DeepCRC: Colorectum and Colorectal Cancer Segmentation in CT Scans via Deep Colorectal Coordinate Transform

Lisha Yao[1,2,*], Yingda Xia[3,*(✉)], Haochen Zhang[1,2], Jiawen Yao[3], Dakai Jin[3], Bingjiang Qiu[1], Yuan Zhang[1], Suyun Li[1,2], Yanting Liang[1,2], Xiansheng Hua[3], Le Lu[3], Xin Chen[2,4], Zaiyi Liu[1,2], and Ling Zhang[3]

[1]Guangdong Provincial People's Hospital,[2]South China University of Technology, [3]Alibaba Group,[4]Guangzhou First People's Hospital

Abstract. We propose DeepCRC, a topology-aware deep learning-based approach for automated colorectum and colorectal cancer (CRC) segmentation in routine abdominal CT scans. Compared with MRI and CT Colonography, regular CT has a broader application but is more challenging. Standard segmentation algorithms often induce discontinued colon prediction, leading to inaccurate or completely failed CRC segmentation. To tackle this issue, we establish a new 1D colorectal coordinate system that encodes the position information along the colorectal elongated topology. In addition to the regular segmentation task, we propose an auxiliary regression task that directly predicts the colorectal coordinate for each voxel. This task integrates the global topological information into the network embedding and thus improves the continuity of the colorectum and the accuracy of the tumor segmentation. To enhance the model's architectural ability of modeling global context, we add self-attention layers to the model backbone, and found it complementary to the proposed algorithm. We validate our approach on a cross-validation of 107 cases and outperform nnUNet by an absolute margin of 1.3% in colorectum segmentation and 8.3% in CRC segmentation. Notably, we achieve comparable tumor segmentation performance with the human inter-observer (DSC: 0.646 vs. 0.639), indicating that our method has similar reproducibility as a human observer.

Keywords: colorectal cancer · colorectal coordinate · segmentation.

1 Introduction

Colorectal cancer (CRC) is the third most common cancer and the second leading cause of cancer-related death worldwide [16]. Computed tomography (CT), which provides location and morphology information of CRC, is routinely performed for cancer detection, diagnosis, staging, surgical planning, and treatment response monitoring [3], as well as structural evaluation of the entire colorectum.

---

* Equal contribution. ✉ Corresponding author (yingda.xia@alibaba-inc.com).

Table 1. A conceptual comparison between our work and existing studies on colorectal cancer segmentation.

Studies	Modality	Scan Range	2D/3D	Preparation	Cost	Challenge
[7] [24]	MRI	rectum	2D	no	high	low
[5] [15] [8]	MRI	rectum	3D	no	high	low
[13] [10]	CTC	colon+rectum	2D	yes	medium	medium
Decathlon [2] Task08 [6, 17]	CT	colon+rectum	3D	yes	low	high
Ours	CT	colon+rectum	3D	no	low	high

Automatic segmentation of colorectum, especially CRC, is the key to achieving automation of these clinical tasks, which can greatly increase the efficiency, reproducibility, and potentially accuracy of the entire clinical workflow.

In this paper, we aim to provide the first exploration of automated segmentation of the colorectum (colon and rectum) and CRC in routine abdominal CT scans (without bowel preparation). The comparison between our work and existing studies in CRC segmentation is shown in Table 1. Compared with MRI [5, 7, 8, 15, 24], CT is much less costly and is generally superior to MRI for the hollow viscera (colon). In clinical practice, MRI mainly allows staging rectal cancer only. For CT Colonography (CTC), it is still not widely implemented because bowel preparation and distention is time-consuming and sometimes causes adverse events, such as examination-related pain, and vasovagal syncope or presyncope [21]. The CRC segmentation task in Medical Segmentation Decathlon [2] uses CT but still after bowel preparation with barium coating for contrast enhancement. As a result, the successful segmentation of the colorectum and tumor in routine abdominal CT without bowel preparation will have a broader application and impact.

However, this task is also more challenging due to the following reasons. (i) The colorectum takes up a large space in the abdomen, and it is always hard to be continually traced caused by the mixing of other organs such as the small intestine. (ii) CRCs are usually small and hard to distinguish from the contents of colorectum in regular CT scans (without bowel preparation), compared with a) CT Colonography, which has clean and adequate colon distention by laxative purgation and colonic insufflation, and b) pelvic MRI, which only contains the rectum structure. Furthermore, the discontinued segmentation of colorectum might eventually cause the misdetection of CRC.

To tackle these two challenges, we propose *Deep Colorectal Coordinate Transform* to improve the accuracy of colorectum and CRC segmentation at the same time. Topologically, the colorectum has a single-path and continual structure that extends between the caecum and the rectum. This special pattern motivates us to propose an auxiliary voxel-wise regression problem to improve segmentation continuity. As shown in an illustrative example in Fig. 1, we set up a new one-dimensional coordinate system based on the centerline of the colorectum. Technically, we transform the 3D voxel space into 1D colorectal coordinate space by projecting each foreground colorectum voxel into the 1D coordinate system,

thus obtaining a normalized scalar for each voxel. In addition to the voxel-wise classification in conventional segmentation task formulation, we directly regress the relative position, *i.e.* the coordinate value in this coordinate system, of each foreground colorectum voxel. This auxiliary task forces the network to learn the global structural information of the colorectum and the positional information of CRCs, and thus helps the network to achieve better segmentation performance. In terms of network architecture design, we add self-attention layers with positional embedding to enhance the ability to model global context. This design not only provides the architectural basis for the proposed coordinate regression task, but improves the ability of tumor and non-tumor distinction as well.

Our approach is related to recent advance in 3D medical image segmentation with deep networks [6, 11, 14] and approaches beyond the voxel space, such as boundary projection [12], distance transform [20], and mesh models [22]. For example, Ni et al. proposed a 3D segmentation framework with an elastic boundary projection to obtain the 2D surface of subjects [12]. Yao et al. learned 3D geometry of organs using a 3D mesh representation for better segmentation performance [22]. These methods take full advantage of topology information and achieve promising segmentation accuracy. Moreover, recent works in attention models [4, 18, 19] also motivate us to enhance the global context of 3D CNNs with self-attention layers.

We validate our proposed method on an in-house dataset, including 107 CT scans with manual colorectum and CRC annotations. Our approach outperforms a strong baseline nnUNet [6] by an absolute DSC margin of 1.3% in colorectum segmentation and 8.3% in CRC segmentation. Moreover, the CRC segmentation performance of our approach (DSC=0.646) is comparable with the inter-observer variability (DSC=0.635), illustrating a strong potential for clinical application.

2   Method

Problem statement. We aim at colorectum and tumor (CRC) segmentation in contrast-enhanced CT scan. For each patient, we have an image $\mathbf{X}$ and its corresponding label $\mathbf{Y}$ in the venous phase. We denote the whole dataset as $S = \{(\mathbf{X_i}, \mathbf{Y_i})|i = 1, 2, ..M\}$, where $\mathbf{X_i} \in \mathbb{R}^{H_i \times W_i \times D_i}$ is a 3D volume representing the CT scan of the i-th patient. $\mathbf{Y_i} \in \mathcal{L}^{H_i \times W_i \times D_i}$ is a voxel-wise annotated label with the same three dimensional size (H_i, W_i, D_i) as $\mathbf{X_i}$. $\mathcal{L} = \{0, 1, 2\}$ represents our segmentation targets, *i.e.*, background, colorectum, and tumor (CRC).

2.1   Colorectal Coordinate Transform

In this section, we discuss how to set up the colorectal coordinate system and how to transform each voxel in $\mathbf{Y}$ into the coordinate system. The output of this process is the groundtruth coordinate map $\mathbf{E}$ which is only used in the training phase. The overall algorithm is described in Algorithm 1 and visually illustrated in Fig. 1 "Colorectal Coordinate Transform".

First, we extract the centerline $\mathbf{C}$ of the colorectum based on the ground truth label map $\mathbf{Y}$. This process is denoted as $\mathbf{C} = f_{cl}(\mathbf{Y})$, where f_{cl} is a centerline
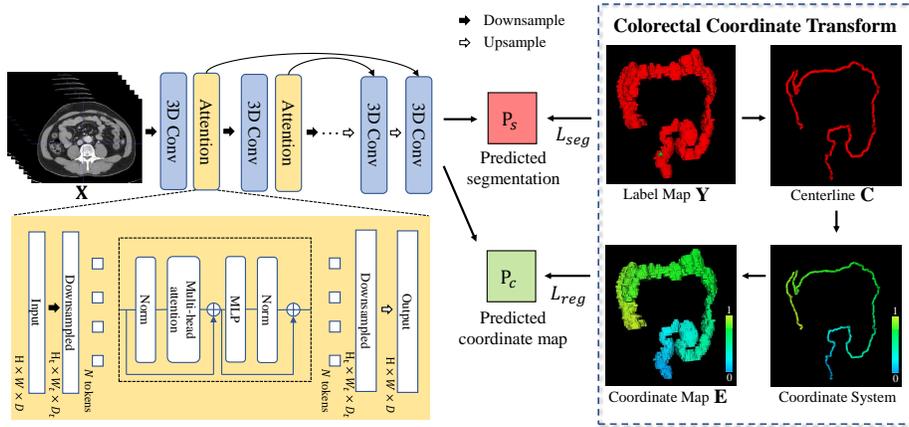
Fig. 1. The pipeline of the proposed method. Our network has an encoder-decoder architecture with self-attention layers during downsampling. In the training phase, we transform the label map $\mathbf{Y}$ into the groundtruth coordinate map $\mathbf{E}$ via "Colorectal Coordinate Transform". Our network takes the input of 3D CT images and outputs the segmentation $\mathbf{P}_s$ (supervised by $\mathcal{L}_{seg}$) and the voxel-wise coordinate map $\mathbf{P}_c$ (supervised by $\mathcal{L}_{reg}$). In the testing phase, we only use $\mathbf{P}_s$ as the segmentation prediction.

extraction algorithm and $\mathbf{C}$ is the extracted 3D centerline image of the same size as $\mathbf{X}$ and $\mathbf{Y}$. We use a robust centerline extraction algorithm [9] to avoid the false centerlines easily produced by the irregular colorectum boundaries. $\mathbf{C}$ is also guaranteed to be one-voxel thick with 26-connectivity. The automated algorithm might fail to extract the correct centerline of some regions where the bowel has severe adhesion. We correct these centerlines semi-automatically by erasing to split these adhesion in $\mathbf{Y}$ and rerunning the algorithm. An example of the extracted $\mathbf{C}$ is shown in Fig. 1 (Centerline $\mathbf{C}$).

Second, we build a colorectal coordinate map $\mathbf{E}$ which is initialized as a all-zero matrix with the same shape as $\mathbf{Y}$. Then we find the lowest foreground position j on the centerline $\mathbf{C}$. We use this position as the starting point to trace the centerline and mark it incrementally on $\mathbf{E}$, which is further normalized to the range of $[0, 1]$ (Fig. 1 Coordinate system).

Finally, we propagate the coordinates along the centerline to the foreground voxels in $\mathbf{Y}$. For each foreground position p, we find the nearest point q on the centerline, and update $\mathbf{E}^p$ with the same coordinate of q on the centerline. After this step, we transform the groundtruth label $\mathbf{Y} \in$ into a coordinate map $\mathbf{E} \in [0, 1]^{H \times W \times D}$ (Fig. 1 Coordinate map $\mathbf{E}$).

2.2   Network Training

Fig. 1 shows the training diagram of the proposed method. Our network has two outputs, *i.e.*, the regular segmentation prediction $\mathbf{P}_s$ and the coordinate map prediction $\mathbf{P}_c$. Following nnUNet [6], the segmentation loss (denoted as $\mathcal{L}_{seg}$) is

---

Algorithm 1 Colorectal Coordinate Transform

Input: Ground truth label map $\mathbf{Y} \in \mathcal{L}^{H \times W \times D}$
Output: Coordinate map $\mathbf{E} \in [0,1]^{H \times W \times D}$

1: Extract the centerline (1-voxel thick and 26-connected): $\mathbf{C} = f_{cl}(\mathbf{Y})$
2: Find the lowest foreground position j on $\mathbf{C}$
3: Initialize zero map $\mathbf{E}$ with the same shape of $\mathbf{Y}$ and set $\mathbf{E}^j \leftarrow 1$
4: **while** $\exists$ unvisited position k in the 26-connectivity of j **do**
5:     $\mathbf{E}^k \leftarrow \mathbf{E}^j + 1$ ▷ Trace the centerline and mark it incrementally
6:     $j \leftarrow k$
7: Normalize D to the range of $[0,1]$: $\mathbf{E} \leftarrow \frac{\mathbf{E}}{\max_i(\mathbf{E}^i)}$.   ▷ Coordinate system
8: **for** each foreground position p on the label map $\mathbf{Y}$ **do**
9:     Find its nearest point q on the centerline C
10:     $\mathbf{E}^p \leftarrow \mathbf{E}^q$ ▷ Project the coordinates to each foreground voxel
11: **return** $\mathbf{E}$

---

defined as a summation of cross-entropy loss and Dice loss [11]. Additionally, we define a regression loss $\mathcal{L}_{reg}$ that minimizes the difference between the predicted coordinate map $\mathbf{P}_c$ and the generated coordinate map $\mathbf{E}$.

$$\mathcal{L}_{reg} = \sum_j ||\mathbf{P}_c^j - \mathbf{E}^j||^2 \tag{1}$$

where j is the j-th voxel of $\mathbf{E}$. With a controlling loss scaler α, the final loss function $\mathcal{L}$ is defined as:

$$\mathcal{L} = \mathcal{L}_{seg} + \alpha \mathcal{L}_{reg} \tag{2}$$

In the testing phase, the trained network takes only an image $\mathbf{X}$ as the input and simply keeps the segmentation prediction $\mathbf{P}_s$ as the final output. The process of colorectal coordinate transform is not needed.

2.3   Network Architecture

In terms of architectural improvement, we integrate the global self-attention layer to enhance the model's ability to model global context. Our proposed auxiliary coordinate regression loss (Eq. 1) requires the network to understand of the colorectum's topology globally. However, vanilla UNet-based segmentation networks heavily rely on local textual change and have a limited receptive field. Thus, an architectural improvement in the global context is desired.

As shown in Figure 1, our network has an encoder-decoder achitecture with skip connections. We add self-attention layers after each downsampling block in the encoder. In each self-attention layer, we first downsample the feature map to a fixed spatial size (H_t, W_t, D_t) and reshape the feature map to obtain N tokens ($N = H_t \times W_t \times D_t$). We then add a learnable positional embedding to the tokens before forwarding the tokens to multi-head attention layers. Finally, we reshape the tokens back to (H_t, W_t, D_t) and upsample it to the original size of the feature map.

3   Experiments

3.1   Dataset and Annotation

In this study, we retrospectively collected 3D volumetric venous phase CT from 107 patients (including 75 males and 32 females, aged from 30 to 89 years old) with colorectal cancers. The patients were injected with the contrast agents at the rate of 2.5-3.5 ml/s and scanned by a GE or Philips CT scanner operated at 120 kVp and 130-250 mAs. The median voxel spacing of the dataset is $0.78 \times 0.78 \times 5$ mm^3. The colorectum and CRC were manually segmented by an experienced radiologist (10-yr) using ITK-SNAP software [23]. More specifically, the colorectum was carefully traced slice-by-slice and annotated from scratch manually, which took about one hour for each volume on average. The CRC was annotated by referring to the corresponding clinical and pathological reports, as well as other CT phases if necessary, and further checked by a senior radiologist specialized in CRC imaging for 23 years. Additionally, to test the inter-observer variability of CRC segmentation, all 107 cases were detected and delineated by another medical student (2-yr in CRC imaging) with only venous phase CT provided.

3.2   Implementation Details

We conduct five-fold cross-validation on 107 cases. In the training phase, we resample all images to the spacing of (2mm, 2mm, 5mm) and randomly crop patches of $(160, 160, 80)$ voxels for the network input. Our approach is built on the nnUNet [6] framework with 5 downsampling blocks and 5 upsampling blocks with skip connections. The self-attention layers are appended after each downsampling block where the spatial sizes are reduced to $(10, 10, 10)$, and thus, the number of tokens is 1000 for the multi-head attention. We use RAdam optimizer with an initial learning rate of 0.001 and polynomial learning decay. The batch size is 2 and the total iteration is 75000. Standard augmentations including random flip, random rotation, and 3D elastic transformation are utilized. We set the regression loss scaler $\alpha = 50$ for the best model performance. We also set $\mathbf{E} \leftarrow \mathbf{E}+1$ and thus $\mathbf{E} \in [1, 2]$ to increase the coordinate value difference between the coordinate starting point (i.e., 1) and the background voxels (i.e., 0). In the testing phase, we resample the test sample to the training spacing and resample the model prediction to the original spacing. We use the sliding-window testing scheme with a stride of $(80, 80, 40)$ voxels. The average testing time for each volume is approximately 10 seconds on an Nvidia Tesla V100 16G GPU.

3.3   Results

Quantitative results. Dice-Sørensen coefficient (DSC), mean surface distance (MSD), Hausdorff distance (HD$_{95}$), and tumor detection rate (TDR) are used to evaluate the colorectum and tumor segmentation, and results over the 5-fold cross-validation of the whole dataset are reported. TDR measures the ratio of the

Table 2. Results on 5-fold cross-validation of the dataset. We compare the results of nnUNet, nnUNet with self-attention blocks (denoted as "nnUNet+self-attn."), nnUNet with regression loss ("nnUNet+$\mathcal{L}_{reg}$"), and our DeepCRC (combined). Results are reported as mean±std. The last two rows demonstrate the leading performances on the Decathlon [2] challenge (CRC segmentation on CT with bowel preparation).

	Methods	DSC↑	MSD(mm)↓	HD$_{95}$(mm)↓	TDR↑
Colorectum	nnUNet	0.849±0.072	1.19±1.13	6.08±5.56	-
	nnUNet+self-attn.	0.851±0.066	1.29±1.14	6.40±5.92	-
	nnUNet+$\mathcal{L}_{reg}$	0.864±0.060	1.08±1.02	5.03±4.91	-
	DeepCRC	0.862±0.058	1.07±0.96	5.07±4.74	-
Tumor	nnUNet	0.563±0.306	14.6±38.4	29.7±46.5	0.841
	nnUNet+self-attn.	0.616±0.287	20.4±49.3	33.0±55.2	0.869
	nnUNet+$\mathcal{L}_{reg}$	0.617±0.289	17.6±42.5	30.4±49.9	0.869
	DeepCRC	0.646±0.275	12.5±36.5	23.3±43.4	0.879
	Inter-observer	0.639±0.280	15.6±40.5	26.1±48.0	0.879
Decathlon-Task08	nnUNet [1]	0.583	-	-	-
Colon Cancer [2]	Swin UNETR [1]	0.595	-	-	-

Note: The median MSD and HD$_{95}$ of DeepCRC are 0.62 mm and 4.12 mm; the inter-observer variability are 1.38 mm and 5.75 mm. The large mean values of MSD and HD$_{95}$ are mainly caused by a small proportion of segmentations with larger errors.

cases that tumor is correctly detected in the dataset, where we use DSC > 0.1 as the criterion for successful detection. Table 2 compares our approach (DeepCRC) with nnUNet [6] and two ablation configurations of our contributions, denoted as "nnUNet+self-attention" and "nnUNet+$\mathcal{L}_{reg}$" respectively. In terms of the colorectum segmentation, the performances with the regression loss (the last two approaches) are higher than the segmentation-loss-only counterpart (the first two approaches), mainly because our auxiliary task to predict the colorectal coordinate provides another topology and location information for colorectum segmentation. As for the tumor segmentation, both self-attention and regression loss improve the segmentation performance (about 5% absolute improvement each in the DSC score). Moreover, these two contributions are complementary to each other, resulting in an even better performance (another 3% absolute improvement) when combined together, reaching a DSC score of 0.646.

In addition, for the state-of-the-art colon cancer segmentation results on a different but comparable dataset (decathlon challenge Task08 [2]), nnUNet is still among the top performers, with a DSC score of 0.583 [1] in CT imaging with bowel preparation (126 training data), which is slightly higher than nnUNet's DSC score of 0.563 on our CT imaging without bowel preparation (107 training data). Our approach (DSC=0.646) evidently outperforms the strong baseline nnUNet by over 8% in tumor DSC and has reached within the human inter-observer variability (0.639).

Qualitative analysis. We analyze the reason for our improvements by visual illustration of two cases segmented by different method configurations,
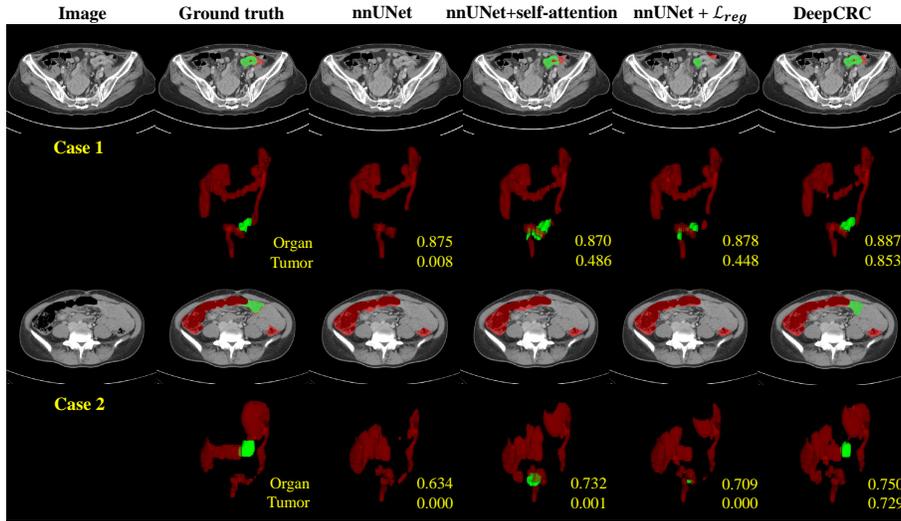
Fig. 2. The visual illustration of colorectum (red) and colorectal tumor (green) segmentation. The DSC scores of the organ and the tumor are marked at the bottom-right of each result. We compare our method (DeepCRC) with the nnUNet baseline and two other ablation experiments in Table 2. In the first case, our approach improves the continuity of the colorectum segmentation and the quality of the tumor segmentation. In the second case, our approach successfully detects the tumor while the other approaches fail, because we correctly predict the colon topology around the tumor.

i.e., nnUNet, +self-attention, $+\mathcal{L}_{reg}$, and DeepCRC (Fig. 2). In the first case, the nnUNet baseline predicts discontinuous colorectum while our DeepCRC improves the integrity of the colorectum prediction. The tumor predictions of the two ablation configurations are only partially correct (about 0.4 in DSC score) due to the discontinuity. Our approach predicts a relatively intact colorectum and an accurate tumor extent. The second case is more challenging due to the irregular colorectum shape. The baseline approaches all fail to detect the tumor, but our approach succeeds because our proposed method can reliably predict a continued topology around the tumor region and distinguishe the tumor from the organ and the background.

From the illustrated qualitative results, we hypothesize that the tumor will affect the appearance of the colorectum and make the affected region harder to be distinguished from non-colorectum regions. Our approach integrates the topological knowledge into the network embedding and enhances the global contextual information. This is beneficial to the improvement of continuity in colorectum segmentation and especially the ability to detect the colorectal tumor.

4   Conclusion

We propose DeepCRC, a colorectum and colorectal tumor segmentation framework in regular contrast-enhanced CT. We introduce an additional auxiliary regression task to directly predict the relative position of each voxel in the colorectal topology and self-attention layers to model global context. Experimental results show that when trained on only small-sized (n<100) data, our approach outperforms nnUNet with improved integrity in colorectum segmentation and substantially better accuracy in tumor segmentation, achieving an accuracy similar to the inter-observer variability. Our approach could serve as the core for the fully-automated diagnosis, treatment, and follow-up of CRC using CT imaging.

References

1. Medical Segmentation Decathlon, "Challenge Leaderboard". https://decathlon-10.grand-challenge.org/evaluation/challenge/leaderboard/
2. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., van Ginneken, B., et al.: The medical segmentation decathlon. arXiv preprint arXiv:2106.05735 (2021)
3. Argilés, G., Tabernero, J., Labianca, R., Hochhauser, D., Salazar, R., Iveson, T., Laurent-Puig, P., Quirke, P., Yoshino, T., Taieb, J., et al.: Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Annals of Oncology **31**(10), 1291–1305 (2020)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. ICLR (2021)
5. Huang, Y.J., Dou, Q., Wang, Z.X., Liu, L.Z., Jin, Y., Li, C.F., Wang, L., Chen, H., Xu, R.H.: 3-d roi-aware u-net for accurate and efficient colorectal tumor segmentation. IEEE Transactions on Cybernetics **51**(11), 5397–5408 (2020)
6. 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)
7. Jian, J., Xiong, F., Xia, W., Zhang, R., Gu, J., Wu, X., Meng, X., Xin, G.: Fully convolutional networks (FCNs)-based segmentation method for colorectal tumors on T2-weighted magnetic resonance images. Australasian Physical Engineering Sciences in Medicine **41**(2), 1–9 (2018)
8. Jiang, Y., Xu, S., Fan, H., Qian, J., Luo, W., Zhen, S., Tao, Y., Sun, J., Lin, H.: ALA-Net: Adaptive Lesion-Aware Attention Network for 3D Colorectal Tumor Segmentation. IEEE Transactions on Medical Imaging **40**(12), 3627–3640 (2021)
9. Jin, D., Iyer, K.S., Chen, C., Hoffman, E.A., Saha, P.K.: A robust and efficient curve skeletonization algorithm for tree-like objects using minimum cost paths. Pattern Recognition Letters **76**, 32–40 (2016)

10. Liu, X., Guo, S., Zhang, H., He, K., Mu, S., Y, G., Li, X.: Accurate colorectal tumor segmentation for CT scans based on the label assignment generative adversarial network. Medical Physics **46**(8) (2019)
11. Milletari, F., Navab, N., Ahmadi, S.A.: V-net: Fully convolutional neural networks for volumetric medical image segmentation. In: 2016 Fourth International Conference on 3D Vision (3DV). pp. 565–571. IEEE (2016)
12. Ni, T., Xie, L., Zheng, H., Fishman, E.K., Yuille, A.L.: Elastic boundary projection for 3d medical image segmentation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2109–2118 (2019)
13. Pei, Y., Mu, L., Fu, Y., He, K., Li, H., Guo, S., Liu, X., Li, M., Zhang, H., Li, X.: Colorectal tumor segmentation of CT scans based on a convolutional neural network with an attention mechanism. IEEE Access **8**, 64131–64138 (2020)
14. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
15. Soomro, M.H., Coppotelli, M., Conforto, S., Schmid, M., Giunta, G., Del Secco, L., Neri, E., Caruso, D., Rengo, M., Laghi, A.: Automated segmentation of colorectal tumor in 3D MRI using 3D multiscale densely connected convolutional neural network. Journal of Healthcare Engineering **2019** (2019)
16. Sung, H., Ferlay, J., Siegel, R.L., Laversanne, M., Soerjomataram, I., Jemal, A., Bray, F.: Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: a cancer journal for clinicians **71**(3), 209–249 (2021)
17. Tang, Y., Yang, D., Li, W., Roth, H., Landman, B., Xu, D., Nath, V., Hatamizadeh, A.: Self-supervised pre-training of swin transformers for 3d medical image analysis. arXiv preprint arXiv:2111.14791 (2021)
18. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, L., Polosukhin, I.: Attention is all you need. arXiv preprint arXiv:1706.03762 (2017)
19. Wang, X., Girshick, R., Gupta, A., He, K.: Non-local neural networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 7794–7803 (2018)
20. Wang, Y., Wei, X., Liu, F., Chen, J., Zhou, Y., Shen, W., Fishman, E.K., Yuille, A.L.: Deep distance transform for tubular structure segmentation in CT scans. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3833–3842 (2020)
21. Wolf, A.M., Fontham, E.T., Church, T.R., Flowers, C.R., Guerra, C.E., LaMonte, S.J., Etzioni, R., McKenna, M.T., Oeffinger, K.C., Shih, Y.C.T., et al.: Colorectal cancer screening for average-risk adults: 2018 guideline update from the American Cancer Society. CA: A Cancer Journal for Clinicians **68**(4), 250–281 (2018)
22. Yao, J., Cai, J., Yang, D., Xu, D., Huang, J.: Integrating 3D geometry of organ for improving medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 318–326. Springer (2019)
23. Yushkevich, P.A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J.C., Gerig, G.: User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. Neuroimage **31**(3), 1116–1128 (2006)
24. Zheng, S., Lin, X., Zhang, W., He, B., Jia, S., Wang, P., Jiang, H., Shi, J., Jia, F.: MDCC-Net: multiscale double-channel convolution U-Net framework for colorectal tumor segmentation. Computers in Biology and Medicine **130**, 104183 (2021)