The Fully Convolutional Transformer for Medical Image Segmentation Supplementary Material

Athanasios Tragakis[1*] Chaitanya Kaul[2*] Roderick Murray-Smith[2] Dirk Husmeier[1]

[1]Mathematics and Statistics, University of Glasgow, United Kingdom, G12 8QW
[2]School of Computing Science, University of Glasgow, United Kingdom, G12 8RZ

1. Additional Tables

Table 1 and Table 2 summarize our results on the Spleen Segmentation and ISIC 2017 skin cancer segmentation datasets respectively.

Table 3 summarizes our results for the ablations conducted to observe the effect of removing skip connections from the model architecture.

Table 4 summarizes the key differences of our Fully Convolutional Transformer, compared to existing medical image segmentation methods in literature.

Table 5 show our results on the ACDC Post-2017-MICCAI online leaderboard. Compared to state-of-the-art methods, our FCT gets the most well rounded segmentation results while also getting the best results for 6 out of 8 classes, and the best overall dice score got from averaging all classes.

Method	Dice
Swin UNet [2]	90.7
UNETR [7]	94.1
SETR PUP [20]	94.9
TransUNet [3]	95.0
CoTr* [17]	94.6
CoTr [17]	95.4
FCT_{224}	95.9
FCT_{384}	**96.6**

Table 1: Segmentation results on Spleen dataset. CoTr* denotes the CoTr model without a CNN encoder. Swin UNet is trained on a 224×224 image size.

Method	Dice	Sensitivity
UNet [13]	76.81	72.14
FocusNet [9]	83.15	76.73
FocusNet++ [10]	84.04	82.22
TransFuse [19]	81.78	80.18
TransUNet [3]	81.74	80.78
BAT [15]	84.86	84.62
FCT_{224}	85.46	84.02
FCT_{384}	**85.98**	**85.29**

Table 2: Segmentation results on ISIC 2017 dataset. All models are trained on image size 384×384 unless specified differently. FCT_{224} and FCT_{384} contain 9.2 million parameters.

Skip connections	Avg.	RV	MYO	LV
0	81.90	82.33	75.23	88.14
1	84.06	83.69	78.19	90.31
2	89.45	88.22	86.18	93.94
3	90.74	89.68	87.77	94.78
4	**92.11**	**91.6**	**89.3**	**95.5**

Table 3: Ablation study on the impact of the number of skip connections on the ACDC dataset. FCT_{224} (with 16.1 million parameters) is used for these ablations.

*Equal Contribution

Method	Positional Encoding	Overlapping Patches	MHSA Projection	Feature Processing
ViT [4]	✓	✗	Linear	Linear
R50 ViT [4]	✓	✗	Linear	Linear
TransUNet [3]	✓	✗	Linear	Linear
UNETR [7]	✓	✗	Linear	Linear
LeVit-UNet-384 [18]	✓	✗	Linear	Linear
Swin UNet [2]	✓	✓	Linear	Linear
nnFormer [21]	✓	✓	Linear	Linear
D-Former [16]	✓	✓	Linear	Linear
FCT (Ours)	✗	✓	Conv	Wide-Focus

Table 4: Comparing representative works with the *Fully Convolutional Transformer*.

Method	Avg.$_{ED}$	RV$_{ED}$	MYO$_{ED}$	LV$_{ED}$	Avg.$_{ES}$	RV$_{ES}$	MYO$_{ES}$	LV$_{ES}$	Overall Avg.
Painchaud [12]	92.50	93.3	88.1	96.1	89.73	88.4	89.7	91.1	91.12
Zotti [22]	92.77	93.4	88.4	96.4	89.70	88.5	89.6	91.2	91.23
Baumgartner [1]	92.90	93.2	89.2	96.3	89.83	88.3	90.1	91.1	91.37
Khened [11]	92.93	93.5	88.9	96.4	89.80	87.9	89.8	91.7	91.37
Simantiris [14]	93.13	-	-	96.7	90.70	-	-	92.8	91.92
Girum [5]	93.37	-	-	-	90.50	-	-	-	91.93
Isensee [8]	94.07	94.6	89.6	96.7	**91.83**	**90.4**	91.9	92.8	92.95
Guo [6]	94.30	-	-	-	91.37	-	-	-	93.02
FCT$_{512}$ (Ours)	**94.33**	**95.0**	**91.0**	**97.0**	91.67	90.0	**92.0**	**93.0**	**93.13**

Table 5: Detailed top 8 results on the ACDC Post-2017-MICCAI online leaderboard. FCT$_{512}$ (with 31.7 million parameters) is used for this experiment. Avg. stands for the Average dice Coefficient. ED Stands for End-Diastolic frames. ES stands for End-Systolic frames. - values are not available on the public leaderboard.

References

[1] Christian F. Baumgartner, Lisa M. Koch, Marc Pollefeys, and Ender Konukoglu. An Exploration of 2D and 3D Deep Learning Techniques for Cardiac MR Image Segmentation. In Mihaela Pop, Maxime Sermesant, Pierre-Marc Jodoin, Alain Lalande, Xiahai Zhuang, Guang Yang, Alistair Young, and Olivier Bernard, editors, *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges*, pages 111–119, Cham, 2018. Springer International Publishing.

[2] 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. *arXiv preprint arXiv:2105.05537*, 2021.

[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] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*, 2020.

[5] Kibrom Berihu Girum, Youssef Skandarani, Raabid Hussain, Alexis Bozorg Grayeli, Gilles Créhange, and Alain Lalande. Automatic myocardial infarction evaluation from delayed-enhancement cardiac MRI using deep convolutional networks. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 378–384. Springer, 2020.

[6] Fumin Guo, Matthew Ng, Maged Goubran, Steffen E. Petersen, Stefan K. Piechnik, Stefan Neubauer, and Graham Wright. Improving cardiac MRI convolutional neural network segmentation on small training datasets and dataset shift: A continuous kernel cut approach. *Medical Image Analysis*, 61:101636, 2020.

[7] Ali Hatamizadeh, Yucheng Tang, Vishwesh Nath, Dong Yang, Andriy Myronenko, Bennett Landman, Holger R Roth, and Daguang Xu. UNETR: Transformers for 3D medical image segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 574–584, 2022.

[8] Fabian Isensee, Paul F Jaeger, Peter M Full, Ivo Wolf, Sandy Engelhardt, and Klaus H Maier-Hein. Automatic cardiac disease assessment on cine-MRI via time-series segmentation and domain specific features. In *International workshop on statistical atlases and computational models of the heart*, pages 120–129. Springer, 2017.

[9] Chaitanya Kaul, Suresh Manandhar, and Nick Pears. Focusnet: An attention-based fully convolutional network for medical image segmentation. In *2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019)*, pages 455–458. IEEE, 2019.

[10] Chaitanya Kaul, Nick Pears, Hang Dai, Roderick Murray-Smith, and Suresh Manandhar. Focusnet++: Attentive aggregated transformations for efficient and accurate medical image segmentation. In *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, pages 1042–1046. IEEE, 2021.

[11] Mahendra Khened, Varghese Alex, and Ganapathy Krishnamurthi. Densely connected fully convolutional network for short-axis cardiac cine MR image segmentation and heart diagnosis using random forest. In *International Workshop on Statistical Atlases and Computational Models of the Heart*, pages 140–151. Springer, 2017.

[12] Nathan Painchaud, Youssef Skandarani, Thierry Judge, Olivier Bernard, Alain Lalande, and Pierre-Marc Jodoin. Cardiac segmentation with strong anatomical guarantees. *IEEE transactions on medical imaging*, 39(11):3703–3713, 2020.

[13] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[14] Georgios Simantiris and Georgios Tziritas. Cardiac MRI Segmentation With a Dilated CNN Incorporating Domain-Specific Constraints. *IEEE Journal of Selected Topics in Signal Processing*, 14(6):1235–1243, 2020.

[15] Jiacheng Wang, Lan Wei, Liansheng Wang, Qichao Zhou, Lei Zhu, and Jing Qin. Boundary-Aware Transformers for Skin Lesion Segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 206–216. Springer, 2021.

[16] Yixuan Wu, Kuanlun Liao, Jintai Chen, Danny Z Chen, Jinhong Wang, Honghao Gao, and Jian Wu. D-Former: A U-shaped Dilated Transformer for 3D Medical Image Segmentation. *arXiv preprint arXiv:2201.00462*, 2022.

[17] Yutong Xie, Jianpeng Zhang, Chunhua Shen, and Yong Xia. Cotr: Efficiently bridging cnn and trans-

former for 3d medical image segmentation. In *International conference on medical image computing and computer-assisted intervention*, pages 171–180. Springer, 2021.

[18] Guoping Xu, Xingrong Wu, Xuan Zhang, and Xinwei He. Levit-unet: Make faster encoders with transformer for medical image segmentation. *arXiv preprint arXiv:2107.08623*, 2021.

[19] Yundong Zhang, Huiye Liu, and Qiang Hu. Transfuse: Fusing transformers and cnns for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 14–24. Springer, 2021.

[20] Sixiao Zheng, Jiachen Lu, Hengshuang Zhao, Xiatian Zhu, Zekun Luo, Yabiao Wang, Yanwei Fu, Jianfeng Feng, Tao Xiang, Philip HS Torr, et al. Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6881–6890, 2021.

[21] Hong-Yu Zhou, Jiansen Guo, Yinghao Zhang, Lequan Yu, Liansheng Wang, and Yizhou Yu. nnFormer: Interleaved Transformer for Volumetric Segmentation. *arXiv preprint arXiv:2109.03201*, 2021.

[22] Clement Zotti, Zhiming Luo, Alain Lalande, and Pierre-Marc Jodoin. Convolutional Neural Network With Shape Prior Applied to Cardiac MRI Segmentation. *IEEE Journal of Biomedical and Health Informatics*, PP:1–1, 08 2018.