

# OutCast: Outdoor Single-image Relighting with Cast Shadows Supplementary Material

<sup>1,2</sup>David Griffiths    <sup>2</sup>Tobias Ritschel    <sup>1</sup>Julien Philip

<sup>1</sup>Adobe Research    <sup>2</sup>University College London

**Table 1:** Relighting error for different loss related ablations across the test dataset according to different metrics. We also report the MSE for shadow prediction. For all metrics, less is better.

Method	Relight			Shadow
	DSSIM↓	LPIPS↓	MSE↓	MSE↓
MSE loss for target shadows (no LPIPS)	.171	<b>.0159</b>	.0437	<b>.150</b>
No PatchGAN	.162	.0190	.0405	.169
Our	<b>.154</b>	.0160	<b>.0399</b>	.175

## 1. Detailed Network Architecture

In Fig. 1 we present a detailed architecture blueprint of our network with all inputs, modules, networks, and outputs.

## 2. Training Data Examples

In Fig. 2, we present random training examples from our dataset. For each viewpoint we provide two lighting conditions and their respective ground truth shadows.

## 3. Real world ground truth evaluation

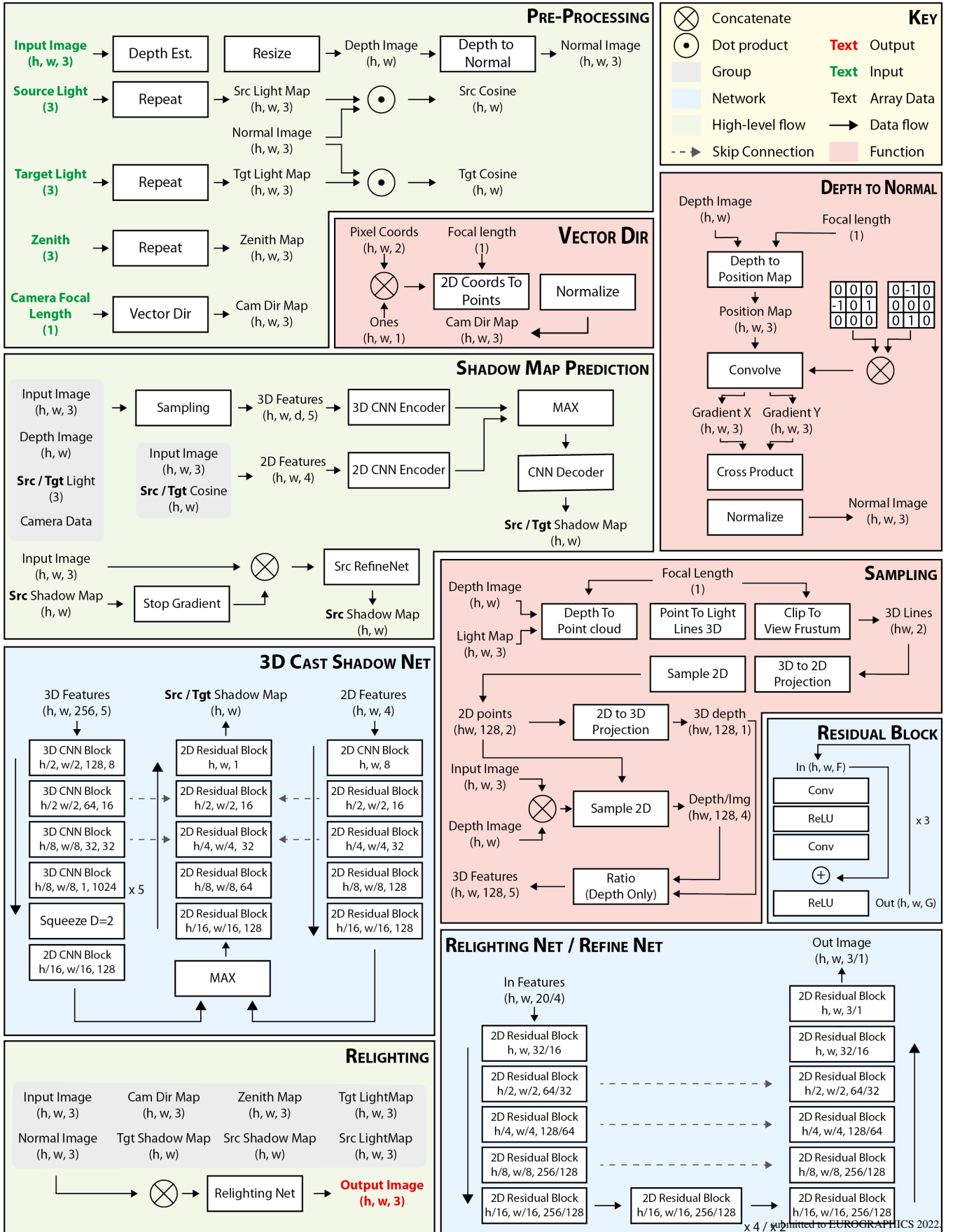
We evaluate our method on a real world lighting scenario. To enable this we utilize separate images taken roughly from the same viewpoint with different lighting conditions. As shown in Fig. 3 we observe that whilst the network output looks plausible, the shadows are misaligned from the target ground truth. We believe this is due to distortions in the predicted depth estimation for this scene.

## 4. Further Ablations

In addition to the ablations presented in the main paper (Sec. 5.1 and Sec. 5.2), we also undertake further ablations specifically evaluating the use of specific loss function components. Tbl. 1 provides the quantitative evaluation for these ablation. As expected, the MSE loss is smaller for shadows when used as the training metric. Overall the pipeline appears to perform marginally better when the PatchGAN loss term is used which seems surprising. While small variations might be due to the randomness of the training process, it is possible that the PatchGAN loss helps escape local minima leading to better convergence. Moreover, as shown in Fig. 4, this loss helps in producing complex localized effects such as high frequency reflections on tree leaves (Fig. 4 first row) or reflection on the water and better looking clouds (Fig. 4 second row). Training with E-LPIPS [KHL19] for shadows does not provide a strong advantage numerically but has a strong impact on the sharpness on elongated shadows as can be seen in Fig. 5. Without the E-LPIPS loss and with a more traditional MSE loss, the shadow network tends to produce overly smooth shadows as a slight misalignment of boundaries is strongly penalized.

## References

- [KHL19] KETTUNEN M., HÄRKÖNEN E., LEHTINEN J.: E-LPIPS: robust perceptual image similarity via random transformation ensembles. *arXiv:1906.03973* (2019).

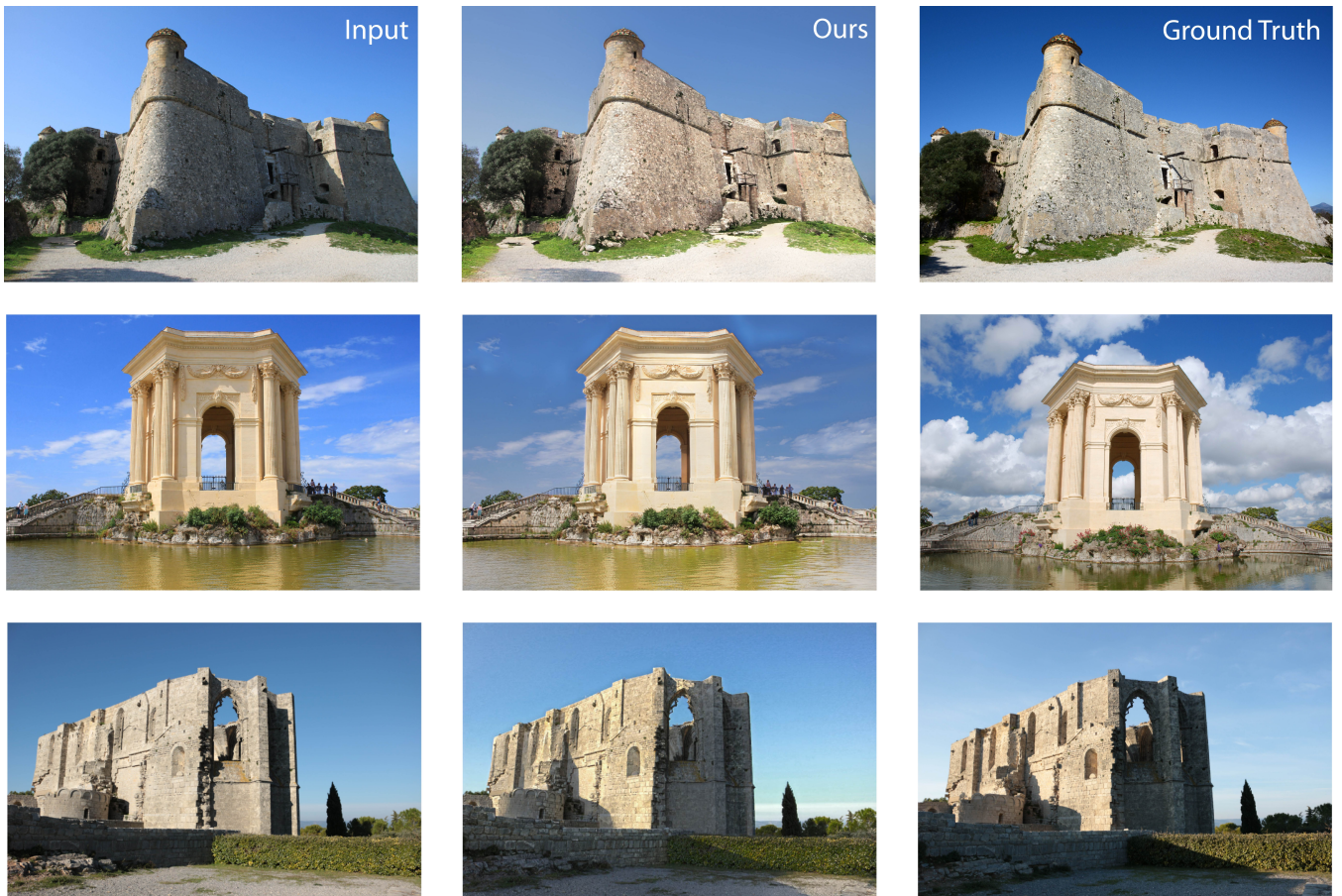


**Figure 1:** Blueprint of our full relighting system. Where arrows are not included assume flow remains in the current direction. All deep learning-based functions were implemented in the PyTorch framework (v1.9).

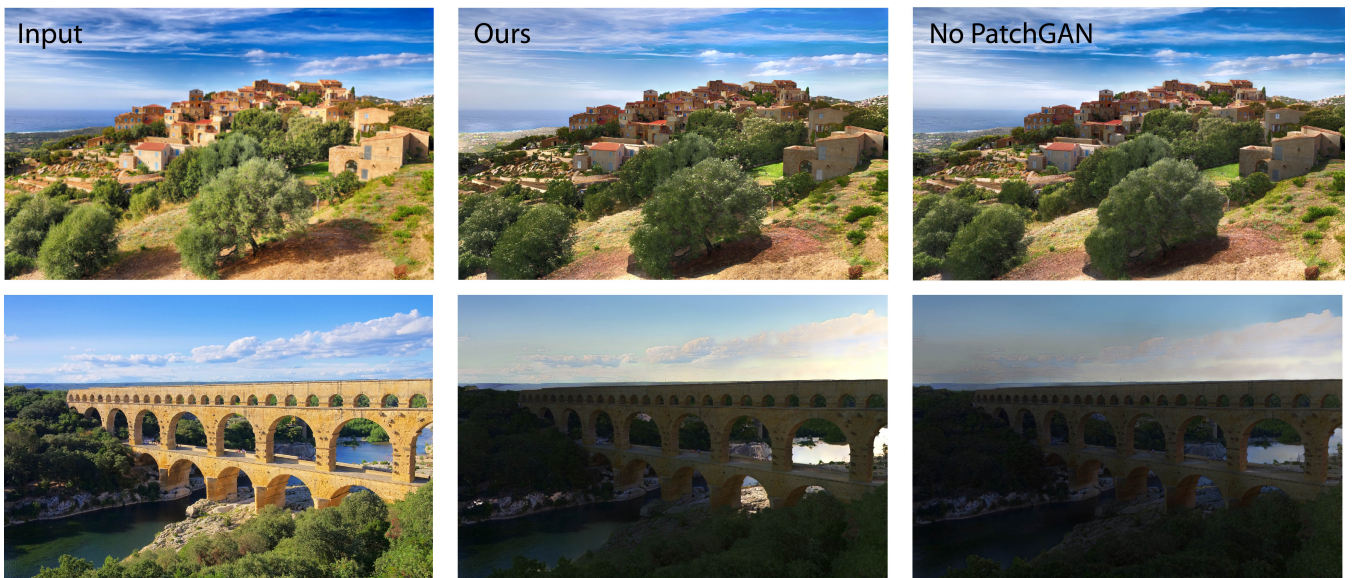


**Figure 2:** Examples of training pairs. From left to right: Input, Ground Truth Source Shadows, Target, Ground Truth Target Shadows.



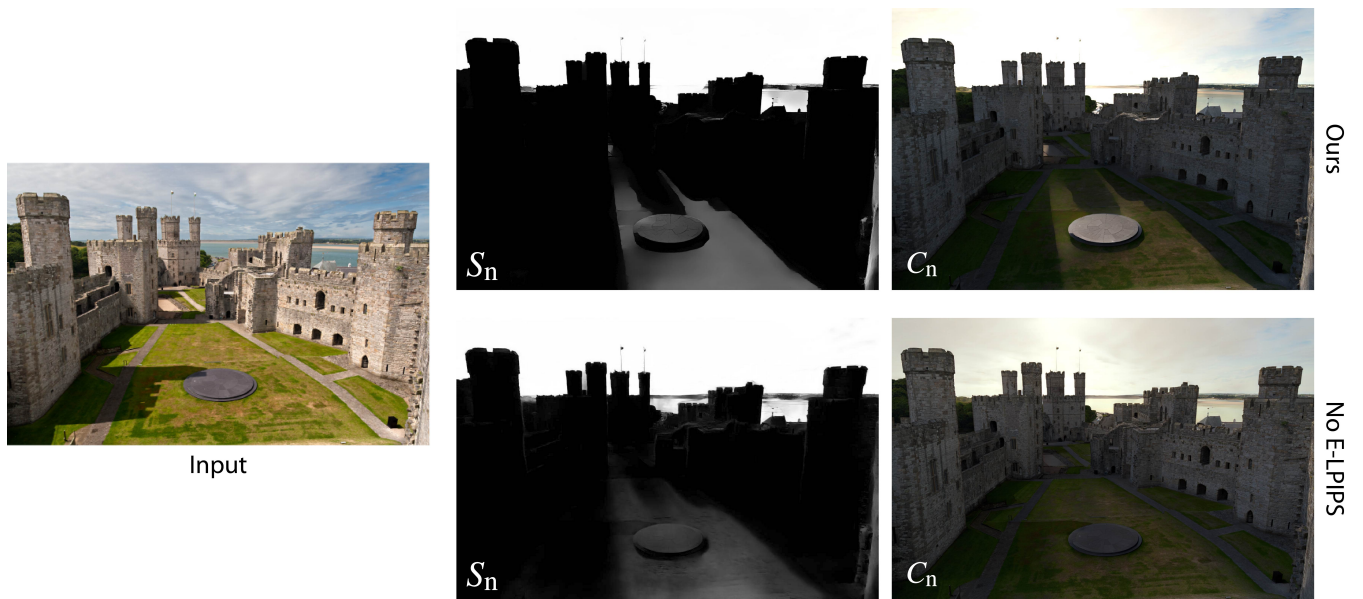


**Figure 3:** Comparison to real images with different lighting conditions. From left to right: Input, Our relighting, Ground Truth.



**Figure 4:** Ablation results when removing the PatchGAN loss. From left to right: Input, Our relighting, Ablation result.





**Figure 5:** Ablation results where the E-LPIPS loss is replaced by an MSE loss. Input to the network is shown on the left. **(top)** Ours for target shadow (left) and output (right). **(bottom)** Ablation for target shadow (left) and output (right).