A Review on Image Dataset Evaluation by Boltzmann Machine under Deep Neural Networks

Monika Mann
M.Tech Scholar
Computer Science & Engg. Department
R. P. Inderaprastha Institute of Technology, Kurukshetra

Parmeet Kaur
Asst. Professor
Computer Science & Engg. Department
R. P. Inderaprastha Institute of Technology, Kurukshetra

***Abstract*- Image description is a good starting point for imparting artificial intelligence to machines by allowing them to analyze and describe complex visual scenes. This work introduces a generic end-to-end trainable Boltzmann architecture to address multi-modal applications. Deep Neural Networks allow each input modality to be independent in terms of architecture, parameters and length of input sequences. This work presents a review on performance evaluation of image dataset using Boltzmann under Deep Neural Networks. The performance will be evaluated in terms of accuracy. All Simulations will be implemented in MATLAB.**

***Keywords*-Deep Neural Networks, MATLAB, Image Processing etc .**

I. INTRODUCTION

Machine learning is a subfield of computer science. At its core, it is the foundation for a set of statistical tools that estimate complicated functions by learning from data. Machine learning can be divided into two main approaches, supervised and unsupervised learning. Supervised learning generally means that the program is given both input and the desired output, for example, pictures of objects with corresponding labels of what is depicted. The goal of the learning (or training) is to construct a map between those two. In contrast to supervised learning, the unsupervised learning approach does not provide the program with the correct output. Here, the goal of training is to find structure inside the given input; it is used, for example, in auto encoders.

Accurate text annotation of image and video content enables more efficient search and retrieval, can aid visual understanding in medical, security, and military applications, and can even be used to describe pictorial content to the visually impaired. Uncertainties about salient content, main subject detection, object recognition, action detection, and scene understanding make this a challenging problem. Despite the difficult nature of this task, computer vision and natural language processing researchers have made significant strides in this area. This work builds upon these previous successes, and introduces a new framework that simultaneously addresses the diverse input modalities, while producing state-of-the-art results.

In the past five years, supervised convolutional models have forever changed the computer vision and machine learning landscape. Due to the recent introduction of large supervised datasets [1] and accelerated training models using Graphic Processing Units (GPUs), the traditional pairing of hand crafted low level vision features with complimentary classifiers has been tested by Convolutional Neural Networks (CNNs). CNNs are deep feed forward networks based upon a hierarchy of abstract layers which simultaneously learn the low level features and classifier. These networks have been shown to equal the performance of neurons in the primate inferior temporal cortex, even under difficult conditions such as pose, scale, and occlusions [2]. CNNs have won competitions in traffic sign detection, house number detection, handwriting recognition, pedestrian detection, object recognition, speech recognition, breast cancer detection, and many more.

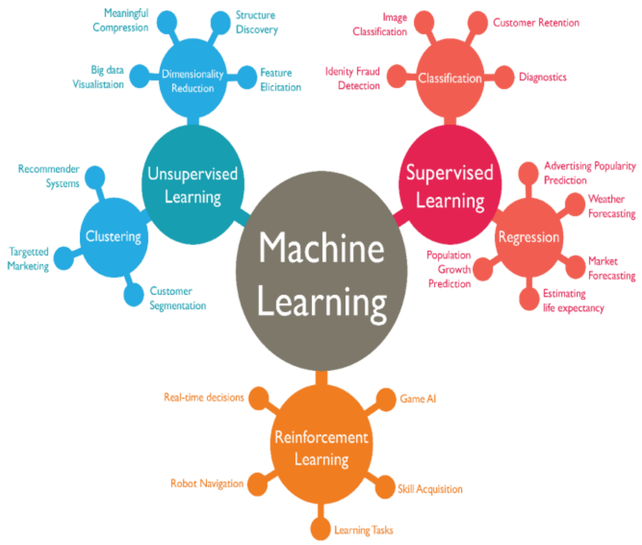

Fig 1: Machine Learning [1]

The human brain does not interpret an image by pixel but it decomposes a problem into sub-problems through multiple levels of interpretations. As shown in [1], the human brains processes visual signals through a structure of multiple layers, well represented by Neural Networks. One of the promises of deep learning is replacing handcrafted features with unsupervised or semi-supervised feature learning and hierarchical feature extraction. Research in this area attempts to make better representations and create models to learn these representations from large-scale data. One of the most striking facts about neural networks is that they can compute any function at all. No matter what the function, it is guaranteed there is a neural network so that for every possible input x, the value f(x) or some close approximation is output from the network.

The paper is ordered as follows. In section II, we discuss correlated work with deep network system. In Section III, It defines models of deep learning system. Finally, conclusion is explained in Section IV.

II. LITERATURE REVIEW

Naylor P. et al. [2017] [7] presented a fully automated workflow to segment nuclei from histopathology image data by using deep neural networks trained from a set of manually annotated images and by processing the posterior probability maps in order to split jointly segmented nuclei. Further, we provide the image data set that has been generated for this study as a benchmark set to the scientific community.

Haozhe Jia et al. [2017] [8] used registration-based coarse segmentation on preprocessed prostate MR images to define the potential boundary region. It then trained four DCNNs as voxel-based classifiers and classify the voxel in the potential region is a prostate voxel when at least three DCNNs made that decision. Finally, it used boundary refinement to eliminate the outliers and smooth the boundary. It evaluated this approach on the MICCAI PROMIS12 challenge dataset and experimental results verify the effectiveness of the proposed algorithms.

Huang D. et al. [2017] [9] proposed a single image dehazing based on deep neural network that is to deal with haze image. In this work, it built up a deep neural network to restore the hazy image. It tested our method both objective and subjective and compare with classical method for dehazing. Our test shows that our method works better than the others in reducing Halo effect and also our method does well in restore colorful of input image. Finally, our method process faster.

Yan S. et al. [2018] [10] developed a new network configuration, Dense-Unet, to achieve optimal performance with low computational cost. Results after the calcium removal process were validated by comparing with gold-standard X-ray angiography. The results demonstrated that removing coronary calcification from images with the proposed approach was feasible, and may potentially improve the diagnostic accuracy of CTA.

Shao C. et al. [2018] [11] adopted a deep neural network image encoder to encode the given input image and generates a compact representation of the image. The representation can be as short as 10 bytes. On the receiver end, it adopted a deep neural network decoder running on a mobile device. When the mobile device receives the BLE broadcasted image data, it decodes the original image. It developed a pair of smart phone applications.

Schlemper J. et. al. [2018] [12] showed that when each 2-D image frame is reconstructed independently, the proposed method outperforms state-of-the-art 2-D compressed sensing approaches, such as dictionary learning-based MR image reconstruction, in terms of reconstruction error and reconstruction speed. Second, when reconstructing the frames of the sequences jointly, we demonstrate that CNNs can learn spatio-temporal correlations efficiently by combining convolution and data sharing approaches.

Choi J. et al. [2018] [13] presented the first deep neural network approach for QIS image reconstruction. Our deep neural network takes the binary bit stream of QIS as input, learns the nonlinear transformation and denoising simultaneously. Experimental results show that the proposed network produces significantly better reconstruction results compared to existing methods.

Chan C. et al. [2018] [14] proposed a feature oriented deep convolutional neural network for PET image denoising that uses weight maps in order to steer the training toward contrast preservation for small features. To obtain the weight maps, we first manually segment the lesions in the target images to create lesion masks. Lesion voxels are assigned stronger weights than the background voxels followed by a Gaussian smoothing.

Ajmal M. et. al. [2019] [15] proposed a weakly supervised approach for complex human activity recognition from realistic videos. The proposed approach required only activity labels for each video to train the model. A novel multilevel contextual features and context estimation procedure from the un-annotated dataset is also introduced. Restricted Boltzmann machine is used to systematically integrate multilevel contextual features. It evaluated the proposed approach on benchmark realistic surveillance video datasets for human and human object interaction activity recognition.

Agarwal T. et. al. [2019] [16] presented four popular convolutional neural network based models namely, VGG16, mobile Net, Resnet50 and InceptionV3 for the comparative study. The efficiency of the considered models is evaluated on various image classification datasets namely; cats and dogs for binary classification and plant seedling dataset for multiclass classification in terms of accuracy.

Kamış S. et al. [2019] [17] presented a comparison of different deep learning methods used for sentiment analysis in Twitter data. In this domain, deep learning (DL) techniques, which contribute at the same time to the solution of a wide range of problems, gained popularity among researchers. It evaluated and compare ensembles and combinations of CNN and a category of RNN the long short term memory (LSTM) networks.

III. MODELS OF DEEP LEARNING

1. Convolutional Neural Network (CNN)

CNN is one of the neural network models for deep learning, which can be described by three specific characteristics, namely locally connected neurons, shared weights and spatial or temporal sub-sampling. The architecture of a CNN is designed to take advantage of the 2D structure of an input image (or other 2D input such as a speech signal). This is achieved with local connections and tied weights followed by some form of pooling which results in translation invariant features. A benefit of CNNs is that they are easier to train and have many fewer parameters than fully connected networks with the same number of hidden units. To provide an overview on the architecture characteristics, CNNs are composed of convolutional and sub sampling layers optionally followed by fully connected layers. The input to a convolutional layer is an m x m x r image where m is the height and width of the image and r is the number of channels, e.g.an RGB image has r = 3. The convolutional layer will have k filters of size n x n x q where n is smaller than the dimension of the image and q can either be the same as the number of channels r or smaller and may vary for each kernel. The size of the filters gives rise to the locally connected structure where each element is convolved with the image to produce k feature maps of size mn + 1.

2. Residual Neural Network

A variation of standard CNNs are Deep Residual Neural Networks. A description and a model for ResNets can be found. ResNets take a standard feed-forward ConvNet and add skip connections that bypass (or shortcut) a few convolution layers at a time. Each bypass gives rise to a residual block in which the convolution layers predict a residual that is added to the following block's input. It proves that these networks can gain accuracy from considerably increased depth.

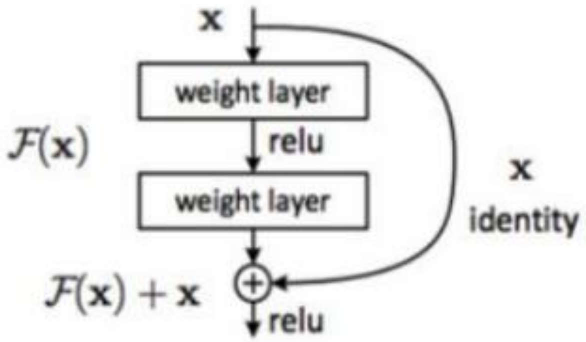

Fig 2: Building Block For Residual Learning [3]

Residual layers address a problem of degradation that occurs as the depth increases and this problem is not due to over fitting. The accuracy saturates first and then degrades rapidly. When extra-layers turn out to be unnecessary or even cause of degradation, an identity mapping between layers is the most effective solution. A residual connection allows to push the residual to zero and propagate along the stack an identity mapping, thus "hiding" the presence of extra (unnecessary) layers. To the extreme, if an identity mapping were optimal to a specific training propagation, it would be easier to push the residual to zero than to fit an identity mapping by a stack of nonlinear layers. Along with the depth degradation, the internal covariate shift issue has been considered with ResNets. A thorough explanation is given. The internal covariate shift is caused by the fact that the distribution of each layer's inputs changes during training along with the change in the network parameter's values.

3. Recurrent Neural Network

The idea behind RNNs is to make use of sequential information. In a traditional neural network we assume that all inputs (and outputs) are independent of each other. But for many tasks that may be a bad idea. To predict the next word in a sentence it is better to know which words came before it. RNNs are called recurrent because they perform the same task for every element of a sequence, with the output made dependent on the previous computations. Another way to think about RNNs is that they have a "memory" which captures information about what has been calculated so far.

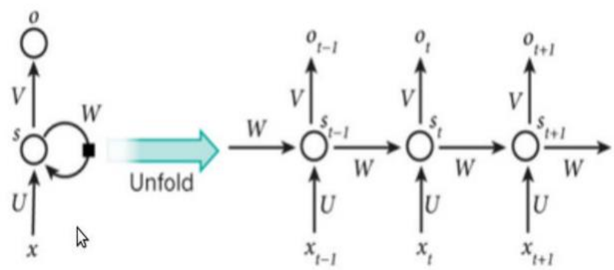

Fig 3: Basic RNN Structure [4]

x t is the input at time step t. For example, x 1 could be a vector corresponding to the second frame of a video. s t is the hidden state at time step t. It is the "memory" of the network. s t is calculated based on the previous hidden state and the input at the current step: st = f (U xt + W st−1) . The function f usually is a nonlinearity such as tanh or a rectifier. s t−1 , ..s t−n , which are required to calculate the first hidden state, are typically initialized to all zeroes. o t is the output at step t. For example, to make prediction on the next frame in a video it would be a vector of probabilities across all possible pixels (the descriptors of the frame) in use in the given system.

4. Long Short-Term Memory

In a traditional recurrent neural network, during gradient back propagation step, the gradient values can end up being multiplied a large number of times by the weight matrix associated with the connections between the neurons of the recurrent hidden layer. This means that the magnitude of weights in the transition matrix can have a strong impact on the learning process. If the weights in this matrix are small (or, more formally, if the leading eigen value of the weight matrix is smaller than 1.0), it can lead to a situation called vanishing gradients where the gradient signal becomes so small that learning either becomes very slow or stops working altogether. It can also make more difficult the task of learning long-term dependencies in the data. Conversely, if the weights in this matrix are large (or the leading eigen value of the weight matrix is larger than 1.0), this can lead to a situation where the gradient signal is so large that it can cause learning to diverge. This event is often referred to as exploding gradients.

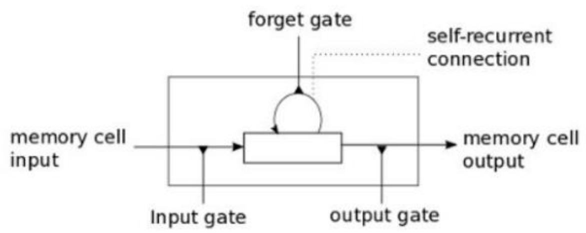

Fig 4: Basic RNN Structure [5]

This model makes use of a new structure called a memory cell. A memory cell is composed of four main elements: an input gate, an output gate, a neuron with a connection to itself and a forget gate. The self-recurrent connection has a weight of 1.0 and ensures that, barring any outside interference, the state of a memory cell can remain constant from one time step to another. The gates serve to modulate the interactions between the memory cell itself and the environment. The input gate can allow incoming signal to alter the state of the memory cell or block it. On the other hand, the output gate can allow the state of the memory cell to have an effect on other neurons or prevent it.

IV. PROBLEM FORMULATION

In the past five years, supervised convolutional models have forever changed the computer vision and machine learning landscape. Due to the recent introduction of large supervised datasets and accelerated training models using Graphic Processing Units, the traditional pairing of hand crafted low level vision features with complimentary classifiers has been bested by Convolutional Neural Networks. CNNs are deep feed forward networks based upon a hierarchy of abstract layers which simultaneously learn the low level features and classifier. These networks have been shown to equal the performance of neurons in the primate inferior temporal cortex, even under difficult conditions such as pose, scale, and occlusions. The Boltzmann models solved the gradient problem by replacing the traditional artificial neuron with a memory cell containing long and short term nonlinear capabilities. The model incredible power was first realized in the speech and natural language processing domains, and more recently to the annotation of image. This Model is a natural fit for temporal sequences of varying lengths and can be trained. Before the training and testing, it will be pre-processed the words by mapping them into the same vector space as the image feature vector extracted such that the dot product of a word vector with its corresponding image vector is maximized.

V. CONCLUSION

From Survey, VGG16 is the accurate model but it takes too much time and computational recourses which can be a problem when we have very limited computational power. Mobile Net, on the other hand, is not as accurate as VGG16 but in terms of resource vs accuracy trade-off, Mobile Net shows superior performance. Due to this, this work presents a review on performance evaluation of image dataset using Botlzmann Machine under deep learning methods. This method will improve the performance of system in terms of accuracy and reduce complexity.

REFFERENCES

[1] D. Linden and T. B. Reddy. (2002) *Handbook of Batteries. McGraw-Hill Professional: New York*.

[2] S. Davis. (2003) *Basics of Design: Battery Power Management*. Supplement to Electronic Design.

[3] K. Minakshi, (2003) "*Digital Image Processing: In: Satellite Remote Sensing and GIS Applications in Agricultural Meteorology*," World Meteorological Organization Publishing, pp. 81-102.

[4] R. Hartley and A. Zisserman. (2003) *Multiple View Geometry in Computer Vision. Cambridge* University Press, second edition.

[5] O. Oktay , E. Ferrante, (2017), " *Anatomically Constrained Neural Networks(ACNNs): Application to Cardiac Image Enhancement and Segmentation*", IEEE Transactions On Medical Imaging, Vol. 37, No. 2, pp. 384-395.

[6] E. Nishani, B. Çiço, (2017), " *Computer Vision Approaches based on Deep Learning and Neural Networks*", Mediterranean Conference on Embedded Computing, pp. 08-11.

[7] P. Naylor, M. Lae, (2017), " *Nuclei Segmentation In Histopathology Images Using Deep Neural Networks*", IEEE Access, pp. 933-936.

[8] H. Jia , Y. Xia, (2017), " *Prostate Segmentation In MR Images Using Ensemble Deep Convolutional Neural Networks*", IEEE Access, pp. 762-765.

[9] D. Huang, K. Chen, (2017), "*Single Image Dehazing Based on Deep Neural Network*", International Conference on Computer Network, Electronic and Automation, pp. 294-299.

[10] S. Yan, F. Shi, (2018), "Calcium Removal From Cardiac Ct Images Using Deep Convolutional Neural Network", IEEE 15th International Symposium on Biomedical Imaging, pp. 466-469.

[11] C. Shao, S. Nirjon, (2018), "*Demo Abstract: Image Storage and Broadcast over BLE with Deep Neural Network Autoencoding*", IEEE/ACM Third International Conference on Internet-of-Things Design and Implementation, pp. 302-303.

[12] J. Schlemper , J. Caballero, (2018), " *A Deep Cascade of Convolutional Neural Networks for Dynamic MR Image Reconstruction*", IEEE Transactions on Medical Imaging, Vol. 37, No. 2, pp. 491-503.

[13] J. Choi, O. Elgendy, (2018), "*Image Reconstruction For Quanta Image Sensors Using Deep Neural Networks*", IEEE Access, pp.6543-6547.

[14] Y. Chen, Y. Xie, Z. Zhou, (2018), "*Brain MRI Super Resolution Using 3d Deep Densely Connected Neural Networks*", IEEE 15th International Symposium on Biomedical Imaging, pp. 739-741.

[15] C. Chan, J. Zhou, (2018), "*Feature Oriented Deep Convolutional Neural Network for PET Image Denoising* ", IEEE Access, pp. 08-11.

[16] M. Ajmal, F. Ahmad, (2019), "*Recognizing Human Activities From Video Using*
Weakly Supervised Contextual Features", IEEE Access, pp. 98420-98435.

[17] T. Agarwal, H. Mittal, (2019), "*Performance Comparison Of Deep Neural Networks On Image Datasets*", IEEE, pp. 05-10.

[18] S. Kamış, D. Goularas, (2019), " *Evaluation of Deep Learning Techniques in Sentiment Analysis from Twitter Data*", International Conference on Deep Learning and Machine Learning in Emerging Applications, pp. 12-17.