Comparison of Computer Vision Approaches for Recognition of Scenarios Suspected of Being Mosquito Breeding Sites in Aerial Images Acquired by UAVs

Rafael Oliveira Cotrin[1], Gustavo Araujo Lima[1], Daniel Trevisan Bravo[1] and Sidnei Alves de Araújo[1]

[1]*Programa de Pós-graduação em Informática e Gestão do Conhecimento (PPGI), Universidade Nove de Julho – UNINOVE, Rua Vergueiro, 235/249 – Liberdade, São Paulo/SP, Brasil*
rafa25.cotrin@gmail.com, gustavoaraujo59@hotmail.com, danieltb3006@gmail.com, saraujo@uni9.pro.br

Abstract. The use of unmanned aerial vehicles (UAVs) for acquisition of aerial images to support health surveillance teams in activities of combatting the mosquito breeding sites has increased a lot in recent years. However, it is still common the manual analysis of such images, requiring much time of the health workers. In this work we investigate two state-of-the-art computer vision approaches which can be employed for recognition of scenarios suspect of being potential mosquito breeding sites from aerial images acquired by UAVs. The first approach, named as BoVW+SVM, is based on Bag of Visual Words (BoVW) technique combined with the Support Vector Machine (SVM) classifier, while the second approach is based on a model of convolutional neural network (CNN) known as YOLO (You Only Look Once). For conducting the experiments, in which the approaches were compared in terms of the mAP-50 measure, we employed a dataset containing 230 images, acquired in urban regions of the city of São Paulo, which contemplate real and simulated suspected scenarios (gutters and roofs with accumulation of objects, open-air inorganic garbage containing old tires, old tires, pet bottles, plastic and paper packaging and other open containers that can accumulate water). The results obtained by YOLO were much superior to those obtained by BoVW+SVM, in terms of precision and processing speed, demonstrating that this CNN model can be employed to compose a computer vision system for automatic inspections in real time.

Keywords: Mosquito Breeding Sites, Unmanned Aerial Vehicle, Convolutional Neural Network, Bag of Visual Words, Support Vector Machine.

1 Introduction

According to the World Health Organization (WHO), the diseases caused by the Aedes aegypti mosquito affect approximately 390 million people per year and, only in Brazil more than 440 thousand cases were registered in 2017 [1]. In addition, over 1,6 million od dengue cases were notified in the Americas only in the first five months of 2020, the majority of them in Brazil [2]. Given this scenario, the fight against mosquitoes have represented a great challenge for Brazilian health authorities, since preventive measures are not always carried out by the population in an appropriate manner even with wide dissemination by the media.

In Brazil, the identification of potential mosquito breeding sites is carried out by technical teams made up of Endemic Disease Control Agents and Community Health Agents. With the exception of few areas, such as the municipality of São Paulo, such teams work only through direct inspections, completing the inspection script, based on the visual identification of risk situations, such as structures, equipment, containers and other objects without protection or with accumulation of water [3]. Inspection activities such as these are expensive, time-consuming, dangerous, in addition to being temporally and spatially limited, resulting in large portions of urban space that end up not being accounted for [4]. In addition, these teams are faced with situations of impediment to carrying out their activities, such as closed, abandoned properties or with access not allowed by the owner. Such facts affect actions to combat the proliferation of the mosquito that transmits Dengue [3].

In parallel with the prevention campaigns, several efforts have been made in Brazil to speed up the search for possible breeding sites of the Aedes aegypti mosquito, mainly in urban areas where there are many places of difficult access for 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 the diseases caused by the mosquito [5,6]. Such equipment, in addition to the low cost, compared to the manned aircraft used in image acquisition tasks, enable remote piloting, flights closer to the ground and acquires images with high spatial and temporal resolutions, allowing the detection of small objects on the Earth's surface and the perception of changes in a given region, in a short period of time. Due to these advantages the UAVs have been widely used to acquire remote sensing images [4,7].

However, although there are many initiatives proposing the use of UAVs for actions of combatting the mosquito breeding sites, usually the images acquired by this equipment are analyzed manually (visually), as in the works of Passos et al. [5] and Diniz and Medeiros [6]. The first work presents the steps to compose a database of annotated images that can be used to evaluate methods for automatic detection of some suspicious objects (tires, bottles and other containers that can accumulate water). However, such database has not been made available in the literature. Diniz and Medeiros [6] address the mapping of the target objects considered in the present work from images acquired by UAvs but, instead of presenting an automatic method, the authors conducted the mapping in a manual way using Geographical Information Systems (GIS) softwares.

Studies in the literature proposing the use of UAVs to detect mosquito breeding sites from automatic analysis of images are actually rare. Agarwal et al. [8], for example, developed a method to detect and visualize possible mosquito breeding sites. The proposed method comprises three steps: evaluation of the quality of the images; image classification using Bag of Visual Words (BoVW) – taking in to account the descriptor Scale Invariant Feature Transform (SIFT) – combined with the Support Vector Machine (SVM) classifier; 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. [9] a framework was proposed for detecting 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 also used with the descriptor Speed Up Robust Features (SURF) and the classification task performed by Bayesian classifiers. In the experiments carried out, the authors obtained an accuracy of 90%. However, the approaches proposed in [8] and [9] indicate only whether or not an image contains a suspicious scenario, without providing a spatial location, which makes them unsuitable for applications that require the precise indication of the location of the possible breeding sites. There are also the works of Carrasco-Escobar et al. [10] and Haas-Stapleton et al. [11] which investigate the automatic detection of mosquito breeding sites based on the analysis of water bodies characteristics and, therefore, quite different from the proposals of [8] and [9] and from those investigated in this work.

In this work, we compared two approaches for detecting and locating scenarios that represent potential mosquito breeding sites from aerial images acquired by UAVs. The scenarios are characterized by the existence of inorganic garbage in external areas containing small objects that can accumulate water such as old tires, pet bottles, plastic and paper packaging, among others.

The first approach, called BoVW+SVM, combines the Bag of Visual Words (BoVW) technique with the Support Vector Machine (SVM) classifier. However, unlike the works [8] and [9], in which only one feature descriptor (SIFT or SURF) was considered, we employed combinations of the following descriptors: Color Histograms (CH), Color Level Co-occurrence Matrix (CLCM), Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP). The second approach (CNN_tiny-YOLOv3) uses a CNN model called tiny-YOLOv3 [12], from the YOLO (You Only Look Once) framework proposed by Redmon et al. [13], which is composed by CNN architectures specially designed to detect objects in images. Currently, such architectures have been used to detect objects in real time, as in the works [14-16]. In fact, both BoVW and YOLOv3 have been widely used in recent works in the literature to detect objects, scenarios and other complex visual patterns in images. Thus, they can be considered as state-of-the-art techniques.

Approaches for detecting and locating scenarios such as those investigated in this work are important because they can serve as subsidies for the implementation of computer systems to assist the planning and executing of activities to control and prevent outbreaks of Aedes aegypti mosquito.

2 Materials and Experimental Setup

For conducting the experiments, a dataset of images acquired in areas located in the city of São Paulo – Brazil was created. The dataset is composed by two sets of images called DS1 and DS2, which include real and simulated scenarios. For generating the simulated scenarios, we placed portions of inorganic waste (small objects that can retain stagnant water such as old tires, pet bottles, plastic and paper containers, among others) on the ground before flying over these same areas for acquiring images. Thus, each portion provides an image.

The DS1 is composed by 119 RGB images of simulated scenarios acquired in an area of the University of São Paulo (USP). Such images, with resolution of 4000×3000 pixels, were acquired using a DJI Phantom 4 advanced equipped with a RGB DJI 20 MP camera (model FC330; 1/2.3" CMOS; FOV 94° 20 mm; aperture of f/2.8). On flights, three distances above the ground were considered: 7, 10 and 13 m. The DS2 contains 111 RGB images of real scenarios acquired in two private houses located in peripheral neighborhoods of the city of São Paulo using a GoPro HERO4 Silver camera (1/2.3" CMOS; FOV 90° 3.0mm; aperture of f/2.8) coupled to the drone. The images of DS2 were captured with distance above the ground ranging from 3 to 5 m and have a resolution of 3000×2250 pixels. Since there are no images of mosquito breeding sites scenarios available in the literature, the image database composed in this work can also be considered as an important scientific contribution since it will be made available to other researchers testing their methods upon request.

For the experiments, the 230 images were divided into 2 parts: 160 images for training (69.6%) and 70 images for the evaluation of compared approaches (30.4%). The performance of BoVW+SVM and CNN_tinny-YOLOv3 were compared in terms of the measure mAP-50 (mean average precision), which takes into account the average precision (AP) of each class and requires that the value resulting from the IoU operation is at least 0.5. The IoU operation computes a ratio of the area of intersection and area of union of the predicted bounding box and ground truth bounding box [18]. Since BoVW+SVM employs different combinations of features descriptors, we considered for the comparison with CNN_tinny-YOLOv3 only the combination of descriptors that provided the best result. The CNN_tiny-YOLOv3 was developed in C/C ++ using the OpenCV and Darknet libraries, while the BoVW+SVM was developed on the Matlab 2018. All experiments were carried out on PC with a 2.5 GHz quad-core Intel i7 processor with 16 GB of RAM and NVIDIA GeForce 930M GPU, and running Windows 10 Pro.

3 Compared Approaches

3.1 BoVW+SVM

The BoVW+SVM approach is based on the works [8,9] and comprises six steps, as follows:

Step 1. Extraction of features from subimages (windows of 200×200 pixels) extracted from the training images. Unlike the works [8,9], which consider only one descriptor (SIFT or SURF), we take into account the descriptors Color Histograms (CH), Color Level Co-occurrence Matrix (CLCM), Histogram of Oriented Gradients (HOG) and Local Binary Patterns (LBP), which were used isolated and combined as follows: CH, CLCM, HOG, LBP, CH+HOG, CH+LBP, CLCM+HOG, HOG+LBP, CH+CLCM+LBP, CH+HOG+LBP, CH+CLCM+HOG+LBP. The descriptors SIFT and SURF were not considered in this work because according to [17] their performances are decreased in images rich in details such as the images from DS1 and DS2. From the CLCM computation, 6 features were extracted (Haralick descriptors): second angular moment, entropy, contrast, variance, correlation and homogeneity. The CH generates 384 features (128 bins for each color channel). HOG and LBP allow to extract, respectivelly, 20736 features (using 8×8 cells) and 2124 features (using 32×32 cells). Both CH and LBP are calculated for the three-color channels separately.

Step 2. Creation of the visual words dictionary from the extracted features using the k-means algorithm. The dictionary size (K) was empirically defined as 440.

Step 3. Representation of each subimage from the dictionary composed of histograms of visual words (coding);

Step 4. Synthesis of visual words histograms into new feature vectors (pooling).

Step 5. Training the SVM classifier employing the new feature vectors;

Step 6. Classification of the windows extracted from each image belonging to the set of tests employing 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 containing other materials (scen_type9). From the 160 images separated for training, features of 900 subimages (100 of each class) were extracted to compose the training sets (each combination of descriptors generates a training set).

3.2 CNN_tiny-YOLOv3

The CNN architecture employed (tiny-YOLOv3) is illustrated in Fig. 1. In its training, the 9 classes mentioned in the previous section were considered. The training dataset was composed of 430 subimages (with several dimensions) manually extracted from 160 training images. It is important to highlight that the data augmentation scheme contemplated by the YOLO framework was employed to increase the amount of samples during the CNN training, aiming to improve the CNN's generalization capacity.

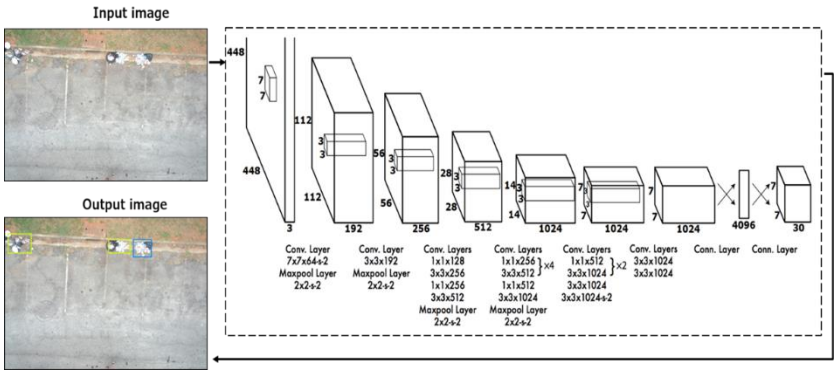

Figure 1. Architecture of CNN used to detect scenarios (CNN_tiny-YOLOv3)

The following training parameters were employed: number of batches = 64; number of subdivisions = 32; maximum number of iterations = 18,000; learning rate = 0.001. These parameters were based on the recommendations of Redmon [12,13]. After 96 hours of training, the value of the validation loss was 0.06 at iteration 17,930, with about 1,147,520 samples generated by the data augmentation scheme.

4 Experimental Results and Discussion

The evaluation of the performance of investigated approaches was made from the classification of the 70 images destined for the tests and that were not used in the training phase. After the classification of the images, the AP is computed for each scenario class (scen_type*) and then the mAP-50 is obtained from these AP values.

The classification using BoVW+SVM is based on the sliding window strategy, in which each subimage extracted from an input image is classified. Empirically, a threshold value of 0.60 was defined for the posterior probability, which is calculated for the predicted class. In this case, only the bounding boxes classified with a probability value equal or greater than to this threshold were considered. The results of the classifications, in terms of mAP-50, for the different combinations of descriptors mentioned in section 3.1 are presented in Table 1 from which one can see that the highest value of mAP-50 (0.6453) was obtained in the classification considering the combination CH+LBP. In addition, the classification produced 1,638 true positive (TP) cases and 2,352 false positive (FP) cases, which lead to a median value of mAP-50. The Fig. 2b illustrate some classified windows representing FP cases marked with red circles.

As the CLCM descriptor computes the occurrences of local transitions between color channels and each scenario can contain a variety of objects with different colors, these local transitions can prejudice instead of helping the discriminative power of the descriptor. Similar analysis can be done for HOG, but taking into account the variety of shapes and texture of the objects. These may be the reasons for the low performance of CLCM and

HOG descriptors. Regarding CH, different from CLCM it reflects average values of color intervals in the RGB space and thus outperforms CLCM. With respect to the LBP, the fact of it was computed for each color channel may have contributed to its performance.

Table 1. Results provided by BoVW+SVM

Descritor(es)	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

The high number of FP cases indicates that BoVW+SVM with considered combinations of descriptors is not the most suitable approach for solving the problem addressed in this study. Another negative aspect of BoVW+SVM approach is the time spent for classification all windows extracted from each analyzed image. In our experiments, the classification of the 70 testing images required several hours of processing in the Matlab environment.

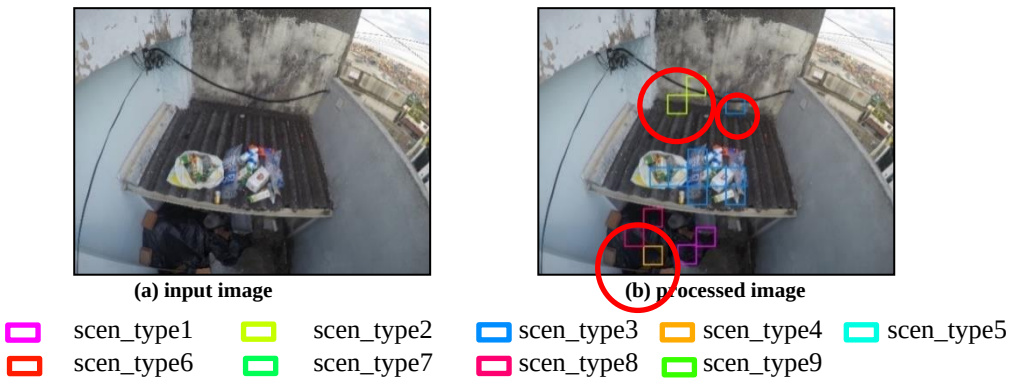

Figure 2. Some cases of FP produced by BoVW+SVM

On the other hand, CNN_tiny-YOLOv3 took only 18 seconds to classify the same 70 images and made 154 detections. From the 111 bounding boxes defined as ground truth, 96 were correctly classified (TP cases) providing a mAP-50 of 0.9028. Some examples of TP cases are illustrated in Fig. 3b. In addition, 15 cases of false negatives (FN) and 11 cases of FP were computed, being two of them indicated in Fig. 3d by red circles.

The results presented in Table 2 demonstrate the good performance of CNN_tiny-YOLOv3 in the detection of 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 less than one of the results obtained by BoVW+SVM (for scen_type6). Its valid to highlight that we also tested BoVW+SVM with smaller windows (100×100), but the results obtained were worse than those described in the tables 1 and 2.

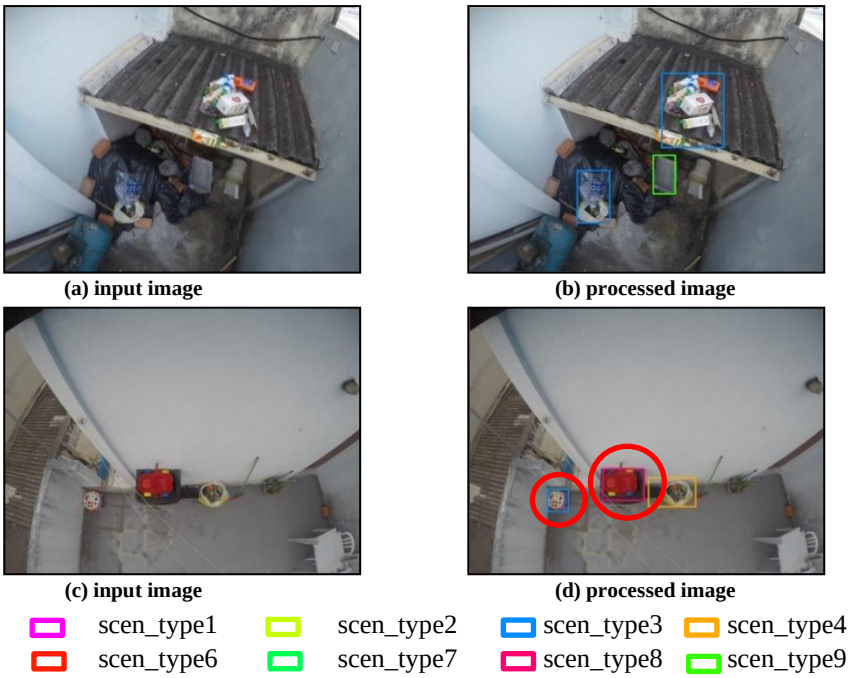

Figure 3. Some results obtained by CNN_tiny-YOLOv3

Table 2. Average Precision (AP) obtained in the detection of scenarios

Class	BoVW+SVM	CNN_tiny-YOLOv3
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**

Finally, it is important to mention that the results obtained by the approaches presented in this work were not compared with the results reported in Agarwal et al. [8] and Mehra et al. [9] due two main reasons: (i) different objectives – while the approaches proposed in [8] and [9] only indicate whether or not an image contains a suspicious scenario, BoVW+SVM and CNN-tinny-YOLOv3 locate and typify the suspicious scenarios in the analyzed images; (ii) the approaches proposed in [8] and [9] were evaluated with a database of images that is not available in the literature. Thus, even if the aim of such approaches was the same as BoVW+SVM and CNN-tinny-YOLOv3, the comparison with different database of images would not be fair. This reinforces the importance of making available the database of images composed in this work.

5 Conclusion

In this work we investigated two approaches for detection of scenarios suspect of being potential mosquito breeding sites from aerial images acquired by UAVs. The first approach (BoVW+SVM) did not present very satisfactory results (mAP-50 = 0.6453) probably because some of the descriptors considered here were impaired in the extraction of local features, due to the variety and agglutination of objects that describe the scenarios in the

images. On the other hand, the CNN_tiny-YOLOv3 approach presented a much superior performance (mAP-50 = 0.9028), probably due to the fact that it works naturally with several scales and increased data. The experiments conducted in this work indicate that CNN tiny-YOLOv3 could be used to compose an intelligent system for tracking and combating possible mosquito breeding sites, mainly in places of difficult access (on slabs or roofs for example), which are usually ignored during inspections made by health workers. In future works we intend to implement a computer vision system for operating in real time and able to automatically provide the geolocation of each scenario detected in the analyzed images.

Acknowledgements. 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).

Authorship statement. The authors hereby confirm that they are the sole liable persons responsible for the authorship of this work, and that all material that has been herein included as part of the present paper is either the property (and authorship) of the authors, or has the permission of the owners to be included here.

References

[1] WHO – World Health Organization. Dengue and severe dengue. 23 June 2020. Available at: *https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue*. Accessed: 19 Fev 2021.
[2] PAHO – Pan American Health Organization. Dengue cases in the Americas reach 1.6 million, which highlights the need for mosquito control during the pandemic. 23 June 2020. 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.
[3] Ministério da Saúde – Governo Federal do Brasil, Secretaria de Vigilância em Saúde, & Departamento de Vigilância Epidemiológica. (2009). Diretrizes nacionais para prevenção e controle de epidemias de dengue. Available at: *https://bvsms.saude.gov.br/bvs/publicacoes/diretrizes_nacionais_prevencao_controle_dengue.pdf*. Accessed: 01 Jul 2021.
[4] Grubesic, T. H., Wallace, D., Chamberlain, A. W., Nelson, J. R.. Using unmanned aerial systems (UAS) for remotely sensing physical disorder in neighborhoods. *Landscape and Urban Planning*, 169, 148-159, 2018.
[5] Passos, W. L., Dias, T. M., Alves Junior, H. M., Barros, B. D., Araujo, G. M., Lima, A. A., Silva, E. A. B., Lima Netto, S. About Automatic Detection of Aedes aegypti Mosquito Focuses. In: *Anais do XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais*, 1–5, 2018.
[6] Diniz, M. T. M., Medeiros, J. B. Mapping of reproductive foci of aedes aegypti in the city of Caicó/RN with the aid of unmanned aerial vehicle. *Revista GeoNordeste*, 2, 196-207, 2018.
[7] Ammour, N., Alhichri, H., Bazi, Y., Benjdira, B., Alajlan, N. e Zuair, M. Deep Learning Approach for Car Detection in UAV Imagery. *Journal Remote Sensing*, 9(4), 1-15, 2017.
[8] Agarwal, A., Chaudhuri, U., Chaudhuri, S., Seetharaman, G. 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, 2014.
[9] Mehra, M., Bagri, A., Jiang, X., Ortiz, J. Image analysis for identifying mosquito breeding grounds. In: *2016 IEEE International Conference on Communication and Networking (SECON Workshops)*, 1–6, 2016.
[10] Carrasco-Escobar, G., Manrique, E., Ruiz-Cabrejos, J., Saavedra, M., Alava, F., Bickersmith, S., ... & Gamboa, D. High-accuracy detection of malaria vector larval habitats using drone-based multispectral imagery. *PLoS neglected tropical diseases*, 13(1), e0007105, 2019.
[11] Haas-Stapleton, E. J., Barretto, M. C., Castillo, E. B., Clausnitzer, R. J., Ferdan, R. L. Assessing Mosquito Breeding Sites and Abundance Using an Unmanned Aircraft. *J. of the American Mosquito Control Association*, 35(3), 228-232, 2019.
[12] Redmon, J., Divvala, S. K., Girshick, R. B., Farhadi, A. You only look once: Unified, real-time object detection. In: *Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 779-788, 2016.
[13] Redmon, J. Yolov3: An incremental improvement. *Computing Research Repository (CoRR)*, 2018.
[14] Yi, Z., Yongliang, S., Jun, Z. An improved tiny-yolov3 pedestrian detection algorithm. *Optik - International Journal for Light and Electron Optics*, 183, 17-23, 2019.
[15] Benjdira, B., Khursheed, T., Koubaa, A., Ammar, A., Ouni, K. Car Detection using Unmanned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3. In: *Proceedings of the 1st International Conference on Unmanned Vehicle Systems (UVS), Muscat, Oman*, 1-6, 2019.
[16] Tian, Y., Yang, G., Wang, Z., Wang, H., Li, E., Liang, Z. Apple detection during different growth stages in orchards using theimproved YOLO-V3 model. *Computers and Eletronics in Agriculture*, 157, 417-426, 2019.
[17] Kim, H. Y., Araujo, S. A. Ciratefi: An RST-invariant template matching with extension to color images. *Integrated Computer-Aided Engineering*, 18(1), 75-90, 2011.
[18] Xia, Y., Ye, G., Yan, S., Feng, Z., & Tian, F. Application Research of Fast UAV Aerial Photography Object Detection and Recognition Based on Improved YOLOv3. *Journal of Physics: Conference Series*, 1550(2020), p. 032075, 2020.