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

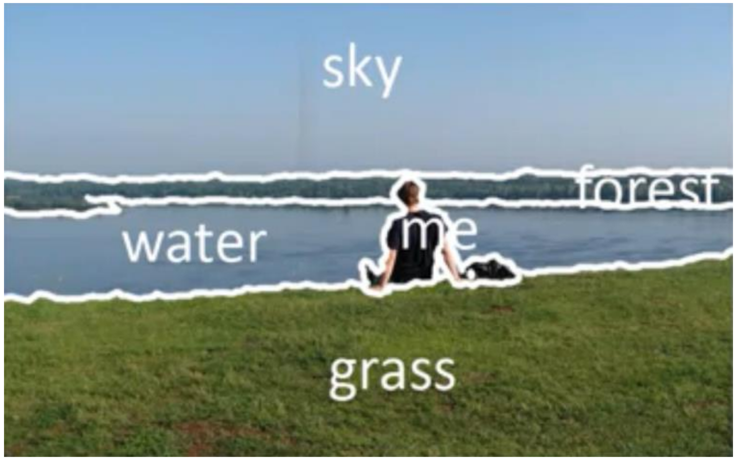


Fig 1. Image segmentation

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]

$$Ncut(A,B) = \frac{cut(A,B)}{assoc(A,V)} + \frac{cut(A,B)}{assoc(B,V)}$$

  - SLIC[2]:

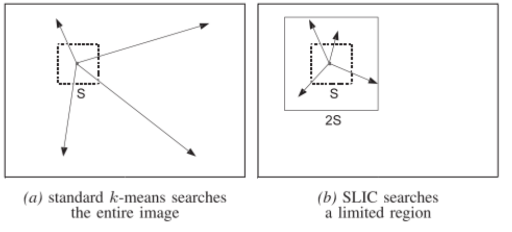


Fig 3. SLIC algorithm

- Drawbacks
  - Hand crafted features

[1]Shi J, Malik J. Normalized cuts and image segmentation[J]. Departmental Papers (CIS), 2000: 107.

[2] Achanta R, Shaji A, Smith K, et al. SLIC superpixels compared to state-of-the-art superpixel methods[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2012, 34(11): 2274-2282.

Kazan Federal UNIVERSITY

Deep Learning method

- Semantic Segmentation
  - Assign a class label to each pixel

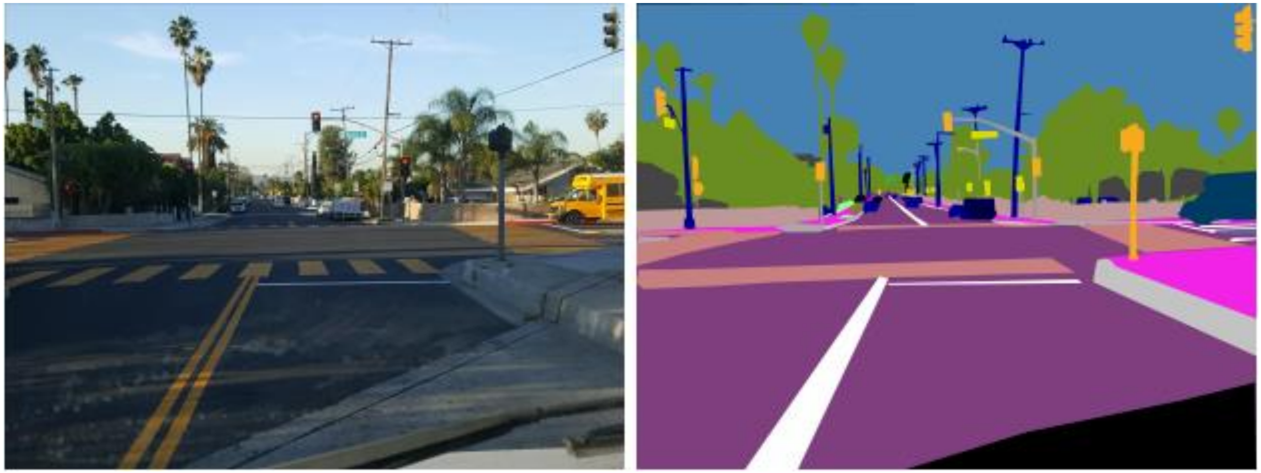


Fig 4. Semantic segmentation in Cityscapes dataset

  - Requirements
    - Pixel accuracy, mIoU...

Deep Learning method

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

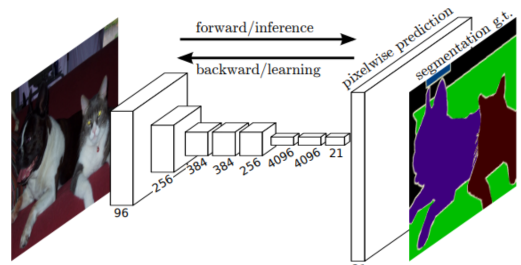


Fig 5. FCN

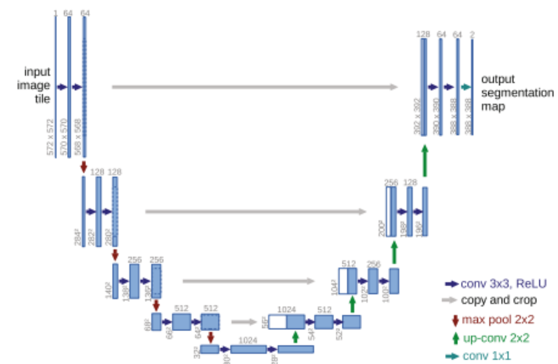


Fig 6. U-Net

- Drawback
  - Require the human annotation, which is hard to collect

[1]Long J, Shelhamer E, Darrell T. Fully convolutional networks for semantic segmentation[C] *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015: 3431-3440.
[2] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C] *International Conference on Medical image computing and computer-assisted intervention*. Springer, Cham, 2015: 234-241.

Motivation

- Summary

Type	Traditional Segmentation	Semantic Segmentation
Level	Superpixel	Pixel
Advantage	Fully explore image itself attributions (intensity, boundary)	Richer representation(semantic)
Disadvantage	Hand-crafted feature	Require Human annotation

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.

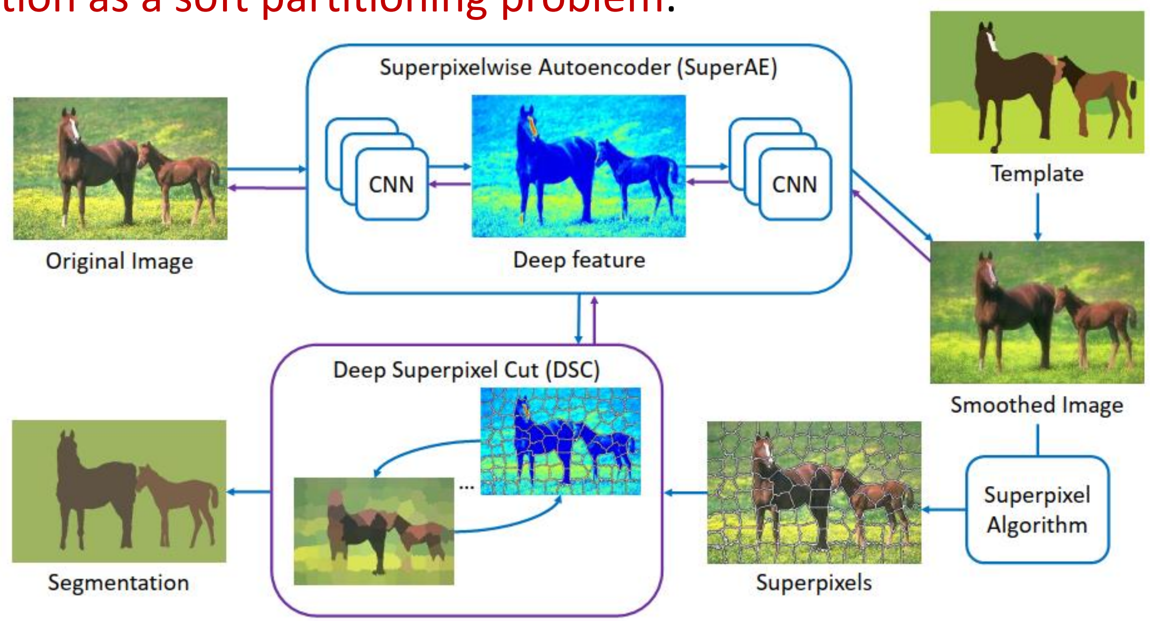


Fig 7. Proposed architecture

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

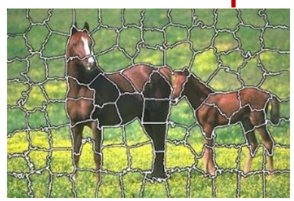


Fig 8. Superpixel example

- 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

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

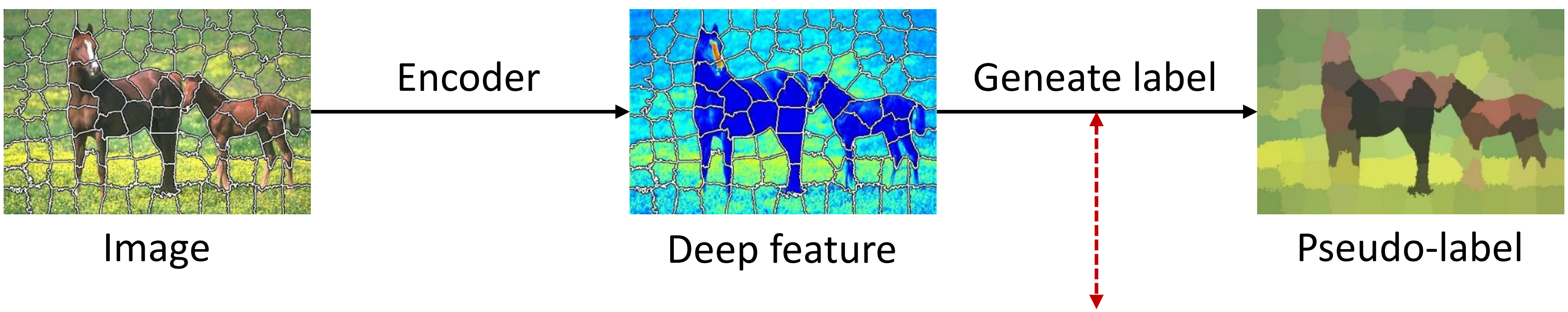
$$v_j = \frac{\sum_{i=1}^{N} t_{ij} x_i}{\sum_{i=1}^{N} t_{ij}}$$

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.
λ: 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.

$$\mathcal{L}_1 = -\sum_{n=1}^{N}\sum_{k=1}^{K} \boxed{y_n^k} \log(p_n^k) \longleftarrow \tilde{y}_n^k \quad \text{Pseudo-label}$$



Image

Deep feature

Pseudo-label

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

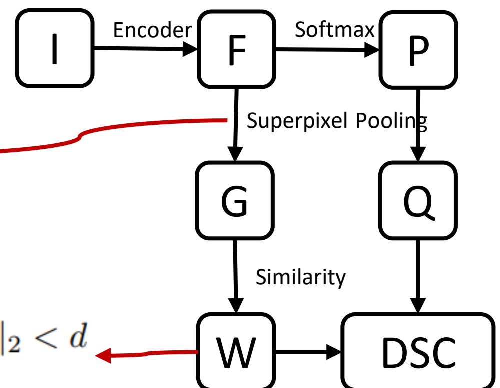
$$\begin{cases} c_n = \operatorname*{argmax}_k p_n^k, & n \in [0, N) \\ C_j = \operatorname*{argmax}_k |c_n| & j \in [0, M), n \in S_j \\ \hat{y}_n^k = \mathbf{1}(c_n = C_j) & j \in [0, M), n \in S_j \end{cases}$$

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.

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

$$q_k^j = \frac{\sum_{i=1}^{N} z_{ij} p_k^i}{\sum_{i=1}^{N} z_{ij}}$$



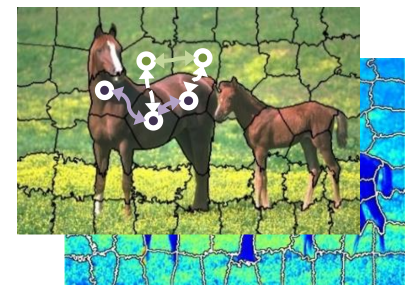


Fig 9. superpixel graph

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

z_{ij}: superpixel indicator
σ: scale factor
l: superpixel distance
d: distance constant

$$\begin{aligned} \mathcal{L}_2 &= \sum_{k=1}^{K} \sum_{a,b \in V} \mathcal{P}(a \in A_k, b \notin A_k) W(a,b) \\ &= \sum_{k=1}^{K} \sum_{a,b \in V} q_k^a w_{ab} (1 - q_k^b) \\ &= \sum_{k=1}^{K} Q_k^T \mathbf{W} (\mathbf{1} - Q_k) \end{aligned}$$

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

$$\mathcal{L}_s = \alpha\mathcal{L}_1 + \beta\mathcal{L}_2$$
$$= \alpha\sum_{n=1}^{N}\sum_{k=1}^{K}\hat{y}_n^k \log(p_n^k) + \beta\sum_{k=1}^{K} Q_k^T\mathbf{W}(\mathbf{1} - Q_k)$$

- Optimization

Cross entropy loss L1 is differential.

DSC loss L2 is also differential.

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

$$\begin{aligned}\frac{\partial\mathcal{L}_2}{\partial Q_k} &= \frac{\partial(Q_k^T\mathbf{W}(\mathbf{1} - Q_k))}{\partial Q_k}\\ &= \frac{\partial(Q_k^T\mathbf{W}\mathbf{1} - Q_k^T\mathbf{W}Q_k)}{\partial Q_k}\\ &= \mathbf{W}\mathbf{1} - \mathbf{W}Q_k - (Q_k^T\mathbf{W})^T\\ &= \mathbf{W}\mathbf{1} - 2\mathbf{W}Q_k\end{aligned}$$

$$\frac{\partial g_j}{\partial f_i}$$

$$\frac{z_{ij}}{\sum_{i=1}^{N} z_{ij}}$$

Softmax operation

Experiment

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

IMG	REC	TEM	$\mathcal{L}_1$	$\mathcal{L}_2$	SC	PRI	VI
✓					0.405	0.761	3.483
✓	✓				0.388	0.756	3.622
✓	✓	✓			0.453	0.769	3.019
✓	✓	✓	✓		0.477	0.769	2.471
✓	✓	✓	✓	✓	0.488	0.764	2.315

Tab 2. The ablation study of different components

Method		SC		RI		VI	
		ODS	OIS	ODS	OIS	ODS	OIS
SLIC	IMG	0.354	0.402	0.698	0.726	2.391	2.306
	SuperAE	0.376	0.415	0.710	0.745	2.360	2.286
	DSC	0.428	0.485	0.709	0.742	2.262	2.079
MS	IMG	0.515	0.535	0.771	0.805	2.450	2.429
	SuperAE	0.497	0.550	0.777	0.795	2.162	2.077
	DSC	0.504	0.543	0.739	0.762	2.014	1.849
EGB	IMG	0.458	0.467	0.770	0.775	2.334	2.325
	SuperAE	0.491	0.500	0.774	0.782	2.190	2.180
	DSC	0.506	0.548	0.737	0.779	1.989	1.853

Tab 3. The ablation study of different superpixels

- Effect of DSC

Method	SC		PRI		VI	
	ODS	OIS	ODS	OIS	ODS	OIS
k-means, RGB	0.340	0.344	0.708	0.732	3.156	3.146
k-means, Deep	0.375	0.382	0.718	0.739	2.865	2.859
NCut, RGB	0.414	0.465	0.717	0.771	2.308	2.213
NCut, Deep	0.386	0.423	0.704	0.754	2.433	2.382
DSC	**0.452**	**0.505**	**0.726**	**0.778**	**2.069**	**2.002**

Tab 4. Compare DSC with other clustering algorithm

[1] Arbelaez P, Maire M, Fowlkes C, et al. Contour detection and hierarchical image segmentation[J]*. IEEE transactions on pattern analysis and machine intelligence*, 2010, 33(5): 898-916.

[2] Pont-Tuset J, Marques F. Supervised evaluation of image segmentation and object proposal techniques[J]. *IEEE transactions on pattern analysis and machine intelligence*, 2015, 38(7): 1465-1478.

Experiment

- Example of different stage

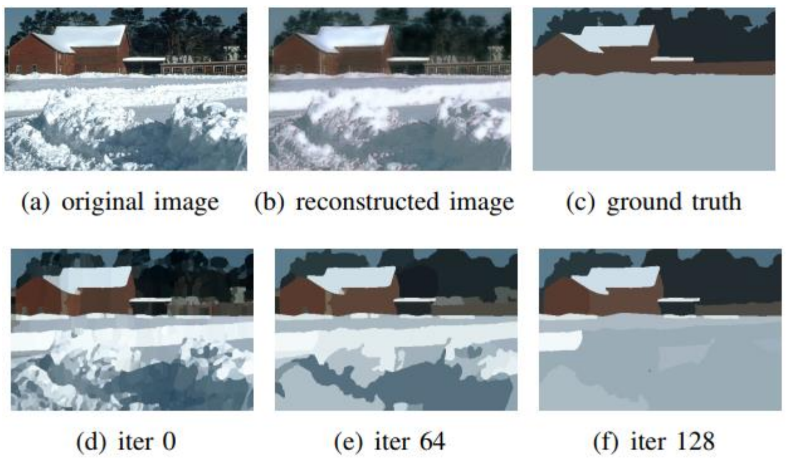


Fig 10. different stage in our method

- BSDS500 benchmarks

Method	SC		PRI		VI	
	ODS	OIS	ODS	OIS	ODS	OIS
SLIC [10]	0.37	0.38	0.74	0.75	2.56	2.50
NCuts [4]	0.45	0.53	0.78	0.80	2.23	1.89
EGB [5]	0.52	0.57	0.80	0.82	2.21	1.87
MS [21]	0.54	0.58	0.79	0.81	1.85	1.64
gPb-owt-ucm [8]	0.59	0.65	0.83	0.86	1.69	1.48
CAE-TVL [22]	0.51	0.56	0.79	0.82	2.11	2.02
Backprop [13]	0.47	0.50	0.75	0.77	2.18	2.15
Mumford-Shah [15]	0.49	–	0.71	–	2.20	–
W-Net [14]	0.57	0.62	0.81	0.84	1.76	1.60
DSC	0.56	0.60	0.80	0.83	1.82	1.62

Tab 5. BSDS500 benchmarks

- Convergence Curve

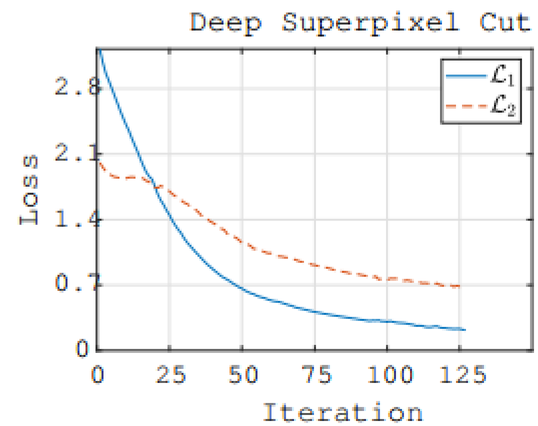


Fig 11. Visualization

- Qualitative Result

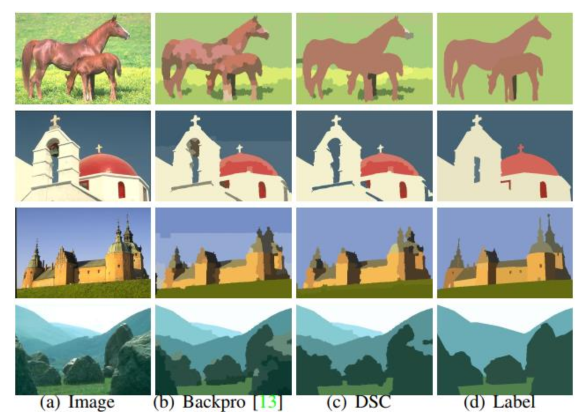


Fig 12. Segmentation Result

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 conductive 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