iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 **Issue:** VII **Month of publication:** July 2020

DOI: https://doi.org/10.22214/ijraset.2020.30261

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 8 Issue VII July 2020- Available at www.ijraset.com

Application of Machine Learning in Computer Vision: A Brief Review

Mr. Shismohammad Mulla[1], Dr. Mahesh Chavan[2]
[1]*M.Tech. in Electronics and Telecommunication Engineering,* [2]*Professor in Electronics Engineering, Department of Electronics Engineering, KIT College of Engineering (Autonomous), Kolhapur*

Abstract: ***A scientific study on the importance of machine learning and its applications in the field of computer vision is carried out in this paper. Recent advancements in Artificial Intelligence, deep learning, computing resources and availability of large training datasets made tasks such as computer vision and natural language processing extremely fast and accurate. Thus Artificial intelligence is a trending topic in the field of computing. Deep learning is a subcategory of machine learning in the field of artificial intelligence. Image processing task can be performed efficiently by using machine learning methods, thus machine learning will provide a better understanding of complex images. Object detection, recognition and tracking are the fields related to computer vision. In the computer vision with the help convolutional neural network-based algorithms like YOLO and R-CNN make a big leap in this field. Algorithms based on machine learning models are excellent at recognizing patterns but typically requires an enormous amount of data sets and lots of computational power. Generally, the neural network requires graphics processing unit for faster execution of machine learning models. This review paper gives a brief overview of real-time object detection and machine learning algorithms implemented by various researchers around the world. Also, this paper consists of a study of various methodology used to detect and recognize a particular object in the image. Real-time object detection algorithms are going to play a vital role in the field of computer vision.***

Keywords: Digital image processing, Computer Vision, Machine Vision, Artificial Neural Networks, Deep Neural Network, Convolutional Neural Network, Applications of Deep Learning Object Detection, Object Recognition, Artificial intelligence, Deep Learning, Machine Learning.

I. INTRODUCTION

The human visual system has always played an important role in human life. Human vision played a very key role in the greatest discoveries human has made. Apparently, Vision is human beings most important sensory. The human visual system is very complicated, fast and accurate; it allows us to perform extremely complex tasks like driving, learning and teaching. Consequently, the development of Computer vision field started to stride. Digital Image processing is the use of digital computers to perform processing on the image with the computer algorithm. Image processing is a subfield of computer vision. In digital image processing, various computer algorithms are used to perform the image processing task. It begins from image acquisition and using complex digital image processing algorithms to perform Feature extractions, Classification, segmentation, object recognition and pattern recognition tasks. Machine learning is going to improve the performance of image processing task and the accuracy of systems. Due to the development of Computer vision field machine can gain high-level understanding from digital images or videos like how humans see the world. This development seeks to automate tasks that a human can perform. Humans look at an image and instantly know what it is, where the objects are and how they can interact with the environment. The ultimate goal of computer vision systems is to understand the content of videos and digital images. This includes techniques of acquiring, analyzing, processing and extraction of relevant data from the real world. Computer vision is improved very fast in the last few decades, studied and developed by many scholars. Deep learning is the fastest growing field of machine learning. It consists of multiple hidden layers of artificial neural networks. The current developments in deep learning in various fields have provided a big role in artificial intelligence. In this review paper, we chronologically present how deep learning algorithms have been employed in image processing and other fields of computing. Purpose of this review paper is to give a general idea of machine learning, how it has been used in computer vision and how it has been improving in the last decades.

The most challenging task in many of the object detection method is a reliance on different computer vision methods for assisting the deep learning-based approach, which leads to slow and much less accurate performance. Many problems in computer vision have been saturating on their accuracy last a decade. However, with the advancement of deep learning techniques, the accuracy of these types of tasks significantly improved. New fast and accurate machine learning models for object detection and recognition

allow computers to drive cars and machines without requiring specialized sensors. In object detection, we have to detect an object in a frame, place a bounding box around it and classify the object. To detect an object in an image or video, these types of systems take classifier for object and evaluate it at several places and scales in a test image. Systems similar to the deformable parts model (DPM) uses the sliding window method where the classifier is run at evenly spaced slides across the whole image. These Historical detection systems repurpose classifiers to perform the detection task. With the development of deep learning techniques, the accuracy for object detection has drastically improved. In this paper, we aim to study available techniques for object detection with high accuracy with real-time performance. YOLO is a new deep learning approach to object detection task before that object detection repurposes classifiers to perform detection. In the YOLO algorithm, the authors frame the object detection problem as a regression problem to spatially separate bounding boxes and an associated class of probabilities [1]. This system is extremely speedy, efficient and accurate. It outperforms other older detection methods which are present over the decade, YOLO model can processes images in real-time at 45 frames per second. A smaller model of the YOLO network is known as fast-YOLO, which can process 155 frames per second. Today machine learning is widespread growing subject amongst the computer vision research community.

II. BACKGROUND AND MOTIVATION

Artificial intelligence (AI) refers to the simulation of humanlike intelligence in machines that can be programmed to think and solve problems like humans and imitate their actions. A more formal definition of an artificial intelligent agent was given by Tom Mitchell as a computer program is said to learn from experience (E) concerning some task (T) and some performance measure (P), if its performance on T, as measured by P, improves with experience E then the program is called a machine learning program [2]. Machine learning is the discipline of Artificial Intelligence that provides systems with the capability to learn and enhance its ability from learning past experience without explicitly been programmed. Machine learning is the development of a computer algorithm that can access information and use it to learn for themselves. The prime aim of this type of system is to allow computers to learn automatically.

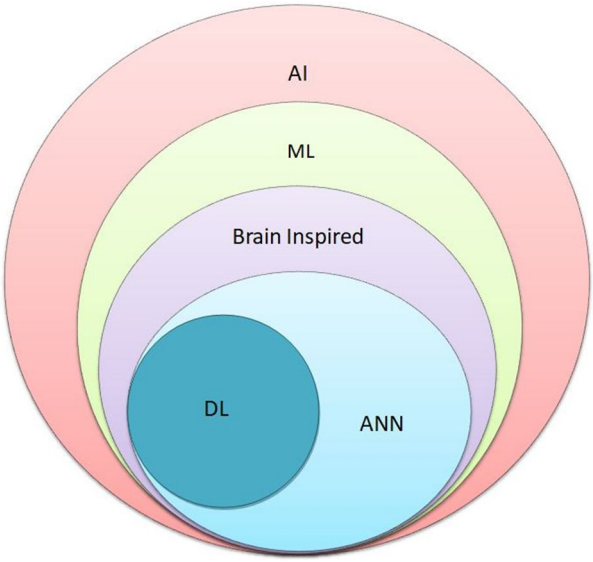


Figure 1. Classifications of Artificial Intelligence.

AI: Artificial Intelligence; ML: Machine Learning; ANN: Artificial Neural Networks; DL: Deep Learning.

A. Machine Learning Methods

1) *Supervised Machine Learning:* Supervised machine learning is a type of machine learning algorithms which can apply what has been learned in the past to new information with the use of labelled examples to predict future events precisely. This type of machine learning method is based on the estimation of computed output and predicted output. Thus learning refers to computing the error and adjusting the error for attaining the expected output. For example, a data set of apartments of specific measurement with real costs is given, then the supervised algorithm can provide answers such as for new apartment what would be the approximate price.

2) *Unsupervised Machine Learning:* In Unsupervised machine learning method, data with no labels are given to the learning algorithm and leaving it on its own to find structure and patterns in its input. Unsupervised learning can be a complete goal in itself or a means towards an end (feature learning) [3]. Unsupervised Learning is harder as compared to Supervised Learning tasks. Unsupervised machine learning takes lots of computational power and very time-consuming.
3) *Reinforcement Machine Learning:* Reinforcement learning is a method based on output with how an intelligent agent ought to take actions in an environment to maximize some kind of long-term reward like humans do in their life [2]. A reward is given for correct output to input provided and a penalty for incorrect output, like how human learns in its childhood. Reinforcement learning is different from supervised learning in the sense that a correct input and output pairs are by no means presented, not all moves explicitly corrected. In short, Reinforcement learning is the type of machine learning model to make a sequence of decisions. The artificial intelligent agent learns to achieve a goal in a very complex and uncertain environment. The computer program use trial and error method to find a solution to the problem. By using the power of trial and test methods, reinforcement learning is the most effective way to manipulate the machine's creativity.

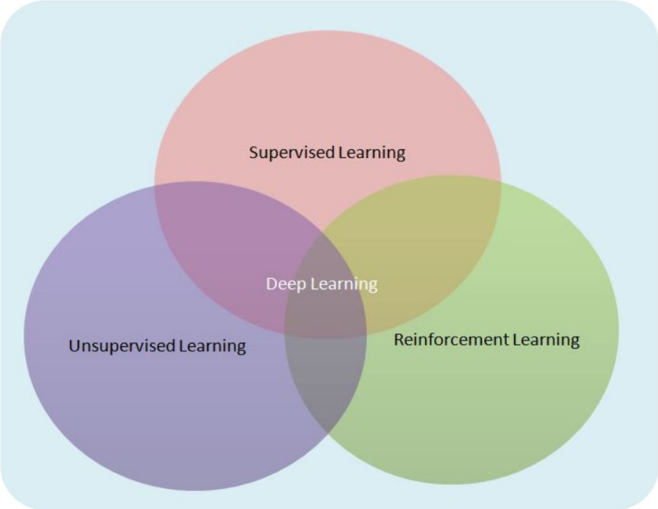


Figure 2. Categories of Machine Learning.

B. Artificial Neural Networks

The goal of an artificial agent is to create a machine that simulates the human brain. The incredibly complex human brain has inspired the development of artificial neural networks. In 1943, McCulloch and Pitts investigated that, how the small brain could produce highly complicated patterns and produce accurate output by using interconnected fundamental cells, known as neurons [4]. The McCulloch and Pitts discovered model of a neuron, called an MCP (McCulloch and Pitts) model. MCP is fundamental of the artificial neuron model. McCulloch and Pitts have a vital contribution to the development of artificial neural networks.

Artificial Neural Networks (ANN) is an information processing unit that is inspired by the way the biological central nervous system functions. It is composed of large numbers of highly interconnected processing elements called neurons. These neurons work together to form the world's complex organ human brain. These neurons are Working together to solve a unique problem. Similar to how brain cell can process information at incredibly fast and accurate speed. Deep Learning refers to Artificial Neural Networks with multiple hidden layers. Over the last few decades, ANN viewed as one of the most powerful tools and has tremendous potential. Indeed this turns out to be very famous in the literature, which is capable to deal with a huge volume of data. The latest research shows that having deeper hidden layers in the neural network has presently started to surpass classical techniques.

C. Deep learning

Deep Learning or Deep Neural Network belongs to Artificial Neural Networks with multiple hidden layers. Deep learning can be known as an approach to improve, optimize output and reduce processing times as compared to a single neural network in various computing processes. In the deep learning techniques has been used for natural language processing, pattern detection, classification, image processing tasks like caption generation, object detection, object recognition, image segmentation and handwriting to text generation [5]. This type of neural network model can be trained on the most challenging publicly available dataset (like PASCAL and VOC).On that dataset, computer vision task such as an object detection can be carried out and their performance is evaluated.

D. Convolutional Neural Network (CNN)

A convolutional neuronal network is another type of deep learning neural networks structure that has multiple hidden layers. Convolutional neural networks are inspired by typical structures present in the human's visual system. The convolutional neural networks are very similar to the deep neural networks as they are produced by neurons that possess values in the form of weights and biases, that weights and biases can be adjusted to perform optimization. A convolutional neural network can have three essential kinds of layers, specifically, (i) convolutional layer, (ii) pooling layer, and (iii) fully connected layer.

CNN has an extraordinary performance in computer vision problems. Fundamentally the applications that deal with image data processing, such as image classification, computer vision, and natural language processing (NLP). Nowadays the popular rise of convolutional neural network (CNN) has created a fabulous performance in object classification and object detection. The exceptional result demonstrated by Girshick et al. [6] proposed that the region proposal network with a convolutional neural network (R-CNN) and Joseph Redmond, Ross Girshick and et al in YOLO Algorithm for object detection and classification. The performance of such systems is outstanding on the state of the art well-known detection benchmarks such as PASCAL VOC, as compared to the traditional handcrafted methods which essentially used deformable part model (DPM) and histogram of oriented gradients (HOG) methods.

III. APPLICATIONS OF MACHINE LEARNING IN COMPUTER VISION

Machine Learning can be used in robotics, medical diagnosis, education, electronic trading platforms, transportation, business decisions and remote sensing and many more. Artificial Neural network has exceptional applications in data science and computing. Machine learning also has applications in field of computer vision, such as object detection and recognition (e.g., [7], [8]), object motion tracking (e.g., [9], [10]), action recognition (e.g., [11], [12]), and image segmentation (e.g., [13], [14]), in which deep learning perform exceptionally. Deep learning designates an abstract layer evaluation and hierarchical problem-solving methods. However, it can be utilized in several real-life problems also. Deep learning can be understood as a technique to improve output and reduce processing time. In the subject of machine vision, natural language processes (NLP) and pattern recognition deep learning techniques have used in the past decade. Due to the scope of this paper, it is very challenging to discuss every application of neural networks. So our main focus is on convolutional neural networks and real-time object detection systems. Nowadays with the emergence of new machine learning algorithms fired fast development in computer vision applications.

A. Automotive

Computer vision plays a very vital role in the emergence of Autonomous vehicles. McKinsey estimated that 15% of cars sold are autonomous by the year 2030. Every year, traffic accidents account for 2.2% of global deaths. This type of Autonomous vehicles can reduce traffic accidents hugely. In the field of transportation, an autonomous vehicle used to streamline traffic flow, with computer vision, we can make traffic management processes more efficient and smart. With advance computer vision technology, we can see a paradigm shift in the automotive industry and can make travelling very safe.

B. Medicine

A deep learning technique has been used for medical applications in the last decade. For example, with the help of deep learning algorithms in neural networks lead to excellent performance in disease detection and prediction. An image generated from magnetic resonance imaging (MRI) from human organs was processed to detect disease. Although in the success of this deep learning method, there are some small problems in these methods. In supervised machine learning, unfortunately, the Training of dataset, high computational resources and dependency on the high quality of training data like MRI images are some limitations to the current method. This can be overcome by the introduction of better training data sets and high GPU power for particular applications. In Optical Coherence Tomography (OCT) deep learning algorithms are showing outstanding results. Conventionally, images can be treated as a manual development of convolutional matrices [15]. In the deep learning method, a framework can predict and detect retinal pathologies and reduce OCT costs. Besides this, CNN can also be applied on retinal images to detect diabetic retinopathy. The machine-learning approach performs accurate detection and prediction than the averages of reputed ophthalmologists. Machine learning models also can be used for drug research by predicting molecular structure and simulating molecular properties like toxicity and the ability to bind to other molecules. Also, it can be used for simulating biological or chemical processes of isolated molecules without using costly softwares and the whole system is extremely faster than traditional methods.

C. Agriculture

Advancement in computer vision algorithms makes the prediction and detection of plant disease through some machine learning algorithm is possible. Deep learning method reduces exhaustive monitoring of large farms of crops. Also, with the help of deep learning very early stage of the disease prediction happens when the occurrence of symptoms when they first appear on plant leaves. John Deere introduced a semi-autonomous combine harvester in CES 2019. This semi-autonomous combine harvester uses artificial intelligence (AI) algorithms and computer vision to inspect grain quality. As semi-autonomous combine harvester gets into a crop field and crop field gets harvested and find the optimal route through the crops to detect anomalies in the field [16]. Especially, computer vision systems can be used for classifying food products into particular grades, detecting defects, and estimating properties such as colour, shape, size, surface defects, and contamination. There's also great potential for deep learning algorithms to identify crop weeds so pesticides and herbicides can be sprayed directly on them instead of on the all crop field.

D. Video Tracking

Video tracking is the process of detecting a moving object and tracking it over a certain time. Humans can recognize the face of a person, therefore, a computer could be capable to detect and track the human subject in real-time. Real-time object detection in an image or video has been getting a lot of attention from the scientific community. Before deep learning appeared, some applications have been carried out within the field of pattern recognition through layer processing, with developments of deep learning the accuracy for object detection has drastically improved. Object tracking has a wide range of applications in computer vision, such as video surveillance, traffic flow monitoring, medical imaging, and human activity recognition. In CCTV surveillance systems detect vehicles in real-time to identify particular events, such as car accidents, recognition of license plates, Vehicle counting, or differentiating between cars, buses, trucks, taxis, and so on. Automatic license plate recognition systems identifies car that commits a crime or accident. Crowd management is another type of security system. Crowd management through video analysis can be useful in shopping malls, hospitals, stadiums, and airports. Recent researches have utilized deep learning method as the primary tool for identification, recognition and tracking. convolutional neural networks effectiveness can be achieved above 90 percent of accuracy in computer vision systems [17].

IV. LIERATURE REVIEW

Konrad Ahlin, Benjamin Joffe and et al. in 2016 presented a paper on Autonomous Leaf Picking robot Using Deep Learning and Visual-Servoing [18]. They have developed Autonomous Leaf Picking robot using the deep learning model. This robot has a camera sensor for an image processing task with Convolutional Neural Networks in image processing and Monoscopic Depth Analysis for visual-servoing. This robotic agent could be used to precisely locate and grasp a leaf from a plant in an unstructured environment. In this experiment, the robot has used only a single camera for identification. Object detection algorithm was able to detect at least one leaf for every frame where the plant was present. The images from the camera had a resolution of 1280x720. The algorithm computational performance tested on NVIDIA GTX 880M GPU. Image Processing took an average of 0.7s per frame.

Liu, Po-Yu and el al in 2018 published a paper on Image reconstruction using deep learning. This paper proposes a deep learning model that can achieve statistically important improvements over traditional algorithms in Poisson image de-noising particularly when the noise is very high. With the booming market of cameras, the necessity for advanced Poisson image de-noising algorithm has been increased. The authors proposed architecture includes a hybrid of convolutional and deconvolutional neural network layers with symmetric connections. De-noising system developed by authors achieved statistically significant 0.38dB, 0.68dB, and 1.04dB average PSNR gains over benchmark traditional algorithms in experiments [19]. This de-noising network also can operate with less computational time while nevertheless outperforming the benchmark algorithm. Due to the computer's computational capability, this network contains only 6 layers. Although, this network performs Poisson de-noising without being explicitly taught the noise characteristics, while still, it can learn the parameters from the data alone.

Hyunkwang Lee, Shahein Tajmir and el al. in 2017, published a paper on Fully Automated Deep Learning System for Bone Age Assessment with the convolutional neural network. Authors has developed a deep learning system to automatically detect and segment the wrist and hand radiograph, to perform automated Bone age assessment. Developed bone age assessment model with a deep learning system uses a fine-tuned convolutional neural network and generate radiology reports. Attention maps generated with using the input occlusion method, this shows that what are the features the trained model uses to perform Bone age assessment. The proposed design uses an Image-Net pre-trained dataset and fine-tuned CNN to obtain 57.32 and 61.40% accuracies for the female and male [20]. The trained algorithm evaluates similar regions of the wrist and hand for Bone age assessment. Fully automated Bone age assessment model can be deployed in the clinical environment as a decision-supporting system. Authors developed machine learning BAA model is more accurate and efficient than the conventional method.

Paul Viola et al in 2001, published a paper on rapid object detection using a boosted cascade of simple features, they developed a method for object detection which minimizes computation time while attaining high detection accuracy. The strategy was used by authors to construct a face detection system is 15 faster than any previous approach [21]. Authors developed model brings new algorithms, representations, and insights which are quite popular and also have a more widespread application in computer vision and image processing. In this paper authors used performs a set of detailed experiments on a tough face detection dataset. This dataset consists of faces under a very large range of conditions such as illumination, pose, scale, and camera position and angle. Author work is divided into three key components. The first component is an integral image which allows the features used by the detector to be computed very quickly. The second component is a learning-based algorithm typically based on ADA-boost. The third component is a cascade method for combining increasingly more complex classifiers. Typically 38-layer cascaded classifier was trained to detect front faces. To train the detector, the authors used a dataset of the face and non-face training images. The face training set consisted of 4916 hand labelled faces scaled and aligned to a base resolution of 24 by 24 pixels. The faces extracted from the World Wide Web. Every classifier in the cascade was trained with the 4916 training faces. In this study, non-face images are also used to train the next layers so that obtained by scanning the partial cascade across the non-face images and collected false positives.

Navneet Dalal, Bill Trigs et al 2005, in the paper, Histograms of Oriented Gradients for Human Detection. An author shows that histograms of oriented gradients descriptors notably outperform current feature sets for human detection. They utilized a linear SVM based human detection model. They used linear SVM as a baseline classifier and their linear SVM detector is quite efficient and thus processing a 320 X 240 image in less than a second. In this research, their model performed perfectly on MIT pedestrian database. So author's use more challenging Dataset, this dataset contains 1800 annotated human images with a large range of poses variations and backgrounds, thus reduced false-positive rates larger than HAAR wavelet-based detector. In this experiment author used locally normalized histogram of the gradient orientations features like SIFT descriptors in a dense overlapping grid which gives very good results for person detection.[22].They concluded that fine-scale gradients, relatively coarse spatial binning, fine Orientation binning, and high-quality local contrast normalization in overlapping descriptor blocks are equally important for good results. Alex Krizhevsky, Ilya Sutskever et al 2012,in paper Image Net Classification with Deep Convolutional Neural Networks [23], developed a supervised machine learning model .In this model they extensively used Convolutional Neural networks. This convolutional neural network has 60 million parameters and 650K neurons. Additionally, this convolutional neural network is consists of five convolutional layers, few of them are followed by max-pooling layers and the other three are fully-connected coats with a final 1000 way soft-max. To make the training of this neural network faster, they used non-saturating neurons and a highly efficient GPU for implementation of the convolution operation. Further to decrease over fitting in the fully-connected Layers author used a recently-developed regularization method called “dropout” that showed to be very effective. They have trained a deep convolutional neural network model to be capable of classifying the 1.2 million high-resolution images in the data set of image-net. Author down-sampled original data set, down-sampled image to 256 x 256 resolution for faster operations. Author used neural network takes a week to train on two GTX 580 3GB GPU. Consequently, authors have developed a Convolutional Neural networks model which can result in the large, deep convolutional neural network which is capable of achieving record-breaking results. This type of deep convolutional neural network is able to achieve high accuracy on a very challenging dataset using completely supervised learning.

Ross Girshick, Jeff Donahue et al in 2014, have developed a supervised machine learning algorithm in Rich feature hierarchies for accurate object detection and semantic segmentation Tech report (v5) [24]. In that machine learning model, they developed regions with convolutional neural networks (R-CNN) model. They tested R-CNN model on PASCAL VOC dataset. In this paper, authors introduced a simplistic and scalable detection algorithm that can improve mean average precision (mAP) with larger than 30 percent. They also compared R-CNN algorithm to Over-Feat model, which is a sliding-window detector based similar CNN architecture. The used CNN input nodes take 227 x 227 pixels image. Their machine learning model works in three modules, which are (1) firstly extracts 2000 bottom-up region proposals, (2) convolutional neural network that computes fixed-length feature Vector model for each proposal and then (3) classifies each region using linear SVM system. Their system is very accurate and extremely efficient; also R-CNN can be scale to hundreds of thousands of object classes without resorting to approximate techniques, like hashing if there were 100k classes, the resulting matrix multiplication takes only 10 seconds on multi-core CPU. R-CNN Pre-training Was performed on open source CAFFE CNN library. After that Domain-specific fine-tuning was made on CNN network. They used stochastic gradient descent (SGD) optimizer model, training of the CNN parameters used warped region proposals technique. After the implementation of the deep learning model, they have tested R-CNN model on PASCAL VOC 2007 dataset and then fine-tuned the convolutional neural network model on VOC 2012 dataset trained and optimized SVMs on the VOC 2012 dataset. Results show that the R-CNN model achieves an mAP of 31.4%, which is considerably leading from 24.3% result of Over-feat.

Joseph Redmon, Ali Farhadi et al, developed a revolutionary YOLO deep learning Algorithm, a state of the art, real-time object detection algorithm based on convolutional neural networks (CNN); they have introduced various enhancements to YOLO detection algorithm. New innovative YOLO algorithm is remarkably fast and accurate. With their proposed YOLO machine learning model more than 9000 categories of objects can be detected [1]. Authors reframe object detection problem as a separate regression problem, right from image pixels to bounding box coordinates and their class probabilities. With YOLO architecture images can processes in real-time up to 45 FPS which is pretty impressive. The smaller version of the YOLO algorithm known as Fast-YOLO which can process images around 155 FPS. It outperforms other well-known detection Methods, such as Faster RCNN with Res-Net and DPM. it does while still running significantly faster. They have trained the YOLO 9000 model on the Image Net dataset and COCO dataset. The Next updated YOLO model can perform detection on PASCAL VOC and COCO dataset. In the training phase, the network is trained on standard Image-Net 1000 class classification dataset. They have used stochastic gradient descent (SGD) optimization with a learning rate of 0:1 in starting for 160 epochs. YOLO views the whole image during training and testing time, which completely encodes contextual information about classes and their appearance. In this YOLO algorithm, the system divides the input image into an S x S grid. Initially, The YOLO trains the classifier network to resolution of 224 x 224 but that can be increased resolution to 448 x 448 for detection. In YOLO VERSION 2 they have fine-tuned the classification network to 448 x 448 resolution for 10 epochs on Image-Net dataset; with the addition of anchor boxes, after that, they changed the resolution to 416 x 416. If the object falls in the centre of the grid cell, that grid cell is responsible for object detection. Every grid cell predicts confidence scores and 'B' bounding boxes for those boxes. They implemented this model as a convolutional neural network and evaluate it on the PASCAL VOC detection dataset. They used NVIDIA Titan X GPU for their training and testing operations. In beginning convolutional layers of the neural network extract feature from an image, while output probabilities and their coordinates predicted by fully connected layers. Authors Neural Network has 24 convolutional layers followed by 2 fully connected layers. They had trained the neural network for 135 epochs in training and validation data sets from PASCAL VOC 2007 and 2012. While this YOLO model is tested in 2012 they also include the VOC 2007 test-data for training. In the test of VOC 2012 data set, YOLO scores approximately 57.9 percent mAP. Resulting YOLO model has good performance on VOC 2007 test set. At high-resolution new YOLOv2 is 78.6 percent of mAP on VOC 2007 set.

Aysegul Ucar, Yakup Demir and el al. 2017 published a paper on Object recognition and detection with deep learning for autonomous driving applications. In an autonomous vehicle, occlusion causes wrong object recognition and object detection. Presented LMCNN- SVM model can handle these challenges easily. In this paper, authors propose a unique model hybrid Local Multiple system (LM-CNNSVM) and Support Vector Machines (SVMs) based on Convolutional Neural Networks for object recognition and pedestrian detection [25]. In this method, they divide first the whole image into local regions and use various Convolutional Neural Networks to extract discriminative features. After that, they select discriminative features used by applying Principal Component Analysis after that PCA applied to the features obtained from CNN in order to de-correlate and reduce them. Lastly, they fuse SVM outputs. They use we used a pre-trained Alex-Net architecture and a new Convolutional Neural Networks architecture including nine layers. After that, they have tested object recognition and pedestrian detection system on Caltech-101 and Caltech Pedestrian datasets.

V. CONCLUSIONS AND FUTURE SCOPE

Through this paper, we tried to present a comprehensive review of prior work conducted in machine learning and utilization of machine learning in computer vision. Also, this paper presents new advancements carried out by various researchers in the field of Real-time object detection, object recognition and deep learning in the past decade. Computer vision technique has utilized to achieve a wide variety of visual tasks, such as object detection, pattern recognition, semantic segmentation, action recognition and face recognition. Especially General-purpose object detection and recognition should be fast, accurate, and able to recognize wide varieties of objects. Similar real-time object detection algorithm like YOLO performs remarkably fast and accurate. YOLO is a real-time object detection algorithm which employs deep convolutional neural network.

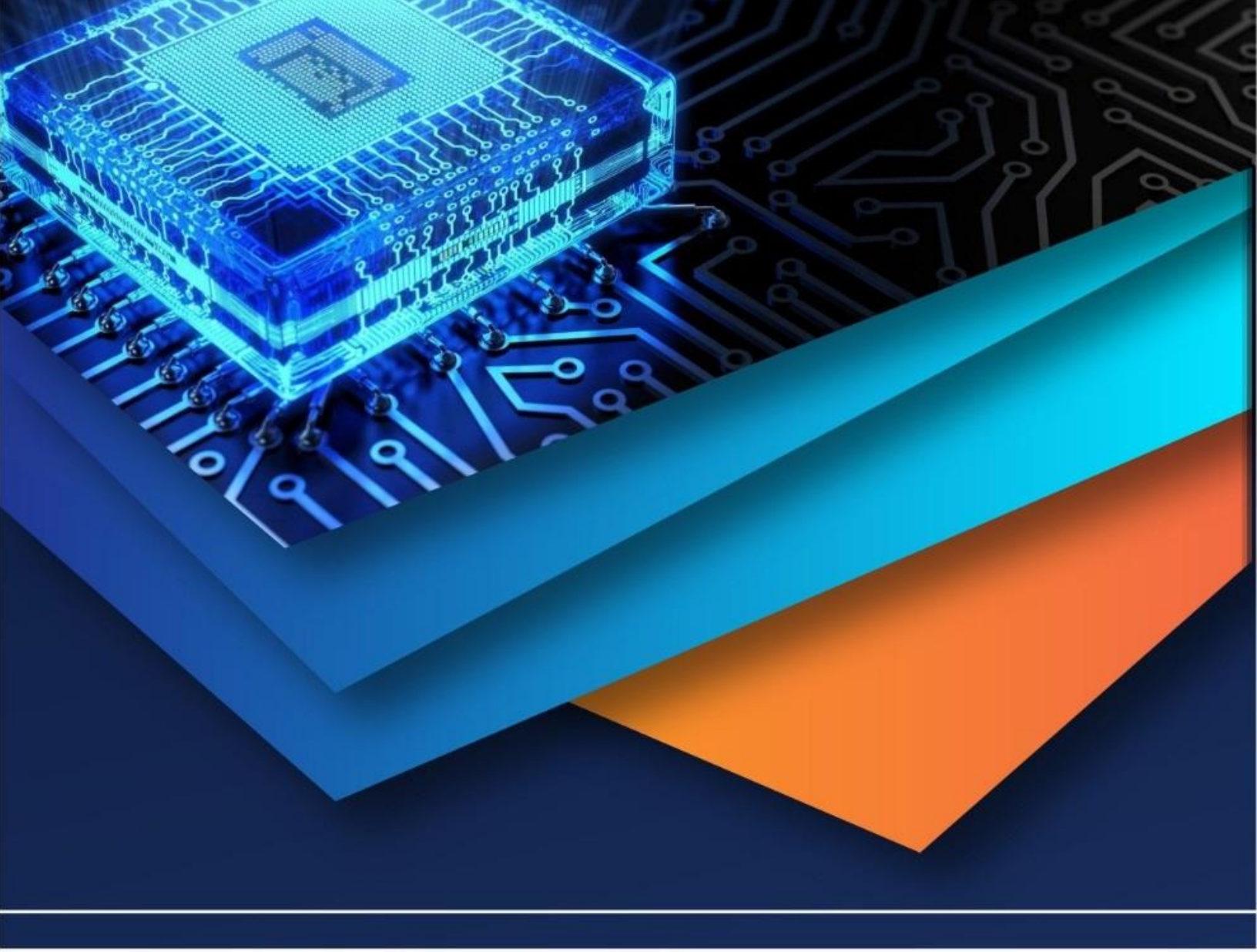
In future, we are going to see more and more applications of YOLO algorithm. It has been seen that artificial intelligence and deep learning methodology emerged in the last few decades which have been explored and utilized in different fields by various researchers with the introduction of artificial neural networks; detection technique has become more fast and accurate. Deep learning is a subset of machine learning; deep learning is the fastest-growing field of machine learning. In the deep learning method, Neural Networks has multiple hidden layers. A convolutional neural network (CNN) is a type of deep learning neural networks structure that has multiple layers. Nowadays the convolutional neural networks (CNN) considered as a powerful tool in the machine learning field. Researches show that deep convolutional neural network is capable of achieving record-breaking results also shows

excellent performance on a most challenging dataset. In some cases, convolutional neural networks effectiveness can achieve more than 90 percent of accuracy. Besides this, a Convolutional neural network is a research field that can be employed in different domains. In downside Majority of machine learning models requires lots of data and computing power. Experiments on such massive dataset are difficult and time-consuming requires lots of computational power. Neural networks perform adequately fast on Graphics Processing Units because They have a large number of cores, which can process multiple calculations simultaneously. GPU can handle multiple computations in parallel which make GPU ideal for neural network training. With the invention of powerful GPUs and lots of standard data set made Deep learning models possible. Due to the scope of this paper, it is very challenging to discuss every application of neural networks. Still, the machine learning field is in its infancy, machine learning undoubtedly shape the future. Indeed we can see that machine learning have tremendous applications in the near future.

REFERENCES

[1] Joseph Redmon, Santosh Divvala, Ross Girshick and Ali Farhadi "You Only Look Once: Unified, Real-Time Object Detection" University of Washington, Allen Institute for AI, Facebook AI Research, arXiv: 1506.02640v5 [cs.CV] 9 May 2016.
[2] Sumit Das and Aritra Dey "Applications of Artificial Intelligence in Machine Learning: Review and Prospect" International Journal of Computer Applications (0975 – 8887) Volume 115 – No. 9, April 2015.
[3] https://towardsdatascience.com/
[4] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," Bulletin of Mathematical Biology, vol. 5, no. 4, pp. 115–133, 1943.
[5] Subhashini Venugopalan , Marcus Rohrbach and others "Sequence to Sequence – Video to Text" arXiv:1505.00487v3 [cs.CV] 19 Oct 2015.
[6] R. Girshick, Jeff Donahue, Trevor Darrell and Jitendra Malik "Fast R-CNN" arXiv preprint arXiv: 1504.08083, 2015.
[7] Shaoqing Ren and R. Girshick "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" arXiv: 1506.01497 [cs.CV].
[8] M. Everingham, L. Van Gool, C. K. I.Williams, J.Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2007 (VOC2007) Results, 2007.
[9] M. Everingham, L. Van Gool, C. K. I.Williams, J.Winn, and A. Zisserman. The PASCAL Visual Object Classes Challenge 2010 (VOC2010) Results, 2010.
[10] P. Dollár, C. Wojek, B. Schiele and P. Perona "Pedestrian Detection: An Evaluation of the State of the Art" PAMI, 2012.
[11] Paul Viola and Michael J Jones. "Robust real-time face detection" IJCV, 57(2):137–154, 2004.
[12] C. Chen, A. Seff, A. Kornhauser and J. Xiao "Deep Driving: Learning Affordance for Direct Perception in Autonomous Driving" 15th IEEE International Conference on Computer Vision ICCV, 2015.
[13] J. E. Hoo, and K. C. Lim. "Accuracy and Error Study of Horizontal and Vertical Measurements with Single View Metrology for Road Surveying" ARPN Journal of Engineering and Applied Sciences. 11(12):7872-6, 2016.
[14] Yonglong Tian, Ping Luo, Xiaogang Wang, and Xiaoou Tang "Pedestrian detection aided by deep learning semantic tasks" In CVPR, 2015.
[15] Lee CS, Tyring AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. "Deep-learning based automated segmentation of macular edema in optical coherence tomography" Biomed Opt Express. 2017;8(7):3440-3448. Published 2017 Jun 23. doi:10.1364/BOE.8.003440.
[16] https://www.forbes.com/sites/bernardmarr/2019/04/08/7
[17] Marra, Francesco & Poggi, Giovanni & Sansone, Carlo & Verdoliva, Luisa (2017),"A Deep Learning Approach for Iris Sensor Model Identification. Pattern Recognition" Letters. 113. 10.1016/j.patrec.2017.04.010".
[18] Ahlin, Konrad et al. "Autonomous Leaf Picking Using Deep Learning and Visual-Servoing." IFAC-Papers On Line 49 (2016): 177-183.
[19] Liu, Po-Yu, and Edmund Y. Lam. "Image reconstruction using deep learning." arXiv preprint arXiv: 1809.10410 (2018).
[20] Lee, H., Tajmir, S., Lee, J. et al. "Fully Automated Deep Learning System for Bone Age Assessment." J Digit Imaging **30,** 427–441(2017).
[21] Paul Viola and Michael Jones "Rapid Object Detection using a Boosted Cascade of Simple Features" in 2001.
[22] Dalal and B. Triggs. "Histograms of oriented gradients for human detection" In CVPR, 2005.
[23] Alex Krizhevsky, Ilya Sutskever ad Geoffrey E. Hinton "ImageNet Classification with Deep Convolutional Neural Networks" January 2012.
[24] R. Girshick, J. Donahue, T. Darrell, and J. Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation" In CVPR, 2014.
[25] Uçar, Ayşegül, et al. "Object Recognition and Detection with Deep Learning for Autonomous Driving Applications." SIMULATION, vol. 93, no. 9, Sept. 2017, pp. 759–769, doi: 10.1177/0037549717709932.
[26] J. Long, E. Shelhamer, and T. Darrell "Fully convolutional networks for semantic segmentation," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR'15), pp. 3431–3440, IEEE, Boston, Mass, USA, June 2015.
[27] S. Ren, K. He, R. Girshick, and J. Sun. "Faster R-CNN: Towards real-time object detection with region proposal networks" arXiv: 1506.01497, 2015.
[28] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," Bulletin of Mathematical Biology, vol. 5, no. 4, pp. 115–133, 1943.
[29] Mxnet, Available online: http://mxnet.io.
[30] W. Ouyang, X. Zeng, X. Wang et al., "Deep ID-Net: Object Detection with Deformable Part Based -amazing-examples-of-computer-and-machine-vision-in-practice/#4da402eb1018.
[31] A. Diba, V. Sharma, A. Pazandeh, H. Pirsiavash, and L. V. Gool, "Theyakly Supervised Cascaded Convolutional Networks," in Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5131–5139, Honolulu, HI, July 2017.
[32] N. Doulamis and A. Voulodimos, "FAST-MDL: Fast Adaptive, Supervised Training of multi-layered deep learning models for consistent object tracking and classification," in Proceedings of the 2016 IEEE International Conference on Imaging Systems and Techniques, IST 2016, pp. 318–323, October 2016.
[33] N. Doulamis, "Adaptable deep learning structures for object labelling/tracking under dynamic visual environments," Multimedia Tools and Applications, pp. 1–39, 2017.
[34] L. Lin, K.Wang,W. Zuo,M.Wang, J. Luo, and L. Zhang, "A deep structured model with radius-margin bound for 3D human activity recognition," International Journal of Computer Vision, vol. 118, no. 2, pp. 256–273, 2016.

[35] S. Cao and R. Nevatia, “Exploring deep learning based solutions in fine grained activity recognition in the wild,” in Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR), pp. 384–389, Cancun, December 2016.
[36] A. Toshev and C. Szegedy, “Deep Pose: Human pose estimation via deep neural networks,” in Proceedings of the 27th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2014, pp. 1653–1660, USA, June 2014.
[37] H.Noh, S.Hong, and B.Han, “Learning de convolution network for semantic segmentation,” in Proceedings of the 15th IEEE International Conference on Computer Vision, ICCV 2015, pp. 1520–1528, Santiago, Chile, December 2015.
[38] J. Long, E. Shelhamer, and T. Darrell, “Fully convolutional networks for semantic segmentation,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR’15), pp. 3431–3440, IEEE, Boston, Mass, USA, June 2015.
[39] K. Fukushima, “Neocognitron: a self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” Biological Cybernetics, 1980.
[40] P. Arbel´aez, B. Hariharan, C. Gu, S. Gupta, L. Bourdev, and J. Malik. "Semantic segmentation using regions and parts" In CVPR, 2012.
[41] H. A. Rowley, S. Baluja and T. Kanade. "Neural network based face detection" TPAMI, 1998.
[42] P. Sermanet, K. Kavukcuoglu, S. Chintala, and Y. LeCun. "Pedestrian detection with unsupervised multi-stage feature learning". In CVPR, 2013.
[43] K. Simonyan and A. Zisserman."Very Deep Convolutional Networks for Large-Scale Image Recognition". arXiv preprint, arXiv:1409.1556, 2014.
[44] R. Girshick, P. Felzenszwalb, and D. McAllester. "Discriminatively trained deformable part models".
[45] Tai Sing Lee and David Mumford "Hierarchical Bayesian inference in the visual cortex" 1434 J. Opt. Soc. Am. A/Vol. 20, No. 7/July 2003.
[46] https://en.wikipedia.org/wiki/Machine_learning
[47] Karpathy, Andrej, and Li Fei-Fei. "Deep visual-semantic alignments for generating image descriptions." Proceedings of the IEEE conference on computer vision and pattern recognition. 2015.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY