OPEN ACCESS INTERNATIONAL JOURNAL OF SCIENCE & ENGINEERING

CLASSIFICATION ACCURACY IMPROVE USING DNN (DEEP NEURAL NETWORK)

Uma Malusare [1] Prof. Dr. B. K. Sarkar[2]

Department of Computer Engineering Padmbhushan Vasantdada Patil Institute of Technology, Bavdhan Pune, Maharashtra

uma1990.malusare@gmail.com, *dr.bksarkar2003@gmail.com*

Abstract: Deep Neural Networks (DNNs) show excellent performance in complex machine learning tasks such as image classification and speech recognition. However, due to their multilayered non-linear structure, they are not transparent, that is, they can be classified into certain categories, which are given new, invisible data samples. To solve this problem the author proposes an RNN (Recurrent Neural Network) algorithm for classifying the trained data given by the CNN (Convolution Neural Network) algorithm. The system uses SUN 397 dataset, heatmap method, and de-convolution method. In system first, perform heat mapping method on data. Heatmap method extracts the regions of the particular image. Then the de-convolution method is a propagation method for calculating the gradient of a function. LRP (layer-wise relevance propagation) algorithm learns the features of an image layer by layer from the input vector of convolution. After that classification performed on the dataset. Hybrid Network model shows how we will use CNN-RNN (LSTM) to the learned dataset. So this way get the required data from CNN local space while training image dataset of the image and provide it to RNN algorithms part Long Short-Term Memory (LSTM) to provide a prediction about what NN has learned in CNN model. Finally, the result shows that the proposed RNN algorithm is more accurate than CNN algorithm.

Keywords: Deep Neural Networks, RNN (Recurrent Neural Network) algorithm, CNN (Convolution Neural Network) algorithm, Heat map method, LRP (layer-wise relevance propagation) algorithm.

---∴∴∴---

I INTRODUCTION

Recently, the deep neural network (DNN) has emerged as a selection method for perceptual tasks such as speech recognition and image classification. A deep neural network (DNN) used as a classifier for pixels. The network uses the intensity of the image in a square window centered on the pixel itself as input to calculate the probability that the Pixel is a membrane. The image is divided by the classification of its pixels [1]. DNN is formed in separate cells with characteristics similar to those that the membrane has been manually annotated. DNN training methods are improving DNNs. DNN is generally considered as a black box method, and users may think this lack of transparency is actually a disadvantage. That is, for each new input data point, what trained DNN model to reach a specific response, intuitive and quantitatively understand the results of DNN inference, it should be noted that this aspect and different feature selection questions: is a feature on average prominent ensemble training data? [1] [2].

The system proposes an RNN algorithm for classifying the trained data given by the CNN algorithm. The system uses SUN 397 dataset. In system first, perform heat mapping method on data. Heatmap method extracts the regions of the particular image. Then the de-convolution method is a propagation method for calculating the gradient of a function. LRP algorithm learns the features of an image layer by layer from the input vector of convolution. After that classification performed on the dataset. Hybrid Network model shows how we will use CNN-RNN (LSTM) to the learned dataset. So this way can get the required data from CNN local space while training image dataset of the image and provide it to RNN algorithms part (LSTM) to provide a prediction about what NN has learned in CNN model.

In this paper study about the related work done, in section II, the proposed approach modules description, mathematical modeling, algorithm and experimental setup in section III .and at final we provide a conclusion in section IV.

II LITERATURE REVIEW

The performance of deep neural networks (DNN) is well in complex machine learning tasks like image classification and speech recognition. Due to the DNN multilayered nonlinear structures, it's not transparent, i.e., given new unseen data samples specific classification or recognition recently, for one test image, various methods are developed to understand the reasoning embodied in DNN. A General method for estimating an ordered set of pixels, such as a heat map, based on perturbations of the region is proposed. Compare the Heat Map data set calculated by the SUN397, ILSVRC2012, and MIT three different methods. The results are explained and their practical implications are discussed [1].

The author addresses the central issue of neuro-anatomical, i.e., automated segmentation of neuronal structures drawn in a stack of electron microscopy (EM) images. This is necessary for efficient mapping of 3D brain structure and connectivity. The segment {\em biological} neuron membrane uses a special type of deep {\em artificial} neural network as a Pixel classifier. Label each pixel (the film is non-film) to predict from the raw pixel value of the angled window to the center. The class classifier is trained by the plain gradient descent on the $512 \times 512 \times 30 $ stack with the known ground truth [2].

The author develops new 3D CNN model for action recognition. From both spatial and temporal dimensions, the developed model accesses features, by performing 3D convolution. This model is developed in many adjacent frames. That frames generates multiple information channels from the input frames, and the final feature representation combines information from all channels. To further improve performance, they propose to normalize the output with high-level features and combine various predictions of different models. It applies a model developed to recognize human behavior in the real environment of the airport surveillance video and achieves superior performance compared to the baseline method [3].

For the last 10 years, nonlinear kernel-based learning has been widely used in science and industry for the solution of ranking problems, classification, regression, etc. Their users are satisfied with the performance of this powerful technology, but opening a non-linear black box is a difficult task, though it requires a deeper understanding of both the learning machine and the problem of data analysis to be solved. In particular, the author evaluates the complexity and noise structure underlying the learning problem and discusses the related dimension estimation (RDE), which makes it possible to distinguish high-low complexity high-low noise scenarios, respectively. In addition, a new local method based on RDE for quantifying the reliability of the learning prediction is introduced. Finally, this paper reports a technique which can explain individual nonlinear prediction. This method not only helps to gain more knowledge about the nonlinear signal processing problem itself but also extends the general usefulness of kernel methods in practical signal processing applications [4].

Deep architecture demonstrates state-of-the-art performance with a variety of settings, particularly using a vision dataset. The deep learning algorithm is based on learning several levels of expression of input. In addition to the performance of the test set, you need to prepare for a solution that has been learned by different deep architectures. One of the objectives of the study is to improve the tool to find a qualitative interpret at ion with a high level of functioning by such a model. It is hoped that the losses will be turned into varies because they were deep in the network. This contrast compares the number of techniques to interpret such problems. Using the technology of stacking the noise removal of the automatic encoders and deep belief networks, they have nurtured a plurality of vision datasets. This consistent filter-like interpretation makes it possible to easily accomplish the unit level. The developed tool allows analyzing deeper models of more depth and achieving traces of variance manifolds for each of the hidden units. This kind of technology allows researchers to deeply understand the architecture of how deep architectural works [5].

In this paper, the author proposes to use unsupervised feature learning as a method for learning features directly from video data. The author extends an independent subspace analysis algorithm for learning invariant Space-Time features from unlabeled video data. They have found that, despite its simplicity, this method can be used to learn hierarchical expressions, such as

Hollywoodd2, Hollywood, UCF, KTH, Hollywood, and Hollywood by substituting pre-learned features with hand-designed features combined with deep learning techniques such as stacking and convolution [6].

III PROPOSED APPROACH

Problem Statement

Deep neural networks (DNNs) are used in image classification and speech recognition. But because of the DNN multilayered non-linear structure, they are not transparent, that is, they can be classified into certain categories, which are given new, invisible data samples. So the system proposes an RNN algorithm for classifying the trained data.

Proposed System Overview

Figure 1 shows, the detailed description of the proposed system.

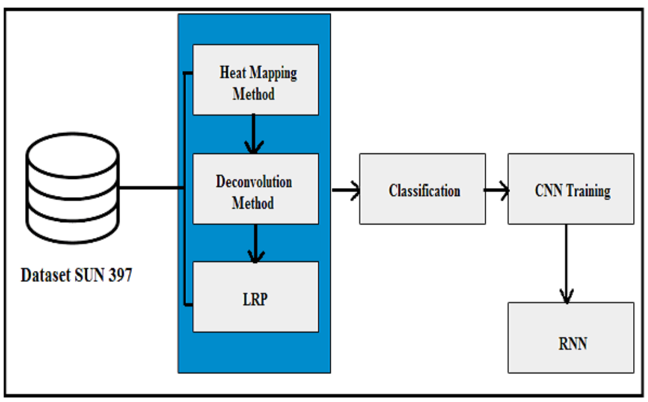


Figure 1. Proposed System Architecture

1) The system uses dataset SUN 397.

2) Heatmap Method [1]: After that, the system performs a heat mapping method to extract features for learning. Achieve transparency in which heatmap is made interesting and intuitive tool.

3) De-convolution Method [1] [2]: The system uses the de-convolution method. This is similar to the back propagation method for calculating the gradient of a function, De-convolution approaches the idea of the back propagation rule using the network's output activation from pixel space to map this way max-pooling and ReLUs with a convolutional net designed for other types of architectures.

4) LRP algorithm learns the features of an image layer by layer from the input vector of convolution. LRP decomposes the classification output, into pixels precisely by observing the principle of conservation for each layer-wise relevance that is it is not lost for evidence or categories. The algorithm does not use a gradient. For this purpose, the applicable general-purpose architecture (including the noncontiguous unit) is set. LRP discusses classification decisions and heat map scores worldwide have a clear interpretation as evidence against the category.

5) Then the system performs classification to classify all data into respective classes.

6) Then using CNN train all dataset. It is a class of DNN; it's commonly used for analyzing visual imagery. The term convolution refers to the mathematical combination of two functions to produce a third function. Merging the two pieces of information in the case of CNN, convolution is performed on the input data using a filter or kernel (these terms are used interchangeably), and a feature map is generated.

7) After that the system, the training data provides an input vector to RNN to know what CNN has learned.

Algorithms

- Algorithm 1: CNN Algorithm [1].
  1) Input it (32x32x3) will hold the raw pixel values of the image.
  2) The CONV layer computes the dot product between each weight and the small area they are connected to the input volume, which is connected to the local area of the input, which could be a volume such as (32x32x12) if you decide to use 12 filters.
  3) The RELU layer will apply the activation element-wise function, such as the maximum threshold (0, x) at zero. The volume size remains unchanged (32x32x12).
  4) A down sampling performed by the pool layers operation along the spatial dimension (width, height) and the volume is (16x16x12).
  5) Fully connected (FC) layer, it computes the class scores, resulting in a volume of size (1x1x10).
- Algorithm 2: RNN (Recurrent Neural Network) Algorithm [Proposed Algorithm]

  Step1: Convolution_network(Input image)

  - Breaks the image into small windows of image
  - start at the left at the top and takes width and height of part
  - It does this all the way across the image and outputs a new image.

  Step2: max pooling

  - It performs operations on small Windows

- The operation runs on this small window is usually (average, max, or min) to combine the small window into a single pixel.

Step3: LSTM (long short term memory)

- Vsnishing gradient problem.
- store and remember things that have happened in the past and help to find patterns across time to make its next guesses make sense.

Step4: Gated Recurrent Units (GRUs)

A gated recurrent unit (GRU) fully writes the contents from its memory cell to the larger net at each time step.

Mathematical Model

I = {Image Dataset}

1) P1= {I}

Read Datasets.

2) P2= {P1}

- Heat mapping

$$h^{(l)} = m_{dec}(h^{(l+1);\theta^{l,l+1}})$$

(1)

3) P3= {P2}

- De-convolution Heatmap

$$R^{(l)} = m_{dec}(R^{l+1}; \theta^{l,l+1})$$

(2)

Where,

R(l), R(l+1) denote the backward signal.

I (l, l+1) is the set of parameters connecting two layers of neurons.

4) P4 = {P3}

- LRP algorithm

$$h^{(l)} = m_{lrp}\left(h^{(l+1);\theta^{l,l+1}}\right)$$

(3)

$$\sum_i h_i^l = \sum_j h_j(l+1) \quad (4)$$

5) P5 = {P4}

- C 4.5

$$Info(S) = -\sum_{i=1}^{k}\{[freq(C_i,S)/|S|]\log_2[freq(C_i,S)/|S|]\}$$

(5)

Where,

|S|= is the number of cases in the training set.

Ci= is a class, i=1,2,…k,

K= Number of cases.

Freq (Ci=S) and is the number of cases in Ci.

6) P6 = {P5}

- CNN

By using CNN train all dataset.

7) P7 = {P6}

- RNN

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h)$$

(6)

$$y_t = \sigma_y(W_y h_t + b_y)$$

(7)

Where,

x_t = input vector

h_t=hidden layer vector

y_t=output vector

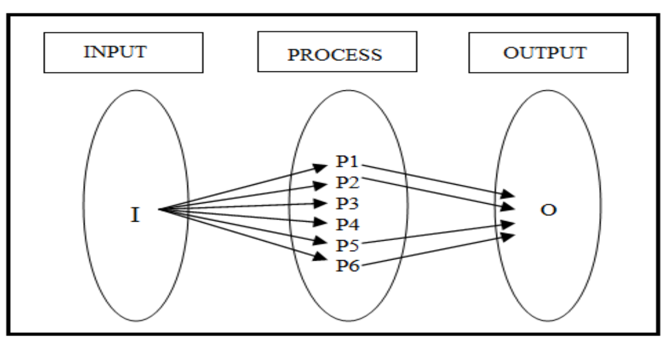
W, U and b: parameter matrices and vector

σ_h and σ_y : Activation function

Output: Output depends on P6 and P7.

O= {Prediction and classification}

VENN DIAGRAM:



IV. RESULTS AND DISCUSSION

A. Experimental Setup

The system is built using Java framework on Windows platform. The Net bean IDE is used as a development tool. The system doesn't require any specific hardware to run; any standard machine is capable of running the application.

B. Expected Result and Discussion

In this section discussed the experimental result of the proposed system.

In this system, we use existing heatmap method, LRP algorithm and de-convolution method and show the results based on SUN 397 dataset. These systems propose an RNN (Recurrent Neural Network) algorithm for classifying the trained data given by the CNN (Convolution Neural Network) algorithm and prove the accuracy of CNN verses RNN+CNN algorithm. Then the system performs classification to classify all data into respective classes. Then by using CNN train all dataset. Then we show the accuracy comparison between CNN and proposed RNN+CNN algorithm. The accuracy is calculated as;

$$h_t = \sigma_h(W_h x_t + U_h h_{t-1} + b_h)$$

(9)

$$y_t = \sigma_y(W_y h_t + b_y)$$

(10)

$Where, y_t$ is an output vector, W, U, and b are parameter matrices and vector, σ_h and σ_y are activation function. Table I shows, the accuracy comparison between the existing and proposed system algorithm. Figure 2 shows, accuracy comparison between the CNN and RNN+CNN algorithms, it shows that the RNN algorithm more accurately classifies the images than CNN algorithm. The RNN uses LSTM (Long Short-Term Memory), its hold information for a much longer period. This information used for future use, when new data are come it compare with the holed information and then it gets the result uniquely and more accurately. So finally conclude that the proposed RNN+CNN algorithm is to perform more accurate and efficient than CNN algorithm.

TABLE I. CNN Vs RNN+CNN ACCURACY COMPARISON

Algorithm	**Accuracy in (%)**
CNN	90
RNN+CNN	93

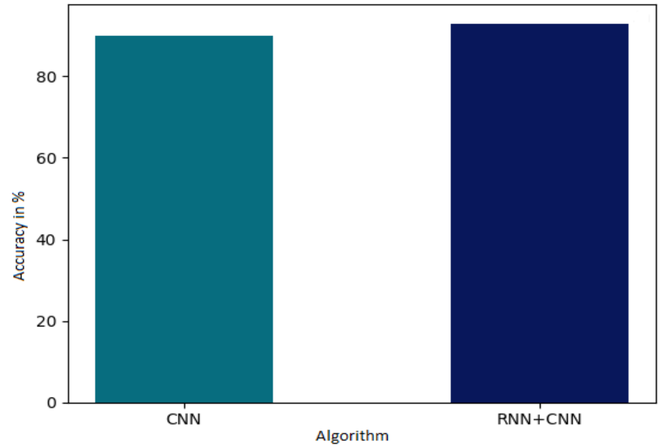


Figure 2: CNN Vs RNN+CNN Accuracy graph

V.CONCLUSION

DNN research has traditionally focused on improving the speed of quality, algorithms, or neural network models. The system uses SUN 397 dataset. System proposes an RNN+CNN algorithm for classifying the trained data given by the CNN algorithm. In this system CNN algorithm is used to training the whole dataset so at the time of training dataset we store the values in local space, this local space gives as input to our proposed RNN algorithm to provide what deep Neural Network has learned. Finally, the result shows that the proposed RNN+CNN algorithm is more accurate than CNN algorithm.

REFERENCES

[1] gregoire montavon sebastian lapuschkin, alexander binder, wojciech samek, and klaus-robert müller, "evaluating the visualization of what a deep neural network has learned", (2016).

[2] l. m. gambardella, a. giusti, j. schmidhuber and d. ciresan, "deep neural networks segment neuronal membranes in electron microscopy images", (2012).

[3] m. yang, w. xu, s. ji, and k. yu, "3d convolutional neural networks for human action recognition", (2010)

[4] t. krueger, m. l. braun, g. montavon, and k.-r muller, "analyzing local structure in kernel-based learning: explanation, complexity, and reliability assessment," (2013)

[5] a. courville, d. erhan, and y. bengio, "understanding representations learned in deep architectures", (2010)

[6] s. y. yeung, w. y. zou, q. v. le, and a. y. ng, "learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis", (2011).

[7] alex krizhevsky, geoffrey e. hinton and ilya sutskever, “recursive deep models for semantic compositionality over a sentiment treebank”, (2013)
[8] l. bottou, k. kavukcuoglu, r. collobert, m. karlen, p. kuksa and j. weston, “natural language processing (almost) from scratch”, (2011).
[9] m. l. braun, g. montavon and k.r. muller, “kernel analysis of deep networks”, (2011)