iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52686

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Efficient 3D Medical Image Segmentation using CoTr: Bridging CNN and Transformer

Sri Lasya Avula[1], Gajjala Akshara[2], Mr. S. Sreehari[3], Dr. SVR. Manimala[4]
[1, 2]*Electronics and Communication, M. V. S. R Engineering College, Osmania University, Hyderabad, India*
[3]*Assistant Professor,* [4]*Associate Professor, Electronics and Communication, M. V. S. R Engineering college, Osmania University, Hyderabad, India*

Abstract: ***Neural networks are a subset of machine learning, and they are at the heart of deep learning algorithms. Before CNNs, identifying objects in images was done manually using time-consuming, manual feature extraction methods. The superior performance of convolutional neural networks, when dealing with images, speech, or audio signals sets them apart from other neural networks. Convolutional neural networks (CNNs) have been the de facto standard for nowadays 3D medical image segmentation. Due to the inductive bias of locality and weight sharing inherent in convolutional operations, these networks lose the ability to model long-range dependency. In this study, a novel framework is presented for accurately segmenting 3D medical images based on the combination of a convolutional neural network and a transformer (CoTr). This framework allows us to construct CNNs for extracting feature representations, and Vision Transformers for modelling long-range dependency on the extracted feature maps. As a self-attention device, the transformer performs a global operation where it draws information from all the information on the system in order to make a decision.***
Keywords: Transformer, Convolutional Neural Networks, Semantic Segmentation, U-net, Medical Image segmentation

Abbreviations
CNN- Convolutional Neural Networks
ReLU- Rectified Linear Unit
MRI- Magnetic resonance imaging
CT- Computed tomography
ConvNet- Convolutional Neural Networks
FCN- Fully convolutional network
VGG- Visual Geometry Group

I. INTRODUCTION

Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. As a result, image segmentation occurs when each pixel in an image is assigned a label that indicates certain characteristics shared among pixels with the same label. To put it simply, segmentation involves labelling pixels. A common label is assigned to each element or pixel belonging to the same category. Until recently, segmenting medical images was a challenging process. A number of the following challenges are encountered when acquiring medical images in the majority of imaging modalities: (i) low resolution (both spatially and spectrally), (ii) high levels of noise, (iii) low contrast, (iv) geometric deformations, and (v) imaging presence.

Image Segmentation Techniques can be classified into five major categories i.e., Thresholding, clustering, Region Based Techniques and Edge Based Techniques and Pixel Based Techniques.

Medical Image Segmentation has become vital in the subject of Medical Analysis of various organs. Medical image segmentation is a process of breaking down an image into multiple segments or regions with similar characteristics. Medical image segmentation recognises boundaries within 2D and 3D images in order to determine essential features and sizes of objects therein, accounting for the variability in medical imaging. This has significantly aided medical research, diagnosis, and computer-assisted surgery. The development of deep learning algorithms has improved the performance and accuracy of medical picture segmentation, which has produced amazing new developments in the medical industry. The Medical Images such as CT or MRI undergo this kind of segmentation to identify and extract regions of interest, such as organs, tissues, or lesions, to aid in diagnosis and treatment. Deep Learning Neural Networks such as CNN, FCN, VGG, RE-Net, AlexNET can be used for this purpose.

© IJRASET: All Rights are Reserved |

However, Fully Convolutional Neural networks, particularly U-Net, have become the most common method for Medical image segmentation. These algorithms, referred to as ConvNets or CNNs, are deep learning networks that can assign significance to different features in an input image. Through the use of learnable weights and biases, CNNs can accurately label individual pixels in an image, making them particularly adept at complex segmentation tasks. Thanks to their high accuracy and flexibility, ConvNets are now a standard tool for image analysis and are widely used in medical imaging, object recognition, and other applications.

A. Background Of Brain Tumour Detection Methods

Machine Learning and Deep Learning algorithms can detect brain tumors through analysis of MRI images, leading to faster and more accurate predictions for treatment. This technology is improving patient outcomes and reducing the time and costs associated with traditional diagnosis methods. MRI scans are a powerful tool for thorough examination of various parts of the body. Compared to other imaging methods, they are particularly useful in detecting early-stage abnormalities in the brain. This allows for early intervention and treatment, potentially saving lives and preventing more serious health issues down the line.

1) AlexNet and Transfer Learning

This study is used for automatically detecting the pathological brain in Magnetic Resonance Images (MRI) based on deep learning structure and transfer learning. But, the volume of brain MRI datasets are usually too small to train the entire deep learning structure. Hence, transfer learning is introduced to train the deep neural network. First, the pre-trained AlexNet structure is obtained and then the parameters of the last three layers are replaced with random weights and the rest of the parameters serve as the initial values. Finally, the modified model is trained with the MRI dataset

2) Automated Brain Tumour Detection using FLAIR MRI

A new automated method is developed to detect and segment abnormal brain tissue from FLAIR MRI scans. The process is fully automated, which means it is quicker and potentially more accurate than previous methods. The technology is particularly useful for detecting lesions that might indicate multiple sclerosis, brain tumors, or other neurological disorders. This methodology could significantly improve the accuracy of diagnoses and help clinicians plan treatments more effectively. A method for classifying brain MRI images is developed using super-pixel technique and feature extraction from each super-pixel. Features such as intensity, Gabor textons, fractal analysis and curvatures are calculated to ensure accurate classification across the entire brain area in FLAIR MRI. The method offers a robust approach to classifying brain MRI images and has potential applications in medical research and clinical practice.

3) Pathological brain detection based on wavelet entropy and Hu moment invariants

Developing an accurate pathological brain detection system, the paper has proposed a novel automatic computer-aided diagnosis (CAD) to detect pathological brains from normal brains obtained by magnetic resonance imaging (MRI) scanning. We used wavelet entropy (WE) and Hu moment invariants (HMI). For feature extraction, and the generalized eigenvalue proximal support vector machine (GEPSVM) for classification. To further enhance classification accuracy, the popular radial basis function (RBF) kernel was employed. The 10 runs of k-fold stratified cross validation result showed that the proposed "WE + HMI + GEPSVM + RBF" method was superior to existing methods w.r.t. classification accuracy.

4) Automatic detection of brain contours in MRI data sets

A software procedure is presented for fully automated detection of brain contours from single-echo 3-D MRI data, developed initially for scans with coronal orientation. The procedure detects structures in a head data volume in a hierarchical fashion. Automatic detection starts with a histogram-based thresholding step, whenever necessary preceded by an image intensity correction procedure. This step is followed by a morphological procedure which refines the binary threshold mask images. the discrimination between desired and undesired structures, is implemented in this step through a sequence of conventional and novel morphological operations, using 2-D and 3-D operations. A final step of the procedure performs overlap tests on candidate brain regions of interest in neighboring slice images to propagate coherent 2-D brain masks through the third dimension.

5) Feed-forward neural network optimized by hybridization of PSO and ABC for abnormal brain detection

A novel automatic classification system based on particle swarm optimization (PSO) and artificial bee colony (ABC), with the aim of distinguishing abnormal brains from normal brains in MRI scanning.

© IJRASET: All Rights are Reserved |

The proposed method used stationary wavelet transform (SWT) to extract features from MR brain images. SWT is translation-invariant and performed well even the image suffered from slight translation. Next, principal component analysis (PCA) was harnessed to reduce the SWT coefficients. Based on three different hybridization methods of PSO and ABC, we proposed three new variants of feed-forward neural network (FNN), consisting of IABAP-FNN, ABC-SPSO-FNN, and HPA-FNN. The 10 runs of K-fold cross validation result showed the proposed HPA-FNN was superior to not only other two proposed classifiers but also existing state-of-the-art methods in terms of classification accuracy.

B. Development of Hybrid Model of CoTr: CNN and Transformer

A. Medical Image Segmentation model using only Transformer

Medical image segmentation using Transformers is a relatively new and emerging area of research. Transformers have shown impressive performance in natural language processing and have the potential to achieve high accuracy in medical image segmentation as well. Transformers can be scaled up to process large volumes of medical image data, which is important in clinical settings where there is often a high demand for accurate and efficient image segmentation. Transformers are designed to be more interpretable than other deep learning models. But Transformers require a large amount of labelled data to train effectively. The self-attention mechanism in transformers allows for efficient and flexible processing of input sequences, resulting in higher accuracy and reduced computation requirements compared to traditional CNNs. With continued advancements in the technology, transformers are expected to revolutionize the field of medical image analysis and improve patient outcomes. Transformers require significant computational resources, including powerful hardware such as GPUs, to train effectively. This can be expensive and limit the availability of these models to some researchers and healthcare institutions.

B. Medical Image Segmentation using only CNN

CNNs are capable of learning complex features and patterns from medical images, making them highly accurate in segmenting the region of interest. CNNs can process large amounts of medical images in a short amount of time. CNNs can adapt to different types of medical images, allowing for their use in a variety of medical imaging applications. But, CNNs require a large amount of labelled data to train effectively. In some cases, acquiring sufficient labelled medical images for training can be challenging. While CNNs can provide highly accurate segmentation, the complexity of the models can make it challenging to interpret the features and patterns that the model has learned. Like any segmentation method, CNNs are not perfect and can produce false positives or false negatives. CNNs are widely used in image recognition, but they have limitations due to their focus on local features and shared weights. As a result, they struggle to capture long-range connections between image elements. This can limit their utility in certain applications

C. Medical Image Segmentation model using both CNN and Transformer

Combining the strengths of both CNNs and Transformers can lead to higher accuracy in medical image segmentation compared to using either model alone. While both CNNs and Transformers require significant computational resources, combining the models can result in more efficient processing of large amounts of medical image data. Combining CNNs and Transformers can result in a more generalizable model that can be applied to different medical imaging datasets and clinical settings. The combination of CNNs and Transformers can be scaled up to process large volumes of medical image data, making it a useful tool in developing models that can be deployed across different healthcare settings. To overcome the drawbacks of CNNs, Transformers are used. Transformers have become a popular alternative to convolutional neural networks (CNNs), particularly in the medical image analysis field. They have been applied successfully to various clinical applications such as image synthesis, registration, segmentation, detection, and diagnosis.

II. DESIGN METHODS

A. Deep Learning

The ability of medical image segmentation has previously been constrained by a number of major problems, including unpredictability in medical imaging, variability in human tissue, noise between image pixels, and intrinsic uncertainty resulting from knowledge gaps in the medical field. Although these problems might always exist, deep learning has made it possible for image segmentation to produce better outcomes than ever before, and its potential is much greater than that of the algorithms that came before it. By labelling each pixel with an item class like "heart," "tumour," or "artery," medical image segmentation is the process of identifying key elements inside medical images.

© IJRASET: All Rights are Reserved |

We can train models using deep learning to automatically and accurately assign labels to pixels. These advancements have allowed the performance of automatic image segmentation to match that of professionally trained radiologists. The main objective to propose a model that has a few predominant features that exceed the pre-existing models. Hence, we've used U-Net to train the network, which provides us with an efficient image segmentation and organ detection

B. CNN

Convolutional Neural Network (CNN or ConvNet) is a deep learning model primarily used for analysing visual data. Unlike traditional neural networks, CNN uses convolutional layers to reduce the input data size and extract important features. The technique involved in CNN is called Convolution, where filters or kernels slide over the input data to identify patterns. This method reduces the number of parameters involved and allows for faster training of the network. CNNs have shown great success in tasks like image recognition, object detection, and natural language processing, making them one of the widely used deep learning models. A CNN can be constructed into three major layers that are a convolutional layer, a pooling layer and a fully connected layer

C. U-Net

The convolutional neural network (CNN) designed and applied in 2015 is specifically tailored for processing biomedical images, which is a more complex task than simple image classification. In addition to identifying whether a disease is present, it must also locate the affected area within the image. In response to these particular requirements, this specialised CNN provides accurate and precise medical diagnosis. It represents a significant improvement in the field of medical imaging
analysis and interpretation due to its capacity to analyse enormous amounts of medical imagery. U-Net is committed to addressing this issue. The reason deep learning models such as object detection networks are able to localise and distinguish borders is because they perform classification on every pixel, resulting in an input and output with the same size. While it is possible to convert an image into a vector and use the same mapping to convert it back into an image, object
detection networks are able to provide more detailed and accurate information due to their ability to perform pixel-wise classification.
U-Net uses a rather novel loss weighting scheme for each pixel such that there is a higher weight at the border of segmented objects. This loss weighting scheme helped the U-Net model segment cells in biomedical images in a discontinuous fashion such that individual cells may be easily identified within the binary segmentation map.
U-NET is a deep learning architecture that is commonly used for image segmentation tasks. The network features a symmetric U-shape design, with four encoder blocks and four decoder blocks connected via a central bridge. The encoder blocks down-sample the image and increase the number of feature channels, while the decoder blocks progressively up-sample and recover the original spatial resolution.

D. Transformer

Transformers are networks that function with data sequences, such a group of words. These word groups are tokenized first, after which they are supplied into the transformers. Transformers add Attention which is a quadratic operation that determines the pairwise inner product of any pair of tokenized words. Stacks of transformer blocks make up transformers. The "Multi-Head Attention" encompass blocks, which are multilayer networks composed of simple linear layers, feedforward networks, and self-attention layers i.e., the main invention of transformers. The number of procedures also rises as the number of words does. Transformers are therefore more difficult to train images on. Pixels make up an image, and each image can include tens of thousands to millions of pixels. Each pixel in a transformer will perform a paired operation with each and every other pixel in the picture. An attention mechanism will expend (500*2) *2 operations on an image with a dimension of 500*500 pixels. Even with many GPUs, this is a massive task. As a result, local focus rather than global attention is typically used when viewing visuals. Additionally, this is accomplished by focusing global attention on various areas of the image rather than the entire picture.

E. Semantic Segmentation

Semantic segmentation is a computer vision technique that assigns a label or category to every pixel in an image. It is commonly used for tasks such as object recognition, scene understanding, and autonomous driving. Semantic segmentation offers an advantage over object detection as it enables the detection of objects that span multiple areas of an image at the pixel level. This technique helps in detecting the object of interest even when it may not be completely visible or surrounded by similar objects.

© IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |

The algorithm is able to distinguish between different objects and segments them accordingly, allowing for more precise analysis and manipulation of the image data. The ability to accurately segment images can enable more advanced computer vision capabilities and improve overall performance in a variety of applications. This is different from other types of image segmentation that group pixels based on their properties such as colour or texture.

Furthermore, semantic segmentation can provide accurate boundary detection, making it ideal for applications such as self-driving cars and medical imaging. Overall, semantic segmentation can be a useful alternative to object detection for a variety of computer vision applications. A further intriguing feature of semantic segmentation is that it is a difficult process because of how variable and complex natural images are. The wide variety of item appearances, forms, and sizes as well as the presence of occlusions, clutter, and other visual artefacts cause the variability. Semantic segmentation is commonly used as a pre-processing step in other computer vision tasks, such as object recognition or scene understanding

F. Ground Truth Images

Ground truth refers to the actual nature of the problem that is the target of a machine learning model, reflected by the relevant data sets associated with the use case in question. These are also called masks. These masks are the ground truth labels for the images. A mask consists of all the pixels belonging to the particular class. One binary mask for dog class output, the other for horse class. For an image of dimension h x w, the output masks for a recognition task for k classes would be h x w x k array. K masks of dimensions h x w each.

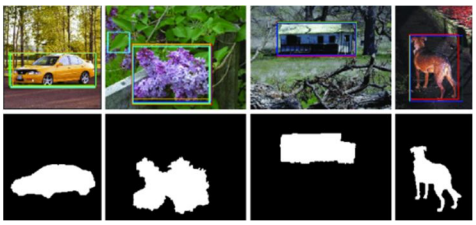


Figure 1: Ground Truth Images

III. SOFTWARE REQUIREMENTS

MatLab Software was used in order to simulate and obtain the results. MATLAB is a widely used platform that enables engineers and scientists to perform numerical computing tasks, build algorithms, analyse data, and develop models. Due to its versatility, MATLAB has gained widespread use across numerous fields, including science, engineering, finance, and data analysis. This integrated environment provides an efficient and effective way to solve complex problems due to its user-friendly interface.

MATLAB is an interactive programming environment for scientific computing such as Data analysis, problem solving, experimentation, and algorithm spanning across many technical fields. A medical image analysis involves analysing medical images in order to extract meaningful information, usually through computation. With MATLAB®, you can easily build algorithms for medical image analysis using the development environment and built-in analysis and data access capabilities. A user-friendly computing platform, MatLAB is equipped with multiple inbuilt image processing and analytical tools strewn across numerous libraries inbuilt into the program

IV. DESIGN IMPLEMENTATION

A. Training a ConvNet: U-Net

The U-Net model used here has 26 convolutional layers in total. The architecture of U-Net comprises of an encoder and decoder segments. The architecture of this particular convolutional neural network starts out with the encoder segment, a dropout layer and a decoder segment. 13 convolutional layers are used in the encoder part of the architecture, while the remaining 13 convolutional layers are used at the decoder part of the architecture. This training model of unit can be known as a mirror structure where the encoder starts out the novel network. The network is built primarily with 2 main layers, a Conv-In layer and a Max Pool layer, where the Conv-In layer consists of a convolutional layer, batch normalisation layer and a Re-Lu layer (Rectified Linear unit). The convolutional layer consists multiple kernel parameters which are to be learned throughout the model. The function of Max pooling layer is to pick out the maximum value from the pool of pixels. This layer returns the most prominent features in the pile of pixels. Therefore, the resultant output image is sharper than the original image. This reduces the number of parameters to learn through the model system and hence reduces the spatial and time complexities. The network encompasses 5 layers of down-sampling at the encoder and 5 layers of up-sampling at the decoder.

© IJRASET: All Rights are Reserved |

After the final layer of up-sampling, a Soft max layer is used. The results generated by a neural network can be challenging to comprehend due to their raw, unprocessed nature. The soft-max activation function simplifies this by making the neural network's outputs easier to interpret. As a result of soft-max activation, the raw outputs from the neural network are transformed into a vector of probabilities, which is essentially a probability distribution over the input classes. Multi-class problems are solved with Soft-max by assigning decimal probabilities to each class. These decimal probabilities add up to 1.0. By imposing this constraint on training, the process converges faster than it would otherwise. The output of this network is a pool of pixel labels which are used to segment the medical image. In conclusion, a U-Net model is successfully trained for a training dataset of 60 images and the validation dataset consists of 8 images. The data paths for training and testing datasets are set. Other hyper-parameters set with the values specified, Epoch=5, Learning rate= 1e-4, Batch size= 8, Height= 224, Width= 224.

B. Inputs and Gray scale conversion

The in-built MATLAB command "uigetfile" is used for selecting an input image from a specific file.

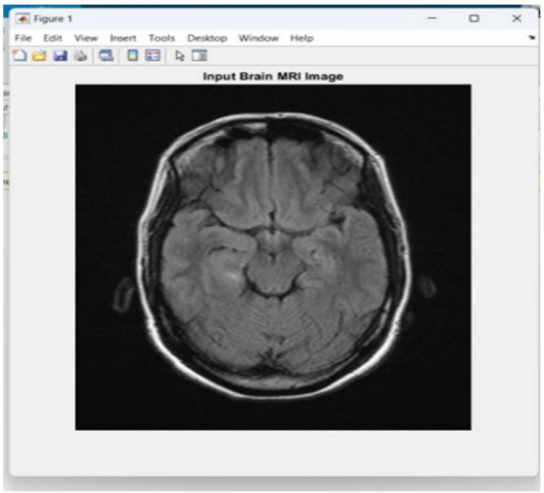


Figure 2: Input Image

If the image is constituted as grayscale, it proceeds into the next stage. But, if the input image is a colour image, then it is being converted into grayscale mainly to decrease the computational complexity, as a colour image operates on 3 planes, say RGB, which calculates each pixel operating cost to be 24 bits. Whereas in grayscale each pixel has an operating cost of 8 bits.

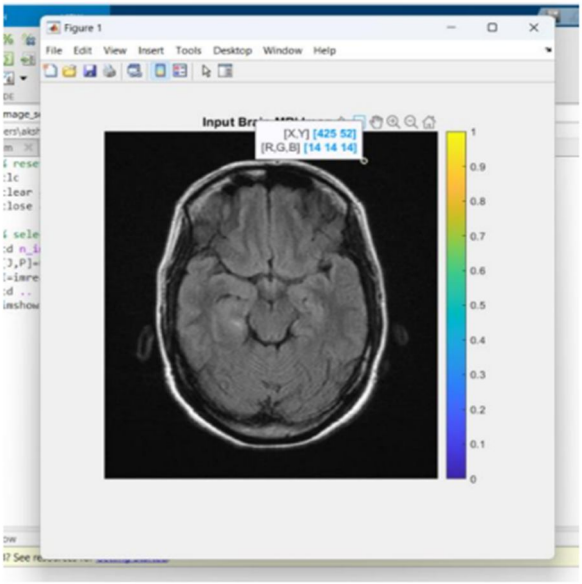


Figure 3: Input Colour Image

© IJRASET: All Rights are Reserved |

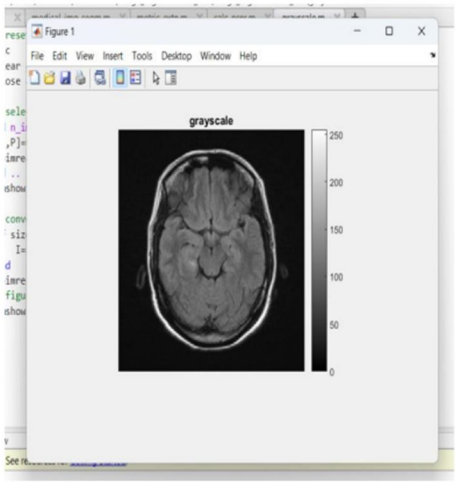


Figure 4: Input Colour Image Converted into Gray-scale Image

C. Transformer

One of the most important features of a transformer is that, it reduces the noise complexity of a image. To test the functionality of the transformer in use, a random noise is added to the input image. The noise is generated by using the MATLAB in-built function "rand", which generates an array of random numbers of the given specific datatype. The command "im2double" is used to convert the image to double precision.

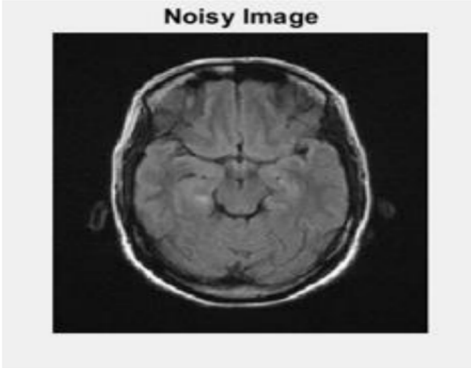


Figure 5: Noisy Image

The noisy image is fed to the transformer input. At this stage, parameters like peak signal-to-noise ratio (PSNR) and mean square error (MSE) are calculated. The transformer output closely resembles the input image that has been converted to grayscale i.e before the random noise has been added to the image.

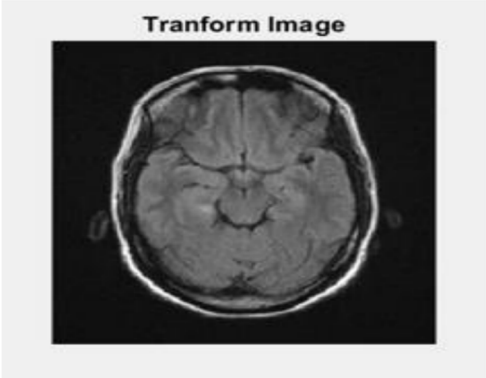


Figure 6: Denoised Image by Transformer

D. Semantic Segmentation

The transformed image is semantically segmented. This segmentation divides the image into different group of pixels that represent distinct categories. After the semantic segmentation is performed, the selection criteria that is considered is, if a group of pixels contains less than 12 pixels, the group is discarded. And consequently, if the group of pixels contains greater than 12 pixels, the group is considered as an organ and undergoes further testing. At this stage, the image is converted into a binary image.

Figure 7: Detected Brain Region

E. Comparison to Optimum Image

The converted binary images are compared to the manual images that are provided by the doctors. The manual images are also called as ground truth images. Ground truth refers to the target of the deep learning model. At this stage, parameters like accuracy, sensitivity, specificity are calculated.

Figure 8: Manual Image

V. DESIGN PERFORMANCE METRICS

A. MSE (Mean Square Error)

MSE is the difference between actual data and obtained data. The formula for calculating MSE is given below.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - \breve{y}_i)^2$$

Where, n is number of data points.

yi is the actual value of data

B. PSNR (Peak Signal to Noise Ratio)

PSNR is calculated with the help of MSE and maximum amplitude. The formula for calculating PSNR is given below.

$$PSNR = 20\ log_{10}(^{MAX}/_{\sqrt{MSE}})$$

C. Confusion matrix

A table called the confusion matrix is used to assess how well a machine learning model performs in classification tasks. The confusion matrix provides a summary of the number of accurate and inaccurate predictions for each class by comparing the predicted labels of the model to the actual labels of the data.

The matrix shows how many true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN) the model generated using the test data. The matrix is a 2X2 table for binary classification. In the case of multi-class classification, the matrix shape is an nXn, where n is the number of classes.

		Predicted class: *P*	Predicted class: *N*
Actual Class	*P*	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

Figure 9: Confusion Matrix

D. SI (Sensitivity)

Absolute quality of the input image is nothing but sensitivity. It is calculated as below by confusion matrix.

$$SI = \frac{2\ Tp}{2\ Tp + Fp + Fn} * 100\%$$

E. Accuracy

Amount of uncertainty calculated by the proposed absolute standard is accuracy. It is calculated as below by confusion matrix

$$AC = \frac{Tp + Tn}{Tp + Fp + Fn + Tn} * 100\%$$

F. Specificity (SP)

Determines how negative the predicted value is with respect to the actual value. It is calculated as below by confusion matrix.

$$SP = \frac{Tn}{Tn + Fp} * 100\%$$

G. Extended Parameters

1) CDR (Correct Detection Ratio)

The degree of trueness for given input image is called as CDR and calculated as below,

$$CDR = \frac{Tp}{Tp + Fn} * 100\%$$

2) USE (Under segmented Error)

This is the ratio of falsely detected pixels to manual detected and can be calculated as,

$$USE = \frac{2Fn}{Tp + Fn} * 100\%$$

3) OSE (Over segmented Error)

OSE is the ratio of pixels not identified by proposed work to manual identified and can be represented as below,

$$OSE = \frac{Fn}{Tp + Fn} * 100\%$$

4) TSE (Total Segmented Error)

TSE is calculated as the sum of USE and OSE.

$$TSE = USE + OSE$$

VI. RESULTS

The below figure shows the different stages, the input image goes through to arrive at the final image where the brain region is detected.

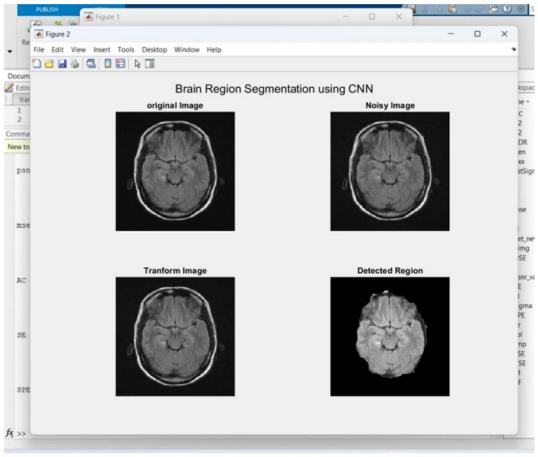


Figure 9: Results

A. *Accuracy, Sensitivity & Specificity*

Metrics	CNN with Transformer (Proposed)	CNN (Existing)
Accuracy	96.37%	96.15%
Sensitivity	96.52%	91.52%
Specificity	96.29%	96.57%

Table for Accuracy, Sensitivity, and Specificity of proposed and existing methods.

B. Extended Parameters

Metric	Value	Optimum value
Peak signal-to-noise ratio (PSNR)	42.0599	30- 50 dB
Mean square error (MSE)	2.5182e-04	Lower the MSE value, better the model
Correct detection Ratio (CDR)	0.9651	0.9- 1.0
Similarity Index (SI)	0.9441	0.9- 1.0
Under segmentation error (USE)	0.0796	Least
Over segmentation error (OSE)	0.0348	Least
Total segmentation error (TSE)	0.1144	Least

Table for existing parameters calculated

VII. CONCLUSION

Medical analysis-computer vision is an active field of research that deals with the deep understanding of image content. Medical Images have been a priliminary factor in survival against many diseases. These medical images also assisted in learning normalities and abnormalities among the human body, which led the medical field to safeguard the nature of predicting and identifying a fatal disease. Learning human anatomy included understanding and researching over the specific organ without any external or internal disturbances such as fluid excess, any factor in the background etc., It includes various subdomains such as object detection, recognition, image understanding, classification, segmentation etc. With the advancement in technologies like CT, MRI etc., images are considered as important part of clinical applications. A hybrid model of CNN Transformer is proposed, namely CoTr, for efficient medical image segmentation and organ detection. In this proposed model, we design a network which employs U-Net as a training model, followed by Transformer which employs self-attention mechanism to reduce the computational and spatial complexities for modelling long range dependency on multiscale and high resolution feature maps. The combination of CNN and Transformer overcome the disadvantages of each other when merged together, resulting in an efficient framework compared to each of them individually. The training model is modified to consider fewer training images in the dataset and yielding a far more precise segmentation, making the model user friendly and easily accessible. We have used semantic segmentation algorithm where only pixels groups which consisted of a value higher than 12 are considered to be an organ and the rest are discarded. In addition, the proposed CoTr is scheduled to achieve balance in keeping details of low-level feature maps, resulting in a superior performance in organ detection, namely brain, for purposes of research and interpretation of analysis. From this perusal on organ detection and its importance, we've understood that brain region can be best segmented and detected with minimal errors by using CNN- UNet and Transformer algorithm This proposed network of CoTr, successfully detects the region of interest, i.e., brain region with an accuracy of 96%.

VIII. FUTURE SCOPE

Future Scope of 3D Medical Image Segmentation using CoTr is reassuring as it allows to compute medical diagnosis easier and accordingly. The most prospective application of this model includes: Brain Tumour detection and Segmentation: Brain tumour segmentation is the process of separating the tumour from the normal brain tissues in clinical routine, it provides useful information for diagnosis and treatment planning. The main goal of brain tumour segmentation is to detect the location and extension of the tumour regions, namely active tumorous tissue, which is done by identifying abnormal areas when compared to normal tissue. The self-attention mechanism employed in the proposed CoTr, has the capacity to model long-range dependencies which is very important for precise brain tumour segmentation.

© IJRASET: All Rights are Reserved |

REFERENCES

[1] Daning Li, Jiaoyang Du, Xiangyu Gao, Wen Gu, Fanfan Zhao, Xiaojie Feng, Hong Yan: An Intelligent Diagnosis Method of Brain MRI Tumor Segmentation Using Deep Convolutional Neural Network and SVM Algorithm

[2] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou,Y.: Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)

[3] Fang, X., Yan, P.: Multi-organ segmentation over partially labeled datasets with multi-scale feature abstraction. IEEE Transactions on Medical Imaging 39(11), 3619–3629 (2020)

[4] Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European conference on computer vision (ECCV). pp. 801–818 (2018)

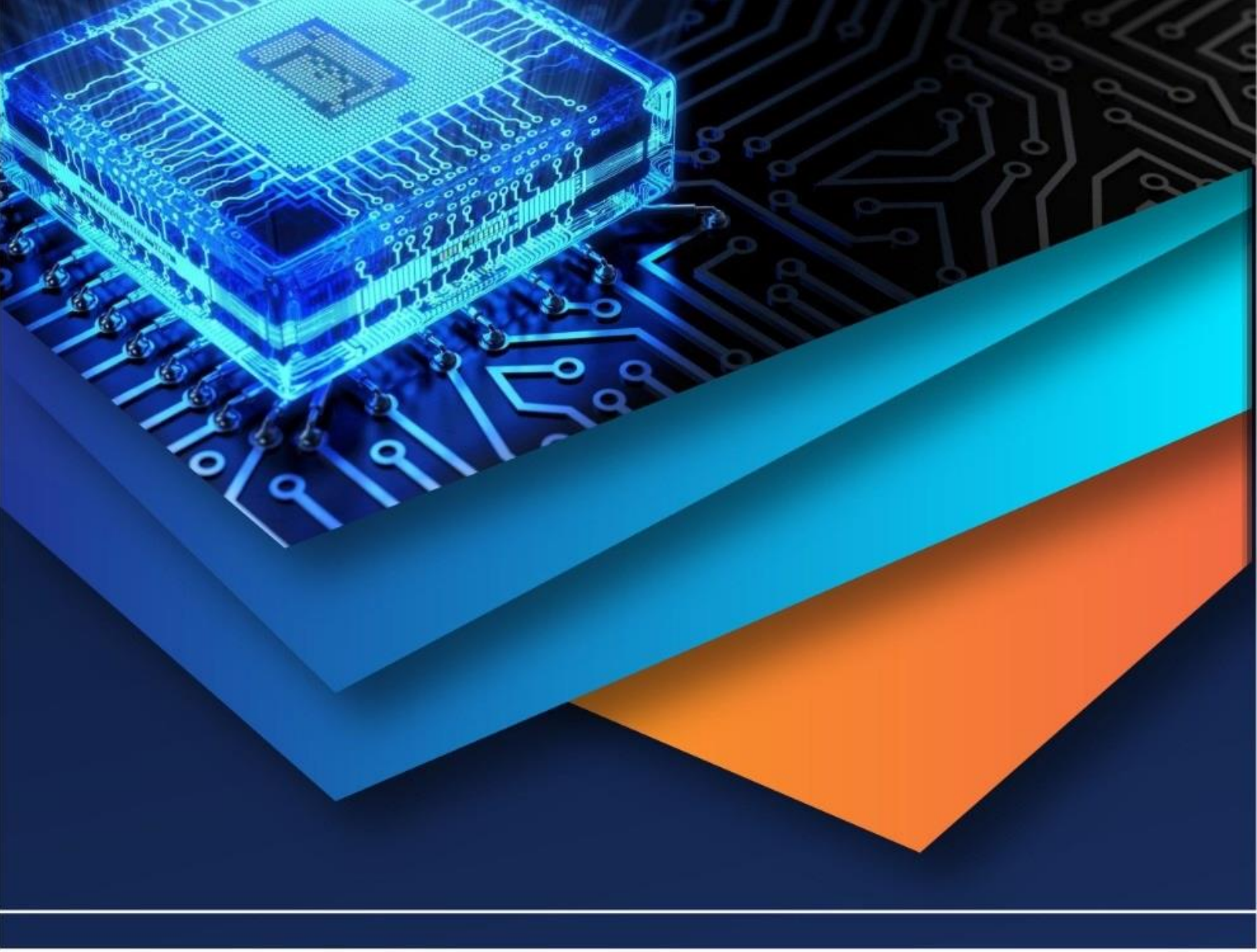
[5] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)

[6] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)

[7] Ytong Xie, Jianpeng Zhang, Chunhua Shen, Yong Xia, CoTr: Efficiently Bridging CNN and Transformer for 3D Medical Image Segmentation, In: International Conference on Medical Image Computing and computer assisted intervention, arXiv: pp 171–180, 2022

[8] Study of Techniques used for Medical Image Segmentation and Computation of Statistical Test for Region Classification of Brain MRI, Anamika Ahirwar, I.J. Information Technology and Computer Science, 2013, 05, 44-53

© IJRASET: All Rights are Reserved |



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)