Estimating Physically-Based Reflectance Parameters From a Single Image With GAN-Guided CNN

CHI-HYOUNG RHEE AND CHANG HA LEE[iD]

School of Computer Science and Engineering, Chung-Ang University, Seoul 06973, South Korea

Corresponding author: Chang Ha Lee (chlee@cau.ac.kr)

ABSTRACT We present a method that estimates the physically accurate reflectance of materials from a single image and reproduces real world materials which can be used in well-known graphics engines and tools. Recovering the BRDF (bidirectional reflectance distribution function) from a single image is an ill-posed problem due to the insufficient irradiance and geometry information as well as the insufficient samples on the BRDF parameters. The problem could be alleviated with a simplified representation of the surface reflectance such as Phong reflection model. Recent works have appealed that convolutional neural network successfully predicts parameters of empirical BRDF models for non-Lambertian surfaces. However, parameters of the physically-based model confront the problem of having non-orthogonal space, making it difficult to estimate physically meaningful results. In this paper, we propose a method to estimate parameters of a physically-based BRDF model from a single image. We focus on the metallic property of the physically-based model to enhance the estimation accuracy. Since metals and nonmetals have very different characteristics, our method processes them separately. Our method also generates auxiliary maps using a cGAN (conditional generative adversarial network) architecture to help in estimating more accurate BRDF parameters. Based on the experimental results, the auxiliary map is selected as an irradiance environment map for the metallic and a specular map for the nonmetallic. These auxiliary maps help to clarify the contributions of different actors, including light color, material color, specular component, and diffuse component, to the surface color. Our method first estimates whether the material on the input image is metallic or nonmetallic. Then, it estimates BRDF parameters using CNN (convolutional neural networks) architecture guided by generated auxiliary maps. Our results show that our method is effective to estimate BRDF parameters both on synthesized as well as real images.

INDEX TERMS Physically-based rendering, BRDF estimation, artificial intelligence, convolutional neural networks, generative adversarial networks, supervised learning.

I. INTRODUCTION

Extracting material properties has been a classic computer vision and graphics problem. The image synthesis is the result of complex physics; shape, reflectance, and illumination need to be known on the processes through the rendering pipeline. The inverse rendering problem, i.e. inferring intrinsic properties from an image is very difficult to solve under most of the assumption since the same visual result can be rendered from the multiple combinations of properties. The estimation

The associate editor coordinating the review of this manuscript and approving it for publication was Haiyong Zheng[iD].

of reflectance from a single monocular color image with uncontrolled illumination is an ill-posed problem even though we have a prior information of the object's shape.

Bidirectional Reflectance Distribution Function (BRDF) is commonly used to encapsulate the material reflective characteristic. Data-driven approaches have advantage for capturing the detailed appearance of real world materials. However, the acquisition and measurement of the BRDF is a time-consuming and delicate process due to the high dimension of the function which is at least 4 (BRDF) and up to 8 (BSSRDF). Alternative approaches to avoid this problem are simplified analytic BRDF models controlled by a set

of parameters. The efficiency and expressiveness of these models are the major merit when the commercial graphic rendering applications reproduce a realistic scene. Despite those merits, the process of adjusting the model parameters by hand is confusing and error-prone task for the application artists.

Deep learning enables the automatic learning the underlying features of both natural and synthetic images from large training data. Recent works demonstrate the capability of solving the non-linear optimization of classic graphics problems such as parameter estimation of rendering equation and environment reproduction, and furthermore convolutional neural networks are particularly known for high performance in the context of a single object. References [1] and [2] estimate BRDF parameters of the material from one or more images of specular objects in a specific class including cars, chairs, and couches. Reference [3] proposed a realtime approach for estimating the surface reflectance from sequential RGBD images in 90ms. These prior works assume the empirical BRDF models such as Phong and Blinn-Phong models that have limited ability to represent the complex appearance of the realistic material.

Our research aims at the handy adjustment of material properties for physically-based rendering applications. Empirical BRDF models have mostly straight-forward and intuitive parameters so that 3D graphic artists can define materials by adjusting them with little difficulty. However, they have limited expressiveness that does not fully capture the physical reflection on the real material surface, especially at grazing angles. In order to synthesize photo-realistic results, state-of-the-art rendering applications support various physically-based models. Disney BRDF [4] is a representative reflectance model that is developed in the principled and art-directable philosophy, but nevertheless it conducts sufficiently realistic results. Our goal is to construct the estimation procedure for major parameters used in Disney BRDF shading model.

In this paper, we propose a method based on generative network guided convolutional neural network to automatically estimate the physically-based BRDF parameters of an isotropic single material in a spherical shape. We assume that acquiring the spherical reflectance map [1] is available prior to our estimation process. We exploit the property which metals and nonmetals (dielectrics) distinctively contain, making two different training networks. Moreover, we enhance the training of the parameter estimator by generating a spherical auxiliary map of the irradiance environment or specular reflection, motivating the cyclical architecture of GAN (generative adversarial networks).

Our contributions are as follows. (1) We propose a deep learning method to estimate BRDF parameters of physically-based shading model from a single image. By using this method, we can synthesize realistic images with physically-based materials captured from real-world objects. (2) We design GAN architecture not only to predict the irradiance map but also to help estimating BRDF parameters

simultaneously, by applying the cooperative structure of map generator and BRDF estimator. The generated irradiance map can be also used for reproducing the illumination environment of the input image.

II. RELATED WORK
A. PARAMETRIC BRDF ESTIMATION
Parametric BRDF estimation has been considered as a reasonable compromise to obtain the visual property of an observed material. Traditional data-driven approaches to acquire material properties are conducted by expensive hardware and slow scanning process in controlled environment. Ngan *et al.* [5] compare the ability of various analytic BRDF models to represent real measured samples, with the supplementary results of parameter fitting. Ghosh *et al.* [6] design an optical setup that allows for basis function illumination of BRDF samples. Dupuy and Jakob [7] use a gonio-photometer to simultaneously manage BRDF acquisition and storage for isotropic and anisotrpic materials. As alternatives to complicated acquisition setup, there has been a variety of approaches to formulate the inverse rendering problem by optimizing parameters through three intrinsic properties, i.e., lighting, geometry, and reflectance. Lombardi and Nishino [8], [9] jointly estimate the reflectance and illumination from a single image of a well-known shape object. Romeiro *et al.* [10] infer BRDF of the isotropic surface through exploiting natural illumination instead of active lighting device. Nielsen *et al.* [11] simplify the measurement procedure for isotropic BRDF reconstruction by reducing the measurement samples based on the regions of importance.

With the progress of the recent studies on deep learning techniques, their performance has been proved to surpass conventional optimizations. Li *et al.* [12] learn spatially varying surface reflectance of a planar material from a single image using self-augmented CNN. Kim *et al.* [3] take RGB-D images as input to estimate the surface reflectance in real-time using two network architectures, HemiCNN and Grouplet. Meka *et al.* [2] also suggest an end-to-end approach for real-time material estimation from a single color image. Georgoulis *et al.* [1] first estimate material intrinsics using a learning-based sparse data interpolation technique, and for the second step, reconstruct reflectance parameters from the estimated material intrinsics. While the objective of prior researches is estimating empirical BRDF models, Vadaurre *et al.* [13] reconstruct physically-based BRDF from two shots of the material. We present the approach that predicts the parameters used in physically-based rendering software from a single shot of the material.

B. ILLUMINATION ESTIMATION
Several studies have estimated illumination of outdoor environment [14] and indoor environment [15] using CNN-based methods. Image transformation techniques [16]–[18] inspire the novel method for generating illumination images. Georgoulis *et al.* [1] estimate an environment map from the

color image of a known shape using U-NET architecture. Boss and Lensch [19] adopt conditional adversarial network that generates differentiable illuminations to be used for predicting BRDF parameters. Our method concerns the two-way process of GAN [16], both illumination estimation and parameter estimation can be simultaneously performed in a single network.

III. PHYSICALLY-BASED RENDERING PARAMETERS
In this section we introduce parameters of the physically-based model for synthesizing training image data. The Cook-Torrance model became the common option in computer graphics since it represents the distinction between metals and nonmetals. The model takes into account both diffuse component that is modeled as Lambertian reflection and the specular component that is modeled as the combination of Fresnel effect, microfacet normal distribution, and geometric visibility. The full model is described as a weighted sum of diffuse and specular components.

$$f_r = k_d(n \cdot l) + f_s. \tag{1}$$

The specular component assumes that the probability of the microfacets aligned in the direction of halfway vector h and the visibility of the facets from different view angles contribute to the surface reflection. These factors are represented through the functions D, F, and G:

$$f_s(v, l) = \frac{F(v)D(h)G(v, l)}{4(n \cdot v)(n \cdot l)}. \tag{2}$$

The normal distribution function $D(h)$ is expressed by GGX [20] which is suitable for modeling the realistic light reflection, providing a subtle softer highlight than Beckmann distribution.

$$D(h) = \frac{\alpha^2 \chi(n \cdot h)}{\pi((n \cdot h)^2(\alpha^2 - 1) + 1)^2}, \tag{3}$$

where α is square of roughness(σ). The geometry function G describes the attenuation of the light caused by the self-shadowing of the microfacets. This self-shadowing is approximated through the Smith factor formulation [21]:

$$G(v, l) = \frac{n \cdot v}{(n \cdot v)(1 - k) + k} \cdot \frac{n \cdot l}{(n \cdot l)(1 - k) + k}, \tag{4}$$

where $k = \frac{\sigma^2}{2}$. The Fresnel function F simulates the light interaction with an angle of incidence, significantly related to the representation of the surface class, i.e. metals or nonmetals. The actual formula is complex and different for conductive (metal) or dielectric (nonmetal) material, and the Schlick's approximation [21] is widely used thanks to the inexpensive computational cost.

$$F(v) = F_0 + (1 - F_0)(1 - (v \cdot h))^5 \tag{5}$$

F_0 is the a proportion of the reflected light when a ray hits surface perpendicularly.

A. ARTIST-FRIENDLY BRDF MODEL
Commercial rendering applications or engines commonly adopt art-directable but not physically strict parameters of the traditional microfacet models. Our model uses 5 parameters for determining the surface reflective appearance, i.e., three for albedo color(A_c), one for metallic(m), and one for roughness(σ). While albedo and roughness have a range of $0.05 \sim 1.0$, our metallic parameter is either 0(nonmetal) or 1(metal). On the periodic table, elements are classified as either metals or nonmetals by their distinctive properties between these two elemental groups. Since most materials around us in our daily life are either metallic or nonmetallic, artist-friendly BRDF models used in modeling tools or game engines use a 0-or-1 metallic variable, which we use in this paper as well. We focus on the metallic properties that act a key role of incident light being reflected on a surface or being transmitted through.

1) METALLIC SURFACE PROPERTY
Metals normally have high reflectivity over the most visible range of the wavelength spectrum of the incident light since the light wave do not penetrate the metallic surface. Only a small fraction of light may be absorbed or refracted, for instance, gold absorbs the blue and violet regions of the spectrum, producing the yellow color when illuminated with white light. This optical aspect can be represented by albedo color (6) that determines the wavelength distribution with respect to the angle of incidence according to Fresnel reflectance.

$$F_{0,c} = m \cdot A_c + (1 - m) \cdot 0.04 \tag{6}$$
$$k_{d,c} = (1 - m) \cdot (1 - F_c) \cdot A_c, \tag{7}$$

where subscript c denotes the color channel. We consider the difference between reflected light and environmental source is the salient feature to infer the intrinsic color of a material.

2) NONMETALLIC SURFACE PROPERTY
Most dielectric (nonmetallic) materials show very different reflectance depending on the angle of incidence. Wood, plastic, rubber, etc. reflect just around 4% of the light at common incident angles and the reflectance gradually increases up to 100% at the grazing angle. In addition, most dielectric materials have similar reflectance curves. Those characteristics are represented by approximating F_0 as 0.04 in the Fresnel reflection (6) for nonmetallic surfaces, where $m = 0$.

In contrast to metals, color representation is dependent on the diffuse reflection. Transmitted light into the interior of the surface experiences a decrease in velocity, leading to the refraction that varies with the wavelength. Equation (7) shows that the diffuse coefficient is proportional to the absorbed and refracted energy, colored by albedo for each color channel. Regarding the above, the full reflection is combination of the diffuse color representation and the specular by Fresnel effect. We believe that the performance of inferring the

intrinsic color would be improved by obtaining separated diffuse or specular component for nonmetallic materials.

IV. BRDF ESTIMATION

There have been several researches to estimate BRDF parameters from a single image using convolutional neural networks [1], [2], [19]. The predicting performance of the BRDF learning is limited due to the ambiguity of the inverse rendering of parameters on naturally illuminated surfaces. The surface color can be interpreted as a result of both the intrinsic albedo of the material and the color of the light source. Since we have only a single input image without the lighting information, it is difficult to distinguish the color of the intrinsic material property from the colors of the light sources, especially for metallic materials. The learning accuracy could be improved with additional data to differentiate these factors.

Our method generates an additional input, an auxiliary map, for guiding the interpretation of the input material image using the conditional generative adversarial network(cGAN) [16]. For metallic materials, we generate an irradiance environment map for the auxiliary map. For nonmetallic materials, it would be more difficult to estimate irradiance environment maps than metals because many of incident lights are refracted inside of the material and scattered out. Learning BRDF parameters could be also limited by the ambiguity between the specular component and the diffuse component of the radiance. For nonmetals, separation of the diffuse reflection and the specular reflection could be an effective clue to reproduce the plausible BRDF parameters. Therefore, we propose to use a specular map, that shows only specular components of the input image, as an auxiliary map for nonmetals. In Section IV-B, we show the results of the experiments to validate this proposition.

Our method first estimates if the material in the input image is metal or nonmetal. For metallic materials, it generates an auxiliary map, which is an irradiance environment map for metals, and estimates BRDF parameters by learning material images as well as the generated auxiliary maps. For nonmetals, it follows the same process except that it uses a specular map as an auxiliary map. Fig. 1 shows the overall process of our method. Our approach to generate auxiliary maps resembles the network of Pix2Pix [16], which is designed for a general purpose image-to-image translation and for generating an image of a given condition. The generator and the discriminator are trained to correspondingly deceive and distinguish the generated image by minimizing the loss function between the generated image and the ground-truth image. For estimating the BRDF parameters of the input material, we intervene the cooperative model named estimator that takes the generated auxiliary map along with the material image as a concatenated input.

We assume that the material in each input image is spherical and homogeneous as shown in Fig. 1. It is also possible to convert general shape material images to spherical images. Georgoulis *et al.* [1] proposed a method to generate a spherical reflectance map from a single-material object of a known class such as cars. We start with the assumption that we obtain either spherical shape objects or general shape objects transformed into spherical shapes.

A. METALLIC CLASSIFIER

The first step of our method is to classify an input material into two categories, metals and nonmetals. The entire subsequent process depends on this classification since our network is trained for materials of each class separately. Fig. 2 shows our metallic classifier architecture, *MetalNet*. It is similar to VGG architecture [22] which starts from 3×3 convolutional layers followed by batch normalization, ReLU, and pooling layers. After these steps, three fully-connected layers are connected to the one-hot encoded output nodes that indicate the metallic parameter. *MetalNet* uses a binary cross-entropy loss function (8), where t_j is the true metallic parameter for the j-th sample and p_j is its predicted metallic parameter. The network is trained by Adam optimizer with 0.0001 learning rate.

$$CE(t, p) = -\frac{1}{N}\left[\sum_{j=1}^{N}[t_j \log(p_j) + (1 - t_j)\log(1 - p_j)]\right] \quad (8)$$

Our metallic classifier, *MetalNet*, presents 98% accuracy with our dataset.

B. PARAMETER ESTIMATOR

In this paper, we estimate five BRDF parameters, that are metallic, albedo colors (RGB), and roughness. The basic approach to acquire the 5 parameters is VGG-like CNN taking the material image as input (CNN in Table 1). It consists of series of 3×3 convolutional layers followed by batch normalization and ReLU. Pooling layers are adopted in every other convolutional layers reducing the size of features by half. The network uses a mean squared loss for regression of 5 parameters and Adam optimizer with 0.0001 learning rate.

As mentioned in Section IV, our method provides an auxiliary map along with the input image to improve the estimation. As the auxiliary map, we use an irradiance environment map or a specular map. The irradiance environment map is a circular image showing the hemi-spherical irradiance environments providing a clue for the incident lights. The second and third columns in Fig. 6 show the true and generated environment maps respectively. The specular map is the specular component of the input image providing a clue for separating the specular components from the diffuse components. The second and third columns in Fig. 7 show the true and generated specular maps respectively.

In our method, the metallic parameter is estimated by *MetalNet* as described in Section IV-A. Our BRDF parameter estimator, *BRDF-Net*, estimates the remaining 4 BRDF parameters, that are albedo colors (RGB) and roughness. Fig. 3 shows the *BRDF-Net* architecture which is similar to the architecture of *MetalNet* in layer connection, activation, and pooling. The major modification is a concatenated input taking a material image and an auxiliary map.
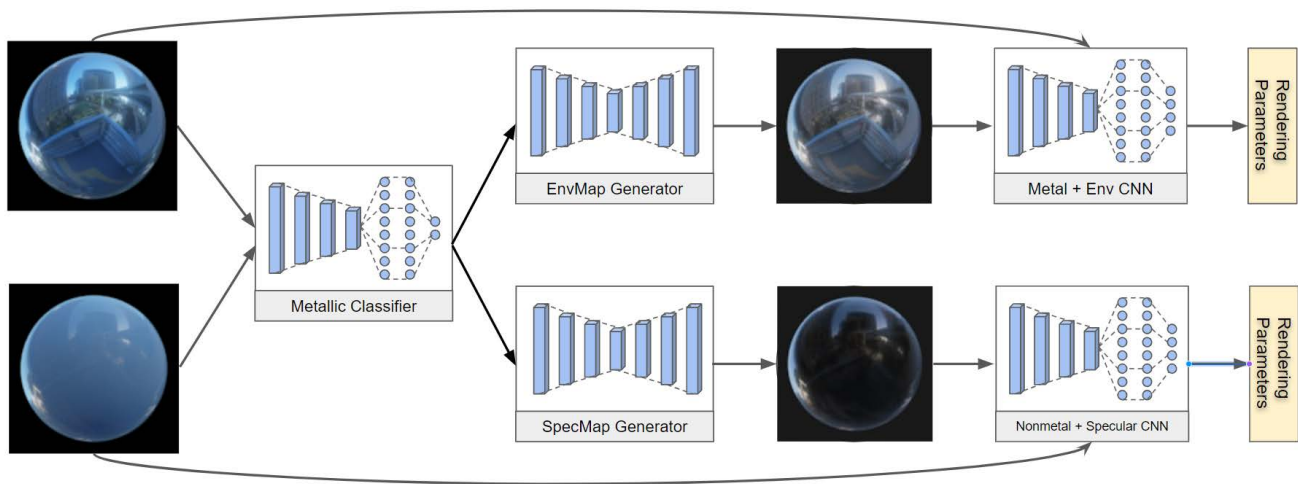
FIGURE 1. Overall process: Our method first determines whether the input material is metal or nonmetal, using a metallic classifier. The metallic materials follow the upper network. First, it generates an auxiliary map, an irradiance environment map for metals, providing an additional input to the parameter estimator. Then, it estimates input material's BRDF parameters. The nonmetallic materials follow the lower network, which is same with the upper network except that the auxiliary map is a specular map.
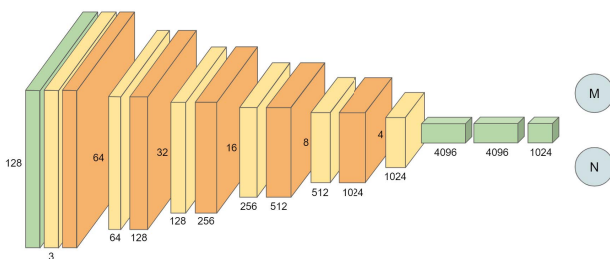

FIGURE 2. *MetalNet*: Architecture of CNN binary classification network. The bottom numbers denote feature channels and the middle numbers represent spatial resolutions of feature maps. Each convolution layer uses 3 × 3 filters until it is transformed into 4 × 4 × 1024 feature maps and is flattened to the subsequent fully-connected layers.

Following layers interpret the relation between the reflection on the material surface and the irradiance environment (specular reflection for nonmetals) and transform them into feature maps using the ReLU activation function. They continue to two fully-connected layers before they finally predict 4 BRDF parameters (albedo RGB and roughness) using the sigmoid activation function. *BRDF-Net* uses a mean squared loss function, L_{L2} as described in (9) and Adam optimizer with 0.0001 learning rate.

$$L_{L2}(E_y) = \mathbb{E}(\|y_p - E_y(x, y_m)\|_2), \quad (9)$$

where x, y_m, and y_p denote input images, true auxiliary maps, and true parameter vectors, respectively. We let E_y be the estimator trained with the true auxiliary map y_m.

We have conducted experiments to determine the appropriate selection of the auxiliary map. In these experiments, we use three networks: a conventional CNN, *BRDF-Net* with irradiance environment maps, and *BRDF-Net* with specular maps. We use three datasets: metals, nonmetals, and mixed

data that contain metals as well as nonmetals in even distribution. All the images in the datasets are provided with corresponding ground-truth environment and specular maps. We have experimented with possible combinations of networks and datasets. Table 1 shows the results. We have found that all the results by providing auxiliary maps along with the input images show smaller errors than the results by the conventional CNN. Metallic materials are best estimated by *BRDF-Net* with irradiance environmental maps. The best estimator for nonmetallic materials is *BRDF-Net* with specular maps. This results also support our hypothesis in Section IV. Therefore, this is the most promising combination for the BRDF parameter estimation. These two networks are followed by the metallic classifier described in Section IV-A.

C. MAP GENERATOR

Our estimator requires not only an input material image but also an additional environment or specular map to precisely predict intrinsic BRDF parameters. Instead of getting the ground-truth maps, we estimate auxiliary maps from the input material image and use it as an input of *BRDF-Net*. Our map generator, *MapGen*, adopts Pix2Pix approach widely known as cGAN which contains generative model competitively contest with discriminative model. The spatial feature resolution of the encoder is gradually reduced by half from 128 to 1, applying 4 × 4 convolutions followed by Leaky-ReLU with stride 2 for downsampling. The first 4 downsampling steps double the number of feature channels and the last 3 layers only reduce the number of feature channels. After a bottleneck layer, the process is reversed. To preserve the low-level information shared between input and output, the network requires skip connections generally
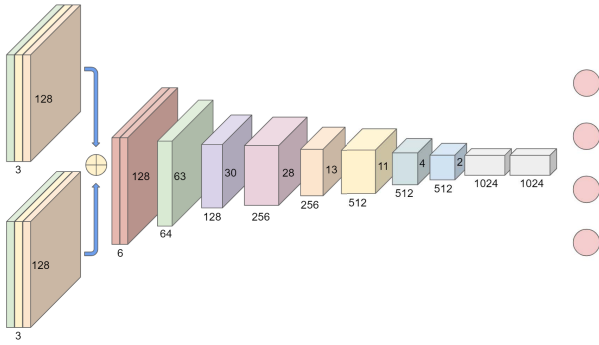
FIGURE 3. *BRDF-Net*: Architecture of the BRDF parameter estimation. Two 128 × 128x3 input images are concatenated into 6 channels and subsequently build similar layers as in Fig. 2 except that the last layer consists of 4 nodes with sigmoid activation.

TABLE 1. Estimation quality comparisons between the basic CNN and our *BRDF-Net* with different types of input and auxiliary maps: Input data can be mixed (both metals and nonmetals), metal, or nonmetal. The auxiliary map can be env (irradiance environment maps) or spec (specular maps). The basic CNN is only tested with mixed input without auxiliary maps, and our *BRDF-Net* is tested with every combination. For both training and testing, we used the ground-truth auxiliary maps. Since metals have no diffuse component, we did not experiment metals with specular maps. p-loss (parameter loss) is the difference between ground-truth BRDF parameters and predicted parameters, and r-loss (render loss) is the difference between ground-truth material images and images rendered with predicted parameters.

	Input	Auxiliary map	p-loss(MAE)	r-loss(RMSE)
CNN	Mixed	-	0.10229	0.03308
BRDF-Net	Mixed	env	0.03799	0.01561
BRDF-Net	Mixed	spec	0.05408	0.02426
BRDF-Net	Metal	env	**0.03121**	**0.01147**
BRDF-Net	Nonmetal	env	0.04613	0.01406
BRDF-Net	Metal	spec	N/A	N/A
BRDF-Net	Nonmetal	spec	**0.03546**	**0.01068**
	Accuracy			
MetalNet	0.98			

used in U-NET [17]. Each skip connection concatenates all feature channels at layer i with those at layer $n - i$.

The objective of the map generator can be expressed as

$$L_{GAN}(G, D) = \mathbf{CE}(1, D(x, y_m)) + \mathbf{CE}(0, D(x, G(x, z))), \quad (10)$$

where discriminator D tries to accept the true map y_m on the observed condition of input image x whereas generator G tries to deceive D into thinking the map $G(x, z)$ generated from a random noise vector z is a true map. In addition, $L1$ loss imposes the generated map not only deceive the discriminator but also imitate the true map in an L1 domain.

$$L_{L1}(G) = \mathbb{E}[\|y_m - G(x, z)\|_1] \quad (11)$$

The previous research [18] suggests mixing those two objectives with the empirically determined weight λ_1 of 100.

$$L_G = \arg \min_G \max_D \{L_{GAN}(G, D) + \lambda_1 L_{L1}(G)\} \quad (12)$$
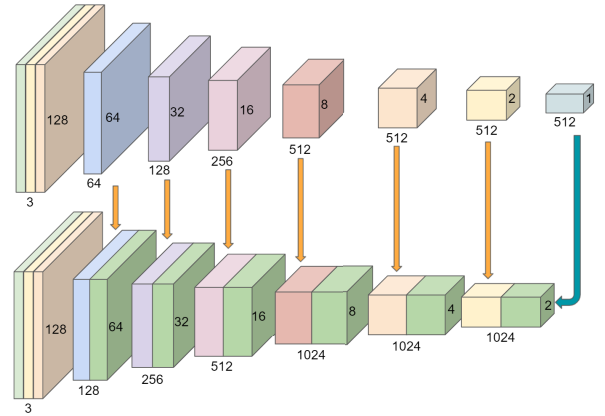

FIGURE 4. *MapGen* follows the architecture of a modified U-Net which consists of an encoder and a decoder. The arrows show the concatenating positions for skip connections between the encoder and the corresponding decoder.

We also use Adam optimizer and follow the suggested learning rate of 0.0002 and momentum term β_1 of 0.5 for the training stability.

D. TRIARCHY GAN

The performance of *BRDF-Net* is enhanced by the additional input generated by *MapGen*. The goal of *MapGen* is to mimic the irradiance environment map of the input image so that it looks plausible. However, since the material image might be drastically changed by a subtle change of the irradiance environment, a less accurate map generator could distract the parameter estimator. Therefore, we propose a triarchy GAN architecture, *triGAN*, that cooperatively estimates auxiliary maps and BRDF parameters. In our architecture, the goal of the map generation is helping to estimate more accurate BRDF parameters in addition to generating more accurate maps. While the network is learning, it improves the auxiliary map as well as the BRDF parameters repeatedly.

In our *triGAN* architecture, *MapGen* first generates the auxiliary map from the input material image. Then *BRDF-Net* takes the generated auxiliary map as an input in addition to the material image, and estimates BRDF parameters. The *BRDF-Net* loss (p-loss) in (13) is added to the *MapGen* total loss function (14) to improve the map generation. In this way, the map generator *MapGen* and the estimator *BRDF-Net* are trained cooperatively so that the generator creates an appropriate auxiliary map to aid the estimator in predicting better BRDF parameters and estimator loss helps the generator to create more helpful auxiliary map. Fig. 5 shows this process.

The objective function of *BRDF-Net* follows (13), which uses a generated auxiliary map instead of a true auxiliary map used in (9).

$$L_{L2}(E_G) = \mathbb{E}[\|y_p - E(x, G(x, z))\|_2] \quad (13)$$

MapGen loss function additionally adopts the L_2-loss of the *BRDF-Net* with weight value $\lambda_2 = 1000$ to adjust the scale.

$$L_{tri}(G) = L_{GAN}(G, D) + \lambda_1 L_{L1}(G) + \lambda_2 L_{L2}(E_G). \quad (14)$$
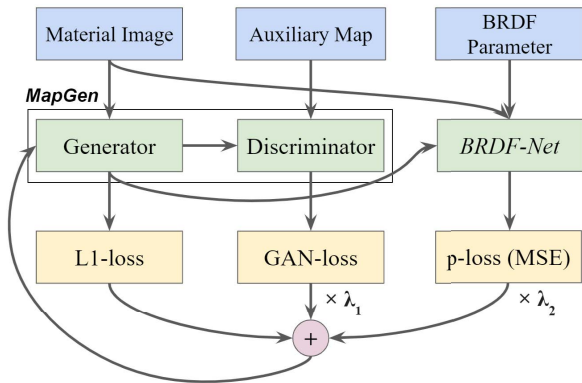
FIGURE 5. Triarchy conditional GAN. *MapGen* generates an auxiliary map providing it to *BRDF-Net* as an additional input. *BRDF-Net* estimates BRDF parameters from the input image and a generated auxiliary map. The loss function of *MapGen* includes the BRDF parameter loss to generate a map that is helpful for the parameter estimation.

TABLE 2. Image difference between ground-truth auxiliary maps and generated maps. *MapGen* is a map generator using cGAN architecture. *triGAN* also generates auxiliary maps while estimating BRDF parameters. The networks with a subscript *env* generate irradiance environment maps and those with a subscript *spec* generate specular maps. The networks are tested with three types of input data: mixed including both metals and nonmetals, metals, and nonmetals. We use RMSE and SSIM metrics for comparing images.

Network		$MapGen_{env}$	$MapGen_{spec}$	$triGAN_{env}$	$triGAN_{spec}$
Mixed	RMSE	0.09201	0.03357	0.09203	0.04949
	SSIM	0.7043	0.9767	0.7047	0.9518
Metal	RMSE	**0.08244**	N/A	0.08297	N/A
	SSIM	**0.7538**		**0.7541**	
Non-metal	RMSE	0.09556	**0.02102**	0.09827	**0.02135**
	SSIM	0.6772	**0.9743**	0.6722	**0.9717**

The extrinsic quality of auxiliary map is not meaningfully improved by the modified objective, but the estimator trained on the generated auxiliary maps produces better BRDF parameters shown in Table 3.

V. DATASETS

We require a large number of paired spherical material images and auxiliary maps (irradiance environment maps and specular maps). Since it is almost impossible to acquire real photos in highly elaborate installations and even pairs, we synthesize rendered images for the training and testing process with 100 HDR irradiance environment maps. HDR environment maps contain 360° panoramic irradiance information in various indoor conditions. The total amount of our synthesized dataset is 40k, where 20k are metals and the other 20k are nonmetals. Each material image has a corresponding environment map and a specular map. The material images are rendered by uniformly sampled parameters of albedo(R, G, B each) and roughness in range from 0.05 to 0.95, following Disney BRDF approach described in Section III-A. For metals or nonmetals, the metallic parameter is determined in 1.0 or 0.0 respectively. Test database has 3.2k synthesized images and 10 real images of homogeneous spherical material photographed indoors. The HDR irradiance environment maps

TABLE 3. The results of BRDF parameter estimation. We have tested with combinations of different networks and input data. The tested networks are *BRDF-Net* and *triGAN*. The networks with a subscript *env* use irradiance environment maps as auxiliary maps and those with a subscript *spec* use specular maps. We use three types of input data: mixed (metal+nonmetal), metal, and nonmetal. The results are MAE between ground-truth parameters and estimated parameters.

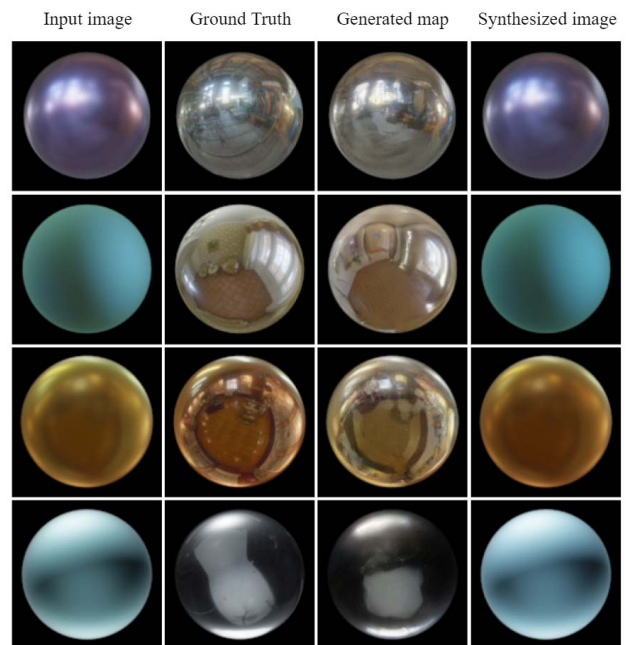| Network	Mixed	Metal	Nonmetal
$BRDF\text{-}Net_{env}$	0.09434	0.08417	0.16865
$BRDF\text{-}Net_{spec}$	0.08158	0.19846	0.07811
$triGAN_{env}$	0.08453	**0.06864**	0.15594
$triGAN_{spec}$	0.08214	0.19821	**0.07286**

FIGURE 6. Results of the BRDF parameter estimation on metals. From left to right: input image, ground-truth environment map, generated environment map, and synthesized image with the estimated parameters. Each material is selected to show various characteristics of materials as well as irradiance environments. From the top row, the difference between input and synthesized images of each material reports 36.51, 35.26, 31.66, and 28.22, in PSNR metric.

for the real photos are captured by DSLR camera compositing multiple exposure images. In addition, the test dataset is synthesized on different irradiance environments from the train dataset.

VI. RESULTS

A. EVALUATION OF ENVIRONMENT AND SPECULAR MAP GENERATION

We evaluate the difference between a ground-truth auxiliary map on the test dataset and a corresponding generated map by *MapGen*. We compare images using RMSE and SSIM metrics with the maximum value = 1.0. Table 2 shows the qualitative comparisons with color space similarity(RMSE) as well as structural similarity(SSIM). Input images categorized as mixed are composed of both metals and nonmetals in same ratio.

TABLE 4. The evaluation of the images rendered by estimated BRDF parameters. The experimented networks are *BRDF-Net* and *triGAN* and the subscript *env* and *spec* mean environment maps and specular maps respectively. These networks are also tested on mixed(metal+nonmetal), metal, and nonmetal separately. We measured RMSE and PSNR between the ground-truth material images and the images rendered by estimated BRDF parameters.

Network		BRDF-Net$_{env}$	BRDF-Net$_{spec}$	triGAN$_{env}$	triGAN$_{spec}$
Mixed	RMSE	0.03603	0.03255	0.03305	0.03257
	PSNR	31.57	32.76	32.16	32.33
Metal	RMSE	0.03556	0.08068	**0.02933**	0.08237
	PSNR	31.43	23.17	**33.14**	22.92
Non-Metal	RMSE	0.07424	0.02898	0.07535	**0.02741**
	PSNR	23.87	33.19	23.88	**33.81**

FIGURE 8. Metallic classification performance with respect to the material roughness. The blue bar represents accuracy of metallic classification and the stacked orange bar shows the binary cross-entropy loss of the predicted class. Each bar covers the range between the material roughness denoted in x-axis.
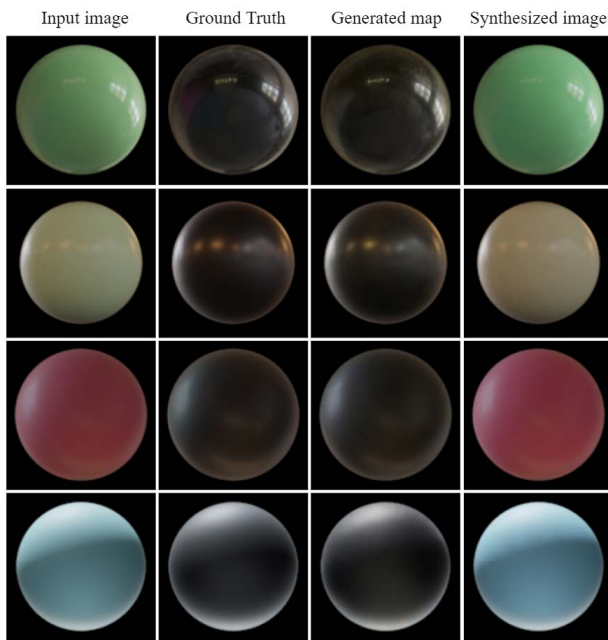

FIGURE 7. Results of the BRDF parameter estimation on nonmetals. From left to right: input image, ground-truth specular map, generated specular map, and synthesized image with the estimated parameters. From the top row, the difference between input and synthesized images of each material reports 33.09, 34.21, 31.77, and 31.79, in PSNR metric.

FIGURE 9. Comparison to the results by Georgoulis *et al.* The images in each column show the input images used for the testing (Input), the ground-truth images in a different environment (GT), their result images (Georgoulis), and our results (Ours). The second and the third columns use the same irradiance environment map, whereas the other columns use difference environment maps. Our results show similar or better estimation especially for metallic materials shown in the third and the fourth rows.

The generator trained on the mixed dataset predicts 5 parameters including metallic, unemploying the metallic classifier result. Notice that the generated results of categorized material dataset present smaller errors than the results of mixed material dataset. We first experimented a map generation with *MapGen* which uses a loss function in (12). Auxiliary maps are also generated in our *triGAN* architecture to guide a BRDF estimation, which uses a loss function (14) including BRDF parameter loss term. Each network is trained separately on irradiance environment maps and specular maps.

The map generator presents improved results on categorized material dataset. However, the quality difference between the results of *MapGen* and *triGAN* is insignificant on both RMSE and SSIM metrics. Nonetheless, the generated
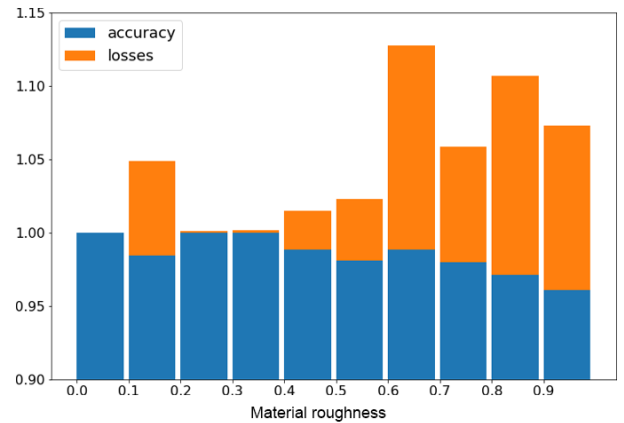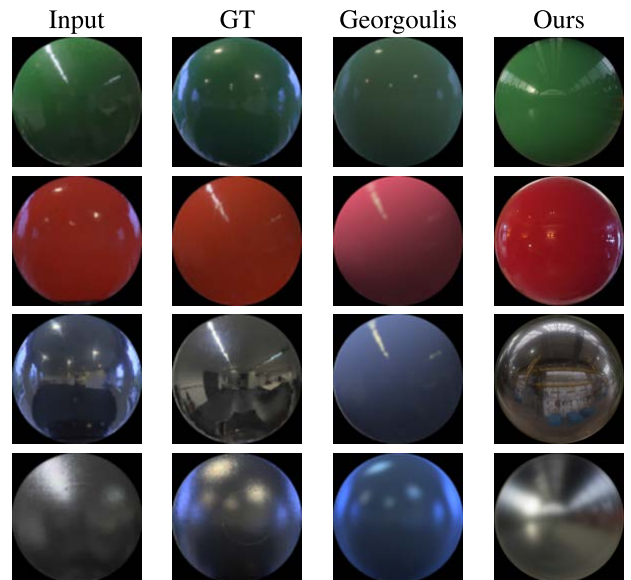
maps in *triGAN* can help the BRDF estimation better than the maps in *BRDF-Net* as described in the next section.

B. EVALUATION ON SYNTHESIZED DATASET

In this section, we evaluate our BRDF estimation method using the categorized test data. We calculated mean absolute error (MAE) between the ground-truth BRDF parameters and their estimated parameters for each mixed, metal, and nonmetal test set. We also tested two networks *BRDF-Net*
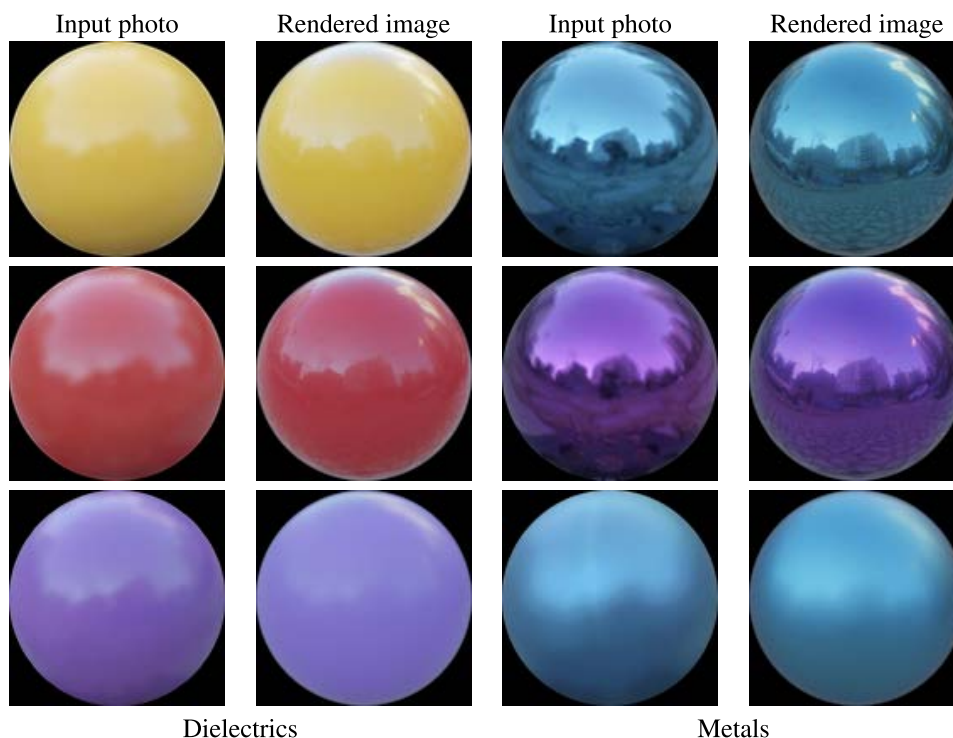
FIGURE 10. Results on real materials: The left two columns present real photos of nonmetal objects in spherical shape and rendered images by our BRDF estimation. The right two columns show results with real metallic objects.

and *triGAN* with environment maps as well as specular maps. Table 3 shows the results.

As derived in Section IV, our approach using environment maps for metals and specular maps for nonmetals shows the best results both for *BRDF-Net* and *triGAN*. Unlike the *BRDF-Net* tested with true environment and specular maps as shown in Table 1, the *BRDF-Net* tested with the auxiliary maps generated by *MapGen* presents less accurate performance as shown in Table 3. Whereas *BRDF-Net* learns the features of true auxiliary maps, which we cannot get, *triGAN* learns the features of generated maps which can be obtained in our method. In *triGAN*, the estimator is trained for intrinsic features of generated maps, presenting better results in both parameter and image spaces.

We synthesized spherical material images with the estimated BRDF parameters using the physically-based rendering method described in Section III. We compare the synthesized results to the ground-truth material images using RMSE and PSNR (Peak Signal-to-noise Ratio). The structural similarity (SSIM) is practically identical for the entire dataset and results since they assume the same spherical shape and illumination. Table 4 shows our method also produces finer synthesized images than previous CNNs. Our final system first determines metalness of the input image, then it runs *triGAN$_{env}$* for metals and *triGAN$_{spec}$* for nonmetals. Based on the accuracy of *MetalNet*, 0.98, the total RMSE error of our system is 0.02937, which shows a preferable result than the error of *triGAN* for the mixed inputs.

C. QUANTITATIVE ANALYSIS WITH RESPECT TO MATERIAL ROUGHNESS

We notice that the quality of estimation is highly related on the material roughness. Extremely rough surface presents practically perfect diffuse, leading to be mistaken classifying its metallic even by human perception. The performance of the metallic classifier consequently affect the performance of the generator and estimator. Fig. 8 depicts both accuracy and binary cross entropy loss of *MetalNet* with respect to material roughness. Accuracy experiences marginal decrease as roughness becomes larger, whereas the loss shows moderate increase. The relatively large losses of rough materials draw question that whether we could obtain the similar accuracy even on real dataset. Since metallic classification is crucial to the performance of the consecutive generation and estimation, this question remains for the future work.

D. COMPARISON TO THE PREVIOUS WORK

We compared our method to the learning based approach by Georgoulis *et al.* [1]. We tested our method with their real dataset and Fig. 9 shows the results. The first column shows the input images used for the testing, and the second column shows the ground-truth images in different environments. The third column shows their result images and the fourth column shows our results generated from the input in the first column. The images in the second and the third columns use the same irradiance environment map. The images in the first column are taken in an environment different from the second and

the third columns. Since we do not have the environment maps used in these images, we use different environment maps for generating our results in the fourth column. Our results show similar or better estimation even though they are rendered in different environments. The last two rows show the results of metallic materials. Since the colors of metals are more affected by the environment, the BRDF estimation might also be affected by the environment. In the third row, The approach by Georgoulis *et al.* estimates the bluish material, which might be affected by the bluish environment of the input image. Our approach estimates more accurate color and reflections by estimating the environment map as well.

E. RESULTS ON REAL WORLD MATERIALS

Our final test is conducted for the real world materials. Due to our limitation of input material shape, we collected 10 spherical-shaped materials, 5 are metals and the others are nonmetals, in diverse surface roughness. Fig. 10 depicts the ground-truth input material images and synthesized images rendered by the estimated BRDF parameters. All materials are photographed in a same outdoor environment. We also captured the irradiance environment where the input photo is taken, and used it for rendering the results. However, the captured environment cannot be exactly same because of the temporal difference, the photographer and the tripod shown in the images, etc. For example, the photographer and the tripod appear only in the input and the clouds appear only in the rendered image. Since it is very difficult to create the strictly identical shape for rendering, we show the visual comparisons instead of quantitative evaluations here.

VII. CONCLUSION

In this paper, we have presented the two-way approach to estimate the realistic radiance properties of metallic and non-metallic materials using the cooperative architecture of CNN and GAN. We have shown that a single material image of known shape is applicable for simultaneously generating an irradiance environment map and estimating material properties. We have also shown that the performance of estimation is improved by the assistance of generated features from GAN. The two different networks working in one dataset with exchange of their gradients can be explained that they have deeper layers in total, whereas they avoid gradient vanishing. To prove our hypothesis that the reflective behavior of metals and nonmetals determines the optimization approach for other parametric properties such as albedo colors and roughness, we have synthesized large scale rendered dataset with various indoor environments and separated into the two groups. We applied a different pipeline of training networks for each group and compared the estimation quality. We confirmed that metals with environment maps and non-metals with specular maps are appropriate combination for our pipeline. Our method shows meaningful improvement of estimation quality on the synthesized test dataset, and

plausible synthesized results from the BRDF estimation of real world materials.

There are many studies that need to be improved in the future. We would like to capture the BRDF parameters from heterogeneous general-shape materials in uncontrolled photos. Our approach also has limitations for the BRDF properties such as anisopropic, tint, sheen, clear-coat, and spatially-varying properties. We would also like to study these properties for more realistic representations of materials.

REFERENCES

[1] 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 Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 8, pp. 1932–1947, Aug. 2018.

[2] A. Meka, M. Maximov, M. Zollhofer, A. Chatterjee, H.-P. Seidel, C. Richardt, and C. Theobalt, "LIME: Live intrinsic material estimation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6315–6324.

[3] K. Kim, J. Gu, S. Tyree, P. Molchanov, M. NieBner, and J. Kautz, "A lightweight approach for on-the-fly reflectance estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 20–28.

[4] B. Burley, "Physically-based shading at Disney," in *Proc. ACM SIGGRAPH*, 2012, pp. 1–7.

[5] A. Ngan, F. Durand, and W. Matusik, "Experimental analysis of BRDF models," in *Proc. EGSR*, 2005, pp. 117–126.

[6] A. Ghosh, W. Heidrich, S. Achutha, and M. O'Toole, "A basis illumination approach to BRDF measurement," *Int. J. Comput. Vis.*, vol. 90, no. 2, pp. 183–197, Nov. 2010.

[7] J. Dupuy and W. Jakob, "An adaptive parameterization for efficient material acquisition and rendering," *ACM Trans. Graph.*, vol. 37, no. 6, pp. 1–14, Jan. 2019.

[8] S. Lombardi and K. Nishino, "Reflectance and natural illumination from a single image," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 117–126.

[9] S. Lombardi and K. Nishino, "Radiometric scene decomposition: Scene reflectance, illumination, and geometry from RGB-D images," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 305–313.

[10] F. Romeiro, Y. Vasilyev, and T. Zickler, "Passive reflectometry," in *Proc. Eur. Conf. Comput. Vis.*, 2008, pp. 859–872.

[11] J. B. Nielsen, H. W. Jensen, and R. Ramamoorthi, "On optimal, minimal BRDF sampling for reflectance acquisition," *ACM Trans. Graph.*, vol. 34, no. 6, pp. 1–11, Nov. 2015.

[12] X. Li, Y. Dong, P. Peers, and X. Tong, "Modeling surface appearance from a single photograph using self-augmented convolutional neural networks," *ACM Trans. Graph.*, vol. 36, no. 4, pp. 1–11, Jul. 2017.

[13] R. Vidaurre, D. Casas, E. Garces, and J. Lopez-Moreno, "BRDF estimation of complex materials with nested learning," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2019, pp. 1347–1356.

[14] Y. Hold-Geoffroy, K. Sunkavalli, S. Hadap, E. Gambaretto, and J.-F. Lalonde, "Deep outdoor illumination estimation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 7312–7321.

[15] M.-A. Gardner, K. Sunkavalli, E. Yumer, X. Shen, E. Gambaretto, C. Gagné, and J.-F. Lalonde, "Learning to predict indoor illumination from a single image," *ACM Trans. Graph.*, vol. 36, no. 6, pp. 1–14, Nov. 2017.

[16] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1125–1134.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. MICCAI*, 2015, pp. 234–241.

[18] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," in *Proc. ICLR*, 2016, pp. 1–16.

[19] M. Boss and H. P. A. Lensch, "Single image BRDF parameter estimation with a conditional adversarial network," 2019, *arXiv:1910.05148*.

[20] B. Walter, S. R. Marschner, H. Li, and K. E. Torrance, "Microfacet models for refraction through rough surfaces," in *Proc. 18th Eurograph. Conf. Rendering Techn.*, 2007, pp. 195–206.

[21] C. Schlick, "An inexpensive BRDF model for physically-based rendering," *Comput. Graph. Forum*, vol. 13, no. 3, pp. 233–246, Aug. 1994.

[22] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Sep. 2014, pp. 1–14.

[23] K. E. Torrance and E. M. Sparrow, "Theory for off-specular reflection from roughened surfaces," *J. Opt. Soc. Amer.*, vol. 57, no. 9, pp. 1105–1114, 1967.

[24] B. T. Phong, "Illumination for computer generated pictures," *Commun. ACM*, vol. 18, no. 6, pp. 311–317, Jun. 1975.

[25] R. L. Cook and K. E. Torrance, "A reflectance model for computer graphics," *ACM Trans. Graph.*, vol. 1, no. 1, pp. 7–24, Jan. 1982.

[26] M. Ashikhmin and P. Shirley, "An anisotropic Phong BRDF model," *J. Graph. Tools*, vol. 5, no. 2, pp. 25–32, 2000.

[27] W. Matusik, H. Pfister, M. Brand, and L. McMillan, "A data-driven reflectance model," *ACM Trans. Graph.*, vol. 22, no. 3, pp. 759–769, Jul. 2003.

[28] K. Hara, K. Nishino, and K. Lkeuchi, "Light source position and reflectance estimation from a single view without the distant illumination assumption," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 4, pp. 493–505, Apr. 2005.

[29] T. Narihira, M. Maire, and S. X. Yu, "Direct intrinsics: Learning albedo-shading decomposition by convolutional regression," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, p. 2992.

[30] D. Guarnera, G. C. Guarnera, A. Ghosh, C. Denk, and M. Glencross, "BRDF representation and acquisition," *Comput. Graph. Forum*, vol. 35, no. 2, pp. 625–650, 2016.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 2, pp. 84–90, Jun. 2012.

CHI-HYOUNG RHEE received the B.S. degree in electrical and electronics engineering and the M.S. degree in computer science from Chung-Ang University, Seoul, South Korea, in 2008 and 2010, respectively, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Engineering.

His research interests include computer graphics, virtual reality, and deep learning.

CHANG HA LEE received the B.S. and M.S. degrees in computer science and statistics from Seoul National University, Seoul, South Korea, in 1995 and 1997, respectively, and the Ph.D. degree in computer science from the University of Maryland, College Park, MD, USA, in 2005.

He was a Research Scientist with the National Capital Area Medical Simulation Center. He is currently an Associate Professor with the School of Computer Science and Engineering, Chung-Ang University, Seoul. His research interests include computer graphics, virtual reality, and deep learning.

• • •