1. A differentiable, end-to-end trainable framework for solving pixel-level grouping problems.

2. [Spherical Embedding Module] Regressing pixels into a hyperspherical embedding space so that pixels from the same group have high cosine similarity while those from different groups have similarity below a specified margin. The choice of embedding dimension and margin is theoretically related to the problem of distributing points uniformly on the sphere.

3. [Recurrent Grouping Module] Recurrently grouping pixels through a variant of mean-shift clustering module implemented as a recurrent neural network parameterized by kernel bandwidth, which is differentiable and enjoying convergent dynamics and probabilistic interpretability.

4. Backpropagating through the recurrent grouping module allows learning to focus on correcting embedding errors that won’t be resolved during subsequent clustering.

5. The overall framework is not only conceptually simple and theoretically abundant, but also practically effective and computationally efficient. We demonstrate substantial improvements over state-of-the-art instance segmentation for object proposal generation, as well as demonstrating the benefits of grouping loss for classification tasks such as boundary detection and semantic segmentation.

Experiments

Proposal detection on VOC2017
Segmented object proposals measured by Average Recall at IoU from 0.5 to 0.95 and step size as 0.5. Average Recall IoU @ [0.5, 0.95] vs. Average per image

Semantic Segmentation on VOC2017

Semantic Instance Segmentation on VOC2017

Recurrent Mean-Shift Grouping

Using von Mises-Fisher distribution $K(x, x') \propto exp(dx, x')$ for non-parametric density estimation.

Margin Selection
Relating to Tammes’ problem or the hard-spheres problems -- maximizing the smallest distance among n points on a sphere.

It is a non-trivial problem to establish a tight analytic bound (margin $\alpha$) for points in a higher dimension embedding space. We adopt a safe (trivial) strategy. For $n$ instances embedded in $n / 2$ dimensions, one can use value of $\alpha = 0.5$ which allows for zero loss by placing a pair of groups antipodally along each of the $n / 2$ orthogonal axes.

Proposal detection measured by Total Average Recall (AR) at IoU=0.5 and various number of proposals per image.

Semantic segmentation measured by Average Recall (AR) at IoU=0.5 and various number of proposals per image.

Semantic instance segmentation measured by Average Recall (AR) at IoU=0.5 and various number of proposals per image.

Acknowledgements
This work was supported by NSF grants DBI-1262547, IIS-1253538, hardware donation from NVIDIA.

Please see the demo and code at https://www.ics.uci.edu/~skong2/SMMMSG.html.