CSV files. Both files are named as the unique source id and frame number of that image. We apply a logarithmic filter to the image to improve image visualization, to scale the values down to readable values, leaving the background and the sources on the image.

B. Data Analysis

To better understand the data collected we analyze the sources separately by their specified labels. Specifically, for this dataset the labels are asteroid (A), comet (C), planet (P), and planetary satellite (S). In order to analyze the data, we calculate the minimum, median, and maximum for each different measurement collected for the source, which can be seen in Table 1.

a) Asteroid Statistics

	RA	Dec.	Absolute Mag.	Visual Mag.	Band 1/Band 2 Difference
Minimum	0.002	-49.99	-1.200	7.456	-7.593
Median	183.9	-2.080	15.13	20.13	0.911
Maximum	360.0	49.99	32.10	37.20	10.63

b) Comet Statistics

	RA	Dec.	Absolute Mag.	Visual Mag.	Band 1/Band 2 Difference
Minimum	0.471	-49.83	2.000	9.579	-3.143
Median	192.6	-2.048	11.83	21.10	0.857
Maximum	359.6	49.00	21.00	34.07	7.208

c) Planet Statistics

	RA	Dec.	Absolute Mag.	Visual Mag.	Band 1/Band 2 Difference
Minimum	0.626	-12.41	-9.400	-1.613	0.240
Median	168.9	-1.590	-6.045	5.720	1.009
Maximum	330.8	19.70	-1.52	8.151	2.010

d) Planetary Satellite Statistics

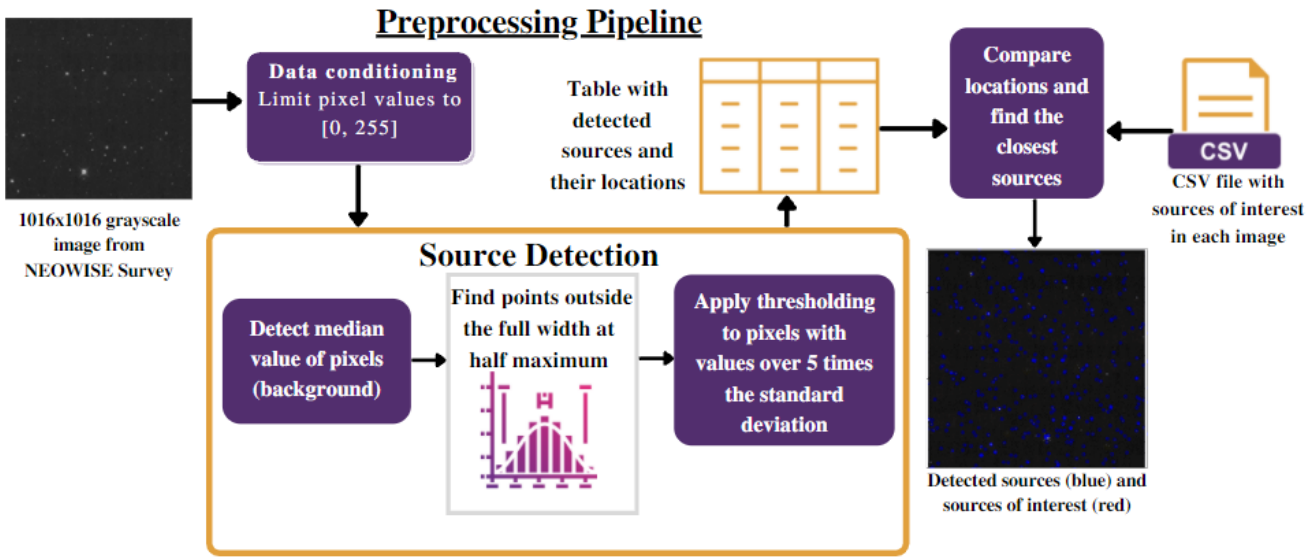
	RA	Dec.	Absolute Mag.	Visual Mag.	Band 1/Band 2 Difference
Minimum	0.101	-12.44	-2.000	5.637	-0.712
Median	91.19	-0.473	6.498	15.83	0.788
Maximum	359.6	19.63	16.90	26.20	5.409

Table 1. Source Data Statistics. This table shows the different data associated with each source type. The minimum, median, and maximum are reported for each feature.

C. Preprocessing Pipeline

Our preprocessing pipeline starts with the input of a 1016x1016 grayscale image collected from the database, and the final goal is the detect the point sources directly on the image and compare them with the known sources collected from the IRSA database, whose data is stored in the CSV file.

The first step of our pipeline is data conditioning. The data contained negative numbers representing broken pixels, so we limited the arrays to contain only positive numbers to allow for image visualization. The second part of the conditioning is limiting the data to values of 0 to 255, the standard values used in image processing. We were able to accomplish this by converting all of the images to 8-bit integers (whole numbers ranging from 0 to 255).



Next, we feed the images into a source detection algorithm called DAOStarfinder [14]. This algorithm was originally developed to detect stars but it can also be used to detect other point-like sources. The sources are detected by calculating the median value of pixels which is assigned as the value of the background. Points that lie outside of the full width half maximum (FWHM) of this value are flagged as possible sources. The last step is to add a threshold. We apply of threshold to pixels with values over 5 times the standard deviation. The output of this algorithm is a table with the x and y coordinates of each detected source on the image.

Lastly, to compare the actual locations of the sources found from the data collected initially and the sources detected from the previous algorithm, we use the coordinates stored in each individual CSV file and compare that with the closes point possible from the source table generated by calculating the smallest Euclidian distance (the length of the line segment between the two points). The final output is an image with all the sources circled in blue, the sources of interest (asteroids, comets, planets, planetary satellites) circled in red as well as the value of the Euclidian distance of original point to the detected point.

The steps of the preprocessing pipeline are pictured in Figure 2.

III. Results

Sources of Interest

Source	Count
Asteroids	548,700
Comets	1,331
Planets	35
Planetary Satellites	198

84.6% of sources of interest detected

Distance between estimated source location and actual location		mean	std. dev
	in pixels	26.76	29.62
	in arcsecs	73.6	81.4

IV. Discussion

We were able to show that the preprocessing pipeline utilized in this research as useful at finding sources because the mean average of the distance between the real source coordinate and the detected source coordinate is around 73.6 arcseconds (around .02 degrees in the sky) or 26.76 pixels. Since we know that the pipeline is able to detect the asteroids with some accuracy, we will be able to use it in the future to train Machine Learning algorithms to detect other asteroids. There is still some work that needs to be done to make sure that more of the sources of interest are detected. One possibility we could explore is to make sure that there are no anomalies in the data set and that there are no mistakes when collecting the coordinates of the sources from the database since there were some calculations that had to be made on the data. Another note is that the data has a large bias towards asteroids, as there is a considerably large amount of data collected for that source as compared to the others. We can also use the data statistics to determine certain parameters that

should be used in training for Machine Learning and what the threshold might be for these parameters.

V. Future Work

In the future, we plan to continue this research by implementing the preprocessing steps into a complete pipeline that involves Machine Learning to detect new asteroids in these images. The data collected in this work will be able to help train the Machine Learning algorithm, as well as be ready for processing, resulting in better accuracy when detecting asteroids, specifically dimmer and slower-moving asteroids. There will be a focus on the images, as well as the metadata from the database on each individual object such as the visible magnitude, absolute magnitude, and the magnitude of the object in the first wavelength band and the second wavelength band. Our work builds on existing efforts in implementing deep learning to detect transients and possible asteroid candidates in the NEOWISE asteroid surveys and improving the ability to recognize irregular moving asteroids. absolute magnitude, and the magnitude of the object in the first wavelength band and the second wavelength band. Our work builds on existing efforts in implementing deep learning to detect transients and possible asteroid candidates in the NEOWISE asteroid surveys and improving the ability to recognize irregular moving asteroids.

VI. Acknowledgements

This research was made possible by Olivet Nazarene University's Pence-Boyce Summer Research experience program and Catalyst ONU. The Pence-Boyce program was established by Elbert Pence and Fanny Boyce through a generous donation that allows undergraduate students to engage in independent research along with a mentor for 10 weeks each summer. This project is also an extension of a project started under the Olivet Honors Program. I would like to thank Dr. José Manjarrés for his expertise in Machine Learning and image processing as well as Dr. Steve Case for his astronomical knowledge and guidance. Funding was provided through the Hippenhammer Faculty grant as well as through the ONU Department of Engineering.

References

[1] Ruprecht, J., "Asteroid Detection Results Using the Space Surveillance Telescope", in *Advanced Maui Optical and Space Surveillance Technologies Conference*, 2015.

[2] D. A. Duev et al., "DeepStreaks: Identifying Fast-Moving Objects in the Zwicky Transient Facility Data with Deep Learning," Mon. Not. R. Astron. Soc., vol. 486, no. 3, pp. 4158–4165, 2019, doi: 10.1093/mnras/stz1096.

[3] A.C. Rabeendran and L. Denneau, "A Two-Stage Deep Learning Detection Classifier for the Atlas Asteroid Survey," Publ. Astron. Soc. Pacific, vol. 133, no. 1021, 2021, doi: 10.1088/1538-3873/abc900.

[4] C. J. Fluke and C. Jacobs, "Surveying the Reach and Maturity of Machine Learning and Artificial Intelligence in Astronomy," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 10, no. 2, pp. 1–40, 2020, doi: 10.1002/widm.1349.

[5] J. Kremer, K. Stensbo-Smidt, F. Gieseke, K. S. Pedersen, and C. Igel, "Big Universe, Big Data: Machine Learning and Image Analysis for Astronomy," IEEE Intell. Syst., vol. 32, no. 2, pp. 16–22, 2017, doi: 10.1109/MIS.2017.40.

[6] J. R. Masiero, A. K. Mainzer, J. M. Bauer, R. M. Cutri, T. Grav, E. Kramer, J. Pittichová, and E. L. Wright, "Asteroid Diameters and Albedos from NEOWISE Reactivation Mission Years Six and Seven," The Planetary Science Journal, vol. 2, no. 4, p. 162, 2021.

[7] A. Mahabal et al., "Machine Learning for the Zwicky Transient Facility," Publ. Astron. Soc. Pacific, vol. 131, no. 997, p. 38002, 2019, doi: 10.1088/1538-3873/aaf3fa.

[8] E. A. Smirnov and A. B. Markov, "Identification of Asteroids Trapped Inside Three-Body Mean Motion Resonances: A Machine-Learning Approach," Mon. Not. R. Astron. Soc., vol. 469, no. 2, pp. 2024–2031, 2017, doi: 10.1093/mnras/stx999.

[9] J. D. Ruprecht, H. E. M. Viggh, J. Varey, and M. E. Cornell, "SST Asteroid Search Performance 2014-2017," IEEE Aerosp. Conf. Proc., vol. 2018-March, pp. 1–8, 2018, doi: 10.1109/AERO.2018.8396388.

[10] R. L. Jones et al., "The Large Synoptic Survey Telescope as a Near-Earth Object Discovery Machine," Icarus, vol. 303, pp. 181–202, 2018, doi: 10.1016/j.icarus.2017.11.033.

[11] M. Lieu, L. Conversi, B. Altieri, and B. Carry, "Detecting Solar System Objects with Convolutional Neural Networks," Mon. Not. R. Astron. Soc., vol. 485, no. 4, pp. 5831–5842, 2019, doi: 10.1093/mnras/stz761.

[12] P. Vereš, R. Jedicke, L. Denneau, R. Wainscoat, M. J. Holman, and H.-W. Lin, "Improved Asteroid Astrometry and Photometry with Trail Fitting," Publ. Astron. Soc. Pacific, vol. 124, no. 921, pp. 1197–1207, 2012, doi: 10.1086/668616.

[13] A. Waszczak et al., "Small Eear-Earth Asteroids in the Palomar Transient Factory Survey: A Real-Time Streak-Detection System," Publ. Astron. Soc. Pacific, vol. 129, no. 973, 2017, doi: 10.1088/1538-3873/129/973/034402

[14] Stetson, P. B., "DAOPHOT: A Computer Program for Crowded-Field Stellar Photometry", *Publications of the Astronomical Society of the Pacific*, vol. 99, p. 191, 1987. doi:10.1086/131977.