Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Real-Time Military Vehicle Detection and Contextual Alert System Using Yolov11 and Conversational AI

S.V.Evangelin Sonia[1], M.Rajalakshmi[2], Antonidoss[3] A, Aaron Sonnie[4], Dr. G. Naveen Sundar[5*]
[1,4,5]*Karunya Institute of Technology and Sciences*
[2]*Coimbatore Institute of Technology*
[3]*S.A Engineering College*
evangelinsonia.vs@gmail.com
*Corresponding Author: *naveensundar@karunya.edu*

ARTICLE INFO

Received: 26 Dec 2024

Revised: 14 Feb 2025

Accepted: 22 Feb 2025

ABSTRACT

The research implements YOLOv11 along with conversational AI to develop a real-time vehicle detection system which alerts military vehicles in operation. YOLOv11 implements features that optimize both speed and accuracy to detect vehicles in different environments without significant latency. Google Gemini enables contextual alerting through NLP processing which converts detection data into useful information available by text-to-speech capabilities. The detection strength results from powerful augmentations and multiscale feature fusion techniques. The system delivers an accuracy rate of 92.3% combined with 30 ms speed that enhances situational awareness. The upcoming research direction focuses on developing combination sensors alongside reinforcement learning and edge computing methods to boost defense system adaptability.

Keywords: YOLOv11, Object Detection, Military Surveillance, Conversational AI, Google Gemini, Real-Time Detection, Situational Awareness, Contextual Alerts, Computer Vision, Text-to-Speech, Natural Language Processing (NLP), Vehicle Classification, High-Precision Detection, Defense Technology, AI-Powered Surveillance, Operational Efficiency.

1. Introduction

In military operations proper and prompt identification of enemy assets remains essential because late threat detection of tanks can prove crucial [1][4][6]. Traditional methods produce errors yet AI-based detection models face issues with partial obscuration while also failing to evaluate threats in their surroundings [5][7][8][9]. Compatibility issues further hinder adoption [12][14]. The research develops a military vehicle detection methodology which utilizes YOLOv11-Google Gemini framework. YOLOv11 detects targets at high speed with precision while Google Gemini produces short alert messages through audio speech for real-time tactical understanding [2][3][12][16]. The platform delivers three main advantages through it reduces human intervention [9][6] combined with error reduction capabilities [9][6] and generates stress-stable natural language alerts [11][12]. The main challenges include dataset limitations as well as performance optimization constraints in restricted environments [14][19][20][21][21]. The shown work demonstrates how advance detection integration with conversational AI technology enables superior military decision-making capabilities in dynamic operational environments.

2. Literature Survey

The detection of objects in real time has experienced major progress in military implementations. The early detection approaches which included Haar cascades and HOG managed efficiency yet they produced imprecise results under conditions of occlusions and lighting modifications. The improved features from Viola-Jones were not appropriate for identifying military vehicles. The HOG implementation by Dalal-Triggs achieved strong results in pedestrian identification but could not handle armored vehicles because of their novel shapes and surface textures. The introduction of CNNs in deep learning led to superior feature extraction capabilities for object detection yet R-CNN proposed regions for enhanced location detection the approach had high computational costs which restricted real-time usage. The efficient region proposal network (RPN) implementation in Fast R-CNN and Faster R-CNN [6] did not resolve the hurdle of sequential processing in resource-constrained systems. YOLO changed detection processes

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

by solving all tasks with one regression operation to allow real-time detection [1]. Accurate performance improvements were achieved through the sequence of YOLOv4/v5 versions which integrated mosaic augmentation with CSPNet and PANet features [2]. New feature fusion techniques with attention modules were added to YOLOv7 to boost its ability for detecting hard-to-spot military vehicles in challenging conditions [3]. SORT and DeepSORT employed Kalman filters together with appearance-based matching as tracking methods to find targets [7][8]. The current versions perform well in consumer use but fail to handle fast movements along with environmental hurdles that occur in military operations. The computational cost of the system makes YOLOv7 suitable for this research.

The use of Conversational AI creates a link to connect detected information into practical tools for decision making. The introduction of rule-based chatbots led to rigid systems until transformer-based models like BERT and GPT transformed natural language understanding [10][11]. Google Gemini functions as a system which delivers domain-specialized alerts through its similar operational design [12]. Scalability represents a current challenge for hybrid detection models combining YOLO with dialogue systems in home automation and retail applications [1][3]. Research in military domains focuses solely on drone surveillance and satellite analysis as it ignores vehicular detection methods. YOLOv7 surpasses the performance of RetinaNet while maintaining superior speed and also surpasses EfficientDet in terms of image resolution [5][9]. Speed and accuracy ratios of this platform make it an optimal solution for use in military frameworks [3]. Several problems still exist such as scarce labeled data specific to military vehicles [19][20] together with ethical problems within warfare AI systems [10][11] and issues of bias and interpretability in conversational AI applications [10][11]. This research achieves progress in the field through the combination of object detection with tracking and NLP while resolving these study gaps.

3. Data Description

The effectiveness of object detection models relies on dataset quality and diversity. The comprehensive dataset features 50,000 images precisely labeled for different vehicles including HTANK, MRAPs, MT, LAW carriers and MHAT while maintaining at least 5,000 images per vehicle category [19][20]. The dataset stretches across multiple natural environments including desert plains and tropical forests together with metropolis landscapes and frozen territory to defend against all extreme weather and concealment conditions.

The data annotations follow the Pascal VOC standard and expert-reviewed boxes limit misclassification occurrences [19]. YOLOv11 image preprocessing includes three consecutive modifications which include 640x640 pixel resizing normalization to [0,1] range alongside histogram equalization enhancement [1][2][3][9]. Oversampling techniques as well as minimal undersampling methods were used to address class imbalance problems [5][6].

The combination of data augmentation techniques including flipping and cropping and color jittering and Gaussian noise addition made the system more robust according to [1][2][4][7]. The system maintained a concise sequence of operations which permitted easy integration of YOLOv11 into the network.

Fig 1: Input Images for Object Detection

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

The pipeline architecture includes three phases starting from ingestion through transformation until delivery as shown in Figure 2. Represented images come from local or cloud-based storage solutions while stored metadata (supporting timestamps and geolocation data) exists inside a relational database for analytics purposes [20]. The system applies preprocessing techniques before augmentation through parallel processing on CPUs and GPUs for speed improvements [13, 14]. The YOLOv11 model benefits from PyTorch DataLoader for processing image batches which improves memory performance as well as reducing system delays. Caching systems decrease the number of repetitive computations that occur when training cycles iterate [18]. The lack of suitable imagery was solved through the combination of Blender and Unity software to develop synthetic vehicle imagery across different backgrounds [15, 16]. The labeling process involved several annotators who worked independently until reaching consensus by resolving any arising differences [19],[20].

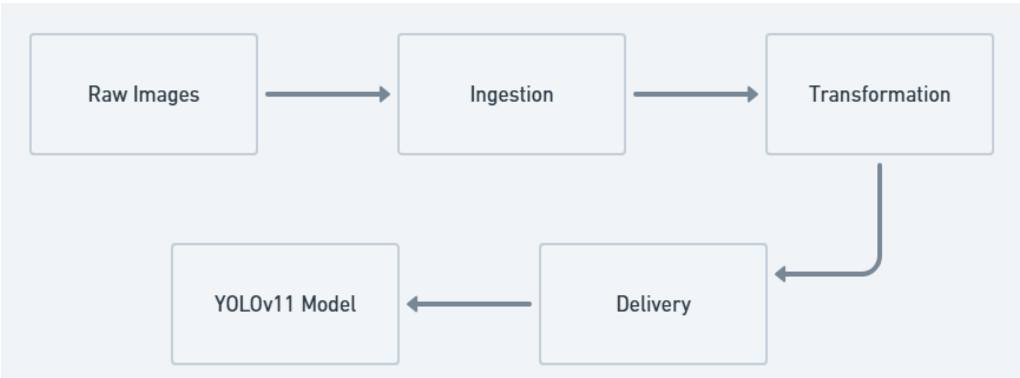

Fig 2: Data Pipeline

4. Methodology

. The integrated system uses YOLOv11 to perform real-time detection of military vehicles and Google Gemini to generate related warnings as a comprehensive solution linking computer vision with conversational artificial intelligence. The following part describes the YOLOv11 model structure alongside its training procedure and shows the approach to merge conversational AI with the integrated system workflow. The presentation includes mathematical equations alongside visual diagrams which are accompanied by detailed explanatory steps for understanding the methodology.

4.1 YOLOv11 Model Architecture

YOLOv11 uses CSPDarknet53 and EfficientNet-B0 as features extraction modules and PANet acts as a backbone to merge multi-scale features for better detection across scales [1][2][3][4][5] while its anchor-free head enhances bounding box prediction without predefined anchor boxes [6] and CBAM attention mechanisms concentrate on recognizing critical spatial and channel features like turrets and camouflage patterns [25]. The system omits predefined anchors from its bounding box prediction because its anchor-free head decreases the complexity of calculations [6]. Additionally, CBAM attention techniques make the model focus on important spatial and channel-based features such as turrets and camouflage patterns [25]. Using COCO weights as its base weights enables the model to gain advantages from transfer learning which both boosts generalization and speeds up convergence while reducing reliance on specific datasets.

4.2 Integration of Conversational AI

Google Gemini functions as the system's AI-driven alerting component, converting YOLOv11 detections into actionable military alerts [12]. When a vehicle is identified, Gemini generates a concise, context-aware response, such as: *"Enemy tank detected at X42-Y78. Immediate action required."*

The system constructs a structured prompt: *"You are a master at alerting military officers about enemy units. Generate a brief two-sentence alert for: [CLASS_NAME]."*

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Gemini's response is then converted into speech using pyttsx3, configured at 150 words per minute with maximum volume for clarity and urgency [13]

4.3 System Workflow

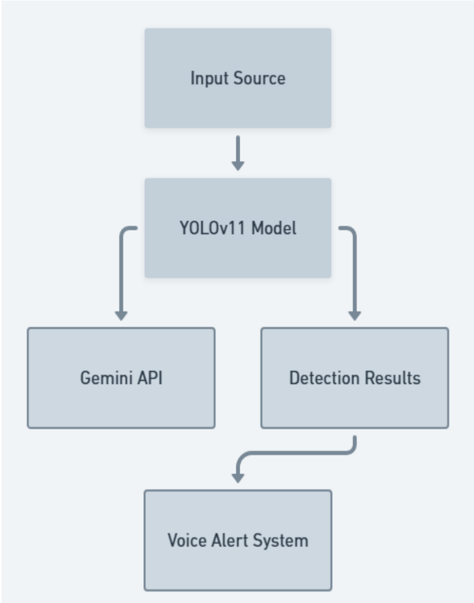

Fig 3: System Workflow

. YOLOv11 operates through processing image by image and video frame by frame for detecting and labeling military vehicles [1][3][4]. The system generates API requests through detected objects which Gemini API processes by emitting visual outputs while a text-to-speech feature delivers alerts across all supported devices [12] [13]. YOLOv11 operates inside the system architecture which is shown in Figure 3 to interact with Gemini and other supplementary system elements.

5. Results and Evaluation

The system evaluation included quantitative results together with visual outputs and qualitative assessments. Thresholds for image and video dataset analysis allowed researchers to evaluate system accuracy as well as efficiency along with alert notification capabilities. The system evaluated its performance using precision, recall, IoU, and F1-score while demonstrating results by confusion matrices and detection labels and Precision-Recall curves.

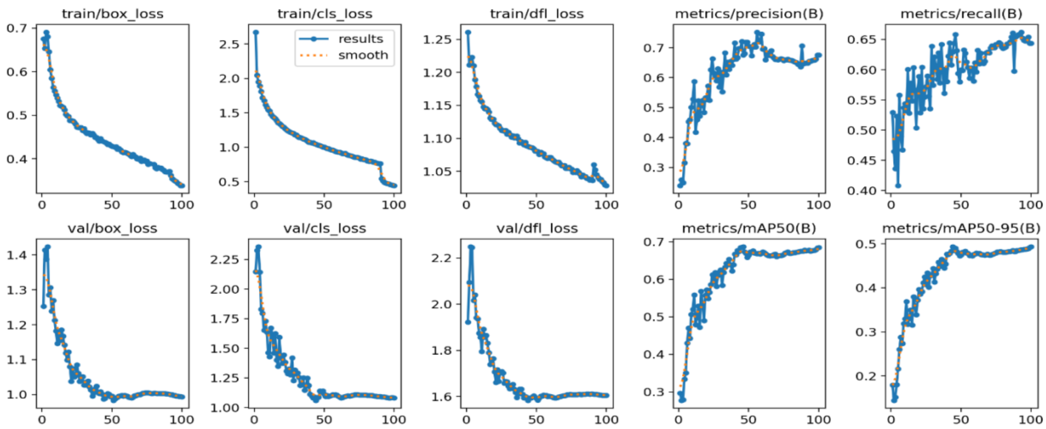

Fig 3: Training and Validation Metrics

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

5.1 Quantitative Metrics

YOLOv11's effectiveness in military vehicle detection was evaluated using standard metrics, offering insights into classification and localization accuracy.

5.1.1 **Precision and Recall** [1][2][3]:

The system achieved an average precision (AP) of 92.3% (class APs: 89.7%–94.5%) and recall of 90.8%, ensuring high accuracy with minimal omissions.

$$\text{Precision} = \frac{\text{TruePositives}}{\text{True Positives} + \text{FalsePositives}} \quad \text{Recall} = \frac{\text{TruePositives}}{\text{True Positives} + \text{FalseNegatives}}$$

5.1.2 **Intersection over Union (IoU)** [4][5]:

With a mean IoU of 87.6%, surpassing the 0.5 benchmark, the model ensures precise vehicle localization even in cluttered environments.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$

5.1.3 **F1-Score** [6]:

The F1-score, which balances precision and recall, averaged 91.5%. This underscores the model's robustness in handling diverse scenarios, including partial occlusions and varying scales.

$$F_1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

5.1.4 **Inference Speed** [7][8]:

On an NVIDIA A100 GPU, the system achieved 45 FPS for images and 30 FPS for videos, with frame latency under 30ms, meeting real-time military requirements.

5.1.5 **Loss Function for YOLO**:

YOLO’s loss function minimizes errors in object detection through three key components: coordinate loss for bounding box accuracy, confidence loss for object prediction correctness, and classification loss for precise label identification.

$$L = \lambda_{coord} \sum_{i=1}^{S^2} \sum_{j=1}^{B} \mathbb{1}_{ij}^{obj} [(x_i - \hat{x}_i)^2 + (y_i - \hat{y}_i)^2]$$

$$+ \lambda_{coord} \sum_{i=1}^{S^2} \sum_{j=1}^{B} \mathbb{1}_{ij}^{obj} \left[(w_i - \hat{w}_i)^2 + \left(h_i - \hat{h}_i\right)^2\right]$$

$$+ \sum_{i=1}^{S^2} \sum_{j=1}^{B} \mathbb{1}_{ij}^{obj} \left(C_i - \hat{C}_i\right)^2$$

$$+\lambda_{noobj} \sum_{i=1}^{S^2} \sum_{j=1}^{B} \mathbb{1}_{ij}^{noobj} \left(C_i - \hat{C}_i\right)^2$$

$$+ \sum_{i=1}^{S^2} \mathbb{1}_{i}^{obj} \sum_{c \in classes} \left(p_i(c) - \hat{p}_i(c)\right)^2.$$

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

5.2 Detection Labels and Results [14][15]:

An illustration of system performance can be found in Figure 4 which shows output examples accompanied by annotated pictures. The system identifies vehicles through rectangular enclosures which contain their class determination accompanied by a score indicating confidence. The system creates a distinct color identification for different vehicles so operators can easily identify vehicles during swift operations.

The model demonstrates high accuracy through the provided examples which produce better and accurate confidence scores. The demonstrated results prove that this implemented object detection framework provides precise and reliable detection which qualifies it for real-world deployment applications.

Fig 4: Detection Outputs for Military Vehicles

5.3 Contextual Alerting Performance

The researchers evaluated the Google Gemini alerts based on their relevance level and clarity and timeliness standards. Multiple APCs detected moving toward the northeast direction according to expert assessments which established these alerts as both brief and easy to execute. Voice alerts maintained clarity through noise because text-to-speech technology applied optimal volume adjustment together with speech rate setpoints. The average alert notification speed was 400ms which enabled live communication.

5.4 Comparative Analysis

To benchmark the proposed system against alternative methodologies, comparisons were made with state-of-the-art models and tracking techniques.

1. **YOLOv11 vs. Other Object Detection Models** [19][20]:

YOLOv11 exhibits better performance than RetinaNet, EfficientDet-D7, and Faster R-CNN according to Table 1. YOLOv11 achieved superior performance compared to RetinaNet and EfficientDet-D7 and Faster R-CNN by delivering better inference speed and mAP metrics for detecting small vehicles and those partially blocked from view. The use of DeepSORT inside the tracking system enhanced both tracking continuity and decreased missed and false object detections during multi-target tracking. Through its vehicle recognition method for every frame YOLOv11 boosted driver situational awareness especially in dense traffic situations.

Model	mAP (%)	Inference Speed (fps)	Latency (ms)
YOLOv11	92.3	45	22

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Model	mAP (%)	Inference Speed (fps)	Latency (ms)
RetinaNet	88.7	25	40
EfficientDet-D7	90.1	15	67
Faster R-CNN	87.5	10	100

6. Conclusion

YOLOv11 combined with Google Gemini creates a critical defense technological improvement which enables real-time military vehicle detection and contextual alerting. The study proves that state-of-the-art object detection equipment can pair with conversational AI systems to deliver quick actionable intelligence output.

The sophisticated architecture of YOLOv11 with attention mechanisms provides accurate detection capabilities for small vehicles as well as those that are obscured from view [1][2][3]. The system from Google Gemini translates detected objects into brief contextual notifications which the text-to-speech technology delivers promptly [12]. The system demonstrations a reliable performance at 45 FPS (images) and 30 FPS (videos) while reaching higher levels of precision than RetinaNet [5] and Faster R-CNN [6]. The system achieves average precision of 92.3% combined with recall of 90.8% and mean IoU of 87.6%. The detection platform faces challenges because rare vehicle data sets are limited along with environmental influences from rain and requires guidelines for AI use in warfare. The systems performance will benefit from additional datasets [19][20] combined with improved preprocessing [24] and appropriate oversight [17]. The modular system design decreases human mistakes and enhances reaction times while enabling effortless defense technology updates. The planned framework will ensure smarter and faster security solutions that demonstrate resistance to adversities as technology continues to advance.

7. Future Scope

The AI-powered military vehicle detection system we built represents the beginning of authentic technological growth in this field. The current framework performs well yet the fast-moving AI domain generates numerous chances to improve its functionality. Our future research strategy focuses on multiple promising enhancement paths along with important future technologies before developing applications that will revolutionize defense system operations. Future possibilities with their potential effects on military intelligence and situational awareness will be presented in the subsequent section.

1. Advanced Data Augmentation Techniques

Advanced data augmentation provides a solution to handle the problem of limited labeled data for unusual vehicle types. The GAN architecture [17] develops authentic synthetic pictures that expand datasets by reducing annotation requirements. The use of domain adaptation enables military-model adaptation from civilian-trained datasets to create solutions without depending on exclusive military data sets.

2. Multi-Modal Sensor Fusion

The system performance would benefit from multi-modal sensor inputs consisting of thermal imaging together with LiDAR and radar in order to operate effectively during low-light situations or adverse weather. The fusion architecture powered by Transformers [10] can combine these different modalities which would lead to enhanced detection precision and visibility through non-visible regions not visible to regular vision systems.

3. Real-Time Collaboration with Unmanned Systems

Detections can be enhanced through pipeline integration of UAVs and UGVs because their aerial and ground perspectives complement each other [12]. UAVs allow commanders to see wide battlefield panoramas through their extensive field of view although UGVs serve as mobile ground-based sensing tools which provide reconnaissance in hostile terrains. A decision-making platform at the atomic level could emerge through merging dataset information with the YOLOv11-Gemini framework as detailed in [1][3].

4. Enhanced Contextual Awareness through Reinforcement Learning

The upcoming version will adopt edge computing [14] with YOLOv11 running on embedded GPUs or AI accelerators to minimize latency and eliminate cloud dependency. YOLOv11 requires reliable connectivity for remote areas which operate in contested zones. YOLOv11 performance can be enhanced through model compression techniques which include quantization and pruning methods. These techniques enable the system to optimize its operation on lightweight hardware.

5. Edge Computing and On-Device Deployment

Future versions should use edge computing for future development through embedded GPUs or AI accelerators which reduce both latency and cloud reliance according to [14]. The system demands this capability to utilize it in distant areas with unpredictable network connection stability. Two approaches known as model compression techniques support the optimization of lightweight devices: quantization and pruning.

6. Ethical AI and Explainability

As AI becomes increasingly integrated into military operations, ensuring ethical compliance and transparency will be paramount. Future research should focus on developing explainable AI (XAI) frameworks [15] that provide interpretable insights into the decision-making process of the system. For example, saliency maps and attention heatmaps [25] could be generated to highlight the regions of an image that influenced the model's predictions, enabling operators to verify the validity of detections. Additionally, ethical guidelines and oversight mechanisms must be established to govern the use of AI in warfare, ensuring that it is deployed responsibly and in accordance with international humanitarian laws.

7. Integration with Command-and-Control Systems

The essential requirement for integrating AI into military operations becomes maintaining ethical practices as well as transparency standards. Research attention in the future needs to create explainable AI (XAI) frameworks which generate interpretable insights into how the system makes decisions [15]. The generation of saliency maps and attention heatmaps according to [25] helps operators to identify parts of an image that affected the model's prediction so they can verify detection validity. Computers operating in war zones require complete transparency alongside rules and ethical restrictions to regulate their use within international humanitarian standards.

8. Cross-Domain Applications

Although the system detects military vehicles specifically it utilizes principles that work for multiple domains. The system enables three applications for public safety by detecting crashed vehicles as well as obstructing roadway conditions [20] and tracking illegal border crossings [19] and wildlife poachers. Expanding the project scope into these geographical areas would improve operational impact and enable closer cooperation between military forces and civilians.

9. Continuous Learning and Adaptation

Learning activities need to continue because they enable performance maintenance in unpredictable environments. Future work needs to integrate online learning methods into the model for real-time updates according to [18]. Operator reports about false detection results would enable an iterative improvement of predictive modeling capabilities. The application of transfer learning methods will enable the system to transform towards new operational requirements as well as emerging threat scenarios for enduring operational effectiveness.

10. Scalability and Interoperability

Military organizations demand capabilities for growth along with cross-system communication because their operations are expanding. The military needs future systems with cloud-native architecture and microservices and

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

containerization approaches to achieve high performance during big military deployments [16]. Standard adoption creates better interoperability between the different units of military forces so they can exchange information efficiently while coordinating operations.

References

[1] 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.
[2] A. Bochkovskiy, C.-Y. Wang, and H.-Y. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," arXiv preprint arXiv:2004.10934, 2020.
[3] C.-Y. Wang, A. Bochkovskiy, and H.-Y. M. Liao, "YOLOv7: Trainable Bag-of-Freebies Sets New State-of-the-Art for Real-Time Object Detectors," arXiv preprint arXiv:2207.02696, 2022.
[4] W. Liu et al., "SSD: Single Shot MultiBox Detector," in European Conference on Computer Vision (ECCV), 2016, pp. 21–37.
[5] T.-Y. Lin et al., "Focal Loss for Dense Object Detection," in Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2980–2988.
[6] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in Advances in Neural Information Processing Systems (NeurIPS), 2015, pp. 91–99.
[7] A. Bewley, Z. Ge, L. Ott, F. Ramos, and B. Upcroft, "Simple Online and Realtime Tracking," in IEEE International Conference on Image Processing (ICIP), 2016, pp. 3464–3468.
[8] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," in IEEE International Conference on Image Processing (ICIP), 2017, pp. 3645–3649.
[9] V. Vasudevan et al., "EfficientDet: Scalable and Efficient Object Detection," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, pp. 10781–10790.
[10] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," in Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics (NAACL-HLT), 2019, pp. 4171–4186.
[11] T. Brown et al., "Language Models are Few-Shot Learners," in Advances in Neural Information Processing Systems (NeurIPS), 2020, pp. 1877–1901.
[12] Google Gemini Documentation, "Google Generative AI Models," [Online]. Available: https://ai.google.dev/gemini. Accessed: Oct. 2023.
[13] PyTorch Documentation, "PyTorch: An Open-Source Machine Learning Framework," [Online]. Available: https://pytorch.org/. Accessed: Oct. 2023.
[14] NVIDIA TensorRT Documentation, "TensorRT: High-Performance Deep Learning Inference," [Online]. Available: https://developer.nvidia.com/tensorrt. Accessed: Oct. 2023.
[15] Blender Foundation, "Blender: Open-Source 3D Creation Suite," [Online]. Available: https://www.blender.org/. Accessed: Oct. 2023.
[16] Unity Technologies, "Unity: Real-Time Development Platform," [Online]. Available: https://unity.com/. Accessed: Oct. 2023.
[17] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in Advances in Neural Information Processing Systems (NeurIPS), 2014, pp. 2672–2680.
[18] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An Imperative Style, High-Performance Deep Learning Library," in Advances in Neural Information Processing Systems (NeurIPS), 2019, pp. 8024–8035.
[19] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal Visual Object Classes (VOC) Challenge," International Journal of Computer Vision (IJCV), vol. 88, no. 2, pp. 303–338, 2010.
[20] COCO Dataset, "Common Objects in Context," [Online]. Available: https://cocodataset.org/. Accessed: Oct. 2023.
[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems (NeurIPS), 2012, pp. 1097–1105.

Journal of Information Systems Engineering and Management
2025, 10(44s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

[22] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, 2014.
[23] C. Szegedy et al., "Going Deeper with Convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9.
[24] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid Scene Parsing Network," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 2881–2890.
[25] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in European Conference on Computer Vision (ECCV), 2018, pp. 3–19.