

Towards a Quantum Field Theory for Optical Artificial Intelligence

Antonio Manzalini

TIM, Via Reiss Romoli, 274 – Turin, Italy
antonio.manzalini@telecomitalia.it

Received: 18th April 2019; Accepted: 13th May 2019; Published: 1st July 2019

Abstract: Today, several socio-techno-economic drivers are steering the evolution of Telecommunications and Internet towards a growing exploitation of ultra-broadband infrastructures (e.g., 5G) and Artificial Intelligence (AI) systems. Focusing on the most promising AI technological approaches, Deep Neural Networks (DNNs) are outperforming in several applications domains. One of the possible explanations, elaborated in literature, is that DNN functioning is deeply rooted in the principles of theoretical Physics, specifically Quantum Field Theory (QFT) and Gauge theory. This is encouraging even more researches and experiments in the direction of a full exploitation of quantum computing and networking for the development of innovative Information Communication Technologies (ICT) and AI systems. In this innovation avenue, given that QFT and Gauge theory have been already proposed for modeling the brain and biological nervous systems, this paper explores the intriguing possibility of exploiting QFT principles also for future DNN, for instance by using electromagnetic waves effects in metamaterials. This appears to be a promising direction of future studies and experiments: therefore, the paper also describes the architecture of a simple optical DNN prototype, based on metamaterials, which is intended as a live test-bed, for simulations and experiments.

Keywords: *Artificial Intelligence; Deep Learning; Deep Neural Networks; Gauge Theory; Quantum Field Theory*

1. Introduction

The transformative role of Computer Science and Information Communication Technologies (ICT) has long been witnessed as a precursor of the scientific progress and economic growth in the modern world. Today, like never before, we are witnessing an increasing exploitation of ultra-broadband infrastructures and Artificial Intelligence (AI) systems, for providing any sort of digital services for our daily life.

The so-called Digital Transformation is going to create a pervasive artificial nervous system [1] embedded into the reality around us. As a matter of fact, a very large adoption of mobile terminals and devices and a pervasive distribution of the Internet of Things will allow collecting enormous big data sets, which will be communicated and transmitted by ultra-broadband, low latency network connections (e.g., 5G) [1] to Cloud and Edge Computing Data Centers (DCs). These DCs, in turn, will

store and elaborate the collected big data in order to infer decisions (with Big Data analytics and AI methods) which will be then exploited, locally, by actuators. This is the typical closed loop of any autonomic or artificial nervous system (e.g., collect-analyze-decide-act). Therefore, big data analytics and Artificial Intelligence (AI) will be at the center of this transformation.

AI is not a new story, of course, but today the socio-techno-economic drivers of the Digital Transformation are creating the boosting conditions for its wider and wider adoption. Among these drivers, in fact, there are: pervasive ultra-broadband connectivity, an increase of performance of IT systems, the lowering costs of hardware, the integration of AI in the management of 5G telecommunications infrastructures, and a plenty of AI services and applications for end Users (e.g. for facial recognition, digital assistant services, web searches or for driving a vehicle).

One of the most demanded task of AI is extracting patterns and features directly from collected big data. Recognizing patterns is the main job of any Neural Network (NN), be it based on digital or quantum technologies. As known, NN are taking inspiration by the human brain architecture: they are like a grid/network of basic computing units, i.e., the neurons. Among the various most promising approaches for accomplishing this goal, Deep Neural Networks (DNNs) [2] are outperforming.

DNNs are biologically-inspired statistical models based on multiple and hierarchical layers of said neurons. It is still unclear the technical reason why DNNs shows this unprecedented capabilities over shallower architectures with a similar number of parameters. One possible explanation, elaborated in literature, is that a DNN is like an iterative coarse-graining scheme. Each high-level layer learns increasingly abstract higher-level features, providing a useful, and at times reduced, representation of the data.

Interestingly, successive coarse-graining reminds an important tool in theoretical physics, called Renormalization Group (RG) [3], used to solve difficult physics problems involving many-length scales. This, and other analysis reported in literature, are suggesting the intriguing possibility that DNN principles are deeply rooted in the theoretical physics.

This paper elaborates in this direction, resuming from the basic principles of Quantum Field Theory (QFT) and Gauge theory, which represent an overall mathematical framework at the basis of the Standard Model. Today, QFT and Gauge theory represent the best recognized candidate framework to bring to the unification of the fundamental four interactions (i.e., gravitational, electromagnetic, strong, and weak). Moreover, and this is even more intriguing, QFT and Gauge theory principles seems to be valid also in other fields, even biology.

Importantly, the common element, share by all these fields, is a problem formulation in terms of free energy. Given also that the Gauge theory has been already concretely proposed for modeling biological nervous systems [4], it seems promising to explore if these principles could be further extended for designing AI systems exploiting quantum effects.

In summary, the main contributions of the paper include:

1. the idea of exploiting QFT and Gauge theory mathematical framework, which has been already adopted for modelling biological brain and nervous systems [4], also for the design and development of future DNN and AI systems. In common there is the formulation in terms of free energy principle;
2. to highlight that, as in the gauge theoretical model of brain or nervous system, the manifolds of the least free energy path are curved; given that, a design of DNN should consider the problem of moving geometric objects, such as feature vectors and kernels, across curved manifold;
3. to show the importance of enhancing the definition of the convolutional kernel (e.g., using constraints and/or Fourier transformations) to take account of curved manifolds, as in Convolutional Neural Networks (CNNs);
4. the proposal of an electromagnetic DNN architecture, based on metamaterials, which is to be considered as a live test-bed, for next simulations and experiments.

The paper is structured as follows. After the introduction, section two describes the basic principles of Restricted Boltzmann Machine-based DNNs (RBM-based DNNs), as their structures are showing strict similarities with the RG. The third section provides an overview of the basic principles

of QFT and Gauge theory and some considerations about how adopting them for DNN design. Sections four describes the high-level architecture and the design principles of an optical DNN prototype based on metamaterials. Conclusions and perspectives on future works are closing the paper.

2. RBM-based DNN

DNNs are biologically-inspired statistical models based on multiple and hierarchical layers of neurons. A DNN has at least two layers which are hidden, where the neurons of one layer receiving inputs from ones in the layer below.

Most used DNNs have up to 50 layers, but there DNNs, for high performance applications, that can reach even one hundred of layers. The number of layers is just one of the characteristics of the complexity of a DNN; others are the numbers of neurons, the numbers of the connections and the related weights.

There are two ways for training a DNN: supervised and un-supervised. The former class includes models such as Convolutional Neural Networks (CNN), Fully Connected DNN (FC DNN), Hierarchical Temporal Memory (HTM). The latter class includes models such as Stacked Auto-Encoders, Restricted Boltzmann Machine (RBM) and Deep Belief Network (DBN).

This section is focusing on RBMs as their structures is showing strict similarities with the RG. A RBM is a two-layer network with stochastic units bidirectionally connected, as shown in figure 1.

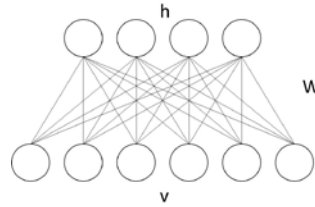


Figure 1. RBM as a bipartite graph.

The V units of the first layer, denoted by the vector $v = (v_1, \dots, v_V)$, correspond to the components related to an observation and are therefore called “visible”, while the H units in the second layer $h = (h_1, \dots, h_H)$ represent “hidden” variables. The term “restricted” refers to the connections between the units: each unit is connected with all hidden ones, but there are no connections between units of the same kind. w_{ij} is the symmetric connection weight between the units.

In analogy to spin models¹ in statistical physics, a RBM models the configuration of visible and hidden units (v, h) in terms of an energy function:

$$E(v, h) = \sum_i b_i v_i + \sum_i c_j h_j + \sum_i w_{ij} v_i h_j \quad (1)$$

A RBM assigns a probability to each joint configuration (v, h), which by convention is high when the energy of the configuration is low (Gibbs distribution):

$$p_\lambda(v, h) = \frac{e^{-E(v, h)}}{Z} \quad (2)$$

where $\lambda \equiv \{b_i, c_j, w_{ij}\}$ are variational parameters and Z the partition function:

$$Z = \sum_{v \in V} \sum_{h \in H} e^{-E_\lambda(v, h)} \quad (3)$$

It has been mentioned that one possible explanation of the unprecedented capabilities of a DNN is like an iterative coarse-graining scheme, e.g., RG [3]: each high-level layer learns increasingly abstract higher-level features, thus providing a useful, and at times reduced, representation of the data. As a

¹ Spin models are used in many studies of complex systems characterized many interacting degrees of freedom: for example, in condensed matter physics, neural networks, or economics. A spin model is specified by a set of degrees of freedom, the spins, and a cost function, or Hamiltonian, which specifies the interaction pattern as well as the type and strength of interactions among the spins.

matter of fact, in a DNN each high-level layer learns increasingly abstract higher-level features, providing a useful, and at times reduced, representation of the data.

These considerations, and others reported in similar literature, are suggesting an intimate connection between RG and a RBM-based DNN. For instance, [5] provides a demonstration that there is a one-to-one mapping between RBM-based DNN and the variational RG. Also [6] makes the conjecture that there might be correspondence between QFT and RBM-based DNN.

This is, moreover, implying even the surprising possibility that the quantum fields, and hence matter (living and not-living) in the universe, can store and elaborate information to pursue different levels of self-organization and cognition.

3. Quantum Field Theory and its relations with Deep Learning

Quantum Field Theory (QFT) is recognized as a fundamental theory of quantum physics for two main reasons: the former concerns its countless empirical and convincing demonstrations; the latter is because it provides a re-interpretation of Quantum Mechanics (QM), overcoming several of its limitations.

QFT is based on a new point of view in quantum physics, where the particles are excited states of underlying fields and as such, they are openly connected to their environment (termed vacuum). The concept of field is rather intuitively well-known: a field is defined as a property of spacetime which could be represented by a scalar, a vector, a complex number, etc. So, in QFT, an excited state of a field is any state with energy greater than the ground state, called quantum vacuum. In other words, QFT, any quantum system is an “open” system, as it is always interacting with the background fluctuations of the quantum vacuum.

As mentioned it seems that there is a relationship between QFT and DNN. This is elaborated in [6], where the Author explains it also through the correspondence of the Euclidean QFT in a flat spacetime with $d+1$ dimensions and the statistical mechanics in $d+1$ dimensional flat space using an imaginary time. Further technical details are outside the scope of the paper.

3.1. The Gauge Theory

A Gauge theory is a type of theory in physics. The Gauge theory, within the context of the QFT and the Standard Model, describes interactions and energy exchanges between quantum particles and their associated wave fields [7].

Field theories are usually defined in terms of a Lagrangian (defined as the difference of the kinetic and the potential energy). In a Gauge theory the Lagrangian is kept invariant under continuous symmetry transformations (called gauge transformation), thanks to the presence of the gauge fields which are sort of compensatory fields (with the related Gauge bosons) able to maintain the symmetry. For example, the Einstein’s General Relativity is a ‘gauge theory of spacetime’.

Today, Gauge theories are used in physics, mathematics, economics, finance and even in biology [8]. Common to most contexts there is a formulation in terms of the free energy principle. In biology and neuroscience, the free-energy principle proposes a unified account of action, perception and learning by arguing that any self-organizing system - in equilibrium with its environment - must minimize its free energy, by solving an optimization problem.

In fact, variational free energy was also introduced by R. Feynman [9] into statistical physics to convert difficult probability-density integration problems into easier optimization problems, and to solve difficult problems in path integral formulations of quantum physics. Interestingly, it has been already exploited in machine learning and statistics to solve many inference and learning problems [10].

In general, it can be said that the minimization of the variational free energy states that a system changes its internal states so that they become the most probable under the influence of the environment external states. In doing so, the system is implicitly learning a generative model of the environment: the better is the model in making predictions the lower is the variational free energy.

In [11] and [12] the free energy principle has been proved valid for describing several aspects of functional brain architectures. In this case the system is the brain (with neuronal or internal states), while the environment (with external states) is equipped with continuous forces and produces local sensory perturbations that are countered through action and perception (i.e., functionals of the gauge field). Concerning the brain's symmetries, it should be mentioned that computational neuroscience community is aiming at providing ways to model symmetry of the network architectures of the connectome, for instance by computing the eigenvalues of different types of connectivity matrices. In [13] the brain symmetry is given the rotational symmetry of the electric dipole of the vibrational field of water molecules.

Therefore, minimizing variational free energy ensures the brain becomes a good model of its environment: this has again a striking correspondence with the concept of "double" described in terms of QFT in [14]. The interactions of a system with its environment determines the doubling of the collective modes of the overall ensemble, i.e., the "system-environment", as the energy flux balance is preserved. In this dissipative model of the brain [14], an external stimulus is responsible for spontaneous symmetry breaking (SSB) thus generating Nambu Goldstone (NG) bosons. We recall that a SSB is a dynamical transformation in which a system moves from a symmetrical state to a state with a lower level of symmetry, and as such a higher level of order. NG bosons are massless and collective vibrational modes coding such order. They do not mediate energy interactions, rather they represent modalities of such interactions. Differently from Gauge bosons, the NG bosons disappear when the coherent state of matter vanish (e.g., phonons in crystals), as there is no energy conservation principle they must obey.

Starting from above considerations, this work has been further motivated by these following two research questions:

- i) if the brain or the nervous systems [4] can be modelled with a QFT/Gauge theory mathematical framework, then it is possible to extend this approach for the design of future DNN / AI systems?
- ii) if, thanks to the QFT/Gauge theory, the minimization of the variational free energy can be transform into a path integral problem on curved manifolds, then it is possible to develop a generalized design model of DNN on curved manifolds?

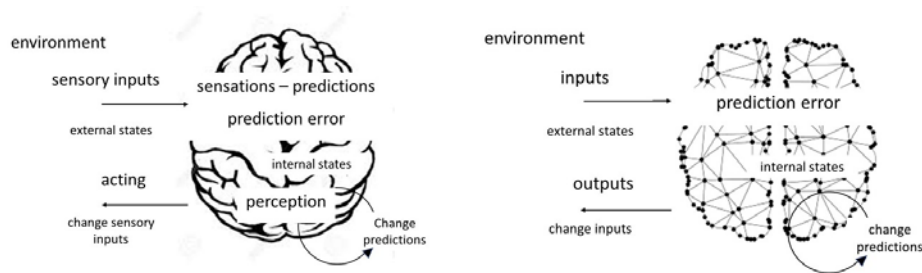


Figure 2. Minimization of free energy for a brain (on the left) and for an ideal AI system (on the right).

3.2. Adopting the Gauge Theory for Deep Learning desing

The design of DNNs for specific applications often leverages on intuition and empirical approaches. Several design principles have been proposed and adopted depending on the specific problem to be solved. Equivariance to symmetry transformations is one of the more general design principles which has brought to the development of DNNs solutions showing excellent levels of performance.

Some remarkable examples of adoption of the equivariance principle are provided by the cases of CNN [15]. In this context, [16] show how the equivariance principle deals with curved manifolds.

We've already mentioned that in the modeling of biological nervous systems with a Gauge theoretic approach [4], the problem of minimizing of the variational free energy can be formulated in terms of a path integral over curved manifolds. Let's analyze this aspect in more details.

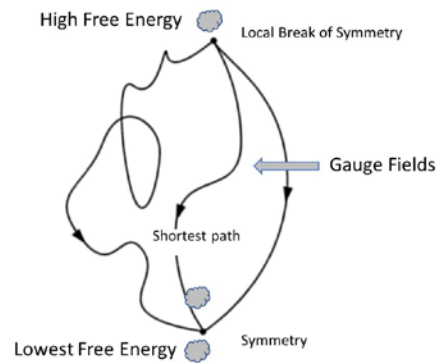


Figure 3. Minimisation of free energy in terms of solving a path integral problem.

When designing a convolution on a manifold, the geometric objects such as feature vectors and kernels should be transported across the manifold. This has a main consequence: the manifold's curvature makes the geometric objects path-dependent, forcing us to consider different local frames.

Any change between local frames is sort of a local gauge transformation implemented by the filters between an input feature map and an output feature map. This is a symmetry breaking, whilst we are looking for a general principle maintaining the symmetry, the equivariance as a design criterion.

We recall that according to a Gauge theoretic approach, to maintain a global symmetry, gauge fields are compensating the broken symmetry by the minimization of the variational free energy. This can be translated in the context of CNN in these terms: to maintain equivariance of the output feature maps (under local transformations), certain constraints must be placed in the convolution kernel [17], mimicking the effects of Gauge fields.

The conclusion is that we need upgrading the convolution kernel with terms capable of taking account of the curvature transformations in the manifolds. This recalls the fact that the distortion of the electromagnetic fields in a metamaterial can be represented as a coordinate (i.e., frame) transformation: these features provide the interesting opportunity of developing electromagnetic convolution kernels over curved manifolds. In fact, Maxwell's equations don't change in any coordinate system, but the refractive index, which means the permittivity and permeability factors which are scaled by a common factor.

The next section describes a proposal of an optical DNN architecture, based on metamaterials, which is to be considered as a live test-bed, for next simulations and experiments.

4. Design and implementation of a prototype

Different technological approaches have been analyzed for designing and implementing a prototype of optical DNN exploiting the principles elaborated in this paper, and in the quoted references.

Among the different approaches, all-optical Diffractive Deep Neural Network (D²NN) architectures [18] appear quite interesting: in particular, D²NN are based on wave communication mechanisms between the layers of a deep network.

Another promising approach regards the use of 3D topological insulators. For instance, in [19] these systems are made of sheets embedding metal nano-antennas acting as split-ring resonators. Said resonators should be designed and tailored so that they interact with electromagnetic waves in specific ways, which allows them to give the structures the topological characteristics they require.

In a similar direction, [20] is proposing the use of metamaterials for analogue computations. Main examples include the use of thin planar metamaterial blocks for making mathematical operations (e.g., convolutions) in the spatial Fourier domain, and the use of multilayered metamaterial slabs to realize desired Green Functions. We recall that metamaterials are artificial materials that owe their properties to subwavelength structures rather than to their chemical composition. Metamaterials are made of periodic oscillators developed with artificial metallic or

dielectric structures. Today it is possible to develop metamaterial with permittivity and permeability designed to change independently and arbitrarily throughout a material, with positive or negative values.

This possibility of manipulating electromagnetic fields in various ways is making metamaterials a very promising direction for designing and developing a prototype of an optical DNN system exploiting the principles described in this paper. For instance, just consider that the distortion of the electromagnetic fields in metamaterials can be represented as a coordinate transformation in a curve manifolds, which can be used for implementing the equivariance principle.

The architecture under study and design is shown in figure 4.

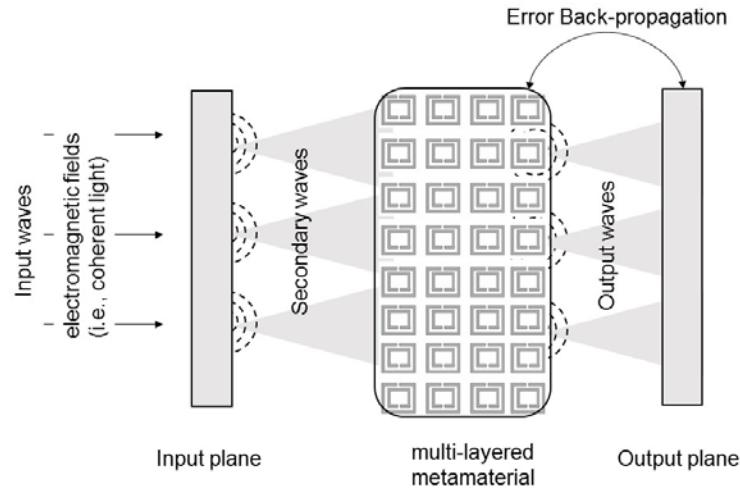


Figure 4. High level description of an optical DNN prototype based on metamaterials.

Some characteristics which make this prototype different with respect to a traditional DNN are:

- the inputs are electromagnetic waves (e.g., coherent light) which bring us dealing with complex numbers (rather than real ones);
- the function of any neuron is expression of wave interference phenomena and electromagnetic interactions with the metamaterial oscillators (rather than sigmoid, linear or non-linear neuron functions);
- coupling of neurons is based on the principles of wave propagation and interference.

Future publications will address ongoing simulations and guidelines for the technical developments of the prototype.

5. Conclusions and perspectives on future work

Today, like never before, we are witnessing an increasing exploitation of ultra-broadband infrastructures (e.g., 5G) and AI systems, for providing any sort of digital services for our daily life. This profound socio-economic Digital Transformation is creating a sort of pervasive artificial nervous system embedded into the reality around us. This will have a profound impact on our economy, society and culture, being as a precursor of a new scientific progress and socio-economic growth.

In this scenario, technology innovation is accelerating so it appears reasonable to imagine that, even in the medium term, quantum computing technologies may offer the opportunity of developing new and innovative ICT and AI systems outperforming the current digital ones. This is also supported by the fact that several studies are arguing the intriguing possibility that DNN/AI principles are deeply rooted in the theoretical quantum physics.

As a matter of fact, QFT and Gauge theory mathematical framework has been already proposed for modeling biological brain and nervous systems. In this direction, this work has explored the

possibility of extending QFT and Gauge theory mathematical framework also for designing and developing future DNN/AI systems based on electromagnetic waves effects.

This appears a promising direction of study as DNN operations can be performed at the speed of light through electromagnetic signals interfering and propagating through metamaterials structures. The paper concluded by describing the architecture of a simple optical DNN prototype based on metamaterials, for future modelling and experimental validations.

References

- [1] D. Soldani, A. Manzalini, "Horizon 2020 and beyond: on the 5G operating system for a true digital society." *IEEE Vehicular Technology Magazine* 10.1 (2015): 32-42.
- [2] G. Hinton, Ruslan R. Salakhutdinov, "Reducing the dimensionality of data with neural networks." *science* 313.5786 (2006): 504-507.
- [3] W. Kenneth G., J. Kogut, "The renormalization group and the ϵ expansion." *Physics reports* 12.2 (1974): 75-199.
- [4] B. Sengupta, et al., "Towards a neuronal gauge theory." *PLoS Biology* 14.3 (2016): e1002400.
- [5] M. Pankaj, D. J. Schwab, "An exact mapping between the variational renormalization group and deep learning." *arXiv preprint arXiv:1410.3831* (2014).
- [6] JW. Lee, "Quantum fields as deep learning." *arXiv preprint arXiv:1708.07408* (2017).
- [7] O. Lochlainn, N. Straumann, "Gauge theory: Historical origins and some modern developments." *Reviews of Modern Physics* 72.1 (2000): 1.
- [8] K. Friston, et al., "Knowing one's place: a free-energy approach to pattern regulation." *Journal of the Royal Society Interface* 12.105 (2015): 20141383.
- [9] R. P. Feynman. *Statistical Mechanics, A set of lectures.* Frontiers in Physics. Perseus Books (1972).
- [10] G. Hinton, D. Van Camp, "Keeping neural networks simple by minimizing the description length of the weights." in *Proc. of the 6th Ann. ACM Conf. on Computational Learning Theory*. 1993.
- [11] K. Friston, "Hierarchical models in the brain." *PLoS computational biology* 4.11 (2008): e1000211.
- [12] K. Friston, "The free-energy principle: a unified brain theory?" *Nature reviews neuroscience* 11.2 (2010):
- [13] W. Freeman, G. Vitiello. "Dissipation and spontaneous symmetry breaking in brain dynamics." *Journal of Physics A: Mathematical and Theoretical* 41.30 (2008): 304042.
- [14] G. Vitiello, "My double unveiled". John Benjamins, 2001.
- [15] T.S. Cohen, M. Geiger, J. Köhler, M. Welling. "Spherical CNNs." *arXiv preprint arXiv:1801.10130*, 2018.
- [16] T. S. Cohen, M. Weiler, B. Kicanaoglu, and M. Welling. "Gauge equivariant convolutional networks and the icosahedral CNN". *CoRR*, 2019.
- [17] D. Pierangeli, et al., "Observation of replica symmetry breaking in disordered nonlinear wave propagation." *Nature communications* 8.1 (2017): 1501.
- [18] X. Lin, et al, "All-optical machine learning using diffractive deep neural networks." *Science* 361.6406 (2018): 1004-1008.
- [19] A. Silva, et al. "Performing mathematical operations with metamaterials." *Science* 343.6167 (2014): 160-163.
- [20] B. Pendry, D. Schurig, and D. R. Smith, "Controlling electromagnetic fields," *science*, vol. 312, no. 5781, pp. 1780–1782, 2006.

