Deep Depth from Focus with Differential Focus Volume

PennState

Fengting Yang, Xiaolei Huang, and Zihan Zhou

College of Information Sciences and Technology, The Pennsylvania State University, University Park, PA, USA

CVPR JUNE 19-24 2022 NEW ORLEANS • LOUISIANA

Introduction

Depth From Focus (DFF):

- Take pictures of the same scene with different focal distances
- Infer depth of the scene through focal analysis

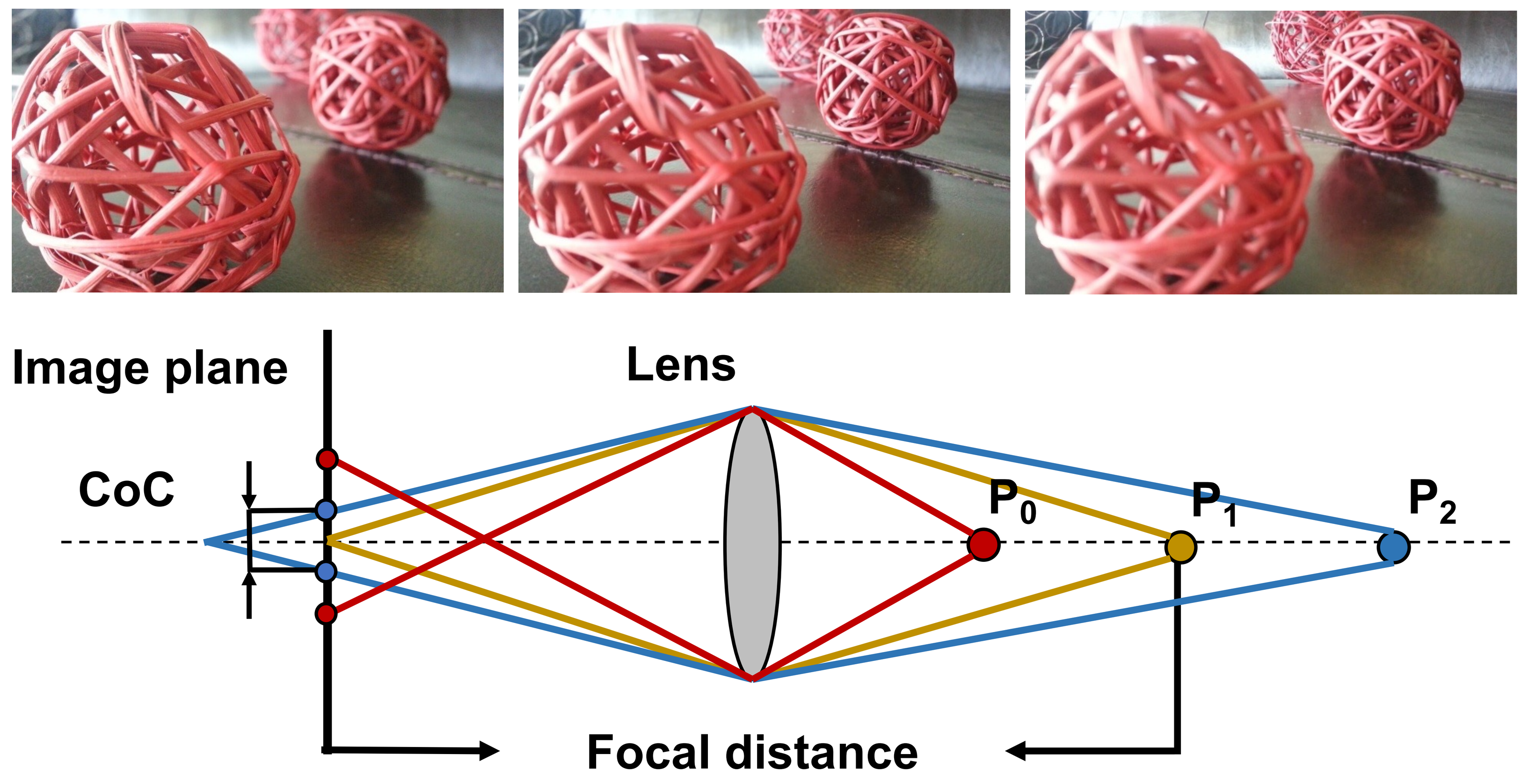


Figure 1. **Top**: DFF input (focal stack), **Bottom**: Optical diagram of DFF.

• Connections between Stereo Matching and DFF

- Stereo matching:
  - Find the best matched pixels among frames
  - Use cost volume to compare pixel similarity features
- DFF:
  - Find the sharpest pixels among frames
  - Use focus volume to compare pixel sharpness features

Challenge and Our Contribution

- **Challenge:** No existing work considers the specialty of DFF in their network designs
- **Contribution:** Introduce the deep differential focus volume by considering the geometric prior of DFF and the success of the deep cost volume in stereo matching

Method

Deep Focus Volume (FV)

- Imitate the deep cost volume in stereo matching
- Stack multi-frame features in a new frame dimension

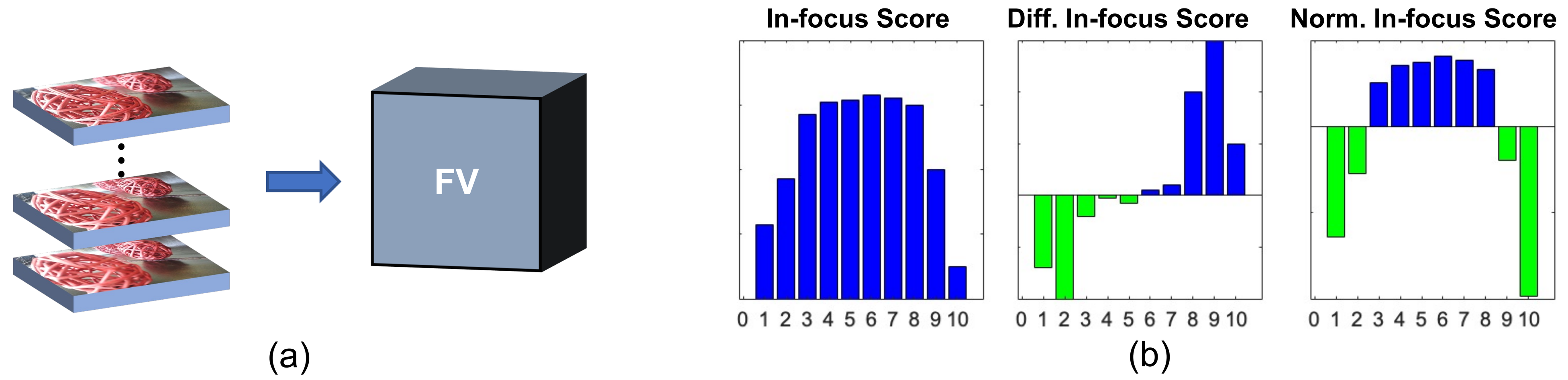
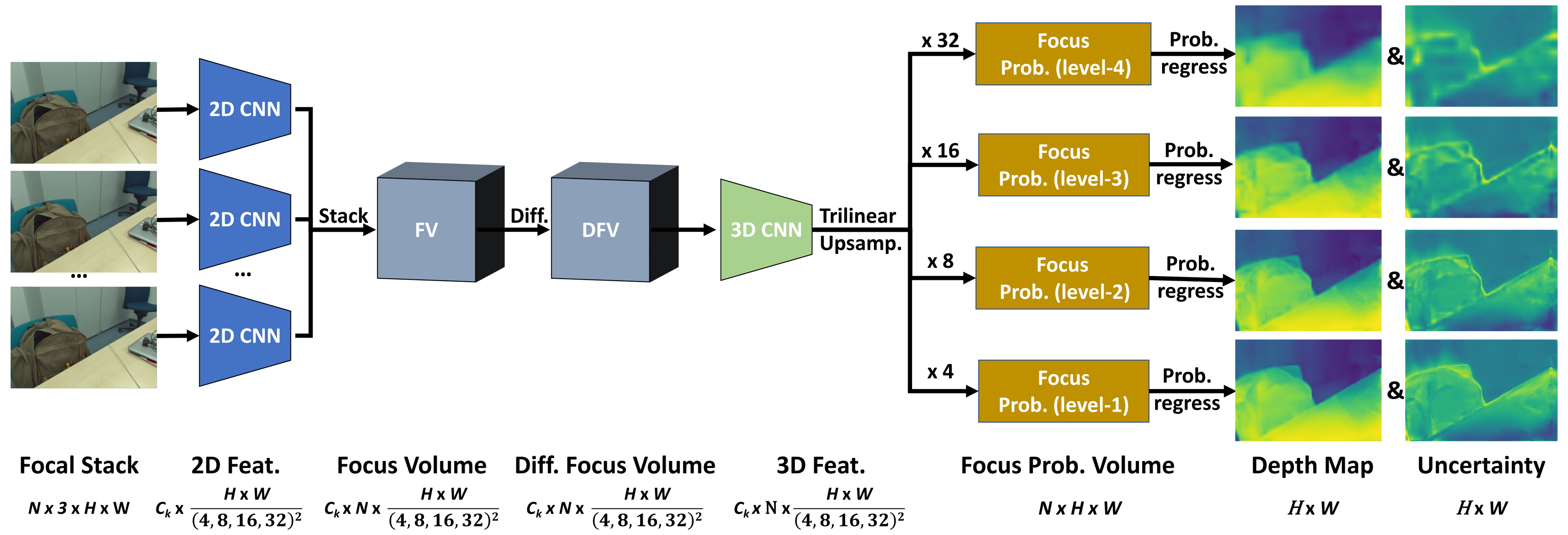


Figure 2. (a) Demonstration of FV, (b) Illustration of differential features and normalized features.

Deep Differential Focus Volume (DFV)

- Consider that a focal stack only has one best-focused pixel of a 3D point
- Differentiate FV along the frame dimension to detect the single extremum



Depth Regression and Uncertainty Estimation

- Given an N-frame focal stack, the depth $\hat{d}_j$ of pixel x_j

$$\hat{d}_j = \sum_{i=1}^{N} p_j^i \cdot l^i \quad (1)$$

  - p_j^i : the probability that the pixel x_j in the i^{th} frame is the best-focused
  - l^i : the focal distance of the i^{th} frame
- The prediction uncertainty $\emptyset_j$ of pixel x_j

$$\emptyset_j = \sqrt{\sum_{i=1}^{N} p_j^i \cdot (l^i - \hat{d}_j)^2} \quad (2)$$

Experiments

Method	MSE ↓	RMS↓	log RMS ↓	Abs. rel.↓	Sqr. rel.↓	δ ↑	δ^2 ↑	δ^3 ↑	Bump.↓	avgUnc.↓	Time(ms)↓
VDFF [29]	$29.66e^{-2}$	$5.05e^{-1}$	0.87	1.18	$85.62e^{-2}$	17.92	32.66	50.31	1.12	–	–
RDF [18]	$11.15e^{-2}$	$3.22e^{-1}$	0.71	0.46	$23.95e^{-2}$	39.48	64.65	76.13	1.54	–	–
DDFF [13]	$3.34e^{-2}$	$1.67e^{-1}$	0.27	0.17	$3.56e^{-2}$	72.82	89.96	96.26	1.74	–	50.6
DefocusNet [27]	$2.18e^{-2}$	$1.34e^{-1}$	0.24	0.15	$3.59e^{-2}$	81.14	93.31	96.62	2.52	–	24.7
Ours-FV	**$1.88e^{-2}$**	**$1.25e^{-1}$**	**0.21**	0.14	$2.43e^{-2}$	81.16	**94.97**	**98.08**	1.45	0.24	**18.1**
Ours-DFV	$2.05e^{-2}$	$1.29e^{-1}$	0.21	**0.13**	**$2.39e^{-2}$**	**81.90**	94.68	98.05	**1.43**	**0.17**	18.2

Table 1. Evaluation results on FoD500 test set.

Method	MSE ↓	RMS↓	log RMS ↓	Abs. rel.↓	Sqr. rel.↓	δ ↑	δ^2 ↑	δ^3 ↑	Bump.↓	avgUnc.↓	Time(ms)↓
VDFF [29]	$156.55e^{-4}$	12.14^{-2}	0.98	1.38	$241.2e^{-3}$	15.26	29.46	44.89	0.43	–	–
RDF [18]	$91.81e^{-4}$	$9.41e^{-2}$	0.91	1.00	$139.4e^{-3}$	15.65	33.08	47.48	1.33	–	–
DDFF [13]	$8.97e^{-4}$	$2.76e^{-2}$	0.28	0.24	$9.47e^{-3}$	61.26	88.70	96.49	0.52	–	191.7
DefocusNet [27]	$8.61e^{-4}$	$2.55e^{-2}$	0.23	0.17	**$6.00e^{-3}$**	72.56	94.15	97.92	0.46	–	34.3
Ours-FV	$6.49e^{-4}$	$2.28e^{-2}$	0.23	0.18	$7.10e^{-3}$	71.93	92.80	97.86	0.42	$5.20e^{-2}$	**33.2**
Ours-DFV	**$5.70e^{-4}$**	**$2.13e^{-2}$**	**0.21**	**0.17**	$6.26e^{-3}$	**76.74**	**94.23**	**98.14**	**0.42**	**$4.99e^{-2}$**	33.3

Table 2. Evaluation results on DDFF-12 validation set. Our-DFV also ranks the 1st on DDFF-12 test set.

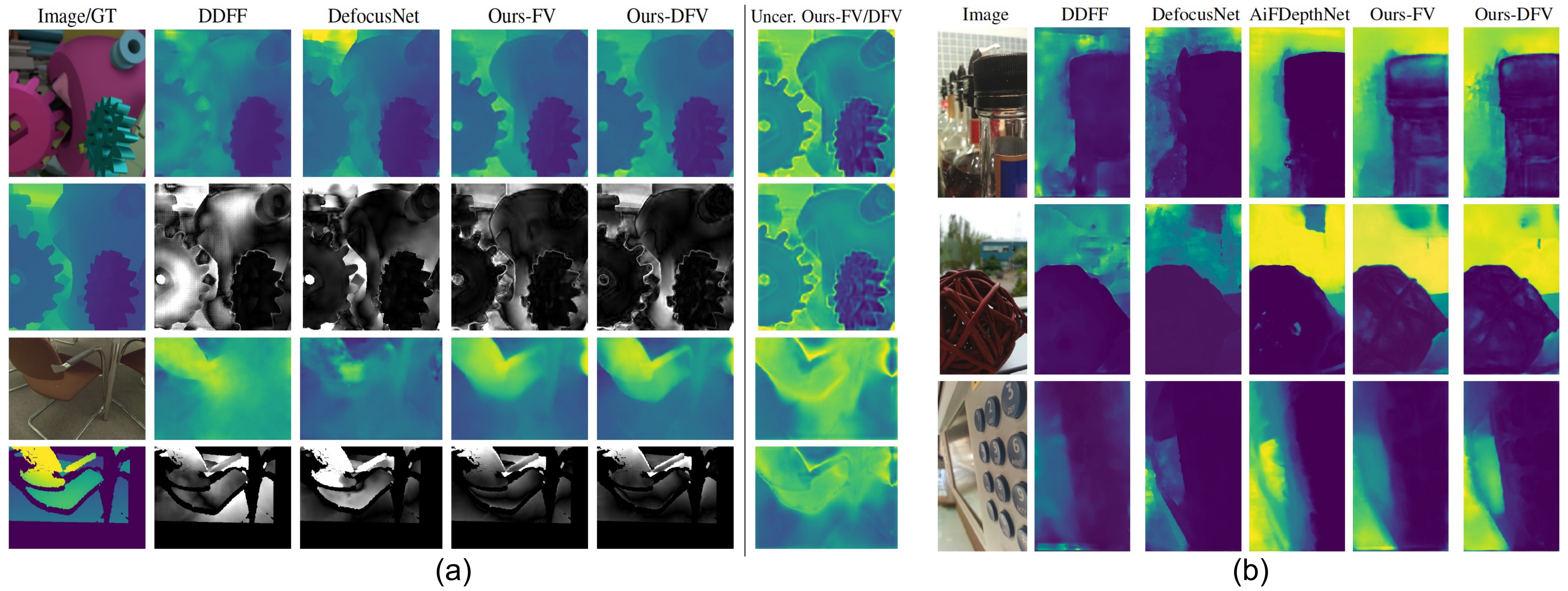


Figure 3. (a) Qualitative results on FoD500 (top two rows) and DDFF-12 (bottom two rows) dataset. (b) Qualitative results on Mobile depth dataset. The warmer or brighter the color, the higher the value.

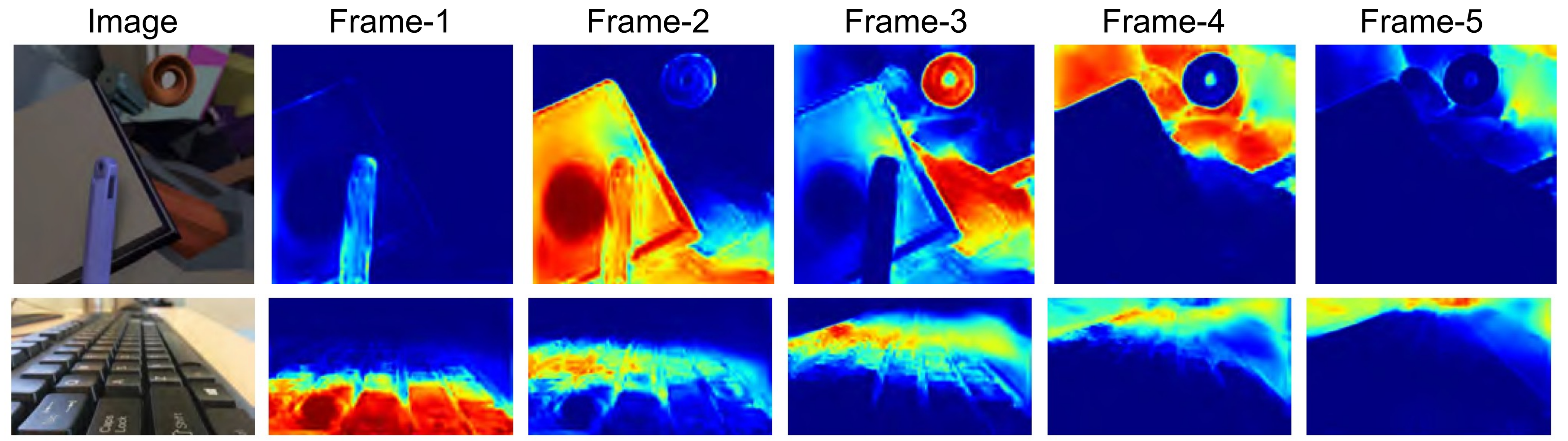


Figure 4. Focus probability visualization on FoD500 (top) and Mobile depth (bottom).

Code and pre-trained model are available at: https://github.com/fuy34/DFV

Acknowledgement: This work is supported in part by NSF award #1815491.