

Generative Models for Multi-Illumination Color Constancy: Supplementary

Partha Das^{1,2}
p.das@uva.nl

Yang Liu²
yang@3duniversum.com

Sezer Karaoglu^{1,2}
s.karaoglu@3duniversum.com

Theo Gevers^{1,2}
th.gevers@uva.nl

University of Amsterdam, The Netherlands¹

3DUniversum, Amsterdam, The Netherlands²

1. Network Details

The architectural details are provided. The inputs to the network are first processed through a shallow convolution head, for feature conditioning. The image processing and the illumination processing head are detailed separately in 1 and 2, respectively. The input to the image processing head is the *RGB* color biased image, while the illumination processing head takes the concatenation of the sparse illumination map and the mask as input. The output for the image head and illumination head is a 64 channel and 32 channel tensor, respectively. All of them have the same spatial dimensions as the input. Tables 1 and 2 details the configurations.

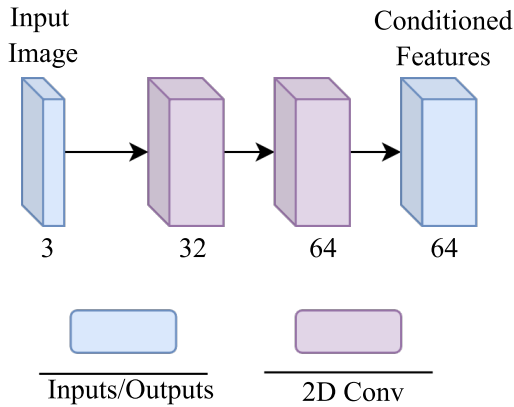


Figure 1. The architecture for the image processing head is shown. The input image is passed through two convolution layers, which squeezes and subsequently expands the channels of the outputs, thus conditioning the features down to the useful features in an end-to-end fashion.

Name	Layer	Kernel Size, Stride, Padding	Output Size
Input	conv1	3x3x32, 1, 1	256x256x32
Output	conv2	3x3x32, 1, 1	256x256x64

Table 1. Overview of the image processing head. The input is 3 channel and the output is a 64 channel tensor with the same spatial dimension.

The Pix2Pix [1] generator architecture is used, with just two changes: 1) The input channels are changed to match

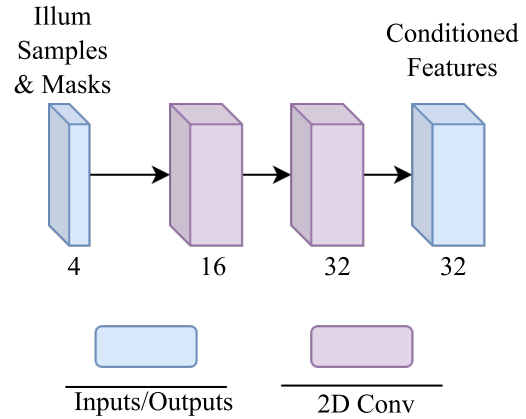


Figure 2. The architecture for the illumination processing head is shown. The input is the sparse illumination samples and the corresponding masks, concatenated channel wise (3 + 1 channels). This is passed through two convolution layers, which squeezes and subsequently expands the channels of the outputs, thus conditioning the features down to the useful features in an end-to-end manner.

Name	Layer	Kernel Size, Stride, Padding	Output Size
Input	conv1	3x3x16, 1, 1	256x256x16
Output	conv2	3x3x16, 1, 1	256x256x32

Table 2. Overview of the illumination processing head. The input is 4 channel concatenation of the sparse illumination samples and illumination masks. The output is a 32 channel tensor with the same spatial dimension.

the incoming tensors' dimension from the processing heads and, 2) The final activation is changed to Sigmoid instead of Tanh, to output a probability map instead of an image.

References

- [1] P. Isola, J. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial nets. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017.