Bidirectional Propagation Makes Sparse Re-Prompting Viable in Zero-Shot SAM2 for Whole-Tumor Glioma Segmentation

Sheikh Ziauddin[1] and Mirza Tahir Ahmed[2]

[1] Wilfrid Laurier University, Waterloo, Ontario, Canada
[2] Queen's University, Kingston, Ontario, Canada

Abstract. Segment Anything Model 2 (SAM2) can segment 3D medical volumes without training by propagating a single prompt across slices, but accuracy degrades with distance from the prompted slice. Two approaches have been proposed independently: bidirectional propagation and periodic re-prompting. We ask how they interact, evaluating SAM2 zero-shot on the whole-tumor region of BraTS2021 ($n = 150$) across a grid of propagation directions, seed criteria, and re-prompt intervals, under a pre-specified analysis with equivalence testing. At a moderate re-prompt interval, bidirectional and forward-only propagation are statistically *equivalent*: once corrections are frequent, propagation direction is immaterial. But the equivalence holds only at frequent re-prompting. As re-prompts grow sparse – the regime that matters when annotation is costly – bidirectional propagation degrades far less, opening a gap of up to 0.14 Dice. We further show that forward-only propagation from an interior seed forfeits accuracy to a structural coverage limit and, in a pilot ablation ($n = 30$), that retaining SAM2's memory bank across re-prompt boundaries silently negates re-prompting entirely. Bidirectional propagation does not add accuracy on top of frequent re-prompting; it makes infrequent re-prompting viable.

Keywords: SAM2 · Volumetric segmentation · Prompt propagation · Glioma · Zero-shot.

1 Introduction

The Segment Anything Model (SAM) [7] established promptable segmentation as a paradigm: a single click or box yields a mask without task-specific training. SAM2 [11] extends this from images to video, propagating a prompt across frames through a memory mechanism. Applied to 3D medical volumes – treating slices as video frames – this offers a training-free route to volumetric segmentation: a single boxed slice can, in principle, be propagated through an entire tumor. In practice, propagation from one prompt accumulates error with distance from the seed slice, and volumetric accuracy falls well short of per-slice prompting. The practical question is how to close that gap while keeping annotation cost low.

Two mechanisms have been proposed independently to narrow it. One: bidirectional propagation seeds from an interior slice and propagates outward in both directions, halving the maximum distance any prediction travels from an anchor [4]. Two: periodic re-prompting supplies a fresh prompt every few slices, capping the drift between corrections [9]. Each helps, but each has been studied in isolation: Dong et al. [4] report bidirectional propagation's benefit aggregated across datasets, and Ludwig et al. [9] demonstrate re-prompting at a single fixed frequency without varying direction. What neither establishes is how the two interact – whether they are redundant (both merely shortening the distance to the nearest anchor), complementary, or whether one dominates. Because both add to the annotation budget, their interaction, not either mechanism alone, determines the best achievable accuracy–cost tradeoff. A separate line of work adapts SAM/SAM2 to medical data through training (dataset-specific fine-tuning [10, 14, 3], 3D-native variants [13, 2, 12]). These can improve accuracy but require task-specific training data, sacrificing the zero-shot, immediately-deployable property that motivates propagation. We deliberately restrict attention to the zero-shot setting as the regime relevant to a practitioner without the data or resources to train.

We characterize this interaction directly on the whole-tumor region of BraTS2021 [1], evaluating SAM2 fully zero-shot across a grid of propagation directions, seed-slice criteria, and re-prompt intervals. Our contributions are:

- We show, using pre-specified equivalence testing, that at a moderate re-prompt interval bidirectional and forward-only propagation are *statistically equivalent*: once re-prompting is frequent enough, propagation direction no longer matters.
- We show the equivalence is confined to frequent re-prompting. In the prompt-efficient regime that matters in practice, bidirectional propagation degrades far more gracefully, making sparse re-prompting viable where forward-only propagation cannot.
- We report, in a pilot ablation (n = 30), a silent implementation pitfall: retaining SAM2's memory bank across re-prompt boundaries makes periodic re-prompting statistically equivalent to not re-prompting at all.

2 Methods

2.1 Data and Model

We evaluate on BraTS2021 [1], restricting to the whole-tumor (WT) label (the union of all non-zero tumor labels) on the FLAIR sequence, following common practice in the BraTS literature. All experiments use $n = 150$ patients drawn uniformly at random from the full 1,251-patient training cohort with a fixed random seed (`seed=42`).

We use SAM2.1 Hiera-Large with `multimask_output=False` (a single mask per prompt, which outperformed a best-of-three-mask alternative on this data)

in a fully zero-shot setting; all prompts, including re-prompts, are oracle bounding boxes derived from ground-truth masks. Experiments were implemented in PyTorch and completed in approximately 12 hours on a single NVIDIA RTX 5060 Ti (16 GB).

Each axial FLAIR slice is normalized independently by linearly rescaling its 1st–99th intensity-percentile range to $[0, 1]$ (values outside are clipped), then written as an 8-bit RGB frame (the grayscale channel replicated three times) at native in-plane resolution. Boxes are the tight bounding box of the ground-truth mask on each slice; slices with an empty ground-truth mask carry no box, and the pipeline advances in the propagation direction to the next slice with a non-empty mask. Bidirectional configurations propagate from the seed toward each end of the tumor's z-extent; the two half-ranges meet at the seed and are concatenated (the seed segmented once, by the forward pass), so no slice is produced by both directions. At each re-prompt boundary in the default `reset` mode, SAM2's memory bank is cleared with `reset_state` before the fresh box is added, so each inter-boundary segment is an independent run rather than one competing with retained features; we examine the alternative (persisting memory) in Section 4.

2.2 Evaluation Metrics

We report the volumetric Dice similarity coefficient, $\mathrm{Dice}(P, G) = 2|P \cap G|/(|P| + |G|)$, computed over the fully reconstructed 3D volume (every predicted slice assembled into a single binary mask and compared against the ground-truth sub-region volume in one pass). Because Dice can mask boundary errors and oversegmentation spikes, we additionally report the 95th-percentile Hausdorff distance (HD95, in mm) and the Normalized Surface Dice (NSD, 1 mm tolerance), both computed on the same reconstructed volumes from ground-truth voxel spacing.

We count a prompt as one distinct boxed slice position – the number of prompted slices, a proxy for annotation effort rather than a direct measure of annotation time or compute cost. A seed re-anchored in both directions counts once, since it is the same slice position. Because all boxes are oracle boxes (Section 2.1), reported accuracies are upper bounds on what manual prompting would achieve.

2.3 Configurations

Table 1 summarizes the eleven configurations evaluated, spanning three factors: propagation direction (forward-only vs. bidirectional), seed-slice criterion (first slice, max-area slice, middle slice of the tumor's z-extent, or the area-weighted z-centroid), and periodic re-prompting (none, or a fresh oracle box supplied every N slices, swept over $N \in \{3, 5, 10, 15, 20, 30\}$).

Configurations seeded away from a volume edge without bidirectionality (Fwd-Max, Fwd-Mid, and their re-prompted counterparts FXR, FMR) have *partial* coverage: forward-only propagation only ever moves from the seed toward the end of the slice range, so tumor-containing slices before the seed are

Table 1. All configurations evaluated. Short codes used in dense tables; descriptive names used in prose.

Short code	Descriptive name	Direction	Seed criterion	Re-prompt interval	Coverage
FF	Fwd-First	forward	first slice	none	full
FX	Fwd-Max	forward	max-area	none	partial
FM	Fwd-Mid	forward	middle	none	partial
BX	Bidir-Max	bidirectional	max-area	none	full
BM	Bidir-Mid	bidirectional	middle	none	full
BC	Bidir-Cent	bidirectional	area-weighted centroid	none	full
FFR	Fwd-First+RP	forward	first slice	3–30 (swept)	full
FXR	Fwd-Max+RP	forward	max-area	3–30 (swept)	partial
FMR	Fwd-Mid+RP	forward	middle	3–30 (swept)	partial
BXR	Bidir-Max+RP	bidirectional	max-area	3–30 (swept)	full
BMR	Bidir-Mid+RP	bidirectional	middle	3–30 (swept)	full

never segmented, regardless of re-prompting frequency. For all re-prompting configurations, SAM2's memory bank is, by default, fully reset (not incrementally added to) at each re-prompt boundary; we return to the consequences of this choice in Section 4.

2.4 Statistics

We pre-specify a primary hypothesis family of two paired comparisons, each paired on patient identity. The first tests whether adding periodic re-prompting to bidirectional propagation improves it (Bidir-Mid+RP vs. Bidir-Mid, the identical configuration with correction off) – the controlled test of re-prompting alone. The second tests whether the best bidirectional re-prompted configuration measurably beats the simplest forward re-prompted one (Bidir-Mid+RP vs. Fwd-First+RP). We use the Wilcoxon signed-rank test throughout because some configurations (notably Fwd-First) exhibit a floor/near-bimodal Dice distribution for which a paired t-test and Cohen's d are unreliable. Within the family, p-values are corrected using the Holm–Bonferroni procedure [5]. We additionally report percentile bootstrap 95% CIs on the paired mean difference (10,000 resamples). Because a claim of practical interchangeability is a claim of equivalence – which non-significance alone cannot establish – we additionally apply a two-one-sided-tests (TOST) equivalence check [8] to the Bidir-Mid+RP vs. Fwd-First+RP contrast, at a 0.01 Dice margin, fixed in advance via an $n = 30$ pilot plan.

The re-prompt-interval sweep (Section 3.3) forms a separate, self-contained exploratory family of five tests (intervals 3, 5, 15, 20, and 30; interval = 10 is already covered by the primary family and is not re-tested here), corrected internally with Holm–Bonferroni. All other comparisons reported in this paper (Table 2's descriptive statistics, the centroid-seeding comparison in Section 4) are exploratory and uncorrected, and are flagged as such rather than treated as confirmatory.

Table 2. Prompt-count vs. accuracy, all configurations, as mean±std over $n = 150$. Prompts = mean oracle boxes per patient; the ceiling prompts every tumor-containing slice. HD95 in mm; NSD at 1 mm tolerance; ↑/↓ mark whether higher or lower is better. Boxes are oracle boxes, so figures are upper bounds. The nnU-Net row is cited, not measured.

Config	Dice ↑	HD95 ↓	NSD ↑	Prompts
Ceiling (per-slice)	0.862 ± 0.092	7.3 ± 5.3	0.621 ± 0.150	67.7
FF (fwd, first)	0.304 ± 0.313	36.2 ± 23.8	0.208 ± 0.215	1
FX (fwd, max-area)	0.582 ± 0.117	27.7 ± 11.7	0.350 ± 0.118	1
FM (fwd, middle)	0.548 ± 0.104	28.2 ± 9.9	0.345 ± 0.124	1
BX (bidir, max-area)	0.792 ± 0.128	15.8 ± 12.9	0.470 ± 0.172	1
BM (bidir, middle)	0.799 ± 0.133	14.3 ± 12.7	0.491 ± 0.190	1
BC (bidir, centroid)	0.798 ± 0.128	14.3 ± 12.2	0.487 ± 0.184	1
FFR (fwd, first, +rp10)	0.830 ± 0.102	7.9 ± 5.9	0.552 ± 0.143	7.2
FXR (fwd, max, +rp10)	0.599 ± 0.109	25.5 ± 10.3	0.392 ± 0.106	4.0
FMR (fwd, mid, +rp10)	0.566 ± 0.093	27.1 ± 8.8	0.385 ± 0.103	3.8
BXR (bidir, max, +rp10)	0.832 ± 0.108	8.6 ± 6.3	0.547 ± 0.156	6.8
BMR (bidir, mid, +rp10)	0.831 ± 0.113	8.4 ± 6.9	0.550 ± 0.163	6.7
nnU-Net (cited, 0 prompts)	∼0.92	–	–	0

3 Results

3.1 Main Comparison

Table 2 reports Dice and prompt count for all eleven configurations. Every bidirectional single-seed configuration substantially outperforms its forward-only counterpart regardless of seed criterion: the three bidirectional variants cluster tightly (0.792–0.799 Dice), well above every forward-only variant (0.304–0.582).

Periodic re-prompting sharply narrows the gap to the per-slice oracle, but only where coverage is complete. The three full-coverage re-prompted configurations (Fwd-First+RP, Bidir-Max+RP, Bidir-Mid+RP) reach within 0.030–0.032 Dice of the oracle (0.862) using roughly 90% fewer prompts (6.7–7.2 vs. 67.7 per patient), and differ from one another by at most 0.002 Dice – a near-equivalence Section 3.2 examines formally. Re-prompting's effect is largest on the weakest baseline (Fwd-First, $0.304 \rightarrow 0.830$) and smallest on the already-strong bidirectional configurations, consistent with periodic correction compensating for long-range propagation drift rather than improving already-reliable tracking. Re-prompting the forward-only, non-edge-seeded configurations (Fwd-Max+RP, Fwd-Mid+RP) instead recovers only a small fraction of the gap: both remain far below their bidirectional, fully-covered counterparts despite a comparable prompt budget, for reasons described in Section 3.4. We include these partial-coverage configurations not as competitive baselines – they segment only part of the volume by construction – but as a controlled probe that isolates coverage from tracking quality (Section 3.4).

Table 3. Pre-specified primary hypothesis family: BMR vs. BM and BMR vs. FFR. Δ = paired mean difference; 95% CI from 10,000 bootstrap resamples; TOST margin = 0.01 Dice.

Contrast	Δ	95% CI	Result
BMR vs. BM	+0.032	$[0.020, 0.044]$	real difference, $p_{\text{Holm}} = 2.7 \times 10^{-12}$
BMR vs. FFR	+0.001	$[-0.005, 0.006]$	**equivalent** (TOST $p = 0.0003$)

For context, Table 2 includes nnU-Net [6], which reports approximately 0.92 Dice on WT fully automatically with zero prompts at inference time. Propagation is not intended to compete with a trained model where one exists; its target is settings without one (Section 1).

Boundary metrics (HD95, NSD) corroborate the Dice findings (Table 2). Re-prompting improves boundary accuracy even more than it improves Dice: Bidir-Mid+RP more than halves the HD95 of single-seed Bidir-Mid (8.4 vs. 14.3 mm). The coverage-limited forward configurations show the worst boundaries by a wide margin (Fwd-Max+RP 25.5 mm vs. Bidir-Max+RP 8.6 mm HD95), consistent with their unreached slices (Section 3.4). Crucially, the primary equivalence holds on boundaries: Bidir-Mid+RP and Fwd-First+RP are indistinguishable on both HD95 and NSD after Holm correction across the three metrics ($p_{\text{Holm}} \geq 0.09$; the raw HD95 gap is 0.5 mm, below the 1 mm voxel size).

3.2 Bidirectionality vs. Periodic Re-Prompting

The first comparison in Table 3 confirms an expected effect: adding periodic re-prompting to bidirectional propagation improves it (Bidir-Mid+RP over Bidir-Mid, the identical configuration with correction off; Wilcoxon $p = 1.4 \times 10^{-12}$). This is the controlled test of re-prompting alone – same seed, same direction, correction on versus off – and it behaves as expected.

The second comparison is the surprising one. The single-seed results predict that Bidir-Mid+RP should also beat Fwd-First+RP: without re-prompting, bidirectional propagation dominates forward-only by a wide margin (0.799 vs. 0.304, Table 2). Yet once both are re-prompted, that advantage vanishes – the two are statistically indistinguishable (Wilcoxon $p = 0.52$). This is not merely a failure to detect a difference: the TOST equivalence check confirms the two are practically interchangeable at the pre-specified 0.01 Dice margin (Table 3), establishing equivalence rather than absence of evidence.

Once periodic re-prompting is in place, the bidirectional propagation and careful seed choice that mattered so much for single-seed segmentation buy nothing measurable over the naive forward, first-slice configuration. We next ask whether this equivalence persists across the full range of practically relevant re-prompt intervals.

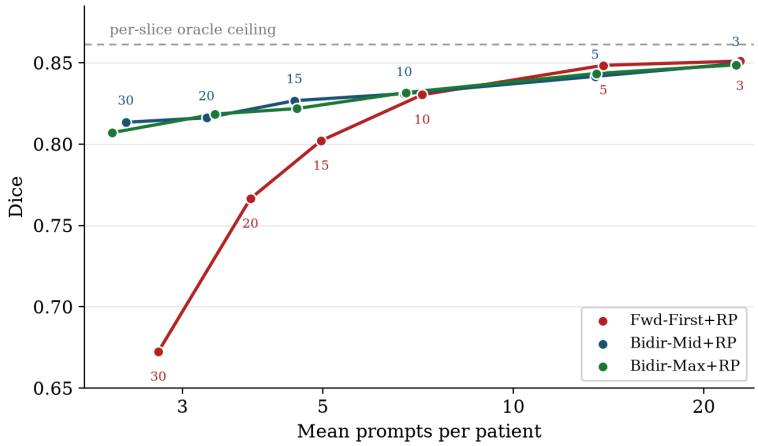

Fig. 1. Dice vs. mean prompts per patient across the re-prompt-interval sweep (log-scaled x-axis; labels = interval in slices). Forward-only, non-edge-seeded configurations (FXR, FMR) are omitted for clarity; they remain flat but capped low (Section 3.4).

3.3 The Effect of Re-Prompt Interval on Bidirectionality

The equivalence established in Section 3.2 holds only at a 10-slice interval; it is one point on a broader relationship, not a general property of periodic re-prompting (Fig. 1). Both bidirectional configurations stay nearly flat as the interval widens from 3 to 30 slices (spans of 0.037 and 0.042 Dice), whereas Fwd-First+RP falls from 0.851 to 0.672 – a span of 0.179, roughly 4–5× larger. The Holm-corrected comparisons (BMR vs. FFR across the five swept intervals as a self-contained family; interval = 10 belongs to the primary family) reveal two regimes rather than a smooth trend. Under dense re-prompting, direction barely matters: at intervals 3 and 5, Fwd-First+RP holds a statistically detectable but practically negligible edge (−0.002 and −0.007 Dice). As re-prompts grow sparse, bidirectionality's advantage rises steeply with Bidir-Mid+RP edging +0.141 Dice at interval = 30 ($p_{\text{Holm}} < 10^{-4}$ at every interval from 15 onward).

Because Bidir-Max+RP and Bidir-Mid+RP agree to within 0.001–0.010 Dice at every interval, this robustness is attributable to bidirectional propagation itself, not the seed criterion. This is what makes sparse annotation practical: at interval = 30, Bidir-Mid+RP retains 0.813 Dice from only 2.4 boxed slices per patient (vs. 67.7 for per-slice labelling), where forward-only propagation has already failed.

3.4 Why Forward-Only Propagation Fails from a Non-Edge Seed

Forward-only propagation from a non-edge seed never reaches tumor-containing slices preceding the seed. Re-prompt frequency cannot change this – re-prompting only re-corrects slices already reached – so Fwd-Max+RP and Fwd-Mid+RP

stay capped well below the bidirectional configurations (0.599 and 0.566 at interval = 10, Table 2). The effect is measurable per patient: Fwd-Max+RP reaches on average 53% of tumor slices, and this coverage correlates strongly with Dice ($r = 0.70$, $p < 10^{-22}$, $n = 150$), as patients whose seed sits nearer a volume edge reach more of the tumor. Because forward-only propagation and the forward pass of bidirectional propagation are identical on the slices they share, this gap reflects missing coverage alone, not worse tracking.

4 Discussion

Practical guidance. For zero-shot SAM2 propagation under an annotation budget, the results give a simple recipe: propagate bidirectionally from an interior seed, and widen the re-prompt interval as far as the accuracy target allows. Bidirectional configurations retain near-peak accuracy even at sparse re-prompting (as few as 2.4 prompts per volume at interval 30), where forward-only propagation pays a steep penalty for the same budget. When re-prompting is frequent, direction is immaterial (Section 3.2), so bidirectional propagation costs nothing where it does not help and is a safe default throughout. Seed criterion can be chosen on convenience: max-area, middle, and area-weighted-centroid seeds are statistically interchangeable (Bidir-Cent vs. Bidir-Mid, $p = 0.665$), so the simplest heuristic suffices. This is not an artifact of the criteria selecting near-identical slices: the max-area and middle seeds sit a mean of 5.6 ± 4.4 slices apart (up to 23), yet yield indistinguishable accuracy – bidirectional propagation absorbs the difference. What matters is coverage: an edge seed, or bidirectional propagation from any interior seed, avoids the structural cap on forward-only propagation from an interior seed (Section 3.4).

A silent implementation pitfall. One implementation detail carries outsized practical consequence: SAM2's memory bank must be reset at each re-prompt boundary, not appended to. In a pilot ablation ($n = 30$), allowing memory to persist across boundaries rendered periodic re-prompting statistically equivalent (TOST, 0.01 Dice margin) to never re-prompting at all for three of four good-seed configurations – the corrective prompts were absorbed by the retained memory and effectively ignored. The failure is silent: the output resembles an ordinary single-seed propagation run, with no error and no obvious degradation to flag the lost corrections. We highlight it because it is easy to introduce unknowingly and costly when the initial seed is poor.

Limitations and future work. Our characterization is deliberately narrow: a single region (whole tumor), dataset (BraTS2021), and model (SAM2, zero-shot), which isolates the direction/re-prompting interaction but leaves its generality across sub-regions, modalities, and anatomies open. Our prompts are also idealized in two ways: the bounding boxes are oracle boxes derived from ground truth, and the seed criteria (max-area, middle, centroid) require the tumor's extent and per-slice areas, which are unavailable before annotation. Both make our

accuracy–cost figures upper bounds on what noisier manual prompting achieves. Under hand-drawn boxes we would expect absolute accuracies to drop, but the *relative* ordering we study – the direction/re-prompting interaction – to persist, since it turns on coverage and drift rather than box precision. The seed idealization, however, matters less in practice because of our null result: since no reasonable interior seed outperforms another (Section 3.3), a practitioner can place the seed on an approximately central slice by visual inspection, rather than computing a specific one. Whether the same direction/re-prompting interaction holds for other tumor sub-regions (tumor core, enhancing tumor), modalities, and pathologies beyond whole-tumor glioma remains open, though the coverage and drift mechanisms it rests on are not specific to this setting. Three extensions follow directly: applying the same grid to the tumor-core and enhancing-tumor sub-regions and other modalities to test generality; replacing oracle boxes with perturbed or detector-generated prompts to quantify robustness to prompt noise; and, motivated by the interval sweep, replacing fixed-interval re-prompting with an adaptive schedule that re-prompts only where propagation begins to degrade.

5 Conclusion

We characterized how propagation direction, seed selection, and periodic re-prompting interact in zero-shot SAM2 propagation for whole-tumor glioma segmentation. The two mechanisms proposed independently to improve propagation – bidirectional propagation and periodic re-prompting – turn out not to be independent levers. When re-prompting is frequent, adding bidirectional propagation on top of it changes nothing: forward-only and bidirectional propagation become statistically equivalent. But this redundancy is confined to frequent re-prompting. In the prompt-efficient regime that matters when annotation is costly, bidirectional propagation degrades far more gracefully, and the gap grows with the re-prompt interval. Forward-only propagation from an interior seed additionally forfeits accuracy to a structural coverage limit that no amount of re-prompting repairs. Periodic re-prompting does not eliminate the value of bidirectional propagation; it merely masks it when re-prompts are sufficiently frequent.

References

1. Baid, U., Ghodasara, S., Mohan, S., Bilello, M., Calabrese, E., Colak, E., Others: The rsna-asnr-miccai brats 2021 benchmark on brain tumor segmentation and radiogenomic classification (2021), https://arxiv.org/abs/2107.02314
2. Chen, C., Miao, J., Wu, D., Zhong, A., Yan, Z., Kim, S., Hu, J., Liu, Z., Sun, L., Li, X., Liu, T., Heng, P.A., Li, Q.: Ma-sam: Modality-agnostic sam adaptation for 3d medical image segmentation. Medical Image Analysis **98**, 103310 (2024). https://doi.org/https://doi.org/10.1016/j.media.2024.103310, https://www.sciencedirect.com/science/article/pii/S1361841524002354

3. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., Sun, H., He, J., Zhang, S., Zhu, M., Qiao, Y.: Sam-med2d (2023), https://arxiv.org/abs/2308.16184
4. Dong, H., Gu, H., Chen, Y., Yang, J., Chen, Y., Mazurowski, M.A.: Segment anything model 2: an application to 2d and 3d medical images. IEEE Transactions on Biomedical Engineering (2026)
5. Holm, S.: A simple sequentially rejective multiple test procedure. Scandinavian journal of statistics pp. 65–70 (1979)
6. Isensee, F., Jäger, P.F., Full, P.M., Vollmuth, P., Maier-Hein, K.H.: nnu-net for brain tumor segmentation. In: Crimi, A., Bakas, S. (eds.) Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries. pp. 118–132. Springer International Publishing, Cham (2021)
7. 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)
8. Lakens, D.: Equivalence tests: A practical primer for t tests, correlations, and meta-analyses. Social psychological and personality science **8**(4), 355–362 (2017)
9. Ludwig, C., Namdar, K., Khalvati, F.: Ai-driven mri-based brain tumour segmentation benchmarking (2025), https://arxiv.org/abs/2506.20786
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 (2025), https://arxiv.org/abs/2504.03600
11. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Others: Sam 2: Segment anything in images and videos (2024), https://arxiv.org/abs/2408.00714
12. Shen, Y., Li, J., Shao, X., Inigo Romillo, B., Jindal, A., Dreizin, D., Unberath, M.: FastSAM3D: An Efficient Segment Anything Model for 3D Volumetric Medical Images . In: proceedings of Medical Image Computing and Computer Assisted Intervention – MICCAI 2024. vol. LNCS 15012. Springer Nature Switzerland (October 2024)
13. Wang, H., Guo, S., Ye, J., Deng, Z., Cheng, J., Li, T., Chen, J., Su, Y., Huang, Z., Shen, Y., Fu, B., Zhang, S., He, J.: Sam-med3d: A vision foundation model for general-purpose segmentation on volumetric medical images. IEEE Transactions on Neural Networks and Learning Systems **36**(10), 17599–17612 (2025). https://doi.org/10.1109/TNNLS.2025.3586694
14. Zhu, J., Hamdi, A., Qi, Y., Jin, Y., Wu, J.: Medical sam 2: Segment medical images as video via segment anything model 2 (2024), https://arxiv.org/abs/2408.00874