Supplementary Materials

AIvaluateXR: An Evaluation Framework for on-Device AI in XR with Benchmarking Results

Abstract

This document presents the supplementary for AIvaluateXR Paper. For mor details please visit our project website: `nanovis.org/AIvaluateXR.html`.

Contents

1 Analysis of On-Device LLMs on Relevant Datasets

As discussed in Section 6 of the main paper, we compared the accuracy and time efficiency of our on-device LLMs against representative cloud-based LLM approaches for XR applications [2, 1]. In this supplementary material, we provide additional details about the two interactive applications and datasets used in the evaluation [2, 1].

1.1 Evaluation Metrics for Interactive Applications Experiments

As described in Section 6.1 (*Evaluation on Interactive Applications*) of the main paper, we evaluate the model's performance on the GeoVis [1] queries using two key metrics: **Formatting** and **Precision**.

Formatting assesses whether the model's response adheres to the expected structure:

- **100%**: The output is in the expected format `[latitude, longitude]`.
- **50%**: The format is incorrect, but the first two numeric values can be parsed as valid coordinates.
- **0%**: The output cannot be parsed to retrieve valid coordinates.

Precision measures the spatial accuracy of the predicted coordinates:

- **100%**: The location is close to the center of the target area.
- **90%**: The location is within the area, closer to the center than the edge.
- **80%**: The location is within the area, near the edge.
- **50%**: The location is outside the area, but the full area is visible at the appropriate zoom level.
- **20%**: The location is outside the area, and only part of it is visible.
- **0%**: The target area is not visible.

Note: If the formatting score is 0%, the precision score is also set to 0% by default, as the response cannot be reliably interpreted.
For the VOICE [2], we have binary answers so accuracy will be 100 % or 0 %.

Here are the prompt details for each experiment.

1.2 VOICE Query List with Prompts

Description of the VOICE Dataset [2]
VOICE (Visual Oracle for Interaction, Conversation, and Explanation) leverages cloud-based LLMs (e.g., ChatGPT) for conversational, interactive exploratory visualization in biological domains. It employs a pack-of-bots architecture, fine-tuning, and prompt engineering to generate coherent responses and flythrough visualizations. Users can issue natural language or voice commands to manipulate 3D molecular models in real time with high accuracy and low latency. The authors introduced a specialized dataset containing queries and prompts to enable this interactive experience. We use this dataset to evaluate the latency and accuracy of on-device LLMs by providing input queries and prompts, comparing the generated responses, and recording inference latency.

VOICE Queries List
The following *VOICE* [2] queries and associated prompts from the XR dataset were used for evaluating LLM interaction performance. Each entry consists of a natural language query and the system prompt that guides model interpretation.

1. Can we get to the post-fusion conformation?
2. Can you show me the closed conformation and then fly to the post-fusion conformation?
3. Can you show me the DNA?
4. Can you show me the RNA?
5. Fly to the next envelope protein.
6. Show me the nucleocapsid.
7. Can you go back to the initial view?
8. Can you show me the infected cell?

9. Can you rotate the view to 45 degrees?
10. Can you zoom in?
11. Can you zoom out to see the full model?
12. Can I see it from the side?
13. Can you rotate it upside down?
14. Can you rotate this virus to see it from the top?
15. Can you zoom out?
16. Zoom in a lot.
17. Zoom in please.
18. Can you rotate this virus to see it from the bottom?
19. I want to return to what I was just looking at
20. Take me back to my last selection
21. Go back one step
22. Take me up a level in the structure
23. Can you display the higher hierarchy?
24. Can you show me a more general view?
25. I want to start fresh
26. Go back to where we began
27. Reset everything and start over

Prompt for the *VOICE* [2] queries

Prompt 1 (for navigation tasks i.e., queries 1 to 17):
I will give you a prompt. The prompt will ask to see something/something to be shown. You will help me determine what the prompter wants to see. Here is how you will reply: follow these step-by-step instructions exactly. If a statement holds true, return the defined answer and don't consider the later statements (hence why they are marked as "else")
If the case mentions a "keyword" (not restricted to but generally scientific terminology), says explicitly

what it wants to see, reply with “0”.
Else, if the prompt explicitly wants to return to the initial / reset position, reply with “3”
Else, if the prompt implies moving back to the last object, reply with “1”
Else, if the prompt implies moving up a level in a hierarchy, reply with “2”
Else reply with “0”
These are a few examples:
Prompt: “Can you show me the Capsid Protein” ⇒ “0”
Prompt: “I want to see the moon” ⇒ “0”
Prompt: “Show me the last thing again” ⇒ “1”
Prompt: “Go back please” ⇒ “1”
Prompt: “Go back to Immunoglobulin C”⇒ “0”
Prompt: “Can you show me an overview” ⇒ “2”
Prompt: “Move up one level please” ⇒ “2”
Prompt: “Start the journey over” ⇒ “3”
Prompt: “Reset my position” ⇒ “3”
Prompt: “Go back to the origin” ⇒ “3”
To summarize, your possible completions are: “0”, “1”, “2”, “3”. Act as an intent classifier and do your best in determining what the prompt author wants to see.
Your completion should always ONLY be that one digit. No explanation, no preface. Don’t add any other characters in your replies.

Prompt 2 (for visual tasks i.e., queries 18 to 27):
When I give you a prompt, I want you to answer the following: What rotation does the prompt imply (in the form of a yaw pitch roll vector), and what degree of zoom does it imply?
Here are the rules you MUST follow without exceptions:
For every single prompt, 0,0,-1 is the initial view direction.
Prompts may reference previous prompts (ex., undo the previous transformation). You should consider previous prompts when explicitly referenced; however, don’t consider previous completions.
Interpret inexact values like “a little” as you see fit, but stay consistent (a little will always be less than a lot).
We are looking at the object from the front, meaning the right, left, top, and bottom are adjacent, and the back is on the other side.
The return answer format is ALWAYS: $\{[zoom\ multiplier], [yaw], [pitch], [roll]\}$
If there is no implied transformation, just use the default values $\{1, 0, 0, 0\}$
EXAMPLES:
“Show me xyz from the top” = $\{1, 0, 90, 0\}$
“make it larger” = $\{2, 0, 0, 0\}$
“tell me something about the capsid” = $\{1, 0, 0, 0\}$
“I want to see a close-up from the right side” = $\{3, 90, 0, 0\}$
“I want to see the right side” = $\{1,90,0,0\}$
“how does the bottom look like” = $\{1, 0, 90, 0\}$
s̈how me the left side of the lipid” = 1,-90,0,0
“show me the back” = 1, 180, 0, 0
“what is a xyz” = $\{1, 0, 0, 0\}$
ONLY reply in the $[zoommultiplyer], [yaw], [pitch], [roll]$ format from now on!

Taken from the *VOICE* [2].

1.3 Conversational GeoVisualization Queries [1] List with Prompts

Description of the Conversational GeoVisualization Dataset [1]
This dataset uses (Cloud-based) LLMs to enable conversational question answering on scientific visualizations focused on global Geo-spatial data. To overcome the LLM's limitations in handling visual context, their approach extracts key visual and descriptive features from rendered visualizations and encodes them into a structured, compact textual representation. This allows the LLM to reason about visual content without fine-tuning. We use this dataset to assess the performance of on-device LLMs in answering context-aware Geo-spatial queries.

Conversational GeoVisualization Queries List with Prompts
To evaluate the capabilities of LLMs when acting as conversational navigation agents, we used the following prompt, concatenated with each of the queries listed below:
Prompt (for queries 1 to 30):
You are a 3D visualization tool in charge of navigation. You will be asked about one location. Based on this request, reply with latitude and longitude in this WGS84 form: [lat,long]. Do no use any spaces in your response. Example: [54.25,15.24]. Be as accurate as possible. Do not include anything else in your response, just the raw coordinates.

Queries List:

1. Take me to Los Angeles
2. Where is Berlin?
3. Show me Shanghai
4. Where is Cairo?
5. Show me Syracuse
6. Where is Mexico?
7. Where is Japan?
8. Show me Andorra
9. Take me to Uruguay
10. Show me Slovakia
11. Where is the Atlantic Ocean?
12. Where is the Mediterranean Sea?
13. Take me to the South China Sea
14. Show me the Red Sea

15. Show me the Hudson Bay
16. Take me to the Nile
17. Where is the Loire?
18. Where is the Volga?
19. Where is the Yangtze?
20. Show me the Guadiana
21. Take me to the Alps
22. Show me the Himalayas
23. Show me the Andes
24. Where are the Rocky Mountains
25. Take me to the Caucasus Mountains
26. Where is the Grand Canyon?
27. Take me to the Pyramids of Giza
28. Take me to the Great Wall of China
29. Take me to the Palace of Versailles
30. Show me the Silfra fissure

Taken from GeoVisualization Dataset [1].
table 1 shows latency of different model in response to the queries.

Model	VIVO Time (s)	ML2 Time (s)
m_8	21.094 ± 11.603	26.290 ± 15.881
m_{11}	26.348 ± 13.638	24.436 ± 14.604
m_{13}	36.438 ± 24.749	55.912 ± 37.799
m_{15}	86.691 ± 53.835	61.206 ± 38.037
m_{17}	111.270 ± 32.320	43.224 ± 26.711

Table 1: Comparison of mean (μ) and standard deviation (σ) of time (in seconds) on VIVO and Magic Leap 2 (ML2) for selected models.

2 Consistency Results

Tables 2 and 3 present performance consistency results for the four device. The values are the mean, standard deviation, Coefficient of Variance and range for 20 runs of each model device pair.

3 Quantitative Results from PP and TG Tests

table 4 presents quantitative results form PP and TG tests. The values reported are the mean (μ), standard deviation (σ), and coefficient of variation (CV). Note that the CV is computed across different parameters (PP/TG = 64, 128, 256, 512, 1024), which reflects the impact of parameter variation (PP/TG = 64, 128, 256, 512, 1024) on results.

Table 2: Performance consistency results for the four devices: Magic Leap 2 (ML2), Meta Quest 3 (MQ3), Vivo X100 Pro (Vivo), and Apple Vision Pro (VPro for CPU and VPro* for GPU). Processing speed was calculated from 20 runs for each model-device pair. The results are reported in terms of the mean (μ) speed (t/s) of 20 runs, its standard deviation (σ), coefficient of variation (CV %), and the range of values: $[min, max]$.

	D	Metric	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	m14	m15	m16	m17
PP Results	ML2	$\mu\pm\sigma$	17.01±4.67	13.99±4.57	10.38±2.74	20.82±0.87	16.81±0.07	13.16±0.47	14.22±0.31	12.63±0.38	13.12±0.29	11.45±0.18	7.92±0.05	7.37±0.19	2.20±0.03	3.14±0.08	1.15±0.05	1.52±0.09
		Range	[10.25, 22.24]	[7.47, 18.75]	[6.27, 13.32]	[18.28, 21.36]	[16.68, 16.95]	[12.04, 13.51]	[12.98, 14.36]	[11.73, 12.87]	[12.00, 13.27]	[10.93, 11.62]	[7.73, 7.99]	[6.73, 7.54]	[2.14, 2.23]	[2.98, 3.27]	[1.10, 1.33]	[1.36, 1.72]
		CV (%)	27.46	32.66	26.41	4.16	0.43	3.56	2.19	2.98	2.22	1.57	0.66	2.54	1.35	2.41	4.70	5.99
	MQ3	$\mu\pm\sigma$	21.00±2.16	25.61±0.10	21.33±1.47	16.33±1.84	10.99±1.65	11.37±1.80	13.74±1.72	10.24±1.34	11.07±0.61	8.17±2.19	7.12±0.04	6.09±0.32	2.65±0.21	2.37±0.12	1.57±0.06	2.04±0.09
		Range	[17.42, 23.59]	[25.03, 25.77]	[14.79, 23.12]	[13.66, 18.72]	[7.62, 13.89]	[7.91, 13.86]	[10.88, 16.74]	[7.79, 11.87]	[7.19, 11.28]	[4.31, 13.05]	[5.96, 7.17]	[5.64, 6.61]	[2.10, 2.92]	[2.12, 2.57]	[1.44, 1.68]	[1.85, 2.34]
		CV (%)	10.27	0.37	6.89	11.24	14.99	15.80	12.51	13.06	5.55	26.81	0.56	5.23	7.84	5.24	3.93	4.23
	Vivo	$\mu\pm\sigma$	15.70±2.88	15.90±2.76	14.98±2.98	13.77±2.13	8.66±0.80	8.27±1.08	10.35±1.10	8.48±1.08	8.82±1.61	11.14±1.21	4.47±0.19	4.35±0.23	2.44±0.07	4.28±0.44	2.61±0.12	5.89±0.47
		Range	[11.16, 25.55]	[9.39, 17.81]	[10.52, 24.36]	[10.65, 15.55]	[7.32, 9.34]	[5.88, 9.53]	[8.27, 11.05]	[7.19, 11.72]	[5.97, 11.60]	[9.13, 12.04]	[3.98, 4.92]	[3.80, 4.86]	[2.27, 2.52]	[2.61, 4.80]	[2.23, 2.76]	[5.17, 6.59]
		CV (%)	18.33	17.38	19.88	15.50	9.26	13.07	10.61	12.75	18.31	10.84	4.24	5.39	3.00	10.30	4.55	7.92
	VPro	$\mu\pm\sigma$	31.03±2.20	41.22±3.04	33.11±1.35	32.96±1.21	20.62±0.56	19.34±0.64	23.14±1.09	19.52±0.73	19.56±0.60	29.22±1.22	9.15±0.40	9.06±0.28	5.52±0.21	7.44±0.40	5.01±0.26	14.05±0.56
		Range	[29, 34.98]	[39.44, 47.03]	[32.98, 34.79]	[32.15, 35.93]	[20.19, 21.63]	[19.09, 20.43]	[22.65, 24.92]	[19.04, 20.98]	[19.15, 20.74]	[28.58, 31.25]	[8.77, 9.67]	[8.88, 9.77]	[5.4, 5.93]	[7.21, 8.13]	[4.83, 5.47]	[13.63, 15.02]
		CV (%)	7.08	7.37	4.07	3.68	2.72	3.31	4.69	3.72	3.05	4.17	4.41	3.12	3.73	5.38	5.16	3.99
	VPro*	$\mu\pm\sigma$	378±3.7	432±3.7	382±4.8	374±4.6	241±1.5	230±2.3	237±2.0	218±5.8	229±4.3	264±3.4	136±9.9	128±12.5	135±2.3	134±2.3	133±1.4	130±6.6
		Range	[369, 384]	[424, 439]	[367, 389]	[363, 380]	[237, 243]	[225, 234]	[232, 240]	[198, 223]	[221, 237]	[255, 268]	[109, 145]	[100, 143]	[130, 138]	[129, 136]	[129, 134]	[109, 135]
		CV (%)	1.0	0.9	1.2	1.2	0.6	1.0	0.8	2.6	1.9	1.3	7.3	9.8	1.7	1.7	1.1	5.1
TG Results	ML2	$\mu\pm\sigma$	9.14±1.92	8.22±1.92	6.53±1.77	8.43±0.07	11.69±0.04	8.52±0.02	7.88±0.01	6.92±0.01	6.25±0.10	5.05±0.02	5.90±0.03	5.59±0.14	2.05±0.01	2.84±0.06	0.24±0.01	0.21±0.01
		Range	[6.92, 11.25]	[5.69, 10.27]	[4.41, 8.39]	[8.33, 8.55]	[11.61, 11.74]	[8.48, 8.55]	[7.85, 7.91]	[6.89, 6.95]	[5.95, 6.31]	[5.01, 5.08]	[5.83, 5.94]	[5.13, 5.73]	[2.04, 2.06]	[2.77, 2.97]	[0.22, 0.27]	[0.19, 0.23]
		CV (%)	20.95	23.36	27.10	0.77	0.32	0.23	0.18	0.21	1.62	0.39	0.48	2.49	0.32	2.24	5.32	5.20
	MQ3	$\mu\pm\sigma$	12.37±1.45	14.19±0.30	11.58±0.98	9.47±1.29	9.00±1.12	8.57±0.70	9.30±0.51	7.61±0.65	8.30±0.26	5.47±0.10	5.62±0.23	5.01±0.24	2.95±0.21	3.06±0.05	1.67±0.15	3.38±0.09
		Range	[9.11, 13.89]	[13.52, 14.67]	[7.77, 12.80]	[7.64, 10.87]	[7.38, 10.39]	[6.85, 9.55]	[7.66, 9.86]	[6.13, 8.24]	[7.40, 8.53]	[5.19, 5.61]	[5.26, 6.17]	[4.61, 5.57]	[2.23, 3.40]	[2.96, 3.13]	[1.40, 1.84]	[3.11, 3.53]
		CV (%)	11.73	2.11	8.43	13.65	12.50	8.17	5.52	8.51	3.18	1.87	4.09	4.83	7.27	1.68	8.93	2.80
	Vivo	$\mu\pm\sigma$	9.39±0.98	9.85±0.42	9.44±1.22	8.66±0.30	6.91±0.08	6.83±0.22	8.21±0.12	6.84±0.61	6.61±0.71	7.51±0.11	3.70±0.05	3.59±0.11	2.09±0.01	3.71±0.04	2.27±0.02	4.72±0.17
		Range	[8.54, 13.35]	[9.17, 10.67]	[7.58, 13.86]	[7.97, 9.26]	[6.75, 7.00]	[6.39, 7.21]	[7.52, 8.35]	[4.37, 7.11]	[5.70, 7.68]	[7.25, 7.64]	[3.60, 3.77]	[3.41, 3.71]	[2.06, 2.11]	[3.65, 3.79]	[2.23, 2.31]	[4.48, 5.00]
		CV (%)	10.46	4.22	12.91	3.48	1.13	3.25	1.48	8.92	10.75	1.41	1.27	3.04	0.68	1.15	1.06	3.61
	VPro	$\mu\pm\sigma$	17.90±1.00	21.15±0.82	18.92±0.74	17.71±0.50	15.72±0.37	14.76±0.26	16.63±0.49	13.93±0.50	14.07±0.19	12.03±0.21	7.20±0.21	7.22±0.17	4.53±0.17	6.24±0.27	4.17±0.18	10.22±0.30
		Range	[16.59, 19.72]	[19.4, 22.49]	[17.31, 19.8]	[16.89, 18.77]	[14.88, 16.36]	[14.32, 15.28]	[15.77, 17.56]	[12.89, 14.99]	[13.71, 14.49]	[11.26, 12.28]	[6.74, 7.52]	[6.98, 7.7]	[4.11, 4.93]	[5.52, 6.72]	[3.86, 4.52]	[9.82, 10.93]
		CV (%)	5.58	3.86	3.89	2.84	2.35	1.73	2.97	3.61	1.35	1.77	2.95	2.41	3.84	4.35	4.31	2.91
	VPro*	$\mu\pm\sigma$	23.1±0.1	27.5±0.1	23.5±0.1	22.6±0.1	23.3±0.1	19.6±0.8	19.9±0.1	18.0±0.4	17.7±0.2	14.6±0.2	13.7±0.8	12.5±1.0	14.5±0.3	15.1±0.2	15.2±0.2	14.7±0.4
		Range	[23, 23.3]	[27.1, 27.7]	[23.2, 23.8]	[22.3, 22.7]	[23.0, 23.5]	[19.2, 22.8]	[19.8, 20.2]	[16.8, 18.5]	[17.2, 17.9]	[14.3, 14.9]	[11.4, 14.6]	[10.2, 13.9]	[13.9, 14.9]	[14.6, 15.4]	[14.9, 15.4]	[13.5, 15.1]
		CV(%)	0.4	0.5	0.5	0.3	0.6	4.3	0.6	2.1	1.0	1.1	5.8	8.3	1.9	1.4	1.1	2.5

Table 3: Performance consistency results for m_1.

Test	Metric	ML2	Vivoo	Meta Q3	VisPro	VisPro*
PP	$\mu \pm \sigma$	51.68 ± 4.22	62.68 ± 16.64	39.32 ± 15.66	292.48 ± 12.87	1603.59 ± 56.12
	Range	[43.62, 54.59]	[21.63, 72.04]	[24.25, 72.59]	[288.5, 312.1]	[1516.9, 1696.1]
	CV (%)	8.17	26.55	39.82	4.40	3.50
TG	$\mu \pm \sigma$	20.26 ± 0.05	22.16 ± 0.90	16.76 ± 3.76	41.94 ± 1.77	42.94 ± 0.80
	Range	[20.17, 20.34]	[20.54, 23.41]	[11.61, 22.53]	[38.90, 46.37]	[41.76, 44.35]
	CV (%)	0.23	4.05	22.44	4.21	1.87

Table 4: Processing speed (tokens per second) of 17 models across four XR devices for varying prompt/string lengths (PP/TG), along with the mean (μ), standard deviation (σ), and coefficient of variation (CV). Here, $*$ denotes a single error occurrence, while $*2$ and $*3$ indicate two and three repetitions due to errors, respectively. Note that the CV is computed across different parameters (PP/TG = 64, 128, 256, 512, 1024), meaning a higher CV does not indicate an error (as in CV for 5-run speed) but rather reflects the impact of parameter variation on results.

Test	Device	$PP\|TG$	Processing speed for each model																	Errors
		Metrics	m_1	m_2	m_3	m_4	m_5	m_6	m_7	m_8	m_9	m_{10}	m_{11}	m_{12}	m_{13}	m_{14}	m_{15}	m_{16}	m_{17}	Count
PP Test	Magic Leap 2	64	68.86	25.23	20.80	14.28	24.94	17.46	14.06	15.03	13.37	14.17	12.50	8.10	7.69	2.24	3.28	2.16	6.82	0
		128	69.07	24.97	20.78	14.25	24.56	17.43	13.93	14.96	13.46	14.04	12.43	8.07	7.66	2.23	3.27	2.16	6.80	0
		256	67.55	24.46	20.35	14.10	24.26	17.02	13.65	14.72	13.13	13.81	12.13	7.96	7.55	2.23	3.26	2.15	6.75	0
		512	64.58	23.31	19.52	13.67	23.04	16.22	13.19	14.08	12.71	13.33	11.77	7.75	7.37	2.22	3.23	2.14	6.55	0
		1024	62.20	22.25	18.73	13.30	22.01	15.57	12.74	13.59	12.25	12.87	11.43	7.54	7.15	2.20	3.20	2.13	6.45	0
		μ	66.85	24.24	20.04	14.14	23.76	16.74	13.91	14.88	13.38	13.84	12.45	7.88	7.47	2.23	3.25	2.15	6.73	-
		σ	2.53	1.19	0.80	0.33	1.18	0.74	0.65	0.58	0.42	0.55	0.40	0.23	0.16	0.01	0.03	0.01	0.14	-
		CV (%)	3.78	4.91	4.00	2.34	4.96	4.42	4.66	3.91	3.14	3.95	3.21	2.92	2.14	0.45	1.02	0.54	2.08	-
	Meta Quest 3	64	118.63	26.39	28.63	23.14	20.53	14.41	13.77	16.74	11.95	11.60	18.78	6.37	6.27	3.58	5.62	2.83	6.83*	1
		128	121.99	26.17	27.04	21.51	19.48	13.08	12.15	14.94	10.52	10.29	16.34	5.98	5.71	3.06*	5.00	2.22	5.98	1
		256	113.77	23.54	24.48	18.99	17.98	12.03	10.93	13.44	9.46	9.57	11.84	4.78	4.56*	2.35	3.74	2.14	4.95*	2
		512	101.85	21.34	21.97	17.44	16.02	9.49	8.39	10.33*	7.05	7.87*	11.01	4.46	3.82	2.19	3.37	1.92	4.06	2
		1024	88.49	19.05	19.77	12.67	12.66*	7.81	7.19	8.68	6.24	6.38	9.58	4.04	3.65	2.18	3.28	1.87	3.82	1
		μ	108.95	23.98	24.77	18.75	17.33	11.35	10.09	12.83	9.05	9.62	13.91	5.33	4.92	2.63	4.80	2.20	4.93	-
		σ	12.77	3.00	3.65	3.54	2.92	2.46	2.37	3.07	1.84	2.02	3.55	0.96	0.92	0.48	0.86	0.14	1.07	-
		CV (%)	11.72	12.51	14.72	18.90	16.84	21.63	23.53	23.94	20.34	21.01	25.55	18.06	18.67	18.25	17.96	6.28	21.68	-
	Vivo X100s Pro	64	94.55	16.43	18.40	15.95	15.90	9.16	9.36	10.83	9.00	9.23	12.02	4.80	4.74	2.56	4.64	2.80	6.43	0
		128	103.07*	16.27	18.35	16.05	15.94	9.17	9.35	10.78	9.07	12.47*	12.05	4.85	4.75	2.57	4.63	2.81	6.46	2
		256	109.44	16.48	18.08	15.92	15.83	9.11	9.25	10.75	8.97	11.89*	11.70	4.80	4.74	2.55	4.61	2.80	6.42	1
		512	101.79	15.86	17.57	15.45	15.37	8.90	9.09	10.47	8.50	8.25	11.12	4.76	4.68	2.53	4.44	2.78	6.29	0
		1024	96.52	15.25	16.73	14.85	14.73	8.66	8.83	10.15	8.05	8.10	10.38	4.68	4.40	2.50	4.32	2.75	5.93	0
		μ	101.07	16.05	17.83	15.65	15.56	9.00	9.18	10.60	8.72	10.59	11.45	4.78	4.67	2.54	4.53	2.79	6.31	-
		σ	5.87	0.48	0.68	0.48	0.47	0.22	0.22	0.30	0.44	1.69	0.84	0.06	0.13	0.03	0.12	0.03	0.19	-
		CV (%)	5.81	2.99	3.82	3.05	3.02	2.44	2.37	2.83	5.07	15.96	7.32	1.26	2.83	1.06	2.57	1.06	3.01	-
	Apple Visions Pro	64	306.88	34.81	48.79	36.14	36.69	21.89	20.46	25.38	21.07	20.04	31.46	10.01	9.54	5.81	8.41	5.23	15.05	0
		128	313.30	34.87	48.13	35.88	36.64	21.39	20.09	25.08	20.74	20.37	31.58	9.74	9.32	5.50	7.66	4.87	14.18	0
		256	309.84	32.62	44.46	32.44	33.59	19.68	19.27	22.95	18.66	18.36	28.49	9.05	8.47	5.35*	7.35	4.75*	13.32	2
		512	287.89	29.74	40.52	29.57	30.63	18.39*	17.89*	21.16	17.21*	17.27*	25.72*	8.70*	8.37*	5.24	7.12	4.66	12.83	7
		1024	254.60	29.28*	38.45*	28.39*	30.10*	17.55*	17.15*	20.29	16.50	16.33	24.10	8.29	8.04	5.04	6.17	4.57	12.31*	7
		μ	294.70	32.26	44.07	32.08	33.74	19.79	18.97	23.57	18.44	18.67	28.67	9.56	8.75	5.39	7.54	4.82	13.54	–
		σ	21.77	2.61	3.79	3.27	2.79	1.78	1.32	2.11	2.07	1.85	3.06	0.67	0.51	0.21	0.92	0.21	1.02	–
		CV (%)	7.39	8.09	8.60	10.19	8.28	8.99	6.96	8.95	11.21	9.90	10.69	7.01	5.85	3.96	12.21	4.36	7.53	–
	Vision Pro (GPU)	64	1529.67	360.11	417.73	368.71	360.02	237.96	218.35	225.86	216.24	230.30	256.45	143.86	144.12	128.94	125.73	129.77	134.74	0
		128	2047.04	394.42	458.48	402.89	397.81	249.57	226.23	230.42	220.94	243.01	263.26	150.84	149.97	141.49	130.77	137.79	140.35	0
		256	2304.05	418.75	474.66	416.41	421.77	259.63	233.24	235.00	227.00	244.50	270.05	153.13	151.58	145.29	135.54	141.05	141.90	0
		512	2255.74	423.83	484.57	421.13	423.35	258.17	232.29	233.41	233.52	244.26	276.66	151.08	150.39	146.56	142.89	141.09	141.64	0
		1024	2117.98	416.88	474.03	409.38	418.41	246.34	224.61	232.08	227.66	225.57	265.67	139.65	140.05	142.85	137.78	138.76	141.32*	1
		μ	2050.90	402.80	461.89	403.70	404.27	250.34	226.94	231.35	225.07	237.53	266.42	147.71	147.22	141.03	134.54	137.69	139.99	–
		σ	309.13	26.40	26.39	20.75	26.79	8.91	6.08	3.50	6.65	8.93	7.55	5.71	4.94	7.04	6.58	4.65	2.99	–
		CV (%)	15.07	6.55	5.71	5.14	6.63	3.56	2.68	1.51	2.95	3.76	2.83	3.86	3.36	4.99	4.89	3.38	2.14	–
TG Test	Magic Leap 2	64	20.42	11.37	10.31	8.45	8.51	12.38	8.88	8.26	7.14	6.47	5.17	6.13	5.89	2.07	3.00	1.99	4.92	0
		128	20.41	11.27	10.32	8.40	8.47	12.21	8.80	8.20	7.09	6.44	5.13	6.08	5.86	2.07	2.99	1.98	4.89	0
		256	20.30	11.13	10.21	8.34	8.39	11.91	8.67	8.07	6.99	6.35	5.08	5.97	5.77	2.07	2.99	1.98	4.87	0
		512	20.11	10.92	10.02	8.23	8.28	11.43	8.40	7.83	6.82	6.21	5.01	5.81	5.61	2.06	2.97	1.97	4.82	0
		1024	19.81	10.53	9.70	7.98	8.04	10.56	7.93	7.43	6.49	5.94	4.83	5.47	5.32	2.06	2.93	1.96	4.71	0
		μ	20.21	11.08	10.11	8.29	8.34	11.71	8.74	8.16	6.91	6.47	5.06	5.99	5.67	2.07	2.98	1.98	4.84	-
		σ	0.22	0.33	0.24	0.17	0.17	0.35	0.34	0.32	0.23	0.21	0.13	0.29	0.18	0.00	0.02	0.01	0.07	-
		CV (%)	1.09	2.96	2.37	2.05	2.04	2.97	3.93	3.90	3.36	3.27	2.62	4.79	3.16	0.23	0.57	0.34	1.48	-
	Meta Quest 3	64	22.96	13.55	13.92	11.85	9.29	11.13	9.59	10.13	8.49	8.32	6.66	5.11	4.95	3.10	4.77	2.01	5.30	0
		128	22.79	12.01	11.83	10.03	8.51	9.06*	7.41*2	8.57	6.57	7.21	5.98	3.93*2	3.67*	2.27	3.12	1.61	3.57*	7
		256	21.14	10.37	9.66*	7.09*2	6.40	6.11	5.14*2	5.74*2	4.29*2	5.10*2	3.43*	3.07*2	2.41	1.73	2.42*	1.43	2.40	14
		512	19.62	7.78*3	5.53	4.49	4.18	4.79	3.85	4.01	3.29	3.53	2.66	2.55	2.16	1.63	2.15	1.37	2.14	3
		1024	12.65*	5.82	4.52	3.75	3.53	3.99	3.32	3.41	2.92	2.88*	2.34	2.31	2.03	1.61	2.04	1.32	2.01	2
		μ	19.43	9.71	8.81	7.04	6.38	7.02	5.87	6.37	5.11	5.41	4.21	3.79	3.44	2.27	3.30	1.75	3.40	-
		σ	4.45	3.47	3.33	3.08	2.62	3.13	2.48	2.82	2.35	2.11	1.92	1.29	1.13	0.72	1.04	0.35	1.29	-
		CV (%)	22.91	35.76	37.76	43.72	41.12	44.57	42.25	44.26	45.99	38.94	45.66	34.14	32.95	31.73	31.52	19.87	38.06	-
	Vivo X100s Pro	64	23.21	9.42	10.29	9.25	9.08	7.31	7.42	8.62	7.22	7.56	7.80	3.84	3.81	2.05	3.80	2.34	4.90	0
		128	22.55	9.26	10.06	9.14	8.93	7.25	7.36	8.53	7.13	7.49	7.66	3.83	3.78	2.07	3.70	2.33	5.00	0
		256	22.71	9.11	9.98	9.03	8.88	7.18	7.18	8.31	6.67	6.95	7.21	3.80	3.66	2.04	3.55	2.24	4.84	0
		512	22.35	8.85	9.53	8.66	8.50	6.70	6.80	7.67	5.97	6.27	6.59	3.61	3.50	1.97	3.44	2.20	4.48	0
		1024	21.69	8.20	8.78	8.00	7.82	6.24	6.15	6.96	5.18	5.32	5.24*	3.44	3.27	1.93	3.29	2.07	4.25	1
		μ	22.10	8.77	9.71	8.81	8.65	6.93	7.00	8.02	6.43	6.92	6.70	3.70	3.60	2.01	3.56	2.24	4.69	-
		σ	0.74	0.46	0.55	0.47	0.42	0.38	0.42	0.63	0.88	1.04	0.94	0.16	0.25	0.05	0.22	0.10	0.25	-
		CV (%)	3.35	5.24	5.67	5.34	4.85	5.48	5.97	7.87	13.71	15.04	14.03	4.39	6.92	2.49	6.18	4.56	5.33	-
	Apple Visions Pro	64	45.14	18.76	23.17	17.56	18.97	15.97	14.84	17.03	14.87	14.22	12.29	7.87	7.65	4.94	6.82	6.83	10.72	0
		128	45.17	19.32	22.30	18.48	18.41	15.15	14.15	16.97	13.76	13.41	12.11	7.27	7.16	4.57	6.05	6.24	10.48*	1
		256	47.83	18.15	20.88	16.85	17.15	14.04	12.83	15.55	12.54	12.12	12.23	6.74	6.83	4.46	6.03	6.11*3	9.02*2	5
		512	47.24	17.97*3	19.69	16.94*2	16.52*2	13.26	11.63	14.17	11.61*3	10.59	11.84*2	3.03*2	3.12*2	4.16	5.52	2.88*	4.06*2	19
		1024	46.35	16.58*3	19.33	15.37*3	13.43*2	11.69	10.19	12.54*	5.68*2	9.80	5.73*	2.83*	5.87*2	4.38*	5.31	2.86	4.57*2	18
		μ	46.75	18.56	21.47	17.04	17.90	14.82	12.73	15.65	11.69	12.83	10.84	5.55	6.13	4.50	6.35	4.98	8.57	–
		σ	1.15	0.97	1.43	1.14	1.89	1.33	1.63	1.55	3.58	1.82	2.47	2.42	1.48	0.54	0.72	1.35	2.16	–
		CV (%)	2.46	5.23	6.67	6.68	10.57	8.97	12.81	9.91	30.61	14.20	22.79	43.69	24.15	12.00	11.34	27.11	25.20	–
	Visions Pro (GPU)	64	43.34	22.71	28.64	22.58	22.14	23.41	19.23	19.47	17.24	16.78	14.68	14.48	13.80	14.74	15.16	14.89	15.06	0
		128	42.35	23.17	28.00	23.42	22.64	23.50	19.31	19.92	17.58	17.78	14.66	14.03	13.57	14.51	14.99	14.91	14.89	0
		256	41.85	23.08	27.01	22.89	22.62	23.81	19.31	19.12*	17.39	17.67	14.30	12.24	12.16	14.34	15.16	14.62	14.70	1
		512	40.73	22.08*	27.67*	23.36	21.85	23.08*	18.08	18.99*	16.75	16.97	13.75	11.37	9.91	13.84	14.65	13.00	13.15	4
		1024	39.80	19.95*	26.66*	21.06*	19.49	19.87*	15.72	17.88	15.58	14.55*	13.55	10.51	9.87*	12.78	13.02	12.50	11.66*	7
		μ	41.61	22.20	27.60	22.66	21.75	22.73	18.33	19.08	16.91	16.75	14.19	12.53	11.86	14.04	14.60	13.98	13.89	–
		σ	1.38	1.33	0.79	0.96	1.31	1.62	1.55	0.76	0.80	1.30	0.52	1.70	1.91	0.78	0.91	1.15	1.46	–
		CV (%)	3.32	5.98	2.85	4.24	6.00	7.13	8.45	3.98	4.75	7.78	3.66	13.57	16.07	5.55	6.20	8.20	10.52	–

References

[1] O. Mena, A. Kouyoumdjian, L. Besançon, M. Gleicher, I. Viola, and A. Ynnerman, "Augmenting a large language model with a combination of text and visual data for conversational visualization of global geospatial data," 2025.

[2] D. Jia, A. Irger, L. Besançon, O. Strnad, D. Luo, J. Björklund, A. Kouyoumdjian, A. Ynnerman, and I. Viola, "VOICE: Visual oracle for interaction, conversation, and explanation," *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–18, 2025.

Thanks.