Hardhat Detection Using IR and Depth Frames

Swaroop[1], Deepika Prabhakar[2]

Electronics and Communication
Dept.[1,2]R.V College of
Engineering, Bangalore, India

swarooppk96@gmail.com[1], deepikaprabhakar@rvce.edu.in[2]

Abstract: *Construction industry is one of the largest sectors and 20% of the total deaths happen here. Safety hardhat helmets are one of the precautionary methods recommended and followed in a construction site. Detecting and classifying people wearing and not wearing helmets is an important task. Safety and Security is one of the most researched fields in recent years and acquiring 3D geometric information from any real environment is an important task for such applications. Well known methods like stereo vision camera system suffers from high time consumption or from the inability to match corresponding points in homogeneous regions. Time of Flight (ToF)technology,being a recent development, fulfills features desired for real-time distance acquisition along with compact size and higher frame rate. A safety application based on the ToF technology ad IR imaging is proposed in the paper. The ToF sensor provides depth information for each pixel as opposed to RGB values in case stereo camera-based systems. The depth sense cameras provide IR images along with the depth information. A YOLO framework is used to classify images. YOLO being faster than RCNN and the faster RCNN is much suitable for real-time classification.The model was trained on 50,000 images. Weights obtained during the 6000th epoch was chosen. A Mean Average Precision of 56% was obtained while testing.*

Keywords: Detection and Classification, Time of Flight, YOLO, Depth

1. INTRODUCTION

Safety is of highest priority in any field of work. Construction industry is one of the largest sectors that employs around 1.1 million people that amounts to only 9% of the world population. But almost 20% of the total deaths happen in the construction industry. Safety hardhat helmets are one of the precautionary methods to be compulsorily followed in a construction site. Detecting and classifying people wearing and not wearing helmets is an important task. Computer vision is one of the highly researched field for the development of safety applications and acquiring Three-dimensional (3D) geometric information from real environments is an essential task for such applications. Many models have been released in-order to distinguish people wearing safety helmets from people not wearing them. Well known methods like stereo vision camera systems suffer from high time consumption or from the inability to match corresponding points in homogeneous regions. Being a recent development in imaging hardware, the Time of Flight (ToF) technology fulfills features desired for real-time distance acquisition along with compact size and higher frame rate. The ToF sensor provides depth information for each pixel as opposed to RGB values in case stereo camera-based systems. Recent development in depth sense technology made the depth sensors to provide not only depth images, but also IR and RGB images. This paper focuses on using the depth and the IR space values from the sensor.

A You Only Look Once (YOLO) framework is used to detect and classify images. YOLO usesonly one convolutional network and predicts bounding boxes along with their respective probabilities. The network doesn't look at the complete image, but looks for regions with higher probability of having an object. YOLO being faster than Region based Convolutional Neural Networks(R-CNN) and faster R-CNN and is much suitable for real-time classification.

Further section of the paper is divided as follows: section 2 studies the previous research work

conducted in this field; section 3 briefs the YOLO framework and the Non-Maximum Suppression algorithm; section 4 briefs on the dataset used and section 5 details on the implementation; section 6 shows the results obtained.

2. RelatedWork

A lot of research has been done in image sensing and object detection. Initially RGB images from stereo vision cameras[4],[5]were used for the detection of objects. RGB images though are easy to obtain and use, they are sensitive to brightness, illumination and exposure. Unlike RGB, IR images are indifferenttowards exposure. Several articles use IR[1] versions of the scene instead of RGB space. Several other papers propose models based on depth information from a Time-of-Flight cameras. The ToF cameras calculate the distance of the object from the sensor using an IR radiation and calculating the phase difference between incident and the reflected light. The depth information though is fast, the suffer problems of low depth precision and low spatial resolution. Also, the measurement accuracy is limited by the emitted IR power [14]Another critical drawback of depth information is the motion blur. In order to combatthis, RGB, IR and depth information are used in pairs or all together. This provides better accuracy atthe cost of computation complexity and time. IR and depth information are used along with 3D Local Binary Pattern [9]to detect a face. Several other articles [10]also use IR and depth information to detect and classify different objects. Several other combinations like RGB and depth [12], [13], RGB and IR [11]are also used by various articles.
Object detection algorithm is an essential part of an object detection problem. Several algorithms were developed over time. Abu et al. [3]uses Histogram Oriented Gradients (HOG) features followed by Circle Hough Transform. HOG is also used to detect helmets in a traffic scenario to ensure safety of the riders [2].Recent algorithms mainly focus on Convolutional Neural Networks (CNNs) and its variants.

Detection problems may have outputs of different length. Therefore, accounting the variable output length into the algorithm is essential.Thus, region-based CNN was proposed[6]. Butthis method takes around 20 seconds to process a single image and this speed is not compatible with real-time applications. Improvements were made to RCNN to formulate Fast and Faster RCNN. Fast RCNN takes in a complete image and generates RoI using the convolutional feature map. Faster RCNN [7]uses a separate network to select RoI unlike selective search algorithm as used in Fast RCNN and is 10 times faster than the later. Though the Faster RCNN is faster than any previous methods, it couldn't be used in real-time situations. Joseph Redmon et al [8]proposed a novel algorithm called You Only Look Once (YOLO). The framework takes SxS grid as an input and provides bounding boxes along with the probability of it having an object.

3. You Only Look Once (YOLO)

Humans, at a single glance, detect and identify objects. Detecting objects at humanspeed helps computers or processors to perform complex functions in less time. Multipledetection algorithms like Regions with Convolutional Neural Networks (R-CNN), FastR-CNN, Faster R-CNN have been developed in recent ears. But these algorithms taketime to process a single image and can't be used in real-time scenario. The time takenby each of these algorithms is provided in the Table 1. This shows that the above-mentioned algorithms are no suitable for real-time scenario.

YOLO [8] is a detection and classification algorithm which was proposed by Joseph Redmondet al.in 2015. The model is a development made to compensate the drawbacks of state-of-the-art model of the time i.e.Faster R-CNN. Faster R-CNN though being extremelyaccurate model, it is time consuming and the acquirable frame rate is low for real-timeapplications. Thus, YOLO was developed. The basic YOLO model reaches upto 45FPSand is very suitable for real-time image sensing applications and the Fast YOLO

reachesan astounding 155FPS with acceptable accuracy. YOLO is based on regression and itpredicts classes and bounding boxes for the whole image in one run, instead of selectingthe interesting part of an image. Unlike its predecessors, YOLO usesonly one convolutional network to predict bounding boxes and the class probabilities of these boxes. Itdivides the image input to the network into SxS grids. Within each grid 'm' bounding boxes and their class probabilities are defined. The value of 'S' is dependent on the quality of detection required. The bounding boxes with probabilities above a pre-defined threshold are selected and to detect objects.Several advantages have been mentionedin the original paper[8] of the algorithm. YOLO is extremely fast when compared toother detection algorithms. It reasons globally when working on an image unlike slidingwindow technique and it learns generalizable features of an image.

Table 1: Comparison of YOLO and other detection algorithms

	Pascal 2007 mAP	**Speed**	
		FPS	**/img**
R-CNN	66.0	0.05	20s
Fast R-CNN	70.0	0.5	2s
Faster R-CNN	73.2	7	140ms
YOLO	63.4	45	22ms

The article [8]compares YOLO with its predecessors with respect to the accuracy and speed of the algorithm as in Table 1. Tough the precision of the model is comparatively lesser than other models, the basic model of YOLO reaches up to 45FPS speed.
During one pass of forward propagation, it determines the probability that cell contains a class. After probabilities are calculated thresholds are applied and unnecessaryanchor boxes are suppressed. The second step removes multiple bounding boxes on anobject and creates a single detection for an object.The YOLO algorithm learns aboutall the parameters required at a time in a single epoch. Pooling is applied at each stageto reduce the number of layers in the network without compromising on the clarity ofthe parameters or the accuracy of the model.Though YOLO makes more localizationerrors, it is less likely to predict a false detection.

3.1 Non-Maximum Suppression

Non-Maximum Suppression (NMS)is a technique used mainly in object detection that helps themodel to select the best bounding box from a set of overlapping bounding boxes. Theselection criteria for selecting the best bounding box can vary according to the scenarioand application. Intersection over Union (IoU) is the most frequently used criteria inYOLOfordeciding the best of the bounding boxes. IoU is calculated as ratio of Area of Intersection of bounding boxes to Area of Unionof the bounding boxes. The value of IoU measures how accurateis the object identified.

NMS uses IoU value to decide if boxes are overlapping or not. If the boxes overlap,then the box with maximum probability of having an object inside it, is selected as thebest among all the overlapping bounding boxes and the remaining bounding boxes areremoved from the detection output.

4. Dataset

Selection or creation of a dataset is one of the key steps in detection and classification application. The dataset has to have a lot of variations in it, so that the model can learn to be more robust against different scenes and scenarios. Several experimentations conducted proved that training the model with RGB images provides a similar effect as when trained with IR image data. IR image can be considered to be a 16-bit grayscale image. Assuming the experiment results obtained to be globally true around 50,000 images of hardhat were collected from different open source data resources. The datasets with bounding box representation

formats other than theYOLO format is converted to theYOLO format to make the dataset more systematic.

5. Image Detection and Classification

In this step, the dataset is passed through the YOLO network. A YOLO version 2 is used as the model. The initial weights were considered from the pretrained model provided by the creators ofYOLO. The training is run through 15000 epochs and a graph between loss and epochs was plotted and weights at 6000 epochs was selected for purpose of testing. Initial stages of testing showed several bounding boxes for a single object. This error was rectified using NMS algorithm as a part of post testing procedures. Depth information is used as a trigger to start the algorithm and classify the images. Thus, power used can be reduced to a minimum.

6. Results

Based on the Mean Average Precision (mAP) of each 1000 epochs and the averageloss value obtained during the process an optimum value of training epochs and weightswere chosen. Figure 1 shows the plot of loss function against the epochs or number ofiterations the model has trained for.Careful observation of the graph in the Figure 1 shows that the mAP of the modeldoes not increase much after 6000 epochs.Also, the loss value stays almost the same.Thus, epoch of 6000 was chosen to be optimum and further process was conducted usingthe weights obtained during the 6000thepoch of training. An average mAP of 56% wasobtained during the process.

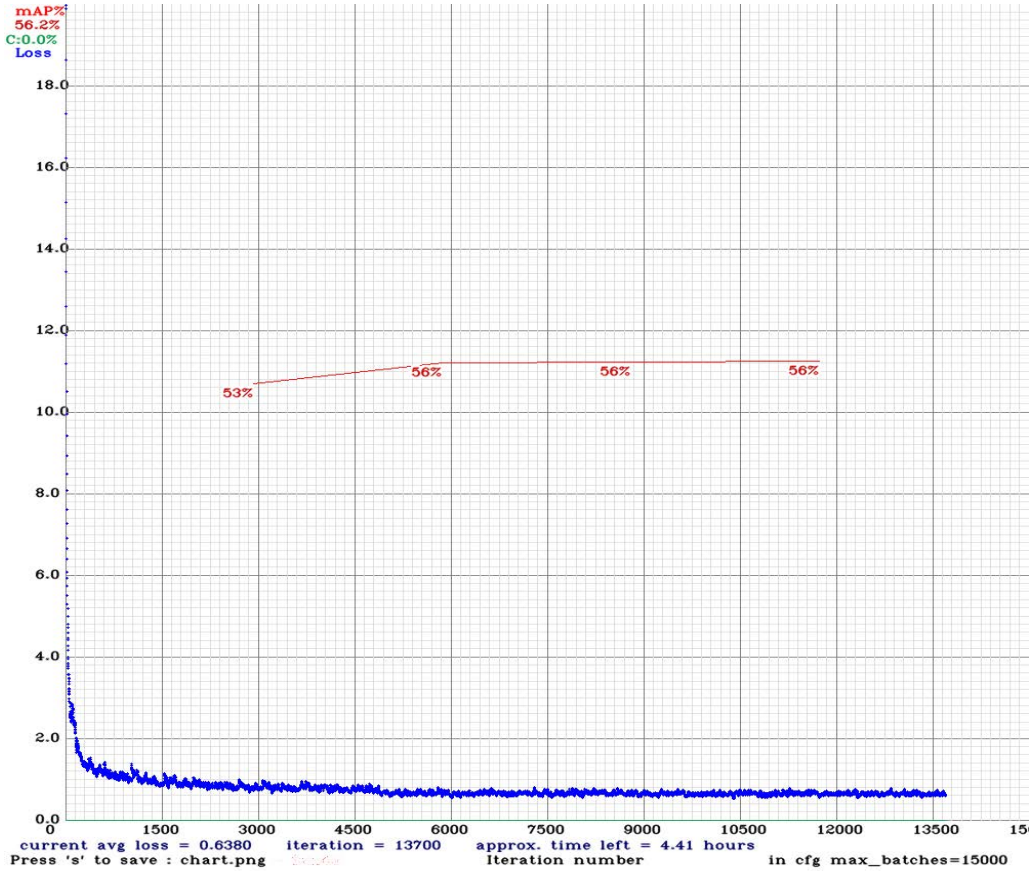
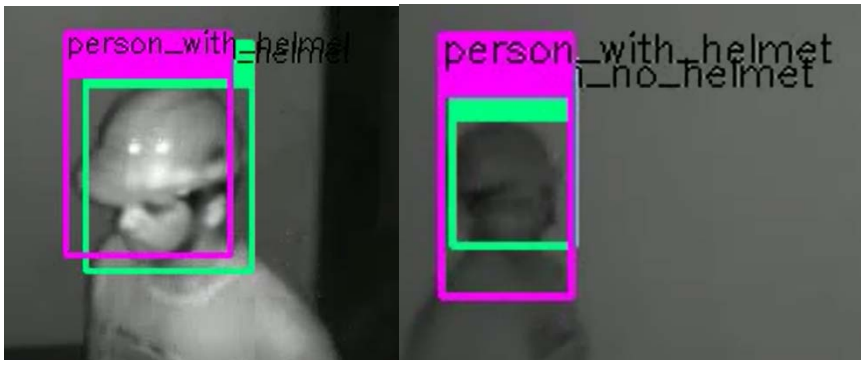


Figure 1. Loss Vs Epoch Plot



(a) **(b)**

Figure 2. Detection Output Before NMS

Testing revealed several problems in the model. Theoutput detection had multiple detections for a single object. Figure 2 shows the imageswhich had multiple detection for a single object. These detections though are accurate, butare unnecessary since there exists only one object in the scene. Thus, removing the extrabounding boxeswas performed using Non-Max Suppression technique. Thiswas performed as a post detection process and output image had all the extra boundingboxes removed. Performing the above said operation not only reduces the memory usedto save bounding boxes but also makes the model more robust and accurate.

(a) **(b)**

Figure 3. Detection Output After NMS

(a) **(b)**

Figure 4. Detection Output for Vehicle Helmets

Figure 3 shows the effect of fine tuning and result of applying Non-Max SuppressionTechnique and he final output of the detection process. The above-mentioned tuning could reduce the multiple detection issue to aconsiderably smaller, almost negligible, number.Detection process was performed on multiple IR image under different illumination.The results showed that the

result of detection is not dependent on illumination of the scene.

Along with hard helmets,vehicle helmets were also detected and is shown in Figure 4. Since the vehicle helmetsare not acceptable for construction sites, they are detected as a person without helmet.

7. Conclusion

Construction industry is one of the foremost and always busy industry in any country of the world.Though the sector creates wonders, there's a high rate of mortalityof workers in this sector. 20% of the total workplace death happen in the constructionsector. Several rules and regulations were created and implemented to prevent unnecessary deaths. One such regulation is wearing a construction helmet, also called the hardhelmet.The paper focuses on implementing a detection algorithm for the hardhats inconstruction area. ToF technology is used to accomplish the same.

The paper focuses on developing a classification and detection algorithm using IR, depth frames and YOLO network.Depth informationof the scene is collected through a ToF sensor. The ToF sensor has options to provideRGB and IR information along with the depth information. A code in C++ languagewas developed to read the depth information from the sensor. The detection algorithmhad YOLO as its base algorithm and was trained for 15,000 epochs and graph betweenloss function and the number of iterations was plotted to determine the weights that needto be used for the working of the model. A dataset containing around 50,000 images wasprepared for the training and validation purpose.The model was tested using the weights obtained on 6000thand a Mean AveragePrecision of 56% was obtained.

REFERENCES

10.1 Journal Articles

[1] Zhang, Hong, Xuzhong Yan, Heng Li, Rui Jin, and HongFeng Fu. "Real-time alarming, monitoring, and locating for non-hard-hat use in construction." Journal of construction engineering and management 145, no. 3 (2019): 04019006.

[2] R. R. V. e. Silva, K. R. T. Aires and R. d. M. S. Veras, "Helmet Detection on Motorcyclists Using Image Descriptors and Classifiers," 2014 27th SIBGRAPI Conference on Graphics, Patterns and Images, 2014, pp. 141-148, doi: 10.1109/SIBGRAPI.2014.28.

[3] A. H. M. Rubaiyat et al., "Automatic Detection of Helmet Uses for Construction Safety," 2016 IEEE/WIC/ACM International Conference on Web Intelligence Workshops (WIW), 2016, pp. 135-142, doi: 10.1109/WIW.2016.045.

[4]Juang, Chia-Feng, Guo-Cyuan Chen, Chung-Wei Liang, and Demei Lee. "Stereo-camera-based object detection using fuzzy color histograms and a fuzzy classifier with depth and shape estimations." Applied Soft Computing 46 (2016): 753-766.

[5]Du, Yi-Chun, Muslikhin Muslikhin, Tsung-Han Hsieh, and Ming-Shyan Wang. "Stereo vision-based object recognition and manipulation by regions with convolutional neural network." Electronics 9, no. 2 (2020): 210.

[6] Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 580-587. 2014.

[7] S. Ren, K. He, R. Girshick and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 39, no. 6, pp. 1137-1149, 1 June 2017, doi: 10.1109/TPAMI.2016.2577031.

[8]Redmon, Joseph, Santosh Divvala, Ross Girshick, and Ali Farhadi. "You only look once: Unified, real-time object detection." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 779-788. 2016.

[9] Kwon, Soon-kak. "Face recognition using depth and infrared pictures." Nonlinear Theory and Its Applications, IEICE 10, no. 1 (2019): 2-15.

[10] Fong, Katherine KaYan. "IR-Depth Face Detection and Lip Localization Using Kinect V2."International Journal of Electrical, Electronics and Data Communication, ISSN: 2320-2084 Volume-3, Issue-9, September.-2015.

[11] Guo-Hua, Chen, Wang Jun-Yi, and Zhang Ai-Jun. "Transparent object detection and location based on RGB-D camera." In Journal of Physics: Conference Series, vol. 1183, no. 1, p. 012011. IOP Publishing, 2019.

[12] Lin, Guichao, Yunchao Tang, Xiangjun Zou, Juntao Xiong, and Jinhui Li. "Guava detection and pose estimation using a low-cost RGB-D sensor in the field." Sensors 19, no. 2 (2019): 428.

[13]U. Asif, M. Bennamoun and F. A. Sohel, "RGB-D Object Recognition and Grasp Detection Using Hierarchical Cascaded Forests," in IEEE Transactions on Robotics, vol. 33, no. 3, pp. 547-564, June 2017, doi:

10.1109/TRO.2016.2638453.

10.2 Books

[14]Miles Hansard, Seungkyu Lee, Ouk Choi, Radu Horaud. Time of Flight Cameras: Principles, Methods, andApplications. Springer, pp.95, 2012, SpringerBriefs in Computer Science, ISBN 978-1-4471-4658- 2. ff10.1007/978-1-4471-4658-2ff. ffhal-00725654f.