# Deep Neural Networks Bring Patterns into Focus

WITH each new version, the typical smartphone's talking digital assistant seems to get better at recognizing spoken language and responding appropriately to requests such as "Make a noon reservation the nearest Thai restaurant." Yet the same digital assistant would be flummoxed if asked to parse the content of a digital photo of, say, a person throwing a ball to a dog on a grass field in a park. The key to such a capability is the ability to recognize patterns, something that humans do quite well, but is still relatively primitive in the computing world. This capability is also greatly needed in applications such as analyzing satellite

photographs, where data collection is far outpacing the ability of human analysts to process the data.

The rapid rise of a branch of machine learning known as deep learning is about to change computing capability. Deep learning algorithms are now being used to train a new generation of artificial neural networks (ANNs) that potentially offer game-changing performance. After years of relatively little attention, the ANN field has recently exploded, and universities and major technology companies such as IBM, Google, Facebook, Baidu, and Apple are investing heavily. A Livermore deep learning research

| 10,000 neurons | 50,000 neurons | 400,000 neurons | Original |
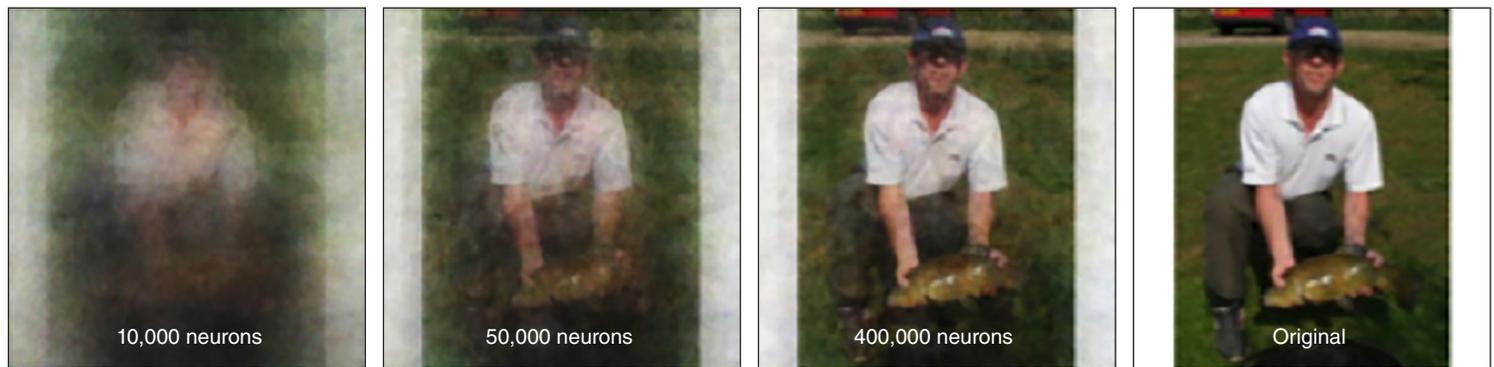
The improving image quality from left to right shows how well a single-layer autoencoder trained with the Livermore Big Artificial Neural Network (LBANN) toolkit is able to reconstruct a training image when using hidden layers consisting of 10,000, 50,000, and 400,000 neurons.

team led by machine learning researcher Barry Chen is working to advance deep learning capabilities and apply them to Livermore's national security missions and basic-science research.

The Livermore team recently developed the Livermore Brain—the world's largest neural network based on unsupervised learning with image data—along with the accompanying software for training massive neural networks. Together, the toolkit is called the Livermore Big Artificial Neural Network (LBANN). In partnership with Yahoo!, Flickr, and the International Computing Science Institute, they also developed and released a massive publicly accessible multimedia data set for pattern recognition and artificial intelligence research called the Yahoo Flickr Creative Commons 100 Million (YFCC100M). Responding to the need for a large, publicly available data set for research, YFCC100M is a "training" database of 100 million images and videos.

Inspired by Living Systems

The computing architecture of the Livermore Brain and other ANNs is inspired by the nervous system of living beings. The exquisitely filigreed mesh of cells called neurons that forms our nervous system and brain processes inputs from the senses and allows us to recognize objects, understand how our environment is changing, and respond accordingly, among myriad other tasks. The computational building block of an ANN is also called a neuron or unit. Groups of units are connected linearly to form layers, with each unit in an input layer connected to each unit in the next layer, and so on through to the final output layer. The layers between input and output are called hidden layers. The deep ANN's predecessor, the shallow ANN, consisted of a few hidden layers, typically one or two. State-of-the-art deep ANNs have many more hidden layers, typically between 15 and 30, but as many as 152. "The deep neural network is one of the technological innovations that allows us to solve problems that shallow networks could not," explains Chen. "The combination of high-performance computing (HPC) power, massive data,

and deep neural networks is what makes possible human-performance-level image recognition with machines."

In the Livermore Brain, a unit is a block of software code. Each unit in the ANN possesses a set of weights (numbers between zero and one) that are "learned" through an optimization procedure to minimize the errors that the ANN makes on a training data set (such as the YFCC100M) consisting of input and desired output target pairs. As data are fed through the ANN during the training process, the resulting output is compared to the desired target. Errors between the ANN output and the target are "back propagated" through the network, assigning blame to the weights responsible for the error. As training proceeds, the weights change and converge toward an optimal configuration for minimizing the ANN's overall error. Feed millions of digitized images through a sufficiently fast and powerful ANN, and these weights begin to represent the underlying common features within the image, forming what Chen calls an "abstract concept space."

The process of training the Livermore Brain relies on an architectural element called the autoencoder, whose training targets are simply the original inputs themselves. That is, the ANN outputs a reconstructed version of the input image. Repeated millions of times with millions of images, this training allows the ANN to get better at reconstructing input. One set of neurons "learns" to recognize edges—boundaries—while the next may register shapes and shadows—the elements of faces, for example, eventually arriving at a set of elements and their relationships that form the class of faces. In 2012, a deep-learning ANN, the Google Brain, with one billion trainable weights running on 1,000 machines, "learned" to distinguish images of faces and cats by training on 10 million 200-by-200-pixel images sampled from YouTube—remarkably, without having labeled training data explicitly designating the category of each image. Researchers at Stanford University replicated the feat, training an ANN on the same number of images using just
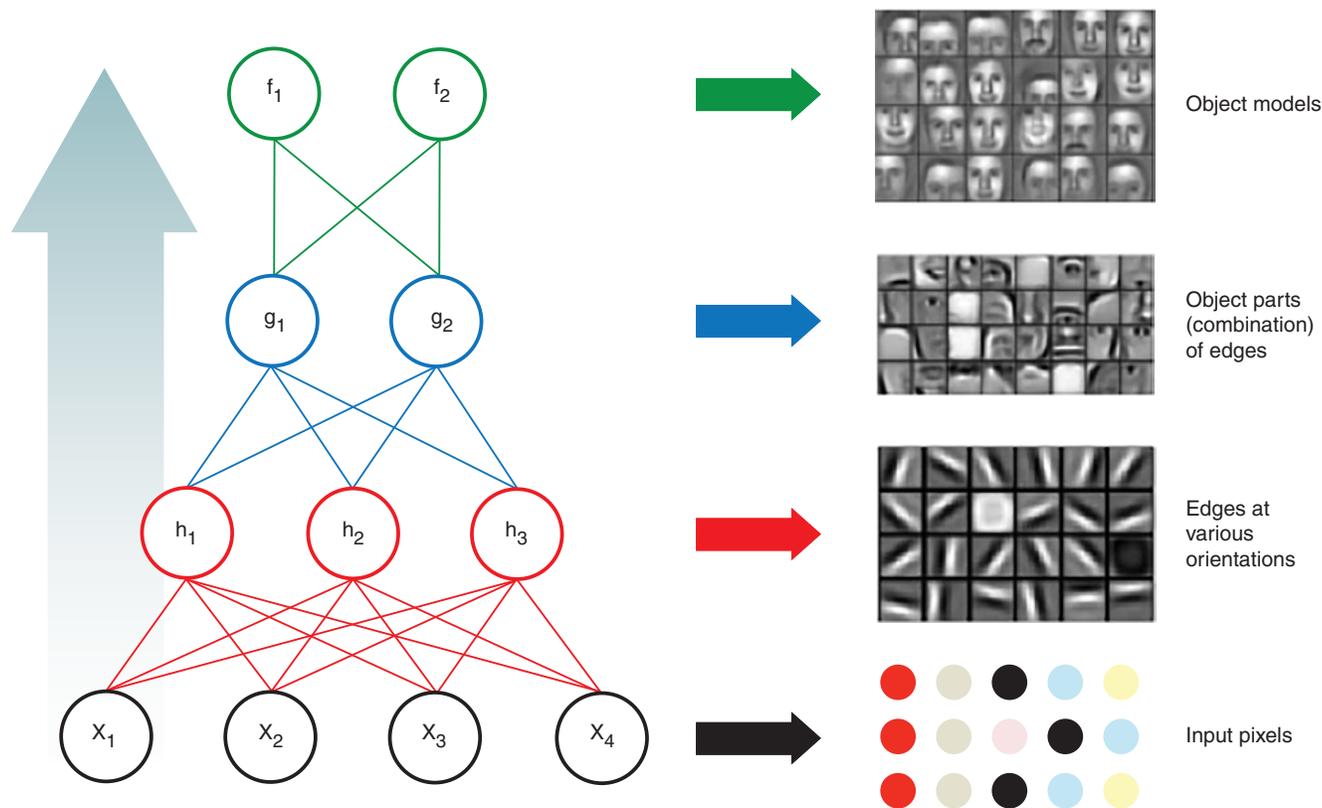
A deep learning neural network such as the Livermore Brain recognizes images through a hierarchy of layers composed of units represented by circles, with each unit connected to units in the layer above it. Starting with (black) input, the network first (red) recognizes the most basic components, such as edges, then (blue) parts containing multiple components, and finally (green) the object itself. (Right-hand images courtesy of Honglak Lee.)

three HPC nodes—CPUs assisted by graphics processing units (GPUs)—thanks to more efficient programming exploiting the massive parallelization afforded by GPUs.

Learning Is Better When Unsupervised

The ability to perform unsupervised learning is what gives the Livermore Brain and others like it so much applicability. "In supervised learning, you have to label the data, which is time consuming and labor intensive," says the Laboratory's Brian Van Essen. "Unsupervised learning allows the neural network to take advantage of massive amounts of unlabeled data to find the description of the data it needs to do a good reconstruction on its own. In other words, the network does its own feature extraction."

Incorporating Stanford code, the Livermore Brain represents the next big leap—the largest unsupervised learning–based deep neural network to date trained on image data. With 15 times as many parameters as Google Brain running on 98 nodes of Livermore's Edge supercomputer, the Livermore Brain has nine layers and

15 billion trainable parameters. Using the YFCC100M database's 99.2 million 300-by-300-pixel images, the network "learned" how to distinguish among a variety of image classes, including city skylines, buildings, aircraft, towers, and text, all without labeled training images.

Van Essen's group is working to further improve the speed and efficiency of LBANN by developing techniques to maximize each node's utilization—so that the network's computing resources are used as fully as possible—while minimizing communication between the nodes, which slows down the network.

Real Data from the Operating Room

Researchers in Chen's group have embarked on several projects funded by the Laboratory Directed Research and Development Program to improve the performance of and develop applications for LBANN. A collaboration between Lawrence Livermore and the University of California at San Francisco (UCSF) is working to apply electrocorticographic (ECoG) data from human brains to neural networks. The ECoG data, collected from epilepsy patients
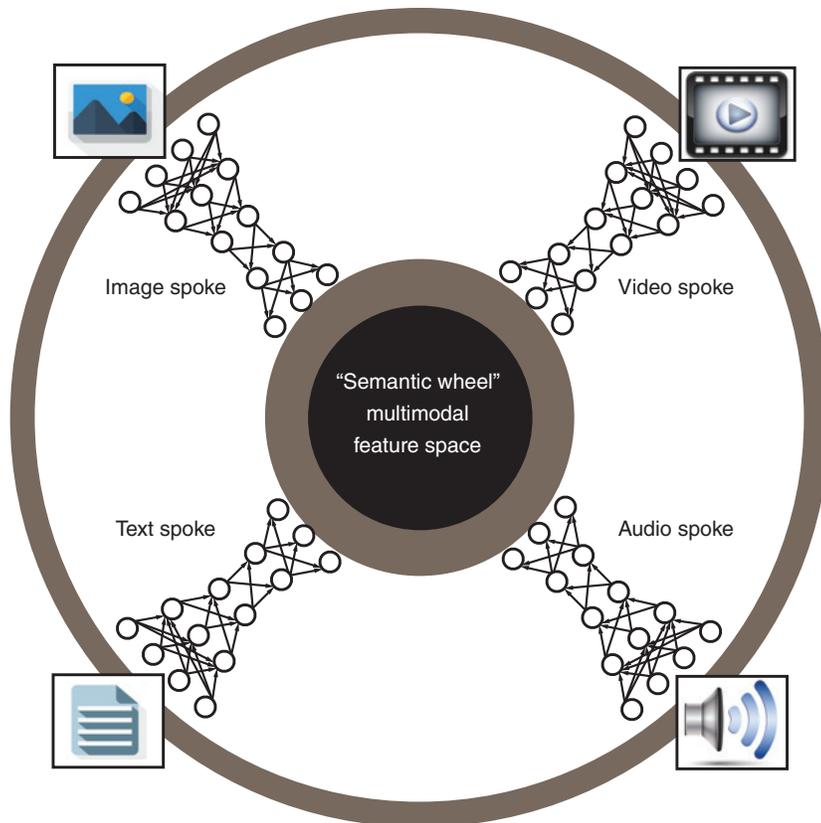
With the help of high-performance computers and the LBANN, Livermore researchers are developing the "semantic wheel" concept. It will map multimodal data—images, audio, text, and video—into a feature space that associates objects within similar data classes, such as words, images, and video all related to buildings.

awaiting brain surgery to treat their condition, are used by surgeons to determine the areas of a patient's brain on which to operate. Electrodes are implanted in their brains, and their activities in their hospital rooms are recorded on video cameras. The ECoG data thus provide a record of which areas of the brain are stimulated when the patient's body engages in activities such as moving arms and hands to eat or using muscles in the legs and elsewhere to shift position in bed.

In this project, Livermore researchers Kofi Boakye, Alan Kaplan, and their UCSF colleagues will feed video of patients through deep neural networks and attempt to correlate brain activity patterns from ECoG data with these specific movements. "We're interested both in what's going on in the brain and how we can use the techniques we develop to improve computer vision and analysis," explains Boakye. "We're going to cast as broad a net as possible and use the project to help us identify applications of interest to a variety of users beyond the medical community."

Things to Come—the Semantic Wheel

"The broader vision of Livermore's neural network research," explains Chen, "is to fuse different types of data—images, audio, video, and text—into a shared feature space where data of related concepts are proximal. Our framework for doing this is called the semantic wheel. The spokes in the wheel are deep neural networks

responsible for learning atomistic representations of individual data modalities. An alternating optimization procedure merges the output of individual spokes, resulting in a shared feature space that will enable the association of elements within images, audio, and video, with text descriptions and vice versa."

The semantic wheel approach could soon be able to find a car or a face—or a person throwing a ball to a dog on a grass field in a park—within thousands of images, or find relationships among variables in millions of data points generated by a high-energy physics experiment, for instance. Livermore's research is leading the way in this approach through the merging of HPC, advanced deep learning architectures, and the largest image data set ever published to create powerful new tools for basic science research and national security applications.

—*Allan Chen*