

Jumping-Off Place

The starting point for several of the team's studies is the MIT computer-vision program. The program implements a model of the primate visual cortex, which is where the brain processes most visual information. The visual cortex is also the best-understood part of the brain. The model is based on experimental studies of how monkeys and humans process visual information.

The team developed a new version of the MIT program to run on "hybrid-architecture" computers such as the Roadrunner supercomputer and a "mini" Roadrunner at the Los Alamos Center for Nonlinear Studies. The new program is called PetANNet, named for the fact that its computer-simulated neurons are connected to compose what's called a neural network, or neural net.

Like the MIT program, PetANNet implements a model of the "what" pathway, the neural pathway that identifies objects in one's field of view. A separate pathway, the "where" pathway, identifies the locations of objects in the visual field. The visual cortex is divided into four major areas from back to front: V1, V2, V4, and IT. The "what" and "where" pathways flow through all four areas, with the "what" pathway on the dorsal side (underside) of the gray matter and the "where" pathway on the ventral (upper) side. In PetANNet, information flows through the "what" pathway almost entirely in the forward direction from V1 to IT, that is, in a "feed-forward" fashion (see "Learn More.")

Visual information enters the "what" pathway through the lens (cornea) of the eye. The cornea focuses images onto the retina, at the back of the eye, where photoreceptors convert light to the electrical signals the brain's neurons use to communicate with each other. "Roughly speaking," Kenyon says, "your eye has about 500,000 photoreceptors, which is about equal to a half-a-megapixel camera."

The electrical signals from the retina go directly to the back of the brain, to V1, and are then processed through the visual cortex, starting with V1 and ending with IT. The field of view is first characterized in terms of simple visual features present in small square sections of the visual field and then in terms of combinations of simple features that represent more-complex features present in larger sections of the visual field. As the information is processed, individual neurons further up the processing hierarchy recognize features that are more and more complex and present in larger and larger sections of the visual field. Near the top of the processing hierarchy, in V4, complex features, such as ears and noses, are recognized by individual neurons that view sizable fractions of the visual field—a fact that has been proven through electrophysiology experiments. In IT, individual neurons respond to objects or types of objects that appear anywhere in the visual field, regardless of how they're lit or oriented. "The magic is that an object, say, a face, is identified in IT as belonging to a distinct category regardless of the scene it happens to be part of," says Bettencourt. The V1-to-IT processing hierarchy is illustrated in the figure in "Learn More." Feed-forward processing is thought to determine the minimum time necessary for primates to see and identify objects. Experiments show that when a scene is presented to the visual cortex of a monkey or a human, information initially flows mainly from the back of the brain to the front—that is, in a feed-forward fashion—rather than laterally or backwards (through "feedback" pathways). The slower processes related to lateral and feedback neural connections kick in after the feed-forward processes do, and those connections are not represented in PetANNet. So it's not surprising that when a scene is presented to a human for up to about 50 milliseconds, the human brain identifies objects with about the same accuracy as the program does—70 to 90 percent. But when presented with a scene for longer times, humans become nearly perfect—accurate to at least 99.999 percent. So the question is, how can feed-forward programs be improved?

The PetANNet Models

The four major areas in the "what" neural pathway of the primate visual cortex—the pathway that identifies objects in the visual field—are V1, V2, V4, and IT, color coded here to map to the PetANNNet model.

Visual processing begins with V1 dividing the field of view, which in this case contains a dog, into many small square sections, each consisting of about 7 × 7 pixels (photoreceptors). V1 contains "stacks" of neurons, and each stack views one of the small squares. Each neuron in a stack is a "feature detector" that responds to a specific feature in the square viewed by that stack.

The most-prevalent feature detectors in V1 are "edge detectors," each of which is a "simple-cell" neuron sensitive to an edge at a specific angle to the horizontal—exactly horizontal, standing vertical, slanted at 70 degrees, and so on. Other feature detectors are sensitive to, for example, color, spatial frequency, direction of motion, and so on. The visual information next goes to a layer of "complex-cell" neurons, where information is sampled and "pooled." One complex-cell neuron will sample the outputs of several edge detectors sensitive to the same edge angle. By "pooling" information in this way, the complex cell determines if that particular edge angle is present in a larger section of the field of view. Thus begins the process of identifying an object, regardless of where it is in the visual field or how it is oriented or lit.

In V2, a new set of simple-cell neurons monitors the outputs of combinations of the V1 complex-cell neurons over a larger field of view. Each combination represents a new feature that is more complex than those viewed by the stacks in V1 and that is present in a larger swath of the visual field. This process is then repeated in the higher levels of the processing hierarchy, where features are even more complex and appear over even larger swaths. Finally, in IT, individual neurons are associated with particular objects and categories of objects. Some of these neurons are activated whenever, say, a specific dog appears anywhere in the entire field of view. Others are activated whenever any kind of dog appears.