Weakly-Supervised Image Annotation and Segmentation with Objects and Attributes Supplementary Material

Zhiyuan Shi, Yongxin Yang, Timothy M. Hospedales, Tao Xiang

1 Further Evaluation

	Free annotation		Segmentation	
	aPascal	ImageNet	LabelMe	aPascal
WS-SIBP	38.6	58.5	22.4	16.6
WS-SIBP+MRF1	38.8	58.6	43.1	22.8
WS-SIBP+MRF2	39.2	59.2	28.4	19.5
WS-SIBP+MRF2$^+$	39.7	60.3	32.6	20.1
WS-MRF-SIBP	40.1	60.7	46.2	27.3

TABLE 1: Evaluation of individual components of our model. We report AP@2 for free annotation on aPascal and ImageNet dataset. For segmentation, per-pixel results are reported for LabelMe and IOU for aPascal. For segmentation results on aPascal 8 attribute annotations are used.

Ablation study: To evaluate the impact of each component of our model, we perform experiments with several stripped-down versions of our model: a) **WS-SIBP**: The basic model without any MRF assumption. b) **WS-SIBP+MRF1**: This enables the spatial MRF to smooth neighbouring superpixels. c) **WS-SIBP+MRF2**: This enables the factorial MRF to model the correlation of factors, but using only the initial image label prior. d) **WS-SIBP+MRF2$^+$**: Same as c) except that we further update the correlation matrix with Eq. (12). e) **WS-MRF-SIBP**: Our full model with all components (b+d).

The results in Table 1 show that each component contributes to our good performance. As expected, the factorial MRF is more helpful (38.6→39.7, 58.5→60.3) for free annotation compared to the spatial MRF (38.6→38.8, 58.5→58.6). In contrast, the spatial MRF significantly improves the semantic segmentation performance (22.4→43.1, 16.6→22.8) compared to the factorial MRF (22.4→32.6, 16.6→20.1). This is expected as the spatial smoothness of the superpixel labels is perhaps the strongest cue exploited by all existing segmentation models.

Iterative learning of factorial correlation: As mentioned in Sec. 3.3 of our main paper, the within-superpixel factorial correlation matrix $\mathbf{M}$ is initialised using image-level co-occurrence statistics and iteratively re-estimated using the superpixel-level inference results using an EM algorithm. Figures 1-3 give an illustration of learning inter-factor, intra-superpixel correlation matrix for the aPascal dataset, where the initial coarsely estimated correlation (Fig. 1) is refined (Fig. 2) to become cleaner and closer to the ground-truth pixel-level correlation (Fig. 3).

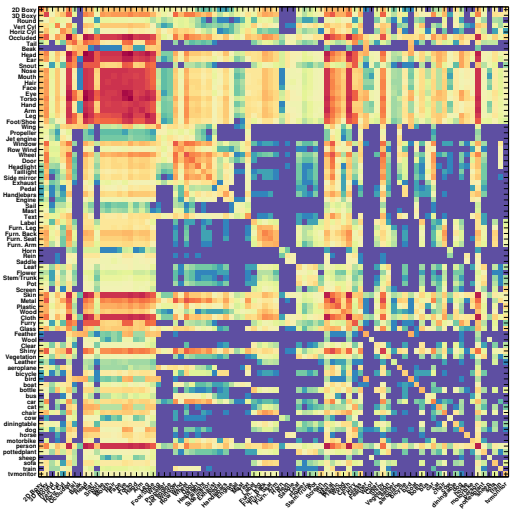

Fig. 1: Initial estimate from image-level annotation

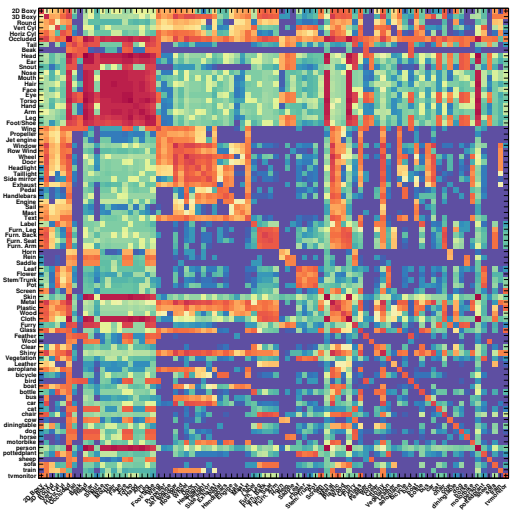

Fig. 2: Our model's refinement

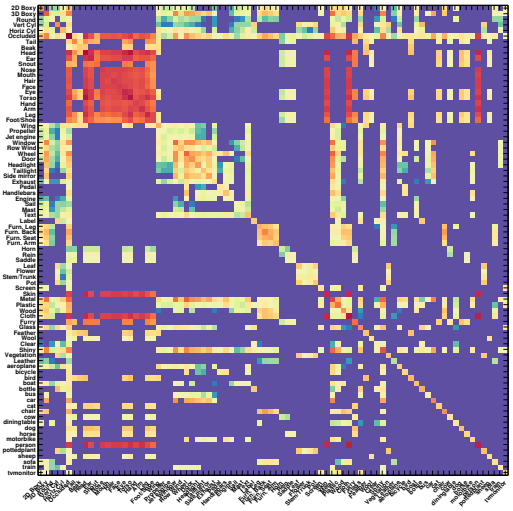

Fig. 3: Ground-truth