International Journal of Applied Engineering Research ISSN 0973-4562 Volume 13, Number 20 (2018) pp. 14657-14663


# Semantic Image Segmentation using Deep Convolutional Neural Networks and Super-Pixels

**Santosh Jangid**
*Research Scholar, JJT University, Jhunjhunu, Rajasthan, India.*

**P S Bhatnagar**
*Director, BKBIET, Pilani, Rajasthan, India.*

**Abstract**

In this paper Deep Convolution Neural Network (DCNN) with multiple layers are projected. Multiple layers work to build an improved feature space. First layer learns 1st order features (e.g. edge).Second layer learns higher order features (combinations of edges, etc.).Some models learn in an unsupervised mode and identify general characteristics of the input space. Final layer of transformed features is fed into supervised layers and an entire network is often subsequently tuned using supervised training using the initial weightings learned in the unsupervised phase. This deep convolution neural network gives a solution in the form of color segmentation. Deep CNN based segmenta-tion model shows 93% accuracy on BSDS300 dataset.

**Keywords:** Deep Convolution Neural Networ (DCNN); Recurrent Neural Network (RNN); Gernal Purpose Unit (GPU); Conditional Random Field (CRF); Recurrent neural network (RNN).

## INTRODUCTION

One primary advantage of convolutional neural networks is the use of shared weight of conventional layers. More or less time delay neural nets have also utilized a very similar structure to convolu-tional neural networks. Convolutional neural networks are utilized for image recognition and sorting tasks. This capability that the community is responsible for gaining knowledge of the filters that in usual algorithms has been hand engineered. The desire of a de-pendence on prior knowledge and the existence of tough to format hand engineered ingredients are a major benefit for CNNs.[2] Con-volutional neural networks are regularly used in photo conscious-ness systems. Some other report on using CNN for photo classifi-cation said that the studying procedure used to be "amazingly fast"; in the identical paper, the first-rate published effects at the time have been executed in the MNIST database and the NORB database. When applied to facial recognition, they were in a position to make a contribution to a massive limit on the error rate. In some other paper, they had been in a position to obtain a 97.6 percentage awareness charge on "5,600 still pictures of greater than 10 sub-jects". In the ILSVRC 2014, which is a large-scale visual focus challenge, nearly every enormously ranked crew used CNN as their simple framework [3-4]. In 2015 a many-layered CNN established the power to distinguish faces from a large assortment of angles, which include upside down, yet when in part occluded with com-petitive performance. The community trained on a database of 200,000 pictures that covered faces at a range of angles and orienta-tions and a in addition 20 million snap shots besides faces. They used batches of 128 photographs over 50,000 iterations. Convolu-tional neural networks with many layers have currently been exam-ined to obtain outstanding results on many high level duties such as image classification, segmentation and object detection.[1] Alexan-der G. Schwing, demonstrate their technique on the semantic pho-tograph segmentation undertaking and show encouraging results. They examine to pastime factor operators, a HOG detector, and three current works aiming at automated object segmentation. They sample a small variety windows in accordance to their objects chance and give an algorithm to rent them. This significantly reduc-es the routine of home windows evaluated by way of the expensive class specific model. They use objects as a complementary score in addition to the category specific model, which leads to fewer false positive. [5] Camille Couprie, practice a multiscale conventional community to study elements immediately from the photographs and the depth information. They acquire accuracy of 64.5% on the NYU-v2 dataset. [6] C.P. Town, demonstrates a strategy to content material primarily based image retrieval situated on the semantically meaningful labelling of pix by means of excessive level visual cate-gories. [7] Christian Szegedy, et.al used Deep Neural Network, that give a simple and yet effective element of target detection as a re-gression hassle to object bounding field masks on Pascal VOC. Clement Farabet et. Al used more than one post processing tech-niques to obtain the remaining labeling. Among those, they plump for a technique to mechanically retrieve, from a pool of segmenta-tion components, a top quality set of looks that best provide an explanation for the characterization. These aspects are arbitrary, e.g. they can be taken away from a partition tree, or from any household of over segmentations.[8] Clement Farabet, proposed scene parsing method here starts via

computing a tree of parts from a graph of pixel dissimilarities. The system produces image labels in less than 1 second on the Stanford Background dataset, Sift Flow Dataset and the Barcelona Dataset. [10] Hongsheng Li, perform forward and backward back propagation CNN for pixel wise classification of pictures such as image segmentation and object spotting. The proposed algorithms dispose redundant calculation in convolution and pooling by presenting D-regularly sparse kernels. Which oper-ate efficiently on GPUs and its computational complexity is regular with recognize to the number of patches sampled from the film. Experiments have proven that their proposed algorithms speed up. The modern main tactics for semantic segmentation take advantage of building statistics by extracting CNN elements from masked image regions. This method introduces synthetic boundaries on the pics and can also impact the niece of the extracted features. [11] Jifeng Dai, recommends a method to hold advantage of structure records by means of protecting conventional features. The concept segments are handled as masks on the conventional characteristic maps. [12] Jonathan Long, constructed "fully conventional" net-works that require input of arbitrary dimension and produce corre-spondingly sized output with environment friendly inference and learning. They used current classification networks (Alex Net, the VGG net, and GoogLeNet) into completely conventional networks and switch their realized representations via fine tuning to the seg-mentation. They achieve good segmentation results on PASCAL VOC, NYUDv2, and SIFT Flow. [13] Jose M. Alvarez, proposed an algorithm for convolutional neural networks to learn nearby facets from training books at different scales and resolutions. Its overall performance is comparable in contrast to the body politic of the art methods, the role of different sources of statistics such as depth and natural process.[14] Joseph J. Lim, proposed a novel method to each reading and detecting nearby contour based repre-sentations for mid- level features. The INRIA and PASCAL data-base were used to detect pedestrian and object detection. [15] Joao arreira, examine to rank the object hypotheses via coaching a con-tinuous model to predict how doable the segments are, given their mid-level area properties. They used VOC09 segmentation dataset and Deep Neural Network segmentation technique. [16] Liang Chieh Chen, proposed DCNN layer with a fully related Conditional Random Field (CRF) to localize phase boundaries. PASCAL VOC-2012 photograph segmentation task, hitting 71.6% IOU accu-racy in the checkers set. [17] Mohammadreza Mostajabi, introduce a simple feed forward structure for semantic segmentation. They achieved 64.4% accuracy on PASCAL VOC-2012 Semantic seg-mentation test set. A. Thakur propose a model based on color illu-mination and perform machine learning to predict correct forgeries. These models give us idea to prepare new algorithms for object segmentation and how to detect correct boundaries of the different objects [18-21].

## PROPOSED ALGORITHM

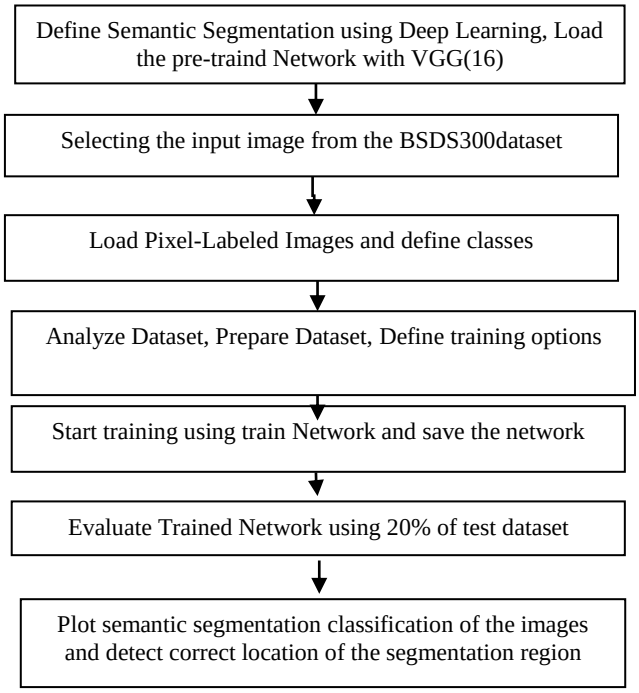


**Figure 1**: The proposed algorithm of Semantic Image Segmentation using Deep Convolutional Neural Networks.

In this Fig.1 first step is to load pre-trained semantic segmentation using Deep Learning model VGG (16) network. The dataset is prepared according to pre trained network. The BSDS300dataset is used for training and testing of the deep neural network. In this pre trained network load pixel labeled images and define classes. Ana-lyze dataset according to the pre trained network, prepare dataset for training and testing. For training 80% data is used whereas for test-ing 20% data is used. Define training options as follows: Training Options SGDM with properties: Momentum: 0.9000; Initial Learn Rate: 1.0000e-03; Learn Rate Schedule Settings: [1×1 struct]; L2Regularization: 5.0000e-04; Gradient Threshold Method: 'l2norm'; Gradient Threshold: Inf; Max Epochs: 100; Mini Batch Size: 2; Verbose: 1; Verbose Frequency: 2; Validation Data; Vali-dation Frequency: 50; Validation Patience: 5; Shuffle: 'every-epoch'; Checkpoint Path: ''; Execution Environment: 'multi-gpu'; Plots: 'training-progress'; Sequence Length: 'longest'; Sequence Padding Value: 0. Train the network with 80% data using train Network and save this network for testing. In testing 20% data is used. Plot the segmentation map after segmentation of the test images.

## BASIC OF NEURAL NETWORK

In single layer structure output price is estimated utilizing the weighted sum of input four variables. We can use any activation function to obtain more honest answers. By varying the weights the accuracy is increased.

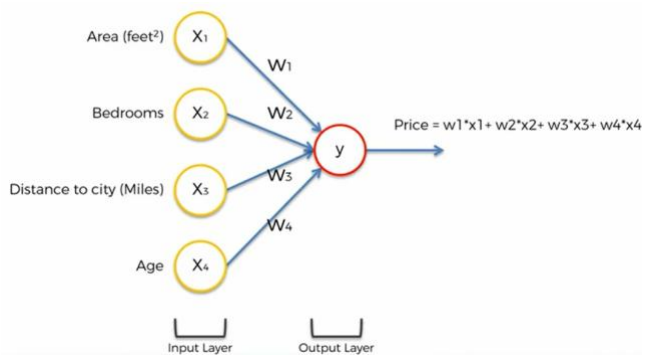


**Figure 2**: Block diagram of hidden layers in neural network

In this Fig. 2 input four variables are applied to first hidden layers of neural network. Top neuron of the hidden layer connected to all the input variables. All variables from input layers have synapsis weights are connecting each one of them to top hidden layer having different values. Some weigh have zero values and some weights have non zero value. Not all input will be valid for every single neuron. x1 and x3 are important for the neuron whereas x2 and x4 are not significant.

**Neural Network Learning Process**

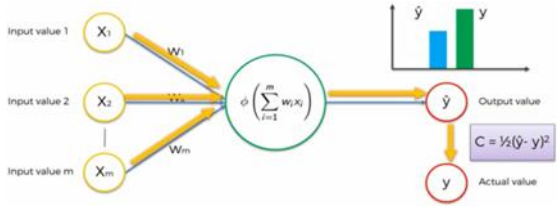


**Figure 3**: Block diagram of neural network learning process

Fig. 3 shows the block diagram of the neural network learning process. In this paper our goal is to create neural net-work, which learns on its own. Let us take an example of dog and cat classification. In machine language we have caused to define dog and cat features like expression, nose, eyes, ears, legs, color, and shape and apply conditions to render decisions. On the other hand, in neural network just code the neural network architecture and point the neural network at a folder of cat and dog which already categorize. Neural network configure on its own, which picture be-longed to which classes. When we give new image it rec-ognize from previous knowledge. The image depicts a sin-gle layer feed forward neural network and it is likewise called a perceptron. Perceptron was first devised in 1957 by frank Rosen blood. His idea was to create a network that learns itself. Input values are applied to perceptron and per-ceptron perform weighted sum using activation function. Output predicted value y' achieved after activation function and it is plotted in chart. Then we apply the cost function C= [1/2 (y'-y) ^2] which calculate the error rate in the pre-dicted output. We have to minimize the cost function so that the actual value is equal to predict output (y'=y).

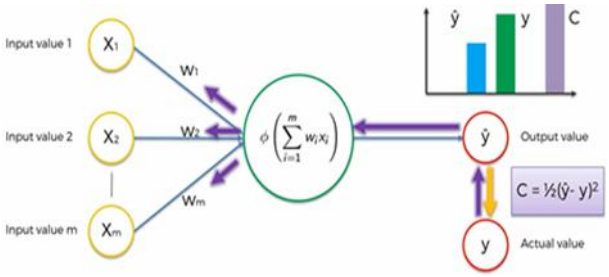


**Figure 4**: Compare both outputs and feed difference backward.

Fig.4 shows the training in both forward and backward direction. Now we compare both outputs and feed difference, backward to the neural network, which update the weights (w1, w2,..., Wm). We have only one row of input value (x1, x2,..., xm) applied to neural network which perform weighted sum and give predicted output. Again, we find cost function and compare the predicted output with actual output and this process continue till (y'=y) and C=0.

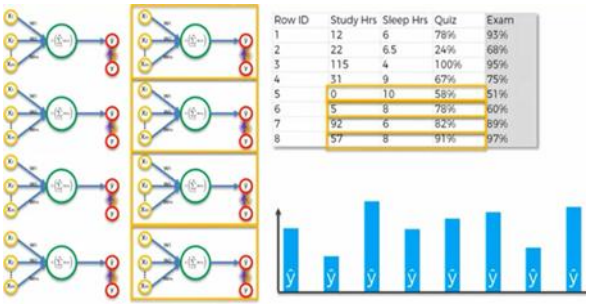


**Figure 5**: Multiple input operation

Fig. 5 shows the multiple input operation takes place at a same time and their cost function is calculated as y'. In multiple input row values we perform all the operation similar to the single row input and plot in chart for every course. We do a comparison between predicted (y') output with the actual output (y) using cost function C=sum [1/2 (y'-y) ^2]. We apply differences in backward direction and update weights (w1, w2, Wm). This process continues until the cost function achieves minimum value (y'=y; C=0).

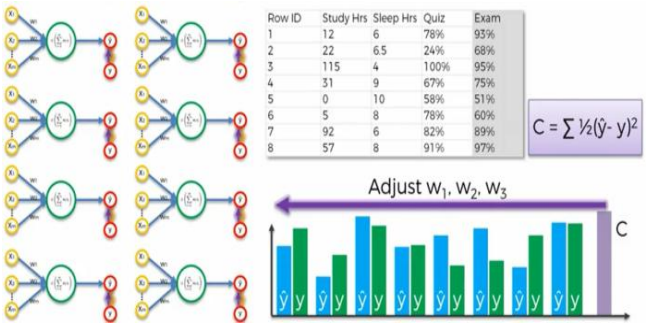


**Figure 6**: Adjust weights to achieve minimum cost function.

Fig. 6 shows that the weights are adjusted in such a manner that the error is reduced to the minimum label. The cost function is the sum of square difference between y' and y. The difference is, back propagated to the neural network and the weights are set accordingly. Fig. 6 shows the sim-ple model of neural network with single input, weight, one hidden layer having an activation function and cost function which calculate the difference between predicted and actual output. The fig. 6 shows the graph of different weights and find out which one is best. X axis represent the forecast output and Y axis represent cost function. The middle value of weight is the best for one input value. But if we have multiple inputs and having multiple numbers of weights we got the curse of dimensionality.

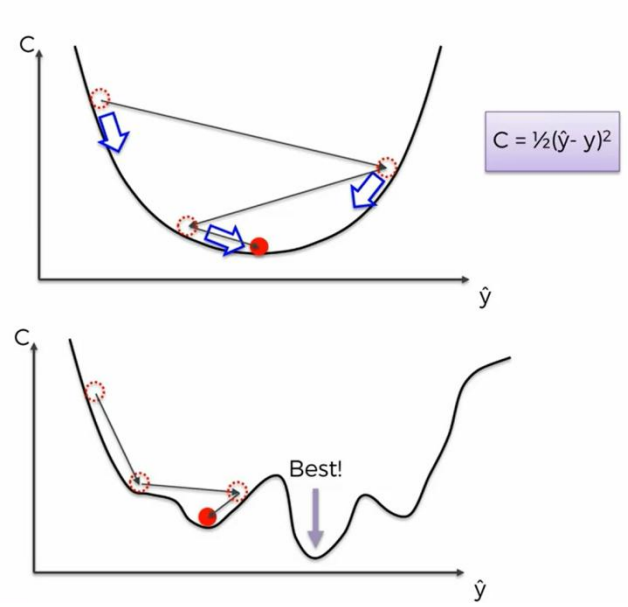


**Figure 7**: Gradient descent.

If neural network is more complex than gradient descent is given.Fig.7shows that the gradient descent operation to the weight on the left side of the graph. We expect at the slant of our cost function at that level. This is called gradient de-scent. In this manner positive and negative value of the slope is found .If the gradient is negative that means it works down toward right side. Again calculate the slope at middle of right side. This time slope is positive which means that it runs down to the odd side. In this way slope value for the best situation is calculated which minimized the cost part.

$$\text{Cost Function} = [1/2(y'-y)^2] \quad (1)$$

In the simple gradient descent only one global minimum is present. This method required convex function. But if sys-tem is not convex then stochastic gradient descent is ap-plied. In this method row1 is applied through neural net-work and update the weight. Take row2 pass through neu-ral network and update weight and so along. In this case weights are updated after each course evaluation. In the neural network there is a process called forward propaga-tion in which data is entered through input layer and then propagated forward to get predicted output. Then compari-son between predicted values with actual values is per-formed. Then error is calculated which is propagated back in the opposite direction and that allows training the net-work by adjusting the weights. Back propagation is able to adjust the weight simultaneously.

**Step 1:** Randomly initialize the weights to small number code to 0 (but not 0).

**Step 2:** Input the first observation of your dataset in the input layer, each feature in one input node.

**Step 3:** Forward propagation from left to right, the neurons are activated in a way that the impact of each neuron's activation is limited by the weights. Propagate the activations until getting the predicted result y.

**Step 4:** Compare the predicted result to the actual result. Eliminate the generated error.

**Step 5:** Back propagation from right to left and update the weights according to how much they are responsible for the error. The learning rate decides by how much we update the weights.

**Step 6:** Repeat step 1 to 5 and update the weights after each observation (Reinforcement Learning). Or: Repeat step 1 to 5 but update the weights only after a batch of observations (Batch Learning).

**Step 7:** When the whole training set passed through the ANN, that makes an epoch. Redo more epochs to get the better results.

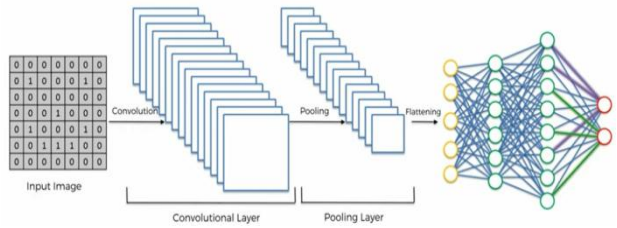


**Figure 8**: Artificial Neural Network training with Stochastic Gradent Descent

Deep learning tries to mimic the human mentality and stick kind of similar functions as the human mind. In our brain millions of neu-rons are present, which connects the brain to the organs, arms and legs and hence along. Human mind learns from experience or through prior observations. Weights are present across neurons in the brain and they represent long term memory that's why CNN and weights are present in the temporal lobe. Recurrent neural network (RNN) remembers things that just happened in the previous couple of observations and apply that knowledge in that going forward that's why we place RNN in the occipital lobe. Frontal lobes have a portion of the short term retention. The parietal lobe is responsible for sensation and perception and constructing a spatial coordination system to represent the world around us and to create a neural net-work which would fit into that category.

## RESULTS

In this paper image segmentation is performed using ma-chine learning and deep learning. The database is divided into three parts. The beginning part is training, second part is validation and the third part is image segmentation. In the first section we have created image segments from la-beled images and gather all icons into single folder named as train. The preparation is done with these segmented im-ages. In the second part segmented and non-segmented labeled images and applied to neural network. This step validates how to segment labeled images. In the third step testing is performed on collected sample of images. In this step neural network automatically learn to segment images. This paper show good accuracy and achieve good segmen-tation results. In the Fig. 9 we can discover how our pro-posed algorithm train neural network to produce better segmentation results.

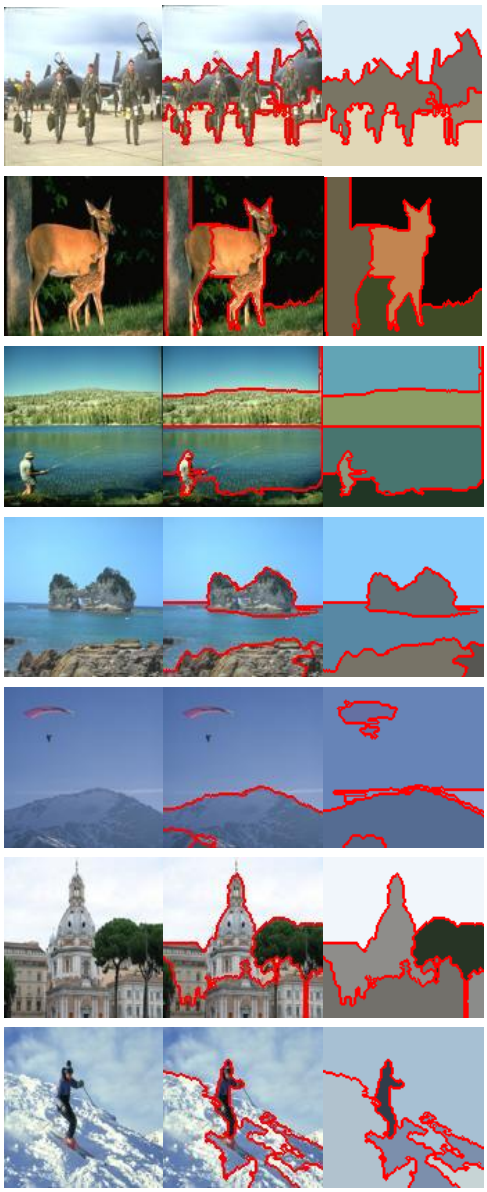

**Figure 9**: Image segmentation results.

Fig. 9 results shows the segmentation of the objects. These results are collected in one folder and applied to deep learning semantic segmentation model which train and test the color segment of the images. N the testing phase the learned pattern of the colors deep learning based semantic segmentation model predict the color pixel label.

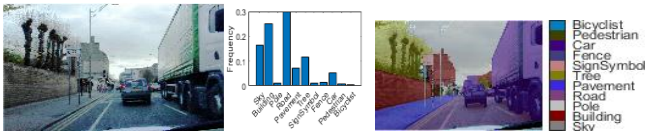


**Figure** 10: a) Original image. b) Graph of automated image segmenta-tion. c) Image Segmentation with Labels

Fig.10 (a) shows the output of original image. Fig.10 (b) shows the graph of automated image segmentation. Fig.10(c) shows the image Segmentation with Labels. These results are obtained after testing the network. The deep segmentation model predict the color labels. We achieved detection accuracy of 93%.

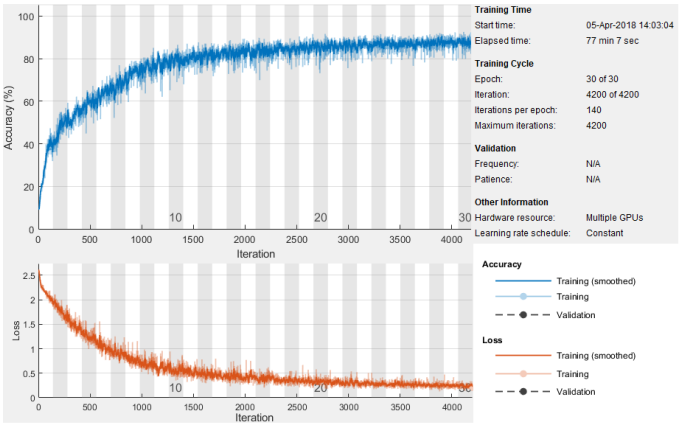


**Figure 11**: Accuracy vs Loss plot. X axis represent number of itera-tion, Y axis represent Accuracy/Loss.

Fig.11 shows the Accuracy vs Loss plot. X axis represent number of iteration, Y axis represent Accuracy/Loss. In the training phase 30 epoch are used. In each epoch 140 iterations takes place. Total 4200 iterations are performed. We have used i7 processor, 16 GB ram, Nvidia 1070 GPUs, SSD for data storage, win10 64 bit, Matlab 18, with Neural network toolbox, Computer vision toolbox, Deep neural network, vggNet toolbox and parallel computing toolbox.

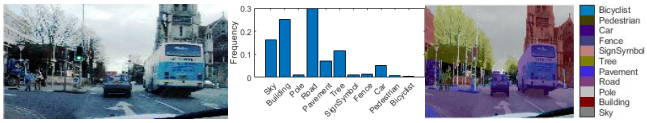


**Figure 12:** a) Original image b) Graph of automated image segmentation. c) Image Segmentation with Labels.

Fig.12 (a) shows the output of original image. Fig.12 (b) shows the graph of automated image segmentation. Fig.12(c) shows the image Segmentation with Labels.

### Major Challenges and Failings

Realistically, it is unclear how nicely the top algorithms work on widespread imagery. It much takes place that the satisfactory meth-ods for a dataset are fine tuned for solely the imagery of a particular office, place or context, so the generality is unclear. Hence, this is clearly one of the major future challenges for the research communi-ty.

How a lot data are critical to teach the algorithm? Some of the great-est approaches require considerable amounts of labeled information. This potential that in some situations, those algorithms will be un-suitable because the labeled datasets are unavailable. For scene clas-sification, the credible datasets generally comprise thousands and thousands of thousands of millions of coaching images; all the same, for most applications the coaching set size is greater probabil-ity to be in the thousands. If the area specialists discovers it tough or not possible to create very giant education sets, then is it viable to design deep getting to know algorithms which require fewer exam-ples?

How a good deal computational assets are involved? More or less of the top methods require quite heavy usage of near super comput-ers for the education phase, which may additionally no longer be used in all settings. Many researchers are therefore looking at the question: For a unique range of parameters, what is the exceptional accuracy that can be achieved?

When will the strategies fail? Achieving higher accuracy is good, but it is critical to possess an understanding of the ramifications of improper segmentations. In some conditions such as driving a car in a city, it is now not challenging to stumble upon segmentation prob-lems that hold been no longer included by means of the education dataset. Having an extremely correct picture segmentation would be really good.

Economy, the following functional strategies are generally developed by the road transport enterprise: the marketing strategy, the financial strategy, the quality strategy, the manufacturing strategy, the social strategy, the strategy of technological and organizational change, the environmental strategy [12].

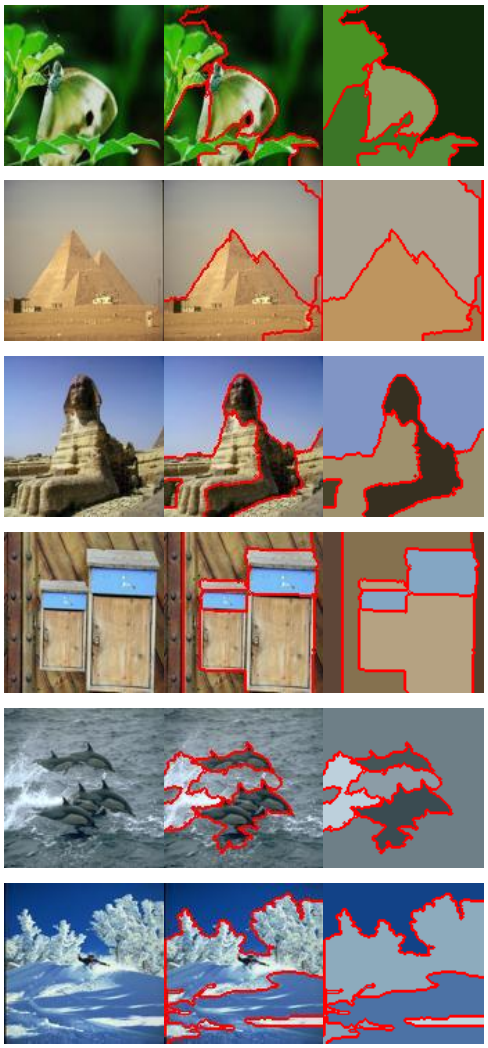

**Figure 13**: a) Original image. B) Example of automated im-age segmentation

For instance, as shown in Fig.13the segmentation has difficulties with the audience members and additionally the objects in the fore-ground. In some cases, the semantic segmentation extends beyond the object boundaries. In the ordinary case, this power that using segmentations also requires perception the effect that the blunders will have on the whole organization.

## CONCLUSION AND FUTURE WORK

Image segmentation has made big improvements in later age. Recent work primarily based on deep gaining knowledge of the techniques which has resulted in groundbreaking improvements in the accuracy of the segmentations currently reported over 93% on the BSDS300 dataset. Because image segmentations are a midlevel representation, they accept the plausible to make major contributions throughout the wide area of visual perception from image classification to picture synthesis; from object attention to object modeling from high per-formance indexing to relevance feedback and interactive search. The Deep Convolution Neural Network can learn the hierarchy of fea-tures (low to mid to high level). Deeper means better feature ab-straction need to regularize the model well and finely tuned network can reduce the domain mismatch.

## REFERENCES

[1] Alexander G. Schwing and Raquel Urtasun (2015), Fully Connected Deep Structured Networks, arXiv:1503.02351v1

[2] Anastasios Doulamis, Nikolaos Doulamis, KlimisNtalianis, and Stefanos Kollias (2003), An Efficient Fully Unsupervised Video Object Segmentation Scheme Using an Adaptive Neural Network Classifier Architecture, IEEE Transactions On Neural Networks, 14(3)

[3] Bharath Hariharan, Pablo Arbel´aez, Ross Girshick, and Jitendra Malik (2014), Simultaneous Detection and Segmentation, arXiv:1407.1808v1.

[4] Bogdan Alexe, Thomas Deselaers, and Vittorio Ferrari (2012), Measuring the Objectness of Image Windows, IEEE Transactions on Pattern Analysis and Machine Intelligence,34(11).

[5] Camille Couprie, Clement Farabet, Laurent Najman and Yann LeCun (2013), Indoor Semantic Segmentation using depth information, arXiv:1301.3572v2.

[6] C.P. Town and D. Sinclair (2001), Content Based Image Retrieval using Semantic Visual Categories, CiteSeer.

[7] Christian Szegedy Alexander ToshevDumitru Erhan (2013), Deep Neural Networks for Object Detection, Advances in Neural Information Processing Systems 26 (NIPS 2013).

[8] Clement Farabet, Camille Couprie, Laurent Najman, Yann Lecun (2013), Learning Hierarchical Features for Scene Labeling, IEEE Transactions on Pattern Analysis and Machine Intelligence, Institute of Electrical and Electronics Engineers (IEEE), 35 (8), 1915–1929.

[9] Clément Farabet, Camille Couprie, Laurent Najman and Yann LeCun(2012), Scene Parsing with Multiscale Feature Learning, Purity Trees, and Optimal Covers, arXiv:1202.2160v2.

[10] Hongsheng Li, Rui Zhao, and XiaogangWang(2014), Highly Efficient Forward and Backward Propagation of Convolutional Neural Networks for Pixelwise Classification, arXiv:1412.4526v2.

[11] Jifeng Dai, Kaiming He and Jian Sun (2015), Convolutional Feature Masking for Joint Object and Stuff Segmentation, arXiv:1412.1283v4

[12] Jonathan Long, Evan Shelhamer and Trevor Darrell (2015), Fully Convolutional Networks for Semantic Segmentation, arXiv:1411.4038v2

[13] Jose M. Alvarez, Yann LeCun, TheoGevers and Antonio M. Lopez (2012), Semantic Road Segmentation via Multi-scale Ensembles of Learned Features, ECCV 2012 Ws/Demos, Part II, LNCS 7584, 586-595

[14] Joseph J. Lim C. Lawrence Zitnick Piotr Doll´ar (2013), Sketch Tokens: A Learned Mid-level Representation for Contour and Object Detection, IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[15] Joao carreira and cristiansminchisescu (2010), constrained parametric min-cuts for automatic object segmentation, IEEE Transactions on Pattern Analysis and Machine IntelliGence.

[16] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy and Alan L. Yuille (2015), Semantic image segmentation with deep convolutional nets and fully connected crfs, ICLR.

[17] MohammadrezaMostajabi, PaymanYadollahpour and Gregory Shakhnarovich (2014), Feedforward semantic segmentation with zoom-out features, arXiv:1412.0774v1

[18] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, Jia Li, Salient Object Detection: A Benchmark, IEEE TIP, 2015.

[19] Ali Borji, Ming-Ming Cheng, Huaizu Jiang, Jia Li, Salient Object Detection: A Survey,arXiveprint, 2014

[20] Ming-Ming Cheng, Niloy J. Mitra, Xiaolei Huang, Philip H. S. Torr, Shi-Min Hu. Global Contrast based Salient Region Detection. IEEE TPAMI, 2015

[21] Thakur, A., Jindal, N. Multimed Tools Appl (2018). https://doi.org/10.1007/s11042-018-5836-5.