LIVE VIDEO DETECTION USING DEEP LEARNING

K.RAMBABU, PEDASINGU NAGESWARA RAO

Assistant Professor(HOD) MCA, DEPT, Dantuluri Narayana Raju College, Bhimavaram, Andhra Pradesh

Email id : kattarambabudnr@gmail.com

PG Student of MCA, DNR COLLEGE P.G COURSES(AUTONOMOUS), Bhimavaram-534202

Email id : pedasingunageswararao75@gmail.com

ABSTRACT

Real time object detection is a vast, vibrant and complex area of computer vision. Ifthere is a single object to be detected in an image, it is known as Image Localization and ifthere are multiple objects in an image, then it is Object Detection. This detects the semanticobjects of a class in digital images and videos.

The applications of real time object detectioninclude tracking objects, video surveillance, pedestrian detection, people counting, self-driving cars, face detection, ball tracking in sports and many more. Convolution NeuralNetworks is a representative tool of Deep learning to detect objects using OpenCV(Opensource Computer Vision), which is a library of programming functions mainly aimed at realtime computer vision.

1 INTRODUCTION

The motive of object detection is to recognize and locate all known objects in a scene. Preferably in 3D space, recovering pose of objects in 3D is very important for robotic control systems.

Imparting intelligence to machines and making robots more and more autonomous and independent has been a sustaining technological dream for the mankind. It is our dream to let the robots take on tedious, boring, or dangerous work so that we can commit our time to more creative tasks. Unfortunately, the intelligent part seems to be still lagging behind. In real life, to achieve this goal, besides hardware development, we need the software that can enable robot the intelligence to do the work and act independently. One of the crucial components regarding this is vision, apart from other types of intelligences such as learning and cognitive thinking. A robot cannot be too intelligent if it cannot see and adapt to a dynamic environment.

The searching or recognition process in real time scenario is very difficult. So far, no effective solution has been found for this problem. Despite a lot of research in this area, the methods developed so far are not efficient, require long training time, are not suitable for real time application, and are not scalable to large number of classes. Object detection is relatively simpler if the machine is looking for detecting one particular object. However, recognizing all the objects inherently requires the skill to differentiate one object from the other, though they may be of same type. Such problem is very difficult for machines, if they do not know about the various possibilities of objects.

1.2 Motivation:

Blind people do lead a normal life with their own style of doing things. But, they definitely face troubles due to inaccessible infrastructure and social challenges. The biggest challenge for a blind person, especially the one with the complete loss of vision, is to navigate around places. Obviously, blind people roam easily around their house without any help because they know the position of everything in the house. Blind people have a tough time finding objects around them. . So we decided to make a REAL TIME OBJECT DETECTION System. We are interested in this project after we went through few papers in this area. As a result we are highly motivated to develop a system that recognizes objects in the real time environment

2 LITERATURE SURVEY

While I can't provide an exhaustive survey, I can offer some key themes, approaches, and notable papers up to my last

knowledge update in September 2021.

Object Detection and Tracking:
The use of deep learning models like Faster R-CNN, YOLO (You Only Look Once), and SSD (Single Shot MultiBox Detector) for real-time object detection and A literature survey on live video detection using deep learning is a comprehensive task that involves reviewing numerous research papers and publications in the field. tracking in live video streams.
Action Recognition:
Research on recognizing human actions in live video, often using 3D CNNs (Convolutional Neural Networks) and RNNs (Recurrent Neural Networks).
Surveillance and Security:
Applications in security, including the detection of intruders, suspicious behavior, and anomalies in live video feeds.
Traffic and Transportation:
The use of deep learning to monitor traffic, detect accidents, and optimize traffic flow in live video from cameras placed on roads and intersections.
Medical Imaging:
Utilizing deep learning for real-time medical image analysis in live video, such as detecting anomalies in X-rays, MRIs, and endoscopic videos.
Human Pose Estimation:
Deep learning models for estimating human poses in real-time video, with applications in sports analysis, gaming, and fitness.
Face Recognition:
The development of facial recognition systems for identifying individuals in live video streams, with a focus on accuracy and privacy concerns.
Social Media and Content Moderation:
The use of deep learning to detect and moderate inappropriate content in live video on platforms like YouTube and Facebook.
Hardware Acceleration:
Research into optimizing deep learning models for live video detection on edge devices, like GPUs and TPUs, for real-time performance.
Datasets and Benchmarks:
Studies creating and benchmarking datasets for live video detection tasks, such as the COCO dataset and UCF101 for action recognition.
Notable Papers (up to September 2021):
"YOLOv4: Optimal Speed and Accuracy of Object Detection" by Alexey Bochkovskiy, et al.
"A Survey on Object Detection in Video" by Zhou, Pan, et al. (2018)
"Two-Stream Convolutional Networks for Action Recognition in Videos" by Karen Simonyan and Andrew Zisserman.
"Deep Residual Learning for Image Recognition" by Kaiming He, et al.
"Real-time Human Action Recognition from Video Streams" by Ahmed Fathi, et al.
"Focal Loss for Dense Object Detection" by Tsung-Yi Lin, et al.
Please note that the field of deep learning is rapidly evolving, and there may have been significant developments and new papers published after my last knowledge update in September 2021. To conduct a thorough literature survey, I recommend searching for recent publications and reviews in academic databases like Google Scholar, IEEE Xplore, and arXiv.

3 EXISTING SYSTEM

Computerized picture preparing is a range portrayed by the requirement for broad test work to build up the practicality of proposed answers for a given issue. A critical trademark hidden the plan of picture preparing frameworks is the huge level of

testing and experimentation that

Typically is required before touching base at a satisfactory arrangement. This trademark infors that the capacity to plan approaches and rapidly model hopeful arrangements by and large assumes a noteworthy part in diminishing the cost and time required to land at a suitableframework execution.

4 PROPOSED WORK AND ALGORITHM

Implementation of proposed strategy includes caffe- model based on Google Image Scenery; Caffe offers the model definitions, optimization settings, pre- trained weights.

5 METHODOLOGIES

MODULES

Data Collection: Gather a diverse dataset of live video footage relevant to your detection task. Label the data to indicate the objects or events you want the model to detect.

Preprocessing: Prepare the video data by resizing frames, normalizing pixel values, and possibly augmenting it to increase diversity and improve model generalization.

Model Selection: Choose a deep learning architecture suitable for video analysis. Convolutional Neural Networks (CNNs) are often used for this purpose. You may also consider 3D CNNs for spatiotemporal analysis.

Model Training: Split your dataset into training, validation, and test sets. Train your deep learning model on the training data, using appropriate loss functions and optimization techniques. Fine-tune hyperparameters to achieve good performance.

Real-Time Processing: Implement a mechanism to process video frames in real-time, typically using a library like OpenCV. This involves breaking the video into frames and feeding them through your trained model.

Detection and Post-processing: Apply object detection or event recognition on each frame. Depending on your task, you might use techniques like YOLO, SSD, or Faster R-CNN. Post-processing can involve filtering and smoothing to improve the stability of detections.

Performance Evaluation: Assess the model's performance using metrics like precision, recall, F1-score, and mean average precision (mAP) for object detection. Fine-tune your model and processing pipeline based on evaluation results.

Deployment: Deploy your live video detection system on the desired platform, which could be a local computer, edge device, or cloud infrastructure.

Monitoring and Maintenance: Continuously monitor the system's performance and update the model as needed to adapt to new scenarios or improve accuracy.

Ethical Considerations: Be aware of potential privacy and ethical concerns, especially when dealing with live video data. Implement measures to protect individuals' privacy and ensure ethical use.

Remember that the success of your live video detection system depends on the quality of your data, the choice of deep learning model, and the fine-tuning of various parameters. It's an iterative process that may require multiple iterations to achieve the desired performance.

5 RESULTS AND DISCUSSION

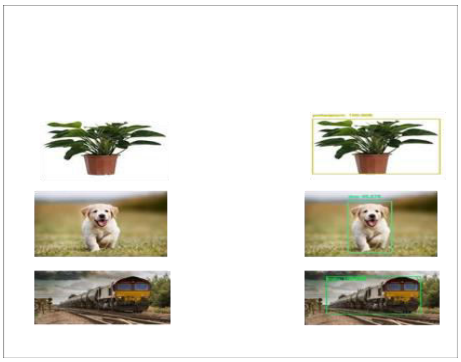

FIG1 : DETECTING LIVE OBJECTS

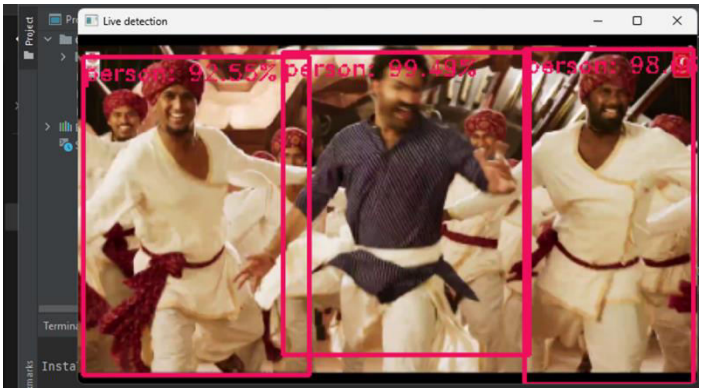

FIG 2: DETECTING EXACT INFO

6.CONCLUSION

CONCLUSION

Deep learning based object detection has been a research hotspot in recent years. This project starts on generic object detection pipelines which provide base architectures for other related tasks. With the help of this the three other common tasks, namely object detection, face detection and pedestrian detection, can be accomplished. Authors accomplished this by combing two things: Object detection with deep learning and OpenCV and Efficient, threaded video streams with OpenCV. The camera sensor noise and lightening condition can change the result as it can create problem in recognizing the object. The end result is a deep learning- based object detector that can process around 6-8 FPS.

7 REFERENCES

1. Bruckner, Daniel. Ml-o-scope: a diagnostic visualization system for deep machine learning pipelines. No. UCB/EECS-2014-99.CALIFORNIA UNIV BERKELEY DEPT OF ELECTRICAL ENGINEERING AND COMPUTER SCIENCES, 2014.
2. K Saleh, Imad, Mehdi Ammi, and Samuel Szoniecky, eds. Challenges of the Internet of Things: Technique, Use, Ethics. John Wiley & Sons, 2018.
3. Petrov, Yordan. Improving object detection by exploiting semantic relations between objects.MS thesis.UniversitatPolitècnica de Catalunya, 2017.
4. Nikouei, SeyedYahya, et al. "Intelligent Surveillance as an Edge Network Service: from Harr-Cascade, SVM to a Lightweight CNN." arXiv preprint arXiv:1805.00331 (2018).
5. Thakar, Kartikey, et al. "Implementation and analysis of template matching for image registration on DevKit- 8500D." Optik-International Journal for Light and Electron Optics 130 (2017): 935-944..
6. Bradski, Gary, and Adrian Kaehler.Learning OpenCV: Computer vision with the OpenCV library." O'Reilly Media, Inc.", 2008.
7. Howard, Andrew G., et al. "Mobilenets: Efficient convolutional neural networks for mobile vision applications." arXiv preprint arXiv:1704.04861 (2017).