AFPSNet: Multi-Class Part Parsing based on Scaled Attention and Feature Fusion
Supplementary Material

Njuod Alsudays Jing Wu Yu-Kun Lai Ze Ji
Cardiff University, UK
{alsudaysn, wuj11, laiy4, jiz1}@cardiff.ac.uk

6. Additional Results on Pascal-Part Datasets

Herein we present further experimental results to compare the proposed AFPSNet with its baseline (DeepLab v3+ [1]), and the state-of-the-art multi-class part parsing methods [4, 2, 3].

In Table 1, we compare the per-part IoU on the 58 part classes achieved using these methods [1]. As can be seen, AFPSNet achieves the highest IoU for 36 out of 58 part classes, more than other methods compared. AFPSNet shows especially superior performance in segmenting bottle cap, person arm, sheep leg, horse/dog/cow tail, etc., achieving 0.05 higher IoU than the second best performed method. These parts are all relatively small in size, indicating our method is especially doing well in detecting small parts of objects.

Fig. 1 shows qualitative results on images with small parts. As can be seen in the first, second, fourth and eighth rows, using our AFPSNet, the detection and segmentation of dog/horse/cow tails are indeed more complete and with more accurate boundaries than the other methods. And similar observations can be made on other small parts, such as the human head and body in the third row where the person rides a motorbike, the aeroplane engine in the fifth row, and the car light in the sixth row, etc.

Fig. 2 shows more qualitative results comparing these state-of-the-art multi-class part parsing methods. AFPSNet shows overall better boundary localization and segmentation results, which confirms the results reported in Table 3 in the main paper. For example, the segmentation result of the bottle in the second row, the sheep head in the eighth row and the cat in the last row are challenging for the other methods, while AFPSNet can successfully detect and segment them. AFPSNet can better predict the boundaries of the bus in the third row, the horse tail in the fifth row and the human parts in the sixth and seventh rows. And it also achieves better segmentation results on the cows in the fourth row.

In Table 2, we compare the per-part IoU on the 108 part classes achieved by AFPSNet, the baseline, and the reported results from state-of-the-art multi-class methods [4, 2]. The results show that our model achieves the highest mIoU for 80 out of 108 part classes (including background), more than other methods compared. AFPSNet shows especially superior performance in segmenting small parts, *e.g.*, bike handlebar, bus door, car headlight, cow horn, dog eye, etc.

Fig. 3 shows qualitative results comparing our method with the baseline and GMNet. AFPSNet shows overall better segmentation results with less missing parts and more accurate boundaries. As can be seen, AFPSNet can better predict the boundaries of the bus in the first row, the car in the fourth row and the horse in the sixth row. Also, it achieves better segmentation results on the bike handlebar in the second row, the cat in the third row, the child hair in the seventh row and the sheep head in the last row. Additionally, AFPSNet shows superior performance in detecting small object parts, *e.g.*, the dog eyes in the fifth row and the human leg in the eighth row.

[1] CSR [3] is not included in the comparison because they did not publish their results on per-part classes.

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 *Proceedings of the European conference on computer vision (ECCV)*, pages 801–818, 2018.

[2] Umberto Michieli, Edoardo Borsato, Luca Rossi, and Pietro Zanuttigh. Gmnet: Graph matching network for large scale part semantic segmentation in the wild. In *European Conference on Computer Vision*, pages 397–414. Springer, 2020.

[3] Xin Tan, Jiachen Xu, Zhou Ye, Jinkun Hao, and Lizhuang Ma. Confident semantic ranking loss for part parsing. In *2021 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2021.

[4] Yifan Zhao, Jia Li, Yu Zhang, and Yonghong Tian. Multi-class part parsing with joint boundary-semantic awareness. In *Pro-*

Table 1. Segmentation performance per-part IoU on the 58 part classes of PASCAL-Part-58 dataset.

Parts name	DeepLab v3	DeepLab v3+	BSANet	GMNet	AFPSNet	Parts name	DeepLab v3	DeepLab v3+	BSANet	GMNet	AFPSNet
	IoU	IoU	IoU	IoU	IoU		IoU	IoU	IoU	IoU	IoU
background	91.1	90.2	91.6	92.7	**94.8**	cow tail	0.0	1.0	7.9	8.1	**21.4**
aeroplane body	66.6	68.4	**70.0**	69.6	69.7	cow leg	46.1	55.1	53.4	53.3	**59.2**
aeroplane engine	25.7	27.8	29.1	25.7	**31.0**	cow torso	69.9	74.0	73.5	77.1	**78.3**
aeroplane wing	33.5	38.3	38.3	34.2	**42.3**	dining table	43.0	43.1	43.7	**51.3**	44.6
aeroplane stern	57.1	52.6	59.2	57.2	**60.5**	dog head	78.7	81.7	82.5	**85.0**	84.5
aeroplane wheel	45.4	50.5	**53.2**	46.8	51.1	dog leg	48.1	50.8	53.8	53.8	**56.2**
bike wheel	78.0	75.7	78.0	**81.3**	79.8	dog tail	27.1	32.6	31.3	31.4	**39.3**
bike body	48.4	52.2	53.4	51.5	**56.3**	dog torso	63.7	62.9	65.7	**68.0**	67.5
bird head	64.6	71.8	**74.0**	71.1	72.5	horse head	74.7	75.4	76.6	73.9	**82.1**
bird wing	35.1	38.3	39.7	38.6	**44.5**	horse tail	47.0	47.2	51.0	50.4	**57.2**
bird leg	29.3	34.1	34.8	28.7	**35.5**	horse leg	55.9	62.3	61.6	59.3	**63.9**
bird torso	66.9	66.8	**70.9**	69.5	70.2	horse torso	70.3	72.8	74.9	73.9	**78.4**
boat	54.4	64.0	60.2	**70.0**	64.0	mbike wheel	70.9	69.9	71.6	**73.5**	73.0
bottle cap	30.7	28.9	29.8	33.9	**39.6**	mbike body	65.1	71.5	71.5	**74.3**	72.6
bottle body	68.8	70.5	68.6	**77.6**	75.8	person head	83.5	84.8	85.0	84.7	**86.2**
bus window	72.7	74.5	74.8	75.4	**78.3**	person	65.9	65.9	68.2	67.0	**71.3**
bus wheel	55.3	55.5	57.1	58.1	**58.2**	person larm	46.9	48.7	52.0	48.6	**55.7**
bus body	74.8	77.6	78.3	**79.9**	79.6	person uarm	51.5	48.6	54.4	52.4	**58.9**
car window	62.6	66.7	68.1	64.8	**71.2**	person lleg	38.6	39.4	43.5	40.2	**46.0**
car wheel	64.8	**72.1**	68.5	70.3	71.9	person uleg	43.8	44.5	47.4	44.5	**50.3**
car light	46.2	53.5	53.7	48.4	**57.6**	pplant pot	45.3	50.0	53.5	56.0	**57.3**
car plate	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	pplant plant	52.4	**59.9**	56.6	56.4	59.3
car body	72.1	76.2	77.0	77.6	**78.0**	sheep head	60.9	70.8	65.4	70.8	**72.1**
cat head	80.2	82.3	83.7	83.8	**85.0**	sheep leg	8.6	19.3	11.7	14.3	**25.4**
cat leg	48.6	47.3	50.1	49.4	**53.2**	sheep torso	68.3	73.0	71.6	**75.6**	73.5
cat tail	40.2	45.9	48.8	46.0	**49.1**	sofa	43.2	42.4	43.1	**56.1**	46.4
cat torso	70.3	69.6	72.6	**73.8**	73.0	train	79.6	82.6	82.2	**85.0**	81.6
chair	35.4	38.2	36.5	**51.4**	39.8	tv screen	69.5	69.1	73.1	**77.0**	74.0
cow head	74.3	77.2	76.4	80.7	**83.7**	tv frame	45.9	46.3	49.8	**54.1**	52.1

ceedings of the IEEE/CVF International Conference on Computer Vision, pages 9177–9186, 2019.

Figure 1. A qualitative comparison of the segmentation results of the state-of-the-art multi-class part parsing methods to show their performance in detecting small object parts.

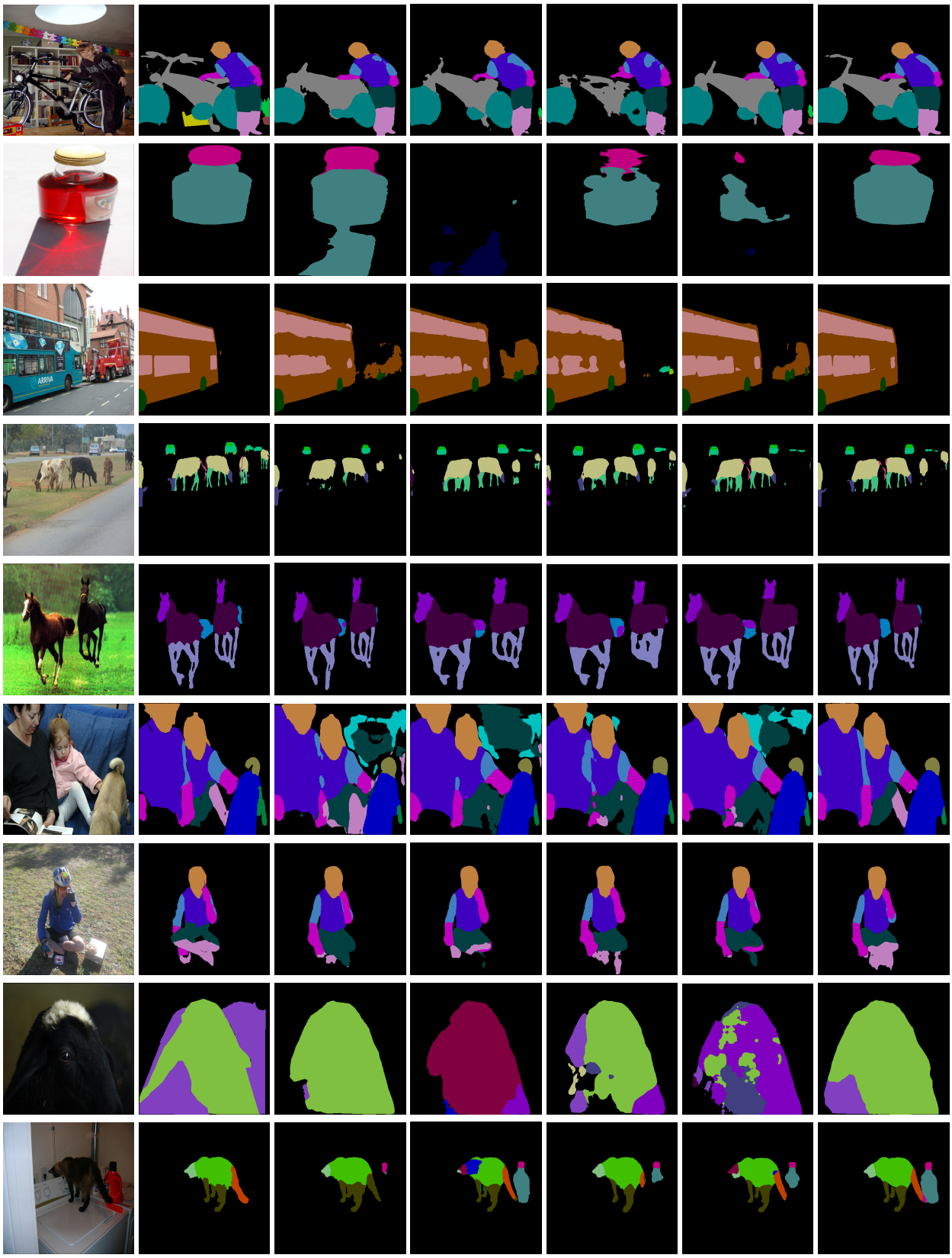

Figure 2. Segmentation results on PASCAL-Part-58 dataset. Our model shows overall better segmentation results with less missing parts and more accurate boundaries compared to the-state-of-the-art models.

Table 2. Segmentation performance per-part IoU on the 108 part classes of PASCAL-Part-108 dataset.

Parts name	DeepLab v3	DeepLab v3+	BSANet	GMNet	AFPSNet	Parts name	DeepLab v3	DeepLab v3+	BSANet	GMNet	AFPSNet
	IoU	IoU	IoU	IoU	IoU		IoU	IoU	IoU	IoU	IoU
background	90.0	94.5	91.6	92.7	**94.9**	dining table	33.0	44.1	45.9	**50.6**	43.0
aero body	61.9	68.1	68.2	61.9	**69.5**	dog head	60.5	63.1	63.8	64.0	**64.9**
aero stern	53.2	59.5	54.2	57.4	**61.2**	dog reye	50.1	50.2	54.1	54.7	**58.5**
aero rwing	28.9	38.3	33.1	34.3	**40.6**	dog rear	54.0	58.0	57.2	56.8	**60.6**
aero engine	24.7	27.0	26.5	27.2	**29.3**	dog nose	63.5	68.2	66.3	66.0	**70.1**
aero wheel	40.9	51.3	44.5	**51.5**	51.3	dog torso	58.4	61.0	62.3	**63.2**	62.2
bike fwheel	78.4	79.1	75.3	80.2	**80.5**	dog neck	27.1	27.8	26.2	28.1	**30.7**
bike saddle	34.1	36.0	31.0	38.0	**42.6**	dog rfleg	39.2	43.1	42.4	43.7	**44.9**
bike handlebar	23.3	22.1	20.6	22.4	**33.6**	dog rfpaw	39.4	44.1	44.2	43.7	**46.4**
bike chainwhell	42.3	44.5	36.5	44.1	**51.1**	dog tail	24.7	35.8	34.9	30.8	**40.1**
birds head	51.5	67.9	66.4	65.3	**68.8**	dog muzzle	65.1	68.5	69.4	68.9	**71.1**
birds beak	40.4	51.9	47.1	44.3	**58.3**	horse head	54.4	64.6	57.1	55.9	**68.5**
birds torso	61.7	62.7	65.2	64.8	**65.3**	horse rear	49.7	56.1	51.1	52.2	**60.3**
birds neck	27.5	38.1	**39.1**	28.4	36.1	horse muzzle	61.3	69.4	65.2	62.9	**72.3**
birds rwing	35.9	40.1	39.3	37.2	**41.3**	horse torso	56.7	62.2	59.5	60.7	**65.1**
birds rleg	23.5	26.0	26.5	23.8	**27.8**	horse neck	42.1	53.3	49.6	47.2	**55.2**
birds rfoot	13.9	13.8	11.6	17.7	**18.3**	horse rfuleg	54.1	60.1	57.0	56.4	**62.0**
birds tail	28.1	32.2	33.0	32.5	**34.7**	horse tail	48.1	53.4	47.6	51.4	**56.6**
boat	53.7	59.5	61.4	**69.2**	61.1	horse rfho	24.1	17.2	12.9	**25.3**	21.9
bottle cap	30.4	31.9	26.2	33.4	**35.8**	mbike fwheel	69.6	72.0	69.3	**73.6**	73.3
bottle body	63.7	67.1	71.5	**78.7**	68.3	mbike hbar	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
bus rightside	70.8	74.8	73.0	75.7	**77.6**	mbike saddle	0.0	0.0	0.0	**0.8**	0.0
bus roofside	7.5	13.9	0.3	13.5	**15.4**	mbike hlight	25.8	23.7	10.6	**28.5**	28.1
bus mirror	2.1	8.6	0.3	6.6	**15.4**	person head	68.2	72.8	69.7	69.3	**74.1**
bus fliplate	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	person reye	35.1	45.2	41.3	38.7	**47.9**
bus door	40.1	43.2	37.2	38.1	**49.1**	person rear	37.4	48.8	41.9	41.4	**52.9**
bus wheel	54.8	49.1	53.1	56.7	**57.6**	person nose	53.0	57.8	54.3	56.7	**62.2**
bus headlight	25.6	27.2	19.9	30.4	**35.7**	person mouth	48.9	54.1	49.5	51.3	**56.3**
bus window	71.8	75.2	73.5	74.6	**78.2**	person hair	70.8	73.2	72.3	71.8	**74.9**
car rightside	64.0	68.8	67.9	70.5	**72.4**	person torso	63.4	67.6	64.3	65.2	**69.9**
car roofside	21.0	15.8	16.1	**22.3**	19.3	person neck	49.7	53.2	50.9	51.2	**55.1**
car fliplate	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	person ruarm	54.7	61.1	55.7	57.4	**63.8**
car door	41.4	45.1	39.6	42.3	**49.6**	person rhand	43.0	48.9	47.4	44.1	**52.2**
car wheel	65.8	67.8	64.0	70.2	**71.7**	person ruleg	50.8	55.1	52.3	53.0	**57.0**
car headlight	42.9	51.1	49.4	46.4	**57.7**	person rfoot	29.8	31.8	28.9	31.3	**35.2**
car window	61.0	68.8	66.5	65.0	**71.9**	pplant pot	43.6	52.8	50.6	56.0	**56.3**
cat head	73.9	76.7	75.6	77.5	**77.7**	pplant plant	42.9	55.2	55.5	56.6	**58.1**
cat reye	58.8	57.1	62.0	62.8	**67.3**	sheep head	45.6	51.2	47.0	**54.0**	52.9
cat rear	65.5	67.7	66.8	67.1	**70.7**	sheep rear	43.2	50.6	47.7	45.3	**54.1**
cat nose	40.3	39.2	41.2	46.3	**46.9**	sheep muzzle	58.2	62.6	61.1	64.9	**65.1**
cat torso	64.2	67.0	66.8	**68.7**	67.9	sheep rhorn	3.0	**46.9**	0.0	5.4	44.4
cat neck	22.8	23.9	19.8	**24.4**	24.0	sheep torso	62.6	65.0	66.4	**68.8**	68.5
cat rfleg	36.5	39.6	38.5	39.1	**41.3**	sheep neck	26.9	**34.5**	25.3	30.3	33.6
cat rfpaw	40.6	42.5	**43.4**	41.7	43.0	sheep rfuleg	8.6	20.6	17.4	11.7	**21.1**
cat tail	40.2	**47.9**	42.6	45.8	47.0	sheep tail	6.7	9.5	1.1	9.1	**15.9**
chair	35.4	37.3	34.1	**49.1**	38.0	sofa	39.2	**47.4**	44.5	53.9	47.2
cow head	51.2	**66.1**	58.2	63.8	66.0	train head	5.3	4.7	**5.6**	4.5	**5.6**
cow rear	51.2	**63.9**	53.0	60.0	61.7	train hrightside	61.9	63.9	63.5	60.8	**64.0**
cow muzzle	61.2	71.9	67.2	**74.9**	73.9	train hroofside	23.0	**22.6**	13.7	21.1	22.0
cow rhorn	28.8	44.7	10.1	44.0	**57.6**	train headlight	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
cow torso	63.4	72.9	69.9	73.2	**75.1**	train coach	28.6	35.2	**42.0**	31.4	36.9
cow neck	9.5	19.9	7.3	20.3	**26.1**	train crightside	15.6	16.2	**19.0**	14.9	18.1
cow rfuleg	46.5	53.8	49.7	54.8	**57.8**	train croofside	10.8	**20.2**	1.0	18.1	15.1
cow tail	6.5	13.6	0.1	13.6	**17.6**	tv screen	60.8	69.7	66.3	70.7	**73.1**

Figure 3. Segmentation results on PASCAL-Part-108 dataset. Our model shows overall better segmentation results with more accurate boundaries ans less missing parts.