Editor: Rubin Landau, rubin@physics.oregonstate.edu

COMPUTER VISION, INSPIRED BY THE HUMAN BRAIN

By Pam Frost Gorder

As scientists work to develop intelligent machines, some are taking their cues from biology. Such is the case at the Center for Biological and Computational Learning (CBCL; http://cbcl.mit.edu/) at the Massachusetts Institute of Technology (MIT), where a computer model is emulating the human brain's vision center. The model replicates what happens during the first few fractions of a second after we see an object—the part of vision performed by the unconscious mind.

In the Blink of an Eye

Thomas Serre was a graduate student at CBCL when he began building the model. It was an extension of work by lab director Tomaso Poggio and former postdoctoral researcher Maximilian Riesenhuber, which they had reported earlier in *Nature Neuroscience* (vol. 2, no. 11, 1999, pp. 1019–1025). Serre's goal was to make the model mirror the human visual system's anatomy and physiology as closely as possible.

To test it, he and his colleagues sat volunteers in a darkened room and asked them to stare at a blank computer screen. An image flashed onscreen for only 20 milliseconds (ms)—more than 10 times faster than the blink of an eye. The image could have been a car, for instance, or a flower or cat, but it went by so fast that most people could hardly see it. Then the researchers asked them whether they saw an animal.

"People would say, 'Oh, I didn't see anything,'" Serre remembers. "And I'd tell them, 'Don't worry about it—just make your best guess.' They were right more often than not." In fact, they were right roughly 80 percent of the time.

So was the computer model. Its collection of algorithms replicated what neuroscientists suspect happens in the brain during those first few milliseconds: a stimulus enters the eye, and neurons carry the signal along a kind of one-way pipeline to the cerebral cortex, where a basic message registers ("I see an animal."). Then, scientists think, the cortex generates a feedback signal; other neurons fire, and the brain processes the image on a conscious level ("I see a cat.") and in context ("I see an orange cat in my living room. This is my cat."). In the MIT experiment, the image disappeared before the feedback could begin to isolate people's ability—and the model's—to identify objects using only one-way, unconscious processing.

The model closely matched the human volunteers' performance. It even tended to make the same mistakes, misidentifying the same objects. In the *Proceedings of the National Academy of Sciences* (vol. 104, no. 15, 2007, pp. 6424–6429) Serre, Poggio, and MIT neuroscientist Aude Oliva reported that their model appears to validate experts' notions of the brain's pipeline architecture.

Scenes Seen

Of the 30 regions of the brain known to contribute to vision, the MIT model only accounts for six that are thought to be key for object identification. Yet it learns to identify objects after examining only a few training images. In a paper in *IEEE Transactions on Pattern Analysis and Machine Intelligence* (vol. 29, no. 3, 2007, pp. 411–426), Serre, Poggio, and their colleagues reported that using fewer than 15 training images, their model matched the performance of state-of-the-art vision systems, some of which were trained with thousands of images. They used the model to recognize objects such as cars and people in busy street scenes.

Serge Belongie, a computer scientist at the University of California, San Diego (http://vision.ucsd.edu/sjb), says that the MIT approach "comprises many components that, taken individually, are familiar to the computer vision community, but there is novelty in the model of the primate visual cortex that they use to integrate and motivate these components." This particular combination of tools might not have been evident, were Poggio and his team not seeking inspiration from biology.

At Johns Hopkins University, mathematician Don Geman (http://cis.jhu.edu/people/faculty/geman/) has long been working on ways to help computers identify objects. A paper he coauthored in 1984 with his brother Stuart Geman (a mathematician at Brown University) is still the most cited paper in image processing (*IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741).

He says Poggio's team deserves a lot of credit for being one of the first to recognize the importance of hierarchical

 Copublished by the IEEE CS and the AIP 1521-9615/08/$25.00 ©2008 IEEE

Web Trends

For a brief look at current events, including program announcements and news items related to science and engineering, check out the following Web sites:

- Computer and Information Science and Engineering Pathways to Revitalized Undergraduate Computing Education (CPATH; www.nsf.gov/pubs/2008/nsf08516/nsf08516.htm). The US National Science Foundation (NSF) is accepting proposals for its CPATH program, which aims to revitalize undergraduate computing education in the US. Grant amounts range from US$50,000 to $1 million. Proposal deadline is 11 March 2008.
- EPSCoR Research Infrastructure Improvement Grant Program (www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf08500). Under the Experimental Program to Stimulate Competitive Research (EPSCor), the NSF is awarding grants to programs in historically underfunded areas. Proposals are due 4 January 2008.
- High-performance computing courses at Purdue University (www.hpcwire.com/hpc/1962653.html). Purdue University's Department of Computer and Information Technology will offer HPC courses in its 2008 spring semester.
- New possibilities for deaf students in computer science (www.nsf.gov/publications/pub_summ.jsp?ods_key=pr07135). This past summer, the NSF sponsored a nine-week program designed to increase the number of deaf or hearing-impaired students pursuing computer science degrees and careers. Instructors taught and communicated with students through sign language and real-time captioning systems. Applications for the upcoming summer program are now being accepted.
- Solicitation for projects that advance innovative computational thinking (www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf07603). The NSF's recent initiative, Cyber-Enabled Discovery and Innovation (CDI), seeks proposals in the following areas: data and knowledge, understanding complexity, and building virtual organizations. The NSF will accept letters of intent from 30 August 2008 to 30 September 2008.
- Sustainable digital data preservation (www.nsf.gov/publications/pub_summ.jsp?ods_key=nsf07061). The NSF is sponsoring an initiative that will create a data infrastructure of organizations to reliably preserve digital data in the science and engineering fields. Preliminary proposals are due 6 October 2008.

structure in vision, back in the 1980s. Hierarchical processing follows a tree-like pattern, from low-level coarse computations to high-level fine ones. "Most of the power-

Computing in SCIENCE & ENGINEERING

EDITOR IN CHIEF
Norman Chonacky, Yale Univ.
cise-editor@aip.org

ASSOCIATE EDITORS IN CHIEF
Jim X. Chen, George Mason Univ.,
jchen@cs.gmu.edu

Denis Donnelly, Siena College,
donnelly@siena.edu

Steven Gottlieb, Indiana Univ.,
sg@indiana.edu

Douglass E. Post, Carnegie Mellon Univ.,
post@ieee.org

David Winch, Kalamazoo College,
winch@TaosNet.com

EDITORIAL BOARD MEMBERS
Klaus-Jürgen Bathe, MIT, kjb@mit.edu
Michael W. Berry, Univ. of Tennessee, berry@cs.utk.edu
Bruce Boghosian, Tufts Univ., bruce.boghosian@tufts.edu
Hans-Joachim Bungartz, Institut für Informatik, bungartz@in.tum.de
Judy Cushing, The Evergreen State College, judyc@evergreen.edu
George Cybenko, Dartmouth College, gvc@dartmouth.edu
Massimo DiPierro, DePaul University, mdipierro@cti.depaul.edu
Jack Dongarra, Univ. of Tennessee, dongarra@cs.utk.edu
Rudolf Eigenmann, Purdue Univ., eigenman@ecn.purdue.edu
William J. Feiereisen, Los Alamos Nat'l Lab, wjf@lanl.gov
Geoffrey Fox, Indiana Univ., gcf@grids.ucs.indiana.edu
Sharon Glotzer, Univ. of Michigan, sglotzer@umich.edu
Anthony C. Hearn, RAND, hearn@rand.org
Charles J. Holland, Darpa, charles.holland@darpa.mil
M.Y. Hussaini, Florida State Univ., myh@cse.fsu.edu
Rachel Kuske, Univ. of British Columbia, rachel@math.ubc.ca
David P. Landau, Univ. of Georgia, dlandau@hal.physast.uga.edu
R. Bowen Loftin, Texas A&M University, Galveston, loftin@tamug.edu
B. Vincent McKoy, California Inst. of Technology, mckoy@its.caltech.edu
Jill P. Mesirov, Whitehead/MIT Ctr. for Genome Research, mesirov@genome.wi.mit.edu
Dianne P. O'Leary, Univ. of Maryland, oleary@cs.umd.edu
William H. Press, Los Alamos Nat'l Lab., wpress@lanl.gov
John Rundle, Univ. of California, Davis, rundle@physics.ucdavis.edu
Henrik Schmidt, MIT, henrik@mit.edu
Greg Wilson, Univ. of Toronto, gvwilson@third-bit.com

CONTRIBUTING EDITORS
Francis Sullivan, The Last Word, fran@super.org
Paul F. Dubois, Café Dubois, paul@pfdubois.com

DEPARTMENT EDITORS

Books: Mario Belloni, Davidson College, mabelloni@davidson.edu
Computing Prescriptions: Isabel Beichl, Nat'l Inst. of Standards and Tech., isabel.beichl@nist.gov
Computer Simulations: Muhammad Sahimi, University of Southern California, moe@iran.usc.edu, and Dietrich Stauffer, Univ. of Köhn, stauffer@thp.uni-koeln.de
Education: Michael Dennin, Univ. of Calif., Irvine, mdennin@uci.edu, and Steven F. Barrett, Univ. of Wyoming, steveb@uwyo.edu
News: Rubin Landau, Oregon State Univ., rubin@physics.oregonstate.edu
Scientific Programming: Konstantin Läufer, Loyola Univ., Chicago, klaufer@cs.luc.edu, and George K. Thiruvathukal, Loyola Univ., Chicago, gkt@cs.luc.edu
Technologies: Michael A. Gray, American Univ., gray@american.edu, and James D. Myers, Collaborative Systems, NCSA, jimmyers@ncsa.uiuc.edu
Visualization Corner: Claudio T. Silva, Univ. of Utah, csilva@cs.utah.edu, and Joel E. Tohline, Louisiana State Univ., tohline@rouge.phys.lsu.edu

STAFF

Senior Editor: Jenny Stout, jstout@computer.org
Senior Editorial Services Manager: Crystal R. Shif
Magazine Editorial Manager: Steve Woods
Staff Editors: Kathy Clark-Fisher, Rebecca L. Deuel, and Brandi Ortega
Production Editor: Monette Velasco
Publications Coordinator: Hazel Kosky, cise@computer.org
Technical Illustrator: Alex Torres

Senior Advertising Coordinator: Marian Anderson
Marketing Manager: Georgann Carter
Senior Business Development Manager: Sandra Brown

AIP Staff:
Circulation Director: Jeff Bebee, jbebee@aip.org
Editorial Liaison: Charles Day, cday@aip.org

IEEE Antennas and Propagation Society liaison: Don Wilton, Univ. of Houston, wilton@uh.edu
IEEE Signal processing Society liaison: Elias S. Manolakos, Northeastevrn Univ., elias@neu.edu

CS PUBLICATIONS BOARD

Sorel Reisman (chair), Angela Burgess, Chita R. Das, Richard H. Eckhouse, Van Eden, Frank E. Ferrante, David A. Grier, Pamela Jones, Phillip A. Laplante, Simon Liu, Paolo Montuschi, Jon Rokne, Linda I. Shafer, Steven L. Tanimoto

CS MAGAZINE OPERATIONS COMMITTEE

Robert E. Filman (chair), David Albonesi, Arnold (Jay) Bragg, Carl Chang, Kwang-Ting (Tim) Cheng, Norman Chonacky, Fred Douglis, Hakan Erdogmus, James Hendler, Carl E. Landwehr, Dejan Milojicic, Sethuraman (Panch) Panchanathan, Maureen Stone, Roy Want, Jeff Yost

EDITORIAL OFFICE

COMPUTING IN SCIENCE & ENGINEERING
10662 Los Vaqueros Circle, Los Alamitos, CA 90720 USA
Phone +1 714 821 8380; www.computer.org/cise/

IEEE Antennas & Propagation Society

IEEE Signal Processing Society

ful machine vision systems today use hierarchical processing in one way or another. Tommy was there early—and vocally," Geman says. "But what separates him from the others in this area today is that he's determined to replicate what the brain does."

This research could also enhance scientists' understanding of the brain, says Tarek El Dokor, director of the Machine Vision Lab at Embry-Riddle Aeronautical University (http://vision.pr.erau.edu). "It provides another critical link in our quest for a better understanding of the underlying neuronal mechanisms of learning in the visual pathways, and the impact of such learning on the science of machine vision," El Dokor says.

He adds that the MIT work holds significant promise for the future, in part because its algorithms can be exploited via massively parallel computing. In fact, his team plans to download the MIT code and implement it using general-purpose computation on GPUs (graphics processing units). So-called general-purpose computing on GPUs (GPGPU) computing offers a massively multithreaded architecture, which El Dokor says is ideal for modeling neural networks.

Taking It All In

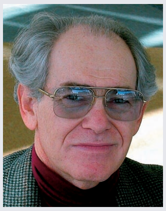
Belongie's lab also has a feed-forward model for computer vision, but adds feedback to identify objects based on context. First, the model picks out individual objects in an image and identifies them. Then a postprocessing step double-checks whether those objects have been identified correctly. Google Sets (http://labs.google.com/sets) provides guidance by generating lists of objects that belong together. For example, in an image of a tennis player, the model might correctly identify the person, the tennis racket, and the tennis court, but misidentify the round yellow object flying across the court as a lemon. Then, during postprocessing, the model cross-references the list of objects with Google Sets, and determines that the yellow object is most likely a tennis ball.

Incorporating context into such models is difficult. Geman calls this the central dilemma of machine vision: highly context-sensitive models are almost intractable computationally. The feed-forward part of the vision system (the one-way pipeline of unconscious processing done by the MIT model) and the feedback part (where higher, conscious processing happens) interact to make the computations extremely complex. Weird effects result.

"If you have a machine vision system that is capable, on average, of identifying nine out of 10 faces in a picture,

Observatoire Landau

Works and Plays Well with Others: From Vision to Wiki



By Rubin Landau, Department Editor

Question: How do you tell a physicist from an engineer at a party?
Answer: The engineer looks at your shoes when he speaks to you.

The unspoken punchline to this joke, which is sometimes told the other way around, is that neither group has a reputation for outstanding social skills, with some of us being downright introverts or one-sided conversationalists. Yet, if the groups that compose a large fraction of the *CiSE* readership are to work and play well together in future cybercommunities, we might need to develop better social skills and communication tools. Although this trite a statement sounds like a line from a self-help book, it's actually a consequence of the Golden Rule: those who have the gold, make the rules. The US National Science Foundation (the pot of gold) has a vision for 21st century science and engineering in which a greatly enhanced cyberinfrastructure (the focus of the last Observatoire) will support the creation of effective, virtual organizations that share networked resources (the rule).

Although several of us already participate in research and education groups with distributed members via conference calls, emails, Web pages, and wikis (more on that later), participation is expected to increase in number and importance as these groupware tools continue to evolve. This, too, might sound like another line from our maligned self-help book, but I believe it could be a great improvement for those of us struggling at resource-challenged institutions. As the cyberinfrastructure continues to develop, I can see us wanting to be members of the virtual user groups that gain real-time networked access to world-class resources such as experimental facilities, distributed sensor networks, high-performance computing (HPC) systems, data collections, and analysis and simulation tools. Such virtual groups will probably be self-organizing, might span multiple communities, and might be called collaboratories, grid communities, science gateways and portals, teragrids, or, in popular culture, social networks.

Visions might be what university presidents and CEOs sell for a living, but how might this work out for practicing scientists and educators, and for our socially challenged colleagues? Most readers have probably heard about social networking sites such as Facebook and MySpace or the peer-to-peer networks that are all the rage with college students and teenagers (speaking of the social-skills challenged) or blogs (Web logs), which are changing the face of journalism. In addition, Really Simple Syndication (RSS) feeds automatically inform people when changes that might interest them are made to their favorite Web sites or blogs. These are all Web 2.0 technologies, and they've stimulated a change in culture right before our (aging) eyes. Furthermore, the business world is already using commercial groupware and project management software produced by giants like Microsoft and Adobe, as well as content and knowledge management systems that permit group interaction. But the Web 2.0 technology that deserves special attention, I believe, is the wiki.

A wiki is a software system installed on a Web server that acts like a collaborative Web site. In contrast to the usual Web pages in which users can only read the pages or interact with them via special constructs such as forms and JavaScripts, wikis permit everyone in a predefined community to edit and create Web pages in truly simple ways. The wiki that most readers have probably heard of is Wikipedia, the online encyclopedia. In it, anyone can edit or create entries, with the large number of interested readers keeping the entries relatively recent and accurate. Although I don't believe that Wikipedia is appropriate for academic scholarship (some entries appear to be written by not-quite experts who are generous in sharing their confusion), I do consult it early when learning a new subject and in finding references, as apparently do some Supreme Court justices and *New York Times* reporters. There are many wikis in addition to Wikipedia, including the original site, WikiWikiWeb created by Ward Cunningham in 1994. In addition, a variety of open source programs exist that you can use to set up a wiki, such as Dokuwiki, Mediawiki, and Swiki, each with its own special features.

continued on p. 10

you'll have one hallucination per picture," Geman says. That is, given a photograph of 10 people, the computer might identify nine of the faces, miss one face, and then "see" at least one face where there is none. "If you try to tune the system to get 99 out of 100 faces, you might have hundreds of hallucinations."

Although Geman concedes that he didn't intend the word "hallucination" literally in that statement, Serre and Poggio are actually hoping to make their computer model hallucinate—to gain insight into human brain disorders such as schizophrenia.

"We have a hypothesis that schizophrenia comes from an imbalance between feed-forward and feedback processing in the brain," Serre says. "And so we think our

continued from p. 9

The importance of wikis is that they let a group of people collaborate on a project simply and quickly via the Internet (or an intranet). Like Wikipedia, the pages are simple but functional, with the focus on text over graphics. Consequently, the source pages are close enough to plaintext to be read and edited without markup symbols getting in the way. When successful, this encourages people to share and record information. Although we can use email to do this, idea streams and data tend to get lost or ignored in amid all the junk mail, and rarely get assembled into a coherent whole that remains in place as the project's archive. Indeed, to many people, wikis are a reincarnation of the democratic, generative approach to the Web that encouraged the original users to build it, an attitude that seems to have been lost in the Web's commercial developments.

Although news stories indicate that wikis aren't as popular as other Web 2.0 technologies, they also indicate that many businesses and organizations have found this mix of technology and sociology useful. Already, wikis play a key role in education, particularly for online courses (even I use one in my computational physics class). The early conclusion on the effectiveness of wikis for education is similar to that for online courses; some are successful and some aren't, with the key being the pedagogy not the technology. The experience of several teachers, myself included, supports the moral of the opening joke. Many science and engineering students don't take naturally to the social interactions inherent in wikis, but if you incorporate one as a key element in a course and require that students use it to be successful, then they'll benefit from increased interactions with the materials and from peer stimulation. For example, some courses encourage students to post their required papers on a wiki so that other students can critique the papers before submission; this tends to improve the grades on the papers (surprise, surprise) and possibly student learning. Other courses have encouraged students to submit exam questions and solutions on a wiki, with the solutions edited by other students but not the teacher. The teacher then chooses a few of these questions for the exam.

But beginnings are hard, and at present, the *CiSE* editorial board doesn't have a wiki.

feed-forward model provides a good basis for making comparisons." By disabling certain algorithms, they hope to simulate the "computing" done by the brain's damaged areas. They're now working with McLean Hospital, a psychiatric hospital affiliated with Harvard Medical School, to give their computerized image identification test to patients with schizophrenia.

Looking Ahead

The vast amount of image data available via Web sites such as Flickr and Google presents an excellent opportunity for the development of massively data-driven computational approaches to object recognition, Belongie says. "Unfortunately, massively data-driven approaches tend to be massively expensive, so much work remains to make such methods viable in practice," he adds. "Cortically inspired approaches such as those from Poggio's lab could provide inspiration for the kind of parallelism needed to achieve this goal."

The payoffs could be enormous, says Pietro Perona, director of Caltech's Center for Neuromorphic Systems Engineering (www.cnse.caltech.edu/). "If you think of MRI machines in medicine—all those volumes of medical data scanned—those images are examined briefly by a doctor, but if there was a computer churning on them, you can imagine the number of cancers that could be found, or the trends happening in a population that you could discover."

He cited as another example the search for missing pilot Steve Fossett, who disappeared on a flight in the Nevada desert in September 2007. Google has made satellite images of the entire area available, and volunteers are searching the images via a Web interface on Amazon.com (www.mturk.com/mturk/welcome). Although nobody has found Fossett's plane yet, they have found the wreckage from other plane crashes that happened long before. Wouldn't Fossett have been found by now, Perona asks, if computers were capable of scanning these images?

"Computers are blind and deaf. They do not see the pictures, they do not hear the sounds. And most of the 'juice'—the information—is in there. A computer should be an expert of all the content," he says. Right now, search engines can only find an image based on the text, or tags, that people use to label data. It's as if all the imagery on the Internet were a vast library in which visitors could only walk down the aisles and read the covers of books on the shelves. "We want to be able to take the books off the shelf and read them," he adds.

"Seeing" computers could form the basis for the first true artificial intelligence. They could support assistive devices for the disabled; they could even lend insight to the workings of the human mind. But for all of machine vision's potential, computers still fall short of the task. Even after 20 years of intense research, the human brain is still the clear winner when it comes to identifying objects.

Geman believes the answer has something to do with our seemingly innate ability to solve problems by first thinking broadly. We identify objects, he says, by using a process similar to a game of "20 questions." We start by examining all the possible answers at once, and we quickly hone in on the right one.

Now a postdoctoral researcher at CBCL, Serre reports that his group is working to incorporate feedback into its model and to add algorithms that emulate eye movements. They're also hoping to develop the model to recognize human movements and objects in motion—two other things our brains do very well.

"If you understand the human visual system well enough, you would hope that one day you could make machines that would work just as well as human brains," he says. "I think that ultimately machines will be even better than the human brain, but if we could just emulate the visual system, that would be a good start."

Pam Frost Gorder is a freelance science writer based in Columbus, Ohio.

Be on the Cutting Edge of Artificial Intelligence!

IEEE Intelligent Systems seeks papers on all aspects of artificial intelligence, focusing on the development of the latest research into practical, fielded applications.

For guidelines, see www.computer.org/mc/intelligent/author.htm.

IEEE Intelligent Systems

The #1 AI Magazine
www.computer.org/intelligent

AAPT 2008 SUMMER MEETING

Physics From the Ground Up July 19-23 • Edmonton, Alberta

Explore the "Festival City" of Edmonton and enjoy the many exciting cultural, arts, dining, and sporting activities:

- Visit the TELUS World of Science™, a hands-on space and science museum
- National Institute of Nanotechnology involves researchers in physics, chemistry, engineering, biology, informatics, pharmacy, and medicine
- Join the University of Alberta in celebrating its 100th anniversary
- Royal Alberta Museum brings 11,000 years of native history to life
- Lovers of speed can watch drivers race at 390 km/hr at the Edmonton Champ Car Race
- Enjoy the long summer evenings and pleasant sunny climate
- Go hiking in the Rocky Mountains, fly-fishing in northern Alberta's lakes, bird watching at Elk Island National Park or simply take in the vast scenic beauty
- West Edmonton Mall: the world's largest shopping center and Alberta's number one tourist attraction; over 800 stores, 100 eateries, and 9 world class attractions

U.S. citizens: Passport required. Apply early to ensure timely processing.

edmonton'08 AAPT | PHYSICS FROM THE GROUND UP

AAPT PHYSICS EDUCATION

www.aapt.org