A Literature Study of Deep learning and its application in Digital Image Processing

Technical Report · June 2017

DOI: 10.13140/RG.2.2.17403.72480

Abdellatif Baba
Türk Hava Kurumu Üniversitesi

UNIVERSITY OF TURKISH AERONAUTICAL ASSOCIATION

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

Digital Image processing (EEE 573)

A Literature Survey of Deep learning and its application in Digital Image Processing

Esam Kamal MARIED

Mansour Abdalla ELDALI

Osama Omar ZIADA

Supervised by

Dr. Eng. Abdellatif BABA

TABLE OF CONTENTS

LIST OF FIGURES

1. Introduction

Deep learning is a subfield of machine learning, which aims to learn a hierarchy of features from input data. Nowadays, researchers have intensively investigated deep learning algorithms for solving challenging problems in many areas such as image classification, speech recognition, signal processing, and natural language processing.

2. Deep learning methods.

Deep learning methods are a group of machine learning methods that can learn features hierarchically from lower level to higher level by building a deep architecture. The deep learning methods have the ability to automatically learn features at multiple levels, which makes the system be able to learn complex mapping function directly from data, without the help of the human-crafted features.

The most characterizing feature of deep learning methods is that their models all have deep architectures. A deep architecture means it has multiple hidden layers in the network. In contrast, a shallow architecture has only a few hidden layers (1 to 2 layers).

Deep Convolutional Neural networks are successfully applied in various areas. Regression, Classification, dimensionality reduction, modeling motion, modeling textures, information retrieval, natural language processing, robotics, fault diagnosis, and road crack detection.

3. Deep Learning Algorithms

Deep learning algorithms have been extensively studied in recent years. As a consequence, there are a large number of related approaches. Generally speaking, these algorithms can be grouped into two categories based on their architectures:

1. Restricted Boltzmann machines (RBMs).
2. Convolutional neural networks (CNNs).

3.1. Restricted Boltzmann Machines. (RBMs).

RBM is an energy-based probabilistic generative model. It is composed of one layer of visible units and one layer of hidden units. The visible units represent the input vector of a data sample and the hidden units represent features that are abstracted from the visible units. Every visible unit is connected to every hidden unit, whereas no connection exists within the visible layer or hidden layer. Figure 1 illustrates the graphical model of restricted Boltzmann machine.

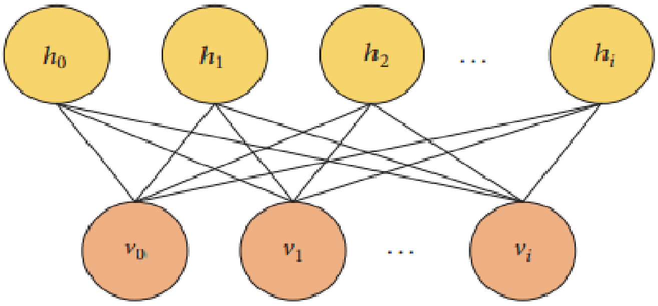

Figure (1) Restricted Boltzmann Machines

3.2. Convolutional Neural Network. (CNNs).

During the last seven years, the quality of image classification and object detection has been dramatically improved due to the deep learning method. Convolutional neural networks (CNNs) brought a revolution in the computer vision area. It not only have been continuously advancing the image classification accuracy, but also play an important role for generic feature extraction such as scene classification, object detection, semantic segmentation, image retrieval, and image caption.

Convolutional neural network (CNNs) is one of the most powerful classes of deep neural networks in image processing tasks. It is highly effective and commonly used in computer vision applications. The convolution neural network contains three types of layers: convolution layers, subsampling layers, and full connection layers. The whole architecture of the convolutional neural network is shown in Figure 2. A brief introduction to each type of layer is provided in the following paragraphs.

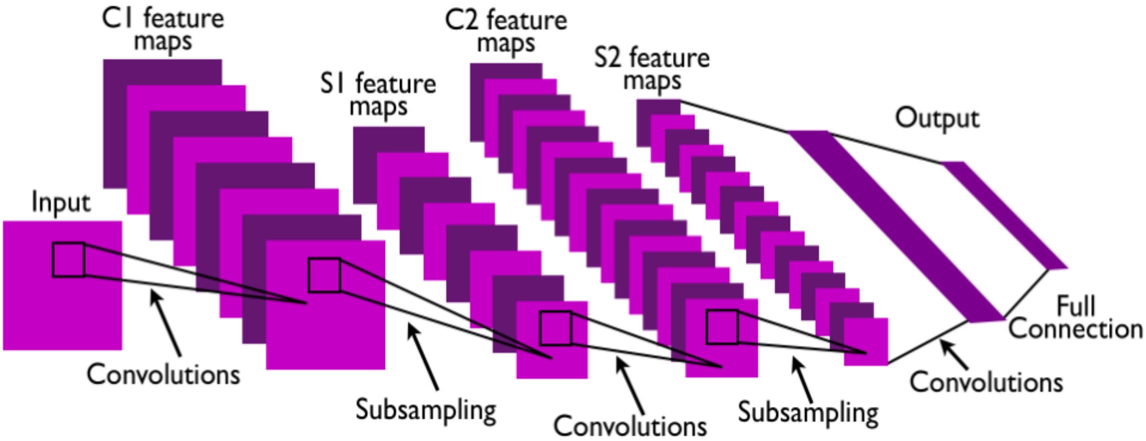

Figure (2) the architecture of convolution neural network

3.2.1 Convolution Layer.

As Figure 3 shows, in convolution layer, the left matrix is the input, which is a digital image, and the right matrix is a convolution matrix. The convolution layer takes the convolution of the input image with the convolution matrix and generats the output image.

3	15	64	22	55	62
92	213	7	32	145	34
17	178	86	33	12	21
231	87	48	5	23	234
59	56	55	45	3	218
82	97	94	33	238	44

1	1	1
1	0	2
1	0	1

Figure (3) Digital image representation and convolution matrix

Usually, the convolution matrix is called filter and the output image is called filter response or filter map. An example of convolution calculation is demonstrated in Figure 4. Each time, a block of pixels is convoluted with a filter and generates a pixel in a new image.

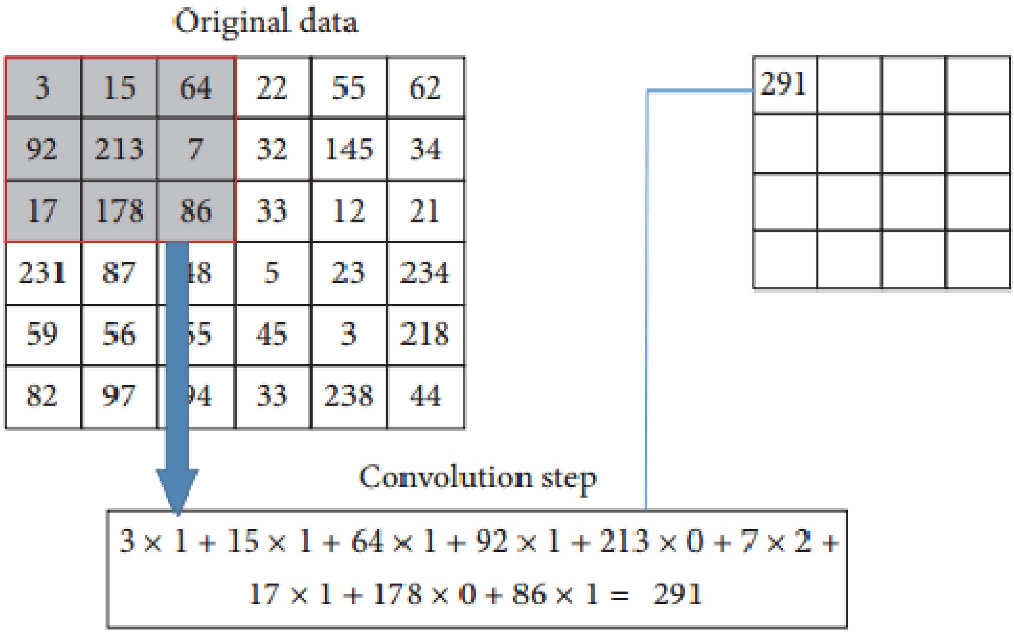

Figure (4) An example of convolution calculation

3.2.2 Subsampling Layer.

The subsampling layer is an important layer to the convolutional neural network. This layer is mainly to reduce the input image size in order to give the neural network more invariance and robustness. The most used method for subsampling layer in image processing tasks is max pooling. So the subsampling layer is frequently called max pooling layer. The max pooling method is shown in Figure 5. The image is divided into blocks and the maximum value of each block is the corresponding pixel value of the output image. The reason to use subsampling layer is as follows. First, the subsampling layer has fewer parameters and it is faster to train. Second, a subsampling layer makes convolution layer tolerate translation and rotation among the input pattern.

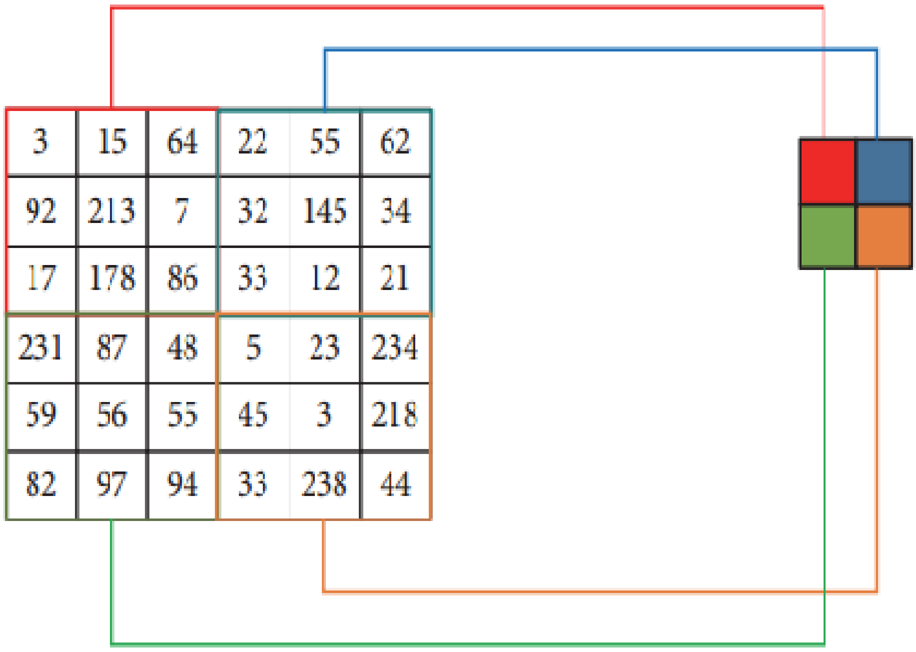

Figure (5) an example of the subsampling Layer.

3.2.3 Full Connection Layer.

Full connection layers are similar to the traditional feed-forward neural layer. They make the neural network fed forward into vectors with a predefined length. We could fit the vector into certain categories or take it as a representation vector for further processing.

4. Applications.

Deep learning has been widely applied in various fields, such as computer vision, signal processing, and speech recognition.

4.1. CNN-Based Applications in Visual Computing.

As we know, convolutional neural networks are very powerful tools for image recognition and classification. These different types of CNNs are often tested on well-known ImageNet LargeScale Visual Recognition Challenge (ILSVRC) dataset and achieved state-of-the-art performance in recent years. After winning the ImageNet competition in 2012, the CNN-based methods have brought about a revolution in computer vision. CNNs have been applied with great success to the object detection, object segmentation, and recognition of objects and regions in images. Compared with hand-crafted features, for example, Local Binary Patterns

(LBP) and Scale Invariant Feature Transform (SIFT), which need additional classifiers to solve vision problems, the CNNs can learn the features and the classifiers jointly and provide superior performance.

4.2. CNN for Face Recognition.

Face recognition has been one of the most important computer vision tasks since the 1970s. Face recognition systems typically consist of four steps. First, given an input image with one or more faces, a face detector locates and isolates faces. Then, each face is pre-processed and aligned using either 2D or 3D modeling methods. Next, a feature extractor extracts features from an aligned face to obtain a low-dimensional representation (or embedding). Finally, a classifier makes predictions based on the low-dimensional representation. The key to getting good performances for face recognition systems is obtaining an effective low-dimensional representation.

Face recognition systems using hand-crafted features include. Lawrence et al. first proposed using CNNs for face recognition. Currently, the state-of-the-art performance of face recognition systems, that is, Facebook's DeepFace and Google's FaceNet, are based on CNNs. Other notable CNN-based face recognition systems are lightened convolutional neural networks and Visual Geometry Group (VGG) Face Descriptor.

Figure 5 shows the logic flow of CNN-based face recognition systems. Instead of using hand-crafted features, CNNs are directly applied to RGB pixel values and used as a feature extractor to provide a low-dimensional representation characterizing a person's face. In order to normalize the input image to make the face robust to different view angles, DeepFace models a face in 3D and aligns it to appear as a frontal face. Then, the normalized input is fed to a single convolution-pooling-convolution filter. Next, 3 locally connected layers and 2 fully connected layers are used to make final predictions.

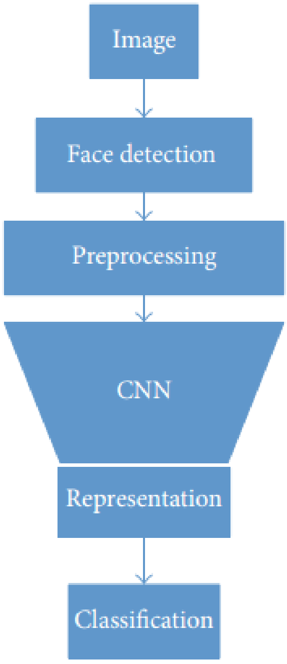

Figure (6) Logic flow of CNN based face recognition.

The architecture of DeepFace is shown in Figure 6. Though DeepFace achieves the best performance on face recognition up to date, its representation is difficult to interpret and use because the faces of the same person are not clustered necessarily during the training process. In contrast, FaceNet defines a triplet loss function directly on the representation, which makes the training procedure learn to cluster face representation of the same person. It should also be noted that OpenFace uses a simple 2D affine transformation to align face input.

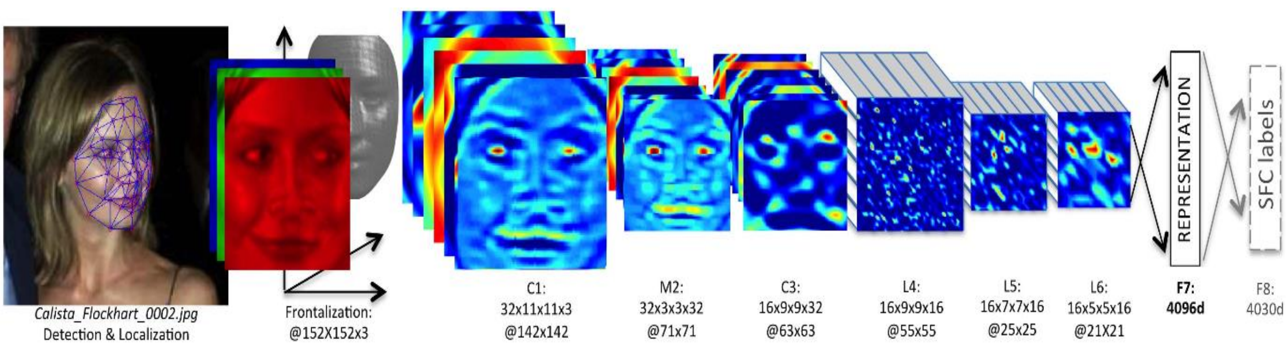

Figure (7) Outline of Deepface architecture.

Nowadays, face recognition in mobile computing is a very attractive topic. While DeepFace and FaceNet remain private and are of large size, OpenFace offers a lightweight, real-time, and open-source face recognition system with competitive accuracy, which is suitable for mobile computing.

5. Conclusion.

This report gives an overview of deep learning algorithms and their applications. Several classic deep learning algorithms such as restricted Boltzmann machines and convolutional neural networks are introduced. In addition to deep learning algorithms, their applications are reviewed and compared with other machine learning methods. Though deep neural networks achieve good performance on many tasks, they still have many properties that need to be investigated and justified.

References

[1]. G. E. Hinton, S. Osindero, and Y.-W. Teh, "A fast learning algorithm for deep belief nets," Neural Computation, vol. 18, no. 7, pp. 1527–1554, 2006.

[2]. D. Yu and L. Deng, "Deep learning and its applications to signal and information processing," IEEE Signal Processing Magazine, vol. 28, no. 1, pp. 145–154, 2011.

[3]. Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: a review," Neurocomputing, vol. 187, pp. 27–48, 2016.

[4]. P. Smolensky, "Information processing in dynamical systems: foundations of harmony theory," Tech. Rep. DTIC Document, 1986.

[5]. Y. Bengio, "Learning deep architectures for AI," Foundations and Trends in Machine Learning, vol. 2, no. 1, pp. 1–27, 2009.

[6]. Y. LeCun and Y. Bengio, "Convolutional networks for images, speech, and time series," The Handbook of Brain Theory and Neural Networks, vol. 3361, no. 10, p. 1995, 1995.