Towards user-centered interactive medical image segmentation in VR with an assistive AI agent

Pascal Spiegler[1], Arash Harirpoush[1] and Yiming Xiao[1]

Department of Computer Science and Software Engineering, Concordia University, Montreal, Quebec, Canada.

Challenges in segmentation

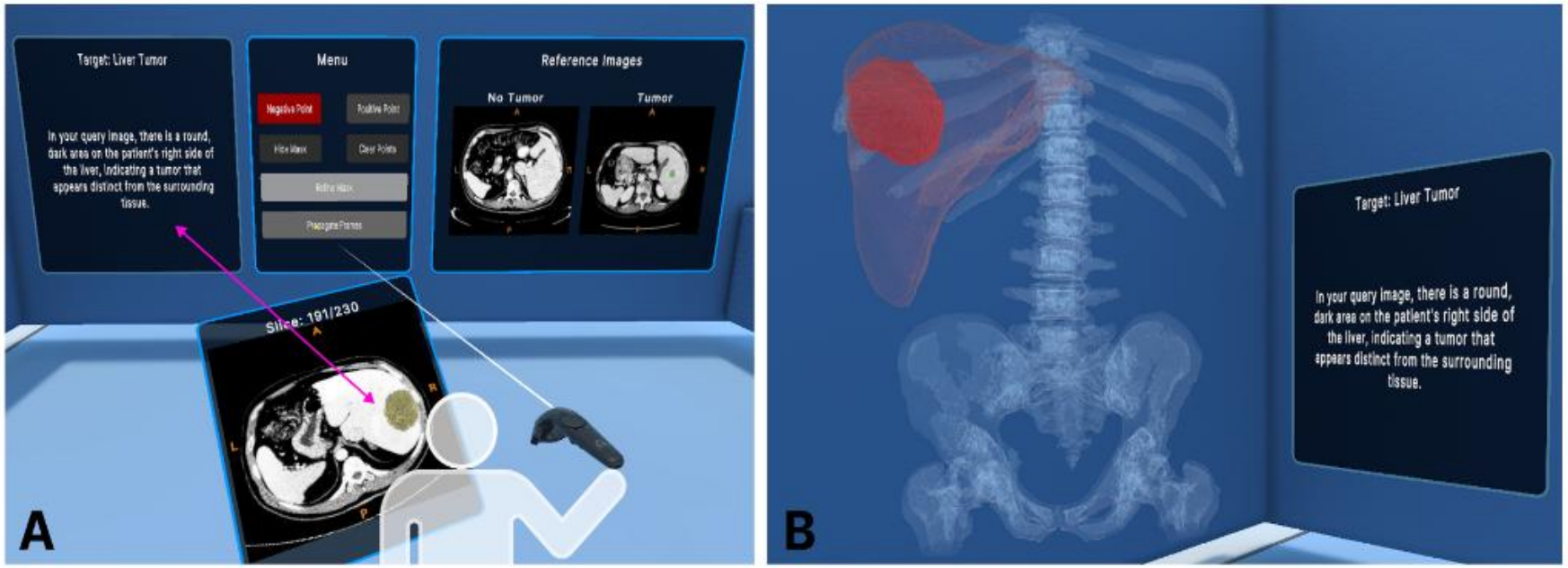
Problem:

- Manual segmentation of 3D medical scans is time-consuming, labor-intensive, and prone to errors
- Fully automatic algorithms lack reliability without user feedback
- Existing tools do not seamlessly integrate AI and VR for interactive, patient-specific segmentation

Goal:

- Develop an immersive VR system with AI assistance that facilitates **efficient, accurate, and user-friendly** segmentation
- Investigate **optimal interaction modes** (controller, head, eye) for mask refinement in VR

Solution proposed : SAMIRA



Development of SAMIRA, a conversational AI agent in VR for medical image segmentation.

Related works

1) VR & Medical Training:

Early VR systems support clinical education but lack AI integration for segmentation (*Chheang et al., 2024*)

2) AI in Medical Imaging:

Foundation models like **BiomedParse** and **SAM2** aid in segmentation but are limited to 2D or desktop applications (*Zhao et al., 2025, Kirillov et al., 2024)*

3) Human-Computer Interaction in VR:

- Studies compare input modes: controller, head pointing, eye tracking (*Sidenmark et al., 2023, Luro et al., 2019, Schultheis et al., 2024*)
- Prior work shows controllers excel in accuracy; head and eye tracking are more natural but less precise

Gap:

No prior VR system integrates **AI-powered, interactive 3D segmentation** with **human-in-the-loop workflows**

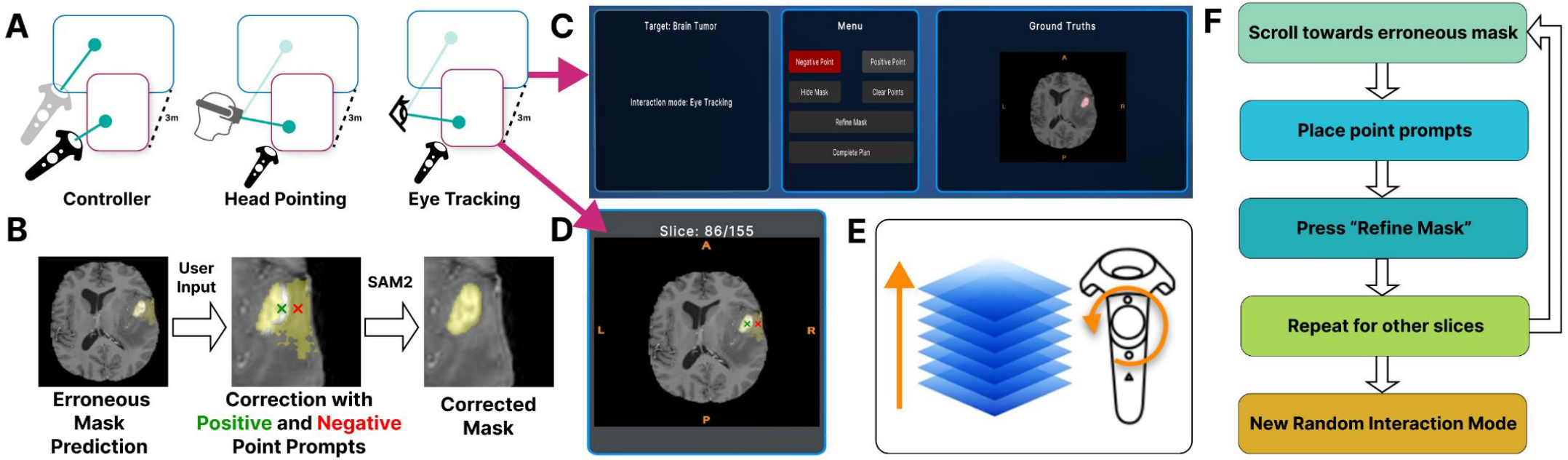
SAMIRA

Fusion of AI & VR: Combines foundational radiological models with an immersive VR interface for real-time, human-in-the-loop segmentation

Key Components:

- **AI Models:** BiomedParse and SAM2 perform initial segmentation, refinement, and mask propagation.
- **VR Interface:** Visualizes slices, reference images, and 3D reconstructions; supports speech and input modes for users.

Workflow : Interaction modes



The AI-Assisted Segmentation System

1) BiomedParse:

- A **Transformer-based vision-language model** trained on 1.2 million medical images with reports.
- It can **detect, classify, and segment** 82 clinical concepts (like tumors) across different imaging modalities using natural language prompts.

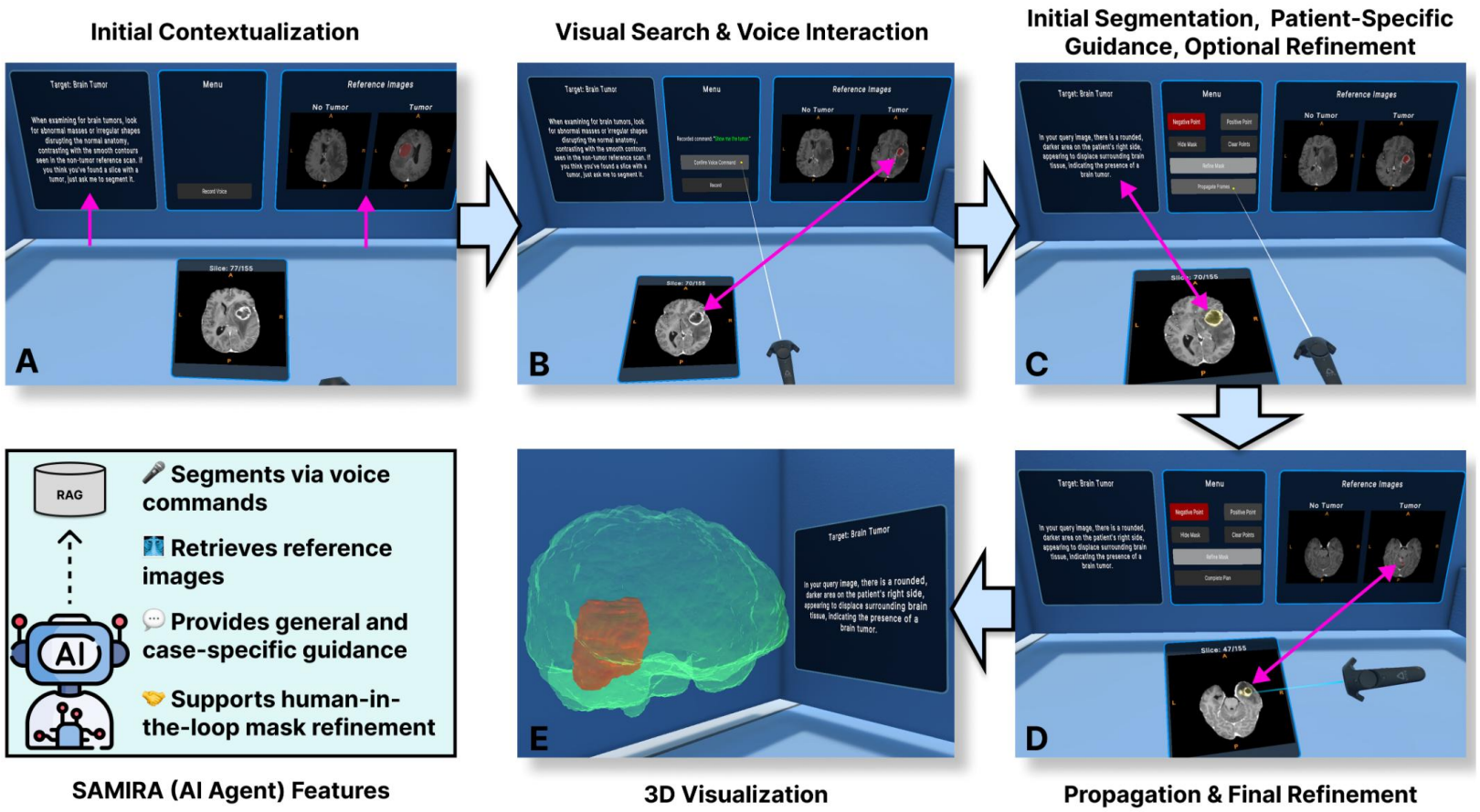
2) SAM2 (Segment Anything Model 2):

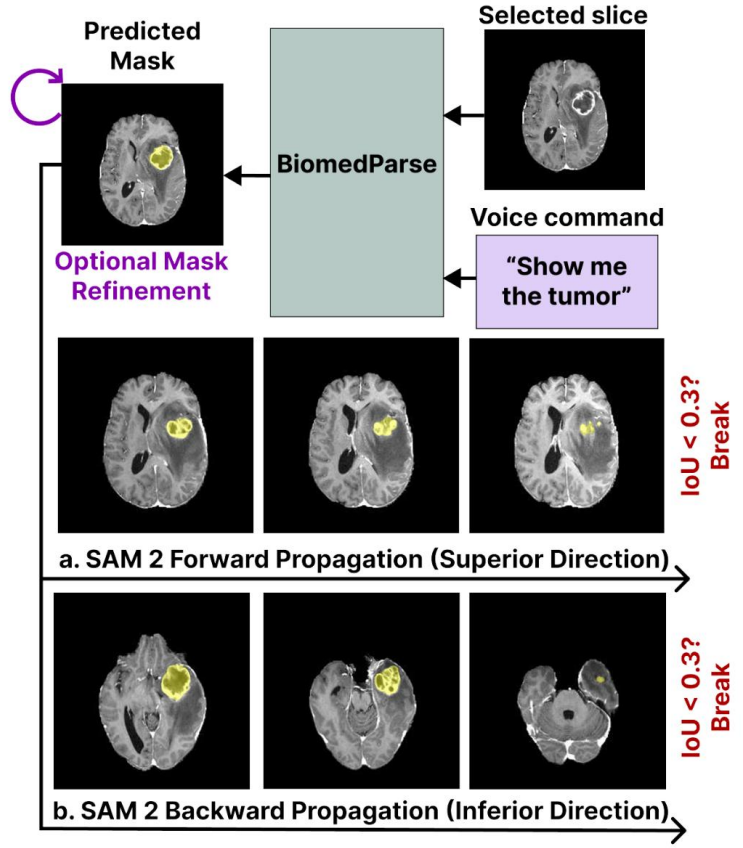
- A **model** designed for **interactive segmentation** using user prompts like points or boxes.
- It produces segmentation masks, which outline objects in images, and can **propagate these masks** across similar images or video frames

3) RAG (Retrieval-Augmented Generation):

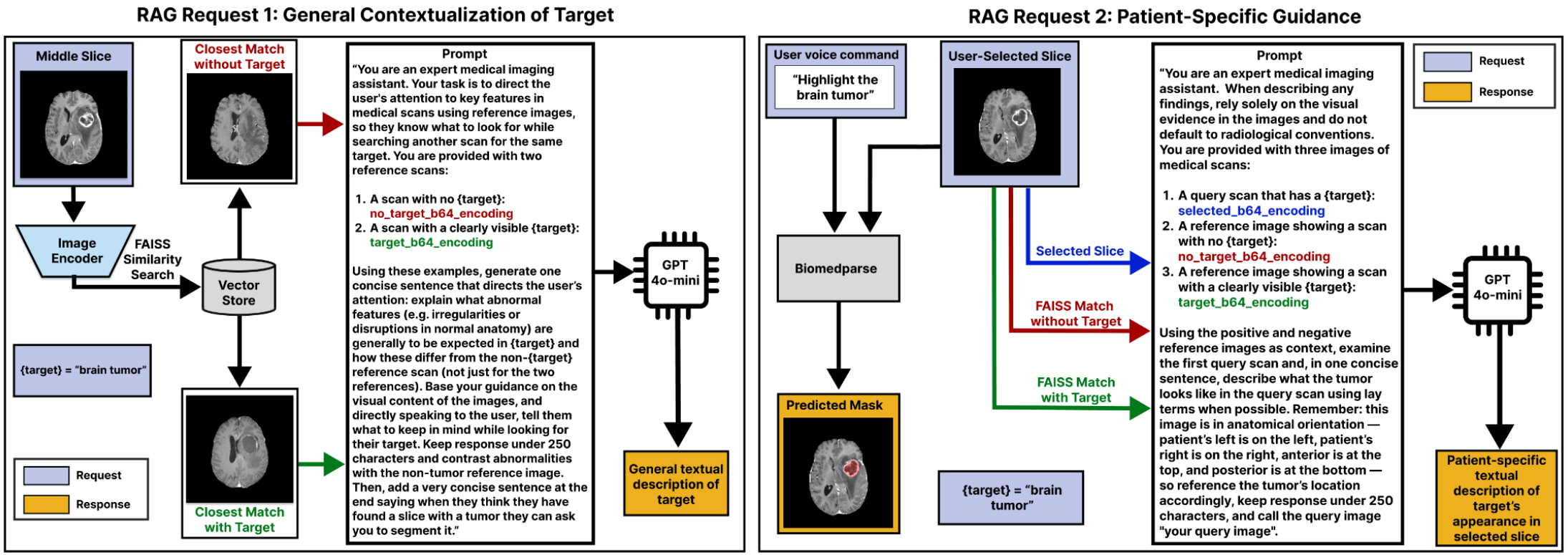
An AI framework that **searches a database** for relevant example images or slices, then **uses those** to generate more accurate, case-specific responses.

Workflow steps





RAG Pipeline

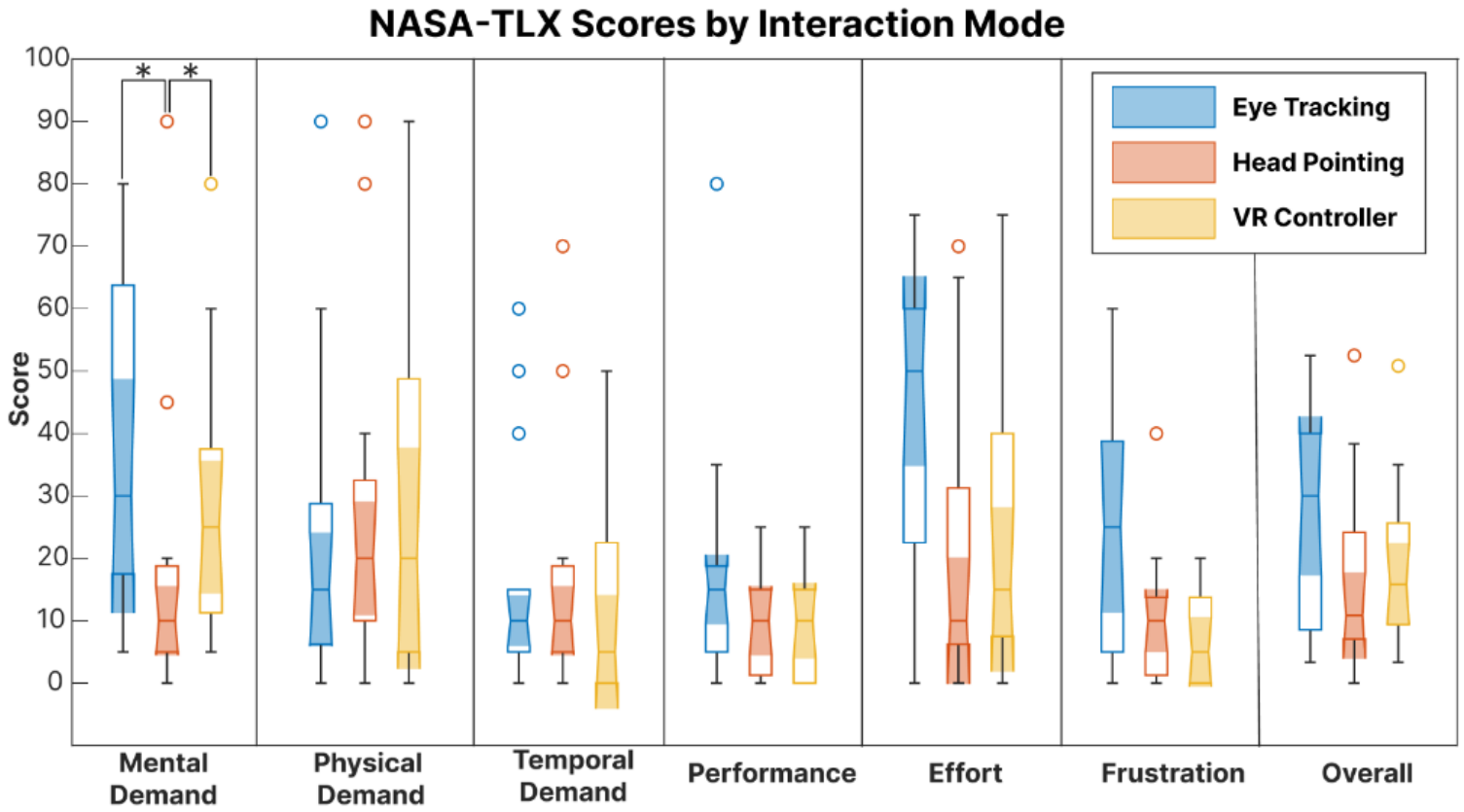


Results

Table 2 Comparison of interaction paradigms for segmentation refinement. Values are mean ± standard deviation. The best score is in bold. NASA-TLX is out of 100.

Interaction Mode	3D Dice (%)	Time (s)	NASA-TLX	Composite Score
Controller	**99.25 ± 0.25**	**220.3 ± 79.3**	18.8 ± 14.5	**0.51 ± 1.91**
Head Pointing	99.21 ± 0.30	248.8 ± 78.5	**16.8 ± 13.9**	0.20 ± 1.56
Eye Tracking	99.13 ± 0.46	251.1 ± 78.7	26.6 ± 15.4	–0.71 ± 1.77

Results


NASA-TLX Scores by Interaction Mode
Score
100
90
80
70
60
50
40
30
20
10
0
*
*
Eye Tracking
Head Pointing
VR Controller
Mental Demand
Physical Demand
Temporal Demand
Performance
Effort
Frustration
Overall

Conclusion

Key Findings:

- User-centered VR interaction paradigms are feasible and effective for medical segmentation
- AI assistance enhances confidence, understanding, and accuracy.
- Head Pointing offers a low mental effort alternative, ideal for single-hand use.

Implications:

- Potential to improve clinical workflows, training, and patient-specific analysis.
- Combines immersive visualization with intelligent guidance for better decision-making.

Future works:

- Further refine interaction techniques for real-world clinical settings.
- Expand to other medical applications and datasets.
- Explore integration with emerging VR and AI technologies.

Thank you!