



HAL
open science

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

Damien Turpin, M. Ganet, S. Antier, E. Bertin, L.P. Xin, N. Leroy, C. Wu, Y. Xu, X.H. Han, H.B. Cai, et al.

► To cite this version:

Damien Turpin, M. Ganet, S. Antier, E. Bertin, L.P. Xin, et al.. Vetting the optical transient candidates detected by the GWAC network using convolutional neural networks. *Monthly Notices of the Royal Astronomical Society*, 2020, 497 (3), pp.2641-2650. 10.1093/mnras/staa2046 . hal-02483987

HAL Id: hal-02483987



<https://hal.science/hal-02483987>

Submitted on 7 Jul 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

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, E. Bertin,⁶ L. P. Xin,¹ N. Leroy,⁷ C. Wu,¹ Y. Xu,^{1,8} X. H. Han,¹ H. B. Cai,¹ H. L. Li,¹ X. M. Lu,¹ Q. C. Feng¹ and J. Y. Wei^{1,8}

¹CAS Key Laboratory of Space Astronomy and Technology, National Astronomical Observatories, Chinese Academy of Sciences, Beijing 100012, China

²Département d'Astrophysique, Astrophysique, Instrumentation et Modélisation de Paris-Saclay, Université Paris-Saclay, CNRS, CEA, F-91191 Gif-sur-Yvette, France

³ENSAE Paris, F-91120 Palaiseau, France

⁴HEC Paris, F-78350 Jouy-en-Josas, France

⁵CNRS, Astroparticule et Cosmologie, F-75013 Paris, France

⁶Institut d'Astrophysique de Paris, UPMC-CNRS, UMR7095, F-75014 Paris, France

⁷IJCLab, Université Paris-Saclay, CNRS/IN2P3, F-91405 Orsay, France

⁸University of Chinese Academy of Sciences, Beijing 100049, China

Accepted 2020 July 7. Received 2020 April 23; in original form 2020 January 8

ABSTRACT

The observation of the transient sky through a multitude of astrophysical messengers has led to several scientific breakthroughs in the 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 multiwavelength and multimessenger follow-up campaigns 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, identification, and 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 (CNNs) to discard many common types of 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 OT sky down to $R = 16$ with a 15 s imaging cadency. We applied our trained CNN classifier on a sample of 1472 GWAC OT candidates detected by the real-time detection pipeline.

Key words: methods: data analysis – surveys – (transients:) neutron star mergers – transients: supernovae.

1 INTRODUCTION

The time-domain astronomy aims at studying transient phenomena with a wide variety of flux and time-scales detected within a very broad range of localization accuracies in the sky depending on the astrophysical messengers emitted [electromagnetic, gravitational waves (GWs), 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 e.g. Gal-Yam 2017, about the SNe classification). In the last century, the SNe were detected only at a rate of few per year,¹ mainly because the observational techniques and strategies were not optimized

to frequently detect such rare events.² Therefore, the workload pressure on the detection pipelines and classification tasks 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-m class telescope.

A first major revolution, in the so-called time-domain 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 (GRBs; Klebesadel, Strong & Olson 1973; Kumar & Zhang 2015) 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, multiwavelength follow-up observations across the whole electromagnetic spectrum became crucial to get a global picture of the physical processes

* E-mail: dturpin@nao.cas.cn, damien.turpin@cea.fr

¹See e.g. <http://www.rochesteerastronomy.org/snimages/snactive.html>

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

at work. The GRBs certainly represent one of the most extreme observational challenges for the follow-up telescopes. First, the short-living initial gamma-ray signal 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 multiwavelength afterglow emission that is fading very quickly so that it usually becomes unreachable for a detection few 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 kinds 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 (e.g. ROTSE, TAROT, BOOTES, and MASTER) that were³ capable of responding to any alert and monitoring a large fraction of the night sky continuously with a high cadence (Marshall et al. 1997; Boër et al. 1999; Castro-Tirado et al. 1999; Akerlof et al. 2003; Klotz, Vachier & Boër 2008; Lipunov et al. 2010). 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 analysed in 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 multimessenger (MM) astronomy (see e.g. Abbott et al. 2017; IceCube Collaboration 2018; Keivani et al. 2018) adding the high-energy neutrinos and the GW events in the global alert broadcasting system.⁴ 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 Rubin Observatory Legacy Survey of Space and Time (LSST; Ivezić et al. 2008, 2019), the Square Kilometer Array (SKA; Taylor 2000), KM3NeT (Adrián-Martínez 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 (Ivezić et al. 2019) will produce 20 TB of data every night with the possibility of having several millions of alerts per night starting from 2021 and running over 10 yr of operation. It should extend the known SNe catalogue by more than three billion 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), ASAS-SN,⁵ PanSTARRS (Chambers et al. 2016), ATLAS (Tonry et al. 2018), ZTF (Bellm et al. 2019), DES (Goldstein et al. 2015), or *Gaia* (Gaia Collaboration 2018). They quite naturally explore the optical transient (OT) sky and also participate in various MM follow-up campaigns (see e.g. Abbott et al. 2017). The data flows generated by those surveys and the subsequent processing to detect transients are already no longer manageable in a reasonable amount of time by the standard techniques previously

used for narrow field-of-view telescopes ($\text{FoV} \leq 1^\circ$) as shown, for example, by the ZTF Collaboration (Mahabal et al. 2019). In general, the transient detection pipelines are based on the point spread function (PSF)-matching and the catalogue 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; Jia et al. 2019; Mahabal et al. 2019; Sánchez 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 photometric and/or spectral properties and their temporal evolutions (e.g. Morii et al. 2016; Narayan et al. 2018; Muthukrishna et al. 2019). Among the zoo of *machine learning* algorithms, convolutional neural networks (CNNs) are known to be well adapted to ingest data containing multiple arrays like images (Bishop 2006; Lecun, Bengio & Hinton 2015). As the interest in using CNNs for image recognition purposes had rapidly grown in a large diversity of research domains during the last few years, extensive studies have been undertaken to explore their fields of applications and lead to very encouraging classification performances (see e.g. Simonyan & Zisserman 2014; Szegedy et al. 2014; Tompson et al. 2014). Generally speaking, CNNs have truly opened new perspectives for performing image recognition and detection tasks to such an extent that they are now the dominant deep machine learning methods used for such a work (see the review of Lecun et al. 2015, and the references therein). The CNNs are now also widely used in astronomy for performing classification tasks and especially the identification of real point-like sources and bogus. They employ multiple interconnected layers, similar to a neuronal network, to efficiently identify patterns in images that make them particularly suitable for the time-domain astronomy needs (Gieseke et al. 2017).

In this paper, we investigate the possibility of using CNNs for the vetting of the OT candidates that will be detected by the Ground Wide field Angle Cameras (GWAC) network. The GWAC system is a synoptic optical survey that is currently able to instantaneously cover 2000 deg² on the sky with a high-imaging cadency of one frame every 15 s. 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 multiwavelength 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 that have to be solved in order to exploit the full capabilities of the instruments.

In Section 2, we will describe the GWAC system and the transient detection pipeline that is currently running. Then, in 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 Section 4, and we finally draw our conclusions and perspectives for this work in Section 5.

³Most of them are still in operation.

⁴Most of the current MM alerting systems use the Global Coordinates Network (GCN) to broadcast alerts in real time; see the GCN archive: https://gcn.gsfc.nasa.gov/gcn3_archive.html

⁵<http://www.astronomy.ohio-state.edu/~assassin/index.shtml>



Figure 1. The GWAC telescope network at the Xinglong Observatory in China. Currently, 4 mounts are operational among 10 at completion. Each mount is equipped with four JFoV cameras (18 cm) and one FFoV camera (3.5 cm) located at the centre of the mount. The total FoV of the current GWAC network is about 2000 deg².

Table 1. Characteristics of the GWAC JFoV cameras.

Parameter	Value
Field of view	150 deg ²
Diameter	18 cm
CCD pixel size	13 μm
Pixel scale	11.7 arcsec
Readout noise	14 e ⁻
R_{lim} (single 10 s/stack frame 1.5 h)	16/18

2 THE GWAC TELESCOPES

2.1 Instrumentation set-up

Since the end of 2017, the GWAC telescopes are under development in China at the Xinglong Observatory. Each GWAC telescope mount is equipped with five cameras: four JFoV cameras (4k × 4k CCD E2V camera with an aperture of 180 mm) and 1 FFoV camera (3k × 3k CCD camera with an aperture of 35 mm) mainly used to monitor the sky seeing and brightness conditions (see Fig. 1). The main scientific instruments, the JFoV cameras, cover a field of view of about 12.4° × 12.4° per camera (~150 deg² per camera), for more characteristic details see Table 1. Taking into account the overlaps between the fields of view of the 4 JFoV cameras, a GWAC mount finally covers 500 deg² on the sky. Each JFoV camera is designed to reach an unfiltered limiting magnitude of about 16 in a dark night for 10 s of exposure. A stacking analysis of the single frames can be performed in real time to reach a maximum limiting magnitude of $R \sim 18$ in clear and dark night as shown in Turpin et al. (2020).

2.2 The GWAC OT 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 Coordinate System) calibration. Those calibrated images are then automatically and independently analysed 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 catalogue cross-matching and the differential image analysis (DIA). Concerning the

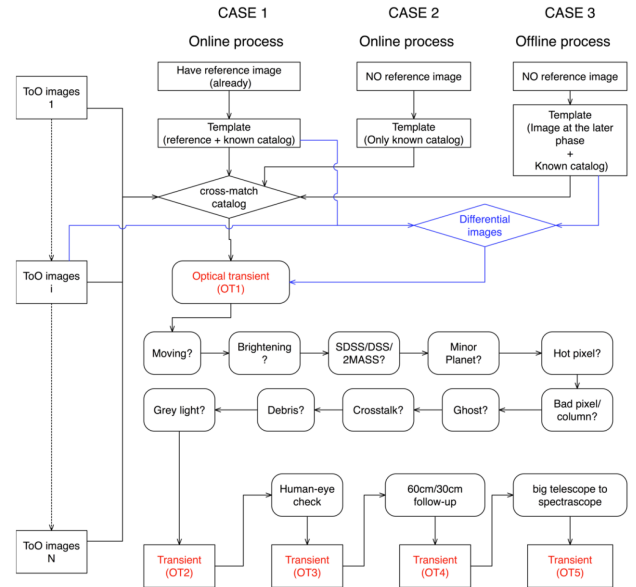


Figure 2. A schematic view of the current GWAC detection pipeline set-up to detect and identify the OT sources in both single and stacked images.

GWAC system, more details can be found in Turpin et al. (2020) and Wang et al. (2019) but typically a new source is detected once it fulfils the following criteria:

- (i) The source has a signal-to-noise ratio (SNR) ≥ 5 and is not detected down an SNR = 5 in the reference images.
- (ii) The source is detected in several successive images.
- (iii) The PSF of the source shall be stellar-like profile, i.e. a 2D Gaussian profile.
- (iv) No 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, the SNR, the optical peak flux, the source position, etc.) on at least five 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 catalogue cross-matching filter using deeper catalogues is then applied to the *OT1 candidates* that passed the first selection criteria (see Fig. 2). Catalogues such as *Gaia* DR2 (Gaia Collaboration 2018), PanStarrs DR1 (Chambers et al. 2016), 2MASS (Skrutskie et al. 2006), *Galex* DR5 (Bianchi et al. 2011), or public data bases on Solar system objects such as the Minor Planet Center⁶ are used to perform this task.

After passing all of those filters, the remaining candidates are grouped in the *OT2 candidates*. Subimages are then cropped from each initial 4k × 4k JFoV image and subtracted from the sky background contribution to make 100 × 100 pixel-sized finding charts centred at the position of each selected *OT2 candidate*. These finding charts are then checked one by one by a human visual inspection. 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

⁶<https://minorplanetcenter.net/iau/mpc.html>

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*. The GWAC-F60 telescopes can perform about 300 individual pointings every night at maximum. 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 arcmin), 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 to search for bright OT events with a minimum of a subminute 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 set-up implies the collection of a huge amount of data every night with typically between 6000 and 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 Fig. 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 deg²) combined with the limited size of the CCD detectors produces a large pixel scale of 11.7 arcsec per pixel and image distortion effects (while corrected in our images). These two factors make the use of the catalogue 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. On average, each night with good observational conditions yields the discovery of around several hundreds up to thousands of *OT2 candidates* for one GWAC mount. As the number of *OT2 candidates* is still large, we then manually reduce this number to several tens (up to one hundred) to be vetted both by humans and by further follow-up observations.

Multiplying this task to the number of GWAC mounts, one can easily understand that this ‘true or false’ classification task becomes no longer manageable both by the GWAC-F60 follow-up telescopes and by the GWAC scientists in a reasonable amount of time. Therefore, our GWAC-F60 telescopes can be rapidly unable to ingest the quantity of GWAC triggers and additionally they can no longer smoothly follow their own observation plans independently of the GWAC activities. Moreover, the increasing workload pressure on the scientists on duty naturally prevents them from focusing their efforts on the most promising events. The identification and classification processes of a genuine transient source then may undergo a long delay that is not compatible with the scientific purposes of the GWAC system that aim to quickly identify short-living OT 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 responsive. As shown previously, there is a crucial need for a classification task that distinguishes the astrophysical sources from the bogus within the GWAC alert stream prior to building the *OT2 candidates* list. Before going deeper into the details, we start to define few acronyms that we will use all along the paper:

- (i) *ROS*: real optical sources in an image.
- (ii) *FOS*: fake optical sources in an image.
- (iii) *ROT*: real optical transients. An ROT is actually an ROS present in a series of images and showing a significant flux variation.
- (iv) *FOT*: fake optical transients.
- (v) *TP*: true positives, i.e. the OT candidates well classified as ROT or ROS.
- (vi) *TN*: true negatives, i.e. the OT candidates well classified as FOT or FOS.
- (vii) *FP*: false positives, i.e. the OT candidates classified as ROT or ROS while there are actually FOT or FOS.
- (viii) *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 on one hand and to avoid too many losses of ROT because of misclassifications on the other hand. The final goal is to obtain a classification accuracy greater than 90 per cent with an FN classification not as great as 2 per cent. 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 CNN algorithm.

This choice is first motivated by the fact that CNNs are very well adapted for pattern recognition in images (Bishop 2006; Krizhevsky, Sutskever & Hinton 2012; Lecun et al. 2015) as we previously mentioned. In addition, the CNNs have demonstrated excellent image classification performances compared to other deep *machine learning* methods, like the random forest (RF) algorithms, with a minimum of implementation (Gieseke et al. 2017). However, compared to RF algorithms, the CNNs require a very large data set to be efficiently trained and a high computational cost for the training steps (a graphics processing unit, GPU, is usually required for such a work). We also point out the fact that one of the great advantages of the CNNs is their ability of classifying on their own once the image features have been learned during the training steps. In other words, our trained CNN will not require any more input information than an image to perform its classification task. On the contrary, the RF algorithms analyse features extracted from the images (data and metadata) before making the classification task. This requires a systematic pre-processing analysis of the images that might bring too much complexity and latency in the overall data processing pipeline. Therefore, the use of the CNN allows us to avoid such an issue and to keep our data processing architecture almost unchanged. Finally, during the last few years, CNNs have been also robustly tested in astronomy with success for many different classification purposes (de la Calleja & Fuentes 2004; Bloom et al. 2012; du Buisson et al. 2015; Mahabal et al. 2019), which reinforces our choice to use them for our astronomical classification purpose.

Table 2. The CNN structure used in this work.

Layers	Sizes	Characteristics
Convolution	32×32	3×3 kernels Activation: ReLU
Max pooling	32×32	2×2
Convolution	32×32	3×3 kernels Activation: ReLU
Max pooling	32×32	2×2
Dense	15	Activation: ReLU
Dense	2	Activation: Softmax

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, and one ReLU and one softmax hidden layer (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 FP, which we strongly want to avoid. The CNN was implemented in PYTHON v3.6, using the KERAS⁷ package with TENSORFLOW2.⁸ The KERAS package has the advantages to provide built-in diagnostic tools and a compact code writing that allow for a relative ease of use. A KERAS Adam optimizer was used with a low learning rate ($\text{lr} = 0.0001$) after witnessing disappointing convergence properties.

As an input, our CNN algorithm uses background-subtracted finding charts (100×100 pixels) of the *OTI candidates*. We then select only the central part of those images (35×35 pixels) for the classification. This choice is made 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 10^5 images) while keeping enough information (background and a minimum number of sources) in the subimages 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 labelling previously made by our expert scientists, which consists in giving either a mark ‘1’ to subimages containing an ROS or ‘0’ if they contain FOS. 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 optimized for our classification purpose or the human labelling is not correct. In such case, the CNN architecture and/or the labelling have to be revised until a good agreement is found.

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 OT 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 Fig. 3, the implementation we set up for both the training and the validation of the CNN algorithm as well as how it should be integrated 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 of images are needed to obtain a well-trained CNN model.

Classifying our detected transients into several astrophysical categories based on additional information 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 are 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 training data set is finally composed of 200 000 subimages (35×35 pixels) with an equal distribution between FOS and ROS. Among them, 180 000 are directly used to train the algorithm while the 20 000 remaining images are used to validate each training epoch.

3.2.1 Details on the ROS image sample

The ROS subimage sample is built from several $4k \times 4k$ GWAC images taken from the same camera during 1 yr of operation. Therefore, we have at our disposal a complete overview of the observational and sky background conditions we can encounter at the GWAC site. The $4k \times 4k$ initial images are chosen randomly and background-subtracted to follow the GWAC detection pipeline process previously described in Section 2.2. In each of the selected images, we extracted the position of the point sources detected by the SEXTRACTOR software (Bertin & Arnouts 1996) at the 3σ confidence level. From this list of sources, we then randomly cropped 35×35 pixel subimages around the SEXTRACTOR positions of 100 randomly chosen sources. However, we make a selection cut on the instrumental magnitudes estimated by SEXTRACTOR as we want to avoid very bright or ‘saturated’ stars that may produce artefacts such as blooming effect. During the source extraction process and the creation of the finding charts, we noticed that the current GWAC detection pipeline can sometimes shift the centroid of the OT candidate from the centre of the finding charts from 1 pixel at maximum in any direction. To be as close as possible to the GWAC

⁷<https://keras.io/>. See also Géron (2019) for a review of the usages of KERAS.

⁸<https://www.tensorflow.org/>
<https://github.com/tensorflow/tensorflow>

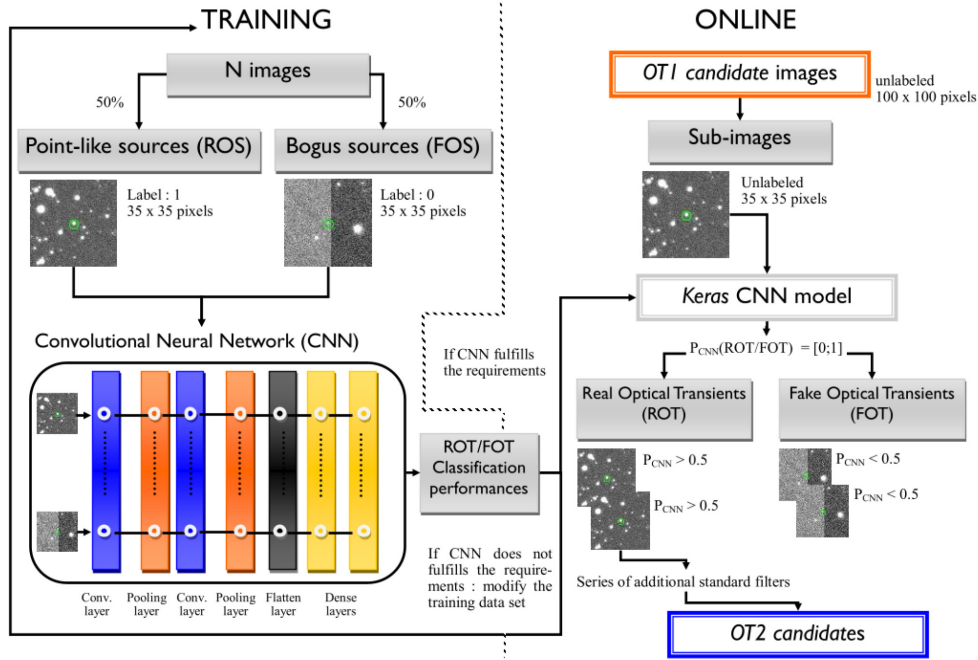


Figure 3. Schematic view of the implementation of the classifier tool from the training of the CNN algorithm (*left-hand side*) to the use of the CNN KERAS model to make the vetting of the GWAC *OTI* candidates in real time. We used a large data set of $N = 200\,000$ images to train the CNN.

pipeline output, we also reproduced this trend for each of our ROS subimages. The centroid of each extracted ROS is therefore shifted in position in all the direction possible by an increment (uniformly) randomly chosen in the range of $[-1; 1]$ pixels. We reproduced this operation on 1000 different $4k \times 4k$ images to obtain a final sample of 100 000 images of ROS. We show, in Fig. 4, a subsample 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, crosstalk 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 into 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 as 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 per cent of the full training data set) in equal proportions between our five categories defined above. Our bogus simulator starts with the same process as for extracting ROS from the $4k \times 4k$ GWAC images. From the background-subtracted initial images, we extract 35×35 pixel subimages and add a bogus in the central position similarly as we did for the ROS. Then, the differences in the process come; depending on the bogus to simulate, we crop different parts of the $4k \times 4k$ images according to the following criteria:

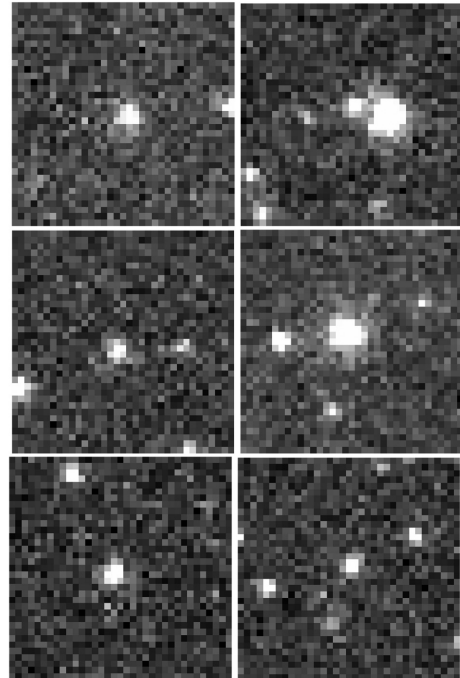


Figure 4. Example of some subimages (35×35 pixels) centred (± 2 pixels) at the position of point-like sources (ROT) extracted from the GWAC $4k \times 4k$ images. 100 000 images similar to these ones are produced to build our ROT training data set. Note that we extract point-like source with no prescription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

- (i) For the *hot pixel* subimages: no *Sextractor* sources should have a position (X, Y) in the subimage consistent within a region of at least 6 pixels around the central pixel (X_0, Y_0) : $(X - X_0)^2 + (Y - Y_0)^2 \geq 36$.

(ii) For the *noisy* subimages: no *Sextractor* sources should be present in the subimages, only the background residual noise.

(iii) For the *bad pixel columns* subimages: any subimage randomly chosen is suitable.

(iv) For the *dark pixel* subimages: any subimage randomly chosen is suitable.

(v) For the *non-uniform sky background* subimages: 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 centre of the image to the position of bright star centroid can span $d \in [6; 15]$ 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 subimages following the pixel intensities we observed from real data. We also add a random increment spanning $[-1; 1]$ pixels to slightly shift the position of the bogus from the centre. The bad column of pixels was simulated as an excess of light observed normalized 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 in the Y direction of the image. In Fig. 5, we compare the 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 labelled ‘1’ while the FOS are labelled ‘0’. We then compared this labelling 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.

4.1 The training

We trained our CNN algorithm on the 200 000 simulated images (50 per cent ROS, 50 per cent FOS) making 10 training epochs to build the final *KERAS* model. Based on the P_{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, and the fraction of the misclassified ones in the FN and FP categories, as shown in Fig. 6. The normalized values of TN, TP, FN, and FP obey the following relations:

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

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. Indeed, the TP and the FP instances are almost maximized up to a normalized value of ‘1’ (i.e. all instances are well classified) while the FN and the FN are close to the minimum ‘0’. To better characterize the classification response of our CNN model, we also computed three diagnosis:

(i) The *receiver operating characteristic* (ROC) curve that displays the *true positive rate* (TPR) as a function of the *false positive rate*.

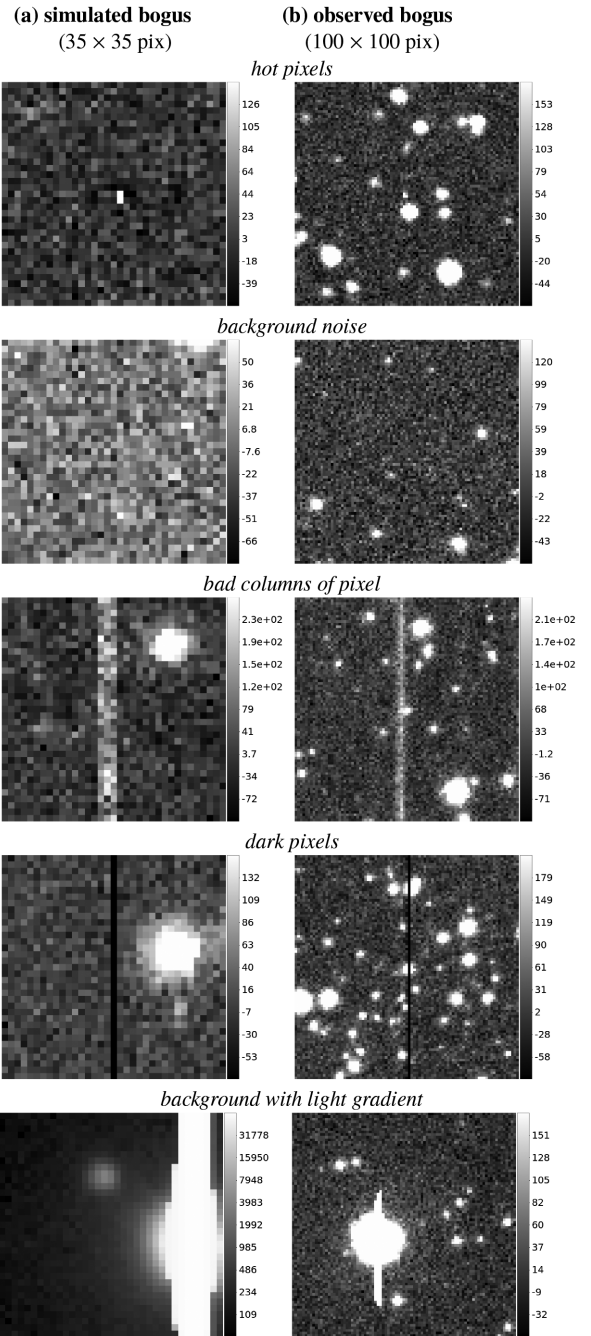


Figure 5. The five categories of bogus simulated for the FOS training data set (a) compared with (b) the same kind of bogus we indeed observed in the GWAC OT1 finding charts (100×100 pixels).

(ii) The *area under the ROC curve* (AUC) that corresponds to the integral of the ROC curve $\in [0; 1]$. ‘0’ or ‘1’ corresponds to an ideal case where 100 per cent of the instances are mis- or well classified.

(iii) The *accuracy coefficient* (AC) $\in [0; 1]$. ‘0’ or ‘1’ corresponds to an ideal case where 100 per cent of the instances are mis- or well classified.

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

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

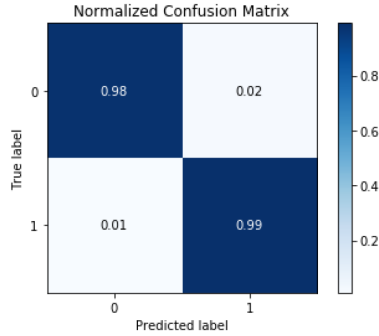


Figure 6. The normalized confusion matrix produced after the training of the CNN algorithm on 200 000 simulated images of bogus and real sources. The numbers in each blue square indicate the fraction of the total instances correctly classified as FP (top left) and TP (bottom right) while in the white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

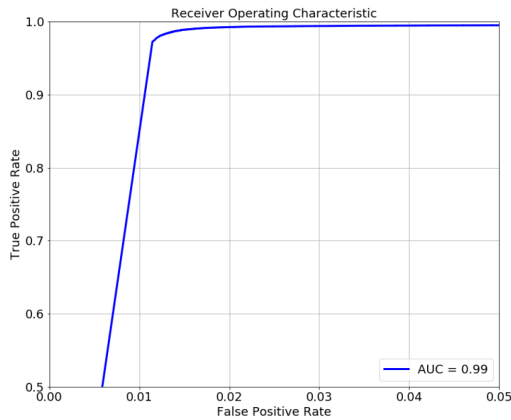


Figure 7. The ROC curve of our CNN model applied to the training data set. The AUC value is also indicated at the bottom.

maximum value ‘1’) while keeping an extremely low FPR (close to ‘0’; see Fig. 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 from real sources.

However, while it allows 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 of data for the simulation that translates in a higher computational cost and a severe complexification of the simulation. 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 as our simulated ones. As an example, we did not simulate 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).

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

Category	N_{images}	N_{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 diag.
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

4.2 The validation

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

As for the training sample, the probability given by the CNN on each image, P_{CNN} , is compared with the image labelling 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 per cent of the total sample. Around only 4 per cent of the images containing a bogus are misclassified as ROS (FP) as shown in Fig. 8. These classification performances are in good agreement with our scientific requirements mentioned in Section 3; 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 OT 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 per cent) of completely new bogus types such as dust obstructions, suspected

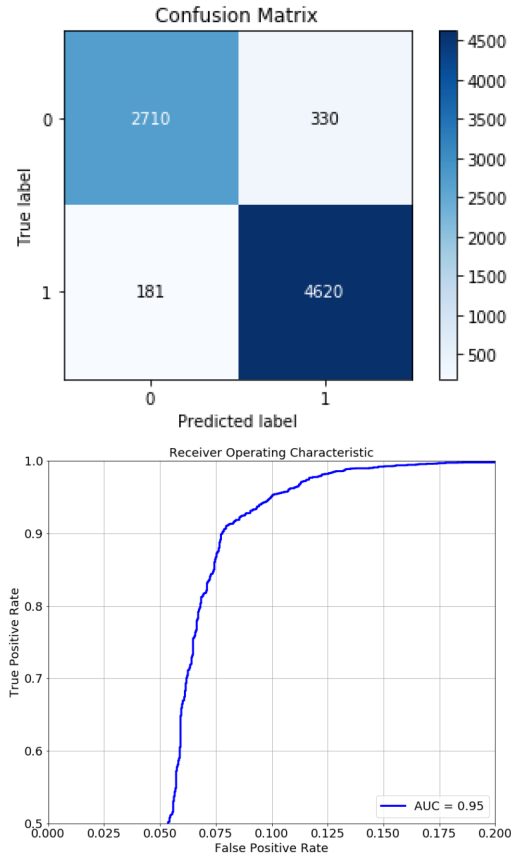


Figure 8. The confusion matrix and the ROC curve of our trained CNN model applied on a complete unseen data set of 7841 images of bogus and real astrophysical sources. The classification diagnosis AUC is around 0.95 in good agreement with our scientific classification requirements.

ROS or low SNR candidates, and a large variety of bad pixels. While the addition of those new bogus makes the classification accuracy drop 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.

4.3 Analysis of the bogus rejection and false positive detections

The analysis presented above only considered the classification of the individual images of bogus and real sources. However, the vetting of the OT candidates must also include the image time series of the source candidates. The rejection of bogus is made by analysing the evolution of the CNN scores across several images. The mean of the CNN probabilities that tracks the stability of the CNN classification over the image time series is used as a criterion of rejection. Playing on this criterion allows to determine the final rate of FP and FN the system will tolerate. The current GWAC pipeline is taking a decision about the genuineness of the candidates after analysing five consecutive images. For comparison, we used the same number of images to take a decision with the CNN. If we have less than five images for a given candidate we computed the mean of P_{CNN} on a minimum of two images. As shown in Equation (3), we choose different rejection criteria in order to analyse the evolution of the FP

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

R	FP (%)	FN (%)
$\leq 5\sigma$	4.6	2.9
$\leq 3\sigma$	7.2	1.7
$\leq 2\sigma$	8.9	1.0
$\leq 1\sigma$	10.9	0.7

and FN as a function of the strictness of our rejection:

$$R = \sum_{i=0}^{N=5} \frac{P_{\text{CNN},i}}{N} \leq 5\sigma, 0.997 (3\sigma), 0.95 (2\sigma), \text{ and } 0.68 (1\sigma). \quad (3)$$

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 and 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 per cent of ROT loss and about 7 per cent of false positive detections (~ 91 per cent of OT candidates 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 an $\text{SNR} \leq 3$, which make them hard to be clearly identified by our CNN algorithm trained on securely detected OT. We find that this ratio of *false positives* is manageable on a whole night by the GWAC data processing pipeline, the scientist on duty, and the GWAC-F60 telescope schedulers.

5 CONCLUSION

The fast identification and classification of the transient sources are the major challenges 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 ($\text{FoV} = 25^\circ, R_{\text{lim}} = 16$ in 10 s) 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 OT candidates. To solve the problem of OT vetting in the GWAC images, we used a convolutional neural network classifier trained on computationally enhanced data relying on the GWAC data base 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 number of images for the decision. The final false positive alarm ratio is about 4 per cent when it is applied to the individual images (94 per cent of OT candidates well classified). When applying the CNN classifier on the image time series of each OT candidate, we end up with about 7 per cent of FP OT 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 per cent 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 OT 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, among others, shows how it is important now for wide field angle telescopes studying the transient sky to use such machine learning techniques to deal with huge data flows and big data analysis.

ACKNOWLEDGEMENTS

DT acknowledges the financial support from the Chinese Academy of Sciences (CAS) PIFI post-doctoral fellowship program (program C). DT was supported by the CNES Postdoctoral Fellowship at Département d'astrophysique at CEA-Saclay. MG acknowledges the financial support of Laboratoire de l'Accélérateur Linéaire (the LAL is now the CNRS/IJCLab) and the National Astronomical Observatories of China (NAOC, Beijing). SA was supported by the CNES Postdoctoral Fellowship at Laboratoire AstroParticule et Cosmologie. This work was supported by the National Natural Science Foundation of China (grant numbers 11533003, 11973055 and U1931133) as well as the Strategic Priority Research Program of the Chinese Academy of Sciences grant no. XDB23040000 and the Strategic Pioneer Program on Space Science, Chinese Academy of Sciences, grant no. XDA15052600.

DATA AVAILABILITY

The data underlying this article will be shared on reasonable request to the corresponding author.

REFERENCES

Abbott B. P. et al., 2017, *ApJ*, 848, L12
 Abbott B. P. et al., 2018, *Living Rev. Relativ.*, 21, 3
 Adrián-Martínez S. et al., 2016, *J. Phys. G: Nucl. Part. Phys.*, 43, 084001
 Akerlof C. W. et al., 2003, *PASP*, 115, 132
 Bellm E. C. et al., 2019, *PASP*, 131, 018002
 Bertin E., Arnouts S., 1996, *A&AS*, 117, 393
 Bianchi L., Herald J., Efremova B., Girardi L., Zabot A., Marigo P., Conti A., Shiao B., 2011, *Ap&SS*, 335, 161
 Bishop C., 2006, *Pattern Recognition and Machine Learning*. Springer-Verlag, New York
 Bloom J. S. et al., 2012, *PASP*, 124, 1175
 Boër M. et al., 1999, *A&AS*, 138, 579

Brown L. M. J., Robson E. I., Gear W. K., Crosthwaite R. P., McHardy I. M., Hanson C. G., Geldzahler B. J., Webb J. R., 1986, *MNRAS*, 219, 671
 Castro-Tirado A. J. et al., 1999, *A&AS*, 138, 583
 Chambers K. C. et al., 2016, preprint ([arXiv:1612.05560](https://arxiv.org/abs/1612.05560))
 de la Calleja J., Fuentes O., 2004, *MNRAS*, 349, 87
 Drake A. J. et al., 2009, *ApJ*, 696, 870
 du Buisson L., Sivanandam N., Bassett B. A., Smith M., 2015, *MNRAS*, 454, 2026
 Gaia Collaboration, 2018, *A&A*, 616, A1
 Gal-Yam A., 2017, *Observational and Physical Classification of Supernovae*. Springer-Verlag, Berlin, p. 195
 Géron A., 2019, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. O'Reilly Media, Inc.
 Gieseke F. et al., 2017, *MNRAS*, 472, 3101
 Goldstein D. A. et al., 2015, *AJ*, 150, 82
 Hartman R. C. et al., 1992, *ApJ*, 385, L1
 Horiuchi S., Beacom J. F., Kochanek C. S., Prieto J. L., Stanek K. Z., Thompson T. A., 2011, *ApJ*, 738, 154
 IceCube Collaboration, 2018, *Science*, 361, 1378
 Ivezić Z. et al., 2008, *Serbian Astron. J.*, 176, 1
 Ivezić Ž. et al., 2019, *ApJ*, 873, 111
 Jia P., Zhao Y., Xue G., Cai D., 2019, *AJ*, 157, 250
 Keivani A. et al., 2018, *ApJ*, 864, 84
 Klebesadel R. W., Strong I. B., Olson R. A., 1973, *ApJ*, 182, L85
 Klotz A., Vachier F., Boër M., 2008, *Astron. Nachr.*, 329, 275
 Krizhevsky A., Sutskever I., Hinton G., 2012, *Neural Inf. Process. Syst.*, 25, 84
 Kumar P., Zhang B., 2015, *Phys. Rep.*, 561, 1
 Law N. M. et al., 2009, *PASP*, 121, 1395
 Lecun Y., Bengio Y., Hinton G., 2015, *Nature*, 521, 436
 Lipunov V. et al., 2010, *Adv. Astron.*, 2010, 349171
 Mahabal A. et al., 2019, *PASP*, 131, 038002
 Marshall S. et al., 1997, *BAAS*, 191, 48.15
 Masci F. J. et al., 2017, *PASP*, 129, 014002
 Morii M. et al., 2016, *PASJ*, 68, 104
 Muthukrishna D., Narayan G., Mandel K. S., Biswas R., Hložek R., 2019, *PASP*, 131, 118002
 Narayan G. et al., 2018, *ApJS*, 236, 9
 Robson E. I., Smith M. G., Ayccock J., Walther D. M., 1988, *Int. Astron. Union Circ.*, 4556, 2
 Sánchez B. et al., 2019, *Astron. Comput.*, 28, 100284
 Simonyan K., Zisserman A., 2014, preprint ([arXiv:1409.1556](https://arxiv.org/abs/1409.1556))
 Skrutskie M. F. et al., 2006, *AJ*, 131, 1163
 Szegedy C. et al., 2014, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), p. 9
 Taylor A. R., 2000, in van Haarlem M. P., ed., *Perspectives on Radio Astronomy: Science with Large Antenna Arrays*. ASTRON, p. 1
 Tompson J., Goroshin R., Jain A., LeCun Y., Bregler C., 2014, preprint ([arXiv:1411.4280](https://arxiv.org/abs/1411.4280))
 Tonry J. L. et al., 2018, *PASP*, 130, 064505
 Turpin D. et al., 2020, *Res. Astron. Astrophys.*, 20, 1
 Wang J. et al., 2019, preprint ([arXiv:1911.06678](https://arxiv.org/abs/1911.06678))
 Wei J. et al., 2016, preprint ([arXiv:1610.06892](https://arxiv.org/abs/1610.06892))

This paper has been typeset from a $\text{\TeX}/\text{\LaTeX}$ file prepared by the author.