IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101248

A Survey on Tool Tracking System

Darvish Davis[1], Anandhu Kumar[2], Alen Jhonson[3]

Universal Engineering College Vallivattom, Thrissur[1,2,3]

Abstract: Object detection refers to the capability of computers and other systems to locate objects present in an image and identify each of them. It has been widely used for face detection in security systems, for vehicle detection in driverless cars, and so on. Existing system performs numerous simpler and complex tasks like real world object detection from videos, concealed object detection…etc. This project focuses on detecting tools from a toolkit. Output achieved an accurate result up to the expectation.

Keywords: YOLO model, Comparer, Computer vision, Machine learning.

I. INTRODUCTION

A tool is a device used to produce a new item from raw input. This single sentence explains how important a tool is. What will be the feeling of one person if he/she losses something useful. Some sort of discomfort will be felt .This is the same situation faced by each and every worker when they loss valuable tools. The currently running model performs concealed object detection using SI Net. Concealed means 'hidden' or 'to keep secret'. It is difficult to identify the objects present in such images with a simple look. This problem is solved using computer vision. It has disadvantages like increase in computation time and inefficiency in identifying all objects present in an image. Computation time is increased due to use of search identification network which is efficient. But it uses convolutional neural network in search phase to process the image and identification phase employs numerous similar functioned refinement blocks. This project deals with a solution for the above mentioned problem. It is done with the help of a machine learning model. In simple words we can say that the principle is machine learning along with yolo model. Almost all computations are performed by yolo model. It works by treating the entire task into two; regression-to identify object position and classification- to determine class of the object.

II. THEORY

A. YOLO model

YOLO is an algorithm that uses neural networks to provide real-time object detection. This algorithm is popular because of its speed and accuracy. It has been used in various applications. YOLO is an abbreviation for the term 'You Only Look Once'. This is an algorithm that detects and recognizes various objects in a picture (in real-time). Object detection in YOLO is done as a regression problem and provides the class probabilities of the detected images. YOLO v4 is an object detector which can be trained on a single GPU with a smaller mini-batch size. This makes it possible to train a superfast and accurate object detector with a single 1080 Ti or 2080 Ti GPU.

B. Comparer

A comparer is a device used to examine in order to note the similarities or differences between the input data provided into it. Comparer takes the input as any of the form as user needed and convert it to sufficient output form after processing it in the comparer. The processing inside a comparer includes the same process as every model does; it takes the input, convert it into machine identifiable code, performs comparison, generates result as 1 if first input is greater than the second one, -1 if the second image is greater than the first one, 0 if both are equal, and finally generates the user understandable output.

C. Computer vision

Computer vision is a field of artificial intelligence that enables computers and systems to derive meaningful information from digital images, videos and other visual inputs and take actions or make recommendations based on that information. If AI enables computers to think, computer vision enables them to see, observe and understand. Computer vision trains machines to perform these functions, but it has to do it in much less time with cameras, data and algorithms rather than optic nerves and a visual cortex of a human eye. Because a system trained to inspect products or watch a production asset can analyse thousands of products or processes a minute, noticing imperceptible defects or issues, it can quickly surpass human capabilities.

IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101248

D. Machine Learning

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed. It is one of the most exciting technologies that one would have ever come across. As it is evident from the name, it gives the computer that makes it more similar to humans; the ability to learn. Machine learning algorithms are used in a wide variety of applications, such as in medicine, email filtering, speech recognition, and computer vision, where it is difficult or unfeasible to develop conventional algorithms to perform the needed tasks.

III.RELATED WORKS

The related works are arranged in the order of technology evolved.

The method used in [13] is traffic congestion control technique to detect congestion and avoid congestion in urban area. It focus on 2 approaches such as traffic congestion detection and avoidance to regulate the traffic flow in urban area. It works based on concept of sliding window and flow density relationship theory. It is to reduce the time delay in a traffic congestion control. First approach is based on relative speed of vehicles and second one is based on sliding window. A code based traffic congestion avoidance mechanisms is used to clear the traffic. Traffic congestion detection (TCD) approach each vehicle, determine its relative speed with one hop neighbors and detect traffic congestion. In traffic congestion avoidance (TCA) approach once the congestion is detected the RSU finds the alternative path to regulate the traffic. The traffic congestion avoidance mechanism considers a four cross roads. Each path having two lanes one for onward and another for downward. When the density increases proportionally the time to avoid the congestion also increases. The framework is evaluated on mean time to detect congestion and mean time to avoid congestion under traffic flow and vehicle density for different scenarios. This approach is implemented only in specific scenarios but it is necessary to consider all possibilities of particular road traffic condition in future. The result shows that the framework provides complete solution for traffic related issues in urban area.

This paper [4] describe the implementation of a 0.435 THz imaging system with 1 mm-scale resolution, which can successfully detect concealed objects. The working frequency is obviously in the transmission band A, which means the THz frequency we choose is very promising in the future long distance remote sensing system. Pyro electric detector based on LiTaO3 crystal, is used for registration of modulated electromagnetic radiation in THz frequency range. Since the detector only responds to changes in the input THz power, a chopper is placed right after the THz source to modulate the input signal of the detector. Also, a filter is designed to reduce detector sensitivity to Mid-IR and visible light. The data acquisition unit block is intended for processing a signal from the pyro electric detector. It provides management of the amplitude modulator (Chopper), synchronous detection, digitization, creation of a data array from the detector, and communication (connection) with the computer through the USB port. A lens is used to focus the THz wave onto the object. And then, another lens will focus the THz wave carrying the information of the object into the pyro electric detector. The object is placed on a scanning table, controlled by a Controller connected to the computer. The scanning range and step interval of the object can be set up by the software in the computer. The positions of all these elements in the quasi-optic path are well designed in order to get a high resolution and aberration-free image of the object.

This paper [14] proposes 2 modifications to previous system; one for accuracy in which an extra convolutional layer is added to network and named it as Fast RCNN type 2. The other for speed in which input channel is reduced from three channel input to one and named as Fast RCNN type 3. Fast RCNN Type 2 all the layers of type 1 Fast RCNN but with an additional 3rd Convolution Layer followed by Re Lu layer which makes total of 13 layers. Fast RCNN Type 3 has a total of 11 layers, same as type 1 Fast RCNN but with a little modification of input channel size, instead of 3 channel input 1 channel has been used (input size = 32x32x1). The training options used are SGDM optimization algorithm with momentum of 0.9, 1.e-06 initial learn rate, 32 max epoch and mini batch size of 128. After generation of ROIs by using edge boxes algorithm, the detector is trained using positive instances. Only those ROI are considered by the network for training whose boundary box has and IOU of 0.7 with the ground truth while the remaining ROIs which are label Led as background during the process are left out

The proposed system in [10] will allow the users to navigate independently using real-time object detection and identification. This system consists of a wearable goggle with earphones based on IOT embedded platform using raspberry Pi 4. It has multiple segments includes image collection, image processing, object detection, expression recognition, real-time ID, text to speech conversion, and navigation system. The entire system is based on goggles, which include a camera and earphones. The whole system works depending on the image collected by the camera. Then the collected images are processed through Tensor Flow to identify different objects in the images. There are four modes of operation after getting the image data. In describe mode, the system uses local or cloud processing to explain the user about the image. In announcement mode, the system will narrate the object in the images in real-time. In search

IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101248

operation mode, object search and find direction are done by using voice recognition. In the proposed system, to identify the object's exact position and distance from the user, the system calculates the angle from the user's current location. In the last mode, real-time identification and expression recognition are made by the system. It has an algorithm to detect new faces and save it in the database for future reference. After all of these steps, the system generates text and process it for text to speech conversion. Then the speech output is sent to the user through a bone conduction earphone.

The method in [20] used faster CNN. It shares full images convolutional features with a detection network. Comparing to previous work, faster RCNN employs a region proposal network and does not require an external method for candidate region proposals. It try to unify the different components of object detection into one single neural network. The network uses features abstracted from the entire image so as to predict each anchor. Also it predicts all bounding boxes for an image simultaneously, which means the network reasons globally about the fall image and all of the object in the image as well. This model try to implement a CNN and then evaluate if on the Pascal VOC detection dataset. The initial convolutional layer of networks extract features from the selected image while the fully connected layers predict the output probabilities and coordinates. Faster RCNN uses a method which called region proposal network to generate candidate regions of interest (ROIs),based on input of selected image. This is opposite to fast RCNN that requires region proposals to be provided by an external source. The RPN is essentially build up by 3 convolution layers and a new layer called proposal layer. During training, the faster RCNN requires two additional new layers like anchor target layer and proposal target layer. Anchor target layer generates the target values. Proposal target layers generates the target class labels for the ROIs. The target regression coefficient per class for final detector which are used in the loss function of detection. To take advantages this four training strategies have been tried. Alternate string, approximately joint training, non approximately joint training, 4 step alternating training. During evaluation we only need proposal layer. Target layers are currently available in Python. Therefore training faster RCNN has currently to be done from the python API.

The purpose of our work [15] is to construct a model using deep convolutional neural network and target detection architecture (DCNN + Faster RCNN) to identify tomato fruit types (healthy fruit and six tomato disease fruit). In order to obtain the high recognition rate of tomato disease, the feature extraction network must accurately extract the characteristics of the disease image. As a model of deep learning, convolutional neural network has the ability of hierarchical learning and perform well in feature extraction. In this paper, three deep convolutional neural networks were selected, namely, VGG16, ResNet50 and ResNet101. The above three networks are especially outstanding in image feature extraction and have been widely used. They were combined with Faster RCNN structure respectively to identify tomato diseases. The overall process of DCNN + Faster R-CNN is divided into four steps. First is DCNN: In this model, deep convolutional neural network is used to extract the feature maps of input image, which is used for subsequent RPN layer and full connection layer. Second is RPN (Region Proposal Networks): RPN network is mainly used to generate region proposals. Third is ROI Pooling: In this layer, Proposals feature map of fixed size is obtained by using the Proposals generated by RPN and the final Feature map obtained by DCNN, which is input into the full connection layer at the back for target recognition and positioning. Fourth is FC and Soft max: The fixed size feature map was input into the full connection layer, and soft max layer was used to classify the input specifically. Meanwhile, the boundary box regression operation was completed to obtain the exact position of the tomato fruit.

This paper [16] proposes a leakage detection method to realize automatic detection for the first time. Faster RCNN is employed instead of humans to handle SF6 infrared thermography image. Faster RCNN is an advanced CNN which has been widely used in numerous practical situations including image identification and face recognition. It can not only detect leakage but also mark the location of leakage. In faster RCNN training, region proposal network is first used to generate a certain amount of candidate rectangular area. The output of RPN has 2 neural networks; one is used to give the candidate rectangular area and the other is used to judge whether there is leakage occurring. With sliding window method, RPN extracts numerous features with in these area using rectangles of different size and different position and conducting it simultaneously. Thus maps the characteristics of the target candidate area into low dimensional vector. This is fed as input into 2 networks. One is boundary regression network and the other is boundary classification network. Boundary regression network gives final position of 4 vectors in target rectangles to determine position of leakage. Boundary classification networks gives the target candidate area a score that is judged as leakage and whether score is higher than threshold value determines that whether there is leakage.

In this paper [17], based on a customized Faster-RCNN model we develop a machine-learning system that can automatically detect parasites in thick blood smear images on smartphones. This paper develops a rapid and robust system for automated malaria diagnosis on smartphones in thick blood smears using a customized Faster-RCNN model. To improve the detection performance of Faster-RCNN for small objects, first split images of $4032\times3024\times3$ pixels into regions of $252\times189\times3$ pixels, and then train a customized Faster-RCNN model with these regions and corresponding

IJARCCE

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101248

manual ground-truth annotations. The customized Faster-RCNN model includes four convolutional layers and two max-pooling layers. For testing, also split blood smear images into regions of 252×189×3 pixels, which are screened for parasites using our cascaded Faster-RCNN model. The detected parasite coordinates are then re-projected into the original image space for visualization and evaluation.

This paper [18] will be investigating the effect of using RGB, CIEXYZ, and CIELAB color space on night-time human detection. To solve the problem of small object size, only the KAIST reasonable subset is used where the annotations have a pixel height of 55 pixels or more. To have better bounding box annotations, the only class used will be the 'person' class since the other class have bad annotations. After filtering, preprocessing is done. The first Pre-processing step for images is to up-sample the image by 2 times. This is done since the size of the image in the KAIST multispectral pedestrian dataset are only 640 X 512 pixels, which is too small for to detect small objects. The second pre-processing step is to convert RGB color space to CIEXYZ or CIELAB. Open CV is used to up-sample the images and to change the color space. Then set Faster RCNN hyper parameters. The Faster RCNN base network used in this experiment is the Resnet101 architecture. The reason Resnet101 is chosen is that it has better performance compared to Resnet50 or VGG16. Resnet101 uses pre trained weights trained using RGB images from the Image Net dataset. To improve the performance, the Faster RCNN anchor generation will be modified to create better anchors. To modify the anchor generation, the feature stride, anchor scale, and anchor ration parameters will be changed. Anchor scales is used by Faster RCNN to generate anchor boxes with different sizes. Anchor ratios is used by Faster RCNN to generate anchor boxes with different height to width ratio. It is important for the performance of Faster RCNN since anchors with different height to width ratio fits some objects better. The evaluation metrics used is the mean average precision (m AP) metrics and log-average miss rate (LAMR). For training the Faster RCNN human detection network, this paper follows Li's experiment. All experiments will use learning rate of 0.001 for the first four epoch while learning rate of 0.0001 is used afterwards. The Faster RCNN human detection network will be trained for 6 epochs.

This research [7] has been carried out for providing a safe environment for drivers, visually impaired people. This paper illustrates experiments conducted on roadside traffic symbols to increase the efficiency and accuracy. Two algorithm faster RCNN and single shot multi box are used here. The candidate objects, regions are proposed by region proposal networks along with VGG-16 detector which further carry out image classification at the top. The bounding box proposals are predicted by the intermediate level layers. In next stage, from the intermediate feature map, use of box proposals are made to crop features. These maps are input to the remaining feature extractor for predicting class labels and for each proposals bounding box refinement. A single feed- forward CN detection framework that predicts classes along with anchor offsets directly without requirements of second stage classification proposal. This framework uses multi scale convolution bounding box. These outputs of bounding box are connected at top to feature maps that are multiple in number.

This paper [1] is an attempt to examine the use of object detection and instance segmentation for emergency vehicle detection. Faster RCNN and Mask RCNN are employed for these respectively. In object detection, the principle of transfer learning is employed i.e. these base networks are pre-trained on large pre-existing datasets to minimize the number of computations. They are also made fully convolutional to accommodate the inputs of random dimensions. Additionally, these base networks are supplemented by object detection networks like Faster RCNN, Single Shot Detectors (SSD) and Region based Fully Convolution Networks (R-FCN). The CNN architecture that has been chosen for implementation is Faster RCNN with a Res Net backbone. The instant segmentation techniques primarily employ the use of a flexible framework called Mask RCNN. It has two principal stages. The first stage anticipates the presence of the object in a region of the input image also known as the Region of Interest (ROI). The second stage forecasts the probability, displays Image over Union (IOU) bounding box and the binate mask around the image based on the results of the first stage. Both stages are fused in the backbone. The network has three components namely FPN, RPN and Backbone network architecture. Here the bottom-up architecture of FPN is implemented for feature extraction from the feed. The RPN is a light network that scans the FPN bottom-up and proposes probable regions in the image where the object is likely to be present. It then recognizes various regions by fitting multiple bounding boxes according to certain IoU values. The backbone is a multilayered neural network which obtains feature maps of the input feed. Here ResNet50 is employed as it is not a very deep architecture. Fine tuning helps the model attain higher accuracy with lesser training time. The Stochastic Gradient Descent algorithm with a mini batch is used to update the weights and momentum for minimizing loss and converging at the most accurate value.

The system in [3] will provide a novel method for detection and classification of types of teeth (viz., incisors, molar, premolar, cannine teeth) and also some underlying oral anomalies such as fixed partial denture and impacted teeth. A python tool for labelling the desired tooth in the images. This is done by drawing a rectangle around each tooth and give a label.Then generate TF records using the labeled images that will act as an input data to the Tensor flow training model. Then convolution is applied in which convolution is the first layer which extract high features from

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101248

given input image. It is also an element wise multiplication and generate an output from multiplication known as feature maps. The feature maps are subject to activation function known as Rectified Linear Unit to increase non-linearity into the input images, then the image(input) is passed on to max pooling layers to reduce the dimensionality of the image and finally given to the fully connected network which used soft max activation function to classify different types of teeth. In the Faster R-CNN architecture, a Region Proposal Network mainly consist of a CNN, a set of anchors and a Region Proposal Layer. At the last convolutional layer a sliding window of size n × n run spatially on the set of feature convolutional maps. A set of 9 anchors are generated for each window with the same center $x_{a,}$ y_a with 3 different aspect ratios and 3 different scales. Anchors is a different sized boxes which help to detect object of various sizes. To evaluate an object detector, ground-truth bounding boxes and predicted bounding boxes are employed. The extracted features (n × n spatial) from the convolutional feature maps will act as an input to other two networks of classification and regression. The regression output will give a predicted bounding box(x, y, w, h). The classification output is a probability P which indicates about the contents of the predicted box.

This paper [8] presents a solution to Traffic monitoring and analysis at Indian toll plazas. Proposed solution is an integration of models trained to help achieve objectives like vehicle detection on live video streams, Vehicle counting, Vehicle model type classification and vehicle registration number detection. This project started with detection and localization of different types of vehicles and form distinguishing boundaries for the located objects into video frames and cropping the detected objects to be stored locally and used for further processing in outlaid milestones. In object detection, we break down the original image into multiple images and perform object localization on each part. To solve the problem of classification and regression, system have different models which uses different approaches such as R-CNN to extract ROI’s. Fast-RCNN extracts ROIs from Feature Map. Faster R-CNN uses Region Proposal Network (RPN) to extract ROIs and in SSD. For object detection we used tensorflow object detection API’s one of the powerful tools to help us achieve the objective detection milestone and use algorithm for implementation. Input image applied with a layers of convolution layers (with RELU activation function). Max pooling layer to concise the derived feature maps by extracting the max values of the feature map values to concentrate on the main features of detected objects. The output of pooling layer flattened feature map; then passed through fully connected neural network layers. Soft Max activation enables to do multi-class classification. For the detected vehicles in multiple video frames using Faster RCNN implementation it became a challenge to monitor the moving vehicles and hence we applied a naming logic to address the detected objects and monitor them in subsequent frames. A combination of location directory, a counter to track number of cropped images, class index id, class index name and object id were used as components to form a name for the detected vehicles. A unique object id talked about is derived via Centroid tracking algorithm are given in the derived name space which is then saved in CSV format and finally merging all for detailed analysis. Such as counting number of vehicles. Further to the processing in the solution the detected vehicles with bounding boxes are cropped and used for vehicle make classification and vehicle registration number detection. We used VGG16 based image classifier for identifying the make of detected vehicles in first stage of our system. The same cropped image of object detection stage of our system was used as an input for vehicle registration number detection. A background subtraction achieved with Otsu thresholding. Background subtracted images were then subjected to segmentation where each character was further predicted with an SVM based character recognition algorithm. Finally, the output of all these models is clubbed to form a CSV file with insights like vehicle counts, vehicle registration number, vehicle make classification description along with timestamp showing the date and time. This insight then redirected over dashboard for live monitoring and further analysis.

The method [19] used SSD (single shot detector) architecture and mobile net model to perform the detection and classification. Localization is the method to determine coordinates of the detected instance and their corresponding boundaries. In this paper, it deduce a model consists of object detection and classification on a platform where the processing power is not powerful. We focus on training the MS COCO dataset which is considered to be one of the finest datasets for performing object detection. The model used tensor flow deep learning API as its backend. Using the power of CNN, the detection accuracy is increased in parallel with the classification accuracy and not by reducing the speed. A video recorded from a drone will be used as input for obtaining the results. This method implements an object based detection system which is able to achieve high accuracy and run at 5-7 frames per second. COCO datasets used in this paper which contains a large number of various kinds of objects. Mobile net is being used as it more suitable for getting implemented where they computational power is less. The images were resized to size 224×224 to lower the cost of the computation. Various filter are applied on the resized images to remove the noise from images. Then, the images are used training and testing stage. The model that is used SSD architecture which has a small 3×3 convolution for feature extraction.

The method used in [21] DSP board embest devkit 8500D. The camera Microsoft life cam HD300 is used in the system. Angstrom operating system installed in it and it is programmed in c ++ languages integrated with open CV library. Frame differencing technique is used in which two consecutive frames are subtracted from one another to remove the stationary background. Background subtraction is a process of identifying any changes in a given frame by subtracting incoming frames from a fixed reference frames. The DSP board making use of its floating and MAC (multiplication accumulation units) operation to increases the system efficiency. This method (DSP) which is based on TI DM3730/AM3715 processor integrating ARM cortex A8 kernel at 1GHz and DSP come at 800MHz. DSP board is a standalone system with Linux angstrom embedded operating system. It comes with inbuilt GCC compiler which can compile C, C++ programs and also has open CV library which facilitates the image processing applications. The DM3730 processor used in the DSP. The programmable DSP helps in performing various signal processing tasks like image processing and analysis, digital filtering and math functions. The DM3730 processor can also be used for high definition video application which involve processing large amount of data. Six general purpose input/ output (GPIO) banks in it. Each GPIO banks comprises dedicated general purpose pins. Display subsystem provides the logic to display a video frame from the memory frame buffer (either SDRAM or SRAM) on a liquid crystal displays (LCD) panel attached to the kit. The software system used in the kit. It consists four parts like xloader, vboost, kernel, rootfs. The operating system supports many packages like open CV library python and java script, web applications and has many inbuilt compilers. It is mainly used in this system as it supports the open CV libraries and also the USB camera can be easily interfaced with it.

This paper [12] proposes an implementation for detecting the identity and location of objects in real-time video preview and overlaying 3D graphics on them in os Apps. The methodology for integrating YOLO model with Augmented Reality consisted of the following steps: Training a tiny YOLO ML model, passing frames to the pre-trained model to request results, handling results to extract predictions, running HitTest to obtain the 3D position of an object, and overlaying 3D graphics at the position. To train the tiny YOLO ML model, the Turicreate engine and the INRIA Annotations for Graz-02(IG02) dataset were used. A stop flag was set up to tell the application when to stop attempting to detect objects. The flag was initially set to false. It could be optionally set to false if only a specific number of instances of an object were to be detected. A view for displaying AR experiences that augment the camera view with 3D Scene Kit content was added to a function that occupied the entire screen. The function would recursively call the other function in its body and then call itself if the stop flag was false. The function would capture a frame by using the screenshot API. The image was converted from the image returned by function to another image. In the classification request, Threshold (Non Maximum Suppression Threshold) is set to 0.5. The tiny YOLO model is prone to predict multiple similar predictions associated with a single object instance in the post pressing of the results, any prediction with Intersection over Union with the highest confidence predictions less than the threshold was eliminated to avoid multiple prediction associated with a single object instance. The results were handled by a successive components.

The main aim of this system [6] is to create the more accurate object detection model for Augmented Reality using communication between the deep learning processing and the micro soft hololens as input/output device. Here paper describes the main process of our system to ensure the object detection using YOLO algorithms via holo lens. The software development contribution is based on client-server interaction connection. For the hardware material, it used two Graphics processing Units (GPU) of NVIDIA type Quadro P4000 to perform the computational time of the server side usually the processing of YOLO algorithms. A micro soft holo lens version 1 was used for the client side to benefit of its camera as an input scene to object detection in real world. This camera mounted on the front of the device which enables applications to see what the user sees. Developers have access to and control of the camera just as they would for color cameras on smartphones, portables, or desktops. In this system, holo Toolkit the library of micro soft is imported to control gaze, gesture and cameras. The server side, included the arket library of Redmon to run the YOLO algorithms which is implemented in C and CUDA programming language.

This research [5] aims to propose a more comprehensive solution. Object detection using YOLO gets input from the camera module, the input from the module will be processed using weights obtained from the training program so as to produce a system that can detect the presence of cigarette objects, the results of the object detection program using YOLO in the form of integer value 1 or 0 which will be done later OR logic gate processes with results obtained from smoke detection programs using smoke sensors. In the process of detecting smoke by a smoke sensor, using the input of smoke circulating in the air in public facilities. The level of CO gas pollution as the dominant gas contained in cigarette smoke is 50 Ppm, if a CO gas of greater than 50 Ppm is detected it will output an integer value of 1 and if not it will issue an integer of 0 which will then be carried out the OR gate logic process with object detection results using YOLO. The output of the system as a whole is the public service announcements about the danger of smoking and also an alert using buzzers triggered by the output bit of the OR operation between object detection using YOLO and smoke detection using sensor readings. In a certain condition, where the output of each detection module is zero (0), so that the output bit of the OR operation is also zero (0), the system will display a random PSA with different topics, depends on the campaign list that is deployed as the default list of the display. That way, that the system is delivering the educative PSAs using the pulsing model as previously described.

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101248

This paper [9] presents a UWB imaging system, including a modified rotating UWB antenna array, RF circuits and the 2D implementation of the delay-and-sum imaging algorithm. In the previous UWB imaging system, there are four receiving antenna elements placed side-by-side as a straight receiving arm at a distance away from the single transmitting antenna to form the antenna array. It is known that if increasing the receiving antennas in the array, the radiated beam-width of the antenna array will be improved, which will improve the cross-resolution of the imaging. In this case, based on the present UWB imaging system, two more antenna elements are added into the rotating antenna array, as well as decreasing the distance between the transmitting antenna and the receiving antenna. Although more antenna elements are added into the array and the receiving antennas are close to each other, the mutual coupling between the antennas is below -15dB during the operating frequency band. Meanwhile the six RX elements array will have narrower radiation beam compared with the four RX elements. The 2D reconstructed image is the combination of each rotation plane of the antenna array. In each plane, the method is based on delay-and-sum algorithm. Here X axis is assumed as the direction along the width of the target while Y axis is along the height of the target and Z axis is along the down range distance. In XY plane, when receiving antenna array A^0, B^0, C^0, D^0, E^0 and F^0 are at the angle of 0^0, the reconstructed image can be achieved. When rotating the antenna array to different angles, the coordinate X-axis and Y-axis will be considered to rotate the same angle with the receiving antenna array so that the reconstructed image will be rotated in terms of the varied coordinate, which is similar to the area. After rotating a circle, the area highlighted will be merged into the final reconstructed result.

This paper [11] proposes a new imaging method by de-aliasing. Here the proposed system perform numerical simulations to evaluate the performance of the imaging method proposed. In the simulation, we choose a planar transceiver antenna array with the height of 2m and width of 1m. The MMW frequency band is set to that from 27.52GHz to 30GHz, where 32 frequency points with the spacing of 80MHz are taken in the simulations. For comparison, we consider two settings on the spacing of the array elements, one is 2mm for the non-aliasing case and the other is 5mm for the aliasing case. Numerical experiments are performed for imaging of two kinds of object targets, one is a plate with sticks, and the other is a pattern with 3-D scattering points. In experiment 1, the plate is taken where seven groups of sticks are placed vertically and horizontally and the plate is parallel to the plane of the array with the distance of 50cm. Each group contains four equal-width sticks, and both the stick width and the spacing between the sticks range from 1mm to 7mm. For simplicity, here four frequency points, i.e., 29.76, 29.84, 29.92, and 30GHz, are employed. We can clearly distinguish the sticks in the groups with 3mm and 4mm width respectively and in both of the vertical and horizontal directions. 4mm width sticks cannot be distinguished, while the imaging results by our proposed de-aliasing method, is almost the same as traditional approach both in resolution and clarity. In experiment 2, the pattern with 3-D scatter points are placed on planes with $z' = 30, 31, 32, 33, 34$ cm. It is seen that by using the traditional method the imaging scattering points within the circle are blurred, while the imaging with our de-aliasing method presents clear scattering point images, and the details where in are distinguishable.

IV. CONCLUSION

This system provides an appropriate solution for loss of tools. The idea about this problem was generated by viewing the adverse effect of workers while losing the tools. Loss of a tool is a painful case because a tool cannot be substituted for other type as the task to be performed by each tool is different. popular libraries like open CV, tensor flow were employed for this purpose. Model was able to identify missing items in a better way. The system performed up to the expectation by accurately identifying the missing tools in a toolkit.

REFERENCES

[1]. Kaushik S, Abhishek Raman, Dr. Rajeswara Rao K.V.S,"Leveraging Computer Vision For Emergency Vehicle Detection-Implementation And Analysis",11th ICCCNT IEEE, 2020 .

[2]. Deng-Ping Fan, Ge-Peng Ji, Ming-Ming Cheng, and Ling Shao, "Concealed Object Detection" , IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE ,july 04,2021.

[3]. Anuradha Laishram, Khelchandra Thongam," Detection and Classification of Dental Pathologies using Faster-RCNN in Orthopantomogram Radiography Image", 2020 7th International Conference on Signal Processing and Integrated Networks.

[4]. Feifei Xin, Hongyan Su, Yong Xiao, "Terahertz Imaging System for Remote Sensing and Security Applications", 2014 3rd Asia-Pacific Conference on Antennas and Propagation.

[5]. Lathifah Arief, Alif Ziden Tantowi, Nefy Puteri Novani, Tri A. Sundara," Implementation of YOLO and Smoke Sensor for Automating Public Service Announcement of Cigarette's Hazard in Public Facilities", 2020 International Conference on Information Technology Systems and Innovation (ICITSI).

[6]. Haythem Bahri, David Krcmarık, Jan Koci, "Accurate object detection system on HoloLens using YOLO algorithm", 2019 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO).

[7]. Samiksha Choyal, Ajay Kumar Singh," An Acoustic based Roadside Symbols Detection and Identification using Faster RCNN and SSD", 2020 International Conference on Emerging Trends in Communication, Control and Computing (ICONC3).

[8]. Narayana Darapaneni, Umang Maheshwari, Anwesh Reddy Paduri, Parikshit Bangade, Sushilkumar C Thorawade, Amit Mane, Rushikesh Borse, "Traffic Monitoring and Analysis At Toll Plaza", 2021 IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS).

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

Vol. 10, Issue 12, December 2021

DOI: 10.17148/IJARCCE.2021.101248

[9]. Min Zhou, Xiaodong Chen, Lei Li, Clive Parini, "The UWB Imaging System with Rotating Antenna Array for Concealed Metallic Object", The 8th European Conference on Antennas and Propagation (EuCAP 2014).
[10]. Amit Ghosh, Shamsul Arefeen Al Mahmud, Thajid Ibna Rouf Uday, Dewan Md. Farid, "Assistive Technology for Visually Impaired using Tensor Flow Object Detection in Raspberry Pi and Coral USB Accelerator", 2020 IEEE Region 10 Symposium (TENSYMP), 5-7 June 2020.
[11]. Weiwei Wang, Kehu Yang, "A METHOD FOR MILLIMETER-WAVE IMAGING OF CONCEALED OBJECTS VIA DE-ALIASING", IEEE 2020.
[12]. Sagar Mahurkar, "Integrating YOLO Object Detection with Augmented Reality for iOS Apps",IEEE 2018.
[13]. Gunasekaran Muthumanickam, Gopalakrishnan Balasubramanian, "A Traffic Congestion Control in Urban Areas with Vehicle-Infrastructure Communications", International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017).
[14]. Asad Ullah, Hongmei Xie2, Muhammad Omer Farooq, Zhaoyun Sun, "Pedestrian Detection in Infrared Images Using Fast RCNN", 2018 IEEE.
[15]. Qimei Wang, Feng Qi, "Tomato diseases recognition based on Faster RCNN", 2019 10th International Conference on Information Technology in Medicine and Education (ITME).
[16].kan Xu, Zheiwen Yuan, Jinli Zhan, Yiping Ji, Xing He, Haosen Yang, "SF6 Gas Infrared ThermalImaging Leakage Detection Based On Faster RCNN", 2019 International Conference on Smart Grid and Electrical Automation (ICSGEA).
[17]. Feng Yang, Hang Yu, Kamolrat Silamut, Richard J Maude, Stefan Jaeger, Sameer Antani," Parasite Detection in Thick Blood Smears Based on Customized Faster-RCNN on Smartphones",2019 IEEE.
[18]. Yap Hong Yeu, Mohd Ibrahim Shapiai, Zool Hilmi Ismail, Hilman Fauzi, "Investigation on Different Color Spaces on Faster RCNN for Night-Time Human Occupancy Modelling", 2019 IEEE 7th Conference on Systems, Process and Control (ICSPC 2019).
[19]. Harit Ahuja, Vedant Kuhar, R. I. Minu," Object Detection and Classification for Autonomous Drones", International Journal of Recent Technology and Engineering (IJRTE).
[20]. Wang Cheng, Peng Zhihao, "Design and Implementation of an Object Detection System Using Faster R-CNN", 2019 International Conference on Robots & Intelligent System (ICRIS).
[21]. Suraj K Mankani, Naman S Kumar, Prasad R Dongrekar, Shreekant Sajjanar, Mohana, H V Ravish Aradhya," Real-Time Implementation of Object Detection and Tracking on DSP for Video Surveillance Applications", IEEE International Conference On Recent Trends In Electronics Information Communication Technology, May 20-21, 2016.