Lightweight Depthwise Separable U-Net for Laptop-Based Abdominal Multi-Organ Segmentation

Haiyan Cheng[0009−0007−8771−7088], Mengshi Tang[0009−0002−3365−8872], Yashuang Zhang[0009−0001−8481−4583], Ruixiang Lei[0009−0008−7413−5860], and Mingjing Yang✉

Intelligent Image processing and Analysis Laboratory, Fuzhou University,Fuzhou 350108, Fujian, China yangmj5@fzu.edu.cn

Abstract. In order to achieve efficient 3D medical image segmentation within resource-constrained environments, we have developed a U-Net-based framework. By integrating depthwise separable convolutions and efficient channel attention mechanisms, our framework is capable of achieving both efficient and accurate segmentation of various abdominal organs within a CPU-based setting. This framework not only optimizes computational efficiency but also enhances the precision of segmentation through sophisticated feature extraction, which is crucial for medical image analysis. The synergy of these techniques not only boosts the performance of our model but also suggests potential for improvement in a range of other medical image segmentation tasks. Our method achieves Dice Similarity Coefficients of 61.1%, 63.4%, and 59.4% on Asia, Europe, and North America datasets, respectively, with an average inference time of under 43 seconds per case.

Keywords: Depthwise Separable Convolution · Efficient Channel Attention · Organ Segmentation.

1 Introduction

In recent years, deep learning technology has made significant breakthroughs in abdominal multi-organ segmentation. However, in resource-constrained environments, such as laptops or edge devices for hospital imaging, the lack of GPU resources remains a key constraint. With the increasingly prominent role of medical image analysis in disease diagnosis and treatment planning, it is particularly critical to develop segmentation algorithms that can operate efficiently in diverse computing environments. Therefore, researchers have begun to focus on the development of lightweight models to accommodate these restrictive conditions.

As a technique to reduce the number of parameters and the amount of computation in convolutional neural networks, Depthwise Separable Convolution has gradually attracted attention in the field of deep learning. François Chollet proposed a new deep convolutional neural network architecture in 2016, called Xception [2], which replaces the Inception module with a deep separable convolution,

thereby maintaining the same number of parameters as Inception V3, achieving better performance. Google proposed MobileNetV1 [10], a lightweight neural network focused on mobile devices. MobileNetV1 is to replace the standard convolutional layer in VGG with the depthwise separable convolution to build a lightweight network, which significantly reduces the number of parameters and computation without significantly decreasing accuracy. The main contributions of this work are summarized as follows:

- Based on U-Net [20], depthwise separable convolution is used to greatly reduce the calculation amount and the number of parameters of the model, to improve the prediction speed and keep the accuracy of the model as much as possible.
- The Efficient Channel Attention (ECA) [7] in the encoder can improve the feature expression ability and enhance the model's attention to important features. At the same time, it is low computational cost and easy to integrate, which can effectively improve the performance of the model without significantly increasing the number of parameters.

2 Method

2.1 Preprocessing

Regarding the preprocessing phase, we carried out four key procedures, detailed as follows:

- Threshold truncation:After an in-depth analysis of many CT images, we selected the [-160, 300] interval as the optimal threshold, in which the visual contrast between different organs and tissues is the most prominent. Using this observation, we applied a threshold truncation technique to enhance the visualization quality of the organs in the images.
- Cropping strategy:In the process of model training, the labeled image slices were selected for processing, and these slices were cropped to eliminate irrelevant data.
- Resampling method for anisotropic data:To reduce GPU memory usage, we optimized the size of the image slices when loading the data. After multiple rounds of validation, we resampled the 2D slice size from 512×512 to 192×192. This adjustment effectively reduced the memory footprint without negatively affecting the segmentation performance of the model.
- Intensity normalization method

2.2 Proposed Method

This study proposes an improved network framework based on the U-Net architecture, with a focus on multi-faceted optimization of the encoder component. In terms of network architecture, this research innovatively adopts depthwise separable convolution, an efficient convolutional operation, by decoupling

traditional convolution into two independent computational steps: depthwise convolution and pointwise convolution, achieving significant improvements in computational efficiency. This design not only preserves the spatial feature extraction capability of traditional convolution but also substantially reduces the model's parameter count. Subsequently, pointwise convolution utilizes 1×1 kernels to achieve linear transformation and feature recombination along the channel dimension, enabling flexible adjustment of channel numbers while maintaining computational complexity. Experimental results demonstrate that this decomposition strategy can significantly reduce the computational cost of standard convolution while preserving model performance, providing feasibility for deployment in resource-constrained environments.

In terms of attention mechanisms, this study introduces an innovative channel attention module Efficient Channel Attention (ECA). This module achieves a balance between performance and efficiency through two key designs: First, it employs global average pooling to compress the channel dimension of feature maps, capturing channel-wise statistics with a global receptive field. Second, it innovatively uses 1D convolution to model local dependencies between channels, with an adaptive kernel size determination mechanism based on the channel count C, ensuring optimal interaction coverage for feature maps of varying scales. Compared to traditional channel attention mechanisms, the ECA module offers three notable advantages: (1) it avoids unnecessary dimensionality reduction, preserving complete channel information; (2) it reduces computational complexity through a local cross-channel interaction strategy; and (3) its adaptive kernel design ensures effective fusion of features at different network depths.Figure 1 shows the Network architecture of our method.

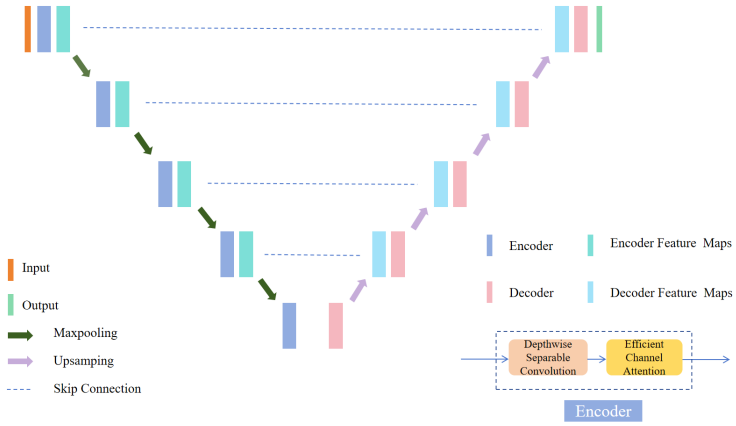

Fig. 1. Network architecture

In the encoder part of U-Net, the traditional convolution layer is replaced by depthwise separable convolution, and an efficient channel attention mechanism is added, which can improve the model's performance without significantly increasing the computational burden. Traditional BatchNorm is replaced by GroupNorm to reduce the model's dependence on batch size and enhance generalization across different batch sizes. Reducing the number of channels in the model to 16 decreases the computational load while maintaining network performance, improving operational efficiency, and making the model more suitable for deployment in environments with limited computing resources. These improvements collectively enhance the performance and efficiency of the network, making it more suitable for medical image segmentation tasks while maintaining the model's lightweight and fast response capabilities.

Loss function: we use the summation between Dice loss and Cross-Entropy loss because compound loss functions have been proven to be robust in various medical image segmentation tasks [13].

2.3 Post-processing

The use of a 3x3x3 closed operation in the post-processing stage of multi-organ segmentation can smooth organ boundaries, fill small holes, and connect adjacent organs, thereby improving the accuracy and robustness of segmentation. This operation helps reduce noise and artifacts, optimizing the segmentation results.

3 Experiments

3.1 Dataset and evaluation measures

The dataset is curated from more than 40 medical centers under the license permission, including TCIA [3], LiTS [1], MSD [21], KiTS [8,9], autoPET [6,5], AMOS [12], AbdomenCT-1K [19], TotalSegmentator [22], and past FLARE challenges [16,17,18]. The training set includes 2050 abdomen CT scans where 50 CT scans with complete labels and 2000 CT scans without labels. The validation and testing sets include 250 and 300 CT scans, respectively. The annotation process used ITK-SNAP [24], nnU-Net [11], MedSAM [14], and Slicer Plugins [4,15].

The evaluation metrics encompass two accuracy measures—Dice Similarity Coefficient (DSC) and Normalized Surface Dice (NSD)—alongside one efficiency measures—runtime. These metrics collectively contribute to the ranking computation. During inference, GPU is not available where the algorithm can only rely on CPU.

3.2 Implementation details

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

Table 1. Development environments and requirements.

System	Windows 10
CPU	Intel(R) Core(TM) i5-12400F CPU@2.50GHz
RAM	4×4GB; 3200MT/s
Programming language	Python 3.8
Deep learning framework	torch 1.7.1, torchvision 0.8.2
Specific dependencies	collections, pandas, scipy

Training protocols We used the dataset of 50 labeled cases provided by the organizers, along with 2000 pseudo-labels, which were generated by the aladdin5 team based on the training of these 50 labeled cases. During model training, we utilized the Adam optimizer with an initial learning rate of 0.001 and employed the CosineAnnealingLR strategy to gradually reduce the learning rate. The training protocols are presented in Table 2.

Table 2. Training protocols.

Network initialization	
Batch size	8
Patch size	3×192×192
Total epochs	100
Optimizer	AdamW
Initial learning rate (lr)	0.001
Lr decay schedule	CosineAnnealingLR
Training time	300 hours
Loss function	CrossEntropyLoss
Number of model parameters	0.405M[1]
Number of flops	1.29G[2]
CO_2eq	1 Kg[3]

4 Results and discussion

4.1 Quantitative results on validation set

The performance metrics on our validation set are showcased in Table 3. We can see the average organs DSC score, organs NSD score are 0.6615, 0.6581. When we employ the original U-Net, the average organs DSC score, organs NSD score are 0.7563, 0.7635.

Table 3. Quantitative evaluation results.

Target	Public Validation		Online Validation	
	DSC(%)	NSD(%)	DSC(%)	NSD(%)
Liver	94.70 ± 5.52	88.27 ± 9.18	95.33	90.22
Right Kidney	77.94 ± 24.03	68.64 ± 23.71	82.78	75.00
Spleen	87.99 ± 19.76	81.92 ± 20.75	90.75	86.18
Pancreas	63.41 ± 15.08	68.89 ± 15.51	65.32	73.09
Aorta	91.22 ± 11.44	91.28± 13.91	93.71	94.49
Inferior vena cava	76.29 ± 18.28	68.97± 19.66	78.88	72.87
Right adrenal gland	2.00 ± 14.00	2.00 ± 14.00	1.00	1.00
Left adrenal gland	4.00 ± 19.60	4.00 ± 19.60	1.00	1.00
Gallbladder	56.99 ± 35.49	51.58 ±35.14	67.53	63.70
Esophagus	70.68 ± 19.57	77.71± 22.23	69.45	79.14
Stomach	75.93 ± 21.08	70.67± 21.13	79.90	74.92
Duodenum	49.60 ± 17.14	67.24 ± 19.61	50.54	69.73
Left kidney	77.26 ± 25.23	65.83 ± 25.61	83.78	74.15
Average	63.69 ± 35.04	62.08± 34.21	66.15	65.81

Table 4. Quantitative evaluation of segmentation efficiency in terms of the running time(Proposed). Evaluation CPU: Intel Xeon(R) W-2133 CPU @ 3.60GHz × 12.

Case ID	Image Size	Running Time (s)
0059	(512, 512, 55)	23.10
0005	(512, 512, 124)	44.83
0159	(512, 512, 152)	53.95
0176	(512, 512, 218)	75.70
0112	(512, 512, 299)	102.11
0135	(512, 512, 316)	108.43
0150	(512, 512, 457)	151.94
0134	(512, 512, 597)	210.54

Table 5. Quantitative evaluation of segmentation efficiency in terms of the running time(Baseline U-Net). Evaluation CPU: Intel Xeon(R) W-2133 CPU @ 3.60GHz × 12.

Case ID	Image Size	Running Time (s)
0059	(512, 512, 55)	51.36
0005	(512, 512, 124)	112.98
0159	(512, 512, 152)	138.17
0176	(512, 512, 218)	202.17
0112	(512, 512, 299)	269.03
0135	(512, 512, 316)	284.78
0150	(512, 512, 457)	411.40
0134	(512, 512, 597)	561.44

4.2 Qualitative results on validation set

Figure 2 presents the segmentation outcomes achieved by our method, clearly demonstrating its superior performance in segmenting larger organs compared to smaller ones. Additionally, over-segmentation phenomena have also occurred.

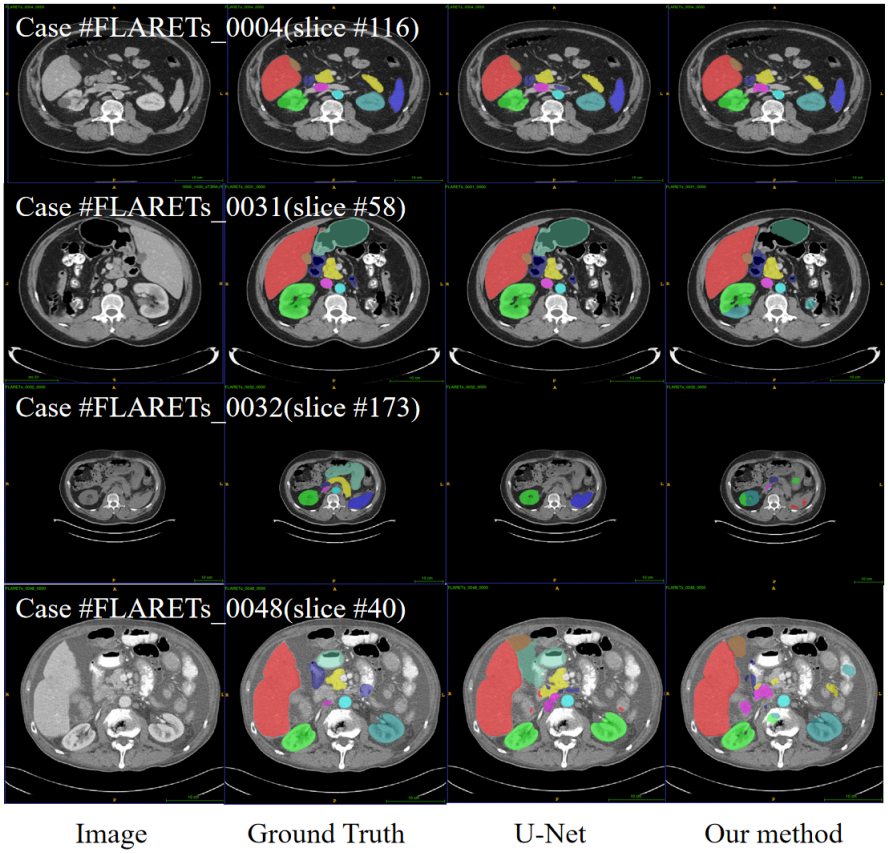

Fig. 2. Two examples with good segmentation results and two examples with bad segmentation results in the validation set.

4.3 Segmentation efficiency results on validation set

The segmentation efficiency results of the validation set(some cases) are shown in Table 4. The segmentation time for multiple organs ranges from a minimum of 23.10 s to a maximum of 210.54 s. The average running time is 73.62 s per case in inference phase(200 cases). However, the average time required for segmentation per case using the original U-Net network is 562 seconds.

The segmentation efficiency results of the ablation study are shown in Table 5. It strongly demonstrates the effectiveness of incorporating depthwise separable convolutions into the network.

4.4 Results on final testing set

As shown in Table 6, our model demonstrates balanced and stable performance across datasets from Asia, Europe, and North America, validating its robust cross-regional generalization capabilities. Despite potential differences in anatomical structures or data distributions across regions, the model exhibits minimal performance variations in core metrics (DSC, NSD) and inference efficiency, highlighting its adaptability to diverse scenarios. Europe achieves slightly higher mean DSC (63.4%) and NSD (63.0%), reflecting its strong robustness in handling complex cases. Meanwhile, Asia and North America, though marginally lower in absolute values, maintain concentrated data distributions (e.g., Asia's DSC standard deviation of ±8.8 and North America's inference time fluctuation of ±9.9 seconds), indicating reliable output consistency under varying data sources. Additionally, North America's average inference speed of 34.9 seconds and Asia's stability at 35.1 seconds provide flexible technical support for diverse clinical applications, such as real-time diagnostics and multi-center collaborative analysis. Overall, the model exhibits neither extreme deviations nor performance degradation due to anatomical variations during cross-regional testing, demonstrating exceptional robustness and universality in multi-center data scenarios, thereby laying a solid foundation for clinical generalization.

Table 6. Quantitative evaluation of test results.

Target	DSC(%)		NSD(%)		Time(s)	
	Mean	Median	Mean	Median	Mean	Median
Asian	61.1 ± 8.8	62.9(55.2,69.0)	61.6 ± 10.0	63.6(55.9,69.6)	35.1 ± 12.7	33.4(25.3,39.1)
European	63.4 ± 9.9	66.3(58.3,71.2)	63.0 ± 11.4	65.9(56.6,71.6)	42.7 ± 12.9	40.6(33.1,50.7)
North America	59.4 ± 7.5	60.3(54.9,64.7)	57.5 ± 8.5	58.9(52.4,63.7)	34.9± 9.9	32.4(29.3,37.7)

4.5 Limitation and future work

In terms of segmentation efficiency, our proposed approach demonstrates a marked enhancement compared to the conventional U-Net. Nevertheless, the segmentation outcomes reveal that our model struggles with the precise delineation of smaller anatomical structures, notably the right and left adrenal glands. Moving forward, we are committed to prioritizing the enhancement of segmentation accuracy for these minute organs.

5 Conclusion

In our current study, we have opted for depthwise separable convolutions to replace the standard convolutional layers, aiming to streamline the compu-

tational process. This modification is particularly effective for the precision demands of medical image segmentation. Following this, we have integrated ECA into our encoder to refine feature extraction by emphasizing the most salient channels. The strategic application of these techniques not only enhances our model's performance but also suggests their potential utility in a broader spectrum of medical image segmentation applications. By refining our approach in this manner, we hope to contribute to the advancement of medical imaging technologies and inspire further research in this critical field.

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 CodaBench [23] for hosting the challenge platform. This work was funded by the National Natural Science Foundation of China (62271149), Fuzhou Science Technology Project(2023-P-001), Fujian Science Technology Project (2022L3003,2021H0013,2020Y9091,2022Y4014).

Disclosure of Interests

The authors declare no competing interests.

References

1. 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) 4
2. Chollet, F.: Xception: Deep learning with depthwise separable convolutions pp. 1251–1258 (2017) 1

3. 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) 4
4. 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) 4
5. Gatidis, S., Früh, M., Fabritius, M., Gu, S., Nikolaou, K., La Fougère, C., Ye, J., He, J., Peng, Y., Bi, L., et al.: The autopet challenge: Towards fully automated lesion segmentation in oncologic pet/ct imaging. Nature Machine Intelligence (in presss) (2024) 4
6. 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) 4
7. Guo, M.H., Xu, T.X., Liu, J.J., Liu, Z.N., Jiang, P.T., Mu, T.J., Zhang, S.H., Martin, R.R., Cheng, M.M., Hu, S.M.: Attention mechanisms in computer vision: A survey. Computational visual media **8**(3), 331–368 (2022) 2
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) 4
9. 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) 4
10. Howard, A.G.: Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv preprint arXiv:1704.04861 (2017) 2
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) 4
12. Ji, Y., Bai, H., GE, C., Yang, J., Zhu, Y., Zhang, R., Li, Z., Zhanng, L., Ma, W., Wan, X., Luo, P.: Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation. Advances in Neural Information Processing Systems **35**, 36722–36732 (2022) 4
13. Ma, J., Chen, J., Ng, M., Huang, R., Li, Y., Li, C., Yang, X., Martel, A.L.: Loss odyssey in medical image segmentation. Medical Image Analysis **71**, 102035 (2021) 4
14. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**, 654 (2024) 4
15. Ma, J., Kim, S., Li, F., Baharoon, M., Asakereh, R., Lyu, H., Wang, B.: Segment anything in medical images and videos: Benchmark and deployment. arXiv preprint arXiv:2408.03322 (2024) 4

16. Ma, J., Zhang, Y., Gu, S., An, X., Wang, Z., Ge, C., Wang, C., Zhang, F., Wang, Y., Xu, Y., Gou, S., Thaler, F., Payer, C., Štern, D., Henderson, E.G., McSweeney, D.M., Green, A., Jackson, P., McIntosh, L., Nguyen, Q.C., Qayyum, A., Conze, P.H., Huang, Z., Zhou, Z., Fan, D.P., Xiong, H., Dong, G., Zhu, Q., He, J., Yang, X.: Fast and low-gpu-memory abdomen ct organ segmentation: The flare challenge. Medical Image Analysis **82**, 102616 (2022) 4
17. Ma, J., Zhang, Y., Gu, S., Ge, C., Ma, S., Young, A., Zhu, C., Meng, K., Yang, X., Huang, Z., Zhang, F., Liu, W., Pan, Y., Huang, S., Wang, J., Sun, M., Xu, W., Jia, D., Choi, J.W., Alves, N., de Wilde, B., Koehler, G., Wu, Y., Wiesenfarth, M., Zhu, Q., Dong, G., He, J., the FLARE Challenge Consortium, Wang, B.: Unleashing the strengths of unlabeled data in pan-cancer abdominal organ quantification: the flare22 challenge. Lancet Digital Health (2024) 4
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) 4
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) 4
20. Moor, M., Banerjee, O., Abad, Z.S.H., Krumholz, H.M., Leskovec, J., Topol, E.J., Rajpurkar, P.: Foundation models for generalist medical artificial intelligence. Nature **616**(7956), 259–265 (2023) 2
21. 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) 4
22. 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) 4
23. 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) 9
24. 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) 4