Deep Sky Modeling for Single Image Outdoor Lighting Estimation

Yannick Hold-Geoffroy*
Adobe Research
holdgeof@adobe.com

Akshaya Athawale*
Indian Institute of Tech. Dhanbad
akshaya.15je001564@am.ism.ac.in

Jean-François Lalonde
Université Laval
jflalonde@gel.ulaval.ca

Abstract

We propose a data-driven learned sky model, which we use for outdoor lighting estimation from a single image. As no large-scale dataset of images and their corresponding ground truth illumination is readily available, we use complementary datasets to train our approach, combining the vast diversity of illumination conditions of SUN360 with the radiometrically calibrated and physically accurate Laval HDR sky database. Our key contribution is to provide a holistic view of both lighting modeling and estimation, solving both problems end-to-end. From a test image, our method can directly estimate an HDR environment map of the lighting without relying on analytical lighting models. We demonstrate the versatility and expressivity of our learned sky model and show that it can be used to recover plausible illumination, leading to visually pleasant virtual object insertions. To further evaluate our method, we capture a dataset of HDR 360° panoramas and show through extensive validation that we significantly outperform previous state-of-the-art.

1. Introduction

The lighting conditions of outdoor scenes can create significant differences in the scene appearance depending on the weather and the time of day. Indeed, one need only consider the striking contrast created by bright highlights and dark shadows at noon, the warm, orange hues of the golden hour, or the gray ominous look of overcast conditions. This wide variety of effects is challenging for approaches that attempt to estimate the lighting conditions from outdoor images.

A popular solution to this problem involves capturing objects of known geometry and reflectance properties (notably, a chrome sphere [6]). Another solution, which does not require access to the scene, is to approximate outdoor lighting with low-dimensional, parametric models. This has the advantage of drastically reducing the dimensionality of the problem down to just a handful of variables, which can more easily be estimated from an image. This insight has recently been exploited to successfully learn to predict lighting from a single outdoor image [12]. In particular, they propose to represent outdoor lighting using the Hošek-Wilkie (HW) sky model [13, 14], which can model high dynamic range (HDR) sky domes using as few as 4 parameters. They learn to predict lighting by fitting the HW model to a large database of outdoor, low dynamic range (LDR) panoramas and training a CNN to regress the HW parameters from limited field of view crops extracted from those panoramas.

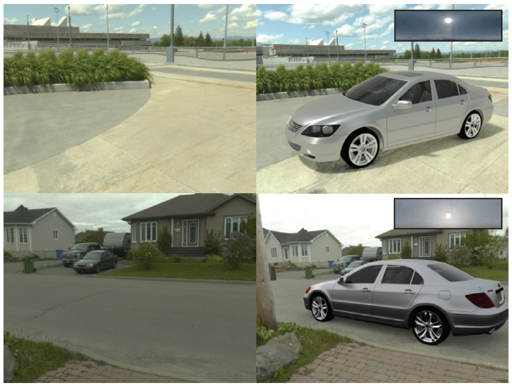


Figure 1. Our method can estimate HDR outdoor lighting conditions from a single image (left). This estimation can be used "as-is" to relight virtual objects that match the input image in both sunny (top-right) and overcast (bottom-right) weather. Our key contribution is to train both our sky model and lighting estimation end-to-end by exploiting multiple complementary datasets during training.

Unfortunately, approximating outdoor lighting analytically comes at a cost. Popular sky models (e.g. [13, 25, 24]) were developed to model *clear* skies with smoothly-varying amounts of atmospheric aerosols (represented by the commonly-used *turbidity* parameter). Therefore, they do not yield accurate representations for other types of common weather conditions such as partially cloudy or completely overcast skies. For example, consider the different lighting conditions in fig. 2, which we represent with the HW parametric model using the non-linear fitting approach of [12]. Note how the HW approximation works well in clear skies (top) but degrades as the cloud cover increases (bottom). Can we obtain a lighting model that is low-dimensional, that can accurately describe the wide variety of outdoor lighting conditions, and that can be estimated from a single image?

In this paper, we propose an answer to this question by *learning* an HDR sky model directly from data. Our non-analytical data-driven sky model can be estimated directly from a single image captured outdoors. Our approach suc-

[*] Parts of this work were completed while Y. Hold-Geoffroy and A. Athawale were at U. Laval.

cessfully models a much larger set of lighting conditions than previous approaches (see fig. 2).

To learn to estimate a non-parametric lighting model from a single photograph, we propose a three-step approach which bears resemblance to the "T-network" architecture proposed by [9], and rely on a variety of existing complementary datasets. First, we train a deep sky autoencoder that learns a data-driven, deep HDR sky model. To train this sky autoencoder, we rely on the Laval HDR sky database [20, 22], a large dataset of unsaturated HDR hemispherical sky images. Second, we project the SUN360 LDR outdoor panorama dataset [28] to HDR, using the "LDR2HDR" network of Zhang and Lalonde [30], and subsequently map each panorama to the latent space of HDR skies from our sky autoencoder. This effectively provides non-parametric sky labels for each panorama. Third, we train an image encoder that learns to estimate these labels from a crop, similarly to [12].

In short, our main contributions are the following:

- we propose a novel sky autoencoder, dubbed "SkyNet"[1], that can accurately represent outdoor HDR lighting in a variety of illumination conditions;
- we show how HDR lighting can be estimated from a single image, modeling a much wider range of illumination conditions than previous work;
- we capture a new dataset of 206 radiometrically calibrated outdoor HDR 360° panoramas;
- we demonstrate, through a series of experiments and a user study, that our approach outperforms the state-of-the-art both qualitatively and quantitatively;
- we offer a technique to bridge the gap between our implicit parameters representation and the versatility of low-dimensional parametric sky models.

2. Related work

Outdoor lighting modeling and estimation have been studied extensively over the past decades. For conciseness, we will focus on outdoor lighting modeling and estimation that is most related to this work.

Outdoor lighting modeling Modeling the sky is a challenging research problem that has been well studied across many disciplines such as atmospheric science, physics, and computer graphics. The Perez All-Weather model [24] was first introduced as an improvement over the previous CIE Standard Clear Sky model, and modeled weather variations using 5 parameters. Preetham et al. [25] later present a simplified model, which relies on a single physically grounded parameter, the atmospheric turbidity. Hošek and Wilkie subsequently proposed an improvement over the Preetham model, which is comprised of both a sky dome [13] and solar disc [14] analytical models. See [17] for a comparison of these analytic sky models.

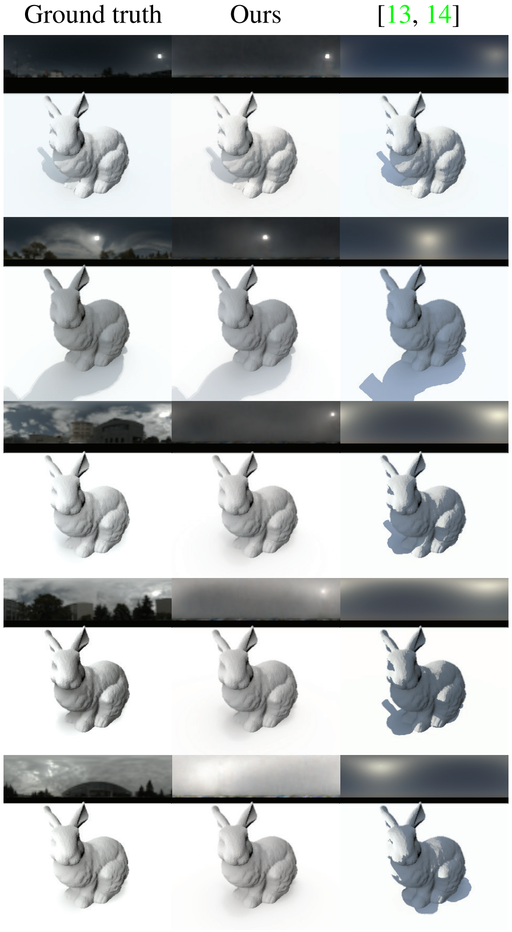


Figure 2. Examples of our 360° unsaturated HDR database (left), our reconstruction using our learned sky model (center) and the Hošek-Wilkie sun and sky models [13, 14] fit using the optimization described in [12] (right). Renders of each method are shown below the panorama. Note how our sky model can accurately produce a wide variety of lighting conditions from sunny (top) to overcast (bottom) and their corresponding shadow contrast and smoothness.

Outdoor lighting estimation Lighting estimation from a single, generic outdoor scene has first been proposed by Lalonde et al. [21]. Their approach relies on the probabilistic combination of multiple cues (such as cast shadows, shading, and sky appearance variation) extracted individually from the image. Karsch et al. [16] propose to match the background image to a large dataset of panoramas [28] and transfer the panorama lighting (obtained through a specially-designed light classifier) to the image. However, the matching metric may yield results that have inconsistent lighting. Other approaches rely on known geometry [23] and/or strong priors on geometry and surface reflectance [1].

[1] Luckily, it has not (yet) gained artificial consciousness [5].

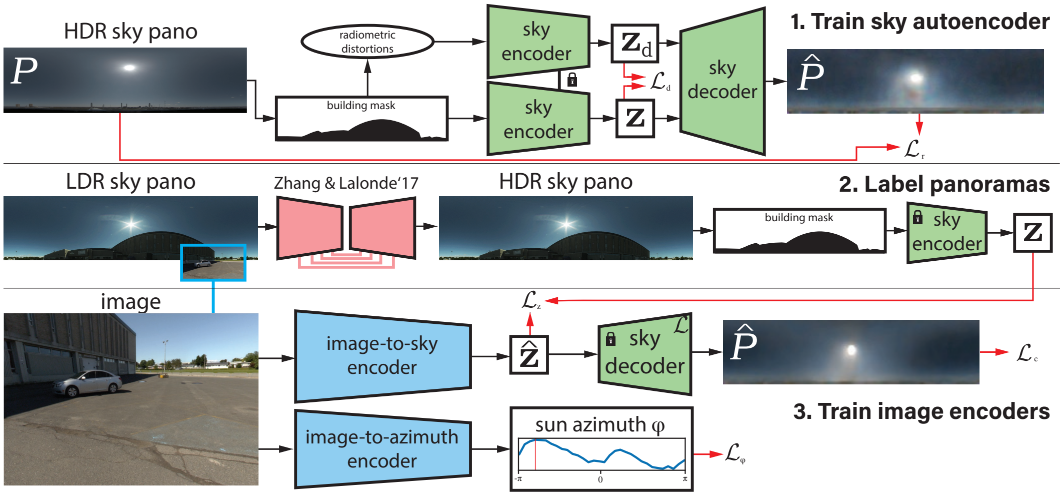


Figure 3. Overview of our proposed three-step approach. First, we train an autoencoder to learn a 64-parameters latent space of skies $\mathbf{z}$ from a large dataset of calibrated skies [20], while enforcing its encoder to be robust to distortions in white balance, exposure and occlusions. Second, we convert the SUN360 LDR panorama dataset to HDR using [30] and obtain their $\mathbf{z}$ vectors with the trained autoencoder. Finally, we train two image encoders to learn the mapping between crops from SUN360, the sun azimuth and their corresponding $\mathbf{z}$. Please see text for the definitions of the loss functions $\mathcal{L}_*$.

Deep learning for lighting estimation Deep learning has also been recently used for lighting estimation. For example, Georgoulis et al. [8] learn to estimate lighting and reflectance from an object of known geometry, by first estimating its reflectance map (i.e., its "orientation-dependent" appearance) [26] and subsequently factoring it into lighting and material properties [7]. Closer to our work, Hold-Geoffroy et al. [12] model outdoor lighting with the parametric, Hošek-Wilkie sky model, and learn to estimate its parameters from a single image. As mentioned above, we take inspiration from this work and significantly improve upon it by proposing to instead use a learned, data-driven outdoor lighting model. Concurrent to this work, Zhang et al. [31] extend [12] with a more flexible parametric sky model. In another closely-related paper, Calian et al. [2] estimate HDR outdoor lighting from a single face image. While they employ a similar deep autoencoder to learn a data-driven model, they rely on a multi-step non-linear optimization approach over the space of face albedo and sky parameters, which is time-consuming and prone to local minima. In contrast, we learn to estimate lighting from a single image of a generic outdoor scene in an end-to-end framework. In addition, our training procedure is more robust to sky occluders (such as buildings and trees) and non-linear radiometric distortions. Cheng et al. [3] estimate lighting from the front and back camera of a mobile phone. However, they represent lighting using low-frequency spherical harmonics, which, as shown in [2], does not appropriately model outdoor lighting.

3. Overview

The goal of our technique is to estimate the illumination conditions from an outdoor image. Directly training such a method in a supervised manner is currently impossible as no large-scale dataset of images and their corresponding illumination conditions is yet available. We therefore propose the following 3-step approach, which is also illustrated in fig. 3.

1. Train the SkyNet autoencoder on HDR skies The first step (fig. 3, top row) is to learn a data-driven sky model from the 33,420 hemispherical sky images in the Laval HDR sky database [20] using a deep autoencoder. The autoencoder, dubbed "SkyNet", learns the space of outdoor lighting by compressing an HDR sky image to a 64-dimensional latent vector $\mathbf{z}$, and reconstructing it at the original resolution. Robustness to white balance, exposure, and occlusions is enforced during training. More details on this step are presented in sec. 4.

2. Label LDR panoramas with SkyNet The second step (fig. 3, middle row) is to use the learned SkyNet autoencoder to obtain $\mathbf{z}$ vectors for a large dataset of panoramas. For this, the Laval HDR sky database cannot be reused as it only contains sky hemispheres. Instead, we take advantage of the wide variety of scenes and lighting conditions captured by the SUN360 panorama dataset [28]. Each panorama is first converted to HDR with the approach of [30] that has been trained specifically for this purpose. Then, sky masks are estimated using the sky segmentation approach of [12] based on a dense CRF [19]. The resulting HDR panoramas, which we dub SUN360-HDR, along with their sky masks are forwarded to the SkyNet encoder to recover $\mathbf{z}$. This has the effect of labeling each panorama in SUN360 with a compact, data-driven representation for outdoor illumination.

3. Train image encoders to predict illumination Finally, the last step (fig. 3, bottom row) is to train an image encoder on limited field of view images extracted from the SUN360

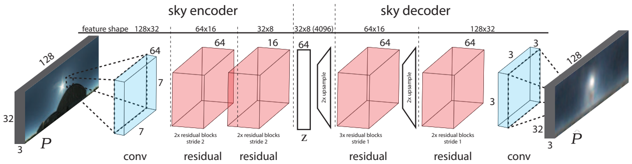


Figure 4. Architecture of our SkyNet deep autoencoder showing the parameters of each layer. ELU [4] activation functions are used after the convolutional layers (blue). Residual blocks (red) [11] have the ReLU activation functions.

dataset, by employing the methodology proposed in [12]. The main difference here is that we train the neural network to predict the $\mathbf{z}$ vector from the previous step corresponding to each crop, instead of the analytical sky parameters as in the previous work. The full HDR sky image can be recovered using the SkyNet decoder. The resulting sky image can be used "as is" as image-based lighting to photorealistically render 3D objects into images with a variety of illumination conditions. We detail this step in sec. 5.

4. Training the SkyNet deep sky model

In this section, we describe SkyNet, our deep autoencoder acting as our sky model, its architecture and training steps.

4.1. Deep autoencoder

To learn our sky model, we adopt an autoencoder architecture which projects a full HDR sky down to 64 parameters (encoder), and subsequently reconstructs it (decoder). This is conceptually similar to [2], with the key differences that we employ a more robust training scheme which includes occlusions and radiometric distortions, making it amenable to full end-to-end learning rather than the non-linear inverse rendering framework of [2]. In addition, we employ a different architecture based on residual layers [11] (see fig. 4).

To represent the sky, we use the equirectangular (latitude-longitude) projection at a resolution of 32×128 in RGB of the up hemisphere. This representation has the advantage of being easy to rotate along the azimuth with a horizontal shift of the image. Similarly to [2, 30], we rotate the panoramas along their azimuth so that the sun is in the center of the image, as we empirically found that training the sky model is simpler and more well-behaved this way. However, unlike its azimuth, we cannot decouple the sun elevation from the sky reconstruction as it influences the sun intensity, color, and overall sky luminance distribution [24].

The SkyNet autoencoder training is mostly performed on the 33,420 panoramas of the Laval HDR sky database [20, 22], which we augment with 7000 panoramas from SUN360-HDR [28, 30] (see sec. 3), both of which include the full dynamic range of the sun. We resize each panorama down to a resolution of 32×128, ensuring that the sky integral remains constant by taking the solid angles into account.

Parameter	Equation	Distribution	Bounds
Exposure (e)	$\mathbf{P}_d = e\mathbf{P}$	$\mathcal{O}(0.2, \sqrt{0.2})$	[0.1, 10]
White bal. ($\mathbf{w}$)	$\mathbf{P}_{d,c} = \mathbf{w}_c \mathbf{P}_c$	$\mathcal{N}(0, 0.06)$	[0.8, 1.2]
Gamma (γ)	$\mathbf{P}_d = \mathbf{P}^{1/\gamma}$	$\mathcal{O}(0.0035, \sqrt{0.2})$	[0.85, 1.2]

Table 1. Parameters used to generate radiometrically distorted versions $\mathbf{P}_d$ of the panoramas $\mathbf{P}$. Here, c denotes the color channel, $\mathcal{N}(\mu, \sigma^2)/\mathcal{O}(\mu, \sigma^2)$ indicate a normal/lognormal distribution.

While the Laval sky database contains unoccluded sky images, panoramas in the SUN360-HDR may contain multiple buildings and other sky occluders which we do not want to learn in our sky model. To prevent SkyNet from learning non-sky features, we reuse the sky segmentation of [12] (based on a CRF refinement [19]) to mask non-sky regions of SUN360-HDR with black pixels. To enforce SkyNet to estimate plausible sky appearance in those regions, we randomly apply black regions to the training images from the Laval sky database and ask the network to recover the original, unoccluded sky appearance. Specifically, we apply, with 50% chance, the non-sky mask from a random SUN360-HDR panorama. This is only done on the Laval sky panoramas, as SUN360-HDR already contains buildings occluding the sky. This requires the neural network to fill in the holes and predict the sky energy distribution under occluded regions.

4.2. Training losses

To obtain robustness to occlusions and radiometric distortions, we train SkyNet using a combination of two losses, as illustrated in the top part of fig. 3. First, two versions of the panorama are fed through the network, one after the other: the original $\mathbf{P}$ and a second one to which we applied random radiometric distortions $\mathbf{P}_d$. These random distortions consist of variations in exposure, white balance and camera response function as described in table 1.

Denoting $\text{enc}(\cdot)$ as the encoder, the first loss used to train the sky autoencoder enforces both the undistorted $\mathbf{z} = \text{enc}(\mathbf{P})$ and distorted $\mathbf{z}_d = \text{enc}(\mathbf{P}_d)$ to be as close as possible by minimizing the L1 norm between them:

$$\mathcal{L}_d = \|\mathbf{z}_d - \mathbf{z}\|_1 \, . \tag{1}$$

This loss encourages the sky encoder to be robust to radiometric distortions that may be present in the input panoramas. Our second loss is the typical autoencoder reconstruction loss, with the difference that both the undistorted and distorted inputs must reconstruct the original panorama using an L1 loss:

$$\mathcal{L}_r = \|\hat{\mathbf{P}} - \mathbf{P}\|_1 + \|\hat{\mathbf{P}}_d - \mathbf{P}\|_1 \, , \tag{2}$$

where $\hat{\mathbf{P}} = \text{dec}(\mathbf{z})$ and $\hat{\mathbf{P}}_d = \text{dec}(\mathbf{z}_d)$ are the panoramas reconstructed by the decoder $\text{dec}(\cdot)$. The reconstruction loss $\mathcal{L}_r$ is only computed on sky pixels in the original panorama $\mathbf{P}$. For example, this loss is not active for regions masked by buildings in panoramas from SUN360-HDR, as no ground truth sky appearance is known for this region. The autoencoder is never penalized for any output in these regions. On the Laval HDR sky panoramas, this loss is active everywhere, even for randomly masked (black) regions. The target appearance for those regions is the original sky pixels of the panorama before the sky was masked, effectively asking the autoencoder to extrapolate—or fill—the region with plausible sky appearance.

Our sky autoencoder is trained with:

$$\mathcal{L}_s = \mathcal{L}_r + \lambda_d \mathcal{L}_d \,, \tag{3}$$

where we empirically set $\lambda_d = 100$ in order to balance the gradient magnitude between $\mathcal{L}_r$ and $\mathcal{L}_d$ during training.

Example sky reconstructions on test panoramas are shown in the middle column of fig. 2. While LDR content such as clouds is lost, the reconstructed panoramas $\hat{\mathbf{P}}$ properly model the energy distribution of the sky and are thus able to faithfully reproduce shadow characteristics like contrast and sharpness. In contrast, while the Hošek-Wilkie sky model properly approximates clear skies, it does not generalize to non-clear skies (right-most column in fig. 2).

4.3. Implementation details

Our sky model holds approximately 1 million parameters which are learned using the Adam [18] optimizer with a learning rate of 10^{-3} and $\beta = (0.5, 0.999)$. We additionally reduce the learning rate by a factor of 10 whenever the minimum error on the validation set has not decreased over the last 10 epochs. Convergence is monitored on the validation set, which is comprised of 14 days (3999 panoramas) from the Laval HDR sky database that we removed from the training set and 2000 panoramas from SUN360-HDR (sec. 3), different from the ones chosen to augment the training set. Training convergence was obtained after 127 epochs in our case, taking roughly 4 hours on a Titan Xp GPU. Sky inference takes approximately 10ms on the same machine. We (un)normalize the input (output) panoramas using the training set mean and standard deviation.

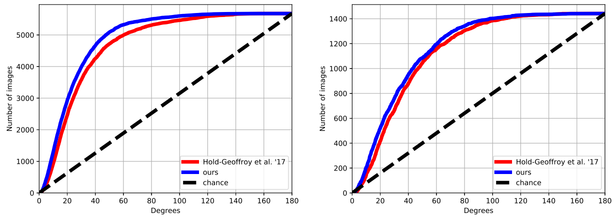
5. Learning to estimate illumination from a single image

In this section, we describe the third step of our approach (c.f. sec. 3 and fig. 3), that is, how we learn to estimate both the sun azimuth φ and the sky parameters $\mathbf{z}$ of our learned sky model from a single, limited field of view image.

5.1. Image lighting estimation

To estimate the sky parameters $\mathbf{z}$ from a limited field of view image, we use a pretrained DenseNet-161 [15] architecture where the last layer was replaced by a fully connected layer of 64 outputs. We finetune this image-to-sky model on sky parameters $\mathbf{z}$ using an L2 loss:

$$\mathcal{L}_z = \|\hat{\mathbf{z}} - \mathbf{z}\|_2 \,. \tag{4}$$



(a) SUN360

(b) HDR panoramas

Figure 5. Comparison of sun position estimations between our proposed model (blue) and Hold-Geoffroy et al. [12] showing the cumulative sun position estimation error on (a) their SUN360 test set and (b) our HDR 360° captures. Using a recent network architecture (DenseNet-161 [15]) grants our technique a slight improvement over the network used by [12].

We observed that this loss on the space of $\mathbf{z}$ alone failed to capture the details in the sky energy distribution and tended to produce average skies without strong sun intensities. To solve this issue, we added an L1 loss on the sky panoramas reconstructed from $\hat{\mathbf{z}}$ and $\mathbf{z}$ by the SkyNet decoder:

$$\mathcal{L}_c = \|(\mathrm{dec}(\hat{\mathbf{z}}) - \mathrm{dec}(\mathbf{z})) \odot \mathbf{d\Omega}\|_1 \,, \tag{5}$$

where $\mathrm{dec}(\cdot)$ denotes the SkyNet decoder, $\mathbf{d\Omega}$ the matrix of solid angles spanned by each pixel in the sky panorama, and $\odot$ the element-wise multiplication operator.

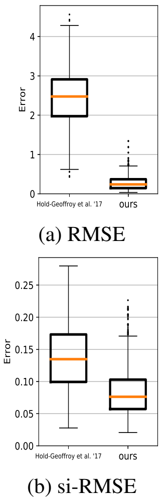
The image-to-sky encoder is trained by summing those two losses: $\mathcal{L}_i = \mathcal{L}_z + \lambda_c \mathcal{L}_c$. Due to the large difference in magnitude between $\mathcal{L}_z$ and $\mathcal{L}_c$, we empirically set $\lambda_c = 3 \times 10^{-10}$ to prevent gradient imbalance during training.

5.2. Sun azimuth estimation

Due to our sky model training (sec. 4.1), the sun will invariably be located in the center column of the estimated sky panorama. We therefore need to estimate the sun azimuth φ to rotate the lighting according to the sun position in the image. Both tasks seem to be closely related, hinting that both could benefit from joint training [12, 29]. However, training a single model to estimate both the sky parameters $\mathbf{z}$ and sun azimuth φ proved difficult. In our experiments, balancing both tasks using a fixed ratio between the losses failed to obtain good generalization performance for both tasks simultaneously. To circumvent this issue, we train a different image-to-azimuth model to estimate a probability distribution of the sun azimuth φ. This sun azimuth distribution is obtained by discretizing the $[-\pi, \pi]$ range into 32 bins, similar to [12]. We use once again a pretrained DenseNet-161 where the last layer is replaced by a fully connected layer of 32 outputs. A Kullback-Leibler divergence loss $\mathcal{L}_\varphi$ with a one-hot target vector is used to train this neural network.

5.3. Implementation details

To train both the image-to-sky and image-to-azimuth encoders, we use the SUN360-HDR dataset which we augment



(a) RMSE

(b) si-RMSE

Figure 6. Quantitative relighting error on the bunny scene (see fig. 7). We compute two metrics comparing renders against ground truth lighting: (a) RMSE and (b) scale-invariant (si-)RMSE [10]. The lighting has been rotated for both methods so the sun is always at the same azimuth. The global intensity of our estimated environment map is generally closer to the ground truth most of the time, leading to an almost 10× improvement in RMSE over [12]. Additionally, our flexible learned sky model allows for increased shadow expressiveness and can handle completely overcast skies, enhancing the si-RMSE by over 60% over the previous state-of-the-art.

with 100 images extracted from 15 captured HDR panoramas (see sec. 6.1 for more details). To counter the severe imbalance between both data sources, we penalize errors committed on captured panoramas by a factor of 4.

To provide more stability to the training, the image-to-sky encoder is first trained for 5 epochs using a learning rate of 3×10^{-4} using only the loss on sky parameters $\mathcal{L}_z$. Afterward, both losses $\mathcal{L}_c$ and $\mathcal{L}_z$ are combined and the learning rate is set to 2×10^{-6}. The image-to-azimuth model was trained with a fixed learning rate of 3×10^{-4}. The Adam optimizer is used with $\beta = (0.4, 0.999)$ and a weight decay of 10^{-7} throughout the training for both the image-to-sky and sun azimuth estimator. Convergence of the image-to-sky and image-to-azimuth models were obtained after 55 and 3 epochs (roughly 5 hours of training each on a Titan Xp GPU), and inference takes roughly 30ms and 24ms, respectively.

6. Experimental validation

This section first presents the dataset used for evaluating and comparing our method to the state-of-the-art method of Hold-Geoffroy et al. [12]. Then, the performance of our proposed method is assessed with qualitative and quantitative results as well as a user study.

6.1. A dataset of outdoor HDR panoramas

The previous state-of-the-art on outdoor illumination estimation [12] proposed an evaluation based solely on SUN360, where the ground truth was obtained using their non-linear optimization on sky pixels to estimate sun intensity. We argue that evaluating on SUN360 does not provide an accurate quantitative relighting performance since it assumes that the non-linear fit accurately models all types of skies present in the panoramas, which is not the case (fig. 2).

To provide a more accurate assessment of the performance of our technique, we captured a new dataset of 206 HDR outdoor panoramas[2]. Following the recommendations of [27], each panorama captures the full 22 f-stops required to record the full unsaturated dynamic range of outdoor scenes. Using a Canon 5D Mark iii camera with an 8mm Sigma fisheye lens, a ND3.0 filter, and mounted on a GigaPan tripod head, we captured 6 sets (at 60° azimuthal increments) of 7 exposures (from 1/8000s to 8s shutter speed at f/14 aperture) in RAW mode. We then automatically stitched the results into a 360° HDR panorama using the

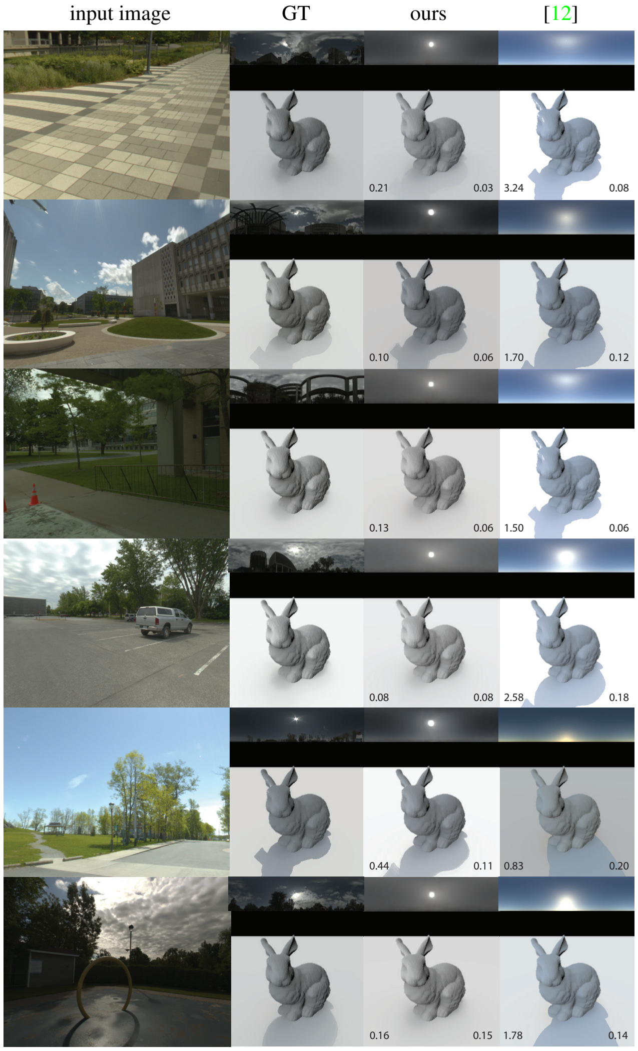


Figure 7. Qualitative relighting comparison between ground truth lighting (GT), our method, and Hold-Geoffroy et al. [12]. RMSE (SI-RMSE) are shown on the bottom left (right). Images from our HDR outdoor panorama dataset were cropped to obtain the input image. The renders using our estimated lighting display a wide variety of cast shadow characteristics such as sharp (sunny), smooth (lightly overcast) and absent (mostly overcast), which the parametric sky model of [12] cannot reproduce.

[2] Available at http://outdoor.hdrdb.com.

	ours	[12]
total votes	536 (69%)	244 (31%)

Table 2. Results of our user study ($N = 39$), which show that users overwhelmingly prefer results obtained with our technique over that of Hold-Geoffroy et al [12].

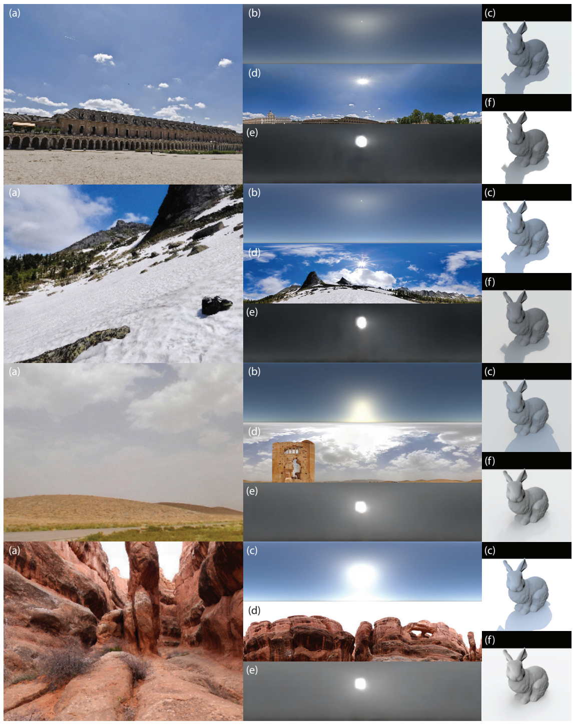


Figure 8. Qualitative relighting evaluation. From (a) an input image, we show the lighting estimation and render from Hold-Geoffroy et al. [12] (b-c) and our method (e-f) on SUN360. Note that no ground truth illumination exists for this dataset, only (d) a saturated LDR panorama. Our method confers a wider variety of shadow characteristics (f) over that of [12].

PTGui commercial software. Since capturing the necessary 42 photos required approximately 3 minutes, care was taken to select scenes with no motion. We repeated this process over 9 different days to capture a diverse set of scenes, resulting in a mix of urban and natural scenes with illumination conditions ranging from overcast to sunny. We select 191 panoramas from this set (non-overlapping in both location and illumination conditions with the 15 used for training in the image-to-sky encoder, see sec. 5.3) and extract 7 crops per panorama for a total of 1,337 images, which we use for evaluation below.

6.2. Quantitative sun position evaluation

We begin by evaluating the performance of our models in estimating the relative position of the sun with respect to the camera from a single limited field of view image. Results on both the SUN360 test set from [12] (left) and our panorama dataset (right) are shown in fig. 5. In both cases, the ground truth is obtained by detecting the center of mass of the brightest region in the panorama, following [12] (who reported a median error of $4.59°$). Since our method only estimates explicitly the sun azimuth, the elevation angle is estimated as the brightest pixel of the reconstructed lighting panorama. Due to the more advanced network architecture employed, we systematically improve sun position estimation over the previous state-of-the-art on both datasets.

6.3. Lighting evaluation on HDR panoramas

The relighting error is compared between [12] and our method using the bunny scene on the 1,337 images from our dataset with ground truth illumination (sec. 6.1). Both the RMSE and the scale-invariant (si-)RMSE [10] are computed, and results are shown in fig. 6. Our technique yields significant improvement in both the RMSE and si-RMSE. For the RMSE, the improvement is mostly due to the fact that the exposure estimation of [12] that seems biased toward bright skies. The render intensity using our estimated lighting is generally much closer to the ground truth. Additionally, the increased versatility of our sky model confers an additional 60% improvement on scale-invariant RMSE.

Qualitative examples of recovered illumination and renders are shown for test images in our HDR panorama dataset in fig. 7 and SUN360 dataset in fig. 8. Both techniques provide plausible estimates yielding strong shadows on sunny days. For both datasets, we observe that lighting from [12] is consistently brighter than the ground truth, resulting in strong cast shadows even on partially cloudy and overcast skies. Our sky model captures the subtle lighting distribution in these conditions more accurately.

We further compare our method by performing virtual object insertions using the Cycles renderer. As fig. 9 shows, our estimated overall lighting intensity is closer to the ground truth than [12] while still providing plausible shading.

6.4. User study on SUN360 LDR panoramas

As no accurate ground truth illumination is available for SUN360, no quantitative relighting evaluation can faithfully be performed on this dataset. Instead, we evaluate performance with a user study where we showed ($N = 39$) participants 20 pairs of images with a virtual object (a bird statue model) composited into the image and lit by the estimates provided by our method and [12]. For each pair, users were asked to select the image where the object looked most realistic. As shown in tab. 2, our method obtained slightly more than 68% of the total votes. Furthermore, our lighting estimations were preferred (more than 50% votes) on 16 of the 20 images.

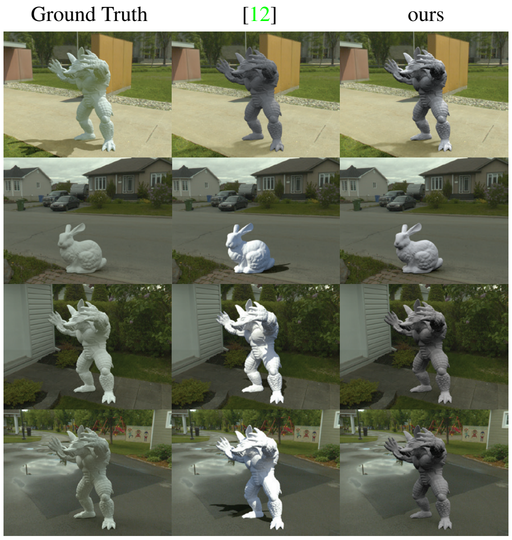


Figure 9. Examples of virtual object insertion comparing our method to [12] on backgrounds extracted from our evaluation dataset (see sec. 6.1).

7. User-guided edits on the sky model

Low-dimensional analytical sky models such as [14, 22] provide explicit parameters for users to interact with, such as sun position and atmospheric turbidity. The main drawback of our implicit sky parameters $\mathbf{z}$ is that they cannot be directly hand-tuned. One could think of generating new skies by interpolating between two sky parameters $\mathbf{z}$. However, doing so yields abrupt transitions that often contain undesirable artifacts such as multiple suns (fig. 10-a).

We propose a new method to browse the parameter space spanned by $\mathbf{z}$ while producing smooth transitions and plausible skies. Our intuition is to start from a known sky parameter $\mathbf{z}$ and iteratively morph this sky toward a desired target using the sky decoder gradient. To generate this gradient, a sky is first forwarded through the sky autoencoder. Then, edits are applied to the reconstructed sky and the resulting gradients on $\mathbf{z}$ are computed. We experiment on two types of edits: the sun elevation and intensity. To change the sun elevation, we move the 5×5 region around the sun either up or down and compute the gradient using the difference between the reconstruction and this modified sky. A similar scheme is used for sun intensity, where the region is multiplied such that its maximum value is the desired sun intensity. Iterating on this scheme using $\mathbf{z}_{n+1} = 4 \times 10^{-10} \cdot \frac{\partial \mathcal{L}_r}{\partial \mathbf{z}} \cdot \mathbf{z}_n$ for a maximum of 300 iterations automatically re-projects this modified sky back to a plausible sky and successfully removes the manually-induced artifacts. The multiplying factor was empirically set as a good balance between stability and convergence speed. Visually smooth transitions are shown in fig. 10.

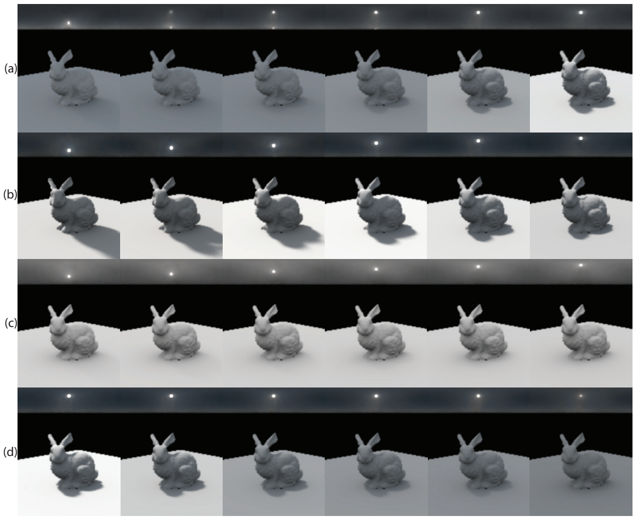


Figure 10. Examples of user-guided edits on the sky model. (a) Interpolating between two sky parameters $\mathbf{z}$ does not produce a smooth and plausible lighting transition. To solve this, we propose a method to enable smooth user-guided lighting edits and show results on changing the sun position on (b) a cloudy and (c) a clear day. Note how the generated skies stay plausible throughout the transition. We further show an example of changing the sun intensity (d), from fully visible to mostly occluded.

8. Discussion

In this paper, we propose what we believe is the first learned sky model trained end-to-end and show how to use it to estimate outdoor lighting from a single limited field of view images. Our key idea is to use three different datasets in synergy: SUN360 [28], Laval HDR sky database [20], and our own HDR 360° captures. Through quantitative and qualitative experiments, we show that our technique significantly outperforms the previous state-of-the-art on both lighting reconstruction and estimation.

While our method proposes state-of-the-art performance, it suffers from some limitations. Notably, the Hošek-Wilkie model employed by [12] tends to produce stronger lighting and sharper shadows than our model, which users seemed to prefer sometimes in our study. Additionally, while our model accurately captures the sky energy, its texture recovery quality is still limited. We hope these limitations can be soon lifted by the current rapid development of deep learning.

Acknowledgements

The authors wish to thank Sébastien Poitras for his help in collecting the HDR panorama dataset. This work was partially supported by the REPARTI Strategic Network and the NSERC Discovery Grant RGPIN-2014-05314. We gratefully acknowledge the support of Nvidia with the donation of the GPUs used for this research, as well as Adobe for generous gift funding.

References

[1] J. T. Barron and J. Malik. Shape, illumination, and reflectance from shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37(8):1670–1687, 2015. 2

[2] D. A. Calian, J.-F. Lalonde, P. Gotardo, T. Simon, I. Matthews, and K. Mitchell. From Faces to Outdoor Light Probes. *Computer Graphics Forum*, 37(2):51–61, 2018. 3, 4

[3] D. Cheng, J. Shi, Y. Chen, X. Deng, and X. Zhang. Learning scene illumination by pairwise photos from rear and front mobile cameras. In *Computer Graphics Forum*, volume 37, pages 213–221, 2018. 3

[4] D.-A. Clevert, T. Unterthiner, and S. Hochreiter. Fast and accurate deep network learning by exponential linear units (ELUs). In *International Conference on Learning Representations*, 2016. 4

[5] S. Cyberdyne. SkyNet: A global information grid/digital defense network, 1997. 2

[6] P. Debevec. Rendering synthetic objects into real scenes : Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *Proceedings of ACM SIGGRAPH*, 1998. 1

[7] S. Georgoulis, K. Rematas, T. Ritschel, M. Fritz, T. Tuytelaars, and L. Van Gool. What is around the camera? In *IEEE International Conference on Computer Vision*, 2017. 3

[8] S. Georgoulis, K. Rematas, T. Ritschel, E. Gavves, M. Fritz, L. Van Gool, and T. Tuytelaars. Reflectance and natural illumination from single-material specular objects using deep learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(8):1932–1947, 2018. 3

[9] R. Girdhar, D. F. Fouhey, M. Rodriguez, and A. Gupta. Learning a predictable and generative vector representation for objects. In *European Conference on Computer Vision*, 2016. 2

[10] R. Grosse, M. K. Johnson, E. H. Adelson, and W. T. Freeman. Ground truth dataset and baseline evaluations for intrinsic image algorithms. In *IEEE International Conference on Computer Vision*, 2009. 6, 7

[11] K. He, X. Zhang, S. Ren, and J. Sun. Deep Residual Learning for Image Recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 4

[12] Y. Hold-Geoffroy, K. Sunkavalli, S. Hadap, E. Gambaretto, and J.-F. Lalonde. Deep outdoor illumination estimation. In *IEEE International Conference on Computer Vision and Pattern Recognition*, 2017. 1, 2, 3, 4, 5, 6, 7, 8

[13] L. Hošek and A. Wilkie. An analytic model for full spectral sky-dome radiance. *ACM Transactions on Graphics*, 31(4):1–9, 2012. 1, 2

[14] L. Hošek and A. Wilkie. Adding a solar-radiance function to the hosek-wilkie skylight model. *IEEE Computer Graphics and Applications*, 33(3):44–52, may 2013. 1, 2, 8

[15] G. Huang, Z. Liu, K. Q. Weinberger, and L. van der Maaten. Densely Connected Convolutional Networks. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2016. 5

[16] K. Karsch, K. Sunkavalli, S. Hadap, N. Carr, H. Jin, R. Fonte, M. Sittig, and D. Forsyth. Automatic scene inference for 3d object compositing. *ACM Transactions on Graphics*, 33(3):32, 2014. 2

[17] J. T. Kider, D. Knowlton, J. Newlin, Y. K. Li, and D. P. Greenberg. A framework for the experimental comparison of solar and skydome illumination. *ACM Transactions on Graphics*, 33(6), nov 2014. 2

[18] D. Kingma and J. Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015. 5

[19] P. Krähenbühl and V. Koltun. Efficient inference in fully connected CRFs with gaussian edge potentials. In *Neural Information Processing Systems*, 2012. 3, 4

[20] J.-F. Lalonde, L.-P. Asselin, J. Becirovski, Y. Hold-Geoffroy, M. Garon, M.-A. Gardner, and J. Zhang. The Laval HDR sky database. `http://sky.hdrdb.com`, 2016. 2, 3, 4, 8

[21] J. F. Lalonde, A. A. Efros, and S. G. Narasimhan. Estimating the natural illumination conditions from a single outdoor image. *International Journal of Computer Vision*, 98(2):123–145, 2012. 2

[22] J.-F. Lalonde and I. Matthews. Lighting estimation in outdoor image collections. In *International Conference on 3D Vision*, 2014. 2, 4, 8

[23] S. Lombardi and K. Nishino. Reflectance and illumination recovery in the wild. *IEEE transactions on pattern analysis and machine intelligence*, 38(1), 2016. 2

[24] R. Perez, R. Seals, and J. Michalsky. All-weather model for sky luminance distribution—preliminary configuration and validation. *Solar Energy*, 50(3):235–245, Mar. 1993. 1, 2, 4

[25] A. J. Preetham, P. Shirley, and B. Smits. A practical analytic model for daylight. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques - SIGGRAPH*, 1999. 1, 2

[26] K. Rematas, T. Ritschel, M. Fritz, E. Gavves, and T. Tuytelaars. Deep reflectance maps. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4508–4516, 2016. 3

[27] J. Stumpfel, A. Jones, A. Wenger, C. Tchou, T. Hawkins, and P. Debevec. Direct hdr capture of the sun and sky. In *Proceedings of ACM AFRIGRAPH*, page 5, 2004. 6

[28] J. Xiao, K. A. Ehinger, A. Oliva, and A. Torralba. Recognizing scene viewpoint using panoramic place representation. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2012. 2, 3, 4, 8

[29] A. R. Zamir, A. Sax, W. Shen, L. Guibas, J. Malik, and S. Savarese. Taskonomy: Disentangling Task Transfer Learning. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018. 5

[30] J. Zhang and J.-F. Lalonde. Learning High Dynamic Range from Outdoor Panoramas. *International Conference on Computer Vision*, 2017. 2, 3, 4

[31] J. Zhang, K. Sunkavalli, Y. Hold-Geoffroy, S. Hadap, J. Eisenman, and J.-F. Lalonde. All-weather deep outdoor lighting estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 3