Multi-View Radar Semantic Segmentation

Arthur Ouaknine[1,2], Alasdair Newson[1], Patrick Pérez[2], Florence Tupin[1], Julien Rebut[2]

(1) Télécom Paris, (2) valeo.ai, arthur.ouaknine@telecom-paris.fr

ICCV 2021 October 11-17 Virtual

Motivation & Contributions

- Automotive radar: **robust to adverse weather conditions** while providing the **position and radial velocity** of surrounding objects.
- Contributions: **architectures** for multi-view radar semantic segmentation; a **combination of losses** adapted to objects' signatures with spatial coherence.

Combination of Losses

Let $f_\theta(\mathbf{x}) = \mathbf{p}$ be a segmentation model with parameters θ, input $\mathbf{x}$, output $\mathbf{p}$ and $\mathbf{y}$ the ground truth. Note $M \times N$ is the size of the view, such as $\Omega = [\![1, M]\!] \times [\![1, N]\!]$; and K the number of classes.

Weighted Cross Entropy

$$\mathcal{L}_{\text{wCE}}(\mathbf{y}, \mathbf{p}) = -\frac{1}{K}\sum_{k=1}^{K} w_k \sum_{(m,n)\in\Omega} \mathbf{y}[m,n,k] \log \mathbf{p}[m,n,k]$$

where $w_k \propto \left(\sum_{\mathbf{y}} \sum_{(m,n)\in\Omega} \mathbf{y}[m,n,k]\right)^{-1}$.

Soft Dice

$$\mathcal{L}_{\text{SDice}} = \frac{1}{K}\sum_{k=1}^{K}\left[1 - \frac{2\sum_{(m,n)} \mathbf{y}[m,n,k]\mathbf{p}[m,n,k]}{\sum_{(m,n)} \mathbf{y}^2[m,n,k] + \mathbf{p}^2[m,n,k]}\right]$$

Coherence

$$\mathcal{L}_{\text{CoL}}(\mathbf{p}^{\text{RD}}, \mathbf{p}^{\text{RA}}) = \|\tilde{\mathbf{p}}^{\text{RD}} - \tilde{\mathbf{p}}^{\text{RA}}\|_{\text{F}}^2$$

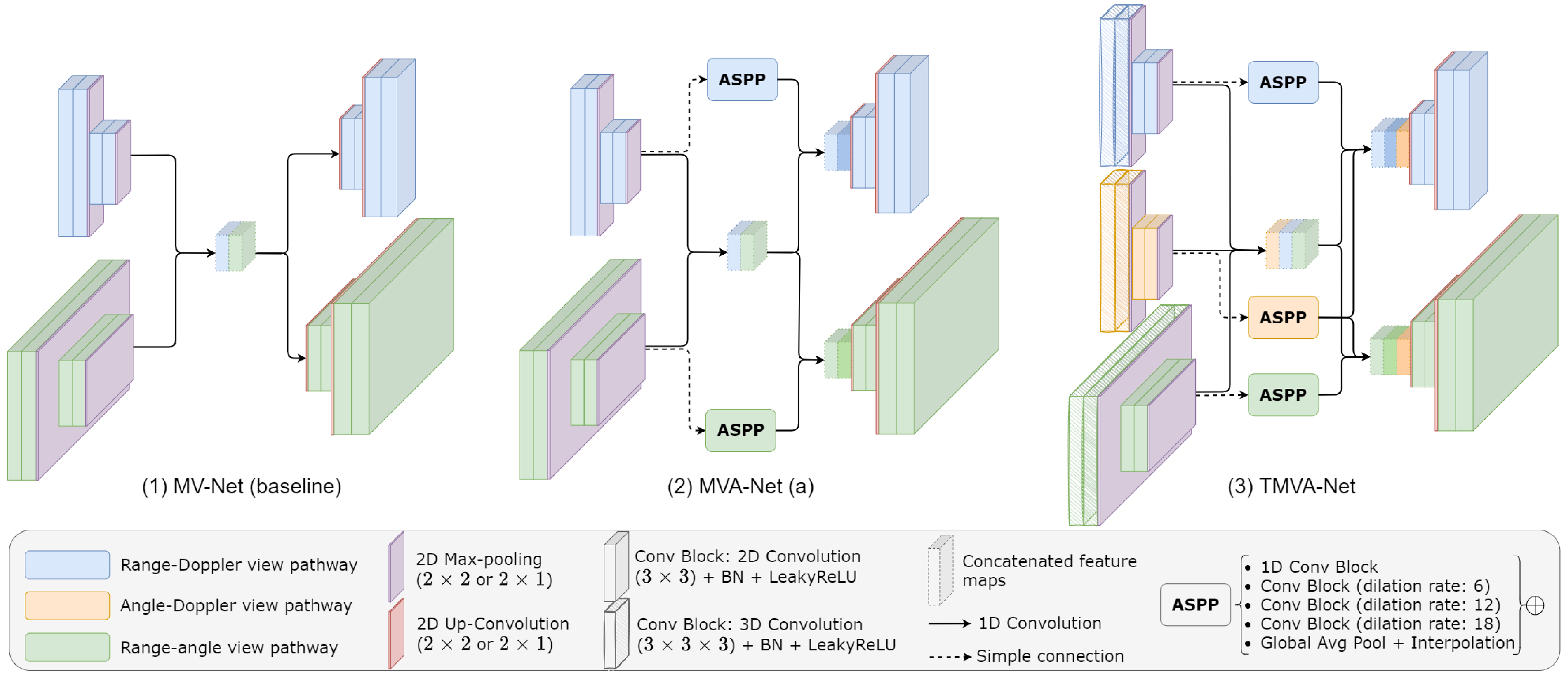
where $\|\cdot\|_{\text{F}}$ denotes the Frobenius norm; $\tilde{\mathbf{p}}^{\text{RD}}$, $\tilde{\mathbf{p}}^{\text{RA}}$ the highest probability maps of each range bin for each class.

Combination

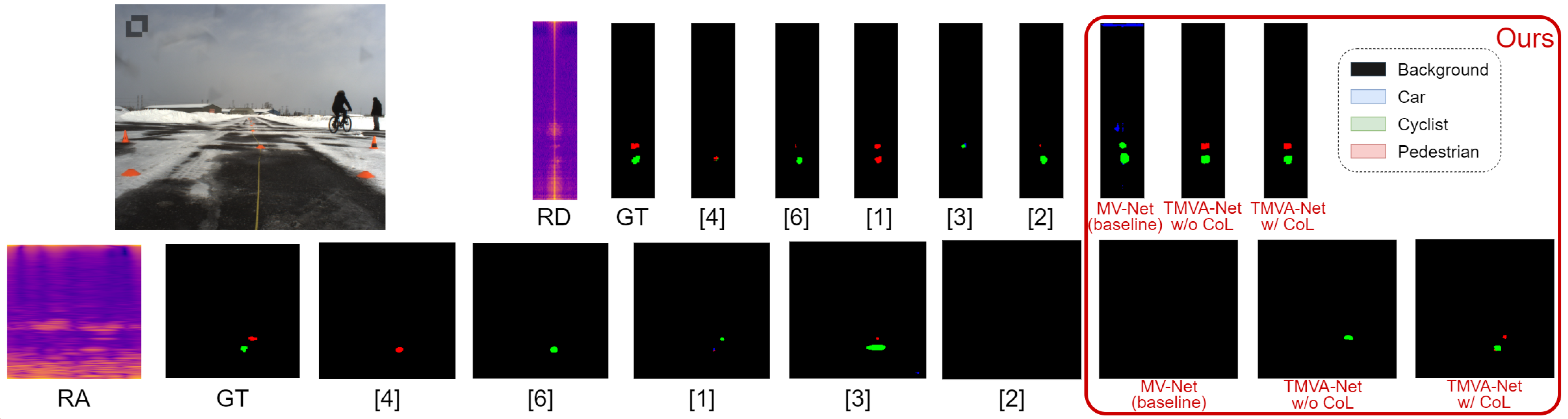
$$\mathcal{L} = \lambda_{\text{wCE}}(\mathcal{L}_{\text{wCE}}^{\text{RD}} + \mathcal{L}_{\text{wCE}}^{\text{RA}}) + \lambda_{\text{SDice}}(\mathcal{L}_{\text{SDice}}^{\text{RD}} + \mathcal{L}_{\text{SDice}}^{\text{RA}}) + \lambda_{\text{CoL}}\mathcal{L}_{\text{CoL}}$$

where λ_{wCE}, λ_{SDice} and λ_{CoL} are weighting factors.

Proposed Architectures



Qualitative Results on the CARRADA-Test Dataset [5]



Ablation Studies

Let RD and RA be the range-Doppler and range-angle views.

View	Method	#Param.↓	mIoU↑	mDice↑
RD	MV-Net (baseline)	2.4M	29.0	39.8
	MVA-Net (a)	3.6M	48.9	60.4
	MVA-Net (b)	4.8M	52.9	64.3
	TMVA-Net	5.6M	**59.3**	**71.5**
RA	MV-Net (baseline)	2.4M	26.8	28.5
	MVA-Net (a)	3.6M	28.1	31.1
	MVA-Net (b)	4.8M	36.7	43.9
	TMVA-Net	5.6M	**40.1**	**49.3**

Table 1: Performances of the proposed architectures.

Loss	RD view mIoU	RD view mDice	RA view mIoU	RA view mDice
CE	56.1	67.8	39.1	48.3
SDice	58.5	70.3	37.1	44.8
wCE	51.1	62.8	34.3	41.1
CE+SDice	45.2	54.0	38.8	46.9
wCE+SDice	**59.3**	**71.5**	40.1	49.3
wCE+SDice+CoL	58.7	70.9	**41.3**	**51.0**

Table 2: Performances of various loss combinations.

Quantitative Results

View	Method	#Param.(M)↓	mIoU↑	mDice↑
RD	FCN-8s [4]	134.3	54.7	66.3
	U-Net [6]	17.3	55.4	68.0
	DeepLabv3+ [1]	59.3	50.8	61.6
	RSS-Net [3]	10.1	32.1	36.9
	RAMP-CNN [2]	106.4	56.6	68.5
	MV-Net (ours-base.)	2.4*	29.0	32.8
	TMVA-Net (ours)	5.6*	58.7	70.9
RA	FCN-8s [4]	134.3	34.5	40.9
	U-Net [6]	17.3	32.8	38.2
	DeepLabv3+ [1]	59.3	32.7	38.3
	RSS-Net [3]	10.1	32.1	37.8
	RAMP-CNN [2]	106.4	27.9	30.5
	MV-Net (ours-base.)	2.4*	26.8	28.5
	TMVA-Net (ours)	5.6*	41.3	51.0

Table 3: Performances on the CARRADA-Test dataset [5] for RD and RA views, ('*') corresponds to a single model that segments both views.

References

[1] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *ECCV*, 2018.
[2] X. Gao, G. Xing, S. Roy, and H. Liu. RAMP-CNN: A Novel Neural Network for Enhanced Automotive Radar Object Recognition. In *Sensor Journal*, 2020.
[3] P. Kaul, D. De Martini, M. Gadd, and P. Newman. RSS-Net: Weakly-Supervised Multi-Class Semantic Segmentation with FMCW Radar. In *IV*, 2020.
[4] J. Long, E. Shelhamer, and T. Darrell. Fully convolutional networks for semantic segmentation. In *CVPR*, 2015.
[5] A. Ouaknine, A. Newson, J. Rebut, F. Tupin, and P. Perez. CARRADA dataset: camera and automotive radar with range- angle- Doppler annotations. In *ICPR*, 2020.
[6] O. Ronneberger, P. Fischer, and T. Brox. U-Net: Convolutional Networks for Biomedical Image Segmentation. In *MICCAI*, 2015.

Github: `https://github.com/valeoai/MVRSS` **Website:** `https://arthurouaknine.github.io/codeanddata/mvrss`