What Makes for a Good Stereoscopic Image?

Supplementary Material

6. Dataset Details

Our dataset is comprised of 2400 examples, each containing a pair of stereo images, resulting in a total of 4800 stereo images, all of which have undergone some form of distortion. Table 3 presents the total amount in which each distortion appears in the dataset. Figure 8 shows visual examples for several distortions, exaggerated for illustration purposes.

Distortion type	Occurrence
2D lifting	1364
MotionCtrl	1156
3D Gaussian splatting	306
SDEdit	241
Uniform White Noise	152
Chromatic Aberration	144
Rotation	141
Keystone	138
Average Blur	127
Gaussian Blur	125
JPEG Compression	124
Gaussian White Noise	123
Checkerboard	117
Warping	115
Brightness	97
Saturation	95
Contrast	80
Hue	79
Magnification	76

Table 3. **Frequency of applied distortions in our proposed dataset.**

7. Training Details

Our model is trained on a single NVIDIA A100 using an Adam optimizer with a learning rate of $3e-5$ and batch size of 16. We maintain the original 1280×720 resolution, only applying a center crop to 1274×714 for compatibility with DINOv2-S's patch size of 14. During training, we finetune the DINOv2 backbone using LoRA, with a rank of 8, alpha of 32, and dropout of 0.1. We use a with a margin of 0.05 for the hinge loss. Additionally, the training data is weighted based on annotator consensus levels, with each epoch taking approximately 12 minutes to train and 1 minute to validate.

8. Detailed Performance on the SCOPE dataset

In Figure 9 we report test set accuracy across several different train, validation and test partitions, categorized by annotation consensus: unanimous ($5-0$ split), majority ($4-1$ split), and divided ($3-2$ split), with the latter being the noisiest and most cognitively penetrable. The sizes of these splits are similar, comprising 32.9%, 34.1%, and 32.9% of the data respectively, confirming that our dataset contains a learnable signal. We train and evaluate our model on five $80\% - 10\% - 10\%$ dataset splits, using five different seeds for each split, and report the mean and standard deviation in Figure 9.

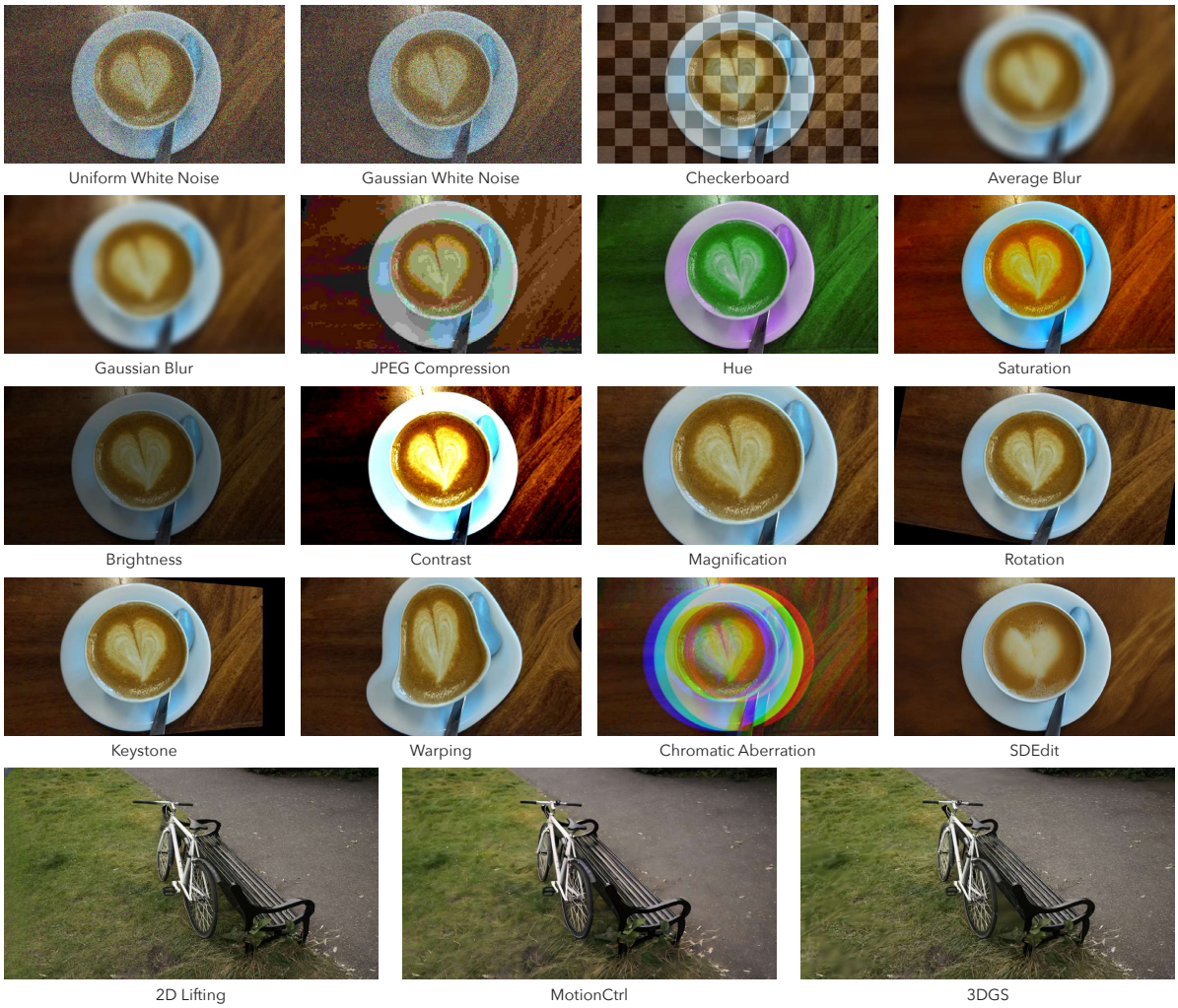

Figure 8. **Distortion examples.** We show examples of image distortions in our dataset, exaggerated for illustration purposes.

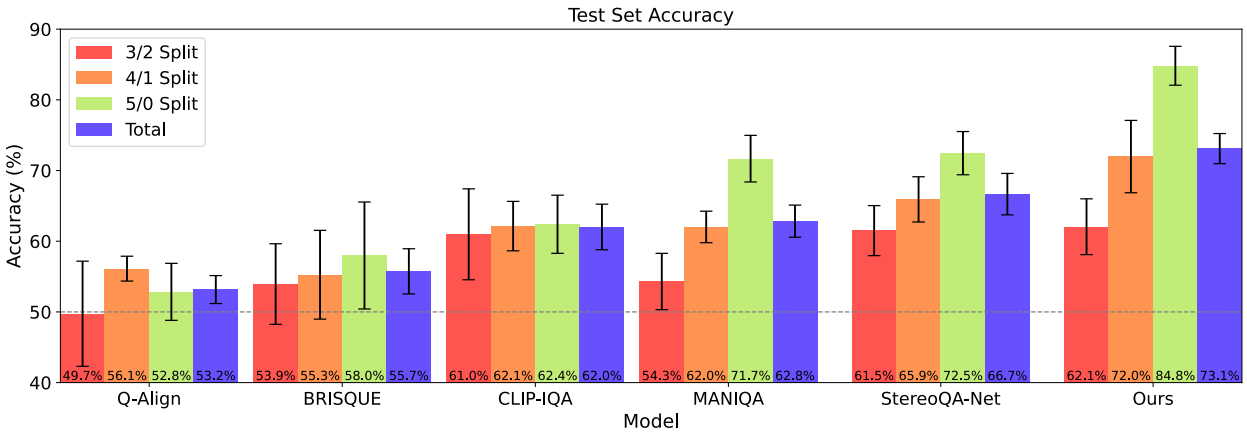

Figure 9. **Test set performance.** We test the performance of existing IQA and NR-SIQA models as well as our proposed model on a held out test sets, and show the results on different splits.

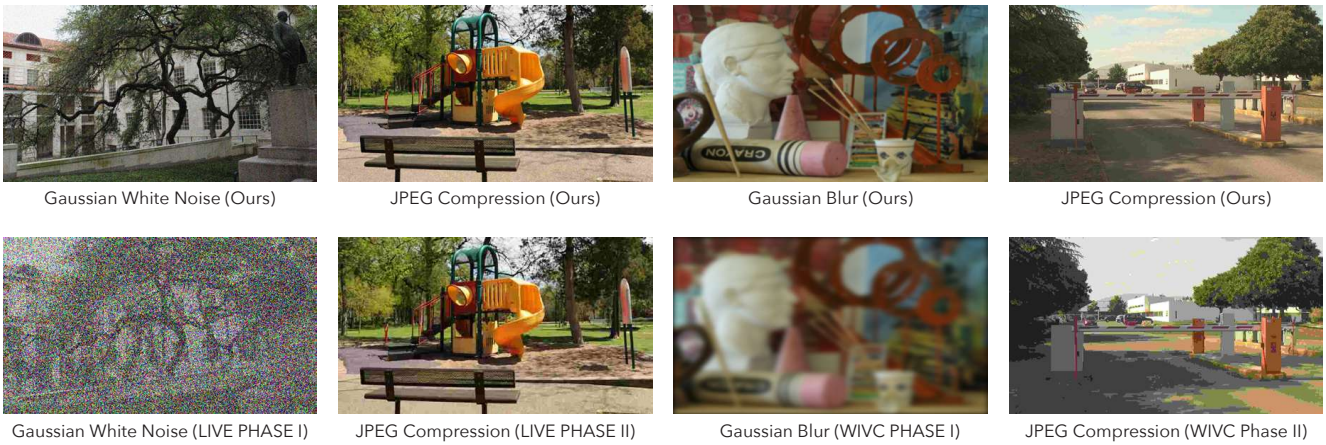

Figure 10. **Distortion strength comparison.** Comparing maximum distortion strengths across existing datasets and our proposed dataset, demonstrating that the distortions applied in our dataset exhibit significantly lower intensity compared to existing SIQA datasets.

		LIVE Phase I		LIVE Phase II		WIVC Phase I		WIVC Phase II	
	Method	**SROCC ↑**	**PLCC ↑**	**SROCC ↑**	**PLCC ↑**	**SROCC ↑**	**PLCC ↑**	**SROCC ↑**	**PLCC ↑**
Manual Feature Based	Chen *et al*. [15]	0.891	0.895	0.880	0.880	–	–	–	–
	Shen *et al*. [74]	0.932	0.936	0.927	0.932	–	–	–	–
	Li *et al*. [45]	0.953	0.965	0.946	0.955	0.937	0.949	0.952	0.960
	Liu *et al*. [49]	0.949	0.958	0.933	0.935	0.928	0.945	0.901	0.913
Deep Learning Based	Zhang *et al*. [102]	0.943	0.947	0.915	0.912	–	–	–	–
	Ding *et al*. [19]	0.942	0.940	0.924	0.930	–	–	–	–
	Fang *et al*. [24]	0.946	0.957	0.934	0.946	–	–	–	–
	Zhou *et al*. [104]	0.965	0.973	0.947	0.957	–	–	–	–
	Shen *et al*. [75]	0.962	0.972	0.951	0.953	–	–	–	–
	Si *et al*. [78]	0.966	0.978	0.953	0.972	0.960	0.969	0.950	0.958
	Zhang *et al*. [100]	0.972	0.977	0.962	0.964	0.972	0.973	0.972	0.973
	iSQoE (Ours)	0.774	0.758	0.763	0.767	0.627	0.687	0.542	0.536

Table 4. **Evaluation on existing datasets for stereoscopic image quality assessment.**

9. Performance on Existing SIQA Datasets

Table 5 provides a comparison between our dataset and existing stereo quality assessment datasets: LIVE 3D Phases I and II [16, 57], Waterloo IVC (WIVC) 3D Phases I and II [84, 85] and IEEE-SA [48]. SCOPE differs from them in several aspects:

1. **Image Quantity:** SCOPE is the largest of the datasets, with more than twice the amount of samples than IEEE-SA - the second largest dataset.
2. **Annotation medium:** The annotations in all these datasets were collected using passive stereoscopic displays or active shutter glasses, while ours were collected on a Vision Pro headset. In Section 4.3 and Figure 6 we demonstrate low correlation between preferences on VR devices and other stereo viewing methods.
3. **Annotation Protocol:** The other datasets collected Mean Opinion Score annotations, an absolute single-image protocol, while SCOPE collected 2AFC which are relative annotations.
4. **Distortion Strengths:** The other datasets applied significantly stronger distortions than SCOPE, see Figure 10.

We evaluate our model on LIVE 3D Phase I and II [15, 16, 57] and Waterloo IVC (WIVC) 3D Phase I and II [84, 85]. For these evaluations, we use standard performance metrics: Spearman rank order correlation coefficient (SROCC) and Pearson linear correlation coefficient (PLCC).

Table 4 shows there is a significant performance gap between our models and the state-of-the-art models reporting performance on these datasets. We attribute this to the difference in annotation mediums between these datasets and

Dataset	Samples	Stereo Images	Clean Images	Annotation Type	Distortions
LIVE Phase I [57]	365	365	20	DMOS	Noise, Blur, Compression, Fast-fading
LIVE Phase II [16]	360	360	8	DMOS	Noise, Blur, Compression, Fast-fading
WIVC Phase I [84]	330	330	6	MOS	Noise, Blur
WIVC Phase II [85]	460	460	10	MOS	Noise, Blur, Compression,
IEEE-SA [48]	800	800	160	MOS	Horizontal disparity
SCOPE (Ours)	2400	4800	2400	2AFC	19 types, see Table 1

Table 5. **Stereoscopic Preference Datasets.** Prior datasets for stereo image evaluation vary in terms of size, the psychophysical experiment in which the annotations were collected, and the distortions they encompass.

SCOPE. Our model is trained and fitted to grade quality as it is perceived on a VR device, rather than on passive stereoscopic displays.

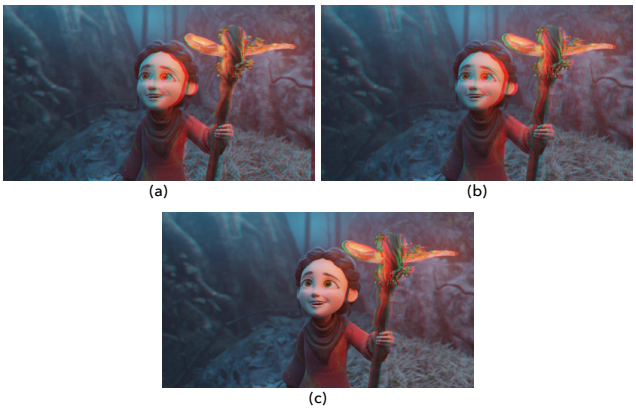

Figure 11. **Example from our off-the-shelf, mono-to-stereo experiment.** Three versions of the same stereo image are generated using different off-the-shelf, mono-to-stereo conversion methods. (a) Depthify.ai (b) Immersity AI (c) Owl3D. The stereo images are presented as anaglyph images for viewing purposes. We recommend viewing the images on a screen and zooming in to better observe the differences.

10. Cross-Medium User Study

Expanding on the user study outlined in Section 4.3, we detail the specific viewing setups for each device. Viewing stereoscopic images with the Apple Vision Pro was done through the native photos app in immersive mode. For the Meta Quest Pro, we employed a third-party application (Pegasus VR media player) due to the absence of a suitable first-party viewer. Both the toggling and anaglyph setups were presented via HTML pages, shown in Figure 12. We opted for full-color anaglyph images, as this convention provided the best stereoscopic 3D experience with our monitor and glasses combination, among all conventions tested.

In addition to Figure 6 that shows the mean correlation, Figure 13 we show the Cohen's kappa coefficient between each of the 10 participants for each viewing device.

11. Off-the-Shelf Mono-to-Stereo Evaluation

We evaluated alignment of human opinion with the different SQoE candidates on the Spring [52] dataset. Figure 11 shows an example from the user study.

12. Licenses

The models and datasets we use are provided under the licenses in Table 6.

Dataset	License	Model	License
Tanks and Temples	CC BY 4.0	MotionCtrl	Apache 2.0
Deep Blending	Apache 2.0	MiDaS	MIT
Mip-NeRF 360	Apache 2.0	Marigold	Apache 2.0
Holopix50k	NC	Depth Anything	Apache 2.0
SPRING	CC BY 4.0	LaMa	Apache 2.0
LIVE 3D Phase I	Custom Academic	3DGS	NC
LIVE 3D Phase II	Custom Academic	DINOv2	Apache 2.0
WIVC 3D Phase I	Custom Academic	Q-Align	S-Lab 1.0
WIVC 3D Phase II	Custom Academic	BRISQUE	Apache 2.0
		CLIP-IQA	S-Lab 1.0
		MANIQA	Apache 2.0
		StereoQA-Net	Custom Academic
		CLIP	MIT
		OpenCLIP	MIT
		Croco	CC BY-NC-SA 4.0
		Depthify.ai	Custom
		Immersity AI	Custom
		Owl3D	Custom

Table 6. **Dataset and model licenses.**

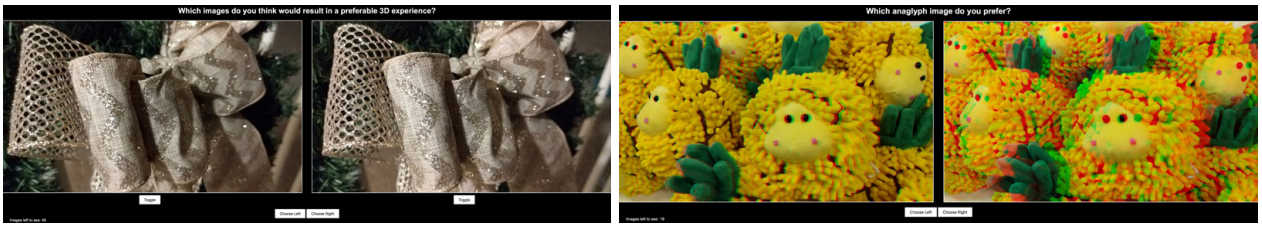

Figure 12. **User study setups: left/right image toggling (top) and anaglyph stereo (bottom)**

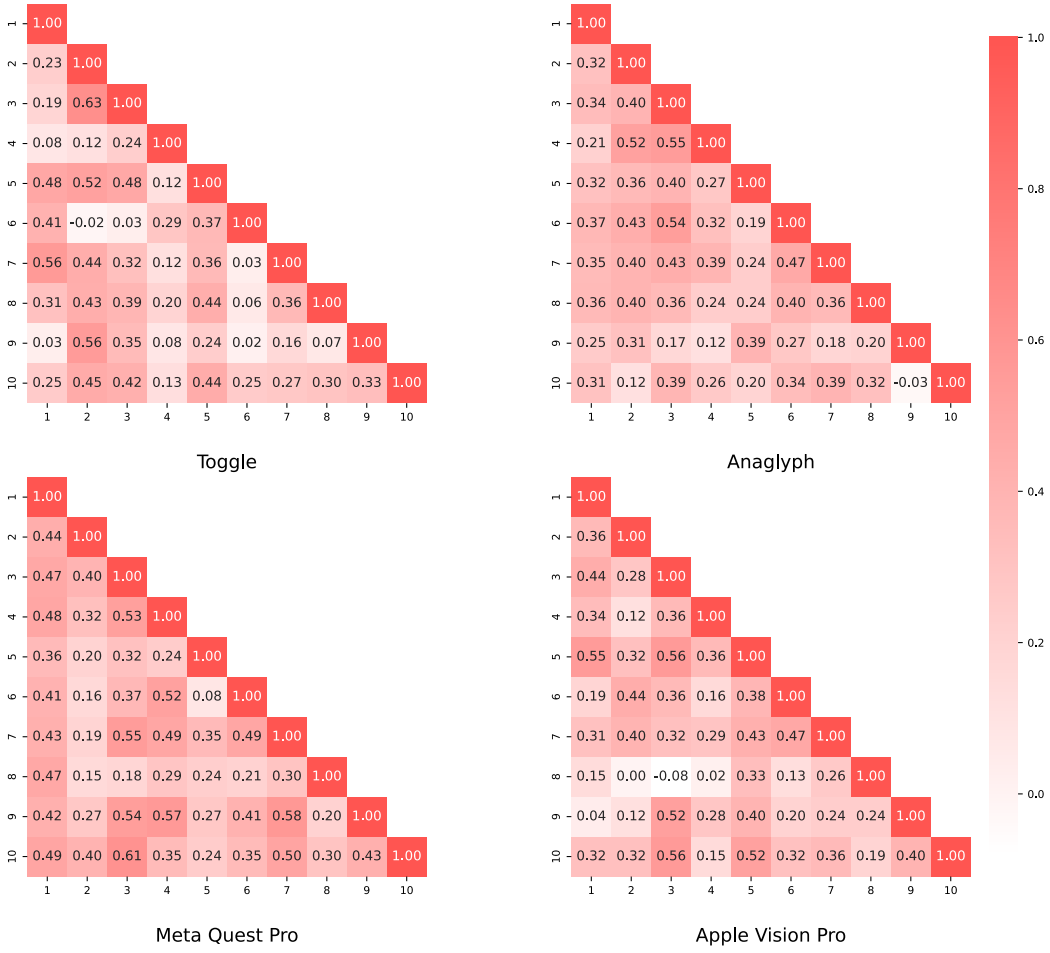

Figure 13. **Inter-rater agreement for each viewing medium, measured using Cohen's kappa coefficient.** The heatmap displays the agreement scores between all pairs of the 10 participants, highlighting the correlation in subjective evaluations across different viewing conditions.