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

Liangjie Cao
2 June 2018

Abstract

Spectral embedding provides a framework for solving perceptual organization problems, including image segmentation and figure/ground organization. From an affinity matrix describing pairwise relationships between pixels, it clusters pixels into regions, and, using a complex-valued extension, orders pixels according to layer. We train a convolutional neural network (CNN) to directly predict the pairwise relationships that define this affinity matrix. Spectral embedding then resolves these predictions into a globally-consistent segmentation and figure/ground organization of the scene. Experiments demonstrate significant benefit to this direct coupling compared to prior works which use explicit intermediate stages, such as edge detection, on the pathway from image to affinities. Our results suggest spectral embedding as a powerful alternative to the conditional random field (CRF)-based globalization schemes typically coupled to deep neural networks.

1. Introduction

Systems for perceptual organization of scenes are commonly architected around a pipeline of intermediate stages. This trend holds even in light of rapid advancements from designs centered on convolutional neural networks (CNNs). Pure CNN approaches for depth from a single image do focus on directly constructing the desired output [4, 3]. However, these works do not address the problem of perceptual grouping without fixed semantic classes. Training the CNN with a target appropriate for the inference procedure eliminates the need for hand-designed intermediate stages such as edge detection. Our strategy parallels recent work connecting CNNs and conditional random fields (CRFs) for semantic segmentation [2]. A crucial difference, however, is that we handle the generic, or class independent, image partitioning problem. In this context, spectral embedding, and specifically Angular Embedding (AE), is a more natural inference algorithm. Figure 1 illustrates our architecture. Angular Embedding, an extension of the spectral relaxation of Normalized Cuts to complex-valued affinities, provides a mathematical framework for solving joint grouping and ranking problems. Previous works established this framework as a basis for segmentation and figure/ground organization as well as object-part grouping and segmentation. The occlusion layering interpretation of figure/ground is the one most likely to be portable across datasets; it corresponds to a mid-level perceptual task. We find this to be precisely the case for our learned model. Trained on BSDS, it generates quite reasonable output when tested on other image sources, including the PASCAL VOC dataset [5].

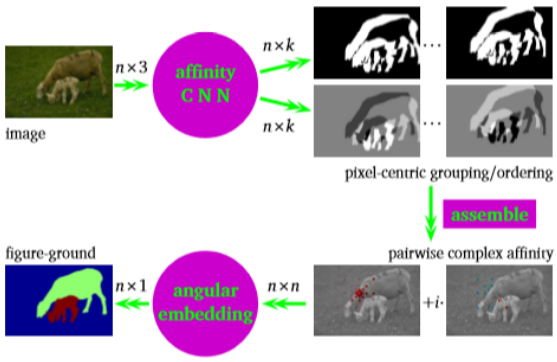

Figure 1. **System architecture.**

2. Spectral Embedding & Generalized Affinity

They abstract the figure/ground problem to that of assigning each pixel ρ a rank $\theta(\rho)$, such that $\theta(.)$ orders pixels by occlusion layer. Assume we are given estimates of the relative order θ(p,q) between many pairs of pixels p and q. The task is then to find $\theta(.)$ that agrees as best as possible with these pairwise estimates. For θ everywhere zero (W = C), this eigenproblem is identical to the spectral relaxation of Normalized Cuts, in which the second and higher eigenvectors encode grouping [1]. With nonzero entries in θ, the first of the now complex-valued eigenvectors is nontrivial and its angle encodes rank ordering while the subsequent eigenvectors still encode grouping.

References

[1] P. Arbelaez, M. Maire, C. C. Fowlkes, and J. Malik. Contour detection and hierarchical image segmentation. In *PAMI*, 2011.

[2] L. C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. Contour detection and hierarchical image segmentation. In *ICLR*, 2015.

[3] D. Eigen and R. Fergus. Predicting depth, surface normals and semantic labels with a common multi-scale convolutional architecture. In *CVPR*, 2015.

[4] D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. In *NIPS*, 2011.

[5] X. He and A. Yuille. Occlusion boundary detection using pseudo-depth. In *ECCV*, 2010.