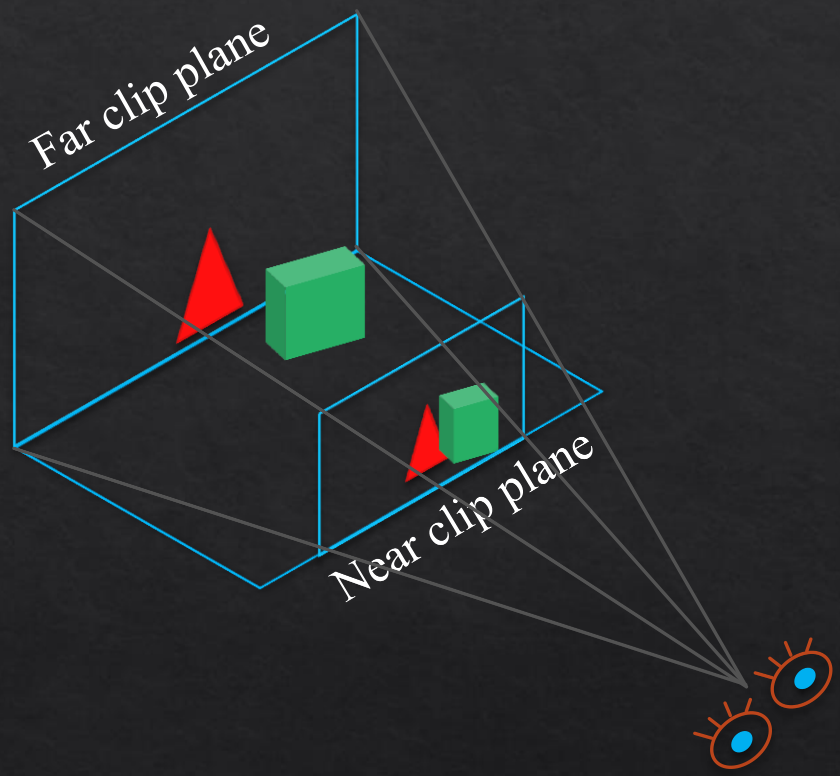
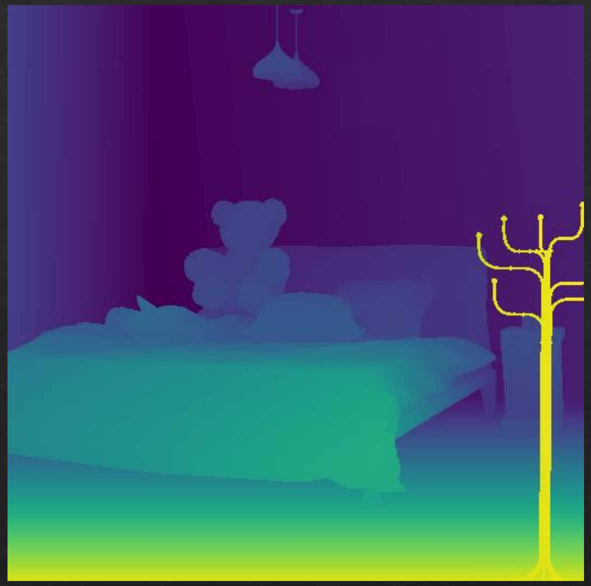
MANet: Multi-scale Aggregated Network for Light Field Depth Estimation

Yan LI[1], Lu Zhang[2], Qiong WANG[3], Gauthier LAFRUIT[1]
[1]University Libre de Bruxelles, Belgium
[2]INSA Rennes, France
[3]Zhejiang University of Technology, China

Depth estimation

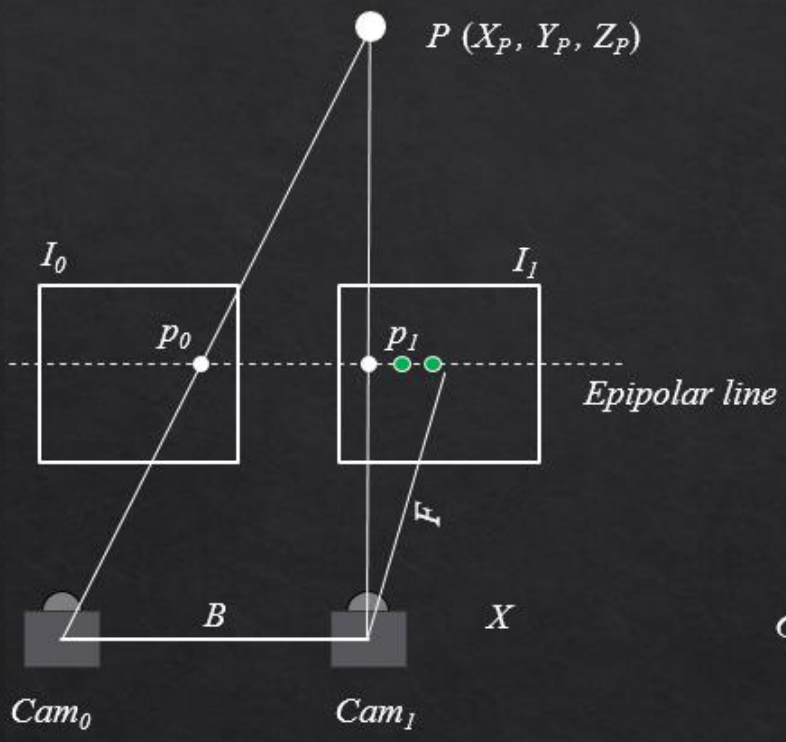


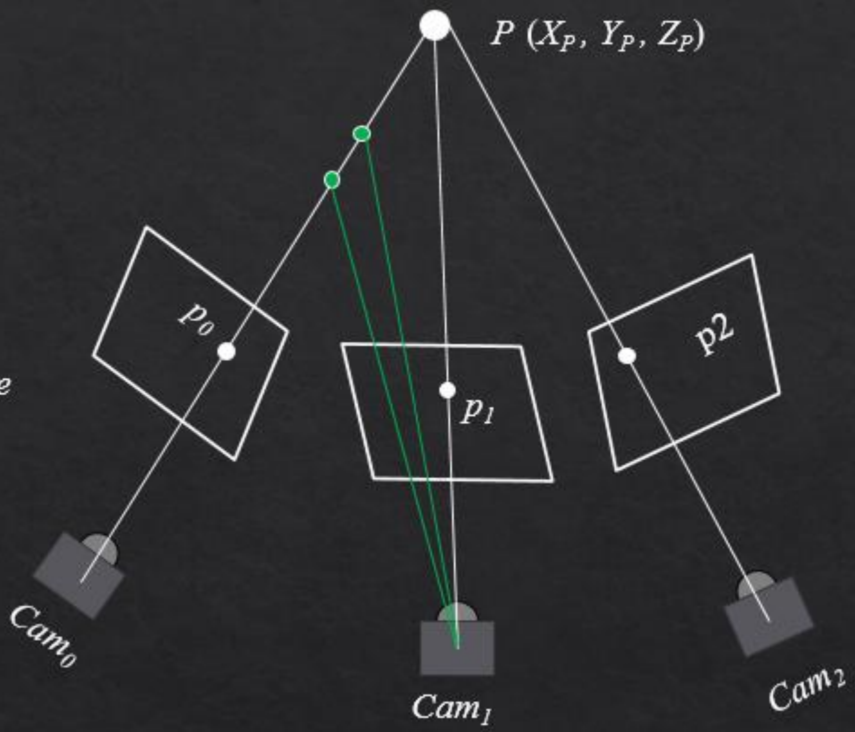
Camera view

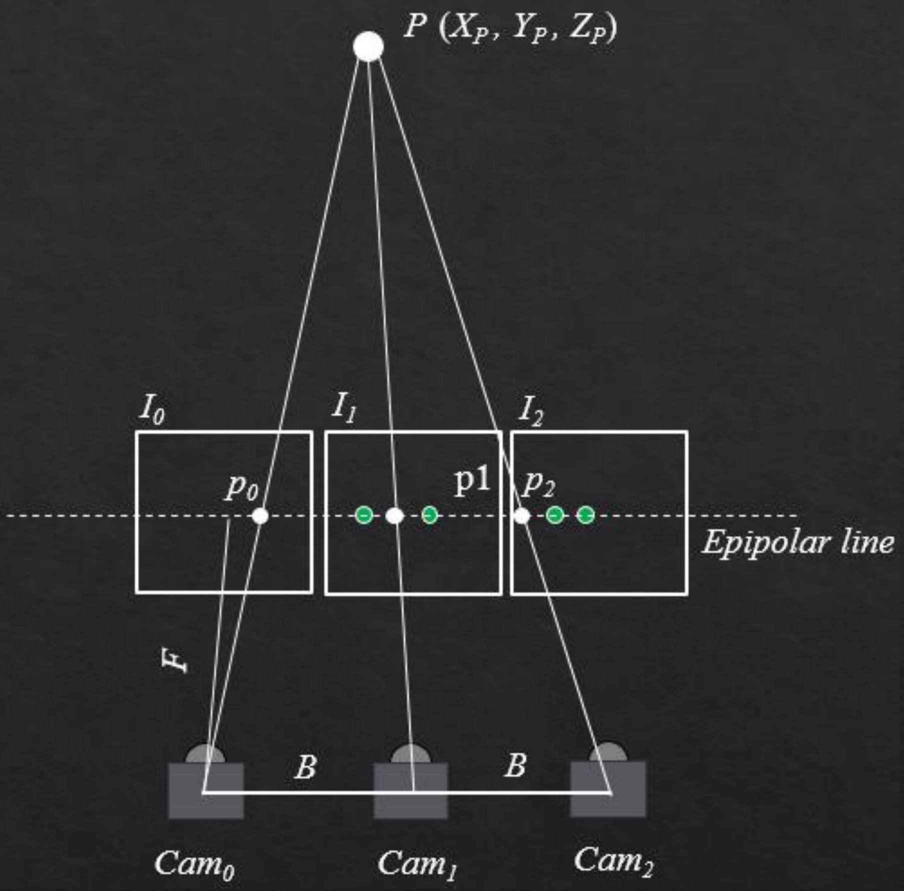


Depth map

Passive depth sensing ➡ Estimate disparity - find pixel offset







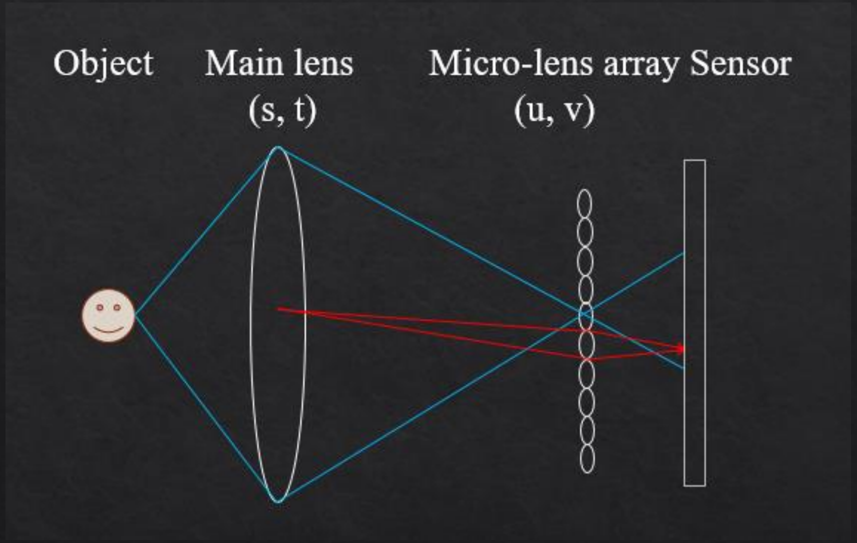
Two-view stereo

Multi-view stereo

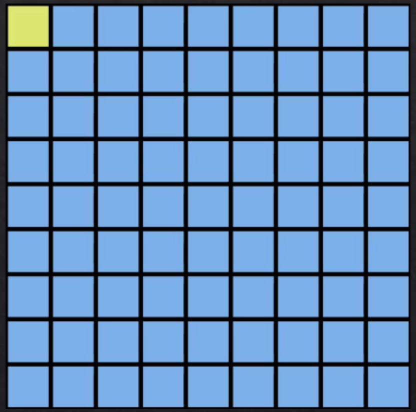
Light fields ✔

Light field camera

Lytro Illum camera[1]



Sub-aperture views



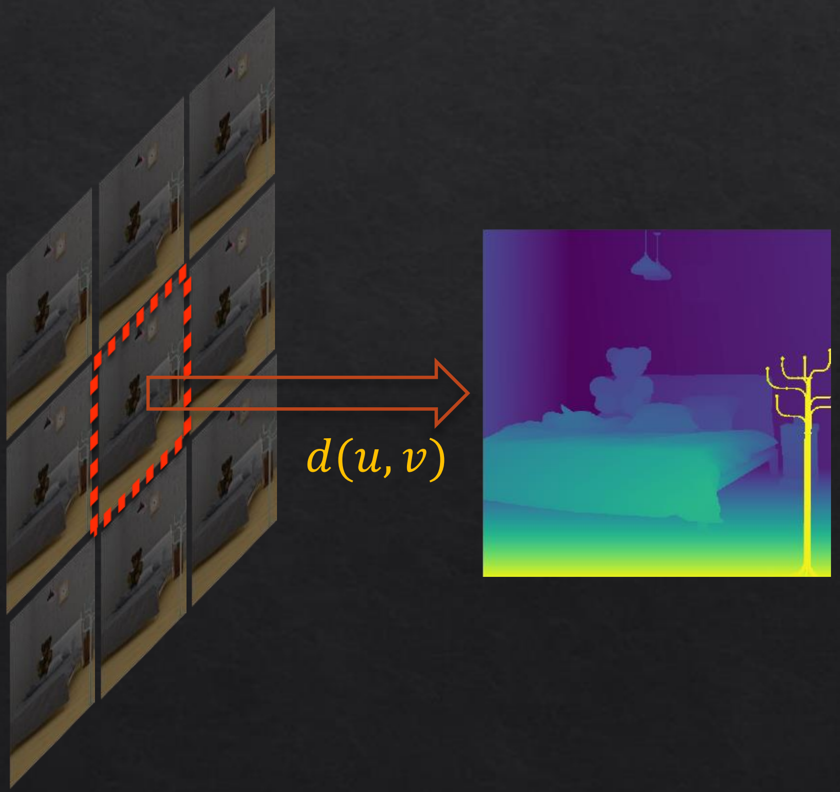
Viewpoints

[1]From google image

Light field depth estimation

The central view and other views in light fields have the following relationship:

$$L(u, v, s, t) = L(u + (s^* - s)d(u, v), v + (t^* - t)d(u, v), s^*, t^*), (s \in [1, M], t \in [1, N])$$



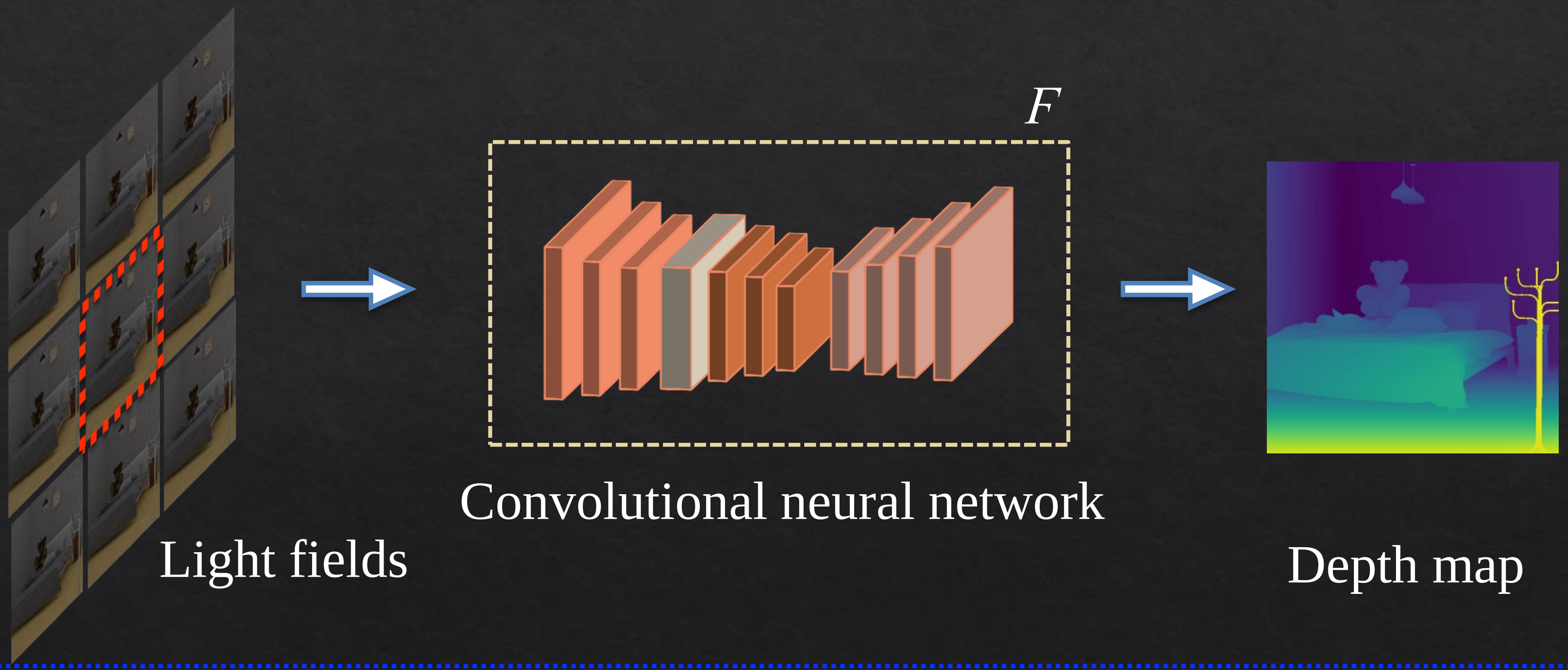
(s^*, t^*): the coordinate of the central view

$d(u, v)$: the disparity of the pixel (u, v) in the central view

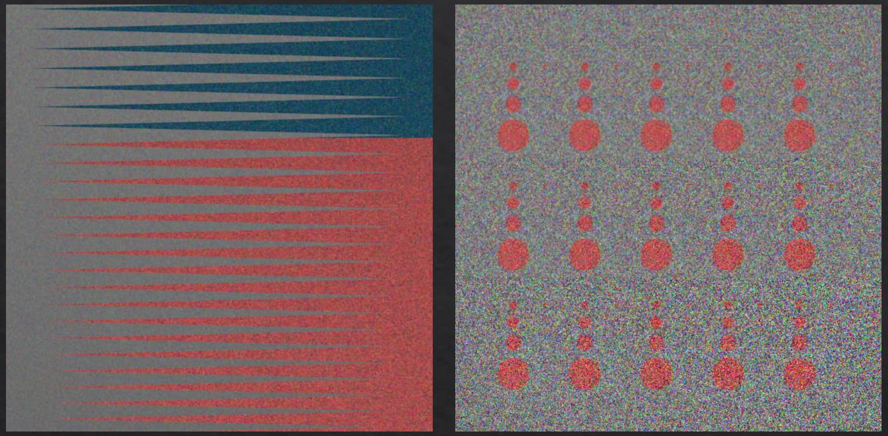
M, N : the number of views along the horizontal and vertical directions in light fields

Traditional → Deep learning ✓ High accuracy and efficiency

Deep learning for light field depth estimation



Issue 1 – Photorealistic contents



Non-photorealistic scene

Photorealistic scene

Real world

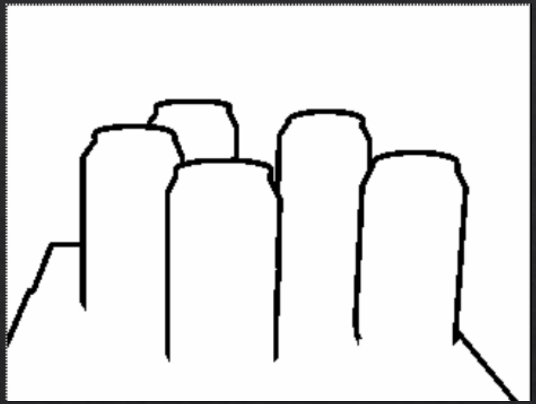
Important !

Multiple everyday objects

Difficult !

Issue 2 – depth discontinuity preserving

Scene



Depth discontinuity

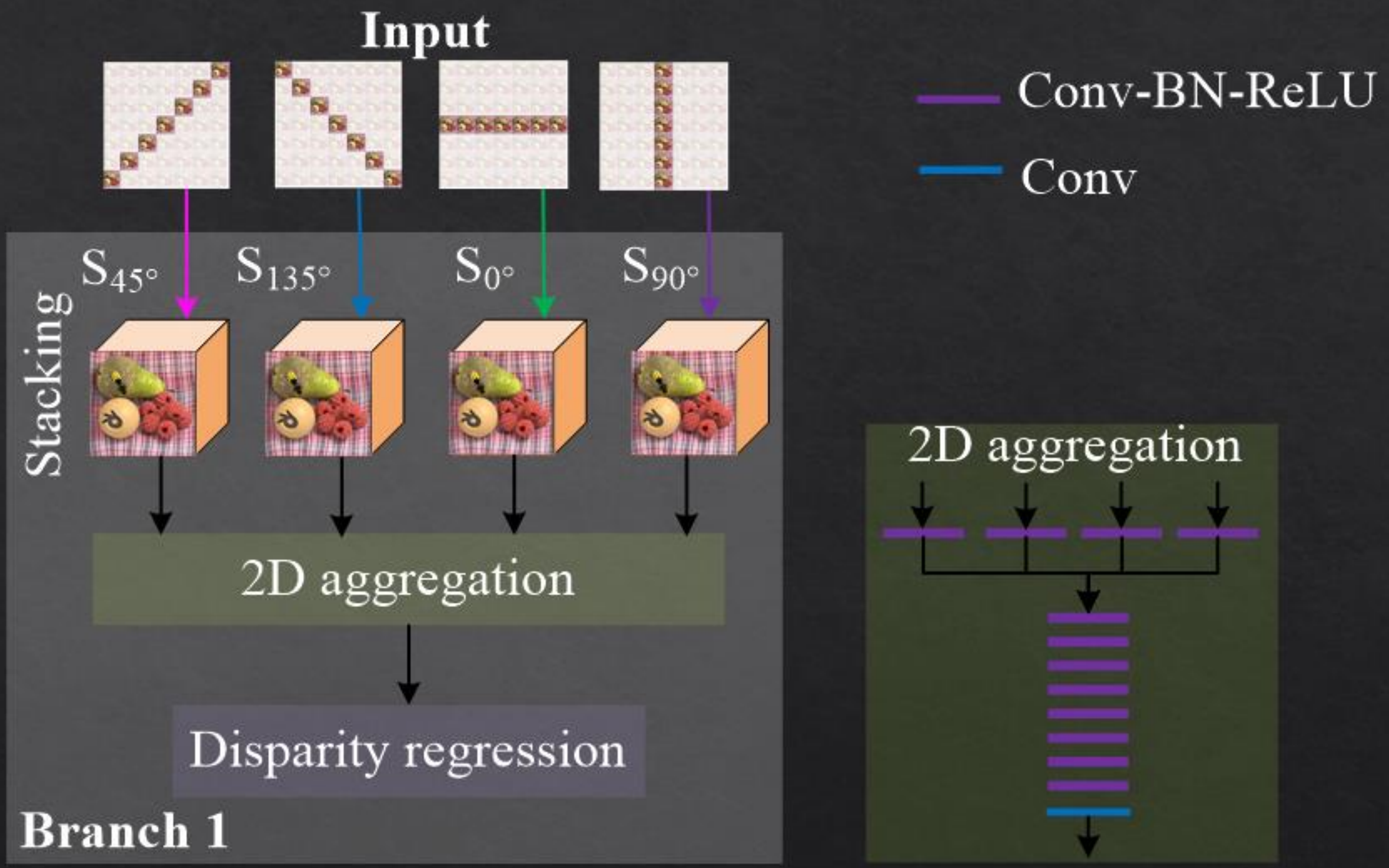
Occlusions
Boundaries of objects

Important !

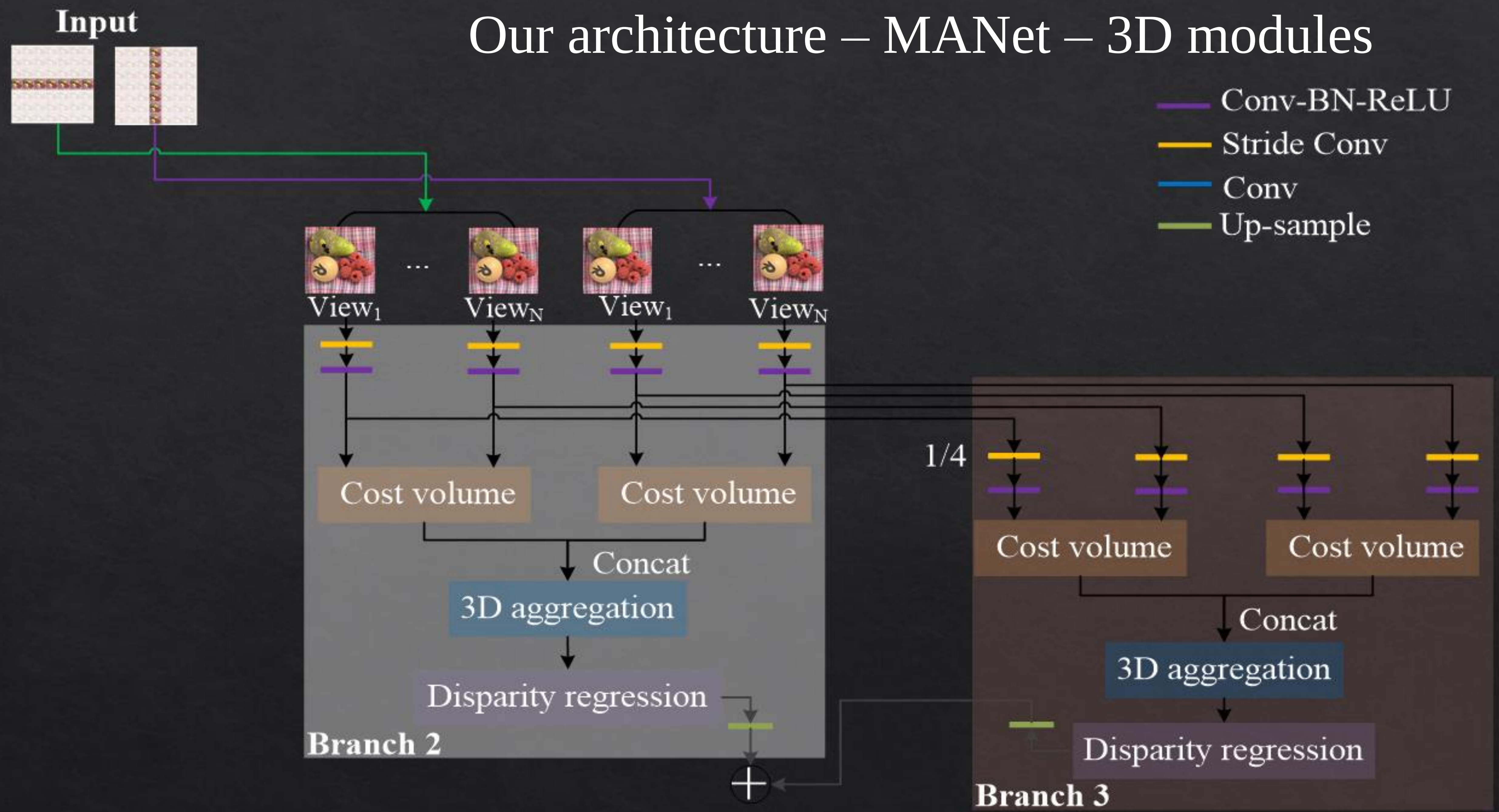
Difficult !

Keep discontinuity but noisy Or Too smooth less noisy ⟶ Trade-off

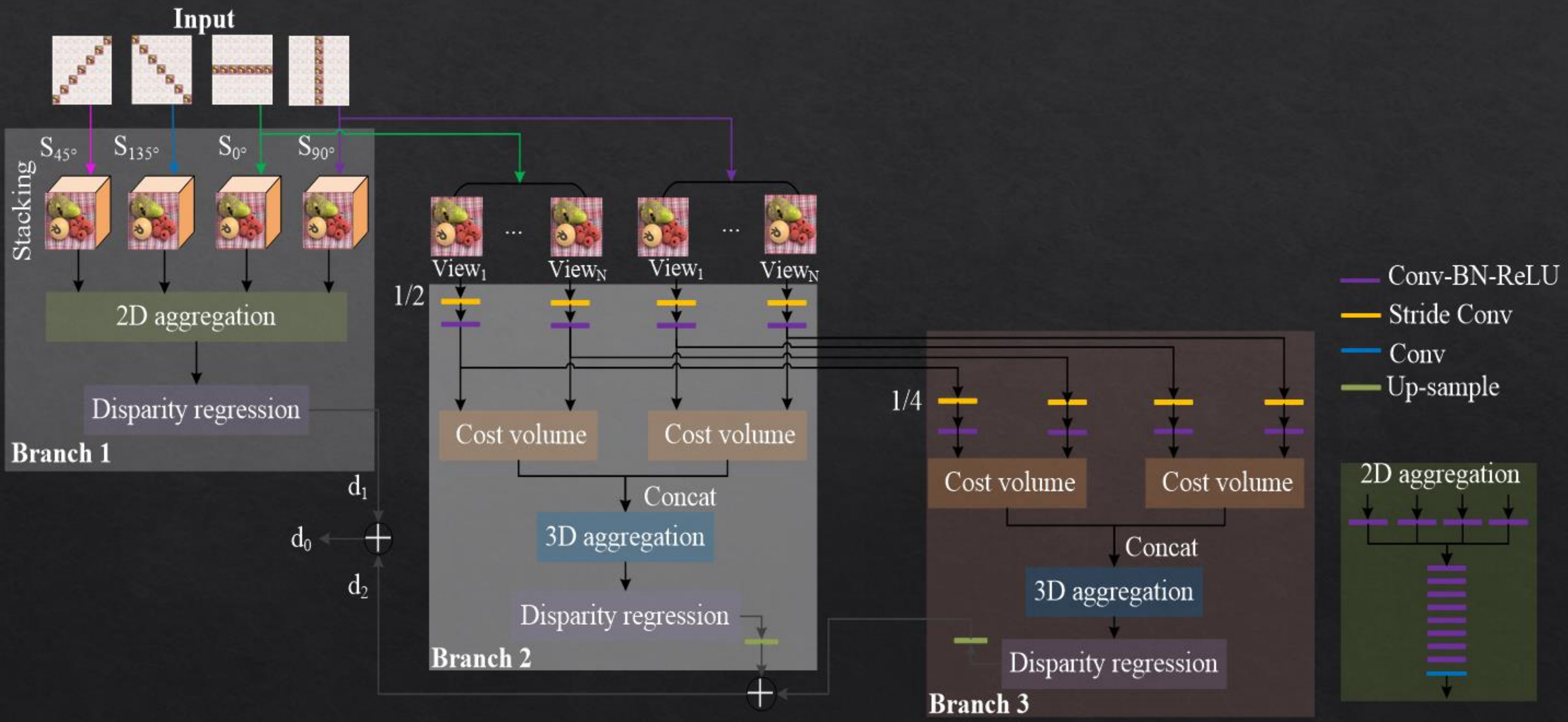
Our architecture – MANet – 2D modules



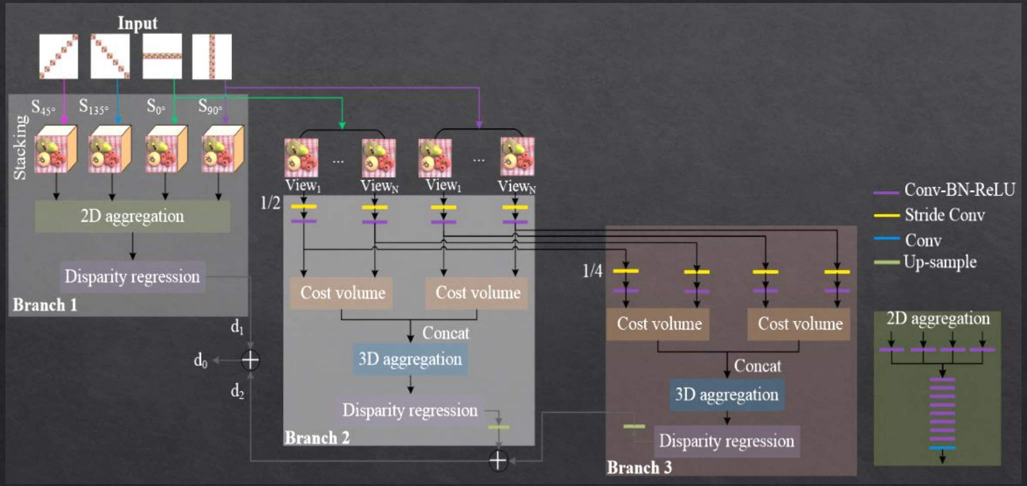
Our architecture – MANet – 3D modules



Our architecture - MANet



Ours vs Epinet Architecture - context information



Epinet

Ours: +deep low-scale features

↓

Bring in large context information

↓

Photorealistic

Input

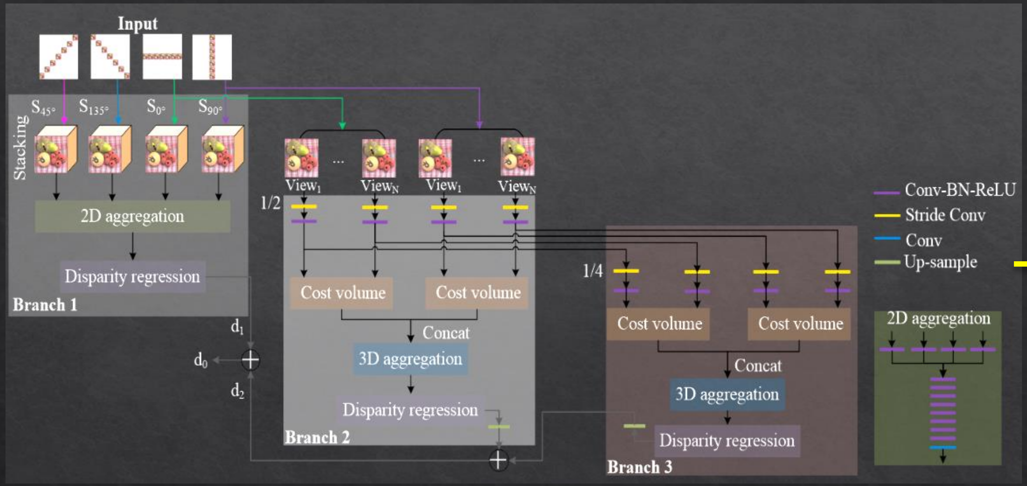
↓

full-scale → 1/2-scale → 1/4-scale

Output1 Output2 Output3

Large context information

Ours vs Epinet Architecture - Coarse to fine fusion



Epinet

Ours: Coarse to fine fusion
Trade-off

Discontinuity preserving and less noisy

Input

full-scale → 1/2-scale → 1/4-scale

Output1 Output2 Output3

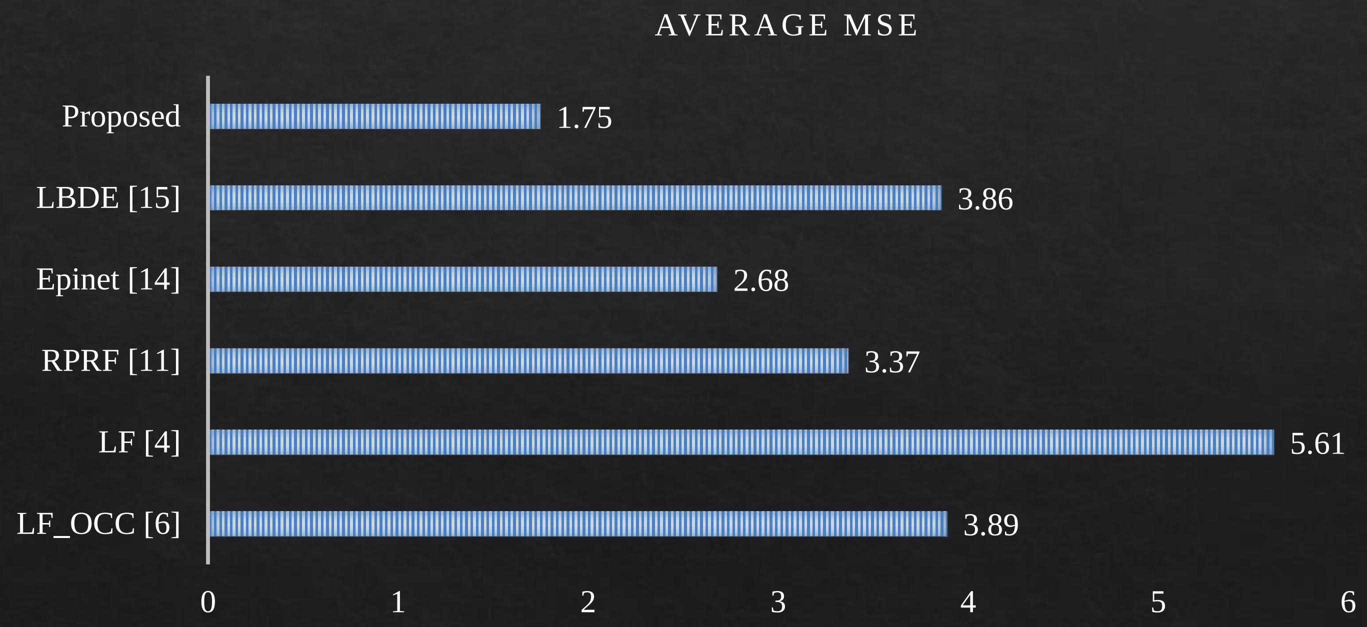
← + ← + ←

Discontinuity but noisy

Too smooth less noisy

Accuracy - Quantitative comparison: Depth map (DM) Vs Ground Truth (GT)

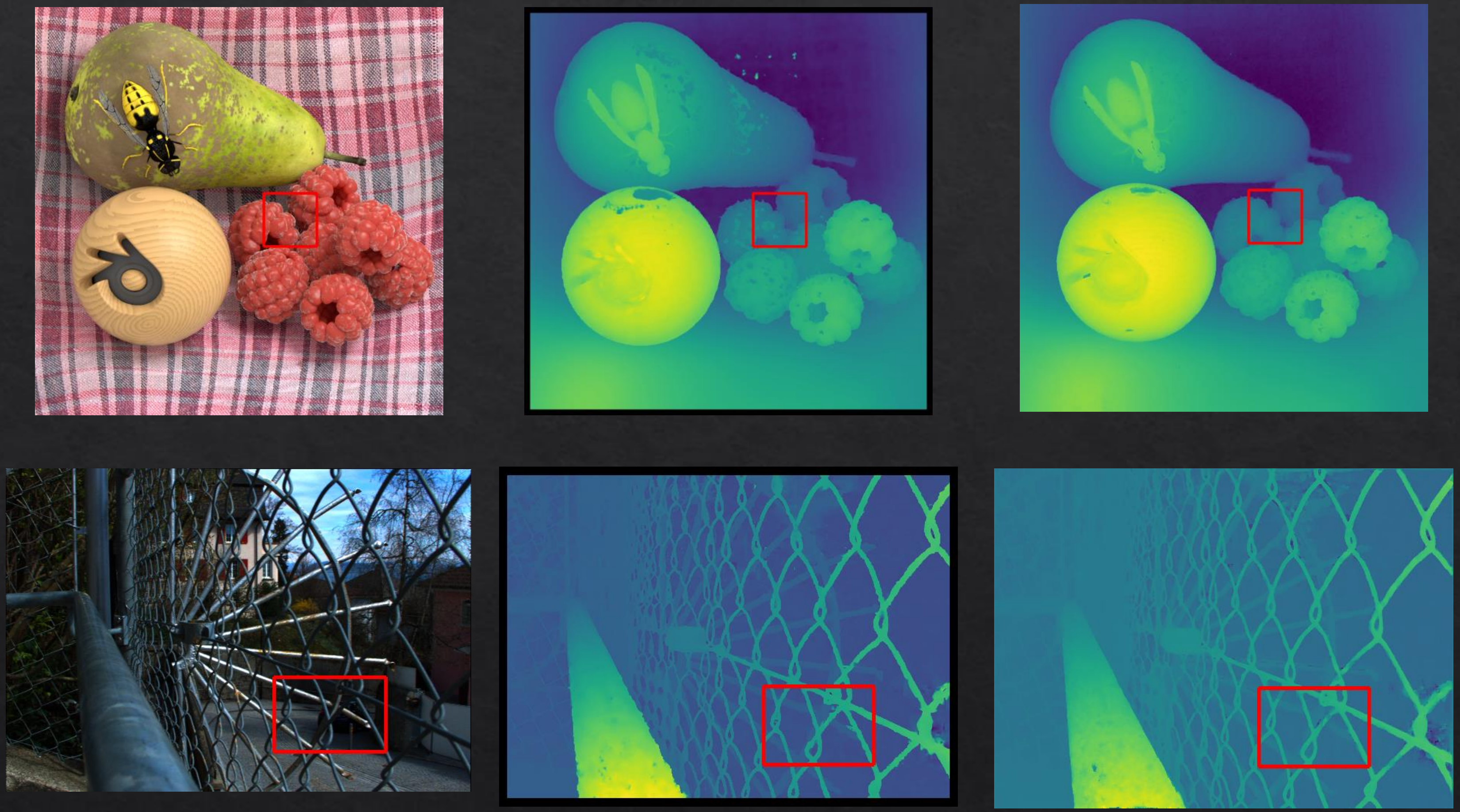
$$\text{Mean Square Error (MSE)} = \frac{1}{h_1 \times w_1} \sum_{i=1}^{h_1 \times w_1} (GT(i) - DM(i))^2$$



Compared traditional methods:
[4] LF (ICCV 2015)
[6] LF_OCC (CVPR 2015)
[11] RPRF (ICCV 2017)

Compared deep-learning methods:
[14] Epinet (CVPR 2018)
[15] LFDE (ICASSP 2019)
Compared dataset: CVIA-HCI and HCI datasets

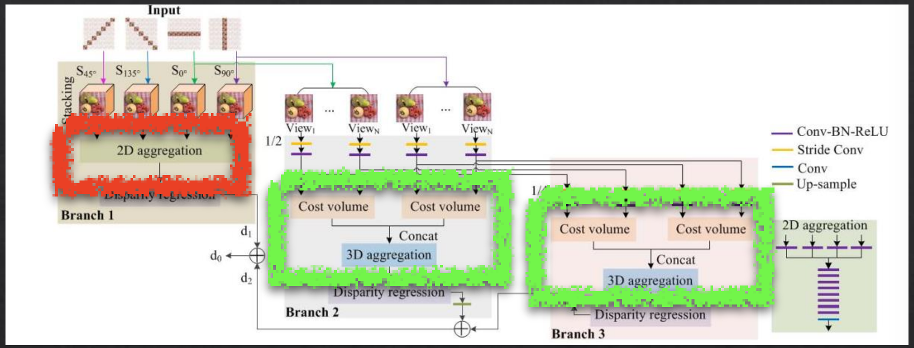
Accuracy - Visual comparison

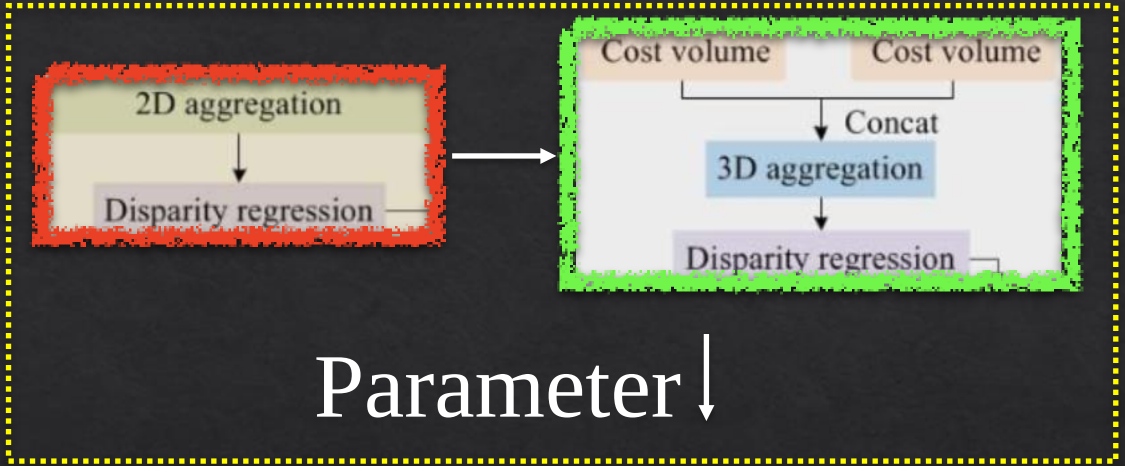


Center view Epinet Proposed

Photorealistic
Discontinuity preserving

Parameter and Runtime





	Parameters	Time (s)
LF_OCC	-	1.05E+04
RPRF	-	34.53
Epinet	5.12M	1.98
LBDE	199M	1.92
Ours	**1.58M**	**0.73**

Few parameters,
high efficiency

Source code: https://github.com/YanWQ/MANet

Reference

[4] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I. So Kweon, “Accurate depth map estimation from a lenslet light field camera,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 1547–1555.

[6] T.-C. Wang, A. A. Efros, and R. Ramamoorthi, “Occlusion-aware depth estimation using light-field cameras,” in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 3487–3495.

[11] C. Huang, “Robust pseudo random fields for light-field stereo matching,” in IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22- 29, 2017, 2017, pp. 11–19.

[14] C. Shin, H.-G. Jeon, Y. Yoon, I. So Kweon, and S. Joo Kim, “Epinet: A fully-convolutional neural network using epipolar geometry for depth from light field images,” in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018, pp.4748–4757.

[15] X. Jiang, J. Shi, and C. Guillemot, “A learning based depth estimation framework for 4d densely and sparsely sampled light fields,” in 2019 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2019, 2019, pp. 2257–2261.

Thanks for your attention!