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

Liangjie Cao
6 June 2018

1. Affinity Learning

Supervised traning of the authors system proceeds from a collection of images and associated ground-truth, {(I_0,S_0,R_0),(I_1,S_1,R_1),...}. Here, I_k is an image defined on domain $\Omega_k \subset N^2$. $S_k : \Omega_k \rightarrow N$ is a segmentation mapping each pixel to a region id, and R_k: $\Omega_k \rightarrow R$ is an rank ordering of pixels according to figure/ground layering. This data defines ground-truth pairwise relationships:

$$\bar{b}_k(p,q) = 1 - \delta(S(p) - S(q)) \tag{1}$$

$$\bar{f}_k(p,q) = (sign(R(q) - R(p)) + 1)/2) \tag{2}$$

As $f(p,q)$ is a conditonal probability, the authors generate training examples $\bar{(f)}_k(p,q)$ for pairs (p,q) satisfying $\bar{b}_k(p,q)$= 1.

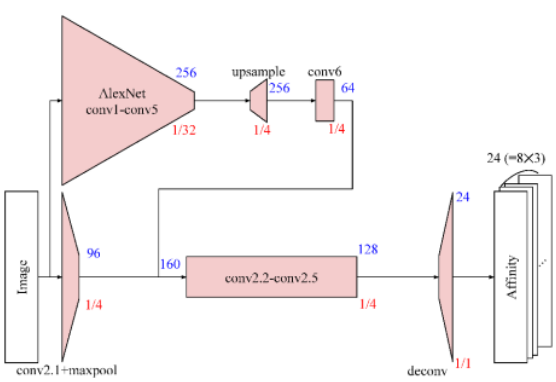

Figure 1. **Deep Affinity Network**

Choosing a CNN to implement these predictors, they regard the problem as mapping an input image to a 48 channel output over the same domain. They adapt prior CNN designs for predicting output quantities at every pixel [1, 2, 5] to their somewhat higher-dimensional prediction task. Specifically, they reuse the basic network design of [5], which first passes a large-scale coarse receptive field through an AlexNet [3]-like subnetwork. It appends this subnetwork's output into a second scale subnetwork acting on a finer receptive field. Figure 1 provides a complete layer diagram. In modifying [5], they increase the size of the penultimate feature map as well as the output dimensionality.

2. Experiments

Training their system for the generic perceptual task of segmentation and figure/ground layering requires a dataset fully annotated in this form. While there appears to be renewed interest in creating lager-scale dataset with such annotation [6], none has yet been released. The following subsetions detail, how, even with such scarcity of training data, their system achieves substantial improvements in figure/ground quality over prior work.
Figure 2 illustrates their method for overcoming this limitation. Given perfect(*e.g.* ground-truth) short-range predictions as input, Angular Embedding generates an extremely high-quality global figure/ground estimate. In a real seting, we want robustness by having many estimates of pairwise relations over many scales. Ground-truth short-range connections suffice as they are perfect estimates. They use the globalized ground-truth figure/ground map as their training signal R in Equation 1. The usual ground-truth segmentation serves as S in Equation 2.

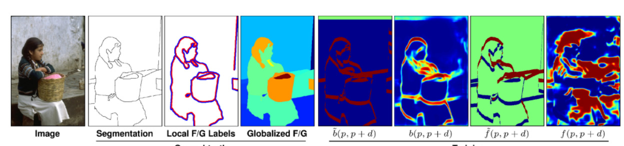

Figure 2. **Affinity learning for segmentation and figure/ground**

3. BSDS:Figure/Ground Benchmark

Table 1 quantitatively compares our ?gure/ground predictions and those of [4] against ground-truth ?gure/ground on our 50 image test subset of BSDS [5]. We consider both projection onto ground-truth segmentation and onto our own systems 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. Figure 3 and the supplementary material provide visual comparisons.

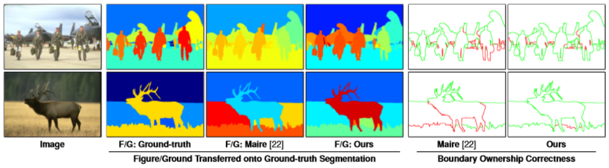

Figure 3. **Figure/ground prediction accuracy measured on ground-truth segmentation**

Ground-truth	R-ACC	B-ACC	B-ACC-50	B-ACC-25
F/G: Ours	0.62	0.69	0.72	0.73
F/G: Maire [4]	0.56	0.58	0.56	0.56

Table 1. **Figure/ground benchmark results**

References

[1] P. Arbelaez. Depth map prediction from a single image using a multi-scale deep network. In *NIPS*, 2014.

[2] P. Arbelaez. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *CVPR*, 2014.

[3] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012.

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

[5] T. Narihira, M. Maire, and S. X. Yu. Direct intrinsics: Learning albedo-shading decomposition by convolutional regression. In *ICCV*, 2015.

[6] Y. Zhu, Y. Tian, D. Metaxas, and P. Dollar. Semantic amodal segmentation. In *CVPR*, 2017.