GMNet: Graph Matching Network for Large Scale Part Semantic Segmentation in the Wild
Supplementary Material

Umberto Michieli[0000−0003−2666−4342], Edoardo Borsato, Luca Rossi, and Pietro Zanuttigh[0000−0002−9502−2389]

Department of Information Engineering, University of Padova, Padova, Italy
{michieli, borsatoedo, rossiluc, zanuttigh}@dei.unipd.it

In this document we show some further experimental results. In particular we report the Intersection over Union (IoU) and Pixel Accuracy (PA) for each per-part-class and some averaged metrics as mean IoU (mIoU), mean PA (mPA) and mean Class Accuracy (mCA) [2]. The results are reported for both the Pascal-Part-58 and the Pascal-Part-108 datasets. Finally, some additional visual results on both datasets are presented.

1 Additional Results on Pascal-Part-58

We start by analyzing the per-part-class IoU and PA on the Pascal-Part-58 dataset. The results are shown in Table 1, where it is possible to see that the proposed method (GMNet) outperforms the baseline [1] approach on almost every part both considering the per-part-IoU and the per-part-PA. With respect to BSANet [3], GMNet can produce clearly higher results on 15 objects out of 21 (such as *bottle*, *bus*, *dog*, *sheep*,...) and can produce comparable results on 2 objects (i.e., on *car* ad *cat*).

We can further verify the ranking of the compared methods analyzing the average metrics reported in Table 2. Here, we can appreciate how GMNet is able to outperform both the baseline and BSANet robustly on all the most widely used metrics for semantic segmentation.

Then, we proceed to analyze some additional qualitative results as reported in Figure 1. The effects of the two main components of our work, namely the object-level semantic embedding network $\mathcal{S}$ and the graph matching module, are clearly visible in the images. The effect of the semantic embedding network is evident in the last 5 rows, where object-level conditioning helps the part-level decoder to accurately segment and label the parts. For instance, in the last row both the baseline and BSANet mislead the dog's parts with cat's parts, while GMNet is able to avoid this error. In row 6, BSANet confuses cow's parts with sheep's parts. In row 7, the baseline confuses sheep's parts with cat's ones and BSANet with dog's parts. GMNet is able to correctly deal with these situations thanks to the object-level guidance.

Table 1. Per-part IoU and PA on the Pascal-Part-58 dataset.

Parts Name	Baseline		BSANet		GMNet		Parts Name	Baseline		BSANet		GMNet	
	IoU	PA	IoU	PA	IoU	PA		IoU	PA	IoU	PA	IoU	PA
background	91.1	96.3	91.6	96.7	**92.7**	**96.9**	cow tail	0.0	0.0	7.9	8.1	**8.1**	**8.4**
aeroplane body	66.6	79.8	**70.0**	**81.4**	69.6	81.2	cow leg	46.1	62.3	53.4	**67.5**	**53.5**	67.2
aeroplane engine	25.7	31.4	**29.1**	**33.8**	25.7	31.2	cow torso	69.9	83.5	73.5	85.9	**77.1**	**87.8**
aeroplane wing	33.5	48.2	**38.3**	**49.1**	34.2	46.4	dining table	43.0	55.4	43.7	54.8	**51.3**	**62.6**
aeroplane stern	57.1	68.2	**59.2**	**72.5**	57.2	70.8	dog head	78.7	88.3	82.5	91.4	**85.0**	**92.7**
aeroplane wheel	45.4	53.3	**53.2**	**62.5**	46.8	53.3	dog leg	48.1	59.9	53.8	63.0	**53.8**	**64.8**
bike wheel	78.0	88.1	78.0	**88.6**	**81.3**	88.5	dog tail	27.1	39.4	31.3	38.0	**31.4**	**41.5**
bike body	48.4	61.2	**53.4**	**68.4**	51.5	64.2	dog torso	63.7	76.8	65.7	79.7	**68.0**	**81.2**
bird head	64.6	72.7	**74.0**	**80.2**	71.1	79.3	horse head	74.7	81.7	**76.6**	**83.3**	73.9	80.5
bird wing	35.1	45.5	**39.7**	**53.2**	38.6	52.9	horse tail	47.0	60.4	**51.0**	59.9	50.4	**62.2**
bird leg	29.3	37.6	**34.8**	**42.6**	28.7	35.4	horse leg	55.9	70.9	**61.6**	**75.8**	59.3	72.9
bird torso	66.9	83.1	**70.9**	**84.4**	69.5	83.1	horse torso	70.3	84.2	**74.9**	86.6	73.9	**87.4**
boat	54.4	64.8	60.2	69.6	**70.0**	**78.5**	mbike wheel	70.9	82.5	71.6	82.1	**73.5**	**84.0**
bottle cap	30.7	35.4	29.8	35.0	**33.9**	**42.5**	mbike body	65.1	80.9	71.5	87.7	**74.3**	**87.8**
bottle body	68.8	78.5	68.6	74.8	**77.6**	**86.1**	person head	83.5	91.6	**85.0**	**92.3**	84.7	91.8
bus window	72.7	83.7	74.8	85.9	**75.4**	**86.1**	person torso	65.9	80.6	**68.2**	**82.7**	67.0	82.3
bus wheel	55.3	66.3	57.1	70.1	**58.1**	**72.1**	person larm	46.9	60.0	**52.0**	**65.6**	48.6	62.8
bus body	74.8	88.2	78.3	88.7	**79.9**	**89.8**	person uarm	51.5	65.8	**54.4**	**68.2**	52.4	66.9
car window	62.6	73.9	**68.1**	**78.2**	64.8	77.5	person lleg	38.6	51.5	**43.5**	**54.6**	40.2	51.5
car wheel	64.8	78.1	68.5	79.7	**70.3**	**79.8**	person uleg	43.8	60.0	**47.4**	**63.5**	44.5	59.9
car light	46.2	54.3	**53.7**	**61.7**	48.4	56.0	pplant pot	45.3	61.0	53.5	64.8	**56.0**	**69.1**
car plate	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	pplant plant	52.4	62.1	**56.6**	65.8	56.4	**66.4**
car body	72.1	86.4	77.0	**88.4**	**77.6**	88.2	sheep head	60.9	69.3	65.4	71.3	**70.8**	**79.0**
cat head	80.2	90.4	83.7	**92.3**	**83.8**	91.6	sheep leg	8.6	11.1	11.7	16.5	**14.3**	**20.2**
cat leg	48.6	**61.2**	**50.1**	58.6	49.4	59.1	sheep torso	68.3	84.4	71.6	86.1	**75.6**	**88.7**
cat tail	40.2	51.3	**48.8**	55.6	46.0	**56.7**	sofa	43.2	58.8	43.1	57.4	**56.1**	**65.0**
cat torso	70.3	85.7	72.6	**88.0**	**73.8**	87.6	train	79.6	86.1	82.2	90.2	**85.0**	**92.0**
chair	35.4	43.3	36.5	42.7	**51.4**	**63.9**	tv screen	69.5	76.0	73.1	78.6	**77.0**	**84.3**
cow head	74.3	85.6	76.4	86.0	**80.7**	**87.8**	tv frame	45.9	56.9	49.8	60.9	**54.1**	**67.4**

Table 2. Comparison in terms of mIoU, mCA and mPA on Pascal-Part-58.

Method	mIoU	mPA	mCA
Baseline [1]	54.45	89.86	65.42
BSANet [3]	58.15	90.76	68.12
GMNet	**59.04**	**91.55**	**69.22**

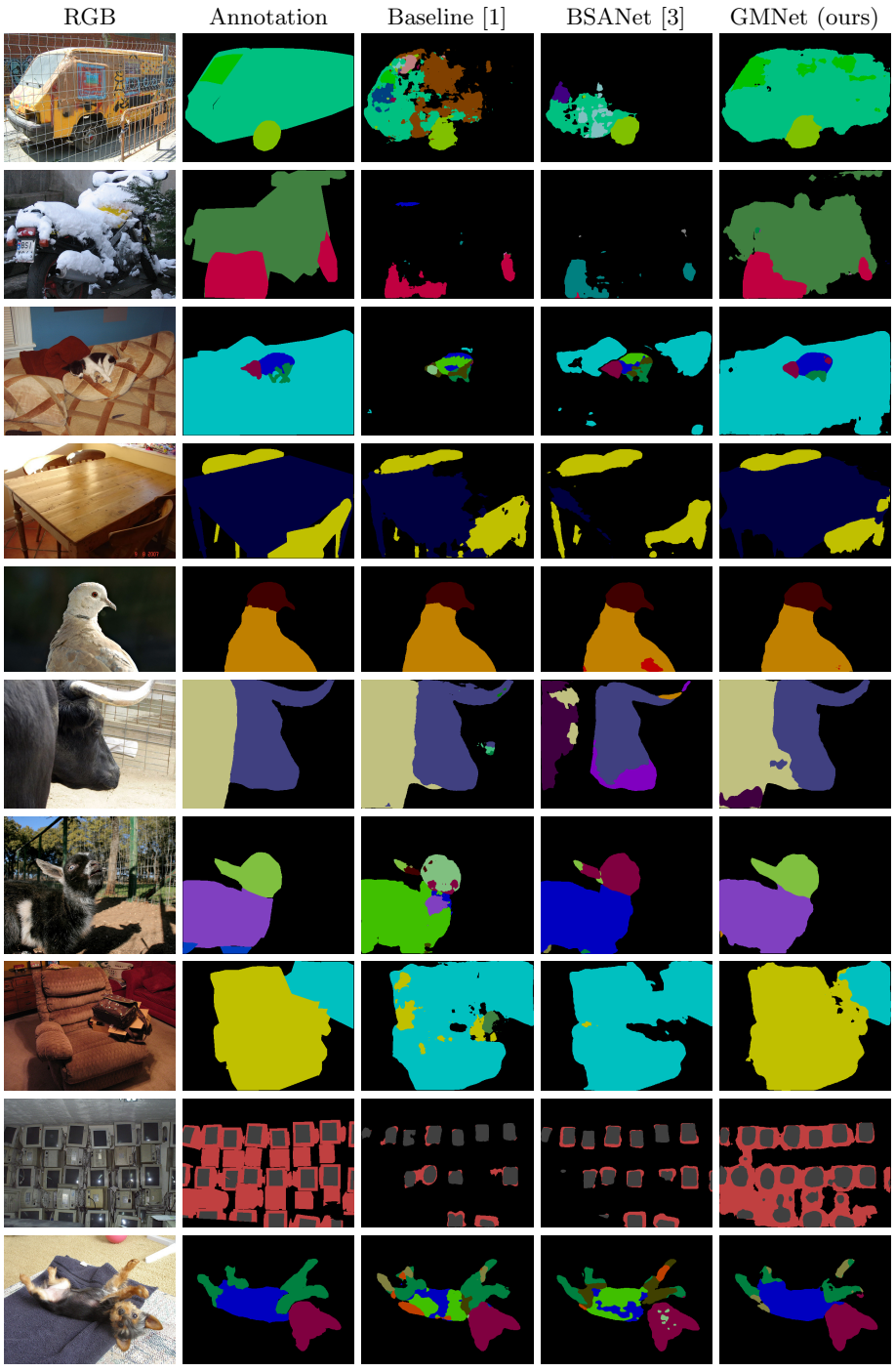

Fig. 1. Qualitative results on some sample scenes on the Pascal-Part-58 dataset (*best viewed in colors*).

The effect of the graph matching module is more appreciable on small parts. For example, we can verify its efficacy in the third row and in the last row. In row 3, both the baseline and BSANet mislead the dog's parts with cat's ones and also their localization is highly imprecise. From one hand, the semantic embedding network corrects the first issue, while the second (i.e., bad localization) is addressed by graph matching. In the last row, graph matching between different reciprocal spatial relationship among parts helps to correctly place the dog's parts.

Moreover, in very challenging images (such as rows 1 to 4), where both the baseline and BSANet partially or completely miss some classes, our method generate superior quality segmentation maps. For instance, in row 1 a vehicle behind a metal grid is being correctly identified and quite well localized in all its parts thanks to the semantic embedding module and to graph matching. The combination of the two modules is also helpful in row 2, where a motorbike covered with snow is being well recognized by our framework. In row 3 we identify the sofa and in row 4 the table, with higher accuracy than the compared methods.

2 Additional Results on Pascal-Part-108

In this section we present some additional results for the Pascal-Part-108 dataset. The per-part-IoU and per-part-PA are reported in Table 3, where we can notice that the gap between the proposed framework and the compared methods is significantly larger than for the Pascal-Part-58 dataset. GMNet achieves higher accuracy than the competitors on almost all the parts. In particular, our framework is able to outperform BSANet [3] in 19 out of 21 object-level classes both with many parts within them (such as *aeroplane*, *bus*, *cat*, *dog*, *person*, *sheep*,...) and with no or few parts within them (such as *boat*, *bottle*, *chair*, *sofa*, *tv*,..).

The mean accuracy results are shown in Table 4 where we can verify that our method clearly outperforms both the baseline [1] and BSANet [3] on all the most popular metrics used to evaluate semantic segmentation architectures. Hence, we prove the robustness of our framework to different evaluation criteria and to different datasets. Additionally, we argue that the proposed framework is able to scale well to even larger sets of parts.

Then, in Figure 2 we report some additional qualitative results. The effect of the object-level semantic embedding network is particularly evident in the first 4 rows. In row 1, a challenging image is presented where both the baseline and BSANet are not able to correctly identify the table. In rows 2 and 3, GMNet generates cleaner segmentation maps exploiting object-level priors which help to disambiguate between cars and buses. In row 4, the baseline and BSANet predict cat's parts in spite of horse's parts which are partially identified by our method.

Table 3. Per-part IoU and PA on the Pascal-Part-108 dataset.

Parts Name	Baseline		BSANet		GMNet		Parts Name	Baseline		BSANet		GMNet	
	IoU	PA	IoU	PA	IoU	PA		IoU	PA	IoU	PA	IoU	PA
background	90.9	**97.2**	91.6	97.1	**92.7**	97.0	dining table	33.0	40.2	45.9	59.7	**50.6**	**62.3**
aero body	61.9	72.3	**68.2**	77.6	61.9	**82.6**	dog head	60.5	75.5	63.8	78.2	**64.0**	**78.9**
aero stern	53.2	68.4	54.2	65.3	**57.4**	**71.0**	dog reye	50.1	61.4	54.1	61.4	**54.7**	**64.7**
aero rwing	28.9	39.8	33.1	**46.5**	**34.3**	46.0	dog rear	54.0	69.4	**57.2**	73.4	56.8	**73.9**
aero engine	24.7	29.0	26.5	32.0	**27.2**	**32.6**	dog nose	63.5	75.0	**66.3**	74.3	66.0	**76.8**
aero wheel	40.9	46.8	44.5	49.6	**51.5**	**61.3**	dog torso	58.4	74.6	62.3	78.4	**63.2**	**79.1**
bike fwheel	78.4	85.7	75.3	86.7	**80.2**	**87.8**	dog neck	27.1	35.4	26.2	30.8	**28.1**	**35.5**
bike saddle	34.1	39.8	31.0	31.9	**38.0**	**43.2**	dog rfleg	39.2	50.6	42.4	53.5	**43.7**	**55.8**
bike handlebar	**23.3**	**26.1**	20.6	22.8	22.4	25.9	dog rfpaw	39.4	47.9	**44.2**	51.7	43.7	**52.9**
bike chainwheel	42.3	50.4	36.5	41.6	**44.1**	**57.0**	dog tail	24.7	37.8	**34.9**	**42.3**	30.8	41.4
birds head	51.5	61.3	**66.4**	**78.0**	65.3	77.7	dog muzzle	65.1	76.1	**69.4**	**82.3**	68.9	80.4
birds beak	40.4	49.5	**47.1**	**54.6**	44.3	54.0	horse head	54.4	67.0	**57.1**	**68.9**	55.9	68.3
birds torso	61.7	77.9	**65.2**	79.4	64.8	**82.6**	horse rear	49.7	58.1	51.1	56.5	**52.2**	**65.6**
birds neck	27.5	32.2	**39.1**	**50.1**	28.4	35.7	horse muzzle	61.3	68.7	**65.2**	**74.0**	62.9	69.5
birds rwing	35.9	50.4	**39.3**	**53.7**	37.2	50.1	horse torso	56.7	75.9	59.5	75.9	**60.7**	**84.3**
birds rleg	23.5	28.6	**26.5**	32.2	23.8	**32.8**	horse neck	42.1	51.3	**49.6**	**64.8**	47.2	55.8
birds rfoot	13.9	16.3	11.6	12.7	**17.7**	**22.5**	horse rfuleg	54.1	68.5	**57.0**	**71.8**	56.4	70.9
birds tail	28.1	39.2	**33.0**	44.1	32.5	**46.1**	horse tail	48.1	63.5	47.6	54.5	**51.4**	**64.4**
boat	53.7	60.3	61.4	71.5	**69.2**	**77.8**	horse rfho	24.1	31.4	12.9	13.7	**25.3**	**32.7**
bottle cap	30.4	35.0	26.2	30.0	**33.4**	**40.0**	mbike fwheel	69.6	78.9	69.3	80.4	**73.6**	**83.3**
bottle body	63.7	69.5	71.5	78.3	**78.7**	**88.3**	mbike hbar	**0.0**	**0.0**	0.0	0.0	0.0	0.0
bus rightside	70.8	85.3	73.0	83.7	**75.7**	**88.4**	mbike saddle	0.0	0.0	0.0	0.0	**0.8**	**0.8**
bus roofside	7.5	7.7	0.3	0.3	**13.5**	**14.4**	mbike hlight	25.8	**32.8**	10.6	11.2	**28.5**	32.4
bus mirror	2.1	2.2	0.3	0.3	**6.6**	**7.6**	person head	68.2	81.9	**69.7**	82.2	69.3	**82.7**
bus fliplate	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	person reye	35.1	39.3	**41.3**	**46.3**	38.7	43.9
bus door	**40.1**	51.2	37.2	**53.2**	38.1	47.3	person rear	37.4	46.0	**41.9**	49.4	41.4	**51.5**
bus wheel	**54.8**	65.5	53.1	63.9	56.7	**69.4**	person nose	53.0	62.1	54.3	63.1	**56.7**	**67.5**
bus headlight	**25.6**	28.3	19.9	20.8	30.4	**34.2**	person mouth	48.9	56.9	49.5	54.9	**51.3**	**60.8**
bus window	71.8	85.2	73.5	86.4	**74.6**	**87.4**	person hair	70.8	83.3	**72.3**	**85.9**	71.8	83.9
car rightside	64.0	78.0	67.9	81.2	**70.5**	**84.5**	person torso	63.4	79.1	64.3	78.3	**65.2**	**80.9**
car roofside	21.0	25.4	16.1	17.6	**22.3**	**26.6**	person neck	49.7	63.8	50.9	65.1	**51.2**	**65.3**
car fliplate	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	person ruarm	54.7	68.6	55.7	70.2	**57.4**	**71.3**
car door	41.4	52.5	39.6	49.0	**42.3**	**53.5**	person rhand	43.0	55.4	**47.4**	**57.6**	44.1	56.8
car wheel	65.8	74.5	64.0	76.6	**70.2**	**80.0**	person ruleg	50.8	66.0	52.3	67.1	**53.0**	**67.9**
car headlight	42.9	48.4	**49.4**	**59.7**	46.4	54.4	person rfoot	29.8	38.9	28.9	32.4	**31.3**	**39.8**
car window	61.0	75.5	**66.5**	**82.4**	65.0	79.0	pplant pot	43.6	54.5	50.6	58.9	**56.0**	**69.0**
cat head	73.9	87.3	75.6	88.5	**77.5**	**88.5**	pplant plant	42.9	48.8	55.5	**68.7**	**56.6**	66.6
cat reye	58.8	69.0	62.0	71.1	**62.8**	**71.8**	sheep head	45.6	56.9	47.0	58.0	**54.0**	**66.9**
cat rear	65.5	77.7	66.8	77.1	**67.1**	**78.8**	sheep rear	43.2	53.0	**47.7**	56.6	45.3	**58.2**
cat nose	40.3	49.1	41.2	45.8	**46.3**	**56.2**	sheep muzzle	58.2	67.0	61.1	72.4	**64.9**	**74.7**
cat torso	64.2	81.4	66.8	84.2	**68.7**	**86.0**	sheep rhorn	3.0	3.6	0.0	0.0	**5.4**	**6.6**
cat neck	22.8	33.8	19.8	25.0	24.4	**34.1**	sheep torso	62.6	78.0	66.4	83.6	**68.8**	**86.3**
cat rfleg	36.5	48.5	38.5	49.2	**39.1**	**50.0**	sheep neck	26.9	38.1	25.3	**41.2**	**30.3**	41.0
cat rfpaw	40.6	50.2	**43.4**	**51.5**	41.7	50.7	sheep rfuleg	8.6	10.6	**17.4**	**24.5**	11.7	14.7
cat tail	40.2	52.2	42.6	49.5	**45.8**	**57.0**	sheep tail	6.7	7.4	1.1	1.1	**9.1**	**11.5**
chair	35.4	42.3	34.1	38.4	**49.1**	**60.4**	sofa	39.2	50.7	44.5	56.9	**53.9**	**66.1**
cow head	51.2	65.5	58.2	74.2	**63.8**	**74.9**	train head	5.3	**6.4**	**5.6**	6.4	4.5	5.3
cow rear	51.2	68.5	53.0	72.9	**60.0**	**75.1**	train hrightside	61.9	77.3	**63.5**	**84.0**	60.8	83.1
cow muzzle	61.2	77.6	67.2	81.9	**74.9**	**86.7**	train hroofside	**23.0**	**28.0**	13.7	17.0	21.1	26.3
cow rhorn	28.8	35.0	10.1	10.2	**44.0**	**50.6**	train headlight	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**	**0.0**
cow torso	63.4	78.6	69.9	85.8	**73.2**	**87.2**	train coach	28.6	33.6	**42.0**	**47.8**	31.4	37.9
cow neck	9.5	12.7	7.3	7.9	**20.3**	**25.9**	train crightside	15.6	24.5	**19.0**	30.6	14.9	**33.8**
cow rfuleg	46.5	60.0	49.7	61.4	**54.8**	**70.7**	train croofside	10.8	11.9	1.0	1.0	**18.1**	**22.6**
cow tail	6.5	7.3	0.1	0.1	**13.6**	**14.9**	tv screen	60.8	71.3	66.3	79.5	**70.7**	**82.9**

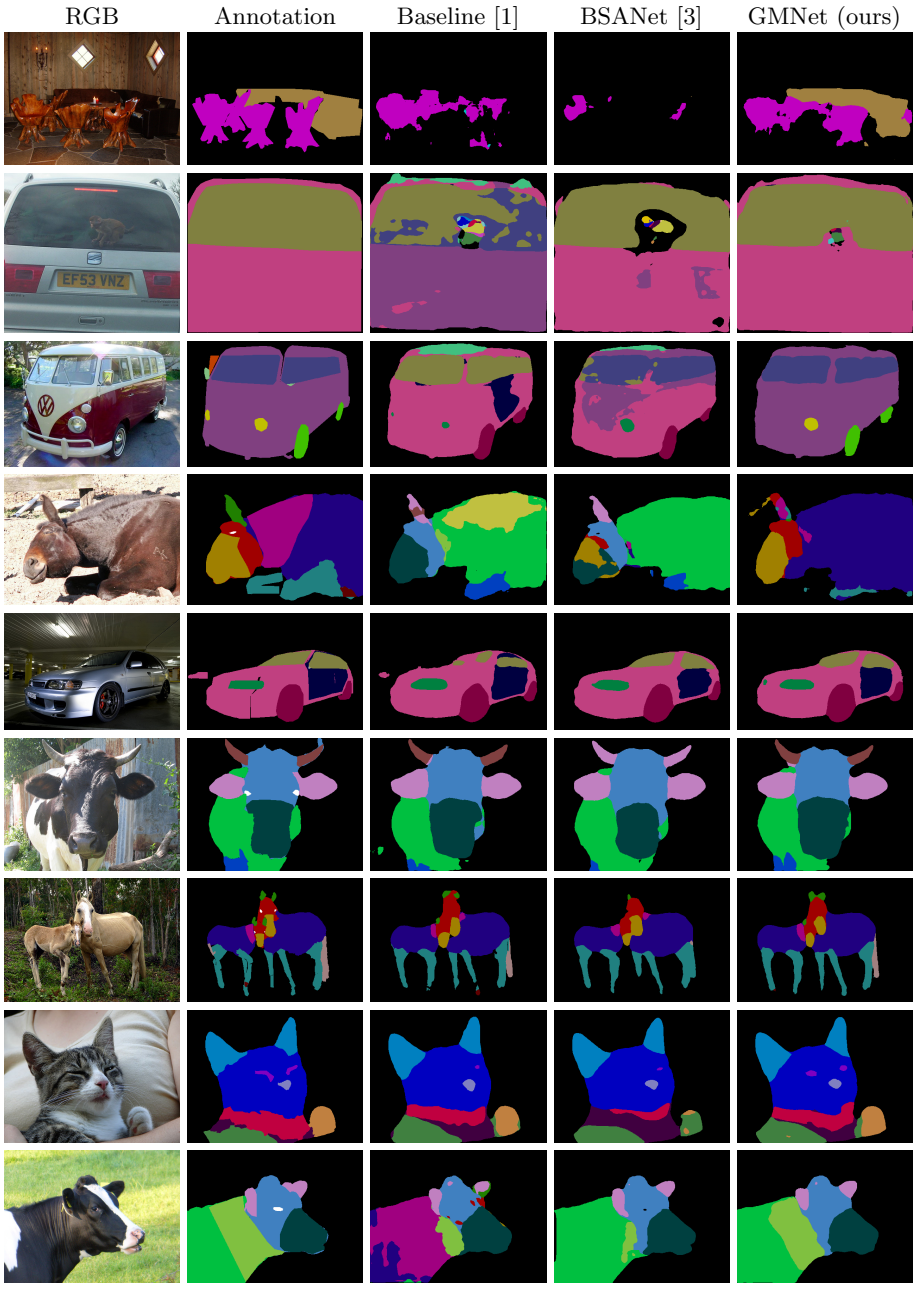

Fig. 2. Qualitative results on sample scenes on the Pascal-Part-108 dataset (*best viewed in colors*).

Table 4. Comparison in terms of mIoU, mCA and mPA on Pascal-Part-108.

Method	mIoU	mPA	mCA
Baseline [1]	41.36	88.57	50.51
BSANet [3]	42.95	89.52	51.71
GMNet	**45.80**	**90.32**	**55.68**

The graph matching module is much more effective on this dataset because contains many small-sized parts. We can verify this from the sixth to the last row. In row 6, the *cow_horns* and *cow_body* are badly localized and labelled both by the baseline and by BSANet. However, the graph matching component on the reciprocal spatial relationship between these parts and the others guides the network to properly localize and label such parts. In row 7 our framework is able to well localize horse's parts and especially the challenging *horse_tail* part. In the second-last row, GMNet correctly identifies difficult cat's parts such as *cat_eyes* and *cat_paws* thanks to the graph matching module. In the last row, the semantic embedding module allows our method to identify the cow and, at the same time, the graph matching module allows to correctly localize the spatial relations among all the parts.

References

1. Chen, L.C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017)
2. Csurka, G., Larlus, D., Perronnin, F., Meylan, F.: What is a good evaluation measure for semantic segmentation? In: Proceedings of British Machine Vision Conference (BMVC). vol. 27, p. 2013 (2013)
3. Zhao, Y., Li, J., Zhang, Y., Tian, Y.: Multi-class part parsing with joint boundary-semantic awareness. In: Proceedings of International Conference on Computer Vision (ICCV). pp. 9177–9186 (2019)