**Title Page:**

An Enhanced Object Detection In Integral Part Of Computer Vision Using Object Localization By Comparing Spatial Pyramid Pooling Net Algorithm Over R-Cnn 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.com

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, Deep Convolutional Neural Network, Novel Pyramid Pooling Layer, Object Localization, Deep Network, Spatial Pyramid Match.

**ABSTRACT**

**Aim:** The aim of the research work is to improve the accuracy of object detection using inventive Deep Convolutional Neural Networks with machine learning algorithms. **Materials and Methods:** The categorisingis performed by adopting a sample size of n = 10 in Spatial Pyramid Pooling net in R-CNN and sample size n = 10 in R-CNN algorithms with a sample size = 10. **Results and Discussion:** The analysis of the results shows that the R-CNN using spatial pyramid pooling net layer has a high accuracy of (85.4350 %) in comparison with the Region based convolutional neural network (78.369 %). There is a statistically significant difference between the study groups with (p<0.001). **Conclusion:** Detection of objects with high accuracy using machine learning shows that the spatial pyramid pooling net layer based R-CNN appears to generate better accuracy than the region based convolutional neural network algorithm.

**Keywords:** Object Detection, Deep Convolutional Neural Network, Novel Pyramid Pooling Layer, Object Localization.

**INTRODUCTION:**

We are witnessing a rapid, progressive extrade in our imaginative and prescient community, in particular because of deep convolutional neural networks (CNNs) and the availability of huge scale education data. 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 networksSpatial Pyramid Pooling(SPP) is a pooling layer that removes the fixed size constraint of the network, i.e. a convolutional neural network does not require a fixed-size input image[(Li et al. 2020)](https://paperpile.com/c/DbTp46/h9Wo). Specifically, we add an SPP layer on top of the last convolutional layer[(Guo et al. 2018)](https://paperpile.com/c/DbTp46/Ox9F). Deep-networks' primarily based totally approach have lately been drastically enhancing upon the nation of the artwork in image classification, item detection, many different popularity tasks, or even non-popularity tasks [(He et al. 2015)](https://paperpile.com/c/DbTp46/fXnC). Specific object detection applications include pedestrian detection, video surveillance, text detection, pose direction, and many more. Through the proposed model the performance of object detectors and trackers has greatly improved, achieving significant standards in object localization[(Zhang et al. 2021)](https://paperpile.com/c/DbTp46/ZboW).

In detecting and classification of objects in real time using spatial pyramidmatch pooling layer by comparing over 17,300 journals from google scholar, 2,334 articles from science direct, 1,206 articles, 1,205 chapters, 794 conference papers from springer link, 135 journals from IEEE xplore digital library. Spatial pyramid pooling layer improved object detection accuracy in region based convolutional neural networks. Among all the articles and journals, the most cited paper is [(He et al. 2015)](https://paperpile.com/c/DbTp46/fXnC) is a most useful and improved . In this work, a novel idea of adding sppnet layer in the convolutional neural network was introduced by [(He et al. 2015)](https://paperpile.com/c/DbTp46/fXnC). By training the spatial pyramid match pooling model on PASCAL VOC 2007 there is an increase in speed up to 64 times faster compared to R-CNN and improved accuracy. SPPs are based on original work more commonly referred to as spatial pyramid matching(SPM)(BoW)[(Ismail et al. 2018)](https://paperpile.com/c/DbTp46/2pZ8). The main working of the SPP layer in the convolutional neural network is the input image with any arbitrary input. The image is put through the convolutional feature pooling layer[(Gao, Shang, and Wu 2021)](https://paperpile.com/c/DbTp46/m8CL). The features pulled by the convolutional layer are passed to the SPP layer, SPP generates a fixed-length output regardless of the size of the input[(Sun, Ni, and Zhao 2022)](https://paperpile.com/c/DbTp46/KkVP).

This technique which was utilised before has less precision on recognizing objects, minuscule items. It is important to distinguish and decide the object in very milliseconds to forestall issues. For instance, self-driving vehicles need to distinguish the object inside a small amount of seconds and dissect the circumstance to push ahead, in any case there will be numerous outcomes. To arrange the strategies and procedures in this exploration for the most part fairs better compared to particular search (R-CNN). It likewise requires some investment to deliver every one of the pictures to prepare the model contrasted with Quicker R-CNN (SPPnet layer). The point of the research work is to improve the accuracy of object detection using novel machine learning algorithms such as the Spatial pyramid pooling net based deep convolutional neural network compared to 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 Spatial Pyramid Pooling net and Region based convolutional neural network, which is used to detect objects in the image with various image datasets and labels. Group 1 is the SPPnet based R-CNN with the sample size of 10 and Group 2 is the 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 and fastly. Sample size has been calculated and it is identified as standard deviation for SPPnet layer R-CNN = 1.39991 and R-CNN = .64140.

**SPATIAL PYRAMID POOLING NET LAYER ALGORITHM**

Spatial Pyramid Pooling (SPP) adds a new layer between the convolutional layers and the fully connected layers. Its job is to map any size input down to fixed size output. The ideal of spatial pyramid pooling, also known as spatial pyramid match or just ‘multilevel pooling’ pre-existed in computer vision, but had not been applied in the context of CNNs.

SPP works by isolating the element maps yield by the last convolutional layer into various spatial canisters with sizes relative to the picture size, so the quantity of containers is fixed paying little heed to the picture size. Receptacles are caught at various degrees of granularity - for instance, one layer of 16 containers isolating the picture into a 4×4 framework, one more layer of 4 canisters partitioning the picture into a 2×2 lattice, and a last layer including the entire picture. In each spatial container, the reactions of each channel are essentially pooled utilizing max pooling.

The SPP method can likewise be utilized for identification. The model used before was the R-CNN method that runs feature extraction on each of 2000 windows extracted from an input image. This is costly and slow. An SPP-net utilized for object detection separates highlight maps just a single time (conceivable at numerous scales). Then, at that point, simply the spatial pyramid pooling piece is run once for every competitor window. This ends up giving equivalent outcomes, yet with running occasions 38x-102x quicker relying upon the quantity of scales.

**Pseudocode for SPPnet**

import torch

import torch.nn as nn

import numpy as np

from spp\_layer import spatial\_pyramid\_pool

import functools

Def \_\_init\_\_(self, opt, input\_nc, ndf=64, gpu\_ids=[]):

super(SPP\_NET, self).\_\_init\_\_()

Self.gpu\_ids → gpu\_ids

Self.ouput\_num → [4,2,1]

Passing input through cnn model

Self.conv1 → nn.conv2d(input\_nc, ndf, 4, 2, 1, bias = False)

self.BN1 → nn.BatchNorm2d(ndf \* 2)

Self.fc1 → nn.Linear(10752, 4096)

Self.fc2 → nn.Linear(4096, 1000)

Spp → spatial\_pyramid\_pool(x,1, [int(x.size(2)), int(x.size(3))], self.output\_num)

Fc1 → self.fc1(spp)

Fc2 → self.fc2(fc1)

Output → s(fc2)

Return output

**R-CNN ALGORITHM**

The R-CNN algorithm has been established as the de facto algorithm for deep learning-based object detection. It significantly outperforms conventional approaches in the PASCAL VOC by capitalizing the following two insights: First, it uses object proposals rather than sliding windows. Before the R-CNN, most object detectors such as DPM adopted a image pyramid plus sliding window approach to generate potential object locations and handle various scales. In the R-CNN pipeline, a fixed number of boxes are proposed per image which most likely contain the target objects. The problem of various scales is also handled automatically by the proposal generation. Fewer but better proposals contribute a lot to the good performance of the R-CNN. Second, it leverages ImageNet pre-trained deep neural network models, which is then fine tuned using the PASCAL VOC. The pre-training process, the R-CNN works poorly. Given the region proposals, training an R-CNN object detector generally composes two major steps: supervised pre-training and domain-specific finetuning. [(Cao et al. 2019)](https://paperpile.com/c/DbTp46/g7dY)

**Pseudocode for R-CNN Algorithm**

import numpy as np

import skimage.io

import matplotlib

import matplotlib.pyplot as plt

Import Mask RCNN

sys.path.append(ROOT\_DIR) # To find local version of the library

from mrcnn import utils

import mrcnn.model as modellib

from mrcnn import visualize

%matplotlib inline

MODEL\_DIR → os.path.join(ROOT\_DIR, "logs")

if not os.path.exists(COCO\_MODEL\_PATH):

utils.download\_trained\_weights(COCO\_MODEL\_PATH)

model→modellib.MaskRCNN(mode="inference",model\_dir=MODEL\_DIR, config=config)

file\_names = next(os.walk(IMAGE\_DIR))[2]

Taking image as a input

image→skimage.io.imread(os.path.join(IMAGE\_DIR,random.choice(file\_names)))

# Run detection

results → model.detect([image], verbose=1)

# Visualize results

r → results[0]

visualize.display\_instances(image, r['rois'], r['masks'], r['class\_ids'],

class\_names, r['scores'])

**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 SPPnet based convolutional neural network by R-CNN by grouping with iterations sample size of 10, mean = 85.4350 Standard Deviation = 1.39991 , Standard Error Mean = .44269. 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= 0.046, Mean Difference= 5.06600 and confidence interval = (4.04297 - 6.08903) of SPPnet(CNN) based Object detection and R-CNN based Object detection is tabulated in Table 3, which shows there is a significant difference between the two groups since P<0.001 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 SPPnet based convolutional neural network have higher accuracy mean (85.4350 %) and Less Loss mean 1.4170 % compared to R-CNN with accuracy (78.369 %) and Less Loss mean 1.7710 % in table-1.

**DISCUSSION**

In this paper we discussed 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/DbTp46/8hXAe). The most important features of detecting objects using “spatial pyramid pooling” based convolutional neural networks [(Ren et al. 2017)](https://paperpile.com/c/DbTp46/b1r4F) is pragmatically proven to be highly effective than R-CNN. The core argument is that to prove that detection of objects like smaller images may be a better method than other methods of object detection. In many of the recent findings, it has been observed that the spatial pyramid pooling layer based deep neural network is the most focused and better method which of detecting objects with more accuracy than rcnn[(Girshick 2015)](https://paperpile.com/c/DbTp46/1RMoE).

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/DbTp46/vCTqR)and assistive robots [(Subudhi 2009)](https://paperpile.com/c/DbTp46/QBdmJ). 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/DbTp46/Egxha). and histogram of oriented gradient (HOG)[(Patel et al. 2020)](https://paperpile.com/c/DbTp46/e9u9r). 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/DbTp46/05fDf) 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/DbTp46/Ixhc0), 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/DbTp46/05fDf).

**CONCLUSION**

Enhanced object detection in integral part of computer vision using novel image localization by comparing SPPnet based cnn algorithm R-CNN algorithm. The current study focused on machine learning algorithms and some deep learning methods, Spatial Pyramid Pooling (CNN) over R-CNN for higher classification of object detection. The accuracy and speed can be slightly improved based on high trained datasets in future. The outcome of the SPPnet based convolutional neural network showed higher accuracy (85.4350 %) than the R-CNN (78.369 %).

**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 organisation 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/DbTp46/vCTqR)[10.1109/access.2018.2807385](http://dx.doi.org/10.1109/access.2018.2807385)[.](http://paperpile.com/b/DbTp46/vCTqR)

[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/DbTp46/Ixhc0)[10.1109/icip.2005.1530096](http://dx.doi.org/10.1109/icip.2005.1530096)[.](http://paperpile.com/b/DbTp46/Ixhc0)

[Cao, Changqing, Bo Wang, Wenrui Zhang, Xiaodong Zeng, Xu Yan, Zhejun Feng, Yutao Liu, and Zengyan Wu. 2019. “An Improved Faster R-CNN for Small Object Detection.” *IEEE Access*. https://doi.org/](http://paperpile.com/b/DbTp46/g7dY)[10.1109/access.2019.2932731](http://dx.doi.org/10.1109/access.2019.2932731)[.](http://paperpile.com/b/DbTp46/g7dY)

[Gao Qunxia, Shang Lijuan, and Wu Kai. 2021. “[Sleep apnea automatic detection method based on convolutional neural network].” *Sheng wu yi xue gong cheng xue za zhi = Journal of biomedical engineering = Shengwu yixue gongchengxue zazhi* 38 (4): 678–85.](http://paperpile.com/b/DbTp46/m8CL)

[Girshick, Ross. 2015. “Fast R-CNN.” *2015 IEEE International Conference on Computer Vision (ICCV)*. https://doi.org/](http://paperpile.com/b/DbTp46/1RMoE)[10.1109/iccv.2015.169](http://dx.doi.org/10.1109/iccv.2015.169)[.](http://paperpile.com/b/DbTp46/1RMoE)

[Guo, Sheng, Tao Yang, Wei Gao, Chen Zhang, and Yanping Zhang. 2018. “An Intelligent Fault Diagnosis Method for Bearings with Variable Rotating Speed Based on Pythagorean Spatial Pyramid Pooling CNN.” *Sensors*  18 (11). https://doi.org/](http://paperpile.com/b/DbTp46/Ox9F)[10.3390/s18113857](http://dx.doi.org/10.3390/s18113857)[.](http://paperpile.com/b/DbTp46/Ox9F)

[He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2015. “Spatial Pyramid Pooling in Deep Convolutional Networks for Visual Recognition.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37 (9): 1904–16.](http://paperpile.com/b/DbTp46/fXnC)

[Ismail, Ahsiah, Mohd Yamani Idna Idris, Mohamad Nizam Ayub, and Lip Yee Por. 2018. “Vision-Based Apple Classification for Smart Manufacturing.” *Sensors*  18 (12). https://doi.org/](http://paperpile.com/b/DbTp46/2pZ8)[10.3390/s18124353](http://dx.doi.org/10.3390/s18124353)[.](http://paperpile.com/b/DbTp46/2pZ8)

[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/DbTp46/05fDf)[10.1007/978-3-319-10602-1\_48](http://dx.doi.org/10.1007/978-3-319-10602-1_48)[.](http://paperpile.com/b/DbTp46/05fDf)

[Li, Qiaoliang, Shiyu Li, Zhuoying He, Huimin Guan, Runmin Chen, Ying Xu, Tao Wang, Suwen Qi, Jun Mei, and Wei Wang. 2020. “DeepRetina: Layer Segmentation of Retina in OCT Images Using Deep Learning.” *Translational Vision Science & Technology* 9 (2): 61.](http://paperpile.com/b/DbTp46/h9Wo)

[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/DbTp46/e9u9r)[10.3390/s20247299](http://dx.doi.org/10.3390/s20247299)[.](http://paperpile.com/b/DbTp46/e9u9r)

[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/DbTp46/8hXAe)[10.1016/j.procs.2018.05.144](http://dx.doi.org/10.1016/j.procs.2018.05.144)[.](http://paperpile.com/b/DbTp46/8hXAe)

[“Real-Time Object Detection and Localization with SIFT-Based Clustering.” 2012. *Image and Vision Computing* 30 (8): 573–87.](http://paperpile.com/b/DbTp46/Egxha)

[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/DbTp46/b1r4F)

[Subudhi, Bidyadhar. 2009. *Computational Intelligence, Control and Computer Vision in Robotics and Automation*.](http://paperpile.com/b/DbTp46/QBdmJ)

[Sun, Yu, Rongrong Ni, and Yao Zhao. 2022. “MFAN: Multi-Level Features Attention Network for Fake Certificate Image Detection.” *Entropy*  24 (1). https://doi.org/](http://paperpile.com/b/DbTp46/KkVP)[10.3390/e24010118](http://dx.doi.org/10.3390/e24010118)[.](http://paperpile.com/b/DbTp46/KkVP)

[Zhang, Dingwen, Junwei Han, Gong Cheng, and Ming-Hsuan Yang. 2021. “Weakly Supervised Object Localization and Detection: A Survey.” *IEEE Transactions on Pattern Analysis and Machine Intelligence* PP (April). https://doi.org/](http://paperpile.com/b/DbTp46/ZboW)[10.1109/TPAMI.2021.3074313](http://dx.doi.org/10.1109/TPAMI.2021.3074313)[.](http://paperpile.com/b/DbTp46/ZboW)

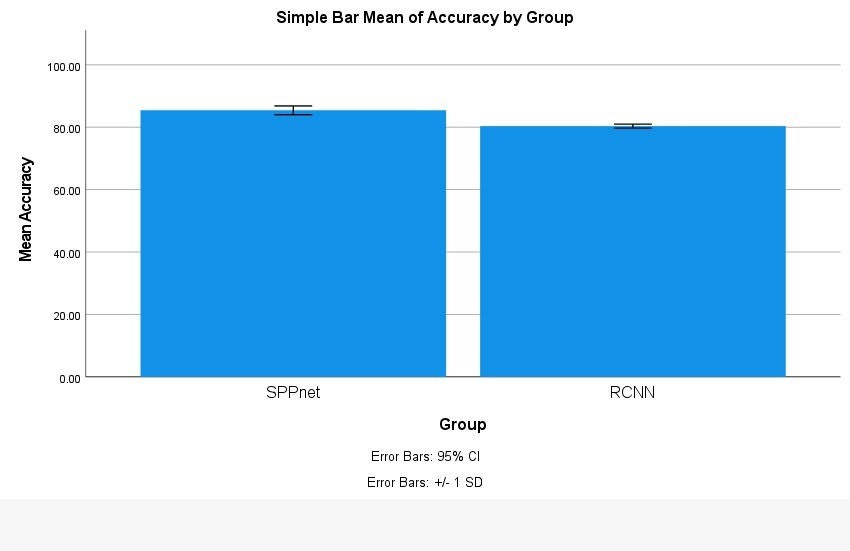
**TABLES AND FIGURES**

**Table-1.** Group Statistics of spatial pyramid pooling layer based convolutional neural network by grouping the iterations with sample size 6, Mean = 85.435, Standard Deviation = 1.39991. 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** | SPPnet | 10 | 85.435 | 1.39991 | 0.44269 |
|  | RCNN | 10 | 78.369 | 0.6414 | 0.20283 |
| **Loss** | SPPnet | 10 | 1.417 | 0.63858 | 0.20194 |
|  | RCNN | 10 | 1.771 | 0.43809 | 0.13854 |

**Table-2.** Independent Sample Test of Accuracy and Loss (calculate P-value = 0.001 and Significant value = 0.046, Mean Difference = 5.066 and confidence interval = (4.04297 - 6.08903). SPPnet CNN and 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** | 4.6 | 0.046 | 10.404 | 18 | <.001 | <.001 | 5.066 | 0.48694 | 4.04297 | 6.08903 |
| **Equal variances not assumed** |  |  | 10.404 | 12.619 | <.001 | <.001 | 5.066 | 0.48694 | 4.01078 | 6.12122 |
| **Loss** | **Equal variances assumed** | 0.176 | 0.68 | -1.446 | 18 | 0.083 | 0.165 | -0.354 | 0.24489 | -0.86849 | 0.16049 |
| **Equal variances not assumed** |  |  | -1.446 | 15.935 | 0.084 | 0.168 | -0.354 | 0.24489 | -0.87331 | 0.16531 |



**Fig. 1.** Comparison of regional proposal network based SPPnet in terms of mean accuracy. It explores that the mean accuracy is slightly better than 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 SPPnet vs R-CNN, Y-axis displaying the error bars with mean accuracy of detection +/-1 SD.