A Dual-Branch Fusion Network with Asymmetric Depthwise Convolutions for Whole-Body Tumor Segmentation

Wenze Fan[1], Yue Liu, Yinyin Luo, Ruizhi Li, and Gang Fang*

Institute of Computing Science and Technology, Guangzhou University, Guangzhou, 510006, China
[1]Co-first authors
*Corresponding author
gangf@gzhu.edu.cn

Abstract. Although deep learning models have made significant progress in medical segmentation, they have encountered challenges in tumor detection and segmentation tasks, particularly in complex whole-body scanning scenarios.To address this issue, this paper proposes an improved nn UNet model. Specifically, an efficient dual branch encoder has been developed, with one branch using a special "Inception style depthwise separable convolution" inspired by the Google Inception network, and the other branch using ordinary residual connections. This model effectively improves segmentation efficiency.Experiments on the MICCAI FLARE 2025 dataset demonstrated significant improvements in both segmentation accuracy and efficiency. Our method achieved an average organ Dice Similarity Coefficient (DSC) of 13.4% and a Normalized Surface Dice (NSD) of 5.98% on the public validation set.The average inference time is 11 seconds.

Keywords: Dual branch · Tumour segmentation · Asymmetric Depthwise Convolutions.

1 Introduction

In recent years, deep learning models have greatly advanced the field of medical image segmentation. Models represented by U-Net [21] have proposed a groundbreaking encoder decoder architecture, which has become an important benchmark for biomedical image segmentation. Subsequently, extended models such as nn-UNet [11] have achieved substantial improvements in various segmentation tasks through their adaptive configuration capabilities, and have become SOTA models in many segmentation tasks. These advances have significantly enhanced the segmentation performance of organs and pathological regions in various medical imaging modes.However, despite these breakthroughs, accurately and efficiently detecting and segmenting tumors in whole-body CT scans still faces significant challenges. Tumors exhibit significant heterogeneity in shape,

size, and appearance across different anatomical regions, making the segmentation task itself extremely complex. In addition, full body scanning involves high-resolution 3D images, which not only imposes strict requirements on computing performance, but also requires a large amount of computing resources. Maintaining high accuracy and speed while processing massive amounts of data further exacerbates the complexity of this challenge.To address these challenges, researchers have explored various advanced methods. For example,TransUNet [3] utilizes the powerful global self attention mechanism of Transformer to capture long-range dependencies, which is of great significance for understanding the diverse morphology of tumors and their contextual relationships in a wide anatomical environment. Swin-UMamba [14] also adopts advanced Mamba architecture, which can efficiently construct long-distance dependency models with linear complexity. This makes it significantly more computationally efficient and consumes less video memory compared to traditional Transformers when processing large-scale, high-resolution whole-body CT scans.Despite these advances, there are still significant challenges in extending these methods to whole-body tumor tasks.

Recently, inspired by a paper [12], it was pointed out that since the emergence of nnUNet, a large number of new models claiming to surpass in the field of 3D medical image segmentation have emerged, especially those based on Transformer and Mamba. However, the author found through a comprehensive, fair, and rigorous benchmark test that the vast majority of these so-called latest technology models cannot surpass**nnU-Net** in terms of performance under fair comparison. At the same time, it was found that the current performance formula for 3D medical segmentation actually uses CNN based **U-Net** models, especially variants of ResNet and ConvNext.

Despite the excellent performance of **nnU-Net**, there is still room for optimization in the highly challenging task of whole-body tumor segmentation. The morphology, size, and texture of tumors throughout the body exhibit extremely high heterogeneity at different anatomical locations. At the same time, high-resolution 3D CT scanning also brings huge computational burden, which puts strict requirements on the feature extraction ability and computational efficiency of the model.To address the above challenges, our main contributions are:

- The inspiration for enhancing the multi-scale feature perception ability of the model mainly comes from the Inception architecture [27] (variants of ResNet and ConvNext). By using convolutional kernels of different sizes and types in parallel within the same module, such methods enable the network to simultaneously capture target features of different sizes, from small lesions to large masses, and effectively deal with the problem of variable tumor size.
- Another important direction is to improve the computational efficiency of models in processing high-resolution 3D data. The convolutions used in the Inception architecture are all depthwise separable convolutions, which greatly reduce the model's parameter count and computational complexity while ensuring the receptive field and feature expression ability. In addi-

tion, asymmetric convolutions (such as 1x11x11) have also been proven to effectively capture features in specific directions at lower costs.
- In order to combine the advantages of different feature extraction paradigms, our encoder adopts a dual branch structure, with one branch using Inception branch to extract features and the other branch using residual branch to extract features. Finally, feature fusion is used to enhance the overall performance of the network.

2 Method

2.1 Preprocessing

We adopted the standard preprocessing process of nn-UNet. Firstly, only select images with tumor annotations for training. Subsequently, all CT images were resampled to an isotropic voxel spacing of 2x2x3 mm ^{3}. Finally, we normalized the Heinz unit (HU) values by truncating them at the 0.05% and 99.5% percentiles, and then subtracting the mean and dividing it by the standard deviation based on the statistical data of the entire dataset.

2.2 Proposed Method

Figure 1 shows the architecture of our proposed model and its various modules. Although many novel U-shaped models have been proposed in recent years to address medical image segmentation problems, especially those based on Transformer and Mamba, we have found that the most advanced performance formulas still use CNN based **U-Net** models and **nnU-Net** frameworks, including Resnet and Convnext variants. Based on this, we have developed a new model to solve the problem of whole-body tumor segmentation.

For the encoder, we used a dual branch structure; At each stage of the IDC branch, a special "Inception like depthwise separable convolution" inspired by the Google Inception network is used. It is divided into five branches in the channel dimension, with one part directly retained and the other four parts extracting features through depthwise separable convolutions of different shapes. There are three main points to using Inception depthwise separable convolution:

1.CT images are usually anisotropic, and the stripe convolution kernels in IDC branches are particularly adept at capturing slender or flat structures along specific axes, which is very effective for segmenting tumors with irregular shapes that may extend in a certain direction.

2. The IDC branch can simultaneously capture local detail features (captured by $3 \times 3 \times 3$ cubic convolution kernels) and contextual information of long-distance and large receptive fields (captured by 3 strip-shaped convolution kernels of $1 \times 11 \times 11$, $11 \times 1 \times 11$, and $11 \times 11 \times 1$) at the same level. This is crucial for identifying tumors of different sizes and shapes.

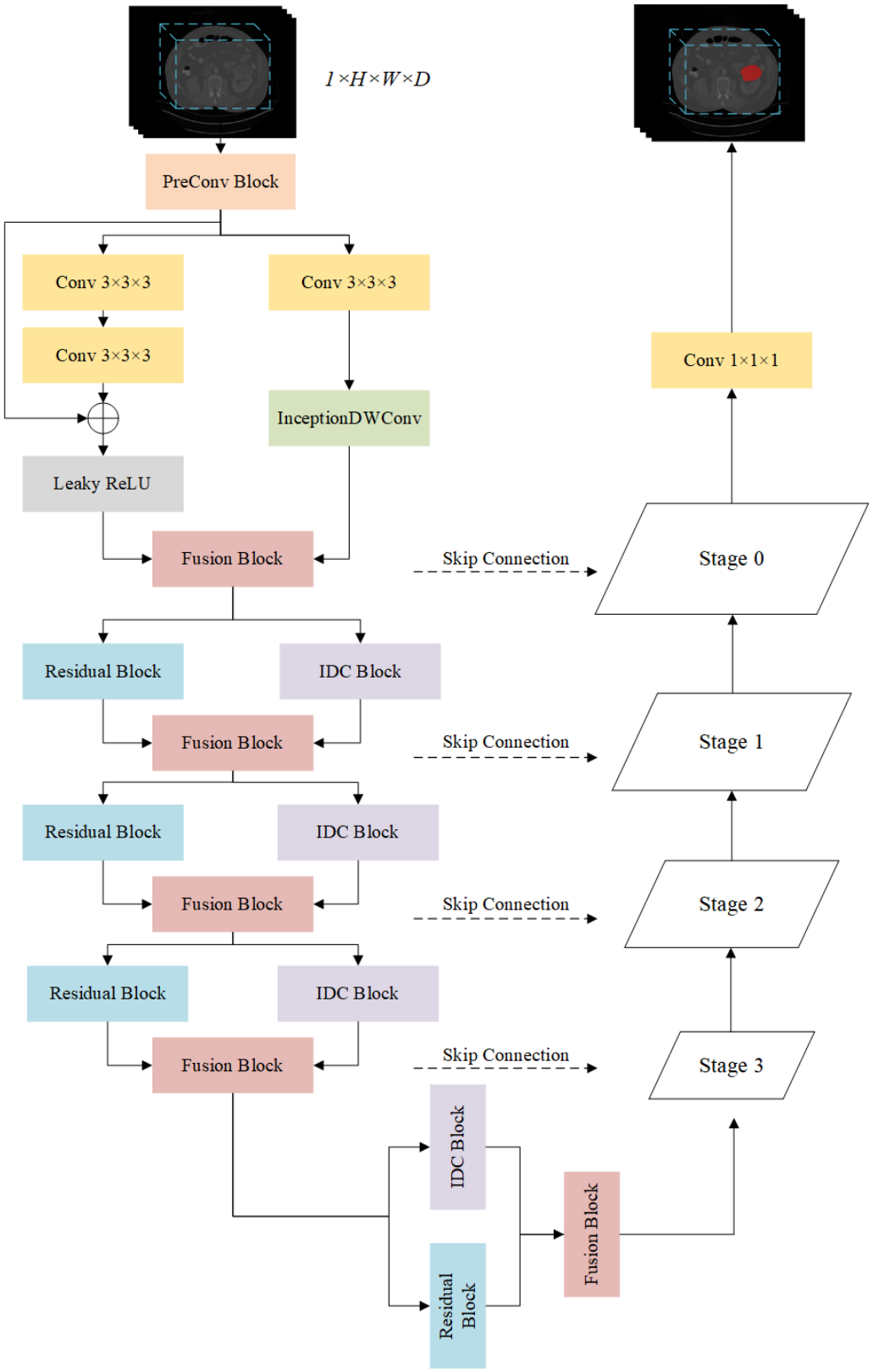

Fig. 1. The model is a five layer U-shaped network consisting of a hybrid encoder with Inception Depthwise Convolution and the original **nnU-Net** decoder.

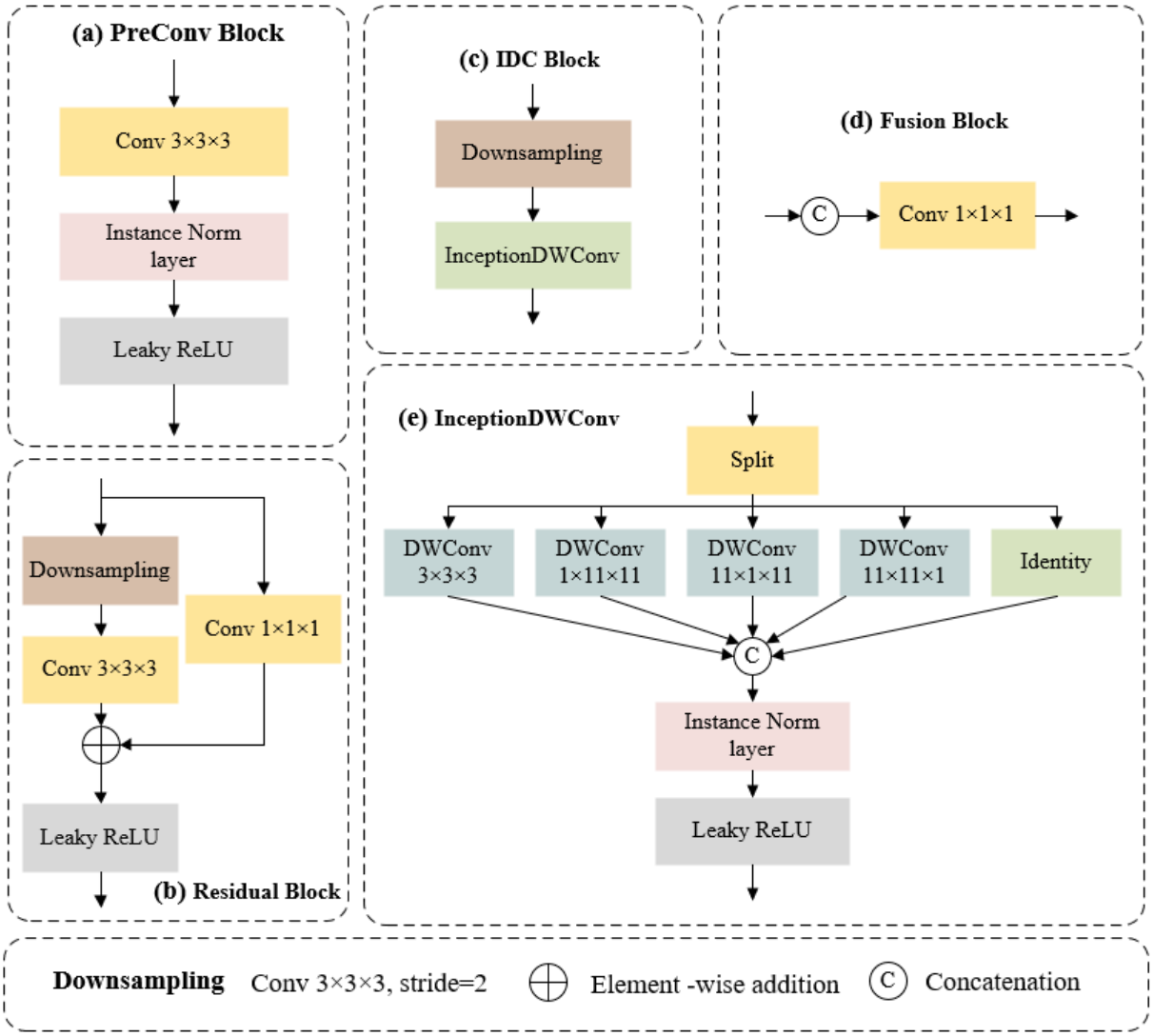

Fig. 2. Model architecture components: (a) **PreConv Block**: Used before the first stage; (b) **Residual Block**:The main modules in a single branch; (c) **IDC Block**: The main modules of another branch; (d) **Fusion Block**: Feature fusion module; (e) **InceptionDWConv**: Core module of the entire network.

3. This branch mainly uses depthwise separable convolution, which can significantly reduce the number of model parameters and computational complexity compared to standard convolution, thereby accelerating model training and inference speed.

However, we found that using only IDC encoder would lack some key 3D tumor voxel information. Although depthwise separable convolution has high efficiency, its feature expression ability is usually considered weaker than standard convolution. Standard convolution integrates information from all input channels when calculating each output channel, enabling learning of richer and more complex feature combinations. The separate processing of IDC branches may to some extent limit their ability to learn the most advanced and abstract features. So we added a residual branch on top of the IDC branch to compensate for its shortcomings. The residual branch uses standard convolution, and during computation, each output feature map integrates information from all input feature maps. This makes up for the shortcomings of IDC branch channel separation processing, enabling the network to learn more comprehensive and holistic feature representations, and better understand the complex spatial relationship between tumors and their surrounding tissues.

Finally, the model can efficiently capture the anisotropic shape, orientation, and multi-scale information of tumors using IDC branches, and learn complex textures and overall contextual features of the tumor's internal and surrounding environment using residual branches. The combination of these abilities is crucial for the task of whole-body cancer segmentation, as tumors of different parts and types exhibit vastly different morphological and textural features on CT.

Loss function: we use the summation between Dice loss and cross-entropy loss because compound loss functions have proven robust in various medical image segmentation tasks.

We did not adopt specific strategies to reduce false positives in CT scans of healthy patients and only used the provided labeled data. In addition, we did not use the pseudo labels generated by the FLARE23 winning algorithm.

3 Experiments

3.1 Dataset and evaluation measures

The segmentation targets cover various lesions. The training dataset is curated from more than 50 medical centers under the license permission, including TCIA [4], LiTS [2], MSD [23], KiTS [8,10,9], autoPET [7,6], TotalSegmentator [24], and AbdomenCT-1K [19], FLARE 2023 [18], DeepLesion [26], COVID-19-CT-Seg-Benchmark [16], COVID-19-20 [22], CHOS [13], LNDB [20], and LIDC [1]. The training set includes more than 10000 abdomen CT scans where 2200 CT scans with partial labels and 1800 CT scans without labels. The validation and testing sets include 100 and 400 CT scans, respectively, which cover various abdominal cancer types, such as liver cancer, kidney cancer, pancreas cancer, colon cancer, gastric cancer, and so on. The lesion annotation process used ITK-SNAP [28], nnU-Net [11], MedSAM [15,17], and Slicer Plugins [5,17].

The evaluation metrics encompass two accuracy measures—Dice Similarity Coefficient (DSC) and Normalized Surface Dice (NSD)—alongside two efficiency measures—running time and area under the GPU memory-time curve. These metrics collectively contribute to the ranking computation. Furthermore, the running time and GPU memory consumption are considered within tolerances of 45 seconds and 4 GB, respectively.

3.2 Implementation details

Environment settings The development environments and requirements are presented in Table 1.

Table 1. Development environments and requirements.

System	Rocky Linux 8.9
CPU	Intel(R) Xeon(R) w5-3435X
RAM	8×32GB; 3200MT/s
GPU (number and type)	2×NVIDIA GeForce RTX 4090 48G
CUDA version	12.2
Programming language	Python 3.10.17
Deep learning framework	torch 2.4.1
Specific dependencies	nnU-Net 2.6.2
Code	https://github.com/Wenze-Fan/FLARE25-Task1

Training protocols Our training protocols followed the default settings of nnU-Net.

1. we used only the provided labeled data without distinguishing between partially labeled or unlabeled data.

2. For data augmentation, we use the built-in data augmentation feature of nnUNet.

3. The patch sampling strategy was also based on the default configuration of nnU-Net.

4. Our model has two types, one is single branch and the other is double branch. In the end, we used the best dual branch model.

4 Results and discussion

4.1 Quantitative results on validation set

Quantitative results are shown in Table 3.On the public validation set, our method achieved an average Dice of 13.4% and an NSD of 5.98%.

In order to visually demonstrate the advantages of our method in terms of

Table 2. Training protocols.

Network initialization	He
Batch size	2
Patch size	128×128×96
Total epochs	2000
Optimizer	Adamw
Initial learning rate (lr)	0.0008
Lr decay schedule	poly
Training time	100 hours
Loss function	DiceLoss and CELoss
Number of model parameters	8.23M[1]
Number of flops	186.75G[2]
CO_2eq	48 Kg[3]

Table 3. Result in Public Validation, Online Validation and Final Testing.

Methods	Public Validation		Online Validation		Testing	
	DSC(%)	NSD(%)	DSC(%)	NSD(%)	DSC(%)	NSD (%)
Algorithm1	13.4	5.98	-	-	-	-

Table 4. Experimental study on ablation of different components of encoder with different effects.

Network	**Method**	**Param**	**Flop**	**DSC**	**NSD**
nn-UNet	Baseline	15.37M	416.78G	7.71	2.79
	IDC branch(Single branch)	8.76M	193.22G	10.29	4.61
	Ours(Dual branch)	**8.23M**	**186.75G**	**13.4**	**5.98**

parameter count, computational complexity, and accuracy, we conducted ablation experiments as shown in Table 4.Compared with traditional methods and encoders using a single branch structure, our method has improved accuracy by 5.69% and 3.19%, and 3.11% and 1.37%, respectively Especially, our method has lower parameter and computational complexity than single branch structures.

Table 5. Quantitative evaluation of segmentation efficiency in terms of the running them and GPU memory consumption. Total GPU denotes the area under GPU Memory-Time curve. Evaluation GPU platform: NVIDIA GeForce RTX 4090 (24G).

Case ID	Image Size	Running Time (s)	Max GPU (MB)	Total GPU (MB)
0001	(512, 512, 55)	19.16	4058	11331
0051	(512, 512, 100)	29.37	3443	42576
0017	(512, 512, 150)	25.22	4147	26724
0019	(512, 512, 215)	28.51	4081	44263
0099	(512, 512, 334)	37.40	3439	54464
0063	(512, 512, 448)	43.76	3435	66108
0048	(512, 512, 499)	48.08	4145	48173
0029	(512, 512, 554)	52.23	3571	41607

4.2 Qualitative results on validation set

Figure 3 shows two successful and two failed tumor segmentation cases.In the good cases (top two rows),the model is mainly designed through parallel multi-scale deep convolution (InceptionDWConv) and is very good at identifying and segmenting tumors with significant features and large volumes.On the contrary,for tumors with atypical morphology and texture (Case0067), the feature patterns learned by the model may not be sufficient to cover them, resulting in incomplete segmentation.For lesions that are too small in size (Case0003), the model will completely lose their feature information during repeated downsampling in the encoder, resulting in complete inability to detect them.

4.3 Segmentation efficiency results on validation set

Table 5 shows inference efficiency on representative cases. Average runtime is around 35 seconds per scan, with peak GPU usage below 4.2 GB, showing good scalability.

4.4 Results on final testing set

4.5 Limitation and future work

Despite achieving good accuracy, our model still has several limitations:

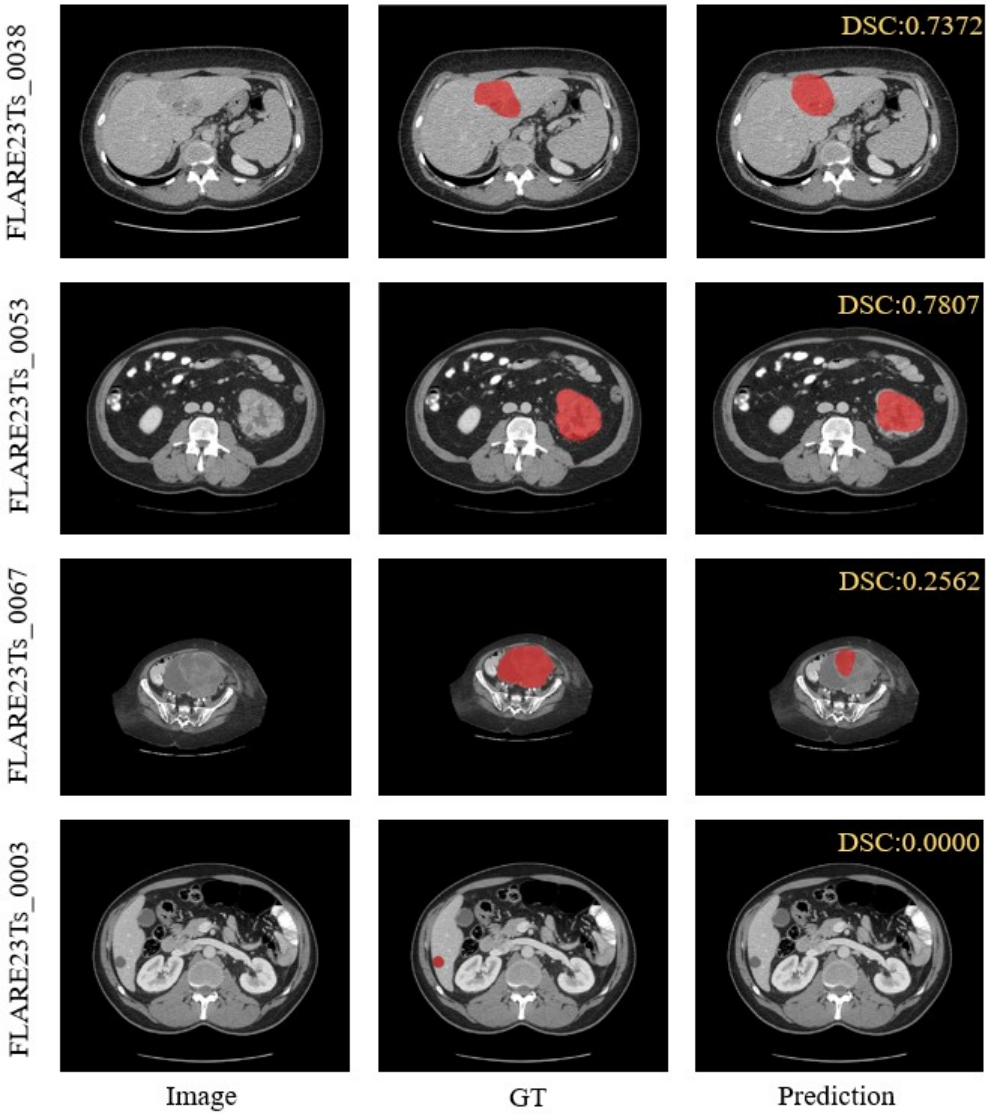

Fig. 3. There are two examples in the validation set with good segmentation performance and two examples with poor segmentation performance.

(1) **Accuracy**: In order to further improve the accuracy of the model, it is possible to consider introducing appropriate attention mechanisms to enhance attention to small or irregular targets, or designing a branch specifically for handling small targets, or through more targeted data augmentation strategies.
(2) **Partial Labels**:We use fully supervised methods, which only use labeled data to train the model. This may result in the model learning most of the patterns in the training data rather than truly learning useful features for tumors with irregular shapes or textures. Future work will include semi supervised learning to develop a student teacher model or pseudo label model to further improve accuracy.
(3) **Lightweight**:Due to the dual branch structure of our encoder, the complexity of the model has been increased, making it difficult to deploy the model in real-world clinical applications. Future work will include lightweighting models while maintaining high accuracy.

5 Conclusion

We propose a dual branch encoder for efficient whole-body tumor segmentation. The model achieved 13.4% DSC and 5.9% NSD. The average inference time reached 11 seconds. Both quantitative and qualitative results indicate that this method can effectively and flexibly learn tumor information from the dataset. Future work will focus on how to use semi supervised learning or a combination of semi supervised learning and nn Unet improvements to more efficiently solve the problem of whole-body tumor segmentation.

Acknowledgements The authors of this paper declare that the segmentation method they implemented for participation in the FLARE 2024 challenge has not used any pre-trained models nor additional datasets other than those provided by the organizers. The proposed solution is fully automatic without any manual intervention. We thank all data owners for making the CT scans publicly available and CodaLab [25] for hosting the challenge platform.

Disclosure of Interests

The authors declare no competing interests.

References

1. Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Zhao, B., Aberle, D.R., Henschke, C.I., Hoffman, E.A., et al.: The lung image database consortium (lidc) and image database resource initiative (idri): a completed reference database of lung nodules on ct scans. Medical Physics **38**(2), 915–931 (2011) 6

2. Bilic, P., Christ, P., Li, H.B., Vorontsov, E., Ben-Cohen, A., Kaissis, G., Szeskin, A., Jacobs, C., Mamani, G.E.H., Chartrand, G., Lohöfer, F., Holch, J.W., Sommer, W., Hofmann, F., Hostettler, A., Lev-Cohain, N., Drozdzal, M., Amitai, M.M., Vivanti, R., Sosna, J., Ezhov, I., Sekuboyina, A., Navarro, F., Kofler, F., Paetzold, J.C., Shit, S., Hu, X., Lipková, J., Rempfler, M., Piraud, M., Kirschke, J., Wiestler, B., Zhang, Z., Hülsemeyer, C., Beetz, M., Ettlinger, F., Antonelli, M., Bae, W., Bellver, M., Bi, L., Chen, H., Chlebus, G., Dam, E.B., Dou, Q., Fu, C.W., Georgescu, B., i Nieto, X.G., Gruen, F., Han, X., Heng, P.A., Hesser, J., Moltz, J.H., Igel, C., Isensee, F., Jäger, P., Jia, F., Kaluva, K.C., Khened, M., Kim, I., Kim, J.H., Kim, S., Kohl, S., Konopczynski, T., Kori, A., Krishnamurthi, G., Li, F., Li, H., Li, J., Li, X., Lowengrub, J., Ma, J., Maier-Hein, K., Maninis, K.K., Meine, H., Merhof, D., Pai, A., Perslev, M., Petersen, J., Pont-Tuset, J., Qi, J., Qi, X., Rippel, O., Roth, K., Sarasua, I., Schenk, A., Shen, Z., Torres, J., Wachinger, C., Wang, C., Weninger, L., Wu, J., Xu, D., Yang, X., Yu, S.C.H., Yuan, Y., Yue, M., Zhang, L., Cardoso, J., Bakas, S., Braren, R., Heinemann, V., Pal, C., Tang, A., Kadoury, S., Soler, L., van Ginneken, B., Greenspan, H., Joskowicz, L., Menze, B.: The liver tumor segmentation benchmark (lits). Medical Image Analysis **84**, 102680 (2023) 6
3. 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), https://arxiv.org/abs/2102.04306 2
4. Clark, K., Vendt, B., Smith, K., Freymann, J., Kirby, J., Koppel, P., Moore, S., Phillips, S., Maffitt, D., Pringle, M., Tarbox, L., Prior, F.: The cancer imaging archive (tcia): maintaining and operating a public information repository. Journal of Digital Imaging **26**(6), 1045–1057 (2013) 6
5. Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al.: 3d slicer as an image computing platform for the quantitative imaging network. Magnetic Resonance Imaging **30**(9), 1323–1341 (2012) 6
6. Gatidis, S., Früh, M., Fabritius, M., Gu, S., Nikolaou, K., Fougère, C.L., Ye, J., He, J., Peng, Y., Bi, L., Ma, J., Wang, B., Zhang, J., Huang, Y., Heiliger, L., Marinov, Z., Stiefelhagen, R., Egger, J., Kleesiek, J., Sibille, L., Xiang, L., Bendazolli, S., Astaraki, M., Schölkopf, B., Ingrisch, M., Cyran, C., Küstner, T.: The autopet challenge: towards fully automated lesion segmentation in oncologic pet/ct imaging. Nature Machine Intelligence (2023) 6
7. Gatidis, S., Hepp, T., Früh, M., La Fougère, C., Nikolaou, K., Pfannenberg, C., Schölkopf, B., Küstner, T., Cyran, C., Rubin, D.: A whole-body fdg-pet/ct dataset with manually annotated tumor lesions. Scientific Data **9**(1), 601 (2022) 6
8. Heller, N., Isensee, F., Maier-Hein, K.H., Hou, X., Xie, C., Li, F., Nan, Y., Mu, G., Lin, Z., Han, M., Yao, G., Gao, Y., Zhang, Y., Wang, Y., Hou, F., Yang, J., Xiong, G., Tian, J., Zhong, C., Ma, J., Rickman, J., Dean, J., Stai, B., Tejpaul, R., Oestreich, M., Blake, P., Kaluzniak, H., Raza, S., Rosenberg, J., Moore, K., Walczak, E., Rengel, Z., Edgerton, Z., Vasdev, R., Peterson, M., McSweeney, S., Peterson, S., Kalapara, A., Sathianathen, N., Papanikolopoulos, N., Weight, C.: The state of the art in kidney and kidney tumor segmentation in contrast-enhanced ct imaging: Results of the kits19 challenge. Medical Image Analysis **67**, 101821 (2021) 6
9. Heller, N., Isensee, F., Trofimova, D., Tejpaul, R., Zhao, Z., Chen, H., Wang, L., Golts, A., Khapun, D., Shats, D., Shoshan, Y., Gilboa-Solomon, F., George, Y., Yang, X., Zhang, J., Zhang, J., Xia, Y., Wu, M., Liu, Z., Walczak, E., McSweeney,

S., Vasdev, R., Hornung, C., Solaiman, R., Schoephoerster, J., Abernathy, B., Wu, D., Abdulkadir, S., Byun, B., Spriggs, J., Struyk, G., Austin, A., Simpson, B., Hagstrom, M., Virnig, S., French, J., Venkatesh, N., Chan, S., Moore, K., Jacobsen, A., Austin, S., Austin, M., Regmi, S., Papanikolopoulos, N., Weight, C.: The kits21 challenge: Automatic segmentation of kidneys, renal tumors, and renal cysts in corticomedullary-phase ct. arXiv preprint arXiv:2307.01984 (2023) 6

10. Heller, N., McSweeney, S., Peterson, M.T., Peterson, S., Rickman, J., Stai, B., Tejpaul, R., Oestreich, M., Blake, P., Rosenberg, J., et al.: An international challenge to use artificial intelligence to define the state-of-the-art in kidney and kidney tumor segmentation in ct imaging. American Society of Clinical Oncology **38**(6), 626–626 (2020) 6
11. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021) 1, 6
12. Isensee, F., Wald, T., Constantin, U., Michael, B., Saikat, R., Klaus, M.H., Paul, J.: nnu-net revisited: A call for rigorous validation in 3d medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention. pp. 488–498. Springer Nature Switzerland, Cham (2024) 2
13. Kavur, A.E., Gezer, N.S., Barış, M., Aslan, S., Conze, P.H., Groza, V., Pham, D.D., Chatterjee, S., Ernst, P., Özkan, S., Baydar, B., Lachinov, D., Han, S., Pauli, J., Isensee, F., Perkonigg, M., Sathish, R., Rajan, R., Sheet, D., Dovletov, G., Speck, O., Nürnberger, A., Maier-Hein, K.H., Bozdağı Akar, G., Ünal, G., Dicle, O., Selver, M.A.: CHAOS Challenge - combined (CT-MR) healthy abdominal organ segmentation. Medical Image Analysis **69**, 101950 (2021) 6
14. Liu, J., Yang, H., Zhou, H.Y., Xi, Y., et al.: Swin-umamba: Mamba-based unet with imagenet-based pretraining. In: Medical Image Computing and Computer Assisted Intervention - MICCAI 2024. pp. 615–625 (2024) 2
15. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 654 (2024) 6
16. Ma, J., Wang, Y., An, X., Ge, C., Yu, Z., Chen, J., Zhu, Q., Dong, G., He, J., He, Z., Cao, T., Zhu, Y., Nie, Z., Yang, X.: Towards data-efficient learning: A benchmark for covid-19 ct lung and infection segmentation. Medical Physics **48**(3), 1197–1210 (2021) 6
17. Ma, J., Yang, Z., Kim, S., Chen, B., Baharoon, M., Fallahpour, A., Asakereh, R., Lyu, H., Wang, B.: Medsam2: Segment anything in 3d medical images and videos. arXiv preprint arXiv:2504.03600 (2025) 6
18. Ma, J., Zhang, Y., Gu, S., Ge, C., Wang, E., Zhou, Q., Huang, Z., Lyu, P., He, J., Wang, B.: Automatic organ and pan-cancer segmentation in abdomen ct: the flare 2023 challenge. arXiv preprint arXiv:2408.12534 (2024) 6
19. Ma, J., Zhang, Y., Gu, S., Zhu, C., Ge, C., Zhang, Y., An, X., Wang, C., Wang, Q., Liu, X., Cao, S., Zhang, Q., Liu, S., Wang, Y., Li, Y., He, J., Yang, X.: Abdomenct-1k: Is abdominal organ segmentation a solved problem? IEEE Transactions on Pattern Analysis and Machine Intelligence **44**(10), 6695–6714 (2022) 6
20. Pedrosa, J., Aresta, G., Ferreira, C., Atwal, G., Phoulady, H.A., Chen, X., Chen, R., Li, J., Wang, L., Galdran, A., et al.: Lndb challenge on automatic lung cancer patient management. Medical Image Analysis **70**, 102027 (2021) 6
21. 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 (2015) 1

22. Roth, H.R., Xu, Z., Tor-Díez, C., Jacob, R.S., Zember, J., Molto, J., Li, W., Xu, S., Turkbey, B., Turkbey, E., et al.: Rapid artificial intelligence solutions in a pandemic—the covid-19-20 lung ct lesion segmentation challenge. Medical Image Analysis **82**, 102605 (2022) 6
23. Simpson, A.L., Antonelli, M., Bakas, S., Bilello, M., Farahani, K., van Ginneken, B., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., Bilic, P., Christ, P.F., Do, R.K.G., Gollub, M., Golia-Pernicka, J., Heckers, S.H., Jarnagin, W.R., McHugo, M.K., Napel, S., Vorontsov, E., Maier-Hein, L., Cardoso, M.J.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. arXiv preprint arXiv:1902.09063 (2019) 6
24. Wasserthal, J., Breit, H.C., Meyer, M.T., Pradella, M., Hinck, D., Sauter, A.W., Heye, T., Boll, D.T., Cyriac, J., Yang, S., Bach, M., Segeroth, M.: Totalsegmentator: Robust segmentation of 104 anatomic structures in ct images. Radiology: Artificial Intelligence **5**(5), e230024 (2023) 6
25. Xu, Z., Escalera, S., Pavão, A., Richard, M., Tu, W.W., Yao, Q., Zhao, H., Guyon, I.: Codabench: Flexible, easy-to-use, and reproducible meta-benchmark platform. Patterns **3**(7), 100543 (2022) 11
26. Yan, K., Wang, X., Lu, L., Summers, R.M.: Deeplesion: automated mining of large-scale lesion annotations and universal lesion detection with deep learning. Journal of Medical Imaging **5**(3), 036501–036501 (2018) 6
27. Yu, W., Zhou, P., Wang, Yan, S.a.W.: Inceptionnext: When inception meets convnext. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5672–5683. IEEE (2024) 2
28. Yushkevich, P.A., Gao, Y., Gerig, G.: Itk-snap: An interactive tool for semi-automatic segmentation of multi-modality biomedical images. In: Annual International Conference of the IEEE Engineering in Medicine and Biology Society. pp. 3342–3345 (2016) 6

Table 6. Checklist Table. Please fill out this checklist table in the answer column.

Requirements	Answer
A meaningful title	Yes
The number of authors ($\leq$6)	4
Author affiliations and ORCID	Yes
Corresponding author email is presented	Yes
Validation scores are presented in the abstract	Yes
Introduction includes at least three parts: background, related work, and motivation	Yes
A pipeline/network figure is provided	2
Pre-processing	3
Strategies to use the partial label	3
Strategies to use the unlabeled images.	
Strategies to improve model inference	3
Post-processing	
The dataset and evaluation metric section are presented	8
Environment setting table is provided	7
Training protocol table is provided	7
Ablation study	8
Efficiency evaluation results are provided	3 4 5
Visualized segmentation example is provided	3
Limitation and future work are presented	Yes
Reference format is consistent.	Yes