Bayesian Joint Topic Modelling for Weakly Supervised Object Localisation (Supplementary Material)

Zhiyuan Shi, Timothy M. Hospedales, Tao Xiang
Queen Mary, University of London, London E1 4NS, UK
{zhiyuan.shi,tmh,txiang}@eecs.qmul.ac.uk

1. Class-level Object Localisation Results

In Table 1 (Sec. 4.1) of the submitted main paper, the proposed methods (Our-Sampling and Our Gaussian) are compared with state-of-the-art competitors [1, 3, 3, 4, 5] on the three variants of the PASCAL VOC 2007 datasets. Due to space limitation, only the results averaged over all the classes in each dataset was shown. In this supplementary material, we provide the more detailed per-class object localisation results in Tables 1, 2, and 3. Note that few of the related works report their per-class results. Only those reported are included in the three tables for comparison.

As mentioned in the main paper (Lines 528–530), refining the localisation by a strong detector [2, 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 (4.5% for Our-Sampling). After refinement using a strong detector, the localisation accuracy becomes even worse (1.3% for Our-Sampling). This is understandable: only a poor detector can be learned with poor localisation, which will not help the further refinement.

References

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

[2] P. Felzenszwalb, R. Girshick, D. McAllester, and D. Ramanan. Object detection with discriminatively trained part-based models. *TPAMI*, 2010. 1

[3] M. Pandey and S. Lazebnik. Scene recognition and

	Initialisation			Refined by detector		
	Our-Sampling	Our-Gaussian	[5]	Our-Sampling	Our-Gaussian	[1]
aeroplane left	57.3	**59.8**	39.1	**71.6**	71.2	58
aeroplane right	**69.2**	67.4	50.0	70.4	**70.8**	59
bicycle left	33.3	**37.8**	28.4	**59.6**	58.6	46
bicycle right	**38.7**	36.5	30.6	**48.7**	47.6	40
boat left	21.8	**25.6**	15.1	43.6	**43.9**	9
boat right	**28.6**	23.4	20.7	45.9	**47.8**	16
bus left	**32.0**	**32.0**	31.0	**49.7**	45.7	38
bus right	47.9	**49.6**	35.1	61.9	52.8	**74**
horse left	**69.7**	**69.7**	48.5	89.4	**90.2**	58
horse right	68.1	**69.7**	45.2	85.2	**88.5**	52
Motorbike left	**68.7**	67.0	46.3	78.6	**82.8**	67
motorbike right	74.5	**78.9**	55.3	82.2	**93.0**	76
Average	50.8	**51.5**	37.1	65.5	**66.1**	50

Table 1. Class-level localisation accuracy for the *VOC07-6×2* dataset

	Initialisation		Refined by detector	
	Our-Sampling	Our-Gaussian	Our-Sampling	Our-Gaussian
aeroplane	**57.5**	54.2	**58.8**	55.5
bicycle	**31.7**	26.4	**32.9**	27.1
boat	**24.9**	21.5	**26.1**	25.1
bottle	**05.3**	**05.3**	05.9	**06.8**
bus	41.6	**44.9**	45.7	**47.8**
chair	**05.6**	**05.6**	**07.5**	06.4
diningtable	**34.0**	29.0	**34.7**	30.3
horse	**56.0**	48.9	**57.6**	53.6
motorbike	**58.0**	57.7	59.9	**60.1**
person	**26.7**	24.5	**28.4**	27.8
pottedplant	**12.7**	10.6	**13.1**	12.7
sofa	**37.3**	36.2	**38.9**	37.9
train	**55.6**	54.8	56.2	**57.1**
tvmonitor	04.7	**06.6**	**06.7**	06.6
Average	**32.2**	30.5	**33.8**	32.5

Table 2. Class-level localisation accuracy for the *VOC07-14* dataset

	Initialisation					Refined by detector		
	Our-Sampling	Our-Gaussian	[5]	[6]	[4]	Our-Sampling	Our-Gaussian	[6]
aeroplane	**58.7**	51.4	38.7	45.4	54.7	**67.3**	61.1	42.4
bicycle	**30.9**	29.6	22.2	20.6	22.7	**54.4**	51.6	46.5
bird	30.9	20.8	27.6	29.7	**33.7**	**34.3**	21.2	18.2
boat	**28.5**	26.5	21.0	12.2	24.5	**17.8**	16.8	08.8
bottle	04.5	05.3	**06.6**	04.1	04.6	01.3	01.8	**02.9**
bus	34.4	36.6	33.3	**37.1**	33.9	46.6	**51.6**	40.9
car	40.4	38.3	39.4	41.0	**42.5**	60.7	40.6	**73.2**
cat	**59.1**	51.5	46.0	53.4	57.0	**68.9**	60.3	44.8
chair	04.9	04.5	**08.1**	06.5	07.3	02.5	01.6	**05.4**
cow	39.0	33.3	34.8	31.9	**39.1**	**32.4**	26.9	30.5
diningtable	28.0	30.0	**31.5**	20.5	24.1	16.2	**21.9**	19.0
dog	**47.7**	41.6	38.0	40.9	43.3	**58.9**	54.0	34.0
horse	**55.2**	48.4	37.6	37.3	41.3	51.5	**53.3**	48.8
motorbike	53.5	**54.3**	43.3	46.5	51.5	64.6	63.9	**65.3**
person	**28.5**	24.3	23.0	22.3	25.3	**18.2**	15.5	08.2
pottedplant	11.4	10.2	11.4	10.2	**13.3**	03.1	05.9	**09.4**
sheep	**35.2**	30.2	28.1	27.1	28.0	20.9	**21.8**	16.7
sofa	29.5	29.7	**34.5**	32.3	29.5	**34.7**	34.2	32.3
train	**56.0**	52.5	43.7	49.0	54.6	**63.4**	61.4	54.8
tvmonitor	05.9	04.7	10.5	09.8	**11.8**	**05.9**	03.2	05.5
Average	**34.1**	31.2	29.0	28.9	32.1	**36.2**	33.4	30.4

Table 3. Class-level localisation accuracy for the *VOC07-20* dataset

weakly supervised object localization with deformable part-based models. *ICCV*, 2011. 1

[4] Z. Shi, P. Siva, and T. Xiang. Transfer learning by ranking for weakly supervised object annotation. In *BMVC*, 2012. 1, 2

[5] P. Siva, C. Russell, and T. Xiang. In defence of negative mining for annotating weakly labelled data. In *ECCV*, 2012. 1, 2

[6] P. Siva and T. Xiang. Weakly supervised object detector learning with model drift detection. In *ICCV*, 2011. 2