

Image Segmentation Using Convolutional Neural Network

Ravi Kaushik, Shailender Kumar

Abstract: Identifying regions in an image and labeling them to different classes is called image segmentation. Automatic image segmentation has been one of the major research areas, which is in trend nowadays. Every other day a new model is being discovered to do better image segmentation for the task of computer vision. Computer vision is making human's life easier by automating the tasks which humans used to do manually. In this survey we are comparing various image segmentation techniques and after comparing them with each other we have explained the merits and demerits of each technique in detail. Detailed analysis of each methodology is done on the basis of various parameters, which are used to provide a comparison among different methods discussed in our work. Our focus is on the techniques which can be optimized and made better than the one which are present before. This survey emphasizes on the importance of applications of image segmentation techniques and to make them more useful for the mankind in daily life. It will enable to us to take full benefits of this technology in monitoring of the time consuming repetitive activities occurring around, as doing such tasks manually can become cumbersome and also increases the possibility of errors.

Index Terms: Computer Vision, Convolution Neural Networks, Deep Learning, Edge Detection Models, Fully Connected Layer, Image Segmentation, Max Pooling.

1. INTRODUCTION

MACHINE learning is the most ideal skill of this digital age. As we dissect the process how a machine learns to classify and obtains the inputs or the raw materials needed for learning the specifics of the desired task. Features or attributes forms the basis of what we feed in the learning algorithm. In the task of image processing and object identification, machine learning plays a vital role. There are many techniques available to do such tasks. We will discuss various methods that can be used to achieve such tasks. In this world of digitization, images play a very important role in various areas of life including scientific computing and visual persuasion tasks etc. Technically images can be binary images, grey scale images, RGB images, hue saturation value or hue saturated lightness images. Each data record can be represented via a huge number of features. But all features are not necessarily significant for analysis or classification. Thus, feature selection and feature extraction are significant research areas. In image segmentation each pixel in image is labeled to different class. This pixel labeling task is also called as dense prediction. Suppose in an image there are various objects available like cars, trees, signals, animals. So, image segmentation will classify all the trees as a single class, all animals and signals to their respective classes. One important thing to consider in image segmentation is that it considers two objects of the same type as a single class. We can differentiate objects of the same type using instance segmentation. CNN is used very frequently for segmenting the image in pattern recognition and object identification. Here, we have discussed some of the real life applications of CNN. Like work done by Olaf Ronneberger to segment the neural structures in microscopic images of human blood [1]. A major work is done by Sadesi Karimpouliet [2] in identifying the different types of rocks present in the images of rocky area. Next, we have discussed

the work done in the field of remote sensing image segmentation by Xiomeng Fu Et al. [3]. Image segmentation is most applicable in medical field, as proper analysis of MRI or X-rays images is very crucial in proper diagnosis. Pim moeskps Et al. [4] has discussed one such application of segmenting the MRI images of brain, breast and cardiac CTA. A similar application is implemented by ZhuolingLi Et al. [5] to detect the deformities in eye cornea images and to identify any object in lungs x-ray images.

Our survey emphasizes on comparing many of such implemented models of CNN in various applications and to produce results which are used to tell which model is best suitable for different applications. Here, we are analyzing each model thoroughly on the basis of efficiency, applications, area of use, size of training data required and many other parameters. In this survey, section 1 describes the introduction part of this research, which explains the motivation behind doing this work and the flow of work explained here. Second section of this survey describes some concepts of image segmentation from the scratch so that reader can get some idea of how this technology actually work and can become useful in identifying the crucial details while reading the survey. Then in section 3, the actual survey related work starts and here a detailed analysis of their work and methodology is explained. Section 4 contains the comparison on the basis of common parameters. Section 5 contains the brief analysis and conclusion of all work discussed here. In the last, section 6 discusses about the future scope of work that is still possible in enhancing the performance and expansion of application areas in the field of image segmentation using CNN.

2 BACKGROUND

In this section, we are going to discuss about the techniques and knowledge required prior to understand the research work we have discussed here in further sections. Here, basics of image segmentation, its usefulness and architecture are being discussed for the better understanding of readers.

2.1 NEED OF IMAGE SEGMENTATION

Suppose, you are crossing a road and you see various objects around like vehicles, traffic lights, footpath, zebra-crossing and pedestrians. Now, while crossing the road your eyes instantly analyze each object and process their locations to take

- Ravi kaushik is currently pursuing masters degree program in Software engineering at Delhi Technological University, India.
E-mail: ravikaushik1095@gmail.com
- Shailender Kumar is currently working as an Associate professor at Delhi Technological University, India.
E-mail: shailenderkumar@dce.ac.in

decision of whether to cross the road at that moment or not. Can computers do this task? The answer to this question was 'No' since a long time before the breakthrough inventions in computer vision and object identification. But now with the help of image segmentation and object identification techniques it is possible for the computers to see the real-world objects and based on their positions they can now take necessary actions. Also, in the medical field, image segmentation plays a very important role. For example, we can consider the task of identifying cancer cells in the microscopic image of human blood. If you ask of identifying cancer cells, it's crucial to identify the shape of cells in blood and any unusual growth in the shape of blood cells will be diagnosed for the presence of cancer cells. This makes the recognition of cancer in blood at an early stage so that it can be cured within time.

2.2 ROLE OF DEEP LEARNING IN IMAGE SEGMENTATION

As all other machine learning techniques work in a way of taking the labelled training data and based on the training, they process new input and take the decisions. But as opposed to this deep learning works with neural networks and can make the conclusions of its own without the need of labeled training data. This method is useful for a self-driving car, so that it can differentiate between a sign board and a pedestrian. Neural networks use algorithms which are present in the network side by side and output of one algorithm is subject to the outcome of another algorithm. This creates a system which can think as a human being for taking the decisions. And this makes a model which is a perfect example of an artificial intelligence system.

2.3 COMPARISON OF VARIOUS IMAGE SEGMENTATION TECHNIQUES

2.3.1 Region Based image segmentation

It is the simplest way of segmenting an image based on its pixel values. This makes use of the fact that there will be a huge difference in pixel values at the edges of an object from the pixel values of its background pixels. So, in such cases we can set a threshold value, the difference in pixel values falling below the threshold can be segmented accordingly. In case there is one object in front of a single background only one threshold value will work. But in case there are multiple object or overlapping among objects we might need multiple threshold values. This technique is also called as threshold segmentation.

2.3.2 Edge Based Segmentation

In an image, what divides an object with its background or other object is an edge. So, this technique makes use of this fact that whenever a new edge is encountered while scanning an image, we are detecting a new object. This detection of edge is done with the help of filters and convolutions. Here, we have a filter matrix which we need to run over the image to detect the specified type of edge for which the filter was made.

2.3.3 Clustering based image segmentation

This technique work in the way that it divides the data points in an image in some clusters having similar pixel values. Based on the number of objects that might present in the image the value of number of clusters is selected. Then during the clustering process similar kind of data points were classified

into a single cluster. In the end image gets classified into several regions. Although this is a time taking process but it provides accurate results on small datasets.

2.3.4 CNN based image segmentation

This method is currently the state of art technique in image segmentation field of research. It works on images which are of 3 dimensions i.e. height, width and number of channels. First two dimensions tell us the image resolution and third dimension represents the number of channels (RGB) or intensity values for red, green and blue colors. Usually images which are fed into the neural network are reduced in dimensions which reduce the processing time and avoid the problem of under fitting. Even though if we take an image of size $224 \times 224 \times 3$ which when converted in to 1 dimension will make an input vector of 150528. So, this input vector is still too large to be fed as input to the neural network.

TABLE 1
Comparison of Image Segmentation Techniques

Advantages			
Region Based	Edge Based	Cluster Based	CNN Based
Simple calculations.	Good for images having better contrast between objects.	Works really well on small datasets and generates excellent clusters.	Provides most accurate results.
Fast operation speed.			
Disadvantages			
Region Based	Edge Based	Cluster Based	CNN Based
If there is no significant grayscale difference or an overlap of the gray scale pixel values, it becomes difficult to get accurate segments.	Not suitable when there are too many edges in the image.	Computation time is too large and expensive.	Takes comparably longer time to train the model.

2.4 DEFINITION OF CNN

CNN stands for Convolutional Neural Network, is used for the classification of images. The input images which are used in CNN are of 3 dimensions i.e. height, width and number of channels. First two dimensions tell us the image resolution. And third dimension represents the number of channels (RGB) or intensity values for red, green and blue colors. Usually images which are fed into the neural network are reduced in dimensions which reduce the processing time and avoid the problem of under fitting. Even though if we take an image of size $224 \times 224 \times 3$ which when converted in to 1 dimension will make an input vector of 150528. So, this input vector is still too large to be fed as input to the neural network.

2.5 WORK FLOW DIAGRAM OF CNN

This section discusses the design structure of CNN based on data and the optimization techniques that can help most of the CNNs with low dependency on the structure. We are actually going to see the techniques in a reverse manner similar to the flow of a back propagation to demonstrate the learning cycle of a CNN.

2.5.1 Convolution Layer

A filter or kernel is used to run over the image in fixed gap intervals called strides. Selecting the size of stride is crucial to achieve desired results. During running the filter over the image, dot product of filter with part of image on which filter lies is calculated. Then sum of all values of product matrix is copied to the corresponding position in convolved feature map matrix. Thus, we get a reduced dimension feature map of image. Filters may be of many kinds, where each filter is used to extract different kind of feature from the image. For e.g. one filter may be responsible for extracting one kind of feature from the image based on shapes and edges and another filter may be used to extract features based on color intensities. Parameters which helps in adjusting CNN's performance: Stride: This defines the number of pixels by which we have to move our filter over the image so that we can focus on a new set of pixels while doing convolution. Stride's value ranges from 1 to 3 depending upon the amount of loss which we can be accommodated during convolution. The amount of loss in image increases with the increasing value of stride. Selecting optimal value of stride is a very crucial task. The accuracy of model highly depends upon this value. Padding: It is a process of adding zeroes around the border of original image symmetrically. This helps us obtaining the feature map output to be of size as per our requirement. Commonly it is used to preserve the dimension of image after convolution. Filters: These are also called kernels. These may be of many types. Some of the types are:

- Sobel filter (horizontal) is used to detect horizontal edges in the image.
- Sobel Filter (vertical) filter is used to detect vertical edges in the image.
- Laplace Filter (both horizontal and vertical) is used to detect horizontal edges as well as vertical edges in the image.
- Blurring filter is used to blur out an image.
- Sharpening filter is used to sharpen an image.

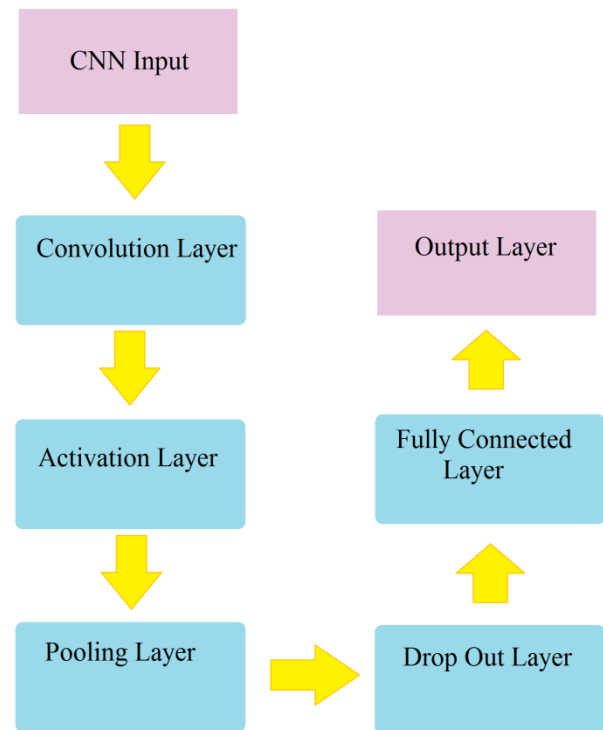


Fig. 1 - Different Layers of Convolutional Neural Network

2.5.2 Pooling Layer

This layer operates a small kernel on the image at a fixed stride. It is used to pick the pixel with the highest intensity and discard other pixels. The resultant matrix will be a reduced dimensional matrix of feature image. This helps in reducing the unnecessary sparse cells of image which are of no use in classification. Max pooling helps reduce the dimensionality of the network (or image), but this may cause some information loss. The concept behind this is that adjacent or nearby pixels can be approximated by the maximum information carrying pixel.

2.5.3 Activation Layer

This layer mostly uses ReLu as an activation function. ReLu is a function which is used to set all negative values to zero and keeps positive value as it is. This step is usually followed by convolution and pooling layer. In numerous fields such as Computer Vision they are becoming the state of art achieving near human or better performance. They sound fascinating but designing a CNN is a herculean task in itself. Till now there is no fixed formula for the design of CNN. Many researchers have come up with the general suggestions such as. But they don't always hold, as the task is dependent as importantly on data as it does on the algorithm. CNN exploits the spatial hierarchical features of data, extracting features and help classify them into different classes. This has led to development to a stream of data augmentation and pre-processing to increase the data, as more data allows for chance of better training and avoiding over fitting. This helps build models that are more robust to new samples as we try to make it more generalized to noise at the training phase.

2.5.4 Drop Out Layer

Dropout Layer is usually applied after the layer containing neurons in the fully connected network. Dropout layer is a regularization layer. With each iteration, it randomly drops out the weights of some input neurons such that the other neurons get more weightage in next iteration. Usually people take dropouts of 20-50%. By using dropout layer, normally the performance of a network increases 1 to 2 %.

2.5.5 Fully Connected Layer

The structure from top to down usually forms a pyramid structure, the number of parameters in these layers keep on converging till they finally reach the number of desired classes. Increasing the number of hidden units in the layer can increase the learning ability of the network, but there is a saturation of the increase in accuracy of the network. There is

no formulation of the units you choose as it is a hit and trial usually. These also pile up with the depth of the fully connected subset of the network. Most of the networks in research usually perform well with number of units in multiple of 64. Two to three layers networks are good if there are enough patterns being passed to the network after flattening the outputs of the Convolutional layers.

2.6 ARCHITECTURE OF CNN

Here, the internal working architecture of CNN is displayed in the figure. The image is represented by a matrix of pixel values and convolution map is obtained by sliding and multiplying filter matrices over the input image.

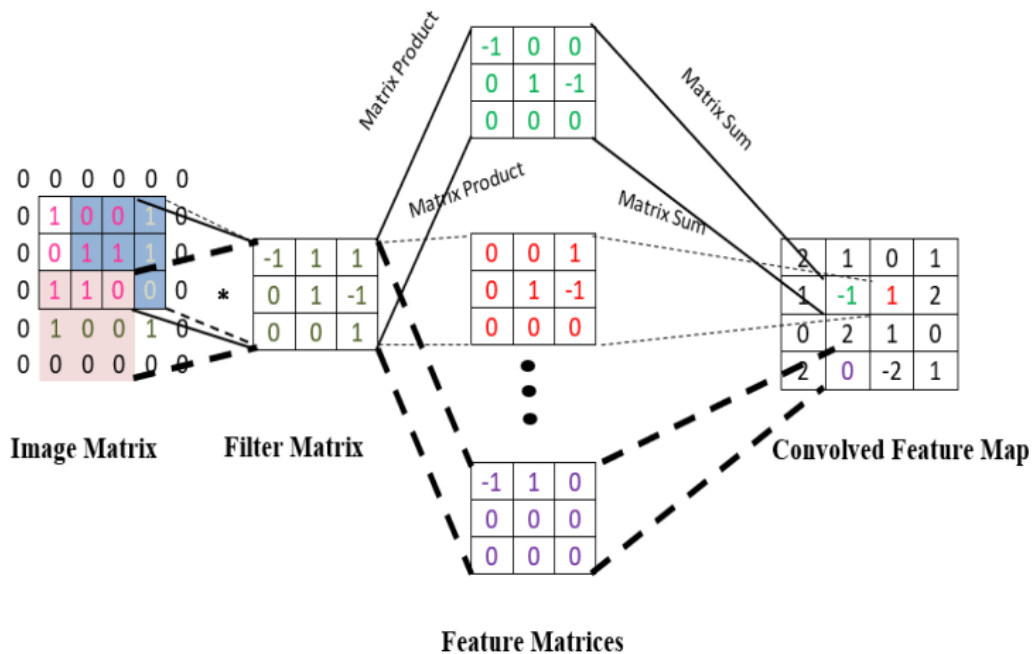


Fig. 2 - CNN Architecture

3 LITERATURE SURVEY

Here, we are going to discuss the research work already done by several authors in the field of image segmentation using CNN. First of all, we'll start with the work done by Olaf Ronneberger Et al. [1]. They have done crucial work on biomedical images dataset. They segmented the neural structures found in microscopic images of human blood samples. Cell growth tracking in a human body is performed automatically using this approach. The typical task of CNN is classification, where output to single image is a class label. But in many other applications, localization information should be included in output. That is each pixel in an image need to be assigned to a class label. In their work they have used CNN to segment the image to identify different cell types and to track the growth of some specific cells in the body over a period of time. Along with identifying the cells in the body, the location of those cells is also important to monitor the cell growth properly. They have applied segmentation task on a raw image which classify objects with different colors then a black & white segmentation mask is generated using white for foreground and black for background. Then the image is

mapped with pixel wise loss weight to identify the border pixels. So, for that reason ceresin et al predicted the class label of pixels in images by providing a region which is local to the pixel in that image and with the help of a sliding window setup to train the network. Accuracy is 77.5% and training time is 10 hours [1]. Next, we are going to discuss the work done by Sadegh Karimpouli Et al. [2]. Here they used CNN to segment the rock images. Segmentation in rock images is a critical step as images presented here are in grey-scale format. Conventional methods of segmentation such as thresholding or watershed algorithm don't work accurately with digital rock images data. As these methods work on color contrast to segment the regions with each other, the problem arises when two phases of similar color and contrast found to be side by side, then these algorithms consider them as a single phase. They have analyzed several machine learning approaches which could be used to solve the problem but Convolutional auto encoder was found to be producing more accurate results than others. Here, a relatively very small dataset is used to train the model, but the data augmentation

method helps to increase the size of training dataset. By creating a training seed which consists of manually and semi manually segmented images. They took the help of a data set of images from Berea sandstone. They divided the dataset among groups of training, testing and validation. From which only 10% of the data was required and used for validation purpose and another 10% was used for the testing purpose and rest of remaining data i.e. 80% was for used to train the model accurately [2]. Finally, this dataset is given to stochastic image generator which makes the size of dataset bigger to train the model more effectively. Then CNN model was implemented which provided results with an accuracy of 96%. Segmentation on the images of High resolution remote sensing data is done by Xiaomeng Fu Et al.[3] with the help of fully convolutional neural network. It is shown that FCN works better than CNN in automatic segmentation of remote sensing images. The accuracy achieved with FCN in this model is above 85%. As we know with increase of satellites, there is a sudden increment in remote sensing image data. And it is almost impossible to segment those images manually to find useful results. Now as the deep learning technique has advanced, it is now possible to extract low level to high level features from raw input data images automatically. Hinton has explained in his work that deep learning methods can obtain much more useful features than the existing methods, and also has the good classification ability. In MRI images or x-ray images, it is very difficult to identify different parts of body. Therefore, image segmentation helps in segmenting different part of the body in those complex images. Pim Moeskops Et al. [4] applied image segmentation on MRI (magnetic resonance imaging) image dataset of brain, breasts and cardiac CTA). The classical procedure of segmentation is to train the model with some hand-crafted features and then use the model on input data to classify the objects, whereas in this approach CNN automatically extracts features which are required at the hand for the task of classification. Here, CNN is used to classify medical images of knee cartilage, brain regions, pancreas and coronary arteries. Four results were obtained from the network, one when the network was trained only for brain MRI, second only for breast MRI, 3rd only for cardiac MRI and 4th when network is trained for all together. Each network is trained using 25000 batches of images per task [12]. ZhuolingLi Et al. [5] have developed a method to adapt domain called CLU-CNN, which is specifically designed for medical images. As oppose to normal object identification task which are designed for pedestrian, cars or trees detection, here the task is detecting object in medical images like some deformities in eye cornea images or some object in lungs x-ray images etc. As we know, medical images have apparently very different features from normal images, so they needed to be handled differently and model training also needed to be done while keeping this thing in mind. Another problem with medical images was that deep neural networks suffer from a very big amount of parameters, this causes slow convergence rate. To overcome this problem sparse methodology and new learning rate schemes have been introduced. Also, medical images are too expensive to obtain. Therefore, the size of training data is not enough to be used as general images training data. The size of training data might be as low as less than 400 which results in not covering all the possible cases while training the model. Also medical images may be taken at different hospitals and different environmental conditions, so images relating to same problem might have a

lot of dissimilarity. As a result of these shortcomings with medical data set, a very deep neural network model cannot be used. There may be many parameters to be trained properly and would need a larger dataset to be able to work accurately. So, as a result they have developed a framework named Clustering Convolution neural networks (CLU-CNN) which aims specifically at medical images segmentation task [5]. Jie Chang Et al. [6] have applied R-CNN to detect and recognize the marine organisms in image restoration sea surface images. R-CNN requires a lot of labeled training data to work accurately and it is impossible to find such a big data set of marine organisms. So, there are three data augmentation methods for underwater imaging has been proposed in this paper. A way of inverse process is used for the simulation of different turbulences in the marine environment. Then the performance of each data augmentation method is compared. Although there have been so many advantages of robotic capturing, and some of the techniques have already been applied in practical. For e.g. Norway developed an ROV (Remote Operated Vehicle) for submarine harvesting but it requires an experienced operator who knows where and how to reach a location where lot of marine organisms can be found. So, this object detection and machine learning algorithm is becoming quite successful during the past decade which helps in automatic object detection and recognition. Cem M. Deniz Et al. [7] have proposed a method to measure bone quality and access fracture risk. As we know, manual segmentation of Magnetic Resonance (MR) is time consuming and error prone process. So, in order to deal with this problem and to take full advantage of MR scan technology an automatic femur segmentation technique has been developed. This method is totally based on Image Segmentation using Convolutional Neuron Network. This method greatly helps in detection of proximal femur in bones. The data set used by them is Microsoft COCO. It contains 2.5 million instances of labeled images [7]. It contains around 20000 images of training set and 2000 images of validation set along with 3000 images of testing set data. They achieved an accuracy of 97.8% and exceeded the performance of 2D CNN by a large margin. The performance of the model improved with adding of more number of layer and feature maps to the model. M. Yang Et al. [8] have used CNN on 3D input to recognize human activity. They have used CNN to process raw input, thus making process of feature extraction automated. This model works on 3D data as most of other CNN models only process 2D data. So, it works on both spatial and temporal dimensions to recognize the activity in a scene. It works on several frames of input thus generates multiple channels of information. The final result is obtained by combining all the input channels together detecting the motion activity in sequence of frames. The accuracy of their model came to be 90.2%. They used a very large data set of 49 hour long video of TRECVID 2008 data [8]. When video was taken certain assumptions about the circumstances of the situation were made because of issues like small scale and changes in the viewpoint. But such kind of assumptions hardly holds in real world scenarios. They designed the model to be able to automatically identify the features which are needed to be extracted from the input. Then, that input is fed to the model to identify the actions. Ahmed Bassiouny Et al. [9] used CNN to categorize images into different categories on giving an input data set of different scenes. They created image segmentation

map of the image for the purpose of representing the images by their segmentation map itself. Image segmentation map is a one to one mapping of each pixel of an image to a set of labels. This approach not only identifies the objects contained in the scene but also identifies shape, location and size of the objects. They used two data sets i.e. MSRC-21 and SIFT flow, to train their model in two different ways. The accuracy achieved by each dataset comes to be 81.9% and 66.8% respectively. Venugopal K.R. Et al. [10] have combined CNN with CC (connected component) algorithm to segment the SEM images. As we know CNN is used to extract features directly from raw images with minimal pre-processing. Also, CNN is able to recognize patterns in the image which are not provided as training data before, provided it resembles one of the training data images. Accuracy (F-score) of the model is found to be 78%. They have done detection of neuron tissues. Neurons are particularly difficult to segment as they are branched, intertwined and are very closely packed. Axons in neural structures are very thin that only electron microscopy has enough resolution to reveal them. In this approach all steps of parameter tuning is automated. So, in order to do automation and avoid human interaction in parameter tuning, CNN has been introduced. In order to recognize 2D patterns with very high degree of images transformations techniques a CNN is used as a multi layer perceptron. Six features maps are constructed in each layer of this CNN model. All the filters in the CNN are given a size of $5 * 5$ [10]. Random selection on the basis of normal distribution of standard deviations is done to calculate the weights of the edges in the model. The input to the CNN is a raw EM image. Detection of features in the image is done automatically by the CNN. CNN automatically learns by adjusting its weights by using stochastic gradient descent learning algorithm. Ji Shunping Et al. [11] have used 3D CNN to automatically classify the crops in remote sensing images taken by satellites. In order to structure the multi-spectral remote sensing data, kernels are designed according to that. Also, fine tuning of 3D CNN's parameters are done in order to train the model with 3D samples of the crops which leads the model to learn spatial and temporal representations. Full crop growth cycle samples are preserved for training the model again for the fully-grown crops. Florent Marie Et al. [12] have segmented the images of kidneys and nephron blastemal with the help of case based reasoning and CNN architecture. They have introduced two methods to achieve the required task. First one is CBR i.e. case based reasoning method for determining a region growth in process of image segmentation. And the second one is the method to perform image segmentation on a small dataset. Input images taken were are CT scans of kidneys of children having tumors in them. The result accuracy obtained is between 88 to 90% overall. The CBR system that they have used comes from CT-scan and it searches for the closest image that has already been segmented in the base case. This technique helps in saving a lot of time and computations. In designing the CNN model the major problem arises is the need of a large training dataset. This however is not the problem in this case as here each tumor is unique and there is only a limited number of tumor segmentations available from the previous tests. Thus, they have developed a method to establish a dedicated deep neural network for each patient that will carry the task of tumor segmentations automatically without the need of human intervention. Bullock Et al. [13] have segmented x-ray

images to identify various regions in an x-ray image like-bones, soft tissues, open beam regions etc. This segmentation is required to quickly identify the region in an x-ray image. As all regions in an x-ray image are of similar shade therefore it becomes very crucial and precisely done task. X-ray image segmentation is an important task in many medical applications such as image enhancement, computer assisted surgery and anomaly detection. These applications normally require medical images to be segmented into three categories i.e. open beam, soft tissue and bone. Current methods rely on classical processing methods such as clustering, line fluctuation analysis or entropy-based methods which normally requires the tuning of hyper-parameters for each body part separately. They have designed a unique CNN architecture to perform the segmentation by extracting fine grained features, alongside with controlling the number of training parameters to control the over fitting problem [13]. This model attains an overall accuracy of 82% (TF: true positives). Yan Song Et al. [14] have discussed segmentation and synthesis methods for sonar images. Also, SSS image segmentation is done for which features are extracted using CNN of multi-pathways. The advantage of CNN with multi pathways is that it can learn local and global features adaptively. SSS i.e. side sonar scanner provides essential high-resolution underwater images which is useful for many tasks like- mine detection, oceanographic research and rescuing underwater. SSS images can be segmented into areas like object highlights, object shadows or sea floor areas. But there is lot of noise and inhomogeneity in SSS images that makes the task of segmenting SSS images very difficult and inaccurate. Shan E Ahmed Raza Et al. [15] have used CNN based deep architecture for the segmentation of cells, nuclei and glands in fluorescence microscopy and histology images. They have used extra Convolutional layers so that it can work on variable input intensities and object sizes to make the model robust to noisy data. This model helps us to build molecular profiles of individual cells. Dan C. Cirean Et al. [16] have segmented neuronal structures in the stacks of electron microscopy images. The label of each pixel in the image is predicted by the pixel centered at square window in which the pixel exists. Then the input layer maps each pixel to a neuron. Then convolution and max pooling layers are applied to preserve the 2D information and to extract features of the image. This approach outperforms the competing techniques with a large margin in all three categories. The solution is based on DNN (Deep neural network) used as a pixel classifier. The probability of a pixel being a membrane is calculated. Wang, S. Et al. [17] have developed a technique to find the Ulva prolifera regions around oceans by extraction of UAV images and applying image segmentation using CNN. Here first of all raw images are processed with super-pixel algorithm to get the local multi-scale patches. Then CNN classification is applied on those patches. By combining super pixel algorithm segmentation and CNN classification they have produced much better results as compared to pixel level segmentation or instance aware segmentations. First of all, a reflection of light is removed as a pre-processing step to process raw UAV images. Then we choose super pixels elements extracted via energy driven sampling. It iteratively refines the super-pixels by updating their boundaries. Here, they defined energy function as the sum of parts $H(s)$ and $G(s)$ [17]. $(E(s) = H(s) + y G(s))$ Where, $H(s)$ is based on the likelihood of the color of super pixels. $G(s)$ is a prior of the shape of the super pixel. In

the end they have compared different numbers of super pixels to evaluate and improve the accuracy and efficiency of the model. Jie Chang Et al. [18] have done segmentation of MR images of the brain using CNN for developing a method of automatic brain tumor detection. They have developed a two-way path model which contains one average pooling layer and other max pooling layer in different paths. Then, finally the CNN model is combined to a fully connected layer to predict optimized results. As MRI scans are very useful in detecting inner body dysfunction, the need of automatic region detection becomes an essential in the medical field. Pre-processing-Not having properly normalized and quantifiable pixel intensities interpretation is one of very big defects in MRI data. So, we need to preprocess the training data in order to extract meaningful quantifiable data. With the growing number of microscopic images automatic Nano particle detection has become an essential task to achieve. Ayse Betul Oktay Et al. [19] have discussed a method to detect Nano particle in microscopic images using segmentation. They have proposed a method for the detection of Nano-particles and detection of their shapes and sizes with the help of deep learning algorithms. This method employees multiple output CNN and has two outputs. First is the detection output which gives the location of the Nano particle in the input image. Other is the segmentation output which outputs the boundary of segmented Nano particles. The final sizes of the Nano particles are determined by the Hough algorithm that works on the segmentation outputs. Here they have used MOCNN i.e. multiple output Convolutional neural network for the purpose of detecting, localizing and segmenting of Nano particles found in microscopic images [19]. MOCNN takes an input image as a window and it produces two outputs for it, from which one output tells us the location of the Nano particle in the image and second output tells us the distance of the object boundary to the window center. Despite very accurate and one of the best ways of image segmentation by CNN, it does have some uncertainties which nobody likes to tackle. Guotai Wang Et al. [20] have analyzed these different kinds of uncertainties for CNN based 2D and 3D image segmentation. Additionally, they have analyzed a test time augmentation-based uncertainty to analyze the effect of different kind of image transformations on

segmentation outputs. They have proposed a test-time augmentation based uncertainties in order to monitor the effect various transformations on the input images. Their task consists of image and noise transformations. Then they compared and combined the proposed aleatoric uncertainty with uncertainty of the model. They concluded that test time augmentation based aleatoric uncertainty provides better results than any other kind of uncertainty calculation method. Nowadays, UAV are most useful in monitoring of ocean areas rather than radars or satellites. As UAVs provides high resolution real-time images and this is not possible with other two sources

4 COMPARISON OF RELATED WORK DISCUSSED

We have compared the literature survey discussed above on basis of various parameters like accuracy, training data size, applicable areas, complexity, methodology and data-source in a tabular form. After comparing these works we have found certain observations. We found the maximum accuracy in the work of Deniz CM and Xiang S [7], where they have achieved an accuracy of 97.8%. These results show that although each model is unique and serve a specific purpose but performance matrices differ in each model substantially. If we talk about training the model on a smaller dataset and still be able to provide useful results work of Sadegh Karimpouli et al. [2] to segment the rock images prove to be very much efficient and useful. This may increase the complexity of model but provides quality results as complexity is not much of an issue with availability of such high end computing hardware systems. Furthermore, Pim Moeskops Et al. [4] proposed model uses image set of 25000 images which provides a good relatively good accuracy of 87%. Their model is trained with a relatively large data set but proves to be lesser complex than other models. Considering area of application, image segmentation is mostly used in medical field and surveillance systems in remote areas. Finally, we have analyzed and compared the work done by many researchers in the field of image segmentation using Convolutional neural networks with the help of table 2.

TABLE 2- Comparison on the basis of various parameters

Paper	Accuracy	Training Data Size	Application area	Complexity	Technique used	Data Source
[1]	77.5%	3200 images	Healthcare, Blood cell growth monitoring	High	Microscopic Imaging	"PhC-U373" and "DIC-HeLa" data set
[2]	96%	500 images	Object classification, Rock image classification	High	Data Augmentation and SegNet Architecture	Digital Rock Samples introduced by Andra et al
[3]	85%	15000 images	Surveillance, Segmentation of Satellite images	Medium	FCN and expanded CNN	VGG16

[4]	87%.	25000 images	Healthcare, To classify images of knee cartridge, brain regions and pancreas.	Low	MRI and X-Ray	Weighted MR brain images from the OASIS project
[5]	72%	400 images	Healthcare, detecting deformities in eye cornea images	High	CLU-CNN	REFUGE CHALLENGE 2018
[6]	60%	1800 images	Marine Organisms Detection	Medium	Faster R-CNN and Data augmentation	Marine organisms data by National Natural Science Foundation
[7]	97.8%	2000 images	Healthcare, To measure bone quality	Medium	MR Scan technology and CNN	Microsoft COCO
[8]	90.2%	49- hour video	automated video surveillance, human action recognition	high	3D CNN	TRECVID 2008 Data
[9]	81.9% and 66.8%	591 + 2886 images	Surveillance, Scene Recognition	Low	Segment Count Feature Vector(SegC) and SegC with Centrist	MSRC-21 and Siftflow
[10]	78%	3000 images	Neuro-Science, Detection of neuronal tissues	High	CNN and Connected Component Algorithm(CC)	Drosophila First in star larva ventral nerve cord (VNC)

5 END SECTIONS

ACKNOWLEDGEMENTS

This work was supported by Computer Science Department of Delhi Technological University. Authors thank the department and all the faculty who provide the opportunity to present this work.

CONCLUSION

In this work we have compared various image segmentation techniques and their state of the art implementations. After researching various techniques we have found that the CNN is one the most powerful tool in image segmentation technique. Detailed analysis of CNN is also done here explaining different layers and workings of each layer. We have explained all the possible advantages and fields where CNN can be used in our daily life. As we know CNN technology is at a boost of implementation nowadays in making the human life more and more convenient and less manual. There have already been a lot of work done in various fields like commutation, medical tasks, crop monitoring, road transportation, activity detection, product quality monitoring. All these fields have seen a great improvement after the use of these techniques, so all the work done is itself state of the art techniques.

5.3 FUTURE SCOPE

There is always room for improvement, innovation or change of existing techniques in any research field. So, despite availability of so much quality research work in this field, yet there is a lot of work to be done in making those automatic monitoring systems more accurate and reliable. There is more scope in handling the uncertainties where images are of bad quality or the boundary pixels of the segmented objects are overlapping. Accuracy of such systems need to be increased to an extent that they can be relied upon to do crucial tasks such as monitoring unidentified activities in restricted areas such as country borders or ministerial offices, where the slightest inaccuracy may prove to be disastrous. Various models need to be implemented by combining two more models into one such system that it can enable a robot or any other automatic system to do tasks more effectively and accurately. For e.g. a system can be designed which is able to identify the objects as well as monitor the activities in surroundings at the same time. This will increase the applicability of such techniques and will make manual tasks much more automatic as compared to they have ever been.

REFERENCES

- [1] Olaf Ronneberger, Philipp Fischer and Thomas Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation", MICCAI 2015: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 pp 234-241.

- [2] Sadegh Karimpouli, Pejman Tahmasebi, Segmentation of digital rock images using deep Convolutional auto encoder networks, *Computers & Geosciences*, Volume 126, 2019, Pages 142-150, ISSN 0098-3004, <https://doi.org/10.1016/j.cageo.2019.02.003>.
- [3] Xiaomeng Fu and Huiming Qu, "Semantic Segmentation of High-resolution Remote Sensing Image Based on Full Convolutional Neural Network", 2018 12th International Symposium on Antennas, Propagation and EM Theory (ISAPE) DOI: 10.1109/ISAPE.2018.8634106
- [4] Pim Moeskops, Jelmer M. Wolterink Bas, H. M. van der Velden, Kenneth G. A. Gilhuijs, Tim Leiner, Max A. Viergever and Ivana Išgum, "Deep Learning for Multi-task Medical Image Segmentation in Multiple Modalities" MICCAI 2016: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016 pp 478-486 047
- [5] Zhuoling Li, Minghui Dong, Shiping Wen, Xiang Hu, Pan Zhou, Zhigang Zeng, CLU-CNNs: Object detection for medical images, *Neurocomputing*, Volume 350, 2019, Pages 53-59, ISSN 0925-2312,
- [6] Hai Huang, Hao Zhou, Xu Yang, Lu Zhang, Lu Qi and Ai-Yun Zang, "Faster R-CNN for marine organisms detection and recognition using data augmentation", *Neurocomputing*, Volume 337, 14 April 2019, Pages 372-384
- [7] Deniz CM, Xiang S, Hallyburton RS, Welbeck A, Babb JS, Honig S, Cho K, Chang G. Segmentation of the Proximal Femur from MR Images using Deep Convolutional Neural Networks. *Sci Rep*. 2018 Nov 7;8(1):16485. doi: 10.1038/s41598-018-34817-6. PubMed PMID: 30405145; PubMed Central PMCID: PMC6220200.
- [8] S. Ji, W. Xu, M. Yang and K. Yu, "3D Convolutional Neural Networks for Human Action Recognition," in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221-231, Jan. 2013. doi: 10.1109/TPAMI.2012.59
- [9] Ahmed Bassiouny, Motaz El-Saban, "Semantic segmentation as image representation for scene recognition", 2014 IEEE International Conference on Image Processing (ICIP).
- [10] M. L.S., V.K. G. (2011) Convolutional Neural Network Based Segmentation. In: Venugopal K.R., Patnaik L.M. (eds) *Computer Networks and Intelligent Computing*. ICIIP 2011. Communications in Computer and Information Science, vol 157. Springer, Berlin, Heidelberg
- [11] Ji Shunping & Chi, Zhang & Xu, Anjian & Shi, Yun & Duan, Yulin. (2018). "3D Convolutional Neural Networks for Crop Classification with Multi-Temporal Remote Sensing Images" 10. 75. 10.3390/rs10010075.
- [12] Florent Marie, Lisa Corbat, Yann Chaussy, Thibault Delavelle, Julien Henriet, Jean-Christophe Lapayre, Segmentation of deformed kidneys and nephroblastoma using Case-Based Reasoning and Convolutional Neural Network, *Expert Systems with Applications*, Volume 127, 2019, Pages 282-294, ISSN 0957-4174
- [13] Bullock, Joseph & Cuesta, Carolina & Quera-Bofarull, Arnau. (2018). XNet: A convolutional neural network (CNN) implementation for medical X-Ray image segmentation suitable for small datasets.
- [14] Yan Song, Bo He, Peng Liu, Tianhong Yan, Side scan sonar image segmentation and synthesis based on extreme learning machine, *Applied Acoustics*, Volume 146, 2019, Pages 56-65, ISSN 0003-682X
- [15] Shan E Ahmed Raza, Linda Cheung, Muhammad Shaban, Simon Graham, David Epstein, Stella Pelengaris, Michael Khan, Nasir M. Rajpoot, Micro-Net: A unified model for segmentation of various objects in microscopy images, *Medical Image Analysis*, Volume 52, 2019, Pages 160-173, ISSN 1361-8415,
- [16] Dan C Cirean, Alessandro Giusti, Luca M Gambardella and Schmidhuber, "Deep Neural Networks Segment Neuronal Membranes in Electron Microscopy Images", *Advances in neural information processing systems* 25 · January 2012
- [17] Wang, Shengke & Liu, Lu & Qu, Liang & Yu, Changyin & Sun, Yujuan & Gao, Feng & Dong, Junyu. (2018). "Accurate Ulva Prolifera Regions Extraction of UAV Images with Superpixel and CNNs for Ocean Environment Monitoring" *Neurocomputing*. 10.1016/j.neucom.2018.06.088.
- [18] Jie Chang, Luming Zhang, Naijie Gu and Xiaoci Zhang, "A Mix-pooling CNN Architecture with FCRF for Brain Tumor Segmentation", *Journal of Visual Communication and Image Representation* 58 · December 2018, DOI: 10.1016/j.jvcir.2018.11.
- [19] Ayse Betul Oktay, Anil Gurses, Automatic detection, localization and segmentation of nano-particles with deep learning in microscopy images, *Micron*, Volume 120, 2019, Pages 113-119, ISSN 0968-4328,
- [20] Guotai Wang, Wenqi Li, Michael Aertsen, Jan Deprest, Sébastien Ourselin, Tom Vercauteren, Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks, *Neurocomputing*, Volume 338, 2019, Pages 34-45.