Depth Estimation on Underwater Omni-directional Images Using a Deep Neural Network

Haofei Kuang, Qingwen Xu and Sören Schwertfeger

{kuanghf, xuqw, soerensch}@shanghaitech.edu.cn

School of Information Science and Technology

ShanghaiTech University, Shanghai

https://robotics.shanghaitech.edu.cn/

Introduction

- Due to the properties of underwater environments, underwater perception is quite different from air. Images captured underwater usually look bluish or greenish. Besides, the underwater images are more blurred than that in air captured by the same camera due to turbidity. These reasons increase the difficulty of depth estimation from images.
- We exploit a deep learning based depth estimation for in-air perspective images to estimate the depth of underwater perspective and omni-directional images. The system overview is shown as Figure 1.

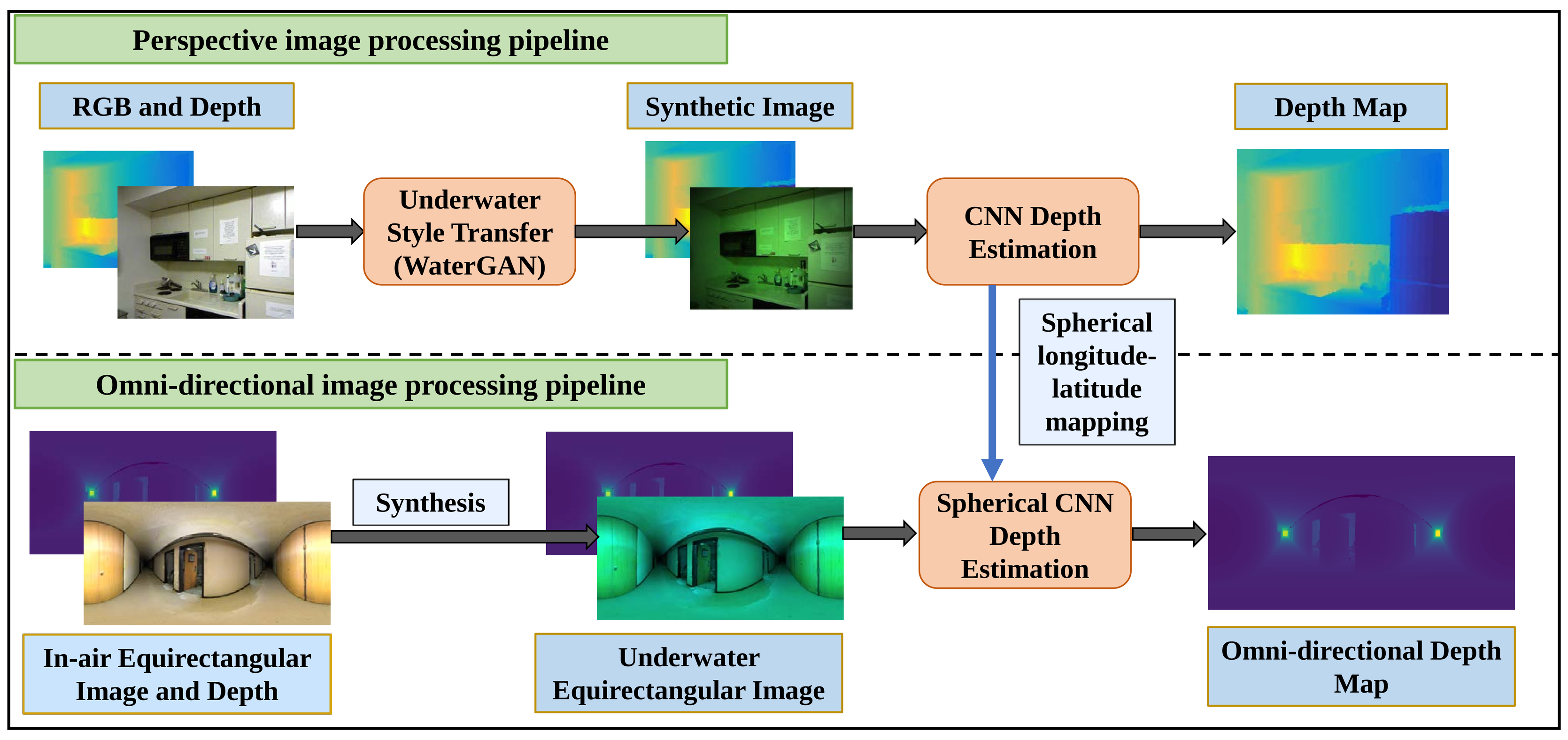

Figure 1: Perspective and omni-directional image processing training pipeline.

Distortion Removed

To handle the distortion of omni-directional images, we replace the square kernel with spherical ones in tangent space as shown in Figure 2. Assume the center of a $n \times n$ square kernel is $(0, 0)$, then the relative coordinates of $n \times n$ kernel can be described as $\left[(\alpha x_{ij}, \beta y_{ij})\right]_{n \times n}$ where

$$x_{ij} = \tan(|i|\,\Delta_\theta), \quad y_{ij} = \frac{\tan(|j|\,\Delta_\phi)}{\cos(|i|\,\Delta_\theta)} \tag{1}$$

, α and β are the symbols consistent with the relative coordinates to center.

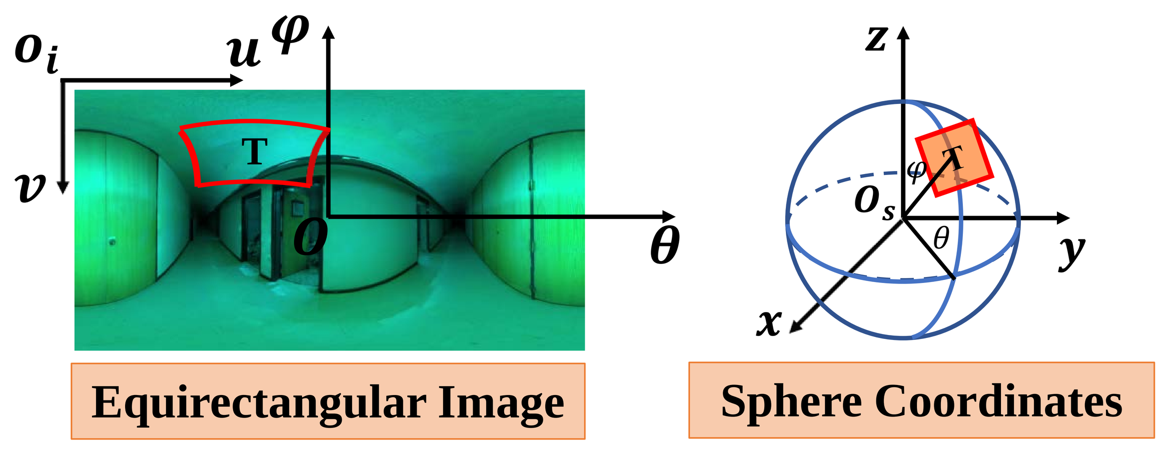

Figure 2: Mapping between longitude-latitude and spherical coordinates.

Synthetic Dataset

- *Underwater NYU*: We transfer NYU-v1 [1] dataset to underwater style via WaterGAN [2], which is an underwater style transfer generative adversarial network (GAN). It contains 32,704 images, of which 26,163 are for training and 6,541 are for testing, and all input data is resized to 228×304.
- *Underwater 360D*: The 360D dataset [3] is transferred to underwater style via . It contains 35,977 images ,of which 34,679 are for training and 1,298 are for testing, and image size is 256×512.
- Preprocessing: random scaling, rotation, center-crop and horizontal flip, as proposed by Eigen et al. [4].

Methods & Models

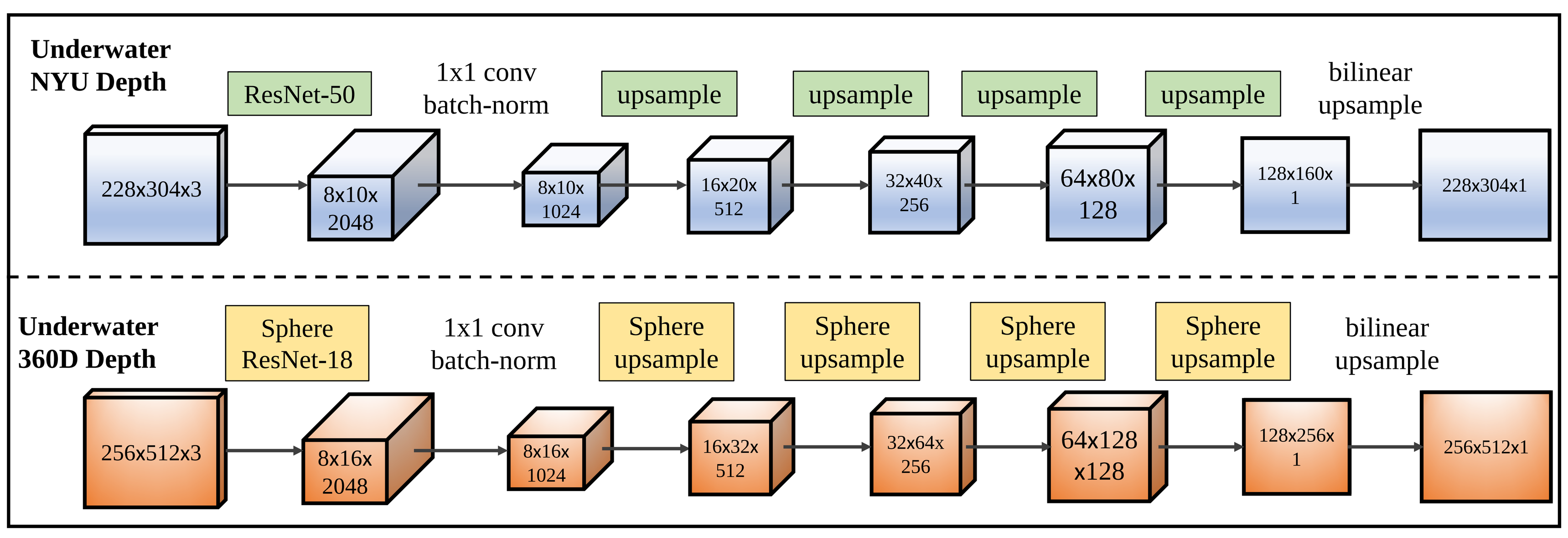

Figure 3: The Fully Convolutional Residual Neural Network (FCRN) [5] architecture of depth estimation networks for perspective and omni-directional images. We train a ResNet-50 FCRN model for the underwater NYU dataset. We are then using a Sphere ResNet-18 architecture to train the model with underwater 360D datasets to reduce the memory and training time.

Experiments and Results

- Perspective Images

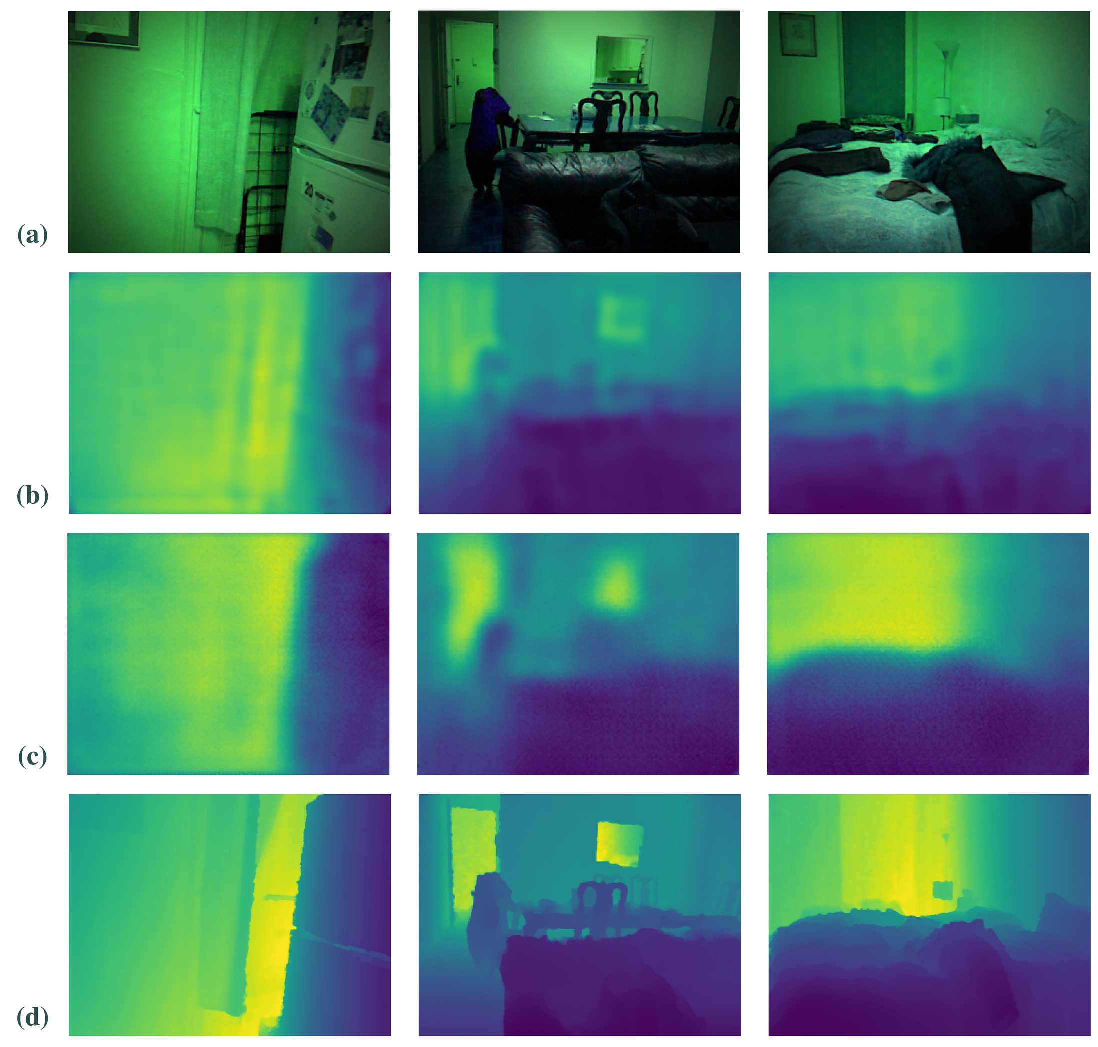

Figure 4: The experimental results of underwater NYU dataset. (a) are RGB images from the testing set. (b) are the predicted depth of Eigen et al.'s [4]; (c) are the predicted depth maps of ours; (d) are the ground truth depth maps.

- Omnidirectional Images

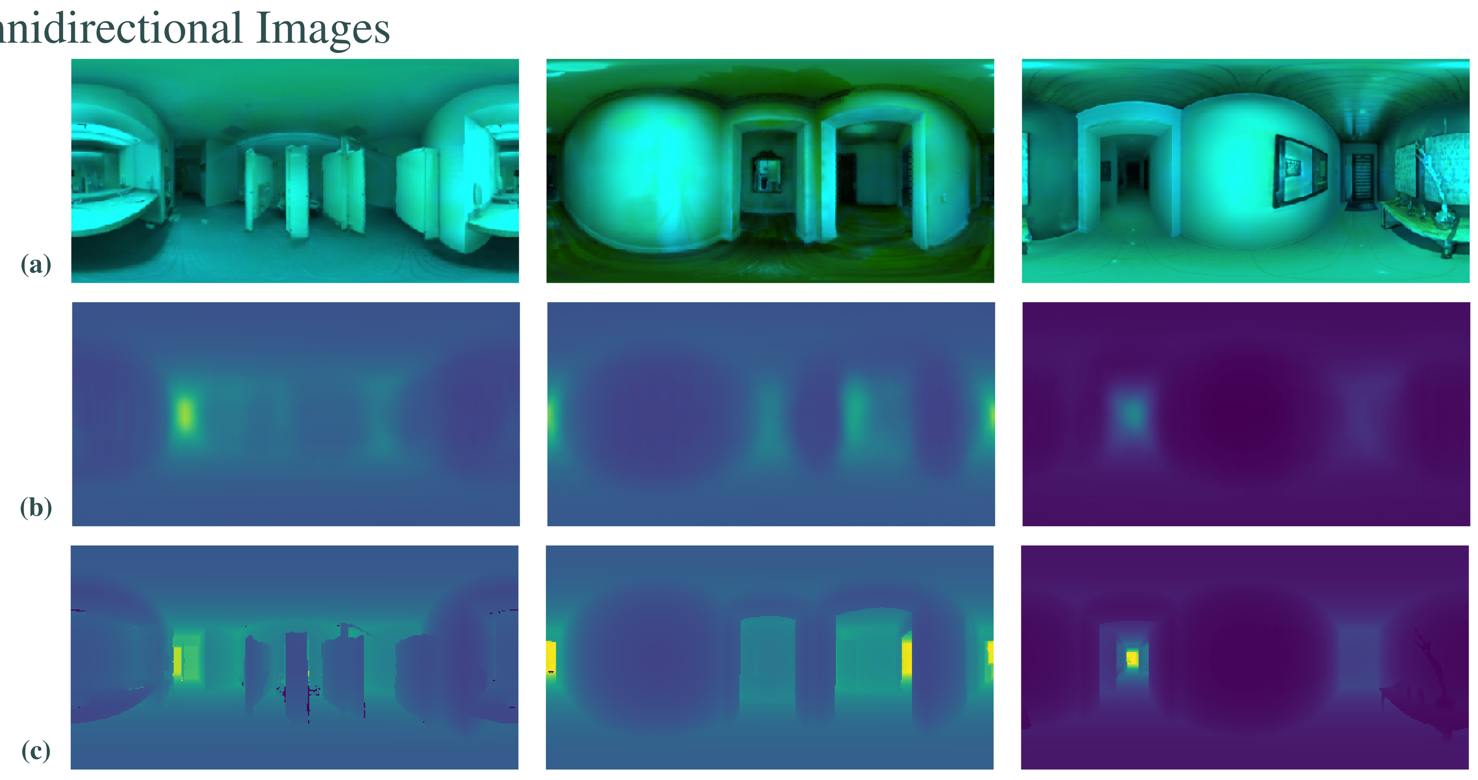

Figure 5: The experimental results of underwater 360D dataset. (a) are RGB images from the testing set. (b) are the predicted depth maps; (c) are the ground truth depth maps.

- Error Metrics: The experiment results shows our model exhibit a very nice performance on the dataset. And our method still performs better than Eigen et al.'s. [4] in perspective scene.

Input	Model	RMSE	MAE	REL	δ_1	t_gpu(s)
Perspective	ResNet-50	0.162	0.117	0.098	0.914	0.0201
	Eigen et al.	0.235	0.184	0.148	0.806	0.005
Omnidirectional	SphereResNet-18	0.604	0.362	0.172	0.711	0.0145

Table 1: The error metrics of each model. t_gpu means the average operation time on each image on GPU.

Conclusion

- Underwater NYU-v1 and 360D dataset: transfer in-air images to underwater style images via WaterGAN and theoretical analysis method respectively.
- Conventional FCRN achieves good performance for underwater perspective images and spherical FCRN works well for underwater omni-directional images.
- Future work:
  - Transfer conventional FCRN to spherical FCRN without retraining.
  - Enlarge the diversity of the training dataset to adapt to different scenarios.
  - Collect more real underwater images to improve the robust of our models.

References

[1] N. Silberman and R. Fergus, "Indoor scene segmentation using a structured light sensor," in *Proceedings of the International Conference on Computer Vision - Workshop on 3D Representation and Recognition*, 2011.

[2] J. Li, K. A. Skinner, R. M. Eustice, and M. Johnson-Roberson, "Watergan: unsupervised generative network to enable real-time color correction of monocular underwater images," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 387–394, 2018.

[3] N. Zioulis, A. Karakottas, D. Zarpalas, and P. Daras, "Omnidepth: Dense depth estimation for indoors spherical panoramas," in *The European Conference on Computer Vision (ECCV)*, September 2018.

[4] D. Eigen, C. Puhrsch, and R. Fergus, "Depth map prediction from a single image using a multi-scale deep network," in *Advances in neural information processing systems*, 2014, pp. 2366–2374.

[5] I. Laina, C. Rupprecht, V. Belagiannis, F. Tombari, and N. Navab, "Deeper depth prediction with fully convolutional residual networks," in *2016 Fourth International Conference on 3D Vision (3DV)*. IEEE, 2016, pp. 239–248.

[6] B. Coors, A. Paul Condurache, and A. Geiger, "Spherenet: Learning spherical representations for detection and classification in omnidirectional images," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 518–533.

[7] K. Tateno, N. Navab, and F. Tombari, "Distortion-aware convolutional filters for dense prediction in panoramic images," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 707–722.