

Discovery of 108 Faint Asteroid Candidates in Archival NEOWISE Infrared Space Telescope Images Using a Custom Deep Learning-Based Pipeline

Gauri Vani Todur¹

¹Santa Clara High School Santa Clara, USA

Publication Date: 2026/07/11

Abstract: Near Earth asteroids (NEAs) are defined as minor planets with orbits close to Earth that could pose potential collision risks. Over 98% of the estimated 3 million small near-Earth asteroids (19-140 meters) remain undiscovered due to their faintness, leaving Earth vulnerable to impacts like the undetected 18-meter-sized 2013 Chelyabinsk meteor that caused injuries to ~1600 people. The purpose of this research is to develop an automated deep learning-based pipeline to accurately discover small, faint, near-Earth asteroids. Because of its ability to pick up faint thermal signals, archival image data in the W2 band from the Near-Earth Object Wide-Field Infrared Survey Explorer (NEOWISE) was used. First, I filtered out stationary objects using a star masking process, and bright artifact pixel patterns by referencing WISE archival bitmask frames. I trained two Convolution Neural Network (CNN) deep learning models on approximately 1 million custom synthetic asteroid images generated from 2D Gaussian functions and blank sky images for classification (0.96 accuracy) and coordinate prediction (1.24-pixel accuracy on real asteroids). I deployed the dual-CNN pipeline on 36 square degrees of sky (49-55 deg. Right Ascension, 19-25 deg. Decl.) over 8.41 days of data, and used a linking algorithm to detect asteroid-like linear movement. I discovered 108 faint asteroid candidates that are 1.98 times fainter (median SNR ~3.3) and have 2.39 times higher apparent angular velocities (median ~0.86 deg/day) than most known asteroids in the search region, indicating the presence of a potentially new population of small and near-Earth asteroids, and recovered 91% of the searched region's known asteroids. To my knowledge, this research is the first to apply deep learning techniques to make new asteroid candidate discoveries from NEOWISE archival imagery. This AI pipeline advances planetary defense efforts, with applications in identifying transients like variable stars.

Keywords: Near-Earth Asteroids, Convolutional Neural Network, NEOWISE, Deep Learning, Infrared Astronomy, Moving Object Detection, Planetary Defense, Synthetic Image Generation, Asteroid Tracklet Linking.

How to Cite: Gauri Vani Todur (2026) Discovery of 108 Faint Asteroid Candidates in Archival NEOWISE Infrared Space Telescope Images Using a Custom Deep Learning-Based Pipeline. *International Journal of Innovative Science and Research Technology*, 11(6), 3212-3225. <https://doi.org/10.38124/ijisrt/26jun1572>

I. INTRODUCTION

Near Earth Asteroids (NEAs) are small rock bodies orbiting the Sun with diameters not large enough to be classified as planets, whose perihelia come within 1.3 Astronomical Units of the Sun (Center for Near Earth Object Studies, NASA). As of November 2025, 39,989 near-Earth objects have been discovered, according to the Minor Planet Center (MPC), the official international organization responsible for collecting and confirming minor planet observations and discoveries (International Astronomical Union).

NEAs pose collision risks to Earth due to their proximity. Catastrophic impacts from large asteroids (like the 10-15 km object that likely caused the Cretaceous-Tertiary

extinction 65 million years ago) are rare. Over 93% of kilometer-sized "civilization-ender" NEOs as of 2025 have been found and tracked, with no identified Earth impact risks (B612 Foundation, 2025).

However, catalog completeness drops drastically for smaller asteroids: As of October 2025, only 50% of the estimated 25,000 near-Earth asteroids between 140-1,000 meters, ~1% of the estimated 500,000 NEAs between 44 and 140 meters, and 0.1% of the NEAs between 19 to 44 meters have been discovered, according to the B612 Foundation, a leading planetary/NEA defense foundation (B612 Foundation, 2025).

Small NEOs remain dangerous despite causing only regional rather than global damage. The 2013 Chelyabinsk meteor (18 meters) injured 1600 people, and the 1908

Tunguska event (30 meters) exploded with the energy equivalent to 3 orders of magnitude greater than the Hiroshima atomic bomb (Kartashova et al. 2018; Brazo & Austin 1984). These impacts were unexpected until shortly before occurrence due to the asteroids' small sizes, faintness, and high velocities during close approaches.

Discovering asteroids enables orbit fitting to assess collision probability and allows time for planetary defense measures, including evacuation or deflection missions like NASA's DART (NASA 2024). Finding and cataloging as many asteroids as possible, particularly small, faint ones, is crucial for planetary defense and provides valuable insights into solar system formation.

➤ *Existing Methods of Asteroid Discovery*

Asteroid discovery initiatives often rely on ground-based surveys including the Zwicky Transient Facility (ZTF, 2018-Present), ATLAS (2015-Present), Catalina Sky Survey (1990s-Present), and Pan-STARRS (2014-Present) (Rigault, 2018; Tonry et al. 2018; Larson et al. 1998; Chambers et al. 2012). However, ground-based telescopes struggle to detect dim asteroids due to sky glow and atmospheric interference. Space-based telescopes like NASA's Wide-Field Infrared Survey Explorer (WISE) solve this issue, but WISE's Moving Object Pipeline Subsystem (WMOPS) imposes a minimum signal-to-noise threshold ($\text{SNR} > 4.5$), causing many faint asteroids to fall below its automated detection limits (Masiero et al. 2018).

Artificial intelligence models have been developed to improve asteroid detection efficiency. In 2018, ZTF launched ZStreak, which uses a Random Forest classifier to analyze asteroid streak features like flux, length, centroid, angle of motion (Ye et al. 2019). However, ZStreak's false positive rate proved problematic: of the 104 streaks it confirms daily from 105-106 detections, only several are confirmed real by human scanners after hours of daily review.

In 2019, Duev et al. created DeepStreaks, a three-CNN pipeline that decreased false positives by 50-100 times. However, out of 33,000 yearly candidates, only 15 were confirmed real. Evidently, there are challenges in distinguishing real asteroids from artifacts. To solve this, more accurate deep learning classifiers are needed.

➤ *WISE Telescope*

NASA's Wide-Field Infrared Survey Explorer (WISE) is a space-based infrared telescope launched in 2009. Its 40 cm telescope captures photometric data in four infrared bands (W1-W4) at 3.4, 4.6, 12, and 22 microns. After cryogen depletion in 2010 disabled the W3 and W4 bands, the telescope was reactivated in 2013 as "NEOWISE" to find asteroids using only W1 and W2 bands (Mainzer et al. 2014). Each WISE frame is 47 x 47 arcminutes, captured every 11 seconds with 10% overlap.

WISE detects infrared thermal radiation emitted by asteroids, which is otherwise undetected or extremely faint in visible wavelengths. Being space-based, WISE avoids atmospheric interference, sky glow, and weather

interruptions that limit ground-based surveys' ability to detect small, faint, low-albedo asteroids.

WISE's Moving Object Processing System (WMOPS) detects sources with $\text{SNR} > 4.5$ and links detections moving in straight lines at constant angular velocity into "tracklets." WMOPS requires five detections to form a tracklet before submission to the Minor Planet Center. This requirement causes rapidly-fading small asteroids to be missed as they fade below detection thresholds before accumulating sufficient observations.

I utilized W2 band data (4.6 microns) for its superior sensitivity to faint thermal radiation. Mainzer et al. (2011) demonstrated that asteroids are more frequently and reliably detected in W2 than W1, as W2 captures thermal signatures from objects too dim in reflected light, making it critical for identifying small, low-albedo asteroids.

➤ *Research Objectives*

The purpose of this research is to develop an accurate, automated process to discover fainter, smaller, and possibly near-Earth asteroids from WISE archival image data. The assumption that smaller asteroids appear faint is supported by the findings of the NEOWISE mission by Mainzer et al. 2014b.

The goals of the research are to develop a software pipeline (in Python) that can: recover >90% of the known asteroids in any given search region in WISE archival image data in the W2 band, require minimal manual inspection necessitating a false positive rate of <5%, and discover at least one new faint near-Earth asteroid candidate.

To achieve these goals, the following overarching steps are taken:

- I use WISE image data in the W2 band, detecting infrared radiation at 4.6 microns. I look for singular moving point sources in the imagery, indicating the presence of a NEA.
- To ignore stars and artifacts, I utilize a procedure to mask out fixed point sources (like stars) by referencing a coaddition ("coadd") frame to ensure remaining sources are only moving objects.
- I automate the process of asteroid discovery by using a "Classification" Convolutional Neural Network image (CNN) classification model. This model is trained on images of faint, synthetically generated asteroids and non-asteroids images of blank sky regions. These training set images have all undergone the star-masking procedure described above. I develop and train a second "Localization CNN" on the same synthetic asteroid images to predict the location of an asteroid within an image.
- To accurately detect small, faint asteroids, I have developed a method that involves generating faint synthetic images of asteroid point sources. This approach allows for the creation of a substantial dataset for training the Classification CNN. Synthetic images are particularly useful because real, faint, small near-Earth asteroids are rare in observational data. By training only on these

synthetic images, the CNN can learn to identify subtle features of faint asteroids more effectively, as a larger training sample size improves the CNN's classification abilities.

- I deploy the Classification and Localization CNNs on a small region of the sky, saving the positive detections' images, coordinates, and the times. I use a linkage software to create tracklets of at least four asteroid detections. Candidate asteroid discoveries are cross-checked with the NEOWISE Known Solar System Object Possible Association List and the Minor Planet Center reportings.
- Discovered asteroid detections go through a vetting procedure, including ensuring their movement paths align with the majority of known asteroids reported in the search region, and have 3+ detections passing a false alarm noise score test.

II. DATA RETRIEVAL

To develop the dual-CNN software pipeline, real WISE W2-band image frames were required as background images of artificially implanted synthetic asteroid images and as part of the Non-Asteroid training set for specifically the Classification-CNN. The following sections detail the procedure I used to select, load, process, and filter all images to be used as suitable training samples for the CNNs, and to load data for the asteroid blind search.

➤ Stationary Source Masking Procedure

I downloaded image frames as FITS files from the NEOWISE archival database hosted by IRSA (NASA/IPAC Infrared Science Archive). The raw FITS files are populated with stars. Since asteroids and stars both appear as circular point sources with varying brightness, I use a standard masking procedure to remove fixed sources, leaving only moving objects.

I create a coadd (average of multiple exposures of the same sky region) from WISE coaddition frames with depths of 12-13 frames across the ecliptic plane. Because the coadd averages all light detected over time, it only contains (magnifies) fixed sources, not moving asteroids.

I estimated background and RMS noise levels for both the single-exposure frame and coadd using a rolling window approach with sigma-clipping and nearest-neighbor interpolation. I smoothed the background and RMS maps using Gaussian blur (kernel size 31, sigma = 20). Using the coadd, I created a binary object map by applying a fixed pixel threshold, then dilated and eroded it to refine stationary object boundaries. I masked the image frame by setting all stationary source pixels to NaNs and filled them with noise derived from the background RMS. I also set pixels below three times the negative RMS to NaN and infilled them. Finally, I normalized the entire image by dividing each pixel by five times the local RMS value to standardize noise levels across the frame and ensure consistent visibility of faint objects.

Because each frame is 47 arcminutes wide while typical asteroids are ~11 arcseconds wide, I cut every frame into 15 x 15 pixel images, creating 4489 smaller images per frame for training and asteroid discovery.

➤ Artifact Removal Process

After processing the single images and masking known stars, it was realized through the course of this research that another process is required to remove artifacts, or sources that may resemble asteroids but are actually spurious detections. To identify artifacts I used contaminated pixel bit-masks provided in the WISE data releases. The bit-masks contain 32-bit signed integer values where pixel values represent the sum of powers of 2, where each power represents a technical issue identified at that pixel location. Through manual analysis of hundreds of image frames and bit-mask comparisons, I found that a bit-mask pixel value of 268435456, identifying the pixel as the victim of an extremely high or low value outlier, caused artifact patterns or hot pixels that the image processing methods describe previously did not fix. In light of this, I wrote a method to remove such patterns, as follows:

- For each 15x15 pixel image, its corresponding bit-mask image is found by cropping it out of the larger bit-mask frame provided by WISE.
- In the bit-mask frame, pixels with bit-flags of 28 in either squares or diagonal or vertical/horizontal adjacent pixels are searched for. This value is identified by 268435456 to indicate the corresponding FITS image pixel. (Bitflags of the value 2^{28} indicate the pixel was an outlier (with an extremely high or low value).
- Create a binary map using the location of the identified bit-mask pattern and dilate it, then apply it to the original image to fill with NaNs in affected area.
- Fill NaNs with noise in original image.

I filter out squares, diagonal pixels, and horizontal/vertically stacked flagged bit-mask pixels. An example of an image being corrected using my artifact removal process can be seen in Figure 1.

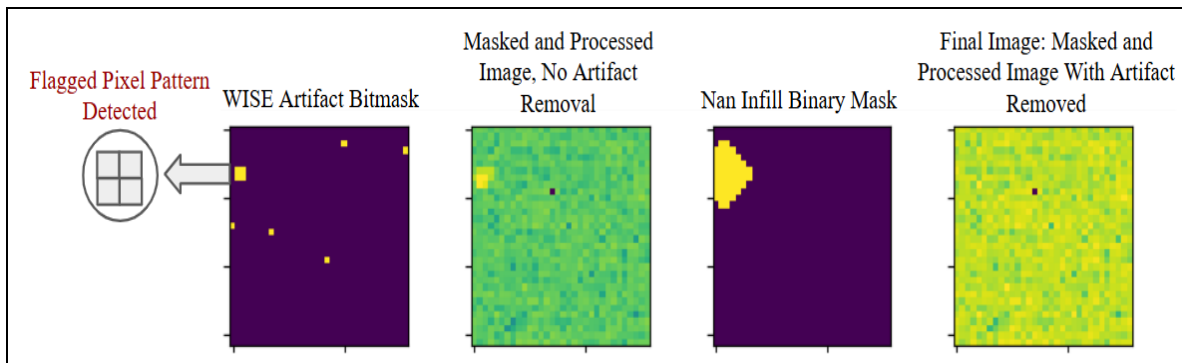


Fig 1 Example of an Image with an Artifact that is a Square Pattern of Bright Pixels. My Automated Removal Process Clears the Surrounding Region of the Image, Resulting in a Clear Image without the Bright Artifact Pattern. Figure Created by the Student Researcher in Matplotlib and Google Slides, 2025

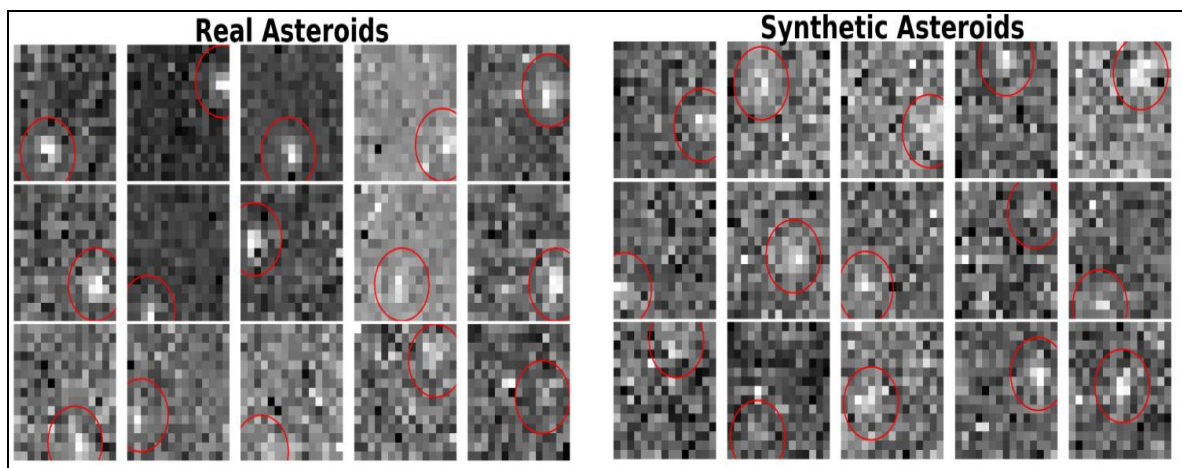


Fig 2 Real vs My Synthetic Asteroid Image Samples. Left: Real Asteroids W2 Band Images. Right: My Synthetically Generated Asteroid Images, Implanted Over a Background Sky Image. These Synthetic Images are Samples of what was used in Training the CNNs. Figures Made by Student Researcher in Matplotlib, and Placed Side-by-Side in Google Slides, 2025.

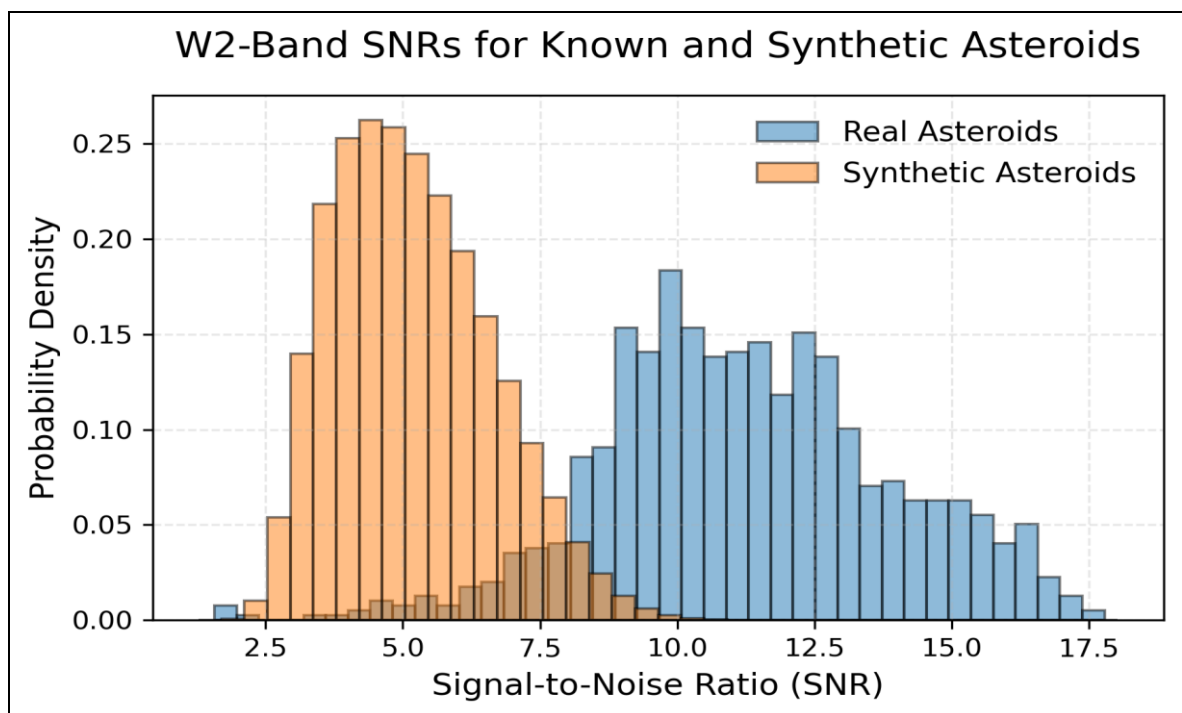


Fig 3 Comparison of SNR Distributions for my Synthetically Generated Asteroid Images Used to Train the CNN Models, and Known Asteroid Images Reported by WISE and the Minor Planet Center. Figure Made by Student in Matplotlib, 2025.

III. SYNTHETIC ASTEROID IMAGE GENERATION

The main goal for this research is to train a machine learning model on very dim images of asteroids, with the aim of discovering fainter asteroids than those already cataloged. Since slower-moving asteroids in WISE data, with shorter exposure times than ground-based telescope data, appear as circular point sources rather than streaks, it is possible to generate a synthetic image of an asteroid using a 2-dimensional Gaussian function.

The equation I used to generate the 2D Gaussian shape in an image is shown in Equation 1.

$$f(x, y) = A \cdot e^{-\left(\frac{(x-x_0)^2}{2\sigma_x^2} + \frac{(y-y_0)^2}{2\sigma_y^2}\right)} \quad \text{Equation 1}$$

This function determines the value of a pixel given its x and y coordinate. A is the amplitude and controls how bright the source will be. X₀ and Y₀ are the coordinates of the center of the Gaussian. σ_x and σ_y are the standard deviations along the x and y axes, respectively.

To generate synthetic asteroid images, I needed a set of parameters to use as the amplitude and standard deviation values. I derived these parameters from 100 real images of asteroids from WISE W2 band images. I specifically used asteroids with W2 magnitudes greater than 14. I loaded, processed, and masked these asteroid's corresponding images and fit a 2D Gaussian function to each of them. By fitting a Gaussian function to each image, I was able to extract the parameters needed to regenerate each asteroid image.

After fitting these asteroid images, I used the extracted parameters to create uniform distributions I could then sample to generate ~ 500,000 images of synthetic asteroids. I ensured every synthetic asteroid image had a signal-to-noise ratio (SNR) over 3, which I deemed the bare minimum for visibility. The center x and y parameters were randomized for every image, since asteroids will not always be perfectly in the center of the image inputted to the machine learning model. The parameter distributions were sampled 447,748 times to create sets of 2D Gaussian parameters to generate faint synthetic asteroid images. I then implanted these images into random cutouts of the sky, all in the W2 band. Figure 2 showcases a side-by-side comparison of real asteroids and my synthetically created asteroid images implanted onto background sky images. The synthetic images are a sample of the images used to train the Classification CNN. Evidently, the real asteroids appear brighter, and the synthetic asteroids fainter but still realistic. This validation step is essential; since the Classification CNN is trained exclusively on synthetic data, the implanted sources must closely mimic real asteroids for the model to generalize to true detections, including previously missed low-SNR objects. Figure 3 presents the SNR distribution of the faint, known asteroid samples, and my generated synthetic asteroid implant images. My synthetic asteroid

images have a peak SNR around 5, while the known asteroids have a peak SNR around 10. This confirms that the synthetic images represent fainter asteroids than those previously catalogued.

IV. DUAL CONVOLUTIONAL NEURAL NETWORK PIPELINE

➤ Introduction

In this research, I aim to detect asteroids in WISE image frame cutouts. To automate this process, I employ the use of deep learning models that I built and trained.

Deep learning is a subfield within machine learning that specifically uses neural networks to comprehend hierarchical features in any type of data, given similar training samples. Each neuron in a neural network can receive multiple input values, calculates a weighted sum of those values, applies a non-linear activation function to the sum, and sends the output to the next neural network layer. The greater the number of neural network layers, the more high-level the network's understanding of the input data becomes.

Convolutional Neural Networks (CNNs) were built to apply neural networks to make decisions about 2-dimensional data: images. Flattening an image into a one-dimensional array for input into a neural network prevents any understanding of the spatial relationships between pixels. For this reason, CNNs were designed to maintain the two-dimensional structure in images and learn spatial features. Specifically, a CNN is made of multiple hidden layers that convolve the image with a kernel to produce a smaller feature map. The feature maps are reduced through pooling layers, flattened, and fed into fully connected neural network layers until an output, or prediction, is produced for the image.

To automate the asteroid discovery process, I built and trained a "Classification CNN" to predict the likelihood of an image containing an asteroid, and a "Localization CNN" to predict its location.

In both of my CNNs, I include the following features:

- **Convolutional Layer:** Extracts local features (like edges and shape) by sliding learnable kernels across the image, computing weighted sums of pixel values.
- **Batch Normalization:** Using the batch's mean and variance, it stabilizes and speeds up training by normalizing activations, then scaling/shifting them with learnable parameters.
- **Max Pooling:** Serves to reduce spatial resolution and noise sensitivity by selecting the maximum values from feature map regions.
- **Global Average Pooling 2D:** Compresses each feature map to one number by averaging spatial elements, directly mapping features to classifications with fewer parameters.

- Dropout Layer: Reduces possibility of the model overfitting by randomly zeroing a fraction of neuron outputs during training.
- Dense Layer: Combines extracted features through fully connected weights in order to compute class scores using weighted sums and an activation function.

I make use of the Adam optimizer with learning rate of $1e-4$, and Binary Cross-entropy loss function. The Adam optimizer is used to update weights with an adaptive learning rate for model convergence. Binary cross-entropy loss measures classification error by penalizing an incorrect confidence score for a two-class output ("Asteroid" vs. "No Asteroid"). The following sections detail the exact architecture of the classification CNN (labels images as having an "Asteroid" vs. "No Asteroid") and the localization CNN (outputs coordinate of predicted asteroid in image).

➤ Classification CNN

The Classification CNN's architecture is shown in Table 1.

Table 1 Classification CNN

Layer (Type)	Output Shape	Param #
Input Layer	(None, 15, 15)	0
Reshape	(None, 15, 15, 1)	0
Conv2D (32 filters)	(None, 15, 15, 32)	320
BatchNormalization	(None, 15, 15, 32)	128
MaxPooling2D	(None, 7, 7, 32)	0
Conv2D (64 filters)	(None, 7, 7, 64)	18,496
Batch Normalization	(None, 7, 7, 64)	256
MaxPooling 2D	(None, 3, 3, 64)	0
Conv2D (128 filters)	(None, 3, 3, 128)	73,856
Batch Normalization	(None, 3, 3, 128)	512
GlobalAveragePooling2D	(None, 128)	0
Dropout	(None, 128)	0
Dense (1 unit)	(None, 1)	129
Total Parameters		93,697
Trainable Parameters		93,249
Non-trainable Parameters		448

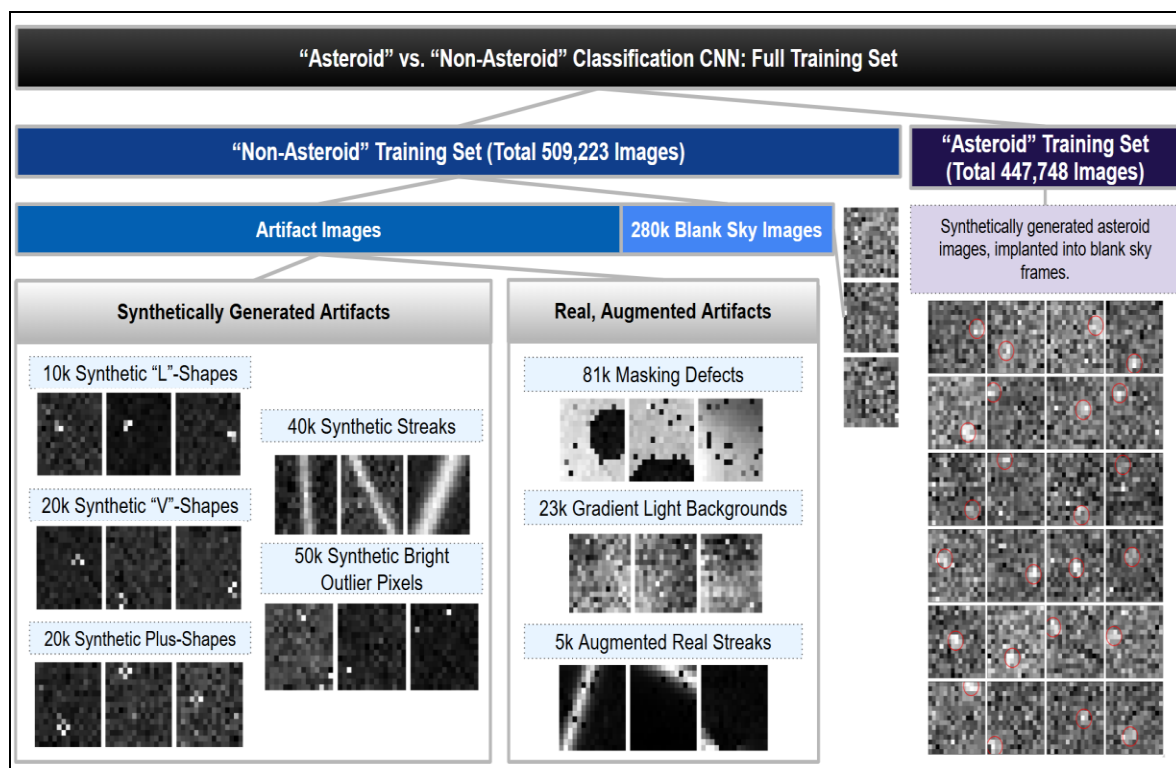


Fig 4 Training Dataset Breakdown for the Classification CNN. The Purpose of this Figure is to Highlight the Various Types of Artifacts Both Generated and Collected/Augmented, as well as the Blank Sky Images Used in the Training Set for the "Non-Asteroid" Class. Figure Created by Student Using Google Slides and Matplotlib for Image Samples, 2025.

I trained the Classification CNN on 509,023 Non-Asteroid images, and 447,748 synthetic asteroid images generated using the process outlined in Section 4, creating a total dataset of 956,771 training images. Roughly half of the Non-Asteroid images are images that contain no source (solely background noise) and were selected from random regions of the sky. The other half of the Non-Asteroid images contain obvious artifacts, either synthetically

implanted or taken from real images and augmented. An overview of the training set is seen in Figure 4.

For the Non-Asteroid Training Set, I began by selecting random background sky images that I had processed, following the procedures outlined in previous sections. Then, I used 280,000 of those images as purely blank sky images to train the model to understand the lack of presence of an asteroid. For the rest of the Non-Asteroid training set images, I found that various types of artifacts

were repeatedly flagged as asteroids by the model, that were not filtered out by my artifact removal process. These artifacts included L-shapes, V-shapes, four pixels in a plus-shape, satellite and diffraction spike streaks, singular bright outlier pixels, masking defects that left dark patches where a star had been removed, and ombre/gradiented light backgrounds. All of these artifact types have features such as gradually increasing brightness, roughly circular shapes, and visible edges that could make them resemble asteroids to the CNN. Since the focus of this research was to build a pipeline with a low false-positive rate, it was imperative that the CNN never flag such artifacts as asteroids. Thus, I decided to generate and curate sets of each type of artifact after viewing tens of thousands of such samples.

The artifact types that I collected included gradient light backgrounds, masking defects, and satellite streaks. The number of images I collected for each of these categories varied, but because CNNs perform better with larger training sample sizes, I decided to augment the images I collected by applying two major transformations to create copies of the collected images: 1) Rotate the image by 90, 180, and 270, and 2) scale the image in brightness by a set of factors ranging from 0.5 to 2. With these augmentation techniques, I was able to curate datasets of 81k masking defects, 23k gradient light backgrounds, and 5k satellite streak images.

Further artifact categories included "L"-shapes, "V"-shapes, and bright outlier pixels that also caused initially high false positive rates. These images, for their slightly circular shape resembling that of a small NEO, proved to be the greatest source of false positives in the course of this research. To address this, I generated a significant number of synthetic samples of these artifacts to add to the "Non-Asteroid" training set.

- "L"-shapes: This pattern is made of three pixels varying in brightness that form a "corner" pattern. I generated samples of these images, and allowed for the random selection of an orientation of the L-shape, a random scale value to be chosen to vary the brightness of the pixels, and a random coordinate in the background image where I implanted it.
- "V"-shapes and "Plus"-shapes: The V-shape is the plus-shape but missing one pixel. To generate these I randomized the orientation, pixel brightnesses, and location where I implanted the shape in the background image.
- Satellite streaks and star diffraction spikes: Streaks left by satellite trails and star diffraction spikes varied in brightness, width, and rotational orientation. To generate artificial streaks, the following steps were taken: First, a random angle θ on the interval $[0, \pi]$ is uniformly sampled and projected across the whole image using the unit direction vector $(\cos\theta, \sin\theta)$. A brightness value is randomly selected from the range $[0.2, 200]$, and the brightness is linearly mapped to a corresponding streak width from ranges I customized after manual analysis of hundreds of real streaks. The streak is then created by applying a Gaussian intensity profile perpendicular to

the streak line, resulting in a smooth appearance. Finally, the streak is implanted into a background image and this process was repeated until a dataset of 50,000 such streaks was created.

- Bright outlier pixels: I injected synthetic bright pixels into 50,000 background images. For each image, 1 to 5 such bright pixels were added, with the number samples from a weighted distribution that favors fewer insertions. The pixel intensities were drawn from a log-normal distribution with parameters $\mu = \log(0.6) + \sigma^2$, $\sigma = 1.0$, constrained to be >0.7 and capped at 600, and implanted at random coordinates.

Following conventional practices in AI research, I split the entire training set into Training, Validation, and Test sets. This ensures that some training samples are set aside for epoch-by-epoch evaluation (Validation set) in order to tune the model's hyperparameters, and some are set aside for post-training evaluation (Test set). This split ensures that the model does not overfit to the training data and provides a way to evaluate the model's classification ability on unseen data. The Training, Validation, and Test split I used was 90%, 5%, and 5%, respectively.

I split the entire training set, comprised of "Asteroid" and "Non-Asteroid" images into batches of 32 images each, which made for 27,191 total batches. I trained the CNN for two epochs, which was reasonable given the small image widths and the large training set size (nearly 1 million). The total training time was 6.06 minutes on an RTX 2080 GPU.

The final training set and validation set accuracies and corresponding losses are shown in Table 2. The validation accuracy being lower than the training accuracy is common, since the model has already seen the training data during learning, allowing it to fit more closely to those samples, whereas the validation data are completely unseen examples. The final performance metrics of the model on the separated test set data are: F1: 96.22%, Accuracy: 96.28% (fraction of correct predictions over all samples), Recall: 94.74% (fraction of real asteroids the model correctly predicts), ROC AUC: 99.36% (Area under the Receiver Operating Characteristic curve, or how well the CNN classifies detections across various confidence thresholds), and False Positive Rate: 2.18% (Fraction of non-asteroid images incorrectly classified as asteroids).

The CNN's performance on real asteroids will only roughly represent its success rate. For this reason a separate test set of 100 real, faint asteroids and 100 no-asteroid images was created. These asteroids were selected as the faintest set of asteroids in a random sky search region with W2 band magnitudes at least 14 (fainter side of the known WISE asteroids). Figure 5 showcases the ROC curve demonstrating a strong ability to classify asteroids vs. non asteroids on both the test split and real asteroid images, the True Positive Rate vs. False Positive Rate curves demonstrating a confidence score of 0.4 or 40% to be an ideal threshold for minimizing TPR-FPR on the real asteroid test set, and the confusion matrix results on the real asteroid test set which yielded a mere 3% FPR.

Table 2 Final Training and Validation Accuracy and Loss of the Classification CNN after Training.

Metric	Training	Validation
Accuracy	0.964	0.959
Loss	0.106	0.138

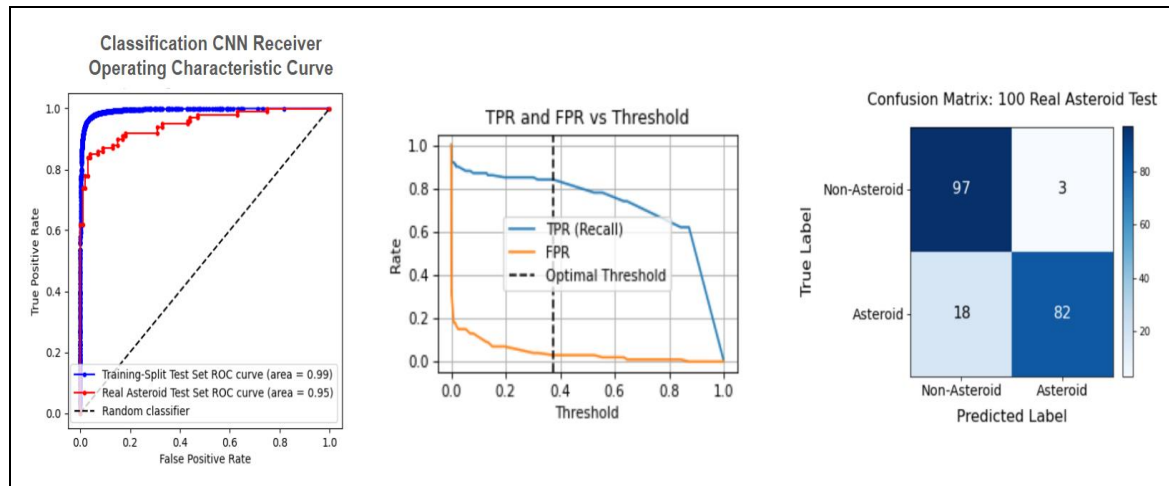


Fig 5 Left: ROC Curve. The blue line shows my model's ROC performance on the Test set initially separated from the total training data, across all confidence threshold values. The red curve represents the model's performance on a test set of 100 faint asteroids (W2 magnitude < 14), randomly selected from the MPC's Numbered Minor Planet list. Middle: True Positive Rate (TPR) vs False Positive Rate (FPR) results on the real asteroid test set, demonstrating a roughly 0.4 confidence threshold is optimal to minimize TPR-FPR. Right: Confusion matrix of Classification CNN's performance on real, ultra faint asteroids.

➤ Coordinate Localization CNN

After a decision is made regarding the potential presence of an asteroid in an image using the Classification CNN, I built and trained a secondary "Localization CNN" to predict the pixel location of the asteroid. (The development of a single CNN to both classify images and localize potential asteroids proved ineffective, since such a model architecture required coordinate labels even for images without asteroids, which led to skewed coordinate predictions). Although the images used for classification are already small (15 pixels-wide = 41.25 arcseconds-wide squares) the precise location of the asteroid given by my Localization CNN is necessary for more accurate and robust detection linking procedures later on.

I trained the Localization CNN on the ~447k synthetic asteroid images used to train the Classification CNN, along with the coordinates of the implanted asteroid in each image. Training was completed on a RTX 2080 GPU in 2.6 minutes. The Localization CNN's architecture is outlined as follows in Table 3. The layers are similar to those in the Classification CNN, however, the final dense layer outputs two values: the predicted horizontal and vertical coordinates of the asteroid within the image.

On the test set of 100 known, faint asteroid images in the WISE W2 band, the model achieved a Mean Absolute Error of merely 1.24 pixels. This demonstrated strong accuracy in predicting asteroid positions.

Table 3 Coordinate Localization CNN Model Architecture

Layer (Type)	Output Shape	Param #
Conv2D	(13, 13, 16)	160
Conv2D	(11, 11, 32)	4,640
MaxPooling2D	(5, 5, 32)	0
Conv2D	(3, 3, 64)	18,496
Conv2D	(1, 1, 128)	73,856
Flatten	(128)	0
Dense	(64)	8,256
Dropout	(64)	0
Dense	(2)	130
Total Parameters		105,538
Trainable Parameters		105,538
Non-trainable Parameters		0

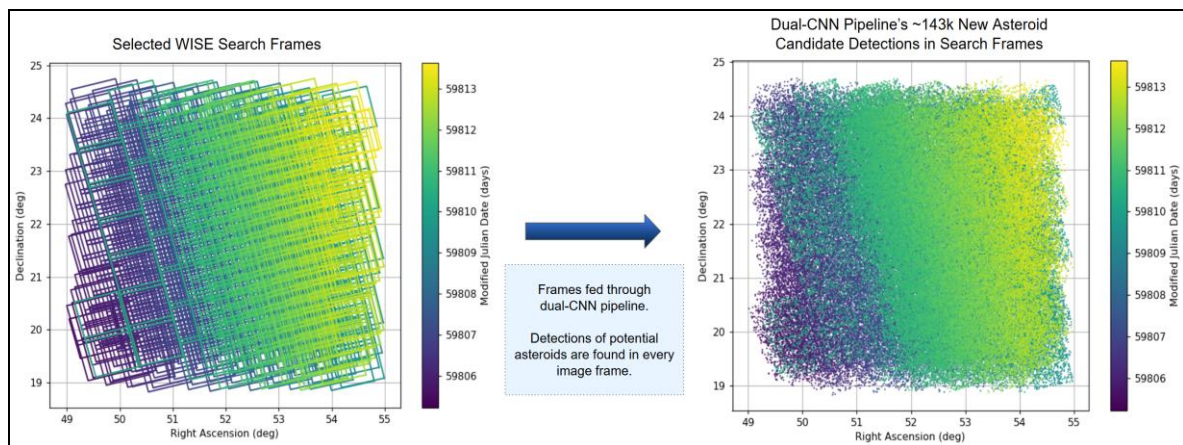


Fig 6 Left: Plot of the Selected WISE W2 Frames for the Search, Color-Mapped to their Respective MJD Values. Right: My CNN Pipeline's Detections of Candidate Asteroids (Images with >40% Confidence of an Asteroid Presence) in the Selected Search Frames. Figures Created by Student Researcher in Matplotlib and Labeled in Google Slides, 2025.

V. ASTEROID SEARCH

I deployed my dual-CNN pipeline on 755 WISE W2-band image frames covering roughly 36 square degrees of sky over 8.41 days of data. The images were processed and masked to remove stationary sources as outlined in the Data Retrieval Section.

From the ~143k candidate detections outputted by the Classification CNN (greater than 30% confidence of being an asteroid), I filtered out ~9k detections that matched with known asteroid detections from the NEOWISE Known Object Possible Association List. This catalog contains detections that may not be real asteroids or simply artifacts, but that overlap with confirmed minor planets' orbits.

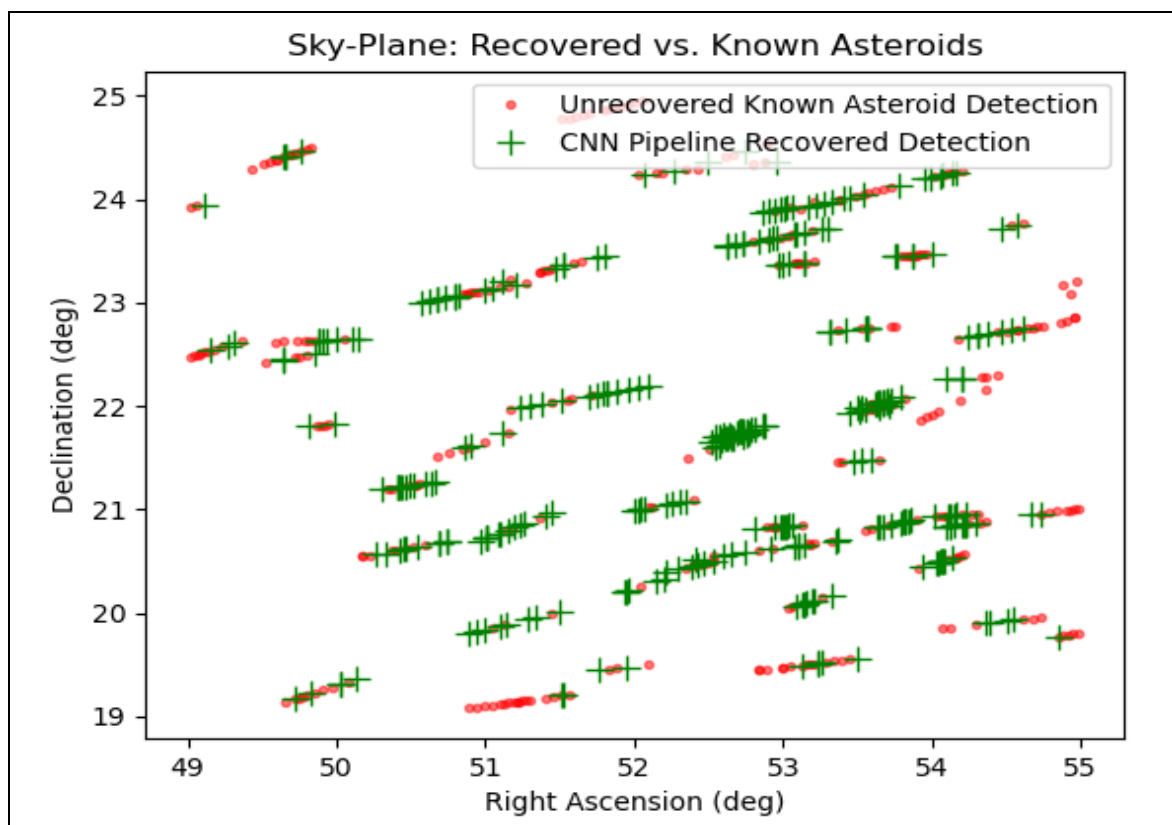


Fig 7 Green Plus Signs Represent My Software's Successful Recovery of a Known Asteroid Detection Reported by the Minor Planet Center and Red Dots Represent Unrecovered Detections. Figure Made by Student Researcher in Matplotlib, 2025.

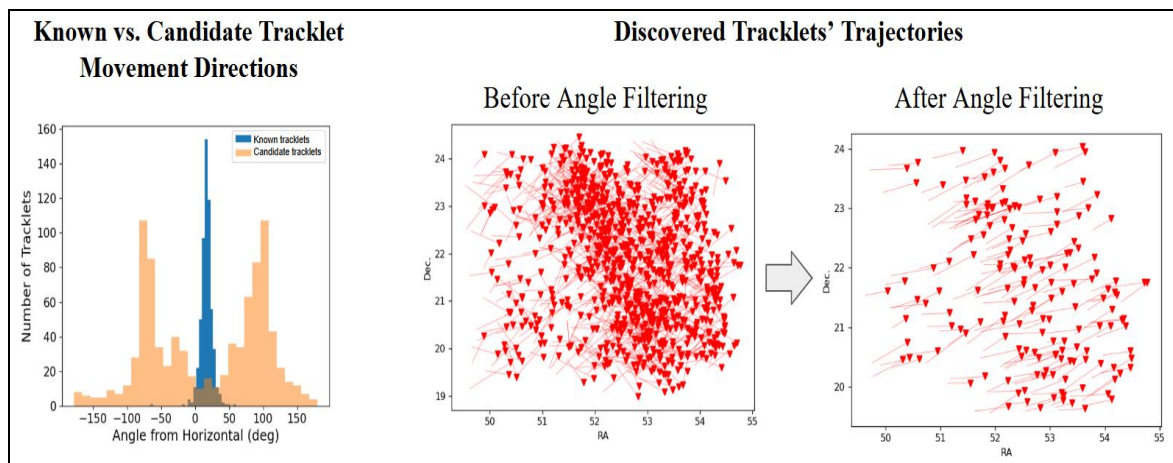


Fig 8 Left: Histogram of Known Asteroids' Movement Direction, Compared with my Initial Tracklets' Movement Directions. Right: Side-by-Side Comparison of all 800+ Tracklets that I Reduced to 171 by Keeping Tracklets with Movement Directions within the Known Asteroids' Range. Figures Created by Student Researcher Using Matplotlib and Google Slides, 2025.

➤ Recovered Asteroids

I searched the Minor Planet Center's list of numbered minor planet detections for all WISE reported observations of a minor planet within the selected region. A sky-plane plot is shown in Figure 7, comparing my software's successful recoveries of those detections, while highlighting the known detections that were not recovered by my software. My CNNs recovered at least one detection for 91% of the different known asteroids in this region.

There are multiple reasons why my software did not pick up every single known WISE detection of a minor planet in this search region. First, my software uses data from only the W2 band, so the reported object may have been only visible in the W1 band. Asteroids are known for having fluctuating magnitudes over time, even from frame to frame due to their rotating nature that changes its reflectivity over time. Second, the confidence threshold I set to determine whether a detection is classified as an asteroid was 40% confidence of there being an asteroid in the image. It is possible that the unrecovered detections were given confidence values slightly lower than this threshold. Third, the detections may have overlapped with an extremely bright star or galaxy, which would have been masked out by the image processing procedure. Regardless of these possibilities, the pipeline's recovery of most known asteroids proved the reliability of training solely on synthetic asteroid images, and the model's effectiveness for a blind search.

➤ Asteroid Discovery

Detecting a source that looks like an asteroid is not enough to confirm its existence. One must create "tracklets", or sequences of detections of the same object over time, at various locations in the sky, to confirm it is a moving object. I utilize a recursive linkage algorithm to "link" the 143k predicted asteroid positions' Right Ascension and Declination coordinates and their respective MJD times into tracklets of 4 or more detections. I imposed several rules on this linkage algorithm: (A) Detections must be moving in a straight line (the angle between three detections must be < 1 degree), (B) there should be a maximum of 0.3 days time difference from detection to detection, and (C) consecutive

pairs of detections within a tracklet should maintain constant apparent angular velocities with the other pairs in the tracklet (maximum difference in angular velocity of 0.01 deg/day).

In roughly 1.5 hours, the linkage algorithm outputted 59,723 tracklets of 3 detections, 886 tracklets of 4 detections, and 11 tracklets of 5 detections. I selected only tracklets of 4 and 5 detections long.

➤ Vetting Procedure

With linked asteroid detections demonstrating the possibility of a moving object, further vetting was necessary to confirm it is a valid asteroid discovery or strong candidate.

First, I queried the IRSA catalog of WISE Possible Association objects for asteroid detections from the search region selected, collecting their Right Ascension/Declination sky coordinates as well as Modified Julian Dates (MJDs). Grouping the detections by asteroid name, I could clearly view the apparent trajectories of the known asteroids from the search region, and realized they move in similar directions due to their location along the ecliptic plane, as shown in Figure 8. Accordingly, I selected only the candidate tracklets with movement angles between -1.12 degrees and 34.44 degrees from horizontal. In Figure 8, the candidate tracklets' movement angle histogram has two clear peaks around -75 and 100 degrees from horizontal. This is due to the maximum frame coverage in a vertical/diagonal strip in the middle of the selected search region, leading to denser predicted asteroid coordinates there. Such regions of clustered detections provided more tracklet possibilities for the linkage algorithm, and since this region was a largely vertical strip, the movement angles are almost 90 degrees inclined.

The resulting 171 candidate tracklets that passed this angle filtering contained linked detections that looked suspiciously like pure noise. So, I wrote a statistical significance assessment pipeline for every image by comparing observed image statistics against null hypothesis

distributions from blank sky regions. This program assigned a p-score to every image to quantify how noise-like the source looks. First, I collected ~2 million blank background sky images using my Classification CNN. Then, I counted the number of pixels exceeding 3-sigma of the median pixel value, along with the total summed flux across the image. I calculated empirical p-values for each discovery by computing the fraction of blank-sky images that equal or exceed the candidate images. Finally, I combined the two independent p-values using Fisher's statistical method, resulting in a single score. I filtered the 171 tracklets only for tracklets with 3+ detections having a 10% combined-p-score, indicating they are brighter than 90% of blank-sky images. This reduced total tracklets to around 110 tracklets.

From the final 110 tracklets, I analyzed the few tracklets that had repeated detections. This yielded three groups of two, four, and two tracklets each. I then selected

one tracklet for each group based on noise scores, visual similarities from one frame to another, tracklet length, and apparent velocity differences. After removing tracklets with repeated detections, I had my final list of 108 strong asteroid candidate discoveries with four detections in each tracklet.

➤ Discoveries Analysis

In Figure 9, the discovered asteroid candidates and known asteroid populations' apparent angular velocity and SNR distributions are shown. The median of my discoveries' SNRs is 3.3, which is 1.98-times lower than the median of the population of known asteroids from the searched region reported by WISE. In addition, the median of my discovered tracklets' apparent angular velocities is 0.86 deg/day, which is 2.39-times higher than the median of the motion rates I calculated from all the WISE-reported asteroids from the search region.

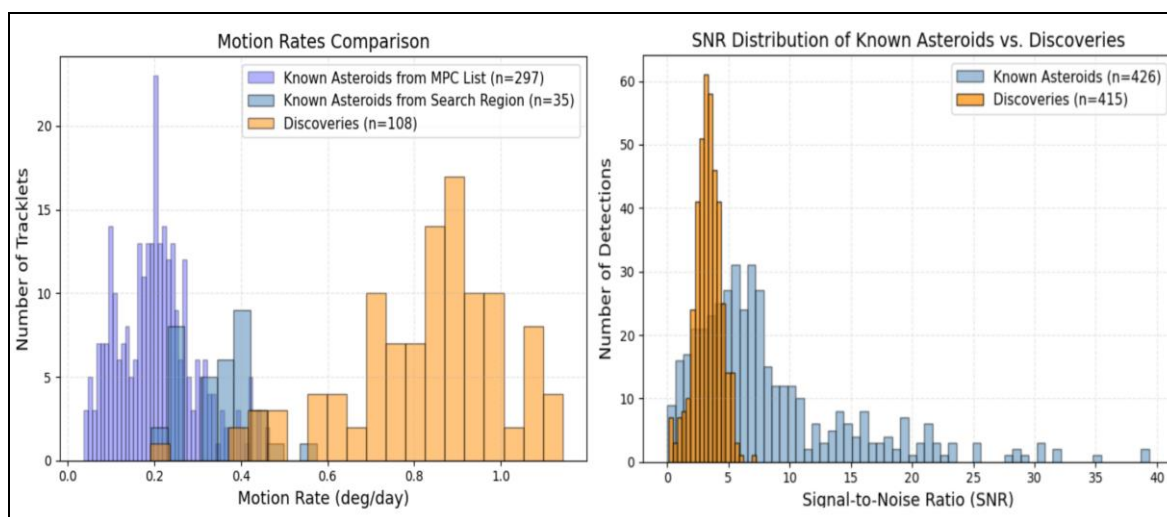


Fig 9 Known vs. Discovered Asteroids' Apparent Angular Velocities (Motion Rates) and Signal-to-Noise Ratios. Details: Known Asteroids' SNR's were Calculated by Loading and Processing Asteroid Detections Reported by the Minor Planet Center (MPC). Known asteroids' Motion Rates were Calculated from the First and Last Detection of Reported Asteroids Coordinates from the MPC. Figures Created by Student Researcher Using Matplotlib, 2025.

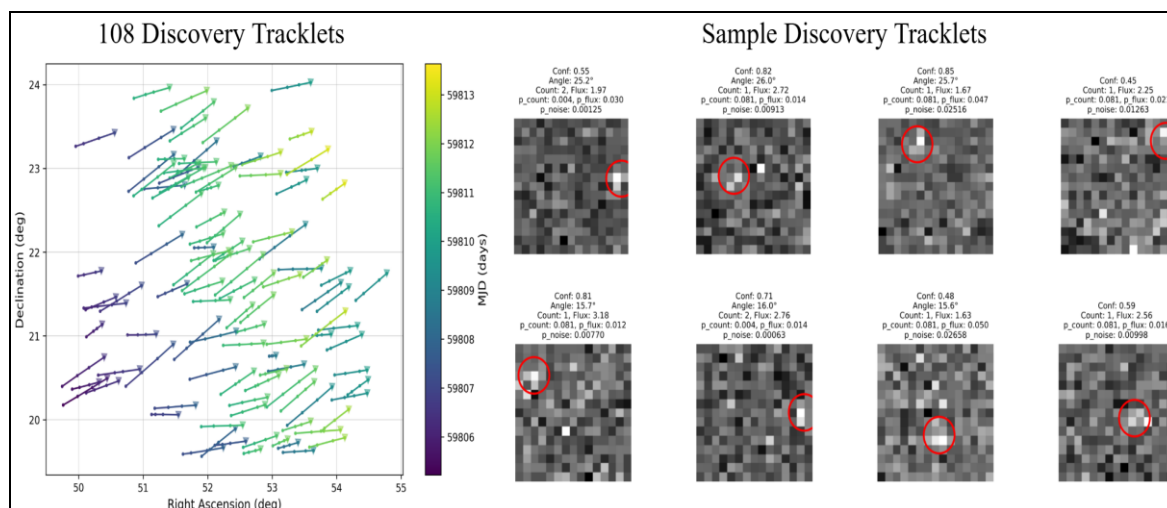


Fig 10 Left: Tracklet Sky-Plane Plot of 108 Discovered Moving Objects, Color-Coded by Time Measured in MJD. Triangles Indicate the Last Detection in a Triangle, Visually Depicting Movement Direction. Right: Two Sample Tracklets' Discovery Images, with Detection Metadata Above. Figures Created by Student Using Matplotlib and Placed Together in Google Slides, 2025.

Figure 10 showcases the sky-plane plot of the asteroid candidates' tracklets, color-coded by time measured in MJD. Triangles indicate the last detection in a tracklet, depicting their movement direction. On the right are two of my tracklets' discovery images, with detection metadata above, including the Classification CNN's confidence score, and angle between the consecutive detections, and p-scores used in the noise score calculations.

VI. DISCUSSION

The newly discovered asteroid candidates are fainter than those found by WISE, proving that training the Classification CNN on synthetic faint images effectively allowed for the detection of fainter moving objects. While the minimum SNR for asteroids discovered by WISE was 4.5, my AI-powered discovery pipeline has no minimum brightness level. Instead, I use a noise score system that combines multiple methods of brightness calculation (summed flux and bright pixel counts) to generate a p-value that allows for the comparison of the detected source with WISE blank-sky frames. This process ensures that detections are not filtered out for being too faint, while ensuring most detections in a tracklet are at least visible.

The faintness of the candidate discoveries indicates that either the objects are very small, or very distant. However, given the high apparent motions of the discoveries, they are likely not distant, as truly distant asteroids, like those in the main belt, often have apparent velocities of ~ 0.3 degrees per day. Thus, the discoveries are likely small and near-Earth. Brightness is also determined by the composition of the asteroid. For example, studies have shown that asteroids with high carbon content are often darker and less reflective than silicate and metal-rich asteroids. The discovered asteroid candidates could have a specific composition that makes them appear dimmer in WISE thermal imagery.

The discovered asteroid candidates have higher apparent velocities than those discovered by WISE likely due to two main reasons:

- The discoveries are potentially near-Earth, compared to the majority of Minor Planet Center-reported asteroids which are located in the main asteroid belt. The sensitivity of this pipeline to faint asteroids indicates that these asteroids may be much smaller than main-belt asteroids, which are often kilometers in diameter. Small asteroids are only visible to WISE if they are closer to Earth. Hence, the asteroids discovered in this research are likely near-Earth or at least closer than ~ 2 AU from Earth where the asteroid belt is located.
- The pipeline's detection requirements are optimized for rapidly-fading objects during brief visibility windows. Fast-moving near-Earth asteroids brighten temporarily during close approaches to Earth before rapidly fading as they recede. WMOPS requires five detections to form a tracklet, but by the time a fast-moving asteroid accumulates its fifth observation over WMOPS's multi-

day baseline, it may have faded below the 4.5 sigma SNR threshold. In contrast, this pipeline requires only three detections with a noise score less than 0.1, allowing tracklets to be completed while the asteroid remains visible. This combination (lower SNR sensitivity, fewer required detections) enables the discovery of small near-Earth asteroids that move rapidly (median 0.86 deg/day) but fade too quickly to accumulate the five detections WMOPS requires.

VII. CONCLUSION

I have developed an accurate dual-CNN detection and tracking pipeline that discovered 108 strong asteroid candidates that are 1.98-times fainter and have 2.39-times higher angular velocities than the vast majority of known asteroids from the searched region, from WISE archival W2-band infrared image data, with minimal manual inspection required. The faintness and high angular velocities of the candidates indicate the discovery of a potentially new population of small, near-Earth asteroids. This is also the first research effort to apply deep learning on WISE imagery and make candidate asteroid discoveries, and achieves a 3% false positive rate on real asteroid images. The candidate discoveries are posted in a PDF hosted on Google Drive at the following link: <https://drive.google.com/file/d/1CKw2RTBZtqETZLk-FYBYK0Gpg3KZ5DnJ/view>. For shorthand, this can be accessed at <https://tinyurl.com/108tracklets>.

The following criteria were used to process/vet the discoveries. Thresholds/statistics are drawn from my analyses of known asteroid data reported by the Minor Planet Center. (A) Detections must not be catalogued before, (B) there must be 4+ detections in a tracklet, (C) there can be a maximum 0.3 days time difference from detection to detection, (D) the differences between angular velocities from consecutive pairs of detections in a tracklet are less than 0.01 deg/day, (E) tracklet movements must be in a straight line, or at most 1 degree off between three consecutive points (aligned with WISE's WMOPS threshold), (F) the asteroids must move counterclockwise in the solar system around the Sun, as the vast majority of minor planets do, (G) the tracklet must be within the correct movement angle range of known asteroids in the same search region, (H) discoveries must have 3+ detections with a noise score < 0.1 , indicating they're brighter than 90% blank sky images (see statistical significance test above), and (I) there must be no repeated detections across different tracklets.

Extrapolating the discoveries across the ecliptic plane (18,000 square degrees), this model could find 54,000 new asteroids, over eight days worth of data. Optimizing my pipeline for large-scale deployment, scientists could find hundreds of thousands, if not millions, of dim asteroid candidates in the 14 years of Wide-field Infrared Survey Explorer archived data.

FUTURE WORK

The moving object candidates discovered using this pipeline are strong candidates based on the validation criteria outlined above. For further confirmation, follow-up observations using data from other space or ground-based surveys like ATLAS are needed, if it is possible to view the small asteroids from Earth. Submitting the tracklets to the Minor Planet Center would allow for targeted follow-ups and orbit determination.

Beyond asteroid discovery, this AI-based pipeline has applications in transient detection. The Classification CNN's ability to distinguish faint sources from background noise makes it suitable for finding variable stars, supernovae, and other time-domain phenomena in WISE. The pipeline could be adapted to detect stellar variability and explosive transients if I were to modify the linking algorithm to search for sources that appear and disappear rather than move linearly.

Future improvements to the pipeline include incorporating multiple WISE-bands to better characterize object properties, and implementing orbit fitting algorithms to assess collision probabilities for the discovered candidates.

The recursive nature of the linkage algorithm means that the time to link all the CNNs' detections grows exponentially, preventing me from being able to deploy this pipeline on all 14 years of WISE imagery without significant optimizations. Deploying an optimized version of this pipeline on all years of WISE archival data is an important step towards potentially discovering tens of thousands of new small and near-Earth asteroid candidates, advancing planetary defense efforts.

ACKNOWLEDGMENTS

This research was conducted over 2.5 years under the virtual mentorship of Dr. Jian Ge, Chair Professor at Shanghai Astronomical Observatory, Chinese Academy of Sciences, and Mr. Kevin Willis, Senior Research Mentor at the Science Talent Training Center. I am deeply grateful for their sustained guidance throughout this project, including their help in shaping the overall research direction, suggesting new ideas to explore, refining the scientific questions, and providing technical feedback on astronomy concepts, data analysis, and programming approaches. Their mentorship and encouragement were invaluable in helping me develop as a student researcher. I am also grateful to the Science Talent Training Center for providing computational resources that supported this work.

This research makes use of data products from the Wide-field Infrared Survey Explorer (WISE) and the Near-Earth Object Wide-Field Infrared Survey Explorer (NEOWISE), which are joint projects of the University of California, Los Angeles, and the Jet Propulsion Laboratory/California Institute of Technology, funded by the National Aeronautics and Space Administration. Archival

data were accessed through the NASA/IPAC Infrared Science Archive (IRSA), operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with NASA.

Asteroid catalog data used for pipeline validation and candidate vetting were sourced from the International Astronomical Union's Minor Planet Center (MPC). I also acknowledge the B612 Foundation's Asteroid Institute for publicly available near-Earth object statistics referenced in this work.

REFERENCES

- [1]. B612 Foundation: 2025 | Asteroid Institute | Annual Report - B612, <https://b612foundation.org/2025-asteroid-institute-annual-report/>
- [2]. Brazo, M., Austin, S.: THE TUNGUSKA EXPLOSION OF 1908, https://www.icr.org/research/index/researchp_sa_r05/
- [3]. Center for Near Earth Object Studies, NASA: Glossary - NEO (Near-Earth Object), <https://cneos.jpl.nasa.gov/glossary/NEO.html>
- [4]. Chambers, K.C., Magnier, E.A., Metcalfe, N., Flewelling, H.A., Huber, M.E., Waters, C.Z., Denneau, L., Draper, P.W., Farrow, D., Finkbeiner, D.P., Holmberg, C., Koppenhoefer, J., Price, P.A., Rest, A., Saglia, R.P., Schlafly, E.F., Smartt, S.J., Sweeney, W., Wainscoat, R.J., Burgett, W.S.: The Pan-STARRS1 Surveys. *The Astrophysical Journal*. 750, (2012). <https://doi.org/10.48550/arXiv.1612.05560>
- [5]. Chambers, K. C., "The Pan-STARRS1 Surveys", arXiv e-prints, Art. no. arXiv:1612.05560, 2016. doi:10.48550/arXiv.1612.05560.
- [6]. Duev, D.A., Ashish Mahabal, Ye, Q., Kushal Tirumala, Belicki, J., Dekany, R., Frederick, S., Graham, M.J., Laher, R.R., Masci, F.J., Prince, T.A., Riddle, R., Philippe Rosnet, Soumagnac, M.T.: DeepStreaks: identifying fast-moving objects in the Zwicky Transient Facility data with deep learning. *Monthly Notices of the Royal Astronomical Society*. 486, 4158–4165 (2019). <https://doi.org/10.1093/mnras/stz1096>
- [7]. International Astronomical Union: IAU Minor Planet Center, <https://www.minorplanetcenter.net/>
- [8]. Kartashova, A.P., Popova, O.P., Glazachev, D.O., Jenniskens, P., Emel'yanenko, V.V., Podobnaya, E.D., Skripnik, A.Y.: Study of injuries from the Chelyabinsk airburst event. *Planetary and Space Science*. 160, 107–114 (2018). <https://doi.org/10.1016/j.pss.2018.04.019>
- [9]. Larson, S., Brownlee, J., Herenrother, C., Spahr, T.: The Catalina Sky Survey for NEOs. *Bulletin of the American Astronomical Society*. 30, 1037 (1998)
- [10]. Mainzer, A., Bauer, J., Cutri, R.M., Grav, T., Masiero, J., Beck, R., Clarkson, P., Conrow, T., Dailey, J., Eisenhardt, P., Fajardo-Acosta, S., Fowler, J., Gelino, C., Grillmair, C., Heinrichsen, I., Kendall, M., Kirkpatrick, J.D., Liu, F., Masci, F.: Initial Performance of the NEOWISE

- Reactivation Mission. *The Astrophysical Journal*. 792, 30 (2014a). <https://doi.org/10.1088/0004-637X/792/1/30>
- [11]. Mainzer, A., Bauer, J., Grav, T., Masiero, J., Cutri, R.M., Wright, E., Nugent, C.R., Stevenson, R., Clyne, E., Cukrov, G., Masci, F.: THE POPULATION OF TINY NEAR-EARTH OBJECTS OBSERVED BY NEOWISE. *The Astrophysical Journal*. 784, 110 (2014b). <https://doi.org/10.1088/0004-637x/784/2/110>
- [12]. Mainzer, A., Grav, T., Bauer, J., Masiero, J., McMillan, R.S., Cutri, R.M., Walker, R., Wright, E., Eisenhardt, P., Tholen, D.J., Spahr, T., Jedicke, R., Denneau, L., DeBaun, E., Elsbury, D., Gautier, T., Gomillion, S., Hand, E., Mo, W., Watkins, J.: NEOWISE OBSERVATIONS OF NEAR-EARTH OBJECTS: PRELIMINARY RESULTS. *The Astrophysical Journal*. 743, 156 (2011). <https://doi.org/10.1088/0004-637x/743/2/156>
- [13]. Masiero, J.R., Redwing, E., Mainzer, A.K., Bauer, J.M., Cutri, R.M., T. Grav, Kramer, E., Nugent, C.R., Sonnett, S., Wright, E.L.: Small and Nearby NEOs Observed by NEOWISE During the First Three Years of Survey: Physical Properties. *The Astronomical Journal*. 156, 60–60 (2018). <https://doi.org/10.3847/1538-3881/aacce4>
- [14]. NASA: Double Asteroid Redirection Test (DART) - NASA, <https://science.nasa.gov/mission/dart/>
- [15]. NASA/IPAC Infrared Science Archive: IRSA NEOWISE Archive, <https://irsa.ipac.caltech.edu/ibe/data/wise/neowiser/>
- [16]. Ridgeway, B.: Marshall Center Astronomer Bill Cooke, Other NASA Researchers Among International Science Coalition Issuing Chelyabinsk Meteor Findings in New Papers - NASA, <https://www.nasa.gov/centers-and-facilities/marshall/marshall-center-astronomer-bill-cooke-other-nasa-researchers-among-international-science-coalition-issuing-chelyabinsk-meteor-findings-in-new-papers/>
- [17]. Rigault, M. (2018). The Zwicky Transient Facility: System Overview, Performance, and First Results. *Publications of the Astronomical Society of the Pacific*. <https://doi.org/10.1088/1538-3873/AAECBE>
- [18]. Tonry, John & Denneau, L. & Heinze, A. & Stalder, B. & Smith, K. & Smartt, S. & Stubbs, Christopher & Weiland, H. & Rest, A.. (2018). ATLAS: A High-cadence All-sky Survey System. *Publications of the Astronomical Society of the Pacific*. 130. [10.1088/1538-3873/aabadf](https://doi.org/10.1088/1538-3873/aabadf).
- [19]. Ye, Q., Masci, F.J., Lin, H.W., Bolin, B., Chang, C.-K., Duev, D.A., Helou, G., Ip, W.-H., Kaplan, D.L., Kramer, E., Mahabal, A., Ngeow, C.-C., Nielsen, A.J., Prince, T.A., Tan, H., Yeh, T.-S., Bellm, E.C., Dekany, R., Giomi, M., Graham, M.J.: Toward Efficient Detection of Small Near-Earth Asteroids Using the Zwicky Transient Facility (ZTF). *Publications of the Astronomical Society of the Pacific*. 131, 078002 (2019). <https://doi.org/10.1088/1538-3873/ab1b18>