Deep Superpixel Cut for Unsupervised Image Segmentation

Qinghong Lin$^{1,2}$, Weichan Zhong$^1$, JiangLin Lu$^{1,2}$
Shenzhen University$^1$, China
Kazan Federal University$^2$, Russia
Outline

- Background
- Motivation
- Method
- Experiment
- Conclusion
Background

- What is Image segmentation?
  - Dividing an image into regions, assigning a class to every pixel in the image

Fig 1. Image segmentation
Traditional method

• **Over segmentation**
  – Groups pixels into regions (**superpixels**) with similar characteristics (color, texture)

• **Requirements**
  – boundary adherence, compactness, effectively...

**Fig 2. Over segmentation**
Traditional method

• Related Work
  – Normalized Cut[1]

\[ N\text{cut}(A, B) = \frac{cut(A, B)}{assoc(A, V)} + \frac{cut(A, B)}{assoc(B, V)} \]

  – SLIC[2]:


Deep Learning method

• **Semantic Segmentation**
  – Assign a class label to each pixel

![Fig 4. Semantic segmentation in Cityscapes dataset](image)

– **Requirements**
  • Pixel accuracy, mIoU...
Deep Learning method

• Related work
  – FCN[1]
  – U-Net[2]

• Drawback
  – Require the human annotation, which is hard to collect


Motivation

• Summary

<table>
<thead>
<tr>
<th>Type</th>
<th>Traditional Segmentation</th>
<th>Semantic Segmentation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Level</td>
<td>Superpixel</td>
<td>Pixel</td>
</tr>
<tr>
<td>Advantage</td>
<td>Fully explore image itself attributions (intensity, boundary)</td>
<td>Richer representation(semantic)</td>
</tr>
<tr>
<td>Disadvantage</td>
<td>Hand-crafted feature</td>
<td>Require Human annotation</td>
</tr>
</tbody>
</table>

Tab 1. Summary of those two type methods

• Inspiration
  – Can we make traditional methods and deep learning methods complementary?
  – Adopt Deep learning model for unsupervised image segmentation is worth thinking
Method

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

Fig 7. Proposed architecture
Superpixelwise Autoencoder

• Intuition
  – The real world image usually contain noise, which would affected the segmentation and needs to be smoothed
  – We observe that pixels within a superpixel are more continuous in color space

• SuperAE
  – K-means loss, pixels => samples and Superpixel => center.
  – SuperAE aim to minimize the variance of pixel features within a superpixel, thus smoothing the regions

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

\[
v_j = \frac{\sum_{i=1}^{N} t_{ij} x_i}{\sum_{i=1}^{N} t_{ij}}
\]

\(N\): pixel number
\(M\): Superpixel number
\(x_i\): original pixel i RGB vector.
\(x_i^R\): reconstructed pixel i RGB vector.
\(v_j\): Superpixel j RGB vector.
\(t_{ij}\): Indicator whether pixel i belong to superpixel j.
\(\lambda\): parameter
Deep Superpixel Cut

• Generate Pseudo Labels
  – Regard the segmentation as a **pixelwise K-class classification problem**.
  – Cross-entropy loss but lack of human annotation, so we use a strategy to generate **pseudo-labels**.

$$
L_1 = -\sum_{n=1}^{N} \sum_{k=1}^{K} y_n^k \log(p_n^k)
$$

$c_n$: class for pixel $n$
$C_j$: class for superpixel $j$

$$
c_n = \arg\max_k p_n^k, \quad n \in [0, N)
$$

$$
C_j = \arg\max_k |c_n|, \quad j \in [0, M), n \in S_j
$$

$$
\hat{y}_n^k = 1(c_n = C_j), \quad j \in [0, M), n \in S_j
$$
Deep Superpixel Cut

- Proposed DSC
  - Regard the segmentation as a superpixelwise graph-partition problem, based on their deep similarity.
  - If two superpixel a and b are more similar in deep features, then the probability they belong to different partitions should be lower.

\[ g_j = \frac{\sum_{i=1}^{N} z_{ij} f_i}{\sum_{i=1}^{N} z_{ij}} \]

\[ q_k^i = \frac{\sum_{i=1}^{N} z_{ij} p_k^i}{\sum_{i=1}^{N} z_{ij}} \]

\[ w_{ij} = \begin{cases} 
  \exp\left(-\frac{\|g_i-g_j\|^2}{\sigma^2}\right), & \text{if } \|l(i)-l(j)\|_2 < d \\
  0, & \text{otherwise} 
\end{cases} \]

\[ L_2 = \sum_{k=1}^{K} \sum_{a,b \in V} P(a \in A_k, b \notin A_k) W(a,b) \]

\[ = \sum_{k=1}^{K} \sum_{a,b \in V} q_k^a w_{ab} (1 - q_k^b) \]

\[ = \sum_{k=1}^{K} Q_k^T W (1 - Q_k) \]

**Fig 9. superpixel graph**

- I: Image
- F: Pixel deep representation
- P: Probability of each pixel belong to which partition
- K: partition number
- G: Superpixel deep representation
- Q: Probability of each superpixel belong to which partition
- W: Similarity matrix
- L: DSC loss
Deep Superpixel Cut

• Loss function

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

\[ = \alpha \sum_{n=1}^{N} \sum_{k=1}^{K} \hat{y}_n^k \log(p_n^k) + \beta \sum_{k=1}^{K} Q_k^T W (1 - Q_k) \]

• Optimization

\[ \frac{\partial \mathcal{L}_2}{\partial F} = \frac{\partial \mathcal{L}_2}{\partial Q} \frac{\partial Q}{\partial G} \frac{\partial G}{\partial F} \]

Cross entropy loss L1 is differential.

DSC loss L2 is also differential.

Softmax operation
Experiment

- Dataset: BSDS500[1]
- Metric: SC, RI, VI[2]
- Ablation study

Tab 2. The ablation study of different components

<table>
<thead>
<tr>
<th>IMG</th>
<th>REC</th>
<th>TEM</th>
<th>$L_1$</th>
<th>$L_2$</th>
<th>SC</th>
<th>PRI</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0.405</td>
<td>0.761</td>
<td>3.483</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0.388</td>
<td>0.756</td>
<td>3.622</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0.453</td>
<td>0.769</td>
<td>3.019</td>
<td></td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0.477</td>
<td>0.769</td>
<td>2.471</td>
<td></td>
</tr>
<tr>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>0.488</td>
<td>0.764</td>
<td>2.315</td>
</tr>
</tbody>
</table>

Tab 3. The ablation study of different superpixels

<table>
<thead>
<tr>
<th>Method</th>
<th>SC</th>
<th>PRI</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ODS</td>
<td>OIS</td>
<td>ODS</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLIC</td>
<td>0.354</td>
<td>0.402</td>
<td>0.698</td>
</tr>
<tr>
<td>SuperAE</td>
<td>0.376</td>
<td>0.415</td>
<td>0.710</td>
</tr>
<tr>
<td>DSC</td>
<td>0.428</td>
<td>0.485</td>
<td>0.709</td>
</tr>
<tr>
<td>MS</td>
<td>0.515</td>
<td>0.535</td>
<td>0.771</td>
</tr>
<tr>
<td>SuperAE</td>
<td>0.497</td>
<td>0.550</td>
<td>0.777</td>
</tr>
<tr>
<td>DSC</td>
<td>0.504</td>
<td>0.543</td>
<td>0.739</td>
</tr>
<tr>
<td>EGB</td>
<td>0.458</td>
<td>0.467</td>
<td>0.770</td>
</tr>
<tr>
<td>SuperAE</td>
<td>0.491</td>
<td>0.500</td>
<td>0.774</td>
</tr>
<tr>
<td>DSC</td>
<td>0.506</td>
<td>0.548</td>
<td>0.737</td>
</tr>
</tbody>
</table>

Tab 4. Compare DSC with other clustering algorithm

<table>
<thead>
<tr>
<th>Method</th>
<th>SC</th>
<th>PRI</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ODS</td>
<td>OIS</td>
<td>ODS</td>
</tr>
<tr>
<td>$k$-means, RGB</td>
<td>0.340</td>
<td>0.344</td>
<td>0.708</td>
</tr>
<tr>
<td>$k$-means, Deep</td>
<td>0.375</td>
<td>0.382</td>
<td>0.718</td>
</tr>
<tr>
<td>NCut, RGB</td>
<td>0.414</td>
<td>0.465</td>
<td>0.717</td>
</tr>
<tr>
<td>NCut, Deep</td>
<td>0.386</td>
<td>0.423</td>
<td>0.704</td>
</tr>
<tr>
<td>DSC</td>
<td>0.452</td>
<td>0.505</td>
<td>0.726</td>
</tr>
</tbody>
</table>


Experiment

- Example of different stage

![Fig 10. different stage in our method](image)

- Convergence Curve

![Fig 11. Visualization](image)

- Qualitative Result

![Fig 12. Segmentation Result](image)

- BSDS500 benchmarks

<table>
<thead>
<tr>
<th>Method</th>
<th>SC</th>
<th>PRI</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>ODS</td>
<td>OIS</td>
<td>ODS</td>
</tr>
<tr>
<td>SLIC [10]</td>
<td>0.37</td>
<td>0.38</td>
<td>0.74</td>
</tr>
<tr>
<td>NCuts [4]</td>
<td>0.45</td>
<td>0.53</td>
<td>0.78</td>
</tr>
<tr>
<td>EGB [5]</td>
<td>0.52</td>
<td>0.57</td>
<td>0.80</td>
</tr>
<tr>
<td>MS [21]</td>
<td>0.54</td>
<td>0.58</td>
<td>0.79</td>
</tr>
<tr>
<td>gpB-owt-ucm [8]</td>
<td>0.59</td>
<td>0.65</td>
<td>0.83</td>
</tr>
<tr>
<td>CAE-TVL [22]</td>
<td>0.51</td>
<td>0.56</td>
<td>0.79</td>
</tr>
<tr>
<td>Backprop [13]</td>
<td>0.47</td>
<td>0.50</td>
<td>0.75</td>
</tr>
<tr>
<td>Mumford-Shah [15]</td>
<td>0.49</td>
<td>—</td>
<td>0.71</td>
</tr>
<tr>
<td>W-Net [14]</td>
<td>0.57</td>
<td>0.62</td>
<td>0.81</td>
</tr>
<tr>
<td>DSC</td>
<td>0.56</td>
<td>0.60</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Tab 5. BSDS500 benchmarks
Conclusion

• We propose a deep learning models for unsupervised image segmentation

• To learn the deep embedding, we design a SuperAE, which also smooths the original image and conducive to the superpixels generation.

• For segmentation, we propose a novel clustering method DSC which measures the deep similarity between superpixels and partitions them into perceptual regions by soft association.

• Experiment results on BSDS500 demonstrate the efficacy of our proposed, and our DSC outperforms most of the unsupervised segmentation methods.
Thank for you listening!

Qinghong Lin
linqinghong@email.szu.edu.cn