Affinity CNN: Learning Pixel-Centric Pairwise Relations for Figure/Ground Embedding

Supplementary Material

Michael Maire
TTI Chicago
mmaire@ttic.edu

Takuya Narihira
Sony Corp.
takuya.narihira@jp.sony.com

Stella X. Yu
UC Berkeley / ICSI
stellayu@berkeley.edu

We provide a self-contained quantitative and qualitative assessment of figure/ground performance, offering additional detail beyond that which appears in the main text.

1. Figure/Ground Benchmark

Given a per-pixel figure/ground ordering assignment, and a segmentation partitioning an image into regions, we can easily order the regions according to figure/ground layering. Simply assign each region a rank order equal to the mean figure/ground order of its member pixels. For robustness to minor misalignment between the figure/ground assignment and the boundaries of regions in the segmentation, we use median in place of mean.

This transfer procedure serves as a basis for comparing different figure/ground orderings. We transfer them both onto the same segmentation. In particular, given predicted figure/ground ordering $\theta(\cdot)$, ground-truth figure/ground ordering $\tilde{\theta}(\cdot)$, and ground-truth segmentation S, we transfer each of $\theta(\cdot)$ and $\tilde{\theta}(\cdot)$ onto S. This gives two orderings of the regions in S, which we compare according to the following metrics:

- Pairwise region accuracy (R-ACC): For each pair of neighboring regions in S, if the ground-truth figure/ground assignment shows them to be in different layers, we test whether the predicted relative ordering of these regions matches the ground-truth relative ordering. That is, we measure accuracy on the classification problem of predicting which region is in front.

- Boundary ownership accuracy (B-ACC): We define the front region as owning the pixels on the common boundary of the region pair and measure the per-pixel accuracy of predicting boundary ownership. This is a reweighting of R-ACC. In R-ACC, all region pairs straddling a ground-truth figure/ground boundary count equally. In B-ACC, their importance is weighted according to length of the boundary.

- Boundary ownership of foreground regions (B-ACC-50, B-ACC-25): Identical to B-ACC, except we only consider boundaries which belong to the foreground-most 50% or 25% of regions in the ground-truth figure/ground ordering of each image. These metrics emphasize the importance of correct predictions for foreground objects while ignoring more distant objects.

Segmentation: Ground-truth	Figure/Ground Prediction Accuracy R-ACC	B-ACC	B-ACC-50	B-ACC-25
F/G: Ours	**0.62**	**0.69**	**0.72**	**0.73**
F/G: Maire [1]	0.56	0.58	0.56	0.56

Segmentation: Ours	Figure/Ground Prediction Accuracy R-ACC	B-ACC	B-ACC-50	B-ACC-25
F/G: Ours	**0.66**	**0.70**	**0.69**	**0.67**
F/G: Maire [1]	0.59	0.62	0.61	0.58

Table 1. **Figure/ground benchmark results.** After transferring figure/ground predictions onto either ground-truth (upper table) or our own (lower table) segmentations, we quantify accuracy of local relative relationships. R-ACC is pairwise region accuracy: considering all pairs of neighboring regions, what fraction are correctly ordered by relative figure/ground? B-ACC is boundary ownership accuracy: what fraction of boundary pixels have correct figure ownership assigned? B-ACC-50 and B-ACC-25 restrict measurement to the boundaries of the 50% and 25% most foreground regions (a proxy for foreground objects). Our system dramatically outperforms [1] across all metrics.

Note that S need not be the ground-truth segmentation. We can project ground-truth figure/ground onto any segmentation (say, a machine-generated one) and compare to predicted figure/ground projected onto that segmentation.

2. Benchmark Results

Table 1 reports a complete comparison of our figure/ground predictions and those of [1] against ground-truth figure/ground on our 50 image test subset of BSDS [2]. We consider both projection onto ground-truth segmentation and onto our own system's segmentation output. For the latter, as our system produces hierarchical segmentation, we use the region partition at a fixed level of the hierarchy, calibrated for optimal boundary F-measure. Figures 1 and 2 provide visual comparison on 10 of the 50 test images.

Across all metrics and both choices for segmentation, our system significantly outperforms [1]. We achieve 69%

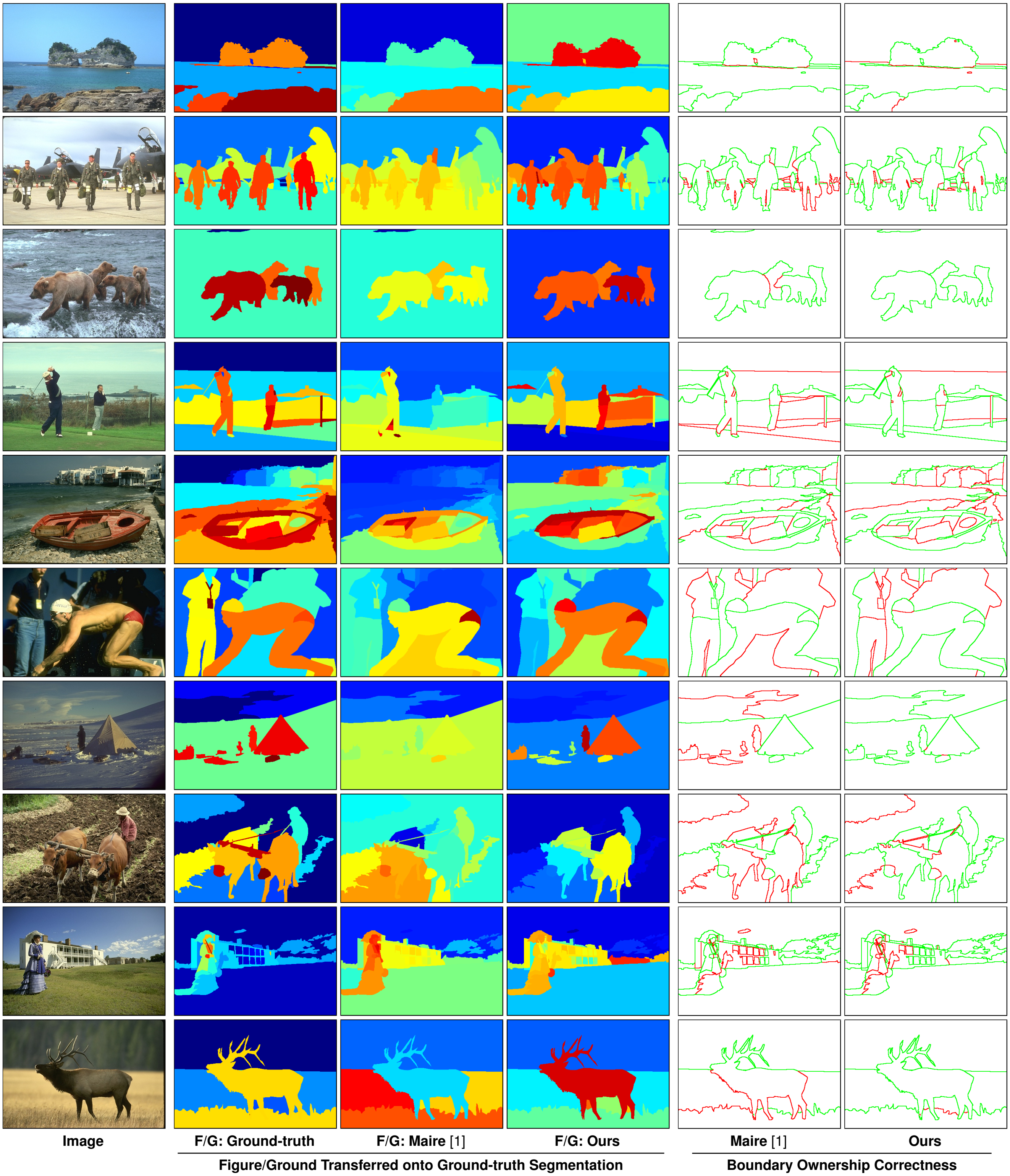

Figure 1. **Figure/ground predictions measured on ground-truth segmentation.** We transfer per-pixel figure/ground predictions (columns 2 through 4 of Figure 7 in the paper) onto the ground-truth segmentation by taking the median value over each region. For boundaries separating regions with different ground-truth figure/ground layer assignments, we check whether the predicted owner (more figural region) matches the owner according to the ground-truth. The rightmost two columns mark correct boundary ownership predictions in green and errors in red for both the results of Maire's system [1] and our system. Note how we correctly predict ownership of object lower boundaries (rows 6, 8, 10) and improve on small objects (row 7). Table 1 gives quantitative benchmarks.

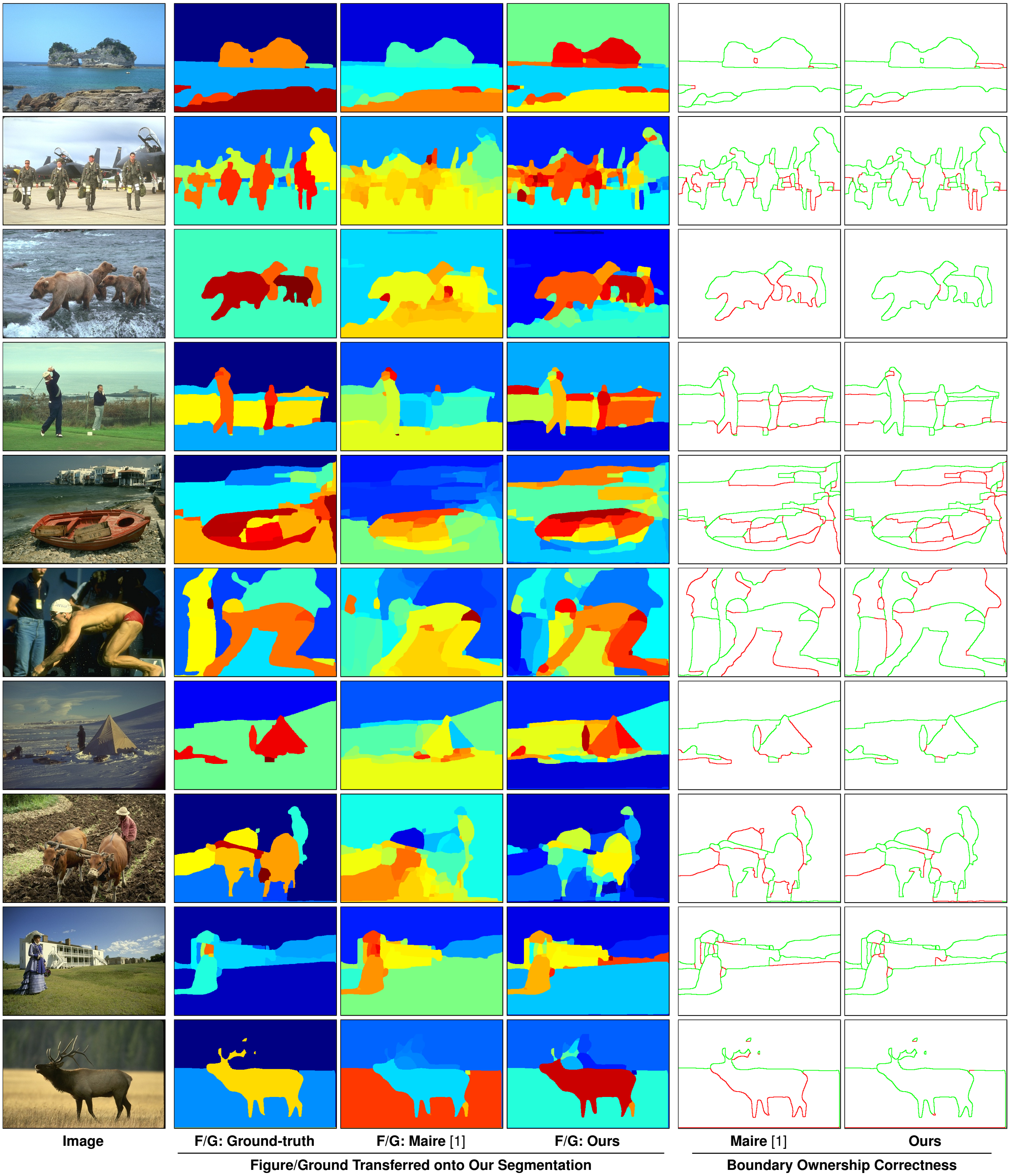

Figure 2. **Figure/ground predictions measured on our segmentation.** As in Figure 1, we transfer ground-truth figure/ground, Maire's figure/ground predictions [1], and our predictions onto a common segmentation. However, instead of using the ground-truth segmentation, we transfer onto the segmentation generated by our system. The ground-truth figure/ground transferred onto our regions defines the boundary ownership signal against which we judge predictions. Comparing with Figure 1, our boundary ownership predictions are mostly consistent regardless of the segmentation (ground-truth or ours) to which they are applied. However, row 3 shows this is not the case for [1]; here, its predicted correct ownership for the lower boundary relies on averaging out over a large ground-truth background region.

and 70% boundary ownership accuracy on ground-truth and automatic segmentation, respectively, compared to 58% and 62% for [1].

References

[1] M. Maire. Simultaneous segmentation and figure/ground organization using angular embedding. *ECCV*, 2010.

[2] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. *ICCV*, 2001.