Motion Detection and Human Activity Recognition for Security

Mohammed Arham
Dept. of Information Science and Engineering
Vidyavardhaka College of Engineering
Mysore
4vv19is049@vvce.ac.in

Amisha Srivastava
Dept. of Information Science and Engineering
Vidyavardhaka College of Engineering
Mysore
4vv19is005@vvce.ac.in

Akshatha G
Dept. of Information Science and Engineering
Vidyavardhaka College of Engineering
Mysore
4vv19is004@vvce.ac.in

Rajendra A B
Dept. of Information Science and Engineering
Vidyavardhaka College of Engineering
Mysore
hodis@vvce.ac.in

***Abstract* — A key problem in computer vision called object detection involves locating and identifying objects in an image or video. Deep learning-based methods for object identification have recently produced cutting-edge results., particularly using convolutional neural networks (CNNs). However, real-time object detection remains a challenging problem, requiring great accuracy and rapid response. In this project, we propose a real-time object detection system using CNNs. We train the model on a large dataset of annotated images, enhancing performance via approaches such as data augmentation and transfer learning. We then optimize the model for deployment on a target hardware platform, such as a mobile device or embedded system. We illustrate the efficacy of our model. in several real-world applications, such as pedestrian detection and tracking in crowded urban environments, vehicle detection and tracking in traffic scenes, and object detection and tracking in industrial settings. Our system provides a scalable and robust solution for multi-object detection and tracking using deep learning and computer vision techniques.**

Keywords—Computer vision, Convolutional neural network (CNNs), deep-learning.

1. INTRODUCTION

Object detection is a crucial task in computer vision that involves identifying and localizing objects within an image or video. This is a challenging problem, as objects may vary in appearance, size, and orientation, and may be occluded by other objects or have complex backgrounds.

Traditionally, object detection involved handcrafting features and designing algorithms to detect specific objects. However, this approach was limited by its reliance on prior knowledge and the ability to design effective features.

In recent years, machine learning, computer vision, and AI techniques have revolutionized object detection, leading to significant advances in accuracy and efficiency. Deep learning-based approaches, in particular, have shown great success in object detection, due to their capacity to learn complicated characteristics straight from data.

Convolutional neural networks (CNNs) are a popular deep learning architecture used in object detection. These networks consist of multiple layers that learn increasingly complex features from the input data. Using a big dataset of annotated images to train a CNN, it can learn to recognize objects and their features, such as edges, corners, and textures. Object detection can be formulated as a classification problem, where the task is to predict the class of each object in an image. It can also be formulated as a regression problem, where the task is to predict the location and size of each object. Many object detection algorithms combine these two approaches, predicting the class of each object and its location within the image.

Object detection is crucial in many applications, but related tasks such as object tracking and multi-object detection and tracking are also significant.sss, such as surveillance, robotics, and autonomous driving. These tasks involve tracking the position and movement of objects over time and can be particularly challenging in crowded or dynamic environments.

Object detection using machine learning, computer vision, and AI has many practical applications, ranging from security cameras to autonomous vehicles. By accurately and efficiently detecting and tracking objects, these systems can improve safety, efficiency, and decision-making in a variety of domains.

2. BACKGROUND AND RELATED WORK

In the research work ‘A Motion Detection System using Python and Opencv’, a motion detection software

system has been introduced which allows us to sense movement about a target or a visual area using python and opencv. This is done by capturing the video frames and then creating a grayscale version of the captured frame. The initial frame will be used as the starting frame. The movement is determined by the phase difference between the old and new frames. Newer frames are going to be called Delta frames. A Dilate function is used to filter the image. Moreover, it could be possible to get the time and date even when the thing is in and out of the frame. A status list is used to record this timestamp. Status list stored value 0 or 1, 0 indicates no results were found, while 1 indicates an item was discovered. The instant the item enters the frame is when this state value switches from 0 to 1, as soon as the object leaves the frame, this value changes from 1 to 0.. The time stamps of these two conversion occurrences are recorded and it is saved into a csv file. In the end, the duration of an object in front of the camera and the frequency with which a moving object is picked up will be shown on a graph. [1].

The main subject of this research 'A Deep Learning Approach for Face Detection using Yolo' is to improve the accuracy of face detection using a deep learning model. To carry out the suggested study, YOLO, a popular deep learning library, is used. The model is trained and tested using the FDDB dataset. Because it is rapid in detecting, YOLO enhances performance. During training and testing, a single neural network is applied to the entire image in YOLO (You Only Look Once). The input to the proposed network is a 448 by 448 colour picture. Seven convolutional layers make up the architecture, which is followed by a pooling layer with a maximum size of two by two. Both the class probabilities as well as the coordinates of the bounding box are predicted by the output layer using NMS (Non-Maximum Suppression) technique. The model was trained over 25 iterations using the gradient decent optimizer method. After 20 epochs, accuracy remained essentially constant at 92.2%, and after experimenting with various values, the ideal learning rate is determined to be 0.0001. As a result, both the network size and the object size have an impact on learning rate. If the object size is small and the network is large or medium-sized, the learning rate should be kept low. It is also observed that the more the number of times the dataset is trained on the network, the better the outcomes. Data overfitting can also occur; hence an optimal value of epoch size should be set at an ideal value that does not result in network overfitting or underfitting. It is also noted that if the GPU configuration is high, the work may be computed at a greater speed. The frame rate rose as the resolution dropped. Because a low-resolution image contains fewer pixels, the GPU processes it faster because there are fewer parameter computations [2].

The main subject of this research 'A Fast and Accurate System for Face Detection, Identification and Verification' is Facial analyticsIt entails extracting data such as focal points, positions, facial expressions, gender, age, identification, etc. In this article, we've presented an overview of current DCNN-based face recognition technologies. We talked about the key elements of our whole face recognition pipeline. The proposed DPSSD face detector does face detection, and the All-in-One CNN performs keypoint localization. We've also detailed our facial verification/ID module, which uses a dual-grid arrangement to represent features. We discussed the training details and data sets for our system and how they relate to existing work on facial recognition. IJBA, IJB-B, IJB-C, and IARPA Janus Challenge Set 5 were the four difficult data sets for which we observed the system's findings (CS5). We demonstrate that our team-based approach produces results that are nearly at the cutting edge. Our face verification accuracy for IJB-C is marginally worse than ArcFace at a FAR of 1e-6 to 1e-1[26]. ArcFace does not publish results for FARs of 10-7 or 10-8 and instead uses a bigger training dataset (5.8 million photos with 85k identities). Even at these ridiculously low FARs, our pipeline still generates respectable figures. Lastly, for all of the identification techniques utilised in the JANUS programme, we have provided thorough performance numbers[3].

The key takeaways from the paper "Refinement Neural Network for High Performance Face Detection" are listed below. In order to narrow the search space for classification, the STC module was developed to remove the majority of single negative samples from the low level. SML module was added to help identify faces and backgrounds on a range of scales. Encourage the FSM module to have additional differentiating characteristics for the classification task. Develop the RFE engine to offer a wider range of receptive fields for face detection in harsh environments. The AFW, PASCAL Face, FDDB, MAFA, and WIDER FACE datasets offer the best performance. The study enhances the capabilities of regression and classification to provide a novel state-of-the-art face detector. For this reason, we develop STR to generally change the positions and dimensions of the anchor from the layers of advanced detection to assure more effective initialization of the following regressor. On the one hand, boosting the regression ability can increase the position accuracy and lower the LOC error. A In the future, we intend to create a simple architecture that uses machine learning (AutoML) techniques to execute RefineFace in real time on both CPU and GPU devices[4].

We track numerous items using contemporary tools like YOLO (You Only Look Once), OpenCV, PyTorch, COCO dataset, Tkinter with MySQL in the paper "Multiple Object Tracking Using Deep Learning with YOLO V5". After object detection, the OpenCV module feeds the real-time or file format video into the algorithm. Moreover, it records and observes the items spotted in the output. We utilise YOLOV5, which is built using PyTorch classifier. Tkinter is used for designing the user interface of the application and makes it easy for user interaction. Object detection happens in three stages i.e., classification, detection and segmentation. In classification the object is classified with a unique id. The subsequent phase, detection, involves utilising trained models to identify the item in frames. The observed object is then segmented for better comprehension in the following step, which is segmentation. This system works on both on live video

and recorded video. The input data is processed into frames for detection. Primarily, three terms—mAP, Precision, and Recall—are used to assess the algorithm's effectiveness. The accuracy of the item is determined by mAP by combining recall and precision. Precision measure the accuracy of the objects that are predicted and recall defines how good the objects. Through enhancing regression and classification capabilities, the paper provides a distinctive state-of-the-art face detector. On the one hand, increasing regression capacity can improve position accuracy and reduce LOC error. To accomplish these objectives, we created STR to approximately reposition and resize the anchor from the high-level detection layers in order to improve the subsequent regressor's initialization. To execute RefineFace in real time on both a CPU device and a GPU device in the future, we aim to build a lightweight architecture using machine learning (AutoML) approaches. [4].

In the article 'Human action recognition using machine learning in uncontrolled environment', detection and classification of human action is done based on the appearance, clothes, illumination, and background. Using videos Human activity for computer vision experts, recognition is an active research area because of the various industrial applications such as traffic surveillance, automotive safety, anti-terrorist applications, rescue missions, real-time tracking Some efficient techniques to classify human actions are by taking actions like getting rid of unnecessary video frames, geodesic distance mining for feature descriptors (GD), extracting Segments of Interest (SoIs), 3D cartesian-plane features (3D-CF), Joints MOCAP (JMOCAP). For classification purpose we use a Neuro Fuzzy Classifier (NFC). In human-model based approaches, the movements and placements of body components are used to recognise human based actions. [6].

The main objective of this research 'Face Detection in Real Time Live Video Using Yolo Algorithm Based on Vgg16 Convolutional Neural Network' is to improve the proposed face detection systems. The suggested method consists of three phases. The initial phase entails create a labelled ground truth database which comprises of images and labels for each image. The next step is the features extraction stage where a VGG16 convolutional neural network that has previously been trained was used. Finally, in order to detect faces, the result of the pre- trained VGG16 network was combined with the YOLOv2 algorithm. in the last portion. To increase the performance of the model, data augmentation is used [7].

In the research work "Human Activity recognition using CNN" we get to know that Human activity recognition is a trending research area in the field of artificial intelligence and computer vision due to its various applications in surveillance, digital entertainment and healthcare. In this research work an intelligent human activity recognition system is developed. Kinetics dataset is used to train using Convolutional Neural Network (CNN) and three-dimensional kernels. RESNET-34 is the 3D CNN model used for this research activity. The length of the video was about a tenth of a second. The trained model gave an accuracy of 79% on the kinetic dataset. Some noteworthy observations were that the accuracy was high for activities like running, standing etc. Whereas for other activities like cooking, doing yoga etc. it was considerably low. For further improvement of results, we can get a detailed dataset which labels different asanas to classify them properly [8].

The paper, 'Smart motion detection surveillance system' proposes a system that includes its own built GUI (Graphical User Interface) based on the needs and its motion detection technology. Motion detection is accomplished primarily through the use of both temporal difference and optical flow approaches in conjunction with a morphological filter. The number of pixels in a picture is minimized via temporal difference. The picture is then analyzed using optical flow detection in the next step. Because optical flow monitors both the speed and direction of moving objects, false alarms caused by fans, trees, and other factors are eliminated. The final step is to apply the morphological filter, which aids in noise reduction. Finally, a binary image is created, with the motion zone colored white and the areas with no motion detected colored black. The device captures images when the motions exceed a particular threshold. As a consequence, the amount of data that has to be analyzed is reduced, making environmental monitoring easier [9].

3. Literature Review

Related Studies	Advantages	Disadvantages	Methodology
[1]	• The system detects the motion if there is any movement in front of the camera. Also, it displays the length of time the object was in front of the camera and a graph is plotted.	• It does not classify the detected object into any class.	• Implemented using python libraries such as OpenCV, Bokeh, Pandas and Datetime. • Captures the video frames. • Finding the difference in phase between the initial frame and the delta frames allows one to calculate movement. Timestamp is recorded using the status list. • Using the recorded timestamp, a graph is plotted.
[2]	• YOLO enhances the performance as it is fast and accurate.	• The huge variance caused by occlusions, lighting, and angles effects face detection hence accuracy, training time, and processing time for recognizing faces in real-time is still a challenge. • Small size of GPU could not accommodate increased batch size.	• Model is trained and tested using the FDDB dataset. • The input to the proposed network is a 448 by 448 colour picture. • The gradient decent optimizer technique was used to train the model.
[3]	• WIDER Face Dataset Results- The faces have a wide range of scale, posture, and occlusion modifications. The UFDD face detection dataset has a number of realistic problems that aren't found in any other dataset. On this dataset, the best mAP, 96.11%, is obtained with the DPSSD face detector.	• IJB-face C's verification accuracy is only slightly worse than ArcFace at FARs between 1e-6 and 1e-1.	• An all-encompassing face algorithm that calculates the facial key-points for every face that DPSSD detects. These key points are used to align the faces to canonical coordinates using the similarity transform, which removes the effects of 2-D rotation and scaling. In order to identify or verify faces, we use DCNNs to extract the deep identity characteristics from the aligned faces. The user's IP address is tracked by the application for the purpose of identifying rogue users.
[4]	• By combining portions of this strategy, you can increase efficiency and accuracy while using less memory and money.	• The cost for developing these techniques maybe high.	• Detecting objects in real-time video can be done using a variety of techniques. One of them is the Frame Difference Method (FDM), which detects moving objects by contrasting the images of the current and preceding frames in a sequence of frames. [4]. The Background Subtraction Method (BSM) leverages the variation from the original background picture to identify moving objects.

[5]	• YOLO, machine learning and CNN provide great accuracy for object tracking and detection	• These algorithms require a local GPU to run efficiently. • They can also be run on cloud-based architecture but preferably to be run on local machines.	• Detect objects and track them. • Can also detect objects in a huge crowd. • Classification of objects is also done to better identify them.
[6]	• Human action recognition can prove useful when identifying threats in a huge crowd.	• Provides a low accuracy score on certain datasets when tested.	• Uses various technologies to perform these actions. • Multiple layers lead to better accuracy.
[7]	• The suggested model's key benefit is the capability of quick real-time detection using high-resolution images.	• Its size, which makes training its settings more time consuming, is a drawback.	• FDDB dataset is used. • Ground truth dataset. • Features extraction. • Face detection. • Data Augmentation.
[8]	• It can be used to produce high accuracy for certain activities like running, standing etc.	• Whereas for other activities like cooking, brushing hair, doing yoga etc. it was low. It can be made better with proper training.	• RESNET-34 is the CNN model used to train the model for this research activity. • For the training for Convolutional Neural Networks(CNN) and three-Dimensional kernels we have used the kinetics dataset.
[9]	• It cuts down the amount of data that has to be analyzed and also saves data space by avoiding recording static photos that frequently do not include the object of interest.	• A GUI must be designed separately in accordance with the specifications.	• Motion detection is done with temporal difference, optical flow detection and morphological filter. • GUI is designed separately. • The resulting image will be a binary image with the motion zone highlighted in white and the regions with no motion detected highlighted in black.

4. CONCLUSION

From the following set of papers, we have understood various ways in which motion detection and human activity recognition-based security systems can be created, thereby contributing to the security of the society in general. These proposed frameworks, technologies and techniques detect any suspicious or unusual activity in real time in large crowds or even in sensitive areas like borders or any other high security areas. Since it is all machine based there is a very low margin of error.

Additionally, the system can also detect human activity to some extent, helping us to tackle the situation in a better way. If tested on a custom dataset this technique provides better accuracies when compared to training on certain existing datasets. Using multiple layers for human activity recognition makes it much better and more accurate producing promising results. Thus, it helps in analyzing the situation in real time with very little flaws and prevents any major catastrophe from happening.

REFERENCES

[1] Suraiya Parveen, Javeria Shah, "A Motion Detection System in Python and Opencv", IEEE Conference Paper, 2021

[2] Dweepna Garg, Parth Goel, Sharnil Pandya, Amit Ganatra, Ketan Kotecha, "A Deep Learning Approach for Face Detection using YOLO", IEEE Conference Paper, 2018

[3] Rajeev Ranjan, Ankan Bansal, Jingxiao Zheng, Hongyu Xu, Joshua Gleason, Boyu Lu, Anirudh Nanduri, Jun-Cheng Chen, Carlos D. Castillo, Rama Chellappa, "A Fast and Accurate System for Face Detection, Identification, and Verification" JOURNAL OF LATEX CLASS FILES, 2019.

[4] Shifeng Zhang, Cheng Chi, Zhen Lei, Senior Member, IEEE, and Stan Z. Li, "Refinement Neural Network for High Performance Face Detection", JOURNAL OF IEEE TRANSACTIONS,JULY 2019.

[5] Abhinu C G, Aswin P, Kiran Krishnan, Bonymol Baby, Dr. K S Angel Viji, "Multiple Object Tracking using Deep Learning with YOLO V5", in International Journal of Engineering Research & Technology (IJERT), 2021.

[6] Inzamam Mashood Nasir, Mudassar Raza, Jamal Hussain Shah, Muhammad Attique Khan, Amjad Rehman "Human Action Recognition using Machine Learning in Uncontrolled Environment", 1st International Conference on Artificial Intelligence and Data Analytics (CAIDA),2021.

[7] Htet Aung, Alexander V. Bobkov, Nyan Lin Tun , "Face Detection in Real Time Live Video Using Yolo Algorithm Based on Vgg16 Convolutional Neural Network", IEEE Conference Paper, 2021.

[8] Ms. Shikha, Rohan Kumar, Shivam Aggarwal, Shrey Jain, "Human Activity Recognition", International Journal of Innovative Technology and Exploring Engineering (IJITEE),2020.

[9] Li Fang, Zhang Meng, "Smart Motion Detection Surveillance System", International Conference on Education Technology and Computer 2009.

[10] Worawut Yimyam, Thidarat Pinthong, Narumol Chumuang, Mahasak Ketcham, "Face detection criminals through CCTV cameras" 14th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS) 2018.

[11] Jeff Donahue, Lisa Anne Hendricks, Marcus Rohrbach, Subhashini Venugopalan, Sergio Guadarrama, Kate Saenko, Trevor Darrell, "Long-term Recurrent Convolutional Networks for Visual Recognition and Description" 2016 IEEE.

[12] Chomtip Pornpanomchai, Fuangchat Stheitsthienchai, Sorawat Rattanachuen, "Object Detection and Counting System", 2008 Congress on Image and Signal Processing.

[13] Dr. Yusuf Perwej, Nikhat Akhtar, Shivam Chaturvedi, Shubham Mishra, "An Intelligent Motion Detection Using OpenCV", International Journal of Scientific Research in Science, Engineering and Technology 2022.

[14] Linus Granath, Andreas Strid, "Detecting the presence of people in a room using motion detection", Teknik och samhälle datavetenskap 2015.

[15] Zameer Gulzar, Anny Leema, "Human Activity Recognition using Machine Learning Classification Techniques", International Journal of Innovative Technology and Exploring Engineering (IJITEE) 2019.