Available online at www.sciencedirect.com
ScienceDirect
ICT Express 4 (2018) 100–106
www.elsevier.com/locate/icte

Are the long–short term memory and convolution neural networks really based on biological systems?

David Balderas Silva*, Pedro Ponce Cruz, Arturo Molina Gutierrez

Tecnologico de Monterrey National Department of Research, Puente 222, Del. Tlalpan, Mexico City, 14380, Mexico

Received 24 January 2018; accepted 10 April 2018
Available online 17 April 2018

Abstract

In general, it is not a simple task to predict sequences or classify images, and it is even more problematic when both are combined. Nevertheless, biological systems can easily predict sequences and are good at image recognition. For these reasons Long–Short Term Memory and Convolutional Neural Networks were created and were based on the memory and visual systems. These algorithms have shown great properties and shown certain resemblance, yet they are still not the same as their biological counterpart. This article reviews the biological bases and compares them.

Keywords: Classification methods; Sequence prediction; Image classification; Vision; Memory

1. Introduction

Two automation problems are classification of complex inputs (i.e. images) and sequences prediction. The first problem requires interpreting the input by understanding what is in it. This can be solved by finding patterns, decomposing them and later appointing a class. On the other hand, sequence prediction requires to have recollection of past events that are similar to the one presented, establish a pattern and make a prediction. Consequently, it is even more problematic to predict sequences of complex inputs. Thus, to solve these problems, it requires an algorithm that can decompose the input and has some sort of memory.

The brain is capable of doing these things easily. Hopfield in 1988 defined the Human brain as computer made out of organic material and wet chemistry but, nonetheless, its one of the world's best computers [1]. This statement should not be taken lightly, while the brain requires more effort for mathematic operations than its silicon computer counterpart, its performance for object recognition and decoding natural language outperform most machines. The brain is so powerful that can quickly separate different inputs (visual, audio, sensory, motor, etc.), process them, discard information and decide an action, while still having control of the organs.

At the moment the brain is far from being understood, nevertheless there are several areas and their basis on how they work that can serve as inspiration for new algorithms. Like, the auditory area that processes information of volume, pitch, sound localization, rhythmic patterns and understanding language; the sensory area involved in cutaneous and other senses; the Motor area, in charge of muscles voluntary control; the Visual Area, which decomposes images to understand information of spatial localization, object detection, recognition, movement, and color. Another area, is the memory area, in charge of collecting information, separating, reinforcing or decreasing connections and many more.

Two algorithms based on brain areas that excel in classification, sequence prediction and decomposing complex data are Long–Short Term Memory and Convolutional Neural Networks. Even more important is that these two algorithms are currently being combined for classification of even more

* Corresponding author.
E-mail addresses: dc.balderassilva@itesm.mx (D. Balderas Silva), Pedro.Ponce@itesm.mx (P. Ponce Cruz), armolina@itesm.mx (A. Molina Gutierrez).
Peer review under responsibility of The Korean Institute of Communications and Information Sciences (KICS).

https://doi.org/10.1016/j.icte.2018.04.001

complex data. As said before, these areas are not completely understood, yet there are many theories on the basis of how they work. This paper intend to give a review of neural, visual and memory areas, to have a better understanding on how they work, and have a view on what they are currently missing.

2. Brain activation

Humans are exposed to abundant information which has to reach many areas to be processed, stored and discarded. The problem is most of this information is redundant, and unneeded, and requires to be discarded. Another problem is that most processes start at different times to accomplish a single task (grasping starts with arm movement before the fingers are completely configured). To accomplish these and understand what is happening inside the brain it is important to distinguish the path each signal follows, starting with nerve activation from external inputs, to the processing center or storage units.

2.1. Neural activation

Neurons are the basic elements in the brain, that adapt their connections to other neurons to create representations or process the information. A neuron can either change their connections or readiness to release transmitter (temporary increased or decreased without requiring either activity in the neuron) to communicate with other neurons. It is important to remember that these changes in connections can be persistent or just temporal [2].

Information first received by a neuron from other neurons or sensors by means of a chemical (neurotransmitters) that arrives at the dendrites. Then, if the input is strong enough, the neuron will generate a pulse which starts on the soma and is transmitted through the axon terminal branches to other neurons.

Nerve electrical pulses, also known as action potentials, are short lasting electrical events which briefly reverse their membrane polarity, from negative (polarized) to positive (depolarized), by changing the ion concentration in and out the cell of Potassium and Sodium. Although this is true, the real mechanism to keep transmitting information is still unknown, there are areas where exist a clear change in the frequency when information is presented. Such as the relationship frequency/strength in efferent (motor) peripheral system and increase of fire rate in afferent (sensory) central nervous system. Also, there are frequency bands where the information is thought to be coded. These frequency bands appear to be operational on different states of the mind. For example, Delta Rhythm (1–3 Hz) is a state that is present in a deep sleep state or coma or Beta Rhythm (13–26 Hz) & Gamma Rhythms (27 Hz >) are seen in awake states.

Notably, there is no full explanation on how the codification is done, and what happened at specific frequencies or why they are limited to them. Also, neurons adapt to constant stimuli with a gradual reduction to the fire rate. Meaning that constant stimuli cannot be associated to a single rate code, in other words coding cannot be implied from a simple stimulus or frequency.

2.2. Brain areas

In normal conditions, an average human brain contains 100 billion neurons, with an average of 1000 synapses connections each (around 10^{14}), which makes it complex to know exactly what path an instruction or memory goes inside the brain. Yet, inside the brain there exist areas that work together. For example, general areas would be: the frontal lobes that are responsible for problem solving, judgment and motor function; the Parietal lobes manage sensation, handwriting and body position; the Temporal lobes are in charge of memory and hearing; and the Occipital lobes are responsible of visual processing system. Another areas are located in different areas of the brain, like memory that is spread in different parts depending on the input or vision that starts at the eyes and ends in the back of the brain.

2.3. Memory

Understanding how the brain creates memory is something that has intrigued many scientists over the years. Hilgard and Marquis defined learning as "the change in strength of an act through training procedures", while Donald Hebb proposed that "memories are stored by networks that strengthen their connections to increase the likelihood of same activity patterns being recreated at a later date" [3,4]. These mean that learning has to be done through a process of repetition to enhance the connections so they could later be used.

Certain approaches to memory hypothesize that items are represented by different patterns of activity among the same neural elements as opposed to occupy different locations that could be linked from associations. This means, that using different patterns of activity, a group of elements could represent different memories which can be retrieved later by generating the same pattern. Then, a specific memory must be represented by a pattern of micro-features, or a particular set of active units. In other words, recollecting a memory must involve the reconstruction of information from different locations of the brain.

Specifically, the traces of these locations are called engrams [5], as the persistence encoding in neural tissue that provides a physical basis for the persistence of memory, in other words a memory trace. Notably, locating these traces is not simple but certain characteristics have been defined. To begin with, engrams are persistent changes in the brain consequence of specific experiences or events, they have the potentials for ecphory (memory retrieval process triggered when a specific cue is proximal to a memory), and, the content of engrams shows what happen at encoding and predicting what can be recovered in subsequent process.[1]

It is believed that the formations of an engram require a strengthening on the synaptic connections in a set of neurons active during encoding, which generates a neuron ensemble. Therefore, increase in synaptic connections increases likelihood that the same path used for encoding will be used for

[1] Engrams may exist dormant (the path exists even before a memory can be created or recovered).

retrieval. Engrams have to be both distributed and localized; highly differentiated and specialized. Therefore no single memory center exists, with many parts participating in individual events. Also it involves limited number of brain systems and pathways with each part contributing differently to the representation.

Additionally, information storage is tied to a processing system in charge of analyzing incoming information. This storage can modify extra data sent to similar analyzing channels. For example, a string characters heard cannot hardly be recalled if another word is presented using the same voice and localization [6], or a flashing light 100 ms after a watched letter can erase its memory [7]. Furthermore, many feats that require enormous memory capacity highly depend upon processing skill. For instance the amount of chess board pieces remembered within a board is larger as the expertise of a player increases, but only if the pattern has been seen before otherwise the number of pieces remembered is similar to a naive [8].

It is important to know that engrams are dynamic (some recordings are plastic), and a memory that has already been established can be updated or changed with new information, or retrieving a memory may transiently destabilized a previously consolidating engram [9]. An example of this dynamic behavior is reversal learning, where a second discrimination is processed, but with the originally "correct" stimulus made "incorrect" and vice-versa.[2]

All the types of learning show that the tracking engrams is not easy, and requires to study the changes in neural substrates (level of molecules, synapses, neurons, neuronal ensembles, and/or brain circuits and networks). One of the first to study this changes was Ramon y Cajal [10] who correctly hypothesized that axons connect with neurons at protrusions (dendritic spines) and proposed that experience-induced modifications would happen at these connections (later identified as synapses). Following the steps of Ramon y Cajal, Donal Hebb [4] said that connection strength will increase between simultaneously firing units creating neurons ensemble.

Learning has been characterized by changes in the structure of neurons and depend on synaptic modifications of a biochemical or biophysical event, and can be accompanied by morphological alterations in neurons structure. Other variations appear in the synaptic compounds; in weight and thickness of the somata; in the number, length and synaptic diameter of dendrite branches; and in the number and shape of dendrite spines and in the neural excitability.

Some of these changes have short duration (i.e. alterations in synaptic compounds), others persist for longer periods (i.e. alterations in synaptic strength) while in rare occasions maybe even transmitted to future generations (alterations in DNA). Yet, these changes are inter-related (variations in compounds might result in alterations in synaptic strength) and inside of them underlies the formation of neural ensembles. For a better understanding on the formation of this traces they can be explained by using the two types of memory: Short-Term Memory and Long-Term Memory.

Short-Term Memory (STM): is an ephemeral memory (∼15 to 30 s), that can be lost if there is a distraction. It is believed that STM is a neural modification accompanying behavioral habituation (waning of response that occurs after repeated stimulation-located early in the pathway) or sensitization (repeated administration of stimulus results in the progressive amplification of response), which includes presynaptic changes in the ability of sensory neurons to release transmitter.

The idea of STM is that it holds information to accomplish an event that is planned to do, like exploring different possible solutions mentally before choosing one to make. The ability to hold information for task completion is a human characteristic, causing areas of the brain to become very active, especially the pre-frontal lobe that is highly developed in humans compared to other species.

Long-Term Memory (LTM): is a recollection that persists. For STM to become LTM it must go through a maturation process, which makes the trace resistant to some agents that can impair or erase the STM [2]. Such maturation is known as the consolidation process, which is either cellular or synaptic and any of them can last for few days to many years.

To convert from STM to LTM information goes through the hippocampus, which sorts out new sensations, compared them to previously saved ones and creates associations. Afterwards, to memorize new facts, information is passed through the hippocampus several times, which strengths associations among new and old facts, until it is no longer necessary. At these points, the corresponding cortex area had already learned proper associations to reconstruct a memory. Notably, although hippocampus is the catalyst to convert information, it is not regarded as the memory center, since engrams are encoded in several places. Yet, subjects with lesions in the hippocampus cannot store new information for more than a few moments.

It is believed that the process to store memory is called Long-Term Potentiation (LTP), which is a persistent strengthening in synaptic connections due to continuous stimulation, which becomes persistent over the synapse connection with just a few minutes of stimulation [11]. The changes are accompanied with temporary changes (∼8 h) in the dendrite spine connections becoming rounder. Some experiments showed that in order to strengthen associations, LTP uses a protein ($PKM\zeta$) which acts on specific synaptic structures, modifying synapse's micro structure increasing the number of functional post-synaptic receptors. Resulting in persistent enrichment of synaptic transmission which encodes the memory.

Assuming a continuous strengthening due to LTP, the synaptic connection would become so tough that it would be impossible to encode new information. Hence, to conserve the neuron plasticity, a process that weakens this connection must exist [12]. This is known as Long-Term Depression (LTD). Unlike LTP which is a brief and high frequency stimulation, LTD occurs when the neurons are stimulated at low rate for long periods. Notably, these two processes are complimentary,

[2] In both cases the learning normally takes longer time since it has to overwrite established connections.

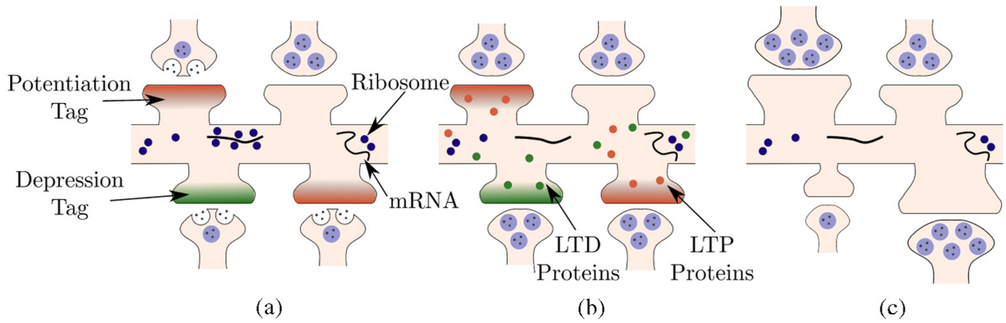


Fig. 1. (a) Potentiation and Depression tags are generated, (b) generation of protein for LTP and LTD. (c) Synapse connection are enhanced and reduced by LTP and LTD.

while LTP can enhance excitatory post-synaptic potential, LTD can reduce it, making the knowledge adjustment possible (see Fig. 1).

2.4. Visual system

Human visual system is undoubtedly one of the world wonders giving the ability to observe its physical environment. Even blurry images or deformed objects can be recognized. The visual system learns how to distinguish between shapes, sizes, wavelengths, orientations and movement directions. It is able to do all these things by dividing the information in two as it enters the eye, sends it back into visual cortex, with millions of neurons, with tens of billions connections between them which dissect the image for processing. The visual cortex is divided into Primary visual cortex (V1) and an entire series of visual cortices (i.e. V2, V3, V4, and MT) that progressively process the more complex images, see Fig. 2.

A little bit more general, the visual system consists mainly of two parts: the eyes and the brain. Whereas brain works as a complex image processing tool, the eye functions as the equivalent of a camera. First, the eye focus on an object and captures the light reflected from it, then, the light hitting directly the eye passes through the cornea (a transparent protecting layer working as a lens and refracting light), going through the iris (which determines the amount of light to let through), passing across the lens, to reach its final destination the retina. There, an electrical signal will be produced through the optical nerve, as long as the light bean is within the range of electromagnetic spectrum (about 300 to 700 nm).

In the retina, light is converted to electrical impulses by two types of photoreceptors: rods, responsible for vision on low light levels (scotopic vision) perceiving only shades of gray, and cones, active on higher light levels (photopic vision) perceiving colors and are responsible for high spatial activity. In fact, to detect colors, there exist three types of cones, and each is sensitive to a different band of electromagnetic spectrum (low-wavelength light red, middle-wavelength light green and short-wavelength light blue).

These photoreceptors work together to produce an image. An example of these is the inhibitory behavior, which is lateral inhibition on the eye that helps localizing boundaries by deblurring, contrast enhancement (inhibition makes changes look more pronounce near the edges, i.e. Mach bands) and edge detection.

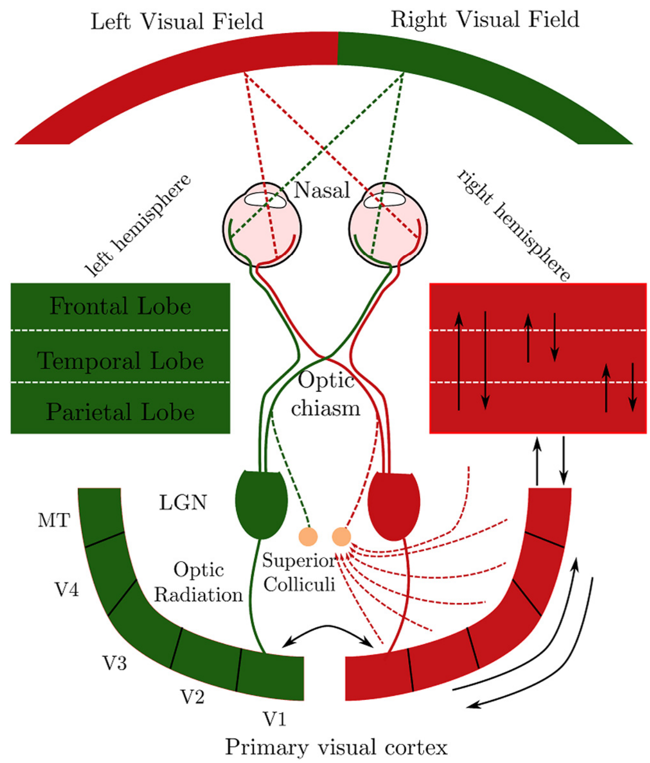


Fig. 2. The visual system with different layers.

The information generated by the photoreceptors is compressed and transfer through two types of ganglion cells: the magno and parvo cells. They differ on size, dimension of its receptive field and conduction rate. In one hand, magno cells are larger, with a big receptive field and higher conduction rate and they are in charge of mediating information about depth and motion. On the other hand, parvo cells are smaller, have a short receptive field and slow conduction rate, and they are in charge of processing information about the color and detail. Evolutionary, these differentiation on the cells could be explained since it is more important to know if something is approaching fast than knowing what exactly is coming.

Notably, there are around 10^8 rods and cones in each eye, but only 10^6 ganglion cells axons in the optic nerve. So, a single

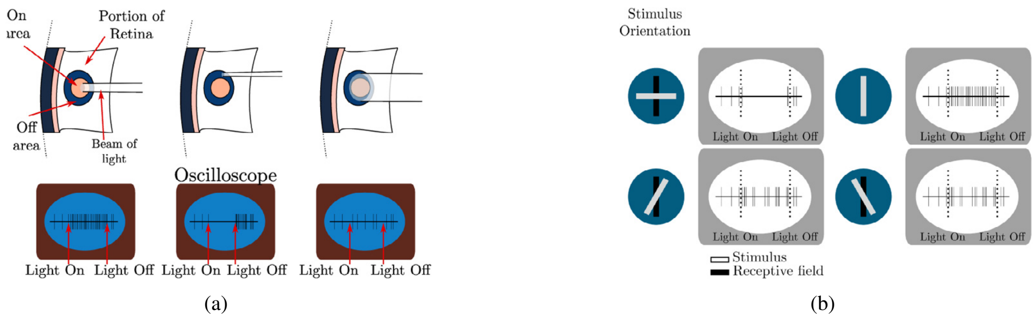


Fig. 3. The ganglion cell activation and lateral activation.

ganglion must receive information from multiple receptor cells. It was explained by Stephen W. Kuffler [13] that: ganglion cells have a slow rate of firing even in the dark; directing diffuse light to the retina has little effect on its rate, but a direct tiny spot falling on a small circular area on the retina or its perimeter can increase or inhibit ganglion cells fire rate; the light that shines on the retina and its perimeter at the same time produces no effect on the firing rate; finally, that other ganglion cells, have a central "off" surrounded by an "on" area. As a result, the optic nerve does not simply tell the brain that light has been detected, rather that contrast between light and dark (i.e. shape) has been detected, see Fig. 3(a).

Information is send to the Lateral Genticulate Nucleus (LGN), where it connects to a new set of inter-neurons. The information is then send to a specialized area in the occipital cortex called the visual cortex. This area was discovered by David Hubel and Torsten Weisel [14], by inserting electrodes in this area while projecting images into the eye of an animal instead of just a direct light, and observing the production of a firing rate (Fig. 3(b)). This experiment explains that even though cells in the LGN responded similar to ganglion cells, they no longer react to circles of light instead they did to either bars of light (or dark), or straight-line edges between light and dark areas. Specifically, these preferences for lines could be explained if they receive information from LGN cells (with circular response areas) arrange in a line (Fig. 3(b)). Additionally, "simple cortical cells" respond to particular area of the screen with a line at a specific angle, while another one will respond to a different position but same angle. It is then that another type of cell, "complex cortical cell", that information is joined using the orientation but with edges moving across the path [15].

Even though the same principle can be used to understand how curves can be detected (using neurons detecting predefined two dimensional shapes), there is a theory by Steven Zucker [16], where, instead of detecting point-wise the curve, it uses connections from semi-similar direction lines to detect the derivative of straight lines (tangent at a certain point). That by using simultaneous firing of these neurons, it would tune into the presence of a continuous curve rather than separate points. These theory can also explain how monocular depths can be detected (with the same tangential angle system) using the rate at which the angles change allows for depth perception in the absence of binocular vision.

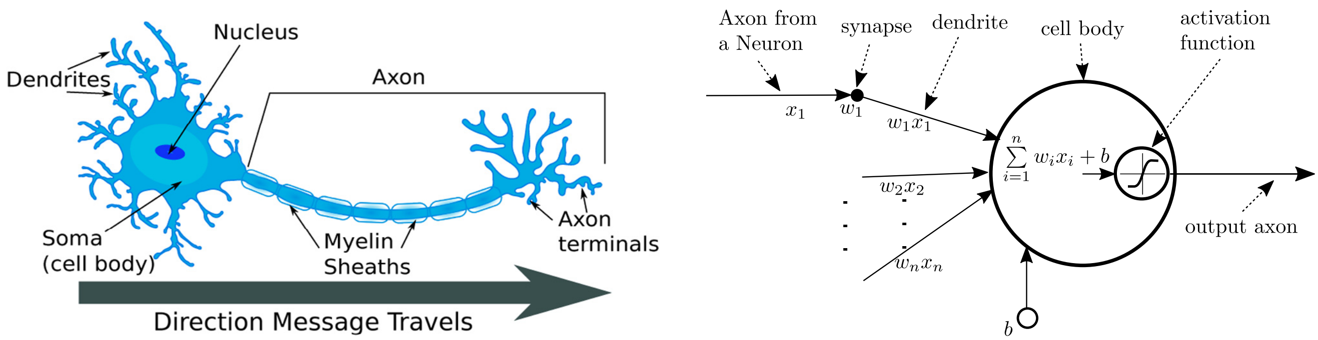
Another important area is the perception of motion, which is detected by different patterns of light in the retinal image. It is important to detect motion for deducting if something is moving in your direction (attract attention); to segment the foreground from background; to see $3D$ shapes and for self-localization in the space (navigation and collision avoidance). The problem is that it can be tricked easily with a simple light change (e.g. rotating spirals). The area responsible to motion is the motor track area. The neurons in this area are selective to velocity (speed and direction) receiving inputs from direction-selective neurons from V1. It can also be tricked by stimulating the area responding to movement, even though the perceive moment is in the opposite direction.

3. Computational neuroscience

Artificial Neural Networks (ANN) are models inspired in biological neurons. It is worth mentioning that the complexity of an ANN neuron is highly abstract, and the networks are still far from the real behavior. Yet, many things have been accomplished by using them and many theories have been created. For instance the first model of the neuron cell was proposed by Warren McCulloch and Walter Pitts [17] which showed that even simple types of networks could compute any arithmetic or logic function (Fig. 4). Comparatively, an ANN unit receives activation inputs (signals through the synapses to the dendrites), if the pulses are strong enough the activation function reacts (threshold is surpassed) as the unit sends the signal to connected units (signal passes through axon to the synapses), and with this mechanism it can understand patterns and remember sequences.

Later, Frank Rosenblatt developed the first successful neuro-computer called perceptron [18] that learn from examples to adapt its weights and that was improved with the work of Paul Werbos in 1974 [19] which adapted the weights based on the error gradient. From these theories many new architectures have been developed based on several areas of the brain, like Long–Short Term memory is based on the memory and Convolution Neural Networks on the visual system.

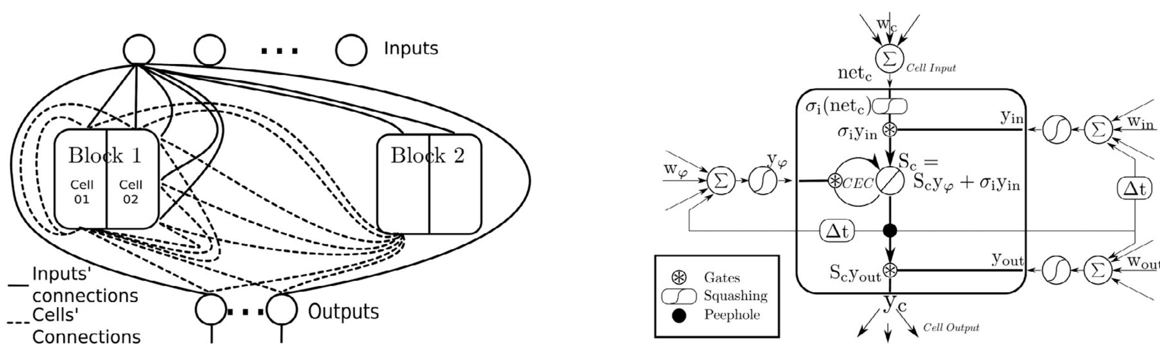
Long–Short Term Memory Networks (LSTM): are a type of ANN that feedback the information to every neuron. The feedback is used to provide some kind of memory that maps sequences. It uses units called *memory blocks* (Fig. 5(a)) which contain *memory cells* and share multiplicative gate units within



(a) Neuron.

(b) Neuron model.

Fig. 4. The basic model of an artificial neural network.



(a)

(b)

Fig. 5. LSTM with memory blocks in the hidden layer.

the block (Fig. 5(b)) and a recurrently self-connected units called "Constant Error Carousels" (CEC) whose activation is the cell state. The CEC enforces the error flow through them and with the gates it controls what is kept, what is left out and what is remembered by controlling the input, output and forgetting of the cells.

To illustrate each of the memories it can be seen all the gates are connected to all the current inputs and all the previous outputs, this could represent a basic short term memory; on the other hand, information from the previous states can be kept or erased (according to forget gate) and accessed (output gate) using the CECs, that would mimic a long term Memory.

Convolution Neural Networks (CNN): are a type of ANN that works by extracting local features at a high resolution and successfully combining them into more complex features at lower resolution. Notably, CNN are based on the work of Hubel and Wiesel [14] and try to mimic how some cells react to certain lines patterns, as well light and dark patterns. Additionally there were some other cells which detected edges no mattering where they were located.

In summary, CNN intention is to use filters to extract important information that is located in the input feature map (i.e. an image) by multiplying a weight matrix across the input feature maps, and get new maps with simpler information. These process can be repeated for several layers until it reaches a final section where there is a simplification of the input feature map, and from there the class can be calculated using other form of training (Fig. 6).

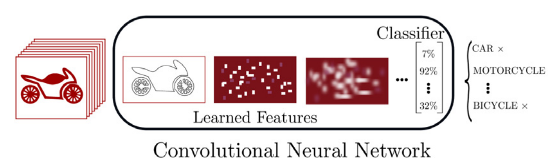


Fig. 6. A convolutional neural network.

4. Discussion

There exist some similitudes and differences between the brain and ANN. In the case of neurons, the brain changes its structure and connexions and how it reacts to each of them adapts its relationships. In the case of ANN, the network adapts its connexions only by changing the values of its weights. Also ANN tends to be densely connected almost all neurons from one layer are connected with the next, while in the brain there are only 0.0001% of connections. Comparing the memory, the brain apparently store information using differences in weights, and morphology, while LSTM can only remember by adapting its weights and which chose how much information will be feedback. Also, on one hand the brain apparently predicts by making comparison with past events even if they are not related, on the other hand LSTM has to have similar sequences to make a prediction.

Correspondingly, human vision uses layers to decompose an image into simpler representation, similar to what CNN does using filters to simple features to classify them. Conversely, the brain has different units for grays and colors that can speed up the process, while in CNN there is no difference in between all the cells. Also, brain can distinguish from objects moving, which is something that is missing from CNN without the use of other topologies.

Thus, some areas of opportunity can be determined:

- Adapt the morphology of the artificial neuron according to the information and not just the weights.
- Intrinsically trained for different tasks simultaneously.
- Make relationships based on previous assumptions.
- Memories change depending on the frequency of the input (high frequency store, low frequency erase).
- Different units depending on the input that speeds the process.
- Use differentiation to detect different shapes and movement.

5. Conclusion

The central theme of this article is reviewing the basis of visual and memory brain activation areas, use two techniques which are based on them, with the final intention of giving their biological foundation. Although, these two algorithms are showing great progress, as far as we understand they do not completely resemble their biological counterparts. While this might change in the future, at the moment, its not perfect, and still the brain has higher capabilities.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] J.J. Hopfield, Artificial neural networks, IEEE Circ. Dev. Mag. 4 (5) (1988) 3–10.
[2] Y. Dudai, The neurobiology of consolidations, or, how stable is the engram? Annu. Rev. Psychol. 55 (2004) 51–86.
[3] E.R. Hilgard, D.G. Marquis, Conditioning and learning, 1940.
[4] D.O. Hebb, The Organization of Behavior: A Neuropsychological Theory, Psychology Press, 2005.
[5] R.W. Semon, The Mneme, G. Allen & Unwin Limited, 1921.
[6] R.G. Crowder, J. Morton, Precategorical acoustic storage (pas), Perception Psychophys. 5 (6) (1969) 365–373.
[7] D. Kahneman, Method, findings, and theory in studies of visual masking, Psychol. Bull. 70 (6p1) (1968) 404.
[8] P. Saariluoma, Aspects of skilled imagery in blindfold chess, Acta Psychol. 77 (1) (1991) 65–89.
[9] A. Ramos, E.L. Schwartz, E.R. John, Stable and plastic unit discharge patterns during behavioral generalization, Science 192 (4237) (1976) 393–396.
[10] S.R. Cajal, A quelle époque apparaissent les expansions des cellules nerveuses de la moëlle épinière du poulet? Gustav Fischer, 1890.
[11] S.A. Josselyn, S. Köhler, P.W. Frankland, Finding the engram, Nat. Rev. Neurosci. 16 (9) (2015) 521.
[12] D. Purves, G.J. Augustine, D. Fitzpatrick, L.C. Katz, A.-S. LaMantia, J.O. McNamara, S.M. Williams, Plasticity of mature synapses and circuits, 2011.
[13] S. Kuffler, Discharge patterns and functional organization of mammalian retina, J. Neurophysiol. 16 (1) (1953) 37–68.
[14] D. Hubel, T. Wiesel, Receptive fields, binocular interaction and functional architecture in the cat's visual cortex, J. Physiol. 160 (1) (1962) 106–154.
[15] E. Kobatake, K. Tanaka, Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex, J. Neurophysiol. 71 (3) (1994) 856–867.
[16] S. Zucker, Differential geometry from the frenet point of view: boundary detection, stereo, texture and color, in: Handbook of Mathematical Models in Computer Vision, Springer, 2006, pp. 357–373.
[17] W. McCulloch, W. Pitts, A logical calculus of the ideas immanent in nervous activity, Bull. Math. Biophys. 5 (4) (1943) 115–133.
[18] F. Rosenblatt, The Perceptron, a Perceiving and Recognizing Automaton Project Para, Cornell Aeronautical Laboratory, 1957.
[19] P.J. Werbos, Beyond regression: new tools for prediction and analysis in the behavioral sciences, in: Applied Mathematics (Doctoral Dissertation), Harvard University, MA, 1974.