IEEE trans. On Visualization and Computer Graphics. May 2025

ESIQA: Perceptual Quality Assessment of Vision-Pro-based Egocentric Spatial Images

Xilei Zhu, Liu Yang, Huiyu Duan, Xiongkuo Min, Guangtao Zhai, and Patrick Le Callet

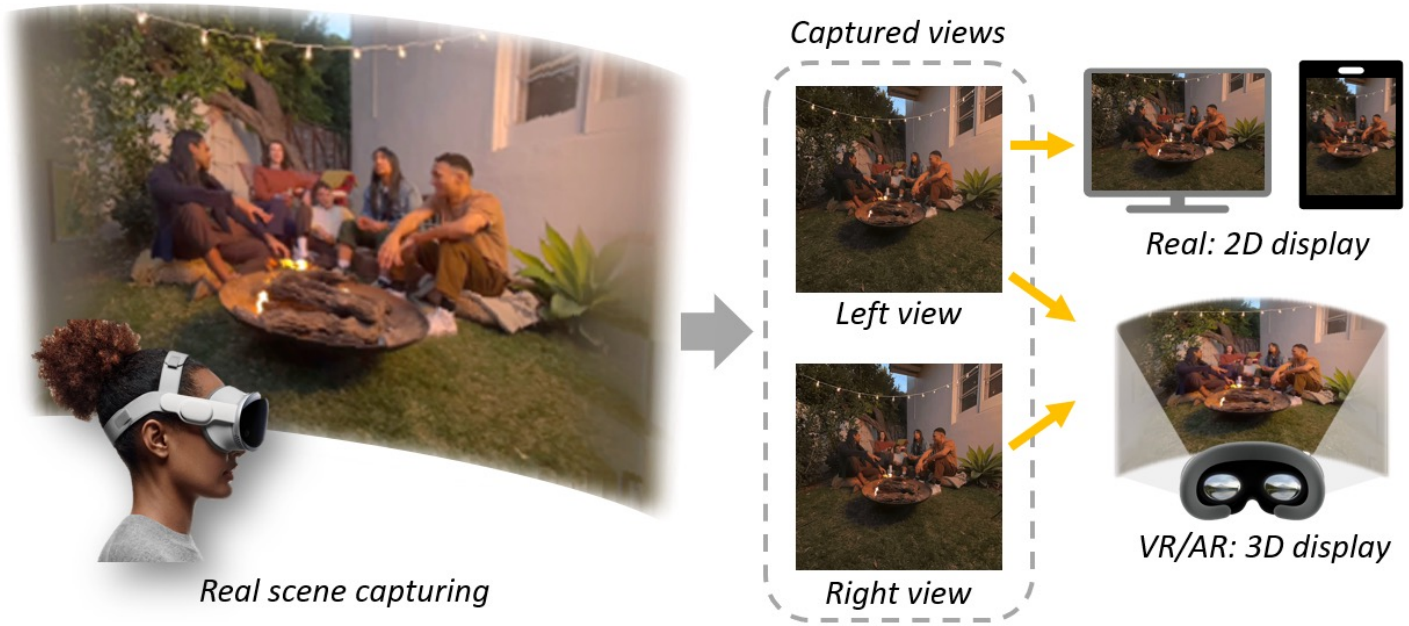
patrick.lecallet@univ-nantes.fr

VQEG spring meeting 2025 – IMG session

Background & Motivation

Why Egocentric Spatial IQA?

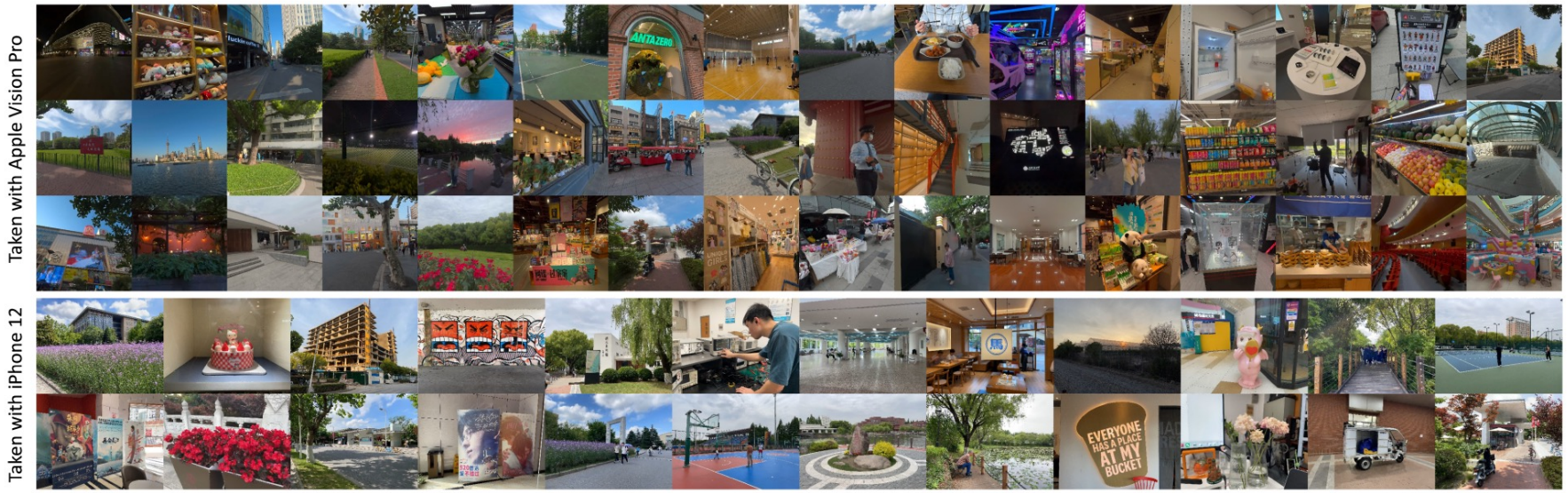
- Limitations of traditional 2D/3D IQA for immersive egocentric content.
- Gap in existing metrics: lack of stereoscopic depth + immersive FoV + first-person perspective.



Database: ESIQAD

Building an Egocentric Spatial IQA Database

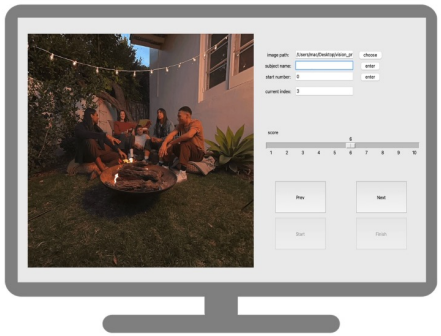
- 500 images: **400 captured** (Apple Vision Pro), **100 synthesized** (iPhone).
- Subjective experiments: 22 participants, 3 display modes (**2D, 3D-window, 3D-immersive**).



Database: ESIQAD

Building the First Egocentric Spatial IQA Database

- 500 images: **400 captured** (Apple Vision Pro) => egocentric video, **100 synthesized** (iPhone).
- Subjective experiments: 22 participants, 3 display modes (**2D, 3D-window, 3D-immersive**).



(a) 2D

(b) 3D-window

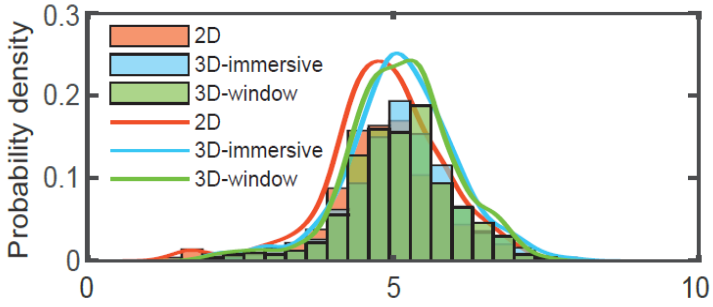
(c) 3D-immersive

Egocentric & Embodiment

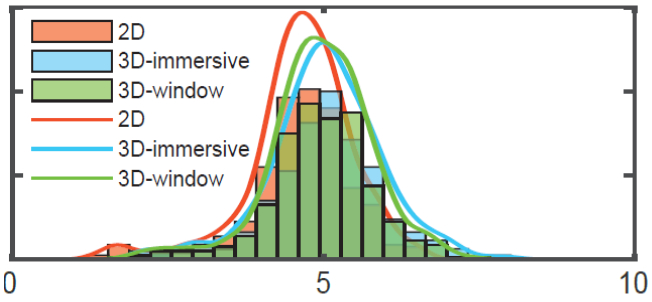
Subjective Experiment

Human Perception Analysis

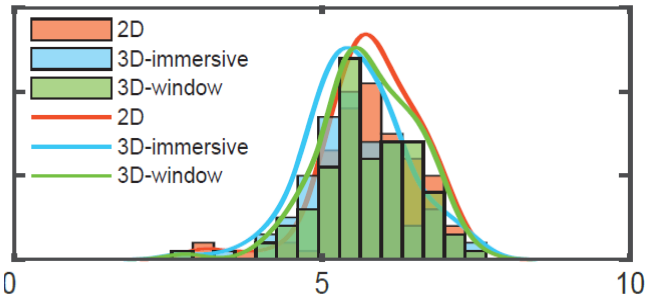
- MOS trends: Higher quality in 3D modes vs. 2D (Fig. (a)).
- **Synthesized** vs. **Captured** images: Better MOS for synthesized (Fig. (b)-(c)).
- Post-questionnaire: Preference for 3D-immersive despite dizziness.



(a) MOS: All images



(b) MOS: Captured images



(c) MOS: Synthesized images

Objective metric

Capturing the different immersive viewing conditions ?

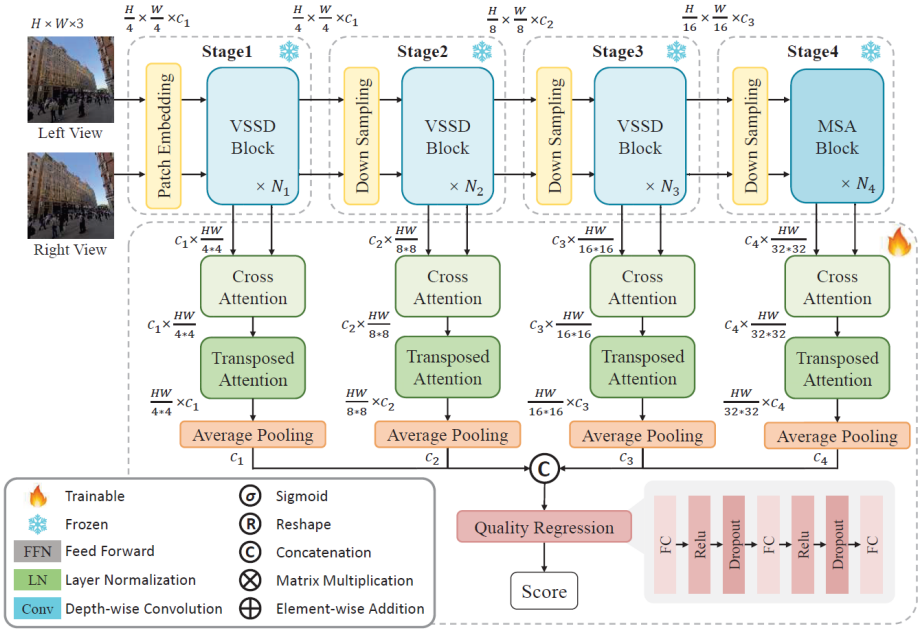
ESIQAnet

ESIQAnet: captures both pixel-level details and immersive perceptual effects that define egocentric content quality

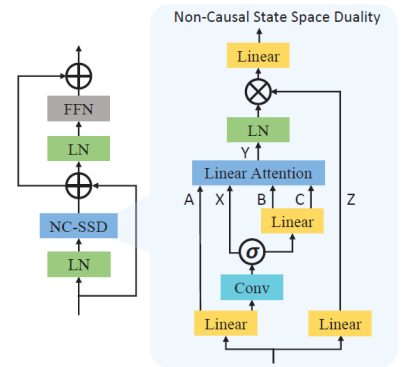
=> predict the perceptual quality of egocentric spatial images across **three display modes (**dynamically adapting to display modes during unified training)

A Mamba2-based Multi-Stage Fusion Model

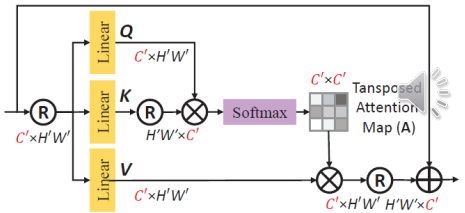
- Architecture: Patch embedding → **VSSD blocks** → Cross/Transposed attention → Quality regression.
- **Cross-attention** for binocular fusion, **Transposed attention** for channel interactions.



(a) Architecture of the ESIQAnet



(b) Visual State Space Duality Block



(c) Transposed Attention Block

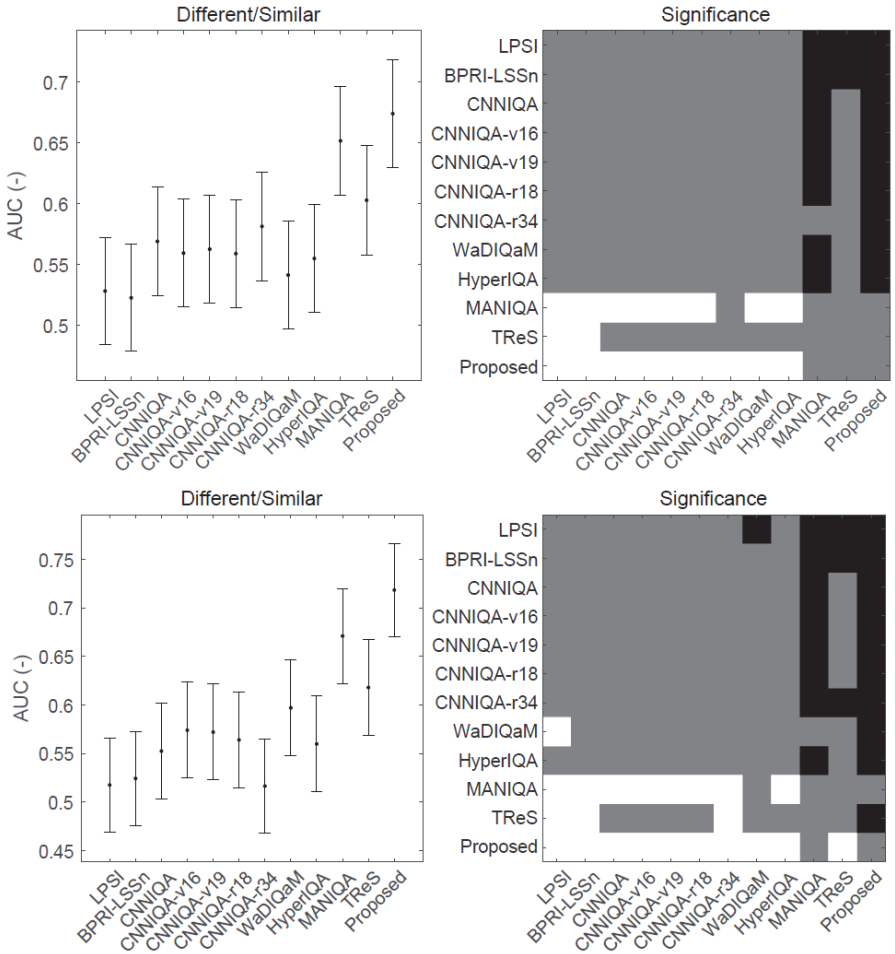
Main Results

ESIQAnet Outperforms ALL Benchmark Models

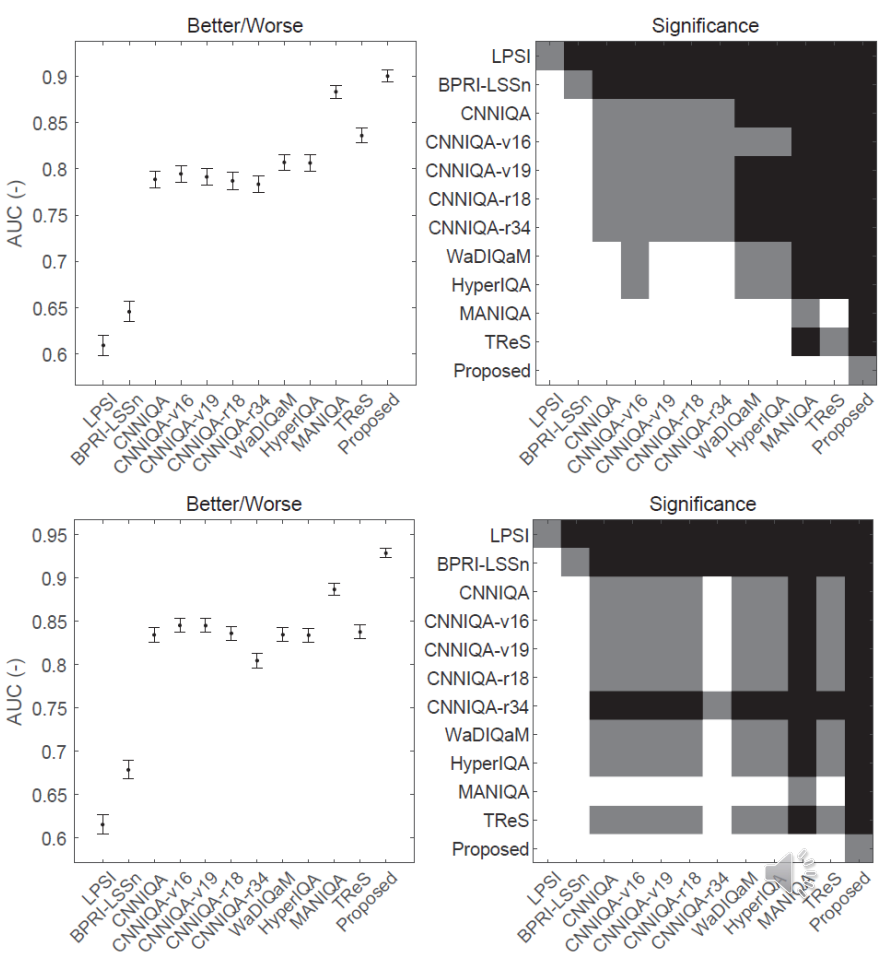
- ESIQAnet-B achieves **2.78%↑ SRCC** in 2D, **4.82%↑ SRCC** in 3D-window, **5.33%↑ SRCC** in 3D-immersive vs. best benchmark Models MANIQA.

Mode	2D			3D-window			3D-immersive		
Method	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC	SRCC	KRCC	PLCC
BMPRI [33]	0.2803	0.1798	0.5525	0.1255	0.0825	0.2841	0.0850	0.0556	0.2780
BPRI [32]	0.1780	0.1127	0.2422	0.0925	0.0637	0.1380	0.0779	0.0535	0.1400
BPRI-LSSn [32]	0.5800	0.4180	0.6255	0.4287	0.3028	0.5768	0.3550	0.2472	0.4703
BPRI-LSSs [32]	0.1400	0.0746	0.3675	0.1036	0.0741	0.1499	0.1189	0.0823	0.1675
BPRI-PSS [32]	0.2115	0.1446	0.4519	0.2348	0.1624	0.4153	0.1923	0.1320	0.2749
BRISQUE [34]	0.5594	0.3972	0.6642	0.3156	0.2186	0.4960	0.2858	0.1943	0.4623
CORNIA [52]	0.4073	0.2837	0.5134	0.1768	0.1188	0.3515	0.1572	0.1043	0.3555
FISBLIM [19]	0.4543	0.3191	0.5767	0.2800	0.1943	0.4888	0.2238	0.1521	0.4743
HOSA [49]	0.5076	0.3599	0.6211	0.2713	0.1849	0.4218	0.2749	0.1864	0.4298
NIQE [35]	0.5665	0.4088	0.6881	0.3081	0.2106	0.4957	0.3133	0.2137	0.4977
ILNIQE [54]	0.3348	0.2317	0.4187	0.3118	0.2144	0.4113	0.2670	0.1827	0.3702
LPSI [48]	0.5710	0.4071	0.6959	0.3986	0.2786	0.5384	0.3396	0.2349	0.4920
QAC [50]	0.4232	0.2867	0.5638	0.2796	0.1884	0.3882	0.2005	0.1346	0.3881
CNNIQA [23]	0.5824	0.4238	0.5404	0.5517	0.4001	0.4738	0.4579	0.3224	0.4050
CNNIQA-v16 [43]	0.6641	0.4747	0.7031	0.5958	0.4436	0.6718	0.5041	0.3666	0.6125
CNNIQA-v19 [43]	0.6883	0.5022	0.6916	0.6015	0.4470	0.6747	0.5106	0.3731	0.6007
CNNIQA-r18 [22]	0.7574	0.5628	0.7006	0.6120	0.4554	0.6678	0.5127	0.3754	0.5974
CNNIQA-r34 [22]	0.6882	0.5107	0.6632	0.6374	0.4839	0.6492	0.6213	0.4582	0.6952
WaDIQaM [4]	0.6598	0.4814	0.6963	0.6493	0.4688	0.6794	0.5010	0.3539	0.5702
HyperIQA [44]	0.6925	0.5188	0.6870	0.4901	0.3491	0.4675	0.4013	0.2808	0.4369
MANIQA [51]	0.8237	0.6376	0.8753	0.7893	0.6016	0.8366	0.7379	0.5506	0.7810
TReS [16]	0.7400	0.5620	0.8223	0.6405	0.4696	0.6974	0.6311	0.4525	0.6030
ESIQAnet-M	0.8334	0.6533	0.8683	0.8297	0.6476	**0.8638**	0.7633	0.5771	0.7980
ESIQAnet-T	0.8427	0.6622	0.8814	0.8212	0.6384	0.8557	0.7739	0.5876	0.8129
ESIQAnet-S	**0.8518**	**0.6766**	**0.8895**	**0.8314**	**0.6488**	0.8610	**0.7797**	**0.5927**	**0.8138**
ESIQAnet-B	**0.8515**	**0.6721**	**0.8900**	**0.8375**	**0.6531**	**0.8639**	**0.7912**	**0.6038**	**0.8180**

Main Results

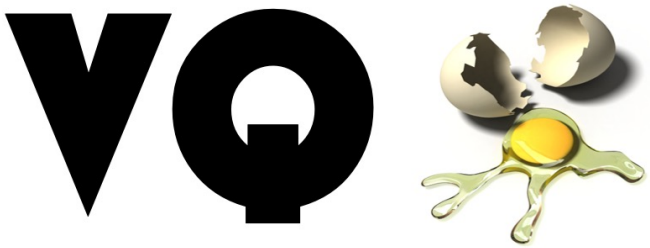


(a) Different vs. Similar Analysis

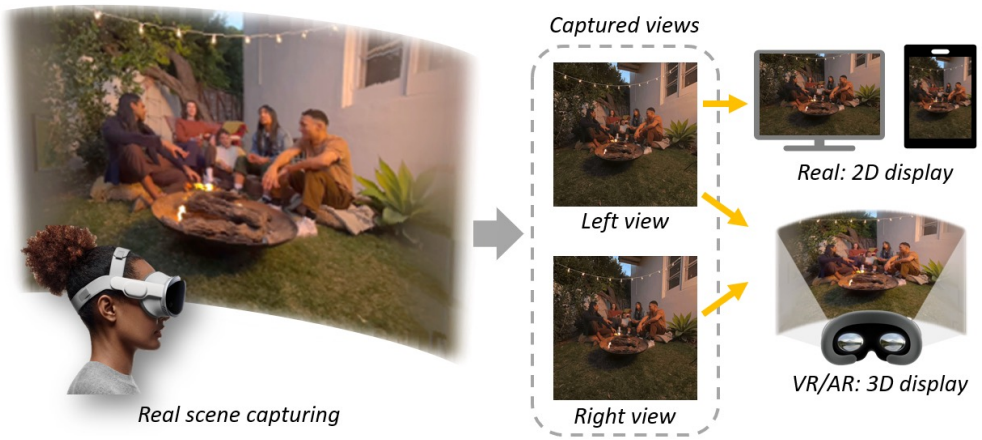


(b) Better vs. Worse Analysis

Conclusion



- **ESIQAD**: First large-scale IQA database for egocentric spatial images
- **Human preference analysis**: Revealed key characteristics under three display modes
- **ESIQAnet**: Novel mamba2-based model achieving SOTA performance across all modes
- **Benchmark validation**: ESIQAnet Outperforms existing IQA methods consistently

Database & Code: https://github.com/IntMeGroup/ESIQA

Contact us: xilei_zhu@sjtu.edu.cn, huiyuduan@sjtu.edu.cn