



A Study on Indirect Performance Parameters of Object Detection

Enoch Arulprakash¹ · A. Martin¹ · T. Miranda Lakshmi²

Received: 5 April 2022 / Accepted: 1 July 2022 / Published online: 21 July 2022
© The Author(s), under exclusive licence to Springer Nature Singapore Pte Ltd 2022

Abstract

Object detection is one of the inevitable tasks in the technological world. When the world started to rely entirely on technological intervention for almost all the tasks, different sectors started to implant artificial intelligence for precise decision making. Object detection is one among the category, which showed its applications in various domains including health care, military and anomaly detection, etc. Since there are many review on object detection, we focus only on the methods which are less expressed but indirectly have a significant performance gain. Notwithstanding, we review predominant methods of object detection including the pre-deep learning era. From the review, we are able to conclude indirect performance parameters of object detector has a significant impact on their performance for different problem scenarios. Finally, we also highlight the best characteristic of object detection in various applications.

Keywords Object detection · Context · Object proposal · Training strategy · Human intervention · Technological intervention

Introduction

From many aspects, technological intervention for human problems has changed its face from assisting to complete depending on the technology, especially, after the evolution of artificial intelligence and deep learning. Object detection is one among the task gaining its reputation almost in all the sectors. There are numerous reviews on the area. Therefore, we tried to avoid reassert the same topics again. Instead, we intensify the least expressed attributes of object detection.

The main motive of our study is to highlight indirect parameters of object detection also provide significant acceleration in performance. Moreover, we also briefly review predominant methods including the pre-deep learning era. Further, we tried to draft the best-researched applications of object detection over the decades from various domains.

The manuscript is organized as follows. In second section, it briefly reviews predominant methods, and third section analyzes indirect parameters of object detection. The fourth section drafts best applications of object detection, and the last section draws conclusion.

This article is part of the topical collection “Advances in Computational Approaches for Artificial Intelligence, Image Processing, IoT and Cloud Applications” guest edited by Bhanu Prakash K N and M. Shivakumar.

✉ Enoch Arulprakash
enocharulprakash03@gmail.com

A. Martin
martin@cutn.ac.in

T. Miranda Lakshmi
miranda@sjctnc.edu.in

¹ Department of Computer Science, Central University of Tamil Nadu, Thiruvurur 610104, India

² Department of Computer Science, St. Joseph's College of Arts and Science, Cuddalore 607001, India

Review on Predominant Methods

Object detection was carried out based on the template matching and object's part-based representation [16]. The focus was on a particular object whose object position layout is roughly adamant (like faces). Then, recognition was based on the object's geometric structure till 1990 [43]. Later, the focus shifted from geometry to the statistical classifier which was based on feature representation [like Adaboost [59], SVM [39] and Neural]. The feature representations through global handcrafted feature extraction-based classifier have set a stage for consecutive research in the ground. The appearance feature representation later shifted

from global representation [37] to local representation. The local representation was invariant to geometric transformations: rotation, occlusion, scale, viewpoint and illumination. Representative methods include SIFT [36], Haar-like features [59], shape contexts [6], local binary patterns (LBP), histogram of gradients (HOG) [12] and region covariance. After extracting local features, features are combined either through straightforward concatenation or feature pooling encoders. Through various methods, including bag of visual words [11], spatial pyramid matching and Fisher vectors [42], local hand feature descriptor methods gained a reputation for their invariant ability to geometric transformation.

In the deep learning era feature descriptor for an object, representation is automatically learned from the convolution neural network. Convolution layers of CNN are responsible for feature extraction; later extracted features are learned in the fully connected layers; and finally classification layer assigns class-specific labels. CNN extracts features layer-by-layer initial layer extracts elementary features and deep layers extract more robust features. Features extracted in the initial layers are combined by the deep layers to extract more discriminative features [54, 57].

Object detector uses CNN as backbone for object detection [7]. Predominant methods in the deep learning era include both single- and two-stage detectors [22, 26, 34, 45]. Single-stage detector associates class label and bounding box regressor into a single pipeline that does

not associate external or internal object proposal. Commonly, it partitions the input images into a coarse grid, and in each cell, objects are classified and bounding boxes are adjusted. Representative methods include Detector-Net OverFeat [48], YOLO [45] and SSD [35] as shown in Table 1. All these methods are identical which resolves class labels in each cell. However, it differs in simultaneously training bounding box regressor and resolving class labels, and YOLO and SSD are the two important detectors in the single-stage detector. YOLO [45] assigns the probability for all the classes in each cell. The class which obtained the highest probability is considered, and bounding boxes are adjusted with respect to the size of the object. Moreover, classification and bounding box training is carried out parallelly end to end. A single-stage detector has the upper hand in real-time applications like pedestrian detection and other moving objects since it is faster compared to the two-stage detector. SSD incorporated the advantage of YOLO and a two-stage detector to build an object detector as fast as YOLO and as accurate as a two-stage detector (faster RCNN). SSD [34] architecture comprises a fixed convolution size 1*1 with stride two throughout the network. Therefore, each consecutive layer decreases the feature map, and SSD associates classifier and detector in each layer to detect an object of varying size. The two-stage detector, on the other hand, associates a preprocessing object proposal, before resolving class

Table 1 Predominant methods of object detection

Detector/method	Proposal method	Pascal VOC 2012 (mAp)	Speed (FPS)	Published in	Pros and cons
RCNN	SS	53.3	<0.1	CVPR14	Pros: Effectively combines DCNN with selective search (SS). SS gave more potential regions for the object, First fully evolved object detector Cons: Heavy computational barrier since it has to scan for large number of frames and aspect ratio
Fast RCNN	SS	68.4	<1	ICCV15	Pros: Re-used the feature extraction from the convolution reduced the computation load to greater extent Cons: Object proposal from external source like SS was disclosed as a computational barrier
Faster RCNN	RPN	70.4	<5	NIPS15	Pros: DCNN has an inherent object localizing ability which is been applied with in a DCNN pipeline to localize as well as classify Cons: Speed of the detector is not applicable for real time applications
OverFeat	No proposal		<0.1	ICLR14	Pros: Non-proposal methods faster than proposal methods Cons: Sequential training for localization and classification required more time for training
YOLO	No proposal	57.9	<25 (VGG)	CVPR16	Pros: First unified framework combined both localization and classification Cons: Difficulty in detecting small objects
SSD	No proposal	74.9	<60	ECCV16	Pros: Combines the advantage of faster RCNN and YOLO. Accuracy wise accurate than YOLO Cons: Not accurate in detecting small object compared to faster RCNN

label and bounding box regression. An external region proposal is the computational barrier in two-stage detector. However, these methods are preferred when accuracy is given major preference over speed, as object proposal search for the clues for an object from the image. Therefore, it is effective in identifying even small objects which led to an effective approach for detecting a small object (nanoparticles, cells) and such application we will discuss in “[Applications of object detection](#)”. However, the object proposal initially was associated with an external proposal using objectness property which was based on object’s edge, color, texture and gradient. Through this approach, search space is reduced to a great extent. But, the external object proposal was not feasible as it occupies considerable time. Therefore, researchers started to associate object proposals within a DCNN pipeline which increased the performance substantially. Representative methods for two-stage detector include RCNN, fast RCNN [23], and faster RCNN as shown in Table 1. Faster RCNN evolved with object proposal within a DCNN pipeline.

To summarize, both single- and two-stage detectors methods: faster RCNN, YOLO and SSD, are frequently used for various applications as we will discuss in “[Applications of object detection](#)”. Faster RCNN is accurate where YOLO is faster. SSD combines both aspects of faster RCNN and YOLO.

Indirect Parameters of Object Detection

The architecture of an object detector plays a key role in the performance as we discussed in “[Review on pre-dominant methods](#)”. Two-stage detector associates object proposal before classification and regression as a different architecture from the single-stage detector which roughly divides the input images to coarse grids omitting the proposal. However, the distinct architecture yields different results as a two-stage detector performs by attaining good accuracy. On the other hand, a single-stage detector performs at a good speed. To sum up, the architecture of object detectors plays a key role in the performance of object detection. As there are numerous reviews from both the detection family, our survey tried to avoid reassert the same methods. Instead, we focus on the other parameters apart from architectural design which can contribute to the performance of object detectors. The indirect parameters includes.

- Context
- Object proposal
- Data augmentation
- Localization error
- Training strategy

Context

Context plays a significant role in object recognition, especially when the represented features are insufficient for prediction, i.e., when the detection framework encounters occlusion, small object or low image quality. Modeling a context provides additional clues for prediction. For instance, for detecting the objects in the kitchen, the possible objects are chimney, gas stove, vessels, cooker, etc.

The context broadly falls into two categories: (a) global context and (b) local context.

- a. **Global context:** It models an entire scene. Detection in office premises will predict the presence of cubicle and laptop and system. Contextual details are combined with the regular feature representation for final prediction [18].
- b. **Local context:** It represents the relationship between the objects. Object’s boundary gives additional details about its interaction with other objects. Expanding objects boundary and exploiting in the boundary regions will provide more supplementary information such as object’s above, below, behind, right and left with other objects which provides a additional clue for prediction from its structural constraints. For example, the proposed object can be a door locker if the object behind is a door, and the proposed object is smaller than a door [18].

DCNN exploits contextual details without explicit modeling since the CNN architectural setup enforces hierarchical feature representation. Notwithstanding, dedicated research has been carried out by explicitly modeling local and global context; the representative framework includes CoupleNet [65], ORN [29], DeepIDNet [30], ION [5]. However, in addition to CNNs hierarchical feature representation, both the detectors (single- and two-stage detectors) have an implicit context modeling. In particular, the single-stage detector looks entire image for detection, thereby modeling a global context. In the two-stage detector, the regressor’s subnetwork appropriates the object boundary by exploiting object boundary.

Object Proposal

Object proposal is a preprocessing step before actual detection. In the absence of object proposal, the detector scans different scale and aspect ratios [12, 15, 59, 66] which leads to computation load and makes the entire process very slow. Object proposal eases the detection framework by selectively giving a few proposals [58] from objectness property (edge, texture, color, gradient) [35]. After the growth of DCCN, the selective search was the computation bottleneck for the

entire detection framework. It is being proved that DCNN has excellent proficiency in locating an object from their conv layers [46]. Later this idea turned to propose the object within the detection framework. DCNN proposal has a computational advantage over external proposal methods (selective search, MCG and EdgeBoxes [30]) and provides a unified framework. Combining proposal, classification and bounding box regression, the first such method of a proposal using DCNN was the region proposal network (RPN) [46] which combined RPN with RCNN and is a milestone in object detection (faster RCNN) [46]. Consequently, many DCNN-based proposal methods have arrived, representative methods include DeepProposal [19], ZIP [32], DeNet [56], etc., which further improved the performance of object proposal. A two-stage detector with RPN is the key for many detection challenges, including Pascal VOC and COCO. Notwithstanding, DeepProposal [19], ZIP [32], DeNet [56], etc., have a performance gain in comparison with RPN with slight computation load.

Data Augmentation

Data augmenting refers to artificially stressing the training data to the various transformations. Such as scaling, cropping, rotating, flipping, distorting and adding noise, leaving the underline category unchanged as augmentation produces more training samples, helps in generalization and avoids over fitting [41, 61]. Researchers [14, 24] proposed adding a datasets by pasting segmented objects into realistic images. Further, Dvornik et al. showed [13] that correctly modeling objects local context is a key to place them in the right surrounding [34].

Localization Error

IOU is an evolution matrix for localization whose performance can eventually affect the detection framework. Intersection over union compares the predicted bounding box and ground truth and, ordinarily, expected to be more than or equal to 0.5. The bounding box regressor optimizes the bounding area aiming to increase IOU in parallel with classification. Bounding boxes are a coarse estimation. Therefore, background pixels are combined with a bounding box, which affects the performance of localization. Usually, some post-processing step, such as non-maximum suppression [8, 28, 34], is applied to remove inappropriate bounding box. But, the excellent localization can be suppressed due to the wrong alignment. However, few approaches are developed to minimize localization error. Representative methods include MRCNN [20], CRAFT [62], cascade RCNN [9]. In MRCNN, RCNN is applied several times to adjust the boundingbox iteratively. CRAFT [62] and AttractionNet [21] adopts a multistage detection to bring the best proposals,

handover to fast RCNN. CaiVasconceolos proposed cascade RCNN, an extension of multistage RCNN, where cascading RCNN is trained sequentially with each RCNN increasing IOU threshold.

Training Strategy

A deep learning detection framework requires massive data to perform well. Moreover, data augmentation is commonly applied during training to alleviate scale variations problems. Training with massive data tends to complicate and overload the process. Effective training and fast convergence are at most concern during training. A few training strategies is proved effective in literature. Singh and Davis proposed SNIP [8, 50–52] that introduced an innovative training technique that decreases scale variations without shrinking the training data. Sing et al. proposed SNIPER, which efficiently processes only context area about ground truth by the relevant scale instead of dealing with the entire image pyramid. MiniBatch size plays a key role in past convergence. Peng et al. proposed MegDet that enabled a large MiniBatch size, effective in faster training and rapid convergence. Further, Peng et al. introduced concurrent GPU training that eases the COCO dataset training by finishing the training in four hours by concurrently processing in 128 GPUs, with the help of GPU batch normalization and novel learning rate policy, impressive in winning the COCO 2017 detection challenge [40].

Comparative Analysis of Indirect-Performance

From sections “Context”, “Object proposal”, “Data augmentation”, “Localization error” and “Training strategy”, we have discussed indirect performance parameters and the corresponding representative methods. To highlight the effectiveness of these parameters and the environment where it can yield more performance in comparison with general detectors are analyzed with the recent works as shown in Table 2. Comparison follows the standard evolution sequences as depicted in Fig. 1.

After the comparative analysis on the indirect performance parameters with different problem scenarios such as insufficient datasets, insufficient feature extraction, detecting small objects, localization error, training massive deep learning models, results from the comparative analysis as shown in Table 2 highlight indirect parameters can significantly contribute to the performance when the generic detector algorithm drops the performance due to the lack of data samples, feature extraction, class imbalance, etc. Moreover, from the comparative analysis we suggest specific parameter for different problem scenarios.

Table 2 Comparative analysis of indirect performance parameters of object detection

Problem scenario	Problem area	Publication	Data set		Performance of generic object detector (faster RCNN, SDD or YOLO)	Indirect Parameters Type	Performance along with indirect parameters		Significance
			Name	Matrix			Name	Value	
Insufficient data-sets	classification of calcaneal fractures in CT images	Injury [1]	CT calcaneus image	mAp	0.53	Data Augmentation	Geometric Transformation	0.72	Augmenting the CT calcaneus images into different Geometric transformation (rotation, distortion and flips), improved 35% more accuracy compared to generic detector
Insufficient feature extraction	Small road object detection	Automotive innovation [10]	TI-DRO	mAp	0.65	Context modeling	Context-Aware Network	0.73	Method fuses the semantic context information along with the generic feature extraction to obtain robust descriptive details which proved to increase the accuracy
Insufficient feature extraction	Salient object detection	CVF [53]	MS-COCO	Avg F-measure	0.6	Context modeling		0.84	A saliency model that exploits semantic scene context for improving saliency reasoning
Detecting small objects	Agriculture pest detection	Computers and electronics in agriculture (R. [60])	AgriPest21	mAp	0.67	Object proposal	Region Proposal Network	0.78	S-region proposal network is effective in locating and detecting small objects
Detecting small objects	Remote sensing images	International Journal of Geo-Information [17]	DIOR datasets	mAp	0.69	Object proposal	Attention-guided dense network	0.79	Attention guided dense network consist DCNN-based object proposal effective in achieving state of the art results in DIOR datasets

Table 2 (continued)

Problem scenario	Problem area	Publication	Data set		Performance of generic object detector (faster RCNN, SDD or YOLO)	Indirect Parameters Type	Performance along with indirect parameters		Significance
			Name	Matrix			Name	Value	
Class imbalance	Vehicle detection	Sensing and imaging [25]	KITTI	mAp	0.65	Localization error	Modified cascade RCNN for vehicle detection	0.83	Cascade RCNN proved to increase accuracy; when target objects lead to wrong detection, small object or when poses occlusion for different weather condition
Concurrency	General object detection	CVPR [40]	MS COCO	Hours and mAp	33 Hours	Training strategy	MegDet: A large mini-batch object detector	0.525	Concurrent GPU training that eases the COCO dataset training by finishing the training in 4 h. By concurrently processing in 128 GPUs, with the help of GPU batch normalization and novel learning rate policy
Training massive data	General object detection	IEEE Transaction [52]	Not Applicable	Speed		Training strategy	Scale normalized image pyramids	2.5 to 5	Scale normalized image pyramid enables fixed size objects during training and thus achieved 2.5 to 5 times speed-up than the normal training

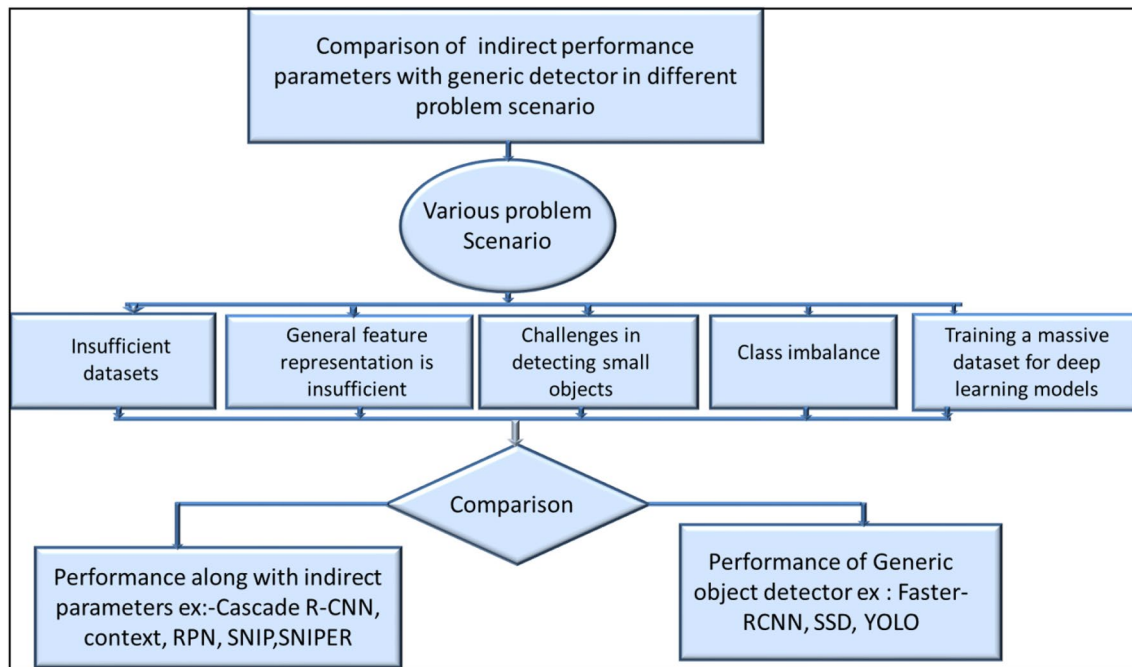


Fig. 1 Comparative analysis of indirect performance parameter

- Data augmentation approach: when lack of training data.
- Context modeling: Feature extraction is not sufficient or quality of the image is not up to the mark.
- Object proposal: When detecting very small or tiny objects.
- Effective localization methods: When there is huge class imbalance among the different class in the dataset.
- Training strategy: Training huge volume of dataset.

Applications of Object Detection

Object detection has been widely used in numerous applications, especially in the field of medical, military, security, anomaly detection and science and engineering as shown in Fig. 2.

Medical Field

Brain Tumor

Manual segmentation of brain tumors is a laborious and time-consuming task for radiologists. A deep learning paradigm is developed as a feasible preference for applications in medical imaging. They can grasp discerning features instinctively, as a neural network can learn a brain's essential features in regulation to classify and segment tumors. This approach outperforms manual segmentation and the classical machine learning approach in stipulations of false-positive decline. Among the

deep learning established ways, CNNs have provided supreme performance for brain tumor segmentation [44].

Radiolucent Lesions

Identification and segmentation of mandibular radiolucent lesions on panoramic radiographs: It aims at five radiolucent lesions (radicular cysts, dentigerous cysts, ameloblastomas, odontogenickeratocysts and simple bone cysts) which takes place regularly in the mandible. A deep learning approach had proven a high standard of detection and classification awareness in the detection of radiolucent lesions of the mandible [3].

Cell Biology

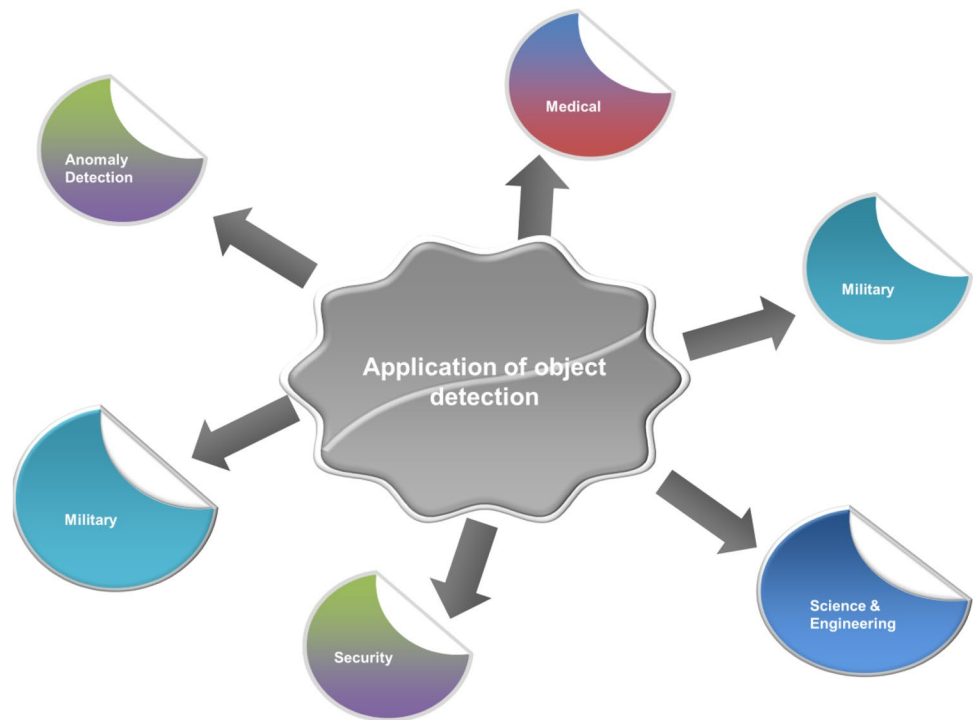
Segmenting a cell from blood or other tissue has significant challenges due to morphology, color intensity, and cell size variability but with the precise accuracy of deep vision. It has displayed outstanding accuracy in classifying and detecting B cells and T cells, detached by a micro-microfluidic chip [55].

Security Field

Luggage Scanner

In the international travel web, the increased traveller throughput and enlarged border security (e.g., postal, sailing

Fig. 2 Applications of an object detector



and freight). The consequences in demanding a well-timed computerized image identification. Convolution neural networks (CNN), a leading in modern object detection problems, are also used in X-ray baggage images for identifying a potential object of threat (gun, shuriken, razor blade and knife) objects. The research result highlights that CNN achieves exceptional accuracy in detecting threat objects [2].

Anomaly Detection

Identification of Defects in Tiny Particles

Tiny tools (less than 1 cm) particles can flaw due to working conditions and poor design. Due to holes, sags, and abrasions, mass-produced products are prone to fault. Decay and lethargy damage arise in day-to-day functions; deep learning's effective feature extraction capability is utilized to identify tiny particles defects. An SSD object detector proved to detect the flaw of 0.8-cm darning needles accurately [63].

Anomaly in Steel Structure

Bolt loosening will affect the safety of the steel arrangement and may lead to severe accidents. Due to the bolt joints, complicated fluctuation properties, it is hard to recognize the bolt loosening in steel structures from a conventional dynamics perspective. However, deep learning's intense feature extraction capability is used to detect the screw and

screw number. From the detected set, bolt loosening is effectively identified using trigonometric relationships [64].

Anomaly in Food Particles

The identification of an external object plays an essential part in the agriculture commodity. In various ways, external particles are brought into food brands; foreign objects in foods are the sole means of customer complaints. The identification of external particles is hugely significant for quality and health. It is a major concern for the food security convention. Manual extraction of foreign objects from food material is a time-consuming and labors task. Foreign object segregation from walnuts using DCNN is applied. DCNN overcomes the cumulating phenomenon between walnuts and external particles, which was a challenging task in manual feature engineering. The DCNN performed with above 99% in more than 100 test images [47].

Science and Engineering

Nanoparticles Segmentation

Images produced from the microscopes are large in number and resolution. Shapes and size distributions and properties of nanoparticles play an essential part in interpreting the material. Hence, each image of nanoparticles should be identified and detected for giving a measurable guide. These particles segmentation is challenging because of the

overlapping instances, changeable particle sizes and shapes. Moreover, manual detection and segmentation of nanoparticles are laborious and time-consuming. A deep learning paradigm is used for detecting and segmenting the nanoparticles from TEM (transmission electron microscopy) images. Multiple output convolution neural networks (MO-CNN) are used, for concurrent recognition and segmentation of nanoparticles. The proposed deep learning approach is powerful and efficient, with immense precision and capable of studying nanoparticles, even in overlapping particles and complex backgrounds [38].

Military

Surveillance

Surveillance plays a crucial role in the continuous analysis of massive amounts of critical visual information. Detecting targets, monitoring security-sensitive areas, and suspects possible suspicious activities lead to an increase in cognitive load and exhaustion in energy level; moreover, it is prone to error. The best alternative is to replace, with a computer vision, to detect suspects and monitor with ease. Furthermore, the probability of error will be lesser [4].

Conclusions

This survey briefly reviewed predominant methods of object detection from pre-deep learning methods. Most importantly, we have given preference for the indirect parameters. From the review and comparative analysis, we conclude indirect performance parameters plays a crucial role in various problems across different areas by boosting the performance in comparison with generic detector. Furthermore, we have highlighted which parameters can be the appropriate choice for different problem conditions.

Moreover, we have shown the best characteristics of object detection in various domains. Results from the various applications showed deep learning methods outperformed conventional methods and goes behind the human visual perception. Therefore, the transition of technological intervention from assisting to completely depending brought serious anxiety on the future role of human intervention for various tasks. Nevertheless, it is essential to associate with the evolving technology to give continuity against the fast-evolving machine-centric period.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Aghnia Farda N, Lai JY, Wang JC, Lee PY, Liu JW, Hsieh IH. Sanders classification of calcaneal fractures in CT images with deep learning and differential data augmentation techniques. *Injury*. 2021;52(3):616–24. <https://doi.org/10.1016/j.injury.2020.09.010>.
2. Akcay S, Kundegorski ME, Willcocks CG, Breckon TP. Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery. *IEEE Trans Inf Forensics Secur*. 2018;13(9):2203–15. <https://doi.org/10.1109/TIFS.2018.2812196>.
3. Arijji Y, Yanashita Y, Kutsuna S, Muramatsu C, Fukuda M, Kise Y, Arijji E. Automatic detection and classification of radiolucent lesions in the mandible on panoramic radiographs using a deep learning object detection technique. *Oral Surg Oral Med Oral Pathol Oral Radiol*. 2019;00(00):1–7. <https://doi.org/10.1016/j.oooo.2019.05.014>.
4. Arulprakash E, Aruldoss M. A study on fight against COVID-19 from Latest Technological Intervention. *SN Comput Sci*. 2020. <https://doi.org/10.1007/s42979-020-00301-0>.
5. Bell S, Zitnick CL, Bala K, Girshick R. Inside-outside net: Detecting objects in context with skip pooling and recurrent neural networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016. pp. 2874–2883.
6. Belongie S, Malik J, Puzicha J. Shape matching and object recognition using shape contexts. *IEEE Trans Pattern Anal Mach Intell*. 2002;24(4):509–22.
7. Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. *IEEE Trans Pattern Anal Mach Intell*. 2013;35:1798–828. <https://doi.org/10.1109/TPAMI.2013.50>.
8. Bodla N, Singh B, Chellappa R, Davis LS. Soft-NMS—improving object detection with one line of code. In: *Proceedings of the IEEE international conference on computer vision*, pp. 5561–5569. 2017.
9. Cai Z, Vasconcelos N. Cascade r-cnn: delving into high quality object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018. pp. 6154–6162.
10. Chen G, Chen K, Zhang L, Zhang L, Knoll A. VCANet: vanishing-point-guided context-aware network for small road object detection. *Autom Innov*. 2021. <https://doi.org/10.1007/s42154-021-00157x>.
11. Csurka G, Dance C, Fan L, Willamowski J, Bray C. Visual categorization with bags of keypoints. In: *Workshop on statistical learning in computer vision, ECCV (Vol. 1, No. 1–22, pp. 1–2)*. 2004.
12. Dalal N, Triggs B. Histograms of oriented gradients for human detection. In: *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05) (Vol. 1, pp. 886–893)*. 2005.
13. Dvornik N, Mairal J, Schmid C. Modeling visual context is key to augmenting object detection datasets. In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018. pp. 364–380.
14. Dwibedi D, Misra I, Hebert M. Cut, paste and learn: Surprisingly easy synthesis for instance detection. In: *Proceedings of the IEEE international conference on computer vision*, 2017. pp. 1301–1310.
15. Felzenszwalb P, McAllester D, Ramanan D. A discriminatively trained, multiscale, deformable part model. In: *2008 IEEE conference on computer vision and pattern recognition*. 2008. pp. 1–8. Ieee.
16. Fischler MA, Elschlager RA. The representation and matching of pictorial structures. *IEEE Trans Comput*. 1973;100(1):67–92.

17. Fu H, Fan X, Yan Z, Du X. Detection of schools in remote sensing images based on attention-guided dense network. *ISPRS Int J Geo Inf*. 2021;10(11):736.
18. Galleguillos C, Belongie S. Context based object categorization: a critical survey. *Comput Vis Image Underst*. 2010;114(6):712–22.
19. Ghodrati A, Diba A, Pedersoli M, Tuytelaars T, Van Gool L. Deepproposal: Hunting objects by cascading deep convolutional layers. In: *Proceedings of the IEEE international conference on computer vision*, 2015. pp. 2578–258.
20. Gidaris S, Komodakis N. Object detection via a multi-region and semantic segmentation-aware cnn model. In: *Proceedings of the IEEE international conference on computer vision*. 2015. pp. 1134–1142.
21. Gidaris S, Komodakis N. Attend refine repeat: active box proposal generation via in-out localization. 2016. arXiv preprint [arXiv:1606.04446](https://arxiv.org/abs/1606.04446).
22. Girshick R, Donahue J, Darrell T, Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014. pp. 580–587.
23. Girshick R. Fast R-CNN. In: *Proceedings of the IEEE International Conference on Computer Vision*, 2015. pp. 1440–1448. <https://doi.org/10.1109/ICCV.2015.169>.
24. Gupta A, Vedaldi A, Zisserman A. Synthetic data for text localisation in natural images. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. pp. 2315–2324.
25. Han X. Modified cascade RCNN based on contextual information for vehicle detection. *Sens Imaging*. 2021;22(1):1–19. <https://doi.org/10.1007/s11220-021-00342-6>.
26. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE Trans Pattern Anal Mach Intell*. 2015;37(9):1904–16.
27. He K, Zhang X, Ren S, Sun J. Spatial pyramid pooling in deep convolutional networks for visual recognition. In: *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 8691 LNCS(PART 3). 2014. pp. 346–361. https://doi.org/10.1007/978-3-319-10578-9_23.
28. Hosang J, Benenson R, Schiele B. Learning non-maximum suppression. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. pp. 4507–4515.
29. Hu H, Gu J, Zhang Z, Dai J, Wei Y. Relation networks for object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018. pp. 3588–3597.
30. Kong T, Sun F, Yao A, Liu H, Lu M, Chen Y. RON: Reverse connection with objectness prior networks for object detection. In: *Proceedings—30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, 2017-Janua*. 2017. pp. 5244–5252. <https://doi.org/10.1109/CVPR.2017.557>.
31. LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, Jackel LD. Backpropagation applied to digit recognition. *Neural Comput*. 1989;1:541–51.
32. Lenc K, Vedaldi A. Understanding image representations by measuring their equivariance and equivalence. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015. pp. 991999.
33. Liu L, Ouyang W, Wang X, Fieguth P, Chen J, Liu X, Pietikäinen M. Deep learning for generic object detection: a survey. *Int J Comput Vis*. 2020;128(2):261–318.
34. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. SSD: Single shot multibox detector. In: *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 9905 LNCS. 2016. Pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
35. Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu CY, Berg AC. SSD: Single shot multibox detector. In: *Lecture notes in computer science (including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics)*, 9905 LNCS. 2016. pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2.
36. Lowe D. Object recognition from local scale-invariant features. In: *Proceedings of the IEEE international conference on computer vision*, 2. 2001.
37. Murase H, Nayar SK. Visual learning and recognition of 3-D objects from appearance. *Int J Comput Vision*. 1995;14(1):5–24.
38. Oktay AB, Gurses A. Automatic detection, localization and segmentation of nano-particles with deep learning in microscopy images. *Micron*. 2019;120:113–9. <https://doi.org/10.1016/j.micron.2019.02.009>.
39. Venkatesan C, Karthigaikumar P, Paul A, Satheeskumaran S, Kumar R. ECG signal pre-processing and SVM classifier-based abnormality detection in remote healthcare applications. *IEEE Access*. 2018;6:9767–73.
40. Peng C, Xiao T, Li Z, Jiang Y, Zhang X, Jia K, Sun J Megdet: A large mini-batch object detector. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018. pp. 6181–6189.
41. Peng X, Sun B, Ali K, Saenko K. Learning deep object detectors from 3d models. In: *Proceedings of the IEEE international conference on computer vision*. 2015. pp. 1278–1286.
42. Perronnin F, Sánchez J, Mensink T. Improving the fisher kernel for large-scale image classification. In: *European conference on computer vision*. 2010. pp. 143–156. Springer, Berlin, Heidelberg.
43. Ponce J, Hebert M, Schmid C, Zisserman A (eds). *Toward category-level object recognition*, Vol. 4170. Springer. 2007.
44. Razzak I, Imran M, Xu G. Efficient Brain tumor segmentation with multiscale two-pathway-group conventional neural networks. *IEEE J Biomed Health Inf* 1:1. <https://doi.org/10.1109/JBHI.2018.2874033>.
45. Redmon J, Divvala S, Girshick R, Farhadi A. You only look once: unified, real-time object detection. In: *Proceedings of the IEEE computer society conference on computer vision and pattern recognition*, 2016;779–788. <https://doi.org/10.1109/CVPR.2016.91>.
46. Ren S, He K, Girshick R, Sun J. Faster R-CNN: towards real-time object detection with region proposal networks. *IEEE Trans Pattern Anal Mach Intell*. 2017;39(6):1137–49. <https://doi.org/10.1109/TPAMI.2016.2577031>.
47. Rong D, Xie L, Ying Y. Computer vision detection of foreign objects in walnuts using deep learning. *Comput Electron Agric*. 2019;162(February):1001–10. <https://doi.org/10.1016/j.compag.2019.05.019>.
48. Rolet P, Sebag M, Teytaud O. Integrated recognition, localization and detection using convolutional networks. In: *Proceedings of the ECML conference*. 2012. pp. 1255–1263.
49. Singh B, Davis LS. An analysis of scale invariance in object detection snip. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018. pp. 3578–3587.
50. Singh B, Li H, Sharma A, Davis LS. R-fcn-3000 at 30fps: Decoupling detection and classification. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2018. pp. 1081–1090.
51. Singh B, Najibi M, Davis LS. Sniper: Efficient multi-scale training. 2018. arXiv preprint [arXiv:1805.09300](https://arxiv.org/abs/1805.09300).
52. Singh B, Najibi M, Sharma A, Davis LS. Scale normalized image pyramids with autofocus for object detection. *IEEE Trans Pattern Anal Mach Intell*. 2021.
53. Siris A, Jiao J, Tam GK, Xie X, Lau RW. Scene context-aware salient object detection. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2021. pp. 4156–4166.
54. Szegedy C, Liu W, Jia Y, Sermanet P, Reed S, Anguelov D, Rabinovich A. Going deeper with convolutions. *IEEE Conf Comput Vis Pattern Recognit*. 2015. <https://doi.org/10.1109/CVPR.2015.7298594>.

55. Turan B, Masuda T, Noor AM, Horio K, Saito TI, Miyata Y, Arai F. High accuracy detection for T-cells and B-cells using deep convolutional neural networks. *ROBOMECH J* 2018;5(1). <https://doi.org/10.1186/s40648-018-0128-4>
56. Tychsen-Smith L, Petersson L. Denet: Scalable real-time object detection with directed sparse sampling. In: *Proceedings of the IEEE international conference on computer vision*. 2017. pp. 428–436.
57. Tygert M, Bruna J, Chintala S, LeCun Y, Piantino S, Szlam A. A mathematical motivation for complex-valued convolutional networks. *Neural Comput*. 2016;28(5):815–25. https://doi.org/10.1162/NECO_a_00824.
58. Vaillant R, Monroq C, Le Cun Y. Original approach for the localisation of objects in images. *IEE Proc Vis Image Signal Process*. 1994;141(4):245–50.
59. Viola P, Jones M. Managing work role performance: challenges for twenty-first century organizations and their employees. *Rapid Object Detection Using a Boosted Cascade of Simple Features*. 2001. <https://doi.org/10.1109/CVPR.2001.990517>.
60. Wang R, Jiao L, Xie C, Chen P, Du J, Li R. S-RPN: sampling-balanced region proposal network for small crop pest detection. *Comput Electron Agric*. 2021;187: 106290.
61. Wang X, Shrivastava A, Gupta A. A-fast-rcnn: Hard positive generation via adversary for object detection. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2017. pp. 2606–2615.
62. Yang F, Choi W, Lin Y. Exploit all the layers: Fast and accurate cnn object detector with scale dependent pooling and cascaded rejection classifiers. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016. pp. 2129–2137.
63. Yang J, Li S, Wang Z, Yang G. Real-time tiny part defect detection system in manufacturing using deep learning. *IEEE Access*. 2019;7:89278–91. <https://doi.org/10.1109/access.2019.2925561>.
64. Zhao X, Zhang Y, Wang N. Bolt loosening angle detection technology using deep learning. *Struct Control Health Monit*. 2019;26(1):1–14. <https://doi.org/10.1002/stc.2292>.
65. Zhu Y, Zhao C, Wang J, Zhao X, Wu Y, Lu H. Couplenet: Coupling global structure with local parts for object detection. In: *Proceedings of the IEEE international conference on computer vision*. 2017. pp. 4126–4134.
66. Zitnick CL, Dollár P. Edge boxes: Locating object proposals from edges. In: *European conference on computer vision*, 2014. pp. 391–405. Cham: Springer.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.