Indian Journal of Computer Science and Technology
Special Issue, Volume 5, Issue 1 (January-April 2026), PP: 178-185.
https://www.doi.org/10.59256/indjcst.20260501C030
International Conference on Cyber and AI Security (ICCAIS-2026)
www.indjcst.com
ISSN No: 2583-5300

An In-Depth Exploration of Object Detection: Techniques, Applications, and Advancements

Saptaparni Chatterjee[1], ShazidWahid Khandakhani[2], Sachikanta Dash[3], Rabinarayan Panda[4], A Murali Krishna[5]

[1] *Department of CSE, Garden City University, Bengaluru, India.*
[2] *Department of Computer Science and Engineering, GIET University, Gunupur, India.*
[3] *Madanapalle Institute of Teccchnology and Science, Madanapalle, Andhra Pradesh, India.*
[4] *Department of Engineering, Garden City University, Bengaluru, India.*
[5] *Department of Electronics and communication engineering, KL University, Andhra Pradesh, India.*

To Cite this Article: *Saptaparni Chatterjee[1], ShazidWahid Khandakhani[2], Sachikanta Dash[3], Rabinarayan Panda[4], A Murali Krishna[5] , "An In-Depth Exploration of Object Detection: Techniques, Applications, and Advancements", Indian Journal of Computer Science and Technology Volume 05, Issue 01 (January-April 2026), PP: 178-185.*

Abstract: *Object detection, which serves as a fundamental element of computer vision, enables users to detect and position objects in pictures. This capability plays an essential role in various applications that include autonomous vehicles, security systems, healthcare solutions, and retail analytics. The article investigates all existing object detection techniques, including their recent advancements that use Convolutional Neural Networks (CNNs) for basic methods and various advanced methods that include few-shot learning, zero-shot learning, synthetic data creation, and domain adaptation. Initially, object detection relied on large annotated datasets and sophisticated model architectures to achieve accurate results. However, recent advancements have introduced methods that reduce data dependency, such as synthetic image generation and text-to-image synthesis, which allow for customizable datasets tailored to unique use cases. This development has been paralleled by the rise of domain adaptation techniques, which enable models to generalize better across diverse conditions and environments. The article investigates how ensemble techniques can improve detection accuracy and system resilience while exploring how Generative Adversarial Networks (GANs) create authentic synthetic data. The article presents few-shot and zero-shot learning methods which enable identification of new classes using only a few labeled examples, which proves valuable in settings that constantly introduce fresh object categories. This article aims to provide an in-depth overview of these cutting-edge techniques, discussing their respective strengths and applications, as well as the limitations and ethical challenges posed by object detection in real-world deployments. The research establishes a complete understanding of current object detection technology during its present phase while including information about future advancements in the field.*

Key Words: *Object detection, advancement, Generative Adversarial Networks, Convolutional Neural Networks.*

I. INTRODUCTION

Object detection has become a foundational element of modern artificial intelligence (AI) applications, enabling machines to interpret and interact with the visual world. The ability to locate and classify objects within an image underpins a broad array of technologies, from smart surveillance and autonomous vehicles to augmented reality and medical diagnostics.

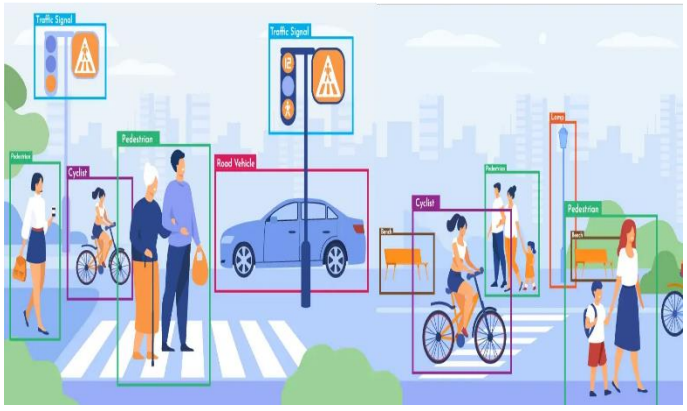

Figure 1 Object detection applications

The traditional method for object detection required engineers to create specific visual detection features through their work. The classical approach to object detection involved splitting the task into two stages: region proposal and classification. The two-stage architectural system developed by CNN-based models R-CNN Fast R-CNN and Faster R-CNN achieved enhanced performance through their use of neural networks which operated in both pathways of the system. The methods show strong results but they require extensive labeled data for their operation and they struggle to handle different environmental conditions which include changing light levels and object obstructionas described in figure 1. Researchers now investigate methods that originate from outside traditional supervised learning methods to solve existing research limitations. The methods of few-shot and zero-shot learning allow models to learn new object classes because these techniques need only minimal or no labeled data for training purposes. Few-shot learning enables the model to learn from limited examples through its use of meta-learning and feature reuse methods which analyze the distinct characteristics of various classes. The zero-shot learning approach employs semantic attributes together with class relationships to enable models to forecast object characteristics for classes that they have not yet studied.

Text-to-image models together with GANs (Generative Adversarial Networks) introduced a major advancement for creating synthetic data. The researchers use synthetic data to develop autonomous driving simulations which test their systems through different weather conditions and traffic patterns. The combination of domain randomization and adaptation methods enables models which learn from synthetic data and restricted real-world data to operate in unknown territories [5]. The method proves useful in fields which require correct results for critical situations, such as medical imaging and autonomous navigation, because these fields depend on accurate results. The advancement of object detection systems brings new ethical and operational difficulties which need to be addressed before these systems can be used. The fields that need to address data privacy and transparency issues together with potential biases include surveillance systems and healthcare environments which involve direct human contact. Object detection systems need to maintain their strength and flexibility while achieving fairness and transparency together with effective protection measures against potential abuses.

The article examines the fundamental technologies behind advanced object detection systems and their practical uses and existing restrictions while showing current developments and upcoming trends in the field. The research paper presents a comprehensive overview of object detection research by showcasing traditional CNN architectures together with recent few-shot learning techniques and synthetic data generation methods.

II. OBJECT DETECTION

Object detection needs two tasks that require finding objects in images while creating bounding boxes to show their locations. The field has experienced significant progress because deep learning methods, especially Convolutional Neural Networks, were introduced. The object detection process delivers essential functions to autonomous driving systems and surveillance operations and medical imaging applications because these fields need fast and accurate object recognition.

III. OBJECT DETECTION WITH CNN'S

1. Object detection using Convolutional Neural Networks (CNNs) The research in reference [6,7] provides a strong method for computer vision which enables the simultaneous detection of objects in images and their respective positions. The process of object detection differs from image classification because it needs to identify multiple objects in an image while also determining their exact positions through bounding boxes.

Key CNN-based Models and Architectures in Object Detection

CNNs establish their strength in object detection because they can understand the spatial structure of images to identify patterns which reveal specific characteristics of objects. The architecture of CNN-based object detection systems enables them to automatically learn all intricate image patterns through their convolutional and pooling network layers. The field has been shaped by several important CNN-based models which have become fundamental to its development.

A. R-CNN Family (Region-based CNNs):

The R-CNN model lineage, initiated with R-CNN (Regions with CNN characteristics), established the foundation for CNN-based object identification. The R-CNN system starts by analyzing images to find probable object regions through its selective search method. The system processes each suggested area through a CNN, which extracts features before conducting classification and bounding box refinement according to [8]..

B. YOLO (You Only Look Once):

YOLO revolutionised object detection by conceptualising it as a singular regression issue instead of a region proposal assignment. In contrast to R-CNN models, which operate in various phases, YOLO analyses the whole picture simultaneously and partitions it into grids. Every grid cell forecasts bounding boxes and the likelihood of each class for objects identified inside that grid. The YOLO system uses a single-pass detection method which enables it to function as one of the fastest object identification systems that works in real-time for applications that include autonomous driving and security systems and live video analysis. The detection system of YOLOv3 and YOLOv4 detects objects with greater accuracy while operating at faster speeds, which has led to their popularity among industrial users.

C. Single Shot MultiBox Detector (SSD):

SSD is another efficient model that performs detection in a single shot, similar to YOLO. SSD's use of feature maps at various scales makes it more robust in detecting small objects and objects with varying aspect ratios, making it useful for applications where object size variation is common.

2. Few-Shot Object Detection

Few-shot object detection aims to detect novel classes with minimal labeled examples, addressing the data dependency challenge inherent in CNN-based object detectors. Few-shot learning methods enable models to generalize to new classes by learning to distinguish key features from a few examples rather than extensive datasets. This capability is particularly valuable in fields where data collection is difficult or costly, such as medical imaging or wildlife monitoring.[11]

Key techniques in few-shot object detection include:

- **Meta-Learning**: Meta-learning functions as an educational process which enables a model to learn new tasks by using only a small set of training data. The meta-learning systems develop universal feature representations which enable them to identify new categories with only limited training data.
- **Prototypical Networks**: These networks use a metric-based approach, where the model learns prototype vectors representing each class.
- **Attention Mechanisms**: Attentiondriven mecha nisms help the learner models focus on relevant features when learning new classes, enabling faster generalization from very few examples.

Few-shot object detection is essential in fields with high data collection costs and is especially useful in dynamic environments where new classes are frequently introduced.

3.Text-to-Image Generators

Text-to-image creation is a sophisticated domain of artificial intelligence and machine learning that focusses on producing visual information from written descriptions. These models seek to comprehend and interpret word cues, converting them into visuals that precisely represent the given material. More recently, diffusion models and transformer-based architectures, text-to-image generators have attained exceptional realism and detail, transforming domains such as design, advertising, content creation, and art.

1. Generative Adversarial Networks (GANs):

Generative Adversarial Networks (GANs) were among the first effective frameworks for text-to-image synthesis. In Generative Adversarial Networks (GANs) two neural networks function through simultaneous training which uses competitive methods to develop their system. The generator produces pictures but the discriminator assesses their authenticity. The generator intends to create images which the discriminator cannot differentiate from authentic ones. The generator develops its skills through time which results in more realistic picture creation.

Text-to-image GAN models such as StackGAN and AttnGAN extend GANs through their text embedding system which generates visual content by using text-based vector information. StackGAN creates pictures in two phases through an initial coarse image which generates a first image based on the text input followed by a refinement step to enhance intricate features. AttnGAN uses attention methods to track specific words from a prompt which enables the system to associate different object components and colors with the prompt description. Despite their accomplishments GANs experience difficulties when they attempt to create visually detailed images which maintain internal coherence for complex tasks. The training process for GANs becomes difficult because of instability problems which result in one of the two systems taking control and producing low-quality results..

2. Diffusion Models:

Lately established diffusion models now present themselves as a strong alternative for creating images. The models achieve their results by executing multiple steps which decrease random noise until they produce an image that matches the given textual input. Theoretical understanding of GANs shows that diffusion models create authentic images which contain complex details through their stable operational methods.

DALL-E 2 and Stable Diffusion demonstrate diffusion-based text-to-image transformation systems through their operational methods. The algorithms train their systems by using extensive image-text databases which enable them to learn how words relate to visual elements. The model starts its development process with complete noise and then builds the image through multiple stages which follow the text prompt. The diffusion models create the image which matches the description because they use text prompts to guide each denoising process. Diffusion models excel at creating solutions which require deep understanding for explanatory concepts like "a futuristic cityscape at dusk with neon lights." The iterative process enables the model to focus on visual element integration through successive steps which improve its ability to depict intricate scenes and fine details [14,15].

3.Transformer Models and CLIP Embeddings:

Research in text-to-image systems has benefited from transformer models which use CLIP for their development. OpenAI's CLIP coordinates visual and textual representations in a common embedding space using contrastive learning. The study enables models to enhance their capacity to detect minute signals which they can link to the correct visual elements Michael claims that DALL-E and Imagen by Google use transformers to convert textual content into embedded formats. Following the description of the prompt's content and style, these embeddings are used to produce visuals that correspond to those specifications. Users may more easily give elaborate, multi-part descriptions and anticipate correct answers since CLIP models are especially effective at deciphering complicated language patterns.

4.Variational Autoencoders (VAEs):

VAEs, while less commonly used on their own in modern text-to-image generators, are sometimes combined with other architectures to create efficient latent spaces. They work by encoding the input text into a compressed latent space from which the

image is generated. Some diffusion models incorporate VAE-like structures to optimize the quality of generated images without consuming excessive computational resources.

IV. SPECULAR OBJECT DETECTION

Specular objects present unique challenges for object detection models due to their highly reflective surfaces. Common examples include mirrors, glass, polished metal, and even some types of liquids. In these cases, the objects often reflect the environment around them, creating complex visual effects that can confuse traditional object detection models[17-20].

Advanced Techniques for Specular Object Detection:

1. **Polarization Techniques in Imaging**: The technique enables multiple image capture through different polarization states using polarized light and filters, which helps to eliminate or decrease reflective elements in photographs. The method finds its primary application in industrial settings that require precise detection of reflective objects.
2. **Deep Learning on Multi-Spectral and Polarization Data**: Training neural networks with polarization and multi-spectral data adds robustness to specular object detection. Specialized networks that combine RGB, polarization, and other spectral channels allow for clearer object boundaries despite reflective interference.
3. **Multi-View and Depth Sensing**: Models achieve object recognition through depth sensors which include LiDAR technology and through multiple viewpoint image capture methods. Depth-based detection enables autonomous vehicles to identify real objects while distinguishing them from their mirrored surroundings.
4. **Adversarial Learning and GANs for Specular Representation**: GANs create synthetic specular images which contain authentic reflections for their visual output. The training process enables models to acquire knowledge about specific characteristics which define specular objects when they experience different types of illumination and environmental situations.
5. **Physics-Based Rendering (PBR) in Synthetic Data**: PBR simulates realistic lighting and reflection behavior, which is valuable for modeling specular properties in synthetic training data. Using PBR-generated synthetic images, models can train under controlled yet varied lighting and reflection conditions, leading to better generalization when detecting specular objects in the real world.

V.SYNTHETIC DATA GENERATION

1. Rule-Based Synthetic Data Generation:

Rule-based methods involve creating data according to predefined rules or distributions. This approach is often used for tabular data and is effective when users have a clear understanding of the relationships between variables. In finance, rule-based systems create transaction simulations through their ability to define connections between different transaction parameters which include transaction amount and time and account type. The implementation of rule-based methods proves to be simple yet they fail to provide the necessary flexibility needed to create complex data types such as images and text because their variable relationships remain difficult to understand.

2.Statistical Methods and Sampling:

The statistical methods create data which replicates actual data characteristics and statistical distribution patterns. The numerical datasets contain variables which can be modeled through distributions such as Poisson and Gaussian that depict their changing patterns and statistical distribution. The Monte Carlo methods function as statistical sampling techniques which use generated random variables to create artificial data for risk assessment and financial prediction and engineering studies. The healthcare industry which requires strict patient privacy protection can use these techniques to protect medical data through anonymization while they create new datasets which maintain original statistical properties [23].

3. Generative Models (GANs, VAEs):

The discriminator creates authentic data through its process of comparing generated samples to real samples which drives the generator to improve its creations [24]. The Variational Autoencoder (VAE) serves as a second widely used generative model which researchers employ to produce high-dimensional data. The VAE system requires data to undergo encoding before it can proceed to its operational workflow which involves creating a compressed representation of the information. The system enables sample generation through the process of decoding which requires the produced data to closely resemble the established training dataset. The VAE system proves beneficial for applications that require interpretable results because it generates synthetic medical images with stable outputs through controlled generation methods. While GANs are more widespread in fields like text and audio production, researchers have modified VAEs for tasks like picture generation as well.

4.Agent-Based Modeling and Simulation:

The purpose of agent-based modeling (ABM) is to create synthetic data through the development of virtual environments which use autonomous agents that follow predefined behaviors. The simulations create accurate representations of complex real-world environments which include traffic systems and economic marketplaces and human social activities. Researchers can create realistic dynamic data by studying how agents interact with each other and with external factors. Application-based modelling (ABM) finds extensive usage in domains like as autonomous vehicle training, where agents (vehicles, pedestrians) engage in regulated situations to provide data for training models. This method demonstrates its effectiveness when handling intricate multi-agent scenarios which require advanced modeling techniques [25].

5.Computer Graphics and 3D Rendering:

Synthetic data is often created utilising 3D modelling and rendering technologies for computer vision applications. This method enables programmers to construct carefully regulated virtual worlds, objects, and scenarios. An example of a synthetic dataset would include 3D-generated images of objects in various lighting conditions, with varying perspectives and backgrounds, used to train object identification algorithms.

Synthetic pictures may now be made to seem almost identical to genuine ones using methods such as imagerealistic rendering. 3D simulations provide safe and cost-effective settings to test models under numerous circumstances, such as weather, road kinds, and traffic patterns, which are essential in training autonomous cars. Unity, Unreal Engine, and Omniverse from NVIDIA are among of the most well-known platforms for creating synthetic 3D data [26-30].

6.Synthetic Data from Differential Privacy Techniques:

Applications with strict privacy needs might benefit from synthetic data that is created using differential privacy techniques by injecting controlled noise to baseline datasets. By ensuring that the generated data closely resembles the original data, sensitive information is protected. Differential privacy solutions are well-suited for sensitive social, financial, and healthcare datasets due to the delicate balance they achieve between data value and privacy. Using methods such as DP-GANs or synthetic datasets with differential privacy guarantees, businesses may use data for machine learning while still meeting legal standards [31-35].

VI.DOMAIN RANDOMIZATION

Domain randomisation is a technique in which synthetic training data is deliberately diversified to include a broad spectrum of visual events, hence enhancing the generalisation capabilities of object identification models when confronted with real-world data. This method is often used to connect synthetic and real-world domains by using differences in colour, lighting, textures, and noise to replicate varied real-world settings [36].

1. Color and Texture Variability:

Randomized Color Palettes: By altering colours in synthetic images, models are trained to concentrate on an object's structural characteristics instead than particular colour indicators. This method assists models in identifying things under diverse lighting conditions, where colours may manifest variably[37].

Texture Changes: Changing the textures within an object class, such as turning a smooth texture into a rough one, helps the model understand that texture isn't always an object-defining feature. This is essential for detecting objects in scenes where the object's texture might be obscured or altered[38].

2. Lighting and Shadow Modifications:

Random Lighting Intensity: Different lighting intensities allow the model to learn how objects appear under dim, bright, or uneven lighting conditions, which is common in natural scenes.

Shadow Effects: Introducing shadows at different angles and intensities, or removing them, helps the model detect objects regardless of shadowing. This adjustment is crucial in applications like autonomous driving, where vehicles pass through shaded areas frequently[39,40].

3. Object Placement and Scale Randomization:

Positional Variability: Objects are placed randomly within the frame or scene, and their orientations are varied to allow detection from multiple angles and distances. The aerial imaging system and the surveillance system require this method because they track objects which present themselves in different scales and angles.

Scale and Size Adjustments: The model learns to handle both distant and close-up views through size alterations of objects in synthetic scenes which show how objects would appear at different distances[42].

4. Background and Environmental Changes:

Background Randomization: The researchers developed a method for creating synthetic images which uses random background generation to compel models to concentrate their attention on objects rather than their surrounding environment. The procedure requires scene background alterations which include transforming between urban and rural and indoor and outdoor environments according to specific application requirements.

5. Occlusions and Noise Addition:

Simulated Occlusions: The system uses random obscuration of object components to improve its ability to identify partially hidden objects which are obstructed by adjacent objects in actual environmental conditions. The process of adding noise to images creates artificial low-quality visual effects which simulate the appearance of images captured by substandard cameras and during difficult nighttime conditions. The models achieve their highest performance because they can handle both image noise and all types of image quality loss.[43]

VII.DISCUSSION

Object detection has evolved from its original advanced state into a fundamental technology which now enables multiple fields of work to operate, including self-driving vehicles and medical evaluation and factory automation. The combination of CNNs and GANs and domain randomisation and few-shot/zero-shot learning methods has created a complicated system which researchers use to develop and test object recognition systems in various challenging environments. The discussion examines both the advantages and disadvantages of different methods while analyzing their overall impact and the existing challenges which scientists face when studying this field.

The CNN-based methods used for object recognition research successfully complete their task by using complex visual data to find and identify objects in images. Through their hierarchical design the system achieves a partial simulation of human visual processing that enables it to detect both fundamental and advanced visual elements. The primary weakness of CNN systems lies in their requirement for large amounts of data and their need for extensive computational power. Object identification training for convolutional neural networks needs large annotated datasets which require expensive and difficult work to gather. Synthetic data, particularly pictures created by GANs, provides a distinct benefit. GANs provide a proficient remedy by generating high-quality synthetic pictures that may address data deficiencies or augment training variety, especially in challenging settings, infrequent instances, or unbalanced datasets. Although GANs generate plausible data, they present inherent problems. Training Generative Adversarial Networks (GANs) is characteristically unstable, often requiring meticulous adjustments and substantial computing resources to get the required quality and realism.

Domain adaption and randomisation facilitate the closure of the synthetic-to-real gap, an essential process for the appropriate use of synthetic data in real-world scenarios. Domain randomisation enhances model generalisation to diverse real-world situations by randomising features such as lighting, texturing, and object placements in synthetic pictures. Domain adaptation, conversely, emphasises the alignment of feature distributions between synthetic and actual data, hence enhancing generalisation. Collectively, these techniques enhance the efficacy of synthetic data; yet, attaining seamless domain adaptation continues to pose difficulties. Certain models based on synthetic data continue to encounter difficulties when faced with real-world changes that were not considered during training. This prompts an inquiry into whether more progress in domain adaptation methodologies, maybe via hybrid GAN-domain adaptation models, might improve the applicability of synthetic data for a broader range of use cases.

Few-shot and zero-shot learning provide alternative methods for addressing data scarcity, enabling models to identify novel objects with limited or no labelled data. These approaches are essential for applications necessitating rapid flexibility, such as medical imaging for unusual illnesses or autonomous navigation in new settings.

The studied methods show significant improvements for object detection yet there remain unsolved problems. Object detection requires three essential factors to function correctly in healthcare and security applications which include real-time performance and data protection and system understanding. The requirements for real-time systems demand development of models which achieve high accuracy and operational efficiency on devices with limited computing power. Data privacy becomes vital when organizations develop models using data which contains sensitive or personal details about individuals. The existing research on federated learning and privacy-preserving synthetic data methods provides solutions for these issues yet both methods still need further investigation. The model assessment process faces challenges because complex systems like CNNs and GANs and ensemble methods create hidden decision pathways which remain hidden from users. The public needs to access transparent understanding of how these models make decisions because this knowledge enables them to trust the system which operates in high-stakes situations.

The presented strategies demonstrate multiple ways that different object detection methods can improve each other. The combination of Convolutional Neural Networks with Generative Adversarial Network-based synthetic data and domain adaptability and few-shot and zero-shot learning and ensemble approaches creates a powerful toolkit for building robust object detection systems. Future advancements will work to optimize the existing methods while enhancing their performance and making them easier to understand for solving existing problems in this important and evolving field.

VIII.CONCLUSION

Object detection models have become essential across multiple industries because they support multiple applications which include driverless cars and medical imaging and security systems and retail analytics. The variety of use cases demonstrates how this field produces widespread impact which requires continuous research to develop solutions for data and computing and deployment challenges. The introduction of GANs has transformed synthetic data creation because it allows researchers to produce training datasets which help build more robust machine learning models. Synthetic data together with domain randomisation and adaptation methods enables models to overcome data limitations by acquiring the ability to operate in multiple real-world environmental contexts. The effective use of synthetic data requires researchers to advance two areas which include improving GAN training stability and developing better domain adaptation techniques. The combination of low-shot and zero-shot learning methods allows models to learn new object categories with minimal data requirements which enables them to operate in environments that involve rare object detection. These strategies allow researchers to use object identification in environments that lack enough data for standard identification while the models need new adjustments to achieve proper performance in different environments.

Ensemble learning achieves its highest accuracy and object recognition reliability through its method of combining various models. The methodology proves most effective in fields that require absolute accuracy, which include autonomous driving and security surveillance because multiple models work together to achieve better detection results and lower false positive rates. The distinct methodologies from different approaches show their ability to work together with ensemble systems, which demonstrate

their practical effectiveness through performance enhancements. The real-time application potential of ensemble methods will increase because of their improved computational efficiency through model compression techniques.

The existing problems in object detection require additional research, which must continue until scientists find solutions. The field needs more research on model interpretability and computational efficiency and data privacy protection, especially for object detection systems that operate in sensitive healthcare environments. The ethical and practical standards of object detection systems require these challenges to be solved before society can accept these technologies. The decision-making process behind complex models needs to be shown through interpretability tools, which should be implemented in CNNs and GAN-based synthetic models to boost user trust and transparent operation of object detection systems.

The future of research will benefit from an integrated approach that combines multiple object detection techniques into flexible systems that work together. The project will develop hybrid systems that unite CNN components with GAN-based synthesis and domain adaptation and ensemble learning to create an adaptable system that handles new object classes and changing environments. The upcoming frameworks will develop an advanced object recognition system, which will work throughout multiple fields, especially with the growth of real-time edge-based systems.

REFERENCES

1. Ali, Mahmoud Atta Mohammed, et al. "Advancing Crowd Object Detection: A Review of YOLO, CNN and ViTs Hybrid Approach." Journal of Intelligent Learning Systems and Applications 16.3 (2024): 175-221.
2. Xie, X., Wu, D., Xie, M., & Li, Z. (2024). GhostFormer: Efficiently amalgamated CNN-transformer architecture for object detection. Pattern Recognition, 148, 110172.
3. Ben Saad, A., Facciolo, G., & Davy, A. (2024). On the Importance of Large Objects in CNN Based Object Detection Algorithms. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (pp. 533-542).
4. Haryono, A., Jati, G., & Jatmiko, W. (2024). Oriented object detection in satellite images using convolutional neural network based on ResNeXt. ETRI Journal, 46(2), 307-322.
5. Sagar, A. S., Chen, Y., Xie, Y., & Kim, H. S. (2024). MSA R-CNN: A comprehensive approach to remote sensing object detection and scene understanding. Expert Systems with Applications, 241, 122788.
6. Amjoud, A. B., & Amrouch, M. (2023). Object detection using deep learning, CNNs and vision transformers: A review. IEEE Access, 11, 35479-35516.
7. Sun, P., Zhang, R., Jiang, Y., Kong, T., Xu, C., Zhan, W., ... & Luo, P. (2023). Sparse r-cnn: An end-to-end framework for object detection. IEEE transactions on pattern analysis and machine intelligence.
8. Ye, T., Qin, W., Zhao, Z., Gao, X., Deng, X., & Ouyang, Y. (2023). Real-time object detection network in UAV-vision based on CNN and transformer. IEEE Transactions on Instrumentation and Measurement, 72, 1-13.
9. Vaishnavi, K., Reddy, G. P., Reddy, T. B., Iyengar, N., & Shaik, S. (2023). Real-time object detection using deep learning. Journal of Advances in Mathematics and Computer Science, 38(8), 24-32.
10. Li, W. Z., Zhou, J. W., Li, X., Cao, Y., & Jin, G. (2023). Few-shot object detection on aerial imagery via deep metric learning and knowledge inheritance. International Journal of Applied Earth Observation and Geoinformation, 122, 103397.
11. Demirel, B., Baran, O. B., & Cinbis, R. G. (2023). Meta-tuning loss functions and data augmentation for few-shot object detection. In proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 7339-7349).
12. Han, J., Ren, Y., Ding, J., Yan, K., & Xia, G. S. (2023, June). Few-shot object detection via variational feature aggregation. In Proceedings of the AAAI Conference on Artificial Intelligence (Vol. 37, No. 1, pp. 755-763).
13. Wang, Y., Zou, X., Yan, L., Zhong, S., & Zhou, J. (2024). SNIDA: Unlocking Few-Shot Object Detection with Non-linear Semantic Decoupling Augmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (pp. 12544-12553).
14. Fan, Q., Zhuo, W., Tang, C. K., & Tai, Y. W. (2024). FSODv2: A Deep Calibrated Few-Shot Object Detection Network. International Journal of Computer Vision, 1-20.
15. Shangguan, Z., & Rostami, M. (2023). Identification of novel classes for improving few-shot object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 3356-3366).
16. Xin, Z., Wu, T., Chen, S., Zou, Y., Shao, L., & You, X. (2024). Ecea: Extensible co-existing attention for few-shot object detection. IEEE Transactions on Image Processing.
17. Noel, G. P. (2024). Evaluating AI-powered text-to-image generators for anatomical illustration: a comparative study. Anatomical Sciences Education, 17(5), 979-983.
18. Horvath, A. S., & Pouliou, P. (2024). AI for conceptual architecture: Reflections on designing with text-to-text, text-to-image, and image-to-image generators. Frontiers of Architectural Research, 13(3), 593-612.
19. Xue, Z., Song, G., Guo, Q., Liu, B., Zong, Z., Liu, Y., & Luo, P. (2024). Raphael: Text-to-image generation via large mixture of diffusion paths. Advances in Neural Information Processing Systems, 36.
20. Khachatryan, L., Movsisyan, A., Tadevosyan, V., Henschel, R., Wang, Z., Navasardyan, S., & Shi, H. (2023). Text2video-zero: Text-to-image diffusion models are zero-shot video generators. In Proceedings of the IEEE/CVF International Conference on Computer Vision (pp. 15954-15964).
21. Baboo, A., Mishra, S. R., & Dash, S. (2024, November). An Improvised Diabetes Prediction System Using Hybrid Ensemble Approach. In 2024 IEEE 11th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) (pp. 1-6). IEEE.
22. Reith, T.P.; D'Alessandro, D.M.; D'Alessandro, M.P. Capability of multimodal large language models to interpret pediatric radiological images. Pediatr. Radiol. 2024, 54, 1729–1737.
23. Martin, S.A.; Zhao, A.; Qu, J.; Imms, P.E.; Irimia, A.; Barkhof, F.; Cole, J.H.; Initiative, A.D.N. Explainable artificial intelligence for neuroimaging-based dementia diagnosis and prognosis. medRxiv 2025.
24. Mishra, S. R., & Dash, S. (2026). AI-Driven Remote Health Monitoring for Predicting Diabetes and Heart Diseases Using ULMCSO and PGND Models. Hyper-Intelligent Networks: Exploring the Future of Connectivity for Society 5.0, 219-247.
25. Mishra, S. R., Dash, S., Padhy, S., & Samuel, P. (2026). Legal Aspects of Operating IoMT Applications in the Fog Computing. In Integrating Cloud, Fog, and Edge Computing in Healthcare: Federated Learning and Blockchain Approaches: Harnessing Distributed Technologies for Enhanced Healthcare Delivery (pp. 211-224). Cham: Springer Nature Switzerland.

26. Xue, C.; Kowshik, S.S.; Lteif, D.; Puducheri, S.; Jasodanand, V.H.; Zhou, O.T.; Walia, A.S.; Guney, O.B.; Zhang, J.D.; Pham, S.T.; et al. AI-based differential diagnosis of dementia etiologies on multimodal data. Nat. Med. 2024, 30, 2977–2989.
27. Schilcher, J.; Nilsson, A.; Andlid, O.; Eklund, A. Fusion of electronic health records and radiographic images for a multimodal deep learning prediction model of atypical femur fractures. Comput. Biol. Med. 2024, 168, 107704.
28. Rath, L., Mishra, S. R., Dash, S., Pradhan, P. C., & Baboo, A. (2025). Predicting diabetic patients coronary artery calcium score, deep learning using retinal images. In Intelligent Computing Techniques and Applications (pp. 113-118). CRC Press.
29. Pattnayak, A. P., Mishra, S. R., Dash, S., & Baboo, A. (2025). Utilization of deep learning and machine learning models to approach high glucose and low glucose prediction with type 1 diabetes mellitus in adult patients. In Intelligent Computing Techniques and Applications (pp. 102-107). CRC Press.
30. Dash, A. B., Dash, S., Padhy, S., Kumar, N., Pati, G. K., & Uthansingh, K. (2025). Leveraging inception-v3 CNN model for efficient image classification. In Intelligent Computing and Communication Techniques (pp. 341-348). CRC Press.
31. Niu, S.; Ma, J.; Bai, L.; Wang, Z.; Guo, L.; Yang, X. EHR-KnowGen: Knowledge-enhanced multimodal learning for disease diagnosis generation. Inf. Fusion 2024, 102, 102069.
32. Dora, N., Dash, S., Baboo, A., & Mishra, S. R. (2025, August). Efficient Nail Disease Diagnosis Using Deep Neural Networks for Predicting Abnormalities. In 2025 International Conference on Next Generation of Green Information and Emerging Technologies (GIET) (pp. 1-5). IEEE.
33. Mishra, S. R., Dash, S., & Rath, L. (2024, November). Effective Diabetes Mellitus Prediction Using a Hybrid Ensemble Machine Learning Model with Iot. In 2024 International Conference on Integrated Intelligence and Communication Systems (ICIICS) (pp. 1-8). IEEE.
34. Dash, A. B., Dash, S., Padhy, S., Mishra, B., & Paikaray, B. K. (2025). Streamlining colorectal cancer diagnosis: leveraging MobileNet-V3 for efficient image classification. Int. J. Internet Manufacturing and Services, 11(4), 317
35. Zeng, L.; Ma, P.; Li, Z.; Liang, S.; Wu, C.; Hong, C.; Li, Y.; Cui, H.; Li, R.; Wang, J.; et al. Multimodal Machine Learning-Based Marker Enables Early Detection and Prognosis Prediction for Hyperuricemia. Adv. Sci. 2024, 11, 2404047.
36. Khuntuli, B., Dash, S., Pradhan, P. C., & Mishra, S. R. Combating food insecurity through remote sensing and machine learning for enhanced crop yield prediction. In Intelligent Computing Techniques and Applications (pp. 135-140). CRC Press.
37. Sahu, P. K., Biswal, B. B., Mishra, S. R., Padhy, J., & Kumar, D. (2025, March). Demand-Based Secured Data Transmission in WSN. In International Conference on Next Generation Computing and Communication Applications (pp. 37-44). Cham: Springer Nature Switzerland.
38. Baboo, A., Mishra, S. R., & Dash, S. (2024, November). An Improvised Diabetes Prediction System Using Hybrid Ensemble Approach. In 2024 IEEE 11th Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON) (pp. 1-6). IEEE.
39. Li, B.; Chen, H.; Lin, X.; Duan, H. Multimodal Learning system integrating electronic medical records and hysteroscopic images for reproductive outcome prediction and risk stratification of endometrial injury: A multicenter diagnostic study. Int. J. Surg. 2024, 110, 3237–3248.
40. Zhu, L.; Lai, Y.; Ta, N.; Cheng, L.; Chen, R. Multimodal approach in the diagnosis of urologic malignancies: critical assessment of ChatGPT-4V's image-reading capabilities. JCO Clin. Cancer Inform. 2024, 8, e2300275.
41. Lin, A.C.; Liu, Z.; Lee, J.; Ranvier, G.F.; Taye, A.; Owen, R.; Matteson, D.S.; Lee, D. Generating a multimodal artificial intelligence model to differentiate benign and malignant follicular neoplasms of the thyroid: A proof-of-concept study. Surgery 2024, 175, 121–127.
42. Baboo, A., Patro, S. P., & Dash, S. (2024, December). A Deep Learning Approach for Enhancing Cardiovascular Disease Prediction Using ECG Data. In 2024 2nd International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES) (pp. 1-5). IEEE.Muzahid, A. A. M., Wanggen, W., Sohel, F., Bennamoun, M., Hou, L., & Ullah, H. (2021). Progressive conditional GAN-based augmentation for 3D object recognition. Neurocomputing, 460, 20-30.
43. Kumpatla, G., Veresi, H., Abhishek, S., & Anjali, T. (2023, October). Revolutionizing brain tumour prediction: A pioneering gan-based framework for synthetic data generation. In 2023 7th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC) (pp. 548-553). IEEE.