Computers, Environment and Urban Systems 90 (2021) 101692

Contents lists available at ScienceDirect

Computers, Environment and Urban Systems

journal homepage: www.elsevier.com/locate/ceus

# Automatic detection of potential mosquito breeding sites from aerial images acquired by unmanned aerial vehicles

Daniel Trevisan Bravo [a], Gustavo Araujo Lima [a], Wonder Alexandre Luz Alves [a], Vitor Pessoa Colombo [b], Luc Djogbénou [c], Sergio Vicente Denser Pamboukian [d], Cristiano Capellani Quaresma [e], Sidnei Alves de Araujo [a,*]

[a] Informatics and Knowledge Management Postgraduate Program, Nove de Julho University, Vergueiro Street, 235/249, Liberdade, São Paulo/SP, Brazil
[b] Communauté d'Études pour l'Aménagement du Territoire (CEAT), École Polytechnique Fédérale de Lausanne, Bâtiment BP – Station 16, CH-1015 Lausanne/VD, Switzerland
[c] Centre de Recherche Pour la Lutte contre les Maladies Infectieuses (CReMIT), Université d'Abomey-Calavi (UAC), Campus d'Abomey-Calavi BP 526, Cotonou, Bénin
[d] Science and Geospatial Applications Postgraduate Program, Mackenzie University, Consolação Street, 896 - building 45, 7th floor - Consolação, São Paulo/SP, Brazil
[e] Smart and Sustainable Cities Postgraduate Program, Nove de Julho University, Vergueiro Street, 235/249, Liberdade, São Paulo/SP, Brazil

ARTICLE INFO

*Keywords:*
Vector control
Mosquito
Unmanned aerial vehicle
Objects detection
Convolutional neural network
Support vector machine
Bag of visual words

ABSTRACT

The World Health Organization (WHO) has stated that effective vector control measures are critical to achieving and sustaining reduction of vector-borne infectious disease incidence. Unmanned aerial vehicles (UAVs), popularly known as drones, can be an important technological tool for health surveillance teams to locate and eliminate mosquito breeding sites in areas where vector-borne diseases such as dengue, zika, chikungunya or malaria are endemic, since they allow the acquisition of aerial images with high spatial and temporal resolution. Currently, though, such images are often analyzed through manual processes that are excessively time-consuming when implementing vector control interventions. In this work we propose computational approaches for the automatic identification of objects and scenarios suspected of being potential mosquito breeding sites from aerial images acquired by drones. These approaches were developed using convolutional neural networks (CNN) and Bag of Visual Words combined with the Support Vector Machine classifier (BoVW + SVM), and their performances were evaluated in terms of mean Average Precision - mAP-50. In the detection of objects using a CNN YOLOv3 model the rate of 0.9651 was obtained for the mAP-50. In the detection of scenarios, in which the performances of BoVW+SVM and a CNN YOLOv3 were compared, the respective rates of 0.6453 and 0.9028 were obtained. These findings indicate that the proposed CNN-based approaches can be used to identify potential mosquito breeding sites from images acquired by UAVs, providing substantial improvements in vector control programs aiming the reduction of mosquito-breeding sources in the environment.

## 1. Introduction

Vector-borne diseases transmitted by mosquitoes pose a great public health challenge today. For instance, the number of notified dengue cases has rapidly increased during the past decades and, according to the World Health Organization (WHO, 2012), "the emergence and spread of all four dengue viruses ('serotypes') from Asia to the Americas, Africa and the Eastern Mediterranean regions represent a global pandemic threat". In the Americas, over 1,6 million cases were notified only in the first five months of 2020, the majority of them in Brazil (PAHO, 2020). Although it is difficult to accurately estimate the real number of cases, as the latter are often underreported, a recent study suggests that, globally, in 2010 there were about 390 million infections of dengue (Bhatt et al., 2013). At present, this figure is sustained, as the WHO estimates that between a 100 and 400 million infections happen every year globally (WHO, 2020a). Malaria is another disease transmitted by mosquitoes that inflicts a relatively high burden, especially in low- and middle-income countries: in 2018, a total of 405′000 deaths were attributed to malaria – of which 380′000 (94%) happened in Africa (WHO, 2019).

The reduction of potential mosquito breeding sites is an important preventive measure to tackle the burden of such diseases and make effort toward their elimination. However, health authorities are often

* Corresponding author.
*E-mail address:* saraujo@uni9.pro.br (S.A. Araujo).

https://doi.org/10.1016/j.compenvurbsys.2021.101692
Received 8 March 2021; Received in revised form 27 July 2021; Accepted 29 July 2021
Available online 12 August 2021

overwhelmed by this challenge, especially in areas where a considerable share of the population have neither access to adequate housing structures nor well-managed, basic services such as water, sanitation and solid waste disposal, which are determinant factors for the risk of transmission of mosquito-borne diseases (Barrera, Navarro, Mora Rodríguez, Domínguez, & González García, 1995; WHO, 2017). Additionally, as the global population becomes increasingly urban, and both the socio-economic and physical consolidation of vulnerable human settlements ("favelas", "bidonvilles", "slums", etc.) prove to be an enduring challenge (UN-Habitat, 2016), it is crucial to develop new techniques that can support vector control in cities characterized by high spatial and social complexity.

In fact, although higher demographic densities and urbanization have been associated with lower incidence of mosquito borne diseases (Araujo et al., 2015; Kabaria, Gilbert, Noor, Snow, & Linard, 2017), it is important to note that such risks vary greatly within cities (De Silva & Marshall, 2012). Typically, mosquito-borne diseases represent an increasing threat to the rapidly growing, poor peri-urban areas around the world (Espinosa, Polop, Rotela, Abril, & Scavuzzo, 2016; Keiser et al., 2004). Also, temperature conditions linked to the urban environment have been associated with dengue incidence (Araujo et al., 2015). Moreover, different vector species have adapted to the urban habitat, like the *Aedes aegypti* – the primary vector of dengue (Simmons, Farrar, van Vinh Chau, & Wills, 2012) as well as other diseases such as Zika, yellow fever or Chikungunya (WHO, 2020b) – or the *Anopheles gambiae* and the *Anopheles stephensis*, main vectors of malaria (Warren, Billing, Bendahmane, & Wijeyaratne, 1999). Vector control is therefore a crucial issue to both public health officials (in terms of epidemiological surveillance) and urban planners (in terms of providing basic infrastructures).

Given the spatial dynamics of rapid and informal settlement growth in the global South, where most cases of mosquito-borne disease happen, there is a need for tools that allow for the rapid localization and elimination of potential mosquito breeding sites in order to prevent outbreaks. In this sense, several efforts have been made, mainly in areas with limited accessibility to health surveillance agents. One of these efforts is the use of Unmanned Aerial Vehicle (UAVs), also known as drones, for the acquisition of aerial images in places with a higher incidence of diseases transmitted by mosquitoes (Carrasco-Escobar et al., 2019; Diniz & Medeiros, 2018; Passos et al., 2018). With such equipment, it is possible to obtain images with high spatial and temporal resolutions, allowing the detection of small objects on the Earth's surface and the detection of changes in a given region and in a short period of time. In addition, this technique requires fewer financial resources than manned aircraft missions and, because the flights are done closer to the ground, it also allows to obtain images with higher spatial resolution. The latter can be used to identify potential mosquito breeding sites such as water containers (Schafrick, Milbrath, Berrocal, Wilson, & Eisenberg, 2013), shaded, peri-domestic areas with poorly managed vegetation (Madzlan, Dom, Tiong, & Zakaria, 2016) or concentrations of solid waste (Barrera et al., 1995).

Indeed, thanks to the high-resolution imagery provided by UAVs, they can be a powerful instrument to support targeted interventions aiming to eliminate mosquito breeding sites and thus prevent diseases such as dengue or malaria. The Zanzibar Malaria Elimination Programme (ZAMEP) is a concrete example of how public authorities can use UAVs to map mosquito habitats (water bodies) to target larval source management efforts (Hardy, Makame, Cross, Majambere, & Msellem, 2017). Another example comes from the government of the city São Paulo which already uses drone to combat the vectors of dengue, chikungunya and zika (PMSP, 2016). Furthermore, Tun-Lin et al. (2009) have shown that, in the case of dengue vector control, targeting only the main breeding sites (roughly 50% of the total number) can be as effective as targeting all sites – at the same time it optimizes human and financial resources. In fact, UAVs offer a great potential to facilitate the identification of the main breeding sites and thus optimize intervention efforts and resources.

Considering these practical advantages, researchers across the globe have explored different methods based on the use of UAVs to detect mosquito breeding sites. However, most current applications often rely on manual or semi-automatic detection methods, and lack of accuracy in terms of the location of these sites. Currently, there are still few studies proposing the use of UAVs to detect potential mosquito breeding sites from automatic analysis of images, such as the works of Agarwal, Chaudhuri, Chaudhuri, and Seetharaman (2014), Mehra, Bagri, Jiang, and Ortiz (2016), Carrasco-Escobar et al. (2019) and Haas-Stapleton, Barretto, Castillo, Clausnitzer, and Ferdan (2019). In the first two studies, the authors advance approaches to analyze an image and indicate if it contains a suspicious scenario, but without providing its location. The works of Carrasco-Escobar et al. (2019) and Haas-Stapleton et al. (2019), on the other hand, investigate the detection of mosquito breeding sites based on the analysis of water bodies characteristics.

In this work, we developed different approaches to detect and locate "objects" and "scenarios" that represent potential mosquito breeding grounds from aerial images acquired by UAVs. The target "objects" are containers used to store water for domestic use ("water tanks"); while the target "scenarios" are defined as exterior areas (of varying sizes) with accumulated inorganic garbage, comprising objects that can hold stagnant water such as old tires, pet bottles, plastic and paper packaging, among others. Many of these objects were reported as productive breeding places in the study conducted by Tun-Lin et al. (2009). To detect target objects, we developed an approach using a Convolutional Neural Network (CNN) YOLOv3 model, called CNN_Objects, from the YOLO (You Only Look Once) framework (Redmon, Divvala, Girshick, & Farhadi, 2016; Redmon & Farhadi, 2018), which is composed by CNN architectures specially designed to detect objects in images. To detect target scenarios, two other approaches were developed and then compared. The first approach, called BoVW+SVM, combines the Bag of Visual Words (BoVW) technique with the Support Vector Machine (SVM) classifier. The second approach, called CNN_Scenarios, employs a CNN tiny-YOLOv3 model. In fact, both BoVW and YOLO have been widely used in recent works to detect objects and scenarios in images and can be considered as state-of-the-art techniques.

## 2. Background

### 2.1. Regulation of unmanned aerial vehicles in Brazil

The current regulation of UAV in Brazil is defined by the following organs: National Telecommunications Agency (ANATEL), National Civil Aviation Agency (ANAC), Department of Airspace Control (DECEA) and Ministry of Defense (MD) (ANAC, 2021). ANATEL is responsible for approving the remote-control devices and frequencies used to fly UAVs. In Brazil, most UAVs are already sold with ANATEL's homologation. However, in some cases, the approval needs to be requested via internet through the system called "Mosaico". The National Civil Aviation Agency (ANAC) is responsible for registering the UAV, which can be done through the "Unmanned Aircraft System - SISANT", being mandatory for any UAVs with a take-off weight greater than 250 g. This Agency is also responsible for regulating the use of UAVs, based on the Brazilian Civil Aviation Special Regulation (RBAC-E n° 94). This regulation was published in 2017 and is based on international practices and standards followed, for example, by the Federal Aviation Administration (United States), the Civil Aviation Safety Authority (Australia) and the European Aviation Safety Agency (European Union).

To meet requirements such as security and privacy, among others, the ANAC's regulations separate UAVs into 3 classes. The UAVs employed in this study belong to class 3, which includes equipment with a maximum take-off weight between 250 g and 25 kg. UAVs in this class do not need project approval for single flights, in which an aircraft can be seen by the pilot during the entire operation. There is also no need for a pilot's license or a license for flights below 120 m. Regardless of the

class, the horizontal distance between the UAV and anyone not involved in the operation cannot be less than 30 m. In addition, it is necessary to take out insurance covering damage to third parties and there is also a need to carry out an operational risk assessment (ANAC, 2021). It is worth mentioning that the flights carried out for this study took place in 2016, one year before ANAC's regulations were ratified. In any case, the flight operations were conducted in a way that preserved the image and privacy of all individuals who were not involved in the project, thus respecting the Brazilian Constitution and the Civil Code, in addition to meeting flight safety requirements.

The DECEA is responsible for registering pilots and issuing authorization for the use of Brazilian airspace. Authorizations must be requested on the internet through the "Request for Access to Remotely Piloted Aircraft - SARPAS" system. Depending on the region, the type of flight, the height of the flight and other factors, the request may take up to 18 days to be processed (DECEA, 2021). The DECEA regulation on "Remotely Piloted Aircraft Systems and Access to the Brazilian Air Space (ICA 100-40)" was published in 2016 and was respected throughout the acquisition of images employed in this work. Finally, the Ministry of Defense (MD) is the organ that regulates the aerophotogrammetric survey activities.

### 2.2. Brief description of the techniques employed in this work

#### 2.2.1. Convolutional neural networks

Convolutional neural networks (CNN), initially proposed by LeCun, Bottou, Bengio, and Haffner (1998), can be described as variations of a multilayer perceptron neural network, developed to demand the least possible pre-processing. This is because CNN has the ability to automatically extract features from patterns, a task that in a traditional pattern recognition method necessarily has to be implemented separately, and which represents one of the main problems of such methods (LeCun et al., 1998). This ability of CNN is one of the main advantages for its application in image analysis tasks (Albawi, Mohammed, & Al-Zawi, 2017; Ball, Anderson, & Chan Sr, 2017; LeCun et al., 1998).

Basically, a CNN consists of three sets of layers: convolutional layers, pooling layers and the fully-connected layers (LeCun et al., 1998). The convolutional layers are responsible for extracting the features from the images. They employ filters that trigger small regions along the entire image. Convolution can be interpreted as a mathematical operation between two functions, from which a third function is produced (Goodfellow, Bengio, & Courville, 2016). In digital image processing, in which the image can be defined as a two-dimensional function, convolution is useful for detecting features such as edges, corners, shapes and textures.

The pooling layers, which are placed between the convolution layers, perform spatial sampling operations using filters that are applied by the image. They produce lower resolution versions of the convolution layers and help make representations invariant to translations (Goodfellow et al., 2016).

Finally, the fully-connected layers work in a similar way to a multilayer perceptron neural network and act as a classifier. They receive input from the previous layer and produce an n-dimensional vector, where $n$ is the number of output classes. Thus, each vector element is used to indicate the probability that the input pattern belongs to that class (Ball et al., 2017; Goodfellow et al., 2016).

#### 2.2.2. Bag of visual words

Bag of Visual Words (BoVW) was introduced by Sivic and Zisserman (2003) for object matching in videos and has since become a popular framework for content-based image indexing and retrieval (CBIR). The BoVW framework has many variants in the literature, but all of them can be roughly divided into three basic steps: feature extraction, codebook construction and vector quantization.

The first step is responsible to perform feature extraction by extracting descriptors (feature vectors) from each image of a given dataset. After, in the second step, a vocabulary of possible visual words is constructed by clustering of the feature vectors obtained in the first step. The vocabulary (or codebook) construction is normally accomplished via the k-means clustering algorithm. Thus, the resulting cluster centers represent the dictionary of visual words. Finally, in the third step a histogram is built, having a length that corresponds to the number of clusters generated, with the frequencies of the visual words. This last step, known as coding or vector quantization, is responsible for representing an image by a single feature vector called bag of visual words (Sivic & Zisserman, 2003).

#### 2.2.3. Support vector machine

Support Vector Machine (SVM) is a supervised learning technique, derived from statistical learning theory, which can be used to solve both classification and regression problems (Cortes & Vapnik, 1995). The basic idea behind this technique is to find a hyperplane (i.e., decision surface) that maximizes the distance (i.e., the "margin") between classes of patterns. However, when the data is non-linearly separable, SVM uses a mapping system known as kernel (for example, polynomial, gaussian and sigmoid) to transform the data into a higher dimensional space where the data is linearly separable (Pisner & Schnyer, 2020).

The classification of a sample $z$ using SVM is given by the distance (called score) from $z$ to the decision boundary ranging from -∞ to +∞. Thus, a positive score indicates that $z$ is predicted for the positive class, while a negative score indicates it is predicted for the negative class. The SVM classifier can be represented by the function $f_{svm}$, defined in Eq. (1).

$$f_{svm}(z) = \sum_{i}^{M} \alpha_i \kappa(z_i, z) - b \tag{1}$$

where $\kappa$ is a kernel function, $M$ is the number of support vectors, $\alpha_i$ is the weights of the support vectors and $b$ is a bias. The support vectors and their weights can be obtained by the sequential minimal optimization - SMO algorithm, proposed by Platt (1998).

### 2.3. Related works

Remote sensing can be defined as the art, science and technology of acquiring information about objects and environments, through the recording, measurement, and interpretation of electromagnetic energy patterns, without the need for direct contact with them (Colwell, 1997).

Thanks to the latest technological innovations, and the evolution of sensors carried on board satellites, the applications of remote sensing in urban studies have been gaining more and more space in the academic sphere. However, at larger geographic scales, even the panchromatic images of WorldView 3 satellite, with 31 cm of ground sample distance (GSD), is still insufficient for the identification of small objects and complex visual patterns that are extremely important for understanding the context of some problems inherent to urban space (Grubesic, Wallace, Chamberlain, & Nelson, 2018), such as the problem addressed in the present study. To give a practical example, we can mention the water tanks we are considering as target object in this study. Taking into account that a typical water tank has on average 120 cm of diameter, it would be represented in an image acquired by the WorldView 3 by a set of approximately 16 pixels, which certainly would bring great difficulties to the automatic identification of such object, even considering modern techniques of machine learning. This problem would be intensified in the case of the target scenarios, for which it is necessary to recognize even smaller objects such as old tires, pet bottles, plastic and paper packaging, among others.

In the same way, temporal resolution also ends up limiting the applications of satellite images in some urban studies, since even WorldView 3 does not allow the acquisition of daily images for all places on the planet, whether due to the greater surface volume of low latitudes, whether due to its limited data collection, which ends up making it difficult to apply this technology in socioeconomic studies,

neighborhood audits and planning (Grubesic et al., 2018). In this case, the use of images from WorldView 3 for solving the problem addressed in this study could also lead to practical limitations, since the scenarios we are considering can easily suffer daily changes.

In recent years, Unmanned Aerial Vehicles (UAVs) have been used extensively as a remote sensing tool, promoting excellent spatial and temporal resolutions, as well as presenting more advantageous economic costs than traditional satellite remote sensing for many applications (Bhola, Krishna, Ramesh, Senthilnath, & Anand, 2018). UAVs also present at least three operational advantages in studies monitoring the physical environment at the neighborhood scale. First, the excellent spatial resolution of this technology allows the identification of small objects, which could not be identified by other remote sensing tools. Considering, for example, that is common images acquired by UAVs with GSD smaller than 5 cm, the same typical water tank with 120 cm of diameter mentioned above would be represented in such images by a set of hundreds or thousands of pixels, allowing its recognition even with the use of simpler techniques of machine learning. Second, UAVs allow to obtain images from multiple angles, different altitudes and at different time points. Third, the use of UAVs limits the exposure of auditors, or agents responsible for data collection, to dangerous situations or places, thus increasing security during data collection activities at neighborhood scale (Grubesic et al., 2018). The reported advantages, as well as the explanations given on the recognition of target objects and scenarios we are considering, justify the use of UAVs to deal with the problem addressed in this study. It should be emphasized that high-resolution imagery provided by satellite, such as WorldView 3, can represent an important alternative for detecting and recognizing many other types of objects and scenarios, especially at smaller geographic scales, so they should be considered whenever possible.

Although UAVs offer the advantages mentioned above, the automatic detection of objects and scenarios in images acquired by such equipment still represents a great challenge, due to the amount of details present in these images, especially those acquired in urban areas.

In this regard, Xu, Yu, Wang, Wu, and Ma (2017) proposed a framework for detecting cars with UAVs using a CNN called Faster R-CNN (Region Convolutional Neural Network), in low altitude images which were acquired at signaled junctions. Ammour et al. (2017) also developed an approach for car detection and counting using a CNN combined with the SVM classifier. To support search and rescue operations in regions with avalanche risk, Bejiga, Zeggada, Nouffidj, and Melgani (2017) developed an approach to extract descriptors of debris images from these regions and to detect objects of interest such as skis or possible victims using CNN and SVM classifiers.

The algorithms used for the detection of objects in the aforementioned studies are based on the concept of Early Deep Learning. Among them, we highlight the R-CNN and Faster R-CNN, which employ a method called Selective Search, which aims to reduce the number of bounding boxes that the algorithm has to test by hierarchical grouping, from similar regions of the image and based on the compatibility of color, texture, size and shape. Even with such reduction, the operations performed to classify the objects using these approaches can be very expensive in computational terms.

YOLO is a framework composed of CNNs specially designed for object detection. The "YOLO - You Only Look Once", proposed by Redmon et al. (2016), has this denomination because it refers to the fact that the CNNs implemented in the framework process the entire image only once at the same time, generating the predictions of the objects. Recently, Redmon and Farhadi (2018) developed a new version of YOLO, called YOLOv3, whose architectures of the CNNs make up the framework capable to recognize 80 different objects in images and videos, in real time.

According to Redmon and Farhadi (2018), in their experiments YOLOv3 outperformed popular and robust methods such as Faster R-CNN with Residual Neural Network (ResNet) developed by He, Zhang, Ren, and Sun (2016) and the Single Shot Multibox Detector (SSD) proposed by Liu et al. (2016), presenting competitive results and being faster. Similarly, Yi, Yongliang, and Jun (2019) developed an approach for pedestrian detection using the tiny-YOLOv3 (a reduced version of the YOLOv3 structure with only 9 layers), in conjunction with the K-means algorithm to filter out the best features of the training set. Benjdira, Khursheed, Koubaa, Ammar, and Ouni (2019) conducted a study comparing the results obtained by Faster R-CNN and YOLOv3 in detection of cars using UAVs and showed that YOLOv3 outperformed Faster R-CNN in terms of accuracy and processing time.

Tian et al. (2019) and Xu, Jia, Sun, Liu, and Cui (2020) presented approaches using YOLOv3 for fruits detection. The former tested real-time detection of apples in orchards, in order to evaluate the growth phases of apples and hence estimate the yield. The authors showed that the proposed YOLOv3-dense (modified version) model is superior to the original YOLO-v3 model and to the R-CNN. Xu et al. (2020) proposed a method based on YOLOv3 for detection of lightweight green mangoes. They explored different architectures of YOLOv3 and proposed mechanisms to improve the speed of the method as well as its robustness in relation to occlusion and invariance to changes in scale and brightness.

Diniz and Medeiros (2018) and Passos et al. (2018) addressed the mapping of the target objects considered in the present study from images acquired by UAVs. Instead of presenting an automatic method, Diniz and Medeiros (2018) conducted the mapping manually using a Geographical Information Systems (GIS) software. Passos et al. (2018) presented the steps for composing a database of annotated images that could be used to evaluate methods for automatic detection of suspicious objects (tires, bottles and other objects that can accumulate water). However, such database is not openly accessible for further assessments.

Regarding the detection of "scenarios", we can mention the works of Hardy et al. (2017), Agarwal et al. (2014), Mehra et al. (2016), Carrasco-Escobar et al. (2019) and Haas-Stapleton et al. (2019). Hardy et al. (2017), for example, explored the use of a low-cost UAV (DJI Phantom 3, which was also used in our study) to map water bodies in seven sites across the Zanzibar archipelago including natural water bodies, irrigated and non-irrigated rice paddies, peri-urban and urban locations, aiming to identify and map aquatic mosquito habitats. However, the authors performed manual interpretation of the images, with the support of the QGIS software, to delineate water body location and size. In the other four works, automatic approaches were proposed.

Agarwal et al. (2014) developed a method to detect and visualize possible mosquito breeding sites. The proposed method comprised three steps: evaluation of the quality of the images; image classification using the technique Bag of Visual Words (BoVW) combined with the SVM classifier, taking into account the descriptor Scale Invariant Feature Transform (SIFT); and the visualization of breeding sites from heat maps, which indicate the regions with the highest risk of mosquito outbreaks. In experiments involving the classification of 500 images, an accuracy of around 82% was obtained. In Mehra et al. (2016) a framework was proposed to detect possible mosquito breeding sites using images from Google and various devices (digital cameras, smartphones and UAVs). For the extraction of features, the BoVW technique was used with the descriptor Speed-Up Robust Features (SURF) and the classification was performed by Bayesian classifiers. In the experiments carried out, the authors obtained an accuracy of 90%.

Although they provide a useful framework, the approaches proposed by Agarwal et al. (2014) and Mehra et al. (2016) only indicate if an image contains a suspicious scenario, without providing their actual spatial location. This makes them unsuitable for applications that require the precise indication of the location of the possible breeding sites.

Carrasco-Escobar et al. (2019) and Haas-Stapleton et al. (2019) investigated the detection of mosquito breeding sites based on the analysis of water bodies' characteristics. Carrasco-Escobar et al. (2019) proposed the use of UAVs to identify Nyssorhynchus darlingi (formerly *Anopheles darlingi*) breeding sites. They developed a method capable of analyzing high-resolution multispectral imagery to determine the

profiles of water bodies where Ny. darlingi is most likely to breed, achieving accuracy rates between 87% and 97%. Haas-Stapleton et al. (2019) proposed the use of UAVs to quantify the accumulated surface water on a 0.54-km$^2$ tidal marsh that abuts San Francisco Bay (USA). The study aimed to provide information for spatially focused inspections of potential mosquito breeding sites and to identify areas where existing ditches needed improvements to ensure water drainage. They also conducted experiments using a CNN-AlexNet model aiming to identify immature mosquitoes in 2 small containers of contrasting colors during simulation tests in a marsh habitat, in which they obtained accuracies varying from 52.8% to 94.1%.

Despite all these technological advancements, there is still a lack of approaches capable of providing accurate locations of predefined objects and scenarios from aerial imagery provided by UAVs. In addition, although CNN is a technique widely applied in object recognition, there are no studies in the literature addressing its use in the identification of mosquito breeding sites taking into account the objects and scenarios considered in the present study. Therefore, we propose different approaches that address these technological gaps, and that can increase the efficiency of the use of UAVs to combat mosquito breeding sites.

## 3. Materials and methods

### 3.1. Image datasets

To conduct the experiments, we composed four datasets with images acquired in different areas located in the state of São Paulo - Brazil, as indicated in Fig. 1. These datasets, named as DS1, DS2, DS3 and DS4 are detailed below. The images from DS1 and DS2 contain many types of water tanks that store water for domestic use, which represent the target objects. These tanks are sometimes uncovered, and thus are potential mosquito breeding sites. The entire flight plannings for acquisition of the images that compose DS1 and DS2 were carried out with the Map Pilot for DJI app. The flight plan was programmed to ensure an overlap of between 60 and 70% between each pair of neighboring images. Each site was surveyed in a single flight that took less than 10 min and did not require changing battery.

The images from DS3 and DS4 contain real and simulated suspicious scenarios. To generate the simulated scenarios, we placed on the ground portions of inorganic waste (small objects that can retain stagnant water such as old tires, pet bottles, plastic and paper containers, among others) and made the flyover of them to acquire the images. Each combination of small objects (portion) generates an image. For the acquisition of images from real and simulated scenarios in DS3 and DS4, the UAV was operated manually to get the UAV closer to the targets. Examples of images from the four datasets are shown in Fig. 2.

To compose DS1, 142 images of 4000 × 3000 pixels were acquired in peripheral regions (São Paulo's district of Guaianases and the municipality of Ferraz de Vasconcelos), using a DJI Phantom 3 UAV equipped with a Sony EXMOR 12.4 MP RGB camera. The UAV flew over an area of 66,311.98 m$^2$, in 8 m 41 s, with an altitude of 50 m above ground level which provided a ground sample distance (GSD) of ~2.2 cm/px. The DS2 consists of 101 images of 4000 × 3000 pixels acquired in an informal settlement named "Porto de Areia", located in the municipality of Carapicuíba. The images were acquired by the same DJI Phantom 3 UAV mentioned above, which flew over an area of 91,344.48 m$^2$, in 6 m 52 s, with an altitude of 70 metes above ground level, providing a GSD of ~2.45 cm/px.

The DS3 is composed by 119 RGB images of simulated scenarios acquired in a vacant site near to the Polytechnic School of the University of São Paulo (USP). These images have a resolution of 4000 × 3000 pixels and were acquired using a DJI Phantom 4 UAV equipped with a DJI 20 MP RGB camera. For the flights, three distances (7, 10 and 13 m) above the ground were considered according to the scenario. These distances provided values of GSD ranging from 0.30 to 0.56 cm/px. The DS4 contains 111 images (resolution of 3000 × 2250 pixels) of real scenarios that were observed in Guaianases and Ferraz de Vasconcelos, using the same DJI Phantom 4 UAV mentioned above, equipped with a GoPro HERO4 Silver camera. The flying altitude ranged from 3 to 5 m above the ground level, providing GSD values below 0.30 cm/px. For

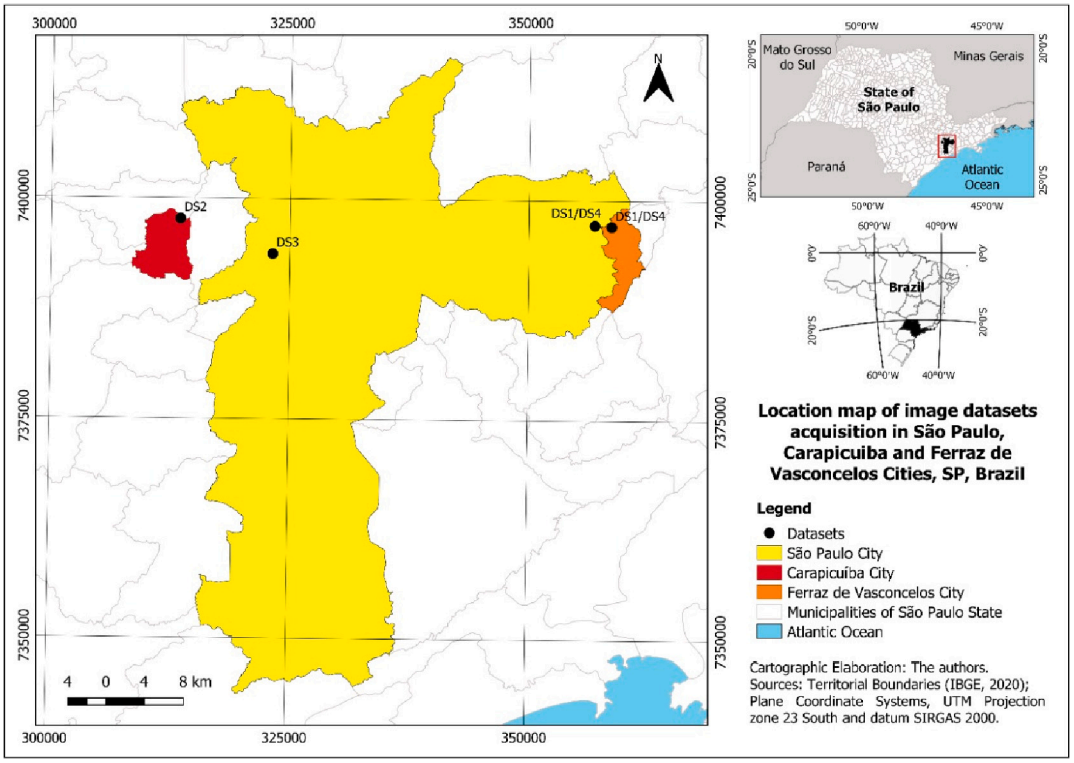
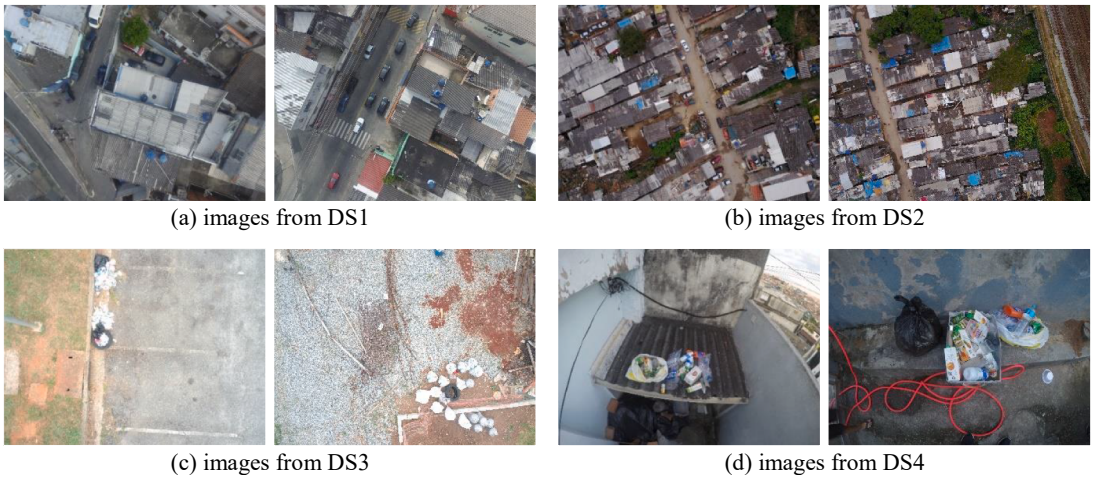



**Fig. 1.** Location map of image datasets acquisition.



(a) images from DS1 (b) images from DS2

(c) images from DS3 (d) images from DS4

**Fig. 2.** Examples of images from the four datasets.

this last dataset we employed the GoPro camera because the original UAV camera presented signs of malfunction. This change eventually allowed us to test the proposed approaches in images with different resolutions.

The 243 images from DS1 and DS2 were used for building and evaluating the approach to detect objects, while the 230 images from DS3 and DS4 were employed for building and evaluating the approaches to detect scenarios. These images were divided into 2 parts: 70% for training and 30% for evaluating the developed approaches.

As mentioned above, although the procedures for acquiring the images took place in 2016 (before ANAC's 2017 regulations were in place), we sought to respect the principles of the Federal Constitution of Brazil, as well as the Brazilian Civil Code. Flight safety requirements were duly respected: we selected sites far from sensitive areas such as airports or prisons. Also, the entire data collection process was conducted respecting people's privacy, in a way that it is not possible to identify people or vehicles in the images composing the four datasets. Moreover, before flying over inhabited areas – such as the settlement Porto de Areia in Carapicuiba, Guaianases and Ferraz de Vasconcelos – we obtained verbal consent from the local inhabitants. These precautions are in accordance with the ethical aspects mentioned by Nelson and Gorichanaz (2019), namely, privacy, security, enforceability, crime, nuisance and professionalism, which must be considered for an emerging technology to be accepted and fully integrated into society.

Going through the datasets, we noted that images contemplating the predetermined objects and scenarios that represent possible mosquito breeding sites are actually rare. This challenge was also reported in previous studies, such as in Passos et al. (2018). Therefore, the image database composed throughout the present study can be considered as an important scientific contribution by itself, since it will be made available upon request to allow other researchers to test their own detection methods.

### 3.2. Detection of objects

The approach described here is focused on the detection of target objects (typical containers for storing water for domestic use, or simply "water tanks"), and employs a CNN architecture from the YOLOv3 framework (Redmon & Farhadi, 2018). This approach was named CNN_Objects, and is illustrated in Fig. 3. The CNN architecture consisted of 106 layers, from which 75 were convolutional layers and 28 were upsampling layers and residual block layers. The layers 82, 94 and 106 were adjusted for detection objects on 3 different scales.

It is important to mention that the characteristics of the water tanks depend on their type and on the year they were manufactured. In addition, the material used for the manufacture of old tanks (asbestos fiber cement) differs a lot from the most recent tanks (plastic polyethylene). Based on the collected images (see Fig. 4) and on the different types of water tanks existing in urban areas in Brazil, these objects were grouped into 7 distinct classes, named as follows: wtank_type1, wtank_type2, wtank_type3, wtank_type4, wtank_type5, wtank_type6 and wtank_type7. Thus, each group included tanks with similar characteristics of color, shape and cover pattern.

The training dataset consisted of 690 sub-images (with several dimensions) manually extracted from 170 images belonging to DS1 and DS2. The 73 remaining images from these two datasets were employed to evaluate the performance of CNN_Objects. In other words, we divided the images from DS1 and DS2 into two parts: 70% for training and 30% for testing (performance evaluation). This division was chosen because a 70–30 ratio is often used in the literature regarding pattern classification tasks. It is important to point out that the data augmentation scheme provided by the framework YOLOv3 was employed to improve the CNN generalization capacity. Thus, at each iteration during the training of the CNN the number of samples is increased automatically by applying different transformations such as zooming, rotation, flipping and noise addition, among others.

The following configuration parameters were adopted for training CNN_Objects: number of batches = 64; subdivisions = 32; maximum number of iterations = 14,000 (2000 * number of classes); and learning rate = 0.001. The parameter adopted to interrupt the training was based on the lack of improvement ("loss of validation"), which is activated if the model runs more than a certain number (10 in our experiments) of epochs without improving the loss. After 288 h, the training was interrupted at the iteration 13,950, with a loss of validation of 0.19. During the training, 1,020,864 sub-images were generated by the data augmentation scheme.

### 3.3. Detection of scenarios

#### 3.3.1. BoVW+SVM

The BoVW+SVM approach (illustrated in Fig. 5) was based on the works of Agarwal et al. (2014) and Mehra et al. (2016), and comprised the following steps:

- **Step 1.** Features extraction from sub-images (windows) belonging to the training images. Unlike the works of Agarwal et al. (2014) and Mehra et al. (2016), which consider only one descriptor (SIFT or SURF), we considered features extracted by several descriptors: Color Histograms (CH), Color Level Co-occurrence Matrix (CLCM), Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). These descriptors were isolated and combined as follows: CH, CLCM, HOG, LBP, CH + HOG, CH + LBP, CLCM+HOG, HOG+LBP,

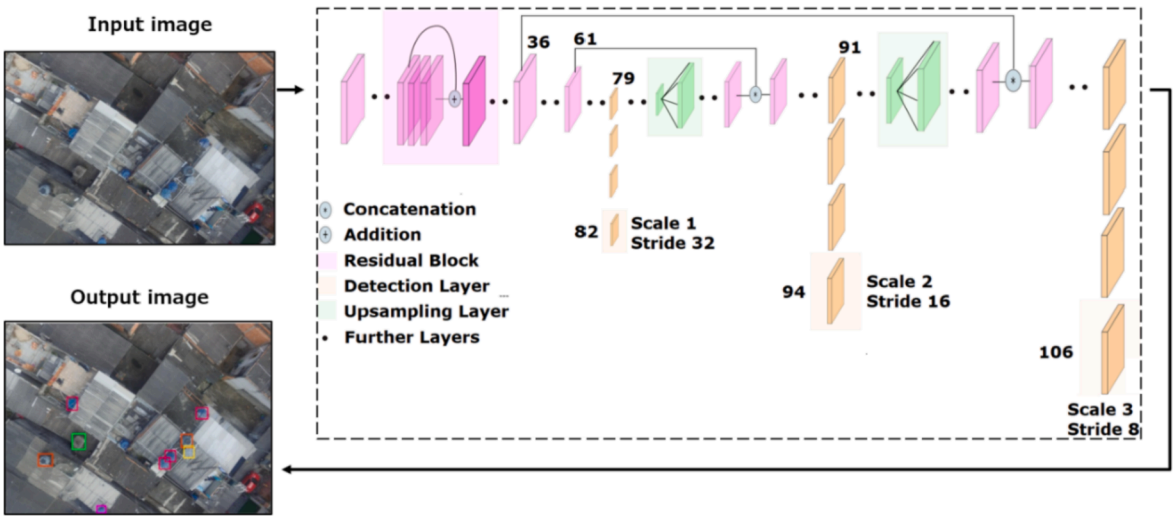



**Fig. 3.** CNN architecture employed to detect target objects (CNN_ Objects). Adapted from Singh et al. (2021)


**Fig. 4.** Types of water tanks considered in this study.

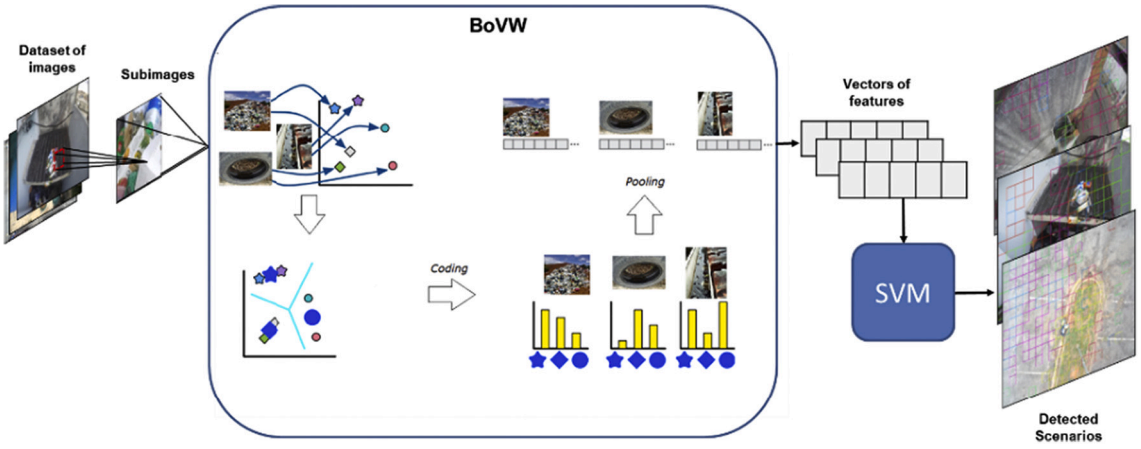



**Fig. 5.** Working of the BoVW+SVM approach.

CH + CLCM+LBP, CH + HOG+LBP, CH + CLCM+HOG+LBP. According to Kim and Araujo (2011), in images containing many details (which is the case of our images), a large amount of inconsistent keypoints can be detected by SIFT and SURF, negatively affecting their performances. For this reason, we did not consider these descriptors in the experiments with BoVW+SVM.

From the CLCM, we computed six Haralick descriptors (features): second angular moment, entropy, contrast, variance, correlation and homogeneity (Haralick et al., 1973). The CH generated 384 features (128 bins for each color channel of the RGB space). From HOG and LBP were extracted, respectively, 20,736 features (considering 8 × 8 cells) and 2124 features (considering 32 × 32 cells). Both CH and LBP were calculated for the three-color channels (Red, Green and Blue) separately.

- **Step 2.** Creation of the visual words dictionary (codebook generation) from the extracted features using the K-means algorithm. The dictionary size (K) was empirically defined as 100.
- **Step 3.** Representation of each window of 200 × 200 pixels extracted from images separated for training by means of a histogram computed from the visual words dictionary. It is worth mentioning that the window size (200 × 200) was defined empirically, based on preliminary experiments.
- **Step 4.** Synthesis of visual words histograms into new vectors of features (quantization or pooling).
- **Step 5.** Training SVM classifier using the new vectors of features belonging to each training set (each combination of descriptors generates a different training set).
- **Step 6.** Classification of the windows belonging to each image separated for test using the trained SVM classifier.

Considering the diversity of scenarios due to the different sizes, shapes, textures and colors of the clustered objects, 9 classes were defined: closed garbage bags (scen_type1), garbage containing old tires (scen_type2), garbage containing paper and small packages (scen_type3), garbage with small containers that can accumulate water (scen_type4), garbage contained in buckets (scen_type5), garbage contained in garbage containers (scen_type6), closed garbage bags mixed with old tires (scen_type7), medium containers with garbage inside (scen_type8) and garbage consisting of other materials (scen_type9).

From the 161 images separated for training, features of 900 sub-images (100 of each class) were extracted to compose the training sets. Regarding the SVM, a cross-validation procedure was carried out aiming to improve the hyperparameters for each training set. As we mentioned previously, the 230 images belonging to DS3 and DS4 were divided into 2 parts: 161 images for training (70%) and 69 images for testing (30%).

#### 3.3.2. CNN_Scenarios

For the identification of scenarios, we employed a simplified model of YOLOv3 that reduces the depth of the convolutional layer, known as tiny-YOLOv3 (see Fig. 6). Obviously, this more compact architecture leads to a significant reduction in the processing time required to process an image. Indeed, the time observed in the training of YOLOv3 for target objects recognition was the main motivation for choosing this model to identify scenarios.

The training dataset was composed of 1430 sub-images (with several dimensions) manually extracted from the 161 training images, considering all classes of scenarios described above. The data augmentation scheme provided by the framework YOLOv3 was applied to improve the CNN_Scenarios generalization capacity.

The following training parameters were employed: number of batches = 64; number of subdivisions = 32; maximum number of iterations = 18,000; and learning rate = 0.001. These parameters were based on the recommendations of Redmon et al. (2016) and Redmon and Farhadi (2018). After 96 h, the training was interrupted with a validation loss value of 0.06, at the iteration 17,930, with about 1,147,520 samples generated by the data augmentation scheme.

### 3.4. Experimental setup

The algorithms described in this work were developed in C/C++ language using the OpenCV[1] and Darknet[2] libraries, and in the Matlab 2018 environment, as presented in Table 1. It is worth mentioning that, for the experiments using the Darknet library, the CUDA (Compute Unified Device Architecture) platform was used to access GPU (Graphics Processing Unit) resources.

The computational experiments were conducted on a PC with 2.5GHz Core i7, 16 GB of RAM, equipped with NVIDIA GeForce 930 M GPU, and Windows 10 Pro operating system. The performances of the approaches were evaluated in terms of the measure mAP-50 (mean average precision), using the images purposively selected for such tests – which were not used in the training phase. In this way, 73 images from DS1 and DS2 were used to evaluate the CNN_Objects and 69 images from DS3 and DS4 were used to evaluate BoVW+SVM and CNN_scenarios. The mAP measure (described in Eq. (2)) was chosen based on previous studies addressing applications of CNN in object detection, including the works of Redmon et al. (2016) and Redmon and Farhadi (2018). Indeed, it is an appropriated measure to express the precision in the location of detected objects.

$$mAP-50 = \frac{1}{nc}\sum_{i=1}^{nc} AP_i \tag{2}$$

where $nc$ is the number of classes and $AP_i$ is the average precision of each class $i$, calculated by Eq. (3), which depends on the precision ($P$) computed for each bounding box (window) $j$ predicted as class $i$ (Eq. (4)) that, in turn, takes into account the value resulting from the $IoU$ operation (Eq. (5)).

$$AP_i = \frac{1}{nd}\sum_{j=1}^{nd} P_{ij} \tag{3}$$

$$P = \frac{TP}{TP + FP} \tag{4}$$

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{5}$$

where $nd$ is the number bounding boxes predicted as class $i$, $TP$ is the number of true positive cases and $FP$ is the number of false positive cases.

The $IOU$ operation computes a ratio from the area of intersection and area of union of the predicted bounding box and the ground truth bounding box (Xia, Ye, Yan, Feng, & Tian, 2020). The numerator is the area of overlap between the predicted bounding box and the ground truth bounding box; the denominator is the area of union between the predicted bounding box and the ground truth bounding box. In the case of mAP-50, each detection is considered as $VP$ if the object or scenario predicted as class $i$ effectively belongs to the class $i$ and if the value resulting from the $IOU$ operation is greater than or equal to 0.5.

Finally, since BoVW+SVM considers different combinations of features descriptors, the comparison with CNN_scenarios was made considering only the combination of descriptors that provided the best result.

## 4. Results

### 4.1. Detection of objects

To evaluate the accuracy of detection of the water tanks (target objects), the 73 test images were submitted to the classification task, which took less than 1 min. After the classification of the images, the AP was computed for each object class (wtank_type*) and, from the AP values, the mAP-50 was calculated. These values are showed in the Table 2.

From the total of 152 ground-truth bounding boxes, 140 were correctly classified. In Fig. 7b it is possible to see that all the water tanks were detected correctly (cases of True Positive - TP). In total, we counted 16 cases of False Positive (FP) and 12 cases of False Negatives (FN). In Fig. 7d it is possible to identify a case of FP, which is highlighted with a red circle and three cases of FN (Fig. 7f), which are highlighted with white circles. The mentioned case of FP occurred probably because the detected object resembles the circular shape and the colors of some types of water tanks. Based on the results of the detection of the water tanks, it was possible to calculate the recall rate (0.9211).

As one can see in Table 2, the performance of YOLOv3 was excellent in the detection of target objects. Its worst performance (0.9040) occurred in the detection of wtank_type5, probably due to the low occurrence of this type of water tank in the images, that may have hindered the learning of the CNN. On the other hand, even in situations where water tanks were too close to each other, the CNN_Objects was able to detect all of them in most cases.

### 4.2. Detection of scenarios

The performance evaluation of the two developed approaches was based on the classification of the 69 images destined for the tests. As in

[1] OpenCV (Open Source Computer Vision Library) – https://www.opencv.org/

[2] https://github.com/AlexeyAB/darknet

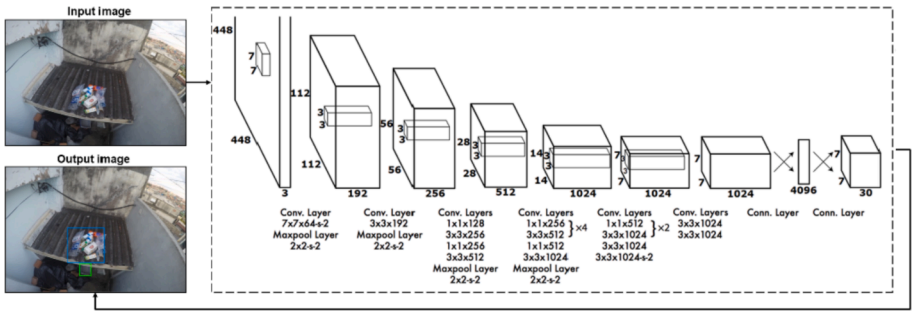



**Fig. 6.** Architecture of CNN used to detect scenarios (CNN_Scenarios). Adapted from Redmon et al. (2016).

**Table 1**
Computational environments and sofwares.

| Approaches | Computational environment |
|---|---|
| Detection of objects and scenarios using YOLOv3 | C/C++ language employing Microsoft® Visual Studio™ 2015 IDE, OpenCV (image processing and computer vision routines) and Darknet (YOLOv3 framework) |
| Detection of scenarios using BoVW+SVM | Matlab 2018 |

**Table 2**
Average Precision in the detection of target objects.

| Class | AP (Average Precision) |
|---|---|
| wtank_type1 | 0.9734 |
| wtank_type2 | 0.9933 |
| wtank_type3 | 0.9763 |
| wtank_type4 | 1.0000 |
| wtank_type5 | 0.9040 |
| wtank_type6 | 0.9091 |
| wtank_type7 | 1.0000 |
| **mAP-50** | **0.9651** |

the case of objects detection, the AP was computed for each scenario class (scen_type*). Then, the mAP-50 was calculated from the AP values.

The classification using BoVW+SVM was based on the sliding window strategy, in which each sub-image of 200 × 200 pixels extracted from a test image was classified. Empirically, a threshold value of 0.60 was defined for the posterior probability, which was calculated for the predicted class. Thus, only the bounding boxes classified with a probability value equal to or greater than this threshold were considered. The results of this classification for each combination of descriptors (described in section 3.3.1) are presented in Table 3.

From Table 3, we see that the highest value of mAP-50 (0.6453) was obtained with the combination CH + LBP. From the classified windows, 1638 were TP cases and 2352 FP cases, resulting in a median value of mAP-50. In Fig. 8b and d it is possible to observe some classified windows representing TP (colored rectangles) and FP (colored rectangles indicated by red arrows).

The CLCM descriptor computes the occurrences of local transitions between color channels. Thus, considering that each scenario could contain a variety of objects with different colors, the calculation of local transitions undermined more than helped in the detection of scenarios. An analogous reasoning can be done for the HOG descriptor, but this time regarding the variety of shapes and textures of the objects. These may be the reasons for the low performance of these two descriptors. With respect to CH, unlike CLCM it reflects average values of color intervals in the RGB space and therefore presented a better performance than CLCM. Regarding the LBP, the fact it was computed for each color channel may have contributed to its relatively high performance.

The high number of FP indicates that BoVW+SVM, using the adopted combinations of descriptors, is not the most appropriate approach for solving the problem investigated in this study. Another negative aspect of this approach is the processing time, since the classification is made for each window extracted from the image. For instance, the classification of the test images required several hours of processing in the Matlab environment.

To classify the same 69 test images, CNN_Scenarios took only 18 s. From the 111 bounding boxes defined as ground truth, 96 were correctly classified (TP cases) providing a mAP-50 value of 0.9028 and a recall rate of 0.8649. Some examples of TP cases are illustrated in Fig. 9b. In addition, 15 cases of false negatives (FN) and 11 cases of FP were computed (see examples in Fig. 9d, indicated by red circles).

The results presented in Table 4 demonstrate the good performance of CNN_ Scenarios. Its worst performance (0.7915) occurred in the detection of scen_type8, probably due to the low occurrence of this type of scenario in the training images. Even so, this low AP value was only lower than one of the results obtained by BoVW+SVM (for scen_type6).

As expected, the CNN with larger architecture (YOLOv3) achieved better results in terms of inference, while tiny-YOLOv3 had a lower performance, but reaching higher speeds, due to their reduced architecture. However, our experiments demonstrate that both models can be employed with the objective of obtaining a system that operates in real time.

It is important to mention that the results obtained by the approaches presented in this study were not compared with the results reported in Agarwal et al. (2014) and Mehra et al. (2016) for two main reasons:

- (i) different objectives - while the approaches proposed in Agarwal et al. (2014) and Mehra et al. (2016) only indicate if an image contains a suspicious scenario, BoVW+SVM and CNN_Scenarios locate and typify the suspicious scenarios in the images. In the case of BoVW+SVM, the analyzed image is divided into sub-images of M × M pixels (in our experiments M = 200), with each sub-image being classified independently, making it possible to locate the scenario in the image. Thus, in an analyzed image there may be many cases of false positives leading to performance degradation.
- (ii) the approaches proposed by Agarwal et al. (2014) and Mehra et al. (2016) were evaluated with databases of images that are not openly accessible. Thus, even if the aim of such approaches was the same as BoVW+SVM and CNN_Scenarios, the comparison with different database of images would not be fair. This emphasizes the importance of making available the datasets of images composed in this study.

Finally, although there is no indication of geolocation for each of the objects and scenarios detected in the analyzed images, the images themselves are actually georeferenced, thanks to the geolocation information recorded automatically by the UAVs on-board GPS. Thus, images acquired in the UAV missions can be imported into a stand-alone software (e.g. Agisoft Photoscan) that performs photogrammetric processing of digital images, extracts georeferenced point clouds, produces digital orthoimage mosaic, and generates 3D spatial data. Orthomosaics


**Fig. 7.** Some results provided by CNN_Objects.

**Table 3**
Results of BoVW+SVM in the detection of suspected scenarios.

| Descriptor(s) | mAP-50 |
|---|---|
| CH | 0.6225 |
| CLCM | 0.4117 |
| HOG | 0.4173 |
| LBP | 0.5019 |
| CH + HOG | 0.5787 |
| **CH + LBP** | **0.6453** |
| CLCM+HOG | 0.4256 |
| HOG+LBP | 0.4443 |
| CH + CLCM+ LBP | 0.6353 |
| CH + HOG+LBP | 0.5902 |
| CH + CLCM+HOG+LBP | 0.5740 |

can further be imported into a geographic information system (e.g. QGIS) for geoprocessing and mapping. It is important to mention that no ground control points were used in the missions we conduct with UAVs for image acquisition. However, the accuracy obtained in the geolocation of the acquired images (less than 1.8 m) was considered satisfactory for the objectives of this study and did not affect negatively the results obtained.

Failing to automatically indicate the geolocation of each object or scenario detected in the analyzed images and not identifying the existence of stagnant water in the objects or scenes are the main limitations of the present study and will be addressed in our future research.

## 5. Conclusions and future research

In this study we proposed different approaches for detecting objects and scenarios likely to become mosquito breeding sites from aerial images acquired by UAVs. The results obtained in the detection of objects (mAP-50 = 0.9651) and scenarios (mAP-50 = 0.9028) indicate that the use of YOLOv3 framework is a good alternative for solving these tasks, especially if we consider the speed and precision provided by the employed CNNs architectures. In addition, these approaches are able not only to detect and locate the objects and scenarios of interest, but also to distinguish very well between each of the classes that were considered in this work.

The methods presented here offer a great potential to improve vector control campaigns and to reduce transmission of mosquito-borne diseases. Indeed, locating and quantifying potential mosquito breeding sites is the first step for the prevention of mosquito-borne diseases through larval control. However, current detection methods are often time-consuming and suffer from inaccuracies related to currently

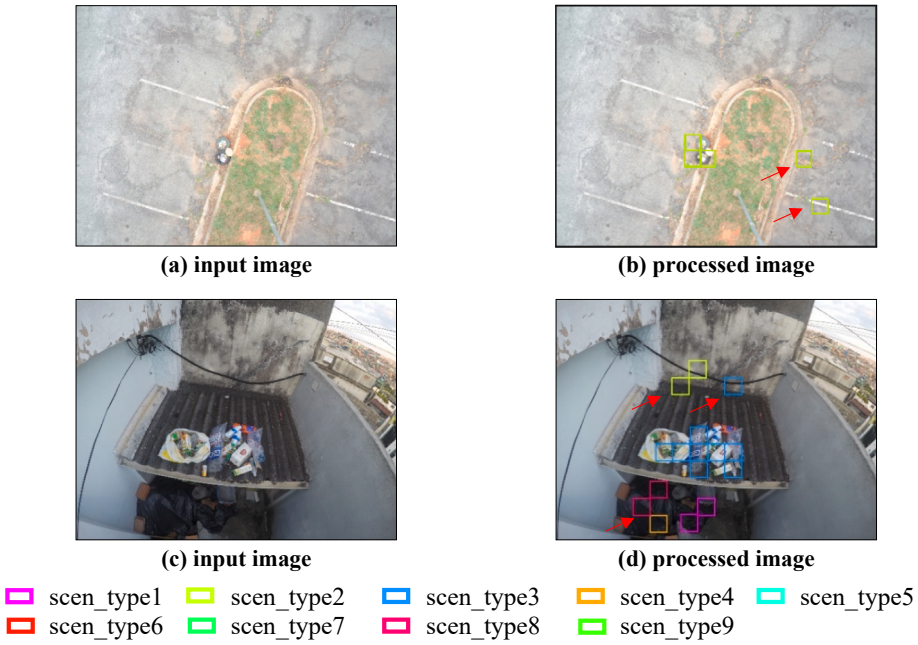



**Fig. 8.** Some results provided by BoVW+SVM.

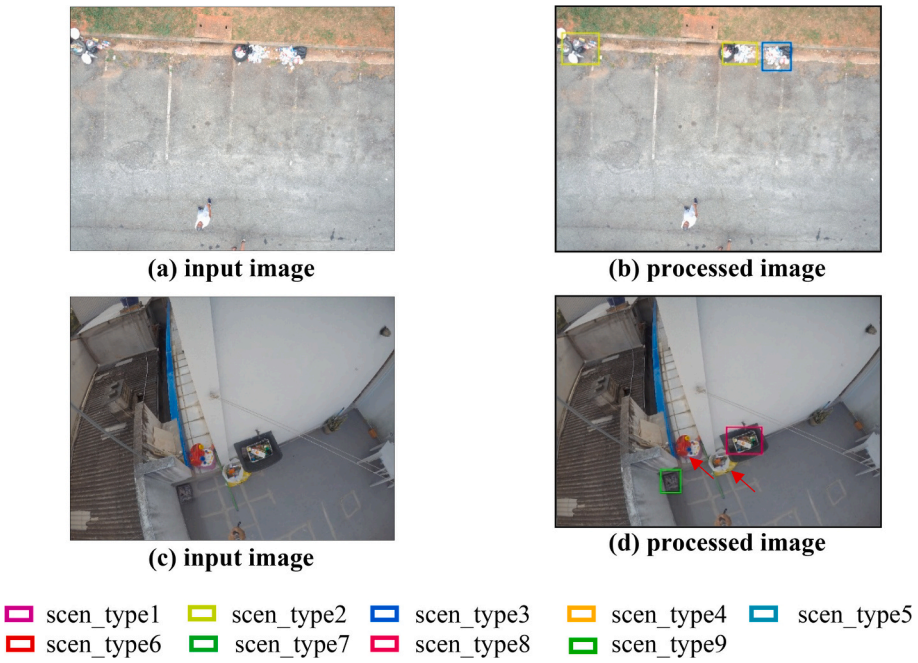



**Fig. 9.** Some results provided by CNN_ Scenarios.

available geographic information. The automated recognition of breeding sites through UAV imagery tackles both aforementioned issues. The approaches presented in this study could easily compose a low-cost software with the capacity to run rapid analysis to locate mosquito breeding sites, in such a way that time and resources are optimized. Especially in poor peri-urban areas of low-and middle-income countries, where essential municipal services (waste collection, water and sanitation infrastructures) are often scarce or even inexistent, such approaches can be an invaluable tool for public health workers to carry out targeted interventions, including sensitization campaigns.

Certainly, the scalability of the methods presented here can be an issue due to limitations in terms of the geographic extent that can be covered by a single UAV. However, recent examples show that it is indeed possible to cover an area as large as the whole Zanzibar Island (see Hardy et al., 2017). The objective of this ambitious initiative is to provide updated, open-access data to support spatial development in Zanzibar led by different sectors of society. Also, local administrations, such as in the São Paulo's city government, are already using UAVs in combating mosquito breeding sites, and as a tool to support local governance and decision-making.

The low-cost of such UAV-based methods allow for constant monitoring and provision of data with unprecedented spatial and temporal resolutions that are crucial for public health and urban planning. For instance, these enhancements in data quality allow for more accurate

**Table 4**
Average Precision (AP) in the detection of suspected scenarios.

| Class | Average Precision (AP) | |
|---|---|---|
| | BoVW+SVM | CNN_ Scenarios |
| scen_type1 | 0.6509 | 0.8182 |
| scen_type2 | 0.5710 | 0.9860 |
| scen_type3 | 0.6544 | 0.8098 |
| scen_type4 | 0.6700 | 0.9924 |
| scen_type5 | 0.4561 | 1.0000 |
| scen_type6 | 0.8511 | 0.8182 |
| scen_type7 | 0.7530 | 1.0000 |
| scen_type8 | 0.5502 | 0.7915 |
| scen_type9 | 0.6510 | 0.9091 |
| **mAP-50** | **0.6453** | **0.9028** |

spatial analyses aiming to understand larger impacts of the physical environment on the risk of mosquito-borne diseases, and on public health in general. In fact, several aspects of settlement morphology (e.g. entropy levels or density) could be explored as predictors of different diseases. From a transdisciplinary perspective, the approaches highlighted in this study could considerably help increasing community involvement in vector control programs based on the reduction of mosquito-breeding sources in the environment. In fact, high-resolution geographic information obtained with UAVs could be used to support community sensitization, and therefore encourage citizen participation in the reduction of mosquito populations through bottom-up action plans.

The images acquired by UAVs are more appropriate than high-resolution images obtained by on-board sensors of satellites to deal with the approaches proposed here since, as already mentioned above, even the panchromatic images of WorldView 3 (with 31 cm of GSD) do not allow to distinguish, at larger geographic scales, the small objects considered in this study. The same limitation was observed by Grubesic et al. (2018), who compared the pachromatic images of the Landsat 8 (15 m of GSD) with the images obtained by an Ebee (senseFly fixed-wing UAV) in two neighborhoods of the city of Phoenix - Arizona. The authors highlighted that the images from that satellite appear pixelated, making it difficult or preventing the distinction of individual elements of the environment, in larger geographic scales.

In future research we intend to o unify objects and scenarios detection in a single CNN architecture that will compose a system for operating in real time and able to automatically provide the geolocation of each object or scene detected in the analyzed images. In addition, we intend to develop an approach allowing the identification of objects and scenarios containing stagnant water, a necessary condition for the reproduction of mosquitoes, in order to increase the public health applicability of the developed approach.

## Acknowledgments

This work was supported by the FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo (Process 2019/05748-0), and by the CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico (research scholarship granted to S. A. Araújo, Process 313765/2019-7).

## References

Agarwal, A., Chaudhuri, U., Chaudhuri, S., & Seetharaman, G. (2014). Detection of potential mosquito breeding sites based on community sourced geotagged images. In *Geospatial infofusion and video analytics IV and motion imagery for isr and situational awareness II* (p. 90890M). https://doi.org/10.1117/12.2058121.

Albawi, S., Mohammed, T. A., & Al-Zawi, S. (2017). Understanding of a convolutional neural network. In *In: Proc. of 2017 international conference on engineering and technology - ICET* (pp. 1–6). https://doi.org/10.1109/ICEngTechnol.2017.8308186.

Ammour, N., Alhichri, H., Bazi, Y., Benjdira, B., Alajlan, N., & Zuair, M. (2017). Deep learning approach for car detection in UAV imagery. *Journal Remote Sensing, 9*(4), 1–15. https://doi.org/10.3390/rs9040312.

ANAC - National Civil Aviation Agency. (2021). Available at: https://www.anac.gov.br/en/drones Accessed: 03 May 2021.

Araujo, R. V., Albertini, M. R., Costa-da-Silva, A. L., Suesdek, L., Franceschi, N. C. S., Bastos, N. M., … Allegro, V. L. A. C. (2015). São Paulo urban heat islands have a higher incidence of dengue than other urban areas. *Brazilian Journal of Infectious Diseases, 19*(2), 146–155. https://doi.org/10.1016/j.bjid.2014.10.004.

Ball, J. E., Anderson, D. T., & Chan, C. S., Sr. (2017). Comprehensive survey of deep learning in remote sensing: Theories, tools, and challenges for the community. *Journal of Applied Remote Sensing, 11*(4), Article 042609. https://doi.org/10.1117/1.JRS.11.042609.

Barrera, R., Navarro, J. C., Mora Rodríguez, J. D., Domínguez, D., & González García, J. E. (1995). Public service deficiencies and Aedes aegypti breeding sites in Venezuela. *Bulletin of the Pan American Health Organization (PAHO), 29*(3). sept. 1995.

Bejiga, M. B., Zeggada, A., Nouffidj, A., & Melgani, F. (2017). A convolutional neural network approach for assisting avalanche search and rescue operations with UAV imagery. *Remote Sensing, 9*(2), 100. https://doi.org/10.3390/rs9020100.

Benjdira, B., Khursheed, T., Koubaa, A., Ammar, A., & Ouni, K. (2019). Car detection using unmanned aerial vehicles: Comparison between faster R-CNN and YOLOv3. In *In: Proceedings of the 1st international conference on unmanned vehicle systems (UVS), Muscat, Oman* (pp. 1–6). https://doi.org/10.1109/UVS.2019.8658300.

Bhatt, S., Gething, P. W., Brady, O. J., Messina, J. P., Farlow, A. W., Moyes, C. L., … Hay, S. I. (2013). The global distribution and burden of dengue. *Nature, 496*(7446), 504–507. https://doi.org/10.1038/nature12060.

Bhola, R., Krishna, N. H., Ramesh, K. N., Senthilnath, J., & Anand, G. (2018). Detection of the power lines in UAV remote sensed images using spectral-spatial methods. *Journal of Environmental Management, 206*, 1233–1242. https://doi.org/10.1016/j.jenvman.2017.09.036.

Carrasco-Escobar, G., Manrique, E., Ruiz-Cabrejos, J., Saavedra, M., Alava, F., Bickersmith, S., … Gamboa, D. (2019). High-accuracy detection of malaria vector larval habitats using drone-based multispectral imagery. *PLoS Neglected Tropical Diseases, 13*(1), Article e0007105. https://doi.org/10.1371/journal.pntd.0007105.

Colwell, R. N. (1997). History and place of photographic interpretation. In W. R. Philipson (Ed.), *Manual of photographic interpretation* (2a ed., pp. 3–47). ASPRS: Maryland.

Cortes, C., & Vapnik, V. (1995). Support vector machine. *Machine Learning, 20*(3), 273–297.

De Silva, P. M., & Marshall, J. M. (2012). Factors contributing to urban malaria transmission in sub-Saharan Africa: A systematic review. *Journal of Tropical Medicine, 2012*. https://doi.org/10.1155/2012/819563.

DECEA - Department of Airspace Control. (2021). ICA 100–40 - Sistemas de Aeronaves Remotamente Pilotadas e o Acesso ao Espaço Aéreo Brasileiro. Available at: https://publicacoes.decea.mil.br/publicacao/ica-100-40.

Diniz, M. T. M., & Medeiros, J. B. (2018). Mapping of breeding sites of aedes aegypti in Caicó/RN city with use of unmanned aerial vehicle. *Revista GeoNordeste, 2*, 196–207.

Espinosa, M. O., Polop, F., Rotela, C. H., Abril, M., & Scavuzzo, C. M. (2016). Spatial pattern evolution of Aedes aegypti breeding sites in an Argentinean city without a dengue vector control programme. *Geospatial Health, 11*(3), 307–317. https://doi.org/10.4081/gh.2016.471.

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning-An MIT Press book. Available in https://www.deeplearningbook.org/.

Grubesic, T. H., Wallace, D., Chamberlain, A. W., & Nelson, J. R. (2018). Using unmanned aerial systems (UAS) for remotely sensing physical disorder in neighborhoods. *Landscape and Urban Planning, 169*, 148–159. https://doi.org/10.1016/j.landurbplan.2017.09.001.

Haas-Stapleton, E. J., Barretto, M. C., Castillo, E. B., Clausnitzer, R. J., & Ferdan, R. L. (2019). Assessing mosquito breeding sites and abundance using an unmanned aircraft. *Journal of the American Mosquito Control Association, 35*(3), 228–232. https://doi.org/10.2987/19-6835.1.

Haralick, R. M., Shanmugam, K., & Dinstein, I. H. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics, 6*, 6103–6621. https://doi.org/10.1109/TSMC.1973.4309314.

Hardy, A., Makame, M., Cross, D., Majambere, S., & Msellem, M. (2017). Using low-cost drones to map malaria vector habitats. *Parasites & Vectors, 10*(1), 1–13. https://doi.org/10.1186/s13071-017-1973-3.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90.

Kabaria, C. W., Gilbert, M., Noor, A. M., Snow, R. W., & Linard, C. (2017). The impact of urbanization and population density on childhood plasmodium falciparum parasite prevalence rates in Africa. *Malaria Journal, 16*(1), 49. https://doi.org/10.1186/s12936-017-1694-2.

Keiser, J., Utzinger, J., De Castro, M. C., Smith, T. A., Tanner, M., & Singer, B. H. (2004). Urbanization in sub-saharan Africa and implication for malaria control. *The American Journal of Tropical Medicine and Hygiene, 71*(2_suppl), 118–127.

Kim, H. Y., & Araujo, S. A. (2011). Ciratefi: An RST-invariant template matching with extension to color images. *Integrated Computer-Aided Engineering, 18*(1), 75–90. https://doi.org/10.3233/ICA-2011-0358.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *In: Proc. of the IEEE, 86*(11), 2278–2324. https://doi.org/10.1109/5.726791.

Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single shot multibox detector. In *European conference on computer vision* (pp. 21–37). https://doi.org/10.1007/978-3-319-46448-0_2.

Madzlan, F., Dom, N. C., Tiong, C. S., & Zakaria, N. (2016). Breeding characteristics of aedes mosquitoes in dengue risk area. *Procedia-Social and Behavioral Sciences, 234*, 164–172. https://doi.org/10.1016/j.sbspro.2016.10.231.

Mehra, M., Bagri, A., Jiang, X., & Ortiz, J. (2016). Image analysis for identifying mosquito breeding grounds. In *Proc. of 2016 IEEE international conference on communication and networking (SECON workshops)* (pp. 1–6). https://doi.org/10.1109/SECONW.2016.7746808.

Nelson, J., & Gorichanaz, T. (2019). Trust as an ethical value in emerging technology governance: The case of drone regulation. *Technology in Society, 59*, 101131. https://doi.org/10.1016/j.techsoc.2019.04.007.

PAHO - Pan American Health Organization. (23 June 2020). Dengue cases in the Americas reach 1.6 million, which highlights the need for mosquito control during the pandemic. Available at: https://www.paho.org/bra/index.php?option=com_content&view=article&id=6205:casos-de-dengue-nas-americas-chegam-a-1-6-milhao-o-que-destaca-a-necessidade-do-controle-de-mosquitos-durante-a-pandemia&Itemid=812 Accessed: 19 Fev 2021.

Passos, W. L., Dias, T. M., Alves Junior, H. M., Barros, B. D., Araujo, G. M., Lima, A. A., … Lima Netto, S. (2018). About automatic detection of aedes aegypti mosquito focuses. In *Proc. of XXXVI Brazilian symposium on telecommunications and signal processing* (pp. 1–5). https://doi.org/10.14209/SBRT.2018.51.

Pisner, D. A., & Schnyer, D. M. (2020). Support vector machine. In *In Machine learning* (pp. 101–121). Academic Press. https://doi.org/10.1016/b978-0-12-815739-8.00006-7.

Platt, J. (1998). Fast training of support vector machines using sequential minimal optimi- zation. Fast training of support vector machines using sequential minimal optimization. In B. Scholkopf, C. J. C. Burges, & A. J. Smola (Eds.), *Advances in kernel methods - support vector learning, Cambridge, MA* (pp. 185–208).

PMSP - Prefeitura Municipal de São Paulo. (12 Feb 2016). Prefeitura promove mobilização de combate ao *Aedes aegypti*. Available in: http://www.capital.sp.gov.br/noticia/prefeitura-promove-mobilizacao-de-combate-ao-aedes.

Redmon, J., Divvala, S. K., Girshick, R. B., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In *Proceedings of 2016 IEEE conference on computer vision and pattern recognition (CVPR)* (pp. 779–788). Avaiable in http://arxiv.org/abs/1506.02640.

Redmon, J., & Farhadi, A. (2018). *Yolov3: An incremental improvement*. Computing Research Repository (CoRR). Available in http://arxiv.org/abs/1804.02767.

Schafrick, N. H., Milbrath, M. O., Berrocal, V. J., Wilson, M. L., & Eisenberg, J. N. (2013). Spatial clustering of Aedes aegypti related to breeding container characteristics in coastal Ecuador: Implications for dengue control. *The American Journal of Tropical Medicine and Hygiene, 89*(4), 758–765. https://doi.org/10.4269/ajtmh.12-0485.

Simmons, C. P., Farrar, J. J., van Vinh Chau, N., & Wills, B. (2012). Dengue. *The New England Journal of Medicine., 366*(15), 1423–1432. https://doi.org/10.1056/nejmra1110265.

Singh, S., Ahuja, U., Kumar, M., Kumar, K., & Sachdeva, M. (2021). Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment. *Multimedia Tools and Applications, 80*(13), 19753–19768. https://doi.org/10.1007/s11042-021-10711-8.

Sivic, J., & Zisserman, A. (2003). Video Google: A text retrieval approach to object matching in videos. In *, 2. In: IEEE international conference on computer vision* (pp. 1470–1477). IEEE Computer: Society. https://doi.org/10.1109/ICCV.2003.1238663.

Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., & Liang, Z. (2019). Apple detection during different growth stages in orchards using theimproved YOLO-V3 model. *Computers and Eletronics in Agriculture, 157*, 417–426. https://doi.org/10.1016/j.compag.2019.01.012.

Tun-Lin, W., Lenhart, A., Nam, V. S., Rebollar-Téllez, E., Morrison, A. C., Barbazan, P., … Kroeger, A. (2009). Reducing costs and operational constraints of dengue vector control by targeting productive breeding places: A multi-country non-inferiority cluster randomized trial. *Tropical Medicine & International Health, 14*(9), 1143–1153. https://doi.org/10.1111/j.1365-3156.2009.02341.x.

UN-Habitat. (2016). *World cities report 2016: Urbanization and development–emerging futures*. Publisher: UN-Habitat. Available at: https://unhabitat.org/world-cities-report Accessed: 19 Fev 2021.

Warren, M., Billing, P., Bendahmane, D., & Wijeyaratne, P. (1999). Malaria in urban and peri-urban areas in sub-Sahara Africa. In *Environmental Health Project, activity report No. 71*.

WHO - World Health Organization. (2012). Global Strategy for dengue prevention and control, 2012–2020. WHO Report. Available at: https://www.who.int/denguecontrol/9789241504034/en/ Accessed: 19 Fev 2021.

WHO - World Health Organization. (2017). Keeping the vector out: housing improvements for vector control and sustainable development. Available at: https://www.who.int/social_determinants/publications/keeping-the-vector-out/en/ Accessed: 19 Fev 2021.

WHO - World Health Organization. (2019). World malaria report 2019. Available at https://apps.who.int/iris/handle/10665/330011 Accessed: 19 Fev 2021.

WHO - World Health Organization. (2020a). Dengue and severe dengue, 23 June. Available at: https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue Accessed: 19 Fev 2021.

WHO - World Health Organization. (2020b). Vector-borne diseases, 2 March. Available at: https://www.who.int/en/news-room/fact-sheets/detail/vector-borne-diseases Accessed: 19 Fev 2021.

Xia, Y., Ye, G., Yan, S., Feng, Z., & Tian, F. (2020). Application research of fast UAV aerial photography object detection and recognition based on improved YOLOv3. *Journal of Physics: Conference Series, 1550*(2020), 032075. https://doi.org/10.1088/1742-6596/1550/3/032075.

Xu, Y., Yu, G., Wang, Y., Wu, X., & Ma, Y. (2017). Car detection from low-altitude UAV imagery with the faster R-CNN. *Journal of Advanced Transportation, 2017*, 1–10. https://doi.org/10.1155/2017/2823617.

Xu, Z. F., Jia, R. S., Sun, H. M., Liu, Q. M., & Cui, Z. (2020). Light-YOLOv3: Fast method for detecting green mangoes in complex scenes using picking robots. *Applied Intelligence*, 1–18. https://doi.org/10.1007/s10489-020-01818-w.

Yi, Z., Yongliang, S., & Jun, Z. (2019). An improved tiny-yolov3 pedestrian detection algorithm. *Optik–International Journal for Light and Electron Optics, 183*, 17–23. https://doi.org/10.1016/j.ijleo.2019.02.038.