XL CILAMCE
IBERO-LATIN AMERICAN CONGRESS ON COMPUTATIONAL METHODS IN ENGINEERING
NOVEMBER 11-14, 2019
Praiamar Natal Hotel & Convention
Natal, RN-BRAZIL

# A CONVOLUTIONAL NEURAL NETWORK-BASED APPROACH FOR DETECTION OF OBJECTS AND SCENARIOS SUSPECTED OF BEING POSSIBLE BREEDING SITES OF AEDES AEGYPTI MOSQUITO

**Daniel Trevisan Bravo**
**Wonder Alexandre Luz Alves**
**Sidnei Alves de Araújo**
*danieltbravo@uni9.pro.br*
*wonder@uni9.pro.br*
*saraujo@uni9.pro.br*
*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*

**Abstract.** In remote sensing field, the use of satellite images to detect small objects becomes impractical due to the low spatial resolution of such images. With the use of unmanned aerial vehicles (UAVs), popularly known as drones, it is possible to acquire aerial images with high spatial and temporal resolutions that allow the detection of small objects on the earth's surface and the perception of changes in a certain region in a short period of time. However, the detection of certain objects in the images acquired by UAVs has been a great challenge due to the amount of details present in these images, especially those acquired in urban areas. This work investigates the automatic detection of objects and scenarios in aerial images acquired by UAVs. The proposed approach aims to detect target objects (typical containers for water storage) and scenarios suspected to be possible breeding sites of the Aedes aegypti mosquito. For the detection of the objects and scenarios, two convolutional neural network (CNN) architectures from the YOLOv3 framework were employed. In the experiments conducted, we considered 142 images acquired in peripheral regions of the city of São Paulo/Brazil containing water tanks; 25 images of real scenarios, obtained by Google image search, containing garbage and old tires and 35 images containing scenarios simulating the main mosquito breeding sites (old tires, gutters, among others). In the experiments performed, in which the proposed approach was evaluated using the mean average precision (mAP), the rates of 0.95 and 0.97, respectively, were obtained for the detection of objects and scenarios, indicating that the proposed approach is a good alternative to solve the investigated problem.

**Keywords:** Convolutional Neural Network, Object Detection, UAV, YOLOv3

# 1 Introduction

In remote sensing area, the use of satellite images to detect small objects becomes impractical due to the low spatial resolution of such images. With the use of unmanned aerial vehicles (UAVs), popularly known as drones, it is possible to acquire aerial images with high spatial and temporal resolutions that allow the detection of small objects on the earth's surface and the perception of changes in a certain region in a short period of time. However, the detection of objects in the images acquired by UAVs has been a great challenge due to the amount of details present in these images, especially those acquired in urban areas.

Regarding the detection of objects in images acquired by UAVs, Xu et al. [1] proposed a framework for detecting cars using a Convolutional Neural Network (CNN), called Faster R-CNN (Region Convolutional Neural Network), in low altitude images, which were acquired at signaled junctions. Ammour et al. [2] also developed an approach to car detection and counting using a CNN combined with the Support Vector Machine (SVM) classifier. In order to support search and rescue operations in regions with avalanche risk, Bejiga et al. [3] 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 works 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 have to test by hierarchical grouping, from similar regions of the image, based on the compatibility of color, texture, size and shape. Even with such reduction, the operations performed to classify the objects using these approaches are very computationally expensive.

YOLO is a framework composed of CNNs specially designed for object detection. The "YOLO - You Only Look Once", proposed by Redmon et al. [4], 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 [5] 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. YOLOv3 outperformed popular and robust methods such as Faster R-CNN with RestNet developed by He et al. [6] and the SSD (Single Shot Multibox Detector) proposed by Liu et al. [7], presenting competitive results and being faster.

Yi et al. [8] developed an approach for pedestrian detection using the tiny-YOLOv3 (reduced version with 9 layers), in conjunction with the k-means algorithm to filter out the best features of the training set. Benjdira et al. [9] conducted a study comparing the results obtained by Faster R-CNN and YOLOv3 in detection of cars using UAVs and showed that YOLOv3 is better than Faster R-CNN. Tian et al. [10] developed an approach using YOLOv3 for the real-time detection of apples in orchards in order to evaluate the growth phases of apples and to estimate the yield. In the experiments conducted they have shown that the proposed YOLOv3-dense (modified version) model is superior to the original YOLO-v3 model and to the R-CNN.

Regarding the detection objects and scenarios suspected to be potential breeding sites of the Aedes aegypti mosquito, although there are many initiatives using UAVs, usually the captured images are analyzed manually (visually), as in the works of Passos et al. [11] and Diniz and Medeiros [12]. The few works found in the literature addressing the automatic analysis are still incomplete, considering only the presence of suspect scenarios in the images, without spatial localization (Agrawal et al. [13]; Fornace et al. [14] and Mehra et al. [15]).

Is in this context that this work is inserted, with the proposal of using two architectures of CNNs belonging to the YOLOv3 framework for the detection of target objects (water tanks) and scenarios (exposed garbage, gutters, combination of old tires and other objects) in urban aerial imagery acquired by UAVs. It should be emphasized that this task is important because it can be part of an approach for the identification of potential breeding sites of the Aedes aegypti mosquito in places with difficult access by residents and health agents, such as slabs and roofs. The main contribution of this paper lies in the fact that the detection of specific objects, such as water tanks, is not treated in any of the approaches for the automatic detection of objects suspected of being possible breeding sites of Aedes aegypti mosquito.

Finally, it is important to mention that the images used in the experiments conducted in the works of Agrawal et al. [13] and Mehra et al. [15] were not made available, making impossible a fair comparison with our approach. However, one can observe that the accuracies of 80% and 92% obtained in [13] and [14], respectively, are lower than the rates obtained in this work.

## 2 Materials and Methods

To conduct the experiments presented in this work, 3 datasets (DS) of images named DS1, DS2 and DS3 were composed. To make up the DS1, the acquisitions of the images were made by using a DJI Phantom 3 professional drone equipped with a Sony EXMOR 12.4 MP RGB camera. The flight was carried out 50 meters from the ground, with a spatial resolution of 2.2 cm. A total of 142 images with size of 4000×3000 pixels were acquired in peripheral regions of the city of São Paulo/Brazil (Guaianases and Ferraz de Vasconcelos). The images from this dataset contain the main target objects, more specifically water tanks, which may be covered or uncovered.

To compose the DS2, 35 images with size of 4000×3000 pixels and spatial resolution of 2.2 cm were acquired, using a DJI Phantom 4 advanced drone equipped with a RGB DJI camera, in a farm located at the city of Mairiporã/Brazil. These images were acquired at 1, 2, 5 and 7 meters away from the scenarios simulated on the ground considering a gutter, old tires and six containers containing different amounts of water in several conditions: clean water, dirty water and silty water.

Finally, the DS3 consists of 25 images that were obtained from Google image search, with real scenarios containing garbage and old tires in the same image.

For the detection of water tanks, defined as target objects and that have various models and colors, a YOLOv3 framework CNN architecture with 106 layers was used. For the scenario detection task, we employed a more compact version of CNN containing 9 layers, called tiny-YOLOv3.

The approach proposed in this work was developed in C++ language using OpenCV and Darknet libraries. The computational experiments were conducted on PC Intel Core i7 with 2.5GHz and 16 GB of RAM.

## 3 Proposed Approach

The proposed approach is composed by 2 steps: detection of target objects and detection of scenarios, which are described in detail in sections 3.1 and 3.2.

### 3.1 Detection of target objects

In this step, the detection of the water tanks is performed employing a CNN architecture of the YOLOv3 framework [5]. The working of this step, called CNN_WaterTanks, is illustrated in Figure 1. The architecture of CNN_WaterTanks is composed by 106 layers, from which 75 are convolutional layers and 28 are upsampling layers and residual block layers. The layers 82, 94 and 106 was adjusted for detection objects on 3 different scales.

It is important to mention that the characteristics of the water tanks depend on the model and the year they were manufactured. In addition, the raw material used for the manufacture of old tanks (asbestos fiber cement) differs a lot from the most current tanks (plastic polyethylene). From the observation of the images and the previous knowledge about the different types of water tanks in urban areas in Brazil, they were subdivided into 7 distinct classes: wtank_type1 (0), wtank_type2 (1), wtank_type3 (2), wtank_type4 (3), wtank_type5 (4) , wtank_type6 (5) and wtank_type7 (6). This division is necessary because each type of tank has specific characteristics (color, shape and cover pattern) that differentiate them from each other.

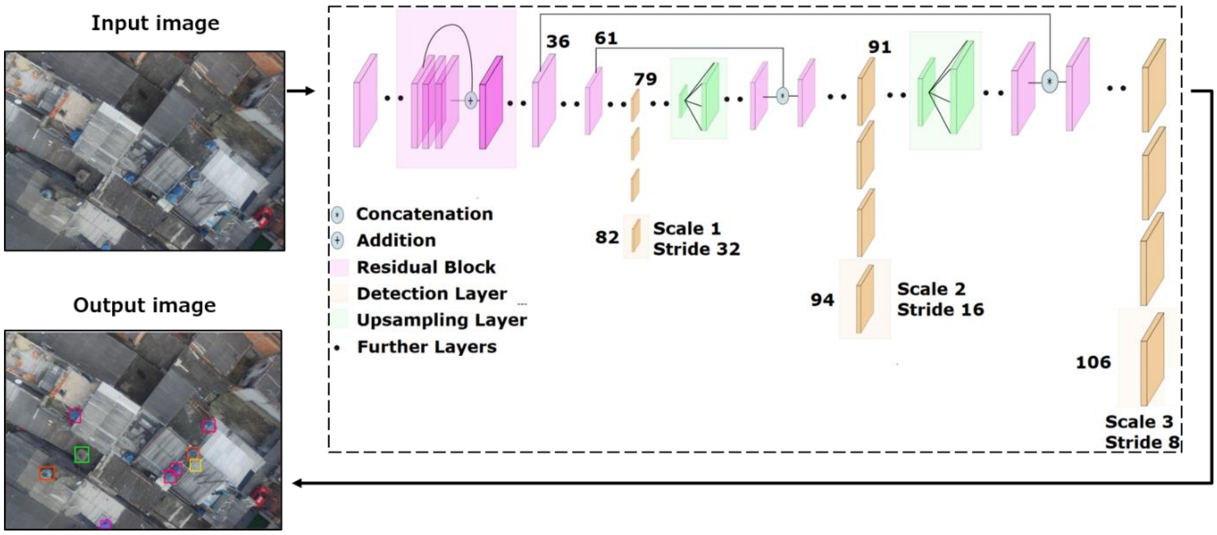


Figure 1. CNN architecture used to detect target objects (CNN_WaterTanks)

The training dataset was composed by 690 sub-images (with several dimensions) manually extracted from 100 images belonging to the DS1, being the remaining images used in the performance evaluation of the CNN_WaterTanks. In other words, we divided the images from DS1 into two parts: 70% for training and 30% for test. This division was chosen because a 70-30 ratio is often used in the literature for classification tasks. It is important to point out that the data augmentation was applied to increase the quantity of samples. Thus, at each iteration during the training of the model the number samples is increased automatically by applying different transformations such as zooming, rotation, flipping, noise addition, etc.

### 3.2 Detection of scenarios

For detection of the scenarios, the tiny-YOLOv3 architecture was used. It is illustrated in Figure 2 and referred here as CNN_Scenarios. The following classes were defined for training it: tire (0), gutter (1), garbage (2) and container (3).

The training dataset was composed of 500 sub-images (with several dimensions) manually extracted from 25 images belonging to DS2 and 15 images from the DS3, which represent approximately 70% of the images from both sets for training. The remaining images of DS2 and DS3 were used in the performance evaluation of CNN_Scenarios.

It is important to highlight that, as in the case previous step the data augmentation scheme was employed to increase the quantity of samples during the training.

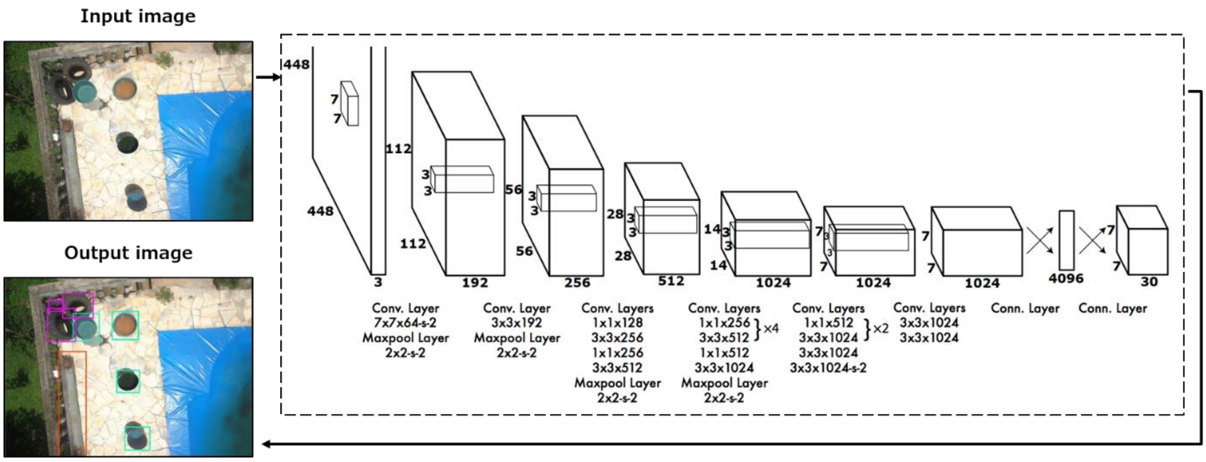


Figure 2. CNN architecture used to detect scenarios (CNN_Scenarios)

# 4 Experimental Results and Discussions

For CNN_WaterTanks, the following configuration parameters were adopted for the YOLOv3 architecture: number of batches = 64; number of subdivisions = 32; maximum quantity of batches = 14000. The parameter selected to stop training is based on lack of improvement in "validation loss", which is activated if the model runs more than five times without decreasing the loss. These parameters were adopted as suggested in Redmon [5]. After 120 hours of training, totaling 5065 iterations, the value of the validation loss was 0.36, which did not decrease during several iterations, proving to be enough.

To evaluate the accuracy of detection of the water tanks, 42 images were submitted to classification task, which spent 21 seconds. From the total of 152 Ground-truth bounding boxes, 140 were correctly classified.

In the Figure 3b it is possible to see that all the water tanks were detected correctly (cases of True Positive – TP).

We counted 19 cases of False Positive (FP) and 12 cases of False Negatives (FN). In Figure 3d it is possible to identify a case of FP, which is highlighted with a red circle and three cases of FN (Figure 3f), 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 (0.88) and Precision (0.92) metrics.

For CNN_Scenarios, the following configuration parameters (suggested in Redmon [5]) were adopted: number of batches = 64; number of subdivisions = 32; maximum quantity of batches = 8000. These parameters were adopted as suggested in Redmon [11]. After 48 hours of training, totaling 7995 iterations, the value of the validation loss was 0.84, which did not decrease during several iterations, proving to be enough.

To evaluate the accuracy of detection of scenarios 20 images were submitted to the CNN_Scenarios, which made all detections in 6 seconds. From the 398 bounding boxes defined as Ground-truth, 391 were correctly detected. In the Figure 4b it is possible to observe that all scenarios were detected correctly (TP).

We identified 7 cases of FN and 9 cases of FP. In Figure 4d it is possible to identify a case of FP, which is highlighted with a red circle and one case of FN (Figure 4f), which is highlighted with white circle. In this task both Recall and Precision were 0.98.


wtank_type1 wtank_type2 wtank_type3 wtank_type4 wtank_type5
wtank_type6 wtank_type7

Figure 3. Some results obtained in the task of water tanks detection using CNN_WaterTanks


Figure 4. Some results obtained in the task of scenarios detection using CNN_Scenarios

For the 42 images classified by CNN_WaterTanks and 20 classified by CNN_Scenarios, the values of 0.95 and 0.97, respectively, were obtained for the mAP-50 (mean Average Precision), considering the AP of each class, which are presented in Table 1.

Table 1. Average Precision in the detection of objects and scenarios

| | Class | AP (Average Precision) |
|---|---|---|
| CNN_WaterTanks | wtank_type1 | 0.97 |
| | wtank_type2 | 0.98 |
| | wtank_type3 | 0.88 |
| | wtank_type4 | 1.00 |
| | wtank_type5 | 0.90 |
| | wtank_type6 | 0.89 |
| | wtank_type7 | 1.00 |
| | mAP-50 | 0.95 |
| CNN_Scenarios | tire | 0.90 |
| | gutter | 0.98 |
| | garbage | 1.00 |
| | container | 1.00 |
| | mAP-50 | 0.97 |

As one can see in Table 1 the performance of YOLOv3 was excellent in detection of objects and scenarios. Its worst performance (0.88) occurred in the detection of wtank_type3, probably due to the low occurrence of this type of water tank in the images, which may have impaired learning of the employed CNN. On the other hand, even in situations where water tanks were too close to each other, CNN_WaterTanks was able to detect them individually in most cases. The worst performance of CNN_Scenarios (0.90) occurred in the detection of some tires. One explanation for this is that in the images extracted from Google the tires, in most cases, are mixed with other objects that may hide some part of them, as the case marked by a red circle in Figure 4d.

## 5 Conclusion

Detection of small objects in images acquired by UAVs is a challenging task, mainly because of high level of detail. Thus, adequate automatic methods of analysis are required. The results presented for detection of the water tanks (mAP-50=0.95) and scenarios (mAP-50=0.97) 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 these CNNs architectures. The approach presented in this work could be employed, for example, for tracking and combating potential breeding sites of the Aedes aegypti mosquito in hard to reach places in urban regions. In future works we intend to extend the database used for training including images containing more objects and scenarios and considering noisy, different illumination, containing partial occlusion and shadows. In addition, we intend to develop an approach to identify objects and scenarios containing stagnant water, in order to increase the applicability of the developed approach.

## Acknowledgements

The authors would like to thank Universidade Nove de Julho – UNINOVE and São Paulo Research Foundation – FAPESP by financial support (Proc. 2019/05748-0).

## References

[1] Y. Xu.; G. Yu; Y. Wang; X. Wu and Y. Ma. Car detection from low-altitude UAV imagery with the faster R-CNN. Journal of Advanced Transportation, v. 2017, 10p, 2017.
[2] N. Ammour; H. Alhichri; Y. Bazi; B. Benjdira; N. Alajlan and M. Zuair. Deep Learning Approach for Car Detection in UAV Imagery. Journal Remote Sensing, v. 9, n.4, p. 1-15, 2017.
[3] M. B. Bejiga; A. Zeggada; A. Nouffidj and F. Melgani. A Convolutional Neural Network Approach for Assisting Avalanche Search and Rescue Operations with UAV Imagery. Journal Remote Sensing, v. 9, n. 2, 100, 2017.
[4] J. Redmon; S. K. Divvala; R. B. Girshick and A. Farhadi. You only look once: Unified, real-time object detection. In: Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), abs/1506.02640, 2016. Avaiable in: <http://arxiv.org/abs/1506.02640>.
[5] J. Redmon. Yolov3: An incremental improvement. Computing Research Repository (CoRR), 2018. Available in: <http://arxiv.org/abs/1804.02767>.
[6] K. He; X. Zhang; S. Ren and J. Sun. Deep Residual Learning for Image Recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pages 770–778, 2016.
[7] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed; C.Y. Fu and A. C Berg. SSD: Single Shot Multibox Detector. In: European conference on computer vision, pages 21–37, 2016.
[8] Z. Yi; S. Yongliang; Z. Jun. An improved tiny-yolov3 pedestrian detection algorithm. Optik - International Journal for Light and Electron Optics, vol. 183, 2019.
[9] B. Benjdira; T. Khursheed; A. Koubaa; A. Ammar and K. Ouni. Car Detection using Un-manned Aerial Vehicles: Comparison between Faster R-CNN and YOLOv3.In: Proceed-ings of the 1st International Conference on Unmanned Vehicle Systems (UVS), Muscat, Oman, 2019.
[10] Y. Tian; G. Yang; Z. Wang; H. Wang; E. Li and Z. Liang. Apple detection during different growth stages in orchards using theimproved YOLO-V3 model. Computers and Eletronics in Agriculture, vol. 157, pages 417-426, 2019.
[11] W. L. Passos; T. M. Dias; H. M. Alves Junior; B. D. Barros; G. M. Araujo; A. A. Lima; E. A. B. Silva and S. Lima Netto. About Automatic Detection of Aedes aegypti Mosquito Focuses. In: Anais do XXXVI Simpósio Brasileiro de Telecomunicações e Processamento de Sinais, p. 1-5, 2018.
[12] M. T. M. Diniz and J. B. Medeiros. Mapping of reproductive foci of aedes aegypti in the city of Caicó / RN with the aid of unmanned aerial vehicle. Revista GeoNordeste, n. 2, p. 196-207, 2018.
[13] A. Agrawal; U. Chaudhuri; S. Chaudhuri and G. Seetharaman. Detection of potential mosquito breeding sites based on community sourced geotagged images. In: SPIE 9089, Geospatial InfoFusion and Video Analytics IV and Motion Imagery for ISR and Situational Awareness II, 2014.
[14] K. M. Fornace; C. J. Drakeley; T. William; F. Espino and J. Cox. Mapping infectious disease landscapes: Unmanned aerial vehicles and epidemiology, Trends in Parasitology, 30(11):514–519, 2014.
[15] M. Mehra; A. Bagri; X. Jiang and J. Ortiz. Image analysis for identifying mosquito breeding grounds. In: 2016 IEEE International Conference on Communication and Networking (SECON Workshops), pp. 1–6, 2016.
[16] Ministry Of Health. National Guidelines for the Prevention and Control of Dengue Epidemics. Available:http://bvsms.saude.gov.br/bvs/publicacoes/Guidelines_nacionais_prevencao_controle_dengue.pdf. Accessed Feb 29, 2019.