Application of Deep Learning with Convolutional Neural Networks in acoustic/seismic image classification

Satyanarayana Yegireddi
Centre for High Energy Systems and Sciences (CHESS), Defence R&D Organisation , Hyderabad

ysatya@yahoo.com

Keywords

Artificial Intelligence, Deep Learning, Convolutional Networks, Acoustic images, Image classification, subbottom profiler

Summary

Classification of acoustic or seismic images is a highly challenging and computationally intensive and efficacy of any algorithm depends on the order of pre-processing and quality of images. Artificial intelligence techniques based on Machine Language and Neural Networks found to be very promising in various fields of engineering and science applications. Deep learning using Convolutional Neural Networks (CNN), a subset of AI reported to be promising and enhances the efficacy of classification. The present study deals with application of CNN to demonstrate its potential in classification of subbottom seismic/acoustic images associated even with a low contrast in their texture. The results are quite promising and CNN classifies over 75 % of images correctly even with limited data used in training of the network. The results can be improved with more number of images data sets in training.

Introduction

Subbottom profilers usually operated relatively at a very high frequency compared to low frequency sources used in seismic exploration, provide high resolution shallow subbottom acoustic image data. Due to high frequency of source, the penetration of acoustics signal limited to very shallower depths ranging from a few tens of meters to a few hundreds of meters depending the sediment characteristics. The data is very useful in offshore geotechnical, geo-engineering applications to understand the subbottom structure, sediment physical and acoustic properties estimation qualitatively. Since the data pertains to shallower depths, often the quality of images are noisy due to multiple reflections. Classification of such images to distinguish weak patterns, textural features represented by pixel intensities variation corresponding changes in the sediments characteristics of will be complex.

A good number of techniques employed in image processing, exploiting the spatial variability pixel values in the image are known for detection of anomalous features with high precision, associated with very weak signals, particularly employed in biomedical, remote sensing applications and other high precision imaging areas. Many image processing techniques exploit statistical measures (Pham Dzung et al, 2000; Hossein Mobahi et al, 2011), thresholding pixel intensities levels (Batenburg and Sibers, 2009, Sonka et al 1993) inqualitative and quantitative analysis. Image texture based classification techniques applied to various types of images like remote sensing, land based applications, medical imaging, etc are well reported (Conners, 1980, Weszka, et al., 1976; Haralick et al.,1973 and Caelli, 1988). The most popular second-order statistical features for texture analysis are derived from the so-called co-occurrence matrix (Haralick 1979).

Artificial Intelligence has been witnessing a phenomenal growth in bridging the gap between the capabilities of humans and machines. In a regular neural network each neuron is connected to all the neurons of previous and following layers, ie, fully connected layer and the last layer is output layer. Satyanarayana and Nithesh (2014) have demonstrated the classification and extraction of structural features of subbottom profiler images combining textural analysis and SOM of neural network. However if input image, and with multiple layers in the networks become computationally intensive and

Deep Learning Networks in acoustic/seismic image classification

time consuming as the number of pixels increase exponentially with adding more layers.

A Convolutional Neural Network (CNN) is a Deep Learning algorithm (Nikhil and Nicholas, 2017)take images as input, assigns importance to various aspects in the image and be able to differentiate one from the other. CNN is capable of capture the spatial and temporal dependencies in an image. In the process, the network reduces the dimension of input images during convolution, retaining the significant features which will be input to following layer. This helps in improving efficacy of the model not only in prediction but also to deal with scalable massive data sets. As there is reduction in number of parameters compared pixel to pixel, the network can be trained better to represent even complex images.

Convolutional Neural Network Architecture

The objective of the convolution operation is to extract the low to high-level features from the input images. It can be more than one convolutional layers. Thus deep learning networks with multiple hidden layers have ability to learn and resolve fine features. They are quite popular in medical image analysis where precision is very crucial ((Justin et al, 2017; Tajbakhsh et al., 2016) Conventionally, the first layer or initial layers are responsible for extract low-level features such as edges, etc. Later layers able to capture the high-Level features and enables the network in better understanding of images in the data set during training. Provision of padding in input image matrix allows to keep dimension of convolved feature map either increased or remain same.

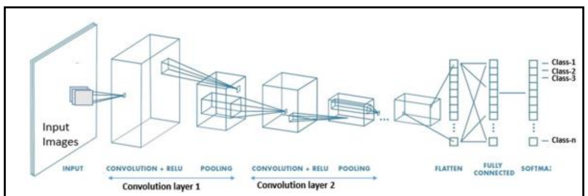

Figure1. Architecture of Convolutional Neural Network

A typical architecture of CNN networks (Figure 1) consists of *Input layer, Convolution Layer, ReLu Layer, Pooling Layer, Fully Connected Layer, Softmax Layer and Classification Layer. Input Image Layer* : 2D-Images to be classified forms the input data to Input layer. The images to be defined by specific size and either in color or grayscale. Convolution layer is to extract and learn the coarse to fine features by convolving the images with suitable filters. In fact, feature learning block consists of *Convolution Layer, Batch normalisation Layer, ReLu Layer* and *pooling Layer*. *Convolution Layer* is defined by number of filters and their size to scan the images. The number filters indicate the number feature maps generated during convolution. To keep same output size after convolution, zero padding can be used. *Batch Normalisation Layer* occur between Convolution and ReLu layers normalises the data to speed up network and reducing sensitivity to network initialization. *ReLu Layer* is the rectified linear unit (*ReLuU*) is the activation function threshold the data. *Convolution layers* including *ReLu*, reduces the spatial size of the feature map and removes some redundant spatial information. Still if we need to condense the spatial size of the representation to reduce the number of parameters**,** *Pooling Layer* is used to speed up the computation of network. Two types pooling commonly used are max pooling and average pooling of layers, defined by filter size and step size of scanning. The CNN network can have one or multiple convolution Layers. As the name suggest *Fully Connected Layer is a* fully connected layer in which the neurons connect to all the neurons of proceeding layer, combines features to classify the images. The output size is equal to number of classes. The *Softmax Layer* normalizes the output of previous layer indicate the probability of classification results. Finally *Classification Layer* uses the probabilities from previous and predicts the labels of classification to validation data and also provides accuracy of prediction. ie percentage of match with actual and predicted.

Data augmentation

In the present study a high resolution subbottom image of offshore continental shelf acquired by a hull-mounted subbottom profiler with operational parametric frequencies ranging from 6 kHz to 12 kHz is used. The image with size 4500x600 pixels represents approximately 500m horizontal distance and depth of nearly 70m along y-axis. The reflection data pertains to normal incidence provide information on vertical section of subsurface structure. It is found that hardly any structural or stratigraphy features present due to small subbottom section represented by image. The variation in pixel intensity values attributed to variability in acoustic impedance contrast and reflectivity is a function of subbottom composition and its characteristics. It may be observed from the image that no significant demarcation of any stratification of the sediment excepting a very weak reflector at shallow depth. If any, due to low impedance contrast, sometimes may get masked in noisy background. However, with increase of depth, there will be definitely at least some minor variation expected in grain size of sediments, compactness, etc., which may not be significantly reflected in the image. The objective of present study is to verify and classify, if any change in the image properties with increase of depth at a shorter spacing, assuming same class of sediments and corresponding image properties remain unchanged along the horizontal axis, at any given depth.

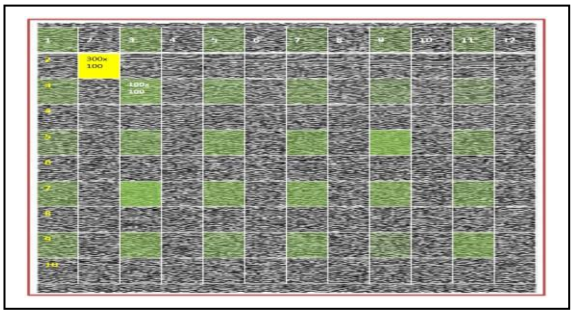

Figure 2. High resolution shallow sub-bottom image used for classification. shaded portion sub-blocks used in training, rest used for validation.

Selection of the images for classification is very critical. With above mentioned criterion, and limitations, the image (4500x600) has been sampled into sub image blocks (100x100) in such way that 6 types of sediment classes along the depth axis and each class is separated by 300 rows. Further, for each class, 12 sub-image block samples selected, to have a total a number of samples of 60 available for classification. It may challenging for any classification approach, as they are closely spaced and associated with low contrast in image properties, particularly, subbottom images. Figure 2. Shows location of samples with in the image. Out of 60 samples 3 samples (shaded images) used to train the Deep Learning Network and remaining 30 used for validation and testing. The images are compressed to 32x32 to reduce computations and run time of network and further converted them into 'tif' image format. Figure 3. Shows the 30 validation samples images (unshaded), not used in training the network for testing the network.

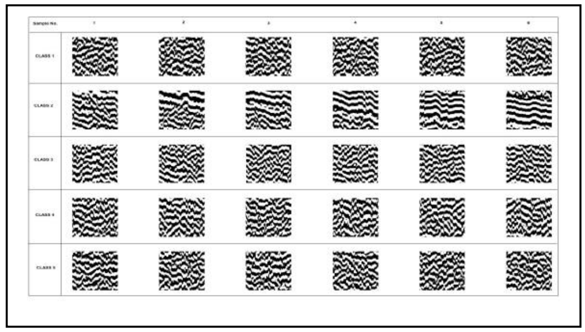

Figure 3. Sample images of 5 classes used for Validation and Testing of Network.

Application of CNN in classification of Images

30 out of 60 sample grayscale images representing 5 classes and each class contain 6 samples form the input images to the network. Remaining 30 images used for validation and testing, which are not used in network training.

Table 1. Some specifications of CNN network	
Input image sample Size (compressed from100 x100)	32x32 (Grayscale)
No of Classes	5
No. of Training data sets (each class 6 sample images)	30
No. of Validation /Test data sets	30
No of convolutional layers	3
Filter sizes	3x3
No of filters	8
Pooling	Max pool
Output Size –fully Connected layer (No of Classes)	5
Network	MLP
Network Training	Stochastic Gradient Descent
Network learning Rate	0.01
No of Epochs	12
Network output size	5Nodes (classes)

In the present network, for extracting feature maps, three convolution layers with 3x3 size filters are used. The number of nodes in the output of fully connected layer set to 5 matching with the number of classes (Table 1). Learning rate of 0.01, stochastic Gradient Descent momentum (SGDM) with feed forward network found to be reasonable good for training the network in less than 12 epochs.

Figure 4 shows the percentage of accuracy with the progress of training the network during the different epochs. The soft max layer finds the probabilities from the output of fully connected layer and further, classification layer assigns the corresponding Class Label to one with maximum probability value. The analysis was carried out using the Computer Vision and Image Processing and Deep Learning Network tools of Matlab (2018).

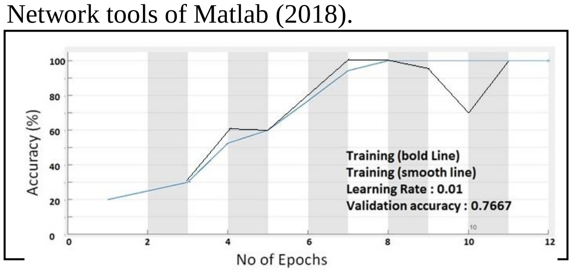

Figure 4. Accuracy of training of image data Vs Epochs

Results and Conclusion

Classification of validation data, not used in training, shows reasonably good results, which may not be expected with the other methods as the sample images (sub- images) selected are very closely spaced vertically and have a little contrast among them. The Network well learned and trained even with 30 sample images representing 5 classes indicates the efficacy of the model. Out of 30 test samples used in validation, 23 samples correctly classified shows 76% of accuracy. Two out 6 of Class1 samples wrongly classified as Class 5, while two samples each of Class3 and Class4 and one of Class 5 wrongly classified as Class2. Interestingly all the samples of Class2 are classified correctly. Mis-classification suggests that all the classes have some common features sharing with Class2.

The efficiency in terms of runtime and improvement in results can be improved by augmenting with larger data base. Increase in the hidden layers help to learn, temporal and spatial variability of between pixels within the images, particularly when large database of complex images to be classified.

References

Batenburg K.J. and J.Sibers, 2009, Optimal threshold selection for tomographic segmentation by projection distance minimization, IEEE Transactions of Medical Imaging, 28(5), pp.676-686.

Conners, R.W., 1980, A theoretical comparison of texture algorithms, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-2, pp. 204–22.

Caelli T. M., 1988, An adaptive computational model for texture segmentation, *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-18, pp. 9–17.

Haralick R. M., K. Shunmugam, and I. Dinstein, 1973, Textural features for image classification, *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-3, pp. 610–621.

Haralick, R.M., 1979, Statistical and Structural Approaches to Texture, *Proceedings of the IEEE*, 67, pp. 786-804.

Hossein Mobahi, Shankar Rao, Allen Yang, Shankar Sastry and Yi May, 2011, Segmentation of natural images by texture and boundary compression, 95(1), pp.86-98.
Justin Ker, Lipo Wang, Jai Rao, and Tchoyoson Lim , 2017, Deep Learning Applications in Medical Image Analysis, IEEE. Translations ,2169-3536 (c)

MATLAB, 2018, Computer Vision and Image processing toolbox, R2018b.

Nikhil Buduma and Nicholas Lacascio, 2017, Fundamentals of Deep Learning Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, CA 95472.

Pham Dzung L, Xu Chenyang, Prince Jerry, 2000, Current methods in medical image segmentation, Annual Review of Biomedical Engineering, 2, pp.315-337.

Satyanarayana Y. and Nitheesh Thomas , 2014 Segmentation and classification of shallow subbottom acoustic data, using image processing and neural networks, Marine Geophysical Research, 35:149-156.

Sonka M., Hlavac V., and Boyley R., 1993, Image processing analysis, and machine vision. Chapman Hall Computing, London.

Suzuki, K., 2017, Overview of deep learning in medical imaging, Radio- logical Physics and Technology, vol. 10, no. 3, pp. 257–273.

Tajbakhsh, N., J. Y. Shin, S. R. Gurudu, R. T. Hurst, C. B. Kendall, M. B. Gotway, and J. Liang, 2016, Convolutional neural networks for medical image analysis: Full training or fine tuning?. IEEE transactions on medical imaging, vol. 35, no. 5, pp. 1299–1312.

Weszka J. S., C. R. Dyer, and A. Rosenfeld, 1976, A comparative study of texture measures for terrain classification, *IEEE Trans. Syst., Man, Cybern.*, vol. SMC-6, pp. 269–285.

Acknowledgement

The author would like to thank Director, CHESS for encouraging and permission to present the work.