Performance Analysis on Deep Learning State of Art Algorithms for Object Recognition

Ms. Yogitha. R[1], Dr. G. Mathivanan[2]

[1,2]*Sathyabama Institute of Science and Technology*
yogitha.ravi1915@gmail.com [1], mathivanan.it@sathyabama.ac.in[2]

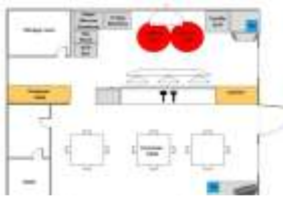
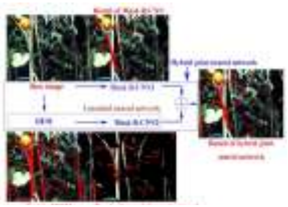
KEYWORDS
Computer Vision, Object Recognition, Artificial Intelligence, Images, Videos

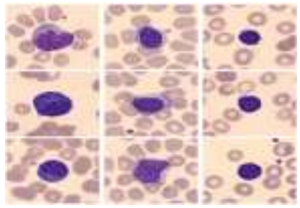
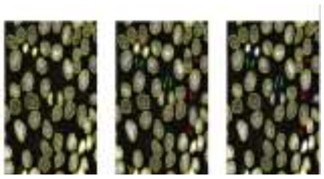
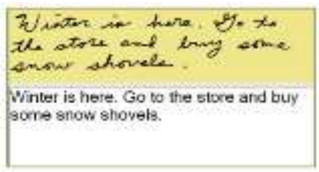
ABSTRACT:
The goal of computer vision, a subfield of computer science, is to replicate some of the intricacies of the human visual system so that machines can recognize and interpret images and videos in the same manner that humans do. Until recently, computer vision was only used in a restricted capacity. In the past few years, artificial intelligence has advanced significantly, outperforming humans in a number of tasks involving object detection, recognition, and classification. This has allowed computer vision to grow exponentially in terms of increasing the precision with which machines can recognize the objects in and around the surrounding environment. A computer vision technology called object recognition helps find and identify objects in a series of images and videos. Despite the fact that the image of the things varies in different viewpoints, different sizes and scales, or when they are translated or rotated, humans can recognise a large number of objects in images with minimal effort. Even when partially obscured from view, human vision system has the greatest capability to identify the objects. Whereas, for computer vision systems, this task is still a difficulty. Over the years, several different approaches and innovations in the algorithm have been tried to impose the human's capability into a computer's vision system. This paper provides a thorough investigation on the evolution of Object Recognition algorithms, datasets used and its performance metrics in a precise manner which will guide the future researchers a direction to proceed their research in innovating algorithms with better accuracy.

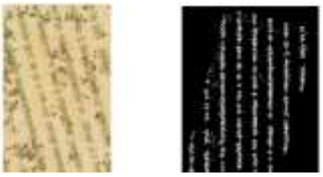
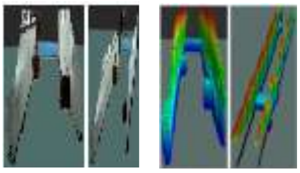
1.Introduction:

Computer Vision is a broad field in computer science which deals with mimicking the human vision in the machine. To perceive how the human eyes and human brain in coordination, work on identifying and recognizing the environment around is a great task, in which researches are still going on. Computer Vision envisions machine doing the same task in a more effective way. If the machine can recreate a human's visual perception, it might be the breakthrough in the field of computer science and artificial intelligence. Back in history, nearly till the end of 19th century Computer Vision [1] was always considered to be one of most notoriously difficult field of research. This was because of the fact that the resources that were available back in those time were very scarce. There were no sight of innovation or improvement in the field of communication, network, hardware or software. But as the innovations were made, the field of Computer Vision paved its way in improvising its algorithms and techniques. Computer Vision can be broadly categorised into many sub-divisions. Object Classification, Object Detection, Object Recognition, Classes Segmentation, Instance Segmentation, Object Localization etc. are some of the categories under it. Each sub-domain deals with its own set of problem. In this paper our focus of work will be on the classic problem Object Recognition.

One of the most fascinating facts on a human being's visual perception is that it can identify and perceive the artifacts around its environment despite of the different variances and disturbance present in the environment. Object Recognition tries to envision human's visual perception in a machine by using different algorithms and framework designed exclusively for it. It has been a persistent and challenging problem statement in the field of research with innovations happening day today. The goal of the Object Recognition algorithms varies according to the application, task and the input that's been considered. Certain kind of applications might have the need to identify difference between the same species of objects like Face Recognition (identifying different people with slight difference in same kind of attributes and features), Sub-Species Identification (identifying and classifying botanical sub classes of same flower), Surveillance (identifying the activities of different people with same features in the given frame), Automatic Image Annotation (identifying different instances of same objects and annotating it with keywords for efficient dataset generation), Automatic Target Recognition (identifying the already decided target object from real time distant images generated) etc. Certain other kind of applications will have the need to identify between different types and species of identification like Character recognition (identifying different types of characters in a text), Counting and monitoring the objects (identifying different objects in the scene, counting and monitoring them like different species of fishes in the sea from satellite images), Autonomous Driving (identifying different instances around the driving environment), Panoramic Image Recognition (identifying objects from panoramic images generated from different angles), Visual Positioning and Tracking (identifying and tracking objects from real time distant images generated by cameras in different angles) etc. Table 1 represents papers associated with these specified applications of Object Recognition along with its concept and output image. Depending on the application or task in hand, availability of input dataset, either of the Object Recognition techniques can be used. Though the field of Computer Vision deals with variety of problems like Object Detection, Image Segmentation, Pose Estimation, Motion Analysis, Scene Reconstruction etc. we primarily focus on Object Recognition. This is because of the fact that to recreate human's visual perception, highest form of accuracy is needed in recognizing the objects and environment around despite the differences and disturbances.

No.	Reference and Application	Year	Content	Output
1	[60] – Object Identification and Classification	2022	Using small CNN with potentially less number of parameters to identify and classify fish species from underwater images and videos	
2	[61] – Activity Detection	2022	Assessing the quality of air by recognizing the activities in a kebab house with the help of sensors and pattern recognition	
3	[62] – Sub Species Recognition	2022	Detecting and recognizing the stem by differentiating the colour similarity between stem and leaves of colour images of tomato plant	

4	[63] – Medical Imaging	2022	Recognizing the morphology of a corresponding cell present in the blood stream and detecting whether its COVID 19 Virus cell structure	
5	[64] – Target Recognition and Tracking	2022	To achieve target tracking, a deep-learning based surveillance and reconnaissance system for autonomous surface vehicles that uses the Siamese network as the fundamental neural network architecture.	
6	[2] – Face Recognition	2021	3D morphable model incorporation into GAN network for creating large scale face recognition datasets	
7	[3] – Sub-Species Identification	2021	Tree species identification mobile app that would detect, recognize and classify based on tree image remotely in real time.	
8	[4] - Surveillance	2021	Activity recognition is performed at four levels using CNN for detection, MOSSE for tracking, LiteFlowNet CNN for feature extraction, DS-GRU for learning activity change in frames	
9	[8] – Automatic Image Annotation	2021	Fully automated system to annotate cell nuclei from custom florescent cell nuclei dataset. Annotated images generated provides equal segmentation performance comparitively	
10	[9] – Automatic Target Recognition	2021	Proposed technique is a combination of channel-shuffle, channel-attention, inverted residual block and ASIR-Block to identify targets from SAR images	
11	[5] – Character Recognition	2021	Proposed method uses SVM in integration with four different methods namely structural pattern recognition,	

			template matching, statistical and graphical method	
12	[6] – Object Counting and Monitoring	2021	Unsupervised Agent-based Learning method to count plants as object. Relies on relationship between objects detected in prior.	
13	[7] – Autonomous Driving	2021	Human vision-based mechanism that maps the image of a scene to a occupancy grid representation which creates a abstract scenario representation for autonomous driving	
14	[10] – Panoramic Image Recognition	2021	Given a panoramic image of google street view, proposed CNN generated meaningful proper views of images for better identification	
15	[11] – Visual Positioning and Tracking	2021	For body-pose estimation and navigation map construction SLAM based method is proposed that uses RGD-B camera to collect images for the same	

There has been a huge improvement in number of research articles published over the years in different Computer Vision tasks like Object Recognition, Object Detection, Object Segmentation and Object Localization. As illustrated in Figure 1, Object Recognition has always been on higher focus when compared with other tasks. This is because of the fact that all these tasks are almost inter related with each other and Object Recognition serves as problem with higher dimension compared with others. So, finding a solution or algorithm for effective recognition of objects will automatically lead to solutions for rest of the CV tasks. This is why main focus of our study is completely based on Object Recognition.

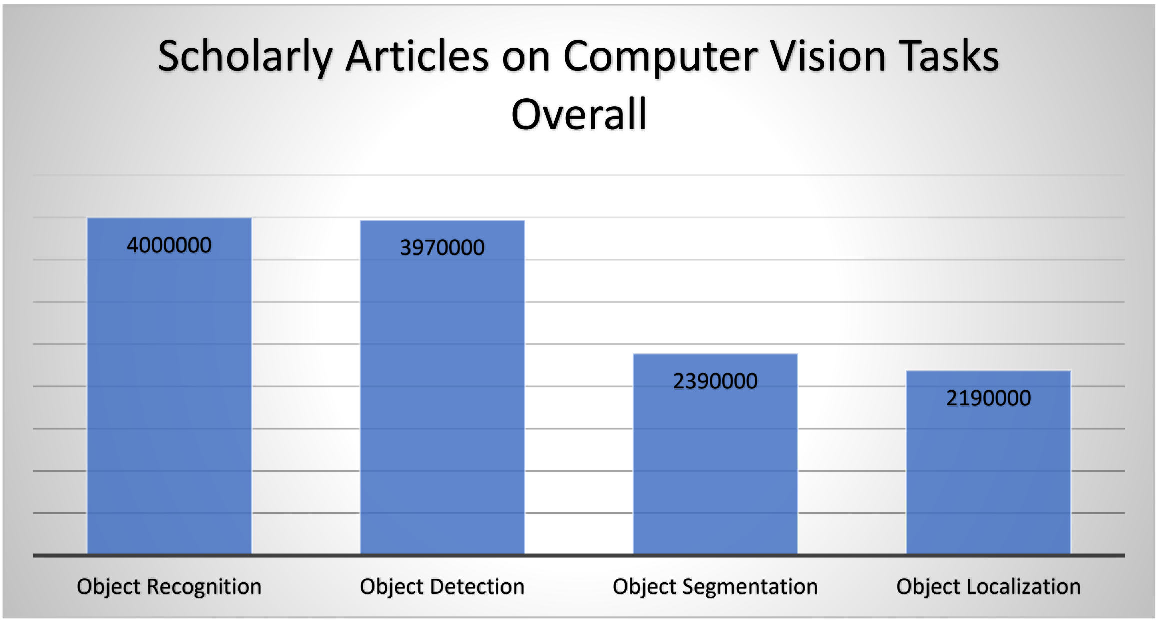
1.1 Comparison with previous reviews:

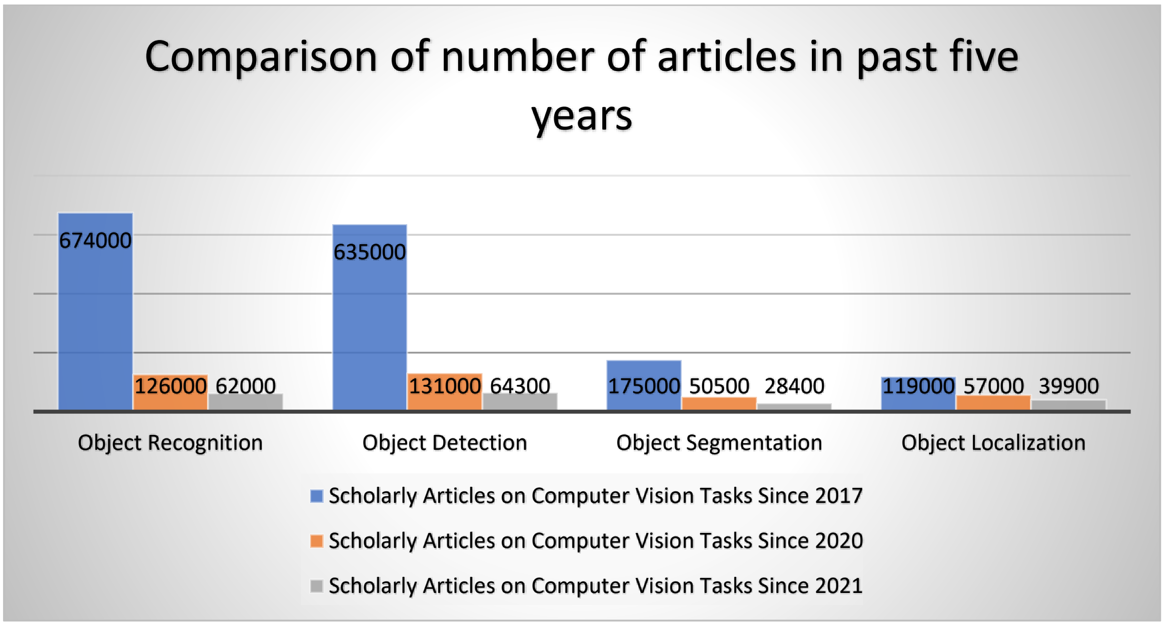
The survey papers on research domain of object recognition have started very early, four to five decades back starting with listing down simple geometrical methods for recognizing objects. Table 1 shows the wide range of survey articles and books that's related to object recognition domain that ranges between 1985 till 2021. It contains survey papers that's specific to one particular application like recognizing 3D objects, recognition for robotic navigation, grasping and vision, recognizing visual scene and open scene, recognizing satellite images, adversarial examples, point cloud images, complex commodity images etc. These papers discuss about methods and techniques that's specifically suitable for any considered application type (Robotic application, 3D modelling, surveillance, tracking objects etc.) or input type (complex commodity images, SAR images, RGDB camera images etc.)

No.	Reference	Year	Venue	Content
1	[12]	2008	MUT	Local (corner, edges, entropy) and Global (model data of whole image) approaches of appearance-based object recognition algorithms
2	[13]	2014	arXiv	List of Neural Network models for Object Recognition along with datasets and their performance on those datasets have been tabulated
3	[14]	2014	IEEE	Comprehensive study on list of available feature based local surface 3D Object Recognition methods
4	[15]	2016	Elsevier	Analysis on object recognition methods that are completely based on available local invariant features in different robotic perspective
5	[21]	2016	ResearchGate	List of feature extraction techniques along with its different types of invariance properties that's used in image retrieval system
6	[22]	2020	ACM	Discussion on how adversarial examples impact on robustness, security and safety of various neural networks
7	[16]	2013	Elsevier	Evolution of Object Recognition system based on computer over the past 50 years
8	[17]	1993	ACM	A comprehensive survey on model-based vision systems for four different types of vision tasks on dense range images
9	[18]	1985	ACM	A clear definition of 3D object recognition problem statement, basic concepts and its relevant literature has been discussed
10	[19]	2001	ACM	A detailed review on processes available on 3D modelling and techniques for matching and identifying objects from imagery of free form
11	[20]	2017	Science Direct	Review on list of cutting-edge strategies to recognize objects in a scene applied on two of very famous state of the art dataset
12	[21]	2017	Wiley	A detailed review and explanation on characteristics and principles of algorithms used in recognizing 3D objects that uses point cloud data
13	[22]	2019	Elsevier	A deep review on trends in object recognition algorithms based on foreground detection and instance recognition with its performance measure on different datasets
14	[23]	2019	MDPI	A classic review on performance analysis of object recognition, segmentation and classification methods specifically designed for MLS data obtained from certain benchmark datasets
15	[24]	2019	Springer	A systematic review conducted from articles published in IEEE Explore, ScienceDirect and ACM in between 2006 and 2016
16	An Introduction to Object Recognition	2010	Book	Introductory concepts and algorithms involved in the domain of object recognition

17	Object Recognition	2002	Book	Introduction, database creation, 3D modelling and detailed review of various case studies in vision systems
18	[27]	2020	IEEE	Review on experimental analysis of training free techniques for recognizing objects from RGDB camera
19	[116]	2020	IEEE	A comparative study of scene recognition techniques that's specifically suitable for satellite images
20	[29]	2019	IEEE	Analysis on CNN and YOLOv3 technique for images taken in complex supermarket commodity environment
21	[30]	2014	IEEE	Conduct of extensive experiments on various standard benchmark datasets using MKL for recognizing objects from a visual scene
22	Representations and Techniques for 3D Object Recognition and Scene Interpretation	2011	Book	Introductory and detailed concept explanation on physical space interpretation, 3D object recognition, integrating and interpreting 3D scene
23	[31]	2020	IEEE	After conducting a detailed survey on target recognition, combination of tactile feedback with vision greatly improves the robustness of grasping in robot
24	[32]	2016	IEEE	A survey on techniques to recognize the visual scene in an environment, these techniques will implicitly/explicitly help the robots to navigate
25	[33]	2021	IEEE	Comprehensive study on various open set recognition techniques available along with its definition, datasets, model representation, comparison and evaluation criteria

For a researcher who is new to the field will need a summary of techniques and algorithms that's generalised and suitable for any give application. This paper work aims to do a comprehensive survey of the algorithms related to the domain, its definition and working, evaluation criteria, different benchmark datasets available to apply these algorithms. This will help any researcher who's completely new to the field to identify their own algorithm or field of interest in this vast domain.





1.2 Scope

Looking at the vast amount of innovation and development in the field of object recognition its nearly impossible to do a survey on all the available historic techniques as it would be extensive and difficult to cover. So, this paper will focus on techniques innovated in the last five that has created a huge impact in the field with highest level of accuracy at each point of time. We have filtered the papers that has been published in famous journals and benchmark conferences to have better clarity of available algorithms and techniques. This will help the fellow researchers of this domain to compare and contrast the advantages and short comings of various algorithms discussed. This will also lead us to cite the open challenges and issues to be resolved in the field that might pave way for the young researchers towards future enhancement of the techniques in object recognition.

The rest of the paper has been organised in the following manner. Section 2 describes the problem statement behind Object Recognition algorithms and its challenges, Section 3 describes the evolution on review of literature and state of art techniques in the domain, Section 4 and 5 describes the Convolution Neural Network's working and frameworks following the same, finally Section 6 concludes the significance behind the comprehensive survey on Object Recognition algorithms

2. Object Recognition

2.1 Problem Statement

Object Recognition: Given an image or set of images or set of images generated from frames of video as class of input data, to detect the objects present in the corresponding input, classify it as structured object connected by lines with high dimension (vehicles, faces of human beings, shapes, etc.) or semi-structured objects connected by joints (whole human body, animals, plants etc.) or unstructured class of objects (sky, land, sand, etc.) and finally recognizing it based on its specific features or identity. Computer Vision tasks also has object detection, segmentation and classification along with object recognition. The rest of the tasks are just the extension of the output from object recognition tasks.

In this generic task of recognizing the object in the picture, after recognizing the object, there are several ways in which the recognized object can be highlighted. A rectangular boundary box can be drawn surrounding the object [34, 35], a mask can be generated around the object [36] or pixels of the recognized object can alone be highlighted using different colors. All these tasks give better clarity about the recognized object's spatial coordinates in the picture. The future work in these tasks mainly focus on pixel wise segmentation and mask generation as it gives more accurate spatial coordinates of the target object.

2.2 Challenges

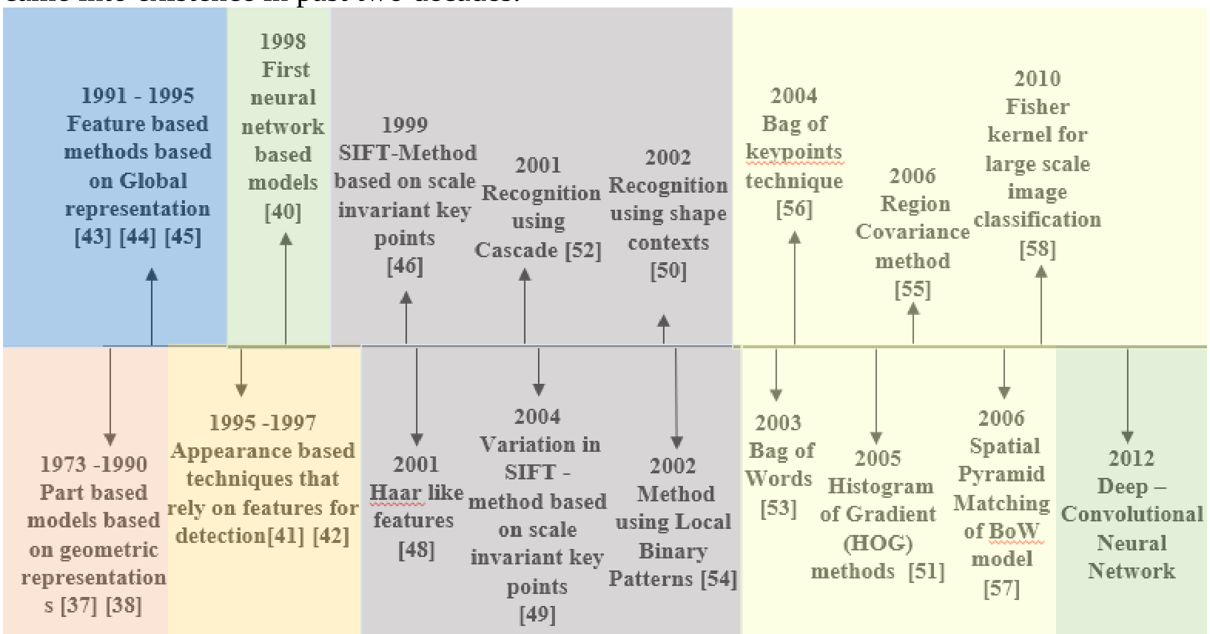
An ideal object recognition algorithm's efficiency can be measured in terms of two criteria in which the first one being High accuracy and the second one being High efficiency. Accuracy can be measured in terms of Localization accuracy and Recognition accuracy in which the former one measures the accuracy of spatial coordinates recognized by the algorithm and the later one measure the accuracy of the target being recognized by the algorithm. To achieve high accuracy is one of the main challenges in object recognition task. To achieve high accuracy the algorithm should be highly robust to various intra class variation. This can be a difficult task because of the following: first one is presence of variations of instances in each class considered whose variation occurs because of difference in texture, shape, color, material etc. Second one is the immense availability of wide range of instances in several categories, Third constraint is due to the conditions and environment in which the input image is taken (Eg. Blurring, Clutter, Illumination, Scaling etc.) and the last one is because of the unnecessary noised that comes along with the dataset. To achieve high accuracy there should also be high focus on distinctiveness of the image being classified. This can be highly affected by the ambiguities that occur in between the interclasses. Also because of the presence of thousands of structured and unstructured data all around us. Efficiency of the algorithm can be measured in terms of time consumption, space consumption and memory usage of the algorithm. Because of thousands to millions of data available in the environment localizing and recognizing the objects become really tiresome. This is due to the fact that there's a need to recognize object and localize their spatial coordinates from large scale image/video. In spite of having good innovative algorithms in the domain of object recognition, there has been no improvement in achieving combined efficiency and accuracy in a good rate. All the past algorithms have been either highly accurate or highly efficient.

	Example for Different Classes of images available in real time
   	Difference in texture, shape, material and colour in same target object i.e., leaf in here
 	Examples for blurred images and clutter image

	Examples for scaled images – the same image has been scaled in three different ways in the given picture
	Image object being giraffe – first picture suffers from too much illumination, second one has too many objects in different size and pose while in the third picture the same object is in motion

3. Review on history of methods

The original history of object recognition techniques started very early, nearly before several decades in 1979 when neocognitron was proposed for the task pattern recognition. It has served as the greatest inspiration behind the discovery of Convolutional Neural Network which became the milestone in the domain of Image Processing and Computer Vision. Before it was recognized as an inspiration to CNN the object recognition algorithms took several turns in between the years 1980 to 2010. As explained in figure there were several inventions and innovations in the techniques in between those years. The evolution began from geometric based models which relied upon simple mathematical shapes and methods for recognition to methods localizing the local features of the object to methods based on global representation. This paper mainly focus on techniques that's based on neural network concept that majorly came into existence in past two decades.



4. Convolutional Neural Network

In the fields of computer vision and image processing, convolutional neural networks are a well-known method for locating, identifying, and detecting objects in any given picture or video. This technique is inspired from the working of receptive field in visual cortex of human being. By getting an input image as a input, CNN takes the input in the form of matrix of pixel values and applies a mathematical concept called convolution operation between the matrix of input pixels and matrix of kernel. The output attained from the operation is further fine-tuned by passing it through a series of convolution and max-pooling layers. The details behind the architecture of CNN is further explained below.

4.1 Kernel and Filter:

Kernel in a CNN represent a series of matrix values that represent any particular geometrical shape (eg. Horizontal line, vertical line, diagonal etc.) or pixel of values used for image reformation (eg. image sharpening kernel, image blurring kernel etc.). These kernels are used to obtain the features in the input image given with the help of convolution operation. The kernel performs the same function as the weight parameter in neural networks. By adjusting the kernel desired features can be extracted from any given input image. Filter on the other hand can be understood as group of kernels put together for convolution operation, in more technical terms concatenation of two or more kernels form a filter. In general kernel is less in dimension when compared with the filter. In common, convolutional neural networks is made up of multiple filters put together with the convolution layers. By differentiating the values of filter, required features can be generated from the given input image.

4.2 Convolution layer

In convolution layer, the input image is taken in the form of matrix of pixels in three channels representing the red, blue and green (primary colours) whose combination makes the whole input image. Then filter matrix is taken and slid through the input image by maintaining a constant stride value at each step. The matrix of pixels obtained from the sum of the product values between the input image and filter marks the first output matrix after convolution. This can be understood in a clear picture using the formula given below.

$$C[m,n] = (i * h)[m,n] = \sum_{1}^{x} * \sum_{1}^{y} * h[j,k]i[m-x][n-y]$$

Where, i represents the input image, h represents the filter, m and n represents the number of rows and columns in the matrix. The dimension of the output matrix after applying convolution depend upon the padding, stride, filter dimensions and input dimensions taken into consideration. Since multiple channels are taken as input, the output matrix will also be in multiple dimension which is represented by below formula.

$$(o,o,nc) = (i,i,nc) * (f,f,nc) = [\frac{n+2p-f}{s} + 1, \frac{n+2p-f}{s} + 1, nf]$$

Where o represents the output, i represents the input, f represents the filter, nc represents the number of channel and nf represents the number of filters.

4.3 ReLU activation

The Rectified Linear Unit activation function comes right after the convolution layer. This activation's primary purpose is to give the entire recognition process some sort of non-linearity. Since convolution operation continuously involves mathematical operations in linear scale, ReLU function balances it by introducing rectification to the final output. f(x) = max(0,x) removes the unnecessary negative values from the final output before giving it to the next pooling layer.

4.4 Pooling layer

The convolution layer's output is given a reduced dimension by making use of a pooling layer. An immediate pooling layer always comes after the convolution layer in order to minimize the dimension and manually choose key attributes of the object included in the input that is being evaluated. Pooling can be done in two ways one being max pooling where maximum pixel is chosen from adjacent pixels in the matrix. The other one being average pooling picks out the average value among the adjacent pixels in the matrix.

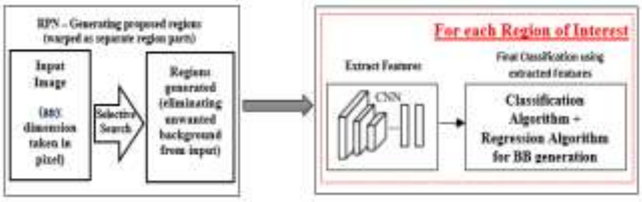
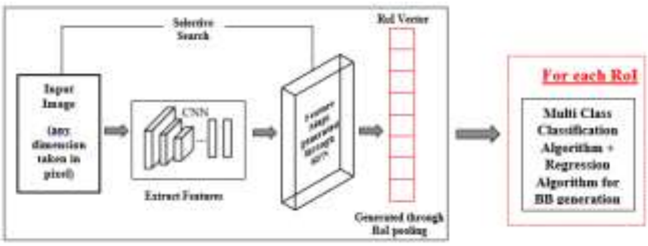
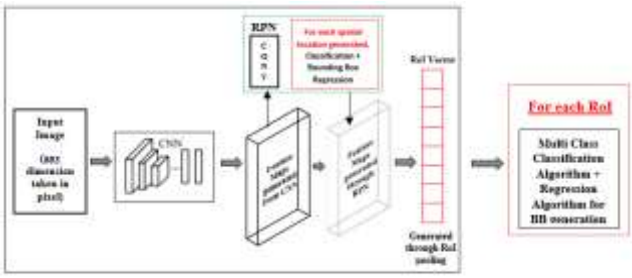
4.5 FC layer

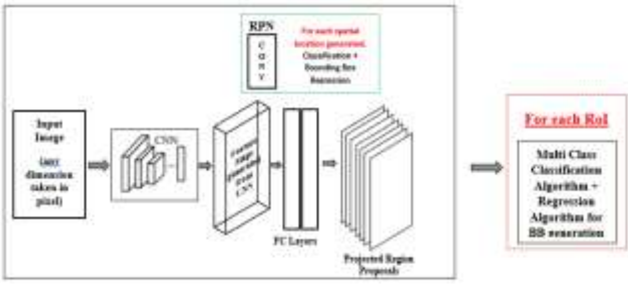
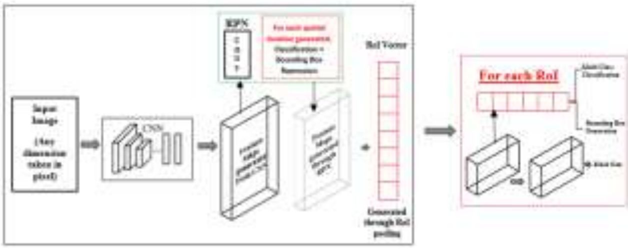
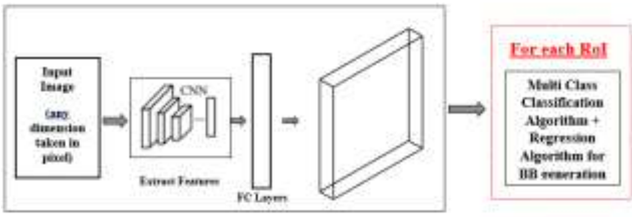
Following a series of convolution operation and pooling operation, the matrix values obtained is flattened to a single column of vector units in which the actual classification is performed.

This densely connected layers work like normal neural network where each unit in the fully connected layer is taken as single unit of neurons and all these neurons contribute to the features of the object present in the image.

5. Popular Frameworks

There are several variations that has been made in the past years in Object Recognition framework pipeline. These variations are made broadly under two categories one is network pipeline based on region proposals and another one is a unified network pipeline. The table shows the most popular frameworks used in both one stage and two stage pipelines. Other than this, there are various small to large variations made throughout that results in wide range of networks. But these architectures results in better accuracy when compared to other frameworks.

No.	**Network Name and Related Works based on it**	**Framework**	**Flow of pipeline**	**Pros and Cons**
1	RCNN – Region Based Convolutional Neural Network Initial Work - [65][66] Improved Version - [67][68][69][70[71]	Two stage		Pro: Candidate regions contains objects most times Con: Slow, hard and expensive training and testing, numerous generated regions makes process very slow
2	Fast RCNN Initial Work - [72] Improved Version - [73][74][75][76][77]	Two stage		Pro: Improved quality and speed of recognition, approximately ten times faster in testing and three times faster in training Con: SS algorithm is still slow and consumes more time
3	Faster RCNN Initial Work - [78][79] Improved Version - [80][81][82][83][84][85][86][87][88][89]	Two stage		Pro: Improved, accurate and efficient RPN for generating regions of interest in a faster rate Con: RPN is applied to several hundreds of RoI which still makes the process time consuming

4	RFCN – Region based Fully Convolutional Neural Network Initial Work - [90] Improved Version - [91][92][93]	Two stage		Pro: RoI subnetwork and crops help in achieving comparable accuracy in faster running time Con: Computationally expensive because of sensitive score map calculation and too many RoI regions.
5	Mask RCNN Initial Work - [94] Improved Version - [95][96][97][98][99][100]	Two stage		Pro: Binary Mask Generation for each and every RoI obtained Con: Works well only in good resolution images without any invariance, blur etc.
6	YOLO – You Only Look Once Initial Work - [101] Improved Version - [102][103][104]	Unified stage		Pro: By design, very fast, runs nearly 45 Frames per Second to 155 Frames per Second Con: At times, predicts False Positive and commonly results in localization error.

Even though a significant number of algorithms are being innovated on a daily basis within the field of Object Recognition, there must be enough input to test the accuracy of those algorithms. Only by testing the algorithm with sufficiently large amount of input images or video, the short comings of the algorithms can be predicted and further innovations can be made. The abundance of data and internet accessibility have led to significant innovation in the creation of datasets for CV algorithms. Some of the very famous datasets with its specification has been mentioned in the Table The performance of the algorithm can be measured by using several performance metrics such as IOU Threshold, Mean Average Precision, Average Precision and Average Recall. All these metrics uses the number of True Positives and False Positives detected by the algorithm with respect to training and testing data to calculate the corresponding algorithm's accuracy.

No.	Dataset Name	Type	Specification
1.	CIFAR10 [105]	Image	60,000 color images in 10 categories
2.	LISA Traffic Sign Detection	Image and Video	7855 annotations on 6610 boundaries
3.	Google Open images	Image	9 million pictures, 16 million bounding boxes, 600 object types
4.	MS COCO	Image	80 classes of objects and 1.5 million instances of objects
5.	Exclusively Dark (ExDark) Image	Image	7,363 low-light images with 12 object types, ranging from extremely low light to twilight
6.	20BN-SOMETHING-SOMETHING Dataset V2	Video	Labelled video clips of people performing basic movements (pre-defined) with variety of objects
7.	ImageNet	Image	Image dataset with 14 million images and 20,000 categories that are classified using the WordNet hierarchy
8.	BDD100K	Video	Uses 100K videos and 10 tasks, to evaluate how far image recognition algorithms have come in autonomous driving.

7. Conclusion

Object Recognition framework has achieved a tremendous improvement in very few decades of time. Wholesome credit goes to the Deep Learning framework which has paved way in reducing the error rate exponentially while the previous algorithms struggled to even reduce it in a linear scale. This survey paper was intended to give a overview on the recent advancements made in the field, its architectures and dataset availability. The inference of this survey has several points that can be promising research filed for future researchers which is explained further. In spite of reduced error rate there is still a huge lag between machine-based recognition and human level recognition. This gap can be filled by proposing more efficient and better detection framework with reduced computation complexity. Most of the algorithms focus only on supervised learning which again needs input in annotated format because of which the capacity of the algorithm stops with available inputs, there are a lot of real-world objects that has not been explored because of the non-availability of annotated data. By moving on to unsupervised learning algorithms this can be avoided. 3D level object recognition can also be explored, given that might be the future and scope of CV algorithms.

This article contains references to the data that were used to support the study's conclusions for future use.

References:

1. Huang, T. (1996-11-19). Vandoni, Carlo, E (ed.). Computer Vision : Evolution And Promise (PDF). 19th CERN School of Computing. Geneva: CERN. pp. 21–25. doi:10.5170/CERN-1996-008.21. ISBN 978-9290830955.
2. A 3D GAN for Improved Large-Pose Facial RecognitionRichard T. Marriott, Sami Romdhani, Liming Chen; Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021, pp. 13445-13455
3. Central Attention and a Dual Path Convolutional Neural Network in Real-World Tree Species Recognition Y Chung, CA Chou, CY Li - … of Environmental Research and Public Health, 2021 - mdpi.com
4. Efficient activity recognition using lightweight CNN and DS-GRU network for surveillance applications A Ullah, K Muhammad, W Ding, V Palade, IU Haq
5. Construction of Statistical SVM based Recognition Model for Handwritten Character Recognition YB Hamdan - Journal of Information Technology, 2021 - irojournals.com

6. Using Agents and Unsupervised Learning for Counting Objects in Images with Spatial Organization. E Jacopin, N Berda, L Courteille, W Grison, L Mathieu… - ICAART (2), 2021 - scitepress.org
7. Occupancy grid mapping with cognitive plausibility for autonomous driving applications A Plebe, JFP Kooij, GPR Papini… - Proceedings of the …, 2021 - openaccess.thecvf.com
8. Automatic image annotation for fluorescent cell nuclei segmentation F Englbrecht, IE Ruider, AR Bausch - PloS one, 2021 - journals.plos.org
9. A Lightweight Fully Convolutional Neural Network for SAR Automatic Target Recognition J Yu, G Zhou, S Zhou, J Yin - Remote Sensing, 2021 - mdpi.com
10. Efficient Search in a Panoramic Image Database for Long-Term Visual Localization S Orhan, Y Bastanlar - Proceedings of the IEEE/CVF …, 2021 - openaccess.thecvf.com
11. Research on Visual Positioning of a Roadheader and Construction of an Environment Map W Zhang, G Zhai, Z Yue, T Pan, R Cheng - Applied Sciences, 2021 - mdpi.com
12. Survey of appearance-based methods for object recognition PM Roth, M Winter - Inst. for computer graphics and vision, Graz …, 2008 - msit.mut.ac.th
13. Object recognition using deep neural networks: A survey S Goyal, P Benjamin - arXiv preprint arXiv:1412.3684, 2014
14. 3D object recognition in cluttered scenes with local surface features: A survey Y Guo, M Bennamoun, F Sohel, M Lu… - IEEE transactions on …, 2014
15. Object recognition using local invariant features for robotic applications: A survey P Loncomilla, J Ruiz-del-Solar, L Martínez - Pattern Recognition, 2016 – Elsevier
16. 50 years of object recognition: Directions forward A Andreopoulos, JK Tsotsos - Computer vision and image understanding, 2013 - Elsevier
17. Model-based object recognition in dense-range images—a review F Arman, JK Aggarwal - Acm Computing Surveys (CSUR), 1993
18. Three-dimensional object recognition PJ Besl, RC Jain - ACM Computing Surveys (CSUR), 1985
19. A survey of free-form object representation and recognition techniques RJ Campbell, PJ Flynn - Computer Vision and Image Understanding, 2001 – Elsevier
20. A survey on learning approaches for Undirected Graphical Models. Application to scene object recognition JR Ruiz-Sarmiento, C Galindo… - International Journal of …, 2017
21. Review of multi-view 3D object recognition methods based on deep learning S Qi, X Ning, G Yang, L Zhang, P Long, W Cai, W Li - Displays, 2021 – Elsevier
22. Adversarial examples on object recognition: A comprehensive survey A Serban, E Poll, J Visser - ACM Computing Surveys (CSUR), 2020 - dl.acm.org
23. A survey on learning approaches for Undirected Graphical Models. Application to scene object recognition JR Ruiz-Sarmiento, C Galindo… - International Journal of …, 2017
24. A Review of Recent Advancements in Appearance-based Object Recognition GP Kusuma, EK Wigati, E Chandra - Procedia Computer Science, 2019 – Elsevier
25. Object recognition, segmentation, and classification of mobile laser scanning point clouds: A state of the art review E Che, J Jung, MJ Olsen - Sensors, 2019
26. 3D object recognition and classification: a systematic literature review LE Carvalho, A Von Wangenheim - Pattern Analysis and Applications, 2019
27. A Survey on Training Free 3D Texture-less Object Recognition Techniques Piyush Joshi;Alireza Rastegarpanah;Rustam Stolkin 2020 Digital Image Computing: Techniques and Applications (DICTA)
28. Object Recognition and Detection in Remote Sensing Images: A Comparative Study Syed Aley Fatima;Ashwani Kumar;Ajay Pratap;Syed Saba Raoof 2020 International Conference on Artificial Intelligence and Signal Processing (AISP)

29. Research on Image Recognition of Supermarket Commodity Based on Convolutional Neural Network Yanmei Liu;Ben Liu;Yuda Chen 2019 12th International Symposium on Computational Intelligence and Design (ISCID)
30. Multiple Kernel Learning for Visual Object Recognition: A Review Serhat S. Bucak;Rong Jin;Anil K. Jain IEEE Transactions on Pattern Analysis and Machine Intelligence
31. Object Detection Recognition and Robot Grasping Based on Machine Learning: A Survey Qiang Bai;Shaobo Li;Jing Yang;Qisong Song;Zhiang Li;Xingxing Zhang IEEE Access
32. Visual Place Recognition: A Survey Stephanie Lowry;Niko Sünderhauf;Paul Newman;John J. Leonard;David Cox;Peter Corke;Michael J. Milford IEEE Transactions on Robotics
33. Recent Advances in Open Set Recognition: A Survey Chuanxing Geng;Sheng-Jun Huang;Songcan Chen IEEE Transactions on Pattern Analysis and Machine Intelligence
34. Everingham M., Gool L. V., Williams C., Winn J., Zisserman A. (2010) The pascal visual object classes (voc) challenge. IJCV 88(2):303–338
35. Russakovsky O., Deng J., Su H., Krause J., Satheesh S., Ma S., Huang Z., Karpathy A., Khosla A., Bernstein M., Berg A., Li F. (2015) ImageNet large scale visual recognition challenge. IJCV 115(3):211–252
36. He K., Gkioxari G., Dollar P., Girshick R. (2017) Mask RCNN. In: ICCV
37. Fischler M., Elschlager R. (1973) The representation and matching of pictorial structures. IEEE Transactions on computers 100(1):67–92
38. Mundy J. (2006) Object recognition in the geometric era: A retrospective. in book Toward Category Level Object Recognition edited by J Ponce, M Hebert, C Schmid and A Zisserman pp. 3–28
39. Rowley H., Baluja S., Kanade T. (1998) Neural network-based face detection. IEEE TPAMI 20(1):23–38
40. Rowley H., Baluja S., Kanade T. (1998) Neural network based face detection. IEEE TPAMI 20(1):23–38
41. Murase H., Nayar S. (1995) Visual learning and recognition of 3D objects from appearance. IJCV 14(1):5–24
42. Schmid C., Mohr R. (1997) Local grayvalue invariants for image retrieval. IEEE TPAMI 19(5):530–535
43. Murase H., Nayar S. (1995) Visual learning and recognition of 3d objects from appearance. IJCV 14(1):5–24
44. Swain M., Ballard D. (1991) Color indexing. IJCV 7(1):11–32
45. Turk M. A., Pentland A. (1991) Face recognition using eigenfaces. In: CVPR, pp. 586–591
46. Lowe D. (1999) Object recognition from local scale invariant features. In: ICCV, vol 2, pp. 1150–1157
47. Mikolajczyk K., Schmid C. (2005) A performance evaluation of local descriptors. IEEE TPAMI 27(10):1615–1630
48. Viola P., Jones M. (2001) Rapid object detection using a boosted cascade of simple features. In: CVPR, vol 1, pp. 1–8
49. Lowe D. (2004) Distinctive image features from scale-invariant keypoints. IJCV 60(2):91–110
50. Belongie S., Malik J., Puzicha J. (2002) Shape matching and object recognition using shape contexts. IEEE TPAMI 24(4):509–522
51. Dalal N., Triggs B. (2005) Histograms of oriented gradients for human detection. In: CVPR, vol 1, pp. 886–893
52. Viola P., Jones M. (2001) Rapid object detection using a boosted cascade of simple features. In: CVPR, vol 1, pp. 1–8
53. Sivic J., Zisserman A. (2003) Video google: A text retrieval approach to object matching in videos. In: International Conference on Computer Vision (ICCV), vol 2, pp. 1470–1477

54. Ojala T., Pietikainen M., Maenp ¨ a¨a T. (2002) Multiresolution gray-scale ¨ and rotation invariant texture classification with local binary patterns. IEEE TPAMI 24(7):971–987
55. Tuzel O., Porikli F., Meer P. (2006) Region covariance: A fast descriptor for detection and classification. In: ECCV, pp. 589–600
56. Csurka G., Dance C., Fan L., Willamowski J., Bray C. (2004) Visual categorization with bags of keypoints. In: ECCV Workshop on statistical learning in computer vision
57. Lazebnik S., Schmid C., Ponce J. (2006) Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In: CVPR, vol 2, pp. 2169–2178
58. Perronnin F., Sanchez J., Mensink T. (2010) Improving the fisher kernel ´ for large scale image classification. In: ECCV, pp. 143–156
59. Fukushima, K.: Self-organization of a neural network which gives position-invariant response. In: Proceedings of the Sixth International Joint Conference on Artificial Intelligence. Tokyo, August 20–23, 1979, pp. 291–293 (1979b)
60. Marius Paraschiv., Ricardo Padrino., Paolo Casari., Eyal Bigal., Aviad Scheinin., Dan Tchernov and Antonio Fernández Anta Classification of Underwater Fish Images and Videos via Very Small Convolutional Neural Networks. In J. Mar. Sci. Eng. 2022, 10, 736.
61. Andrzej Szczurek., Andi Azizah and Monika Maciejewska The Detection of Activities Occurring Inside Quick Service Restaurants That Influence Air Quality Sensors 2022, 22, 4056.
62. Xiang, R.; Zhang, M.; Zhang, J Recognition for Stems of Tomato Plants at Night Based on a Hybrid Joint Neural Network Agriculture 2022, 12, 743.
63. Rodellar, J.; Barrera, K.; Alférez, S.; Boldú, L.; Laguna, J.; Molina, A.; Merino, A. A Deep Learning Approach for the Morphological Recognition of Reactive Lymphocytes in Patients with COVID-19 Infection. Bioengineering 2022, 9, 229.
64. Lee, M.-F.R.; Lin, C.-Y. Object Tracking for an Autonomous Unmanned Surface Vehicle. Machines 2022, 10, 378.
65. Girshick R., Donahue J., Darrell T., Malik J. (2014) Rich feature hierarchies for accurate object detection and semantic segmentation. In: CVPR, pp. 580–587
66. Girshick R., Donahue J., Darrell T., Malik J. (2016) Region-based convolutional networks for accurate object detection and segmentation. IEEE TPAMI 38(1):142–158
67. Li Z., Peng C., Yu G., Zhang X., Deng Y., Sun J. (2018) Light head RCNN: In defense of two stage object detector. In: CVPR
68. D. Cai, X. Sun, N. Zhou, X. Han and J. Yao, "Efficient Mitosis Detection in Breast Cancer Histology Images by RCNN," *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, 2019, pp. 919-922, doi: 10.1109/ISBI.2019.8759461.
69. Y. Li *et al.*, "A Joint Model of Clinical Domain Classification and Slot Filling Based on RCNN and BiGRU-CRF," *2019 IEEE International Conference on Big Data (Big Data)*, 2019, pp. 6133-6135, doi: 10.1109/BigData47090.2019.9005449.
70. B. Sridevi and A. Elumalai, "Evaluation of RCNN parameters for Senile dementia diagnosis using ADNI dataset," 2021 International Conference on Recent Trends on Electronics, Information, Communication & Technology (RTEICT), 2021, pp. 87-90, doi: 10.1109/RTEICT52294.2021.9573969.
71. N. Yao, G. Shan and X. Zhu, "Substation Object Detection Based on Enhance RCNN Model," 2021 6th Asia Conference on Power and Electrical Engineering (ACPEE), 2021, pp. 463-469, doi: 10.1109/ACPEE51499.2021.9437086.
72. Girshick R. (2015) Fast R-CNN. In: ICCV, pp. 1440–1448
73. K. S. Htet and M. M. Sein, "Event Analysis for Vehicle Classification using Fast RCNN," 2020 IEEE 9th Global Conference on Consumer Electronics (GCCE), 2020, pp. 403-404, doi: 10.1109/GCCE50665.2020.9291978.
74. L. Quan, D. Pei, B. Wang and W. Ruan, "Research on Human Target Recognition Algorithm of Home Service Robot Based on Fast-RCNN," 2017 10th International

Conference on Intelligent Computation Technology and Automation (ICICTA), 2017, pp. 369-373, doi: 10.1109/ICICTA.2017.88.

75. A. Ullah, H. Xie, M. O. Farooq and Z. Sun, "Pedestrian Detection in Infrared Images Using Fast RCNN," 2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA), 2018, pp. 1-6, doi: 10.1109/IPTA.2018.8608121.
76. L. Sommer, N. Schmidt, A. Schumann and J. Beyerer, "Search Area Reduction Fast-RCNN for Fast Vehicle Detection in Large Aerial Imagery," 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 3054-3058, doi: 10.1109/ICIP.2018.8451189.
77. K. Wang, Y. Dong, H. Bai, Y. Zhao and K. Hu, "Use fast R-CNN and cascade structure for face detection," 2016 Visual Communications and Image Processing (VCIP), 2016, pp. 1-4, doi: 10.1109/VCIP.2016.7805472.
78. Ren S., He K., Girshick R., Sun J. (2015) Faster R-CNN: Towards real time object detection with region proposal networks. In: NIPS, pp. 91– 99
79. Ren S., He K., Girshick R., Sun J. (2017) Faster RCNN: Towards real time object detection with region proposal networks. IEEE TPAMI 39(6):1137–1149
80. W. Zhao, M. Xu, X. Cheng and Z. Zhao, "An Insulator in Transmission Lines Recognition and Fault Detection Model Based on Improved Faster RCNN," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-8, 2021, Art no. 5016408, doi: 10.1109/TIM.2021.3112227.
81. Q. Xu, X. Zhang, R. Cheng, Y. Song and N. Wang, "Occlusion Problem-Oriented Adversarial Faster-RCNN Scheme," in IEEE Access, vol. 7, pp. 170362-170373, 2019, doi: 10.1109/ACCESS.2019.2955685.
82. Z. Zhou, Y. Cao, M. Wang, E. Fan and Q. M. J. Wu, "Faster-RCNN Based Robust Coverless Information Hiding System in Cloud Environment," in IEEE Access, vol. 7, pp. 179891-179897, 2019, doi: 10.1109/ACCESS.2019.2955990.
83. Z. He, P. Peng, L. Wang and Y. Jiang, "Enhancing Seismic P-Wave Arrival Picking by Target-Oriented Detection of the Local Windows Using Faster-RCNN," in IEEE Access, vol. 8, pp. 141733-141747, 2020, doi: 10.1109/ACCESS.2020.3013262.
84. A. Pramanik, S. K. Pal, J. Maiti and P. Mitra, "Granulated RCNN and Multi-Class Deep SORT for Multi-Object Detection and Tracking," in IEEE Transactions on Emerging Topics in Computational Intelligence, vol. 6, no. 1, pp. 171-181, Feb. 2022, doi: 10.1109/TETCI.2020.3041019.
85. F. Li et al., "An Automatic Detection Method of Bird’s Nest on Transmission Line Tower Based on Faster_RCNN," in IEEE Access, vol. 8, pp. 164214-164221, 2020, doi: 10.1109/ACCESS.2020.3022419.
86. Y. Zhao, W. Cui, S. Geng, B. Bo, Y. Feng and W. Zhang, "A Malware Detection Method of Code Texture Visualization Based on an Improved Faster RCNN Combining Transfer Learning," in IEEE Access, vol. 8, pp. 166630-166641, 2020, doi: 10.1109/ACCESS.2020.3022722.
87. Z. Liu, Y. Lyu, L. Wang and Z. Han, "Detection Approach Based on an Improved Faster RCNN for Brace Sleeve Screws in High-Speed Railways," in IEEE Transactions on Instrumentation and Measurement, vol. 69, no. 7, pp. 4395-4403, July 2020, doi: 10.1109/TIM.2019.2941292.
88. C. Dai et al., "Video Scene Segmentation Using Tensor-Train Faster-RCNN for Multimedia IoT Systems," in IEEE Internet of Things Journal, vol. 8, no. 12, pp. 9697-9705, 15 June15, 2021, doi: 10.1109/JIOT.2020.3022353.
89. D. Zhang, J. Li, L. Xiong, L. Lin, M. Ye and S. Yang, "Cycle-Consistent Domain Adaptive Faster RCNN," in IEEE Access, vol. 7, pp. 123903-123911, 2019, doi: 10.1109/ACCESS.2019.2938837.
90. Dai J., Li Y., He K., Sun J. (2016) RFCN: object detection via region based fully convolutional networks. In: NIPS, pp. 379–387

91. Y. Du and Q. Wang, "Multi-Angle Face Detection Based on Improved RFCN Algorithm Using Multi-Scale Training," 2021 6th International Conference on Intelligent Computing and Signal Processing (ICSP), 2021, pp. 319-322, doi: 10.1109/ICSP51882.2021.9408676.
92. A. Masood, B. Sheng, P. Li, P. Yang and J. Kim, "Automatic Computer Aided System for Lung Cancer in Chest CTs Using MD-RFCN Combined with Tri-Level Region Proposal Network," 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), 2019, pp. 1377-1380, doi: 10.1109/INDIN41052.2019.8972133.
93. A. Masood et al., "Automated Decision Support System for Lung Cancer Detection and Classification via Enhanced RFCN With Multilayer Fusion RPN," in IEEE Transactions on Industrial Informatics, vol. 16, no. 12, pp. 7791-7801, Dec. 2020, doi: 10.1109/TII.2020.2972918.
94. He K., Gkioxari G., Dollar P., Girshick R. (2017) Mask RCNN. In: ICCV
95. Q. Zhang, X. Chang and S. B. Bian, "Vehicle-Damage-Detection Segmentation Algorithm Based on Improved Mask RCNN," in IEEE Access, vol. 8, pp. 6997-7004, 2020, doi: 10.1109/ACCESS.2020.2964055.
96. J. Park and H. Moon, "Lightweight Mask RCNN for Warship Detection and Segmentation," in IEEE Access, vol. 10, pp. 24936-24944, 2022, doi: 10.1109/ACCESS.2022.3149297.
97. Z. Zhou, M. Zhang, J. Chen and X. Wu, "Detection and Classification of Multi-Magnetic Targets Using Mask-RCNN," in IEEE Access, vol. 8, pp. 187202-187207, 2020, doi: 10.1109/ACCESS.2020.3030676.
98. B. Wang, Y. Zhao and C. L. P. Chen, "Hybrid Transfer Learning and Broad Learning System for Wearing Mask Detection in the COVID-19 Era," in IEEE Transactions on Instrumentation and Measurement, vol. 70, pp. 1-12, 2021, Art no. 5009612, doi: 10.1109/TIM.2021.3069844.
99. C. Tao, Z. Gao, J. Yan, C. Li and G. Cui, "Indoor 3D Semantic Robot VSLAM Based on Mask Regional Convolutional Neural Network," in IEEE Access, vol. 8, pp. 52906-52916, 2020, doi: 10.1109/ACCESS.2020.2981648.
100. J. Shi, Y. Zhou and W. X. Q. Zhang, "Target Detection Based on Improved Mask Rcnn in Service Robot," 2019 Chinese Control Conference (CCC), 2019, pp. 8519-8524, doi: 10.23919/ChiCC.2019.8866278.
101. Redmon J., Divvala S., Girshick R., Farhadi A. (2016) You only look once: Unified, real-time object detection. In: CVPR, pp. 779–788
102. Redmon J., Farhadi A. (2017) YOLO9000: Better, faster, stronger. In: CVPR
103. M. Mahendru and S. K. Dubey, "Real Time Object Detection with Audio Feedback using Yolo vs. Yolo_v3," 2021 11th International Conference on Cloud Computing, Data Science & Engineering (Confluence), 2021, pp. 734-740, doi: 10.1109/Confluence51648.2021.9377064.
104. T. -H. Wu, T. -W. Wang and Y. -Q. Liu, "Real-Time Vehicle and Distance Detection Based on Improved Yolo v5 Network," 2021 3rd World Symposium on Artificial Intelligence (WSAI), 2021, pp. 24-28, doi: 10.1109/WSAI51899.2021.9486316.
105. Krizhevsky A. (2009) Learning multiple layers of features from tiny images. Master's thesis, University of Toronto
106. A. S. Lundervold and A. Lundervold, "An overview of deep learning in medical imaging focusing on MRI", *Zeitschrift für Med. Phys.*, vol. 29, no. 2, pp. 102-127, May 2019.
107. S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami and M. K. Khan, "Medical image analysis using convolutional neural networks: A review", *J. Med. Syst.*, vol. 42, no. 11, pp. 226, Nov. 2018.
108. P. Lakhani, D. L. Gray, C. R. Pett, P. Nagy and G. Shih, "Hello world deep learning in medical imaging", *J. Digit. Imag.*, vol. 31, no. 3, pp. 283-289, Jun. 2018

109. S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Alnowami and M. K. Khan, "Medical image analysis using convolutional neural networks: A review", *J. Med. Syst.*, vol. 42, no. 11, pp. 226, Nov. 2018.
110. H. Gao, X. Zeng and C. Yao, "Application of improved distributed naive Bayesian algorithms in text classification", *J. Supercomput.*, vol. 75, no. 9, pp. 5831-5847, Sep. 2019.
111. A. Saez, J. Sanchez-Monedero, P. A. Gutierrez and C. Hervas-Martinez, "Machine learning methods for binary and multiclass classification of melanoma thickness from dermoscopic images", *IEEE Trans. Med. Imag.*, vol. 35, no. 4, pp. 1036-1045, Apr. 2016.
112. A. Rehman, S. Naz, M. I. Razzak, F. Akram and M. Imran, "A deep learning-based framework for automatic brain tumors classification using transfer learning", *Circuits Syst. Signal Process.*, vol. 39, no. 2, pp. 757-775, Dec. 2013.
113. H. Irshad, A. Veillard, L. Roux and D. Racoceanu, "Methods for nuclei detection segmentation and classification in digital histopathology: A review—Current status and future potential", *IEEE Rev. Biomed. Eng.*, vol. 7, pp. 97-114, 2014.
114. A. Brunetti, L. Carnimeo, G. F. Trotta and V. Bevilacqua, "Computer-assisted frameworks for classification of liver breast and blood neoplasias via neural networks: A survey based on medical images", *Neurocomputing*, vol. 335, pp. 274-298, Mar. 2019.
115. T. Hussain, K. Muhammad, S. Khan, A. Ullah, M. Y. Lee and S. W. Baik, "Intelligent baby behavior monitoring using embedded vision in IoT for smart healthcare centers", *J. Artif. Intell. Syst.*, vol. 1, no. 15, pp. 2019, 2019.
116. S. A. Fatima, A. Kumar, A. Pratap and S. S. Raoof, "Object Recognition and Detection in Remote Sensing Images: A Comparative Study," 2020 International Conference on Artificial Intelligence and Signal Processing (AISP), Amaravati, India, 2020, pp. 1-5, doi: 10.1109/AISP48273.2020.9073614.