Minerals Engineering 115 (2018) 68–78

Contents lists available at ScienceDirect

Minerals Engineering

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

Froth image analysis by use of transfer learning and convolutional neural networks

Yihao Fu[a], Chris Aldrich[a,b,*]

[a] Department of Mining Engineering and Metallurgical Engineering, Western Australian School of Mines, Curtin University, GPO Box U1987, 6845 WA, Australia
[b] Department of Process Engineering, Stellenbosch University, Private Bag X1, Matieland, 7602 Stellenbosch, South Africa

A R T I C L E I N F O

Keywords:
Froth flotation
Image analysis
Convolution Neural Networks
AlexNet
Deep learning
Machine vision

A B S T R A C T

Deep learning constitutes a significant recent advance in machine learning and has been particularly successful in applications related to image processing, where it can already surpass human accuracy in some cases. In this paper, the use of a convolutional neural network, AlexNet, pretrained on a database of images of common objects was used as is to extract features from flotation froth images. These features could subsequently be used to predict the conditions or performance of the flotation systems. Two case studies are considered. In the first, froth regimes in an industrial flotation plant could be identified significantly more reliably with the features generated by AlexNet than with previous state-of-the-art approaches, such as wavelets, grey level co-occurrence matrices or local binary patterns. In the second case study, the arsenic concentration in the batch flotation of realgar-orpiment-quartz mixtures could be predicted more accurately than was possible with features extracted by wavelets, grey level co-occurrence matrices, local binary patterns or by use of colour. These results suggest that feature extraction with convolutional neural networks trained on complex data sets from other domains can serve as more reliable methods than previous state-of-the-art approaches to froth image analysis.

1. Introduction

Froth flotation is well established as the most widely used and important separation technology in mineral processing (Fuerstenau et al., 2007). Selective floating of mineral species in an aerated pulp can be achieved for a wide range of minerals and particle sizes. Although considerable advances have been made in the modelling and understanding of flotation systems, control of froth flotation processes is still complicated by the fact the key performance indicators of the process, namely recovery and grade, can generally not be measured online (Shean and Cilliers, 2011).

However, images of the froth itself is a rich source of information regarding the operational state of the flotation system (Aldrich et al., 2010) and froth image analysis provides a viable approach to the development of sensors that can be used in online control systems, such as discussed by Shean and Cilliers (2011). These sensors can infer the state of the process (Aldrich et al., 1997; Xu et al., 2015; Zhang et al., 2016) or in some cases can be used to estimate grades online, such as estimating the ash content in coal (Cruz and Adel, 1998; Zhang et al., 2014; Tan et al., 2016), platinum in PGM flotation (Marais and Aldrich, 2010, 2011a, 2011b), as well as base metals (Kaartinen and Koivo, 2002; Duchesne et al., 2003; Kaartinen et al., 2006a, 2006b; Runge et al., 2007).

The basic approach to accomplish this generally consists of two stages. In the first stage, features are extracted from the froth image and in the second stage these features are used as predictors for some froth condition or key performance indicator. The feature engineering process in the first stage could be as simple as capturing the colour of the froth, as an indication of the mineral species being floated, or it could be more complicated, where the condition of the froth can be inferred from the bubble size distribution, the froth viscosity, stability, etc.

This assumes the availability of reliable knowledge with regard to the underlying physical mechanisms of the flotation process that may not be realistic. Therefore feature extraction in the first stage may not necessarily capture all the relevant information required by the model, as it may be difficult to know a priori what features would be reliable predictors of the froth state.

For example, when the goal is to estimate the concentration of the species being floated, the colour of the froth may not necessarily give a reliable indication of the concentration of the species in the froth, as multiple species may contribute to the colour. The same goes for other features, such as the froth stability that may be increased by increased

* Corresponding author at: Department of Mining Engineering and Metallurgical Engineering, Western Australian School of Mines, Curtin University, GPO Box U1987, 6845 WA, Australia.
E-mail address: chris.aldrich@curtin.edu.au (C. Aldrich).

http://dx.doi.org/10.1016/j.mineng.2017.10.005
Received 18 July 2017; Received in revised form 18 September 2017; Accepted 7 October 2017
0892-6875/

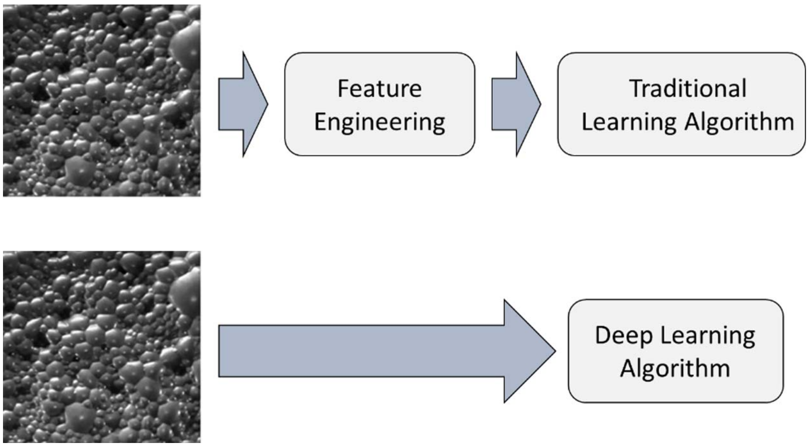

Fig. 1. Development of froth image models, using traditional methods (top) and deep learning algorithms (bottom).

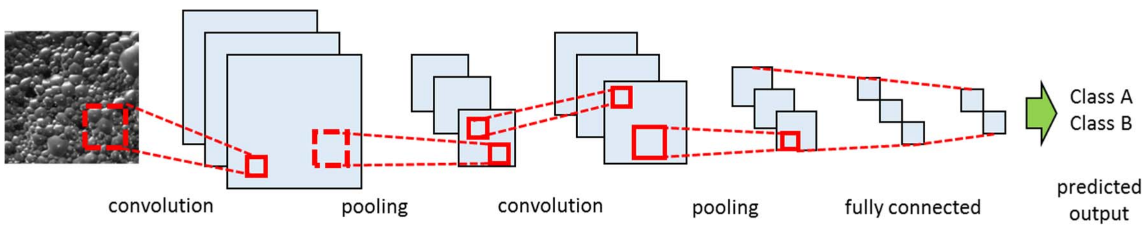

Fig. 2. Typical architecture of a convolutional neural network.

solids loading, but which again may not necessarily be a reliable indication of the mineral species in the froth. Under these circumstances, feature engineering can become a costly and cumbersome trial and error process.

The second stage, following feature extraction is comparatively inexpensive, with a suitable machine learning algorithm applied to construct a model that can predict the labels associated with the froth images. These labels can be discrete, e.g. to identify different froth classes or operating conditions, or they can be continuous, or numeric, to indicate the concentrations of reagents or mineral species in the froth or some other quantitative key performance indicator of the flotation cell. Models based on artificial neural networks have been popular choices in this regard (Moolman et al., 1995; Marais and Aldrich, 2010; Jahedsaravani et al., 2014).

In principle, a better approach would be to directly and automatically guide the feature extraction process based on the predictive power of the features in the model itself, as indicated in Fig.1. In this figure, the traditional two-stage approach (top) is contrasted with a single stage approach, where feature extraction is done internally by the model itself (bottom).

In this paper it is shown that supervised feature extraction, i.e. the extraction of features to maximise image recognition, can yield significantly better results than what could be achieved with features not directly extracted to achieve the same goal. In addition, it is shown that this can be achieved by making use of convolutional neural networks that have been pretrained on image data from a different domain.

Image analysis with convolutional neural networks is briefly discussed in the following section. In Section 3, the analytical methodology is summarised, together with four other algorithms that are used here as a basis for comparison regarding the performance of the neural network feature extractor. In Sections 4 and 5, application of these algorithms to two different froth systems are considered. Further discussion and conclusions are provided in Sections 6 and 7.

2. Image analysis with convolutional neural networks

Convolutional neural networks (CNNs) belong to a class of deep, feed-forward artificial neural networks that have successfully been applied to image analysis. They are biologically inspired variants of multilayer perceptrons that emulate the animal visual cortex, the most powerful visual processing system in existence.

The main advantage of CNNs over traditional fully connected neural networks is that they have comparatively fewer parameters to learn. Convolutional layers with small kernels are an effective means of extracting high level features that are fed to fully connected layers. Training of CNNs is accomplished by use of backpropagation and stochastic gradient descent (Rumelhart et al., 1986).

CNNs typically consist of the following types of layers, as indicated in Fig.2. The first is an input layer, where data (images) are presented to the network. This is followed by convolutional layers that contain a series of fixed sized filters to perform convolution on the image data to generate so-called feature maps. These filters can highlight some patterns helpful for image characterization, such as edges and textures.

Pooling layers summarise the data by sliding windows across the feature maps and applying some linear or nonlinear operations, such as calculating the local mean or maximum values, to ensure that the network focuses only on the most important patterns. Fully connected layers are used to interpret patterns generated by the previous layers. In addition to these, rectified linear units can be used to facilitate training of CNNs by applying non-linear functions to the outputs for faster convergence. Finally, so-called loss layers are used to specify how network errors are penalised during training of the network and can include loss functions, such as softmax and sigmoidal cross-entropy.

As deep neural networks, CNNs can have a very large capacity to capture complex features from image data and over the last few years, these networks have emerged as state-of-the-art approaches to image recognition, often outperforming traditional approaches by a large margin (Simonyan and Zisserman, 2014; Kheradpisheh et al., 2016).

The annual ImageNet Challenge (Russakovsky et al., 2015) is a good example of this. In this competition, some 1.2 million images are available to train models to recognise any of a 1000 everyday objects in diverse settings, while testing is done on 100,000 images not previously seen by the model. The error rates of the winners over the last seven years are shown in Fig. 3.