iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** IX **Month of publication:** September 2022

DOI: https://doi.org/10.22214/ijraset.2022.46702

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: 7.538
Volume 10 Issue IX Sep 2022- Available at www.ijraset.com

Literature Review of Moving Object Detection Using Machine Learning

Sanjiw Kumar[1], Dr. Anant Kumar Sinha[2], Dr. Narendra Kumar[3]
[1]*Research Scholar, Magadh University, Bodh-Gaya*
[2]*Associate Prof. & Head, Dept. of Physics, A. M. College, Gaya*
[3]*Assistant Prof., Dept. of Computer Applications and IT, A. M. College, Gaya*

Abstract: ***In order to answer three different research questions, this paper conducts a literature review on the subject of machine learning in object detection security. The work done covers more than 34 research papers that are relevant to this topic. Charts, tables, and statistics are used to summarize the data and provide the reader with an easily readable summary of the papers that are relevant to this topic.***

The detection of moving objects in videos and video surveillance, which are both significant and difficult tasks in many computer vision applications, are covered in this chapter's review and systematic investigation. Detection algorithms for humans, cars, threats, and security, for example. One of the most difficult research topics in computer vision right now is video surveillance in dynamic environments, especially for people, vehicles, and specific objects in cases of security. This technology is essential to the fight against terrorism, crime, and public safety, as well as for the effective management of accidents and the crime that is seen happening these days. The concept of real-time computing task implementation in video surveillance systems is also presented in the study. In this review study, numerous systems are evaluated in order to determine how well they can track moving objects in an indoor or outdoor area in real time.

Keywords: Computer Vision, Human Detection, Mobile Video Surveillance, Moving Object Detection, Background Subtraction, ML Algorithms.

I. INTRODUCTION

In order to detect, recognize, and track things throughout a series of photos as well as to interpret and explain object behavior, video surveillance systems are being developed to replace the dated conventional practice of having human operators watch cameras. Many computer vision applications, such as surveillance, vehicle navigation, and autonomous robot navigation, involve the crucial and difficult tasks of object identification and tracking. Finding items within a frame of a video sequence is the task of object detection. A mechanism for object detection is necessary for every tracking technique, either in each frame or at the moment the item first shows in the video.

The practice of tracking an object or group of objects over time with a camera is known as object tracking. There is a lot of interest in object tracking algorithms because of the powerful computers, the accessibility of high-quality, low-cost video cameras, and the growing need for automated video analysis. Three essential processes are involved in video analysis: finding intriguing moving objects, following them from frame to frame, and analyzing the behavior of the tracked items. Consequently, the utilization of object tracking is relevant for motion-based recognition tasks.

Machine Learning (ML), a branch of Artificial Intelligence (AI) that focuses on teaching a computer how to learn, is crucial to the success of object detection. Supervised Learning (SL), Unsupervised Learning (UL), and Reinforcement are the three types of machine learning. The objective behind SL is to provide an input-output pair for the ML model to train on. UL refers to the concept of providing a high volume of unlabeled data to an ML model so that it can learn from and identify it. The premise behind reinforcement learning is that the ML model will improve with time.

The term "image processing" refers to the processing of an image or video frame that is used as input, with the outcome set of processing possibly being a set of related picture parameters. Visualization, or observing things that are not visible, is the goal of image processing. One of the newest and most well-liked research areas in digital image processing is the study of human motion. The goal is to identify human motions from the backdrop image in a video sequence where human movement is a key component of human detection and motion analysis.

Along with some moving items from the video frame, it also detects, tracks, and recognises human activity. Understanding human activity from a video is a significant area of computer vision research that has become increasingly important in recent years.

©IJRASET: All Rights are Reserved

Recent advancements in computer vision, the accessibility of inexpensive hardware like video cameras, and a range of new and intriguing applications like visual surveillance and personal identification are all major drivers of the expanding interest in human motion analysis. Recognizing motion of objects in the two provided photos is the aim of motion detection. Finding an object's motion can also help with object recognition.

Automated monitoring, sometimes referred to as Intelligence Visual Surveillance (IVS), involves object analysis and interpretation in addition to object detection and tracking to identify the scene's visual actions. Wide area surveillance control and scene interpretation are the primary duties of IVS. Numerous changes in lighting and well-known difficulties must be taken into account while object tracking. Generally speaking, video analysis can be divided into three main stages: identifying moving objects, determining an object's journey from one frame to the next, and examining entity tracks to assess their performance. It is significantly simpler to track things in a static environment than in a dynamic one.

II. APPLICATIONS OF OBJECT DETECTION

There are numerous publications that discuss the use of object detection in security. Object detection has applications in many different fields. Despite the fact that object detection is frequently utilized in physical security applications such as airport baggage, x-ray object detection, fraud detection, moving vehicle object detection, abandoned item detection, face detection, and pedestrian object detection. Additionally, it has uses in cyber security, such as smart-phone, facial detection. Additionally, they pose certain challenges for online human verification tools like reCAPTCHA, which enables experts in that sector to develop more effective defenses.

III. MACHINE LEARNING IN OBJECT DETECTION

Machine Learning (ML), a branch of Artificial Intelligence (AI) that focuses on teaching computers how to learn, is crucial to the success of object detection. There are three Machine Learning sub-types. These are:

A. Supervised Learning

Supervised learning is a sort of machine learning in which the output is predicted by the machines using well-labeled training data that has been used to train the machines. The term "labelled data" refers to input data that has already been assigned the appropriate output. In supervised learning, the labeled training data that is given to the computers serves as the supervisor, instructing them on how to correctly predict the output. It employs the same idea that a pupil would learn under a teacher's guidance.

The method of supervised learning involves giving the machine learning model the right input data as well as the output data. Finding a mapping function to link the input variable (x) with the output variable is the goal of a supervised learning algorithm (y)

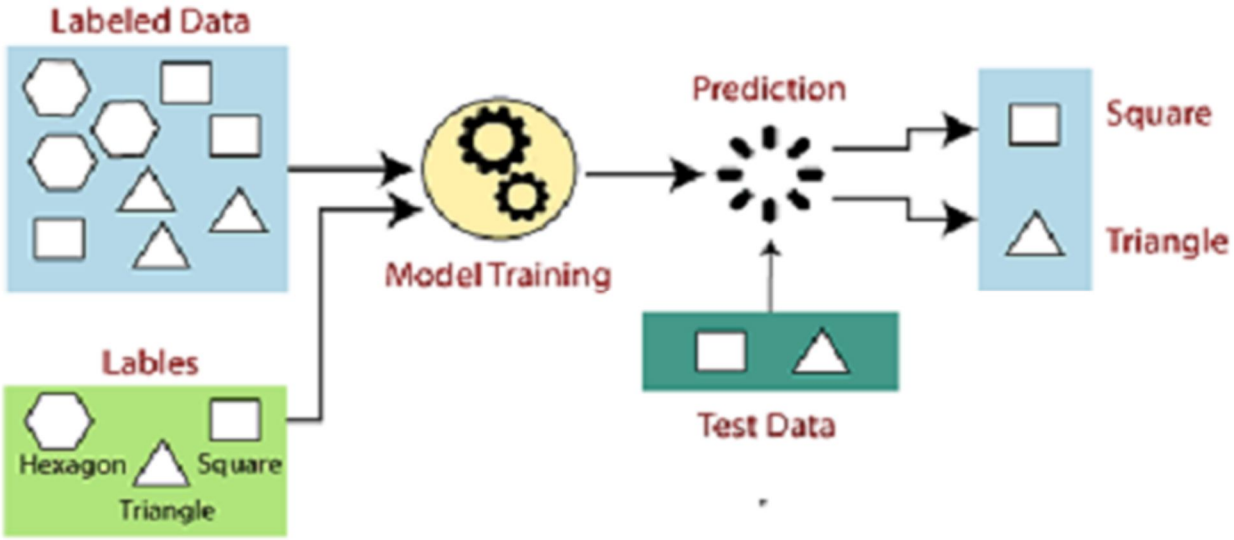


Figure 1: Supervised ML
Source: “https://www.javatpoint.com/supervised-machine-learning”

B. Unsupervised Learning

Unsupervised learning is a subcategory of machine learning in which models are trained using un-labeled datasets and are free to operate on the data without involving any supervision. Unsupervised learning is a type of machine learning in which models are not supervised using training datasets, as the name implies. Instead, models themselves decipher the provided data to reveal hidden patterns and insights. It is comparable to the learning process that occurs in the human brain while learning something new.

Unlike supervised learning, we have the input data but no corresponding output data, unsupervised learning cannot be used to solve a regression or classification problem directly. Finding the underlying structure of a dataset, classifying the data into groups based on similarities, and representing the dataset in a compressed format are the objectives of unsupervised learning.

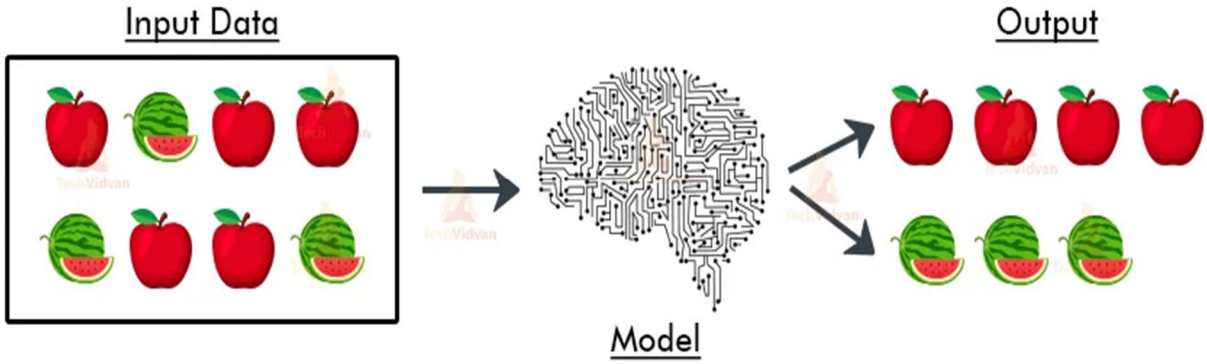


Figure 2: Unsupervised ML
Source: https://techvidvan.com/tutorials/unsupervised-learning/

C. Reinforcement Learning

A machine learning training method called reinforcement learning rewards desired behaviors and/or penalizes undesirable ones. A reinforcement learning agent can typically perceive and comprehend its surroundings, act, and learn by making mistakes. By executing actions and observing the outcomes of those actions, an agent learns how to behave in a given environment via reinforcement learning, a feedback-based machine learning technique. The agent receives compliments for each positive activity, and is penalized or given negative feedback for each negative action.

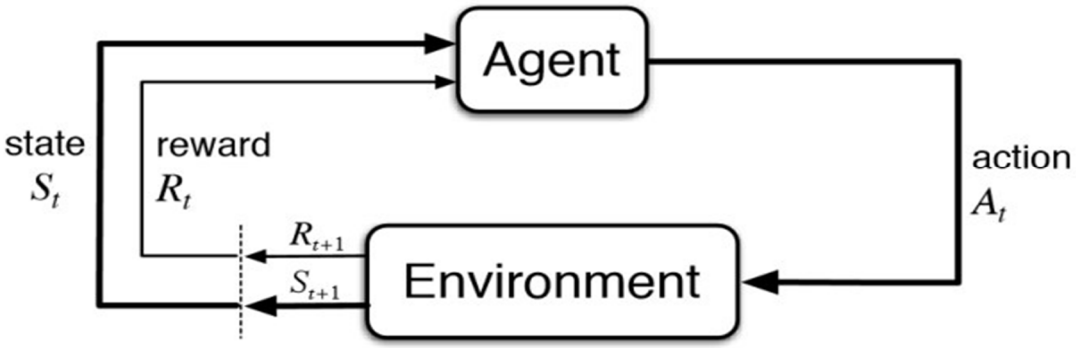


Figure 3: Reinforcement ML
Source: "https://www.kdnuggets.com/2018/03/5-things-reinforcement-learning.html"

IV. LITERATURE REVIEW

Among the many uses for object detection, security is one of the most crucial ones. Of course, there are many applications of object detection utilizing machine learning in the realm of security. The topic that we opted to conduct a literature review on has not yet been covered by a publication, as far as we can tell. And to our knowledge, no one has conducted a thorough literature study on the subject we have chosen. The two studies that are the closest to being a survey that we have discovered are either specific or cover particular methodologies rather than analyze other papers.

A. A study on several techniques for object detection and tracking in video surveillance footage was done in one work by Murugan et al. [1]. The paper highlights how video surveillance has existed as a technology since the 1950s and reinforces the point stated in the Introduction section about how watching security cameras can be tiresome and have an adverse effect on a person's mental health. It also illustrates how automating the procedure was the answer to that tedious chore by introducing a sophisticated surveillance system. This paper's main goal is to describe the various techniques for object detection and tracking in videos. The first technique described in the study is backdrop subtraction. The report claims that one of the most popular techniques for spotting moving items is backdrop removal. To show only the moving object's pixels, background subtraction includes identifying the background and removing it. The issue with it is that the results of the procedure are impacted if the background is not static and changes as a result of illumination or specific weather conditions. There are other algorithms for background subtraction as well.

B. A study by Flitton et al. [2] compares various 3D interest point descriptors for CT images of baggage at airports. Finding interesting stuff during baggage x-ray inspections is the core concept here. Five distinct methods were compared in the paper: density, density histogram, density gradient histogram, rotation invariant feature transform, and scale invariant feature transform. The report does a decent job of assessing each while providing crucial indicators, although the research was only done on a relatively narrow issue. Both of the aforementioned studies do a great job of describing the various strategies, but in comparison to our paper, they are too narrow in scope. Instead of separating the focus on machine learning and object detection, both publications provide deeper explanation of the object detection techniques. The two articles cover fewer papers than our Systematic Literature Review does because of their narrower scopes.

C. Bezak, P. (September - 2016) suggests a deep learning approach for item recognition in historical architecture images in Trnava. For jobs requiring object recognition, it employs deep learning architectures based on CNN (Convolutional Neural Network) [3]. Architecture is improved by using activation functions and a cascade of convolutional layers. Setting up the number of layers and the number of neurons in each layer is crucial. For this, the TRNAVA LeNet 10 model was created and trained. This model is based on a dataset of 460 training images and 140 validation images, or a 3:1 ratio. The images are 28x28 pixels in size, were color photographs, and were encoded in jpg format. The model correctly identified the correct item in the picture of the Trnava historical building. The proposed model's prediction accuracy increased to 98.88%.

D. Jung, H., Lee, S. et al (Jan – 2015), suggested that instead of using manually produced characteristics, deep learning techniques should be used to recognize face expressions. Convolutional neural networks (CNN) and deep neural networks (DNN) are two types of deep networks that are used to tackle recognizing difficulties [4]. The deep networks were created quickly utilizing deep learning toolkits that enable CUDA, such as Caffe and CudaConvnet2. Additionally, they made use of the OpenCV library to create the Haar-like face detection technique. The photographs were reduced in size and cropped to 64 * 64. The 327 face photos were then separated into ten groups, one of which was utilized for training and the other nine for testing. For six emotions, the recognition rates were good, but the disgust label had a low recognition rate. Because there were only 547 training photos for the disgust label in the FER 2013 database. Over-fitting is a possibility with the DNN.

E. Tenguria, R., Parkhedkar et al (April – 2017), Convolutional neural networks have been replaced by more precise yet sophisticated approaches that can recognize things in real-time, according to the study [5]. This paper promises a significant advance in object recognition and tagging. However, development in this area has been somewhat modest. It intends to combine the domains of computer vision and robotics, with a particular emphasis on the implementation of image description applications on an embedded system platform. A fixed number of items are allowed in the image, according on the data set used to train the model. Shaoqing Ren et al. claim that the development of the Region Proposal Network (RPN) permits sharing of the entire image's convolutional characteristics with the network, resulting in nearly free region proposals. In this case, the algorithm is directed by the region suggestion technique in order to find objects present in the image. Second, the application of this technique in our system makes it computationally efficient and tailored to function on low-powered platforms.

F. Etemad, E., & Gao, Q. (Sep. – 2017): In order to improve the effectiveness of current object recognition algorithms, the research [6] presents an object localization method that makes use of image edge information as a cue to pinpoint the locations of the objects. The perceptual organisation components of human vision are used to extract the Generic Edge Tokens (GETs) of the image. To precisely locate objects, these edge tokens are parsed using the Best First Search method, with the detection score provided by the Deep Convolutional Neural Network serving as the goal function. The search space is a collection of edge elements whose overlaps with the current candidate object are greater than zero when the BFS is applied to the object localization and its search space. Real-time testing revealed that the model was more effective than the RCNN, and there is still

room for improvement by enhancing object localization by combining picture edge, color and texture information, and the learnt properties of the image.

G. Mazumdar, M., Sarasvathi, V. and Kumar, A. (Aug. - 2017), suggested a technique for creating an interactive application to identify things from films; upon user input, it is also capable of identifying the specific object now displayed on the screen [7]. For this challenge, which has a 77% accuracy rate, a sequential frame extraction technique for films as well as a deep learning strategy utilizing Convolutional and Fully Connected Neural Networks are applied. The work of computer vision is still fairly difficult even when the item is somewhat warped, translated, rotated, or partially obscured from view and can still be easily spotted by humans. Taking advantage of the fact that videos are composed of frames synced with some playback audio, the analysis of the video can be done in much more detail by looking at the objects present in the frame images themselves, running the classifier to obtain probabilities for various classes, and then classifying the genre as well as identifying any objects in the video. By adding more datasets and optimizing the hardware setup, this model's operational accuracy is increased, enabling faster and more accurate item categorization over a wider variety of classes.

H. Sujana, S. R., Abisheck, S. S., Ahmed, A. T., & Chandran, K. S. (2017), suggests a method for using convolutional neural networks and the idea of deep learning to identify things [8]. It generates an output with the collection of recognized objects using the input video. For each of the objects, the convolutional neural network calculates a confidence score. It employs Single Shot Multi-box Detector, which had a high accuracy rate and used convolutional networks to identify many items simultaneously. In order to raise the object's confidence score and produce only one detection for each object, respectively, it employs Hard Negative Mining and Non-Maximum Suppression. Thus, these aid in selecting the highest score and prevent multiple item detection. Thus, using neural networks in conjunction with deep residual networks improves item identification's computing efficiency and accuracy.

I. Cheng, S. C. (2005) has formulated an approach based on motion estimation with a block-based moment-preserving edge detector is proposed. In this study [9] as an object-based coding technique for very low bit-rate channels. The object-based coding techniques that leverage global motion components are the most popular. When images have noise and quickly moving objects, such as cars, the global motion components have a difficulty with significant prediction errors even after motion compensation using the discrete cosine transform (DCT). The segmented objects cannot include sub-objects that move in separate directions if there is a modest prediction error. The method put forth in this paper uses a hybrid object-based video-coding approach that minimizes the drawbacks of both the object-based and block-based coders while retaining their relative benefits, namely the ability to compactly represent objects by visual-pattern approximations of their boundaries and segment moving objects from video sequences. It is accomplished by employing the moment preserving edge detector to identify the line edge from a square block. Block matching is needed for motion estimation due to the high computational cost. A quick block-matching method that is based on visual patterns is employed to simplify this problem. According to the results, the suggested method is effective in terms of subjective quality, peak signal-to-noise ratio (PSNR), and compression ratio.

J. Alexeev, A., Matveev, Y., and Kukharev, G. (2018), In this study [10], an advanced and-new object detection algorithm is proposed that makes use of a neural network with a Network in Network (NiN) type convolution kernel to enable highly parallel processing. Only when the convolution kernel is applied in the form of a fully linked network does its non-linear approach allow for a big stride and the abandonment of pooling. Detection is the process of simultaneously locating items on a picture and recognizing them. Detector can work with pictures of any size. Due to the algorithm's strong computational efficiency, the processing time may rise by up to 300 ms while processing HD frames on a single CPU core. The high degree of regularity in network operations that results in massively parallel data processing on GPU is likely to shorten the operating time to less than 10 ms. The suggested technique can handle minor overlaps and average-quality photos of objects being recognized. An image's bounds and classes of objects serve as output delimiters for an end-to-end learner model. In order to assess the algorithm, an open access image database is used. This technique may simultaneously detect a variety of items; it is not constrained to the use of a single type of object. The suggested algorithm efficiently processes images at a high rate that is comparable to that of previous algorithms.

K. Wong, S. C et al (2017). The issue of tracking and classifying many objects in an image sequence online is dealt with in this study. The answer is to first track every object in the image without relying on object-oriented prior information. This can be done by using hand-crafted features or user-based track initialization. The tracked objects are classified using a fast-learning image classifier with a shallow convolutional neural network design [11]. When this classifier is paired with object state data from the tracking technique, object recognition accuracy increases. By transferring the past information from the detection and tracking stages to the classification stage, a reliable, all-purpose object identification system with the capacity to detect and

track a range of object kinds can be achieved. The Neovision2 Tower data set contains a number of tracked objects, and the algorithm adapts to learn their shape and motion. By utilizing object-specific previous information in identification and tracking, the strategy is competitive, and an examination shows that it also offers extra practical benefits due to its generality.

L. Yang, L., Wang, L., & Wu, S. (2018). The presented article [12] suggests Object detection with an image is essentially what object confirmation is. Three steps are involved in conventional object identification algorithms: region selection, feature extraction, and classification. By applying a single deep convolutional neural network to the image, YOLOv2 reframes object detection as a single regression problem that goes straight from image pixels to bounding box coordinates and class probabilities at the same time. Multi-feature fusion has emerged as a new paradigm in the architecture of deep convolutional neural networks in recent years. Low-layer filters capture the detail texture information of objects, while high-layer filters extract the semantic information. The high sensitivity of radar detection and the high object confirmation accuracy are both features of the intelligent radar perimeter security system.

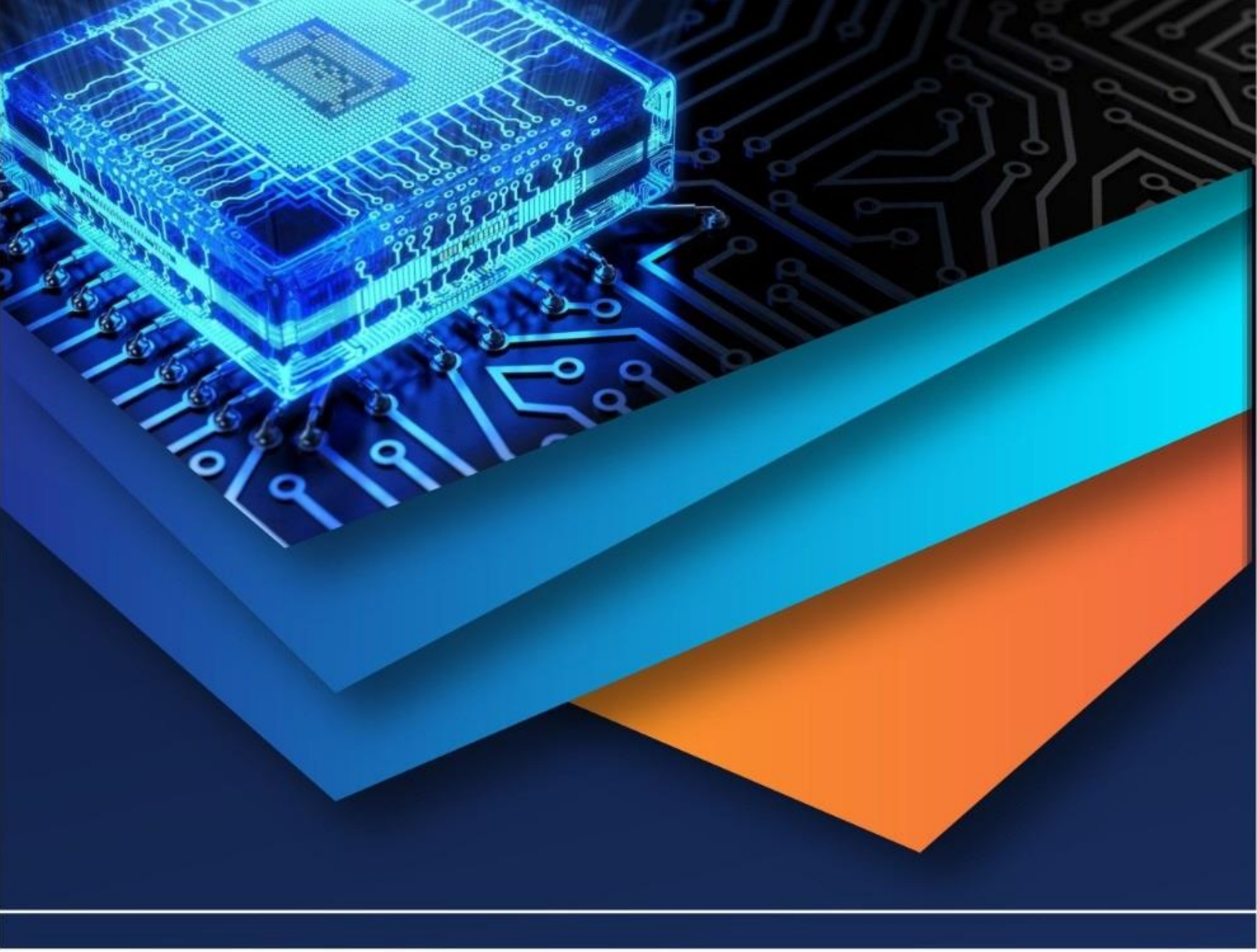
V. CONCLUSION

One of the newest and most fascinating fields in machine learning is object recognition. Face detection is a popular use of object detection that is present in practically all smart-phone cameras. However, the main drawbacks in all of these articles are the real-time applications and accuracy rates attained, which can be fixed by using ImageNet's superior Inception Model architecture. These systems can be combined with other tasks like pose estimation, where detecting the item comes first in the pipeline and estimating pose in the observed region comes second. It may be used to track objects, making robots and medical applications possible. This problem therefore has a wide range of uses. Due to their short training times, low latency, quick evaluation, etc., machine/deep learning models are the most effective for object detection.

REFERENCES

[1] S. Murugan, K. S. Devi (2018), A. Sivaranjani, and P. Srinivasan, “A study on various methods used for video summarization and moving object detection for video surveillance applications,” Multimed. Tools Appl., vol. 77, no. 18, pp. 23273– 23290.
[2] G. Flitton, T. P. Breckon, and N. Megherbi, “A comparison of 3D interest point descriptors with application to airport baggage object detection in complex CT imagery,” Pattern Recognit., vol. 46, no. 9, pp. 2420–2436, 2013.
[3] Bezak, P. (2016, September). Building recognition system based on deep learning. In 2016 Third International Conference on Artificial Intelligence and Pattern Recognition (AIPR) (pp. 1-5). IEEE.
[4] Jung, H., Lee, S., Park, S., Kim, B., Kim, J., Lee, I., & Ahn, C. (2015, January). Development of deep learning-based facial expression recognition system. In 2015 21st Korea-Japan Joint Workshop on Frontiers of Computer Vision (FCV) (pp. 1-4). IEEE.
[5] Tenguria, R., Parkhedkar, S., Modak, N., Madan, R., & Tondwalkar, A. (2017, April). Design framework for general purpose object recognition on a robotic platform. In 2017 International Conference on Communication and Signal Processing (ICCSP) (pp. 2157-2160). IEEE.
[6] Etemad, E., & Gao, Q. (2017, September). Object localization by optimizing convolutional neural network detection score using generic edge features. In 2017 IEEE International Conference on Image Processing (ICIP) (pp. 675-679). IEEE.
[7] Mazumdar, M., Sarasvathi, V., & Kumar, A. (2017, August). Object recognition in videos by sequential frame extraction using convolutional neural networks and fully connected neural networks. In 2017 International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS) (pp. 1485-1488). IEEE.
[8] Sujana, S. R., Abisheck, S. S., Ahmed, A. T., & Chandran, K. S. (2017, April). Real time object identification using deep convolutional neural networks. In 2017 International Conference on Communication and Signal Processing (ICCSP) (pp. 1801-1805). IEEE.
[9] Cheng, S. C. (2005). Visual pattern matching in motion estimation for object-based very low bit-rate coding using moment-preserving edge detection. IEEE transactions on multimedia, 7(2), 189-200.
[10] Alexeev, A., Matveev, Y., and Kukharev, G. (2018, October). Using a Fully Connected Convolutional Network to Detect Objects in Images. In 2018 Fifth International Conference on Social Networks Analysis, Management and Security (SNAMS) (pp. 141-146). IEEE.
[11] Wong, S. C., Stamatescu, V., Gatt, A., Kearney, D., Lee, I., and McDonnell, M. D. (2017). Track everything: Limiting prior knowledge in online multi-object recognition. IEEE Transactions on Image Processing, 26(10), 4669-4683.
[12] Yang, L., Wang, L., & Wu, S. (2018, April). Real-time object recognition algorithm based on deep convolutional neural network. In 2018 IEEE 3rd International Conference on Cloud Computing and Big Data Analysis (ICCCBDA) (pp. 331-335). IEEE.
[13] https://www.javatpoint.com/supervised-machine-learning
[14] https://techvidvan.com/tutorials/unsupervised-learning/
[15] https://www.kdnuggets.com/2018/03/5-things-reinforcement-learning.html

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)