This is the accepted version of a paper presented at First Swedish Symposium on Deep Learning.

Citation for the original published paper:

Semantic Labeling using Convolutional Networks coupled with Graph-Cuts for Document binarization.
In:

N.B. When citing this work, cite the original published paper.

Permanent link to this version:
http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-336174
Semantic Labeling using Convolutional Networks coupled with Graph-Cuts for Document binarization

Kalyan Ram Ayyalasomayajula  
Department of Information Technology  
Uppsala University  
kalyan.ram@it.uu.se

Anders Brun  
Department of Information Technology  
Uppsala University

Abstract—Most data mining applications on collections of historical documents require binarization of the digitized images as a pre-processing step. Historical documents are often subjected to degradations making mathematical modeling of appearance of the text, background and all kinds of degradations challenging. In the current work we try to tackle binarization as pixel classification problem. We first apply semantic segmentation, using fully convolutional neural networks. In order to improve the sharpness of the result, we further apply a graph cut algorithm. The labels from the semantic segmentation are used as approximate estimates of the text and background, with the probability map of background used for pruning the edges in the graph cut thus achieving state of the art results in binarization.

I. INTRODUCTION

In historical document image analysis, binarized images record each pixel in parchment/paper as background (BG) and pixel in text/ink as foreground (FG) by preserving most of the relevant visual information in the image. Historical documents have a lot of artifacts as shown in Fig.1 making the task of binarization very challenging. The method builds upon the ideas borrowed from Howe’s approach [1] of defining the binarization as labeling problem and further developing the approach to refine source-sink estimates to graph-cut based on our previous work [2].

II. METHOD

From Howe’s method [1] and the source-sink refinement approach [2], it can be inferred that results of graph-cut can benefit a lot from an approach that can produce better edge estimates. To meet this end we propose an approach to assist the graph-cut through source-sink and edge estimates from a Fully Convolutional Neural Network (FCNN). The idea is built from the fact that FCNNs have been very effective in producing state-of-the-art results in semantic segmentation [3]. The binarization results can be posed as a semantic segmentation problem with two class instances. The architecture of a FCNN is taken from [3]. It learns to combine coarse, high layer information with fine, low layer information.

• The labels obtained from FCNN output act as very good source and sink estimates.
• The network performs very well in estimating the background, but gives a very conservative estimate of the foreground labels. This is due to the class imbalance between FG and BG pixels, as the FCN is optimized for overall accuracy [3] on class label prediction.
• The probability map of the BG class thresholded on mean BG class probability is good indication of FG/BG separation.
• A graph is constructed by including edges in Fig.2(d) within this thresholded probability mask give a good estimate of the FG spread.
• These source, sink and edge map estimates serve as optimal inputs for a graph-cut algorithm to be applicable.

III. EXPERIMENTS

The experiments were conducted on the DIBCO [4] datasets for binarization consisting of 76 images. The dataset was divided into training and validation sets with 70-30 split. The ground truth and input images were converted into 500 × 300 pixels of cropped images with 100 pixels horizontally and vertically overlap. This creates more data for training through augmentation of cropped overlapping input images as mentioned previously. The cropped images of 500 × 300 pixels allow the trained model to fit into memory in Convolutional Architecture for Fast Feature Embedding framework [5].
Fig. 2. (a) Input Image (b) Represents FCN-8s as described in [3] (c) Output from FCNN indicated in thick arrows (d) Pruned edge map using the class probabilities from FCNN (e) Representation of source (ink) and sink (background) for graph-cut in dashed lines. Ink, background and edges indicated in red, blue and green pixels respectively (f) Representation of graph-cut (g) Final binarized output

(commonly known as CAFFE) on a on an NVIDIA Titan GTX. The model was then initialized with weights from a model pre-trained on PASCAL-VOC dataset available at [6]. Results for DIBCO data-set are provided for Howe’s method (HW) and FCNN with graph-cut (NNGC) for few DIBCO metrics in the Table I.

<table>
<thead>
<tr>
<th>File Name</th>
<th>FMeasure</th>
<th>PSNR</th>
<th>DRD</th>
</tr>
</thead>
<tbody>
<tr>
<td>2011_HW6</td>
<td>78.9</td>
<td>86.4</td>
<td>15.8</td>
</tr>
<tr>
<td>2011_PR2</td>
<td>74.9</td>
<td>80.1</td>
<td>11.4</td>
</tr>
<tr>
<td>2012_H13</td>
<td>63.0</td>
<td>87.4</td>
<td>15.5</td>
</tr>
<tr>
<td>2013_HW5</td>
<td>69.3</td>
<td>91.7</td>
<td>15.9</td>
</tr>
<tr>
<td>2013_HW7</td>
<td>51.4</td>
<td>74.3</td>
<td>18.1</td>
</tr>
<tr>
<td>2013_PR6</td>
<td>70.0</td>
<td>84.8</td>
<td>10.3</td>
</tr>
<tr>
<td>Mean DIBCO</td>
<td>81.7</td>
<td>89.8</td>
<td>17.8</td>
</tr>
</tbody>
</table>

IV. CONCLUSION

The current experimental framework shows relevance of FCNNs in segmentation based tasks such as binarization in documents to obtain state-of-the-art results. The possibilities as stated above are speculations into intended directions to pursue our future research into this field. The current approach can be extended to other tasks such as layout analysis and de-noising of historical documents. The method can benefit greatly from training on synthetic data generated by mimicking various degradations through filter operation thus enabling zero shot learning. Investigation into end-to-end training using a total variation frame-work would prove to be fruitful with possible extension into general segmentation and denoising framework for historical document collections.

ACKNOWLEDGMENT

Research sponsors "Vetenskapsrådet D.Nr 2012-5743) and Riksbankens Jubileumsfond (R.Nr NHS14-2068:1).

REFERENCES