Expert Systems With Applications 85 (2017) 114–122

Contents lists available at ScienceDirect

Expert Systems With Applications

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

# Deep learning for biological image classification

Carlos Affonso[a,*], André Luis Debiaso Rossi[a], Fábio Henrique Antunes Vieira[a], André Carlos Ponce de Leon Ferreira de Carvalho[b]

[a] UNESP - Universidade Estadual Paulista, Julio de Mesquita Filho, Brazil
[b] ICMC - USP Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, Brazil

ARTICLE INFO

*Article history:*
Received 3 June 2016
Revised 11 May 2017
Accepted 16 May 2017
Available online 17 May 2017

*Keywords:*
Wood classification
Deep learning
Image classification
Machine learning

ABSTRACT

A number of industries use human inspection to visually classify the quality of their products and the raw materials used in the production process, this process could be done automatically through digital image processing. The industries are not always interested in the most accurate technique for a given problem, but most appropriate for the expected results, there must be a balance between accuracy and computational cost. This paper investigates the classification of the quality of wood boards based on their images. For such, it compares the use of deep learning, particularly Convolutional Neural Networks, with the combination of texture-based feature extraction techniques and traditional techniques: Decision tree induction algorithms, Neural Networks, Nearest neighbors and Support vector machines. Reported studies show that Deep Learning techniques applied to image processing tasks have achieved predictive performance superior to traditional classification techniques, mainly in high complex scenarios. One of the reasons pointed out is their embedded feature extraction mechanism. Deep Learning techniques directly identify and extract features, considered by them to be relevant, in a given image dataset. However, empirical results for the image data set have shown that the texture descriptor method proposed, regardless of the strategy employed is very competitive when compared with Convolutional Neural Network for all the performed experiments. The best performance of the texture descriptor method could be caused by the nature of the image dataset. Finally are pointed out some perspectives of futures developments with the application of Active learning and Semi supervised methods.



## 1. Introduction

Quality analysis activities are often used by industries to ensure the quality of their products. These activities are usually carried out by human inspection, mainly by visually scanning the products in a production line. The inspection allows correction of problems and discards of defective products, resulting in a better quality of the final production. However, the use of human beings in the quality assessment adds subjective factors to this process and, due to problems like distraction, stress, and fatigue, can accept products whose quality is below the desired level. These problems show the importance of the use of efficient image classification techniques to improve the quality control in production lines (Affonso, Sassi, & Barreiros, 2015).

Frequent obstacles that arise when using these techniques are the design and tuning of automated image classification system since various aspects must be taken into consideration. Besides, considering that wood presents, as natural raw material, a variety of macroscopic and physical features, such as weight (different moisture content), color (variation), odor, hardness, texture, and surface appearances, its distinction becomes even harder.

In recent years, important efficiency gains have been achieved by machine vision systems, due to the development of high technology camera sensors and advances in processing capacity. Meanwhile, the price of systems based on cameras has decreased, enabling a cost-efficient classification solution environment for the quality of an extensive variety of products.

In complex problems as image classification, the capture of the essential features must be carried out without a priori knowledge of the image. Therefore, modeling by traditional computational techniques is quite difficult, considering the complexity and non-linearity of image systems.

Although texture has not a clear definition, such descriptors have a wide application on image classification, computer vision, and similar fields. Hossain and Serikawa (2013) surveyed a group of texture datasets from related to different areas of medicine and natural environment.

* Corresponding author.
*E-mail addresses:* affonso@itapeva.unesp.br, affonso.unesp@gmail.com (C. Affonso), alrossi@itapeva.unesp.br (A.L.D. Rossi), fhavieira@itapeva.unesp.br (F.H.A. Vieira), andre@icmc.usp.br (A.C.P.d.L.F. de Carvalho).

http://dx.doi.org/10.1016/j.eswa.2017.05.039
0957-4174/

Local binary patterns (LBP) is one of the most used descriptor considering its resistance to light changes, low computational cost and ability to classify using fine details (Nanni, Brahnam, & Lumini, 2012). Different texture descriptors have been proposed such as, Histograms of Oriented Gradient (Dalal & Triggs, 2005), wavelets (Unser, 1995) and Gabor filter (Mehrotra, Namuduri, & Ranganathan, 1992). However the most traditional is Haralick's texture descriptor (Haralick, Shanmugam, & Dinstein, 1973).

Due to the shortcomings of the manual process, Machine Learning (ML) algorithms have been widely used for classification and clustering of wood materials (Gonzaga, de Franca, & Frere, 1999). The representation of the data provided as the "experience" to these algorithms has strong influence on their performance (Bengio, 2009).

A number of features usually requires computational complexity and even greater runtime. Moreover, the noise in the database caused by excessive image features can cause a reduction in its capacity of representation.

In recent studies, Deep Learning (DL) techniques presented better predictive performance than state-of-the-art algorithms in many domains, including image classification (Krizhevsky, Sutskever, & Hinton, 2012). DL deals with the problem of data representation by introducing simpler intermediate representations that allow to combine them in order to build complex concepts. Therefore, it is unnecessary to apply many preprocessing techniques to extract features, which represent the image data (Bengio, 2009).

Artificial Neural Networks (ANN) are a quintessential example of DL model. Although ANN dates back to the 1950s, researchers now are able to train deeper structures than it had been possible before. The idea of using a higher number of layers, the multilayer network, is justified by the learning algorithm used to train a network in image classification (Al-Allaf, Abdalkader, & Tamimi, 2013).

On the other hand, because of its complex structure, DL needs a large volume of data to generate models with high predictive performance and, consequently, has high computational cost. Whereas recent works suggest that deep architectures might be more accurate, their training was unsuccessful until the recent uses of unsupervised pre-training (Bengio, 2009). That could happen because the gradient-based training of Convolutional Neural Network (CNN) achieves some local minimum and additionally for deeper architecture became more difficult to obtain satisfactory result (Bengio, Lamblin, Popovici, Larochelle et al., 2007; Erhan, Manzagol, Bengio, Bengio, & Vincent, 2009).

A successful DL algorithm based on ANN is the CNN. Some studies suggest CNN is superior to traditional learning algorithms, such as K-Nearest Neighbors (KNN), Multilayer Perceptron (MLP), and Support Vector Machines (SVM) for image classification (Chen, Xiang, Liu, & Pan, 2014; Ferreira & Giraldi, 2017; Makantasis, Protopapadakis, Doulamis, Doulamis, & Loupos, 2015; Neubauer, 1998; Norlander, Grahn, & Maki, 2015; Park, Kwon, Park, & Kang, 2016).

The literature points out DL Architecture as the best performance solution for image classification. However, its computational cost is much higher than the traditional techniques using texture descriptors (Gibert, Patel, & Chellappa, 2017; Krizhevsky, Sutskever, & Hinton, 2012).

The main objective of this paper was to analyze CNN to classify wood boards regarding their quality. The models were trained on a image data set, where each instance is a collection of pixel values representing a wood board image in grayscale. This data set presents three classes, according to their quality, with restricted examples.

Moreover, the CNN performance is compared to traditional learning algorithms, namely DT, ANN, and KNN. Each instance of the training data for these algorithms is a set of texture descriptors extracted from a wood board.

Since there is a "limited" number of instances to train the CNN, it is hypothetical that CNN achieves similar predictive performance compared to these algorithms using texture descriptors.

In this way, the traditional techniques using texture descriptors seems to be the smart choice for this investigation problem, once they present the advantage of lower computational cost.

In order to test the hypothesis, the classification accuracy for these two approaches was validated, such as DL techniques (through CNN) and Texture Descriptors (through Haralick's descriptors).

The next sections are organized as follows. Section 2 gives a brief introduction of DL techniques and focuses on relevant similar studies. Section 2.3 describes the texture descriptor methods proposed and investigated in this paper. Section 4 presents the experimental evaluation of DL architectures and discusses the results. Finally, Section 6 presents the main conclusions and future research directions.

## 2. Deep learning

The recent revival of DL techniques was triggered by the works on learning representations, or more traditional models (Hinton et al., 2012). DL architectures appear to solve problems that require complex highly-varying functions. Besides that, they usually involve such problems with very large, and in most cases, non-labeled data set.

In order to deal with it, DL techniques learn characteristic hierarchies with features from higher levels of hierarchy formed by a composition of lower level features (Bengio, 2009).

DL assimilates complex behaviors with expansive information sets to select effective characteristics automatically by neural network structures in quite profound layers. The model achieves such goals adopting unsupervised layers succeeded by supervised ones, applying learning-teaching to signal data (Kim, Choi, & Lee, 2015).

### *2.1. Convolutional neural network*

CNN has attracted a high interest in the image and speech classification scientific communities, since its topology is more similar to biological systems. Another main characteristic is its receptive fields, which was inspired by the cat's visual cortex (Hubel & Wiesel, 1962).

The CNN topology is based on three main concepts, namely: local receptive fields, shared weights and spatial or temporal sampling (LeCun, Bengio, & Haffner, 1998). CNN can eliminate the feature extraction process imputing the network directly with normalized images. Typically, an image data set contains many hundred pixels.

If a full network is considered, each neuron is connected to every pixel. Therefore, the computational cost and the memory requirements would be unfordable. Another deficiency on the unstructured fully connected network for image classification is its non-acceptance for local distortions on receptive fields.

The spatial invariance is obtained through the shared weight across the image (LeCun et al., 1990). Subsampling is an important strategy in object recognition, as it helps achieve invariance to distortions of the visual image.

Because of its own nature, image data set has a strong spatial correlation. In order to deal with that, CNN restricts the receptive fields. Another characteristic of CNN is the shared weights. In this way, a set of pixels in a receptive field located at different places on an image, has identical weight vectors, which outputs constitute a feature map.

This operation could be considered as the convolutional transformer. Each feature map is followed by a layer that performs a lo-