

# Microsoft

### Abstract

We present a CNN-based method for outdoor high-dynamic-range (HDR) environment map prediction from low-dynamic-range (LDR) portrait images. Our method relies on two different CNN architectures, one for light encoding and another for face-to-light prediction. Outdoor lighting is characterised by an extremely high dynamic range, and thus our encoding splits the environment map data between low and high intensity components, and encodes them using tailored representations. The combination of both network architectures constitutes an endto-end method for accurate HDR light prediction from faces at real-time rates, inaccessible for previous methods which focused on low dynamic range lighting or relied on non-linear optimisation schemes. We train our networks using both real and synthetic images, we compare our light encoding with other methods for light representation, and we analyse our results for light prediction on real images. We show that our predicted HDR environment maps can be used as accurate illumination sources for scene renderings, with potential applications in 3D object insertion for augmented reality.

#### Motivation

Fast inference of lighting from a portrait allows for realistic mixed-reality applications, which can be leveraged for real-time scene manipulation in video-conference software and general augmented reality tasks, such as 3D object insertion, face relighting, highlight extraction, background editing and shadow removal.





3D Object Insertion [S. B. Knorr et al. 2014] Real-tim nination estimation from faces for coherent rendering.

**Face Relighting** [H. Zhou et al. 2019] Deep single-image portrait relighting



**Shadow Removal** [X. Zhang et al. 2020] Portrait Shadow Manipulation



**Highlight Extraction** [C. Li et al. 2017] Specular Highlight Removal in Facial Images.

## High-Dynamic-Range Lighting Estimation From Face Portraits

Alejandro Sztrajman<sup>1</sup> Alexandros Neophytou<sup>2</sup> Tim Weyrich<sup>1</sup>

<sup>1</sup>University College London

### Light Encoding

Our method for outdoor HDR environment map representation splits the light data between low and high intensity components, and uses different encodings for each of them:

- The low intensity part, comprising a dense and low contrast image, contributes a uniform illumination to the scene. It is encoded using EnvNet, an autoencoder architecture for environment maps.
- The high intensity part, sparse and with high contrast, contains the sun's corona, and produces a directional illumination on the scene. We encode itusing a non-linear optimisation with a 2D Gaussian prior.



Training of the EnvNet autoencoder is done with the Laval Outdoor HDR Dataset, which contains 205 high-dynamic-range environment maps of outdoor scenes from different geographical settings, weather conditions and times of the day. Environment maps are augmented by applying random rotations of the azimuth.



We evaluate our encodings in terms of lighting and shading metrics, comparing against spherical harmonics of different degrees, and with EnvNet alone without low/high splitting + Gaussian fit:

- Lighting: direct comparison of the GT and predicted environment maps.
- Shading: comparison of scenes rendered using GT and predicted environment maps as illumination source.

Eric Sommerlade<sup>2</sup>

<sup>2</sup>Microsoft

Light prediction is performed by a second CNN architecture, which takes the luminance of a portrait image as input, and outputs embeddings for predicted low and high intensity components of illumination. The predicted environment map is then reconstructed by evaluating the low embedding Z<sub>Low</sub> through the EnvNet autoencoder, and the high embedding Z<sub>High</sub> through a Gaussian function.

![](_page_0_Figure_33.jpeg)

![](_page_0_Figure_34.jpeg)

Training of the light prediction network is done with:

- Real portrait images and environment maps from the Laval Face & Lighting HDR Dataset.
- Synthetic data generated by combining scanned faces from the ICT 3D Relightable Facial Expression Database, with azimuth augmented environment maps from the Laval Face & Lighting HDR Dataset.

We evaluate light prediction on a test set of the Laval Face & Lighting HDR Dataset, and compare with an optimisation-based approach by Calian et al.

![](_page_0_Figure_39.jpeg)

Calian et al.	0.031	0.89	0.08
Table: Light pro	ediction com	parison, usir	ng pixel-t

EnvMap High (Enc.) EnvMap Low (Enc.)

> Table: Results taken over 41 environment maps from the Laval Outdoor HDR Dataset (test set). For Lighting metrics we include a geometric factor to account for spherical coordi-

![](_page_0_Picture_44.jpeg)

This project has received funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 642841.

![](_page_0_Picture_46.jpeg)

### **Light Prediction**

![](_page_0_Picture_50.jpeg)

![](_page_0_Picture_51.jpeg)

Durs)	Pred Envma (Calian et al.)	p GT Render	Pred (Ours)	Pred (Calian et al.)
		S	F	B
		:		
			Z	
			30	35
			B	
S (s)	RMSE-dw (l)	Sun Altitude (l)	Sun Azimuth (l)	
29	0.15	0.12	0.48	
86	0.13	0.41	0.12	

-to-pixel and perceptually-based metrics. Comparison with [D. Calian et al. 2018] From Faces to Outdoor Light Probes.