Controllable Light Diffusion for Portraits

David Futschik[1,2] Kelvin Ritland[1] James Vecore[1] Sean Fanello[1]
Sergio Orts-Escolano[1] Brian Curless[1,3] Daniel Sýkora[1,2] Rohit Pandey[1]

[1]Google Research [2]CTU in Prague, FEE [3]University of Washington

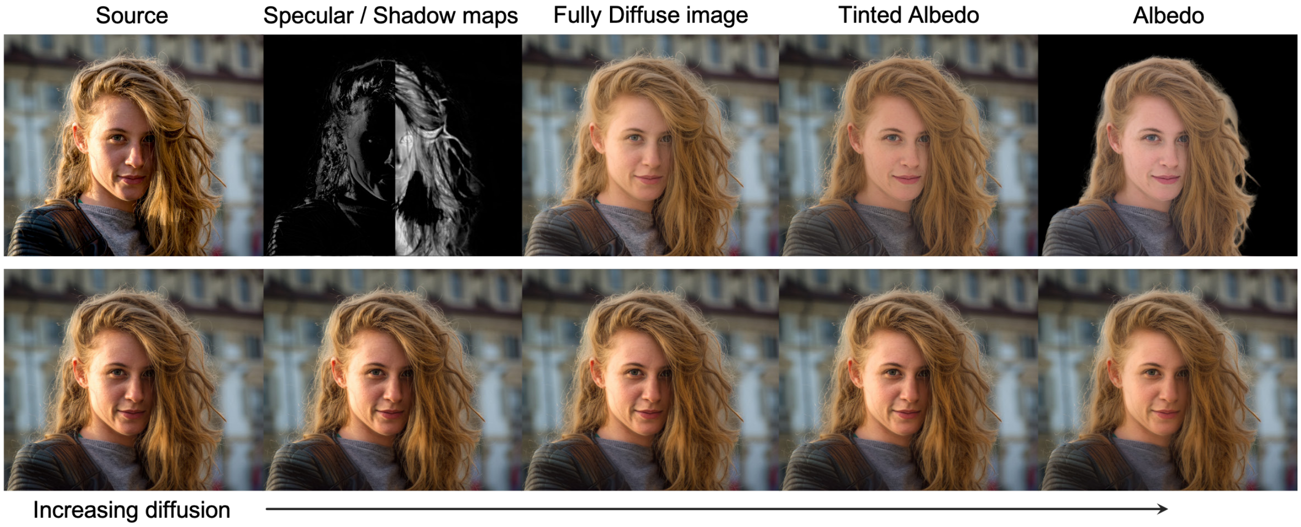

Figure 1. Our method enables control over the diffuseness of light in arbitrary portrait images. We first extract specular/shadow maps from the input image and then produce a fully diffuse image. Additionally, we show how to recover a uniformly lit image, *i.e.* tinted by the average light color, from which we can estimate the untinted albedo. The bottom row illustrates the application of editing the input photo by gradually increasing the amount of light diffusion.

Abstract

We introduce light diffusion*, a novel method to improve lighting in portraits, softening harsh shadows and specular highlights while preserving overall scene illumination. Inspired by professional photographers' diffusers and scrims, our method softens lighting given only a single portrait photo. Previous portrait relighting approaches focus on changing the entire lighting environment, removing shadows (ignoring strong specular highlights), or removing shading entirely. In contrast, we propose a learning based method that allows us to control the amount of light diffusion and apply it on in-the-wild portraits. Additionally, we design a method to synthetically generate plausible external shadows with sub-surface scattering effects while conforming to the shape of the subject's face. Finally, we show how our approach can increase the robustness of higher level vision applications, such as albedo estimation, geometry estimation and semantic segmentation.*

1. Introduction

High quality lighting of a subject is essential for capturing beautiful portraits. Professional photographers go to great lengths and cost to control lighting. Outside the studio, natural lighting can be particularly harsh due to direct sunlight, resulting in strong shadows and pronounced specular effects across a subject's face. While the effect can be dramatic, it is usually not the desired look. Professional photographers often address this problem with a scrim or diffuser (Figure 2), mounted on a rig along the line of sight from the sun to soften the shadows and specular highlights, leading to much more pleasing portraits [9]. Casual photographers, however, generally lack the equipment, expertise, or even the desire to spend time in the moment to perfect the lighting in this way. We take inspiration from professional photography and propose to diffuse the lighting on a subject in an image, i.e., directly estimating the appearance of the person as though the lighting had been softer, enabling anyone to improve the lighting in their photos after the shot

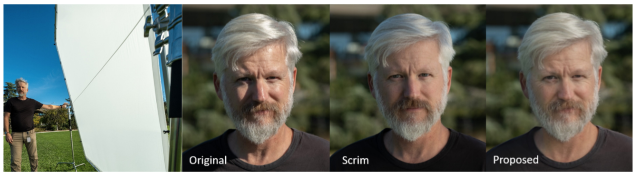

Figure 2. Using a bulky scrim (left), a photographer can reduce strong shadows and specularities. Our proposed approach operates directly on the original image to produce a similar softening effect.

is taken.

Deep learning approaches have led to great advances in relighting portraits [12, 21, 23, 26, 27, 29, 33, 35, 37, 38]. Our goal is different: we want to improve the existing lighting rather than replace it entirely. This goal has two advantages: the resulting portrait has improved lighting that is visually consistent with the existing background, and the task is ultimately simpler, leading to a more robust solution than completely relighting the subject under arbitrary illumination. Indeed, one could estimate the lighting [14, 15], diffuse (blur) it, and then relight the subject [12, 23, 33], but lighting estimation and the full relighting task themselves are open research questions. We instead go directly from input image to diffused-lighting image without any illumination estimation.

Past works [11, 37] specifically focused on removing shadows from a subject via CNNs. However, these methods do not address the unflattering specularities that remain which our work tackles.

In the extreme, lighting can be diffused until it is completely uniform. The problem of "delighting," recovering the underlying texture (albedo) as though a subject has been uniformly lit by white light[1], has also been studied (most recently in [30]). The resulting portrait is not suitable as an end result – too flat, not visually consistent with the background – but the albedo map can be used as a step in portrait relighting systems [23]. Delighting, however, has proved to be a challenging task to do well, as the space of materials, particularly clothing, can be too large to handle effectively.

In this paper, we propose *light diffusion*, a learning-based approach to controllably adjust the levels of diffuse lighting on a portrait subject. The proposed method is able to soften specularities, self shadows, and external shadows while maintaining the color tones of the subject, leading to a result that naturally blends into the original scene (see Fig. 1). Our variable diffusion formulation allows us to go from subtle shading adjustment all the way to removing the shading on the subject entirely to obtain an albedo robust to shadows and clothing variation.

Our overall contributions are the following:

- A novel, learning-based formulation for the light diffusion problem, which enables controlling the strength of shadows and specular highlights in portraits.
- A synthetic external shadow generation approach that conforms to the shape of the subject and matches the diffuseness of the illumination.
- A robust albedo predictor, able to deal with color ambiguities in clothing with widely varying materials and colors.
- Extensive experiments and comparisons with state-of-art approaches, as well as results on downstream applications showing how light diffusion can improve the performance of a variety of computer visions tasks.

2. Related Work

Controlling the illumination in captured photos has been exhaustively studied in the context of portrait relighting [12, 21, 23, 26, 27, 29, 33, 35, 37, 38], which tries to address this problem for consumer photography using deep learning. Generative models and inverse rendering [1, 7, 17, 22, 28] have also been proposed to enable face editing and synthesis of portraits under any desired illumination.

The method of Sun et al. [27], was the first to propose a self-supervised architecture to infer the current lighting condition and replace it with any desired illumination to obtain newly relit images. This was the first deep learning method applied to this specific topic, overcoming issues of previous approaches such as [26].

However, this approach does not explicitly perform any image decomposition, relying on a full end-to-end method, which makes its explainability harder. More recent methods [12, 21, 23, 29] decompose the relighting problem into multiple stages. These approaches usually rely on a geometry network to predict surface normals of the subject, and an albedo predictor generates a *de-lit* image of the portrait, that is close to the actual albedo (i.e. if the person was illuminated by a white diffuse light from any direction). A final neural renderer module combines geometry, albedo and target illumination to generate the relit image. Differently from previous work, Pandey et al. [23] showed the importance of a per-pixel aligned lighting representation to better exploit U-Net architectures [25], showing state-of-art results for relighting and compositing.

Other methods specifically focus on the problem of image decomposition [2, 3, 13, 18, 19, 24, 30–32], attempting to decompose the image into albedo, geometry and reflectance components. Early methods rely on model fitting and parametric techniques [2–4, 18], which are limited in capturing high frequency details not captured by these models, whereas more recent approaches employ learned based strategies [13, 19, 24, 31].

[1] Technically, uniform lighting will leave ambient occlusion in the recovered albedo, often desirable for downstream rendering tasks.

In particular, the method of Weir et al. [30] explicitly tackles the problem of *portrait de-lighting*. This approach relies on novel loss functions specifically targeting shadows and specular highlights. The final inferred result is an albedo image, which resembles the portrait as if it was lit from diffuse uniform illumination. Similarly, Yang et al. [32] propose an architecture to remove facial make-up, generating albedo images.

These methods, however completely remove the lighting from the current scene, whereas in photography applications one may simply want to control the effect of the current illumination, perhaps softening shadows and specular highlights. Along these lines the methods of Zhang et al. [37] and Inouei and Yamasaki [11] propose novel approaches to generate synthetic shadows that can be applied to in-the-wild images. Given these synthetically generated datasets, they propose a CNN based architecture to learn to remove shadows. The final systems are capable of removing harsh shadows while softening the overall look. Despite their impressive results, these approaches are designed to deal with shadow removal, and, although some softening effect can be obtained as byproduct of the method, their formulations ignore high order light transport effects such as specular highlights.

In contrast, we propose a novel learning based formulation to control the amount of light diffusion in portraits, without changing the overall scene illumination while softening or removing shadow and specular highlights.

3. A Framework for Light Diffusion

In this section, we formulate the light diffusion problem, and then propose a learning based solution for in-the-wild portraits. Finally we show how our model can be applied to infer a more robust albedo from images, improving downstream applications such as relighting, face part segmentation, and normal estimation.

3.1. Problem Formulation

We model formation of image I of a subject P in terms of illumination from a HDR environment map $E(\theta, \phi)$:

$$I = R[P, E(\theta, \phi)] \tag{1}$$

where $R[\cdot]$ renders the subject under the given HDR environment map. We can then model light diffusion as rendering the subject under a smoothed version of the HDR environment map. Concretely, a light-diffused image I_d is formed as:

$$I_d = R\left[P, E(\theta, \phi) * \frac{\cos_+^n(\theta)}{\sum_{i,j}^{H,W} \cos_+^n(\theta_{i,j})}\right] \tag{2}$$

where $*$ represents spherical convolution, and the incident HDR environment map is smoothed with normalized kernel $\cos_+^n(\theta) \equiv \max(0, \cos^n(\theta))$, effectively pre-smoothing the HDR environment map with the Phong specular term. The exponent n controls the amount of blur or diffusion of the lighting. Setting n to 1 leads to a diffusely lit image, and higher specular exponents result in sharper shadows and specular effects, as seen in Figure 3.

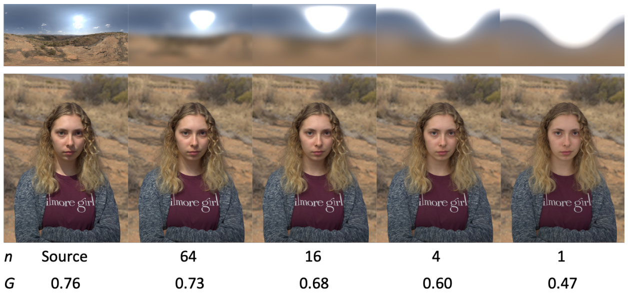

Figure 3. Illumination convolution. Shown are the original environment and relit image, followed by convolution with $\cos_+^n \theta$ with varying exponent n, and the resulting Gini coefficient G for each diffused environment. Note the gradual reduction in light harshness while still maintaining the overall lighting tone.

Our goal then is to construct a function f that takes I and the amount of diffusion controlled by exponent n and predicts the resulting light-diffused image $I_d = f(I, n)$. In practice, as described in section 3.2.1, we replace n with a parameter t that proved to be easier for the networks to learn; this new parameter is based on a novel application of the Gini coefficient [8] to measure environment map diffuseness.

3.2. Learning-based Light Diffusion

We perform the light diffusion task in a deep learning framework. We can represent the mapping f as a deep network f_β:

$$I_d = f_\beta(I, t) \tag{3}$$

where β represents the parameters of the network. To supervise training, we capture subjects in a light stage and, using the OLAT images [6], synthetically render each subject under an HDRI environment $E(\theta, \phi)$ and a diffused version of that environment $E(\theta, \phi) * \cos_+^n(\theta)$, providing training pair I and I_d.

In practice, we obtain better results with a sequence of two networks. The first network estimates a specular map S, which represents image brightening, and a shadow map D, which represents image darkening, both relative to a fully diffusely lit ($n = 1$) image. Concretely, we generate the fully diffused image I_{diffuse} as described in Equation 2 with $n = 1$ and then define the shadow D and specular S maps as:

$$S = \max(\min(1 - I_{\text{diffuse}}/I, 1), 0) \tag{4}$$
$$D = \max(\min(1 - I/I_{\text{diffuse}}, 1), 0) \tag{5}$$

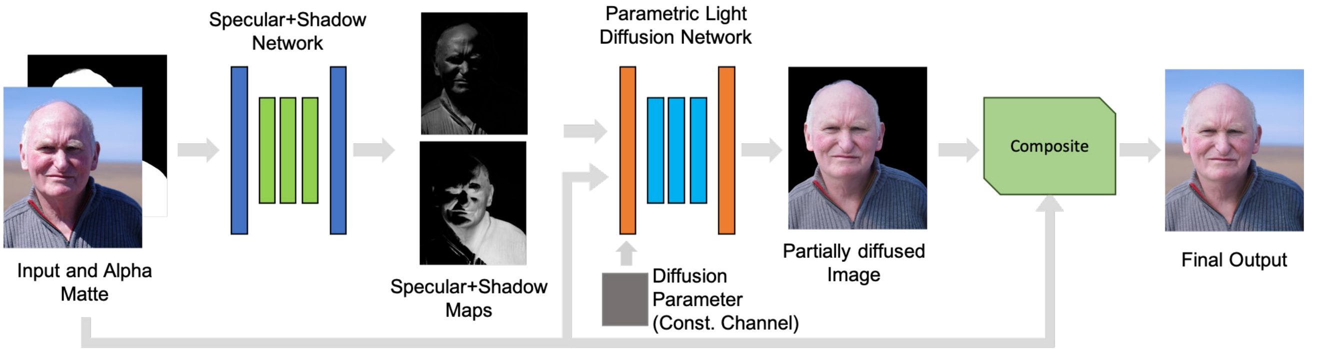

Figure 4. Architecture for parametric diffusion. Taking a portrait image with an alpha matte, the first stage predicts specular and shadow maps. The second stage uses these maps and the source image to produce an image with light diffused according to an input diffusion parameter. The result is composited over the input image to replace the foreground subject with the newly lit version.

Given the light stage data, it is easy to additionally synthesize I_{diffuse} and compute S and D for a given HDR environment map E to supervise training of a shadow-specular network g_{β_s}:

$$\{S, D\} = g_{\beta_s}(I) \tag{6}$$

The light diffusion network then maps the input image along with the specular and shadow maps to the final result:

$$I_d = h_{\beta_d}(I, S, D, t) \tag{7}$$

Note that, as we are not seeking to modify lighting of the background, we focus all the computation on the subject in the portrait. We thus estimate a matte for the foreground subject and feed it into the networks as well; I then is represented as the union of the original image and its portrait matte. The overall framework is shown in Figure 4.

In addition, we can extend our framework to infer a more robust albedo than prior work, through a process of repeated light diffusion. We now detail each of the individual components of the light diffusion and albedo estimation networks.

3.2.1 Network details

Specular+Shadow Network The specular+shadow network g_{β_s} is a single network that takes in the source image I along with a pre-computed alpha matte [23], as a $1024 \times 768 \times 4$ dimensional tensor. We used a U-Net [25] with 7 encoder-decoder layers and skip connections. Each layer used a 3×3 convolution followed by a Leaky ReLU activation. The number of filters for each layer is $24, 32, 64, 64, 64, 92, 128$ for the encoder, 128 for the bottleneck, and $128, 92, 64, 64, 64, 32, 24$ for the decoder. We used blur-pooling [36] layers for down-sampling and bilinear resizing followed by a 3×3 convolution for upsampling. The final output - two single channel maps - is generated by a 3×3 convolution with two filters.

Parametric Diffusion Net The diffusion network h_{β_d} takes the source image, alpha matte, specular map, shadow map, and the diffusion parameter t (as a constant channel) into the Diffusion Net as a $1024 \times 768 \times 7$ tensor. The Diffusion Net is a U-Net similar to the previous U-Net, with $48, 92, 128, 256, 256, 384, 512$ encoder filters, 512 bottleneck filters, and $384, 384, 256, 256, 128, 92, 48$ decoder filters. The larger filter count accounts for the additional difficulty of the diffusion task.

Diffusion parameter choice The diffusion parameter t indicates the amount of diffusion. While one can naively rely on specular exponents as a control parameter, we observed that directly using them led to poor and inconsistent results, as the perceptual change for the same specular convolution can be very varied for different HDR environments, for instance, a map with evenly distributed lighting will hardly change, whereas a map with a point light would change greatly. We hypothesize the non-linear nature of this operation is difficult for the model to learn, and so we quantified a different parameter based on a measure of 'absolute' diffuseness.

To measure the absolute diffusivity of an image, we observed that the degree of diffusion is related to how evenly distributed the lighting environment is, which strongly depends on the specific scene; e.g., if all the lighting comes from a single, bright source, we will tend to have harsh shadows and strong specular effects, but if the environment has many large area lights, the image will have soft shadows and subdued specular effects. In other words, the diffusivity is related to the inequality of the lighting environment. Thus, we propose to quantify the amount of diffusion by using the *Gini coefficient* [8] of the lighting environment, which is designed to measure inequality. Empirically, we found that the Gini coefficient gives a normalized measurement of the distribution of the light in an HDR map, as seen in Figure 5, and thus we use it to control the amount of dif-

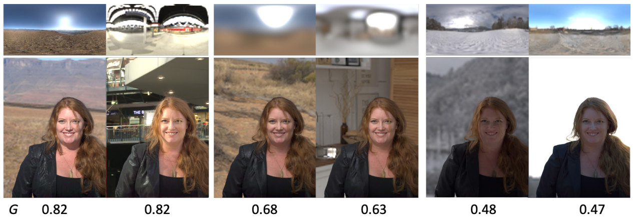

Figure 5. Gini coefficients G of some HDR maps and their relit images. Similar Gini coefficients approximately yield a similar quality of lighting, allowing a consistent measure of diffusion.

fusion.

Mathematically, for a finite multiset $X \subset \mathbb{R}^+$, where $|X| = k$, the Gini coefficient, $G \in [0, 1]$, is computed as

$$G = \frac{\sum_{x_i, x_j \in X} |x_i - x_j|}{2k \sum_{x_i \in X} x_i}. \tag{8}$$

For a discrete HDR environment map, we compute the Gini coefficient by setting each $x_i \in X$ to be the luminance from the i^{th} sample of the HDR environment map. For instance, on a discrete equirectangular projection $E(\theta, \phi)$ where $(\theta, \phi) \in [0, \pi] \times [0, 2\pi)$, the i^{th} sample's light contribution is given by $E(\theta_i, \phi_i) \sin(\theta_i)$, where $\sin(\theta_i)$ compensates for higher sampling density at the poles. If we indicate the i^{th} sample of E by E_i, the coefficient is then given by

$$G = \frac{\sum_i \sum_j |E_i \sin(\theta_i) - E_j \sin(\theta_j)|}{2k \sum_i E_i \sin(\theta_i)} \tag{9}$$

where i, j range over all samples of the equirectangular map and k is the total number of samples in the map.

Finally, as an input parameter, we re-scaled this absolute measure based on each training example: $t = (G_t - G_d)/(G_s - G_d)$, where G_t is the Gini coefficient for the target image, G_d is the Gini coefficient for the fully diffused image (diffused with specular exponent 1), and G_s is the Gini coefficient for the source image. Parameter t ranges from 0 to 1, where 0 corresponds to maximally diffuse ($G_t = G_d$) and 1 corresponds to no diffusion ($G_t = G_s$).

3.2.2 Albedo Estimation

We observed that the primary source of errors in albedo estimation in state-of-the-art approaches like [23] arises from color and material ambiguities in clothing and is exacerbated by shadows. The albedo estimation stage tends to be the quality bottleneck in image relighting, as errors are propagated forward in such multistage pipelines. Motivated by this observation, we propose to adapt our light diffusion approach to albedo estimation (Figure 6).

While the fully diffuse image (diffused with $n = 1$) removes most shading effects, the approach can be pushed further to estimate an image only lit by the average color of the HDRI map, i.e., a *tinted albedo*. Since the diffuse convolution operation preserves the average illumination of the HDR environment map and acts as a strong smoothing operation, repeated convolution converges to the average color of the HDR environment map. We found that iterating our diffusion network just three times (along with end-to-end training of the iteration based network) yielded good results. An alternative formulation of this problem is to pass the fully diffuse image into a separate network which estimates this tinted albedo, and we show a comparison between these two in the supplementary material.

To remove the color tint, we crop the face – which resides in the more constrained space of skin tone priors – and train a CNN to estimate the RGB tint of the environment, again supervised by light stage data. We then divide out this tint to recover the untinted albedo for the foreground.

3.3. Data Generation and Training Details

To train the proposed model, we require supervised pairs of input portraits I and output diffused images I_d. Following previous work [23], we rely on a light stage [10, 20] to capture the full reflectance field of multiple subjects as well as their geometry. Our data generation stage consists of generation of images with varying levels of diffusion as well as the tinted and true albedo maps, to use as ground truth to train our model.

Importantly, we also propose a synthetic shadow augmentation strategy to add external shadow with subsurface scattering effects that are not easily modeled in relit images generated in the light stage. We extend the method proposed by [37] to follow the 3D geometry of the scene, by placing a virtual cylinder around the subject with a silhouette mapped to the surface. We then project the silhouette over the 3D surface of the subject – reconstructed from the light stage dataset – from the strongest light in the scene followed by blurring and opacity adjustment of the resulting projected shadow map, guided by the Gini coefficient of the environment (smaller Ginis have more blur and lower shadow opacity). The resulting shadow map is used to blend between the original image and the image after removing the brightest light direction contribution. This shadow augmentation step is key to effective light diffusion and albedo estimation.

We also augmented with subsurface scattering effects, since the light stage dataset does not include hard shadows cast by foreign objects. We implemented a heuristic approach which uses the shadow map and a skin segmentation map to inpaint a red tint around shadow edges inside detected skin regions.

For more details on our training dataset, our approach for shadow augmentation, and specifics of model training and loss functions, please refer to our supplementary material.

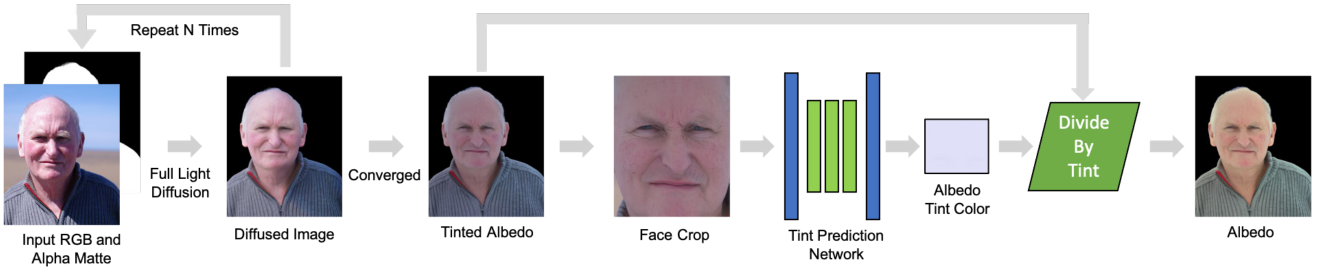

Figure 6. Architecture of our proposed extension of light diffusion to albedo prediction. We recurrently apply diffusion network N times to an image, yielding an albedo map tinted by the average color of scene light. We train a model to estimate this tint, then divide it out to produce an untinted albedo image. In this case, a warmer albedo color arises after removing the blue tint introduced by the sky illumination.

4. Experiments

In this section we experimentally verify the design choices of our architecture with qualitative and quantitative comparisons, and also show the effect of light diffusion on inputs for tasks like geometry estimation, semantic segmentation and portrait relighting.

4.1. Evaluation Datasets

We created two evaluation datasets, one from light stage images, where ground truth is available, and the other from a selection of in the wild images with harsh lighting conditions. Despite having no ground truth, qualitative results on the in-the-wild set are critical in evaluating the generalization ability of our model.

The light stage dataset consists of six subjects with diverse skin tones and genders lit by ten challenging lighting environments. The subjects as well as the HDRI maps are withheld from training and used to compute quantitative metrics against the ground truth.

The in-the-wild dataset consists of 282 diverse portrait images in various realistic lighting conditions that highlight the usefulness of the proposed Light Diffusion. We use this set to show qualitative results for ablation studies and comparison against the state-of-the-art.

4.2. Full Diffusion Results

First, we trained models that predicted fully diffuse results – the most difficult case – to compare various design choices of the proposed algorithm. Figure 7 shows a comparison among three architectures. From left to right these are, (1) A model trained to predict a fully diffuse image on data without any of our proposed external shadow augmentation techniques, (2) A large model with twice as many filters but without the shadow/specular maps prediction network, and (3) Our proposed architecture. Note that the model trained without shadow augmentation does significantly worse with shadow removal, and even a larger model trained on this data struggles with shadow edges, despite having a comparable number of parameters to our proposed approach. Table 1 shows quantitative metrics computed on the light stage data. The large U-Net (with no shadow/specular maps) does marginally better on average metrics, however, as shown in Figure 7, this does not hold on qualitative results on in-the-wild data, suggesting the increased parameter count caused overfitting to light stage images.

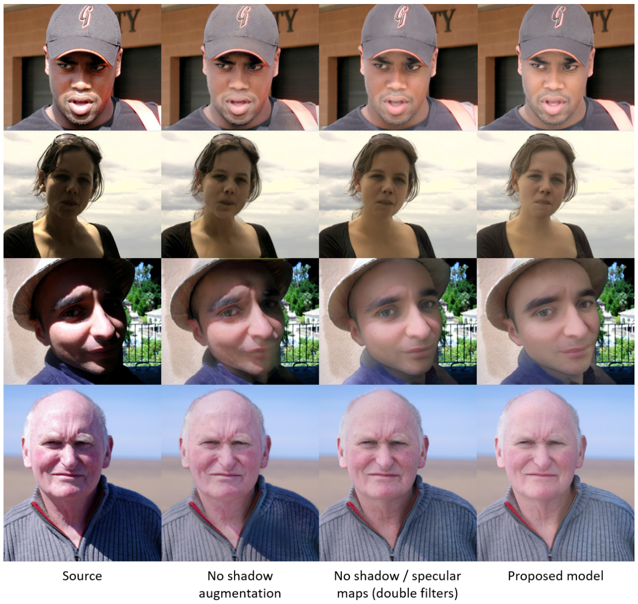

Figure 7. Fully diffuse ablation study. Left to right: source image, model with no shadow augmentation, model with twice as many filters but no shadow or specular map, proposed model.

4.3. In-the-Wild Applications

In this section, we show the results of our approach on a variety of applications. As byproduct, the proposed framework can be used for multiple computational photography experiences as well as to improve downstream computer vision tasks.

Model	MAE ↓	MSE ↓	SSIM ↑	LPIPS ↓
Proposed	0.0098	0.0006	0.9692	0.0340
No shadow aug.	0.0123	0.0009	0.9563	0.0503
Large model	0.0094	0.0005	0.9749	0.0304

Table 1. Quantitative metrics for fully diffuse prediction ablation study. Although the large model with no shadow/specular maps seems to do slightly better on average, results on in-the-wild data (Figure 7) suggests that it substantially overfits to the light stage data.

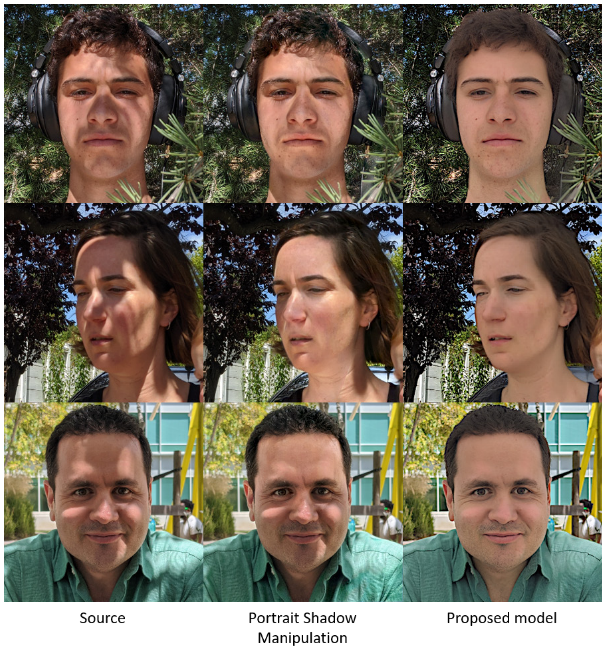

Figure 8. Comparison with [37]. Note how our method better softens portraits, removing shadows and specular highlights.

Shadow Removal. In Figure 8 we compare our fully diffuse output with that of the shadow removal approach proposed in [37]. Note that our model not only diffuses hard shadow edges better, but can also reduce harsh specular effects on the skin.

Albedo Prediction. In Figure 9 we compare our approach to state-of-the-art delighting approaches [23, 30, 34]. Note that all previous approaches suffer from artifacts on clothing due to color ambiguities or harsh shadows, whereas our proposed algorithm correctly removes shadows and produces accurate skin tones, with no artifacts on clothing. We also provide an ablation on albedo prediction architecture choices in our supplementary text.

Parametric Diffusion. In Figure 10 we also show the range of outputs our parametric diffusion model can produce. This is a critical feature of our proposed application,

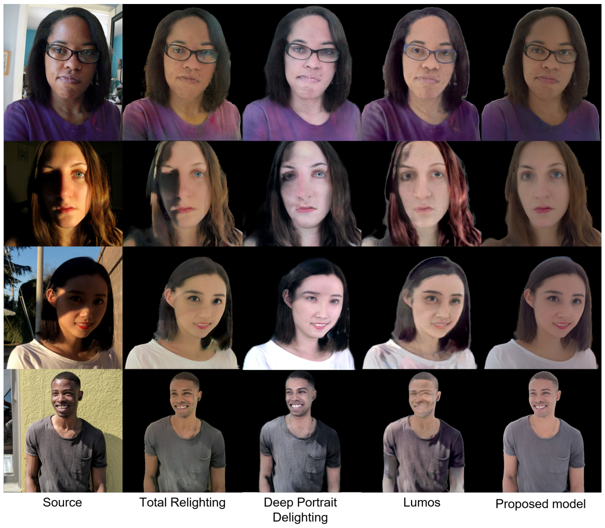

Figure 9. Albedo prediction comparisons against state of the art. Our approach has markedly better color stability, shadow removal, and skin tone preservation across subjects. From left to right: source, Total Relighting [23], Deep Portrait Delighting [30], Lumos [34], our model.

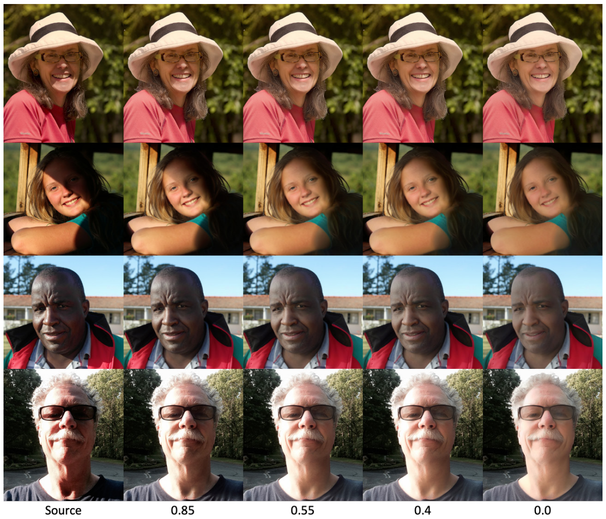

Figure 10. Parametric diffusion results. The model is able to gradually remove harsh shadows and specular effects, across skin tones, complex and deep shadows, and highly saturated images. Shown is the source image, followed by our model output at the shown diffusion parameters.

since the fully diffuse output may appear too "flat" for compositing back into the original scene. With such control scheme, a user can choose the level of diffusion according to the level of contrast / drama they might be going for in the portrait. Note that our model can produce an aesthetically pleasing range of diffusion levels, with speculars being gradually reduced and harsh shadow edges being realistically blurred. A naive interpolation approach between the source and fully diffuse output would leave behind unnatural shadow edges.

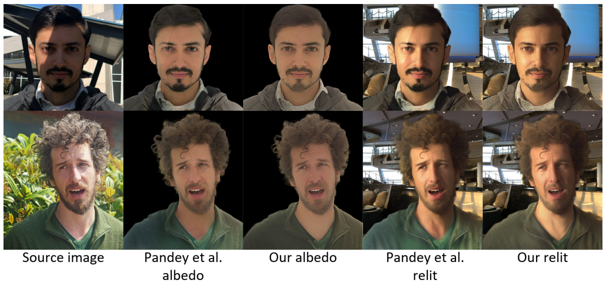

Figure 11. Example of replacing albedo prediction within Total Relighting [23] with our albedo. From left to right: input image, albedo estimated by Total Relighting, albedo estimated by our method, image relit by Total Relighting using original albedo and image relit by Total Relighting using our albedo. Our approach is notably better at dealing with external shadows (top row) and clothing discoloration (bottom row), resulting in more realistic relighting results.

Portrait Relighting. As mentioned in the previous sections, the albedo estimation stage tends to be the bottleneck in quality for state-of-the-art portrait relighting approaches like [23]. In Figure 9 we show that our albedo estimation approach is significantly more robust to artifacts that arise from color ambiguities and harsh shadows. In Figure 11 we show the effect of using our estimated albedo and feeding it into the relighting module of [23]. Note that the relit result quality greatly benefits from our albedo, and no longer shows artifacts on clothing or harsh shadow regions.

Other Applications. While controlling the amount of light diffusion in a portrait is a crucial feature itself for computational photography, it can also be used as preprocessing step to simplify other computer vision tasks. The importance of relighting as a data augmentation strategy has been demonstrated in various contexts [5, 28], and here we show that reducing the amount of unwanted shadows and specular highlights has a similar beneficial effect to downstream applications. In Figure 12 we show the effect of using a diffuse image instead of the original for an off-the-shelf normal map estimator [23] and semantic segmenter for face parsing [16]. In both cases, artifacts due to the external shadow are removed by using the fully diffuse input.

5. Discussion

We proposed a complete system for *light diffusion*, a novel method to control the lighting in portrait photography by reducing harsh shadows and specular effects. Our approach can be used directly to improve photographs and can aid numerous downstream tasks. In particular, we have shown that light diffusion generalizes well to albedo prediction, greatly improving on the state of the art. We have also shown that geometry estimation and semantic segmentation is improved, and we expect that the process should improve many other downstream portrait-based vision tasks.

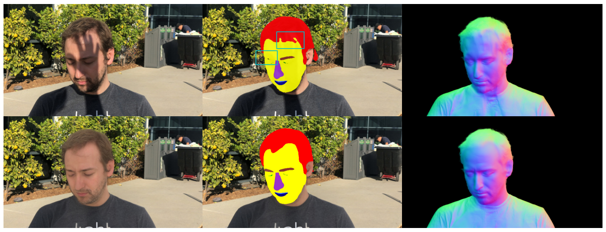

Figure 12. Light diffusion (bottom left) can improve results of state-of-the-art image processing methods, such as face parsing [16] (middle, improvement in green boxes) or normal map estimation (right, removal of shadow embossing).

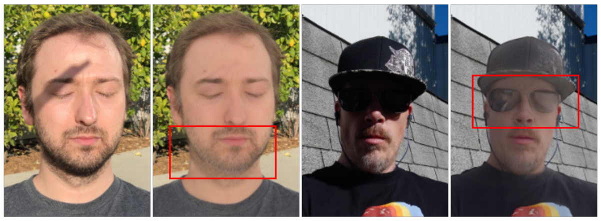

Figure 13. Limitations of our approach. After full diffusion, the model can sometimes lighten facial hair. The model also has trouble with dark sunglasses, tending to inpaint them with skin shades.

Limitations. Despite vastly increasing the domain of materials that can have lighting adjustments, our model has some limitations, as shown in Figure 13. In particular, we notice that dark facial hair tends to be lightened, perhaps due to its resemblance to a shadow region. In addition, our model has trouble with dark sunglasses, tending to add skin tones to them. Other limitations include over-blurring excessively diffused images where fine details should be synthesized instead, failure to remove very strong specularities and occasional confusion of objects for shadows.

Fairness. Our results have shown that our proposed approach works well across a variety of skin tones. To validate this, we also ran a detailed fairness study to analyze results across different Fitzpatrick skin tones. Please refer to our supplementary material for details.

Acknowledgements

David Futschik and Daniel Sýkora were partly supported by the Research Center for Informatics, grant No. CZ.02.1.01/0.0/0.0/16_019/0000765 and by the Grant Agency of the Czech Technical University in Prague, grant No. SGS22/173/OHK3/3T/13.

References

[1] Rameen Abdal, Peihao Zhu, Niloy J. Mitra, and Peter Wonka. Styleflow: Attribute-conditioned exploration of stylegan-generated images using conditional continuous normalizing flows. *ACM Transactions on Graphics*, 2021. 2

[2] Abdelrehim Ahmed and Aly Farag. A new statistical model combining shape and spherical harmonics illumination for face reconstruction. In *Advances in Visual Computing*, 2007. 2

[3] Jonathan T. Barron and Jitendra Malik. Shape, Illumination, and Reflectance From Shading. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015. 2

[4] Volker Blanz and Thomas Vetter. A morphable model for the synthesis of 3d faces. In *SIGGRAPH*, 1999. 2

[5] George Chogovadze, Rémi Pautrat, and Marc Pollefeys. Controllable data augmentation through deep relighting, 2021. 8

[6] Paul Debevec, Tim Hawkins, Chris Tchou, Haarm-Pieter Duiker, Westley Sarokin, and Mark Sagar. Acquiring the reflectance field of a human face. SIGGRAPH '00, page 145–156, USA, 2000. ACM Press/Addison-Wesley Publishing Co. 3

[7] Yu Deng, Jiaolong Yang, Dong Chen, Fang Wen, and Xin Tong. Disentangled and controllable face image generation via 3d imitative-contrastive learning. In *IEEE Computer Vision and Pattern Recognition*, 2020. 2

[8] Corrado Gini. *Variabilità e mutabilità: contributo allo studio delle distribuzioni e delle relazioni statistiche.[Fasc. I.]*. Tipogr. di P. Cuppini, 1912. 3, 4

[9] Christopher Grey. Master lighting guide for portrait photographers. In *Amherst Media*, 2004. 1

[10] Kaiwen Guo, Peter Lincoln, Philip Davidson, Jay Busch, Xueming Yu, Matt Whalen, Geoff Harvey, Sergio Orts-Escolano, Rohit Pandey, Jason Dourgarian, Danhang Tang, Anastasia Tkach, Adarsh Kowdle, Emily Cooper, Mingsong Dou, Sean Fanello, Graham Fyffe, Christoph Rhemann, Jonathan Taylor, Paul Debevec, and Shahram Izadi. The relightables: Volumetric performance capture of humans with realistic relighting. *ACM Trans. Graph.*, 38(6), nov 2019. 5

[11] Naoto Inoue and Toshihiko Yamasaki. Learning from synthetic shadows for shadow detection and removal. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021. 2, 3

[12] Chaonan Ji, Tao Yu, Kaiwen Guo, Jingxin Liu, and Yebin Liu. Geometry-aware single-image full-body human relighting. 2022. 2

[13] Yoshihiro Kanamori and Yuki Endo. Relighting Humans: Occlusion-Aware Inverse Rendering for Full-Body Human Images. *ACM Transactions Graphics (Proc. SIGGRAPH Asia)*, 2018. 2

[14] Chloe LeGendre, Wan-Chun Ma, Graham Fyffe, John Flynn, Laurent Charbonnel, Jay Busch, and Paul E. Debevec. Deeplight: Learning illumination for unconstrained mobile mixed reality. *CVPR*, 2019. 2

[15] Chloe LeGendre, Wan-Chun Ma, Rohit Pandey, Sean Fanello, Christoph Rhemann, Jason Dourgarian, Jay Busch, and Paul Debevec. Learning illumination from diverse portraits. In *SIGGRAPH Asia 2020 Technical Communications*, 2020. 2

[16] Yiming Lin, Jie Shen, Yujiang Wang, and Maja Pantic. Roi tanh-polar transformer network for face parsing in the wild. *Image and Vision Computing*, 112:104190, 2021. 8

[17] B R Mallikarjun, Ayush Tewari, Abdallah Dib, Tim Weyrich, Bernd Bickel, Hans-Peter Seidel, Hanspeter Pfister, Wojciech Matusik, Louis Chevallier, Mohamed Elgharib, et al. Photoapp: Photorealistic appearance editing of head portraits. *ACM Transactions on Graphics*, 40(4):1–16, 2021. 2

[18] Abhimitra Meka, Gereon Fox, Michael Zollhöfer, Christian Richardt, and Christian Theobalt. Live User-Guided Intrinsic Video for Static Scene. *IEEE Transactions on Visualization and Computer Graphics*, 2017. 2

[19] Abhimitra Meka, Maxim Maximov, Michael Zollhoefer, Avishek Chatterjee, Hans-Peter Seidel, Christian Richardt, and Christian Theobalt. LIME: Live Intrinsic Material Estimation. In *Proc. Computer Vision and Pattern Recognition*, 2018. 2

[20] Abhimitra Meka, Rohit Pandey, Christian Häne, Sergio Orts-Escolano, Peter Barnum, Philip David-Son, Daniel Erickson, Yinda Zhang, Jonathan Taylor, Sofien Bouaziz, Chloe LeGendre, Wan-Chun Ma, Ryan Overbeck, Thabo Beeler, Paul Debevec, Shahram Izadi, Christian Theobalt, Christoph Rhemann, and Sean Fanello. Deep relightable textures: Volumetric performance capture with neural rendering. *ACM Transactions on Graphics*, 2020. 5

[21] Thomas Nestmeyer, Jean-François Lalonde, Iain Matthews, and Andreas M. Lehrmann. Learning physics-guided face relighting under directional light. In *CVPR*, 2020. 2

[22] Xingang Pan, Xudong Xu, Chen Change Loy, Christian Theobalt, and Bo Dai. A shading-guided generative implicit model for shape-accurate 3d-aware image synthesis. In *NeurIPS*, 2021. 2

[23] Rohit Pandey, Sergio Orts Escolano, Chloe Legendre, Christian Haene, Sofien Bouaziz, Christoph Rhemann, Paul Debevec, and Sean Fanello. Total relighting: learning to relight portraits for background replacement. *ACM Transactions on Graphics (TOG)*, 40(4):1–21, 2021. 2, 4, 5, 7, 8

[24] Peiran Ren, Yue Dong, Stephen Lin, Xin Tong, and Baining Guo. Image Based Relighting Using Neural Networks. *ACM Transactions on Graphics*, 2015. 2

[25] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing. 2, 4

[26] Zhixin Shu, Sunil Hadap, Eli Shechtman, Kalyan Sunkavalli, Sylvain Paris, and Dimitris Samaras. Portrait lighting transfer using a mass transport approach. *ACM Transactions on Graphics (TOG)*, 36(4):1, 2017. 2

[27] Tiancheng Sun, Jonathan T Barron, Yun-Ta Tsai, Zexiang Xu, Xueming Yu, Graham Fyffe, Christoph Rhemann, Jay Busch, Paul Debevec, and Ravi Ramamoorthi. Single image

portrait relighting. *ACM Transactions on Graphics (TOG)*, 38(4):79, 2019. 2

[28] Feitong Tan, Sean Fanello, Abhimitra Meka, Sergio Orts-Escolano, Danhang Tang, Rohit Pandey, Jonathan Taylor, Ping Tan, and Yinda Zhang. Volux-gan: A generative model for 3d face synthesis with hdri relighting. *ACM SIGGRAPH*, 2022. 2, 8

[29] Zhibo Wang, Xin Yu, Ming Lu, Quan Wang, Chen Qian, and Feng Xu. Single image portrait relighting via explicit multiple reflectance channel modeling. *ACM SIGGRAPH Asia and Transactions on Graphics*, 2020. 2

[30] Joshua Weir, Junhong Zhao, Andrew Chalmers, and Taehyun Rhee. Deep portrait delighting. *ECCV*, 2022. 2, 3, 7

[31] Zexiang Xu, Kalyan Sunkavalli, Sunil Hadap, and Ravi Ramamoorthi. Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics*, 2018. 2

[32] Xingchao Yang and Takafumi Taketomi. BareSkinNet: Demakeup and De-lighting via 3D Face Reconstruction. *Computer Graphics Forum*, 2022. 2, 3

[33] Yu-Ying Yeh, Koki Nagano, Sameh Khamis, Jan Kautz, Ming-Yu Liu, and Ting-Chun Wang. Learning to relight portrait images via a virtual light stage and synthetic-to-real adaptation. *ACM Transactions on Graphics (TOG)*, 2022. 2

[34] Yu-Ying Yeh, Koki Nagano, Sameh Khamis, Jan Kautz, Ming-Yu Liu, and Ting-Chun Wang. Learning to relight portrait images via a virtual light stage and synthetic-to-real adaptation. *ACM Transactions on Graphics (TOG)*, 2022. 7

[35] Longwen Zhang, Qixuan Zhang, Minye Wu, Jingyi Yu, and Lan Xu. Neural video portrait relighting in real-time via consistency modeling. *CoRR*, 2021. 2

[36] Richard Zhang. Making convolutional networks shift-invariant again. In *ICML*, 2019. 4

[37] Xuaner Zhang, Jonathan T. Barron, Yun-Ta Tsai, Rohit Pandey, Xiuming Zhang, Ren Ng, and David E. Jacobs. Portrait shadow manipulation. volume 39, 2020. 2, 3, 5, 7

[38] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David Jacobs. Deep single image portrait relighting. In *ICCV*, 2019. 2