International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

# INFRARED IMAGE PEDESTRIAN DETECTION VIA YOLO v3

[1] SVMG Phani Kumar C, [2] Daram Sravani, [3] *Kummari Sruthi*

[1] *Assistant Professor, Department of Electronics and Communication Engineering, BhojReddy Engineering College for Women, Hyderabad, Telangana, India.*
[2,3,4] *Students, Department of Electronics and Communication Engineering, BhojReddy Engineering College for Women, Hyderabad, Telangana, India.*

**Abstract**

The principle of infrared image is thermal imaging technology. Infrared pedestrian detection technology can be applied to the safety monitoring of the elderly, which can not only protect personal privacy, but also realize pedestrian identification at night, which has strong application value and social significance. A method of infrared image pedestrian detection with improved YOLOv3 algorithm is proposed to increase the detection accuracy and solve the problem of low detection accuracy caused by infrared pedestrian target edge blurring. And according to the characteristics of infrared pedestrian, a complex sample data set is established which is applied to infrared pedestrian detection. The infrared image enhancement method with WDSR-B is adopted to improve the clarity of the data set. In addition, based on YOLOv3 algorithm, the output of the 4-time down-sampling layer is added to obtain richer context information for small targets and improve the detection performance of the network for small-target pedestrians. And the improved YOLOv3 network is trained by the enhanced infrared data set.

## I INTRODUCTION

In today's life, people's demands for safety are getting higher and higher. Video-based security systems are widely used in banks, transportation, military, and even homes. However, most video-based security systems currently require manual monitoring, and human fatigue or negligence introduces uncertainty to the security system. Therefore, an intelligent security system based on pedestrian detection is of great significance for ensuringpeople's normal life and work. The traditional pedestrian detection method is to design some hand-designed algorithms and features, such as ACF and LDCF algorithms. Thanks to the development of image classification tasks based on deep learning, the target recognition and detection technology in computer vision has made great progress in recent years. RCNN applied deep learning technology to target detection tasks for the first time.the Fast R-CNN algorithm, which greatly improves the detection speed ofR-CNN through the region candidate mechanism Liu proposed an end-to-end targetdetection algorithm SSD, by setting features of different scales and resolutions the accuracy of target detection. The YOLO series target detection network proposed by Redmon takes into account the accuracy and speed of target detection and others applied the target detection algorithm based on deep learning to the task of pedestrian detection, which achieved accuracy far exceeding traditional methods. Although the current pedestrian detection algorithm based on deep learning has achieved amazing recognition results, the traditional RGB camera video stream data collection requires certain lighting conditions and cannot adapt to the dark environment atnight. The night is the time when security accidents are most likely to occur, which createsloopholes in the security system. How to detect illegal intrusions in important target areas atnight has become an urgent problem to be solved. Besides the automotive safety, pedestriandetection has also been widely used in many other applications, such as robotics and surveillance.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362


Figures (a) and (b) are pictures obtained through an RBG camera, (c) and (d) are theimages collected by the infrared camera.

Infrared and RBG cameras capture images. Figures (a) and (b) are pictures obtainedthrough an RBG camera at night when there is no external light source, and it is basically impossible to collect target information. (c) and (d) are the images collected by the infraredcamera. As shown in (d), the edges of infrared imaging human contours are relatively blurred, and detailed information such as textures are lacking; the spatial distance of the target is different at any time, the imaging scale difference is obvious; the human target is easily annihilated by the environmental background in a complex environment.

Background According to the preliminary data from Governs Highway Safety Association, the number of pedestrian fatalities in the United States increased significantly recent years. About 6,000 pedestrian fatalities are estimated to have occurred in 2016, which could make 2016 the first year in more than two decades with more than 6,000 pedestrian deaths. On the other hand, intelligence techniques, such as self-driving technology, develop rapidly tobenefit people's daily life. Therefore, automotive safety has been an important issue that people are concerned about. In order to guarantee their safety, the accuracy of pedestrian detection is much more significant. Besides the automotive safety, pedestrian detection has also been widely used in many other applications, such as robotics and surveillance. Thesecritical applications attract great attention to theresearchers. They come up with different methods to improve the accuracy for pedestrian detection. From designing the algorithms for the base features of pedestrians to improving the learning algorithms, researchers have made great contributions to this area. However, pedestrian detection is a challenging task due to its complexity background as well as various body sizes and postures.

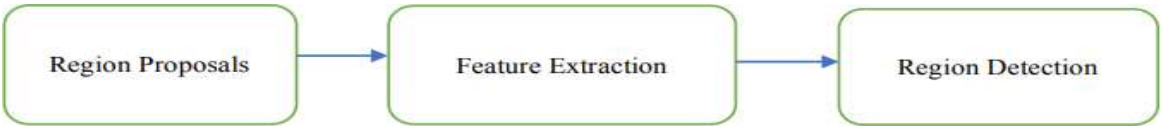
Traditionally, researchers focus more on capturing low-level feature extraction of pedestrians by manually designing some algorithms. Such hand-crafted method has two steps to compute features, selecting the region of interest in the image, like the corners, and then using a descriptor to calculate the characteristics of the region. The descriptor can distinguish the characteristics from others. However, using hand-crafted method has to finda good trade-off between the accuracy and efficiency of computation. Recently, with the development of deep learning technology, researchers take advantage of deep neural network, especially the convolution neural network to automatically extract the features from original images. Instead of extracting the low level feature as traditional methods, deep learning method extracts high-level feature due to deep layers of convolution neural network. For the first several layers, convolution neural network just learns the low-level features, such as edge, dots and colors, while the later layers will learn to recognize the general shape of the objects and get a high-level representation of the image. So, deep learning method can discover multiple levels of representation of the images through multiple layers and can be considered as feature extractor.These features extracted by convolution neural network are directly learn from the data, which are different from the traditional hand-crafted methods that the features are designedby the experts. Besides feature extraction for the pedestrians, classification and detection of the region is also an important part that researchers are focus on. For hand- crafted methods, typically it should combine with the classifiers to classify and detect the features that are extracted before. For the deep learning methods, it can be trained from end-to-end,extracting features from the original image and classifying the result from the last layer of convolution neural network, which means that convolution neural network can be considered as both feature extractor and classifier.In order to get higher efficiency and accuracy on detecting pedestrians, utilize deeplearning method could be an effective way. In the previous work, researchers combine thehand-crafted methods with deep neural network, leading

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
ISSN: 2457-0362

good results of pedestrian detection, but hand-crafted methods are still slower than the efficient network, since these are implemented on CPU while the region-based CNN can be implemented on GPU.

Recently, Faster RCNN, can purely use convolution neural network to achieve the detection of objects without any hand-crafted method. The result has shown that it has great performance on object detection. Pedestrian detection, a specific category of objects, can also utilize part of network in Faster RCNN to implement the process of detection in deep neural network.

Pedestrian Detection has been an important research area in computer vision for decades of years and researchers propose a lot of methods. Typically, they solve such problem have similar pipeline: region proposals, feature extraction and region classification




Pedestrian detection pipeline

At the first stage, regions of interest are selected in the images. Researchers come up with some methods to select the regions as candidates, some of them include parts of pedestrians, others do not. Then the features of these candidate regions will be extracted. These features can be considered as representations of the whole image and feed to the classifiers. Finally, train the classifiers to recognize these features. These classifiers can provide a binary flag to indicate whether pedestrians exist in the candidate regions.

For region proposals, traditionally researchers use hand-crafted method, selecting the region of interests in the image based on the low-level features. Sliding window is the most common method that people use in the research. In general, the detected image is fixed andthe sliding window with different scales will traverse the whole image from the top left to the bottom right, this method is simple but time-consuming. Edge Box is another method based on sliding windows. Instead of traversing every location of the image, this method accelerates the process of proposing the candidate region. The authors use edges information to generate bounding box for the objects.

The main idea is that they observed the number of contours that totally enclosed in the bounding box can indicate the probability that the bounding box includes objects. The edgescorrespond to the boundaries of the objects and when all the pixels of the edges are in the bounding box, the edges are thought to be wholly enclosed in it. Since the number ofthe bounding box could be so large, the authors came up with a method that they must scorethese candidates to select the best one. Firstly, they utilized the Structured Edge detector proposed in and to get the initial edge-map. After that they grouped the neighbor pixels with the similar orientations as the edge groups and calculated the affinity for the neighboring groups, shown on the third image of the Figure 1.3. According to the affinity, the boundary and the score of the box can be calculated. They used sliding window method to find the potential candidate bounding boxes with different scale and aspect ratio. Figure1 shows the process of edge box method that find the example of correct bounding box

from left to right: Original image, initial edge map used Structured Edges, edge groups and example of correct bounding box.Selective search is a hand-crafted method that based on super-pixels. It generally groupsthe pixels of similar colors together and generates the bounding box around these regions. There are three main design considerations they proposed in the paper: capture all scales, diversification and fast to compute. For capturing all scales, they proposed the hierarchicalgrouping that combined with segmentation method, generating all location switch differentscales and grouping these regions until the single region is left. The way to group the regions is the following: 1) use the segmentation method to create the initial image. 2) calculate the similarity between all these neighbouring regions 3) group the two most similar regions and calculate the new similarity

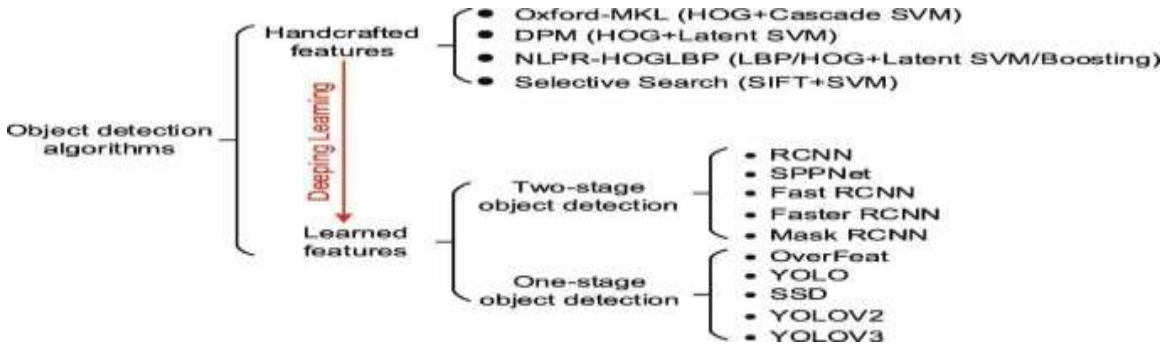
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

between this new region and its neighbouring regions. 4) repeat the process until the whole image becomes a single region.The steps from 2) to 4) are considered as greedy algorithm.

## II LITERATURE SURVEY

At present, researches on infrared image pedestrian detection are mainly divided into two categories: one is based on traditional machine vision method, and the other is based on deep learning method. The application of the latter in infrared image pedestrian detection technology is gradually maturing and improving.

In the field of traditional machine vision, the background subtraction method based on theimproved Gaussian mixture model (GMM) to segment human targets,and at the same time, he realized the correct detection of human targets in complex scenesby using Support vector machine (SVM). The three-frame differencemethod and adopted the three-frame differencemethod based on regional estimation to realize pedestrian detection in vehicle-mounted infrared images. The significance detection principle based on frequency domain to generatethe region of interestgraph, and trained the neural network to generate the pedestrian targetprobability graph, thus realizing pedestrian detection. With the development of target detection algorithms inthe field of deep learning, many algorithms are used to solve the problem of infrared pedestrian detection. The SE- MSSD framework based on SSD improvement. At the target classification ability of YOLOv3 model was notperfect, so he integrated the idea of weighted demarcation of features in SENet into YOLOv3 to better describe pedestrian characteristics. The LenET-7 system which contains 3 convolutional layers, 3 pooling layers and 1 output layer, solves the problem of miscellaneous parameters of full convolutional neural network and improves the detection rate of infrared pedestrian images.Pedestrian detection in infrared (IR) aerial night vision has gradually become an important research direction in recent years, and is critically applied in the fields of object search, person re-identification, marine trash detection, intelligent driving, and so on. In recent years, the rapid development of deep learning (DL) has injected new blood into object detection. Thus, the deep-learning-based object detection method has become the




A review of Object Detection based on deep Learning

mainstream. Although various detection algorithms have been proposed successively, the unstructured scenarios are still the main challenge to these detection algorithms, such as rigid target deformation, rapid target movement, obstacle occlusion, and drastic light changes in real applications.Analysis on YOLO v3

## III EXISTING METHODS

Image processing is a method to perform some operations on an image, in order to get an enhanced image or to extract some useful information from it. Although the current pedestrian detection algorithm based on deep learning has achieved amazing recognition results, the traditional RGB camera video stream data collection requires certain lighting conditions and cannot adapt to the dark environment at night. The night is the time when security accidents are most likely to occur, which creates loopholes in the security system. The infrared image is used instead of the traditional RGB image to detect pedestrians in thedark at night. Experiments have proved that it has high detection accuracy and has a excellent recognition effect in the case of overlapping pedestrians. Especially for the category imbalance phenomenon of night image frames, the category balance loss item is added to the loss function to

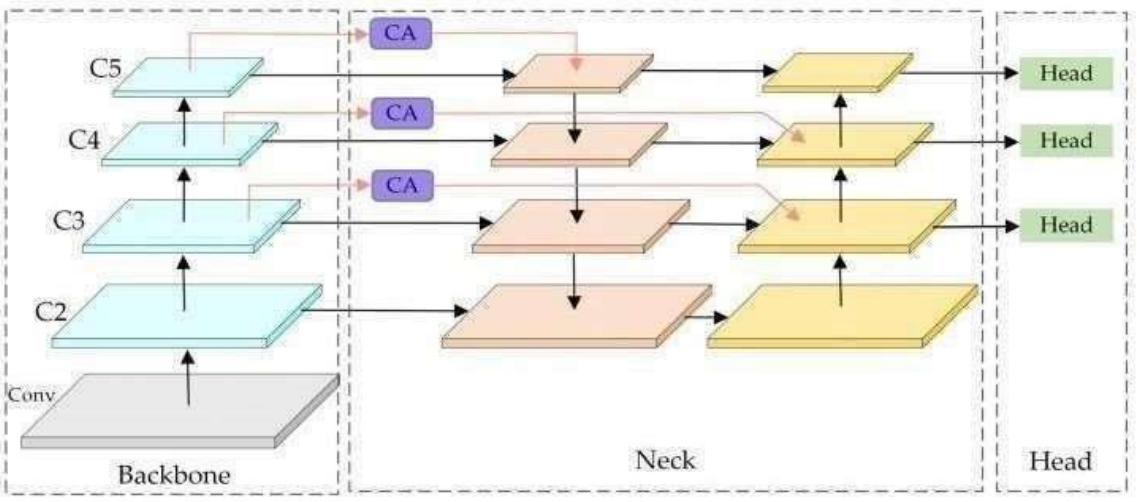
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

optimize. And the effectiveness of the design is proved throughquantitative experiments.

Example of Infrared image

## IV PROPOSED METHODS

The overall structure of the infrared pedestrian detection network proposed is shown in Figure 3.2. After the raw infrared images are input to the network, they first carryout featureextraction through a backbone network composed of AFEM to obtain a pyramid of feature maps in different scales. Then, the feature map is fused with features through our designedfeature fusion network CA-FPN, which can make the information in each layer more balanced and enhance the feature representation at different levels. At last, the result detection at different scales is achieved through three detection heads.




The overall structure of IPD-Net

Backbone

The backbone network is mainly used to extract the feature information of pedestrians in infrared images. As shown in Figure 3.3(a), the backbone network of IPD-Net consists ofa stack of Conv and AFEM. The Conv contains three operations, standard convolution, normalization, and activation functions. The structure of the AFEM module is shown in Figure 3.3(b), where the input feature maps are operated in two separate ways. The featuremap of one path is first passed through convolutions to adjust the number of channels to 0.5times C2 and then through a residual structure consisting of an SSK module. Another way adjusts to 0.5 times the number of channels of C2 by a convolution. Then the two parts areconcatenated to obtain the output with the number of channels of C2, and finally the concatenate features are fused using a Conv convolution block. By using the AFEM convolution module, a richer combination of gradients can be achieved, the learning capability of the CNN is effectively enhanced, and the computational effort is reduced.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

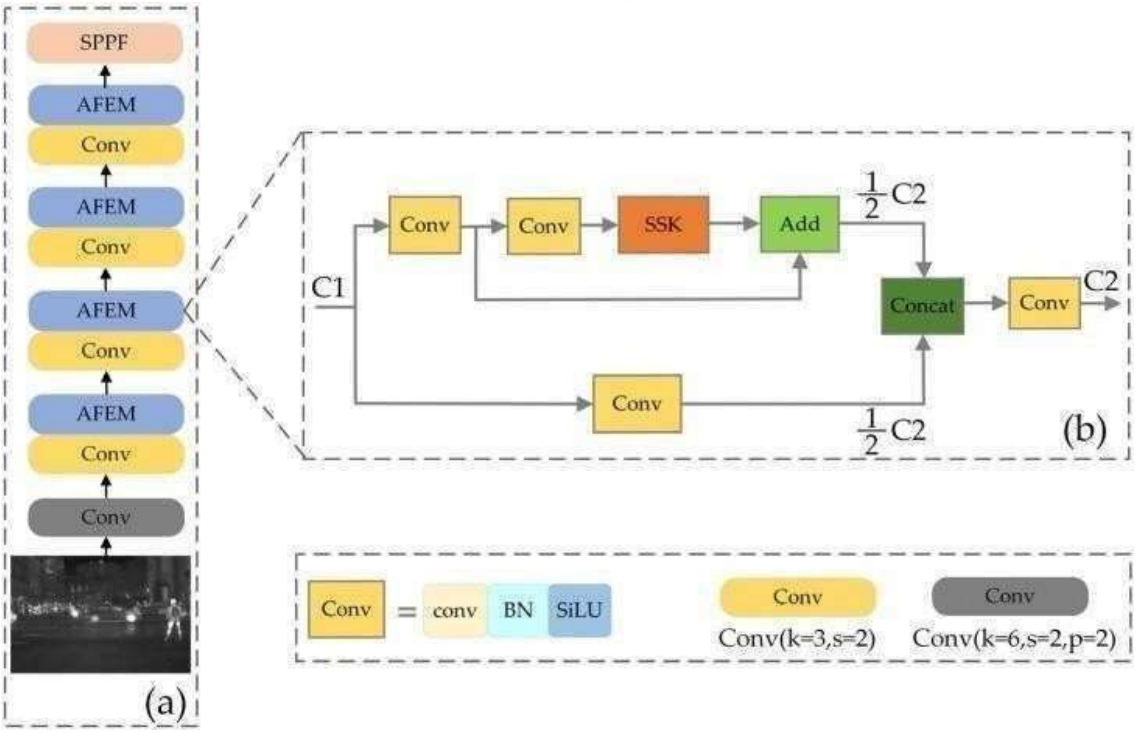



Fig 4.3 The specific structure of backbone and AFEM block: (a) structure of the IPD-Net backbone; (b) structure of the AFEM.

The SSK is a module based on an improved selective kernel (SK). In standard CNN, it is flawed that the receptive fields of each layer of artificial neurons are designed to have the same scale. Each neuron should be able to adaptively adjust its receptive field size according to the input information, so that convolution kernels with different receptive fields can extract richer feature information.

Therefore, an SK model is designed to capture the feature information of objects, which can adaptively adjust the convolution kernel size to 3, 5, and 7. However, introducing larger-scale convolution kernels results in a heavier number of parameters. To address this problem, we designed an SSK block with the structure shown in Figure 3.3. In a convolutional neural network, two cascaded 3 × 3 convolutional kernels have the same receptive fields as a 5 × 5 convolutional kernel and will consume fewer computation resources. Therefore, we can use two cascaded 3 × 3 convolution kernels in series instead of one 5 × 5 convolution kernel to reduce the computational effort while obtaining the samereceptive fields.

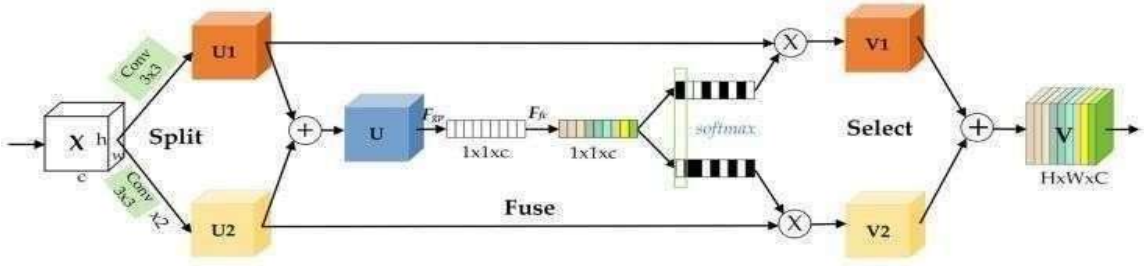
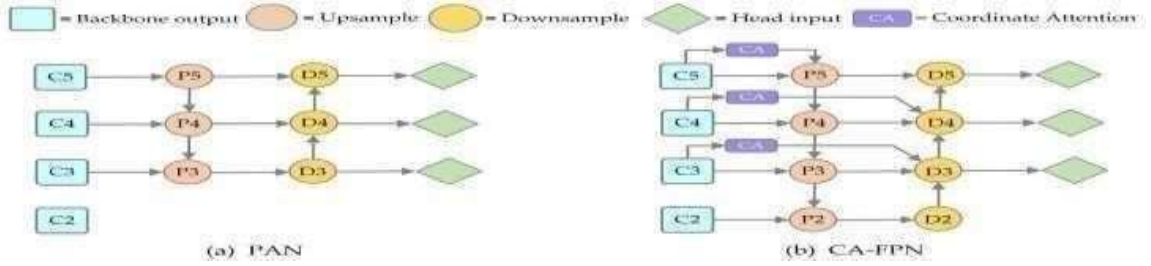



Fig 4.4 The specific structure of the SSK block.

The SSK module is shown in Figure 4.4. Three main operations are carried out in the SSK module: splitting, fusion, and selection. The input feature map X is split into two pathways, passed through a 3 × 3 convolution kernel and two stacked 3 × 3 convolution kernels to obtain feature maps U1 and U2, respectively. Then, U1 and U2 are summed to obtain the fused feature map U. In the fusion stage, U is compressed to 1 × 1 × C by a global averagepooling, and the corresponding weight encoding is extracted by the SoftMax function aftertwo full connection layers. Finally, the obtained weight-encoding values are multiplied with U1 and U2,

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal www.ijarst.in
IJARST
ISSN: 2457-0362

respectively, in the select stage and added together to obtain the feature map V, which contains all weight-encoding information. After splitting, fusion, and selection, the obtained result V incorporates the feature information extracted from the receptive fieldso that the network adaptively adjusts the receptive field using a similar way to channel attention. Compared with the 3 × 3 convolution kernel in the original residual structure, SSK obtains a multi-scale receptive field and has better feature information extraction capability. It can extract pedestrian feature information from infrared images more effectively and receive feature maps with richer feature information. It solves the problemof YOLOv5s having insufficient ability to extract feature information in infrared images.NeckYOLOv5s uses the structure of FPN and PAN for the multi-scale fusion of features, as shownin Figure 4.5(a). The FPN structure is up sampled by a top-down method and then fused with each feature map layer through lateral connections to introduce high-level semantic information from the deep feature map into the shallow network. The PAN is structured sothat the layers of feature maps contain more balanced information and are more conducive to pedestrian detection by the detection head block. However, the PAN feature fusion network still has two problems: (1) Infrared images contain many weak and small objects. As the number of convolution layers increases, some weak and small objects will be lost, so the PAN network does not make full use of the shallow feature maps for fusion. (2) The PAN network needs to dig deeper into the deep feature maps for localization information. Inaddition, the PAN structure fuses the location information and small object in formation in the shallow feature map into the deep feature map by down sampling. Still, a large amountof information is lost in the down sampling process.




Comparison of feature fusion networks in YOLOv5s and IPD-Net.

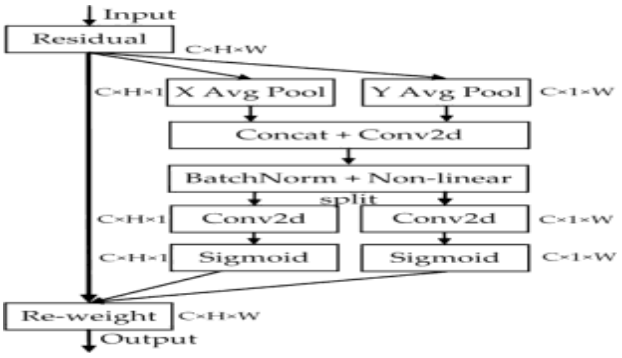
To address these two main problems, inspired by PAN and BiFPN, we designed CA-FPN to obtain better feature fusion. The structure of CA-FPN which enhances feature reuse by adding lateral skip connections, uses the coordinate attention module to further exploit the localization information of targets in the deep feature map, andachieves full fusion of feature information and location information of weak and small targets.

Enhanced Fusion of Shallow Feature Maps:To solve the problem that the deepening of the YOLOv5 network leads to a missing weak pedestrian object in the feature map, we make full use of the information in the shallow feature maps. As shown we added a lateral connection of layer of layer C2 to the bottom of the feature fusion network. The P3 layer feature map is up sampledand fused with the C2 layer feature map to obtain P2, which is then down sampled and further fused to obtain the final prediction. The introduction of C2 and P2 layer feature maps enhances the use of weak objects in the shallow feature maps. It improves the model'sdetection accuracy for weak and small objects in infrared images.

**Feature Fusion with Coordinate Attention Model**:

We use the coordinate attention (CA) module in CA-FPN to enhance the extraction of location information in the deep network feature maps. The structure of the CA module is shown in Figure 4.6. After averaging pooling along the *x*-direction (H) and *y*-direction (W),respectively, the CA block extracts weights for both the *x* and *y* directions, respectively, to obtain global location encoding information. Then, the extracted location encoding information fuses with the original feature map to enhance the location information in the feature map.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

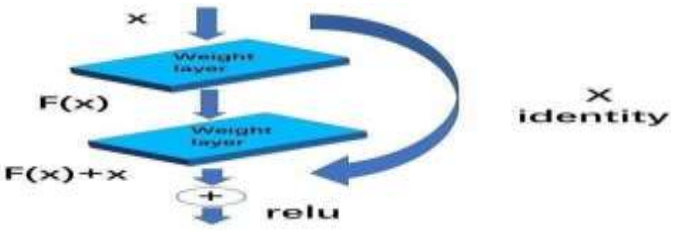



Structure of the Coordinate Attention Module.

To fully dig into the deeper feature map location information, we introduced the CA module into the feature fusion network. C3, C4, and C5 feature maps are enhanced with position information by the CA module and multiplexed using skip connections. C3 and C4 are passed through the CA module and then concatenated to obtain D3 and D4 feature maps, and C5 is passed through CA and concatenated to obtain the P5 feature map. The position information in the C3, C4, and C5 feature maps are enhanced by using the coordinate attention module so that the feature maps contain more information on object positioning.In the CA-FPN, the shallow feature map is first introduced to make full use of the weak and small object information in the shallow feature map and improve the detection accuracyof the model for weak objects. Secondly, the position information in the feature map is encoded by the coordinate attention module to enhance the ability to mine the localization information in the deep feature map, improve the localization capability of the infrared detection model, and enhance the detection accuracy.


DarknetConv2D structure

In addition, Residual blocks in Residual network skip connection in Residual network no longer limit the increase depth of the neural network, so that the accuracy can be improvedand the optimization can be improved more easily 416*416*3 image was input into the Yolov3 model and three different scales of prediction were output via the Darknet53

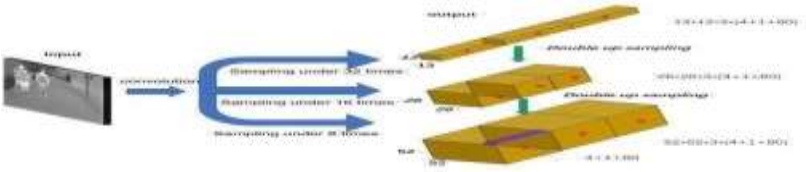



Residual block structure

Each scale for each of the N channels contains bra anchors for each grid and each size. Sowe have 13 times 13 times 3 plus 26 times 26 times 3 plus 52 times 52 times 3. Each prediction corresponds to 85 dimensions, with 4 representing coordinate value, 1 representing confidence score and 80 representing coco categories. The

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

input and output block diagram of the whole Yolov3 is as follows.



In put and output block diagram of the whole Yolov3

A. Use diou loss function

Yolov3 adopts iou loss, which consists of three parts: the intersection ratio between the prediction frame box and the ground-truth is defined as iou, and then the loss function is defined as

However, it can be seen from the definition that IoU loss is always 0 and IoU loss will be constant in

$$L_{IoU} = 1 - \frac{|B \cup B^{gt}|}{|B \cap B^{gt}|} \qquad (1)$$

the case that the prediction box and ground-truth are non-intersecting. If the two boxes do not intersect, IoU loss is always 0. In addition, the two boxes have the same intersection area and there are many ways of intersection, so it is impossible to determine which way the two boxes intersect, which will result in the failure to determine the optimization direction. Although the Subsequent Giou made up for this deficiency. But there are still many shortcomings. RDIoU on the basis of IOU, which is define as follows

$$R_{DIoU} = \frac{\rho^2(b, b^{gt})}{c^2} \qquad (2)$$

**V ARCHITECTURE OF YOLOV4**

YOLOv4 (You Only Look Once version 4) is an object detection model that achieved state- of-the-art performance at the time of its release. The "ECA" in YOLOv4 with ECA stands for "Efficient Channel Attention." ECA is a module that was introduced in YOLOv4 to enhance the model's ability to capture contextual information from feature maps, leading toimproved object detection accuracy.

The ECA module is designed to capture long-range dependencies among different channels of feature maps. It helps the model to effectively model the relationships between channels, which can be crucial for object detection tasks. The ECA module achieves this by performing channel-wise attention, where attention weights are learned for each channel individually. The input layer in YOLOv4 is the first layer of the neural network model, which receives the input image data and processes it to generate the subsequent feature maps. The purpose of the input layer is to transform the raw image data into a format that can be processed by the subsequent layers of the network. In the context of the Darknet framework, the "conv2d" refers to the 2-dimensional convolutional layer. Darknet is a popular open-source neural network framework that is often used for training and deploying deep learning models, including YOLO (You Only Look Once) models. Convolutional and max pooling layers are two essential components of convolutional neural networks (CNNs)used for tasks such as image classification, object detection, and image segmentation. These layers help in extracting and abstracting relevant features from the input data. Concatenation, also known as concatenation or feature concatenation, is the process of combining feature maps from different layers or branches of a neural network.

It involves stacking the feature maps along the channel dimension. Convolution is a fundamental operation in neural networks that applies a set of filters to input feature maps. Itis used to extract local patterns and capture spatial relationships. The "YOLO head" refers tothe final layers of the YOLO (You Only Look Once) object detection architecture. It is responsible for generating bounding box predictions, class probabilities, and confidence scores for detected objects in an input image. YOLO (You Only Look Once) is a popular object detection algorithm that aims to simultaneously predict bounding boxes and class probabilities for multiple

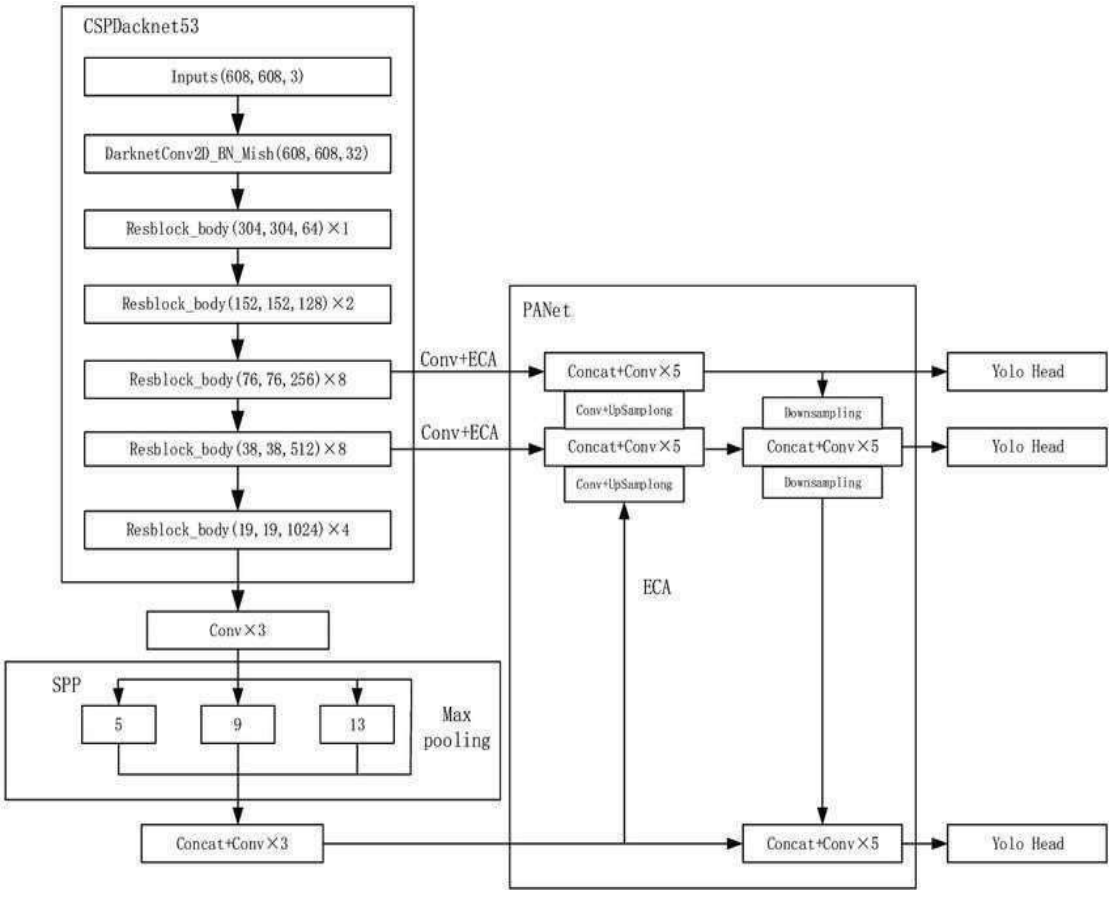
International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
IJARST
ISSN: 2457-0362

objects in an image. YOLO v3 is the third version of the YOLO algorithm, which introduced several improvements over its predecessors. Here's an overview of the architecture. These layers help in extracting and abstracting relevant features from the input data. Concatenation, also known as concatenation or feature concatenation, is the process of combining feature maps from different layers or branches of a neural network. It involves stacking the feature maps along the channel dimension. Convolution is a fundamental operation in neural networks that applies a set of filters to input feature maps. It is used to extract local patterns and capture spatial relationships.

The input layer in YOLOv4 is the first layer of the neural network model, which receives the input image data and processes it to generate the subsequent feature maps. The purpose of the input layer is to transform the raw image data into a format that can be processed by the subsequent layers of the network. In the context of the Darknet framework,the "conv2d" refers to the 2-dimensional convolutional layer.

YOLOv4 utilizes a modified backbone network called CSPDarknet53, which is a more efficient and deeper variant of Darknet architecture.

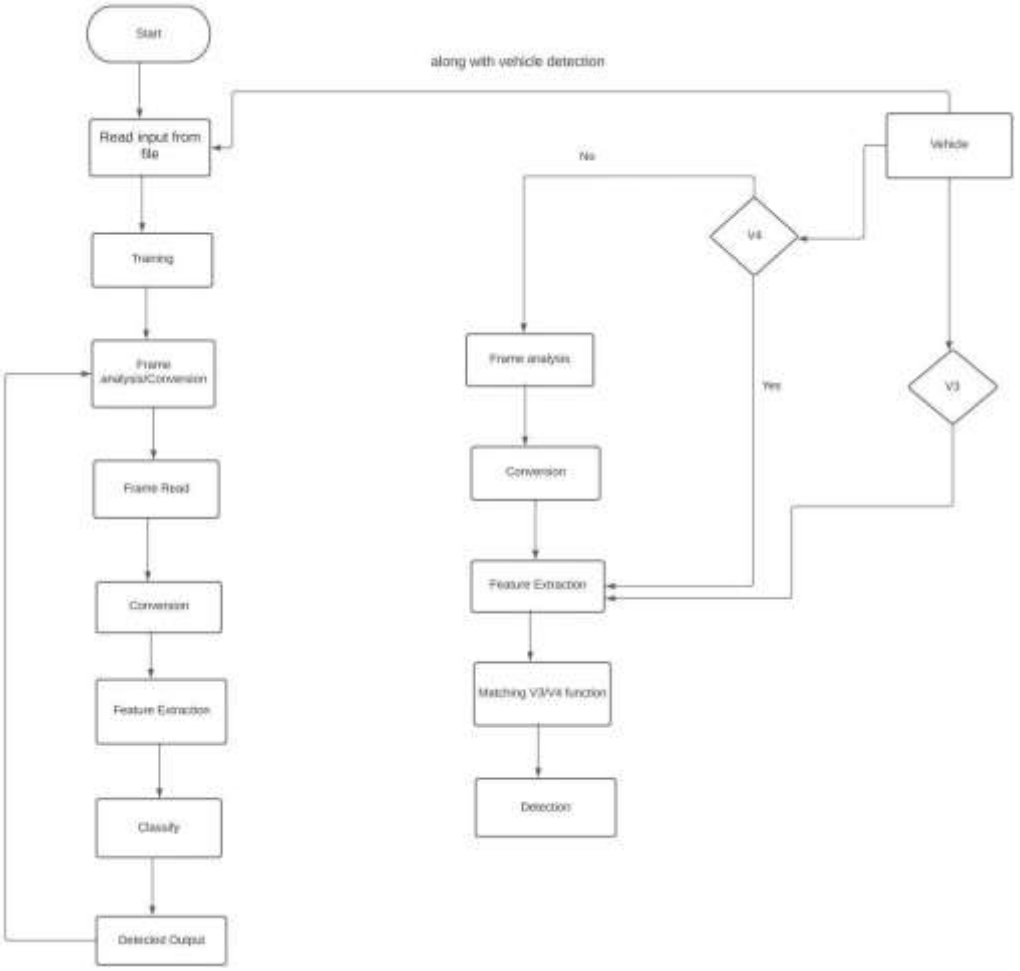
The CSPDarknet53 backbone network helps capture more complex and abstract features from input images. YOLOv4 incorporates the Feature Pyramid Network (FPN), which enables multi-scale feature fusion. FPN combines features from different layers of the backbone network to handle objects of various sizes and scales effectively. Yolov3 employs techniques like multi-scale training, data augmentation, and anchor boxes to improve detection performance. YOLOv4 introduces advanced training techniques such as Cut Mix and Mosaic data augmentation, which enhance the generalization capability of the model. It also uses CIOU (Complete Intersection over Union) loss for better bounding box regression. Yolov3 achieves real-time object detection with decent accuracy. It strikes a balance between speed and accuracy.




Architecture of YOLOv4

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362

**VI FLOW CHART**




Basic Flow chart of Pedestrian detection

**VII RESULTS**

Output of YOLO-V3


Single Pedestrian detection

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362


Multiple Pedestrian Detection

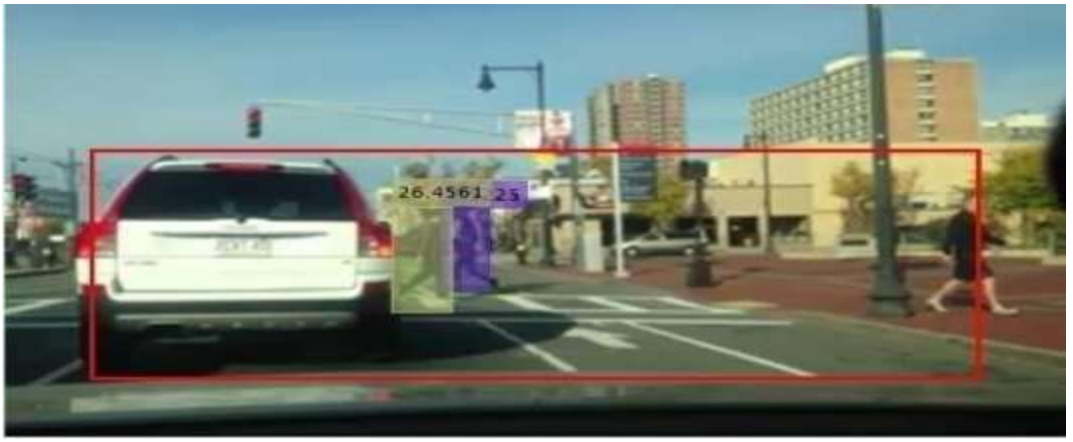

Pedestrian Detection While Overlapping

***OUTPUT OF YOLO-V4***


Single Pedestrian detection

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362


Multiple Pedestrian Detection



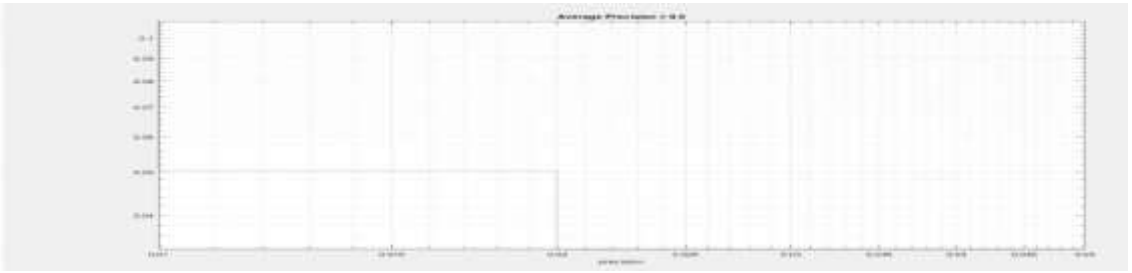

Pedestrian Detection While Overlapping

***COMPARISON BETWEEN YOLOV3 AND YOLOV4***

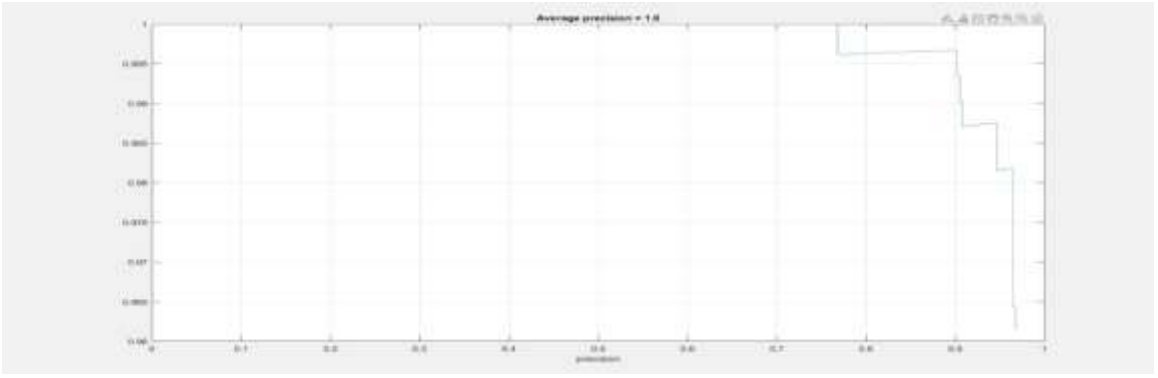
YOLOv3 enables real-time pedestrian detection, making it suitable for applications that require efficient and timely analysis of pedestrian presence in video or image data.YOLOv3 incorporates a Feature Pyramid Network (FPN) that allows multi-scale feature fusion. This helps handle pedestrians of different sizes and scales, ensuring robust detectionperformance across varying scenarios. YOLOv3 utilizes a Darknet-53 backbone network, which captures features from the input image at different scales and resolutions. This enables the model toextract meaningful representations for pedestrian detection. YOLOv3 employs various training techniques such as data augmentation, multi-scale training, and the use of anchor boxes to improve the model's ability to detect pedestrians accurately. YOLOv3 achieves competitive pedestrian detection accuracy, although it may not matchthe precision of more recent models like YOLOv4. However, it still provides reliable detection performance for many practical applications.

YOLOv3 and YOLOv4 does not make much difference but the only deference is accuracy. The clearly do not visible in the output. We are showing in the form of graphs.Inthe below two graphs we can see the slight difference. We can observe the average precisionin the first graph is 0.0 that is by using YOLOv3. And We can observe the average precisionin the second graph is 1.0.Thi is only difference of YOLOv3 and YOLOv4.

International Journal For Advanced Research In Science & Technology
A peer reviewed international journal
www.ijarst.in
ISSN: 2457-0362



output graph of YOLOv3



Output graph of YOLOv4

## VII CONCLUSION

In conclusion, YOLOv3 has proven to be an effective and popular choice for pedestrian detection. It offers real-time object detection capabilities while maintaining a good balancebetween accuracy and inference speed. Here are some key points regarding YOLOv3-based Real-Time Performance: YOLOv3 enables real-time pedestrian detection, making it suitable for applications that require efficient and timely analysis of pedestrian presence in video orimage data. YOLOv3 incorporates a Feature Pyramid Network (FPN)that allows multi- scale feature fusion. This helps handle pedestrians of different sizes and scales, ensuring robust detection performance across varying scenarios.

YOLOv3 utilizes a Darknet-53 backbone network, which captures features from the input image at different scales and resolutions. This enables the model to extract meaningfulrepresentations for pedestrian detection. YOLOv3 employs various training techniques such as data augmentation, multi-scale training, and the use of anchor boxes to improve themodel's ability to detect pedestrians accurately. YOLOv3 achieves competitive pedestrian detection accuracy, although it may not match the precision of more recent models like YOLOv4 or YOLOv5. However, it still provides reliable detection performance for many practical applications. YOLOv3 is well-documented and has a large user base, making it relatively easier to implement and fine-tune for pedestrian detection tasks. It's worth noting that the choice of the specific version of YOLO (e.g., YOLOv3, YOLOv4) may depend onthe specific requirements of your application, available computational resources, and the desired trade-off between accuracy and speed. Consider evaluating and comparing the performance of different models to select the one that best suits your pedestrian detection needs.

## REFERENCES

1. Bochkovskiy, C.-Y. Wang, and H. M. Liao, "YOLOv4: Optimal Speed and Accuracy of Object Detection," rXiv preprint arXiv:2004.10934, 2020.
2. Bochkovskiy, C.-Y. Wang, and H. M. Liao, "YOLOv4: Optimal Speed and Accuracy of ObjectDetection (v3.0)," A. Bochkovskiy, "YOLOv4 in the Darknet Framework," Medium,
3. Bochkovskiy, "YOLOv4: Optimal Speed and Accuracy of Object Detection (Presentation),"2020.

4. P. Maheshwari and A. Vashisht, "YOLOv4 Object Detection in TensorFlow 2.x," 2020.
5. 5Hurney P., Waldron P., Morgan F., Jones E., Glavin M. Review of pedestrian detection techniques in automotive far-infrared video. *IET Intell. Transp. Syst.* 2015;**9**:824–832. doi: 10.1049/iet-its.2014.0236.
6. Hurney P., Waldron P., Morgan F., Jones E., Glavin M. Review of pedestrian detection techniques in automotive far-infrared video. *IET Intell. Transp. Syst.* 2015;**9**:824–832. doi: 10.1049/iet-its.2014.0236.
7. Hwang S., Park J., Kim N., Choi Y., So Kweon I. Multispectral pedestrian detection:Benchmark dataset and baseline; Proceedings of the IEEE Conference on ComputerVisionand Pattern Recognition (CVPR); Boston, MA, USA. 7–12 June 2015;pp. 1037– 1045.
8. Girshick R., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Columbus, OH, USA. 23–28 June 2014; pp. 580–
9. 587.
10. Redmon J., Divvala S., Girshick R., Farhadi A. You only look once: Unified, real-time objectdetection; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Las Vegas, NV, USA. 27–30 June 2016.
11. Redmon J., Farhadi A. YOLO9000: Better, Faster, Stronger; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); Honolulu, HI, USA. 21– 26 July 2017; pp. 6517–6525.

## Authors Biography

**SVMG Phani Kumar C.**
AssistantProfessor(ECE),

He has completed his degree B. Tech (ECE) in Ramappa Engineering College, Warangal and M.Tech (DC) from Kakatiya University, Warangal. He has more than 7 years of teaching experience.

**Daram Sravani**
Btech scholar,Department of Electronics and Communication Engineering,Bhoj Reddy Engineering College for Women. Santhosh nagar Cross Roads, Vinaynagar,Saidabad,Hyderabad,Telangana-500059
mail:sravanidaram07@gmail.com
Btech scholar in Department of Electronics and Communication Engineering, Bhoj Reddy Engineering College for Women. I completed my intermediate in Maram Kethan Reddy junior college, Suryapet. An enthusiastic learner with highly motivational and leadership skills. Always willing to innovate the New things,

**Kummari Sruthi**
Btech scholar, Department of Electronics and communication engineering,BhojReddy Engineering College for Women. Santhosh nagar cross roads, Vinay Nagar,Saidabad, Hyderabad,Telangana-500059
Email:ksruthi0902@gmail.com
Btech final year in Department of Electronics and Communication Engineering, BhojReddy Engineering College for Women. I completed my intermediate from SR junior college for women karimnagar.A highly skilled talented and knowledge candidate with expensive knowledge in the filed of electronics and eager to learn innovative things which can develop the existing technology.