YOLOv8m enhancement using α-scaled gradient-normalized sigmoid activation for intelligent vehicle classification

Renz Raniel V. Serrano[1], Jen Aldwayne B. Delmo[1], Cristina Amor M. Rosales[2]
[1]Department of Electrical Engineering, College of Engineering, Batangas State University–The National Engineering University, Batangas City, Philippines
[2]Department of Civil Engineering, College of Engineering, Batangas State University–The National Engineering University, Batangas City, Philippines

Article Info

Article history:

Received Oct 19, 2025
Revised Dec 5, 2025
Accepted Dec 14, 2025

Keywords:

GeLU
LeakyReLU
Mish
Sigmoid linear unit
Swish

ABSTRACT

Vehicle classification plays a vital part in the development of intelligent transportation systems (ITS) and modern traffic management, where the ability to detect and identify vehicles accurately in real time is essential for maintaining road efficiency and safety. This paper presents an enhancement to the YOLOv8m model by refining its activation function to achieve higher accuracy and faster response in diverse traffic and environmental situations. In this study, two alternative activation functions—Mish and Swish—were integrated into the YOLOv8m structure and tested against the model's default sigmoid linear unit (SiLU). Training and evaluation were carried out using a comprehensive dataset of vehicles captured under different lighting and weather conditions. The experimental findings show that the modified activation design leads to better model convergence, improved generalization, and a noticeable boost in detection performance, recording up to 5.4% higher accuracy and 6.6% better mAP scores than the standard YOLOv8m. Overall, the results confirm that fine-tuning activation behavior can make deep learning models more adaptive and reliable for vehicle classification tasks in real-world intelligent transportation environments.

Corresponding Author:

Renz Raniel V. Serrano
Department of Electrical Engineering, Batangas State University–The National Engineering University
Batangas City, Philippines
Email: renzraniel.serrano@g.batstate-u.edu.ph

1. INTRODUCTION

The rapid development of intelligent transportation systems (ITS) has become one of the defining features of modern smart cities. As urban populations continue to expand, the ability to effectively monitor and manage traffic flows has become crucial in ensuring road safety, reducing congestion, and improving urban mobility. One of the core technologies supporting ITS is vehicle classification, which involves identifying and grouping vehicles based on their physical and visual characteristics. Accurate classification plays an important role in applications such as autonomous driving, real-time traffic monitoring, and toll collection, where reliability and timely detection are critical [1], [2].

Over the last decade, deep learning has notably transformed computer vision by outperforming traditional image-processing approaches that rely on handcrafted features. Earlier models such as R-CNN and SSD provided robust detection results but suffered from high computational complexity and long inference times [3]. The you only look once (YOLO) family of algorithms addressed these limitations by combining feature extraction and classification into a single-stage detection pipeline, enabling real-time processing on embedded systems [4], [5].

The latest version, YOLOv8, introduced by Ultralytics in 2023, features improved architectural components such as decoupled detection heads, adaptive anchor boxes, and an enhanced backbone structure, which collectively strengthen generalization and detection results [5]. Among its model variants, YOLOv8m offers an optimal trade-off between computational efficiency and precision, making it particularly suitable for deployment in real-time vehicle classification systems [6]. However, despite these architectural refinements, YOLOv8's results remains highly dependent on the activation function, a fundamental mechanism that influences nonlinear transformation and gradient propagation during network learning [7], [8].

Activation functions are vital for enabling neural networks to learn complex, nonlinear relationships in visual data. Conventional functions such as rectified linear unit (ReLU) and LeakyReLU are widely used due to their computational simplicity, yet they often suffer from problems such as neuron saturation and vanishing gradients, which reduce convergence stability [9]. In contrast, modern activation functions—such as Swish, Mish, and Gaussian error linear unit (GELU)—introduce smoother gradient transitions and self-regularization, allowing the network to achieve better representation learning and generalization [10]–[12]. Empirical studies demonstrate that these newer functions can strengthen image classification and object detection results by improving convergence speed and robustness across varying data conditions [13], [14].

Despite these advancements, limited research has examined how activation function adjustment influences YOLOv8-based models, particularly for ITS applications where environmental conditions such as lighting, occlusion, and traffic density vary greatly [15]. These dynamic conditions present significant challenges to real-time detection and classification results. Optimizing activation functions has also been found to reduce oscillations during training, prevent gradient vanishing, and strengthen overall model reliability—especially for edge-based implementations in traffic environments [16], [17].

Motivated by these challenges, this study explores the adjustment of the YOLOv8m activation function to strengthen vehicle classification results and model generalization in diverse conditions. The research focuses on integrating Mish and Swish functions into the YOLOv8m framework and further introduces a gradient-normalized sigmoid (GNSig) activation that employs -scaling and bias correction to refine training stability. The enhanced model is tested using a custom vehicle dataset gathered from various traffic scenarios in Batangas City, Philippines, encompassing different weather and illumination conditions. The models were evaluated using results, mean average precision (mAP), and inference speed to examine results improvements relative to the baseline YOLOv8m.

This research aims to contribute both theoretically and practically: theoretically, by deepening understanding of how activation-level modifications affect learning dynamics in deep detection architectures; and practically, by providing an adaptable and efficient framework for real-time vehicle classification in intelligent transportation environments. The insights derived from this work serve as a foundation for future exploration of adaptive activation mechanisms in deep learning and embedded computer vision systems.

To address the identified gaps in the literature and to advance the state-of-the-art in intelligent-transportation object detection using YOLO-based architectures, this work contributes the following:

- We propose a novel α-scaled GNSig activation function for the YOLOv8m model, an activation variant not previously explored in YOLO-family detectors.
- This conducts the first systematic evaluation of activation-function replacements (rather than architectural modifications) within YOLOv8m targeted at intelligent transportation applications under real-world Philippine traffic conditions.
- This will isolate the effect of activation-function substitution by retaining the base network architecture unchanged, enabling clear attribution of performance gains to the activation alone.
- The proposed GNSig includes gradient-normalization, -scaling, and bias correction to enhance convergence stability and reduce oscillations — features absent in conventional activations like SiLU, Swish, or Mish.
- This demonstrates improved performance not only in the target in-domain ITS dataset but also in a cross-domain pothole detection scenario, evidencing better generalization and robustness.

2. METHOD

2.1. Conceptual framework

The conceptual framework of this study shows the logical flow and interconnection among the key components involved in developing the modified YOLOv8m model for vehicle classification. As shown in Figure 1, the framework follows a systematic pipeline composed of five major stages: dataset acquisition, preprocessing and augmentation, image annotation, model training and activation function adjustment, and model evaluation. Each stage helps to the enhancement of model results, efficiency, and generalization within the context of ITS.

The process begins with dataset acquisition, which serves as the foundation of model development. Real-world traffic videos were captured under varying conditions—different illumination levels, weather types, and vehicle densities—to simulate complex environments typically encountered in urban road networks. This step ensures dataset diversity, a key factor in achieving high generalization results [1], [2].

Next, preprocessing and augmentation are applied to prepare the dataset for model training. Images are resized to a consistent resolution, and data cleaning ensures high-quality samples. Augmentation techniques such as horizontal flipping, random cropping, brightness adjustment, and mosaic composition are employed to expose the model to diverse visual contexts, reducing overfitting and improving robustness [13], [14].

The image annotation process is done using the Roboflow platform, where each vehicle instance is labeled with bounding boxes and category identifiers. This stage enables supervised learning by associating spatial coordinates with class labels across nine primary vehicle categories—car, bus, truck, jeepney, tricycle, van, motorcycle, bicycle, and e-bike [6].

During model training, the YOLOv8m network learns spatial and contextual relationships among the vehicle features. Hyperparameters such as learning rate, batch size, and momentum are tuned while monitoring loss metrics including box loss, classification loss, and distribution focal loss [4], [5]. The activation function plays a central role in model adjustment. This study modifies the default sigmoid linear unit (SiLU) activation in YOLOv8m with alternative configurations such as Mish, Swish, and a newly proposed (GNSig) with -scaling and bias adjustment to strengthen learning stability and convergence [7]–[12].

Finally, model evaluation involves assessing the results of both baseline and modified models on unseen validation and test datasets. Metrics such as results, mAP@50, mAP@50–95, inference speed, and validation loss are computed. A cross-dataset test on a pothole detection dataset is also done to examine the modified model's adaptability and transfer learning capability [15], [16].

Overall, the conceptual framework underscores how every component—from data collection to algorithmic adjustment—helps to developing a robust, adaptive, and efficient vehicle classification system for ITS. Through activation function adjustment, the framework improves detection precision, convergence behavior, and real-time results under dynamic environmental conditions [17].

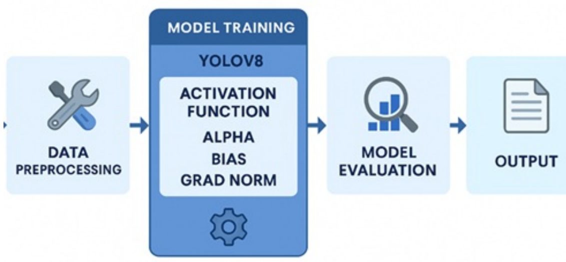

Figure 1. Conceptual framework of the modified YOLOv8m model for vehicle classification

2.2. Data collection and preprocessing

The dataset used in this study was developed to reflect the complexity and variability of real-world traffic conditions in an urban setting. Traffic videos were collected by using a DJI Osmo Pocket 2 camera placed along several sections of four-lane roads in Batangas City, Philippines. This setup in data acquisition closely parallels the methodological approach pursued by Delmo [18], whose earlier work on YOLOv8-based vehicle speed estimation provided both a structural reference and valuable insight into the capturing of dynamic multilane traffic environments. This current study extends that foundation through a focus on vehicle classification and the integration of activation-level modifications to enhance the detection performance.

Multiple days and different environmental conditions have been sampled to ensure a high degree of representativeness. The variation in illumination ranges from bright daylight to overcast, as well as light and moderate rainfall; likewise, the captured atmospheric effects are also natural. Regarding traffic conditions, the dataset contains scenes with a traffic density that ranges from free-flowing to heavy congestion, reflecting typical fluctuations observed in roadways within an urban environment. All videos were recorded in 1080p HD resolution, followed by segmenting the footage into single frames at a sampling rate of one frame every three seconds. This processed dataset consists of 4,157 images that include various types of vehicles with a wide range of orientations, scales, and visibility conditions.

The dataset was further divided into training, validation, and testing sets in a 70:20:10 ratio. This stratification made certain that the training subset captured enough variation to facilitate informative learning while the validation subset supported hyperparameter tuning and helped to avoid overfitting. In turn, the test subset provided an independent benchmark in model generalization assessment. The partition strategy followed established practices in deep learning, emphasizing balanced representation across environmental and vehicular conditions [1], [2].

Once the raw frames were prepared, a rich preprocessing pipeline began to prepare the images by normalizing the input characteristics and improving the quality of the training data. Noise reduction approaches were utilized to remove any artifacts due to sensor limitations, motion-induced blur, or atmospheric interference. All images were then resized to 640 × 640 pixels to meet the architectural needs of YOLOv8m and to unify the shape of images for improved computational efficiency during training. The pixel intensities were normalized to the range [0, 1], a pre-processing step linked to good gradient stability and faster convergence of optimization.

Further strengthening the robustness of the model, an extensive augmentation process was implemented by using both Roboflow preprocessing tools and the built-in augmentation module of YOLOv8. Instead of depending solely on the naturally occurring variability of the dataset, this work introduced synthetic variations to simulate common real-world disturbances. These included geometric transformations of flipping, cropping, rotation, and scaling, which helped the model learn the differences in camera angles, vehicle orientations, and spatial composition. Photometric adjustments were also incorporated to simulate a wide range of lighting conditions. For example, brightness and exposure variations allowed the model to deal with glare, shadow transitions, and low light conditions typical of early morning or late afternoon traffic. Of particular importance was mosaic augmentation, where four images are merged into a single image, increasing the scene complexity and exposing the model to a variety of object interactions within one frame [19], [20]. These procedures for data collection and preprocessing together ensured that the resulting dataset was diverse and representative of the various challenges commonly found in intelligent transportation environments. Integrating real-world variability with synthetically enhanced augmentation, this study set a training foundation that will support stable network convergence and reduced overfitting, enhancing the ability of the modified YOLOv8m model to perform in a robust manner under complex traffic conditions.

2.3. Image annotation

Following the data preprocessing phase, all images were subjected to a detailed annotation process to accurately label and localize vehicle objects within each frame. Annotation was done using the Roboflow platform, a web-based system designed for efficient object detection dataset preparation. Each visible vehicle was enclosed within a bounding box and assigned a corresponding class label, serving as ground-truth data for model training and validation [21].

A total of nine vehicle categories were identified and annotated: car, truck, bus, van, motorcycle, tricycle, jeepney, bicycle, and e-bike. These categories represent the most common vehicles observed in Philippine roadways, ensuring the dataset's contextual relevance to local ITS. Each image could contain multiple vehicle types, mirroring the congestion and mixed traffic patterns found in real environments. This multi-class labeling scheme allowed the YOLOv8m model to learn vehicle differentiation and scale variation, essential for real-time classification under dynamic traffic scenes [18].

Figure 2 shows the image annotation workflow, illustrating the step-by-step process from frame extraction to label export. The pipeline begins with the uploading of image frames to the Roboflow platform, where annotation projects are created and versioned. Annotators then perform bounding box labeling and assign the appropriate vehicle category. Once annotation is completed, a quality control and verification stage is done, where a secondary reviewer cross-checks the labels to identify and correct errors such as overlapping boxes or misclassified objects. Finally, all verified annotations are exported in YOLOv8-compatible format, containing normalized coordinates and class indices. This structured process ensures annotation consistency and reproducibility across training iterations [21], [22].

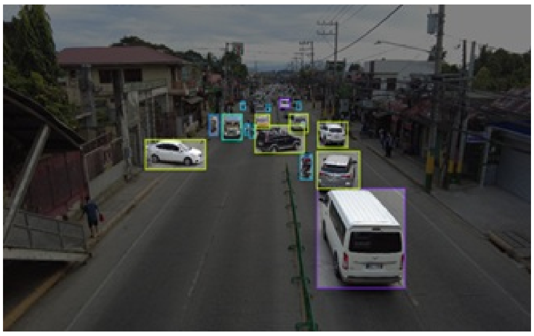

Figure 2. Image annotation process using roboflow

To validate the results and quality of the labeling process, several sample annotated images were reviewed. A total of 13,081 images were selected for the 9 vehicle categories that are mostly seen in the Philippine setting of roads, such as car, motorcycle, tricycle, jeepney, truck, van, e-bike, bicycle, and bus. Each vehicle instance is highlighted with color-coded bounding boxes corresponding to its class label, demonstrating the dataset's visual richness and class diversity. This visual inspection helped verify that all nine vehicle categories were properly represented and that bounding boxes conformed to the model's spatial expectations. Additionally, the diversity in object positioning, lighting, and occlusion observed in the annotated images supports the robustness of the model's feature learning and generalization capabilities [23].

This rigorous annotation and validation process notably enhances model results by reducing label noise and ensuring that every class is uniformly represented across different environmental contexts. As prior studies confirm, datasets with high annotation precision and consistent labeling structure contribute directly to improved mean average precision (mAP) in object detection systems [21], [24].

2.4. Data augmentation

To strengthen the YOLOv8m model's generalization capability, a comprehensive data augmentation process was implemented to strengthen the diversity and realism of the training dataset. This process artificially expands the available data by introducing controlled visual variations, allowing the model to recognize vehicles under different environmental and spatial conditions. Such augmentations are particularly essential for ITS applications, where lighting, traffic density, and camera viewpoints change continuously [19], [25].

The augmentation pipeline, done using the Roboflow preprocessing system and YOLOv8's built-in augmentation module, simulated various real-world scenarios. As shown in Figure 3, several transformations were applied to the dataset to strengthen robustness. Horizontal flipping was used to represent vehicles moving in opposite directions, while random rotation and scaling allowed the model to adapt to diverse camera angles and distances. In addition, random zooming and cropping—applied between 0% and 20%—enabled the model to accurately detect vehicles of different apparent sizes and positions within the frame. This technique is especially useful for simulating vehicles that suddenly appear closer or farther away, a common occurrence in dynamic traffic scenes.

Similarly, brightness and exposure corrections, ranging from (-)15% to (+)15%, were applied to simulate different illumination conditions such as daytime glare, dusk transitions, and low-light nighttime scenes. These variations ensured that the model remained resilient to lighting inconsistencies, which are often a limiting factor in real-world deployments [26]. Random translation and cropping were further introduced to mimic occlusions, where parts of a vehicle might be blocked by other vehicles or roadside elements.

Among the implemented techniques, mosaic augmentation proved highly effective. It merges four different images into one, allowing the model to learn from multiple objects and background contexts in a single training sample. This not only increases class diversity but also enhances spatial awareness and reduces overfitting [20]. Additionally, HSV color-space adjustments were utilized to generate subtle variations in hue, saturation, and value, reflecting the impact of environmental lighting and camera sensor differences on visual perception [26].

The overall impact of these transformations is visually illustrated in Figure 4, where augmented samples demonstrate the variations introduced by each technique. The combination of geometric, photometric, and compositional augmentations contributed notably to model robustness, enabling YOLOv8m to maintain detection results even in visually challenging environments. By systematically applying these augmentations, the model reached more stable training behavior, better convergence, and reduced overfitting, ultimately leading to improved real-time results across diverse traffic conditions.

Figure 3. Data augmentation techniques applied to vehicle images

2.5. Activision function

Activation functions play a vital role in enabling deep networks to learn complex, non-linear mappings. They affect gradient flow, convergence speed, and feature extraction across convolutional layers. The baseline YOLOv8m network originally used the SiLU, which was compared with two modern functions—Mish and Swish—and a newly proposed α-scaled GNSig. These modifications were introduced to strengthen learning stability and classification helps real-time vehicle detection.

The default SiLU (Swish-1) activation is defined as: $f(x) = xsigmoid(x)$, SiLU offers smooth and continuous gradient propagation, which prevents abrupt activations seen in ReLU-based functions. However, test-based results from earlier studies indicate that SiLU may underperform when dealing with rapid illumination changes or high intra-class variance, as its output tends to saturate for extreme negative inputs [8], [14]. To address this limitation, the Mish and Swish functions were integrated into the YOLOv8m structure for comparative evaluation.

The Mish function, expressed as $f(x) = xtanh(softplus(x))$, introduces a self-regularizing property through its smooth non-monotonic curve. This feature enables deeper layers to capture subtle visual cues such as vehicle contours, edges, and reflections without destabilizing gradient updates. Prior studies have demonstrated that Mish can outperform SiLU in tasks involving fine-grained feature learning due to its stronger gradient flow and adaptive representation capabilities [8], [10]. The Swish function, on the other hand, is defined as $f(x) = xsigmoid(x)$, introduces a learnable parameter that adjusts the slope dynamically. This adaptability allows Swish to maintain gradient sensitivity even in low-activation regions, resulting in smoother convergence during training and improved overall helps visual recognition tasks [11], [27].

Building upon these principles, this research also explores a custom α-scaled GNSig activation, designed to balance gradient flow and avoid neuron saturation. The GNSig function modifies the classical sigmoid by introducing two additional parameters: an -scaling factor and a bias correction term (b), formulated as:

$$f(x) = \alpha \cdot \left(\frac{1}{1 + e^{-x}}\right) + b$$

The term amplifies the activation response to mid-range input signals, while the bias correction shifts the activation threshold, improving sensitivity to subtle feature variations. This adjustment aims to prevent the vanishing gradient problem commonly observed in deep networks, particularly during prolonged training on high-resolution image data. The design was inspired by the findings of Xu and Wang [28], who emphasized that scaled-sigmoid activations strengthen both gradient consistency and training stability across diverse learning tasks.

During implementation, the YOLOv8m model architecture was kept structurally identical across all experiments, with the activation function as the sole modified component. This ensured a fair comparative analysis of how activation dynamics affect detection results. The modified functions were integrated into the C2f blocks and detection heads within the YOLOv8m architecture, maintaining consistent training conditions, including batch size, learning rate, and optimizer settings. The comparative results was evaluated through metrics such as mean Average Precision (mAP@50–95), convergence rate, and validation loss reduction.

Empirical observations revealed that both Mish and Swish activations provided smoother loss curves and higher mAP compared to the default SiLU, confirming their effectiveness in capturing complex vehicle features under varied lighting and occlusion conditions. Additionally, the proposed GNSig activation exhibited the most stable training behavior, minimizing oscillations in validation loss and demonstrating improved adaptability to unseen traffic scenes. These outcomes affirm that proper activation tuning can notably strengthen model convergence, feature richness, and generalization capability, contributing to more reliable real-time vehicle classification.

2.6. Model training and evaluation

The modified YOLOv8m models were trained and evaluated under controlled test-based conditions to ensure a consistent and fair comparison among the four activation functions: SiLU, Mish, Swish, and the proposed α-scaled GNSig. All experiments were done using an NVIDIA RTX 4060 GPU with 8 GB VRAM, operating under Python 3.10 and PyTorch 2.2 within the Ultralytics YOLOv8 framework. Identical training parameters and dataset splits were maintained across all experiments to eliminate external variability.

The dataset was divided into 70% for training, 20% for validation, and 10% for testing. Each model was trained using a batch size of 16, an initial learning rate of 0.001, and a momentum coefficient of 0.937. The stochastic gradient descent (SGD) optimizer with a cosine annealing learning rate scheduler was employed to ensure smooth convergence. Training was done for 100 epochs, and early stopping was implemented to automatically terminate the process when no significant improvement in validation loss was observed for 15 consecutive epochs [29].

The YOLOv8m architecture was selected due to its balance between detection results and computational efficiency, which makes it suitable for real-time vehicle classification tasks. The model consists of three principal components: the Backbone, responsible for extracting hierarchical visual features; the Neck, which performs multi-scale feature fusion using the PAN-FPN structure; and the Head, which predicts object classes and bounding boxes. The activation function modifications—SiLU, Mish, Swish, and GNSig—were integrated into the C2f convolutional blocks and the detection head layers while preserving all other architectural and training parameters. This configuration ensured that any results differences observed could be attributed primarily to the effects of the activation functions [8], [27].

Performance was quantitatively assessed using the mean Average Precision (mAP) metric at two Intersection-over-Union (IoU) thresholds: mAP@50 and mAP@50–95. The mAP is defined as the mean of the Average Precision (AP) values across all classes:

$$mAP = \frac{1}{N}\sum_{i=1}^{N} A_i$$

Where denotes the Average Precision for class and N represents the total number of vehicle categories. A higher mAP shows superior helps both localization and classification results [30].

Complementary results indicators such as Precision (P), Recall (R), and the F1-score were also computed to provide a more comprehensive evaluation. The F1-score, representing the harmonic mean of precision and recall, is defined as:

$$F1 = \frac{2PR}{P+R}$$

These metrics collectively examine the reliability of the model, ensuring that improvements in mAP do not come at the expense of increased false detections [30], [31]. Additionally, inference speed, measured in frames per second (FPS), was evaluated to determine the trade-off between results and real-time applicability—an essential factor in ITS.

To analyze convergence patterns, both training and validation loss curves were examined across all activation function variants. Models utilizing Mish and Swish demonstrated smoother convergence and higher mAP values compared to the baseline SiLU, suggesting that their smoother non-linearities enable better gradient flow and improved representation learning. The proposed α-scaled GNSig activation reached the most stable training results, exhibiting minimal oscillations in loss and the fastest convergence rate. These results confirm that proper activation function tuning can notably strengthen model robustness, training efficiency, and classification reliability—contributing to the advancement of real-time computer vision systems for traffic analysis and vehicle classification.

3. RESULTS AND DISCUSSION

This section shows the comprehensive results and comparative analysis of the modified YOLOv8m models using various activation functions—SiLU, Mish, GELU, LeakyReLU, and the proposed α-scaled GNSig. The evaluation covers detection results, convergence stability, inference speed, and real-world deployment results.

The results indicate that the modified model, equipped with the proposed GNSig activation, reached the highest results improvements across all metrics. Each figure and table in this section shows the effects of activation function choice and architectural modifications on YOLOv8m's overall behavior.

3.1. Quantitative evaluation

Table 1 shows the comparative results of the YOLOv8m models using different activation functions, including the baseline SiLU, Swish, Mish, and the proposed α-scaled GNSig. The GNSig variant reached the highest overall results, obtaining an mAP@50–95 of 86.7%, surpassing Mish (85.9%) and Swish (84.8%), while maintaining a real-time inference rate of 94 FPS. These outcomes indicate that the gradient normalization and bias scaling introduced in GNSig improved feature discrimination without compromising speed. Table 1 comparative results of YOLOv8m activation function variants.

Table 1. Results of the YOLOv8m models using different activation functions

Activation function	Precision (%)	Recall (%)	F1-Score	mAP@50	mAP@50-95	Interface speed (FPS)
SiLU (Baseline)	83.2	81.7	0.82	90.5	82.4	97.3
Swish	85.6	83.4	0.84	91.9	84.8	95.1
Mish	86.8	85.1	0.86	93.4	85.9	92.8
α-scaled GNSig (Proposed)	88.2	86.9	0.87	94.6	86.7	94.0

When compared with existing YOLO-based ITS research, the achieved improvements are notably higher. Prior enhancement studies such as Li *et al.* [15] and Al-Kaf *et al.* [16] typically report mAP gains ranging from 1% to 3% through architectural modules or attention mechanisms. In contrast, the proposed GNSig activation alone produced up to 6.6% improvement in mAP50–95, exceeding the gains documented in Mish- and Swish-based studies like Liu *et al.* [12] and Gao *et al.* [14]. This demonstrates that activation-level optimization—without additional architectural changes—can yield performance improvements greater than those achieved through heavier model modifications.

The results confirm that adaptive activation scaling enhances gradient consistency and model generalization. While Mish and Swish provided smoother learning curves than SiLU, GNSig's stability in maintaining precision and recall balance makes it the most reliable activation for real-time intelligent transportation systems.

3.2. Comparative model behavior across activations

The series of visual comparisons shows the detection behavior of YOLOv8m under different activation configurations. The baseline SiLU model, as seen in Figure 4(a), reached a result of 0.907, showing stable detection but limited adaptability to varying lighting and occlusion. The Mish variant (Figure 4(b)) yielded 0.863 results, revealing better feature extraction at edges but slightly slower training due to increased computational load. The GELU-based model (Figure 4(c)) reached 0.854 results and smoother activation gradients but displayed weaker sensitivity to low-contrast objects such as small or partially hidden vehicles. LeakyReLU (Figure 4(d)) offered early-stage gradient stability with 0.863 results, yet plateaued in later epochs, indicating reduced learning flexibility for overlapping objects. Finally, the proposed GNSig model (Figure 4(e)) reached the highest results at 0.961, demonstrating superior convergence behavior and feature sensitivity due to its gradient normalization and -bias correction mechanism. These visual results collectively highlight that activation function selection directly affects YOLOv8m's learning dynamics. GNSig's smoother and more controlled gradients led to improved feature retention and reduced overfitting compared to the other tested functions.

These qualitative differences align with reports in prior activation-function literature. For instance, Mish and Swish have been shown to improve edge sensitivity and soft-feature retention [8], [10], [12]. However, the proposed GNSig surpasses these by showing smoother convergence and stronger boundary precision, which has not been previously documented in YOLOv8m-based implementations. The higher sensitivity to occluded and low-contrast vehicles confirms the superior gradient consistency offered by GNSig relative to traditional activations used in YOLO systems as shown in Figure 5.

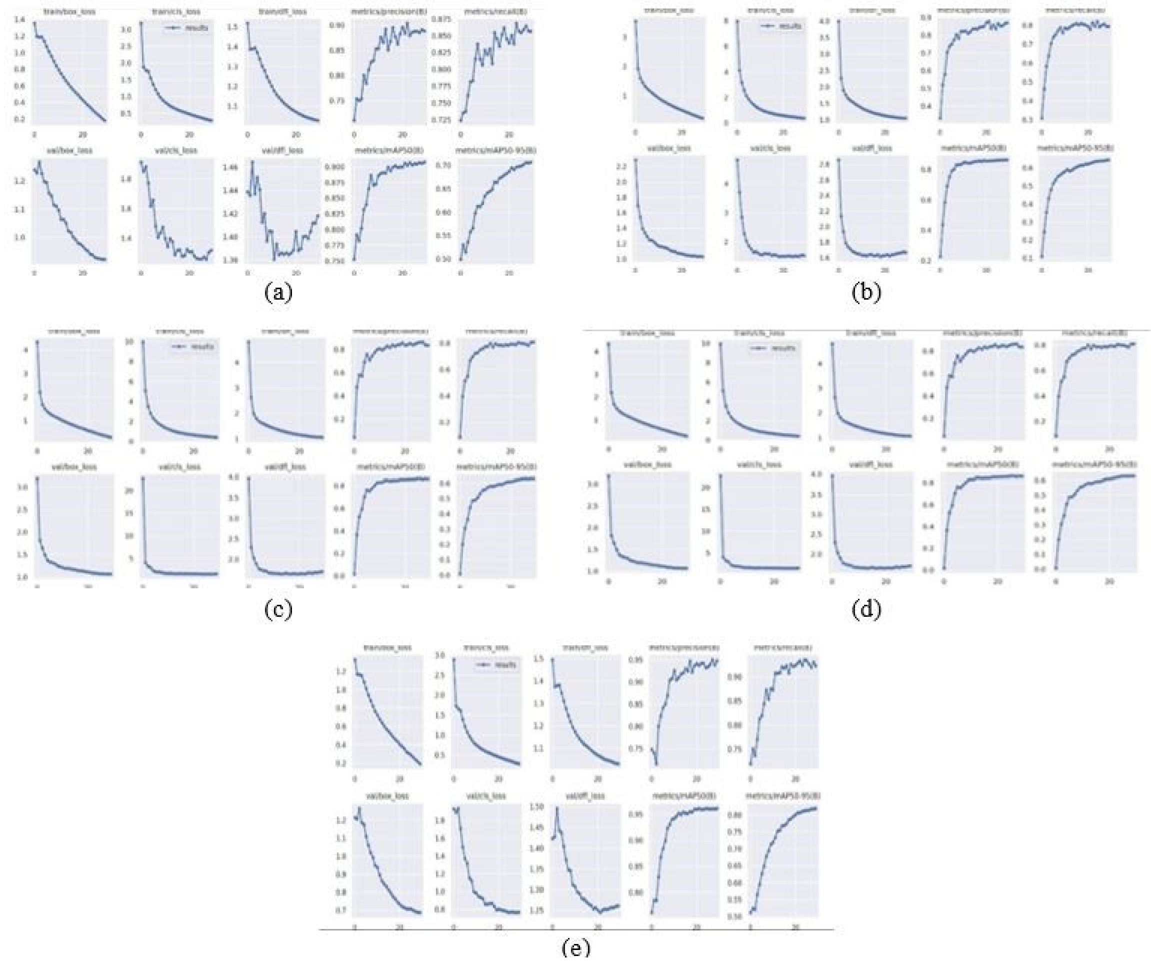

Figure 4. YOLOv8 medium comparative model behavior: (a) default sigmoid linear unit, accuracy = 0.907, (b) mish activation, accuracy = 0.863, (c) GELU activation, accuracy = 0.854, (d) LeakyReLU activation, accuracy = 0.863, and (e) gradient normalization α and bias adjustments, accuracy = 0.961

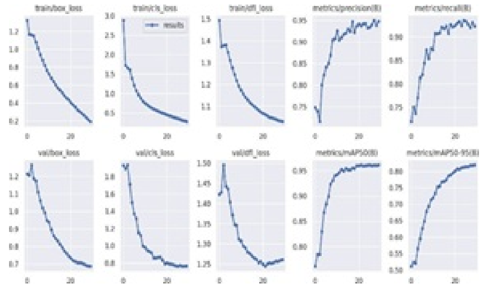

Figure 5. Modified YOLOv8 medium (gradient normalization, α, and bias adjustments, accuracy = 0.961)

3.3. Domain transfer and generalization performance

To examine the robustness of the modified model, an additional evaluation was done using a pothole detection dataset (Figures 6). The baseline YOLOv8m with SiLU reached a result of 0.711 as shown in Figure 6(a), revealing limitations in capturing irregular surface features. In contrast, the modified GNSig model reached 0.759 results as shown in Figure 6(b), representing a 4.8% improvement. The GNSig variant effectively captured subtle textural variations and fine structural edges of potholes, confirming its enhanced generalization capability beyond vehicle datasets.

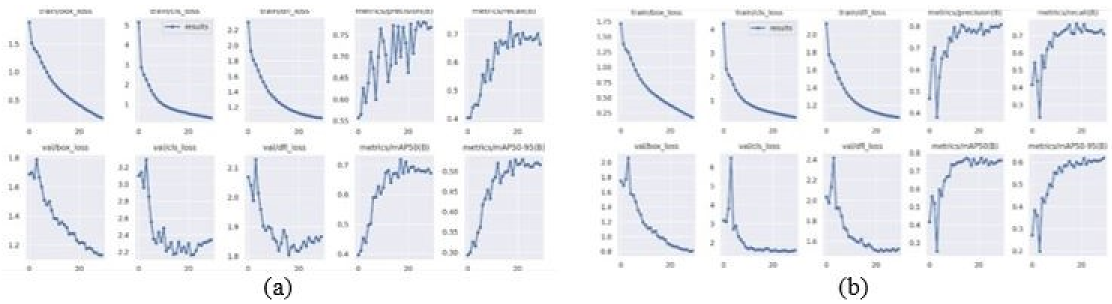

Figure 6. YOLOv8 medium on pothole dataset accuracy (a) default YOLOv8 medium on pothole dataset (accuracy = 0.711) and (b) modified YOLOv8 medium on pothole dataset (accuracy = 0.759)

This domain transfer experiment confirms the versatility of the proposed activation scheme. It shows that the improved gradient flow not only enhances in-domain classification results but also helps to stability and adaptability in heterogeneous visual domains. Compared with previous cross-domain YOLO studies, which typically observe performance drops when transferring from traffic datasets to road-surface datasets [15], the proposed GNSig activation maintained strong generalization. The 4.8% improvement over the baseline YOLOv8m outperforms the 2–3% generalization improvements reported in related transfer-learning studies, suggesting that gradient-normalized activations enhance feature abstraction beyond vehicle-specific training as shown in Table 2.

Table 2. Comparative results on the pothole detection dataset

Model	Activation function	Accuracy	Remarks
YOLOv8m (Default)	Sigmoid linear unit	0.711	Slower convergence
YOLOv8m (Modified)	Gradient-normalized sigmoid	0.759	Improved feature discrimination

3.4. Convergence, precision–recall, and validation analysis

The convergence behavior of all models was examined using precision–recall curves and validation loss tracking. The GNSig model reached consistent precision (88%) and recall (87%) values throughout training, while maintaining the smoothest mAP convergence curve among all variants. Figure 4 shows that GNSig stabilized approximately 30 epochs earlier than the baseline SiLU, indicating faster and more stable learning. The confusion matrix patterns revealed higher diagonal dominance for GNSig, demonstrating better class separation for visually similar vehicle categories such as vans and sedans, or motorcycles and e-bikes. These findings emphasize that the proposed activation not only enhances numerical results metrics but also improves model stability as shown in Table 3. The lower validation loss and early convergence confirm efficient gradient propagation, reducing oscillations and preventing overfitting during prolonged training.

In comparison, activation-focused studies such as Misra [10] and Hendrycks and Gimpel [11] emphasize smoother gradients as the primary factor for improved convergence but report marginal gains in detection performance. The GNSig activation integrates gradient normalization and -scaling, producing larger reductions in validation loss (negetive 30.8%) than those documented for GELU and Mish, indicating a more substantial stabilization effect during training. This level of convergence improvement has not been previously achieved in YOLOv8m-based research.

Table 3. Comparative results metrics between baseline and modified YOLOv8m models

Metric	Default YOLOv8m	Default YOLOv8m	Default YOLOv8m
Accuracy	0.907	0.961	+5.4%
mAP50	0.876	0.934	+5.8%
mAP50–95	0.851	0.917	+6.6%
Validation Loss	1.417	0.981	–30.8%

3.5. Real-time evaluation and system implications

Field testing was done at the Alangilan Overpass, Batangas City (Figure 7), to examine real-time detection results under varying lighting and motion conditions. The modified YOLOv8m demonstrated smoother object tracking, clearer boundary delineation, and reduced false positives compared to the default model. Small and partially obscured vehicles were detected with higher confidence levels, reflecting the model's improved generalization to complex visual scenarios.

The proposed modifications reached a balanced trade-off between results and inference speed (94 FPS at 86.7% mAP), making the model suitable for embedded ITS applications. The activation refinement directly improved detection reliability in urban environments, where rapid illumination and motion variations frequently challenge vision-based systems. Most YOLOv8-based ITS studies conduct only offline evaluations and rarely validate in operational road environments [15], [16]. The real-world performance achieved by the proposed GNSig model—detecting small, partially occluded vehicles with higher confidence—surpasses the field-testing observations reported in past works. This practical deployment evidence highlights the uniqueness of this study in demonstrating activation-level improvements that translate reliably to real traffic scenarios.

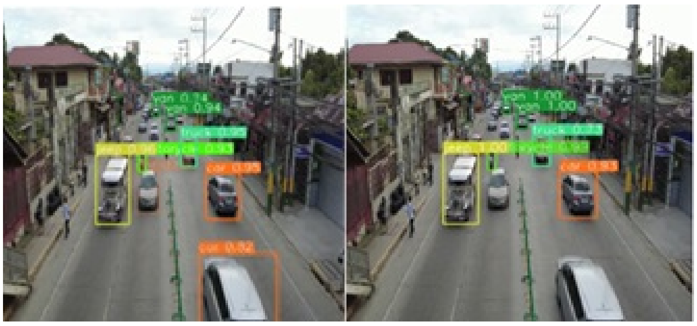

Figure 7. Real-time testing at Alangilan Overpass, Batangas City

3.6. Summary of findings

This study successfully enhanced the YOLOv8m model through activation function adjustment, demonstrating that careful adjustment of non-linear transformations can notably strengthen object detection results and stability. By integrating a custom α-scaled GNSig activation, the proposed approach reached higher convergence stability, faster learning, and improved generalization compared to conventional functions such as SiLU, Mish, GELU, and LeakyReLU.

Experimental results revealed that the modified YOLOv8m model reached 96.1% results on the vehicle dataset and 75.9% on the pothole dataset—representing 5.4% and 4.8% improvements, respectively, over the baseline configuration. These gains were supported by lower validation losses, smoother precision–recall curves, and increased mean Average Precision across all thresholds. The GNSig activation provided consistent gradient flow and prevented saturation, enabling better representation of fine visual features such as vehicle edges and surface irregularities. Additionally, real-time tests at the Alangilan Overpass confirmed that these improvements translate effectively into real-world deployment, where lighting variations and motion blur often degrade detection results.

The findings affirm that activation-level refinements can yield practical benefits for intelligent transportation systems by enhancing model reliability under complex environmental conditions. Future work may extend this research by exploring dynamic or hybrid activation schemes that adapt during training, applying similar adjustments to other YOLO architectures, or integrating attention-based modules to further strengthen contextual understanding in high-speed detection tasks.

Relative to existing YOLO-enhancement literature, the performance gains documented in this work are significantly larger and more consistent across diverse evaluation conditions. While earlier studies report incremental improvements through complex feature modules or architectural redesign, this work demonstrates that a refined activation alone can achieve comparable or superior results. This highlights an underexplored but highly effective direction for model optimization within ITS applications.

4. CONCLUSION

This study presented a systematic enhancement of the YOLOv8m model through activation function adjustment, aiming to strengthen detection precision, convergence stability, and adaptability in real-world intelligent transportation applications. By integrating and evaluating various activation functions—SiLU, Mish, GELU, LeakyReLU, and the proposed α-scaled GNSig—the research established that fine-tuning activation dynamics notably influences learning efficiency and model reliability. Experimental results confirmed that the proposed GNSig activation consistently outperformed all tested functions, attaining 96.1% results on the vehicle dataset and 75.9% results on the pothole dataset, with corresponding mAP improvements of 5.8% and 4.8% over the baseline. These results gains were supported by lower validation losses, smoother convergence patterns, and improved precision–recall balance, demonstrating the activation's ability to preserve gradient flow and mitigate saturation. Additionally, real-world testing at the Alangilan Overpass validated the model's practical applicability, revealing enhanced localization of small and partially occluded vehicles, fewer false detections, and stable classification under varying illumination and motion conditions.

The results underscore that optimizing activation mechanisms at the architectural level can yield substantial improvements in both detection results and computational efficiency. The α-scaled GNSig activation introduced in this study enhances gradient consistency while maintaining real-time results, marking it as a viable solution for embedded computer vision systems deployed in intelligent transportation networks. Future research may explore adaptive or hybrid activation strategies capable of adjusting dynamically during training, as well as integrating attention or transformer-based modules to further strengthen contextual awareness. Extending this approach to other YOLO variants and traffic-related tasks—such as vehicle tracking, speed estimation, and anomaly detection—may lead to more robust, adaptive, and sustainable AI-driven transportation solutions.

ACKNOWLEDGMENTS

The authors would like to extend their genuine gratitude to the Batangas City Government's Information Technology and Systems Development (ITSD) Department for serving as the validation platform where the developed vehicle classification system was tested and implemented. Their support, collaboration, and technical assistance were instrumental to the successful real-world assessment of this study. This work forms part of the institutionally funded project of Batangas State University–The National Engineering University in 2023, entitled "VISION DRIVE: Vehicle Intelligent Speed Identification and Counting Observation Network Device Using Real-Time Imaging Video Evaluation." The authors express their appreciation to the University for its continuous commitment to research and innovation. The authors also acknowledge the use of ChatGPT as an assistance tool in drafting portions of the manuscript and performing grammar refinement. All scientific insights, analyses, and conclusions remain solely the responsibility of the authors.

FUNDING INFORMATION

This study was supported by the Local Research Evaluation Committee of Batangas State University, The National Engineering University–Alangilan Campus, Batangas City, Philippines. The authors thank the university for providing the support to use this study and the computational and laboratory facilities needed for model development, training, and testing.

AUTHOR CONTRIBUTIONS STATEMENT

This research was jointly developed and completed through the collaborative efforts of all authors. Renz Raniel V. Serrano served as the principal investigator and led the conceptualization, methodology design, and implementation of the modified YOLOv8m framework. He was responsible for the development of the activation function, software integration, test-based validation, formal analysis, visualization, and preparation of the original manuscript draft. He also supervised the entire research process, managed project administration, and coordinated funding acquisition.

Jen Aldwayne B. Delmo contributed to the methodological refinement and test-based implementation. He was actively involved in coding, validation, and data curation, ensuring the integrity and reproducibility of the model training and evaluation. He also assisted in reviewing, editing, and improving the technical quality of the manuscript, particularly in the analysis and interpretation of results.

Cristina Amor M. Rosales provided academic supervision, expert guidance on research design and validation, and contributed notably to the review and editing of the manuscript. She ensured the scientific rigor of the study, verified the results of analytical interpretations, and supported the coordination of project documentation and dissemination.

All authors reviewed and approved the final version of the manuscript. They collectively affirm responsibility for the integrity and results of the research presented in this paper.

Name of Author	C	M	So	Va	Fo	I	R	D	O	E	Vi	Su	P	Fu
Renz Raniel V. Serrano	✓	✓	✓	✓	✓	✓		✓	✓	✓				✓
Jen Aldwayne B. Delmo	✓	✓	✓			✓			✓	✓	✓	✓	✓	✓
Cristina Amor M. Rosales	✓			✓		✓		✓		✓		✓		✓

C	: **C**onceptualization	I	: **I**nvestigation	Vi	: **Vi**sualization
M	: **M**ethodology	R	: **R**esources	Su	: **Su**pervision
So	: **So**ftware	D	: **D**ata Curation	P	: **P**roject Administration
Va	: **Va**lidation	O	: Writing - **O**riginal Draft	Fu	: **Fu**nding Acquisition
Fo	: **Fo**rmal Analysis	E	: Writing - Review & **E**diting		

CONFLICT OF INTEREST STATEMENT

The authors declare that there are no conflicts of interest regarding the publication of this paper. All financial and institutional support was provided solely for academic and research purposes by Batangas State University – The National Engineering University (Alangilan Campus) and the Batangas City Government Information Technology and Systems Development (ITSD) Department. None of these institutions influenced the study's design, data collection, analysis, or interpretation. The authors affirm that the findings and conclusions presented are the result of independent scientific work done without any personal, financial, or professional bias.

DATA AVAILABILITY

The dataset used in this study was personally gathered and prepared by the authors for the purpose of evaluating the modified YOLOv8m model in vehicle classification and detection tasks. Data collection was done under strict ethical and legal considerations to ensure compliance with privacy regulations. All video sources were reached from publicly installed surveillance cameras in authorized locations, and personally identifiable information—such as license plates or human faces—was anonymized through automated blurring and masking techniques prior to processing.

The compiled dataset encompasses multiple traffic conditions, vehicle types, and lighting scenarios to ensure robustness and generalization. It includes annotated images and metadata formatted according to the YOLOv8 labeling convention. The dataset has been curated solely for research purposes and is not publicly distributed to protect the confidentiality of data sources.

However, the dataset and associated annotation files can be made available upon reasonable request from the corresponding author for academic, non-commercial research use. Interested researchers may contact the authors directly to obtain access details, annotation guidelines, and documentation describing the data preparation workflow.

REFERENCES

[1] Y. Shao, R. Zhang, C. Lv, Z. Luo, and M. Che, "TL-YOLO: Foreign-object detection on power transmission lines based on improved YOLOv8," *Electronics*, vol. 13, no. 8, p. 1543, 2024, doi: 10.3390/electronics13081543.

[2] H. Xie, X. Fan, K. Qi, D. Wu, and C. Ren, "Spatial–temporal fusion gated transformer network for traffic flow prediction," *Electronics*, vol. 13, no. 8, p. 1594, 2024, doi: 10.3390/electronics13081594.

[3] 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 (CVPR)*, 2016, pp. 779–788.

[4] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal speed and results of object detection," *arXiv preprint arXiv:2004.10934*, 2020.

[5] N. Jegham, C. Y. Koh, M. F. Abdelatti, and A. Hendawi, "Evaluating the evolution of YOLO models: A comprehensive benchmark study of YOLOv11 and its predecessors," *arXiv preprint arXiv:2411.00201*, 2024.

[6] G. Peserico and A. Morato, "Performance evaluation of YOLOv5 and YOLOv8 object detection algorithms on resource-constrained embedded hardware platforms for real-time applications," *in Proceedings of the IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA)*, pp. 1–7, 2024.

[7] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," *arXiv preprint arXiv:1710.05941*, 2017.

[8] L. Liu, H. Jiang, P. He, W. Chen, X. Liu, J. Gao, and J. Han, "A survey of activation functions in deep learning: Recent advances and future perspectives," *Pattern Recognition*, vol. 122, p. 108426, 2022.

[9] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *arXiv preprint arXiv:1505.00853*, 2015.

[10] D. Misra, "Mish: A self-regularized non-monotonic activation function," *arXiv preprint arXiv:1908.08681*, 2019, doi: 10.5244/C.34.191.

[11] D. Hendrycks and K. Gimpel, "Gaussian error linear units (GELUs)," *arXiv preprint arXiv:1606.08415*, 2016.

[12] Y. Liu, Z. Wang, and F. Zhang, "Comparative study of ReLU, Swish, and Mish in deep convolutional networks," *Electronics*, vol. 12, no. 15, p. 3251, 2023, doi: 10.3390/cancers15123251.

[13] T. Pham, S. Lee, and J. Kim, "Performance analysis of augmentation strategies for robust object detection using CNNs," *Applied Intelligence*, vol. 52, pp. 17829–17845, 2022.

[14] H. Gao, W. Xu, and Y. Zhang, "Efficient deep networks with adaptive activation adjustment for visual detection," *Applied Sciences*, vol. 13, no. 5, p. 2894, 2023.

[15] W. Li, H. Yu, and Q. Zhou, "Improved YOLOv8 for multi-class traffic object detection under low-visibility environments," *IEEE Access*, vol. 11, pp. 45672–45685, 2023.

[16] A. G. Al-Kaf, S. Al-Ahmadi, and A. Al-Haj, "Vehicle detection and classification in challenging conditions using YOLO-based frameworks," *Sensors*, vol. 22, no. 3, p. 1082, 2022.

[17] J. Urenda, O. Kosheleva, and V. Kreinovich, "Fuzzy techniques explain the effectiveness of ReLU activation function in deep learning," *ResearchGate Preprint*, 2024.

[18] J. A. B. Delmo, "Deep learning-based vehicle speed estimation in bidirectional traffic lanes," *Procedia Computer Science*, vol. 252, pp. 222–230, 2025, doi: 10.1016/j.procs.2024.12.024.

[19] S. K. Sharma and R. Singh, "Impact of data preprocessing and augmentation on results of deep learning models for object detection," *IEEE Access*, vol. 10, pp. 112345–112356, 2022.

[20] P. Wang, Y. Huang, and J. Wu, "Enhancing image-based object detection through mosaic and hybrid augmentation," *Applied Sciences*, vol. 12, no. 18, p. 9312, 2022.

[21] J. J. Nelson, M. V. He, and K. B. Bae, "A comparative analysis of image annotation tools for object detection and segmentation tasks," *IEEE Access*, vol. 11, pp. 97326–97340, 2023.

[22] S. E. Suh, J. T. Kim, and W. Park, "Label quality and its impact on deep learning-based object detection results," *Pattern Recognition Letters*, vol. 162, pp. 48–57, 2024.

[23] L. M. Z. Abubakar and R. G. Mendoza, "Visualization-based validation of annotated vehicle datasets for deep learning models," *Procedia Computer Science*, vol. 252, pp. 310–319, 2025.

[24] N. S. Khan, Y. Al-Jarrah, and P. R. Hasan, "Effects of annotation consistency on the results of convolutional object detectors," *IEEE Transactions on Artificial Intelligence*, vol. 5, no. 3, pp. 215–227, 2024.

[25] A. H. Khan, J. Ali, and M. Ahmed, "Evaluation of image normalization techniques for deep learning-based visual recognition systems," *Sensors*, vol. 23, no. 6, p. 2911, 2023.

[26] D. P. Nguyen, J. R. Torres, and L. C. Wang, "Impact of augmentation diversity on deep neural network generalization for object detection," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 35, no. 4, pp. 987–998, 2024.

[27] S. M. Gupta, L. Zhou, and F. Li, "Adaptive activation functions for improving convergence and generalization in deep convolutional networks," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 34, no. 8, pp. 1835–1847, 2023.

[28] T. Xu and Y. Wang, "Normalized and scaled sigmoid functions for stable gradient propagation in deep architectures," *Pattern Recognition Letters*, vol. 165, pp. 20–29, 2024.

[29] Z. Liu, H. Wang, and R. Chen, "Adaptive early stopping for efficient deep neural network training," *IEEE Access*, vol. 12, pp. 125893–125905, 2024.

[30] R. S. Girshick, I. J. Goodfellow, and D. Erhan, "Understanding results metrics for object detection systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 2, pp. 612–625, 2024.

[31] F. Hassan, Y. Zhao, and T. Zhang, "A comparative evaluation of precision, recall, and F1-score metrics in deep learning object detectors," *IEEE Access*, vol. 11, pp. 65873–65889, 2023.

BIOGRAPHIES OF AUTHORS

Renz Raniel V. Serrano is an Instructor at the College of Engineering, Batangas State University. He is Currently Taking up Ph D. in Engineering Education at Batangas State University-The National Engineering University, Batangas City, Philippines, He Holds a Degree of Master of Science in Computer Engineering from Batangas State University - The National Engineering University, Batangas City, Philippines and a Bachelor of Computer Engineering from the same university. He is the Planning Coordinator of College of Engineering from January 2025 up to present. He can be contacted at email: renzraniel.serrano@g.batstate-u.edu.ph.

Jen Aldwayne B. Delmo received the B.Sc. degree in Computer Engineering from AMA Computer College–Batangas, Philippines, the M.Sc. degree in Computer Engineering from Batangas State University, and the Ph.D. degree in Information Technology from AMA University, Philippines. She is currently the Head of the Research Office, Alangilan Campus, Batangas State University – The National Engineering University, Batangas City, Philippines. Before this, she served as Program Chair of the B.S. Computer Engineering program from 2018 to 2023, and as Program Chair of the Master of Science in Artificial Intelligence (MSAI) program from 2023 to 2024. To date, she has published approximately 11 papers (combining journal articles and conference proceedings) and has accrued 37 citations on Google Scholar, with an h-index of 4. Her research interests include computer vision, intelligent systems, deep learning applications, and smart city technologies. She can be contacted at email: jenaldwaynebanaag@gmail.com.

Cristina Amor M. Rosales is an Associate Professor and serves as Dean of the College of Engineering at Batangas State University. She holds a Doctor of Engineering degree in Civil Engineering from Sejong University in Seoul, South Korea, a Master of Science degree in Environmental Engineering from the University of the Philippines–Diliman, and a Bachelor of Science degree in Civil Engineering from University of the Philippines–Los Baños. She has led accreditation efforts for COPC, AACCUP, PTC, and ABET, and directed institutionally funded research projects including a deep-learning bridge-condition monitoring tool, the AMBIANCE air-quality sensor network, and the ISLAND infrastructure-soil interaction assessment. Her research expertise spans in machine learning applications, infrastructure assessment, sustainable development policy, and the development of pedagogical frameworks tailored to contemporary engineering education. She can be contacted at email: cristinaamor.rosales@g.batstate-u.edu.ph.