Detection of sharp objects using deep neural network based object detection algorithm

R Kayalvizhi
Department of ECE,
SRM IST
Chennai, India
kayalvir1@srmist.edu.in

S Malarvizhi
Department of ECE
SRM IST
Chennai, India
malarvig@srmist.edu.in

Siddhartha Dhar Choudhury
Department of CSE
SRM IST
Chennai, India
sc6501@srmist.edu.in

Anita Topkar
Electronics Division,
BARC
Mumbai, India
anita@barc.gov.in

P Vijayakumar
Department of ECE,
SRM IST
Chennai, India
vijayakp@srmist.edu.in

2020 4th International Conference on Computer, Communication and Signal Processing (ICCCSP) | 978-1-7281-6509-7/20/$31.00 ©2020 IEEE | DOI: 10.1109/ICCCSP49186.2020.9315272

***Abstract*—Deep learning algorithms have the ability to learn complex functions and provide state-of-the-art results for com- puter vision problems. In recent times, these algorithms far exceeded the existing computer vision based techniques for object detection in X-ray imaging systems. So far, in literature single class of object namely gun and its parts were considered for detection using the SIXray10 database. We propose deep learning-based solution for the detection of sharp objects namely knife, scissors, wrench, pliers in the SIXray 10database.We propose two models namely model A and model B using a common object detection algorithm-YOLOv3 (You Only Look Once) with InceptionV3 and ResNet-50. YOLO is a deep neural network based object detection algorithm that performs the task in one-shot which allows real time inference in video of 15-30 fps. The model is FCN (Fully Convolutional Network) as has the capacity to perform both regression and classification by sharing weights for both the tasks. The network predicts a rectangular box called bounding box around the predicted object of interest along with the associated class. We analyze the performance of both model in terms of mAP. We achieve mean accuracy of 59.95% for model-A and 63.35% for Model-B. The most daunting part of the project is the low ratio of harmful to non-harmful items. By performing rigorous experiments we came up with the best set of possible results which uses varied pre-trained neural networks for feature extraction in tandem with YOLO model for object detection. We endeavor to improve on these existing results so as these systems can be successfully deployed in airports to minimize human error and improve security in such environments.**

***Index Terms*—Baggage screening, Deep learning, Object de- tection, YOLOv3, X-ray testing, Deep learning, Object detection, YOLOV3, X-ray testing.**

I. INTRODUCTION

In recent years, desire to commute and travel has increased, the number of people using the airports to travel has also increased, so it has become necessary to conduct security checks on passenger bags. X-ray baggage screening is the technique used to protect public space from dangerous explosives, which is prevalent due to the growth of the population in large cities and crowd density in public transport hubs. The key challenge in object detection is cluttered X-ray images, hence there is a need to automate and recognize sharp objects in screening. This proves to be a difficult task for the human operators available at the time to filter and approve the items in the limited time available. In suitcases, the items are randomly stacked and may overlap each other, which makes it difficult for the operator to identify the objects within limited time available as they may be sharp in nature or may be rotated, which the naked human eyes might not be able to identify. It is known that human eye has excellent vision capabilities, but in cases where objects are stacked one over the other, it becomes difficult for us to identify each object with high accuracy.Deep neural network based detection systems have not yet been deployed in airports due to their unacceptably low accuracy scores and the absence of uniformly distributed dataset of harmful and non-harmful items.. Our aim is to employ deep learning to reduce inspection time, the human force and effort in the baggage screening process and also to improve the security level.

II. RELATEDWORK

The recent interest in deep learning [1] has attracted a large audience from various disciplines, leading to an exponential improvement in the techniques used for creating neural networks (which lies at the core of deep learning) architectures. In particular, Convolutional Neural Network has shown significant progress in image processing and computer vision problems. One of the classic computer vision problem, X-ray imaging [2][3][4] has gained further study on CNN applied over baggage images. Some of the previous works done inthis area are based on the Bag of visual words [5][6][7][8][9]. A small chunk of the work done in X-ray imaging also used sparse representations and fuzzykNN[10][11].

Since the advent of applying the deep learning and Convolu tional Nets to various applications, there a need for analyzing color images from natural pictures. Recent work shows that the deep learning works well for analysis of X-ray security images. In [11], authors introduced Convolutional NeuralNet-works to X-ray baggage imaging problem. A similar machine learning-based Histogram Oriented Gradients (HOG) along with Support Vector Machine

978-1-7281-6509-7/20/$31.00 ©2020 IEEE

(SVM) classifier approach has been used in [12],[22]. Comparison of Convolutional Neural Network (CNN) with Bag of Visual Words (BoVW) has been made in [12]. The authors of [8], conducted a wide array of experiments in evaluating the quality of features learned from the Convolutional Layers and used handcrafted features such as SURF, KAZE, SIFT, and FAST. In [13] authors proposed many models for object detection and achieved promising results.

Recurrent-Convolutional Neural Network R-CNN, Regional based fast CNN namely, R-FCN, and Faster-RCNN networks were employed by authors [14] for object detections on X-ray images. In [15], Faster-RCNN with a ResNet as backend was employed to detect objects in PASCAL VOC 2012 dataset. In [16] work was done to evaluate various deep learning-based object detection neural network architectures such as R-CNN, FR-CNN, Faster R-CNN, YOLO, Single Shot Detector SSD, and they concluded that for their application case SSD achieved the best results out of the chosen set of networks. In [20] authors, integrates ResNet-50 for feature extraction of the YOLOv3 framework, and the object detection results demonstrate that it is efficient for detecting multi-objects from an image of complex natural scenes.

Recent work used YOLOv2 with Inceptionv3 and ResNet-50 backend for single object detection namely gun on the SIXray10 database. They achieved overall mean Average Precision (mAP) accuracy of 60.4% using InceptionV3 backend and 63.7% using a Resnet-50 backend. In this work, we propose YOLOv3 single-stage detector model forSIXray10 to detect four classes of objects namely knife, scissors, wrench, pliers.

III. PROPOSEDWORK

In this work, we employ the YOLOv3 (You Only Look Once version3) model [19] on dataset SIXray10 [21] for four class detection. A significant obstacle presented by researchers with this particular dataset is the imbalance of classes (harmful and non-harmful items) which has a ratio of 1:10. This ratio introduces immense overfitting, through various experimentations, we have tried to overcome the overfitting as much as possible, at the expense of scarifying the accuracy. For training a deep neural network based model You Only Look Once v3 (YOLOv3) has been used, the most commonly used architecture for feature extraction in this architecture is DarkNet which allows rapid training and faster inference. In our case we have used InceptionV3 and ResNet-50 instead of DarkNet which have the ability to extract features better than DarkNet.

Both the above mentioned feature extractors are deep convolutional networks trained on the famous ImageNet dataset which consists of 1.4 million images that correspond to 1,000 different classes. The networks (InceptionV3 and ResNet-50) have the ability to extract highly complex features from the input images, thus saving time learning these features ourselves and allowing us to rather focus on high- level task of object detection, this process is termed as "Transfer Learning". The model thus learns only the classification and regression heads which actually contribute to the object detector's output. Since, these neural nets involve handling large matrices and demands high-computational capabilities, we ran the experiments in a i7 desktop with a primary memory of 64 gigabytes and a powerful graphics processor - Nvidia RTX 2080Ti.

A. YOLOV3

You Only Look Once are an object detection algorithm that draws bounding boxes around a particular object of interest or multiple objects. The variant of YOLO that we used for our problem is YOLOv3. The original YOLOv3 used a DarkNet backend for feature extraction, which was trained on the ImageNet database. The DarkNet backend had 53 layers and the YOLOv3 itself had 53 layers of convolution networks, resulting 106 layers of network to be trained. Instead of using 106 layers of network, we replaced the DarkNet backend with an Inceptionv3 and ResNet-50 backends for extracting the basic features from the image. The performance of this network is as good as the DarkNet architecture for extracting features from the image. We froze the convolution layers of both these architectures with the ImageNet pre-trained weights so as to minimize the number of training parameters, and it proved to be a wise decision as the important features were extracted with less computations.

1) InceptionV3: The InceptionV3 network has been trained on images from ImageNet which is a large scale image database. It has 42 layers out of which the last layer contains softmax activation for N-way classification into 1000 different classes corresponding to the classes in this particular database. The final dense layers' weights were removed as they were not applicable in our use case, in a way this can be compared as deleting those layers from the resulting computational graph. The weights of the model till last layer of inception module was used which was found to be sufficient in extraction of relevant features from a collection of both complex and simple images which were encoded by these layers into feature maps of smaller dimensions and more depth than the original image. The features extracted from this backbone were passed to the object detector - YOLOv3 which was responsible for computation of bounding boxes around four different classes which were of interest to us - scissors, pliers, knife, and wrench to distinguish between them.

2) Resnet-50: As it is evident from the name of the architecture, ResNet-50 contains 50 layers which is created by combining fully connected dense layers and convolutional layers. A istinct feature of ResNets is the residual block - these blocks allow the neural network to feed low-level features to later layers so as to remind them of the importance of these features in computation of complex features much like building blocks which results in significant increase in performance. Like InceptionV3, ResNet-50 was also trained using the ImageNet database.

The final softmax activated dense layer with 1000 nodes denotes the different classes in ImageNet. This layer and the fully connected layers that come before it are not of interest to our use case and hence we throw away their weights. The output from this feature extractor backend is passed onto the YOLOv3 architecture giving for further classification and computation of bounding boxes around the aforementioned objects of interest.

978-1-7281-6509-7/20/$31.00 ©2020 IEEE

It is seen that in both the Inceptionv3 ,Resnet-50,models, we remove some layers of the pre-trained network ImageNET as it can confuse the model. This is because those layers contain features that are not specific to the six-ray 10 database for our objective of training the four classes.

With this frame work we propped two combined models as Model A: YOLOv3 with InceptionV3 backend and Model B: YOLOv3 with ResNet-50

Model A: YOLOv3 with InceptionV3 backend : Model A uses Inceptionv3 for extracting features out of the image. This architecture is pre-trained and available for use as open- source deployments in various deep learning frameworks. The InceptionV3 model was trained on the famous Imagenet set, that has images from 1000 different classes. The final layers of the neural network lead to classification of the image into these 1000 different categories, so we removed them as our task was slightly more complex than classification, for the remaining convolutional layers we retained the pre-trained weights till the final inception block. This was then used as a feature extractor which can extract important features from an input image which is passed to 53 layered YOLOv3 network. All images are passed through InceptionV3 to extract low dimensional complex features from the images, this is passed for object detection to the YOLOv3 model which performs regression (for bounding box computation) and classification (for predicting class of object). Figure 1: gives the architectural details of Model A.

Model B: YOLOv3 with ResNet-50 backend. : Model B uses YOLOv3 architecture ResNet-50 for feature extraction from an image. ResNet-50 has also been train on the same database as InceptionV3 - ImageNet with 1000 classes belonging to 1000 different items. The last fully connected dense layers in this network is removed too as they do not help in our case. The weights are frozen for the inception blocks made out of convolutional layers which saves computational cost of training even the feature extractors from scratch and lets us focus on the high level task of predicting bounding box and class of objects of interest.

IV. EVALUATIONMETHODS

Performance of the our proposed models for 4 classes using SIXray10 database are studied with mAP (mean Average Precision) and accuracy metrics. The Average Precision is calculated from recall value whose value should lie between 0 and 1. mAP value is calculated by averaging the Average Precision values. The important terms used here are listed below:

A. Intersection over union(IoU)

IoU is the measurement of common things presents between two sets. So we use that metric to find the difference between the predicted boundary metric and the ground truth that is the real object boundary. In some cases, we use data-sets that predefine an IoU threshold marking (assume 0.5) in telling whether the result is positive, negative, or neutral.

B. Precision and recall

Precision measures the accuracy of our predictions. i.e. number of correct predictions.

$$\text{Precision} = \frac{TP}{TP+FP}$$

TP = True Positive, FP = False Positive

Recall metric is used to measure the accuracy with which you find all the positives. For example, if the top K predictions have 80% of the positive predictions.

$$\text{Recall} = \frac{TP}{TP+FN}$$

TP = True Positive, FN = False Negative

V. RESULT ANDDISCUSSIONS

We got an overall mAP (mean Average Precision) of the model A is found to be 59.95% and individual class mAP were tabulated in Table1.

Table 1: mAP scores obtained using Model A

Four classes			
Knife	Wrench	Pliers	Scissors
59.43%	61.26%	62.79%	56.32%

The overall mAP of model B is calculated to be 61.35% and the individual mAP for each class were tabulated as Table2.

Table 2: mAP scores obtained using ModelB

Four classes			
Knife	Wrench	Pliers	Scissors
62.17%	63.33%	59.51%	60.38%

VI. CONCLUSION

We performed experiments using YOLOv3 on SIX-ray 10 database for object identification. So, far in literature only gun and items were taken as object for identification. We extend this over other four objects namely Knife, Wrench, Pliers and Scissors. Two models namely A and B were proposed and its performance were studied. Out of the two models, model B: YOLOv3 with ResNet-50 backend provides better results approximately 2% increase in mAP than model A: YOLOv3 with InceptionV3 backend.

So, from our experimentation we conclude that the results from ResNet-50 based feature extractor for YOLOv3 worked better than the one with InceptionV3 backend as it was able to capture the features of the X-ray scans much better than InceptionV3 and we can safely assume that ResNet-50 or other members of ResNet family of models can be used in problems related to X-ray scans of baggage. In future, we would like to improve the quality of our results by fine tuning and tweaking the hyper parameters of the neural network even further to generate better results.

REFERENCES

[1] LeCun, Yann, Yoshua Bengio, and Geoffrey Hinton. ”Deep learning. nature 521.” (2015):530-531.

[2] Mery, Domingo. ”X-ray testing by computer vision.” In Proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp. 360-367.2013.

978-1-7281-6509-7/20/$31.00 ©2020 IEEE

[3] Mery, Domingo, Erick Svec, Marco Arias, Vladimir Riffo, Jose M. Saavedra, and Sandipan Banerjee. ”Modern computer vision techniques for x-ray testing in baggage inspection.” IEEE Transactions on Systems, Man,andCybernetics:Systems47,no.4(2016):682-692.

[4] Mery, Domingo, and Carlos Arteta. ”Automatic defect recognition in x-ray testing using computer vision.” In 2017 IEEE Winter Conference on Applications of Computer Vision (WACV), pp. 1026-1035. IEEE, 2017.

[5] Bas¸tan, Muhammet, Mohammad Reza Yousefi, and Thomas M. Breuel. ”Visual words on baggage X-ray images.” In International Conference on Computer Analysis of Images and Patterns, pp. 360-368. Springer, Berlin, Heidelberg,2011.

[6] Turcsany, Diana, Andre Mouton, and Toby P. Breckon. ”Improving feature-based object recognition for X-ray baggage security screening using primed visualwords.” In 2013 IEEE International Conference on IndustrialTechnology(ICIT),pp.1140-1145.IEEE,2013.

[7] Bastan, Muhammet, Wonmin Byeon, and Thomas M. Breuel. ”Object recognition in multi-view dual energy x-ray images.” In BMVC, vol. 1, no. 2, p. 11.2013.

[8] Kundegorski,MikolajE.,SametAkçay,MichaelDevereux,Andre Mouton, and Toby P. Breckon. ”On using feature descriptors as visual words for object detection within x-ray baggage security screening.” (2016):12-6.

[9] Bas¸tan, Muhammet. ”Multi-view object detection in dual-energy X-ray images.” Machine Vision and Applications 26, no. 7-8 (2015): 1045-1060.

[10] Mery, Domingo, Erick Svec, and Marco Arias. ”Object recognition in baggage inspection using adaptive sparse representations of X-ray images.” In Image and Video Technology, pp. 709-720. Springer, Cham, 2015.

[11] Roomi, M., and R. Rajashankarii. ”Detection of concealed weapons in x-ray images using fuzzy k-nn.” International Journal of Computer Science, Engineering and Information Technology 2, no. 2 (2012): 187- 196.

[12] Akçay,Samet,MikolajE.Kundegorski,MichaelDevereux,andToby P. Breckon. ”Transfer learning using convolutional neural networks for object classification within x-ray baggage security imagery.” In 2016 IEEE International Conference on Image Processing (ICIP), pp. 1057-1061. IEEE, 2016.

[13] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. ”Imagenet classification with deep convolutional neural networks.” In Advances in neuralinformationprocessingsystems,pp.1097-1105.2012.

[14] Gaus, Yona Falinie A., Neelanjan Bhowmik, and Toby P. Breckon. ”On the use of deep learning for the detection of firearms in x-ray baggage security imagery.”(2019).

[15] Dai, Jifeng, Yi Li, Kaiming He, and Jian Sun. ”R-fcn: Object detection via region-based fully convolutional networks.” In Advances in neural informationprocessingsystems,pp.379-387.2016.

[16] Akcay,Samet,MikolajE.Kundegorski,ChrisG.Willcocks,andToby P. Breckon. ”Using deep convolutional neural network architectures for object classification and detection within x-ray baggage security imagery.” IEEE transactions on information forensics and security 13, no. 9 (2018):2203-2215.

[17] Jain, Deepak Kumar. ”An evaluation of deep learning based object

[18] detection strategies for threat object detection in baggage security imagery.” Pattern Recognition Letters 120 (2019):112-119.

[19] Wei, Yuanxi, and Xiaoping Liu. ”Dangerous goods detection based on transfer learning in X-ray images.” Neural Computing and Applications (2019):1-14.

[20] Zhang, Xiuling, Xiaopeng Dong, Qijun Wei, and Kaixuan Zhou. ”Real- time object detection algorithm based on improved YOLOv3.” Journal ofElectronicImaging28,no.5(2019):053022.

[21] Lu, Zhenyu, Jia Lu, Quanbo Ge, and Tianming Zhan. ”Multi-object Detection Method based on YOLO and ResNet Hybrid Networks.” In2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics(ICARM),pp.827-832.IEEE,2019.

[22] Akcay, Samet, and Toby Breckon. ”Towards Automatic Threat Detection: A Survey of Advances of Deep Learning within X-ray Security Imaging.” arXiv preprint arXiv:2001.01293(2020).

[23] Kundegorski,MikolajE.,SametAkc¸ay,MichaelDevereux,Andre Mouton, and Toby P. Breckon. ”On using feature descriptors as visual words for object detection within x-ray baggage security screening.” (2016):12-6.

978-1-7281-6509-7/20/$31.00 ©2020 IEEE

VII. APPENDIX

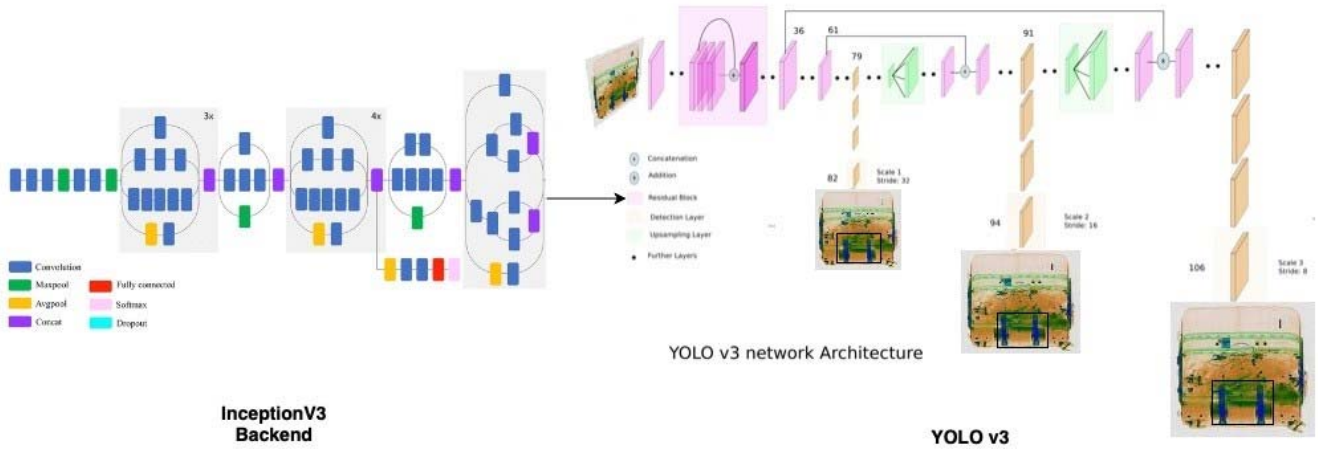

Fig. 1 YOLOv3 with Inception V3 backend

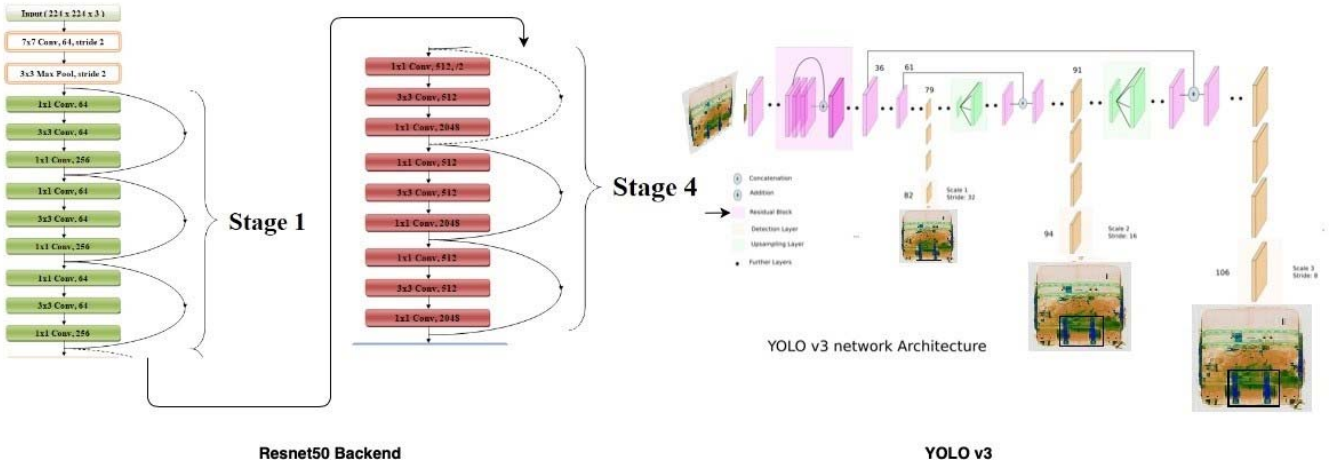

Fig. 2 YOLOv3 with ResNet-50 backend

978-1-7281-6509-7/20/$31.00 ©2020 IEEE