



# Guiding the Guidance: A Comparative Analysis of User Guidance Signals for Interactive Segmentation of Volumetric Images

Zdravko Marinov<sup>1,2(✉)</sup>, Rainer Stiefelhofen<sup>1</sup>, and Jens Kleesiek<sup>3,4</sup>

<sup>1</sup> Karlsruhe Institute of Technology, Karlsruhe, Germany  
{zdravko.marinov,rainer.stiefelhofen}@kit.edu

<sup>2</sup> HIDSS4Health - Helmholtz Information and Data Science School for Health,  
Karlsruhe, Heidelberg, Germany

<sup>3</sup> Institute for AI in Medicine, University Hospital Essen, Essen, Germany  
jens.kleesiek@uk-essen.de

<sup>4</sup> Cancer Research Center Cologne Essen (CCCE), University Medicine Essen, Essen,  
Germany

**Abstract.** Interactive segmentation reduces the annotation time of medical images and allows annotators to iteratively refine labels with corrective interactions, such as clicks. While existing interactive models transform clicks into user guidance signals, which are combined with images to form (image, guidance) pairs, the question of how to best represent the guidance has not been fully explored. To address this, we conduct a comparative study of existing guidance signals by training interactive models with different signals and parameter settings to identify crucial parameters for the model's design. Based on our findings, we design a guidance signal that retains the benefits of other signals while addressing their limitations. We propose an adaptive Gaussian heatmap guidance signal that utilizes the geodesic distance transform to dynamically adapt the radius of each heatmap when encoding clicks. We conduct our study on the MSD Spleen and the AutoPET datasets to explore the segmentation of both anatomy (spleen) and pathology (tumor lesions). Our results show that choosing the guidance signal is crucial for interactive segmentation as we improve the performance by 14% Dice with our adaptive heatmaps on the challenging AutoPET dataset when compared to non-interactive models. This brings interactive models one step closer to deployment in clinical workflows. Code: <https://github.com/Zrrr1997/Guiding-The-Guidance/>.

**Keywords:** Interactive Segmentation · Comparative Study · Click Guidance

---

**Supplementary Information** The online version contains supplementary material available at [https://doi.org/10.1007/978-3-031-43898-1\\_61](https://doi.org/10.1007/978-3-031-43898-1_61).

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2023  
H. Greenspan et al. (Eds.): MICCAI 2023, LNCS 14222, pp. 637–647, 2023.  
[https://doi.org/10.1007/978-3-031-43898-1\\_61](https://doi.org/10.1007/978-3-031-43898-1_61)

# 1 Introduction

Deep learning models have achieved remarkable success in segmenting anatomy and lesions from medical images but often rely on large-scale manually annotated datasets [1–3]. This is challenging when working with volumetric medical data as voxelwise labeling requires a lot of time and expertise. Interactive segmentation models address this issue by utilizing weak labels, such as clicks, instead of voxelwise annotations [5–7]. The clicks are transformed into guidance signals, e.g., Gaussian heatmaps or Euclidean/Geodesic distance maps, and used together with the image as a joint input for the interactive model. Annotators can make additional clicks in missegmented areas to iteratively refine the segmentation mask, which often significantly improves the prediction compared to non-interactive models [4, 13]. However, prior research on choosing guidance signals for interactive models is limited to small ablation studies [5, 8, 9]. There is also no systematic framework for comparing guidance signals, which includes not only accuracy but also efficiency and the ability to iteratively improve predictions with new clicks, which are all important aspects of interactive models [7]. We address these challenges with the following contributions:

1. We compare 5 existing guidance signals on the AutoPET [1] and MSD Spleen [2] datasets and vary various hyperparameters. We show which parameters are essential to tune for each guidance and suggest default values.
2. We introduce 5 guidance evaluation metrics (**M1**)-(**M5**), which evaluate the performance, efficiency, and ability to improve with new clicks. This provides a systematic framework for comparing guidance signals in future research.
3. Based on our insights from 1., we propose novel adaptive Gaussian heatmaps, which use geodesic distance values around each click to set the radius of each heatmap. Our adaptive heatmaps mitigate the weaknesses of the 5 guidances and achieve the best performance on AutoPET [1] and MSD Spleen [2].

**Related Work.** Previous work comparing guidance signals has mostly been limited to small ablation studies. Sofiuk et al. [9] and Benenson et al. [8] both compare Euclidean distance maps with solid disks and find that disks perform better. However, neither of them explore different parameter settings for each guidance and both work with natural 2D images. Dupont et al. [12] note that a comprehensive comparison of existing guidance signals would be helpful in designing interactive models. The closest work to ours is MIDeepSeg [5], which proposes a user guidance based on exponentialized geodesic distances and compare it to existing guidance signals. However, they use only initial clicks and do not add iterative corrective clicks to refine the segmentation. In contrast to previous work, our research evaluates the influence of hyperparameters for guidance signals and assesses the guidances’ efficiency and ability to improve with new clicks, in addition to accuracy. While some previous works [20–22] propose using a larger radius for the first click’s heatmap, our adaptive heatmaps offer a greater flexibility by adjusting the radius at each new click dynamically.

## 2 Methods

### 2.1 Guidance Signals

We define the five guidance signals over a set of  $N$  clicks  $\mathcal{C} = \{c_1, c_2, \dots, c_N\}$  where  $c_i = (x_i, y_i, z_i)$  is the  $i^{th}$  click. As disks and heatmaps can be computed independently for each click, they are defined for a single click  $c_i$  over 3D voxels  $v = (x, y, z)$  in the volume. The **disk** signal fills spheres with a radius  $\sigma$  centered around each click  $c_i$ , which is represented by the equation in Eq. (1).

$$\text{disk}(v, c_i, \sigma) = \begin{cases} 1, & \text{if } \|v - c_i\|_2 \leq \sigma \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

The **Gaussian heatmap** applies Gaussian filters centered around each click to create softer edges with an exponential decrease away from the click (Eq. (2)).

$$\text{heatmap}(v, c_i, \sigma) = \exp\left(-\frac{\|v - c_i\|_2}{2\sigma^2}\right) \quad (2)$$

The **Euclidean distance transform** (EDT) is defined in Eq. (3) as the minimum Euclidean distance between a voxel  $v$  and the set of clicks  $\mathcal{C}$ . It is similar to the disk signal in Eq. (1), but instead of filling the sphere with a constant value it computes the distance of each voxel to the closest click point.

$$\text{EDT}(v, \mathcal{C}) = \min_{c_i \in \mathcal{C}} \|v - c_i\|_2 \quad (3)$$

The **Geodesic distance transform** (GDT) is defined in Eq. (4) as the shortest path distance between each voxel in the volume and the closest click in the set  $\mathcal{C}$  [14]. The shortest path in GDT also takes into account intensity differences between voxels along the path. We use the method of Asad et al. [10] to compute the shortest path which is denoted as  $\Phi$  in Eq. (4).

$$\text{GDT}(v, \mathcal{C}) = \min_{c_i \in \mathcal{C}} \Phi(v, c_i) \quad (4)$$

We also examine the **exponentialized Geodesic distance** (exp-GDT) proposed in MIDeepSeg [5] that is defined in Eq. (5) as an exponentiation of GDT:

$$\text{exp-GDT}(v, \mathcal{C}) = 1 - \exp(-\text{GDT}(v, \mathcal{C})) \quad (5)$$

*Note:* We normalize signals to  $[0, 1]$  and invert intensity values for Euclidean and Geodesic distances  $d(x)$  by  $1 - d(x)$  for better highlighting of small distances.

We define our **adaptive Gaussian heatmaps** ad-heatmap( $v, c_i, \sigma_i$ ) via:

$$\sigma_i = \lfloor ae^{-bx} \rfloor, \text{ where } x = \frac{1}{|\mathcal{N}_{c_i}|} \sum_{v \in \mathcal{N}_{c_i}} \text{GDT}(v, \mathcal{C}) \quad (6)$$

Here,  $\mathcal{N}_{c_i}$  is the 9-neighborhood of  $c_i$ ,  $a = 13$  limits the maximum radius to 13, and  $b = 0.15$  is set empirically<sup>1</sup> (details in supplementary). The radius  $\sigma_i$  is

<sup>1</sup> We note that  $b$  can also be automatically learned but we leave this for future work.

smaller for higher  $x$ , i.e., when the mean geodesic distance in the neighboring voxels is high, indicating large intensity changes such as edges. This leads to a more precise guidance with a smaller radius  $\sigma_i$  near edges and a larger radius in homogeneous areas such as clicks in the center of the object of interest. An example of this process and each guidance signal can be seen in Fig. 1a).

## 2.2 Model Backbone and Datasets

We use the DeepEdit [11] model with a U-Net backbone [15] and simulate a fixed number of clicks  $N$  during training and evaluation. For each volume,  $N$  clicks are iteratively sampled from over- and undersegmented predictions of the model as in [16] and represented as foreground and background guidance signals. We implemented our experiments with MONAI Label [23] and will release our code.

We trained and evaluated all of our models on the openly available AutoPET [1] and MSD Spleen [2] datasets. MSD Spleen [2] contains 41 CT volumes with voxel size  $0.79 \times 0.79 \times 5.00 \text{mm}^3$  and average resolution of  $512 \times 512 \times 89$  voxels with dense annotations of the spleen. AutoPET [1] consists of 1014 PET/CT volumes with annotated tumor lesions of melanoma, lung cancer, or lymphoma. We discard the 513 tumor-free patients, leaving us with 501 volumes. We also only use PET data for our experiments. The PET volumes have a voxel size of  $2.0 \times 2.0 \times 2.0 \text{mm}^3$  and an average resolution of  $400 \times 400 \times 352$  voxels.

## 2.3 Hyperparameters: Experiments

We keep these parameters constant for all models: learning rate =  $10^{-5}$ , #clicks  $N = 10$ , Dice Cross-Entropy Loss [24], and a fixed 80–20 training-validation split ( $\mathcal{D}_{\text{train}}/\mathcal{D}_{\text{val}}$ ). We apply the same data augmentation transforms to all models and simulate clicks as proposed in Sakinis et al. [16]. We train using one A100 GPU for 20 and 100 epochs on AutoPET [1] and MSD Spleen [2] respectively.

We vary the following four hyperparameters (H1)–(H4):

**(H1) Sigma.** We vary the radius  $\sigma$  of disks and heatmaps in Eq. (1) and (2) and also explore how this parameter influences the performance of the distance-based signals in Eq. (3)–(5). Instead of initializing the seed clicks  $\mathcal{C}$  as individual voxels  $c_i$ , we initialize the set of seed clicks  $\mathcal{C}$  as all voxels within a radius  $\sigma$  centered at each  $c_i$  and then compute the distance transform as in Eq. (3)–(5).

**(H2) Theta.** We explore how truncating the values of distance-based signals in Eq. (3)–(5) affects the performance. We discard the top  $\theta \in \{10\%, 30\%, 50\%\}$  of the distance values and keep only smaller distances closer to the clicks making the guidance more precise. Unlike MIDeepSeg [5], we compute the  $\theta$  threshold for each image individually, as fixed thresholds may not be suitable for all images.

**(H3) Input Adaptor.** We test three methods for combining guidance signals with input volumes proposed by Sofiuk et al. [9] - Concat, Distance Maps Fusion (DMF), and Conv1S. Concat combines input and guidance by concatenating their channels. DMF additionally includes  $1 \times 1$  conv. layers to adjust the channels to match the original size in the backbone. Conv1S has two branches for the guidance and volume, which are summed and fed to the backbone.

**Table 1.** Variation of hyperparameters **(H1)** – **(H4)** in our experiments.

$\sigma$	$\{0, 1, 5, 9, 13\}$	$\theta$	$\{0\%, 10\%, 30\%, 50\%\}$
Input Adaptor	$\{\text{Concat, DMF, Conv1S}\}$	$p$	$\{50\%, 75\%, 100\%\}$

**(H4) Probability of Interaction.** We randomly decide for each volume whether to add the  $N$  clicks or not, with a probability of  $p$ , in order to make the model more independent of interactions and improve its initial segmentation. All the hyperparameters we vary are summarized in Table 1. Each combination of hyperparameters corresponds to a separately trained DeepEdit [11] model.

## 2.4 Additional Evaluation Metrics

We use 5 metrics **(M1)**–**(M5)** (Table 2) to evaluate the validation performance.

**Table 2.** Evaluation metrics for the comparison of guidance signals.

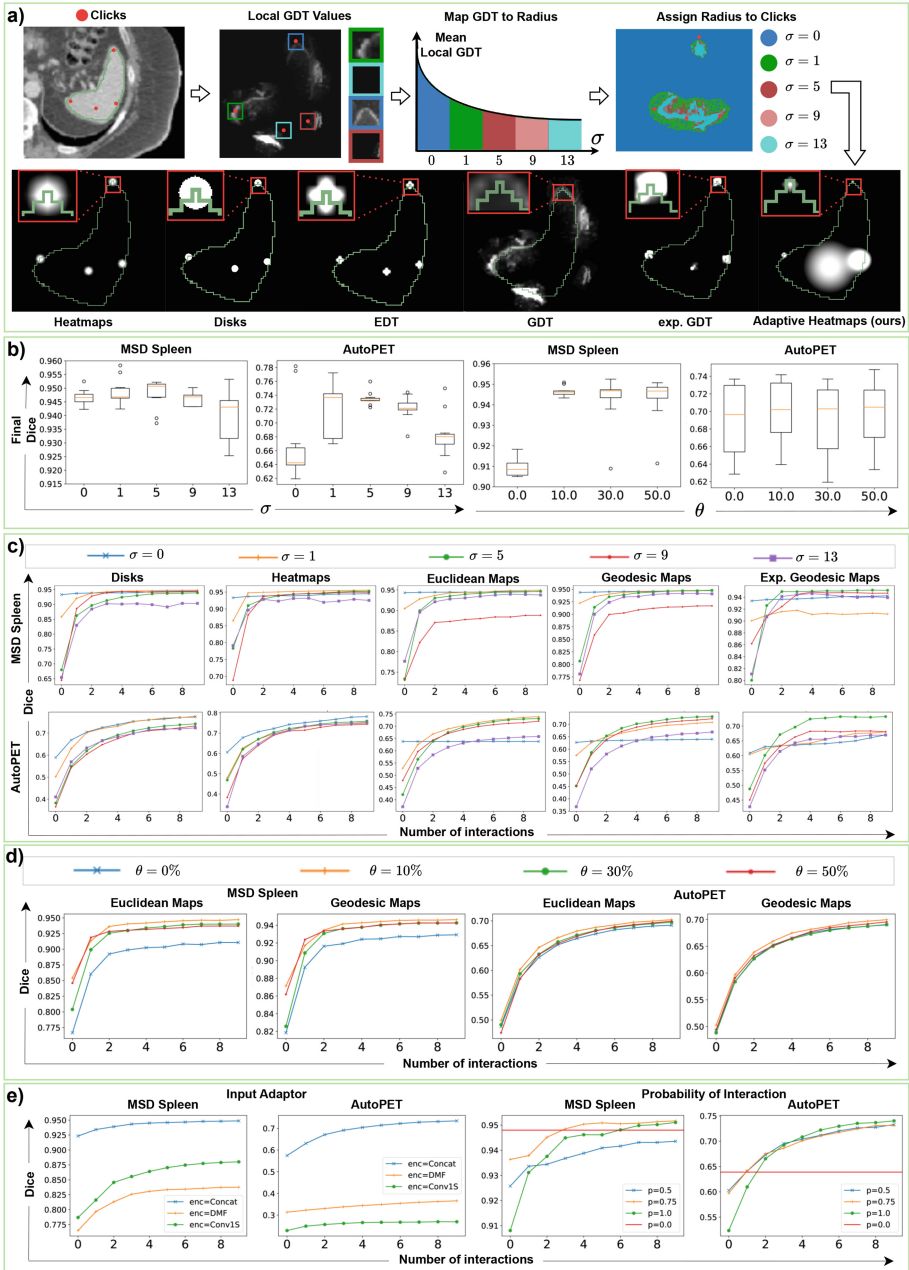
Metric	Description	
<b>(M1)</b>	Final Dice	Mean Dice score after $N = 10$ clicks per volume
<b>(M2)</b>	Initial Dice	Mean Dice score before any clicks per volume ( $N = 0$ ). A higher initial Dice indicates less work for the annotator
<b>(M3)</b>	Efficiency	Inverted* time measurement ( $1 - T$ ) in seconds needed to compute the guidance. Low efficiency increases the annotation time with every new click. Note that this metric depends on the volume size and hardware setup. *Our maximum measurement $T_{\max}$ is shorter than 1 second
<b>(M4)</b>	Consistent Improvement	Ratio of clicks $\mathcal{C}^+$ that improve the Dice score to the total number of validation clicks: $\frac{ \mathcal{C}^+ }{N \cdot  \mathcal{D}_{\text{val}} }$ , where $N = 10$ and $\mathcal{D}_{\text{val}}$ is the validation dataset
<b>(M5)</b>	Ground-truth Overlap	Overlap of the guidance $G$ with the ground-truth mask $M$ : $\frac{ M \cap G }{ G }$ . This estimates the guidance precision as corrective clicks are often near boundaries, and if guidances are too large, such as disks with a large $\sigma$ , there is a large overlap with the background outside the boundary

## 3 Results

### 3.1 Hyperparameters: Results

We first train a DeepEdit [11] model for each  $(\sigma, \theta)$  pair and set  $p = 100\%$  and the input adaptor to Concat to constrain the parameter space.

**(H1) Sigma.** Results in Fig. 1b) show that on MSD Spleen [2], the highest Dice scores are at  $\sigma = 5$ , with a slight improvement for two samples at  $\sigma = 1$ , but performance decreases for higher values  $\sigma > 5$ . On AutoPET [1],  $\sigma = 5$  and two samples with  $\sigma = 0$  show the best performance, while higher values again demonstrate a significant performance drop. Figure 1c) shows that the best final



**Fig. 1.** a) Our adaptive heatmaps use the mean local GDT values around each click to compute click-specific radiuses and form larger heatmaps in homogeneous regions and smaller near boundaries. b)–e) Results for varied hyperparameters. b) The final Dice of  $\sigma$  (left) and  $\theta$  (right) **aggregated** for all guidance signals. c) The influence of  $\sigma$  and d) the influence of  $\theta$  on individual guidances. e) Results for the different input adaptors (left) and probability of interaction  $p$  (right).

**Table 3.** Optimal parameter settings of our interactive models and the non-interactive baseline (non-int.) and their Dice scores. \*Same for both datasets.

	MSD Spleen [2]/AutoPET [1]						
	non-int	D	H	EDT	GDT	exp-GDT	Ours
$\sigma$	-/-	1/0	1/0	1*	5*	5*	adaptive*
$\theta$	-/-	-/-	-/-	10%*	10%*	-/-	-/-
Adaptor	Concat*	Concat*					
$p$	0%*	75%/100%					
Dice	94.9/64.9	95.9/78.2	95.8/78.2	95.8/75.2	95.2/74.5	95.2/73.2	<b>96.9/79.9</b>

and initial Dice for disks and heatmaps are with  $\sigma = 1$  and  $\sigma = 0$ . Geodesic maps exhibit lower Dice scores for small  $\sigma < 5$  and achieve the best performance for  $\sigma = 5$  on both datasets. Larger  $\sigma$  values lead to a worse initial Dice for all guidance signals. Differences in results for different  $\sigma$  values are more pronounced in AutoPET [1] as it is a more challenging dataset [17–19, 25].

**(H2) Theta.** We examine the impact of truncating large distance values for the EDT and GDT guidances from Eq. (3) and (4). Figure 1a) shows that the highest final Dice scores are achieved with  $\theta = 10$  for MSD Spleen [2]. On AutoPET [1], the scores are relatively similar when varying  $\theta$  with a slight improvement at  $\theta = 10$ . The results in Fig. 1d) also confirm that  $\theta = 10$  is the optimal parameter for both datasets and that not truncating values on MSD Spleen [2], i.e.  $\theta = 0$ , leads to a sharp drop in performance.

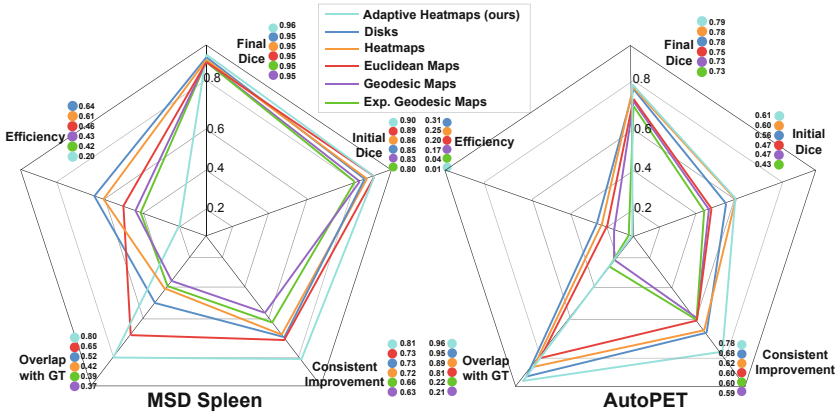
For our next experiments, we fix the optimal  $(\sigma, \theta)$  pair for each of the five guidances (see Table 3) and train a DeepEdit [11] model for all combinations of input adaptors and probability of interaction.

**(H3) Input Adaptor.** We look into different ways of combining guidance signals with input volumes using the input adaptors proposed by Sofiuek et al. [9]. The results in Fig. 1e) indicate that the best performance is achieved by simply concatenating the guidance signal with the input volume. This holds true for both datasets and the difference in performance is substantial.

**(H4) Probability of Interaction.** Figure 1e) shows that  $p \in \{75\%, 100\%\}$  results in the best performance on MSD Spleen [2], with a faster convergence rate for  $p = 75\%$ . However, with  $p = 50\%$ , the performance is worse than the non-interactive baseline ( $p = 0\%$ ). On AutoPET [1], the results for all  $p$  values are similar, but the highest Dice is achieved with  $p = 100\%$ . Note that  $p = 100\%$  results in lower initial Dice scores and requires more interactions to converge, indicating that the models depend more on the interactions. For the rest of our experiments, we use the optimal hyperparameters for each guidance in Table 1.

### 3.2 Additional Evaluation Metrics: Results

The comparison of the guidance signals using our five metrics (M1)–(M5) can be seen in Fig. 2. Although the concrete values for MSD Spleen [2] and



**Fig. 2.** Comparison of all guidance signals with our 5 metrics. The circles next to each metric represent the ranking of the guidances (sorted top-to-bottom).

AutoPET [1] are different, the five metrics follow the same trend on both datasets.

**(M1) Initial and (M2) Final Dice.** Overall, all guidance signals improve their initial-to-final Dice scores after  $N$  clicks, with AutoPET [1] showing a gap between disks/heatmaps and distance-based signals. Moreover, geodesic-based signals have lower initial scores on both datasets and require more interactions.

**(M3) Consistent Improvement.** The consistent improvement is  $\approx 65\%$  for both datasets, but it is slightly worse for AutoPET [1] as it is more challenging. Heatmaps and disks achieve the most consistent improvement, which means they are more precise in correcting errors. In contrast, geodesic distances change globally with new clicks as the whole guidance must be recomputed. These changes may confuse the model and lead to inconsistent improvement.

**(M4) Overlap with Ground Truth.** Heatmaps, disks, and EDT have a significantly higher overlap with the ground truth compared to geodesic-based signals, particularly on AutoPET [1]. GDT incorporates the changes in voxel intensity, which is not a strong signal for lesions with weak boundaries in AutoPET [1], resulting in a smaller overlap with the ground truth. The guidances are ranked in the same order in **(M3)** and in **(M4)** for both datasets. Thus, a good overlap with the ground truth can be associated with precise corrections.

**(M5) Efficiency.** Efficiency is much higher on MSD Spleen [2] compared to AutoPET [1], as AutoPET has a  $\times 2.4$  larger mean volume size. The time also includes the sampling of new clicks for each simulated interaction. Disks are the most efficient signal, filling up spheres with constant values, while heatmaps are slightly slower due to applying a Gaussian filter over the disks. Distance transform-based guidances are the slowest on both datasets due to their complexity, but all guidance signals are computed in a reasonable time ( $< 1$  s).



**Adaptive Heatmaps: Results.** Varying (H1)–(H4) and examining (M1)–(M5), we find disks/heatmaps as the best signals, but with inflexibility near edges due to their fixed radius (Fig. 1a)). Using GDT as a proxy signal to adapt  $\sigma_i$  for each click  $c_i$  mitigates this weakness by imposing large  $\sigma_i$  in homogeneous areas and small, precise  $\sigma_i$  near edges (Fig. 1a)). This results in substantially higher consistent improvement and overlap with ground truth and the best initial and final Dice (Table 3). Thus, our comparative study has led to the creation of a more consistent and flexible signal with a slight performance boost, albeit with an efficiency cost due to the need to compute both GDT and heatmaps.

## 4 Conclusion

Our comparative experiments yield insights into tuning existing guiding signals and designing new ones. We find that smaller radiuses ( $\sigma \leq 5$ ), a small threshold ( $\theta = 10\%$ ), more iterations with interactions ( $p \geq 75\%$ ), and traditional concatenation should be used. Weaknesses in existing signals include overly large radiuses near edges and inconsistent improvement for geodesic-based signals that change with each click. This analysis inspires our adaptive heatmaps, which adapt the radiuses of the heatmaps according to the geodesic values around the clicks, mitigating the inflexibility and inconsistency of existing guidances. We emphasize the importance of guidance representation in clinical applications, where a consistent and robust model is critical. Our study provides an overview of potential pitfalls, important parameters to tune, and how to design future guidance signals, along with proposed metrics for systematic comparison.

**Acknowledgements.** The present contribution is supported by the Helmholtz Association under the joint research school “HIDSS4Health - Helmholtz Information and Data Science School for Health. This work was performed on the HoreKa supercomputer funded by the Ministry of Science, Research and the Arts Baden-Württemberg and by the Federal Ministry of Education and Research.

## References

1. Gatidis, S., et al.: A whole-body FDG-PET/CT Dataset with manually annotated Tumor Lesions. *Sci. Data* **9**(1), 1–7 (2022)
2. Simpson, A.L., et al.: A large annotated medical image dataset for the development and evaluation of segmentation algorithms. *arXiv preprint arXiv:1902.09063* (2019)
3. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Trans. Med. Imaging* **34**(10), 1993–2024 (2014)
4. Kleesiek, J., et al.: Virtual raters for reproducible and objective assessments in radiology. *Sci. Rep.* **6**(1), 1–11 (2016)
5. Luo, X., et al.: MIDeepSeg: minimally interactive segmentation of unseen objects from medical images using deep learning. *Med. Image Anal.* **72**, 102102 (2021)
6. Gotkowsky, K., et al.: i3Deep: efficient 3D interactive segmentation with the nnU-Net. In: *International Conference on Medical Imaging with Deep Learning*. PMLR (2022)

7. Asad, M., Fidon, L., Vercauteren, T.: ECONet: efficient convolutional online likelihood network for scribble-based interactive segmentation. In: Medical Imaging with Deep Learning-MIDL 2022 (2022)
8. Benenson, R., Popov, S., Ferrari, V.: Large-scale interactive object segmentation with human annotators. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2019)
9. Sofiiuk, K., Petrov, I.A., Konushin, A.: Reviving iterative training with mask guidance for interactive segmentation. In: 2022 IEEE International Conference on Image Processing (ICIP), Bordeaux, France, pp. 3141–3145 (2022). <https://doi.org/10.1109/ICIP46576.2022.9897365>
10. Asad, M., Dorent, R., Vercauteren, T.: FastGeodis: fast generalised geodesic distance transform. *J. Open Source Softw.* **7**(79), 4532 (2022)
11. Diaz-Pinto, A., et al.: DeepEdit: deep editable learning for interactive segmentation of 3D medical images. In: Nguyen, H.V., Huang, S.X., Xue, Y. (eds.) Data Augmentation, Labelling, and Imperfections: Second MICCAI Workshop, DALI 2022, Held in Conjunction with MICCAI 2022, Singapore, 22 September 2022, Proceedings, pp. 11–21. Springer, Cham (2022). [https://doi.org/10.1007/978-3-031-17027-0\\_2](https://doi.org/10.1007/978-3-031-17027-0_2)
12. Dupont, C., Ouakrim, Y., Pham, Q.C.: UCP-net: unstructured contour points for instance segmentation. In: 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE (2021)
13. Mahadevan, S., Voigtlaender, P., Leibe, B.: Iteratively trained interactive segmentation. In: British Machine Vision Conference (BMVC) (2018)
14. Criminisi, A., Sharp, T., Blake, A.: GeoS: geodesic image segmentation. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008. LNCS, vol. 5302, pp. 99–112. Springer, Heidelberg (2008). [https://doi.org/10.1007/978-3-540-88682-2\\_9](https://doi.org/10.1007/978-3-540-88682-2_9)
15. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). [https://doi.org/10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28)
16. Sakinis, T., et al.: Interactive segmentation of medical images through fully convolutional neural networks. arXiv preprint [arXiv:1903.08205](https://arxiv.org/abs/1903.08205) (2019)
17. Heiliger, L., et al.: AutoPET challenge: combining nn-Unet with swin UNETR augmented by maximum intensity projection classifier. arXiv preprint [arXiv:2209.01112](https://arxiv.org/abs/2209.01112) (2022)
18. Zhong, S., Mo, J., Liu, Z.: AutoPET challenge 2022: automatic segmentation of whole-body tumor lesion based on deep learning and FDG PET/CT. arXiv preprint [arXiv:2209.01212](https://arxiv.org/abs/2209.01212) (2022)
19. Ye, J., et al.: Exploring vanilla U-net for lesion segmentation from whole-body FDG-PET/CT scans. arXiv preprint [arXiv:2210.07490](https://arxiv.org/abs/2210.07490) (2022)
20. Lin, Z., et al.: Interactive image segmentation with first click attention. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (2020)
21. Pirabakaran, R., Khan, N.: Interactive segmentation using U-Net with weight map and dynamic user interactions. In: 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). IEEE (2022)
22. Lin, C.-T., et al.: Interactive object segmentation with dynamic click transform. In: 2021 IEEE International Conference on Image Processing (ICIP). IEEE (2021)
23. Diaz-Pinto, A., et al.: Monai label: a framework for ai-assisted interactive labeling of 3d medical images. arXiv preprint [arXiv:2203.12362](https://arxiv.org/abs/2203.12362) (2022)

24. Isensee, F., et al.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**(2), 203–211 (2021)
25. Hallitschke, V.J., et al.: Multimodal interactive lung lesion segmentation: a framework for annotating PET/CT images based on physiological and anatomical cues. [arXiv:2301.09914](https://arxiv.org/abs/2301.09914) (2023)