Domain Adaptation via Minimax Entropy for Real/Bogus Classification of Astronomical Alerts

Guillermo Cabrera-Vives [1 2 3] **César Bolivar** [1] **Francisco Förster** [4 3 5] **Alejandra M. Muñoz Arancibia** [3 5]
Manuel Pérez-Carrasco [2 1 3] **Esteban Reyes** [6]

Abstract

Time domain astronomy is advancing towards the analysis of multiple massive datasets in real time, prompting the development of multi-stream machine learning models. In this work, we study Domain Adaptation (DA) for real/bogus classification of astronomical alerts using four different datasets: HiTS, DES, ATLAS, and ZTF. We study the domain shift between these datasets, and improve a naive deep learning classification model by using a fine tuning approach and semi-supervised deep DA via Minimax Entropy (MME). We compare the balanced accuracy of these models for different source-target scenarios. We find that both the fine tuning and MME models improve significantly the base model with as few as one labeled item per class coming from the target dataset, but that the MME does not compromise its performance on the source dataset.

1. Introduction

Time-domain survey telescopes are providing astronomers with vast amounts of data on celestial objects and phenomena. Surveys such as the Asteroid Terrestrial-impact Last Alert System (ATLAS; Tonry et al., 2018) or the Zwicky Transient Facility (ZTF; Bellm et al., 2018) emit when the brightness or location of a source change, producing a continuous astronomical alert stream. The aggregation, annotation, and classification of alerts in a rapid and consistent fashion is done by astronomical alert brokers (Narayan et al., 2018; Nordin et al., 2019; Smith, 2019; Förster et al., 2021; Möller et al., 2021). An important number of these alerts are *bogus* artifacts created by the image reduction pipelines, hence, the importance of creating real/bogus classification algorithms which have proven to be extremely useful for detecting real astrophysical phenomena. During the last decade, most of these algorithms have been based on Convolutional Neural Networks (Cabrera-Vives et al., 2016; 2017; Reyes et al., 2018; Duev et al., 2019; Turpin et al., 2020; Yin et al., 2021; Rabeendran & Denneau, 2021) which need a significant amount of data to be trained. Domain adaptation (DA) techniques such as the Minimax Entropy (MME; Saito et al., 2019) approach are an alternative that help training such models with fewer amount of data. Furthermore, DA allows models to perform inference simultaneously for multiple dataset that may follow different distributions. This is particularly important when developing multi-stream models for alert streams from next-generation telescopes such as the Vera Rubin Observatory as soon they start producing data. By effectively working across various alert streams and accounting for domain shifts, these models can leverage labeled and unlabeled data from multiple domains, enhancing their learning capabilities. Moreover, conducting inference on multiple alert streams using a single model facilitates performance monitoring across surveys.

In this work, we evaluate the use of fine tuning and MME for the real/bogus classification of alert stamps and their availability to transfer knowledge from models trained on *source* surveys to different *target* surveys using few shots of labeled sources. We start by describing the four datasets we use (HiTS, DES, ATLAS, and ZTF) in Section 2. In Section 3, we provide a comprehensive description of our feature extraction and classification models, as well as the domain adaptation techniques employed. We outline the details of our experiments in Section 4, followed by the presentation of the obtained results in Section 5. Finally, we draw conclusions based on these findings in Section 6.

[1] Department of Computer Science, Universidad de Concepción, Concepción, Chile [2] Data Science Unit, Universidad de Concepción, Concepción, Chile [3] Millennium Institute of Astrophysics, Chile [4] Data and Artificial Intelligence Initiative (IDIA), Faculty of Physical and Mathematical Sciences, University of Chile, Chile [5] Center for Mathematical Modeling (CMM), Universidad de Chile, Chile [6] Fintual Administradora General de Fondos S.A., Santiago, Chile. Correspondence to: Guillermo Cabrera-Vives <guillecabrera@inf.udec.cl>.

ICML 2023 Workshop on Machine Learning for Astrophysics, Honolulu, Hawaii, USA. PMLR 202, 2023. Copyright 2023 by the author(s).

2. Data

We use image stamps from four surveys: the High Cadence Transient Survey (HiTS; Förster et al., 2016), the Dark Energy Survey (DES; Goldstein et al., 2015), ATLAS (Tonry et al., 2018), and ZTF (Dekany et al., 2020). These four datasets consist of astronomical alerts represented as 3-channel images: 1) a reference image, 2) a science image taken at the time of observation, and 3) a difference image created by matching the point-spread-function of the science and reference images and subtracting them. Each alert within every dataset was assigned a corresponding label, indicating whether it is deemed ”real” (representing an astronomical event of interest) or ”bogus” (indicating a false detection). All images were cropped to 21 × 21 pixels and were normalized to have a mean of 0 and a standard deviation of 1.

The primary goal of the HiTS survey was to detect supernovae during their earliest hours of explosions. Their real/bogus dataset consist of a total of 1,437,684 images of 21 × 21 pixels (Cabrera-Vives et al., 2017). Bogus stamps were directly taken by the Dark Energy Camera (Flaugher et al., 2015) while real stamps were simulated within their pipeline. By construction, this dataset contains a total of 718,842 “real” stamps and 718,842 “bogus” stamps.

The DES dataset was obtained from Goldstein et al. 2015[1] and it contains 51× 51 pixels stamps from 898,963 source candidates. Of these candidates, 454,092 are simulated supernovae labeled as ”real”, and 444,871 are ”bogus” sources that came out of the DES pipeline (Abbott et al., 2018).

ATLAS is a sky survey system that aims at finding dangerous near-Earth asteroids. We use 61×61 pixels stamps coming from 3,678 candidate sources. This dataset was visually labeled and is composed of 500 persistent burn trails, 500 cosmic rays, 500 spike artifacts, 500 noise fluctuations, 500 sources labeled as asteroids, and 678 candidates labeled as asteroid streaks. The ”real” dataset was created by joining the labeled asteroids and asteroid streaks, while the ”bogus” class was created by combining the burn trails, cosmic rays, spikes, and noise.

We gathered ZTF stamps following the procedure described by Carrasco-Davis et al. 2021. The raw dataset consists of 63 × 63 pixels stamps for a total of 36,262 source candidates, but 467 images were of a smaller resolution and were discarded. This dataset originally had 9,996 images labeled as active galactic nuclei, 1,079 labeled as supernovae, 9,938 labeled as variable stars, 9,899 labeled as asteroids and 5,350 labeled as bogus. All the non bogus labels were combined into the single “real” label. Some images contained bad pixels, that were replaced by the median of the image where they were found.

3. Model

Our baseline classification model was taken from Ganin et al. 2016 and consists of a *feature extractor* component and a *predictor* component. The feature extractor component is composed of two 2-dimensional convolutional layers, each followed by a max-pooling layer. Batch normalization and ReLU activation functions are applied to both convolutional layers. The predictor component consists of three linear layers, each using batch normalization and ReLU activation functions, with the exception of the last layer which employs a Softmax function. As a benchmark, this model was trained with each dataset separately using a binary cross entropy loss function.

The MME model, similar to the baseline model, comprises a feature extractor and a predictor component. The feature extractor component shares the same architecture as the base model. The predictor component includes a L2 normalization layer, succeeded by a linear layer scaled by a temperature hyperparameter, and a Softmax activation function. Each class is represented in the feature space as an estimated vector *prototype*. This model is trained using two datasets: a fully labeled *source* dataset and a partially labeled *target* dataset. It is worth mentioning that while MME allows for unsupervised learning using the target dataset, our work focuses on the semi-supervised approach. During training, the model parameters are optimized using a two-term loss function:

$$\mathcal{L} = H(y_s, \hat{y}_s) + \lambda H(\hat{y}_u), \tag{1}$$

where $H(y_s, \hat{y}_s)$ represents the cross-entropy loss between the true (y_s) and predicted ($\hat{y}_s$) labels for the labeled (supervised) dataset, and $H(\hat{y}_u)$ denotes the entropy of the predicted labels for the unlabeled (unsupervised) dataset. The weight λ controls the balance between the two terms in the loss function. A gradient reversal layer (Ganin et al., 2016) is inserted between the feature extractor and predictor components of the model. This layer flips the sign of the gradient value during backpropagation, but only the unlabeled data passes through this layer. Consequently, the entropy term is minimized for the unlabeled target examples, encouraging the model to learn discriminative features that cluster around the estimated prototypes, while its maximization encourages feature representations that are invariant to domain shifts. This mechanism helps the model effectively adapt to the target dataset, leveraging information from both the labeled and unlabeled data sources.

4. Experiments

Three sets of experiments were defined: a *baseline* training, a *fine tuning* training and a *domain adaptation* train-

[1] https://portal.nersc.gov/project/dessn/autoscan/

Source / Target	HiTS	DES	ATLAS	ZTF
HiTS	**0.983 ± 0.004**	0.811 ± 0.026	0.626 ± 0.019	0.548 ± 0.015
DES	0.945 ± 0.011	**0.955 ± 0.003**	0.703 ± 0.006	0.606 ± 0.011
ATLAS	0.777 ± 0.031	0.697 ± 0.058	**0.967 ± 0.008**	0.502 ± 0.036
ZTF	0.765 ± 0.023	0.752 ± 0.022	0.633 ± 0.019	**0.945 ± 0.007**

Table 1. Baseline results. Each row/column corresponds to the mean and standard deviation of the balanced accuracy for that source/target experiment, calculated across 10 different random splits.

ing. A single round of these experiments use the same training/validation/testing set partitioning. To compare the models performance, the balanced accuracy metric (average recall) was used. In order to avoid the overuse of target labels, 10 labeled items per class were used for validation for the fine-tuning and MME experiments. To address imbalanced data, oversampling was applied to prevent bias towards the overrepresented class.

The baseline model was trained independently for each separate source dataset. We then used these models to evaluate their performances on all four datasets. The fine tuning training, consists of taking a baseline model and further training it using a smaller labeled set from the other three target datasets that were not trained on. We perform fine tuning using 1, 5, 10, 20 and 40 labeled items per class. We evaluate the performance of the fine-tuned models in both the source and target datasets in order to evaluate their domain adaptation capacities.

For DA via MME, we used the full labeled source training set, a small number of labeled items per class sampled from the target dataset, and an unlabeled dataset comprising the remaining target objects. The feature extractor weights are initialized with those of the corresponding baseline model following the approach of (Saito et al., 2019). Different instances of MME are run for each source-target scenario, varying the amount of target labeled data (1, 5, 10, 20, and 40 items per class, the same as the fine-tuning experiment). To assess the domain adaptation capabilities of MME, we evaluate its performance on both the source and target data. The optimal value for the hyperparameter λ in Eq. 1 is determined by training MME with multiple values (0.01, 0.02, 0.03, 0.05, 0.1, 0.5, and 10) and selecting the one yielding the highest balanced accuracy on the target validation set. All aforementioned experiment instances were repeated in a 10-fold manner, with each iteration employing a distinct random data partitioning.

5. Results

We start by evaluating the transferring learning capability of the baseline models when trained on each dataset separately by calculating their performances when applied on stamps from all surveys. Table 1 shows the balanced accuracy for these experiments. All models achieve an accuracy over 94% over the source dataset. Our results are consistent with the literature for DES (∼96%, Acero-Cuellar et al., 2022) and ATLAS (∼95.2%, Rabeendran & Denneau, 2021) while for the other datasets we achieve a slightly lower accuracy than the state-of-the-art (∼99.5% for HiTS, Cabrera-Vives et al. 2017; ∼98% for ZTF, Duev et al. 2019). We attribute this decrease in performance to the capacity of our model, yet we deem it of lesser significance in light of the primary focus of this paper, which is the evaluation of domain adaptation techniques. We notice that the greater transferring capacity of models is achieved by DES→HiTS (DES as source, HiTS as target) and HiTS→DES, which is to be expected given that both datasets were obtained using the Dark Energy Camera and both aimed at searching for supernovae. The rest of the experiments show worse performances, posing the need of transfering these models to the target datasets.

We compare the capacity of transferring the learned representations to other datasets by using fine tuning and MME in Figure 1 in terms of the number of labeled target objects presented to the model (shots). Each plot represents a source/target scenario. We show the accuracy of fine tuning and MME in terms of the shots on the source data (left plot of each panel) and on the target data (right plot of each panel). As noticed previously when discussing Table 1, when transferring DES→HiTS, fine tuning and MME are able to achieve competitive performances with few shots both for source and target stamps. In the target data, both fine tuning, and MME are able to surpass the baseline after only one or five shots of labeled objects from the target. Fine tuning and MME achieve comparable results on the target for most experiments, the exception being ATLAS→ZTF, ZTF→HiTS, and ZTF→ATLAS where fine tuning outperforms MME. However, it is worth noting that fine-tuning the model on the target data consistently leads to a deterioration in performance for the source dataset, regardless of the number of shots employed. This is of particular importance when aiming at developing generalist models for multi-stream classification.

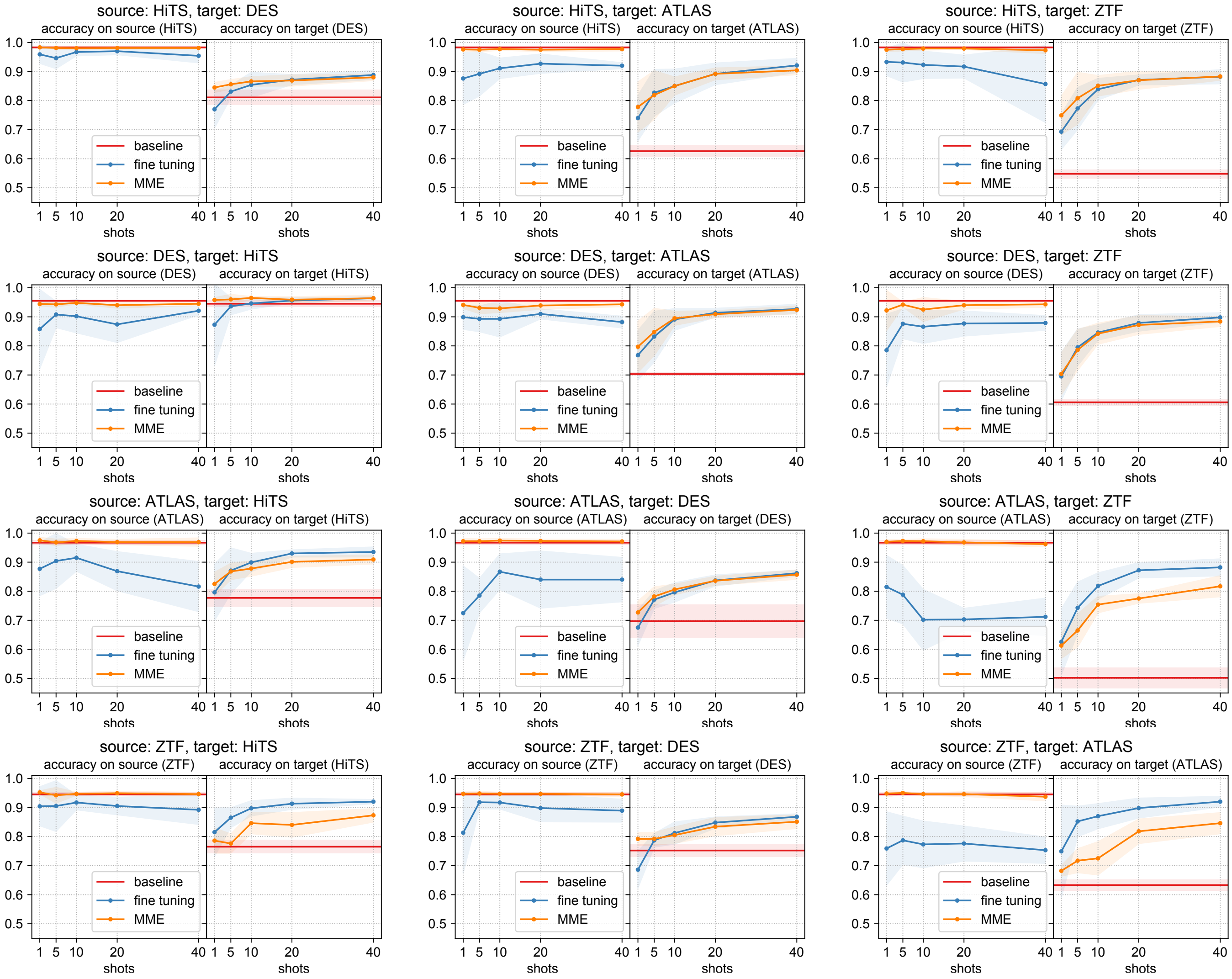

Figure 1. Balanced accuracy for different source/target scenarios for HiTS, DES, ATLAS, and ZTF. Each pair of plots represent one source→target experiment. The left plot shows the performance of fine tuning (blue) and MME (orange) on the source dataset as the number of shots increase, while the right plot shows their performance on the target dataset. The baseline result from Table 1 are shown as a horizontal red line.

6. Conclusions

We evaluate the transferability of convolutional neural networks (CNN) trained to classify image stamps of astronomical alerts on a source dataset to target data coming from various surveys. Using real/bogus stamps coming from HiTS, DES, ATLAS, and ZTF, we show that even though the CNN models are able to achieve over ∼94% balanced accuracy on the source dataset, they struggle to achieve competitive performances on stamps coming from surveys with a slightly different distribution. To address this, we examine two transfer learning techniques for the task: fine-tuning and domain adaptation via Minimax Entropy (MME). We show that both methods exhibit rapid learning capabilities from the target data with only a few labeled shots. However, MME maintains a high level of accuracy on the source domain, whereas fine tuning leads to a degradation of results on such domain. This is of special importance when considering the training of generalist models capable of performing inference in a multi-stream scenario, especially in the context of the first-light of upcoming instruments like the Vera C. Rubin Observatory.

Acknowledgements

The authors acknowledge support from the National Agency for Research and Development (ANID) grants: Millennium Science Initiative Program – ICN12 009 (GCV, FF, AMMA, MPC); FONDECYT regular 1231877 (GCV), 1200710 (FF); PIA/BASAL projects FB210005 (FF, AMMA), AFB-170001 (FF, AMMA), and FB210003 (FF).

References

Abbott, T., Abdalla, F., Allam, S., Amara, A., Annis, J., Asorey, J., Avila, S., Ballester, O., Banerji, M., Barkhouse, W., et al. The dark energy survey: Data release 1. *The Astrophysical Journal Supplement Series*, 239(2):18, 2018.

Acero-Cuellar, T., Bianco, F., Dobler, G., Sako, M., and Qu, H. There's no difference: Convolutional neural networks for transient detection without template subtraction. *arXiv preprint arXiv:2203.07390*, 2022.

Bellm, E. C., Kulkarni, S. R., Graham, M. J., Dekany, R., Smith, R. M., Riddle, R., Masci, F. J., Helou, G., Prince, T. A., Adams, S. M., et al. The zwicky transient facility: system overview, performance, and first results. *Publications of the Astronomical Society of the Pacific*, 131(995): 018002, 2018.

Cabrera-Vives, G., Reyes, I., Förster, F., Estévez, P. A., and Maureira, J.-C. Supernovae detection by using convolutional neural networks. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pp. 251–258. IEEE, 2016.

Cabrera-Vives, G., Reyes, I., Förster, F., Estévez, P. A., and Maureira, J.-C. Deep-hits: Rotation invariant convolutional neural network for transient detection. *The Astrophysical Journal*, 836(1):97, 2017.

Carrasco-Davis, R., Reyes, E., Valenzuela, C., Förster, F., Estévez, P. A., Pignata, G., Bauer, F. E., Reyes, I., Sánchez-Sáez, P., Cabrera-Vives, G., et al. Alert classification for the alerce broker system: The real-time stamp classifier. *The Astronomical Journal*, 162(6):231, 2021.

Dekany, R., Smith, R. M., Riddle, R., Feeney, M., Porter, M., Hale, D., Zolkower, J., Belicki, J., Kaye, S., Henning, J., et al. The zwicky transient facility: Observing system. *Publications of the Astronomical Society of the Pacific*, 132(1009):038001, 2020.

Duev, D. A., Mahabal, A., Masci, F. J., Graham, M. J., Rusholme, B., Walters, R., Karmarkar, I., Frederick, S., Kasliwal, M. M., Rebbapragada, U., et al. Real-bogus classification for the zwicky transient facility using deep learning. *Monthly Notices of the Royal Astronomical Society*, 489(3):3582–3590, 2019.

Flaugher, B., Diehl, H., Honscheid, K., Abbott, T., Alvarez, O., Angstadt, R., Annis, J., Antonik, M., Ballester, O., Beaufore, L., et al. The dark energy camera. *The Astronomical Journal*, 150(5):150, 2015.

Förster, F., Maureira, J. C., San Martín, J., Hamuy, M., Martínez, J., Huijse, P., Cabrera, G., Galbany, L., De Jaeger, T., González-Gaitán, S., et al. The high cadence transient survey (hits). i. survey design and supernova shock breakout constraints. *The Astrophysical Journal*, 832(2):155, 2016.

Förster, F., Cabrera-Vives, G., Castillo-Navarrete, E., Estévez, P., Sánchez-Sáez, P., Arredondo, J., Bauer, F., Carrasco-Davis, R., Catelan, M., Elorrieta, F., et al. The automatic learning for the rapid classification of events (alerce) alert broker. *The Astronomical Journal*, 161(5): 242, 2021.

Ganin, Y., Ustinova, E., Ajakan, H., Germain, P., Larochelle, H., Laviolette, F., Marchand, M., and Lempitsky, V. Domain-adversarial training of neural networks. *J. Mach. Learn. Res.*, 17(1):2096–2030, January 2016. ISSN 1532-4435.

Goldstein, D., D'Andrea, C., Fischer, J., Foley, R., Gupta, R., Kessler, R., Kim, A., Nichol, R., Nugent, P., Papadopoulos, A., et al. Automated transient identification in the dark energy survey. *The Astronomical Journal*, 150 (3):82, 2015.

Möller, A., Peloton, J., Ishida, E. E., Arnault, C., Bachelet, E., Blaineau, T., Boutigny, D., Chauhan, A., Gangler, E., Hernandez, F., et al. Fink, a new generation of broker for the lsst community. *Monthly Notices of the Royal Astronomical Society*, 501(3):3272–3288, 2021.

Narayan, G., Zaidi, T., Soraisam, M. D., Wang, Z., Lochner, M., Matheson, T., Saha, A., Yang, S., Zhao, Z., Kececioglu, J., et al. Machine-learning-based brokers for real-time classification of the lsst alert stream. *The Astrophysical Journal Supplement Series*, 236(1):9, 2018.

Nordin, J., Brinnel, V., Van Santen, J., Bulla, M., Feindt, U., Franckowiak, A., Fremling, C., Gal-Yam, A., Giomi, M., Kowalski, M., et al. Transient processing and analysis using ampel: alert management, photometry, and evaluation of light curves. *Astronomy & Astrophysics*, 631: A147, 2019.

Rabeendran, A. C. and Denneau, L. A two-stage deep learning detection classifier for the atlas asteroid survey. *Publications of the Astronomical Society of the Pacific*, 133(1021):034501, 2021.

Reyes, E., Estévez, P. A., Reyes, I., Cabrera-Vives, G., Huijse, P., Carrasco, R., and Forster, F. Enhanced rotational invariant convolutional neural network for supernovae detection. In *2018 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. IEEE, 2018.

Saito, K., Kim, D., Sclaroff, S., Darrell, T., and Saenko, K. Semi-supervised domain adaptation via minimax entropy. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 8050–8058, 2019.

Smith, K. Lasair: the transient alert broker for lsst: Uk. *The Extragalactic Explosive Universe: the New Era of Transient Surveys and Data-Driven Discovery*, pp. 51, 2019.

Tonry, J., Denneau, L., Heinze, A., Stalder, B., Smith, K., Smartt, S., Stubbs, C., Weiland, H., and Rest, A. Atlas: a high-cadence all-sky survey system. *Publications of the Astronomical Society of the Pacific*, 130(988):064505, 2018.

Turpin, D., Ganet, M., Antier, S., Bertin, E., Xin, L., Leroy, N., Wu, C., Xu, Y., Han, X., Cai, H., et al. Vetting the optical transient candidates detected by the gwac network using convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 497(3):2641–2650, 2020.

Yin, K., Jia, J., Gao, X., Sun, T., and Zhou, Z. Supernovae detection with fully convolutional one-stage framework. *Sensors*, 21(5):1926, 2021.