Artificial Intelligence – State of Art Convolution Neural Network Architectures in a Nutshell

F. Catherine Tamilarasi, J. Shanmugam

Abstract*: *It is a well-known fact that all the Artificial Intelligence (AI)researches happening across multiple verticals such as Neuro Imaging, Computer Vision, Deep learning etc point to one master goal of modelling the human brain function by understanding how each part of the brain works. The Convolution neural network (CNN) is one of best deep architecture suitable to handle variety of inputs. In this paper we explore the different types of input data the CNN deep architecture can process and some of the CNN configuration changes that has proved good Accuracy. We have highlighted those specialized CNN architectures along with different types of data inputs they handle including the Functional Magnetic Resonance (fMRI) Neuro Image brain data input.

Keywords: **Artificial Intelligence, Deep Learning, CNN, Visual, Audio and Speech Recognition.**

I. INTRODUCTION

Mapping the human brain functions to a computer is otherwise called Brain decoding. To achieve this the learning capability of human brain need to be replicated by the AI system. The Machine learning and Deep learning algorithms are continuously aiming for such accurate human visual and auditory systems self-learning models. [1] Though we have many Deep learning architectures, the CNN being Multilayered and hierarchical has proven effective in modelling human Visual Cortex\auditory cortex and audio\video\image classification. CNN has proven capability to process below 7 types of input data. [2] In this paper we will discuss recently developed CNN special architectures that are enhanced to accommodate 7 different types of input data. [3]

1. Audio
2. Music
3. Video
4. Text
5. Speech, Language
6. cognitive inputs
7. Image – fMRI scan, the Brain Neuro Image data.

II. DNN – LARGE SCALE AUDIO INPUT

The Audio Set is a large audio dataset [8] with above 1 millionacoustic events. The Deep Neural Network (DNN) and its subarchitectures when compared on Large scale audio classification accuracy, they all performed well and highest performance shown in Inception V3 model next being ResNet Model.

- Fully connected
- Alexnet
- VGG
- Inception V3
- ResNet-50

III. CDNN- MUSIC INPUT

The convolutional Deep neural network is used in automaticmusic content analysis. From ‘million songs ‘dataset theCDNN is used for Music identification, genre classification,author identification, playlist generation etc advanced conceptsinclude music composing and computational musicology. [18]

IV. TASK OPTIMIZED CNN - NATURAL SOUND INPUT

One of the core objective of Deep learning AI systems is to develop human auditory system model that decide output task based on given input sound stimuli. In such systems the CNN predicts the neural activity from features learnt from input sound stimuli. In the novel Task optimized CNN model proposed in [5] the first layer input was a ‘cochleagram’ that is output of ‘cochlear model’. Unique synthetic sound task can be generated by the cochlear model for each input natural sound and same could be used to train the CNN. It has been proved that this type CNN with task optimized input best models the Primary Auditory cortex of human beings than other conventional CNN models. This model is further enhanced by using Dual –task input to CNN where after 4th Layer the CNN branches into 5 task specific layer. The classification accuracy of this model is compared with human behaviour and resultswere reasonably good.

V. CLDNN – SPEECH INPUT

Convolution LSTM DNN model is a combination of CNN, LSTM and Deep Neural network and it is a model specifically suitable for speech recognition related tasks. This model is proposed in [10] to overcome the individual limitations of existing speech models by bootstrapping 2 CNN layers, linear layer for feature reduction, 2 LSTM layers and few fully connected layers. The input speech dataset passed into CNN+LSTM,

Revised Manuscript Received on September 22, 2019.
F. Catherine Tamilarasi, Research Scholar, BIHER, catherinetamilarasi@gmail.com
Dr. J. Shanmugam, Research Supervisor, BIHER, jshanmugam@yahoo.com

LSTM+DNN and finally to CNN+LSTM+DNN shown to achieve better feature extraction efficiency, good temporal modelling and minimal frequency variations. [11].

VI. RECURRENCE BASED CNN MODELS

RNN –Cognitive Input

Recurrent Neural Network and Reinforcement learningcombination are used in analysis related to Human Working memory, attention and decision making models. This combination will fit well to model to explore mapping between sensory input and output cognitive behavior routed through cortex or hippocampus in a closed loop [13].

Long Term RCN –Video input

Long Term Recurrence Convolution Network model combines the convolution layer and long range temporal recursion model called LSTM. Specifically, suitable for time series video learning. [6] Here the CNN performs Visual feature extraction and feature space representation. The LSTM then handles sequential learning of visual data. This same combination is proven to perform well in activity recognition, image description and video description tasks. [14]

RCNN – Visual Object Recognition

The concept of recurrence indicates time dependent feature variations. ie, nth time instant output is a function of (n-1) th time instant output and current input. Hence the Recurrent neural networks are based on combination of LSTM and Convolution Neural Network architectures. RCNN consist of several recurrent convolution layers followed by a global max pooling and softmax layer. This model was mainly proposed for Visual object recognition. One more advanced version of this model was Long- term Recurrent Convolutional Network. In all recurrent networks the features. [7]

IRCNN –Cognitive Input

The Inception Recurrent Neural Networks are motivated by recurrent connectivity of human brain synapses. The IRCNN is a RCNN built into Inception model. Architecture wise here a transaction block defined by set of 3 layers ie convolution, pooling and drop out are repeated after IRCNN block [7] and a final softmax layer added as final layer. In applications targeting minimal computational parameters the Inception layer is replaced simple models like Alexnet and VGNet.

VII. CNN - FMRI INPUT

In brain activity mapping studies, CNN is used to extractactivity patterns through fMRI decoding. CNN hierarchy and visual cortex hierarchy mapping is used to understand neural activity patterns. [15]. From fMRI the CNN combined with Autoencoder is used to extract features and feature maps related to target disease and build the classification model. [9] Tensorflow, and Python based Packages such as Niftynet,Pytorch and NiPy support fMRI Deep Learning Analysis.

VIII. G-CNN - FMRI INPUT

Graph Convolution Neural Network G-CNN a Bootstrapping based deep architecture proved classification accuracy of 70.86 in classifying Autism Brain Imaging Data Exchange (ABIDE). It uses CPAC Pre-processing Pipeline and 'Harvard Oxford'atlas and it is a binary graph based classification approach with showing improved accuracy due to use of 2 hyper parameters namely the size of ensemble and edge drop out probability.[19]. This has proven more suitable for fMRI processing.

IX. ENSEMBLE DEEP LEARNING FOR SPEECH RECOGNITION

CNN and RNN are the best proven DNN models in speech recognition. In [19] the ensemble technique of various combination is shown to prove higher classification accuracy. When CNN is used alone it has accuracy of 80% whereas when used as Linear and log linear ensemble ((DNN+CNN+RNN) its accuracy improved to 81.6% and 81.7% respectively. [20]

X. RESULT AND CONCLUSION

This paper summarizes the 9 different types of specialized Convolution neural network (CNN) architectures and their various configurations that are successfully used to process 7 different types of input data. Particularly it is highlighted that CNN architecture is more suitable to process and classify Functional Magnetic Resonance Imaging (fMRI) scan brain Image data and thereby enabling research in various Cognitive disorder diagnosis. Also it is understood that the Classification Accuracy of CNN can be improved by various combinations of Ensembling with other Deep Networks.

REFERENCES

1) Li Deng, "Deep Learning for Speech/Language Processing", The Journal of Neuroscience, Sept 6, 2015 Deep Learning Technology Center Microsoft Research, Redmond, USA Tutorial given at Interspeech,
2) Pierre Sermanet, "A Deep Learning Pipeline for Image Understanding and Acoustic Modeling", dissertation for the degree of Doctor of Philosophy, New York University January 2014

3) Daniel LK,Yamins,James, JDiCarlo, "Eight open questions in the computational modeling of higher sensory cortex",Current Opinion in Neurobiology Volume 37, April 2016,Pages 114-120, https://doi.org/10.1016/j.conb.2016.02.001.
4) Jenelle Feather , Josh H. McDermott, "Auditory texture synthesis from task-optimized convolutional neural networks"2018 Conference on Cognitive Computational Neuroscience.
5) Kell et al., 2018, Neuron 98, 630–644, May 2, 2018 ª 2018 Elsevier Inc. , "A Task-Optimized Neural Network Replicates Human Auditory Behavior, Predicts Brain Responses, and Reveals a Cortical Processing Hierarchy",https://doi.org/10.1016/j.neuron.2018.03.044
6) Yann Lecun, Yoshua.Bengio , "Convolutional Networks for Images, Speech, and Time-Series", page 255--258. The MIT Press, (1995)
7) Md Zahangir Alom ALO, Mahmudul HasanChris Yakopcic, TarekM Taha ," Inception RecurrentConvolutional Neural Network for Object Recognition",arXiv:1704.07709v1 [cs.CV] 25 Apr 2017.
8) Shawn Hershey, Sourish Chaudhuri, Daniel P. W. Ellis, Jort F. Gemmeke, Aren Jansen, R. Channing Moore, Manoj Plakal,Devin Platt, Rif A. Saurous, Bryan Seybold, Malcolm Slaney, Ron J. Weiss, Kevin Wilson,,"CNN ARCHITECTURES FOR LARGE-SCALE AUDIO CLASSIFICATION", 978-1-5090- 4117-6/17/$31.00 ©2017 IEEE ICASSP 2017
9) Rushil Anirudh, Jayaraman J. Thiagarajan," Bootstrapping Graph Convolutional Neural Networks for Autism Spectrum Disorder Classification", 24 Apr 2017 arXiv:1704.07487v2 [stat.ML] 30 Oct 2018.
10) Tara N. Sainath, Oriol Vinyals, Andrew Senior, Has¸im Sak, "CONVOLUTIONAL, LONG SHORT-TERM MEMORY, FULLY CONNECTED DEEP NEURAL NETWORKS"
11) Suniya.V.S, Dominic Mathew, "Acoustic Modeling Using Auditory Model Features and Convolutional Neural Network", 2015 IEEE International Conference on Power, 978-1-4673- 8072-0/15/$31.00 ©2015 IEEE Instrumentation, Control and Computing (PICC)
12) Elisabeth Anderson , "Automated Audio Content Analysis with Deep Convolutional Neural Networks" 21 April 2017 Thesis for degree of Master by Advanced Study in Software Engineering and Internet Architecture.
13) Geoffrey Hinton, Li Deng, Dong Yu, George E. Dahl, Abdel-rahman Mohamed, Navdeep Jaitly, Andrew Senior,Vincent Vanhoucke, Patrick Nguyen, Tara N. Sainath, andBrian Kingsbury , "Deep Neural Networks in acoustic modeling in Speech Recognition",IEEE SIGNAL PROCESSING MAGAZINE [82] NOVEMBER 2012 DOI : 10.1109/MSP.2012.2205597 Date of publication: 15 October2012 1053-5888/12/$31.00©2012IEEE
14) Jeffrey Donahue Lisa Hendricks Sergio GuadarramaMarcus Rohrbach Subhashini Venugopalan Kate SaenkoTrevor Darrell, "Long-term Recurrent Convolutional Networks for Visual Recognition and Description", Technical Report No. UCB/EECS-2014-180http://www.eecs.berkeley.edu/Pubs/TechRpts/2014/EECS2014-180.html November 17, 2014
15) K. Seeliger∗, M. Fritsche, U. G¨uc¸l¨u, S. Schoenmakers,J.-M. Schoffelen, S. E. Bosch, M. A. J. van Gerven"CNN-based Encoding and Decoding of Visual ObjectRecognition in Space and Time", bioRxiv preprint first postedonline Mar. 18, 2017; doi: http://dx.doi.org/10.1101/118091TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL.22, NO.10, OCTOBER2014.
16) Corey Kereliuk, , Bob L. Sturm, Jan Larsen , "Deep Learning and Music Adversaries" , 16 Jul 2015 arXiv:1507.04761v1 [cs.LG].
17) Rushil Anirudh and Jayaraman J. Thiagarajan,"BOOTSTRAPPING GRAPH CONVOLUTIONAL NEURAL NETWORKS FOR AUTISM SPECTRUM DISORDER CLASSIFICATION", Conference: Presented at: Machine Learning for Health Care, Boston, MA, United States, Aug 18 - Aug 19, 2017.
18) Li Deng and John C. Platt, "Ensemble Deep Learning for Speech Recognition", Microsoft Research, One MicrosoftWay, Redmond, WA, USA. .