A Display Simulation

Here, we provide additional details related to the implementation and visualization of our TMO.

A.1 Tone vs. Relative Luminance

Tone mappers typically operate on relative luminance. As stated in the main manuscript, this provides no guarantees that the tone-mapped luminances are within the gamut of the target display. Here, we show that operations on *tone*, on the other hand, guarantee that the output of tone mapping is within gamut. Consider an RGB color vector $\mathbf{c}$ that we want to tone map. Colors may go out of gamut after the color correction step. We have the following Schlick [1995] correction computation when operating on tone,

$$t = \max(\mathbf{c}_r, \mathbf{c}_g, \mathbf{c}_b) \tag{7}$$

$$\mathbf{c}' = \frac{t'}{t} \cdot \mathbf{c} \tag{8}$$

$$= t' \cdot \frac{\mathbf{c}}{\mathbf{c}_{\max}}; \text{ say } \mathbf{c}_{\max} = \mathbf{c}_r \text{ W.L.O.G.,} \tag{9}$$

$$= t' \cdot (1, \frac{\mathbf{c}_g}{\mathbf{c}_{\max}}, \frac{\mathbf{c}_b}{\mathbf{c}_{\max}}) \tag{10}$$

where t' is t after tone mapping. t' is guaranteed to have value $\leq L_{\max}$, and both $\frac{\mathbf{c}_g}{\mathbf{c}_{\max}}$ and $\frac{\mathbf{c}_b}{\mathbf{c}_{\max}}$ have value ≤ 1 given our assumption that $\mathbf{c}_{\max} = \mathbf{c}_r$. As such, all tone-mapped colors $\mathbf{c}'$ are within $L_{\max}$.

A.2 Smoothness Heuristic

In the Chen et al. [2023] TMO (and our Fixed TMO), smoothness is set to $s = 0.3$. A heuristic function was used to compute the smoothness parameter s in the Content-Aware TMO,

$$s = \frac{1}{W \cdot H} \sum_{t \in \log_{10}(\mathcal{T})} \mathbb{1}(t), \tag{11}$$

where $\mathbb{1}(t)$ is an indicator function which equals 1 if $t > L_{\text{init}}$ and 0 otherwise, and (W, H) are image resolution. This heuristic increases the contrast of highlights (closer to log-linear) if they are a large portion of the image and vice versa (highlights clipped).

A.3 Example Tone-Mapped Stimuli

We present the results of our display simulation tone mapping in Figures 17 to 20, where the Fixed TMO is shown on top and Content-Aware TMO on bottom for representative frames in each of the four categories described in Section 3.3. Please note that images here are not exactly what users saw in the study, and are all tone-mapped using the Reinhard et al. [2002] Photographic TMO to an SDR display (roughly 300 nits peak luminance) for presentation in a PDF format. Please see our supplementary webpage for a more accurate depiction of the scenes.

A.4 Tone Mapping Pseudocode

Pseudocode describing both the Fixed TMO and Content-Aware TMO are shown in Algorithm 1. Note that the only difference between the two TMOs is the way in which the start luminance of highlight compression L_{init} and the smoothness s are computed. We omit the implementations of Spline and VR_TMO because they are described in detail in Chen et al. [2023] and Tariq et al. [2023].

Algorithm 1: Tone mapping operator

Function `FixedTMO`($\mathcal{I}$, $D_{\min}$, $D_{\max}$, $L_{\min}$, $L_{\max}$):
- **Input**: $\mathcal{I}$; linear RGB image,
  - $L_{\min}$; reference display black level,
  - $L_{\max}$; reference display peak luminance,
  - $L_{\min}$; target display black level,
  - $L_{\max}$; target display peak luminance
- **Output**: tone-mapped image $\mathcal{I}'$
- `// compute tone`
- $\mathcal{T} = \max(\mathcal{I}_r, \max(\mathcal{I}_g, \mathcal{I}_b))$; `// Eq. (1)`
- `// compress highlight contrast [Chen et al. 2023]`
- init $\mathcal{T}' = \mathcal{T}$;
- $s = 0.3$;
- **if** $L_{\max} < 100$; **then**
  - $L_{\text{init}} = 25$;
- **else**
  - **if** $L_{\max} < 200$; **then**
    - $L_{\text{init}} = 50$;
  - **else**
    - $L_{\text{init}} = 120$;
  - **end**
- **end**
- $\mathcal{T}'[T > s] = \textsc{Spline}(L_{\text{init}}, s, D_{\min}, D_{\max}, L_{\min}, L_{\max})$;
- `// Schlick [1995] color correction`
- $\mathcal{I}' = \mathcal{I} * \mathcal{T}'/\mathcal{T}$; `// Eq. (2)`
- `// simulate black level`
- $\mathcal{I}' = \mathcal{I}' * (L_{\max} - L_{\min})/L_{\max} + L_{\min}$; `// Eq. (3)`
- **return** $\mathcal{I}'$

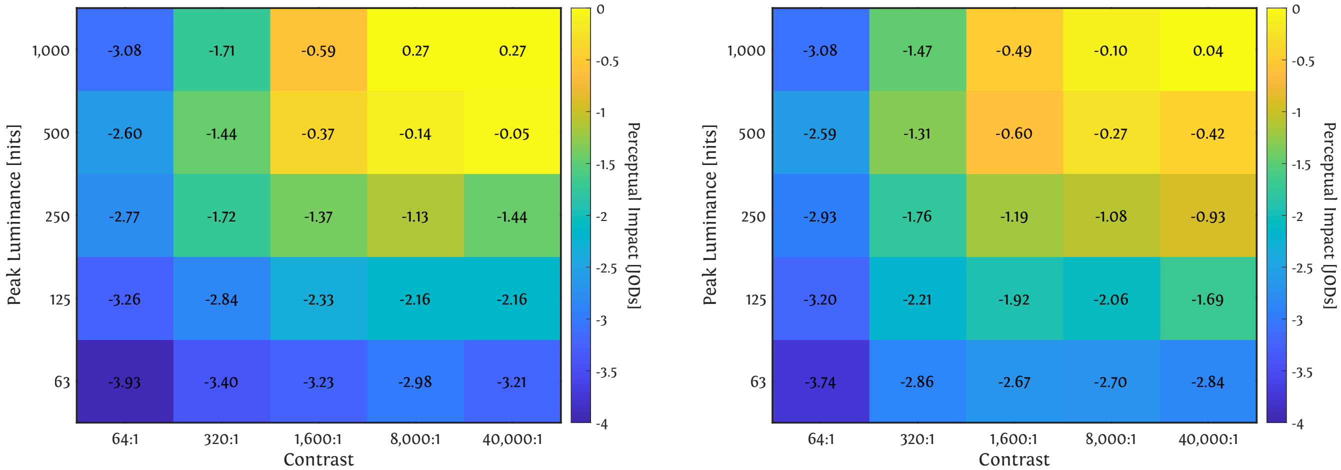
Function `ContentAwareTMO`($\mathcal{I}$, $D_{\min}$, $D_{\max}$, $L_{\min}$, $L_{\max}$):
- `// compute tone`
- $\mathcal{T} = \max(\mathcal{I}_r, \max(\mathcal{I}_g, \mathcal{I}_b))$; `// Eq. (1)`
- `// compute optimal starting luminance [Tariq et al. 2023]`
- $L_{\text{init}} = \text{VR\_TMO}(\mathcal{I}, \textsc{Spline}, D_{\min}, D_{\max}, L_{\min}, L_{\max})$;
- `// smoothness heuristic`
- $s = 1 - \text{sum}(\text{log10}(\mathcal{T}) < L_{\text{init}})$ / size($\mathcal{T}$); `// Eq. (11)`
- `// compress highlight contrast [Chen et al. 2023]`
- init $\mathcal{T}' = \mathcal{T}$;
- $\mathcal{T}'[T > s] = \textsc{Spline}(L_{\text{init}}, s, D_{\min}, D_{\max}, L_{\min}, L_{\max})$;
- `// Schlick [1995] color correction`
- $\mathcal{I}' = \mathcal{I} * \mathcal{T}'/\mathcal{T}$; `// Eq. (2)`
- **return** $\mathcal{I}'$

B Additional User Study Results

The complete study results with exact JOD values are shown in Figure 11 for both tone mapping techniques we studied. Colors correspond to magnitude of JOD values, as shown in the color bar.

C Computational Model

The parameters of our computational model, described in Section 5, are listed in Table 1 below. We also show additional results of our model evaluation in this section.



(a) Fixed TMO

(b) Content-Aware TMO

Fig. 11. *Complete user study results.*Results of the main user study are plotted here for both the (a) Fixed TMO and the (b) Content-Aware TMO. Colors represent JOD values (defined in the color bar), with values labeled in each cell of the matrix. x-axis represents contrast, and y-axis peak luminance (nits). Here, the reference condition (not shown in these plots) was mapped to 0 JODs.

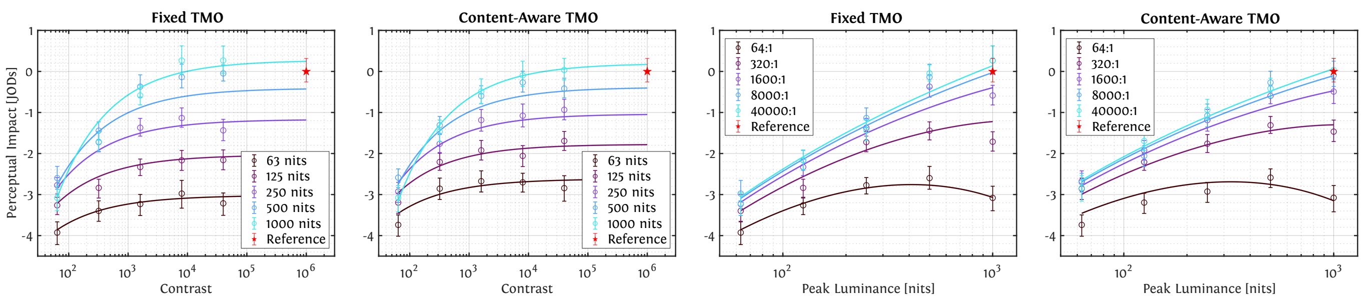


Fig. 12. *Model evaluation for all conditions.* This plot is identical to the result in Appendix C.1, except all conditions are plotted in a single figure for ease of comparison. The left two plots are projections on the contrast axis, and the right two are for peak luminance.

Table 1. Parameters of our computational model.

TMO	k_1	k_2	k_3	k_4
Fixed TMO	7.897e3	7.902e3	0.845	8.151e-4
Content-Aware TMO	9.971e3	9.975e3	0.847	5.518e-4

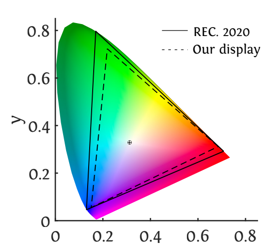
C.1 Additional Model Evaluation Results

The color of the curves in Figure 6 corresponds to the data (whether Fixed TMO or Content-Aware TMO) the model was trained on. Note here that the model evaluation captures interesting effects evident in our data. For example, in the bottom left plot (64:1 contrast), our model shows a decrease in quality scores with increased peak luminance. Our model also exhibits a saturation effect for increasing contrast (first row). We show additional plots of our model evaluation in Figure 12, including all study conditions in a single plot to compare the effect of each parameter.

D Haploscope Testbed

In this section, we include additional details of our haploscope.

D.1 Display Calibration



Both displays were calibrated to a peak luminance of 1,000 nits with ITU-R Recommendation BT.2020 (Rec. 2020) primaries, D65 whitepoint, with the perceptual quantizer (PQ/SMPTE ST 2084) electro-optical transfer function [Miller et al. 2013; Standard 2014]. A comparison between our display's primaries and the REC. 2020 primaries are shown in the inset. A Konica Minolta CS-2000 spectroradiometer[9] was

[9]Konica Minolta CS-2000 spectroradiometer display calibration device, for more details: sensing.konicaminolta.us/us/products/cs-2000-spectroradiometer/.

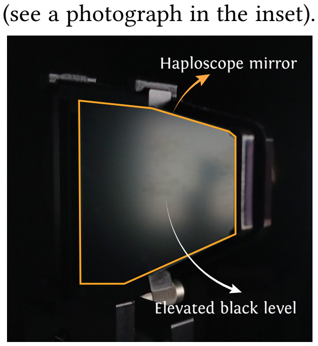
used to interface with EIZO's ColorNavigator 7 software which automatically displays uniform patches for measurement. The CS-2000 imaged each display with 1° measuring angle through the viewing mirrors. Once complete, the software stores the result as a 3D lookup table for later reproduction of the calibration target. Color reproduction accuracy degrades with time, and a display requires recalibration every so often to maintain performance. As such, we performed near-daily calibration of both EIZO displays using its built-in colorimeter. Additional validation of our system with gray square patches rendered in PsychToolbox, measured with a Spectrascan PR-745 spectroradiometer, are shown in Figure 13 where we find alignment between measured and input luminances.

D.2 Viewing Conditions

Haploscope displays were configured to have a virtual display distance of 55 cm, or 1.82 diopters (D), in front of the user. This provides a 65.7° field of view and 58.5 pixels per degree (ppd). See Fig. 2b for a schematic of this setup. Because this is a relatively high ppd for commercial VR, we reduce the display resolution to 1080p, which consequently halves the ppd to 29.2. For context, popular commercial options like the Meta Quest 3 have a display with an estimated 25 ppd[10], and field of view up to 110° [Mehrfard et al. 2019]. Our setup's field of view is lower in comparison, but is slightly higher than that of the HDR VR prototype from Matsuda et al. [2022b]. During testing, we found the display distance was the closest for comfortable binocular fusion.

D.3 Mitigation of Cross-talk

The opposing arrangement of our displays leads to crosstalk (light from one display reflects off the other). When studying high contrasts (low black levels) and bright peak luminances, this reflection can raise the black level by several orders of magnitude, reducing the contrast of the display significantly. Assuming the worst case when one display outputs full-frame white and the other is off, black level is raised by 21.25 nits (see a photograph in the inset). We mitigated this effect by enclosing the viewing optics using dark sheets, which converge to an optical exit near the user's eye. The opening of the enclosure is covered by the user's head, effectively eliminating crosstalk in our system. An additional baffle is placed between the mirrors (not pictured), and blocks light from passing near the user's nose bridge. Refer to Figure 2b for the optical arrangement.



E Main Study Scenes

Descriptions and length of our HDR video dataset are as follows. Note that there are no publicly available stereoscopic HDR video datasets. All stimuli in our study showed the same video to the left and right eye.

[10]vr-compare.com/headset/metaquest3

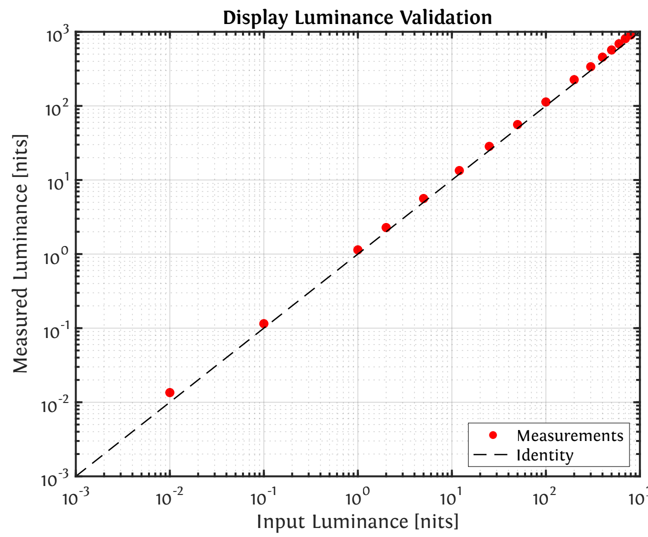


Fig. 13. *Additional haploscope calibration results.* A Spectrascan PR-745 was used to measure luminance patches rendered in PsychToolbox. The x-axis is the input patch luminance, and the y-axis is the luminance measured by the PR-745 (both on a log-nit scale). Red points are measurements, and the dashed line is identity.

E.1 Productivity

- Blender (7 seconds): scrolling through the Blender site[11].
- Earth (7 seconds): a widget with a rotating Earth and a brief text description.
- Messenger (8 seconds): a user likes a message and clicks on images in a messaging application.

E.2 Faces

- Face 1 (7 seconds): a female with dark skin tone, glasses.
- Face 2 (7 seconds): a female with tan skin tone, earrings.
- Face 3 (7 seconds): a male with light skin tone and a beard.

E.3 UGC/Passthrough

- Street (7 seconds): a street at night with people walking about, including bright streetlights and signage.
- Courtyard (8 seconds): tree in a courtyard, camera zoom-in.
- Porsche (5 seconds): zoom-out on the wheel of a red car.

E.4 Entertainment

- Smith (6 seconds): a blacksmith hammers a piece of metal, with bright sparks flying.
- Showgirl (5 seconds): a camera pan to a woman preparing for a performance.
- Werewolf (4 seconds): a cartoon scene where a teddy bear fights a werewolf.

[11]https://www.blender.org/features/

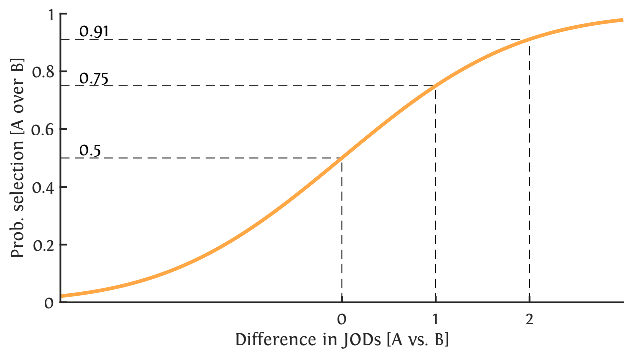


Fig. 14. We map JODs (x-axis) to units of percentage preference (y-axis).

F JOD Definition

The JOD unit is defined in Perez-Ortiz and Mantiuk [2017]. JODs can be mapped to percentage preference, as shown in Figure 14. Here, we show that a difference in 1 JOD between some condition A and another B equals a percentage selection of A of 75% over B.

G Validation Study

This section describes additional details related to the study conducted to validate our computational model (Section 5.1).

G.1 Experiment Protocol

We used the ITU P.910 5-point rating scale, which ranges between Bad, Poor, Fair, Good, Excellent [Installations and Line 1999]. Users employed a continuous slider, which was implemented as a selection between 100 discrete values on this scale. The selection marker is initialized at the center of the scale during each rating session. Similar to our main study, a 500 ms grey blank is inserted when switching stimuli.

G.2 Validation Study Scenes

Six HDRI probes stored in .exr format were used. One scene in each of the Productivity, Faces, and Entertainment categories was selected, and three were chosen in the UGC/Passthrough category. Scenes in the UGC/Passthrough category were sourced from Poly Haven[12]. The background in the Productivity scene is also from Poly Haven, but with the Blender webpage and UI buttons overlayed. The Face scene is taken from a frame of the Face 2 video in the main study. The Entertainment scene was modeled in Blender, with assets coming from the main study's Werewolf scene. The HDRIs are shown in Figure 15; note that images were manually tone-mapped to show relevant details in this PDF format.

G.3 Contrast Measurement

Simultaneous contrast measurements for the HDR VR display used in our evaluation study are shown in Figure 16. A Konica Minolta CS-2000 was positioned at the center of the HMD, and measured both black and white squares in checkerboard test patterns with

[12]Poly Haven webpage: polyhaven.com

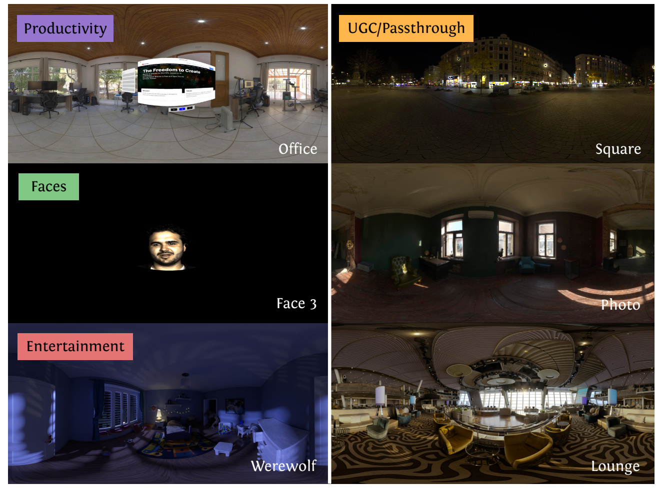


Fig. 15. Here we display the 360° HDRIs used in the subjective model validation study. Images were manually tone-mapped for visualization. Image credits Greg Zaal & Sergej Majboroda, Blender.

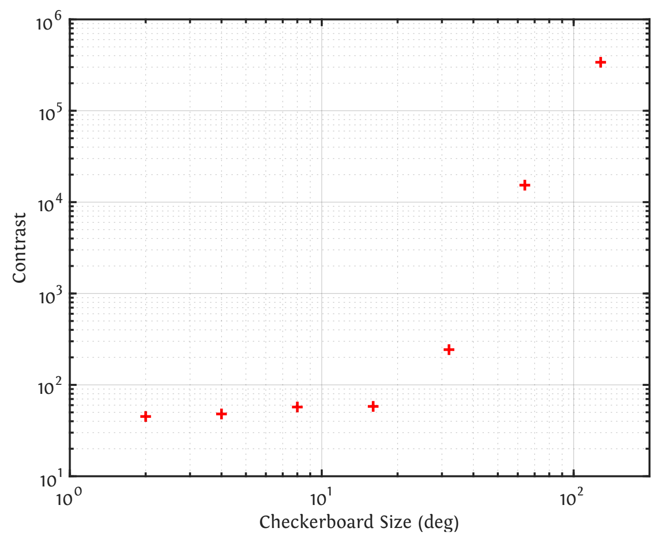


Fig. 16. *HDR VR display contrast measurement.* The contrast (y-axis) of our HDR VR prototype were measured for increasing checkerboard size (x-axis). Both axes are plotted on a log scale.

squares of size 2, 4, 8, 16, 32, 64, and 128 degrees of visual angle. Our HMD has a 62° field of view; the 64 degree condition was a split screen black/white pattern while the 128 degree condition was either full screen white or black (essentially a sequential contrast measurement). Contrast was computed by dividing the measurement of the white square by the black square.

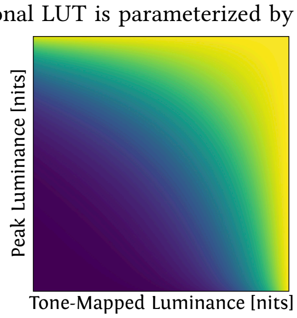
The contrast of each scene was measured using the CS-2000 by placing black and white squares (3° of visual angle) in bright scene regions. Full data for each scene at the corresponding peak luminances (mapped using our Fixed TMO) are shown in Table 2. Higher contrasts correspond to larger scatter points in Figure 8.

Table 2. *Validation study scene contrast measurements.* Black and white represent the measurements of the black and white patches, respectively. Contrast is equal to the white divided by the black measurement. L_{max} is the peak luminance the display is mapped to using the FIXED TMO.

scene	black [nits]	white [nits]	contrast	L_{max} [nits]
Office	1.27E+00	6.71E+01	52.9:1	60
Office	4.85E+00	2.73E+02	56.3:1	250
Office	1.36E+01	7.91E+02	58.0:1	750
Office	1.79E+01	1.04E+03	58.1:1	1000
Lounge	1.47E+00	6.72E+01	45.7:1	60
Lounge	4.85E+00	2.73E+02	56.2:1	250
Lounge	1.23E+01	7.89E+02	64.3:1	750
Lounge	1.58E+01	1.04E+03	65.8:1	1000
Face 3	1.40E+00	6.70E+01	47.7:1	60
Face 3	4.74E+00	2.72E+02	57.4:1	250
Face 3	1.07E+01	7.87E+02	73.5:1	750
Face 3	1.35E+01	1.03E+03	76.6:1	1000
Photo	1.30E+00	6.69E+01	51.6:1	60
Photo	3.71E+00	2.72E+02	73.3:1	250
Photo	5.96E+00	7.84E+02	132:1	750
Photo	7.64E+00	1.03E+03	135:1	1000
Square	2.25E-01	6.57E+01	292:1	60
Square	5.96E-01	2.68E+02	449:1	250
Square	1.26E+00	7.78E+02	618:1	750
Square	1.60E+00	1.02E+03	638:1	1000
Werewolf	2.95E-01	6.59E+01	224:1	60
Werewolf	6.15E-01	2.68E+02	436:1	250
Werewolf	1.06E+00	7.78E+02	733:1	750
Werewolf	1.32E+00	1.02E+03	773:1	1000

G.4 Real-Time TMO

In order to implement the FIXED TMO on our HDR VR HMD, we had to make it real-time to enable head-tracking. The spline curve defined by Chen et al. [2023] cannot be computed analytically, and evaluating it via e.g. binary search is too costly within a shader. Real-time performance is accomplished by storing the tone curve as a lookup table (LUT) and applying the TMO as a post-processing fragment shader in Unity, similar in principle to Tariq et al. [2023]. A two-dimensional LUT is parameterized by target peak luminance L_{max} and input luminance L, and outputs tone-mapped luminance (see inset for visualization). We define a mapping $(L_{\text{max}}, L) \rightarrow (i, j)$, where i, j index into a cell of the 2D LUT to output the tone-mapped luminance, as defined by the shape of the Chen et al. [2023] TMO. More specifically, L_{max} indexes into a row of the LUT. The first cell of the row maps the input luminance $L = L_{\text{init}} = 120$, and the last cell of the row maps L=1,000 nits to L_{max}. A row consists of 256 elements interpolated at equal steps in a log space; we did not find any contouring artifacts with this LUT size. Two additional LUTs were created for $L_{\text{init}} < 200$ and $L_{\text{init}} < 100$, but in practice a 3D LUT could have been defined.



H Application Details

The displays, their specifications, and predicted JOD scores for application Figure 14 are as follows:

- DCI Cinema Standard (SDR); -0.3 JODs (48 nits, 2,000:1)
- Epson 3800 (Projector); 0.9 JODs (258 nits, 147:1);
- Dolby Cinema; 1.0 JODs (106 nits, 7.5k:1);
- DCI Cinema Standard (HDR); 2.2 JODs (300 nits, 60k:1);
- Sony X90L; 2.9 JODs (711 nits, 42,222:1);
- Dell Inspiron 15 3000; 1.6 JODs (239 nits, 1,098:1);
- Dell U2723QE; 2.2 JODs (415 nits, 1,978:1)

I Traditional Display Specifications

Budget televisions can even have a higher contrast than commercial VR display. For example, the Hisense A7N LED TV ($200) has a 5000:1 contrast[13]. The high-end models often go over 380,000:1, such as the Sony BRAVIA 9 QLED TV[14].

[13] Hisense A7N LED $200 TV Review
[14] Sony BRAVIA 9 QLED $3,000 TV Review

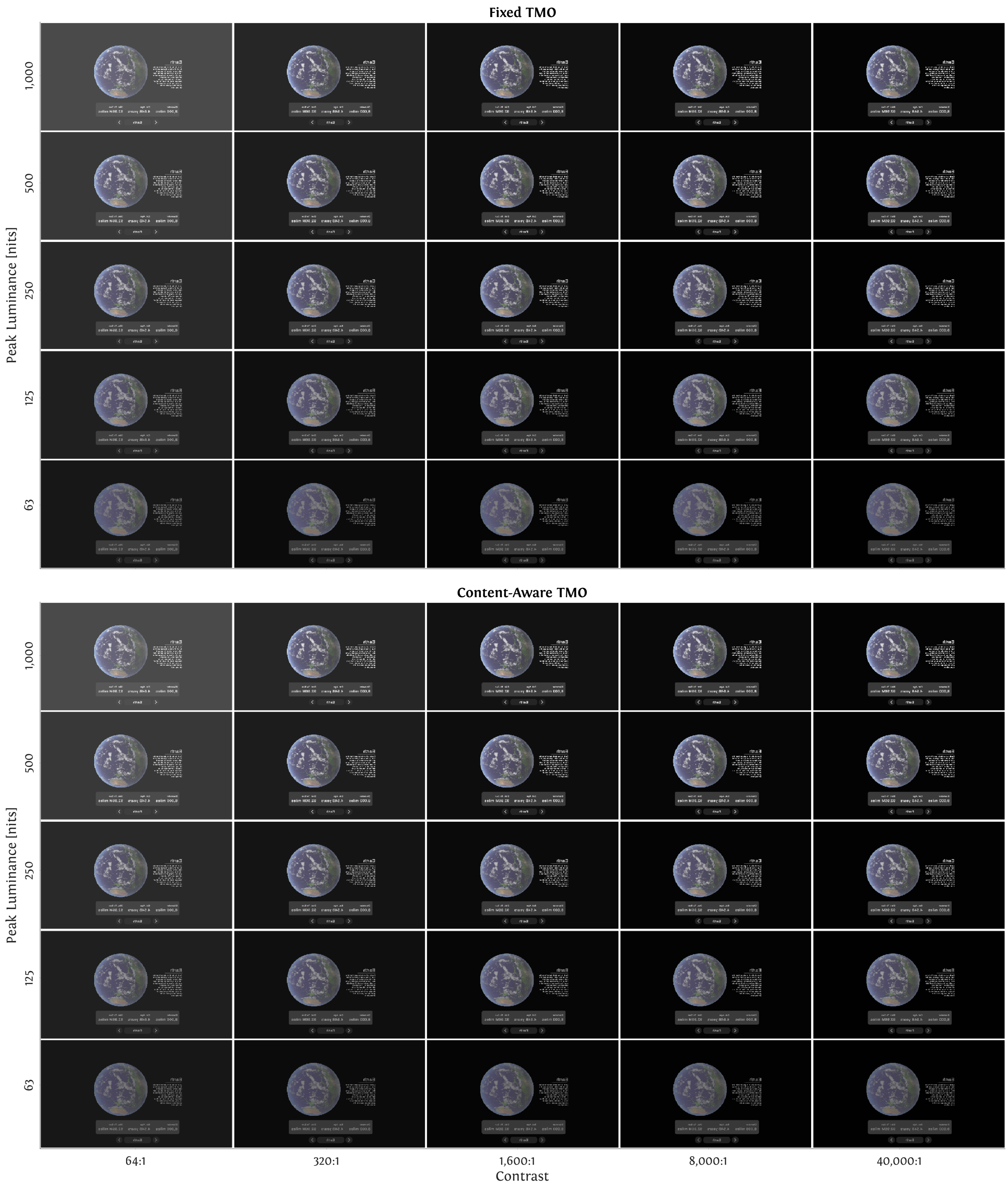


Fig. 17. *Display simulation for a Productivity scene.* Representative frames with TMOs applied for the Earth scene.

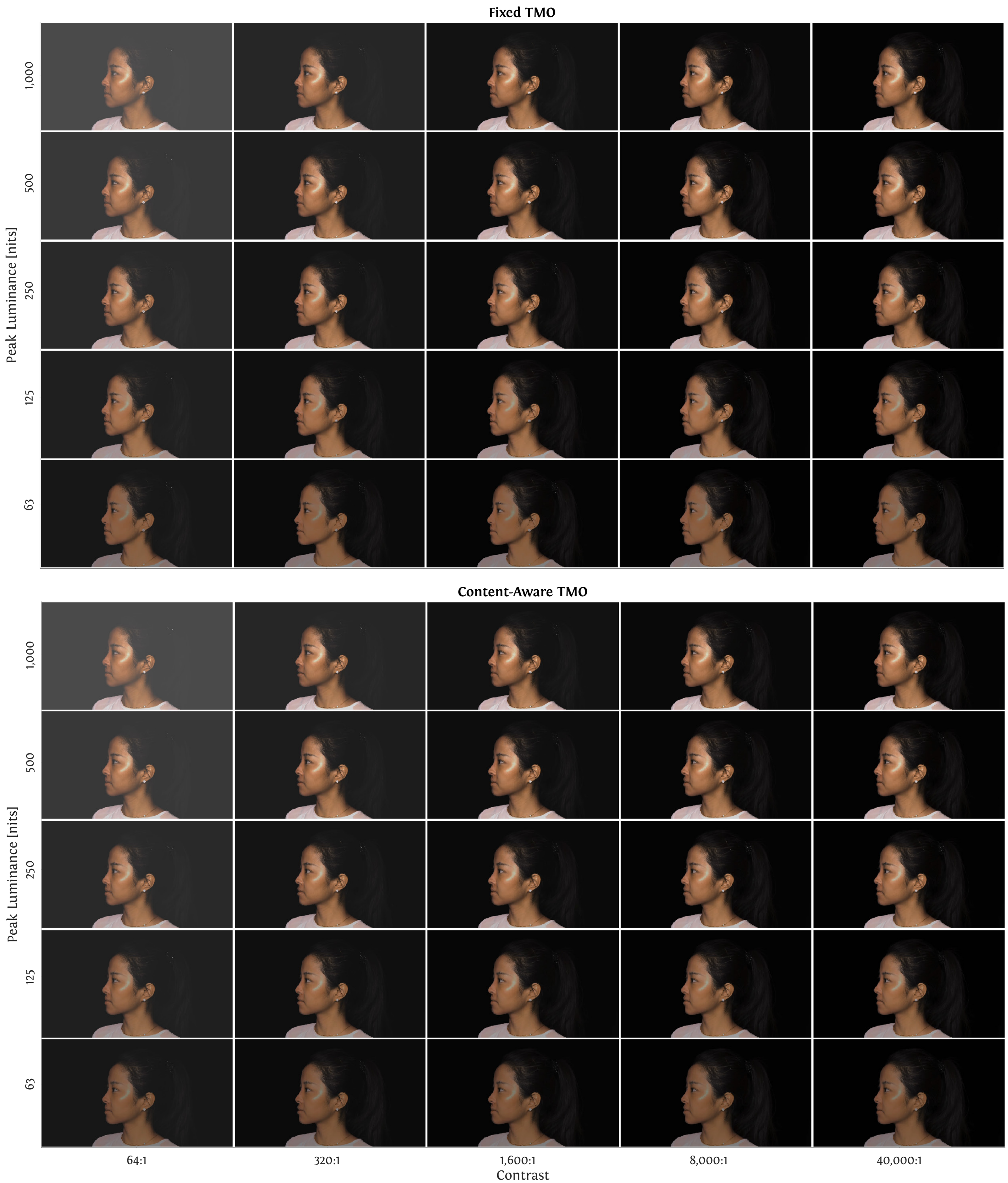


Fig. 18. *Display simulation for a Face scene.* Representative frames with TMOs applied for the Face 3 scene.

Fig. 19. *Display simulation for a UGC/Passthrough scene.* Representative frames with TMOs applied for the Night Street scene. Image credits SJTU.

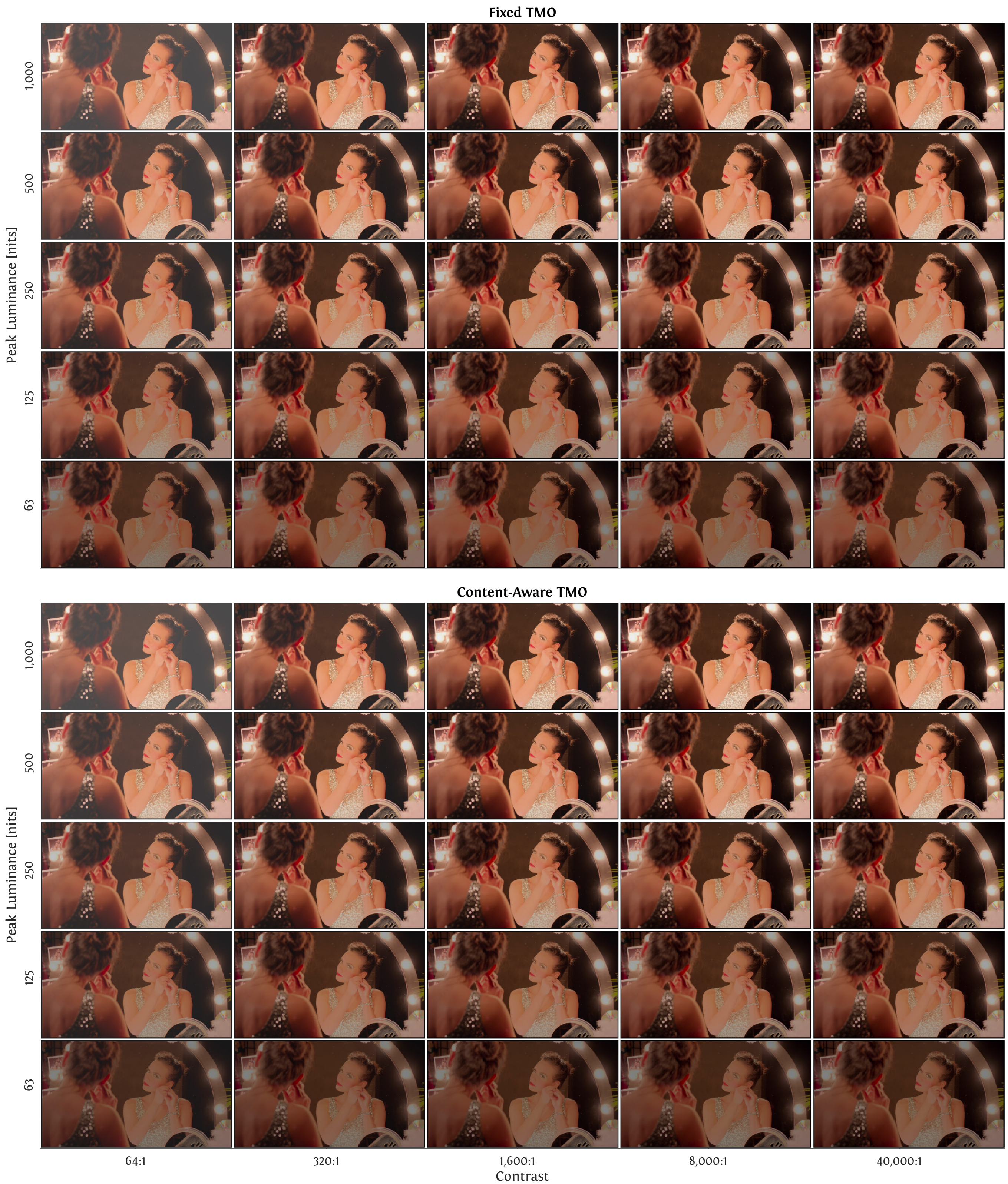


Fig. 20. *Display simulation for an Entertainment scene.* Representative frames with TMOs applied for the Showgirl scene. Image credits HdM-Stuttgart.