Bridging Unsupervised and Supervised Depth from Focus via All-in-Focus Supervision

Ning-Hsu Wang[1,2] Ren Wang[1] Yu-Lun Liu[1] Yu-Hao Huang[1] Yu-Lin Chang[1] Chia-Ping Chen[1] Kevin Jou[1]

[1]MediaTek Inc. [2]National Tsing Hua University

arXiv Paper: https://arxiv.org/pdf/2108.10843.pdf

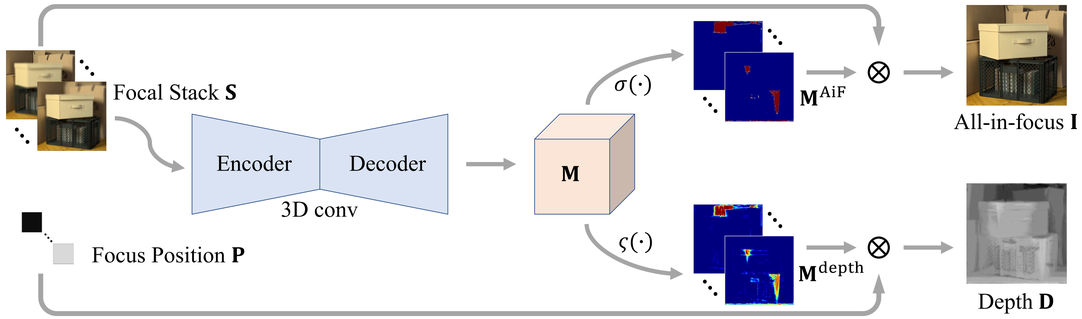
Github Code: https://github.com/albert100121/AiFDepthNet/

Summary

We reach SOTA in depth from focus, and the contributions are:

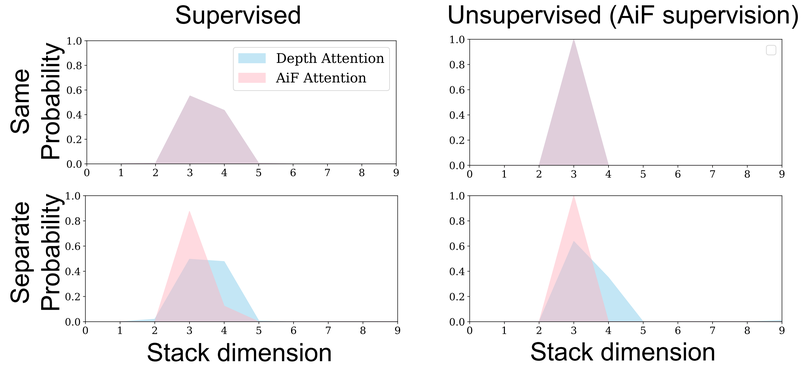
- We outperform SOTA methods in various comparisons, and also runs faster.
- Our method is the first that can learn depth estimation **unsupervisedly from only all-in-focus (AiF) images**, and perform favorably against SOTA methods.
- Our method allows test-time optimization on real-world data to mitigate the domain gap, especially when ground truth depth data are unavailable.

Overview



- Given a focal stack **S**, a 3D ConvNet produces an attention map **M** shared between depth and AiF estimation.
- With different normalization functions $\sigma(\cdot)$ and $\varsigma(\cdot)$, the attention map can be further manipulated to generate either depth or AiF results.

Separate Normalization



We adopt separate normalization functions for depth and AiF estimation to pursue a proper distribution on the attention map for sparse focal stacks.

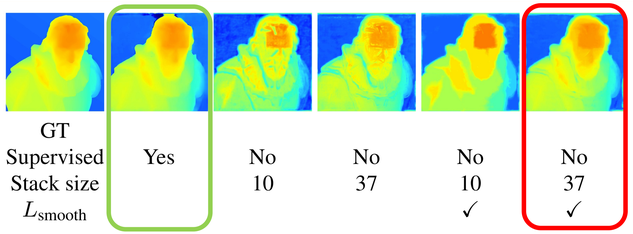
Training Loss

We can train our network either supervisedly with GT depth data, or unsupervisedly with AiF images as well as a smoothness loss.

- $L_{supervised} = L_{depth} = \mathbb{E}\left[\|D - D_{gt}\|_1\right]$
- $L_{unsupervised} = L_{AiF} + \alpha L_{smooth}$, where $L_{AiF} = \mathbb{E}\left[\|I - I_{gt}\|_1\right]$

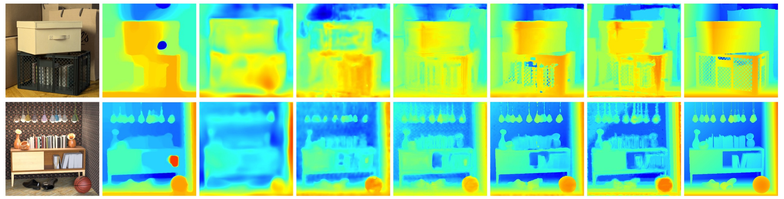
[1] Maximov et al. CVPR 2020 [2] Hazirbas et al. ACCV 2018 [3] Zhao et al. CVPR 2017 [4] Möller et al. TIP 2015 [5] Honauer et al. ACCV 2016 [6] Scharstein et al. GCPR 2014 [7] Suwajanakorn et al. CVPR 2015

Experimental Results

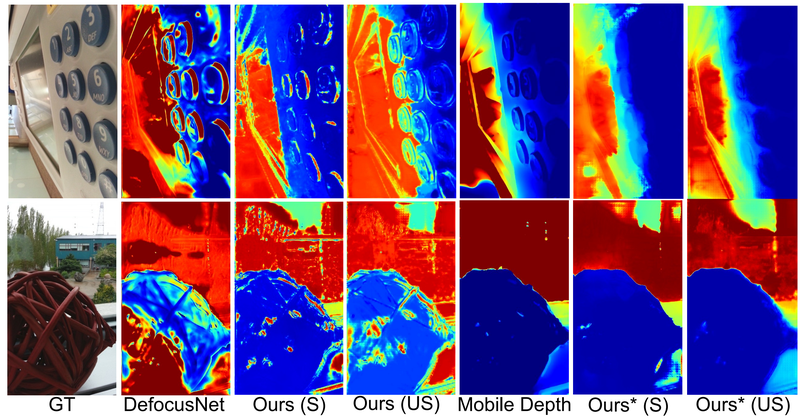


The result from unsupervised learning performs favorably againt the one trained with GT depth data.

Dataset	Supervised	MSE↓	RMSE↓	Bump↓
Ours	Yes	**0.0472**	**0.2014**	1.58
DefocusNet [1]	Yes	0.0593	0.2355	2.69
DDFF [2]	Yes	0.19	0.42	1.92
PSPNet [3]	Yes	0.37	0.53	**1.21**
VDFF [4]	Yes	1.3	1.15	1.58
Ours	No	0.0746	0.2398	2.58

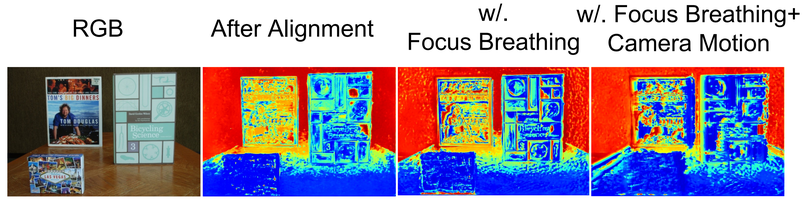


Test-time Optimization



After test-time optimization, our model produces better visual results by closing the domain gap between training data and test data. (* indicates test-time optimization)

Limitation



An alignment process is need for out-of-alignment images.