APPENDIX

1 KNOWLEDGE BASE

1.1 Testing Knowledge Base Construction

We test the output of knowledge retrieval using different methods: LLM alone, LLM with RAG, and LLM with Web Search. The top nine regions of interest are derived through LLM summary from key papers on scientific visualization of flame combustion-related data in the past two decades, as listed in Tab. 1

Table 1: Papers considered to derive the regions of interest for flame combustion data.

Index	Paper	Year
1	Visualizing Multivariate Volume Data from Turbulent Combustion Simulations [1]	2007
2	Knowledge-Assisted Visualization of Turbulent Combustion Simulations [8]	2008
3	Application-Driven Compression for Visualizing Large-Scale Time-Varying Data [9]	2009
4	In Situ Visualization for Large-Scale Combustion Simulations [12]	2010
5	A preview and exploratory technique for large-scale scientific simulations [7]	2011
6	Dual space analysis of turbulent combustion particle data [10]	2011
7	Feature-Based Statistical Analysis of Combustion Simulation Data [2]	2011
8	Visual Analytics for Finding Critical Structures in Massive Time-Varying Turbulent-Flow Simulations [3]	2012
9	Tracking features in embedded surfaces: Understanding extinction in turbulent combustion [11]	2015
10	On-The-Fly Tracking of Flame Surfaces for the Visual Analysis of Combustion Processes [5]	2018
11	Radiative Transport Based Flame Volume Reconstruction from Videos [6]	2018
12	Visual Analytics of Combustion on Time-Varying Turbulent Flow [4]	2022

1.2 Knowledge Retrieval Evaluation

The following is the respective detailed output from various settings of the forager agent for knowledge retrieval:

LLM Alone:

1. Flame anchoring behavior
2. Variable relationships and stabilization
3. Main flame structure (stoichiometric surface)
4. Particle trajectory analysis
5. Turbulent combustion dynamics

LLM + RAG:

1. Flame Dynamics
2. Combustion Chemistry
3. Turbulence Modeling
4. Heat Transfer
5. Reaction Kinetics

LLM + Web Search:

In flame combustion simulations, several key subregions are critical for accurate modeling and analysis:

1. Reaction Zone
2. Preheat Zone
3. Flame Front
4. Turbulent Mixing Region
5. Boundary Layer
6. Soot Formation Zone
7. Extinction and Re-ignition Regions

These subregions are essential for understanding the complex interactions within combustion processes and are commonly investigated in computational fluid dynamics (CFD) studies.

2 DATASET DETAILS

We use five scientific datasets to validate the usability of SASAV. They are Abdomen, Chameleon, Miranca, Flame, and Richtmyer. The Abdomen is one scan from AbdomenAtlas 1.1 Mini Dataset, which is a fully-annotated abdominal CT dataset to date, including 9,262 CT volumes with annotations for 25 different anatomical structures. The Chameleon dataset is a CT scan of a chameleon. The Miranda is a time step of a density field in a simulation of the mixing transition in Rayleigh-Taylor instability. The Flame dataset is a simulated combustion 3D scalar field. The Richtmyer dataset is the entropy field (timestep 160) of Richtmyer-Meshkov instability simulation.

3 COLOR AND OPACITY MAPPING EXAMPLE

For the Abdoment (AbdomenAtlas 1.1 Mini) dataset, the detailed color and opacity suggested by the Transfer Function Designer (TFD) together with the explanation are listed in Tab. 2.

REFERENCES

[1] H. Akiba, K.-l. Ma, J. H. Chen, and E. R. Hawkes. Visualizing multivariate volume data from turbulent combustion simulations. *Computing in Science Engineering*, 9(2):76–83, 2007. doi: 10.1109/MCSE.2007.42 1

[2] J. C. Bennett, V. Krishnamoorthy, S. Liu, R. W. Grout, E. R. Hawkes, J. H. Chen, J. Shepherd, V. Pascucci, and P.-T. Bremer. Feature-based statistical analysis of combustion simulation data. *IEEE Transactions on Visualization and Computer Graphics*, 17(12):1822–1831, 2011. doi: 10.1109/TVCG.2011.199 1

[3] K. P. Gaither, H. Childs, K. W. Schulz, C. Harrison, W. Barth, D. Donzis, and P.-K. Yeung. Visual analytics for finding critical structures in massive time-varying turbulent-flow simulations. *IEEE Computer Graphics and Applications*, 32(4):34–45, 2012. doi: 10.1109/MCG.2012.63 1

[4] X. Han, Y. An, G. Shan, J. Liu, and B. Yang. Visual analytics of combustion on time-varying turbulent-flow. *Advances in Aerodynamics*, 4(1):9, 2022. 1

[5] T. Oster, A. Abdelsamie, M. Motejat, T. Gerrits, C. Rössl, D. Thévenin, and H. Theisel. On-the-fly tracking of flame surfaces for the visual analysis of combustion processes. *Computer Graphics Forum*, 37(6):358–369, 2018. doi: 10.1111/cgf.13331 1

[6] L. Shen, D. Zhu, S. Nadeem, Z. Wang, and A. E. Kaufman. Radiative transport based flame volume reconstruction from videos. *IEEE Transactions on Visualization and Computer Graphics*, 24(7):2209–2222, 2018. doi: 10.1109/TVCG.2017.2712688 1

[7] A. Tikhonova, H. Yu, C. D. Correa, J. H. Chen, and K.-L. Ma. A preview and exploratory technique for large-scale scientific simulations. In *Proceedings of the 11th Eurographics Conference on Parallel Graphics and Visualization*, EGPGV '11, p. 111–120. Eurographics Association, Goslar, DEU, 2011. 1

[8] C. Wang, H. Yu, and K.-L. Ma. Knowledge-assisted visualization of turbulent combustion simulations. *extended abstract, KAV*, 7. 1

[9] C. Wang, H. Yu, and K.-L. Ma. Application-driven compression for visualizing large-scale time-varying data. *IEEE Computer Graphics and Applications*, 30(1):59–69, 2010. doi: 10.1109/MCG.2010.3 1

[10] J. Wei, H. Yu, R. W. Grout, J. H. Chen, and K.-L. Ma. Dual space analysis of turbulent combustion particle data. In *2011 IEEE Pacific Visualization Symposium*, pp. 91–98, 2011. doi: 10.1109/PACIFICVIS.2011.5742377 1

[11] W. Widanagamaachchi, J. Chen, P. Klacansky, V. Pascucci, H. Kolla, A. Bhagatwala, and P.-T. Bremer. Tracking features in embedded surfaces: Understanding extinction in turbulent combustion. In *2015 IEEE 5th Symposium on Large Data Analysis and Visualization (LDAV)*, pp. 9–16, 2015. doi: 10.1109/LDAV.2015.7348066 1

[12] H. Yu, C. Wang, R. W. Grout, J. H. Chen, and K.-L. Ma. In situ visualization for large-scale combustion simulations. *IEEE Computer Graphics and Applications*, 30(3):45–57, 2010. doi: 10.1109/MCG.2010.55 1

Table 2: Color and opacity suggestion by the Transfer Function Designer (TFD) from the output of the Sematric Analyzer (SA). SA decides if a structure can be recognized. TFD assigns positive opacity to recognizable structures and zero opacity to unrecognizable structures.

Label	Structure	Output	Alpha
3	Kidney (left)	[0.82, 0.24, 0.28] — A muted reddish tone resembles the natural color of kidney tissue and contrasts well with surrounding organs and grayscale CT structures, making the kidney easy to distinguish in 3D visualization.	1.0
4	Kidney (right)	[0.82, 0.18, 0.24] — A slightly muted red is commonly used for kidneys because it resembles the organ's natural vascular tissue color while remaining distinct from nearby organs (like liver or spleen) and provides strong contrast against typical grayscale CT backgrounds.	1.0
5	Liver	[0.80, 0.30, 0.25] — A warm reddish-brown tone closely resembles the natural appearance of liver tissue and stands out clearly against surrounding organs (often grayish in CT), making the structure easy to recognize in 3D visualization.	0.4
8	Spleen	[0.78, 0.16, 0.20] — A deep reddish tone is commonly used for the spleen in medical visualization because the organ is highly vascularized and naturally appears dark red in anatomy, which also helps it stand out clearly from nearby organs like the liver and stomach in CT renderings.	1.0
9	Stomach	[0.88, 0.35, 0.32] — A warm reddish-pink color works well for the stomach because it resembles soft tissue coloration while still contrasting clearly with typical liver (brown), spleen (purple), and pancreas (yellowish) colors in abdominal CT visualizations.	0.5
12	Bladder	[0.95, 0.85, 0.25] — A soft yellow helps the bladder stand out clearly against surrounding reddish organs (like liver or muscle) and darker tissues in CT visualizations while remaining bright and visually intuitive for a fluid-filled structure.	1.0
14	Colon	[0.85, 0.35, 0.30] — A warm reddish color highlights the colon clearly against typical CT visualizations (which often use cool or neutral tones for surrounding tissues), making the organ easy to distinguish while still resembling natural anatomical coloring.	0.2
16	Esophagus	[0.85, 0.35, 0.30] — A warm reddish tone works well for the esophagus because it resembles soft-tissue color and visually distinguishes the thin tubular structure from nearby organs (often colored in cooler tones like liver/bone) in CT 3D renderings.	1.0
17	Femur (left)	[0.92, 0.89, 0.78] — A light warm ivory tone closely resembles the natural appearance of cortical bone in medical renderings, providing good contrast against surrounding soft tissues while remaining visually intuitive for anatomical interpretation.	1.0
18	Femur (right)	[0.92, 0.86, 0.76] — A slightly warm off-white/beige resembles the natural color of cortical bone and provides strong contrast against soft tissue and darker anatomical structures in CT volume visualizations.	1.0
20	Intestine	[0.93, 0.64, 0.37] — A warm soft orange resembles the natural color tone of intestinal tissue and provides strong contrast against surrounding organs (often rendered in reds, purples, or grays) while remaining visually comfortable for 3D CT visualization.	0.3
21	Lung (left)	[0.70, 0.85, 1.00] — A soft light-blue is commonly used for lungs because it visually suggests air/lightness and contrasts clearly with typical organ colors (reds, yellows, browns) in CT anatomical visualizations.	0.3
22	Lung (right)	[0.65, 0.85, 1.0] — A light bluish color resembles the visual association of air and oxygen, helping the lungs appear distinct from surrounding soft tissues (usually reddish/pink) while remaining soft and easy to perceive in medical CT visualizations.	0.3