3D Vision: From Stereo Matching to Point Cloud Understanding

Yulan Guo

GAMES

2020·12·03

3D Vision: From Stereo Matching to Point Cloud Understanding

Supervised Stereo Matching

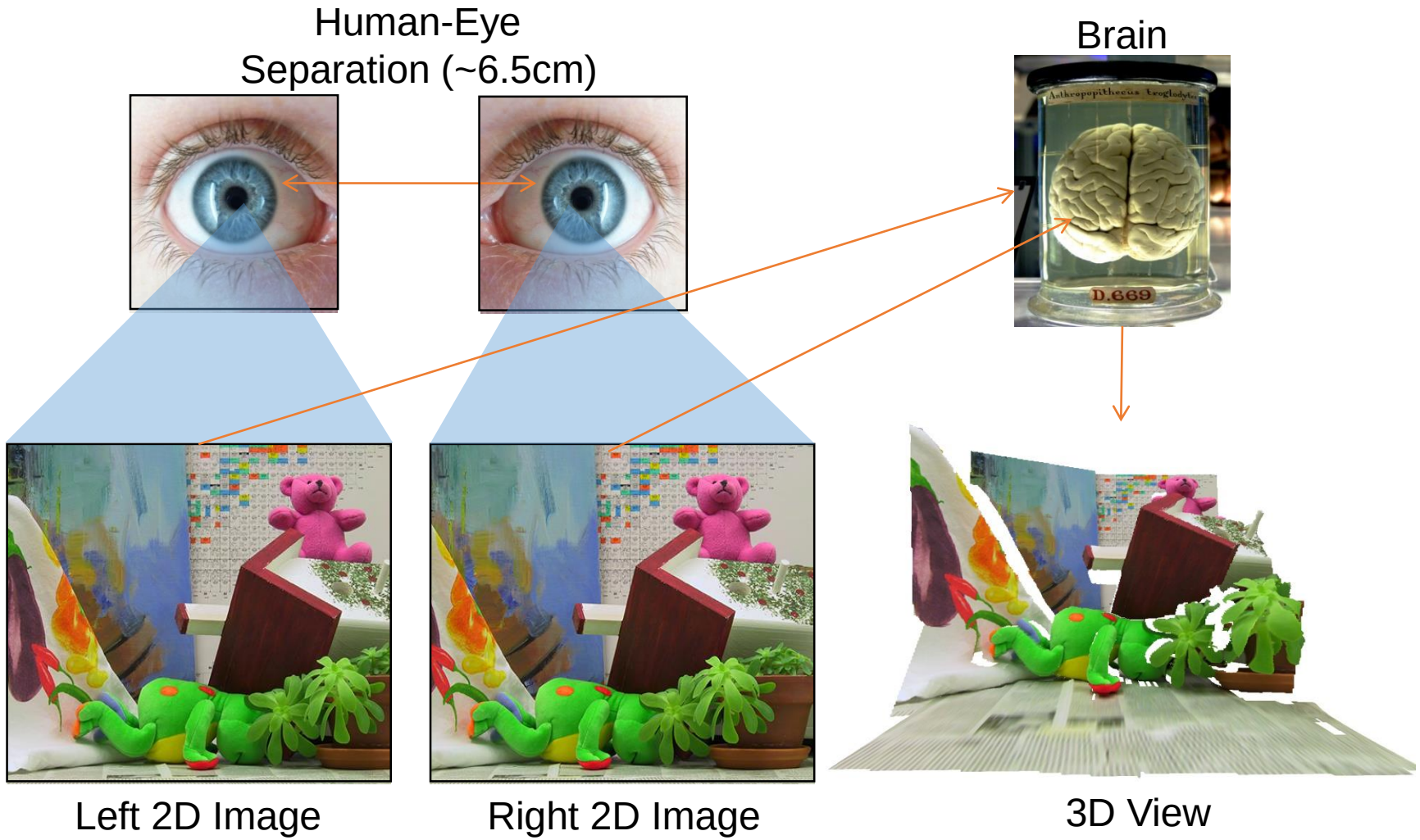
Unsupervised Stereo Correspondence Learning

Semantic Segmentation of Large-scale Point Clouds

01

Supervised Stereo Matching

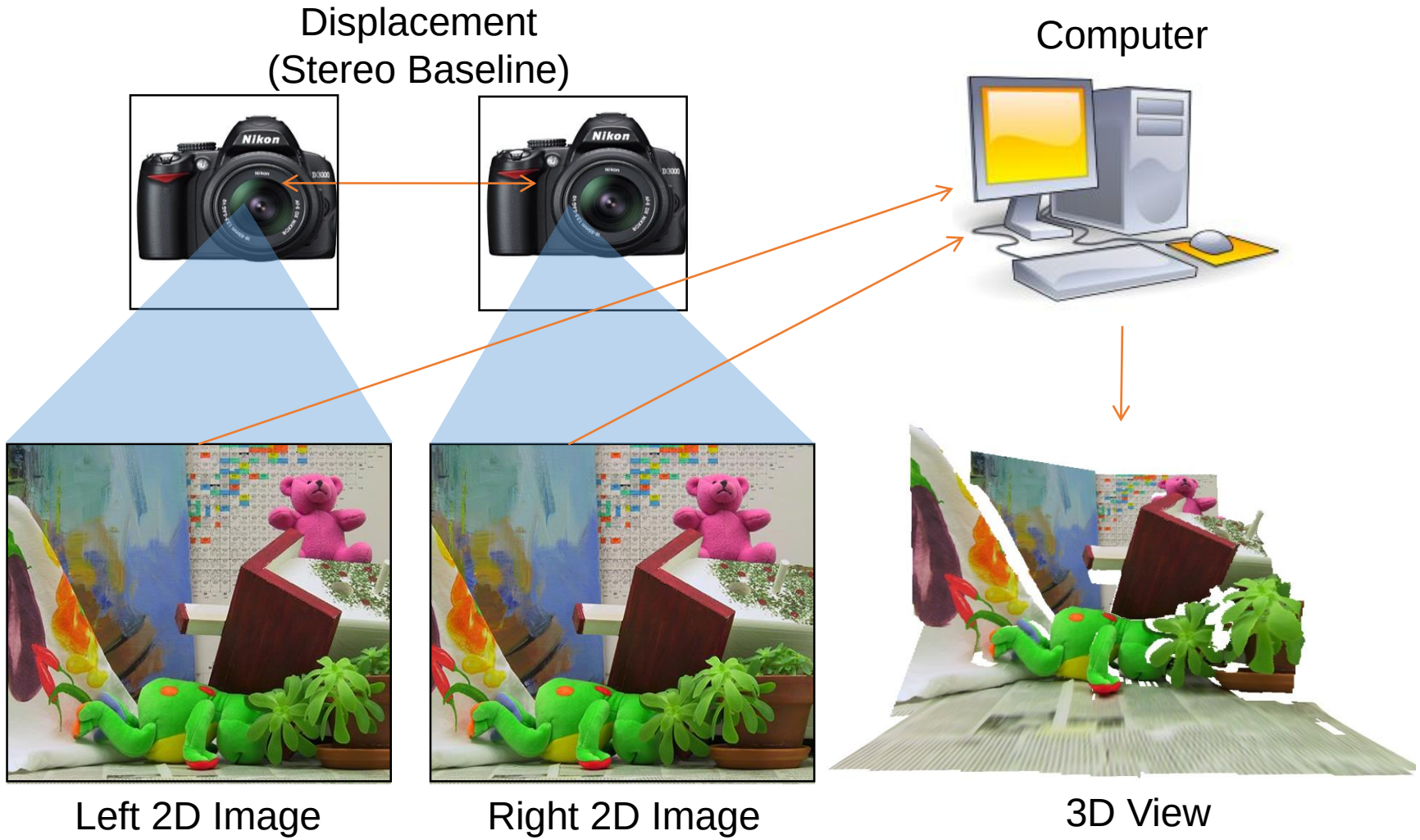
Background



01

Supervised Stereo Matching

Background



01

Supervised Stereo Matching

Related Work

Matching Cost Computation

- Distances: L1, L2
- Correlation: NCC， ZNCC
- Non-parametric measures: rank and census transforms

Cost Aggregation

- Square window
- Gaussian
- 3D aggregation
- Shiftable window

Disparity Computation

- Local methods: winner-take-all (WTA)
- Global methods: an energy minimization framework
  - dynamic programming, max-flow and graph-cut
  - cooperative algorithms

Disparity Refinement

- Subpixel enhancement: iterative gradient descent, curve fitting
- Median filter
- Bilateral filter

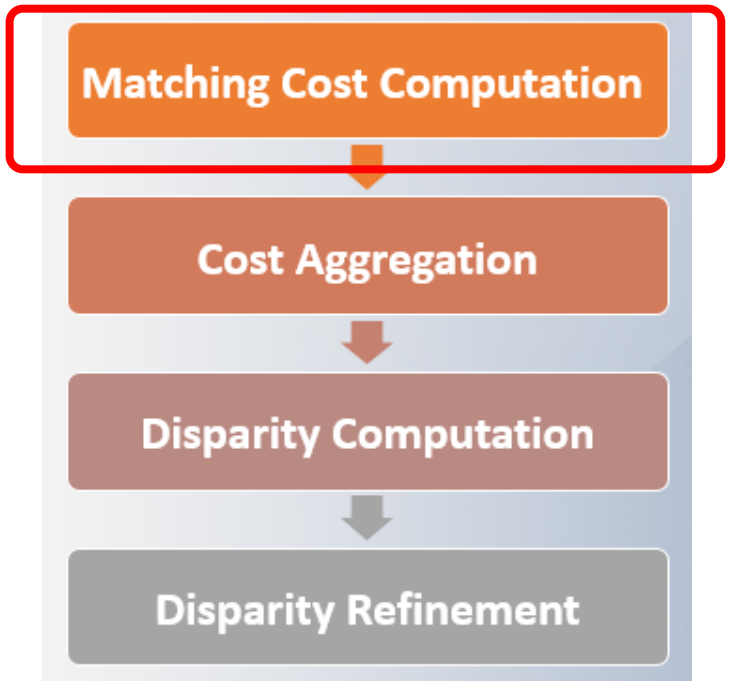
Scharstein & Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. IJCV, 2002, 47, 7-42

01 Supervised Stereo Matching

Related Work

DL for Depth Estimation

- **Matching Cost Computation**
  - Žbontar & Lecun, CVPR 2015 & JMLR 2016
  - Luo et al, CVPR 2016
  - Shaked & Wolf, CVPR 2016

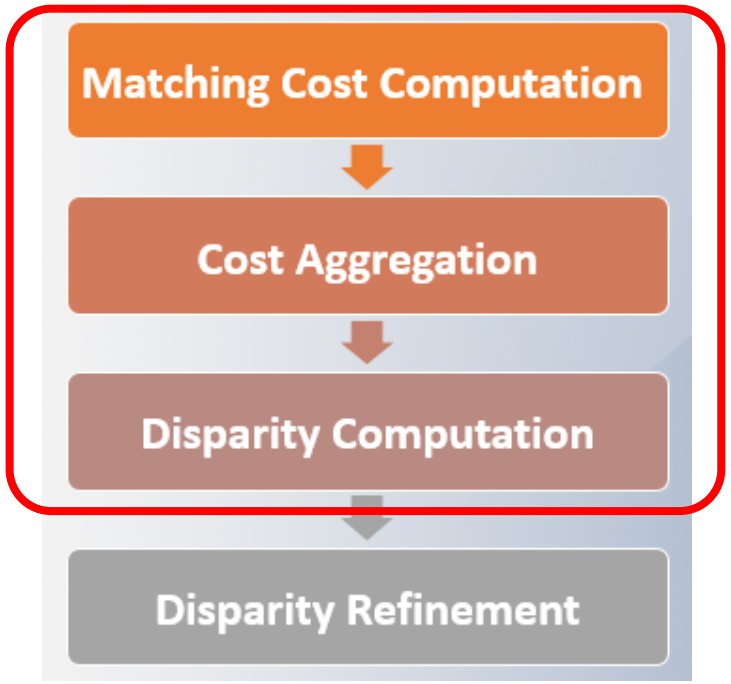


01 Supervised Stereo Matching

Related Work

DL for Depth Estimation

- **Matching Cost Computation**
  - Žbontar & Lecun, CVPR 2015 & JMLR 2016
  - Luo et al, CVPR 2016
  - Shaked & Wolf, CVPR 2016
- **3-In-1 CNN**: for Matching Cost Computation, Cost aggregation, and Disparity Computation
  - Mayer et al. CVPR 2016
  - Kendall et al. CVPR 2017

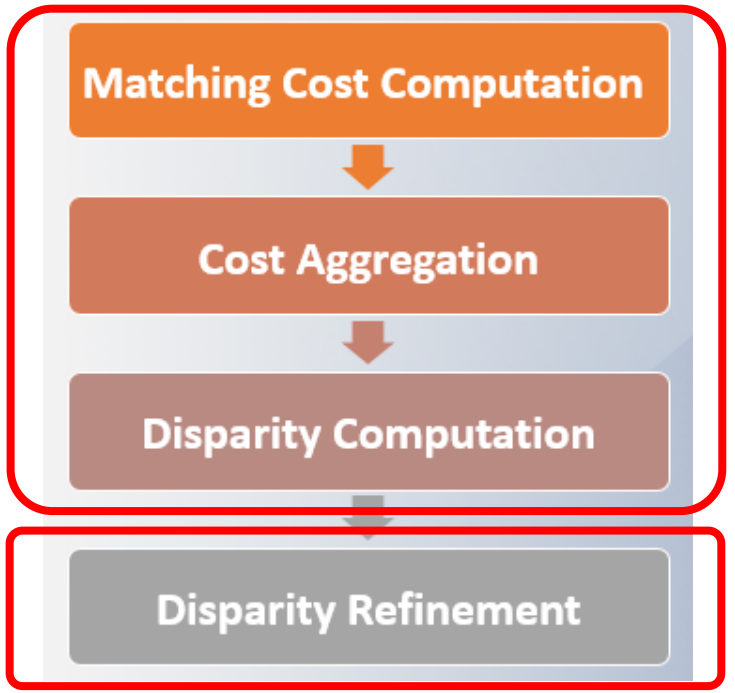


01 Supervised Stereo Matching

Related Work

DL for Depth Estimation

- **Matching Cost Computation**
  - Žbontar & Lecun, CVPR 2015 & JMLR 2016
  - Luo et al, CVPR 2016
  - Shaked & Wolf, CVPR 2016
- **3-In-1 CNN**: for Matching Cost Computation, Cost aggregation, and Disparity Computation
  - Mayer et al. CVPR 2016
  - Kendall et al. CVPR 2017
- **Additional CNN for Disparity Refinement**
  - Gidaris & Komodakis, CVPR 2017
  - Pang et al. ICCVW 2017



01 Supervised Stereo Matching

Our work

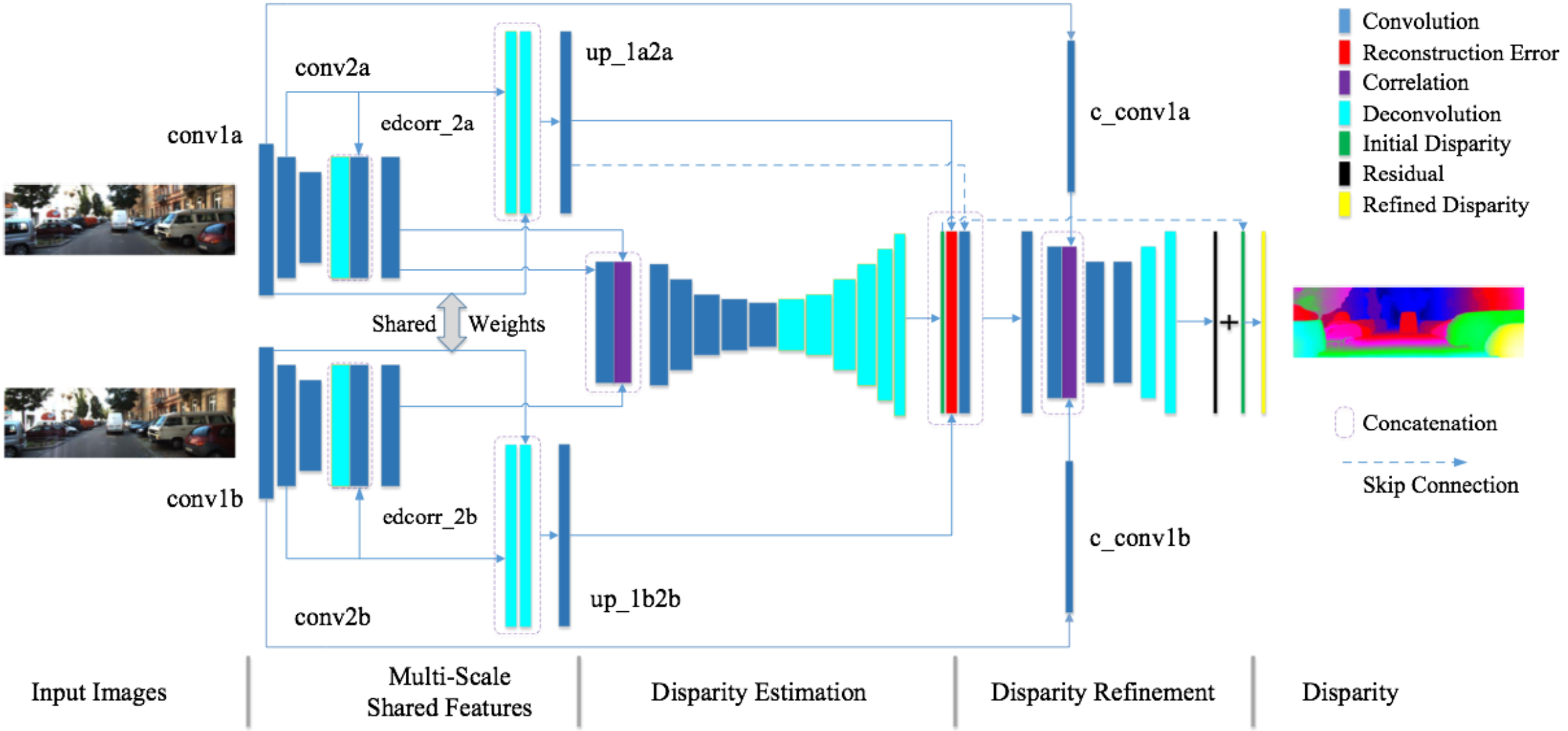
- **All-In-One: Integrate All Steps of Stereo Matching into One Network**
  - Improve Accuracy
  - Improve Efficiency
- **Residual Learning: Integrate Disparity Refinement into CNN**
  - Traditional Methods
    - left-right check: find correct, mismatched and occluded regions
    - interpolation, sub-pixel enhancement, filtering
    - hard to be modelled by CNN
  - Residual Learning
    - use initial disparity $disp_i$ to reconstruct left image from right image

*Z. Liang, **Y. Guo***, et al. Stereo Matching Using Multi-level Cost Volume and Multi-scale Feature Constancy. IEEE TPAMI 2019.*

01

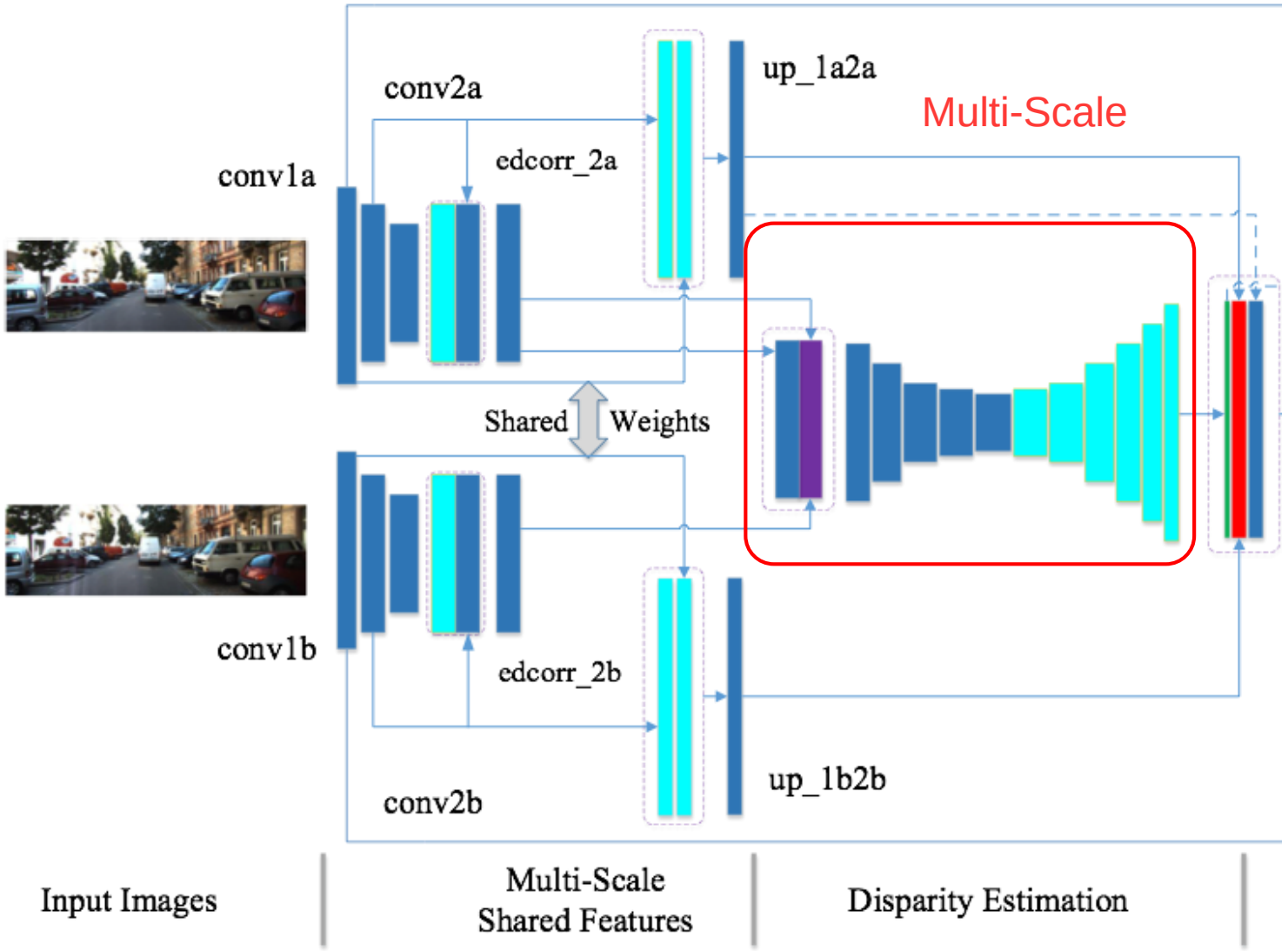
Supervised Stereo Matching

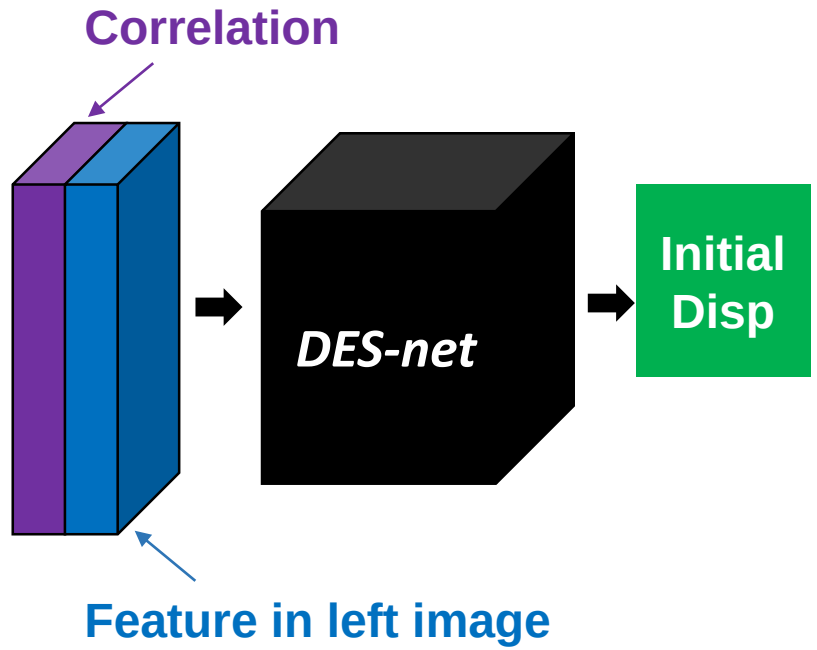
Overall Architecture



01 Supervised Stereo Matching

Overall Architecture

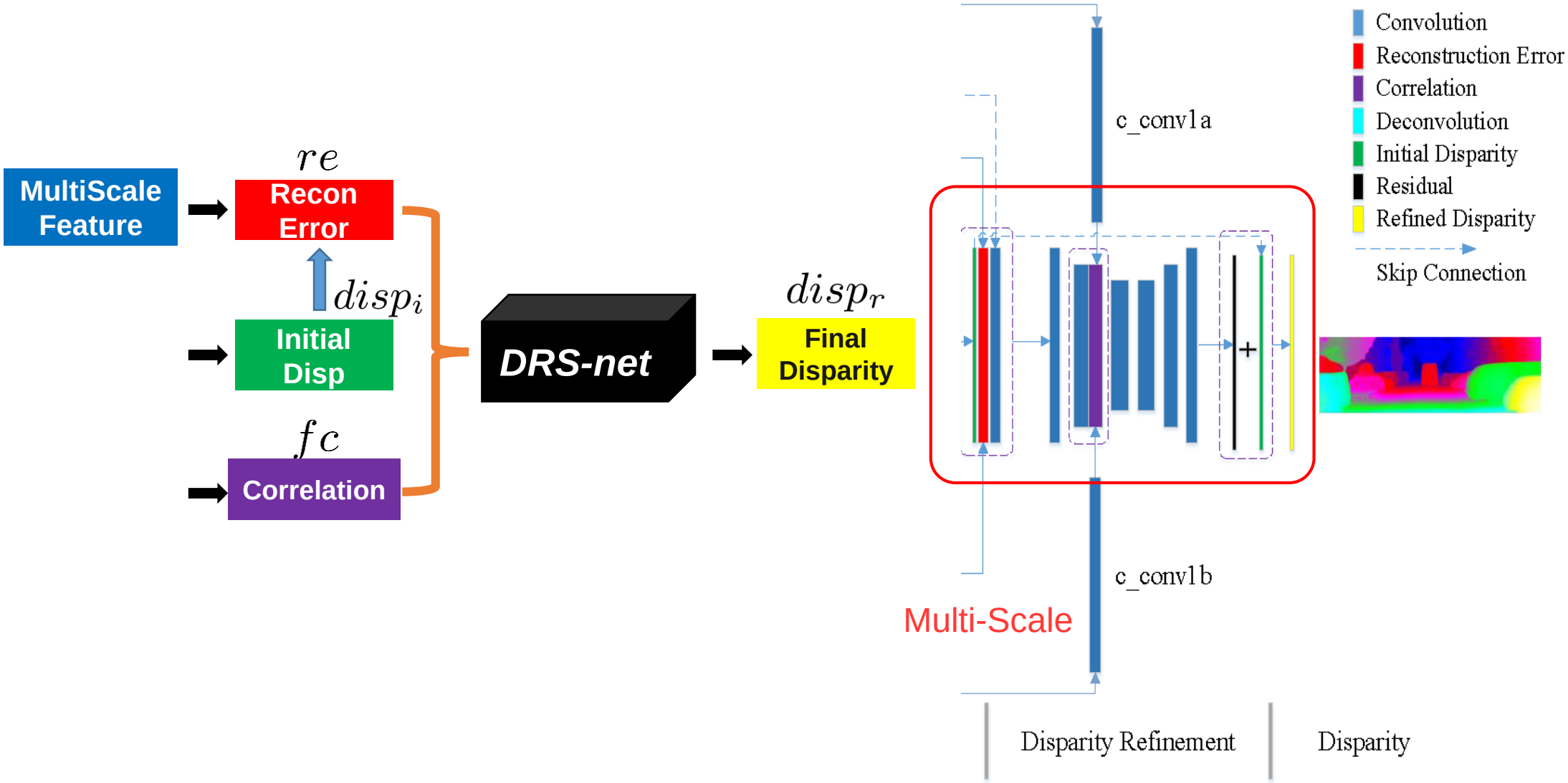




01

Supervised Stereo Matching

Overall Architecture



01 Supervised Stereo Matching

Results on SceneFlow

Model	> 1px	> 3px	>5px	EPE	Params.	Time (ms)
DES-net (without disparity refinement)	16.78	6.49	4.31	2.81	**42.76M**	**61**
DES-net + DRS-net without feature correlation fc and re	15.65	6.12	4.11	2.72	43.30M	100
DES-net + DRS-net without reconstruction error re	15.54	6.10	4.10	2.70	43.34M	111
DES-net + DRS-net without feature correlation fc	11.13	5.32	3.79	2.56	43.31M	103
DES-net + DRS-net without initial disparity $disp_i$	11.64	5.37	3.81	2.61	43.34M	114
DES-net + DRS-net (iResNet)	10.24	4.93	3.54	2.50	43.34M	114
Refinement × 2 (iResNet-i2)	9.42	4.64	3.37	2.46	43.34M	131
Refinement × 3 (iResNet-i3)	**9.28**	**4.57**	**3.32**	**2.45**	43.34M	148

- The EPE can be significantly reduced **with disparity refinement** using feature correlation and reconstruction error
- **Feature reconstruction error** plays the major role in performance improvement

*Z. Liang, **Y. Guo***, et al. Stereo Matching Using Multi-level Cost Volume and Multi-scale Feature Constancy. IEEE TPAMI 2019.*

01 Supervised Stereo Matching

Results on SceneFlow

Model	> 1px	> 3px	>5px	EPE	Params.	Time (ms)
DES-net (without disparity refinement)	16.78	6.49	4.31	2.81	**42.76M**	**61**
DES-net + DRS-net without feature correlation fc and re	15.65	6.12	4.11	2.72	43.30M	100
DES-net + DRS-net without reconstruction error re	15.54	6.10	4.10	2.70	43.34M	111
DES-net + DRS-net without feature correlation fc	11.13	5.32	3.79	2.56	43.31M	103
DES-net + DRS-net without initial disparity $disp_i$	11.64	5.37	3.81	2.61	43.34M	114
DES-net + DRS-net (iResNet)	10.24	4.93	3.54	2.50	43.34M	114
Refinement × 2 (iResNet-i2)	9.42	4.64	3.37	2.46	43.34M	131
Refinement × 3 (iResNet-i3)	**9.28**	**4.57**	**3.32**	**2.45**	43.34M	148

- The EPE can be significantly reduced **with disparity refinement** using feature correlation and reconstruction error
- **Feature reconstruction error** plays the major role in performance improvement
- **Iterative refinement** helps to further improve the performance

*Z. Liang, **Y. Guo***, et al. Stereo Matching Using Multi-level Cost Volume and Multi-scale Feature Constancy. IEEE TPAMI 2019.*

01 Supervised Stereo Matching

Results on KITTI 2015

	All Pixels			Non-Occluded Pixels			Runtime
	D1-bg	D1-fg	D1-all	D1-bg	D1-fg	D1-all	(s)
CRL [19]	2.48	**3.59**	2.67	2.32	**3.12**	2.45	0.47
GC-NET [12]	2.21	6.16	2.87	2.02	5.58	2.61	0.9
DRR [4]	2.58	6.04	3.16	2.34	4.87	2.76	0.4
SsSMnet [35]	2.70	6.92	3.40	2.46	6.13	3.06	0.8
L-ResMatch [28]	2.72	6.95	3.42	2.35	5.74	2.91	48
Displets v2 [5]	3.00	5.56	3.43	2.73	4.95	3.09	265
SGM-Net [27]	2.66	8.64	3.66	2.23	7.44	3.09	67
MC-CNN-acrt [32]	2.89	8.88	3.88	2.48	7.64	3.33	67
DispNetC [17]	4.32	4.41	4.34	4.11	3.72	4.05	**0.06**
iResNet-i2e2 (ours)	**2.10**	3.64	**2.36**	**1.94**	3.20	**2.15**	0.25

*Z. Liang, **Y. Guo***, et al. Stereo Matching Using Multi-level Cost Volume and Multi-scale Feature Constancy. IEEE TPAMI 2019.*

01 Supervised Stereo Matching

CVPR 2018 Robust Vision Challenge

	Method	Middlebury (Detailed subrankings)	KITTI (Detailed subrankings)	ETH3D (Detailed subrankings)
1	iResNet_ROB	6	2	2
	Learning for Disparity Estimation through Feature Constancy [Project page] - Submitted by Zhengfa Liang (National University of Defense Technology)			
2	DN-CSS_ROB	3	7	1
	Submitted by Tonmoy Saikia (University of Freiburg)			
3	DLCB_ROB	5	4	3
	Submitted by Sebastien Drouyer (CMLA - ENS Cachan)			
4	NaN_ROB	2	5	9
	Submitted by Yao Lu (Australian National University)			
4	PSMNet_ROB	10	1	5
	Pyramid Stereo Matching Network [Project page] - Submitted by Hsiao-Chien Yang (National Chiao Tung University (NCTU))			
6	NOSS_ROB	1	16	4
	Submitted by Anonymous			

3D Vision: From Stereo Matching to Point Cloud Understanding

Supervised Stereo Matching

Unsupervised Stereo Correspondence Learning

Semantic Segmentation of Large-scale Point Clouds

02

Unsupervised Stereo Correspondence Learning

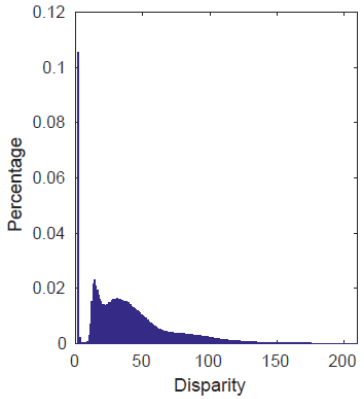
Objectives

- **Good Flexibility and Scalability** to Different Disparities

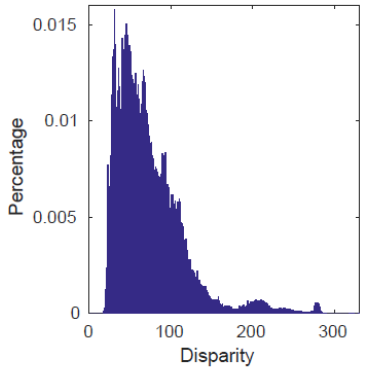
  Disparities between stereo images can vary significantly

  - ❑ different baselines
  - ❑ different focal lengths
  - ❑ different depths
  - ❑ different resolutions

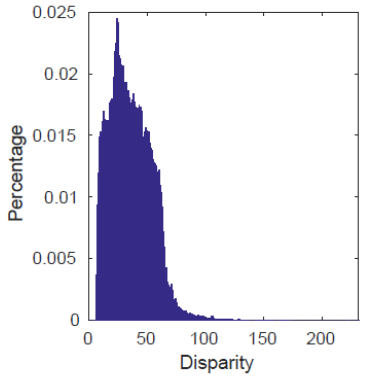
- **Low Computational and Memory Cost**



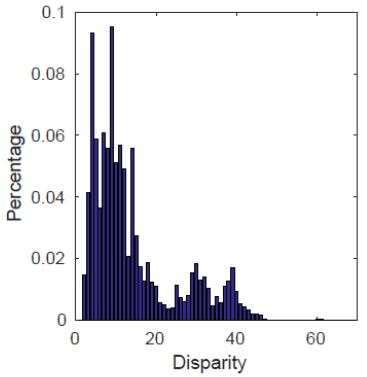
(a) *SceneFlow*



(b) *Middlebury 2014*



(c) *KITTI 2015*



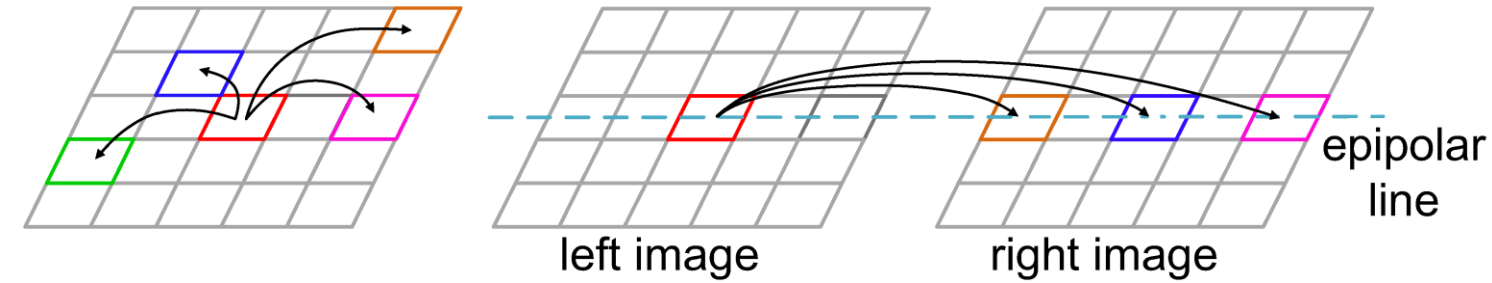
(d) *ETH3D 2017*

02

Unsupervised Stereo Correspondence Learning

Parallax-attention Module (PAM)

- **PAM to Capture Stereo Correspondence**
  - ❑ Achieve global receptive field along the epipolar line to handle different stereo images with large disparity variations
  - ❑ Improve efficiency



(a) self-attention (b) parallax-attention

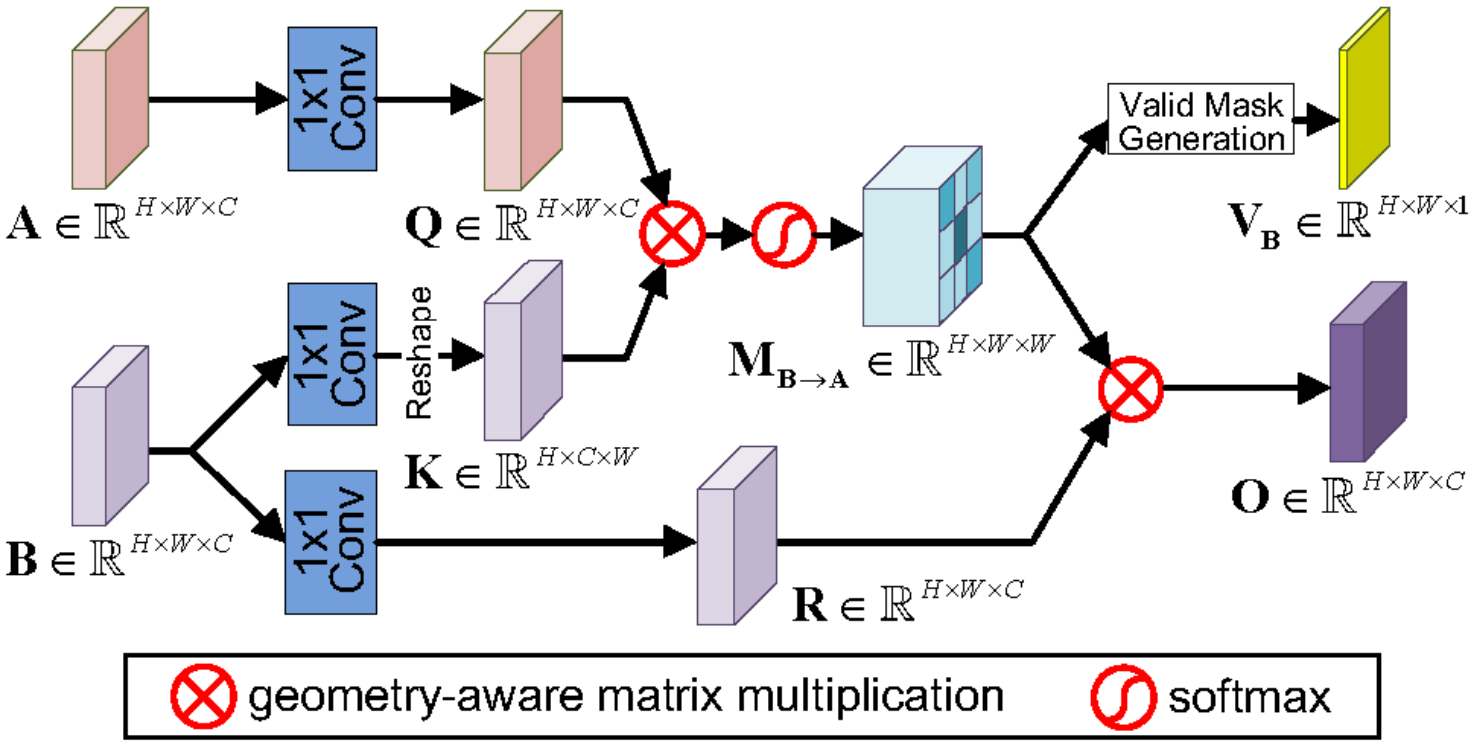
*L. Wang, **Y. Guo***, et al. Parallax Attention for Unsupervised Stereo Correspondence Learning. IEEE TPAMI, 2020*

02

Unsupervised Stereo Correspondence Learning

Parallax-attention Module (PAM)

Overview



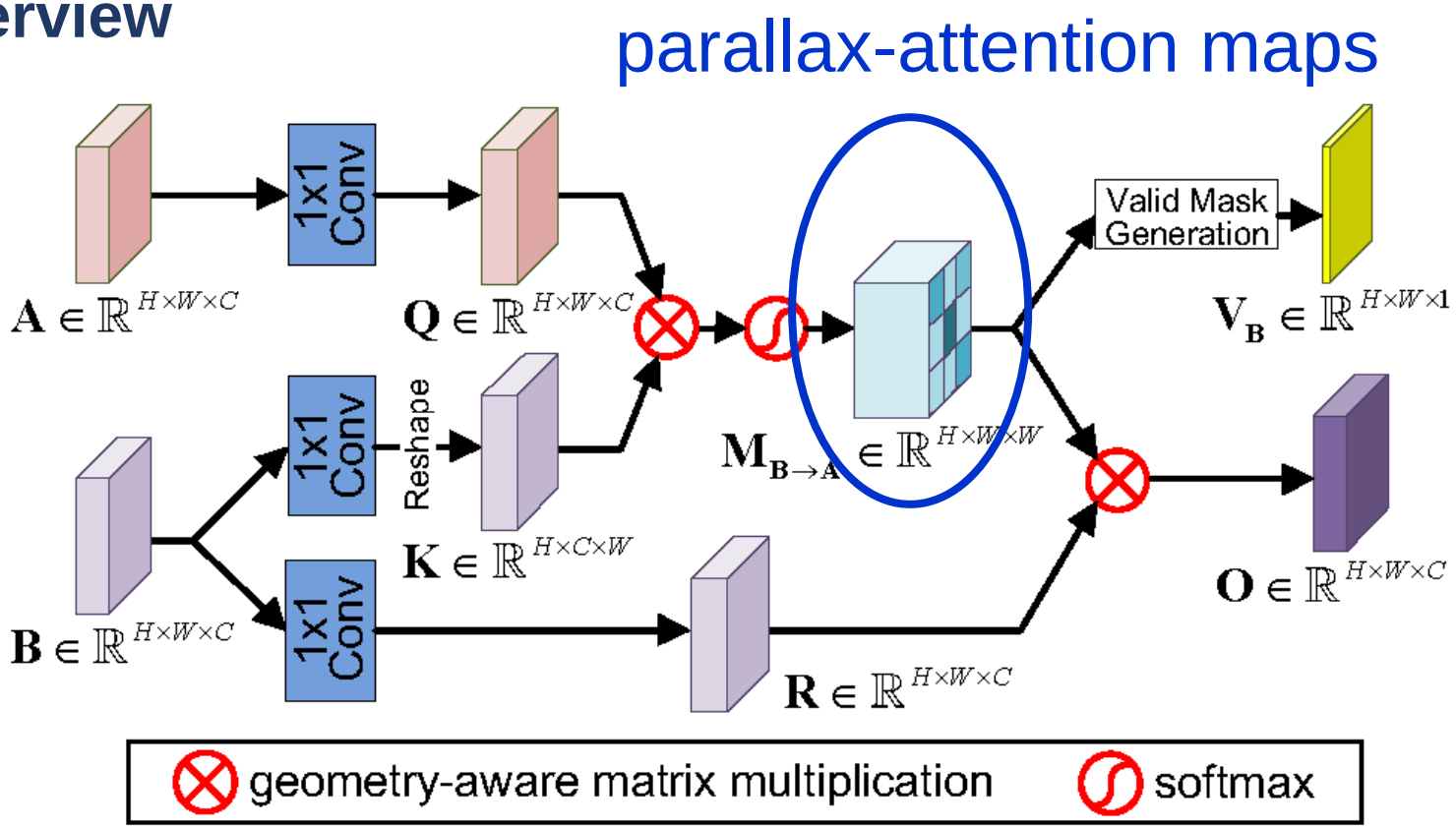
*L. Wang, **Y. Guo***, et al. Parallax Attention for Unsupervised Stereo Correspondence Learning. IEEE TPAMI, 2020*

02

Unsupervised Stereo Correspondence Learning

Parallax-attention Module (PAM)

Overview



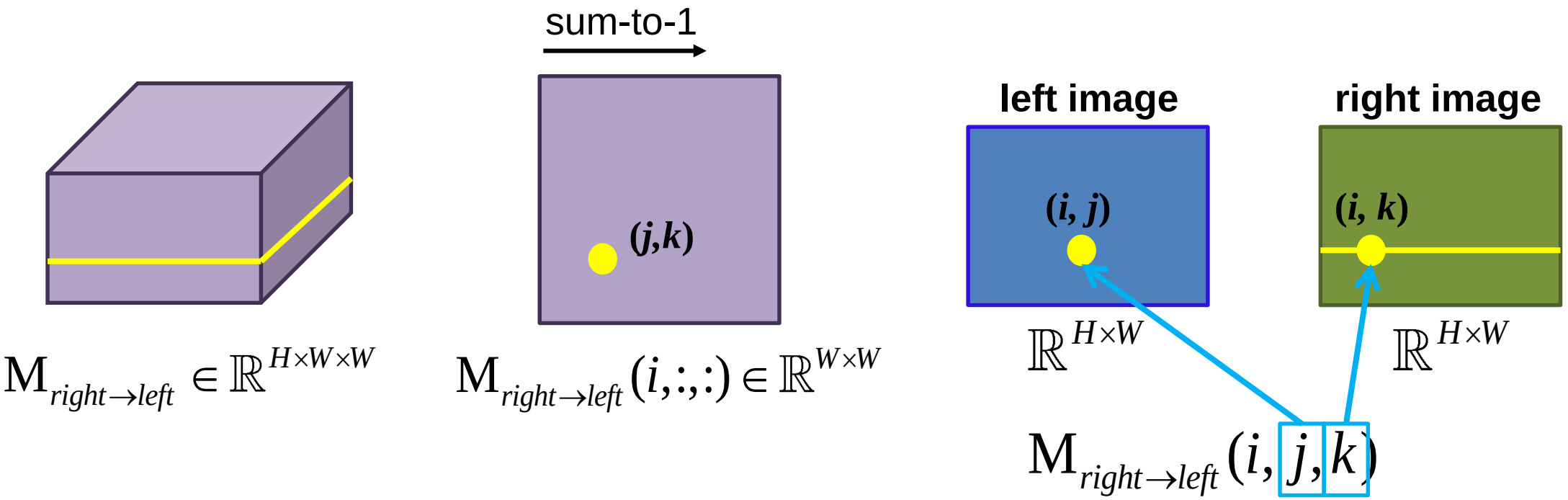
*L. Wang, **Y. Guo***, et al. Parallax Attention for Unsupervised Stereo Correspondence Learning. IEEE TPAMI, 2020*

02

Unsupervised Stereo Correspondence Learning

Parallax-attention Module (PAM)

How PAM Works?



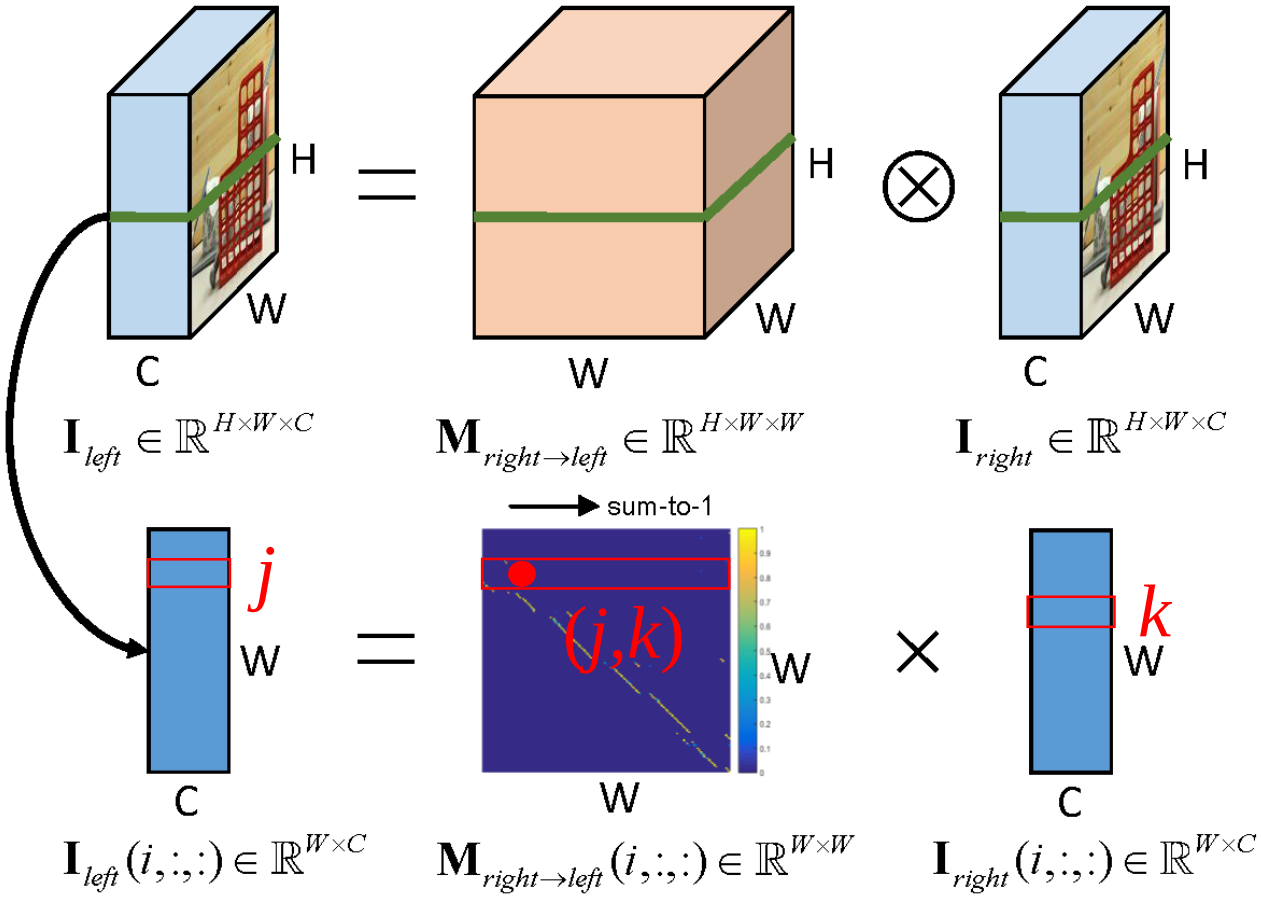
*L. Wang, **Y. Guo***, et al. Parallax Attention for Unsupervised Stereo Correspondence Learning. IEEE TPAMI, 2020*

02

Unsupervised Stereo Correspondence Learning

Parallax-attention Module (PAM)

■ Geometry-Aware Matrix Multiplication as Warping



*L. Wang, **Y. Guo***, et al. Parallax Attention for Unsupervised Stereo Correspondence Learning. IEEE TPAMI, 2020*

02

Unsupervised Stereo Correspondence Learning

Parallax-attention Module (PAM)

❑ Loss Design

(1) Left-Right Consistency

$$\begin{cases} \mathbf{I}_{left} = \mathbf{M}_{right \to left} \otimes \mathbf{I}_{right} \\ \mathbf{I}_{right} = \mathbf{M}_{left \to right} \otimes \mathbf{I}_{left} \end{cases}$$

(2) Cycle Consistency

$$\begin{cases} \mathbf{I}_{left} = \mathbf{M}_{left \to right \to left} \otimes \mathbf{I}_{left} \\ \mathbf{I}_{right} = \mathbf{M}_{right \to left \to right} \otimes \mathbf{I}_{right} \end{cases}$$

$$\begin{cases} \mathbf{M}_{left \to right \to left} = \mathbf{M}_{right \to left} \otimes \mathbf{M}_{left \to right} \\ \mathbf{M}_{right \to left \to right} = \mathbf{M}_{left \to right} \otimes \mathbf{M}_{right \to left} \end{cases}$$

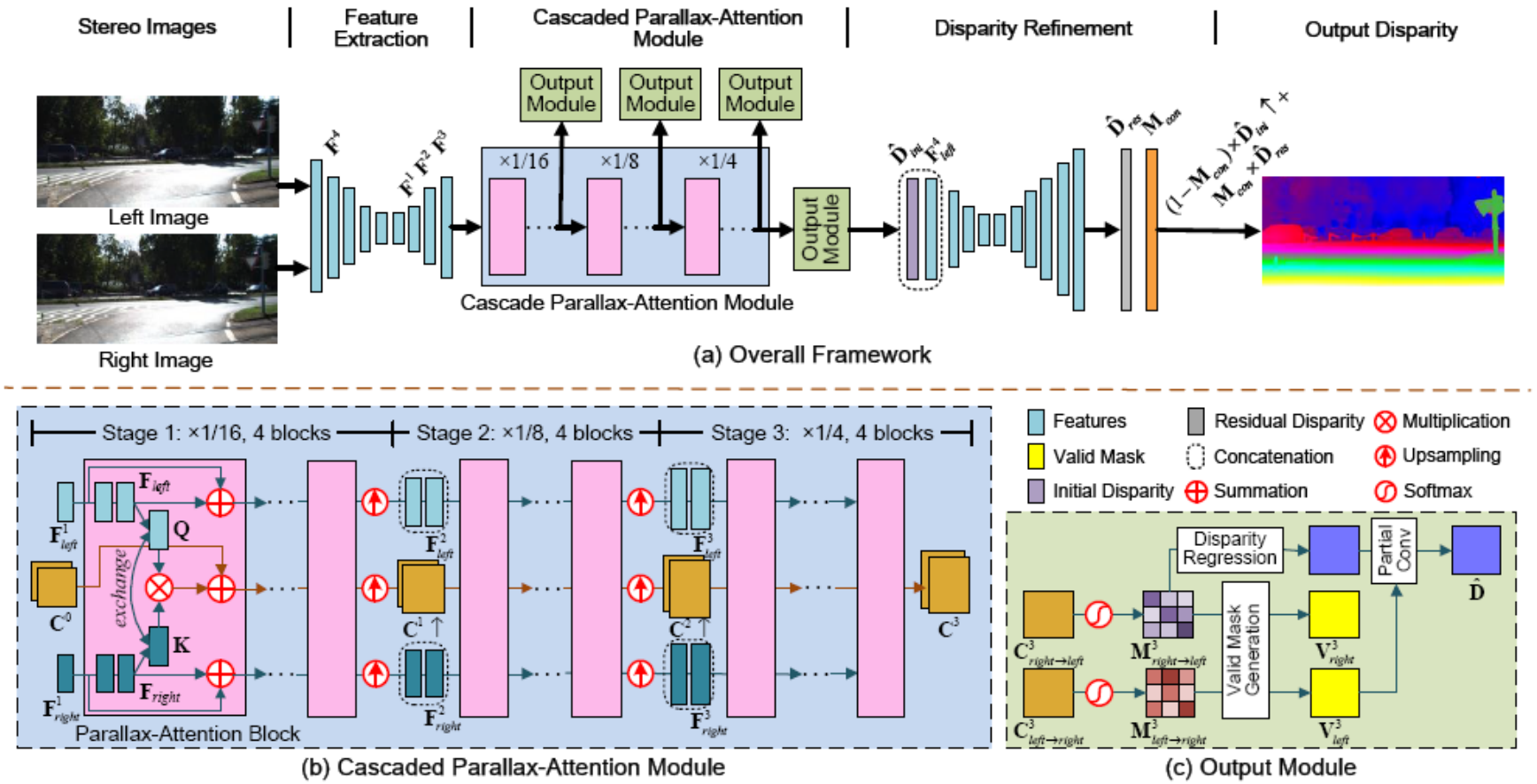
*L. Wang, **Y. Guo***, et al. Parallax Attention for Unsupervised Stereo Correspondence Learning. IEEE TPAMI, 2020*

02

Unsupervised Stereo Correspondence Learning

PAM for Unsupervised Stereo Matching (PASMnet)

Stereo Matching Network



*L. Wang, **Y. Guo***, et al. Parallax Attention for Unsupervised Stereo Correspondence Learning. IEEE TPAMI, 2020*

02

Unsupervised Stereo Correspondence Learning

PAM for Unsupervised Stereo Matching (PASMnet)

■ Comparison to stereo matching methods on KITTI 2015

		Noc			All		
		D1-bg	D1-fg	D1-all	D1-bg	D1-fg	D1-all
Supervised	DispNet [11]	4.11	3.72	4.05	4.32	4.41	4.34
	GC-Net [4]	2.02	5.58	2.61	2.21	6.16	2.87
	CRL [52]	2.32	3.12	2.45	2.48	3.59	2.67
	iResNet [14]	2.07	**2.76**	2.19	2.25	**3.40**	2.44
	PSMNet [5]	**1.71**	4.31	**2.14**	**1.86**	4.62	**2.32**
	PASMnet_192 (ours)	1.88	3.91	2.22	2.04	4.33	2.41
Unsupervised	USCNN [53]	-	-	11.71	-	-	16.55
	Yu et al. [54]	-	-	8.35	-	-	19.14
	Zhou et al. [27]	-	-	8.61	-	-	9.91
	SegStereo [24]	-	-	7.70	-	-	8.79
	OASM [28]	5.44	17.30	7.39	6.89	19.42	8.98
	PASMnet (ours)	5.35	15.24	6.99	5.89	16.74	7.70
	PASMnet_192 (ours)	**5.02**	**15.16**	**6.69**	**5.41**	**16.36**	**7.23**

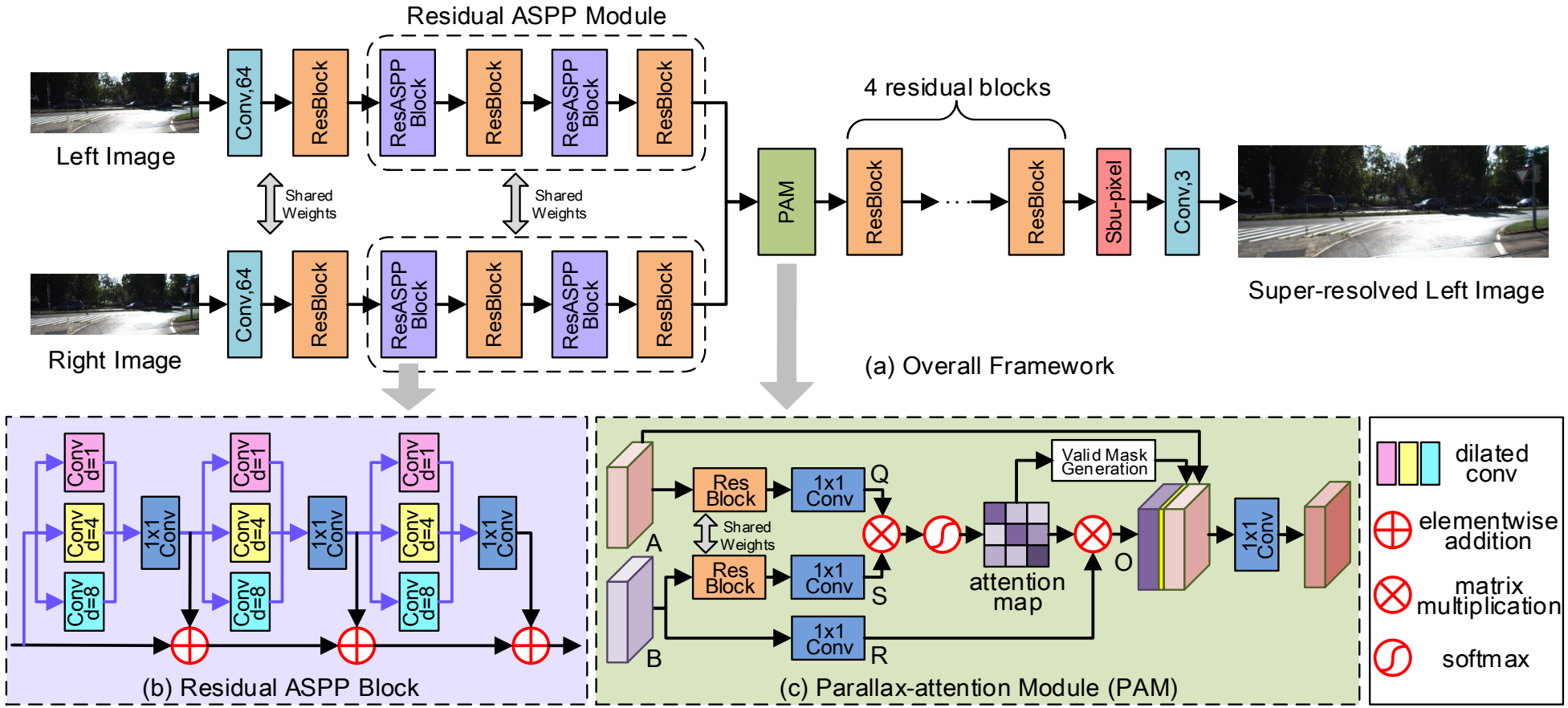
*L. Wang, **Y. Guo***, et al. Parallax Attention for Unsupervised Stereo Correspondence Learning. IEEE TPAMI, 2020*

02

Unsupervised Stereo Correspondence Learning

PAM for Stereo Image Super-Resolution (PASSRnet)

Stereo Super-Resolution Network



*L. Wang, **Y. Guo***, et al. Parallax Attention for Unsupervised Stereo Correspondence Learning. IEEE TPAMI, 2020*

02

Unsupervised Stereo Correspondence Learning

PAM for Stereo Image Super-Resolution (PASSRnet)

■ Comparative Results on Middlebury, KITTI 2012 and KITTI 2015

Dataset	Scale	Single Image SR					Stereo Image SR	
		SRCNN [1]	VDSR [19]	DRCN [42]	LapSRN [5]	DRRN [20]	StereoSR [6]	Ours
Middlebury (5 images)	×2	32.05/0.935	32.66/0.941	32.82/0.941	32.75/0.940	32.91/0.945	33.05/0.955*	**34.05/0.960**
	×4	27.46/0.843	27.89/0.853	27.93/0.856	27.98/0.861	27.93/0.855	26.80/0.850*	**28.63/0.871**
KITTI 2012 (20 images)	×2	29.75/0.901	30.17/0.906	30.19/0.906	30.10/0.905	30.16/0.908	30.13/0.908	**30.65/0.916**
	×4	25.53/0.764	25.93/0.778	25.92/0.777	25.96/0.779	25.94/0.773	-	**26.26/0.790**
KITTI 2015 (20 images)	×2	28.77/0.901	28.99/0.904	29.04/0.904	28.97/0.903	29.00/0.906	29.09/0.909	**29.78/0.919**
	×4	24.68/0.744	25.01/0.760	25.04/0.759	25.03/0.760	25.05/0.756	-	**25.43/0.776**

*L. Wang, **Y. Guo***, et al. Parallax Attention for Unsupervised Stereo Correspondence Learning. IEEE TPAMI, 2020*

02

Unsupervised Stereo Correspondence Learning

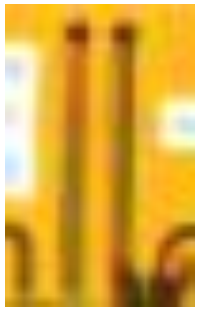
PAM for Stereo Image Super-Resolution (PASSRnet)

Bicubic

02

Unsupervised Stereo Correspondence Learning

PAM for Stereo Image Super-Resolution (PASSRnet)



StereoSR, CVPR 2018

02

Unsupervised Stereo Correspondence Learning

PAM for Stereo Image Super-Resolution (PASSRnet)

Our PASSRnet

02

Unsupervised Stereo Correspondence Learning

PAM for Stereo Image Super-Resolution (PASSRnet)

Groundtruth

3D Vision: From Stereo Matching to Point Cloud Understanding

Supervised Stereo Matching

Unsupervised Stereo Correspondence Learning

Semantic Segmentation of Large-scale Point Clouds

03

Semantic Segmentation of Large-scale Point Clouds

Task Definition

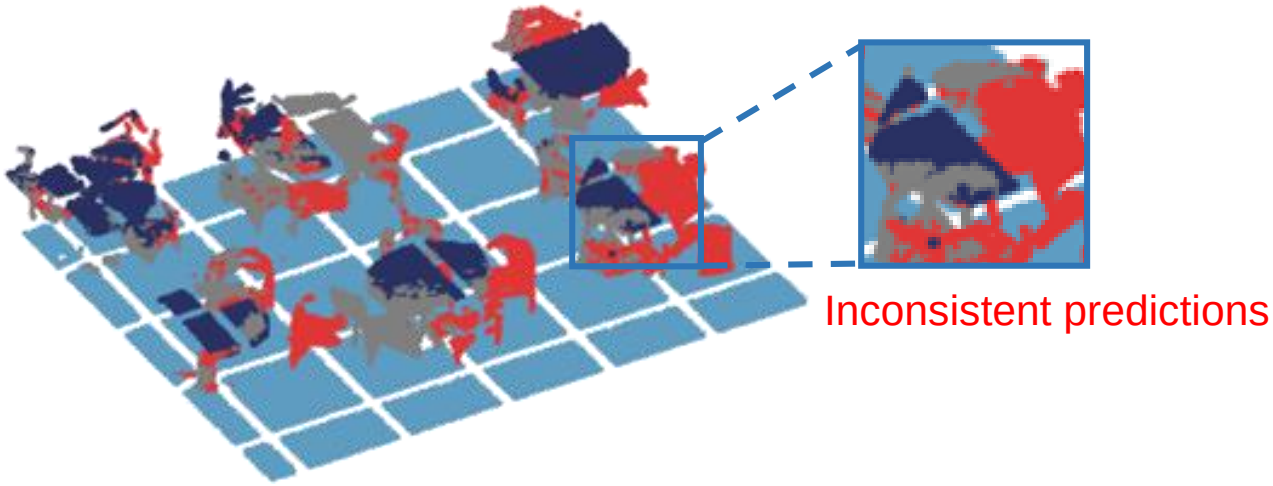
Credit: C. R. Qi, H. Su, K. Mo, and L. J. Guibas. PointNet: Deep learning on point sets for 3D classification and segmentation. CVPR 2017.

03 Semantic Segmentation of Large-scale Point Clouds

Motivation

PointNet

Input point clouds



Output Semantic Results

floor table chair clutter window door bookcase

Limitations of existing methods:

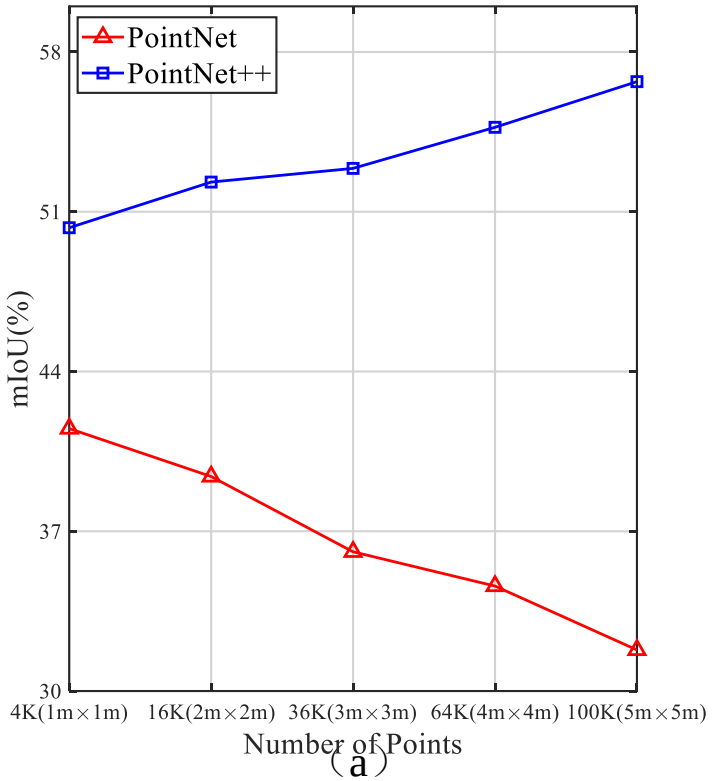
(1) Most approaches are limited to extremely small 3D point clouds (e.g. 1mx1m blocks with 4K points).
PointNet (CVPR'17); PointNet++ (NeurIPS'17); PointCNN (NeurIPS'18); PCCN (CVPR'18); ShellNet (ICCV'19)

(2) Few methods can directly process large-scale point clouds, but they either rely on time-consuming preprocessing or expensive voxelization steps. SPG (CVPR'18); FCPN (ECCV'18); TagentConv (CVPR'18); PCT (ICIP'19)

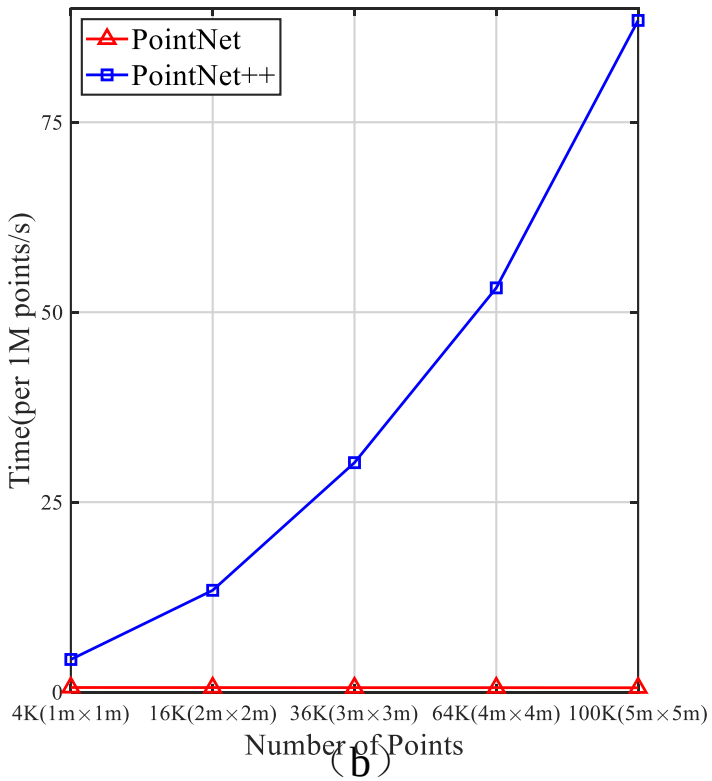
03

Semantic Segmentation of Large-scale Point Clouds

Motivation



(a)



(b)

Increase the block size:

(1) PointNet only learn independent point features without considering local geometric relationships. The max-pooling operation used for capturing global features discards the majority of information from point features.

(2) The inference time of PointNet++ increases dramatically for a larger number of points, since the computational complexity of farthest point sampling is quadratically related to the number of input points

03

Semantic Segmentation of Large-scale Point Clouds

Motivation

- Difficulties of large-scale point cloud segmentation:

- ***Complex geometry**: A large-scale point cloud usually contains dozens of object classes, hundreds of instances and millions of points.*
- ***GPU Memory limitations**: It is difficult to process a large-scale point cloud (million-scale points) in a single pass in todays' GPUs.*
- ***Variable and diverse** : The spatial size and number of points of a point cloud acquired by real-world depth sensors can change significantly*

Q. Hu et al. RandLA-Net: Efficient Semantic Segmentation of Large-scale Point Clouds. CVPR, Oral, 2020

03 Semantic Segmentation of Large-scale Point Clouds

Objectives

- Can we develop a method that is:
  - **Process large-scale point clouds directly**
    - *Without block partition and block merging*
    - *Keep the original geometry as much as possible*
  - **Computationally efficient & Memory efficient**
    - *Without time-consuming preprocessing or memory-cost voxelization steps*
    - *Inference a large-scale point cloud in a single pass*
  - **Accurate & Scalable**
    - *Capture and preserve the prominent features from complex geometric structures*
    - *Able to process input point clouds with different spatial size and number of points*

Q. Hu et al. RandLA-Net: Efficient Semantic Segmentation of Large-scale Point Clouds. CVPR, Oral, 2020

03 Semantic Segmentation of Large-scale Point Clouds

Overview

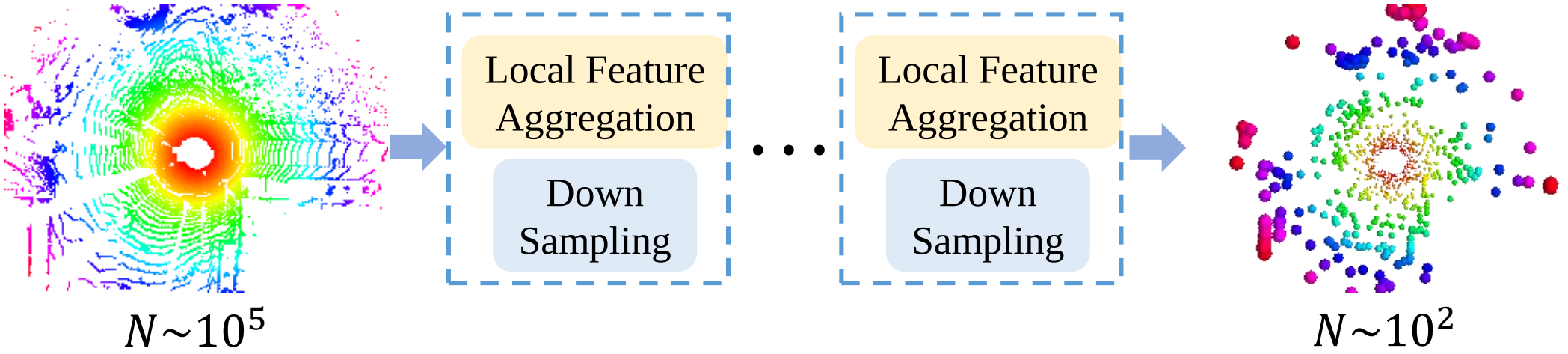


Figure 2. In each layer of the network, the large-scale point cloud is significantly downsampled, yet is capable of retaining features necessary for accurate segmentation.

- What we need:
  - Efficient point sampling to reduce memory footprint and computational cost
  - Effective local feature aggregation to capture the geometrical patterns

Q. Hu et al. RandLA-Net: Efficient Semantic Segmentation of Large-scale Point Clouds. CVPR, Oral, 2020

Semantic Segmentation of Large-scale Point Clouds

Sampling

The quest for efficient sampling

- FPS, IDIS and GS are too computationally expensive
- CRS approaches have an excessive memory footprint and PGS has extremely large exploration space
- Random sampling is by far the most suitable approach to process large-scale point clouds in terms of efficiency

New question:

➢ How to preserve useful features?

Q. Hu et al. RandLA-Net: Efficient Semantic Segmentation of Large-scale Point Clouds. CVPR, Oral, 2020

03 Semantic Segmentation of Large-scale Point Clouds

Local Feature Aggregation

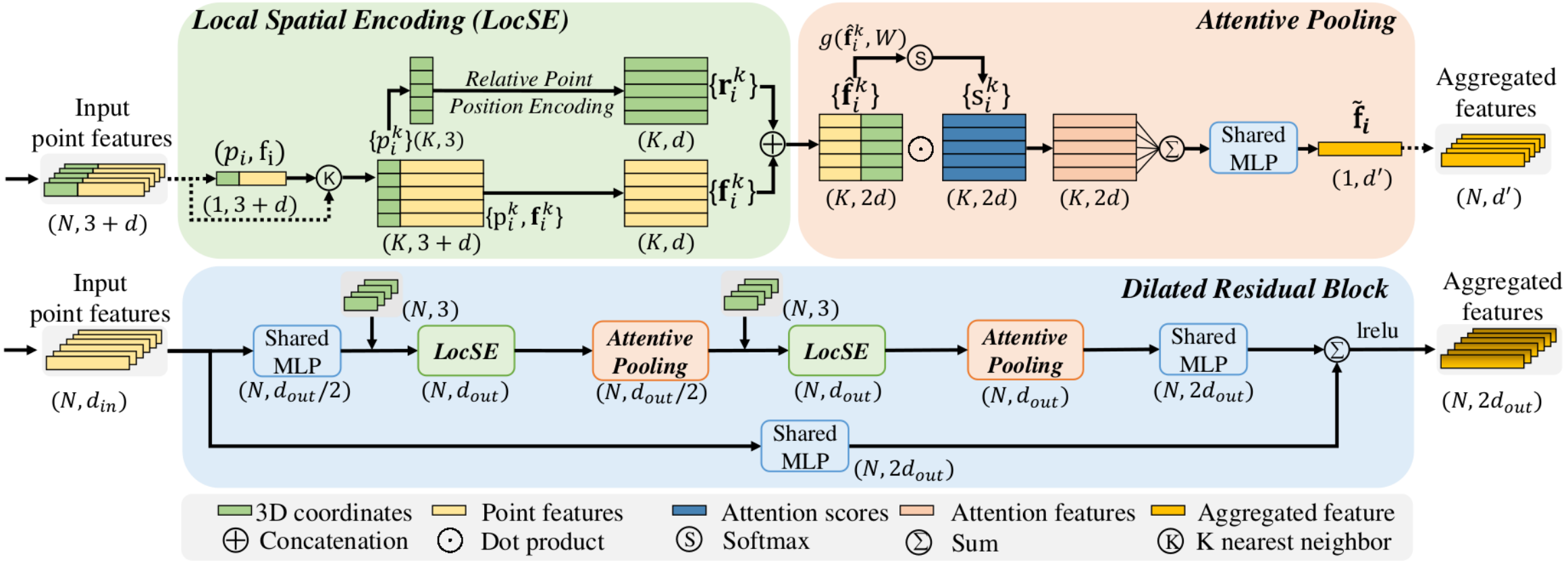


Figure 3. The proposed local feature aggregation module. The top panel shows the location spatial encoding block that extracts features, and the attentive pooling mechanism that weights the most important, based on the local context and geometry. The bottom panel shows how two of these components are chained together, to increase the receptive field size, within a residual block.

Q. Hu et al. RandLA-Net: Efficient Semantic Segmentation of Large-scale Point Clouds. CVPR, Oral, 2020

03 Semantic Segmentation of Large-scale Point Clouds

Network Architecture

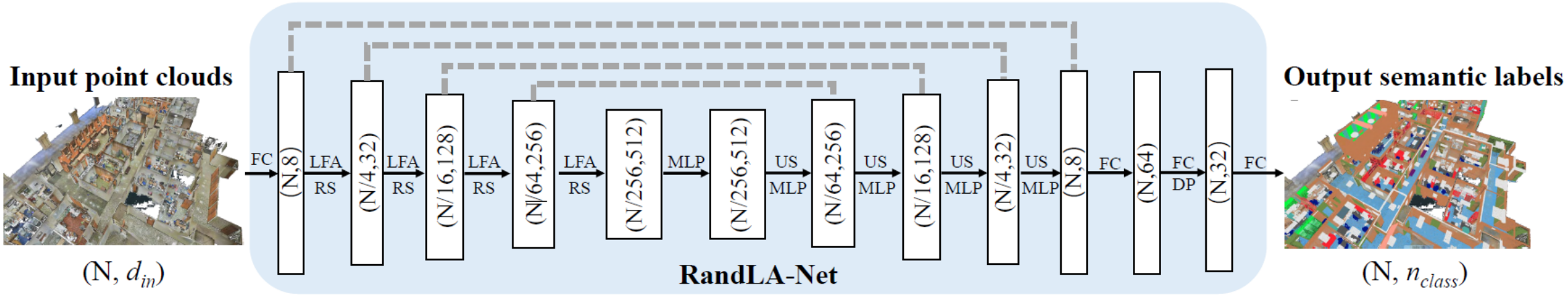


Figure 7. The detailed architecture of our RandLA-Net. (N, D) represents the number of points and feature dimension respectively. FC: Fully Connected layer, LFA: Local Feature Aggregation, RS: Random Sampling, MLP: shared Multi-Layer Perceptron, US: Up-sampling, DP: Dropout.

Highlights:

(1) 3D encoder-decoder architecture with skip connections

(2) Four encoding and decoding layers + three fully-connected layers + dropout

(3) The down-sampling ratio is set to 4 in each layer;

Q. Hu et al. RandLA-Net: Efficient Semantic Segmentation of Large-scale Point Clouds. CVPR, Oral, 2020

03

Semantic Segmentation of Large-scale Point Clouds

Efficiency of Random Sampling

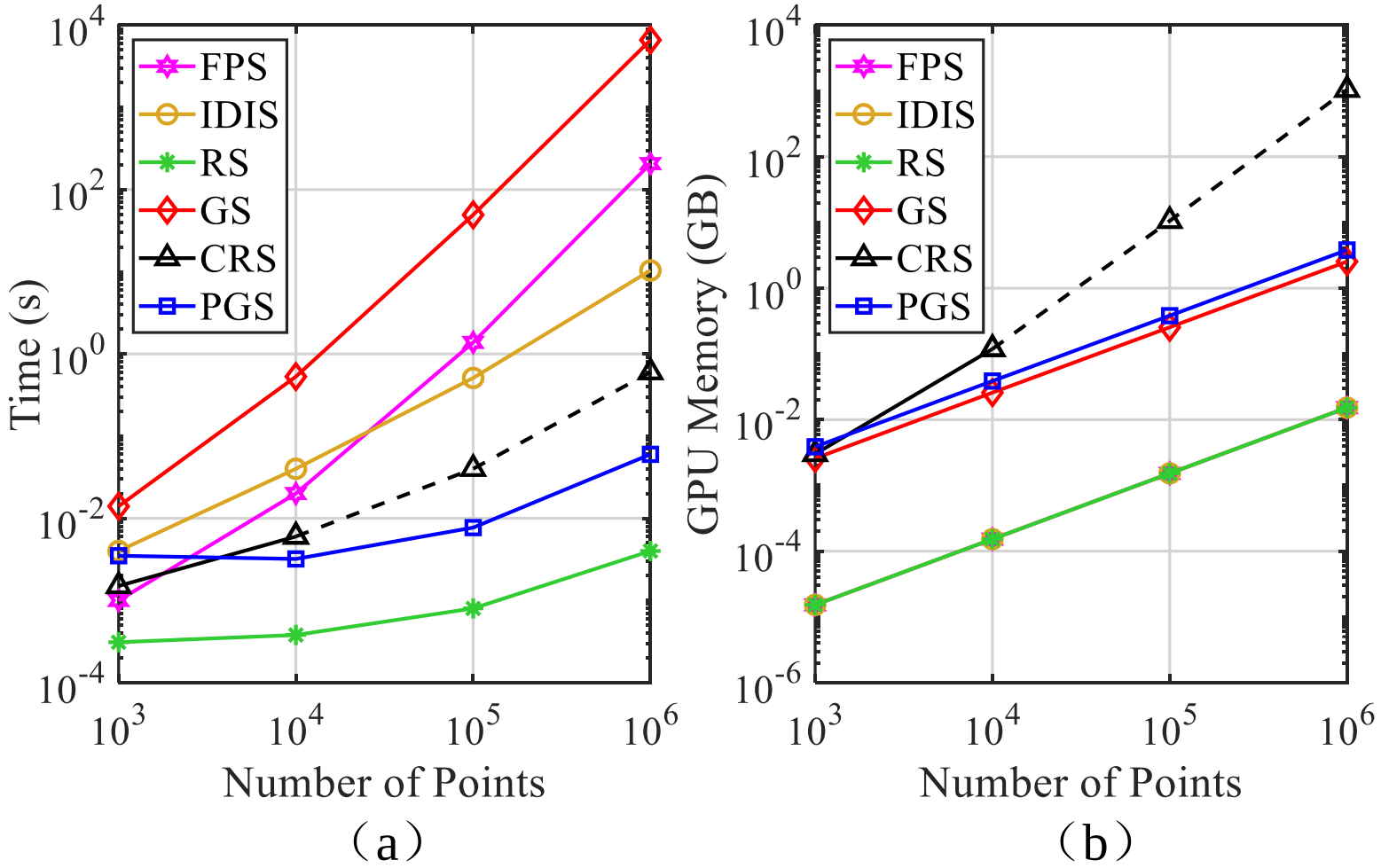


Figure 5. Time and memory consumption of different sampling approaches. The dashed lines represent estimated values due to the limited GPU memory.

03

Semantic Segmentation of Large-scale Point Clouds

Efficiency of RandLA-Net

	Total time (seconds)	Parameters (millions)	Maximum inference points (millions)
PointNet (Vanilla) [37]	192	0.8	0.49
PointNet++ (SSG) [38]	9831	0.97	0.98
PointCNN [29]	8142	11	0.05
SPG [23]	43584	**0.25**	-
KPConv [48]	717	14.9	0.54
RandLA-Net (Ours)	**176**	0.95	**1.15**

Table 1. The computation time, network parameters and maximum number of input points of different approaches for semantic segmentation on Sequence 08 of SemanticKITTI dataset.

- Evaluate on real-world large-scale dataset (Sequences 08 of SemanticKITTI)
- 4071 scans of point clouds in total, 81920 points from each scan were fed to each network
- All experiments conducted on a PC with an AMD 3700X @3.6GHz and an NVIDIA RTX2080Ti GPU

03 Semantic Segmentation of Large-scale Point Clouds

Evaluation on Semantic3D

	mIoU (%)	OA (%)	man-made.	natural.	high veg.	low veg.	buildings	hard scape	scanning art.	cars
SnapNet_ [4]	59.1	88.6	82.0	77.3	79.7	22.9	91.1	18.4	37.3	64.4
SEGCloud [46]	61.3	88.1	83.9	66.0	86.0	40.5	91.1	30.9	27.5	64.3
RF_MSSF [47]	62.7	90.3	87.6	80.3	81.8	36.4	92.2	24.1	42.6	56.6
MSDeepVoxNet [40]	65.3	88.4	83.0	67.2	83.8	36.7	92.4	31.3	50.0	78.2
ShellNet [63]	69.3	93.2	96.3	90.4	83.9	41.0	94.2	34.7	43.9	70.2
GACNet [50]	70.8	91.9	86.4	77.7	**88.5**	**60.6**	94.2	37.3	43.5	77.8
SPG [23]	73.2	94.0	**97.4**	**92.6**	87.9	44.0	83.2	31.0	63.5	76.2
KPConv [48]	74.6	92.9	90.9	82.2	84.2	47.9	94.9	40.0	**77.3**	**79.7**
RandLA-Net (Ours)	**76.0**	**94.4**	96.5	92.0	85.1	50.3	**95.0**	**41.1**	68.2	79.4

Table 2. Quantitative results of different approaches on Semantic3D (reduced-8) [16]. Only the recent published approaches are compared. Accessed on 15 November 2019.

- 15 point clouds for training and 15 for online testing, with more than 4 billion points
- 8 semantic categories, both 3D coordinates and RGB information are available
- Covering up to 160×240×30 meters in real-world 3D space and up to 10^8 points

03 Semantic Segmentation of Large-scale Point Clouds

Evaluation on SemanticKITTI

	Methods	Size	mIoU(%)	Params(M)	road	sidewalk	parking	other-ground	building	car	truck	bicycle	motorcycle	other-vehicle	vegetation	trunk	terrain	person	bicyclist	motorcyclist	fence	pole	traffic-sign
Point-based Methods	PointNet [37]	50K pts	14.6	3	61.6	35.7	15.8	1.4	41.4	46.3	0.1	1.3	0.3	0.8	31.0	4.6	17.6	0.2	0.2	0.0	12.9	2.4	3.7
	SPG [23]		17.4	**0.25**	45.0	28.5	0.6	0.6	64.3	49.3	0.1	0.2	0.2	0.8	48.9	27.2	24.6	0.3	2.7	0.1	20.8	15.9	0.8
	SPLATNet [43]		18.4	0.8	64.6	39.1	0.4	0.0	58.3	58.2	0.0	0.0	0.0	0.0	71.1	9.9	19.3	0.0	0.0	0.0	23.1	5.6	0.0
	PointNet++ [38]		20.1	6	72.0	41.8	18.7	5.6	62.3	53.7	0.9	1.9	0.2	0.2	46.5	13.8	30.0	0.9	1.0	0.0	16.9	6.0	8.9
	TangentConv [45]		40.9	0.4	83.9	63.9	33.4	15.4	83.4	90.8	15.2	2.7	16.5	12.1	79.5	49.3	58.1	23.0	28.4	**8.1**	49.0	35.8	28.5
Projection Methods	SqueezeSeg [52]	64*2048 pixels	29.5	1	85.4	54.3	26.9	4.5	57.4	68.8	3.3	16.0	4.1	3.6	60.0	24.3	53.7	12.9	13.1	0.9	29.0	17.5	24.5
	SqueezeSegV2 [53]		39.7	1	88.6	67.6	45.8	17.7	73.7	81.8	13.4	18.5	17.9	14.0	71.8	35.8	60.2	20.1	25.1	3.9	41.1	20.2	36.3
	DarkNet21Seg [3]		47.4	25	91.4	74.0	57.0	26.4	81.9	85.4	18.6	**26.2**	26.5	15.6	77.6	48.4	63.6	31.8	33.6	4.0	52.3	36.0	50.0
	DarkNet53Seg [3]		49.9	50	**91.8**	**74.6**	**64.8**	**27.9**	**84.1**	86.4	25.5	24.5	**32.7**	22.6	78.3	50.1	**64.0**	36.2	33.6	4.7	**55.0**	38.9	**52.2**
	RandLA-Net (Ours)	50K pts	**50.3**	0.95	90.4	67.9	56.9	15.5	81.1	**94.0**	**42.7**	19.8	21.4	**38.7**	**78.3**	**60.3**	59.0	**47.5**	**48.8**	4.6	49.7	**44.2**	38.1

Table 3. Quantitative results of different approaches on SemanticKITTI [3]. Only the recent published methods are compared and all scores are obtained from the online single scan evaluation track. Accessed on 15 November 2019.

- Sequential LiDAR point clouds, with 19 semantic categories
- 19130 scans for training, 4071 scans for validation and 20351 scans for testing
- Only have 3D coordinates without color information

03

Semantic Segmentation of Large-scale Point Clouds

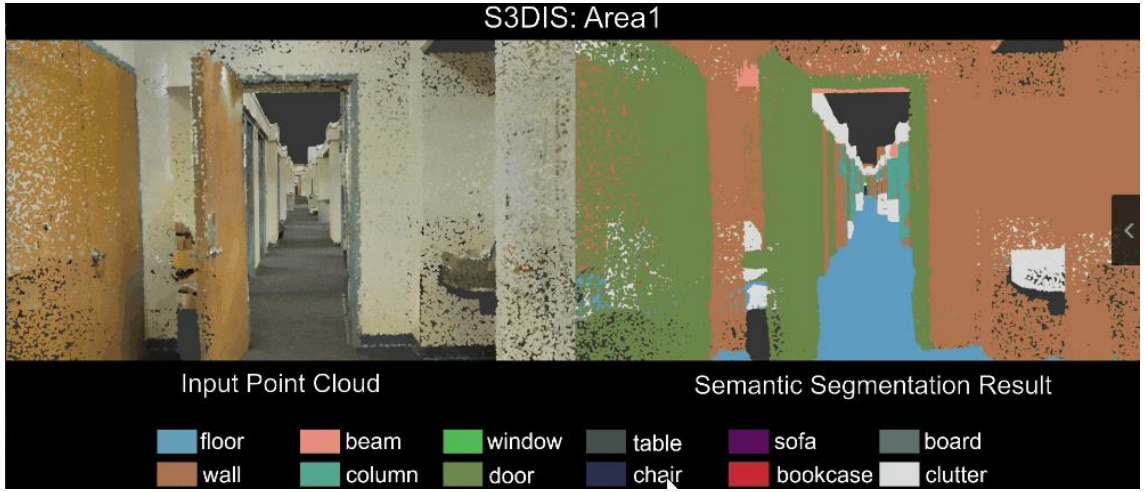
Evaluation on S3DIS

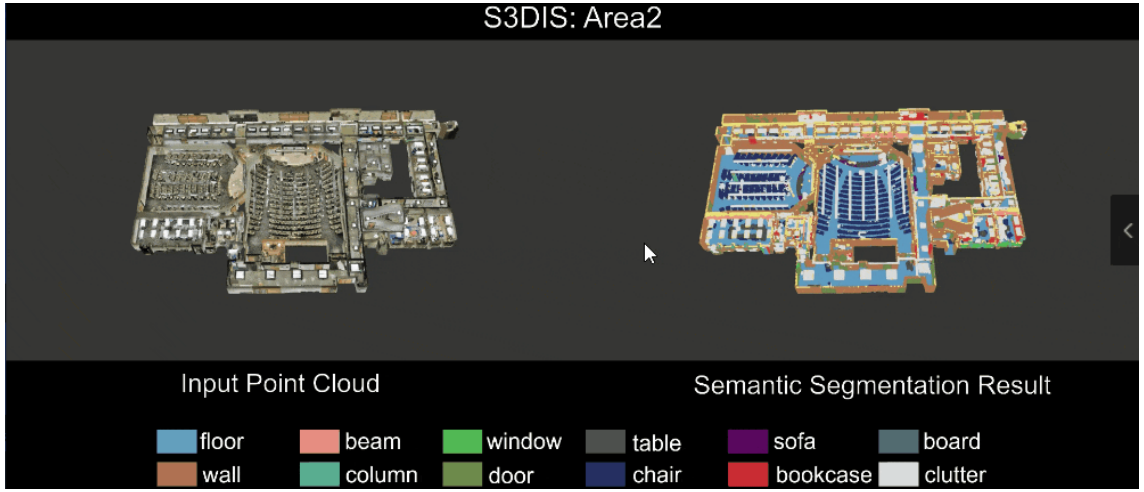
	OA(%)	mAcc(%)	mIoU(%)
PointNet [37]	78.6	66.2	47.6
PointNet++ [38]	81.0	67.1	54.5
DGCNN [51]	84.1	-	56.1
3P-RNN [61]	86.9	-	56.3
RSNet [19]	-	66.5	56.5
SPG [23]	85.5	73.0	62.1
LSANet [6]	86.8	-	62.2
PointCNN [29]	88.1	75.6	65.4
PointWeb [64]	87.3	76.2	66.7
ShellNet [63]	87.1	-	66.8
HEPIN [20]	**88.2**	-	67.8
KPConv [48]	-	79.1	**70.6**
RandLA-Net (Ours)	87.2	**81.5**	68.5

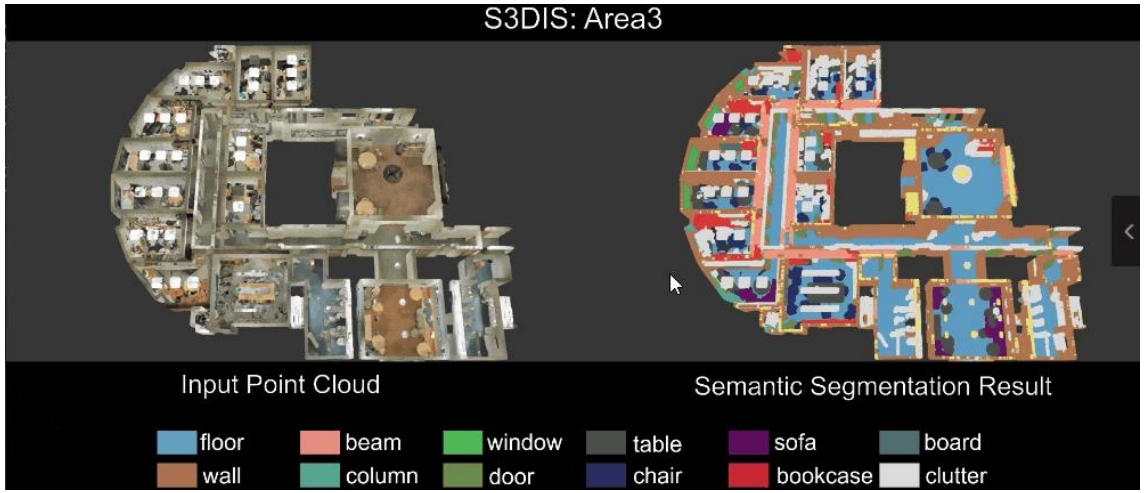
Table 4. Quantitative results of different approaches on S3DIS dataset [2] (6-fold cross validation). Only the recent published methods are included.

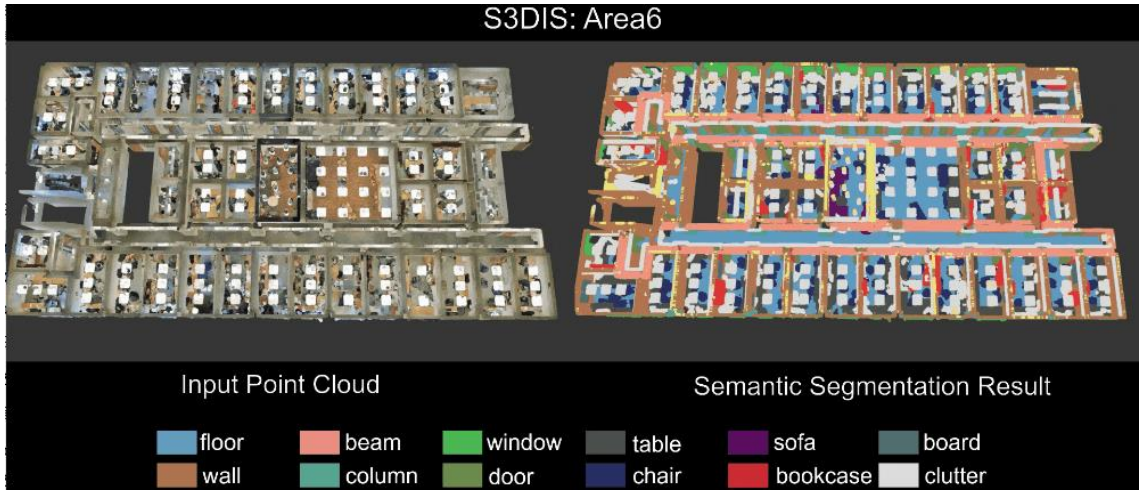
03 Semantic Segmentation of Large-scale Point Clouds

Evaluation on S3DIS







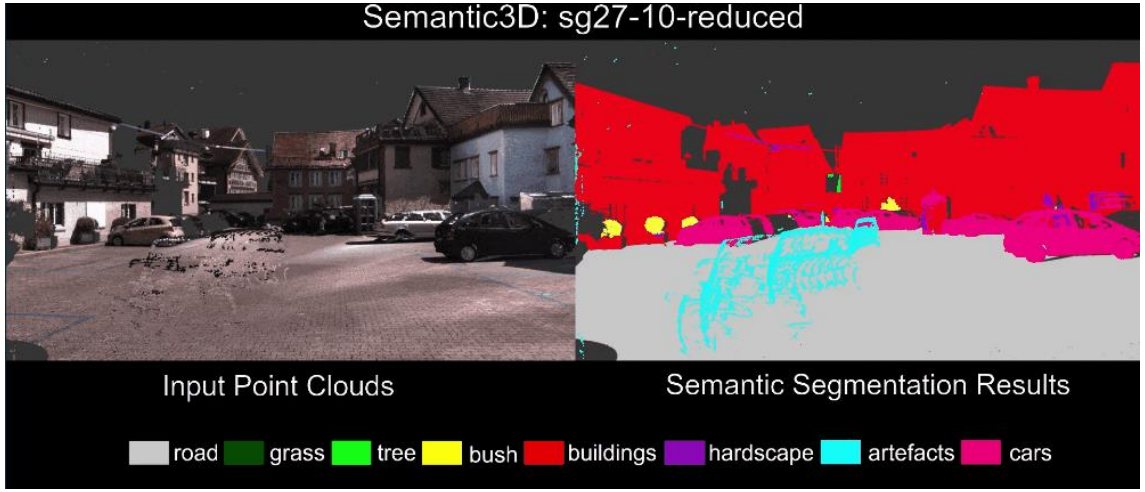


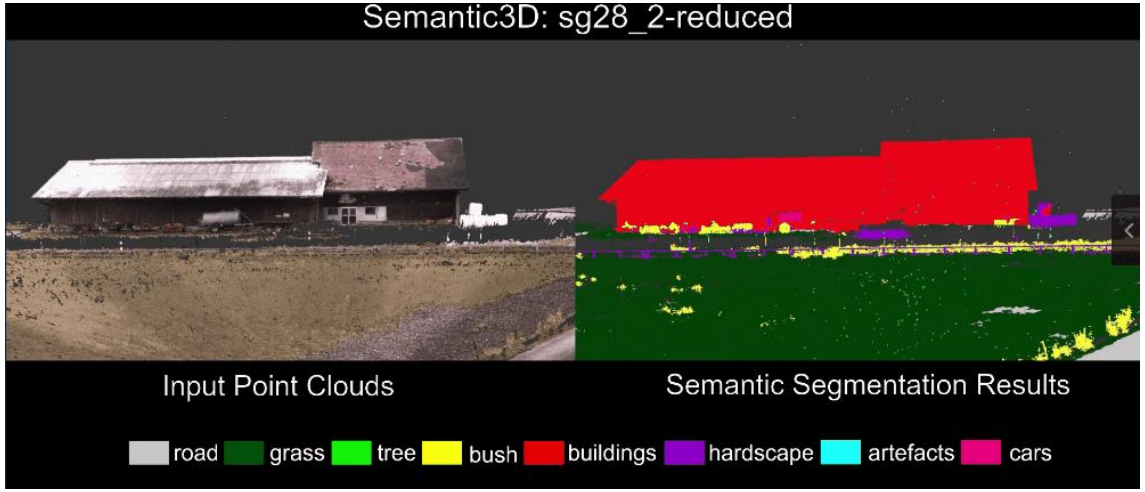
Q. Hu et al. RandLA-Net: Efficient Semantic Segmentation of Large-scale Point Clouds. CVPR, Oral, 2020

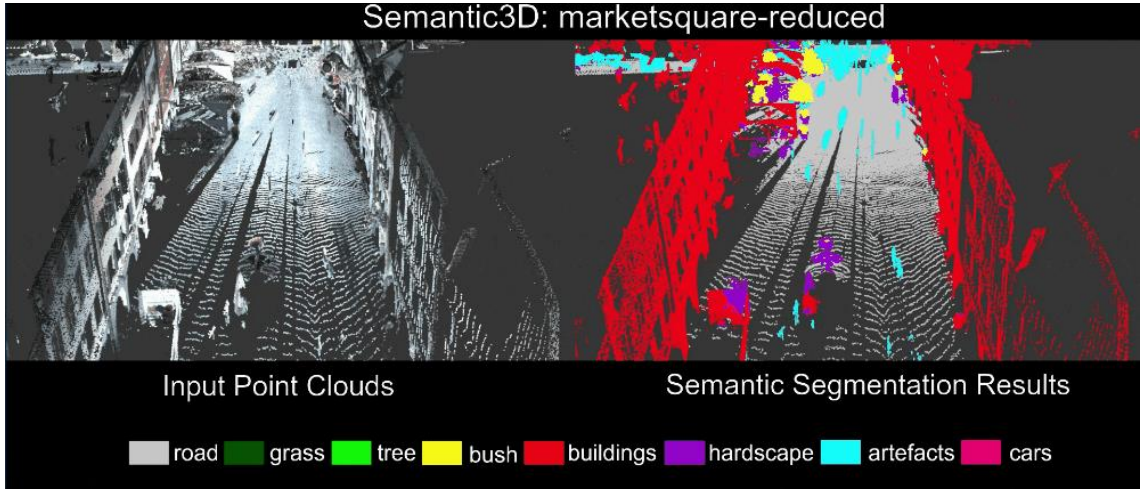
03

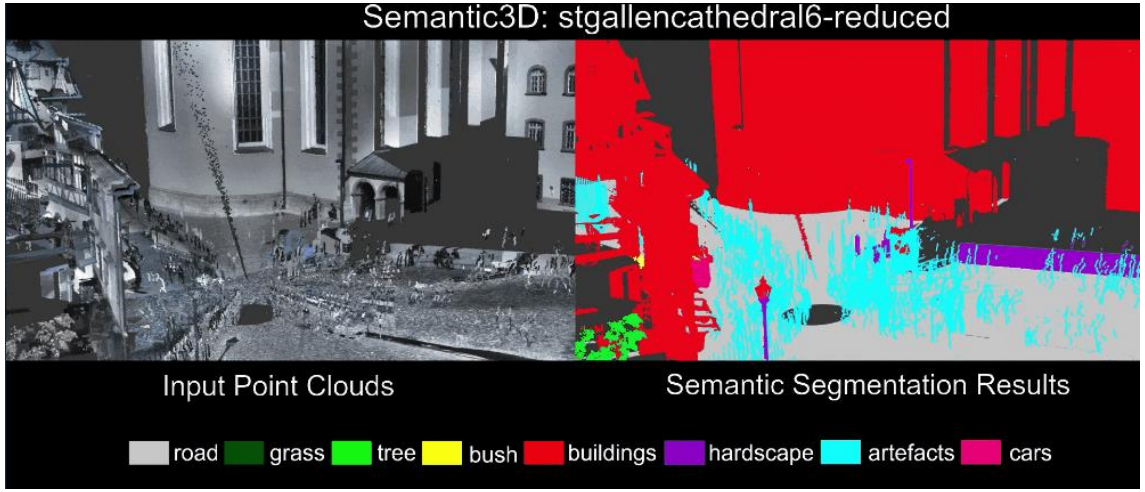
Semantic Segmentation of Large-scale Point Clouds

Evaluation on Semantic3D









Q. Hu et al. RandLA-Net: Efficient Semantic Segmentation of Large-scale Point Clouds. CVPR, Oral, 2020

Summary

Codes & Datasets

- Supervised Stereo Matching
  - CVPR 2018, IEEE TPAMI 2019
  - Codes: https://github.com/Gary66/iResNet
- Unsupervised Stereo Correspondence Learning
  - CVPR 2019, IEEE TPAMI 2020
  - Codes: https://github.com/Gary66/PASSRnet
  - Flickr1024 Dataset: https://yingqianwang.github.io/Flickr1024
- Semantic Segmentation of Large-scale Point Clouds
  - CVPR 2020 Oral
  - Codes: https://github.com/Gary66/RandLA-Net

THANKS