Comments and Corrections

Corrections to “Multi-Modal Sensor Fusion-Based Semantic Segmentation for Snow Driving Scenarios”

Sirawich Vachmanus, *Graduate Student Member, IEEE*, Ankit A. Ravankar, *Member, IEEE*, Takanori Emaru, *Member, IEEE*, and Yukinori Kobayashi

IN THE above article [1], we had mentioned the SSMA in many result tables, but unfortunately, we had made a mistake in the citation, which would have made the readers unable to link to the original material of the method. The faults in the above article have been corrected in this article. The self-supervised model adaptation (SSMA) [2] is a fusion module combining two feature maps. This module was developed for multi-modal semantic segmentation and achieved the best efficiency in many conditions.

Table II, Table V, Table VI, and Table VII in [1] are corrected to be in sequential order as Table I, Table III, Table II, and Table IV, respectively. The citation index of SSMA in all tables is corrected to SSMA [2]. In addition, the citation in Fig. 12 in the article [1] is also changed to SSMA [2], as shown in Fig. 1.

References

[1] S. Vachmanus, A. A. Ravankar, T. Emaru, and Y. Kobayashi, “Multi-modal sensor fusion-based semantic segmentation for snow driving scenarios,” *IEEE Sensors J.*, vol. 21, no. 15, pp. 16839–16851, Aug. 2021.

[2] A. Valada, R. Mohan, and W. Burgard, “Self-supervised model adaptation for multimodal semantic segmentation,” *Int. J. Comput. Vis.*, vol. 128, pp. 1239–1285, Jul. 2019.

[3] H. Zhao, X. Qi, X. Shen, J. Shi, and J. Jia, “ICNet for real-time semantic segmentation on high-resolution images,” *CoRR*, vol. abs/1704.08545, pp. 1–16, Apr. 2017.

[4] J. Fu, J. Liu, H. Tian, Z. Fang, and H. Lu, “Dual attention network for scene segmentation,” *CoRR*, vol. abs/1809.02983, pp. 1–10, Sep. 2018.

[5] Z. Tian, T. He, C. Shen, and Y. Yan, “Decoders matter for semantic segmentation: Data-dependent decoding enables flexible feature aggregation,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3121–3130.

[6] H. Zhao *et al.*, “PSANet: Point-wise spatial attention network for scene parsing,” in *Computer Vision,* V. Ferrari, M. Hebert, C. Sminchisescu, and Y. Weiss, Eds. Cham, Switzerland: Springer, 2018, pp. 270–286.

[7] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, “Pyramid scene parsing network,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6230–6239.

[8] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, “Rethinking atrous convolution for semantic image segmentation,” *CoRR*, vol. abs/1706.05587, pp. 1–14, Jun. 2017.

Manuscript received September 27, 2021; accepted September 27, 2021. Date of current version November 12, 2021. *(Corresponding author: Sirawich Vachmanus.)*

Sirawich Vachmanus is with the Laboratory of Robotics and Dynamics, Hokkaido University, Sapporo, Hokkaido 060-8628, Japan (e-mail: vachmas@gmail.com).

Ankit A. Ravankar is with the Department of Robotics, Division of Mechanical Engineering, Tohoku University, Sendai 980-8579, Japan (e-mail: ankit@srd.mech.tohoku.ac.jp).

Takanori Emaru is with the Department of Human Mechanical System and Design Engineering, Research Laboratory of Robotics and Dynamics, Hokkaido University, Sapporo, Hokkaido 060-8628, Japan (e-mail: emaru@eng.hokudai.ac.jp).

Yukinori Kobayashi is with the National Institute of Technology, Tomakomai College, Tomakomai, Hokkaido 059-1275, Japan (e-mail: kobay@eng.hokudai.ac.jp).

Digital Object Identifier 10.1109/JSEN.2021.3116408

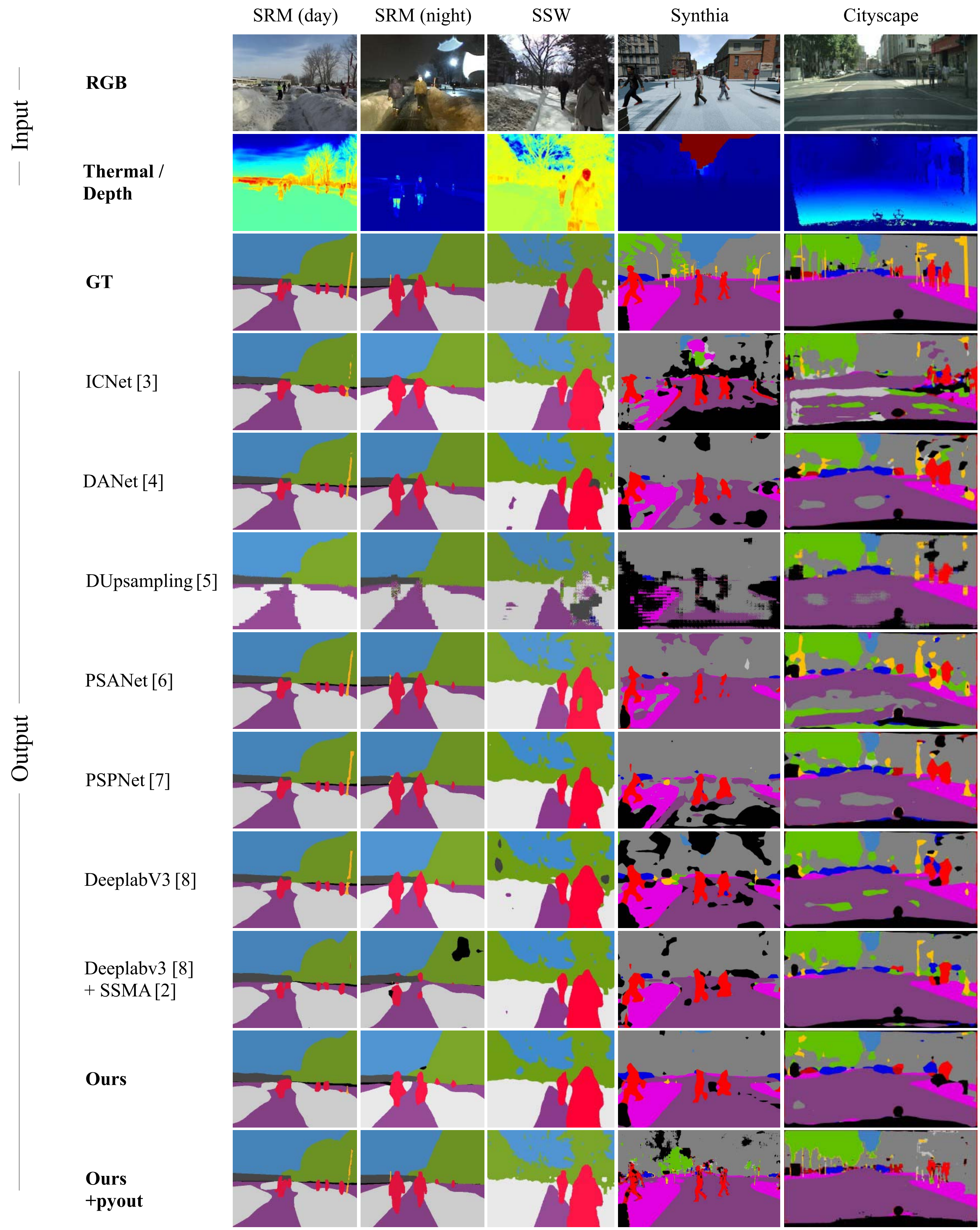

Fig. 1. Results of segmentation on all datasets.

TABLE I
THE COMPLEXITY COMPARISON OF NETWORKS

Network	Dataset RGB	Dataset T	Parameters	Memory	MACs	Time
ICNet [3]	✓	-	28.29M	0.93G	18.7G	52ms
	✓	✓	30.73M	1.15G	36.65G	112ms
DANet [4]	✓	-	49.62M	1.78G	75.3G	31ms
	✓	✓	49.63M	2.46G	150.6G	63ms
DUpsampling [5]	✓	-	34.53M	1.49G	57.19G	28ms
	✓	✓	39.47M	2.35G	109.78G	58ms
PSANet [6]	✓	-	52.97M	2.19G	144.85G	31ms
	✓	✓	71.85M	3.81G	289.67G	61ms
PSPNet [7]	✓	-	48.76M	1.69G	69.29G	31ms
	✓	✓	67.63M	2.79G	138.58G	59ms
DeeplabV3 [8]	✓	-	41.81M	1.46G	65.3G	31ms
	✓	✓	42.4M	2.21G	130.6G	59ms
DeeplabV3 [8] + SSMA [2]	✓	✓	42.54M	2.53G	130.77G	60ms
Ours	✓	✓	45.5M	2.27G	134.31G	58ms
Ours + pyout	✓	✓	109M	3.53G	393.29G	75ms

M is $\times 10^6$, G is $\times 10^9$

TABLE II
THE SEGMENTATION PERFORMANCE ON THE SRM DATASET AND SSW DATASET

ID	IoU (%) 1	2	3	4	5	6	7	8	9	mIoU(%)	mF1(%)
SRM dataset											
ICNet [3]	83.2	77.0	74.9	24.1	86.6	54.1	66.1	11.7	22.4	55.6	67.0
DANet [4]	85.4	91.3	85.7	37.7	87.7	91.2	72.4	18.9	35.9	67.4	76.9
DUpsampling [5]	84.2	90.4	81.0	0.1	87.0	93.7	64.2	0.9	39.9	60.2	67.0
PSANet [6]	82.4	88.5	80.3	38.7	87.9	91.8	58.9	16.2	47.2	65.8	76.1
PSPNet [7]	87.3	89.8	83.9	34.6	90.1	88.9	74.1	20.7	52.6	69.1	78.6
DeeplabV3 [8]	85.0	90.6	82.6	33.4	93.2	95.0	72.9	19.8	59.4	70.2	79.3
DeeplabV3 [8] + SSMA [2]	85.7	91.4	83.9	36.7	92.5	95.3	74.8	26.6	53.7	71.2	80.4
Ours	**87.5**	**92.7**	**85.4**	**36.4**	**94.4**	**95.9**	**75.8**	**27.5**	**63.3**	**73.2**	**81.9**
Ours + pyout	**91.8**	**95.5**	**91.4**	**53.5**	**95.6**	**96.5**	**81.8**	**31.9**	**67.4**	**78.4**	**86.5**
SSW dataset											
ICNet [3]	89.4	87.4	88.3	8.2	89.5	92.7	79.7	79.7	24.9	71.1	78.3
DANet [4]	88.2	86.9	87.2	16.1	86.4	88.1	82.8	78.9	26.2	71.2	79.4
DUpsampling [5]	87.3	83.7	80.2	0.5	86.9	90.5	1.5	22.2	0.0	50.3	55.7
PSANet [6]	89.0	87.4	87.5	18.0	87.7	89.7	85.8	70.2	19.4	70.5	78.5
PSPNet [7]	89.5	87.5	88.5	18.8	86.7	87.8	86.1	83.3	32.1	73.4	81.3
DeeplabV3 [8]	88.5	86.6	81.6	17.2	83.6	88.7	85.5	81.8	30.6	71.6	80.0
DeeplabV3 [8] + SSMA [2]	91.9	90.5	89.5	14.1	87.7	89.5	84.5	74.7	21.3	71.5	79.8
Ours	**88.0**	**86.5**	**88.5**	**18.2**	**87.0**	**89.2**	**88.9**	**85.3**	**32.2**	**73.8**	**81.5**
Ours + pyout	**94.1**	**93.5**	**92.2**	**33.8**	**93.8**	**95.3**	**93.7**	**84.5**	**49.1**	**81.1**	**87.6**

TABLE III
THE COMPARISON OF USING THE THERMAL MAP IN THE SRM DATASET

Network	Human IoU (%)	
	RGB	RGB-T
ICNet [3]	57.6	66.1
DANet [4]	67.2	72.4
DUpsampling [5]	2.3	64.2
PSANet [6]	68.4	58.9
PSPNet [7]	64.5	74.1
DeeplabV3 [8]	64.9	72.9
DeeplabV3 [8] + SSMA [2]	-	74.8
Ours	-	**75.8**
Ours + pyout	-	**78.4**

TABLE IV
THE SEGMENTATION PERFORMANCE ON THE CITYSCAPE DATASET AND SYNTHIA DATASET

Network	mIoU(%)	mAcc(%)	mF1(%)
		Cityscape	
ICNet [3]	40.0	96.0	51.7
DANet [4]	58.6	97.7	70.9
DUpsampling [5]	53.1	97.4	65.3
PSANet [6]	48.3	96.0	62.7
PSPNet [7]	56.1	97.5	68.6
DeeplabV3 [8]	59.8	97.7	72.3
DeeplabV3 [8] + SSMA [2]	58.2	97.5	71.4
Ours	**59.6**	**97.7**	**73.2**
Ours + pyout	**62.6**	**98.1**	**73.8**
		Synthia	
ICNet [3]	27.9	89.5	36.5
DANet [4]	51.3	94.0	64.2
DUpsampling [5]	33.0	91.0	44.1
PSANet [6]	55.0	97.2	67.9
PSPNet [7]	47.1	92.0	59.7
DeeplabV3 [8]	62.4	97.1	73.9
DeeplabV3 [8] + SSMA [2]	61.5	98.0	72.5
Ours	**65.1**	**98.1**	**76.1**
Ours + pyout	**69.4**	**98.7**	**78.5**