IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

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

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 11, Issue 4, April 2022

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.118

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104133 |

Object Detection and Recognition and Age-Gender Prediction using YOLO and OpenCV

Musaddik Moulavi[1], Anuprita Barge[2], Rajan Vishwakarma[3], Bhushan Dhengle[4], Prof. Anisaara Nadaph[5]

U.G. Student, Department of Computer Engineering, Trinity College of Engineering & Research, Pune, Maharashtra, India[1,2,3,4]

Associate Professor, Department of Computer Engineering, Trinity College of Engineering & Research, Pune, Maharashtra, India [5]

ABSTRACT: Using video cameras for watching the field is common in day-today life. Most of those police investigation use human to watch the activities that's happening within the space of interest. But victimization human in surveillance has its own disadvantage; to beat that limitation researchers are operating in a much automated visual surveillance systems. The visual surveillance method includes of the subsequent steps: setting modelling, motion segmentation, object classification, tracking, behaviour understanding, human identification and knowledge fusion. The primary and foremost step in visual surveillance is characteristic moving objects in a video sequence. The moving object of interest could also be human being, vehicle, etc. Object detection is that the technology that deals with characteristic the linguistics category of the moving object within the video sequence. Hence Object Detection is extremely very important for pursuit moving object and behaviour analysis in the given video sequence. Considering the importance of object detection in visual police investigation, this paper presents numerous strategieson the market for object detection. Automatic Visual surveillance has wide space of applications resembling human identification at a distance, watching the congestion, detection of abnormal behaviours and so on

KEYWORDS: Object detection, combining YOLO and Fast R-CNN, OpenCV, Object tracking.

I. INTRODUCTION

Humans look at a picture and instantly understand what objects are within the image, wherever they are, and the way they interact. The human sensory system is quick and correct, permitting United States of America to perform complicated tasks like driving with very little acutely aware thought. Fast, accurate algorithms for object detection would enable computers to drive cars while not specialised sensors, modify helpful devices to convey period of time scene data to human users, and unlock the potential for general purpose, responsive robotic systems. Current detection systems repurpose classifiers to perform detection. To detect an object, these systems take a classifier for that object and measure it at numerous locations and scales in a very take a look at image. Systems like deformable components models (DPM) use a window approach wherever the classifier is run at equally spaced locations over the whole image. Newer approaches like R-CNN use region proposal ways to initial generate potential bounding boxes in a picture so run a classifier on these projected boxes. When classification, post-processing is employed to refine the bounding boxes, eliminate duplicate detections, and rescore the boxes supported different objects within the scene. Reframe object detection as one regression problem, straight from image pixels to bounding box coordinates and class chances. Victimisation system like, you merely look once (YOLO) at a picture to predict what objects are gift and wherever they are. YOLO is refresh fully simple. One convolutional network at the same time predicts multiple bounding boxes and sophistication probabilities for those boxes. YOLO trains on full pictures and directly optimizes detection performance. This unified model has many advantages over ancient ways of object detection. First, YOLO is extraordinarily fast. merely run neural network on a brand new image at take a look at time to predict detections. Base network runs at forty five frames per second with no execution on a Titan X GPU and a quick version runs at over one hundred fifty fps. to urgea lot of accuracy we have a tendency to mix YOLO and fast R-CNN.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104133 |

II. RELATED WORK

Object detection (OD) system finds objects in the real world by making use of the object models which is known a priori. This task is comparatively difficult to perform for the machines as compared to Humans who perform OD very effortlessly and instantaneously. This paperwill give a review of the various techniques and approaches that are used to detect objects in images. Basically an OD system can be described easily by seeing, which shows the basic stages that are involved in the process of OD. The basic input to the OD system can be an image or a scene in case of videos. The basic aim of this system is to detect objects that are present in the image or scene or simply in other words the system needs to categorise the various objects into respective object classes [2]. Fast R-CNN first fine-tunes a CommNet on object proposals using log loss. Then, it fits SVMs to ConvNet features. These SVMs act as object detectors, replacing the softmax classifier learnt by fine-tuning. In the third training stage, bounding-box repressors are learned.R-CNN is slow because it performs a ConvNet forward pass for each object proposal, without sharing computation. Spatial pyramid pooling networks (SPPnets) were proposed to speed up R-CNN by sharing computation. The SPPnet method computes a convolutional feature mAP for the entire input image and then classifies each object proposal using a feature vector extracted from the shared feature mAP. Features are extracted for a proposal by maxpooling the portion of the feature mAP inside the proposal into a fixed-size output (e.g., 6 × 6)multiple output sizes are pooled and then concatenated as in spatial pyramid pooling. SPPnet accelerates R-CNN by 10 to 100× at test time. Training time is also reduced by 3× due to faster proposal feature extraction.

III. METHODOLOGY

OpenCV-Python: OpenCV 4.0 supports the deep learning frameworks like Fast R-CNN, YOLO, and easy integration with an OpenCV application, OpenCV‟s CPU implementation of the DNN module is astonishingly fast. For example, Darknet when used with OpenCV takes about 2 seconds on a CPU for inference on a single image. In contrast, OpenCV‟s implementation runs in a mere 0.22 Compared to other programming languages python is suitable for my project, OpenCV supports Python 3.7, and it is a package based language. It is very easy and simple.

1. **Real-time detection:**Unify the separate components of object detection into a single neural network. The network uses features of the entire image to predict each bounding box. It also predicts all bounding boxes in all classes for an image at once. This means that the network makes global judgments about the entire image and all objects in the image. YOLO's design allows for end-to-end training and real-time speeds while maintaining high average accuracy.The system divides the input image into an S×S grid. When the centre of an object falls on a grid cell, that grid cell is responsible for detecting that object. Each grid cell predicts bounding boxes B and confidence values for those boxes. These confidence values reflect how confident the model is that the box contains an object and also how accurate it thinks the predicted box is.If there is no object in that cell, the confidence values must be zero. Otherwise, you want the confidence value to be equal to the intercept over the union (IOU) between the predicted box and the ground truth. Each bounding box consists of 5 predictions: x, y, w, h and confidence. The (x,y) coordinates represent the center of the box relative to the grid cell boundaries.The width and height are predicted relative to the entire image. Finally, the confidence prediction represents the IOU between the predicted frame and each ground truth frame. Each grid cell also predicts conditional class probabilities C. These probabilities depend on the grid cell containing an object that predicts only one set of class probabilities per grid cell, regardless of the number of cellsB.At test time, multiply the conditional class probabilities and the individual box confidence predictions, giving us class-specific confidence values for each box. These values encode both the likelihood that that class will appear in the box and how well the predicted box fits the object.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104133 |

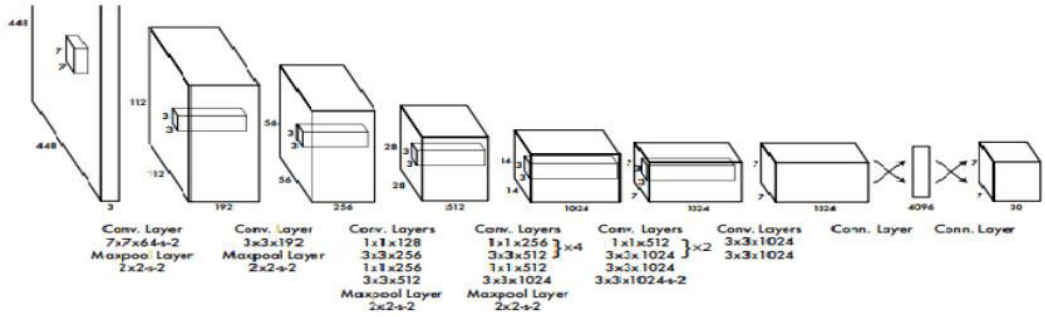


Figure 1: Shows Architecture detection network has 24 convolutional layers followed by 2 fully connected layers. Alternating 1 × 1 convolutional layers reduce the features space from preceding layers. Network pre train the convolutional layers on the PASCAL VOC 2012 classification task at half the resolution (224 × 224input-image) and then double the resolution for detection.

2. **Network design**: Implementing this model as a convolutional neural network and valuate it on the PASCAL VOC detection dataset. The initial convolutional layers of the network extract features from the image whereas the totally connected layers predict the output chances and coordinates. Network architecture is galvanized by the GoogleNet model for image classification; network has twenty four convolutional layers followed by two fully connected layers. Rather than the origin modules employed byGoogleNet, merely use one × 1 reduction layers followed by three × 3 convolutional layers, almost likeLin et al. the total network is shown in Figure 1. conjointly train a quick version of YOLO designed to push the boundaries of fast object detection. Quick YOLO uses a neural network with fewer convolutional layers (9 insteadof 24) and fewer filters in those layers. Exceptthe scale of the network, all coaching and testing parameters are constant between YOLO and quick YOLO.

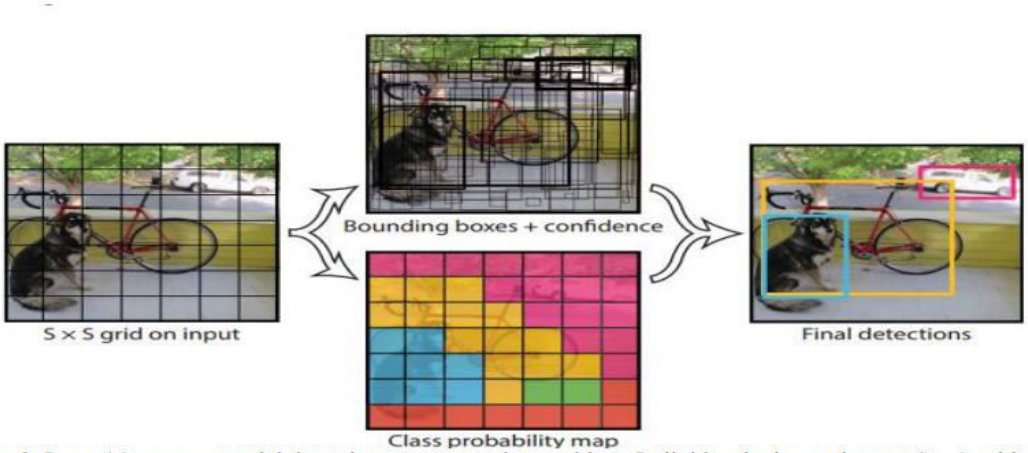


Figure 2: Input (a), system model detection as a regression problem. It divides the image into an S × S grid and for each grid cell predicts B bounding boxes, confidence for those boxes, and C class probabilities. Thesepredictions are encoded as an S × S × (B ∗ 5 + C) tensor. For evaluating YOLO on PASCAL VOC, use S = 7, B = 2. PASCAL VOC has 20 labelled classes so C = 20.final prediction is a 7 × 7 × 30 tensor as output (b).

3. **Face and Gender Prediction:**

CAFFE Model: CAFFE (Convolutional design for quick Feature Embedding) could be a deep learning framework, originally developed at University of California, Berkeley. It's open source, beneath a BSD license. It's written in C++, with a Python interface. Caffe supports sorts of deep learning ideasconnectedwithin the fields of image classification and image segmentation. It supports CNN and totally connected neural network designs. Supports kernel libraries cherish NVIDIA, CNN and Intel MKL. During this project Caffe model helps United States of Americaoutlinethe interior states of the parameters of the layers[27]. Protocol Buffer Files: Protocol Buffers (Protobuf) is a free associated open supply cross-platform library. They're used for knowledge serialization. These are tensorflow files that are accustomed describe the network configuration. The protobuf files are written in xml which has .pbtxt extension. Whereas the files with .pb extension contain data in binary format which is difficult to read. Google developed Protocol Buffers for internal use and provided a code generator for multiple languages beneath an open source license. These

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104133 |

Protocol Buffers were designed with an aim for simplicity and higher performance. Also were aimed to be quicker than XML. But these are used at Google to store and interchange numerous types of data. Also used for several inter-machine communication. Limitations to YOLO: YOLO imposes robust spatial constraints on bounding box predictions since every grid cell solely predicts 2 boxes and might only have one class. This spatial constraint limits the quantity of close objects that model can predict. This model struggles with tiny objects that seem in groups, cherish flocks of birds. Since model learns to predict bounding boxes from data, it struggles to generalize to things in new or uncommon side ratios or configurations. This model also uses comparatively coarse options for predicting bounding boxes since design has multiple down sampling layers from the input image. Finally, whereas train on a loss perform that approximates detection performance, loss function treats errors an equivalent in small bounding boxes versus massive bounding boxes. A little error in an exceedingly large box is usually benign however a small error in a small box features a lot of bigger result on IOU, main supply of error is wrong localizations.

IV. EXPERIMENTAL RESULTS

First, compare YOLO to other real-time detection systems in PASCAL VOC 2007. To understand the differences between YOLO and RCNN variants, examine the bugs in VOC 2007 caused by YOLO and Fast RCNN, one of the RCNN versions. Error profiling shows that YOLO can be used to re-evaluate fast RCNN detections and reduce false positive background errors, resulting in a significant performance boost. Also present the results from VOC 2012 and compare mAPto the most advanced current methods. Finally, show that YOLO generalizes to new domains on two datasets of artworks better than other detectors. To further explore the differences between YOLO and state-of-the-art detectors, take a look at a detailed breakdown of the results in VOC 2007. Compare YOLO to Fast RCNN, as Fast RCNN is one of the most powerful detectors in PASCAL VOC and its detections are publicly available. Using the methodology and tools of Hoiem et al. Check for each category at the time of the test the best N predictions for that category. Each prediction is correct or classified by the type of error.

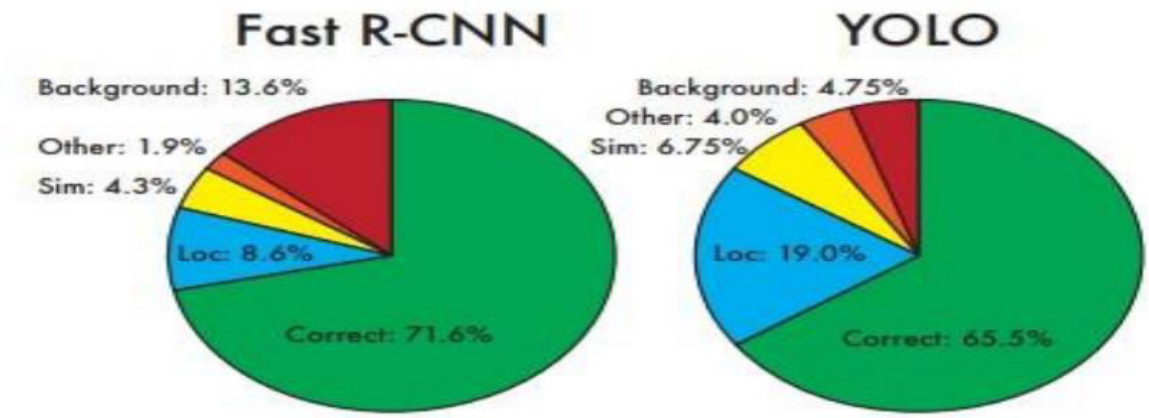
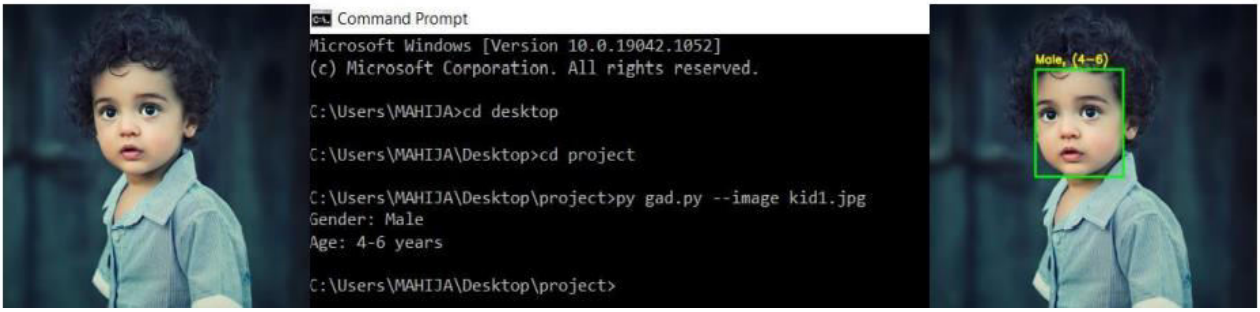


Figure 3: Error Analysis: Fast R-CNN vs. YOLO these charts show the percentage of localization and background errors in the top N detections for various categories (N = # objects in that category).

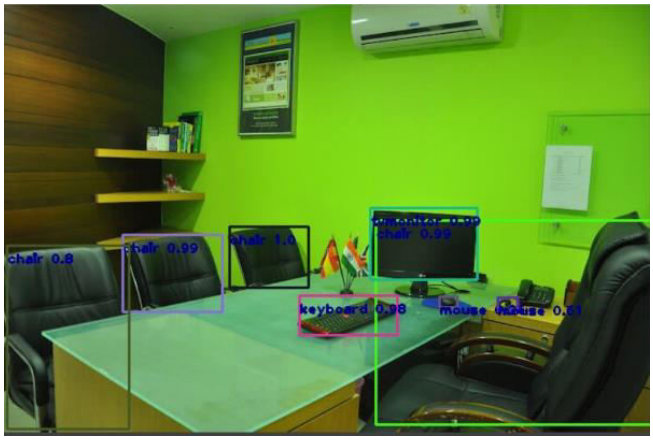


From the above image when given a picture, the child's face is detected and the age, gender are shown. The child in the image is of the age group 4-6 who is a male. The output satisfies the age and gender. The output is shown with a green square box that shows the face of the child and above the box are the gender and age.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104133 |



In the above image, the objects are detected by giving an image file as an input

In the above image, the objects are detected by webcam

V. CONCLUSION

Object detection is a keycap potential for maximum laptop and robotic vision systems. Although superb development has been found with inside the closing years, and a few present strategies at the moment area part of many purchaser electronics (e.g., face detection for auto-cognizance in smartphones) or were included in assistant riding technologies, People none the less a long way from attaining human-stage overall performance, especially in phrases of open-global learning. Being a person of YOLO, it's far a unified version for item detection and clean to construct and may beskilledat once on images. Unlike classifier-primarily based totally methods, YOLO is properly skilled on a loss characteristic that at once corresponds to detection overall performance and the entire version is skilled jointly. Fast YOLO is the brand new and quickest general-motiveitem detector within side the literature. YOLO has pushed the state-of-the-art in real-time item detection. YOLO can Generalizes properly to new domain names making it perfect for software. In spite of speedy improvement and accomplished development of item detection, there are nonetheless many open troubles for destiny work. Multi-assignment joint optimization and multi-modal statistics fusion. Due to the correlations among special obligations inside and outdooritem detection, multi-assignment joint optimization has already been studied by many researchers. Scale adaption. Objects typically exist in special scales, which is extra obvious in face detection and pedestrian detection. To boom the robustness to scale changes, it's far demanded to educate scale-invariant, multi-scale or scale adaptive detectors. Spatial correlations and contextual modelling. Spatial distribution performs an essential position in item detection. So area idea era and grid regression are taken to reapprobably item locations. However, the correlations among more than one proposal and item categories are ignored. Besides, the worldwide shape statistics is deserted through the position-touchy rating mAPs in R-FCN. Unsupervised and weakly supervised learning. It's very time eating to manually draw huge portions of bounding boxes. The age and gender detection of the use of OpenCVcould be very useful in authorization functions, clinicalfunctions or surveillance functions. The CNN and OpenCVmixed can provide superb results. The OIU-Audience dataset used withinside the assignment offer send result with more accuracy. We used protocol buffer and Caffeversion files. This assignment indicates how OpenCV may be used for face detection with nodifferent complex process.

The Future scope of work: This assignment may be superior in few methods such that this assignment may be used to its fullest: 1) Application- The assignment may beevolved into an internet software or a cellularsoftware such that it's far without problems accessible. 2)In public locations- the use of sensors will be utilized in public locations like restaurants, ATM locations, and stores such then while a robbery occurs the scope of locating the character will be plenty extra clean. 3) Enhancing this assignment to locate more than one people-this assignment may be superior such it may estimate age and gender even for a set of people with inside the image. This version does locate the face of people in a sethowever cannot provide the correct age and gender estimation

REFERENCES

[1] M. Bertalmio,G.Sapiro,V.Caselles,and C. Ballester, "Image inpainting", inProc.SIGGRAPH, pp. 417–424, 2000.
[2] A. Criminisi, P. Perez, and K. Toyama, "Region filling and object removal by exemplar-based image inpainting.", IEEE Transactions on Image Processing, vol. 13, no.9, pp. 1200–1212, 2004.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104133 |

[3] Marcelo Bertalmio, LuminitaVese, Guillermo Sapiro, Stanley Osher, "Simultaneous Structure and Texture Image Inpainting", IEEE Transactions On Image Processing, vol. 12, No. 8, 2003.
[4] Yassin M. Y. Hasan and Lina J. Karam, "Morphological Text Extraction from Images",IEEE Transactions On Image Processing, vol. 9, No. 11, 2000
[5] Eftychios A. Pnevmatikakis, Petros Maragos "An Inpainting System For Automatic Image Structure-Texture Restoration With Text Removal", IEEE trans. 978-1-4244-1764, 2008
[6] S.Bhuvaneswari, T.S.Subashini, "Automatic Detection and Inpainting of Text Images", International Journal of Computer Applications (0975 – 8887) Volume 61– No.7, 2013
[7] Aria Pezeshk and Richard L. Tutwiler, "Automatic Feature Extraction and Text Recognition from Scanned Topographic Maps", IEEE Transactions on geosciences and remote sensing, VOL. 49, NO. 12, 2011
[8] Xiaoqing Liu and JagathSamarabandu, "Multiscale Edge-Based Text Extraction From Complex Images", IEEE Trans., 1424403677, 2006
[9] Nobuo Ezaki, Marius Bulacu Lambert , Schomaker , "Text Detection from Natural Scene Images: Towards a System for Visually Impaired Persons" , Proc. of 17th Int. Conf. on Pattern Recognition (ICPR), IEEE Computer Society, pp. 683-686, vol. II, 2004
[10] Mr. Rajesh H. Davda1, Mr. Noor Mohammed, " Text Detection, Removal and Region Filling Using Image Inpainting", International Journal of Futuristic Science Engineering and Technology, vol. 1 Issue 2, ISSN 2320 – 4486, 2013
[11] UdayModha, Preeti Dave, " Image Inpainting-Automatic Detection and Removal of Text From Images", International Journal of Engineering Research and Applications (IJERA), ISSN: 2248-9622 Vol. 2, Issue 2, 2012
[12] Muthukumar S, Dr.Krishnan .N, Pasupathi.P, Deepa. S, "Analysis of Image Inpainting Techniques with Exemplar, Poisson, Successive Elimination and 8 Pixel Neighborhood Methods", International Journal of Computer Applications (0975 – 8887), Volume 9, No.11, 2010

INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118

doi cross ref

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

Scan to save the contact details