Deep Learning and the Tri-Level Hypothesis

Cosmo Harrigan[*]

Department of Mathematics, University of Washington

1 Introduction

I will present Marr's Tri-Level Hypothesis on cognitive systems, which states that a cognitive system should be considered and studied at three distinct levels: computation, representation and implementation. I will then analyze the convolutional neural network model from deep learning under this paradigm. Finally, I will present Poggio's argument for the inclusion of an additional level of description to explain how learning takes place and reasons why I find this to be a compelling argument.

2 Overview of the Tri-Level Hypothesis

Initially, studies of the human visual system and the brain focused on how neurons fired and how signals were transmitted in response to sensory stimuli. Eventually, it was realized that in order to more productively understand these processes, it was necessary to consider them as information-processing systems operating at multiple levels of abstraction [1]. It was proposed that a cognitive system like vision could not be understood simply in terms of the functioning of its component parts; that, rather, it was necessary to "contemplate different kinds of explanation at different levels of description" which, taken together, produce "a cohesive whole" [1].

Along with Poggio, Marr formulated the *Tri-Level Hypothesis*, which outlined the following relevant levels of study for cognitive systems [1]:

[*] cosmoh@uw.edu

1. **Hardware implementation:** How can the representation and algorithm be realized physically?

2. **Representation and algorithm:** How can this computational theory be implemented? In particular, what is the representation for the input and output, and what is the algorithm for the transformation?

3. **Computational theory:** What is the goal of the computation, why is it appropriate, and what is the logic of the strategy by which it can be carried out?

3 Convolutional Neural Networks

3.1 Summary

Convolutional neural networks are a particular architecture of artificial neural networks and are a subset of the field of deep learning [2] [3], in which models possessing multiple levels of abstraction learn properties of the world. Their objective is to provide a model and an implementation of a system that can effectively and efficiently learn to extract and represent a high-level description of salient features of a visual scene using artificial neural networks. Once this high-level description is produced, it should then be possible for it to be utilized effectively by an information-processing system. The study of convolutional neural networks began in the late 1980s [4] [5] and the model has demonstrated dramatic successes in recent years in the field of computer vision [6]. They demonstrate some similarities to the human visual cortex, although they also exhibit many differences [7] [8].

3.2 Analysis using the Tri-Level Hypothesis

3.2.1 Implementation

The first level of the tri-level hypothesis refers to the level of implementation; the physical substrate upon which a cognitive system is realized. For a convolutional neural network, as with modern deep learning architectures in general, the substrate is typically a digital computer, relying on a central processing unit which performs varying operations sequentially and a graphics processing unit which performs similar operations in parallel [2]. This level is constructed of electronic components, such as transistors, which implement a level of operation wherein information is represented as binary words and operated on in discrete steps.

3.2.2 Representation and algorithm

Convolutional neural networks consist of groups of *artificial neurons* called *layers* organized in a particular fashion. Artificial neurons operate by applying various mathematical operations to transform their inputs. Within each layer there exists a set of *feature detectors*, each of which responds to the presence of a particular pattern in an image. Each feature detector is applied at every location in the image. This application of identical feature detectors across the image is referred to as *weight sharing.* At each location, the extent to which the feature is present is calculated by using the mathematical operation of *convolution.* To represent the information more compactly, *dimensionality reduction* techniques which aim to extract the most relevant components are utilized between layers; usually *max pooling* layers [4], which only consider the most prominent feature within a particular location, or, more recently, *strided convolutions* [9], which spatially separate the application of the feature detectors in such a way as to reduce the dimensionality of the output by reducing the amount by which they overlap.

3.2.3 Computational theory

Having discussed the level of implementation and the level of representation and algorithm, we can now turn to the high-level computations that the convolutional neural network architecture addresses. In doing so, we can discuss an abstraction that is separate from the prior levels and could potentially be realized by alternative lower-level architectures.

The key properties that are responsible for the effectiveness of the convolutional neural network model are the following insights, which are general statements about regularities present in natural scenes:

- The world contains many statistical regularities
- Similar patterns tend to appear in multiple regions of space
- Natural scenes have a hierarchical, compositional structure

I will now discuss these properties in more detail. The presence of statistical regularities is important for any algorithm which intends to encode information efficiently, since the absence of such a property would imply that no general encoding could exist that would be more compact than the entities which it aims to represent. Convolutional neural networks are one instance of a broader family of algorithms which aim to encode visual information effectively.

Convolutional neural networks maintain multiple levels of description of the image that they are processing, with each layer representing a level of abstraction. Within each layer is a set of feature detectors, and on higher-level layers, these feature detectors operate on the output of previous feature detectors; hence they are hierarchical and compositional.

The weight-sharing within layers reduces sensitivity to shifts in spatial position, which aligns with the observation that similar patterns repeat themselves at different locations in space.

In this analysis of the third level of description in the hierarchy, it is appropriate to note that these are general computational concepts, which do not rely on the specific representational, algorithmic and implementational details described in the preceding levels.

In fact, we can imagine other implementations of the preceding levels which would be able to achieve functionally equivalent properties on the computational level. For instance, in his afterward to Marr's *Vision* [1], Poggio describes attempts at using probabilistic models referred to as directed graphical models [10] [11] for this purpose. Recurrent neural networks [2] and information theoretic methods [12] [13] are also active areas of research on other formal structures that could be used to achieve similar objectives.

4 On the Need for a Fourth Level: Learning

4.1 Overview

In his afterward to Marr's *Vision* [1], Poggio argued for the inclusion of learning as the very top level of understanding, above the computational level. He boldly pronounced, "Only then may we be able to build intelligent machines that could learn to see–and think–without the need to be programmed to do it." Taking this argument further, he continued, "One could even argue that a description of the learning algorithms and their a priori assumptions is deeper and more useful than a description of the details of what is actually learned." He then stated provocatively: "Learning, I think, should have been included explicitly in Turing's operational definition of intelligence–his famous Turing test." He outlined the situation very clearly when he then explained, "Of course it is important to understand the computations and the representations used by the brain [...] but it is also important to understand how an individual organism, and in fact a whole species, learns and develops them from experience of the natural world."

We can find a stark example supporting this point of view in the convolutional neural network model. The artificial neurons in a convolutional neural network model are initialized with random parameters, which are the numbers that describe the specific computations which are performed. This initialization will perform poorly if applied directly to any problems without training. Hence, a vital element of the architecture is the learning process by which the model learns *the right parameters* to perform well.

4.2 Paradigms of Learning

In this section, I will briefly present three major paradigms of learning that are applicable to convolutional neural networks, in order to better clarify how learning takes place in support of the argument for a fourth level of analysis.

1. **Backpropagation:** Backpropagation [14] [4] is a supervised learning method that uses labeled examples, and computes the difference between the network's output and the target output, and iteratively makes small adjustments to the network until its predictions better match the labeled examples.

2. **Evolutionary algorithms:** Evolutionary algorithms [15] are an optimization method inspired by evolution in nature, whereby a population of slightly different individuals evolves and reproduces according to its fitness when applied to a particular problem. The concept can be applied to convolutional neural networks by encoding their structure and weights into a genetic algorithm.

3. **Reinforcement Learning:** Reinforcement learning [16] studies the problem of learning to assign credit to different causal elements of an architecture based on reward signals received from an environment. Convolutional neural networks can be trained using reinforcement learning [17] [18].

5 Objections

One potential objection to the argument for a fourth level of description is that the paradigms of learning could be sufficiently described using the algorithmic and computational levels of description. For instance, each of the three examples above refer to a specific learning algorithm, which could potentially be included as part of the second level of description in Marr's tri-level hypothesis; and, each of these learning algorithms in turn, are a specific implementation of a broader computational concept, such as supervised learning. I can see the motivation for this objection, since it preserves the structure of the tri-level hypothesis; however, I think that separating learning into an additional level of analysis is potentially a useful and productive approach, by clarifying its foundational role in the success of a cognitive architecture.

Another objection is that the tri-level hypothesis as a whole only applies to a particular type of system which is sufficiently modular [19], consisting of clearly distinct components. However, it seems that with careful thinking, systems that are non-modular could also be adequately analyzed at these levels of abstraction, since the functional properties of such a system could still be differentiated from their component parts.

A final objection to the tri-level hypothesis is the issue of computational tractability [19]. The algorithms described in the second level must be efficient in the sense that they must be implementable in such a way as to yield useful results in practice. This requires that the underlying components on the first level be capable of doing so. However, as Turing [20] showed, there are broad classes of equivalences of machines that could implement computation, and the field of computational complexity shows that there are broad classes of equivalent complexity, which weakens the strength of this objection.

6 Conclusion

I have presented Marr's tri-level hypothesis regarding multiple levels of analysis of cognitive systems, and applied this framework to the analysis of the convolutional neural network architecture of vision from the field of deep learning. I have identified the ways in which this framework is appropriate for such an analysis, and have also argued that it may be appropriate to introduce a fourth level of description to address the foundational role of learning in cognitive systems as suggested by Poggio.

References

[1] David Marr. “Vision: A computational investigation into the human representation and processing of visual information, henry holt and co”. In: *Inc., New York, NY* 2 (1982).

[2] Jürgen Schmidhuber. “Deep learning in neural networks: An overview”. In: *Neural Networks* 61 (2015), pp. 85–117.

[3] Yoshua Bengio, Ian J Goodfellow, and Aaron Courville. *Deep Learning.* 2015.

[4] Yann LeCun et al. “Gradient-based learning applied to document recognition”. In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2323. ISSN: 00189219. DOI: 10.1109/5.726791. arXiv: 1102.0183.

[5] Y. LeCun et al. “Backpropagation Applied to Handwritten Zip Code Recognition”. In: *Neural Comput.* 1.4 (Dec. 1989), pp. 541–551.

[6] Olga Russakovsky et al. “ImageNet Large Scale Visual Recognition Challenge”. In: *International Journal of Computer Vision* 115.3 (2015), pp. 211–252.

[7] David H Hubel and Torsten N Wiesel. “Receptive fields and functional architecture of monkey striate cortex”. In: *The Journal of physiology* 195.1 (1968), pp. 215–243.

[8] Fabio Anselmi and Tomaso Poggio. *Representation learning in sensory cortex: a theory.* Tech. rep. Center for Brains, Minds and Machines (CBMM), 2014.

[9] Jost Tobias Springenberg et al. “Striving for Simplicity: The All Convolutional Net”. In: (2014). arXiv: 1412.6806. URL: http://arxiv.org/abs/1412.6806.

[10] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques.* MIT press, 2009.

[11] Pedro Domingos. *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World.* Basic Books, 2015.

[12] Ray J Solomonoff. “A formal theory of inductive inference. Part I”. In: *Information and control* 7.1 (1964), pp. 1–22.

[13] Marcus Hutter. *Universal Artificial Intelligence.* Vol. 1. 2. 2005, pp. 1–82. ISBN: 9783540221395. DOI: 10.1145/1358628.1358961.

[14] David E. Rumelhart, Geoffrey E. Hinton, and Ronald J. Williams. “Learning Internal Representations by Error Propagation”. In: (1985).

[15] Melanie Mitchell. *An Introduction to Genetic Algorithms.* MIT Press, 1998, p. 209. ISBN: 0262631857.

[16] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction.* Vol. 1. 1. MIT press Cambridge, 1998.

[17] Martin Riedmiller. *Neural fitted q iteration first experiences with a data efficient neural reinforcement learning method.* Vol. 3720. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, pp. 317–328.

[18] Volodymyr Mnih et al. “Human-level control through deep reinforcement learning”. In: *Nature* 518.7540 (2015), pp. 529–533.

[19] José Luis Bermúdez. *Cognitive science: An introduction to the science of the mind.* Cambridge University Press, 2014.

[20] Alan M Turing. “Computing machinery and intelligence”. In: *Mind* 59.236 (1950), pp. 433–460.