NoiBox: Safeguard Frontier via Reverse Adversarial Diffusion for Noisy-Box-Supervised 3D Tumor Segmentation

Kailun Ji[1,2†], Shaoqing Zhai[1†], Cheng Lian[1], Yixin Su[1], Baoru Huang[3], Feng Xiao[4], and Quan Zhou[1(✉)]

[1] Wuhan University of Technology
[2] Northwestern Polytechnical University
[3] University of Liverpool
[4] Zhongnan Hospital of Wuhan University
{330146, zhaishaoqing, chenglian, suyixin, zhouquan910}@whut.edu.cn, baoru.huang@liverpool.ac.uk, xiao.feng@whu.edu.cn

Abstract. Voxel-wise 3D tumor segmentation is clinically essential yet prohibitively expensive to annotate at scale. In clinical workflows, radiologists can instead provide a *single* coarse 3D bounding box as a fast region-of-interest cue. However, such boxes are typically *Noisy*, introducing background contamination near the inner boundary and tumor leakage beyond the box. Naive approaches that enforce hard box-as-mask constraints or use rigid pseudo-labels in self-training amplify confirmation bias and degrade boundary accuracy. To overcome this, we propose a **safeguard frontier** paradigm that learns 3D tumor segmentation from a noisy box by constructing a *soft frontier supervision* target. Specifically, we partitions 3D space into a far-field *safe background* and an *inner/outer uncertainty band* around the box. Within such band, we mine reliable foreground-core and safe-background seeds via hybrid reliability sampling, then generates a detached soft frontier target through *reverse adversarial diffusion* integrating the affinity-driven diffusion evolution with annealed prediction gate to jointly build the safeguard frontier between *outward-diffusing foreground evidence* and *inward-diffusing background evidence*. The model training combines a distance-decayed safe-background constraint with a adversarial diffusion consistency constraint. Importantly, the approach is plug-and-play with any 3D backbones. Experiments on MSD-Lung and MSD-Pancreas demonstrate that our method outperforms state-of-the-art (SOTA) box-supervised approaches by at least 2.77% and 0.77% in Dice score, respectively. It highlights the practical value of scalable single-box annotation. Codes are at https://github.com/Radioactive-jkl/Noibox.

Keywords: 3D tumor segmentation · Noisy-box supervision · Hybrid reliability sampling · Soft frontier target · Reverse adversarial diffusion

† Equal contribution;
✉Corresponding author.

1 Introduction

Accurate 3D tumor segmentation is a critical prerequisite for clinical decision-making, including diagnosis, radiotherapy planning, and longitudinal monitoring [18, 1]. Despite remarkable advances in deep learning–based medical image analysis, state-of-the-art (SOTA) 3D segmentation models still depend heavily on dense voxel-wise annotations. These labels are not only time-consuming and costly to acquire but also lack scalability across anatomies, imaging protocols, and clinical sites. [22, 3, 11, 8, 7]. In contrast, routine clinical practice often provides a far more accessible form of supervision: a coarse region-of-interest (ROI) delineated by a clinician, typically as a *single* axis-aligned 3D bounding box [23, 21, 13, 17]. Such box prompts can be annotated in seconds during reading, triage, or quality-assurance workflows, offering a pragmatic and scalable pathway toward building large and diverse clinically cohorts [14, 20].

Learning 3D segmentation from a single box remains challenging in clinical practice since radiologist-provided boxes are *loose* rather than tight [27, 15]. Such loose boxes based methods introduce two inevitable uncertainties that extend beyond standard box-supervised segmentation. First, the voxels *inside* a loose box are frequently confused with normal tissue, particularly near the box boundaries, so treating all interior voxels as foreground leads to unexpected over-segmentation [28]. Second, a small fraction of tumor voxels may lie *outside* the box due to imperfect placement, anatomical ambiguity or multi-foci lesions, enforcing hard background constraints on all exterior voxels thus induces false negatives [4]. Crucially, these uncertainties are *coupled*, since enforcing hard constraints such as assigning all interior voxels as foreground and all exterior voxels as background can induce confirmation bias during training. In this setting, initial boundary errors are reinforced during the learning loops and progressively amplified, particularly under patch-based training where limited contextual awareness hinders global consistency [24, 12].

Existing box-supervised segmentation methods typically follow one of three paradigms, all of which struggle under the uncertainty caused by noisy bounding boxes. (i) *Hard constraints* treat the box as binary supervision or tight priors, which tends to suppress true positives outside the box and over-penalize leakage [17, 26, 27]. (ii) *Pseudo-mask self-training* converts model predictions into pseudo labels, yet boundary errors near box boundaries can accumulate and become irreversible due to error propagation [2, 6, 16]. (iii) *Unidirectional propagation or smoothing* propagates foreground labels outward or background labels inward without explicit constraints, resulting in blurred boundaries and an inability to model regions where both foreground and background are plausible [9, 15]. Overall, there is a lack of an effective supervision mechanism that explicitly models the *inner/outer* uncertainty band around a noisy box and produces a *soft* training target that preserves boundary ambiguity instead of forcing premature hard constraints.

In this paper, we propose to supervise 3D tumor segmentation from a noisy bounding box using a *safeguard frontier*. It is an estimated conservative boundary between reliable background evidence and uncertain tumor regions. Rather than

interpreting the box as a proxy mask, we construct a soft target that creates a adversarial boundary within the box-induced uncertainty band. Specifically, given a 3D volume and a noisy box, we first partition 3D space into a far-field *safe background* region and a box-centered *uncertain region* that comprises inner and outer boundary bands, then explore high-confidence foreground core seeds and safe background seeds. Moreover, soft target is generated through *reverse adversarial diffusion*, where foreground evidence propagates outward from reliable core seeds and background evidence propagates inward from safe background anchors. The two evidence streams yields a detached frontier target in uncertain regions, enabling stable learning without imposing hard labels near the noisy box boundary and suppressing true positive voxels that lie outside the box. To further enhance robustness, we train the model with a distance-decayed safe-background loss and a adversarial diffusion consistency loss. Notably, during the training, the diffusion module functions exclusively as a safeguard frontier generator, iteratively refining the boundary to improve supervision quality. Experimentally, we conduct extensive experiments on MSD-Lung and MSD-Pancreas, achieving Dice score improvements of 2.77% and 0.77% over SOTA box-supervised methods.

The major contributions are as follows:

- **Noisy 3D box-induced uncertainty field partition.** We explicitly model the coupled uncertainties inherent in clinician-provided noisy boxes, *i.e.*, background contamination near the inner boundary and potential tumor leakage beyond the box, by partitioning space into a far-field safe background region and a box-centered *inner/outer* uncertainty band.
- **Safeguard frontier generation via reverse adversarial diffusion.** We design a frontier supervision target generated through reverse adversarial interaction between *outward-diffusing* foreground evidence and *inward-diffusing* background evidence, enabling soft boundary-aware learning without reliance on hard pseudo-masks.

2 Method

2.1 Problem Setup

Let $V \in \mathbb{R}^{D\times H\times W}$ denote a 3D image volume on a voxel lattice Ω, and let $v \in \Omega$ index a voxel location. Given a clinician-provided noisy bounding box B, we train a 3D segmentation network $f(\cdot;\theta_t)$ at epoch t to predict a voxel-wise tumor probability map

$$P_t(v) = f(V;\theta_t) \in [0,1]^{D\times H\times W} \tag{1}$$

where $P_t(v)$ denotes the probability at voxel v.

Fig. 1 illustrates the overview of our proposed framework. Instead of converting B to hard labels, we construct a dynamic soft supervision target within a box-induced uncertainty band. Specifically, it consists of two key modules: (i) Box-induced uncertainty field partition (BUFP) to separate reliable foreground/background seed regions from the box-centered inner/outer uncertainty

bands, (ii) Reverse adversarial diffusion (RAD) to generate a soft frontier target π_t, where foreground and background evidence evolve via adversarial diffusion.

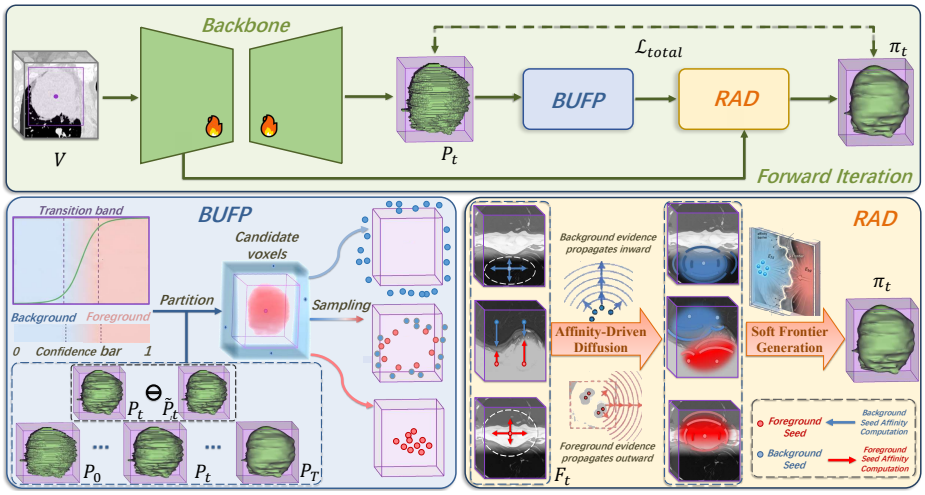

Fig. 1. The pipeline of the proposed method. It consists of two key modules: BUFP and RAD. BUFP constructs a box-induced spatial partition that separates reliable foreground/background seeds from the box-centered inner/outer uncertainty bands. RAD yields a soft frontier target by adversarial propagating foreground evidence outward and background evidence inward.

2.2 Box-Induced Uncertainty Field Partition

We construct a continuous uncertainty field parameterized by box geometry, which softens supervision near the margin and enforces strong constraints only in far-field background region and identified tumor area.

We first model the 3D space as a continuous uncertainty field derived from the geometry of the bounding box. Let $d_B(v) \in \mathbb{R}$ be the signed shortest distance from voxel v to the *surface* of box B: $d_B(v) > 0$ if $v \in B$ and $d_B(v) < 0$ if $v \notin B$. We map $d_B(v)$ to a soft box-induced spatial prior $C_{sp}(v) = \text{sigmoid}(d_B(v)/\tau)$ where τ controls the transition band width. Intuitively, $C_{sp}(v)$ is high inside/near the box and decays smoothly outside.

Separately, we define a consistency-based reliability prior $C_{se}(v) = 1 - \left|P_t(v) - \tilde{P}_t(v)\right|$, where $\tilde{P}_t(v)$ is the prediction from an EMA teacher [25]. To ensure stable propagation, we employ hybrid reliability sampling to identify high-confidence foreground S_{fg} and background seed voxels S_{bg}. We assess voxel reliability $C(v)$ by fusing $C_{sp}(v)$ with $C_{se}(v)$:

$$C(v) = C_{sp}(v) + C_{se}(v) \tag{2}$$

Foreground seeds. We extract sparse foreground seeds by greedy non-maximum suppression (NMS) with radius r, applied to gated foreground candidates using reliability score $C(v)$:

$$S_{fg} \in \arg\max_{S \subseteq \Omega} \sum_{v \in S} C(v) \quad \text{s.t.} \quad \|v - u\|_2 > r \;\; \forall v \neq u, \; v, u \in S \tag{3}$$

Background seeds. We uniformly sample background seeds from the gated background candidates.

2.3 Reverse Adversarial Diffusion

To generate the soft frontier, we propose the reverse adversarial diffusion, which is an adversarial boundary refinement process in which foreground evidence diffuses outward from the box interior while background evidence propagates inward from the far field. This is realized as a seed-constrained Dirichlet minimization on an affinity-weighted voxel matrix.

Affinity-Driven Diffusion Evolution. We preserve anatomical boundaries by guiding diffusion through an affinity-weighted operator, where affinities are computed from encoder feature embeddings $\hat{F}_t(v) \in \mathbb{R}^C$. we construct an undirected voxel graph on Ω with edges connecting local neighbors, and define the symmetric affinity as:

$$w_{vu} = \exp\left(-\frac{\|\hat{F}_t(v) - \hat{F}_t(u)\|_2^2}{\sigma^2}\right) \tag{4}$$

where σ controls affinity sensitivity. When $\hat{F}_t(v)$ and $\hat{F}_t(u)$ are highly similar, enabling free diffusion within homogeneous regions, when dissimilarity increases at anatomical boundaries, suppressing diffusion and preventing label leakage.

Soft Frontier Target. We obtain the safeguard frontier target $\pi_t : \Omega \to [0, 1]$ by solving a seeded Dirichlet minimization:

$$\pi_t = \arg\min_{\pi} \; \frac{1}{2} \sum_{(v,u) \in E} w_{vu} \big(\pi(v) - \pi(u)\big)^2 \quad \text{s.t.} \quad \pi(v) = S(v) \;\; \forall v \in S \tag{5}$$

where $S = S_{\text{fg}} \cup S_{\text{bg}}$, with $S(v) = 1$ for foreground seeds and $S(v) = 0$ for background seeds. We treat π_t as a detached target during segmentation network optimization. Moreover, voxels connected to foreground seeds tend to take values close to 1, while those strongly connected to background seeds tend to approach 0. Consequently, regions where $\pi_t(v) \approx 0.5$ correspond to high-ambiguity zones where foreground and background influences are balanced.

Table 1. Comparison with SOTA methods on MSD-Lung and MSD-Pancreas under noisy-box supervision. The optimal values are shown in bold.

Method	MSD-Lung				MSD-Pancreas			
	SwinUNETR		UNETR		SwinUNETR		UNETR	
	Dice ↑	HD95 ↓	Dice ↑	HD95 ↓	Dice ↑	HD95 ↓	Dice ↑	HD95 ↓
Fully Supervised	71.69	10.85	60.27	11.21	77.24	10.12	66.74	10.32
Box2Mask [17]	62.57	13.80	51.07	15.47	73.55	15.10	62.05	14.11
BoxInst [26]	64.40	11.93	52.90	13.39	74.04	12.73	64.54	11.89
BBTP [27]	62.98	12.52	51.48	14.69	74.27	12.49	64.77	11.67
SANPNet [2]	62.24	13.02	50.74	16.76	73.33	12.81	63.83	12.96
MonoBox [9]	64.96	12.14	54.35	13.79	73.42	12.49	64.92	11.67
ReliableMD [6]	64.77	12.45	53.27	14.42	74.89	12.38	65.39	11.26
LooBox [15]	65.19	11.46	54.69	13.44	75.18	12.27	65.68	10.46
GeoCoBox [16]	65.25	11.57	53.14	13.46	75.06	12.25	64.82	11.08
Ours	**67.06**	**10.11**	**55.31**	**12.20**	**75.64**	**10.89**	**66.22**	**10.37**

2.4 Objective Optimization with One-ramp Hardening

We optimize the model using a unified objective that jointly enforces a safe-background loss $\mathcal{L}_{bg}$ and a soft frontier consistency loss $\mathcal{L}_{sf}$. To mitigate confirmation bias in early training, we adopt a one-ramp hardening strategy to gradually increase λ_t:

$$\mathcal{L}_{total} = \mathcal{L}_{bg} + \lambda_t \mathcal{L}_{sf} \tag{6}$$

$$\mathcal{L}_{bg} = \sum_{v \in \Omega} (1 - C_{sp}) \mathrm{BCE}(P_t, 0) \tag{7}$$

$$\mathcal{L}_{sf} = \sum_{v \in \Omega} C_{sp} C_{se} \mathrm{BCE}(P_t, \pi_t) \tag{8}$$

3 Experiments

3.1 Datasets and Evaluation Metrics

We evaluate our method on Task06-Lung and Task07-Pancreas from the Medical Segmentation Decathlon (MSD) [1], targeting 3D tumor segmentation from a single noisy 3D bounding box provided by local radiologists. Task06-Lung and Task07-Pancreas contain 96 CT volumes and 420 CT volumes, respectively. Each CT volume is resampled to isotropic spacing of $1 \times 1 \times 1$ mm^3. We use a 4:1 ratio for training and testing. All quantitative results are obtained through 5-fold cross-validation. The Dice Coefficient (Dice) [5] and the 95th-percentile Hausdorff Distance (HD95) [10] are selected as quantitative evaluation metrics.

3.2 Implementation Details

We adopt SwinUNETR [7] as the default 3D backbone and additionally table results with UNETR [8] as backbone to demonstrate plug-and-play compatibility.

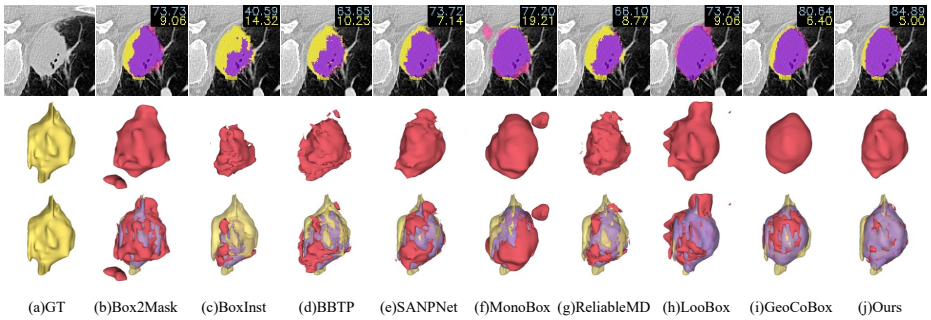

Fig. 2. Qualitative comparison on MSD-Lung. True positives, false positives, and false negatives are marked by purple, pink, and yellow, respectively.

All experiments are conducted on NVIDIA A100 GPU. The model is trained for 40 epochs using the AdamW [19] optimizer with an initial learning rate of $1e^{-4}$ and batch size is set to 4. Moreover, we set λ_t to Gaussian warming up function.

3.3 Comparisons with State-of-the-art Methods

We compare against some SOTA 3D box-supervised baselines under the same backbone. Table 1 summarizes quantitative results on MSD-Lung and MSD-Pancreas using SwinUNETR and UNETR as backbones. As shown in Table 1, our method consistently improves Dice and reduces HD95 across both datasets and backbones. Notably, methods relying on hard box-as-foreground priors degrade when the prompt box contains background contamination, whereas our safeguard frontier supervision remains robust by explicitly modeling the inner–outer uncertainty band and learning with a soft frontier target. Qualitative comparisons in Fig.2 further demonstrate that our approach better preserves fine tumor boundaries and reduces false positives near the box margin, yielding segmentations that more closely align with the ground truth (GT).

3.4 Ablation Studies

Effectiveness of the key components Table 2 validates the contribution of each component on MSD-Pancreas. Excluding spatial uncertainty partition (SUP) causes over-segmentation, while omitting hybrid reliability sampling (HRS) increases confirmation bias. Removing affinity-driven diffusion evolution (ADE) substantially degrades HD95. Collectively, these results demonstrate that our model synergistically integrates these mechanisms, achieving more competitive results compared to other variants.

Effectiveness of the Key Hyperparameter Table 3 presents the performance across different hyperparameter configurations (τ from 1.0 to 3.0, σ from 0.1 to 0.5) in two essential components: BUFP and RAD. We observe that the

Table 2. Ablation study of key components on MSD-Pancreas dataset.

Component	BUFP		RAD	Metrics	
	SUP	HRS	ADE	Dice ↑	HD95 ↓
w/o SUP		✓	✓	66.05	10.18
w/o HRS	✓		✓	65.72	10.92
w/o ADE	✓	✓		66.34	10.55
Ours	✓	✓	✓	**67.06**	**10.11**

Table 3. Sensitivity analysis of hyperparameters τ (BUFP) and σ (RAD) on MSD-Lung. Bold denotes the optimal value.

Setting	τ(1.0-3.0)					σ(0.1-0.5)				
	1.0	1.5	**2.0**	2.5	3.0	0.1	0.2	**0.3**	0.4	0.5
Dice ↑	65.21	66.45	**67.06**	66.77	65.83	66.14	66.79	**67.06**	66.93	66.16
HD95 ↓	12.94	11.02	**10.11**	10.49	11.35	11.82	10.78	**10.11**	10.34	11.43

model exhibits superior performance when we set τ and σ to 2.0 and 0.3, respectively. They are adopted in subsequent experiments. Consequently, we set $\tau = 2.0$ and $\sigma = 0.3$ for all subsequent experiments, as this configuration offers the most robust balance.

Effectiveness of Strategy Selection To explore how our strategy contributes to supervision, we visualize the evolution of the RAD in Figure 3. It can be found that foreground evidence propagates outward and background evidence propagates inward until converging at the frontier. This behavior aligns with the intended safeguard mechanism, enabling reliable boundary learning under noisy-box supervision.

4 Conclusion

We introduce a safeguard frontier paradigm for 3D tumor segmentation from noisy bounding boxes. To address background contamination and tumor leakage from naive box-to-mask constraints, our framework explicitly models coupled uncertainties via adversarial evidence propagation, constructing a soft frontier supervision target. The method partitions 3D space into a far-field safe background, high-confidence foreground, and an uncertainty-aware band, where reliable seeds are identified through hybrid reliability sampling. Additionally, by affinity-driven reverse adversarial diffusion, a soft frontier target is yielded to provide explicit supervision. Experiments on MSD-Lung and MSD-Pancreas demonstrate that our approach outperforms SOTA methods by at least 0.77% in Dice score. This

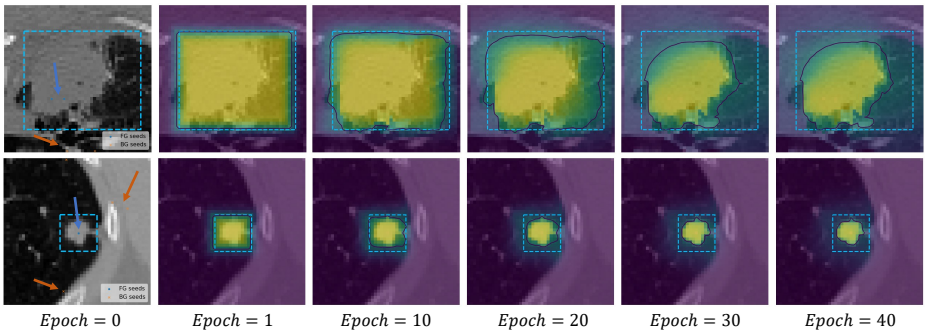

Fig. 3. Visualization of the RAD during different epochs.

work establishes a clinically viable paradigm for scalable 3D tumor segmentation under noisy-box supervision.

5 Acknowledgments.

This work was supported by the National Natural Science Foundation of China (No: 62401414) and the Joint Funds of the Natural Science Foundation of Hubei Province, China (No. 2026AFC0648) .

6 Disclosure of Interests.

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

References

1. Antonelli, M., Reinke, A., Bakas, S., Farahani, K., Kopp-Schneider, A., Landman, B.A., Litjens, G., Menze, B., Ronneberger, O., Summers, R.M., et al.: The medical segmentation decathlon. Nature Communications **13**(1), 4128 (2022)
2. Chi, J., Lin, G., Li, Z., Zhang, W., Chen, J.h., Huang, Y.: Coarse for fine: Bounding box supervised thyroid ultrasound image segmentation using spatial arrangement and hierarchical prediction consistency. IEEE Journal of Biomedical and Health Informatics **29**(6), 4186–4199 (2025)
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3d u-net: learning dense volumetric segmentation from sparse annotation. In: International Conference on Medical Image Computing and Computer-assisted Intervention. pp. 424–432. Springer (2016)
4. Deng, Q., Hui, L., Xie, J., Yang, J.: Sketchy bounding-box supervision for 3d instance segmentation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 8879–8888 (2025)

5. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology **26**(3), 297–302 (1945)
6. Fang, C., Wang, Q., Cheng, L., Gao, Z., Pan, C., Cao, Z., Zheng, Z., Zhang, D.: Reliable mutual distillation for medical image segmentation under imperfect annotations. IEEE Transactions on Medical Imaging **42**(6), 1720–1734 (2023)
7. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H.R., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. In: International MICCAI Brainlesion Workshop. pp. 272–284. Springer (2021)
8. Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D.: Unetr: Transformers for 3d medical image segmentation. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 574–584 (2022)
9. Hu, Q., Yi, Z., Zhou, Y., Huang, F., Liu, M., Li, Q., Wang, Z.: Monobox: Tightness-free box-supervised polyp segmentation using monotonicity constraint. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 3572–3580 (2025)
10. Huttenlocher, D.P., Klanderman, G.A., Rucklidge, W.J.: Comparing images using the hausdorff distance. IEEE Transactions on Pattern Analysis and Machine Intelligence **15**(9), 850–863 (2002)
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)
12. Jeon, Y.S., Yang, H., Fu, H., Kway, Y., Feng, M.: No more sliding window: efficient 3d medical image segmentation with differentiable top-k patch sampling. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 376–386. Springer (2025)
13. Kervadec, H., Dolz, J., Wang, S., Granger, E., Ayed, I.B.: Bounding boxes for weakly supervised segmentation: Global constraints get close to full supervision. In: Medical Imaging with Deep Learning. pp. 365–381. PMLR (2020)
14. Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al.: Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4015–4026 (2023)
15. Lan, T., Yi, Z., Xu, X., Zhu, M.: Loobox: Loose-box-supervised 3d tumor segmentation with self-correcting bidirectional learning. In: Proceedings of the 33rd ACM International Conference on Multimedia. pp. 8077–8086 (2025)
16. Lan, T., Yi, Z., Xu, X., Zhu, M.: Geocobox: Box-supervised 3d tumor segmentation via geometric co-embedding. In: Proceedings of the AAAI Conference on Artificial Intelligence (2026)
17. Li, W., Liu, W., Zhu, J., Cui, M., Yu, R., Hua, X., Zhang, L.: Box2mask: Box-supervised instance segmentation via level-set evolution. IEEE Transactions on Pattern Analysis and Machine Intelligence **46**(7), 5157–5173 (2024)
18. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., Van Der Laak, J.A., Van Ginneken, B., Sánchez, C.I.: A survey on deep learning in medical image analysis. Medical image analysis **42**, 60–88 (2017)
19. Loshchilov, I., Hutter, F.: Decoupled weight decay regularization. arXiv preprint arXiv:1711.05101 (2017)
20. Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B.: Segment anything in medical images. Nature Communications **15**(1), 654 (2024)

21. Rajchl, M., Lee, M.C., Oktay, O., Kamnitsas, K., Passerat-Palmbach, J., Bai, W., Damodaram, M., Rutherford, M.A., Hajnal, J.V., Kainz, B., et al.: Deepcut: Object segmentation from bounding box annotations using convolutional neural networks. IEEE Transactions on Medical Imaging **36**(2), 674–683 (2016)
22. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention. pp. 234–241. Springer (2015)
23. Rother, C., Kolmogorov, V., Blake, A.: " grabcut" interactive foreground extraction using iterated graph cuts. ACM Transactions on Graphics **23**(3), 309–314 (2004)
24. Song, B., Wang, Q.: Sdcl: Students discrepancy-informed correction learning for semi-supervised medical image segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 567–577. Springer (2024)
25. Tarvainen, A., Valpola, H.: Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. Advances in Neural Information Processing Systems **30** (2017)
26. Tian, Z., Shen, C., Wang, X., Chen, H.: Boxinst: High-performance instance segmentation with box annotations. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 5443–5452 (2021)
27. Wang, J., Xia, B.: Bounding box tightness prior for weakly supervised image segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 526–536. Springer (2021)
28. Wei, J., Hu, Y., Cui, S., Zhou, S.K., Li, Z.: Weakpolyp: You only look bounding box for polyp segmentation. In: International Conference on Medical Image Computing and Computer-assisted Intervention. pp. 757–766. Springer (2023)