Small Object Detection in Indoor Environment

Tan Teng Hui, 1211102289

Shahbe Mat-Desa, Wan-Noorshahida Mohd-Isa

Abstract

The detection of small objects indoors has various challenges, especially because of the low level of resolution and the diminished rate of detection accuracy. Although the area has been highly advanced in various ways, a lot of methods still have multiple complications.A key challenge in achieving real-time performance is minimizing computational costs while maintaining an acceptably high level of accuracy. Activation functions are essential for increasing computational efficiency and detection accuracy in YOLO-based methods. This paper evaluates a number of activation functions: SiLU, ReLU, ELU, LeakyReLU, and Sigmoid, and assesses their impact on the performance of YOLOv8 and YOLOv11 models in indoor environments. A custom dataset containing small objects such as Lego pieces, dice, and cubes has been created and labeled using bounding boxes. In the experiments, YOLOv8 with the ELU activation function has the best mAP50-95 performance, reaching the value of 0.801. The results of the experiment show that the choice of a proper activation function to be used for YOLO may significantly improve the possibility of detection accuracy.

Problem Statement & Objectives

Detecting small objects in indoor environments is a challenging task due to cluttered backgrounds, varying lighting, object occlusion, and low object resolution. Traditional detection methods often struggle to accurately identify and localize these small items, leading to poor performance in applications like robotics, smart homes, and surveillance. Despite advances in deep learning, existing models still face limitations in meeting the specific demands of small object detection indoors, highlighting the need for improved techniques tailored for such environments.

- To conduct a literature review on small object detection techniques and methodologies.
- To implement machine learning algorithms to tackle small object detection in an indoor environment.
- To analyze the results and discuss the findings.

Literature Review

- Efficient YOLOv8: Improved with bottleneck layers and anchor-free head for better small object detection. Achieved 45.9% mAP and 76.8% precision on the VisDrone dataset.
- DS-YOLO: Enhanced with multi-scale feature fusion and dynamic sampling. Gained over 4.9% increase in Recall and 4.2% improvement in mAP@0.5 on CrowdHuman dataset. 4.6% boost in Recall and 5% rise in mAP@0.5 on VisDrone2019 dataset
- Composite Backbone Model: Introduced Composite Dilated Convolution and Attention Module (CDAM) to combine contextual and attention-based features. Improved small object detection accuracy by 2.7% on the MS COCO dataset.

Research Methodology

Data Preparation & Preproccessing

A custom dataset comprising ten object classes including dice, Lego, cube, hexagon, cylinder, battery, screw, eraser, marble, and paperclip, was created using images sourced from Roboflow. All images were annotated with bounding boxes to ensure accurate representation of the target objects. To improve model robustness and detection performance under varying conditions, data preprocessing and augmentation techniques were applied, including random rotations, horizontal flips, and 90-degree rotations. The final dataset, consisting of 3,219 images, is divided into training, validation, and test sets using a 70:20:10 split ratio

Transfer Learning & Change of Activation Function

Transfer learning was used to take advantage of the knowledge gained when a model is trained on another dataset. The pre-trained YOLOv8 model was fine-tuned on the MS COCO combined with our custom dataset. First, the default training setting of the model was used, i.e., employing the default SiLU activation function. Further sets of experiments were conducted using YOLOv8 to observe the change in performance due to the variation of activation functions: ReLU, ELU, Leaky ReLU, and Sigmoid. Another set of experiments is also conducted to measure the difference in performance between YOLOv8 and YOLOv11. In contrast to YOLOv8, YOLOv11 was taken in its default architecture and its default activation function, without any changes.

Mathematical Formula for Different Activation Function

$$SiLU(x) = x \cdot \sigma(x)$$

$$ReLU(x) = \begin{cases} 0, x < 0 \\ 0, x \geq 0 \end{cases}$$

$$LeakyReLU(x) = \begin{cases} x, \ if\ x \geq 0 \\ negative\_slope * x, \ otherwise \end{cases}$$

$$ELU(x) = \begin{cases} x, \ x > 0 \\ \alpha * (exp(x) - 1), x \leq 0 \end{cases}$$

$$Sigmoid(x) = \sigma(x) = \frac{1}{1 + exp(-x)}$$

Evaluation Metrics

$$Precision = \frac{TP}{TP + FP}$$

$$Recall = \frac{TP}{TP + FN}$$

$$\text{mAP} = \frac{\sum_{i=1}^{K} \text{AP}_i}{K}$$

$$\text{F1Score} = 2 \cdot \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Results & Discussion

Comparison of Yolov8 Model with Different Activation Function and Different Epochs

TABLE 1 Validation results of all classes using YOLOv8 with different activation functions

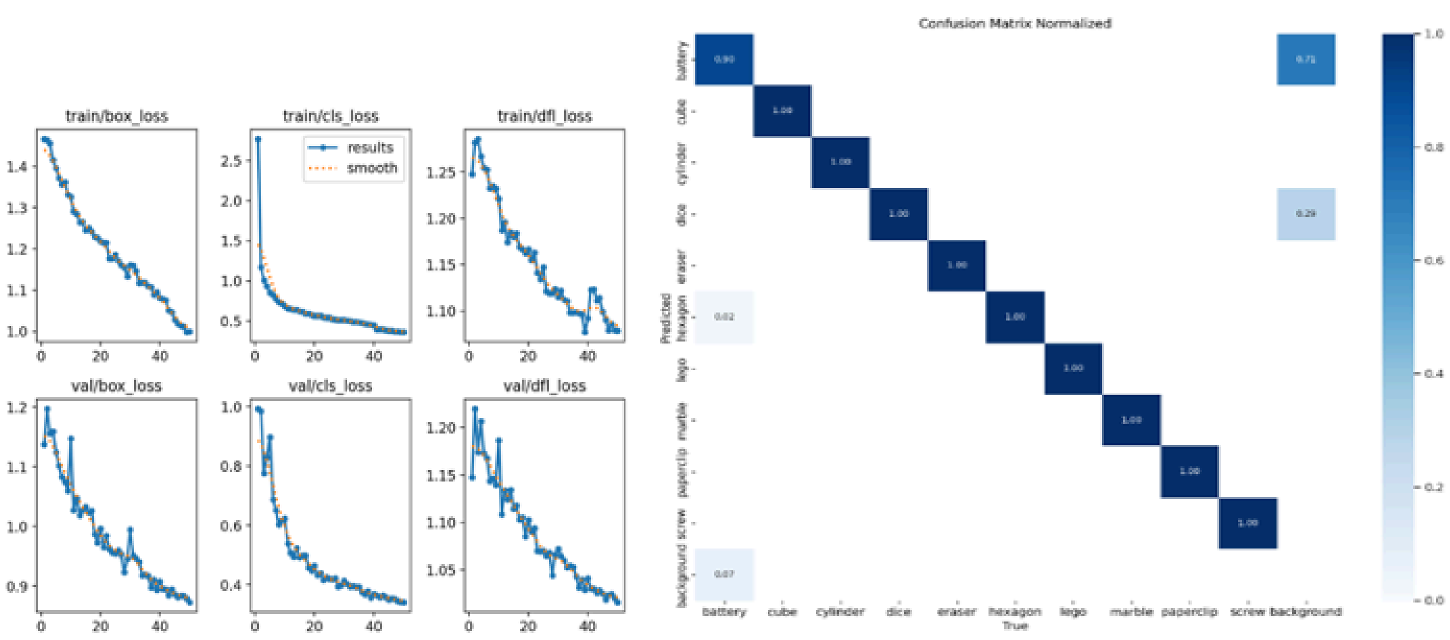
Activation function	Images	Instances	Measurement metrics at 20 epochs (above) and 50 epochs (below)			
			Precision	Recall (R)	mAP50	mAP50-95
SiLU (default)	379	661	0.990 **0.992**	0.984 0.988	0.991 0.990	0.793 0.793
ReLU	379	661	0.983 0.991	0.981 0.984	0.987 0.988	0.781 0.799
LeakyReLU	379	661	0.979 0.986	0.980 0.986	0.983 **0.991**	0.778 0.795
ELU	379	661	0.986 0.990	0.988 **0.990**	0.988 0.989	0.784 **0.801**
Sigmoid	379	661	0.943 0.972	0.944 0.982	0.962 0.979	0.702 0.735

Yolov11 model with default activation function and different epochs

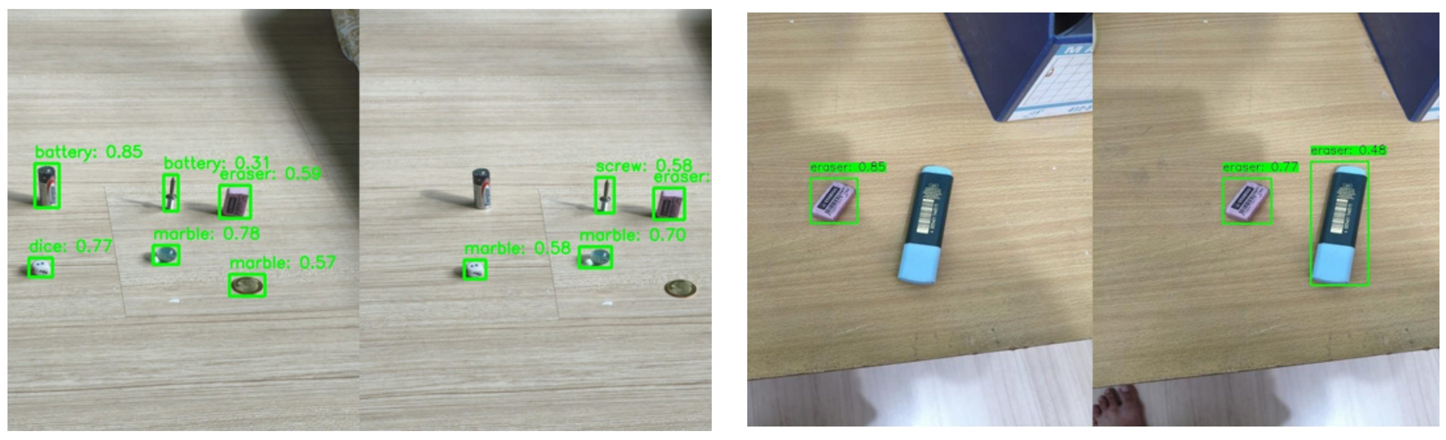
TABLE 2 Validation results of YOLOv11 model

Class	Images	Instances	Measurement metrics at 20 epochs (above) and 50 epochs (below)			
			Precision	Recall (R)	mAP50	mAP50-95
all	379	661	0.988 0.990	0.982 0.983	0.987 0.989	0.791 0.799

Loss of boxes, objects, and classes in best YOLOv8 model, (b) Confusion Matrix of Best YOLOv8 model



Comparison Results of Best and Worst YOLOv8 Model



Conclusion

In summary, the current research tested the effectiveness of the YOLOv8 object detection algorithm for detecting small objects in indoor environments with significant focus on performance evaluation by using different activation functions, namely ReLU, Leaky ReLU, ELU, and Sigmoid. A custom dataset with ten different types of small objects demonstrated that activation functions significantly contribute to the variation in detection accuracy as well as generalization performance. Among the tested versions, ELU consistently performed better than its counterparts in both quantitative measures (like mAP, precision, and recall) and qualitative evaluations of real test images.

The findings revealed that YOLOv8 using ELU showed greater robustness in challenging situations, such as cluttered environments, changing lighting, and partial occlusion, while models with Sigmoid activation often produced lower confidence scores and classification errors. A comparative analysis also highlighted the importance of activation functions in improving spatial localization and reducing false positive rates.

MMU MULTIMEDIA UNIVERSITY

FACULTY OF COMPUTING & INFORMATICS