

AI-DRIVEN AUTOMATED TOOL FOR ABDOMINAL CT BODY COMPOSITION ANALYSIS IN GASTROINTESTINAL CANCER MANAGEMENT

Xinyu Nan¹ Meng He² Zifan Chen¹ Bin Dong^{3,4,5*} Lei Tang^{2*} Li Zhang^{1*}

¹Center for Data Science, Peking University, China

²Department of Radiology, Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Peking University Cancer Hospital and Institute, Beijing, China

³Beijing International Center for Mathematical Research (BICMR), Peking University, Beijing, China

⁴Center for Machine Learning Research, Peking University, Beijing, China

⁵National Biomedical Imaging Center, Peking University, Beijing, China

ABSTRACT

The incidence of gastrointestinal cancers remains significantly high, particularly in China, emphasizing the importance of accurate prognostic assessments and effective treatment strategies. Research shows a strong correlation between abdominal muscle and fat tissue composition and patient outcomes. However, existing manual methods for analyzing abdominal tissue composition are time-consuming and costly, limiting clinical research scalability. To address these challenges, we developed an AI-driven tool for automated analysis of abdominal CT scans to effectively identify and segment muscle, subcutaneous fat, and visceral fat. Our tool integrates a multi-view localization model and a high-precision 2D nnUNet-based segmentation model, demonstrating a localization accuracy of 90% and a Dice Score Coefficient of 0.967 for segmentation. Furthermore, it features an interactive interface that allows clinicians to refine the segmentation results, ensuring high-quality outcomes effectively. Our tool offers a standardized method for effectively extracting critical abdominal tissues, potentially enhancing the management and treatment for gastrointestinal cancers. The code is available at <https://github.com/NanXinyu/AI-Tool4Abdominal-Seg.git>.

Index Terms— Abdominal CT analysis, gastrointestinal cancer management, AI-driven tool, body composition analysis, multi-view localization

1. INTRODUCTION

Gastrointestinal cancers pose a significant global health challenge, with approximately 1.2 million new cases diagnosed annually worldwide, 40% of which occur in China [1]. This high incidence underscores the critical need for effective evaluation of patient prognosis and treatment efficacy. Recent studies have revealed a strong association between abdominal body composition and patient outcomes in gastrointestinal cancers [2–5]. Sarcopenia, for instance, has been linked to poor prognosis,

particularly in patients receiving targeted therapies or immune checkpoint inhibitor (ICI) treatments [3, 4]. Conversely, the “obesity paradox” suggests that increased subcutaneous and visceral fat may serve as a protective factor, especially for patients undergoing ICI therapies [5].

Given the significant impact of abdominal tissue composition on patient outcomes, accurate measurement and analysis of these tissues are crucial. However, current clinical methods primarily rely on manual annotation by specialists, a process that is time-consuming, costly, and limits large-scale clinical research [6, 7]. This limitation highlights the urgent need for an automated tool to streamline body composition assessment, enabling large-scale analysis and potentially improving patient care through more accurate and efficient evaluation.

Automated analysis of abdominal tissue composition relies on accurate region-of-interest (ROI) localization and precise tissue segmentation. Recent advancements in artificial intelligence (AI) have significantly improved both areas. Chen et al. demonstrated AI’s potential in anatomical region identification with their multi-scale context-guided model for spinal tissue localization [8]. In segmentation, Ronneberger et al. introduced the U-Net architecture [9], later expanded into nnU-Net [10], which has shown robust performance across various clinical tasks. Jothiraj et al. also proposed a deep learning framework [11] consisting of localization and semantic segmentation of polyp for early diagnosis of colorectal cancer. Leveraging these advances, we developed an automated tool for efficient abdominal tissue analysis in gastrointestinal cancer management. Our proposed tool integrates two key components: (1) A multi-view abdominal localization model, which accurately detects the starting and ending slice positions of the abdomen within a CT scan using a novel multi-view fusion module. (2) A comprehensive segmentation and analysis system, which is initialized with an automatic segmentation model driven by a 2D nnU-Net-based module, precisely delineating muscle area (MA), subcutaneous fat area (SFA), and visceral

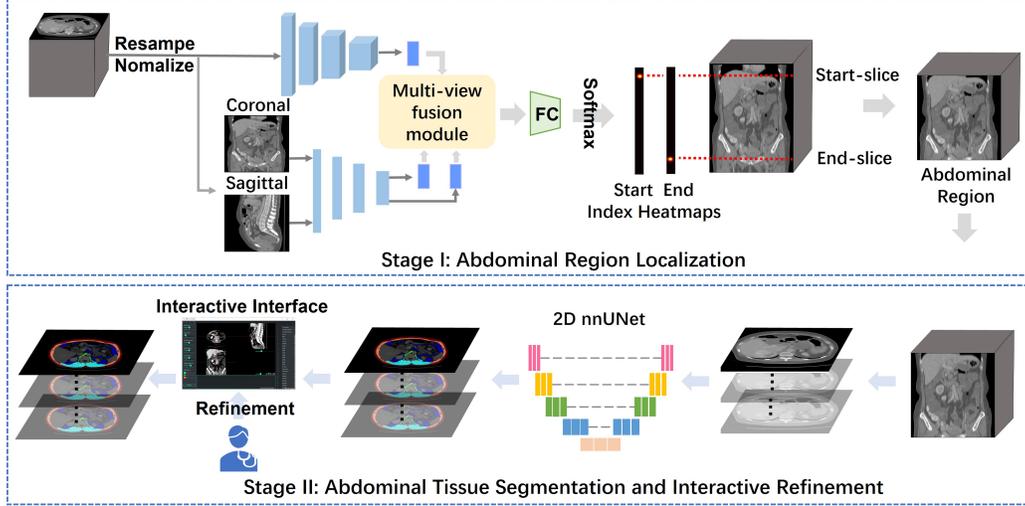


Fig. 1. An overview of our automatic annotation tool for abdominal components. (a) Stage I: A multi-view localization model consisting of a 3D ResNet18, a lightweight downsampling network and a multi-view fusion module to localize the start and end slice indexes of abdominal slices. (b) Stage II: A two-step process involving (i) a high-precision segmentation model based on 2D nnUNet architecture to segment abdominal muscle, SFA, and VFA of each slice, producing high-quality initial results, and (ii) an optional interactive interface for clinicians to further refine these segmentation results if needed.

fat area (VFA). The system then provides an interactive interface allowing clinicians to further refine the automated segmentation results with minimal manual effort, ensuring high-quality outcomes. Additionally, the tool automatically computes quantitative parameters for these abdominal tissues, providing clinicians with valuable metrics for patient assessment and treatment planning.

In summary, our main contributions include:

- Development of an AI-based tool for automated analysis of abdominal CT body composition, enabling precise localization and segmentation of three abdominal tissue types. This tool automates parameter calculation, streamlining the management of gastrointestinal cancer patients.
- Introduction of a novel multi-view fusion localization model that accurately extracts the abdominal region from 3D CT scans.
- Implementation of a high-precision 2D nnU-Net-driven segmentation model for abdominal tissues, complemented by an interactive interface. This combination provides accurate initial results while allowing clinicians to efficiently refine segmentations with minimal effort.

2. METHODOLOGY

2.1. Overview of the Proposed Automated Tool

The overview of our AI-driven tool is illustrated in Fig.1 and consists of three stages:

- Stage I: Abdominal Region localization. Firstly, the user selects a medical image covering the abdomen and feeds it

into a multi-view abdomen localization model, aiming to detect the start and end position of abdomen and extract all abdominal slices.

- Stage II: Abdominal Tissue Segmentation and Interactive Refinement. Next, all the extracted abdominal slices are individually processed through a well-trained 2D nnUNet-based segmentation model [10] to segment abdominal tissues: muscle, subcutaneous fat area (SFA), and visceral fat area (VFA). Although the segmentation results are already of high quality, users have the option to further optimize these results using an interactive interface based on PyQt5, allowing for the generation of even more refined segmentation labels if needed.

2.2. Step I: Abdominal Region Localization

The core component of Stage I is the abdominal region localization model. This model begins with preprocessing the 3D CT image through resampling, normalization, and windowing enhancement. The preprocessed image $I_{vol} \in 1 \times D \times H \times W$ is then input into a 3D ResNet-18 [12] to extract volumetric features $F_{vol} \in 512 \times D' \times H' \times W'$ ($D'=1/16D$). However, the multiple downsampling stages within the 3D ResNet can cause interference in locating the start and end slice positions, potentially leading to substantial errors in the final predictions. To address this issue, we introduced multi-view CT images to mitigate the loss of positional information during the 3D CT feature extraction process. Inspired by the work of Fahdi et al. [13], we employ a lightweight four-stage 2D downsampling network to extract features from

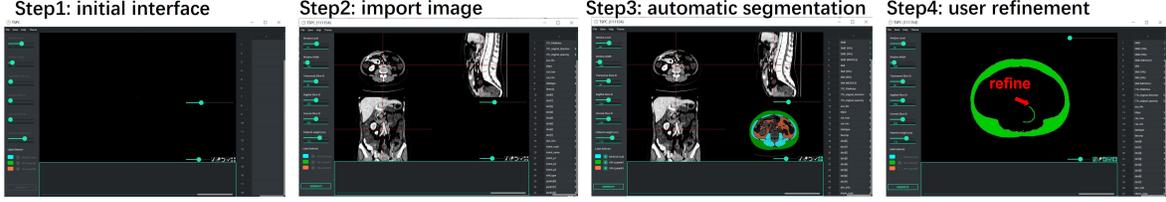


Fig. 2. Workflow of the interactive automatic annotation tool.

both coronal-view image $I_{cor} \in 1 \times D \times H$ and sagittal-view image $I_{sag} \in 1 \times D \times W$ centered on the CT volume. Each stage consists of a single convolutional layer, preserving the original positional information of the images as much as possible. To integrate features from different views, we utilize a multi-view fusion module based on cross-attention, as shown in Eq.1. This module comprises two parallel cross-attention modules [14] to generate the volume-coronal fused feature $F_{vol-cor}$ and volume-sagittal fused feature $F_{vol-sag}$. Here, F_{vol} serves as the value and key vectors for both cross-attention modules, while F_{cor} and F_{sag} act as the query vectors for their respective modules.

$$\begin{aligned}
 \tilde{F}_{vol} &= F_{vol-cor} + F_{vol-sag} + F_{vol} \\
 &= \text{SoftMax} \left(\frac{F_{cor} F_{vol}^T}{\sqrt{d_k}} \right) F_{vol} + \text{SoftMax} \left(\frac{F_{sag} F_{vol}^T}{\sqrt{d_k}} \right) F_{vol} \\
 &\quad + F_{vol}
 \end{aligned} \tag{1}$$

where, F_{cor} and F_{sag} are features extracted from coronal image and sagittal image, respectively. \tilde{F}_{vol} is the fused feature. The fused feature \tilde{F}_{vol} is then processed through a linear layer followed by a *SoftMax* activation function to produce the final 1D Gaussian heatmap predictions of the start and end slice positions of the abdominal region. Inspired by the work of Li et al., [15] the probability distribution output vectors are represented as 1D Gaussian heatmaps centered around the predicted start and end positions, respectively.

2.3. Stage II: Abdominal Tissue Segmentation and Interactive Refinement

We extracted abdominal slices based on the start and end positions predicted by the localization model in Stage I. In Stage II, we employ a two-step approach: automatic segmentation followed by interactive refinement.

First, we utilize a 2D segmentation model to perform segmentation predictions on each slice of the volume. We chose nnUNet [10] as our 2D segmentation model due to its exceptional performance in various medical imaging tasks without requiring manual intervention. nnUNet is a deep learning-based segmentation method that automatically configures itself, including pre-processing, network architecture, training, and post-processing for any new task [10]. To achieve higher segmentation accuracy and facilitate application to different

segmentation tasks, we extracted 2D slices from CT volumetric images to generate a comprehensive 2D abdominal dataset.

This dataset comprises image-mask pairs representing various abdominal components, including muscle, subcutaneous fat area (SFA), and visceral fat area (VFA). The 2D nnUNet was then trained on this dataset to predict segmentation masks for these abdominal tissues on each slice.

Following the automatic segmentation, we provide an interactive refinement interface, as illustrated in Fig.2. This interface allows users, particularly professional doctors, to review and refine the segmentation results generated by the nnUNet model. Users can view the results from transverse, coronal, and sagittal perspectives, enabling a comprehensive assessment of the segmentation quality.

3. EXPERIMENTS

3.1. Data Acquisition and pre-processing

We utilized two datasets from Peking Cancer Hospital to train and test our localization and segmentation models. The AbdLoc dataset, used for the localization model, consists of 199 CT volumetric images, each annotated with the starting and ending slice positions of the abdominal region within the volumes. For the segmentation model, we employed the AbdSeg dataset, which encompasses 1,230 image-mask pairs derived from CT images of patients diagnosed with various gastrointestinal cancers. Each slice in the dataset was extracted from abdominal region and is accompanied by a non-empty mask. For the localization model, we preprocessed the AbdLoc data by resampling each volume to a fixed size of $512 \times 64 \times 64$ before feeding it into the model. For the segmentation model, we extracted multiple 2D slices from each 3D volume in the AbdSeg dataset, resulting in a total of 85,944 2D image-mask pairs. The preprocessing of these 2D images followed the automated pipeline established by nnUNet [10].

3.2. Implementation Details

For the localization model, we implemented a 3D ResNet-18 for feature extraction from volumetric input, complemented by a four-stage lightweight downsampling network for coronal and sagittal inputs. We divided the AbdLoc dataset into training and testing sets in a 4:1 ratio. The model was trained for 250 epochs using the Adam optimizer with an initial learning rate of $1e-3$. We employed Kullback-Leibler divergence as

Rep.	start				end			
	Ave. (mm)	Max (mm)	$\leq 5mm$	$\leq 10mm$	Ave. (mm)	Max (mm)	$\leq 5mm$	$\leq 10mm$
Gaussian ($-m$)	3.08	21.97	90.0%	95.0%	2.60	12.31	90.0%	92.5%
Gaussian ($-o$)	3.19	27.25	85.0%	92.5%	3.53	34.28	85.0%	90.0%
0-1 ($-m$)	5.20	17.578	57.5%	82.5%	5.25	26.695	67.5%	82.5%
0-1 ($-o$)	5.79	21.97	47.5%	87.5%	7.62	39.06	57.5%	72.5%

Table 1. Evaluation of localization model across various metrics. **Gaussian** and **0-1** indicate that the training targets are the Gaussian heatmap and 0-1 distribution, respectively. $-m$ and $-o$ indicate introducing multi-view and one-view, respectively.

the loss function, calculated between the predicted and ground truth 1D Gaussian heatmaps. Experiments for both training and testing of the localization model were conducted on two NVIDIA Tesla A800 GPUs, with a batch size of 8 for each GPU. For the segmentation model, we trained a 2D nnUNet using 68,756 samples for training and 17,188 samples for validation. nnU-Net automatically configures the entire segmentation pipeline, encompassing all steps from preprocessing to model configuration, model training and post-processing. The experiments for the segmentation model were conducted on a single NVIDIA Tesla A800 GPU.

3.3. Evaluation Metrics

For the localization model, we employed the $L1$ distance metric as the primary evaluation criterion. Our process began with resampling all input images to a uniform size of $512 \times 64 \times 64$, recording the spacing values both before (s_{ori}) and after (s_{res}) resampling. Using these spacing values, along with the predicted ($pred$) and ground-truth (gt) slice indexes, we calculated the prediction error for the start and end positions, denoted as $L_{err}(mm)$. The $L1$ error between ground truth and prediction is calculated as follows:

$$L_{err} = \|pred \times s_{res} - gt \times s_{ori}\| \quad (2)$$

Additionally, we used the Dice Score Coefficient (DSC), the 95th percentile Hausdorff Distance (95th HD) and Intersection over Union (IoU) to evaluate segmentation quality on the testing dataset.

3.4. Localization and Segmentation Results Analysis

We evaluated the accuracy of the localization model using 40 CT volumetric samples. The assessment involved four metrics: the average $L1$ error, the maximum $L1$ error, the proportion of $L1$ error $\leq 5 mm$, and the proportion of $L1$ error $\leq 10 mm$. An $L1 \leq 5mm$ can be considered indicative of a correct localization prediction.

To validate the efficacy of representing the start and end slice positions with Gaussian heatmaps, we also conducted experiments using 0-1 distribution (where 1 indicates the start or end index position), as shown in Tab.1. The results clearly demonstrate that the Gaussian heatmap-based localization model achieved superior accuracy, outperforming the 0-1

distribution-based model across nearly all evaluated metrics. Furthermore, we performed ablation experiments presented in Tab.1, which indicate that the introduction of multi-view approaches effectively enhances localization performance, as evidenced by the reduction in both average and maximum localization errors.

For the segmentation performance, we assessed 17,188 2D CT scans using the DSC, 95th HD and IoU. Tab.2 presents the evaluation results for segmenting muscle, SFA, and VFA. The average DSC of the three tissue types exceeds 0.95, indicating highly reliable segmentation predictions with only slight discrepancies between the predicted masks and the ground truth masks. The average 95th HD of approximately $2mm$ further confirms the high precision of our segmentation results. The average IoU of the three tissue types also exceeds 90%, exhibiting highly reliable segmentation predictions as well.

	muscle	SFA	VFA	Ave.
DSC	0.979	0.979	0.944	0.967
95 HD	1.179	1.427	3.612	2.073
IoU	95.95%	94.98%	88.84%	93.26%

Table 2. Evaluation of the segmentation model for muscle, SFA and VFA according to DSC, 95 HD and IoU.

4. CONCLUSION

This study presents a novel AI-driven tool for automated analysis of abdominal CT body composition, specifically designed to assist in the management of gastrointestinal cancer patients. By integrating a multi-view abdominal localization model and a high-precision segmentation model based on a 2D nnU-Net, the proposed tool offers promising improvements in efficiency and accuracy for abdominal tissue analysis. Moreover, the interactive interface enables clinicians to refine segmentation results with minimal manual effort, further enhancing the quality of analysis. This combination of automation, precision, and collaborative features has the potential to standardize abdominal tissue measurement, reduce clinician workload, and ultimately improve the diagnosis and treatment of gastrointestinal cancer patients. At present, this work only focuses on the gastric cancer and localization and segmentation of abdominal cavity, and future work can be extended to various types of cancer and organ locations.

5. COMPLIANCE WITH ETHICAL STANDARDS

This study was approved by the Institutional Review Board of Beijing Cancer Hospital(2024KT74), and was conducted according to the tenets of the Declaration of Helsinki. Written informed consent was revoked for this retrospective study by IRB of Beijing Cancer Hospital.

6. ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China(12090022 to B.D., 11831002 to B.D., 81801778 to L.Z.), Clinical Medicine Plus X-Young Scholars Project of Peking University (PKU2023LCXQ041 to L.Z.).

7. REFERENCES

- [1] S. S. Joshi and B. D. Badgwell, "Current treatment and recent progress in gastric cancer," *CA: a cancer journal for clinicians*, vol. 71, no. 3, pp. 264–279, 2021. 1
- [2] M. He, Z. Chen, L. Zhang, X. Gao, X. Chong, H. Li, L. Shen, J. Ji, X. Zhang, B. Dong, Z. Li, and T. Lei, "Associations of subcutaneous fat area and systemic immune-inflammation index with survival in patients with advanced gastric cancer receiving dual pd-1 and her2 blockade," *Journal for Immunotherapy of Cancer*, vol. 11, 2023. 1
- [3] Y.-Y. Kim, J. Lee, W. K. Jeong, S. T. Kim, J.-H. Kim, J. Y. Hong, W. K. Kang, K.-M. Kim, I. Sohn, and D. Choi, "Prognostic significance of sarcopenia in microsatellite-stable gastric cancer patients treated with programmed death-1 inhibitors," *Gastric Cancer*, vol. 24, pp. 457–466, 2021. 1
- [4] B.-B. Chen, P.-C. Liang, T. T.-F. Shih, T.-H. Liu, Y.-C. Shen, L.-C. Lu, Z.-Z. Lin, C. Hsu, C.-H. Hsu, A.-L. Cheng, *et al.*, "Sarcopenia and myosteotosis are associated with survival in patients receiving immunotherapy for advanced hepatocellular carcinoma," *European Radiology*, vol. 33, no. 1, pp. 512–522, 2023. 1
- [5] H. Xu, D. Cao, A. He, and W. Ge, "The prognostic role of obesity is independent of sex in cancer patients treated with immune checkpoint inhibitors: A pooled analysis of 4090 cancer patients," *International Immunopharmacology*, vol. 74, p. 105745, 2019. 1
- [6] S. Koitka, L. Kroll, E. Malamutmann, A. Oezcelik, and F. Nensa, "Fully automated body composition analysis in routine ct imaging using 3d semantic segmentation convolutional neural networks," *European radiology*, vol. 31, pp. 1795–1804, 2021. 1
- [7] H. J. Park, Y. Shin, J. Park, H. Kim, I. S. Lee, D.-W. Seo, J. Huh, T. Y. Lee, T. Park, J. Lee, *et al.*, "Development and validation of a deep learning system for segmentation of abdominal muscle and fat on computed tomography," *Korean journal of radiology*, vol. 21, no. 1, pp. 88–100, 2020. 1
- [8] Z. Chen, J. Zhao, H. Yu, Y. Zhang, and L. Zhang, "Multi-scale context-guided lumbar spine disease identification with coarse-to-fine localization and classification," in *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, pp. 1–5, IEEE, 2022. 1
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical image computing and computer-assisted intervention—MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18*, pp. 234–241, Springer, 2015. 1
- [10] F. Isensee, P. F. Jaeger, S. A. Kohl, J. Petersen, and K. H. Maier-Hein, "nnu-net: a self-configuring method for deep learning-based biomedical image segmentation," *Nature methods*, vol. 18, no. 2, pp. 203–211, 2021. 1, 2.1, 2.3, 3.1
- [11] J. S and J. A. Kandaswami, "Localization and semantic segmentation of polyp in an effort of early diagnosis of colorectal cancer from wireless capsule endoscopy images," in *2022 Seventh International Conference on Parallel, Distributed and Grid Computing (PDGC)*, pp. 749–754, 2022. 1
- [12] K. Hara, H. Kataoka, and Y. Satoh, "Can spatiotemporal 3d cnns retrace the history of 2d cnns and imagenet?," in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6546–6555, 2018. 2.2
- [13] F. Kanavati, S. Islam, E. O. Aboagye, and A. G. Rockall, "Automatic 13 slice detection in 3d ct images using fully-convolutional networks," *ArXiv*, vol. abs/1811.09244, 2018. 2.2
- [14] H. Lin, X. Cheng, X. Wu, F. Yang, D. Shen, Z. Wang, Q. Song, and W. Yuan, "Cat: Cross attention in vision transformer," *2022 IEEE International Conference on Multimedia and Expo (ICME)*, pp. 1–6, 2021. 2.2
- [15] Y. Li, S. Yang, P. Liu, S. Zhang, Y. Wang, Z. Wang, W. Yang, and S.-T. Xia, "Simcc: A simple coordinate classification perspective for human pose estimation," in *Computer Vision – ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part VI*, (Berlin, Heidelberg), p. 89–106, Springer-Verlag, 2022. 2.2