Supplementary Material

Javier Rodríguez-Puigvert[1*] Víctor M. Batlle[1*] J.M.M. Montiel[1] Rubén Martinez Cantín[1]
Pascal Fua[2] Juan D. Tardós[1] Javier Civera[1]

[1]I3A - Universidad de Zaragoza [2]École Polytechnique Fédérale de Lausanne

1. Network architectures

LightDepth U-Net We use a U-Net architecture with skip connections and two decoders. Our encoder is a ResNet18 [5] initialized with the weights from pre-training in ImageNet [3]. Regarding the decoders, our albedo decoder uses a sigmoid activation function and our depth decoder an ELU+1 activation function after the last convolution.

LightDepth DPT We extend LightDepth DPT [6] adding a branch for the prediction of albedo decoder. For the depth estimation, we initialize the encoder and depth decoder with DPT Hybrid weights. For albedo estimation, we train the albedo decoder from scratch. In our pipeline, we use the half of resolution than the original images for training, upsampling the outputs with bilinear interpolation.

Figure 1 presents the head for the albedo decoder that includes a sigmoid activation function.

2. Datasets

Table 2 shows which sections of the C3VD [2] were used for training / testing. We split into sections to ensure a fair comparison along the dataset. Regarding real endoscopy images, we use with the sequences 051, 009 and 058 of the EndoMapper dataset [1].

*equal contribution

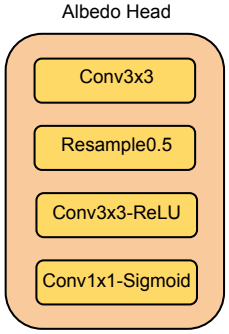

Figure 1. Albedo estimation head

Model	Texture	Video	Frames	Stage
Cecum	1	b	765	Train
Cecum	2	b	1120	Train
Cecum	2	c	595	Train
Cecum	4	a	465	Train
Cecum	4	b	425	Train
Sigmoid Colon	1	a	800	Train
Sigmoid Colon	2	a	513	Train
Sigmoid Colon	3	b	536	Train
Transcending Colon	1	a	61	Train
Transcending Colon	1	b	700	Train
Transcending Colon	2	b	102	Train
Transcending Colon	2	c	235	Train
Transcending Colon	3	b	214	Train
Transcending Colon	4	b	595	Train
Descending Colon Down	4	a	74	Train
Cecum	1	a	276	Test
Cecum	2	a	370	Test
Cecum	3	a	730	Test
Sigmoid Colon	3	a	610	Test
Transcending Colon	2	a	194	Test
Transcending Colon	3	a	250	Test
Transcending Colon	4	a	384	Test
Descending Colon Up	4	a	74	Test

Table 1. Dataset Split for C3VD

3. Normals from Depth

Figure 2 shows examples of Open3d [7], in-house, U-Net and TFtN [4] used in the analysis.

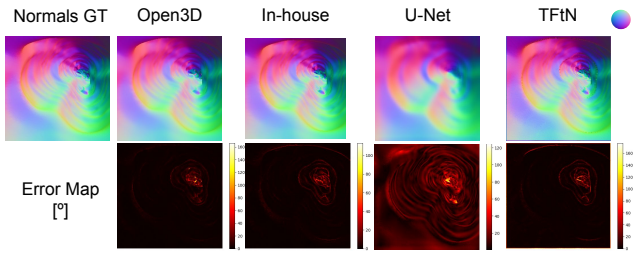

Figure 2. Quantitative results of different approaches to obtaining surface normals from a depth map.

4. Additional results

We present additional quantitative and qualitative results. Figure 3 shows additional qualitative results of LightDepthDPT TTR in real colonoscopy and gastroscopy procedures. Figure 4 shows quantitative results of LightDepth U-Net in the transverse and cecum sections of the C3VD.

Finally, in Figure 5 we show examples of LightDepth U-Net in our in-house synthetic dataset. The predicted depth and normals capture the shape of the colon sections, as shown in the 3D reconstruction. The albedo map appears brighter as we fix Value Channel to $V = 100$. Our method recovers the different albedo of mucosa and blood vessels.

References

[1] Pablo Azagra et al. EndoMapper dataset of complete calibrated endoscopy procedures. *arXiv:2204.14240*, 2022. 1

[2] Taylor L. Bobrow, Mayank Golhar, Rohan Vijayan, Venkata Akshintala, Juan R. Garcia, and Nicholas J. Durr. Colonoscopy 3d video dataset with paired depth from 2D-3D registration. *arXiv:2206.08903*, 2022. 1

[3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009. 1

[4] Rui Fan, Hengli Wang, Bohuan Xue, Huaiyang Huang, Yuan Wang, Ming Liu, and Ioannis Pitas. Three-filters-to-normal: An accurate and ultrafast surface normal estimator. *IEEE Robotics and Automation Letters*, 6(3):5405–5412, 2021. 1

[5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016. 1

[6] René Ranftl, Alexey Bochkovskiy, and Vladlen Koltun. Vision transformers for dense prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 12179–12188, October 2021. 1

[7] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Open3D: A modern library for 3D data processing. *arXiv:1801.09847*, 2018. 1

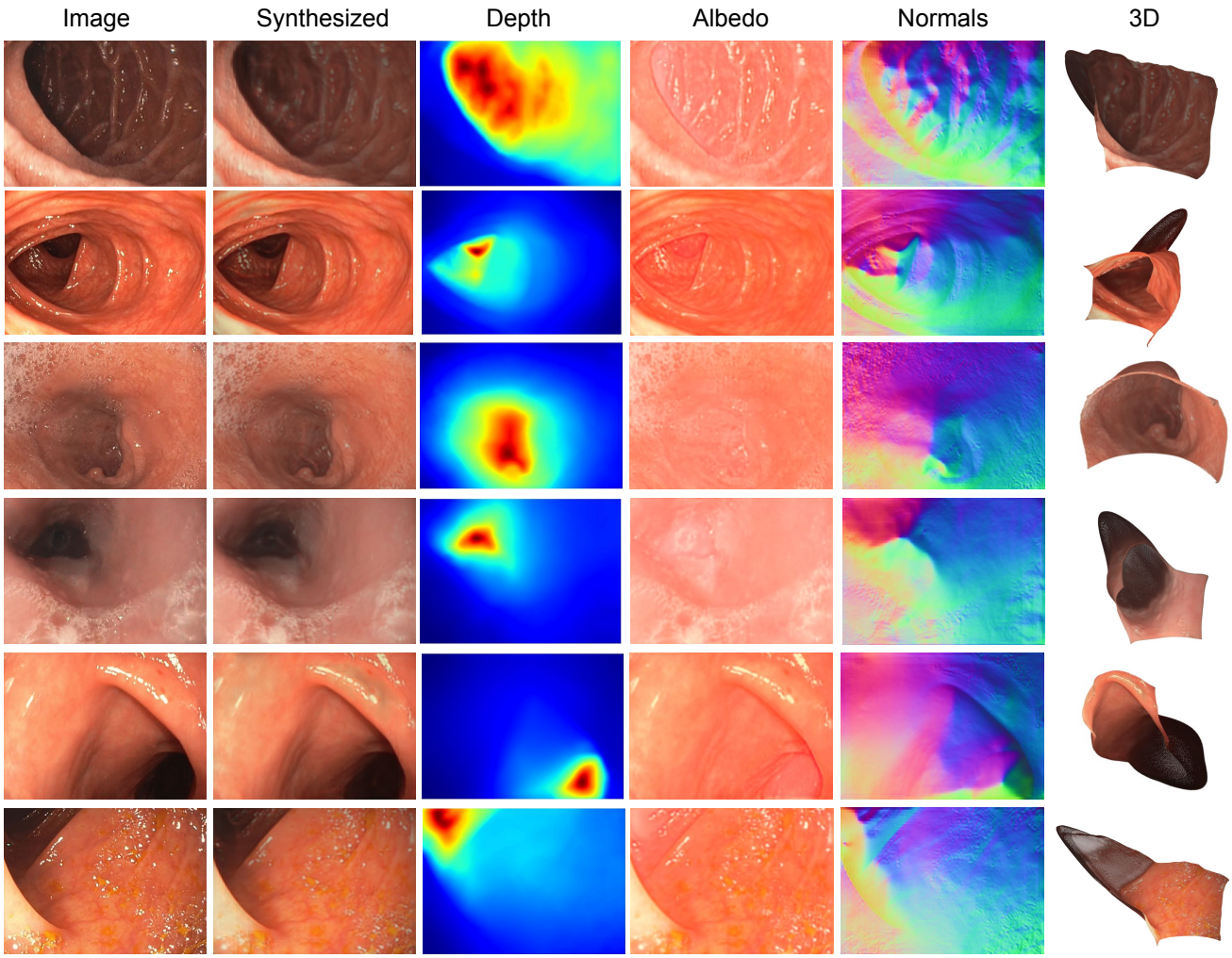

Figure 3. Additional qualitative examples of LightDepthDPT in real colonoscopy and gastroscopy procedures.

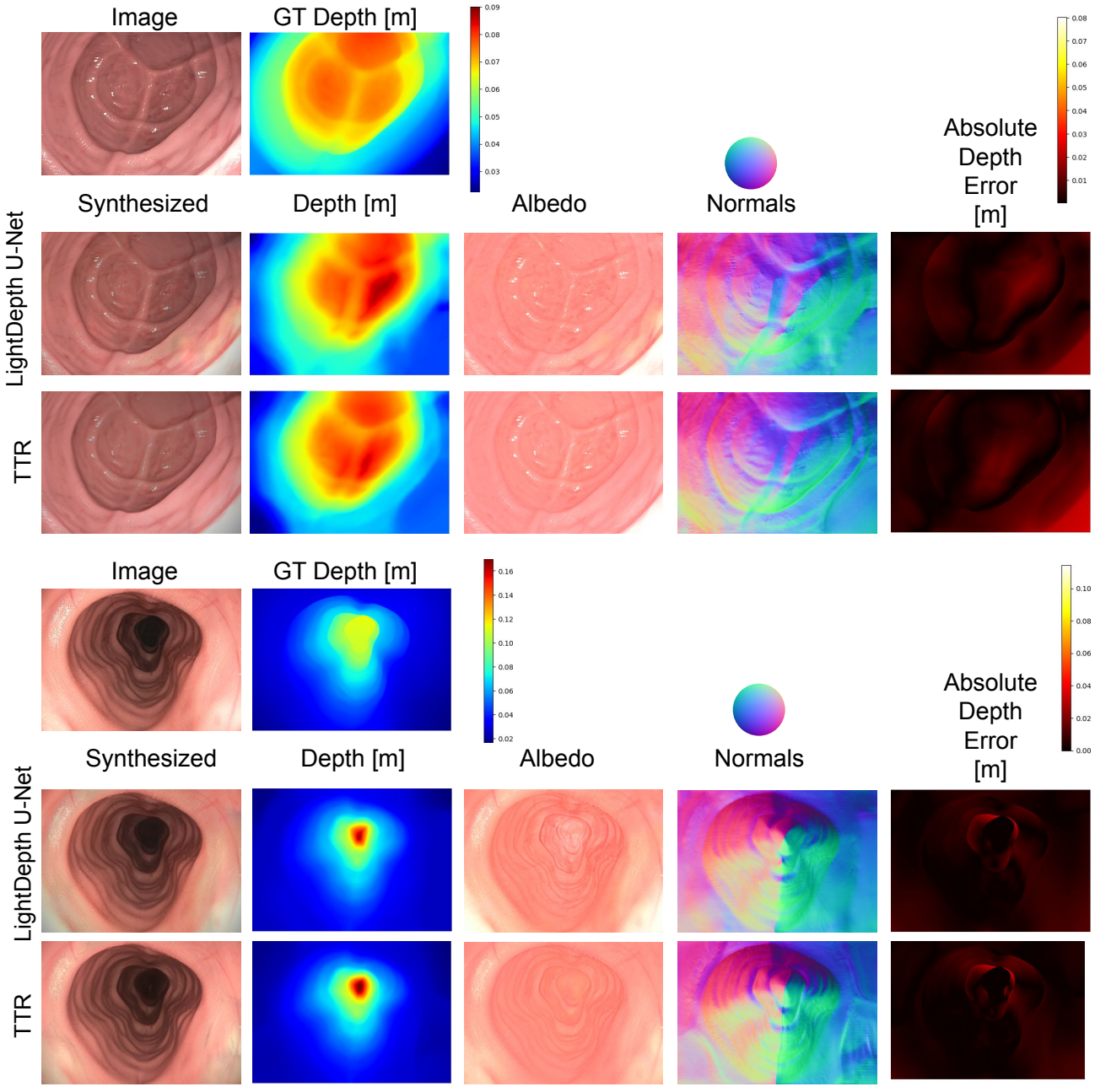

Figure 4. Additional quantitative examples of LightDepth U-Net in C3VD.

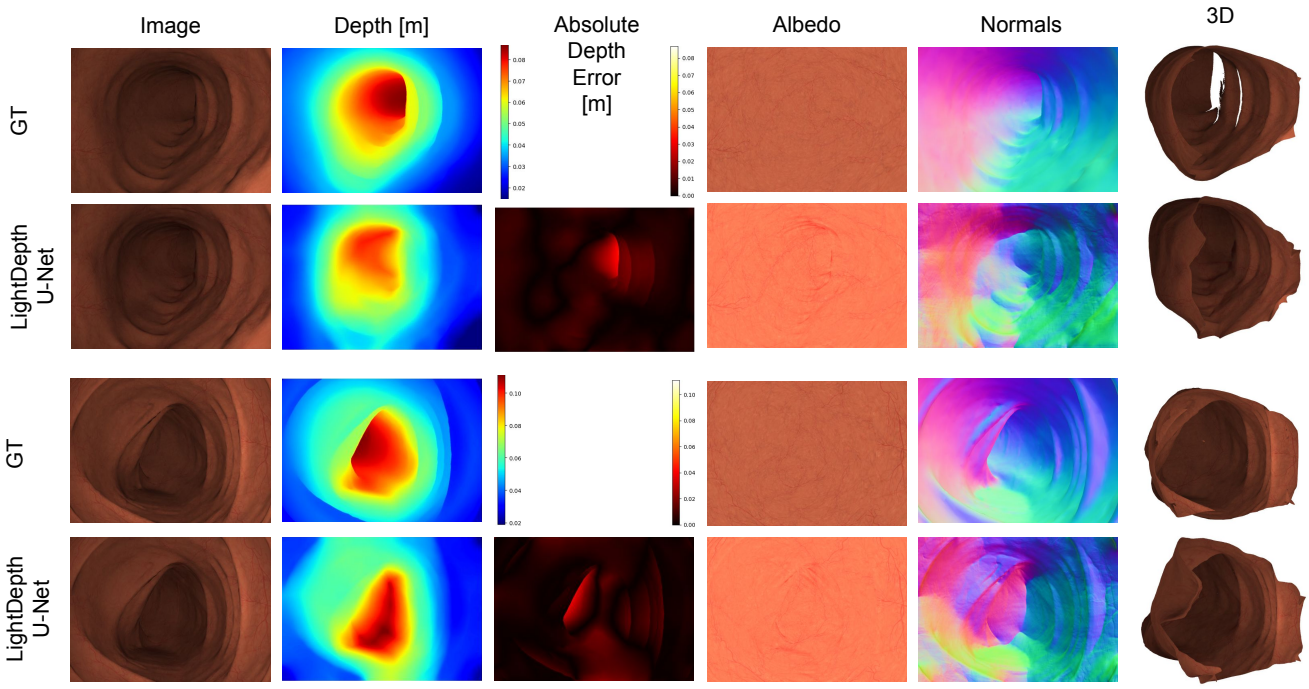

Figure 5. Qualitative examples of LightDepth U-Net in Synthetic dataset.