Mimicking How the Brain Recognizes Street Scenes

February 6 2007

At last, neuroscience is having an impact on computer science and artificial intelligence (AI). For the first time, scientists in Tomaso Poggio's laboratory at the McGovern Institute for Brain Research at MIT applied a computational model of how the brain processes visual information to a complex, real world task: recognizing the objects in a busy street scene. The researchers were pleasantly surprised at the power of this new approach.

"People have been talking about computers imitating the brain for a long time," said Poggio, who is also the Eugene McDermott Professor in the Department of Brain and Cognitive Sciences and the co-director of the Center for Biological and Computational Learning at MIT. "That was Alan Turing's original motivation in the 1940s. But in the last 50 years, computer science and AI have developed independently of neuroscience. Our work is biologically inspired computer science."

"We developed a model of the visual system that was meant to be useful for neuroscientists in designing and interpreting experiments, but that also could be used for computer science," said Thomas Serre, a former PhD student and now a post-doctoral researcher in Poggio's lab and lead author a paper about the street scene application in the 2007 *IEEE Transactions on Pattern Analysis and Machine Intelligence*. "We chose street scene recognition as an example because it has a restricted set of object categories, and it has practical social applications."

Near-term applications include surveillance and automobile driver's

assistance, and eventually visual search engines, biomedical imaging analysis, robots with realistic vision. On the neuroscience end, this research is essential for designing augmented sensory prostheses, such as one that could replicate the computations carried by damaged nerves from the retina. "And once you have a good model of how the human brain works," Serre explained, "you can break it to mimic a brain disorder." One brain disorder that involves distortions in visual perception is schizophrenia, but nobody understands the neurobiological basis for those distortions.

"The versatility of the biological model turns computer vision from a trick into something really useful," said co-author Stanley Bileschi, a post-doctoral researcher in the Poggio lab. He and co-author Lior Wolf, a former post-doctoral associate who is now on the faculty of the Computer Science Department at Tel-Aviv University, are working with the MIT entrepreneur office, the Deshpande Center in the Sloan School. This center helps MIT students and professors bridge the gap between an intriguing idea or technology and a commercially viable concept.

Recognizing Scenes

The IEEE paper describes how the team "showed" the model randomly selected images so that it could "learn" to identify commonly occurring features in real-word objects, such as trees, cars, and people. In so-called supervised training sessions, the model used those features to label by category the varied examples of objects found in digital photographs of street scenes: buildings, cars, motorcycles, airplanes, faces, pedestrians, roads, skies, trees, and leaves. The photographs derive from a Street Scene Database compiled by Bileschi.

Compared to traditional computer-vision systems, the biological model was surprisingly versatile. Traditional systems are engineered for specific object classes. For instance, systems engineered to detect faces

or recognize textures are poor at detecting cars. In the biological model, the same algorithm can learn to detect widely different types of objects.

To test the model, the team presented full street scenes consisting of previously unseen examples from the Street Scene Database. The model scanned the scene and, based on its supervised training, recognized the objects in the scene. The upshot is that the model learned from examples, which, according to Poggio, is a hallmark of artificial intelligence.

Modeling Object Recognition

Teaching a computer how to recognize objects has been exceedingly difficult because a computer model has two paradoxical goals. It needs to create a representation for a particular object that is very specific, such as a horse as opposed to a cow or a unicorn. At the same time the representation must be sufficiently "invariant" so as to discard meaningless changes in pose, illumination, size, position, and many other variations in appearances.

Even a child's brain handles these contradictory tasks easily in rapid object recognition. Pixel-like information enters from the retina and passes in a fast feed-forward, bottom-up sweep through the hierarchical architecture of the visual cortex. What makes the Poggio lab's model so innovative and powerful is that, computationally speaking, it mimics the brain's own hierarchy. Specifically, the "layers" within the model replicate the way neurons process input and output stimuli – according to neural recordings in physiological labs. Like the brain, the model alternates several times between computations that help build an object representation that is increasingly invariant to changes in appearances of an object in the visual field and computations that help build an object representation that is increasingly complex and specific to a given object.

The model's success validates work in physiology labs that have measured the tuning properties of neurons throughout visual cortex. By necessity, most of those experiments are made with simplistic artificial stimuli, such as gratings, bars, and line drawings that bear little resemblance to real-world images. "We put together a system that mimics as closely as possible how cortical cells respond to simple stimuli like the ones that are used in the physiology lab," said Serre. "The fact that this system seems to work on realistic street scene images is a concept proof that the activity of neurons as measured in the lab is sufficient to explain how brains can perform complex recognition tasks."

Making it More Useful

The model used in the street scene application mimics only the computations the brain uses for rapid object recognition. The lab is now elaborating the model to include the brain's feedback loops from the cognitive centers. This slower form of object recognition provides time for context and reflection, such as: if I see a car, it must be on the road not in the sky. Giving the model the ability to recognize such semantic features will empower it for broader applications, including managing seemingly insurmountable amounts of data, work tasks, or even email. The team is also working on a model for recognizing motions and actions, such as walking or talking, which could be used to filter videos for anomalous behaviors – or for smarter movie editing.

The Street Scene database is freely available at Center for Biological and Computational Learning website (cbcl.mit.edu/). Co-author Maximilian Riesenbuber began the work on the model for visual recognition for his Ph.D. dissertation in Poggio's lab and continues this work as assistant professor of neuroscience at the Georgetown University Medical Center. This research was partially funded by the U.S. Defense Advanced Research Projects Agency (DARPA), U.S. Office of Naval Research, and the U.S. National Science Foundation-National Institutes of Health.