Page Segmentation for Historical Document Images Based on Superpixel Classification with Unsupervised Feature Learning

(2016)

Kai Chen, Cheng-Lin Liu, Mathias Seuret, Marcus Liwicki, Jean Hennebert, and Rolf Ingold

Resume

December 4, 2018

Abstract

an efficient page seg- mentation method for historical document images based on superpixels as basic units of segmentation. An image is first oversegmented into superpixels with SLIC algorithm. Then, each superpixel is represented by the features of its central pixel. The features are learned from pixel intensity values with stacked convolutional autoencoders in an unsupervised manner. A support vector machine (SVM) classifier is used to classify superpixels.

1 Introduction

In this work, rather than using pixels, superpixels, i.e., small regions obtained from an over segmentation are considered as the elementary units of the page segmentation task. Then features are learned directly from randomly selected image patches by using stacked convolutional autoencoders. With a support vector machine (SVM) trained with the features of the central pixels of the superpixels, an image is segmented into four regions: periphery, background, text block, and decoration. Finally, the segmentation results are refined by a connected components based smoothing procedure.

The advantages of the proposed method are: (1) The page segmentation is efficient compared to the previous methods. (2) fewer pixels for classifier training. (3) instead of using randomly selected pixels for classifier training, and given that these pixels may contain redundancy. The proposed method selects only the central pixels of the superpixels as training samples which are informative and representative.

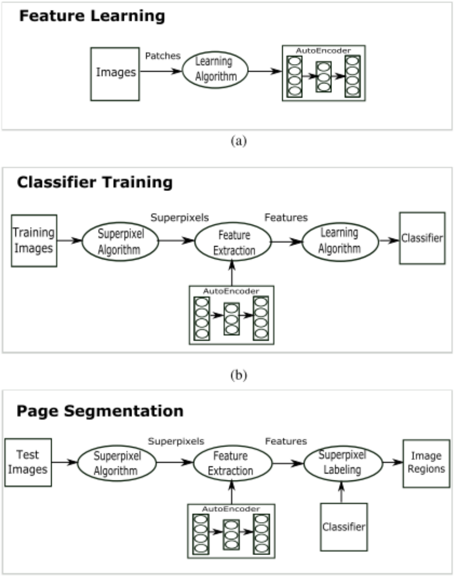
2 Method

The proposed method consists of four steps. The first step relies on superpixel algorithms, i.e., we segment an image into superpixels. In the second step, features are learned with the unsupervised learning method. In the third step, the learned features are used to train an SVM. With this SVM, the superpixels are classified into: *periphery, background, text block, and decoration.* Finally, we refine the segmentation results by eliminating the connected components (CCs) with predefined rules.

2.1 Superpixel

Superpixel algorithm aims at grouping pixels into perceptually meaningful patches that belong to the same object. The motivation of embedding the superpixel algorithm into our system is to reduce the computational complexity without degrading the quality of the segmentation. For the purpose of increasing the speed, instead of predicting each pixel's class, the method predicts each superpixel's class.

Side note: SLIC For color images in the CIELAB color space, the clustering procedure begins with an initialization step where k initial cluster centers $C_i = [l_i a_i b_i x_i y_i]^{\mathrm{T}}$ are sampled on a regular grid spaced S pixels apart. To produce roughly equally sized superpixels, the grid interval is $S = \sqrt{N/k}$. The centers are moved to seed locations corresponding to the lowest gradient position in a $3x3$ neighborhood. This is done to avoid centering a superpixel on an edge and to reduce the chance of seeding a superpixel with a noisy pixel.
Once each pixel has been associated to the nearest cluster center, an update step adjusts the cluster centers to be the mean $[labxy]^{\mathrm{T}}$ vector of all the pixels belonging to the cluster. The L_2 norm is used to compute a residual error E between the new cluster center locations and previous cluster center locations. The assignment and update steps can be repeated iteratively until the error converges.



(a)

(b)

Compared to other superpixel algorithms (Watershed - The mean shift - a graph-based approach), SLIC reaches a compromise between speed and segmentation quality. Furthermore, the generated superpixels'shapes are regular and approximately equally sized. Benefiting from this property, a graphical model can be applied to refine the segmentation results.

2.2 Feature learning

Using a single-layer fully-connected neural network as an autoencoder (AE) which learns to reconstruct its input data. Features can be discovered in the hidden layer.
In order to learn high-dimensional feature representations from unlabeled pixels, the feature learning system stacks three levels of AEs. We denote $x^{(k)}$ as the input vector and $W^{(k)}$ as the weights of the AE on the k-th level.

- Level 1: We randomly select 10 millions 55 pixels image patches $P^{(1)}$ from the training set. We set the number of hidden units of the AE to 40.
- Level 2: A 15x15 pixels image patch $P^{(2)}$ is composed vector of each $P^{(1)}$ by 3x3 patches $P^{(1)}$ without overlapping. Each 3x3 pixels from the inputs 15x15 are encoded using the learned weights $W_{(1)}$ are used obtaining 9 vectors $x_1^{(1)}....x_9^{(1)}$, the vectors are concatenated and then the new weights $W^{(2)}$ are learned, the AE has 30 hidden neurones.
- Level 3: The same things above is repeated using 45x45 pixels, he AE has 20 hidden neurones.

2.3 Classifier training and pixel labeling

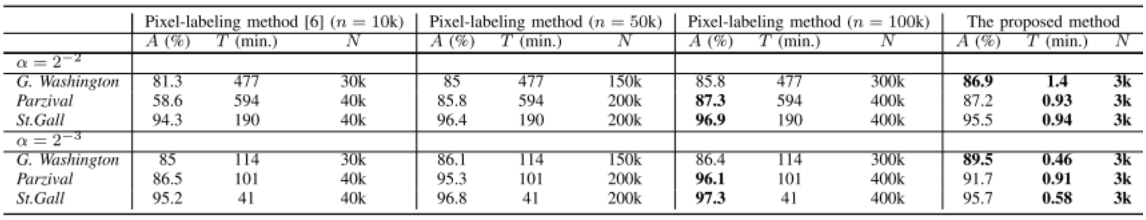
An SVM is trained with the labels and the learned features of the central pixels of the superpixels on the training images.
To segment an image, we first generate superpixels on that image. Then each superpixel is represented by the feature vector of its central pixel. With the trained SVM, the superpixels are classified into four classes, where the central pixels class is considered as the class of its corresponding superpixels.

3 Results

The original image sizes are: 2200x3400, 2000x3008, and 1664x2496 pixels for the datasets: G. Washington, Parzival, and Saint Gall respectively. Four classes of layout elements are defined in the Parzival and Saint Gall datasets, i.e., periphery, background, text block, and decoration. And three classes are defined in the G. Washington dataset, i.e., periphery, background, and text block.

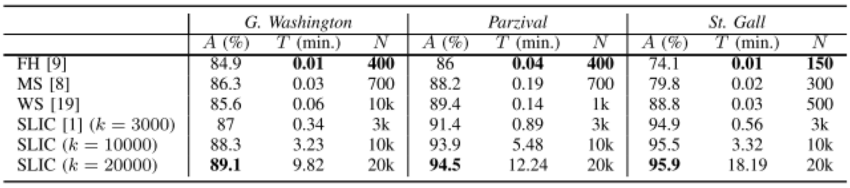
The number of the generated superpixels by SLIC is set to k = 3000 in order to achieve the compromise between the speed and quality of the segmentation results. Experiments are performed on the images of two resolutions with the scaling factors: $\alpha = 2^{-2}$ and $\alpha = 2^{-3}$.

Table I: Summary of segmentation results by using pixel-level and superpixel-level methods. Experiments are performed on two image scaling factors $\alpha \in \{2^{-2}, 2^{-3}\}$. The criteria used for evaluation are: classification accuracy A (%) [4], run time per image T (min.), and the number of training pixels per image N. For the pixel-level method, the number of training pixels per class is denoted as n. The number of superpixels generated by SLIC [1] is predefined as $k = 3000$.



	Pixel-labeling method [6] ($n = 10$k)			Pixel-labeling method ($n = 50$k)			Pixel-labeling method ($n = 100$k)			The proposed method		
	A (%)	T (min.)	N	A (%)	T (min.)	N	A (%)	T (min.)	N	A (%)	T (min.)	N
$\alpha = 2^{-2}$												
G. Washington	81.3	477	30k	85	477	150k	85.8	477	300k	**86.9**	**1.4**	**3k**
Parzival	58.6	594	40k	85.8	594	200k	**87.3**	594	400k	87.2	**0.93**	**3k**
St.Gall	94.3	190	40k	96.4	190	200k	**96.9**	190	400k	95.5	**0.94**	**3k**
$\alpha = 2^{-3}$												
G. Washington	85	114	30k	86.1	114	150k	86.4	114	300k	**89.5**	**0.46**	**3k**
Parzival	86.5	101	40k	95.3	101	200k	**96.1**	101	400k	91.7	**0.91**	**3k**
St.Gall	95.2	41	40k	96.8	41	200k	**97.3**	41	400k	95.7	**0.58**	**3k**

Table II: Comparison of four state-of-the-art superpixel algorithms on page segmentation for historical document images. The criteria used for evaluation are: classification accuracy A (%) [4], run time per image T (min.), and the number of training pixels per image N. Experiments are performed on the three datasets without using the post-processing procedure. Images are scaled down with the factor $\alpha = 2^{-3}$. The number of superpixels generated by SLIC [1] is predefined as k.



	G. Washington			*Parzival*			*St. Gall*		
	A (%)	T (min.)	N	A (%)	T (min.)	N	A (%)	T (min.)	N
FH [9]	84.9	**0.01**	**400**	86	**0.04**	**400**	74.1	**0.01**	**150**
MS [8]	86.3	0.03	700	88.2	0.19	700	79.8	0.02	300
WS [19]	85.6	0.06	10k	89.4	0.14	1k	88.8	0.03	500
SLIC [1] ($k = 3000$)	87	0.34	3k	91.4	0.89	3k	94.9	0.56	3k
SLIC ($k = 10000$)	88.3	3.23	10k	93.9	5.48	10k	95.5	3.32	10k
SLIC ($k = 20000$)	**89.1**	9.82	20k	**94.5**	12.24	20k	**95.9**	18.19	20k

References

[1] Kai Chen, Cheng-Lin Liu, Mathias Seuret, Marcus Liwicki, Jean Hennebert, and Rolf Ingold, *Page Segmentation for Historical Document Images Based on Superpixel Classification with Unsupervised Feature Learning*