Modeling Image Segmentation as Epidemic Spread in an Interdependent Network

Connor Greenwell, Emory Hufbauer
Computer Science Dept.
University of Kentucky

Abstract

We propose formulating the problem of image segmentation as a form of phenomena spread on an interdependent network of superpixels. We define two independent metrics for superpixel similarity and use them to generate two unique, but connected topologies describing the image. We then model image segmentation as the spread of the phenomenon of similarity, using two different propagation models. We then extensively optimize these models against the PASCAL VOC training dataset. Finally we compare our method against a clustering-based baseline algorithm operating on the same superpixelizations.

1 Introduction

Image segmentation is an important and popular topic in computer vision [7, 6, 11]. It is important to note that image segmentation is a distinct but related task to object detection. Notably, image segmentation is not concerned with *what* the objects in an image might be, only *where* they are. Typically this is formulated as assigning a unique label to each pixel in an image, where the label indicates belonging to a particular object. For example, an image containing two cats would be expected to have three labels: one for each cat, and a background label. The current state of the art uses convolutional neural networks.

Interdependent network theory is the branch of network theory devoted to studying the behavior and properties of networks composed of two or more subnetworks, each with different topology or complex interactions. Interdependent network theory is commonly used for analyzing the propagation of phenomena like power failure or social influence through complex networks such as cyber-physical systems and social networks. We propose an interdependent network model of image topology, and explore the use of propagation models to segment the image.

We first create an interdependent network model of an image. The weights of edges between adjacent pixels in the lattice represent their degree of similarity to each other. The weights of edges between nodes in the overlay graph represent their overlap and estimated potential to belong to the same object. The weights of edges between the overlay and lattice represent ownership of pixels by superpixels. We then perform a simultaneous, competitive propagation of multiple phenomena through this network using a variety of models, with the ultimate goal of developing a propagation model with property that, after propagation has completed, the regions of the lattice affected by each phenomena correspond well to the segments of the image.

The following paper is laid out as follows: In Section 1.1 we survey the current state of the art in image segmentation, as well as methods related to ours. In Section 2 we describe our method in detail. In Section 3 we describe the metrics by which we judge our methods and those we compare against. In Section 3.1 we describe the benchmark datasets against which we compare. Finally, in Section 4 we describe our resulting performance and evaluation.

1.1 Related Work

[12] establishes theory for talking about epidemic spread in interdependent graphs. [8, 10] deals more specifically with epidemic spread in complex networks. [2, 5] deals more broadly with the dynamics of interdependent graphs.

[9] use a Markov-Random-Field on pre-computed super pixels to perform image segmentation. [7, 6, 11] present a variety of neural network based approached to instance segmentation on natural images.

2 Approach

2.1 Image Network Extraction

To construct our interdependent network model of image topology, first we apply an established superpixelization algorithm (specifically SLIC0) to the image, clustering nearby pixels together into contiguous regions.

Because superpixelization groups similar pixels together, we can approximate the topology of the underlying image using the superpixel network, greatly simplifying future operations.

First we encode an input image as a pair of identical graphs where each node is represents a pixel, and edges represent adjacency between pixels in the image. We create an interdependent network model of an image. The weights of edges between adjacent pixels in each graph represents their degree of similarity to each other, which we describe in the next subsection. Since pixels are represented by nodes in both graphs, we connect their respective nodes with an interdependence link with a uniform weight.

We then perform a simultaneous, competitive propagation of multiple phenomena through this network using a variety of models, with the ultimate goal of developing a propagation model with property that, after propagation has completed, the regions of the lattice affected by each phenomena correspond well to the segments of the image. This produces a proposal segmentation for each image pixel, which is defined as a per-pixel likelihood of belonging to the same object as the source pixel.

Finally, each proposed segmentation is thresholded to produce a binary mask which we combine with a naive segmentation merging algorithm which iteratively applies a logical OR to each successive pair of proposed segmentations.

2.1.1 Alpha superpixel similarity metric

We then define two independent metrics for the similarity of a pair of superpixels. They are ultimately used to generate the topologies of the respective layers of the interdependent network model, so their mutual independence ensures the dissimilar topology of the network layers.

The first metric (the alpha metric) measures the difference in average value (or "brightness") between a pair of superpixels. Initially, each pixel in the image is represented as an RGB color vector x. The alpha metric first computes the euclidean norm of each pixel's vector, and normalizes the result to the range $[0, 1]$. The result is a measure of the pixel's brightness which is independent of its hue or saturation:

$$\alpha_p(\underline{x}) = \frac{\underline{x} \cdot \underline{x}}{\sqrt{3}} = \frac{1}{\sqrt{3}}(\underline{x}_r^2 + \underline{x}_g^2 + \underline{x}_b^2)$$

For a pair of superpixels X and Y, we then define the alpha metric to be the normalized reciprocal of the absolute value of the difference between the average values of the pixels in each:

$$\alpha(X,\, Y) = \frac{2}{1 + |\bar{X}_\alpha - \bar{Y}_\alpha|} - 1$$

$$\bar{X}_\alpha = \sum_{\underline{x}\,\in\, X} \alpha_p(\underline{x}),\ \bar{Y}_\alpha = \sum_{\underline{y}\,\in\, Y} \alpha_p(\underline{y})$$

The result is a real value in the range $[0, 1]$, with lower values representing less similar superpixels.

2.1.2 Beta superpixel similarity metric

The second (or "beta") superpixel similarity metric instead rejects value information entirely, and measures the similarity in saturation and hue between two superpixels. First, it subtracts the vector $[0.5, 0.5, 0.5]$ from the each pixel's color vector, and then divides it by it's own magnitude:

$$\beta_p(\underline{x}) = \frac{\underline{x} - [0.5, 0.5, 0.5]}{|\underline{x} - [0.5, 0.5, 0.5]|}$$

This effectively transforms the color space from the cube between $[0, 0, 0]- > [1, 1, 1]$, to the hollow sphere with radius 1 and centered at the origin. Points on the surface of the sphere represent unique saturation/value combinations. We then average and renormalize these vectors for each superpixel, and take the dot product between them. Because their magnitudes are unit, the dot product effectively measures the cosine of the average angle between the two superpixels in saturation/value space. We finally normalize to the range $[0, 1]$:

$$\beta(X,\, Y) = \frac{1}{2}[1 + \bar{X}_\beta \cdot \bar{Y}_\beta]$$

$$\bar{X}_\beta = \sum_{\underline{x}\,\in\, X} \beta_p(\underline{x}),\ \bar{Y}_\beta = \sum_{\underline{y}\,\in\, Y} \beta_p(\underline{y})$$

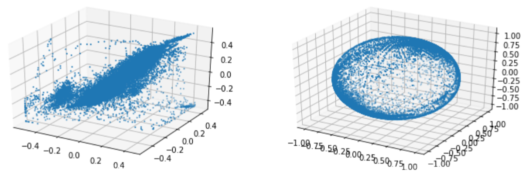


Figure 1: The color space before (left) and after (right) transformation to the sphere

2.1.3 Constructing the network

Next, for each layer we construct the adjacency network over the set of superpixels, and assign weights to its edges based on the above metrics. The result is an α layer and a β layer, which share the same structure, but have independent weightings.

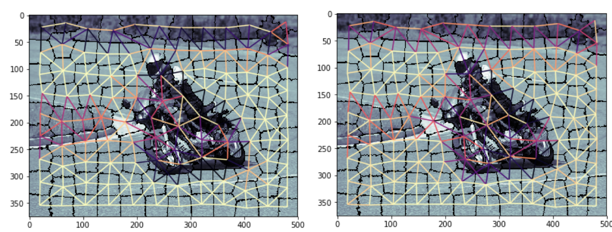


Figure 2: The two layers (α left, β right) of our network model for a sample image

The two layers are then connected one-to-one to form a single network model.

2.2 Propagation

Finally, we are able to perform propagation. We used two models: a markov random-walk model, and a weighted SI-model.

2.3 Baseline

A naive baseline method has been developed for us to compare our actual method against. It is based on using DBSCAN [3] to cluster and merge superpixels produced by SLIC [1].

3 Evaluation

As a preprocessing step, we pre-segment each image into superpixels using the SLIC-Zero parameter-free superpixelization method [1]. We hypothesize that this does not significantly degrade our results.

The goal of this project is mostly to explore the space and point to future research possibilities. Although the results are compared with those of existing algorithms, they are not expected to be competitive with the state of the art. To that end we evaluate our performance on a classic image segmentation benchmark dataset, PASCAL VOC, which contains 2,913 full resolution images and per-pixel instance segmentations.

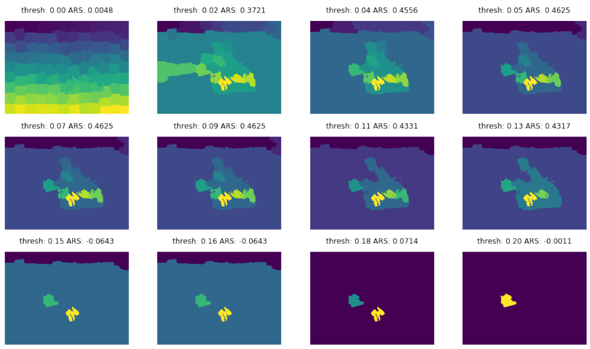
We evaluate two different methods for simulating information propagation in our proposed interdependent graph. The first is based on random walks which we simulate with Markov chains. We experiment with different walk horizons. The second is based on an approximation of epidemic spread.

3.1 Dataset

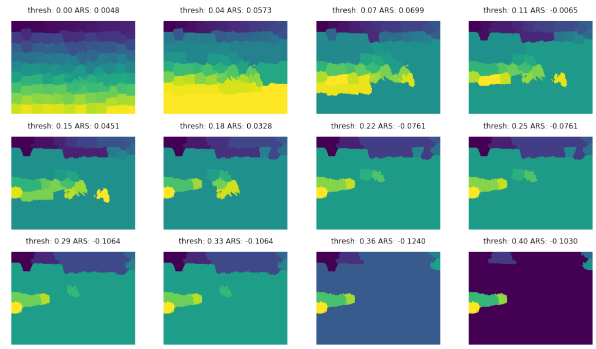
The dataset upon which we evaluate our method is the PASCAL Visual Object Classification (VOC) dataset [4]. Over the years in which the VOC challenge was run, a couple different variations of the problem we presented, ranging from whole-image object classification to human pose estimation. In 2012 they introduced a subset of per-pixel image segmentations for 2,913 of the images in the dataset. In each segmentation, there is a background class, some number of instance classes, and a "don't care" region around each object instance. The image segmentation task is distinct from object classification in that the goal is not to actually classify to which object category each pixel belongs to, but to separate the image into its constituent parts, similar to foreground/background segmentation. We believe that this task presents a unique challenge that can be addressed by our proposed method of propagating information through a interdependent graph defined by pixel/region similarities.

4 Results and Discussion

We evaluate our methods on the Adjusted Rand Score (ARS) from [13]. In Figure 4 we visualize the distribution of ARS for each of our described methods (baseline, Markov-based, and Epidemic-based). In Table 1 we show the average ARS of each method on the PASCAL VOC dataset. Both of the methods proposed in this paper perform worse than the baseline; quantitatively, qualitatively, and in terms of computation time.



(a) Sample segmentation using only the alpha scores.



(b) Sample segmentation using only the beta scores.

Figure 3: Each of our proposed similarity metrics provides enough information to perform rudimentary segmentation. To demonstrate this we take a simple thresholding approach, cutting edges below a given threshold, shown above each image. We observe that in this setting, alpha scores seem useful for differentiating between regions in the foreground, and beta scores seem useful for regions in the background. Our intuition is that combining these two metrics will yield superior segmentations.

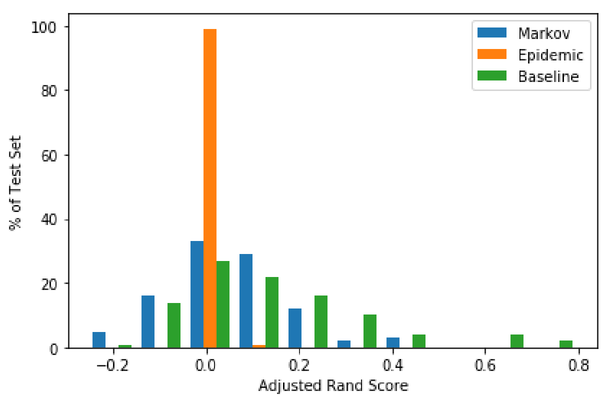


Figure 4: Comparison of Adjusted Rand Score on 100 images from the training set. Having more area to the right of the graph indicates a higher proportion of high quality segmentations.

Qualitatively, segmentations produced by the Markov-based and Epidemic-based methods leave much to be desired. Figure 5 shows the images and resulting segmentations for the five highest scoring images under each method. The Markov-based segmentations are noisy and typically centered over points of interest in an image. This noise can be attributed to the fact that the likelihood of arriving in a particular node during a random walk is not uniform when starting in other nodes in a single object.

Each method described is subject to a number of hyper-parameters which ultimately dictate the methods performance with regards to ARS. In Figure 6 we show the results of a hyper-parameter sweep over both the Markov-based and Epidemic-based methods. They share the a parameter controlling the threshold at which each proposed segmentation is binarized at. Interestingly, each method demonstrates its own sensitivity to the parameter. Specifically, the Epidemic model is relatively insensitive to it. We attribute this to a quirk of the Epidemic-based method, proposal segmentations that begin in areas with low neighboring similarity will not spread far. Due to the way that we combine proposed segmentations, this produces many smaller, less informative segments. Effectively we are amplifying noise in the signal. In Figure 7 we demonstrate what happens when we use a simple heuristic to select the best proposals.

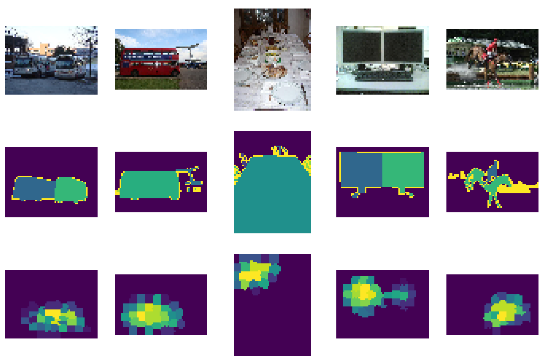
Table 1: Mean Adjusted Rand Score for each method covered in this paper. Higher values indicate higher quality segmentations.

Method	Mean ARS
Baseline	0.1385439
Markov-based	0.0536634
Epidemic-based	0.0124122

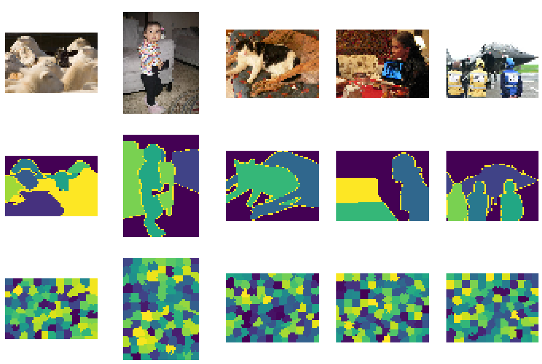
5 Conclusion

In this document we presented a method for performing instance segmentation on photographs by model-

(a) Baseline



(b) Markov random walk based



(c) Epidemic based

Figure 5: Sample results from each of our proposed methods.

ing instance membership as an information spread or epidemic model in an interdependent graph. We introduced two simple similarity metrics which are used as the basis for transmitting instance membership between neighboring pixels. Finally we compare our method to a rudimentary baseline.

5.1 Future Work

These methods did not perform better than the proposed baseline. Some possible directions for future work include the following:

- Improving the method by which we combine candidate segmentations.
- Incorporating additional similarity metrics, and additional layers to the interdependent graph.
- Explore the effect of using other superpixel segmentation strategies.

References

[1] Radhakrishna Achanta, Appu Shaji, Kevin Smith, Aurelien Lucchi, Pascal Fua, and Sabine Süsstrunk. Slic superpixels. Technical report, 2010.

[2] Stefano Boccaletti, Ginestra Bianconi, Regino Criado, Charo I Del Genio, Jesús Gómez-Gardenes, Miguel Romance, Irene Sendina-Nadal, Zhen Wang, and Massimiliano Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544(1):1–122, 2014.

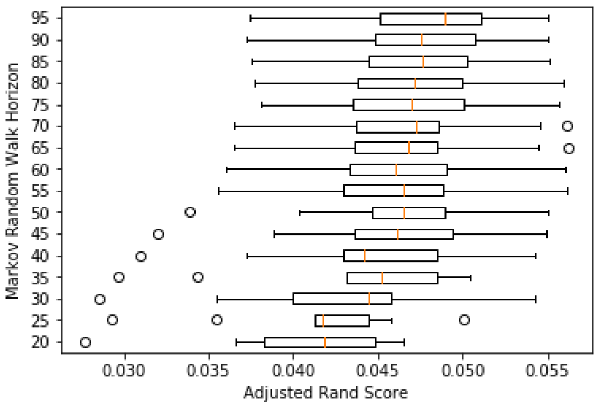
[3] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm for discovering clusters in large spatial databases with noise.

[4] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn, and A. Zisserman. The pascal visual object classes (voc) challenge. *International Journal of Computer Vision*, 88(2):303–338, June 2010.

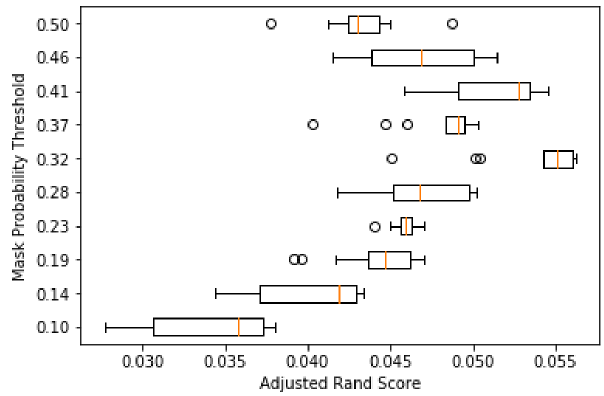
[5] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. Multilayer networks. *Journal of complex networks*, 2(3):203–271, 2014.

[6] Yi Li, Haozhi Qi, Jifeng Dai, Xiangyang Ji, and Yichen Wei. Fully convolutional instance-aware semantic segmentation. In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pages 2359–2367, 2017.

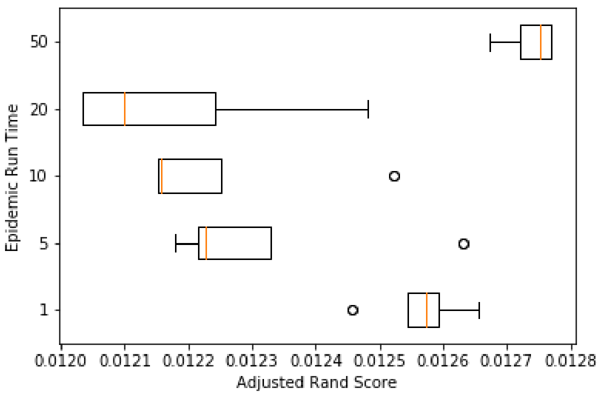
[7] Alejandro Newell, Zhiao Huang, and Jia Deng. Associative embedding: End-to-end learning for joint detection and grouping. In *Advances in*



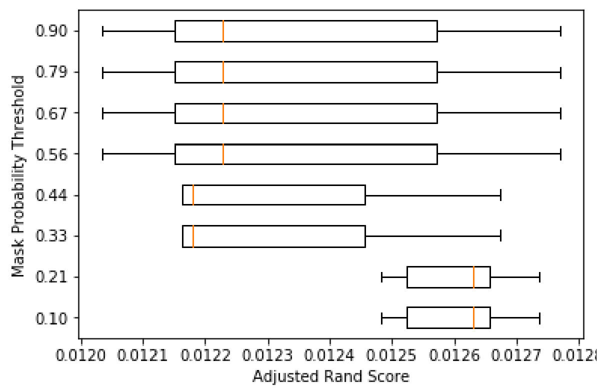
(a) Markov random walk, walk horizon



(b) Markov random walk, merge threshold



(c) Epidemic, simulation run-time



(d) Epidemic, merge threshold

Figure 6: Analysis of performance under different hyper-parameters for each of our proposed methods. (a) and (b) show that threshold is the most important parameter for the Markov-based method. (c) and (d) show that simulation run time is the most important parameter for the Epidemic-based method.

Neural Information Processing Systems, pages 2274–2284, 2017.

[8] Romualdo Pastor-Satorras, Claudio Castellano, Piet Van Mieghem, and Alessandro Vespignani. Epidemic processes in complex networks. *Reviews of modern physics*, 87(3):925, 2015.

[9] Soo-Chang Pei, Wen-Wen Chang, and Chih-Tsung Shen. Saliency detection using superpixel belief propagation. In *Image Processing (ICIP), 2014 IEEE International Conference on*, pages 1135–1139. IEEE, 2014.

[10] Lorenzo Pellis, Frank Ball, Shweta Bansal, Ken Eames, Thomas House, Valerie Isham, and Pieter Trapman. Eight challenges for network epidemic models. *Epidemics*, 10:58–62, 2015.

[11] Mengye Ren and Richard S Zemel. End-to-end instance segmentation with recurrent attention.

[12] Seung-Woo Son, Golnoosh Bizhani, Claire Christensen, Peter Grassberger, and Maya Paczuski. Percolation theory on interdependent networks based on epidemic spreading. *EPL (Europhysics Letters)*, 97(1):16006, 2012.

[13] Ranjith Unnikrishnan, Caroline Pantofaru, and Martial Hebert. A measure for objective evaluation of image segmentation algorithms. In *Computer Vision and Pattern Recognition-Workshops, 2005. CVPR Workshops. IEEE Computer Society Conference on*, pages 34–34. IEEE, 2005.

Figure 7: Binarized segmentations proposed by the epidemic-based method, sorted by our simple heuristic.