



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 5, May 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com

Object Detection Using Yolo and Tensor Flow

M. Chinnarao, R. Goutham Sai Kalyan, T. Naga Pravallika, B. Srinivas

Professor & HOD, Department of CSE, Lingayas Institute of Management and Technology, Madalavarigudem,
Andhra Pradesh, India

UG Students, Department of CSE, Lingayas Institute of Management and Technology, Madalavarigudem,
Andhra Pradesh, India

ABSTRACT: Object detection methods aim to identify all target objects in the target image and determine the categories and position information in order to achieve machine vision understanding. Numerous approaches have been proposed to solve this problem, mainly inspired by methods of computer vision and deep learning. However, existing approaches always perform poorly for the detection of small, dense objects, and even fail to detect objects with random geometric transformations. In this study, we compare and analyse mainstream object detection algorithms and propose a multi-scaled deformable convolutional object detection network to deal with the challenges faced by current methods.

Our analysis demonstrates a strong performance on par, or even better, than state of the art methods. We use deep convolutional networks to obtain multi-scaled features, and add deformable convolutional structures to overcome geometric transformations. We then fuse the multi-scaled features by up sampling, in order to implement the final object recognition and region regress. Experiments prove that our suggested framework improves the accuracy of detecting small target objects with geometric deformation, showing significant improvements in the trade-of between accuracy and speed.

KEYWORDS: Deep Learning, Neural Networks, Object Detection, YOLOv3, Residual Networks.

I. INTRODUCTION

The main purpose of object detection is to identify and locate one or more effective targets from still image or video data. It comprehensively includes a variety of important techniques, such as image processing, pattern recognition, artificial intelligence and machine learning. It has broad application prospects in such areas such as road traffic accident prevention, warnings of dangerous goods in factories, military restricted area monitoring and advanced human-computer interaction. Since the application scenarios of multi-target detection in the real world are usually complex and variable, balancing the relationship between accuracy and computing costs is a difficult task.

The object detection process is traditionally established by manually extracting feature models, where the common features are represented by HOG (histogram of oriented gradient), SIFT (scale-invariant feature transform), Haar (Haar-like features) and other classic algorithms based on grayscale. Following feature extraction, the SVM (support vector machine) or Adaboost algorithms are used for classification in order to obtain target information. These traditional extracting feature models are only able to determine low-level feature information, such as contour information and texture information, and have limitations in detecting multiple targets under complex scenes due to their poor generalization performance. However, object detection models, such as the R-CNN (region-based convolutional neural networks) series and the YOLO (you only look once) or SSD (single shot multiBox detection) models based on the deep learning CNN (convolutional neural network) features are more well-known.

Deep learning CNN models not only extract the detail texture features from pre-level convolution networks, but are also able to obtain higher-level information from the post-level convolution layer. Following the traditional CNN process, the R-CNN series uses an enumeration method to presuppose the target candidate region in the feature map, gradually fine-tuning the position information and optimizing the object position for classification and recognition. In contrast, other object detection models will simultaneously predict the bounding

box and classification directly in the feature map by applying different convolution sets. the R-CNN model has two operation stages (candidate region proposal and further detection) that allow for higher detection accuracy, while SSD and YOLO are able to directly detect the classification and position information, improving the detection speed. We propose a novel multi-scaled deformable convolution network model to deal with the trade-of between accuracy and

speed in object detection. The multi-scaled deformable convolutional neural network uses a new convolution method that has two offsets for image feature generation that are more sensitive to object deformation information. Additionally, the ability to detect objects that have geometrical deformations is improved. Secondly, feature fusion operations are performed on the multiple scale feature maps in the final detection. The image information of different scaled feature maps is simultaneously used to predict the classification and position information. This modification ensures the detection speed, enhances the target information of small objects, and also improves the accuracy of object detection.

The key contributions of our work are as follows:

1. The novel deformable convolution structure replaces the ordinary normal convolution operation for object detection. It effectively lets the CNN improve the generalization ability of extracting image features under different geometric deformations. Also, the new network automatically trains the offset of the convolution without wasting a large amount of computation time and cache space. Thus, significant performance gains on computer vision tasks, such as object detection and semantic segmentation, are observed.
2. An up-sample is applied to the feature pyramid to merge the multi-scaled feature information. This increases the accuracy of small target object detection by avoiding the loss of information of small target objects after multiple convolution and pooling operations. It also provides an important scheme for the detection of dense objects with overlapping occlusion in complex scenes.

KEYWORDS

The main keywords are: Object detection, Deformable convolution, Computer vision

OBJECT DETECTION

- **Object detection** is merely to recognize the object with bounding box in the image, where in image classification, we can simply categorize(classify) that is an object in the image or not in terms of the likelihood (Probability).
- **Object detection** is considered one of the noteworthy areas in the **deep learning and Computer vision**. **Object detection** has been determined the numerous applications in computer vision such as **object tracking, retrieval, video surveillance, image captioning, Image segmentation, Medical Image** and several greater number other applications as well.

DEFORMABLE CONVOLUTION

- Deformable convolution introduces the mechanism of deformable module, which has learn-able shape to adapt to the changes of features. Conventionally, the shapes of kernels and samples in convolution, defined by the sampling matrix, is fixed from the start.

COMPUTER VISION

- Computer Vision (CV) is a field of Artificial Intelligence (AI) that deals with computational methods to help computers understand and interpret the content of digital images and videos. Hence, computer vision aims to make computers see and understand visual data input from cameras or sensors.

YOLO

- 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 to detect traffic signals, people, parking meters, and animals.

II. METHODOLOGIES

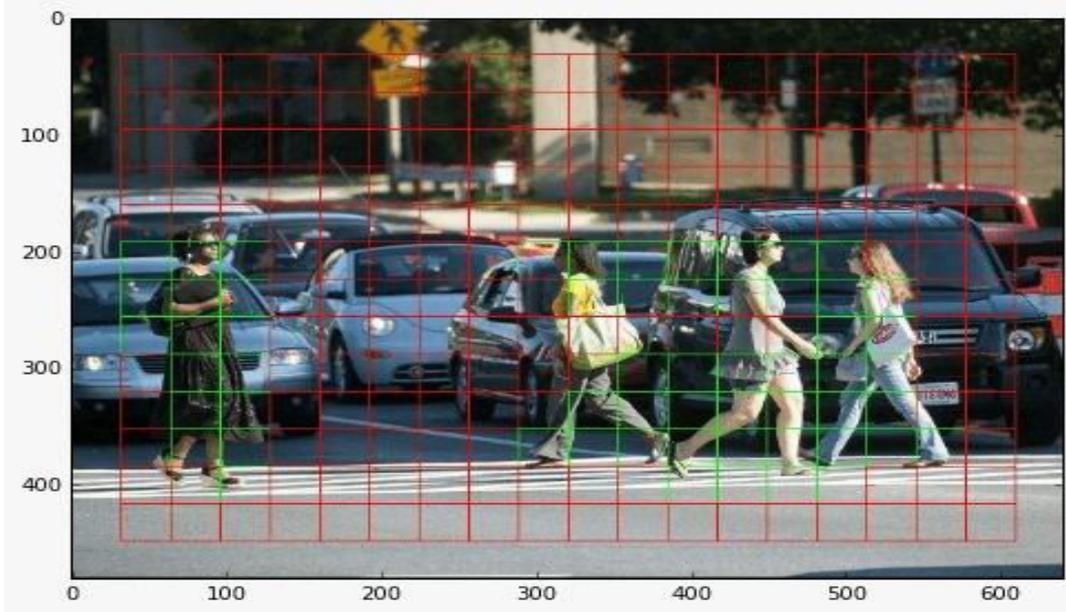
ALGORITHM

YOLO algorithm is used here.

- YOLO means You Only Look Once.
- YOLO algorithm works using the following three techniques:
 1. Residual blocks
 2. Bounding box regression
 3. Intersection Over Union (IOU)

1. Residual blocks

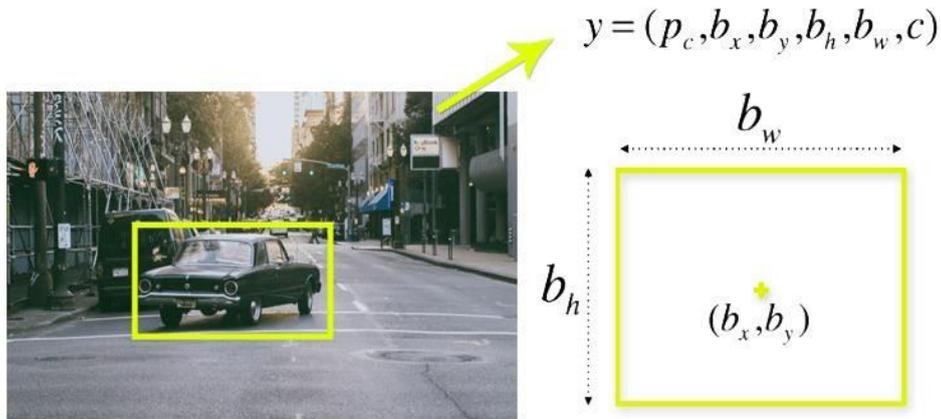
First, the image is divided into various grids. Each grid has a dimension of S x S. The following image shows how an input image is divided into grids.



Residual Blocks

2. Bounding box regression

- A bounding box is an outline that highlights an object in an image.
- Every bounding box in the image consists of the following attributes:
Width (bw)
Height (bh)
- Bounding box center (bx,by)
- The following image shows an example of a bounding box. The bounding box has been represented by a yellow outline.

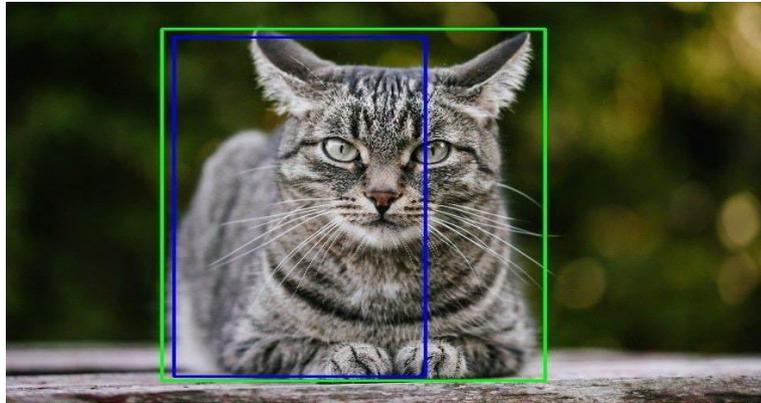


Bounding Box Regression

3. Intersection over union (IOU)

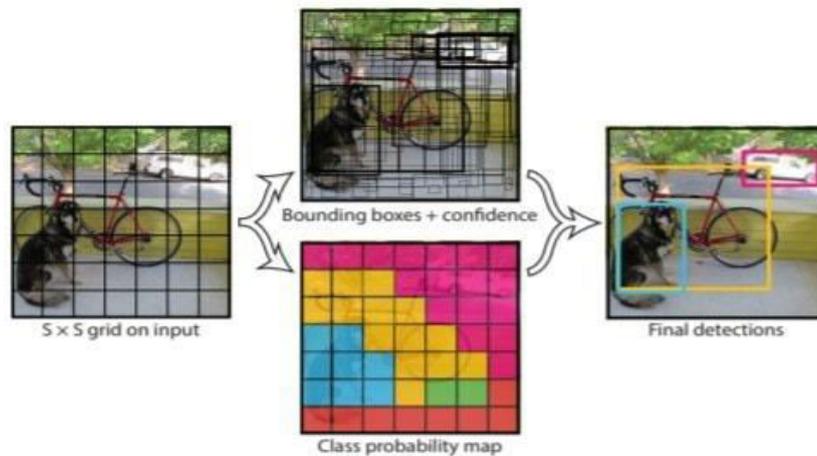
- Intersection over union (IOU) is a phenomenon in object detection that describes how boxes overlap.
- YOLO uses IOU to provide an output box that surrounds the objects perfectly.
- Each grid cell is responsible for predicting the bounding boxes and their confidence scores.

- The IOU is equal to 1 if the predicted bounding box is the same as the real box. This mechanism eliminates bounding boxes that are not equal to the real box. The following image provides a simple example of how IOU works.



Intersection over union (IOU)

- Combination of the three techniques:
The following image shows how the three techniques are applied to produce the final detection results.



Our image object detector adopts YOLO's backbone network and adds the new trick in convolution operation and feature information fusion. The overall framework is shown as Fig. 1. The first backbone network is the Darknet53 network. As a new network for performing feature extraction, it is a hybrid approach combining the network used in YOLO v2, Darknet-19, and the newer residual network tactics. The network, which is larger, uses successive 3x3 and 1x1 convolutional layers, with shortcut connections. In addition, we add three deformable convolution layers before the convolutional layers

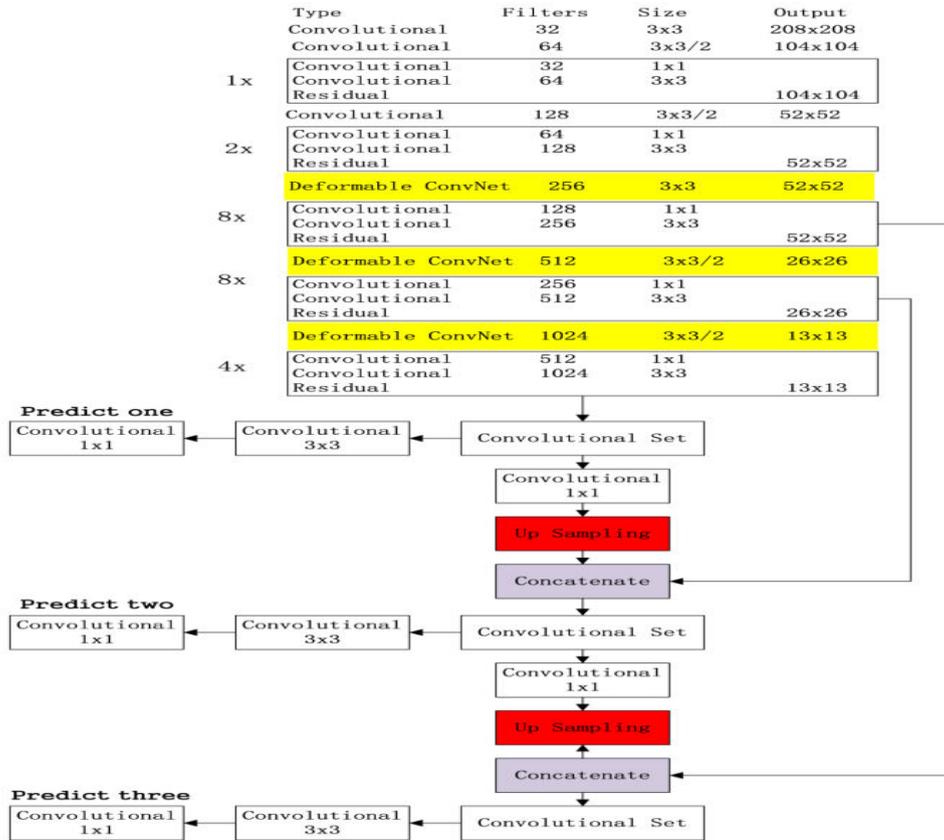


FIGURE 1

Architecture of the multi-scaled deformable convolutional neural network framework. The framework mainly consists of two components. (1) The backbone network based on Darknet53, which includes residual structures and deformable convolution. (2) The object detection network, based on multi-scaled detection and feature fusion.

A common convolution operation performs sampling in the input feature map X with a regular grid R , and sums the sample values under the weights, w . The grid R defines the size and expansion of the receptive field. For example, a 3×3 convolution kernel with an expansion size of 1 can be defined as follows:

$$R = \{(-1, -1), (-1,0) \dots (0,1), (1,1)\}, \quad (1)$$

For every output $y(P_o)$, the sampling must be performed with nine positions from X . These nine positions are in the shape of a grid diffused around a centre position $X(P_o)$. The coordinates $(-1, -1)$ represent the upper left corner of $X(P_o)$, while $(1,1)$ represent the lower right corner, with the remaining follow the same representation. Under traditional convolution, for each position P_o on the output feature map Y , we output the feature map formula Y as:

$$Y(P_o) = \sum w(P_n) \cdot X(P_o + P_n) \quad (2)$$

Under the deformation convolution, for each output $y(P_o)$, nine positions are sampled from X , in the shape of the grid that is diffused around centre position $X(P_o)$. Subsequently, a new parameter is added. The parameter ΔP_n allows the points of sampling to be diffused into an irregular shape as follows:

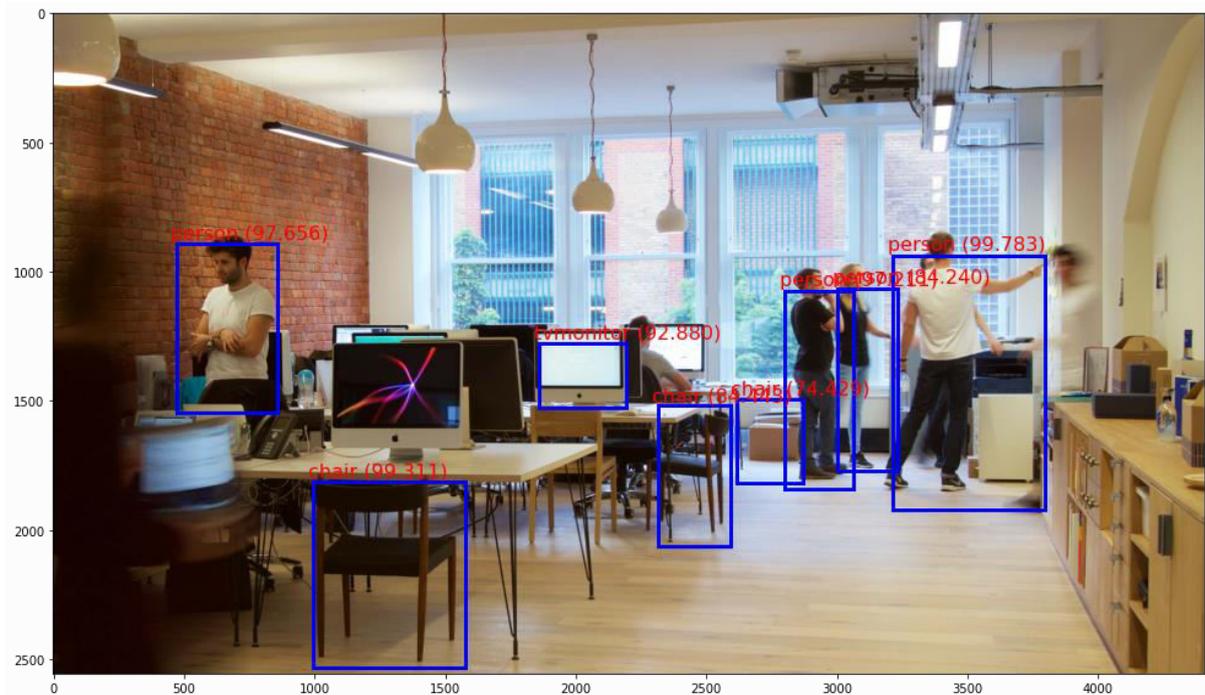
$$Y(P_o) = \sum w(P_n) X(P_o + P_n + \Delta p_n) \quad (3)$$

The new sampling is located at an irregular position with offset $P_n + \Delta P_n$. The offset ΔP_n is usually a decimal, while the $X(P_o)$ on the feature map is always a whole number. Formula (3) can be realized by bilinear interpolation:

$$X(P) = \sum G(q,p) \cdot X(q) \quad (4)$$

where P represents an arbitrary (decimal) position ($P = P_o + P_n + \Delta P_n$ in Eq. (2)),
 q enumerates all global spatial positions in the feature map X , and G is a bilinear interpolated kernel.

III. RESULTS



- If we pass an image as an input, then we will get the input as shown in above figure.
- We will get the output image with the objects identified in it.
- In the above figure, we can able to see that the objects are detected and are named with the corresponding object.

IV. CONCLUSION AND FUTURE SCOPE

Based on the trick of deformable convolutional networks, this study proposes a new multi-scaled deformable convolutional object detection network structure. This network uses a deformable convolution structure instead of an ordinary convolution operation in order to increase the learning ability of the model with respect to object geometric deformation, as well as increasing the accuracy of object detection. This study also uses multi-scaled feature maps that combine low-level features by up-sampling to extract target object position information. This increases the ability of the model to detect small target objects and dense objects, and also greatly makes up for the defect in missing detections, which is always present in other object detection models.

This object detection is used at

- **Autonomous driving:** YOLO algorithm can be used in autonomous cars to detect objects around cars such as vehicles, people, and parking signals. Object detection in autonomous cars is done to avoid collision since no human driver is controlling the car.
- **Wildlife:** This algorithm is used to detect various types of animals in forests. This type of detection is used by wildlife rangers and journalists to identify animals in videos (both recorded and real-time) and images. Some of the animals that can be detected include giraffes, elephants, and bears.
- **Security:** YOLO can also be used in security systems to enforce security in an area. Let's assume that people have been restricted from passing through a certain area for security reasons. If someone passes through the restricted area, the YOLO algorithm will detect him/her, which will require the security personnel to take further action.

REFERENCES

1. Xia, Z. (2019). An Overview of Deep Learning. *Deep Learning in Object Detection and Recognition*, 1-18. doi:10.1007/978-981-10-5152-4_1
2. Leung, H., & Haykin, S. (1991). The complex backpropagation algorithm. *IEEE Transactions on Signal Processing*, 39(9), 2101-2104. doi:10.1109/78.134446
3. Real-Time Object Detection for Aiding Visually Impaired using Deep Learning. (2020). *International Journal of Engineering and Advanced Technology Regular Issue*, 9(4), 1600-1605. doi:10.35940/ijeat.d8374.049420
4. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You Only Look Once: Unified, Real-Time Object Detection. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. doi:10.1109/cvpr.2016.91
5. Budiharto, W., Gunawan, A. A., Suroso, J. S., Chowanda, A., Patrik, A., & Utama, G. (2018). Fast Object Detection for Quadcopter Drone Using Deep Learning. *2018 3rd International Conference on Computer and Communication System (ICCCS)*. doi:10.1109/ccoms.2018.8463284
6. Li, X., Wang, J., Xu, F., & Song, J. (2019). Improvement of YOLOv3 Algorithm in Workpiece Detection. *2019 IEEE 9th Annual International Conference on CYBER Technology in Automation, Control, and Intelligent Systems (CYBER)*. doi:10.1109/cyber46603.2019.9066490
7. Zhao, L., & Wan, Y. (2019). A New Deep Learning Architecture for Person Detection. *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*. doi:10.1109/iccc47050.2019.9064172
8. Lu, Z., Lu, J., Ge, Q., & Zhan, T. (2019). Multi-object Detection Method based on YOLO and ResNet Hybrid Networks. *2019 IEEE 4th International Conference on Advanced Robotics and Mechatronics (ICARM)*. doi:10.1109/icarm.2019.8833671
9. Zhang, D. (2018). Vehicle target detection methods based on color fusion deformable part model. *EURASIP Journal on Wireless Communications and Networking*, 2018(1). doi:10.1186/s13638-018-1111-8
10. Kang, M., Leng, X., Lin, Z., & Ji, K. (2017). A modified faster R-CNN based on CFAR algorithm for SAR ship detection. *2017 International Workshop on Remote Sensing with Intelligent Processing (RSIP)*. doi:10.1109/rsip.2017.7958815
11. Hung, P. D., & Kien, N. N. (2019). SSD-MobileNet Implementation for Classifying Fish Species. *Advances in Intelligent Systems and Computing Intelligent Computing and Optimization*, 399-408. doi:10.1007/978-3-030-33585-4_40
12. Castiblanco, C., Rodriguez, J., Mondragon, I., Parra, C., & Colorado, J. (2014). Air Drones for Explosive Landmines Detection. *ROBOT2013: First Iberian Robotics Conference Advances in Intelligent Systems and Computing*, 107-114. doi:10.1007/978-3-319-03653-3_9
13. Ma, H., Liu, Y., Ren, Y., & Yu, J. (2019). Detection of Collapsed Buildings in Post-Earthquake Remote Sensing Images Based on the Improved YOLOv3. *Remote Sensing*, 12(1), 44. doi:10.3390/rs12010044
14. Khong, L. M., Gale, T. J., Jiang, D., Olivier, J. C., & Ortiz-Catalan, M. (2013). Multi-layer perceptron training algorithms for pattern recognition of myoelectric signals. *The 6th 2013 Biomedical Engineering International Conference*. doi:10.1109/bmeicon.2013.6687665
15. Andre, T., Neuhold, D., & Bettstetter, C. (2014). Coordinated multi-robot exploration: Out of the box packages for ROS. *2014 IEEE Globecom Workshops (GC Wkshps)*. doi:10.1109/glocomw.2014.7063639



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details