Bayesian Joint Modelling for Object Localisation in Weakly Labelled Images Supplementary Material

Zhiyuan Shi, Timothy M. Hospedales, Tao Xiang

1 Per-Class Object Localisation Results

In Sec. 8.2 of the main paper, we evaluated object localisation from weakly labeled data. Table 2 (main paper) compared our proposed methods (Our-Sampling and Our Gaussian) with the state-of-the-art competitors [1], [2], [3], [4] on the three variants of the PASCAL VOC 2007 dataset: $VOC07-6\times2$, $VOC07-14$ and $VOC07-20$. Due to the space limitation, only the results averaged over all the classes in each dataset were shown. In this supplementary section, we provide more detailed per-class object localisation results in for the three VOC variants in Tables 1, 2, and 3 respectively. Note that few previous studies report their per-class results. Those reported per-class results are included in the tables for comparison.

As mentioned in the main paper, refining the localisation by a strong detector [5], [1] brings overall improvements on the localisation accuracy. However, the improvement can be very limited or even negative for some classes when the initialisation performance is poor. For instance, Table 3 shows that for the challenging *bottle* class, the initial weakly supervised object localisation accuracy is weak (6.3% for Our-Sampling+prior). After refinement using a strong detector, the localisation accuracy becomes even worse (4.2% for Our-Sampling+prior). This is understandable: with poor localisation, only a poor detector will be learned, which will not help refine the localisation.

	Initialisation			Refined by detector		
	Our-Sampling+prior	Our-Gaussian+prior	[4]	Our-Sampling+prior	Our-Gaussian+prior	[1]
aeroplane left	58.7	**55.3**	39.1	**72.0**	72.0	58
aeroplane right	**64.2**	72.6	50.0	71.0	**71.9**	59
bicycle left	29.0	**34.4**	28.4	**60.2**	58.8	46
bicycle right	**36.3**	38.0	30.6	**48.5**	48.3	40
boat left	20.7	**27.8**	15.1	44.4	**44.2**	9
boat right	**27.8**	19.5	20.7	46.1	**48.2**	16
bus left	**38.1**	**32.8**	31.0	**49.7**	46.0	38
bus right	52.8	**47.3**	35.1	61.7	54.2	**74**
horse left	**71.4**	**67.1**	48.5	89.8	**90.1**	58
horse right	69.6	**77.7**	45.2	85.6	**88.5**	52
Motorbike left	**68.4**	68.9	46.3	79.3	**83.8**	67
motorbike right	77.9	**80.3**	55.3	82.8	**94.7**	76
Average	51.2	**51.8**	37.1	65.9	**66.7**	50

TABLE 1: Per-class localisation accuracy for the *VOC07-6×2* dataset

2 Per-Class Object Detection Results

In the main paper, we further evaluated object detection performance on test data given detectors trained from weakly supervised images. Table 3 (main paper) reports the mAP of detection performance on both VOC-6×2 and VOC-20. Here we provide the per-class AP for VOC-6×2 and VOC-20 in Table 4 and Table 5 respectively. We can see that for some classes (e.g. bicycle right, motorbike left in Table 4; Bicycle, bus, car, motorbike, train in Table 5) our approach achieves comparable performance to the fully supervised detector. This is a very encouraging result. It shows that with our framework, automatic localisation can replace manual location annotation to train detectors for these classes. However, for those with very low localisation accuracy (e.g. bottle and pottedplant), the weakly supervised detector fails completely.

	Initialisation		Refined by detector	
	Our-Sampling+prior	Our-Gaussian+prior	Our-Sampling+prior	Our-Gaussian+prior
aeroplane	**58.8**	54.1	**61.0**	57.2
bicycle	**32.8**	27.7	**34.8**	27.8
boat	**25.8**	23.4	**28.5**	24.9
bottle	**06.8**	**05.7**	07.3	**06.4**
bus	42.7	**45.3**	47.5	**48.3**
chair	**06.8**	**06.0**	**09.9**	07.6
diningtable	**33.6**	30.5	**36.6**	31.9
horse	**57.8**	48.9	**58.1**	54.3
motorbike	**59.5**	59.6	61.2	**60.5**
person	**28.6**	24.1	**30.6**	28.2
pottedplant	**14.2**	10.2	**14.8**	11.0
sofa	**37.4**	36.8	**39.1**	39.2
train	**56.8**	56.0	58.1	**57.0**
tvmonitor	06.5	**07.4**	**08.4**	08.3
Average	**33.4**	31.1	**35.4**	33.0

TABLE 2: Per-class localisation accuracy for the *VOC07-14* dataset

	Initialisation					Refined by detector		
	Our-Sampling+prior	Our-Gaussian+prior	[4]	[6]	[3]	Our-Sampling+prior	Our-Gaussian+prior	[6]
aeroplane	**62.0**	53.1	38.7	45.4	54.7	**68.3**	63.0	42.4
bicycle	**33.8**	33.0	22.2	20.6	22.7	**56.8**	54.9	46.5
bird	32.9	20.9	27.6	29.7	**33.7**	**37.5**	24.7	18.2
boat	**30.8**	26.2	21.0	12.2	24.5	**20.2**	17.5	08.8
bottle	06.3	08.0	**06.6**	04.1	04.6	04.2	05.1	**02.9**
bus	36.5	37.8	33.3	**37.1**	33.9	48.8	**53.7**	40.9
car	42.7	41.8	39.4	41.0	**42.5**	63.3	42.6	**73.2**
cat	**60.5**	53.5	46.0	53.4	57.0	**71.7**	60.6	44.8
chair	07.4	07.6	**08.1**	06.5	07.3	61.0	04.6	**05.4**
cow	39.0	34.7	34.8	31.9	**39.1**	**33.7**	31.1	30.5
diningtable	30.4	31.7	**31.5**	20.5	24.1	16.2	**26.4**	19.0
dog	**50.1**	43.3	38.0	40.9	43.3	**61.5**	56.5	34.0
horse	**57.7**	51.9	37.6	37.3	41.3	55.5	**54.7**	48.8
motorbike	56.9	**56.8**	43.3	46.5	51.5	65.4	67.5	**65.3**
person	**30.3**	26.2	23.0	22.3	25.3	**21.2**	17.5	08.2
pottedplant	12.1	14.1	11.4	10.2	**13.3**	03.6	07.3	**09.4**
sheep	**35.6**	32.8	28.1	27.1	28.0	24.4	**25.8**	16.7
sofa	30.6	32.8	**34.5**	32.3	29.5	**37.3**	35.3	32.3
train	**58.1**	56.8	43.7	49.0	54.6	**63.5**	62.1	54.8
tvmonitor	08.2	07.0	10.5	09.8	**11.8**	**07.8**	05.7	05.5
Average	**36.1**	33.5	29.0	28.9	32.1	**38.3**	35.8	30.4

TABLE 3: Per-class localisation accuracy for the *VOC07-20* dataset

REFERENCES

[1] T. Deselaers, B. Alexe, and V. Ferrari, "Weakly supervised localization and learning with generic knowledge," *IJCV*, 2012.

[2] M. Pandey and S. Lazebnik, "Scene recognition and weakly supervised object localization with deformable part-based models," in *ICCV*, 2011.

[3] Z. Shi, P. Siva, and T. Xiang, "Transfer learning by ranking for weakly supervised object annotation," in *BMVC*, 2012.

[4] P. Siva, C. Russell, and T. Xiang, "In defence of negative mining for annotating weakly labelled data," in *ECCV*, 2012.

[5] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *TPAMI*, 2010.

[6] P. Siva and T. Xiang., "Weakly supervised object detector learning with model drift detection," in *ICCV*, 2011.

	[2]	[1]	Our	FSL
aeroplane left	7.5	5	12.4	23
aeroplane right	21.1	18	26.2	32
bicycle left	38.5	49	48.4	59
bicycle right	44.8	62	63.6	64
boat left	0.3	0	0.2	0
boat right	0.5	0	0.5	1
bus left	0	0	0.8	21
bus right	3	16	18.2	20
horse left	45.9	29	39.3	45
horse right	17.3	14	28.5	39
Motorbike left	43.8	48	53.3	55
motorbike right	27.2	16	22.3	42
Average	20.8	21.4	26.1	33.4

TABLE 4: Per-class average precision for object detection on *VOC07-6×2* dataset

	[6]	Our	FSL
aeroplane	13.4	25.5	29.0
bicycle	44	50.0	53.6
bird	3.1	0.4	0.6
boat	3.1	9.0	13.4
bottle	0	0	26.2
bus	31.2	35.6	39.4
car	43.9	45.6	46.4
cat	7.1	14.4	16.1
chair	0.1	1.1	16.3
cow	9.3	13.4	16.5
diningtable	9.9	8.2	24.5
dog	1.5	3.2	5.0
horse	29.4	38.4	43.6
motorbike	38.3	37.3	37.8
person	4.6	16.5	35.0
pottedplant	0.1	0	8.8
sheep	0.4	2	17.3
sofa	3.8	10.0	21.6
train	34.2	34.0	34.0
tvmonitor	0	0.2	39.0
Average	13.9	17.2	26.3

TABLE 5: Per-class average precision for object detection on the *VOC07-20* dataset