Toward the Automated Detection of Light Echoes in Synoptic Surveys: Considerations on the Application of Deep Convolutional Neural Networks

Xiaolong Li1,2, Federica B. Bianco1,2,3,4, Gregory Dobler1,2,3, Roee Partoush5, Armin Rest5,6, Tatiana Acero-Cuellar7,8, Riley Clarke1, Willow Fox Fortino1, Somayeh Khakpash1,8, and Ming Lian1,9

1 Department of Physics and Astronomy, University of Delaware, Newark, DE 19716-2570, USA
2 Science Institute, University of Delaware, Newark, DE 19716-2570, USA
3 Biden School of Public Policy and Administration, University of Delaware, Newark, DE 19716-2570, USA
4 Vera C. Rubin Observatory, Tucson, AZ, 85719, USA
5 Department of Physics and Astronomy, Johns Hopkins University, 3400 North Charles Street, Baltimore, MD 21218, USA
6 Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, MD 21218, USA
7 Observatorio Astronómico Nacional, Universidad Nacional de Colombia, Bogotá, Colombia
8 Department of Physics and Astronomy, Rutgers, State University of New Jersey, 136 Frelinghuysen Road, Piscataway, New Jersey 08854, USA
9 Department of Physics, Lehigh University, Bethlehem, PA 18015, USA

Received 2021 December 6; revised 2022 September 13; accepted 2022 September 19; published 2022 November 15

Abstract

Light echoes (LEs) are the reflections of astrophysical transients off of interstellar dust. They are fascinating astronomical phenomena that enable studies of the scattering dust as well as of the original transients. LEs, however, are rare and extremely difficult to detect as they appear as faint, diffuse, time-evolving features. The detection of LEs still largely relies on human inspection of images, a method unfeasible in the era of large synoptic surveys. The Vera C. Rubin Observatory Legacy Survey of Space and Time (LSST) will generate an unprecedented amount of astronomical imaging data at high spatial resolution, exquisite image quality, and over tens of thousands of square degrees of sky: an ideal survey for LEs. However, the Rubin data processing pipelines are optimized for the detection of point sources and will entirely miss LEs. Over the past several years, artificial intelligence (AI) object-detection frameworks have achieved and surpassed real-time, human-level performance. In this work, we leverage a data set from the Asteroid Terrestrial-impact Last Alert System telescope to test a popular AI object-detection framework, You Only Look Once, or YOLO, developed by the computer-vision community, to demonstrate the potential of AI for the detection of LEs in astronomical images. We find that an AI framework can reach human-level performance even with a size- and quality-limited data set. We explore and highlight challenges, including class imbalance and label incompleteness, and road map the work required to build an end-to-end pipeline for the automated detection and study of LEs in high-throughput astronomical surveys.

Unified Astronomy Thesaurus concepts: Convolutional neural networks (1938); Interstellar dust (836); Transient detection (1957)

1. Introduction

Light echoes (LEs) are the reflections of light emitted by transients off of interstellar dust. While photons from a transient can reach us traveling along the line of sight, photons originally directed away from us can be reflected back to earth by favorably oriented dust sheets. The reflected image of the transient will inherit the complexity of the underlying dust structure, with morphological features at arcsecond or even arcminute spatial scales.

LEs provide invaluable information about the dust and the transient sources that originate them. They enable the study of dust and can reveal the dust structure around a transient in detail (Patat 2005). Through LEs, we have the unique opportunity to resee ancient transients detected in direct light decades or even centuries ago, and study them with new technology and instrumentation (Rest et al. 2011) so that LEs can be used, for example, to classify historical supernovae (Rest et al. 2005, 2008a, 2008b). This was in fact how the Tycho and Cas A supernovae were classified (Krause et al. 2008a, 2008b; Rest et al. 2008b). Finally, LEs can also provide a 3D view of an astronomical phenomenon, enabling the study of asymmetry of individual supernovae and supernova classes (Rest et al. 2011; Finn et al. 2016).

LEs are rare phenomena, as they require the serendipity of a bright transient and of the presence of a dust sheet oriented correctly to reflect the light toward the Earth. They are also extremely difficult to detect. The magnitude of an LE is ~10 times fainter than its transient source (Patat 2005). Thus, LEs bright enough for detection are extremely rare. Furthermore, the dynamic and complex dust environment makes their shape irregular and their identification more difficult than the identification of point-source transients. To date, no algorithm has proven successful in simultaneously detecting and localizing LEs in untargeted image surveys, and visual inspection of template-subtracted images is the framework generally adopted for detection (see Section 3).

The next generation of ground-based synoptic surveys, and in particular the Rubin Observatory Legacy Survey of Space and Time (LSST), will make visual inspection of images unsuitable for practical purposes. Rubin LSST is scheduled to observe the sky continuously for 10 yr starting in 2024. The telescope will have an 8.4 m (6.5 m effective) primary mirror, a 9.6 deg2 field of view, and a 3.2 Gigapixel camera, which will produce 20 Tb of information-dense data each night (Ivezic et al. 2019). LSST’s unique survey capability in the time...
domain and exquisite image quality will enable a vast and diverse range of scientific investigations. The LSST, with its unique combination of sensitivity, wide field of view, and dense observing cadence, has the capacity to revolutionize LE studies, pushing the field from novelty detections into statistically relevant regimes. However, with nearly 1000 images each night for 10 yr, it will become impossible to visually inspect the images to detect the LEs. An automated detection pipeline for LEs is needed to analyze the enormous outputs of this telescope.

The LSST data processing pipeline under development is scoped to detect millions of alerts every night: anything that significantly changes (at a 5σ level) in the night sky from a reference image. However, the pipeline is designed for point-like sources, and it will entirely miss diffuse and extended transients, such as LEs. Our ultimate goal is to create an end-to-end pipeline for the detection and study of LEs in the LSST era. With this paper, we set the stage by exploring the potential of region-based convolutional neural networks (CNNs) for LE discovery.

Automating the detection of LEs is a complex computer-vision problem. Over the past several years, deep learning has proven successful in discovering intricate structures in high-dimensional data, and especially in images (LeCun et al. 2015). Deep-learning-based object-detection frameworks have achieved real-time, high-accuracy performance (Zhao et al. 2019) that matches or surpasses that of humans. Deep neural networks (DNNs) have recently been used in the detection and study of astronomical sources, including the detection of galaxy clusters (Chan & Stott 2019), gravitational lenses (Davies et al. 2019), supernova remnants (Liu et al. 2019), and more. DNNs have also been applied to the detection of LEs in Bhullar et al. (2021). This work produced a sliding-window detection model, ALED, which is based on a capsule CNN architecture (Sabour et al. 2017). We compare this model in detail with our region-based model in the Results section of this paper (see Section 5).

In this paper, we present a proof-of-concept LE detection pipeline based on a robust and proven AI object-detection framework, the third publicly available version of the “You Only Look Once” (YOLOv3) model (Redmon & Farhadi 2018), applied to the Asteroid Terrestrial-impact Last Alert System (ATLAS; Tonry et al. 2018) telescope data. ATLAS provides a wide-field multiepoch survey. Compared to the LSST, the survey for which we ultimately want to build a pipeline, the ATLAS survey is shallow (mag range 11–19.5) and coarse (resolution 1786 pixel−1); however, these data provide an excellent test bed that allows us to demonstrate the potential of AI to detect LEs and to explore the subtleties and nuances of architectures suitable to accomplish this task.

This paper is organized as follows. We introduce the LE phenomenon and its history of discovery in Section 2. In Section 3 we describe how our data set is prepared. In Section 4, we present our detection framework, the definition of detection, and the method of evaluation. The results of our model are discussed in Section 5. Finally, in Section 6 we highlight the challenges and the future work required to build an end-to-end automated pipeline for the detection and study of LEs.

2. Historical Note on Light Echoes

LEs were first discovered around Nova Persei 1901 (Ritchey 1901, 1902). Since then, LEs have been observed from a variety of sources, including variable stars: Galactic Nova Sagittarii (Swope 1940), Galactic Cepheid RS Puppis (Havlen 1972), Nova Cygni (Bode & Evans 1985), OH 231.8+4.2 (Kastner et al. 1992, 1998), V838 Monocerotis (Bond et al. 2003), the T Tauri star S CrA (Ortiz et al. 2010), the Herbig Ae/Be star R CrA (Ortiz et al. 2010); historical supernovae, of which we have centuries-old records or that we only know from their remnants: SNR 0519-69.0 (Rest et al. 2005), SNR 0509-67.5 (Rest et al. 2005), SNR N103B (Rest et al. 2005), Cas A (Rest et al. 2007, 2008a; Krause et al. 2008a), Tycho (Rest et al. 2007, 2008a; Krause et al. 2008b); modern supernovae, observed directly with modern astrophysical instrumentation as well, as in LEs: 1987A (Crotts 1988; Suntzeff et al. 1988), 1980K (Sugerman et al. 2012), 1991T (Schmidt et al. 1994; Sparks et al. 1999), 1993J (Sugerman & Crotts 2002; Liu et al. 2003), 1995E (Quinn et al. 2006), 1998bu (Cappellaro et al. 2001), 1999ev (Maund & Smartt 2005), 2002hh (Welch et al. 2007; Otsuka et al. 2011), 2003gd (Sugerman 2005; Van Dyk et al. 2006; Otsuka et al. 2011), 2004et (Otsuka et al. 2011), 2006X (Crotts & Yourdon 2008; Wang et al. 2008), 2006bc (Otsuka et al. 2011; Gallagher et al. 2012) 2006gy (Miller et al. 2010), 2007it (Andrews et al. 2011), 2007af ( Drozdov et al. 2015), 2008bk (Van Dyk 2013), 2012aw (Van Dyk et al. 2015), 2014J (Crotts 2015; Yang et al. 2017), 2016ad (Sugerman & Lawrence 2016) and the massive eruptive stellar system η Carinae (Rest et al. 2012; Prieto et al. 2014; Smith et al. 2018a, 2018b).

The first systematic study of the phenomenon appeared in 1939 (Couderc 1939), and the theoretical framework was formalized further in Sugerman (2003), Tylenda (2004), Patat (2005), Patat et al. (2006), and Rest et al. (2011). As shown in Figure 1, at each point in time, for any given transient there is an ellipsoidal surface with foci at the Earth and the transient. If there is dust with the proper orientation on the ellipsoidal surface, photons that reach it are reflected toward the Earth. All points on the ellipsoid reflect light from the transient that arrives at the Earth at the same time (Patat 2005; Patat et al. 2006; Rest et al. 2011). Because of the increased path length, reflected light arrives at a delay compared to light traveling the direct path from the transient to the Earth. As time goes by, the ellipsoid expands and so does the travel path of LE light, giving us the opportunity to reobserve transients, even historical transients, that were directly observed centuries ago (Rest et al. 2005, 2007, 2008a, 2011, 2012).

In astronomical images, LEs appear as faint, morphologically diverse, diffuse, resolved features. When they are reflected by dust directly surrounding the transient, circumstellar dust ejected from stellar eruptions, for example, the LEs may appear as rings (see top plot of the right panel in Figure 1). Most commonly, however, and when the source is a centuries-old transient, they take the shape of the reflecting dust filaments along a segment of the reflection ellipsoid (see bottom plot of the right panel in Figure 1).

3. Data

In this work, we trained an AI model to detect LEs based on a data set of visually inspected template-subtracted images from the ATLAS survey.

Template-subtracted images can reveal transients that, due to their faint nature or the complex environment in which they arise (such as supernovae near the center of a galaxy or a star-forming region), are invisible to the naked eye in the original images. Template subtraction, or difference image analysis
portion of the ellipsoid will have dust that reflects the light from the original transient did. On the plane of the sky, we see the cross section of the reflection ellipsoid; yet not all LEs are circles: in most cases only a portion of the ellipsoid will have dust that reflects the light echo toward Earth. Most observed LEs are only arcs of a circle.

Figure 1. Slice of the geometry of the LE phenomenon (left) and its projection on the plane of the sky (right) in the case of a continuous dust sheet (top) and dust filaments (bottom). At each point in time, for every transient, there is an ellipsoidal surface with foci at the Earth and the transient. All points on the ellipsoid reflect light from the transient that arrive on Earth at the same time, and the ellipsoid itself expands over time. Due to the increase in path length, LEs reach the observer after the light from the original transient did. On the plane of the sky, we see the cross section of the reflection ellipsoid; yet not all LEs are circles: in most cases only a portion of the ellipsoid will have dust that reflects the light echo toward Earth. Most observed LEs are only arcs of a circle.

(DIA), was developed as a technique for detecting transients and monitoring microlensing events (Norgaard-Nielsen et al. 1989; Crotts 1992; Phillips & Davis 1995). Models based on the optimal image subtraction (OIS) method (Alard & Lupton 1998) are widely deployed to survey-based searches for moving objects and changing astronomical phenomena. Two images, of which one is considered a “template”, are aligned, point-spread function (PSF)-matched, and subtracted from one another. The template itself can be a single, or more commonly a composite, image (generally from the same survey) with image properties that reach or approach the highest image quality the survey is capable of acquiring in terms of overall noise, spatial resolution, and depth. After subtraction, the static background vanishes while any time-dependent phenomena, such as transients or moving objects, appear as a deviation from the zero-flux average.

In practice, the DIA process is prone to the generation of artifacts that complicate the detection of transients in general, and LEs in particular. We visually inspected images within a sky region ranging from decl. 40° to 90°, R.A. 0° to 360°. This corresponds to about 234,000 inspected images as each field is observed at five different epochs (see Table 2). We found 17 images that host LEs with sufficient signal-to-noise ratio (S/N) to enable unambiguous detection by human visual inspection.10 These LEs primarily originate from the Cass A and Tycho SNe, but we also found and included several LE groups in our training set that are inconsistent with originating from either of those SNe and whose source is being investigated by our team. Note that hereafter we define an LE as a single contiguous dust filament lit by a transient’s light, and an LE group as an ensemble of individual LEs originating from a complex structure of dust with many filaments lit at the same time. Most LEs in our images are in LE groups. Figure 2 shows an example image.

**Table 1**

| Parameters       | Value        |
|------------------|--------------|
| Camera           | Acam         |
| Format (pixels)  | 10560 × 10560|
| Pixel Scale      | 1.86         |
| Field of view    | 5.375 × 5.375|
| Number of filters| 7            |

Note. Additional specifications are available in https://atlas.fallingstar.com/specifications.php.

**Table 2**

| Parameters       | Value        |
|------------------|--------------|
| R.A.             | 0 360        |
| decl.            | −40 90       |
| Epochs (MJD)     | 58400, 58450, 58500, 58600, 58650 |
| Total number of images | 46800 × 5 |
| Filter           | r            |
| Image size after DIA processing | 1800 × 1800 pixels (1° × 1°) |
| Images with LEs  | 17 × 5       |

We visually inspected images within a sky region ranging from decl. 40° to 90°, R.A. 0° to 360°. This corresponds to about 234,000 inspected images as each field is observed at five different epochs (see Table 2). We found 17 images that host LEs with sufficient signal-to-noise ratio (S/N) to enable unambiguous detection by human visual inspection.10 These LEs primarily originate from the Cass A and Tycho SNe, but we also found and included several LE groups in our training set that are inconsistent with originating from either of those SNe and whose source is being investigated by our team. Note that hereafter we define an LE as a single contiguous dust filament lit by a transient’s light, and an LE group as an ensemble of individual LEs originating from a complex structure of dust with many filaments lit at the same time. Most LEs in our images are in LE groups. Figure 2 shows an example image.

10 We note here that defining the S/N of an LE is a challenging task: the morphology is complex and the edges of the features are not trivial to identify. We find that a functional definition of S/N can exploit the characteristic “dipole” morphology. As the light travels across a sheet of dust, it illuminates different regions. In a difference image, this results in clusters of pixel values positive on one side and negative on the other (pointing in the direction of the original transient). Our functional definition of LE S/N then exploits the contrast in the dipole and the size of the region covered by the LE, measured through a simple segmentation procedure, as $S/N_{LE} = IQR/\log_{10}(N_{pix})$, where $N_{pix}$ is the number of pixels covered by an LE, and IQR indicates the interquartile range of the values of those pixels. Under this metric the range of $S/N_{LE}$ in the 17 images we selected is (0.15, 0.85). Ultimately, we find that this metric generally tracks the score (see Section 4) that our model assigns to our LEs.
from ATLAS with an LE group composed of two LEs in the top left corner.

Figure 3 shows the sky locations of the 17 LEs; their coordinates and transient sources are listed in Table 3. These constitute our data set, the size of which is significantly smaller than those traditionally used to train region-based CNN object-detection models in the computer-vision literature (e.g., the original implementation of YOLO was trained on ImageNet; Russakovsky et al. 2015, which consisted of more than 1,000,000 images with 1000 different object classes).

In each image, we draw bounding boxes for three classes, LEs, stars, and “other”, which is a catchall class for additional features and artifacts that may be present in an image, such as star streaks, and satellite tracks. Notice that there is a certain amount of arbitrariness in the choice of the location and extent of an LE bounding box. Our bounding boxes strive to delimit each LE but end up, in many cases, including multiple LEs from the same LE group. We will return to this point when we describe our strategy for model evaluation in Section 4.2. We build our data set in the COCO (Lin et al. 2014) format using labelme.11

![Figure 2. Example of a template-subtracted (hereafter “difference”) image from the ATLAS survey (see Section 3 for details). The native resolution of the difference image is 1800 × 1800 pixels, corresponding to about 1 square degree. The images are template-subtracted to aid detectability of the faint, time-evolving LE features. The full-frame image (left) contains an LE in the top left corner, as well as stars and artifacts throughout. Because of the rarity of LEs, in order to mitigate class imbalance between stars and LEs in our data set, we split each ATLAS image into nine tiles and select only the tiles that contain LEs. The top left tile of this image that contains an LE is marked by an orange frame and is reproduced to the right with boxes corresponding to saturated stars, star streaks, and LEs shown in blue, teal, and red, respectively.](image)

![Figure 3. Sky positions of 17 LEs contained in our final data set: 10 of them are from the Tycho SN, and 2 are from Cas A. The rest have positions and orientations that are inconsistent with both Tycho and Cas A and are of as-of-yet unknown origin. LEs are preferentially found in high-dust-content regions, such as near the Galactic plane, which is indicated by a thick blue line in this plot. The right panel shows a zoom-in view of the region containing the majority of the LEs recovered in our data. Table 3 contains the coordinate of each of the LEs in this figure.](image)

### Table 3
| R.A. | Decl. | Gal l | Gal b | Source          |
|------|-------|-------|-------|-----------------|
| 4.0  | 57.0  | 118.1 | −5.5  | Tycho           |
| 6.0  | 57.0  | 119.2 | −5.7  | unknown         |
| 12.0 | 58.0  | 122.5 | −4.9  | Tycho           |
| 13.0 | 59.0  | 123.0 | −3.9  | Tycho           |
| 15.0 | 59.0  | 124.0 | −3.9  | Tycho           |
| 16.0 | 59.0  | 124.6 | −3.8  | Tycho           |
| 17.0 | 59.0  | 125.1 | −3.8  | Tycho           |
| 82.0 | 12.0  | 192.3 | −12.4 | unknown         |
| 320.0| 59.0  | 98.5  | 6.6   | unknown         |
| 325.0| 57.0  | 99.1  | 3.3   | unknown         |
| 326.0| 57.0  | 99.5  | 2.9   | unknown         |
| 339.0| 57.0  | 105.3 | −1.2  | Cas A           |
| 348.0| 64.0  | 112.3 | 3.2   | Cas A           |
| 350.0| 60.0  | 111.7 | −0.9  | Tycho           |
| 352.0| 60.0  | 112.7 | −1.2  | Tycho           |
| 353.0| 59.0  | 112.8 | −2.3  | Tycho           |
| 354.0| 59.0  | 113.3 | −2.5  | Tycho           |

**Note.** The coordinates reported are for the difference image centroid.

11 https://github.com/wdentaro/labelme
Because LEs are rare phenomena, much rarer than stars, there exists a large class imbalance between the number of objects in our three classes, and we note that imbalance in the training data set can potentially introduce a bias in learning (Oksuz et al. 2020) and should therefore be addressed. The root of this potential bias lies in the fact that overrepresented classes have more influence on the loss function, when the loss function is calculated as an average or median across all examples. Class imbalance in the training of machine-learning models might thus lead to better performance on over-represented classes relative to underrepresented ones. However, we also note that in-class diversity and the S/N of individual objects are other important factors that may either reduce or enhance this effect. While we will implement a modification of the YOLO loss function that will help mitigate the risk of this bias (see Section 4), we also address the class imbalance in data preparation. We split the images into 576 × 576 pixel subimages and only include in our training and test data sets the image segments that host LEs, for a total of 28 576 × 576 images.

We apply three data augmentation techniques to expand our data set: we first flip each image horizontally, and then rotate each image by 90°, 180°, and 270°. Figure 4 shows an example of the results of image augmentation, and Table 4 shows the number of images available in our data set after each step of our augmentation process. While other augmentations are certainly possible, including rotation by angles not a multiple of 90°, augmentations that only modify the region of the image containing the LE, blurring, cut-and-paste, etc., at this stage we conservatively only apply augmentations that do not require interpolation nor modeling of the distribution of noise or signal. In future work we will explore more aggressive augmentations as well as simulations to increase the size of the training data. The final data set contains 224 images, 576 boxes labeled as LEs, and 1248 labeled as stars.

On average, each image contains three LEs, six stars, and four “other” objects. Figure 5 shows the distribution of number of bounding boxes for both LEs, stars, and “other” in each image, and the size distribution of the boxes. The size of the bounding box is measured as the ratio of the diagonal of the box and the diagonal of the image. The mean size of LE boxes (as well as “other” boxes) in these units is ∼0.1 (∼3'), while the stars are smaller, ∼0.06 (∼1/8).

Finally, the pixel values in images that are fed into a neural network need to be appropriately scaled. Unlike normal 8-bit color RGB images, imaging data from the ATLAS telescope are stored as FITS files (Wells & Greisen 1979). In our data, which is entirely comprised of template-subtracted images, the FITS pixel values range from about −10,000 to 5000. To feed the images as input to a traditional neural network architecture, however, a normalization is required. In order to constrain the pixel value range while minimizing the loss of information, we clip the raw FITS array to [−10, 10], a range that contains about 95% of the pixel values in our data. We then standardize the clipped images by subtracting the mean and dividing the result by the standard deviation of the distribution of pixel values in the clipped FITS images.

---

12 YOLOv3 accepts input images of any size; however we choose 576 to facilitate potential future work and model comparisons as some CNN models require the input size to be divisible by 64.

13 We will reintroduce images without LEs at the end of our model development process to test and discuss how our model performs with image sets that, for the most part, do not contain any LEs, such as the ones we expect from LSST and any other survey (see Section 5).
LEs are not finite objects: they exist at all levels of S/N, and their edges blend into the background as the dust that reflects them blends into empty space and the light they reflect fades. Thus we cannot aspire to be complete in our human labeling because where LEs blend into the noise, human detection fails. As our training set is not complete, we have to teach the CNN both what is and what is not an LE: a multilabel model. In this initial work we only train the CNN, we cannot expect that the model will generate bounding boxes that exactly overlap each one of the label boxes.

1. LEs are not finite objects: they exist at all levels of S/N, and their edges blend into the background as the dust that reflects them blends into empty space and the light they reflect fades. Thus we cannot aspire to be complete in our human labeling because where LEs blend into the noise, human detection fails. As our training set is not complete, we have to teach the CNN both what is and what is not an LE: a multilabel model. In this initial work we only train the model to identify LEs, stars, and a limited number of artifacts.

2. Of the many features in an image, LEs are an extremely rare occurrence. Stars and galaxies, transients and variable phenomena, moving objects, and several artifacts overwhelm the image searches. Detection of rare features is a problem often encountered in medical applications of computer vision (Sánchez Fernández et al. 2020, and references therein). We explore the detection of LEs in a subset of our data where each image contains at least one LE (as described in Section 3) because we want to focus on the architectural choices that enable the detection of LE features.

3. Putting bounding boxes around LEs requires us to make choices that are, to some extent, arbitrary: each LE inherits a complex structure from the underlying dust, and most LE filaments are connected into complex structures (an LE group; Section 3). While we strive for consistency in setting the bounding boxes around isolated filaments, we cannot expect that the model will generate bounding boxes that exactly overlap each one of the label boxes. We address this issue by modifying the region-based CNN loss (see Section 4.1) exploring various approaches to mitigating this problem, including the application of focal loss (Lin et al. 2017; Vaswani et al. 2017), and with a specific definition of true and false positives (see Section 4.2) that differ in the training and evaluation phases.
YOLO (Redmon & Farhadi 2018) is a popular single-stage regression/classification region-based framework, where “regression” refers to the identification of regions of interest through regression of bounding-box coordinates and “classification” refers to the classification of the object inside of those bounding-box regions. Feature extraction and object localization are integrated into a single deep neural network model. An input image is split into an $S \times S$ grid of cells. Each cell is responsible for both class prediction and bounding-box regression of objects whose centers fall inside the cell. The YOLOv3 model makes predictions at three size scales (1/8, 1/16, 1/32 of the image size).

### 4.1. YOLOv3

YOLO (Redmon & Farhadi 2018) is a popular single-stage regression/classification region-based framework, where “regression” refers to the identification of regions of interest through regression of bounding-box coordinates and “classification” refers to the classification of the object inside of those bounding-box regions. Feature extraction and object localization are integrated into a single deep neural network model. An input image is split into an $S \times S$ grid of cells. Each cell is responsible for both class prediction and bounding-box regression of objects whose centers fall inside the cell. The YOLOv3 model makes predictions at three size scales (1/8, 1/16, 1/32 of the image size).

The predicted bounding boxes are first filtered by thresholding on the probabilistic classification score $p_i$, which is the confidence $C$ times the probability $p_i$ that the object is of a certain class, $p_i = C p_{G>i}$, in order to eliminate boxes with a low score. Nonmaximum suppression (NMS; Neubeck & Van Gool 2006) is then applied to keep only one box when several boxes overlap with each other.

YOLOv3 minimizes a loss function that consists of three parts: confidence loss $L_{\text{conf}}$, classification loss $L_{\text{cls}}$, and bounding-box regression loss $L_{\text{box}}$.

$$L_{\text{total}} = L_{\text{conf}} + L_{\text{cls}} + L_{\text{box}}.$$  \hspace{1cm} (1)

The confidence loss $L_{\text{conf}}$ represents the loss from the confidence probability $C$ (1 if there is an object in the bounding box and 0 for background), while $L_{\text{cls}}$ measures the ability to make a correct classification. We use a simple cross-entropy loss,

$$L_{\text{cls}} = \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\text{obj}} f_{CE}(p_{ij}, \hat{p}_{ij}),$$  \hspace{1cm} (2)

where $S^2$ denotes the three scales of the anchor boxes, and $B$ is the total number of anchor boxes assigned to each grid cell. $I_{ij}^{\text{obj}}$ denotes whether there is an object in the cell (it is 1 when there is an object and 0 otherwise), while $f_{CE}$ stands for the cross-entropy function given by:

$$f_{CE}(x, y) = -y \log(x) - (1 - y) \log(1 - x).$$  \hspace{1cm} (3)

The number of background anchor boxes is typically much larger than the number of foreground anchor boxes. To address this problem, Lin et al. (2017) proposed to use the “focal” loss, which adds a factor to the cross entropy loss. We adopt a similar approach for the detection of LEs because, as shown in Section 3, on average there are only three LEs in each image. We redesign $L_{\text{conf}}$ by using logarithmic loss to down-weight the contribution of background anchor boxes, resulting in a focal loss.

$$L_{\text{conf}} = \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\text{obj}} (1 - C_{ij})^\gamma f_{CE}(C_{ij}, \hat{C}_{ij})$$

$$+ \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij}^{\text{no-obj}} (1 - C_{ij})^\gamma f_{CE}(C_{ij}, \hat{C}_{ij}),$$ \hspace{1cm} (4)

where $I_{ij}^{\text{obj}}$ and $I_{ij}^{\text{no-obj}}$ are mask factors that are the same as in $L_{\text{cls}}$, except $I_{ij}^{\text{no-obj}} = 1$ when there are no objects in the cell. $f_{CE}$ is the cross entropy function defined in Equation (3), and $(1 - C_{ij})^\gamma$ is the focal loss term. We set the value $\gamma = 2$ as suggested in Lin et al. (2017).

$L_{\text{box}}$ quantifies the coordinate difference between predicted and labeled bounding boxes. The most commonly used method for comparing the similarity between two bounding boxes is by computing the intersection over union (IoU), the ratio between the area of the intersection and the area of the union of two boxes. The more overlap, the greater the IoU, and the more closely aligned are the two boxes. Then the loss based on IoU

---

14 As pointed out in Redmon & Farhadi (2018), Darknet-53 is a powerful model, which has comparable accuracy to ResNet-101 (He et al. 2016) used in Faster RCNN (Ren et al. 2015), but the speed is 1.5× faster. Our preliminary work comparing results based on YOLO with results obtained from Faster RCNN does not suggest that the one-stage YOLO architecture is less suitable for the detection of LEs than a more traditional two-stage RCNN model when comparing model accuracies for the two.
is just \( L_{\text{IoU}} = 1 - \text{IoU} \). In YOLOv3, the IoU loss is weighted by a factor related to the area of the boxes to improve the performance for small boxes. Overall,

\[
L_{\text{box}} = \sum_{i=0}^{S^2} \sum_{j=0}^{B} I_{ij} (1 - \text{IoU}) (2 - \hat{w}_{ij} \hat{h}_{ij}),
\]

where \( \hat{w}_{ij} \) and \( \hat{h}_{ij} \) are the width and height of the labeled box.

However, as we have described above, it is particularly tricky to label the LEs in an unambiguous way. Inside of a label box there are often multiple LEs and the LE grouping method is somewhat arbitrary. In the face of this ambiguity, the IoU loss is not able to properly reflect the properties of LEs, and a traditional IoU may penalize good detections that split an LE group differently than a visual inspector, but in just as valid a way. To enable the model to learn features inside the box and avoid penalizing fair choices that happen to be different than the ones implemented by the human labeler, we replace the IoU with the IoP, intersection over prediction (where prediction is the area of the annotation label), in the evaluation phase. Simultaneously we generate larger annotation boxes for evaluation. In the training phase, we prepare annotations designed to separate each LE in a group. In evaluation, however, we use annotations designed to include all LEs in an LE group, so that a different segmentation of the LE group would not be penalized by our model (see Figure 7).

### 4.2. Detection Definition and Evaluation

While the loss minimization determines the parameter values of a model, some parameters have to be set by the user to select the optimal model for object detection. In the case of our CNN, in addition to the hyperparameter choices indicated in Table 5, this means choosing the thresholds for detection and classification above which a detection is considered valid. There are two thresholds that need to be chosen: (1) a “score threshold”, which is a threshold set on the probabilistic classification score \( p_s \), and (2) an IoP (or IoU) threshold.

Selecting the threshold values defines true and false positives based on the overlap of a label and a bounding box and the probabilistic classification of the object inside the box. We consider a detected bounding box as a true positive (TP) if

1. the predicted class is the same as labeled (star or LE) and with score \( p_s \) larger than the score threshold;
2. the overlap between predicted and ground truth bounding boxes measured as their intersection over the prediction (or union) is larger than the IoP (or IoU) threshold.

Conversely, we count it as a false positive (FP), if

1. the predicted class is the same as labeled with \( p_s \) larger than a threshold, but IoP (or IoU) is below the threshold.

Based on the properties of LEs discussed above, the standard IoU is not a good evaluation metric for LEs. The LEs are extended features without clear boundaries, and valid prediction boxes that fall within an annotation box would have low IoU. Taking this into consideration, when searching for LEs, we replace the IoU with the IoP for LEs, while for the stars and the “other” category we continue to use the standard IoU as the location and boundaries of the bounding box are generally not ambiguous. This reduces the number of false negatives (FNs). Finally, predicted boxes that overlap with the same label box

![Figure 7](image-url)

**Table 5. Hyperparameters of YOLOv3**

| Parameters       | Value          |
|------------------|----------------|
| Backbone         | Darknet-53     |
| Feature strides  | [8, 16, 32]    |
| Anchor per scale | 3              |
| Initial learning rate | \(10^{-4}\) |
| Final learning rate  | \(10^{-6}\)   |

The Astronomical Journal, 164:250 (13pp), 2022 December

Li et al.
(see Figure 7) need to be consolidated so as to not double count TPs.

Once we define the TP and FP, we are able to determine the receiver operating characteristic (ROC) curve, which indicates the number of TPs versus FPs at different thresholds. The two-dimensional area underneath the curve (AUC) measures the overall performance of the model: the larger the area, the better the model. In addition, we can also then calculate the precision–recall (PR) curve for various IoU (or IoP for LEs) and classification score thresholds.

Precision and recall are defined as

\[ P = \frac{TP}{TP + FP} = \frac{TP}{N_{\text{det}}} \]
\[ R = \frac{TP}{TP + FN} = \frac{TP}{N_{\text{ann}}} \]

where \( N_{\text{det}} \) is the total number of detected boxes, and \( N_{\text{ann}} \) is the total number of labeled boxes. Precision measures the fraction of positive detections that are correctly classified, while recall measures the completeness fraction. Precision and recall are often in tension: improving precision by increasing the classification and/or detection threshold, typically reduces recall, and vice versa. Therefore, the F1 score, defined as the harmonic mean of precision and recall, is often used as a measure of balance between precision and recall.

The evaluation steps for our models are summarized in Algorithm 1.

**Algorithm 1. LE detection model evaluation**

\[ N_{\text{ann}} \leftarrow \text{number of labeled objects}; \]
\[ N_{\text{det}} \leftarrow \text{number of predicted bounding boxes}; \]
set classification score threshold;
set IoP detection threshold;
\[ N_{TP} \leftarrow 0; \]
\[ N_{FP} \leftarrow 0; \]
for each predicted bounding box do
if classification score > threshold then
if IoP > IoP threshold then
\[ N_{TP} \leftarrow N_{TP} + 1 \]
else
\[ N_{FP} \leftarrow N_{FP} + 1 \]
end
end
\[ P = \frac{N_{TP}}{N_{\text{det}}}, \]
\[ R = \frac{N_{TP}}{N_{\text{ann}}}, \]
\[ F1 = \frac{2PR}{P + R} \]

4.3. Training

We trained our model on the Google Colaboratory platform. We adopt a threefold cross-validation scheme in which 28 images containing LEs were randomly split into 10, 10, and 8 subsets. We kept one subset as a test set, and trained the model on the other two subsets. We shuffled the images once and repeated this process, thus obtaining six cross-validation sets. Due to the very limited size of our data set, we did not create a holdout set but only a training and a test set for each cross-validation instance. The results we present in Section 5 represent the average of the results across these six cross-validation sets. We used an Adam optimizer (Kingma & Ba 2014) with an adaptive learning rate, setting the initial and final values to \( 10^{-4} \) and \( 10^{-6} \), respectively. We applied a learning rate schedule to improve the training progress as described in He et al. (2019). The learning rate is decreased from its initial value following a cosine function.

5. Results

Figure 7 shows an example of the YOLO predictions from our test set. The figure demonstrates how replacing the IoU with the IoP in model evaluation leads to a better definition of FPs for LEs: the model successfully detected several LEs belonging to the same LE group within the labeled bounding box. Note that we replace the IoU with the IoP only for LEs. For the stars and “other” we continue to use the IoU as the prediction box annotations are much less unambiguous, and the predictions are never in fact contained in the label boxes, so replacing the IoU is unnecessary.

Figure 8 shows our performance over multiple training epochs averaged over the six cross-validation sets. The top panel in the figure shows the total loss (Equation (1)) for the training and test sets. The training loss continues to decrease throughout all epochs, while the test loss reaches a minimum around 50 epochs and then rises slowly due to overfitting. The precision and recall for LEs for the test set are shown in the bottom panel. In this figure, the precision and recall are calculated with a classification score threshold of 0.2 and IoP threshold of 0.2. The F1 score is the harmonic mean of precision and recall, demonstrating how a balance between the two is achieved near 50 epochs.
IoP threshold of 0.2. After \( \sim 50 \) epochs, the \( F_1 \) score stabilizes, and we choose the model state at 50 epochs as the final model for further evaluation.

The ROC and PR curves show the model performance for a given box-overlap threshold as the threshold on the classification probability is modified. The ROC and PR curves for IoP (IoU) thresholds for LEs (stars) from 0 to 0.5 are shown in Figure 9. The top panel in Figure 9 shows that, for this particular cross-validation instance, if we choose an \( IoP > 0.2 \), \( R > 70\% \) recall is achieved at a precision of about \( P \sim 95\% \) (score = 0.4). This means that we have to tolerate a 5\% FP rate to detect 70\% of LEs.

With the large amount of astrophysical images to be produced with our target survey, LSST, we must control FPs in our model or else we will be overwhelmed with an unmanageable number of images to visually inspect. Or worse, we may deploy and waste follow-up resources if we trust the automated detections without vetting them. As indicated in Section 3, we prepared a training set that contains LEs in every image, but LEs are rare events and the vast majority of images from any surveys contain none. To test the robustness of the model, we added 100 images to the test set that contain stars, potentially artifacts, but no LEs. Following the same evaluation methods described in Section 4.2, we found that precision drops with the score threshold, but the rate of FPs grows sublinearly with the number of added images. The slope at score 0.5 is about 0.1, while at score 0.2, it is 0.3. We also found that these FPs largely appear near the edge of stars or where dipole structures—structures associated with poor image subtraction due to mismatched PSF between template and search image or poor local image alignment—are present (see Figure 10).

Both our model and ALED (Bhullar et al. 2021), the only other automated LE detection method we are aware of, can achieve a high \( P \sim 90\% \) precision and high recall \( R \sim 90\% \), although, as the models were applied to different data sets, it is not straightforward to compare their performance at this time.

---

**Figure 9.** Results from YOLOv3 for LEs (a) and stars (b) for a chosen cross-validation set. In both (a) and (b), the left panel shows a scatter plot of classification score vs. IoP (top, for LEs) or IoU (bottom, for stars) for all predicted bounding boxes. The middle panels show ROC curves and the right panels show the PR curves. The gray regions in the left plots delimit the region of rejected predictions (below a score threshold of 0.2) and the orange regions represent the location where FPs are found (IoP or IoU smaller than 0.2). The ROC (middle) and PR curves (right) are plotted for different score thresholds (points along each curve) and different IoP or IoU (as labeled). In the ROC curve plots, the three black lines indicate a ratio of TP to FPs of \( 1/1 \), \( 1/5 \) and \( 1/10 \). If we choose an \( IoP > 0.2 \) and a classification score threshold of 0.4, larger than 70\% recall is achieved at a precision of about 95\%. This means that we have to tolerate a 5\% FPs to detect 70\% of LEs.
We note that the final results reported for our model are the median across the six cross-validation instances, while it appears that ALED did not perform cross validation. On a single set, our precision and recall can be as high as 0.9 simultaneously (see Figure 9), but perhaps the most fair comparison is that with the model reported in Table 4 of Bhullar et al. (2021) as “ALED-m”: a precision of $P = 1$ and a recall of $R = 0.25$. When taking the median of the score across the test sets for each cross-validation instance, at $P_{\text{median}} = 1$ we measure $R_{\text{median}} = 0.3$ for score and IoP thresholds of 0.7 and 0.2, respectively. While both data sets used for training the models are by no means large, our data set contains only 28 of 576 images with LEs (before augmentation) compared to 175 of 200 × 200 images in Bhullar et al. (2021). There are significant differences in the methodological approach. First, the ALED model identifies the presence of a light echo in an image (via binary classification) but does not locate the LE directly. The YOLO framework performs the detection and localization simultaneously and returns the class type as well as the exact coordinates of positions, which is important if we envision further automation of the LE study pipeline, such as automatic selection of LEs for follow-up and automatic follow-up on robotic systems, an increasingly common practice in astrophysics (e.g., Street et al. 2018). We note, however, that the location of the LEs in the images can be estimated in a follow-up step through the analysis of the ALED feature maps. In addition, the training of ALED requires a balanced data set, which must contain an equal number of LE and non-LE images, while our implementation can tolerate class imbalance through the application of a focal loss. Furthermore, the usage of region-based CNNs is particularly suited to application on image sequences, and applying our model to time-resolved LE image sequences is an ongoing project.

All plots and associated model outputs can be found at the project GitHub repository AILE.15

6. Conclusion and Future Work

In this paper, we described a novel LE detection model based on an AI object-detection framework applied to a data set of template-subtracted images from the ATLAS survey. As described in Section 3, we labeled three classes: LEs (diffuse features from reflections of transients off of interstellar dust), stars (point-like sources), and “other” (a catchall category used mostly for artifacts). We designed the detection model based on the architecture of YOLOv3, as described in Section 4. Unlike traditional object-detection tasks, LEs are diverse, extended features, their shapes vary over sky positions as well as time, their edges are not sharp and blend into the background, and their structure is complex. Thus, human labels can be quite subjective. Therefore, we modified the loss function used to train the model to incorporate both single LEs and LE groups. We found that an $F_1 \sim 0.7$ can be achieved at precision 70% even with an extremely limited training set, establishing region-based CNNs as viable detection model architectures for localization and classification of LEs within an astronomical image.

This work demonstrates that modern region-based object-detection architectures—i.e., architectures that identify and then classify regions of interest as opposed to sliding-window, classification-only architectures that classify every subregion of an image—produce detection accuracies for LEs that allow for the searching of extremely large upcoming data sets for these objects, a task which may very well be unfeasible with sliding-window classification models. While it is true that these region-based models are rapidly evolving (YOLO itself is now at version 7; Wang et al. 2022, and other more recent models include Chen et al. 2019; Wu et al. 2019 for example), the core conceptual foundations are common to all such models. That is, the models presented in this paper have demonstrated that region-based models are applicable to astronomical data (low S/N, very high spatial resolution, etc.) for the detection of diffuse, morphologically complex features at the limit of the image S/N, suggesting that as these models are improved and new models (e.g., transformer-based architectures) are developed within the computer-vision community, they will likely also be applicable to this problem when appropriate steps are taken to perform specific image preparation and definition of appropriate model architecture elements, including specialized loss functions.

While the aim of this paper was to provide a proof-of-concept, representing the first step toward the goal of building an AI-based LE detection platform in the Rubin LSST era, we

---

15 https://github.com/xiaolong/AILE
have in fact made significant progress despite limitations in the presently available data. We achieved relatively high model accuracies ($F1 \approx 0.7$) with a very limited training data set size. We demonstrated that the number of false positives scales sublinearly with the number of images Figure 10 suggesting that our method can be successfully trained on and applied to large data sets without being overwhelmed by false positives or losing detection power. Furthermore, our model is also robust to significant class imbalance between labeled classes (~600 LEs, ~1300 stars).

Given that the ATLAS survey has relatively few LEs with uniquely characteristic image properties, this raises the question as to the transferability of the models presented in this work. We note however that this is in fact a strength of our study as, while it is certainly true that any new data set will require new training and testing sets to be created and it is used to train the model, we have demonstrated that achieving reasonable testing accuracy is possible even with very few training examples with significant image noise and confounding non-LE artifacts and objects. We expect then that for upcoming data sets with higher S/N and more potential LE and non-LE objects on which to train, models can only be refined further. That is, as the models presented here are able to achieve reasonable accuracy, this provides confidence that at least comparable (but likely much higher) accuracy on larger and cleaner data sets will be possible.

The ambitious scientific goals that LSST has set to pursue (Ivezić et al. 2019) impose unprecedented requirements on image quality, including PSF characterization, and control over systematics in the images (see for example The LSST Dark Energy Science Collaboration et al. 2018). Thus, the noise properties of the Rubin LSST images are expected to be far superior to current synoptic surveys, including ATLAS. By developing a model able to detect bright LEs in ATLAS images we have implicitly demonstrated that we will be able to push to far fainter detection limits in the LSST era. Particularly, the $5\sigma$ limit of the survey is expected to be reached at a magnitude of $g \sim 24.5$ and $r \sim 24.0$, enabling the detection of far fainter LEs than the ATLAS survey; the plate scale of 0.245\arcsec will lead to seeing limited images at a site where subarcsecond seeing is common, enabling an accurate characterization of the shape of the PSF that allows for an LSST DIA pipeline expected to produce a small number of image artifacts, including in the high-density, high-dust regions of the Galactic plane that LEs prefer, leading to fewer FP detections; the LSST survey will collect a median of 815 images for every position in the sky, which will enable the creation of increasingly more accurate templates over the lifetime of the survey, enabling progressive improvements in the detection of LEs in LSST.

In future work, we will expand our training/testing data sets with both additional observational data as well as additional augmentations and simulations to increase the accuracy of the model and its performance against a variety of morphologies and to push into ever lower S/N regimes.

Two important LE features remain to be explored: their color and their time evolution. The ATLAS data included five observing epochs, while each Rubin LSST field will be observed at least 800 times over the 10 yr of the survey (with more epochs available in selected fields). The presence of an LE in multiple images collected at different times will confirm the existence of the LE and help remove FPs (for example those caused by light reflection inside of the telescope optical tube). Images taken from different epochs also provide an opportunity to study the time evolution of LEs. Leveraging the multiepoch nature of the ATLAS data set will be explored in future work. Furthermore, while ATLAS provides images in a single optical band while Rubin will observe in six filters. Exploiting the color properties of LEs in their detection and classification is also part of our plan for future work.

This paper was supported by the National Science Foundation grant No. 2108841: Detecting and studying light echoes in the era of Rubin and Artificial Intelligence NSF; and No. 1814993: An Astronomical Time Machine: Light Echoes from Historic Supernovae. X.L. and F.B.B. were additionally supported by the University of Delaware General University Research (GUR) grant No. 20A00782. The authors acknowledge the support of the Vera C. Rubin Legacy Survey of Space Time Science Collaborations and particularly of the Transient and Variable Star Science Collaboration (TVS SC) that provided opportunities for collaboration and exchange of ideas and knowledge. This work has made use of data from the Asteroid Terrestrial-impact Last Alert System (ATLAS) project. The ATLAS project is primarily funded to search for near-Earth asteroids through NASA grants NN12AR55G, 80NSSC18K0284, and 80NSSC18K1575; by products of the Near-Earth Object (NEO) search include images and catalogs from the survey area.

We used the following software packages: astropy (Astropy Collaboration et al. 2013, 2018), numpy (Harris et al. 2020), pandas (Wes McKinney 2010; pandas development team 2020), matplotlib (Hunter 2007), openpy (Bradski 2000), TensorFlow (Abadi et al. 2015), and labelme (Wada 2016).

Our data set is augmented using modules published in a dedicated GitHub repository. The implementation of the YOLOv3 architecture is based on code published in a dedicated GitHub repository.

**ORCID iDs**

Xiaolong Li https://orcid.org/0000-0002-0514-5650

Federica B. Bianco https://orcid.org/0000-0003-1953-8727

Gregory Dobler https://orcid.org/0000-0002-9276-3261

Roei Partoush https://orcid.org/0000-0001-6149-4209

Armin Rest https://orcid.org/0000-0002-4410-5387

Tatiana Acero-Cuellar https://orcid.org/0000-0002-5947-2454

Riley Clarke https://orcid.org/0000-0001-9273-5036

Willow Fox Porto https://orcid.org/0000-0001-7559-7890

Somayeh Khakpash https://orcid.org/0000-0002-1910-7065

Ming Lian https://orcid.org/0000-0002-4670-4930

**References**

Abadi, M., Agarwal, A., Barham, P., et al. 2015, TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, https://tensorflow.org

Alard, C., & Lupton, R. H. 1998, ApJ, 503, 325

https://www.lsstcorporation.org/science-collaborations

https://lsst-tvssc.github.io/

https://github.com/Paperspace/DataAugmentationForObjectDetection

https://github.com/pythonlessons/TensorFlow-2.x-YOLOv3
