JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# A survey of UNet in medical images

**Riya K Azeez[1], Bismi P R[2]**
[1,2]Department of Computer Applications- M E S College Marampally,
[1]riyaazeez1997@gmail.com, [2]bismi15011@gmail.com

***Abstract***: UNet has become a very popular architecture for medical image segmentation due to its high accuracy and ability to handle complex anatomical structures. Medical image segmentation is the process of identifying and delineating specific structures or regions of interest within medical images, such as CT scans or MRI images. UNet has been used for a variety of medical image segmentation tasks, including segmentation of brain tumors, liver and spleen segmentation, cardiac segmentation, and segmentation of blood vessels in angiography images. The use of UNet in medical image segmentation has led to significant improvements in accuracy and speed compared to traditional segmentation methods and has enabled more accurate diagnosis and treatment planning for a variety of medical conditions. One advantage of UNet is its ability to handle the class imbalance problem that is commonly encountered in medical image segmentation tasks, where the number of pixels belonging to the target class is much smaller than those belonging to the background. This paper conducts a formal review of the concept of unet in medical images.

**Keywords: UNet, deep learning, machine learning**

## I. I INTRODUCTION

Machine learning plays a crucial role in the UNet architecture for image segmentation tasks. UNet is a deep neural network that is trained using a machine learning approach known as supervised learning. In supervised learning, the model is trained using a labeled dataset, which consists of pairs of input images and corresponding output masks or labels. The model learns to map the input images to their corresponding masks or labels by minimizing a loss function that measures the difference between the predicted masks and the ground truth masks. The UNet architecture consists of a contracting path and an expanding path, which are both composed of convolutional layers. The contracting path captures the context of the input image by reducing its spatial resolution while increasing its feature maps, while the expanding path upsamples the feature maps back to the original spatial resolution of the input picture. During training, the UNet model updates its weights by backpropagating the error from the output to the input, using an optimizer such as stochastic gradient descent. The objective of the optimization is to minimize the loss function, which measures the difference between the predicted masks and the ground truth masks. Once the model is trained, it can be used to segment new images by passing them through the network and obtaining the predicted masks or labels. The accuracy of the segmentation depends on the quality of the training dataset and the architecture of the UNet model.

The foundational element of the UNet architecture for picture segmentation is CNN or Convolutional Neural Network. A fully convolutional network, the foundation of the UNet architecture, is made up of several convolutional layers that capture the spatial and contextual details of the input picture. To extract features from the input image and create feature maps, which reflect the learned features at various scales, the UNet uses convolutional layers. The UNet architecture also has skip connections, which enable the network to keep the incoming image's fine-grained details. The network can combine high-level features from the contracting path with low-level features from the expanding path by connecting matching layers in the contracting and expanding pathways using skip connections. By using pairs of input images and matching output masks or labels, the CNNs used in UNet are trained using a supervised learning methodology. The CNN adjusts its weights during training to reduce the discrepancy between the expected and actual masks. Usually, a backpropagation algorithm variant like stochastic gradient descent is used for this training procedure.

UNet is a deep-learning architecture that has been widely used for medical image segmentation tasks. The process of locating and separating particular areas or structures within medical images, such as organs, tumors, or blood vessels, is known as medical image segmentation. For several medical uses, such as diagnosis, treatment planning, and image-guided interventions, accurate segmentation is essential. Because of its high precision and capacity for handling complicated anatomical structures, the UNet architecture has grown in popularity for biomedical image segmentation tasks. A fully convolutional neural network with a contracting route and an expanding path form the foundation of the UNet architecture. While the expanding path up samples the feature maps back to the input image's initial spatial resolution, the contracting path down samples the feature maps to capture the context of the input image while increasing its feature maps. The UNet architecture also has skip connections, which enable the network to keep the incoming image's fine-grained details. The network can combine high-level features from the contracting path with low-level features from the expanding path by connecting matching layers in the contracting and expanding pathways using skip connections. Medical image segmentation jobs involving the segmentation of organs in CT or MRI scans, blood vessels in angiography or retinal images, and tumors in PET or ultrasound images have all been tackled using UNet. It has been demonstrated that UNet performs these jobs with accuracy and efficiency, frequently outperforming other segmentation algorithms. Finally, UNet is a powerful deep-learning architecture for medical image segmentation tasks, enabling accurate and efficient segmentation of complex anatomical structures within medical images.

## II U-Net and Its Variants for Medical Image Segmentation

The "UNET 3+: A Full-scale Connected UNET for Medical Image Segmentation" is a method proposed in a research paper published in 2020 by[1] Huimin Huang, Lanfen Lin, Ruofeng Tong, Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen, and Jian Wu. The method is designed for medical image segmentation, which is the task of delineating regions of interest in medical images, such as identifying tumors or organs. The UNET 3+ model is an extension of the original UNET architecture, which is a convolutional neural network (CNN) widely used for image segmentation tasks. The UNET 3+ model incorporates several modifications to enhance its performance. The key features of the UNET 3+ model are, Full-scale skip connections, It is the UNET 3+ model uses skip connections that connect all the encoder and decoder blocks, allowing for better information flow across different levels of abstraction in the network. This helps to preserve fine details and contextual information, which are important for accurate segmentation. Next Dilated convolution, it is the UNET 3+ model employs dilated convolutions in the encoder blocks, which allows for increased receptive field without sacrificing spatial resolution. This helps the model capture both local and global contextual information for improved segmentation accuracy. Deep supervision, it is the UNET 3+ model introduces deep supervision by adding auxiliary classifiers at different depths of the decoder blocks. These auxiliary classifiers provide additional supervision signals during training, helping the model to learn more robust and accurate representations. Next the Attention mechanism, it is the UNET 3+ model incorporates attention mechanisms, such as channel attention and spatial attention, to selectively emphasize informative features and suppress irrelevant information, which helps to improve the model's ability to focus on important regions during segmentation. Post-processing techniques, it is the UNET 3+ model applies post-processing techniques, such as morphological operations and conditional random fields (CRF), to refine the segmentation results and generate more accurate and visually appealing segmentations. The UNET 3+ model was evaluated on various medical image datasets, and the results showed that it achieved state-of-the-art performance in terms of segmentation accuracy and computational efficiency. The method proposed in the paper is expected to have potential applications in medical image analysis and clinical practice. the future scopes of the proposed method could involve further validation, comparison, exploration of different techniques, clinical integration, and optimization, to enhance its applicability, reliability, and effectiveness in medical image segmentation and other related tasks.

Later [2] Yuanfeng Ji, Ruimao Zhang, Zhen Li, Jiamin Ren, Shaoting Zhang, and Ping Luo propose a general framework for searching neural architecture for segmentation of 3D medical images, called UXNet, which searches scale-related feature aggregation strategies and block-wise operators in the encoder-decoder path. The carefully designed search space achieves robust segmentation results. In addition, the discovered segmentation architecture reveals the properties of information propagation for the specific data set. The further will explore UXNet under real-world hardware constraints, such as memory, speed, and power consumption. In addition, the task-oriented search space design will alsorepresent a possible research direction. UXNet proposes an approach that combines multi-level feature aggregation, NAS, skip connections, 3D CNNs, and loss functions to automatically search for an optimal architecture and achieve accurate segmentationof 3D medical images, which has the potential to improve the performance of medical image segmentation tasks. In[9] 2020 Yutong Cai and Yong Wang proposed a multiscale mechanism**.** Comparing the ma unet model with the unet series models proposed in recent years, the experimental results show the results of the model. This paper has achieved better results than the previous models confirming the effectiveness of the method. This model uses the more lightweight attention unet as basic networkarchitecture. The multiscale mechanism achieves better results than other latest segmentation networks**.**

Next in[4]2021 levit unet was proposed by Guoping Xu, Xingrong Wu, Xuan Zhang, and Xinwei He. In this article, levit unet integrates a levit transformer module into the unet architecture for fast and accurate segmentation of medical images. Levit unet achieves better performance comparing various methods. Levit is passed into the decoder via a skip connection, which can effectively reuse the spatial information of the feature map. Later[5] Nahian Siddique, Sidike Parading, Colin p elkin, and Vijay Devabhaktuni proposed a U-net based architecture that is quite ground-breaking and valuable in medical image analysis. The growth of U-net papers since 2017 lends credibility to its status as the leading deep learning technique in medical image diagnosis. Thus, despite the many challenges remaining in deep learning-based image analysis, we expect U-net to be one of the major paths forward.[7] In 2021 Ange Lou, Shuyue Guan, and Murray Loew proposeddc unet. Dc unet is a potential successor to the unet architecture. It is an enhanced version of unet which is the most popular in the d successful deep learning model. Dcunet is more efficient compared with multresunet and classical unet. The performance of dc unet is better compared to 2 other unets. So dc unet architecture can be an effective, model for medical image segencoder-

decodermentation. Future work of this papers test the model for more datasets. In 2021[8] Song-Toan Tran, Ching-Hwa Cheng Minh-Hai Le, and Don-Gey Liu proposed another paper tmd unet. Tmd-unet, which had three main improvements compared to unet. This paper uses 7 different datasets and they have better efficiency.

Later in 2022[6] Launch of AFter-UNet, an end-to-end framework for segmenting medical images.The proposed framework uses an axial fusion mechanism to fuse intra-layer and inter-layer context information to drive the final segmentation process. Experiments on three datasets demonstrate the effectiveness of our models compared to previous work. Later in [3]2023, Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang presented a Swin-Unet that uses a U-shaped encoder-decoder-like pure transformer-based U-shaped Medical imaging encoder-decoder is segmentation. The tokenized image blobs are fed into the transformer-based U-shaped coder-decoder architecture with skip connections for local-to-global semantic feature learning. Use hierarchical swin transformer with shifted windows as the encoder to extract context features and symmetric swin transformer based decoder with patch expanding layer is designed to perform the upsampling operation to restore the spatial resolution of the feature maps dataset in synapse multiorgan ct image. In this swin unet best performance with segmentation accuracy 79.13 %.compared with transunet and att unet. Swin unet has excellent performance with an accuracy of 90.00 %. This paper uses a 2d medical image dataset. Swin unet has excellent performance and generalization ability. Future work of the paper is swin unet used the 3d dataset images.

## *IV.CONCLUSION*

In this paper, we have discussed the concepts of different types of units in medical image segmentation as of 2020. The U-Net 3+ is an improved version of the original U-Net that incorporates skip connections, feature fusion modules, and other enhancements to improve segmentation performance. In UXNet is a CNN architecture that combines U-Net with the eXtremeNet architecture, which is designed for object detection tasks. UXNet uses a U-Net-like encoder-decoder structure with additional feature fusion modules to improve accuracy. In Levit U-Net is a CNN architecture that combines the Levit architecture, which is designed for image classification tasks, with the U-Net architecture for image segmentation. It uses a hierarchical feature extraction approach and incorporates skip connections for better performance. After U-Net is a variant of the U-Net architecture that introduces additional residual connections after each stage of the encoder and decoder. These residual connections are designed to improve the flow of gradients during training, leading to better convergence and segmentation accuracy. DC U-Net, or Dilated Convolution U-Net, is a variant of the U-Net architecture that incorporates dilated convolutions, which have larger receptive fields and can capture more context, in both the encoder and decoder. This can help improve the accuracy of image segmentation, especially for objects with varying scales. TMD U-Net stands for Temporal Multi-scale Dilated U-Net and is designed for time-series data, incorporating dilated convolutions in the temporal dimension to capture temporal dependencies in addition to spatial features. MA U-Net stands for Multi-attention U-Net, and it incorporates multiple self-attention mechanisms to capture long-range dependencies and improve feature representation in the encoder and decoder pathways.. Swin U-Net combines the U-Net architecture with the Swin Transformer, which is a recent and promising architecture for large-scale image classification, to capture long-range dependencies and global context efficiently. From this, the swin-unet generated in 2023 is more efficient, accurate, and better performing. But in this paper, 2D medical images are used. In the future, we can use 3D images

.

## REFERENCES

[1] Huimin Huang, Lanfen Lin, Ruofeng Tong Hongjie Hu, Qiaowei Zhang, Yutaro Iwamoto, Xianhua Han, Yen-Wei Chen3,4,1, Jian Wu1 "UNET 3+: A FULL- SCALE CONNECTED UNET FOR MEDICAL IMAGE SEGMENTATION "(2020)

[2] Yuanfeng Ji, Ruimao Zhang, Zhen Li, Jiamin Ren, Shaoting Zhang, and Ping Luo "UXNet: Searching Multi-level Feature Aggregation for 3D Medical Image Segmentation" (2020)

[3] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang "Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation"(2023)

[4] Guoping Xu, Xingrong Wu, Xuan Zhang, Xinwei He "LEVIT-UNET: MAKE FASTER ENCODERS WITH TRANSFORMER FOR MEDICAL IMAGE SEGMENTATION"(2021)

[5] NAHIAN SIDDIQUE, SIDIKE PAHEDING, COLIN P. ELKIN AND VIJAY DEVABHAKTUNI "U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications"(2021)

[6] Xiangyi Yan Hao Tang Shanlin Sun Haoyu Ma Deying Kong Xiaohui Xie "AFTer-UNet: Axial Fusion Transformer UNet for Medical Image Segmentation" (2022)

[7] Ange Lou, Shuyue Guan, Murray Loew "DC-UNet: Rethinking the U-Net Architecture with Dual Channel Efficient CNN for Medical Images Segmentation"(2021)

[8] Song-Toan Tran, Ching-Hwa Cheng Minh-Hai Le, and Don-Gey Liu "TMD-Unet: Triple-Unet with Multi-Scale Input Features and Dense Skip Connection for Medical Image Segmentation" (2021)

[9] Yutong Cai, Yong Wang "MA-Unet: An improved version of Unet based on multi-scale and attention mechanism for medical image segmentation" (2020)

[10] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L. Yuille, Yuyin Zhou "TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation"(2021)

[11] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl1, Jens Petersen and Klaus H. Maier-Hein "nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation"(2021)

[12] Ze Liu1, Yutong Lin1, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin Baining Guo "Swin Transformer: Hierarchical Vision Transformer using Shifted Windows"(2021)

[13] Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, Ling Shao "Pyramid Vision Transformer: A Versatile Backbone for Dense Prediction without Convolutions"(2021)

[14] Wenxuan Wang, Chen Chen, Meng Ding, Hong Yu, Sen Zha, Jiangyun Li "TransBTS: Multimodal Brain Tumor Segmentation Using Transformer"(2021)

[15] Yutong Xie, Jianpeng Zhang, Chunhua Shen, Yong Xia "CoTr: Efficiently Bridging CNN and Transformer for 3D Medical Image Segmentation" (2021)