A Survey on Techniques used in Medical Imaging Processing

Priyadarshini Chatterjee¹ and Sushama Rani Dutta²

¹Research Scholar, Koneru Lakshmaiah Education Foundation, Hyderabad
²Assistant Professor, Koneru Lakshmaiah Education Foundation, Hyderabad
Email: jinipriya@klh.edu.in

Abstract. Automated diagnosis of diseases in the recent years have gain lots of advantages and potential. Specially automated screening of cancers has helped the clinicians over the time. Sometimes it is seen that the diagnosis of the clinicians is biased but automated detection can help them to come to a proper conclusion. Automated screening is implemented using either artificial inter connected system or convolutional inter connected system. As Artificial neural network is slow in computation, so Convolutional Neural Network has achieved lots of importance in the recent years. It is also seen that Convolutional Neural Network architecture requires a smaller number of datasets. This also provides them an edge over Artificial Neural Networks. Convolutional Neural Networks is used for both segmentation and classification. Image dissection is one of the important steps in the model used for any kind of image analysis. This paper surveys various such Convolutional Neural Networks that are used for medical image analysis.

1. Introduction
Understanding and finding the correct region of interest in medical images helps to correctly diagnose the disease. Image segmentation is one of the prominent steps in analyzing medical images using machine learning algorithms. Traditional image segmentation algorithms always do not give correct result and does not help to dig the region of interest correctly [1]. All anatomical structures depicted in the medical image needs to be correctly segmented. There are range of applications such as surgical planning, anatomical structure modeling and image-based diagnosis where correct segmentation is very necessary and important. Automatic segmentation methods have been researched earlier but it is also proven that they can give hardly correct results. There can be various reasons like poor image quality, the pathological parameters of the patients differ to a large extent and are inhomogeneous, the predictions made by clinicians also varies. All these varying parameters gives different structure boundary of the image to be segmented. The results of segmentation are incorrect as there is no unanimity in all the above-mentioned functional areas. Although there are methods like augmenting the medical images with user interactions, but this step seems to be a burden on the user interactions. As per the researchers a good segmentation should have few user interactions. This leads to increase in interaction efficiency. Machine learning tools are generally used to decreases user interaction. Examples of machine learning methods are Gaussian mixture models, GrabCut algorithms etc. There are some extra techniques that are required to be applied on the image like bounding box and scribbles to refine the result. These increases the computational cost. Recently deep learning tools and techniques along with Convolutional inter connected system is used for image dissection [2]. They provide results much faster than what one gets using manual techniques. There are many types of Convolutional Neural Networks used amongst which fully convolutional neural network has gained lots of importance in
medical image segmentation. It gives the correct result by providing forward processing only once and that to at the testing time. Recent improvement on Neural Network technology has been done on two aspects: first overcomes the problem of reduced imager caused by repeated combination of down sampling and highest pooling. Although certain up sampling layers recovers the resolution easily, this leads to binary like segmentation. The results are of low accuracy. In case of dilated convolutional network, the down sampling layers are replaced and exponential increase of the receptive field happens with no loss of resolution. It always requires a generalized segmentation model which receives further attention. The second aspect enforces inter-pixel dependencies in order to get a temporally correct result. This helps to understand edge details and also reduces unwanted disturbances in pixel differentiation [3]. Another important architecture that has gained immense response in the field of image processing is called as U-Net. It is applied in cell detection from two dimensional to three dimensional images. There are certain Conditional Random Field’s used as a step of post processing along with segmentation in convolutional neural networks.

2. Related Work

Typical CNN’s such as Google Net, AlexNet, VCG and ResNet were initially analyzed to do image classification tasks [41]. There are certain works earlier that have adapted these networks and progressed on with their pixel classification. These networks help in pixel labelling, having patch or region-based methods. These networks achieved higher rate of accuracy than the traditional methods. These networks have certain disadvantages also and that is shown at the time of testing. Fully convolutional network works by considering an entire image as input and gives the result in the form of dense segmentation. There is also a problem of loss of spatial design because of more than one stage of max pooling and also down sampling. To overcome these disadvantages this method uses stack of deconvolution and activation function to up-sample the layers [4].

Inspired by these features of convolutional network and deconvolutional framework, a U-shaped network called U-Net was proposed. The three-dimensional version of this U-Net architecture was specially used for biomedical image segmentation. A similar kind of architecture called V-Net is proposed that is used for segmentation of prostate MRI images. There were certain drawbacks of successive down sampling and max-sampling [5]. There is certain loss in feature map resolution. So dilated convolution was proposed that would keep safe the architectural parameters of feature maps and thus enhance the receptive field so that it contains larger textual information. In a pile of dilated convolution, object tracing and semantic segmentation is performed. Enlarged convolution is also used for instance-sensing segmentation and also detects actions from video graphic frames [6].

Convolution network are also used to extract multiple scale features. This is done to improve the segmentation accuracy. Multi scale features are obtained by passing multiple forms of the input image throughout the same network. The features thus obtained can be used is used for pixel classification. The features belonging to each pixel are obtained from two homocentric patches having varying sizes. In case of multi scale images, these are fed at number of stages into a recurrent convolutional network. Another popular use of multiple scale feature is in using the feature maps considering disproportional levels of convolutional neural network. The features from intermediate layers are augmented for the purpose of localization and segmentation.

Another procedure in segmentation that is widely used is called Interactive Image Segmentation that is used in various application. The user interactions are varied such as contour-based, click-based and bounding box methods. Scribbles can be also drawn in Graph Cuts, Random Walks, GeoS. However, most of these are used on lower-level features and requires larger amount of user markers. These helps to deal with images having darker profile and equivocal boundaries. There are machine learning tools that are proposed to learn from user markers. These algorithms cannot achieve better dissection accuracy with lesser user communications. But they need to rely upon limited manual features that relies on one’s experience. The deep convolutional network has improved interactive segmentation. The convolution neural network has spontaneous feature learning and also shows high performance. Some of the examples are a three-dimensional U-Net that learns from marked images and later on used
in not fully automatic segmentation [7]. Scribbles are also used to train the CNNs. Deep cut uses user-provided boundary box as markings to train the convolutional neural network as required for the segmentation of MRI images.

There are procedures which are not fully communicative and are also not suitable for testing, as they restrain themselves to accept farther annotations for refinement. In case of deep communicative objects selection tools are proposed where images captured are changed into Euclidean distance maps and then they are added with the input of fully convolutional neural networks. In contrast the geodesic distance transforms and encodes the spatial dimensions and contrasts the sensitivity, though it has not been used for CNNs. Graphical models such as Conditional Reference Fields are immensely used to increase dissection accuracy by using spatial accuracies. In case of spatial regularization, the Potts energy is minimized with minimum max/cut flow algorithm. In the maximum-flow problem mapping is done to provide optimization formulations [8]. They provide segmentation consistency of neighboring pair pixels having similarity in features. In the process to include long distance connections within image segmentation, a totally connected CRF is used. This establishes pair wise potential amongst all images. The adjacent edge capabilities as defined by a straight mixture of Gaussian Kernels. Other methods include gradient based optimization and integrated output support vector machines and. These are also used to learn from parameters in CRFs. After we have successfully segmented the image, we need to proceed with feature extraction. For feature extraction of medical images there are various deep learning algorithms. Convolutional Neural network can be used for feature extraction. Many researches on feature extraction of medical data deals with making a model that will analyze the text features of the medical images using convolutional neural network. After we have done the feature extraction, last step as part of image processing, we can go with a classifier to analyze the features based on their abnormalities. We can use either a convolutional neural network or any traditional classification algorithm for classifying the processed images. There are many states of the art algorithms that can be used for image recognition. It is known that Support Vector Machine algorithm is one of the most common classification algorithms that is used in many medical image processing and analysis [9].

3. Overview of Techniques

When we speak about techniques used in medical imaging, we mean preprocessing, classification, segmentation and feature extraction. This paper provides a comparison (advantages and disadvantages) of techniques and also tries to bring upon solutions to the problems for which a technique is discarded. The reader will get a broader picture of almost all the processes adopted till date in bio medical image analysis. Neural Network, is an ideal technique for image analysis and segmentation. To start with, this paper reviewed the U-Net structure for image dissection. U-Net is a CNN having the shape of U. Based on the drawbacks of process who proposed a sliding-window architecture to foretell the label of the class of each pixel, has shown how a U-Net is to be used for image dissection of a 513×513 image in less than a second on a GPU. The main principal of U-Net is that it relies totally on pixel-based segmentation. It is important in case of bio-medical images to have a class label as an output on every pixel. The U-Net architecture also had the advantage of being able to work on very few training images. This paper deals with fully connected neural network that is modified so that it takes very few trainings set to output more accurate segmentations. In case of U-Net as the training set is very few so excessive data augmentation is required. The network learns to be invariant to such changes and does not follow the deformations in the fully supervised image fraternity. This is extremely important for bio medical images [10]. The similar objects of the same class are separated using large weights in the loss function. The U-Net was applied for three different dissection tasks [42]. In order to train the data, a set of 40 images were taken of size (512×512). Each image was annotated and was separated into white and black membranes. As test set was widely available but the mapping of segmentation was kept secret. The U-Net architecture achieved a rand error of 0.0003529 and a warping error of 0.0382.
Next technique that can be mentioned in this paper is called DeepCut technique for segmenting object from bounding box annotation using CNN. This technique was proposed by Martin Rajchl et al [“DeepCut: Object Segmentation from Bounding Box Annotations using Convolutional Neural Networks”, 2016] [43]. The GrabCut technique of image dissection is modified and extended to work on image repository. An energy minimization problem was formulated over a thickly joined conditional random field. The proposed DeepCut method can be used for energy minimization as similar to GrabCut. There are two key stages in the proposed algorithm, label update and model estimation. In case of DeepCut technique the GMM used in GrabCut is replaced with Neural Network model. The graph cut solver forms a densely connected graph. Transfer Learning technique is used to reinitialize the CNN with the results of last iteration. The interconnected network used in this proposed methodology is a forward CNN having at-least one layer. The max-pooling layer reduces the sample size of input information that helps in learning object representations at different scales. This reduces the output to a limited variation of classes. If there is a database of size N having images in the form I = \{I_1, I_2, ...IN\} with the corresponding bounding boxes denoted as B=\{B_1,B_2,...BN\}, pixelwise segmentation is done on the objects present in the set I but limited to set B. A CNN with parameter Θ is used to classify the image patches that are centered around a single point located at ‘j’ into background and foreground. The network configuration consists of two sets of convolutional and highest-pooling layers. The two subsequent levels are again connected to a layer containing thickly connected neurons. There is a resultant layer having neurons related with foreground and background [11]. Training and Organization phase consists of extracting patches from the training database that are equally distributed between the classes. The lossy function is categorized as cross entropy between the coding distribution. The training data set is augmented by Gaussian -distributed intensity offset for better learning of the network. The DeepCut approach trains the Neural network model with patches Y obtained by sampling C and H for background and foreground respectively. When this method was compared with other methods, this technique gave a better result in accuracy in segmentation of the images of brain and lungs. Figure 2 is the blockdiagram of the algorithm.

Another useful technique in case of bio medical images is called Interactive Image Segmentation with Image- Specific fine tuning proposed by Guotai Wang et.al [“Interactive Medical Image Segmentation Using Deep Learning with Image-Specific Fine Tuning”, 2018, IEE transaction on medical imaging][44]. The proposed algorithm uses interactive segmentation by embedding Convolutional Neural Networks into bounding box and using scribble-related segmentation pipeline. Image specific fine tuning is performed so that the CNN model adapts to particular test images. The algorithm was used for two-dimensional dissection of multiple organs and three- dimensional dissection of a core brain tumor and the whole brain tumor [12].
The Convolutional Neural Networks here can also deal with unseen objects in the past in the framework of image segmentation. The Convolutional interconnected neural network system here takes as input the content of a boundary box of a particular image and provides binary dissection of that content. At the time of testing the user provides the content of a boundary box into the Convolutional Neural Network [13]. The segmentation model (BIFSeg) draws out the region inside the bounding box. This extracted part acts as an input to the pre trained the Convolutional interconnected system to acquire the initial dissection. The Convolutional interconnected system is trained to adapt itself to various parameters like contrast, saliency etc so that they can generalize their learnings to unseen objects. The method can either use supervised or unsupervised image respective fine tuning to refine the segmentation. For two-dimensional image segmentation P-Net is adopted. This network preserves the resolution preserving and uses dilated convolution. There is a classifier that accepts the output after dilation. For the testing stage the model is updated using image specific fine tuning. For three dimensional images an extension of P-Net is used. The network has a dipolar receptive field that uses less memory and can be applied to Magnetic Resonance Imaging. This network is named as PC-Net [14]. Various CNNs result is compared in this paper and it is established that BIFSeg can handle some of the cases that other architecture cannot handle.

![Figure 3. Interactive Image Processing](image)

Next technique that this paper reviews is called as A Deep Interactive Geodesic Framework for medical image segmentation by Guotai Wang et.al [“DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation”, 2017] [45]. In this method one CNN is used to provide initially spontaneous segmentation and user communications are added to identify deviations of segmentation. The other Convolutional Neural Network takes this as input along with user communications to provide refine results. The initial network is a P-Net and the next network used is a R-Net [16]. The P-Net having C1 takes input an unprocessed image and does the initial spontaneous segmentation. This initial segmentation is analyzed by users and further augmented with scribbles and clicks are again sent for refinement through the R-Network. Both R-Network and P-Network preserves the dimension of the image by using Conditional Random Field technique [15]. The R-Net and the P-Net are together trained by backward propagation. The scribbles are used as user interactions which divides an image into foreground and back ground pixels. Then Geodesic distance is used to convert dependencies of variables in the future space. This is combined with Random Forest to perform semantic segmentation. The geodesic distance transform along with the scribbles is used as user segmentation. The convolutional interconnected network in the proposed method can take high -level features using receptive fields [17-18]. The CRF is implemented using recurrent neural network and the result of each block of the network is dilated with 0 parameter. The proposed algorithm is compared with FCN and DeepMedics network for doing two-dimensional segmentation and DeepMedics and High-Res3DNet for three-dimensional dissection. A detailed comparison is provided by Guotai Wang et.al. It is also shown that Geodesic distance with two CNNs P-Net and R-Net provides better result for two-dimensional placenta dissection captured using three-dimensional brain tumor bifurcations and MRI images using FLAIR images. Figure 4 below shows the algorithm overview.
Next technique this paper reviews is on ["Inf-Net: Automatic COVID-19 Lung Infection Segmentation from CT Images", 2020, IEEE transactions on medical imaging] proposed by Deng-Ping Fan et al [46]. This paper proposes an algorithm that detects lung infections from Tomographic images in a spontaneous way [19-20]. As segmentation and collection of data for training the network, so the algorithm proposed a unique lung infection network known as “Inf-Net” that automatically identifies infected areas from the slices of c images. The Inf-Net uses a decoder parallel in nature to assemble higher level features after which it generates a universal map. The particular edge augmented and implicit reverse augmented is used to identify the boundaries and thus enhances the representations. In order to compensate for lesser data, a half-supervised bifurcation model is used. This segmentation framework is based on haphazardly chosen communication strategy which requires very few labelled images and augments unlabeled data. Elaborating on the mechanism of Inf-Net, the Tomographic images are initially fed into two convolution layers in order to process high resolution and low-level features. The edge concentration module is added to improve the boundary regions. Features having lower levels are then fed into the CNN to acquire the high-level features. There is a parallel partial decoder to assemble these features to assemble the global map for the fine localization of the lung infection. Then these features are combined with the low-level feature are then given as input into multi backward concentration module under the guidance of the universal map. The technique of making the edge attention module relies on feeding the low-level features with limited resolution in order to learn the explicit edge attention. The Inf-Net is designed to contain two different network components that are used as fine labeller and rough indicator [21]. There is also an augmented framework that acts as a fine labeller. Instead of accumulating features from all levels, there can be adaptive learning i.e the reverse attention can be done in three parallel levels. The Inf-Net is improved using semi supervised learning that helps in augmenting large number of un labelled images so that the training data set can be made more powerful. In this work 150 axial Tomographic images of different COVID-19 patients were collected and the data set was augmented with semi-supervised COVID-19 segmented data set. The structure of the proposed method is shown below. This proposed method was compared with five traditional segmentation models. After comparison with the baseline methods, it is shown that the method’s performance is better than the other methods. Figure 5 below is the block diagram of the algorithm.

**Algorithm:**

Non-Supervised Inf-Net

**Input:** Calibrated trained data $K_{labeled}$ and not labelled training data $M_{unlabeled}$

**Output:** Trained Inf-Net $L$

1: Constructing a trained data $D_{Training}$ having all the labelled Tomographic images from $K_{labeled}$
2: Training the model $N$ by $D_{Training}$
3: repeat
4: Do testing by the training structure $N$ and $P$ Tomographic images that are randomly selected from $M_{unlabeled}$, that outputs network-specific data $L_{Net-labeled}$ having $P$

Tomographic images having false labels
5: Enhance the dataset for training with $L_{Net-labeled}$

\text{\textit{i.e.,}}
\[ D_{\text{training}} = D_{\text{training}} \cup D_{\text{Net-labeled}} \]

6: Delete the \( P \) testing Tomographic slides from \( M_{\text{Unlabeled}} \)
7: Perfectly-tune \( n \) by \( D_{\text{training}} \)
8: until \( M_{\text{Unlabeled}} \) is empty
9: return After trained structure

Zhijie Zhang et al in the year 2019 proposed [“ET-Net: A Generic Edge-aTention Guidance Network for Medical Image Segmentation”] [47]. This uses the edge-specific model to assist the network. The E-Net is a decoder encoder network. The E-Net has the edge guidance model and the weighted average modules attached at the end. The ResNet-50 is used as an encoding interconnected system that consists of Encoder-block one for each feature. The input initially goes into a feature extracted stream that consists of convolutional layer. The final output is generated by summing up the output with the short circuiting of the inputs. Having this interconnection, the proposed model will generate class oriented higher-level features from the Encoder-block [22]. There is also a D-block that supports depth wise convolution along with a \( 2 \times 2 \) convolution to increase the quantity of channels. The edge specific module is applied on the early layers. This module has mainly two important functions:
- It provides an edge specification representation to accelerate the procedure of segmentation. It monitors the CNN layers by edge detection loss. In order to accept various shapes and sizes of different objects, the proposed system sums up multi-scale outputs for final predictions. This module highlights important features, attenuates increased number information and edge specification to better the segmentation.
- For data augmentation random scale, random mirror and random rotation is applied to all datasets. The proportional weights of the encoder network are trained previously on ImageNet and the features of the all layers are spontaneously initialized. A dilation technique is used as the output of Encoder-block with the output parameter of \( 1/16 \). At the time of training the batch size is set to 17. This approach is applied to some major varieties of images such as X-Ray images, Tomographic images and retinal images. As an evaluation metrics, the die coefficients of the optic cup and optic disc as well as average union over intersection are used. The components of this method are compared with other base method, the accuracy percentage is computed considering images of under different settings. It is shown that with the addition of edge guidance module, the accuracy percentage is achieved to be 91% to 97% other than 84% to 90% obtained by other method. Below Figure 5 is the block diagram of the algorithm.

![Figure 5. ET-Net Architecture for Segmentation of Retinal Image](image)

“Transfer Learning with Edge Attention for Prostate MRI Segmentation” by Xiangxiang Qin [48] analysed how learning implemented using deep neural network along with multiple levels edge attention can bring in much more accuracy in segmentation of MRI pictures. The components required for are a 3D-pretrained encoder, 3-D multiscale decoder, multiple-levels edge specific module and a Pyramid attention structure. The three-dimensional encoder is previously trained on the MRI images. There is a difficulty in the 3D prostate segmentation which is non clarity of the boundary regions. At the last three level of the decoder, convolutional layer is used to extract many-level edge data. The addition mechanism is used to monitor how the decoder acquires the information. The multiple-level edge specific model acquires three levels of edge information [23-24]. This information also guides the paths of extracting features from the encoder. This information adds with edge feature maps to predict output of the last layer. The Pyramid attention module has two parts, the out put product of all encoded images.
level and the unsampled result of each decoding level. The proposed method was implemented in a Pytorch framework requiring 11 GB GPU. At the time of pre-training, the spatial design of each dimension is made equal to 0.626×0.626×1.4 mm. An SGD optimizer is used having a frequency of 0.9 and the initial learning rate was decreased by a weight decay after a particular period of time. At all training iteration the input is fed into the network, deformed images. The sub volumes were randomly cropped to the size of $97 \times 97 \times 33$ voxels. The data enhancement was performed online that can reduce overfitting by limited training images.

Figure 6. Segmentation Using Transfer Learning

“Deep Learning Applications in Medical Image Analysis,2018 [49]” by Justin Ker et al gives overview about the applications of different convolutional neural network on medical imaging. As per the study, both 3 d and 2 d structures can be analyzed with the help of CNNs. CNNs can be well used to output image recognition tasks. CNNs can perform many tasks like detection, localization, classification, segmentation and registration. As per the paper, Convolution Neural Networks can perform the image classification task due their unique characteristics of preserving local image relations and at the same time they can also perform dimensionality reduction. Relationship between important feature needs to be captured thereby reducing the number of parameters and increasing the computational efficiency [25]. CNNs can take as input both two- and three-dimensional images. This becomes an important application for hospital use. Supervised and Unsupervised deep learning algorithms are popularly used in analysis of medical images. Among the supervised learning models, this paper throws light on convolutional neural networks. The CNN is defined to be an operation between two functions. While doing image processing, one function is considered to be voxel sizes at a particular position in the image and the other function is a kernel or filter. All input may is represented as array of matrices [26]. The convolution operation is represented as a * symbol. The output for dimensional convolutional operation can be defined as

$$Y(k) = \sum_a I(m). M(t - m) \quad (1)$$

A two dimensional convolution operation with the image input $I(x,y)$ with respect to kernel$(m,n)$ is defined to be

$$X(k) = \sum_m \sum_n I(a,b). K(x - a, y - b) \quad (2)$$

Apart from the convolutional layer there is a RELU layer that functions to set the negative input values to zero. This helps to speed up the calculation and training. Then there is a pooling layer that decreases the number of parameters to be calculated. Finally, there is a totally connected layer, takes the result from preceding layer. It computes a probability to classify to categorize the output into various classes. This paper has also surveyed Transfer learning and Recurrent neural network methods to detect cancer cell. It has also surveyed unsupervised learning methods and dealt in details how CNNs are used in medical image analysis. Figure 7 below is the blockdiagram of the algorithm.

Figure 7. Classification using CNN
“A Noise-Robust Framework for Automatic Segmentation of COVID-19 Pneumonia Lesions from CT Images [50]” by Guotai Wang et al in August 2020 describes how analysis of pneumonia abrasions detected from the Tomographic scans of COVID 19 patients can be made to be important for correct diagnosis of lung infections. This is a noise robust framework that uses noisy images and make them to learn from used for the segmentation task. The architecture of the work is explained using Block diagram of figure 8.

\[ \text{L}_{\text{NR-Dice}} = \sum_{j}^{M} \frac{|p_{j} - q_{j}| \lambda}{\sum_{j}^{M} p_{j}^2} + \sum_{j}^{M} g j^2 + \beta \quad (3) \]

This Cople-Net was compared with four networks and the result of the comparison is tabulated in the paper. It was also shown that the Cople-Net achieved the best result in terms of segmentation [28].

Next paper that is used for studying is “Supervised Edge Attention Network for Accurate Image Instance Segmentation” by Xier Chen et al. This paper is about Instance segmentation for which a mask called R-CNN is used [29]. There is a fully convolutional box head for detecting branches and also to find out correct boundary box for segmentation. The supervised edge specific module is used as an addition to the original mask. The architecture of the model is shown in figure 9 below.

\[ R = R_{\text{box}} + R_{\text{seg}} \quad (4) \]
Where $H_p$ and $W_p$ are height and width of the proposal. In order to minimize the non-specific content and to protect the boundaries of the objects from becoming hazy, the attention module is applied [30].

There are four convolutional layers to achieve the edge augmentation features to nearly similar depth as that of the feature it uses to add with on the mask head. This helps to put the features in almost equal data set making them more adaptive. The multiple task loss function is defined as

$$L=L_{bbox} + L_{seg}$$

(5)

Where $L_{bbox}$ includes three parts:

$$L_{bbox} (c, B, I) = \lambda_1 L_{cls} (c, c *, _) + \lambda_2 L_{reg} (B, B^*) + \lambda_3 L_{IoU} (I, I^*)$$

(6)

It is also required to make strong the hard sample learning; a logarithmic function is used as follows:

$$L_{reg} = - \ln (G_{IoU} + 1)$$

(7)

The loss of mask $L_{seg}$ combines two terms

$$L_{seg} = \lambda_4 L_{mask} + \lambda_5 L_{edge}$$

(8)

The experiment is performed on an increased-scale object detection and example segmentation model called COCO. Standard evaluation metrics of COCO is used as an estimation to analyse the performance of the proposed system.

“Enhancing U-Net with Spatial-Channel Attention Gate for Abnormal Tissue Segmentation in Medical Imaging [51]” by Trinh Le Ba Khanh et al in August 2020. This paper proposes spatial-channel specification gate that increases the segmentation capacity of the U-Net structure and having minimum computation overhead [30]. This paper proposes that the mingling between the encoder and decoder features can be made more efficient by increasing the textual information. The Spatial-Channel Specification Gate can overcome the drawbacks of traditional U-Net network. The Spatial-Channel Attention Gate is having four parts:

Encoder module, Decoder module, Fusion module, Forecasting module.

The data taken as input as an image is directly given into the encoder module. The features resulting from the decoder is generated by multiplying the Spatial-Channel Specification Gate with the result of the encoder. The resultant output of the decoder is fed into the prediction module, for predicting the abnormal tissues. The Spatial-Channel Attention Gate consists of two types of attention modules that is used to guide the model to pay attention to spatial and detailed structure of important regions [31]. It helps to increase the textual data in the low-level encoder features. The output of the encoder can be summarized as

$$M_k ' = M_k ' \Theta M_s (M_e, M_d) \Theta M_c (M_e, M_d)$$

(9)

Where $\Theta$ denotes element wise multiplication. The proposed Spatial-Channel Attention Gate tries to find out the important regions. This important region will be segmented. The spatial specification map is constructed based on the relation of important regions. This process of refining the input characters is depicted below in the diagram. There is also a Channel Specification gate that supports to analyze what is important in the input feature. The channel specification map is constructed taking into account the
interrelation between the channels of the convolutional features. Figure 10 below is the overview of the Fusion Mode

![Image](image1.png)

**Figure 10. Fusion Model**

“SpineParseNet: Spine Parsing for Volumetric MR Image by a Two-Stage Segmentation Framework with Semantic Image Representation [52]” by Shumao Pang et al in IEEE transactions on medical image processing, January 2021 deals with multiclass dissection of volumetric MRI images of intervertebral and vertebrae discs [32]. This plays crucial role in detection of different spinal diseases and treatment of spine disorders. SpineParseNet is a two-stage framework that parses the volumetric MR images of spines. The SpineParseNet is made of a three-dimensional graphical convolutional segmentation network, and a two- dimensional residual U-Net. In three-dimensional graphical convolutional segmentation network, each node of the graph is specific spinal structure. Region pooling is converted to graph representation [33]. The adjacency matrix of the graph is represented as per the spinal structure. The proposed region is projected to resemble the graphical representation of a semantic image representation. This helps the SpineParseNet to generate coarse segmentation. Finally, the two-dimensional ResUNet transforms the segmentation. Figure below shows how spine parsing is done. Figure 11 is the block diagram of the algorithm.

![Image](image2.png)

**Figure 11. Spine parsing (multi class segmentation)**

Spine parsing can be defined as multiple-class segmentation of the intervertebral and vertebrae discs with respect to spine images. In order to remove some of the short comings of the clinical analysis of the volumetric MR images of spines, a three-dimensional graph convolutional segmentation network with the following configuration is used:

- 3D Res U Network
- 3D U-Network
- 3D Graphonomy
- 3D DeepLabv3+ Network

A double-stage training strategy is used to make the SparseNet. To achieve 3D coarse segmentation, the input given was 19×256×128 MR volume and 20 × 18 × 256 × 128 thinly populated probability maps [34-35]. The activation function used for the resultant layer is Softmax. The three-dimensional Res UNetwork, 3D DeepLabv3+ network and 3D U-Network are made to learn with a team size of 4 to 101 iterations using Adam optimizer. Dice similarity coefficient (DSC) and Precision and are used to evaluate the dissection performance. As per the result analysis it is shown that SpineParseNet can achieve a mean of 90.01±6.45%, 92.44±6.55%, and 88.08±6.20% for the IVDs and the vertebrae, and all the other 19 spinal structures dissection respectively. The consequent average precisions with values of 88.3±4.54%, 87.56±4.82%, and 85.38±3.60% are lesser than the average, that depicts that the false-positive rates of the bifurcation are more than the false-negative rates [34]. Moreover, the SpineParseNet acquires mean Dice similarity coefficient of 86.33±3.66%, 84.79±3.64%, and 88.50±2.81% for the IVDs, vertebrae, and all 18 belong to spinal segmentation respectively.

“Flexible Prediction of CT Images from MRI Data Through Improved Neighborhood Anchored Regression for PET Attenuation Corrections [53]” by Liming Zhong et al, in IEEE journal of biomedical and health informatics, April 2020 deals with attenuation correction in PET/MRI images.
This paper proposes a method for foretelling the false computed tomographic pictures from T-2 weighted MRI and T-1 and data is done [35]. The data content of the proposed method is limited. The proposed approach also uses good neighborhood motivated regression as a primary method to calculate projected matrices. This helps to foretell the pseudo image Tomographic patches [36]. The steps of the proposed algorithm follow as enhancement of MR/CT data set, understanding of nonlinear markers of the MR images, hierarchical search of the nearest neighbors, optimization and lastly multi-regressor ensemble. The data space is obtained from the local hospital. Essential pre-processing step includes removing the bias fields advertisements in the MR images followed with intensity normalization. Spatial normalization was also done using linear affine registration. The INAR-E method is used to predict the p Tomographic clips from MRI data [37]. The INAR method is an improvement over the classic method as it handles the not a linear mapping between the MR features and the Tomographic images. If an anchor is given as $\text{AN}^{\text{MR}} \in \mathbb{R}^{k \times d}$, then the training of the surrounding grounded regression can be formulated as

$$\text{arg min} \| \text{AN}^{\text{MR}} - \text{D}^{\text{MR}} \|$$  \hspace{1cm} (10)

The image setups of the intermingled sparse sampling scheme and sampling scheme can be projected based on dense sampling light-based tomography systems as one can use the signals acquired at the coarse detector positions and discard the rest.

There is a round shape transducer array consisting of 128 elements covering a $270^\circ$ angular degree having a radius of 40.5mm. The transducer characters have a mean frequency of 5MHz and a 7-dB bandwidth of 99%. This method when compared with other methods gives a better accuracy rate $\lambda$.

![Image Setup](image.png)

Figure 12. Multispectral Interlaced Sparse sampling

“Contrast Enhancement of Medical Images Using Statistical Methods with Image Processing Concepts [54]” by Zohair Al-Ameen et al, in 6th international engineering conference on sustainable technology and development in 2020, deals with how the foil of medical images can be improved taking into consideration statistical analysis and statistical methods. It says that medical imaging quality is not as per the standard, and that needs to be improved. There are device limitations that makes the quality of the images to have poor contrast [38]. It is also said that using the image processing concepts and some statistical methods the quality of the medical images can be improved for better processing. The dataset for the proposed algorithm consists of degraded quality of images. They are assessed using one specific and without reference metric. Then they are compared with the known contrast enhancement algorithms [39]. It is also shown that this algorithm performed better than the other known algorithms. The proposed algorithm is made using some initial image processing and statistical and methods. This algorithm accomplishes the task by changing the local and the global contrast of the image. Then it combines the outcome using logarithmic image processing. This produces a result that is again processed using flexible linear stretching to form the enhanced image. The low visualized image $i(m,n)$ is firstly analyzed using a CST technique to change the local contrast.

The comparison algorithms used are single scale retina, CLAHE, fuzzy-II and DOTHE. An anonymous data spatial quality determiner is used as a quality evaluator metrics. This assesses the quality of the contrast of the image. The proposed work is simulated using MATLAB 2018a version. The comparison chart of the proposed algorithm with the known four algorithms is also given in the paper that shows the improved performance of this proposed algorithm than the other different algorithm.
“Adaptive Medical Image Deep Color Perception Algorithm [55]” by Xianhua Zeng et al in IEEE Access in March 2020 deals in improving the color of medical images. It is seen most of the medical images are grey. There are various pitfalls in the traditional hand-craft method. This paper proposes a unique technique that continuously colors the grey scale medical images by protecting the content. This is done using fine-tuned deep neural network. This paper claims that it is the first method that colorizes a medical image. The proposed work defines a Yloss parameter that is defined as a nonlinear variable of $l_1$ and $l_2$ norm [40]. This preserves the content invariant between colorized and target data. There is another algorithm called adaptive reference image search algorithm that codes target medical image and reference images with search reference image using hash code. This helps in manual selection of reference images. This method is evaluated in terms of improvement in highest signal to noise ratio and architectural similarity index over the baseline method. The medical images are colored using color representations and Content using deep neural network [56]. The Yloss parameter helps the doctors to do the diagnosis correctly. The reference image is acquired using adaptive method that helps in fully automatic coloring. The flow diagram mentioned below depicts the proposed algorithm.

3.1 Comparison of Techniques on the different accuracy rate

We mention below the table that gives the comparison of different convolutional neural networks used in medical image processing. The study is done based on the accuracy of the output.

| Dataset          | Method Used          | Image Type          | accuracy   |
|------------------|----------------------|---------------------|------------|
| Skin Cancer      | ResNet101            | Clinical Photography| 84.09 %    |
| Brain Images     | 3D-CNN               | MRI                 | 96 %       |
| Heart Disease    | ID3-CNN              | CT                  | 91%        |
| Brain Tumor      | Capsule Network Model| MRI                 | 86.5 %     |
| Retina           | Transfer Learning    | OCT                 | 96.6 %     |
| Liver lesion     | GAN+CNN              | CT                  | 92.4 %     |
| Brain Tumor      | Fuzzy C-means        | MRI                 | 98%        |
| Brain Image      | 2D-CNN               | MRI                 | 94%        |
| Lung Nodules     | Deep Fusion Feature  | CT                  | 96%        |
| Brain Tumor      | DLM                  | MRI                 | 98%        |
| Lung Disease     | Auto Encoder         | CT                  | 94%        |

Table 1. Comparison of different techniques
3.2 Comparison Graph of different CN

![Figure 13. Percentage of Error](image1)

![Figure 14. Accuracy rate](image2)

4. Conclusion and Future Work

This survey has shown the accuracy of different convolutional network that are used as segmentation and classification tool for medical image data. It is also showing the percentage of error and the accuracy rate of different convolutional networks. In the future we propose to use our own convolutional network to analyze cancerous images and also to show that this model provides with more accurate results than the models whose survey is done.

References

[1] Dutta, Sushama Rani, et al. "IoT in Autism Detection in Its Early Stages." Internet of Things: Enabling Technologies, Security and Social Implications. Springer, Singapore, 2021. 47-58.
[2] Dutta, Sushama Rani, Sujoy Datta, and Monideepa Roy IEEE, 2019. "Using cogency and machine learning for autism detection from a preliminary symptom." 2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence).
[3] Dutta, Sushama Rani, et al IEEE, 2017. "A Machine Learning-Based Method for Autism Diagnosis Assistance in Children." 2017 International Conference on Information Technology (ICIT).
[4] Giri, Soumyajit, Sushama Rani, et al IEEE, 2017. "A real time epidemic alert generation system for rural areas using WBANs and kiosks." 2017 International Conference on Information Technology (ICIT).
[5] Dutta, Sushama Rani, and Monideepa Roy IEEE, 2017. "EDOT: Context-aware tracking of similar data patterns of patients for faster diagnoses." 2017 Second International Conference on Electrical, Computer and Communication Technologies (ICECCT).
[6] Dutta, Sushama Rani, and Monideepa Roy IEEE, 2016. "Providing context-aware healthcare services using circular geofencing technique." 2016 3rd International Conference on Computing for Sustainable Global Development (INDIACom).
[7] Dutta, Sushama Rani, et al May 2019." Handling Scarcity of Data in Autism Identification using Binary Imputation Method” International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8, Issue-1.
[8] S.Pawar, M. Bhusan, and M. Wagh,2020 “The plant leaf disease diagnosis and spectral data analysis using machine learning – A review”, International Journal of Advanced Science and Technology, vol, 29, no. 9s, pp. 3343-3359, May 2020..
[9] A. Nalavade, A. Bai, M. Bhusan, 2020”Deep Learning Techniques and Models for improving Machine Reading Comprehension System”, IJAST, vol. 29, no. 04, pp. 9692–9710, Oct. 2020.
[10] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath Jan. 2018, “Generative adversarial networks: An overview,” IEEE Signal Process. Mag., vol. 35, no. 1, pp. 53–65.
[11] M.-Y. Liu and O. Tuzel.2016 “Coupled generative adversarial networks,” in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 469–477.
[12] A. Ghosh, V. Kulharia, V. Namboodiri, P. H. S. Torr, and P. K. Dokania.2018 “Multi-agent diverse generative adversarial networks,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2018, pp. 8513–8521.
[13] P. Isola. J. Zhu, T. Zhou, and A. A. Efros.2017 “Image-to-image translation with conditional adversarial networks,” in Proc. IEEE Conf. Comput. Vision Pattern Recognit., 2017, pp. 1125–1134.
[14] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros.2017 “Unpaired image-to-image translation using cycle-consistent adversarial networks,” in Proc. IEEE Int. Conf. Comput. Vision, 2017, pp. 2223–2232.
[15] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen.2016 “Improved techniques for training GANs” in Proc. Adv. Neural Inf. Process. Syst., 2016, pp. 2234–2242.
[16] M. Heusel, H. Ramsauer, T. Unterthiner, B. Nessler, and S. Hochreiter.2017 “GANs trained by a two time-scale update rule converge to a local Nash equilibrium,” in Proc. Adv. Neural Inf. Process. Syst., 2017, pp. 6626–6637.
[17] Dhungel N, Carneiro G, Bradley AP.2015 “Deep learning and structured prediction for the segmentation of mass in mammograms”. In:International Conference on Medical Image Computing and Computer-Assisted Intervention.Springer, 2015, pp 605-612.
[18] Dou Q, Chen H, Jin Y, Yu L, Qin J, Heng PA.2016“3D deeply supervised network for automatic liver segmentation from CT volumes”. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2016, pp 149–157.
[19] Dou Q, Chen H, Jin Y, Yu L, Qin J, Heng PA.2016“3D deeply supervised network for automatic liver segmentation from CT volumes”. In International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, 2016, pp 149–157.
[20] Cheng JZ, Ni D, Chou YH, Qin J, Tiu CM, Chang YC, Huang CS, Shen D, Chen CM.2016. “Computer-aided diagnosis with deep learning architecture: applications to breast lesions in US images and pulmonary nodules in CT scans”. Sci Rep 6:24454, 2016
[21] Dou Q, Yu L, Chen H, Jin Y, Yang X, Qin J, Heng PA.2017. “3D deeply supervised network for automated segmentation of volumetric medical images”. Med Image Anal 41:40–54, 2017.
[22] L. Ford and D. Fullkerson. Flows in Networks. Princeton University Press, 1962.
[23] A. Goldberg and R. Tarjan. 1988 “A new approach to the maximum flow problem”. Journal of the Association for Computing Machinery, 35(4):921–940, October 1988.
[24] D. Greig, B. Porteous, and A. Seheult.1989. “Exact maximum a posteriori estimation for binary images”. Journal of the Royal Statistical Society, Series B, 51(2):271–279, 1989.
[25] L. D. Griffin, A. C. F. Colchester, S. A. R’oll, and C. S. Studholme.1994. “Hierarchical segmentation satisfying constraints”. In British Machine Vision Conference, pages 135–144, 1994.
[26] R. M. Haralick and L. G. Shapiro.1992. “Computer and Robot Vision”. Addison-Wesley Publishing Company, 1992.
[27] M. Moseley and G. Donnan.2004. “Multimodality imaging: Introduction,” Stroke, vol. 35, no. 11 S1, pp. 2632–2634, 2004.
[28] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. Deep Learning, vol. 1. Cambridge, MA, USA: MIT Press, 2016.
[29] G. van Tulder and M. de Bruijne.2013. “Combining generative and discriminative representation learning for lung CT analysis with convolutional restricted Boltzmann machines,” IEEE Trans. Med. Imag., vol. 35, no. 5, pp. 1262–1272, May 2016.
[30] B. Hu, Y. Tang, E. I. Chang, Y. Fan, M. Lai, and Y. Xu.2019 “Unsupervised learning for cell-level visual representation with generative adversarial networks,” IEEE J. Biomed. Health Inform., vol. 23, no. 3, pp. 1316–1328, May 2019.
[31] C. Lu and M. Mandal.2014 “Toward automatic mitotic cell detection and segmentation in multispectral histopathological images.” IEEE J. Biomed. Health Inform., vol. 18, no. 2, pp. 594–605, Mar. 2014.
[32] J. E. Iglesias, E. Konukoglu, D. Zikic, B. Glocker, K. Van Leemput, and B. Fischl.2013. “Is synthesizing MRI contrast useful for inter-modality analysis?” in Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervention, 2013, pp. 631–638.
[33] K. Kim and H. Myung.1998. “Autoencoder-combined generative adversarial networks for synthetic image data generation and detection of jellyfish swarm,” IEEE Access, vol. 6, pp. 54207–54214, 2018.
[34] D. L. Collins et al., “Design and construction of a realistic digital brain phantom,” IEEE Trans.Med. Imag., vol. 17, no. 3, pp. 463–468, Jun. 1998.
[35] E. Hodneland, E. Hanson, A. Zanna, A. Lundervold, and J. M Nordbotten.2016. “Physical models for simulation and reconstruction of human tissue deformation fields in dynamic MRI,” IEEE Trans. Biomed. Eng., vol. 63, no. 10, pp. 2200–2210, Oct. 2016.
[36] M. Chen, A. Carass, A. Jorj, J. Lee, S. Roy, and J. L. Prince.2017. “Cross contrast multi-channel image registration using image synthesis for MR brain images,” Med. Image Anal., vol. 36, pp. 2–14, 2017.
[37] P. Costa et al.2018. “End-to-end adversarial retinal image synthesis,” IEEE Trans. Med. Imag., vol. 37, no. 3, pp. 781–791, Mar. 2018.
[38] A. Charitsis, T. Joyce, M. V. Giuffrida, and S. A. Tsafaris.2018. “Multimodal MR synthesis via modality-invariant latent representation,” IEEE Trans. Med. Imag., vol. 37, no. 3, pp. 803–814, Mar. 2018.
[39] N. Li, Z. Zheng, S. Zhang, Z. Yu, H. Zheng, and B. Zheng. 2018. “The synthesis of unpaired underwater images using a multistyle generative adversarial network,” *IEEE Access*, vol. 6, pp.

[40] A. F. Frangi, S. A. Tsaftaris, and J. L. Prince. “Simulation and synthesis in medical imaging,” *IEEE Trans. Med. Imag.*, vol. 37, no. 3, pp. 673–679, Mar. 2018.

[41] Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You only look once. 2016 “Unified, real-time object detection”. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

[42] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, arXiv:1505.04597v1 [cs.CV] 18 May 2015.

[43] Martin Rajchl, Matthew C. H. Lee, Ozan Oktay, Konstantinos Kamitsas, Jonathan Passerat-Palmbach, Wenhua Bai, Melissa Damodaram, Mary A. Rutherford, Joseph V Hajnal, Bernhard Kainz, and Daniel Rueckert, “Comparison of Medical Image Segmentation Using Statistical Methods with Image Processing and Artificial Neural Networks.” *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2653–2663, Aug. 2020.

[44] G. Wang et al., “Interactive Medical Image Segmentation Using Deep Learning With Image-Specific Fine Tuning,” in *IEEE Transactions on Medical Imaging*, vol. 37, no. 7, pp. 1562-1573, July 2018, doi: 10.1109/TMI.2018.2791721.

[45] Guotai Wang, Maria A. Zuluaga, Wenchu Li, Rosalind Pratt, Premal A. Patel, Michael Aertsen, Tom Doel, Anna L. David, Jan Deprest, Sebastien Ourselin, Tom Vercauteren, “DeepIGeoS: A Deep Interactive Geodesic Framework for Medical Image Segmentation”, arXiv:1707.00652v3 [cs.CV] 19 Sep 2017.

[46] Deng-Ping Fan , Tao Zhou , G-E Peng Ji , Yi Zhou , Geng Chen , Huazhu Fu , Jianbing Shen, Ling Shao,” Inf-Net:Automatic COVID-19 Lung Infection Segmentation From CT Images”, IEEE transactions on medical imaging, vol. 39, no. 8, August 2020.

[47] Zhijie Zhang , Huazhu Fu , Hang Dai , Jianbing Shen , Yanwei Pang and Ling Shao, “ET-Net: A Generic Edge-aTtention Guidance Network for Medical Image Segmentation”, arXiv:1907.10936v1 [cs.CV] 25 Jul 2019.

[48] Xiangxiang Qin,“ Transfer Learning with Edge Attention for Prostate MRI Segmentation” East China University of Science and Technology.

[49] Justin Ker, Lipo Wang, Hai Rao, and Tchyooson Lim, “Deep Learning Applications in Medical Image Analysis”, *IEEE Access* - December 2017.

[50] G. Wang et al., “A Noise-Robust Framework for Automatic Segmentation of COVID-19 Pneumonia Lesions From CT Images,” in *IEEE Transactions on Medical Imaging*, vol. 39, no. 8, pp. 2653-2663, Aug. 2020, doi: 10.1109/TMI.2020.3000314.

[51] Trinh Le Ba Khanh , Duy-Phuong Dao, Ngoc-Huyinh Ho , Hyung-Jeong Yang * , Eun-Teeun Baek , Gueesang Lee, Soo-Hyung Kim and Seek Bong Yoo, “Enhancing U-Net with Spatial-Channel Attention Gate for Abnormal Tissue Segmentation in Medical Imaging”in *MDPI*.

[52] S. Pang et al., "SpineParseNet: Spine Parsing for Volumetric MR Image by a Two-Stage Segmentation Framework with Semantic Image Representation,” in *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 262-273, Jan. 2021, doi: 10.1109/TMI.2020.3025087.

[53] L. Zhong et al., “Flexible Prediction of CT Images from MRI Data Through Improved Neighborhood Anchored Regression for PET Attenuation Correction,” in *IEEE Journal of Biomedical and Health Informatics*, vol. 24, no. 4, pp. 1114-1124, April 2020, doi: 10.1109/JBHI.2019.2927368.

[54] Z. Al-Ameen, “Contrast Enhancement of Medical Images Using Statistical Methods with Image Processing Concepts,” 2020 6th International Engineering Conference “Sustainable Technology and Development” (IEC), 2020, pp. 169-173, doi: 10.1109/IEC49899.2020.9129295.

[55] X. Zeng, S. Tong, Y. Lu, L. Xu and Z. Huang, "Adaptive Medical Image Deep Color Perception Algorithm," in *IEEE Access*, vol. 8, pp. 56559-56571, 2020, doi: 10.1109/ACCESS.2020.2982187.

[56] K. Verma, S. Bhardwaj, R. Arya, M.S.U. Islam, M. Bhusan, A. Kumar, and P. Samant, “Latest tools for data mining and machine learning”, International Journal of Innovative Technology and Exploring Engineering (IJITEE), vol. 8, no. 9S, pp. 18-23, ISSN: 2278-3075, 2019. Available