Accident Scene Image Identification Technique using Convolutional Neural Network

Omogbhemhe M.I.
Department of Computer Science
Ambrose Alli University, Ekpoma

Odegua R.O.
Department of Computer Science
Ambrose Alli University, Ekpoma

ABSTRACT
Building intelligent software that can effectively detect accident scene with the help of Google map has suffered set back because of the poor ability of the currently used software to effectively detect, identify and classify accident scene images from non accident scene images. Hence there is need for a better technique of implementing this software. In this paper, Convolutional neural networks (CNN) which is a part of deep learning algorithm was used to provide a better classification technique that any software to be developed for the purpose of detecting accident scene image can adopt. The algorithm was tested on 4000 accident scene images with other kind of images (cats and dogs) by adopting models of other researchers. In this paper, classification accuracy and Mean Squared Error (MSE) were used to evaluate the algorithm in identifying accident scene images accurately. The result was further presented using a graph of MSE against a number of trained epochs. The result of the experiment shows accuracy in the image classification and identification.

Keywords
Image Identification, Accident Scene, Convolutional Neural Network, Classification, Algorithm

1. INTRODUCTION
According to [1], image processing is the manipulation of digital images with the use of computer. Manipulating images is broad and complex mathematically analyses are used. Image manipulation, classification, fusion, restoration and computation are some of the main operations that must be done while processing an image if accurate result is needed. Classification of images forms an important aspect of image processing that software that must process image should prioritize. [1] noted that the main objective of image classification is the automatic allocation of image to thematic classes which will aid easy processes and manipulation. In image classification there are supervised classification and unsupervised classification. It is therefore important for any image processing system to adopt any one that best meet its objective. There are two steps for image classification; training of the system followed by testing. In training process, we take the properties of the images (form a class) and form a unique description for a particular class that other same properties will adopt. The classification process is done for all classes depending on the type of problem been classified which can be binary-class classification or multi-class classification. The categorization of the test images under various classes for which system was trained is the testing step. Based on the training properties, assigning class is done using the classes' partitions. [2] noted that deep is a new area of machine learning research. Deap learning is an area of machine learning that provide good processing power for signals and images. It is a class of machine learning techniques that dig deep into many layers of non-linear information processing for supervised or unsupervised feature that is extracted and transformed to forming pattern analysis and classification. In light of the advantages of using these tools for successful classification, using for classifying accident images will be benefit to building good software that can identify accident scene. This paper therefore adopts Convolutional Neural Network (CNN) for image classification for processing. The image data used for testing the algorithm includes images of accident scene, cats and dogs.

2. THEORITICAL BACKGROUND
The idea behind Computational models of neural networks has long been existed. McCulloch and Pitts were the first researchers that proposed these models [3]. It is a well known fact that neural networks comprises of number of layers with each layer connected to the other layers forming the network in form of human neurons. [4] describes feed-forward neural network (FFNN) as a term of neural activation and the strength of the connections between each pair of neurons. The input layer is the start place in the FFNN while the output layer is its stop place and the neurons are implemented with a direct connection. The hidden layer mediates between the input and the output layer and performs the computations that produce the needed output. Through the adjustment of weights learning occurs and the objective is to minimize error between the output generated through the output layer and the input given to the input layer of the model. Adjustment of the weights is done by process of back propagation (in which the partial derivative of the error with respect to last layer of weights is calculated). Through a recursive arrangement, there is repetition in the weight adjustment till the weight and input layers connection is updated. Artificial neural network develops techniques in software (machine learning using a programming language like python) and hardware in the manner nervous system process information for taking the best decision by human.

Artificial Neural Networks learn new tasks in generating solution to problems using the human brain behavior. Neurons are actually the set of elements that assist the brain in good decision making. Each neuron performs its independent task for the benefit and overall objective of the brain in making accurate decision. There exist a very large number of neurons, however, they possess the ability to coordinate themselves for the advantage of taking the best decision by the brain. The communication abilities posses by neuron empowers the brain to coordinate the neuron easily and apply such in making best decision. The different variants of the multi layer perceptron imbibed in CNN are the filters which are localized and used in the input layer [5].

In the work of [6], it was noted that Convolutional neural networks are designed to process two-dimensional (2-D) image. [7] describes the CNN architecture used in this paper. The network is divided into three layers which are

convolution layer, sub sampling layer and output layer. Convolutional neural networks are special purpose neural network applied in processing data in form of a grid-like topology in its representation, some of which are time-series data. Convolutional networks have been highly successful in practical applications and implementation. By the name "convolutional neural network" describe a network that uses mathematical computation called convolution and involves special kind of linear computation. Convolutional networks can also be seen as replacing the general matrix multiplication with the convolution in at least one or two of the network layers.

3. CONVOLUTION LAYER

The first layer of the CNN network is the convolution layer. The structure of this layer as adopted from [8] is shown in the figure 1 in this paper. It consists of a convolution mask, bias terms and a function expression. Together, these generate output of the layer. Figure 1 below shows a 5x5 mask that perform convolution over a 32x32 input feature map. The resultant output is a 28x28 matrix. Then bias is added and sigmoid function is applied on the matrix [8].
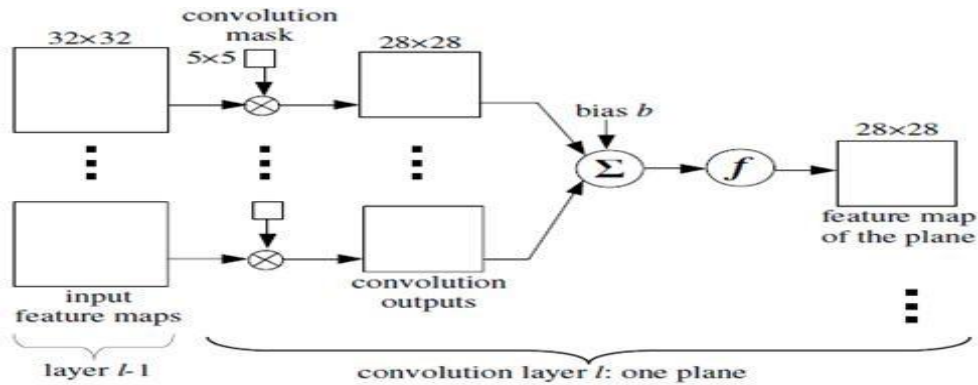

Figure 1: Convolutional layer working. Source: [8]

4. METHODOLOGY

It is important to state clearing that the used algorithm was adopted from the work of [8] but been used for accident image classification. In this algorithm, when the input is given as an image during the data set preparation they converted to gray scale as data information and not color information. Also, the images are resized to 150x150. The image pyramid is a data structure designed to support efficient scaled convolution through reduced image representation. The image pyramid consists copies of the original image in which both sample density and resolution are decreased in regular steps [9]. Similarly there is need to train the algorithm on the data used. Training of the algorithm as stated in [7] is to minimized errors between the network output and the desired output. The error function adopted from [8] is as defined by the equation below and is same for weights as well as bias terms.

$$E(w) = \frac{1}{K \times N} \sum_{k=1}^{K} \sum_{n=1}^{N_L} (y_n^k - d_n^k)^2$$

Here Y_n^k is the actual output of the network, K is the number of input image and the output vectors desired. X^k is the Kth training image and d^k corresponding to the desired output vector. The error gradient is computed through error sensitivities, which are defined as the partial derivatives of the error function with respect to the weighted sum input to a neuron. Once the error gradient E(t) is derived, numerous optimization algorithms for minimizing the energy function can be applied to train the network. Here, Root Mean Square Propagation (RMSPROP) is used as adopted from [8].

5. RESULTS

The result aspect of this paper shows the classification accuracy generated when the CNN is used on the dataset (accident scene images, dog images and cat images)

considered. Classification accuracy and Mean Squared Error (MSE) are used to present the output of the algorithm for the data considered. The changes made on the MSE compared to the training epochs were depicted in the graph. Worthy of note is that the MSE metric are considered to be the best and easy to use quality metric. It calculates the mean of squared difference between original and trained approximation. To generate lower MSE with original image, the image has to undergo better training. As the value for Mean Squared Error (MSE) tends to decrease, the variation in the final reconstructed output and the original image is very less. MSE indicates the close proximity between underlying true image and the final reconstructed output. The idea here is to use enough number of epochs that would result in low MSE, high classification accuracy and with least duration for training the network. The network is tested on accident scene dataset and for different number of iterations (epochs).

6. ACCIDENT SCENE IMAGE CLASSIFICATION

The data used in this experiment were also taken from the kaggle platform. Kaggle is an open source data science platform that contains numerous datasets for machine learning. The dataset consist of 4000 images divided into four classes (Sparse traffic, Dense Traffic, Accident and Vehicle on Fire). This dataset is further divided into 900 classes for each category. This makes the problem a multiclass problem and as such, we use a multimodal output with our CNN. Some sample images are shown in figure 2.
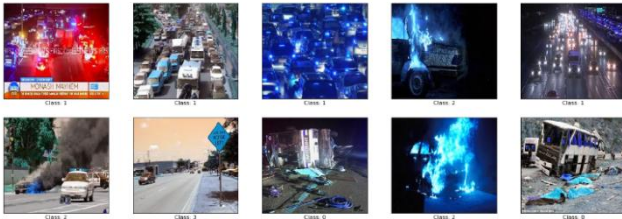
Figure 2: Sample images from data set

This is a 4 class image dataset. There are 900 images for each class. Each image measures 400x400 pixels. The images were manually extracted from large images of road scenes around the world. The pixel resolution of this public domain imagery is about 1 foot. The classification is done between two sets of data at a time. And the results were evaluated using classification accuracy for train and test data as well as MSE curves for different epochs used.
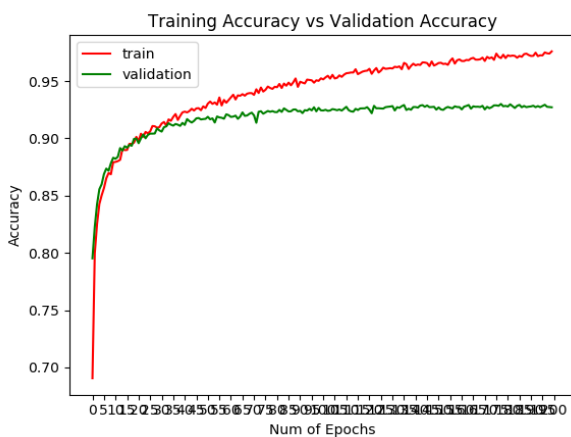
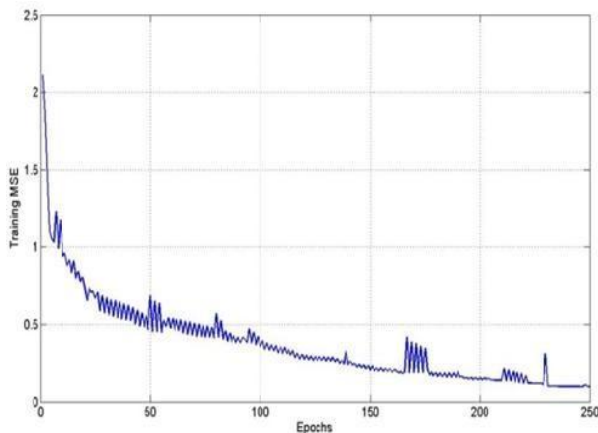

Figure 3: Accuracy Plot for Training

Figure 4: MSE Error Plot for Training

7. CONCLUSION

Results observed in the comparative study with other traditional methods suggest that CNN gives better accuracy and boosts the performance of the system due to unique features like shared weights and local connectivity. CNN is better than other deep learning methods in applications pertaining to computer vision and natural language processing because it mitigates most of the traditional problems. This paper gave a brief overview of Convolutional Neural Network as a sub class of neural networks. The paper explained how the CNN algorithm works and tested it in a four class classification problems to detect and classify accident scenes. We displayed error and accuracy metrics after training our algorithms for specific number of iterations.

8. REFERENCES

[1] Lillesand, T.M. and Kiefer, R.W. and Chipman, J.W.(2004). in "Remote Sensing and Image Interpretation" 5th ed. Wiley, 2004

[2] Li Deng and Dong Yu (2014). "Deep Learning: methods and applications" by Microsoft research [Online] available at: http://research.microsoft.com/pubs/209355/NOW-Book Revised- Feb2014-online.pdf

[3] McCulloch, Warren; Walter Pitts"A Logical Calculus of Ideas Im- manent in Nervous Activity", Bulletin of Mathematical Biophysics 5 (4): 115–133(1943)

[4] An introduction to convolutional neural networks [Online]availableat:http://white.stanford.edu/teach/index. php/An_Introduction_to _Convolutional_Neural_Networks

[5] Hubel, D. and Wiesel, T. (1968). Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology (Lon- don), 195, 215–243C. J. Kaufman, Rocky Mountain Research Labora- tories, Boulder, Colo., personal communication, 1992. (Personal communication)

[6] Yann LeCun, Leon Bottou, Yodhua Bengio and Patrick Haffner, "Gradient -Based Learning Applied to Document Recognition", Proc. Of IEEE, November 1998.

[7] S. L. Phung and A. Bouzerdoum,"MATLAB library for convolutional neural network," Technical Report, ICT Research Institute, Visual and Audio Signal Processing Laboratory, University of Wollongong. Available at: http://www.uow.edu.au/~phung

[8] Deepika J, Sowmya.V and Soman K.P (2015). Image Classification Using Convolutional Neural Networks. International Journal of Advancements in Research & Technology, Volume 3, Issue 6

[9] Adelson, Edward H., Charles H. Anderson, James R. Bergen, Peter J. Burt, and Joan M. Ogden. "Pyramid methods in image processing." RCA engineer 29, no. 6 (1984): 33-41.