

Article

# How Neurons in Deep Models Relate with Neurons in the Brain

Arianna Pavone \* and Alessio Plebe

Department of Cognitive Science, Università di Messina, via Concezione n.6/8, 98122 Messina, Italy; aplebe@unime.it

\* Correspondence: apavone@unime.it

**Abstract:** In dealing with the algorithmic aspects of intelligent systems, the analogy with the biological brain has always been attractive, and has often had a dual function. On the one hand, it has been an effective source of inspiration for their design, while, on the other hand, it has been used as the justification for their success, especially in the case of Deep Learning (DL) models. However, in recent years, inspiration from the brain has lost its grip on its first role, yet it continues to be proposed in its second role, although we believe it is also becoming less and less defensible. Outside the chorus, there are theoretical proposals that instead identify important demarcation lines between DL and human cognition, to the point of being even incommensurable. In this article we argue that, paradoxically, the partial indifference of the developers of deep neural models to the functioning of biological neurons is one of the reasons for their success, having promoted a pragmatically opportunistic attitude. We believe that it is even possible to glimpse a biological analogy of a different kind, in that the essentially heuristic way of proceeding in modern DL development bears intriguing similarities to natural evolution.

**Keywords:** brain–computer analogy; algorithmic explainability; incommensurable algorithmic solutions; black-box



**Citation:** Pavone, A.; Plebe, A. How Neurons in Deep Models Relate with Neurons in the Brain. *Algorithms* **2021**, *14*, 272. <https://doi.org/10.3390/a14090272>

Academic Editors: Frank Werner and Giuseppe Lancia

Received: 6 August 2021

Accepted: 13 September 2021

Published: 17 September 2021

**Publisher's Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

The analogy between brain and computer has always been a driving force in the field of Artificial Intelligence (AI), as a source of inspiration for the design of any algorithm underlying intelligent systems on the one hand, and as a justification for why some algorithmic technologies were able to achieve similar (or superior) performance to human ones in solving certain tasks on the other.

Both of these accounts of the brain–computer analogy, however, seem to be called into question by recent advances of modern AI based on deep learning techniques. On the one hand, it has emerged that the adoption of algorithmic strategies other than those presumed to be adopted in a biological brain has led to wider success, with results that significantly exceed human performance; on the other hand, the analogy between modern artificial neural models and biological ones appears increasingly inconsistent [1], in the physical structure, in the type of abstraction operated on the input data and in the implemented algorithmic solutions, so much so as to argue that the two cognitive models may be *incommensurable* [2]. The term *incommensurable* is probably a problematic term in this context, as it can be interpreted with multiple meanings, but we will cover the subject in more detail in Section 3.

The weakening of the analogy between the brain and the computer, which could be considered a value in itself in the design of the algorithmic aspects of neural networks, changes things. With the abandonment of the *analogy with the brain at all costs* in the design of the algorithms underlying a cognitive architecture, we have returned to an opportunistic attitude, whereby the effectiveness of a cognitive model is measured again only based on its performance: if it fulfills the task for which it was designed, then it is a good application model, otherwise not.

Regarding the second aspect, the need for a plausible explanation of the efficiency of the method seems to be a purely theoretical instance, but there are also more practical implications. For several applications there is a requirement that is known as “Explainable AI” (XAI) [3–5], which implies that the results of the solution can be understood by humans. This need for explanation has prompted research on automatic methods that provide interpretations of individual predictions made by DL models. For example, *Local Interpretable Model-agnostic Explanations* (LIME) [6] is a simplified approximation of the original DL model, which is interpretable; *SHapley Additive exPlanation* (SHAP) [7] assigns scores to the original model features that mostly contribute to an individual prediction. Even if these methods provide interesting insights on DL model behavior, they are exposed to the very same problem: how to ensure the faithfulness of the automatic generated explanations for a prediction made by inherently opaque models. Despite their opacity [8], the analogy with the functioning of a biological brain allowed the algorithms’ ancestors of modern deep learning systems to be more acceptable, and not fully understanding them could be justified by our limited knowledge of how a biological brain works [1].

Technically speaking, the crux of the XAI problem in deep learning lies in the fact that artificial neural networks do not provide programmers with clear representations of their algorithmic functioning [9–11]. Deep neural networks lack the interpretability of the model, so when we consider why a computer makes a particular decision or prediction instead of another, we remain ignorant about the reason for the decision. This lack of knowledge about decisions made by an intelligent system has important implications, ranging from the practical to the ethical to the legal.

In this work, we address the issue of providing some explanations of the reason for the failure of the analogy between biological cognitive models and the algorithmic logics adopted by modern deep learning systems, embracing the emerging thesis of *incommensurability* of the two cognitive models proposed by the scholar Beatrice Fazi [2], which is in contrast to the position of many scholars. Then, we also try to provide a justification for the success of modern artificial cognitive systems despite their persistent opacity, highlighting some aspects that make them closer to the biological cognitive model, more than what can emerge from recent studies.

Specifically, we discuss the fact that modeling a successful cognitive system, biological or artificial, can be the result of an opportunistic approach to design. After all, even nature, in the implementation of the notion of adaptation, works in an opportunistic way: if a biological model works better than another, it has more chances of surviving over time, regardless of the elegance of the solution adopted for adaptation.

A cognitive system that *self-adapts* works opportunistically by providing solutions that are not (necessarily) elegant and do not need to be explained by logical rules. Evolution, through natural selection, therefore produces cognitive models that work, but which nevertheless can remain opaque to any attempt at explainability.

## 2. Deep Learning and the Explainability Problem

The most striking change in AI in recent years is due to the family of algorithmic techniques used in modern deep learning systems [12]. These techniques represent the latest evolution of artificial neural networks organized in layers [13]. Deep learning is responsible for the current resurgence of AI [14] after several decades of slow and unsatisfactory progress. Thanks to efficient algorithmic techniques, together with their enormous computing capabilities, modern computers seem able to act without being explicitly programmed, modifying their underlying mathematical functions and defining their behavior on the basis of the data provided in input and producing decisions or forecasts.

The levels of efficiency achieved by intelligent systems based on deep learning were completely unexpected, given that they are based on algorithmic techniques that derive from artificial neural networks, an approach to AI that was stagnant at the beginning of this century. After decades of intense development, artificial neural networks seemed to

have exhausted their potential. However, applying minor tweaks to the initial design led to a resurrection so impressive that it remained largely inexplicable.

Deep learning actually works well from an engineering point of view; so good that it is difficult to understand why [9]. The decision-making of quasi-autonomous artificial agents powered by deep learning affects millions of people every day. The range of decisions covered by deep learning is vast. Thus there are certainly sociological and marketing factors contributing to the current fortunate period of deep learning, but this is obviously not sufficient to justify the effectiveness of such algorithmic techniques. For this reason, in recent years, there has been growing attention on finding valid explanations for this success. The main two explanations generally point to hardware performance and neuroscience [9], but scholars have long agreed that both of these explanations are not adoptable. First, there is a lot of skepticism about attributing the success of deep learning to computing power alone.

The second justification, that traces the effectiveness of such systems to their connection with neuroscience, has been very common in the past and is often stated as an undisputed fact. More recently, however, the role of neuroscience in explaining the efficacy of deep learning appears to have failed in its goal [1].

Technically, this aspect is referred to as the *opacity* of an algorithm, a widespread problem in modern computer science that can be associated with many widely used algorithms, such as social classification and positioning mechanisms, spam filters, credit card fraud detection, results listings in a search engine, market segmentation and advertising, just to name a few, and it is no coincidence that these classification mechanisms are in many cases based on deep learning and machine learning algorithms.

The scholar Jenna Burrell, in her work [8], distinguishes three forms of opacity: opacity as a business secret or intentional state, opacity as technical illiteracy and an opacity that derives from a lack of explainability of the algorithm. The latter is the one that interests us most in this work, a form of opacity that derives, for example, from the mismatch between the characteristic mathematical optimization of machine learning and the needs of reasoning on a human scale and of semantic interpretation styles. It is a form of opacity that is often confused with the latter as part of the general sense that algorithms and code are very technical, complex and difficult to understand [8].

*Explainability is a key word for present and future algorithmic cultures, raising equally unique social and ethical challenges. (Jenna Burrell [8])*

However, it is interesting to note that what makes deep learning techniques powerful often also makes their theoretical support opaque. In this context, deep learning algorithms can be traced back to black boxes [15], tools that appear to work very well but for which it is not as clear how and why they work. The knowledge generated and encapsulated in these models remains, in part, implicit due to the very nature of these tools: the non-linear structure of the hidden mathematical functions, the abstraction of the data they produce, the distributed character of the network representations, and the presence of numerous configurations of its large sets of variables.

Thus, strange as it may seem, there is still no formal explanation for why deep learning math is so effective. Perhaps, the domain that appears most appropriate for a justification of the functioning of deep learning is that of mathematics [1]. In a recent paper, Berner et al. [16] describe the new field of mathematical analysis of deep learning, giving partial answers to some relevant questions concerning such techniques: the outstanding generalization power of overparametrized neural networks, the role of depth in deep architectures, the apparent absence of the curse of dimensionality, the surprisingly successful optimization performance despite the non-convexity of the problem, understanding what features are learned, why deep architectures perform exceptionally well in physical problems, and which fine aspects of an architecture affect the behavior of a learning task in which way.

### 3. Incommensurable Cognitive Models

The brain provides animals with a behavioral flexibility that is unmatched in our most sophisticated digital computers. Biological brains are constantly adapting to the uncertain, noisy and rapidly changing world they find themselves immersed in. It is therefore not surprising that the idea of drawing inspiration from biological neural architectures both for the design of microprocessors—the so-called *brain in silicon* [17]—and for the design of artificial intelligent systems, has been a driving force since the birth of the first computers.

Regarding the design of the neuromorphic hardware, Plebe and Grasso [17] identify three main periods during which this idea was transformed into practical projects. The first neural hardware was designed by Marvin Minsky in 1951 [18], based on the logical interpretation of the neural activity of McCulloch and Pitts [19], and was followed by a few other attempts. After a long period of an almost total lack of progress, a renewed interest sparked in the late 1980s, driven by the success of neural algorithms in software [13], with several funded projects in Europe, the United States and Japan [20]. However, by the beginning of this century, almost none of the results of all that effort had reached maturity. In recent years, a new wave of enthusiasm for neural hardware has spread, driven by some large projects funded in Europe and the United States for realistic brain simulations [21]. In this case, a revolution in the design of microprocessors was also foreseen, in terms that closely resemble those of the previous two periods. A revolution that has never happened. Although we are not currently able to predict the future of neuromorphic hardware, the principles used to promote this venture appear theoretically wrong, and therefore the premises for the success of this approach are weak [17].

On the other hand, the discourse relating to algorithmic aspects of neural networks is different, even if the conclusions seem to go in the same direction.

Initially, the similarity with the functioning of a biological brain was considered a useful cue for the efficacy of the purpose: if the analogy worked then it was considered a good thing, otherwise different ways of dealing with the problem were investigated. However, this way of thinking had become less popular over the past two or three decades. More recently, however, with the success of deep neural networks, things seem to have changed again and the initial opportunistic attitude seems to be back in vogue again. The resonance of the successes of deep learning has sparked intense reflections and discussions within the cognitive sciences and philosophy [9,22–26] and the analogy between the biological brain and the artificial cognitive models has been even more heavily questioned, mainly because of the persistent opacity of such models [8,15].

Yet deep learning is by no means a perspective that conflicts with the idea of designing a certain cognitive architecture in a biologically inspired way. On the contrary, we argue it ultimately derives from the central role that connectionism has given to learning and this, in turn, is an idea that is based on the way the brain's learning works and that comes from the biological environment.

However, recent achievements in deep learning have provided some surprising counter-proofs of the biological-brain analogy, in which the adoption of strategies other than those adopted by the brain is successful.

Perconti and Plebe [1] suggest two possible justifications for the XAI problem. The first justification is linked to the opacity of the biological cognitive model and therefore to the fact that our knowledge of how the brain deals with certain problems is incomplete. From this follows our inability to translate this solution into the correct algorithmic approach for an artificial agent. An alternative justification is that the algorithmic solutions adopted by a biological brain are not the most effective to be implemented in an artificial cognitive model.

It should be added that there are neural models that attempt to adhere much more rigorously than DL to brain mechanisms and processes that have been functionally identified and are empirically justified. One of the best example is SPAUN [27], based on the Neural Engineering Framework [28], the functional and anatomical architecture of which broadly matches the organization of the cortex and of several subcortical nuclei. Models of this sort

are excellent for the exploration of cognitive abilities, but they are not at all competitive with DL in application fields.

Here we argue a third justification that embraces the recent hypothesis of the incommensurability of the two cognitive models advanced by the researcher Beatrice Fazi [2]. Specifically, we discuss how the mechanisms put in place by biological evolution, linked to the specific constraints of an organic system that imposes more tortuous paths, have led to solutions that are not comparable to the solutions taken in the modeling of digital cognition. However, it is important to underline that the solutions adopted by biological systems cannot be defined as better or worse than the solutions implemented in an artificial cognitive model: they are simply incommensurable.

The key idea is that recognizing the incommensurability between the way nature and machines construct cognitive models implies recognizing the ontological and epistemological disparity between the way humans and computational agents make decisions. Thus, inevitably, this discrepancy is reflected in the way such decisions might be told or represented by humans and artificial algorithmic agents, respectively.

In defense of this thesis, we highlight some of the profound differences between the two cognitive models by evaluating, in the following sections, the physical structure of the models, the differences in the learning paradigm and the differences in the abstraction model.

### 3.1. Structural Issues

It is self-evident that, when comparing brains and computers, the first difference that stands out is at a structural level, although some similarities can be observed.

We can observe that both biological brain and a computer base their functioning on electricity, but this analogy is ephemeral. The electrical energy for digital computing is in fact conducted by metals, such as copper, the fastest conductor available, and semiconductors, such as silicon and germanium, which allow the flow of electrons to be controlled at the maximum possible speed. The biological model has instead faced a huge disadvantage in dealing with electricity, compared to human-made devices, as metals and semiconductors cannot be used inside living organisms. Given the abundant availability of sodium in the marine environment during the Paleozoic, nature opted for the use of ions, the only electrical conductors compatible with organic materials available at that time. The biophysical breakthrough in the exploitation of electricity in animals was in fact the ion channel, a sort of natural electrical device, which first appeared as a potassium channel about three billion years ago in bacteria, then evolved into the sodium channel 650 million years ago, which is currently the most important neural channel [29].

Of course, evolution has had its problems in allowing biological cognitive models to deal with electricity, but it has also had much more imagination in designing such models. In fact, neurons are not all the same. There are hundreds of different types of neurons that have been identified in the mammalian brain. Neurons can range in size from few micrometers up to several centimeters in length [30]. The number of inputs to a cell can range from around 500 or less (in the retina) to well over 200,000 (Purkinje cells in the cerebellum). Even in the new era of three-dimensional semiconductors, which could adapt to the density of a massive connection pool in the nervous system, technology would not be sufficient as a basis for engineering the layered structure of a biological brain.

It has been a traditionally held view that the ways of exploiting electricity in the brain and in the computer lead to two different kind of computation: analogical and digital. This is undoubtedly an important differentiation, but its actual scope has been widely reviewed in the light of advances in philosophy of computation. There is currently a convergence on a broader notion of computation, which Piccinini and Bahar [31] call *generic computation* that can encompass brains, computers, and possibly other physical devices. It is defined as the processing of vehicles—entities or variables that can change state according to rules that are sensitive to certain properties of the vehicle and, in particular, to then differences between

different portions of the vehicles [32]. Assuming that brains process medium-independent vehicles, it follows that both computers and brains are generic computation devices.

### 3.2. The Learning Paradigm

Learning is the behavioral modification that follows, or is induced by, an interaction with the environment and is the result of experiences that lead to the establishment of new response configurations to external stimuli. It is one of the fundamental cognitive phenomena for evolution and involves many animal species in addition to man; the development and survival of individuals is based on their ability to learn.

The learning paradigm is also one of the most successful loans from the mind and brain to the world of digital computing, a loan that has accompanied the design of artificial intelligent systems since their inception. Alan Turing [33] himself, with extraordinary foresight, imagined the creation of artificial brains which, instead of being rigidly programmed in their tasks, could learn them on their own through a series of experiences. Inspired by neuroscience, he also imagined the presence of distributed interconnected elements and the possibility of organizing the system by strengthening or cutting connections based on experience.

From a simple idea, the learning paradigm became a reality about 30 years after the appearance of the first digital computers with the invention of the first artificial neural networks. Today, the “deep” version of those early neural networks is responsible for the astonishing resurgence of artificial intelligence.

One might therefore think that the analogy with the learning of a biological brain has paid off. Once again, however, this analogy seems illuminating, but only at an appropriate distance.

The success of the artificial neural network [13] has in fact nothing to do with neuroscience, and instead derives from the gradient methods, developed in mathematics for the minimization of the continuously differentiable function [34]. The resulting learning algorithm is very efficient and is called the backpropagation algorithm. This method, conceived by Geoffrey Hinton [35], the father of deep learning, however drastically differs from biological learning.

If  $\mathbf{w}$  is the vector of all the learnable parameters in a network, and  $\mathcal{L}(\mathbf{x}, \mathbf{w})$  is the measure of the error of the network with parameters  $\mathbf{w}$  when applied to sample  $\mathbf{x}$ , the back-propagation updates the parameters iteratively, according to the following formula:

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \nabla_{\mathbf{w}} \mathcal{L}(\mathbf{x}_t, \mathbf{w}_t), \quad (1)$$

where  $t$  spans all available samples  $\mathbf{x}_t$ , and  $\eta$  is the learning rate.

Today, deep neural networks can still learn with algorithms quite close to the good old backpropagation, even if this term has fallen into disuse, replaced by Stochastic Gradient Descent (SGD) [36]. Its basic formulation is as follows:

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \eta \nabla_{\mathbf{w}} \frac{1}{M} \sum_i^M \mathcal{L}(\mathbf{x}_i, \mathbf{w}_t), \quad (2)$$

and the similarity with the standard backpropagation equation is immediately noticed.

Instead of calculating the gradients on a single sample  $t$ , in Equation (1) a stochastic estimate is made on a random subset of dimension  $M$  of the entire dataset and at each iteration step  $t$  a different subset is sampled, with the same size.

There have been attempts to more closely reproduce the learning mechanisms put in place by the biological brain, but every time we try to copy biological learning techniques in more detail, the advantages have disappeared [1]. Hinton himself, despite his fundamental contribution in the invention of the back-propagation algorithm, worked on an alternative neural model called the Boltzmann Machine, which used a more plausible learning method [37]. However, the performances obtained by these models have never been up to those obtained with the learning of networks with back propagation. Even

today, Hinton continues to test whether there are variations in the SGD in the direction of some sort of biological plausibility, but the results on large-scale benchmarks are well below those of the SGD [38].

Despite wide efforts, the effectiveness of SGD in deep learning remains elusive. There are, indeed, theoretical results on the convergence of SGD [39]. Let  $\mathbf{w}^*$  be the best possible vector of parameters, selected by those with a norm less than a scalar  $B > 0$ ,  $\mathcal{W} = \{w \mid \|w\| \leq B\}$ :

$$\mathbf{w}^* = \min_{\mathbf{w} \in \mathcal{W}} \left\{ \frac{1}{N} \sum_i^N \mathcal{L}(\mathbf{x}_i, \mathbf{w}), \right\} \quad (3)$$

where  $N$  is the number of available samples, in the general case  $\mathbf{w}^*$  is unknown. Further let  $\rho > 0$ , and assume Equation (2) is run for  $T$  iterations with learning rate  $\eta = \frac{B}{\rho\sqrt{T}}$ . Then:

$$\frac{1}{N} \sum_i^N \mathcal{L}(\mathbf{x}_i, \mathbf{w}_T) - \frac{1}{N} \sum_i^N \mathcal{L}(\mathbf{x}_i, \mathbf{w}^*) \leq \frac{B\rho}{\sqrt{T}}. \quad (4)$$

Therefore, it is possible to achieve an arbitrary convergence with a large enough number  $T$  of iterations. This result, however, has been demonstrated only for the case where the loss function is convex in  $\mathbf{w}$ , and this assumption is rarely met in deep learning models.

Furthermore, the sharp advantage of deep over shallow networks remains puzzling, given that a higher number of layers is likely to lead to complex, non-convex loss functions. Partial answers to this questions are coming from topological analysis. Bianchini and Scarselli [40,41] used Betti numbers to investigate the topological properties of the space of functions generated by neural networks. The term ‘‘Betti number’’ was introduced by Henri Poincaré after the work of Enrico [42], and, informally, refers to the number of holes on a topological surface, in a given dimension. Bianchini and Scarselli found different asymptotic expressions for the Betti numbers of the topology generated by all functions in neural networks for shallow and deep networks. The analysis is limited to networks with a single binary output, and the topology is investigated for the set of all  $\vec{x}$  for which the output of the network is positive, as is typical in a binary classifier. Calling such a set  $\mathcal{S}_n$  for shallow networks, and  $\mathcal{D}_n$  for deep ones, with  $n$  the overall number of units, the sum of Betti numbers  $B()$  is ruled by the following equations:

$$B(\mathcal{S}_n) \in O(n^D), \quad (5)$$

$$B(\mathcal{D}_n) \in \Omega(2^n), \quad (6)$$

where  $D$  is the dimension of the input vector. Equation (5) states that for shallow networks Betti numbers grows at most polynomially with respect to the number of the hidden units  $n$ , while from Equation (6) it turns out that for deep architectures Betti numbers can grow exponentially in the number of the hidden units. Therefore, by increasing the number of hidden units, more complex functions can be generated when the architectures are deep.

Betti numbers based complexity has been later extended to more general neural models, as classifiers into  $c$  classes [43]. For a model  $\mathcal{M}$  the sum of Betti numbers can have the following complexity:

$$B(\mathcal{M}_n) \in O\left((c(h-1)+1)^{c+n}\right), \quad (7)$$

where  $n$  is the total number of the hidden units as in the previous equations, and  $h$  is the number of hidden layers. Equation (7) indicates that, for multi-class neural models too, the Betti numbers based complexity grows with the increasing depth  $h$ . In [44] there is the demonstration that a simple family of functions on  $\mathbb{R}^d$  is expressible by a feedforward neural network with two hidden layers and not by a network with one hidden layer, unless its width will grow with  $O(e^d)$ . The same group later extended [45] the results to hyperspherical and hyperelliptical functions. In the end, whatever the essential reason why

SGD works so well for deep networks, these studies clearly show that this reason lies in mathematical intricacies, which have very little to do with the functioning of brain neurons.

### 3.3. Information Abstraction

In analyzing the analogy between the biological brain and its digital counterpart, it is therefore also necessary to pay attention to how a human being (or an animal) receives and processes stimuli and information, and how it can be said that a digital machine can do the same. There is also a significant difference between the abstraction choices adopted by man and those adopted by a computer. As we will see also in this respect, the two models appear immeasurable because they cannot be measured against each other or compared with a common standard.

This difference not only concerns the different capacity at which the two models have to receive input data, but also concerns a specifically computational relationship with the intelligible, that is, with what is apprehensible only through forms of abstract activity.

Using the metaphor of vision, since much of the most successful scientific research on deep learning has focused on computer vision, we can say that artificial cognitive agents see and process inputs differently. Indeed, relatively simple and immediate human intuitions on how to identify shapes, structures and geometric primitives are not easily expressed in computational terms.

Here we return to an example reported by Beatrice Fazi [2] relating to the automatic recognition of human handwriting, a task notoriously difficult to perform computationally and for which the more traditional programming techniques do not work well but where, however, deep learning algorithms succeed very well. Suppose we want a program to recognize a handwritten digit, such as zero. In the case of supervised learning, thousands of scans of handwritten zeros are sent to the machine as training examples. The program then learns to recognize the digit, not as a human might do (e.g., by determining that a zero resembles a vertical oval), but by mechanically detecting complex patterns of darker and lighter pixels expressed as arrays of numbers.

This is certainly a different form of perception or reception of the input data. However, beyond the physical reception of data, we are also seeing a specific form of abstractive ability, similar to an automated mode of conceptualization, that is, an automated mode of forming internal representations intended to generalize while abstracting from observed facts.

For example, returning to the case of algorithmic recognition of handwritten zeros, the deep learning model identifies and constructs more relevant representations than any human programmer could have identified and given to the machine. In fact, these are representations that a human would not have (and could not have) abstracted in the first place. The way in which the program extracts and organizes information in terms of characteristics and then generalizes this information to form the desired “concept”—or, in computational terms, the representation of the desired output of zero—is therefore entirely and exclusively computational.

## 4. A Role for Evolution, and an Artificial Metaphor for It

From what has been discussed in the previous sections, it is evident that the attempts to compare biological brain models and modern artificial cognitive models have failed except in specific sectors such as models related to computer vision.

This certainly had repercussions on the explainability of the functioning of artificial cognitive models. At present, progress in understanding some computational activities is achieved by trial and error, and operations are often rationalized retrospectively. To put it differently, “many algorithms used in artificial neural networks are understood only at the heuristic level, where scientists know empirically that certain training protocols employing large datasets will result in excellent performance” [46].

However, one must always keep in mind that neural networks share a general cognitive principle that is fundamental to biological neural circuits as well. They are both



formidable devices for learning sophisticated functions based on experience. Organisms whose brain circuits have the highest capacity to learn, such as man, owe it to natural evolution. The ability of neural models to learn, despite the efficiency of deep learning, does not come for free, nor is there a natural evolution that helps. However, we argue that the patient and painstaking effort of deep neural model designers can be framed as a metaphor for natural evolution.

The basic idea of biological evolution is that populations and species of organisms change over time. The British naturalist Charles Darwin, in his most influential and controversial book entitled “The Origin of Species” [47], was the first to propose the idea that living species can evolve and to suggest a mechanism for evolution: natural selection and genetic drift, that is, the change in the frequency of an existing genetic variant. Based on this mechanism, hereditary traits that help organisms survive and reproduce become more common in a population over time.

Such mechanisms of evolution work with the random variation generated by the mutation of the genetic code of organisms in the passage from one generation to the next. Although environmental factors are thought to influence the mutation rate, scholars agree that they are unable to influence the direction of the mutation. Hence, for example, exposure to harmful chemicals can increase the mutation rate, but it will not be able to stimulate those mutations that make the body resistant to those chemicals. For this reason the mutations are random, that is, the fact that specific mutations occur or not is not related to the usefulness of the mutation itself.

Here, we argue that biological and artificial cognitive models have a lot in common regarding the process of evolution whereby such models are modified until they converge to a successful version in solving certain problems. In this context, it is necessary to take into account the right differences in carrying out this analogy. The metaphor of evolution should be considered not in the strict sense but serves to help us understand how certain dynamics can coincide, as well as the results obtained in the process of building a cognitive model.

For instance, one of the interesting things is that evolution, in its general sense, does not necessarily converge to the same solution nor the optimal solution but, based on the conditions and contexts in which it operates, it can produce significantly different solutions to the same problem. Such solutions, although different, are shaped by the same forces and are produced by the same laws and, at times, may have nothing in common and can hardly be compared, therefore they are incommensurable.

#### 4.1. Variable Tuning and Natural Selection

The cognitive model of a living species, such as man, is defined at birth and is written in its genetic code, the DNA, inherited from the parents. We are talking about a complex and particularly large piece of code. The human genome, for instance, is made of 3.2 billion bases but other organisms have different genome sizes, and even much simpler living beings are characterized by a very complex genetic code: the Sars-Cov-2 virus, for example, with over 29 million bases is one of the RNA viruses with the longest and most complex genetic code.

We refer to it as a *piece of code* since it contains instructions needed to make the proteins and molecules essential for the growth, development and health of the organism. However, it also contains the *parameters* of the biological system (which are the indicators of the features of the living organism) including the brain, indicating which cells it will contain, the number of each cell, the size of those cells or how they will interact, just to enumerate a few of these parameters.

However, if we think of the various changes that occurred in the brain from the first mammals to the appearance of homo sapiens, we can see how this model has evolved into the very complex version that we (partially) know today. This took place not only through changes that made it possible to converge to the current version of the model, but also through a series of trials and errors implemented by the evolutionary process and by natural selection. This is a phenomenon known as *genetic drift* in which some genes

may occasionally change randomly as they pass from one generation to the next. If these mutations lead to changes in the model that make it more environmentally friendly then they are more likely to consolidate in subsequent generations. Evolution then proceeds by applying random modifications to the biological parameters encoded in the genome and natural selection chooses which of the proposed mutations is the most suitable for survival in the environment.

Similarly to what happens in the evolution of a biological cognitive model, we can say that the modeling of an artificial cognitive system follows a similar path. Neural network modeling can often involve tuning many different hyperparameters, which are typically external to the model, which means that they cannot be learned and modified by the network itself through data processing. They are referred to as optimization parameters because there is no analytical formula available to compute their appropriate values [48].

For this reason, such values must often be specified by the programmer. Their definition is based, in most cases, on experience or through the setting of some initial values and, subsequently, proceeding by trials and errors to modify them until the identification of a satisfactory solution for a given problem.

Often one of the most painful aspects of Deep Neural Networks modeling and training is the large number of hyperparameters you are dealing with. These could be, for instance, the learning rate  $\alpha$ , the discount factor  $\rho$  and the epsilon  $\varepsilon$  if the RMSprop optimizer is used [49,50], or the exponential decay rates  $\beta_1$  and  $\beta_2$  if the Adam optimizer [51] is used. It is also necessary to define the number of levels in the network or the number of hidden units for each level. In addition it may be possible to use learning rate planners. However, those just exposed are just some of the hyperparameters that may need to be configured when modeling a neural network.

Usually, it can help to classify hyperparameters into two groups: hyperparameters used for training and those used for model design. A correct choice of hyperparameters related to model training would allow neural networks to learn faster and achieve improved performance making the optimization process definitely something you would like to worry about. Hyperparameters for model design are more related to the structure of neural networks. A trivial example is the number of hidden layers and the width of these layers.

A really useful summary of the order of importance of hyperparameters was compiled by Tong Yu and Hong Zhu in their paper [52].

Thus, the evolution of a cognitive model can be seen (in all contexts) as a *tuning* problem of the model variables. It is therefore an engineering problem and not an algorithmic design problem.

However, unlike what happens in the design of algorithms operating in intelligent artificial systems, biological cognitive models are not driven by a programmer or any higher entity. There is no grand plan or design of nature. Evolution has no purpose and there are a lot of mutations in genes that are neutral and have absolutely no positive or negative effect on survival, and they just happen to be passed on and it is easy to understand how these neutral aspects are the ones that most resist explainability, both in a biological system and in an artificial one.

In fact, it is a very common misconception about natural selection, that over time, evolution selects the characteristics of an organism that are most perfectly suited to its environment, just as a programmer selects the values of the hyperparameters that allow for better performance of the neural network. The misunderstanding may be partly due to the term *natural selection* itself which conjures up parallels with, for example, a dog breeder “selecting” the desirable traits in their animals. Indeed, nature is not actually “selecting” anything: natural selection is a process, not a conscious force. For this reason, there are good reasons why the natural selection process may not always lead to a “perfect” solution. First, selection can only act on the available genetic variation. A cheetah, for example, cannot evolve to run faster if a genetic variant is not available that allows for greater speed. Secondly, the biological cognitive model is limited in working with the tools that nature has made available to it. This means that an artificial cognitive system is able to converge

in a more targeted and faster way to a version capable of obtaining good results in solving certain tasks. However, there are already some first attempts to leave the evolution of artificial models to chance by implementing neural networks that automatically modify their hyperparameters assisted by genetic algorithms [53,54].

Thus, another aspect that disagrees with this analogy could be the fact that, in an artificial system, evolution is a relatively fast process, unlike what happens in the evolution of biological models which require unfathomable times. However, it might come as a surprise to know that biological evolution can also be very fast. We can also see it happen in real time, if we're lucky (or unlucky).

Let us take some very small life forms like bacteria, for example, which are actively becoming resistant to antibiotics. Some have become resistant to so many drugs that they are rapidly becoming nearly impossible to treat. Mutations in the genes of a virus responsible for a pandemic, such as that of COVID-19, can occur within a few weeks, giving rise to several new variants. There is a similar situation even in the not so small life forms. Just to cite an example, we spend a lot of money on pesticides and herbicides to eradicate pests and weeds, but both have created all sorts of adaptations over the past half century to escape death and continue to thrive.

#### 4.2. A World of Different Cognitive Models

The idea of the existence of very different (or even incommensurable) cognitive models is not something that only shapes the comparison between the biological and the artificial brains. The world we live in could already provide us with notable examples of significantly different cognitive models, each with its own peculiar features, each of which offers different solutions to the same problems, sometimes impenetrable to explainability.

Charles Darwin's idea that man is essentially a "big-brain ape" and the basic mammalian brain uniformity theory was abandoned in the 1990s when the turn to microscopic study of the human brain [55] highlighted significant differences in the very structure of the various cognitive models. It was discovered, for example, that in one layer of man's primary visual cortex, nerve cells are organized in a complex mesh pattern that is very different even from that of primates, our closest relatives.

Obviously, the fact that man is an animal implies that his brain has a lot in common with that of many other animals, for example the components that allow us to manage movement, thought and the perception of the world around us. However, every animal's brain does something a little different and special. For example, cats have very good eyesight and have more brain capacity for their sense of sight. Likewise, dogs have a very good sense of smell, so the part of their brain that can identify different smells is very powerful compared to other animals. However, these examples concern senses that are familiar to us even if they are considerably more sensitive. Thus it is possible to give a sort of explanation of the cognitive abilities of these animals, even if they are hardly imaginable for us.

Taking a step further towards our thesis, we can observe that the cognitive models developed by other animals make it possible to process information that humans are unable to perceive. Bats, for example, are able to map their surroundings through sound, thanks to a sonar system used also in other animals such as dolphins and whales. It has been discovered that the sounds produced by these species allow them to form a mental map of their surroundings in three dimensions. More recently it has been discovered that this ability is even more incredible: the sound waves emitted can penetrate through certain objects and soft tissues, providing the animal with a kind of X-ray vision. Even electricity, invisible to us, represents for some animal species a signal that guides them towards their goal. Sharks and other fish can detect it thanks to a series of channels filled with a gelatinous material, which open outward through the pores of the skin.

In these specific cases it is much more difficult for us to give an understandable interpretation of how these species manage to solve certain problems thanks to their

perceptual models. There is a scientific interpretation but explaining it to those unfamiliar with the language of science is particularly difficult.

In nature there are cognitive models that are even less penetrable to explainability. For example, it has recently been discovered that some insects, and also other animals such as octopus, thanks to the dorsal edge of their eyes, are able to distinguish polarized light, that is, light that has a specific vector orientation in space. This cognitive ability appears to be also used by ants and bees to travel great distances to return to their burrows.

Finally there are some cases that still resist explainability. We cite the tendency of certain animals to align their bodies with the north–south axis of the earth’s magnetic field during movement. Scientists have been observing this phenomenon for a decade, but have found no explanation for this ability, which, according to what has been confirmed, is present in many creatures, from bacteria to vertebrates.

Scientists argue that different species use different systems to do so. For example, there is a type of pigment called cryptochrome that is present in the eye of some species and is activated with blue light by a magnetism-sensitive quantum mechanism. It is thought that this would allow these animals to see the Earth’s magnetic field in the form of a blue trail.

## 5. Discussion

In order for the algorithmic aspects of a cognitive model, biological or artificial, to be effectively explained and therefore expressed and shared, there must be a common experience between the communicator and the recipient of the communication. It is clear that this is not possible in the case of interactions between man and a computer, since there is no common phenomenological or existential ground between human abstractions and those of a computational agent.

The same can be said in the case of interactions between animal species that perceive the environment differently. Although scientists are able to give an explanation of why certain cognitive and perceptual models can work, it is not possible to allow an experiential sharing of the algorithmic aspects that characterize the cognitive model. Physics tells us how bats see the world around us, but we are unable to imagine or share that kind of perception.

We must therefore be careful when dealing with, on the one hand, how a human being receives and processes stimuli or information, and how, on the other, it can be said that an artificial system (or another animal species) does the same. It is therefore important to speak of incommensurability between abstract choices made by different cognitive models.

The confirmation of what has been said can be found in the analysis of artificial cognitive systems capable of self-training without the use of training datasets provided in input. Let us take the example of AlphaGo Zero, a version of DeepMind’s AlphaGo software developed in 2017 [56] and created without using training data. AlphaGo Zero was stronger than any previous version and reached the level of AlphaGo Master in 21 days. The interesting thing is that this software was successful not because it behaved like a human player, but because it played differently to a human. This condition is particularly interesting from both a philosophical and a socio-cultural point of view. The cognitive model has autonomously developed its own completely personal sensitivity in playing games, far exceeding human performance in this task.

If we accept the fact that artificial cognitive models are incommensurable compared to the human cognitive model and, consequently, that they are able to develop independent solutions significantly divergent from those adopted by a biological brain, we could be able to exploit them as a source of inspiration to guide human investigation.

For example, in her recent work “In Defense of the Black Box”, scholar Elizabeth Holm reports that in an innovative medical imaging study, scientists trained a deep learning system to diagnose diabetic retinopathy, achieving performance that has outdated a committee of ophthalmological experts, exploiting a variety of factors including cardiological risk, age and sex. Interestingly, no one had ever noticed gender differences in human retinas before,

so the results from the artificial cognitive model inspired researchers to investigate how and why male and female retinas differ.

Alternative cognitive models available in nature (implemented in different animal species) have often inspired our technology and science in general. For example, the dual frequency band biosonar of bats inspired scientists to develop a new processing technique to improve geophysical observations based on the interpretation of two-dimensional radar images of the subsoil [57].

It follows therefore that incommensurable cognitive models, precisely because they are such, can contribute substantially and productively to science, technology, engineering and mathematics to “provide value, optimize results and stimulate inspiration” [15].

## 6. Conclusions

We have tried to assess the relation between the so-called “neurons” in DL—the most successful family of algorithms nowadays—and their biological counterparts in the brain. We found that the common explanation of the outstanding efficiency of DL thanks to brain inspiration is largely unwarranted. Indeed, DL inherits from brains the paradigm of learning, bringing models to work by the modification of parameters induced by experience, but the ways to actually modify the parameters are profoundly different in the brain and DL models. The lack of a sound mathematical explanation of how learning works efficiently in DL models is one reason for the mostly heuristic search for improvements. Sometimes the tuning of the hyperparameters that rule the modification of parameters during learning proceeds so tentatively as to deserve an analogy with natural selection. A critical consequence of our analysis is that the glaring opacity of DL models cannot be justified—therefore accepted—with the excuse of mimicking the brain, which is inherently opaque. We have discussed a position outside the chorus that highlights the gap between DL and the human mind, to the point of even being incommensurable. We suggested that designing DL models without pursuing biological plausibility, even being incommensurable with human cognition, is a component of their success.

In the end, even the fact that we are unable to provide a logical explanation for why the algorithms underlying state-of-the-art DL systems work so well should not in itself be a problem with the applicability of such tools.

**Author Contributions:** Both authors has contributed equally to this work. Both authors have read and agreed to the published version of the manuscript.

**Funding:** This research received no external funding.

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Perconti, P.; Plebe, A. Brain Inspiration Is Not Panacea. In *Brain-Inspired Cognitive Architectures for Artificial Intelligence*; Advances in Intelligent Systems and Computing; Springer: Cham, Switzerland, 2021; Volume 1310, pp. 359–364.
2. Fazi, M.B. Beyond Human: Deep Learning, Explainability and Representation. *Theory Cult. Soc.* **2020**, 1–23, [CrossRef]
3. Dösilović, F.K.; Brčić, M.; Hlupić, N. Explainable artificial intelligence: A survey. In Proceedings of the 2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), Opatija, Croatia, 21–25 May 2018; pp. 0210–0215.
4. Molnar, C. *Interpretable Machine Learning. A Guide for Making Black Box Models Explainable*; Lulu Press: Morrisville, NC, USA, 2019.
5. van Lent, M.; Fisher, W.; Mancuso, M. An Explainable Artificial Intelligence System for Small-unit Tactical Behavior. In Proceedings of the AAAI Conference on Artificial Intelligence, San Jose, CA, USA, 25–29 July 2004; pp. 900–907.
6. Ribeiro, M.T.; Singh, S.; Guestrin, C. “Why Should I Trust You?”: Explaining the Predictions of Any Classifier. *arXiv* **2016**, arXiv:1602.04938.
7. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. In Proceedings of the 31st International Conference on Neural Information Processing Systems, (NIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 4768–4777.

8. Burrel, J. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data Soc.* **2016**, *3*, 1–12. [[CrossRef](#)]
9. Plebe, A.; Grasso, G. The unbearable shallow understanding of deep learning. *Minds Mach.* **2019**, *29*, 515–553. [[CrossRef](#)]
10. Samek, W.; Wiegand, T.; Müller, K.-R. Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models. *arXiv* **2017**, arXiv:1708.08296.
11. Zednik, C. Solving the black box problem: A normative framework for explainable artificial intelligence. *Philos. Technol.* **2021**, *34*, 265–288. [[CrossRef](#)]
12. Schmidhuber, J. Deep learning in neural networks: An overview. *Neural Netw.* **2015**, *61*, 85–117. [[CrossRef](#)] [[PubMed](#)]
13. Rumelhart, D.E.; McClelland, J.L. (Eds.) *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*; MIT Press: Cambridge, MA, USA, 1986.
14. Tan, K.H.; Lim, B.P. The artificial intelligence renaissance: Deep learning and the road to human-level machine intelligence. *APSIPA Trans. Signal Inf. Process.* **2019**, *7*, e6. [[CrossRef](#)]
15. Holm, E.A. In defense of the black box. *Science* **2019**, *364*, 26–27. [[CrossRef](#)]
16. Berner, J.; Grohs, P.; Kutyniok, G.; Petersen, P. The Modern Mathematics of Deep Learning. *arXiv* **2021**, arXiv:2105.04026.
17. Plebe, A.; Grasso, G. The Brain in Silicon: History, and Skepticism. In Proceedings of the 3rd International Conference on History and Philosophy of Computing (HaPoC), Pisa, Italy, 8–11 October 2015; pp. 273–286.
18. Minsky, M.L. Neural Nets and the Brain-Model Problem. Ph.D. Thesis, Princeton University, Princeton, NJ, USA, 1954.
19. McCulloch, W.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [[CrossRef](#)]
20. Taylor, J. *The Promise of Neural Networks*; Springer: London, UK, 1993.
21. Markram, H.; Muller, E.; Ramaswamy, S.; Reimann, M.W.; Abdellah, M.; Sanchez, C.A.; Ailamaki, A.; Alonso-Nanclares, L.; Antille, N.; Arsever, S.; et al. Reconstruction and simulation of neocortical microcircuitry. *Cell* **2015**, *163*, 456–492. [[CrossRef](#)]
22. Lake, B.M.; Ullman, T.D.; Tenenbaum, J.B.; Gershman, S.J. Building machines that learn and think like people. *Behav. Brain Sci.* **2017**, *40*, e253. [[CrossRef](#)]
23. Landgrebe, J.; Smith, B. Making AI meaningful again. *Synthese* **2019**, *198*, 2061–2081. [[CrossRef](#)]
24. Marcus, G. Deep learning: A critical appraisal. *arXiv* **2018**, arXiv:1801.00631.
25. Perconti, P.; Plebe, A. Deep learning and cognitive science. *Cognition* **2020**, *203*, 104365. [[CrossRef](#)] [[PubMed](#)]
26. Schubach, A. Judging machines: Philosophical aspects of deep learning. *Synthese* **2019**, *198*, 1807–1827. [[CrossRef](#)]
27. Eliasmith, C.; Stewart, T.C.; Choo, X.; Bekolay, T.; DeWolf, T.; Tang, Y.; Rasmussen, D. A Large-Scale Model of the Functioning Brain. *Science* **2012**, *338*, 1202–1205. [[CrossRef](#)]
28. Eliasmith, C. *How to Build a Brain: A Neural Architecture for Biological Cognition*; Oxford University Press: Oxford, UK, 2013.
29. Zakon, H.H. Adaptive evolution of voltage-gated sodium channels: The first 800 million years. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 10619–10625. [[CrossRef](#)]
30. Squire, L.R.; Bloom, F.; McConnell, S.; Roberts, J.; Spitzer, N.; Zigmond, M. *Fundamental Neuroscience*; Academic Press: New York, NY, USA, 2003.
31. Piccinini, G.; Bahar, S. Neural Computation and the Computational Theory of Cognition. *Cogn. Sci.* **2013**, *34*, 453–488. [[CrossRef](#)]
32. Piccinini, G. *Physical Computation: A Mechanistic Account*; Oxford University Press: Oxford, UK, 2015.
33. Turing, A. Intelligent machinery. In *Collected Works of A. M. Turing: Mechanical Intelligence*; Ince, D.C., Ed.; Technical Report for National Physical Laboratory; Edinburgh University Press: London, UK, 1969.
34. Cauchy, A.L. Methode generale le pour la resolution des systemes d'equations simultatees. *Compt. Rend. Seances Acad. Sci. Paris* **1847**, *25*, 536–538.
35. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning representations by backpropagating errors. *Nature* **1986**, *323*, 533–536. [[CrossRef](#)]
36. Bottou, L.; LeCun, Y. Large scale online learning. In *Advances in Neural Information Processing Systems*; MIT Press: Cambridge, MA, USA, 2004; pp. 217–224.
37. Hinton, G.E.; Sejnowski, T.J.; Ackley, D.H. *Boltzmann Machines: Constraint Networks That Learn*; Technical Report 84-119; Carnegie-Mellon University, Computer Science Department: Pittsburgh, PA, USA, 1984.
38. Bartunov, S.; Santoro, A.; Richards, B.A.; Marris, L.; Hinton, G.E.; Lillicrap, T. Assessing the scalability of biologically-motivated deep learning algorithms and architectures. In Proceedings of the 32nd Conference on Neural Information Processing Systems (NeurIPS 2018), Montreal, QC, Canada, 3–8 December 2018.
39. Shai, S.S.; Shai, B.D. *Understanding Machine Learning: From Theory to Algorithms*; Cambridge University Press: Cambridge, UK, 2014.
40. Bianchini, M.; Scarselli, F. On the complexity of shallow and deep neural network classifiers. In Proceedings of the European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, Bruges, Belgium, 23–25 April 2014; pp. 371–376.
41. Bianchini, M.; Scarselli, F. On the complexity of neural network classifiers: A Comparison Between Shallow and Deep Architectures. *IEEE Trans. Neural Netw. Learn. Syst.* **2014**, *25*, 1553–1565. [[CrossRef](#)]
42. Betti, E. Teoria della elasticita. *Il Nuovo Cimento (1869–1876)* **1872**, *7*, 158–180. [[CrossRef](#)]
43. Sun, S.; Chen, W.; Wang, L.; Liu, X.; Liu, T.Y. On the Depth of Deep Neural Networks: A Theoretical View. In Proceedings of the AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 2066–2072.

44. Eldan, R.; Shamir, O. The Power of Depth for Feedforward Neural Networks. *J. Mach. Learn. Res.* **2016**, *49*, 1–34.
45. Safran, I.; Shamir, O. Depth-Width Tradeoffs in Approximating Natural Functions with Neural Networks. *arXiv* **2017**, arXiv:1610.09887.
46. Lin, H.W.; Tegmark, M.; Rolnick, D. Why does deep and cheap learning work so well? *J. Stat. Phys.* **2017**, *168*, 1223–1247. [[CrossRef](#)]
47. Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*; Routledge: London, UK, 1859.
48. Kuhn, M.; Johnson, K. *Applied Predictive Modeling*; Springer: New York, NY, USA, 2013.
49. Dauphin, Y.H.; de Vries, H.; Chung, J.; Bengio, Y. RMSProp and equilibrated adaptive learning rates for non-convex optimization. *arXiv* **2015**, arXiv:1502.04390.
50. Xu, D.; Zhang, S.; Zhang, H.; Mandic, D.P. Convergence of the RMSProp deep learning method with penalty for nonconvex optimization. *Neural Netw.* **2021**, *139*, 17–23. [[CrossRef](#)] [[PubMed](#)]
51. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization. *arXiv* **2014**, arXiv:1412.6980.
52. Yu, T.; Zhu, H. Hyper-Parameter Optimization: A Review of Algorithms and Applications. *arXiv* **2020**, arXiv:2003.05689.
53. Ma, B.; Li, X.; Xia, Y.; Zhang, Y. Autonomous deep learning: A genetic DCNN designer for image classification. *Neurocomputing* **2020**, *379*, 152–161. [[CrossRef](#)]
54. Pan, Y.; Yang, Y.; Li, W. A Deep Learning Trained by Genetic Algorithm to Improve the Efficiency of Path Planning for Data Collection with Multi-UAV. *IEEE Access* **2021**, *9*, 7994–8005. [[CrossRef](#)]
55. Preuss, T.M.; Coleman, G.Q. Human-specific organization of primary visual cortex: Alternating compartments of dense Cat-301 and calbindin immunoreactivity in layer 4A. *Cereb. Cortex* **2002**, *12*, 671–691. [[CrossRef](#)] [[PubMed](#)]
56. Silver, D.; Schrittwieser, J.; Simonyan, K.; Antonoglou, I.; Huang, A.; Guez, A.; Hubert, T.; Baker, L.; Lai, M.; Bolton, A.; et al. Mastering the game of Go without human knowledge. *Nature* **2017**, *550*, 354–359. [[CrossRef](#)]
57. Carrer, L.; Bruzzone, L. Solving for ambiguities in radar geophysical exploration of planetary bodies by mimicking bats echolocation. *Nat. Commun.* **2017**, *8*, 2248. [[CrossRef](#)]