A New Approach for Liver Plus Its Tumor Segmentation in CT Image by TransNUNet

Yaqi Hu
Faculty of Information Technology
Macau University of Science and Technology
Macau SAR, China
1809853di011003@student.must.edu.mo

Xiaoniu Yang
Faculty of Information Technology
Macau University of Science and Technology
Macau SAR, China
2009853gii20009@student.must.edu.mo

Xiaolin Tian
Faculty of Information Technology
Macau University of Science and Technology
Macau SAR, China
xltian@must.edu.mo

***Abstract*—Computer aided diagnosis (CAD) of human liver and its tumor can provide information for both patients and doctors and help them in disease diagnosis, disease treatment, disease tracking, etc. However, manually done a medical reporting of liver and tumor will cost plenty of time and need expertise to finish the work and reduce the errors. Automatic liver and its tumor segmentation provide efficient way to solve the problem, but there still exist difficulties due to CT image shape, like there is little difference between the healthy part and the diseased part. Our research focuses on computer-only segmentation of liver and its liver tumor in human abdominal CT image. Based on the traditional U-Net network with one attention mechanism, the research has further added one Cbam architecture into the model. We divided the dataset into two parts, first 100 patient samples are used as the training dataset, and last 30 patient samples are used as the valid and testing dataset. (Official dataset provided 130 patient samples in total). We should notice that the dataset not only contains the patient samples that have the tumors, but also samples that do not have. The model was first been trained several times, then validated, and tested with last patient samples using the Liver Tumor Segmentation Challenge (LiTS-2017) dataset. For dice score of liver and its tumor segmentation in training, validated and tested was 98%, 91%, respectively, from the images of LiTS dataset. We provided a solution to deal with the liver and its tumor segmentation problem with computer-only method.**

***Index Terms*—CT Image, Liver Segmentation, Tumor Segmentation, Convolutional neural networks, Deep Learning**

I. Introduction

Up to now, the prevalence of liver diseases is gradually increasing, especially the liver tumor which caused lots of death. At present, the datasets used for training and testing are all labelled and well-tested by experienced doctors, but there still are several challenges such as the contrast difference of images and low image quality. Moreover, manually identification and segmentation of liver tumor and its volumes is a challenging work, in that case, automatic liver and its tumor segmentation system is widely used in hospitals. To overcome the challenges, several methods have been implemented to deal with the problems, including convolutional neural networks like U-Net [1], 3D U-Net [2] architecture and random forest classifier, which provided a solution to segment the liver and tumors.

At first, researchers were all work at solving the segmentation problems to see whether can get a full segmentation of liver. After getting a solution to segment the liver and tumor, many research teams were aimed at improving the accuracy of models using combined machine learning architectures and techniques, therefore, many DL methods are proposed in different experiment. However, limited by the computational power and model design techniques, most CNN models performed poorly than the traditional machine learning algorithm, i.e., SVM or RF. Nowadays, with the great increasing of computational power and appearance of new models, deep learning models performs best in biomedical image segmentation challenge than traditional ways and take nearly most fields. Plenty of s-o-t-a CNN models are proved perform best in their regions, including nature language processing (relation extraction, event extraction), image processing (object classification), computer vision, etc. What's more, the demand of huge datasets to train the deep learning models to let them perform well in different tasks also influence many traditional industries like manufacturing and medicine.

However, big datasets and new deep learning models with deeper layers make deep learning a time-consuming and complexity work, and due to data collection and other technologies, it is also difficult to get a well-labelled dataset to train, test and valid the model, also, it is difficult to get a good solution. In addition, deep learning suffers from a lack of generalization and overfitting when trained with a small dataset. Therefore, training a deep learning model perfectly is hard in real time experiments.

To get a better result, our researchers not only optimize the model with its network architecture but add new mechanisms such as attention mechanism to help model learn image features. The attention mechanism is aimed at increasing the weight of the feature learned by the model. It can also ignore some unnecessary noise information.

U-Net is one deep learning model currently used in biomedi-

This research is supported by the Science and Technology Development Fund of Macao (FDCT-18-032-FI-001A/0009/2018/A).

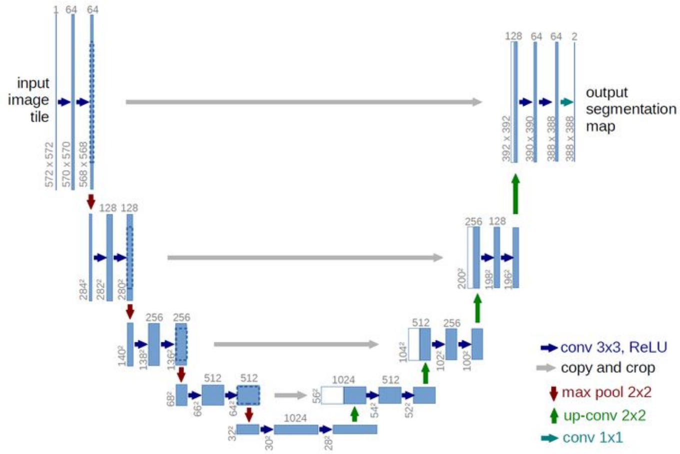


Fig. 1: The U-Net Architecture [1]

cal image segmentation and provides bases for several state-of-the-art outcomes. Fig.1 shows a symmetric "U" shape design of U-Net, consist of two components : down-sampling, the encoder (left) and up-sampling, the decoder (right). The output of each layer of up-sampling (encoder) is connected in series with the corresponding layer of down-sampling (decoder). Instead, output from the second layer are also connected. In addition, U-Net performs well in image location even with a compared small dataset.

II. Related Work

Overall, for liver and its tumor segmentation challenge, several research teams have proposed their own models and methods to solve the problem. Most of them use machine learning and deep learning models to split the liver and its tumors. Segmentation of liver and liver tumor is a challenging task, especially the segmentation of liver tumor. For most research teams, they can segment the overall contour of the liver very well, and the accuracy is often very high. However, for the segmentation of liver tumor, due to the low contrast between the tumor and the surrounding liver, the size of the tumor is different, and the distribution is uneven, research teams often cannot get a good segmentation result. According to that, several preprocessing methods is used to enhance the image. Most of them use CNN model to deal with the tough problem in different ways.

Christ et al. [9] given a FCN with 3D dense conditional Random Field method which combined them together. What's more, Gruber, N. et al. [8] proposed a two UNet joint way to finish the task. They first using one-step UNet to get a discrete class label output and then using another UNet to further segment the tumor in the liver.

In addition, after changing the network shape of different models, to better enhance the performance of networks, several new mechanisms are introduced into the medical image processing region, especially the transformer architecture [4]. Georg Hille et al. proposed one hybrid CNN with transformer layers named SWTR-UNet to solve joint liver and hepatic lesion segmentation tasks. They have included MSA blocks, a kind of self-attention with multi-head method in the network between the encoder and decoder stages to help them get a better outcome. Asima Bibi et al. proposed one added attention mechanism U-Net for automatic liver tumor segmentation. They have added one high resolution inception block with an attention block to conduct one Acon-U-Net. Li Jiajian et al. proposed a 2.5D Network with scale attention awareness. A 2.5D ResNet 34 was used as the encoder and one scale attention perception module was added between encoder and decoder stages.

What's more, researchers also found that context related features also enhance network performance in image segmentation tasks. It can not only helps model capture image features, but also assist convolutional network recognize objects from complex background. As shown in Fig.2, Cbam [5] is a convolutional block attention module found by researchers nowadays and added to different networks. It uses both channel attention and spatial attention and change the connection method from parallel to serial. For medical image segmentation tasks, we are focus on organ shape and tumor/lesion region, therefore, inserting Cbam module into the network can improve the weight of target features, suppress other irrelevant information, and help the model obtain better segmentation outcome.

III. Methodology

The method TransNUNet used in this research is changed from the U-Net based transformer network named TransUNet. Like the U-Net structure, TransNUNet also includes two stages, down-sampling, and up-sampling, as shown in Fig. 3. However, unlike TransUNet, we select CNN as its encoder and decoder, with a new attention mechanism named Cbam in the up-sampling stage (four blocks) to enhance the performance. More specifically, the network using CNN for preliminary feature extraction, and get object maps at the end of the down-sampling stage, which size is one sixteen of original one. After finish down-sampling stage, with several sequentially layers for feature extraction, up-sampling stage of initial image is completed.

Especially, we use a Cbam attention mechanism that can support network better get the features of original image even under different sizes, and further analysis them in detail. The Cbam mechanism work in up-sampling stage. After done one convolution, it will do the channel attention first, and then do the spatial attention before going to next convolution block. Meanwhile, one skip-connection from the input feature map directly to the outpt is used to get the result. With two max-pooling and two fully connected layers, a same channel sized vector will be generated. Spatial attention is more focus on where is the meaningful part, which means we will reduce the channel number. Also, we implemented one 7*7 convolutional

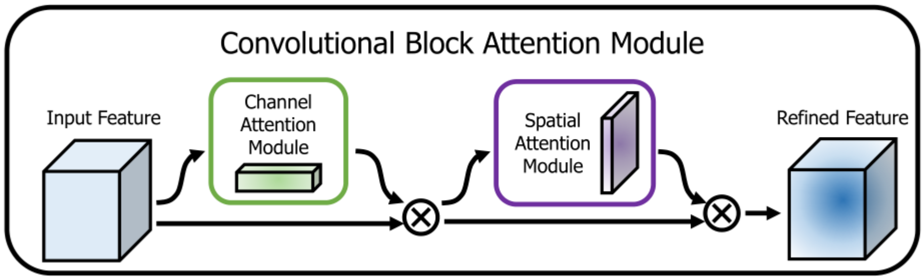


Fig. 2: The Cbam Architecture [5]

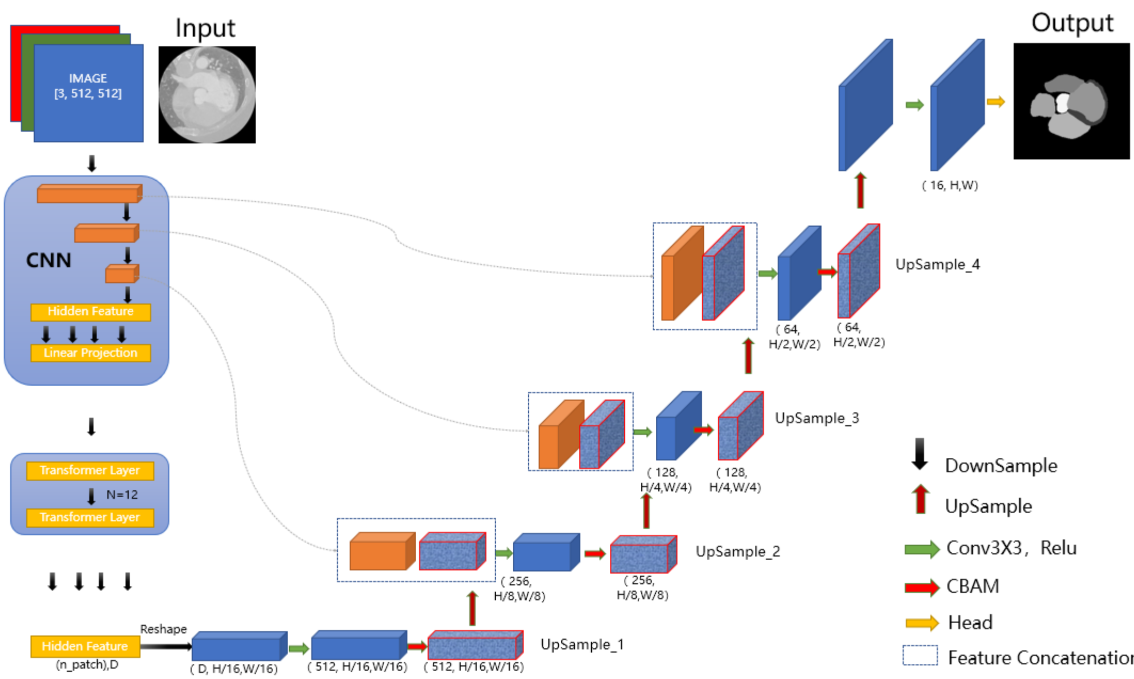


Fig. 3: The structure of TransNUNet [10]

kernel sized layer with two max-pooling, and finally we will get a represented matrix. The matrix will be multiplied with the vector to update the feature map.

What's more, the concatenation operation is performed in each layer, which makes the image size of each layer of Cbam increases from one sixteen of original one to the same size of initial one. Finally, after completing four parts of enhancement of feature map, the segmentation ability of the network will be significantly improved.

IV. Experimental Setup

A. Datasets Preparation

All the images used in training and testing processes are contained in The Liver Tumor Segmentation Challenge (L-i-T-S) dataset [6] from the MICCAI 2017 challenge which is publicly available and free to download. As the raw dataset is provided via Nifty (.nii) files, and the image size is 512 * 512, it is complex and time-consuming costed by the GPU memory constraints. Therefore, we reduced the image size to 256*256 as the input to the network.What's more, we also do the Nifty files transformation (PNG to Nifty & Nifty to PNG) as our model only can receive the PNG files.

The official provides a challenge competition website to submit the results of model segmentation result and give the corresponding scores (automatically given by the official website algorithm). At the beginning, we used all 130 patient samples in training dataset to train the model, and 70 samples of all testing datasets to predict the results and submit them to the official website. However, because the official website takes more time to evaluate our experimental results than its specified time, we cannot obtain the results predicted by our model.

After that, as the test dataset provided by the challenge has no labels, we choose to separate the training dataset into two parts which contains 130 patient samples. 100 patient samples images are used in training process and 30 patient samples are used in testing process, respectively.

There are three classes in the dataset mask, background, liver, liver, liver tumor, with background's pixel value as '0', liver's pixel value as '1', liver tumor's as '2'. Our network not only segments the liver from the background, but also segments liver tumors from liver. Also, we should notice that there are several cases that do not contains tumors, which means there is no label '2' in these cases GTs image. We are not sure how these cases influence the predict result, but see from the result, our model segmented the liver (no tumor) perfectly.

B. Experiment Environment

This research runs on Python 3.7 and uses several libraries/packages like torch, Numpy, OpenCV, Scikit-image, etc. All the networks trained and tested in the research were implemented on Pytorch. Moreover, the experiments were conducted on Windows 10 system with CPU i7-11700 and 8GB RAM GPU 3070ti.

C. Evaluation Metric

We give a multi-branch function to check the model accuracy. The basic function included in the research contains the cross-entropy function. This function is implemented to optimize the model, and dice function, we used it to evaluate. The cross-entropy function is aimed at optimizing the segmentation ability of the model, meanwhile, the Dice function aimed at evaluated distance of the predicted result with the label.

As a result, the smaller the dice loss is, the better the model performance will be, and closer the predicted outcome to the GT (groundtruth).

V. Preliminary Results

To better see the results, we have provided comparison of dice coefficient of models, in Table I. As the Table I shows below, among four different models, TranNUNet has achieved best predicting accuracy which is 97.93 in training and 91.96 in testing. The experiment shows that TransNUNet has a strong ability in dealing with liver and its tumor segmentation challenge and can overcome difficulties like image contrast differences, low image quality, etc.

It should be noted that the experimental results shown in this part are all the results before data preprocessing. Therefore, the training set / test set used in the four models are not preprocessed.

Fig 4. (End of the paper) shows the comparison of the final segmentation outcome of our model and UNet. Four patient samples are listed, and four serial diagram which contains original image, ground truth image, UNet result, TranNUNet result are given. For the ground truth image, the blue color represents the tumor in the liver and white color represents liver. It can be seen from the images that our method can clearly segment the liver and its tumor from the background and sort them into different classes. Obviously, our model TransNUNet performs better than UNet and TranNUNet performance excellent in segmenting liver and its tumors.

TABLE I: Comparison of Dice coefficient of models

Models	Training Accuracy (DSC) %	Testing Accuracy (DSC) %
UNet	86.19	71.85
UNet3+	95.31	82.61
TransUNet	94.56	87.13
TransNUNet	97.93	91.96

VI. Conclusion

The paper proposed a methodology for liver and liver tumor segmentation named TransNUNet. Based on the TransUNet, we have added Cbam attention mechanism and used multi-branch loss function to check the accuracy of models and to obtain better performance. Compared with traditional U-Net, U-Net3+, TransUNet, our network performs best among these networks, which achieves 91.96% in liver and liver tumor segmentation. The result shows that with Cbam mechanism, using a channel attention which is focus on which channel is meaningful and spatial attention which is focus on where is the meaningful channel, the global feature solves the limitations of CNN networks in feature extraction. What's more, attention mechanism also gives the connection between pixels of different classes. Our experiments also reveals that attention mechanism can significantly improve 'U' shape network performance in medical image processing tasks and challenges. However, there are still exist several troubles to be solved in the future like hard to segment small and low contrast liver tumor, the easy overfitting caused by deeper network, and huge time cost during training process, etc.

References

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI.

[2] Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., & Ronneberger, O. (2016). 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. ArXiv, abs/1606.06650.

[3] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306, 2021.

[4] Yutong Xie, Jianpeng Zhang, Chunhua Shen, and Yong Xia. Cotr: Efficiently bridging cnn and transformer for 3d medical image segmentation. arXiv preprint arXiv:2103.03024, 2021.

[5] SanghyunWoo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam: Convolutional block attention module. In Proceedings of the European conference on computer vision (ECCV), pages 3–19, 2018.

[6] Bilic, P., Christ, P.F., Vorontsov, E., Chlebus, G., Chen, H., Dou, Q., Fu, C., Han, X., Heng, P., Hesser, J.W., Kadoury, S., Konopczynski, T.K., Le, M., Li, C., Li, X., Lipková, J., Lowengrub, J.S., Meine, H., Moltz, J.H., Pal, C.J., Piraud, M., Qi, X., Qi, J., Rempfler, M., Roth, K., Schenk, A., Sekuboyina, A.K., Zhou, P., Hülsemeyer, C., Beetz, M., Ettlinger, F.,

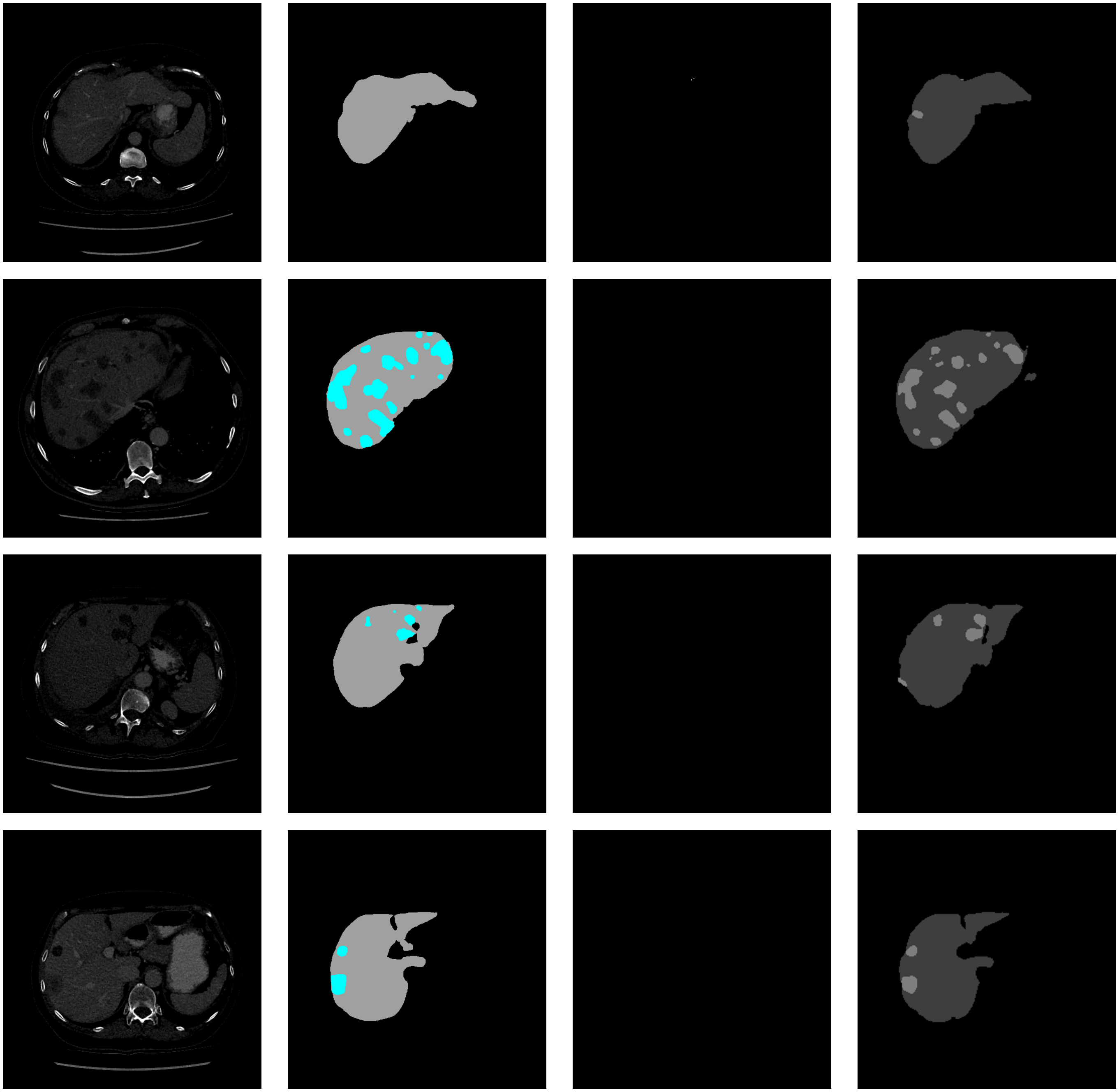
Grün, F., Kaissis, G., Lohöfer, F.K., Braren, R.F., Holch, J.W., Hofmann, F., Sommer, W.H., Heinemann, V., Jacobs, C., Mamani, G.E., Ginneken, B.V., Chartrand, G., Tang, A., Drozdzal, M., Ben-Cohen, A., Klang, E., Amitai, M.M., Konen, E., Greenspan, H., Moreau, J., Hostettler, A., Soler, L., Vivanti, R., Szeskin, A., Lev-Cohain, N., Sosna, J., Joskowicz, L., & Menze, B.H. (2019). The Liver Tumor Segmentation Benchmark (LiTS). ArXiv, abs/1901.04056.

[7] Huang, H., Lin, L., Tong, R., Hu, H., Zhang, Q., Iwamoto, Y., Han, X., Chen, Y., & Wu, J. (2020). UNet 3+: A Full-Scale Connected UNet for Medical Image Segmentation. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 1055-1059.

[8] Gruber, N., Antholzer, S., Jaschke, W., Kremser, C., & Haltmeier, M. (2019). A Joint Deep Learning Approach for Automated Liver and Tumor Segmentation. 2019 13th International conference on Sampling Theory and Applications (SampTA), 1-5.

[9] P. F. Christ et al., "Automatic Liver and Lesion Segmentation in CT Using Cascaded Fully Convolutional Neural Networks and 3D Conditional Random Fields," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), vol. 9901 LNCS, pp. 415–423, Oct. 2016, doi: 10.1007/978-3-319-46723-$8_4$8.

[10] Yang, Xiaoniu and Tian, Xiaolin. " TransNUNet: Using Attention Mechanism for Whole Heart Segmentation" 2022 IEEE 2nd International Conference on Power, Electronics and Computer Applications (ICPECA) Power, Electronics and Computer Applications (ICPECA), 2022 IEEE 2nd International Conference on. :553-556 Jan, 2022.



(a) Initial Image (b) Groundtruth (c) UNet Result (d) TransNUNet Result

Fig. 4: Comparison of different models predicted outcome for liver/tumor segmentation