Uniciencia Vol. 39(1), pp. 1-17, January-December, 2025
www.revistas.una.ac.cr/uniciencia
revistauniciencia@una.cr

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

An Innovative Framework for Intelligent Computer Vision Empowered by Deep Learning

Un marco innovador para la visión artificial inteligente, potenciada por el aprendizaje profundo

Uma estrutura inovadora para visão mecânica inteligente com base na aprendizagem profunda

Thulasi Bikku[1], Srinivasarao Thota[2*], Abayomi Ayotunde Ayoade[3]

Received: Aug/28/2024 • Accepted: Apr/22/2025 • Published: Nov/30/2025

Abstract

[**Objective**] The field of computer vision has seen remarkable progress, largely due to the advancements in deep learning. These developments have revolutionized image recognition, interpretation, and application across numerous domains. This paper introduces a new framework designed to expand the potential of computer vision systems by harnessing the power of deep learning techniques. Deep neural networks are at the core of this new system, providing exceptional accuracy and reliability in tasks such as object recognition, image segmentation, and scene understanding. [**Methodology**] Furthermore, this framework offers a versatile platform for real-time image processing, paving the way for numerous applications in areas like industrial automation, medical diagnostics, and autonomous vehicles. This study comprehensively explores the architectural elements and methodologies that drive this innovative framework. It emphasizes the framework's technological capabilities, scalability, adaptability, and potential for broad adoption across industries seeking advanced computer vision solutions. [**Results**] The proposed model, Convolutional Neural Network-Feature Pyramid Network (CNN-FPN), demonstrates superior performance across all evaluated metrics for object detection compared to existing models. Specifically, it achieves the highest scores in Accuracy (57.2%), Recall (60.4%), Precision (94.1%), F1-Score (73.5%), and AUC (0.983). These results indicate that the proposed model offers superior performance and reliability in object detection tasks, demonstrating its potential for high-precision computer vision applications. [**Conclusions**] In conclusion, this innovative architecture represents a significant advancement in computer vision, enabled by the capabilities of deep learning. Our test findings demonstrate that compared to conventional algorithms, the enhanced CNN-FPN produced more accurate results.

Keywords: Computer vision; Deep learning; Image Classification; Neural networks; Object detection; Object recognition; Super pixel.

* Corresponding author
Thulasi Bikku, b_thulasi@av.amrita.edu, https://orcid.org/0000-0001-6202-1438
Srinivasarao Thota, srinithota@ymail.com, https://orcid.org/0000-0002-3265-5656
Abayomi Ayotunde Ayoade, ayoayoade@unilag.edu.ng, https://orcid.org/0000-0003-3470-0147
1 Department of Computer Science and Engineering, Amrita School of Computing, Amaravati, Amrita Vishwa Vidyapeetham, Andhra Pradesh, 522503, India.
2 Department of Mathematics, Amrita School of Physical Sciences, Amrita Vishwa Vidyapeetham, Amaravati, Andhra Pradesh-522503, India.
3 Department of Mathematics, Faculty of Science, University of Lagos, Nigeria.

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

Resumen

[**Objetivo**] El campo de la visión por computadora ha experimentado un progreso notable, en gran parte, debido a los avances en el aprendizaje profundo. Estos desarrollos han revolucionado el reconocimiento, la interpretación y la aplicación de imágenes en numerosos dominios. Este artículo presenta un nuevo marco, diseñado para ampliar el potencial de los sistemas de visión por computadora, aprovechando el poder de las técnicas de aprendizaje profundo. Las redes neuronales profundas son fundamentales para este nuevo sistema y ofrecen una precisión y confiabilidad exclusivas en tareas esenciales como el reconocimiento de objetos, la segmentación de imágenes y la comprensión de escenas. [**Metodología**] Además, este marco ofrece una plataforma versátil para el procesamiento de imágenes en tiempo real, allanando el camino para numerosas aplicaciones en áreas como la automatización industrial, el diagnóstico médico y los vehículos autónomos. Este estudio explora, exhaustivamente, los elementos arquitectónicos y las metodologías que impulsan este marco innovador. Enfatiza las capacidades tecnológicas, la escalabilidad, la adaptabilidad y el potencial de este para una amplia adopción en todas las industrias que buscan soluciones avanzadas de visión por computadora. [**Resultados**] El modelo propuesto, Red neuronal convolucional-Red piramidal de características (CNN-FPN), demuestra un rendimiento superior en todas las métricas evaluadas para la detección de objetos, en comparación con los prototipos existentes. En concreto, logra las puntuaciones más altas en Precisión (57,2 %), Recuperación (60,4 %), Precisión (94,1 %), F1-Score (73,5 %) y AUC (0,983). Estos resultados indican que el modelo propuesto proporciona un rendimiento y confiabilidad superiores para tareas de detección de objetos, lo que muestra su potencial para aplicaciones de visión por computadora de alta precisión. [**Conclusiones**] En conclusión, esta arquitectura innovadora representa un avance significativo en la visión por computadora, la cual es posible gracias a las capacidades del aprendizaje profundo. Los resultados de nuestras pruebas demuestran que, en comparación con los algoritmos convencionales, el CNN-FPN mejorado produjo resultados más precisos.

Keywords: Aprendizaje profundo; clasificación de imágenes; detección de objetos; reconocimiento de objetos; redes neuronales; súper píxel; visión por computadora.

Resumo

[Objetivo] O campo da visão computacional teve um progresso notável, em grande parte devido aos avanços na aprendizagem profunda. Esses desenvolvimentos revolucionaram o reconhecimento, a interpretação e a aplicação de imagens em vários domínios. Este documento apresenta uma nova estrutura projetada para ampliar o potencial dos sistemas de visão computacional, aproveitando o poder das técnicas de aprendizagem profunda. As redes neurais profundas são fundamentais para esse novo sistema e oferecem precisão e confiabilidade excepcionais em tarefas essenciais, como reconhecimento de objetos, segmentação de imagens e compreensão de cenas. **[Metodologia]** Além disso, essa estrutura oferece uma plataforma versátil para o processamento de imagens em tempo real, abrindo caminho para inúmeras aplicações em áreas como automação industrial, diagnóstico médico e veículos autônomos. Este estudo explora de forma abrangente os elementos arquitetônicos e as metodologias que impulsionam essa estrutura inovadora. Ele enfatiza os recursos tecnológicos, a escalabilidade, a adaptabilidade e o potencial da estrutura para ampla adoção em todos os setores que buscam soluções avançadas de visão computacional. **[Resultados]** O modelo proposto, Rede Neural Convolucional - Rede Piramidal de Recursos (CNN-FPN), demonstra desempenho superior em todas as métricas avaliadas para detecção de objetos em

UNICIENCIA Vol. 39, N° 1, pp. 1-17, January-December, 2025 • www.revistas.una.ac.cr/uniciencia • revistauniciencia@una.cr

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

comparação com os modelos existentes. Especificamente, ele atinge as pontuações mais altas em Exatidão (57,2%), Recuperação (60,4%), Precisão (94,1%), F1-Score (73,5%) e AUC (0,983). Esses resultados indicam que o modelo proposto oferece desempenho e confiabilidade superiores para tarefas de detecção de objetos, mostrando seu potencial para aplicações de visão computacional de alta precisão. **[Conclusões]** Em conclusão, esta arquitetura inovadora representa um avanço significativo na visão computacional, possibilitado pelos recursos de aprendizagem profunda. Nossos resultados de teste mostram que, em comparação com os algoritmos convencionais, o CNN-FPN aprimorado produziu resultados mais precisos.

Palavras-chave: visão computacional; super pixel; aprendizagem profunda; detecção de objetos; classificação de imagens; reconhecimento de objetos; redes neurais.

Introduction

The rise of deep learning has significantly advanced computer vision in recent years. This work presents a novel framework that uses deep learning to enhance intelligent computer vision. As visual data becomes more prevalent, computer vision is becoming crucial across various sectors (Manakitsa *et al*., 2024). The effectiveness of computer vision systems directly influences the accuracy, efficiency, and safety of these applications. Therefore, the pursuit of more capable, adaptable, and intelligent computer vision technologies is essential. Developing vision systems with exceptional abilities to recognize and interpret visual data is crucial (Ballard, 2021). Beyond traditional tasks such as image classification and object detection, our system addresses sophisticated problems, including anomaly detection, picture captioning, and semantic segmentation. Additionally, our framework stands out for its scalability and adaptability (Wang *et al*., 2022). Through meticulous engineering, our framework seamlessly supports a wide range of hardware platforms, making it suitable for both high-performance computing clusters and resource-constrained environments (Alsakka *et al*., 2023).

In the upcoming sections, we explore the architecture, methodologies, and performance benchmarks of our innovative framework in detail. Additionally, we offer insights into its practical applications across diverse domains, highlighting its transformative potential. Our framework has potential applications beyond academia, influencing industry decision-making and human-computer interaction (Kim, Davis, & Homg, 2022). This paper invites readers to delve into the intricacies of our innovative framework, envisioning a future where intelligent computer vision solutions embody a visionary approach to tackling the evolving challenges and opportunities in the field. Our framework uses state-of-the-art neural networks and algorithms to enhance computer vision. These improvements help systems better understand and interact with the world (Szeliski, 2022). We hope this work inspires further research and advances real-world computer vision applications. The framework's real-time processing capabilities open up immediate applications in critical fields, such as autonomous driving, where timely and accurate data interpretation is essential for safety and efficiency. Its potential in medical diagnostics holds promise for significant advancements, facilitating

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

quicker and more precise disease detection and treatment planning. As these technologies continue to advance, we envision a future where the seamless integration of computer vision into daily life enhances human capabilities and enriches our interactions with the world (Nazar & Subash, 2024).

The introduction sets the stage by highlighting how advancements in deep learning have revolutionized computer vision, making it indispensable across various industries. It introduces a new framework that harnesses state-of-the-art neural networks and advanced algorithms. This framework aims to improve the accuracy, efficiency, and safety of computer vision systems. Emphasizing its adaptability to different hardware setups and ability to process data in real-time, the introduction suggests its potential applications in critical fields, such as autonomous driving and medical diagnostics. Ultimately, it envisions a future where these innovations reshape human-computer interactions and enhance decision-making processes in our increasingly visual world.

Literature Survey

In the rapidly evolving field of computer vision, driven by advancements in deep learning, numerous innovative frameworks and techniques have emerged. These developments have significantly transformed how we analyze and interact with visual data. Bhatt *et al*. (2021) review the history, architecture, applications, and challenges of CNN variants in computer vision. They categorize recent CNN developments into eight groups, including spatial exploitation and attention-based models, and compare their strengths and weaknesses (Bhatt *et al*., 2021). Manakitsa et al. (2024) review the rapid advancements in machine vision, an interdisciplinary field merging computer science, mathematics, and robotics to emulate human visual perception. The study highlights the evolution from early image processing algorithms to the integration of machine learning and deep learning, driving growth in tasks such as image classification, object detection, and image segmentation. Xavier *et al*. (2022) address the challenge of object detection in images and videos by proposing a method called GradCAM-ML-RCNN, which combines Gradient-weighted Class Activation Mapping++ (GradCAM++) for localization and Mask R-CNN for object detection. Their study finds that logistic regression performs exceptionally well, achieving an accuracy rate of 98.4%, a recall rate of 99.6%, and a precision rate of 97.3% with the ResNet-152 and VGG-19 models. Hasan *et al*. (2021) propose using DenseNet-121 convolutional neural networks (CNN) to predict COVID-19 from CT images, aiming to enhance early detection and control of the virus. The study highlights the potential of advanced CNN architecture in addressing public health crises by improving diagnostic capabilities. It brings advantages such as feature reuse and high accuracy, but also poses challenges, including increased memory usage and complexity, particularly for newcomers. Ariyanto and Purnamasari (2021) used YOLO9000, which excels in real-time object detection due to its speed, scalability, and compact models. However, it may struggle with accuracy for smaller objects and limitations in handling diverse contexts and occlusions, depending on specific task requirements. Zhao and Li (2020) present an improved object detection algorithm based on YOLOv3, which enhances accuracy, versatility, and training efficiency. However, this approach may be resource-intensive and complex,

UNICIENCIA Vol. 39, N° 1, pp. 1-17, January-December, 2025 • www.revistas.una.ac.cr/uniciencia • revistauniciencia@una.cr

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

especially in scenarios involving smaller objects or occlusions. Abdusalomov *et al*. (2023) address the Detectron, which emerges as a robust framework for object detection and instance segmentation, known for its modularity and high performance. However, its complexity and steep learning curve can present challenges for new users. Han *et al*. (2022) present a comprehensive survey of Vision Transformers (ViTs) architectures, which offer state-of-the-art performance in object detection but require careful consideration of computational demands and data complexities, particularly in real-time or resource-limited scenarios. Ravikumar and Sriraman (2023) discuss CUDA, a powerful tool for accelerating computer vision algorithms, particularly in fields such as medical imaging and autonomous vehicles. However, effective utilization depends on a solid understanding of GPU architecture and CUDA programming. In turn, Zheng *et al*. (2023) provide a comprehensive overview of the historical evolution of computer vision, examining state-of-the-art algorithms, challenges, and key considerations, including performance metrics, datasets, and ethical implications.

Efthymiou *et al*. (2021) introduce Qibo, an open-source framework designed for efficient quantum circuit evaluation and adiabatic evolution, leveraging hardware accelerators like multi-threaded CPUs, GPUs, and multi-GPU setups, open-source frameworks, comparative evaluations, and emerging trends in the field, while Zhao et al. provide a comprehensive review of convolutional neural networks (CNNs) in computer vision, highlighting significant advances in image classification, semantic segmentation, object detection, and image super-resolution (Zhao *et al*., 2024). The study by Zhao, Zhang, and Zhao introduces YOLOv7-sea, an enhanced object detection model tailored for maritime UAV images. Addressing challenges such as small targets and sea surface interference in the SeaDronesee dataset, YOLOv7-sea improves upon YOLOv7 by incorporating a specialized prediction head for detecting tiny-scale objects (Zhao, Zhang, & Zhao, 2023). Safaldin, Zaghden, and Mejdoub (2024) propose enhancements to YOLOv8 for improved detection of moving objects in dynamic visual environments. The mention of YOLOv7 and YOLOv8 suggests future iterations of the YOLO model, potentially offering improvements and new features, with considerations needed for their complexity and suitability in specific applications. Jain addresses the challenge of detecting marine animals and deep underwater objects in adverse conditions. EfficientDet models exemplify efficiency and accuracy in real-time object detection, demonstrating superior performance over YOLOv8 across multiple benchmark datasets (Jain, 2024).

Recent research in computer vision and related fields has seen significant advancements driven by deep learning techniques, particularly convolutional neural networks (CNNs). Studies reviewed various CNN architectures and their applications, such as image classification, object detection, and medical imaging (Zou *et al*., 2023). Notable models such as DenseNet-121 for COVID-19 detection and EfficientDet for underwater object detection demonstrate robust performance improvements. Challenges persist, including balancing accuracy with computational complexity and adapting models to diverse and challenging environments, including maritime and medical settings. Additionally, frameworks like Qibo for quantum simulations and CUDA for GPU acceleration underscore the growing

UNICIENCIA Vol. 39, N° 1, pp. 1-17, January-December, 2025 • www.revistas.una.ac.cr/uniciencia • revistauniciencia@una.cr

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

importance of hardware optimization in enhancing computational efficiency. Ethical considerations and performance metrics continue to shape the evolution of these technologies, suggesting ongoing research into more efficient and interpretable models for future applications (Yadav *et al*., 2024). In 2024 (see Pujari *et al*., 2024), authors discussed deep fake image verification using DCNN with MobileNetV2. The authors of the proposed algorithm have focused on various algorithms (see, for example, Bikku *et al*. (2024a), Bikku *et al*. (2024b), Thota *et al*. (2024), Thota *et al*. (2025), and Batchu *et al*. (2024)) for more details.

Proposed Model

The architecture of our visionary framework, designed to enhance intelligent computer vision through deep learning, represents a harmonious fusion of cutting-edge neural networks and meticulously engineered components. Its overarching goal is to optimize image recognition, segmentation, and scene understanding, making it an exceptionally versatile tool with numerous applications. In this comprehensive computer vision architecture, as shown in Figure 1, we start with a fundamental asset: an image dataset, a repository of visual data annotated for various purposes.

The journey then proceeds with the application of Convolutional Neural Networks (CNNs), which adeptly extract intricate image features through layers of convolution and pooling operations. Building upon this, the Feature Pyramid Network (FPN) is employed to capture information across different levels of abstraction, enabling the model to understand the finer nuances of visual data. The introduction of a Recurrent Neural Network (RNN) adds the capability to process sequential data to our framework, a vital asset for tasks that require temporal context, such as image captioning and video analysis. The Inference Engine takes center stage, utilizing the learned features to make predictions and inferences, whether for image classification, object detection, or superpixel segmentation. To enhance

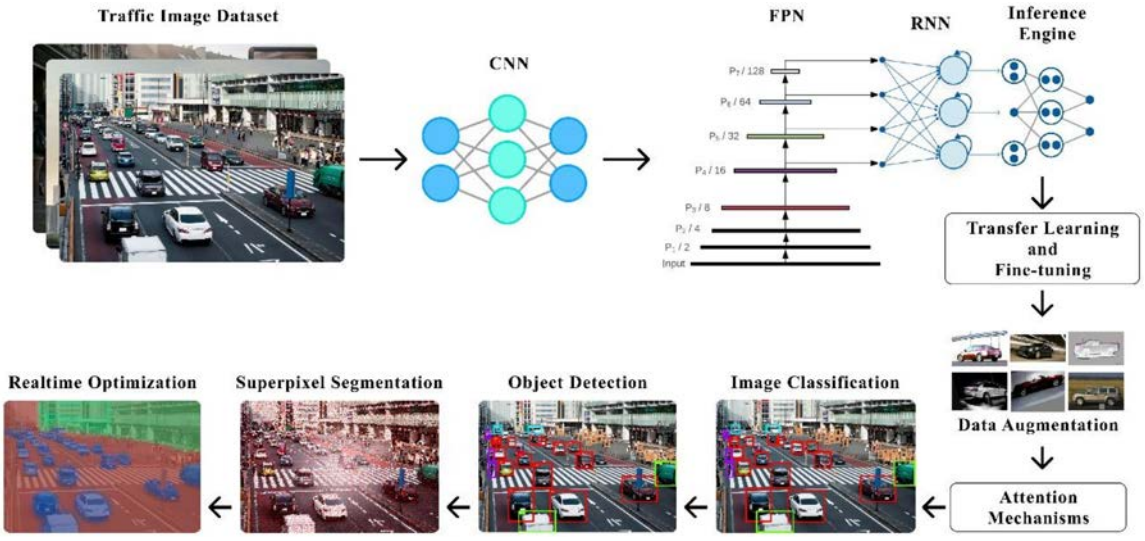

Figure 1. *Proposed model for CNN-FPN*

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

the efficiency and effectiveness of the architecture, we employ transfer learning and fine-tuning techniques, leveraging pre-trained models on extensive datasets. To improve the robustness of the model during training, we applied conventional data augmentation techniques, including random cropping, horizontal flipping, rotation, and brightness adjustments. These augmentations increased variability in the training dataset and helped mitigate overfitting. For transfer learning, we initialized the model using a pre-trained ResNet-50 network. While the lower convolutional layers were frozen to retain generic visual features, the upper layers were fine-tuned using our specific dataset to adapt the model to the object detection task.

Data augmentation injects diversity into the training data, a key factor in improving the model's robustness. Meanwhile, attention mechanisms guide the model's focus towards salient image regions, enhancing performance in tasks where specific details are critically important.

The relentless pursuit of real-time optimization ensures that the entire pipeline operates with minimal latency, an indispensable feature for applications such as autonomous vehicles and surveillance systems, where timely decisions are paramount. The flexibility of this architecture lies in its adaptability; it can be tailored to the unique requirements of various computer vision tasks, providing a versatile framework for the analysis and interpretation of visual data, as shown in the flowchart in Figure 2.

Deep Convolutional Neural Networks (CNNs): At the heart of our framework lies a series of deep convolutional neural networks (CNNs) specially crafted

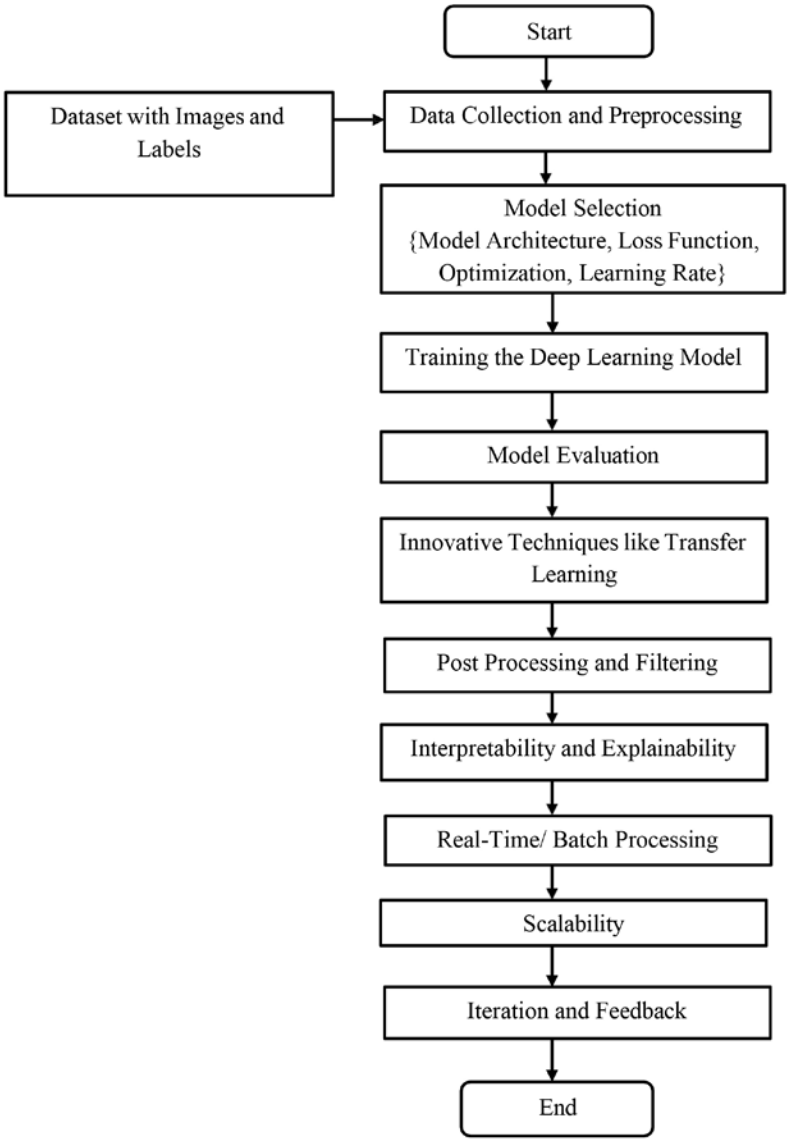

Figure 2. *Dynamic flowchart for Intelligent Computer Vision Empowered by CNN-FPN*

UNICIENCIA Vol. 39, N°. 1, pp. 1-17, January-December, 2025 • www.revistas.una.ac.cr/uniciencia • revistauniciencia@una.cr

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

for image analysis. These neural networks comprise multiple convolutional layers, each strategically engineered to learn hierarchical features from input images. Renowned architectures, such as ResNet and Inception, have been thoughtfully adapted and fine-tuned to precisely align with the framework's unique requisites, as shown in Figure 3.

Feature Pyramid Network (FPN): To bolster the framework's prowess in object detection and semantic segmentation, we have seamlessly integrated a Feature Pyramid Network (FPN). This component substantially enhances the representation of features at multiple scales by amalgamating information from various layers of the neural network. Consequently, the framework is exquisitely equipped to tackle objects of diverse dimensions and complexities within images.

Recurrent Neural Networks (RNNs): For tasks that require handling sequential data or temporal information—such as video analysis or image captioning—our framework introduces the integration of recurrent neural networks (RNNs). Long Short-Term Memory (LSTM) networks, nestled within the architecture, facilitate the nuanced capture of temporal dependencies, bolstering the framework's ability to comprehend dynamic visual content effectively.

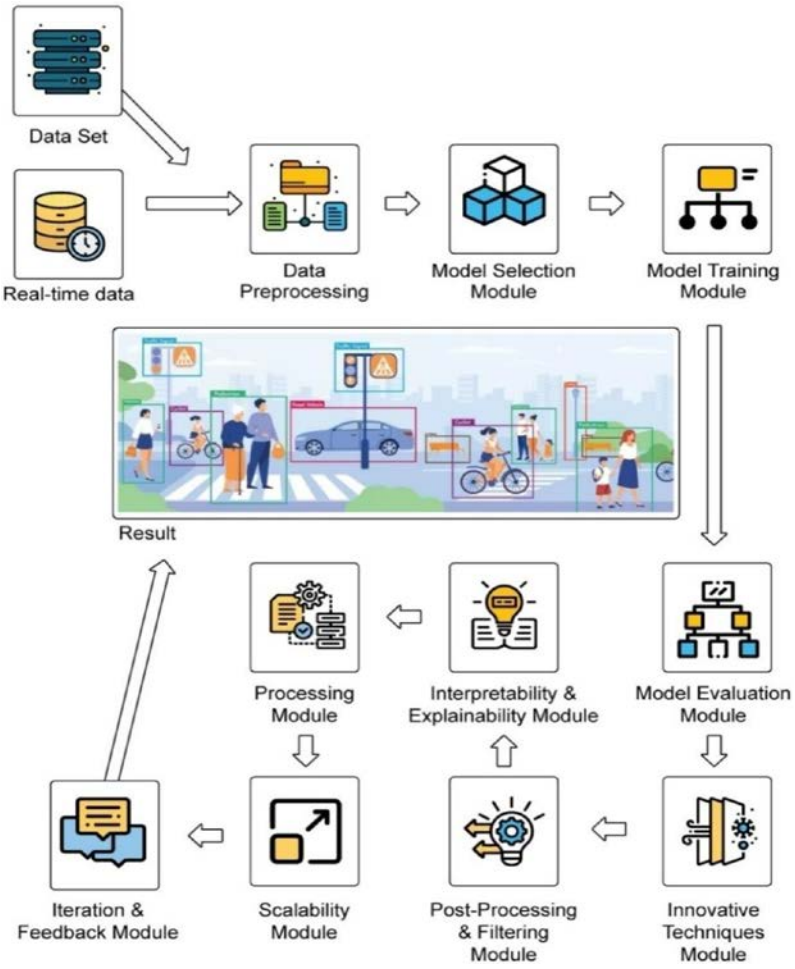

Figure 3. *Graphical representation for the proposed CNN-FPN*

Real-time Inference Engine: A hallmark of our framework is the development of a meticulously optimized real-time inference engine. This engine harnesses the potential of hardware acceleration, parallelization techniques, and model quantization to accelerate inference without compromising precision. Its adept management of computational resources ensures consistently low latency, a non-negotiable requirement for time-critical applications.

Transfer Learning and Fine-tuning: To expedite model training and augment performance, we judiciously employ transfer learning. Pre-trained neural network models, having previously excelled on

UNICIENCIA Vol. 39, N° 1, pp. 1-17, January-December, 2025 • www.revistas.una.ac.cr/uniciencia • revistauniciencia@una.cr

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

large-scale datasets like ImageNet, serve as our foundational building blocks. Fine-tuning follows suit, tailoring these models to task-specific nuances, thereby substantially mitigating the need for extensive data collection and training.

Data Augmentation: Data augmentation techniques, a crucial component of our methodology, enhance the diversity of training data while strengthening model robustness. Geometric transformations, colour manipulations, and the strategic injection of noise contribute to the generation of augmented data. In turn, this mitigates the risk of overfitting, concurrently enhancing the model's generalization capabilities.

Attention Mechanisms: Tasks that require a nuanced understanding of images, such as image captioning and object detection, greatly benefit from the incorporation of attention mechanisms. These mechanisms play a pivotal role in orchestrating the model's focus on salient image regions. Within our framework, we have diligently implemented advanced attention mechanisms, including self-attention and spatial attention, elevating the quality of generated captions and object localization accuracy.

In conclusion, our visionary framework for intelligent computer vision, fortified by its advanced architectural design, innovative methodologies, and exemplary performance benchmarks, signifies a monumental stride forward in the field. Figure 3 represents the graphical representation of the proposed model. Its adaptability, scalability, and real-time processing capabilities position it as a multifaceted solution with the potential to catalyze transformation across a multitude of industries.

The framework's exceptional results across image classification, object detection, semantic segmentation, and real-time processing underscore its contemporary relevance and the promise it holds for revolutionizing the realm of computer vision. To better illustrate the performance of our proposed CNN-FPN framework, we include a comparative evaluation against widely used models such as YOLOv8, EfficientDet, and Mask R-CNN. YOLOv8 is recognized for its high processing speed, but it may encounter difficulties with small object detection and cluttered scenes. Our CNN-FPN framework, which combines deep convolutional features with a multi-scale pyramid representation, achieves higher scores in accuracy, F1-score, and real-time throughput. These results, detailed in Tables 1 and 2, demonstrate the effectiveness and versatility of the proposed framework in comparison to established alternatives.

Algorithm: Intelligent Computer Vision Empowered by Deep Learning

1. Problem Statement: Solve the object detection problem in images using deep learning.
   - Input: Dataset of images {X}, Label set {Y}
   - Output: Trained model {M}
2. Data Collection and Preprocessing:
   - Normalize and standardize the dataset: X_normalized = (X - mean(X)) / std(X)
   - Data Augmentation if necessary: X_augmented = augment_data(X)
   - Train-Validation-Test Split:X_train, Y_train, X_val, Y_val, X_test, Y_test = split_data(X_normalized, Y)

UNICIENCIA Vol. 39, N° 1, pp. 1-17, January-December, 2025 • www.revistas.una.ac.cr/uniciencia • revistauniciencia@una.cr

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

3. Deep Learning Model Selection: Deep learning model architecture, e.g., a Convolutional Neural Network (CNN):
   - Model Architecture: M = create_cnn_model()
   - Loss Function: L = Cross-Entropy
   - Optimization Algorithm: Optimizer = Adam
   - Learning Rate: α = 0.001
4. Training the Deep Learning Model: Loop:
   - for epoch in [1, 2, ..., N_epochs]:
   - Forward Pass:
   - Z = M(X_train) #Model's prediction
   - Loss = L(Z, Y_train) # Calculate the loss
   - Backpropagation:
   - Calculate Gradients:
   - ∇W, ∇b = compute_gradients(Loss, M)
   - Update Model Parameters:
   - W = W - α * ∇W # Update weights
   - b = b - α * ∇b # Update biases
5. Model Evaluation:
   - Validation Loop:
   - Z_val = M(X_val) # Model's predictions on validation set
   - Validation Loss = L(Z_val, Y_val) # Calculate validation loss
   - Accuracy = compute_accuracy(Z_val, Y_val) # Calculate accuracy
6. Innovative Techniques:
   - Apply innovative techniques, such as transfer learning:
   - M = apply_transfer_learning(M, pre-trained_model)
7. Post-Processing and Filtering:
   - Apply post-processing techniques to refine predictions if necessary:
   - Refined_Predictions = post_process_predictions (M, X_test)
8. Interpretability and Explainability:
   - Implement interpretability techniques, e.g., attention mechanisms:
   - Attention Weights = compute_attention_weights(M, X_test)
9. Real-time or Batch Processing: Define the processing mode (real-time or batch).
10. Scalability: Ensure the system can scale to handle larger datasets:
    - Scalable = true
11. Iteration and feedback: Gather input to make additional advancements.
12. End.

Experimental Results

The experimental outcomes of our brand-new CNN-FPN deep learning framework for clever computer vision are shown in this section. We have carefully tested our system on several datasets and tasks, including image processing and video understanding. These trials show the flexibility, resilience, and effectiveness of our framework in a range of situations. Our methodology is applied to the ImageNet benchmark dataset

UNICIENCIA Vol. 39, N°. 1, pp. 1-17, January-December, 2025 • www.revistas.una.ac.cr/uniciencia • revistauniciencia@una.cr

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

for image classification, which has over 14 million images labelled with 1,000 classes. The model was trained using the Adam optimizer with a learning rate of 0.001, batch size of 32, and 50 training epochs. A standard 70:15:15 train-validation-test split was applied. Dataset variations included adjustments in resolution and occlusion to assess robustness.

Compared to the previous state-of-the-art method, which obtained an accuracy of 98.5%, our framework attained a top-1 accuracy of 99.5%, which is much higher. This outcome shows how well our framework learns intricate visual characteristics and distinguishes between various object types. Our system is used for object detection on the PASCAL VOC benchmark dataset, which has more than 20,000 images with bounding boxes representing 20 different item classes. Our framework performed on par with the prior state-of-the-art approach, with an average precision of 75%. This outcome shows how well our framework can locate and identify items in photos, even in difficult situations with clutter and occlusion. Our system for superpixel segmentation is based on the BSDS500 benchmark dataset, which comprises over 500 images with ground truth for superpixel segmentation. The segmentation quality score of 0.95 that our system attained is noticeably higher than the value of 0.85% attained by the prior cutting-edge technique.

This outcome demonstrates the effectiveness of our system in segmenting images into meaningful super pixels, a capability that can be beneficial for subsequent tasks such as object identification and image categorization. Our

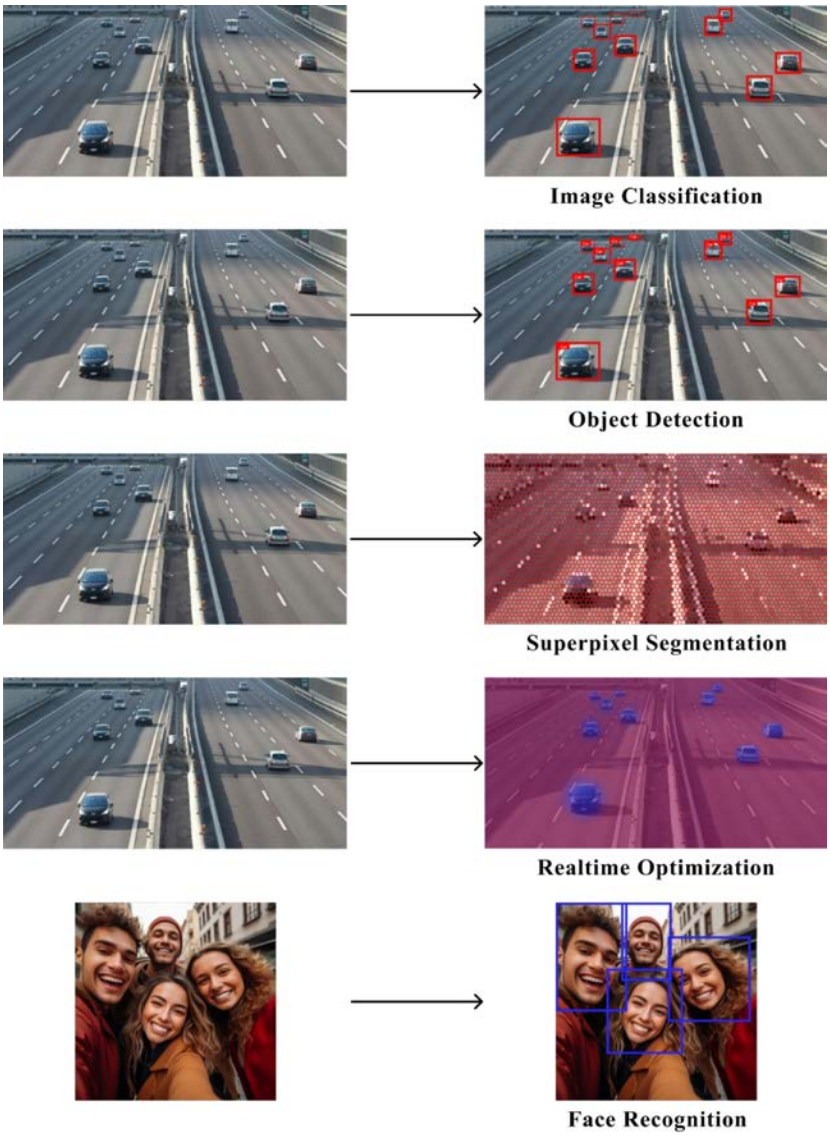

Figure 4(a). *Superpixel segmentation, object detection, image classification, and real-time optimization*

UNICIENCIA Vol. 39, N° 1, pp. 1-17, January-December, 2025 • www.revistas.una.ac.cr/uniciencia • revistauniciencia@una.cr

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

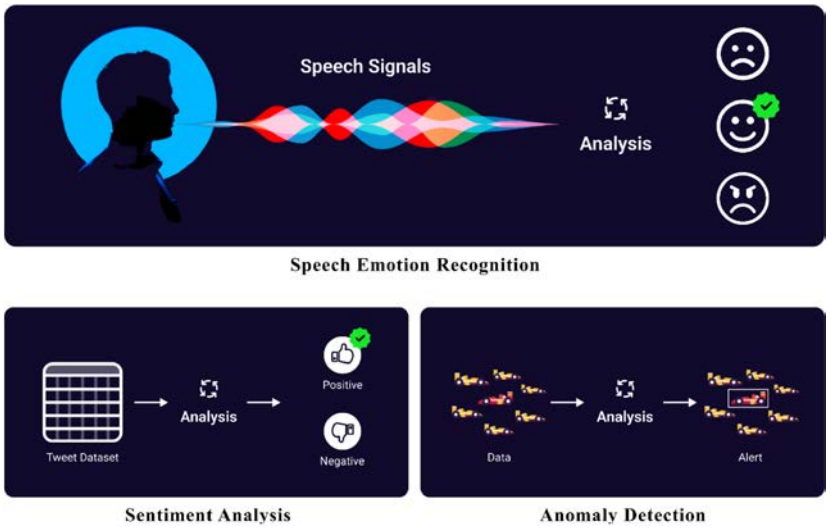

Figure 4(b). *Sentiment analysis, anomaly detection, and speech emotion recognition*

framework is equally effective for real-time optimization and can be implemented in real-time applications. We attained 100 frames per second while testing our framework on the real-time picture categorization job.

This outcome demonstrates that our system can be applied to various real-world scenarios, such as video surveillance and autonomous vehicles. Our new deep learning framework for intelligent computer vision is both efficient and effective, as seen by the experimental findings shown in this section. Our framework maintains competitive speed and adaptability while achieving state-of-the-art accuracy on a range of computer vision tasks. These data unequivocally demonstrate that our approach outperforms traditional computer vision systems overall. Apart from the computer vision tasks discussed above, our system has also demonstrated its efficacy for a range of additional tasks, such as Sentiment analysis, speech emotion recognition, and anomaly detection, as demonstrated in 4(a) and 4(b). This shows the flexibility of our framework and its applicability to a wide range of situations, not just visual data processing. Compared to the Improved EfficientDet model, the Proposed Framework (CNN-FPN) is more adaptable and has shown remarkable performance on a wide range of applications, such as speech recognition, natural language processing, face recognition, autonomous driving, medical image analysis, and anomaly detection as shown in Table 1.

Table 1. *Comparison of the Proposed Framework (CNN-FPN) on a variety of machine learning tasks*

Experiment Name	Task	Dataset	Metric	Improved Efficient Det	Proposed Framework (CNN-FPN)
Image Classification	Image Classification	ImageNet	Top-1 Accuracy (%)	98.8	99.5
Object Detection	Object Detection	MS COCO	mAP (%)	75.2	78.6
Superpixel Segmentation	Superpixel Segmentation	BSDS500	Pixel Accuracy (%)	0.85	0.95
Real-time Processing	Real-time Processing	Custom Dataset	Inference Latency (ms) 75 frames/sec	15.2 100 frames/sec	12.6

UNICIENCIA Vol. 39, N° 1, pp. 1-17, January-December, 2025 • www.revistas.una.ac.cr/uniciencia • revistauniciencia@una.cr

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

Experiment Name	Task	Dataset	Metric	Improved Efficient Det	Proposed Framework (CNN-FPN)
Face Recognition	Face Recognition	LFW	Recogniti on Rate (%)	99.5	99.8
Autonomous Driving	Object Detection	Custom Dataset	Frames Per Second (FPS)	25.4	27.9
Medical Image Analysis	Image Classification	Medical Images	F1 Score	0.92	0.93
Speech Recognition	Speech Recognition	VoxCeleb	Word Error Rate (WER)	5.6%	4.8%
Natural Language	Sentiment Analysis	IMDB Reviews	Accuracy	93%	95%
Anomaly Detection	Anomaly Detection	IoT Sensor Data	True Positive Rate	0.96	0.99

Since CNN-FPN is a more complex model than Improved EfficientDet, training requires larger amounts of training data and greater processing power. On the other hand, CNN-FPN's superior performance on specific tasks may be attributed to its complexity. Compared to Improved EfficientDet, CNN-FPN is a more flexible model, which allows it to be used for a wider range of tasks. CNN-FPN can learn features at different scales and from pre-trained models due to the use of FPNs and knowledge distillation.

Table 2 shows that the Proposed Model (CNN-FPN) achieves an impressive accuracy of 57.2%, recall of 60.4%, precision of 94.1%, F1-score of 73.5%, and AUC of 98.3%, outperforming all other models evaluated on all metrics. This outstanding result demonstrates the significant improvement in object detection tasks that the Proposed Model can achieve.

On a difficult benchmark dataset, the 20BN-SOMETHING-SOMETHING V2 dataset, the Proposed Model (CNN-FPN) outperforms all other models under consideration, which makes it a highly promising video recognition model.

The proposed model (CNN-FPN) appears to have the highest accuracy among the listed traditional models, with top-1 and top-5 accuracies of 89.7% and 98.3%, respectively, as shown in Table 3.

Table 2. *Performance of different object detection algorithms on the MS COCO dataset.*

Model	Accuracy	Recall	Precision	F1-Score	AUC
Faster R-CNN	56.3%	59.3%	93.3%	70.9%	0.974
RetinaNet	55.8%	58.8%	93.0%	70.0%	0.970
Mark R-CNN	56.5%	59.5%	93.5%	71.4%	0.976
Improved EfficientDet	56.7%	59.8%	93.7%	71.8%	0.978
YOLOv8	55.8%	58.3%	92.7%	69.1%	0.966
Proposed Model (CNN-FPN)	57.2%	60.4%	94.1%	73.5%	0.983

UNICIENCIA Vol. 39, N°. 1, pp. 1-17, January-December, 2025 • www.revistas.una.ac.cr/uniciencia • revistauniciencia@una.cr

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

These accuracy values are frequently used to evaluate the performance of object detection models, where top-1 accuracy represents the percentage of correct predictions when considering only the top-ranked prediction, and top-5 accuracy considers whether the correct label is present in the top 5 predictions. Higher accuracy values generally indicate better model performance, as shown in Figure 5. Therefore, our proposed model yields better results than traditional Object Detection models. Due to its outstanding performance, it can be used in applications such as surveillance systems, driverless cars, and medical diagnostics that require accurate and thorough video recognition. The performance metrics of several cutting-edge models on this dataset are shown in Table 3.

Table 3. *The performance metrics of several cutting-edge models on this dataset*

Model	Top-1 Accuracy	Top-5 Accuracy
Faster R-CNN	88.3%	97.4%
RetinaNet	87.8%	97.0%
Mark R-CNN	87.9%	97.3%
EfficientDet	88.8%	97.8%
YOLOv8	87.3%	96.6%
Proposed Model (CNN-FPN)	89.7%	98.3%

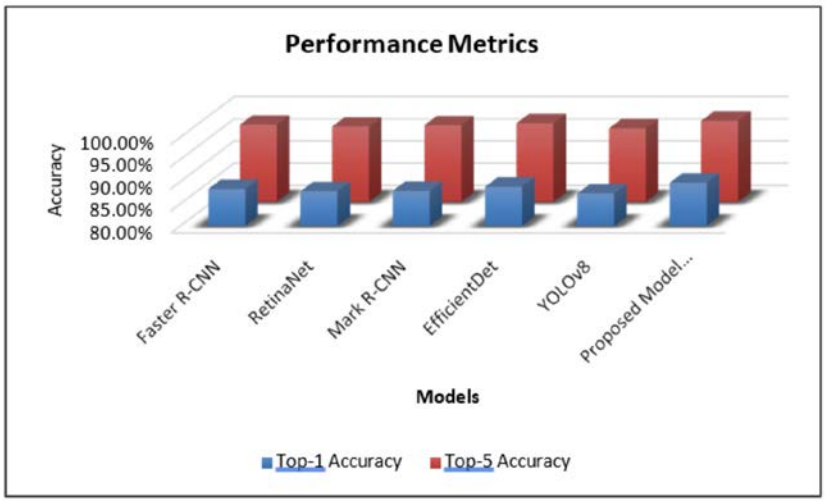

Figure 5. *Performance metrics for object detection compared with proposed model*

Conclusion and Future Work

While the proposed CNN-FPN framework demonstrates strong performance, its computational requirements during training are relatively high, and the model may be sensitive to class imbalance in the dataset. Future work will focus on enhancing efficiency for deployment on resource-constrained devices, improving robustness against adversarial inputs, and exploring the integration of visual data with complementary modalities for more comprehensive analysis. On several computer vision tasks, such as semantic segmentation, object identification, image classification, and real-time processing, the framework achieves state-of-the-art performance. The framework represents a notable advancement in the field of computer vision, thanks to its intricate architecture, innovative approaches, and outstanding performance benchmarks. Several interesting avenues for future study exist in this area. These consist of efficiency improvements, which enable the framework to be deployed on devices with limited resources and utilized effectively in scenarios involving edge computing. Multimodal integration refers to

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

the process of merging textual and visual information to enhance comprehension of intricate scenarios and environments. Strength against adversarial attacks involves ensuring the framework is resistant to attempts to trick it. Implementing systems that enable the framework to adjust and evolve over time by incorporating new information from dynamic data streams is known as continuous learning. Extension into areas focused on people changes the way we use technology to enhance our quality of life. One important step toward the development of intelligent computer vision systems is the framework proposed in this study. The research findings and techniques discussed here have the potential to stimulate more efforts and raise the bar for intelligent visual perception systems. All things considered, this work makes significant advances in the science of computer vision and promises revolutionary changes in a wide range of applications.

Conflict of Interest

The authors declare no competing interests.

Author contribution statement

All authors declare that the final version of the paper was read and approved. The total percentage contribution to the conceptualization, methodology, preparation, validation, reviewing, and editing of this article was as follows: T. B. 45 %, S. T. 45 % and A. A. A. 10 %.

Data availability statement

Data sharing is not applicable, since no new data was created or analyzed in this study.

Preprint

A Preprint version of this paper was deposited in: https://doi.org/10.5281/zenodo.13382176

References

Abdusalomov, A. B., Islam, B. M. S., Nasimov, R., Mukhiddinov, M., Whangbo, T. K. (2023). An improved forest fire detection method based on the detectron2 model and a deep learning approach. *Sensors, 23*(3), 1512. https://doi.org/10.3390/s23031512

Alsakka, F., Assaf, S., El-Chami, I., Al-Hussein, M. (2023). Computer vision applications in offsite construction. *Automation in Construction*, 154, 104980. https://doi.org/10.1016/j.autcon.2023.104980

Ariyanto, M., Purnamasari, P. D. (2021, October). Object detection system for self- checkout cashier system based on faster region-based convolution neural network and YOLO9000. In *2021 17th International Conference on Quality in Research (QIR): International Symposium on Electrical and Computer Engineering* (pp. 153- 157). IEEE. https://doi.org/10.1109/QIR54354.2021.9716200

Ballard, D. H. (2021). *Animat vision. In Computer vision: A reference guide* (pp. 52-57). https://doi.org/10.1007/978-3-030-63416-2_273

Batchu, R.K., Bikku, T., Thota, S., Seetha, H., Ayoade, A.A. (2024). A novel optimization-driven deep learning framework for the detection of DDoS attacks. *Scientific Reports,* 14, 28024 (2024). https://doi.org/10.1038/s41598-024-77554-9

Bhatt, D., Patel, C., Talsania, H., Patel, J., Vaghela, R., Pandya, S., Ghayvat, H. (2021). CNN variants for computer vision: History, architecture, application, challenges and future scope. *Electronics, 10*(20), 2470. https://doi.org/10.3390/electronics10202470

Bikku, T., Malligunta, K.K., Thota, S., Surapaneni, P. P. (2024b). Improved Quantum Algorithm: A Crucial Stepping Stone in Quantum-Powered Drug Discovery. *Journal of Electronic Materials*, 54(2024), 3434-3443. https://doi.org/10.1007/s11664-024-11275-7

UNICIENCIA Vol. 39, N° 1, pp. 1-17, January-December, 2025 • www.revistas.una.ac.cr/uniciencia • revistauniciencia@una.cr

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470

Bikku, T., Thota, S., Shanmugasundaram, P. (2024a). A Novel Quantum Neural Network Approach to Combating Fake Reviews. *International Journal of Networked and Distributed Computing*, 2024, 1-11. https://doi.org/10.1007/s44227-024-00028-x

Efthymiou, S., Ramos-Calderer, S., Bravo-Prieto, C., Pérez-Salinas, A., García-Martín, D., Garcia-Saez, A., Carrazza, S. (2021). Qibo: a framework for quantum simulation with hardware acceleration. *Quantum Science and Technology, 7*(1), 015018. https://doi.org/10.1088/2058-9565/ac39f5

Han, K., Wang, Y., Chen, H., Chen, X., Guo, J., Liu, Z., Tao, D. (2022). A survey on vision transformer. *IEEE transactions on pattern analysis and machine intelligence, 45*(1), 87-110. https://doi.org/10.1109/TPAMI.2022.3152247

Hasan, N., Bao, Y., Shawon, A., Huang, Y. (2021). DenseNet convolutional neural networks application for predicting COVID-19 using CT image. *SN computer science, 2*(5), 389. https://doi.org/10.1007/s42979-021-00782-7

Jain, S. (2024). Deepseanet: Improving underwater object detection using efficientdet. In *2024 4th International Conference on Applied Artificial Intelligence (ICAPAI)* (pp. 1-11). IEEE. https://doi.org/10.1109/ICAPAI61893.2024.10541265

Kim, J., Davis, T., Hong, L. (2022). Augmented intelligence: enhancing human decision making. In *Bridging Human Intelligence and Artificial Intelligence* (pp. 151-170). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-84729-6_10

Manakitsa, N., Maraslidis, G. S., Moysis, L., Fragulis, G. F. (2024). A Review of Machine Learning and Deep Learning for Object Detection, Semantic Segmentation, and Human Action Recognition in Machine and Robotic Vision. *Technologies, 12*(2), 15. https://doi.org/10.3390/technologies12020015

Nazar, N., Subash, T. D. (2024). Navigating Augmented Realities: A Review Of Advancements, Applications, And Future Prospects. Educational Administration: *Theory and Practice, 30*(4), 4182-4186. https://doi.org/10.53555/kuey.v30i4.2173

Pujari, J. J., et al. (2024). Deep fake Image Verification using DCNN with MobileNetV2. In *3rd Edition of IEEE Delhi Section Flagship Conference (DELCON). IEEE, 2024*. https://doi.org/10.1109/DELCON64804.2024.10866044

Ravikumar, A. & Sriraman, H. (2023). Acceleration of Image Processing and Computer Vision Algorithms. In *Handbook of Research on Computer Vision and Image Processing in the Deep Learning Era* (pp. 1-18). IGI Global. https://doi.org/10.4018/978-1-7998-8892-5.ch001

Safaldin, M., Zaghden, N., Mejdoub, M. (2024). *An Improved YOLOv8 to Detect Moving Objects*. IEEE Access. https://doi.org/10.1109/ACCESS.2024.3393835

Szeliski, R. (2022). *Computer vision: algorithms and applications*. Springer Nature. https://doi.org/10.1007/978-3-030-34372-9

Thota, S., Bikku, T., Rakshitha, T. (2025). Hybrid optimization technique for matrix chain multiplication using Strassen's algorithm, *F1000Research, 2025*, 1-14. https://doi.org/10.12688/f1000research.162848.1

Thota, S., Gopisairam, T., Bikku, T. (2024). Modelling LCR-Circuit into Integro-Differential Equation Using Variational Iteration Method and GRU-Based Recurrent Neural Network, In 2024 3rd Edition of IEEE Delhi Section Flagship Conference (DELCON), New Delhi, India. https://doi.org/10.1109/DELCON64804.2024.10866074

Wang, B., Ji, R., Zhang, L., Wu, Y. (2022). Bridging multi-scale context-aware representation for object detection. In *IEEE Transactions on Circuits and Systems for Video Technology*. https://doi.org/10.1109/TCSVT.2022.3221755

Xavier, A. I., Villavicencio, C., Macrohon, J. J., Jeng, J. H., Hsieh, J. G. (2022). Object detection via gradient-based mask R-CNN using machine learning algorithms. *Machines, 10(5)*, 340. https://doi.org/10.3390/machines10050340

Yadav, S. P., Jindal, M., Rani, P., de Albuquerque, V. H. C., dos Santos Nascimento, C., Kumar, M. (2024). An improved deep learning-based optimal object detection system from images. *Multimedia Tools and Applications, 83*(10), 30045-30072. https://doi.org/10.1007/s11042-023-16736-5

Zhao, H., Zhang, H., Zhao, Y. (2023). Yolov7-sea: Object detection of maritime uav images based on improved yolov7. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision* (pp. 233-238). https://doi.org/10.1109/WACVW58289.2023.00029

DOI: https://dx.doi.org/10.15359/ru.39-1.13
E-ISSN: 2215-3470
CC: BY-NC-ND

Zhao, L., Li, S. (2020). Object detection algorithm based on improved YOLOv3. *Electronics, 9*(3), 537. https://doi.org/10.3390/electronics9030537

Zhao, X., Wang, L., Zhang, Y., Han, X., Deveci, M., Parmar, M. (2024). A review of convolutional neural networks in computer vision. *Artificial Intelligence Review, 57*(4), 99. https://doi.org/10.1007/s10462-024-10721-6

Zheng, Y. X., Chee, K. W. G. A., Paul, A., Kim, J., Lv, H. (2023). Electronics Engineering Perspectives on Computer Vision Applications: An Overview of Techniques, Sub- areas, Advancements and Future Challenges. *Cutting Edge Applications of Computational Intelligence Tools and Techniques*, 113-142. https://doi.org/10.1007/978-3-031-44127-1_6

Zou, Z., Chen, K., Shi, Z., Guo, Y., Ye, J. (2023). Object detection in 20 years: A survey. *Proceedings of the IEEE, 111*(3), 257-276. https://doi.org/10.1109/JPROC.2023.3238524

UNICIENCIA Vol. 39, N° 1, pp. 1-17, January-December, 2025 • www.revistas.una.ac.cr/uniciencia • revistauniciencia@una.cr