

## Supplemental Materials for Paper 1874

### Characterizing Layouts of Outdoor Scenes Using Spatial Topic Processes

#### Abstract

This document compiles more experimental results as supplement to the main paper. Overall, the supplemental materials comprise three parts:

- 1) Detailed results of scene classification (including tables and confusion matrices)
- 2) More segmentation results
- 3) Videos to demonstrate the hallucination results, which show how the inferred layouts evolve as the observed part grow

In particular, part (1) and part (2) are included in this document, while part (3) is a video file (layout\_hallu.mp4)

#### Detailed Results of Scene Classification

Due to page limit, the main paper only provides a summary of the results (as shown in Fig 4 and Fig 5). More details are provided here.

Following are classification correct rates under different settings on both data sets (MSRCv2 and SUN), organized as tables.

**Table 1: Scene classification correct rates on MSRC (v2)**

<i>npi</i>	SPM-L0	SPM-L1	SPM-L2	STP-3	STP-4	STP-6
500	0.2780	0.3017	0.3085	0.5458	0.5729	0.5424
1000	0.3695	0.4237	0.4271	0.5831	0.6373	0.6780
2000	0.5017	0.5220	0.5559	0.6305	0.6949	0.7322
5000	0.6203	0.6610	0.6847	0.6712	0.7220	0.7390

**Table 2: Scene classification correct rates on SUN**

<i>npi</i>	SPM-L0	SPM-L1	SPM-L2	STP-3	STP-4	STP-6
500	0.3689	0.3711	0.3745	0.6208	0.6509	0.7075
1000	0.4331	0.4895	0.5034	0.6992	0.7375	0.7810
2000	0.5874	0.6196	0.6223	0.7434	0.7830	0.8005
5000	0.7129	0.7295	0.7500	0.7593	0.8108	0.8213

Here, *npi* is the number of visual words extracted per image. *SPM-L0*, *SPM-L1*, and *SPM-L2* refer to Spatial Pyramid Matching with no cell division (like bag of features), one division level (into 4 cells), and two division levels (into 16 cells). *STP-3*, *STP-4*, and *STP-6* refer to Spatial Topic Process with 3-by-3, 4-by-4, and 6-by-6 grids.



## More Segmentation Results

The following four pages show more segmentation results. Each page contains 16 groups of results, two on each row. Each group comprises three images: input image, inferred layout, and semantic segmentation (from left to right).

The following is the color map used in coloring semantic segmentations.

building	ground	snowy-mt	bridge
balcony	road	water	snow
fence	path	sea	ice
sky	grass	wave	animal
cloud	vehicle	sand	person
trees	ship	stream	wheat
plants	mountain	board	flowers







