Segmentation Techniques of Deep Learning for Medical Image Analysis

Vivek Kumar Sharma[1*], Nisha Vasudeva[2], Garima Sharma[3]

Abstract:

In artificial intelligence, deep learning is the advance process that duplicates the working mechanism of human brain, and it also creates decision making thinking for classification and clustering process. The Deep learning techniques have been used in different fields, like computer vision, diagnosis, prognosis of any disease in medical field and natural language processing. It is become an important research field of computer vision. With the development of deep learning, medical imaging becomes very important and vast research areas. These deep learning algorithms have the facility to avoid outbreaks of illness, recognize and diagnose illnesses, minimize running expenses for hospital management and patients. This paper discusses different deep learning segmentation techniques used in healthcare fields. First, the basic ideas of medical image segmentation based on deep learning are discussed. By discussing its research scope and summarizing the methods of medical image segmentation and their own limitations, the future development direction is expanded. The different pathological tissues and organs are discussed, calculating the specificity between them and their classic segmentation algorithms are summarized. Despite the good achievements of medical image segmentation in recent years, medical image segmentation supported deep learning has still encountered difficulties in research. The wrong segmentation results are unable to meet the actual clinical requirements.

Keywords: Medical imaging, Deep learning, image segmentation

DOI Number : 10.4704/nq.2022.20.14. NQ88033 **Neuro Quantology 2022 ; 20(14):240-251**

1. Introduction

Medical image segmentation aims to form anatomical or pathological structures changes in clearer in images; it often plays a key role in computer-aided diagnosis and smart medicine due to the great improvement in diagnostic efficiency and accuracy. Popular medical image segmentation tasks include liver and liver-tumor segmentation [1], brain and brain-tumor segmentation, blind spot segmentation, cell segmentation, lung segmentation, pulmonary nodules, and cardiac image segmentation. With the event and popularization of medical imaging equipment's, X-ray, computerized tomography (CT), resonance Imaging (MRI) and ultrasound have become four important image-assisted means to help clinicians diagnose diseases, to gauge prognosis, and to plan operations in medical institutions. In practical applications, although these ways of imaging have advantages also as disadvantages, they're useful for the medical examination of different parts of human body. To assist clinicians, make accurate diagnosis, it's necessary to segment some crucial objects in medical images and extract features from segmented areas. Early approaches to medical image segmentation often depend upon edge detection, template matching techniques, statistical shape models, active contours, machine learning, etc. In medical imaging, computerized tomography (CT), MRI imaging, and ultrasound images are prevalent.

These imaging modalities are often used for diagnosis, treatment planning, disease monitoring, and evaluation of response to therapy. Traditionally, radiologists perform these analyses, counting on human-discernible

***Corresponding Author:** Vivek Kumar Sharma
Address: *1,2Department of Computer Science Engineering, Parul University,
3 Department of Computer Science Engineering, Vivekananda Global University
Relevant conflicts of interest/financial disclosures: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest
Received: **Accepted:**

visual features. However, this manual approach requires years of specialized training, and therefore the inherent complexity of these types of images can make certain manual tasks and the subjective process laborious, time consuming, and susceptible to inter observer variability. Furthermore, increasing evidence suggests that the complex quantitative information on medical images is underutilized using current approaches. Deep learning thus offers multiple benefits compared with previous techniques: the power to perform medical image analysis for identification of high-order complex features, performing different classification tasks or predictive modeling, and accelerating image processing tasks. These advantages will facilitate deployment the clinical workflow. This text provides the fundamental background required to understand and develop deep learning models used for medical image processing. This paper covers the different deep learning architectures such as feed forward, recurrent neural networks and their variants. They also provide use cases of such application in medical image processing.

Image segmentation has been a fundamental problem in computer vision since the first days of the field [2]. an important component of many visual understanding systems, it involves partitioning images (or video frames) into multiple segments and objects and plays a central role during a broad range of applications [3], including medical image analysis (e.g., tumor boundary extraction and measurement of tissue volumes), autonomous vehicles (e.g., navigable surface and pedestrian detection), video surveillance, and augmented reality to call a few. Image segmentation are often formulated as the problem of classifying pixels with semantic labels (semantic segmentation), or partitioning of individual objects (instance segmentation), or both (Panoptic segmentation). Semantic segmentation performs pixel-level labeling with a group of object categories (e.g., human, car, tree, sky) for all image pixels; thus, it's generally a more demanding undertaking than whole-image classification, which predicts one label for the entire image. Instance segmentation extends the scope of semantic segmentation by detecting and delineating each object of interest within the image (e.g., individual people).

Objective of the paper

- To provide a comprehensive review and analysis of deep-learning-based image segmentation algorithms
- To overview popular image segmentation datasets, grouped into 2D and 2.5D (RGB-D) images
- To summarize the performances of the reviewed segmentation methods on popular benchmarks
- To discuss several challenges and future research directions for deep-learning-based image segmentation

This paper is organized as follows: Section 2 overviews of machine learning and deep learning techniques in medical imaging. Section 3 reviews the most significant state-of-the-art deep learning-based segmentation models. Section 4 overviews some of the most popular image segmentation datasets and their characteristics. Section 5 lists popular metrics for evaluating deep-learning-based segmentation models and tabulates model performances. Section 6 discusses the main challenges and opportunities of deep learning-based segmentation methods. Section 7 concludes this paper.

2. Techniques in medical imaging

Machine learning algorithms are very effective in medical imaging to review specific diseases. Differing types of entities such as lesions and organs in medical image processing can be too complicated and cannot be shown correctly by a simple mathematical solution. In [4], the author used the pixel-based investigation to research medical images for diseases. The pixel analysis in machine learning appeared in medical image processing, which uses certain values in images right away instead of features extraction from chunks as input data. The enactment of this method could be better than that of simple feature-based classifiers for specific problems. The image with low contrast may be a challenging problem in order to investigate its properties. The feature calculation and segmentation aren't required for pixel-based machine learning, unlike ordinary classifiers which avoid errors generated from inaccurate segmentation and have calculation. The pixel analysis utilizes long training time due to the high dimensionality of data (a large number of pixels in an image) in [5]; the author targeted the low contrast medical images for the analysis. The furthermost efficient technique used for contrast improvement is Histogram Equalization (HE).

The authors proposed a way named "Modified Histogram-Based Contrast Enhancement using Homomorphic Filtering" (MHFIL). It used two phases handling process, within the first phase, global contrast is improved using histogram modification. Further, second phase homomorphic filtering is projected for image sharpening. The low contrast chest X-ray 10 medical images are investigated within the experiment. The MH-FIL has minimum values altogether 10 images computer to other techniques. The medical image clarification is that the highest responsibility of radiologists, with the assignments involving equally images with better quality and its analysis. The CAD has developed for several years. There are numerous machine learning methods analyzed through medical images, for instance, linear discriminant analysis, support vector machines, decision trees, etc.

In [6], the author used machine learning approaches in medical image evaluation. In specific, they used local binary patterns extensively contemplated among texture descriptors. The dataset of neonatal facial images for categorizing pain conditions beginning from facial descriptions. Especially, the outcomes on the extensively premeditated 2D-HeLa dataset and therefore the suggested descriptor gains the maximum implementation including all the numerous texture descriptors. A linear support vector machine classifier is applied on the 2D-HeLa dataset and within the PAP dataset. The 92.4 % accuracy got which is that the highest values among all other descriptors on the mentioned dataset. The neural network technique is employed in medical images to investigate the disease details [7].

The neural network groups are retained for cancer discovery. It's used to critic where a cell is normal with excessive assurance where each distinct network has only two outcomes either it'll be a normal cell or cancer cell. The predictions of those cells' network are merged by a predominant method, i.e., plurality voting. The results showed that the neural network collectively accomplished a high rate of accuracy and a coffee value of false negative analysis.

The machine learning expert systems provide contrivances for the assembly of premises from patients' information. Different rules are mined from the knowledge of specialists to paradigm an expert system. The group of clinical problems which will be used as examples, knowledge in intelligent systems may achieve by machine learning approaches which will be used to generate a methodical description of clinical characters that distinctively describe the clinical circumstances. Therefore, information is often articulated in the arrangement of simple rules, or often as a choice tree. A typical example of this category of the scheme is KARDIO, which is grown to translate ECGs [8].

In medical image analysis, the great standard for evaluating image feature is a statistical analysis. The channelized Hoteling observer (CHO) is usually used for specifically in nuclear medicine imaging. The channels are enthused by the thought of amenable subjects in the human visual structure. This method is employed to detect image quality evaluation and further, the CHO has defensibly and positive influence on the medical imaging. The next algorithm is named a channelized SVM (CSVM). There are two medical physicists assessed the flaw discernibility in 100 noisy images then the score confidence of a lesion actuality contemporary on a six-point scale. Then, a training session is employed to involve an extra 60 images. The human spectators achieved this assignment for 6 diverse selections of the flattening filter with two dissimilar choices of the number of repetitions in the OS-EM rebuilding algorithm [5].

Deep learning in medical imaging

To guide computers to find out features that can characterize the data for the given issue. This concept lies at the foundation of several deep learning procedures. The models comprised of varied layers that transmute input images to give outputs about the specific diseases because of cramming gradually high-level features. The higher type of these models for image analysis is Convolutional neural networks (CNNs). The CNNs comprise several layers that convert the input with convolution filters. The task of employing deep learning methods to the medical field frequently use in familiarizing current architectures in distinctive input formats like three-dimensional data.

Previously, the needs of CNNs to big data, full 3D convolutions and therefore the subsequent huge number of constraints are avoided by separating the volume of Interest into portions [8].

Different methods for segmentation of medical image:

There are different deep learning methods for segmentation of medical image analysis. The methods for segmentation are listed below.

- Region based method
  1. Thresholding
     - 1.1. Local
     - 1.2. Global
  2. Region growing
- Clustering Methods
  1. K-means
  2. Fuzzy C Means
  3. Estimation Maximization
- Classifier Methods
  1. KNN
  2. Maximum likelihood
- Hybrid methods

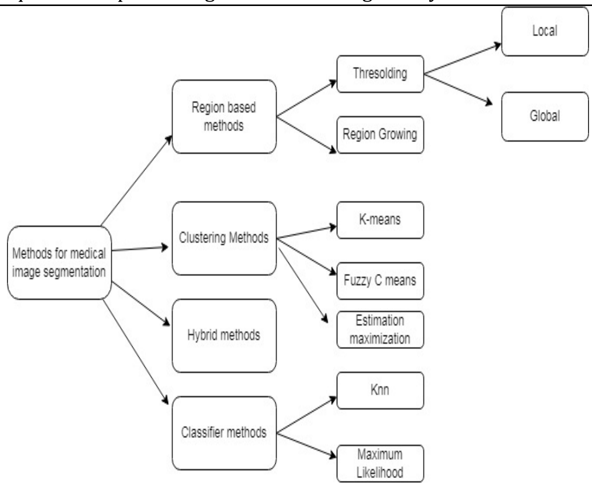

Figure1. Methods of Segmentation

Region based methods

Region-based segmentation algorithms divide the image into sections with similar features. These regions are only a gaggle of pixels and the algorithm find these groups by first locating a seed point which could be a small section or a large portion of the input image. After finding the seed points, a region-based segmentation algorithm would either add more pixels to them or shrink them so it can merge them with other seed points. Based on these two methods, we will classify region-based segmentation into the following categories:

Region Growing

In this method, we start with a little set of pixels and then start iteratively merging more pixels according to particular similarity conditions. A neighborhood growing algorithm would pick an arbitrary seed pixel in the image, compare it with the neighboring pixels and begin increasing the region by finding matches to the seed point.

When a specific region can't grow further, the algorithm will pick another seed pixel which could not belong to any existing region. One region can have too many attributes causing it to require over most of the image. To avoid such a mistake, region growing algorithms grow multiple regions at the identical time.

Threshold Method

The simplest method for segmentation in image processing is the threshold method. It divides the pixels in a picture by comparing the pixel's intensity with a specified value (threshold).

It's useful when the required object has a higher intensity than the background. We can consider the threshold value (T) to be a constant but it would only work if the image has very little noise, unnecessary information and data. We will keep the threshold value constant or dynamic according to our requirements.

The thresholding method converts a grey-scale image into a binary image by dividing it into two segments. According to the different threshold values, we will classify Thresholding segmentation in the following categories:

Simple Thresholding

In this method, we replace the image's pixels with either white or black. The intensity of a pixel at a specific position is less than the threshold value, we replace it with black. On the opposite hand, if it's above the threshold, we'd replace it with white. This is often simple thresholding and is particularly suitable for beginners in image segmentation.

Otsu's Binarization

In simple thresholding, we take a continuing threshold value and used it to perform image segmentation. While the simple method for this is to test different values and choose one, it's not the most efficient one.

Take a picture with a histogram having two peaks, one for the foreground and one for the background. By using Otsu binarization, we will take the approximate value of the middle of those peaks as threshold value.

In Otsu binarization, we calculate the edge value from the image's histogram if the image is bimodal. This process is sort of popular for scanning documents, recognizing patterns, and removing unnecessary colors from a file.

Adaptive Thresholding

We use one constant threshold value won't be a suitable approach to take with every image. Different images have different backgrounds and conditions which affect their properties. Thus, rather than using one constant threshold value for performing segmentation on the entire image, we will have the threshold value

variable. During this technique, we use different threshold values for various sections of an image. This method works well with images that have varying lighting conditions. We'd like to use an algorithm that segments the image into smaller sections and calculates the threshold value for each of them.

Clustering-Based Segmentation Algorithms

They are unsupervised algorithms and help us in finding hidden data in the image that might not be visible to a normal vision. This hidden data includes information like clusters, structures, shadings, etc. A clustering algorithm divides the image into clusters (disjoint groups) of pixels that have similar features. It might separate the data elements into clusters where the elements in a cluster are more similar in comparison to the elements present in other clusters.

Some of the popular clustering algorithms include fuzzy c-means (FCM), k-means, and improved k-means algorithms. In image segmentation, we'd mostly use the k-means clustering algorithm as it's quite simple and efficient. On the opposite hand, the FCM algorithm puts the pixels in several classes according to their varying degrees of membership. The most important clustering algorithms for segmentation in image processing are:

K-means Clustering

K-means may be a simple unsupervised machine learning algorithm. It classifies a picture through a specific number of clusters. It starts the method by dividing the image space into k pixels that represent k group centroids. Then they assign each object to the group supported the distance between them and the centroid. When the algorithm has assigned all pixels to all or any the clusters, it can move and reassign the centroids.

Fuzzy C Means

With the fuzzy c-means clustering method, the pixels within the image can get clustered in multiple clusters. This suggests a pixel can belong to more than one cluster. However, every pixel would have varying levels of

similarities with every cluster. The fuzzy c-means algorithm has an optimization function which affects the accuracy of results. Clustering algorithms can make sure of most of image segmentation needs.

Classification Methods

Classification may be a pattern recognition technique which uses training data to find the patterns. Training data includes a sample of image features with their labels. This system is also called Supervised Learning technique, because it involves training data which are segmented manually then presented to the automatic process. Variety of classifier methods have been used for image processing.

The 2 methods are

• K nearest Neighbor (KNN)

• Maximum Likelihood

The disadvantages of those methods are that they do not take into consideration the spatial information. Another problem is training data which need to be segmented through human interaction. Segmentation of sample data not only takes longer, but also depends on human abilities.

KNN

k-nearest-neighbor (KNN) may be a regular non-parametric and commonly used classification method. This method is understood as a non-parametric method because the KNN algorithm does not need any information about statistical properties of pixels. The KNN algorithm needs an outsized amount of sample data which are labeled as training data. As shown within the figure, each pixel is assessed according to the number of nearest neighbors.

Maximum likelihood

Maximum likelihood estimates the parameter of statistical model. During this method, it's supposed that there is a huge amount of data and we just have a sample set of these data. This method tries to seek out the best estimation for sample data to produce a nearest data model to original data.

Advantages of this method are

- It is often used in wide range of situations
- The approximation of Gaussian distribution and sample variances that can be applied to produce confidence bonds and suggestion tests for parameters.
- Sensitive to starting value selection

One of the disadvantages of this method is that it cannot be applied to small samples of data. Using maximum likelihood method with small training data results in insignificant results.

Deep learning architecture for imaging

Convolutional Neural Network: A Convolutional neural network or CNN consists of a stack of three main neural layers: Convolutional layer, pooling layer, and fully connected layer [9]. Each layer has its own role. The convolution layer detects distinct features like edges or other visual elements in a picture. Convolution layer performs mathematical process of multiplication of local neighbors of an image pixel with kernels. CNN uses different kernels for convolving the given image for generating its feature maps. Pooling layer reduces the spatial (width, height) dimensions of the input file for the next layers of neural network. It doesn't change the depth of the data. This operation is named as sub sampling. This size reduction decreases the computational requirements for upcoming layers. The fully connected layers perform high-level reasoning in NN. These layers integrate the varied feature responses from the given input image so as to provide the final results. Different CNN models are reported in the literature, including Alex Net, Google Net, VGG, Inception, Squeeze Net, and Dens Net. Here, each network uses different number of convolutions and pooling layers with important process blocks in between them. The CNN models are employed mostly for classification task. In [8], Squeeze Net and Google Net are employed to classify brain MRI images into three different categories. the CNN segmentation models performance is restricted by the following: the fully connected layers in CNN cannot manage different input sizes A Convolutional neural network with a fully connected layer cannot be employed for object segmentation task, because the presence of number of objects of interest in the image segmentation task is not fixed; therefore, the length of the output layer cannot be constant.

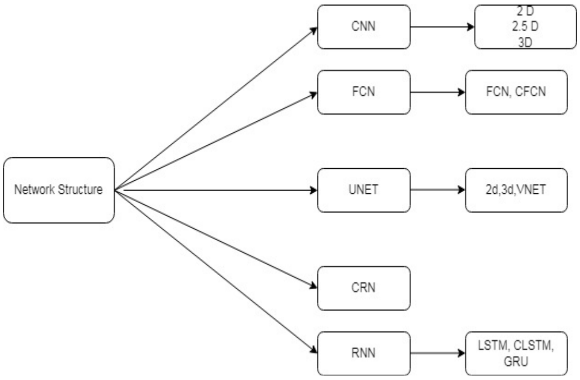

Figure 2: Deep Learning Models for medical Imaging

Fully Convolutional Network: In fully convolutional network (FCN), only convolutional layers exist. The various existing in CNN architectures can be modified into FCN by converting the last fully connected layer of CNN into a fully convolutional layer. The model designed by [11] can output spatial segmentation map and may have dense pixel-wise prediction from the input image of full size instead of performing patch-wise predictions. The model uses skip connections which perform up sampling on feature maps from final layer and fuses it with the feature map of previous layers. The model thus produces an in-depth segmentation in just one go. The traditional FCN model however has the following limitations: It is not fast for real time inference and it does not consider the global context information efficiently. In FCN, the resolution of the feature maps generated at the output is down sampled thanks to propagation through alternate convolution and pooling layers. These leads to low resolution predictions in FCN with fuzziness in object boundaries. A complicated FCN called Parse Net [12] has been also reported; it utilizes global average pooling to attain global context. The approaches incorporating models like conditional random fields and Markov random field into DL architecture have been also reported.

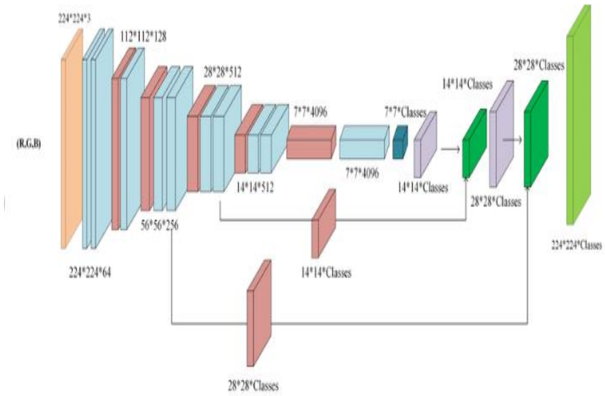

Figure 3: FCN architecture [11]

Encoder-Decoder Models: Encoder-decoder based models employ two-stage model to map data points from the input domain to the output domain. The encoder stage compresses the given input, x to latent space representation, while the decoder predicts the output from this representation. The various types of encoder-decoders based models generally employed for medical image segmentation are discussed as follows:

U-Net: U-Net model [14] features a down sampling and up-sampling part. The down sampling section with FCN like architecture extracts features using 3 × 3 convolutions to capture context. The up-sampling part performs deconvolution to decrease the amount of computed feature maps. The feature maps generated by down-sampling or contracting part are fed as input to up-sampling part so on avoid any loss of information. The symmetric up-sampling part provides precise localization. The model generates a segmentation map which categorizes each pixel present within the image. The U-Net model offers the subsequent advantages: U-Net model can perform efficient segmentation of images using limited number of labeled training images U-Net architecture combines the location

information obtained from the down sampling path and the contextual information obtained from up-sampling path to predict a fair segmentation map U-Net models also have few limitations, stated as follows: Input image size is restricted to 572 × 572 In the middle layers of deeper UNET models, the training generally slows down which causes the network to ignore the layers with abstract features the skip connections of the model impose a restrictive fusion scheme which causes accumulation of the same scale feature maps of the encoder and decoder networks.

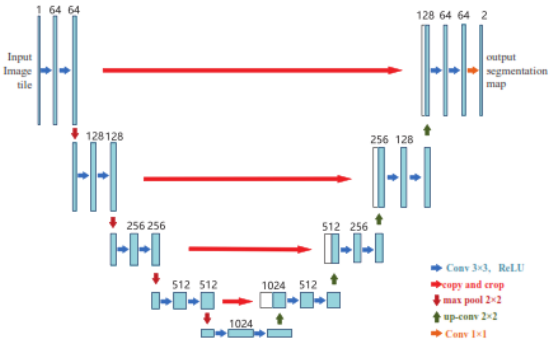

Figure 4: U-Net architecture

VNet: It is also an FCN-based model employed for medical image segmentation [15]. VNet architecture has two parts, compression and decompression network. The compression network comprises convolution layers at each stage with residual function. These convolution layers utilized volumetric kernels. The decompression network extracts feature and expands the spatial representation of low-resolution feature maps. It gives two-channel probabilistic segmentation for both foreground and background regions.

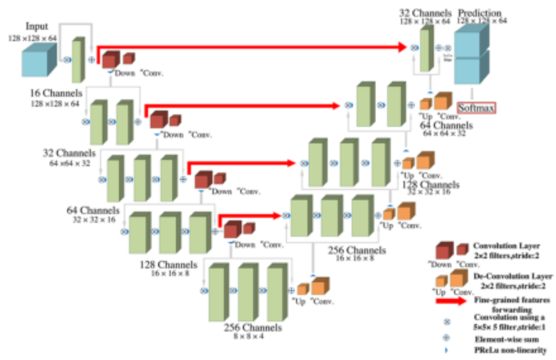

Figure 5: VNet architecture [14]

Regional Convolutional Network: Regional Convolutional network has been utilized for object detection and segmentation task. The R-CNN architecture presented in [14] generates region proposal network for bounding boxes using selective search process. These region proposals are then warped to plain squares and are forwarded to a CNN so as to generate feature vector map as output. The output dense layer consists of features extracted from the image and these features are then fed to classification algorithm so on classify the objects lying within the region proposal network. The algorithm also predicts the offset values for increasing the precision level of the region proposal or bounding box.

Table1: comparison of different algorithm

Algorithm	Description	Strength	Weakness
CNN	This structure consists of three layers- Convolutional layer, pooling layer and fully connected layer	Simple to understand It segments the input image by labeling the pixel	If input image is of different size, then segmentation task is not accurate.
U-Net	This architecture combines information from up sampling path and down sampling path for predicts the segmentation.	Segmentation of image is done by labeling limited number of training image.	The size of input image is fixed and the skip connection of this model.
FCN	Fully Convolutional layers are used.	This model gives a spatial segmentation	Training process is very hard.
V-Net	Convolution is done by kernels.	Applied on 3D data for segmentation.	Less accuracy in large amount of data.
R-CNN	This architecture extracts regions from image.	Prediction of presence of object within the region is accurate.	It uses only selective search algorithm and time taken is more.

3. Medical Image Segmentation Datasets

Data is vital in deep learning models. Deep learning models require great deal of data. The info plays an important role. it's difficult to collect the medical image data as there are data privacy rules governing collection and labeling of data and also it requires time-consuming explanation to be performed by experts [15].

The medical image datasets are often categorized into three different categories: 2D images, 2.5D images, and 3D images. In 2D medical images, each information element in image is named pixels. In 3D medical images, each element is named voxel. The 3D images also are sometimes represented as a sequential series of 2D slices. CT, MR, PET, and ultrasound pixels represent 3D voxels. The pictures may exist in JPEG, PNG, or DICOM format. The medical imaging is performed in different sorts of modalities [16], like CT scan, ultrasound, MRI, mammograms, positron emission tomography (PET), and X-ray of different body parts. MR imaging allows achieving variable contrast image by employing different pulse sequences. MR imaging gives the interior structure of chest, liver, brain, pelvis, abdomen, etc. CT imaging uses X-rays to get the information about the structure and function of the body parts. CT imaging is employed for diagnosis of disease in brain, abdomen, liver, pelvis, chest, spine, and CT based angiography. Mammography may be a technique that uses X-rays to capture the images of the internal structure of the breast. Chest X-rays (CXR) imaging may be a photographic image depicting internal composition of chest which is produced by passing X-rays through the chest and these rays are being absorbed by different amounts of different components in the chest.

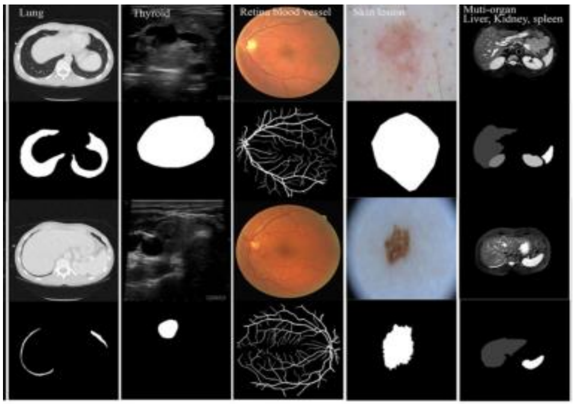

Figure 5: Different medical image dataset [3]

4. Evaluation Metrics

The metric helps in evaluating the performance of any designed model. The metrics provide the accuracy of the designed model. The popular metrics employed for assessing effectiveness of any designed segmentation algorithm is represented in terms of the following:

- True positive (TP) represents that both the actual data class and the class of predicted data are true.
- True negative (TN) represents that both the actual data class and the class of predicted data are false.
- False positive (FP) represents that the actual data class is false while the class of predicted data is true.
- False negative (FN) represents that the actual data class is true while the class of predicted data is false.

Precision: Precision is an evaluation metric that tells us about the proportion of input data cases that are reported to be true and represented in [17].

Precision=$TP/(TP + FP)$

Recall: Recall gives the percentage of the total relevant results which had been correctly classified by the model [2].

Recall=$TP/(TP + FP)$

F1 Score: F1 score talks about models' accuracy as represented in the following equation. It is defined as the harmonic average of the precision and recall values.

$$F1\ score = 2 * precision * recall$$

Pixel Accuracy: It gives the percentage of pixels in a given input image which are correctly classified by the model:

$$Pixel\ Accuracy = no. of\ pixels\ properly \frac{classified}{total\ no\ of\ pixels}$$

Intersection over Union. Intersection over union (IOU) or Jaccard index is a metric commonly used for checking the performance of image segmentation algorithm. It is the amount of intersecting area between the predicted image segment and the ground truth mask, divided by the total area of union between the predicted segment mask and the ground truth mask:

$$Dice = modulus(a \cap b) \div modulus(a \cup b)'$$

Where A represents ground truth. B represents predicted segmentation. Mean IOU is employed for evaluating modern segmentation algorithm. Mean IOU is the average of IOU for each class.

Dice coefficient: It is defined in the following equation and termed as twice the amount of intersection area between the segment predicted and the ground truth divided by the total number of pixels in both the predicted segment and ground truth image:

$$Dice\ coeffcient = 2 * modulus(A \cap B)$$

PA: Pixel accuracy simply finds the ratio of pixels properly classified, divided by the total number of pixels. For $K + 1$ classes (K foreground classes and the background), pixel accuracy is defined as

$$PA = \frac{\sum_{i=0}^{K} p_{ii}}{\sum_{i=0}^{K} \sum_{j=0}^{K} p_{ij}},$$

MSD: It is also known as the Symmetric Hausdorff Distance, and is similar to *ASD* except that the maximum distance that is taken instead of the average:

$$MSD(A,B) = \max\left\{ \max_{s_A \in S(A)} d(s_A, S(B)),\ \max_{s_B \in S(B)} d(s_B, S(A)) \right\}$$

5. Challenges

The medical image segmentation field has gained advantage from deep learning, but still, it's a challenging task to employ deep neural networks due to the following.

Dataset: Deep learning network models require great deal of data. The info required for training is well annotated. The dataset plays a crucial role in various DL based medical procedures [18]. In medical image processing, the gathering of large amounts of annotated medical images is tough. Also, performing annotation on fresh medical images is tedious and expensive and requires expertise. Several large-scale datasets are publicly available. There's still a need of more challenging datasets which can enable better training of DL models and are capable of handling dense objects.

Typically, the prevailing 3D datasets [19] are not so large and few of them are synthetic, so tougher datasets are required. The dimensions of the existing medical image datasets can be increased by:

(a) Application of image augmentation transformations like rotating image by different angles, flipping image vertically or horizontally, cropping, and shearing image. These augmentation techniques can boost the system performance.
(b) The appliance of transfer learning from efficient models can provide solution to the problem of limited data.
(c) Finally comes synthesizing data collected from various sources.

Class Imbalance in Datasets: Class imbalance is intrinsic in various publicly available medical image datasets. A highly imbalanced data poses great difficulty in training DL model and makes model accuracy misleading, for instance, during a patient data, where the disease is comparatively rare and occurs only in 10% of patients screened. The general designed model accuracy would be high as most of the patients do not have the disease and will reach local minima [20]. The matter of class imbalance can be solved by:

(a) Oversampling the data; the quantity of oversampling depends on the extent of imbalance in the dataset.
(b) Second, by changing the evaluation or performance metric, the matter of dataset imbalance can be handled.
(c) Data augmentation techniques are often applied to create new data samples.
(d) By combining minority classes, dataset class imbalance problem also can be handled.

Sparse Annotations: It provides full annotation for 3D images may be a time-consuming task and is not always possible. So, partial labelling of data slices in 3D images is done. it's really challenging to train DL model based on these sparsely annotated 3D images [20]. Just in case of sparsely annotated dataset, weighted loss function is often applied to the dataset. The weights for the unlabeled data within the available dataset are all set to zero, so on learn only from the pixels which are labelled.

Intensity Inhomogeneities: In pathology images, color and intensity inhomogeneities [21] are common. Intensity inhomogeneities cause shading over the image. It's more specific in the segmentation of MR images. Also, the TEM images have brightness variations thanks to presence of nonuniform support films. The segmentation process becomes tedious thanks to these variations.

For correcting intensity inhomogeneities [21], different algorithms are employed and lots of nonparametric techniques are proposed in the literature. Prefiltering operation are often employed before segmentation to remove inhomogeneities. Also, intensity inhomogeneities are taken care of by improvement in scanning devices. In medical images, there could also be different artifacts present during manipulation of images. The different sensors and electronic components used for capturing images create noise within the image. Within the captured image, gray levels are often very close to each other and there may be weak image boundaries.

There could also be overlap in tissues and presence of irregularities like skin lines and hair in dermoscopic images. Of these complexities cause difficulty in identification of region of interest in medical images. To get rid of different artifacts and noises from the image, different image enhancement techniques are used before segmentation. The image enhancement technique suppresses the noise within the image and preserves the integrity of the edges of the image.

Challenges with DL Models: The important challenging issues associated with the training of DNN for robust segmentation of the medical images are as follows: Overfitting the Model. Overfitting of the model refers to the instance when the model learns the small print and regularities in training dataset with high accuracy compared with the unprocessed data instance. It mainly occurs while training the model with a little size training data. Overfitting are often handled by (a) increasing the size of dataset by applying augmentation techniques. (b) Dropout techniques also help in handling overfitting by discarding the output of a number of the random set of network neurons during each iteration.

Memory Efficient Models. Medical image segmentation models require great deal of memory. So as to make these models compatible with certain devices like mobile phones, the models are required to be simplified. Simpler models and model compression techniques can reduce memory requirements for a DL model. The training of deep neural specification needs time. In image segmentation, fast convergence of coaching time for deep NN is required. (E solution to the present problem is (a) application of batch normalization. It refers to locating the pixel values around 0 by subtracting the pixel values from the mean of the image. It's effective in providing fast convergence. (b) Also, adding pooling layers to scale back dimension of parameters can also provide faster convergence.

Vanishing Gradient: Deep neural network faces the matter of vanishing gradient. It occurs because the final gradient loss is not able to be back propagated to earlier layers. The vanishing gradient problem is more pronounced in 3D models. There are several solutions to the matter of gradient vanishing.

(a) By upscaling the intermediate hidden layer output using deconvolution and SoftMax, the auxiliary losses and therefore the original loss of hidden layers are combined to strengthen the gradient value.
(b) Also, by carefully initializing weights, for the network, we will combat the problem of vanishing gradient.

Computational Complexity: Deep learning algorithm performing feature analysis must operate at a high level of computational efficiency. These algorithms need high performance computing devices and GPU. A number of the top algorithms may require supercomputers for training the model, which cannot be available. To combat these issues, the researcher has got to consider the specific number of parameters to attain a limited level of accuracy.

6. Conclusion and future scope

Deep learning-based automated diagnosis of diseases from medical images had become the newest area of research. Within the present work, we had summarized the foremost popular DL based models employed for segmentation of medical images with their underlined advantages and disadvantages. A summary of the different medical image dataset employed for segmentation of diseases and the various performance metrics utilized for evaluating the performance of image segmentation algorithm is also provided. This paper also investigates the different challenges faced in segmentation of medical images using the deep networks and discusses the different state-of-the-art solutions to beat these challenges. With advances in technology, deep learning plays a really important role in segmentation of images. These different studies reviewed the applications of deep neural networks in medical image segmentation task outperform the normal image segmentation techniques. This work will help the researchers in designing neural network architectures in the medical field for diagnosis of disease. Also, the researchers will become aware with the possible challenges within the field of deep learning based medical image segmentation and the state-of-the-art solutions. This review paper provides the reference material and therefore the valuable research in the area of medical image segmentation. The image segmentation techniques have come distant from manual image segmentation to automated segmentation using machine learning and deep learning approaches. These ML/DL based approaches can generate segmentation on large set of images. It helps in identification of meaningful objects and diagnosis of diseases within the images. The image segmentation techniques discussed within the paper can be explored by future researchers for application to various datasets. This future work may include a comparative study of the different existing deep learning models discussed within the paper on the publicly available datasets. Also, different combination of layers and classifiers are often explored to improve the accuracy of image segmentation model. There's still a requirement of an efficient solution to improve performance of image segmentation model. So, the varied new deep learning model designs can be explored by future researchers.

References

[1] Minaee, Shervin, et al. "Image segmentation using deep learning: A survey." IEEE transactions on pattern analysis and machine intelligence (2021).
[2] Li, Wen. "Automatic segmentation of liver tumor in CT images with deep convolutional neural networks." Journal of Computer and Communications 3.11 (2015) : 146.
[3] Wang, Risheng, et al. "Medical image segmentation using deep learning: A survey." IET Image Processing 16.5 (2022): 1243-1267.
[4] Fu, Huazhu, et al. "Joint optic disc and cup segmentation based on multi-label deep network and polar transformation." IEEE transactions on medical imaging 37.7 (2018): 1597-1605.
[5] Song, Tzu-Hsi, et al. "Dual-channel active contour model for megakaryocytic cell segmentation in bone marrow trephine histology images." IEEE Transactions on Biomedical Engineering 64.12 (2017): 2913-2923.
[6] Amyar, Amine, et al. "Multi-task deep learning based CT imaging analysis for COVID-19 pneumonia: Classification and segmentation." Computers in Biology and Medicine 126 (2020): 104037.
[7] GAO, Lianli, et al. "Unsupervised urban scene segmentation via domain adaptation." Neurocomputing 406 (2020): 295-301.
[8] Fathi, Alireza, et al. "Semantic instance segmentation via deep metric learning." arXiv preprint arXiv: 1703.10277 (2017).
[9] Dai, Jifeng, and et al. "R-fcn: Object detection via region-based fully convolutional networks." Advances in neural information processing systems 29 (2016).
10] Li, Xia, et al. "Expectation-maximization attention networks for semantic segmentation." Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
[11] Huang, Chao, et al. "3D U $$^ $$-Net: A 3D Universal U-Net for Multi-domain Medical Image Segmentation." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2019.
[12] Xu, Xiaowei, et al. "Quantization of fully convolutional networks for accurate biomedical image segmentation." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.
[13] Li, Xiaomeng, et al. "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes." IEEE transactions on medical imaging 37.12 (2018): 2663-2674.
[14] He, Kaiming, et al. "Mask r-cnn." Proceedings of the IEEE international conference on computer vision. 2017.
[15] Mahapatra, Dwarikanath, et al. "Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network." International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer, Cham, 2018.
[16] Kavur, Ali Emre, et al. "Chaos-combined (ct-mr) healthy abdominal organ segmentation challenge data." Med. Image Anal 69 (2019): 101950.
[17] Wong, Chun Lok. "3D segmentation with exponential logarithmic loss for highly unbalanced object sizes." U.S. Patent No. 10,803,591. 13 Oct. 2020.
[18] Chen, Xu, et al. "Learning active contour models for medical image segmentation." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2019.
[19] Nandi, A., et al. "DefED-Net: Deformable Encoder-Decoder Network for Liver and Liver Tumor Segmentation." (2021).
[20] Ros, German, et al. "The synthia dataset: A large collection of synthetic images for semantic segmentation of urban scenes." Proceedings of the IEEE conference on computer vision and pattern recognition. 2016.
[21] Abrams, Jesse F., et al. "Habitat-Net: Segmentation of habitat images using deep learning." Ecological informatics 51 (2019): 121-128.