THE PASTRY A.I. THAT LEARNED TO FIGHT CANCER

In Japan, a system designed to distinguish croissants from bear claws has turned out to be capable of a whole lot more.

By James Somers

March 18, 2021

Illustration by Julian Glander

One morning in the spring of 2019, I entered a pastry shop in the Ueno train station, in Tokyo. The shop worked cafeteria-style. After taking a tray and tongs at the front, you browsed, plucking what you liked from heaps of baked goods. What first struck me was the selection, which seemed endless: there were croissants, turnovers, Danishes, pies, cakes, and open-faced sandwiches piled up everywhere, sometimes in dozens of varieties. But I was most surprised when I got to the register. At the urging of an attendant, I slid my items onto a glowing rectangle on the counter. A nearby screen displayed an image, shot from above, of my doughnuts and Danish. I watched as a set of jagged, neon-green squiggles appeared around each item, accompanied by its name in Japanese and a price. The system had apparently recognized my pastries by sight. It calculated what I owed, and I paid.

I tried to gather myself while the attendant wrapped and bagged my items. I was still stunned when I got outside. The bakery system had the flavor of magic—a feat seemingly beyond the possible, made to look inevitable. I had often imagined that, someday, I'd be able to point my smartphone camera at a peculiar flower and have it identified, or at a chess board, to study the position. Eventually, the tech would get to the point where one could do such things routinely. Now it appeared that we were in this world already, and that the frontier was pastry.

Computers learned to see only recently. For decades, image recognition was one of the grand challenges in artificial intelligence. As I write this, I can look up at my shelves: they contain books, and a skein of yarn, and a tangled cable, all inside a cabinet whose glass enclosure is reflecting leaves in the trees outside my window. I can't help but parse this scene—about a third of the neurons in my cerebral cortex are implicated in processing visual information. But, to a computer, it's a mess of color and brightness and shadow. A computer has never untangled a cable, doesn't get that glass is reflective, doesn't know that trees sway in the wind. A.I. researchers used to think that, without some kind of model of how the world worked and all that was in it, a computer might never be able to distinguish the parts of complex scenes. The field of "computer vision" was a zoo of algorithms that made do in the meantime. The prospect of seeing like a human was a distant dream.

All this changed in 2012, when Alex Krizhevsky, a graduate student in computer science, released AlexNet, a program that approached image recognition using a technique called deep learning. AlexNet was a neural network, "deep" because its simulated neurons were arranged in many layers. As the network was shown new images, it guessed what was in them; inevitably, it was wrong, but after each guess it was made to adjust the connections between its layers of neurons, until it learned to output a label matching the one that researchers provided. (Eventually, the interior layers of such networks can come to resemble the human visual cortex: early layers detect simple features, like edges, while later layers perform more complex tasks, such as picking out shapes.) Deep learning had been around for years, but was thought impractical. AlexNet showed that the technique could be used to solve real-world problems, while still running quickly on cheap computers. Today, virtually every A.I. system you've heard of—Siri, AlphaGo, Google Translate —depends on the technique.

The drawback of deep learning is that it requires large amounts of specialized data. A deep-learning system for recognizing faces might have to be trained on tens of thousands of portraits, and it won't recognize a dress unless it's also been shown thousands of dresses. Deep-learning researchers, therefore, have learned to collect and label data on an industrial scale. In recent years, we've all joined in the effort: today's facial recognition is particularly good because people tag themselves in pictures that they upload to social networks. Google asks users to label objects that its A.I.s are still learning to identify: that's what you're doing when you take those "Are you a bot?" tests, in which you select all the squares containing bridges, crosswalks, or streetlights. Even so, there are blind spots. Self-driving cars have been known to struggle with unusual signage, such as the blue stop signs found in Hawaii, or signs obscured by dirt or trees. In 2017, a group of computer scientists at the University of California, Berkeley, pointed out that, on the Internet, almost all the images tagged as "bedrooms" are "clearly staged and depict a made bed from 2-3 meters away." As a result, networks have trouble recognizing real bedrooms.

It's possible to fill in these blind spots through focussed effort. A few years ago, I interviewed for a job at a company that was using deep learning to read X-rays, starting with bone fractures. The programmers asked surgeons and radiologists from some of the best hospitals in the U.S. to label a library of images. (The job I interviewed for wouldn't have involved the deep-learning system; instead, I'd help improve the Microsoft Paint-like program that the doctors used for labelling.) In Tokyo, outside the bakery, I wondered whether the pastry recognizer could possibly be relying on a similar effort. But it was hard to imagine a team of bakers assiduously photographing and labelling each batch as it came out of the oven, tens of thousands of times, for all the varieties on offer. My partner suggested that the bakery might be working with templates, such that every pain au chocolat would have precisely the same shape. An alternative suggested by the machine's retro graphics—but perplexing, given the system's uncanny performance—was that it wasn't using deep learning. Maybe someone had gone down the old road of computer vision. Maybe, by really considering what pastry looked like, they had taught their software to see it.

Hisashi Kambe, the man behind the pastry A.I., grew up in Nishiwaki City, a small town that sits at Japan's geographic center. The city calls itself Japan's navel; surrounded by

mountains and rice fields, it's best known for airy, yarn-dyed cotton fabrics woven in intricate patterns, which have been made there since the eighteenth century. As a teen-ager, Kambe planned to take over his father's lumber business, which supplied wood to homes built in the traditional style. But he went to college in Tokyo and, after graduating, in 1974, took a job in Osaka at Matsushita Electric Works, which later became Panasonic. There, he managed the company's relationship with I.B.M. Finding himself in over his head, he took computer classes at night and fell in love with the machines.

In his late twenties, Kambe came home to Nishiwaki, splitting his time between the lumber mill and a local job-training center, where he taught computer classes. Interest in computers was soaring, and he spent more and more time at the school; meanwhile, more houses in the area were being built in a Western style, and traditional carpentry was in decline. Kambe decided to forego the family business. Instead, in 1982, he started a small software company. In taking on projects, he followed his own curiosity. In 1983, he began working with NHK, one of Japan's largest broadcasters. Kambe, his wife, and two other programmers developed a graphics system for displaying the score during baseball games and exchange rates on the nightly news. In 1984, Kambe took on a problem of special significance in Nishiwaki. Textiles were often woven on looms controlled by planning programs; the programs, written on printed cards, looked like sheet music. A small mistake on a planning card could produce fabric with a wildly incorrect pattern. So Kambe developed SUPER TEX-SIM, a program that allowed textile manufacturers to simulate the design process, with interactive yarn and color editors. It sold poorly until 1985, a series of breaks led to a distribution deal with Mitsubishi's fabric division. Kambe formally incorporated as BRAIN Co., Ltd.

For twenty years, BRAIN took on projects that revolved, in various ways, around seeing. The company made a system for rendering kanji characters on personal computers, a tool that helped engineers design bridges, systems for onscreen graphics, and more textile simulators. Then, in 2007, BRAIN was approached by a restaurant chain that had decided to spin off a line of bakeries. Bread had always been an import in Japan—the Japanese word for it, "*pan*," comes from Portuguese—and the country's rich history of trade had left consumers with ecumenical tastes. Unlike French boulangeries, which might stake their reputations on a handful of staples, its

bakeries emphasized range. (In Japan, even Kit Kats come in more than three hundred flavors, including yogurt sake and cheesecake.) New kinds of baked goods were being invented all the time: the "carbonara," for instance, takes the Italian pasta dish and turns it into a kind of breakfast sandwich, with a piece of bacon, slathered in egg, cheese, and pepper, baked open-faced atop a roll; the "ham corn" pulls a similar trick, but uses a mixture of corn and mayo for its topping. Every kind of baked good was an opportunity for innovation.

Analysts at the new bakery venture conducted market research. They found that a bakery sold more the more varieties it offered; a bakery offering a hundred items sold almost twice as much as one selling thirty. They also discovered that "naked" pastries, sitting in open baskets, sold three times as well as pastries that were individually wrapped, because they appeared fresher. These two facts conspired to create a crisis: with hundreds of pastry types, but no wrappers—and, therefore, no bar codes—new cashiers had to spend months memorizing what each variety looked like, and its price. The checkout process was difficult and error-prone—the cashier would fumble at the register, handling each item individually—and also unsanitary and slow. Lines in pastry shops grew longer and longer. The restaurant chain turned to BRAIN for help. Could they automate the checkout process?

AlexNet was five years in the future; even if Kambe and his team could have photographed thousands of pastries, they couldn't have pulled a neural network off the shelf. Instead, the state of the art in computer vision involved piecing together a pipeline of algorithms, each charged with a specific task. Suppose that you wanted to build a pedestrian-recognition system. You'd start with an algorithm that massaged the brightness and colors in your image, so that you weren't stymied by someone's red shirt. Next, you might add algorithms that identified regions of interest, perhaps by noticing the zebra pattern of a crosswalk. Only then could you begin analyzing image "features" —patterns of gradients and contrasts that could help you pick out the distinctive curve of someone's shoulders, or the "A" made by a torso and legs. At each stage, you could choose from dozens if not hundreds of algorithms, and ways of combining them.

For the BRAIN team, progress was hard-won. They started by trying to get the cleanest picture possible. A document outlining the company's early R. & D. efforts contains a triptych of pastries:

a carbonara sandwich, a ham corn, and a "minced potato." This trio of lookalikes was one of the system's early nemeses: "As you see," the text below the photograph reads, "the bread is basically brown and round." The engineers confronted two categories of problem. The first they called "similarity among different kinds": a bacon pain d'épi, for instance—a sort of braided baguette with bacon inside—has a complicated knotted structure that makes it easy to mistake for sweet-potato bread. The second was "difference among same kinds": even a croissant came in many shapes and sizes, depending on how you baked it; a cream doughnut didn't look the same once its powdered sugar had melted.

In 2008, the financial crisis dried up BRAIN's other business. Kambe was alarmed to realize that he had bet his company, which was having to make layoffs, on the pastry project. The situation lent the team a kind of maniacal focus. The company developed ten BakeryScan prototypes in two years, with new image preprocessors and classifiers. They tried out different cameras and light bulbs. By combining and rewriting numberless algorithms, they managed to build a system with ninety-eight per cent accuracy across fifty varieties of bread. (At the office, they were nothing if not well fed.) But this was all under carefully controlled conditions. In a real bakery, the lighting changes constantly, and BRAIN's software had to work no matter the season or the time of day. Items would often be placed on the device haphazardly: two pastries that touched looked like one big pastry. A subsystem was developed to handle this scenario. Another subsystem, called "Magnet," was made to address the opposite problem of a pastry that had been accidentally ripped apart.

A major development was the introduction of a backlight—the forerunner of the glowing rectangle I'd noticed in the Ueno store. It helped eliminate shadows, including the ones cast by a doughnut into a doughnut hole. (One of BRAIN's patent applications explains how a pastry's "chromatic dispersion" can be analyzed "to permit definitive extraction of contour lines even where the pastry is of such hole-containing shape.") At one point, when it became clear that baking times were never consistent, Kambe's team made a study of the phenomenon. They came up with a mathematical model relating bakedness to color. In the end, they spent five years immersed in bread. By 2013, they had built a device that could take a picture of pastries sitting on a backlight, analyze their visual features, and distinguish a ham corn from a carbonara sandwich.

That year, BakeryScan launched as a real product. Today, it costs about twenty thousand dollars. Andersen Bakery, one of BRAIN's biggest customers, has deployed the system in hundreds of bakeries, including the one in Ueno station. The company says it's cut down on training time and has made the checkout process more hygienic. Employees are more relaxed and can talk to customers; lines have been virtually eliminated. At first, BakeryScan's performance wasn't perfect. But the BRAIN team included a feedback mechanism: when the system isn't confident, it draws a yellow or red contour around a pastry instead of a green one; it then asks the operator to choose from a small set of best guesses or to specify the item manually. In this way, BakeryScan learns. By the time I encountered it, it had achieved an even higher level of accuracy.

I first spoke to Kambe two summers ago, via Skype. It was early morning at BRAIN headquarters, in Nishiwaki, and he'd drawn beige blinds over the windows in a conference room. Kambe is seventy, and he was wearing a short-sleeved striped dress shirt. He has slightly gray hair, wire-frame glasses, and a laugh that comes easily. He seemed relatively relaxed for a founder of a technology company—just pleased with what he'd built so far and eager to build more. Today, BRAIN has twenty-six employees, nearly half of them software engineers. "Being a good company is more important than becoming a big company," he told me. BRAIN attracts talent, in part, because some young engineers, after leaving for university in Tokyo, must come back "to continue the house," as Kambe did. "There is no job in the countryside," he said. "Fortunately, there is BRAIN."

BakeryScan has been covered widely by the Japanese media; it's also familiar to many bakery patrons. In ads, the device appears as a cartoon figure with big googly eyes on top of its camera and gloved hands pointing happily at different pastries. In a recent entrance exam for high school in the Hyogo prefecture, an English-language reading-comprehension question took the form of a dialogue between two friends, Eric and Saori. "Last week, I saw an interesting machine at a bakery," Eric says. He describes the BakeryScan without naming it. "The system will be used more in many fields," he concludes.

In early 2017, a doctor at the Louis Pasteur Center for Medical Research, in Kyoto, saw a television segment about the BakeryScan. He realized that cancer cells, under a microscope,

looked kind of like bread. He contacted BRAIN, and the company agreed to begin developing a version of BakeryScan for pathologists. They had already built a framework for finding interesting features in images; they'd already built tools allowing human experts to give the program feedback. Now, instead of identifying powdered sugar or bacon, their system would take a microscope slide of a urinary cell and identify and measure its nucleus.

BRAIN began adapting BakeryScan to other domains and calling the core technology AI-Scan. AI-Scan algorithms have since been used to distinguish pills in hospitals, to count the number of people in an eighteenth-century ukiyo-e woodblock print, and to label the charms and amulets for sale in shrines. One company has used it to automatically detect incorrectly wired bolts in jet-engine parts. At the SPring-8 Angstrom Compact Free Electron Laser (SACLA), in Hyogo, a seven-hundred-metre-long experimental apparatus produces high-intensity laser pulses; since reading the millions of resulting pictures by hand would be impractical, a few scientists at the SACLA facility have started using algorithms from AI-Scan. Kambe said that he never imagined that BakeryScan's technology would be applied to projects like these.

In the spring of 2018, Kambe was invited to speak about the A.I. identification of cancer cells at a conference in Sapporo. The other speakers had degrees from Harvard and Stanford. "High-class people," he said. He felt out of place. But when he saw that they were all using deep-learning systems, he felt that he had something to contribute. At the conference, Kambe argued that there are tasks for which deep learning is still impractical. A few years earlier, BRAIN had tried replacing their hand-tuned system with a deep neural network. The new system managed to recognize pastries just as effectively, but the catch was the amount of data it required. A Japanese bakery, Kambe said, might introduce a new pastry variety every week, but the deep-learning system required thousands of training examples. Where would all the pictures come from? Show BakeryScan a pastry never seen on Earth, and it'll recognize the next one of its kind about forty per cent of the time; according to BRAIN, after just five examples, it is ninety per cent accurate, and after twenty it's nearly perfect. Moreover, whereas deep-learning systems are relatively inscrutable—you can't look at a neural network and say exactly why a decision emerged from it— BakeryScan's judgements, based as they are on a hand-engineered system, are more articulable. If the system misidentifies something, you can figure out why.

These days, it is unusual to develop A.I. in the way that BRAIN developed BakeryScan. The approach requires a mastery of fine details; it is in spirit artisanal. It takes years, during which parameters must be tuned and special cases accounted for. Deep learning relieves you from having to understand how the seasons affect the shadows in a doughnut hole; you merely plug in enough examples and the network figures it out. And, with deep learning, the same "brain" can accomplish different tasks when you feed it different data. DeepMind, the Alphabet subsidiary, used different data sets to train a single neural network to beat humans at chess, Shogi, and Go. Systems that depend on domain-specific knowledge, as BakeryScan does, need not just new data but new filters, new features, and new algorithms before they can be used elsewhere. Today, solving the pastry problem without deep learning would seem impossible; it's a wonder that, in 2007, when neural networks weren't a viable option, Kambe even took it on. The system that he and his team managed to build over the following fifteen years must surely be one of the more sophisticated achievements in "classical" computer vision—a fact obscured, perhaps, by its origin in baked goods.

Alex Krizhevsky, of AlexNet, told me he thought that a modern neural-network approach to the pastry problem would work just fine. Deep learning has come a long way, even since 2019, when I first entered the bakery in Ueno; training a pastry network might not require as many examples as I'd imagined. "In these neural nets," Krizhevsky said recently, over Zoom, "there's a lot of low-level features that need to be learned that are more or less common across images." Virtually all images have edges, for instance, and color gradients. A procedure called "transfer learning" takes a neural network trained on a vast data set and specializes it with a small supplement; a network pre-trained on the fundamentals might only need a dozen examples of each pastry variety to start telling them apart. "Look, even if it is the case that with deep learning it's difficult to get it to work with four examples per class, well then you get twenty examples per class," he said.

It's possible that Krizhevsky underestimates the pastry problem or overestimates the ability of deep learning to solve it. Deep learning works best on the often-seen; maybe there is room in the blind spots for something else. But it's equally possible that he's right. BakeryScan, and the immense effort that went into it, might have been an accident of timing: had Kambe started on the problem just a few years later, it might not have been so hard.

I suspect that, on some level, Kambe is glad to have taken the longer road. Last November, I reconvened with him over Zoom. Since we last spoke, BRAIN seemed to have invested heavily in the A/V setup at its Nishiwaki headquarters: Kambe was now wearing a three-piece suit and tie and standing at a podium in front of a wall checkered with "BRAIN" and "AI-Scan" logos. To his left was a large L.C.D. screen. When I asked a question, Kambe would dart through folders on the computer in front of him to pull up one of thousands of slides, or sometimes show me a live tech demo. There was someone operating the camera. Another assistant, who was helping to translate when necessary, was wearing a mask, but took it off after a few minutes; the coronavirus was on the rise again in Japan, but in the Nishiwaki area there hadn't been more than a handful of confirmed cases since March.

Kambe explained that the cancer-cell detector, now called Cyto-Aiscan, was being tested in two major hospitals in Kobe and Kyoto. It had become capable of "whole-slide" analysis: instead of analyzing a cell at a time, it could look at an entire microscope slide and identify the cells that might be cancerous. He pulled up an example, and the assistant handed him a retractable pointer. "This is Level IV," he said, referring to a sample from a stage-four cancer patient. The microscope slide looked like a tabletop speckled with black spots. Overlaid on the speckled areas were beige, yellow, and red rectangles. Kambe pointed out the red ones. "There is a cancer cell," he said. He clicked on one, and a zoomed-in image appeared. There were a few blobs, and singled out among them was the tiny cell deemed cancerous. Beside it, a set of sliders showed that cell's score on the most relevant visual features: the color tone of the nucleus, and its size and texture; the cell's overall roundness; its "center of gravity," based on where the genetic material was densest. With the sliders, Kambe said, it's easy for a doctor to judge which cells are cancerous. The system was apparently working at ninety-nine per cent accuracy. I asked Kambe how it worked—did it use deep learning? "Original way," he said. Then, with a huge smile, "Same as bread."

There had been other projects. The yarn-dyed textile industry was going the way of Kambe's father's lumber business, but BRAIN had adapted; it was working with a Toyota subsidiary on software for designing side airbags, which are actually woven in complex patterns, like textiles, with synthetic fibers. And BakeryScan, too, was evolving. Because of COVID-19, many pastry shops had violated one of their golden rules and begun wrapping items individually in cellophane.

This helped ease customers' fears, but the unpredictable reflections from the plastic threw off BakeryScan's algorithms. It turned out that, for this problem, they had a big data set. For years, they'd been collecting images from customers like me, as we checked out; they also had more recent pictures, in which the same pastries were covered in cellophane. They had trained a neural network to see through the plastic. Deep learning took the wrappers off; BakeryScan did the rest.

James Somers is a writer and a programmer based in New York.

More: **Artificial Intelligence**   **Technology**   **Computers**   **Japan**