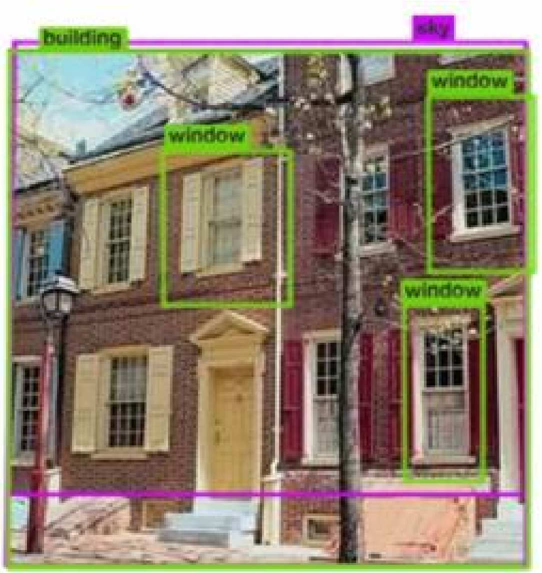
Context is ev ... well, something, anyway

March 5 2010, by Larry Hardesty



Standard object recognition software mistakenly detects a sofa, a cabinet, and a mirror in a street scene (right), but a new MIT system corrects those errors (left) using statistical information about how often certain types of objects occur together. Image: Myung 'Jin' Choi

Today, computers can't reliably identify the objects in digital images. But if they could, they could comb through hours of video for the two or three minutes that a viewer might be interested in, or perform web searches where the search term was an image, not a sequence of words. And of course, object recognition is a prerequisite for the kind of home assistance robot that could execute an order like "Bring me the stapler." Now, MIT researchers have found a way to improve object recognition systems by using information about context. If the MIT system thinks it's identified a chair, for instance, it becomes more confident that the rectangular thing nearby is a table.

A typical object recognition system will scan a digital image for groups of pixels that differ from those around them; those pixels could define an edge, a corner or some other feature of an object. Usually, the system has been trained on a set of sample images, which teaches it how to correlate feature patterns with particular objects.

Some researchers have tried to use context information to refine those correlations. But according to Myung "Jin" Choi, a grad student in MIT's Laboratory for Information and Decision Systems and one of the leaders of the new project, those researchers were generally working with a standard training set that included examples of only about 20 different types of object. In that case, it was fairly straightforward to specify how frequently each object co-occurred with every other object in the set.

Upping the ante

A system that could recognize only 20 different objects, however, wouldn't be very useful. And with a large number of objects, it becomes computationally impractical to consider the frequency of all possible two-object combinations. In work to be presented at the IEEE Conference on Computer Vision and Pattern Recognition this summer, Choi and her colleagues — including graduate student Joseph Lim and Professors Antonio Torralba and Alan Willsky — describe a different approach. Working with a training set that included more than 4,000 images and 107 different types of objects, they created algorithms that pored through the images and automatically constructed a hierarchical map of the object categories — kind of like the organizational chart for a large company, which shows who reports to whom. In the map, each object is connected to at most one object above it in the hierarchy (everyone in the organization reports to only one person), drastically reducing the number of connections that the system has to consider. The connection between any two objects is given a weight that indicates how often the objects appear together in the training images. The map also encodes information about the typical relative locations of two connected objects: buildings generally appear above roads, for instance, not below them.

When the system analyzes a new image, it uses standard object recognition algorithms to generate a list of candidate objects, together with each object's "confidence score" — a statistical measure of how likely the object is to have been correctly identified. Then it revises those scores on the basis of the information encoded in the contextual map.

In experiments, Choi compared the performance of the bare object recognition algorithms with their performance when augmented by the contextual map. In both cases, she considered the three objects per image with the highest confidence scores. The bare algorithms correctly identified all three objects roughly 14 percent of the time; with the addition of the contextual map, the success rate jumped to about 25 percent.

Long row to hoe

Of course, that means that the system still failed to correctly identify three objects per image about 75 percent of the time, which shows just how difficult the problem of object recognition remains. "Context really is essential," says Serge Belongie, an associate professor of computer science at the University of California, San Diego who has worked on both object recognition in general and context-based object recognition in particular. "It deserves a proper treatment, and Jin is doing that." But Belongie cautions that context awareness will never be more than an augmentation of an underlying system that recognizes objects from visual features. “We absolutely cannot afford to take our eye off the ball of the component recognition systems that need to feed these context engines,” he says. And, he adds, to be useful, object recognition systems will need to be much more precise than today’s prototypes are. “Imagine that you take a picture of a wild mushroom while you’re hiking,” Belongie says, “and then you send it to the system to find out what it is. And it says, ‘Mushroom!’ You’re like, Thanks. That’s really useful. I knew that part.”

Nonetheless, Choi is continuing to improve her contextual-map system, against the day when the underlying algorithms are more reliable. The next version of the system, she says, will add entries to the map that, in effect, represent higher-level scene descriptions. Street scenes, for instance, may frequently feature sky, buildings and roads, while building interiors may frequently feature floors, walls and windows. The system won’t need to explicitly label these additional map entries, however; it will simply register them as foci around which certain types of objects regularly cluster. She’s confident that this modification will make the added benefits of context awareness even more acute.

Provided by Massachusetts Institute of Technology

Citation: Context is ev ... well, something, anyway (2010, March 5) retrieved 26 April 2024 from https://phys.org/news/2010-03-context-ev.html