Application of Deep Learning Algorithms to Image Classification

J. D. Gallego-Posada,[†] D. A. Montoya-Zapata,[*] D. E. Sierra-Sosa[§] and O. L. Quintero-Montoya[‡]
Research Group on Mathematical Modeling
School of Mathematical Sciences
Universidad EAFIT
Medellín, Colombia
{[†]*jgalle29, [*]dmonto39, [§]dsierras, [‡]oquinte1*} *@eafit.edu.co*

***Index Terms*—Deep learning, image classification, medical images.**

I. Problem Statement

One of the main problems in computer vision is the image classification problem, which is concerned with determining the presence of visual structures in an input image. More precisely, we are given a set of input images $X = \{X_1, ..., X_n\}$ and a matrix of labels

$$Y' = \left[y_1' \cdots y_n' \right],$$

where each $y_i \in \mathbb{B}^k$ for all $i = 1, 2, ..., n$, and k denotes the number of classes. The objective is to generate a set of predictions $\hat{Y} \in \mathbb{R}_{nxk}$ based on the set X so that they resemble Y as much as possible [1]. For the kind of analysis we are interested in, we will assume that the sample labels y_i satisfy $\sum_j y_{ij} = 1$ for all $i = 1, 2, ..., n$.

Deep Learning is a subfield of machine learning which uses computational models, with hierarchical architectures composed by multiple processing layers, to learn representations of complex data such as images, sound and text [2]. Since its emergence, Deep Learning has been widely used for many applications in some artificial intelligence fields, such as natural language processing and computer vision [3].

The image classification problem has been typically solved using machine learning techniques [1], additionally, in the last decade, the use of Deep Learning has increased dramatically, due to three reasons: the increased chip processing abilities, the decrease of cost of computing hardware and the advances in the machine learning algorithms [4].

II. Objectives

General Objective

To asses the performance of Deep Learning techniques applied to the detection of specific structures in medical images.

Specific Objectives

- To perform a review on the state-of-the-art in Deep Learning.
- To synthesize the theoretical foundations for the Deep Learning techniques to be used.
- To implement a Deep Learning algorithm and benchmark it against analogue implementations of the same algorithm.

III. Justification

The interest in the application of Deep Learning techniques has dramatically increased in the last decade and it has become one of the main research fields in Computer Science around the globe [3]. Thus, any academic contribution that allows a better understanding of Deep Learning, and particularly, its applications to the classification of medical images would be a valuable resource for current and future research projects.

In addition, one of the main interests of the Research Group on Mathematical Modeling - GRIMMAT, is the problem of facial features detection, in which Deep Learning is a promising approach.

By means of this project, we want to gain understanding in Deep Learning in order to be qualified to attend *Cornell's Program for Research Experience on Deep Learning* taught by

Prof. John E. Hopcroft during summer 2016, as part of joint efforts to strengthen relationships between Cornell University and Universidad EAFIT.

IV. Preceding Research

The methods based on Bags of Visual Words (BoW), which describe the images as histograms of keywords, were some of the first methods used in image classification problems [5].

These methods were improved by the incorporation of spatial geometry, through the use of spatial pyramids [6]. In this method, the image is divided into subregions and the visual words are counted in each subregion instead of the whole image.

In 2010, Lin et al. [7] used sparce coding, a basic Deep Learning algorithm, and obtained the best performance in the ImageNet: Large Scale Visual Recognition Challenge (ILSVRC).

Perronnin et al. [8] achieved the state-of-the-art image classification result in 2011 employing the Fisher kernel to extract higher order statistics, an previously unexplored alternative due to the zero order statistics base for BoW methods.

The ILSVRC 2012 was won by the model proposed in [9] by Krizhevsky et al., which represents a break point in large scale object recognition, since a large Convolutional Neural Network had a well performing on natural image classification.

Since 2013, further improvements and different architectures of Convolutional Neural Networks had been created. Zeiler et al. [10] introduced a new visualization strategy to provide further understanding on the behaviour and function of intermediate feature layers. GoogLeNet [11] and VGG [12], who used deeper architectures and reached a better performance level. Wu et al. [13] developed new strategies to avoid overfitting and underfitting when there is little data.

Other research projects have focused on getting a further understanding of what Deep Neural Networks learn, addressing the problem from both a theoretical and a empirical perspective. For instance, Li et al. [14] have recently studied convergent learning, aiming to analyze cases in which different neural networks learn similar representations. In this work, they propose a method for quantifying the similarity between deep neural networks and showed that there exist basic features which are learned by multiple networks with the same architectures but different random initialization.

V. Scope

There exists a wide variety of Deep Learning algorithms such as Convolutional Neural Networks and Restricted Boltzmann Machines. In this case, we will focus our efforts in the implementation of a single Deep Learning algorithm for the image classification problem of determining the presence of specific structures in medical images. The particular algorithm to be implemented will be defined after performing the review of the state-of-the-art.

VI. Methodology

The review of the state-of-the-art in Deep Learning will begin with a database search on papers related to the different applications and theoretical developments in this field, extracting the most relevant aspects from each paper after having examined them. Finally, we will order chronologically the selected information to write the state-of-the-art review in Deep Learning.

To complete the second objective of the project, we will search, select and read additional papers containing the mathematical structure needed to define Deep Learning theoretically.

In third place, in order to perform the implementation of a Deep Learning algorithm and to evaluate its performance, we will first write the pseudo-code and code a preliminary version. Once we have a working implementation, we will make proofs to calibrate the parameters of this computational model and benchmark our algorithm against an analogue implementation of the same algorithm.

Furthermore, the group will hold weekly meetings, in which we will discuss progress and possible project related issues.

VII. Budget

Table I describes the budget allotments required to conduct this research.

TABLE I
Budget Requirements

Item	Cost
Research Practice Course	$ 2.500.000
Weekly meetings with tutor	$ 3.500.000
Technical computing software	$ 12.000.000
Computer usage	$ 100.000
Books and papers	$ 300.000
Office materials	$ 20.000
Total	$ 18.420.000

TABLE II
PROPOSED ACTIVITY SCHEDULE BY WEEK.

Specific Goal	Week	Activity
O1. State of the Art Review	1	Search and of papers related to the state of the art in Deep Learning
	2	Initial reading of papers and writing of the Research Proposal
	3	Preparation of the Proposal Presentation
	4	Further reading of papers related to the state of the art in Deep Learning
	5	Writing of the review on the state of the art in Deep Learning
O2. Theoretical Foundations	6	Analysis and pre-processing of input images on which the algorithm(s) will be applied
	7	Review on possibly different theoretical structures on Deep Learning
	8	Writing of the mathematical foundations for Deep Learning definitions, techniques and algorithms
	9	Making of the slides and preparation of the Progress Presentation
O3. Implementation of Algorithm(s)	10	Coding of the preliminary version of the Deep Learning algorithm(s)
	11	Coding of the preliminary version of the Deep Learning algorithm(s)
	12	Coding the final version of the Deep Learning algorithm(s)
	13	Performance evaluation of the implemented algorithm(s) against existing implementations
	14	Writing of the preliminary version of the Article
Article Finishing	15	Writing of the final version of the Article - Preparation of the Project Presentation
	16	Writing of the final version of the Article
	17	Creation of the slides for the Project Presentation
	18	Preparation of the Project Presentation
	19	Final Project Presentation

VIII. SCHEDULE

Table II displays the proposed activities by week in order to reach the mentioned project objectives.

IX. INTELLECTUAL PROPERTY

According to the internal regulations on intellectual property within Universidad EAFIT, the results of this practice are product of the co-autorship between Prof. Dr. Olga Lucia Quintero-Montoya, Prof. Dr. Daniel Esteban Sierra-Sosa, and students Jose Daniel Gallego-Posada and Diego Alejandro Montoya-Zapata.

In case further products, beside academic articles, should be generated from this work, the intellectual property distribution related to them will be directed under the current regulation of this matter determined by Universidad EAFIT.

REFERENCES

[1] D. Novotný, "Large Scale Object Detection," Ph.D. dissertation, Czech Technical University, 2014.

[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[3] Y. Guo, Y. Liu, A. Oerlemans, S. Lao, S. Wu, and M. S. Lew, "Deep learning for visual understanding: A review," *Neurocomputing*, 2015.

[4] L. Deng, "A tutorial survey of architectures, algorithms, and applications for deep learning," *APSIPA Transactions on Signal and Information Processing*, vol. 3, no. January, p. e2, 2014.

[5] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," *Proceedings of the ECCV International Workshop on Statistical Learning in Computer Vision*, pp. 59–74, 2004.

[6] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, 2006, pp. 2169–2178.

[7] Y. Lin, F. Lv, S. Zhu, M. Yang, T. Cour, K. Yu, L. Cao, and T. Huang, "Large-scale image classification: Fast feature extraction and SVM training," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2011, pp. 1689–1696.

[8] F. Perronnin, Y. Liu, J. Sánchez, and H. Poirier, "Large-scale image retrieval with compressed Fisher vectors," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3384–3391.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances In Neural Information Processing Systems*, pp. 1–9, 2012.

[10] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proceedings of the ECCV International Workshop on Statistical Learning in Computer Vision*. Springer, 2014, pp. 818–833.

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[12] K. Simonyan and A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition,” *Proceedings of the ICLR*, pp. 1–14, 2015.

[13] R. Wu, S. Yan, Y. Shan, Q. Dang, and G. Sun, “Deep Image: Scaling up Image Recognition,” *Arxiv*, p. 12, 2015.

[14] Y. Li, J. Yosinski, J. Clune, H. Lipson, and J. Hopcroft, “Convergent Learning: Do different neural networks learn the same representations?” in *ICLR*, 2016, pp. 1–21.