SqueezeSeg: Convolutional Neural Nets with Recurrent CRF for Real-Time Road-Object Segmentation from 3D LiDAR Point Cloud

Longtian Qiu qiult@shanghaitech.edu.cn

REFERENCES

[1] A. Geiger, P. Lenz, and R. Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on. IEEE, 2012, pp. 3354–3361.

[2] B. Douillard, J. Underwood, N. Kuntz, V. Vlaskine, A. Quadros, P. Morton, and A. Frenkel, "On the segmentation of 3d lidar point clouds," in Robotics and Automation (ICRA), 2011 IEEE International Conference on. IEEE, 2011, pp. 2798–2805.

[3] M. Himmelsbach, A. Mueller, T. Lu¨ttel, and H.-J. Wu¨nsche, "Lidar- based 3d object perception," in Proceedings of 1st international workshop on cognition for technical systems, vol. 1, 2008.

[4] D. Z. Wang, I. Posner, and P. Newman, "What could move? finding cars, pedestrians and bicyclists in 3d laser data," in Robotics and Automation (ICRA), 2012 IEEE International Conference on. IEEE, 2012, pp. 4038–4044.

[5] D. Zermas, I. Izzat, and N. Papanikolopoulos, "Fast segmentation of 3d point clouds: A paradigm on lidar data for autonomous vehicle applications," in Robotics and Automation (ICRA), 2017 IEEE Inter- national Conference on. IEEE, 2017, pp. 5067–5073.

[6] S. Thrun, M. Montemerlo, H. Dahlkamp, D. Stavens, A. Aron, J. Diebel, P. Fong, J. Gale, M. Halpenny, G. Hoffmann et al., "Stanley: The robot that won the darpa grand challenge," Journal of field Robotics, vol. 23, no. 9, pp. 661–692, 2006.

[7] F. Moosmann, O. Pink, and C. Stiller, "Segmentation of 3d lidar data in non-flat urban environments using a local convexity criterion," in Intelligent Vehicles Symposium, 2009 IEEE. IEEE, 2009, pp. 215– 220.

[8] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in CVPR, 2015.

[9] P. Kra¨henbu¨hl and V. Koltun, "Efficient inference in fully connected crfs with gaussian edge potentials," in Advances in neural information processing systems, 2011, pp. 109–117.

[10] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," arXiv preprint arXiv:1606.00915, 2016.

[11] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in Proceedings of the IEEE International Conference on Computer Vision, 2015, pp. 1529–1537.

[12] F. N. Iandola, S. Han, M. W. Moskewicz, K. Ashraf, W. J. Dally, and K. Keutzer, "SqueezeNet: Alexnet-level accuracy with 50x fewer parameters and <0.5mb model size," arXiv:1602.07360, 2016.

[13] M.-O. Shin, G.-M. Oh, S.-W. Kim, and S.-W. Seo, "Real-time and accurate segmentation of 3-d point clouds based on gaussian process regression," IEEE Transactions on Intelligent Transportation Systems, 2017.

[14] L. Caltagirone, S. Scheidegger, L. Svensson, and M. Wahde, "Fast lidar-based road detection using fully convolutional neural networks." in Intelligent Vehicles Symposium (IV), 2017 IEEE. IEEE, 2017, pp. 1019–1024.

[15] X. Chen, H. Ma, J. Wan, B. Li, and T. Xia, "Multi-view 3d object detection network for autonomous driving," arXiv preprint arXiv:1611.07759, 2016.

[16] J. Schlosser, C. K. Chow, and Z. Kira, "Fusing lidar and images for pedestrian detection using convolutional neural networks," in Robotics and Automation (ICRA), 2016 IEEE International Conference on. IEEE, 2016, pp. 2198–2205.

[17] B. Li, T. Zhang, and T. Xia, "Vehicle detection from 3d lidar using fully convolutional network," arXiv preprint arXiv:1608.07916, 2016. [18] D. Maturana and S. Scherer, "3d convolutional neural networks for landing zone detection from lidar," in Robotics and Automation (ICRA), 2015 IEEE International Conference on. IEEE, 2015, pp.

3471–3478.
[19] S. R. Richter, V. Vineet, S. Roth, and V. Koltun, "Playing for data:

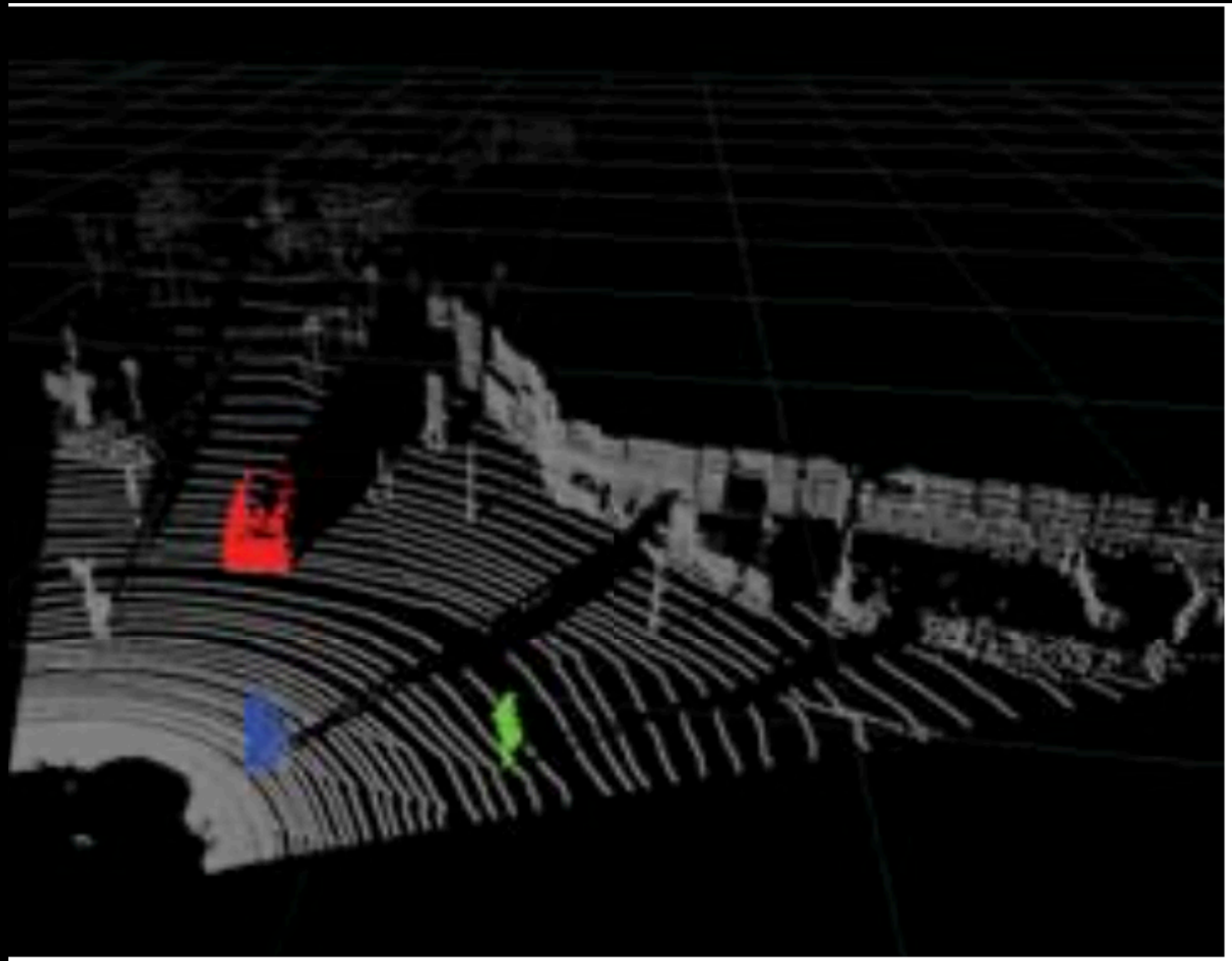
Ground truth from computer games," in European Conference on Computer Vision (ECCV), ser. LNCS, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., vol. 9906. Springer International Publishing, 2016, pp. 102–118.

[20] M. Johnson-Roberson, C. Barto, R. Mehta, S. N. Sridhar, and R. Vasudevan, "Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks?" CoRR, vol. abs/1610.01983, 2016. [Online]. Available: http://arxiv.org/abs/1610. 01983

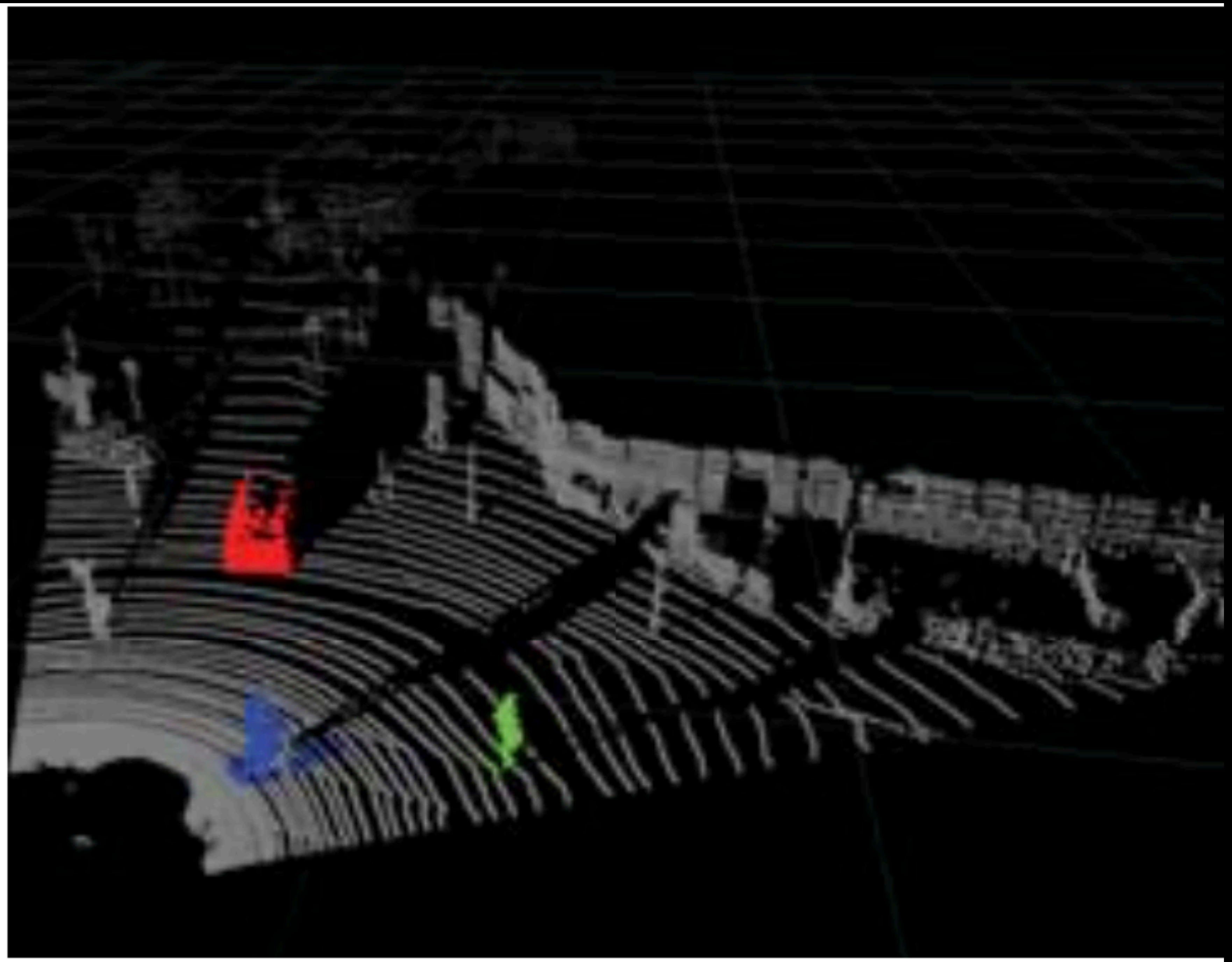
[21] B. Wu, F. Iandola, P. H. Jin, and K. Keutzer, "Squeezedet: Uni- fied, small, low power fully convolutional neural networks for real-time object detection for autonomous driving," arXiv preprint arXiv:1612.01051, 2016.

Introduction

Address semantic segmentation of road-objects from 3D LiDAR point clouds. In particular, we wish to detect and categorize instances of interest, such as cars, pedestrians and cyclists.



Ground truth segmentation



Predicted segmentation

RELATED WORK

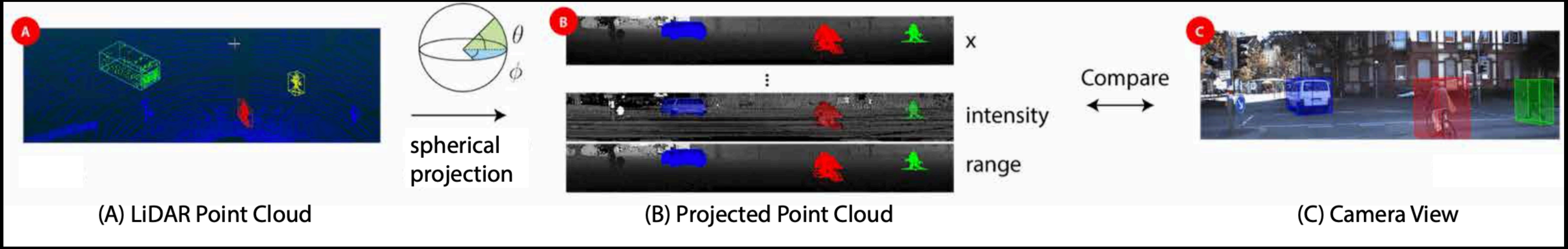
- Semantic segmentation for 3D LiDAR point clouds
- CNN for 3D point clouds
- Semantic Segmentation for Images
- Data Collection through Simulation

METHOD DESCRIPTION

Point Cloud Transformation

$$\theta = \arcsin \frac{z}{\sqrt{x^2+y^2+z^2}}, \quad \tilde{\theta} = \lfloor \theta / \triangle\theta \rfloor,$$
$$\phi = \arcsin \frac{y}{\sqrt{x^2+y^2}}, \quad \tilde{\phi} = \lfloor \phi / \triangle\phi \rfloor.$$

project the LiDAR point cloud onto a sphere for a dense, grid-based representation



METHOD DESCRIPTION

Network structure

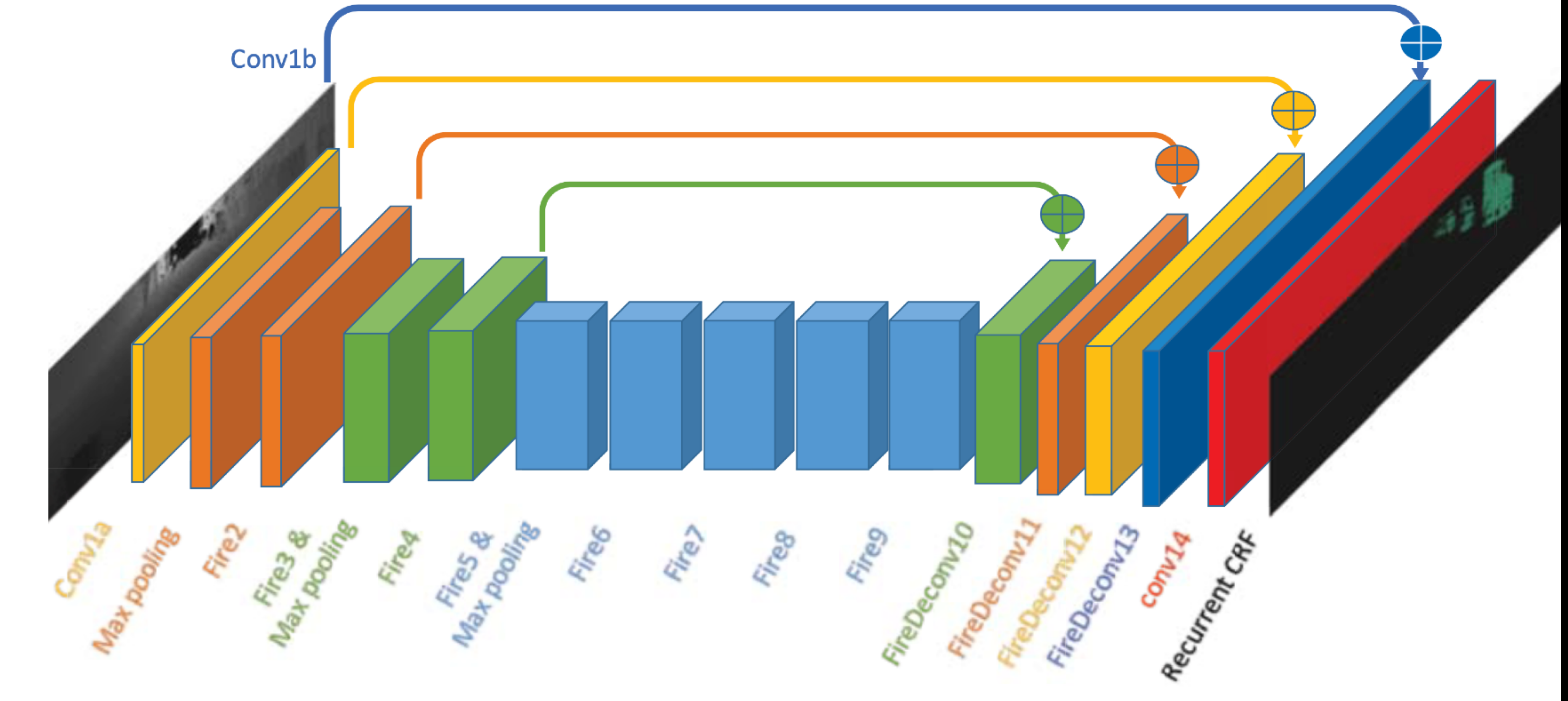
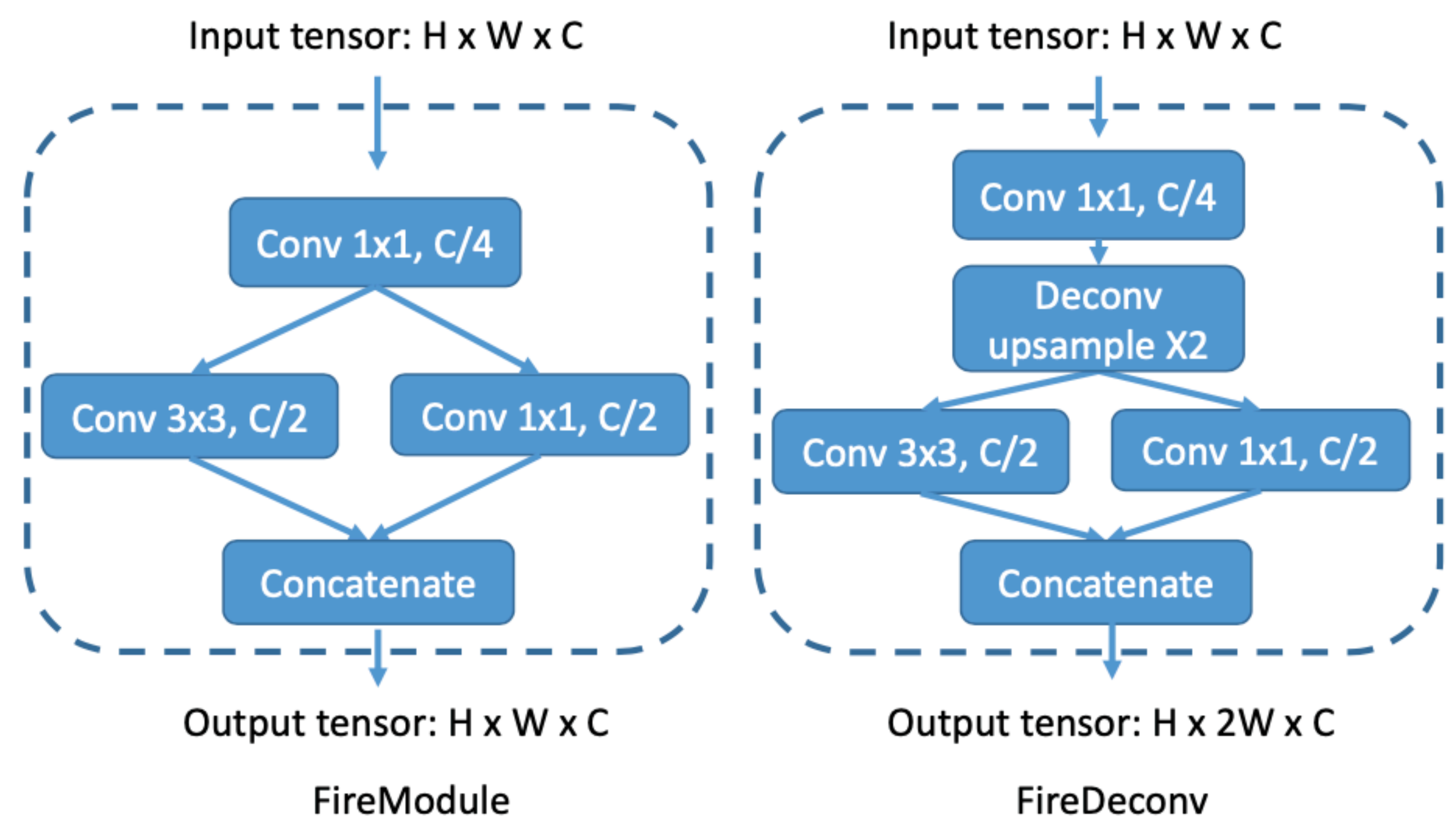


Fig. 3: Network structure of SqueezeSeg.



METHOD DESCRIPTION

Conditional Random Field

Energy function of CRF model

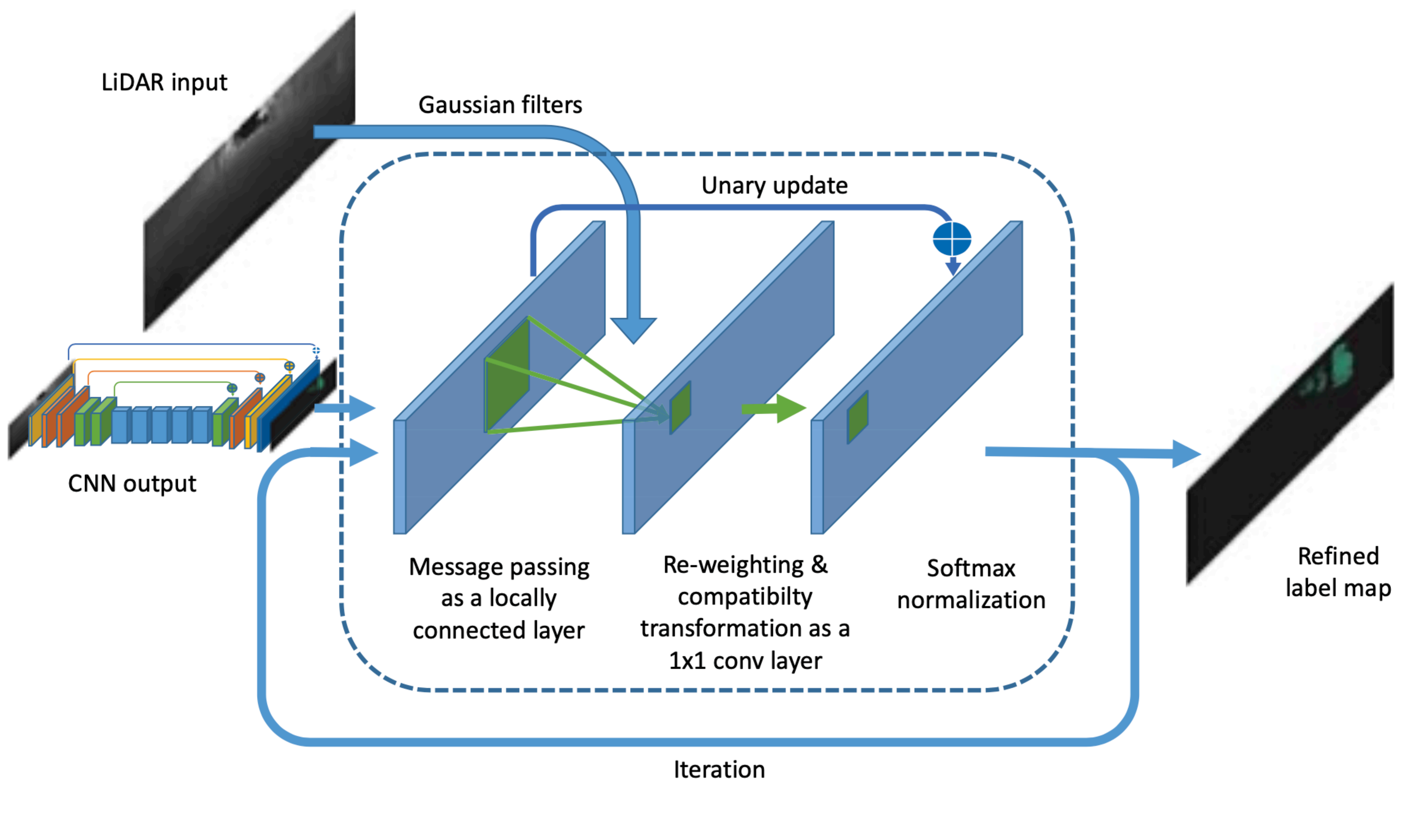
$$E(\mathbf{c}) = \sum_i u_i(c_i) + \sum_{i,j} b_{i,j}(c_i, c_j).$$

$$u_i(c_i) \;=\; -\log P(c_i)$$

$$b_{i,j}(c_i, c_j) = \mu(c_i, c_j) \sum_{m=1}^{M} w_m k^m(\mathbf{f_i}, \mathbf{f_j})$$

$$\mu(c_i, c_j) \;=\; 1 \text{ if } c_i \neq c_j \text{ and } 0 \text{ otherwise.}$$

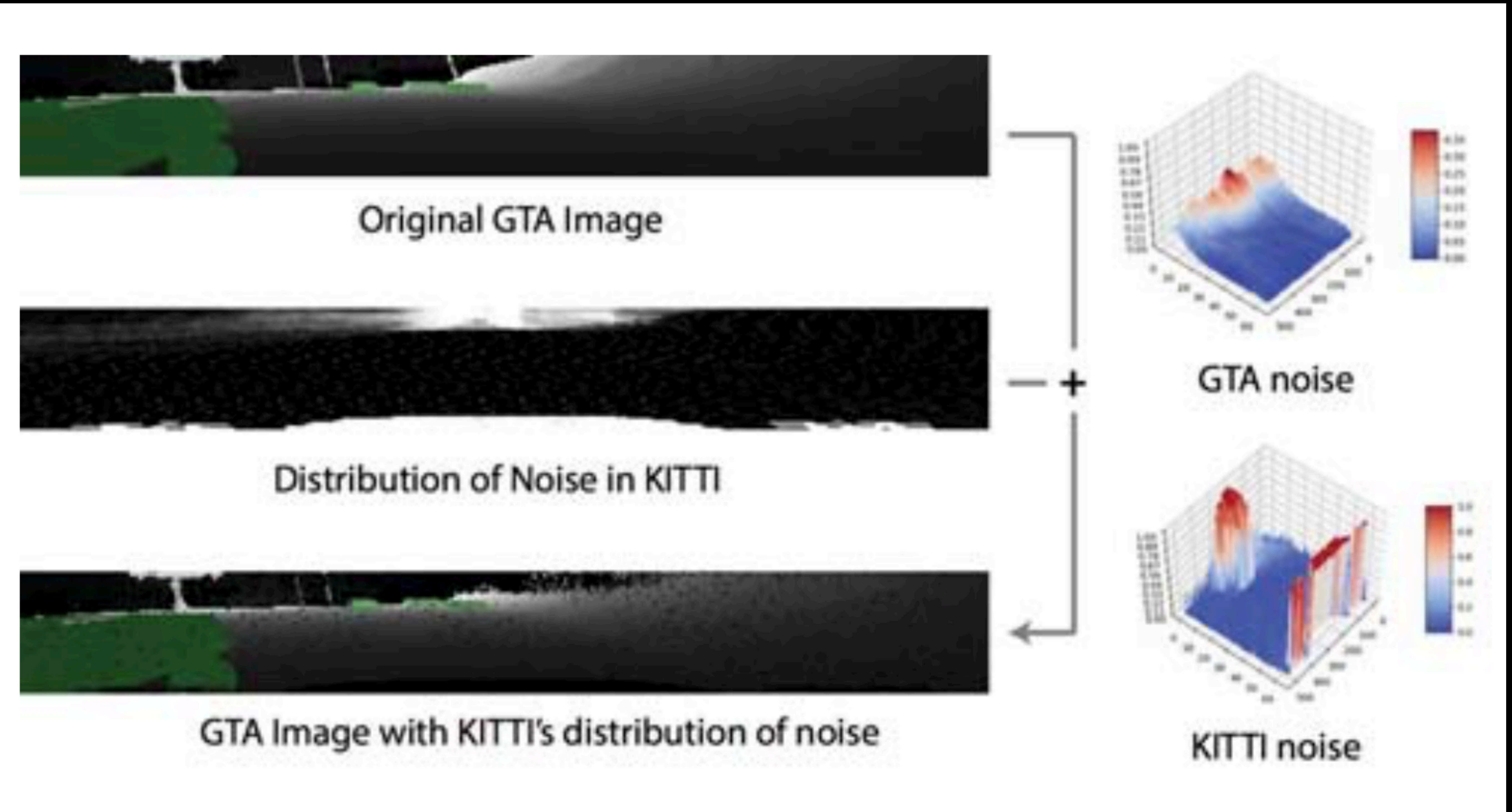
Conditional Random Field (CRF) as an RNN layer.



METHOD DESCRIPTION

Data collection

- KITTI raw dataset
- Synthesized dataset from GTAV
  - Add noise by KITTI’s distribution of noise



Experimental Result

TABLE I: Segmentation Performance of SqueezeSeg

		Class-level			Instance-level		
		P	R	**IoU**	P	R	**IoU**
car	w/ CRF	66.7	95.4	64.6	63.4	90.7	59.5
	w/o CRF	62.7	95.5	60.9	60.0	91.3	56.7
pedestrian	w/ CRF	45.2	29.7	21.8	43.5	28.6	20.8
	w/o CRF	52.9	28.6	22.8	50.8	27.5	21.7
cyclist	w/ CRF	35.7	45.8	25.1	30.4	39.0	20.6
	w/o CRF	35.2	51.1	26.4	30.1	43.7	21.7

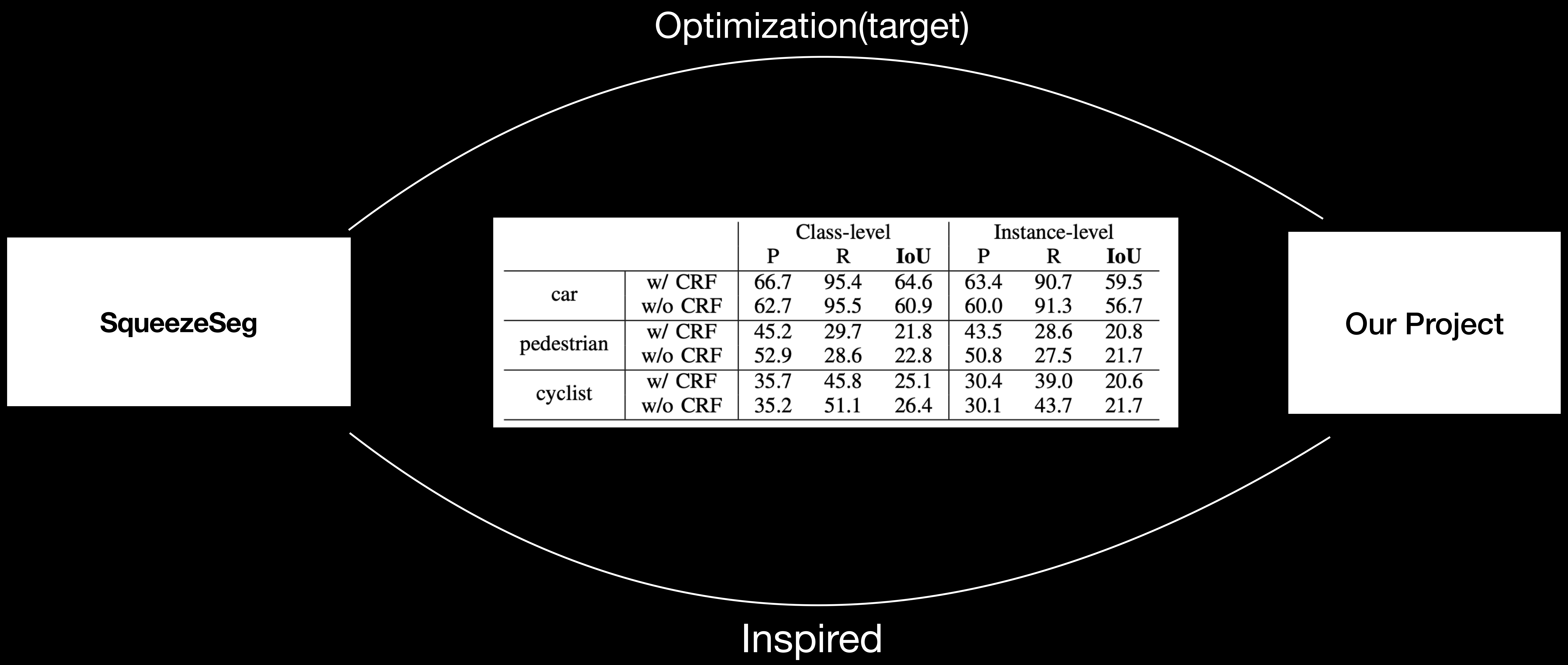
TABLE II: Average runtime and standard deviation of the SqueezeSeg Pipeline

unit: ms	Titan X	Drive PX2 AutoCruise	Drive PX2 AutoChauffeur	Xeon E5 CPU
SqueezeSeg	13.6/0.8	74.0/0.8	37.8/1.7	-
SqueezeSeg w/o CRF	8.7/0.5	52.0/1.3	25.1/0.8	-
DBSCAN	-	-	-	27.3/45.8

TABLE III: Segmentation Performance on the Car Category with Simulated Data

	Class-level			Instance-level		
	P	R	**IoU**	P	R	**IoU**
KITTI	58.9	95.0	57.1	56.1	90.5	53.0
GTA	30.4	86.6	29.0	29.7	84.6	28.2
KITTI + GTA	69.6	92.8	66.0	66.6	88.8	61.4

how this paper is relevant to our project



		Class-level			Instance-level		
		P	R	**IoU**	P	R	**IoU**
car	w/ CRF	66.7	95.4	64.6	63.4	90.7	59.5
	w/o CRF	62.7	95.5	60.9	60.0	91.3	56.7
pedestrian	w/ CRF	45.2	29.7	21.8	43.5	28.6	20.8
	w/o CRF	52.9	28.6	22.8	50.8	27.5	21.7
cyclist	w/ CRF	35.7	45.8	25.1	30.4	39.0	20.6
	w/o CRF	35.2	51.1	26.4	30.1	43.7	21.7