**Title Page:**

# A Real Time Object Detection In Integral Part Of Computer Vision Using Novel Image Classification By Comparing Faster R-CNN Algorithm Over Fast-RCNN Algorithm

# M.Srikar¹, N.Manikandan²

M. Srikar ¹,

Research Scholar,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Tamil Nadu, India, PinCode: 602105

srikarm18[@saveetha.co](mailto:nellurunikhilchowdary17@saveetha.com)m

Dr.N.Manikandan²

Project Guide, Corresponding Author,

Department of Computer Science and Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Tamil Nadu, India, PinCode: 602105

Manikandann.sse@saveetha.com

**Keywords:** Object Detection, Region Proposal Networks, Convolutional Neural Network, Anchors, Novel image classification, Translation-Invariant Anchors, Soft Non Maximum Suppression Algorithm.

**ABSTRACT**

**Aim:** The aim of the research work is to improve the accuracy of object detection using novel image classification using machine learning algorithms. **Materials and Methods:** The categorisingis performed by adopting a sample size of n = 10 in Faster R-CNN(RPN) and sample size n = 10 in Fast R-CNN(Selective Search) algorithms with a sample size = 10. **Results and Discussion:** The analysis of the results shows that the Faster R-CNN using regional proposal networks has a high accuracy of (81.7280) in comparison with the Selective Search based Fast R-CNN (79.6180). There is a statistically significant difference between the study groups with (p<0.05). **Conclusion:** Detection of objects with high accuracy using machine learning algorithms shows that the regional proposal network based Faster R-CNN appears to generate better accuracy than the Selective search(Fast RCNN) algorithm.

**Keywords:** Object Detection, Region Proposal Networks, Convolutional Neural Network, Novel Image Classification, Anchors, Translation-Invariant Anchors, Soft Non Maximum Suppression Algorithm.

### **INTRODUCTION**

Recent advances in object detection are driven by the success of region proposal methods and region-based convolutional neural networks(R-CNN). The purpose of this research is to improve the Accuracy of object detection in computer vision through image classification using machine learning algorithms by using regional proposal networks[(Lv et al. 2018)](https://paperpile.com/c/istsWQ/yH7x). Object Detection is an important task in computer vision, it is one of the most complex tasks for a computer system[(Arulprakash and Aruldoss 2021)](https://paperpile.com/c/istsWQ/m1MJ). Since computer vision is a major thing in the trending world with self-driving cars, traffic detections, tracking objects and many security purposes[(Jiang et al. 2018)](https://paperpile.com/c/istsWQ/UZHJ). It was focused to develop a low cost object detection model and enhance the accuracy and speed in order to detect objects with accuracy and identifying correctly[(Novotny and Matas 2015)](https://paperpile.com/c/istsWQ/YTuT).

Predicting objects using object detection algorithms for over past years and several surveys and detection and recognition have been published in the last years over 17,600 articles from Google Scholar, 7761 journals IEEE Xplore digital library, 975 research articles from ScienceDirect. Among all the articles and journals, the most cited paper is [(Ren et al. 2017)](https://paperpile.com/c/istsWQ/xMnC). The model produced by the [(Ren et al. 2017)](https://paperpile.com/c/istsWQ/xMnC) is very accurate and with improved speed compared to the Selective search model. RPN has a classifier and a regressor. Classifier determines the probability of a proposal having the target object. Regression regresses the coordinates of the proposals[(Zheng, Chen, and Hu 2019)](https://paperpile.com/c/istsWQ/n8XX). The authors have introduced the concept of anchors. Anchor is the central point of the sliding window[(Laban et al. 2019)](https://paperpile.com/c/istsWQ/FZay). So, this algorithm is robust against translations, therefore one of the key properties of this algorithm is translational invariance [(Soni 2019)](https://paperpile.com/c/istsWQ/3GOk).

This method which was used before has less accuracy on detecting objects. It is necessary to detect and determine the object in very milliseconds to prevent problems. For example, self-driving cars need to detect the object within a fraction of seconds and analyse the situation to move forward, otherwise there will be many consequences. In order to sequence the methods and techniques in this research generally fairs better than Selective search(Fast R-CNN) [(Girshick 2015)](https://paperpile.com/c/istsWQ/GU40). It also takes a lot more time to render all the images to train the model compared to Faster R-CNN(SNMS)[(Cai and Vasconcelos 2018)](https://paperpile.com/c/istsWQ/bR8B). The aim of the research work is to improve the accuracy of object detection using novel regional proposal networks over Selective search machine learning algorithms such as Faster region convolutional neural network compared to Fast region convolutional neural network to improve accuracy.

**MATERIALS AND METHODS**

The research work was performed in the Image Processing Lab, Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS. Basically it is considered with two groups of classifiers namely Faster R-CNN(REGIONAL PROPOSAL NETWORK) and Fast R-CNN(Selective search), which is used to detect objects in the image with various image datasets and labels. Group 1 is the Faster R-CNN with the sample size of 10 and Group 2 is the Fast R-CNN with sample size of 10 and it was used to compare for more accuracy score and loss values for choosing the best algorithm to detect objects correctly. Sample size has been calculated and it is identified as standard deviation for Faster R-CNN = 1.68131 and Fast R-CNN = 2.81191.

**FASTER R-CNN**

The Object Detection system, called Faster R-CNN, is compared to two modules. The first module is a deep fully convolutional network that proposes regions(Faster R-CNN), and the second module is the Fast R-CNN detector that uses the proposed regions. The entire system is a single, unified network with ‘attention’ mechanisms, the RPN module tells the Fast R-CNN module where to look. An RPN takes as input an image and returns a set of rectangles, these are the object proposals and each of them has an objectness score. This is done using a fully convolutional network, which shares some convolutional layers with a Fast R-CNN object detection network. The regions are generated by using a small network over the convolutional feature map.This feature map is the output of the last shared convolutional layer. The basic architecture of Faster R-CNN can be shown in Fig-1.

**Pseudocode for Faster R-CNN**

Initialize the parameters

Import caffe

Import tensorflow

Import \* python modules

Conf Threshold → 0.5

Load the models

from weights import weightsPath

weightsPath→.../frozen\_inference\_graph.pb

From config import configPath

configPath → .../rcnn\_inception\_v2\_coco\_Data.pbtxt

Grabbed,frame → vs.read()

Blob → cv2.dnn.blobFromImage(frame)

net.setInput(blob) #passing the blob as an input to the ConvNets

for i in range(numDetections):

box → boxes[0,0,i]

left → int(frameW\*box[3]) #Acquiring bounding boxes

top → int(frameH\*box[4])

right → int(frameW\*box[5])

bottom → int(frameH\*box[6])

cv2.rectangle(frame,(startX,startY),(endX,endY),color,2) #drawing bounding boxes

Resize and reshape the image and form a cluster pixel

Accuracy of the Faster R-CNN regional proposal network.

**FAST R-CNN**

Fast R-CNN is similar to Faster R-CNN but the only difference is that the Fast R-CNN is based on Selective search and Faster R-CNN is based on regional proposal networks. Let’s illustrate the Fast R-CNN architecture, a Fast R-CNN network takes as input an entire image and a collection of object proposals. The network first processes the total image with many convolutional and max pooling layers to supply a conv feature map. Then, for every object proposal a neighbourhood of interest (Roi) pooling layer extracts a fixed-length feature vector from the feature map. Each feature vector is fed into a sequence of totally connected (fc) layers that finally branch into 2 relation output layer, (i) That produces softmax chance estimates over K object categories and a catch-all “background” category and another layer that outputs four real-valued numbers for every K object categories. Every set of four values encodes refined bounding-box positions for one in all the K categories.

**Pseudocode of Fast R-CNN**

Import parameters

Import selective search layer

Import tensorflow

Import \* python

Conf Threshold → 0.5

maskThreshold → 0.3

Load the models

from weights import weightsPath

weightsPath→.../frozen\_inference\_graph.pb

From config import configPath

configPath → .../rcnn\_inception\_v2\_coco\_Data.pbtxt

Grabbed,frame → vs.read()

Blob → cv2.dnn.blobFromImage(frame)

net.setInput(blob) #passing the blob as an input to the ConvNets

for i in range(numDetections):

box → boxes[0,0,i]

mask → masks[i]

left → int(frameW\*box[3]) #Acquiring bounding boxes

top → int(frameH\*box[4])

right → int(frameW\*box[5])

bottom → int(frameH\*box[6])

cv2.rectangle(frame,(startX,startY),(endX,endY),color,2) #drawing bounding boxes

Resize and reshape the image and form a cluster pixel

Accuracy of the Fast R-CNN regional proposal network.

**STATISTICAL ANALYSIS**

The analysis was done using IBM SPSS version 21. It is a statistical software tool used for data analysis. For both proposed and existing algorithms 10 iterations were done with a maximum of 10 samples and for each iteration the predicted accuracy was noted for analysing accuracy. The value obtained from the iterations of the Independent Sample T-test was performed. The dependent data sets are ImageNet, Microsoft COCO test-dev, PASCAL VOC 2007,PASCAL VOC 2012. The independent values are AlexNet, VGGNet, RetinaNet, ResNeXt-101-FPN. The fragmented analysis has been done with independent and dependent variables to find the objects with more accuracy and speed.

**RESULTS**

The Datasets used to train models are the COCO dataset, PASCAL VOC 2007, 2012 datasets. The model has trained through more than 22000 images on specific labels. Group statistics of Faster R-CNN by Fast R-CNN by grouping with iterations sample size of 10, mean = 81.7280 Standard Deviation = 1.68131 , Standard Error Mean = .53168. Descriptive Independent Sample Test of Accuracy and Loss is applied for the dataset in SPSS. Here it specifies equal variances with and without assuming a T-Test Score of two groups with each sample size of 10 in Table 2. The Significant value= 0.600, Mean Difference= 2.11000 and confidence interval = (-.06662 - 4.28662) of Faster R-CNN based Object detection and Fast R-CNN based Object detection is tabulated in Table 3, which shows there is a significant difference between the two groups since P<0.05 with an independent sample T-Test. Images, labels and tested image datasets independent variables. The dependent variables in object detection are detected with the help of the independent variables. The statistical analysis of two independent groups shows that the Faster R-CNN have higher accuracy mean (81.7280) and Less Loss mean 1.68131 compared to selective search based Fast R-CNN with accuracy (79.6180) and Less Loss mean .64955 in table-1.

**DISCUSSION**

In this research work detecting objects in real time based on images, subsequently termed Object Detection using computer vision by machine learning algorithms is very important in many industries in order to process different scenarios [(Pathak, Pandey, and Rautaray 2018)](https://paperpile.com/c/istsWQ/tz2b). The most important features of detecting objects using Faster R-CNN [(Ren et al. 2017)](https://paperpile.com/c/istsWQ/xMnC) is pragmatically proven to be highly effective than Fast R-CNN. The core argument is that to prove that detection of objects in low light images may be a better method than other methods of object detection. In many of the recent findings, it has been observed that the Region proposal network is the most focused and better method of detecting objects with more accuracy than Fast rcnn[(Girshick 2015)](https://paperpile.com/c/istsWQ/GU40).

Object detection aims to acknowledge and localize each object instance with a bounding box. As a classical problem within the field of computer vision, it's widely utilized in autonomous vehicles [(Akhtar and Mian 2018)](https://paperpile.com/c/istsWQ/xBpE)and assistive robots [(Subudhi 2009)](https://paperpile.com/c/istsWQ/dBHI). The normal object detection methods are generally supported scale invariant feature transform (SIFT)[(“Real-Time Object Detection and Localization with SIFT-Based Clustering” 2012)](https://paperpile.com/c/istsWQ/Aubq). and histogram of oriented gradient (HOG)[(Patel et al. 2020)](https://paperpile.com/c/istsWQ/Pu2B). These methods extract the thin features and breeze through the image to seek out regions with a class-specific maximum response. However, these methods perform well only on constrained object categories and are sensitive to noise. These problems limit the appliance range of the normal object detection methods.

The evidence from recent success of cascade for object detection (Cai and Vasconcelos 2018; Cheng et al. 2018a, b) and instance segmentation on COCO [(Lin et al. 2014)](https://paperpile.com/c/istsWQ/FTSb) and other challenges has shown that multistage object detection could be a future framework for a speed-accuracy trade-off. The bounding boxes are most widely used in the evaluation of generic object detection algorithms[(Bauckhage and Tsotsos 2005)](https://paperpile.com/c/istsWQ/74kl), therefore this is the approach we adopt in this survey. However, as the research community moves towards the deeper scene understanding from image level classification to single object localization, to generic object detection, and to pixel wise object segmentation, it is expected that future challenges will be at the pixel level [(Lin et al. 2014)](https://paperpile.com/c/istsWQ/FTSb).

**CONCLUSION**

A real time object detection in integral part of computer vision using novel image classification by comparing faster r-cnn algorithm over fast-rcnn algorithm. The current study focused on machine learning algorithms, Region proposal network(Faster R-CNN) over Selective search (Fast R-CNN) for higher classification of object detection. It can be slightly improved based on high trained datasets in future. The outcome of the Faster R-CNN based on regional proposal networks showed higher accuracy (81.7280) than the Selective search based Fast R-CNN (79.6180).

**DECLARATION**

**Conflict of Interests**

No conflict of interest

**Authors Contribution**

Author MS was involved in data collection, data analysis, manuscript writing. Author NM was involved in the Action process, Data verification and validation, and Critical review of manuscript.

**Acknowledgments**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences ( Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this research work successfully.

**Funding:** We thank the following organization for providing financial support that enabled us to complete the study.

1. SNEW.AI Technologies, Chennai.
2. Saveetha University
3. Saveetha Institute of Medical and Technical Sciences
4. Saveetha School of Engineering

**REFERENCES**

[Akhtar, Naveed, and Ajmal Mian. 2018. “Threat of Adversarial Attacks on Deep Learning in Computer Vision: A Survey.” *IEEE Access*. https://doi.org/](http://paperpile.com/b/istsWQ/xBpE)[10.1109/access.2018.2807385](http://dx.doi.org/10.1109/access.2018.2807385)[.](http://paperpile.com/b/istsWQ/xBpE)

[Arulprakash, Enoch, and Martin Aruldoss. 2021. “A Study on Generic Object Detection with Emphasis on Future Research Directions.” *Journal of King Saud University - Computer and Information Sciences*. https://doi.org/](http://paperpile.com/b/istsWQ/m1MJ)[10.1016/j.jksuci.2021.08.001](http://dx.doi.org/10.1016/j.jksuci.2021.08.001)[.](http://paperpile.com/b/istsWQ/m1MJ)

[Bauckhage, C., and J. K. Tsotsos. 2005. “Bounding Box Splitting for Robust Shape Classification.” *IEEE International Conference on Image Processing 2005*. https://doi.org/](http://paperpile.com/b/istsWQ/74kl)[10.1109/icip.2005.1530096](http://dx.doi.org/10.1109/icip.2005.1530096)[.](http://paperpile.com/b/istsWQ/74kl)

[Cai, Zhaowei, and Nuno Vasconcelos. 2018. “Cascade R-CNN: Delving Into High Quality Object Detection.” *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. https://doi.org/](http://paperpile.com/b/istsWQ/bR8B)[10.1109/cvpr.2018.00644](http://dx.doi.org/10.1109/cvpr.2018.00644)[.](http://paperpile.com/b/istsWQ/bR8B)

[Girshick, Ross. 2015. “Fast R-CNN.” *2015 IEEE International Conference on Computer Vision (ICCV)*. https://doi.org/](http://paperpile.com/b/istsWQ/GU40)[10.1109/iccv.2015.169](http://dx.doi.org/10.1109/iccv.2015.169)[.](http://paperpile.com/b/istsWQ/GU40)

[Jiang, Xiaoyue, Abdenour Hadid, Yanwei Pang, Eric Granger, and Xiaoyi Feng. 2018. *Deep Learning in Object Detection and Recognition*. Springer.](http://paperpile.com/b/istsWQ/UZHJ)

[Laban, Noureldin, Bassam Abdellatif, Hala M. Ebeid, Howida A. Shedeed, and Mohamed F. Tolba. 2019. “Convolutional Neural Network with Dilated Anchors for Object Detection in Very High Resolution Satellite Images.” *2019 14th International Conference on Computer Engineering and Systems (ICCES)*. https://doi.org/](http://paperpile.com/b/istsWQ/FZay)[10.1109/icces48960.2019.9068145](http://dx.doi.org/10.1109/icces48960.2019.9068145)[.](http://paperpile.com/b/istsWQ/FZay)

[Lin, Tsung-Yi, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. “Microsoft COCO: Common Objects in Context.” *Computer Vision – ECCV 2014*. https://doi.org/](http://paperpile.com/b/istsWQ/FTSb)[10.1007/978-3-319-10602-1\_48](http://dx.doi.org/10.1007/978-3-319-10602-1_48)[.](http://paperpile.com/b/istsWQ/FTSb)

[Lv, Xiaogang, Xiaotao Zhang, Yinghua Jiang, and Jianxin Zhang. 2018. “Pedestrian Detection Using Regional Proposal Network with Feature Fusion.” *2018 Eighth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. https://doi.org/](http://paperpile.com/b/istsWQ/yH7x)[10.1109/ipta.2018.8608159](http://dx.doi.org/10.1109/ipta.2018.8608159)[.](http://paperpile.com/b/istsWQ/yH7x)

[Novotny, David, and Jiri Matas. 2015. “Cascaded Sparse Spatial Bins for Efficient and Effective Generic Object Detection.” *2015 IEEE International Conference on Computer Vision (ICCV)*. https://doi.org/](http://paperpile.com/b/istsWQ/YTuT)[10.1109/iccv.2015.137](http://dx.doi.org/10.1109/iccv.2015.137)[.](http://paperpile.com/b/istsWQ/YTuT)

[Patel, Chirag I., Dileep Labana, Sharnil Pandya, Kirit Modi, Hemant Ghayvat, and Muhammad Awais. 2020. “Histogram of Oriented Gradient-Based Fusion of Features for Human Action Recognition in Action Video Sequences.” *Sensors*  20 (24). https://doi.org/](http://paperpile.com/b/istsWQ/Pu2B)[10.3390/s20247299](http://dx.doi.org/10.3390/s20247299)[.](http://paperpile.com/b/istsWQ/Pu2B)

[Pathak, Ajeet Ram, Manjusha Pandey, and Siddharth Rautaray. 2018. “Application of Deep Learning for Object Detection.” *Procedia Computer Science*. https://doi.org/](http://paperpile.com/b/istsWQ/tz2b)[10.1016/j.procs.2018.05.144](http://dx.doi.org/10.1016/j.procs.2018.05.144)[.](http://paperpile.com/b/istsWQ/tz2b)

[“Real-Time Object Detection and Localization with SIFT-Based Clustering.” 2012. *Image and Vision Computing* 30 (8): 573–87.](http://paperpile.com/b/istsWQ/Aubq)

[Ren, Shaoqing, Kaiming He, Ross Girshick, and Jian Sun. 2017. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 39 (6): 1137–49.](http://paperpile.com/b/istsWQ/xMnC)

[Soni, Divyanshu. 2019. “Translation Invariance in Convolutional Neural Networks.” Medium. November 13, 2019.](http://paperpile.com/b/istsWQ/3GOk) <https://divsoni2012.medium.com/translation-invariance-in-convolutional-neural-networks-61d9b6fa03df>[.](http://paperpile.com/b/istsWQ/3GOk)

[Subudhi, Bidyadhar. 2009. *Computational Intelligence, Control and Computer Vision in Robotics and Automation*.](http://paperpile.com/b/istsWQ/dBHI)

[Zheng, Jingye, Dihu Chen, and Haifeng Hu. 2019. “Multi-Scale Proposal Regression Network for Temporal Action Proposal Generation.” *IEEE Access*. https://doi.org/](http://paperpile.com/b/istsWQ/n8XX)[10.1109/access.2019.2933360](http://dx.doi.org/10.1109/access.2019.2933360)[.](http://paperpile.com/b/istsWQ/n8XX)

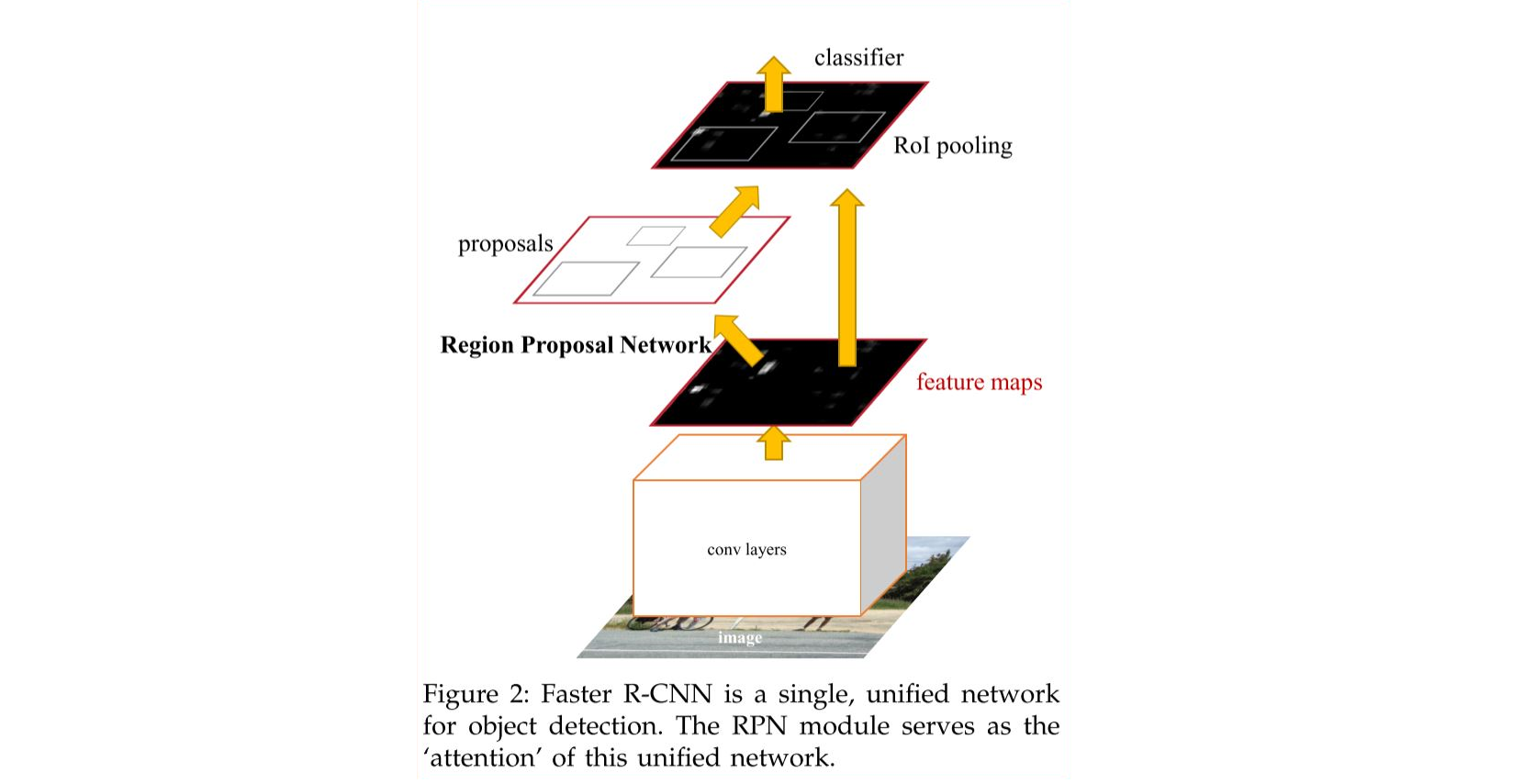
**TABLES AND FIGURES**

**Table-1.** Group Statistics of regional proposal network based Faster R-CNN by grouping the iterations with sample size 6, Mean = 81.7280, Standard Deviation = 1.68131. Descriptive Independent Sample Test of Accuracy and Loss is applied for the dataset in SPSS. Here it specifies Equal variances with and without assuming a T-Test Score of two groups with each sample size of 10.

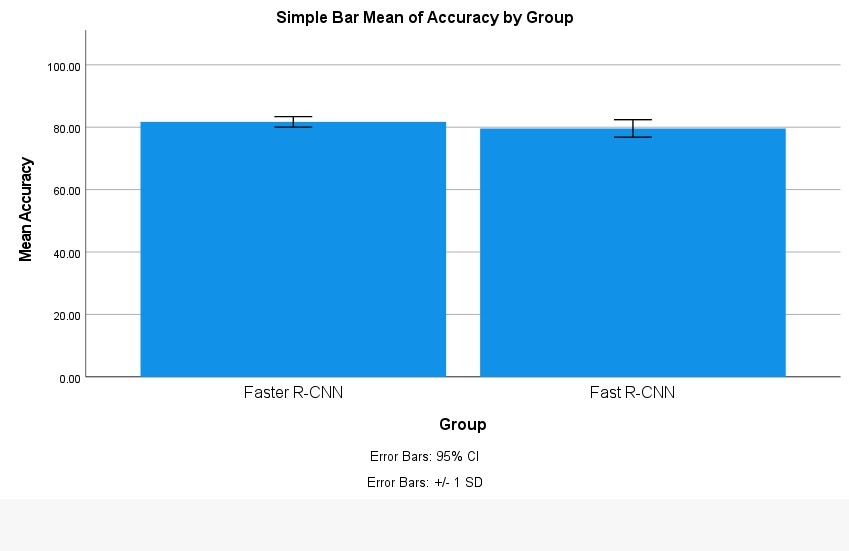
|  | **Group** | **N** | **Mean** | **Std. Deviation** | **Std.Error Mean** |
| --- | --- | --- | --- | --- | --- |
| **Accuracy** | Faster R-CNN | 10 | 81.7280 | 1.68131 | .53168 |
|  | Fast R-CNN | 10 | 79.6180 | 2.81191 | .88920 |
| **Loss** | Faster R-CNN | 10 | 1.5300 | .64955 | .20540 |
|  | Fast R-CNN | 10 | 2.0960 | .37787 | .11949 |

**Table-2.** Independent Sample Test of Accuracy and Loss calculate P-value = 0.001 and Significant value = 0.600, Mean Difference = 2.1100 and confidence interval = (-0.6662 - 4.28662). Faster R-CNN and Fast R-CNN are significantly different from each other.

|  |  |  | |  |  | **Significance** | |  |  | **95% confidence interval of the difference** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | **F** | **Sig.** | **t** | **df** | **One-sided p** | **Two-Sided p** | **Mean Difference** | **Std. Error Difference** | **Lower** | **Upper** |
| **accuracy** | **Equal variances assumed** | .285 | .600 | 2.037 | 18 | <.028 | <.057 | 2.1100 | 1.03603 | -.06662 | 4.28662 |
| **Equal variances not assumed** |  |  | 2.037 | 14.706 | <.030 | <.060 | 2.11000 | 1.03603 | -.10210 | 4.32210 |
| **Loss** | **Equal variances assumed** | 2.877 | .107 | -2.382 | 18 | <.014 | <.028 | -.56600 | .23763 | -1.06525 | -.06675 |
| **Equal variances not assumed** |  |  | -2.382 | 14.466 | <.016 | <.031 | -.56600 | .23763 | -1.07414 | -.05786 |

****

**Fig. 1.** The Faster R-CNN is a simple, unified network for object detection. The RPN module serves as the ‘attention’ of this unified network.

****

**Fig. 2.** Comparison of regional proposal network based Faster R-CNN in terms of mean accuracy. It explores that the mean accuracy is slightly better than Fast R-CNN with Selective search and the standard deviation is moderately improved compared to logistic regression. Graphical representation of the bar graph is plotted using group id as X-axis Faster RCNN vs Fast R-CNN, Y-axis displaying the error bars with mean accuracy of detection +/-1 SD.