Supplementary Materials:
Discriminative Re-ranking of Diverse Segmentations

Payman Yadollahpour
TTI-Chicago

Dhruv Batra
Virginia Tech

Gregory Shakhnarovich
TTI-Chicago

Abstract

In this supplementary document, we present how we characterize diversity in solutions along with some oracle statistics (Section 1.1) , and show some example success and failure cases of our re-ranker (Section 2).

1. Analyzing Diverse Segmentations

In this section, we characterize the diversity achieved in the multiple segmentations. Specifically, we investigate the sources of diversity, and attempt to quantify the extent to which diversity enables potential gain in accuracy over the MAP solution.

1.1. Diversity and Oracles

For the analysis reported in this section, we used the VOC 2012 training and validation sets. ALE and O_2P models were trained on VOC2012 training set, and the models were used to produce 10 segmentations for each image in the validation set. Following [1], we tuned the Lagrangian multipliers via cross-val ($\lambda_{ALE} = 1.25$ and $\lambda_{O_2P} = 0.08$).

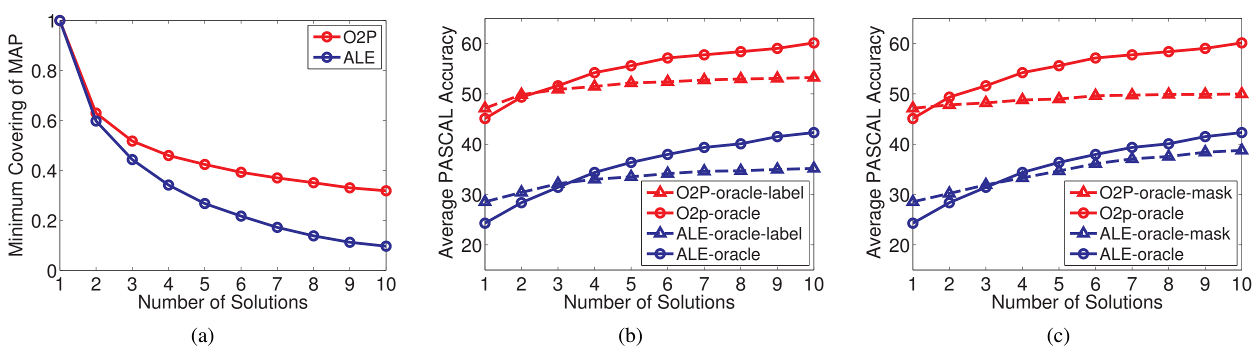

Figure 1: (a) Average minimum-covering (2) of MAP in the first M solutions vs. M. (b) Accuracy of an oracle restricted to labels present in the MAP, or (c) restricted to masks present in MAP. See text for details.

Diversity of Solutions. In order to characterize the diversity in *Div*MBEST solutions, we define a *covering* measure, which for a given image i captures how much of the MAP segmentation is covered by one of the subsequent solutions. Let $\{s_{i,1}^{(m)}, \ldots, s_{i,K}^{(m)}\}$ denote the set of K segments in the m^{th} solution for image i and $\{s_{i,1}^{(1)}, \ldots, s_{i,K'}^{(1)}\}$ denote the set of segments in MAP. The category-independent covering score is given by:

$$D_1(\mathbf{y}_i^{(m)}) = \frac{1}{\sum_{k'} |s_{i,k'}^{(1)}|} \sum_{k'=1}^{K'} |s_{i,k'}^{(1)}| \max_{k \in [K]} O(s_{i,k'}^{(1)}, s_{i,k}^{(m)}), \tag{1}$$

	$\mathbf{D_1}$ **Oracle Covering**	$\mathbf{D_2}$ **Oracle Covering**
ALE	0.55	0.45
$\mathbf{O_2P}$	0.61	0.58

Table 1: Average covering score between oracle solutions and MAP: (*left*) show the category-independent measure and (*right*) shows the category-specific measure. See text for details.

where $|s_{i,k}^{(m)}|$ denotes the size of the segment and $O(\cdot,\cdot)$ is the intersection-over-union measure of the two segments. For O_2P these segments correspond to CPMC segments [3], while in ALE the segments are connected components in the segmentation.

To get an idea of how different the *most* diverse solution is, we can define the minimum cover of the MAP solution by the M segmentations for image i as:

$$D_1^{(i,M)} \doteq \min_{m=1,\ldots,M} \{D_1(\mathbf{y}_i^{(m)})\}. \tag{2}$$

A plot of average minimum diversity in the dataset, *i.e.* $\sum_i D_1^{(i,M)}/n$ for $M = 1,\ldots,10$ is shown in Fig. 1a. We can see that both models produce at least one solution that is significantly different from the MAP.

Diversity of Oracle. The previous measure simply captures diversity and can be easily affected by poor quality solutions that are different from MAP. We can also try to characterize the diversity in the oracle solutions. This measure tells us how different the oracle solution is from the MAP solution on average. Analogous to (1) we can compute $D_1(\mathbf{y}_i^*)$ to measure by how much the segments in the oracle segmentation cover the MAP segments. We can also use a category-specific covering measure:

$$D_2(\mathbf{y}_i^*) = \frac{1}{\sum_{k'} |s_{i,k'}^{(1)}|} \sum_{k'=1}^{K'} |s_{i,k'}^{(1)}| \max_{\substack{k\in[K]:\\ y_{i,k}^*=y_{i,k'}^{(1)}}} O(s_{i,k'}^{(1)}, s_{i,k}^*), \tag{3}$$

where $y_{i,k}^*$ and $y_{i,k'}^{(1)}$ are the labels of the oracle and MAP segments respectively. Tbl. 1 summarizes these results which show that the oracle segmentations are not simply minor perturbations of the MAP segmentations.

Gain from Diversity. The previous measure tells us that the oracle solution is indeed quite different from the MAP. We now try to study *how* it is different – do the additional solutions introduce new categories or new masks or both? In order to answer this question, we measure the performance of a restricted oracle that chooses in each additional solution the best label possible for all segments, albeit restricted to the set of labels found in MAP. Fig. 1b shows that such a restricted oracle (O_2P`-oracle-label` and `ALE-oracle-label`) performs worse than the unrestricted oracle, indicating that the additional solutions do in fact introduce categories present in ground-truth but not in MAP.

Similarly, we can restrict the oracle to only take segment masks found in MAP and assign to them the best possible labels found in $\mathbf{y}_i^{(m)}$. Again Fig. 1c shows that such a restricted oracle significantly under-performs, indicating the MAP masks are not ideal and that the additional solutions do in fact introduce useful masks.

Thus, we can conclude that there are clear differences in both the labels and segments of the oracle segmentations compared to the MAP.

2. Example Re-ranking Results

We now show example re-ranking results (with the O_2P model). We divide these examples into three cases:

- **Success Case #1:** where MAP is not the most accurate solution and the re-ranker successfully picks a solution better than MAP. This case demonstrates the benefit of using a diverse list of segmentations that are useful to the re-ranker.
- **Success Case #2:** where MAP is the most accurate solution and the re-ranker successfully picks MAP.
- **Failure Case:** where the re-ranker picks a solution worse than MAP. These are typically pathological confusing cases like dog vs. cat or bus vs. car.

Figure 2: Color map for reading VOC segmentation results.

Success Case #1:
MAP is NOT the most accurate; Re-ranker picks a solution better than MAP

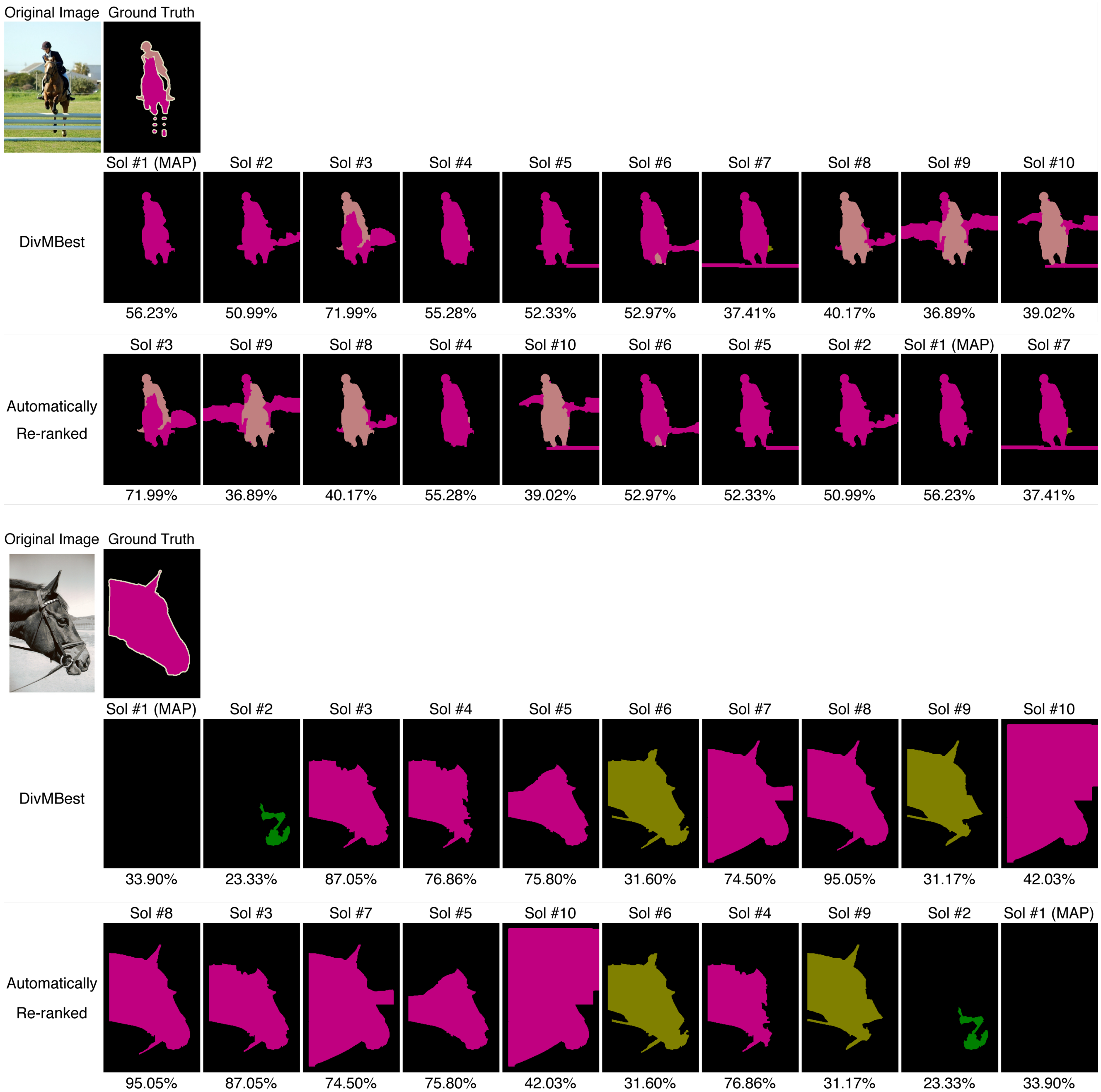

Figure 3: Success Case #1: MAP is not the most accurate; Re-ranker picks a solution better than MAP. In each group, the first row shows the original image and ground-truth segmentation. The second row shows the 10 *Div*MBEST solutions produced from the CRF in stage 1. The third row shows the solutions re-ranked by our proposed re-ranker.

Success Case #1 (Continued):
MAP is NOT the most accurate; Re-ranker picks a solution better than MAP

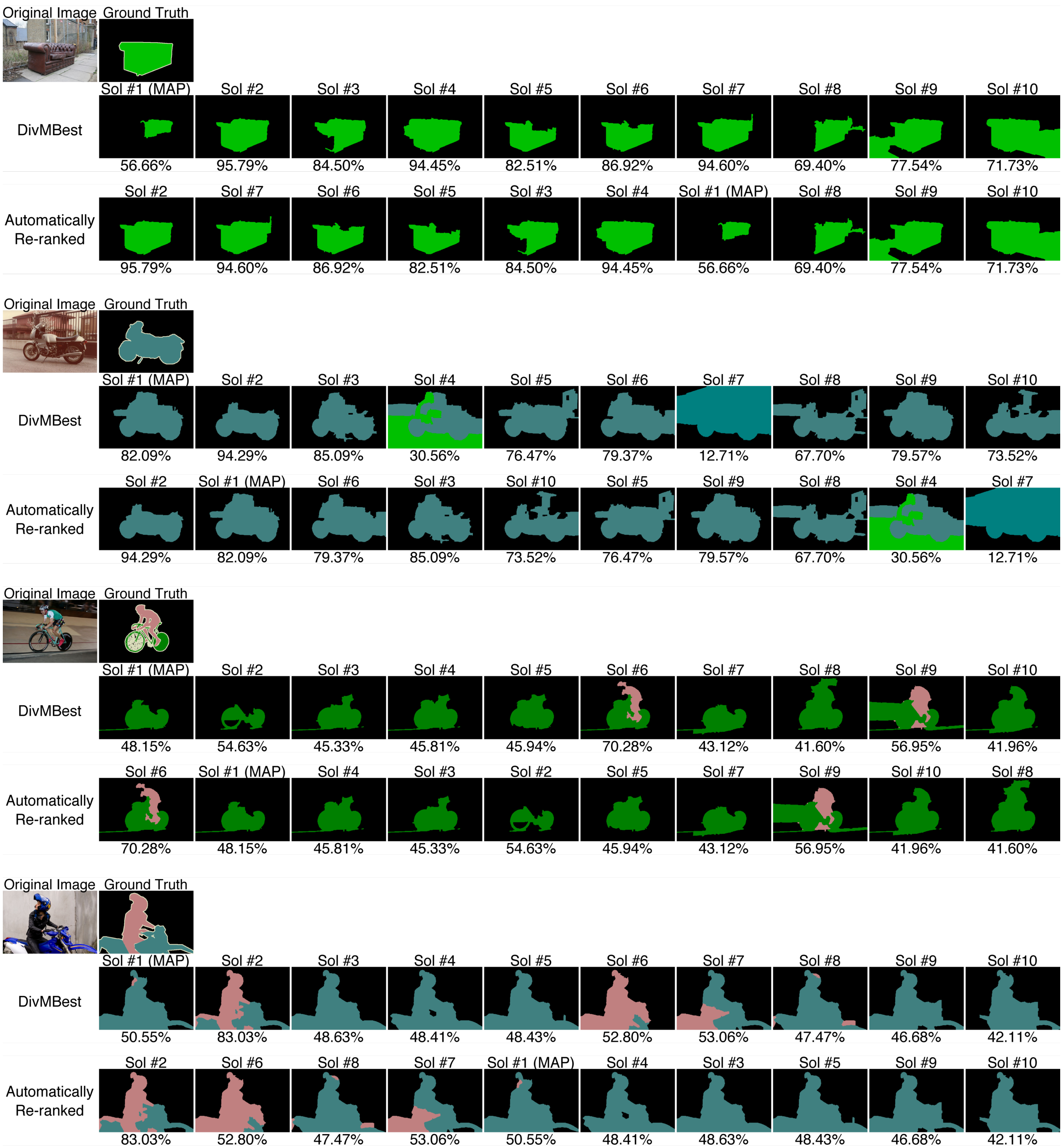

Figure 4: Success Case #1: MAP is not the most accurate; Re-ranker picks a solution better than MAP. In each group, the first row shows the original image and ground-truth segmentation. The second row shows the 10 *Div*MBEST solutions produced from the CRF in stage 1. The third row shows the solutions re-ranked by our proposed re-ranker.

Success Case #1 (Continued):
MAP is NOT the most accurate; Re-ranker picks a solution better than MAP

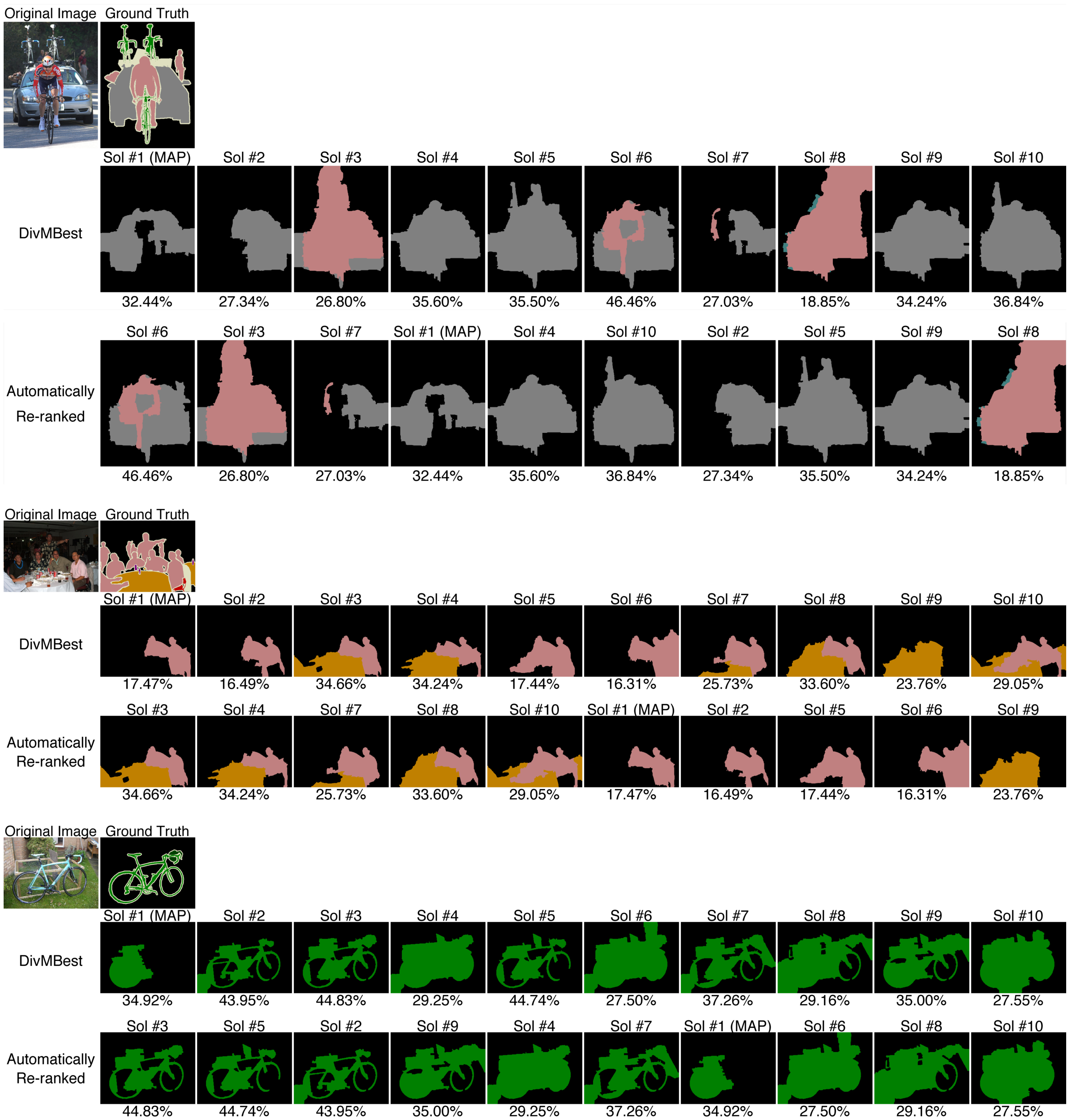

Figure 5: Success Case #1: MAP is not the most accurate; Re-ranker picks a solution better than MAP. In each group, the first row shows the original image and ground-truth segmentation. The second row shows the 10 *Div*MBEST solutions produced from the CRF in stage 1. The third row shows the solutions re-ranked by our proposed re-ranker.

Success Case #1 (Continued): MAP is NOT the most accurate; Re-ranker picks a solution better than MAP

Figure 6: Success Case #1: MAP is not the most accurate; Re-ranker picks a solution better than MAP. In each group, the first row shows the original image and ground-truth segmentation. The second row shows the 10 *Div*MBEST solutions produced from the CRF in stage 1. The third row shows the solutions re-ranked by our proposed re-ranker.

Success Case #1 (Continued):
MAP is NOT the most accurate; Re-ranker picks a solution better than MAP

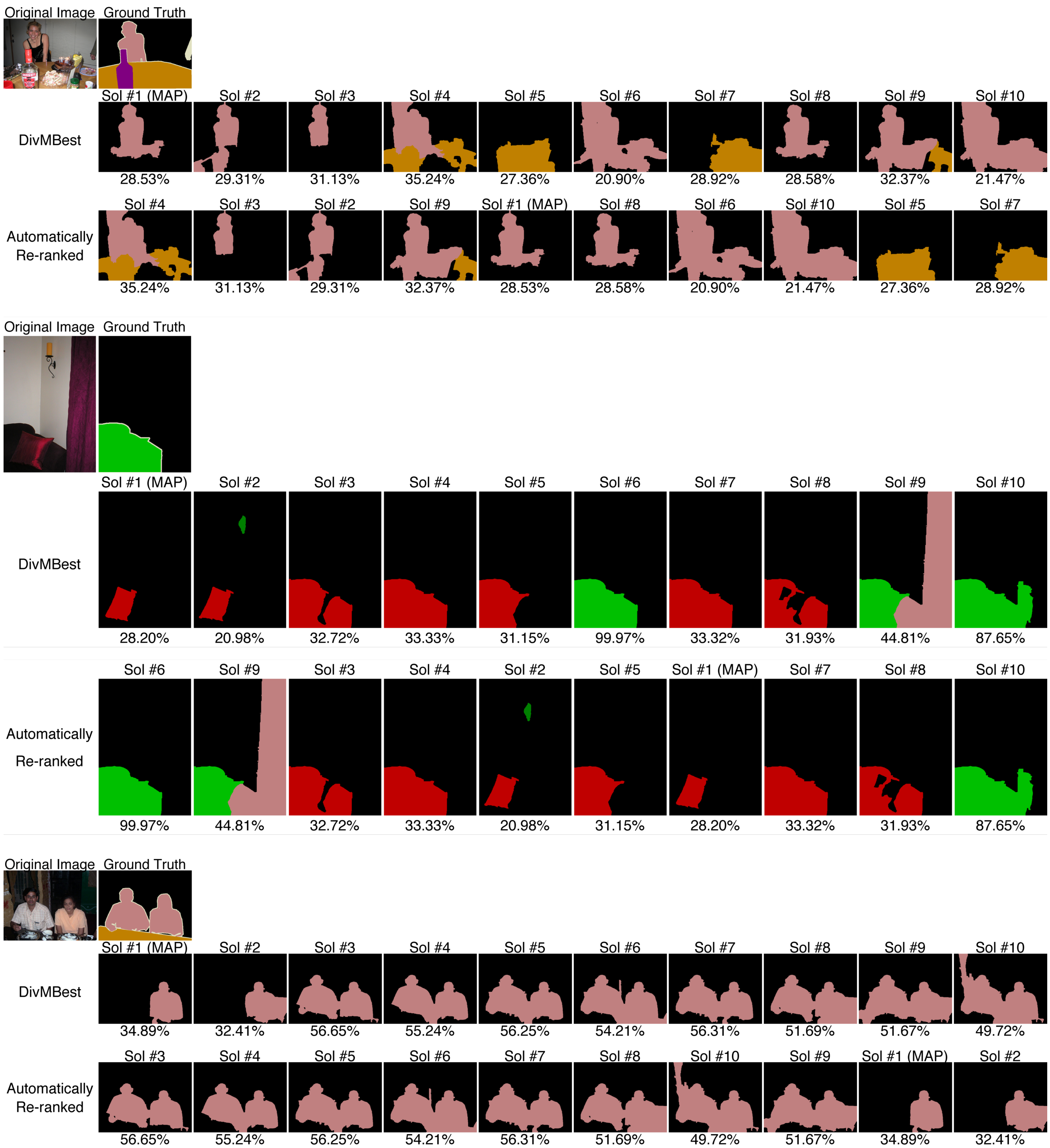

Figure 7: Success Case #1: MAP is not the most accurate; Re-ranker picks a solution better than MAP. In each group, the first row shows the original image and ground-truth segmentation. The second row shows the 10 *Div*MBEST solutions produced from the CRF in stage 1. The third row shows the solutions re-ranked by our proposed re-ranker.

Success Case #1 (Continued):
MAP is NOT the most accurate; Re-ranker picks a solution better than MAP

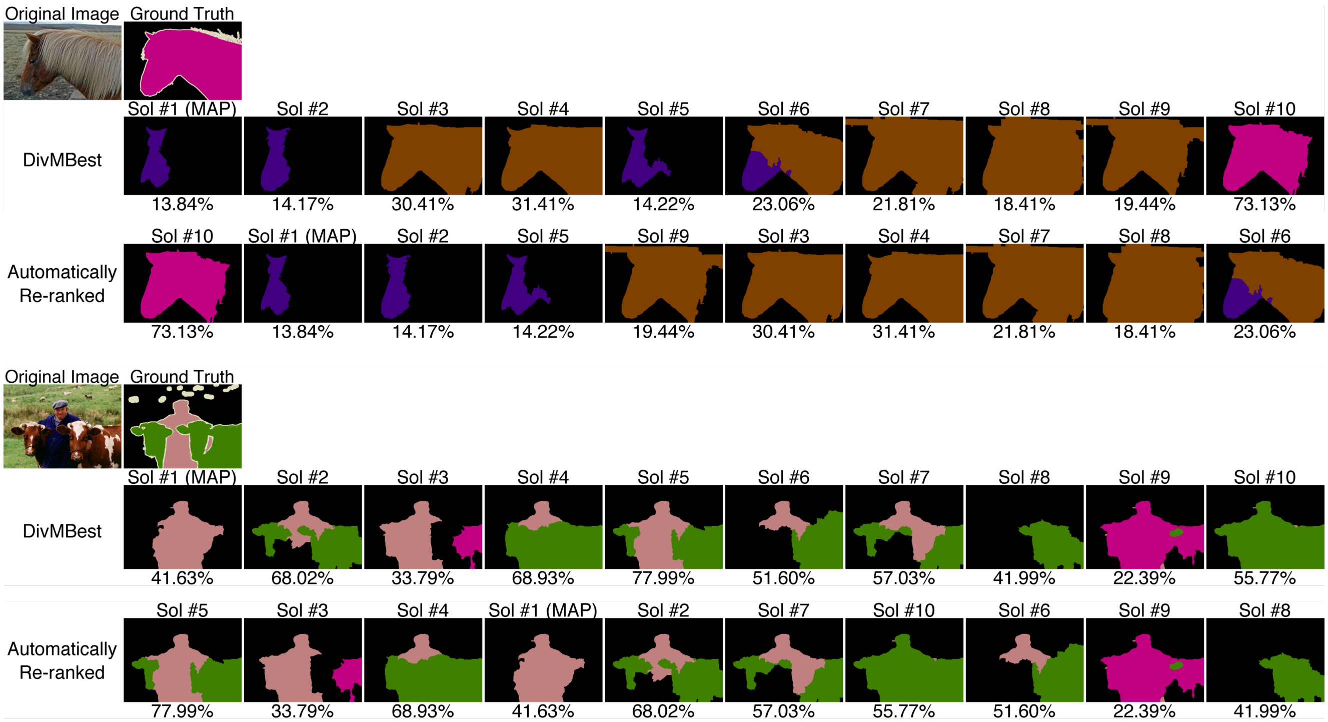

Figure 8: Success Case #1: MAP is not the most accurate; Re-ranker picks a solution better than MAP. In each group, the first row shows the original image and ground-truth segmentation. The second row shows the 10 *Div*MBEST solutions produced from the CRF in stage 1. The third row shows the solutions re-ranked by our proposed re-ranker.

Success Case #2:
MAP is the most accurate; Re-ranker picks MAP

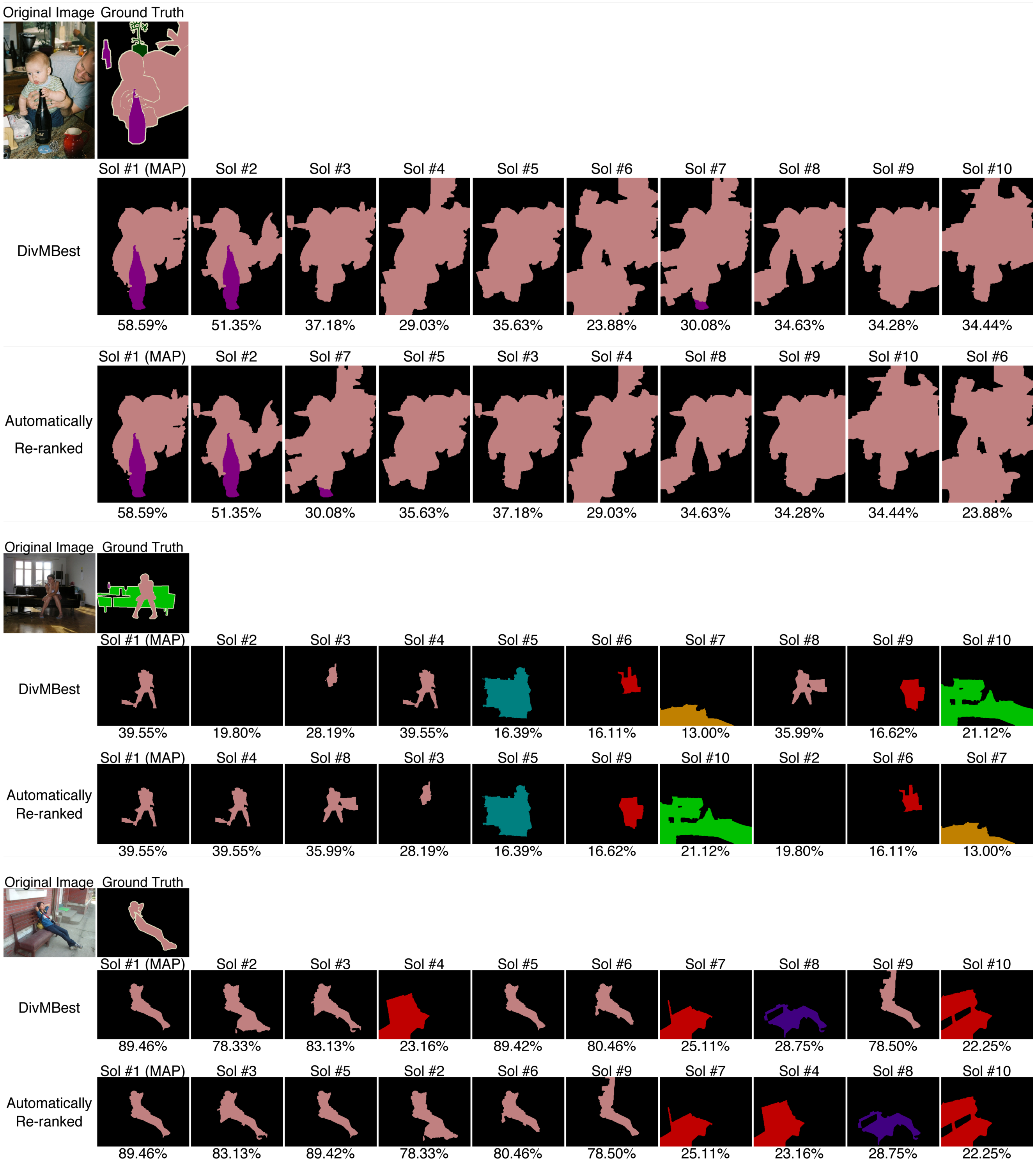

Figure 9: Success Case #2: MAP is the most accurate; Re-ranker picks MAP. In each group, the first row shows the original image and ground-truth segmentation. The second row shows the 10 *Div*MBEST solutions produced from the CRF in stage 1. The third row shows the solutions re-ranked by our proposed re-ranker.

Success Case #2 (Continued):
MAP is the most accurate; Re-ranker picks MAP

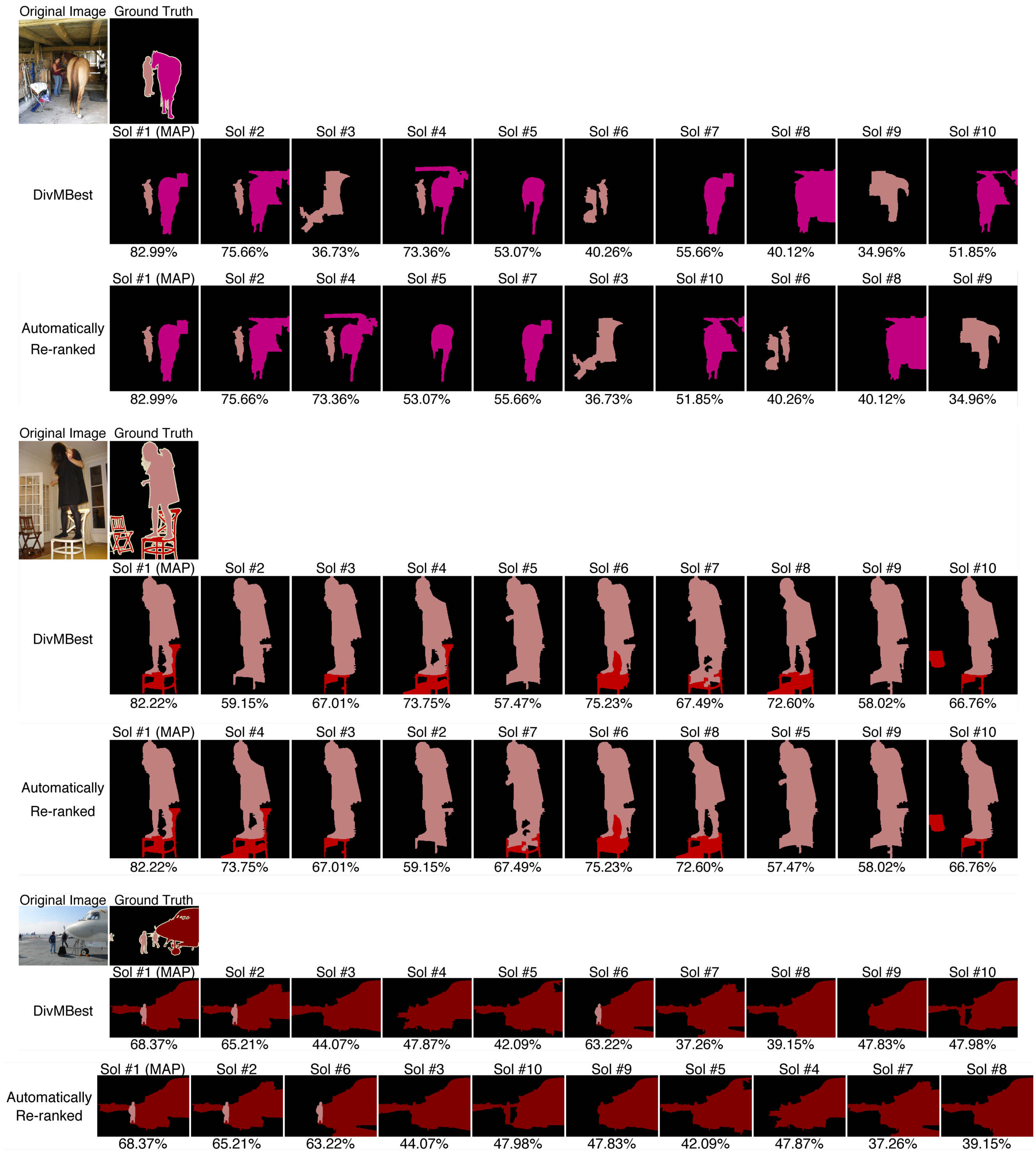

Figure 10: Success Case #2: MAP is the most accurate; Re-ranker picks MAP. In each group, the first row shows the original image and ground-truth segmentation. The second row shows the 10 *Div*MBEST solutions produced from the CRF in stage 1. The third row shows the solutions re-ranked by our proposed re-ranker.

Success Case #2 (Continued):
MAP is the most accurate; Re-ranker picks MAP

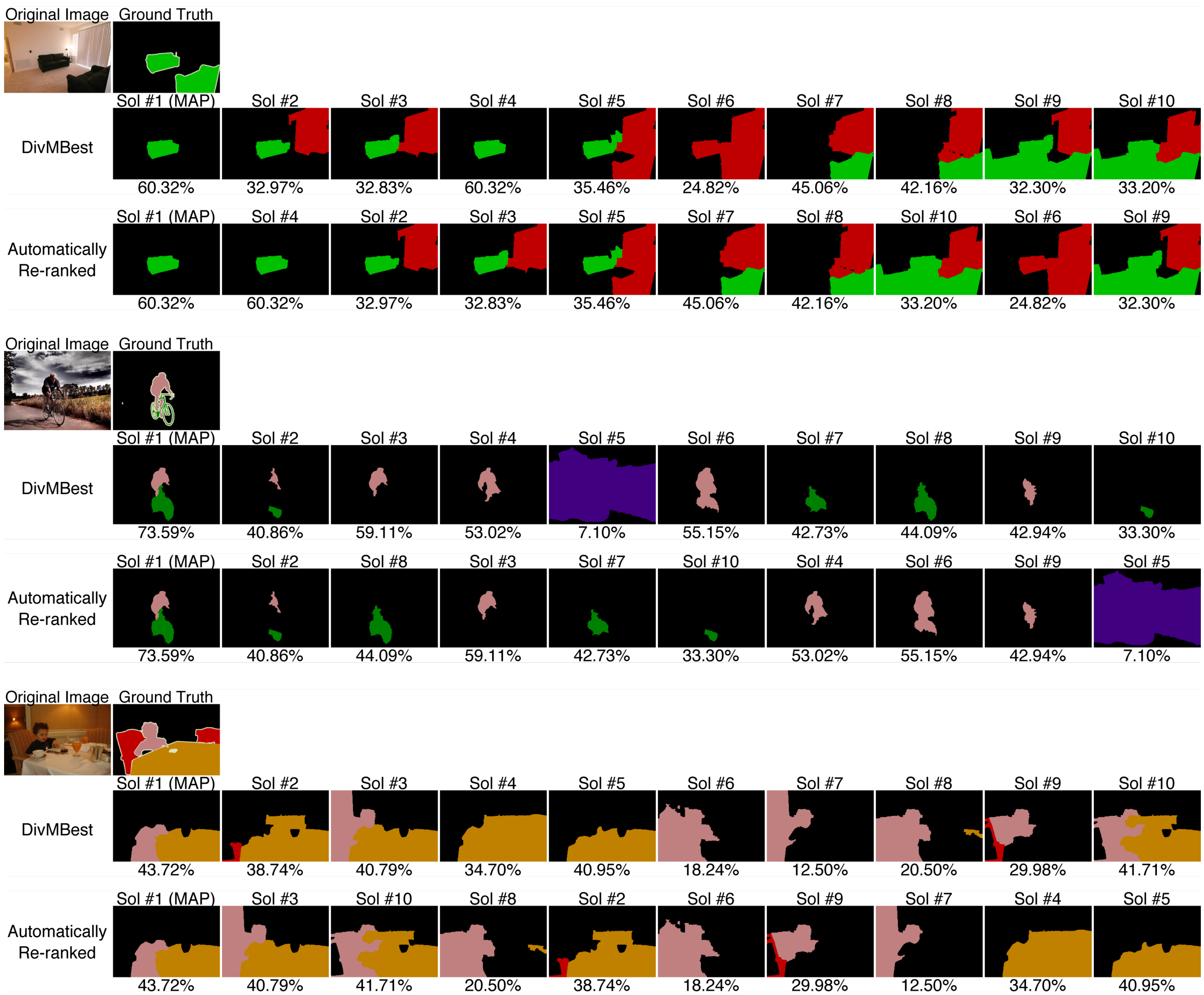

Figure 11: Success Case #2: MAP is the most accurate; Re-ranker picks MAP. In each group, the first row shows the original image and ground-truth segmentation. The second row shows the 10 *Div*MBEST solutions produced from the CRF in stage 1. The third row shows the solutions re-ranked by our proposed re-ranker.

Failure Case:
Re-ranker picks a solution worse than MAP

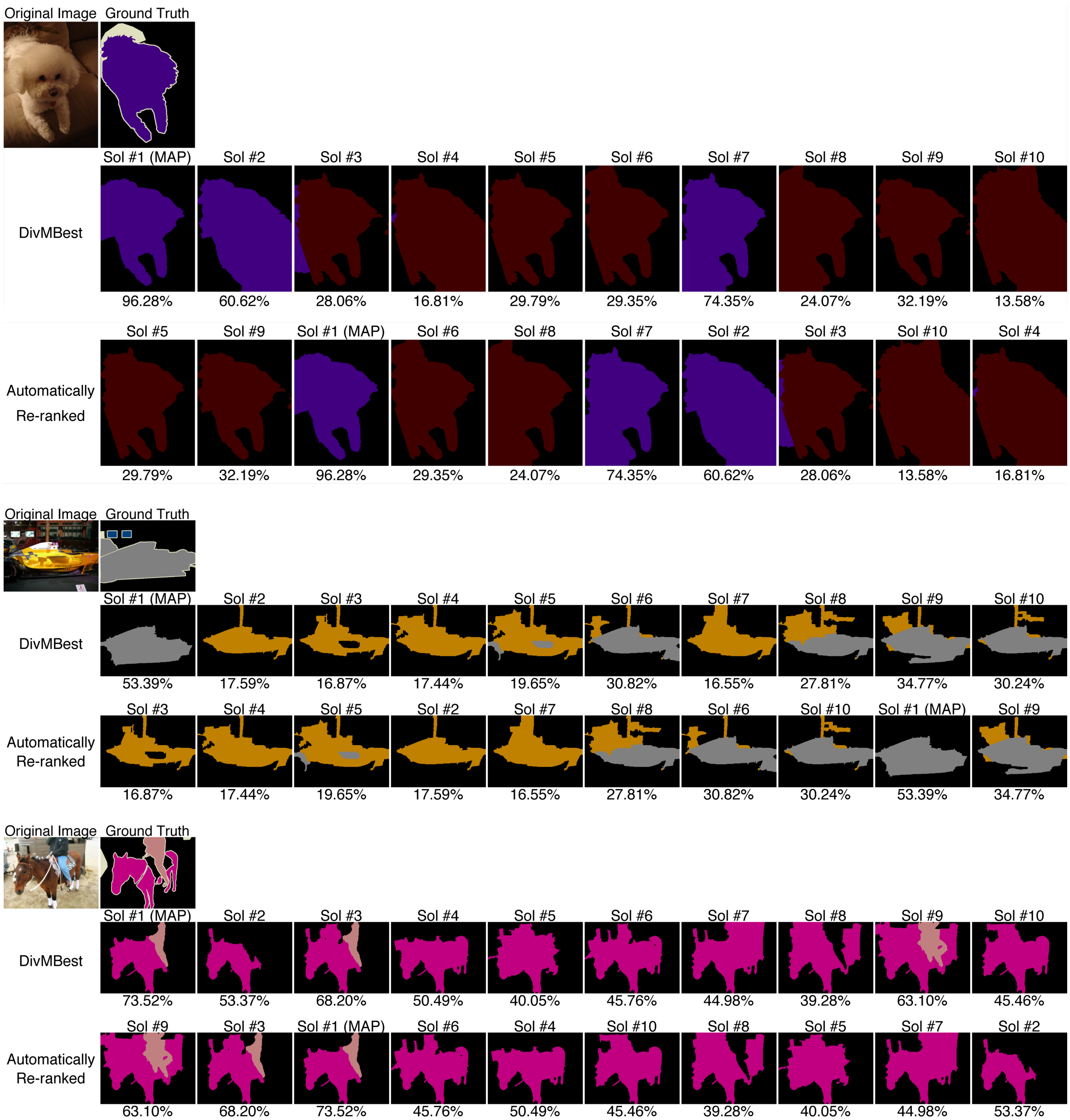

Figure 12: Failure Case: Re-ranker picks a solution worse than MAP. In each group, the first row shows the original image and ground-truth segmentation. The second row shows the 10 *Div*MBEST solutions produced from the CRF in stage 1. The third row shows the solutions re-ranked by our proposed re-ranker.

Failure Case:
Re-ranker picks a solution worse than MAP

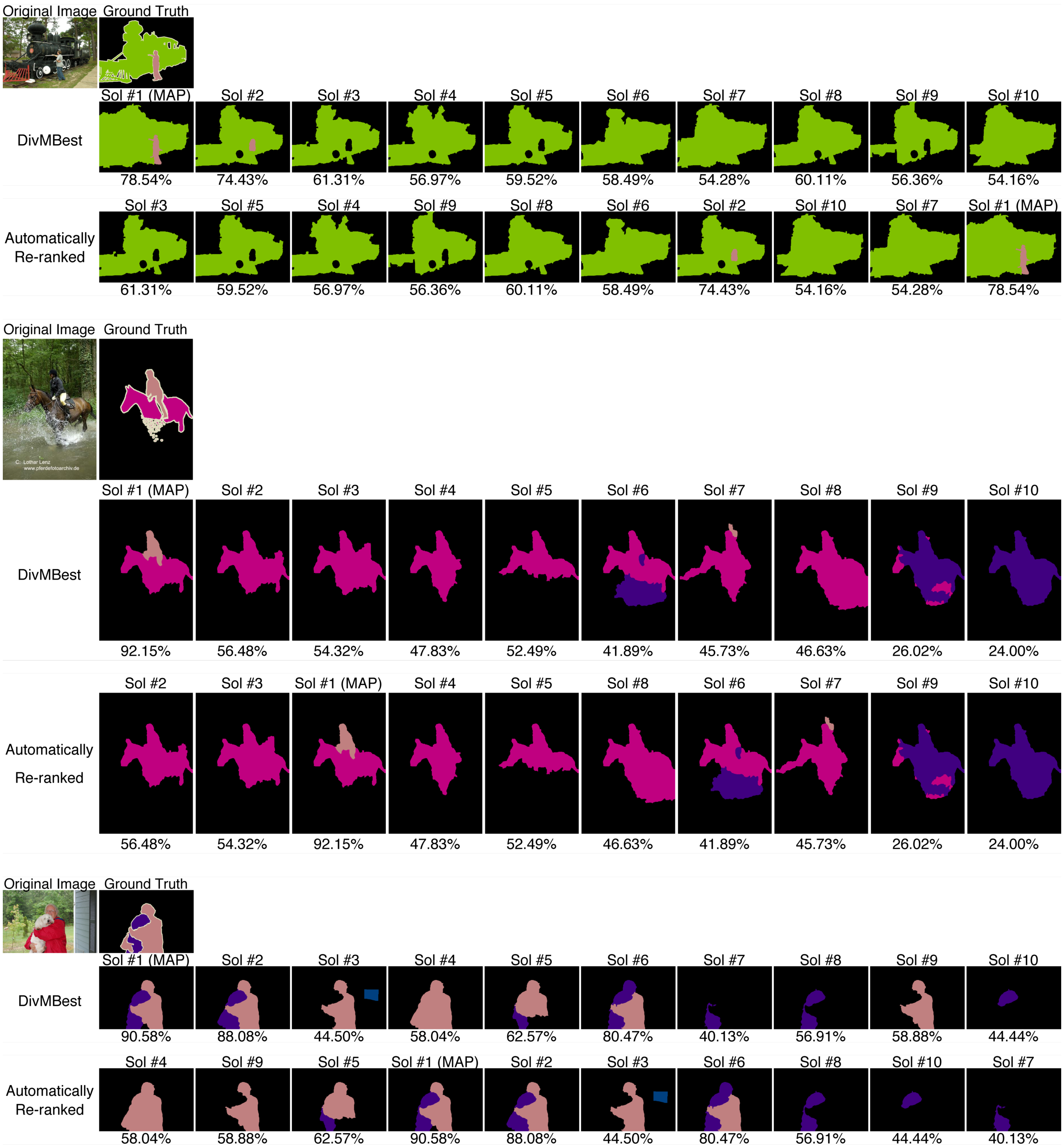

Figure 13: Failure Case: Re-ranker picks a solution worse than MAP. In each group, the first row shows the original image and ground-truth segmentation. The second row shows the 10 *Div*MBEST solutions produced from the CRF in stage 1. The third row shows the solutions re-ranked by our proposed re-ranker.

Failure Case:
Re-ranker picks a solution worse than MAP

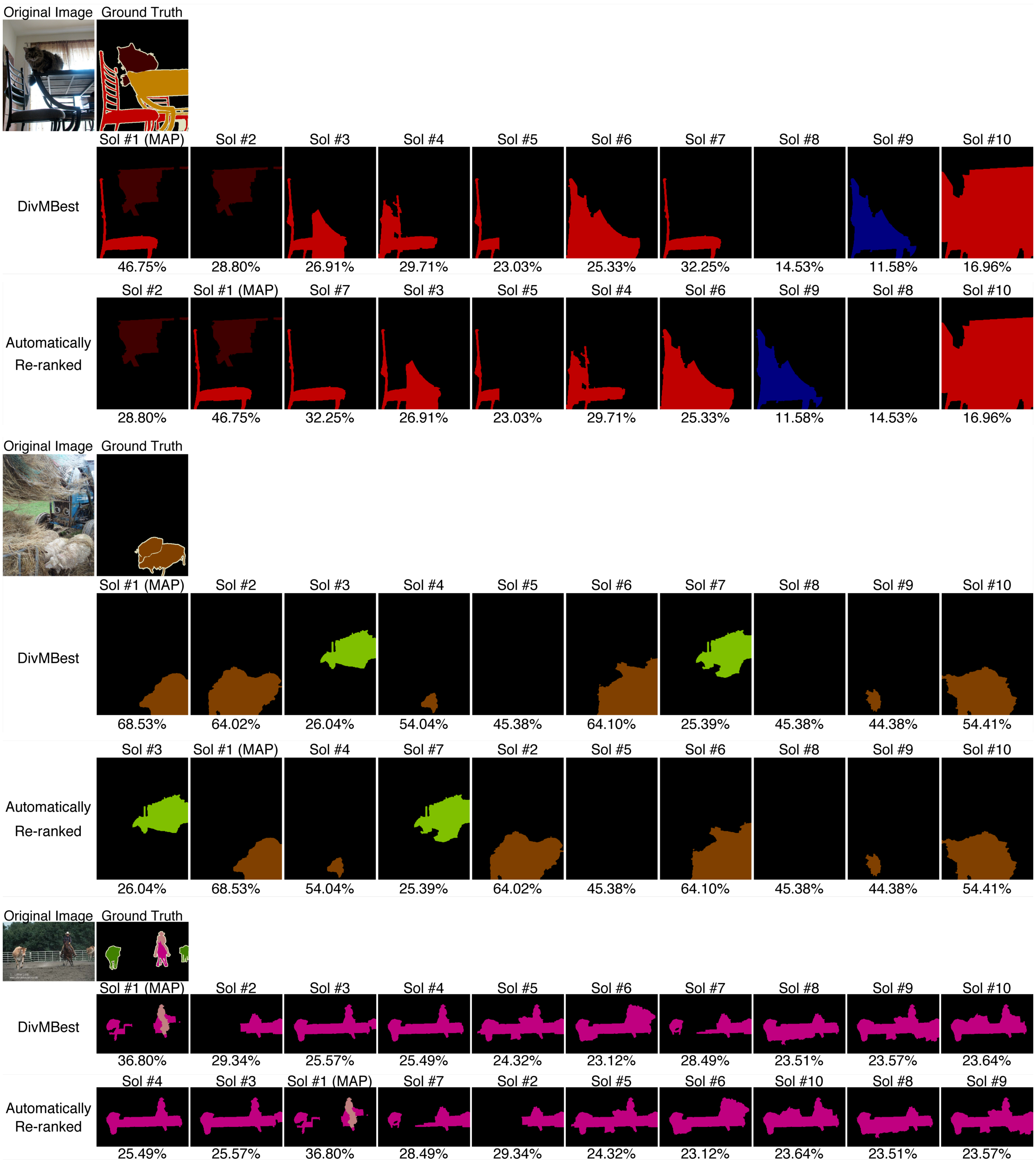

Figure 14: Failure Case: Re-ranker picks a solution worse than MAP. In each group, the first row shows the original image and ground-truth segmentation. The second row shows the 10 *Div*MBEST solutions produced from the CRF in stage 1. The third row shows the solutions re-ranked by our proposed re-ranker.

References

[1] D. Batra, P. Yadollahpour, A. Guzman-Rivera, and G. Shakhnarovich. Diverse M-Best Solutions in Markov Random Fields. In *ECCV*, 2012. 1
[2] J. Carreira, R. Caseiro, J. Batista, and C. Sminchisescu. Semantic segmentation with second-order pooling. In *ECCV*, pages 430–443, 2012.
[3] J. Carreira and C. Sminchisescu. Constrained parametric min-cuts for automatic object segmentation. In *CVPR*, 2010. 1
[4] L. Ladickỳ, C. Russell, P. Kohli, and P. H. S. Torr. Associative hierarchical CRFs for object class image segmentation. *ICCV*, 2009.
[5] L. Ladicky, C. Russell, P. Kohli, and P. H. S. Torr. Graph cut based inference with co-occurrence statistics. In *ECCV*, 2010.