International Journal on Document Analysis and Recognition (IJDAR)
https://doi.org/10.1007/s10032-018-0310-5

EDITORIAL

Special issue on deep learning for document analysis and recognition

Cheng-Lin Liu[1] · Gernot A. Fink[2] · Venu Govindaraju[3] · Lianwen Jin[4]

Deep learning—designing models and learning algorithms for deep neural networks—has achieved great success in various areas of artificial intelligence. Owing to the advances in theory and learning algorithms, the availability of big training data, and GPU computing, deep learning has yielded superior performance in many applications of pattern recognition and artificial intelligence. Examples include speech recognition, character and text recognition, image segmentation, object detection and recognition, traffic sign recognition, and face recognition. The exploration of new deep-learning models and algorithms as well as their potential applications has attracted great interest and attention.

Deep learning has significantly reshaped Document Analysis and Recognition (DAR) research, a field that analyzes digital contents of document images and handwriting. The adoption of deep learning has improved greatly the performance of character and text recognition (particularly, handwritten and scene-text recognition), text localization and document segmentation. Some of the most successful deep learning models include the convolutional neural network (CNN), the recurrent neural network with long short-term memory (RNN–LSTM), and the fully convolutional network (FCN). Now researchers are applying deep learning to additional document analysis problems, including layout analysis, writer identification and document retrieval.

This special issue acknowledges new advances in DAR that are using deep learning methods. The editors received 15 full submissions by the September 2017 deadline. The topics ranged from document image segmentation, layout analysis, text and object localization to character and text recognition, language modeling and handwritten mathematics recognition. They included signature verification, document retrieval and document understanding. The guest editors created a strict, peer-review process and invited guest reviewers to consider all submissions. At least two reviewers reviewed each paper, and most accepted papers underwent second-round review. Finally, the editors and reviewers accepted five papers for publication in this special issue. The outlined contents follow:

In "Learning to Detect, Localize and Recognize Many Text Objects in Document Images from Few Examples," Moysset et al. propose a new neural model, which directly predicts object coordinates for text detection in document images. Key components of the model are spatial 2D-LSTM recurrent layers, which convey contextual information between the regions of the image. They use a new form of local parameter sharing to keep the overall number of trainable parameters low. This makes the model more powerful than state-of-the-art applications where training data are not abundant. The model also facilitates the detection of many objects in a single image and can deal with inputs of variable sizes without resizing. The researchers propose two regression strategies that would limit the amount of information produced by the local model components and enhance the localization precision of the coordinate regressor. These strategies: (1) separately predict lower-left and upper-right corners of each object bounding box, followed by combinatorial pairing; (2) only predict the left side of the objects and estimate the right position jointly with text recognition. This has led to good full-page text recognition results in heterogeneous documents. The researchers have performed experiments on the text-line localization task in the Maurdor dataset.

In "Fully Convolutional Network for Handwritten Text Line Segmentation," Renton et al. present a learning-based

✉ Cheng-Lin Liu
liucl@nlpr.ia.ac.cn

Gernot A. Fink
Gernot.Fink@tu-dortmund.de

Venu Govindaraju
govind@buffalo.edu

Lianwen Jin
eelwjin@scut.edu.cn

[1] Institute of Automation of Chinese Academy of Sciences, Beijing, China

[2] TU Dortmund University, Dortmund, Germany

[3] University at Buffalo, Buffalo, USA

[4] South China University of Technology, Guangzhou, China

Published online: 20 August 2018

method for handwritten text-line segmentation in document images. They use a variant of deep fully convolutional networks (FCN) with dilated convolutions, which allow to never reduce the input resolution and produce a pixel-level labeling. They have trained the FCN to identify X-height labeling as text-line representation, which has many advantages for text recognition. In experiments on a public dataset, they show that the proposed method outperforms the most popular variants of FCN, based on deconvolution or unpooling layers. They provide results that investigate various settings, concluding with a comparison of recent approaches in the cBAD international competition.

In "Integrating Scattering Feature Maps with Convolutional Neural Networks for Malayalam Handwritten Character Recognition," Manjusha et al. use scattering-transform-based wavelet filters as first-layer convolutional filters in CNN architecture. A series of scattering-transform operations generates the scattering networks. The scattering coefficients generated in the first few layers are effective in capturing the dominant energy contained in the input data patterns. This architecture is equivalent to using scattering wavelet filters as first-layer receptive fields in CNN architecture. They used the proposed hybrid CNN architecture in Malayalam handwritten character recognition. The experimental results confirm that the proposed hybrid CNN architecture, based on scattering feature maps, could outperform CNN's equivalent self-learning architecture regarding problems of handwritten character recognition.

In "Attribute CNNs for Word Spotting in Handwritten Documents," Sudholt and Fink present an approach for word spotting in document images using learning attribute representations with convolutional neural networks (CNNs). By taking a probabilistic perspective on training CNNs, they derive two different loss functions for binary and real-valued word string embeddings. They also propose two different CNN architectures specifically designed for word spotting, which can be trained end-to-end. In a number of experiments, they investigate the influence of different word string embeddings and optimization strategies. Their work shows that the proposed 'Attribute CNNs' achieve state-of-the-art results for segmentation-based word spotting on a large variety of datasets.

In "Fixed-Sized Representation Learning from Offline Handwritten Signatures of Different Sizes," Hafemann et al. propose modifying the network architecture of deep convolutional neural network using spatial pyramid pooling to learn about fixed-sized representation from variable-sized signatures. They also are investigating the impact of image resolution used for training, and the impact of adapting (fine-tuning) the representations to new operating conditions (different acquisition protocols, such as writing instruments and scan resolution). On the GPDS dataset, their results compare to state-of-the-art work, while not needing a maximum size to process the signatures. They also show that using higher resolutions (300 or 600 dpi) can improve performance when skilled forgeries from a subset of users are available for feature learning, but lower resolutions (around 100 dpi) can be used only with genuine signatures. When the operating conditions change, the researchers show that fine-tuning can improve performance.

We thank all the authors who shared their invaluable work and all reviewers for their insightful comments in considering the submissions for this special issue. We also want to thank the Editors-in-Chief of "International Journal on Document Analysis and Recognition" for giving us an opportunity to guest-edit this special edition and address the exciting advances in this field. Finally, we want to recognize the hard work of journal assistant Saranya Karunakaran for her ongoing support and assistance in this tremendous effort.