Towards Physically-Based Sky-Modeling For Image Based Lighting

Ian J. Maquignaz

Université Laval

Québec, Québec

Canada

ian.maquignaz.1@ulaval.ca
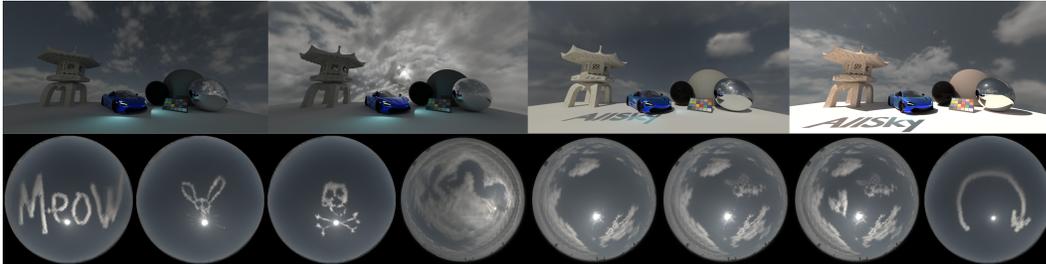
Figure 1. IBL renders of AllSky environment maps and AllSky environment maps generated from user-drawn labels.

Abstract

Accurate environment maps are a key component for rendering photorealistic outdoor scenes with coherent illumination. They enable captivating visual arts, immersive virtual reality, and a wide range of engineering and scientific applications. Recent works have extended sky-models to be more comprehensive and inclusive of cloud formations but, as we demonstrate, existing methods fall short in faithfully recreating natural skies. Though in recent years the visual quality of DNN-generated High Dynamic Range Imagery (HDRI) has greatly improved, the environment maps generated by DNN sky-models do not re-light scenes with the same tones, shadows, and illumination as physically captured HDR imagery. In this work, we demonstrate progress in HDR literature to be tangential to sky-modelling as current works cannot support both photorealism and the 22 f-stops required for the Full Dynamic Range (FDR) of outdoor illumination. We achieve this by proposing AllSky, a flexible all-weather sky-model learned directly from physically captured HDRI which we leverage to study the input modalities, tonemapping, conditioning, and evaluation of sky-models. Per user-controlled positioning of the sun and cloud formations, AllSky expands on current functionality by allowing for intuitive user control over environment maps and achieves state-of-the-art sky-model performance. Through our proposed evaluation, we demonstrate existing DNN sky-models are not interchangeable with physically captured HDRI or parametric sky-models, with current limitations being prohibitive of scalability and accurate illumination in downstream applications.

1. Introduction

Illumination is central to human perception of physical spaces and the visual quality of media and film [14, 42, 54]. Early works modelling skydomes combined data from varied sources into pre-computed and parametric sky models for engineering and scientific applications, with the first sky models [36, 43] modelling only luminance. With the advent of the digital age, a new paradigm of applications spurred interest in sky models to enable a wide range of emerging digital applications. Nishita et al. [38] proposed the first colour sky model enabling the generation of extraterrestrial views of the Earth for space flight simulators, and Image-Based Lighting (IBL) techniques [10] were proposed to render synthetic objects into real and virtual scenes.

Though conventional Low Dynamic Range (LDR) imagery is suitable for some applications, High Dynamic Range (HDR) [46] imagery is integral to sky-models and the capture of outdoor scenes. HDR images capture a greater range of illumination and, in the particular case of outdoor lighting, can capture the estimated 22 f-stops of exposure necessary for the highlights and shadows of an average real-world outdoor scene [27, 46, 53]. To distinguish the proportionality of scene exposure captured by an HDRI, we define the following:

1. **Low Dynamic Range (LDR)** Imagery: Display-referenced images with compressed dynamic range which can be clipped and displayed in 8-bit colour.
2. **High Dynamic Range (HDR)** Imagery: Scene-referenced measures of illumination with uncompressed dynamic range and precision greater than LDR 8-bit colour for later display as LDR. This includes imagery captured by conventional cameras in 12-bit RAW.
3. **Extended Dynamic Range (EDR)** Imagery: HDR images captured using techniques such as LDR-bracketing for greater exposure range than a singular image from a conventional camera.
4. **Full Dynamic Range (FDR)** Imagery / Physically-Captured Imagery: HDR images that ***fully***-capture the dynamic range of a reference scene without saturation of the exposure range.

This distinction between HDRIs is necessary given the

challenges in applying HDR literature to sky-modeling [12, 20, 34, 44, 57, 62]. In this work, we measure HDR as the Exposure Value (EV) of HDRI given $EV = log_2(|I|_{max} - |I|_{min} + 1)$, where $|I|$ is grayscale intensity. We find that HDR literature generally does not provide sufficient detail to determine supported exposure ranges, focuses on exposure ranges $\leq 6EV$ when real-world outdoor scenes are $\geq 13EV$, omit daytime outdoor imagery in results and, if included, demonstrates saturated clouds and solar features. Such limitations are problematic, given exposure range is key to the illumination provided by environment-maps and current sky-models struggle to reproduce the photorealism and illumination of real-world skies. As demonstrated by Fig. 2, incrementally clipping the EV of an HDRI while equalizing exposure to the FDR 15EV ground truth results in visually indiscernible alterations to the environment-maps (Fig. 2, top), but significant alteration to illumination in IBL scenes through softer tones, shadows, and light transmission (Fig. 2, bottom). Despite advances in HDR image generation using deep neural networks (DNNs), current sky-models struggle photorealistically reproduce the physical accuracy and illumination of real-world skies. These limitations hinder their effectiveness in downstream applications such as IBL, where accurate lighting is crucial.
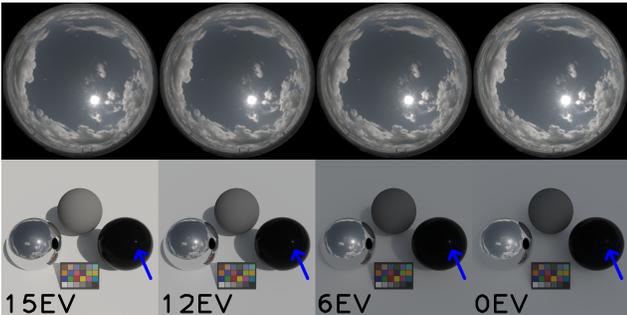

Figure 2. Each column illustrates the impact of an incremental clipping of exposure with exposure equalization to the 15EV ground truth. Though the environment-maps are visually unaltered, renderings have altered tones, shadows and light transmission (sun transmission; black glass orb).

In this work, we propose AllSky, a conditional generative sky-model trained directly on physically captured FDR imagery and enabling user-controlled positioning of the sun and cloud formations. We leverage this model to study the input modalities, tonemapping, conditioning, and evaluation of sky-models and propose a novel framework for evaluating DNN sky-models for coherence with real-world luminance. We demonstrate AllSky achieves state-of-the-art (SOTA) illumination with sustained visual quality, and identify shortfalls in DNN sky-modeling which impede their scalability and interchangeability with physically captured imagery and parametric counterparts. Our findings improve performance in downstream applications, but show recent improvements to DNN HDRI generation to be tangential to sky-modeling. Significant work remains to make sky-models scalable to higher resolutions while sustaining the exposure range of the sun.

2. Background

Physically-captured FDR images offer unsurpassed quality and photorealism for IBL renderings, but physical capture is difficult, labour-intensive [30] and the resulting imagery is both inflexible and of fixed temporal- and geo-locality. Though HDR and EDR images can be suitable, partial-capture of skydome dynamic range generally saturates the solar spectrum in favor of capturing lucid cloud formations. Sky-models have progressed over the past four decades to offer a wide range of accurate, versatile, and portable models providing artists, engineers and scientists a definitive cost-advantage to physical capture [7]. Early works included numerical and parametric models [8, 13, 17, 23, 24, 33, 39, 41, 43, 45, 59] which generated clear and overcast skies via a limited set of user-parameters tied to physical properties but were unable to model atmospheric formations.

With the advent of deep learning, the fitting of analytical models to features and modalities can be automated with parameterization to a finite set of latent or user-intuitive parameters. Notably, this concept has been proposed for lighting estimation [21, 22, 60, 61, 63], where LDR images can be used to guide the generation of HDR environment-maps for relighting virtual objects and scenes.

Recent deep learning approaches have been proposed to augment parametric clear-skies with photorealistic clouds. CloudNet [48] proposed a method for augmenting Hošek-Wilkie [23] parametric skies per user-controlled cloud placement. SkyGAN [35] proposed fitting the Prague Sky Model (PSM) [59] model to real-world photographs, enabling the re-generation of PSM skies with cloud formations. This formulation was also demonstrated by LM-GAN [56], fitting physical captures to the 11-parameter Lalonde-Matthews (LM) outdoor lighting model [33] and producing photorealistic (though lower-resolution) weathered skies. Most recently, Text2Light [9] introduced textually controlled generation of indoor and outdoor environment-maps via diffusion, though its textual conditioning and diversity are dubious.

The objective of physically-accurate sky-models is to generate FDR environment-maps which emulate the illumination of clear (13EV+), cloudy and/or overcast daylight skies of FDR environment-maps, but many challenges remain to achieve this objective. To mitigate problems modeling illumination, many works propose post-generation 'tricks' to augment EDR to FDR solar illumination, including compositing with a parametric model sun and further aggregated inverse-tonemapping [9, 35, 48]. Modelling and control over non-uniform atmospheric formations such as

clouds remain a challenge and works that augment clear-skies with atmospheric formations are not proven to retain a tractable relationship to the parameters of the underlying clear sky-models. A common alternative is to include complex cloud formations through volumetric cloud rendering [3, 28] and cloud simulations [5] post sky-model generation. While this multi-step approach can be versatile and photorealistic, it is labour and computationally expensive [19].

As a result, the modeling of outdoor FDR imagery remains a challenge without a clear solution [51]. Regardless of the approach, the current limitations result in many situations where physical capture is unsurpassed and the preferable method to provide both photorealism, weather variations and illumination [49, 58].

3. Methodology

We define our tonemapping operators, sky-model, and losses in terms of design considerations for sky-modelling.

3.1. Tone-mapping Dynamic Range

Tone-mapping operators (T_m) compress HDRI to a visible, displayable, or latent colour-space more favourable for DNN training given $I' = T_m(I)$. We investigate a range of operators, including: Power-Law (T_γ), logarithmic (T_{log_n}), μ-law (T_μ), and variations thereof as shown in Appendix N.2 and Fig. 3 (left). With the growing popularity of mixed-tone-mappers (combinations of tone-mappers) such as Hybrid Log-Gamma for display of 8K broadcasts [37], we propose μ-lawLog$_2$ ($T_{\mu \log_2}$) to better preserve texture-rich cloud and skydome components (generally $EV \lessgtr 1$) and aggressively compress texture-poor high-exposure regions which are generally saturated in LDR imagery.

$$\mu\text{-lawLog}_2: T_{\mu \log_2}(I) = \log_2 \left[\frac{\log_e(1 + \mu I)}{\log_e(1.0 + \mu)} + 1 \right] \quad (1)$$
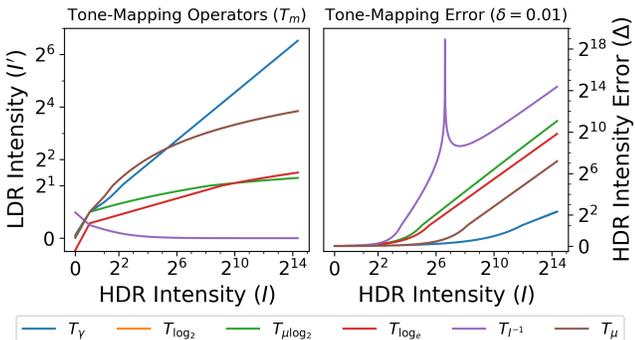

Figure 3. Tone-Mapping operator (T_m) compression (left) and respective non-linearity between error (δ=0.01) in LDR compressed space and error (Δ) in uncompressed HDR space (right).

Each operator is a bijection allowing for the recovery of the original image via $I = T_m{}^{-1}(I')$, but given $\Delta I = |I - T_m{}^{-1}(T_m(I) - \delta)|$ these operators introduce a non-linearity

between error (δ) in LDR compressed space and error (Δ) in uncompressed HDR space (right). Most pronounced with the solar disk, a small error (δ) in LDR space results in a large error (Δ) in HDR space with a profound impact to illumination in IBL rendering.

3.2. Model

As a baseline and backbone, we implement UNetFixUp [15] as a configurable and proven U-Net architecture for outdoor image editing. The model can be configured for arbitrary discrete ($\mathbb{Z}^+$) or continuous ($\mathbb{R}$) c_i-channel input and c_o-channel output. We propose organizing sky-model metrics as shown in Fig. 4, allowing dynamic-range-aware evaluation through losses and metrics in LDR compressed space, cLDR compressed and clipped LDR-space, and HDR inverse-tonemapped linear HDR-space. Though losses can be placed in LDR-, cLDR- and/or HDR-space, choosing saturation points (clipping) is ambiguous and dependent on tonemapper selection, and HDR-space losses are prone to exploding gradients.
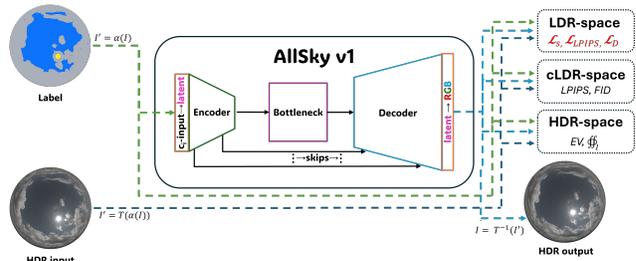

Figure 4. AllSky baseline architecture, accepting arbitrary discrete ($\mathbb{Z}^+$) or continuous ($\mathbb{R}$) c_i-channel input labels. The c_o-channel output can be evaluated by losses and metrics in LDR compressed space, $cLDR$ compressed and clipped LDR-space, and HDR inverse-tone-mapped linear space.

3.3. Losses

In experimentation, we observed that an L_1-loss on a dataset of clear-sky imagery guides the model toward a median clear-sky image with few visual artifacts, but IBL renderings exhibited poor illumination. To explain this discrepancy, we explored sky-model losses and the guidance they provided given a dataset of FDR skydomes.

3.3.1 Supervised Losses

As shown in Fig. 5, the sun (red) is represented by a small subset of pixels whose intensities belittle the remainder of the skydome (blue/green). The four-pixels solar region with intensity $EV \geq 11$ of the 512^2 skydome represents 55% of the environment-maps' illumination. The remaining 45% of illumination is provided by the remaining 99%+ of pixels with intensities $EV \leq 8$, representing an amalgamation of refractive mediums (i.e. skydome, clouds, etc.). Conventional global losses (e.g. an L_1-loss) are insensitive to the
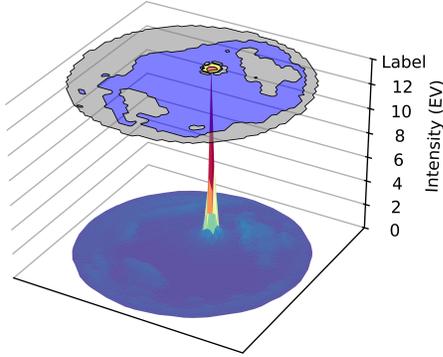
Figure 5. 3D surface of skydome illumination, coloured from low to high intensity. Disk above 3D surface illustrates skydome segmentation. The sun (red) spike is a small subset of pixels whose intensities belittle the remainder of the skydome. HDRDB sample from June 7th, 2016 at 1:54PM [26].
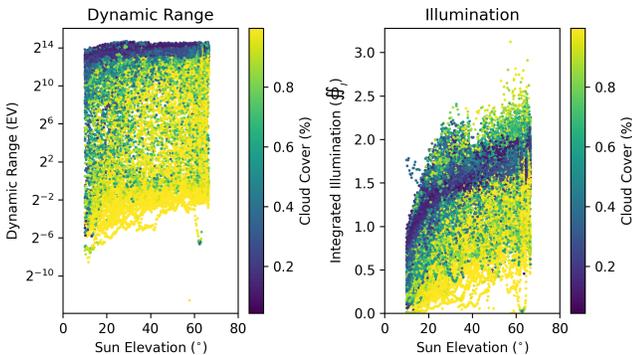

Figure 6. Distribution of exposure range and illumination by sun elevation and cloud cover in HDRDB [26].

solar exposure range of due to the class-imbalance it entail. Although greater sensitivity is a potential solution (e.g. L_2 instead of L_1) the potential loss-imbalances and/or gradient explosion must be mitigated. In this regard, we propose implement class-segmented losses, as Eq. (2):

$$\mathcal{L}(I_{real}, I_{fake}, M) = \mathcal{L}(I_{real} \subset M, I_{fake} \subset M) \quad (2)$$

This allows conventional metrics (e.g. L_1) to be selectively applied via a segmentation mask (M) to target regions which would otherwise be saturated by pixels from outside their distributions (e.g. border, skydome).

3.3.2 Unsupervised Losses

From an extraterrestrial point-of-view (POV), the sun is a $0.5°$ angular-diameter disk with near-constant illumination [50] but, from a terrestrial POV, its size and radiant intensity are attenuated by a stochastic atmosphere. As illustrated in Fig. 6, there is no intuitive relationship between solar-elevation, cloud cover, dynamic range and/or illumination. Similarly, cloud formations consist of 27 categories [40] with variable textures and altitude-specific classification. In the absence of labelled datasets and cloud-formation classification tools, current DNN sky-models un-

conditionally generate cloud formations and exhibit little diversity in cloud formations. Unintentionally, where segmentation morphology is deterministic of cloud formation, some models achieve a greater diversity by overfitting to input masks.

In the absence of input modalities to condition solar radiant intensity and/or cloud textures, skies are nondeterministic systems with intractable variability. As a result, supervised losses (e.g. L_1, LPIPS [64]) guide per an invalid assumption of exactness in reconstruction. To mitigate, we propose a conditional n-Layer discriminator inspired by [4, 25] with adaptive gradient penalty [2, 16, 31], adaptive weighting of the adversarial loss [47] and a modified hinge loss. Given DNN sky-models are ignorant of their operation within LDR-space and the non-linear relationship between LDR- and HDR-space, unsupervised learning also mitigates ill-adapted conditioning by visual losses created for alternate LDR-spaces (e.g. LPIPS [64] adaptability to logarithmic LDR-space has not been previously studied).

4. Experimental Configuration

For the purpose of comparison, we train all AllSky, Cloud-Net, SkyGAN and SkyNet against the Laval HDR Sky database (HDRDB [26], Appendix L), which consists of approximately 32k FDR HDRI. We evaluate Text2Light on the FDR Laval Outdoor Dataset (LOD) [22] using the author's textual CLIP prompts. Images herein are gamma ($\gamma = 2.2$) tone-mapped and target the four primary skydome configurations: clear, cloudy with unobstructed sun, cloudy with obstructed sun, and overcast skies.

4.1. Image-Processing

We develop a pre-processing pipeline to re-size and augment HDRI, optimizing for retention of physically captured characteristics. For evaluation, we quantify pixel intensity as linear luminance per BT.709 [55] and further quantify global-illumination as Integrated Illumination ($\oiint_I$):

$$\oiint_I (I) = \sum \Omega |I| \quad (3)$$

Where Ω is the environment-map's solid angles. As HDRDB, LOD and other datasets previously used to train DNN sky-models are not radiometrically calibrated, we assume a linear camera response and do not report illumination in metric quantities.

4.1.1 Scaling environment-maps

We observe DNN sky-models are inherently scale-sensitive due to the impact of interpolation on the exposure range of physically captured HDRI and study this observation in Appendix J. As plotted in Fig. 7, interpolation methods available through PyTorch (tv.) and OpenCV (cv2.) impact

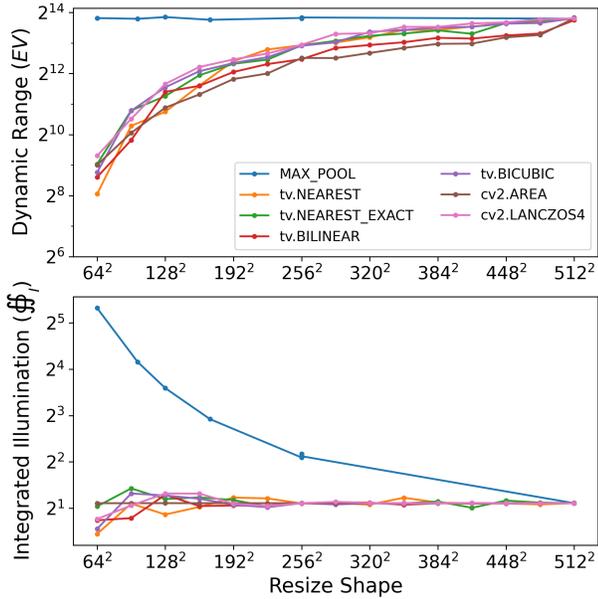environment-map illumination and exposure range [1, 6].[1]

Figure 7. Impact of resizing a 512^2 physically captured HDRI on exposure range and illumination. Mean of 100 augmentations of an HDRI from June 7th, 2016 at 12:52PM [26].

Environment-map format conversions are avoided where possible and, along with augmentation, are applied prior to downsampling. Where otherwise unavoidable, linear spline interpolation is applied and downsampling favours factors of $f(x) = 1/2^x$ to enable inter-area interpolation (retaining over 99% of illumination energy). To prevent discontinuities, environment-maps are skyangular formatted (See Appendix R.1 for more detail on seams).

4.1.2 Segmentation

Solar positioning is refined from ephemeris calculations [52], labelling the extraterrestrial $0.5°$ sun as the solar disk. To reflect terrestrial imagery, we extend the masked solar region to a diameter of $5°$ to include the solar corona and atmospheric attenuation of the extraterrestrial solar disc.

Segmentation of cloud formations was achieved by thresholding RGB via the ratio $Y = \frac{B-R}{B+R}$ proposed by Dev et al. [11]. We find that applying this ratio to μ-lawLog$_2$ ($T_{\mu \log_2}$) tone-mapped HDRI provides visually robust segmentation, but note variability with illumination intensity (e.g. sunrise/sunset) and seasonality [32]. Cloud masks are morphologically processed to produce 'crude' masks which can be further eroded by a parametric circular brush of kernel size k to reduce complexity. We characterize erosion by k=15 as emulating a 'hand-drawn' sketch.

---

[1]See Appendix J.1 for the impact to shadows, tones, and light transmission in renderings.

The final 'discrete' ($\mathbb{Z}^+$) label is a 1-channel composite of solar, cloud, skydome (clear-sky), and border masks. We take inspiration from CloudNet [48] and create augmented 'continuous' ($\mathbb{R}$) labels, infilling the solar regions as cosine similarity to the sun, the skydome as solid angles (Ω), and clouds as distance fields modulated by Perlin noise. A colourized illustration of 'discrete' and 'continuous' labels is provided in Fig. 8.

4.2. Baseline Models

In this work we compare CloudNet, SkyGAN, Text2Light, and AllSky through training against HDRDB [26]. We reproduce CloudNet and SkyGAN per the authors' source code with minimal modifications to the data-processing pipeline. We evaluate Text2Light's *outdoor* functionality the author's published model and the LOD dataset [22].

4.2.1 CloudNet

CloudNet was reproduced per the authors' source code with model and parameters unchanged [48]. We adapted the data-processing pipeline to our segmentation of HDRDB, reproducing the input histogram-equalized Hošek-Wilkie [23] clear skies, cloud distance field maps, and inverted tone-mapping ($T_{I^{-1}}$) per Eq. (4). Originally trained on a small dataset of 629 HDRI for 7,500 epochs, we implemented a framework to facilitate a 3,500 epoch training.

$$\text{Inverted: } T_{I^{-1}}(I) = 1/(1 + I + 0.01) \qquad (4)$$

We evaluated CloudNet both with and without the sun-pass-through mechanism, thus repeating the model's evaluation with the Hošek-Wilkie clear sky solar disk summed into generated environment-maps. To evaluate the impact of CloudNet's input modalities, we train a modified baseline per our discrete ($\mathbb{Z}^+$) input labels.

4.2.2 SkyGAN

SkyGAN was reproduced per the authors' source code with model and parameters unchanged [35]. We adapt the data-processing pipeline to HDRDB, reproducing the input PSM [59] clear skies, and natural logarithmic ($T_{\log_e}$) tone-mapping per Eq. (5) with $\alpha = 0.22$, $\beta = 2.5$ and $\epsilon = 1^{-3}$).

$$\text{Natural Logarithmic: } T_{\log_e}(I) = [log(I + \epsilon) + \beta] * \alpha \quad (5)$$

We completed a 3,500 epoch training of SkyGAN and a variant which bypassed the clear sky encoder. In addition, we adapt StyleGAN3 [29] (SkyGAN's backbone) to HDRDB and ablate select tone-mapping operators through 1,000 epoch trainings.

4.2.3 Text2Light

Evaluation of Text2Light's *outdoor* functionality was completed via unaltered implementation the author's published

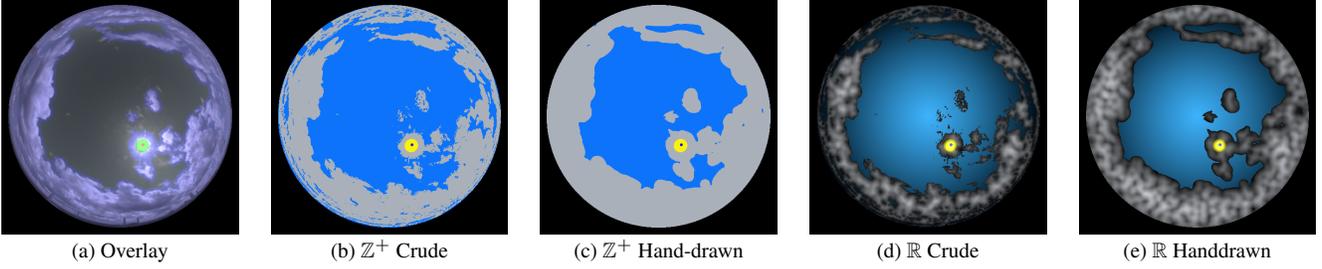| (a) Overlay | (b) $\mathbb{Z}^+$ Crude | (c) $\mathbb{Z}^+$ Hand-drawn | (d) $\mathbb{R}$ Crude | (e) $\mathbb{R}$ Handdrawn |

Figure 8. Discrete ($\mathbb{Z}^+$, Figs. 8b and 8c) labels are 1-channel composites of solar, cloud, skydome (clear-sky), and border masks. Continuous ($\mathbb{R}$, Figs. 8d and 8e) labels augment the segmentation with distance-field modulation of perlin noise. Cloud segmentation is produced via colour ratio [11] and eroded from 'crude' to 'hand-drawn'.

model, checkpoints and textual CLIP prompts with no up-scaling (β=1) [9]. Text2Light's generation pipeline consists of three stages: 1) LDR generation via diffusion, 2) LDR to HDR augmentation via MLP Inverse Tone-Mapping Operator (iTMO), and 3) HDR to FDR parametric boosting per Eq. (6) with constants ρ=4, ϑ=0.83, γ=1 and β= 0.7.[2]

$$M = \left[\overline{\overline{I_b}}/\max\left(\overline{\overline{I_b}}\right) - \vartheta\right]_0^1$$
$$I_b' = I_b + (I_b M \rho) \qquad (6)$$
$$I_b' = e^{(I_b' - \overline{I_b'})\gamma - \beta}$$

We evaluated each stage of Text2Light's generation pipeline, normalizing the resulting HDR image pairs for exposure via $|I_{real}| = \alpha|I_{fake}|$ and reformatting the environment-maps to skyangular for visual metrics.

4.2.4 AllSky

In Appendix M, we leverage AllSky's configurability to ablate the input modalities, tone-mapping operators, and the losses of SkyNet, CloudNet, and SkyGAN. As a baseline, we recreate the training of *SkyNet* [22] which implemented a U-Net architecture guided only by a global L_1-loss. We create a variants emulating CloudNet (continuous labels ($\mathbb{R}$) and LPIPS loss), and SkyGAN (implementing our n-Layer discriminator). All variants implemented class-selective L_1-losses for border and skydome regions.

Per the findings of our ablations in Appendix M, we complete 400-epoch trainings of AllSky at 256^2, and 512^2 resolution and report their performance in Tab. 1. These models are trained with $\mathbb{Z}^+$ discrete labels per guidance from LPIPS, our n-Layer discriminator, and class-selective L_1-losses for border and skydome regions.

4.3. Metrics

An objective of this work is to demonstrate the inadequacy of current DNN sky-model evaluations in regards to solar illumination and characterize the trade-off between accurate

---

[2] I_b denotes a batch of images, $\overline{\overline{I_b}}$ denotes the per-image-mean of I_b, $\overline{I_b'}$ is the mean of I_b' and $[X]_a^b$ clips the values of X such that $a \leq x \leq b$.

illumination and visual quality exhibited by current models. In this regard, we evaluate the sensitivity of commonly reported metrics in Appendix K, finding that SSIM, Multi-Scale SSIM, and HDR-VDP3 offer no meaningful sensitivity to exposure range, while other supervised losses such as L_1, L_2, and EMD offer limited and diminishing sensitivity with increasing resolution. We also note an assumption of reconstruction exactness introduces significant noise from penalizes valid natural stochasticity in style and positioning.

We evaluate the performance of DNN sky-models for visual quality (LPIPS [64] and FID [18]), exposure range (EV), and illumination ($\oiint_I$). Given the generalization that atmospheric textures are low dynamic range, we assume visual quality metrics are insensitive to tone-mapping and, for consistency, evaluate all visual metrics in Gamma (γ) cLDR-space. We quantify exposure range and illumination as EV and $\oiint_I$ for environment-maps in uncompressed HDR space. To demonstrate the importance and sensitivity of these metrics, we report them alongside L_1, L_2, and PSNR as many works evaluate using these metrics exclusively [9, 22, 48]. For greater sensitivity, we implement PSNR with base $\log_2$ ($\text{PSNR}_{\log_2}$).

5. Experimental Results

In this work we adapt CloudNet, and SkyGAN to HDRDB without modification to authors' defined pre- and post-processing pipelines. As a result, each baseline models exhibit different ground truth exposure ranges (EV) and illumination ($\oiint_I$) despite being tested against an identical subset of samples.

5.1. CloudNet

CloudNet and *CloudNet w/o Clear Sky* differ little in Tab. 1, suggesting clear sky and cloud distance field map input modalities contribute negligibly to model performance. Though the Hošek-Wilkie model provides intuitive control over clear skies generation through parameters for turbidity and ground albedo, CloudNet keeps these parameters constant and after histogram equalization, retains no relationship to these parameters. The aggregation of the clear-sky

Table 1. Comparative results for Text2Light [9], CloudNet [48], SkyGAN [35], and AllSky (ours). AllSky offers unsurpassed visual quality (LPIPS, FID) for both crude (C) and hand-drawn (H) labels, and accurate FDR illumination (EV, $\oiint_I$) superseded only by SkyGAN. The results further demonstrate that L_1, L_2 and PNSR offer little correlation to visual quality or illumination. <u>Underlined</u> values indicate category best. Exposure matching is denoted by (*).

	LDR	T_γ**-cLDR**		**HDR**			
	$\text{PSNR}_{\log_2}$ ↑	LPIPS ↓	FID ↓	L_1 ↓	L_2 ↓	EV ←	$\oiint_I$ ←
Ground Truth 512^2	-	-	-	-	-	12.85	3.54
Text2Light SRiTMO Boosted	46.5	0.51	-	0.27	<u>816</u>	8.4	0.16
Text2Light SRiTMO Boosted*	46.5	<u>0.42</u>	-	0.32	1380	<u>12.92</u>	3.73
Text2Light LDR+Boost*	<u>47.3</u>	<u>0.42</u>	-	<u>0.25</u>	827	7.96	<u>3.69</u>
Ground Truth 512^2	-	-	-	-	-	12.7	1.05
CloudNet	<u>99.4</u>	<u>0.17</u>	35.2	<u>0.07</u>	<u>414</u>	6.3	0.53
CloudNet w/ Sun	<u>99.4</u>	<u>0.17</u>	<u>35.1</u>	0.17	1907	<u>14.4</u>	2.07
CloudNet w/o Clear Sky	92.1	0.18	41.5	0.08	418	6.7	<u>0.64</u>
CloudNet w/o Clear Sky w/Sun	92.1	0.18	41.5	0.18	1909	<u>14.4</u>	2.18
Ground Truth 256^2	-	-	-	-	-	11.12	1.15
SkyGAN	-	-	<u>55</u>	-	-	10.7	0.83
SkyGAN w/o Clear Sky	-	-	70	-	-	<u>11.3</u>	<u>1.01</u>
StyleGAN $T_{\mu\log_2}$	-	-	<u>49</u>	-	-	<u>7.7</u>	<u>1.04</u>
Ground Truth 256^2	-	-	-	-	-	8.09	1.13
AllSky $\mathbb{Z}^+$ $T_{\mu\log_2}$ C	<u>104.3</u>	<u>0.13</u>	<u>14.9</u>	<u>0.07</u>	965	<u>7.24</u>	<u>0.85</u>
AllSky $\mathbb{Z}^+$ $T_{\mu\log_2}$ H	98.1	0.16	22.6	0.43	344704	9.58	4.49
Ground Truth 512^2	-	-	-	-	-	9.11	1.14
AllSky $\mathbb{Z}^+$ $T_{\mu\log_2}$ C	<u>102.8</u>	<u>0.14</u>	<u>17.1</u>	<u>0.06</u>	<u>467</u>	<u>8.23</u>	<u>0.82</u>
AllSky $\mathbb{Z}^+$ $T_{\mu\log_2}$ H	93.3	0.18	24.8	0.11	9919	9.95	1.52

sun (see *w/ Sun*) significantly improves EV, but overshoots illumination and introduces a solar disc that pierces through cloud formations. This differs from natural terrestrial imagery where the solar disk is attenuated by the atmosphere and diffused through cloud formations. Given the disparity between generated and FDR ground truth illumination (EV, $\oiint$), environment maps are characterizable as EDR.

5.2. SkyGAN

SkyGAN's exposure range and illumination are shown to be adversely impacted by its clear sky input modality, as its removal improves both exposure range and illumination. SkyGAN exhibits EDR exposure range (75% of ground truth in linear space) but offers poor visual quality and illumination in comparison to StyleGAN3.

Visually in Appendix Q, SkyGAN and StyleGAN both exhibit texture-less cloud 'smears'. We believe this to be the result of discriminator emphasis on solar features. In this regard, StyleGAN3 baselines with less aggressive tone-mapping operator (T_γ, $T_{\log_2}$) collapsed and we exclusively report StyleGAN3 with $T_{\mu\log_2}$.

5.3. Text2Light

As shown by *Text2Light SRiTMO Boosted** in Tab. 1, Text2Light generates EDR environment-maps which can be boosted and exposure matched to provide an exposure range and illumination close to the FDR ground truth. As demonstrated through *Text2Light SRiTMO Boosted*, exposure matching to the ground truth is imperative to correct under-exposure. We note that our exposure matching is dependent on reference FDR imagery and not manually reproducible through conventional exposure and/or offset sliders.

We ablate Text2Light's iTMO by directly augmenting generated LDR imagery to FDR via the parametric boost Eq. (6). As shown by *Text2Light LDR+Boost**, this sample-specific manually tuned heuristic boosts LDR to EDR with lower exposure range, but matches or improves visual quality. We find the results to be inconsistent and scene dependent, producing incoherent results when imagery lacks texture below the horizon. Though inconsistency and manual parameter tuning may be accommodatable in applications such as IBL rendering, it is prohibitive of many downstream scientific and engineering applications which require real-world accuracy.

In Appendix R.2 we demonstrate user control via CLIP prompts to be erratic and with little stochasticity. One run of the authors' textual prompts generated 80 image pairs consisting of 17 unique images and the remaining reappearing between 2 and 12 times within the generated set.

5.4. AllSky

Through our extended work in Appendix M, we ablate All-Sky for discrete ($\mathbb{Z}^+$) and continuous ($\mathbb{R}$) input modalities, tone-mapping operators (T_m), and losses. Continuous ($\mathbb{R}$) labels quantitatively provide an advantage with adversarial losses, but visual examination shows perlin noise propagating into the textures of cloud formations. As this is visually unappealing, we find discrete ($\mathbb{Z}^+$) labels to be favourable.

We perform an ablation study to determine the tractability provided by cloud masks by iteratively eroding the morphology of cloud segmentation. Though illumination (EV, $\oiint_I$) is sustained, visual performance (LPIPS, FID) is incrementally lost as cloud masks are eroded from crude (C) to hand-drawn (H). We repeat this study to determine performance when used by a human-operator, providing models with hand-drawn (H) cloud masks regardless of cloud-mask morphology given during training. This resulted in lost visual performance (LPIPS, FID) proportional to the difference in erosion between training and hand-drawn (H) cloud masks, demonstrating models trained for favourable visual performance with crude (C) cloud masks do not sustain the same performance when used by a human operator.

In Appendix O.1, we find tone-mapping directly impacts the visual quality, exposure range and illumination of sky-model, with our $T_{\mu \log_2}$ tone-mapper improving visual quality and illumination but note aggressive tonemapping introduces instability in exposure range due to the non-linearity between error in LDR and HDR space. Though unideal, aggressive tonemapping mitigates the 'holes' in place of the solar disk produces by weaker tonemappers ($T_{I^{-1}}$, and T_γ).

Through our ablation of L_1, LPIPS, and adversarial losses, we find alleviating the comparative improves visual quality and illumination, but adversarial losses disproportionately differentiating through solar features and impede visual quality.

Combining these findings, we created AllSky as reported in Tab. 1. We apply our discrete ($\mathbb{Z}^+$) labels, our $T_{\mu \log_2}$ tone-mapping, our class-selective L_1 for border and sky-dome regions, LPIPS, and our n-Layer discriminator. As shown, AllSky demonstrates visual quality (LPIPS, FID) for both crude (C) and hand-drawn (H) labels and accurate illumination (EV, $\oiint_I$). AllSky sustains visual quality with FDR illumination accuracy is exceeded only by our variant of SkyGAN and without manually-tuned boosting as required by Text2Light.

6. Discussion

The work herein studies state-of-the-art (SOTA) sky-models and presents ablations of current methodologies with the intention of deepening our understanding of the shortfalls of current DNN sky-models. Sky-modelling extends beyond the conventional paradigm of High Dynamic Range literature through a requirement for greater exposure range and we demonstrate in Tab. 1 that current methods exhibit a trade-off between photorealism (visual quality) and real-world Full Dynamic Range (FDR). We demonstrate this to directly impact Image Based Lighting (IBL) where renderings with saturated illumination produce skewed tones, shadows, and light transmission.

Though our proposed model AllSky out-performs current SOTA for all visual benchmarks, we note that higher-resolutions will exacerbate solar class-imbalance and exposure range, leading to lost visual quality and greater instability in exposure range and illumination. This can be observed through AllSky in Tab. 1 as resolution is doubled from 256^2 to 512^2.

We demonstrate that careful consideration should be taken in evaluating the role and functionality of sky-model input modalities. We demonstrate cloud masks can provide tractability to deterministically reconstruct cloud formations and, if not manually replicatable by a user, result in lost visual quality. We further demonstrate that augmenting inputs with clear-sky and/or other priori does not add user-configurability or improve performance. Though textual prompts can facilitate human interaction in many applications, we show Text2Light's implementation to be dubious and unsupporting of diversity.

Through our study of metrics, we find that commonly reported metrics provide little insight into the exposure range and illumination of generated environment-maps. Given the stochasticity of clouds formations and variability of solar intensity, we find that comparative metrics penalize natural variation which is unimpeding of photorealism and illumination. As shown in Tab. 1, our proposed metrics for exposure range (EV) and illumination ($\oiint_I$) offer sensitivity which correlate to better visual quality in downstream applications such as IBL rendering.

7. Conclusions

In this work we propose a sky-model trained directly on physically captured FDR imagery (AllSky) and leverage its configurability to study the limitations and evaluation of DNN sky-models. Through our framework, we demonstrate DNN sky-models are subject to class-imbalances and exposure range limitations which are uncharacterized by common metrics and can result in a trade-off between visual quality and illumination. Though AllSky surpasses the current SOTA in both visual quality and user-controllability, we demonstrate existing models cannot generate environment-maps with the same visual quality and illumination as physically captured HDRI or parametric sky-models. Significant work remains to enable DNN sky-models to simultaneously emulate photorealistic weathered skies and accurate natural illumination at higher resolutions.

References

[1] Jason Ansel, Edward Yang, Horace He, Natalia Gimelshein, Animesh Jain, Michael Voznesensky, Bin Bao, Peter Bell, David Berard, Evgeni Burovski, Geeta Chauhan, Anjali Chourdia, Will Constable, Alban Desmaison, Zachary DeVito, Elias Ellison, Will Feng, Jiong Gong, Michael Gschwind, Brian Hirsh, Sherlock Huang, Kshiteej Kalambarkar, Laurent Kirsch, Michael Lazos, Mario Lezcano, Yanbo Liang, Jason Liang, Yinghai Lu, CK Luk, Bert Maher, Yunjie Pan, Christian Puhrsch, Matthias Reso, Mark Saroufim, Marcos Yukio Siraichi, Helen Suk, Michael Suo, Phil Tillet, Eikan Wang, Xiaodong Wang, William Wen, Shunting Zhang, Xu Zhao, Keren Zhou, Richard Zou, Ajit Mathews, Gregory Chanan, Peng Wu, and Soumith Chintala. PyTorch 2: Faster Machine Learning Through Dynamic Python Bytecode Transformation and Graph Compilation. In *29th ACM International Conference on Architectural Support for Programming Languages and Operating Systems, Volume 2 (ASPLOS '24)*. ACM, 2024. 5

[2] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein gan, 2017. 4

[3] David Åsberg. Real-time rendering of dynamic baked clouds, 2024. 3

[4] Andreas Blattmann, Tim Dockhorn, Sumith Kulal, Daniel Mendelevitch, Maciej Kilian, Dominik Lorenz, Yam Levi, Zion English, Vikram Voleti, Adam Letts, Varun Jampani, and Robin Rombach. Stable video diffusion: Scaling latent video diffusion models to large datasets, 2023. 4

[5] Antoine Bouthors, Fabrice Neyret, Nelson Max, Eric Bruneton, and Cyril Crassin. Interactive multiple anisotropic scattering in clouds. In *Proceedings of the 2008 symposium on Interactive 3D graphics and games*, pages 173–182, 2008. 3

[6] G. Bradski. The OpenCV Library. *Dr. Dobb's Journal of Software Tools*, 2000. 5

[7] Eric Bruneton. A qualitative and quantitative evaluation of 8 clear sky models. *IEEE Transactions on Visualization and Computer Graphics*, 23(12):2641–2655, 2017. 2

[8] Eric Bruneton and Fabrice Neyret. Precomputed atmospheric scattering. *Computer Graphics Forum*, 27(4):1079–1086, 2008. 2

[9] Zhaoxi Chen, Guangcong Wang, and Ziwei Liu. Text2light: Zero-shot text-driven hdr panorama generation. *ACM Transactions on Graphics (TOG)*, 41(6):1–16, 2022. 2, 6, 7

[10] Paul Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proc. SIGGRAPH 98*, pages 189–198, 1998. 1

[11] S. Dev, F. M. Savoy, Y. H. Lee, and S. Winkler. High-dynamic-range imaging for cloud segmentation. *Atmospheric Measurement Techniques*, 11(4):2041–2049, 2018. 5, 6

[12] Gabriel Eilertsen, Joel Kronander, Gyorgy Denes, Rafal K. Mantiuk, and Jonas Unger. HDR image reconstruction from a single exposure using deep cnns. *CoRR*, abs/1710.07480, 2017. 2

[13] Oskar Elek and Petr Kmoch. Real-time spectral scattering in large-scale natural participating media. In *Proceedings of the 26th Spring Conference on Computer Graphics*, pages 77–84, 2010. 2

[14] Carolina Espinoza-Sanhueza, Marc Hébert, Jean-François Lalonde, and Claude MH Demers. Exploring light and colour patterns for remote biophilic northern architecture. *Indoor and Built Environment*, 33(2):359–376, 2024. 1

[15] David Griffiths, Tobias Ritschel, and Julien Philip. Outcast: Outdoor single-image relighting with cast shadows. *Computer Graphics Forum*, 41(2):179–193, 2022. 3

[16] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, and Aaron Courville. Improved training of wasserstein gans, 2017. 4

[17] Jörg Haber, Marcus Magnor, and Hans-Peter Seidel. Physically-based simulation of twilight phenomena. *ACM Transactions on Graphics (TOG)*, 24(4):1353–1373, 2005. 2

[18] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. Gans trained by a two time-scale update rule converge to a local nash equilibrium. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017. 6

[19] Stephen Hill, Stephen McAuley, Laurent Belcour, Will Earl, Niklas Harrysson, Sébastien Hillaire, Naty Hoffman, Lee Kerley, Jasmin Patry, Rob Pieké, Igor Skliar, Jonathan Stone, Pascal Barla, Mégane Bati, and Iliyan Georgiev. Physically based shading in theory and practice. In *ACM SIGGRAPH 2020 Courses*, 2020. 3

[20] Jack Hilliard, Adrian Hilton, and Jean-Yves Guillemaut. 360u-former: Hdr illumination estimation with panoramic adapted vision transformers, 2024. 2

[21] Yannick Hold-Geoffroy, Kalyan Sunkavalli, Sunil Hadap, Emiliano Gambaretto, and Jean-François Lalonde. Deep outdoor illumination estimation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7312–7321, 2017. 2

[22] Yannick Hold-Geoffroy, Akshaya Athawale, and Jean-François Lalonde. Deep sky modeling for single image outdoor lighting estimation. In *CVPR*, 2019. 2, 4, 5, 6

[23] Lukas Hosek and Alexander Wilkie. An analytic model for full spectral sky-dome radiance. *ACM Trans. Graph.*, 31(4), 2012. 2, 5

[24] Lukáš Hošek and Alexander Wilkie. Adding a solar-radiance function to the hošek-wilkie skylight model. *IEEE Computer Graphics and Applications*, 33(3):44–52, 2013. 2

[25] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017. 4

[26] J.-F. Lalonde, L.-P. Asselin, J. Becirovski, Y. Hold-Geoffroy, M. Garon, M.-A. Gardner, and J. Zhang. The Laval HDR sky database, 2016. 4, 5

[27] Henrik Wann Jensen, Frédo Durand, Julie Dorsey, Michael M. Stark, Peter Shirley, and Simon Premože. A physically-based night sky model. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, page 399–408, New York, NY, USA, 2001. Association for Computing Machinery. 1

[28] Simon Kallweit, Thomas Müller, Brian McWilliams, Markus Gross, and Jan Novák. Deep scattering: Rendering atmospheric clouds with radiance-predicting neural networks. *ACM Trans. Graph. (Proc. of Siggraph Asia)*, 36(6), 2017. 3

[29] Tero Karras, Miika Aittala, Samuli Laine, Erik Härkönen, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Alias-free generative adversarial networks. In *Proc. NeurIPS*, 2021. 5

[30] Joseph T. Kider, Daniel Knowlton, Jeremy Newlin, Yining Karl Li, and Donald P. Greenberg. A framework for the experimental comparison of solar and skydome illumination. *ACM Trans. Graph.*, 33(6), 2014. 2

[31] Naveen Kodali, Jacob Abernethy, James Hays, and Zsolt Kira. On convergence and stability of gans, 2017. 4

[32] TL Koehler, RW Johnson, and JE Shields. Status of the whole sky imager database. In *Proceedings of the Cloud Impacts on DOD Operations and Systems, 1991 Conference*, pages 77–80, 1991. 5

[33] Jean-François Lalonde and Iain Matthews. Lighting estimation in outdoor image collections. In *2014 2nd International Conference on 3D Vision*, pages 131–138. IEEE, 2014. 2

[34] Ben Mildenhall, Peter Hedman, Ricardo Martin-Brualla, Pratul P Srinivasan, and Jonathan T Barron. Nerf in the dark: High dynamic range view synthesis from noisy raw images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16190–16199, 2022. 2

[35] Martin Mirbauer, Tobias Rittig, Tomáš Iser, Jaroslav Křivánek, and Elena Šikudová. Skygan: Realistic cloud imagery for image-based lighting. *Computer Graphics Forum*, 43(1):e14990, 2024. 2, 5, 7

[36] Parry Moon. Proposed standard solar-radiation curves for engineering use. *Journal of the Franklin Institute*, 230(5):583–617, 1940. 1

[37] Yuji Nagata, Kenichiro Ichikawa, Takayuki Yamashita, Seiji Mitsuhashi, and Hiroyasu Masuda. Content production technology on hybrid log-gamma. In *SMPTE 2017 Annual Technical Conference and Exhibition*, pages 1–12. SMPTE, 2017. 3

[38] Tomoyuki Nishita, Takao Sirai, Katsumi Tadamura, and Eihachiro Nakamae. Display of the earth taking into account atmospheric scattering. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques*, pages 175–182, 1993. 1

[39] Tomoyuki Nishita, Yoshinori Dobashi, and Eihachiro Nakamae. Display of clouds taking into account multiple anisotropic scattering and sky light. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques*, pages 379–386, 1996. 2

[40] National Oceanic and Atmospheric Administration. NWS Cloud Chart, 2023. 4

[41] Sean O'Neil. Accurate atmospheric scattering. *Gpu Gems*, 2:253–268, 2005. 2

[42] Mojtaba Parsaee, Claude Mh Demers, Marc Hébert, Jean-François Lalonde, and André Potvin. Biophilic, photobiological and energy-efficient design framework of adaptive building façades for northern canada. *Indoor and Built Environment*, 30(5):665–691, 2021. 1

[43] R. Perez, R. Seals, and J. Michalsky. All-weather model for sky luminance distribution—preliminary configuration and validation. *Solar Energy*, 50(3):235–245, 1993. 1, 2

[44] Pakkapon Phongthawee, Worameth Chinchuthakun, Nontaphat Sinsunthithet, Varun Jampani, Amit Raj, Pramook Khungurn, and Supasorn Suwajanakorn. Diffusionlight: Light probes for free by painting a chrome ball. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 98–108, 2024. 2

[45] A. J. Preetham, Peter Shirley, and Brian Smits. A practical analytic model for daylight. In *Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques*, page 91–100, USA, 1999. ACM Press/Addison-Wesley Publishing Co. 2

[46] Erik Reinhard, Wolfgang Heidrich, Paul Debevec, Sumanta Pattanaik, Greg Ward, and Karol Myszkowski. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010. 1

[47] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2021. 4

[48] Pinar Satilmis, Demetris Marnerides, Kurt Debattista, and Thomas Bashford-Rogers. Deep synthesis of cloud lighting. *IEEE Computer Graphics and Applications*, 2022. 2, 5, 6, 7

[49] Chris Scullion. Forza horizon 5's art team explains how it got the skies to look so realistic, 2021. 3

[50] Michael A. Seeds. *Foundations of Astronomy*. Thomson Brooks/Cole, 2006. 4

[51] Gowri Somanath and Daniel Kurz. HDR environment map estimation for real-time augmented reality. *CoRR*, abs/2011.10687, 2020. 3

[52] Brandon Stafford. pysolar, 2018. 5

[53] Jessi Stumpfel, Andrew Jones, Andreas Wenger, Chris Tchou, Tim Hawkins, and Paul Debevec. Direct hdr capture of the sun and sky. In *ACM SIGGRAPH 2006 Courses*, page 5–es, New York, NY, USA, 2006. Association for Computing Machinery. 1

[54] Ermo Säks. Lighting in cinematic virtual reality. *International Journal on Stereo & Immersive Media*, 7(1):122–135, 2024. 1

[55] International Communications Union. Parameter values for the hdtv standards for production and international programme exchange, 2015. 4

[56] Lucas Valença, Ian J. Maquignaz, Hadi Moazen, Rishikesh Madan, Yannick Hold-Geoffroy, and Jean-François Lalonde. Lm-gan: A photorealistic all-weather parametric sky model, 2023. 2

[57] Chao Wang, Ana Serrano, Xingang Pan, Bin Chen, Karol Myszkowski, Hans-Peter Seidel, Christian Theobalt, and Thomas Leimkühler. Glowgan: Unsupervised learning of hdr images from ldr images in the wild. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10509–10519, 2023. 2

[58] Alexander Wilkie and Lukas Hošek. Predicting sky dome appearance on earth-like extrasolar worlds. In *Proceedings of the 29th Spring Conference on Computer Graphics*, page 145–152, New York, NY, USA, 2013. Association for Computing Machinery. 3

[59] Alexander Wilkie, Petr Vevoda, Thomas Bashford-Rogers, Lukáš Hošek, Tomáš Iser, Monika Kolářová, Tobias Rittig, and Jaroslav Křivánek. A fitted radiance and attenuation model for realistic atmospheres. *ACM Trans. Graph.*, 40(4), 2021. 2, 5

[60] Piaopiao Yu, Jie Guo, Fan Huang, Cheng Zhou, Hongwei Che, Xiao Ling, and Yanwen Guo. Hierarchical disentangled representation learning for outdoor illumination estimation and editing. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 15313–15322, 2021. 2

[61] Piaopiao Yu, Jie Guo, Longhai Wu, Cheng Zhou, Mengtian Li, Chenchen Wang, and Yanwen Guo. Dual attention autoencoder for all-weather outdoor lighting estimation. *Sci. China Inf. Sci.*, 64, 2021. 2

[62] Jinsong Zhang and Jean-Francois Lalonde. Learning high dynamic range from outdoor panoramas. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017. 2

[63] Jinsong Zhang, Kalyan Sunkavalli, Yannick Hold-Geoffroy, Sunil Hadap, Jonathan Eisenman, and Jean-François Lalonde. All-weather deep outdoor lighting estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10158–10166, 2019. 2

[64] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric, 2018. 4, 6