Chapter 10

Artificial Intelligence in Neuroscience

Will Xiao, Mengmi Zhang, and Gabriel Kreiman

Neurosurgeons have the privilege of peeking inside the most precious and the most mysterious device on earth: the human brain (Crick et al., 2004). The human brain is also the most expensive device on earth given that mental health problems constitute the largest health care cost. By deciphering the inner secrets of brain computations, scientists and engineers have taken inspiration to develop smart artificial intelligence (AI) algorithms. These AI algorithms in turn provide much help to understanding brain function and to multiple applications in brain disorders, including neurosurgery.

10.1 From Neural Circuits to Artificial Intelligence

Humanity has long imagined automated machines that can do work for us. The ultimate frontier is to construct machines that can emulate the human brain, or perhaps even surpass human intelligence. The development of AI commenced in the 1950s and has gained momentum in the last decade. Throughout its short history, AI research has been inspired by notions of neuroscience. Visual processing constitutes a paradigmatic example of AI and the links between AI and neuroscience.

Consider a hypothetical algorithm that is capable of taking as input a magnetic resonance image of a subject's brain and detecting the presence of a tumor. At the heart of AI algorithms are neural networks, that is, interconnected neuron-like units that receive inputs and progressively transform those inputs into forms that are more useful to solve the task. The output of the algorithm indicates the probability that the image contains a tumor and its location. A unit in a neural network is a highly simplified model of a neuron: it receives inputs from other units, weighs and sums those inputs, applies a non-linear transformation, and produces a scalar output (McCulloch and Pitts, 1943). Neuronal firing rates are represented by scalar activation values, the strengths of synapses between neurons are replaced by weights, and the biophysics of action potential generation is captured by a single non-linearity that converts the weighted inputs into an output.

The power of neural networks derives from connecting these simple units into large ensembles that have interesting emergent properties. Here again, AI takes inspiration from the brain. The visual cortex uses a divide-and-conquer strategy by layering neurons in an approximately hierarchical fashion from the retina, to the lateral geniculate nucleus, onto primary visual cortex (V1), onto visual cortical area V2, and so on (Felleman and Van Essen, 1991; Figure 10.1A). A particularly successful computational architecture, known as deep neural networks, follows the same principle by having multiple layers that sequentially transfsorm information (Figure 10.1B).

In the visual cortex, the same image features are extracted at different locations throughout the visual field. For example, each V1 neuron selectively responds to bars of a specific orientation at a specific part of the visual field (Hubel and Wiesel, 1962). The population of V1 neurons tile the entire visual field, with neurons specific to each orientation at, say, the center of gaze and also in the visual periphery. To create an artificial layer with similar properties, computational models recur to a convolution operation, such that units with the same weights are effectively repeated in different locations. Meanwhile, units with different weights account for different orientation tuning. Similarly, features other than orientation are extracted in different units and layers using the convolution operation. The resulting networks are known as deep convolutional neural networks (CNNs). Many other neurobiologically inspired operations are included, such as non-linear pooling of inputs (Riesenhuber and Poggio, 1999) and normalization (Carandini and Heeger, 2011).

Humans learn to recognize cars, chairs, or brain tumors in images. It is generally thought that such learning largely amounts to modifying synaptic strengths in visual cortex. Equivalently, neural networks are trained by modifying the weights between units. Powerful learning algorithms have been developed to modify synaptic weights,

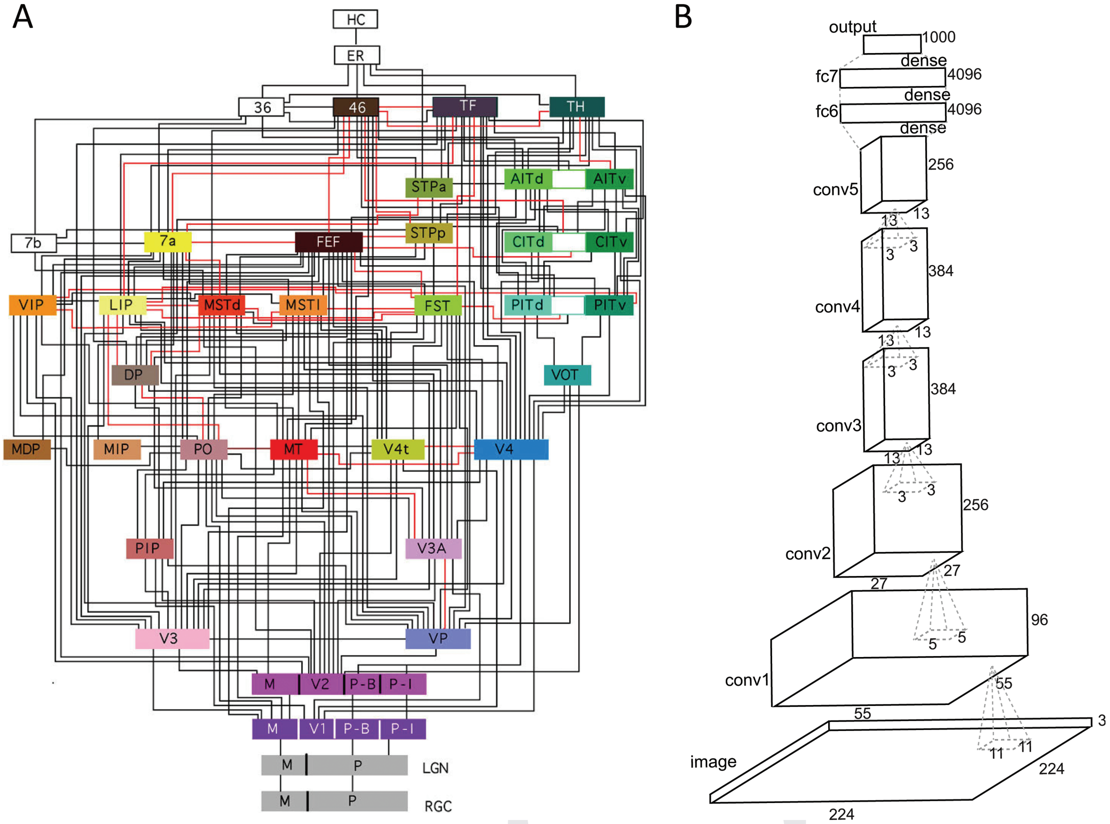

Figure 10.1 Biological and artificial visual networks. (A) Mesoscopic anatomical connections between different visual areas in the macaque cortex (Felleman and Van Essen, 1991). Each colored box shows a brain area and lines represent anatomical connections. The bottom shows the retinal ganglion cells (RGCs). Visual areas are arranged in an approximately hierarchical fashion. This architecture has inspired the idea that information flows in a semi-sequential fashion from the bottom to the top of this diagram. See Felleman and Van Essen (1991) for the definition of each anatomical abbreviation. (B) Example deep convolutional neural network (CNN) architecture known as AlexNet (Krizhevsky et al., 2012). Each solid box represents a layer in a neural network (loosely equivalent to a brain area). The dashed box denotes a unit (neuron) in a layer. The converging dash lines bring information to the next layer (representation of anatomical connections across layers). The input to the model is an image (bottom), the output is a series of numbers, 1000 in this case, which represent the probabilities that the image contains one of 1000 possible object categories. The numbers indicate the dimensions of each layer, "conv" stands for convolutional layer, and "fc" indicates a fully connected layer. See Krizhevsky et al. (2012) for further details. Author's artwork

including back-propagation. Back-propagation is often used in a supervised learning setting: the network produces tentative (initially random) labels for the training examples, and if the output is wrong, the error signal is propagated back throughout the network to modify the weights in the direction that makes the output closer to the correct answer. Trained with many correctly labeled examples, the network can gradually learn to separate, for example, images with or without tumors.

10.2 Artificial Intelligence Contributions to Visual Neuroscience

Just as neuroscientific knowledge inspired advances in AI, so does AI feed back to neuroscience, providing new tools, models, and ways of understanding the brain (see Hassabis et al., 2017; Richards et al., 2019; Serre, 2019; Zador, 2019 for reviews on the synergisms between neuroscience and AI). We highlight three active research directions in vision that draw on AI as tools and models.

10.2.1 Goal-Driven Convolutional Neural Networks Explain Neuronal Responses Underlying Object Recognition

Object recognition is a well-defined visual process underlying our ability to make sense of the visual world. Clinical case studies exemplify the profound consequences of failures in visual recognition (Sacks, 2002). Neurons along the ventral visual cortex are especially important for visual recognition, as demonstrated by studies in non-human primates (DiCarlo et al., 2012) and using invasive neurophysiological recordings from human neurosurgical patients (Liu et al., 2009). Understanding how these neurons process visual information is of both medical and practical interest. Responses of early visual cortex have been explained by hand-designed models, such as Gabor wavelet filters (Hubel and Wiesel, 1959), or by models based on first-principles such as efficient coding (Olshausen and Field, 1996). However, similar approaches have achieved limited success in describing the responses of neurons in higher-order visual areas, likely because of the complexity of high-level visual features.

In parallel, for a better characterization of visual cortex, CNNs loosely based on neurobiological knowledge were being developed (see previous section). CNNs are image-computable models, meaning that we can show the same images to the models as we show to a monkey or a human. Thus, we can directly compare the responses of artificial units and biological neurons to the same inputs. As described in the previous section, the synaptic weights of the CNNs are adjusted via supervised learning on a computer vision data set created to assess object recognition, notably without using any neural data. One approach to comparing AI units to biological neurons is to use a linear mapping to fit the responses of a biological neuron's response patterns (Yamins et al., 2014). Validation of these models was only possible relatively recently, as AI systems, fueled by several technological advances, became powerful enough to do realistic visual recognition tasks. Remarkably, CNNs as models to quantitatively explain neuronal responses have outperformed all previous models in describing the responses of neurons in the retina, in primary visual cortex, and in other areas along the ventral visual stream, including the top echelons of inferior temporal cortex. For example, the famous AlexNet neural network (Krizhevsky et al., 2012), at its time the best-performing algorithm by far on a challenging image recognition task, has been shown to be a good model of visual neuronal responses, as have later CNNs with increasingly better performance (Schrimpf et al., 2018). Furthermore, CNN models with higher performance on computer vision object categorization tasks tend to better explain neuronal responses in ventral visual cortex. This result suggested that brain-like computations may be specified by brain-related tasks, rather than by neural data directly. Ongoing research highlights the remaining gap in the quantitative description of neuronal responses (Schrimpf et al., 2018), and aims to address deviations of CNNs from biological vision in architecture (Kar et al., 2019; Kubilius et al., 2019; Tang et al., 2018), learning rules (Lillicrap et al., 2020), and behavior (Nguyen et al., 2015; Szegedy et al., 2020).

10.2.2 Convolutional Neural Network-Learned Features Reveal Stimulus Preferences and Organizing Principles in Visual Cortex

As discussed above, visual features learned by neural networks agree well with those used in the brain. In addition, learned features – unlike hand-picked features such as the orientation of a grating, surface texture, shape curvature, or face appearance – apply generally throughout visual areas and types of selectivity. Thus, such visual features can serve as a general framework for defining stimulus preferences and explaining the organizing principles throughout the visual cortex.

A recent study (Ponce et al., 2019) showed that the tuning properties of V1 and inferior temporal (IT) cortex neurons can be discovered using a generative neural network (Goodfellow et al., 2014), without resorting to arbitrary investigator-chosen features like oriented gratings or faces. In a typical CNN, the input is an image and the output is a vector of image features or labels. A generative neural network inverts the process, taking features as inputs and producing images. Combining the generative network with neuronal recordings in a closed loop (Figure 10.2), this method allowed the investigators to discover image features that activated the neurons as well as, and in some cases better than, the best natural images. Notably, although the generative network was trained on natural images, it was not limited to images it has seen. Thus, this approach could discover neuronal tuning in an unbiased fashion, without imposing an investigator's existing notions of what a neuron might be tuned to. Similar approaches have been used to reveal visual neuronal tuning in other areas in monkeys (Abbasi-Asl et al., 2018; Bashivan et al., 2019), mice (Walker et al., 2019), and artificial neural networks (Olah et al., 2017; Zeiler and Fergus, 2014).

An interesting property of the cortex is its topographical organization whereby nearby neurons show similar tuning properties. A recent study (Bao et al., 2020) used CNN-encoded image features to confirm this topographical organization of macaque IT cortex and describe its feature selectivity. The IT cortex contains continuous patches selective for directions of image variation defined by a CNN. Patches selective for different feature directions are arranged in a consistent order, repeated from

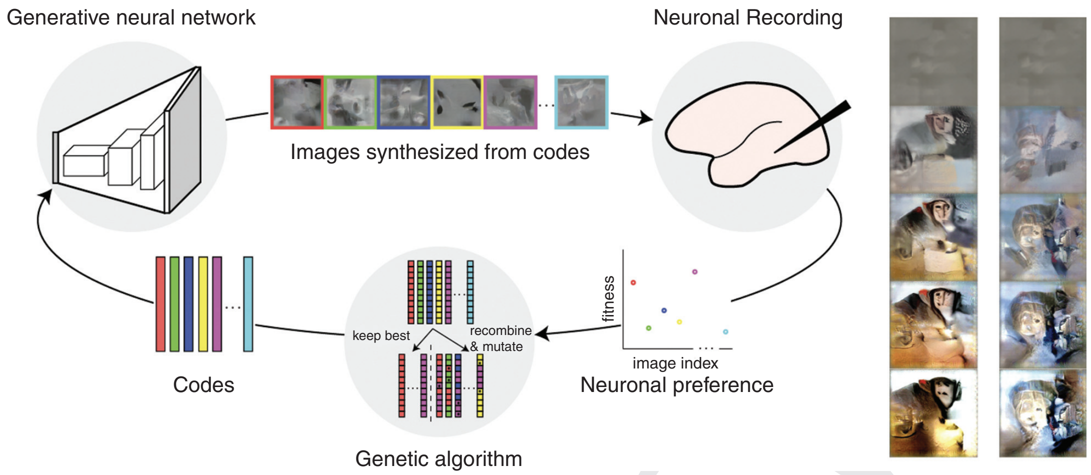

Figure 10.2 Example of how AI has helped advance vision research. A generative neural network coupled to a genetic algorithm can be used in a closed-loop system to uncover neuron selectivity in an automatic and unbiased manner. A generative neural network is an inverted CNN that takes feature codes as inputs and creates images. The synthetic images are presented to a monkey while recording neuronal responses. These neuronal responses are used as a "fitness" function to rank the images from best to worst. Finally, a genetic algorithm mutates and recombines the feature codes to feed back to the image generator. The loop is iterated until the algorithm converges on images that effectively trigger high activation for the neurons (Ponce et al., 2019). The two columns on the right show two example evolutions of preferred images for two face-selective neurons. Author's artwork

posterior to anterior IT cortex. Hearkening to layers in deep neural networks, this organization of the IT cortex is also hierarchical: responses in more anterior patches are more invariant to the same object viewed from different angles. Moreover, previous accounts of IT tuning to faces, size (big–small), and 3D shape can be explained in a common framework based on the CNN-derived features. In fact, although these features partially correlated with semantic categories, CNN-derived features explain neuronal responses better than do semantic categories.

In summary, artificial neural networks can be used to meaningfully interpret the computations along visual cortex and parsimoniously explain neuronal response properties. Such an understanding may one day be used to develop more specific and powerful visual prosthetics (see the last section).

10.2.3 Artificial Intelligence-Based Models Illuminate Complex Visual Behaviors

Artificial intelligence algorithms can now achieve human-level performance on some specialized visual tasks, such as the ImageNet challenge for object recognition (He et al., 2015; Russakovsky et al., 2015) and clinical tasks like detecting whether an image contains a tumor or not (Lotter et al., 2021). Yet, the general visual behaviors of humans are much richer, more complex, and more versatile, and there is still a long way to go for AI systems to emulate higher-order human vision. Nevertheless, domain-specific AI systems provide useful building blocks for modeling such behaviors. For example, humans can perform efficient visual search. One example is searching for a tumor in a magnetic resonance image. This search is efficient, because humans do not have to exhaustively sample all possible locations; and invariant, because humans can search for objects regardless of their position, scale, illumination, or rotation. In addition, humans can perform a zero-shot search, namely searching for novel objects. In comparison, object detection algorithms in computer vision require extensive training for each object class and extensive sampling of image locations during detection. To start closing the gap, investigators developed a biologically inspired computational model for visual search (Zhang et al., 2018) composed of an object recognition module and an attention module. The model produces a series of eye movements that are specific to the sought target and the search image. Without training on human data, this algorithm approximates human eye movements and search performance. Like humans, this model does not need to be retrained for new search targets or conditions, working out-of-the-box to search for objects in an array on a uniform background, in natural photographs, and in Where's Waldo images. Presumably, this is possible because the underlying object recognition model has learned generally useful visual features and invariances. Similar methods of using simple models as building blocks hold promise for explaining a large range of complex visual behaviors.

The reach of AI algorithms now extends well beyond pure sensory processes. Artificial intelligence-based systems have been used to model a large range of behaviors including reinforcement learning (Dabney et al., 2020), decision making, and game-play. In some cases, AI algorithms have achieved superhuman performance in problems not long ago thought to be still out of reach for machines – including Atari games (Mnih et al., 2015), chess, the ancient Chinese game of Go (Silver et al., 2016, 2017, 2018), video games like StarCraft II (Vinyals et al., 2019), and Texas hold 'em (Brown and Sandholm, 2019).

10.3 Prospects of Artificial Intelligence-Based Tools in Neuroscience and Neurosurgery

The ability of AI systems to perform accurate pattern recognition holds the promise to radically transform many aspects of clinical practice. One of the important domains that has been revolutionized by advances in AI is image-based diagnosis, but AI is also likely to impact many other domains such as the development of better brain–machine interfaces.

10.3.1 Discovering the Unexpected: Disease Diagnosis from Retinal Fundus Photographs with Artificial Intelligence Systems

An example application of CNNs in clinical diagnosis is the analysis of retinal fundus photographs (see also work in cancer detection; Lotter et al., 2021). Such images can be used to diagnose conditions such as diabetic retinopathy (Gulsham et al., 2016; Ting et al., 2017). In the same fashion that a CNN can be trained to discriminate images of chairs versus tables via supervised learning, these neural networks can be fed with multiple examples of fundus photographs with or without diabetic retinopathy. These expert-annotated images are used to train the network (that is, change the synaptic weights between units). The CNNs excel at this task, and their performance is similar to that of expert ophthalmologists (Poplin et al., 2018). Subsequent work extended CNNs to diagnosing glaucoma and age-related macular degeneration (AMD) using retinal images (Ting et al., 2017).

Surprisingly, CNNs can be trained to accurately predict other information from these images, including a subject's gender or age. Clinicians never knew that there was such information in fundus photographs; after all, gender and age are not particularly interesting variables to decipher from clinical images given that such information is always available to the doctor.

What is even more astounding, CNNs could use fundus photographs to predict smoking status (71% of the time) and systolic blood pressure (11.23 mmHg in mean absolute error). Next, the investigators then asked whether it is possible to predict the risk of cardiovascular disease from the same images. Strikingly, CNNs could estimate the risk of cardiovascular disease as well as the Framingham score, without any information other than the images themselves.

10.3.2 Detecting Spatiotemporal Patterns in Data: Seizure Prediction Using Deep Learning

The success of CNNs in pattern classification from images has encouraged the development of a plethora of other types of deep-learning algorithms that extend beyond image processing, for example, incorporating analysis of temporal information in the classification of video data and other time-varying signals. A noteworthy domain of applications is seizure detection and seizure prediction (Fergus et al., 2016). A recent review paper by Siddiqui et al. (2020) shows that a wide range of machine-learning methods are capable of classifying seizures given a segment of time-series data.

The main challenges in seizure detection and seizure prediction involve the selection of appropriate classifiers and features. The type of CNN architectures discussed in Sections 10.1 and 10.2 do not readily incorporate temporal information. Rather, signals go from one layer to the next in a sequential fashion. In stark contrast, information flow in the brain includes complex dynamics arising from horizontal connections between neurons in a given brain area as well as the interplay of bottom-up and top-down signals (Felleman and Van Essen, 1991). Akin to such biological connectivity, recurrent neural networks (RNNs), which belong to a family of deep-learning models, have shown success in describing dynamic aspects of neural firing in the visual system (Kar et al., 2019; Tang et al., 2018) and in natural-language processing (Vaswani et al., 2017).

Similarly, several studies have started exploring the feasibility and efficiency of applying RNNs in extracting temporal signatures of seizures from time-series data. Cho and Jang (2020) surveyed and compared current deep-learning techniques in seizure classification. These techniques typically take inputs from either of three major classes: 5-s segments of raw time-series data from electroencephalographic (EEG) data, 2D images of raw EEG waveforms, and 2D images in frequency domain using short-term Fourier transform of the waveforms. Seizure classification using AI algorithms achieves high accuracy (99.3% of the time). These results encourage the exploration of the more challenging problem of seizure prediction.

10.3.3 Restoring Vision for the Blind: Artificial Intelligence-Assisted Brain–Machine Interactions

In the famous "frog galvanoscope" experiment of 1781, Luigi Galvani injected electric currents to twitch a frog's legs. Since then, humans have never stopped pursuing the dream of communicating between nervous systems and machines. Until now, courageous and initial efforts have been taken in developing medical implants to treat Parkinson's disease via deep brain stimulation (Cagnan et al., 2019) and neuronal signal decoders to decipher movement intentions in quadriplegic patients (Tam et al., 2019). With the advent of powerful AI tools, major revolutions could come for brain–machine interfaces (BMIs) for a wide variety of clinical applications.

A tantalizing prospect is the development of prosthetics for visual restoration, which could be life-changing for the large number of patients who are blind or have severe visual impairment. Economical yet sophisticated digital cameras that can rapidly capture and transmit visual information are now prevalent in smartphones and wearable devices. Convolutional neural networks can extract meaningful information from such images. The main challenge and missing link is how to transmit this information to a patient's brain. A review paper by Niketeghad and Pouratian (2019) outlines opportunities and challenges of the cortical visual prosthesis under investigation. Two main approaches are taken, based on electrical stimulation of either the retina or V1. Both approaches suffer from the limited number of electrodes and non-uniform arrangement of their ensuing phosphenes in the visual field.

An intriguing alternative would be to use the type of computer vision approaches highlighted in the previous section to preprocess the images and submit to the cortex a digested version of neural codes compatible with brains (Figure 10.3). This highly processed information could then be fed into higher visual centers via electrical stimulation. Instead of eliciting individual phosphenes, this approach may be used to evoke a complete visual precept, by triggering neuronal populations tuned to more complex visual features or even aspects of visual scene understanding. For example, depending on the subject's goals, the AI algorithm could either read text, help avoid obstacles during navigation, identify faces, or even provide a description of the visual scene. A schematic illustration of a hypothetical high-level BMI for the blind is presented in Figure 10.3. One could imagine that AI tools, combined with an understanding of the visual cortex, can enable the "inception" of specific precepts by a neuro-visual prosthesis (Roe et al., 2020).

An even more intriguing alternative is to feed this highly processed information into readily available information channels, such as auditory and somatosensory modalities. For example, the AI algorithms could interpret the scene surrounding a blind subject's environment and then provide explicit directions and instructions via a virtually generated avatar that provides directional verbal commands. Rather than a BMI device, this approach merely requires portable and virtual reality (VR)-compatible headphones that are readily available through the gaming community (Liu et al., 2018).

10.3.4 Artificial Intelligence-Specific Concerns in Neuroscience and Neurosurgery

Artificial intelligence technologies come with risks; they face criticisms including privacy threats, security pitfalls, potential biases, issues of legal liability, as well as other ethical and moral concerns (Grote and Berens, 2020; Rigby, 2019; Vayena et al., 2018). For example, many AI models need access to massive data sets for training. Transferring gigabytes of data across multiple healthcare organizations might lead to data breaches. Moreover, biases intrinsic to the data sets (such as those pertaining to race, gender, or other stereotypes) may become ingrained into the algorithm during training. Even if all privacy and bias concerns were adequately addressed, algorithms can fail to provide the right answer. Of course, clinicians do not always make the correct diagnosis, either. But, if an algorithm fails, who is to blame?

With great power comes great responsibility. As humanity continues to develop and benefit from intelligent machines, researchers and practitioners alike should heed potential ethical concerns, address them proactively, and shape the future of AI.

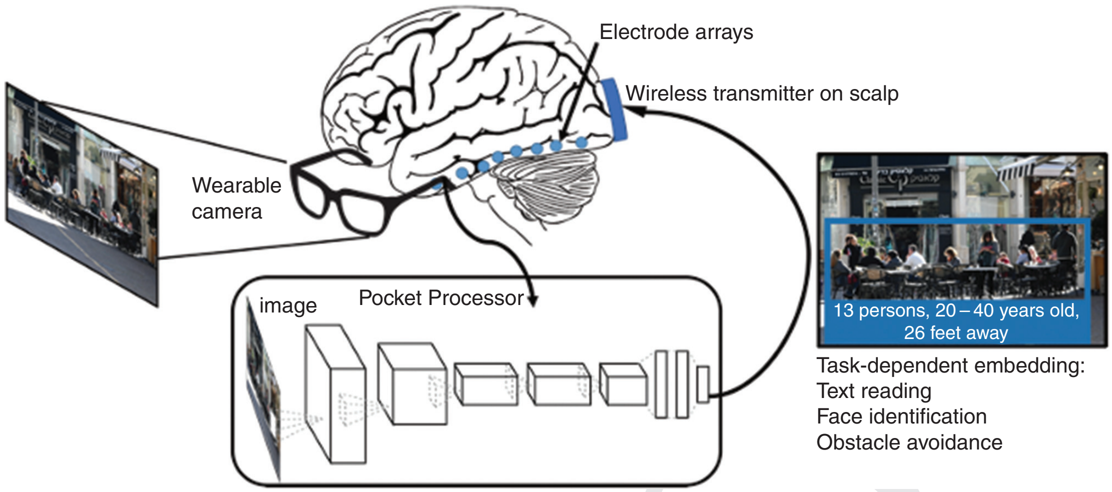

Figure 10.3 Schematic illustration of a hypothetical future high-level brain–machine interface for the blind. A wearable camera sends the visual input to the CNN algorithm for image processing implemented in a pocket processor. The CNN algorithm could extract in real time task-relevant information including reading text, identifying faces, and avoiding obstacles during navigation. The main challenge is how to convey the extracted information to the blind person. This figure illustrates a potential wireless transmitter that communicates with implanted electrodes arrays across multiple levels of ventral visual cortex that would electrically stimulate the brain to pass on the AI-processed information. Author's artwork

References

Abbasi-Asl R, Chen Y, Bloniarz A, et al. The DeepTune framework for modeling and characterizing neurons in visual cortex area V4. *bioRxiv* 2018:465534. https://doi.org/10.1101/465534

Bao P, She L, McGill M, and Tsao DY. A map of object space in primate inferotemporal cortex. *Nature* 2020;**583**:103–08.

Bashivan P, Kar K, and DiCarlo JJ. Neural population control via deep image synthesis. *Science* 2019;**364**(6439):eeav9436.

Brown N and Sandholm T. Superhuman AI for multiplayer poker. *Science* 2019;**365**:885–90. https://doi.org/10.1126/science.aay240.

Cagnan H, Denison T, McIntyre C, and Brown P. Emerging technologies for improved deep brain stimulation. *Nat Biotechnol* 2019;**37**:1024–33. https://doi.org/10.1038/s41587-019-0244-6.

Carandini M and Heeger D. Normalization as a canonical neural computation. *Nat Rev Neurosci* 2011;**23**:51–62. https://doi.org/10.1038/nrn3136.

Cho KO and Jang HJ. Comparison of different input modalities and network structures for deep learning-based seizure detection. *Scientific Rep* 2020;**10**:1–11. https://doi.org/10.1038/s41598-019-56958-y.

Crick F, Koch C, Kreiman G, and Fried I. Consciousness and neurosurgery. *Neurosurgery* 2004;**55**:272–82. https://doi.org/10.1227/01.neu.0000129279.26534.76.

Dabney W, Kurth-Nelson Z, Uchida N, et al. A distributional code for value in dopamine-based reinforcement learning. *Nature* 2020;**577**:671–5. https://doi.org/10.1038/s41586-019-1924-6.

DiCarlo JJ, Zoccolan D, and Rust NC. How does the brain solve visual object recognition? *Neuron* 2012;**73**:415–34. https://doi.org/10.1016/j.neuron.2012.01.010.

Felleman D and Van Essen D. Distributed hierarchical processing in the primate cerebral cortex. *Cereb Cortex* 1991;**1**:1–47. https://doi.org/10.1093/cercor/1.1.1-a.

Fergus P, Hussain A, Hignett D, Al-Jumeily D, Abdel-Aziz K, and Hamdan H. A machine learning system for automated whole-brain seizure detection. *Appl Comput Informat* 2016;**12**:70–89. https://doi.org/10.1016/j.aci.2015.01.001.

Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2014: 2672–80.

Grote T and Berens P. On the ethics of algorithmic decision-making in healthcare. *J Med Ethics* 2020;**46**:205–11. https://doi.org/10.1136/medethics-2019-105586.

Gulshan V, Peng L, Coram M, et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA* 2016;**316**:2402–10. https://doi.org/10.1001/jama.2016.17216.

Hassabis D, Kumaran D, Summerfield C, and Botvinick M. Neuroscience-inspired artificial intelligence. *Neuron* 2017;**95**:245–58. https://doi.org/0.1016/j.neuron.2017.06.011.

He K, Zhang X, Ren S, and Sun J. Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE International Conference on Computer Vision*, Santiago, Chile, 2015, pp. 1026–34. https://doi.org/10.1109/ICCV.2015.123.

Hubel DH and Wiesel TN. Receptive fields of single neurones in the cat's striate cortex. *J Physiol* 1959;**148**:574. https://doi.org/10.1113/jphysiol.1959.sp006308.

Hubel DH and Wiesel TN. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *J Physiol* 1962;**160**:106. https://doi.org/10.1113/jphysiol.1962.sp006837.

Kar K, Kubilius J, Schmidt K, Issa EB, and DiCarlo JJ. Evidence that recurrent circuits are critical to the ventral stream's execution of core object recognition behavior. *Nature Neuroscience* 2019;**22**:974–83. https://doi.org/10.1038/s41593-019-0392-5.

Krizhevsky A, Sutskever I, and Hinton GE. ImageNet classification with deep convolutional neural networks. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2012:1097–105.

Kubilius J, Schrimpf M, Kar K, et al. Brain-like object recognition with high-performing shallow recurrent ANNs. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2019: 12805–16.

Lillicrap TP, Santoro A, Marris L, Akerman CJ, and Hinton G. Backpropagation and the brain. *Nat Rev Neurosci* 2020;**21**:335–46. https://doi.org/10.1038/s41583-020-0277-3.

Liu H, Agam Y, Madsen J, and Kreiman G. Timing, timing, timing: fast decoding of object information from intracranial field potentials in human visual cortex. *Neuron* 2009;**62**:281–90. https://doi.org/10.1016/j.neuron.2009.02.025.

Liu, Yang, Stiles NRB, Meister M. Augmented reality powers a cognitive assistant for the blind.*eLife* 2018;**7**:e37841. https://doi.org/10.7554/eLife.37841.

Lotter W, Diab A, Haslam B, et al. Robust breast cancer detection in mammography and digital breast tomosynthesis using annotation-efficient deep learning approach. *Nat Med* 2021;**27**:244–9. https://doi.org/10.1038/s41591-020-01174-9.

McCulloch W and Pitts W. Logical calculus of the ideas immanent in nervous activity. *Bull Math Biophys* 1943;**5**:115–33. https://doi.org/10.1007/BF02478259.

Mnih V, Kavukcuoglu K, Silver D, et al. Human-level control through deep reinforcement learning. *Nature* 2015;**518**:529–33. https://doi.org/10.1038/nature14236.

Nguyen A, Yosinski J, and Clune J. Deep neural networks are easily fooled: high confidence predictions for unrecognizable images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Santiago, Chile, 2015, pp. 427–36. https://doi.org/10.1109/CVPR.2015.7298640.

Niketeghad S and Pouratian N. Brain machine interfaces for vision restoration: the current state of cortical visual prosthetics. *Neurotherapeutics* 2019;**16**:134–43. https://doi.org/10.1007/s13311-018-0660-1.

Olah C, Mordvintsev A, and Schubert L. Feature visualization. *Distill* 2017;**2**:e7. https://distill.pub/2017/feature-visualization/.

Olshausen BA and Field DJ. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature* 1996;**381**:607–09. https://doi.org/10.1038/381607a0.

Ponce CR, Xiao W, Schade PF, Hartmann TS, Kreiman G, and Livingstone MS. Evolving images for visual neurons using a deep generative network reveals coding principles and neuronal preferences. *Cell* 2019;**177**:999–1009. https://doi.org/10.1016/j.cell.2019.04.005.

Poplin R, Varadarajan AV, Blumer K, et al. Prediction of cardiovascular risk factors from retinal fundus photographs via deep learning. *Nat Biomed Eng* 2018;**2**:158. https://doi.org/10.1038/s41551-018-0195-0.

Richards BA, Lillicrap TP, Beaudoin P, et al. A deep learning framework for neuroscience. *Nat Neurosci* 2019;**22**:1761–70. https://doi.org/10.1038/s41593-019-0520-2.

Riesenhuber M and Poggio T. Hierarchical models of object recognition in cortex. *Nat Neurosci* 1999;**2**:1019–25. https://doi.org/10.1038/14819.

Rigby MJ. Ethical dimensions of using artificial intelligence in health care. *AMA J Ethics* 2019;**21**:121–4. https://doi.org/10.1001/amajethics.2019.121.

Roe AW, Chen G, Xu AG, and Hu J. A roadmap to a columnar visual cortical prosthetic. *Curr Opin Physiol* 2020;**16**:68–78. https://doi.org/10.1016/j.cophys.2020.06.009.

Russakovsky O, Deng J, Su H, et al. ImageNet large scale visual recognition challenge. *Int J Comput Vis* 2015;**115**:211–52. https://doi.org/10.1007/s11263-015-0816-y.

Sacks O. *The Man who Mistook His Wife for a Hat*. Picador, 2002.

Schrimpf M, Kubilius J, Hong H, et al. Brain-score: which artificial neural network for object recognition is most brain-like? *BioRxiv* 2018:407007. https://doi.org/10.1101/407007.

Serre T. Deep learning: the good, the bad, and the ugly. *Annu Rev Vision Sci* 2019;**5**:399–426. https://doi.org/10.1146/annurev-vision-091718-014951.

Siddiqui MK, Morales-Menendez R, Huang X, and Hussain N. A review of epileptic seizure detection using machine learning classifiers. *Brain Informat* 2020;**7**:1–18. https://doi.org/10.1186/s40708-020-00105-1.

Silver D, Huang A, Maddison CJ, et al. Mastering the game of Go with deep neural networks and tree search. *Nature* 2016;**529**:484–9. https://doi.org/10.1038/nature16961.

Silver D, Hubert T, Schrittwieser J, et al. A general reinforcement learning algorithm that masters chess, shogi, and Go through self-play. *Science* 2018;**362**:1140–4. https://doi.org/10.1126/science.aar6404.

Silver D, Schrittwieser J, Simonyan K, et al. Mastering the game of Go without human knowledge. *Nature* 2017;**550**:354–9. https://doi.org/10.1038/nature24270

Szegedy C, Zaremba W, Sutskever I, et al. Intriguing properties of neural networks. *arXiv* 2013; preprint arXiv:1312.6199. https://doi.org/10.48550/arXiv.1312.6199.

Tam Wk, Wu T, Zhao Q, Keefer E, and Yang Z. Human motor decoding from neural signals: a review. *BMC Biomed Eng* 2019;**1**:22. https://doi.org/10.1186/s42490-019-0022-z.

Tang H, Schrimpf M, Lotter W, et al. Recurrent computations for visual pattern completion. *PNAS*2018;**28**:8835–40. https://doi.org/10.1073/pnas.1719397115.

Ting DSW, Cheung CYL, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA* 2017;**318**:2211–23. https://doi.org/10.1001/jama.2017.18152..

Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need. In *Advances in Neural Information Processing Systems*. Curran Associates, Inc., 2017: 5998–6008. https://doi.org/10.48550/arXiv.1706.03762.

Vayena E, Blasimme A, and Cohen IG. Machine learning in medicine: addressing ethical challenges. *PLoS Med* 2018;**15**:e1002689. https://doi.org/10.1371/journal.pmed.1002689.

Vinyals O, Babuschkin I, Czarnecki WM, et al. Grandmaster level in StarCraft II using multi-agent reinforcement learning. *Nature* 2019;**575**:350–4. https://doi.org/10.1038/s41586-019-1724-z.

Walker EY, Sinz FH, Cobos E, et al. Inception loops discover what excites neurons most using deep predictive models. *Nat Neurosci* 2019;**22**:2060–5. https://doi.org/10.1038/s41593-019-0517-x.

Yamins DL, Hong H, Cadieu CF, Solomon EA, Seibert D, and DiCarlo JJ. Performance-optimized hierarchical models predict neural responses in higher visual cortex. *PNAS* 2014;**111**:8619–24. https://doi.org/10.1073/pnas.1403112111.

Zador AM. A critique of pure learning and what artificial neural networks can learn from animal brains. *Nat Commun* 2019;**10**:1–7. https://doi.org/10.1038/s41467-019-11786-6.

Zeiler MD and Fergus R. Visualizing and understanding convolutional networks. In Fleet D, Pajdla T, Schiele B, Tuytelaars T (Eds.), *Computer Vision – European Conference on Computer Vision 2014*. Lecture Notes in Computer Science, vol. 8689. Springer, 2014: pp. 818–33. https://doi.org/10.1007/978-3-319-10590-1_53.

Zhang M, Feng J, Ma KT, Lim JH, Zhao Q, and Kreiman G. Finding any Waldo with zero-shot invariant and efficient visual search. *Nat Commun* 2018;**9**:1–15. https://doi.org/10.1038/s41467-018-06217-x.