**2020** **­­Reference Papers:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Paper Title** | **Conference/Journal, Publisher** | **Authors** | **Methods** | **Remarks** |
| Research on Small Target Detection in Driving Scenarios Based on Improved Yolo Network. | [**Artificial Intelligence -Empowered Intelligent Transportation Systems**](https://ieeexplore.ieee.org/xpl/topics.jsp?isnumber=8948470&punumber=6287639&refinements=SpecialSection:Artificial%20Intelligence%20(AI)-Empowered%20Intelligent%20Transportation%20Systems) **(J) / IEEE** | Qiwei Xu, Runzi Lin, Han Yue, Hong Huang, Yun Yang, Zhigang Yao | YOLO v3, 2080 Ti machine, Dataset used is ApolloScape (Baidu’s autopilot dataset). | Improvised YOLO v3 and it showed better results compared to YOLO v3. Accuracy is 84.76%. |
| Tinier-YOLO: A Real-Time Object Detection Method for Constrained Environments. | **(J) / IEEE** | Wei Fang, Lin Wang, Peiming Ren | Tinier-YOLO-v3, PASCAL VOC (2007 + 2012), COCO. | Faster runtime speed compared to other lightweight models. But, is suitable for embedded systems. Accuracy will be a concern. |
| YOLO-ACN: Focusing on Small Target and Occluded Object Detection. | **(J) / IEEE** | Yongjun Li, Shasha Li, Haohao Du, Lijia Chen, Dongming Zhang, Yao Li | YOLO-ACN, MS COCO, Infrared pedestrian dataset KAIST, NVIDIA Tesla K40. | Doesn’t improve much performance with the proposed algorithm, compared to YOLO v3. It’s mainly focused on small objects detection. |
| Object Detection and Classification System for Visually Impaired. | **ICCSP (C) / IEEE** | Rashika Joshi, Meenakshi Tripati, Amit Kumar, Manoj Singh Gaur | MobileNetSSD(SSD -Single Shot-Detector)**,** PASCAL VOC 2007. | Got pretty good accuracy, but the dataset is small. According to current trends, it’s not sufficient. The algorithm is used only for embedded systems. |
| Robot Eye: Automatic Object Detection and Recognition Using Deep Attention Network to Assist Blind People | **ICPAI (C) / IEEE** | Ervin Yohannes, Paul Lin, Chih-Yang Lin, Timothy K. Shih | Self-designed model (DarkNet-53 based), ZED Stereo camera, PASCAL VOC, MS COCO datasets. | Accuracy is 81%, better than YOLO v3. Used PASCAL VOC for classes, and mixed MS COCO with them. Number of attributes are too small compared to COCO. |
| Real-Time Object Detection for Visually Challenged People | **ICICCS (C) / IEEE** | Sunit Vaidya, Naisha Shah, Niti Shah, Radha Shankarmani | YOLO v3, Datasets (Tiny YOLOv3, YOLOv3) | Showed better performance on Tiny YOLOv3 dataset compared to YOLOv3. They haven’t really trained the model, they used the pretrained weights. |
| An Evaluation of RetinaNet on Indoor Object Detection for Blind and Visually Impaired Persons Assistance Navigation | **Neural Processing Letters (J) / Springer** | Mouna Afif, Riadh Ayachi, Yahia Said, Edwige Pissaloux, Mohamed Atri | RetinaNet (ResNet, DenseNet, VGGNet based), Self prepared Dataset (Contains 8000 images). | Attained 84.61% mAP. Focused on only indoor navigation. A decent dataset to train a DCNN, but the number of objects it can detect is very small. Got pretty good results with the proposed algorithm. |
| Indoor object detection and recognition for an ICT mobility assistance of visually impaired people. | **Multimedia Tools and Applications (J) / Springer** | Mouna Afif, Riadh Ayachi, Edwige Pissaloux, Yahia Said, Mohamed Atri | YOLOv3, DarkNet-53. Dataset contains 8000 images and contains 16 indoor object classes. | Attained 73.19% mAP, and it’s only focused on indoor navigation. Later they presented a paper which showed much better  performance. Used pretrained weights of COCO and trained on the new dataset. |
| Deep Learning based Object Detection and Recognition Framework for the Visually-Impaired | **ICCMC (C) / IEEE** | Swapnil Bhole, Aniket Dhok | PASCAL VOC 2007 dataset, SSD, Inception v3 model. | They added currency detection to the dataset and achieved 90.2% accuracy on currency. But the dataset contains only 20 classes. |
| Assisting the Visually Impaired in Multi-object Scene Description Using OWA-Based Fusion of CNN Models | **Arabian Journal for Science and Engineering (J) / Springer** | Haikel Alhichri, Yakoub Bazi, Naif Alajlan | VGG16, SqueezeNet, Included OWA approach, CMOS based camera. | Dataset is too small (Less than 1000 images), and is completely focused on indoor objects. Trained with very low hyper-parameters, since the data is too small. Hence, the algoritm provided higher accuracy. |
| Intelligent Navigation System for the Visually Impaired - A Deep Learning Approach | **ICCMC (C) / IEEE** | Deepak Kumar Yadav, Somsankar Mookherji, Joanne Gomes, Siddarth Patil | Basic 4-Layered Convolutional Neural Network (CNN). | Achieved high accuracy on the dataset used. But the dataset is not sufficient. And the model is not deep enough for an Image recognition tasks. |

* Research on Small Target Detection in Driving Scenarios Based on Improved Yolo Network.
* **Publisher –** IEEE
* **13/01/2020** ([Artificial Intelligence (AI)-Empowered Intelligent Transportation Systems](https://ieeexplore.ieee.org/xpl/topics.jsp?isnumber=8948470&punumber=6287639&refinements=SpecialSection:Artificial%20Intelligence%20(AI)-Empowered%20Intelligent%20Transportation%20Systems)).
* **Authors –** Qiwei Xu, Runzi Lin, Han Yue, Hong Huang, Yun Yang, Zhigang Yao
* Used YOLO V3 , and made some advances to it, and achieved an accuracy of 84.76%, and the real-time detection on 2080 Ti reaches 41FPS, accuracy increased by 5.43%.
* Tinier-YOLO: A Real-Time Object Detection Method for Constrained Environments.
* **Publisher –** IEEE
* **24/12/2019**
* **Authors –** Wei Fang, Lin Wang, Peiming Ren
* Used Tiny-YOLO-V3, which can run on small embedded systems.
* YOLO-ACN: Focusing on Small Target and Occulded Object Detection.
* **Publisher –** IEEE
* **22/12/2020**
* **Authors –** Yongjun Li, Shasha Li, Haohao Du, Lijia Chen, Dongming Zhang, Yao Li.
* Used YOLO V3, and presented YOLO-ACN, and achieved decent mAPs.
* Object Detection and Classification System for Visually Impaired.
* **Publisher –** IEEE
* **01/09/2020** (2020 International conference on communication and signal processing).
* **Authors –** Rashika Joshi, Meenakshi Tripati, Amit Kumar, Manoj Singh Gaur
* Used **MobileNetSSD** (SSD -Single Shot Detector)**,** in an embedded system.
* Robot Eye: Automatic Object Detection and Recognition Using Deep Attention Network to Assist Blind People
* **Publisher –** IEEE
* **30/12/2020** (2020 International Conference on Pervasive Artificial Intelligence).
* **Authors –** Ervin Yohannes, Paul Lin, Chih-Yang Lin, Timothy K. Shih
* Used self-designed model (**DarkNet-53** as the backbone), which gave better results than **YOLO (You Only Look Once) v3**.
* Real-Time Object Detection for Visually Challenged People.
* **Publisher –** IEEE
* **19/06/2020** (2020 4th International Conference on Intelligent Computing and Control Systems)
* **Authors –** Sunit Vaidya, Naisha Shah, Niti Shah, Radha Shankarmani
* Used **YOLO v3**.
* An Evaluation of RetinaNet on Indoor Object Detection for Blind and Visually Impaired Persons Assistance Navigation
* **Publisher –** Springer
* **23/01/2020** (Neural Processing Letters)
* **Authors –** Mouna Afif, Riadh Ayachi, Yahia Said, Edwige Pissaloux, Mohamed Atri
* Used **RetinaNet.**
* Indoor object detection and recognition for an ICT mobility assistance of visually impaired people.
* **Publisher –** Springer
* **22/08/2020** (Multimedia Tools and Applications)
* **Authors –** Mouna Afif, Riadh Ayachi, Edwige Pissaloux, Yahia Said, Mohamed Atri
* Used **YOLOv3, DarkNet-53.**
* Deep Learning based Object Detection and Recognition Framework for the Visually-Impaired
* **Publisher –** IEEE
* **23/04/2020** ([2020 4th International Conference on Computing Methodologies and Communication](https://ieeexplore.ieee.org/xpl/conhome/9070072/proceeding))
* **Authors –** Swapnil Bhole, Aniket Dhok
* Used **PASCAL VOC 2007**  dataset, **SSD, Inception v3** model**.**
* Assisting the Visually Impaired in Multi-object Scene Description Using OWA-Based Fusion of CNN Models
* **Publisher –** Springer
* **29/07/2020** (Arabian Journal for Science and Engineering)
* **Authors –** Haikel Alhichri, Yakoub Bazi, Naif Alajlan
* Used **VGG16, SqueezeNet,** Included **OWA** approach (Ordered Weighted Average)
* Intelligent Navigation System for the Visually Impaired - A Deep Learning Approach
* **Pubisher –** IEEE
* **23/04/2020** (2020 [Fourth International Conference on Computing Methodologies and Communication](https://ieeexplore.ieee.org/xpl/conhome/9070072/proceeding))
* **Authors –** Deepak Kumar Yadav, Somsankar Mookherji, Joanne Gomes, Siddarth Patil
* Used basic **4-Layered Convolutional Neural Network** (CNN), but achieved high accuracy on the dataset used. But the dataset is not sufficient.

**Introduction**

Visually impaired people face a lot of difficulties in doing their daily activities. There is a say that, Out of all the five sense organs, eyes are most important. It’s also accepted by the scientists. Because more than 90% of our knowledge about the world comes from eyes.

They always need the help of other persons or either a stick. These methods are not always fruitful. Hence, we took up this project to help visually impaired people to recognize their surroundings. This can be done by using deep learning techniques.

Detecting and recognizing the objects and generating speech about the objects helps visually impaired in a great way in understanding their surroundings.

**Literature Review**

Hence, we took up this project to help visually impaired people to recognize their surroundings. In recent years, deep learning has become more popular technique for solving these problems of identifying objects.

The deep learning systems achieve high accuracy rates at lower cost. Convolution Neural Network (CNN) is one deep learning method that is being used in computer vision problems, such as autonomous car driving, helping visually impaired people walk outside.

There are many CNN-based methods to solve these detection and recognition, like Single Shot Detector (SSD), You Only Look Once (YOLO), these are one-stage detection architectures that predict the coordinates and the classes of the objects at the same time for fast processing.

There are also two-stage detection architectures like Faster R-CNN and Mask R-CNN. But YOLO algorithm is faster, accurate, light-weight compared to the methods mentioned above.

In this project we are going to use YOLO v5 algorithm, which is more faster, more accurate, and light-weight compared to other versions of YOLO.

Dataset we are planning to work-on is MS-COCO dataset, which has 80 classes/labels. Which means it can recognize 80 different objects.

After detecting and recognizing the objects, we are planning to generate speech, about the recognized algorithm. This can be achieved by using Recurrent Neural Network Techniques (RNN).

**Next Step**

**Image Classification –** Image classification is a supervised learning problem. It is a process of labelling objects in the image, classify or predict the class of a specified object in an image. The main goal is to accurately identify the feature in an image.

**Image Detection –** Image or Object Detection is a computer technology that processes the image and detects objects in it. And determine where objects are located in given image such as object localisation and which category each belongs to. If you need to classify the image detected, you need classification. But, if you just need to locate them, for example – find out the number of objects in the picture, you should use Image Detection.

**Image Recognition –** Image Recognition is the ability of AI to detect the object in an image, classify the detected object, and recognize it. It is the ability of software to identify objects, places, people, writing and actions in images.

Source:

* <https://medium.com/ai-techsystems/image-detection-recognition-and-image-classification-with-machine-learning-92226ea5f595>
* <https://analyticsindiamag.com/what-is-the-difference-between-image-classification-object-detection-techniques/>

**Dataset’s**

* <https://ieee-dataport.org/open-access/annotated-image-dataset-household-objects-robofeihome-team>
  + This data set contains two sets of pictures of household objects, created by the **RoboFEI@Home** team to develop object detection systems for a domestic robot.
  + The first data set was created with objects from a local supermarket. Product brands are typical from Brazil. The second data set is composed of objects from the **RoboCup@Home** 2018 OPL competition.
  + The data set contains the basic resources for the creation of custom object detection systems. Users can use the provided annotated images to train their own models and validate them in a set of test images, or they can analyze the inference time of the proposed methods in the provided videos.
* <http://web.mit.edu/torralba/www/indoor.html>
  + Indoor scene recognition is a challenging open problem in high level vision. Most scene recognition models that work well for outdoor scenes perform poorly in the indoor domain. The main difficulty is that while some indoor scenes (e.g. corridors) can be well characterized by global spatial properties, others (e.g., bookstores) are better characterized by the objects they contain. More generally, to address the indoor scenes recognition problem we need a model that can exploit local and global discriminative information.
* <https://cocodataset.org/#download>
  + MS COCO is among the most detailed image datasets as it features a large-scale object detection, segmentation, and captioning dataset of over 200,000 labeled images, 80 Object classes.
* <https://image-net.org/download.php>
  + The most highly-used subset of ImageNet is the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012-2017 image classification and localization dataset. This dataset spans 1000 object classes and contains 1,281,167 training images, 50,000 validation images and 100,000 test images. This subset is available on [Kaggle](https://www.kaggle.com/c/imagenet-object-localization-challenge/overview/description).
* <https://storage.googleapis.com/openimages/web/factsfigures.html>
  + Open Images is a dataset of ~9M images annotated with image-level labels, object bounding boxes, object segmentation masks, visual relationships, and localized narratives.

**CSPDarknet53:** it is a convolutional neural network and backbone for object detection that uses Darknet-53. It employes CSPNet (Cross Stage Partial Network) strategy to partition the feature map of the base layers into two parts and then merge them through a cross-stage hierarchy.

**PANet:** Path Aggregation Network, amins to boost information flow in a proposal-based instance segmentation framework. Specifically, the feature hierarchy is enhanced with accurate localization signals in lower layers by bottom-up path augmentation, which shortens the information path between lower layers and topmost features.

**YOLO Layer:** It’s a fully convolutional network. It has 75 convolutional layers, with skip connections and upsampling layers. No form of pooling is used, and a convolutional layer with stride 2 is used to downsmaple the feature maps.

**EfficientNet:** Making layers wider, deeper, and input image at higher resolution, or combination of all these improvements. As you can imagine, the exploration of all these possibilities can be quite tedious for machine learning researchers. EfficientNet set out to define an automatic procedure for scaling ConvNet model architectures.

**BiFPN Layer:** Weighted Bi-directional Feature Pyramid Network, which allows easy and fast multi-scale feature fusion.

**Class Loss:** Is also called **Classification loss.** It’s the loss that measures the correctness of the classification of each predicted bounding box, that contains an object or empty. This loss is usually called cross entropy loss.

**Box Loss:** Is also called **Localization loss.** It’s the loss that measures how tight the prediction bounding boxes are to the ground truth object. Measures the error in the predicted boundary box location and sizes.

**Object Loss:** Is also called **Confidence loss.** It’s the loss that measures wrong box-object IoU Prediction.

**Precision:** Measures how accurate your predictions are. i.e., the percentage of your predictions are correct.

Precision = TP / (TP + FP)

**Recall:** Measures how well you find all the positives.

Recall = TP / (TP + FN)

**mAP:** mean Average Precision is the average of AP (Average Precision for each class).

**mAP\_0.5:** mAP at (Intersection over Union) IoU threshold 0.5.

**mAP\_0.5:0.95:** mAP over different IoU thresholds, from 0.5 to 0.95