Pattern Recognition 61 (2017) 537–538

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Editorial: Special Section on deep image and video understanding

Deep learning has gained attention in the past years due to its ability in learning features in an unsupervised fashion. Since such techniques are based on the assumption that data is processed hierarchically in the brain, they can be used to encode different information at each step of the learning process, thus obtaining coarser-to-finer representation from images, signals and video.

Convolutional Neural Networks (CNNs), Deep Belief Networks (DBNs) and Auto-encoders are among the most used ones to learn features from any kind of data, to be further employed to feed supervised pattern recognition techniques. As a matter of fact, we have a number of discriminative versions of the aforementioned techniques, which are capable of both learning features and perform classification. A CNN, for instance, is composed of two main modules: a feature learner, and an Artificial Neural Network (ANN). The former is used to extract features from the input data, and the latter gets a high-dimensional feature vector to perform traditional classification. Restricted Boltzmann Machines (RBMs) are often used in a generative fashion, but discriminative versions are available either.

The main idea of this special issue is to compile recent and high quality works related to deep learning, as well as to foster the community regarding such research area. We focused mainly on image- and video-based data, but other works were welcomed as well. Additionally, the special issue aimed at gathering a few selected papers presented at the SIBGRAPI 2015 – Conference on Graphics, Patterns and Images. We received a number of works from people worldwide, such as USA, Brazil, Australia, China, India and Malaysia, among others. Papers that range from image segmentation, pedestrian detection, remote sensing image classification, biometrics and medical-based applications were submitted and reviewed by three reviewers, at least.

The first paper, by Lin Ma, Zhuo Chen, Long Xu and Yihua Yan, used a multimodal deep learning approach for solar radio burst classification. The main idea is to learn a joint representation of the solar data acquired from different spectra, which are treated as a different modality. The authors used an autoencoder together with a structured regularization to force the learning over the modality-specific architectures. The second paper by Anush Sankaran, Gaurav Goswami, Mayank Vatsa, Richa Singh, and Angshul Majumdar, proposed a semi-supervised RBM, which combines an unsupervised learning with a supervised sparsity regularizer. The work evaluates the proposed approach in both RBMs and DBNs, obtaining results comparable to the ones achieved in some well-known datasets.

The work by Chen et al. used CNNs to classify lung nodule malignancy suspiciousness using Computed Tomography images. The authors proposed the Multi-crop Convolutional Neural Network (MC-CNN), which extracts nodule salient information by means of a multi-crop pooling strategy instead of the conventional max-pooling approach. A softmax layer with two output units (high and low suspiciousness) was used at the very end of the pipeline. The proposed MC-CNN achieved state-of-the-arts results with respect to lung nodule malignancy suspiciousness. Zhao et al. presented a Discriminant Deep Belief Network to classify SAR images. Given a set of image patches, weak classifiers are used for labeling purposes, which are then used to feed a DBN model for feature learning. The proposed approach is compared against a number of techniques, and with more accurate results.

David Ribeiro, Jacinto C. Nascimento, Alexandre Bernardino and Gustavo Carneiro used CNNs to detect pedestrians in natural images. The work showed one can improve non-deep state-of-the-art pedestrian detections by means of CNNs, as well as better results can be achieved by means of pre-trained models. The proposed approach cascaded state-of-the-art non-deep detectors with a deep compositional architecture, showing very promising results. Ohn-Bar and Trivedi employed Convolutional Neural Networks for object detection and localization. The authors considered capturing the contextual information among scales by learning to pool features across multiples scales at the same layer, thus resulting in significant gains in both detection performance and localization quality of objects on the PASCAL VOC dataset and a Multi-View Highway Vehicles dataset.

van Noord and Postma proposed to evaluate scale-variant and scale-invariant features for deep image classification by means of a multi-scale CNN. The authors showed it can outperform single-scale CNNs, and also they highlighted the importance in building datasets with scale-variant and scale-invariant representations. The work by Lore, Akintayo and Sarkar presented a deep autoencoder-based approach to adaptively brighten images without saturating lighter regions. The authors used a sparse denoising autoencoder that can learn from darkened and noise-added training images how to produce brighter versions with less degradation.

Nogueira, Penatti and dos Santos analyzed CNNs in different scenarios concerning the task of remote sensing imagery classification. Their work was presented at Sibgrapi'2015, and it was invited to this special issue. The authors evaluated CNNs in distinct scenarios in an extensive comparison against state-of-the-art image descriptors, achieving promising results for all datasets employed in the work. Ferrari, Lombardi and Signoroni employed CNNs for bacterial colony counting, being one of the few application-oriented papers. The authors showed that deep learning techniques could outperform handcrafted features with respect to the aforementioned context.

http://dx.doi.org/10.1016/j.patcog.2016.09.001

Franco and Oliveira proposed a deep learning-based approach to deal with the problem of person re-identification in images extracted from videos. The authors proposed the Convolutional Covariance Features (CCF), which basically describe the feature maps learned from CNNs. The proposed work achieved the best top 1 result when compared against 18 state-of-the-art methods over three well-known datasets. The last paper by Lopes et al. employed CNNs for facial expression recognition. Such work was presented at Sibgrapi' 2015 as well, and it proposed a simple solution based on a combination of CNNs and some image-specific pre-processing steps. Competitive results were achieved over CK+, JAFFE and BU-3DFE datasets.

On the whole, this special issue covered a number of works that highlighted the importance in using deep learning representations for image- and video-based data. The problem of obtaining large collection of data is a concern in all works considered in this issue, and each paper presented its own perspective to deal with such shortcoming. Although there is a consensus that deep learning representations often outperform handcrafted ones, there is still room for improvements and novelties. We hope the readers of this special issue will benefit from recent and interesting advances related to CNNs and DBNs, as well as how they can be applied to a variety of different applications.

João Paulo Papa[1,2]: Currently, Prof. Papa is with Department of Computing at São Paulo State University, Brazil. His main interests are related to model selection in machine learning techniques using meta-heuristics, pattern recognition and image processing. Prof. Papa has obtained the best paper award at the 2014 International Conference on Pattern Recognition (ICPR) due to his paper "Efficient supervised optimum-path forest classification for large datasets" published at *Pattern Recognition* during 2012. Nowadays, Prof. Papa is co-editing a book related to image processing applications and meta-heuristic optimization to be published by Springer in the next semester, and it has authored more than 150 peer reviewed paper in conferences and journals.

Ryan Farrell[3,4]: Currently, Prof. Farrell is an Assistant Professor at Brigham Young University, and his main interests include computer vision and object recognition. He has been the Co-chair for the First (CVPR 2011) and Second (CVPR 2013) Workshops on Fine-Grained Visual Categorization (FGVC), and a Program Committee Member for CVPR 2010/2011/2012/2013/2014, ECCV 2010/2012 and ICCV 2011/2013. Additionally, Prof. Farrell has been the website Chair/Co-Chair for CVPR 2010, CVPR 2012, CVPR 2013, ICCV 2013 and CVPR 2014.

Guest Editor
João Paulo Papa
Department of Computing, São Paulo State University, Bauru 17033-360, Brazil
E-mail address: papa@fc.unesp.br

Guest Editor
Ryan Farrell
Computer Science Department, Brigham Young University, Provo 84602, USA
E-mail address: farrell@cs.byu.edu

[1] https://scholar.google.com/citations?user=8hU3dn8AAAAJ&hl=en
[2] http://wwwp.fc.unesp.br/~papa
[3] https://scholar.google.com/citations?user=56EZh6YAAAAJ&hl=en
[4] https://faculty.cs.byu.edu/~farrell