

# PICTURE 3 IMPERFECT

How deep learning with less data can drive a huge leap in computer imaging and optical systems

**Our world is blanketed by cameras, every square metre recorded and catalogued in pixel upon endless pixel of data. We've got cameras on doorbells, cameras on car-backup screens, security cameras on street corners, GoPros on bike helmets, not to mention cameras on satellites and swallowable pill-cameras that take medical images of our internal organs.** Add in screenshots, video calls and smartphones — which record nearly two trillion photos worldwide per year by one estimate — and the volume of recorded images is overwhelming.

And yet, statistically speaking, almost every one of those pictures is garbage. The point of taking all those photos and videos is presumably to use them; however, our cameras, as good as they are, are far from perfect. Low light, bad focus, improper framing and blurring conspire to give us incomplete representations of the world we're trying to accurately capture with our constant clicking and streaming.

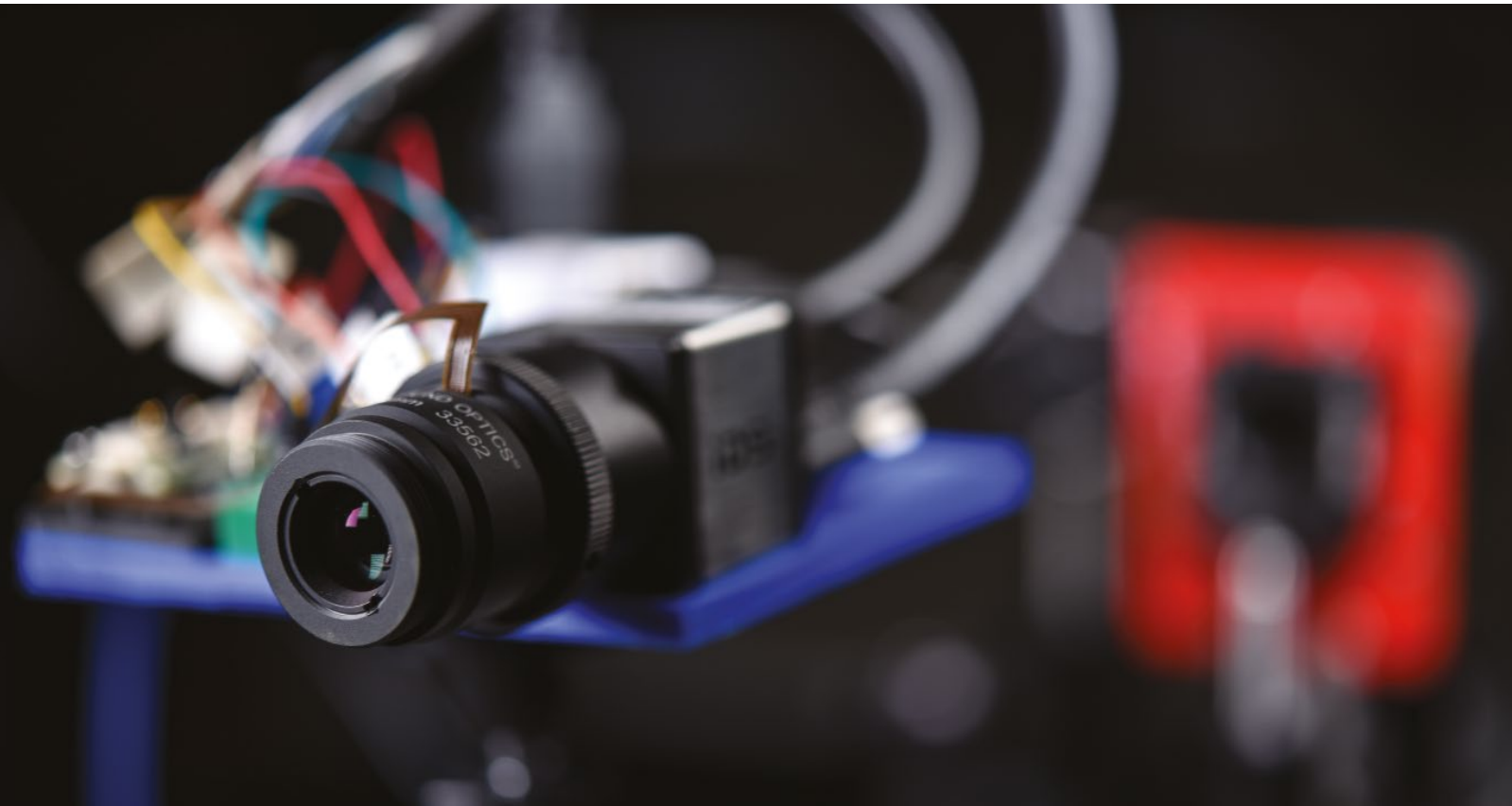
This is where Raja Giryes comes into the picture. An electrical engineering professor at Tel Aviv University, Giryes runs the Deep Learning Lab, which combines complex computing and cutting-edge camera design to make sense of all this photographic data. To put it simply — although there isn't much about this field that's simple — “deep learning” is the process by which computers digest and learn from massive amounts of data. On first pass, for example, a computer can't tell you whether it's looking at a picture of an elephant or an orca. But give the computer enough accurately labelled images, and soon it will identify patterns and reliably

recognize which animal is which just by looking at a few centimetres of a trunk or a bit of dorsal fin. That same principle can be used to create order and coherence out of the untold billions of photographs taken every day.

“One of the main problems in artificial intelligence is that you need lots of data to train neural networks,” says Giryes, who was an Azrieli Graduate Studies Fellow between 2010 and 2013 while working toward his PhD in computer science at Technion–Israel Institute of Technology. Neural networks are the computerized algorithms that mimic how the human brain processes information. “What we're trying to overcome,” says Giryes, “is how to learn with less data, how to be efficient, how to adapt your data to apply to new domains and new problems.” In essence, Giryes and his team are helping computers get smarter quickly by reducing the amount of data necessary for training.

The human eye and brain are constantly assimilating data, unconsciously learning about the world. Nobody taught you that trees stand still, but by the time you were a toddler you had seen enough to know that you can lean against one and not fall over. For computers, acquiring even this kind of rudimentary knowledge — what is a tree, for instance, and is swaying in the wind the same as moving? — requires massive data. Providing that information for every category and situation means an enormous workload for the humans tasked with uploading and correctly tagging countless images.

By Zac Unger  
Photographs by Boaz Perlstein



Raja Giryes (previous pages) and his lab use deep learning to design optical elements that improve imaging, such as the “phase mask” that allows this camera to reconstruct depth information and change focus while acquiring images, which can be used to remove motion blur from photos.

**‘Our work means that you don’t need to train a neural network for each new camera lens from scratch. You only need to model the blurriness of the lens and then modify the already trained neural network inference steps and get great results.’**

Giryes’s work is broadly applicable to all manner of circumstances where datasets are sketchy or incomplete, solving problems not only for still images but also for video and even medical systems that determine which human cells are cancerous and which are healthy. In one side project, he and his colleagues trained a neural network to accurately assign archaeological artifacts to the correct location and period. Using a publicly available photo database from the Israel Antiquities Authority, the AI was fed a massive amount of information on everything from the Lower Paleolithic period (which started about 1.4 million years ago) to the Late Islamic period (14th century). Using a neural network, the AI extracted features known as “embeddings” from the images of the artifacts. A computer then translated the colours, shapes and other aspects of the artifacts’ appearance into a series of numbers that can be compared from artifact to artifact, eliminating the subjectivity that human brains may be prone to when analyzing a piece of pottery.

Archaeology is a particularly difficult field because two temporally or spatially adjacent societies might have tools or weapons that look remarkably similar. In tests, Giryes’s algorithm handily beat two trained archaeologists, correctly identifying artifacts with much greater accuracy. This technique is valuable in illuminating new avenues of study for finding meaningful relationships between ancient sites — and much more.



This camera system uses a beam splitter to capture the same scene with two different sensors. Because the same data can be collected using a high-resolution and low-resolution sensor at the same time, researchers can use this information to learn how to increase the resolution of the latter.

“Getting lots of examples requires lots of effort,” says Giryes. “The question for us is do we need to collect all the data every time, or can we develop a technique that would achieve the same result in a faster way?”

Although Giryes’s work with large datasets and deep learning has a wide range of applications, his main focus is computational imaging and designing new optical systems using artificial intelligence. The practical outcomes of this work include everything from improving picture quality to image recognition, training a computer to identify real-world images from tiny fragments.

Representation of three-dimensional data presents a particularly thorny problem that his lab is tackling. “You take a photo with your phone,” Giryes says, “and you want it to be a 3D photo, but you’re only using a single camera. We’re figuring out how to design the optics of the camera to get better depth estimation.” The easiest way to make a three-dimensional image is by combining regular images from different angles. But computers building models of lifelike objects don’t always have the luxury of data inputs from all directions. Instead, Giryes developed an AI technology that teaches a camera (by automatically designing its optical element) how to get 3D information from a single direction. In addition to

## EDITING IMPLICIT REPRESENTATIONS

1

Using neural networks, computers transform images made up of 2D pixels into 3D representations composed of “voxels.”



2

These representations show what a common object such as a chair looks like not just head-on but from all angles.



3

Traditionally, to see how an object responds to digital manipulation, you need to scan the space and check all points.



4

By disentangling component parts into geometric data, you can transform, interpolate and combine shape segments.



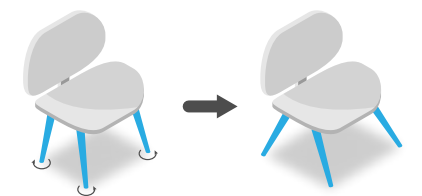
5

This process can allow researchers to separate and mix the individual parts of objects and generate new types of shapes.



6

This technique holds promise for designers trying to create new products and test how they might perform under different stressors.



capturing 3D data, Giryes also studies how to better represent it. His team uses “implicit representation” to accurately depict or manipulate objects with realistic depth.

We’re all familiar with pixels, the smallest element in a digital display. An alternative approach uses “voxels,” the 3D equivalent. Think of a three-dimensional picture of a stack of blocks; each voxel is one block. (Etymology is helpful here: the word pixel derives from “picture” plus “element,” whereas voxel comes from “volume” plus “element.”) Using neural networks, computers assemble these voxels into images of, say, an airplane and have an understanding not just of what an airplane would look like head-on but from all sides. Although voxels are useful, they are of limited resolution, says Giryes, “but with implicit representation, you can simulate a function with any resolution you want.”

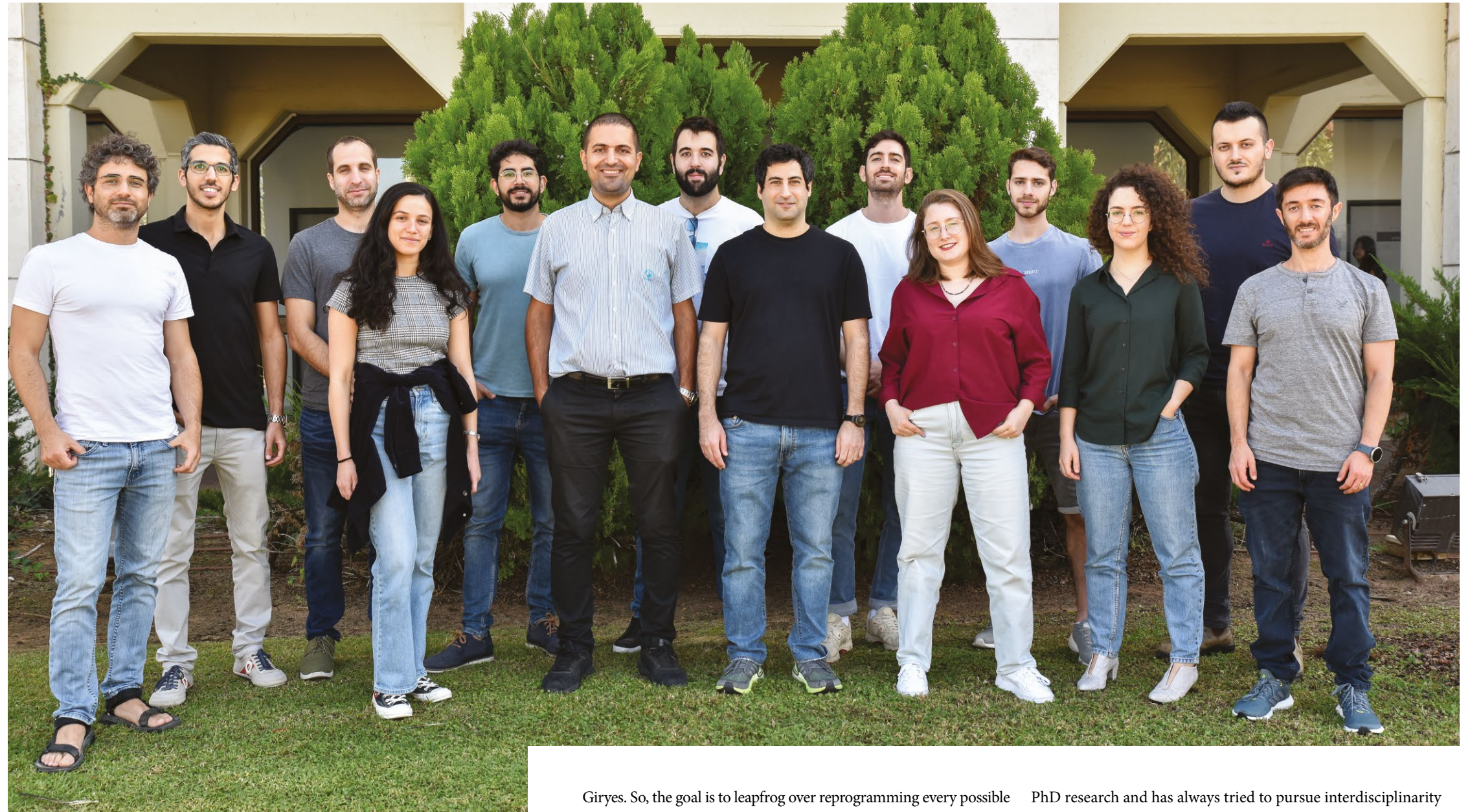
In the paper he wrote with colleagues on the subject, the technique is described as allowing for “manipulation of implicit shapes by means of transforming, interpolating and combining shape segments together without requiring explicit part supervision.” In lay terms, “imagine that you have two chairs,” says Giryes. “I can take the back of one chair and the legs of another and edit them together, mix up the parts and generate new types of shapes you’ve never seen before.”

The neural network learns to separate out one piece of the model from another, modifying individual parts without disrupting the rest of the image. By taking a three-dimensional representation of a common object and disentangling its component parts into essential geometric data, the model allows users to more easily edit these implicit shapes in high resolution. An image of a chair then ceases to be a chair, but is reduced to its most basic shape segments, which can be reassembled in multiple configurations without requiring direct supervisions of each discrete part. This technique holds promise for designers and engineers trying to imagine new products and test detailed representations of how they might perform under different stressors.

Shady Abu-Hussein, one of Giryes’s PhD students, describes how the research team has also worked to combat low resolution in images, a significant factor that reduces the usefulness of many pictures. The goal is to create a “super resolution network” that can decode the fuzz and automatically enhance the quality of what lies beneath. “These networks are usually trained for a specific lens structure,” says Abu-Hussein. “Our work means that you don’t need to train a neural network for each new camera lens from scratch. You only need to model the blurriness of the lens and then modify the already trained neural network inference steps and get great results.”

One of the most exciting real-world applications for Giryes’s brand of deep learning is in the field of self-driving cars, which must use computer vision to assess the world around them with critical levels of accuracy. Giryes consults with the Israeli company Innoviz to develop guidance and obstacle-detection systems for autonomous vehicles. Innoviz — which recently signed a \$4 billion USD deal with Volkswagen — develops LIDAR (laser imaging, detection and ranging), a technology that employs laser-light rebounds to measure distance and create 3D images of objects ahead. But even the best computerized eyes are only as good as the knowledge base they use to process their inputs.

“It costs enormous amounts of money and time to collect 10,000 hours of driving data and then have a person tell you what’s in each frame,” says



Giryes (sixth from left) and his large research team combine complex computing and cutting-edge camera design to make sense of the world’s deluge of photographic data.

**Giryes’s work is broadly applicable to all manner of circumstances where datasets are sketchy or incomplete, solving problems not only for still images but also for video and even medical systems that determine which human cells are cancerous and which are healthy.**

Giryes. So, the goal is to leapfrog over reprogramming every possible situation a car might encounter and get it to learn from similar experiences. For example, a human driver has probably never seen a five-metre cube painted gold with blue polka dots, but we’d still automatically know to swerve if it fell off the truck in front of us. Getting autonomous systems to extrapolate from the data they already have with minimal or no adaptation is the ultimate goal.

“Making decisions using multiple sensors such as both cameras and LIDAR is important as we need to learn how to use data to be adaptive to different environments,” Giryes says. “We have a car that works well in the day, but we also need it to drive at night. It has to work as well in the city as it does in the country. It has to drive well in Europe and also drive well in India.”

In both academia and industry, Israel has been a crucial hub in the development of deep learning techniques for processing all manner of inputs. The country is home not only to dozens of successful high-tech start-ups but also Tel Aviv University and Technion’s groundbreaking electrical engineering and computer science research groups. Giryes was in one of the latter labs while conducting his

PhD research and has always tried to pursue interdisciplinarity in his work, a collaborative attitude he now extends to the many graduate and postdoctoral students he supervises.

“Raja isn’t just a great mentor and teacher,” says Abu-Hussein, “but he also likes to help out his students with the smallest details, even down to looking at the lines of code, which a lot of professors would just assign another more experienced student to do.”

That drive for interdisciplinarity carries over into Giryes’s work, as he combines deep learning techniques with advances in optical hardware. “Right now, you have engineers who design optics and engineers who design sensors, and then you have the people who program the post-processing algorithms that give you better results,” he says. “But the ultimate goal is to use both deep learning and new optics to design a totally new camera that is innovative and revolutionary.”

With so much of our lives recorded on screen, we’re going to need exactly that kind of leading-edge thinking to make the images do justice to the magnificent complexity of the real world surrounding us. ▲●■