High-Dynamic-Range Lighting Estimation From Face Portraits

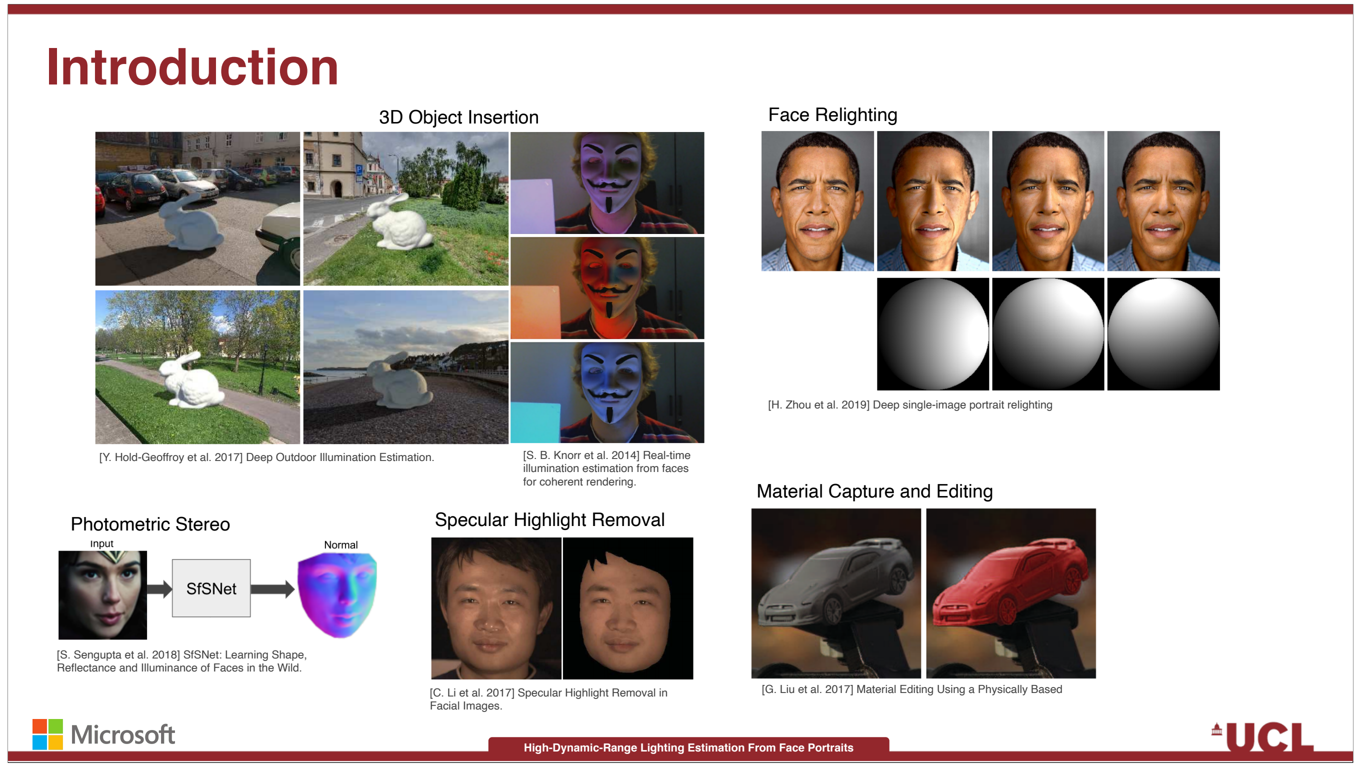
Alejandro Sztrajman[1] Alexandros Neophytou[2] Tim Weyrich[1] Eric Sommerlade[2]

International Conference on 3D Vision, November 2020

[1]University College London [2]Microsoft

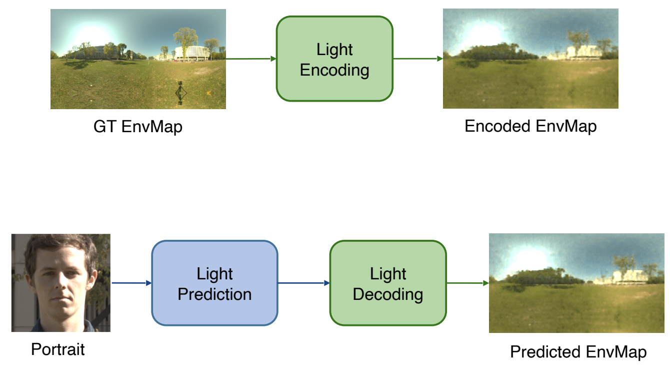
Microsoft UCL

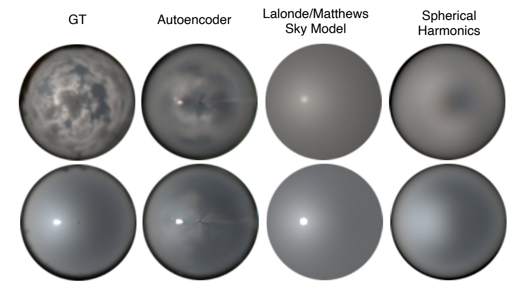
* We present High-Dynamic-Range Lighting Estimation From Face Portraits, a joint collaboration by University College London and Microsoft, presented at the 8th International Conference on 3D Vision.



* Illumination is one of the main components in the process of image formation.
<Show>
* Its fast estimation from images has many direct applications, such as 3D object insertion and scene re-illumination,
which can be leveraged for real-time scene manipulation in video-conference software and general augmented reality applications.
<Show>
* In addition, an accurate estimation of light can be used to capture and edit other components of the image formation process.
* This leads to a great number of potential applications, such as normal estimation, specular highlight removal, shadow removal, and material capture and editing.

Introduction



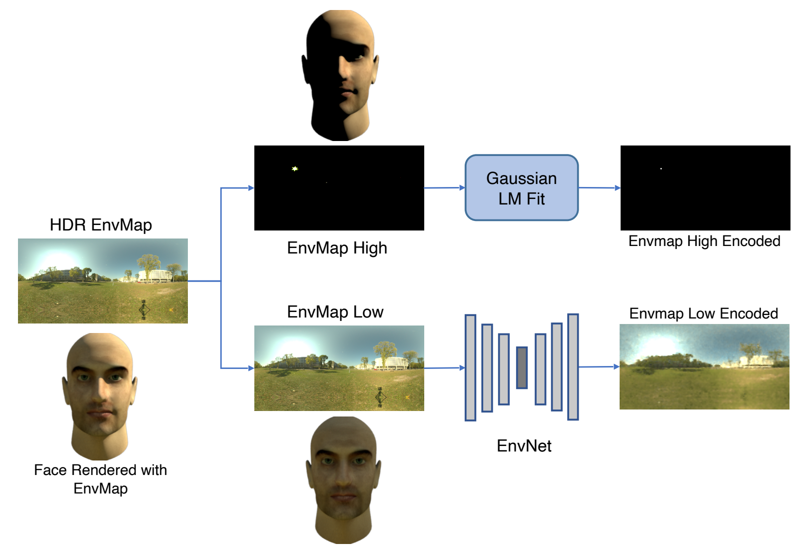


[Calian et al. 2018] From faces to outdoor light probes.

* Our approach for light prediction from portraits is based on two CNN architectures: one for environment map encoding/decoding,
<Show>
and another for environment map prediction from portraits.
<Show>
* Previous methods for light estimation represented the illumination by either using an autoencoder,
or a parametric model, such as Spherical Harmonics, or tailored models such as the Lalonde/Matthews and Hosek/Wilkie models.

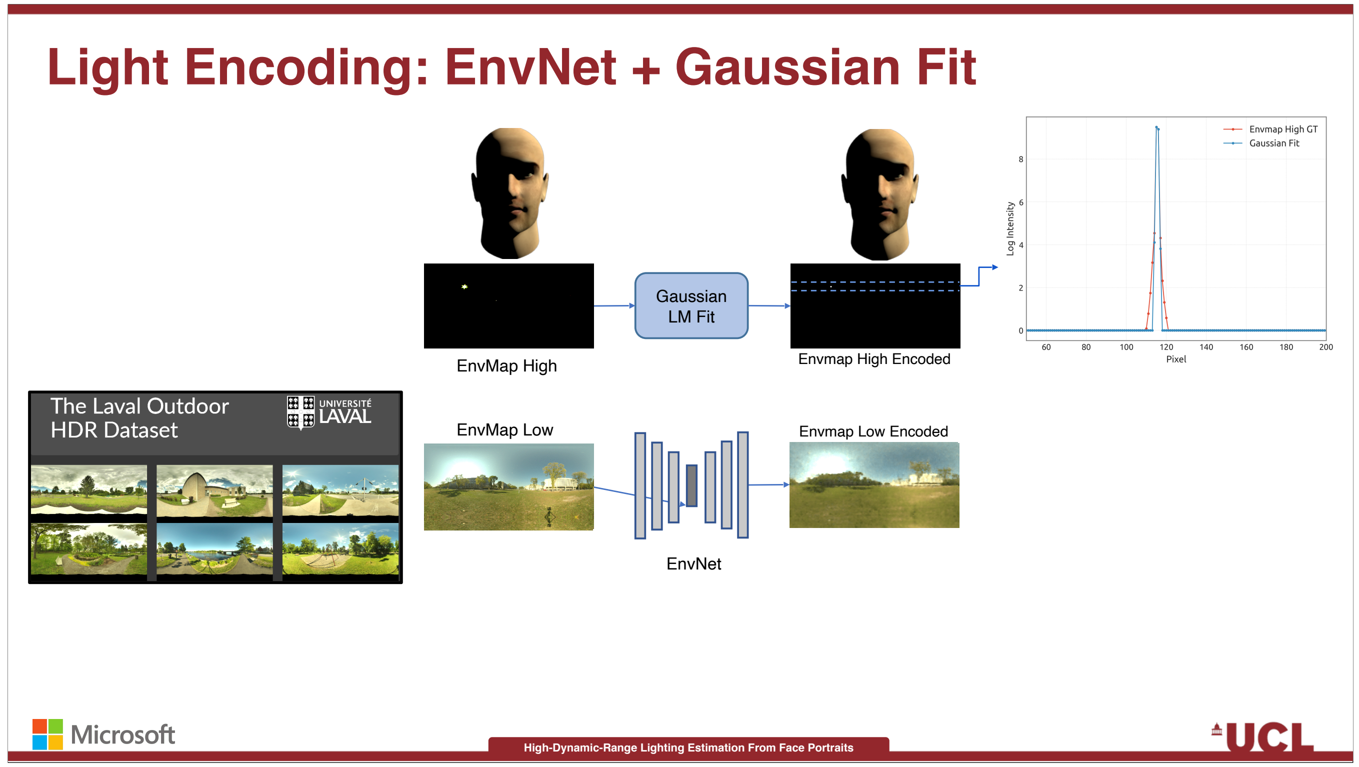
* Autoencoders are prone to oversmooth the ground truth data, which makes them suboptimal for the extremely high-dynamic-range of outdoor illumination.
* Parametric models, on the other hand, are not flexible in terms of representing different scenes, and usually focus only on the upper-hemisphere corresponding to the sky illumination.

Light Encoding: EnvNet + Gaussian Fit



* Our approach for light prediction uses a hybrid encoding, which combines data-driven and parametric-based representations.
<Show>
* The encoding of light is based on the observation that outdoor illumination has two very distinctive components, which contribute differently to the illumination of a scene, and which are fit for different representations.
<Show>
* Most of the environment map is formed by low intensity light, which provides a uniform illumination on the scene.
<Show>
* The sun disc, on the other hand, spreads over only a few pixels, but it’s extremely intense, comprising a crucial source of directional light.

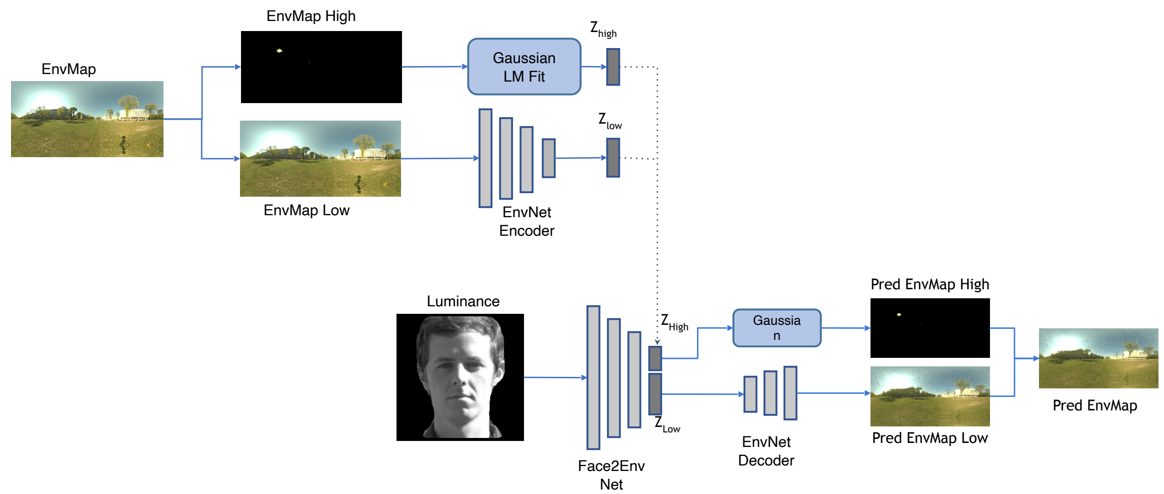
* We thus split light maps into high and low intensity components and use different encodings for each part.
<Show>
* The low intensity part, comprising a dense and low contrast image, is encoded using EnvNet, an autoencoder architecture for environment maps.
<Show>
* The high intensity part, sparse and with high contrast, is encoded using 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.
* Captured scenes vary in terms of geographical setting, weather conditions and time of the day.
<Show>
* Environment maps are further augmented by applying random rotations of the azimuth.
<Show>

* For the high part of the illumination we fit a Gaussian function with a fixed variance obtained empirically.
* The optimisation fits the amplitude and centre of the Gaussian function.
<Show>
We find that this produces a good fitting, especially of the most intense pixels of the environment map, and leads to a good reproduction of the directional light.

Light Prediction



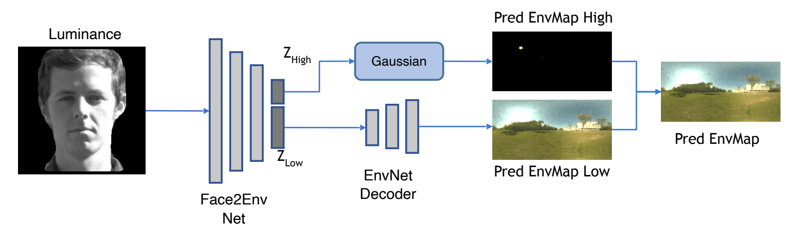
* After training EnvNet, we can generate compact encodings of the low and high parts of an environment map, by evaluating each part through the corresponding encoding method.

<Show>
* Light prediction is performed by a second CNN architecture, which takes the luminance of a portrait image as input.
* The network outputs embeddings for predicted high and low intensity components of illumination.
<Show>
* The predicted environment map is then reconstructed by evaluating the embeddings through the EnvNet decoder and the gaussian function.

Light Prediction

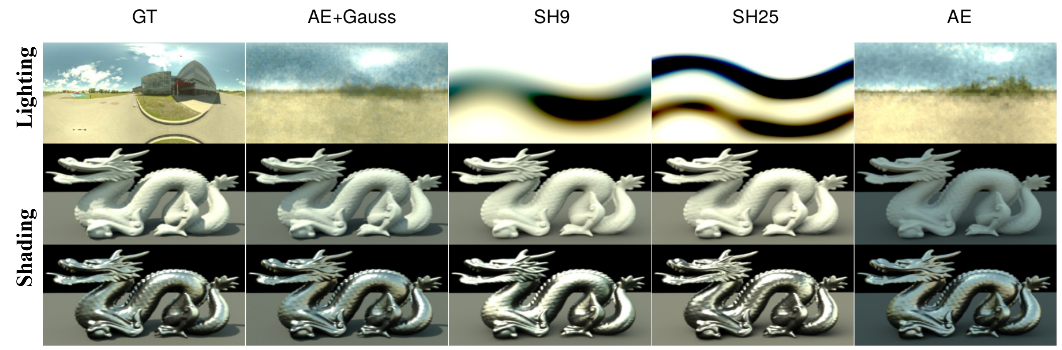
Laval Face+Lighting Dataset (Real)

ICT-3DRFE Dataset (Synth)



* To train the light prediction network we use real portrait images and environment maps from the Laval Face & Lighting HDR Dataset,
<Show>
* and synthetic data from renderings.
* These are 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.

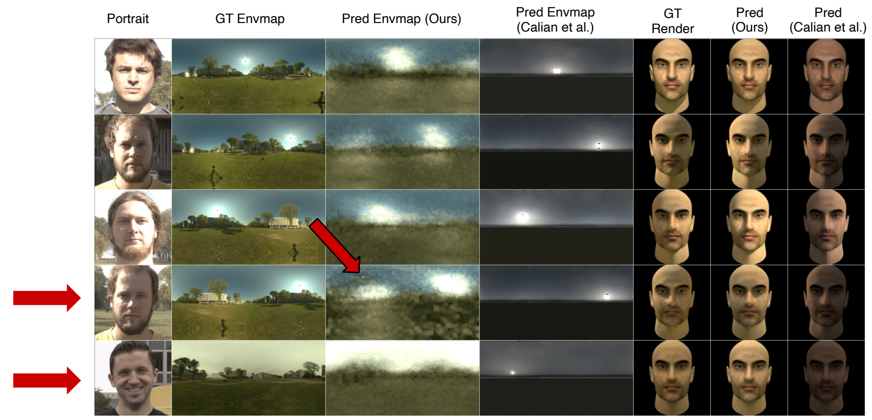
Results: Light Encoding



	Shading		Lighting	
	MAE	RMSE	MAE–dω	RMSE–dω
SH9	0.030 ± 0.024	0.031 ± 0.025	0.20 ± 0.09	0.21 ± 0.09
SH25	0.025 ± 0.024	0.026 ± 0.022	0.21 ± 0.11	0.21 ± 0.11
AE	0.032 ± 0.024	0.033 ± 0.024	0.12 ± 0.04	0.12 ± 0.04
AE+Gauss	$\mathbf{0.022 \pm 0.015}$	$\mathbf{0.023 \pm 0.015}$	$\mathbf{0.08 \pm 0.03}$	$\mathbf{0.08 \pm 0.03}$

* We show results for light encoding, both in terms of lighting metrics, which directly compare environment maps,
<Show Figure>
* and shading metrics, which compare renderings generated with the predicted environment maps as illumination sources.
* For the shading metrics computation we use rendered scenes with purely diffuse and specular materials.
<Show Figure>
* We compare our representation with spherical harmonics of different degrees, and with an autoencoder alone, without splitting the environment map between low and high intensities.
<Show Figure>
* We summarise the encoding results, taken over 41 previously unseen environment maps from the Laval Outdoor HDR Dataset.
* Our representation presents smaller errors, both in terms of evaluated lighting and shading metrics.
* A qualitative inspection of the dragon renderings shows that our encoding is better at preserving shadows, due to the accurate representation of the sun's directional light.

Results: Light Prediction

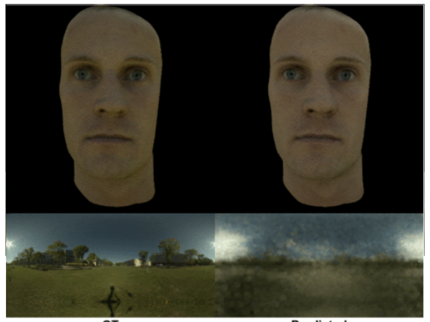


	RMSE (s)	SSIM (s)	LPIPS (s)	RMSE-dw (l)	Sun Altitude (l)	Sun Azimuth (l)
Our method	**0.017**	**0.95**	**0.029**	0.15	**0.12**	0.48
Calian et al.	0.031	0.89	0.086	**0.13**	0.41	**0.12**

[D. Calian et al. 2018] From Faces to Outdoor Light Probes.

* For light prediction we evaluate on a test set of the Laval Face & Lighting HDR Dataset, and compare with an optimisation-based approach by Calian et al.
* Once again we utilise lighting and shading metrics for evaluation.
* For shading metrics we include both pixel-to-pixel and perceptually-based errors.
* In the case of lighting metrics, we compare ground truth and predicted environment maps, adding a geometric factor to account for the spherical coordinates mapping.
* We additionally measure the average error, in radians, of the estimation of the sun's altitude and azimuth.
* Calian's method shows superior accuracy in the prediction of the sun's azimuth, while our method does it for the sun's altitude, which is usually placed too close to the horizon line in Calian's method.
* Moreover, our method shows a clear advantage in terms of shading metrics, both point-to-point and perceptually-based, as shown in the illuminated rendered faces.
<Show>
* In row 4 we show an example of a failure case.
* Our method is able to estimate the position of only one high-intensity light source, and the prediction is expected to degrade in cases with additional sources.
<Show>
* This can be the case of environment maps with highlights on windows, or buildings reflecting the sun.
<Show>
* The method does, however, perform well when the sun is not visible, due to overcast weather or being occluded by a building, as shown in row 5.

* As future work we hope to extend the method to predict multiple sources of light.
* This would address current issues and additionally enable the estimation of both indoor and outdoor scenes.



High-Dynamic-Range Lighting Estimation From Face Portraits

Alejandro Sztrajman, Alexandros Neophytou, Tim Weyrich, Eric Sommerlade

Microsoft

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.

UCL