International Journal of Technology Research and Management
ISSN (Online): 2348-9006
Vol 11 Issue 08 August 2024

Study of various Approaches used for Object Detection and Recognition by AI and Natural Language Processing

Prabhati Bharti[1], Asst. Prof. Ayush Kumar[2]
Research Scholar[1], Assistant Professor[2]
Department Of Computer Science & Engineering, Radha Raman Engineering College, Bhopal (M.P.)[1, 2]
itsprabhati16@gmail.com[1]

Abstract: *In recent years, the convergence of computer vision and artificial intelligence has significantly enhanced object identification and recognition systems. This research explores contemporary computer vision methodologies aimed at advancing these systems. The study begins with a discussion on the foundational concepts of object detection, highlighting the transition from traditional methods to deep learning techniques. It delves into convolution neural networks (CNNs) and their role in revolutionizing object recognition by automating the hierarchical extraction of features from visual data. The importance of dataset annotation and the emergence of large-scale annotated datasets, which are crucial for training and evaluating robust object identification models. The utilization of transfer learning and domain adaptation is examined to improve model generalization across diverse environments. Additionally, the study considers the challenges posed by occlusion, scale variations, and different environmental conditions in real-world applications. The paper reviews region-based and anchor-based methods designed to enhance the accuracy and efficiency of object detection. It also explores the integration of geographical and semantic information, highlighting the synergy between object detection and recognition. By examining AI algorithms and natural language processing (NLP) for contextual awareness, the study suggests pathways to more sophisticated systems capable of interpreting the relationships and interactions between identified objects. In this paper we discusses current trends and future directions in the field, such as 3D object detection, federated learning for decentralized systems, and ethical considerations in AI-based object detection and recognition technologies. The potential impact of these advancements on autonomous vehicles, surveillance, healthcare, and robotics is also considered, emphasizing the transformative possibilities of artificial intelligence in object detection and recognition.*

Keywords: *AI, Part-of-speech (POS), Parsing, NLP.*

1. INTRODUCTION

Object identification and recognition using artificial intelligence (AI) constitute a significant advancement in computer vision, a branch of AI focused on training computers to interpret and understand visual input. This technology's growing popularity spans numerous industries, including agricultural monitoring, security surveillance, healthcare, and self-driving cars, due to its wide-ranging applications [1].

Object detection involves training AI systems to locate and recognize specific objects within videos or photographs. This is achieved through complex algorithms that analyze visual input and identify patterns unique to objects. AI algorithms typically learn from large datasets containing labeled images, which enable the systems to distinguish between various items accurately.

Recognition goes beyond merely identifying objects to understanding what those objects are [2]. This requires more sophisticated algorithms capable of categorizing objects into predefined groups. For instance, an AI system can accurately recognize traffic signs in a street setting, a crucial function for autonomous driving applications, as illustrated in Figure 1.

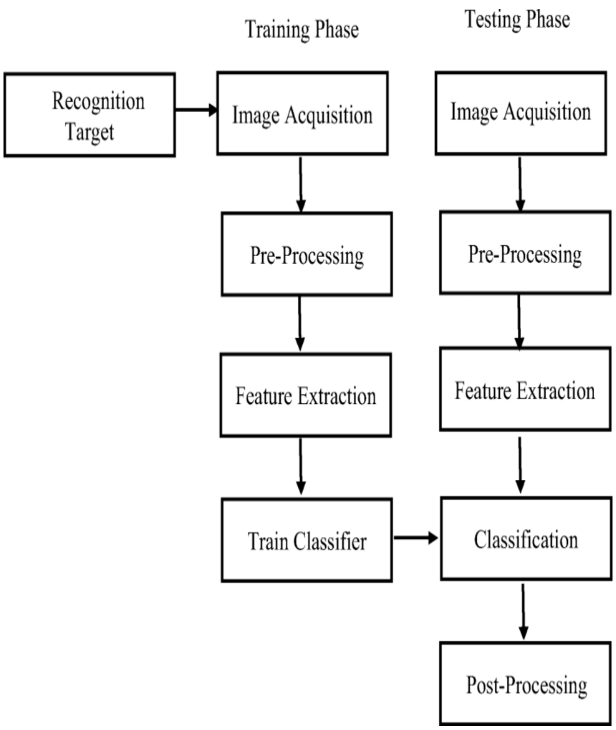

Figure 1: The AI-based object manufacturing process

The integration of AI into object identification and detection has been propelled by advancements in machine learning, particularly deep learning. Convolution neural networks (CNNs), a type of neural network, drive these advancements. CNNs mimic the way the human brain processes visual information, allowing AI systems to detect and differentiate objects with remarkable precision and efficiency.

The impact of AI in this field is profound. It not only streamlines and enhances the efficiency of tasks traditionally performed by humans but also opens up new opportunities that were previously unattainable. In healthcare, for example, AI-enabled object recognition can assist in diagnosing illnesses based on medical imaging. In retail, it can aid in inventory management and customer behavior tracking.

A closer examination of contemporary computer vision methodologies reveals the foundational concepts of object detection and the shift from traditional approaches to deep learning. Deep learning has revolutionized object recognition by automating the hierarchical extraction of features from visual data. This automation is crucial for handling the vast amount of data required to train robust object identification models.

Dataset annotation and the rise of large-scale annotated datasets play a pivotal role in this process. These datasets are essential for training and evaluating models [3], ensuring they can generalize well across diverse environments. Transfer learning and domain adaptation further enhance model generalization by leveraging pre-existing knowledge to improve performance in varied settings.

Real-world applications present challenges such as occlusion, scale variations, and different environmental conditions. Addressing these challenges requires innovative approaches to ensure AI systems perform reliably in diverse scenarios. Region-based and anchor-based methods have been developed to enhance the accuracy and efficiency of object detection. These methods focus on refining the process of locating and identifying objects within an image or video, improving both speed and precision.

The integration of geographical and semantic information highlights the synergy between object detection and recognition. By combining these elements, AI systems can achieve contextual awareness, interpreting the relationships and interactions between identified objects. This is particularly relevant in applications where understanding the context is as important as recognizing the objects themselves.

Exploring the intersection of AI algorithms and natural language processing (NLP) opens up possibilities for even more sophisticated systems [4]. These systems can interpret not only visual data but also the context and meaning behind it. For instance, in autonomous driving, recognizing a stop sign is important, but understanding its significance in the context of the surrounding traffic environment is crucial for safe navigation.

Current trends and future directions in the field include 3D object detection, which offers a more comprehensive understanding of objects in a three-dimensional space. Federated learning, another emerging trend, enables decentralized systems to learn from data distributed across multiple devices without compromising privacy. This approach is particularly valuable in applications where data security and privacy are paramount.

Ethical considerations are also becoming increasingly important as AI-based object detection and recognition technologies advance. Ensuring these technologies are developed and deployed responsibly is crucial to avoid potential biases and ensure they benefit society as a whole. This includes addressing issues such as data privacy, algorithmic fairness, and transparency in AI decision-making processes.

2. LITERATURE REVIEW

The intersection of artificial intelligence (AI) and computer vision has led to significant advancements in object detection and recognition, influencing various applications such as autonomous driving, healthcare, surveillance, and more. This literature review explores key studies and methodologies in the field, examining how deep learning and machine learning have revolutionized these technologies.

Yang et al. (2017) proposed an image captioning approach that integrates object detection and localization, highlighting the role of deep learning in generating descriptive captions for images by identifying and localizing objects within them [5]. This method improves the contextual understanding of visual scenes, essential for applications requiring detailed image descriptions.

Zhang et al. (2017) introduced discriminative bimodal networks that enhance visual localization and detection through natural language queries [6]. This approach allows for more accurate object identification by leveraging both visual and textual data, demonstrating the effectiveness of multimodal techniques in improving object detection accuracy.

Pathak et al. (2018) reviewed the application of deep learning in object detection, focusing on the transformative impact of convolution neural networks (CNNs) [7]. They emphasized how deep learning algorithms have outperformed traditional methods by automatically extracting hierarchical features from visual data, leading to significant improvements in detection accuracy and efficiency.

Othman and Aydin (2018) explored a deep learning application using the Movidius Neural Compute Stick (NCS) for embedded object detection and recognition [8]. Their work demonstrated the feasibility of deploying deep learning models on edge devices, highlighting the potential for real-time, low-power object detection applications in various fields.

Sharma et al. (2019) discussed practical approaches to image and text processing using machine learning techniques [9]. They highlighted the integration of image processing with natural language processing (NLP) to enhance the contextual understanding of detected objects, paving the way for more intelligent and interactive systems.

Yang et al. (2019) examined deep learning applications in NLP, illustrating how these techniques can be applied to improve object recognition by providing contextual insights from textual data [10]. This integration enhances the overall understanding of visual scenes, making AI systems more robust and versatile.

Zhao et al. (2019) provided a comprehensive review of object detection with deep learning, detailing the evolution of techniques from traditional methods to advanced deep learning models [11]. They discussed various architectures, such as region-based CNNs (R-CNN) and single-shot detectors (SSD), which have significantly improved detection performance.

Gollapudi (2019) explored the synergy between AI and computer vision, emphasizing the importance of combining deep learning with other AI techniques to enhance object detection and recognition [12]. Their study provided insights into how AI can be leveraged to interpret complex visual scenes and improve system accuracy.

Haidar et al. (2019) were early innovators in applying object recognition techniques to text detection, showcasing how these methods can be effectively adapted to identify textual elements within images [13]. Their work marked a significant advancement in the field by extending the capabilities of object recognition beyond traditional object identification to include the extraction and recognition of text from visual data. This pioneering approach demonstrated the versatility of object recognition technologies and paved the way for more sophisticated systems that can seamlessly integrate text detection into broader image analysis applications. Their research contributed to expanding the scope of object recognition methodologies.

Kumar et al. (2019) advanced the field by integrating deep learning techniques into image caption generation systems [14]. Their work highlights the crucial role of precise object recognition in crafting coherent and contextually accurate image descriptions. By leveraging deep learning, they demonstrated how improved object identification can enhance the quality of generated captions, ensuring that descriptions are not only relevant but also reflective of the visual content. This approach underscores the significance of accurate object recognition in generating meaningful and contextually appropriate image descriptions, thereby contributing to the broader advancement of automated image captioning systems.

The use of adaptive convolution neural networks (CNNs) was explored by Bapu et al. (2019), who utilized N-gram models for spatial object recognition, highlighting the effectiveness of CNNs in handling spatial data . This approach has been fundamental in advancing object detection accuracy and efficiency [15].

Torfi et al. (2020) performed an extensive survey on the progress of natural language processing (NLP) propelled by deep learning, highlighting the increasing integration between NLP and object detection techniques [16]. Their review details how advancements in deep learning have significantly enhanced NLP capabilities and explores the collaborative potential of combining NLP with object detection. By examining the evolving landscape of these technologies, the study underscores the growing synergy between NLP and object detection, demonstrating how deep learning methodologies are driving innovations and improvements across both fields. This work reflects the dynamic interplay between these areas of artificial intelligence.

The integration of deep learning with object detection is essential for applications such as autonomous driving. Li et al. (2020) exemplify this by presenting a hybrid framework that combines these technologies to enable real-time decision-making in vehicles [17]. Their approach leverages deep learning to enhance the accuracy of object detection systems, which is critical for the safe and effective operation of autonomous vehicles. By integrating these advanced techniques, their framework provides a robust solution for interpreting complex driving environments, illustrating the importance of merging deep learning and object detection in modern autonomous driving systems.

Vashisht and Kumar (2020) conducted a comprehensive survey of object detection methods within image processing, presenting a thorough comparative analysis of both traditional and contemporary techniques [18]. Their study offers a detailed examination of various approaches, highlighting the evolution from earlier methods to advanced modern techniques. By exploring the strengths and limitations of different object detection strategies, the survey provides valuable insights into how these methods have developed and the advancements that have been made in the field. This extensive review serves as a critical resource for understanding the progress and current state of object detection technologies.

Object detection has proven highly valuable in assistive technologies. Joshi et al. (2020) developed an advanced system for multi-object detection and navigation specifically designed to assist visually impaired individuals [19]. By utilizing artificial intelligence, their system enhances mobility and independence for users. The technology enables real-time detection and recognition of multiple objects in the environment, providing crucial information that aids in navigation and spatial awareness. This innovative application demonstrates the potential of AI-driven object detection systems to significantly improve the quality of life for those with visual impairments, offering greater autonomy and support in everyday activities.

Van Le et al. (2020) created a multimodal assistive solution by integrating image and speech recognition into their object detection system [20]. This innovative approach combines visual and auditory inputs to enhance user interaction and accessibility. The system leverages image recognition to identify objects and speech recognition to provide spoken descriptions and instructions. By merging these two modalities, the technology offers a more comprehensive assistive experience, improving usability for individuals who rely on both visual and auditory cues. This integration underscores the potential of combining different sensory inputs to create more effective and user-friendly assistive technologies.

Deep learning methodologies have greatly enhanced object detection systems. Grover et al. (2021) developed AI Optics, a groundbreaking system that leverages deep learning for object recognition and caption generation, specifically aimed at assisting the blind [21]. This system offers real-time descriptive feedback by analyzing and identifying objects in the user's environment. The use of advanced deep learning techniques enables accurate and timely descriptions, significantly improving the accessibility and usability of the system for visually impaired individuals. AI Optics exemplifies how deep learning can transform object detection technologies to provide valuable assistance and support.

Srivastava et al. (2021) present a sign language recognition system utilizing Tensor Flow's object detection API [22]. Their research is significant in the field of assistive technologies, particularly for individuals with hearing and speech impairments. The system demonstrates the capability of machine learning models to recognize various hand gestures associated with sign language, leveraging the robust features of TensorbFlow to achieve high accuracy and real-time processing capabilities. The study highlights the integration of object detection frameworks with deep learning techniques to facilitate effective communication for the differently-abled community.

Kocaleva and Koceski (2021) provide a comprehensive overview of image recognition and real-time object detection [23]. Their survey covers various methodologies and advancements in the field, discussing the evolution of algorithms and technologies that have significantly improved the accuracy and efficiency of object detection systems. They delve into the practical applications of these technologies across different domains, emphasizing their importance in enhancing automation and decision-making processes in real-time scenarios.

Liu et al. (2021) explore an object detection and recognition system based on computer vision analysis [24]. The system employs advanced computer vision techniques to analyze and interpret visual data, enabling the detection and recognition of objects with high precision. The study underscores the integration of machine learning algorithms with computer vision to develop robust object detection frameworks that can be applied in various industrial and research settings.

Skepetaris (2022) in his master's thesis delves into deep learning techniques for object detection [25]. He provides an in-depth analysis of various deep learning models and their applications in object detection, highlighting the advancements and challenges in the field. His work is a crucial addition to the existing literature, offering insights into the practical implementation and performance evaluation of deep learning models in detecting and recognizing objects.

Olszewska (2022) introduces an explainable artificial intelligence approach for automatic object detection and recognition [26]. Her research focuses on the development of interpretable models that not only perform object detection but also provide explanations for their predictions. This approach enhances the transparency and trustworthiness of AI systems, making them more reliable for critical applications where understanding the decision-making process is essential.

Kaur et al. (2022) present a survey on deep learning approaches to medical image analysis and real-time object detection [27]. Their work systematically reviews various deep learning models and their applications in the medical field, highlighting the potential of these models to improve diagnostic accuracy and efficiency. They also discuss the challenges and future directions in the integration of real-time object detection in medical imaging, emphasizing the need for further research to enhance the performance and applicability of these systems.

Abdelreheem et al. (2022) present the 3DRefTransformer, a model developed for detailed object identification in real-world scenes through natural language processing [28]. This model utilizes transformer architectures to enhance the accuracy of object recognition, particularly in complex environments. By integrating advanced transformer techniques, 3DRefTransformer improves the precision of identifying and describing objects based on detailed natural language inputs. The approach demonstrates significant effectiveness in navigating and interpreting intricate scenes, highlighting its potential to advance fine-grained object identification and enhance the usability of object detection systems in varied and challenging contexts.

Mondal et al. (2022) explore the application of deep learning for handwritten English word recognition [29]. They propose a deep learning-based object detection architecture tailored for this purpose, showing that such models can significantly enhance the accuracy and efficiency of handwritten text recognition compared to traditional methods.

Rupa et al. (2022) present an innovative application of object detection for public health [30]. Their research focuses on medicine drug name detection using augmented reality (AR). By integrating AR with object recognition technologies, their system aims to improve the accessibility and accuracy of drug information, which is crucial for both healthcare professionals and patients.

Zou et al. (2023) present a comprehensive survey that spans two decades of advancements in object detection [31]. Their analysis explores a wide range of methodologies and technological progress in the field. By reviewing historical developments and current trends, they provide valuable insights into how object detection techniques have evolved over time. The survey also identifies emerging research areas and future directions, offering a thorough understanding of the field’s trajectory. This extensive review serves as a crucial resource for researchers and practitioners aiming to grasp the past achievements and anticipate future advancements in object detection.

Kaur and Singh (2023) provide a thorough review of object detection using deep learning techniques [32]. Their work highlights the major advancements deep learning has brought to the field, focusing on different architectures and models that have significantly enhanced both detection accuracy and speed. By examining the impact of these techniques, their review underscores the transformative role deep learning plays in improving object detection systems. This comprehensive analysis offers valuable insights into how deep learning innovations have shaped the current landscape and advanced the effectiveness of object detection technologies.

Kaur and Singh (2023) provide an in-depth review of object detection advancements driven by deep learning [33]. Their analysis highlights the substantial impact of deep learning techniques on enhancing detection accuracy and speed. The review covers a range of architectures and models that have been pivotal in advancing the field, illustrating how these innovations have contributed to more effective and efficient object detection. By examining key developments and breakthroughs, their work underscores the transformative role of deep learning in improving performance and capabilities within the object detection domain.

Alzahrani and Al-Baity (2023) propose a novel object recognition system aimed at aiding the visually impaired [34]. Their approach leverages deep learning techniques combined with Arabic annotations to improve accessibility and usability specifically for Arabic-speaking users. By incorporating language-specific annotations, the system enhances the effectiveness of object recognition and provides more relevant feedback to users. This development marks a significant advancement in creating inclusive technology solutions, addressing the unique needs of Arabic-speaking visually impaired individuals and demonstrating progress towards more universally accessible assistive technologies.

Wase et al. (2023) investigate the integration of object detection models with large language models (LLMs) to boost safety and security applications [35]. Their research showcases how combining these advanced technologies can enhance the robustness and accuracy of object detection systems in critical situations. By fusing object detection capabilities with the linguistic understanding of LLMs, the study highlights significant improvements in system performance, making it more reliable and effective in complex and high-stakes environments. This approach represents a promising advancement in developing sophisticated safety and security solutions through the synergy of object detection and language processing technologies.

Chiu et al. (2024) present a study on integrating object detection and natural language processing (NLP) models to develop a personalized attraction recommendation agent [36]. Their research, published in Advanced Engineering Informatics, emphasizes the use of combined object detection and NLP technologies within a smart product service system. This integration aims to enhance user experience by providing personalized recommendations based on visual and textual data. The study illustrates how advancements in these technologies can be leveraged to create more sophisticated and user-centric applications in recommendation systems.

Ma (2024) explores the application of artificial intelligence (AI) technologies in both NLP and image processing, focusing on their ethical and social impacts [37]. Published in the International Journal of Computer Science and Information Technology, this paper reviews various AI technologies and their implications. The author provides a comprehensive overview of how AI-driven advancements in NLP and image processing are reshaping industries and the associated ethical considerations, highlighting the broader societal impact of these technologies.

Manakitsa et al. (2024) offer a review of machine learning and deep learning techniques applied to object detection, semantic segmentation, and human action recognition within machine and robotic vision systems [38]. Their article in Technologies provides an extensive examination of recent developments in these areas, focusing on how these technologies enhance the capabilities of machine vision systems. The review covers both theoretical and practical aspects, offering insights into the current state and future directions of research in these fields.

Mangalika (2024) provides a comprehensive study on the progression from object recognition to content-based image retrieval within computer vision [39]. This review examines key developments and applications in the field, highlighting advancements in technologies that enable more effective image recognition and retrieval. The paper discusses how object recognition has evolved to support content-based image retrieval systems, enhancing their ability to search and retrieve images based on visual content rather than metadata alone. Mangalika's study underscores the significance of these advancements in improving image processing capabilities and offers insights into the ongoing evolution and future directions of computer vision technologies.

Mahesh Babu et al. (2024) investigate the integration of object recognition into conversational chatbots using deep learning and machine learning techniques [40]. Their research, published in Conversational Artificial Intelligence, explores how object recognition can enhance chatbot interactions by enabling them to understand and respond to visual inputs. The study demonstrates the potential of combining NLP and computer vision to create more interactive and responsive conversational agents.

Buettner and Kovashka (2024) examine the role of attribute context in vision-language models for object recognition and detection [41]. Presented at the IEEE/CVF Winter Conference on Applications of Computer Vision, their research delves into how contextual information influences the performance of vision-language models. The paper highlights the significance of incorporating attribute context to improve the accuracy and efficiency of object recognition and detection tasks.

Chen et al. (2024) propose Taskclip, a method for extending large vision-language models to enhance task-oriented object detection [42]. Their preprint on arXiv introduces a novel approach that integrates vision-language models with task-specific objectives to improve object detection performance. The study demonstrates the potential of leveraging large-scale vision-language models for more precise and context-aware detection tasks.

3. CONCLUSION

The convergence of computer vision and artificial intelligence has markedly advanced object identification and recognition systems in recent years. This research provides a comprehensive overview of contemporary methodologies that are driving progress in this field. Starting with foundational concepts in object detection, the study outlines the shift from traditional approaches to more sophisticated deep learning techniques, particularly the use of convolution neural networks (CNNs). These techniques have revolutionized object recognition by automating the extraction of hierarchical features from visual data, thereby enhancing the efficiency and accuracy of object identification.

The paper highlights the critical role of dataset annotation and the development of large-scale annotated datasets, which are essential for training and evaluating robust object identification models. The discussion extends to the application of transfer learning and domain adaptation strategies, which improve model generalization across varied environments and scenarios. It also addresses the persistent challenges faced in real-world applications, such as occlusion, scale variations, and environmental conditions.

An in-depth review of region-based and anchor-based methods reveals significant strides in improving object detection accuracy and efficiency. The study also emphasizes the integration of geographical and semantic information, showcasing the synergy between object detection and recognition. By incorporating AI algorithms and natural language processing (NLP) for contextual awareness, the research proposes pathways for developing more sophisticated systems that can better interpret the relationships and interactions between identified objects.

The discussion of current trends and future directions, including 3D object detection, federated learning, and ethical considerations, underscores the dynamic nature of the field. The potential impact of these advancements on various sectors, such as autonomous vehicles, surveillance, healthcare, and robotics, is substantial. The transformative possibilities of AI in object detection and recognition are evident, with ongoing research promising to further enhance the capabilities and applications of these technologies.

REFERENCES

[1] Torfi, Amirsina, Rouzbeh A. Shirvani, Yaser Keneshloo, Nader Tavaf, and Edward A. Fox. "Natural language processing advancements by deep learning: A survey." arXiv preprint arXiv: 2003.01200 (2020).

[2] Li, Yanfen, Hanxiang Wang, L. Minh Dang, Tan N. Nguyen, Dongil Han, Ahyun Lee, Insung Jang, and Hyeonjoon Moon. "A deep learning-based hybrid framework for object detection and recognition in autonomous driving." IEEE Access 8 (2020): 194228-194239.

[3] Joshi, Rakesh Chandra, Saumya Yadav, Malay Kishore Dutta, and Carlos M. Travieso-Gonzalez. "Efficient multi-object detection and smart navigation using artificial intelligence for visually impaired people." Entropy 22, no. 9 (2020): 941.

[4] Zhang, Fan, Hasan Fleyeh, Xinru Wang, and Minghui Lu. "Construction site accident analysis using text mining and natural language processing techniques." Automation in Construction 99 (2019): 238-248.

[5] Yang, Zhongliang, Yu-Jin Zhang, Sadaqat ur Rehman, and Yongfeng Huang. "Image captioning with object detection and localization." In Image and Graphics: 9th International Conference, ICIG 2017, Shanghai, China, September 13-15, 2017, Revised Selected Papers, Part II 9, pp. 109-118. Springer International Publishing, 2017.

[6] Zhang, Yuting, Luyao Yuan, Yijie Guo, Zhiyuan He, I-An Huang, and Honglak Lee. "Discriminative bimodal networks for visual localization and detection with natural language queries." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 557-566. 2017.

[7] Pathak, Ajeet Ram, Manjusha Pandey, and Siddharth Rautaray. "Application of deep learning for object detection." Procedia computer science 132 (2018): 1706-1717.

[8] Othman, Nashwan Adnan, and Ilhan Aydin. "A new deep learning application based on movidius ncs for embedded object detection and recognition." In 2018 2nd international symposium on multidisciplinary studies and innovative technologies (ISMSIT), pp. 1-5. IEEE, 2018.

[9] Sharma, Shorya, Arpit Sodhi, Nitesh Nagpal, Aman Dureja, and Ajay Dureja. "Practical approaches: machine learning techniques for image processing and text processing." International Journal of Information Systems & Management Science 2, no. 2 (2019).

[10] Yang, Haiqin, Linkai Luo, Lap Pong Chueng, David Ling, and Francis Chin. "Deep learning and its applications to natural language processing." Deep learning: Fundamentals, theory and applications (2019): 89-109.

[11] Zhao, Zhong-Qiu, Peng Zheng, Shou-tao Xu, and Xindong Wu. "Object detection with deep learning: A review." IEEE transactions on neural networks and learning systems 30, no. 11 (2019): 3212-3232.

[12] Gollapudi, Sunila, and Sunila Gollapudi. "Artificial intelligence and computer vision." Learn Computer Vision Using OpenCV: With Deep Learning CNNs and RNNs (2019): 1-29.

[13] Haidar, Siba, Ihab Sbeity, and Marwa Ayyoub. "Text detection using object recognition techniques." COJ Robotics & Artificial Intelligence 1, no. 1 (2019): 1-6.

[14] Kumar, N. Komal, D. Vigneswari, A. Mohan, K. Laxman, and J. Yuvaraj. "Detection and recognition of objects in image caption generator system: A deep learning approach." In 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), pp. 107-109. IEEE, 2019.

[15] Bapu, J. Joshua, D. Jemi Florinabel, Y. Harold Robinson, E. Golden Julie, Raghvendra Kumar, Vo Truong Nhu Ngoc, Le Hoang Son, Tran Manh Tuan, and Cu Nguyen Giap. "Adaptive convolutional neural network using N-gram for spatial object recognition." Earth Science Informatics 12, no. 4 (2019): 525-540.

[16] Torfi, Amirsina, Rouzbeh A. Shirvani, Yaser Keneshloo, Nader Tavaf, and Edward A. Fox. "Natural language processing advancements by deep learning: A survey." arXiv preprint arXiv:2003.01200 (2020).

[17] Li, Yanfen, Hanxiang Wang, L. Minh Dang, Tan N. Nguyen, Dongil Han, Ahyun Lee, Insung Jang, and Hyeonjoon Moon. "A deep learning-based hybrid framework for object detection and recognition in autonomous driving." IEEE Access 8 (2020): 194228-194239.

[18] Vashisht, Manisha, and Brijesh Kumar. "A survey paper on object detection methods in image processing." In 2020 International Conference on Computer Science, Engineering and Applications (ICCSEA), pp. 1-4. IEEE, 2020.

[19] Joshi, Rakesh Chandra, Saumya Yadav, Malay Kishore Dutta, and Carlos M. Travieso-Gonzalez. "Efficient multi-object detection and smart navigation using artificial intelligence for visually impaired people." Entropy 22, no. 9 (2020): 941.

[20] Van Le, Chung, Vikram Puri, and Sandeep Singh Jagdev. "Object Detection System with Image and Speech Recognition." In Handbook of IoT and Blockchain, pp. 85-94. CRC Press, 2020.

[21] Grover, Moksh, Rajat Rathi, Chinkit Manchanda, Kanishk Garg, and Ravinder Beniwal. "AI Optics: Object recognition and caption generation for Blinds using Deep Learning Methodologies." In 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), pp. 354-359. IEEE, 2021.

[22] Srivastava, Sharvani, Amisha Gangwar, Richa Mishra, and Sudhakar Singh. "Sign language recognition system using TensorFlow object detection API." In International conference on advanced network technologies and intelligent computing, pp. 634-646. Cham: Springer International Publishing, 2021.

[23] Kocaleva, Mirjana, and Saso Koceski. "An overview of image recognition and real-time object detection." Balkan Journal of Applied Mathematics and Informatics 4, no. 2 (2021): 41-50.

[24] Liu, Haitao, Yuge Li, and Dongchang Liu. "Object detection and recognition system based on computer vision analysis." In Journal of Physics: Conference Series, vol. 1976, no. 1, p. 012024. IOP Publishing, 2021.

[25] Skepetaris, Konstantinos. "Deep learning-Object detection." Master's thesis, Πανεπιστήμιο Πειραιώς, 2022.

[26] Olszewska, Joanna Isabelle. "Snakes in Trees: An Explainable Artificial Intelligence Approach for Automatic Object Detection and Recognition." In ICAART (3), pp. 996-1002. 2022.

[27] Kaur, Amrita, Yadwinder Singh, Nirvair Neeru, Lakhwinder Kaur, and Ashima Singh. "A survey on deep learning approaches to medical images and a systematic look up into real-time object detection." Archives of Computational Methods in Engineering (2022): 1-41.

[28] Abdelreheem, Ahmed, Ujjwal Upadhyay, Ivan Skorokhodov, Rawan Al Yahya, Jun Chen, and Mohamed Elhoseiny. "3dreftransformer: Fine-grained object identification in real-world scenes using natural language." In Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp. 3941-3950. 2022.

[29] Mondal, Riktim, Samir Malakar, Elisa H. Barney Smith, and Ram Sarkar. "Handwritten English word recognition using a deep learning based object detection architecture." Multimedia Tools and Applications 81, no. 1 (2022): 975-1000.

[30] Rupa, Ch, Gautam Srivastava, Bharath Ganji, Sai Praveen Tatiparthi, Karthik Maddala, Srinivas Koppu, and Jerry Chun-Wei Lin. "Medicine drug name detection based object recognition using augmented reality." Frontiers in Public Health 10 (2022): 881701.

[31] Zou, Zhengxia, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. "Object detection in 20 years: A survey." Proceedings of the IEEE 111, no. 3 (2023): 257-276.

[32] Kaur, Ravpreet, and Sarbjeet Singh. "A comprehensive review of object detection with deep learning." Digital Signal Processing 132 (2023): 103812.

[33] Sharada, K., Wajdi Alghamdi, K. Karthika, Ahmed Hussein Alawadi, Gulomova Nozima, and V. Vijayan. "Deep Learning Techniques for Image Recognition and Object Detection." In E3S Web of Conferences, vol. 399, p. 04032. EDP Sciences, 2023.

[34] Alzahrani, Nada, and Heyam H. Al-Baity. "Object recognition system for the visually impaired: a deep learning approach using Arabic annotation." Electronics 12, no. 3 (2023): 541.

[35] Wase, Zeba Mohsin, Vijay K. Madisetti, and Arshdeep Bahga. "Object Detection Meets LLMs: Model Fusion for Safety and Security." Journal of Software Engineering and Applications 16, no. 12 (2023): 672-684.

[36] Chiu, Ming-Chuan, Cheng-Zhou Tsai, and Yu-Chen Huang. "Integrating object detection and natural language processing models to build a personalized attraction recommendation agent in a smart product service system." Advanced Engineering Informatics 61 (2024): 102484.

[37] Ma, Wenxiang. "The Application of Artificial Intelligence Technology in Natural Language Processing and Image Processing and Its Ethical and Social Impact." International Journal of Computer Science and Information Technology 2, no. 2 (2024): 124-128.

[38] Manakitsa, Nikoleta, George S. Maraslidis, Lazaros Moysis, and George F. Fragulis. "A review of machine learning and deep learning for object detection, semantic segmentation, and human action recognition in machine and robotic vision." Technologies 12, no. 2 (2024): 15.

[39] Mangalika, Udula. "Object Recognition to Content Based Image Retrieval: A Study of the Developments and Applications of Computer Vision." (2024).

[40] Mahesh Babu, A., Malik Jawarneh, José Luis Arias-Gonzáles, Meenakshi, Kishori Kasat, and K. P. Yuvaraj. "Conversational Chatbot With Object Recognition Using Deep Learning and Machine Learning." Conversational Artificial Intelligence (2024): 335-352.

[41] Buettner, Kyle, and Adriana Kovashka. "Investigating the Role of Attribute Context in Vision-Language Models for Object Recognition and Detection." In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 5474-5484. 2024.

[42] Chen, Hanning, Wenjun Huang, Yang Ni, Sanggeon Yun, Fei Wen, Hugo Latapie, and Mohsen Imani. "Taskclip: Extend large vision-language model for task oriented object detection." arXiv preprint arXiv:2403.08108 (2024).