Vetting the optical transient candidates detected by the GWAC network using convolutional neural networks

Damien Turpin,1,2⋆ M. Ganet,3,4 S. Antier,5,6 E. Bertin,7 L. P. Xin,1 N. Leroy,5 C. Wu,1 Y. Xu,1 X. H. Han,1 H. B. Cai,1 H. L. Li,1 X. M. Lu1 and J. Y. Wei1,8

1CAS Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100101, China
2AIM, CEA, CNRS, Université Paris-Saclay, Université Paris Diderot, Sorbonne Paris Cité, F-91191 Gif-sur-Yvette, France
3ENSAE Paris, F-91120 Palaiseau, France
4HEC Paris, F-78350 Jouy-en-Josas, France
5LAL, Univ. Paris-Sud, CNRS/IN2P3, Orsay, France
6APC, Univ. Paris Diderot, CNRS/IN2P3, CEA/bfu, Obs de Paris, Sorbonne Paris Cité, France
7UPMC-CNRS, UMR7095, Institut d’Astrophysique de Paris, 75014 Paris, France
8School of Astronomy and Space Science, University of Chinese Academy of Sciences, 101408 Beijing, China

Accepted XXX. Received YYY; in original form ZZZ

ABSTRACT

The observation of the transient sky through a multitude of astrophysical messengers has led to several scientific breakthroughs these last two decades thanks to the fast evolution of the observational techniques and strategies employed by the astronomers. Now, it requires to be able to coordinate multi-wavelength and multi-messenger follow-up campaign with instruments both in space and on ground jointly capable of scanning a large fraction of the sky with a high imaging cadency and duty cycle. In the optical domain, the key challenge of the wide field of view telescopes covering tens to hundreds of square degrees is to deal with the detection, the identification and the classification of hundreds to thousands of optical transient (OT) candidates every night in a reasonable amount of time. In the last decade, new automated tools based on machine learning approaches have been developed to perform those tasks with a low computing time and a high classification efficiency. In this paper, we present an efficient classification method using Convolutional Neural Networks (CNN) to discard any bogus falsely detected in astrophysical images in the optical domain. We designed this tool to improve the performances of the OT detection pipeline of the Ground Wide field Angle Cameras (GWAC) telescopes, a network of robotic telescopes aiming at monitoring the optical transient sky down to R=16 with a 15 seconds imaging cadency. We applied our trained CNN classifier on a sample of 1472 GWAC OT candidates detected by the real-time detection pipeline. It yields a good classification performance with 94% of well classified event and a false positive rate of 4%.

Key words: methods: data analysis; surveys - supernovae: general; Astrophysics - Instrumentation and Methods for Astrophysics

1 INTRODUCTION

The time domain astronomy aim at studying transient phenomena having a wide variety of flux and time scales and detected with a very broad range of localization accuracies in the sky depending on the astrophysical messengers emitted (electromagnetic, gravitational waves and high-energy particles). For several centuries, the main observed transient phenomena were the supernovae (SNe) in the optical domain, tracing the violent fate of the most massive stars undergoing a core collapse or the thermonuclear explosion of white dwarfs accreting the matter of a companion star (see for example Gal-Yam 2017, about the SNe classification). In the last century, the SNe were detected only at a rate of few per year1, mainly because the observational techniques and strategies were not optimized to frequently

⋆ E-mail: damien.turpin@cea.fr

1 See for example http://www.rochesterastronomy.org/snimages/snactive.html

© 2020 The Authors
detect such rare events\(^2\). Therefore, the workload pressure on the detection pipelines and classification procedures of those transients were easily manageable by involving human actions in several steps, especially knowing that SNe can be observed during several days to months after the initial explosion with a 1-meter class telescope.

A first major revolution in the transient sky astronomy came with the development of the high-energy x-ray and gamma-ray telescopes and the detection of new classes of transients such as the Gamma-ray bursts (GRB)\(^3\) or the flaring blazars (Brown et al. 1986; Robson et al. 1988; Hartman et al. 1992). In addition to the high-energy emission, those transients also produce low energy broadband emission up to the radio wavelengths. Hence, multi-wavelength follow-up observations across the whole electromagnetic spectrum became crucial to get a global picture of the physical processes. The GRBs certainly represent one of the most extreme observational challenge for the follow-up telescopes as the short-living initial gamma-ray signal (see the review on GRB physics by Kumar & Zhang 2015) can be very poorly localized within up to several tens of square degrees depending on the trigger instrument. Then, a race against time is engaged to catch the so-called multi-wavelength afterglow emission that is fading very quickly so that it usually becomes unreachable for a detection 1-2 days after the trigger time by any x-ray or optical facility. This kind of transient event has definitely led to the birth of a new type of astronomy where different type of electromagnetic facilities have to work together in near real-time to complete the scientific data sets. Two decades ago, in the optical domain, several groups started to develop networks of small aperture robotic telescopes (for example ROTSE, TAROT, BOOTES, MASTER) that were\(^4\) capable to respond to any alert and scan a large fraction of the night sky continuously with a high cadence (Marshall et al. 1997; Akarsu et al. 2016). The multiplication of the synergies between the space and ground-based telescopes, all broadcasting alerts about a large variety of transient sources, has largely contributed to increase the flow of data to be analyzed on real-time (photometry, spectroscopy and polarimetry).

Currently, the increasing pressure on the data processing of the follow-up telescopes studying the transient sky is significantly accelerating with the recent birth of the multi-messenger (MM) astronomy adding the high energy-neutrinos (HEN) and gravitational wave (GW) events in the optical alert broadcasting system. With the constant sensitivity improvements of the electromagnetic and the MM facilities, one can now regularly deal with the reception of several valuable transient alerts of any astrophysical type every night. In the next decade, the multiplication of the facilities dedicated to the study of the transient sky and being able to make an all-sky monitoring at even deeper sensitivities will continue to progress, e.g. the Large Synoptic Survey Telescope (LSST)\(^5\) (Ivezic et al. 2008), the Square Kilometer Array (SKA) (Taylor 2000), KM3NeT (Adrian-Martinez et al. 2016), SVOM (Wei et al. 2016) or the next generation of GW detectors LIGO/Virgo and Kagra (Abbott et al. 2018). Those projects will definitely make the time-domain astronomy enter into the big data era. As an example, the LSST project (Ivezic et al. 2019) would produce 20 terabytes of data every night with the possibility of having several hundreds of thousands alerts per night starting from 2021 and running over ten years of operation. It should extend the known SNe catalog with more than three billions of new entries (more than two orders of magnitude in terms of detection rate compared to any current survey).

In the optical domain several groups already developed synoptic surveys, like the Catalina Real-time Transient Survey (CRTS; Drake et al. 2009), PTF (Law et al. 2009), ASASSN\(^4\), PanSTARRS (Chambers et al. 2016), ATLAS (Tonry et al. 2018), ZTF (Bellm et al. 2019), DES (Goldstein et al. 2015) or Gaia (Gaia Collaboration et al. 2018), that explore the transient sky in addition to their participation to the various multi-messenger follow-up campaigns (see for example Abbott et al. 2017). The data flow generated by those surveys are already no longer manageable in a reasonable amount of time by the standard techniques previously used for narrow field of view telescopes as shown for example for ZTF (Mahabal et al. 2019). The standard transient detection pipelines were usually based on PSF-matching and the catalog cross-matching methods for the detection of new sources, followed by a human validation of each transient candidate for the classification task. The growing alert rates and data flows now force the astronomers to develop new observational strategies and techniques to quickly detect, identify and classify the numerous uncatalogued sources they catch every night in their extensive searches.

New techniques using machine learning algorithm are developed to perform robust automated classifications of hundreds up to thousands of sources every night in real-time. The classification task is usually split into two steps independently performed. First, the goal is to filter out the bogus sources from the real uncatalogued sources of interest (e.g. Masci et al. 2017; Sánchez et al. 2019; Mahabal et al. 2019; Jia et al. 2019) immediately after the detection. The second step goes deeper in the classification procedure by associating an astrophysical category to an identified transient based on its temporal and/or spectral properties (e.g. Morii et al. 2016; Narayan et al. 2018; Muthukrishna et al. 2019). Among the zoo of machine learning algorithms, convolutional neural networks (CNN) are now massively used for such tasks as they are well-adapted to ingest data containing multiple arrays like images (Bishop 2006; Lecun et al. 2015). They employ multiple interconnected layers, similar to a neuronal network, to efficiently identify patterns in images and are therefore particularly suitable for the time-domain astronomy (Gieseke et al. 2017).

In this paper, we investigate the possibility of using CNN for the vetting of the optical transient (OT) candidates that will be detected by the Ground Wide field Angle Cameras network (GWAC). The GWAC system is a synoptic optical survey which is currently able to instantaneously cover 2000 square degrees on the sky with a high imaging cadency of

\(^2\) The observed local (within 100 Mpc) supernovae rate is about \(10^{-4}\) SNe - yr\(^{-1}\) - Mpc\(^{-3}\) (Horishita et al. 2011).

\(^3\) Most of them are still in operation.

\(^4\) http://www.astronomy.ohio-state.edu/~assassin/index.shtml
one frame every 15 seconds. In operation since 2017, GWAC is a part of the ground-based follow-up system of the SVOM mission (Wei et al. 2016), the next generation of space mission dedicated to the study of the multi-wavelength transient sky. It already provides a large data flow that must be smoothly digested by the real-time data processing pipeline as well as a significant amount of OT candidates sometimes well identified as real transients such as dwarf novae outbursts recently discovered in the GWAC survey (Wang et al. 2019). The GWAC network is a perfect example of the evolution of the optical facilities that emerge nowadays to study the transient phenomena. It brings new observational challenges which have to be solved in order to exploit the full capabilities of the instruments. In the section 2, we will describe the GWAC system and the transient detection pipeline which is currently running. Then, in the section 3, we will introduce the deep machine learning classifier we set up for the vetting of the GWAC OT. The classification results and performances will be presented in the section 4 and we finally draw our conclusions and perspectives for this work in the section 5.

2 THE GWAC TELESCOPES

2.1 Instrumentation setup

Since the end of 2017, the Ground Wide field Angle Cameras telescopes are under development in China at the Xinglong Observatory. Each GWAC telescope mount is equipped with five cameras: four JFoV cameras (4k CCD E2V camera with an aperture of 180 mm) and one FFoV camera (3k CCD camera with an aperture of 35 mm) used to monitor the sky seeing and brightness conditions, see Figure 1. The main scientific instruments, the JFoV cameras, cover a field of view of about 12.4° × 12.4° per camera (~150 square degrees per camera). Taking into account the overlaps between the fields of view of the 4 JFoV cameras, a GWAC mount finally covers 500 square degrees on the sky. Each JFoV camera is designed to reach an unfiltered limiting magnitude of about 16 in a dark night for 10 seconds of exposure. A stacking analysis of the single frames can be performed on real-time to reach a maximum limiting magnitude of R~18 in clear and dark night as shown in (Turpin et al. 2020).

2.2 The GWAC optical transient detection pipeline

The search for OT in GWAC data is made through several steps from the detection of candidates to their identification as being real variable/transient sources. The raw images are first pre-processed camera per camera to correct them from the Dark and the Bias offsets and to make the WCS (World Coordinates System) calibration. Those calibrated images are then automatically and independently analyzed by two pipelines to search for OT candidates. These two pipelines make use of standard methods comparing the scientific images with reference images taken much earlier such as the catalog cross-matching and the differential image analysis (DIA). Concerning the GWAC system more details can be found in (Turpin et al. 2020; Wang et al. 2019) but typically a new source is detected once it fulfills the following criteria:

(i) The source has a signal-to-noise (SNR) ratio ≥ 5 and is not detected down a SNR = 5 in the reference images
(ii) The source is detected in several successive images
(iii) The point spread function (PSF) of the source shall be stellar-like profile, i.e. a 2D gaussian profile.
(iv) no any CCD defect is detected in a region of 6 pixels around the source.

The uncatalogued sources extracted from those analysis form the preliminary OT candidate list named OT1 candidates. Then, several filters are applied on the source candidate parameters (the Full Width at Half Maximum -FWHM-, the SNR, the optical peak flux, the source position, etc.) on at least 5 successive images. Practically speaking, these filters aim to clean the OT1 candidates from most of the spurious sources like the hot pixels or cosmic ray tracks. If at least 2/5 images pass the selection criteria, the OT candidates is kept otherwise it is rejected. A catalog cross matching filter using deeper catalogs is then applied to the OT1 candidates that passed the first selection criteria, see Figure 2. Catalogs such as Gaia DR2 (Gaia Collaboration et al. 2018), PanStars DR1 (Chambers et al. 2016), 2MASS (Skrutskie et al. 2006), Galex DR5 (Bianchi et al. 2011) or public databases on solar system objects such as the Minor Planet Center are used to perform this task.

After passing all of those filters, the remaining candi-
Figure 2. A schematic view of the current GWAC detection pipeline setup to detect and identify the optical transient sources in both single and stacked images.

...candidates. Sub-images are then cropped from each initial 4k × 4k JFoV images and subtracted from the sky background contribution to make 100 × 100 pixel-sized finding charts centered at the positions of each selected OT2 candidate. These finding charts are then checked one by one by a human eyed-check analysis. Simultaneously, two 60 cm robotic telescopes (GWAC-F60A and GWAC-F60B) located beside the GWAC telescopes at the Xinglong Observatory automatically perform follow-up observations of any source found by the GWAC system in order to help the GWAC scientist on duty to finally confirm the genuineness of a given OT2 candidate. Once the OT candidates are confirmed as being real transient sources, additional follow-up observations can be triggered with larger telescopes and public alerts can be released. This kind of detection pipeline is commonly used in the time domain astronomy. However, while it is robust enough for telescopes with a very limited field of view (typically few tens of arcminute), it turns to be no longer the optimal solution for telescopes covering hundreds of square degrees in the sky like the GWAC system as explained in the next section.

2.3 Data flow and false detection rate

The GWAC telescope network is operated in a sky survey mode following a pre-defined sky grid pointing strategy searching for bright optical transient events with a minimum of sub-minute time scale. Since the beginning of 2019, four GWAC mounts (16 JFoV cameras) are operational but at completion, the full system will be composed of 10 mounts. This setup implies the collection of a huge amount of data every night with typically between 6000-8000 images taken each night for a single telescope mount. When the observational conditions are optimum, the current network can generate as a whole as many as 24 000 images per night (up to 80k images per night for the complete network). When using the detection pipeline described above and in Figure 2, the difficulties encountered with the GWAC telescopes system mainly come from the data flow and subsequently, the large false detection rate it can produce. The data flow is generated by the image cadency (the exposure time) and the number of operated cameras. The false detection rate is partly due to the data flow itself but it is also strongly dependent on the optical sensitivity of the instruments, their field of views and the strictness of the transient selection criteria. In addition, the large field of view of the single GWAC cameras (150 sq·deg) combined with the limited size of the CCD detectors produces a large pixel scale of 11.7 arcsec·pix⁻¹ and image distortion effects (while corrected in our images). These two factors make the use of the catalog cross matching and the differential image analysis even more complicated. This usually results in the production of additional fake detections populating the OT1 candidates category. While the standard filtering algorithms are able to clean many fake OT1 candidates, there is still a large fraction of them that pass through the filters. Typically, for one GWAC telescope mount, the number of OT1 candidates can be as numerous as several hundreds in a single night depending on the observational conditions. Our standard filtering algorithms then reduces this number to several tens up to a few hundreds. Those ones then must be manually vetted both by humans and further follow-up observations. Multiplying this task to the number of GWAC mounts and one can easily understand that this “true or false” classification task becomes no longer manageable both by the GWAC-F60 follow-up telescopes and the GWAC scientists in a reasonable amount of time. Therefore, our GWAC-F60 telescopes can be rapidly unable to ingest the quantity of triggers and additionally they can no longer smoothly follow their own observation plans independently of the GWAC...
camera activities. Moreover, the increase in our duty scientists workload finally make them no longer being able to focus their efforts on the most promising events. The identification and classification processes of a genuine transient source then undergo a long delay which is not compatible with the scientific purposes of the GWAC system that aim to quickly identify short-lived optical transient sources.

3 A DEEP LEARNING CLASSIFIER

Our goal is to improve the current detection pipeline of the GWAC system, especially in easing the OT1 candidates classification and making the human decision-taking process more responsible. As shown previously, there is a crucial need for a classification that distinguishes the astrophysical sources from the GWAC alert stream prior to build the OT2 candidates list. Before going deeper into the details, we start to define few acronyms that we will use all along the paper:

- ROS: real optical sources in an image.
- FOS: fake optical sources in an image.
- ROT: real optical transients. A ROT is actually a ROS present in a series of images and showing a significant flux variation.
- FOT: fake optical transients.
- TP: true positives, i.e. the OT candidates well classified as ROT or ROS.
- TN: true negatives, i.e. the OT candidates well classified as FOT or FOS.
- FP: false positives, i.e. the OT candidates classified as ROT or ROS while there are actually FOT or FOS.
- FN: false negatives, i.e. the OT candidates classified as FOT or FOS while there are actually ROT or ROS.

One immediately understands that our classifier must minimize the number of FP and FN to limit the contamination of the OT2 candidates sample by any bogus in one hand and to avoid too many losses of ROT because of misclassifications in the other hand. The final goal is to obtain a classification accuracy greater than 90% with a FN classification not as great as 2%. Indeed, we prefer to keep more false positives (FP) instead of losing too many transients falsely classified as bogus (FN) in the classification process. To perform this task, we used a Convolutional Neural Networks algorithm. This choice is firstly motivated by the fact that CNNs are very well adapted for pattern recognition in images (Bishop 2006; Lecun et al. 2015) and have been already robustly tested with success for many different purposes in astronomy (Bloom et al. 2012; de la Calleja & Fuentes 2004; du Buisson et al. 2015; Mahabal et al. 2019). Secondly, the CNNs have demonstrated excellent classification performances compared to other standard and deep machine learning methods with a minimum of implementation (Gieseke et al. 2017).

3.1 The CNN model architecture and implementation

While this kind of “true or False” classification game does not require in principle a very deep and complex network structure, a too basic network may also have limited performances even considering such a “simple” task as noticed by Gieseke et al. (2017). We therefore built a CNN code using an architecture composed of two convolutional layers, two pooling layers, one ReLu and one softmax hidden layers, see the details in Table 2. The pooling layers were kept to 2 × 2 bin size due to the small size of some objects projected in the large GWAC pixel scale. The cross-entropy function was used as a loss function to give a high weight for very confident false positives, which we strongly want to avoid. The CNN was implemented in Python v3.6, using the Keras6 package with TensorFlow2. The Keras package has the advantages to provide built-in diagnostic tools and a compact code writing which allow for a relative ease of use. A Keras Adam optimizer was used with a low learning rate (lr=0.0001) after witnessing disappointing convergence properties.

As an input, our CNN algorithm uses background-subtracted finding charts (100 × 100 pixels) of the OT1 candidates. We then select only the central part of those images (35 × 35 pixels) for the classification. This choice is motivated to have a high learning rate as the CNN requires to be trained on an extensive amount of data (typically of the order of a minimum of 105 images) while keeping enough information (background and a minimum number of sources) in the sub-images for the pattern recognition. Before being able to give any classification on our OT candidates, the CNN must be trained to recognize patterns in our images. When a CNN layer receives an input, an output is then produced to feed the next layer. As long as the inputs are transformed into outputs, a series of several weights is produced to finally converge and build a final probabilistic rank between 0 and 1. The training phase contains several epochs of test to make the final convergence. The CNN ranking is then compared to the image labeling previously made by our expert scientists which consists in giving either a mark “1” to sub-images containing a ROT or “0” if they contain FOT. Therefore, this comparison method gives an idea on the level of agreement or disagreement of the CNN decision with the human classification. If a disagreement is frequently observed, it means either the CNN architecture is not optimised for our classification purpose or the human labeling is not correct. In such case, the CNN architecture and/or the labeling has to be revised until a good agreement is found.

Table 2. The CNN structure used in this work.

| Layers  | Sizes  | Characteristics     |
|---------|--------|---------------------|
| convolution | 32 × 32 | 3 × 3 kernels relu  |
| max pooling | 32 × 32 | 2 × 2               |
| convolution | 32 × 32 | 3 × 3 kernels relu  |
| max pooling | 32 × 32 | 2 × 2               |
| dense    | 15     | activation:relu     |
| dense    | 2      | activation:Softmax  |

6 https://keras.io/. See also (Géron 2019) for a review of the usages of Keras.
7 https://www.tensorflow.org/ https://github.com/tensorflow/tensorflow
At the end of the training, we build a Keras python model of our CNN to be used later to classify any OT candidate detected by the GWAC real-time detection pipeline. Before using the CNN model in production, a final test of the classification performances is usually performed on an image sample that has never been used previously. If the CNN model reaches the classification requirements, i.e. if the number of misclassified sources is consistent with our scientific requirements, it can then be used to classify the genuineness of any GWAC optical transient candidates. On the contrary, if the classification is not good enough, a new training with a data set more representative of the GWAC OT candidate images must be performed until the classification requirements are fulfilled.

We illustrate, in figure 3, the implementation we set up for both the training and the validation of the CNN algorithm as well as how it should be inserted in the detection pipeline of GWAC during real-time data taking.

3.2 Training data set

The classification of the sources into different astrophysical categories can be challenging. Indeed, transient sources are rare events and one might not have collected enough images of transient sources for the training. Some techniques can be used for the augmentation of the training data set such as simulating images of transient sources with a physical or empirical model or adding rotated images of real transient sources which artificially produces a new background and source distribution compared to the initial images as suggested by Gieseke et al. (2017). Typically, several tens of thousands images are needed to obtain a well trained CNN model.

Classifying our detected transients into several astrophysical categories based on additional informations such as the spectral and flux time evolution is actually beyond the scope of this work. For our purpose, our bogus/real source classification tasks is independent of the nature of the transient as long as it is supposed to be a point-like source in the images. As a consequence, we avoid the problem of having too few images of real GWAC transients to train the CNN. Instead, we can directly extract point-like sources in GWAC images to build our sample of ROS images. Our five categories defined above. Our bogus simulator starts to validate each training epoch.

We reproduced this operation on 1000 different 4k × 4k images to obtain a final sample of 100 000 images of ROS. We show, in figure 4, a sub-sample of ROS images we used for the training of our CNN.

3.2.2 Details on the FOS image sample

While the ROS should all have a similar 2D gaussian profile (in the ideal case with negligible distortion effects) it is no longer the case for FOS. Indeed, a large variety of bogus can lead to false detections such as cosmic-ray tracks, hot pixels, bad pixels, crosstalks artefacts, dusts, irregularities in the sky background contribution, etc. Therefore, our FOS training data set must reproduce as close as possible such bogus shape distribution.

As it is actually very complicated to exactly mimic all the types of bogus we may encounter, we finally divided our bogus in several categories that are easily reproducible and correspond to the most frequent type of the bogus we encounter in GWAC images: hot pixels, background noise, bad column of pixels, dark pixels and a sky background with a significant light gradient. To reach the same statistics than the ROS training sample we had to use data augmentation techniques as we did not get enough images of all the categories of bogus. We simulated 100 000 images of bogus (50% of the full training data set) in equal proportions between our five categories defined above. Our bogus simulator starts with the same process than for extracting ROS from the 4k × 4k GWAC images. From the background-subtracted initial images, we extract 35 × 35 pixel sub-images and add a bogus in the central position similarly as we did for the ROS. Then comes the difference in the process, depending on the bogus to simulate we crop different parts of the 4k × 4k images according to the following criteria:

(i) For the hot pixel sub-images: no Sextractor sources should have a position (X,Y) in the sub-image consistent with a region of at least 6 pixels around the central pixel (X0,Y0): (X − X0)² + (Y − Y0)² ≥ 36

(ii) For the noisy sub-images: no Sextractor sources should be present in the sub-images only the background residual noise.

(iii) For the bad pixel columns sub-images: any sub-image randomly chosen is suitable.
GWAC OT candidate vetting using CNN

![Diagram of the classifier tool implementation](image)

**Figure 3.** Schematic view of the implementation of the classifier tool from the training of the CNN algorithm (left side) to the use of the CNN Keras model to make the vetting of the GWAC OT1 candidates on real-time. We used a large data set of N = 200 000 images to train the CNN.

**Figure 4.** Example of some sub-images (35x35 pixels) centered (± 2 pixel) at the position of point-like sources (ROTs) extracted from the GWAC 4k × 4k images. 100 000 images similar to these ones are produced to build our ROT training data set. Note that we extract point-like source with no prescription relative to their position in the original image (close to the edge or not, located in a dense star field, etc.).

(iv) For the dark pixels sub-images: any sub-image randomly chosen is suitable.

(v) For the non uniform sky background sub-images: we select only the position of brightest stars (estimated by Sextractor), even the saturated stars, that produce a light gradient in the surrounding pixels. The distance from the center of the image to the position of bright star centroid can span from $d \in [6; 15]$ pixels.

For the hot pixels, we actually choose to randomly put a single or a group of hot pixels (2 × 2 pixels at maximum) at the central position of the sub-images following the pixel intensities we observed from real data. We also add a random increment spanning in [-1;1] pixel to slightly shift the position of the bogus from the center. The bad column of pixels were simulated as an excess of light observed normalised to the pixel intensities we observed for this kind of bogus in real data. Also according to the real data the number of bad columns ranges from 1 to 3 either in the X or Y direction of the image. In figure 5, we compare typical bogus we simulated with the observed ones in the OT1 candidates finding charts.

4 ANALYSIS AND RESULTS

The analysis of the classification performance of our CNN is made in two steps: the training to build the keras model and the validation step of the classification procedure on a previously unseen image sample. For the training, the ROS images are labeled “1” while the FOS are labeled “0”. We
then compared this labeling with the CNN model probabilistic prediction spanning in the range $P_{\text{CNN}} \in [0; 1]$. Therefore, a source in a given image is considered as an FOS if $P_{\text{CNN}} < 0.5$ and as an ROS if $P_{\text{CNN}} \geq 0.5$. The mid value 0.5 represents a perfect random guess by the CNN model between the two categories.

![Figure 5](image_url)

**Figure 5.** The five categories of bogus simulated for the FOS training data set (a) compared with (b) the same kind of bogus we indeed observed in the GWAC OT1 finding charts (100 x 100 pixel-sized).

We trained our CNN algorithm on the 200 000 simulated images (50% ROS, 50% FOS) making 10 training epochs to build the final Keras model. Based on the $P_{\text{CNN}}$ criteria, we can build the normalized confusion matrix for a quick look of the classification results. The normalized confusion matrices allow to display the fraction of the well classified instances as $TN$ and $TP$ while in the white squares are shown the mis-classified instances as $FN$ (top right) and $FP$ (bottom left). For a perfect classifier, the blue squares would indicate "1" while the white squares would indicate "0".

![Figure 6](image_url)

**Figure 6.** The normalized confusion matrix produced after the training of the CNN algorithm on 200 000 simulated images of bogus and real sources. The numbers in each blue square indicate the fraction of the total instances correctly classified as $FP$ (top left) and $TP$ (bottom right) while in the white squares are shown the mis-classified instances as $TN$ (top right) and $FN$ (bottom left). For a perfect classifier, the blue squares would indicate "1" while the white squares would indicate "0".

### 4.1 The training

We trained our CNN algorithm on the 200 000 simulated images (50% ROS, 50% FOS) making 10 training epochs to build the final Keras model. Based on the $P_{\text{CNN}}$ criteria, we can build the normalized confusion matrix for a quick look of the classification results. The normalized confusion matrices allow to display the fraction of the well classified instances as $TN$, $TP$ and the fraction of the mis-classified ones in the $FN$ and $FP$ categories, as shown in figure 6. The normalized values of $TN$, $TP$, $FN$, $FP$ obey to the following rule:

$$\frac{TN + FN}{N_{\text{FOS}}} = 1, \quad N_{\text{ROS}} = 10^5 : \quad \frac{TP + FP}{N_{\text{ROS}}} = 1, \quad N_{\text{ROS}} = 10^5 \quad (1)$$

Based on the training data set, the normalized confusion matrix shows that the CNN algorithm has been well trained to recognize bogus and real sources with classification performances close the ideal case where the $TP$ and the $FP$ instances would be maximized up to a normalize value of "1" while the $FP$ and the $FN$ would have been minimize to "0". To better characterize the classification response of our CNN model, we also computed three diagnosis:

(i) The **receiver operating characteristic** (ROC) curves that display on graph the True Positive Rate (TPR) as function of the False Positive Rate.

(ii) The **Area Under the ROC Curve** (AUC) which corresponds to the integral of the ROC curve $\in [0; 1]$. "0" or "1" correspond to an ideal case where 100% of the instances are mis- or well classified.

(iii) The **Accuracy** coefficient $(AC) \in [0; 1]$. "0" or "1" correspond to an ideal case where 100% of the instances are mis- or well classified. :

$$AC = \frac{TN + TP}{TN + FP + FN + TP} \quad (2)$$

The ROC curve of the CNN model applied to the training data set also shows that we obtain a very high TPR (close to the maximum value "1") while keeping a extremely low FPR (close to "0"), see figure 7. This trend is a very convincing proof that a classifier is behaving well as it falsely classifies a very few number of detected events.
Finally, the corresponding AUC and AC are 0.99 and 0.986, respectively, and also point out a very good classification performance of our CNN model. All these diagnosis confirm that the architecture of our trained CNN model is well adapted to distinguish bogus form real sources. However, while it helps to confirm that the architecture of the CNN is robust enough to perform this kind of classification task, it does not guarantee at all that our CNN model will have the same performance on real GWAC images as our implementation has a limitation. Indeed, we could not simulate all the types of bogus we encounter in the real GWAC images as it would require a too large amount data for the simulation which translates in a higher computational cost and a severe worsening of the simulation complexity. Nevertheless, our FOS simulations and the architecture of our CNN are expected to be generic enough to deal with unseen bogus that may share the same properties than our simulated ones. As an example, we did not simulated any cosmic-ray track in our bogus sample but we simulated some groups of defective pixels that share common properties with those of cosmic-ray tracks (elongated shape with no PSF model or having a very sharp PSF model).

### 4.2 The validation

To finally validate the classification performance of our CNN, we confront it with a new sample of images representative of the zoo of bogus and real sources the GWAC pipeline generally detects. Each of those images have been previously labeled by our expert scientists following the same labeling rule described in 3.1. In addition to this labeling, each image has been manually classified into representative categories such as real moving objects, hot pixels, flaring stars, variable stars, bad pixels, dark pixels, incorrectly processed columns of pixels. This categorization fits the different groups of bogus used in the simulated training data set and are the most common bogus encountered in GWAC images. Our validation image sample is finally composed of 7841 images of 1472 objects detected by the GWAC transient search pipeline in 2017 and 2018. The detail of the object distribution into each source category is shown in the table 3.

As for the training sample, the probability given by the CNN on each image, $P_{CNN}$, is compared with the image labeling to make the diagnosis of the classification performance. When we applied our trained CNN model on those images, we finally found that the overall accuracy of the classifier is AC = 0.94 with a very low number of FN classification, around 2% of the total sample. Around 4% of the images containing a bogus are misclassified as ROS (FP) as shown in Figure 8. These classification performances are in good agreement with our scientific requirements mentioned in section 3 and hence, we consider that our generic deep learning classifier is robust enough to be automatized in the transient detection pipeline as a tool to vet the GWAC optical transient candidates. In table 4, we give more details on the classification performance for each source category used in the validation sample.

In addition, we also explored the capabilities of our generic CNN model in classifying bogus images that were not included in the simulation of the training sample. We added to our initial validation samples around 1700 images (a data augmentation of ~ 20%) of completely new bogus types such as dust obstructions, suspected ROS or low signal-to-noise ratio candidates and a large variety of bad pixels. While the addition of those new bogus make the classification accuracy dropped to AC = 0.91, we found that the performances are still good enough with respect to our scientific requirements. It reinforces the validation test and overall shows how powerful and generic are the CNN algorithms, even with relatively simple layer architectures, in distinguishing any type of bogus from real point-like sources.

**Table 3.** The different categories of the image sample used to validate the classification performances of our CNN.

| category            | $N_{\text{images}}$ | $N_{\text{object}}$ | typical label |
|---------------------|---------------------|---------------------|---------------|
| moving objects      | 878                 | 29                  | 1 (real)      |
| hot pixels          | 1798                | 862                 | 0 (bogus)     |
| flaring stars       | 102                 | 20                  | 1 (real)      |
| variable stars      | 3909                | 23                  | 1 (real)      |
| bad pixels          | 267                 | 95                  | 0 (bogus)     |
| dark pixels         | 333                 | 166                 | 0 (bogus)     |
| bad pixel columns   | 554                 | 277                 | 0 (bogus)     |
| **Total**           | 7841               | 1472                |               |

**Table 4.** The results of the CNN classification in the different categories of real/bogus sources tested during the validation step.

| category            | TP  | TN  | FP  | FN  | AC  |
|---------------------|-----|-----|-----|-----|-----|
| moving objects      | 0.96| 0.01| 0.02| <0.01| 0.97|
| hot pixels          | <0.01| 0.93| 0.06| <0.01| 0.94|
| flaring stars       | 0.82| 0.09| 0.09| 0.0  | 0.91|
| variable stars      | 0.93| <0.01| 0.06| <0.01| 0.94|
| bad pixels          | <0.01| 0.94| 0.05| 0.0  | 0.94|
| dark pixels         | 0.09| 0.83| 0.07| 0.01 | 0.91|
| bad pixel columns   | 0.02| 0.69| 0.29| <0.01| 0.71|

As for the training sample, the probability given by the CNN on each image, $P_{CNN}$, is compared with the image labeling to make the diagnosis of the classification performance. When we applied our trained CNN model on those images, we finally found that the overall accuracy of the classifier is AC = 0.94 with a very low number of FN classification, around 2% of the total sample. Around 4% of the images containing a bogus are misclassified as ROS (FP) as shown in Figure 8. These classification performances are in good agreement with our scientific requirements mentioned in section 3 and hence, we consider that our generic deep learning classifier is robust enough to be automatized in the transient detection pipeline as a tool to vet the GWAC optical transient candidates. In table 4, we give more details on the classification performance for each source category used in the validation sample.

In addition, we also explored the capabilities of our generic CNN model in classifying bogus images that were not included in the simulation of the training sample. We added to our initial validation samples around 1700 images (a data augmentation of ~ 20%) of completely new bogus types such as dust obstructions, suspected ROS or low signal-to-noise ratio candidates and a large variety of bad pixels. While the addition of those new bogus make the classification accuracy dropped to AC = 0.91, we found that the performances are still good enough with respect to our scientific requirements. It reinforces the validation test and overall shows how powerful and generic are the CNN algorithms, even with relatively simple layer architectures, in distinguishing any type of bogus from real point-like sources.
The confusion matrix and the ROC curve of our Figure 8.

The analysis presented above only considered the classification of the individual images of bogus and real sources. However, the vetting of the OT candidate must also include the time evolution of the source candidates, i.e. taking into account the image time series. Therefore, the rejection of bogus is made on the basis of the evolution of the CNN score across several images. The mean of the CNN probabilities accounts the image time series. Therefore, the rejection of bogus and real sources.

Table 5. The evolution of the false positives (FP) and false negatives (FN) as function of the rejection criterion $R \geq 0.99, 0.95, 0.90$.

| $R$         | $FP$ | $FN$ |
|-----------|------|------|
| $R \leq 5\sigma$ | 4.6  | 2.9  |
| $R \leq 3\sigma$ | 7.2  | 1.7  |
| $R \leq 2\sigma$ | 8.9  | 1.0  |
| $R \leq 1\sigma$ | 10.9 | 0.7  |

A candidate is finally classified as a bogus if it satisfies the rejection criterion otherwise it is classified as a real point-like source. We applied these criteria to the full validation sample of candidates (1861 candidates including the bogus not simulated in the training data set) and show, in Table 5, the evolution of FN, FP. The goal is to find the good trade-off in the rejection criterion in order to minimize both FP and FN. A too strict rejection may enhance too much the FN while keeping the FP very low, i.e. we miss some real events but do not get any fake. On the contrary, a too shallow rejection will go into the opposite direction, i.e. we would keep many bogus by ensuring to keep all the real events.

We find that a rejection criterion at $R=3\sigma$ confidence level is finally a good trade-off with less than 2% of ROT loss and about 7% of false positive detections (~ 91% of OT candidate well classified as bogus or real sources). We noticed that the FN candidates are actually sources having brightness very close to the detection threshold with a SNR $\leq 3$ which make them hard to be clearly identified by our CNN algorithm trained on securely detected OT.

These results translate into the following scenario. In a typical night, where a hundred of candidates would have been detected, only 7 bogus would have been eventually stored in the OT candidates list before passing through the series of filters described in section 2.2 and the human/GWAC-F60 telescope vetting step. We find that this number of false positives is now easily manageable on a whole night by the GWAC data processing pipeline, the scientist on duty and in addition it will significantly reduce the workload pressure on the GWAC-F60 telescopes scheduler.

5 CONCLUSION

The fast identification and classification of the transient sources are the major challenge to take up for the current and the near upcoming wide field angle facilities dedicated to the time-domain astronomy. In this paper, we have presented a method to distinguish real astrophysical sources from many types of bogus detected by the GWAC survey telescopes ($\text{FoV} = 25^\circ$, $R_{lim} = 16$ in 10s) based on a deep machine learning approach. The machine learning methods are usually easy-to-set-up, cost-effective, time-effective and bring a valuable automated classification procedure to any transient detection pipeline. The first “True or False” classification step is now unavoidable to obtain efficient transient search pipeline and quick human reaction to validate the optical transient candidates. To solve the problem of optical
transient vetting in the GWAC images, we used a convolutional neural network classifier trained on computationally-enhanced data relying on the GWAC database to generate images of real sources and bogus.

The CNN classifier proved to be very efficient in filtering out many types of bogus using a few amount of images for the decision. The final false positive alarm ratio is less than 5% when it is applied to the individual images. When applying the CNN classifier on the image time series of each OT candidate, we end up with about 7% of FP optical transient classifications at the level of the *OT1* candidate sample. The great advantage of our classifier is that it keeps the loss of real OT (FN) as low as 2% of the total transient candidate sample. This is a key parameter to maintain a high level of transient detection rate every night.

Including such classifier tool in the transient detection pipeline of GWAC will significantly lighten the workload pressure of the pipeline itself and the GWAC duty scientists. These performances are in well agreement with the scientific requirements of the GWAC system that aim at detecting and quickly identifying optical transient sources. Therefore, the output CNN score is a precious information for the scientists who will have to take important decisions and actions with respect to any detected OT candidate. Our classifier is generic enough so that a quick configuration of the CNN parameters can also make it usable for other kind of optical facilities.

This work amongst others shows how it is important now for wide field angle telescopes in the time domain astronomy to use such machine learning techniques to deal with huge data flow and big data analysis.

ACKNOWLEDGEMENTS

D.Turpin acknowledges the financial support from the Chinese Academy of Sciences (CAS) PIFI post-doctoral fellowship program (program C). D. Turpin is now supported by the CNES Postdoctoral Fellowship at Département d’astrophysique du CEA-Saclay. M. Ganet acknowledges the financial support of LAL and NAOC. S. Antier represents the financial support of LAL and NAOC. M. Ganet acknowledges the financial support from the CNES Postdoctoral Fellowship at Département d’astrophysique du CEA-Saclay. M. Ganet acknowledges the financial support of LAL and NAOC. S. Antier represents the financial support of LAL and NAOC.

REFERENCES

Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2018, Living Reviews in Relativity, 21, 3
Abbott, B. P., Abbott, R., Abbott, T. D., et al. 2017, ApJ, 848, L12
Adrián-Martínez, S., Ageron, M., Aharonian, F., et al. 2016, Journal of Physics G Nuclear Physics, 43, 084001
Akerlof, C. W., Kehoe, R. L., McKay, T. A., et al. 2003, PASP, 115, 132
Bellm, E. C., Kulkarni, S. R., Graham, M. J., et al. 2019, PASP, 131, 018002
Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
Bianchi, L., Herald, J., Efremova, B., et al. 2011, Ap&SS, 335, 161
Bishop, C. 2006, Pattern Recognition and Machine Learning (Springer)
Bloom, J. S., Richards, J. W., Nugent, P. E., et al. 2012, PASP, 124, 1175
Boër, M., Bringer, M., Klotz, A., et al. 1999, A&AS, 138, 579
Brown, L. M. J., Robson, E. I., Gear, W. K., et al. 1986, MNRAS, 219, 671
Castro-Tirado, A. J., Soldán, J., Bernas, M., et al. 1999, A&AS, 138, 583
Chambers, K. C., Magnier, E. A., Metcalfe, N., et al. 2016, arXiv e-prints, arXiv:1612.05560
de la Calleja, J. & Fuentes, O. 2004, MNRAS, 349, 87
Drake, A. J., Djorgovski, S. G., Mahabal, A., et al. 2009, ApJ, 696, 870
du Buisseno, L., Sivanandam, N., Bassett, B. A., & Smith, M. 2015, MNRAS, 454, 2026
Gaia Collaboration, Brown, A. G. A., Vallenari, A., et al. 2018, A&A, 616, A1
Gal-Yam, A. 2017, Observational and Physical Classification of Supernovae, 195
Géron, A. 2019, Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow (O’Reilly Media, Inc.)
Gieseke, F., Bloemen, S., van den Bogaard, C., et al. 2017, MNRAS, 472, 3101
Goldstein, D. A., D’Andrea, C. B., Fischer, J. A., et al. 2015, AJ, 150, 82
Hartman, R. C., Bertsch, D. L., Fichtel, C. E., et al. 1992, ApJ, 385, L1
Horúchi, S., Beacom, J. F., Kochanek, C. S., et al. 2011, ApJ, 738, 154
Ivezic, Z., Axelrod, T., Brandt, W. N., et al. 2008, Russian Astronomical Journal, 176, 1
Ivezic, Z., Kahn, S. M., Tyson, J. A., et al. 2019, ApJ, 873, 111
Jia, P., Zhao, Y., Xue, G., & Cai, D. 2019, AJ, 157, 250
Klebesadel, R. W., Strong, I. B., & Olson, R. A. 1973, ApJ, 182, 785
Klotz, A., Vachier, F., & Boër, M. 2008, Astronomische Nachrichten, 329, 275
Kumar, P. & Zhang, B. 2015, Phys. Rep., 561, 1
Law, N. M., Kulkarni, S. R., Dekany, R. G., et al. 2009, PASP, 121, 1395
Lecun, Y., Bengio, Y., & Hinton, G. 2015, Nature, 521, 436
Lipunov, V., Kornilov, V., Gorbovskoy, E., et al. 2010, Advances in Astronomy, 2010, 349171
Mahabal, A., Rebbapragada, U., Walters, R., et al. 2019, PASP, 131, 038002
Marshall, S., Akerlof, C., Kehoe, R., et al. 1997, in American Astronomical Society Meeting Abstracts, Vol. 191, American Astronomical Society Meeting Abstracts, 48.15
Masci, F. J., Laher, R. R., Rebbapragada, U. D., et al. 2017, PASP, 129, 014002
Mori, M., Ikeda, S., Tominaga, N., et al. 2016, Publications of the Astronomical Society of Japan, 68, 104
Muthukrishna, D., Narayan, G., Mandel, K. S., Biswas, R., & Hložek, R. 2019, PASP, 131, 118002
Narayan, G., Zaidi, T., Soraisam, M. D., et al. 2018, ApJS, 236, 9
Robson, E. I., Smith, M. G., Aycock, J., & Walther, D. M. 1988, IAU Circ., 4556, 2
Sánchez, B., Domínguez R., M. J., & Aragon, F., et al. 2019, Astronomy and Computing, 28, 100284
Skrutskie, M. F., Cutri, R. M., Stiening, R., et al. 2006, AJ, 131, 1163
Taylor, A. R. 2000, in Perspectives on Radio Astronomy: Science with Large Antenna Arrays, ed. M. P. van Haarlem, 1

GWAC OT candidate vetting using CNN

Bertin, E. & Arnouts, S. 1996, A&AS, 117, 393
Tonry, J. L., Denneau, L., Heinze, A. N., et al. 2018, PASP, 130, 064505
Turpin, D., Wu, C., Han, X. H., et al. 2020, RAA, 20, 1
Wang, J., Li, H. L., Xin, L. P., et al. 2019, arXiv e-prints, arXiv:1911.06678
Wei, J., Cordier, B., Antier, S., et al. 2016, arXiv e-prints, arXiv:1610.06892