This appendix provides additional implementation details to support the reproducibility of our WCG HDR video dataset, display-simulation pipeline, color-enhancement algorithms, perceptual model derivation, and user study designs.

A WCG HDR Video Dataset

This section provides additional details on the WCG HDR video dataset introduced in Section 3.1, including preprocessing steps, scene characteristics, output video metadata, and the computation of color-volume coverage.

A.1 Video Preprocessing

The source videos are available at different resolutions: *Fruits* from Pexels[7] is 4K UHD (3840 × 2160); *Sol Levante* from Netflix Open Content[8] is 1080p (1920 × 1080); *Cosmos Laundromat*, also from Netflix Open Content, is DCI 2K Scope (2048 × 858); *Showgirl* from HdM-HDR-2014 [Froehlich et al. 2014] is 1080p (1920 × 1080); *Forest*, *Drawing*, and *Knitting* from AVT-VQDB-UHD-2-HDR [Keller et al. 2024] are DCI-4K (4096×2160); and the newly created videos (*Bunny*, *Surface*, *Werewolf*, and *Particles*) are rendered and exported at 1080p (1920 × 1080). Source videos not already at 1080p are standardized to 1080p through spatial resampling with the Spline36 filter implemented in `zimg`[9], followed by center cropping when necessary. From each source video, we sample one or more representative clips (3–7 seconds each; see below for details) without scene cuts, ensuring that each clip contains a single continuous shot. This avoids abrupt content changes during stimulus presentation, which could confound observers' color judgments. The original temporal resolution of each clip is preserved (24–60 frames per second, FPS). These processing choices are aligned with the user-study setups and requirements described in Sections 3.3 and 4. Because the source videos collected from existing datasets are not available in stereoscopic format, identical content is presented to the left and right eyes in the stereoscopic OST-AR studies (Section 4).

A.2 Scene Characteristics

Descriptions of the 13 video clips in our dataset are provided below. These clips span real-world footage, cinematic scenes, animated films, and synthetic graphics, with varied color distributions, dynamic ranges, and scene dynamics.

- *Showgirl* (5 seconds, 60 FPS): a cinematic dressing-room shot of a female performer applying makeup, with warm skin tones, saturated reds and cyans, localized highlights, and slow camera/subject motion.
- *Forest* (5 seconds, 60 FPS): an outdoor nature scene of a person in a denim jacket and jeans walking along a forest path, dominated by foliage greens with broad low-to-high luminance variation.
- *Fruits* (6 seconds, 24 FPS): a slow camera pan over an outdoor picnic scene with fruits and food items under bright daylight illumination, featuring high-chroma reds, oranges, and yellows at high luminance levels.
- *Drawing* (5 seconds, 60 FPS): an over-the-shoulder shot of a person sketching on white paper, combining high local contrast around the sketch, dark blue clothing, saturated red socks, and saturated green foliage in the background.
- *Knitting* (5 seconds, 60 FPS): a close-up shot of an elderly woman's hands knitting yellow-orange fabric while wearing a dark blue sweater, emphasizing warm hues and skin tones.
- *Sol Levante 1* (5.7 seconds, 24 FPS): an animated action shot with saturated blue, cyan, and purple energy effects, strong local contrast, rapid subject motion, and broad low-to-high luminance coverage.
- *Sol Levante 2* (3.4 seconds, 24 FPS): an animated volcanic scene with saturated magenta, purple, blue, and orange-red illumination, spanning a wide dynamic range.
- *Sol Levante 3* (3.3 seconds, 24 FPS): an animated aerial panning shot of a bright blue sea, green landscape, and a saturated magenta-purple foreground character at high luminance levels, with broad camera and subject motion.
- *Bunny* (4.2 seconds, 30 FPS): a stylized animation of a bunny running through a grassy field with sky, clouds, and a horizon line in the background, featuring saturated greens, yellow-oranges, and blue-purples, smooth color gradients, and large subject motion.
- *Surface* (5 seconds, 30 FPS): an abstract synthetic scene of multicolored surfaces waving and morphing over time, providing smooth color gradients and broad coverage of highly saturated colors across many hues.
- *Werewolf* (7 seconds, 30 FPS): an animated scene transitioning from dark saturated blues and reds to a moderately bright scene with localized saturated yellow and red highlights, accompanied by large character and camera motion.
- *Particles* (5 seconds, 30 FPS): an abstract particle animation with dense green, magenta, red, and yellow distributions against a dark background, providing saturated color samples and dynamic structure.
- *Cosmos Laundromat* (4 seconds, 24 FPS): an animated fantasy scene with saturated red, magenta, purple, cyan, and green elements, combining stylized rendering with high chroma diversity.

A.3 Perceptual Color Volume

To quantify the color coverage of our dataset, we follow the perceptual color volume (PCV) formulation introduced by Dolby [2023]. PCV operates in ICtCp color space, which is designed to be approximately linear in just-noticeable-difference (JND) units and perceptually uniform for HDR/WCG imagery. ICtCp is defined with respect to the perceptual quantizer (PQ) electro-optical transfer function (EOTF), which maps normalized signal values in $[0, 1]$ to absolute luminance in $[, 10{,}000]$ nits. The component ranges of ICtCp are defined such that the BT.2020 gamut at 10,000-nit peak luminance and 0-nit black level fits within $I \in [0, 1]$ and $C_t, C_p \in [-0.5, 0.5]$. Following the PCV formulation, we convert ICtCp coordinates into JND units by scaling each axis as:

$$\tilde{I} = 720\, I, \quad \tilde{C}_t = 360\, C_t, \quad \tilde{C}_p = 720\, C_p, \tag{19}$$

[7] https://www.pexels.com/
[8] https://opencontent.netflix.com/
[9] https://github.com/sekrit-twc/zimg

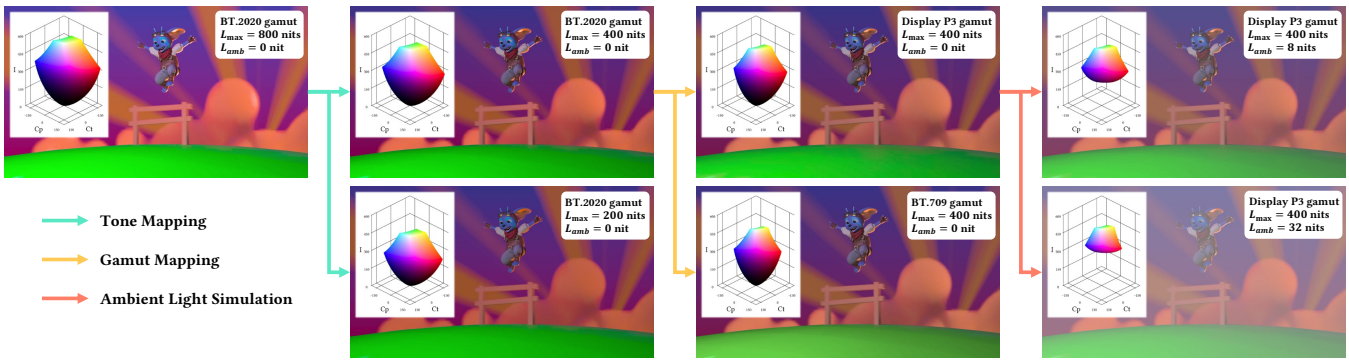

Fig. 13. *Display-simulation pipeline.* Our pipeline combines tone mapping, gamut mapping, and ambient-light modeling to emulate target viewing conditions with varying color gamut, dynamic range, and ambient luminance on a WCG/HDR reference display, avoiding the need to instantiate multiple physical setups. Insets visualize the corresponding effective color gamut in ICtCp color space after accounting for all three factors.

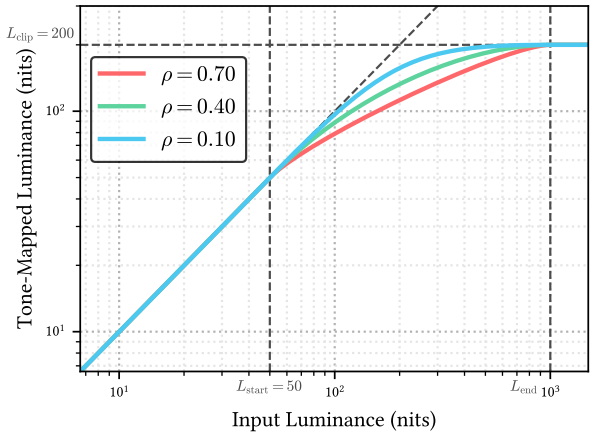

Fig. 14. *Bézier-based tone-mapping curve [Chen et al. 2023].* The curve follows a Bézier formulation and is parameterized by $\theta_{TM} = (\rho, L_{start}, L_{end}, L_{clip})$, where ρ controls the curve slope, and L_{start}, L_{end}, and L_{clip} define the starting, ending, and clipping luminance levels, respectively. Input luminance values below L_{start} are preserved, values between L_{start} and L_{end} are smoothly compressed, and values above L_{end} are clipped to L_{clip}. A larger ρ yields a smoother, more log-linear compression curve that better preserves highlight contrast, whereas a smaller ρ yields a curve closer to clipping. The figure shows example curves for three values of ρ.

This JND-scaled ICtCp space discretizes color volume into unit voxels, and the number of occupied voxels is reported in millions of distinguishable colors (MDCs).

The reference color volume used in Figure 2 is the BT.2020 gamut at 1,000-nit peak luminance and 0.001-nit black level, shown as the dotted outer shell. We first enumerate all integer-bounded voxels in JND-scaled ICtCp space that fall inside this reference color volume. We then aggregate all pixels from all 13 video clips, convert each pixel from BT.2020 RGB to ICtCp, apply the JND-unit scaling, and assign the resulting transformed pixel to its corresponding voxel. A voxel is considered covered if at least one dataset pixel falls inside it. The spatial and temporal subsampling described in Figure 2 is used only for visualization and not in this computation. The color-volume coverage of our dataset relative to the reference color volume is therefore computed as:

$$\text{Coverage} = \frac{|V_{\text{dataset}} \cap V_{\text{BT.2020}}|}{|V_{\text{BT.2020}}|}, \tag{20}$$

where $V_{\text{BT.2020}}$ denotes the set of voxels inside the reference BT.2020 color volume, and V_{dataset} denotes the set of voxels occupied by pixels from our dataset. Based on this procedure, our dataset covers 50% of the reference color volume.

B Tone Mapping

This section details the tone-mapping operator introduced in Section 3.2.1, including its four processing stages, tone-curve parameterization (visualized in Figure 14), and key parameter selection.

Given the target display with black level L_{min} and peak luminance L_{max}, the tone-mapping operator proceeds as follows:

(1) Extract the input tone as the channel-wise maximum of I_{in}. While tone mapping can operate on different luminance proxies, recent work shows that the maximum RGB channel helps avoid out-of-gamut clipping and associated color distortions [Burke et al. 2020; Chen et al. 2025].

$$T_{\text{in}} = \max(I_{\text{in}}^R, I_{\text{in}}^G, I_{\text{in}}^B), \qquad \text{[Tone Extraction]} \tag{21}$$

(2) Apply the tone curve proposed by Chen et al. [2023], $\mathcal{T}(\cdot\,; \theta_{\text{TM}})$, to the input tone T_{in}. The curve follows a Bézier formulation and is parameterized by $\theta_{\text{TM}} = (\rho, L_{\text{start}}, L_{\text{end}}, L_{\text{clip}})$, where ρ controls the curve slope, and L_{start}, L_{end}, and $L_{\text{clip}} = L_{\text{max}}$ define the starting, ending, and clipping luminance levels, respectively (see Figure 14).

$$T_{\text{out}} = \mathcal{T}(T_{\text{in}}; \theta_{\text{TM}}), \qquad \text{[Tone Mapping]} \tag{22}$$

(3) Perform Schlick color correction [1995] to preserve hue by maintaining relative RGB channel ratios.

$$\tilde{I}_{\text{TM}} = (T_{\text{out}}/T_{\text{in}})\, I_{\text{in}}, \qquad \text{[Color Correction]} \tag{23}$$

(4) Simulate the target black level L_{min} through affine luminance rescaling, following Mantiuk et al. [2024].

$$I_{\text{TM}} = L_{\text{min}} + (1 - L_{\text{min}}/L_{\text{max}})\, \tilde{I}_{\text{TM}}, \quad \text{[Black-Level Simulation]} \tag{24}$$

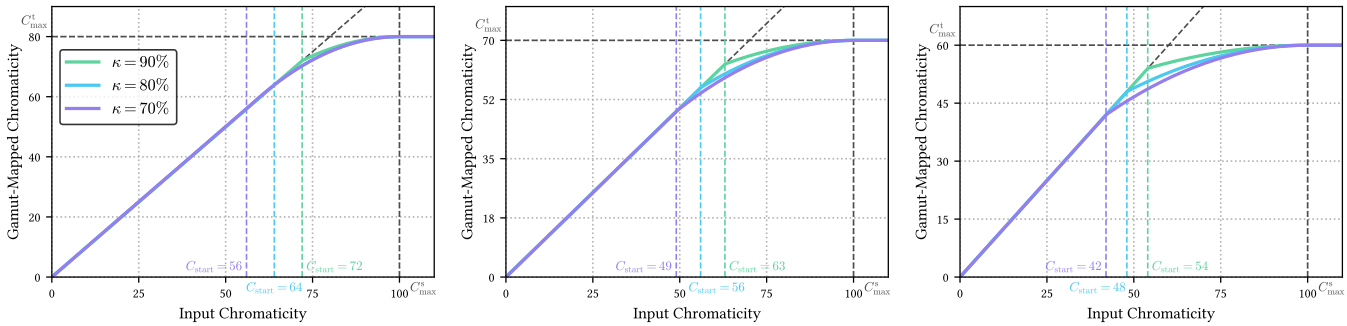

Fig. 15. *Knee functions for smooth gamut mapping [MacDonald et al. 2002].* The function is parameterized by $\theta_{\mathrm{GM}} = (\kappa, C^{\mathrm{s}}_{\mathrm{max}}, C^{\mathrm{t}}_{\mathrm{max}})$, where $C^{\mathrm{s}}_{\mathrm{max}}$ and $C^{\mathrm{t}}_{\mathrm{max}}$ denote the maximum input and output values, respectively, and $\kappa \in [0, 1]$ controls the knee position $C_{\mathrm{start}} = \kappa\, C^{\mathrm{t}}_{\mathrm{max}}$. The function maps an input scalar value C_{in} to an output value C_{out}: values below C_{start} are preserved, values between C_{start} and $C^{\mathrm{s}}_{\mathrm{max}}$ are smoothly compressed toward $C^{\mathrm{t}}_{\mathrm{max}}$, and values above $C^{\mathrm{s}}_{\mathrm{max}}$ are clipped to $C^{\mathrm{t}}_{\mathrm{max}}$. The three panels illustrate different target maxima $C^{\mathrm{t}}_{\mathrm{max}}$, and the curves within each panel show different knee positions controlled by κ.

Following Tariq et al. [2023], we perform scene-adaptive tone-curve adjustment by optimizing the starting luminance L_{start} of the nonlinear compression region to minimize contrast distortions. To improve temporal stability, we optimize L_{start} only at keyframes sampled at 2 FPS and interpolate the optimized values in the log-luminance domain, yielding temporally smooth tone-mapping parameters for intermediate frames. This simple temporal smoothing strategy produced no visible temporal artifacts in our processed videos while maintaining high computational efficiency. For videos containing more complex and dynamic content, more sophisticated temporally coherent tone-mapping methods [Aydin et al. 2014; Boitard et al. 2012] may be required. Following Chen et al. [2025], we compute the slope parameter ρ using a heuristic based on the input tone distribution:

$$\rho = \frac{1}{|\Omega|} \sum_{\mathbf{x} \in \Omega} \mathbb{1}(T_{\mathrm{in}}(\mathbf{x}) > L_{\mathrm{start}}), \tag{25}$$

where Ω denotes the set of pixels in the input frame, $T_{\mathrm{in}}(\mathbf{x})$ denotes the input tone at pixel $\mathbf{x}$, and $\mathbb{1}(\cdot)$ denotes the indicator function. Intuitively, ρ measures the fraction of pixels whose input tone exceeds the optimized starting luminance L_{start}. A larger value indicates that highlights occupy a larger portion of the frame, yielding a smoother, more log-linear compression curve that better preserves highlight contrast. Conversely, a smaller value indicates sparse highlights and yields a curve closer to clipping.

C Gamut Mapping

This section describes the knee-function formulation used by the gamut-mapping algorithms introduced in Section 3.2.2.

We adopt the knee function $\mathcal{S}(\cdot\,; \theta_{\mathrm{GM}})$ introduced by MacDonald et al. [2002]. The function is parameterized by $\theta_{\mathrm{GM}} = (\kappa, C^{\mathrm{s}}_{\mathrm{max}}, C^{\mathrm{t}}_{\mathrm{max}})$, where $C^{\mathrm{s}}_{\mathrm{max}}$ and $C^{\mathrm{t}}_{\mathrm{max}}$ denote the maximum input and output values, respectively, and $\kappa \in [0, 1]$ controls the knee position $C_{\mathrm{start}} = \kappa\, C^{\mathrm{t}}_{\mathrm{max}}$ (refer to Figure 15). The function maps an input scalar value C_{in} to an output value C_{out}, preserving values below C_{start} and smoothly compressing values above C_{start} toward $C^{\mathrm{t}}_{\mathrm{max}}$. Let

$$\zeta = \frac{1}{2} \frac{C_{\mathrm{in}} - C_{\mathrm{start}}}{C^{\mathrm{s}}_{\mathrm{max}} - C_{\mathrm{start}}}, \tag{26}$$

The knee function $C_{\mathrm{out}} = \mathcal{S}(C_{\mathrm{in}}; \theta_{\mathrm{GM}})$ is defined as:

$$C_{\mathrm{out}} = \begin{cases} C_{\mathrm{in}}, & C_{\mathrm{in}} \le C_{\mathrm{start}}, \\ C_{\mathrm{start}} + 4(C^{\mathrm{t}}_{\mathrm{max}} - C_{\mathrm{start}})(\zeta - \zeta^2), & C_{\mathrm{start}} < C_{\mathrm{in}} < C^{\mathrm{s}}_{\mathrm{max}}, \\ C^{\mathrm{t}}_{\mathrm{max}}, & C_{\mathrm{in}} \ge C^{\mathrm{s}}_{\mathrm{max}}. \end{cases} \tag{27}$$

Thus, C_{start} controls the extent of the unchanged region: values below C_{start} are preserved, values between C_{start} and $C^{\mathrm{s}}_{\mathrm{max}}$ are smoothly compressed, and values above $C^{\mathrm{s}}_{\mathrm{max}}$ are clipped to $C^{\mathrm{t}}_{\mathrm{max}}$.

In our implementation of **RGB-SC**, we define $\kappa = 0.8$, $C^{\mathrm{s}}_{\mathrm{max}} = 1.25\, L_{\mathrm{max}}$, and $C^{\mathrm{t}}_{\mathrm{max}} = L_{\mathrm{max}}$. For **ICtCp-CHL**, we also set $\kappa = 0.8$, while $C^{\mathrm{s}}_{\mathrm{max}}(L, h)$ and $C^{\mathrm{t}}_{\mathrm{max}}(L, h)$ denote the hue- and luminance-dependent maximum chroma values of the source (BT.2020) and target (G_{target}) gamuts, respectively. These values are obtained from the precomputed 2D lookup tables (LUTs) described in Section 3.2.2.

D Direct-View User Study

This section provides additional details on the stimuli selection process for the direct-view user study (Section 3.3) and reports the full ANOVA statistics (Table 5). We first summarize the display types considered in our study (Appendix D.1), then analyze ambient-luminance distributions in indoor and outdoor viewing environments (Appendix D.2). Finally, we describe how peak-luminance and ambient-luminance levels were sampled to define the study stimuli (Appendix D.3).

D.1 Display Types

We considered display types commonly used in practical applications, including laptops, phones/tablets, desktop monitors, televisions, cinema displays, VR headsets, and OST-AR glasses. Table 1 summarizes representative models in each category and their key color-related characteristics.

D.2 Ambient-Luminance Distributions

We used the indoor and outdoor ambient-luminance distributions reported by Zhang et al. [2023] to estimate representative ambient-luminance quantiles. Rather than using incident ambient luminance

Table 1. *Representative models of common display types and their color-related characteristics.* Peak luminance is reported as sustained full-screen luminance rather than transient or local peak luminance. Contrast is defined as the ratio between peak luminance and minimum luminance, i.e., black level. For all models except OST-AR glasses, contrast is reported under ideal near-dark viewing conditions. For OST-AR glasses, contrast is typically close to or below 10:1 [Chapiro et al. 2024], because these devices are driven as a function of ambient luminance rather than continuously at peak luminance to improve power efficiency. Unavailable exact values are denoted as N/A.

Category	Model	Peak Luminance (nits)	Color Gamut	Contrast	Technology
Laptop	MacBook Pro 2024	1,000	Display P3	1,000,000:1	Mini-LED
Phone & Tablet	iPhone 17 Pro	1,000	Display P3	2,000,000:1	OLED
	iPad Pro 2022	1,000	Display P3	1,000,000:1	LCD (IPS)
Desktop Monitor	Dell U2723QE	400	Custom (98% of DCI-P3)	2,000:1	LCD (IPS)
	ASUS PG27AQDM	400	Custom (99% of DCI-P3)	1,500,000:1	W-OLED
	Apple Pro Display XDR	1,000	Display P3	1,000,000:1	LCD (IPS)
	Apple Studio Display	600	Display P3	1,200:1	LCD (IPS)
Television	TCL S325	200	BT.709	3,500:1	LCD (VA)
	Samsung S95C	250	Custom (99% of DCI-P3)	1,000,000:1	QD-OLED
Cinema Display	Dolby Cinema	106	Custom (122% of DCI-P3)	7,500:1	Laser Projectors
VR Headset	Meta Quest 3	100	Custom (99% of BT.709)	64:1	LCD
	Meta Quest Pro	100	Custom (99% of BT.709)	64:1	Mini-LED
	Apple Vision Pro	100	Custom (92% of DCI-P3)	100:1	Micro-OLED
OST-AR Glasses	Meta Orion AI Glasses	N/A	N/A	N/A	Micro-LED
	Meta Ray-Ban Display	5,000	N/A	N/A	LCoS

Table 2. *Estimated ambient-luminance quantiles for indoor and outdoor viewing environments.* Incident ambient luminance is estimated from the distributions reported by Zhang et al. [2023] using log-domain interpolation within each luminance bin. For conventional flat-panel displays, eye-reaching ambient luminance is estimated using 1–5% panel reflectance; for OST-AR displays, it is estimated using 20–90% optical transmittance.

Quantile	Indoor Ambient Luminance (nits)					Outdoor Ambient Luminance (nits)				
	Incident	Flat 1%	Flat 5%	AR 20%	AR 90%	Incident	Flat 1%	Flat 5%	AR 20%	AR 90%
10%	1.67	0.017	0.084	0.33	1.51	2.15	0.022	0.11	0.43	1.94
20%	2.72	0.027	0.14	0.54	2.45	13.6	0.14	0.68	2.71	12.2
30%	4.23	0.042	0.21	0.85	3.80	56.2	0.56	2.81	11.2	50.6
40%	5.83	0.058	0.29	1.17	5.25	158.5	1.59	7.93	31.7	143
50%	8.41	0.084	0.42	1.68	7.57	299	2.99	14.9	59.7	269
60%	12.3	0.12	0.61	2.46	11.1	531	5.31	26.5	106	478
70%	18.2	0.18	0.91	3.63	16.3	1,000	10.0	50.0	200	900
80%	28.1	0.28	1.40	5.62	25.3	1,995	20.0	99.8	399	1,796
90%	50.5	0.51	2.53	10.1	45.5	4,786	47.9	239	957	4,308
100%	251	2.51	12.6	50.2	226	10,000	100	500	2,000	9,000

Table 3. *Estimated additive contrast for flat-panel displays in indoor and outdoor viewing environments.* Eye-reaching ambient luminance is estimated from the incident ambient luminance in Table 2 by assuming 3% panel reflectance. Additive contrast is computed using Equation (8) with $L_{\text{min}} = 0$ nit for representative peak-luminance settings: 200 nits (SDR) and 1,000 nits (HDR). Cell colors indicate additive-contrast ranges: green > 1,000:1, yellow 100:1–1,000:1, orange 10:1–100:1, and red < 10:1.

Environment	Peak Luminance	Ambient-Luminance Quantile									
		10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
Indoor	200 nits (SDR)	3,993:1	2,452:1	1,577:1	1,145:1	794:1	543:1	367:1	238:1	133:1	28:1
	1,000 nits (HDR)	19,961:1	12,256:1	7,881:1	5,719:1	3,965:1	2,711:1	1,833:1	1,187:1	661:1	134:1
Outdoor	200 nits (SDR)	3,102:1	491:1	120:1	43:1	23:1	14:1	8:1	4:1	2:1	2:1
	1,000 nits (HDR)	15,505:1	2,452:1	594:1	211:1	112:1	64:1	34:1	18:1	8:1	4:1

Table 4. *Additive contrast of sampled direct-view user-study conditions.* Additive contrast is computed using Equation (8) with $L_{\text{min}} = 0.001$ nit for two peak-luminance settings, 200 nits (SDR) and 900 nits (HDR), across the sampled eye-reaching ambient luminance levels. Cell colors follow the convention in Table 3: green > 1,000:1, yellow 100:1–1,000:1, orange 10:1–100:1, and red < 10:1. The gray cell denotes the excluded SDR condition at $L_{\text{amb}} = 100$ nits.

Peak Luminance	Eye-Reaching Ambient Luminance L_{amb} (nits)					
	0	**1**	**4**	**16**	**64**	**100**
200 nits (SDR)	200,001:1	201:1	51:1	13.5:1	4.125:1	3:1
900 nits (HDR)	900,001:1	901:1	226:1	57.25:1	15.06:1	10:1

directly, we estimated the ambient light reaching the observer's eyes by accounting for display-dependent reflection or transmission: 1–5% panel reflectance for conventional flat-panel displays and 20–90% optical transmittance for OST-AR displays with transition lenses. The 1% lower bound for panel reflectance is motivated by the on-axis reflectivity of flat-panel displays with anti-reflective coatings, such as 1.65% for Apple Pro Display XDR, 1.5% for Samsung Galaxy S24 Ultra, and 1.1% for LG C3 OLED. For flat-panel displays without anti-reflective coatings, on-axis reflectivity can approach 5%, such as 4.8% for iPhone 15 Pro Max and 4.8% for Google Pixel 8 Pro. A 20–90% optical-transmittance range is assumed for OST-AR displays with transition lenses. Table 2 summarizes the quantiles for indoor and outdoor viewing environments and for both display classes.

D.3 Direct-View User-Study Conditions

With the ambient-luminance quantiles summarized in Table 2, we computed the corresponding additive contrast for each display class using Equation (8). For flat-panel displays commonly used under non-negligible ambient illumination, including laptops, phones/tablets, desktop monitors, and televisions, we considered two representative peak-luminance settings, 200 nits (SDR) and 1,000 nits (HDR), based upon the display characteristics summarized in Table 1. Table 3 reports the computed additive contrast for these settings. For display classes typically used under near-dark viewing conditions, such as cinema displays and VR headsets, additive contrast is expected to remain close to the native contrast reported in Table 1. In contrast, OST-AR displays often operate at much lower additive contrast (close to or below 10:1 [Chapiro et al. 2024]), because these devices are driven as a function of ambient luminance rather than continuously at peak luminance to improve power efficiency.

Based on these additive-contrast estimates and a pilot study with two expert participants, we sampled six eye-reaching ambient luminance levels for L_{amb}: 0, 1, 4, 16, 64, and 100 nits. These conditions cover more than the 90th percentile of eye-reaching ambient luminance for both flat-panel and OST-AR displays in indoor viewing environments, and approximately the 80th percentile for flat-panel displays and the 60th percentile for OST-AR displays in outdoor viewing environments (see Table 2). We used two peak-luminance settings: 200 nits (SDR) and 900 nits (HDR). The HDR peak luminance was limited to 900 nits because both the target display emission and the eye-reaching ambient component were simulated on the same EIZO display; therefore, the sum of the simulated display peak luminance L_{max} and eye-reaching ambient luminance L_{amb} must not exceed the EIZO peak luminance, 1,000 nits. The target black level L_{min} was set to the EIZO black level, 0.001 nit.

Table 5. *N-way ANOVA results for the direct-view user study.* The dependent variable is perceptual quality, measured in just-objectionable-difference (JOD) units [Perez-Ortiz et al. 2019]. Factors include scene, peak luminance, color gamut, ambient luminance, and gamut-mapping algorithm (GMA). The table reports main effects and two-way interaction effects. SS denotes sum of squares, and df denotes degrees of freedom.

Effect	**SS**	**df**	***F***	***p***
Scene	2289.8	8	555.09	< .001
Peak Luminance	4.91	1	9.52	.002
Color Gamut	92.73	2	89.92	< .001
Ambient Luminance	1999.4	3	1292.6	< .001
GMA	27.14	1	52.64	< .001
Scene × Peak Luminance	758.16	8	183.80	< .001
Scene × Color Gamut	44.38	16	5.38	< .001
Scene × Ambient Luminance	118.55	24	9.58	< .001
Scene × GMA	14.41	8	3.49	< .001
Peak Luminance × Color Gamut	0.15	2	0.14	.867
Peak Luminance × Ambient Luminance	18.71	3	12.09	< .001
Peak Luminance × GMA	0.20	1	0.38	.537
Color Gamut × Ambient Luminance	14.88	6	4.81	< .001
Color Gamut × GMA	5.39	2	5.23	.006
Ambient Luminance × GMA	1.78	3	1.15	.329

Table 4 summarizes the additive contrast for all sampled conditions. These conditions span an additive-contrast range of 4.125:1–900,001:1, covering the various displays discussed in Table 1. The 100-nit ambient condition was excluded for the 200-nit peak setting because its additive contrast is similar to that of the 64-nit condition. The 0-nit condition isolates intrinsic perceptual differences across color gamuts under near-dark viewing. Although the EIZO display cannot reproduce the full BT.2020 gamut, more than 93% of the ambient-blended BT.2020 gamut falls within the EIZO gamut under 64–100 nits of ambient luminance, enabling its simulation. This resulted in 5 SDR and 6 HDR ambient conditions for each of the EIZO, Display P3, and BT.709 gamuts, and 1 SDR and 2 HDR ambient conditions for the BT.2020 gamut, yielding 36 conditions per scene and GMA, and 648 stimuli in total (9 scenes × 2 GMAs × 36 conditions).

E Perceptual Model

This section provides additional details on the perceptual-model formulation introduced in Section 3.4 and reports the fitted model parameters (Table 6).

Motivated by the strong agreement between perceptual color volume (MDC) and perceived color quality (JOD), as indicated by consistently high linear, rank-based, and distribution-sensitive correlations (refer to Section 3.4 for details), we develop an analytical model that predicts perceived color quality from perceptual color volume. As illustrated in Figure 5, perceived color quality increases with perceptual color volume and gradually saturates as the test color volume approaches the reference color volume, suggesting a quasi power-law relationship. To model this behavior, we evaluate

Table 6. *Fitted parameters for the selected power-model formulation.*

GMA	Peak Luminance	k_1	k_2	k_3
RGB-SC	200 nits (SDR)	4.7263	0.2259	−8.4946
	900 nits (HDR)	2.5812	0.3819	−7.8089
ICtCp-CHL	200 nits (SDR)	2.8183	0.3304	−6.8539
	900 nits (HDR)	2.4101	0.3923	−7.7905

three candidate functional forms—power, logarithmic, and rational—and fit each of them using nonlinear least squares.

Although peak luminance is incorporated into the MDC computation, Figure 5 shows that the JOD–MDC relationship remains luminance-dependent. Because the two sampled peak-luminance settings are insufficient to directly fit a robust unified luminance-dependent model, we first build separate models for each peak-luminance setting and then merge the fitted parameters through log-luminance interpolation. The resulting unified formulations for the three candidate models are:

$$Q_{\text{power}}(V, L) = k_1(L)\, V^{\,k_2(L)} + k_3(L), \tag{28}$$

$$Q_{\log}(V, L) = k_1(L)\, \log(V + \epsilon) + k_2(L), \tag{29}$$

$$Q_{\text{rational}}(V, L) = \frac{k_1(L)\, V}{V + k_2(L)} + k_3(L), \tag{30}$$

with luminance-dependent parameters defined by

$$k_i(L) = (1 - \alpha) k_i^{\text{SDR}} + \alpha k_i^{\text{HDR}}, \tag{31}$$

$$\alpha = \frac{\log\left(L / L_{\max}^{\text{SDR}}\right)}{\log\left(L_{\max}^{\text{HDR}} / L_{\max}^{\text{SDR}}\right)}, \quad L \in [L_{\max}^{\text{SDR}}, L_{\max}^{\text{HDR}}]. \tag{32}$$

Here, Q denotes perceived color quality in JOD units, and V denotes perceptual color volume in MDCs. $L_{\max}^{\text{SDR}}$ and $L_{\max}^{\text{HDR}}$ denote the two peak-luminance levels used for fitting: 200 nits for SDR and 900 nits for HDR. For the power and rational models, $i \in \{1, 2, 3\}$; for the logarithmic model, $i \in \{1, 2\}$. We set $\epsilon = 10^{-6}$ in the logarithmic model for numerical stability.

We fit separate models for the two GMAs. For **ICtCp-CHL**, the power model achieved an RMSE of 0.2157 and $R^2 = 0.9762$, outperforming the logarithmic model (RMSE 0.2703, $R^2 = 0.9627$) and the rational model (RMSE 0.2311, $R^2 = 0.9727$). For **RGB-SC**, the power model also achieved the best fit, with an RMSE of 0.2374 and $R^2 = 0.9729$, outperforming the logarithmic model (RMSE 0.2798, $R^2 = 0.9624$) and slightly improving over the rational model (RMSE 0.2402, $R^2 = 0.9723$).

Model selection was performed using the Akaike Information Criterion (AIC), a metric used to evaluate and compare statistical models by balancing goodness of fit against model complexity. For **RGB-SC**, the power, logarithmic, and rational models achieved AIC values of −85.77, −78.62, and −84.98, respectively. For **ICtCp-CHL**, the corresponding AIC values were −92.29, −80.96, and −87.61. Although the rational model achieved comparable RMSE and R^2, its higher AIC suggests mild over-parameterization without a commensurate gain in explanatory power. The logarithmic model produced consistently higher error, indicating insufficient flexibility to capture the observed saturation behavior. Based on these results, we adopt the power model throughout this work.

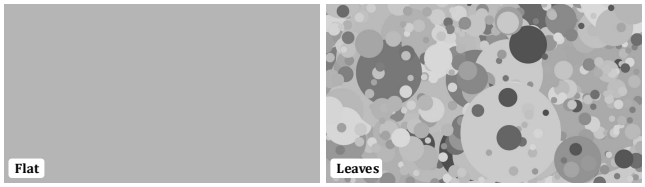

Fig. 16. *Backgrounds used in the OST-AR color-quality study (Section 4.1).* We used two backgrounds: *flat*, a uniform D65 white background, and *leaves*, a high-contrast leaves pattern following Chapiro et al. [2024]. Both backgrounds were presented with mean luminance calibrated such that, after beam-splitter and ND-filter transmission, the light reaching the observer produced the specified ambient luminance.

Table 6 reports the fitted parameters of the power model for each GMA. These parameters define the SDR and HDR endpoints used in the log-luminance interpolation in Equations (31) and (32). The exponent parameter k_2 is below 1 in all cases, consistent with the observed saturating relationship between perceptual color volume and perceived color quality: increasing perceptual color volume yields progressively smaller quality gains as the test color volume approaches the reference color volume.

F Color Enhancement

This section provides additional implementation details and parameter settings for the three color-enhancement algorithms introduced in Section 3.5.

For **RGB-CE**, we set the scaling factor $s \in [1, 2]$ as a function of additive contrast $c_{\text{add}} = (L_{\max} + L_{\text{amb}})/(L_{\min} + L_{\text{amb}})$:

$$s = 1.45 - 0.025\, c_{\text{add}}, \tag{33}$$

For **Schlick-CC**, we define the compensated ambient term L_{comp} as a fraction of the display peak luminance $L_{\max}$, with the fraction decreasing log-linearly with $L_{\max}$:

$$L_{\text{comp}} = \min\left(L_{\text{amb}},\ L_{\max}\left(0.05 - 0.03 \log\left(\frac{L_{\max}}{70}\right)\right)\right), \tag{34}$$

This makes the compensated ambient term display-luminance dependent while gradually reducing the compensated fraction at higher peak luminance, mitigating over-enhancement and chromatic artifacts.

For **ICtCp-CE**, we similarly define L_{comp} as a fraction of $L_{\max}$, but include a scene-dependent scaling factor $p \in [0, 1]$ to adapt to the color characteristics of individual scenes:

$$L_{\text{comp}} = \min\left(L_{\text{amb}},\ p\, L_{\max}\left(1.0 - 0.4 \log\left(\frac{L_{\max}}{70}\right)\right)\right). \tag{35}$$

For the OST-AR color-enhancement study in Section 4.2, we set p to 0.03, 0.04, 0.10, and 0.12 for *Fruits*, *Drawing*, *Cosmos Laundromat*, and *Sol Levante 3*, respectively.

Unlike **RGB-CE** and **Schlick-CC**, **ICtCp-CE** does not require an additional dynamic-range clipping step after chroma expansion. The chroma-expansion operation is bounded by the effective color gamut after accounting for the compensated ambient term; therefore, the transformed pixels remain within the target display range by construction.

Table 7. *Additive contrast of sampled OST-AR color-quality study conditions.* Additive contrast is computed using Equation (8) with $L_{\text{min}} = 0.0008$ nit (the EIZO black level after accounting for the 80% beam-splitter reflectance) for two peak-luminance settings, 200 nits (SDR) and 800 nits (HDR), across the sampled eye-reaching ambient luminance levels. Cell colors follow the convention in Table 3: green > 1,000:1, yellow 100:1–1,000:1, orange 10:1–100:1, and red < 10:1.

Peak Luminance	Ambient Luminance L_{amb} (nits)			
	0.01	**5**	**25**	**100**
200 nits (SDR)	18,520:1	41:1	9:1	3:1
800 nits (HDR)	74,075:1	161:1	33:1	9:1

Table 8. *N-way ANOVA results for the OST-AR color-quality study.* The dependent variable is perceptual quality, measured in just-objectionable-difference (JOD) units [Perez-Ortiz et al. 2019]. Factors include scene, peak luminance, color gamut, ambient luminance, and background type. The table reports main effects and two-way interaction effects. SS denotes sum of squares, and df denotes degrees of freedom.

Effect	**SS**	**df**	***F***	***p***
Scene	1.65	3	4.37	.007
Peak Luminance	0.66	1	5.29	.025
Color Gamut	0.93	1	7.44	.008
Ambient Luminance	109.01	2	433.71	< .001
Background Type	7.30	1	58.07	< .001
Scene × Peak Luminance	0.89	3	2.35	.080
Scene × Color Gamut	0.75	3	2.00	.123
Scene × Ambient Luminance	3.86	6	5.12	< .001
Scene × Background Type	0.51	3	1.35	.266
Peak Luminance × Color Gamut	0.01	1	0.06	.801
Peak Luminance × Ambient Luminance	0.17	2	0.68	.509
Peak Luminance × Background Type	0.54	1	4.31	.042
Color Gamut × Ambient Luminance	0.06	2	0.24	.786
Color Gamut × Background Type	0.13	1	1.05	.310
Ambient Luminance × Background Type	2.26	2	9.00	< .001

G User Study: Perceived Color Quality under OST-AR Viewing

This section provides additional details on the stimuli selection process for the OST-AR color-quality study (Section 4.1) and reports the full ANOVA statistics (Table 8).

Based on the additive-contrast estimates for representative display types in indoor and outdoor viewing environments (Appendix D.3), we sampled four eye-reaching ambient luminance levels for L_{amb}: 0.01, 5, 25, and 100 nits. The 0.01-nit level was the lowest luminance achievable with the DynaScan display after accounting for the beam-splitter and ND-filter transmittance. These levels cover more than the 90th percentile of eye-reaching ambient luminance for OST-AR displays in indoor viewing environments, and approximately the 60th percentile in outdoor viewing environments (see Table 2). We used two peak-luminance settings: 200 nits (SDR) and 800 nits (HDR). The HDR peak luminance was limited to 800 nits because video luminance values were scaled by 1.25 before presentation to compensate for the 80% beam-splitter reflectance; after this scaling, the displayed peak luminance must not exceed the 1,000-nit peak luminance of the EIZO displays. The target black level L_{min} was set to the EIZO black level after accounting for the beam-splitter reflectance, 0.0008 nit. Table 7 summarizes the additive contrast for all sampled conditions, which span 3:1–74,075:1.

Table 9. *Peak luminance required for sampled OST-AR color-enhancement viewing conditions.* Peak luminance values are computed from the sampled additive-contrast and ambient-luminance levels using Equation (8). Gray cells denote excluded conditions: $c_{\text{add}} = 4{:}1$ with $L_{\text{amb}} = 10$ nits requires an impractically low peak luminance, whereas $c_{\text{add}} = 12{:}1$ with $L_{\text{amb}} = 100$ nits requires an infeasibly high peak luminance that exceeds the EIZO display capability after accounting for the 80% beam-splitter reflectance.

Additive Contrast	Ambient Luminance (nits)		
	10	**40**	**100**
4:1	30	120	300
8:1	70	280	700
12:1	110	440	1,100

Table 10. *N-way ANOVA results for the OST-AR color-enhancement study.* The dependent variable is perceptual quality, measured in just-objectionable-difference (JOD) units [Perez-Ortiz et al. 2019]. Factors include scene, viewing condition, and enhancement method. Viewing condition denotes the sampled additive-contrast × ambient-luminance combinations. Enhancement method includes the no-enhancement baseline and the three color-enhancement algorithms. The table reports main effects and two-way interaction effects. SS denotes sum of squares, and df denotes degrees of freedom.

Effect	**SS**	**df**	***F***	***p***
Scene	4.75	3	10.49	< .001
Viewing Condition	7.17	6	7.91	< .001
Enhancement Method	7.84	3	17.31	< .001
Scene × Viewing Condition	6.98	18	2.57	.004
Scene × Enhancement Method	3.13	9	2.30	.029
Viewing Condition × Enhancement Method	5.28	18	1.94	.031

Two backgrounds were used (see Figure 16): *flat*, a uniform D65 white background, and *leaves*, a high-contrast leaves pattern following Chapiro et al. [2024]. Both backgrounds were presented with mean luminance calibrated such that, after beam-splitter and ND-filter transmission, the light reaching the observer produced the specified ambient luminance. This resulted in 4 SDR and 4 HDR ambient conditions for each of the Display P3 and BT.709 gamuts, yielding 16 conditions per scene and background, and 128 stimuli in total (4 scenes × 2 backgrounds × 16 conditions).

We used the same 2IFC protocol and ASAP framework described in Section 3.3 to schedule pairwise comparisons. Comparisons were allowed across color gamuts, ambient luminance, and backgrounds, but not across scenes or peak luminance. For each scene–peak-luminance block, the reference was set to the Display P3, zero-ambient, *flat* condition. Its *leaves* counterpart, which produced the same appearance under zero ambient luminance, was excluded to reduce experiment duration.

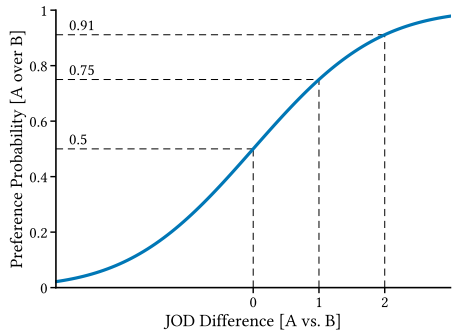

Fig. 17. *JOD difference vs. observer preference probability.* Just-objectionable-difference (JOD) [Perez-Ortiz et al. 2019] defines a relative quality scale in which score differences correspond to observer preference probabilities. A difference of 1 JOD between two conditions indicates that 75% of observers prefer the higher-scoring condition, 2 JODs correspond to 91% preference, and so on.

H User Study: Color Enhancement under OST-AR Viewing

This section provides additional details on the stimuli selection process for the OST-AR color-enhancement study (Section 4.2) and reports the full ANOVA statistics (Table 10).

OST-AR displays often operate at relatively low additive contrast (close to or below 10:1 [Chapiro et al. 2024]), because these devices are commonly used in bright indoor or outdoor environments and are driven as a function of ambient luminance rather than continuously at their peak luminance to improve power efficiency. Based on these considerations, we sampled three additive-contrast levels, 12:1, 8:1, and 4:1, and three eye-reaching ambient luminance levels, 10, 40, and 100 nits. The corresponding peak luminance was computed from the sampled additive-contrast and ambient-luminance levels using Equation (8). Two additive-contrast × ambient-luminance combinations were excluded: c_{add} = 4:1 with L_{amb} = 10 nits required an impractically low peak luminance, whereas c_{add} = 12:1 with L_{amb} = 100 nits required an infeasibly high peak luminance exceeding the EIZO display capability after accounting for the 80% beam-splitter reflectance. The target black level L_{min} was set to the EIZO black level after accounting for the beam-splitter reflectance, 0.0008 nit. Table 9 summarizes the peak luminance values for all sampled viewing conditions.

In addition to the three color-enhancement algorithms, we included a no-enhancement baseline in which no ambient-light compensation was applied. For the enhancement algorithms, the target display was assumed to have a Display P3 gamut, providing additional gamut headroom to compensate for ambient-induced color-fidelity losses. This setup reflects practical scenarios where visual content is mastered for a standard gamut such as BT.709, while display devices use custom color primaries that exceed these standards (e.g., the EIZO gamut shown in Figure 4, left). As in Section 4.1, video luminance values were scaled by 1.25 before presentation. The *flat* background, a uniform D65 white background, was used for all conditions. Its luminance was calibrated such that, after beam-splitter and ND-filter transmission, the light reaching the observer produced the specified ambient luminance. In total, this yielded 112 stimuli: 4 scenes × 4 enhancement methods (including the no-enhancement baseline) × 7 viewing conditions (3 additive contrast levels × 3 ambient luminance levels - 2 excluded combinations).