Note: Deep Spectral Methods: A Surprisingly Strong Baseline for Unsupervised Semantic Segmentation and Localization

sgc

June 19, 2022

1 Introduction

Instead of using pure deep learning in Unsupervised localization and segmentation like other works do, this paper is inspired from traditional spectral segmentation methods by reframing image decomposition as a graph partitioning problem. This work examines the eigenvectors of the Laplacian of a feature affinity matrix from self-supervised networks. The eigenvectors of the Laplacian of the affinity matrix already decomposed the image into meaningful parts. Surprisingly outperformed the SOTA.
The proposed method first utilizes a self-supervised network to extract dense features corresponding to image patches.
And theoretically authors find that eigenvectors of the Laplacian of weighted affinity graph directly correspond to semantically meaningful image regions. Notably, the eigenvector with the smallest nonzero eigenvalue generally corresponds to the most prominent object in the scene.t convert the eigensegments into discrete image regions by thresholding and associate each region with a semantic feature vector from the network.

2 Method

2.1 The choice of W

The adjacency matrix W is generated through KNN:

$$W_{\text{knn}}(u, v) = \begin{cases} 1 - \|\psi(u) - \psi(v)\|, & u \in \text{KNN}_\psi(v), \\ 0, & \text{otherwise}, \end{cases}$$

2.2 Semantic Spectral Decomposition

The model begin by extracting deep features $f = \phi(I) \in RC \times M/P \times N/P$ using a network ϕ. In this paper authors use the transformer and P is the patch size. Additionally features from the keys of the last attention layer work especially well.
The feature affinity map contains coarse information, so this method combines it with low level information from features of the 0th layer of the network(color-level).
The color affinity matrix, the paper uses the sparse KNN-matting matrix using the formula mentioned.
Then take the eigenvectors of its Laplacian $L = D^{1/2}(DW)D^{1/2}$ to decompose an image into soft segments: $\{y_0, , y_{n1}\} = eigs(L)$.

2.3 Semantic Segmentation

Step1: Discretize the first m eigenvectors $y_1, , y_m$ of L by clustering them across the eigenvector dimension using K-means clustering(over-cluster)
Step2: Take a crop around each segment, and compute its feature vector fs using the self-supervised transformer.
Step3: Cluster the set of all feature vectors f_s across all images using K-means clustering with k clusters. This step also assign labels.
At last, train a standard segmentation model with a self-supervised backbone using the segmentation obtained above as pseudo labels.

3 Summary

This method take use of the spectral graph theory which is commonly used in pre deep learning era. The eigenvectors of image already have good semantic information. With this information this method achieved new SOTA.

References