Current Biology 24, R921–R929, September 22, 2014 ©2014 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2014.08.026

Neural Networks and Neuroscience-Inspired Computer Vision

Review

David Daniel Cox[1,2,3,*] and Thomas Dean[4,5]

Brains are, at a fundamental level, biological computing machines. They transform a torrent of complex and ambiguous sensory information into coherent thought and action, allowing an organism to perceive and model its environment, synthesize and make decisions from disparate streams of information, and adapt to a changing environment. Against this backdrop, it is perhaps not surprising that computer science, the science of building artificial computational systems, has long looked to biology for inspiration. However, while the opportunities for cross-pollination between neuroscience and computer science are great, the road to achieving brain-like algorithms has been long and rocky. Here, we review the historical connections between neuroscience and computer science, and we look forward to a new era of potential collaboration, enabled by recent rapid advances in both biologically-inspired computer vision and in experimental neuroscience methods. In particular, we explore where neuroscience-inspired algorithms have succeeded, where they still fail, and we identify areas where deeper connections are likely to be fruitful.

Introduction

The human brain is a staggeringly complex computational system, consisting of some 100 billion neurons, connected by an estimated 100 trillion synapses [1]. The brain allows us to make sense of a complex and ever-changing sensory world, to plan complex actions, to navigate our social environment and intuit the minds of others, and to learn and remember across our entire lifespans. It can be said, without exaggeration, that the complexity of our brains has given rise to every aspect of our collective civilization and our technology. In many ways, the brain represents one of the greatest frontiers in our understanding of ourselves.

Throughout history, our understanding of the brain, and the language we use to describe it, has leaned heavily on the language and understanding of our contemporary man-made technologies. Descartes explained the mind in terms of hydraulic analogies and the movement of fluids [2]. To Freud the brain was like a steam engine, distributing and releasing pressure [3]. In the era of radio, brains were increasingly described in terms of 'channels' and frequencies. Perhaps not surprisingly, today we also use the language of modern-day technologies. Neuroscientists increasingly speak of neuronal 'computations' and the 'circuits' responsible for behaviors; distant brain regions communicate to form 'networks' of activity.

For all of the superficial resemblance between silicon computers and brains, the two systems are in many ways a study in contrasts. The individual computational elements in man-made silicon computers typically run at amazingly fast clock speeds, up to billions of cycles per second, with electrical signals being transmitted from one part of the chip to another at nearly the speed of light. Meanwhile, biological neurons are positively sluggish by comparison. Action potentials in mammals propagate at speeds ranging from less than 1 m/s, up to just over 100 m/s [4]. As a result, visual signals from the eye would take on the order of tens of ms to traverse from one side of the brain to another [5]. However, this number is slowed even further by the delays of synaptic transmission in intervening synapses along the way. With all of these delays, signals don't reach primary visual cortex in humans until around 50 ms after photons reach the retina [6,7]. Signals don't reach later stages of visual processing until almost 200 ms after the retina is stimulated [8]. By the standards of a silicon computer, such propagation times are glacially slow. However, what brains lose in the speed of individual elements, they potentially make up for in parallelism and connectivity. While an advanced GPU might have thousands of processing cores that operate on data in parallel [9], the brain has billions of neurons operating simultaneously. Moreover, while our fastest parallel computing architectures today are primarily limited by their ability to move the right data to the processors at the right time to serve a given algorithm [10–12], the human brain is densely interconnected, with its billions of neurons sending signals to one another across a network containing trillions of connections. The sheer number of these connections, and their structure, allow information to rapidly flow from one part of the brain to another, often requiring only a few synaptic steps to span between distant brain regions [13]. Strikingly, our brains perform their incredible feats while only consuming about 20 watts of power [14] — roughly the power consumption of an average laptop.

Yet, while it would be easy to dismiss the 'computational' perspective on neuroscience as another passing metaphor, it is a metaphor that runs deeper: beyond the metaphor of 'the brain is a computer', computational science provides a rigorous formal framework and tools for reasoning about information-processing systems, separating what gets computed ('algorithm') from how it gets computed ('implementation').

Furthermore, we live today in a world where enormous computational power is available. With the advent of the internet, we routinely interact with vast networks of computers and we possess technologies to harness the collective power of massive server farms. Several groups have launched large, multinational efforts to simulate parts of or whole brains *in silico* [15]. While such efforts are surrounded by controversy about whether they are biologically realistic or focus at the appropriate level of biological detail [16], the exponential nature of the growth of computing power makes it entirely plausible that we'll soon be able to routinely marshall computing power rivaling or exceeding that of the brain. Meanwhile, a number of groups are working on producing silicon architectures whose elemental building blocks work more like neurons [17,18]. Barring big surprises

[1]Center for Brain Science, Harvard University, Cambridge, MA 02138, USA. [2]Department of Molecular and Cellular Biology, Harvard University, Cambridge, MA 02138, USA. [3]School of Engineering and Applied Science, Harvard University, Cambridge, MA 02138, USA. [4]Google Research, Mountain View, CA 94043, USA. [5]Department of Computer Science, Brown University, Providence, RI 2912, USA.
*E-mail: davidcox@fas.harvard.edu

in the fundamental nature of neuronal function, we can contemplate a world where simulating an entire brain becomes commonplace.

However, even as the barriers of raw computational power fall away, knowing what to do with all of that power is a greater problem. Despite significant progress in neuroscience, we still know little about how brain circuits organize themselves to give rise to behavior and learning. In the absence of a clear mandate for what to build, the interaction between neuroscience and biologically inspired computing has been a co-evolution, with each field providing tantalizing, but ultimately incomplete, clues to the other.

Here, we review the interplay between neuroscience and computing, focusing on connections between visual neuroscience and the fields of computer vision and machine learning, with particular attention to visual object recognition, where the recent progress has been especially quick. In many ways, vision lies at the leading edge of both neuroscience and machine perception; we arguably know more about the brain's visual system than we know about almost any other brain subsystem, and computer vision has played a leading role in the development of machine learning, machine perception, and biologically inspired computing in general [19]. While a full exploration of all connections between neuroscience and computer science is beyond the scope of the present article, vision in general, and object recognition in particular, nonetheless provides an interesting test case in the intersection of neuroscience and computing. Here, we explore the past and present of this interface, and we suggest possible avenues for future cross-pollination.

A Brief History of the Artificial Neural Network

The history of biologically inspired algorithms stretches surprisingly far back into the history of computing. McCulloch and Pitts formalized the notion of an 'integrate and fire' neuron in 1943, and Hebb first proposed the idea of associative learning in neurons — "what fires together, wires together" — in the late 1940s [20,21]. Meanwhile, the transistor was only invented in 1947; practical integrated circuits emerged only in the late 1950s; mainframes and 'minicomputers' were not commonplace until the 1960s; and personal computers did not appear until the 1980s and 1990s. That theory would precede practical application by so many years is a testament to the foresight of these early pioneers.

One of the earliest instantiations of a neural network that could learn was the 'perceptron' of Rosenblatt [22–24], who proposed a simple arrangement of input and output neurons that could make decisions on the basis of input vectors. The initial form of the perceptron proved to be fundamentally limited, only being capable of learning linear functions of the inputs, and neural network research faced a temporary setback at the hands of the rival 'symbolic artificial intelligence' camp, which sought to model intelligence through abstract symbolic operations, rather than drawing direct inspiration from the machinery of the brain.

The addition of nonlinear activation functions and a 'hidden' layer of units between the inputs and outputs of the networks overcame the theoretical limitations of the perceptron, and over the next two decades, a wide range of different forms of artificial neural networks (ANNs) emerged [25]. While the inclusion of a hidden layer made it possible, in principle, for an ANN to compute any function, it was less clear how to train a network to compute an arbitrary function of interest. In the 1960s and 1970s, the back-propagation algorithm [26–28] was introduced, which provided a concrete mechanism for propagating error signals back through a multi-layer neural network. Back-propagation also lacks a clear story connecting it to biology — it is not known how neurons might propagate signals 'backward' through multiple synapses to adjust their strengths. However, it allows networks with hidden layers to be trained efficiently, and this alone was enough to drive its popularity.

Artificial neural networks flourished through the 1980s and optimism ran high. 'Connectionism' became a popular term for describing the study of various kinds of early neural networks aimed at solving a wide range of problems, from vision to language. A host of investigators (e.g., LeCun, Bengio, Hinton, Schmidthuber, to name just a few) made seminal contributions to the state-of-the-art in neural networks, and they were increasingly applied to a range of practical problems. The convolutional neural network emerged as a powerful tool in the analysis of images and played an important role in the young field of computer vision, achieving excellent performance on the problem of hand-written digit recognition, a real-world application of neural networks. Meanwhile, neuroscience has provided guiding force for the development of artificial neural networks, providing inspiration for architectural features of neural networks (e.g., simple-to-complex pooling in Fukushima's neocognitron [29,30]).

Casualties of the A.I. Winter

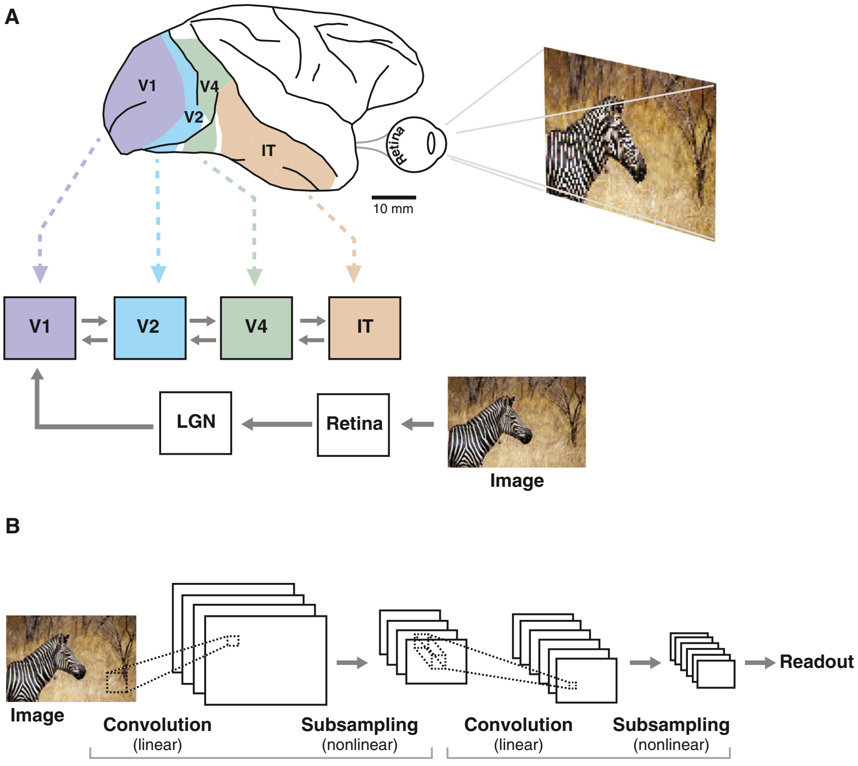
While the 1980s saw enormous enthusiasm and hope around the idea that machines could model and recreate perceptual and cognitive abilities of humans, this enthusiasm waned significantly in the 1990s. The promise (and promises) of the neural network community in the 1980s were great, but in many ways these approaches failed to deliver practical results, as did parallel efforts to model higher-level cognition from the 'symbolic A.I.' camp. This period came to be known as the 'A.I. winter', since it represented a significant cooling off in both interest and funding for both artificial neural networks and symbolic A.I. research.

In the domain of computer vision, while neural networks saw some early successes, a diverse range of conventional, engineered solutions to specific computer vision problems emerged and gained prominence. Many of these approaches could have been tied either implicitly or explicitly to neuroscience ideas, but the community largely eschewed such connections, emphasizing intuitive and theoretical appeal over biological inspiration. For instance, David Lowe's widely influential Scale Invariant Feature Transform (SIFT) was originally described in analogy to the primate ventral visual pathway [31], but although it quickly became a ubiquitous component of conventional computer vision systems, its biological inspiration is rarely mentioned.

In the domain of machine learning, the 1990s saw the development and rise of a variety of machine learning approaches that would supplant the neural network. Support vector machines (SVMs), in particular, offered excellent generalization performance and relative freedom from mysterious and difficult-to-choose training parameters [32]. Because SVMs rely on the mathematics of convex optimization at their heart, they can guarantee efficient arrival at a global optimum, even for problems that are large in the number of training examples. Neural networks, by comparison, required seemingly arbitrary decisions about the number of units in the network, how long to spend training a network, and how big a change to make in the network connection

Figure 1. A rough correspondence between the areas associated with the primary visual cortex and the layers in a convolutional network.

(A) Four Brodmann areas associated with the ventral visual stream along with a block diagram showing just a few of the many forward and backward projections between these areas. (B) A simple feedforward convolutional network [105] in which the two bracketed pairs of convolution operator followed by a pooling layer are roughly analogous to the hierarchy of the biological visual system. Adapted from [106].



Current Biology

weights at each training step. Neural networks came to be painted as slow and finicky to train, beset by voodoo-parameters, and simply inferior to other approaches.

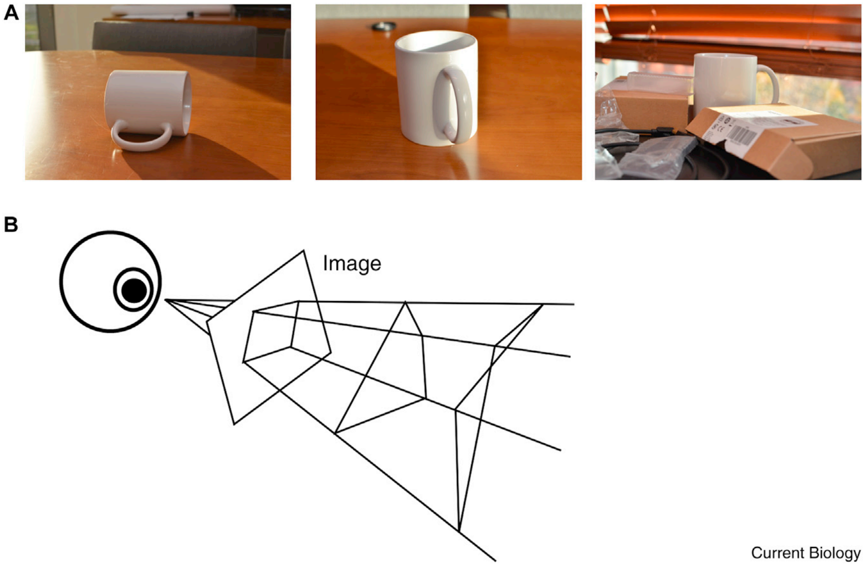
Deep Learning and the Second A.I. Spring

One of the principal limitations of traditional artificial neural networks has been that methods for training a system with multiple layers were not straightforward or not available. From a visual neuroscience perspective, however, the appeal of having multilayer networks is obvious. In primates, the ventral visual pathway (Figure 1A) is thought to subserve visual form and object vision, and it is organized as a hierarchical series of interconnected visual areas. Neurons in early areas, such as area V1, respond to comparatively simple, spatially local features of the retinal image, while later areas, such as area V4 and inferotemporal cortex, respond to increasingly complex visual features over larger regions of visual space. While the exact nature of population representations in the visual cortex are still poorly understood, it is clear that as one progresses along the ventral visual pathway, neurons begin to represent visual objects in a way that is tolerant to variation in the exact appearance of that object on the retina. Because a visual object can be viewed from different vantage points and under different lighting conditions, it can cast an effectively infinite number of different images onto the retina (Figure 2A). The converse is also true: any given image on the retina can correspond to infinitely many possible objects in the world (Figure 2B). The idea that the ventral visual pathway exists to transform images into a better format, one that allows the brain to reason about objects in spite of this level of variation, is an old idea in neuroscience, and one that continues to serve as a foundational working hypothesis in the study of high level vision.

While many artificial neural networks in the 1980s were largely treated as classifiers, responsible for mapping high-dimensional input vectors (e.g., images) onto class labels or some other output function, when seen through the lens of visual systems neuroscience, the role of a visual system is not so much to classify images, but to successively transform images from one format of representation into a different, more flexible one that better reflects the structure of the external world. Rather than simply mapping one function onto another, the goal of a visual hierarchy is to discover latent structure and make it explicit, such that it can be manipulated to serve a number of different tasks. Some aspects of this distinction are largely semantic — any neural network will, by definition, map from one function to another, but a representation-learning perspective dictates a very different set of priorities in the design of a neural network. For one, since this perspective seeks to discover representations of the external world, unsupervised pre-training — using unlabeled examples to train an initial state of a network — makes increasing sense, especially where the network might need to serve multiple end goals. In addition, following a ventral stream-inspired plan places a premium on networks with many layers of processing — so-called ‘deep’ networks.

While the A.I. Winter took a broad toll on research in machine perception and machine intelligence — including neural networks and biologically-inspired vision — many of the original stalwarts of classical neural network approaches, such as Geoff Hinton, Yann LeCun and Yoshua Bengio, continued their work and increasingly rallied around ‘deep learning’ approaches [33]. Researchers working on deep networks began to accumulate a steady stream of practical successes. In the domain of vision, convolutional architectures — which scan a set of filters across an image at each level of a deep hierarchy — proved to be especially effective during this period. These systems have fewer weights to train, work well even when the weights aren’t trained, and naturally capture the spatial stationarity in natural images (a set of similar visual features tend to appear at different spatial locations in an image).

With the introduction of the Restricted Boltzmann Machine and its variants [34], Hinton, Bengio and their collaborators



Current Biology

Figure 2. The robustness of the human visual system.

(A) We are capable of recognizing objects across a wide variation in pose, lighting conditions and partial occlusion, and (B) we deal effortlessly with the ambiguities that routinely arise in the process of projecting 3-D images on the approximately 2-D retina. As an illustration, we note that the image cast on the retina when viewing a simple line drawing is consistent with an infinite number of wire frame objects (B), and yet we have no trouble making the right interpretation in all but contrived situations.

revived interest in multilayer perceptrons, but the training procedures tended to be complex, and the update rules, such as contrastive divergence, lacked clear theoretical grounding. A high-visibility publication by Hinton and Salakhutdinov in the journal *Science* added credibility to the field [35], but still only a relatively small community was working in the area.

Meanwhile, neuroscience made important theoretical moves forward during this period, providing new clues as to how the earlier generation of neural networks could be improved: max pooling has all but replaced winner-take-all in pooling layers, thereby providing impressive performance gains [36]; surround suppression in classical receptive fields is routinely applied in the form of local non-max suppression for edge and contour detectors and localization in object recognition [37]; rectified linear units have substantially outperformed sigmoidal activation functions to obtain the best results in several benchmark problems [38]; and local (divisive) normalization which appears to operate in a number of neural systems [39] applied in the form of local contrast normalization is one of the most important components in state-of-the-art object recognition systems [40].

The rapid ascendance of deep learning approaches reached critical mass in 2013, when Hinton and colleagues demonstrated a deep network that yielded exceptional performance on the ImageNet object classification challenge data set. Variants on this network would subsequently be applied to a wide range of different problems, scoring top results effectively everywhere it was applied [41]. While it can be difficult to quantitatively chart the ups and downs of scholarly zeitgeist, in the case of the new field of deep learning, its popularity can be measured in dollars, with industry titans such as Google, Facebook, and Baidu hiring up a significant fraction of the field of experts in deep learning, with vast sums of money changing hands. Arguably never before has such a large fraction of a research community been so rapidly privatized, resulting in shockwaves through the field.

While it would be appealing to paint the rise of deep learning in terms of some key breakthrough theoretical advance, in truth, some of the most successful deep learning systems are not so different from the back-propagation networks of the 1980s. Certainly, theoretical advances have been made, but in large part the enabling factor in the latest deep learning is the availability of computational power and of vast quantities of data. Google and Facebook each handle enormous volumes of images (e.g., 100 hours of video are uploaded to YouTube every minute). This provides an unprecedented pool of data to use in training networks, and models can now be trained on datasets orders of magnitude larger than previously available. Meanwhile, GPUs have made certain key computations necessary for deep learning approaches very fast, particularly convolutions. Modern data centers also made it possible to train many models simultaneously, and thus search the space of models more effectively.

Simply put, the community seized the opportunity presented by advances in hardware, figured out efficient numerical recipes for performing algorithms like back-propagation effectively, and were able to perform thousands of experiments quickly. Over a short period of time, neural network models went from an obscure and maligned artifact of the past to a dominant force in nearly every field of machine learning and perception. A godfather of this field, Geoff Hinton, is fond of saying that it took 17 years to get deep learning right; one year thinking and 16 years of progress in computing, praise be to Intel.

Do Neuroscience and Deep Learning Still Need Each Other?

Given the current excitement surrounding modern deep learning approaches, an obvious question that one might ask is whether machine learning still needs anything from neuroscience. Certainly, the gains of deep learning approaches have been impressive, and there is a great deal of enthusiasm that deep learning approaches will continue to overcome many current problems in machine learning. Meanwhile, the flow of ideas from neuroscience to computer science has been sporadic, and not always responsible for the greatest progress. Against this backdrop, it might be easy to assume that machine learning doesn't need neuroscience anymore. Indeed, Yann LeCun, a major player in the new A.I. Spring, has even been recently quoted as saying that while we can get inspiration from the biology, we shouldn't be blinded by it.

While there is much cause for optimism for deep learning, there is also substantial evidence that should temper this enthusiasm. While deep learning approaches are beginning to rival human performance in certain situations, the gap between humans and machines is still great. One important divide between humans and current deep learning systems is in the size of required training datasets. Humans and animals can rapidly learn concepts, often from single training examples. Studies of human concept learning show that humans can accurately learn complex visual object categories from fleeting numbers of examples [42,43]. In contrast, current deep learning approaches require vast quantities of data to work. For instance, the Krishevsky *et al.* model that was used to achieve high levels of performance in the ImageNet challenge [38] was trained using 1,000 labeled examples each from 1,000 categories of objects, for a total of 1 million labeled images. This number begins to approach the scale of the number of visual fixations a human makes in a year (assuming three saccades per second during waking hours), and only a fleeting few of those fixations could be counted as being 'labeled' in any sense.

A related issue for modern computer vision is out-of-set generalization. One fundamental challenge in computer vision is in evaluating performance. Computer vision performance is typically assessed against benchmark datasets. However, Torralba and Efros [44] have elegantly shown that most systems trained on one data set perform better on that data set than on another that contains the same categories of objects, suggesting some degree of bias in these benchmark datasets. The more effective tests of deep learning approaches will come when they are deployed in real operational settings and real applications. Tests of this sort are underway, albeit largely in commercial settings where the resulting performance data may not always be made available.

Moreover, most computer vision test benchmarks are examples of so-called 'closed-set' problems — problems where all of the classes that a system will encounter are known in advance. Thus when building a system that identifies the category 'cars', we not only have a number of labeled examples of cars, but we also have a large number of examples of the other categories that we might encounter (e.g., faces, people, houses, etc.). In contrast, in the real world, we rarely enumerate all of the possible negative classes of objects that we might encounter. Indeed, a large fraction of patches of the visual world that we encounter cannot even be unambiguously labeled. Attempts to come to terms with such 'open-set' problems shows them to be much more difficult. Indeed, even the venerable, largely solved MNIST hand-written digit recognition dataset [45] (current systems achieve in excess of 99% accuracy [46]) becomes difficult when the system doesn't have access during training to examples of all of the possible digits it might encounter during testing [47]. This concern is also echoed in large-scale object datasets — the ImageNet challenge includes a 'detection' variant of the challenge wherein objects must be located within images [48]. Because the negative set of image patches that the system must reject includes an enormous diversity of image content, not all of which can be easily labeled, it is effectively an open-set problem. Performance on this variant protocol for ImageNet is still uniformly poor for current artificial systems [49]. Several computational efforts have sought to break free of this mold, tackling extremely large numbers of categories by treating object recognition as a mapping rather than a classification problem [50]; however, such approaches today remain the exception rather than the rule.

Other, more subtle signs of trouble also exist in the deep learning literature. For instance, Szegendy and colleagues [51] showed that one can add carefully crafted 'noise' to images and cause them to be arbitrarily misclassified by a current deep learning system. While the original image and the altered image are classified as completely different objects by the deep learning system, they are effectively indistinguishable by humans (they would be considered to be 'metamers' in the language of visual psychophysics [52]). This suggests that the nature of representations in humans and deep neural networks are still qualitatively different. Taken together, the gap between the performance of artificial and biological systems suggests there is more that neuroscience can teach deep learning.

Forging New Links

So what's next? How can the neuroscience, computer vision, and machine learning communities communicate more effectively with one another?

Historically, the dialogue between neuroscience and machine learning has been hampered by limitations in technology and differences in culture. On the one hand, neuroscience has historically lacked experimental tools that could provide new inspiration and constraint for existing neural network architectures. Electrophysiology, a gold-standard in characterizing neuronal responses, has typically only allowed relatively brief sessions with cells, and it has not generally been possible to target the same cells across long spans of time. Using traditional tracing techniques, the connectivity between cells could only be probed sparsely, providing comparatively macroscopic information about the projections of neurons between areas, but providing little detail about the fine-grained organization of brain circuitry.

On the other hand, following an initial flurry of excitement, neural network approaches lagged behind other machine learning methods for decades. While various neuroscience-inspired vision models have been fit to neuroscience data, given the paucity of data to fit and the large number of free parameters to be fit, it is difficult to draw strong conclusions from such an exercise, since one would expect almost any sufficiently expressive model to be able to fit the experimental results. While such efforts are clearly important, few testable predictions have emerged to date. Further, it is unclear what conclusions to draw from models that can explain neuronal firing rates, but can't reproduce the function of the larger system. If we built a model that could explain some small number of measurements taken from the inside of a car engine, but the model car itself was unable to operate, then we are left in uncertain territory on how to interpret the model.

Today, both of these barriers have been removed. Increasingly, neuroscience tools give access to the activity of large populations of cells [53,54], and the same cell across large spans of time [55]. We can directly measure the connectivity between identified cells, and genetic tools give us unprecedented cell-type-specific access to neuronal networks, with the ability to measure, stimulate, and silence cells with exquisite precision [56,57]. Meanwhile, deep neural networks have become a dominant approach in many machine learning domains [58], and high performance computing tools give us the power to test new ideas at scale. This creates several promising avenues for further exploration.

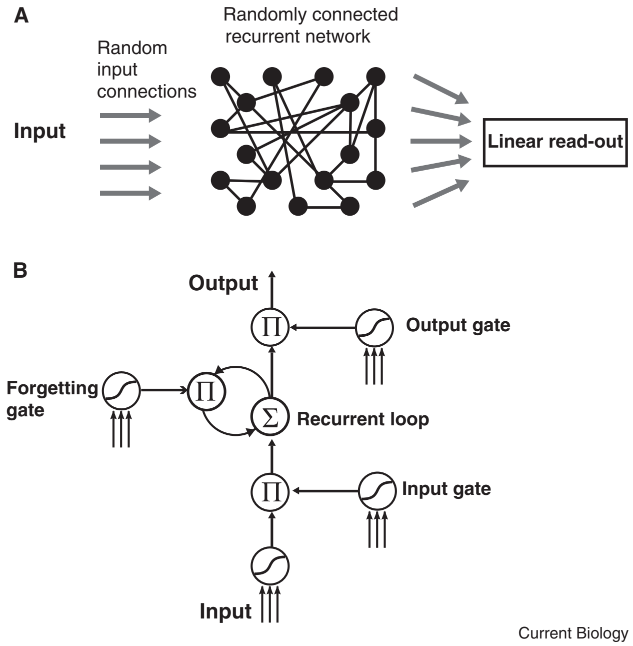


Figure 3. Examples of extant recurrent network architectures.

(A) 'Reservoir computing' avoids the difficulties of training recurrent architectures by feeding inputs into randomly connected recurrent networks and then reading out a result via a simple linear learning rule. Such architectures have found uses in a variety of applications with time-varying signals. (B) Another kind of recurrent network that is gaining popularity is the 'long-short-term memory' network. The detailed function of such networks is beyond the scope of the present article, but at a high level the network allows information to be stored and accumulated in a recurrent loop, while multiplicative 'gates' control the flow of information into, out of, and within the loop. Networks of this sort have the ability to learn over long timescales and to produce outputs at irregular time intervals.

Recurrence, Feedback, and Dynamics

At their root, the current crop of successful deep networks are actually quite simple in their final structure — most are still largely feedforward in their organization, and many of the key operations can be conceptualized as a generalized convolution — a computation that operates on a local neighborhood of inputs. Meanwhile, we know that the real visual cortex is quite a bit more complex. Local cortical microcircuits contain myriad local recurrent connections and ubiquitous feedback connections between cortical areas, not to mention subcortical loops [59] and long-range modulatory connections (e.g., [60]). Visual cortex is organized into six cell layers with stereotyped patterns of connectivity, collectively comprising at least dozens of genetically distinguishable cell types, which presumably subserve distinct functions within the network [61]. We don't argue here that artificial networks need to slavishly copy this complexity to be computationally 'like' the brain or to be useful for machine vision. Indeed, from an applied, engineering perspective, copying superficial features of the architecture of cortex is not necessary, and might even be distracting from the goal of building better vision systems. However, the extent of the complexity found in the brain relative to artificial networks suggests qualitative, rather than quantitative, gaps that need to be spanned, or at least, understood.

This is not to say that no work has been done on networks that include recurrence or feedback. Several major families of current deep neural networks, such as Restricted Boltzmann Machines, incorporate algorithmic forward and reverse passes to perform inference, and these provide some concrete hypotheses for roles that feedback might play in real neuronal networks [34,62,63]. Meanwhile, recurrent neural networks (RNNs; e.g., Figure 3), which contain loops in their connectivity graph, have long been a topic of study [64,65]; though, with a few exceptions, they have proven more difficult to train than feedforward networks. RNNs allow for the incorporation of feedback and support models that essentially remember the results of prior computations and are capable of establishing long-range temporal dependencies within visual, auditory and text data [66,67].

A variety of functional roles have been proposed for recurrent connections, by both the neuroscience and computer vision communities. One natural idea is that recurrence enables contextual information to be incorporated to enhance otherwise ambiguous inputs. Humans are able to recognize highly degraded images when external context provides additional clues. Incorporating such context, for instance in a Bayesian framework, is a popular idea, although one that remains to be fleshed out. Similarly, a variety of models posit specific roles for top-down feedback connections in allocating attention to different parts of a scene [68]. New tools in neuroscience increasingly provide experimental access to study these connections directly. For instance, viruses now exist that can jump across a single synapse [69], delivering genetically encoded indicators and opsins that enable the activity of neurons that provide input to a given target to be measured and manipulated. We believe that coming to terms with the nature of these connections and their roles will be one area where neuroscience and computer vision might enjoy special synergy.

Another elephant in the room is the role of spiking in neuronal information processing. Real neuronal systems exchange information through a chatter of discrete action potentials, or 'spikes'. However, the current wave of deep learning success does not include any notion of spiking, instead propagating scalar-valued 'activation' through the network in discrete time steps. Even within neuroscience, while no one doubts that there are many timing-dependent phenomena in neurons (such as spike timing-dependent plasticity [70]), there remains substantial debate about whether understanding detailed spike timing is critically important to understanding sensory coding, or whether slower timescale rate codes suffice [71]. Such concerns become even more pronounced when considering recurrent networks, and a growing subfield of theoretical neuroscience is using the tools of dynamical systems and statistical mechanics to describe and understand the behavior of populations of interconnected spiking neurons [72]. While it is safe to say that spiking networks have not participated as top performers in machine vision at any point in history to date, this could easily change as theory and available computational power catch up.

Beyond Still Images

Another clear area for growth in neuroscience-inspired computer vision is in the processing of time-varying images. To date, many of the greatest successes in computer vision and object recognition have been with still images — this is perhaps not surprising given that we've only just now attained a level of computational power to handle images effectively, and video multiplies the scale of raw data involved. There are other practical challenges with video. Still images on the web are often accompanied by linguistic cues in the form of anchoring text linked to photos or captions in the case of figures in more traditional documents. Video on the other hand, especially the wild type found on such sites as YouTube, often has little or no annotation and what annotation it does have, for example comments written by visitors to the website, are ambiguous, often spurious and generally have little to do with the specific visual categories shown in the frames of the video. This makes it difficult to train systems that require a great deal of supervision. The analysis of motion in video has been shown to be of value in several areas, such as categorizing different kinds of human action in video [73–75]. The analysis of time-varying signals for object perception has long been the subject of convergent interest across computational neuroscience and computer vision [76–86], but we argue that there is a good deal of room to make these connections stronger going forward.

Real neuronal systems perform a variety of different kinds of temporal filtering and processing on incoming inputs, and the visual system is no exception. Neurons in area V1, for instance, have obvious tuning in both space and time, with many responding optimally to moving edges [87]. Some visual areas, such as area MT, are clearly specialized for analyzing motion [88], but even ventral stream areas that are thought to be involved in representing object form show interesting temporal structure in their responses. For example, the responses of neurons in inferotemporal cortex, at the end of the ventral visual hierarchy, show complex patterns of adaptation based on previously seen images [89]. In an extreme example, Meyer *et al.* showed that responses of inferotemporal neurons to particular stimuli could be nearly completely suppressed if the animal learned to expect their appearance in a particular ordered sequence; these same stimuli evoked strong responses when seen out of the learned order [90]. This suggests that object processing in visual cortex is sensitive to temporal contingencies, though systems neuroscience has only begun to scratch the surface of understanding what role time plays in the ventral pathway, and little is known about what mechanisms might underlie these phenomena.

On the machine learning side, various kinds of recurrent networks are enjoying the beginnings of a resurgence in interest for temporal learning. For instance, a conceptually simple framework known as ‘reservoir’ computing [91] (Figure 3A), which uses unstructured, randomly connected recurrent networks, paired with a simple linear read-out, has been applied to a variety of temporal recognition problems outside of vision with a surprising degree of success (though effective applications to vision are still rare). In addition, more complex networks, such as long-short-term memory models (LSTMs) are increasingly showing impressive performance in tasks that require sequence learning [92–94]. Even more promising, early work on LSTMs is already feeding back onto neuroscience, having spawned several theories about neural structures responsible for sequence learning involving prefrontal cortex and the basal ganglia [95,96]. Another promising point of contact between neuroscience and computation in the context of temporal data streams is in the simulation of eye movements to simulate focused serial sampling of an otherwise cluttered and difficult to parse scene [97–100]. A number of efforts to utilize object tracking in conjunction with object recognition have been proposed.

Towards Better Representation Learning

A dominant theme in the recent resurgence of neural networks has been the importance of learning good, flexible representations of the external visual world. While from an engineering perspective there is no strict requirement that the representations found in artificial networks be similar to those found in nature, we argue that biology provides a potentially exciting and rich source of ideas for representation learning. Importantly, neuroscience increasingly has tools that allow large populations of neurons to be monitored over long periods of time, offering hope that we can begin to ‘watch’ biological learning in progress as an animal learns to perform a given task. Such efforts would provide both static snapshots of the properties of how ‘good’ representations are organized, along with dynamic information about what kinds of learning rules might give rise to them. With the BRAIN initiative fueling even greater interest in large-scale methods for recording neuronal activity [101], the tools for undertaking such work will only get better.

Beyond Visual Cortex

Most current artificial neural networks for vision exist as isolated visual systems, which take in an image as input, and output a category label or a vector representation that can be given to a classifier to provide a label. However, real visual systems do not exist in a vacuum, but rather exist integrated into larger networks concerned with guiding motor action, monitoring and distributing signals about reward value, and integrating disparate senses together. Several efforts are underway to study vision in the context of larger networks [102,103] that include other components such as working memory, retinas and/or attentional spotlights that can move to sample different portions of the scene, and motor effectors that allow the system to interact with the environment. We know from decades of neuroscience research that brains devote vast networks of neuronal hardware to driving such active feedback loops, and we know that the de facto activity of ventral visual cortical responses are heavily shaped by saccadic eye movements during natural viewing behavior. Incorporating active sensing and flexible task requirements will no doubt shape the nature of representations in the deep learning systems, and it represents a promising direction for interplay between neuroscience and computer vision.

Conclusion

While the interchange of ideas between neuroscience and computer vision has experienced ups and downs, it is hard not to be enthusiastic about the future of neuroscience-inspired computer vision. In many ways, the current environment is a perfect storm of opportunity, with recent successes in machine learning and recent advances in neuroscience technology coinciding almost perfectly, and with the two fields perhaps poised to take advantage of each other’s insight at an unprecedented scale. However, seizing this opportunity will require effort and a cultural shift, as the two fields often have very different goals and approaches. With elevated enthusiasm also come elevated expectations; the broader field of brain-inspired A.I. has already gone through one boom–bust cycle, and some observers worry that hopes are already running too high [104]. The cyclic nature of passing academic trends may very well be unavoidable, but we argue that in many ways we stand at a very different place today relative to the beginning of the first A.I. Winter. For one, in contrast to the first A.I. Winter, where the 'product' being sold in commercial contexts was arguably largely hype, today, the current crop of neural networks are being used to solve a wide variety of real-world problems of core interest to companies like Apple, Facebook, Google, IBM, and Microsoft. But perhaps even more salient, never before have the fields of neuroscience, computer vision, and machine learning had so much to say to one another. The trick will be making sure that we listen.

References

1. Herculano-Houzel, S. (2012). The remarkable, yet not extraordinary, human brain as a scaled-up primate brain and its associated cost. Proc. Natl. Acad. Sci. USA *109* (Supp 1), 10661–10668.
2. Stanford encyclopedia of philosophy. http://plato.stanford.edu/entries/pineal-gland. Accessed: 2014/07/14.
3. Leary, D. (1994). Psyche's muse. In Metaphors in the History of Psychology, D. Leary, ed. (Cambridge: Cambridge University Press).
4. Rushton, W. (1951). A theory of the effects of fibre size in medullated nerve. J. Physiol. *115*, 101.
5. Ogden, T.E., and Miller, R.F. (1966). Studies of the optic nerve of the rhesus monkey: nerve fiber spectrum and physiological properties. Vis. Res. *6*, 485–506.
6. Vanni, S., Warnking, J., Dojat, M., Delon-Martin, C., Bullier, J., and Segebarth, C. (2004). Sequence of pattern onset responses in the human visual areas: an fMRI constrained VEP source analysis. Neuroimage *21*, 801–817.
7. Di Russo, F., Martínez, A., Sereno, M.I., Pitzalis, S., and Hillyard, S.A. (2002). Cortical sources of the early components of the visual evoked potential. Hum. Brain Mapping *15*, 95–111.
8. Bötzel, K., Schulze, S., and Stodieck, S.R. (1995). Scalp topography and analysis of intracranial sources of face-evoked potentials. Exp. Brain Res. *104*, 135–143.
9. Lindholm, E., Nickolls, J., Oberman, S., and Montrym, J. (2008). Nvidia tesla: A unified graphics and computing architecture. IEEE Micro *28*, 39–55.
10. McCalpin, J.D. (1995). A survey of memory bandwidth and machine balance in current high performance computers. IEEE TCCA Newsletter, 19–25.
11. Owens, J.D., Houston, M., Luebke, D., Green, S., Stone, J.E., and Phillips, J.C. (2008). GPU computing. Proc. IEEE *96*, 879–899.
12. Armbrust, M., Fox, A., Griffith, R., Joseph, A.D., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I., *et al*. (2010). A view of cloud computing. Commun. ACM *53*, 50–58.
13. Bassett, D.S., and Bullmore, E. (2006). Small-world brain networks. Neuroscientist *12*, 512–523.
14. Mink, J.W., Blumenschine, R.J., and Adams, D.B. (1981). Ratio of central nervous system to body metabolism in vertebrates: its constancy and functional basis. Am. J. Physiol *241*, 203–212.
15. Markram, H. (2012). The human brain project. Sci. Am. *306*, 50–55.
16. Open message to the european commission concerning the human brain project. http://neurofuture.eu. Accessed: 2014/07/14.
17. Boahen, K. (2005). Neuromorphic microchips. Sci. Am. *292*, 55–63.
18. Indiveri, G., Linares-Barranco, B., Hamilton, T.J., Van Schaik, A., Etienne-Cummings, R., Delbruck, T., Liu, S.-C., Dudek, P., Häfliger, P., Renaud, S., *et al*. (2011). Neuromorphic silicon neuron circuits. Front. Neurosci. *5*, 73.
19. Prince, S. (2012). Computer Vision: Models Learning and Inference (Cambridge: Cambridge University Press).
20. McCulloch, W.S., and Pitts, W.H. (1943). A logical calculus of ideas immanent in nervous activity. Bulletin Mathematical Biophys. *5*, 115–133.
21. Hebb, D. (1949). The Organization of Behavior (New York: Wiley).
22. Rosenblatt, F. (1958). The Perceptron: A probabilistic model for information storage and organization in the brain. Psychol. Rev. *65*, 386–408.
23. Rosenblatt, F. (1961). Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms (Washington, DC: Spartan Books).
24. Rosenblatt, F. (1962). A comparison of several perceptron models. In Self-Organizing Systems, Yovits, Jacobi, and Goldstein, eds. (Spartan Books).
25. Rumelhart, D.E., Hinton, G.E., and Williams, R.J. (1986). Learning internal representations by error propagation. In Parallel Distributed Processing: Explorations in the Microstructure of Cognition, Volume I: Foundations, D.E. Rumelhart and J.L. McClelland, eds. (Cambridge, MA: MIT Press).
26. Bryson, A.E., Denham, W.F., and Dreyfus, S.E. (1963). Optimal programming problems with inequality constraints. AIAA J. *1*, 2544–2550.
27. Amari, S. (1967). A theory of adaptive pattern classifiers. IEEE Trans. Electronic Computers *3*, 299–307.
28. Werbos, P. (1974). Beyond regression: New tools for prediction and analysis in the behavioral sciences. Ph.D. thesis, Harvard University.
29. Fukushima, K. (1980). Neocognitron: A self organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. Biol. Cybernet. *36*, 93–202.
30. Fukushima, K., Miyake, S., and Ito, T. (1988). Neocognitron: A neural network model for a mechanism of visual pattern recognition. In Artificial Neural Networks: Theoretical Concepts (IEEE Computer Society Press), pp. 136–144.
31. Lowe, D.G. (1999). Object recognition from local scale-invariant features. In Computer vision, 1999. The Proceedings of the Seventh IEEE International Conference on, volume 2 (IEEE), pp. 1150–1157.
32. Cortes, C., and Vapnik, V. (1995). Support-vector networks. Machine Learning *20*, 273–297.
33. Bengio, Y. (2009). Learning deep architectures for ai. Foundations Trends Machine Learning *2*, 1–127.
34. Hinton, G.E. (2002). Training products of experts by minimizing contrastive divergence. Neural Comput. *14*, 1771–1800.
35. Hinton, G.E., and Salakhutdinov, R.R. (2006). Reducing the dimensionality of data with neural networks. Science *313*, 504–507.
36. Riesenhuber, M., and Poggio, T. (1999). Hierarchical models of object recognition in cortex. Nat. Neurosci. *2*, 1019–1025.
37. Hyvärinen, A. (2010). Statistical models of natural images and cortical visual representation. Top. Cogn. Sci. *2*, 251–264.
38. Krizhevsky, A., Sutskever, I., and Hinton, G. (2012). Imagenet classification with deep convolutional neural networks. In Advances in Neural Information Processing Systems 25, P. Bartlett, F. Pereira, C. Burges, L. Bottou, and K. Weinberger, eds., pp. 1106–1114.
39. Carandini, M., and Heeger, D.J. (2012). Normalization as a canonical neural computation. Nat. Rev. Neurosci. *13*, 51–62.
40. Jarrett, K., Kavukcuoglu, K., Ranzato, M., and LeCun, Y. (2009). What is the best multi-stage architecture for object recognition? In Proceedings of the International Conference on Computer Vision (IEEE Computer Society).
41. Le, Q., Ranzato, M., Monga, R., Devin, M., Chen, K., Corrado, G., Dean, J., and Ng, A. (2012). Building high-level features using large scale unsupervised learning. In Proceedings of the 29th International Conference on Machine Learning, J. Langford and J. Pineau, eds., pp. 81–88.
42. Ashby, F.G., and Maddox, W.T. (2005). Human category learning. Annu. Rev. Psychol. *56*, 149–178.
43. Lake, B.M., Salakhutdinov, R., Gross, J., and Tenenbaum, J.B. (2011). One shot learning of simple visual concepts. In Proceedings of the 33rd Annual Conference of the Cognitive Science Society, pp. 2568–2573.
44. Torralba, A., and Efros, A.A. (2011). Unbiased look at dataset bias. In 2011 IEEE Conference on Computer Vision and Pattern Recognition (IEEE), pp. 1521–1528.
45. LeCun, Y. and Cortes, C. The mnist database of handwritten digits, 1998. Available electronically at http://yann.lecun.com/exdb/mnist.
46. Ciresan, D., Meier, U., and Schmidhuber, J. (2012). Multi-column deep neural networks for image classification. In 2012 IEEE Conference on Computer Vision and Pattern Recognition (IEEE), pp. 3642–3649.
47. Scheirer, W.J., Jain, L.P., and Boult, T.E. (2014). Probability models for open set recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI), To appear, preprint at http://www.wjscheirer.com/papers/wjs_pami2014_probability.pdf.
48. Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. (2009). Imagenet: A large-scale hierarchical image database. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (IEEE), pp. 248–255.
49. Imagenet large scale visual recognition challenge 2013. http://www.image-net.org/challenges/LSVRC/2013/results.php. Accessed: 2014/07/14.
50. Dean, T., Ruzon, M.A., Segal, M., Shlens, J., Vijayanarasimhan, S., and Yagnik, J. (2013). Fast, accurate detection of 100,000 object classes on a single machine. In 2013 IEEE Conference on Computer Vision and Pattern Recognition (Los Alamitos, CA: USA. IEEE Computer Society), pp. 1814–1821.
51. Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., Erhan, D., Goodfellow, I., and Fergus, R. (2013). Intriguing properties of neural networks. arXiv, preprint arXiv:1312.6199.
52. Freeman, J., and Simoncelli, E.P. (2011). Metamers of the ventral stream. Nat. Neurosci. *14*, 1195–1201.
53. Ohki, K., Chung, S., Ch'ng, Y.H., Kara, P., and Reid, R.C. (2005). Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex. Nature *433*, 597–603.
54. Barretto, R.P., Messerschmidt, B., and Schnitzer, M.J. (2009). In vivo fluorescence imaging with high-resolution microlenses. Nat. Methods *6*, 511–512.
55. Margolis, D.J., Lütcke, H., Schulz, K., Haiss, F., Weber, B., Kügler, S., Hasan, M.T., and Helmchen, F. (2012). Reorganization of cortical population activity imaged throughout long-term sensory deprivation. Nat. Neurosci. *15*, 1539–1546.
56. Deisseroth, K. (2011). Optogenetics. Nat. Methods *8*, 26–29.

57. Bernstein, J.G., and Boyden, E.S. (2011). Optogenetic tools for analyzing the neural circuits of behavior. Trends Cogn. Sci. *15*, 592–600.
58. Goodfellow, I.J., Erhan, D., Carrier, P.L., Courville, A., Mirza, M., Hamner, B., Cukierski, W., Tang, Y., Thaler, D., Lee, D.-H., *et al*. (2013). Challenges in representation learning: A report on three machine learning contests. CoRR, abs/1307.1414.
59. Briggs, F., and Usrey, W.M. (2008). Emerging views of corticothalamic function. Curr. Opin. Neurobiol. *18*, 403–407.
60. Yu, A.J., and Dayan, P. (2005). Uncertainty, neuromodulation, and attention. Neuron *46*, 681–692.
61. Gilbert, C.D. (1983). Microcircuitry of the visual cortex. Annu. Rev. Neurosci. *6*, 217–247.
62. Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. In Parallel Distributed Processing: Explorations in the Microstructure of Cognition, *vol 1*, (Cambridge, MA: MIT Press).
63. Salakhutdinov, R., and Hinton, G.E. (2009). Deep boltzmann machines. In International Conference on Artificial Intelligence and Statistics, pp. 448–455.
64. Williams, R.J., and Zipser, D. (1989). A learning algorithm for continually running fully recurrent neural networks. Neural Comput. *1*, 270–280.
65. Pineda, F.J. (1987). Generalization of back-propagation to recurrent neural networks. Phys. Rev. Lett. *59*, 2229.
66. Socher, R., Lin, C.C.-Y., Ng, A.Y., and Manning, C.D. (2011). Parsing natural scenes and natural language with recursive neural networks. In Proceedings of the 28th International Conference on Machine Learning, L. Getoor and T. Scheffer, eds., pp. 129–136.
67. Socher, R., Perelygin, A., Wu, J.Y., Chuang, J., Manning, C.D., Ng, A.Y., and Potts, C. (2013). Recursive deep models for semantic compositionality over a sentiment treebank. In Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (Stroudsburg, PA, USA: Association for Computational Linguistics), pp. 1631–1642.
68. Tsotsos, J., Liu, Y., Martinez-Trujillo, J., Pomplun, M., Simine, E., and Zhou, K. (2005). Attending to visual motion. Computer Vision Image Understanding *100*, 3–40.
69. Wickersham, I.R., Finke, S., Conzelmann, K.-K., and Callaway, E.M. (2006). Retrograde neuronal tracing with a deletion-mutant rabies virus. Nat. Methods *4*, 47–49.
70. Dan, Y., and Poo, M.-m. (2004). Spike timing-dependent plasticity of neural circuits. Neuron *44*, 23–30.
71. Shadlen, M.N., and Movshon, J.A. (1999). Synchrony unbound: a critical evaluation of the temporal binding hypothesis. Neuron *24*, 67–77.
72. Vogels, T.P., Rajan, K., and Abbott, L. (2005). Neural network dynamics. Annu. Rev. Neurosci. *28*, 357–376.
73. Le, Q.V., Zou, W.Y., Yeung, S.Y., and Ng, A.Y. (2011). Learning hierarchical invariant spatio-temporal features for action recognition with independent subspace analysis. In 2011 IEEE Conference on Computer Vision and Pattern Recognition (IEEE), pp. 3361–3368.
74. Wang, H., Klaser, A., Schmid, C., and Liu, C.-L. (2011). Action recognition by dense trajectories. In 2011 IEEE Conference on Computer Vision and Pattern Recognition (IEEE), pp. 3169–3176.
75. Jain, M., Jégou, H., and Bouthemy, P. (2013). Better exploiting motion for better action recognition. In 2013 IEEE Conference on Computer Vision and Pattern Recognition (IEEE), pp. 2555–2562.
76. Adelson, E.H., and Movshon, J.A. (1982). Phenomenal coherence of moving visual patterns. Nature *300*, 523–525.
77. Cadieu, C.F., and Olshausen, B.A. (2012). Learning intermediate-level representations of form and motion from natural movies. Neural Comput. *24*, 827–866.
78. Rust, N.C., Mante, V., Simoncelli, E.P., and Movshon, J.A. (2006). How MT cells analyze the motion of visual patterns. Nat. Neurosci. *9*, 1421–1431.
79. Berkes, P., and Wiskott, L. (2005). Slow feature analysis yields a rich repertoire of complex cell properties. J. Vis. *5*, 579–602.
80. Wiskott, L. (2003). How does our visual system achieve shift and size invariance? In Problems in Systems Neuroscience, J.L. van Hemmen and T.J. Sejnowski, eds. (Oxford: Oxford University Press).
81. Wiskott, L., and Sejnowski, T. (2002). Slow feature analysis: unsupervised learning of invariances. Neural Comput. *14*, 715–770.
82. van Hateren, J.H., and Ruderman, D.L. (1998). Independent component analysis of natural image sequences yields spatiotemporal filters similar to simple cells in primary visual cortex. Proc. Biol. Sci. *265*, 2315–2320.
83. van Hateren, J.H., and van der Schaaf, A. (1998). Independent component filters of natural images compared with simple cells in primary visual cortex. Proc. Biol. Sci. *265*, 359–366.
84. Laptev, I., and Lindeberg, T. (2004). Velocity adaptation of spatio-temporal receptive fields for direct recognition of activities: an experimental study. Image Vis. Comput. *22*, 105–116.
85. Hildreth, E.C. (1984). Measurement of Visual Motion (Cambridge, MA: MIT Press).
86. Dean, T., Corrado, G., and Washington, R. (2009). Recursive sparse, spatiotemporal coding. In Proceedings of the Fifth IEEE International Workshop on Multimedia Information Processing and Retrieval.
87. Ringach, D.L. (2004). Mapping receptive fields in primary visual cortex. J. Physiol. *558*, 717–728.
88. Albright, T.D. (1984). Direction and orientation selectivity of neurons in visual area mt of the macaque. J. Neurophysiol. *52*, 1106–1130.
89. De Baene, W., and Vogels, R. (2010). Effects of adaptation on the stimulus selectivity of macaque inferior temporal spiking activity and local field potentials. Cereb. Cortex *20*, 2145–2165.
90. Meyer, T., and Olson, C.R. (2011). Statistical learning of visual transitions in monkey inferotemporal cortex. Proc. Natl. Acad. Sci. USA *108*, 19401–19406.
91. Schrauwen, B., Verstraeten, D., and Van Campenhout, J. (2007). An overview of reservoir computing: theory, applications and implementations. In Proceedings of the 15th European Symposium on Artificial Neural Networks, pp. 471–482.
92. Hochreiter, S., and Schmidhuber, J. (1997). Long short-term memory. Neural Comput. *9*, 1735–1780.
93. Gers, F.A., Schmidhuber, J., and Cummins, F. (2000). Learning to forget: Continual prediction with lstm. Neural Comput. *12*, 2451–2471.
94. Gers, F.A., Schraudolph, N.N., and Schmidhuber, J. (2003). Learning precise timing with lstm recurrent networks. J. Machine Learning Res. *3*, 115–143.
95. Krueger, K.A., and Dayan, P. (2007). Flexible shaping: how learning in small steps helps. Cognition *110*, 380–394.
96. O'Reilly, R.C., and Frank, M.J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. Neural Comput. *18*, 283–328.
97. Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. IEEE Trans. Pattern Analysis Machine Intelligence *20*, 1254–1259.
98. Vig, E., Dorr, M., and Barth, E. (2009). Efficient visual coding and the predictability of eye movements on natural movies. Spat. Vis. *22*, 397–408.
99. Judd, T., Ehinger, K., Durand, F., and Torralba, A. (2009). Learning to predict where humans look. In 2009 IEEE Conference on Computer Vision and Pattern Recognition (IEEE), pp. 2106–2113.
100. Vig, E., Dorr, M., and Cox, D. (2014). Large-scale optimization of hierarchical features for saliency prediction in natural images. In 2014 IEEE Conference on Computer Vision and Pattern Recognition (IEEE).
101. Insel, T.R., Landis, S.C., and Collins, F.S. (2013). The NIH brain initiative. Science *340*, 687–688.
102. Shi, X., Bruce, N.D., and Tsotsos, J.K. (2012). Biologically motivated local contextual modulation improves low-level visual feature representations. In Image Analysis and Recognition, volume 7324 of Lecture Notes in Computer Science, A. Campilho and M. Kamel, eds. (Berlin, Heidelberg: Springer), pp. 79–88.
103. Eliasmith, C., Stewart, T.C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., and Rasmussen, D. (2012). A large-scale model of the functioning brain. Science *338*, 1202–1205.
104. Marcus, G. (2012). Is "Deep Learning" a Revolution in Artificial Intelligence? The New Yorker *25*.
105. LeCun, Y., Jackel, L., Bottou, L., Brunot, A., Cortes, C., Denker, J., Drucker, H., Guyon, I., Muller, U., Sackinger, E., *et al*. (1995). Comparison of learning algorithms for handwritten digit recognition. In International Conference on Artificial Neural Networks, *volume 60* .
106. Deep learning tutorial. http://deeplearning.net/tutorial/lenet.html. Accessed: 2014/07/14.