**Title Page:**

# A Supervised Stable Object Detection With Image Feature Extraction Using Image Segmentation By Comparing Histogram Of Oriented Gradients (HOG) Algorithm Over Scale Invariant Feature Transform (SIFT) Algorithm Model.

# 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

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, Support Vector Machine(SVM), Novel Feature Descriptor, Invariant Feature Transforms.

**ABSTRACT**

**Aim:** The aim of the research work is to improve the accuracy of object detection using novel image segmentation using machine learning algorithms. **Materials and Methods:** The categorisingis performed by adopting a sample size of n = 10 in Histogram Of Oriented Gradients (HOG) and sample size n = 10 in Scale Invariant Feature Transform (SIFT) algorithms with a sample size = 10. **Results and Discussion:** The analysis of the results shows that the Histogram Of Oriented Gradients (HOG) has a high accuracy of (92.4919 %) in comparison with the Scale Invariant Feature Transform (SIFT) (86.304 %). There is a statistically significant difference between the study groups with (p<.001). **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, Support Vector Machine(SVM), Novel Feature Descriptor, Invariant Feature Transforms.

**INTRODUCTION**

The advances of object detection in computer vision fields required a lot of image processing and segmentation to extract and identify the objects in the image. The purpose of the research is to improve the Accuracy of object detection and gradient extraction from the input through image segmentation using machine learning algorithms by using regional proposal networks[(Anitha et al. 2016)](https://paperpile.com/c/5J1YUJ/st86). 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/5J1YUJ/W6xO). Computer vision is a major thing in this daily changing world with content-based image retrieval, machine vision, medical imaging, including volume rendered images from computed tomography and magnetic resonance imaging, self-driving cars to face recognition in our mobiles, traffic detections, and many more security purposes [(El-Baz, Jiang, and Suri 2016)](https://paperpile.com/c/5J1YUJ/Rm6i). The goal was to develop an inexpensive object detection model and improve its accuracy and speed to detect objects faster and more accurately[(Novotny and Matas 2015)](https://paperpile.com/c/5J1YUJ/dvaZR).

Identifying objects using different image segmentation algorithms for over past years and several surveys and detection and segmentation have been published in the last years over 20,600 articles from Google Scholar, 7782 journals from IEEE Xplore, 18,339 research articles from ScienceDirect. Among all the research articles and journals, the most cited paper is [(Nilsback and Zisserman 2008)](https://paperpile.com/c/5J1YUJ/Gwjh). The model proposed by Nilsback and Zisserman is very advanced and improved to extract low level features of a given input image compared to the models with different classifiers and features. The proposed feature descriptor creates a very discriminating feature for describing image content[(Wang et al. 2021)](https://paperpile.com/c/5J1YUJ/ej3w). Four image datasets are used to evaluate the proposed descriptor. Extensive experiments have been conducted using different classifiers and different image features to demonstrate the superiority of the proposed method[(Vohra and Prodanov 2021)](https://paperpile.com/c/5J1YUJ/df6o). A new image feature descriptor for content based image retrieval using scale invariant feature transforms and local derivative pattern[(Giveki, Soltanshahi, and Montazer 2017)](https://paperpile.com/c/5J1YUJ/IAXt).

This method which was used before has less accuracy in segmenting objects. It is necessary to determine and segment the object to identify in very less time to prevent issues. For example, a common use for image segmentation in medical purpose images is to detect and label 3D volume images or pixels in voxels that represent tumours in the patient’s body. In order to sequence the methods and techniques in this proposed model is generally much better than the convolutional neural network because it takes a lot more time to train the dataset. Image feature descriptors are used to extract homogeneous features in an image. Texture features are a very useful characterization for a wide range of images used in this model. Texture features may be widely classified into spatial texture function extraction strategies and spectral texture features extraction strategies primarily based totally at the domain from which they may be extracted. In the primary approach, texture features have become extracted with the aid of computing the pixel data or with the aid of locating the neighbourhood pixel systems in the original photograph domain [(Hung, Song, and Lan 2019)](https://paperpile.com/c/5J1YUJ/i0Jd).

**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 Histogram of Oriented Gradients(HOG) and Scale Invariant Feature Transform(SIFT) algorithms, which is used to detect objects in the image with various image datasets and labels. Group 1 is the Histogram of Oriented Gradients(HOG) with the sample size of 10 and Group 2 is the Scale Invariant Feature Transform(SIFT) 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 Histogram of Oriented Gradients(HOG) =.32554 and Scale Invariant Feature Transform(SIFT) = .45535.

**HISTOGRAM OF ORIENTED GRADIENTS (HOG) ALGORITHM.**

The primary algorithm used in this proposed model is Histogram of Oriented Gradients(HOG) algorithm. HOG are feature descriptors that can be used in image segmentation and computer vision for object detection purposes. In this paper, we extract the HOG feature from infrared images and use this feature as the basis for classification[(Liang, Wang, and Zhang 2015)](https://paperpile.com/c/5J1YUJ/7XSq). This technique counts the number of occurrences of the gradient direction in the localised part of the image. This method is quite similar to edge plotting, The HOG descriptor focuses on the structure or shape of an object. It is better than any edge descriptor because it uses the magnitude as well as the angle of the gradient to compute the features. For image regions, it creates a histogram using the magnitude and direction of the gradient. From the input image the gradient will be calculated. Gradient is obtained by combining the magnitude and angle of the image. Considering a block of 3×3 pixels, the first Gx and Gy are calculated for each pixel. The first Gx and Gy are calculated using the formulas below for each pixel value.(The Gx and Gy are the change in gradients respective to the direction. The technique counts occurrences of gradient orientation in localised portions of an image and trained classifiers using different datasets will recognize the objects in the image. By extracting the features in the image and these features obtained will be given to the classifier. In this model we take the Support Vector Machine(SVM) classifier. Support Vector Machine(SVM) classifiers look for optimal hyperplanes as a decision function. The HOG features of the probe and the input image are taken by the Support Vector Machine(SVM), the classifier looks for the closest feature matching with the trained datasets and gives as output.

**Pseudocode for Histogram of Oriented Gradients Algorithm.**

Import numpy as np

Import keras

Import os

Import cv2

From keras.models import sequential

From keras.layers import Dense, Dropout, Flatten, Conv2D, MaxPooling2D

Path\_test → [ ]

Path\_train → [ ]

For filename in os.listdir(train\_path):

if (filename.split(‘.’)[1]==”jpg”):

labels\_train.append(filename.split(‘\_’)[0])

path\_train.append(os.path.join(train\_path, filename))

Label\_train\_unique → np.unique(np.array(labels\_train))

Label\_test\_unique → np.unique(np.array(labels\_test))

Image → cv2.imread(path\_train[5])

Rgb\_img →cv2.cvtColor(image,cv2.color\_BGR2RGB)

plt.imshow(rgb\_img)

Y\_train → keras.utils.np\_utils.to\_categorical(temp\_train,4)

Y\_test → keras.utils.np\_utils.to-categorical(temp\_test, 4)

Svm\_hog → sequential()

svm\_hog.add(Conv2D(32,(3,3), padding=’same’, input\_shape=(50,50,3),activision=”relu”))

svm\_hog.add(Maxpooling2D(pool\_size=(2,2)))

svm\_hog.add(Dropout(0.25))

hog\_image\_rescaled → exposure.rescale\_intensity(hog\_image, in\_range=(0, 10))

ax2.imshow(hog\_image\_rescaled, cmap=plt.cm.gray)

plt.show()

**SCALE INVARIANT FEATURE TRANSFORM (SIFT) ALGORITHM.**

One of the other approaches to achieving the local descriptor is Scale Invariant Feature Transform(SIFT) Algorithm. SIFT method converts image data into relative scale invariant coordinates according to the specific characteristics of the locality and revolves around four main stages: detect extreme points in space, position of key points, orientation like signature descriptor and keypoint(Invariant Feature Transforms) [(Goncalves, Corte-Real, and Goncalves 2011)](https://paperpile.com/c/5J1YUJ/5DJc). The detection is performed by searching over all scales and image locations in order to identify potential interest points that are invariant to scale and orientation. Once a set of keypoint candidates is obtained, the next step is to accurately localise them. This is performed by rejecting those keypoints, which have low contrast or are poorly localised along an edge, by a detailed fit to the nearby data for location, scale, and ration of principal curvatures. So simply one should strive matching patches round the salient characteristics points, however those patches will themselves alternate if there’s alternate in object pose or illumination. So those patches will result in numerous false matches and correspondences.

**Pseudocode for Scale Invariant Feature Transform Algorithm.**

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

import keras

from keras.layers import Dense, Dropout, Flatten, Conv2D, MaxPooling2D

from keras.utils.np\_utils import to\_categorical

image = cv2.imread(path\_train[5])

rgb\_img = cv2.cvtColor(image, cv2.COLOR\_BGR2RGB)

X\_train = []

for path in path\_train:

img = cv2.imread(path)

X\_train = np.array(X\_train)

X\_test = []

for path in path\_test:

img = cv2.imread(path)

X\_test = np.array(X\_test)

For all octaves

{

List keypoint\_list

For all scales

{

ConvolveImageGaussParallel();

BuildDoGParallel();

//Detect Keypoint

#pragma omp parallel for

For all pixels p in image

{

If (IsKeypoint(p))

#pragma omp critical

keypoint\_list.add(p);

}

}

#pragma omp parallel for

For all pixels kp in keypoint\_list

{

ExtractFeature(kp);

}

DownSampleImageParallel();

}

import os

import cv2

from sklearn import preprocessing

fig, (ax1, ax2) = plt.subplots(1, 2, figsize=(8, 4), sharex=True, sharey=True)

ax1.axis('off')

ax1.imshow(image, cmap=plt.cm.gray)

**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.

**RESULT**

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 Histogram Of Oriented Gradients (HOG) of by Scale Invariant Feature Transform (SIFT) by grouping with iterations sample size of 10, mean = 92.4919 Standard Deviation = 0.36961 , Standard Error Mean = 0.11688. 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= 0.125, Mean Difference= 6.1879 and confidence interval = (5.52833 - 6.84747) of Histogram Of Oriented Gradients (HOG) based Object detection and Scale Invariant Feature Transform (SIFT) based Object detection is tabulated in Table 3, which shows there is a significant difference between the two groups since P<0.005 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 Histogram Of Oriented Gradients (HOG) have higher accuracy mean (92.4919 % ) and Less Loss mean 1.1475 % compared to selective search based Scale Invariant Feature Transform (SIFT) with accuracy (86.304 %) and Less Loss mean 1.762 % 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/5J1YUJ/Jr1ZY). The most important features of detecting objects using Histogram Of Oriented Gradients (HOG) [(Ren et al. 2017)](https://paperpile.com/c/5J1YUJ/A6LMr) is pragmatically proven to be highly effective than Scale Invariant Feature Transform (SIFT). The core argument is that to prove that detection of objects using image segmentation techniques may be a better method than other methods to extract image features which helps to identify the objects easily. In many of the recent findings, it has been observed that this is the most focused and better method of extracting image gradients with more accuracy than SIFT[(Zhou et al. 2021)](https://paperpile.com/c/5J1YUJ/okga).

The purpose of computer vision is to reach a high level of understanding through photos and videos. Traditional view algorithms that take a fully rendered colour image as input. However, in situations where colour is not needed such as gradient-based algorithms discussed in this article, decoding is redundant. It is not only time consuming but also wasteful of storage space to get almost the same result if we consider colours of the image. So these algorithm models consider the input image and convert it into grey and extract image features and gradients. For colour images, grey scale images are generated for gradient extraction.

To compare the performance of HOG descriptors extracted from colour images and pattern images over the SIFT descriptors, the traditional HOG with support vector machine(SVM) framework has been proposed in this model and used to detect the fruit images from various datasets provided. Where models are trained and tested on each dataset separately. Precision-recall curves along with average precision are used to present the detection results. And the SIFT feature extraction, extremums are searched among a 5 × 5 neighbourhood instead of 3 × 3. To confirm the scale and the rotation invariant property of the generated SIFT features key points detected from the transformed images, i.e the image is scaled, rotated and blurred. Main attractions matched with what was detected from pictures. The repeatability criteria introduced in were used to evaluate the performance of the SIFT descriptor to find the dots match.

**CONCLUSION**

A supervised stable object detection with image feature extraction using image segmentation by comparing Histogram of Oriented Gradients (HOG) algorithm over Scale Invariant Feature Transform (SIFT) algorithm model. The current study focused on machine learning algorithms, Histogram of Oriented Gradients (HOG) over Scale Invariant Feature Transform (SIFT) algorithm for higher classification of object detection. It can be slightly improved based on high trained datasets in future. The outcome of the Histogram of Oriented Gradients (HOG) showed higher accuracy (92.4919 % ) than the Scale Invariant Feature Transform (SIFT) algorithm (86.304 %).

**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**

[Anitha, U., S. Malarkkan, J. Premalatha, and P. Grace Kanmani Prince. 2016. “Study of Object Detection in Sonar Image Using Image Segmentation and Edge Detection Methods.” *Indian Journal of Science and Technology*. https://doi.org/](http://paperpile.com/b/5J1YUJ/st86)[10.17485/ijst/2016/v9i42/104608](http://dx.doi.org/10.17485/ijst/2016/v9i42/104608)[.](http://paperpile.com/b/5J1YUJ/st86)

[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/5J1YUJ/W6xO)[10.1016/j.jksuci.2021.08.001](http://dx.doi.org/10.1016/j.jksuci.2021.08.001)[.](http://paperpile.com/b/5J1YUJ/W6xO)

[El-Baz, Ayman, Xiaoyi Jiang, and Jasjit S. Suri. 2016. *Biomedical Image Segmentation: Advances and Trends*. CRC Press.](http://paperpile.com/b/5J1YUJ/Rm6i)

[Giveki, Davar, Mohammad Ali Soltanshahi, and Gholam Ali Montazer. 2017. “A New Image Feature Descriptor for Content Based Image Retrieval Using Scale Invariant Feature Transform and Local Derivative Pattern.” *Optik*. https://doi.org/](http://paperpile.com/b/5J1YUJ/IAXt)[10.1016/j.ijleo.2016.11.046](http://dx.doi.org/10.1016/j.ijleo.2016.11.046)[.](http://paperpile.com/b/5J1YUJ/IAXt)

[Goncalves, Hernâni, Luís Corte-Real, and José A. Goncalves. 2011. “Automatic Image Registration Through Image Segmentation and SIFT.” *IEEE Transactions on Geoscience and Remote Sensing*. https://doi.org/](http://paperpile.com/b/5J1YUJ/5DJc)[10.1109/tgrs.2011.2109389](http://dx.doi.org/10.1109/tgrs.2011.2109389)[.](http://paperpile.com/b/5J1YUJ/5DJc)

[Hung, Chih-Cheng, Enmin Song, and Yihua Lan. 2019. “Image Texture, Texture Features, and Image Texture Classification and Segmentation.” *Image Texture Analysis*. https://doi.org/](http://paperpile.com/b/5J1YUJ/i0Jd)[10.1007/978-3-030-13773-1\_1](http://dx.doi.org/10.1007/978-3-030-13773-1_1)[.](http://paperpile.com/b/5J1YUJ/i0Jd)

[Liang, Ying, Luping Wang, and Luping Zhang. 2015. “Infrared Image Segmentation Using HOG Feature and Kernel Extreme Learning Machine.” *AOPC 2015: Image Processing and Analysis*. https://doi.org/](http://paperpile.com/b/5J1YUJ/7XSq)[10.1117/12.2199352](http://dx.doi.org/10.1117/12.2199352)[.](http://paperpile.com/b/5J1YUJ/7XSq)

[Nilsback, Maria-Elena, and Andrew Zisserman. 2008. “Automated Flower Classification over a Large Number of Classes.” *2008 Sixth Indian Conference on Computer Vision, Graphics & Image Processing*. https://doi.org/](http://paperpile.com/b/5J1YUJ/Gwjh)[10.1109/icvgip.2008.47](http://dx.doi.org/10.1109/icvgip.2008.47)[.](http://paperpile.com/b/5J1YUJ/Gwjh)

[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/5J1YUJ/dvaZR)[10.1109/iccv.2015.137](http://dx.doi.org/10.1109/iccv.2015.137)[.](http://paperpile.com/b/5J1YUJ/dvaZR)

[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/5J1YUJ/Jr1ZY)[10.1016/j.procs.2018.05.144](http://dx.doi.org/10.1016/j.procs.2018.05.144)[.](http://paperpile.com/b/5J1YUJ/Jr1ZY)

[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/5J1YUJ/A6LMr)

[Vohra, Sumit K., and Dimiter Prodanov. 2021. “The Active Segmentation Platform for Microscopic Image Classification and Segmentation.” *Brain Sciences* 11 (12). https://doi.org/](http://paperpile.com/b/5J1YUJ/df6o)[10.3390/brainsci11121645](http://dx.doi.org/10.3390/brainsci11121645)[.](http://paperpile.com/b/5J1YUJ/df6o)

[Wang, Lei, Chunhong Chang, Zhouqi Liu, Jin Huang, Cong Liu, and Chunxiang Liu. 2021. “A Medical Image Fusion Method Based on SIFT and Deep Convolutional Neural Network in the SIST Domain.” *Journal of Healthcare Engineering* 2021 (April): 9958017.](http://paperpile.com/b/5J1YUJ/ej3w)

[Zhou, Wei, Ling Zhang, Shengyu Gao, and Xin Lou. 2021. “Gradient-Based Feature Extraction From Raw Bayer Pattern Images.” *IEEE Transactions on Image Processing: A Publication of the IEEE Signal Processing Society* 30 (May): 5122–37.](http://paperpile.com/b/5J1YUJ/okga)

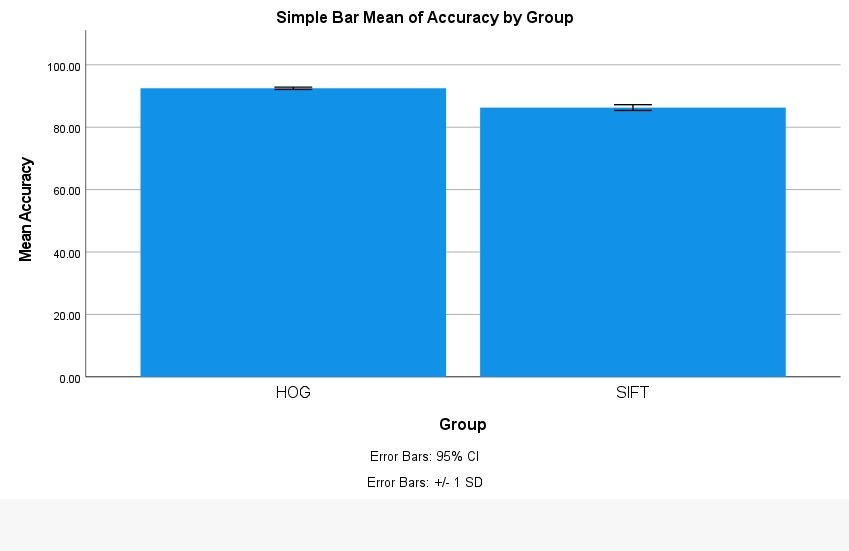
**TABLES AND FIGURES**

**Table-1.** Group Statistics of Histogram Of Oriented Gradients (HOG) by grouping the iterations with sample size 6, Mean = 92.4919, Standard Deviation = 0.36961. 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** | HOG | 10 | 92.4919 | 0.36961 | 0.11688 |
|  | SIFT | 10 | 86.304 | 0.9214 | 0.29137 |
| **Loss** | HOG | 10 | 1.1475 | 0.66829 | 0.21133 |
|  | SIFT | 10 | 1.762 | 0.39662 | 0.12542 |

**Table-2.** Independent Sample Test of Accuracy and Loss (calculate P-value = 0.001 and Significant value = 0.125, Mean Difference = 6.1879 and confidence interval = (5.52833 - 6.84747). Histogram Of Oriented Gradients (HOG) and Scale Invariant Feature Transform (SIFT) 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** | 2.59 | 0.125 | 19.71 | 18 | <.001 | <.001 | 6.1879 | 0.31394 | 5.52833 | 6.84747 |
| **Equal variances not assumed** |  |  | 19.71 | 11.823 | <.001 | <.001 | 6.1879 | 0.31394 | 5.50274 | 6.87306 |
| **Loss** | **Equal variances assumed** | 3.101 | 0.095 | -2.501 | 18 | 0.011 | 0.022 | -0.6145 | 0.24575 | -1.1308 | -0.0982 |
| **Equal variances not assumed** |  |  | -2.501 | 14.64 | 0.012 | 0.025 | -0.6145 | 0.24575 | -1.13942 | -0.08958 |

****

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