CSE 152: Computer Vision

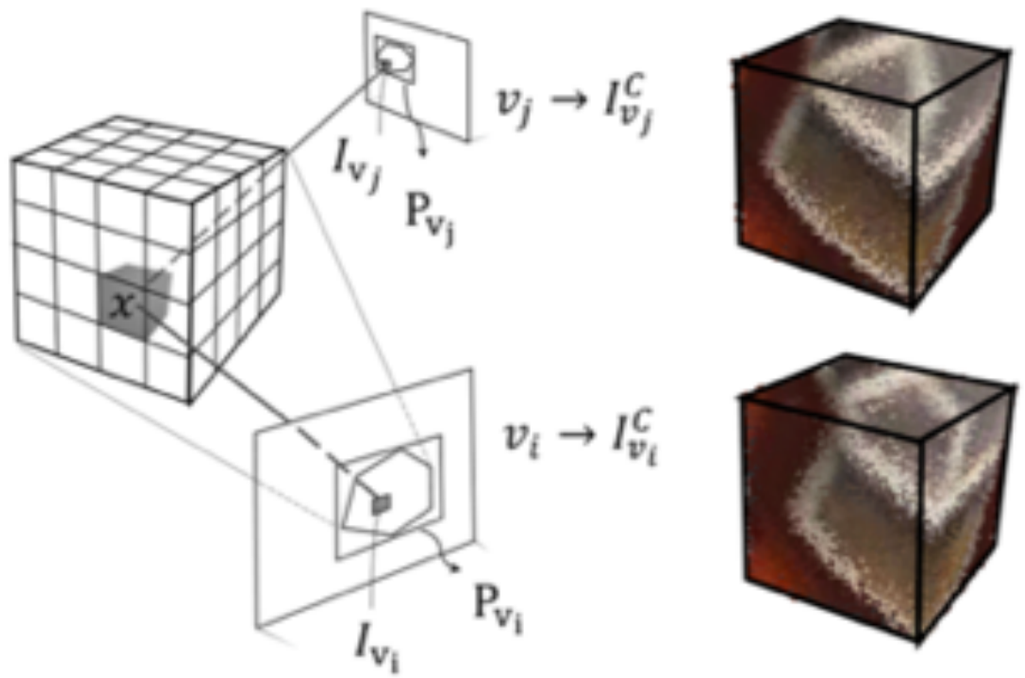
Hao Su

Lecture 14: Deep Multi-View Stereo

Some slides in this lecture are courtesy to Prof. Silvio Savarese

Surface Reconstruction as Voxel Occupancy Prediction

Unprojection along viewing rays to build colored voxel cubes.



Ji et al., “**SurfaceNet: An End-to-End 3D Neural Network for Multiview Stereopsis**”, *ICCV 2017*

Predict the surface confidence for each voxel:

$$L(I^C_{v_i}, I^C_{v_j}, \hat{S}^C) =$$
$$-\sum_{x \in C} \{\alpha \hat{s}_x \log p_x + (1-\alpha)(1-\hat{s}_x)\log(1-p_x)\}$$

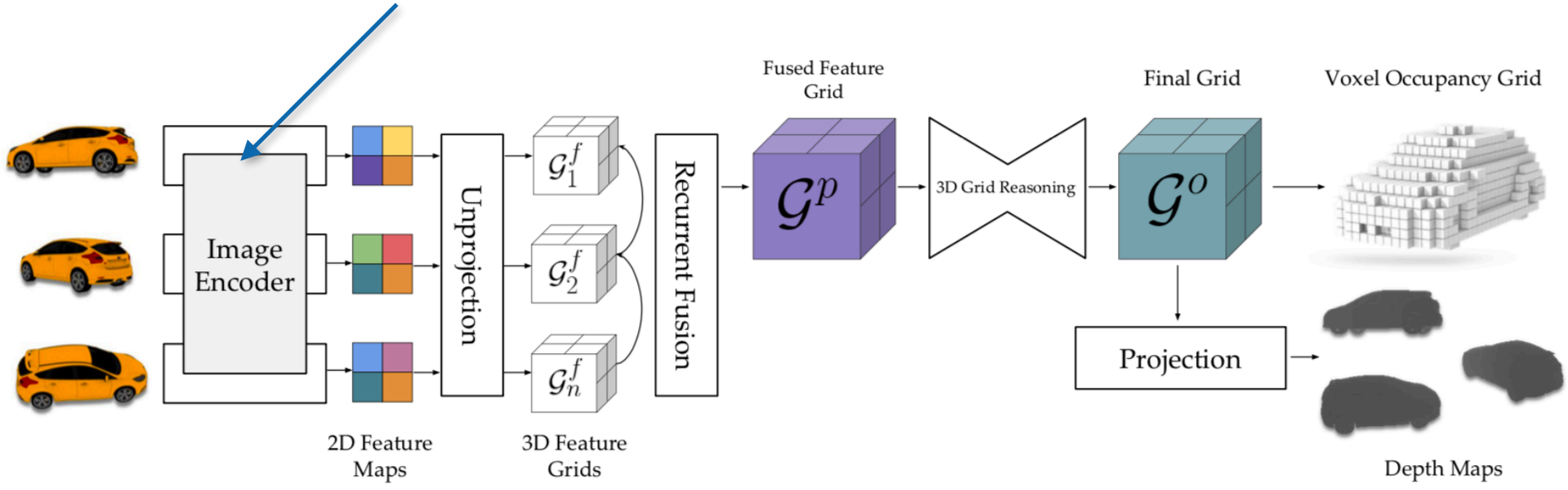
Ji et al., "**SurfaceNet: An End-to-End 3D Neural Network for Multiview Stereopsis**", *ICCV 2017*

Limitations:

- Pre-computed grids can only take RGB colors at coarse resolution
- Voxel binarization introduces quantization errors.

Learning-Based Stereopsis

- End-to-end learning of deep features for each pixel.

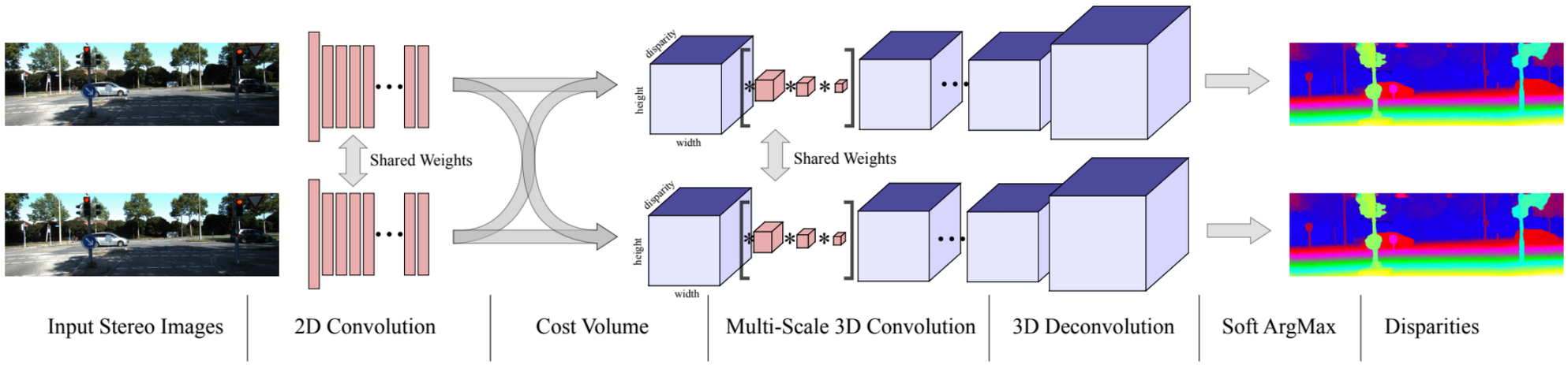


Kar et al., "**Learning a Multi-View Stereo Machine**",
NeurIPS 2017

Still very coarse resolution (32x32x32) due to volumetric representation.

Kar et al., “**Learning a Multi-View Stereo Machine**”,
NeurIPS 2017

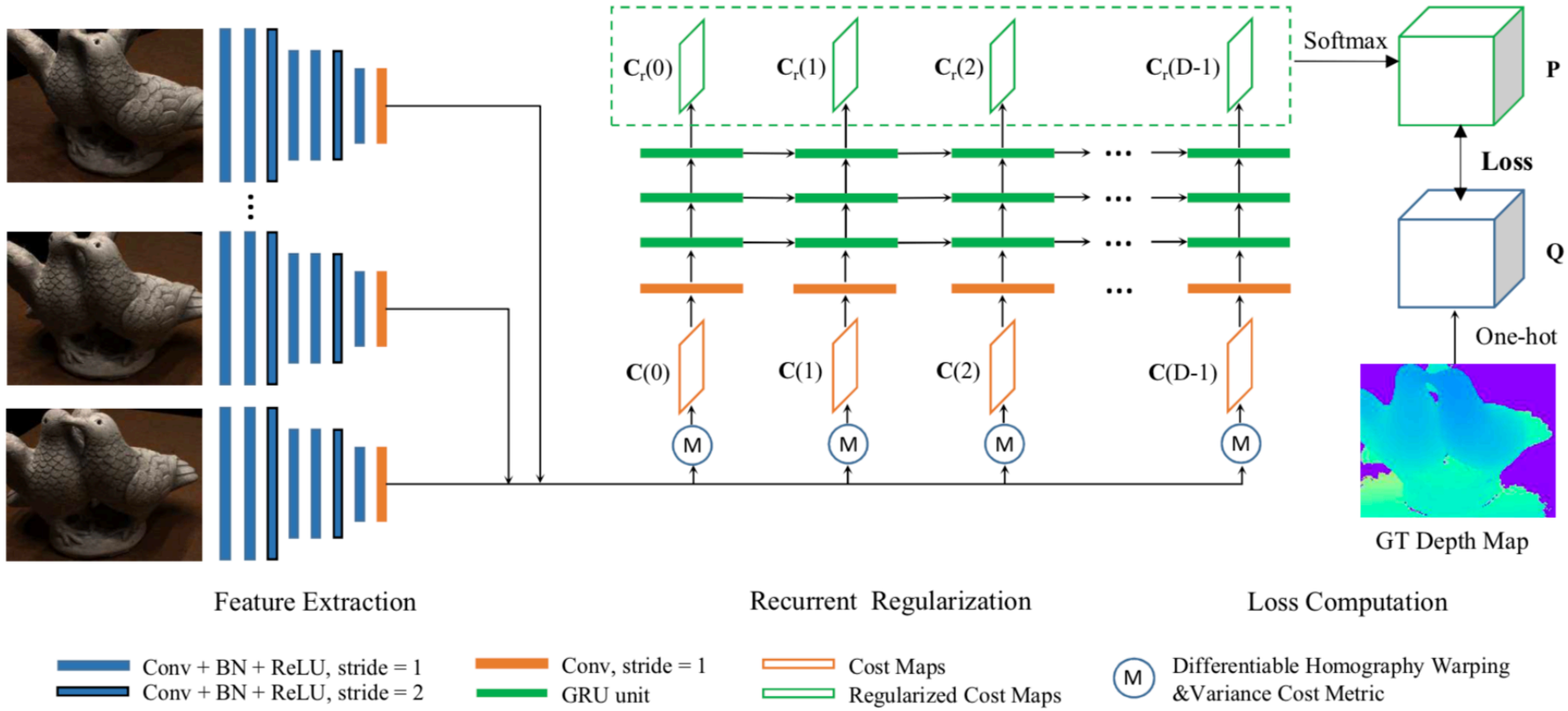
- Differentiable soft-argmin to achieve sub-pixel accuracy.
- View-aligned cost-volume construction.



$$soft\ argmin := \sum_{d=0}^{D_{max}} d \times \sigma(-c_d)$$

Kendall et al., "**End-to-End Learning of Geometry and Context for Deep Stereo Regression**", *ICCV 2017*

Idea 1: Slide-by-Slide Processing of Cost Volume by Recurrent Neural Network



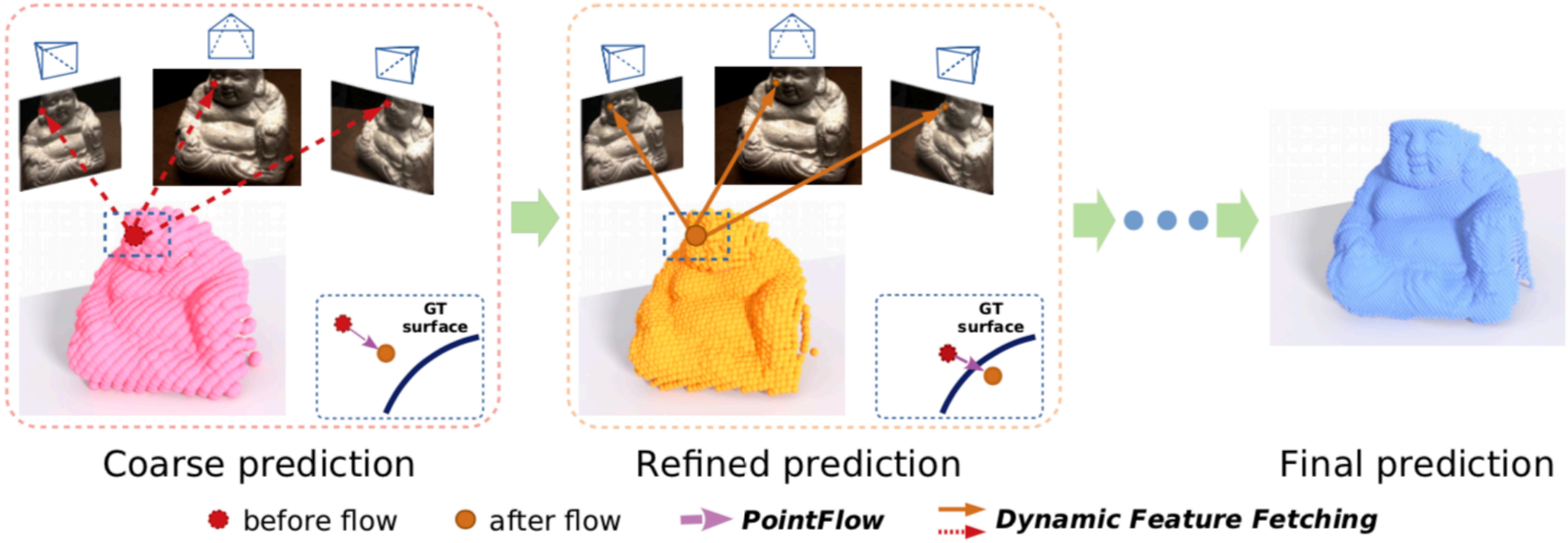
The cost volume is sequentially regularized along the depth direction.

Yao et al., "**MVSNet: Depth Inference for Unstructured Multi-view Stereo**", *ECCV 2018*

Yao et al., "**Recurrent MVSNet for High-resolution Multi-view Stereo Depth Inference**", *CVPR 2019*

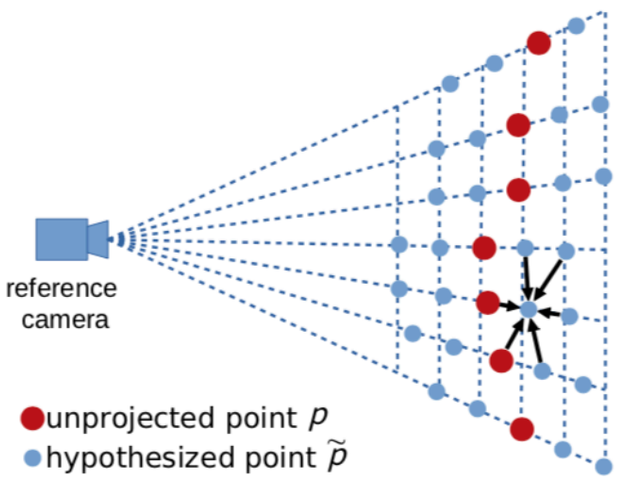
Idea 2: Point-based MVS

- Point-based representation for computational efficiency.
- Iteratively update the location of points and spawn more points.
- More flexible and accurate.

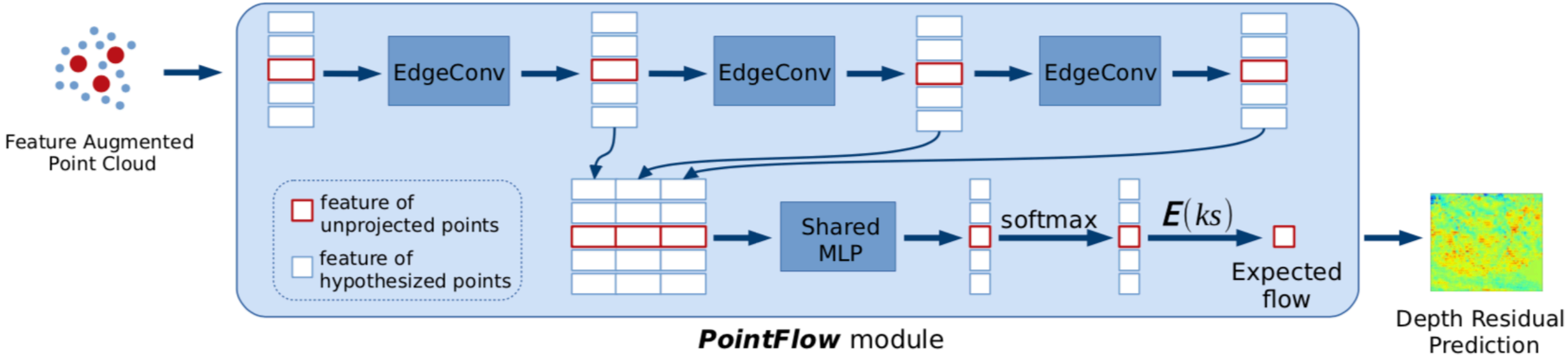


Chen et al., “**Point-Based Multi-View Stereo Network**”, *ICCV 2019*

hypothesized point generation
along camera direction:



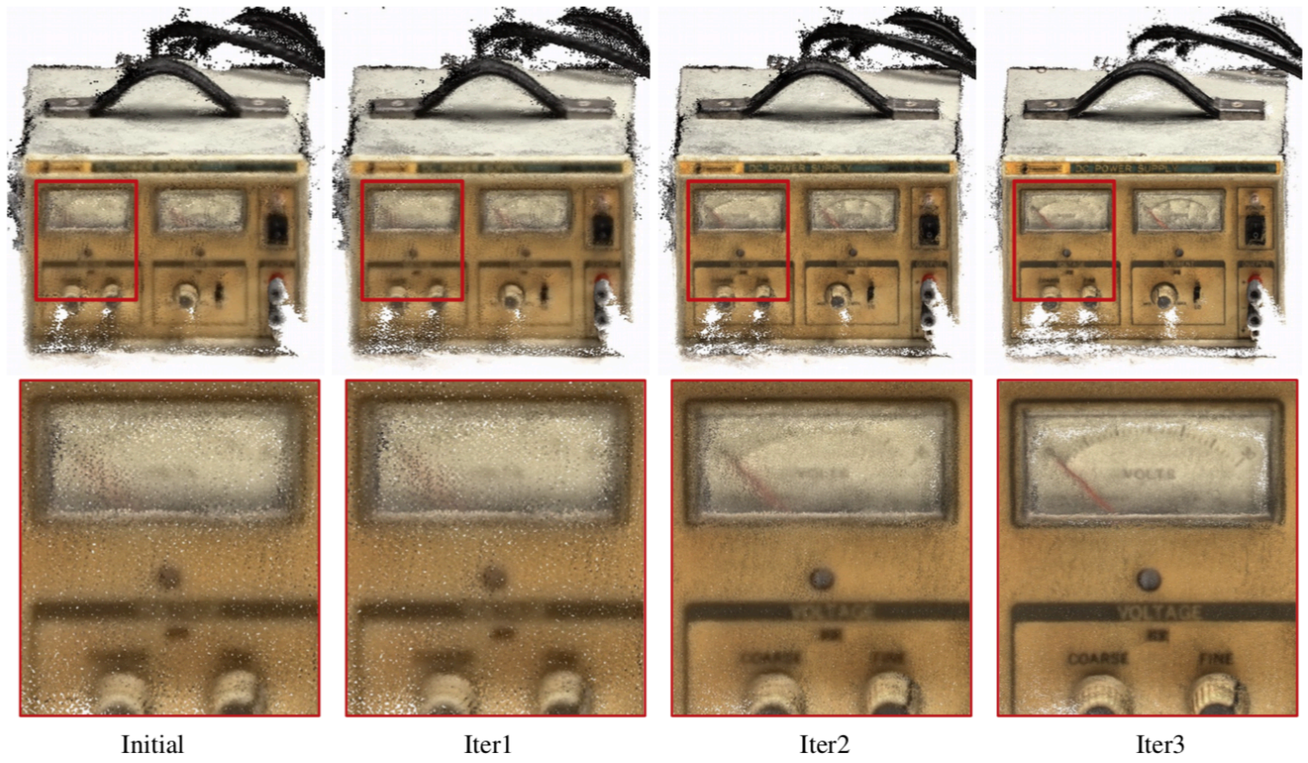
- PointFlow: iteratively pull all the points to the right position.



PointFlow module

Chen et al., "**Point-Based Multi-View Stereo Network**", *ICCV 2019*

Iterative refinement:



Results on DTU benchmark

Iter.	Acc. (mm)	Comp. (mm)	Overall (mm)	0.5mm *f-score*	Depth Map Res.	Depth Interval (mm)	GPU Mem. (MB)	Runtime (s)
-	0.693	0.758	0.726	47.95	160×120	5.30	**7219**	**0.34**
1	0.674	0.750	0.712	48.63	160×120	5.30	7221	0.61
2	0.448	0.487	0.468	76.08	320×240	4.00	7235	1.14
3	**0.361**	**0.421**	**0.391**	**84.27**	**640×480**	**0.80**	8731	3.35
MVSNet[29]	0.456	0.646	0.551	71.60	288×216	2.65	10805	1.05

Chen et al., “**Point-Based Multi-View Stereo Network**”, *ICCV 2019*

Learning for SfM

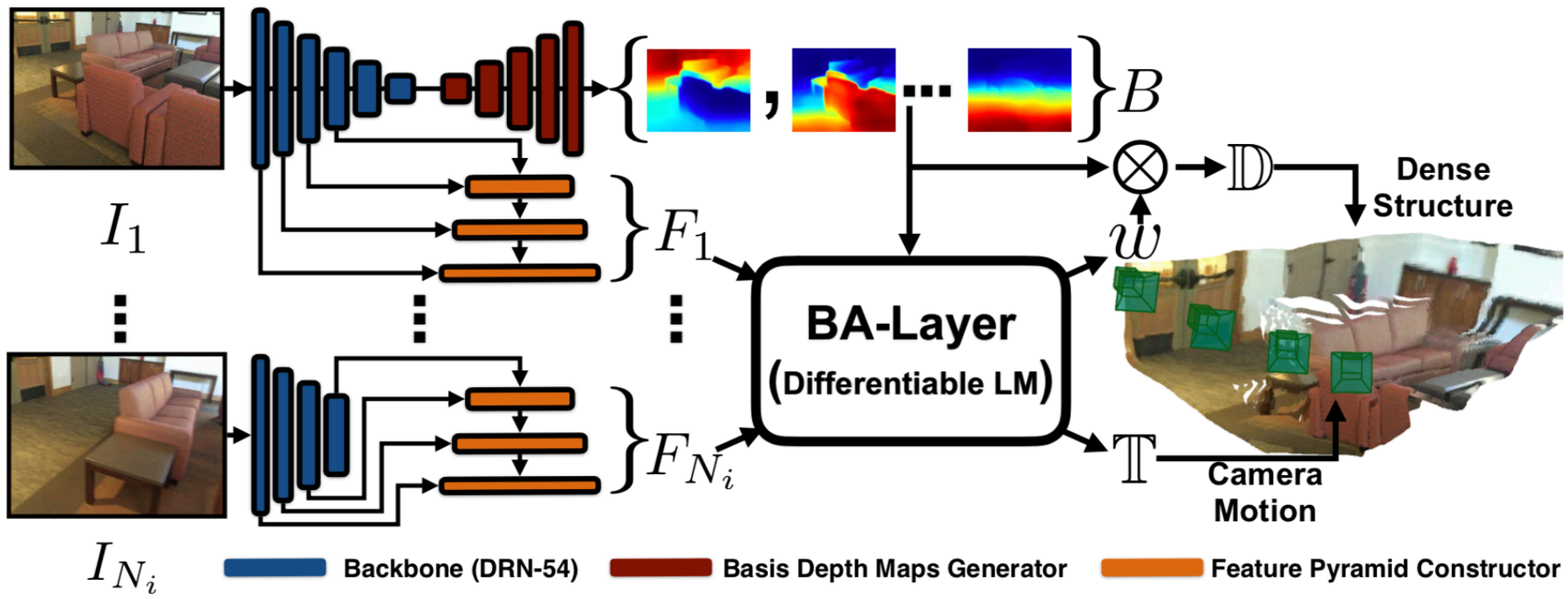
- Above learning-based MVS methods all **assume relative camera pose**

- What if not?
  - Classic 3D: Bundle Adjustment

- Learning-based bundle adjustment

BA-Net

End-to-end pipeline for SfM with differentiable bundle adjustment.

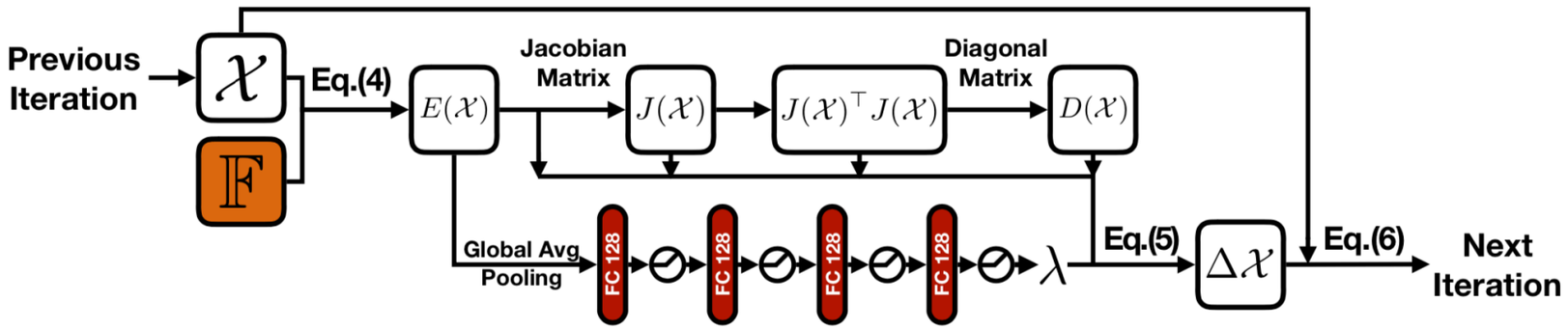


Tang et al., "**BA-Net: Dense Bundle Adjustment Network**", *ICLR 2019*

Differentiable LM algorithm:

- Iterative update as rollout of network layers
- Use network to predict the damping factor lambda.

BA-layer:



$$\Delta\mathcal{X} = (J(\mathcal{X})^\top J(\mathcal{X}) + \lambda D(\mathcal{X}))^{-1} J(\mathcal{X})^\top E(\mathcal{X}).$$

Tang et al., "**BA-Net: Dense Bundle Adjustment Network**", *ICLR 2019*