Sparse Needlets for Lighting Estimation with Spherical Transport Loss

Fangneng Zhan [1,2,4], Changgong Zhang [2], Wenbo Hu [3], Shijian Lu [1],
Feiying Ma [2], Xuansong Xie [2], Ling Shao [4]

[1] Nanyang Technological University   [2] DAMO Academy, Alibaba Group
[3] The Chinese University of Hong Kong   [4] Inception Institute of Artificial Intelligence

Abstract

Accurate lighting estimation is challenging yet critical to many computer vision and computer graphics tasks such as high-dynamic-range (HDR) relighting. Existing approaches model lighting in either frequency domain or spatial domain which is insufficient to represent the complex lighting conditions in scenes and tends to produce inaccurate estimation. This paper presents NeedleLight, a new lighting estimation model that represents illumination with needlets and allows lighting estimation in both frequency domain and spatial domain jointly. An optimal thresholding function is designed to achieve sparse needlets which trims redundant lighting parameters and demonstrates superior localization properties for illumination representation. In addition, a novel spherical transport loss is designed based on optimal transport theory which guides to regress lighting representation parameters with consideration of the spatial information. Furthermore, we propose a new metric that is concise yet effective by directly evaluating the estimated illumination maps rather than rendered images. Extensive experiments show that NeedleLight achieves superior lighting estimation consistently across multiple evaluation metrics as compared with state-of-the-art methods.

1. Introduction

Lighting estimation aims to recover illumination from a single image with limited field of view. It has a wide range of applications in various computer vision and computer graphics tasks such as high-dynamic-range (HDR) relighting in mixed reality, etc. However, lighting estimation is a challenging and ill-posed problem as it needs to predict the illumination coming from a full sphere of directions including those unobserved from the current view in the scene. Additionally, it often requires to infer HDR illuminations from low-dynamic-range (LDR) observations so as to light virtual objects realistically while inserting them into real scene images as illustrated in Fig. 1.

Lighting estimation has been tackled by regressing the parameters of various lighting representations in either frequency domain [7, 15]) or spatial domain [14, 13, 34, 35]. However, lighting estimation in frequency domain normally represents illumination with Spherical Harmonics (SH) which lack spatial localization capabilities. Thus it tends to capture global lighting instead of the exact spatial locations of the light sources which often leads to weak shading and shadow effects as illustrated in Garon et al. [15] of Fig. 1. Lighting estimation in spatial domain has been addressed by direct generation of the illumination maps or indirect reconstruction through spherical Gaussian function. However, direct generation of illumination maps often leads to worse generalization as lighting estimation is an under-constrained problem by itself, and spherical Gaussian often involves a complicated training process as described in [13]. Both types of approaches in spatial domain do not explicitly consider lighting frequency, and thus lead to inaccurate relighting performance as illustrated in Gardner et al. [13] of Fig. 1. The high frequency information also tends to be blurred due to the use of naive L2 loss in the training. Additionally, existing evaluation metrics in lighting estimation usually assess the objects rendered with the predicted illumination maps, which is time-consuming and sensitive to the test setting.

In this work, we propose NeedleLight, a new model that introduces needlet for accurate and robust lighting estimation from a single image. As a new generation of spherical wavelets, needlet enjoys good localization properties in both frequency and spatial domain which makes it ideal to be the basis for illumination representation. Moreover, to remove the redundant parameters in needlet coefficients which will disturb the regression of principle light sources, we design an optimal thresholding function to achieve sparse needlets which improve the lighting estimation greatly.

Unlike spherical harmonic coefficients, needlet coefficients are spatially localized over a unit sphere. To utilize the spatial information in regression, we propose a Spherical Transport Loss (STL) based on optimal transport theory. STL is able to capture spatial information via a cost matrix

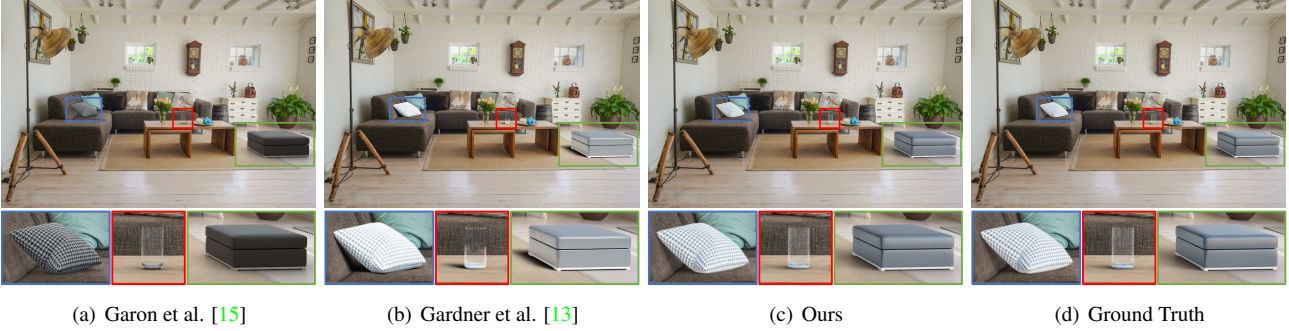| (a) Garon et al. [15] | (b) Gardner et al. [13] | (c) Ours | (d) Ground Truth |

Figure 1. The proposed NeedleLight estimates a parametric lighting representation from a single scene image which is critical to many tasks such as virtual object insertion. Unlike previous methods that predict lighting in either frequency domain [15] (losing spatial localization) or spatial domain [13] (losing frequency information) only, we introduce a novel needlets basis which is capable of representing and estimating lighting accurately in both frequency and spatial domains.

and estimate the needlet coefficients more accurately than a naive L2 loss. Besides, STL employs auxiliary point strategy to preserve high frequency information and greatly reduce the dimension of required parameters. Based on STL, we design a new metric for the evaluation of lighting estimation by measuring the discrepancy between illumination maps. The new metric highly simplifies the evaluation procedure and provides concise yet effective evaluation with regard to the lighting color, intensity and position.

The contribution of this work can be summarized in three aspects. First, we introduce a novel needlet basis for illumination representation which allows to regress the parameters in both frequency and spatial domains simultaneously. Second, we develop an optimal thresholding function to achieve sparse needlets which effectively removes the redundant needlet coefficients and improves the lighting estimation. Third, we design a novel Spherical Transport Loss (STL) that effectively utilizes the spatial information of needlet coefficients in regression. With STL, we also design a new evaluation metric that is more concise and effective than existing evaluation metrics.

2. Related Works

Lighting Estimation: Lighting estimation is a classic challenge in computer vision and computer graphics, and it is critical for realistic relighting in virtual objects insertion [11, 41, 18, 54, 43, 2, 16, 47, 29, 42, 5, 49] and image synthesis [45, 40, 44, 50, 38, 46, 51, 39, 48]. Traditional approaches require user intervention or assumptions about the underlying illumination model, scene geometry, etc. For example, Karsh et al. [17] recovers parametric 3D lighting from a single image but requires user annotations for initial lighting and geometry estimates. Zhang et al. [52] requires a full multi-view 3D reconstruction of scenes. Lombardi et al. [24] estimates illumination from an object of known shape with a low-dimensional model.

The recent works estimate lighting by regressing representation parameters or generating illumination maps [21, 37, 16, 53, 23, 6, 31]. For example, Cheng et al. [7] regresses the SH parameters of global lighting with a render loss. Maier et al. [25] recover spherical harmonics illumination with additional depth information. Garon et al. [15] estimate lighting by predicting SH coefficients from a background image and local patch. Gardner et al. [13] estimate the positions, intensities, and colours of light sources and reconstructs illumination maps with a spherical Gaussian function. Li et al. [20] represent illumination maps with multiple spherical Gaussian functions and regresses the corresponding Gaussian parameters for lighting estimation. Gardner et al. [14] generate illumination maps directly with a two-steps training strategy. Legendre et al. [19] regress HDR lighting from LDR images by comparing the ground-truth sphere image to the rendered one with the predicted illumination. Srinivasan et al. [35] estimate a 3D volumetric RGB model of a scene and uses standard volume rendering to estimate incident illuminations. Zhan et al. [?] propose to formulate the regression of illumination as the regression of a spherical distribution.

Needlets: Needlets are a new generation of spherical wavelets [33] and have desirable localization capabilities in both spatial and frequency domains. Narcowich et al. [30] first introduces needlets in the Functional Analysis literature, and Baldi et al. [1] further analyses the statistical properties of needlets. Due to the good localization property in the multipole and pixel spaces, needlets have been widely applied to the research of Cosmic Microwave Background [32, 27].

The aforementioned works estimate lighting in either frequency domain or spatial domain which is insufficient to capture the complex illumination in real scenes. Additionally, most existing works regress the illumination with a naive L2 loss or its variant which struggles to regress high frequency information and often introduces blurs. We intro-

duce needlet basis for lighting representation which allows regression of illumination in both frequency and spatial domains jointly. A novel spherical transport loss is proposed to achieve stable and accurate regression of needlet coefficients. Details on lighting representation, spherical transport loss are to be presented in the ensuing sections.

3. Proposed Method

The proposed NeedleLight estimates illumination by regressing needlet coefficients from a single image as illustrated in Fig. 2. A novel spherical transport loss is designed to achieve stable and effective regression of needlet coefficients. The following subsections describe how needlets and spherical transport loss work together for accurate and robust lighting estimation.

3.1. Needlets based Lighting Representation

Needlets [1, 30] are a new type of spherical wavelets that has been successfully used in microwave signal analysis. They can be localized at a finite number of frequencies, and decay quasi-exponentially fast away from the global maximum. Thus they enjoy good localization properties in both frequency and spatial domains. As described in [1], a signal $I(x)$ (e.g. lighting signals of interest in this research) at a given frequency $j \in \mathbb{N}$ can be represented by the spherical needlets basis $\psi_{jk}(x)$ and the needlet coefficients β_{jk} as follows:

$$
\psi_{jk}(x) = \sqrt{\lambda_{jk}} \sum_{l=\lceil B^{j-1} \rceil}^{\lfloor B^{j+1} \rfloor} b(\frac{l}{B^j}) \sum_{m=-l}^{l} Y_{lm}(\xi_{jk}) \overline{Y}_{lm}(x)
$$
$$
\beta_{jk} = \sqrt{\lambda_{jk}} \sum_{l=0}^{\infty} b(\frac{l}{B^j}) \sum_{m=-l}^{l} a_{lm} Y_{lm}(\xi_{jk})
$$

(1)

where $x \in \mathbb{S}^2$, ξ_{jk} and λ_{jk} are pre-defined cubature points that spread over the unit sphere as shown in Fig. 3, and the associated cubature weights, respectively, $b(\cdot)$ is a window function, B is a free parameter larger than 1, Y_{lm} is spherical harmonic function with degree l and order m, a_{lm} is the corresponding spherical harmonic coefficients, $\overline{Y}_{lm}$ is the complex conjugation of Y_{lm}. ξ_{jk} represents the spatial location of needlet basis ψ_{jk}, thus the needlet coefficients are spatially localized on the unit sphere. The signal $I(x)$ can be reconstructed via $I(x) = \sum_{j,k} \beta_{jk} \psi_{jk}(x)$.

Compared with SH, needlets have compact supports for the localization in spatial domain. As a result, they can easily and parsimoniously represent signals over the unit sphere that exhibits local sharp peaks or valleys, which are commonly presented in HDR illumination maps. Consequently, needlets serve as a more suitable basis for the representation of illumination maps.

3.2. Sparse Needlets

A signal is said to be sparse if it can be reconstructed with only a small amount of basis functions. An illumination map consists of several dominant light sources with high radiance energy and an ambient residual. The light sources are obviously sparse under the needlet basis while the ambient component is not. For the lighting estimation, the reconstruction of dominant light sources is more significant than the reconstruction of the ambient component especially in high frequency sections. Besides, there are dozens of needlet coefficients in high frequency section (252 coefficients for j_{max}=3), most of which are redundant parameters of ambient. Those redundant parameters will severely disturb the regression of principal light sources and lead to diffuse illumination (or low frequency illumination) which is undesired for relighting. Thus we deduce an optimal sparse function for needlets to separate the principle coefficients of light sources from the redundant ambient component.

We derive the sparse function from a Bayesian perspective and form the problem as a maximum posterior estimator. Normally, we assume that the needlet coefficients of light sources s follows Laplace distribution prior as Laplace prior is well adapted to model sparse signals [4], which is also proved in [12]. The needlet coefficients of ambient can be ideally treated as a Gaussian distribution [26]. The needlet coefficients β of the illumination map can thus be modelled as:

$$
\beta = \underbrace{s}_{light\ sources} + \underbrace{\phi}_{ambient} + \underbrace{\eta}_{noise}
$$

(2)

where s denotes the needlet coefficients of sparse light sources which follow Laplace distribution, ϕ denotes the needlet coefficients of ambient which is a Gaussian distribution, and η denotes noises that follow a Gaussian distribution. According to Vansynge et al. [36], the Bayesian formulation of the problem can be written as:

$$
P(s|\beta) \propto P(s)\mathcal{L}(\beta|s)
$$
$$
\mathcal{L}(\beta|s, \phi) = N(\beta; \phi + s, M_\eta)
$$

(3)

where M_η denotes the covariance matrices of the noise. As ϕ follows Gaussian distribution, we can derive:

$$
\mathcal{L}(\beta|s) = \int ... \int \mathcal{L}(\beta|s, \phi) P(\phi) d\phi \propto
$$
$$
\exp\left[-\frac{1}{2}\beta^T M_\eta^{-1}\beta + s^T M_\eta^{-1}\beta - \frac{1}{2}s^T M_\eta^{-1}s \right] \cdot \exp
$$
$$
\left[\frac{1}{2}(M_\eta^{-1}\beta - M_\eta^{-1}s)^T (M_\eta^{-1} + M_\phi^{-1})(M_\eta^{-1}\beta - M_\eta^{-1}s) \right]
$$

(4)

where M_ϕ denotes the covariance matrices of the Gaussian distribution ϕ. As $P(s)$ follows the Laplace distribution namely $P(s) \propto \exp[-\lambda ||s||]$, the maximum posterior estimator is obtained by maximizing $P(s|\beta) = \mathcal{L}(\beta|s) * P(s)$.

Needlets Coefficients

FC

$j=1$

Cubature Points for $j=k$

$j=k$

$j=n$

Spherical Transport Loss

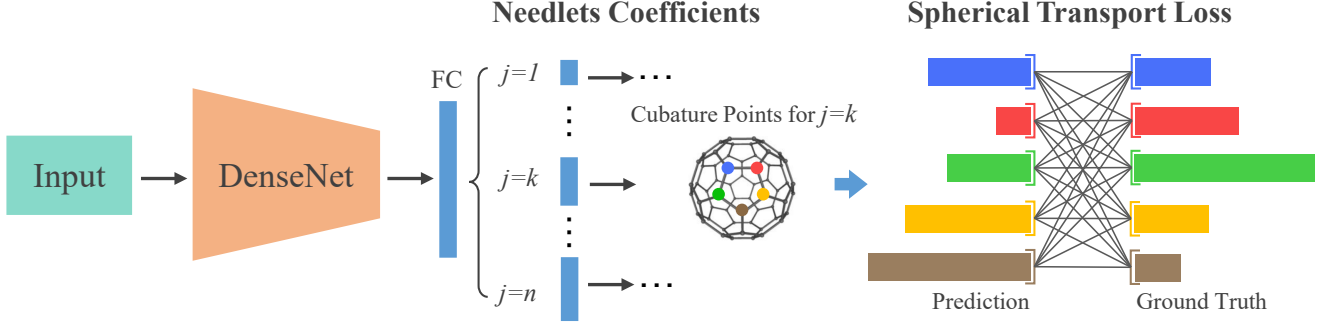Input → DenseNet →

Prediction   Ground Truth

Figure 2. The architecture of the proposed NeedleLight: We introduce sparse needlets to represent lighting. The frequency bands in *Needlet Coefficients* are denoted as $j = 1, \cdots, n$. The needlet coefficients in each frequency band are spatially localized on a set of *Cubature Points* on the unit sphere (illustration for $j = k$ only). For stable lighting regression, we design a spherical transport loss to capture the divergence between the predicted needlet coefficients (*Prediction*) and the ground-truth coefficients (*Ground Truth*). The values of needlet coefficients are illustrated by color bars (only draw 5 for illustration).
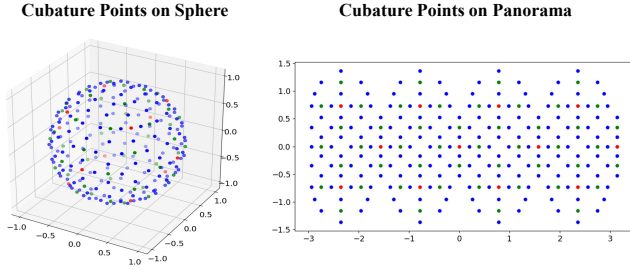

Cubature Points on Sphere   **Cubature Points on Panorama**

Figure 3. The visualization of cubature points (j_{max}=3) on sphere and panorama. The red, green and blue points denote the cubature points of frequency bands from low to high frequency, namely j=1, j=2 and j=3 respectively.

We take the partial derivative with respect to s:

$$\partial_s(-log(P(s|\beta))) = -(M_\eta + M_\eta M_\phi^{-1} M_\eta)^{-1}s + M_\eta^{-1}s \\ + (M_\eta + M_\eta M_\phi^{-1} M_\eta)^{-1}\beta - M_\eta^{-1}\beta + \lambda\partial_s||s|| \quad (5)$$

By making $\partial_s(-log(P(s|\beta))) = 0$, the following solution (or sparse function) can thus be derived:

$$s = \beta - \left[M_\eta^{-1} - (M_\eta + M_\eta M_\phi^{-1} M_\eta)^{-1}\right]^{-1} \lambda\partial_s||s|| \quad (6)$$

which is a soft thresholding operator with threshold $(M_\eta^{-1} - (M_\eta + M_\eta M_\phi^{-1} M_\eta)^{-1})\lambda$. More details about the derivation of the sparse function are provided in the supplemental file.

We apply the sparse function largely to the *high-frequency needlet coefficients* so as to shrink the redundant coefficients of the ambient component. The sparse needlets provide sparse representation of illumination maps, which is desirable for the regression of light sources by using the proposed spherical transport loss to be described in the following subsection.

3.3. Spherical Transport Loss

Different from spherical harmonics, needlet coefficients are spatially localized in a unit sphere. A simple MSE loss cannot utilize the spatial information. Besides, as an under-constrained task, the optimization in lighting estimation is severely challenging since it aims to recover the environment lighting from all directions based on a single scene image with a limited field of view. Thus the training of lighting estimation models struggles with the regression of high frequency information.

In this work, we propose a novel Spherical Transport Loss (STL) to achieve the stable and effective lighting regression. Because the needlet coefficients are spatially localized on the unit sphere as specified by cubature points, we treat the regression of needlet coefficients as an Unbalanced Optimal Transport (UOT) [22, 8] problem on the unit sphere. Intuitively, UOT computes the cost for transporting a measure distributed on a space to another measure of possibly different total masses. As the needlet coefficients may contains negative values, we take the natural exponent of them before deriving the spherical transport loss. Then we can define two sets of needlet coefficients represented by two positive vectors $a = (a_1, \cdot, a_n) \in \mathbb{R}^n_+$ and $b = (b_1, \cdots, b_n) \in \mathbb{R}^n_+$ and their spatial layouts are specified by cubature points on the unit sphere. A distance matrix C and a transportation plan matrix P can thus be derived, where each entry C_{ij} in C gives the cost of moving point a_i to point b_j which can be defined by radian distance between points on the unit sphere, and P_{ij} in P represents the probability of assigning a point a_i to a point b_j. The regularized UOT problem (namely STL) can thus be defined as follows:

$$\min_P \left[\langle C, P \rangle + \tau KL(P \cdot \vec{1}||a) + \tau KL(\vec{1} \cdot P||b) - \gamma H(P)\right] \quad (7)$$

where τ and γ are regularization parameters, KL is the KerKullback-Leibler Divergence, $H(P)$ is the entropic regularization for efficient approximation of original UOT
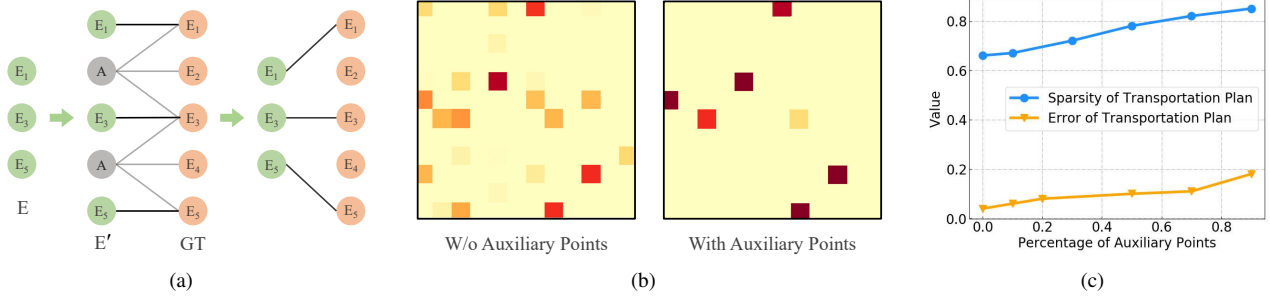
Figure 4. (a) The process to achieve sparse transportation plan through auxiliary points. E and GT are the estimated needlet coefficients and the ground truth, respectively. 'A' denotes the auxiliary point. E′ denotes the predicted coefficients extended with auxiliary points. The connection between E′ and GT signifies the resulting transportation plan. (b) Transportation plans for the needlet coefficients (j=1) w/o and w/ auxiliary points, respectively. (c) The sparsity and error of transportation plan matrix with different percentage of auxiliary points.

problem [8]. The regularized UOT can be solved by Sinkhorn iteration [10].

To construct clear and sharp instead of diffuse (i.e., low frequency) light sources, we expect the transportation plan to be sparse. Otherwise the diffuse light sources will result in weak shading and shadow effects as presented in Garon et al. [15] of Fig. 6. Besides, the dimension of output layer increases quickly when the frequency of needlets becomes large (252 needlet coefficients when $j_{max} = 3$). We propose an auxiliary point strategy to achieve sparse transportation plan and reduce the dimension of output layer. Auxiliary points are assigned with small value (we select $\frac{1}{(number\ of\ coefficients)}$) and 0 cost for transport, which can be used for absorbing unused probability mass in cases of partial transport. As shown in Fig. 4(a), we only estimate partial needlet coefficients (E), and then use the auxiliary points to replace other needlet coefficients to obtain a new set of coefficients (E′). After obtaining the optimal transport between E′ and the ground truth GT, we can extract a sparse transportation plan by removing the connection with the auxiliary points. The two samples in Fig. 4(b) show the transportation plan matrix for the needlet coefficients ($j = 1$) with and without auxiliary points, respectively. Fig. 4(c) shows the sparsity and error of the transportation plan matrix ($n \times n$) with different percentage of auxiliary points (The evaluation metrics for the sparsity and error of transportation plan are described in [3]). When more auxiliary points are used, the sparsity of the transportation matrix increases but the error also increases as shown in Fig. 4 (c). As a trade-off, we select 66% of coefficients as auxiliary points.

Using spherical transport loss for needlet coefficients regression has two clear advantages. First, it makes the regression sensitive to the global geometry, thus effectively penalizing predicted activation that is far away from the ground truth distribution. Second, it can preserve the high frequency information during training with the proposed auxiliary points.

Evaluation Metric: Lighting estimation has been widely evaluated by using root mean square error (RMSE) and scale-invariant RMSE (si-RMSE) that measure the standard deviation of residuals of the rendered images. RMSE mainly evaluates the estimated lighting intensity, and si-RMSE focuses more on the evaluation of lighting positions. In addition, lighting estimation has also been evaluated by using Amazon Mechanical Turk (AMT) that performs crowdsourcing user study for subjective assessment of empirical realism of rendered images.

Existing metrics evaluate estimated lighting largely by applying them to the rendered objects. Thus the performance of an estimation model is highly affected by test settings such as materials and 3D shape of the rendered objects. Based on the proposed spherical transport loss, we design a spherical transport distance (STD) metric that directly evaluates the optimal transport distance between the predicted illumination map and the ground-truth map on the unit sphere. The only difference from STL is that STD discards the auxiliary point strategy. The proposed STD highly simplifies the evaluation procedure and provides concise yet effective evaluations regarding to the lighting color, lighting intensity, and lighting position jointly.

Figure 5. The scene for quantitative evaluations consists of three spheres with diffuse gray, matte silver and mirror silver materials.

Table 1. Comparison of NeedleLight with several state-of-the-art lighting estimation methods: The evaluation metrics include the widely used RMSE, si-RMSE, AMT, and our proposed STD. D, S, M denote a diffuse, a matte silver and a mirror material of the rendered objects, respectively.

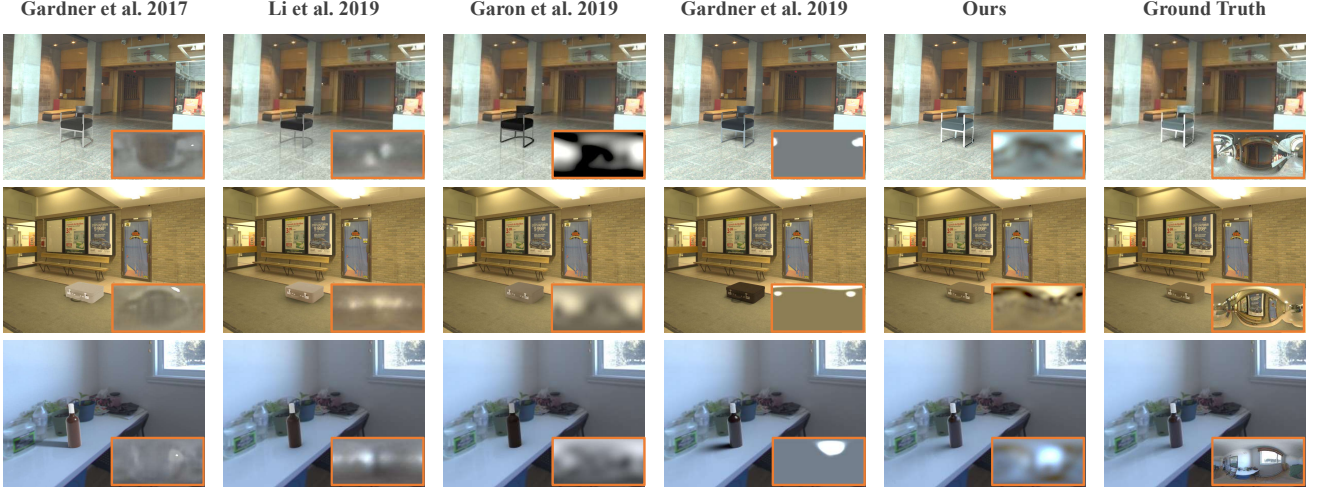| Metrics	Gardner et al. [14]			Gardner et al. [13]			Li et al. [20]			Garon et al. [15]			NeedleLight		
	D	S	M	D	S	M	D	S	M	D	S	M	D	S	M
RMSE	0.13	0.16	0.18	**0.06**	0.10	0.15	0.21	0.23	0.26	0.18	0.20	0.24	0.07	**0.07**	**0.09**
si-RMSE	0.15	0.15	0.17	0.07	0.09	0.12	0.19	0.21	0.23	0.21	0.24	0.26	**0.05**	**0.06**	**0.08**
AMT	28%	23%	21%	34%	33%	30%	28%	27%	23%	29%	26%	24%	**41%**	**39%**	**36%**
STD	6.84			5.52			7.01			7.14			**4.21**		

Figure 6. Visual comparison of NeedleLight with state-of-the-art lighting estimation methods: With the illumination maps predicted by different methods (at bottom-right corner of each rendered image), the rendered virtual objects demonstrate different lighting intensity, color, shadow and shade.

4. Experiments

4.1. Dataset and Implementation

We evaluate NeedleLight by using the Laval Indoor HDR Dataset [14] that consists of 2,100 HDR panoramas taken in a variety of indoor environments. Similar to [14], we extract eight limited field of view crops from each panorama which produces 19,556 images as used in our experiments. The image warping operation as described in [14] is applied to the panoramas. We apply the proposed sparse needlets with $j_{max} = 3$ to extract needlet coefficients as the ground truth for training. Similar to [13, 15], DenseNet121 is used as the backbone network and produces a 4096-dimensional latent vector which is further forwarded to a fully-connected layer with 1024 units. Three separate output layers are added to regress needlet coefficients in frequency bands $j = 1, 2, 3$. In the experiment, we randomly select 200 images as testing set and the rest for training set. All the objects are rendered with Blender [9].

The proposed NeedleLight is implemented by the PyTorch framework. The Adam is adopted as optimizer which employs a learning rate decay mechanism (initial learning rate is 0.001). The network is trained in 100 epochs with a batch size of 64. In addition, the network training is performed on two NVIDIA Tesla P100 GPUs with 16GB memory.

4.2. Quantitative Evaluation

We compare NeedleLight with a number of state-of-the-art lighting estimation methods including Garon et al. [15] that estimates lighting in frequency domain and Gardner et al. [14, 13] and Li et al. [20] that estimate lighting in spatial domain. To perform quantitative evaluations, we create three spheres with gray diffuse, matte silver and mirror silver materials for evaluation as shown in Fig. 5, which is consistent with the evaluation setting in [19]. Then we render 300 images of objects (100 images for each material) by using the illumination maps that are predicted from testing set by each compared method. Table 1 shows experimental results by using 4 evaluation metrics as described in *Evaluation Metrics*, where *D*, *S* and *M* denote a diffuse, a matte silver and a mirror material of the objects to be rendered, respectively. The AMT user study is conducted by showing two images rendered by the ground truth and one of

the methods in Table 1 to 20 users who will pick the more realistic image. The score is the percentage of images rendered by the method that is deemed as more realistic than the ground-truth rendering.

We can observe that NeedleLight outperforms other methods in most cases under different evaluation metrics and materials as it allows regression in frequency and spatial domain jointly. The only exception is for diffuse material by [13] while evaluated using RMSE, largely because the parameterization in Gardner et al. [13] simplifies the scene illumination in spatial domain to achieve accurate prediction of light intensity while diffuse material is largely affected by light intensity and RMSE is most sensitive to light intensity. Gardner et al. [14] predicts illumination maps directly by a two-stage training strategy. As an under-constrained problem, the direct generation methods like Gardner et al. [14] tend to over-fit the training set and present worse generalization performance. Besides, both Gardner et al. [14] and Gardner et al. [13] estimate lighting in spatial domain which cannot recover frequency information and tends to generate inaccurate shading and shadow that are largely measured by si-RMSE. Garon et al. [15] recovers lighting in frequency domain by regressing the SH coefficients, which tend to capture global instead of localized lighting. Thus Garon et al. [15] struggles to regress accurate lighting position and recover high frequency information. Li et al. [20] adopts spherical Gaussian functions to reconstruct illumination in spatial domain, thus it cannot recover accurate illumination frequency. Besides, it uses a masked L2 loss to preserve high frequency information though it cannot solve the missing of high frequency information essentially as illustrated in Fig. 6. Instead, our proposed spherical transport loss with auxiliary points improves the regression of high frequency information significantly.

In addition, we can observe that the performance of the state-of-the-art methods is not consistent under different evaluation metrics. For example, Li et al. [20] outperforms Garon et al. [15] in si-RMSE but the situation becomes the other way around in RMSE. The divergence of different metrics makes it hard to provide consistent evaluations. The proposed spherical transport distance (STD) instead provides relatively consistent and comprehensive evaluations in light intensity, color, position, etc. as shown in Table 1.

4.3. Qualitative Evaluation

We also compare NeedleLight with four state-of-the-art lighting estimation methods qualitatively. we design 25 3D scenes with objects for insertion and render them with the predicted illumination maps. Fig. 6 shows several rendered images and the predicted illumination maps. We can observe that NeedleLight predicts realistic illumination maps with plausible light sources, thus producing realistic rendering with clear and accurate shade and shadows that are

Figure 7. Object relighting on a variety of photos from the Internet. In all cases, light estimation is performed completely automatically by our model. The predicted illumination maps are utilized to relight the virtual objects with Blender [9]. More rendering samples and the analysis of **spatially-varying** rendering are available in supplementary file.

very close to the ground truth. As a comparison, the illumination maps predicted by Gardner et al. [14] and Gardner et al. [13] tend to present clearly higher illumination frequency than the ground truth, largely because they recover illumination in spatial domain without explicitly taking frequency information into account. Garon et al. [15] and Li et al. [20] only predict illumination of low frequency, thus the produced renderings present very weak shade and shadow effects as illustrated in Fig. 6.

Besides the testing set, we also validate the proposed method on natural images collected from the Internet as shown in Fig. 7. The proposed method achieves accurate estimation of scene illumination, thus the 3D objects can be embedded into the images with real shading and shadow effects. We also include the analysis of **spatially-varying** illumination prediction in the supplementary file.

4.4. Ablation Study

We further evaluate NeedleLight by developing four NeedleLight variants as listed in Table 2, including a baseline model which regress spherical harmonic coefficients with L2 loss (*SH+L2*), regressing original needlet coefficients with L2 loss (*SN+L2*), regressing needlet coefficients

Table 2. Ablation study of the proposed NeedleLight: *SH* denotes using spherical harmonics for representation; *SN* denotes using spherical needlets for representation; *ST* denotes applying the derived soft thresholding function to needlet coefficients; *L2* and *STL* denotes regressing coefficients with L2 loss and spherical transport loss; *SN+ST+L2+STL* denotes standard NeedleLight with sparse needlets which are regressed with both L2 and spherical transport loss.

Models	RMSE			si-RMSE			AMT			STD
	D	S	M	D	S	M	D	S	M	-
SH+L2	0.19	0.20	0.25	0.22	0.24	0.28	23%	21%	18%	7.14
SN + L2	0.14	0.14	0.16	0.11	0.14	0.15	33%	31%	28%	4.93
SN+ST+L2	0.11	0.12	0.14	0.09	0.09	0.13	34%	32%	30%	4.74
SN+ST+STL	0.10	0.11	0.13	0.07	0.10	0.12	37%	34%	32%	4.51
SN+ST+L2+STL	**0.07**	**0.07**	**0.09**	**0.05**	**0.06**	**0.08**	**41%**	**39%**	**36%**	**4.21**

Table 3. Ablation studies over different bases including spherical harmonics (SH), spherical gaussian (SG), spherical distribution (SD), Haar, spherical needlets (SN). HT and ST denote apply hard thresholding and the derived soft thresholding functions to the spherical needlet coefficients. j_{max} denotes the order of needlets for representation. SN(j_{max}=3)+ST is the standard setting of NeedleLight.

Models	L2	L2 + STL
SH	7.14	7.13
SG	6.01	5.74
SD	5.75	5.54
Haar	6.07	5.83
SN (j_{max}=1)	7.52	7.18
SN (j_{max}=2)	6.71	6.20
SN (j_{max}=4)	5.37	4.96
SN(j_{max}=3)	5.32	4.93
SN(j_{max}=3)+**HT**	4.92	4.51
SN(j_{max}=3)+**ST**	**4.74**	**4.21**

after applying soft thresholding (namely sparse needlets) with L2 loss (*SN+ST+L2*), regressing sparse needlets coefficients with spherical transport loss (*SN+ST+STL*). The standard NeedleLight regresses spares needlets with both L2 and spherical transport loss (*SN+ST+L2+STL*). Similar to the setting in Quantitative Evaluation, we apply the five variants and the standard NeedleLight to render 300 images with 15 objects of different materials. As Table 2 shows, using needlets for illumination representation (SN+L2) helps to achieve better lighting estimation compared with spherical harmonics function (SH+L2). In addition, the performance of *SN+ST+L2* is improved clearly as sparse needlet coefficients helps the regression of light sources significantly by the zero setting of redundant coefficients. Additionally, the standard *SN+ST+L2+STL* outperforms *SN+ST+L2* and *SN+ST+STL*, demonstrating the L2 and the proposed STL are complementary for the regression of needlet coefficients.

We also study effect of thresholding function and differ-

ent orders of needlets and compare spherical needlets (SN) with other basis functions for illumination representation such as spherical harmonics (SH), spherical gaussian (SG), spherical distribution (SD) [**?**] and Haar [28] as shown in Table 3. We followed the experimental setting as in Table 2 and measure the proposed STD as the evaluation metric. The number of coefficients of SH, SG, SD and Haar are set to be consistent with spherical needlets with 3 orders (about 250 coefficients). As Table 3 show, needlets SN(j_{max}=3) outperforms other representation bases when regressing with both L2 loss and spherical transport loss. Compared with spherical gaussian, spherical distribution and spherical harmonics functions, needlets enable representation in both spatial and frequency domains, thus achieving accurate regression in both domains. Compared with other wavelets such as Haar, needlets is a new generation of spherical wavelets and is more suitable for the representation of spherical image with sharp local peak valleys which are commonly presented in HDR illumination map. Regressing spherical harmonic coefficients with spherical transport loss doesn't improve the performance as there is no spatial information in spherical harmonic coefficients. For spherical gaussian, spherical distribution and Haar, the performance with the spherical transport loss included is clearly improved as the corresponding coefficients are spatially localized, which demonstrates the effectiveness of the proposed spherical transport loss. Further, the prediction performance drops when j_{max}=1,2 are used, and increasing the order to j_{max}=4 doesn't improve the performance. We conjecture that the larger number of parameters with j_{max}=4 affects the regression accuracy negatively. Besides, the performance of SN(j_{max}=3)+ST outperforms SN(j_{max}=3)+HT, demonstrating the superiority of the derived soft thresholding function.

5. Conclusions

This paper presents NeedleLight, a lighting estimation model that introduces needlet basis for illumination representation and prediction. In NeedleLight, we deduce an op-

timal thresholding function from Bayesian framework for a sparse representation in terms of needlet basis. To tackle the regression of needlet coefficients with spatial localization, a novel spherical transport loss with auxiliary points is designed which performs regression by minimizing the discrepancy between two spherical distributions. Both the quantitative and qualitative experiments show that Needle-Light is capable of predicting illumination accurately from a single indoor image. We will continue to investigate needlet basis for more efficient illumination representation and explore optimal transport for better network training.

References

[1] Paolo Baldi, Gérard Kerkyacharian, Domenico Marinucci, Dominique Picard, et al. Asymptotics for spherical needlets. *The Annals of Statistics*, 37(3):1150–1171, 2009. 2, 3

[2] Jonathan T Barron and Jitendra Malik. Intrinsic scene properties from a single rgb-d image. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 17–24, 2013. 2

[3] Mathieu Blondel, Vivien Seguy, and Antoine Rolet. Smooth and sparse optimal transport. In *International Conference on Artificial Intelligence and Statistics*, pages 880–889. PMLR, 2018. 5

[4] Jérome Bobin, Yassir Moudden, J-L Starck, J Fadili, and Nabila Aghanim. Sz and cmb reconstruction using generalized morphological component analysis. *Statistical Methodology*, 5(4):307–317, 2008. 3

[5] Mark Boss, Varun Jampani, Kihwan Kim, Hendrik Lensch, and Jan Kautz. Two-shot spatially-varying brdf and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3982–3991, 2020. 2

[6] Andrew Chalmers, Junhong Zhao, Daniel Medeiros, and Taehyun Rhee. Reconstructing reflection maps using a stacked-cnn for mixed reality rendering. *IEEE Transactions on Visualization and Computer Graphics*, 2020. 2

[7] Dachuan Cheng, Jian Shi, Yanyun Chen, Xiaoming Deng, and Xiaopeng Zhang. Learning scene illumination by pairwise photos from rear and front mobile cameras. In *Computer Graphics Forum*, pages 213–221. Wiley Online Library, 2018. 1, 2

[8] Lenaic Chizat, Gabriel Peyré, Bernhard Schmitzer, and François-Xavier Vialard. Scaling algorithms for unbalanced transport problems. *arXiv preprint arXiv:1607.05816*, 2016. 4, 5

[9] Blender Online Community. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam, 2018. 6, 7

[10] Marco Cuturi. Sinkhorn distances: Lightspeed computation of optimal transport. In *Advances in neural information processing systems*, pages 2292–2300, 2013. 5

[11] Paul Debevec. Rendering synthetic objects into real scenes: Bridging traditional and image-based graphics with global illumination and high dynamic range photography. In *ACM SIGGRAPH 2008 classes*, pages 1–10, 2008. 2

[12] Ron O Dror, Alan S Willsky, and Edward H Adelson. Statistical characterization of real-world illumination. *Journal of Vision*, 4(9):11–11, 2004. 3

[13] Marc-André Gardner, Yannick Hold-Geoffroy, Kalyan Sunkavalli, Christian Gagné, and Jean-François Lalonde. Deep parametric indoor lighting estimation. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 7175–7183, 2019. 1, 2, 6, 7

[14] 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. *arXiv preprint arXiv:1704.00090*, 2017. 1, 2, 6, 7

[15] Mathieu Garon, Kalyan Sunkavalli, Sunil Hadap, Nathan Carr, and Jean-François Lalonde. Fast spatially-varying indoor lighting estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6908–6917, 2019. 1, 2, 5, 6, 7

[16] 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

[17] Kevin Karsch, Varsha Hedau, David Forsyth, and Derek Hoiem. Rendering synthetic objects into legacy photographs. *ACM Transactions on Graphics (TOG)*, 30(6):1–12, 2011. 2

[18] Jean-François Lalonde, Alexei A Efros, and Srinivasa G Narasimhan. Estimating the natural illumination conditions from a single outdoor image. *International Journal of Computer Vision*, 98(2):123–145, 2012. 2

[19] Chloe LeGendre, Wan-Chun Ma, Graham Fyffe, John Flynn, Laurent Charbonnel, Jay Busch, and Paul Debevec. Deeplight: Learning illumination for unconstrained mobile mixed reality. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5918–5928, 2019. 2, 6

[20] Mengtian Li, Jie Guo, Xiufen Cui, Rui Pan, Yanwen Guo, Chenchen Wang, Piaopiao Yu, and Fei Pan. Deep spherical gaussian illumination estimation for indoor scene. In *Proceedings of the ACM Multimedia Asia*, pages 1–6, 2019. 2, 6, 7

[21] Zhengqin Li, Mohammad Shafiei, Ravi Ramamoorthi, Kalyan Sunkavalli, and Manmohan Chandraker. Inverse rendering for complex indoor scenes: Shape, spatially-varying lighting and svbrdf from a single image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2475–2484, 2020. 2

[22] Matthias Liero, Alexander Mielke, and Giuseppe Savaré. Optimal entropy-transport problems and a new hellinger–kantorovich distance between positive measures. *Inventiones mathematicae*, 211(3):969–1117, 2018. 4

[23] Celong Liu, Zhong Li, Shuxue Quan, and Yi Xu. Lighting estimation via differentiable screen-space rendering. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 575–576. IEEE, 2020. 2

[24] Stephen Lombardi and Ko Nishino. Reflectance and illumination recovery in the wild. *IEEE transactions on pattern analysis and machine intelligence*, 38(1):129–141, 2015. 2

[25] Robert Maier, Kihwan Kim, Daniel Cremers, Jan Kautz, and Matthias Nießner. Intrinsic3d: High-quality 3d reconstruction by joint appearance and geometry optimization with spatially-varying lighting. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3114–3122, 2017. 2

[26] Davide Maino, A Farusi, Carlo Baccigalupi, Francesca Perrotta, AJ Banday, L Bedini, Carlo Burigana, Gianfranco De Zotti, KM Górski, and E Salerno. All-sky astrophysical component separation with fast independent component analysis (fastica). *Monthly Notices of the Royal Astronomical Society*, 334(1):53–68, 2002. 3

[27] D Marinucci, D Pietrobon, A Balbi, P Baldi, P Cabella, G Kerkyacharian, Paolo Natoli, D Picard, and N Vittorio. Spherical needlets for cosmic microwave background data analysis. *Monthly Notices of the Royal Astronomical Society*, 383(2):539–545, 2008. 2

[28] Colm Mulcahy. Image compression using the haar wavelet transform. *Spelman Science and Mathematics Journal*, 1(1):22–31, 1997. 8

[29] Lukas Murmann, Michael Gharbi, Miika Aittala, and Fredo Durand. A dataset of multi-illumination images in the wild. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4080–4089, 2019. 2

[30] Francis J Narcowich, Pencho Petrushev, and Joseph D Ward. Localized tight frames on spheres. *SIAM Journal on Mathematical Analysis*, 38(2):574–594, 2006. 2, 3

[31] Jinwoo Park, Hunmin Park, Sung-Eui Yoon, and Woontack Woo. Physically-inspired deep light estimation from a homogeneous-material object for mixed reality lighting. *IEEE Transactions on Visualization and Computer Graphics*, 26(5):2002–2011, 2020. 2

[32] Davide Pietrobon, Amedeo Balbi, and Domenico Marinucci. Integrated sachs-wolfe effect from the cross correlation of wmap3 year and the nrao vla sky survey data: New results and constraints on dark energy. *Physical Review*, 2006. 2

[33] Peter Schröder and Wim Sweldens. Spherical wavelets: Efficiently representing functions on the sphere. In *Proceedings of the 22nd annual conference on Computer graphics and interactive techniques*, pages 161–172, 1995. 2

[34] Shuran Song and Thomas Funkhouser. Neural illumination: Lighting prediction for indoor environments. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 6918–6926, 2019. 1

[35] Pratul P Srinivasan, Ben Mildenhall, Matthew Tancik, Jonathan T Barron, Richard Tucker, and Noah Snavely. Lighthouse: Predicting lighting volumes for spatially-coherent illumination. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8080–8089, 2020. 1, 2

[36] Flavien Vansyngel, Benjamin D Wandelt, Jean-François Cardoso, and Karim Benabed. Semi-blind bayesian inference of cmb map and power spectrum. *Astronomy & Astrophysics*, 588:A113, 2016. 3

[37] Henrique Weber, Donald Prévost, and Jean-François Lalonde. Learning to estimate indoor lighting from 3d objects. In *2018 International Conference on 3D Vision (3DV)*, pages 199–207. IEEE, 2018. 2

[38] Chuhui Xue, Shijian Lu, and Fangneng Zhan. Accurate scene text detection through border semantics awareness and bootstrapping. In *Proceedings of the European conference on computer vision (ECCV)*, pages 355–372, 2018. 2

[39] Yingchen Yu, Fangneng Zhan, Rongliang Wu, Jianxiong Pan, Kaiwen Cui, Shijian Lu, Feiying Ma, Xuansong Xie, and Chunyan Miao. Diverse image inpainting with bidirectional and autoregressive transformers. *arXiv preprint arXiv:2104.12335*, 2021. 2

[40] Fangneng Zhan and Shijian Lu. Esir: End-to-end scene text recognition via iterative image rectification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2059–2068, 2019. 2

[41] Fangneng Zhan, Shijian Lu, and Chuhui Xue. Verisimilar image synthesis for accurate detection and recognition of texts in scenes. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 249–266, 2018. 2

[42] Fangneng Zhan, Shijian Lu, Changgong Zhang, Feiying Ma, and Xuansong Xie. Adversarial image composition with auxiliary illumination. In *Proceedings of the Asian Conference on Computer Vision*, 2020. 2

[43] Fangneng Zhan, Shijian Lu, Changgong Zhang, Feiying Ma, and Xuansong Xie. Towards realistic 3d embedding via view alignment. *arXiv preprint arXiv:2007.07066*, 2020. 2

[44] Fangneng Zhan, Chuhui Xue, and Shijian Lu. Ga-dan: Geometry-aware domain adaptation network for scene text detection and recognition. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 9105–9115, 2019. 2

[45] Fangneng Zhan, Yingchen Yu, Kaiwen Cui, Gongjie Zhang, Shijian Lu, Jianxiong Pan, Changgong Zhang, Feiying Ma, Xuansong Xie, and Chunyan Miao. Unbalanced feature transport for exemplar-based image translation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2021. 2

[46] Fangneng Zhan, Yingchen Yu, Rongliang Wu, Kaiwen Cui, Aoran Xiao, Shijian Lu, and Ling Shao. Bi-level feature alignment for semantic image translation & manipulation. *arXiv preprint*, 2021. 2

[47] Fangneng Zhan, Yingchen Yu, Rongliang Wu, Changgong Zhang, Shijian Lu, Ling Shao, Feiying Ma, and Xuansong Xie. Gmlight: Lighting estimation via geometric distribution approximation. *arXiv preprint arXiv:2102.10244*, 2021. 2

[48] Fangneng Zhan and Changgong Zhang. Spatial-aware gan for unsupervised person re-identification. In *2020 25th International Conference on Pattern Recognition (ICPR)*, pages 6889–6896. IEEE, 2021. 2

[49] Fangneng Zhan, Hongyuan Zhu, and Shijian Lu. Scene text synthesis for efficient and effective deep network training. *arXiv preprint arXiv:1901.09193*, 2019. 2

[50] Fangneng Zhan, Hongyuan Zhu, and Shijian Lu. Spatial fusion gan for image synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3653–3662, 2019. 2

[51] Changgong Zhang, Fangneng Zhan, and Yuan Chang. Deep monocular 3d human pose estimation via cascaded dimension-lifting. *arXiv preprint arXiv:2104.03520*, 2021. 2

[52] Edward Zhang, Michael F Cohen, and Brian Curless. Emptying, refurnishing, and relighting indoor spaces. *ACM Transactions on Graphics (TOG)*, 35(6):1–14, 2016. 2

[53] 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 Conference on Computer Vision and Pattern Recognition*, pages 10158–10166, 2019. 2

[54] Daniel Zoran, Dilip Krishnan, Jose Bento, and Bill Freeman. Shape and illumination from shading using the generic viewpoint assumption. In *Advances in Neural Information Processing Systems*, pages 226–234, 2014. 2