Deep Learning in Image Computing: An Overview

Jomy John
Assistant Professor, Department of Computer Science
KKTM Government College, Pullut, Kodungallur, Kerala (Affiliated to the University of Calicut)
jomyeldos@gmail.com

Abstract— Deep learning is a growing trend in computing. It is an improvement to artificial neural network. Deep Neural Networks are used in image classification, detection and segmentation. In this paper, an overview is carried out about the usage of deep neural network in various areas of image computing including image quality assessment, document imaging, object recognition, medical imaging, content based image retrieval and microscopy images with representative notable works in these areas. These works reveal that promising results are emerging with the help of deep learning architecture.

Keywords— Deep Learning, CNN, LSTM, RNN, CBIR, Image quality assessment, Document Images, Object Recognition, Medical Images, Microscopy Images.

1. INTRODUCTION

Deep learning is emerging as the leading machine-learning tool in the general imaging and computer vision domains. Deep learning has a long history, and its basic concept is originated from artificial neural network research. The feed forward neural networks with many hidden layers are indeed a good example of the models with a deep architecture [1]. Deep learning consists of more layers that permit higher levels of abstraction and improved predictions from data. Convolutional Neural Networks (CNNs) have proven to be powerful tools for a broad range of computer vision tasks. Deep CNNs automatically learn mid-level and high-level abstractions obtained from images. In the case of document image analysis, a deep learning method called, Long Short Term Memory (LSTM), a recurrent neural network are proven to be more effective.

Deep learning architecture is used extensively in image computing domain. This paper provides an overview about the works in image quality assessment, document imaging, object recognition, medical imaging, content based image retrieval and microscopy images with representative notable works in these areas.

This paper is divided into four sections. The next section covers description about deep learning. Section III covers the existing works in different image computing domain with subsections on including image quality assessment, document imaging, object recognition, medical imaging, content based image retrieval and microscopy images and Section IV concludes the paper.

2. DEEP LEARNING

Deep Learning is a new way to train multiple neural networks. A standard neural network consists of many simple connected processors called neurons, each producing a sequence of real valued activations. Input neurons get activated through sensors perceiving the environment; other neurons get activated through weighted connections from previous active neurons. Some neurons may influence the environment by triggering actions. Learning is about finding weights that make the neural network exhibit desired behaviour. Depending on the problem and how the neurons are connected, such behaviour may require long casual chains of computational stages where each stage transforms the aggregate activation of networks [2]. It has multiple layers of nonlinear processing units and supervised or unsupervised learning of feature representation in each layer forming a hierarchy from low level to high level features [3]. Deep learning is about accurately assigning credit across many such stages. Deep learning algorithm trans- form their inputs through more layers than shallow learning algorithms. At each layer, the signal is transformed by a processing unit where parameters are learned through training [2]. A chain of transformations from input to output is a credit assignment path which describe potentially causal connections between input and output. Each of the non-output layers are trained to be an auto-encoder and it is forced to learn good features that describe what comes from the previous layer. An auto encoder is trained with an absolutely standard weight adjustment algorithm to reproduce the input.

Deep learning architectures may include deep generative models such as the nodes in deep neural networks, convolu- tional deep neural networks, deep belief networks, deep Boltz- mann machines and recurrent neural networks. Convolutional Neural Networks (CNNs) have proven to be powerful tools for a broad range of computer vision tasks. A deep belief network (DBN) is a probabilistic, generative model made up of multiple layers of hidden units. It can be considered a composition of simple learning modules that make up each layer [4]. Convolutional deep belief networks have structure very similar to a convolutional neural network and are trained similar to deep belief networks. Therefore, they exploit the

2D structure of images and make use of pre-training like deep belief networks. A deep Boltzmann machine (DBM) is a type of binary pairwise Markov random field (undirected probabilistic graphical model) with multiple layers of hidden random variables. It is a network of symmetrically coupled stochastic binary units. A deep learning method called, Long Short Term Memory (LSTM), a recurrent neural network are proven to be more effective handwriting recognition.

Deep neural networks also have become relevant for more general field of reinforcement learning where there is no supervising teacher.

Recurrent Themes of Deep Learning include [2]:

1. Dynamic Programming for Supervised Reinforcement Learning (SL/RL)
2. Unsupervised Learning Facilitating Supervised Learning and Reinforcement Learning.
3. Learning Hierarchical Representation through Deep Super- vised Learning, Unsupervised Learning and Reinforcement Learning
4. Occams Razor: Compression and Minimum Description Length
5. Fast Graphics Processing Units for Deep Learning in Neural Networks

3. EXPERIM ENTS ON VARIOUS IMAGE COMPUTING TASKS

3.1. Image Quality Assessment

Good quality images are a prerequisite for many image computing tasks. In this paper, the authors [5] proposed a deep learning approach for document image quality assessment. Given a noise corrupted document image, its quality score was estimated as a prediction of character recognition accuracy. First the document image was divided into patches and non- informative patches were sifted out using Otsus´ binarization technique. Second, quality scores were obtained for all selected patches using a Convolutional Neural Network, and the patch scores were averaged over the image to obtain the document score. The proposed CNN contained two layers of convolution, location blind max-min pooling, and rectified linear units in the fully connected layers. They showed that their method achieved state of the art performance through experiments.

Another work on deep learning for distortion-generic blind image quality assessment was reported by Bianco et al [6]. They have used features extracted from pre-trained Convolutional Neural Networks as a generic image description. Their best proposal, named DeepBIQ, estimated the image quality by average pooling the scores predicted on multiple sub-regions of the original image. The score of each sub-region was computed using a Support Vector Regression (SVR) machine taking as input features extracted using a CNN fine-tuned for category-based image quality assessment. Experimental results showed that DeepBIQ outperforms the state-of-the-art methods and in most of the cases the quality score predictions of DeepBIQ were closer to the average observer than those of a generic human observer.

Deep Max-Pooling convolutional Neural Networks are Deep Neural Networks (DNN) with convolutional and max-pooling layers. When used to scan images by means of a sliding window, however, their high computational complexity can slow down the most powerful hardware. Giusti et al. [7] showed how dynamic programming can speedup the process by orders of magnitude, even when max-pooling layers are present.

3.2. Document Imaging

Long Short Term Memory(LSTM) is a recurrent neural network (RNN) architecture. In book chapter [8], the authors have given a short overview of the underlying algorithms of LSTM and describes numerous recent successful applica- tions in the area of document image analysis. In 2009, deep multidimensional LSTM networks won three ICDAR 2009 competitions in connected handwriting recognition, without any prior knowledge about the three languages to be learned [9].

In object and scene analysis, deep neural nets were capable of learning a hierarchical chain of abstraction from pixel inputs to concise and descriptive representations. In ref [10], the authors presented document image classification and retrieval, using features learned by deep convolutional neural networks. This work explored this capacity in the realm of document analysis, and confirmed that this representation strategy was superior to a variety of popular hand-crafted alternatives. Using experiments it was showed that (i) features extracted from CNNs were robust to compression, (ii) CNNs trained on non-document images transferred well to document analysis tasks, and (iii) enforcing region-specific feature-learning was unnecessary given sufficient training data.

3.3. Object Recognition

Object detection is one of the fundamental tasks in computer vision. A common paradigm to address this problem is to train object detectors which operate on a sub-image and apply these detectors in an exhaustive manner across all locations and scales. Deep convolutional neural networks have recently achieved state-of-the-art performance on a number of image recognition benchmarks, including the ImageNet Large-Scale Visual Recognition Challenge (ILSVRC-2012). The winning model on the localization sub-task was a network that could predict a single bounding box and a confidence score for each object category in the image. Such a model captures the whole- image context around the objects but cannot handle multiple instances of the same object in the image without naively replicating the number of outputs for each instance. Erhan et al.[11] proposed a saliency-inspired neural network model for detection, which could predict a set of class-agnostic bounding boxes along with a single score for each box, corresponding to its likelihood of containing any object of interest. The model naturally handled a variable number of instances for each class and allowed for crossclass generalization at the highest levels of the network. They were able to obtain competitive recognition performance on VOC2007 and ILSVRC2012, while using only the top few predicted locations in each image and a small number of neural network evaluations.

The success of CNNs is attributed to their ability to learn rich midlevel image representations as opposed to hand- designed low-level features used in other image classification methods. Learning CNNs, however, amounts to estimating millions of parameters and requires a very large number of annotated image samples. This property currently prevents application of CNNs to problems with limited training data. In this work [12], the authors showed how image representations learned with CNNs on large-scale annotated datasets can be efficiently transferred to other visual recognition tasks with limited amount of training data. They designed a method to reuse layers trained on the ImageNet dataset to compute mid-level image representation for images in the PASCAL VOC dataset. They showed that the transferred representation lead to significantly improved results for object and action classification, outperforming the current state of the art on Pascal VOC 2007 and 2012 datasets.

Chan et al.[13], proposed a very simple deep learning network for image classification which comprised only the very basic data processing components: Cascaded Principal Component Analysis (PCA), binary hashing, and block-wise histograms. In the proposed architecture, PCA was employed to learn multistage filter banks. It was followed by simple binary hashing and block histograms for indexing and pooling. This architecture was thus named as a PCA network (PCANet) and was designed and learned extremely easily and efficiently. For comparison and better understanding, they also introduced two simple variations to the PCANet, namely the RandNet and LDANet. They share the same topology of PCANet but their cascaded filters were either selected randomly or learned from LDA. They tested these basic networks extensively on many benchmark visual datasets for different tasks, such as LFW for face verification, MultiPIE, Extended Yale B, AR, FERET datasets for face recognition, as well as MNIST for hand-written digits recognition. For all tasks, PCANet model as on par with the state of the art feature.

3.4. Medical Imaging

Deep learning methods are most effective when applied to large training sets, but in the medical domain large data sets are not always available. These areas are therefore faced with major challenges including (a) Can deep networks be used effectively for medical tasks? (b) Is transfer learning from general imagery to the medical domain relevant? (c) Can rely on learned features alone or may combine them with handcrafted features for the task? These issues are addressed in the special issue of IEEE-Transactions on Medical Imaging (IEEE-TMI) on deep learning in medical imaging [1]. Therefore, in this paper, the review is conducted excluding papers from the special issue.

Mammographic scoring of density and texture are established methods to relate the risk of breast cancer. To address the tasks of breast tissue segmentation, scoring of percentage mammographic density and scoring of mammographic texture, Petersen et al. [14] presented a method that learns descriptive features from unlabelled mammograms as an input to deep learning architecture.

In ref [15] authors provided the idea of using deep neural network architecture with dynamically programmed layers for brain connectome prediction problem. Understanding the brain connectome structure is important in the research for epilepsy and other neuropathological diseases. They introduced a new deep learning architecture that used the spatial and temporal nature of the neuronal activation data. That architecture consisted of a combination of Convolutional layer and a Recurrent layer for predicting the connectome of neurons based on their time-series of activation data with a dynamically programmed layer for determining the alignment between the neuronal activations of pair-wise combinations of neurons.

Combining multi-modality brain data for disease diagnosis commonly leads to improved performance. A challenge in using multi-modality data is that some modality might be missing for some subjects. In the work cited [16], the authors proposed a deep learning based framework for estimating multi-modality imaging data. Their method took the form of convolutional neural networks, where the input and output are two volumetric modalities. The network contained a large number of trainable parameters that captured the relationship between input and output modalities. When trained on subjects with all modalities, the network could estimate the output modality given the input modality. They valuated their method on the Alzheimers Disease Neuroimaging Initiative (ADNI) database, where the input and output modalities were MRI and PET images, respectively. Results showed that their method significantly outperformed prior methods.

Segmentation of anatomical structures in medical images is often based on a voxel/pixel classification approach. Deep learning systems, such as convolutional neural networks (CNNs), can infer a hierarchical representation of images that fosters categorization. In work reported [17], the authors proposed a system for voxel classification integrating three

2D CNNs, which have a one-to-one association with the xy, yz and zx planes of 3D image, respectively. They applied their method to the segmentation of tibial cartilage in low field knee MRI scans and tested it on 114 unseen scans. Even though they used only 2D features at a single scale, it was reported to perform better than a state-of-the-art method using

3D multi-scale features by a deep learning architecture that autonomously learns the features from the images.

For the last decade, it has been shown that neuroimaging can be a potential tool for the diagnosis of Alzheimers Disease (AD) and its prodromal stage, Mild Cognitive Impairment (MCI), and also fusion of different modalities can further provide the complementary information to enhance diagnostic accuracy. The problems of both feature representation and fusion of multimodal information from Magnetic Resonance Imaging (MRI) and Positron Emission Tomography (PET) were addressed in [18]. The authors proposed a method for a high-level latent and shared feature representation from neuroimaging modalities via deep learning. Specifically, they used Deep Boltzmann Machine (DBM), a deep network with a restricted Boltzmann machine as a building block, to find a latent hierarchical feature representation from a 3D patch, and then devised a systematic method for a joint feature representation from the paired patches of MRI and PET with a multimodal DBM.

Lung cancer has a poor prognosis when not diagnosed early and unresectable lesions are present. The management of small lung nodules noted on computed tomography scan is controversial due to uncertain tumor characteristics. A conventional computer-aided diagnosis (CAD) scheme requires several image processing and pattern recognition steps to accomplish a quantitative tumor differentiation result. Deep learning techniques have the intrinsic advantage of an automatic exploitation feature and tuning of performance in a seamless fashion. In ref [19], authors introduced models of a deep belief in the context of nodule classification in computed tomography images. The experimental results suggested that deep learning methods could achieve better discriminative results. It is reported to outperform compared to feature computing methods.

3.5. Content Based Image Retrieval

Content-based image retrieval (CBIR) is one of the fundamental research challenges extensively studied in multimedia community for decades. CBIR aims to search for images through analysing their visual contents. The key challenge has been attributed to the well-known semantic gap issue that exists between low-level image pixels captured by machines and high-level semantic concepts perceived by human. Inspired by recent successes of deep learning techniques for computer vision and other applications, in this paper [20], the authors attempted to address an open problem: if deep learning is a hope for bridging the semantic gap in CBIR and how much improvements in CBIR tasks can be achieved by exploring the state-of-the-art deep learning techniques for learning feature representations and similarity measures. They investigated a framework of deep learning with application to CBIR tasks with a set of empirical studies and came out with encouraging results.

3.6. Microscopy Images in Drug Diagnosis

High-content screening (HCS) technologies have enabled large scale imaging experiments for studying cell biology and for drug screening [21]. These systems produce hundreds of thousands of microscopy images per day and their utility depends on automated image analysis. Recently, deep learning approaches that learn feature representations directly from pixel intensity values have dominated object recognition challenges. These tasks typically have a single centered object per image and existing models are not directly applicable to microscopy datasets. The authors developed an approach that combines deep convolutional neural networks (CNNs) with multiple instance learning (MIL) in order to classify and segment microscopy images.

Immunohistochemistry (IHC) staining is a widely used technique in the diagnosis of abnormal cells such as cancer. For instance, it can be used to determine the distribution and localization of the differentially expressed biomarkers of immune cells (such as T-cells or B-cells) in cancerous tissue for an immune response study. Typically, the immunological data of interest includes the type, density and location of the immune cells within the tumor samples. Manually counting each subset of immune cells under a bright field microscope for each piece of IHC stained tissue is extremely tedious and time consuming. This makes automatic detection very attractive, but it can be very challenging due to the wide variety of cell appearances resulting from different tissue types, block cuttings, and staining processes. In ref [22], a method for automatic immune cell counting on digitally scanned images of IHC stained slides was presented. The method first uses a sparse color unmixing technique to separate the IHC image into multiple color channels that correspond to different cell structures. Since the immune cell biomarkers were membrane markers, the detection problem was formulated into a deep learning framework using the membrane image channel. The algorithm was evaluated on a clinical data set containing a large number of IHC slides and demonstrated more effective detection than the existing technique and the result was also in accordance with the human observers output.

4. CONCLUSION

In this paper, an overview about the works in image computing using deep learning is carried out. Representative works from each image computing domain is reported. In the majority of the works presented, use of deep learning is proven to improve performance over the state-of-the-art methods. Even though deep learning can be done using supervised and unsupervised learning, most of works are focused on supervised learning. Deep learning networks can improve significantly with large training data.

REFERENCES

[1] H. Greenspan, B. van Ginneken, and R. M. Summers, “Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique,” IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1153–1159, 2016.
[2] J. Schmidhuber, “Deep learning in neural networks: An overview,” Neural Networks, vol. 61, pp. 85–117, 2015.
[3] D. Yu, L. Deng, and D. Yu, “Deep learning methods and applications,” Foundations and Trends in Signal Processing, 2014.
[4] G. Hinton, “Deep belief networks,” Scholarpedia, 4(5): 5947.
[5] L. Kang, P. Ye, Y. Li, and D. Doermann, “A deep learning approach to document image quality assessment,” in 2014 IEEE International Conference on Image Processing (ICIP), pp. 2570–2574, IEEE, 2014.

[6] S. Bianco, L. Celona, P. Napoletano, and R. Schettini, "On the use of deep learning for blind image quality assessment," arXiv preprint arXiv:1602.05531, 2016.

[7] A. Giusti, D. C. Cires¸an, J. Masci, L. M. Gambardella, and J. Schmidhu- ber, "Fast image scanning with deep max-pooling convolutional neural networks," arXiv preprint arXiv:1302.1700, 2013.

[8] M. Liwicki12, V. Frinken, and M. Z. Afzal, "Latest developments of lstm neural networks with applications of document image analysis," Handbook of Pattern Recognition and Computer Vision, p. 293, 2015.

[9] A. Graves and J. Schmidhuber, "Offline handwriting recognition with multidimensional recurrent neural networks," Advances in Neural Infor- mation Processing Systems 22 (NIPS'22), 2009.

[10] A. W. Harley, A. Ufkes, and K. G. Derpanis, "Evaluation of deep convolutional nets for document image classification and retrieval," in Document Analysis and Recognition (ICDAR), 2015 13th International Conference on, pp. 991–995, IEEE, 2015.

[11] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2147–2154, 2014.

[12] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1717–1724, 2014.

[13] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. Ma, "Pcanet: A simple deep learning baseline for image classification?," IEEE Transactions on Image Processing, vol. 24, no. 12, pp. 5017–5032, 2015.

[14] K. Petersen, M. Nielsen, P. Diao, N. Karssemeijer, and M. Lillholm, "Breast tissue segmentation and mammographic risk scoring using deep learning," in International Workshop on Digital Mammography, pp. 88–94, Springer, 2014.

[15] V. Veeriah, R. Durvasula, and G.-J. Qi, "Deep learning architecture with dynamically programmed layers for brain connectome prediction," in Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 1205–1214, ACM, 2015.

[16] R. Li, W. Zhang, H.-I. Suk, L. Wang, J. Li, D. Shen, and S. Ji, "Deep learning based imaging data completion for improved brain disease diagnosis," in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 305–312, Springer, 2014.

[17] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 246–253, Springer, 2013.

[18] H.-I. Suk and D. Shen, "Deep learning-based feature representation for ad/mci classification," in International Conference on Medical Image Computing and Computer-Assisted Intervention, pp. 583–590, Springer, 2013.

[19] K.-L. Hua, C.-H. Hsu, S. C. Hidayati, W.-H. Cheng, and Y.-J. Chen, "Computer-aided classification of lung nodules on computed tomog- raphy images via deep learning technique," OncoTargets and therapy, vol. 8, 2015.

[20] J. Wan, D. Wang, S. C. H. Hoi, P. Wu, J. Zhu, Y. Zhang, and J. Li, "Deep learning for content-based image retrieval: A comprehensive study," in Proceedings of the 22nd ACM international conference on Multimedia, pp. 157–166, ACM, 2014.

[21] O. Z. Kraus, J. L. Ba, and B. J. Frey, "Classifying and segmenting mi- croscopy images with deep multiple instance learning," Bioinformatics, vol. 32, no. 12, pp. i52–i59, 2016.

[22] T. Chen and C. Chefdhotel, "Deep learning based automatic immune cell detection for immunohistochemistry images," in International Workshop on Machine Learning in Medical Imaging, pp. 17–24, Springer, 2014.