

Beyond the Pixel: a Photometrically Calibrated HDR Dataset for Luminance and Color Prediction

Christophe Bolduc, Justine Giroux, Marc Hébert, Claude Demers, and Jean-François Lalonde
Université Laval

Abstract

Light plays an important role in human well-being. However, most computer vision tasks treat pixels without considering their relationship to physical luminance. To address this shortcoming, we introduce the Laval Photometric Indoor HDR Dataset, the first large-scale photometrically calibrated dataset of high dynamic range 360° panoramas. Our key contribution is the calibration of an existing, uncalibrated HDR Dataset. We do so by accurately capturing RAW bracketed exposures simultaneously with a professional photometric measurement device (chroma meter) for multiple scenes across a variety of lighting conditions. Using the resulting measurements, we establish the calibration coefficients to be applied to the HDR images. The resulting dataset is a rich representation of indoor scenes which displays a wide range of illuminance and color, and varied types of light sources. We exploit the dataset to introduce three novel tasks, where: per-pixel luminance, per-pixel color and planar illuminance can be predicted from a single input image. Finally, we also capture another smaller photometric dataset with a commercial 360° camera, to experiment on generalization across cameras. We are optimistic that the release of our datasets and associated code will spark interest in physically accurate light estimation within the community. Dataset and code are available at <https://lvsn.github.io/beyondthepixel/>.

1. Introduction

Natural light has shaped the way our human visual system evolved [13], plays a key role in driving our circadian rhythm [16], and affects our mental health [47] and social organization [11]. It has also been shown [54] that human vision relies on stable properties of light, measured in terms of luminance (in cd m^{-2}), in order to perceive object features such as shape and color.

Natural light is also at the heart of photography and computer vision. However, most if not all computer vision approaches consider pixel values as a 3-channel input to be processed without considering the relationship be-

tween pixel intensity and luminance. This is understandable since modern digital cameras pursue a goal different from measuring physically accurate perceived brightness: they strive to create visually pleasing photographs. In doing so, their internal image signal processors (ISP) perform a series of operations on the measured light (denoising, contrast enhancement, tonemapping, etc. [34]) in order to produce pixel values which, while visually appealing, no longer correspond to the physical properties of the environment.

Modeling camera ISPs and inverting their image formation process has been the subject of many works (e.g. [67, 29, 55]). Here, most approaches aim at recovering an image where pixel values are linearly proportional to the scene radiance (or luminance). Closely related are approaches for capturing high dynamic range (HDR) images [15], or predicting HDR from low dynamic range (LDR) photographs [17, 39, 49, 41, 45]. While linear pixel values can be extremely useful for physics-based vision applications (e.g. [27]), the scale factor to absolute luminance is still unknown. Can we go beyond (linear) pixel values and recover per-pixel luminance from a single image?

In this paper, we propose the Laval Photometric Indoor HDR Dataset: what we believe to be the first large-scale dataset to help the community answer this question. The novel dataset of physically accurate luminance and colors acquired in a wide variety of indoor scenes. Our key idea is to leverage the camera and RAW captures of an existing dataset of HDR indoor 360° panoramas that was previously captured [20]. We contribute by carefully calibrating the camera with a chroma meter to determine the per-channel correction factors to be applied to each panorama. Our analysis shows that the Laval Photometric Indoor HDR Dataset contains a wide range of illuminance (e.g. [0 lx, 7000 lx]) and color, expressed in correlated color temperature (CCT) (e.g. [2000 K, 8000 K]), capturing the diversity of indoor environments. We also explore the luminance and color distributions of individual light sources in the dataset, which span several orders of magnitude of average luminance.

We present three novel learning tasks that are enabled by our calibrated dataset. Given a single image as input, we explore how per-pixel luminance, per-pixel color, and planar

illuminance can be estimated. More importantly, we also consider what information must be available for accurate light prediction. Indeed, democratizing the process of capturing physical luminance begs several important questions: can luminance be accurately estimated using conventional, uncalibrated cameras? If so, is HDR imagery needed or is a single, well-exposed shot sufficient? Is a generic approach appropriate or do methods need to be finetuned to specific cameras? We provide initial answers to these challenging questions by presenting learning experiments on our novel dataset, as well as on another, smaller photometric dataset captured with an off-the-shelf 360° camera. By publicly releasing the calibrated datasets and associated code, we hope to spur interest in the community and help it consider the physical light measurements that lie beyond the pixels.

2. Related work

Radiometric camera calibration A large body of work has tackled the recovery of the camera response function, the (usually proprietary) non-linear tonemapping curve applied by cameras [26]. This can be done for example from multiple exposures [15], an image sequence [35], or even from a single image [43, 44, 42]. This can also be done jointly with other tasks, e.g. vignetting correction [36].

HDR reconstruction Fusing multiple low dynamic range (LDR) images at different exposures into one high dynamic range (HDR) image has been studied extensively [60, 15]. The overlapping exposures can be leveraged to simultaneously linearize the input images and reconstruct the radiance. Images must be aligned if the camera moves [65], and more complex treatment must be given to avoid ghosting if the scene is in movement [24, 33, 56]. HDR images can also be reconstructed from image bursts [38], specialized optics [64, 51], or even during NeRF [53] training: from RAW inputs [52], non-overlapping exposures [31], or by using a dedicated network [22]. While we leverage HDR reconstruction techniques to build the Laval Photometric Indoor HDR Dataset, we wish to recover absolute luminance values: we must therefore use specialized tools for acquiring these measurements.

Inverse tonemapping Algorithms for recovering HDR from LDR images (known as inverse tone mapping) have been proposed. This was done by inverting tonemapping operators [58, 5, 6], expanding the dynamic range via edge stopping functions [59] or by employing scene-specific iTMO [37]. Recently, deep learning methods have been proposed, for example by predicting the exposure stack that is then fused with HDR methods [17, 39] or by directly predicting the HDR image [70, 41, 49, 62, 68]. Of note, [45] models each stage of the camera pipeline using individual

networks. In this work, we employ a Unet (similar to [45]) to explore novel tasks enabled by our photometric dataset.

Photometric calibration HDR images store relative luminance values. Multiple techniques can be used to identify the (linear) photometric calibration coefficients of an imaging system to retrieve absolute luminance: measure the illuminance of a scene using a chroma meter and compare it to the integration of a calibrated 180° fisheye lens [32], use a luminance meter to measure the luminance of an incoming direction and compare directly with the corresponding pixel of the camera [57], or employ a calibrated display [46]. Here, we follow [32] to calibrate our dataset.

Color prediction and post-capture white balance To our knowledge, no current method allows for the prediction of photometric color from a single LDR image. Closely related, automatic white balance (or illuminant estimation, or color constancy) has been thoroughly explored in computer vision [7, 9, 10, 18, 12, 21, 63]. For example, correcting the white balance based on presets [3] allows the network to understand the color temperature of a scene. While most of these works assume a single illuminant, correcting for multiple illuminants has also been explored [30, 23, 4]. In these works, no absolute color values are obtained.

Luminance prediction Prediction of HDR at physical luminance is a relatively new task. Of note, Wei et al. [66] tackle the problem, but their dataset is limited to stores, only luminance without its chromaticity is available, and the exposure is given as input in the experiments, hence the network does not learn to predict illuminance from the visual features in the scene. We believe that ours is the first large-scale HDR dataset containing absolute colorimetric information of the luminance maps.

3. The Laval Photometric Indoor HDR Dataset

3.1. Base dataset

We rely on the existing Laval Indoor HDR dataset [20] (herein referred to as “Laval dataset”) which comprises over 2300 HDR panoramas captured in a large variety of indoor scenes. The data was captured with a Canon 5D Mark III camera and a Sigma 8 mm fisheye lens mounted on a tripod equipped with a robotic panoramic tripod head, and programmed to shoot 7 bracketed exposures at 60° increments along the azimuth. The resulting 42 photos were shot in RAW mode, and automatically stitched into a 22 f-stop HDR 360° panorama using the PTGui Pro commercial software. As with other HDR datasets (e.g. [15, 1, 33, 45]), the Laval dataset captures up-to-scale luminance values. In this work, we explicitly calibrate for the unknown, per-channel



Figure 1. Setup used for the capture of the calibration dataset. The original Canon 5D Mark III camera used for the Laval dataset (left, graciously provided by the original authors) captures a bracket of images at different exposures while the chroma meter CL-200a (right) measures the illuminance and chromaticity of the scene.

scale factors in order to recover calibrated luminance and color values at every pixel.

3.2. Capturing a calibration dataset

We can photometrically calibrate the Laval dataset by first capturing a “calibration dataset” to determine the per-channel scale factors for the camera. The estimated scale factors can then be applied to the panoramas to obtain accurate luminance and color values.

To capture the calibration dataset, we place the camera side-by-side with a Konica Minolta CL-200a chroma meter, as shown in fig. 1. We then simultaneously acquire a bracket of 7 RAW exposures from the camera and a reading from the chroma meter. To ensure a match to the images from the Laval dataset, the exact same camera parameters (retrieved from the EXIF header in the RAW files) were used. We found that a total of 5 exposure configurations (differing mainly by the aperture used¹) were used to capture the Laval dataset — we therefore captured brackets for each of the 5 camera configurations at each scene to ensure proper calibration across all panoramas. The chroma meter measures both the scene illuminance (in $\text{cd sr}^{-1} \text{m}^{-2}$, or lx) and its chromaticity (in the CIE xy color space).

The PTGui Pro software was then used to merge the different exposure images into one HDR image. Vignetting correction is applied by PTGui using a polynomial model optimized during a previous panorama stitching from overlapping images. Since vignetting varies with the aperture, the model’s parameters are computed for each of the 5 exposure configurations. This process was repeated in a variety of scenes (different from the scenes in the original dataset)

¹f/4, f/11, f/13, f/14, f/18, with f/4 and f/14 representing the vast majority (98 %) of the Laval dataset.

with diverse illumination conditions. In all, 135 scenes were captured to establish the calibration dataset.

3.3. Illuminance computation

To compute the illuminance from the captured HDR image, we first geometrically calibrate the camera using [40] with the projection model from [50] (implemented in OpenCV). Using the recovered lens parameters, the captured fisheye images are re-projected to an orthographic projection. The (uncalibrated) illuminance E can then be computed for each color channel using

$$E = \frac{\pi}{N} \sum_i^N L(i), \quad (1)$$

where $L(i)$ is the value of pixel i , and N the number of pixels in the (circular) image. Eq. (1) derives from the CIE illuminance equation and is explained in the supplementary.

3.4. Calibration coefficients

The uncalibrated illuminance from eq. (1) can be compared to the absolute illuminance measured with the chroma meter. To obtain per-channel illuminance, the xyY color value captured by the chroma meter is converted to RGB (see supplementary). A linear regression identifies the coefficients to be applied to the Laval dataset to obtain the photometrically accurate HDR panoramas. For example, the regression for the f/14 capture configuration for each RGB channel has coefficients of determination (R^2) of (0.985, 0.987, 0.989) for (R, G, B) respectively, indicating the high reliability of the fits. The uncertainty on the calibration is discussed in the supplementary.

3.5. Photometric correction

All panoramas from the Laval dataset were regenerated with PTGui Pro, and each one corrected according to the linear coefficients of its capture configuration. In total, after filtering out the few panoramas that were oversaturated, we have 2362 HDR photometric panoramas of physically accurate luminance and color at 3884×7768 pixel resolution.

4. Visualizing the dataset

In this section, physical properties of the entire scene and of individual light sources are derived for each panorama in the Laval Photometric Indoor HDR Dataset. For visualization of photometric colors, each pixel is expressed as the correlated color temperature (CCT) of its luminance (cf. supplementary). We now explore the diversity in the dataset by computing statistics of the different physical parameters of scenes and light sources present therein.

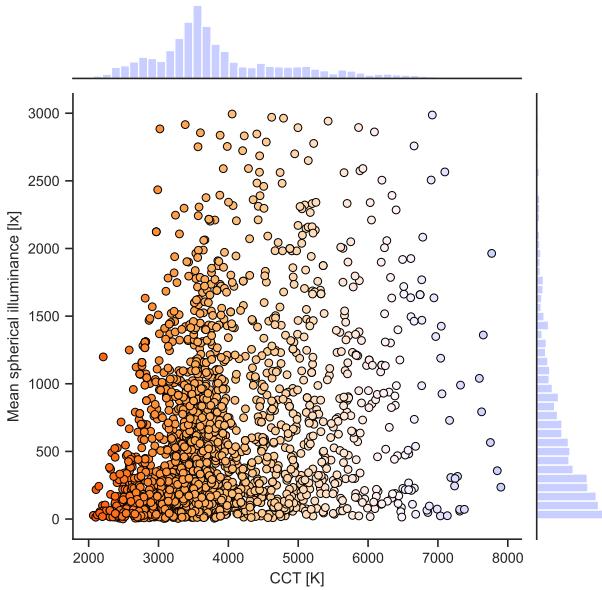


Figure 2. Correlation between the CCT (K) and the mean spherical illuminance (lx) of the photometric panoramas of the dataset. The distributions of CCT (top) and mean spherical illuminance (right) of the data are also displayed. Only the data with a CCT in [2000 K, 8000 K] and a mean spherical illuminance in [0 lx, 3000 lx] are included to better see the trends (2187 out of 2362 panoramas). Points are color-coded according to their CCT.

4.1. Entire scene

We represent an entire 360° panorama as its mean spherical illuminance (MSI) [14] and CCT (cf. supplementary). The correlation between both quantities and their distributions over the entire dataset is presented in fig. 2. As can be observed in fig. 2, the majority of panoramas in the dataset possess relatively low MSI and CCT (bottom-left quadrant in the plot). Indeed, the median MSI and CCT of the dataset are 461 lx and 3654 K respectively. Figs. 3 and 4 illustrate the dataset, showing visual examples sorted by MSI and CCT respectively. For example, fig. 3 shows that scenes with lower MSI correspond to basements and bedrooms with curtains and artificial lighting (incandescent, compact fluorescent and/or “soft white” LED), whereas scenes with higher MSI correspond to well-lit public spaces, often containing very bright ceiling lights or large windows on a sunny day. Similarly, fig. 4 illustrates that scenes with lower CCT have predominantly artificial lighting, whereas higher CCT can be due to strong outdoor lighting from windows and/or more neutral surface reflectance.

4.2. Individual light sources

To provide a more fine-grained analysis of the dataset, we detect and segment light sources in panoramas using the approach by Gardner et al. [19]. In total, 11 060 light

sources are detected, for which the mean CCT and luminance are computed. Overall, the average mean luminance for all the light sources included in the dataset is $27\,874 \text{ cd m}^{-2}$ (median of 3854 cd m^{-2}), and the average mean CCT is 3648 K (median of 3380 K).

A total of 406 randomly selected light sources were manually labeled in order to study the physical properties and diversity of three types of light sources differentiated by their visual appearance: elongated “tubes” (190 samples), point-type “bulbs” (105 samples), and large “windows” (111 samples), including skylights and windows with curtains. Fig. 5 shows examples of detected sources and the correlation between their CCT and average luminance values for each category.

Of all three categories, “tubes” have the most compact distributions, with standard deviations of 465 K and 2957 cd m^{-2} for CCT and luminance respectively, as they all tend to use the same fluorescent lighting technology. These types of lights are mostly used in public spaces, thus their properties are expected to be similar. In comparison, the standard deviations of the CCT and luminance for the sources labeled as “bulbs” are 1090 K/95 381 cd m^{-2} and 1395 K/6821 cd m^{-2} for the “windows”. As these numbers suggest, the average luminance of the “bulbs” varies considerably more than the “tube” category, as their area (size) and purpose differ and can be found both in public and private spaces. The “bulb” category also shows a wide diversity in CCT, as it contains different types of lighting technologies (tungsten, compact fluorescent, or LED) that are difficult to separate visually. They are the category with the highest average luminance, with mean/median values of $59\,701/26\,058 \text{ cd m}^{-2}$, compared to “tubes” ($3212/2330 \text{ cd m}^{-2}$) and “windows” ($6087/3032 \text{ cd m}^{-2}$), as they are small but very bright light sources that are usually scattered by diffusers or reflectors.

The “windows” category is the most diverse group of light sources, especially regarding the CCT, as the spectral properties of the incoming light depends on multiple factors, such as the time of day, weather, geographical position and orientation of the room, the scene outside the window, human-made light modifiers (curtains and blinds), etc. That category also contains the hottest light sources. The total intensity of the light also varies greatly, as panoramas were taken during the day and at night, during sunny and cloudy days. The first and fourth image of the bottom row of fig. 5 show the impact of curtains on the average luminance of the light penetrating in the scene, compared to the windows (other examples on the third row) without curtains. Human-made light modifiers also have an impact on the “bulb” category due to the lampshade and type of fixture used, reducing the average luminance perceived, as it can be seen in the first and fifth examples of the second row of fig. 5. The skylight shown in the fifth example of the third row of fig. 5 has

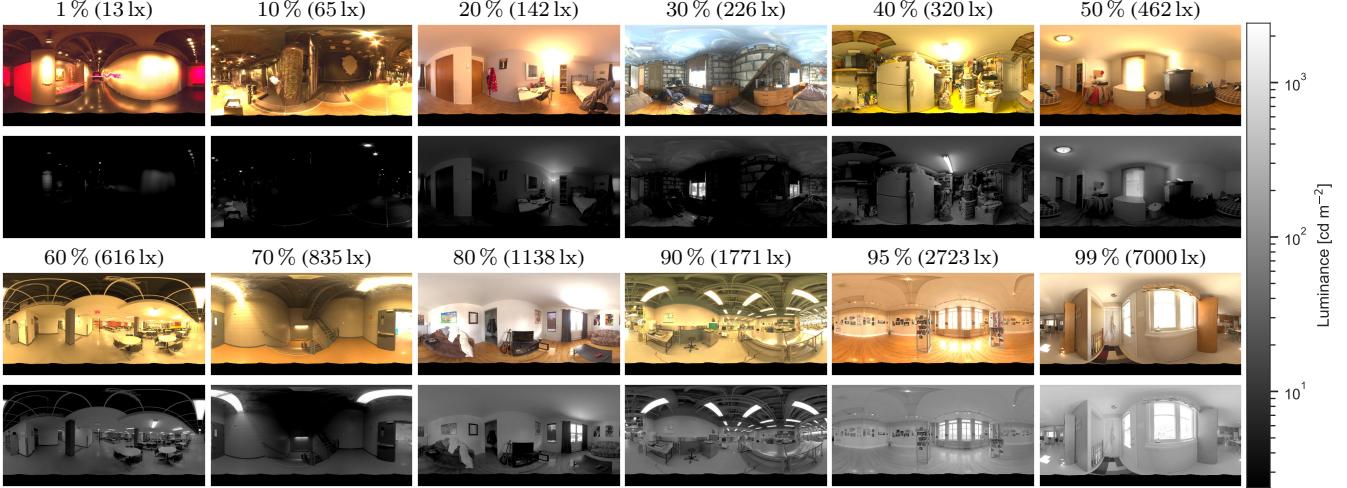


Figure 3. Example scenes with mean spherical illuminance (MSI) close to the quantile values. Greyscale images below show the corresponding log-luminance maps (color scale shown on the right). The percentiles and corresponding measured MSI are indicated above the images. Images are reexposed and tonemapped ($\gamma = 2.2$) for display.

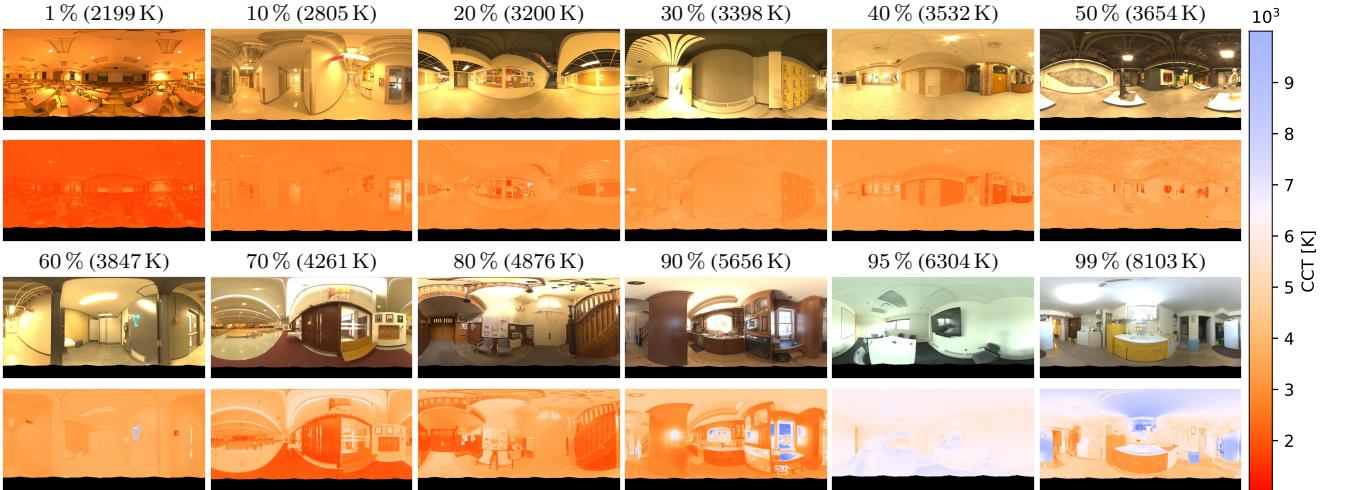


Figure 4. Example scenes with CCT close to the quantile values. Colored images below show the CCT map of the scenes, with corresponding scale shown on the right. The percentiles and corresponding measured scene CCT are indicated above the images. Images are reexposed and tonemapped ($\gamma = 2.2$) for display.

a very high CCT, due to the blue color of the sky, compared to the other lateral windows, indicating that the orientation in space of the window (and the view accessible from it) has an impact on the spectral properties of the light in the scene.

5. Learning to predict photometric values

Our main goal is to develop algorithms that perform physically accurate lighting predictions from real-world photographs captured “in the wild.” It is our hope that the proposed Laval Photometric Indoor HDR Dataset helps the community make strides towards this goal. Here, we introduce new tasks that are enabled by our dataset, and analyze the conditions necessary for accurate light prediction.

5.1. Prediction tasks

We present three novel learning tasks that are enabled by our dataset. Given a single image as input, each task aims to predict the following values.

1. **Per-pixel luminance:** we wish to recover the luminance (cd m^{-2}) at each pixel in the input. For clarity, losses are attributed independently to two subtasks: extrapolating HDR values from LDR inputs (similar to [45], see sec. 2); and predicting the scalar exposure to appropriately scale the HDR values to luminance. Here, we wonder how different degradations (e.g. noise, quantization, tonemapping) on the input affect the prediction.

2. **Per-pixel color:** we wish to estimate the color at each pixel in the input by predicting its CCT. We augment the

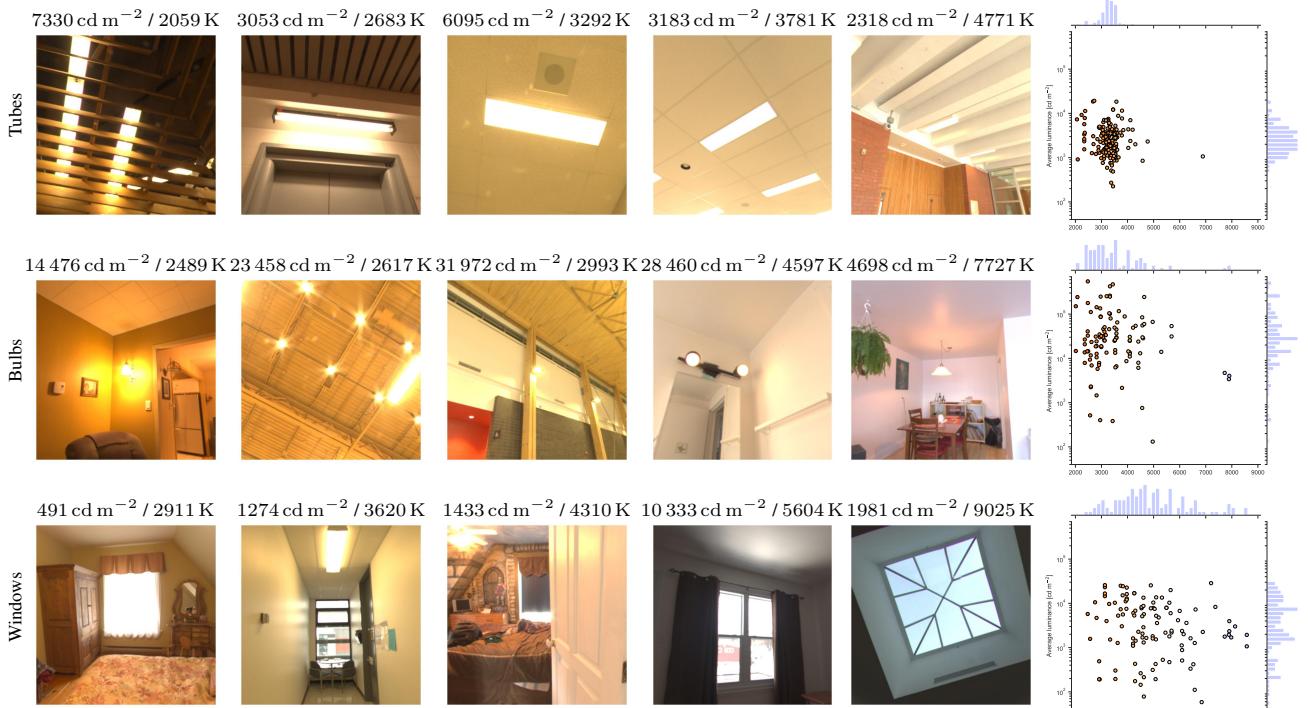


Figure 5. From top to bottom: examples of light sources labeled as tubes, bulbs, and windows (left), and correlation between the CCT and the average luminance of the light sources (right). Each light source is centered in the frame, with its average luminance/CCT indicated above. The mean/median values of each category are: “window” 4946 K/4626 K; “tubes” 3293 K/3290 K; “bulb” 3506 K/3294 K. Images are reexposed and tonemapped ($\gamma = 2.2$) for display.

white balance (WB) of the input using [2] so that the colorimetry of the scene is unknown and wish to see if the network is able to correctly identify the true CCT, as well as predicting the CCT for saturated pixels.

3. Planar illuminance: we wish to predict the (scalar) planar illuminance. This is computed using eq. (1) from a 180° photometric HDR image, but can it also be done from a narrower field of view (FOV)? We also explore the impact of the information provided in the input: can a single LDR image, at arbitrary exposure, be sufficient? Is HDR necessary, or alternatively, is the ground truth exposure needed? The input to each of these tasks are adapted to measure if a deep learning architecture can understand the photometry of a scene from limited data.

5.2. Learning architecture and per-task data

Fig. 6 summarizes the architecture used for the experiments. A UNet [61] with fixup initialization [69] (implementation from [25]) takes an image as input and outputs an image of the same resolution. It consists of 5 down/up sampling with skip connections, with 6 residual blocks [28] at each level and 6 bottleneck layers. The decoder part of the UNet is used for pixel prediction tasks. Additionally, a subnetwork is added at the center of the bottleneck layers of the UNet, consisting of a 4-layer MLP, and outputting

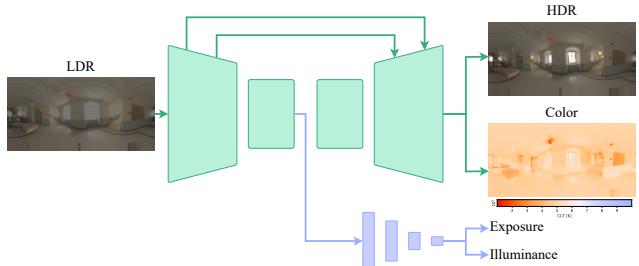


Figure 6. Architecture used for learning tasks. The learning architecture is based on a UNet with fixup initialization (green), with an added subnetwork for scalar prediction tasks (blue). The input is an LDR image of varying degradations and field of view (FOV). The outputs depend on the task: per-pixel luminance prediction outputs an HDR image and its (scalar) exposure; per-pixel color prediction outputs a CCT map; and illuminance prediction outputs an illuminance scalar only.

a single scalar. This subnetwork is used for scalar prediction tasks. Every inner layer uses ReLU activation, and the last layers of the decoder and subnetwork have tanh activation function. The outputs are normalized in logarithmic scale accordingly. Each network is trained with the Adam optimizer with learning rates of 10^{-6} for luminance and illuminance tasks, and of 10^{-5} for color.

Input	RMSE \downarrow	siRMSE \downarrow	HV3 \uparrow
Linear	116.2	83.4	96.4
Gamma	121.9	84.1	96.4
Quantized	125.4	83.8	95.9
Noise	116.3	84.2	96.6
All	121.0	84.7	96.0

Table 1. The effect of input degradations on the prediction of the per-pixel luminance. The rows indicate the degradation applied to the input of the network. Each input is reexposed and clipped in the range [0,1], and different transformations are applied: none ("Linear"), tonemapping (with $\gamma = 2.2$) ("Gamma"), 8-bit quantization ("Quantized"), additive gaussian noise with variance drawn uniformly in the [0, 0.03] interval ("Noise") and all 3 degradations compounded ("All"). "HV3" is the HDR-VDP-3 metric [48].

Input	180°		120°		60°		60°–120°	
	RMSE \downarrow	R $^2\uparrow$						
HDR	22.0	0.969	20.6	0.969	51.8	0.861	49.7	0.839
LDR+scale	47.8	0.830	50.9	0.834	57.4	0.819	52.8	0.876
LDR	123.7	0.406	124.1	0.374	122.0	0.402	126.5	0.423

Table 2. Illuminance prediction at different FOV with different levels of information as input. The network is trained and evaluated on: the photometric HDR ("HDR"), the reexposed photometric HDR clipped in the range [0, 1] with ("LDR+scale") and without ("LDR") knowledge of the exposure. Additionally, the network is trained on multiple FOVs: full hemispherical 180° (in equirectangular projection), 120°, 60°, and randomly varying in the [60°, 120°] interval (using perspective projection).

The architecture is adapted to each task. For the per-pixel luminance task, an HDR image in the same exposure as the given input is predicted. The subnetwork predicts the exposure needed to scale the predicted HDR to absolute luminance. Here, different degradations are applied to the input: clipping, reexposing, gamma tonemapping, 8 bits quantization and additive gaussian noise. For the per-pixel color task, the subnetwork is omitted and the decoder outputs a CCT map from a WB-augmented input. For the planar illuminance task, the subnetwork outputs the illuminance and the decoder is discarded. Since the planar illuminance is not defined for 360°, the input is a hemisphere, in equirectangular projection or rectangular projection with a given FOV. The photometric dataset is randomly split 80 %–10 %–10 % for train-val-test respectively for all experiments below. For computational efficiency, all HDR panoramas are rescaled to 64 × 128 resolution with an energy preserving scaling function, except for illuminance prediction where we extract perspective projection at 160 × 120.

5.3. Experimental results

Per-pixel luminance Here, the input is reexposed so that its 90th percentile corresponds to 0.8 and is clipped in the [0, 1] range. We then explore the effect of degrading the input on the prediction. We experiment by tonemapping the

input ($\gamma = 2.2$), by quantizing it to 8 bits, by adding gaussian noise (with variance drawn uniformly in the [0, 0.03] interval), and by combining all three degradations. Here, the decoder predicts an HDR image in the same scale as the input LDR, and the subnetwork predicts the scalar which multiplies the HDR to obtain the luminance map. The results in tab. 1 show the RMSE and its scale-invariant version (siRMSE) [8] (both weighted by solid angles), and HDR-VDP-3 [48] to measure the visual quality of the prediction. We observe that the network is robust to the degradations.

Per-pixel color Here, we explore the capacity to predict the per-pixel CCT, having as input a LDR image with random WB augmentation. The HDR input is first reexposed so that its 90th percentile maps to 0.8, clipped to [0, 1], and the WB is augmented to a random preset using [2]. The predictions are scored using RMSE as well as the relative error between the prediction and the ground truth. Fig. 7 shows qualitative results of color prediction. The mean relative error and RMSE are 4.25 % and 173.0 on the entire test set. We observe that the network struggles with larger color variations across the image. However, CCT is accurately predicted despite color changes in the input.

Planar illuminance Here, we experiment with three types of inputs: a photometric HDR image ("HDR"), a linear LDR image (reexposed HDR clipped to the [0, 1] interval) with ("LDR+scale") and without ("LDR") knowledge of the exposure. In addition, we also evaluate the impact of the FOV of the input. We experiment with FOVs of 180°, 120°, 60°, and uniformly random in the [60, 120]° interval. The image is stored in an equirectangular representation for 180°, and perspective projection for the other, lower FOVs.

Tab. 2 shows the results of these series of experiments. We report the RMSE and R 2 for each combination of input type and FOV. First, observe that the experiment with a FOV of 180° with the HDR image (top-left in tab. 2) amounts to learning the illuminance integration (eq. (1)). Unsurprisingly, narrowing the FOV results in decreased performance, due to the hidden lights beyond the FOV which may directly affect the planar illuminance. Limiting the photometric input information by clipping the HDR but keeping the correct scale (LDR+scale), reduces the scores moderately. The network now has to predict the amount of light beyond the clipped pixels. Discarding the scale from the input significantly worsens the results, indicating that full dynamic range (HDR) and/or knowing the absolute exposure are both necessary conditions for accurate illuminance inference.

6. Generalization to another camera

We now present experiments to evaluate the usefulness of our proposed dataset for predicting luminance and color

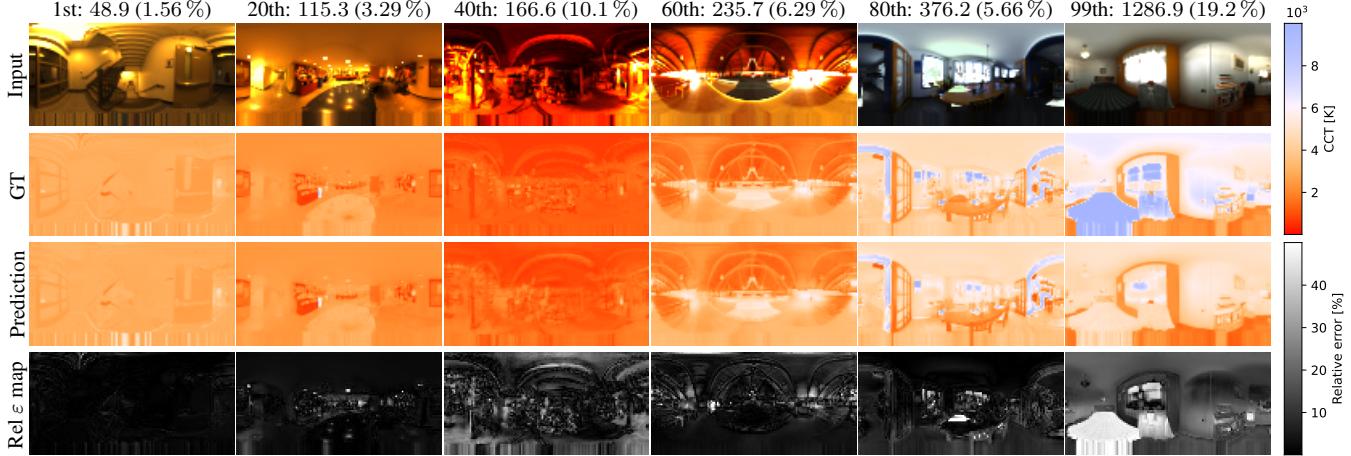


Figure 7. Examples of color prediction. The first row indicates the RMSE percentile: the RMSE (relative error). The “input” is the calibrated HDR reexposed, clipped, with a random WB augmenter [2]. Other rows show the ground truth and predicted CCT maps, and the relative error map. Colormaps for both the CCT and the relative error are shown at the right.

values from images captured with another camera. To this end, we rely on the Ricoh Theta Z1, an off-the-shelf 360° camera, and captured a small dataset of 74 calibrated HDR panoramas (referred to as “Theta dataset”) using the same process as in sec. 3. In addition to the bracketed RAW images, we also capture well-exposed LDR images (in jpeg format) produced by the camera. This small experimental dataset will also be released publicly. As opposed to [46] which calibrated this camera for photometric measurements, we calibrate for color as well as luminance.

Architecture and data The network architecture for each task is kept the same as in sec. 5. The networks are first trained on the calibrated Laval dataset with synthetically degraded LDR inputs (all degradations). The Theta dataset is split 40%-10%-50% for train-val-test respectively. The pre-trained networks are then fine-tuned on the Theta dataset (with jpeg images as input) with the same learning rate.

Experimental results Tab. 3 shows the results of the experiments for each task (cf. sec. 5.1). We experiment with a degraded LDR as input to the models pretrained on our dataset and the jpeg image of the camera as input to the pre-trained and fine-tuned models. The input images have 120° FOV for the planar illuminance prediction task, and all jpeg images are captured with the same white balance setting.

Directly providing the jpeg images as inputs to the pre-trained model (2nd row in tab. 3) results in significantly degraded performance across most tasks as compared to using degraded LDR images. This shows that the domain gap between produced jpeg images and simulated LDR images is still wide. Fine-tuning the networks on jpeg inputs is therefore necessary to obtain performance similar or sometimes slightly better than those obtained on the synthetic LDR.

	Luminance			Color		Illuminance	
Input	RMSE \downarrow	siRMSE \downarrow	HV3 \uparrow	RMSE \downarrow	rel $\varepsilon\downarrow$	RMSE \downarrow	$R^2\uparrow$
PT LDR	130.6	86.9	95.5	334.0	11.97	153.2	0.469
PT Jpeg	170.0	100.6	90.3	676.4	25.06	141.9	0.314
FT Jpeg	156.7	96.4	91.4	177.9	5.52	143.3	0.385

Table 3. Domain adaptation on a real-world dataset for which a jpeg image of a scene, as well as the calibrated luminance map, are captured with a Ricoh Theta Z1. Here, “HV3” refers to the HDR-VDP-3 metric [48]. We report performance on all three tasks from sec. 5.1. Each row corresponds to degraded LDR and jpeg as input to the pre-trained model (“PT LDR” and “PT Jpeg” resp.), and jpeg as input to the fine-tuned model (“FT Jpeg”).

7. Conclusion

We present the Laval Photometric Indoor HDR Dataset, the first photometrically accurate, large-scale dataset of HDR panoramic images. Our calibration method relies on a carefully curated calibration dataset of RAW exposure brackets captured with the original camera and a chroma meter. We also capture another small calibrated dataset with a Ricoh Theta Z1 for experiments on jpeg inputs. We present baselines for three novel tasks: per-pixel luminance, per-pixel color and planar illuminance predictions. We hope this new dataset will spur and catalyze research by empowering others to explore novel photometric and colorimetric tasks in computer vision, such as white balance prediction under multiple illuminations, physically based inverse rendering and “in the wild” image relighting.

Acknowledgements This research was supported by Sentinel North, NSERC grant RGPIN 2020-04799, and the Digital Research Alliance Canada. The authors thank Mojtaba Parsaei and Anthony Gagnon for their help with the chroma meter and the Theta Z1 calibration.

References

- [1] Polyhaven HDRIs. <https://polyhaven.com/hdris>, 2023. 2
- [2] Mahmoud Afifi and Michael S. Brown. What else can fool deep learning? addressing color constancy errors on deep neural network performance. In *Int. Conf. Comput. Vis.*, 2019. 6, 7, 8
- [3] Mahmoud Afifi and Michael S. Brown. Deep white-balance editing. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 2
- [4] Mahmoud Afifi, Marcus A Brubaker, and Michael S Brown. Auto white-balance correction for mixed-illuminant scenes. In *IEEE Winter conf. App. Comput. Vis.*, 2022. 2
- [5] Francesco Banterle, Patrick Ledda, Kurt Debattista, and Alan Chalmers. Inverse tone mapping. In *Int. Conf. Comput. Graph. Int. Tech.*, 2006. 2
- [6] Francesco Banterle, Patrick Ledda, Kurt Debattista, Alan Chalmers, and Marina Bloj. A framework for inverse tone mapping. *The Vis. Comput.*, 23(7):467–478, 2007. 2
- [7] Jonathan T Barron. Convolutional color constancy. In *Int. Conf. Comput. Vis.*, 2015. 2
- [8] Jonathan T Barron and Jitendra Malik. Shape, illumination, and reflectance from shading. *IEEE Trans. Pattern Anal. Mach. Intell.*, 37(8):1670–1687, 2014. 7
- [9] Jonathan T Barron and Yun-Ta Tsai. Fast fourier color constancy. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2017. 2
- [10] Simone Bianco and Claudio Cusano. Quasi-unsupervised color constancy. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2019. 2
- [11] Jane Brox. *Brilliant: the evolution of artificial light*. Houghton Mifflin Harcourt, 2010. 1
- [12] Dongliang Cheng, Dilip K Prasad, and Michael S Brown. Illuminant estimation for color constancy: why spatial-domain methods work and the role of the color distribution. *JOSA A*, 31(5):1049–1058, 2014. 2
- [13] Thomas W Cronin, Sönke Johnsen, N Justin Marshall, and Eric J Warrant. Visual ecology. In *Visual Ecology*. Princeton University Press, 2014. 1
- [14] Commission Internationale de L'Eclairage (CIE). Ilv: International lighting vocabulary, 2nd edition. Technical report, 2020. 4
- [15] Paul E Debevec and Jitendra Malik. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH*. 2008. 1, 2
- [16] Jeanne F Duffy and Charles A Czeisler. Effect of light on human circadian physiology. *Sleep medicine clinics*, 4(2):165–177, 2009. 1
- [17] Yuki Endo, Yoshihiro Kanamori, and Jun Mitani. Deep reverse tone mapping. *ACM Trans. Graph.*, 36(6), 2017. 1, 2
- [18] Graham D Finlayson, Michal Mackiewicz, and Anya Hurlbert. Color correction using root-polynomial regression. *IEEE Trans. Image Process.*, 24(5):1460–1470, 2015. 2
- [19] Marc-André Gardner, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Christian Gagné, and Jean-François Lalonde. Deep parametric indoor lighting estimation. In *Int. Conf. Comput. Vis.*, 2019. 4
- [20] Marc-André Gardner, Kalyan Sunkavalli, Ersin Yumer, Xiaohui Shen, Emiliano Gambaretto, Christian Gagné, and Jean-François Lalonde. Learning to predict indoor illumination from a single image. *ACM Trans. Graph.*, 9(4), 2017. 1, 2
- [21] Peter Vincent Gehler, Carsten Rother, Andrew Blake, Tom Minka, and Toby Sharp. Bayesian color constancy revisited. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2008. 2
- [22] Pulkit Gera, Mohammad Reza Karimi Dastjerdi, Charles Renaud, P. J. Narayanan, and Jean-François Lalonde. Casual indoor HDR radiance capture from omnidirectional images. In *Brit. Mach. Vis. Conf.*, 2022. 2
- [23] Arjan Gijsenij, Rui Lu, and Theo Gevers. Color constancy for multiple light sources. *IEEE Trans. Image Process.*, 21(2):697–707, 2011. 2
- [24] Miguel Granados, Kwang In Kim, James Tompkin, and Christian Theobalt. Automatic noise modeling for ghost-free hdr reconstruction. *ACM Trans. Graph.*, 32(6):1–10, 2013. 2
- [25] David Griffiths, Tobias Ritschel, and Julien Philip. Outcast: Single image relighting with cast shadows. *Comput. Graph. Forum*, 43, 2022. 6
- [26] Michael D Grossberg and Shree K Nayar. Modeling the space of camera response functions. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(10):1272–1282, 2004. 2
- [27] Bjoern Haefner, Simon Green, Alan Oursland, Daniel Andersen, Michael Goesele, Daniel Cremers, Richard Newcombe, and Thomas Whelan. Recovering real-world reflectance properties and shading from HDR imagery. In *Int. conf. 3D Vis.*, 2021. 1
- [28] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2016. 6
- [29] Felix Heide, Markus Steinberger, Yun-Ta Tsai, Mushfiqur Rouf, Dawid Pajak, Dikpal Reddy, Orazio Gallo, Jing Liu, Wolfgang Heidrich, Karen Egiazarian, Jan Kautz, and Kari Pulli. FlexISP: A flexible camera image processing framework. *ACM Trans. Graph.*, 33(6), 2014. 1
- [30] Eugene Hsu, Tom Mertens, Sylvain Paris, Shai Avidan, and Frédéric Durand. Light mixture estimation for spatially varying white balance. In *ACM Trans. Graph.*, volume 27, pages 1–7. 2008. 2
- [31] Xin Huang, Qi Zhang, Feng Ying, Hongdong Li, Xuan Wang, and Qing Wang. Hdr-nerf: High dynamic range neural radiance fields. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022. 2
- [32] B Jung and M Inanici. Measuring circadian lighting through high dynamic range photography. *Lighting Res. & Tech.*, 51(5):742–763, 2019. 2
- [33] Nima Khademi Kalantari and Ravi Ramamoorthi. Deep high dynamic range imaging of dynamic scenes. *ACM Trans. Graph.*, 36(4), 2017. 2
- [34] Hakki Can Karaimer and Michael S. Brown. A software platform for manipulating the camera imaging pipeline. In *Eur. Conf. Comput. Vis.*, 2016. 1
- [35] Seon Joo Kim, Jan-Michael Frahm, and Marc Pollefeys. Radiometric calibration with illumination change for outdoor

- scene analysis. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2008. 2
- [36] Seon Joo Kim and Marc Pollefeys. Robust radiometric calibration and vignetting correction. *IEEE Trans. Pattern Anal. Mach. Intell.*, 30(4):562–576, 2008. 2
- [37] Pin-Hung Kuo, Chi-Sun Tang, and Shao-Yi Chien. Content-adaptive inverse tone mapping. In *Vis. Comm. Image Proc.*, 2012. 2
- [38] Bruno Lecouat, Thomas Eboli, Jean Ponce, and Julien Mairal. High dynamic range and super-resolution from raw image bursts. *ACM Trans. Graph.*, 41(4), jul 2022. 2
- [39] Siyeong Lee, Gwon Hwan An, and Suk-Ju Kang. Deep chain HDR: Reconstructing a high dynamic range image from a single low dynamic range image. *IEEE Access*, 2018. 1, 2
- [40] Bo Li, Lionel Heng, Kevin Koser, and Marc Pollefeys. A multiple-camera system calibration toolbox using a feature descriptor-based calibration pattern. In *IEEE/RSJ Int. Conf. Intell. Robots Syst.*, 2013. 3
- [41] Jinghui Li and Peiyu Fang. HDRNET: Single-image-based hdr reconstruction using channel attention cnn. In *Int. Conf. Multimedia Sys. Sig. Process*, 2019. 1, 2
- [42] Haiting Lin, Seon Joo Kim, Sabine Süsstrunk, and Michael S Brown. Revisiting radiometric calibration for color computer vision. In *Int. Conf. Comput. Vis.*, 2011. 2
- [43] Stephen Lin, Jinwei Gu, Shuntaro Yamazaki, and Heung-Yeung Shum. Radiometric calibration from a single image. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2004. 2
- [44] Stephen Lin and Lei Zhang. Determining the radiometric response function from a single grayscale image. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2005. 2
- [45] Yu-Lun Liu, Wei-Sheng Lai, Yu-Sheng Chen, Yi-Lung Kao, Ming-Hsuan Yang, Yung-Yu Chuang, and Jia-Bin Huang. Single-image HDR reconstruction by learning to reverse the camera pipeline. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 1, 2, 5
- [46] Ian MacPherson, Richard F Murray, and Michael S Brown. A 360° omnidirectional photometer using a Ricoh Theta Z1. In *Color Imag. Conf.*, 2022. 2, 8
- [47] Andres Magnusson and Diane Boivin. Seasonal affective disorder: an overview. *Chronobiology international*, 20(2):189–207, 2003. 1
- [48] Rafal Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. Hdr-vdp-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Trans. Graph.*, 30(4), jul 2011. 7, 8
- [49] Demetris Marnerides, Thomas Bashford-Rogers, Jonathan Hatchett, and Kurt Debattista. ExpandNet: A deep convolutional neural network for high dynamic range expansion from low dynamic range content. *Comput. Graph. Forum*, 37(2), 2018. 1, 2
- [50] Christopher Mei and Patrick Rives. Single view point omnidirectional camera calibration from planar grids. In *Int. Conf. Robot. Aut.*, 2007. 3
- [51] Christopher A Metzler, Hayato Ikoma, Yifan Peng, and Gordon Wetzstein. Deep optics for single-shot high-dynamic-range imaging. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 2
- [52] 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 *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022. 2
- [53] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Comm. of the ACM*, 65(1):99–106, 2021. 2
- [54] Richard F Murray and Wendy J Adams. Visual perception and natural illumination. *Current Opinion in Behavioral Sciences*, 30:48–54, 2019. 1
- [55] Seonghyeon Nam, Abhijith Punnapurath, Marcus A Brubaker, and Michael S Brown. Learning srgb-to-raw-rgb de-rendering with content-aware metadata. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2022. 1
- [56] Yuzhen Niu, Jianbin Wu, Wenxi Liu, Wenzhong Guo, and Rynson WH Lau. HDR-GAN: HDR image reconstruction from multi-exposed LDR images with large motions. *IEEE Trans. Image Process.*, 30:3885–3896, 2021. 2
- [57] C. Pierson, C. Cauwerts, M. Bodart, and J. Wienold. Tutorial: Luminance maps for daylighting studies from high dynamic range photography. *LEUKOS*, 17(2):140–169, 2021. 2
- [58] Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. Photographic tone reproduction for digital images. *ACM Trans. Graph.*, 21(3):267–276, 2002. 2
- [59] Allan G. Rempel, Matthew Trentacoste, Helge Seetzen, H. David Young, Wolfgang Heidrich, Lorne Whitehead, and Greg Ward. LDR2HDR: On-the-fly reverse tone mapping of legacy video and photographs. *ACM Trans. Graph.*, 26(3):39, 2007. 2
- [60] Mark A. Robertson, Sean Borman, and Robert L. Stevenson. Dynamic range improvement through multiple exposures. In *IEEE Int. Conf. Image Process.*, 1999. 2
- [61] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Med. Img. Comp. Comput.-Ass. Int.*, 2015. 6
- [62] Marcel Santana Santos, Ing Ren Tsang, and Nima Khademi Kalantari. Single image hdr reconstruction using a cnn with masked features and perceptual loss. *ACM Trans. Graph.*, 39(4), 2020. 2
- [63] Wu Shi, Chen Change Loy, and Xiaoou Tang. Deep specialized network for illuminant estimation. In *Eur. Conf. Comput. Vis.*, 2016. 2
- [64] Qilin Sun, Ethan Tseng, Qiang Fu, Wolfgang Heidrich, and Felix Heide. Learning rank-1 diffractive optics for single-shot high dynamic range imaging. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2020. 2
- [65] Greg Ward. Fast, robust image registration for compositing high dynamic range photographs from hand-held exposures. *ACM Trans. Graph.*, 8(2):17–30, 2003. 2
- [66] Wei Wei, Li Guan, Yue Liu, Hao Kang, Haoxiang Li, Ying Wu, and Gang Hua. Beyond visual attractiveness: Physically plausible single image hdr reconstruction for spherical panoramas. *ArXiv*, abs/2103.12926, 2021. 2
- [67] Ying Xiong, Kate Saenko, Trevor Darrell, and Todd Zickler. From pixels to physics: Probabilistic color de-rendering. In *IEEE Conf. Comput. Vis. Pattern Recog.*, 2012. 1

- [68] Hanning Yu, Wentao Liu, Chengjiang Long, Bo Dong, Qin Zou, and Chunxia Xiao. Luminance attentive networks for hdr image and panorama reconstruction. *Comput. Graph. Forum*, 40(7), 2021. 2
- [69] Hongyi Zhang, Yann N Dauphin, and Tengyu Ma. Fixup initialization: Residual learning without normalization. In *Int. Conf. Learn. Represent.*, 2019. 6
- [70] Jinsong Zhang and Jean-François Lalonde. Learning high dynamic range from outdoor panoramas. In *Int. Conf. Comput. Vis.*, 2017. 2