Supplemental Material for Single Image 3D Without a Single 3D Image

David F. Fouhey[1], Wajahat Hussain[2], Abhinav Gupta[1], Martial Hebert[1]
[1] Robotics Institute, Carnegie Mellon University, USA
[2] University of Zaragoza, Spain

Contents: This supplementary material contains additional results for our method on both NYUv2 and Places-205. These include quantitative sensitivity analysis as well as qualitative results demonstrating the method.

1. **Sensitivity Analysis:** We report fuller sensitivy analysis of our method, which demonstrates that it is robust to parameter settings.
2. **Random Results on NYUv2:** We report more results on NYUv2.
3. **Random Results on Places-205:** We report results on the Places dataset.
4. **Style elements from Places-205:** We show some style elements learned on each category in Places-205.
5. **Illustration of Top/Bottom NYUv2 Style Elements:** We show the top and bottom style elements learned on NYUv2 with the aim of illustrating the selection criterion.

1. Sensitivity Analysis

We report sensitivity analysis in Table 1 for the important parameters of our method, in particular its prior. We check parameters controlling its beliefs) (i.e., how it assigns probabilities) and how it is factored into the final probability (i.e., how many votes it gets, in terms of ELDA detection scores). These aim to cover a large range of values, including ones that are unlikely.

Parameter settings: Where a linear grid makes sense, we try a grid of $\frac{1}{4}, \frac{1}{2}, 1, 2, 4$ times the parameter's value. For the box's maximum aspect ratio, we treat it as being $1 + d$ (as otherwise, we'd try 0.5, which is equivalent to 2); we thus try $1.25, 1.5, 2, 3, 5$. For the bandwith of our squared negative expontenial, a bandwidth of 0 or of 1 times the image height is obviously incorrect, so we use a tighter grid of $\frac{3}{8}, \frac{3}{8}, \ldots, \frac{6}{8}$, excluding $\frac{2}{8}$ as an impossible setting (it assigns 2% chance to a horizontal surface in middle of the image).

Results: The results are stable and all even using implausible values leaves us competitive with 3DP and substantially outperforming other unsupervised methods. If we double or halve the parameters, all results are well below 1° and 1%. If we go as much as $4x$ or $1/4$, which includes parameters that are implausible, most results are again within 1° or 1%. The most sensitive parameter is the number of votes the vertical/horizontal prior gets; even here, changing things by a factor of four gives only a drop of 1.7° in median error, and only at most 1.3%.

Table 1. Sensitivity Analysis

		Summary Stats. (°) (Lower Better)			% Good Pixels (Higher Better)		
		Mean	Median	RMSE	11.25°	22.5°	30°
Box Max. Aspect Ratio	5	38.6	21.3	52.9	37.4	51.0	55.6
	3	38.2	21.1	52.3	37.3	51.1	55.9
	2	38.6	21.7	52.6	36.8	50.6	55.4
	1.5	38.7	22.1	52.6	36.5	50.3	55.1
	1.25	38.9	22.3	52.8	36.3	50.1	54.9
Max. Difference Within	Factor of 2	0.4	0.6	0.3	0.5	0.5	0.5
	Factor of 4	0.4	0.6	0.3	0.6	0.5	0.5
Vertial Prior Strength	40	38.5	22.0	52.4	36.4	50.4	55.3
	20	38.4	21.6	52.4	36.7	50.7	55.5
	10	38.6	21.7	52.6	36.8	50.6	55.4
	5	38.6	21.8	52.7	36.8	50.6	55.3
	2.5	38.8	21.9	52.9	36.8	50.5	55.2
Max. Difference Within	Factor of 2	0.2	0.1	0.2	0.1	0.1	0.1
	Factor of 4	0.2	0.3	0.3	0.4	0.3	0.2
Vert./Horiz. Prior Strength	200	39.1	22.3	53.3	36.6	50.2	54.8
	100	38.8	22.3	52.7	36.4	50.2	54.9
	50	38.6	21.7	52.6	36.8	50.6	55.4
	25	38.8	22.0	52.9	36.7	50.4	55.1
	12.5	39.4	23.4	53.2	35.5	49.3	54.2
Max. Difference Within	Factor of 2	0.2	0.6	0.3	0.4	0.4	0.5
	Factor of 4	0.8	1.7	0.7	1.3	1.3	1.2
Vert. Prior Exp. Bandwidth	0.75	38.7	21.6	52.9	37.1	50.7	55.3
	0.625	38.7	21.6	52.8	37.0	50.7	55.3
	0.5	38.6	21.7	52.6	36.8	50.6	55.4
	0.375	38.9	22.5	52.8	36.1	50.0	54.8
Max. Difference	—	0.3	0.8	0.2	0.7	0.6	0.6

2. Random Results on NYUv2

We now show results on NYUv2. These were chosen and sorted randomly (i.e., using `randperm`) for fairness. **Notes for interpreting:** Unusual-looking colors in NYUv2 correspond to vanishing-point estimation failure.

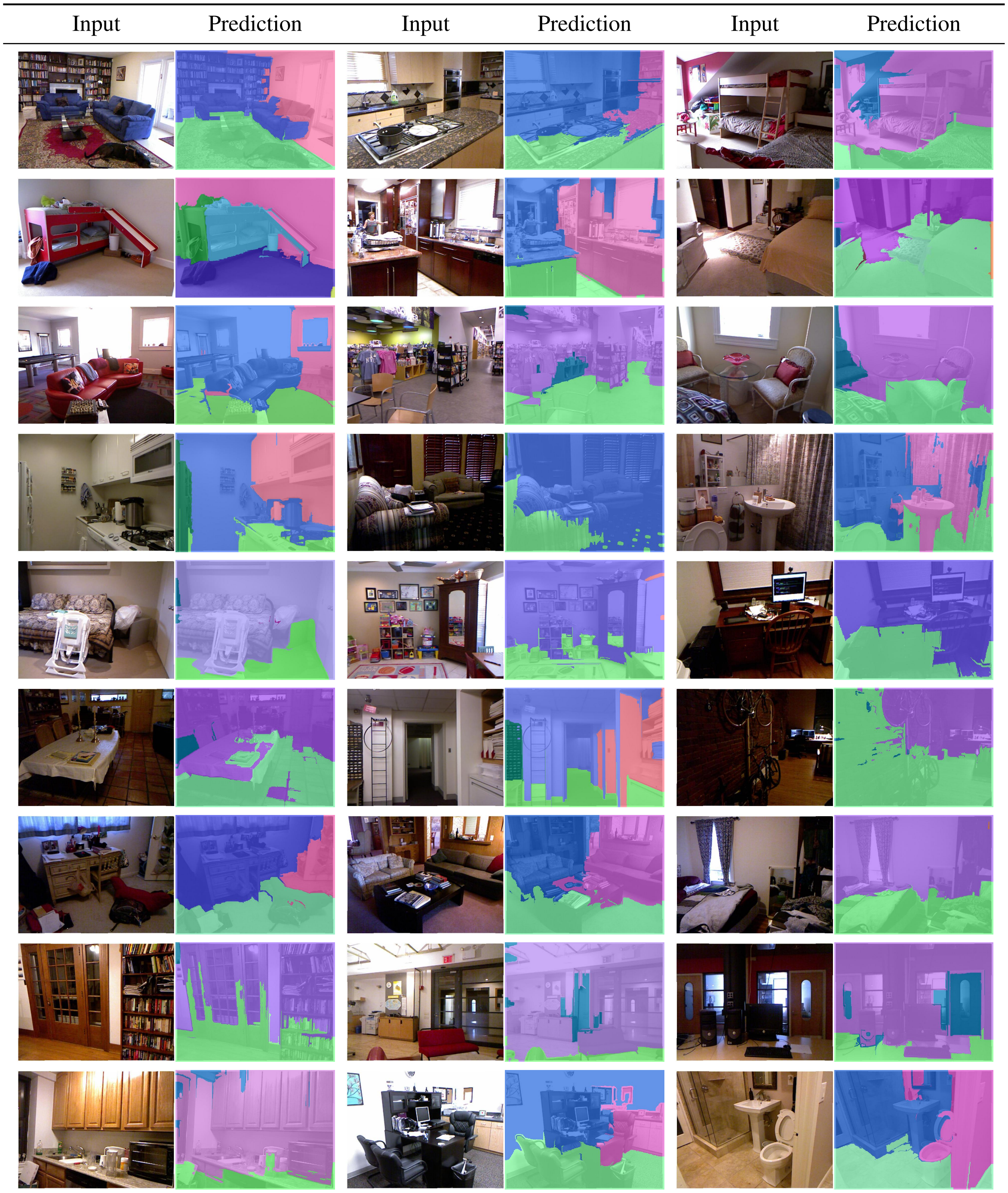


Figure 1. Random Results on NYUv2

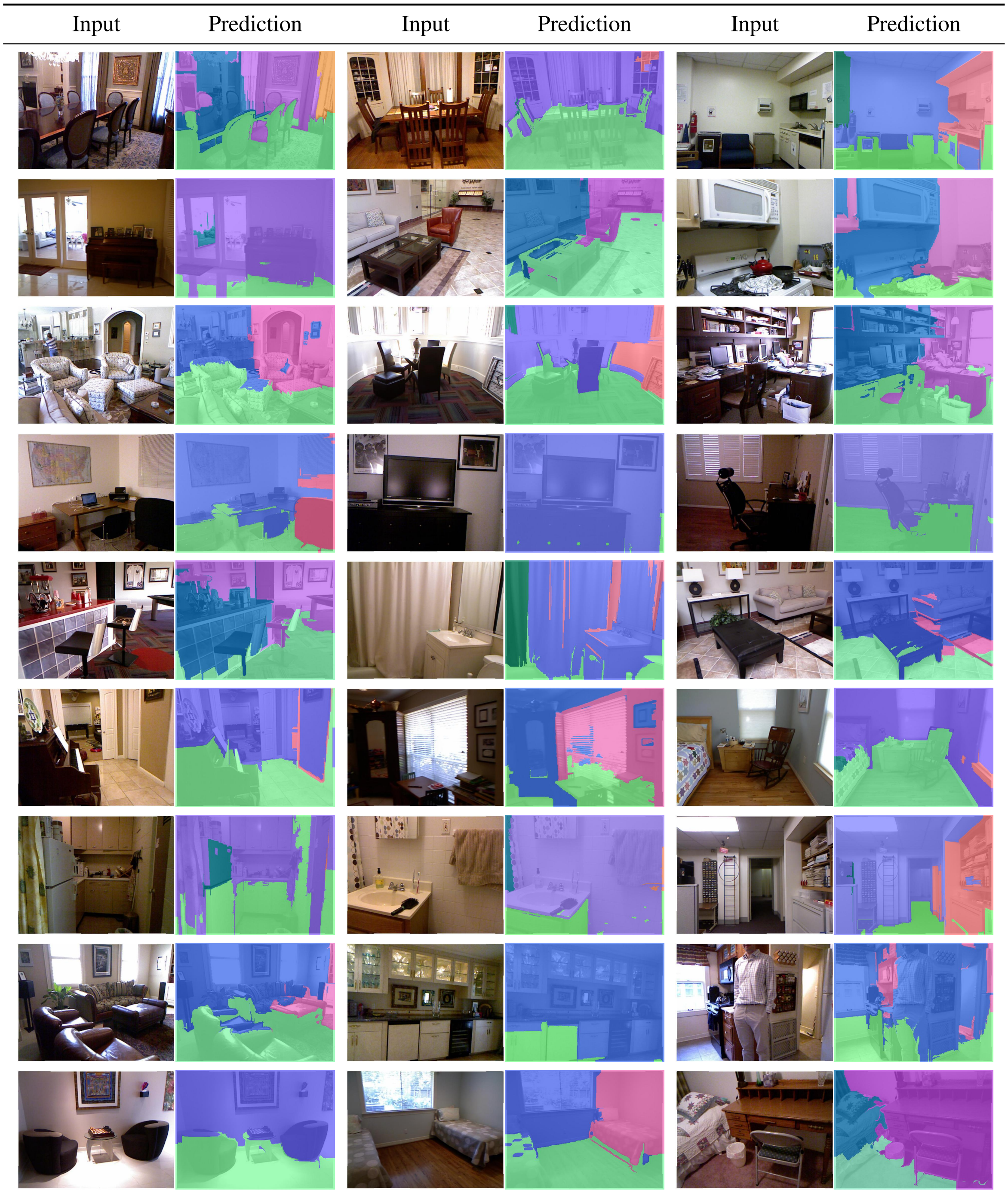


Figure 2. Random Results on NYUv2

3. Random Results on Places-205

We now show random results from each Places-205 category we used. Our method generally interprets the results well; again, as in NYU, sometimes the segmentation propagation is too aggressive.

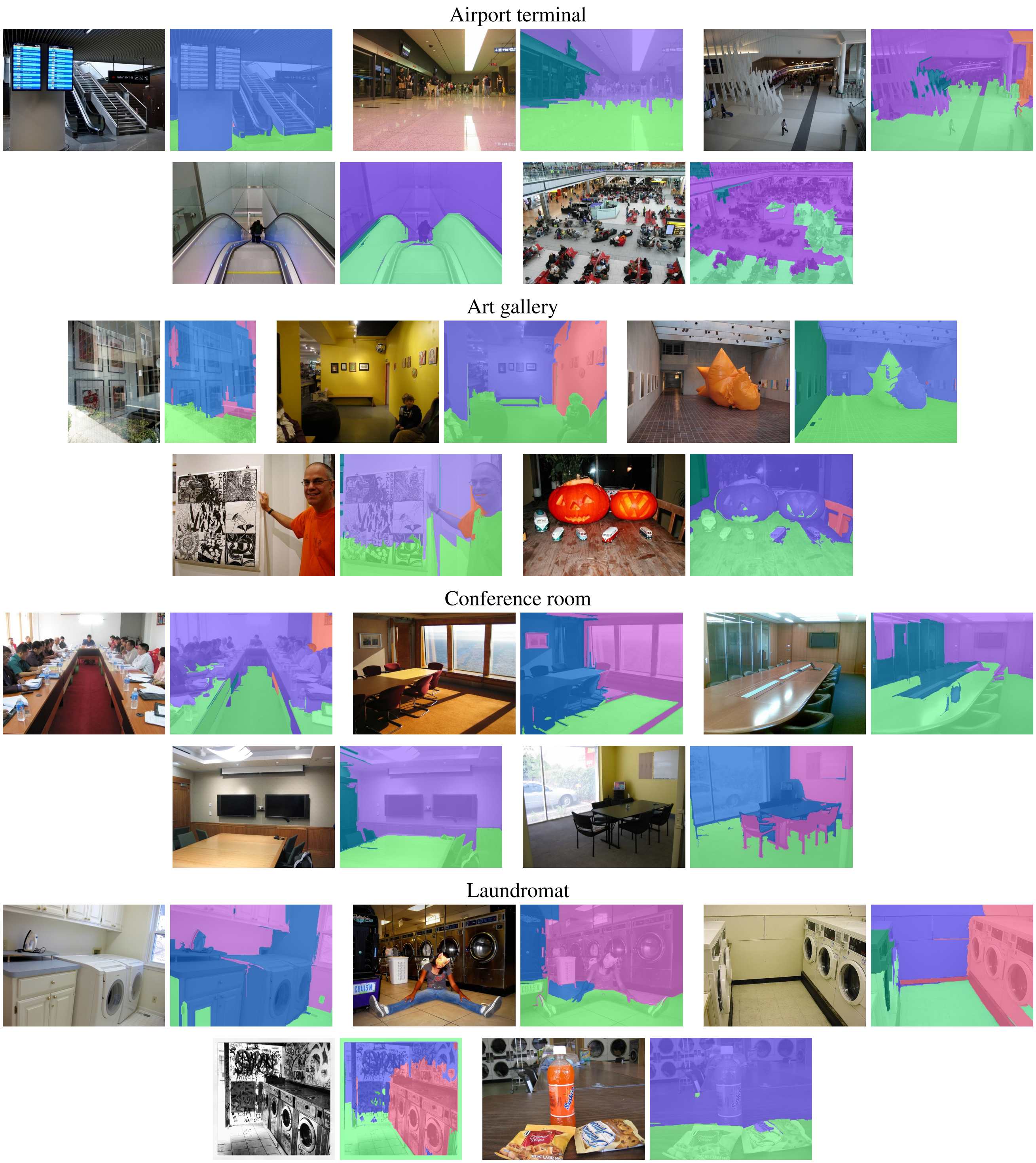


Figure 3. Random Results on Places-205, subdivided by category

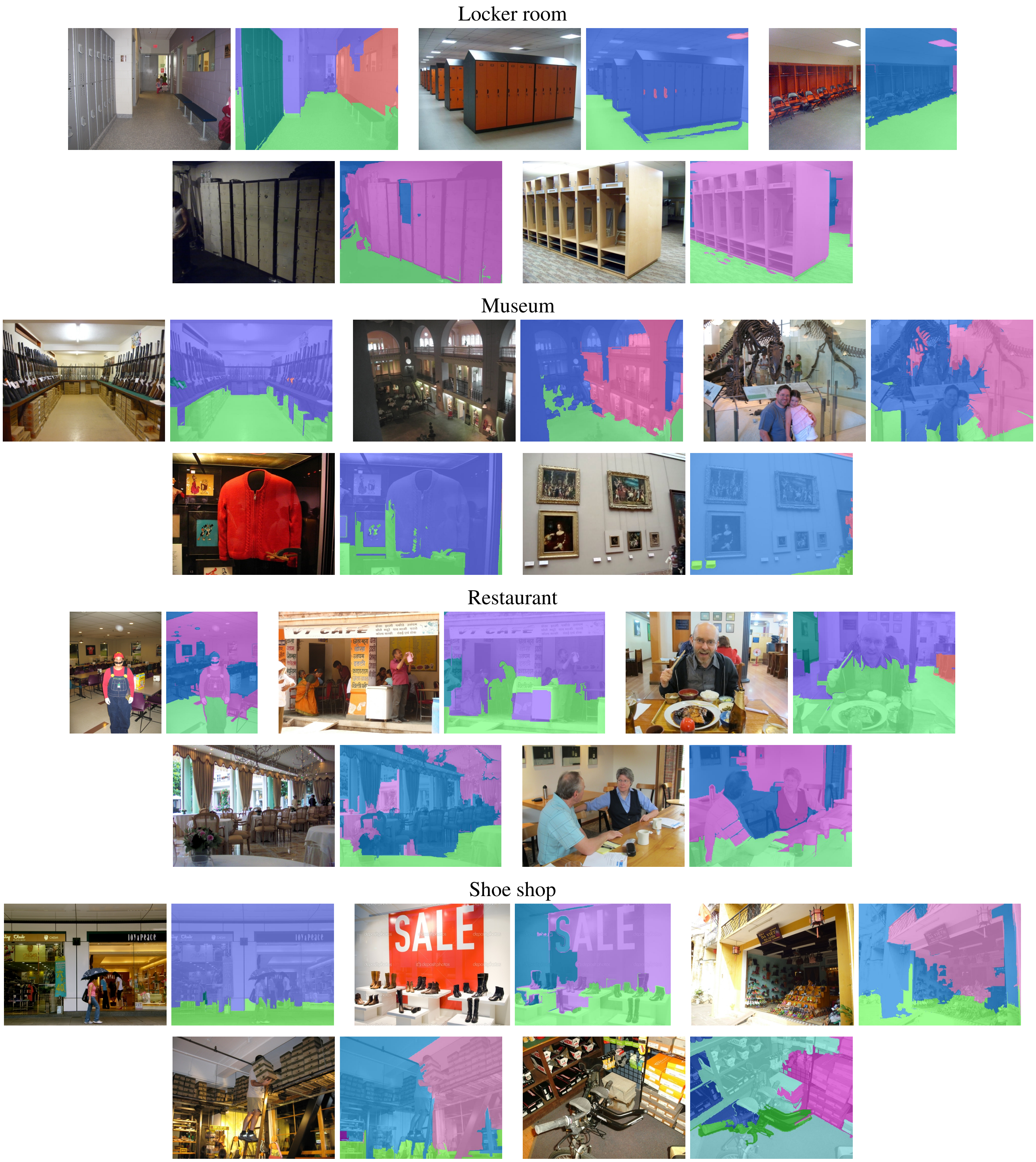


Figure 4. Random Results on Places-205, subdivided by category

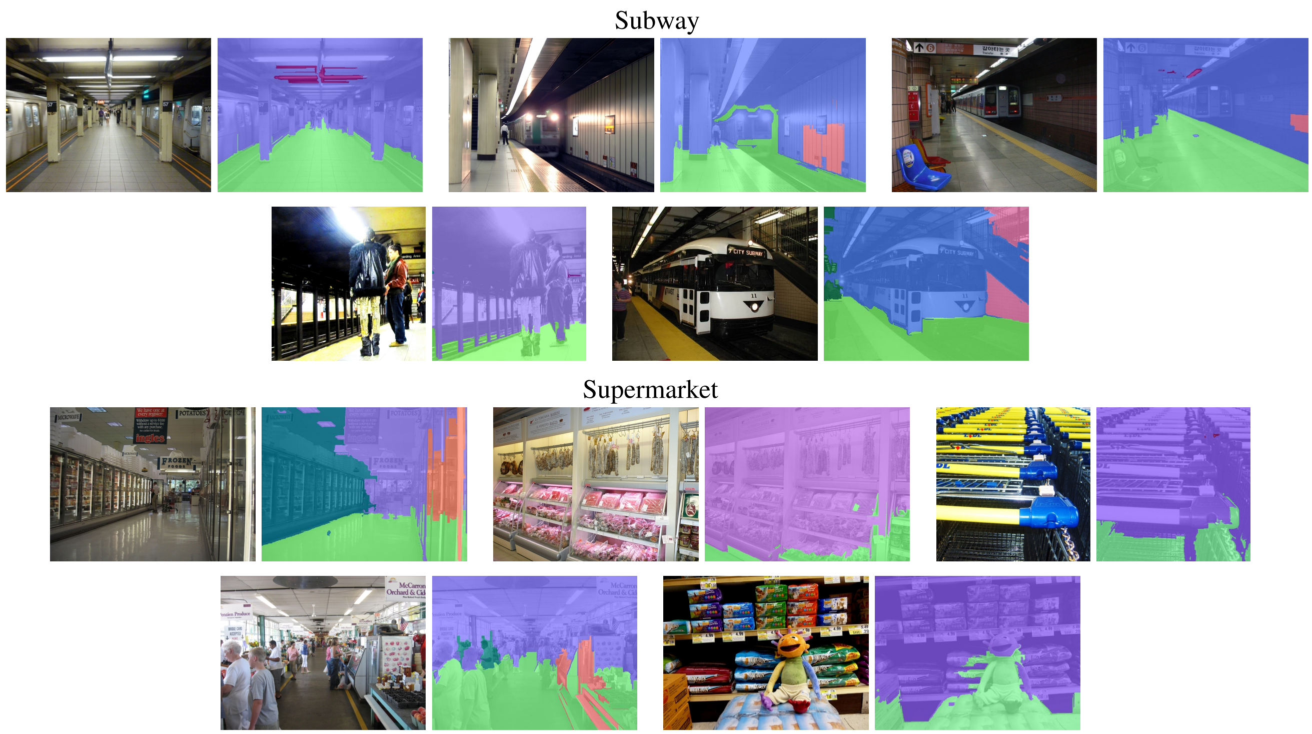


Figure 5. Random Results on Places-205, subdivided by category

4. Elements from Places-205

We now show randomly selected vertical elements learned on the Places-205 dataset. These are typically fairly scene-specific: in laundromats, our method latches on to the fronts of washers and dryers, and in supermarkets, it picks up on shelving units. One frequent failure mode is when the element is rectified nearly correctly; then, it is close enough that its HOG representation yields an effective style element, but appears slightly off to viewers.

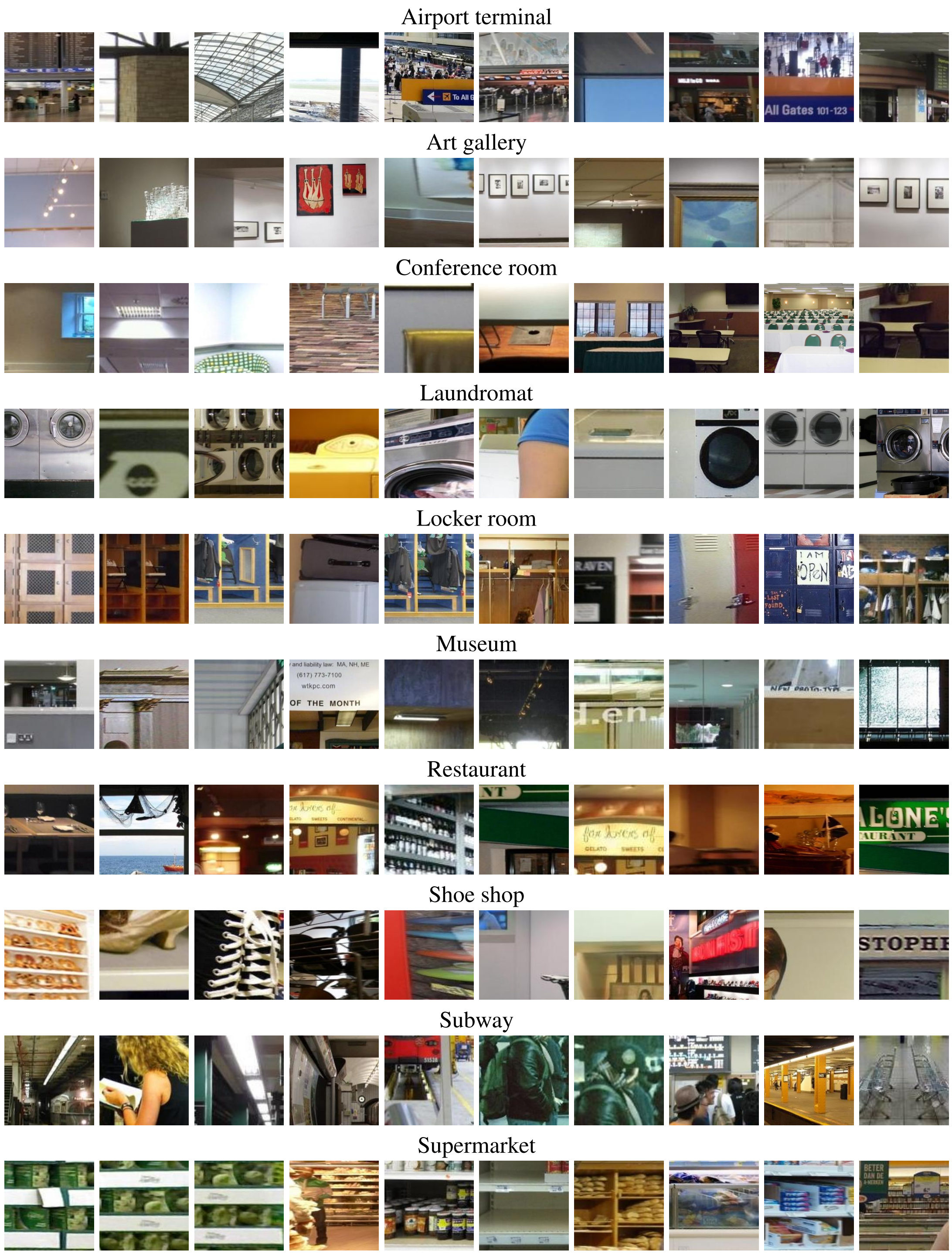
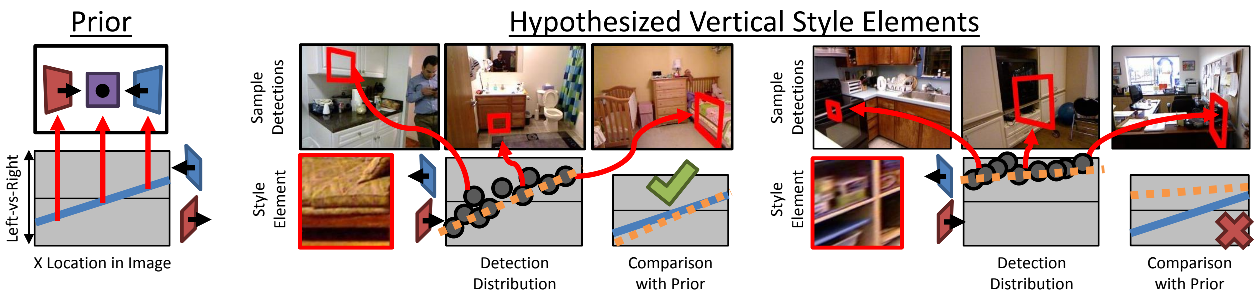


Figure 6. Random Style Elements Learned on Places-205



5. Top/Bottom Vertical Style Elements from NYUv2

We now show the top 5 ranked patches and the bottom 5 vertical patches from our pool of hypothesized style elements. For each patch, we show:

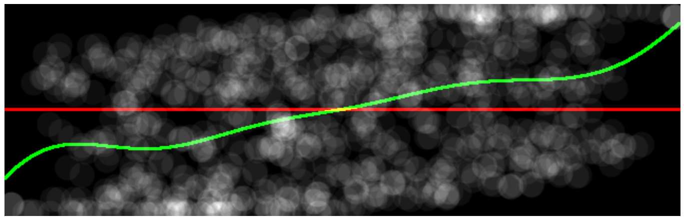
1. (top-left) a hypothesized element (top-left);
2. (top-right) a scatter plot of its detections' orientations (right-to-left) as a function of location on the X axis. We show a fit to this data in green; we show fronto-parallel in red. The green line represents the average orientation with respect to the x-axis for the style element's detections. **The y-axis goes from right-to-left; the x-axis goes across the x axis of the image**.
3. (bottom) some automatically selected top detections of the style element on the discovery set. We show both the rectified image on which the element had a detection as well as the original image.

For reference, we reproduce Fig. 5 of the paper, which shows how this fitted orientation-with-respect-to-X-location is used; in order to identify good and bad style elements, the green fit to the detection data is compared to the prior. Those that match the prior are ranked high; those that do not are ranked low.

Notice that bottom-ranked patches are not rectified to the correct direction and often fire consistently in one orientation across the whole image.

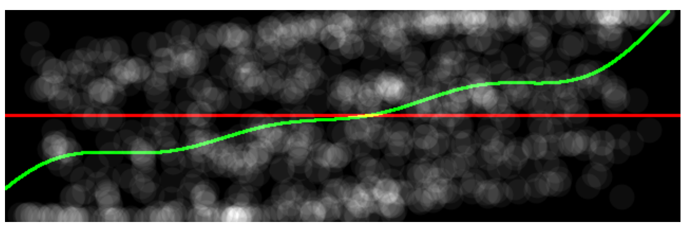
Top Ranked Results

Style Element

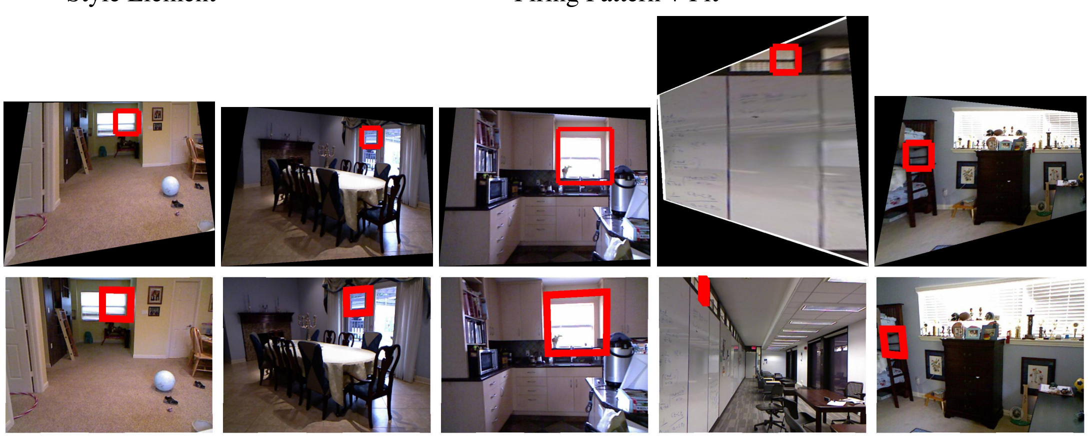


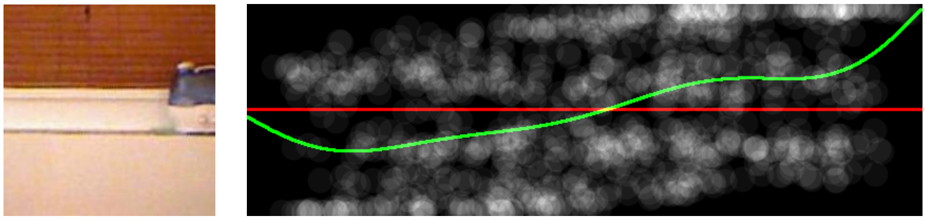
Firing Pattern + Fit

Style Element



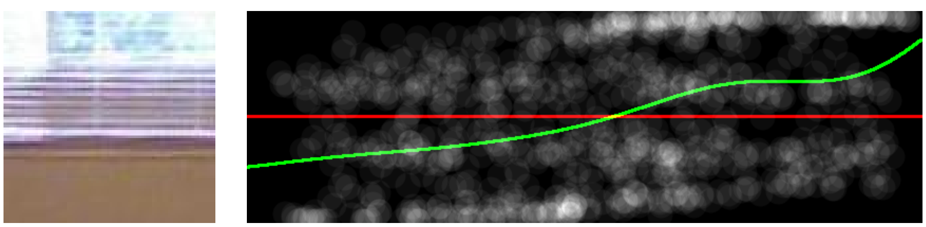
Firing Pattern + Fit





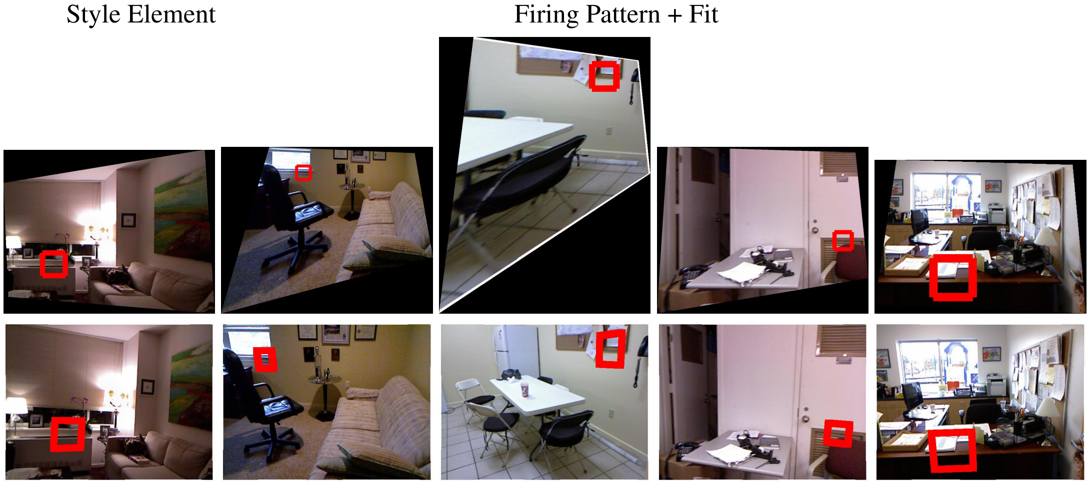
Style Element

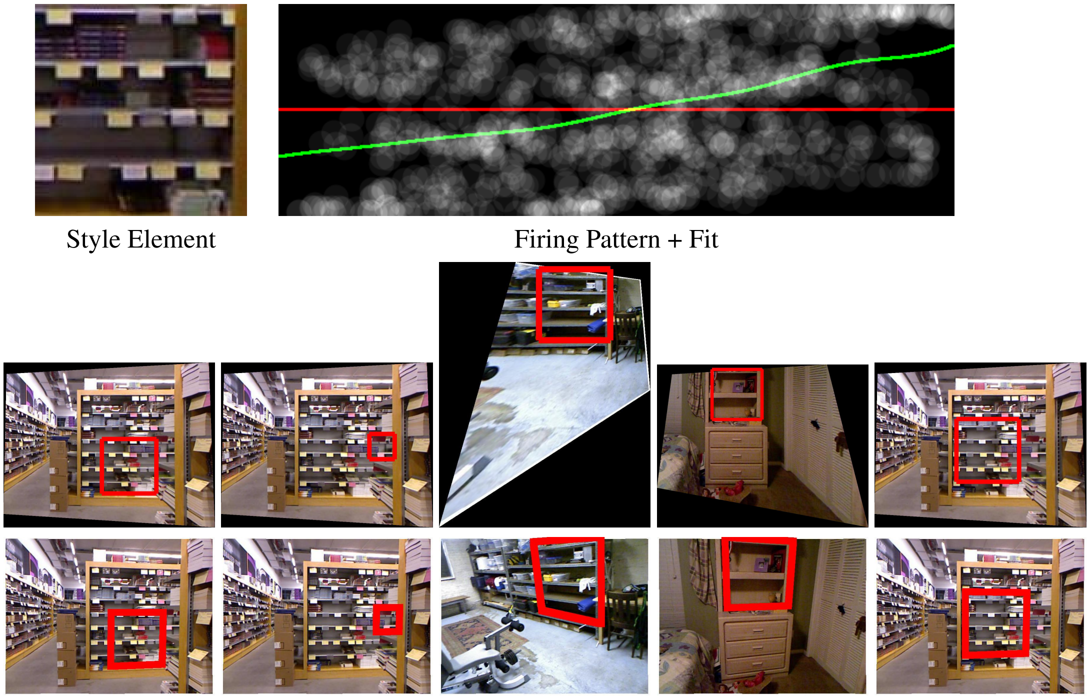
Firing Pattern + Fit



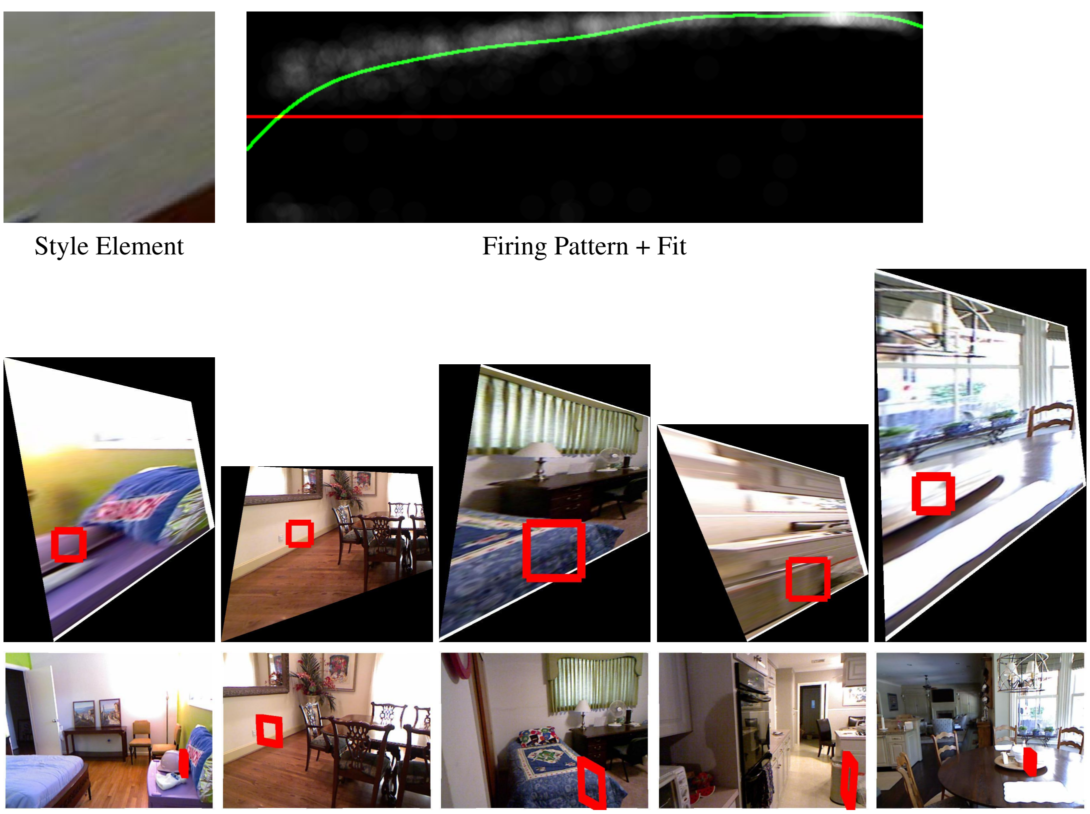
Style Element

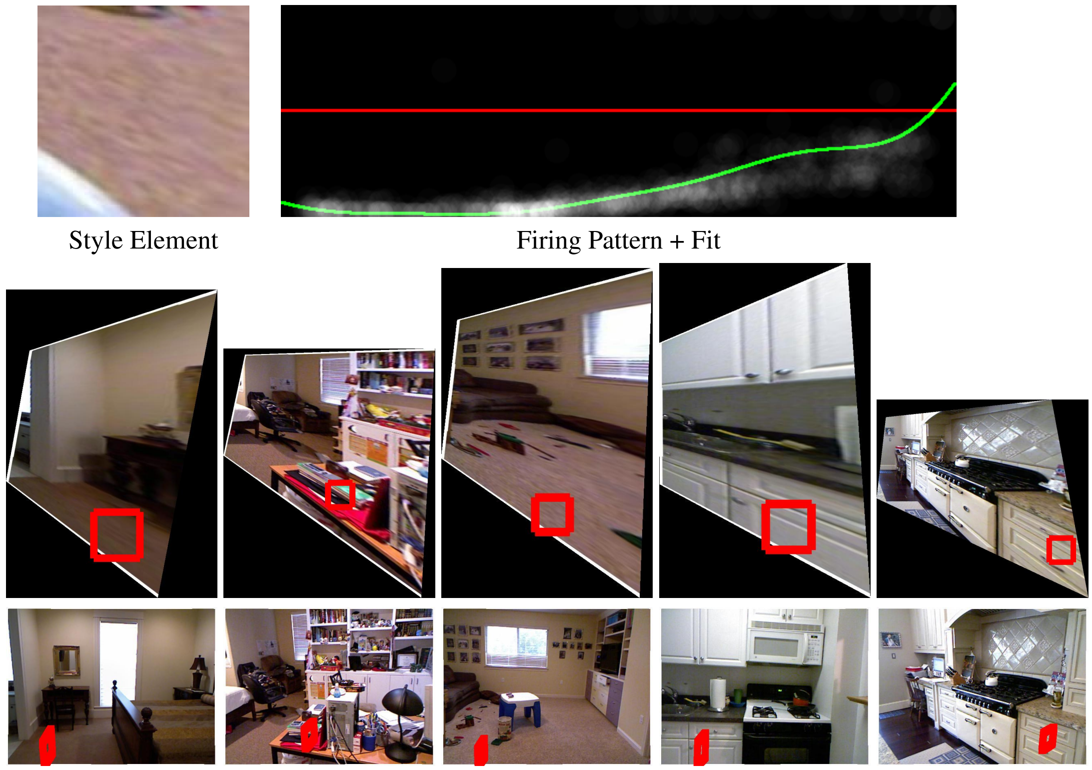
Firing Pattern + Fit

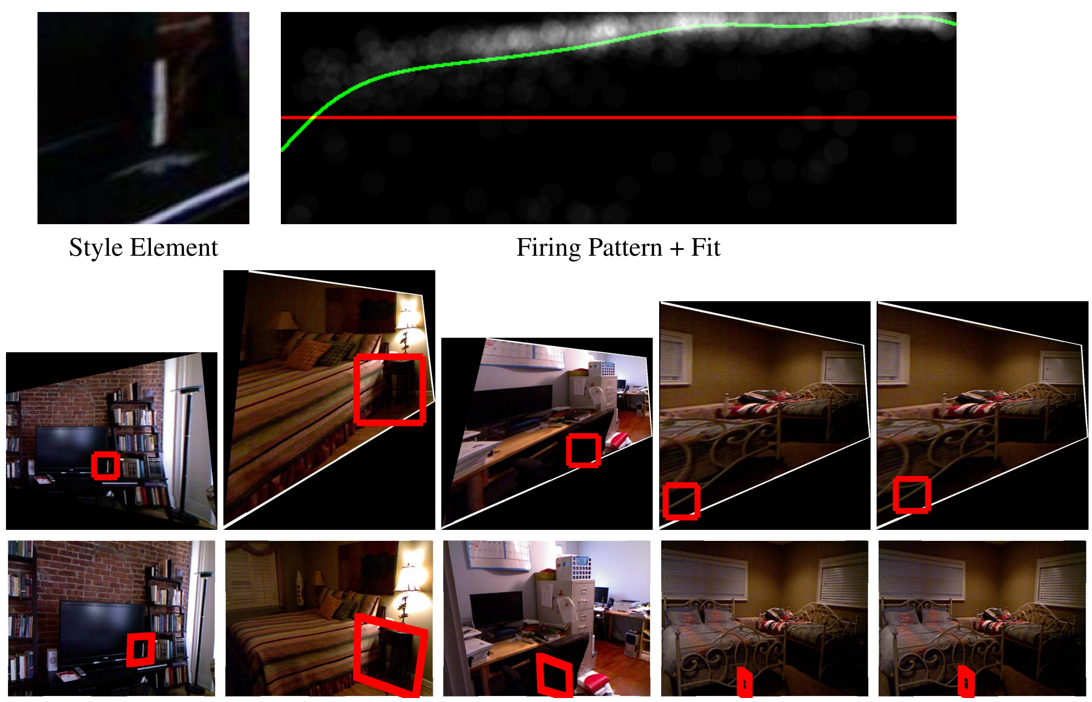


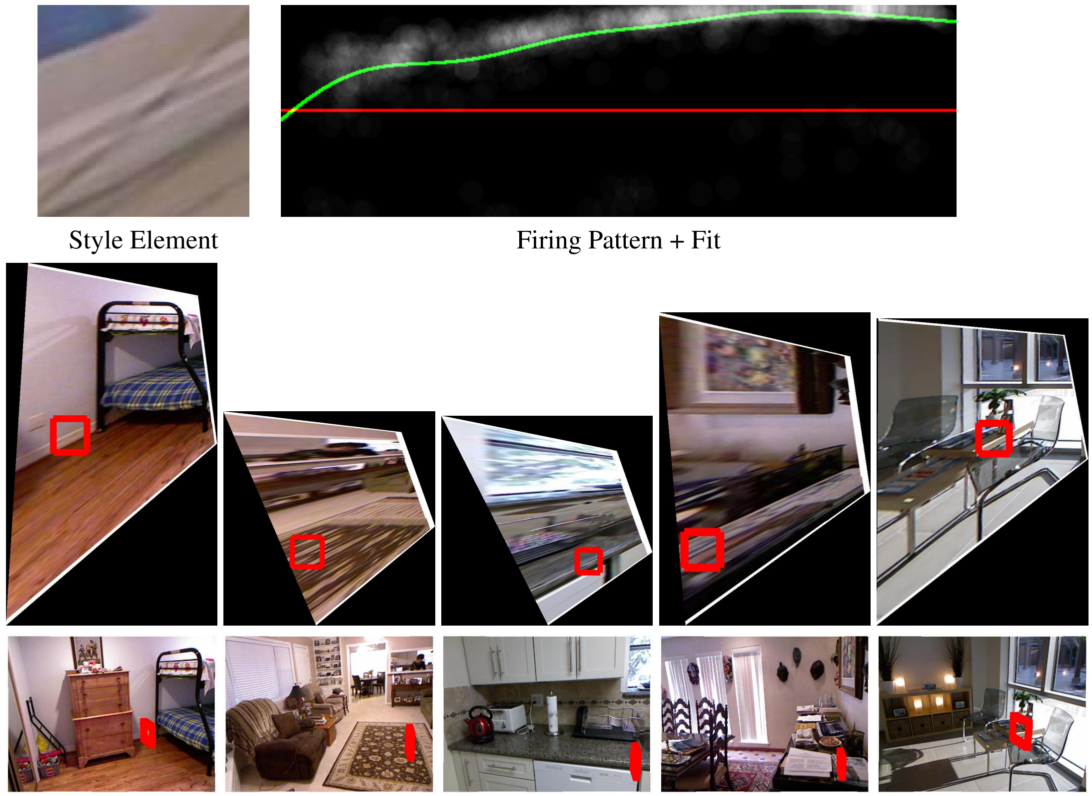


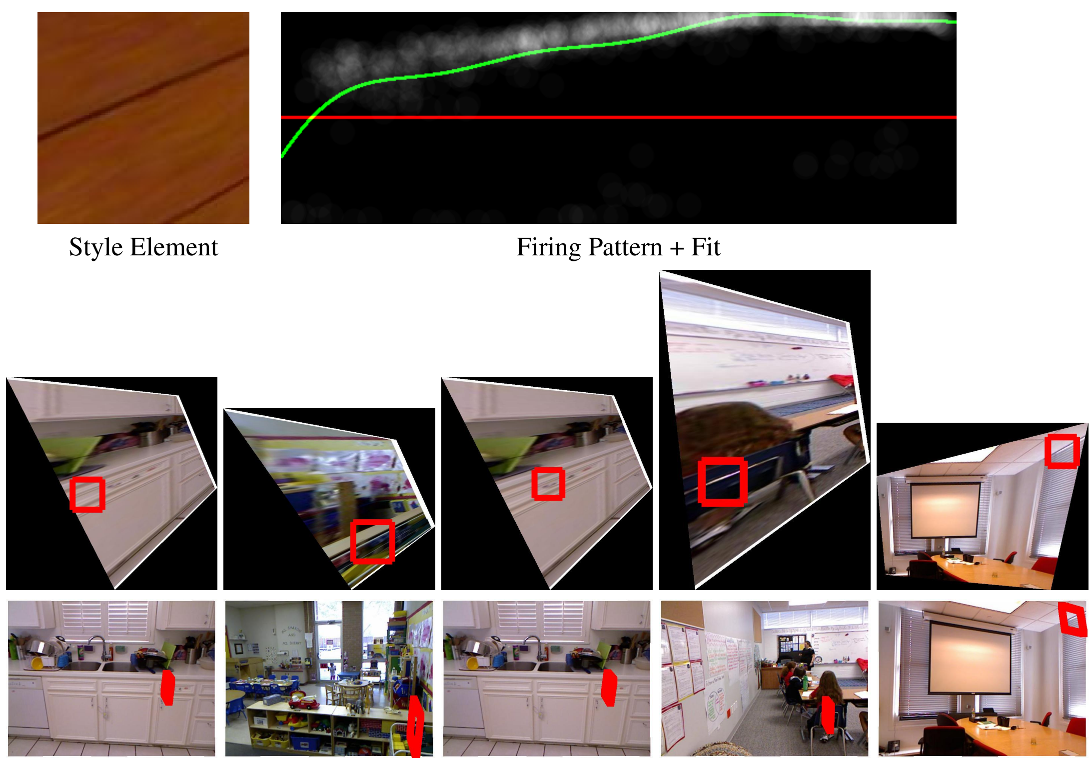
Lowest Ranked Results




Style Element
Firing Pattern + Fit


Style Element
Firing Pattern + Fit


Style Element
Firing Pattern + Fit


Style Element
Firing Pattern + Fit