

# Supplementary Material of Inter-Region Affinity Distillation for Road Marking Segmentation

Yuenan Hou<sup>1</sup>, Zheng Ma<sup>2</sup>, Chunxiao Liu<sup>2</sup>, Tak-Wai Hui<sup>1</sup>, and Chen Change Loy<sup>3†</sup>

<sup>1</sup>The Chinese University of Hong Kong <sup>2</sup>SenseTime Group Limited <sup>3</sup>Nanyang Technological University  
<sup>1</sup>{hy117, twhui}@ie.cuhk.edu.hk, <sup>2</sup>{mazheng, liuchunxiao}@sensetime.com, <sup>3</sup>ccloy@ntu.edu.sg

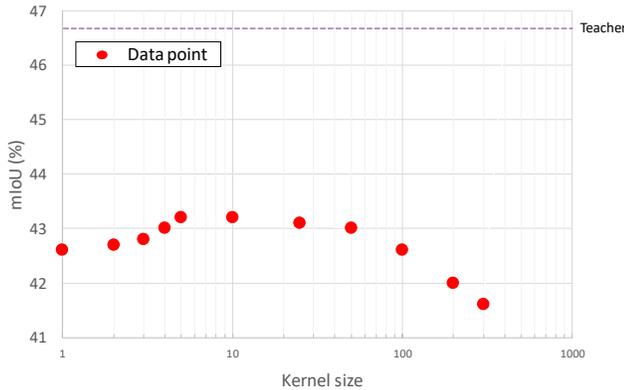


Figure 1: Performance of ERFNet-IntRA-KD with different kernel sizes on ApolloScape testing set.

## 1. Effect of the size of the smoothing kernel

From Fig. 1, we can see that when the kernel size is  $1 \times 1$ , the performance of IntRA-KD is slightly degraded since the neighbouring areas around road markings are missed in the distillation process. When the kernel size ranges from 2 to 50, the performance remains almost the same, suggesting the stability of IntRA-KD. However, when the kernel size continues to increase, the performance of IntRA-KD drops quickly since too many background areas are included and it will make the moment vector of each class inaccurate. In the experiment, the kernel size is set as 5.

## 2. Performance of ENet and ResNet-18 using different distillation methods on three benchmarks

Tables 1- 3 record the performance of different algorithms using ENet and ResNet-18 as backbones on three benchmarks.

†: Corresponding author.

## 3. Comparison between IntRA-KD and previous distillation methods

We compare the learning complexity and performance gains of using different forms of knowledge in Table 4. Suppose the input image is  $\mathbf{X} \in \mathbb{R}^{h \times w \times 3}$ , the feature map is  $\mathbf{F} \in \mathbb{R}^{h_f \times w_f \times c}$  and  $n$  is the number of classes. Since we use high-level and middle-level feature maps as distillation targets,  $h_f$  is either  $\frac{1}{4}h$  or  $\frac{1}{8}h$ . The number of classes  $n$  is usually smaller than  $\frac{1}{10}c$  and  $\frac{1}{10}h_f$ . It is evident that the introduced affinity graph has the least learning complexity but brings most performance gains over other distillation methods. The above result apparently showcases the importance of utilizing the structural knowledge in road marking segmentation.

## 4. More qualitative results of different algorithms on three benchmarks

In this section, we show more qualitative results of our IntRA-KD and BiFPN [14] (the most competitive baseline) on three benchmarks. As shown in (a) and (c) of Fig. 2, ERFNet-IntRA-KD makes more accurate predictions on long and thin road markings. As for other challenging scenarios (*e.g.*, crowded roads and poor light conditions), predictions made by ERFNet and ERFNet-BiFPN are inaccurate and messy. By contrast, the predictions of ERFNet-IntRA-KD are more complete and accurate in these challenging conditions. The experimental results strongly showcase the effectiveness of IntRA-KD.

## 5. Visualization of the affinity graph of different algorithms

From Fig. 3, we can see that IntRA-KD not only improves the predictions of ERFNet, but also makes the generated affinity graph more rational compared with BiFPN. More specifically, with IntRA-KD, those spatially close and visually similar road markings are pulled closer and those spatially distant and visually different markings are pulled apart in the feature space. The provided samples

Table 1: Performance of ENet and ResNet-18 using different distillation methods on ApolloScape testing set.

(a) ENet			(b) ResNet-18		
Type	Algorithm	mIoU	Type	Algorithm	mIoU
Teacher	<b>ResNet-101 [3]</b>	<b>46.6</b>	Teacher	<b>ResNet-101 [3]</b>	<b>46.6</b>
Student	ENet [10]	39.8	Student	ResNet-18 [3]	40.0
Self distillation	ENet-DKS [11]	40.1	Self distillation	ResNet-18-DKS [11]	40.3
	ENet-SAD [5]	40.3		ResNet-18-SAD [5]	40.5
Teacher-student distillation	ENet-KD [4]	40.1	Teacher-student distillation	ResNet-18-KD [4]	40.2
	ENet-SKD [8]	40.2		ResNet-18-SKD [8]	40.4
	ENet-PS-N [12]	40.0		ResNet-18-PS-N [12]	40.1
	ENet-IRG [7]	40.4		ResNet-18-IRG [7]	40.5
	ENet-BiFPN [14]	40.6		ResNet-18-BiFPN [14]	41.1
	<b>ENet-IntRA-KD (ours)</b>	<b>41.6</b>		<b>ResNet-18-IntRA-KD (ours)</b>	<b>42.4</b>

Table 2: Performance of ENet and ResNet-18 using different distillation methods on CULane testing set.

(a) ENet			(b) ResNet-18		
Type	Algorithm	$F_1$ -measure	Type	Algorithm	$F_1$ -measure
Teacher	<b>ResNet-101 [5]</b>	<b>72.8</b>	Teacher	<b>ResNet-101 [5]</b>	<b>72.8</b>
Student	ENet [10]	69.8	Student	ResNet-18 [3]	69.4
Self distillation	ENet-DKS [11]	70.3	Self distillation	ResNet-18-DKS [11]	69.8
	ENet-SAD [5]	70.8		ResNet-18-SAD [5]	70.5
Teacher-student distillation	ENet-KD [4]	70.0	Teacher-student distillation	ResNet-18-KD [4]	69.7
	ENet-SKD [8]	70.2		ResNet-18-SKD [8]	70.0
	ENet-PS-N [12]	70.1		ResNet-18-PS-N [12]	69.8
	ENet-IRG [7]	70.4		ResNet-18-IRG [7]	70.2
	ENet-BiFPN [14]	71.0		ResNet-18-BiFPN [14]	70.8
	<b>ENet-IntRA-KD (ours)</b>	<b>71.8</b>		<b>ResNet-18-IntRA-KD (ours)</b>	<b>71.4</b>

Table 3: Performance of ENet and ResNet-18 using different distillation methods on LLAMAS testing set.

(a) ENet			(b) ResNet-18		
Type	Algorithm	mAP	Type	Algorithm	mAP
Teacher	<b>ResNet-101 [3]</b>	<b>0.607</b>	Teacher	<b>ResNet-101 [3]</b>	<b>0.607</b>
Student	ENet [10]	0.562	Student	ResNet-18 [3]	0.565
Self distillation	ENet-DKS [11]	0.564	Self distillation	ResNet-18-DKS [11]	0.568
	ENet-SAD [5]	0.566		ResNet-18-SAD [5]	0.569
Teacher-student distillation	ENet-KD [4]	0.564	Teacher-student distillation	ResNet-18-KD [4]	0.566
	ENet-SKD [8]	0.567		ResNet-18-SKD [8]	0.570
	ENet-PS-N [12]	0.566		ResNet-18-PS-N [12]	0.569
	ENet-IRG [7]	0.567		ResNet-18-IRG [7]	0.571
	ENet-BiFPN [14]	0.572		ResNet-18-BiFPN [14]	0.576
	<b>ENet-IntRA-KD (ours)</b>	<b>0.585</b>		<b>ResNet-18-IntRA-KD (ours)</b>	<b>0.590</b>

strongly indicate the effectiveness of IntRA-KD in transferring the structural knowledge from the teacher model to the student model.

## 6. Performance of ERFNet and SwiftNetRN-18 on four benchmarks

We test IntRA-KD on Cityscapes [2] using DeepLabv3+ [1] as teacher and ERFNet and SwiftNetRN-18 [9] as students. IntRA-KD consistently improves the mIoU of ERFNet and SwiftNetRN-18 from {69.8, 75.5} to {71.4, 76.6}, respectively. Besides, we have also ap-

Table 4: Comparison between learning complexity and performance gains on mIoU when using different forms of knowledge. The performance gains are calculated on ApolloScape testing set using ERFNet as student and ResNet-101 as teacher.

Knowledge form	Complexity	Gains
Probability map [4]	$\mathcal{O}(hwn)$	0.3
Feature map [6]	$\mathcal{O}(h_f w_f c)$	0.4
Pairwise similarity map [8]	$\mathcal{O}(h_f^2 w_f^2)$	0.3
Attention map [13]	$\mathcal{O}(h_f w_f)$	0.7
<b>Inter-region affinity graph (ours)</b>	$\mathcal{O}(n^2)$	<b>1.9</b>

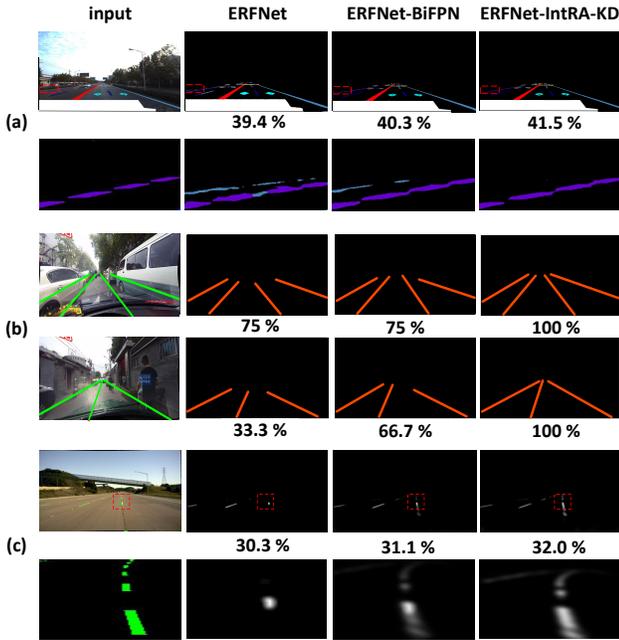


Figure 2: Qualitative results of different algorithms on (a) ApolloScape, (b) CULane and (c) LLAMAS testing sets. The number below each image denotes the accuracy for (a) and (c),  $F_1$ -measure for (b). Ground-truth labels are drawn on the input image. The second row of (a) and (c) are the enlarged areas covered by the red dashed rectangle.

plied IntRA-KD to SwiftNetRN-18 on all benchmarks by taking ResNet-101 as teacher. From Table. 5, it is evident that IntRA-KD brings significant performance gains to SwiftNetRN-18 in all benchmarks. The promising results validate the effectiveness and generality of the proposed IntRA-KD.

## References

[1] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. Encoder-decoder with atrous separable convolution for semantic image segmentation. In *European conference on computer vision*, pages 801–818,

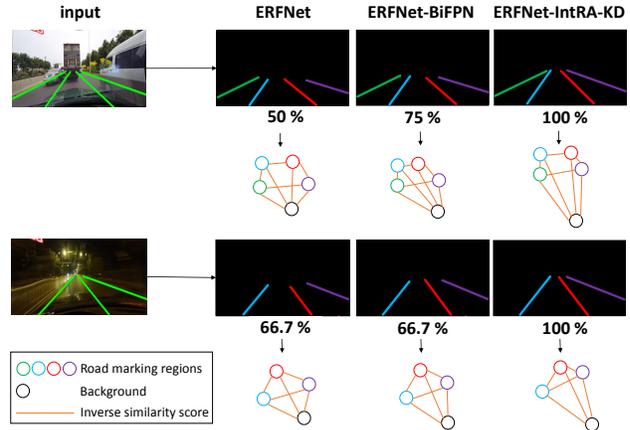


Figure 3: Visualization of the affinity graph generated by different methods. For visualization, we represent edges in an affinity graph by the inverse similarity score. The number below the segmentation map is  $F_1$ -measure.

2018. 2

[2] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The Cityscapes dataset for semantic urban scene understanding. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 3213–3223, 2016. 2

[3] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 2

[4] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *Statistics*, 1050:9, 2015. 2, 3

[5] Yuenan Hou, Zheng Ma, Chunxiao Liu, and Chen Change Loy. Learning lightweight lane detection CNNs by self attention distillation. In *IEEE International Conference on Computer Vision*, pages 1013–1021, 2019. 2

[6] Yuenan Hou, Zheng Ma, Chunxiao Liu, and Chen Change Loy. Learning to steer by mimicking features from heterogeneous auxiliary networks. In *Association for the Advancement of Artificial Intelligence*, volume 33, pages 8433–8440, 2019. 3

[7] Yufan Liu, Jiajiong Cao, Bing Li, Chunfeng Yuan, Weiming Hu, Yangxi Li, and Yunqiang Duan. Knowledge distillation via instance relationship graph. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 7096–7104, 2019. 2

[8] Yifan Liu, Ke Chen, Chris Liu, Zengchang Qin, Zhenbo Luo, and Jingdong Wang. Structured knowledge distillation for semantic segmentation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2604–2613, 2019. 2, 3

[9] Marin Orsic, Ivan Kreso, Petra Bevandic, and Sinisa Segvic. In defense of pre-trained ImageNet architectures for real-time semantic segmentation of road-driving images. In *IEEE*

Table 5: Performance of ERFNet and SwiftNetRN-18 on four benchmarks.

Algorithm	ApolloScape	CULane	LLAMAS	Cityscapes
	mIoU	$F_1$	mAP	mIoU
ERFNet	40.4	70.2	0.570	69.8
ERFNet + IntRA-KD	43.2	72.4	0.598	71.4
SwiftNetRN-18	41.3	70.8	0.578	75.5
SwiftNetRN-18 + IntRA-KD	44.0	72.7	0.603	76.6

*Conference on Computer Vision and Pattern Recognition*, pages 12607–12616, 2019. 2

- [10] Eduardo Romera, José M Alvarez, Luis M Bergasa, and Roberto Arroyo. ERFNet: Efficient residual factorized convnet for real-time semantic segmentation. *IEEE Transactions on Intelligent Transportation Systems*, 19(1):263–272, 2017. 2
- [11] Dawei Sun, Anbang Yao, Aojun Zhou, and Hao Zhao. Deeply-supervised knowledge synergy. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 6997–7006, 2019. 2
- [12] Junho Yim, Donggyu Joo, Jihoon Bae, and Junmo Kim. A gift from knowledge distillation: Fast optimization, network minimization and transfer learning. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4133–4141, 2017. 2
- [13] Sergey Zagoruyko and Nikos Komodakis. Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer. In *International Conference on Learning Representations*, 2017. 3
- [14] Lei Zhu, Zijun Deng, Xiaowei Hu, Chi-Wing Fu, Xuemiao Xu, Jing Qin, and Pheng-Ann Heng. Bidirectional feature pyramid network with recurrent attention residual modules for shadow detection. In *European Conference on Computer Vision*, pages 121–136, 2018. 1, 2