Artificial Neural Networks: a bio-inspired revolution Or a long lasting misconception and self-delusion

Emanuel Diamant
VIDIA-mant
Kiriat Ono, Israel
emanl.245@gmail.com

Abstract: Ali Rahimi, best paper award recipient at NIPS 2017, labelled the current state of Deep Learning (DL) headway as "alchemy". Yann LeCun, one of the prominent figures in the DL R&D, was insulted by this expression. However, in his response, LeCun did not claimed that DL designers know how and why their DL systems reach so surprising performances. The possible reason for this cautiousness is: No one knows how and in which way system input data is transformed into semantic information at the system's output. And this, certainly, has its own reason: No one knows what information is! I dare to offer my humble clarification about this obscure and usually untouchable matter. I hope someone would be ready to line up with me.

Keywords: Artificial Neural Networks (ANNs), Deep Learning Networks (DLNs), Information and information processing, Machine Learning and Machine Teaching.

1. Introduction

These days, we witness a tide of activity in Deep Learning Neural Networks (DLNNs) research and development. The press is proud to announce that five most valuable public companies (Apple, Google, Microsoft, Facebook and Amazon) are heavily relying in their new flag projects on the deep learning neural networks concept and capabilities [1].

Actually, the hype around DLNNs is not entirely unjustified: In recent years, they have turned out as a powerful tool for application in a variety of domains, such as language and speech modeling, image and video analysis, protein structure prediction, and so on [2].

However, despite their widespread and growing popularity, the operational principles of the DLNNs remain unclear and ambiguous. There is still little insight into their internal operations, and lack of knowledge on why and how they achieve their impressive performances. From a scientific point of view, this is entirely unsatisfactory. Without a clear understanding of how and why the DLNNs work, the development of better models inevitably reduces to trial-and-error experimentation [3].

Intended to cope with uncertainces in the DLNN research and development, we have to recall some primary concepts about the Artificial Neural Networks (ANNs). (There is no doubt that DLNNs are an evolutionary extension of the ANNs). Inspired by human brain functionality, ANNs were at first proposed by McCulloch and Pitts in 1943. Very soon after this, the Artificial Intelligence (AI) community has adopted them as a primary tool of brain-based human-level intelligence modeling. (Artificial Intelligence as a new research branch was announced at the Dartmouth College meeting in the summer of 1956).

Despite of its famous founding fathers (J. McCarthy, M.L. Minsky, N. Rochester, and C.E. Shannon) AI researchers have never been skilful enough to clarify what do they have in mind by putting in use the term "Intelligence". Without any further explanation, it was accepted that Intelligence is the human ability to cope with different cognitive tasks, that its location is in the brain, and that it is a product of activity of neuronal assemblies situated therein.

The lack of precise definitions had not bothered the ANNs developers and explorers. Very soon McCulloch and Pitts' model was superseded by a more powerful approach – In 1958 Rosenblatt had proposed the perceptron, an ANN that was able to learn from a given set of input-output examples (undeniably, learning is a cognitive ability trait). That spurs an era of new ideas and problems destined to be solved by the perceptron. However, at the end of the 1960s it became clear that, after all, the novelties are not so good, and new, more complex multilayered architectures are desired and looked-for.

Since then, the ANN research (tightly coupled with the AI research) has gone over many ups and downs. However, the central idea that the brain is an information processing engine and the ANN is a tool for modeling brain information processing abilities has never been challenged or objected [4], [5].

In this regard, it would be interesting to scrutinize to what extent the concept of ANN as an information processing device is exposed in the huge amount of papers and special publications devoted to ANN design issues. A Google search returns, as usual, an endless list of possible hits, but only few of them are really relevant – for an illustration purpose only I put some of them on the Reference list: [6], [7], [8], [9].

A striking feature common to all of them is that intended to deal with ANNs information processing issues they all are actually busy with ANNs data processing challenges. Any reservations or stipulations for that are not provided. Because that is a very common and a traditional mix-up by which the terms "information" and "data" are persistently swapped, replaced and used interchangeably.

This tradition is rooted in Shannon's "Mathematical Theory of Communication", published in 1948, where the notion of "Information" was initially introduced. The Shannon's theory was devised to improve communication systems performance and to assure an efficient and reliable message exchange (data exchange) over a communication channel. In such a context, the question "what is information" per se has never been asked and was irrelevant to the engineering problems under consideration. (Recall the known Shannon's quotation about meaning needlessness for a case of a data message transmission). The newly invented notion of "information measure" has served the design tasks pretty well. That led to a long lasting improper mixing and merging between notions of "information" and "information measure", which, in turn, made the relations between notion of "information" and notions of "data", "knowledge", and "semantics", undefined, blurred, and intuitive.

However, recent advances in almost all scientific fields put an urgent demand for an explicit definition of what information is. And the ANNs are not an exception in this regard. For that reason, we have to pause here and to scrutinize in depth the data/information exchanges, swaps and substitutions.

2. In quest of Information definition

There is a widespread conviction that a consensus definition of Information does not exist, and considering the multitude of its applications, it may be not even supposed to exist. I do not agree with this convention. On several occasions, I have already published my opinion on that subject [10], [11], [12]. This time, with all fitting excuses, I would like to repeat some fractions of these previous publications (in order to preserve the consistency of our discussion).

Contrary to the widespread use of Shannon's Information Theory, my research builds up on the Kolmogorov's ideas on information, [13]. In the mid-60s of the past century, Kolmogorov has proposed an algorithmic approach to a quantitative information definition [13]. According to Kolmogorov, a not random binary string (called a separate finite object) can be represented by a compressed description of it (produced by a computer program in an algorithmic fashion) "in such a way that from the description, the original message can be completely reconstructed" [14]. "The amount of information in the string is then defined as the size of the shortest computer program that outputs the string and then terminates" [14]. (For a really random string such a condensed description cannot be provided and "the shortest program for generating it is as long as the chain itself" [15]). The compressed description of a binary object has been dubbed as "algorithmic information" and its quantitative measure (the length of the descriptive program) has been dubbed as the description "Complexity".

Taking Kolmogorov's insights as a starting point, I have developed my own definition of information that can be articulated in the following way: "**Information is a linguistic description of structures observable in a given data set**".

To make the scrutiny into this definition more palpable I propose to consider a digital image as a given data set. A digital image is a two-dimensional set of data elements called picture elements or pixels. In an image, pixels are distributed not randomly, but, due to the similarity in their physical properties, they are naturally grouped into some clusters or clumps. I propose to call these clusters **primary or physical data structures**.

In the eyes of an external observer, the primary data structures are further arranged into more larger and complex agglomerations, which I propose to call **secondary data structures**. These secondary structures reflect human observer's view on the grouping of primary data structures, and therefore they could be called **meaningful or semantic data structures**. While formation of primary (physical) data structures is guided by objective (natural, physical) properties of the data, the subsequent formation of secondary (semantic) data structures is a subjective process guided by human conventions and habits.

As it was said, **Description of structures observable in a data set should be called "Information".** In this regard, two types of information must be distinguished – **Physical Information and Semantic Information**. They are both language-based descriptions; however, physical information can be described with a variety of languages (recall that mathematics is also a language), while semantic information can be described only by means of natural human language. (More details on the subject could be find in [16]).

Those, who will go and look in [16], would discover that every information description is a top-down-evolving coarse-to-fine hierarchy of descriptions representing various levels of description complexity (various levels of description details). Physical information hierarchy is located at the lowest level of the semantic hierarchy. The process of sensor data interpretation is reified as a process of physical information extraction from the input data, followed by an attempt to associate this physical information (about the input data) with physical information already retained at the lowest level of the semantic hierarchy.

If such an association is attained, the input physical information becomes related (via the physical information retained in the system) with a relevant linguistic term (see also the block diagram in [10]), with a word that places the physical information in the context of a phrase, which provides the semantic interpretation of it. In such a way, the input physical data object becomes named with an appropriate linguistic label and framed into a suitable linguistic phrase (and further – in a story, a tale, a narrative), which provides the desired meaning for the input physical information. (Again, more details can be found on the website [16]).

3. If that is Information, how ANN is supposed to handle it?

Because no one knows what information is, it is fair to assume that no one knows how to handle it in a right manner. For that reason, I would like to draw your attention to a hypothetical brain visual information processing architecture, which I have proposed about 10 years ago [10]. Considering this hypothetical information processing architecture as a reference, it will be more handy to us to launch a preliminary study of the ANN (DLNN) information processing abilities and to explore to what extent the ANNs are suitable to carry out information processing tasks. After all, modeling the brain means to reveal the way in which it is processing information. And that is what we are trying to investigate in this part of the paper: the similarities and dissimilarities between two brain inspired cognitive architectures, the ANN (DLNN) and the Visual Information Processing arrangement.

What falls first in the eyes is the hierarchical multi layered architecture common to both of the systems. Although the early versions of ANNs were designed as a single-layer implementation, contemporary DLNNs are full-blown multi-layered accomplishments (for example, GoogLeNet is a 22 layer CNN [17]).

Both systems receive volumes (vectors) of raw sensor data at their input and this data has to be processed and transformed into semantic information description, which is the output of an information processing system, or into cognitive object labels that are the output of a DLNN system. It is self-understood that object labeling is equivalent to a brief semantic description.

There is a slight problem (in our discussion) with finding appropriate comparable terms corresponding to those which are being used in these two different, although compatible, domains – the ANNs application and the Information theory application. The two have evolved as totally independent fields of knowledge with their own curriculum vitae, specific dogmas, and particular languages. Therefore, notions very close in their meaning are expressed in the two domains with different words and terms, which have to be carefully interrelated and adjusted in the course of our discussion. I hope that by paying attention to these differences we would be able to manage this problem.

So, to this end, the purpose of the first input layer is common to both scrutinized architectures: to reveal and disclose data structures hidden in the input feed. If we restrict ourselves to the case of image information processing, we can say that at this occasion, the input data is image pixels, and the purpose of the first layer is to reveal the hidden pixel structures present in the input image.

As it was explained above, in the case of image information processing, the low-level pixel structures discovery is dubbed as primary or physical data structures discovery. The process is finalized by an immediate structure description production, which is the physical information about the low-level pixel bundles. This physical information is advanced to the next processing layers for further treatment.

In contrast to this, the DLNN architectures gradually (that is, in a level-by-level fashion) create a whole set of structures: low-level, middle-level, and high-level structures [18].

Despite of the widespread use of the term "information", ANN-DLNN designers know nothing about information. Evidently, they do not know about information duality, the complex nature of Physical and Semantic Information. They do not know that higher-level semantic structures are created first by grouping primary structures into secondary structures, and then regrouping secondary structures into even more complex arrangements and compositions. They do not know that the primary structures aggregation into secondary structures is a process performed solely by an observer, who is watching and scrutinizing the image. And all this is an observer's exceptional privilege and duty. And any other way to arrange primary structures into a meaningful secondary structure does not exist. (That is the reason why we call the secondary structures semantic structures and their description we call semantic information)

Now, it becomes clear why DLNN designers experience such enormous difficulties moving from low-level structures to middle-level and high-level structures. In ANN-DLNN jargon they are called "data features" and along all the processing path they are treated routinely as data. However, middle-level and high-level data features are in fact semantic structures (semantic objects and proto-objects), and therefore deserve a different way of treatment.

A lot of problems causes the request for invariance of location, rotation, and scale properties of data features. Huge efforts are spent by ANN-DLNN designers to fulfil this requirement. At the same time, Information processing approach is free from these difficulties because semantic information processing is busy with physical information processing (interpretation). Physical information by definition is a **description of structures** observable in a given data set. Original data features are become dissolved in the description and do not take part in the image understanding process.

It is exactly for that reason:
- We get the meaning of a written word irrelevant to the font size or the style of the letters.
- We recognize equally well a portrait of a known person on a huge-size advertising billboard, on a magazine front page, or on a postage stamp – perceptual information is dimensionless, (while data features are not).
- We get the meaning of a scene irrelevant to its illumination. We look on the old black-and-white photos and we do not perceive the lack of colors.
- The same is true for voice perception and spoken utterance understanding – we understand what is being said irrelevantly to who is speaking (a man, a women, a child) and irrelevant to the volume levels of the speech (loudly or as a whisper).
- Blind people read Brail-style writings irrelevant to the size or the form, or the warmth of the touch-code.

The process of assigning semantic labels to data objects is performed by an information processing system in a following way: a group of primary physical structures revealed in the input data is attempted to be associated with a similar group of physical structures (a known secondary structure) retained in the system knowledge base. (System's knowledge base is simply another name for the system's information hierarchy, with levels of semantic information at its upper part and levels of physical information at its lower part). If the similarity with an associated secondary structure is resolved, the group composition is become fixed and the group as a whole is branded with a label borrowed from the associated secondary structure, which is retained (memorized) in the system's knowledge base. Under this label, the physical information group (transformed into a secondary information item) is inserted into a semantic structure located at the next, higher level of semantic information hierarchy, where it becomes a part of a sentence, a phrase, a linguistic expression. That is the way, in which input data objects become detected and delineated, and branded with their semantic labels.

Then, as a very important issue comes here the question of the system knowledge base origin. As it was explained earlier, the creation of secondary structures is an entirely subjective process. What primary structures are chosen to be composed into a secondary structure and what primary structures interrelations endorse such a composition is a matter of observer's subjective habits and preferences. His pacts and agreements with other observers, which are involved in the image viewing process. For that reason, system's semantic information hierarchy – in other words, system's reference knowledge base – cannot be autonomously established (acquired, learned, done – take what seems to you as the most suitable title). System's semantic information hierarchy must be always supplied, delivered, granted from the outside. And that is a trap for all "learning" theories and "training" practices. The acquisition of a new semantic hierarchy for the purposes of information processing system would be more appropriate to call a "teaching process" rather to call it a "learning process".

ANN-DLNN designers know nothing about semantic information and semantic information hierarchy usage. They have developed their own rules and methods how to assign semantic labels to given data patterns and data objects. The process is called ANN model training mode and it is implemented in the following way: A set of human selected and labeled images containing an intentionally chosen visual object is submitted to a given ANN-DLNN system. The system extracts from each image a set of low-level, middle-level, and high-level features, which are then combined and merged to produce a composed object representation. Different input images are processed in such a way that their representations are sufficiently close between them. The finally achieved representation is tied with the label under which the training images were provided.

ANN-DLNN designers don't know that creation of a representative feature composition is a subjective observer dependent task, and they are trying hard to formalize and computationally support this design goal. But, it can not be done this way. Subjective semantic descriptions could not be formalized. What follows from this – the ANN-DLNN design assumptions are wrong. The system will never reach its goals.

4. Conclusions

The ANN-DLNN were conceived to model human brain information processing. What information is and what does it means to process it – such questions have never been asked even from the very beginning of the ANN era. The things do not changed too much since then. Today, the term "information" also pops up sporadically here and there, merely in the titles of some publications [19]. In the rest of occasions, the term "information" is used interchangeably with the term "data". Of course, that leads to omnipresent confusion and unresolved misunderstanding. For example, at the system input we have only raw data, however, at the system output we have already labeled semantic objects. How the system achieves that? – nobody knows. There is a common agreement that this is the weakest point of the ANN R&D exploration.

The unpredictable way, in which ANN receives its results, led even to a fierce argument exchange between two very respected figures in ANN research community – Ali Rahimi and Yann LeCun [20]. In his award talk at NIPS 2017, Ali Rahimi has declared that "**the current practice in machine learning is akin to "alchemy**". Yann LeCun has responded in his Facebook page: "Criticizing an entire community (and an incredibly successful one at that) for practicing "alchemy", simply because our current theoretical tools haven't caught up with our practice is dangerous". And citing further: "It's insulting, yes. But never mind that: **It's wrong!"**

In his response, LeCun does not deny and, on the contrary, confirms that "our current theoretical tools haven't caught up with our practice". LeCun, essentially, approves that ANN-DLNN research has a critical lack of understanding of how input data feed translates itself into a persistent information flow through all system levels, and finally appears at the system output as a labeled semantic object. Alchemy is not the worst word for such a wonder.

The purpose of this paper was to make obscure issues of the ANN-DLNN research a little more lucid and comprehensible.

References

[1] Jurgen Schmidhuber, **Falling Walls: The Past, Present and Future of Artificial Intelligence,** Scientific American, November 2, 2017. https://blogs.scientificamerican.com/observations/falling-walls-the-past-present-and-future-of-artificial-intelligence/

[2] Klaus Greff, *et al*, **LSTM: A Search Space Odyssey,** Transactions on Neural Networks and Learning Systems, Volume: 28, Issue: 10, Oct. 2017, page(s): 2222 – 2232, http://arxiv.org/abs/1503.04069

[3] M. D. Zeiler and R. Fergus, **Visualizing and Understanding Convolutional Networks,** http://arxiv.org/abs/1311.2901

[4] Nikolaus Kriegeskorte, **Deep neural networks: a new framework for modelling biological vision and brain information processing,** Bio-archive, October 2015, https://www.biorxiv.org/content/biorxiv/early/2015/10/26/029876.full.pdf

[5] Giulio Ruffini, **Information, complexity, brains and reality (Kolmogorov Manifesto)**, https://arxiv.org/abs/0704.1147

[6] Ravid Shwartz-Ziv and Naftali Tishby, **Opening the Black Box of Deep Neural Networks via Information**, https://arxiv.org/abs/1703.00810

[7] Giacomo Indiveri and Shih-Chii Liu, **Memory and information processing in neuromorphic systems,** https://arxiv.org/abs/1506.03264

[8] Alexander Alemi, *et al*, **An Information-Theoretic Analysis of Deep Latent-Variable Models,** https://arxiv.org/abs/1711.00464

[9] Joel Zylberberg, *et al*, **Robust information propagation through noisy neural circuits,** https://arxiv.org/abs/1608.05706

[10] Emanuel Diamant, **Unveiling the mystery of visual information processing in human brain,** *Brain Research*, vol. 1225, pp. 171-178, 15 Aug. 2008, https://arxiv.org/abs/0807.0337

[11] Emanuel Diamant, **The brain is processing information, not data. Does anybody care?,** ISIS Summit Vienna 2015, http://sciforum.net/conference/isis-summit-vienna-2015/paper/2842

[12] Emanuel Diamant, **Shannon's definition of information is obsolete and inadequate. It is time to embrace Kolmogorov's insights on the matter**, Conference Paper, November 2016. https://www.researchgate.net/publication/311223095

[13] A. N. Kolmogorov, **Three approaches to the quantitative definition of information**, Problems of Information and Transmission, Vol. 1, No. 1, pp. 1-7, 1965. http://alexander.shen.free.fr/library/Kolmogorov65_Three-Approaches-to-Information.pdf

[14] Peter Grunwald and Paul Vitanyi, **Algorithmic Information Theory**, 2008. http://arxiv.org/pdf/0809.2754.pdf

[15] Peter Grunwald and Paul Vitanyi, **Shannon Information and Kolmogorov Complexity**, 2004. http://arxiv.org/pdf/cs/0410002.pdf

[16] Emanuel Diamant, **Brain, Vision, Robotics and Artificial Intelligence**. http://www.vidia-mant.info

[17] Adit Deshpande, **The 9 Deep Learning Papers You Need To Know About,** https://adeshpande3.github.io/adeshpande3.github.io/

[18] Zeiler, M. D., et al., **Adaptive deconvolutional networks for mid and high level feature learning,** In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 2018–2025). Red Hook, NY. http://www.matthewzeiler.com/wp-content/uploads/2017/07/iccv2011.pdf

[19] Samuel S. Schoenholz, et al., **Deep Information Propagation**, International Conference on Learning Representations ICLR 2017, https://research.google.com/pubs/pub45931.html

[20] Yann LeCun, **Response to Ali Rahimi's NIPS lecture** , December 6, 2017 https://www.facebook.com/yann.lecun/posts/10154938130592143