Supplemental materials for “ImageSpirit: Verbal Guided Image Parsing”

Ming-Ming Cheng[1] Shuai Zheng[1*] Wen-Yan Lin[2] Vibhav Vineet[2] Paul Sturgess[2]
Nigel Crook[2] Niloy J. Mitra[3] Philip Torr[1]
[1]University of Oxford [2]Oxford Brookes University [3]University College London

CR Categories: I.3.6 [Computer Graphics]: Methodology and Techniques—Interaction Techniques I.4.8 [Image Processing and Computer Vision]: Scene Analysis—Object Recognition

Keywords: Image manipulation, speech control, object class segmentation, semantic attributes.

Links: DL PDF WEB VIDEO DATA CODE

1 Image parsing results

To evaluate per-pixel object class prediction results, we compared our scene parsing results [Cheng et al. 2014] with H-CRF [Ladicky et al. 2009] and DenseCRF [Krähenbühl and Koltun 2011] on aNYU dataset. The statistics in the main paper show that our method produces more accurate object class prediction results. Visual comparisons for all results of 725 testing images in aNYU dataset are show in Fig. 6-36. Besides more accurate object class prediction, our method also produces attribute predictions which are important for using verbal commands to improve scene parsing. Verbal guided image parsing results are shown for both aNUY dataset (Fig. 1-3) as well as Google images (Fig. 4-5).

References

CHENG, M.-M., ZHENG, S., LIN, W.-Y., VINEET, V., STURGESS, P., CROOK, N., MITRA, N., AND TORR, P. 2014. Imagespirit: Verbal guided image parsing. *ACM Transactions on Graphics*.

KRÄHENBÜHL, P., AND KOLTUN, V. 2011. Efficient inference in fully connected CRFs with gaussian edge potentials. In *NIPS*.

LADICKY, L., RUSSELL, C., KOHLI, P., AND TORR, P. H. S. 2009. Associative hierarchical CRFs for object class image segmentation. In *ICCV*.

---

*joint first author.

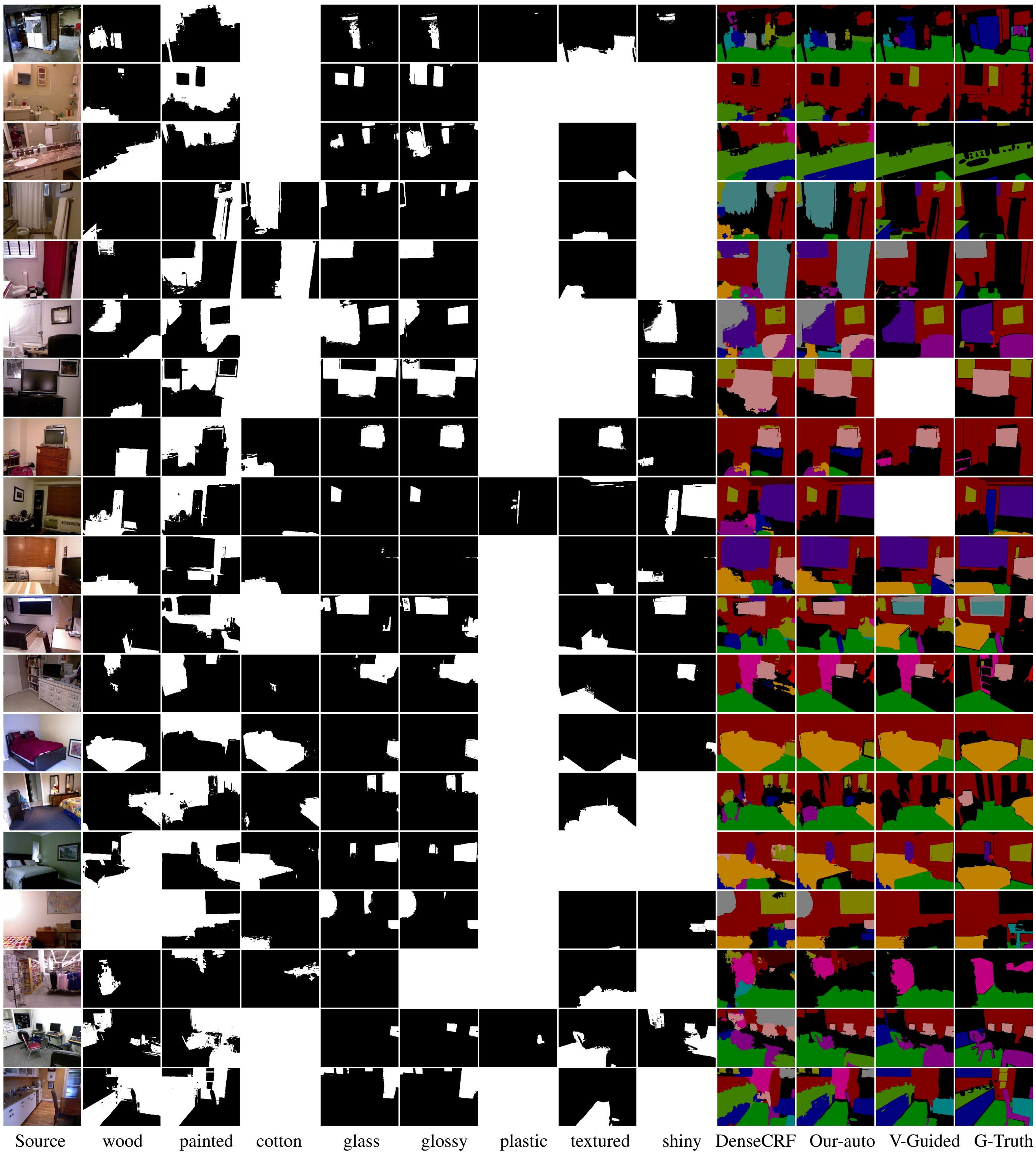


Figure 1: *Compared with state of art methods, our approach gives more accurate per-pixel object classification. More over, even if our automatic attribute prediction results are of low quality, they can still be effectively used to verbally control the image parsing process.*

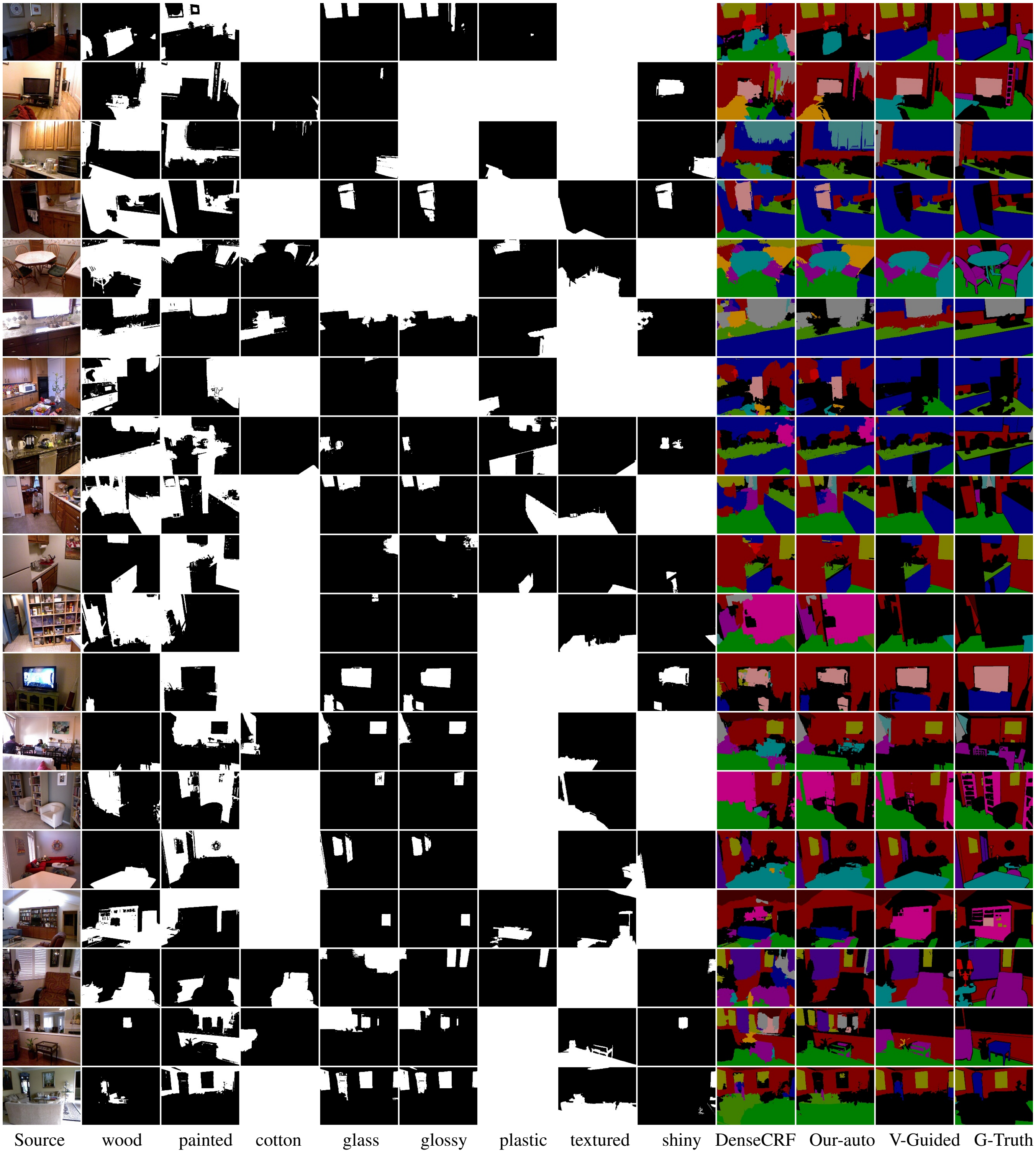


Figure 2: *Compared with state of art methods, our approach gives more accurate per-pixel object classification. More over, even if our automatic attribute prediction results are of low quality, they can still be effectively used to verbally control the image parsing process.*

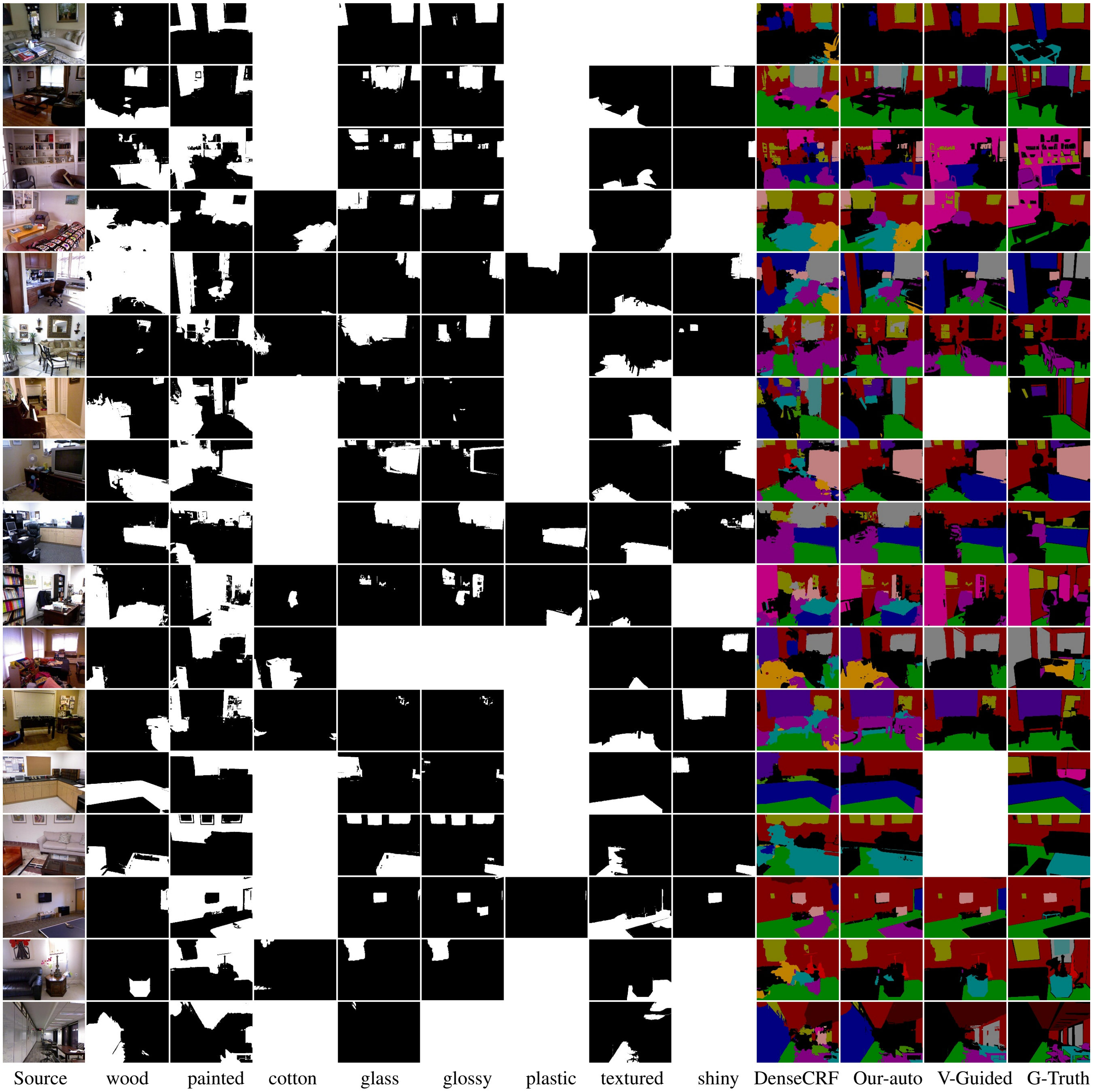
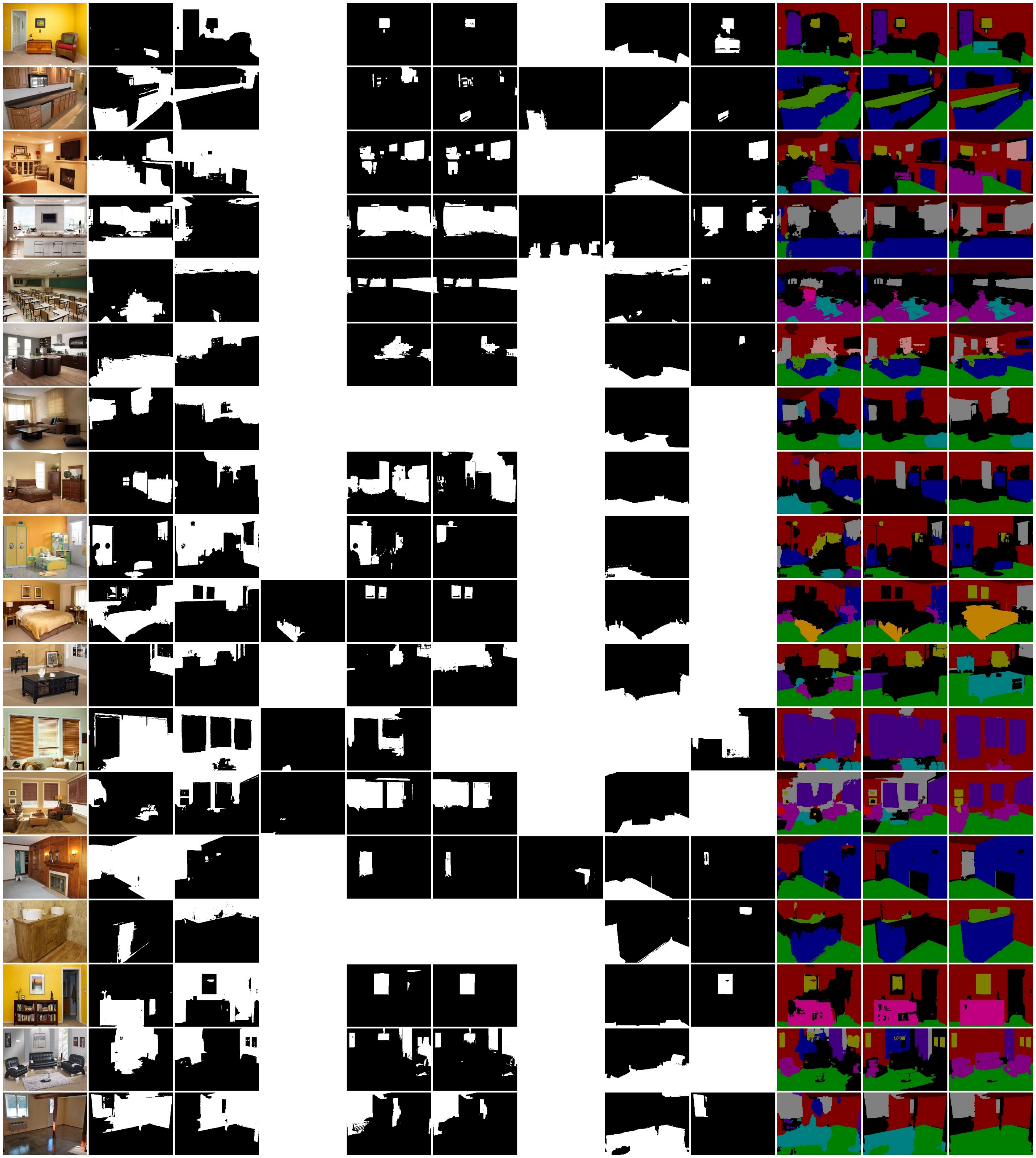


Figure 3: *Compared with state of art methods, our approach gives more accurate per-pixel object classification. More over, even if our automatic attribute prediction results are of low quality, they can still be effectively used to verbally control the image parsing process.*



Source wood painted cotton glass glossy plastic textured shiny DenseCRF Our-auto V-Guided

Figure 4: *Our system that is trained on the aNYU indoor dataset, generalizes to images of similar scene types obtained from Google. Our system allows users to refine this initial results verbally (see V-Guided for results).*

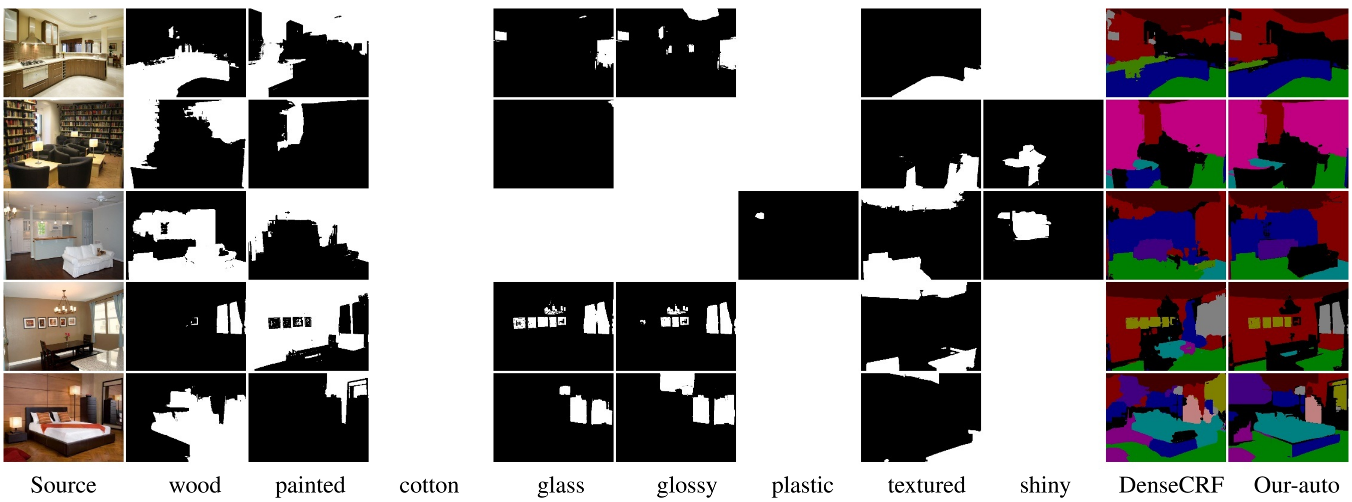
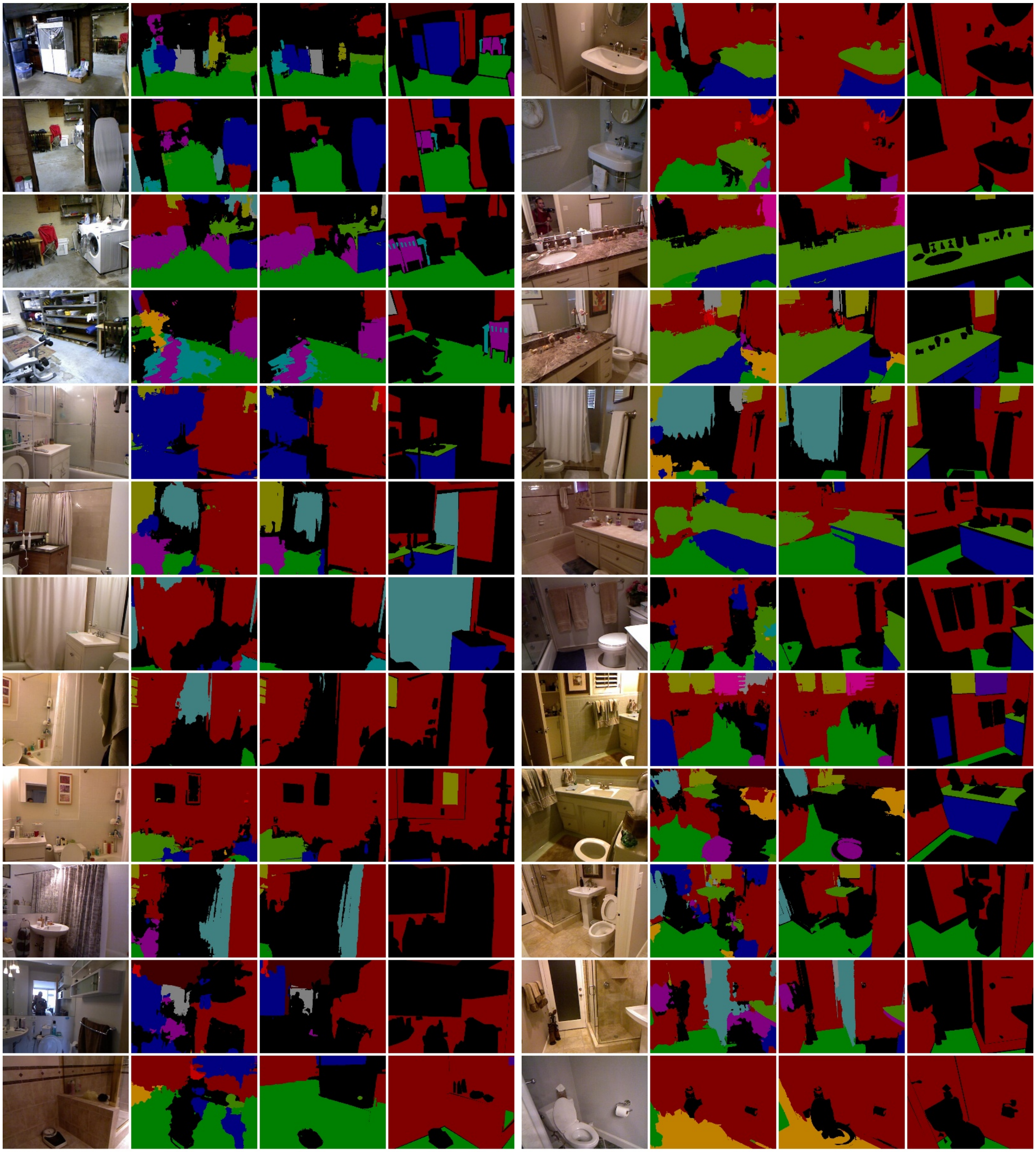


Figure 5: *Google images where verbal parsing is difficult: when there is no suitable configuration of attributes (predicted attributes, color attributes, position attributes, or their combinations) related to the desired object region, verbal interaction does not help.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 6: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

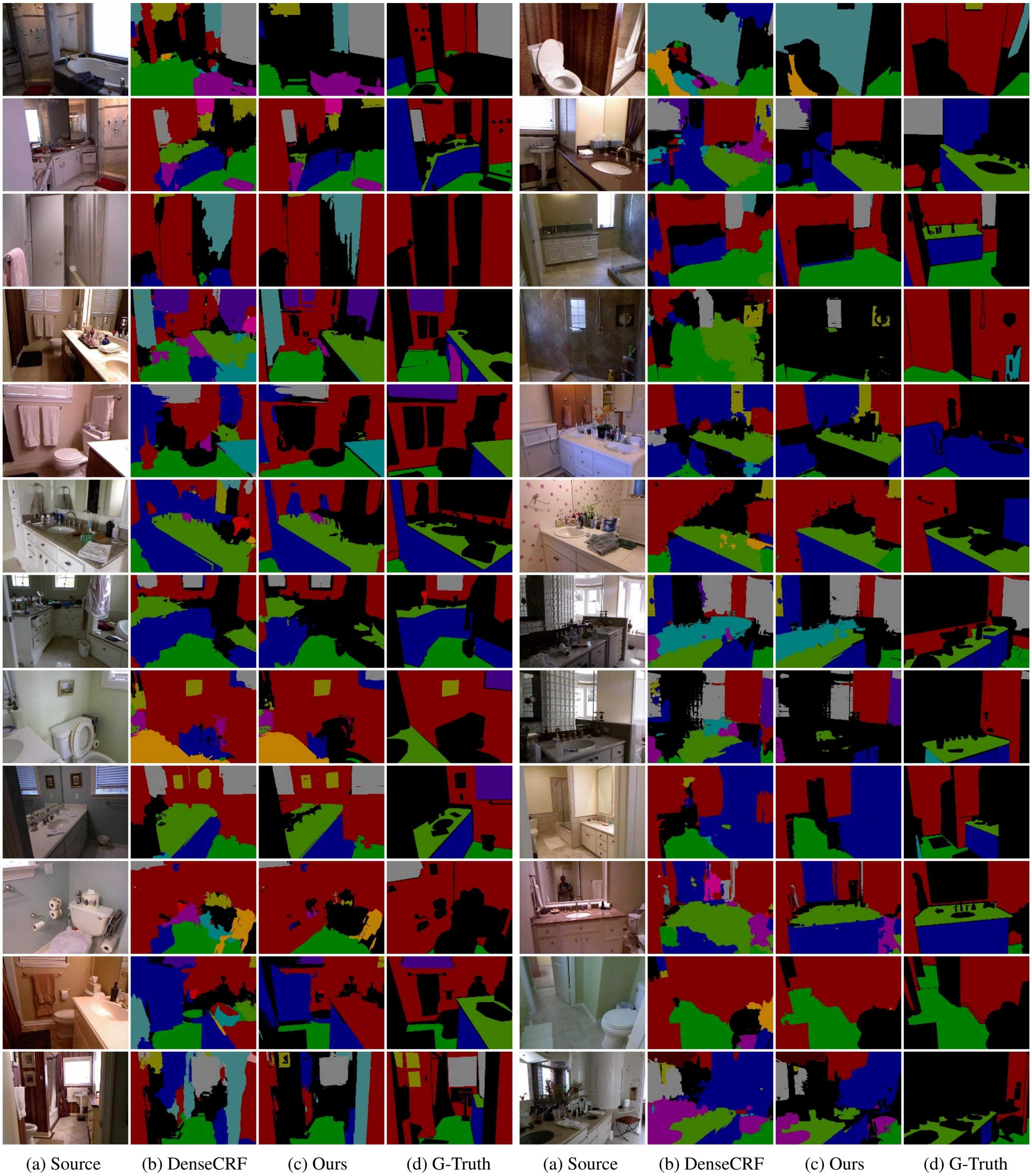
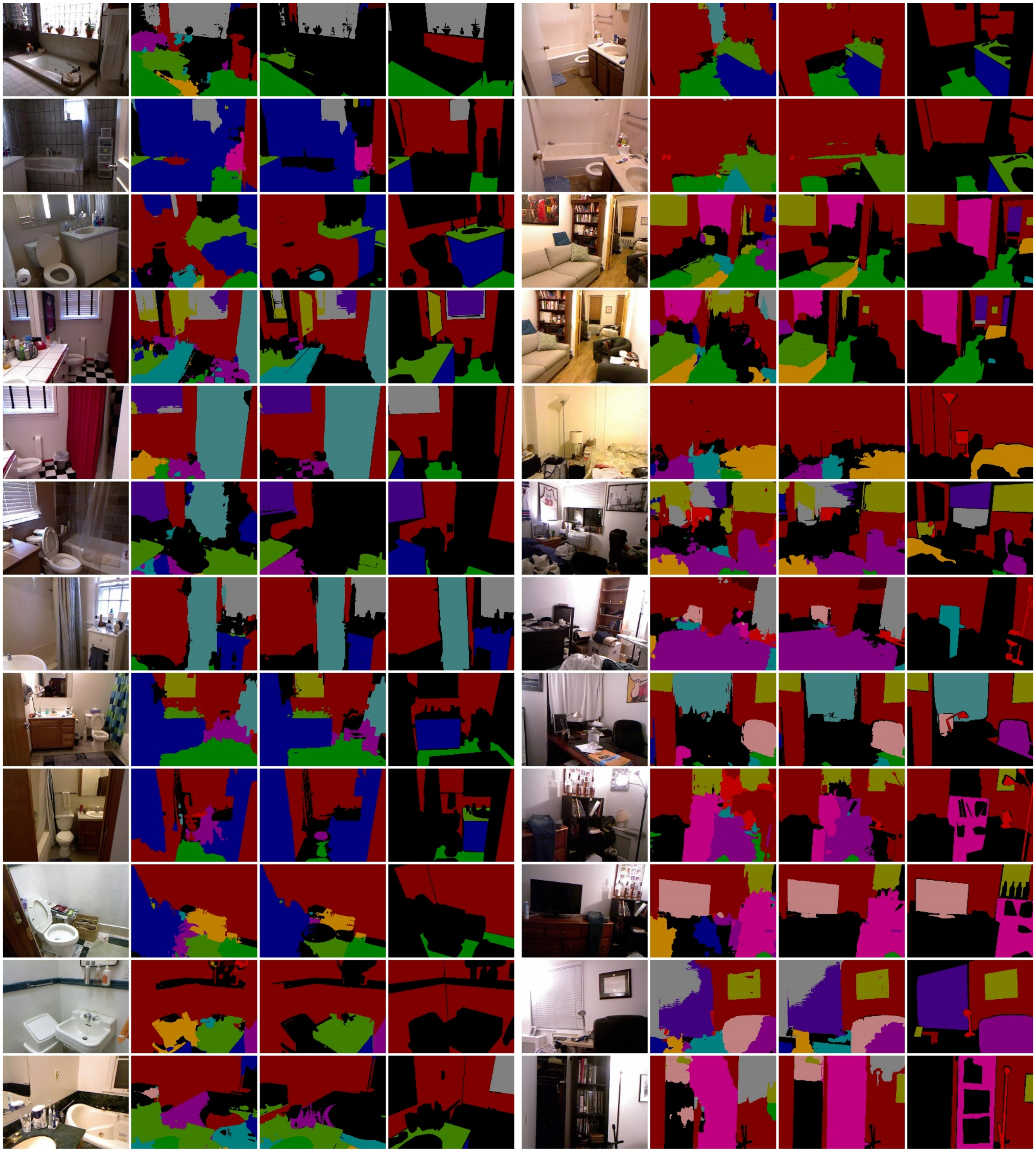
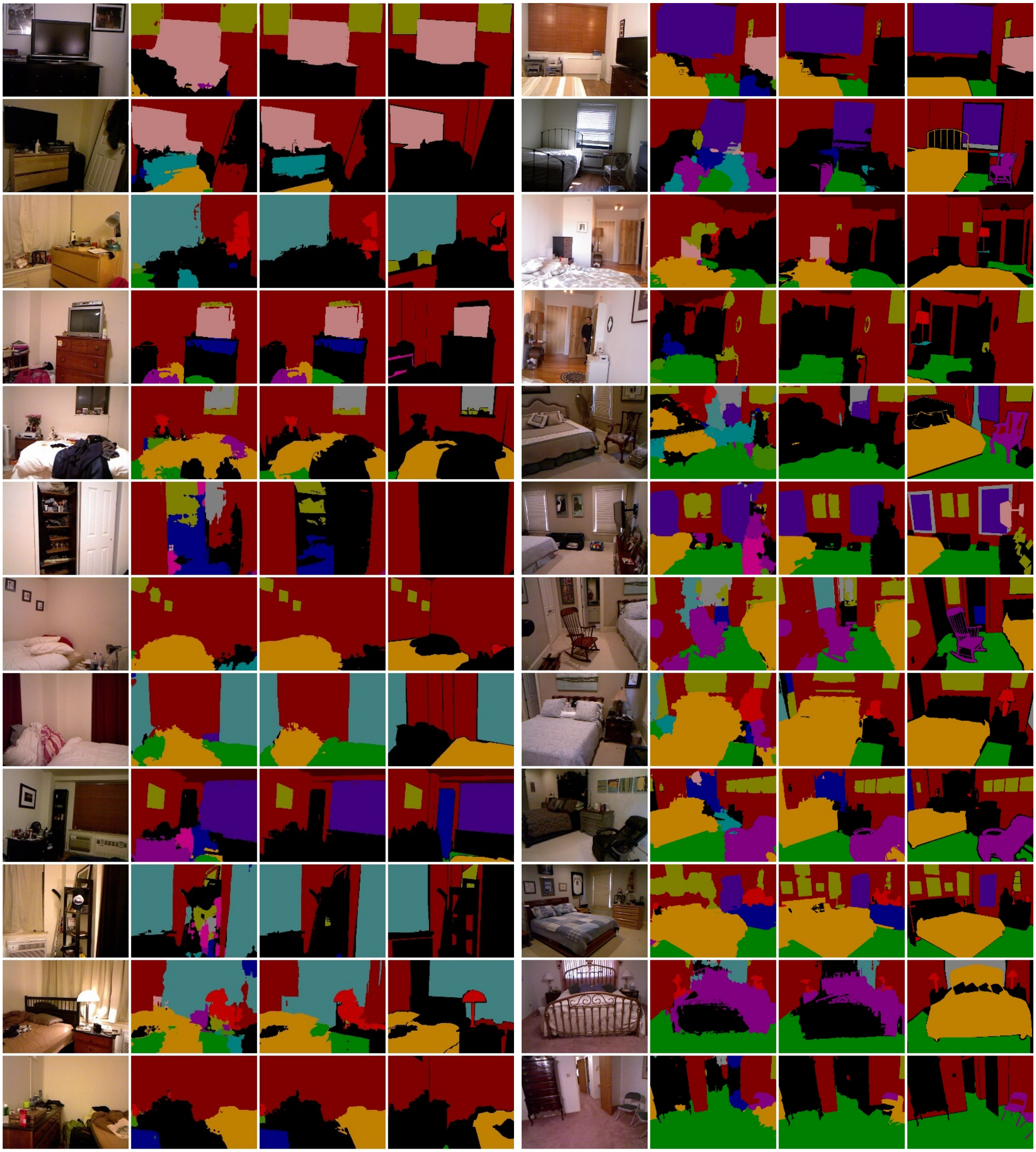


Figure 7: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 8: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 9: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

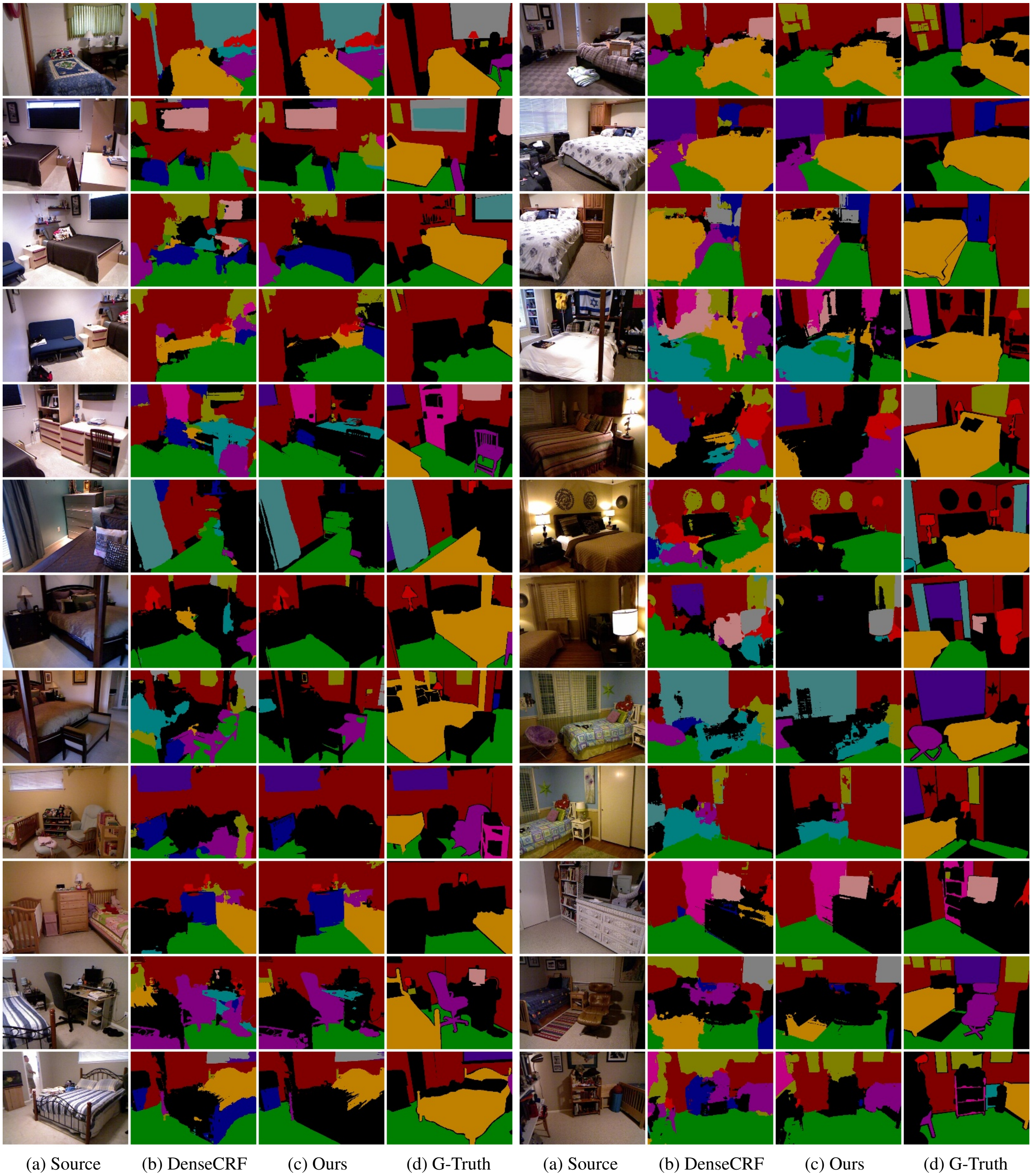
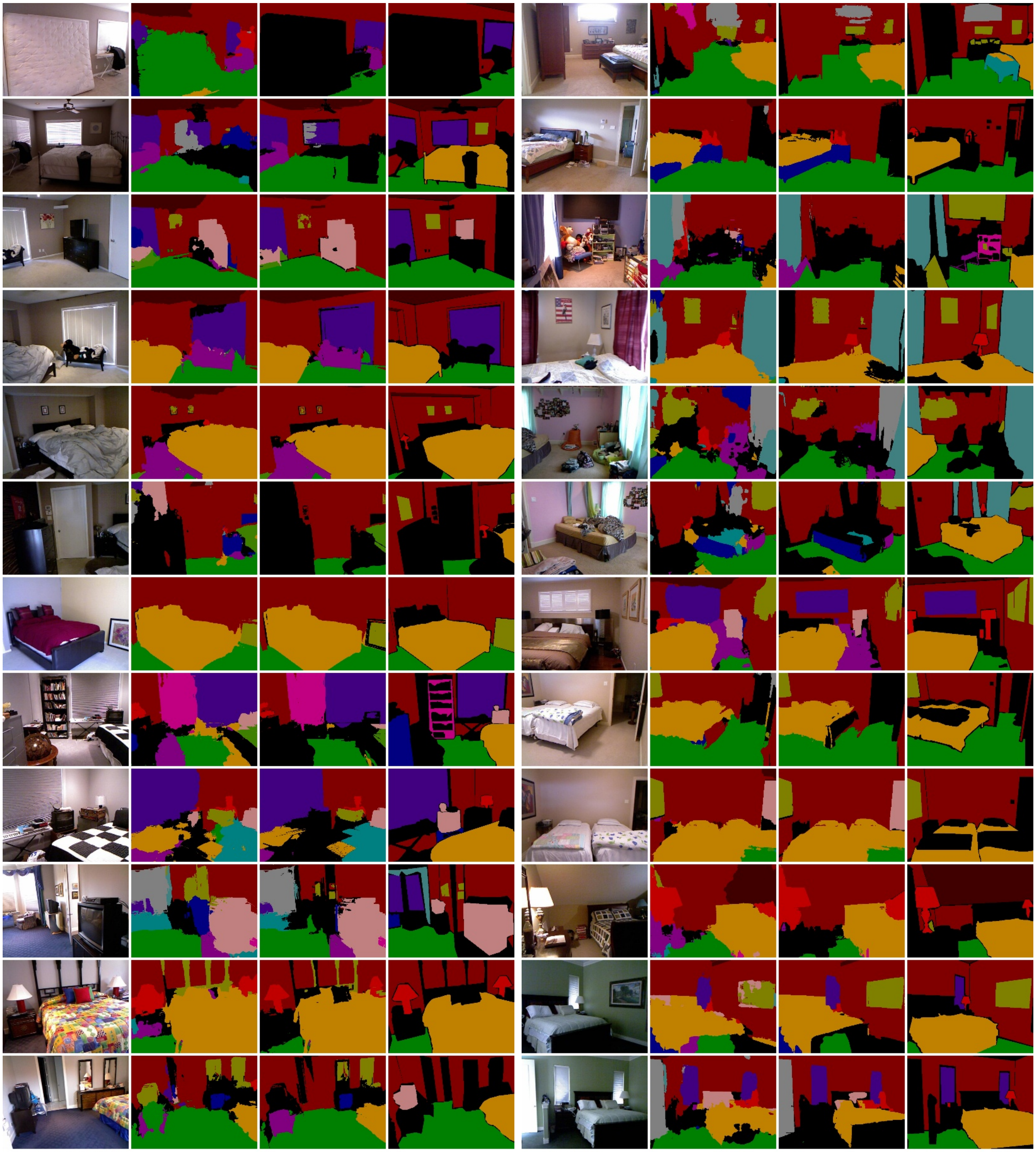


Figure 10: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 11: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

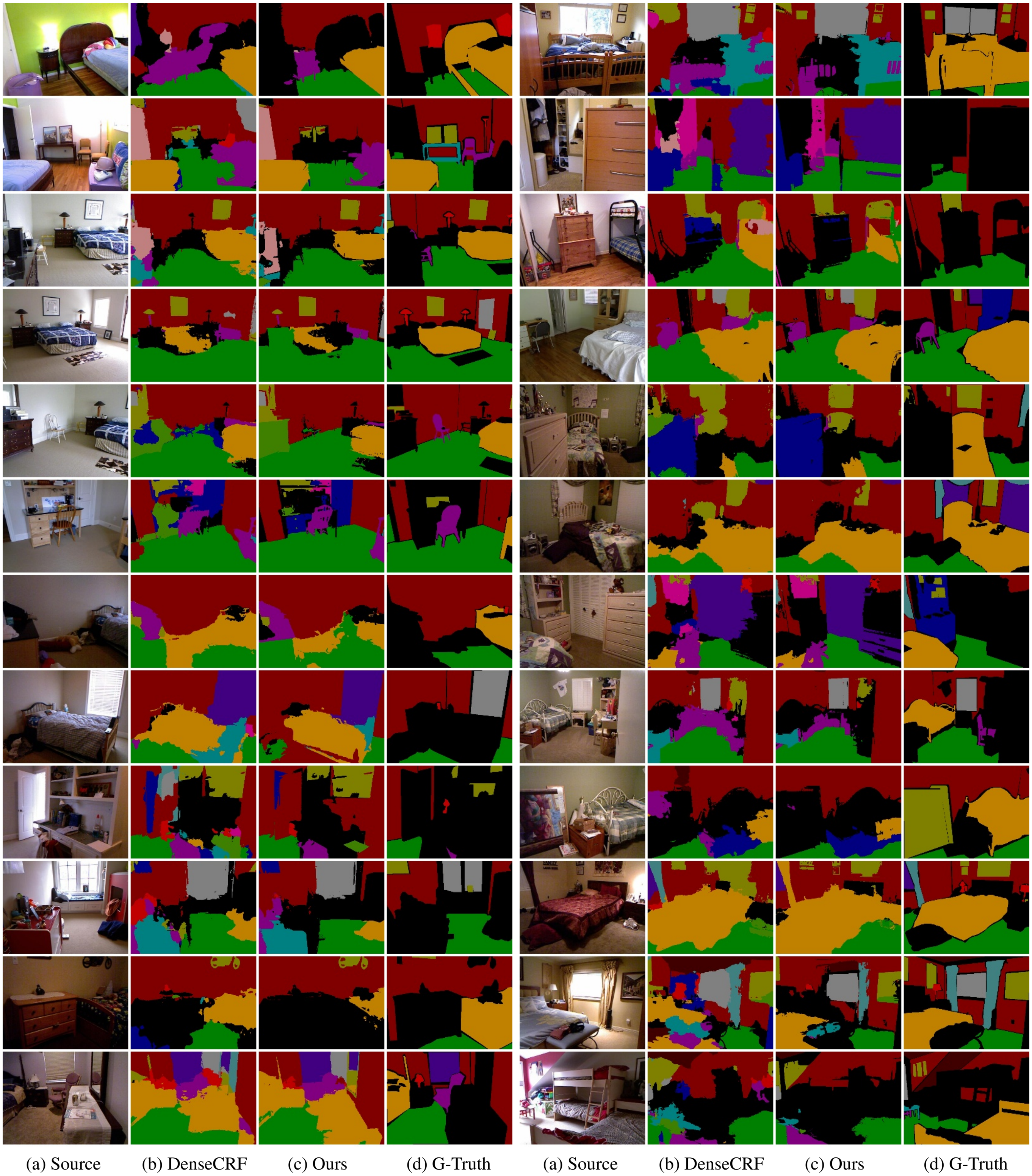


Figure 12: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 13: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

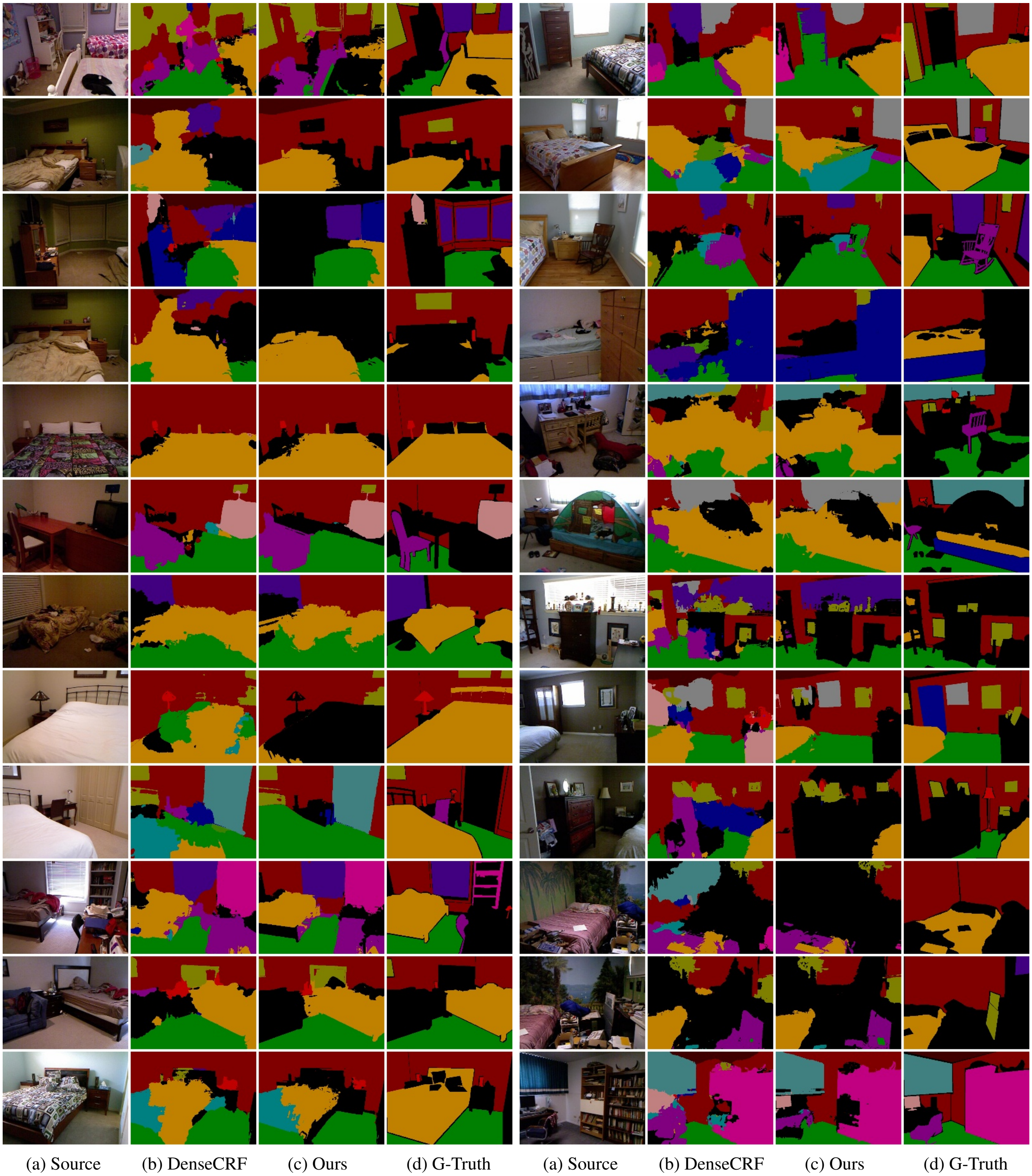
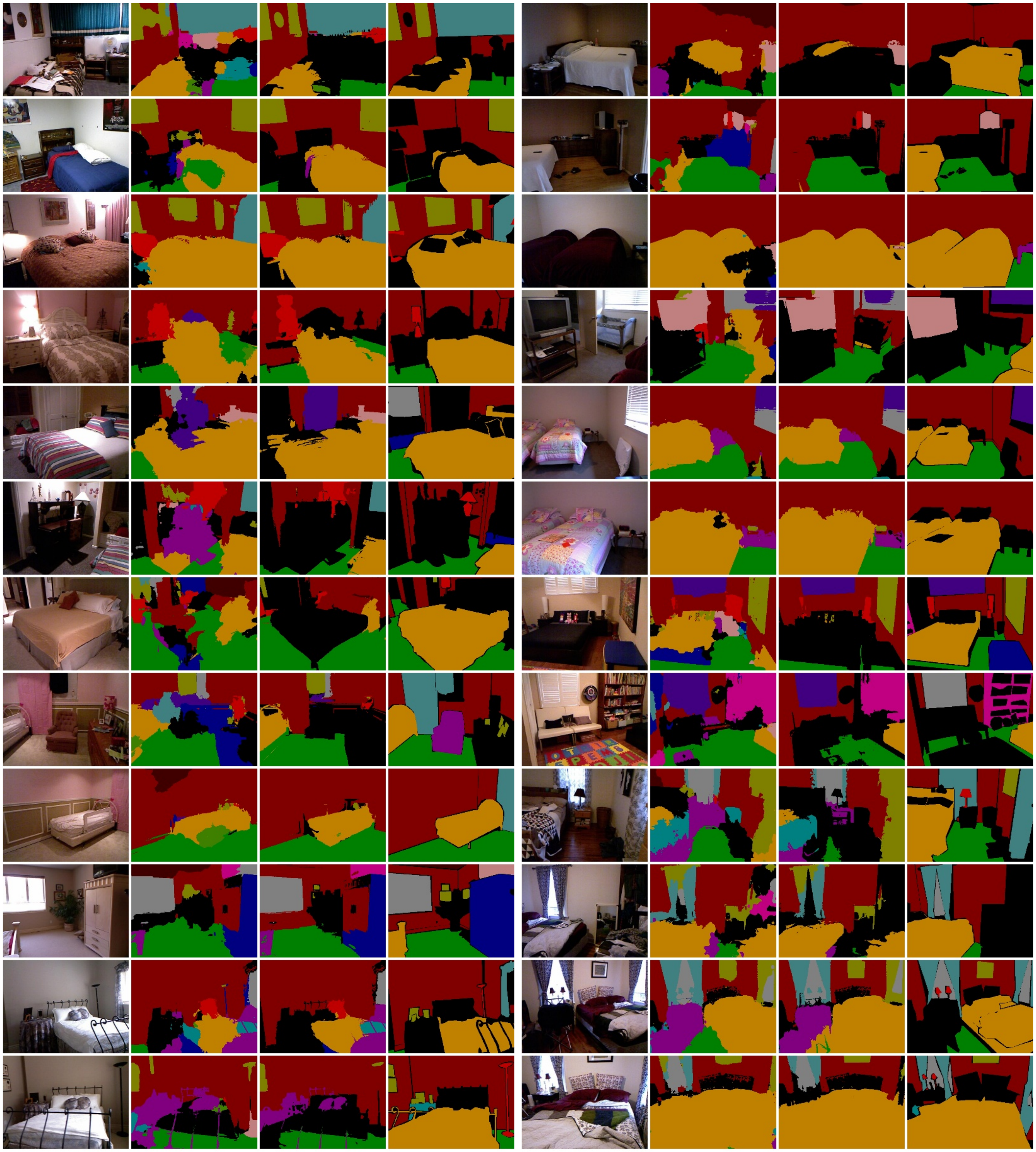


Figure 14: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 15: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

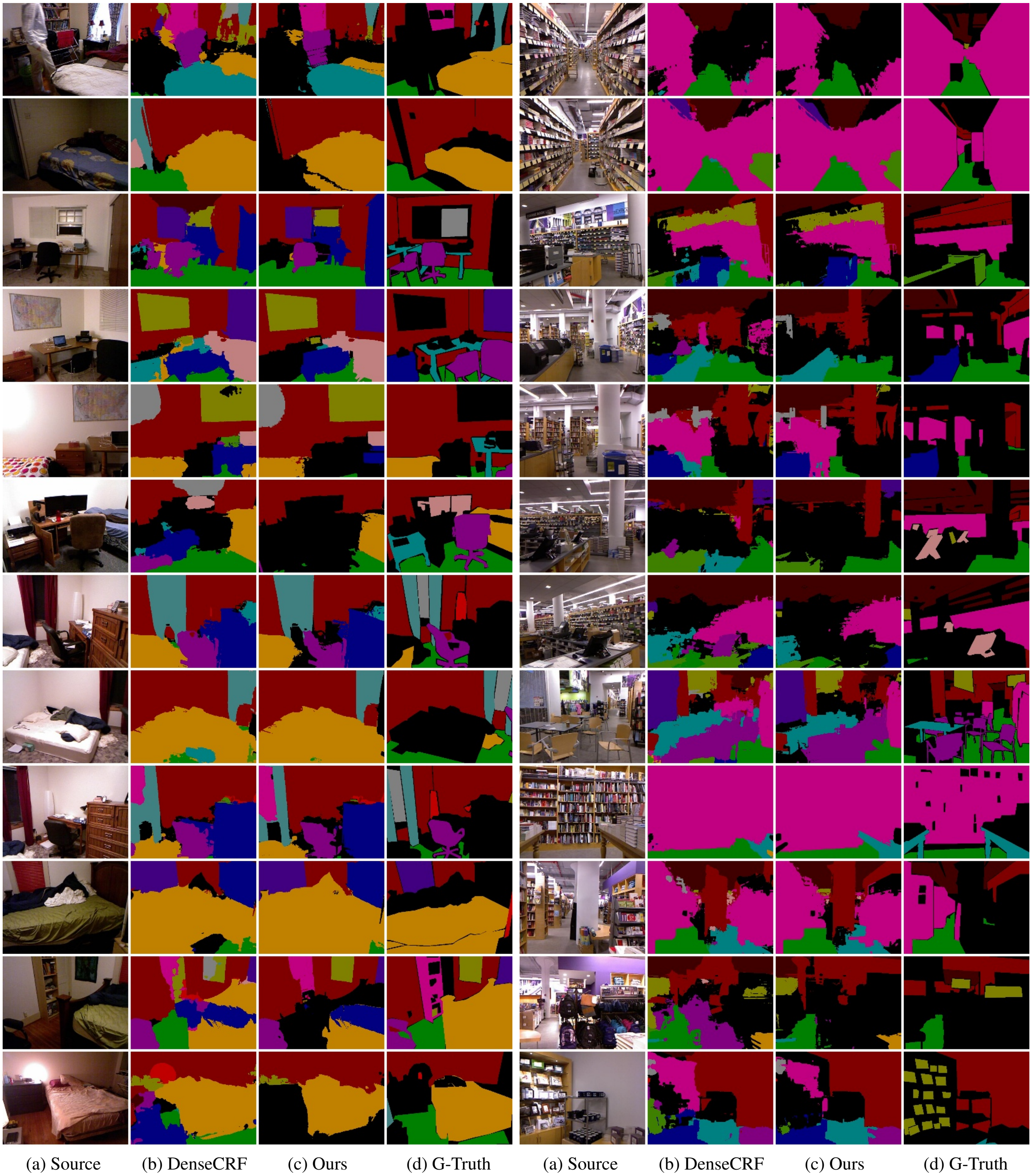


Figure 16: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

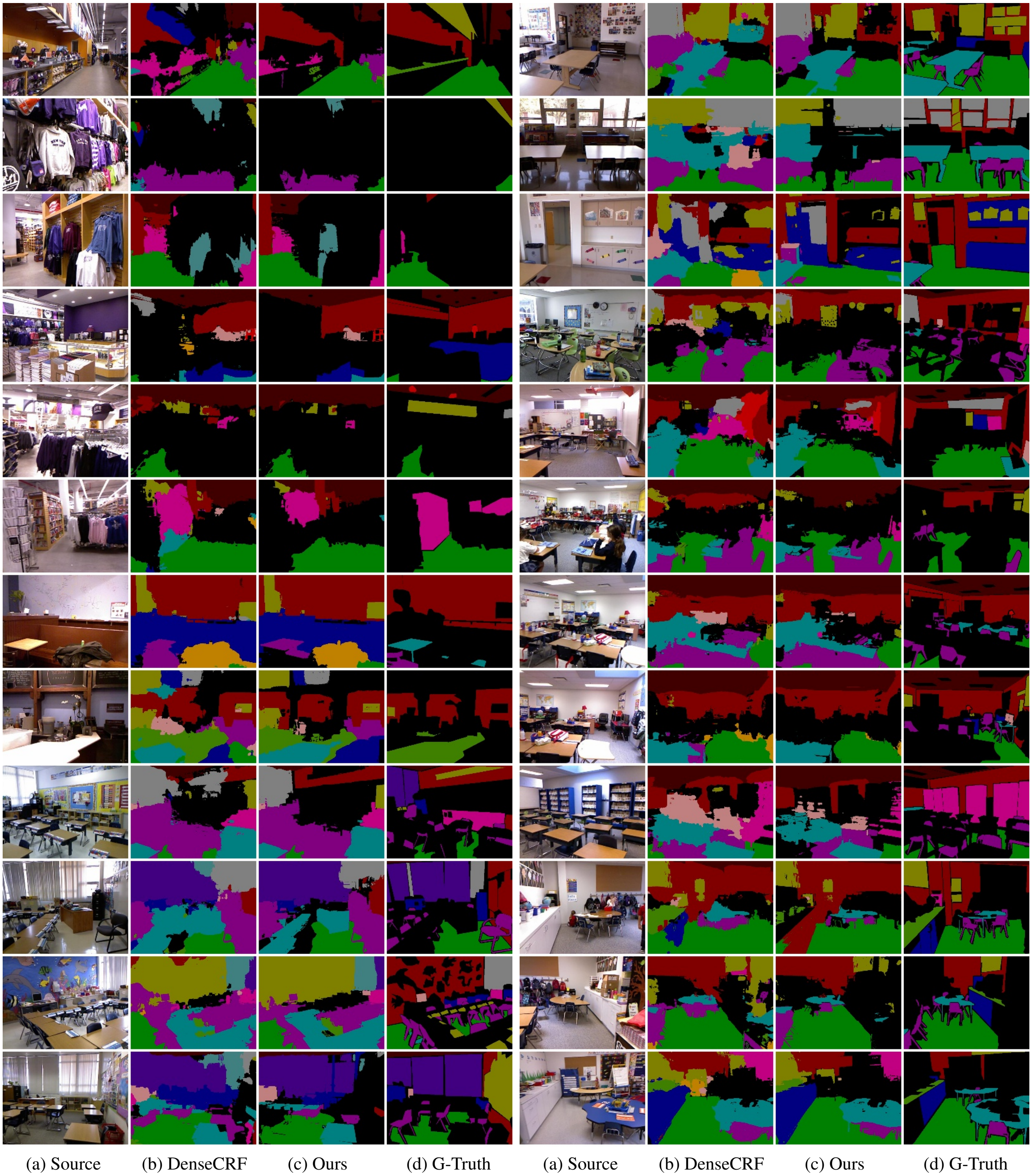
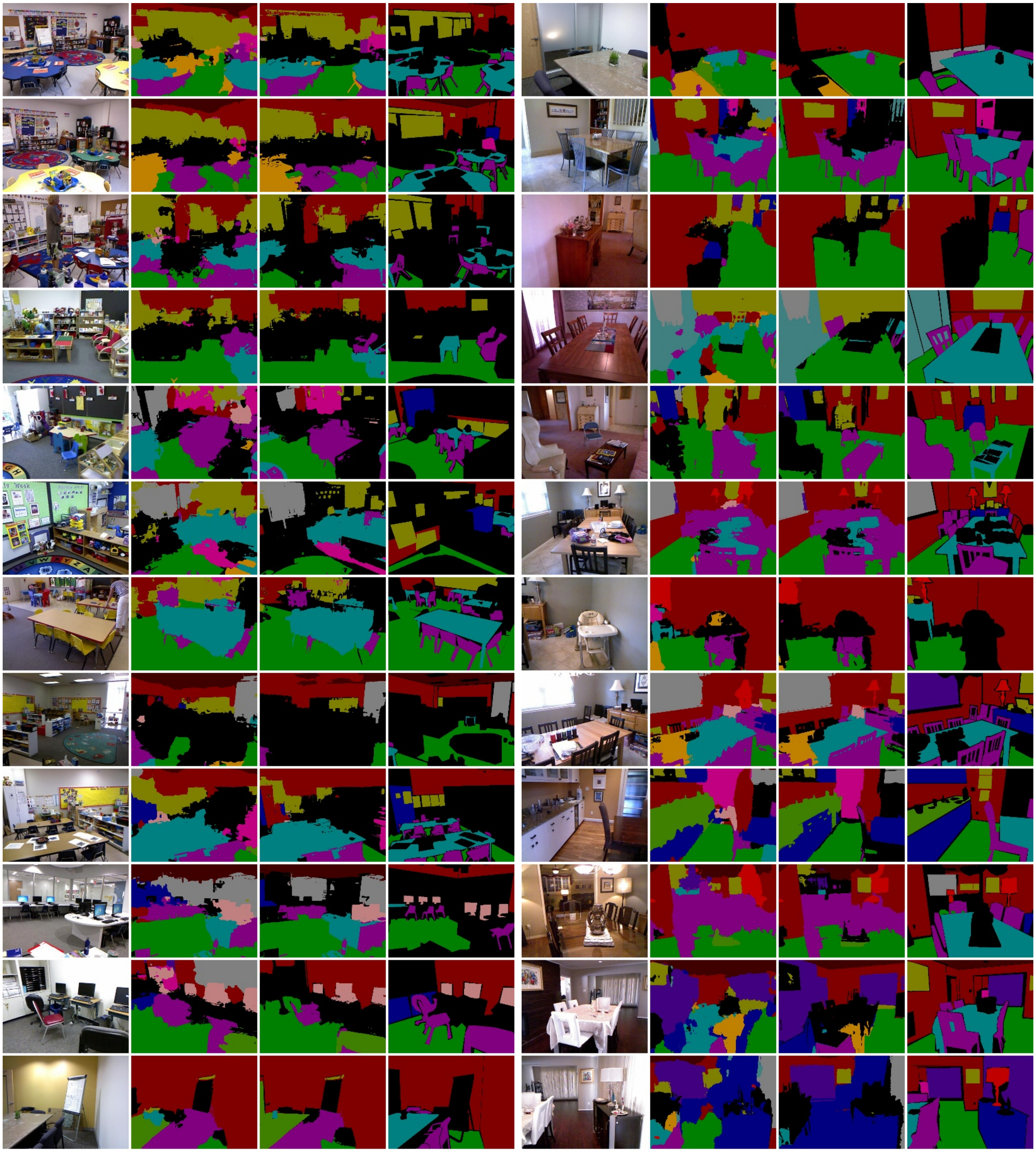


Figure 17: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 18: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

Figure 19: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 20: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

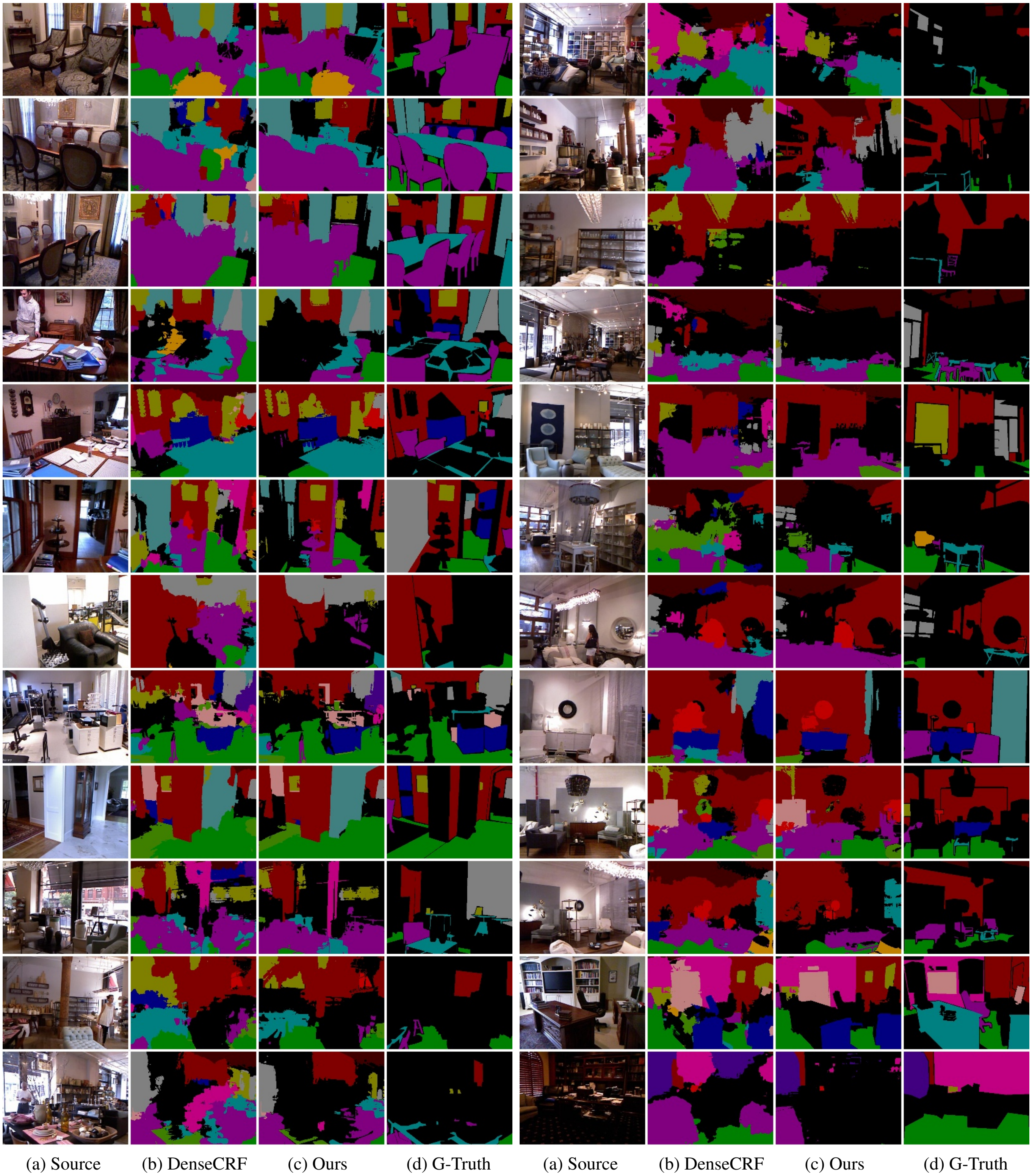
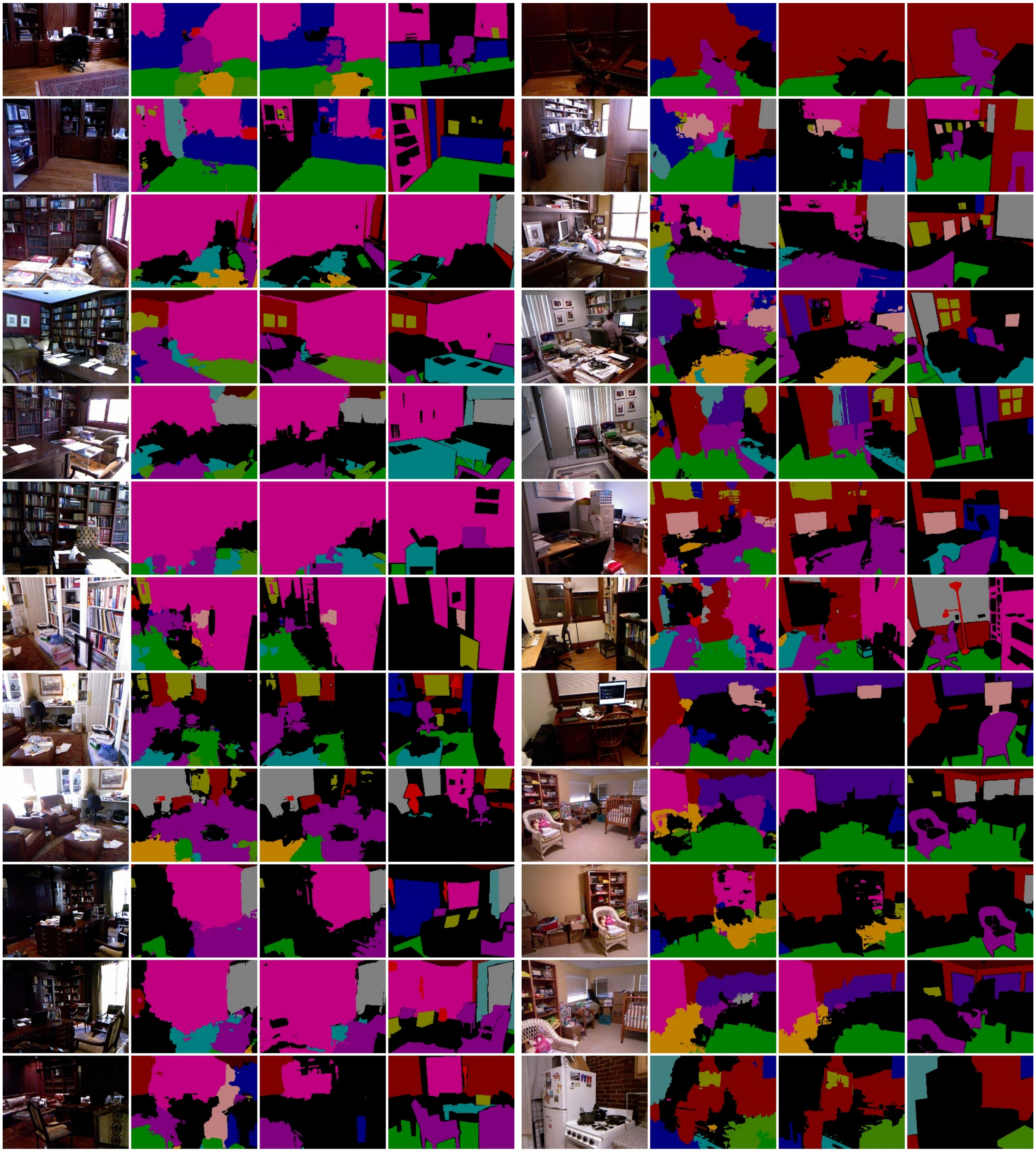


Figure 21: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 22: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

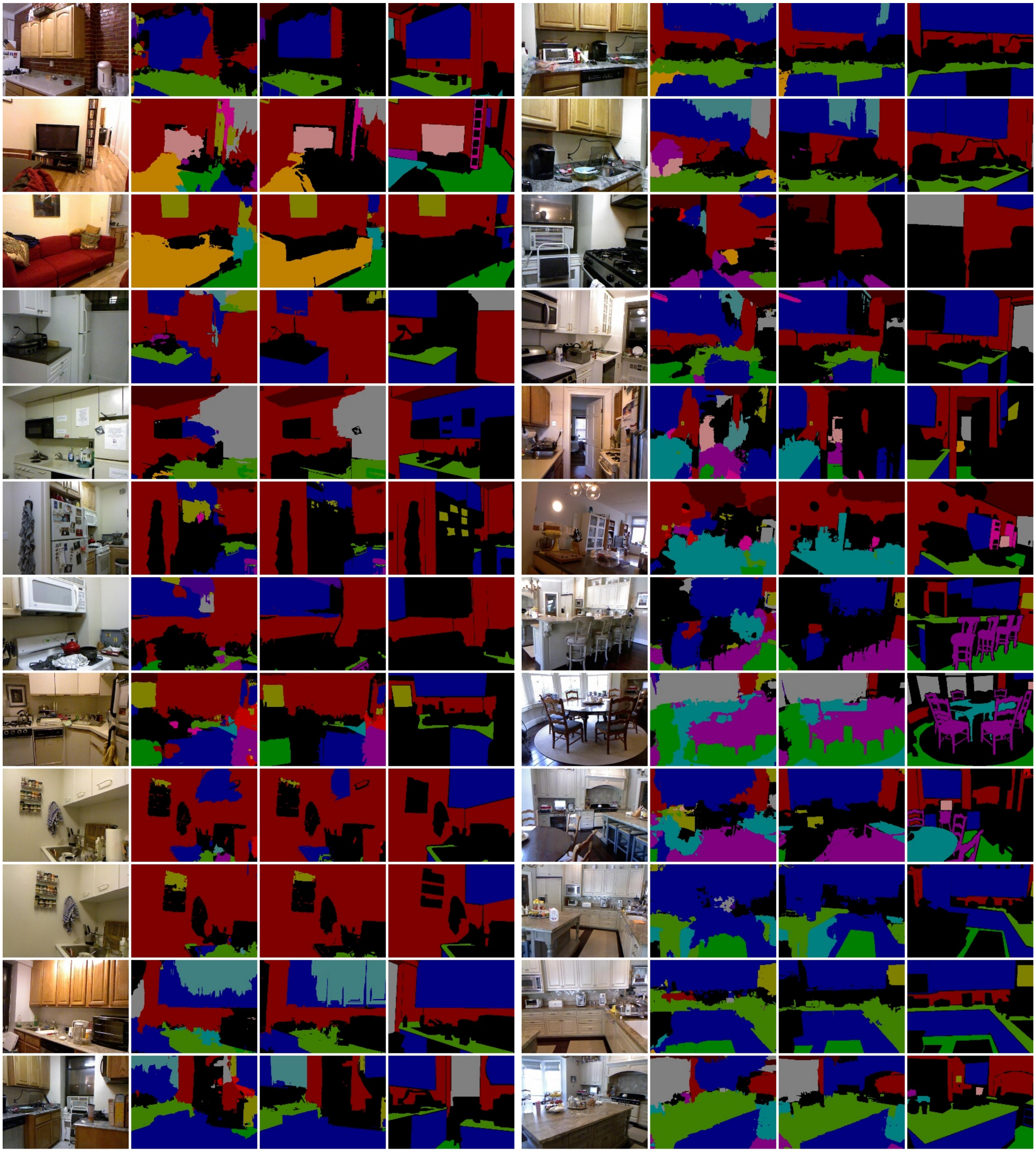
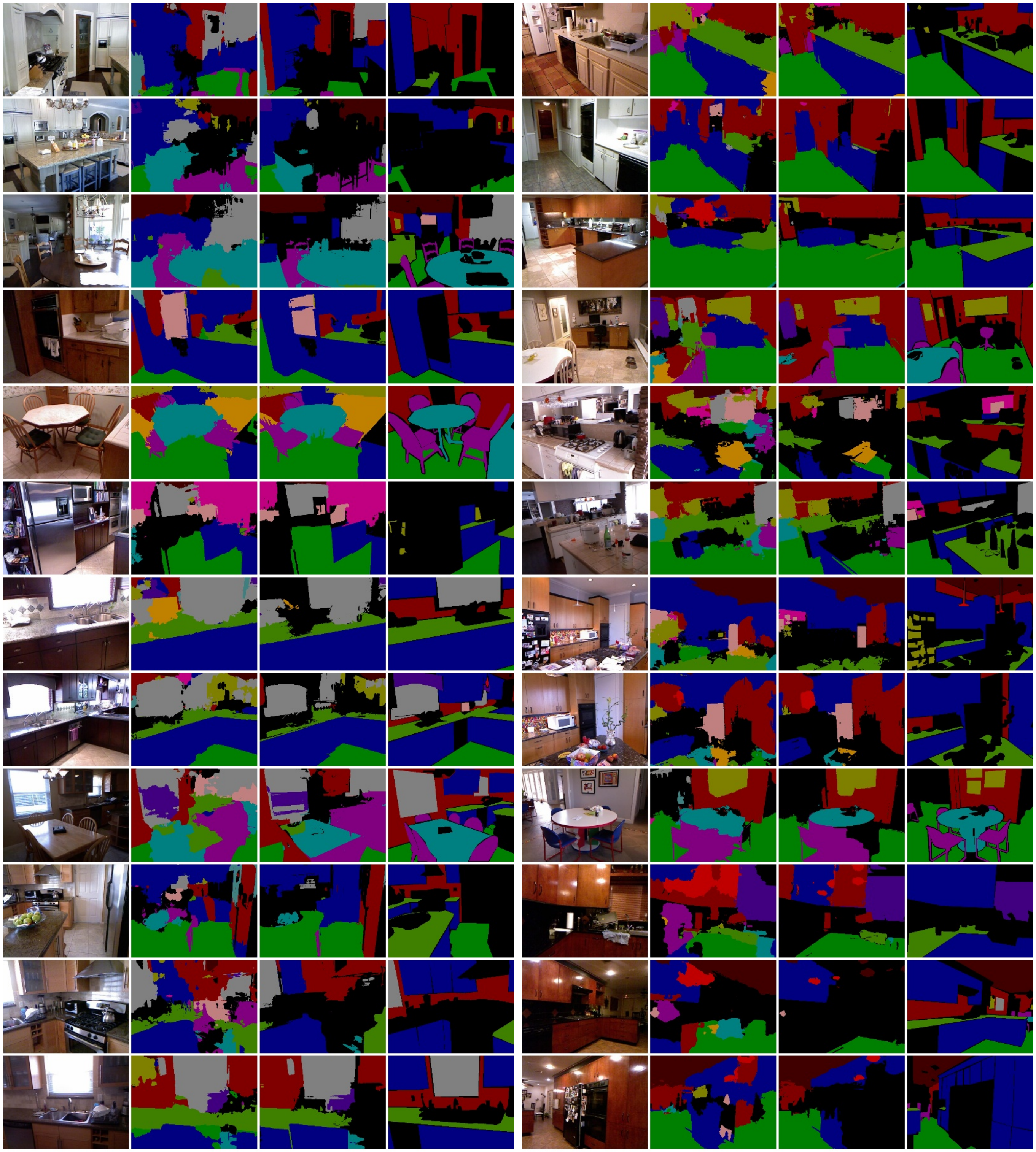


Figure 23: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 24: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

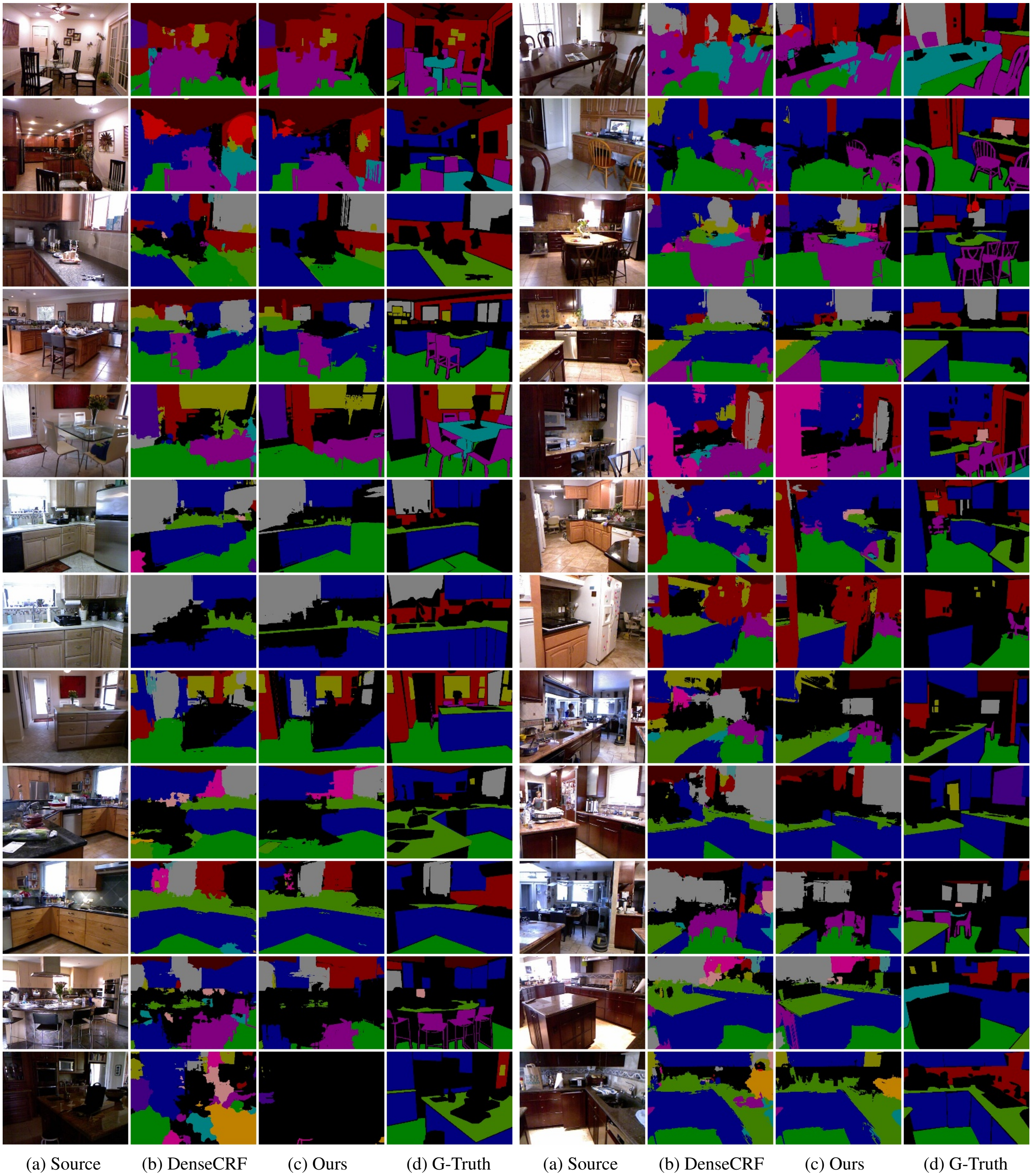
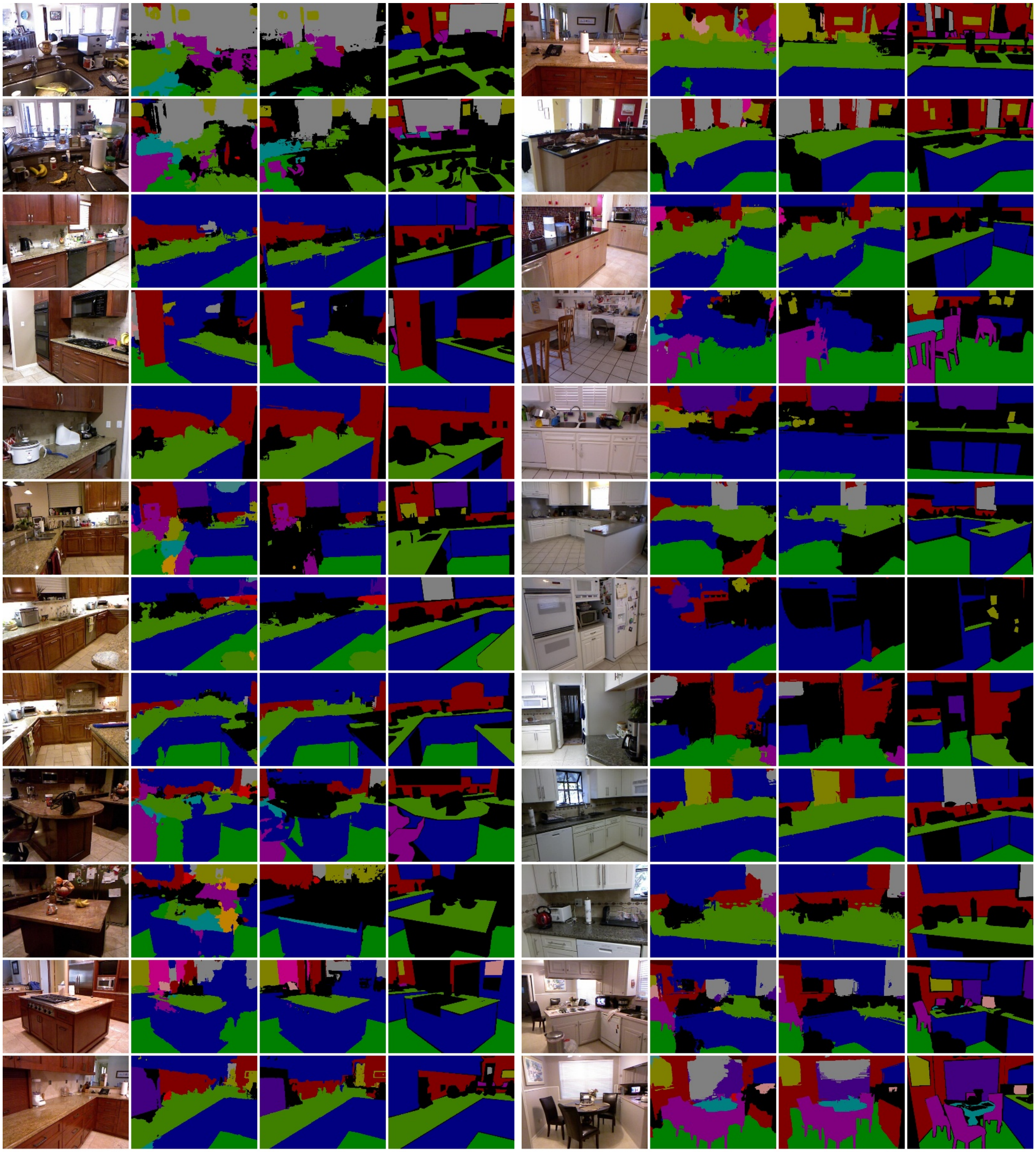


Figure 25: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 26: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

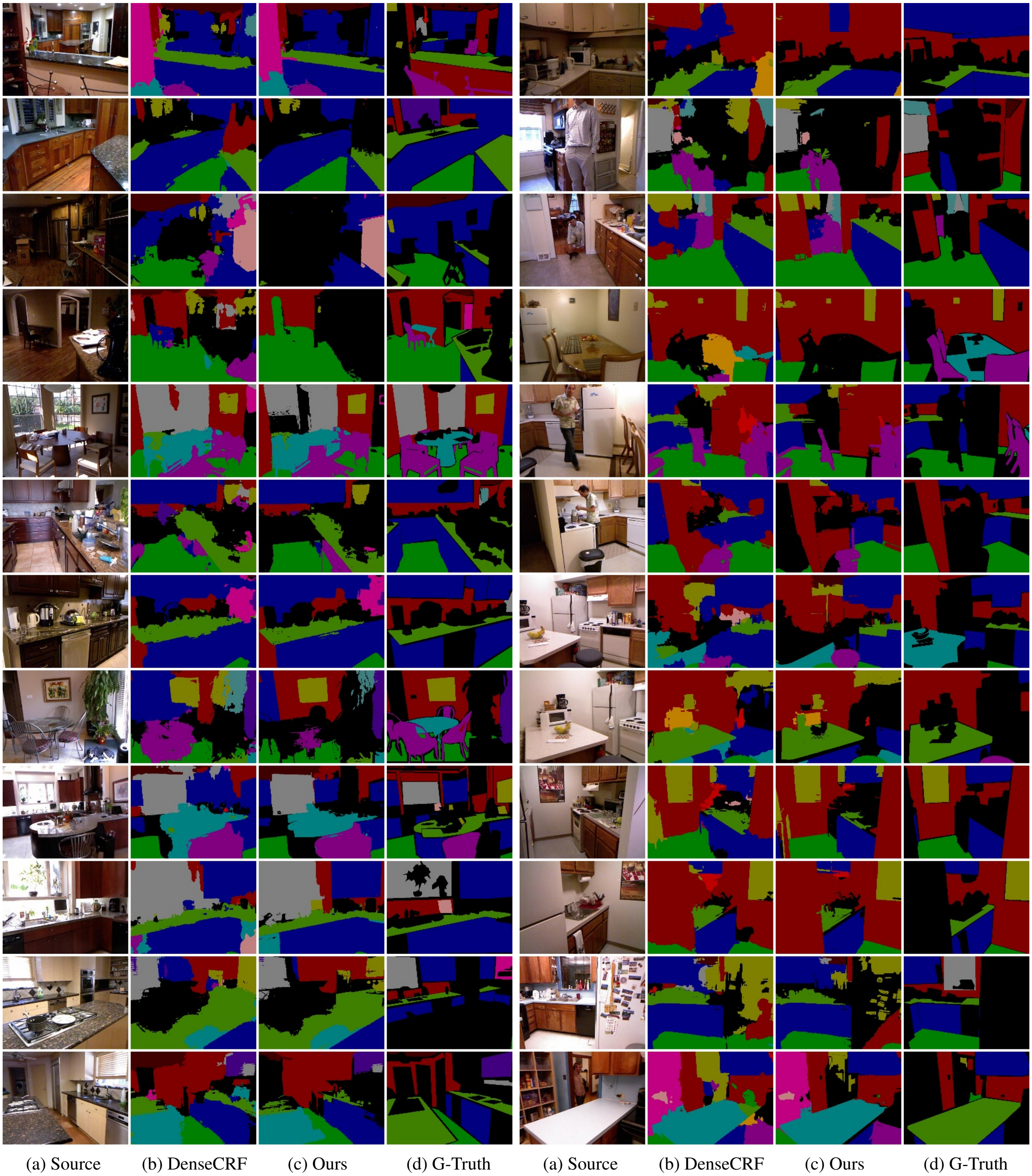
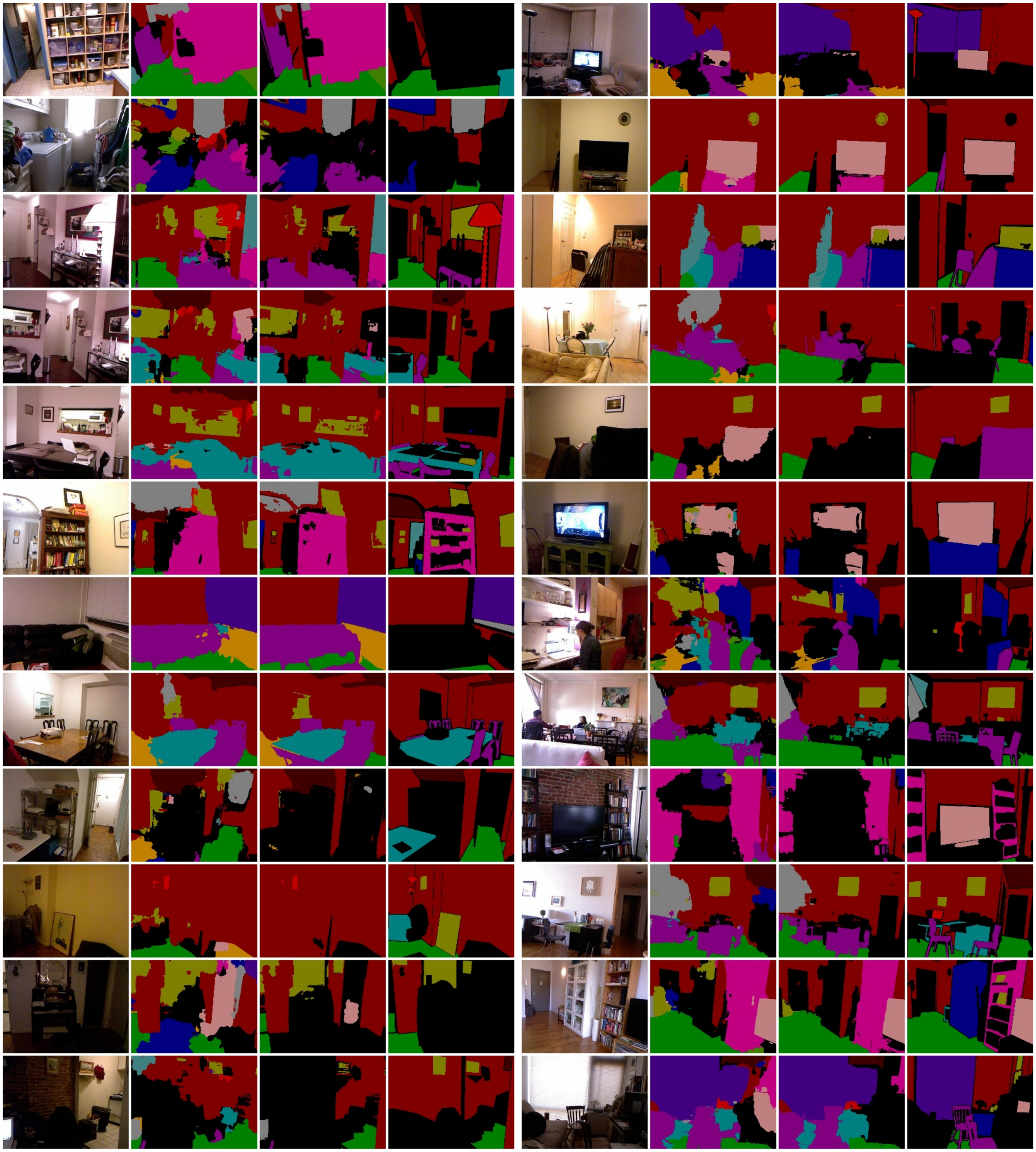
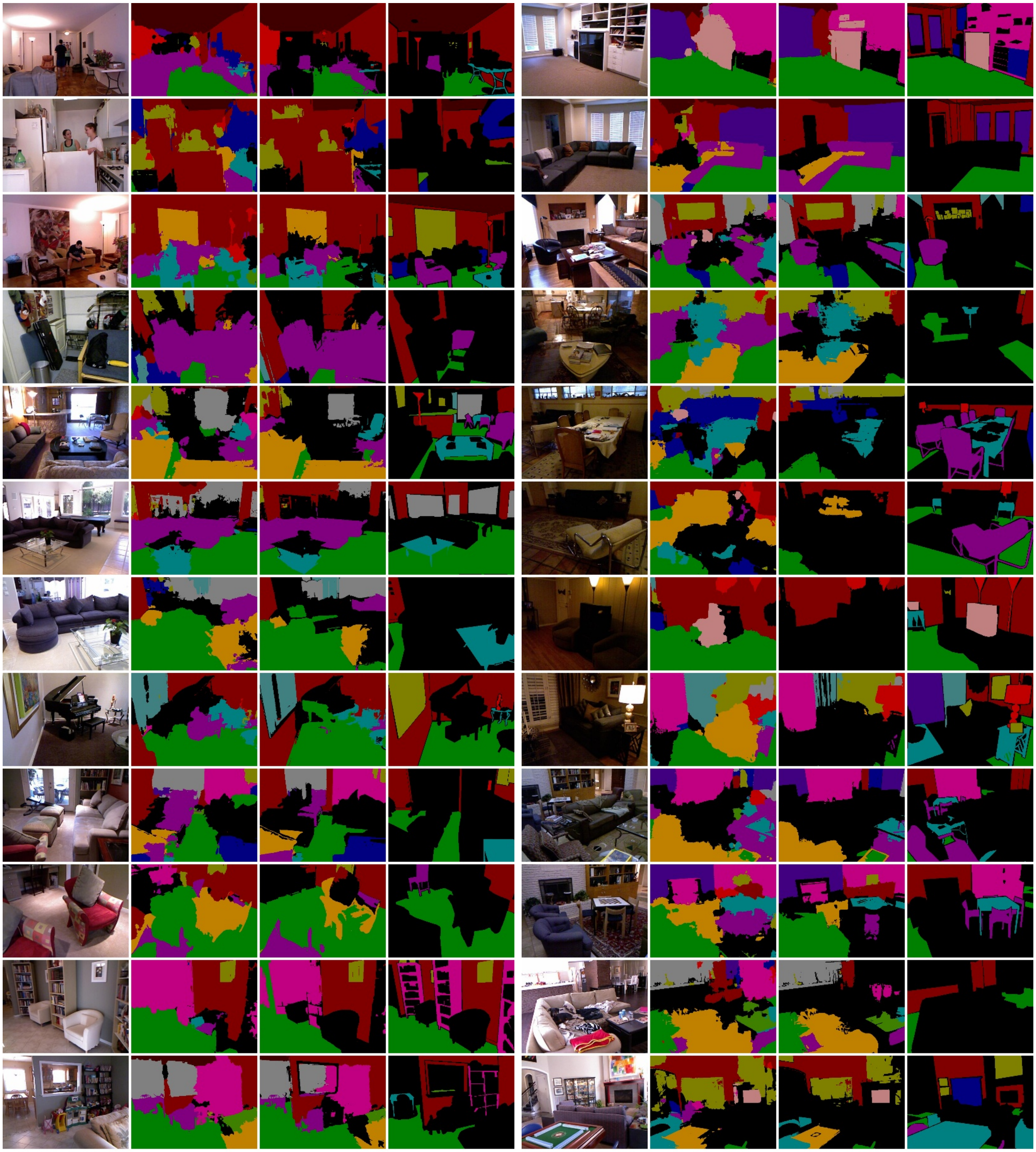


Figure 27: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 28: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

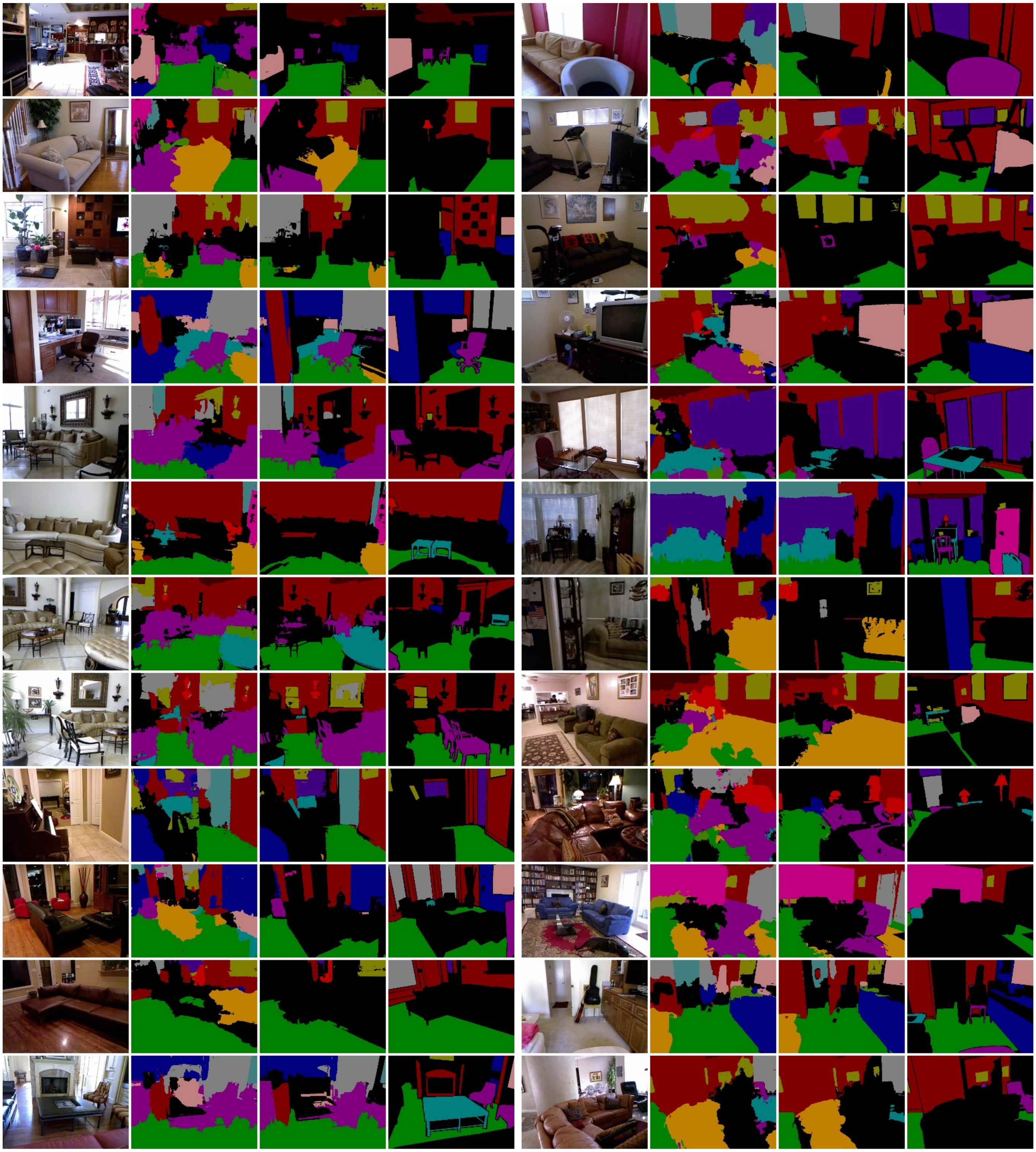


(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 29: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

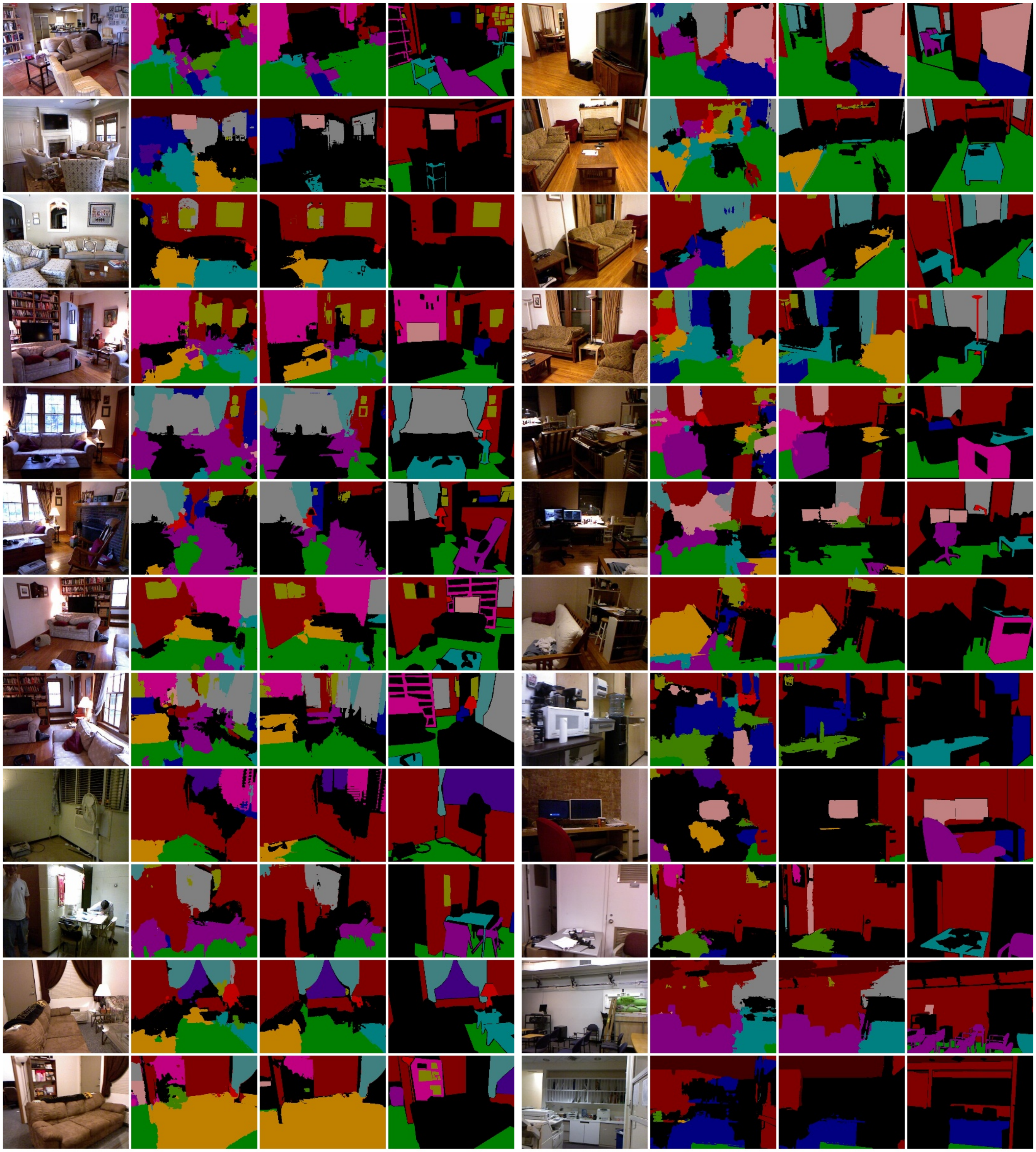
(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 30: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



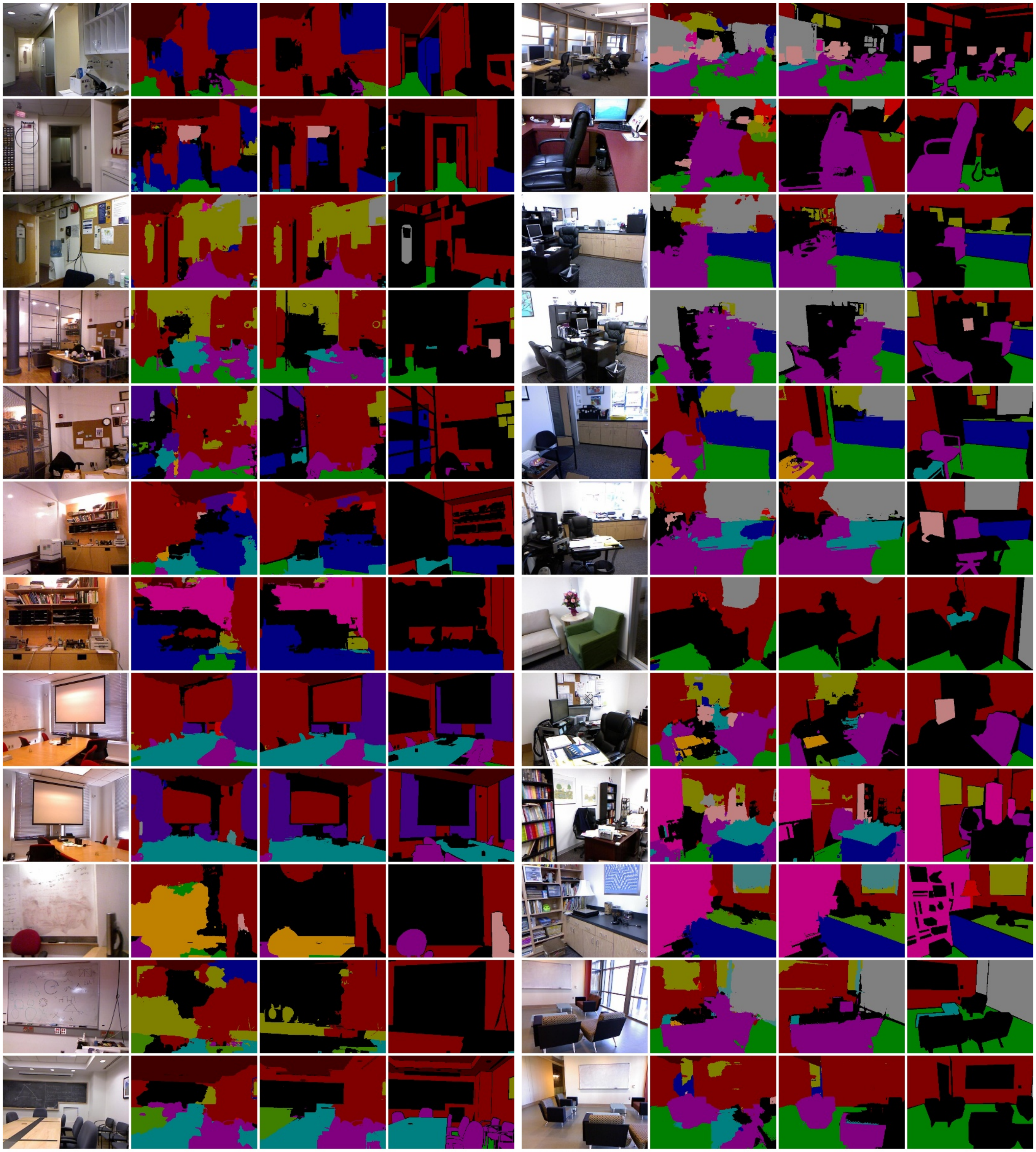
(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 31: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 32: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 33: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

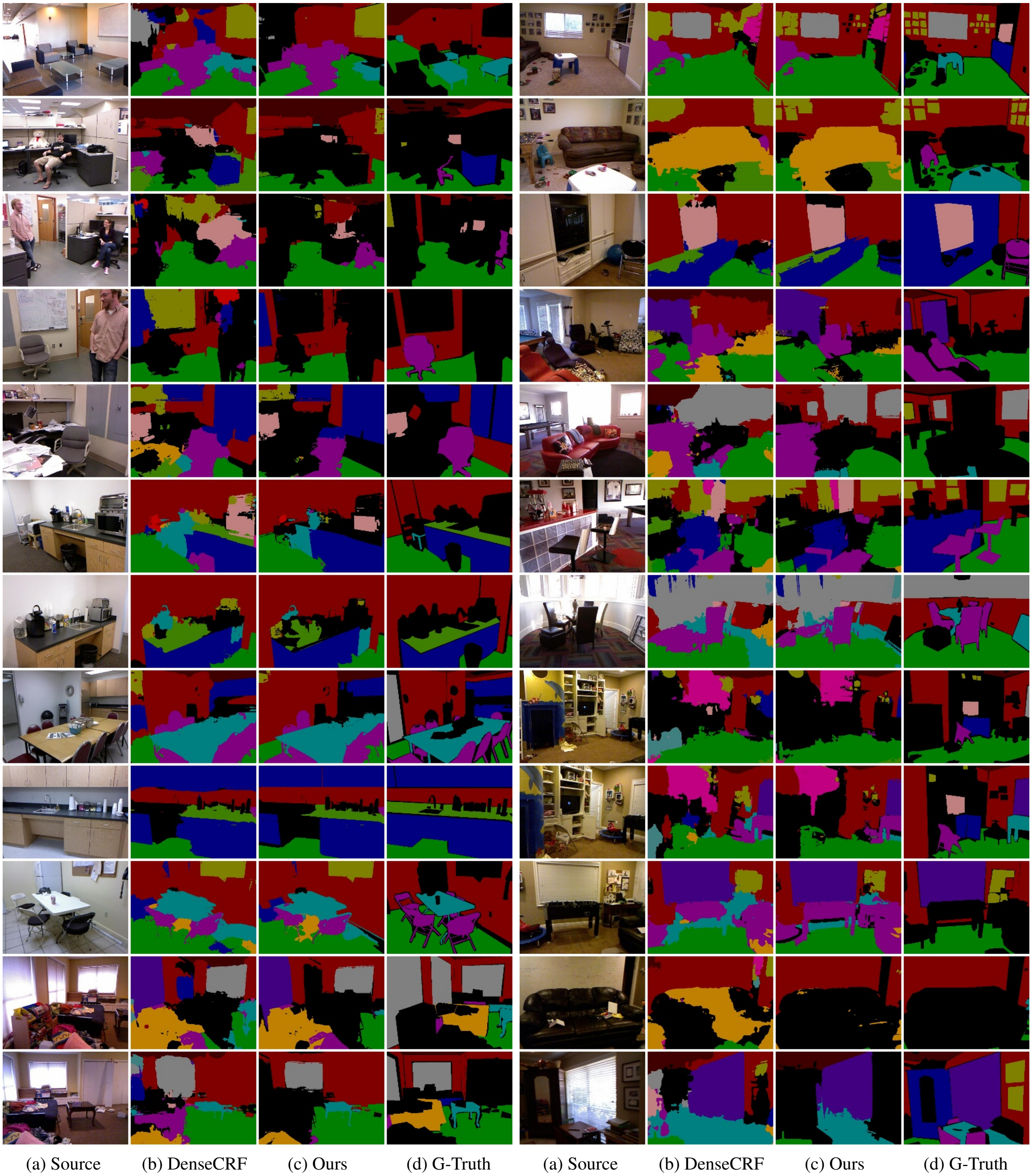
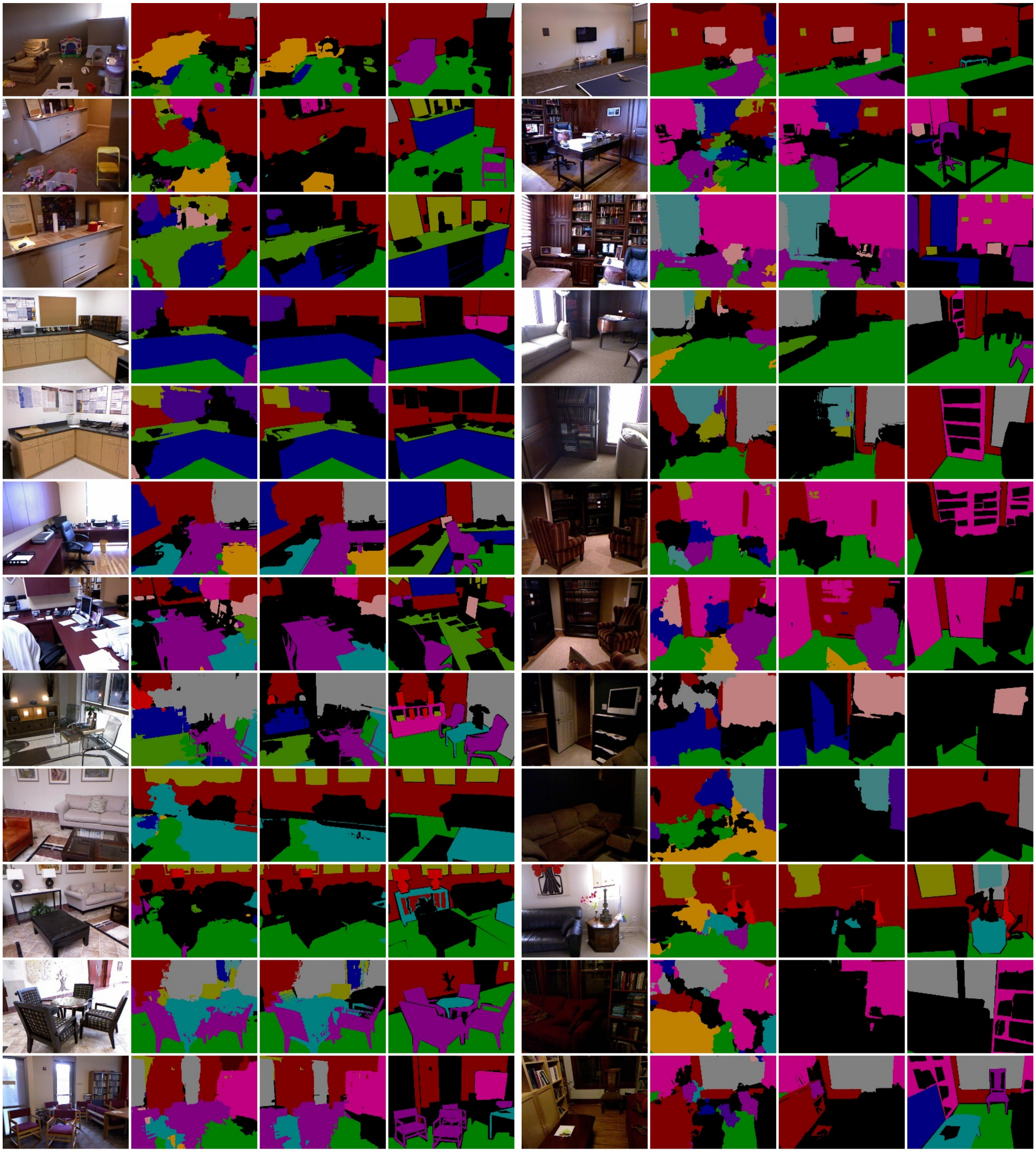


Figure 34: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*



(a) Source (b) DenseCRF (c) Ours (d) G-Truth (a) Source (b) DenseCRF (c) Ours (d) G-Truth

Figure 35: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*

Figure 36: *Comparision of automatic per-pixel object class prediction. Our results generally have higher quality than DenseCRF. Moreover, the additional attribute prediction allow users of our system to verbally refine the results, enabling an intuitive editing mode.*