Visual Identification System for Aedes Detection using Drones

Henrique Soares*, Zenilton Kleber G. do Patrocínio Jr., Silvio Jamil F. Guimarães

*Undergraduate student in Computer Science
Laboratory of Image and Multimedia Data Science (IMScience)
Pontifícia Universidade Católica de Minas Gerais - Minas Gerais, Brazil 30535-901
henrique.soares.2811@gmail.com, zenilton@pucminas.br, sjamil@pucminas.br

***Abstract*—The proliferation of diseases transmitted by the Aedes aegypti mosquito, such as dengue, poses a persistent challenge to public health systems worldwide. Traditional surveillance methods for identifying breeding sites are often labor-intensive and limited in scale. This paper introduces VISADE (Visual Identification System for Aedes Detection), a system designed to automate this process through the use of Unmanned Aerial Vehicles (UAVs) and deep learning. By capturing and analyzing aerial imagery, VISADE employs a state-of-the-art object detection model, YOLOv11m, to identify and georeference potential breeding sites like discarded tires, buckets, and water tanks. A key aspect of our methodology is the use of high-resolution images (1200x1200 pixels) to enhance the detection of small objects, a significant challenge in aerial surveys. Our model, trained on a custom dataset collected in urban areas of Minas Gerais, Brazil, achieved a mean Average Precision (mAP@0.5) of 0.934, demonstrating high efficacy in identifying potential breeding sites. VISADE aims to provide public health authorities with an efficient, scalable, and low-cost tool, generating risk maps to direct vector control actions more effectively.**

***Index Terms*—Object Detection, Deep Learning, Public Health, Aedes aegypti, Unmanned Aerial Vehicles, Drones, YOLO, Computer Vision.**

I. Introduction

Arboviral diseases transmitted by the *Aedes aegypti* mosquito, such as dengue, Zika, and chikungunya, constitute a serious public health problem, especially in tropical and subtropical regions [1]. The World Health Organization (WHO) points to vector control as the main prevention strategy. This control depends on the elimination of breeding sites—places with standing water where mosquitoes reproduce. However, these sites are typically identified by health agents through manual inspections, which are costly and lack scalability [2].

In recent years, the field of Artificial Intelligence, and more specifically Deep Learning, has revolutionized computers' ability to interpret complex data [3]. Through Convolutional Neural Networks (CNNs), which are architectures inspired by the human visual cortex, deep learning models can learn hierarchies of features directly from raw data, such as the pixels of an image. This capability has enabled advancements in various areas of Computer Vision, including Object Detection, which consists of training algorithms to not only classify an image but also to identify the precise location of multiple object instances within it.

The application of object detection to aerial images captured by Unmanned Aerial Vehicles (UAVs) is an active and promising area of research, but it presents unique challenges. The detection of targets from an aerial perspective is hindered by factors such as variations in altitude and lighting, partial occlusion of objects by vegetation or structures, and, especially, the small size of the targets of interest (e.g., bottles, tires) [4]. The detection of small objects remains a frontier of research, as models can struggle to extract discriminative features from targets that occupy only a few pixels in the image.

In this context, this paper presents VISADE (Visual Identification System for Aedes Detection), a system that automates the detection of potential mosquito breeding sites through the collection and analysis of aerial images. Our approach uses a deep learning model, YOLOv11m, trained on a custom dataset to recognize objects that commonly serve as breeding sites. Unlike approaches that modify model architecture [5], [6], our main contribution investigates **the effectiveness of using high-resolution images as a practical and direct strategy to improve accuracy in small object detection**. This study quantifies the approach's pros and cons, offering insights for UAV-based surveillance.

This work is organized as follows. Section II presents related works that use similar approaches but with different models and datasets. Section III details the proposed methodology, including data collection, the theoretical basis of the detection model, and the evaluation protocol. In Section IV, the quantitative and qualitative results are presented. Section V conducts a critical analysis of these results, comparing them with the state of the art and discussing their implications. Finally, Section VI presents the conclusions and points to directions for future work.

II. Related Work

The automated detection of mosquito breeding sites using UAVs and deep learning has been gaining attention. Models from the YOLO (You Only Look Once) family are frequently employed due to their excellent balance between speed, accuracy, and ease of implementation. Laranjeira et al. [2] used YOLOv7 to detect and track mosquito hotspots, highlighting challenges with small objects and imbalanced classes in the public MBG dataset. Similarly, Perumal et al. [7] compared several models, including YOLOv8, on the same dataset and achieved a mAP of 0.92, using Generative Adversarial Networks (GANs) to increase the number of images in the 'pool' class, a technique to combat class imbalance.

The evolution of the YOLO family models has brought continuous improvements, and the scientific community has extensively explored how to adapt them for specific challenges. Gao et al. [5] proposed MultiDCCSP-YOLO, a model based on YOLOv9 with specific modules (Multi-DC Block) and a custom loss function (Adaptive Threshold Focal Loss - ATFL) to handle small objects in their custom dataset (PBSD), achieving a mAP@0.5 of 0.952. This type of approach focuses on modifying the network architecture to extract better features from small objects. Another successful line of research is the integration of attention mechanisms. The work of Apu et al. [6] with CBAM-enhanced YOLOv9, which integrates a Convolutional Block Attention Module (CBAM), reported a mAP@0.5 of 0.995 and an impressive mAP@0.5:0.95 of 0.864. This highlights that directing the model's attention toward the most informative regions significantly improves the accuracy of bounding box localization.

These studies highlight two main trends: the customization of architectures for specific problems and the importance of the quality and suitability of the dataset, which justifies our approach of creating our own dataset for the Brazilian context and investigating a more direct method (increasing the input resolution) to tackle the challenge of small objects.

III. Visual Identification System for Aedes Detection

The VISADE system was developed in three main stages, which will be detailed in this section: (1) data acquisition and dataset development; (2) definition and training of the detection model; and (3) the quantitative evaluation protocol.

A. Data Acquisition and Dataset Development

The images for the development of VISADE were captured using a DJI Mini 2 drone, equipped with a 12 Megapixel camera, operating at an altitude of 50 m over urban and peri-urban areas in Belo Horizonte/MG and Sete Lagoas/MG. The choice of this altitude represented a balance between covering a significant area per image and maintaining a sufficient Ground Sample Distance (GSD) for the identification of small objects.

We created a custom dataset since public ones lacked adequate object and scenario representation. The object classes were defined based on public health guidelines on the main artificial containers that accumulate water [1]: `bottle`, `bucket`, `pool`, `tire`, and `water_tanks`. The choice of these classes aims to cover a broad spectrum of common breeding sites in urban environments, from small containers to large water reservoirs.

The annotation of bounding boxes was performed manually using the web tool makesense.ai [8], which allows for a precise labeling process. The final dataset was divided into training (80%) and validation (20%) subsets. This division is a standard practice in machine learning, aiming to train the model on a larger portion of the data to learn patterns and evaluate its ability to generalize to an independent dataset, thus preventing overfitting. The absence of a separate test set is a limitation of the current scope, with results being reported on the validation set.

B. Object Detection Model: YOLO

The core of VISADE is YOLO (You Only Look Once), a family of single-shot object detection models introduced by Redmon et al. [9]. Unlike older two-stage approaches that first proposed regions and then classified them, YOLO treats detection as a single regression problem, going directly from image pixels to bounding box coordinates and class probabilities. To do this, it divides the input image into a grid, and each grid cell is responsible for detecting objects whose center falls within it. For each object, the model predicts the bounding box location (x, y coordinates, width, height), a confidence score (which indicates how certain the model is that the box contains an object), and the probabilities of belonging to each of the possible classes. This architecture allows YOLO to be extremely fast, making it ideal for real-time applications or analyzing large volumes of data.

For this work, the YOLOv11m model was chosen, using the implementation from the high-level Ultralytics library [10]. This library was selected because it is a state-of-the-art, well-maintained implementation that abstracts many of the complexities of the training pipeline and provides robust and optimized models, allowing the researcher to focus on experimentation and analysis. The model was initialized with weights pre-trained on the COCO dataset and then fine-tuned with our custom dataset for 100 epochs on an NVIDIA GeForce 2060 Super GPU. To handle the challenge of small objects, the input image resolution (`imgsz`) was set to 1200×1200 pixels. Due to GPU memory limitations, this high resolution required reducing the batch size to 8.

C. Quantitative Evaluation

The model's performance is measured by standard metrics in the field. A detection is considered a True Positive (TP) if it overlaps with a ground-truth annotation with an Intersection over Union (IoU) above a certain threshold, and the class is correct. The IoU measures the ratio of the overlap area to the total union area between the predicted box and the ground-truth box. A detection is a False Positive (FP) if the IoU is lower than the threshold or the class is incorrect. A False Negative (FN) occurs when a ground-truth object is not detected.

Precision (Equation 1) measures the accuracy of the detections, while Recall (Equation 2) measures the model's ability to find all relevant objects.

$$\text{Precision} = \frac{TP}{TP + FP} \tag{1}$$

$$\text{Recall} = \frac{TP}{TP + FN} \tag{2}$$

The main metric, mean Average Precision (mAP), is the average of Precision over multiple Recall levels, calculated by the area under the Precision-Recall curve. It provides a single, consolidated measure of performance. We calculate mAP with an IoU threshold of 0.5 (denoted mAP@0.5) and the average over multiple IoU thresholds, from 0.5 to 0.95 with increments of 0.05 (denoted mAP@0.5:0.95).

IV. Experimental analysis

In this section, we present the quantitative and qualitative results obtained by the trained VISADE model. All results were extracted from the validation set after 100 epochs of training.

A. Quantitative Results

The consolidated performance metrics of the model are presented in Table I. The model achieved high precision (0.912) and recall (0.904), resulting in a mAP@0.5 of 0.934.

Table I
Performance Metrics of the VISADE Model.

Metric	Value
Precision (B)	0.912
Recall (B)	0.904
mAP@0.5 (B)	0.934
mAP@0.5:0.95 (B)	0.505
Box Loss (Val)	1.63639
Class Loss (Val)	0.60404
DFL Loss (Val)	1.00854

Figure 1 illustrates the convergence of the loss functions on the validation set, showing that the model stabilized its learning over the epochs. Figure 2 shows the evolution of the mAP metrics, highlighting the rapid growth of mAP@0.5 and a slower growth and lower plateau for mAP@0.5:0.95.

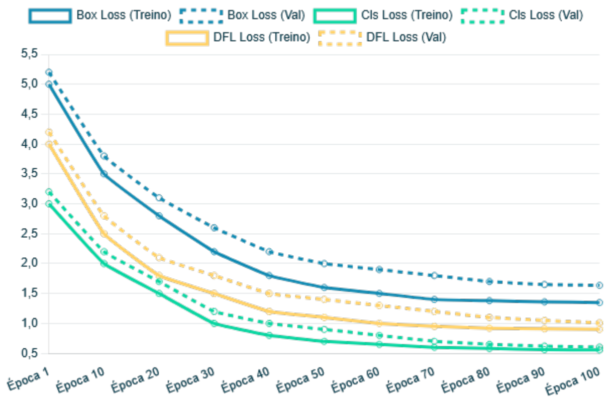

Figure 1. Validation loss curves (box, class, dfl) during the 100 training epochs.

B. Quantitative analysis

Table II positions the results of VISADE in relation to the works discussed in Section II, providing a clear overview of our system's performance when compared to the state-of-the-art.

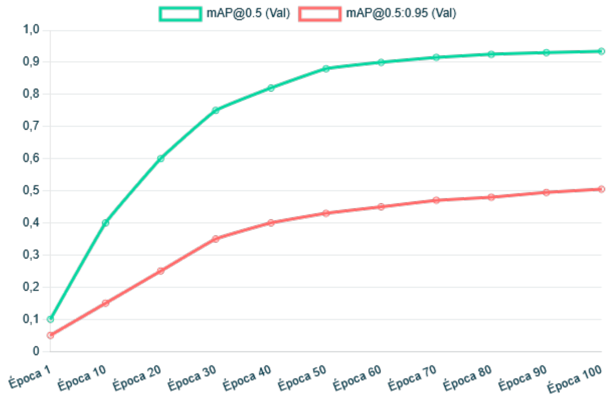

Figure 2. mAP@0.5 and mAP@0.5:0.95 curves on the validation set during training.

V. Discussion

The analysis of the results reveals a promising performance for the VISADE system, especially considering its practical objective. The high mAP@0.5 value (0.934), comparable to that achieved by Perumal et al. [7], demonstrates that the model is effective in identifying locations with potential dengue mosquito hotspots. This good performance can be attributed mainly to the creation of a custom and representative dataset for the application scenario, which reduces the "domain gap" between training and inference, as well as the use of high-resolution images (1200×1200), which provide sufficient visual details to distinguish objects even at an altitude of 50 meters.

Despite the robust performance in mAP@0.5, the observed difference in relation to mAP@0.5:0.95 (0.505) indicates that the precision in object localization can still be improved. This more rigorous metric evaluates both object presence and the precision of the bounding box alignment.. In the context of VISADE, however, this limitation has a reduced impact, as the main purpose of the system is to alert a health agent to the presence of possible hotspots, not necessarily to provide precise segmentation. Thus, mAP@0.5, which confirms the presence and approximate location of the object, proves to be more aligned with the application's objectives.

This limitation in bounding box precision may be largely related to the challenges in annotating small, irregularly shaped, or partially occluded objects, a recurring difficulty in object detection in aerial images [11]. In this sense, improvements in the annotation process could contribute to raising the mAP@0.5:0.95, although this gain is more relevant in scenarios that require greater accuracy in target delineation.

Finally, the qualitative results (Figure 3) reinforce these conclusions: most objects are correctly identified, but the boxes do not always accurately outline their boundaries. Works like that of Apu et al. [6], which used attention mechanisms (CBAM) to achieve a mAP@0.5:0.95 of 0.864, demonstrate that incorporating modules focused on refined feature extraction may be essential for gains in this aspect. Thus, while high resolution allows the model to "see better", the use of

Table II
PERFORMANCE COMPARISON OF VISADE WITH STATE-OF-THE-ART MODELS.

Feature	VISADE (YOLOv11)	Laranjeira et al. [2]	Perumal et al. [7]	Gao et al. [5]	Apu et al. [6]
Base Model	YOLOv11m	YOLOv7	YOLOv8	YOLOv9 (MultiDCCSP)	YOLOv9 (with CBAM)
Dataset	Custom	MBG	MBG (with GAN)	PBSD (Custom)	Custom (1500 img)
Input Resolution	1200×1200	960×960	640×640	960×960	800×600 (dataset)
mAP@0.5 (Overall)	≈ 0.934	N/R	0.92	0.952	0.995
mAP@0.5:0.95 (Overall)	≈ 0.505	N/R	N/R	N/R	0.864

architectural techniques like attention enhances its capability to localize objects with greater precision, being a promising direction for future evolutions of VISADE.

Figure 3. Example of model predictions on a validation batch. Correct detection of multiple objects is observed, but with variability in the precision of the bounding boxes.

VI. CONCLUSION

In this work, we presented VISADE, a visual detection system for *Aedes aegypti* breeding sites based on aerial images captured by drones and analyzed by a YOLOv11m model. The main contribution of this research lies in demonstrating that the use of high-resolution images, combined with a custom dataset representative of the Brazilian context, is a viable and effective strategy to increase the detection capability of small objects in complex urban environments.

The quantitative results show the system's effectiveness in the general identification of potential hotspots, with a mAP@0.5 of 0.934, a competitive value compared to the state of the art. On the other hand, the observed discrepancy in relation to mAP@0.5:0.95 (0.505) reveals challenges related to the precision of object localization, especially in contexts where targets are small, occluded, or have irregular contours, a recurring problem in aerial image detection tasks.

The qualitative analysis confirmed that, for the practical purpose of VISADE, which is to provide visual support for field actions by health agents, the approximate location of the objects is sufficient. However, we recognize that gains in precision can make the system even more reliable and scalable.

We conclude that VISADE represents a significant step in the use of accessible technologies, such as drones and deep learning, to support public health surveillance strategies. Its continuous development can contribute concretely to the early identification of breeding sites and, consequently, to the more effective control of epidemics transmitted by the *Aedes aegypti* mosquito.

VII. FUTURE WORK

As future directions, we propose the following improvements: (1) incorporating attention mechanisms, such as CBAM [6], with the goal of refining feature extraction and improving the precision of bounding boxes; (2) investigating more sophisticated loss functions, such as Wise-IoU [12], which consider the quality of box regression; (3) expanding the dataset, both in size and diversity, including underrepresented classes like `puddle`, as well as applying balancing and data augmentation strategies; and (4) fully integrating the georeferencing module, allowing for the automated generation of risk maps that can be used by public health managers in a practical and actionable way.

Another promising direction is the use of multiple UAVs working collaboratively. Cooperative strategies can significantly expand coverage and reduce mission time, allowing larger urban regions to be inspected simultaneously. Several approaches have been proposed in the literature, such as area partitioning based on Voronoi diagrams or polygon decomposition [13]–[15], optimized trajectory allocation using heuristics and evolutionary algorithms [16]–[18], and centralized or distributed communication architectures for mission coordination [19], [20]. In the context of VISADE, this would enable the development of a software platform for health agents where inspection routes are automatically divided among multiple drones, and the detections are aggregated into a single risk map, marked with georeferenced pins and image evidence.

ACKNOWLEDGMENT

The authors thank the Pontifícia Universidade Católica de Minas Gerais – PUC-Minas, Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES – (CAPES STIC-AMSUD 88887.878869/2023-00, 23-STIC-10 and Finance Code 001), the Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (Grants 407242/2021-0, 306573/2022-9, 442950/2023-3) and Fundação de Apoio à Pesquisa do Estado de Minas Gerais – FAPEMIG (Grant APQ-01079-23, APQ-05058-23 and PCE-00301-25).

REFERENCES

[1] Organização Mundial da Saúde, "Dengue and severe dengue," https://www.who.int/news-room/fact-sheets/detail/dengue-and-severe-dengue, 2025, acessado em: Jun 2025.

[2] C. Laranjeira, D. Andrade, and J. A. dos Santos, "YOLOv7 for Mosquito Breeding Grounds Detection and Tracking," in *arXiv preprint arXiv:2310.10423*, Oct 2023.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[4] W. Liu, W. Gu, F. Yang, S. Liu, H. Chen, and J. Liu, "RLRD-YOLO: Research on Lightweight Real-Time Detection Algorithm for Small Unmanned Aerial Vehicle Objects Based on Improved YOLOv8," *Sensors*, vol. 24, no. 7, p. 2199, 2024.

[5] Y. Gao, Z. Chen, J. Li, T. Wu, W. Xu, L. Wang, Y. Liu, and D. Chen, "Detection of Potential Breeding Sites Based on UAV Remote Sensing Imagery and MultiDCCSP-YOLO Network," *Remote Sensing*, vol. 16, no. 7, p. 1198, 2024.

[6] M. S. H. Apu, S. Ahmed, M. T. Ahmed, D. D. Gnakamene, M. Bhatt, S. Bhatia, S. Kadry, R. Kumar, and G. Meena, "Smart System for Real Time Monitoring and Diagnosis of Dengue Surfaces in Bangladesh Using CBAM-Enhanced YOLOv9," *Drones*, vol. 8, no. 5, p. 194, 2024.

[7] V. Perumal, R. Sasana, P. Rakshitha, and J. F. F. Doss, "Mosquito Breeding Grounds Detection Using Deep Learning Techniques," in *2023 International Conference on Advances in Computing, Communication and Applied Informatics (ACCAI)*, 2023, pp. 1–6.

[8] P. Skalski, "Make Sense," https://github.com/SkalskiP/make-sense/, 2019.

[9] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788.

[10] G. Jocher, A. Chaurasia, J. Qiu, and U. contributors, "Ultralytics YOLO," 2023. [Online]. Available: https://github.com/ultralytics/ultralytics

[11] T. Jiang and Y. Zhong, "ODverse33: Is the New YOLO Version Always Better? A Multi Domain benchmark from YOLO v5 to v11," *arXiv preprint arXiv:2502.14314*, Feb 2025.

[12] Z. Tong, Y. Chen, Z. Xu, and L. Lu, "Wise-IoU v3: A More Powerful Bounding Box Loss Function," *arXiv preprint arXiv:2301.10055v3*, Sep 2023.

[13] I. Maza and A. Ollero, "Multiple uav cooperative searching operation using polygon area decomposition and efficient coverage algorithms," in *Distributed Autonomous Robotic Systems 6*. Springer, 2004, pp. 221–230.

[14] F. Balampanis, K. Patsianis, E. Kosmatopoulos, and S. Chatzichristofis, "Coastal areas division and coverage with multiple uavs for remote sensing," *Sensors*, vol. 17, no. 4, p. 808, 2017.

[15] C. Gao *et al.*, "A hierarchical multi-uav cooperative framework for infrastructure inspection and reconstruction," *Control Theory and Technology*, vol. 22, pp. 394–405, 2024.

[16] M. Luna *et al.*, "Fast multi-uav path planning for optimal area coverage in aerial sensing applications," *Sensors*, vol. 22, no. 6, p. 2297, 2022.

[17] Y. Yu and S. Lee, "Efficient multi-uav path planning for collaborative area search operations," *Applied Sciences*, vol. 13, no. 15, p. 8728, 2023.

[18] H. Zeng *et al.*, "Multi-uav cooperative coverage search for various regions based on differential evolution algorithm," *Biomimetics*, vol. 9, no. 7, p. 384, 2024.

[19] F. Causa *et al.*, "Multi-drone cooperation for improved lidar-based mapping," *Sensors*, vol. 24, no. 10, p. 3014, 2024.

[20] M. Rahman *et al.*, "A survey on multi-uav path planning: Classification, algorithms, open research problems, and future directions," *Drones*, vol. 9, no. 4, p. 263, 2025.