Nanotechnology Perceptions
ISSN 1660-6795
www.nano-ntp.com

Comprehensive Review On Recent Progress, Challenges, And Future Prospects In Deep Learning For Robust Deformable Object Detection And Segmentation

Sonali Pulate[1] , Dr. Rekha labade[2] , Dr. S.V.Chaudhari[3] , Dr. B.S.Agarkar[4]

[1]PhD Scholar Department Of Electronics and telecommunication Engineering (Affiliated to SPPU Pune),Kopargaon,Maharashtra sonalipulate@gmail.com
[2]Research Guide Department of Electronics and telecommunication Amruthvahini College of Engineering(Affiliated to SPPU Pune),Sangamner, Maharashtra
[3]Research Coordinator Depth.Of Electronics and telecommunication Engineering (Affiliated to SPPU Pune),Kopargaon,Maharashtra chaudharisachinetc@sanjivani.org.in
[4]Head of Department Department Of Electronics and telecommunication Engineering (Affiliated to SPPU Pune),Kopargaon,Maharashtra hodetccoe@sanjivani.org.in

In this review article, we provide an extensive overview of the most recent advancements, prevailing challenges, and promising future prospects in the domain of deep learning techniques for robust deformable object detection and segmentation in images. We explore the state-of-the-art methods, highlighting their capabilities and limitations, and discuss the various challenges that researchers encounter in this domain, such as handling occlusions, variations in scale, and deformations. Furthermore, we explore the potential applications of these technologies in areas like medical imaging, autonomous vehicles, and robotics. Our analysis incorporates not only the achievements to date but also the directions in which this field is headed, explaining emerging trends and future research possibilities. By offering a thorough summary of the current situation, This review offers a holistic perspective on the deformable object detection and segmentation, serving as a valuable resource for researchers, professionals, and business experts to understand the complexities of deep learning-based deformable object detection and segmentation, guiding them towards a better understanding of the field's current state and its promising future developments.

Keywords: Convolutional Neural Networks, Deep Learning, Recurrent Neural Networks, deformable object, Occlusions, Cluttered Scenes.

1. INTRODUCTION

The field of computer vision has witnessed a remarkable transformation in recent years, owing much of its success to the unprecedented advancements in deep learning techniques. In particular, the application of deep learning models for deformable object detection and segmentation in images has opened new horizons for a wide range of applications, from medical imaging to autonomous robotics. This review article embarks on an exploration of this rapidly evolving landscape, offering a comprehensive analysis of the current state of the art, the persisting challenges, and the exciting prospects that lie on the horizon. Multiple identification of object in photos or videos is a key component of numerous actual-world issues. In particular, when the objects to be detected differ in terms of size, colour, shape, and texture, creating practical detectors for these kinds of problems can be accomplished through the use of contemporary deep learning models. All the same, this task becomes more difficult when the target objects are small (represented by fewer pixels, similar in size, shape, colour, and texture), and handled similarly [1].

The accurate detection and segmentation of deformable objects within images is a fundamental task in computer vision with profound implications for both academia and industry. Deformable objects, such as human bodies, animals, and flexible materials, exhibit complex, non-rigid shapes and often vary in appearance due to changes in pose, illumination, and occlusions. Traditional computer vision techniques, relying on handcrafted features and rule-based algorithms, have struggled to robustly address these challenges. Enter deep learning, a subfield of machine learning that leverages the power of artificial neural networks to automatically learn hierarchical representations of data. Deep learning, with its ability to capture intricate patterns and variations in data, has proven to be a game-changer in deformable object analysis. In the domain of machine/computer vision, object detection is a fundamentally important and extremely encouraging field from a scientific and mechanical standpoint. The object-detection task categorizes various kinds of objects in an image. Although significant progress has been made in the structure to address object detection issues, hidden elements like occlusion, arbitrary viewpoints, and cluttered environments still persist in the uncontrolled environment [15].

In the context of deformable object detection, convolutional neural networks (CNNs) have risen to prominence as a dominant approach. CNNs can extract local features from images and are capable of learning representations that adapt to changes in object appearance, making them exceptionally well-suited for this task. Meanwhile, recurrent neural networks (RNNs) have played a pivotal role in deformable object segmentation, as they can model the temporal dependencies and spatial relationships within sequences of images. These deep learning architectures have demonstrated impressive results in a wide range of applications, including human pose estimation, medical image analysis, and industrial automation.

While these advancements are certainly promising, they are not without their challenges. The review will delve into the persisting issues that continue to challenge researchers and practitioners in the field. Dealing with occlusions, variations in object appearance, and the scarcity of annotated data remain significant hurdles. Moreover, the interpretation of deep learning models, often referred to as the "black box" problem, has raised concerns about model transparency and accountability in critical applications.

In the latter part of this review, we will explore the future prospects for deformable object detection and segmentation. These prospects include the integration of multi-modal data sources, which can enrich the information available to deep learning models, leading to more robust and accurate results. Furthermore, the development of architectures that are explicitly designed to address deformable objects' unique characteristics and challenges is expected to be a focal point for future research. Finally, we will discuss the incorporation of explainable AI techniques, allowing researchers and practitioners to gain deeper insights into model decisions and to enhance trust and interpretability in deformable object analysis systems.

In summary, the review aims to offer an in-depth analysis of the state-of-the-art deep learning techniques in the context of deformable object detection and segmentation, highlighting both the remarkable progress made and the significant challenges that persist. It will also shed light on the promising prospects and research directions that hold the potential to further revolutionize this exciting field.

2. RELATED WORK

Francisco Pérez et al. [1] proposed a two-level methodology called Object Detection with Binary Classifiers (ODeBiC) based on deep learning for detecting small objects that can be confused with a handgun or a knife when manipulated with hand. The first level selects candidate regions from the input frame, while the second level applies a binarization technique based on a CNN classifier with One-Versus-All or One-Versus-One. The proposed methodology is evaluated using the Sohasweapon dataset and compared with other classification approaches. The experimental analysis shows that the ODeBiC methodology reduces the number of false positives and improves precision compared to the baseline detection model. Yiqian Wang et al. [2] suggested a deep learning framework for multimodal retinal image registration, which includes a segmentation network, feature detection and description network, and an outlier rejection network. The proposed framework demonstrates a significant improvement in robustness and accuracy compared to other coarse alignment methods, as reflected by a higher success rate and Dice coefficient. Lujun Zhai et al. [3] provides a comprehensive review of real-world image restoration algorithms, benchmark datasets, image quality assessment methods, and four major categories of deep learning-based image restoration methods. It discusses the latest developments and advances in each category of network architecture, comparing representative state-of-the-art image restoration methods visually and numerically.

Athina Ilioudi et al. [4] provides a systematic analysis of existing methods in the field of computer vision tasks, with a focus on object localization and identification in video frames. It presents an overview of DL-based video computer vision methods and highlights the challenges associated with these approaches. The study also analyzes the advancements of deep learning techniques for computer vision tasks in videos and addresses their current weaknesses. Lubna Aziz et al. [5] conduct a comprehensive survey of recent advances in visual object detection with deep learning, covering about 300 publications. It discusses various region proposal-based object detection methods such as R-CNN, SPPnet, Fast R-CNN, Faster R-CNN, Mask RCN, RFCN, FPN, as well as classification/regression-based methods like YOLO, SSD, DSSD, RetinaNet, RefineDet, CornerNet, EfficientDet. The paper also explores the application of deep learning architectures to five major fields: Object Detection in Surveillance, Military, Transportation, Medical, and Daily Life. It covers factors affecting detection performance, such as a wide range of object categories and limited storage capacity and computational power. Jaswinder Singh and B.K. Sharma [6] suggested R-CNN approaches that have been being presented for detecting objects in the strange environment. This method reduces the complexity needed to identify an object from a complex circumstance. Real-time frame rates will be used in the suggested detection of objects system. The quality of region schemes is enhanced by the learned RPN method. Overall, object detections are achieving the required level of precision.

Aysxegu Ucxar et al. [7] suggested a hybrid Local Multiple system (LM-CNN-SVM) based on Convolutional Neural Networks (CNNs) and Support Vector Machines (SVMs) for object recognition and detection in autonomous driving applications. The proposed system divides the whole image into local regions and employs multiple CNNs to learn local object features. Discriminative features are selected using Principal Component Analysis (PCA). Multiple SVMs are used to increase the generalization ability of the classifier system. SVM outputs are fused to improve the results. The LM-CNN-SVM system significantly surpasses other methods in terms of higher recognition performance and lower miss rate. It shows promise for improving object recognition and detection in real-time applications. Connor Shorten and Taghi M. Khoshgoftaar [8] discusses the advancements in deep neural networks and their successful application in computer vision tasks such as image classification, object detection, and image segmentation. The researchers highlight the use of data augmentation in the AlexNet CNN architecture, which increased the dataset size and reduced overfitting, resulting in improved model performance. Overall, author provides a comprehensive survey of image data augmentation techniques for deep learning, highlighting their effectiveness in improving model performance and reducing overfitting. Li Liu et al. [9] conduct comprehensive survey of the recent achievements in the field of generic object detection using deep learning techniques. The survey aims to offer a taxonomy and high-level perspective on the basis of popular datasets, evaluation metrics, context modeling, and detection proposal methods, helping researchers understand the current research and identify open challenges for future research. Sreyasee Das Bhattacharjeea and Anurag Mittalb [10] proposed a new algorithm for object detection using a single reasonably good sketch as a reference to build a model for the object. The method hierarchically segments the sketch into parts using an automatic algorithm and estimates a different affine transformation for each part while matching. A Hough-style voting scheme collects evidence for the object from the leaves to the root in the part decomposition tree for robust detection. Missing edge segments, clutter and generic object deformations are handled by flexibly following the contour paths in the edge image that resemble the model contours. Efficient data structures and a two-stage matching approach assist in yielding an efficient and robust system. The full multi-stage detection system includes the use of integral image concept for smoothening the vote response, screening matches based on the number of votes received, and stitching together the individual CSPs for further confirmation of object continuity. Zhong-Qiu Zhao et al. [11] provides a review on deep learning-based object detection frameworks, starting with a brief introduction to the history of deep learning and Convolutional Neural Networks (CNN). It discusses traditional object detection methods that rely on handcrafted features and shallow trainable architectures, which can easily stagnate in performance. The paper explores the use of deep learning models that learn semantic, high-level, and deeper features to address the limitations of traditional architectures. Gyumin Shim et al. [12] demonstrated a novel reference-based super-resolution (RefSR) method called Similarity Search and Extraction Network (SSEN) for extracting aligned relevant features from a reference image to enhance single image super-resolution (SISR) performance. The proposed algorithm is end-to-end trainable without additional supervision or heavy computation, and it predicts the best match with a single network forward operation. Chao Zuo et al. [13] provide an overview of the current status and latest progress of deep-learning technologies in the field of optical metrology. It discusses traditional image-processing algorithms in optical metrology and the basic concepts of deep learning. The review highlights the applications of deep learning in various optical metrology tasks, including fringe denoising, phase retrieval, phase unwrapping, subset correlation, and error compensation. Xiaogang Wang [14] explored a historical overview of deep learning and its applications in object recognition, detection, and segmentation, focusing on computer vision challenges and their applications to images and videos. Sushma Jaiswal and Tarun Jaiswal [15] demonstrated comprehensive investigation and evaluation of various object detection approaches, covering both first and second stage detectors **[**1**]**. It introduces the Mask R-CNN method, which is known for its practicality, efficiency, and ability to detect objects and produce segmentation masks for each instance. Mask R-CNN has shown superior outcomes in COCO challenges for instance segmentation, bounding box, and key point detection. Naoual El-Djouher et al. [16] presented a novel approach based on deep neural networks and saliency features for fusing RGB visible images with thermal images for robust and accurate object detection under challenging conditions. The architecture consists of two branches, each containing a deep network based on the YOLO v2 object detector. One network learns from thermal images and the other from color images, as they have different characteristics. The YOLO v2 architecture is used for object detection, which performs classification probabilities and localization of objects simultaneously. It consists of convolutional layers, fully connected layers, and predefined anchors for generating bounding boxes. Thorsten Hoeser et al. [17] provides a comprehensive review of 429 studies on image segmentation and object detection with convolutional neural networks (CNNs) in Earth observation (EO) applications. The authors extensively examine the spatial distribution of study sites, employed sensors, used datasets, and CNN architectures in EO applications. The impact of datasets like ImageNet, PASCAL VOC, MSCOCO, and Cityscape on the evolution of CNNs in EO is discussed.

The paper highlights the sparse applications of CNNs with a multi-temporal perspective in EO and suggests combining CNNs with other model types, such as LSTMs, to increase applications and analyze object dynamics.

Ravpreet Kaur and Sarbjeet Singh [18] conducted a detailed review of object detection and its different aspects, including object detection frameworks, backbone convolutional neural networks, common datasets, and evaluation metrics. It also discusses the challenges in designing deep neural networks for object detection and compares the performance of object detection models on PASCAL VOC and MS COCO datasets. The researcher emphasizes the rapid progress of object detection following the introduction of deep learning and provides a thorough analysis of state-of-the-art object detection models, backbone architectures, and their performance on standard datasets. It also discusses the challenges, applications, and future research directions in object detection. Sunil and Gagandeep [19] provide a review of machine learning approaches for object detection, including template-based, part-based, region-based, and contour-based methods. It also discusses the wide range of applications for object detection, such as image recovery, safety, investigation, machine system assessment, and computerized vehicle structure. P. Devaki et al. [20] suggested existing network architectures such as VGG and ResNet are larger in size, making them unsuitable for computationally less powered systems. The paper explores the use of MobileNet, a network architecture developed by Google for mobile and less power embedded vision applications. The researcher also discusses the use of region conventional neural networks (CNN) for object detection, which provides more accurate results than HOG-based implementations. However, the sliding window method used in RCNN can be computationally expensive and complex.

Ajeet Ram Pathak et al. [21] discusses the role of deep learning techniques based on convolutional neural networks (CNN) for object detection. It explores the use of part-based methods, which represent objects as collections of local parts and spatial structures, combined with CNN for object detection. Xiongwei Wu et al. [22] provides a comprehensive survey of recent advances in visual object detection with deep learning, covering three major parts: detection components, learning strategies, and applications & benchmarks. The authors review a large body of recent related work in literature and systematically analyze the existing object detection frameworks. They discuss various factors affecting detection performance, such as detector architectures, feature learning, proposal generation, and sampling strategies. Gong Cheng, Junwei Han [23] conduct a review of the recent progress in the field of object detection in optical remote sensing images, covering about 270 publications. It discusses different object detection methods, including template matching-based methods, knowledge-based methods, object-based image analysis (OBIA)-based methods, and machine learning-based methods. The author also discusses five publicly available datasets and three standard evaluation metrics for object detection. It highlights the challenges in current studies and proposes two promising research directions: deep learning-based feature representation and weakly supervised learning-based geospatial object detection. Syed Sahil Abbas et al. [24] provides a comprehensive review of deep learning-based object detectors and lightweight classification architectures, focusing on recent developments in the field. The authors compare the performance of several object detection models, including SSD, Faster RCNN, and R-FCN, using different feature extractors such as Inception V2, Resnet-101, and Mobilenet-V1. It highlights the changes in network designs and their implications on the performance of object detectors, providing insights into the working of prominent networks.

Wei Hua et al. [25] provide a comprehensive presentation of the object detection datasets used in aerial remote sensing images and the evaluation metrics employed. It analyzes and compares the experimental results of the latest network models based on deep learning methods for small target detection in aerial images. It reviews the optimization strategies used for small target detection in aerial images, including deep learning techniques. It focuses on the challenges of small object detection in aerial images and the application of deep learning methods to overcome these challenges. Li Yi et al. [26] implemented deep learning for object detection in a factory using images of machines to train four models: SSDInception-v2 (Inception), SSDMobilenetv2 (MobileNet), SSDResNet50 (ResNet), and Faster R-CNN Inception-ResNet50v2 (R-CNN). Performance evaluation was done based on loss curves, mean average precision curves, and performance metrics. Mei-Ling Huang and Yi-Shun Wu [27] focuses on the use of image processing and computer vision techniques for the classification and identification of fortunella margarita, a fruit crop. The dataset provided in the paper can be used for the development and training of fruit detection systems, with applications in segmentation and classification. The dataset consists of a total of 6611 annotation images in XML format, manually labeled with the growth stage and location of the fruit. The images were captured using an iPhone 11 Pro in Jiaoxi, Yilan County, Taiwan, under both clear and cloudy weather conditions. Abhishek Gupta et al. [28] conduct a comprehensive survey of deep learning applications for object detection and scene perception in autonomous vehicles. It examines the theory underlying self-driving vehicles from a deep learning perspective and evaluates current implementations. The survey discusses the role of deep learning in interpreting complex vision, enhancing perception, and enabling kinematic maneuvers in self-driving cars. It explores the use of deep learning techniques, such as convolutional neural networks (CNNs), for object detection and scene perception in self-driving cars.

Zhong-Qiu Zhao et al. [29] conduct a review on deep learning-based object detection frameworks, starting with a brief introduction to the history of deep learning and Convolutional Neural Networks (CNNs). It discusses traditional object detection methods that rely on handcrafted features and shallow trainable architectures, which easily stagnate in performance. The author highlights the advantages of deep learning models in learning semantic, high-level, and deeper features. Manhuai Lu and Liqin Chen [30] mentions conventional object detection methods such as Viola Jones detectors, histogram of oriented gradients (HOD) detector, and deformable part-based model (DPM). It also highlights the advancements in deep learning that have led to the development of numerous object detection methods. These deep learning-based methods can be categorized into R-CNN-based, SSD-based, and YOLO-based object detection methods. Licheng Jiao et al. [31] provide a comprehensive overview of a variety of object detection methods, including one-stage and two-stage detectors, as well as traditional and new applications. It analyzes the methods of existing typical detection models and describes benchmark datasets. The survey discusses the architecture of exploiting these object detection methods to build an effective and efficient system. It points out a set of development trends to better follow the state-of-the-art algorithms and further research. Jwalin Bhatt et al. [32] provides a comprehensive overview of traditional and deep learning-based approaches for graphical page object detection in document images. This paper discusses the importance of deep neural networks in improving the detection of graphical page objects such as tables, formulas, and figures in scanned document images. It also addresses the challenges and future directions in the field of graphical page object detection, emphasizing the need for continued research and development in this area. M. Indirani and S. Shanka [33] graphical model-based techniques have been highlighted in the literature as efficient for modeling and yielding promising results in saliency and object segmentation studies. The research paper emphasizes the use of graphical models for efficient modeling and achieving promising results in the field of saliency and object segmentation. Ayoub Benchabana et al. [34]. proposed approach in this paper addresses these challenges by utilizing adaptive super pixel segmentation and deep learning techniques to enhance building detection accuracy in high-resolution remote sensing images The proposed approach has shown superior accuracy compared to previous works, with an average F1 score of 98.83 , demonstrating its potential for fast and accurate urban monitoring and city planning in urban areas

Ch. Radhika et al. [35] proposed model in the paper utilizes Residual Neural Network101 (ResNet101) and ZFNet for feature extraction and providing additional information about objects in remote sensing images. Single scale and multiscale object detection are implemented using You Only Look Once (YOLOV5) and Faster Region based Convolutional Neural Network (Faster RCNN). Zheng Li et al. [36] provides a comprehensive review of deep learning-based object detection techniques for remote sensing images, focusing on recent advancements and strategies. The review also discusses the challenges faced by current Real-time small object detection (RSOD) algorithms based on deep learning techniques and suggests areas for further research and development, such as network structures and direction prediction strategies. Imran Ahmed et al. [37] have explored various methods for object detection and segmentation in remote sensing images, including conventional hand-crafted features, machine learning, and deep learning techniques. The researchers introduce an IoT-enabled surveillance system for multiple object detection, emphasizing real-time surveillance applications. Simegnew Yihunie Alaba [38] reviews deep learning-based image 3D object detection models for autonomous driving, focusing on feature learning from images rather than hand-crafted feature extractors, which improves performance and training processes. Challenges and future directions in 3D object detection are discussed, highlighting the importance of accurate and robust perception systems for autonomous driving and robots

Summary of Review:

Recent advancements in deep learning have substantially improved robust deformable object detection and segmentation. Key progress includes the use of advanced convolutional neural networks (CNNs) and transformer-based architectures, enhancing accuracy and robustness in varying object shapes and appearances. Techniques such as data augmentation and transfer learning have further boosted performance across diverse datasets. Despite these strides, significant challenges persist, including the demand for extensive annotated datasets, high computational costs, and real-time processing limitations. Future directions focus on harnessing unsupervised and semi-supervised learning, integrating multi-modal data, and improving model interpretability. There is also an emphasis on developing lightweight models suitable for edge devices and enhancing generalization across applications like medical imaging, robotics, and autonomous vehicles.

Comparison of Various Methods used for Deformable Object Detection and Segmentation

In this paper we rigorously explored the different methods for Deformable Object Detection and Segmentation and performed the comparison between them.

The table I depict the comparison of various methods used for Deformable Object Detection and Segmentation for various metrics. Figure 1 displays comparison of various methods used for Deformable Object Detection and Segmentation using various Accuracy, Recall Precision, mAP, GLOPS and FPS. Figure 2 displays comparison of various methods used for Deformable Object Detection and Segmentation.

Table I: Comparison of various methods used for Deformable Object Detection and Segmentation

Method	**Dataset Used**	**Accuracy (%)**	**Recall (%)**	**Precision (%)**	**mAP (%)**	**GFLOPs**	**FPS**
Faster R-CNN [39]	COCO	74.9	73.2	75.8	74.9	200	5-7
Mask R-CNN [40]	COCO	78.5	76.7	80.3	78.5	220	5
YOLOv4 [41]	COCO	65.7	70.0	67.5	65.7	130	50-60
SSD [42]	COCO	74.3	72.1	76.3	74.3	100	46
U-Net [43]	ISBI Challenge	85.5	84.0	86.8	N/A	100	--
TransUNet [45]	Synapse Multi-organ	90.6	89.2	91.1	N/A	250	--
Swin Transformer [46]	ADE20K	88.1	86.5	89.7	N/A	245	--

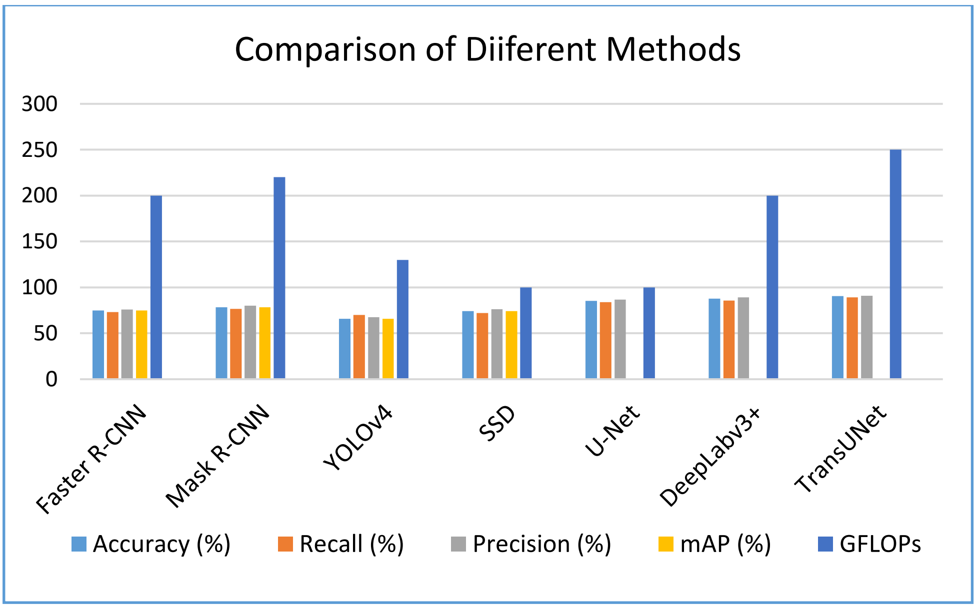

Fig 1: Comparison of different methods used for Deformable Object Detection and Segmentation with metrics Accuracy, Recall Precision, mAP, GLOPS and FPS.

3. Challenges

The following are the challenges in Deep Learning for Robust Deformable Object

1. Handling the wide range of possible deformations in object shapes and sizes requires complex and flexible models.
2. Accurately detecting and segmenting objects that are partially occluded or only partially visible is a significant challenge.
3. Achieving high accuracy in real-time applications necessitates efficient and optimized algorithms.
4. The scarcity of labeled training data for deformable objects limits the performance of deep learning models.
5. Ensuring robustness to varying lighting conditions and environmental changes is essential for reliable detection and segmentation.

Conclusion

Deep learning for robust deformable object detection and segmentation has seen substantial progress, particularly with the development of advanced neural network architectures and enhanced data augmentation techniques. Despite these advancements, challenges such as the need for large annotated datasets, managing occlusions, achieving real-time performance, and ensuring robustness to noise and environmental changes remain significant. Future research must focus on creating more efficient algorithms, leveraging unsupervised and semi-supervised learning, and developing models that can generalize across diverse conditions. Enhancing model interpretability and explainability is crucial for their deployment in critical applications. Interdisciplinary approaches, combining insights from computer vision, robotics, and human-computer interaction, could provide innovative solutions to existing challenges. Additionally, advancements in hardware accelerators and edge computing will further enable the practical application of these models. Collaboration between academia and industry will be essential to translate research breakthroughs into real-world applications. Overall, while the field faces complex challenges, the ongoing innovations and collaborative efforts promise a bright future for robust deformable object detection and segmentation.

The future scope for this review lies in deep learning for robust deformable object detection and segmentation will focus on creating adaptive models capable of handling diverse conditions with limited labeled data, driven by innovations in unsupervised learning and hardware acceleration. These developments will make the technology more practical and widely applicable across various fields.

REFERENCES

[1] F. Pérez-Hernández, S. Tabik, A. Lamas, R. Olmos, H. Fujita, and F. Herrera, "Object Detection Binary Classifiers methodology based on deep learning to identify small objects handled similarly: Application in video surveillance," Knowledge-Based Systems, vol. 194, p. 105590, Apr. 2020, doi: 10.1016/j.knosys.2020.105590.

[2] Y. Wang et al., "A Segmentation Based Robust Deep Learning Framework for Multimodal Retinal Image Registration," ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), May 2020, doi: 10.1109/icassp40776.2020.9054077..

[3] L. Zhai, Y. Wang, S. Cui, and Y. Zhou, “A Comprehensive Review of Deep Learning-Based Real-World Image Restoration,” IEEE Access, vol. 11, pp. 21049–21067, 2023, doi: 10.1109/access.2023.3250616.

[4] A. Ilioudi, A. Dabiri, B. J. Wolf, and B. De Schutter, “Deep Learning for Object Detection and Segmentation in Videos: Toward an Integration With Domain Knowledge,” IEEE Access, vol. 10, pp. 34562–34576, 2022, doi: 10.1109/access.2022.3162827..

[5] L. Aziz, Md. S. B. Haji Salam, U. U. Sheikh, and S. Ayub, “Exploring Deep Learning-Based Architecture, Strategies, Applications and Current Trends in Generic Object Detection: A Comprehensive Review,” IEEE Access, vol. 8, pp. 170461–170495, 2020, doi: 10.1109/access.2020.3021508.

[6] Jaswinder Singh #1, Dr. B.K. Sharma, “ R-CNN Based Object Detection and Classification Methods for Complex Sceneries,” International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 20 (2018) pp. 14505-14509

[7] A. Uçar, Y. Demir, and C. Güzeliş, “Object recognition and detection with deep learning for autonomous driving applications,” SIMULATION, vol. 93, no. 9, pp. 759–769, Jun. 2017, doi: 10.1177/0037549717709932.

[8] C. Shorten and T. M. Khoshgoftaar, “A survey on Image Data Augmentation for Deep Learning,” Journal of Big Data, vol. 6, no. 1, Jul. 2019, doi: 10.1186/s40537-019-0197-0.

[9] L. Liu et al., “Deep Learning for Generic Object Detection: A Survey,” International Journal of Computer Vision, vol. 128, no. 2, pp. 261–318, Oct. 2019, doi: 10.1007/s11263-019-01247-4.

[10] S. Das Bhattacharjee and A. Mittal, “Part-based deformable object detection with a single sketch,” Computer Vision and Image Understanding, vol. 139, pp. 73–87, Oct. 2015, doi: 10.1016/j.cviu.2015.06.005.

[11] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu, “Object Detection with Deep Learning: A Review,” IEEE Transactions on Neural Networks and Learning Systems. Apr 2019, pp. 1-12.

[12] G. Shim, J. Park, and I. S. Kweon, “Robust Reference-Based Super-Resolution With Similarity-Aware Deformable Convolution,” 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Jun. 2020, doi: 10.1109/cvpr42600.2020.00845.

[13] C. Zuo et al., “Deep learning in optical metrology: a review,” Light: Science & Applications, vol. 11, no. 1, Feb. 2022, doi: 10.1038/s41377-022-00714-x.

[14] X. Wang, “Deep Learning in Object Recognition, Detection, and Segmentation,” 2016, doi: 10.1561/9781680831177.

[15] Sushma Jaiswal and Tarun Jaiswal, “Deep Learning Approaches for Object Detection,” Artificial Intelligence Evolution, pp. 122–144, Nov. 2020, doi: 10.37256/aie.122020564.

[16] N. E.-D. Mebtouche and N. Baha, “Robust Object Detection Based on Deep Neural Network and Saliency Features from Visible and Thermal Images,” Advances in Intelligent Systems and Computing, pp. 529–540, 2022, doi: 10.1007/978-3-030-90639-9_43.

[17] T. Hoeser, F. Bachofer, and C. Kuenzer, “Object Detection and Image Segmentation with Deep Learning on Earth Observation Data: A Review—Part II: Applications,” Remote Sensing, vol. 12, no. 18, p. 3053, Sep. 2020, doi: 10.3390/rs12183053.

[18] R. Kaur and S. Singh, “A comprehensive review of object detection with deep learning,” Digital Signal Processing, vol. 132, p. 103812, Jan. 2023, doi: 10.1016/j.dsp.2022.103812.

[19] Sunil and Gagandeep, “Study of Object Detection Methods and Applications on Digital Images”, International Journal of Scientific Development and Research, Volume 4, Issue 5, May 2019, pp. 491-497.

[20] P. Devaki, S. Shivavarsha, G. Bala Kowsalya, M. Manjupavithraa, and E.A. Vima, “Real-Time Object Detection using Deep Learning and Open CV,” Special Issue, vol. 8, no. 12S, pp. 411–414, Dec. 2019, doi: 10.35940/ijitee.l1103.10812s19.

[21] Ajeet Ram Pathaka, , Manjusha Pandeya, and Siddharth Rautaraya, “ Application of Deep Learning for Object Detection”, Procedia Computer Science 132 (2018), pp. 1706–1717

[22] X. Wu, D. Sahoo, and S. C. H. Hoi, “Recent advances in deep learning for object detection,” Neurocomputing, vol. 396, pp. 39–64, Jul. 2020, doi: 10.1016/j.neucom.2020.01.085.

[23] G. Cheng and J. Han, “A survey on object detection in optical remote sensing images,” ISPRS Journal of Photogrammetry and Remote Sensing, vol. 117, pp. 11–28, Jul. 2016, doi: 10.1016/j.isprsjprs.2016.03.014.

[24] S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, “A survey of modern deep learning based object detection models,” Digital Signal Processing, vol. 126, p. 103514, Jun. 2022, doi: 10.1016/j.dsp.2022.103514.

[25] W. Hua and Q. Chen, “A Survey of Small Object Detection Based on Deep Learning in Aerial Images,” Jun. 2023, doi: 10.21203/rs.3.rs-3074407/v1.

[26] L. Yi, C. Siedler, Y. Kinkel, M. Glatt, P. Kölsch, and J. C. Aurich, “Object detection in factory based on deep learning approach,” Procedia CIRP, vol. 104, pp. 1029–1034, 2021, doi: 10.1016/j.procir.2021.11.173.

[27] M.-L. Huang and Y.-S. Wu, “A dataset of fortunella margarita images for object detection of deep learning based methods,” Data in Brief, vol. 38, p. 107293, Oct. 2021, doi: 10.1016/j.dib.2021.107293.

[28] A. Gupta, A. Anpalagan, L. Guan, and A. S. Khwaja, “Deep learning for object detection and scene perception in self-driving cars: Survey, challenges, and open issues,” Array, vol. 10, p. 100057, Jul. 2021, doi: 10.1016/j.array.2021.100057

[29] Zhong-Qiu Zhao, Peng Zheng, Shou-tao Xu, and Xindong Wu, “Object Detection with Deep Learning: A Review”, Ieee Transactions on Neural Networks and Learning Systems, Apr 2019.

[30] “Retracted: Efficient Object Detection Algorithm in Kitchen Appliance Scene Images Based on Deep Learning,” Mathematical Problems in Engineering, vol. 2023, pp. 1–1, Jul. 2023, doi: 10.1155/2023/9891509.

[31] Licheng Jiao et al., “A Survey of Deep Learning-Based Object Detection”, IEEE Access, VOLUME 7, 2019, doi: 10.1109/ACCESS.2019.2939201

[32] J. Bhatt, K. A. Hashmi, M. Z. Afzal, and D. Stricker, “A Survey of Graphical Page Object Detection with Deep Neural Networks,” Applied Sciences, vol. 11, no. 12, p. 5344, Jun. 2021, doi: 10.3390/app11125344.

[33] M. Indirani, S. Shankar,“Research on Salient Object Detection using Deep Learning and Segmentation Methods,” Special Issue, vol. 8, no. 2S11, pp. 280–287, Nov. 2019, doi: 10.35940/ijrte.b1046.0982s1119.

[34] A. Benchabana, M.-K. Kholladi, R. Bensaci, and B. Khaldi, “Building Detection in High-Resolution Remote Sensing Images by Enhancing Superpixel Segmentation and Classification Using Deep Learning Approaches,” Buildings, vol. 13, no. 7, p. 1649, Jun. 2023, doi: 10.3390/buildings13071649.

[35] Ch. Radhika, D. Nikhila, D. Shivani, B. SiriPriya, and B. Reddy, “MultiScale Object Detection in Remote Sensing Images using Deep Learning,” International Journal for Research in Applied Science and Engineering Technology, vol. 11, no. 7, pp. 708–712, Jul. 2023, doi: 10.22214/ijraset.2023.54643.

[36] Z. Li et al., “Deep Learning-Based Object Detection Techniques for Remote Sensing Images: A Survey,” Remote Sensing, vol. 14, no. 10, p. 2385, May 2022, doi: 10.3390/rs14102385.

[37] I. Ahmed, M. Ahmad, A. Chehri, M. M. Hassan, and G. Jeon, “IoT Enabled Deep Learning Based Framework for Multiple Object Detection in Remote Sensing Images,” Remote Sensing, vol. 14, no. 16, p. 4107, Aug. 2022, doi: 10.3390/rs14164107.
[38] S. Y. Alaba and J. E. Ball, “Deep Learning-Based Image 3-D Object Detection for Autonomous Driving: Review,” IEEE Sensors Journal, vol. 23, no. 4, pp. 3378–3394, Feb. 2023, doi: 10.1109/jsen.2023.3235830.
[39] S. Ren, K. He, R. Girshick, and J. Sun, “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/tpami.2016.2577031.
[40] K. He, G. Gkioxari, P. Dollar, and R. Girshick, “Mask R-CNN,” 2017 IEEE International Conference on Computer Vision (ICCV), Oct. 2017, doi: 10.1109/iccv.2017.322.
[41] Alexey Bochkovskiy et al., “YOLOv4: Optimal Speed and Accuracy of Object Detection,” 23 Apr 2020
[42] W. Liu et al., “SSD: Single Shot MultiBox Detector,” Lecture Notes in Computer Science, pp. 21–37, 2016, doi: 10.1007/978-3-319-46448-0_2.
[43] O. Ronneberger, P. Fischer, and T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation,” Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, pp. 234–241, 2015, doi: 10.1007/978-3-319-24574-4_28.
[44] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, “Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation,” Lecture Notes in Computer Science, pp. 833–851, 2018, doi: 10.1007/978-3-030-01234-2_49.
[45] A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, and D. Zhang, “DS-TransUNet: Dual Swin Transformer U-Net for Medical Image Segmentation,” IEEE Transactions on Instrumentation and Measurement, vol. 71, pp. 1–15, 2022, doi: 10.1109/tim.2022.3178991.
[46] Z. Liu et al., “Swin Transformer: Hierarchical Vision Transformer using Shifted Windows,” 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Oct. 2021, doi: 10.1109/iccv48922.2021.00986.