

Significance of Learning and Memories in Computational Methods

^{*1}Kuraku Nirmala, ²Dr. K. Saruladha

^{1,2}Department of Computer Science and Engineering, Pondicherry Engineering College, Puducherry, India

Email: avnirmalas@pec.edu, charuladha@pec.edu

Received: 06th December 2019, Accepted: 20th January 2020, Published: 30th April 2020

Abstract

Learning is perpetual, support for learning is memories. Making machine learn requires abundant utilization of memory structures. Neural Networks plays catalytic role for the artificial intelligence and learning processes. Deeper learning was made possible with multiple neural networks playing in synchronization with artificial intelligence, machine learning and convolutional neural networks. Convolution is the basic psychological phenomenon right from the infancy to the adulthood in learning and classifying the real world inputs. Images can be understood by classifying the color palletes and the schemes. Voice is differentiated with the help of acoustic models and the infra sound detection models. Contribution of Convolution Neural Networks is a great gift to the world in making the machine understands the vitals aspects of sound and images. In this paper a study is presented on the influences of convolutional neural networks and long short-term memory networks that impose new ideas to develop intensive learning frameworks.

Keywords

Neural Networks, Artificial Intelligence, Machine Learning, CNN, LSTM.

Introduction

Memory is the perennial facility of the intellect which subsumes physical frameworks in brain where information is encoded, stored and retrieved when needed. Memory is mirror of vital and musing experiences; it has mechanisms of time variant conglomerate methods of attention and retention of information which is evolved purposeful influencing future action.

In the world of computations and applications, memory is conceptualized as types, stages, and processes. The types of memory are explicit and implicit memories, and then the stages are three in major namely the stages: sensory, short-term, and long-term (Atkinson & Shiffrin, 1968). The long-term memory requires emphasis on the cognitive techniques for memory improvisation. The processes are central to long-term memory are encoding, storage, and retrieval and they are three.

As types	Explicit Memory
	Implicit Memory
	Sensory Memory
As stages	Short-term Memory
	Long-term Memory
As processes	Encoding
	Storage
	Retrieval

Table 1: Memory Conceptualization

Almost the lifestyle of everyone has embraced with new technologies and particularly in social communication. The new technologies shape the lifestyle, how the way they are used. There is a recurrent theme and concern about the lifestyles, since they emerge in ancient civilizations theoretically and scientifically. These methods were first employed in 19th century as called positivism. The term positivism is coined by Auguste Comte, the French philosopher in 1847. Comte considered as first intellectual, who articulated positivism systematically in order to present empirical methods for metaphysics. Learning is education and thus learning theories are introduced in the cognitive psychology. Though education is about transfer of knowledge, and for centuries it has been “teachers dictate and pupil learn”. Since times immemorial, efforts have been put to identify the developing theories of learning. Some of the learning theories are based on behaviorism, humanism, cognitive theory and constructivism. Learning theories are conceptualized as frameworks; behaviorists look at the learning as an aspect of conditioning and advocate a system of rewards and targets.

Much significant progress has been enhanced by the using frugal computer emulations. The natural phenomenon of learning is induced into the computers, the way that a learner thinks and acquires knowledge for future actions of decision making. Computational Intelligence is to make the computer able to learn and work on a specific task from the supporting data and the experimental observations. Computational Intelligence and Artificial Intelligence seek common goals. Crusading the opinions and arguments of the facts around the world and mathematically inducing principles of learning and connectivism is artificial intelligence. A set of nature-inspired algorithms and phenomena that address the common real-world problems easing the complexity with mathematical reason is developed as a programming paradigm which enables a computer to learn are neural networks.

Advent of Neural Networks

A neural network is often alternatively termed as artificial neural network is an information processing paradigm inspired by the process of human brain. A novel structure of information processing system evolves as the key element of the neural networks [1].

Warren McCulloch and the logician Walter Pitts were the neurophysiologists, who developed a neural network model in 1943.

Neural networks looms into varied methods of problem solving when compared to conventional computers. A set of statements/instructions that solve a specific problem is an algorithm which is followed by Conventional computers. A conventional computer cannot solve the problem unless the specified steps of algorithm are followed [1]. Thus conventional computers are restricted in the problem solving capability of particular instances. But conventional computers would do so much more useful that they could do things that they are actually programmed how to do. In neural networks neurons are the elementary components that are highly interconnected processing components to solve problem. As the computers cannot be programmed to execute a specific task, they are built to learn by example and selected examples prove their performance. However, as it does by learning, the way how the problem is solved is unpredictable which a disadvantage in it is.

Problems can be solved by cognitive approaches using conventional computers because; the small unambiguous instructions, the procedure how the problem is to be solved must be known. Such are converted into a high level language and subsequently as a program, into machine understandable code. These practices on machines are predictable; however, some software or hardware fault may cause error in the process. Thus, neural networks and conventional computers working on algorithms are complimentary to each other [1].

Traditional von Neumann architecture simplified the total computing scenario, in lieu the success of neural networks. Ironically, even John von Neumann has suggested the imitation model of neural functions by using scientifically traditional telegraph relays or vacuum tubes.

Similar network independent of one another using matrix mathematics and arrays using of analog ADLINE circuits is inspired by Kohonen and Anderson in 1972, where the neurons play the role of activating a set of outputs.

John Hopfield of Caltech, in their consensus, and a paper presented at National Academy of Sciences, caused a renewal of interests in neural computing with mathematics. Modeling the brain using the mathematics and computing devices targeting the mathematical analysis of thoughts, which is a baseline technological myth for neural computing is developed.

A US-Japan Joint Conference propounded in Kyoto, Japan, on Cooperative or Competitive Neural Networks. A Fifth Generation effort in Neural Computing has been announced with abundant funding.

In 1985, the American Institute of Physics competed with - Neural Networks for Computing and forthwith in 1987, the Institute of Electrical and Electronic Engineer's (IEEE) first International Conference on Neural Networks which drew much attention for neural computing engineers.

Bernard Widrow, in 1989 at the Neural Networks for Defense engagement of World War IV, "World War III never happened," where the battlefields are world trade and manufacturing. In 1990, the defense department of US, raised as Department of Defense Small Business Innovation Research Program has announced plethora of topics which ideally had targeted neural networks.

Role of AI

Artificial Intelligence of Charvaka (Cārvāka) leads the precursors of the Indian Materialism, which was alternatively known as Lokāyata and Bṛhaspatya. Direct perceptions, empiricism, and conditional inference has been directly slashed by Charvaka. Rejecting Vedas and supporting philosophical skepticism was observed as super sources of knowledge. The Greek mythology introduces Mechanical men and artificial beings, as the golden robots. Thus, even in the present, artificial intelligence has constantly encouraged as an important element of science fiction. Chinese and allied Egypt, Greece have carved Realistic humanoid automatons.

The myth or fact that the mechanization of human thought have been observed as Artificial Intelligence, though the meaning of mechanical or format has a long told histories. Formal deduction and deductive intelligence are developed in the first millennium BCE by Chinese, Indian and Greek philosophers. But in 1940s and 50s handful of scientists from variety of field believed mathematics, psychology, engineering and economics are the sources of possibilities of creating an artificial brain. Further, the field of artificial intelligence research was confounded as an academic discipline in 1956.

Claude Shannon and Nathan Rochester of IBM and Marvin Minsky, John McCarthy are the two senior scientists, at the Dartmouth Conference in 1956 roosted that they were in the era of discovery, of springing across new ground of imitating human intelligence. Programs to solve algebra, word puzzle problems, proving theorems of geometry and abilities to learning and speak English were developed in their research laboratories [1][2][3][10]. The denizens of the research at times have to believe that “intelligence” is a behaviour by machines and it was possible.

Machine Learning

Programs that create programs to accomplish specific tasks. The artificial intelligence with a set of statistical inferences from the huge collection of data summarizing and bringing the quintessence of the information useful for decision making. The concept of a program that thinks, rather a program that can build an algorithm for a given known task. Programs that learn about programs or real world problems. To critically understand this concept, statistical thinking of a program with artificial intelligence and learning are embellished into a phenomenon. Arthur Samuel of IBM's Laboratory, Some of the very beginning, the machine learning programs were practiced by Poughkeepsie, they play checkers. Perceptron is invented by Frank Rosenblatt at the Cornell Aeronautical Laboratory.

The optimistic definition that developed about Machine learning (ML) is, it is the meticulous study of algorithms and statistics. The algorithms and statistical models drive the computer systems progressively and further improve their performance about a specific task. A mathematical basis is vented in the machine learning to build a model and samples of data used to be known as “training data”, that develops predictions or decisions logically and without the dire need of explicit programming to perform the task.

A set of data, mathematical model that contains inputs and known outputs, based on this phenomenon Machine learning tasks are classified into several broad categories. The algorithm that is build from a mathematical model based on the set of data is a supervised learning model. The algorithm that is build from a mathematical model even, based on the characteristics of data rather the values is an unsupervised learning model.

Deep Learning

Deep learning is a theme oriented learning combining artificial neural networks with multi layered perceptrons and machine learning [2]. Identifying the learning capability and which paths of learning network are responsible for success and as well as failure. Incorporating changes to the paths in the network that leads to failure in order to improve the performance. Minsky, 1963 – had developed *fundamental credit assignment problem*.

The anatomy of a standard neural network (NN) contains of so many simple neurons which act as, connected processors, each neuron produces sequences of activations [1]. Sensors activate the input neurons perceived by the environment; weighted connections are active neurons which are otherside associated in the neural network. Neurons also are triggered by the environment actions. Long causal chains and transformations (non-linear) are seen as development of computational stages in learning, which is depending on the problem and the neurons, their constitution of connections. The aggregation of activation of the networks is based on the weights or assigned credits for the long chains. The concept of Deep Learning is also about; how accurately credit assignment is done across many such stages.

Models of neural networks academics and research are shallow neural network having limited capabilities [2]. Even models are generated to overcome hindrances using non-linear layers of neurons which are at least dated back to 1960s. The decension of the paths of a non-linear neural network to the increase or decrease the level of learning is guided by a gradient. A supervised learning method proposed to guide the network as a “teacher-like model” is ailed with arbitrary depth of learning overcome by the back propagation. Such a contribution in the research of neural networks with thick resources of networks of back propagation leads to deep learning networks.

Thus, deep learning consents new computational models which are composed of multiple processing layers that have learning capabilities of different representations of data characterized with multiple levels of data abstraction.

Convolutional Neural Network

The major applications of intelligent computing and cognitive architectures are targeting on informal media of videos, images and undocumented texts. A randomly taken video footage or a snapshot of image gives clues to the intelligent teams to dig out the culprits, or historically significant fossils that tell stories about the predecessors. Nowadays data is not mere text and numbers, rather a cloud of images, videos and undocumented texts piled in all the vaults of hobbyists, secret agents and huge research agencies. What is done to these collected data or information? Is a big question that arises for the common researchers? The ERT teams mine patterns and understand the cause and effect of the incidents in the daily life. The medical images that read so many health related characteristics are very difficult to analyze and sort them and enable them to be available for the studies. Learning is the most important task that should a choice for data analyst to understand the hidden characteristics.

For which, the learning mechanisms miniaturize the process and bring forth the essence of stored data. Learning is not guaranteed as accurate, the question of accuracy goes to a level of belief that the identified patterns and their usefulness. Yet an intensive learning in such scenarios is deep learning. Deep learning requires deep neural networks and no matter they are linear or non-linear [2][4][5]. With a minimal preprocessing and learning rich content from videos and images with convolution of process. Thus a better learning process that deals with huge collections of perceptrons and multi-layered networks is Convolutional neural network. Though neural network is an analogy of the human brain neuron, the Convolutional networks were also been inspired by the biological processes, where the connectivity pattern between neurons bear a resemblance to the organization of visual cortex. The Convolutional Neural network has another important analogy of the individual cortical neurons responding to the stimuli in the restricted visual regions known as the receptive field the same is engulfed in the architectures. The convolution is a partial overlap of the reception of the different neurons covering the complete visual field [4][5].

The convolution operation is escalated in convolutional layers by passing the result to the next layers. The response is emulated to visual stimuli in the convolution process by each individual neuron. Convolutional neuron processes data for respective receptive field. It is not practically application for image classification, although these contain feedforward neural networks that are fully connected, hence, it is not practical to apply this architecture to images. A shall and deep architecture is necessary for congregating the neurons to work on large input sizes of associated images, where each pixel is considered to be a variable with relevance.

Hubel and Wiesel worked in the 1950s and 1960s on the cat and monkey and showed that visual cortexes contain neurons that individually respond to small regions of the visual field. Convolutional neural networks, thus proven to solve critical problems in vision, robotics comprising of convolution layer, activation layer, normalizing layer, pooling layer, loss layer and regularization. These steps refine the learned images or inputs and derive the essential portions of the images or policies of decisions.

Convolutional Neural Networks a.k.a. ConvNets has its most profound applications in speech recognition. The structure of CNNs directly accommodates the variances in speech. Defining weight parameters to the vocal tracts and the sound that they produce with time invariability, with a degree of invariance to small frequency shifts, which generally are available in natural speaker speech, are handled in CNNs very effectively [10][11]. Like in image processing the CNNs also share limited weights and perform convolutions and mappings to produce quality recognition architectures. CNN composing with DNN and RNN develop into resilient architectures. At a rudimentary test bed of speech recognition platform, CNNs are even used in acoustic modeling. Acoustic modeling is used to identify the human speech with various models of tones and gender.

Spectrogram play a very important role in supplying specific and condensed input to CNNs in speech recognition architectures [10][11][12]. CNNs even are with little sound inputs help of SAD (Speech Activity Detection) methods are useful in modeling the infra sounds of creatures like insects and small animals. Using the SAD methodology the CNNs can be used to establish sound channels and apply the noise filters and enhance the candidate sound data. Pooling of inputs reiterated in the CNNs improvise the channel models.

Speech Command Recognition is a popular application using CNNs. As the rapid development of mobile and hand held device usher in the information technology influencing the civil lives, the CNNs has greater contributions of command recognition [11][12][13][14]. The phrasal commands are interpreted using the channelization of the sounds and the temporal delimitation is observed in the speech command recognition applications. Keyword based filters in speech command recognition is a improvised achievement in the research of CNNs using hidden markov chains. The HMM (Hidden Markov Model) synthesizes the speech into separation of words and voice command dictionaries and grammar can be elevated in the commands in order to fulfill the end user requirements [12]. Research is due in the areas of expression of word shouts and commands are being processed with formulated feelings using the HMM and SAD models in the way to the sophistication of CNN architectures for Speech Recognition.

Statistics About Research Contributions

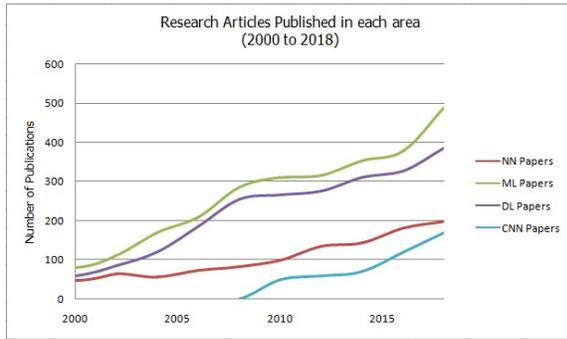


Fig 1: Research Articles Publish in the areas of Neural Networks, Machine Learning, Deep Learning and Convolutional Neural Networks in the years between 2000 and 2018.

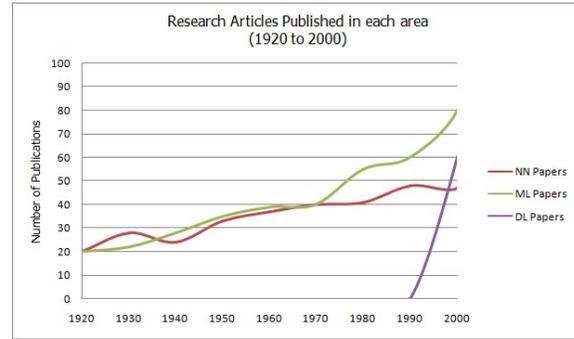


Fig 2: Research Articles Publish in the areas of Neural Networks, Machine Learning and Deep Learning in the years between 1920 and 2000.

Long Short-Term Memories

Amongst the applications of neural networks, recurrent neural networks become standard tools for natural language processing. "... out of all the applications of language processing and modeling RNN are used for translations and answering the questions", by Sutskever et al. (2014), Jozefowicz et al. (2016), and Hewlett et al. (2016). In particular, for the faculty of Natural Language Processing the concepts of Long Short Term Memory (LSTM) architectures developed by Hochreiter & Schmidhuber (1997) plays an important role of acting as the rudimentary building blocks.

Learning compact structures in Convolutional Neural Networks (CNNs) have been widely explored in the past few years. Han et al. (2015b) proposed connection pruning for sparse CNNs. Pruning method also works successfully in coarse-grain levels, such as pruning filters in CNNs (Li et al. (2017)) and reducing neuron numbers (Alvarez & Salzmann (2016)). Wen et al. (2016) presented a general framework to learn versatile compact structures in Deep Neural Networks such as neurons, filters, filter shapes, channels and even layers [6][7][8][9].

Convolution neural networks alternatively or with enhancements with the support of memory and iterative inference yielding is carried into a Long Short-Term memory. The combination is also referred to as LSTM network models a memory based cognitive architecture [7][8][9]. Feature layers used to extract the elementary knowledge and embedding layer maps features to real-valued vector representations. Inherent the networks of these layers are looped using recurrent network model. The LSTM layers work on the activation functions and work on gradient problems. Element wise computation takes in the LSTM layers to calculate the new states of memory cells. High dimensions of input feature vector comprises of the inputs that even process the inputs in the reverse orders using special learning parameters.

LSTMs are stacked in Recurrent Neural Network for analyzing the component wise knowledge in high dimension input vector models. Smaller hidden LSTM training units are packed into Recurrent Neural Networks and in order to impose learning from scratch. High profiled learning with comprehensive data is possible with contextually embedding layers to extract representations from input vectors.

LSTM Contributions

LSTM has a very latest applications and much variance of combinations with CNN and RNN have been contributed in the research.

James Murdoch and Arthur Szlam of Berkeley and Facebook in Feb, 2017, New York have proposed "Automatic Rule Extraction from Long Short-Term Memory Networks". A novel method of visualizing the importance of specific inputs determining the output of LSTM. [7]

Hai Li et. al, in March 2018, in the works "learning intrinsic sparse structures with in Long Short-Term memory", proposed significant model compression for the wide adoption of recurrent neural networks in devices with limited resources and in business clusters. The concept of intrinsic sparse structures will simultaneously evaluate the compact models without losing any perplexity of a model.[8]

Angrosh Mandya et. al., in November 2018, in their works proposed combined orchestration of CNN with LSTM, entitled “Combining Long Short Term Memory and Convolutional Neural Network for Cross-Sentence n -ary Relation Extraction”. The framework has proved the results for the “that exploits both word embeddings and position embeddings for the task of cross-sentence n -ary relation extraction” [9].

Table 2: A Chronological Order of Achievements in Machine Learning, Deep Learning and AI.

Year	Contribution	
1950	Turing Test (or) Turing’s Learning Machine	To determine if a computer has intelligence, A human being should be unable to distinguish the machine from another human being by using the replies to questions put to both.
1950	Genesis of Machine learning	Simple algorithms are used in the research of Pioneering machine learning.
1952	First learning program of game checkers	Arthur Samuel begun working on the first algorithms of machine learning and developed programs that play checkers, where the programs made a winning strategy. At IBM's Poughkeepsie Laboratory.
1957	First Neural Network for Computers	Frank Rosenblaut invented the perceptron (the first neural network) to simulate the thought process of the human brain. Great deal of excitement is development in the research of neural networks with the invention of the perceptron.
1963	Machines Playing Tic-Tac-Toe	Introduction of reinforcement learning to play Tic-Tac-Toe (also famously known as noughts and crosses), machine learning by Donald Michie creates a 304 match boxes and beads.
1967	The nearest neighborhood algorithm is written	For basic pattern recognition by computers, used to map routes for travelling salesperson (starting at a random city but ensuring they visit cities during a short tour).
1969	Limitations of Neural Networks	Marvin Minsky and Seymour Papert has published their book on Neural Networks and Perceptrons. The book describes some of the limitations of perceptrons and neural networks. The fundamental thought of neural network seems to have hindrances.
1970	Automatic Differentiation (Backpropagation)	A general method for automatic differentiation (AD) using discrete connected networks with nesting of differential functions is published by Seppo Linnainmaa. A modern back propagation method is fused into the research.
1979	Remote control vehicle or Mobile Robot (or) Stanford Cart.	A cart that can navigate and avoid obstacles intelligently in a restricted room using visual guidance is developed by the students at Stanford University.
1980	Neocognitron	The first book on Convolutional Neural Networks and as a type of Artificial Neural Networks is published by Kunihiko Fukushima. The fact in the book is about the real time comments on the neocognitron which is a varied artificial neural network (ANN) that inspires a convolutional neural network.
1981	Explanation-based Learning	This makes the computer analyze data (also known as training data) to create a general rule so that it can discard unnecessary data.
1982	Recurrent Neural Network	A model of content-addressable memory system as a recurrent neural network is developed by John Hopfield.
1985	Speech Recognition	The computer learns to sound the words the same way a kid does. Phoneme input are given.
1986	Backpropagation (NN)	Usage of internal representations and application neural networks as Back propagation is first applied by Paul Werbos,
1989	Reinforcement Learning	The concept of Q-learning is developed by Christopher Watkins, which is feasible for reinforcement learning.
1990	Machine learning shifted from knowledge-driven to data-driven approach.	Computer programs are developed by scientists to analyze voluminous training data to generate conclusions and learn from data results.
1995	Random Forest Algorithms	Papers on random decision forests are published by Tin Kam Ho.
1995	Support Vector Machines	Publications on support vector machines by Corinna Cortes and Vladimir Vapnik.
1997	IBM Deep Blue Beats Kasparov	World Champion Chess is won by IBM's Deep Blue.
1997	Long Short-Term Memory	LSTM the Long Short-Term Memory is invented by Sepp Hochreiter and Jürgen Schmidhuber, which improvises the efficiency and practicality of RNN.
2002	Torch Machine Learning Library	First software library for machine learning, Torch API and SDK.
2006	Deep Learning	To make computer visually distinguish objects and text in pictures and videos.

Conclusions

In the field of computer science Cognitive technology have ushered into various dimension that mimics functions of the human brain through various means, further empowered with enhancements of learning and obtaining pin-point comprehensions in various high dimensions of vectored and non-vectored data, orchestrations of DNN, RNN, CNN and LSTM has to take place no matter they are specific in their applications. The intense mixes of these cognitive technologies derive a novel concept or learning and comprehension to the next generation.

References

- [1] Simon Haykin, “*Neural Networks and Learning Machines*”, (C) 2009 by Pearson Education, Inc., Upper Saddle River, New Jersey 07458.
- [2] Goodfellow, Ian, et al. “*Deep learning*”. Vol. 1. Cambridge: MIT press, 2016.
- [3] Kim, Kwang-Ki K., Ernesto Ríos Patrón, and Richard D. Braatz. “*Standard representation and unified stability analysis for dynamic artificial neural network models.*” *Neural Networks* 98 (2018): 251-262.
- [4] Deng, Li, and Dong Yu. “*Deep learning: methods and applications.*” *Foundations and Trends® in Signal Processing* 7.3-4 (2014): 197-387.
- [5] Kwang Gi Kim, “*Deep Learning*”, *Healthcare Information Research*. 2016 October;22(4):351-354.
- [6] Sahu, Sunil Kumar, and Ashish Anand. “*Drug-drug interaction extraction from biomedical texts using long short-term memory network.*” *Journal of biomedical informatics* 86 (2018): 15-24.
- [7] Murdoch, W. James, and Arthur Szlam. “*Automatic rule extraction from long short term memory networks.*” arXiv preprint arXiv:1702.02540 (2017).
- [8] Wen, Wei, et al. “*Learning intrinsic sparse structures within long short-term memory.*” arXiv preprint arXiv:1709.05027 (2017).
- [9] Mandya, Angrosh, et al. “*Combining Long Short Term Memory and Convolutional Neural Network for Cross-Sentence n-ary Relation Extraction.*” arXiv preprint arXiv:1811.00845 (2018).
- [10] Ossama Abdel-Hamid, Abdel-rahman Mohamed, Hui Jiang, Li Deng, Gerald Penn, and Dong Yu, “*Convolutional Neural Networks for Speech Recognition*”, *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, Vol. 22, No. 10, October 2014.
- [11] LeCun, Yann, and Yoshua Bengio. “*Convolutional networks for images, speech, and time series.*” *The handbook of brain theory and neural networks* 3361.10 (1995): 1995.
- [12] Thomas, Samuel, et al. “*Analyzing convolutional neural networks for speech activity detection in mismatched acoustic conditions.*” 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2014.
- [13] Warden, Pete. “*Speech commands: A dataset for limited-vocabulary speech recognition.*” arXiv preprint arXiv:1804.03209 (2018).
- [14] Takahashi, Naoya, et al. “*Deep convolutional neural networks and data augmentation for acoustic event detection.*” arXiv preprint arXiv:1604.07160 (2016).
- [15] Wikipedia.