Supplementary Material: Rotation Equivariant Orientation Estimation for Omnidirectional Localization

Chao Zhang[1], Ignas Budvytis[1,2], Stephan Liwicki[1], and Roberto Cipolla[1,2]

[1] Cambridge Research Lab, Toshiba Europe Ltd, Cambridge, UK
{chao.zhang, stephan.liwicki}@crl.toshiba.co.uk
[2] Department of Engineering, University of Cambridge, Cambridge, UK
{ib255, rc10001}@cam.ac.uk

1 Data Augmentation

In this section, we provide the details of several data augmentations used in our experiments on 2D3DS[1] to improve localization performance on real images.

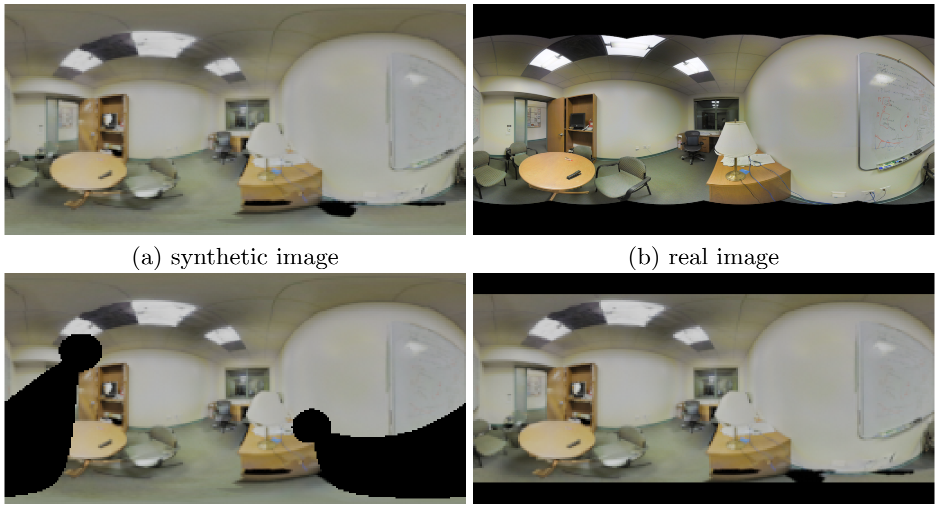
Color Augmentation This augmentation adjusts the brightness, contrast, saturation and hue of an image in random order. The scalar factors for these augmentations are set to 0.4, 0.4, 0.4 and 0.1, respectively.

Gamma Augmentation This augmentation performs gamma correction of an image. It is also known as Power Law Transform. γ value larger than 1 makes the shadow darker, while γ smaller than 1 makes the dark region brighter. The value of γ is uniformly chosen in the interval [0.5, 2.0].

Gaussian Blur This augmentation randomly blurs the image using a Gaussian kernel with standard deviation uniformly sampled in the interval [0, 0.5]. The augmentation is randomly enabled for 50% of the training images.

Gaussian Noise This augmentation randomly adds pixel-wise noise to an image. The noise is sampled from Gaussian distribution with zero mean and standard deviations in the interval [0, 12.75]. The augmentation is randomly enabled for 50% of the training images.

Cropping Augmentation This augmentation crops an image using a rectangular bounding box. For an input image of 128×256, the height and width of the box is uniformly sampled in the pixel interval [60, 120] and [160, 240], respectively. The bounding box is also rotated randomly in $\mathcal{SO}(3)$ space. The aim of "cropping augmentation" is to reduce over-fitting via structured "drop out". This augmentation is randomly enabled for 20% of the training images. Note, this augmentation is performed on input images prior to resizing. An example is shown in Fig.1(c).



(a) synthetic image (b) real image

(c) synthetic with random cropping (d) synthetic with random masking

Fig. 1. An example of synthetic rendering (a) and real image (b) with same camera pose from 2D3DS[1] are shown to illustrate the effect of data augmentation. The color differences in terms of brightness and contrast are prominent. To reduce over-fitting, we apply cropping augmentation (c) to introduce structured drop-out. Note at the top and bottom part, black artifacts are present in real image (b). To match this visible artifacts, masking augmentation (d) of random size is applied on synthetic images during training. (Only one random size is shown in (d))

Masking Augmentation For training, synthetic renderings are produced using Blender and are artifacts-free. While for testing, the real images from 2D3DS have black artifacts at the top and bottom. To reduce the effect of artifacts, we randomly mask the top and bottom of an image to zero values. The mask size is uniformly selected in the pixel interval [0, 10] for input of 64×64. The purpose of using "masking augmentation" is to better match the artifacts visible in the real testing data. This augmentation is randomly enabled for 50% of the training images. This augmentation is applied on resized input. As an illustration example, result done on original input image is shown in Fig.1(d).

References

1. Armeni, I., Sax, S., Zamir, A.R., Savarese, S.: Joint 2d-3d-semantic data for indoor scene understanding. arXiv preprint arXiv:1702.01105 (2017)