International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

IJIEMR Transactions, online available on 23rd Dec 2019. Link :http://www.ijiemr.org/downloads.php?vol=Volume-08&issue=ISSUE-12

Title: **CONCEPTUAL UNDERSTANDING OF CONVOLUTIONAL NEURAL NETWORK- A DEEP LEARNING APPROACH**

Volume 08, Issue 12, Pages: 313–322.

Paper Authors

KALLA KIRAN, N RAMA KRISHNA

RAMACHANDRA COLLEGE OF ENGINEERING, ELURU

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

CONCEPTUAL UNDERSTANDING OF CONVOLUTIONAL NEURAL NETWORK- A DEEP LEARNING APPROACH

[1]KALLA KIRAN, [2]N RAMA KRISHNA
[1]PROFESSOR, DEPARTMENT OF CSE,RAMACHANDRA COLLEGE OF ENGINEERING, ELURU
[2]ASST. PROFESSOR, DEPARTMENT OF CSE, RAMACHANDRA COLLEGE OF ENGINEERING, ELURU
[1]kallakiran1974@gmail.com [2]nandigum.mrk@gmail.com

Abstract

Deep learning has become an area of interest to the researchers in the past few years. Convolutional Neural Network (CNN) is a deep learning approach that is widely used for solving complex problems. It overcomes the limitations of traditional machine learning approaches. The motivation of this study is to provide the knowledge and understanding about various aspects of CNN. This study provides the conceptual understanding of CNN along with its three most common architectures, and learning algorithms. This study will help researchers to have a broad comprehension of CNN and motivate them to venture in this field. This study will be a resource and quick reference for those who are interested in this field.

INTRODUCTION

With the rapidly growing demand for learnable machines for solving many complex problems, deep learning has evolved itself as an area of interest to the researchers in the past few years. As researchers tend to mimic human behavior, a major question arises that how do the humans acquire knowledge? The answer to this question is an essential ability of humans i.e. learning, which needs to be incorporated in machines, hence the term machine learning was coined. Machine learning promises to reduce the efforts by making the machines to learn themselves through past experiences [2] using three approaches of learning namely, learning under supervision, without supervision and semi-supervised learning [4]. The conventional machine learning techniques need feature extraction as the prerequisite, and this requires a domain expert [16]. Furthermore, selection of appropriate features for a given problem is a challenging task. Deep learning techniques overcome the problem of feature selection by not requiring pre-selected features but extracting the significant features from raw input automatically for a problem in hand. Deep learning model consists of a collection of processing layers that can learn various features of data through multiple levels of abstraction [15]. Multiple levels allow the network to learn distinct features. Deep learning has emerged as an approach for achieving promising results in various applications like image recognition, speech

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

recognition [9], topic classification, sentiment analysis, language translation, natural language understanding, signal processing, face recognition, prediction of bioactivity of small molecules etc. There are different deep learning architectures such as deep belief networks, recurrent neural networks, convolution neural networks etc.

Convolution Neural Network (CNN), often called ConvNet, has deep feed-forward architecture and has astonishing ability to generalize in a better way as compared to networks with fully connected layers. [8] Describes CNN as the concept of hierarchical feature detectors in biologically inspired manner. It can learn highly abstract features and can identify objects efficiently. The considerable reasons why CNN is considered above other classical models are as follows. First, the key interest for applying CNN lies in the idea of using concept of weight sharing, due to which the number of parameters that needs training is substantially reduced, resulting in improved generalization [1]. Due to lesser parameters, CNN can be trained smoothly and does not suffer overfitting. Secondly, the classification stage is incorporated with feature extraction stage, both uses learning process. Thirdly, it is much difficult to implement large networks using general models of artificial neural network (ANN) than implementing in CNN. CNNs are widely being used in various domains due to their remarkable performance such as image classification; [5], object detection, face detection, speech recognition, vehicle recognition, diabetic retinopathy, facial expression recognition and many more. The motivation of this study is to establish a theoretical framework while adding to the knowledge and understanding about CNN. The purpose of this study is to present the amalgamation of the elementary principles of CNN and providing the details about the general model, three most common architectures and learning algorithms. A new learning technique, ADAM proposed by [11] has also been elucidated. In addition to that, it computes learning rate for every individual parameter. The complete structure of the sections is as follows. Section 1 gives the introduction and provides the purpose of the study. Describes the general model of CNN along with all its elementary concepts. Introduces various architectures of CNN. Portrays the learning algorithm and illustrates the conclusion and future scope.

Deep learning (deep structured learning, hierarchical learning or deep machine learning) is a branch of machine learning based on a set of algorithms that attempt to model high-level abstractions in data by using multiple processing layers with complex structures, or otherwise composed of multiple non-linear transformations. Deep learning is part of a broader family of machine learning methods based on learning representations of data. An observation (e.g., an image) can be represented in many ways such as a vector of intensity values per pixel, or in a more abstract way as a set of edges, regions of particular shape, etc. Some representations make it easier to learn tasks (e.g., face recognition or facial expression recognition) from examples. One of the potentials of deep learning is replacing handcrafted features with efficient algorithms for unsupervised or semi-supervised feature learning and hierarchical feature extraction. Studies in this area attempts to make better representations and create models to learn these representations from large-scale unlabeled data. Some of the representations are inspired by advances in neuroscience and are loosely based on interpretation of information processing and communication patterns in a nervous system, such as neural coding which attempts to define a relationship between various stimuli and associated neuronal responses in the brain.

Various deep learning architectures such as deep neural networks, convolutional deep neural networks, deep belief networks and recurrent neural networks have been applied to fields like computer vision, automatic speech recognition, natural language processing, audio recognition and bioinformatics where they have been shown to produce state-of-the-art results on various tasks.

Alternatively, deep learninghas been characterized as a buzzword, or a rebranding of neural networks. Deep learning could be characterized as a class of machine learning algorithms that Use a cascade of many layers of nonlinear processing units for feature extraction and transformation. Each successive layer uses the output from the previous layer as input. The algorithms may be supervised or unsupervised and applications include pattern analysis (unsupervised) and classification (supervised).

It's a term that covers a particular approach to building and training neural networks. Neural networks have been around since the 1950s, and like nuclear fusion, they've been an incredibly promising laboratory idea whose practical deployment has been beset by constant delays. They take an array of numbers (that can represent pixels, audio waveforms, or words), run a series of functions on that array, and output one or more numbers as outputs. The outputs are usually a prediction of some properties you're trying to guess from the input, for example whether or not an image is a picture of a cat.

The functions that are run inside the black box are controlled by the memory of the neural network, arrays of numbers known as weights that define how the inputs are combined and recombined to produce the results. Dealing with real-world problems like cat-detection requires very complex functions, which mean these arrays are very large, containing around 60 million numbers in the case of one of the recent computer vision networks. The biggest obstacle to using neural networks has been figuring out how to set all these massive arrays to values that will do a good job transforming the input signals into output predictions.

A convolutional neural network consists of an input and an output layer, as well as multiple hidden layers. The hidden layers of a CNN typically consist of a series of convolutional layers that *convolve* with a

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

multiplication or other dot product. The activation function is commonly a RELU layer, and is subsequently followed by additional convolutions such as pooling layers, fully connected layers and normalization layers, referred to as hidden layers because their inputs and outputs are masked by the activation function and final convolution. The final convolution, in turn, often involves backpropagation in order to more accurately weight the end product.

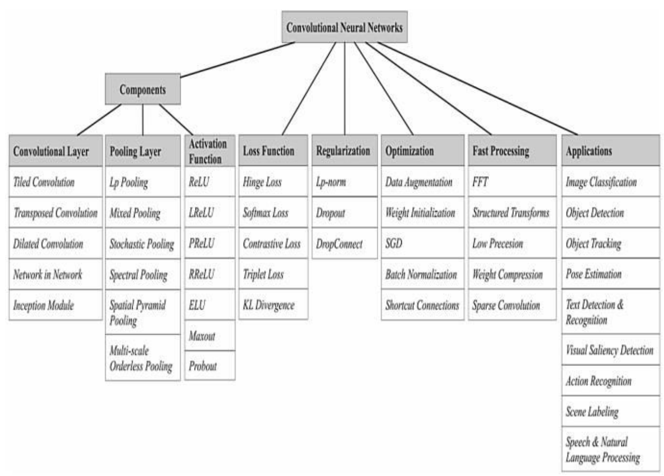
Though the layers are colloquially referred to as convolutions, this is only by convention. Mathematically, it is technically a *sliding dot product* or cross-correlation. This has significance for the indices in the matrix, in that it affects how weight is determined at a specific index point.

Fundamental Concepts

Deep learning algorithms are based on distributed representations. The underlying assumption behind distributed representations is that observed data is generated by the interactions of factors

Organized in layers. Deep learning adds the assumption that these layers of factors correspond to levels of abstraction or composition. Varying numbers of layers and layer sizes can be used to provide different amounts of abstraction. Deep learning exploits this idea of hierarchical explanatory factors where higher level, more abstract concepts are learned from the lower level ones. These architectures are often constructed with a greedy layer-by-layer method. Deep learning helps to disentangle these abstractions and pick out which features are useful for learning.

For supervised learning tasks, deep learning methods obviate feature engineering, by translating the data into compact intermediate representations akin to principal components, and derive layered structures which remove redundancy in representation. Many deep learning algorithms are applied to unsupervised learning tasks. This is an important benefit because unlabeled data is usually more abundant than labeled data. An example of a deep structure that can be trained in an unsupervised manner is a deep belief network. Deep neural networks are generally interpreted in terms of: Universal approximation theorem or Probabilistic inference.



Deep Learning Applications

There have been several studies demonstrating the effectiveness of deep learning methods in a variety of application domains. In addition to the Mixed National Institute of Standards and Technology (MNIST) handwriting challenge, there are applications in face detection, speech

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

recognition and detection, general object recognition, natural language processing, and robotics.

The reality of data proliferation and abundance of multimodal sensory information is admittedly

a challenge and a recurring theme in many military as well as civilian applications, such as sophisticated surveillance systems. Consequently, interest in deep machine learning has not been

Limited to academic research. Recently, the Defense Advanced Research Projects Agency (DARPA) announced a research program exclusively focused on deep learning. Several private organizations have focused their attention on commercializing deep learning technologies with applications to broad domains.

Lenz et al recently presented a system for detecting robotic grasps from RGB-D data using a deep learning approach which has several advantages over current state-of-the-art methods. Their approach firstly proved that using deep learning allows you to avoid using hand-engineering features, but learning them instead. Secondly, their results showed that deep learning methods significantly outperformed even well designed hand-engineered features from previous work.

Hence deep learning system with group regularization is capable of robustly detecting grasps for a wide range of objects,

Existing method

Google Flu Trends (GFT), which was introduced in 2009 was a highly influential paper in digital disease detection and inspired a lot of work in the field (Ginsberg et al., 2009). It illustrated that data which was not necessarily organized or collected for health related purposes could be used for health analysis. In our current age of big data, this is an important finding. In recent years, social media, especially Twitter, has been used for health analysis with positive results (Cassandra Harrison et al., Lamb et al., 2013; Li and Cardie, 2013; Broniatowski et al., 2013). Many of the papers detailing this sort of Twitter analysis make use of a TF-IDF representation for Tweets. These feature vector representations do not consider word semantics and are limited by the vocabulary of the dataset. One way to get around this issue is the application of deep learning. Deep learning is a branch of machine learning that has seen a lot of interest lately, having displayed state-of-the-art performance in many difficult tasks. In recent literature, deep learning has been widely applied to twitter for sentiment analysis and it has shown promising results. In America, it has been used for the surveillance of flu trends.

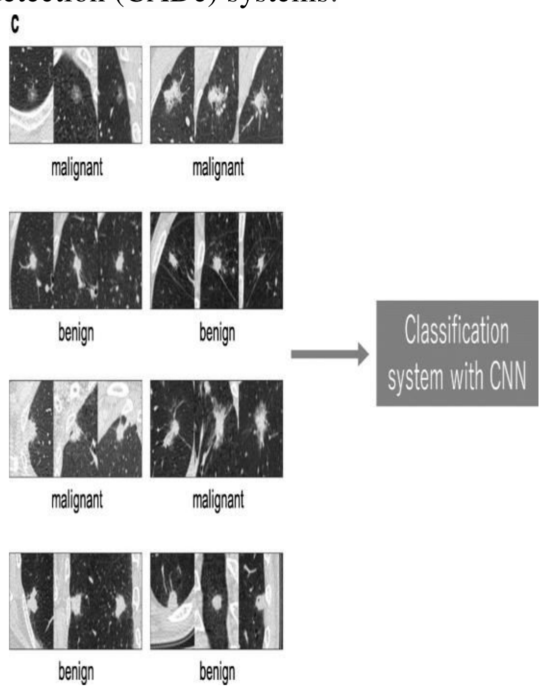
However, in addition to the lack of context and semantics, there could be a problem with the initial choice of keywords for searching and collecting Tweets. To tackle this, we want to employ deep learning methods for the exploration of an adaptive automatic keyword system. In such a system, an initial set of keywords is chosen and used to stream Tweets. The keywords most associated with relevance (i.e. the keywords that are observed to exist in the text of Tweets that are relevant) can be promoted. Words that are similar in meaning to these keywords could potentially bring in more relevant Tweets, which are currently not being collected. Semantic information obtained from deep learning would enable us to find such words. We use deep learning to find semantic information encoded in word vectors learned from deep embeddings and we use the semantic information to automatically generate alternative keywords based on word similarity. Additionally, deep neural word models are learned in an unsupervised manner. They do not require expensive labelling, but can be derived from large unannotated corpora that are easily obtainable. This means that these algorithms are prime candidates for tasks with small amounts of labelled data. We have a double objective: (i) to robustly and accurately classify Tweets for the purpose of syndromic.

Proposed method

Concepts are the foundation of human deep learning, understanding, and knowledge integration and transfer. The current technology of machine deep learning is largely at the level of surface learning in human learning, focusing on rote memorization of factual knowledge in the form of feature representations. To elevate machine deep learning toward the level of human deep learning, we proposed concept-oriented deep learning (CODL) which extends (machine) deep learning with concept representations and conceptual understanding capability. CODL leverages Microsoft Concept Graph, or something comparable, as the common / background conceptual knowledge base and the framework for conceptual understanding. In particular, concept names and concept taxonomies (isA relationships) originate from Microsoft Concept Graph. In CODL, feature representations are always learned semantically segmented in a concept-oriented manner. Concept representations are the same as concept-oriented feature representations, but from a top-down, concept-driven perspective which is the focus of CODL. It can be difficult to gather and create labeled concept representation datasets to use for training. Due to the semantically-segmented nature of concepts, a good alternative is to use concept exemplars. Concept representation learning systems provide the platforms and tools for use in CODL. They support supervised concept representation learning as well as unsupervised concept representation learning based on concept exemplars. Since, 11 in real-world scenarios, concepts and their associated data are almost always collected in an incremental manner, a good concept representation learning system must support incremental and continual learning (using concept exemplars). Also, to be effective, in CODL one should focus on learning and using concept representations for basic-level concepts.

Methodology

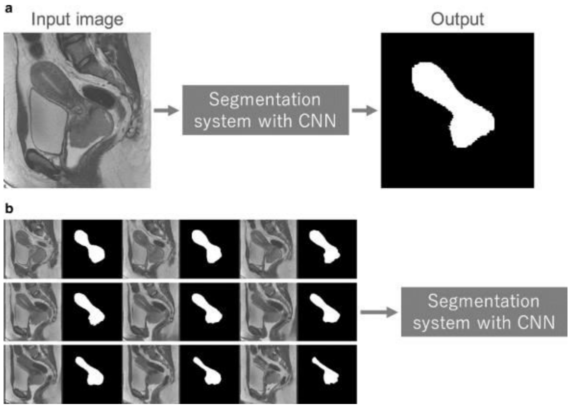
In medical image analysis, classification with deep learning usually utilizes target lesions depicted in medical images, and these lesions are classified into two or more classes. For example, deep learning is frequently used for the classification of lung nodules on computed tomography (CT) images as benign or malignant. As shown, it is necessary to prepare a large number of training data with corresponding labels for efficient classification using CNN. For lung nodule classification, CT images of lung nodules and their labels (i.e., benign or cancerous) are used as training data. Two examples of training data of lung nodule classification between benign lung nodule and primary lung cancer; shows the training data where each datum includes an axial image and its label, and shows the training data where each datum includes three images (axial, coronal, and sagittal images of a lung nodule) and their labels. After training CNN, the target lesions of medical images can be specified in the deployment phase by medical doctors or computer-aided detection (CADe) systems.



DATA CONTENTS

Segmentation of organs or anatomical structures is a fundamental image processing technique for medical image analysis, such as quantitative evaluation of clinical parameters (organ volume and shape) and computer-aided diagnosis (CAD) system. In the previous section, classification depends on the segmentation of lesions of interest. Segmentation can be performed manually by radiologists or dedicated personnel, a time-consuming process. However, one can also apply CNN to this task as well. Shows a representative example of segmentation of the uterus with a malignant tumor on MRI. In most cases, a segmentation system directly receives an entire image and outputs its segmentation result. Training data for the segmentation system consist of the medical images containing the organ or structure of interest and the segmentation result; the latter is mainly obtained from previously performed manual segmentation. A representative example of training data for the segmentation system of a uterus with a malignant tumor. In contrast to classification, because an entire image is inputted to the segmentation system, it is necessary for the system to capture the global spatial context of the entire image for efficient segmentation.

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org



Carrier Key Assumption

Training a network is a process of finding kernels in convolution layers and weights in fully connected layers which minimize differences between output predictions and given ground truth labels on a training dataset. Backpropagation algorithm is the method commonly used for training neural networks where loss function and gradient descent optimization algorithm play essential roles. A model performance under particular kernels and weights is calculated by a loss function through forward propagation on a training dataset, and learnable parameters, namely kernels and weights, are updated according to the loss value through an optimization algorithm called backpropagation and gradient descent.

Available data are typically split into three sets: a training, a validation, and a test set, though there are some variants, such as cross validation. A training set is used to train a network, where loss values are calculated via forward propagation and learnable parameters are updated via backpropagation. A validation set is used to evaluate the model during the training process, fine-tune hyper parameters, and perform model selection. A test set is ideally used only once at the very end of the project in order to evaluate the performance of the final model that was fine-tuned and selected on the training process with training and validation sets.

RESULT AND DISCUSSION

Class activation map (CAM)

Conclusion and Future Work:

Major advantage of deep learning over conventional machine learning technique is that it can independently detect relevant features in high dimensional data as compared to shallow networks. There exists sufficient literature on different deep learning techniques such as recurrent neural network, deep belief networks and CNN. This study has thrown light upon the basic understanding of CNN, which is a deep learning approach to solve many complex problems. This study has described the general model, various architectures, and two important learning algorithms of the CNN. CNN has emerged as a prominent technique used for classification based on contextual information. It has immense ability to learn contextual features and thereby has overcome the issues involved in pixel wise classification. It reduces number

International Journal for Innovative Engineering and Management Research
A Peer Revieved Open Access International Journal
www.ijiemr.org

of parameters required to a great extent. CNN is extensively being used for classification in remote sensing, ocean front recognition task, high-resolution data, traffic sign recognition [10], audio scene, and segmentation of MR brain images. This study will provide broad understanding to researchers who want to venture in this field. It will act as a means to learners, researchers and to those who are interested in this field. The adaptive keyword system could collect Tweets with an initial set of keywords and then modify this set by including words it knows are similar to words that appear often in Tweets the relevance filter finds relevant. It would also exclude words that do not tend to appear in Tweets that the relevance filter finds relevant. By repeatedly doing this over time, the set of keywords used to collect Tweets will change.

REFERENCES:

[1] Arel, I., Rose, D. C., and Karnowski, T. P. (2010) "Deep machine learning-a new frontier in artificial intelligence research [research frontier]." *IEEE computational intelligence magazine* **5** (**4**): 13-18.

[2] Carbonell, J. G., Michalski, R. S., and Mitchell, T. M. (1983) "An overview of machine learning. In Machine learning." *Springer Berlin Heidelberg* (pp. 3-23).

[3] Chen, H., Tang, Y., Li, L., Yuan, Y., Li, X., & Tang, Y. (2013) "Error analysis of stochastic gradient descent ranking." *IEEE transactions on cybernetics* **43** (**3**): 898-909.

[4] Dharmadhikari, S. C., Ingle, M., and Kulkarni, P. (2011) "Empirical studies on machine learning based text classification algorithms." *Advanced Computing* **2** (**6**): 161.

[5] Donahue, J., Jia, Y., Vinyals, O., Hoffman, J., Zhang, N., Tzeng, E., and Darrell, T. (2014) "Decaf: A deep convolutional activation feature for generic visual recognition." *In International conference on machine learning* (pp. 647-655).

[6] Fang, J., Zhou, Y., Yu, Y., and Du, S. (2017) "Fine-grained vehicle model recognition using a coarse-to-fine convolutional neural network architecture." *IEEE Transactions on Intelligent Transportation Systems* **18** (**7**): 1782-1792.

[7] Feng, J., Li, F., Lu, S., Liu, J., and Ma, D. (2017) "Injurious or Noninjurious Defect Identification From MFL Images in Pipeline Inspection Using Convolutional Neural Network." *IEEE Transactions on Instrumentation and Measurement* **66** (**7**): 1883-1892.

[8] Fieres, J., Schemmel, J., and Meier, K. (2006) "Training convolutional networks of threshold neurons suited for low-power hardware implementation." In *Neural Networks, 2006. IJCNN'06. International Joint Conference on. IEEE*. (pp. 21-28).

[9] Hinton, G. E., Srivastava, N., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. R. (2012) "Improving neural networks by preventing co-adaptation of feature detectors". *arXiv preprint arXiv*:1207.0580.

[10] Jin, J., Fu, K., and Zhang, C. (2014) "Traffic sign recognition with hinge loss trained convolutional neural networks."

IEEE Transactions on Intelligent Transportation Systems **15** (**5**): 1991-2000.
[11] Kingma, D. P., & Ba, J. (2014). “Adam: A method for stochastic optimization. ”*arXiv preprint arXiv:1412.6980*.