Perceptual Organization and Recognition of Indoor Scenes from RGB-D Images Supplementary Material

Saurabh Gupta, Pablo Arbeláez, and Jitendra Malik
University of California, Berkeley - Berkeley, CA 94720
{sgupta, arbelaez, malik}@eecs.berkeley.edu

In the supplementary material, we complete the derivation for estimation of gravity direction (Section 1), and provide more qualitative examples of the output of our system (Section 2).

1. Extracting a Geocentric Coordinate Frame

Recall that in iteration i, we have an old estimate of the gravity vector $\boldsymbol{g}_{i-1}$ and we make hard assignments of the local surface normals at all points into 2 sets - aligned set $\mathcal{N}_{\|}$ and orthogonal set $\mathcal{N}_{\perp}$, (based on a threshold d on the angle made by the local surface normal with the gravity vector) as follows and then stack the vectors in $\mathcal{N}_{\|}$ to form a matrix $N_{\|}$, and stack the ones in $\mathcal{N}_{\perp}$ to form $N_{\perp}$.

$$\begin{aligned}\mathcal{N}_{\|} &= \{\boldsymbol{n} : \theta(\boldsymbol{n}, \boldsymbol{g}_{i-1}) < d \text{ or } \theta(\boldsymbol{n}, \boldsymbol{g}_{i-1}) > 180^\circ - d\} \\ \mathcal{N}_{\perp} &= \{\boldsymbol{n} : 90^\circ - d < \theta(\boldsymbol{n}, \boldsymbol{g}_{i-1}) < 90^\circ + d\} \\ & \quad \text{where, } \theta(\boldsymbol{a}, \boldsymbol{b}) = \text{Angle between } \boldsymbol{a} \text{ and } \boldsymbol{b}.\end{aligned}$$

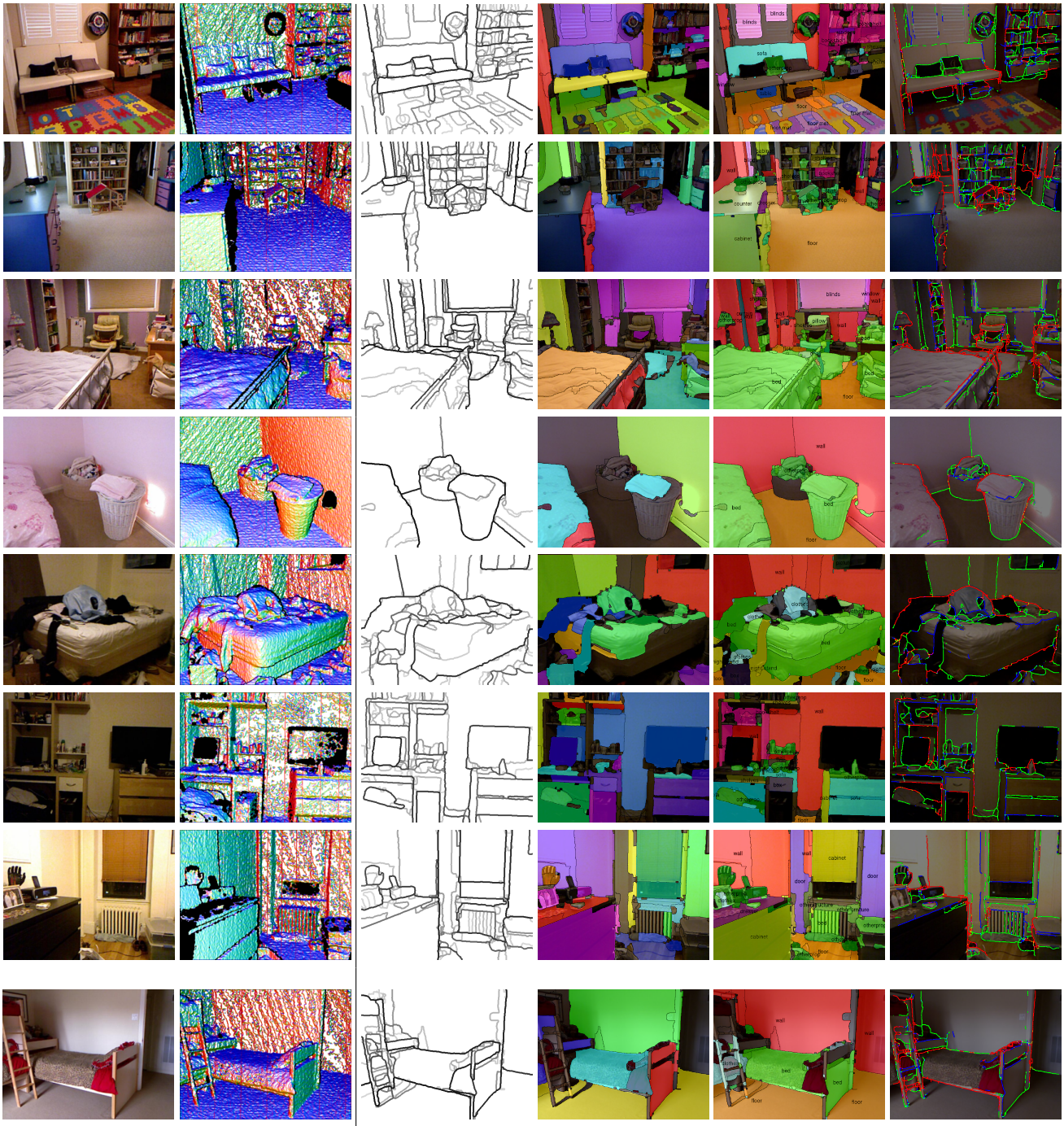
We then find our new estimate for the gravity vector $\boldsymbol{g}_i$, by finding a vector that is the as aligned to vectors in the aligned set and as orthogonal to the vectors in the orthogonal set by solving the following optimization problem which as we show corresponds to solving for the eigen vector with the smallest eigen value of the matrix $N_{\perp}N_{\perp}^t - N_{\|}N_{\|}^t$.

$$\begin{aligned}\boldsymbol{g}_i &= \underset{\boldsymbol{g}:\|\boldsymbol{g}\|_2=1}{\operatorname{argmin}} \sum_{\boldsymbol{n}\in\mathcal{N}_{\perp}} \cos^2(\theta(\boldsymbol{n}, \boldsymbol{g})) + \sum_{\boldsymbol{n}\in\mathcal{N}_{\|}} \sin^2(\theta(\boldsymbol{n}, \boldsymbol{g})) \\ &= \underset{\boldsymbol{g}:\|\boldsymbol{g}\|_2=1}{\operatorname{argmin}} \sum_{\boldsymbol{n}\in\mathcal{N}^{\perp}} \left(\boldsymbol{g}^t\boldsymbol{n}\right)^2 + \sum_{\boldsymbol{n}\in\mathcal{N}^{\|}} \left(1 - (\boldsymbol{g}^t\boldsymbol{n})^2\right) \\ &= \underset{\boldsymbol{g}:\|\boldsymbol{g}\|_2=1}{\operatorname{argmin}} \|N_{\perp}^t\boldsymbol{g}\|_2^2 - \|N_{\|}^t\boldsymbol{g}\|_2^2 \\ &= \underset{\boldsymbol{g}:\|\boldsymbol{g}\|_2=1}{\operatorname{argmin}} \boldsymbol{g}^t \left(N_{\perp}N_{\perp}^t - N_{\|}N_{\|}^t\right) \boldsymbol{g}\end{aligned}$$

2. Results

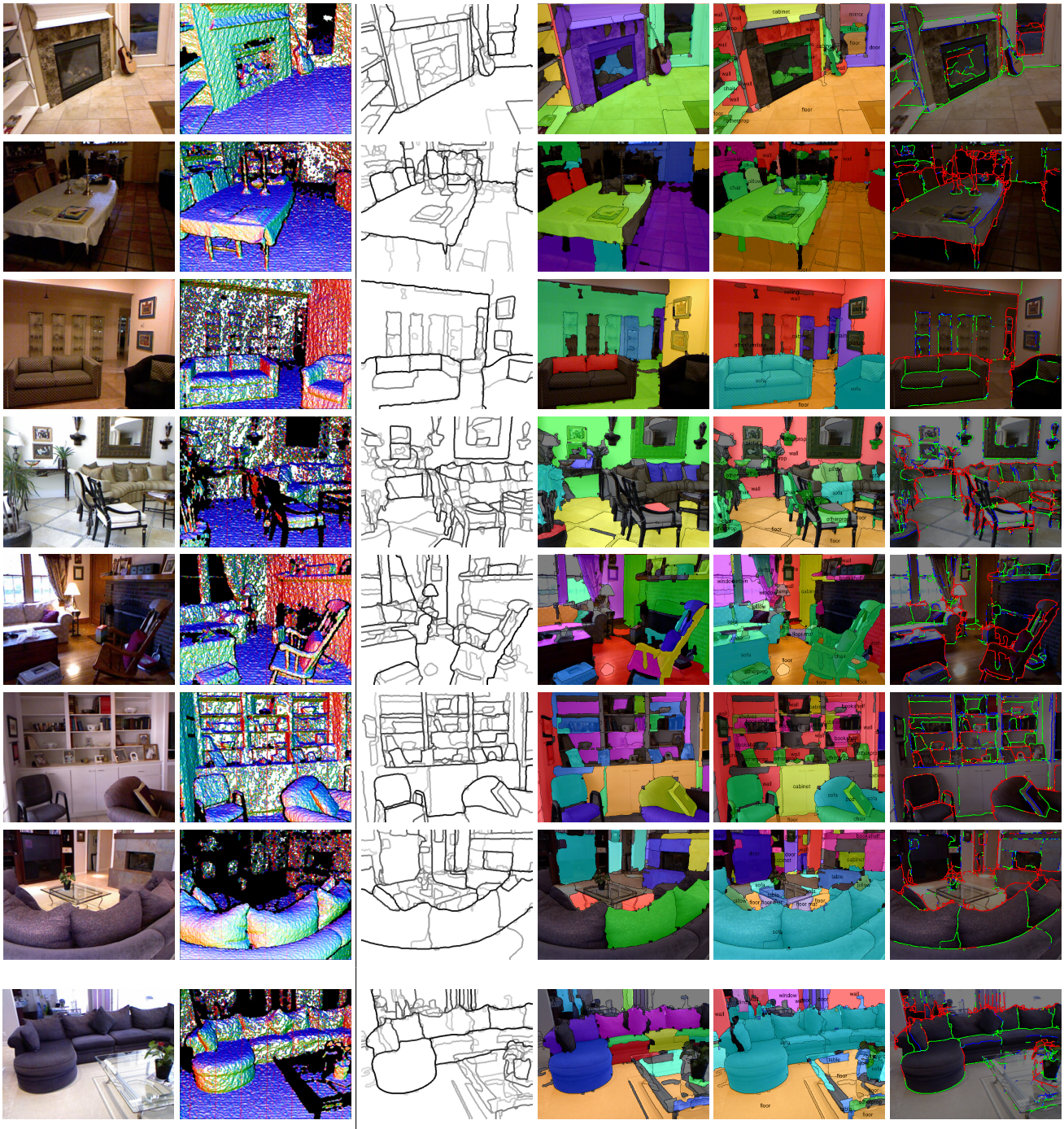
2.1. Semantic Segmentation

We provide more visualizations for the results in the following Figures. We pick the major scene categories from the test set and show few random images from each of the categories - bedroom (Figure 1), living room (Figure 2), dining room (Figure 3), kitchen (Figure 4), bathroom (Figure 5), office (Figure 6), classroom (Figure 7), and other (Figure 8).



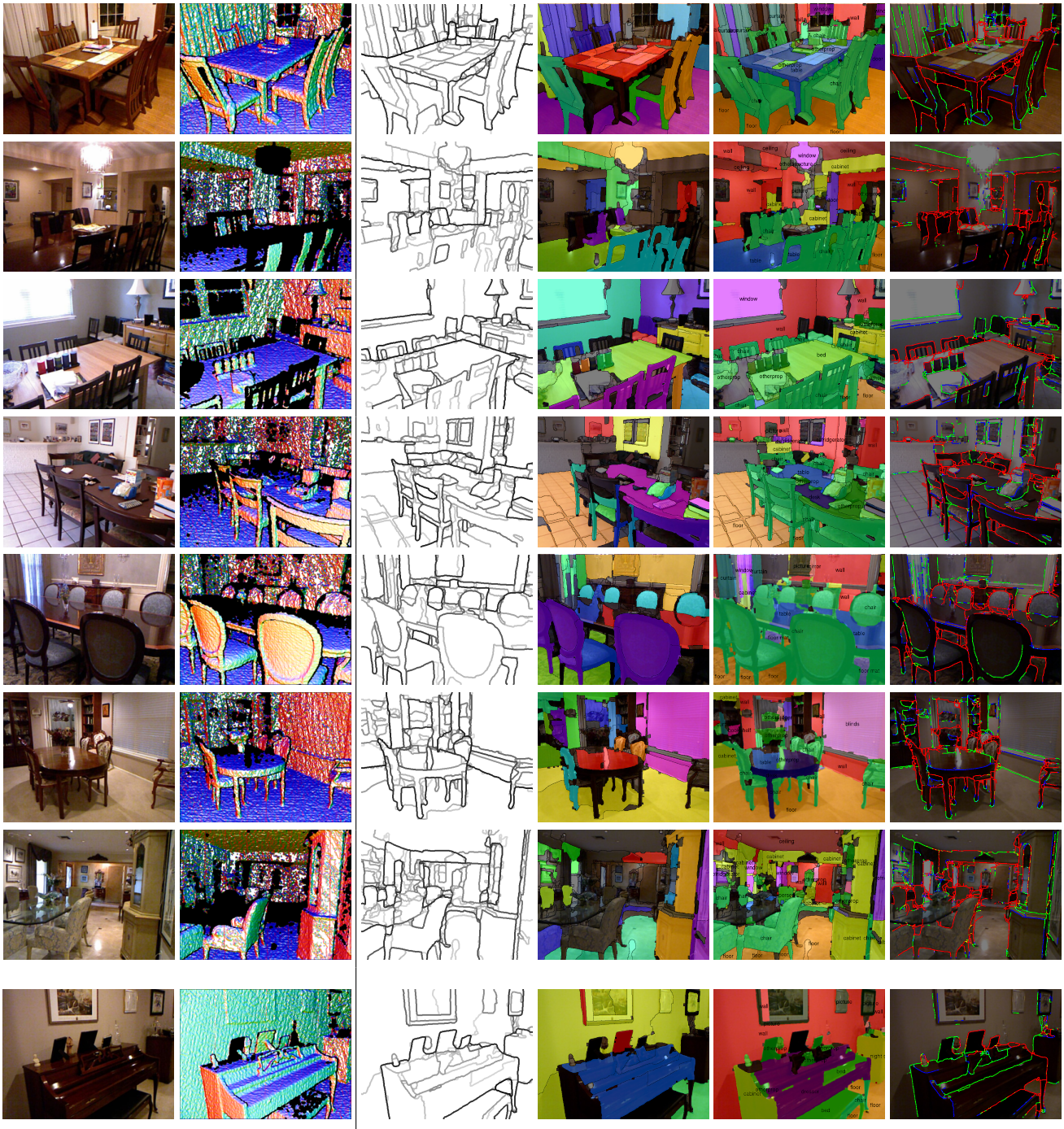
(a) Color Image (b) Depth (c) Contours (d) Amodal Completion (e) Semantic Segmentation (f) Contour Classification

Figure 1. **Output of our system on Bedroom scenes**:We take in as input a single color and depth image and produce as output bottom up segmentation, long range completion, semantic segmentation and contour classification (into depth discontinuities (red), concave surface normal discontinuities (green) and convex surface normal discontinuities (blue)).



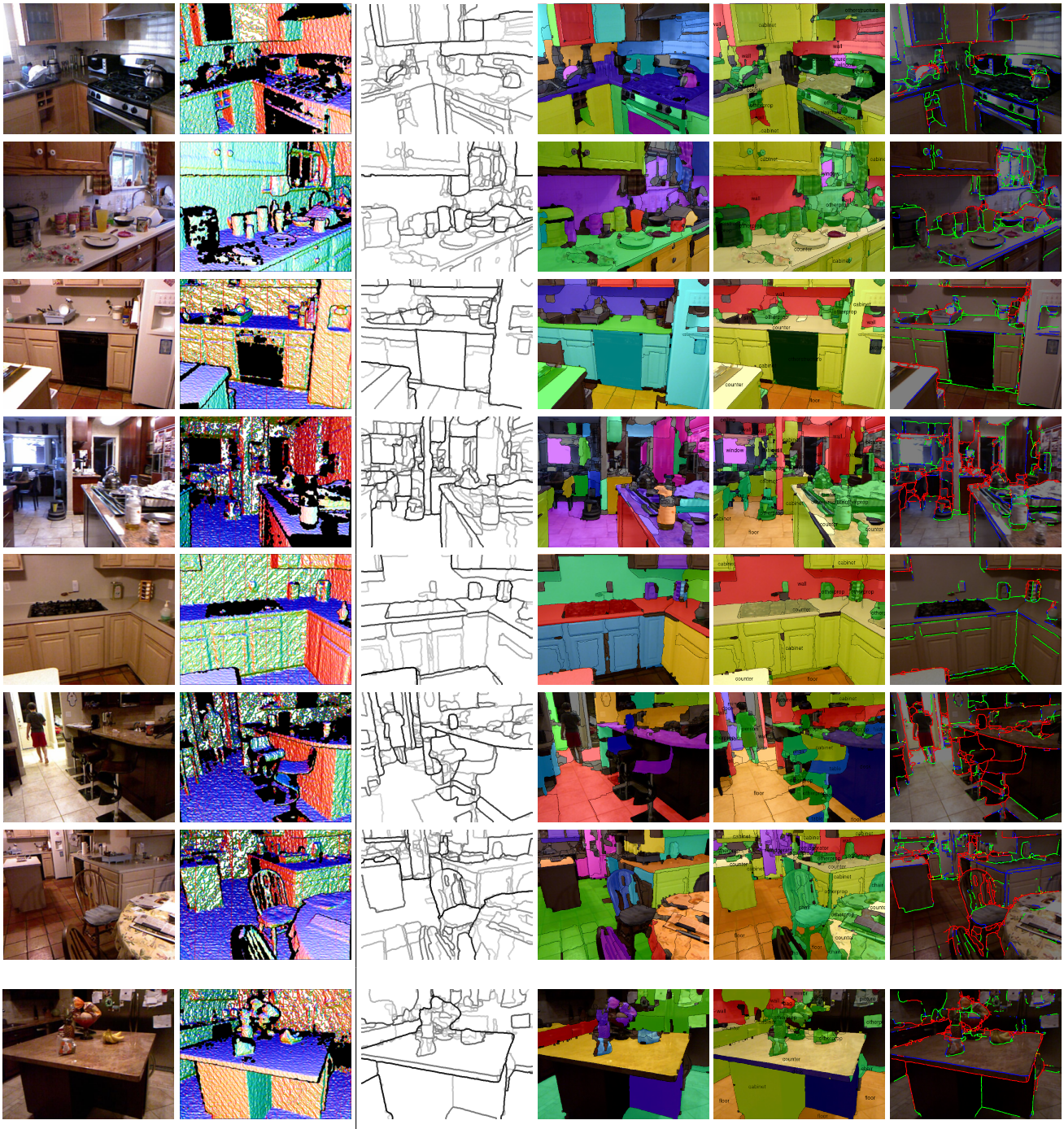
(a) Color Image (b) Depth (c) Contours (d) Amodal Completion (e) Semantic Segmentation (f) Contour Classification

Figure 2. **Output of our system on Living Room scenes**:We take in as input a single color and depth image and produce as output bottom up segmentation, long range completion, semantic segmentation and contour classification (into depth discontinuities (red), concave surface normal discontinuities (green) and convex surface normal discontinuities (blue)).



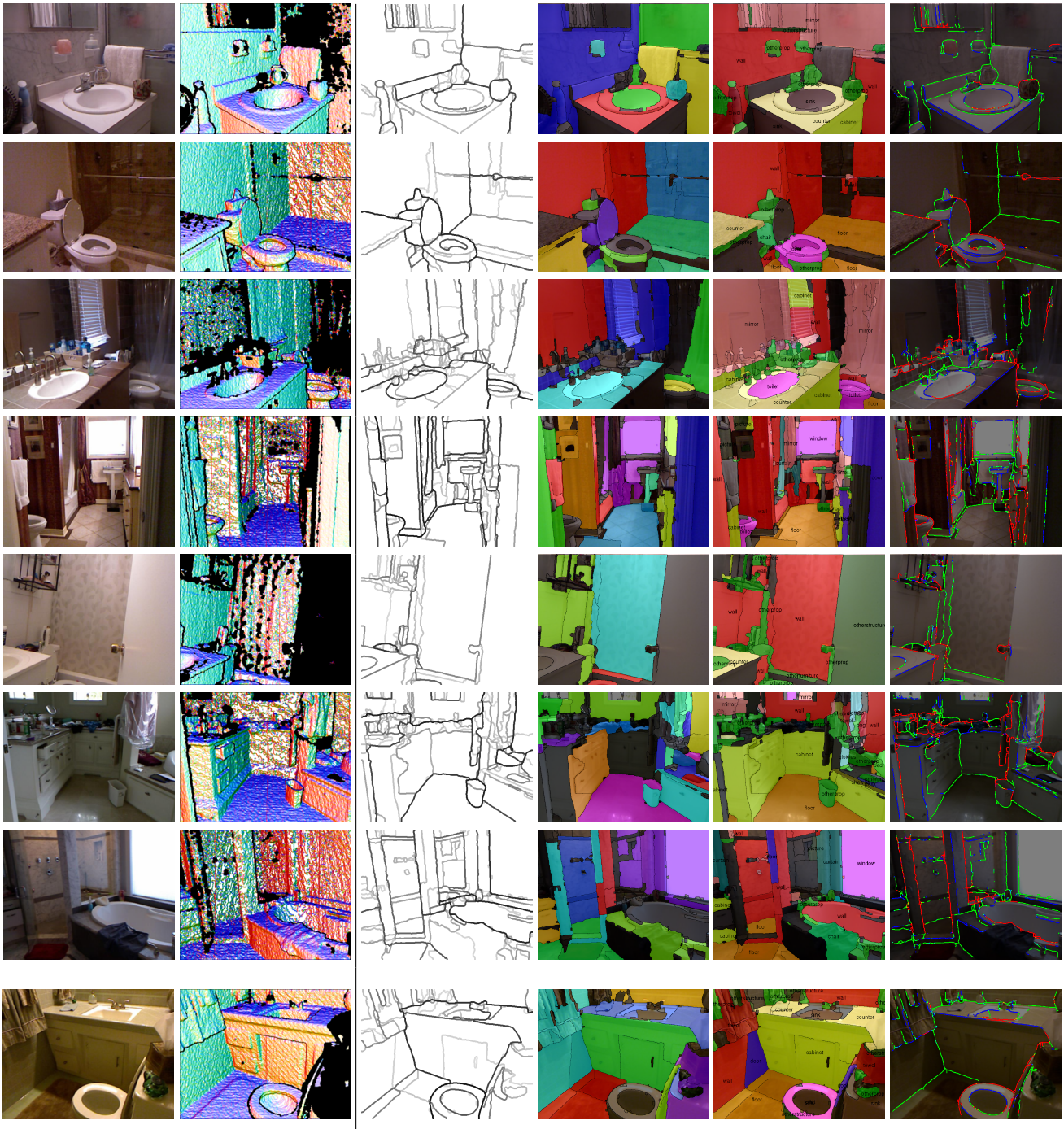
(a) Color Image (b) Depth (c) Contours (d) Amodal Completion (e) Semantic Segmentation (f) Contour Classification

Figure 3. **Output of our system on Dining Room scenes**:We take in as input a single color and depth image and produce as output bottom up segmentation, long range completion, semantic segmentation and contour classification (into depth discontinuities (red), concave surface normal discontinuities (green) and convex surface normal discontinuities (blue)).



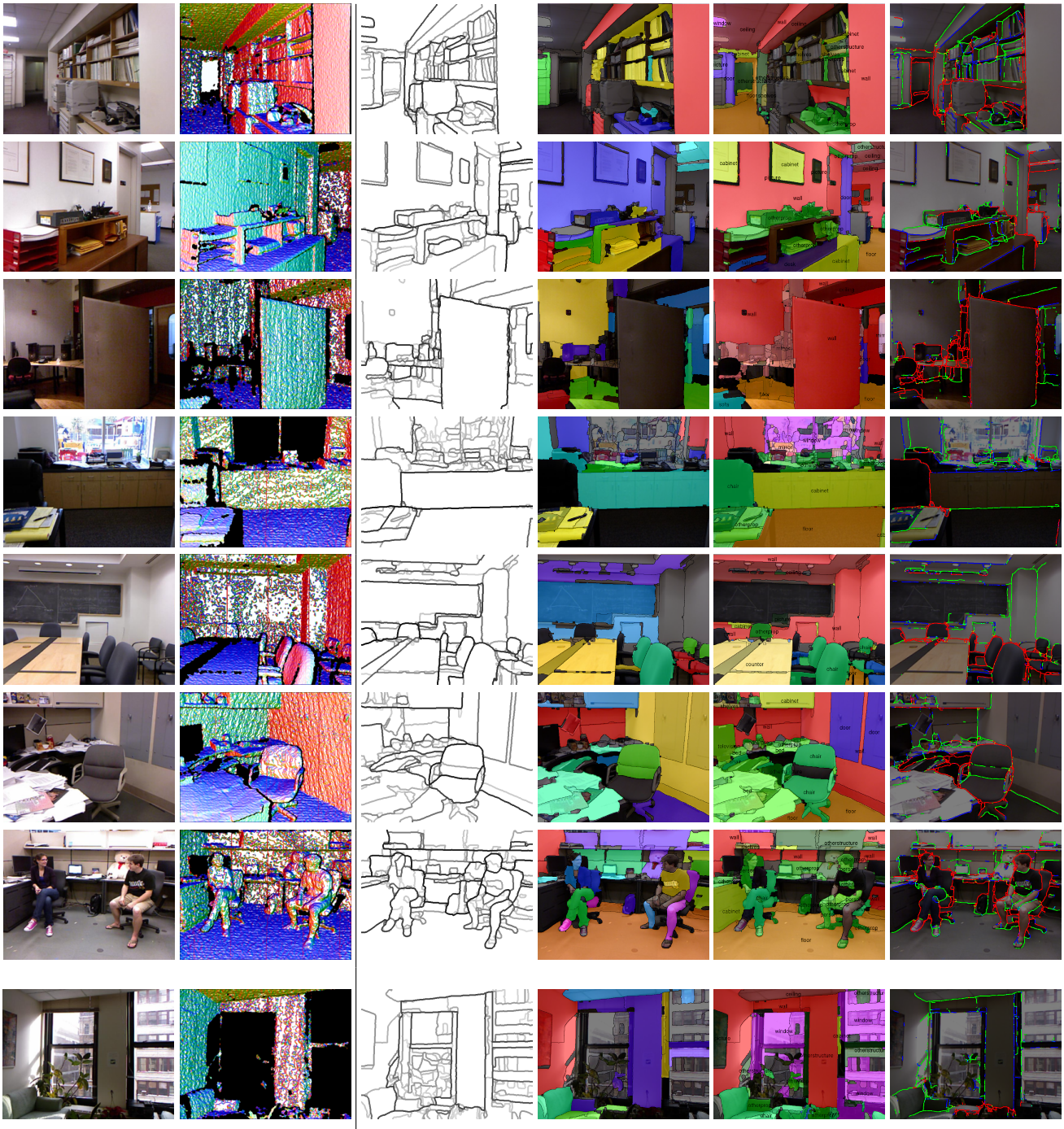
(a) Color Image (b) Depth (c) Contours (d) Amodal Completion (e) Semantic Segmentation (f) Contour Classification

Figure 4. **Output of our system on Kitchen scenes**:We take in as input a single color and depth image and produce as output bottom up segmentation, long range completion, semantic segmentation and contour classification (into depth discontinuities (red), concave surface normal discontinuities (green) and convex surface normal discontinuities (blue)).



(a) Color Image (b) Depth (c) Contours (d) Amodal Completion (e) Semantic Segmentation (f) Contour Classification

Figure 5. **Output of our system on Bathroom scenes**:We take in as input a single color and depth image and produce as output bottom up segmentation, long range completion, semantic segmentation and contour classification (into depth discontinuities (red), concave surface normal discontinuities (green) and convex surface normal discontinuities (blue)).



(a) Color Image (b) Depth (c) Contours (d) Amodal Completion (e) Semantic Segmentation (f) Contour Classification

Figure 6. **Output of our system on Office scenes**:We take in as input a single color and depth image and produce as output bottom up segmentation, long range completion, semantic segmentation and contour classification (into depth discontinuities (red), concave surface normal discontinuities (green) and convex surface normal discontinuities (blue)).

(a) Color Image (b) Depth (c) Contours (d) Amodal Completion (e) Semantic Segmentation (f) Contour Classification

Figure 7. **Output of our system on Classroom scenes**:We take in as input a single color and depth image and produce as output bottom up segmentation, long range completion, semantic segmentation and contour classification (into depth discontinuities (red), concave surface normal discontinuities (green) and convex surface normal discontinuities (blue)).

(a) Color Image (b) Depth (c) Contours (d) Amodal Completion (e) Semantic Segmentation (f) Contour Classification

Figure 8. **Output of our system on Other scenes**:We take in as input a single color and depth image and produce as output bottom up segmentation, long range completion, semantic segmentation and contour classification (into depth discontinuities (red), concave surface normal discontinuities (green) and convex surface normal discontinuities (blue)).