We propose a deep learning architecture for unsupervised image segmentation, which contains two stages:

First stage, a Superpixelwise Autoencoder (SuperAE) is designed to learn the deep embedding and reconstruct a smoothed image, then the smoothed image is passed to generate superpixels.

Second stage, we present a novel clustering algorithm called Deep Superpixel Cut (DSC), which measures the deep similarity between superpixels and formulates image segmentation as a soft partitioning problem.

The real-world image usually contains noise, which would affect the segmentation and needs to be smoothed. We observe that pixels within a superpixel are more continuous in color space, so we regard this as constraint in reconstruction.

Our SuperAE formulation:

\[ \mathcal{L}_r = \|x - x^R\|^2 + \lambda \sum_{i=1}^{N} \sum_{j=1}^{M} t_{ij} \|x_i^R - v_j\|^2 \]

\[ v_j = \frac{\sum_{i=1}^{N} t_{ij} x_i}{\sum_{i=1}^{N} t_{ij}} \] (Superpixel RGB vector)

SuperAE aims to minimize the variance of pixel features within a superpixel, thus smoothing the regions.

Regard the segmentation as a pixelwise K-class classification problem by generate pseudo-labels.

\[ \mathcal{L}_1 = -\sum_{n=1}^{N} \sum_{k=1}^{K} y_{nk} \log(p_k^n) \] (Cross-entropy Loss)

Regard the segmentation as a superpixelwise graph-partition problem, based on their deep similarity.

\[ \mathcal{L}_2 = \sum_{k=1}^{K} \sum_{a \in A_k, b \notin A_k} W(a, b) \]

\[ = \sum_{k=1}^{K} \sum_{a \in A_k} \sum_{b \in V \setminus A_k} \mathcal{Q}_{ab}(1 - Q_k) \]

\[ = \sum_{k=1}^{K} Q_k W(1 - Q_k) \] (Differential) Intuition: If two superpixel a and b are more similar in deep features, then the probability they belong to different partitions should be lower.

Loss function

\[ \mathcal{L}_s = \alpha \mathcal{L}_1 + \beta \mathcal{L}_2 \]

\[ = \alpha \sum_{n=1}^{N} \sum_{k=1}^{K} y_{nk} \log(p_k^n) + \beta \sum_{k=1}^{K} Q_k W(1 - Q_k) \]

Evaluate the effectiveness of SuperAE and DSC on different superpixel algorithms.

REFERENCES