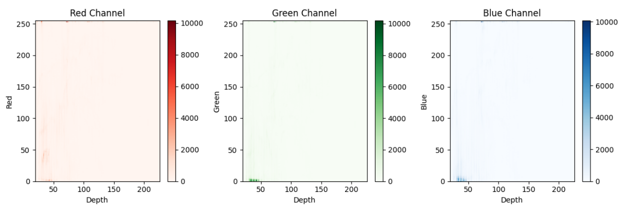
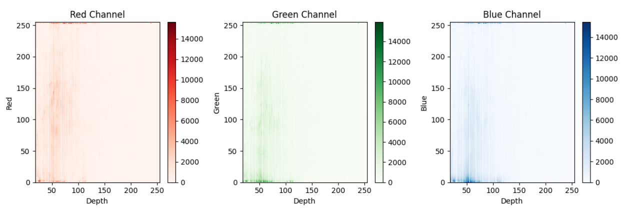
How does the Machine Perceive Depth for Indoor Single Images with CNN?

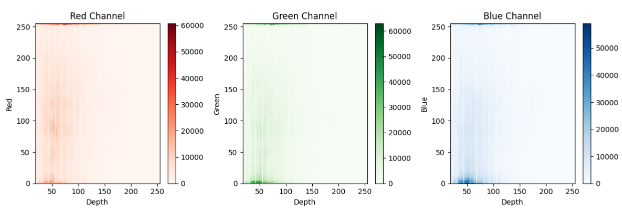
Supplementary Material



(a) 10 Images



(b) 100 Images



(c) 500 Images

Figure 16. Heatmap of the Relationship between the Distribution of RGB Three-channel Values and the Depth Map. The horizontal axis represents the depth range, while the vertical axis corresponds to the pixel count of the R, G, and B channels within the respective depth ranges. The colour bar values represent the pixel counts for three respective channels from different numbers of images randomly selected from the NYU dataset.

7. Phase Scrambling

Listing 1. Pseudocode of Phase Scrambling [?]

```
imFourier = fft2(input)
Amp = abs(imFourier)
Phase = angle(imFourier)
Phase = Phase + RandomPhase
imScrambled = ifft2(Amp * exp(1j * Phase))
imScrambled = GetRealPart(imScrambled)
```

List 1 presents the pseudo-code for the phase scrambling process.

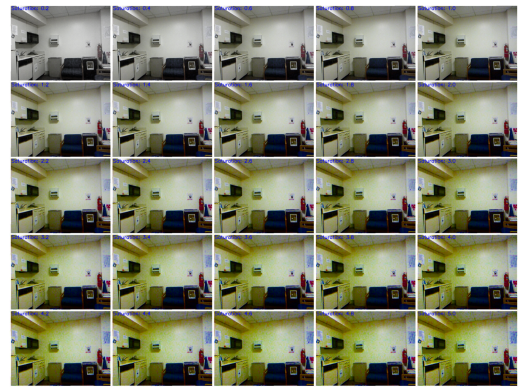


Figure 17. Saturation Maps with Different Saturation Values

8. Colour

Figure 16 represents the results of the accumulated values obtained from datasets of 50, 100, and 500 randomly sampled images. The depth range is depicted on the horizontal axis, while the vertical axis indicates the number of pixels in the R, G, and B channels corresponding to the specific depth ranges. These images show that the factors affecting depth are not significantly related to the distribution of pixels on the RGB channel.

9. Saturation

Figure 17 illustrates saturation maps with different saturation values, while Figure 18 displays various RGB images with different saturation values alongside their corresponding model performance. It can be observed that the model's performance does not exhibit a strong sensitivity to different saturation values. As the saturation values increase, there is a slight decline in performance. We hypothesised that this decline is due to the presence of more noise in images with high saturation values, as depicted in Figure 17, which negatively impacts the model's performance (shown in Sec. 12).

Note that during training, the channel order is BGR. However, for the sake of convenience in checking, the images have been converted to RGB channel order.

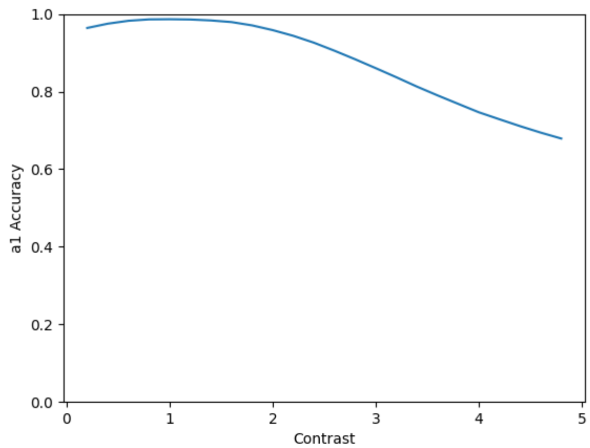


Figure 18. Different Saturation RGB Images and Model Performance

10. Shape

Figure 19 illustrates the RGB images alongside their respective shape maps, as well as the output depth maps generated by the trained model using these inputs. Despite the substantial disparity in information content between the RGB images and shape maps, their contributions to depth estimation appear to be similar.

11. Contrast

Due to the inclusion of shape, shading, and other information, Contrast cannot be extracted independently. The adopted method involves utilising a trained model and incrementally adjusting the contrast of the test set images during the reasoning process. This enables observation of the performance of the model's depth estimation and facilitates analysis.

Figure 20 shows images with different contrast values ranging from 0.2 to 5.

Figure 21 illustrates that when the contrast remains relatively stable compared to the original image, such as within the range of 0.6-1.6, we observed minimal changes in performance. This observation leads us to suspect that the narrower depth range typically found in indoor scenes may contribute to this phenomenon, as the variations within this small range might not be noticeable.

Considering the contrast formula, output = saturate $(src*alpha+beta)$, excessive or insufficient contrast values can result in a loss of picture details, leading to a significant decline in performance.

12. Discussion

However, Figure 17 and Figure 20, show that these approaches merely appeared to mirror the acquired knowledge of the data-driven model. The model attained its optimal performance when presented with input data characterised by the same levels of original saturation and contrast as those found in the training dataset.

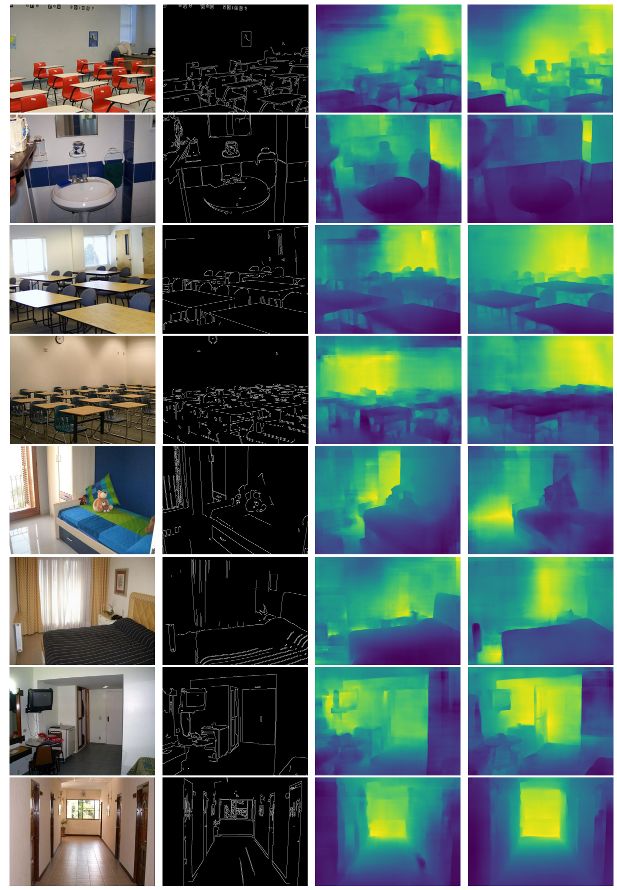


Figure 19. Performance of Shape ONLY model with New Indoor Scenes from other Domains. The left column displays RGB scene images, the second column presents corresponding edge maps, and the third column showcases the results generated by the pre-trained shape-input model. The right column exhibits the outcomes produced by the pre-trained RGB-input model.

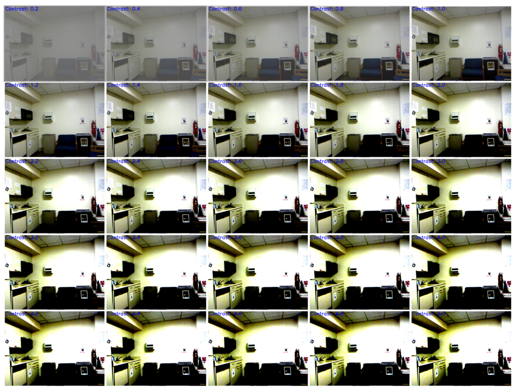


Figure 20. Contrast Maps with Different Contrast Values

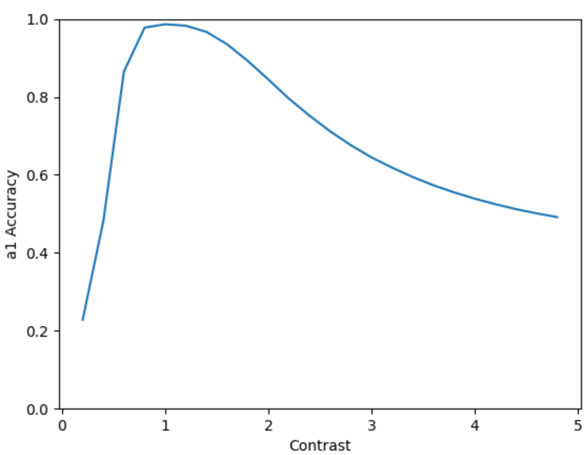


Figure 21. Different Contrast RGB Images and Model Performance