ScribbleDose: Scribble-Guided Dose Prediction in Radiotherapy

Zhenxi Zhang[1⋆], Yitao Zhuang[1⋆], Yao Pu[1⋆], Peixin Yu[1⋆], Zirong Li[3], Yan Xia[3], Hui Li[1], Bin Li[4], Fuchen Zheng[5], and Ge Ren[1,2✉]

[1] Department of Health Technology and Informatics, The Hong Kong Polytechnic University, Hong Kong SAR
[2] The Hong Kong Polytechnic University Shenzhen Research Institute, The Hong Kong Polytechnic University, China
[3] Department of Orthodontics and Orofacial Orthopedics, Friedrich-Alexander-University Erlangen-Nuremberg, Germany
[4] Institute of Scientific Instrumentation, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, China
[5] Department of Computer and Information Science, University of Macau, Macau SAR

Abstract. Anatomical structure masks are widely adopted in radiotherapy dose prediction, as they provide explicit geometric constraints for structure-dose modeling. However, conventional manual delineation of these masks requires precise annotation of radiotherapy-relevant anatomical boundaries, which is time-consuming and labor-intensive. To reduce the dependence on dense structural masks, we propose ScribbleDose, a scribble-guided dose prediction framework that learns dose distributions from sparsely annotated anatomical structures. Specifically, we design a Scribble Completion Module (SCM) to reconstruct dense anatomical priors by propagating sparse scribble labels to semantically correlated voxels. During label propagation, a supervoxel-based regularization is introduced to preserve boundary consistency and improve anatomical plausibility. Furthermore, we propose a Structure-Guided Dose Generation Module (SGDGM) to incorporate the reconstructed structural priors into dose prediction through mask-guided attention. The completed masks derived from scribbles serve as intermediate structural guidance, forming an end-to-end scribble-mask-dose learning pipeline under sparse annotation. Experiments on the GDP-HMM dataset demonstrate that ScribbleDose achieves competitive dose prediction performance using only scribble-level structural annotations. The source code and reannotated scribble annotations are publicly available at https://github.com/iCherishxixixi/ScribbleDose.

Keywords: Radiotherapy Dose Prediction · Scribble Annotation · Structure Guided Modeling.

⋆ Equal contribution.
✉ Corresponding author. E-mail: gary-ge.ren@polyu.edu.hk.

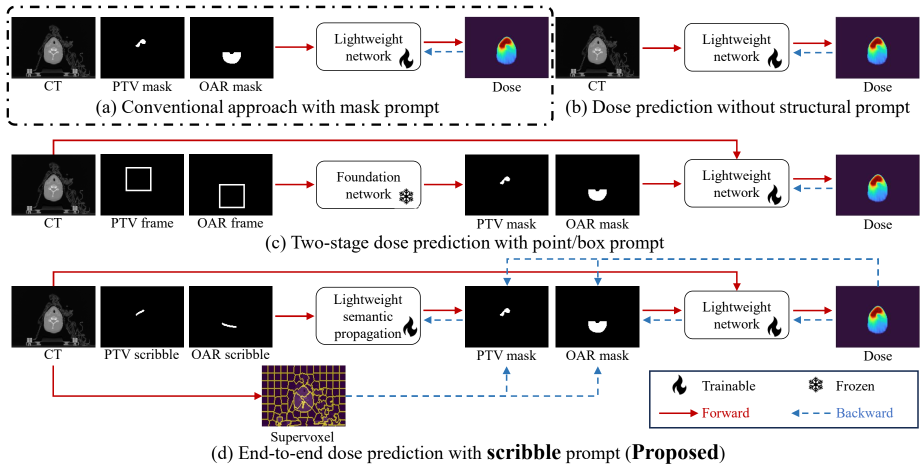

Fig. 1. Comparison of existing dose prediction paradigms and the proposed scribble-guided framework. (a) represents the conventional mask-based paradigm that relies on densely delineated structural masks. (b) and (c) aim to reduce annotation burden through CT-only modeling and two-stage foundation-model-based pipelines with box prompts, respectively. (d) presents the proposed unified scribble-guided framework.

1 Introduction

Radiotherapy planning plays a central role in modern cancer treatment, as treatment outcomes largely depend on the delivered dose distribution [4]. A high-quality plan requires sufficient coverage of the Planning Target Volume (PTV) while minimizing irradiation to surrounding Organs-at-Risk (OAR) [11]. Traditional treatment planning relies on manual iterative optimization guided by expert-defined objectives, which is time-consuming and may lead to inter-planner variability [21]. Recently, deep learning-based approaches built upon CNN [20,12], Transformers [5,18], Mamba [27], and diffusion model [25] have been widely adopted for automated dose prediction. These models can substantially shorten planning time while achieving competitive prediction accuracy, making deep learning-based dose prediction an important direction for automated radiotherapy planning [13].

Most deep learning-based dose prediction models rely on fully delineated anatomical masks as explicit geometric priors for structure-dose modeling, as illustrated in Fig. 1(a). However, obtaining high-quality masks requires precise manual delineation, which is labor-intensive and highly dependent on clinical expertise. To alleviate this burden, Willems et al. first proposed predicting dose distributions directly from CT images, as shown in Fig. 1(b), without requiring explicit PTV or OAR masks [19]. Subsequently, Jiao et al. introduced a CT-dose framework that incorporates supervoxel-based representations to enrich anatomical information while still avoiding manual contour inputs [6]. While such mask-free approaches alleviate the need for manual delineation, the absence of explicit structural guidance may compromise the modeling of fine-grained geometric constraints that are critical for clinically plausible dose allocation. To incorporate explicit structural constraints while reducing manual contouring, subsequent studies have adopted a two-stage framework based on foundation models [23,24]. In this framework, foundation models are first employed to extract dense PTV and OAR masks from CT images, often guided by sparse prompts such as points and boxes, as illustrated in Fig. 1(c). The generated masks are subsequently used as inputs for downstream dose prediction. Although this approach alleviates manual contouring during deployment, the two-stage design optimizes anatomical structure extraction and dose prediction separately, which may limit their joint adaptation for structure-aware dose modeling.

To address these limitations, we explore sparse scribbles as a lightweight alternative to fully delineated anatomical masks. Compared with dense mask annotations, scribbles substantially reduce annotation effort by avoiding precise boundary delineation of PTVs and OARs while preserving essential structural cues for dose modeling [7,22]. Recent scribble-supervised medical image segmentation studies have demonstrated the potential of sparse annotations for dense structure reconstruction through knowledge distillation, contrastive regularization, and foreground-background feature discrimination [14,8]. Different from these segmentation-oriented methods, our goal is not to develop a standalone scribble-supervised segmentation model, but to exploit scribbles as lightweight structural guidance for radiotherapy dose prediction. To this end, we design a scribble-guided dose prediction framework (Fig. 1(d)) that completes sparse scribble annotations into dense anatomical priors and jointly integrates them into dose modeling. Specifically, the Scribble Completion Module (SCM) propagates scribble labels to semantically correlated voxels, while supervoxel-derived boundary priors regularize the propagated masks toward anatomical boundaries. The completed masks are then incorporated by the Structure-Guided Dose Generation Module (SGDGM) through mask-guided attention, enabling continuous scribble-mask-dose learning. During training, dose prediction losses are further propagated back to the scribble completion stage, allowing structural completion and dose generation to be optimized jointly. Experiments on the benchmark dataset validate the feasibility of accurate dose prediction under sparse structural annotation. Our contributions are summarized as follows:

1. We reannotate and publicly release scribble-level structural labels on the GDP-HMM dataset, providing a sparse-annotation benchmark for scribble-guided radiotherapy dose prediction.
2. We propose a unified scribble-guided dose prediction framework that reduces the dependence on fully delineated structural masks by completing sparse scribbles into dense anatomical priors.
3. We introduce a Scribble Completion Module with supervoxel boundary regularization and a Structure-Guided Dose Generation Module with mask-guided attention, enabling continuous scribble-mask-dose learning for structure-aware dose prediction.

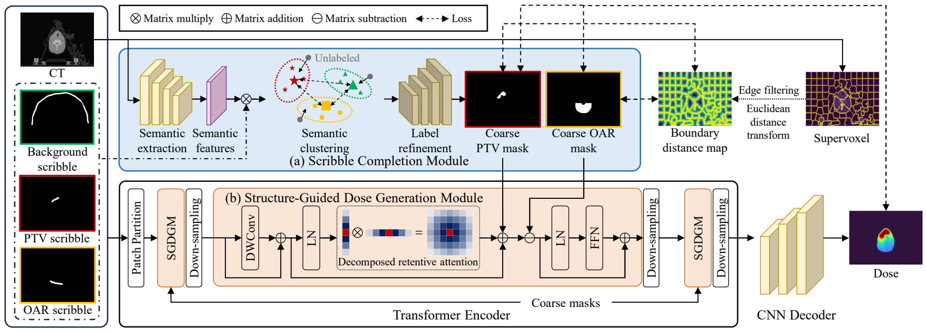

Fig. 2. Overview of the proposed scribble-guided dose prediction framework. The framework includes two main components: (a) the Scribble Completion Module (SCM) and (b) the Structure-Guided Dose Generation Module (SGDGM).

2 Method

In this section, we present the proposed scribble-guided dose prediction framework (Fig. 2), which reformulates structure-aware dose modeling under sparse structural supervision. Rather than relying on fully delineated masks, the proposed framework establishes a continuous scribble-mask-dose learning pipeline that incorporates geometric constraints under sparse annotation and jointly optimizes structural reconstruction and dose generation. The framework is built around two principal modules. The Scribble Completion Module (SCM, Fig. 2(a)) reconstructs dense structural masks from sparse scribbles under supervoxel-based geometric constraints. The Structure-Guided Dose Generation Module (SGDGM, Fig. 2(b)) then integrates the reconstructed masks into dose modeling via mask-aware attention. This design allows structural cues and dose supervision to mutually reinforce each other during training.

2.1 Scribble Completion Module

The Scribble Completion Module aims to reconstruct dense structural masks from sparse scribble annotations. To achieve this, SCM consists of two complementary components: (1) a scribble-guided semantic clustering strategy to propagate sparse labels in feature space, and (2) a supervoxel-based boundary constraint to enforce anatomical consistency in the spatial domain.

Scribble-Guided Semantic Clustering aims to propagate sparse scribble annotations to unlabeled voxels in feature space and generate dense structural masks. Inspired by prototype-based feature clustering in semantic segmentation [17] and semantic feature aggregation in medical structure segmentation [26], we perform centroid-based semantic clustering by computing class-wise feature centroids from scribble-labeled voxels and propagating sparse labels in the feature space. Specifically, given a CT volume $X \in \mathbb{R}^{D\times H\times W}$, we extract volumetric features $F \in \mathbb{R}^{C\times D\times H\times W}$ utilizing a lightweight 3D encoder, where C is the feature dimension. For each semantic class $k \in \{\mathrm{PTV}, \mathrm{OARs}, \mathrm{Background}\}$, we compute a class-wise feature centroid from scribble-labeled voxels. Let Ω_k denote the voxel set annotated by scribbles of class k, and N_k its cardinality. The centroid is defined as:

$$c_k = \frac{1}{N_k} \sum_{i\in\Omega_k} \frac{F_i}{\|F_i\|}, \tag{1}$$

where feature vectors are ℓ_2-normalized to stabilize similarity computation. We then measure the similarity between each voxel feature and the class centroids using cosine similarity:

$$s_{k,i} = \frac{c_k^\top F_i}{\tau}, \tag{2}$$

where τ is a temperature parameter. The similarity scores $s_{k,i}$ are computed for unlabeled voxels and serve as coarse class logits. These logits are subsequently refined by a lightweight decoder to obtain dense mask predictions.

To encourage intra-class compactness and inter-class separability in feature space, we introduce two regularization terms. The compactness loss reduces feature variance within scribble regions:

$$\mathcal{L}_{\mathrm{compact}} = \sum_k \frac{1}{N_k} \sum_{i\in\Omega_k} \|F_i - c_k\|^2, \tag{3}$$

while the separation loss penalizes similarity between different class centroids:

$$\mathcal{L}_{\mathrm{sep}} = \sum_{k\neq j} \exp\bigl(-\|c_k - c_j\|^2\bigr)\,. \tag{4}$$

Through this centroid-driven semantic propagation, sparse scribble annotations are progressively expanded into coarse structural masks in feature space.

Supervoxel-Based Boundary Constraint introduces geometry-aware spatial regularization into mask reconstruction. In particular, supervoxels partition the CT volume into locally homogeneous intensity regions, preserving intrinsic anatomical boundaries. By leveraging supervoxel-derived boundary cues, the reconstructed masks are encouraged to align with anatomically consistent spatial structures.

Direct alignment with hard boundary maps treats all non-boundary deviations uniformly, without distinguishing between small and large spatial discrepancies. To provide distance-sensitive geometric supervision, we construct a distance-based boundary representation derived from supervoxel boundaries. Specifically, we apply Euclidean Distance Transform (EDT) to the supervoxel boundary map to compute a distance field D_{sv}, where each voxel encodes its minimal distance to the nearest supervoxel boundary. The resulting distance map satisfies the Eikonal property $|\nabla D_{\mathrm{sv}}| = 1$ almost everywhere, yielding a spatially continuous geometric supervision signal. In practice, Gaussian smoothing is employed as an efficient approximation of distance-based weighting to obtain a differentiable soft boundary representation $\tilde{D}_{\mathrm{sv}}$. The boundary constraint is formulated as:

$$\mathcal{L}_{\mathrm{boundary}} = \left\| B_{\mathrm{pred}} - \tilde{D}_{\mathrm{sv}} \right\| \tag{5}$$

where B_{pred} denotes the predicted boundary map. This distance-based regularization stabilizes boundary refinement and promotes anatomically plausible mask reconstruction beyond scribble-labeled regions.

2.2 Structure-Guided Dose Generation Module

The Structure-Guided Dose Generation Module aims to explicitly couple anatomical structures with dose modeling while preserving spatial continuity of dose distributions. To model such spatial coherence, SGDGM is implemented using Retentive Transformers (RMT) [2]. Specifically, RMT adopts decomposed retentive attention with distance-aware decay along spatial directions. This distance-decay mechanism encourages each patch to focus on neighboring regions while attenuating distant interactions, introducing an inductive bias aligned with the smoothness of dose distributions.

Building upon the RMT backbone, we introduce a structure-guided attention mechanism based on the reconstructed coarse masks. Specifically, let M_{PTV} and M_{OAR} denote the coarse masks of PTV and OAR regions, respectively. The structure-guided modulation map G is defined as:

$$G = 2 \cdot \sigma(1 + \alpha M_{\mathrm{PTV}} - \beta M_{\mathrm{OAR}}), \tag{6}$$

where α and β are learnable positive coefficients and $\sigma(\cdot)$ denotes the sigmoid function. The modulation map is applied multiplicatively to the multi-scale retentive attention maps:

$$\mathrm{Attn}'_l = \mathrm{Attn}_l \odot G_l, \tag{7}$$

where G_l is resized to match the spatial resolution of Attn_l. This modulation enhances attention responses within PTV regions while suppressing responses in OAR regions, thereby establishing explicit structure-dose coupling.

Moreover, to further reinforce structural consistency in dose allocation, we introduce a ranking-based regularization term:

$$\mathcal{L}_{\mathrm{rank}} = \max\left(0, \bar{D}_{\mathrm{OAR}} - \bar{D}_{\mathrm{PTV}}\right), \tag{8}$$

where $\bar{D}_{\mathrm{PTV}}$ and $\bar{D}_{\mathrm{OAR}}$ denote the average predicted dose within PTV and OAR regions, respectively. This auxiliary constraint provides a coarse global preference for higher average dose in target volumes than in OAR regions, serving as a supplementary regularization rather than a replacement for clinically specified dose-volume constraints.

Table 1. Performance comparison under different structure prompt types on the GDP-HMM dataset. * denotes a two-stage framework where segmentation is first performed to generate dense structural masks, followed by MedNeXt for dose prediction. Values are reported as mean (standard deviation).

Method	Prompt	Dose Score ↓	DVH Score ↓	HI ↓	CI ↓	D_{95} ↓
3D UNet [19]	None	3.275(1.295)	5.778(2.070)	0.082(0.049)	0.340(0.185)	5.136(3.053)
RMT [2]	None	3.516(1.368)	7.085(3.215)	0.157(0.158)	0.526(0.223)	6.709(7.324)
MedNeXt [15]	None	3.152(1.651)	6.425(2.294)	0.100(0.076)	0.393(0.185)	5.142(3.662)
MedSAM2*[10]	Box	2.670(1.356)	5.382(2.099)	0.157(0.064)	0.587(0.236)	7.537(4.521)
SAMMed3D*[16]	Point	2.732(1.275)	5.219(1.784)	0.143(0.061)	0.576(0.248)	6.558(3.957)
C3D [9]	Mask	1.993(1.139)	2.642(1.766)	**0.040(0.091)**	0.557(0.285)	2.398(1.980)
C3D [9]	Scribble	2.348(1.232)	6.716(2.280)	0.274(0.186)	0.361(0.253)	4.332(3.994)
RMT [2]	Scribble	2.557(1.263)	4.600(2.504)	0.109(0.073)	0.515(0.288)	6.138(4.914)
MedNeXt [15]	Scribble	2.268(1.245)	4.082(2.331)	0.100(0.076)	0.434(0.247)	5.675(4.517)
ScribbleDose	Scribble	**1.974(1.149)**	**2.471(1.641)**	<u>0.046(0.043)</u>	**0.171(0.188)**	**2.350(2.137)**

3 Experiments

3.1 Dataset and Implementation

Experiments were conducted on the dataset released by the GDP-HMM 2025 challenge [3], which contains 3231 radiotherapy plans. Following the official partition, 2627 cases were used for training, 148 for validation, and 356 for testing. For each case, the dataset provides CT volumes together with PTV and OAR masks, as well as the corresponding clinical dose distributions. Additionally, scribbles were generated slice-wise from manual masks following [1], ensuring they remain strictly inside the original contours.

Model training was conducted using PyTorch 2.6 on four NVIDIA A6000 GPUs. Optimization was performed with Adam for 300 epochs. The learning rate was initialized at 1×10^{-4} and gradually annealed to 1×10^{-6}.

3.2 Experimental Results

We evaluate different structure prompt types using Dose score, Dose Volume Histogram (DVH) score, Homogeneity Index (HI), Conformity Index (CI), and D_{95}. Here, D_{95} represents the dose received by 95% of the PTV volume. As shown in Table 1, introducing structural prompts consistently improves dose prediction compared with CT-only modeling. For the two-stage foundation-model-based pipelines, MedSAM2 and SAMMed3D reduce the Dose Score of MedNeXt without structural prompts from 3.152 to 2.670 and 2.732, respectively, indicating that converting sparse prompts into dense masks provides useful geometric constraints for dose estimation. Among these sparse-prompt settings, the proposed ScribbleDose achieves the lowest Dose and DVH scores. Notably, ScribbleDose achieves performance comparable to the mask-based C3D baseline, suggesting that sparse scribble guidance can recover effective structural cues for dose prediction while substantially reducing annotation requirements.

Table 2. Ablation study on the Semantic Guidance (SG) and Morphological Guidance (MG). Values are reported as mean (standard deviation).

SG	MG	Dose Score ↓	DVH Score ↓	HI ↓	CI ↓	D_{95} ↓
×	×	2.557(1.263)	4.600(2.504)	0.109(0.073)	0.515(0.288)	6.138(4.914)
✓	×	2.209(1.202)	3.329(2.049)	0.078(0.057)	0.268(0.228)	3.913(3.476)
×	✓	2.287(1.299)	4.774(1.626)	0.051(0.041)	0.487(0.209)	5.212(2.300)
✓	✓	**1.974(1.149)**	**2.471(1.641)**	**0.046(0.043)**	**0.171(0.188)**	**2.350(2.137)**

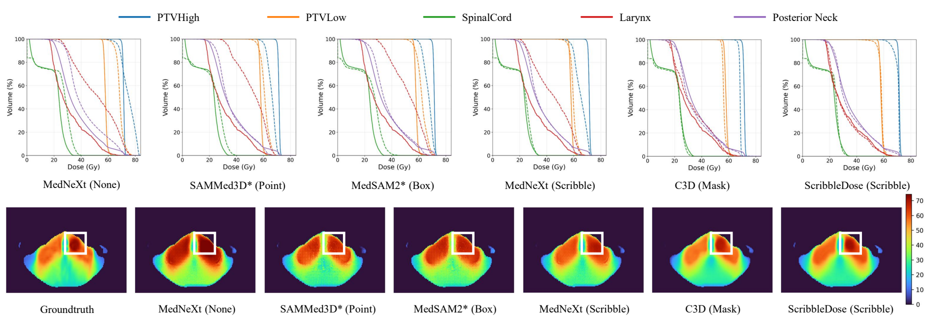

Fig. 3. Qualitative comparison of predicted dose distributions. **Top:** DVH of PTVs and OARs. Solid lines represent the reference dose distributions, while dashed lines denote the predicted dose distributions. **Bottom:** The axial dose maps.

Ablation Study: Table 2 evaluates the contributions of Semantic Guidance (SG) and Morphological Guidance (MG). SG corresponds to the semantic propagation mechanism in the Scribble Completion Module, while MG introduces supervoxel-based boundary constraints for anatomical plausibility. Introducing SG substantially reduces both Dose and DVH scores, demonstrating that semantic propagation effectively recovers structure-dose relationships from sparse scribbles. Adding MG further improves geometric consistency and leads to better overall performance. The combination of SG and MG achieves the best results across most evaluation metrics, confirming that semantic and morphological guidance are complementary for accurate scribble-guided dose prediction.

Visualization: Qualitative results in Fig. 3 further demonstrate the effectiveness of the proposed scribble-guided framework. The DVH curves produced by ScribbleDose are closer to the reference distributions for both PTVs and OARs, indicating improved dose-volume consistency. The axial dose maps show more concentrated and anatomically consistent high-dose regions, as highlighted by the white box. These observations suggest that sparse scribble annotations, when completed into dense anatomical priors, can provide effective structural guidance for dose prediction.

4 Conclusion

In this work, we propose ScribbleDose, a scribble-guided dose prediction framework for radiotherapy under sparse structural annotation. By reconstructing dense anatomical priors from sparse scribbles through SCM and incorporating them into dose modeling via SGDGM, the proposed method establishes an end-to-end scribble-mask-dose learning pipeline. Experiments on the GDP-HMM dataset demonstrate that ScribbleDose can reduce the dependence on fully delineated structural masks while maintaining competitive dose prediction performance. As an initial step toward scribble-guided dose prediction, this study uses scribbles simulated from manual masks. Future work will further investigate real clinical scribbles with potential positional shifts, missing slices, and inconsistent stroke patterns. We will also explore more clinically interpretable dose objectives and improve the robustness of supervoxel-based priors under ambiguous anatomical boundaries.

Acknowledgments. This research was partly supported by the Shenzhen Medical Research Fund (A2503079), Health and Medical Research Fund (11222456, 10240147) of the Health Bureau, Guangdong Basic and Applied Basic Research Foundation (2025A15 15012926), and Shenzhen Science and Technology Program (JCYJ20230807140403007).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Boettcher, W., Hoyer, L., Unal, O., Lenssen, J.E., Schiele, B.: Scribbles for all: Benchmarking scribble supervised segmentation across datasets. In: Advances in Neural Information Processing Systems. vol. 37, pp. 46002–46024 (2024)
2. Fan, Q., Huang, H., Chen, M., Liu, H., He, R.: Rmt: Retentive networks meet vision transformers. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5641–5651 (2024)
3. Gao, R., Diallo, M., Liu, H., Magliari, A., Sackett, J., Verbakel, W., Meyers, S., Zarepisheh, M., Mcbeth, R., Arberet, S., et al.: Automating high quality rt planning at scale. arXiv e-prints pp. arXiv–2501 (2025)
4. Hernandez, V., Hansen, C.R., Widesott, L., Bäck, A., Canters, R., Fusella, M., Götstedt, J., Jurado-Bruggeman, D., Mukumoto, N., Kaplan, L.P., et al.: What is plan quality in radiotherapy? the importance of evaluating dose metrics, complexity, and robustness of treatment plans. Radiotherapy and Oncology **153**, 26–33 (2020)
5. Hu, C., Wang, H., Zhang, W., Xie, Y., Jiao, L., Cui, S.: Trdosepred: A deep learning dose prediction algorithm based on transformers for head and neck cancer radiotherapy. Journal of Applied Clinical Medical Physics **24**(7), e13942 (2023)
6. Jiao, Z., Peng, X., Wang, Y., Xiao, J., Nie, D., Wu, X., Wang, X., Zhou, J., Shen, D.: Transdose: Transformer-based radiotherapy dose prediction from ct images guided by super-pixel-level gcn classification. Medical Image Analysis **89**, 102902 (2023)

7. Lin, D., Dai, J., Jia, J., He, K., Sun, J.: Scribblesup: Scribble-supervised convolutional networks for semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3159–3167 (2016)
8. Liu, J., Tan, S.Y., Yang, X., Xu, Y., Yeo, S.Y.: Effdnet: A scribble-supervised medical image segmentation method with enhanced foreground feature discrimination. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 194–204. Springer (2025)
9. Liu, S., Zhang, J., Li, T., Yan, H., Liu, J.: A cascade 3d u-net for dose prediction in radiotherapy. Medical Physics **48**(9), 5574–5582 (2021)
10. 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)
11. Monshouwer, R., Hoffmann, A.L., Kunze-Busch, M., Bussink, J., Kaanders, J.H., Huizenga, H.: A practical approach to assess clinical planning tradeoffs in the design of individualized imrt treatment plans. Radiotherapy and Oncology **97**(3), 561–566 (2010)
12. Nguyen, D., Long, T., Jia, X., Lu, W., Gu, X., Iqbal, Z., Jiang, S.: A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning. Scientific Reports **9**(1), 1076 (2019)
13. Nijkamp, J., Knäusl, B., Aznar, M., Georg, D., Thorwarth, D., Thwaites, D., Muren, L.P., van der Heide, U.A.: Current advances in automation in radiotherapy. Radiotherapy and Oncology **205** (2025)
14. Qu, Y., Lu, T., Zhang, S., Wang, G.: Scribsd+: Scribble-supervised medical image segmentation based on simultaneous multi-scale knowledge distillation and class-wise contrastive regularization. Computerized Medical Imaging and Graphics **116**, 102416 (2024)
15. Roy, S., Koehler, G., Ulrich, C., Baumgartner, M., Petersen, J., Isensee, F., Jaeger, P.F., Maier-Hein, K.H.: Mednext: Transformer-driven scaling of convnets for medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 405–415. Springer (2023)
16. Wang, H., Guo, S., Ye, J., Deng, Z., Cheng, J., Li, T., Chen, J., Su, Y., Huang, Z., Shen, Y., et al.: Sam-med3d: A vision foundation model for general-purpose segmentation on volumetric medical images. IEEE Transactions on Neural Networks and Learning Systems (2025)
17. Wang, K., Liew, J.H., Zou, Y., Zhou, D., Feng, J.: Panet: Few-shot image semantic segmentation with prototype alignment. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9197–9206 (2019)
18. Wen, L., Xiao, J., Tan, S., Wu, X., Zhou, J., Peng, X., Wang, Y.: A transformer-embedded multi-task model for dose distribution prediction. International Journal of Neural Systems **33**(08), 2350043 (2023)
19. Willems, S., Crijns, W., Sterpin, E., Haustermans, K., Maes, F.: Feasibility of ct-only 3d dose prediction for vmat prostate plans using deep learning. In: Workshop on Artificial Intelligence in Radiation Therapy. pp. 10–17. Springer (2019)
20. Xiong, T., Ren, G., Chen, Z., Huang, Y.H., Ma, Z., Li, Z., Sheng, Y., Wu, Q.J., Cai, J.: A generalizable dose prediction model for automatic radiotherapy planning based on physics-informed priors and large-kernel convolutions. Medical Physics **53**(1), e70272 (2026)
21. Xiong, T., Zeng, G., Chen, Z., Huang, Y.H., Li, B., Zhou, D., Liu, X., Sheng, Y., Ren, G., Wu, Q.J., et al.: Automatic planning for functional lung avoidance

radiotherapy based on function-guided beam angle selection and plan optimization. Physics in Medicine & Biology **69**(15), 155007 (2024)
22. Zhang, K., Zhuang, X.: Cyclemix: A holistic strategy for medical image segmentation from scribble supervision. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11656–11665 (2022)
23. Zhang, L., Liu, Z., Zhang, L., Wu, Z., Yu, X., Holmes, J., Feng, H., Dai, H., Li, X., Li, Q.: Generalizable and promptable artificial intelligence model to augment clinical delineation in radiation oncology. Medical Physics **51**(3), 2187–2199 (2024)
24. Zhang, L., Liu, Z., Zhang, L., Wu, Z., Yu, X., Holmes, J., Feng, H., Dai, H., Li, X., Li, Q., et al.: Segment anything model (sam) for radiation oncology. arXiv preprint arXiv:2306.11730 (2023)
25. Zhang, Y., Li, C., Zhong, L., Chen, Z., Yang, W., Wang, X.: Dosediff: Distance-aware diffusion model for dose prediction in radiotherapy. IEEE Transactions on Medical Imaging **43**(10), 3621–3633 (2024)
26. Zhang, Z., Zheng, F., Iltaf, A., Han, Y., Cheng, Z., Du, Y., Li, B., Liu, T., Zhou, S.: Adaptive morph-patch transformer for aortic vessel segmentation. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 40, pp. 13015–13024 (2026)
27. Zhou, P., Peng, Q., Li, Y., Li, C.: A novel mamba-based 3d dose prediction model with channel-aware scan for nasopharyngeal carcinoma radiotherapy. International Journal of Radiation Oncology, Biology, Physics **123**(1), e134 (2025)