

Supplementary Material: “Separating Objects and Clutter in Indoor Scenes”

S. H. Khan* Xuming He[†] M. Bannamoun* F. Sohel* R. Togneri[‡]

*School of CSSE UWA [†]NICTA and ANU [‡]School of EECE UWA

{salman.khan,mohammed.bennamoun,ferdous.sohel,roberto.togneri}@uwa.edu.au, xuming.he@nicta.com.au

This document comprise of supplementary material accompanying [1].

1. Inference as MILP

The complete set of linear inequality constraints for $\mathbf{c}$ and $\mathbf{m}$ is as follows:

$$c_i \geq y_{i,j}, c_j \geq y_{i,j}, y_{i,j} \geq c_i + c_j - 1, \quad (1)$$

$$\forall i, j : o_i \text{ and } o_j \in \mathcal{O}_{oc}, 0 \leq \mu_{i,j}^{obs} < \alpha_{obs},$$

$$\forall i, j : o_i \text{ or } o_j \in \mathcal{O}_{sbc}, 0 \leq \mu_{i,j}^{obs} < \alpha'_{obs}. \quad (2)$$

$$c_i \geq x_{i,j}, c_j \geq x_{i,j}, x_{i,j} \geq c_i + c_j - 1, \quad (3)$$

$$\forall i, j : o_i \text{ and } o_j \in \mathcal{O}_{oc}, 0 \leq \mu_{i,j}^{int} < \alpha_{int},$$

$$\forall i, j : o_i \text{ or } o_j \in \mathcal{O}_{sbc}, 0 \leq \mu_{i,j}^{int} < \alpha'_{int}. \quad (4)$$

$$c_i + c_j \leq 1, \quad (5)$$

$$\forall i, j : o_i \text{ and } o_j \in \mathcal{O}_{oc}, \mu_{i,j}^{int} \geq \alpha_{int} \vee \mu_{i,j}^{obs} \geq \alpha_{obs},$$

$$\forall i, j : o_i \text{ or } o_j \in \mathcal{O}_{sbc}, \mu_{i,j}^{int} \geq \alpha'_{int} \vee \mu_{i,j}^{obs} \geq \alpha'_{obs}. \quad (6)$$

$$m_i \geq w_{i,j}, m_j \geq w_{i,j}, w_{i,j} \geq m_i + m_j - 1, \forall i, j \quad (7)$$

$$m_j \leq \sum_{k: s_j \in o_k} c_k. \quad \forall j \quad (8)$$

$$c_k \geq z_{j,k}, m_j \leq 1 - z_{j,k}, z_{j,k} \geq c_k - m_j, \quad (9)$$

$$\forall k : o_k \in \mathcal{O}_{oc}, 0 \leq \mu_{j,k}^{occ} < \alpha_{int},$$

$$\forall k : o_k \in \mathcal{O}_{sbc}, 0 \leq \mu_{j,k}^{occ} < \alpha'_{int}. \quad (10)$$

2. Parameter Learning

Algorithm 1 Parameter Learning using the Structured SVM Formulation

Input: Training set: $\mathcal{T} = \{(\mathbf{y}^n, \mathbf{x}^n)\}_{1 \times N}$; ϵ convergence threshold; initial parameters λ_0

Output: Learned parameters λ^*

```

1:  $\mathbf{S} \leftarrow \emptyset$  // initialize working set of low energy labelings which will be used as active constraints
2:  $\lambda \leftarrow \lambda_0$  // initialize the parameter vector
3: while  $\Delta\lambda \geq \epsilon$  do
4:   for  $n = 1 \dots N$  do
5:      $\mathbf{y}^* \leftarrow \underset{\mathbf{y} \in \mathcal{Y}}{\operatorname{argmin}} E(\mathbf{y}, \mathbf{x}^{(n)}; \lambda) - \Delta(\mathbf{y}^{(n)}, \mathbf{y})$ 
6:     if  $\mathbf{y}^* \neq \mathbf{y}^{(n)}$  then
7:        $\mathbf{S}^{(n)} \leftarrow \mathbf{S}^{(n)} \cup \{\mathbf{y}^*\}$ 
8:     end if
9:   end for
10:   $\lambda^* \leftarrow \underset{\lambda}{\operatorname{argmin}} \frac{1}{2} \|\lambda\|^2 + \frac{C}{N} \sum_n \xi_n$ 
11:  s.t.  $\lambda \geq 0, \xi_n \geq 0$ , // update the parameters such that
12:   $E(\mathbf{y}, \mathbf{x}^n; \lambda) - E(\mathbf{y}^n, \mathbf{x}^n; \lambda) \geq \Delta(\mathbf{y}^{(n)}, \mathbf{y}) - \xi_n \quad \forall \mathbf{y} \in \mathbf{S}^{(n)} \forall n$  // ground truth has lowest energy
13: end while

```

The training set consists of input image ($\mathbf{x}$) and annotation ($\mathbf{y}$) pairs. The annotations $\mathbf{y}$ have labeled cluttered/non-cluttered regions as well as the ground truth cuboids. The energy minimization step in Algorithm 1 (line 5) is solved using the branch and bound method. The weight update step in Algorithm 1 (lines 10 - 12) can be solved using any standard quadratic program solver.

We use the re-scaled margin energy function formulation of Taskar et al. [3] in the above algorithm. The re-scaled margin cutting plane algorithm efficiently adds low energy labelings to the active constraints set and updates the parameters such that the ground-truth has lowest energy. $\Delta(\cdot)$ is the IOU loss function for cuboid matching, defined as :

$$\Delta(\mathbf{y}^{(n)}, \mathbf{y}) = \sum_i \left(1 - \frac{|y_i^{(n)} \cap y_i|}{|y_i^{(n)} \cup y_i|} \right).$$

In our case, the initial parameters (λ_0) are estimated using the piece-wise training method described in [2]. Reasonable estimates of initial parameters make the parameter learning process efficient and less prone to sticking into local minima.

References

- [1] S. H. Khan, X. He, M. Bennamoun, F. Sohel, and R. Togneri. Separating objects and clutter in indoor scenes. In *CVPR*. IEEE, 2015. 1
- [2] J. Shotton, J. Winn, C. Rother, and A. Criminisi. Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *IJCV*, 2009. 2
- [3] B. Taskar, V. Chatalbashev, and D. Koller. Learning associative markov networks. In *ICML*, page 102. ACM, 2004. 2