Quest Journals
Journal of Electronics and Communication Engineering Research
Volume 8 ~ Issue 9 (2022) pp: 13-24
ISSN(Online) : 2321-5941
www.questjournals.org

Research Paper

A Review on Image Classification and Object Detection Using Artificial Neural Networks

K D Mohana Sundaram[1], T Shankar[2], N Sudhakar Reddy[3]
[1]*Research Scholar, Department of ECE, Annamalai University, Chidambaram, Tamil Nadu, India.*
[2]*Assistant Professor, Department of ECE, Government college of Engineering (Deputed from Annamalai University), Srirangam, Trichy, Tamil Nadu, India.*
[3]*Professor, Department of ECE, M.J.R.College of Engineering & Technology, Piler, Andhra Pradesh, India.*

Abstract- *This paper deals with all artificial neural networks that have been implemented for object detection in the past decade. Although there is extensive work on object Detection based on ANN [1], we have made a few selections. For this reason, only articles published by peer reviewed journals are taken into account. In addition to magazine articles, this paper also contains good conference contributions to ANN for object detection. All articles are grouped according to implementation techniques and further divided into sub-techniques. Our main goal is to provide every detail related to ANN for object Detection so that this paper is very helpful for readers who are looking for the complete literature on ANN for object Detection.*

Index Terms— *ANN-Artificial Neural Networks, Object Detection.*

R*eceived 12 Sep, 2022;* ***R****evised 26 Sep., 2022;* ***A****ccepted 28 Sep., 2022 © The author(s) 2022.*
P*ublished with open access at* ***www.questjournals.org***

I. INTRODUCTION

This review is completely different from previous reviews because it only deals with articles that relate to ANN object detection schemes. It covers implementation techniques used by ANN researchers for object Detection, as well as articles published by well-known publishers of international journals. All research has been broken down into the number of groups, and each group contains only the work dealing with the same techniques. Each group is worked out in detail and at the end of each group a short summary method is described, which covers important parts of their work. Finally, this review concludes with some existing challenges in implementing the ANN for object Detection. Finally, it shows the future trend and the limits. As in previous years, significant efforts have been made to investigate research in traditional applications such as Mobile Vision Application [39], Image Detection [37] and Large Format Image Detection [36], Small Object Detection [26] and Image Segmentation [14], real-time Object detection [16] etc. The proposed review included all papers including their scope. The following table: 1 is an index showing which main points for describing the ANN for object Detection are dealt with.

II. ARTIFICIAL NEURAL NETWORKS

The human brain is the connection of a large number of biological neurons, as shown in Fig. 1. The human brain is one of the most complicated things. The brain consists of approximately 10^{10} neurons. Each neuron is connected to around 10^4 other neurons. A neuron is a small cell that receives an electrochemical signal as an input from its various sources and then sends electrical impulses to other neurons.

Table 1: Index

I	Introduction
II	ANN
III	Characteristics and Application of ANN
IV	Object Detection Using ANNs
V	Classification of ANN for Object Detection
VI	Implementation Techniques of ANN

A neuron is made up of a nucleus - a cell body known as soma. Long, irregularly shaped filaments, known as dendrites, are attached to soma as shown in Fig. 2. The dendrites behave like input channels. Another type of connection associated with the soma is the axon. The axon is electrically active and serves as an output channel. The axon ends in a special contact called Synapse, which connects the axon to the dendrite connection of another neuron. A neuron receives an electrochemical signal as input, simply sums up the input and produces an output.

The behaviour of a biological neuron can be determined by a simple model, as shown in Fig. 3. Each component of the model has a direct analogy to the actual components of a biological neuron and is therefore referred to as an artificial neuron. It is the model that forms the basis for artificial neural networks (ANN) [2]. The word ANN refers to the connections between the artificial neurons in the different layers of each system. An example system consists of three layers. The first layer has input neurons that send data to the second layer of neurons via synapses and then to the third layer of output neurons via further synapses. More complex systems have more layers of neurons, with some having more layers of input and output neurons. The synapses store parameters called "weights" that manipulate the data in the calculations. An ANN is usually defined by three types of parameters:

1.The interconnection pattern between the different layers of neurons.
2. The learning process for updating the weights of the interconnections.
3. The activation function that converts a neuron's weighted input to its output activation.

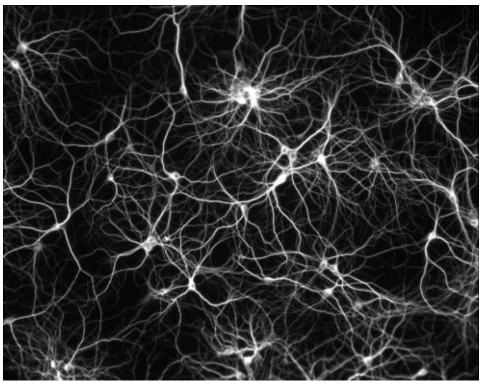

Fig 1: Neural Network in human brain

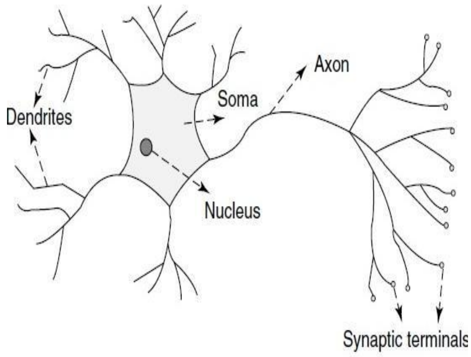

Fig 2: Biological Neuron

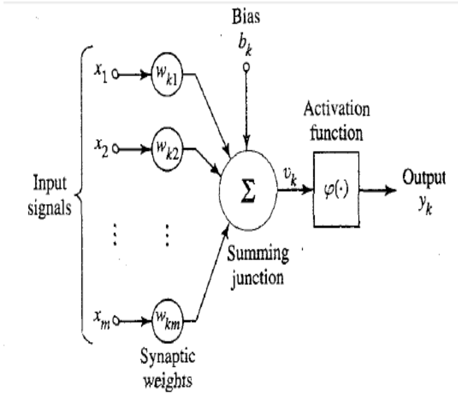

Fig 3: Artificial Neuron

III. CHARACTERISTICS AND APPLICATION OF ANN

An ANN consists of a large number of neuron-like processing elements. All of these processing elements have a large number of weighted connections between them. The connection between the elements offers a distributed representation of data. A learning process [2] is implemented to acquire knowledge. ANNs learn using examples. The ANN architecture can thus be trained using known examples. Therefore, they can identify the new objects that have not been trained before. ANN can process information in parallel, at high speed and in a distributed manner.

The two most important areas in which ANNs have great application potential are Speech and Image processing. In speech processing, ANNs are used for vowel classification, the Detection of vowel-consonant segments, the Detection of stop-consonant-vowel utterances in Indian languages, Net Talk, Phonetic Type Writer etc. In image processing, ANNs are used to recognize symbols, to recognize handwriting, to recognize objects in an image, to segment images, to classify and segment textures, etc. The other general areas in which ANNs can be used are art, bioinformatics, forecasting, healthcare, intrusion detection, communication, robotics and pattern Detection.

IV. OBJECT DETECTION USING ANNs

Object Detection is a computer vision technique for locating instances of objects in images or videos. Object detection algorithms can be implemented either in machine learning or in deep learning to achieve meaningful results. When people look at pictures or videos, we can recognize and localize objects of interest within a few moments. The goal of object detection is to replicate this intelligence using a computer. A variety of techniques can be used to perform object Detection. Popular deep learning-based approaches using convolutional neural networks (CNNs) [5] such as R-CNN [14] and YOLO v2 [30] automatically learn to recognize objects in images.

There are two important approaches to start object detection using deep learning:

• *Create and train a custom object detector*. To train a custom object detector from scratch, the network architecture must be designed to learn the functions for the most interesting objects. In addition, a very large set of labeled data must be compiled to train the CNN. The results of a custom object detector can be remarkable. That means you have to manually set up the levels and weights in the CNN, which takes a lot of time and training data.

• *Use a pretrained object detector.* Many object Detection workflows use deep learning leverage transfer learning, an approach that allows you to start with a pre-trained network and then optimize it for the right application. This method can deliver faster results because the object detectors have already been trained on thousands or even millions of images.

V. CLASSIFICATION OF ANNs FOR OBJECT DETECTION

The ANN techniques for object Detection can be classified as shown in Fig 4. All research work that deals with object Detection is grouped and further divided according to the implementation techniques used. The first group is object Detection based on the Convolutional Neural Network (CNN). The second group is based on regional CNN (R-CNN). The third group is based on a single stage network. The fourth group is based on very deep CNN.

VI. IMPLEMENTATION TECHNIQUES OF ANNs

Object detection is a key technology behind Advanced Driver Assistance Systems (ADAS) that allow cars to recognize lanes or pedestrians to improve traffic safety. Object detection is also useful in applications such as video surveillance or image recovery systems. Many object detection algorithms based on ANN have been proposed in recent years and have achieved good results. The important unit for measuring the success rate of these algorithms is the top 1 and top 5 error rate.

Top 1 error rate: Defined as a percentage of the time that the classifier did not give the correct class the highest probability rating.

Top 5 Error Rate: The percentage of time that the classifier did not include the correct class among the top 5 probabilities or guesswork.

VI A. Convolutional Neural Network (CNN)

Keiron O'Shea et al. [5] described a brief introduction to CNNs and discussed recently published articles and newly developed techniques for developing these brilliantly fantastic image Detection models. This introduction will familiarize you with the basics of ANNs and machine learning.

Alex Krizhevsky et al. [8] used a Deep CNN to divide the 1.2 million high-resolution images in the ImageNet LSVRC 2010 competition into 1000 different classes. They achieved a top 1 and top 5 error rate of 37.5% and 17.0%, which is significantly better than in the previous work. They used five folding layers, some of which are followed by max pooling layers and three fully bonded layers with a final 1000-way softmax. To reduce the over fitting, they developed a regularization method called Dropout, which turned out to be very effective.

Hideaki Yanagisawa et al. [10] compared different CNN-based object Detection algorithms for object Detection in manga images. Manga images are popular Japanese comic book content. In order to obtain metadata from manga images, techniques for automatic detection of the manga content were examined. The Convolutional Neural Network (CNN) was recently used for object Detection in manga images. Their experimental results show that CNN is effective for object Detection in manga images.

Yiming Zhang et al. [12] described a real-time object Detection system for 360-degree panoramic images using a convolutional neural network (CNN). You have introduced a CNN-based Detection framework for object Detectionwith a post-processing phase to fine-tune the result. In addition, they proposed a new method to extract the existing data sets of ordinary images, e.g. B. ImageNet and PASCAL VOC to reuse in object Detection for 360-degree panoramic images. They achieved higher accuracy and recall rates than traditional object detection methods for 360-degree panoramic images. The summarized work of researchers on object Detection based on CNN is described in Table 2.

VI B. (i) Region Based Convolutional Neural Network (R-CNN)

Ross Girshick et al. [14] proposed a simple and scalable detection algorithm that improves the mean average accuracy (mAP) by more than 30% compared to the previous best result of the PASCAL VOC data set and achieves a mAP of 53.3%. Their approach combines two important insights: (1) high-capacity convolutional neural networks (CNNs) can be applied to bottom-up region suggestions to locate and segment objects, and (2) if pre-labeled training data is scarce, monitored pre -Training For an auxiliary task, followed by a domain-specific fine-tuning, there is a significant performance.

Joshua Herrera et al. [16] proposed a distributed edge cloud RCNN pipeline. By dividing the object detection pipeline into components and dynamically distributing these components in the cloud, optimal performance was achieved to enable object detection in real time. As a proof of concept, they evaluated the performance of the proposed system on a distributed computer platform, including cloud servers and edge embedded devices, for real-time object detection in live video streams.

(ii). Fast R-CNN

Ross Girshick et al. [17] proposed a Fast Region-based Convolutional Network method (Fast R-CNN) for object Detection. Fast R-CNN builds on earlier work on the efficient classification of object proposals using deep convolution networks. Compared to previous work, Fast R-CNN used various innovations to improve the training and test speed while increasing the detection accuracy. Fast R-CNN trains the very deep VGG16 network 9 × faster than R-CNN, is 213 × faster at test time and achieves a higher mAP on PASCAL VOC 2012. Compared to SPPnet, fast R-CNN VGG16 trains 3 × faster, tests 10 × faster and more accurate.

(iii). Faster R-CNN

Shaoqing Ren et al. [18] introduced a Region Proposal Network (RPN), which shares full-screen convolution functions with the Detection network and enables almost free region suggestions. An RPN is a fully folded network that simultaneously predicts object boundaries and objectivity values at any position. The RPN is trained consistently to generate high quality region suggestions that Fast R-CNN used for Detection. They have merged RPN and Fast R-CNN into a single network by sharing their convolutional features. For the very deep VGG-16 model, the Detection system had a frame rate of 5 fps on a GPU, while with PASCAL VOC 2007, 2012 and MS COCO data sets, with only 300 suggestions per object, a highly modern object Detection accuracy was achieved. In the ILSVRC and COCO 2015 competitions, Faster R-CNN and RPN won first place on several routes.

Syed Mazhar Abbas et al. [20] focused on improving the R-CNN, Fast R-CNN and YOLO architecture. The work focused on using the Region Proposals Network (RPN) to extract the region of interest in an image. RPN outputs an image based on the property rating. The output objects are subjected to roll polling for classification. Her research focused on training Faster R-CNN using a custom set of images. Her trained network efficiently recognized objects from an image that consisted of several objects. Your network requires a minimum GPU capacity of 3.0 or higher.

Irina Mocanu et al. [21] presented a new convolution architecture for neural networks to perform object Detection based on RGB-D images. The network is an extension of the Faster RCNN network, in which an additional input network branch has been added to process the depth image. The network was evaluated based on the SUN RGB-D data set for object Detection and a positive difference in the mAP score of around 4% compared to the original value was achieved.

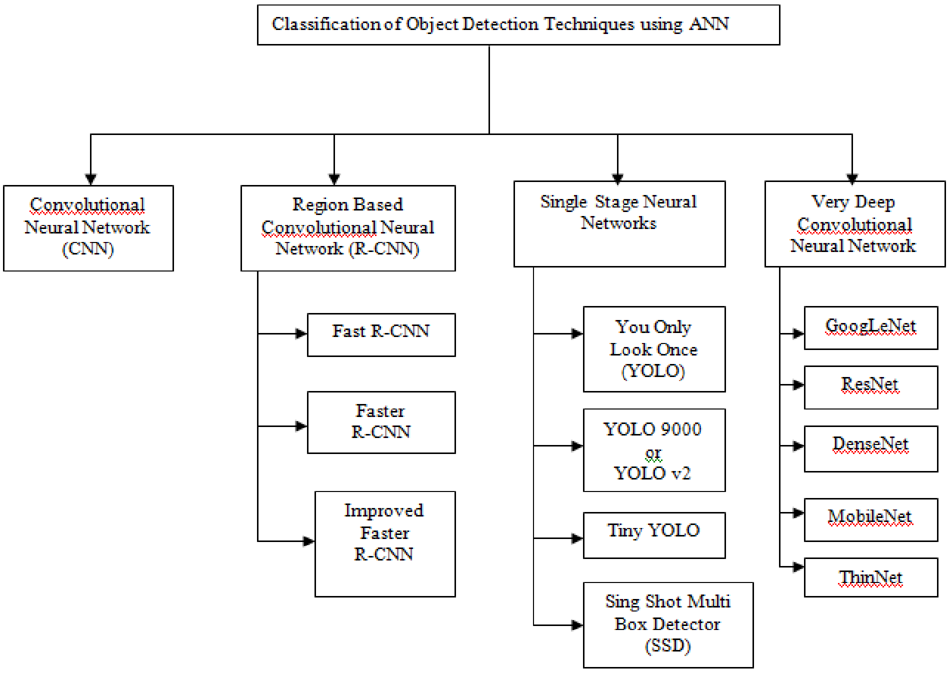

Fig 4: Classification of Object Detection Techniques using ANN

Table 2: Summary of Object Detection based on CNN

Paper	Techniques Used	Data Set Used to Train	Application
[8]	CNN	ImageNet LSVRC-2010	#
[10]	CNN	Manga109	Manga Images
[12]	CNN	ImageNet and PASCAL VOC	360^0 Panoramic Images

$-Did not clearly mentioned the data set trained on.

#- Did not clearly mentioned the application

Pengcheng Fan et al. [24] proposed a Faster R-CNN-based algorithm to improve the accuracy of small object detection. They used a subset of the COCO data set to create a benchmark database. This benchmark database was specially developed to evaluate the performance of object detection algorithms when detecting small objects. They improved Faster R-CNN with a more flexible method of integrating context information. The experiments show that the improved Faster R-CNN algorithm performs well in terms of accuracy and recall rate of small object detection. Your small object detection algorithm can strike a balance between detection speed and detection accuracy.

Wendi cai et al. [25] proposed a new algorithm based on a faster R-CNN to recognize the street objects. With the increasing number of traffic accidents, research and development of smart cars was promoted. The detection of street objects has become an important research topic. A generic model Detection algorithm based on the Convolution Neural Network (CNN) has to design the training model, while training and testing the model takes a long time. Transfer learning is used to optimize the pre-trained models using COCO's image task data sets and to transfer a generic deep learning model to a specific model with different weights and results. In addition, the CNN structure is adjusted to improve overall performance and the road environment is trained on the specific scene. Their results showed that the fine-tuned network is effective.

(iv). Improved Faster R-CNN

Changing Caoet al. [26] proposed an improved algorithm based on a faster regional CNN (Faster R-CNN) for the detection of small objects. Using the two-step detection idea, they proposed an improved loss function based on intersection over union (IoU) for the bounding box regression in the positioning phase and used bilinear interpolation to do the regions of interest (RoI) pooling operation to solve the problem Improve Positioning Deviation Issue In the detection phase, they used the multi-scale convolutional merger to provide more information to the feature map and the improved non-maximum suppression (NMS) algorithm to avoid losing overlapping objects. The results showed that the proposed algorithm performed well on traffic signs with a resolution in the range of (0.32), the algorithm recall rate reaching 90% and the accuracy rate reaching 87%. The detection performance is significantly better than with faster R-CNN. Therefore, their algorithm is an effective way to recognize small objects.

Yu Liu et al. [27] introduced an alternative training algorithm to address the problem of faster R-CNN. Among the various targeting algorithms, Faster R-CNN is an algorithm with excellent performance in terms of both detection accuracy and speed. However, it still has some shortcomings like too many negative samples. To address the problem of faster R-CNN, two strategies are introduced, hard negative sample mining and alternate training. Hard Negative Sample Mining is used to obtain hard negative samples that re-train the model to improve the trained model. By alternating training, RPN and Fast R-CNN split folding layers in faster R-CNN instead of learning two independent networks. The simulation result showed that the proposed algorithm has great advantages in terms of accuracy of detection. The summarized work of researcherson object Detection based on R-CNN is described in Table 3.

Table 3: Summary of Object Detection based on R-CNN

Paper	Techniques Used	Data Set Used to Train	Application
[14]	R-CNN	PASCAL VOC	#
[16]	R-CNN	MOS COCO	Real Time Object Detection
[17]	Fast R-CNN	PASCAL VOC	#
[18]	Faster R-CNN	PASCAL VOC and MOS COCO	Real Time Object Detection
[20]	Faster R-CNN	$	#
[21]	Faster R-CNN	SUN RGB-D	RGB-D Images
[24]	Faster R-CNN	MOS COCO	Small Object Detection
[25]	Faster R-CNN	MOS COCO	Street Object Detection
[26]	Improved Faster R-CNN	TT100K	Small Object Detection
[27]	Improved Faster R-CNN	PASCAL VOC	#

VI C Single Stage Network

(i) YOLO-You Only Look Once

Redmon et al. [28] introduced YOLO, a new approach to object Detection. With this algorithm, a single neural network predicts boundaries and class probabilities directly from frames in an evaluation. Because the entire detection pipeline is a single network, it can be optimized directly for detection performance. Their uniform architecture is extremely fast and processes images in real time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astonishing 155 frames per second and still achieves twice the mAP of other real-time detectors. Compared to state-of-the-art detection systems, YOLO makes more localization errors, but predicts incorrect detections far less frequently if nothing is available. Finally, YOLO learns very general representations of objects. It far outperforms all other Detection methods, including DPM and RCNN, when generalizing natural images to works of art in both the Picasso and People Art data sets.

(ii) YOLO 9000 (or) YOLO v2

Joseph Redmon et al. [29] introduced YOLO9000, a state-of-the-art real-time object Detection system that can recognize over 9000 object categories. First, they proposed various improvements to the YOLO detection method, both new and from earlier work. The improved model YOLOv2 is trained on PASCAL VOC and COCO. With a novel, multi-scale training method, the same YOLOv2 model can be executed in different sizes, which offers a simple compromise between speed and accuracy. With 67 FPS, YOLOv2 achieves 76.8 mAP for VOC 2007. With 40 FPS, YOLOv2 reaches 78.6 mAP, surpassing the latest methods such as Faster R-CNN with ResNet and SSD, while still running significantly faster. Finally, they proposed a method to train object Detection and classification together. With this method, you train YOLO9000 simultaneously in the COCO Detection data record and in the ImageNet classification data record. Thanks to the joint training, YOLO9000 can predict detections for object classes for which no detection data is marked. They validate their approach to the ImageNet discovery task. YOLO9000 receives 19.7 mAP in the validation set for ImageNet Detection, although only Detection data is available for 44 of the 200 classes. On the 156 classes that are not included in COCO, YOLO9000 receives 16.0 mAP. YOLO9000 predicts real-time detections for more than 9000 different object categories.

(iii) YOLO Lite (or) Tiny YOLO

Rachel Huang et al. [32] focused on YOLO-LITE, a real-time object Detection model that was developed for portable devices such as laptops or mobile phones without a graphics processing unit (GPU). The model was trained first on the PASCAL VOC data set and then on the COCO data set and achieved anmAP of 33.81% and 12.26%, respectively. YOLO-LITE runs on a non-GPU computer with around 21 FPS and 10 FPS after being implemented on a website with only 7 layers and 482 million FLOPS. This speed is 3.8x faster than the fastest, state-of-the-art model, SSD MobileNetv1. Based on the original object detection algorithm YOLOV2, YOLOLITE was developed to create a smaller, faster and more efficient model and to improve the accessibility of real-time object detection for a wide range of devices.

(iv) SSD-Single Shot Multi Box Detector

Wei Liu et al. [33] presented a method for recognizing objects in images using a single deep neural network. Their approach, called SSD, describes the output space of bounding boxes in a series of standard fields across different aspect ratios and scales per location of the feature map. At prediction time, the network generates ratings for the existence of each object category in each standard box and makes adjustments to the box to better match the shape of the object. In addition, the network combines predictions from multiple feature maps with different resolutions to naturally process objects of different sizes. SSD is straightforward compared to methods that require object suggestions, as proposal generation and subsequent pixel or feature re-sampling phases are completely eliminated and all calculations are combined in a single network. This makes SSD easy to train and easy to integrate into systems that require a detection component. Experimental results using the PASCAL VOC, COCO, and ILSVRC data sets confirm that SSD is competitive in accuracy against methods that use an additional object suggestion step and is much faster while providing a unified framework for training and inference. When entering 300 x 300, the SSD achieved 74.3% mAP in the VOC2007 test at 59 FPS on an Nvidia Titan X, and when entering 512 x 512 the SSD achieved a mAP of 76.9% and exceeded it comparable, faster, state-of-the-art R-CNN model. Compared to other single-stage methods, SSD is much more accurate even with a smaller input image size. The summarized work of researchers on object Detection on the basis of a one-step network is described in Table 4.

VI D Very Deep CNN

(i) GoogLeNet

Christian Szegedy et al. [35] proposed a Deep Convolutional Neural Network Architecture with the code name Inception, which achieves the newest state of the art for classification and detection in the ImageNet Large-Scale Visual Detection Challenge 2014 (ILSVRC14). The main feature of this architecture is the improved use of computer resources within the network. Thanks to a carefully designed design, they have increased the depth and breadth of the network and at the same time kept the computing budget constant. In order to optimize the quality, the architectural decisions were based on the Hebrew principle and the intuition of multi-scale processing. A special incarnation that is used in its filing for ILSVRC14 is called GoogLeNet, a 22-layer network, the quality of which is assessed as part of the classification and Detection.

Table 4: Summary of Object Detection based on Single Stage Networks

Paper	Techniques Used	Data Set Used to Train	Application
[28]	YOLO	ImageNet1000	Real Time Object Detection
[29]	YOLO9000 or YOLO v2	PASCAL VOC and MOS COCO	Real Time Object Detection
[32]	YOLO Lite or Tiny YOLO	PASCAL VOC and MOS COCO	Real Time Object Detection
[33]	SSD	PASCAL VOC and MOS COCO and ImageNet LSVRC	Real Time Object Detection

(ii) ResNet

Kaiming He et al. [37] presented a residual learning framework to facilitate the training of networks that are much deeper than those previously used. Deeper neural networks are more difficult to train. They rephrase the levels as residual learning functions in relation to the level inputs, instead of learning functions that are not referenced. They provide extensive empirical evidence that these residual networks are easier to optimize and can gain accuracy through a considerably greater depth. In the ImageNet dataset, residual networks with a depth of up to 152 layers are evaluated - 8 times deeper than VGG networks, but still with less complexity. An ensemble of these residual networks achieved an error of 3.57% in the ImageNet test set. This result won 1st place in the ILSVRC 2015 classification task. They also present analyzes of CIFAR-10 with 100 and 1000 layers. The depth of the representations is of central importance for many visual Detection tasks. Simply because of their extremely deep representations, they achieve a relative improvement of 28% compared to the COCO object Detection data record. Deep residual networks are the basis for their entries in ILSVRC & COCO 2015 competitions, where they also took first place in the tasks of ImageNet detection, ImageNet localization, COCO detection and COCO segmentation.

(iii) *DenseNet*

Gao Huang et al. [38] introduced the Dense Convolutional Network (DenseNet), which connects every layer in a feed-forward manner with every other layer. While conventional convolution networks with L layers have L connections - one between each layer and its subsequent layer - the proposed network has $\frac{L(L+1)}{2}$ direct connections. For each level, the feature maps of all previous layers are used as inputs, and their own feature maps are used as inputs to all subsequent layers. DenseNets offer several convincing advantages: they alleviate the problem of the disappearance gradient, increase the spread of features, promote the reuse of features and significantly reduce the number of parameters. They evaluate their proposed architecture using four highly competitive benchmark tasks for object Detection (CIFAR-10, CIFAR-100, SVHN and ImageNet). DenseNets achieve significant improvements over most of the prior art in most of them.

(iv) MobileNet

Andrew G. Howard et al. [39] presented a class of efficient models called MobileNets for mobile and embedded image processing applications. MobileNets are based on an optimized architecture that uses deeply separable windings to build light deep neural networks. They introduced two simple global hyper parameters that represent an efficient compromise between latency and accuracy. These hyper parameters allow the model builder to select the correct size model for its application based on the limitations of the problem. They presented extensive experiments on trade-offs between resources and accuracy and performed well compared to other popular models for ImageNet classification. They demonstrate the effectiveness of MobileNets in a variety of applications and use cases, including object detection, fine grain classification, face attributes and large scale geolocation.

(v) ThinNet

Sen Cao et al. [40] introduced a class of efficient network architectures called ThinNet, primarily for object Detection applications on memory and computational platforms. This architecture is based on two proposed modules: front module and tinier module. The front module reduces the loss of information due to raw input images by using more convolution layers with small filters. The Tinier module uses point folding layers in front of the conventional folding layer to reduce the model size and the calculation while ensuring the accuracy of detection. Experimental evaluations of the ImageNet classification and the PASCAL VOC object Detection data records show the superior performance of ThinNet compared to other common models. Your pre-trained classification model (ThinNet_C) achieves the same top 1 and top 5 performance as the classic AlexNet, but only with 1/50 of the parameters. The detection model also makes significant improvements over other detection methods, while a smaller model size is required to achieve high performance. The summarized work of researchers on object Detection based on Deep CNN is described in Table 5.

(vi) Comparison Table

The comparison of all techniques of the artificial neural network including their applications and the data set used for the training is shown in Table 6.

Table 5: Summary of Object Detection based on Very Deep CNN

Paper	Techniques Used	Data Set Used to Train	Application
[35]	GoogLeNet	ILSVRC	#
[37]	ResNet	MOS COCO and ILSVRC	#
[38]	DenseNet	ImageNet	#
[39]	MobileNet	MOS COCO	Mobile Application
[40]	ThinNet	PASCAL VOC	Real Time Object Detection

Table 6: Summary of Object Detection based on Various ANN Techniques

Paper	Techniques Used	Data Set Used to Train	Application
[8]	CNN	ImageNet LSVRC-2010	#
[10]	CNN	Manga109	Manga Images
[12]	CNN	ImageNet and PASCAL VOC	360^0 Panoramic Images
[14]	R-CNN	PASCAL VOC	#
[16]	R-CNN	MOS COCO	Real Time Object Detection
[17]	Fast R-CNN	PASCAL VOC	#
[18]	Faster R-CNN	PASCAL VOC and MOS COCO	Real Time Object Detection
[20]	Faster R-CNN	$	#
[21]	Faster R-CNN	SUN RGB-D	RGB-D Images
[24]	Faster R-CNN	MOS COCO	Small Object Detection
[25]	Faster R-CNN	MOS COCO	Street Object Detection
[26]	Improved Faster R-CNN	TT100K	Small Object Detection
[27]	Improved Faster R-CNN	PASCAL VOC	#
[28]	YOLO	ImageNet1000	Real Time Object Detection
[29]	YOLO9000 or YOLO v2	PASCAL VOC and MOS COCO	Real Time Object Detection
[32]	YOLO Lite or Tiny YOLO	PASCAL VOC and MOS COCO	Real Time Object Detection
[33]	SSD	PASCAL VOC and MOS COCO and ImageNet LSVRC	Real Time Object Detection

[35]	GoogLeNet	ILSVRC	#
[37]	ResNet	MOS COCO and ILSVRC	#
[38]	DenseNet	ImageNet	#
[39]	MobileNet	MOS COCO	Mobile Application
[40]	ThinNet	PASCAL VOC	Real Time Object Detection

VII. CONCLUSION

In this literature, various object Detection techniques based on the artificial neural network (ANN) were checked and grouped based on the implementation techniques. For each group of object Detection techniques created, a brief summary was mentioned to make the work easier for the readers at the end of each group. However, many challenges are still being considered, e.g. B. Training a very deep network because it is difficult and accurate and achieves good top 1 and top 5 error rates. As already mentioned, most of the articles in this overview have been described in which object Detection is implemented using various ANN techniques.

VIII. FUTURE SCOPE OF ANN FOR OBJECT DETECTION

(i). The ANN-based object detection technology can be implemented on portable devices such as drones and robots if the size is reduced.
(ii). Video is a visual multimedia source that combines a sequence of images into one moving image. The various ANN-based object Detection techniques can be extended to video.

REFERENCES

[1]. Vidushi Sharma, Sachin Rai, Anurag Dev "A Comprehensive Study of Artificial Neural Networks", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 2, Issue 10, October 2012 ISSN: 2277 128X.

[2]. Er.Parveen Kumar, Er.Pooja Sharma "Artificial Neural Networks-A Study", International Journal of Emerging Engineering Research and Technology, Volume 2, Issue 2, May 2014, PP 143-148.

[3]. Ms. Sonali. B. Maind et. al., "Research Paper on Basic of Artificial Neural Network", International Journal on Recent and Innovation Trends in Computing and Communication, Volume: 2 Issue: 1 January 2014.

[4]. M.R.M. Veera Manickam, M. Mohanapriya et al "RESEARCH STUDY ON APPLICATIONS OF ARTIFICIAL NEURAL NETWORKS AND E-LEARNING PERSONALIZATION", International Journal of Civil Engineering and Technology (IJCIET), Volume 8, Issue 8, August 2017, pp. 1422–1432, Article ID: IJCIET_08_08_156, ISSN Print: 0976-6308 and ISSN Online: 0976-6316.

[5]. Keiron O'Shea and Ryan Nash "An Introduction to Convolutional Neural Networks", , https://www.researchgate.net/publication/285164623.

[6]. Ciresan, D. Meier, U. Schmidhuber "Multi-column deep neural networks for image classification", In: Computer Vision and Pattern Detection (CVPR), 2012 IEEE Conference on. pp. 3642–3649. IEEE (2012).

[7]. Reagan L. Galvez, Argel A. Bandala et al "Object Detection Using Convolutional Neural Networks" Proceedings of TENCON 2018 - 2018 IEEE Region 10 Conference, (Jeju, Korea, 28-31 October 2018).

[8]. A. Krizhevsky, I. Sutskever and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in neural information processing systems, 2012.

[9]. J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In CVPR09, 2009.

[10]. Hideaki Yanagisawa, Takuro Yamashita, Hiroshi Watanabe "A Study on Object Detection Method from Manga Images using CNN", 978-1-5386-2615-3/18, IEEE 2018.

[11]. W.T. Chu, W.W. Li, "Manga Face Net: Face Detection in Manga based on Deep Neural Network," Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, pp. 412-415, June 2017.

[12]. Yiming Zhang, Xiangyun Xiao, Xubo Yang "Real-time Object Detection for 360-degree Panoramic Image using CNN", 2017 International Conference on Virtual Reality and Visualization (ICVRV), 978-1-5386-2636-8, IEEE 2019.

[13]. Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe: Convolutional architecture for fast feature embedding," in Proceedings of the22nd ACM international conference on Multimedia. ACM, 2014, pp. 675–678.

[14]. R. Girshick, J. Donahue, T. Darrel, J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in Conference on Computer Vision and Pattern Detection, pp. 580-587, IEEE, 2014.

[15]. R. Girshick, "Fast r-cnn," in Inernational Conference on Computer Vision, pp. 1440-1448, IEEE, 2015.

[16]. Joshua Herrera, Mevlut A Demir, ParsaYousefi, John J Prevost, and Paul Rad "Distributed Edge Cloud R-CNN for Real Time Object Detection", 978-1-5323-7791-4/2018/ @ 2018 TSI Press, Inc.

[17]. Ross Girshick Fast R-CNN IEEE International Conference on Computer Vision IEEE Transaction, IEEE 2015 DOI 10.1109/ICCV.2015.169

[18]. S. Ren, K. He, R. Girshick, J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks" in Advances in Neural Information Processing Systems, pp. 91-99, 2015.

[19]. R. Girshick, J. Donahue, T. Darrell, and J. Malik. Regionbased convolutional networks for accurate object detection and segmentation. TPAMI, 2015. 5, 7, 8.

[20]. Syed Mazhar Abbas, Dr. Shailendra Narayan Singh "Region-based Object Detection and Classification using Faster R-CNN", International Conference on "Computational Intelligence and Communication Technology" (CICT 2018).

[21]. Irina Mocanu, Cosmin Clapon "Multimodal Convolutional Neural Network for Object Detection Using RGB-D Images" IEEE Transaction, 978-1-5386-4695-3/18, IEEE 2018.

[22]. A. Eital, J.T. Springenberg, L. Spinello, M. Riedmiller, and W. Burgard, "Multimodal deep learning for robust RGB-D object Detection", IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Hamburg, Germany, pp. 681–687, 2015.

[23]. X. Xu, Y. Li, Gangshan Wu, and J. Luo. 2017. Multi-modal deep feature learning for RGB-D object detection", Pattern Detection, vol. 72, pp. 300–313. DOI: https://doi.org/10.1016/j.patcog.2017.07.026.
[24]. Pengcheng Fang, Yijie Shi "Small Object Detection Using Context Information Fusion in Faster R-CNN", 4th International Conference on Computer and Communications, 978-1-5386-8339-2, IEEE 2018.
[25]. Wendi Cai, Jiadie Li, ZhongzhaoXie, Tao Zhao, Kang Lu "Street Object Detection Based on Faster R-CNN", Proceedings of the 37th Chinese Control Conference, July 25-27, 2018, Wuhan, China.
[26]. CHANGQING CAO, BO WANG et al "An Improved Faster R-CNN for Small Object Detection", IEEE Transaction, DOI 10.1109/ACCESS.2019.2932731, IEEE Access.
[27]. Yu Liu "An Improved Faster R-CNN for Object Detection", 11th International Symposium on Computational Intelligence and Design, IEEE Transaction, DOI 10.1109/ISCID.2018.10128.
[28]. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in Proceedingsof the IEEE Conference on Computer Vision andPattern Detection, 2016, pp. 779–788.
[29]. J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," http://pjreddie.com/yolo9000/.
[30]. Jun Sang, Zhongyuan Wu et al "An Improved YOLOv2 for Vehicle Detection" Sensors 2018, 18, 4272; doi:10.3390/s18124272.
[31]. Yang-Lang Chang, Amare Anagaw et al "Ship Detection Based on YOLOv2 for SAR Imagery", Remote Sens. 2019, 11, 786; doi:10.3390/rs11070786.
[32]. Rachel Huang, Jonathan Pedoeem, Cuixian Chen "YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers".
[33]. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.Y. Fu, A. C. Berg, "Ssd: Single shot multibox detector," in European Conference on Computer Vision, pp. 21-37, Springer, 2016.
[34]. Malay Shah, Prof. RupalKapdi **"**Object Detection Using Deep Neural Networks" International Conference on Intelligent Computing and Control Systems, 978-1-5386-2745-7/17, IEEE Transaction, IEEE 2017.
[35]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. ILSVRC, 2014.
[36]. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image Detection", ILSVRC, 2014.
[37]. K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image Detection, IEEE Conference on Computer Vision and Pattern Detection, 2016.
[38]. Gao Huang, Zhuang Liu, Laurens van der Maaten "Densely Connected Convolutional Networks", IEEE Transaction.
[39]. Andrew G. Howard, Menglong Zhu et al "MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications".
[40]. Sen Cao, Yazhou Liu et al "ThinNet: An Efficient Convolutional Neural Network for Object Detection" 24th International Conference on Pattern Detection (ICPR),Beijing, China, August 20-24, 2018.
[41]. Forrest N. Iandola, Song Han et al "SQUEEZENET: ALEXNET-LEVEL ACCURACY WITH 50X FEWER PARAMETERS AND <0.5MB MODEL SIZE", conference paper at ICLR 2017.
[42]. K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image Detection," International Conference on Learning Representations, 2015.
[43]. J. Deng, W. Dong, R. Socher, L.J Li, K. Li, L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in Computer Vison and Pattern Detection 2009 (CVPR2009) IEEE Conference on, pp. 248- 255, IEEE, 2009.
[44]. C. Szegedy,W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in The IEEE Conference on ComputerVision and Pattern Detection (CVPR), June 2015.
[45]. M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results," http://www.pascalnetwork. org/challenges/VOC/voc2007/workshop/index.html.
[46]. J. Deng, A. Berg, S. Satheesh, H. Su, A. Khosla, and L. Fei-Fei. ImageNet Large Scale Visual Detection Competition 2012 (ILSVRC2012). http://www.image-net.org/ challenges/LSVRC/2012/.
[47]. Y. Jia. Caffe: An open source convolutional architecture for fast feature embedding. http://caffe.berkeleyvision.org/, 2013.
[48]. K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image Detection. In ICLR, 2015. 1, 5.
[49]. Object Detection using fasterR-CNN,2017[Online] Available: https://in.mathworks.com/help/vision/examples/object detection-using-faster-r-cnn-deeplearning. html?requestedDomain=www.mathworks.com [Accessed 10 Oct-2017].
[50]. C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, and A. Rabinovich, "Going deeper with convolutions," in IEEE Conference on Computer Vision and Pattern Detection (CVPR), 2015.
[51]. C. Szegedy, A. Toshev, and D. Erhan, "Deep neural networks for object detection," in Neural Information Processing Systems (NIPS), 2013.
[52]. Deng J, Berg A, Satheesh S, et al. Imagenet large scale visual Detection competition 2012 (ILSVRC2012). 2012.
[53]. Deng J, DongW, Socher R, et al. Imagenet: A large-scale hierarchical image database. Computer Vision and Pattern Detection, 2009.
[54]. Chen X, Fang H, Lin T Y, et al. Microsoft COCO captions: Data collection and evaluation server, arXiv preprint arXiv:1504.00325, 2015.
[55]. Girshick R, Donahue J, Darrell T, et al. "Region-based convolutional networks for accurate object detection and segmentation." IEEE Transactions on Pattern Analysis and Machine Intelligence, 2016, vol. 38, no.1, pp. 142-158.
[56]. Roh, Myung-Cheol, and Ju-young Lee, "Refining faster-RCNN for accurate object detection," in Proc. of Fifteenth IAPR International Conference on Machine Vision Applications (MVA), pp. 514-517, 2017.
[57]. Han, Guangxing, Xuan Zhang, and Chongrong Li, "Revisiting Faster R-CNN: A Deeper Look at Region Proposal Network," in Proc. Of International Conference on Neural Information Processing, Springer, Cham, pp. 14-24, 2017.
[58]. DAI J, LI Y, HE K, et al, "R-FCN: object detection via region-based fully convolutional networks", in Advances in neural information processing system, pp. 379-387, 2016.
[59]. Chen, C., Liu, M., Tuzel, O., Xiao, J.: R-cnn for small object detection. (2016) 214–230.
[60]. Ren Y, Zhu CR, Xiao SP. "Small Object Detection in Optical Remote Sensing Image via Modified Faster R-CNN." Journal Citation Reports., vol.8, no.5, 2018.
[61]. Huang Jipeng, Shi Yinghuan, Gao Yang. "Multi-Scale Faster-RCNN Algorithm for Small Object Detection." Computer Research and Development.2019, vol.56, pp.319-327.
[62]. Liang Zhenwen, Shao Jie, Zhang Dongyanget al. "Small object detection using deep feature pyramid networks."19th Pacific-Rim Conference on Multimedia. 2018, 11166 LNCS, pp.554-564.
[63]. E. L. Denton, W. Zaremba, J. Bruna, Y. Lecun, and R. Fergus, "Exploiting Linear Structure Within Convolutional Networks for Efficient Evaluation," in Neural Information Processing Systems, 2014, pp. 1269-1277.

[64]. S. Han, J. Pool, J. Tran, and W. J. Dally, “Learning both weights and connections for efficient neural networks,” in Neural Information Processing Systems, 2015, pp. 1135-1143.

[65]. S. Han, H. Mao, and W. J. Dally, “Deep Compression: Compressing Deep Neural Networks with Pruning, Trained Quantization and Huffman Coding,” in International Conference on Learning Representations, 2016.

[66]. F. Yu, and V. Koltun, “Multi-Scale Context Aggregation by Dilated Convolutions,” in International Conference on Learning Representations, 2016.

[67]. Z. Liu, J. Li, Z. Shen, G. Huang, S. Yan, and C. Zhang, “Learning Efficient Convolutional Networks through Network Slimming,” in International Conference on Computer Vision, 2017.

[68]. C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, “Rethinking the Inception Architecture for Computer Vision,” in Computer Vision and Pattern Detection, 2016, pp. 2818-2826.

[69]. C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inceptionv4, Inception-ResNet and the Impact of Residual Connections on Learning,” in National Conference on Artificial Intelligence, 2016, pp. 4278-4284.

[70]. R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization,” in International Conference on Computer Vision, 2017, pp. 618-626.