IMPROVED OBJECT DETECTION WITH NEURAL NETWORK

Dave Krupali Arvindbhai[1], A. R. Kazi[2]

[1] Student, Computer Engineering, Gujarat Technical University, Gujarat, India
[2]Assistant Professor, Computer Engineering, Gujarat Technical University, Gujarat, India

ABSTRACT

Object detection *is a computer technology related to computer vision and image processing that deals with detecting instances of semantic objects of a certain class (such as humans, buildings, or cars) in digital images and videos.In Image Processing two methods are used CLAASIFICATION and DETECTION.* ***CNN*** *is a one type of Feature Extraction but it is used for detect multiple shape and object. CNN is used as a classification.In this paper we focuses on convolution neural networks and Deep Pyramid technique allowing further insights into their internal operation. After giving a brief introduction to neural networks and its layers, we review supervised training of neural networks in detail. The second section introduces the different types of layers present in recent convolutional neural networks. Based on these basic building blocks, we discuss the architecture of the convolutional neural network. In the third section we focuses on the deep pyramid technique. In* ***Deep Pyramid****, or* ***pyramid representation****, is a type of multi-scale signal representation developed by the computer vision, image processing and signal processing communities, in which a signal or an image is subject to repeated smoothing and sub sampling.*

Keywords:- *Object detection, Pre-processing, Feature Extraction, Convolutional Neural network, Deep Pyramid, Classification*

1. INTRODUCTION:

In image processing the object detection is generated large amount of data and class of the image in multiple level .In this survey used of CNN technique to detect the right portion of the object that the input comes. CNN have associated terminology and a set of concepts that is unique to them, and that sets them apart from other types of neural network architecture. CNN are usually applied to image data. Every image is a matrix of pixel values. The range of value that can be encoded in each pixel depends upon its bit size. CNN is biologically-inspired models[7]. The most popular application for CNNs in the recent times has been image analysis[8]. DEEP PYRAMID is the technique to shows the support of the CNN that uses to generated the double accuracy of the object. In future predicted that by use of this two technique image detection has clear to detect the object.

2. SYSTEM MODEL:

Object Detection is the process of selecting, exploring and classify large amounts of images in order to discover unknown techniques or relationships which provides a clear and useful result to the image providers.

2.1 Object Detection with neural network involves the following Steps:

1. Input Image: The first step is to identify image.
2. Pre-Processing: Its processes the input image to produce output that used as input to another image also.
3. Feature Extraction: The purpose of this step is to detect and select the important data of the image and its done with DEEP PYRAMID method.
4. Classification: In used of the CNN the important part is detected and the other chunk of data is deleted. For next step the detect part is classify its class by the classification step.

5. Object Detection: Based on the results of the CNN technique, an analysis is conducted to determine the image detection from the analysis and create a detection of the image and the final image part is detected.

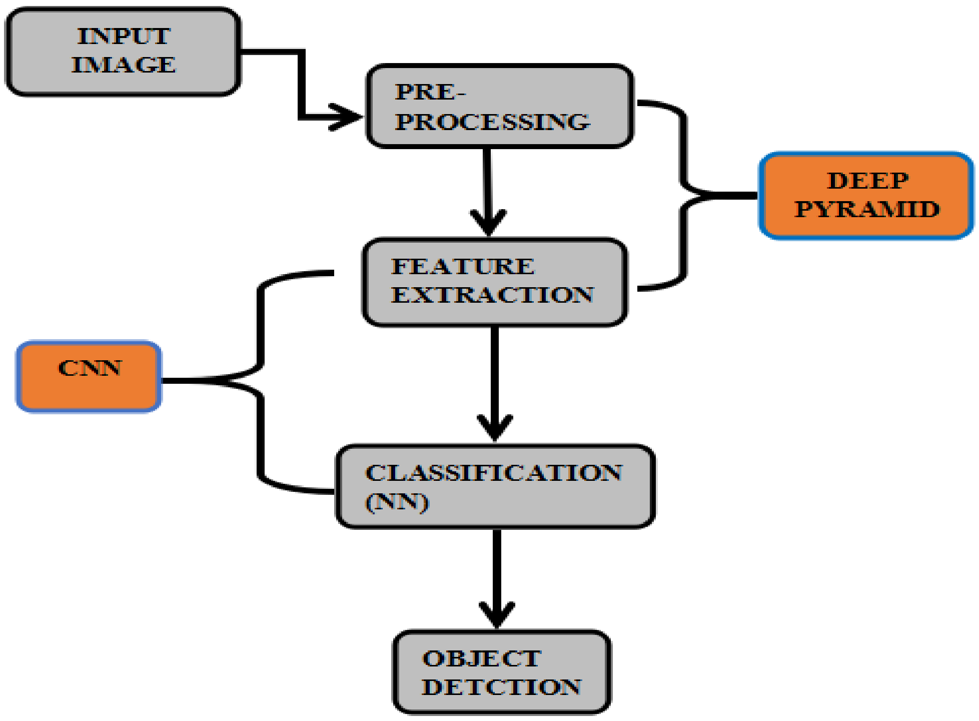


Fig -1: Object Detection with Neural Network Process

3. RELATED BACKGROUND:

3.1 Convolutional Neural Network(CNN):

In Image Processing Object Detection is most important. There are many techniques in object detection like SVM, HOG and many more. But CNN is the best in performance and time in multiple ways.

There are two types of Image Processing Techniques:

Classification: It don't know the object but it knows the class of the object.

Detection: It detect the object in the image.(ex. dog , people)

CNN is one type of machine learning algorithm.CNN is also called as CONVOLUTIONAL NEURAL NETWORK (ConvNet). In machine learning algorithm there is always usefull Feature Extraction.It is a techniques to use to take only important data (color, texture , shape) from the image. Sometimes it is difficult to generate Feature Extraction in a huge dataset like multiple images. So by that I would like to use CNN in object detection.CNN is a one type of Feature Extraction but it is used for detect multiple shape and object. CNN is used as a classification.In the object detection if we used the CNN so anyone else like HOG,SVM and many more algorithm is not need to generate and it also increase the performance of the process. Convolutional networks were inspired by biological processes.

In CNN it takes chunks of data and then detect the image. And generated the nodes in 3D forms. In CNN using deep pyramid. It is use in CNN as a layer like HOG,SVM. But not a Feature Extraction. Deep Pyramid is mostly used to capable the CNN in multiple nodes. In Feature Extraction it used to generated separately in system but this is good to generated CNN. CNN have three important layers:(1) Convolution layer, (2) Pooling layer and (3) Fully-connected layer.

3.2 Deep Pyramid:

Pyramid, or pyramid representation, is a type of multi-scale signal representation developed by the computer vision, image processing and signal processing communities, in which a signal or an image is subject to repeated smoothing and sub sampling. Pyramid representation is a predecessor to scale-space representation and multiresolution analysis. Sometimes Feature Extraction do not used in multiple image detection So the main advantage of Deep Pyramid is that it is like Feature Extraction but it is used in CNN in all bid and small data types and images.

Rectifier Linear Unit: F(x)=max(0,x)

Sigmoid Linear Unit: $S(t) = \frac{1}{1+e^{-t}}.$

Convolution Formulas

4. PREDICTION OF OBJECT DETECTION:

In this paper, Multi-resolution image features may be approximated via extrapolation from nearby scales, rather than being computed explicitly. This fundamental insight allows us to design object detection algorithms that are as accurate, and considerably faster, than the state-of-the-art.

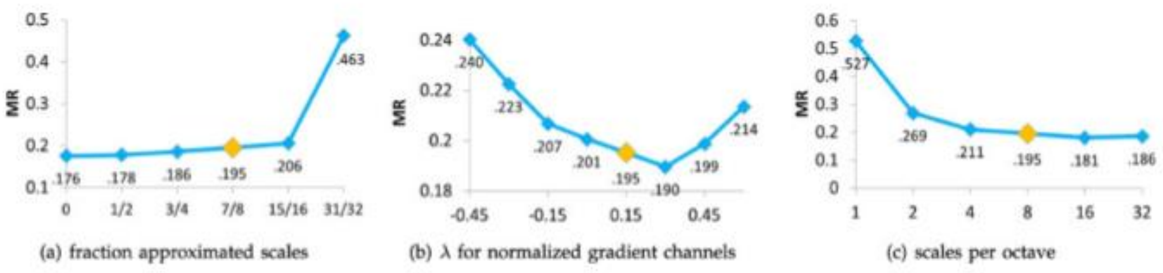


Fig: Fractions scale

Our results are not restricted to object detection nor to visual recognition. The foundations we have developed should readily apply to other computer vision tasks where a fine-grained scale sampling of features is necessary as the image processing front end[1]. In this survey, We have presented a framework for detecting multiple object instances in images which is similar to the traditional Hough transform.

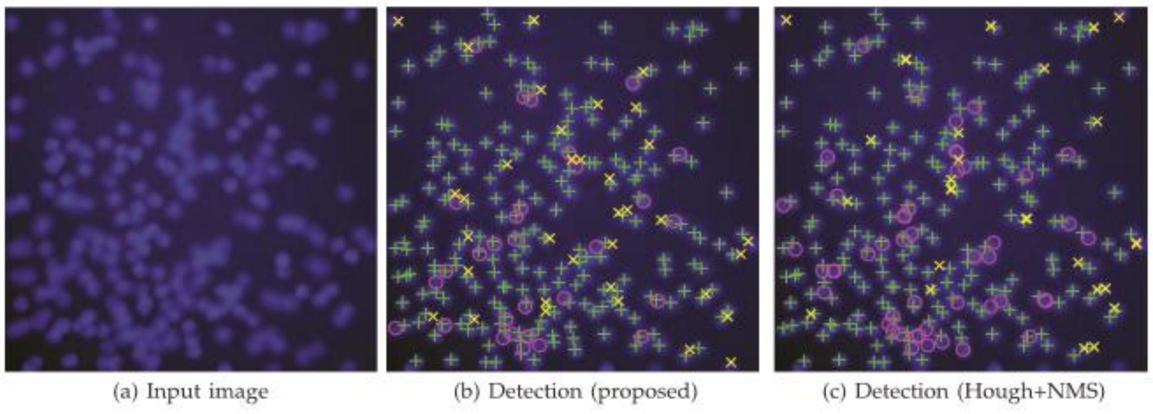


Fig. 9 HBM Process

This is investigated in our follow up work [6], where we consider a multilayer graphical model for the parsing of an image of a man-made environment. The model [6] includes layers for 1) estimation of straight lines based on edge pixels (this layer corresponds to the model derived and evaluated in this paper), 2) grouping of the detected lines in parallel line families, 3) the estimation of the horizon and the zenith of the scene[2]. In this survey, We investigate the problem of incorporating color for object detection. Most state-of-the-art object detectors rely on shape while ignoring color. Recent approaches to augmenting intensity-based detectors with color often provide inferior results for object categories with varying importance of color and shape. We propose the use of color attributes as an explicit color representation for object detection Finally, we introduce a new dataset of cartoon characters where color plays an important role[3]. In this paper,In this paper we have presented the Shape Boltzmann Machine, a strong generative model of object shape. The SBM is based on the general DBM architecture, a form of undirected graphical model that makes heavy use of latent variables to model high-order dependencies between the observed variables. We believe that the *combination of (a)* carefully chosen connectivity and capacity constraints, along with (b) a hierarchical architecture, and (c) a training procedure that allows for the joint optimization of the full model, is key to the success of the SBM.

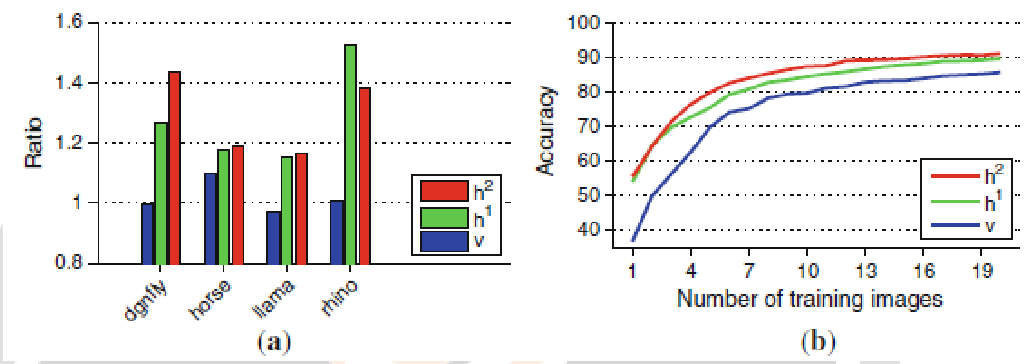
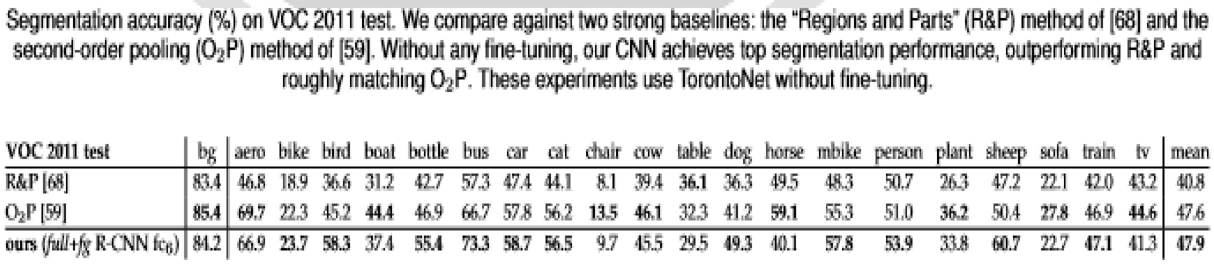


Fig: (a)The ratio of interand intra-class distances, (b) GLM classification

Overall we believe that by integrating powerful component models like the SBM into comprehensive generative models of images, performance in many computer vision tasks can be improved. We believe this to be a very promising direction of research[4]. In this surveying, In recent years, object detection performance had stagnated.The best performing systems were complex ensembles combining multiple low-level image features with high-level context from object detectors and scene classifiers. This paper presents a simple and scalable object detection algorithm that gives more than a 50% relative improvement over the best previous result on PASCAL VOC 2012.


Segmentation accuracy (%) on VOC 2011 test. We compare against two strong baselines: the "Regions and Parts" (R&P) method of [68] and the second-order pooling (O_2P) method of [59]. Without any fine-tuning, our CNN achieves top segmentation performance, outperforming R&P and roughly matching O_2P. These experiments use TorontoNet without fine-tuning.

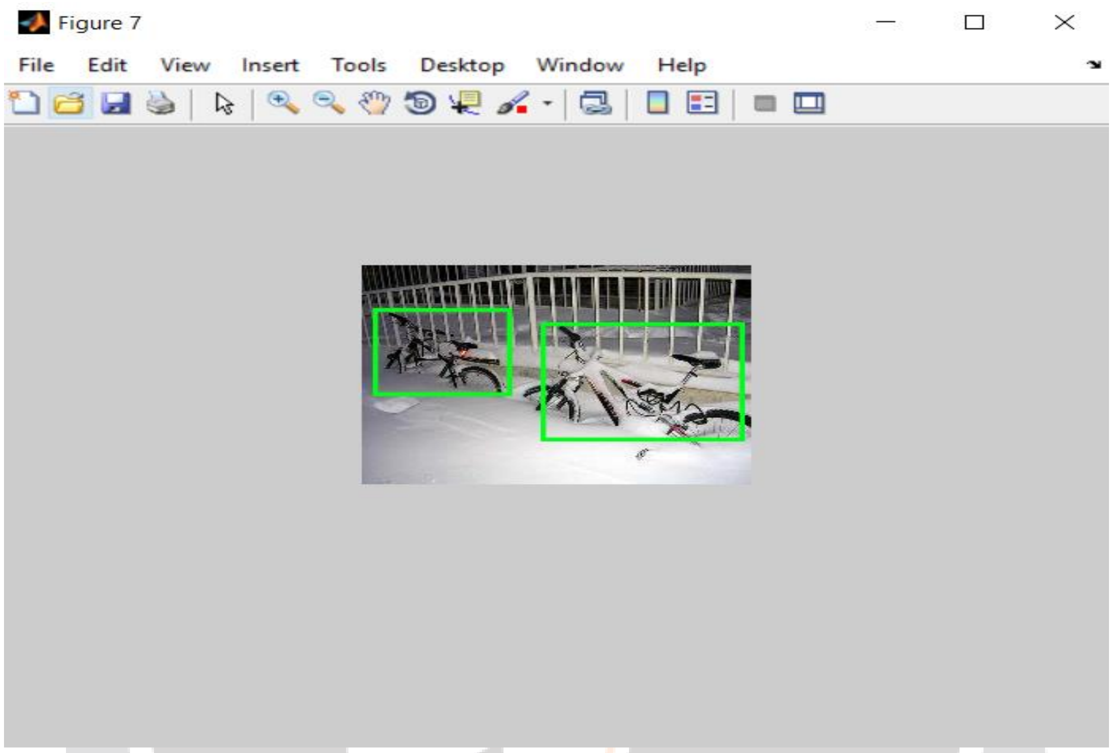
VOC 2011 test	bg	aero	bike	bird	boat	bottle	bus	car	cat	chair	cow	table	dog	horse	mbike	person	plant	sheep	sofa	train	tv	mean
R&P [68]	83.4	46.8	18.9	36.6	31.2	42.7	57.3	47.4	44.1	8.1	39.4	**36.1**	36.3	49.5	48.3	50.7	26.3	47.2	22.1	42.0	43.2	40.8
O_2P [59]	**85.4**	**69.7**	22.3	45.2	**44.4**	46.9	66.7	57.8	56.2	**13.5**	**46.1**	32.3	41.2	**59.1**	55.3	51.0	**36.2**	50.4	**27.8**	46.9	**44.6**	47.6
ours (*full+fg* R-CNN fc_6)	84.2	66.9	**23.7**	**58.3**	37.4	**55.4**	**73.3**	**58.7**	**56.5**	9.7	45.5	29.5	**49.3**	40.1	**57.8**	**53.9**	33.8	**60.7**	22.7	**47.1**	41.3	**47.9**

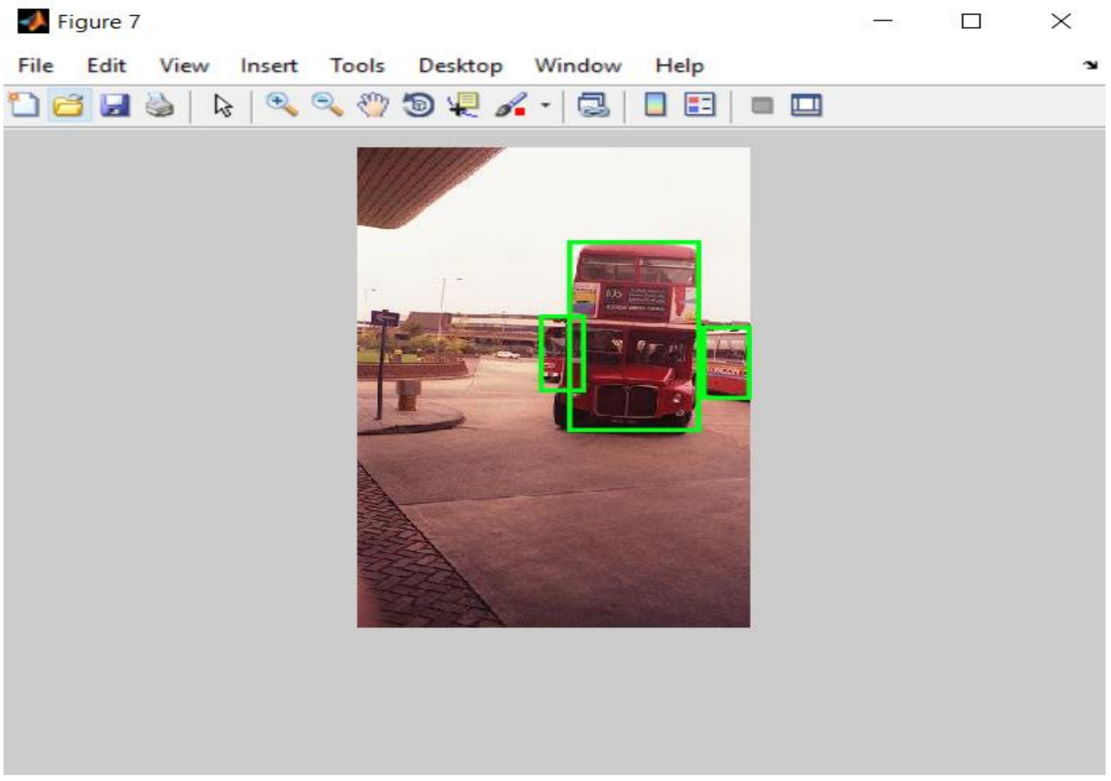
Table: Segmentation Accuracy(%)

We achieved this performance through two insights. The first is to apply high-capacity convolutional networks to bottom-up region proposals in order to localize and segment objects. The second is a paradigm for training large CNNs when labeled training data are scarce[5]. By this , in this survey paper we predict that the accuracy(%) of the object is increase use of the CNN and DEEP PYRAMID technique.

5. IMPLEMENTATION AND RESULTS:

Let's consider following outputs of the image for mention system.





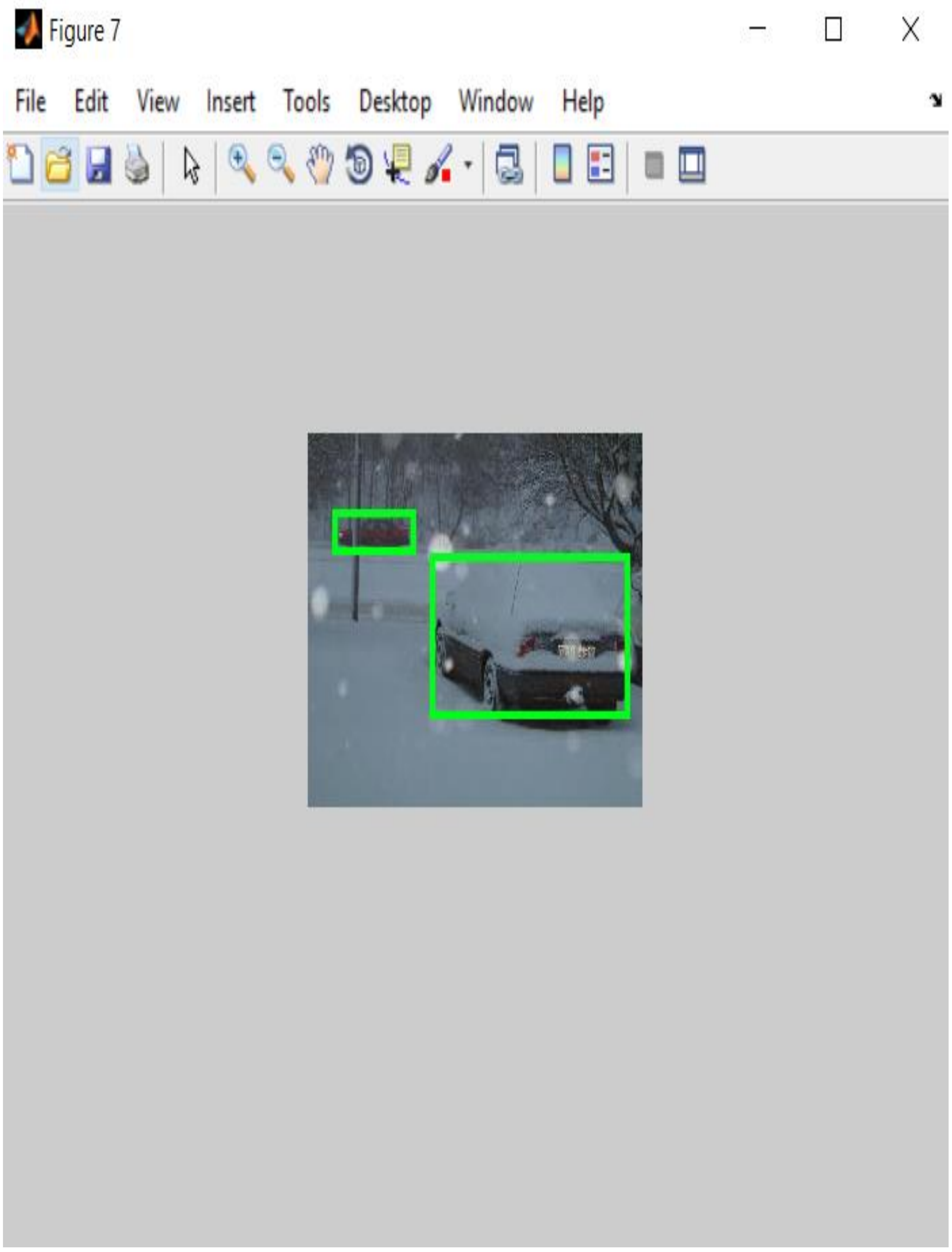


Fig: Detect multiple object in multiple images from image using CNN and DEEP PYRAMID

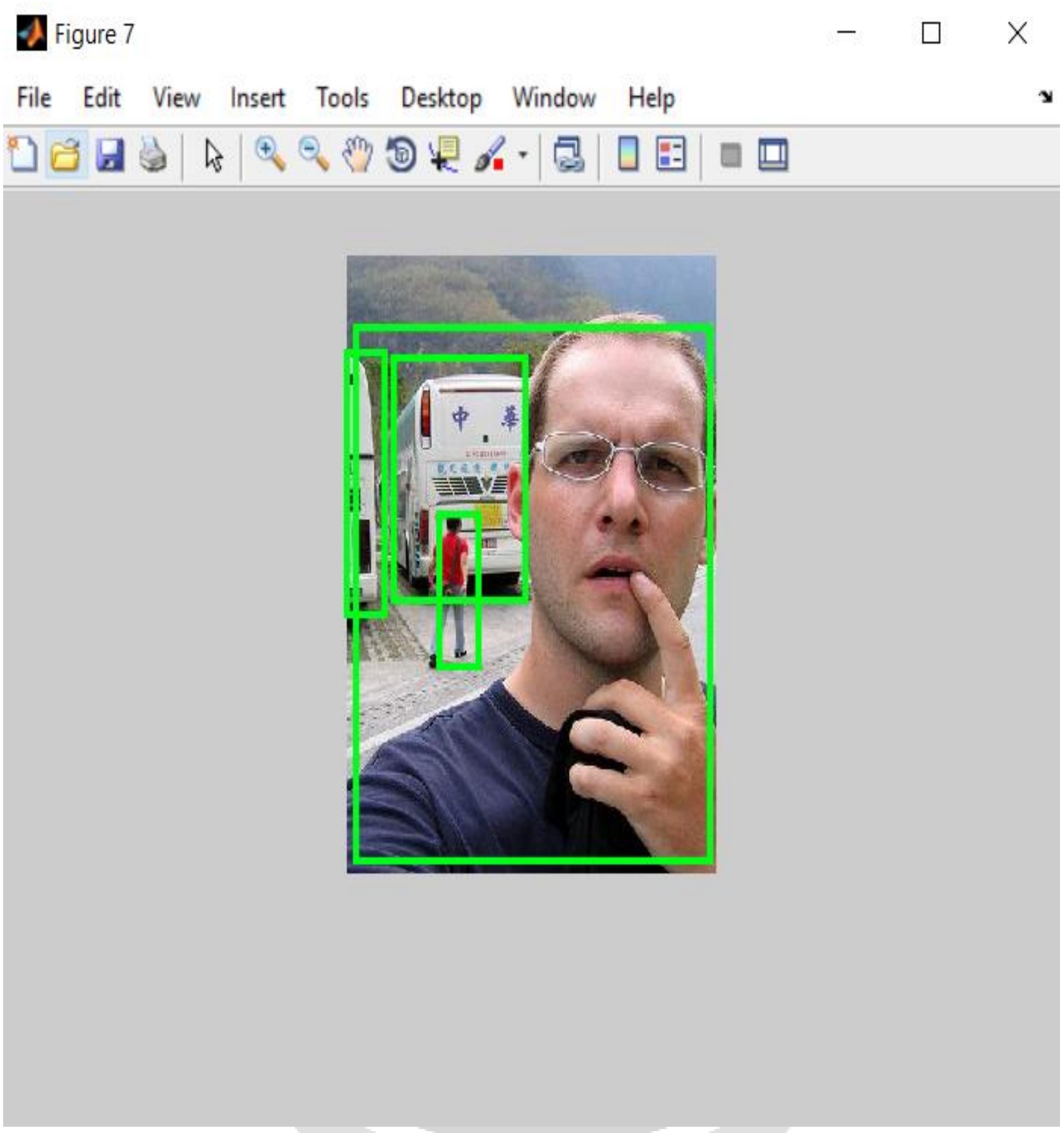


Fig: Detect multiple object from image using CNN and DEEP PYRAMID

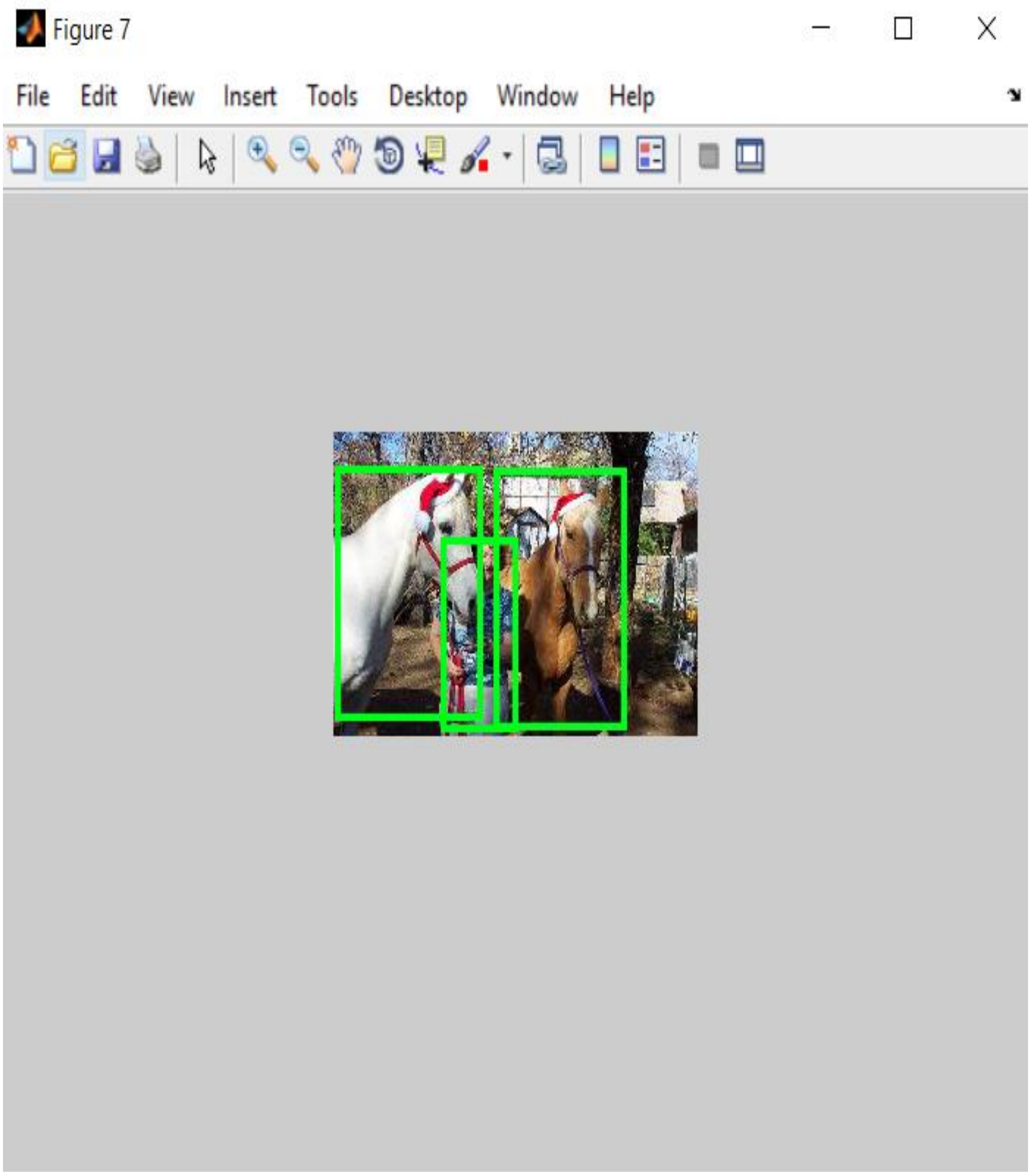


Fig: Detect multiple object and hide objects from image using CNN and DEEP PYRAMID

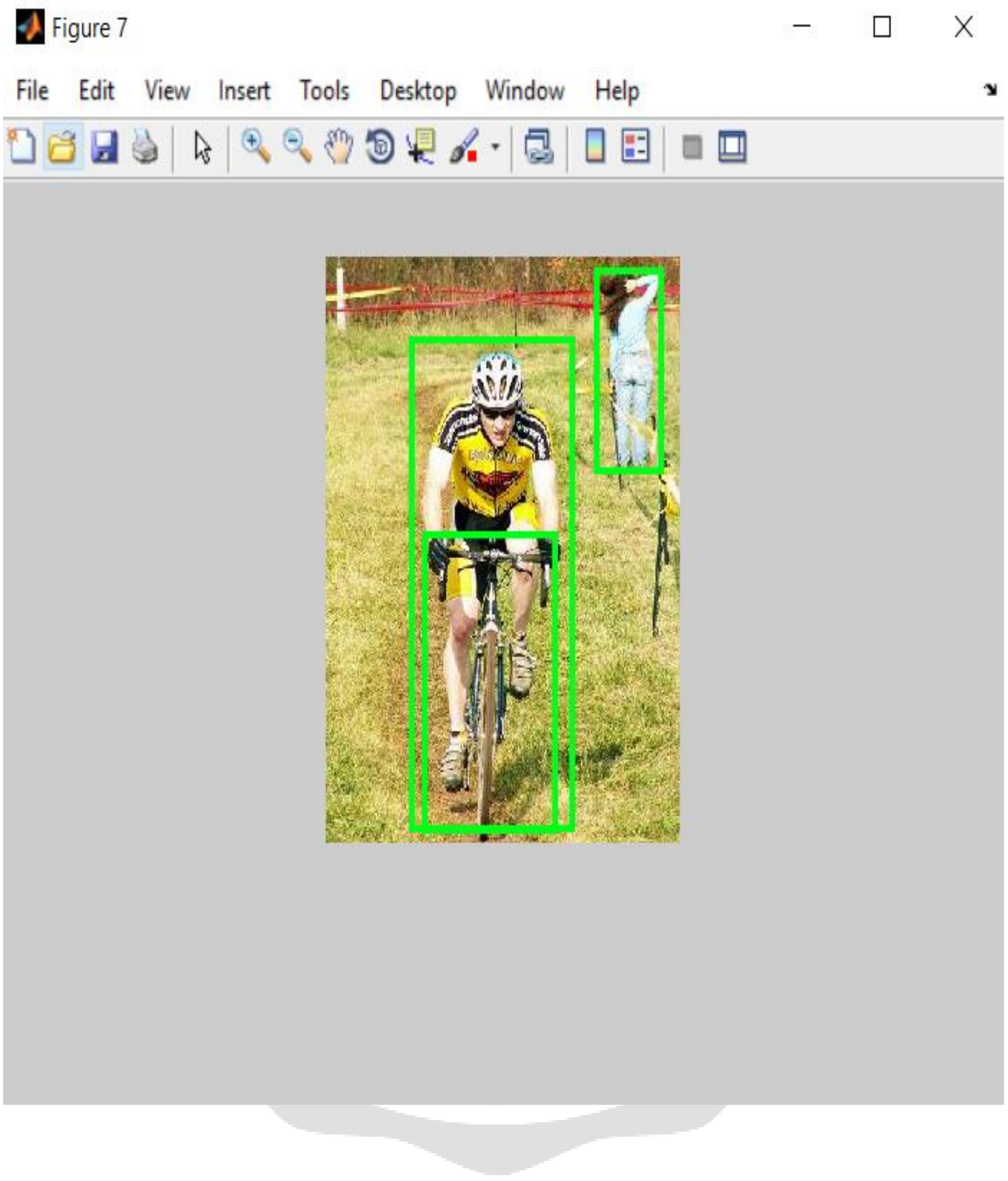


Fig: Detect multiple object from image using CNN and DEEP PYRAMID

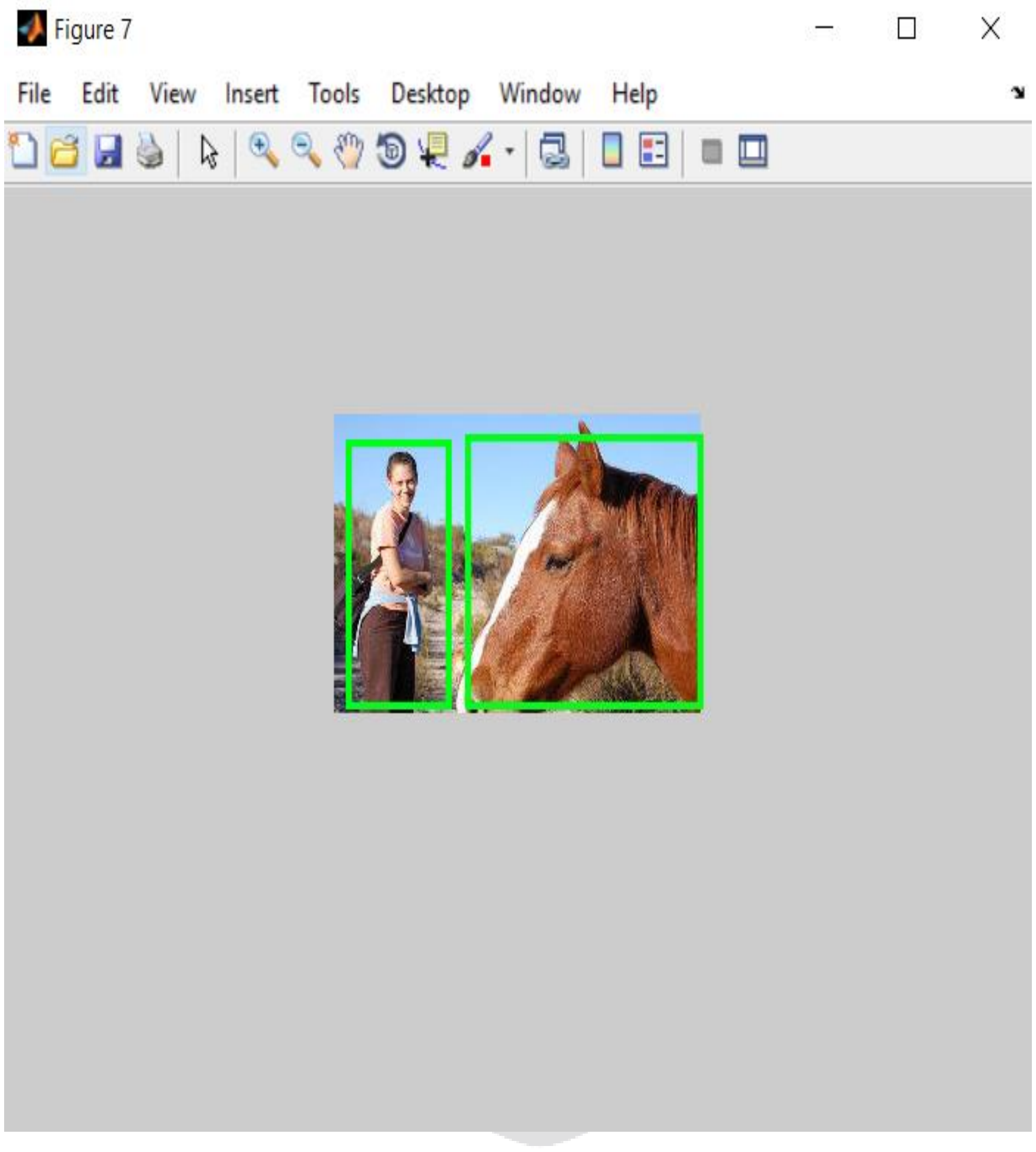
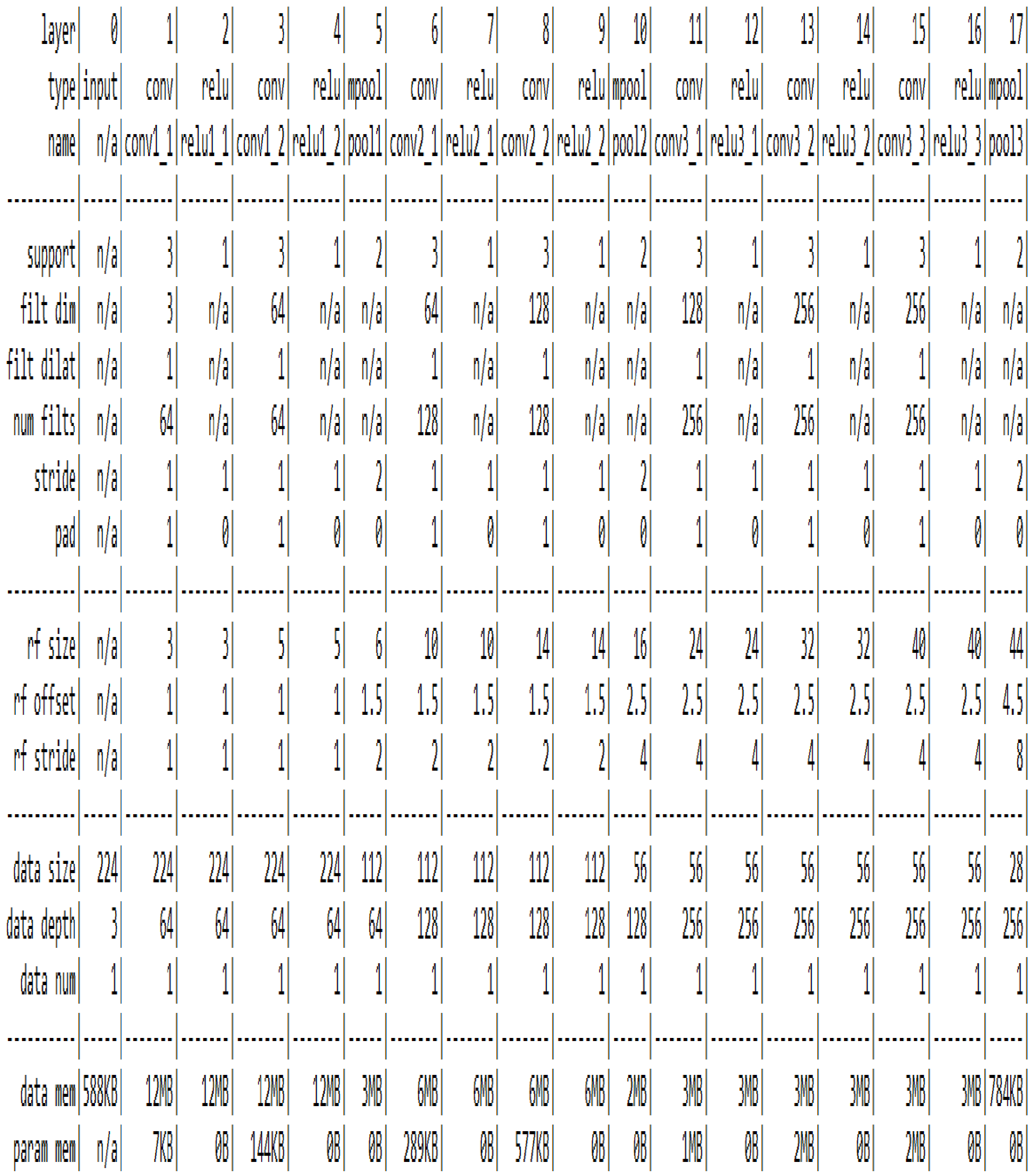
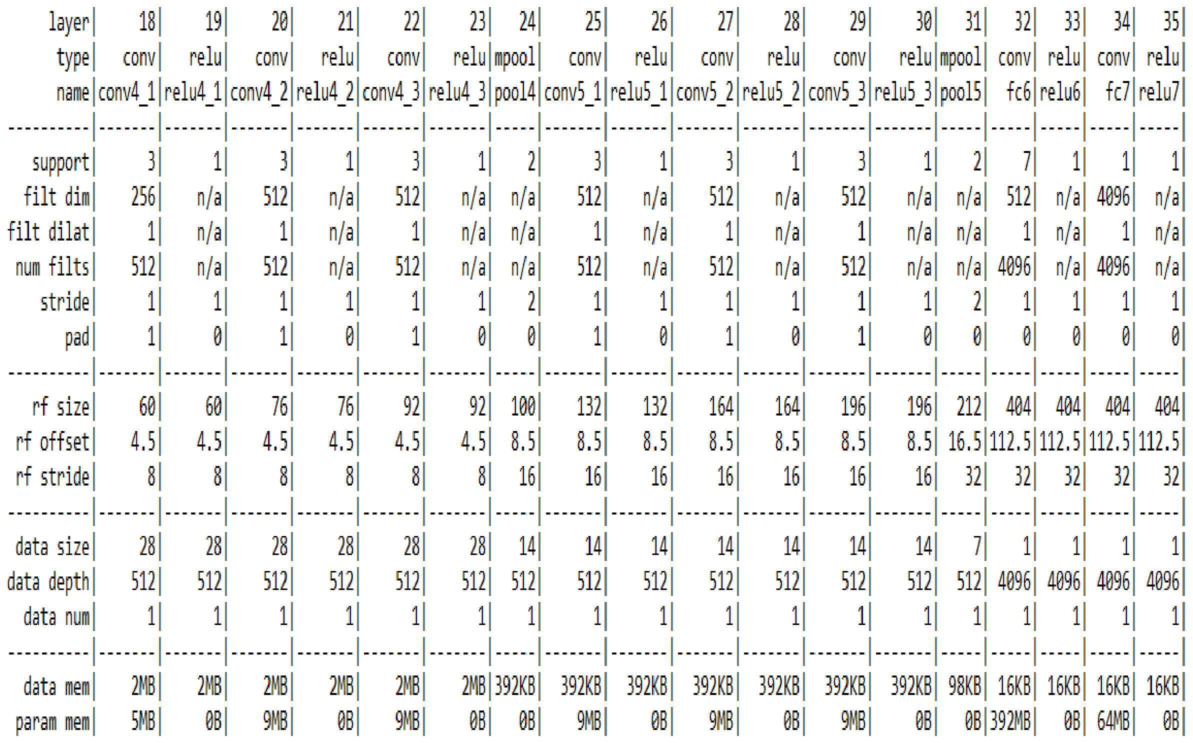


Fig: Detect multiple object and from image using CNN and DEEP PYRAMID

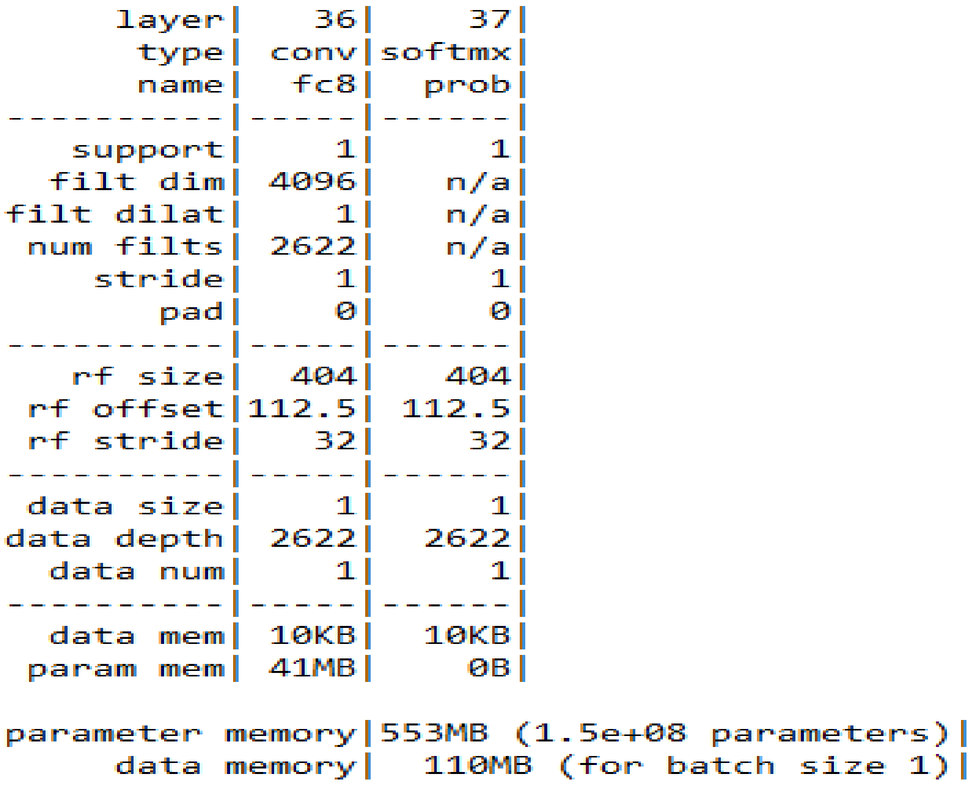
DATASET:



```
      layer|    0|      1|      2|      3|      4|    5|      6|      7|      8|      9|   10|     11|     12|     13|     14|     15|     16|   17|
       type|input|   conv|   relu|   conv|   relu|mpool|   conv|   relu|   conv|   relu|mpool|   conv|   relu|   conv|   relu|   conv|   relu|mpool|
       name|  n/a|conv1_1|relu1_1|conv1_2|relu1_2|pool1|conv2_1|relu2_1|conv2_2|relu2_2|pool2|conv3_1|relu3_1|conv3_2|relu3_2|conv3_3|relu3_3|pool3|
-----------|-----|-------|-------|-------|-------|-----|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|-------|-----|
    support|  n/a|      3|      1|      3|      1|    2|      3|      1|      3|      1|    2|      3|      1|      3|      1|      3|      1|    2|
   filt dim|  n/a|      3|    n/a|     64|    n/a|  n/a|     64|    n/a|    128|    n/a|  n/a|    128|    n/a|    256|    n/a|    256|    n/a|  n/a|
 filt dilat|  n/a|      1|    n/a|      1|    n/a|  n/a|      1|    n/a|      1|    n/a|  n/a|      1|    n/a|      1|    n/a|      1|    n/a|  n/a|
  num filts|  n/a|     64|    n/a|     64|    n/a|  n/a|    128|    n/a|    128|    n/a|  n/a|    256|    n/a|    256|    n/a|    256|    n/a|  n/a|
     stride|  n/a|      1|      1|      1|      1|    2|      1|      1|      1|      1|    2|      1|      1|      1|      1|      1|      1|    2|
        pad|  n/a|      1|      0|      1|      0|    0|      1|      0|      1|      0|    0|      1|      0|      1|      0|      1|      0|    0|
-----------|-----|-------|-------|-------|-------|-----|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|-------|-----|
    rf size|  n/a|      3|      3|      5|      5|    6|     10|     10|     14|     14|   16|     24|     24|     32|     32|     40|     40|   44|
  rf offset|  n/a|      1|      1|      1|      1|  1.5|    1.5|    1.5|    1.5|    1.5|  2.5|    2.5|    2.5|    2.5|    2.5|    2.5|    2.5|  4.5|
  rf stride|  n/a|      1|      1|      1|      1|    2|      2|      2|      2|      2|    4|      4|      4|      4|      4|      4|      4|    8|
-----------|-----|-------|-------|-------|-------|-----|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|-------|-----|
  data size|  224|    224|    224|    224|    224|  112|    112|    112|    112|    112|   56|     56|     56|     56|     56|     56|     56|   28|
 data depth|    3|     64|     64|     64|     64|   64|    128|    128|    128|    128|  128|    256|    256|    256|    256|    256|    256|  256|
   data num|    1|      1|      1|      1|      1|    1|      1|      1|      1|      1|    1|      1|      1|      1|      1|      1|      1|    1|
-----------|-----|-------|-------|-------|-------|-----|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|-------|-----|
   data mem|588KB|   12MB|   12MB|   12MB|   12MB|  3MB|    6MB|    6MB|    6MB|    6MB|  2MB|    3MB|    3MB|    3MB|    3MB|    3MB|    3MB|784KB|
  param mem|  n/a|    7KB|     0B|  144KB|     0B|   0B|  289KB|     0B|  577KB|     0B|   0B|    1MB|     0B|    2MB|     0B|    2MB|     0B|   0B|
```



```
     layer|     18|     19|     20|     21|     22|     23|   24|     25|     26|     27|     28|     29|     30|   31|   32|   33|   34|   35|
      type|   conv|   relu|   conv|   relu|   conv|   relu|mpool|   conv|   relu|   conv|   relu|   conv|   relu|mpool| conv| relu| conv| relu|
      name|conv4_1|relu4_1|conv4_2|relu4_2|conv4_3|relu4_3|pool4|conv5_1|relu5_1|conv5_2|relu5_2|conv5_3|relu5_3|pool5|  fc6|relu6|  fc7|relu7|
----------|-------|-------|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|-------|-----|-----|-----|-----|-----|
   support|      3|      1|      3|      1|      3|      1|    2|      3|      1|      3|      1|      3|      1|    2|    7|    1|    1|    1|
  filt dim|    256|    n/a|    512|    n/a|    512|    n/a|  n/a|    512|    n/a|    512|    n/a|    512|    n/a|  n/a|  512|  n/a| 4096|  n/a|
filt dilat|      1|    n/a|      1|    n/a|      1|    n/a|  n/a|      1|    n/a|      1|    n/a|      1|    n/a|  n/a|    1|  n/a|    1|  n/a|
 num filts|    512|    n/a|    512|    n/a|    512|    n/a|  n/a|    512|    n/a|    512|    n/a|    512|    n/a|  n/a| 4096|  n/a| 4096|  n/a|
    stride|      1|      1|      1|      1|      1|      1|    2|      1|      1|      1|      1|      1|      1|    2|    1|    1|    1|    1|
       pad|      1|      0|      1|      0|      1|      0|    0|      1|      0|      1|      0|      1|      0|    0|    0|    0|    0|    0|
----------|-------|-------|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|-------|-----|-----|-----|-----|-----|
   rf size|     60|     60|     76|     76|     92|     92|  100|    132|    132|    164|    164|    196|    196|  212|  404|  404|  404|  404|
 rf offset|    4.5|    4.5|    4.5|    4.5|    4.5|    4.5|  8.5|    8.5|    8.5|    8.5|    8.5|    8.5|    8.5| 16.5|112.5|112.5|112.5|112.5|
 rf stride|      8|      8|      8|      8|      8|      8|   16|     16|     16|     16|     16|     16|     16|   32|   32|   32|   32|   32|
----------|-------|-------|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|-------|-----|-----|-----|-----|-----|
 data size|     28|     28|     28|     28|     28|     28|   14|     14|     14|     14|     14|     14|     14|    7|    1|    1|    1|    1|
data depth|    512|    512|    512|    512|    512|    512|  512|    512|    512|    512|    512|    512|    512|  512| 4096| 4096| 4096| 4096|
  data num|      1|      1|      1|      1|      1|      1|    1|      1|      1|      1|      1|      1|      1|    1|    1|    1|    1|    1|
----------|-------|-------|-------|-------|-------|-------|-----|-------|-------|-------|-------|-------|-------|-----|-----|-----|-----|-----|
  data mem|    2MB|    2MB|    2MB|    2MB|    2MB|    2MB|392KB|  392KB|  392KB|  392KB|  392KB|  392KB|  392KB| 98KB| 16KB| 16KB| 16KB| 16KB|
 param mem|    5MB|     0B|    9MB|     0B|    9MB|     0B|   0B|    9MB|     0B|    9MB|     0B|    9MB|     0B|   0B|392MB|   0B| 64MB|   0B|
```



```
     layer|   36|    37|
      type| conv|softmx|
      name|  fc8|  prob|
----------|-----|------|
   support|    1|     1|
  filt dim| 4096|   n/a|
filt dilat|    1|   n/a|
 num filts| 2622|   n/a|
    stride|    1|     1|
       pad|    0|     0|
----------|-----|------|
   rf size|  404|   404|
 rf offset|112.5| 112.5|
 rf stride|   32|    32|
----------|-----|------|
 data size|    1|     1|
data depth| 2622|  2622|
  data num|    1|     1|
----------|-----|------|
  data mem| 10KB|  10KB|
 param mem| 41MB|    0B|

parameter memory|553MB (1.5e+08 parameters)|
     data memory|  110MB (for batch size 1)|
```

- ✓ Total no. of layers=8
- ✓ Subsampling layer=3
- ✓ Convolutional layer=3
- ✓ Fully connected layer=2

In this CNN, we used the **standard-MIT Database** , this database is open source.
To collect dataset of any image please click on following link.

http://host.robots.ox.ac.uk/pascal/VOC/databases.html#MIT

5. CONCLUSION:

In image processing many techniques are used but we used CNN algorithm for object detection. By, CNN the multiple objects are detected in image but it is not useful to detect very small and blur object so with the help of DEEP PYRAMID we can detect that small and blur object. In base paper used of classical tools from computer vision and deep learning in R-CNN technique. In this implementation predict that the DEEP PYRAMID used with CNN layers to generate the multiple shape and object of live things and also the detection of invisble object. Here first by using CNN, now detect the multiple object in image work used of DEEP PYRAMID with CNN the small and blur object also detected and the accuracy of the object is increased in % by filtering the object.In this last dissertation phase implementation of using DEEP PYRAMID to detect multiple object in multiple images.

6. REFFERENCES:

1) Piotr Dollar, Ron Apple, Serge Belongie, and Pietro Perona, “Fast Feature Pyramids for Object Detection”, IEEE transactions on pattern analysis and machine intelligence, vol 36,no.8 - August 2014.
2) Olga Barinova, Victor Lempitsky, and Pushmeet Kholi, “On Detection of Multiple Object Instances Using Hough Transforms”, IEEE transactions on pattern analysis and machine intelligence, vol 34,no.9 - September 2012.
3) Fahad Shahbaz Khan, Rao Muhammad Anwer, Joost van de Weijer, Andrew D. Bagdanov, Maria Vanrell, Antonio M. Lopez, “Color Attributes for Object Detection”, 978-1-4673-1228-8/12/$31.00 ©2012 IEEE - 2012.
4) S. M. All Eslami, Nicolas Heess, Christopher K. I. Williams, John Winn, “The Shape Boltzmann Machine: A Strong Model of Object Shape, Springer Science+Business Media New York - 2013.
5) Ross Girshick, Jeff Donahue, Trevor Darrell, Jitendra Malik, “Region-based Convolutional Networks for Accurate Object Detection and Segmentation”, DOI 10.1109/TPAMI.2015.2437384, IEEE Transactions on Pattern Analysis and Machine Intelligence- 2015.
6) O. Barinova, V. Lempitsky, E. Tretiak, and P. Kohli, “Geometric Image Parsing in Man-Made Environments,” Proc. 11th European Conf. Computer Vision, 2010.
7) LeCun, Yann, et al. “Gradient-based learning applied to document recognition.” Proceedings of the IEEE 86.11 (1998): 2278-2324.
8) Abineet Saxena, “ Convolutional Neural Networks: An Illustrates Explanation”, June-2016.