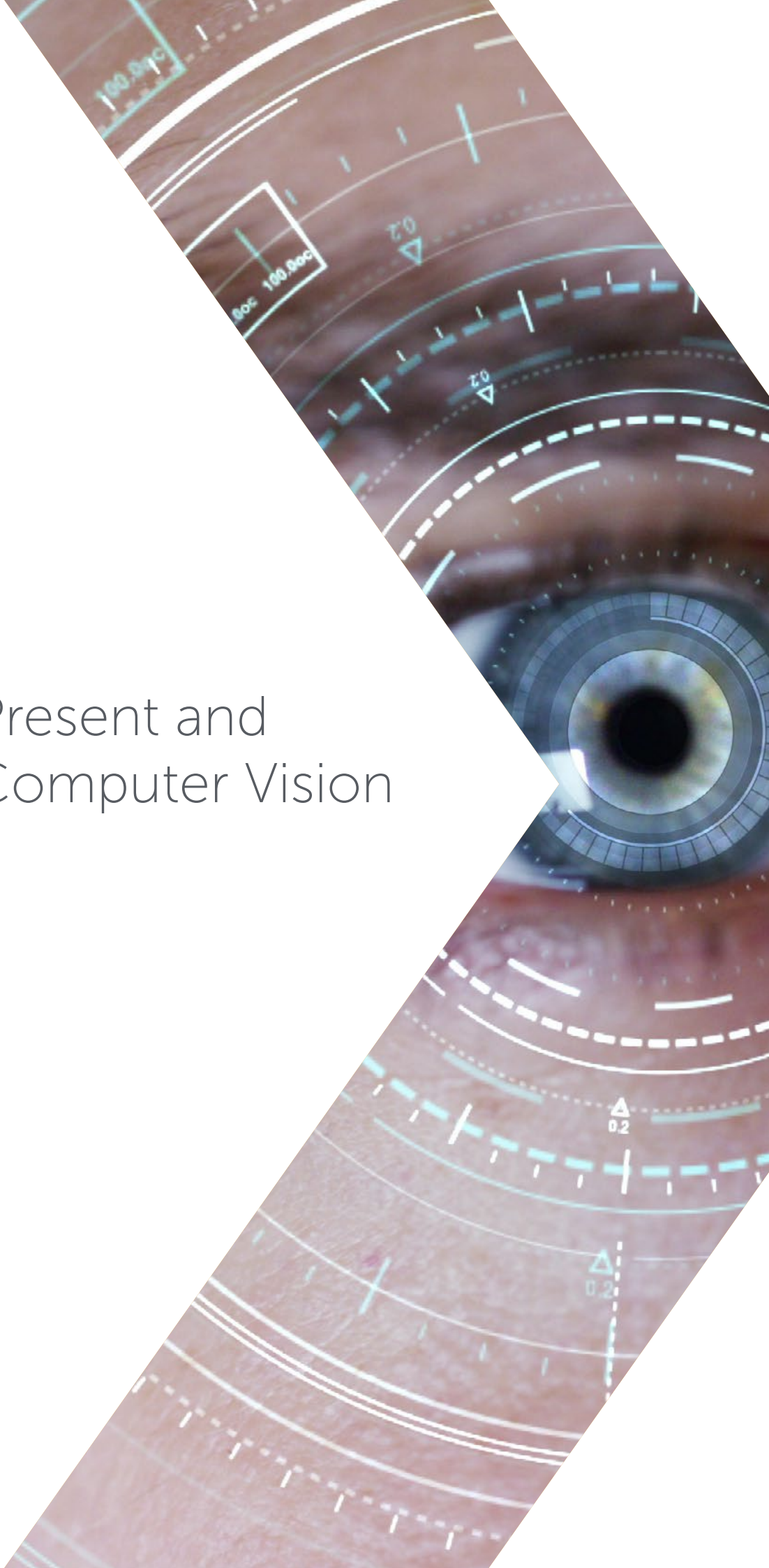
trax

WHITEPAPER

The Past, Present and Future of Computer Vision



Introduction

One of the most widely known tropes of science fiction is that of a civilization being unable to make sense of a futuristic technology, captured in this adage from sci-fi author Arthur C. Clarke:

"Any sufficiently advanced technology is indistinguishable from magic."

If one could travel back in time even a few decades to show people some of today's technologies, chances are that they would confuse it for magic. Self-driving cars, robots that move heavy objects, in-game video analyzes during football matches, drones that can analyze crop quality and soil suitability in agriculture and cameras that see what is happening on retail store shelves – these are all quite commonplace today.

Today's consumers are becoming more familiar with technology concepts like artificial intelligence (AI), machine learning (ML) and augmented reality (AR) that are at work here. But at the heart of some of these applications is a technology that's not talked about as much – Computer Vision.

Computer Vision is an interdisciplinary field of science that enables computers or other machines to see, identify and process images and videos much like the human eye does. It involves a number of tasks including signal processing, image enhancement, object detection and classification, motion analysis and 3D image reconstruction.

This is the story of how the field of computer vision has evolved, starting from student projects in research labs in the 1950s to the explosion of Convolutional Neural Networks (CNN) this decade and the frenzy of activity happening today around putting it to use to solve everyday problems.

Chapter 1: Human and Computer Vision

Many of the technological advancements in history have been the result of humans drawing inspiration from nature. The shape of aeroplanes bears a resemblance to the shape and wing span of birds while the invention of velcro was based on how burdock burrs stick to clothes. Even the Shinkansen bullet train in Japan has its nose inspired by the beak-shaped nose of the kingfisher. Computer vision also derives its inspiration from how the human brain sees the outside world to make meaning out of the visual input received.

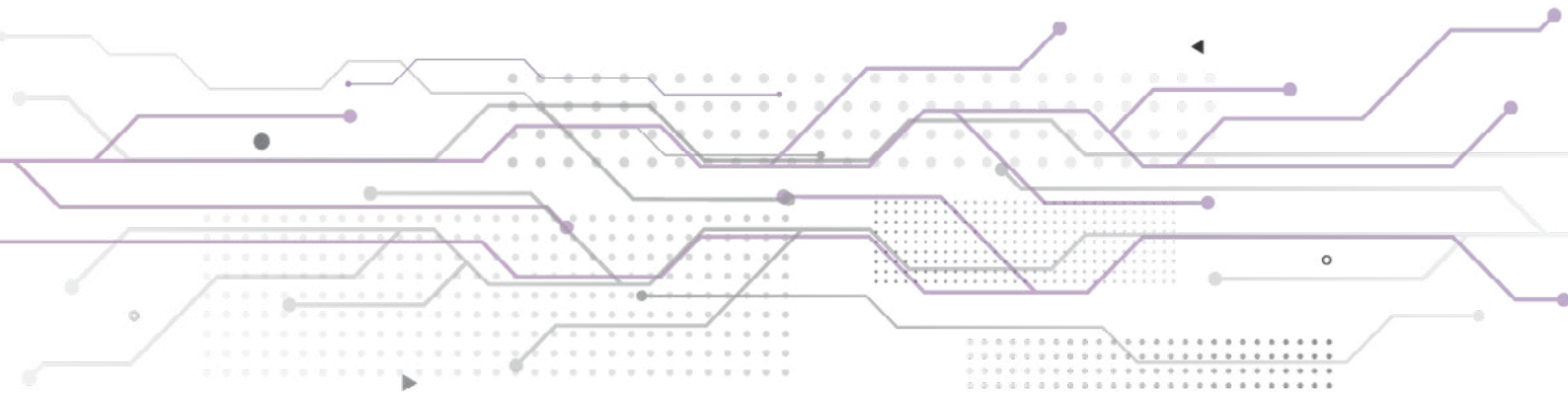
INSPIRATION FROM THE HUMAN EYE

The human eye is one of the most complex systems in the body. Light from the outside world enters the eyes through the transparent covering called the cornea. The cornea bends the light rays passing through the round hole of the pupil. The iris, which is the coloured part of the eye surrounding the pupil, opens and closes to adjust the size of the pupil and regulate the amount of light that passes through.

The light rays then pass through the lens and fall on the retina, which contains millions of light-sensitive nerve cells, called rods and cones. Cones, located at the centre of the retina are responsible for providing clear, sharp central vision and detecting fine colours and details. Rods, on the other hand, are located in the outer edges of the retina and help with peripheral or side vision and in detecting motion in dim light. Together, these cells help convert the light into electrical signals which are sent to the brain for processing the final image we see.

At a high-level, computer vision systems also go through a similar process. A camera or a sensor captures light from the external world, converts the light photons into signals, and a transfer mechanism carries the signal to a place for processing and eventual interpretation.

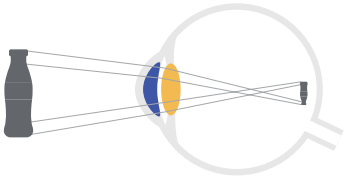
Both the human eye and computer vision systems also rely on comparison with reference data to process the signal. The human eye compares the visual input with what it has seen in memory and through experience, while the computer accesses a set of reference images from a database to verify and learn. The similarities however end there.



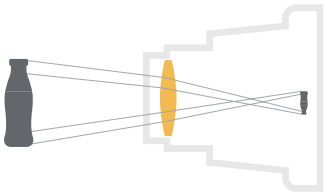
THE HUMAN EYE AND COMPUTER VISION

Since the field of computer vision draws inspiration from the workings of the human eye, both share some high-level similarities.

Human Vision | **Computer Vision**



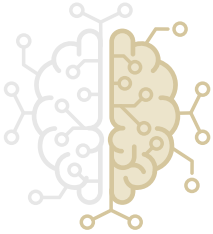
1. Capturing external input



Light rays pass through the cornea, the pupil and the eye lens and fall on the retina.

The machine takes visual input through a camera or sensor, which may use lenses to filter or expand the amount of light

2. Processing input in a "brain"



Nerve cells on the retina convert light into electrical signals which are then carried to the brain for processing

Light photons converted to signals are transferred to an "engine" for processing

3. Interpretation using reference data

The brain takes the signals from the optic nerve and processes the information to produce the image that we see, with help from memory

The computer vision system compares the input with a large data set of reference images for verification

HOW COMPUTER VISION DIFFERS FROM HUMAN VISION

Unlike machines, the human eye suffers from certain innate biases, formed as a result of mechanisms used by humans to survive among predators over time. For example, the eye tends to see faces in non-human objects, a phenomenon called paraidolia. This could be a result of centuries of pattern-seeking tendencies of our ancestors - imagining faces of dangerous creatures on a tree or assuming the rustling of grass to be due to footsteps of an approaching predator rather than the wind.

Another facet of the human eye is that its field of view is restricted to 200 degrees, and it also does not detect colour uniformly within this range. We tend to see colour better at the centre of the field while peripheral vision is used to detect low light objects and movement.

On the other hand, since machines process all the light captured as an array of numbers, computer vision does not suffer from these biases and inefficiencies, while 360-degree camera systems enable capture of the entire field of view.

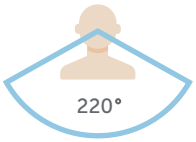
Despite similarities, there are key differences between the eye and how computer vision systems operate, some of which arguably make the latter more accurate.

Human Vision | **Computer Vision**

The human brain suffers from biases owing to our long history of evolution

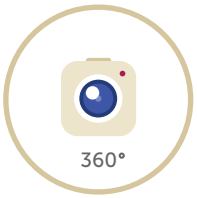
1. Biases

Machines don't have any innate biases

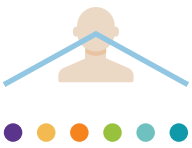


The human eye can only see within a 220-degree field

2. Field of view



With the help of multiple cameras and even 360-degree cameras, machines can cover the entire field of view



The eye sees colour better at the centre of the field

3. Non-uniform colour detection

Machines see colour uniformly across the field of view

THE COMPLEXITY OF TEACHING COMPUTERS TO SEE

That's not to say that computer vision systems are fool proof. Much like the optical illusions that befuddle the human brain, computer vision systems can also be tricked using what are called "adversarial images". These are patterns and pictures that exploit weaknesses in computer vision algorithms to fool them into mistaking a panda for a gibbon, or a cat for a guacamole. In fact, a team of students at MIT published a study in 2017 which showed how they could fool a system into wrongly classifying a photo of a 3D-printed turtle as a rifle!

Malicious actors could use this to cause harm, like manipulating face recognition tools into recognising the wrong people, or to attack the computer vision systems that enable self-driving cars. For example, a small patch on the side of the motorway could make a self-driving car think that it is looking at a stop sign.

The root of all the complexity in making computer vision work lies in the fact that we don't even fully understand how the human eye works. We know the building blocks of how the eye captures light and how neurons fire to perform calculations based on input. But there are so many interconnected neurons, and making sense of this network and understanding what's going on and why is a huge challenge.

The process by which the brain converts a 2D image to build the correct 3D model is also a mystery. This is a problem which has no clear solutions, often referred to as an "ill-defined problem" in academic science.

All of this makes it hard to understand computer vision as well, so much so that researchers and practitioners often resort to intuitions to overcome some problems.

Research at Trax:

Dolev Pomeranz

Chief Architect & Head of Research

A long-standing interest in software engineering took Dolev to the Ben-Gurion University in Israel. In particular, he always found problems being tackled by computer vision intriguing, and once he signed up for a course on the topic, he fell in love with it and hasn't looked back.

Talking about the complexity involved in getting a computer to accomplish a task like detecting new faces, he says, "The human brain hides so much calculation as it helps us detect, understand and recognise a face". But for computers the task is even more complex. When you understand what an image is – essentially a big array of numbers – how do you go about making a computer understand that there is a face in it?". It was this monumental challenge of going into the unknown that pulled Dolev deeper into the field of computer vision.

During his Masters Degree, Dolev conducted a research project that saw him teach computers to solve jigsaw puzzles. Though the problem had been studied progressively since the 1960s, Dolev's team was the first to have devised a computational solution that was fully automatic and without any manual hints.

Expert Q&A: Is Computer Vision a mystery?
Dolev Pomeranz, Chief Architect & Head of Research at Trax explains the underlying challenge of trying to get computers to do what the human eye does

How similar is computer vision to the workings of human vision?
I think the similarities are not something everyone agrees upon. There is still some debate about whether computer vision is biologically inspired or not. A good analogy is comparing the wings of a bird with that of plane. But the bird's wings are non-rigid, have muscles and move about while there's rigidity in the plane's wings. So you could say computer vision draws some inspiration from the human eye but when you look at the details there are many differences.

How does the complexity in computer vision systems compare to that of human vision?
In some sense, computer vision algorithms are much simpler than human vision. For example, computers can be tricked easily using simple techniques which don't fool the human eye. So the human vision is system is still better in that sense.

But on the other hand in some tasks, we have achieved superhuman levels of vision with computers. For example, in the famous ImageNet challenge, you feed a system thousands of classes of objects like "container ship", "mite", "mushroom" or "cherry", and the computer has to classify images into each of these classes. And what we have seen is that the accuracy of the best candidates in the competition has improved drastically – from around 74 percent in 2012 to 95 percent in 2017. This is considered in the field to be as good as human level. In simple terms, this means that computers are getting better than even humans at classifying objects correctly like this.

Why do some people, even researchers within the field, say that no one really understands computer vision?
This goes back to the same problem of our lack of complete understanding of how human vision works. We understand the basic building blocks – how a neuron works, that it gets inputs and performs calculations. But when you have so many neurons connected, then understanding the network generated with such units becomes hard.

And with machines it's still complex enough. Because of this, sometimes it is perceived as "dark magic". People build intuitions on what should be done to overcome some problems. It's not like you can guarantee solutions all the time and even when you try, you might not understand why a technique doesn't work.

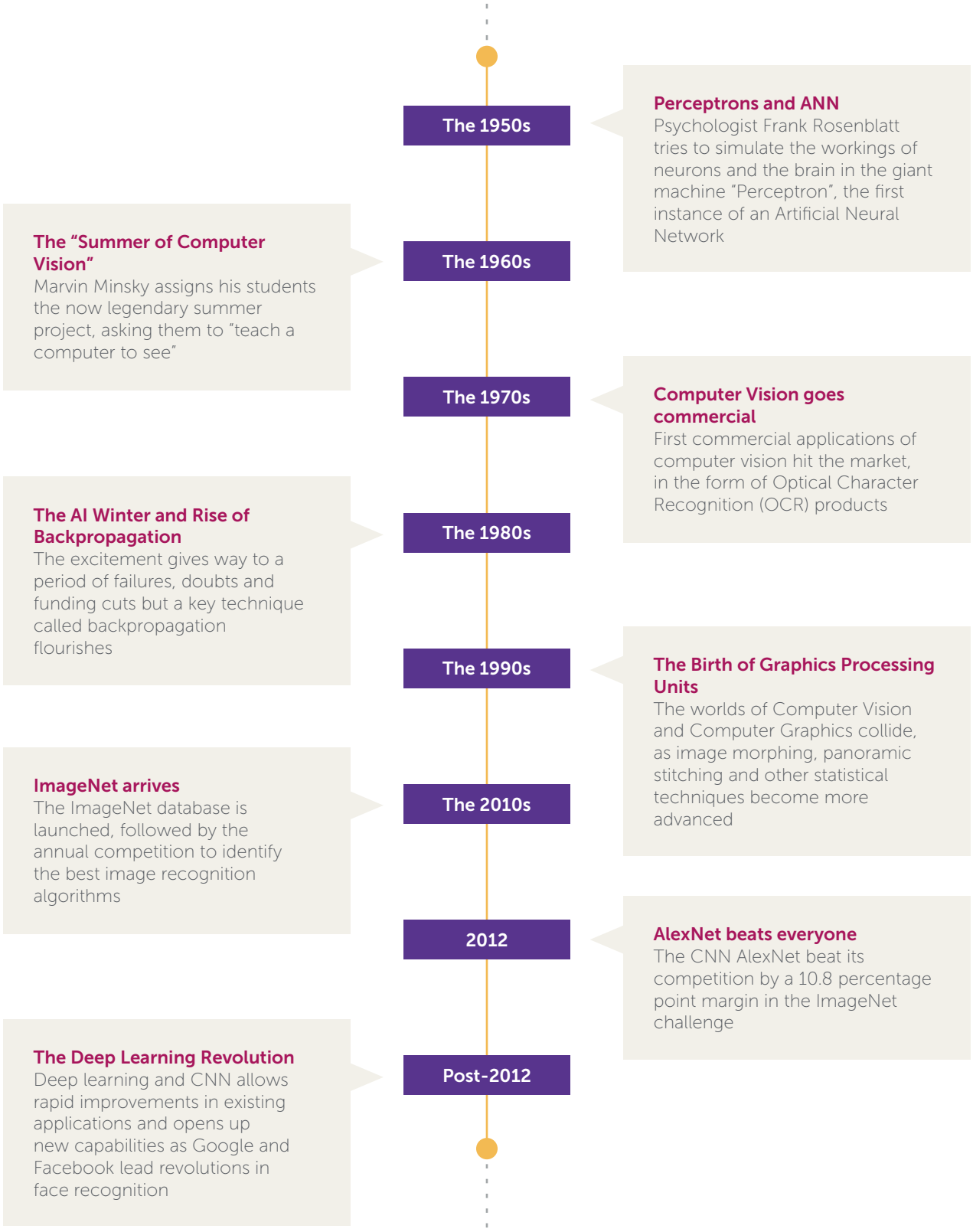
There are some best practices that help practitioners and even some work being done in the theoretical domain of neural networks to try to explain the high-level principles and limitations that these complex structures adhere to. But unlike a classical analytical function it isn't so easy to use mathematical tools to solve problems. So while researchers are seeing some success in explaining, it is still a relatively new field and there's much more work to be done.

Chapter 2: A Brief History of Computer Vision

Due to its interdisciplinary nature, the history of computer vision is closely tied to the related fields of AI, ML and robotics. Although computers were being trained to recognize documents, photomicrographs and high-altitude views of the earth's surface in the early 1950s, experts agree that it was the invention of perceptrons that kicked the field into top gear.

A BRIEF HISTORY OF COMPUTER VISION

Computer vision as a field has evolved over time, through contributions from numerous scientists, building on their predecessors' works. Here are some of the key milestones in this ongoing journey.



PERCEPTRONS AND THE BIRTH OF ARTIFICIAL NEURAL NETWORKS

In the 1950s, biologists were trying to understand how signals passing between neurons gave rise to intelligence and learning. Psychologist Frank Rosenblatt tried simulating this process electronically in software and hardware using thick wires attached to giant machines. His design "Perceptron", was one of the earliest instances of an "artificial neural network" and aimed to learn how to sort simple images into triangles and squares.

The device had 8 simulated neurons and dials connected to 400 light detectors. Upon receiving light signals, the neutrons would describe what it saw in binary digits of "1" or "0". While this failed at the beginning, Rosenblatt was able to train the system to correctly call out the shapes using a method called supervised learning. In this method, he trained the perceptron by showing an image along with the correct answer and the machine would tweak how much each neuron "weighted" the incoming signals to eventually arrive at the correct description.

THE LEGEND OF MARVIN MINSKY AND THE SUMMER OF COMPUTER VISION

A seminal milestone in the evolution of computer vision occurred in the late 1960s through the work of the pioneer Marvin Minksy. Having founded the Artificial Intelligence Lab at MIT in 1959, he assigned his students a summer project in 1966. The task was to link a camera to a computer and get it to describe what it saw.

Although no one managed to solve the problem completely, some of the work done in the process set the field in motion. For instance, it was soon possible to understand the edges of an object within an image. By the 1970s, the first commercial applications of computer vision hit the market, in the form of OCR products that enabled computers to read printed text.

Pioneers of Computer Vision:
Frank Rosenblatt

Frank Rosenblatt was one of the earliest influential figures in the field of computer vision. He was originally trained in psychology and cognitive science. Enamoured by how learning emerged from the firing of neurons in the human brain, Rosenblatt wanted to see if he could get machines to do the same. He had the following proposition in mind: "I want to construct an electronic or electromechanical system which would learn to recognize similarities or identities between patterns of optical, electrical, or tonal information, in a manner which may be closely analogous to the perceptual processes of a biological brain."

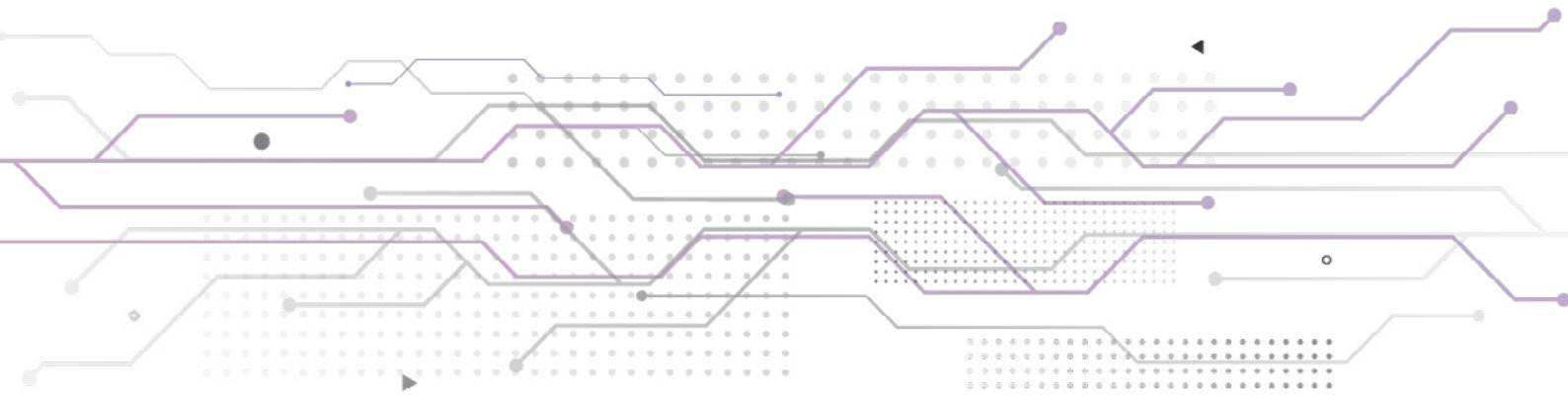
Rosenblatt's efforts resulted in the Perceptron, a machine that would become the precursor for today's artificial neural networks.

Pioneers of Computer Vision:
Marvin Minsky

Perhaps no name is more renowned in the field of AI and computer vision than that of Marvin Minsky. Legend has it that during an undergraduate course in 1966, he gave his students a summer project with this simple instruction that was way ahead of the time: "Spend the summer linking a camera to a computer. Then get the computer to describe what it sees."

Never mind a summer – half a century later, we're still working on perfecting this.

However, the work done during the summer did help progress the field of computer vision, leading to the first commercial applications in OCR products.



THE AI WINTER AND THE RISE OF BACKPROPAGATION

It is common for breakthrough technologies to go through a phase of intense research, hype and excitement, followed by a period of failed attempts, pessimism and funding cuts. The field of AI too experienced such a lull in the 1970s, a period termed "The AI Winter". It was sparked by a few failed attempts at machine translation from Russian to English during the Cold War, a critique of the concept of perceptrons by Marvin Minsky and funding cuts for AI research in the US and UK.

One topic of research called "backpropagation" however survived this winter. Backpropagation is shorthand for "backward propagation" and is a technique that helps in the calculation of the weights used in a neural network's layers. In the context of learning, it figures out how far off the final answer or output was from the input and sends feedback about this error to the previous layers one by one, like a game of telephone. Using this feedback loop, the network is able to identify subsequent images that go through it slightly better every time.

Pioneers of Computer Vision:

Yann LeCun

Yann LeCun is a leading figure in AI and computer vision research today and has seen the field grow and evolve from the 1980s. Surprised at the abandonment of Rosenblatt's perceptron theory, he teamed up with a group of researchers to reinvigorate the idea of neural networks with multiple layers. In 1989 he gave the first practical demonstration of backpropagation at Bell Labs and developed a system that could read handwritten checks and zip codes.

LeCun is currently the Chief AI Scientist at Facebook and an influential voice on the state of research, applications and ongoing trends in AI.

STATISTICAL TECHNIQUES AND COMPUTER GRAPHICS

The focus during the subsequent decades was on developing better mathematical techniques to analyze scenes and images. The artificial neural networks became more sophisticated, as more "layers" were added in the process of interpreting the image.

During the 1990s, there was considerable interaction between the fields of computer graphics and computer vision. The task of manipulating real-world imagery to create animations for video games and films became more sophisticated with image morphing. This was the decade during which statistical techniques were first used to recognize faces in images.

Panoramic image stitching, where multiple images are combined with overlapping fields of view to create one single image, also became more feasible. The growth of the internet and the access to cheap computing power during this period were major factors that contributed massively to the growth of the field.

The arrival of the graphical processing unit (GPU) in the 90s video game industry proved to a be a vital catalyst. Unlike the central processing unit (CPUs), which were multitasking brains inside computers, GPUs were dedicated units focussing on specific, repetitive tasks like generating complex shapes to create 3D graphics in gaming environments.

GPUs are responsible firstly for converting data in video game environments like locations of objects, angle relative to other objects, colour and texture into pixels on a screen. Secondly, every time the screen refreshes or the scene changes within the game, GPUs have to perform a multitude of calculations in rapid quick time to generate the next set of pixels. This very processing power is what comes in handy to feed neural networks thousands of images to learn from and identify patterns for improving its accuracy.

THE BIRTH OF IMAGENET AND THE DEEP LEARNING EXPLOSION

As GPUs made neural networks work faster in the 2000s, it became easier to get machines to learn from large labelled data sets with multiple layers transmitting information to the next. But there was still the challenge of building a large enough dataset that reflected the real world.

Enter Fei-Fei Li, a rising computer science professor at the University of Illinois Urbana-Champaign. Li realized that existing datasets didn't capture how complex and variable the real world was. For example, an algorithm that referenced just five pictures of cats, would only work off five camera angles, lighting conditions, and maybe a small variety of cat breed. But if it could tap into 500 pictures of cats, there would be many more examples to draw patterns and commonalities from. "We're going to map out the entire world of objects," Li said at the time. Building on an existing idea of associating an image with a word, she imagined a much larger dataset called ImageNet, with many images associated with a word.

What started out as a research poster at a conference soon evolved into an annual competition to see which algorithms could identify objects in the dataset's images with the lowest error rate. In just two years after launching the competition in 2012, ImageNet become a benchmark for how well image classification algorithms fared against the most complex visual dataset assembled at the time.

Perhaps the single biggest event that accelerated the field of computer vision occurred in 2012 when the deep CNN called AlexNet beat its competition by a 10.8 percentage point margin in the ImageNet challenge. This development paved the way for rapid improvement in every application of computer vision. For example, Facebook's face recognition capabilities became 97.35 percent accurate, growing a whopping 27 percent more than what was previously possible and self-driving cars began to use CNN to detect objects.

Expert Q&A: How Deep Learning Revolutionized Computer Vision

Ziv Mhabary, Vice President of Computer Vision Algorithms at Trax talks about how the arrival of deep learning completely revolutionized the field of computer vision.

Why has the phrase "Computer Vision" not caught on in popular media compared to Artificial Intelligence?
AI refers to the ability of machines to process and understand data of any kind while computer vision deals primarily with visual data. You can think of computer vision as one of many components of AI based products, for example autonomous cars. Popular media hasn't always focused on the differences between the two, and has instead adopted AI as a catch-all term.

It's worth mentioning that while AI's popularity is relatively recent, computer vision has long fueled a host of day to day applications, from your digital camera, in MRIs for medicine to many more.

Research at Trax:

Ziv Mhabary
VP, Computer Vision Algorithms

"What I like most about computer vision is that you can actually see the results of the algorithms that you develop", says Ziv. Having been attracted to the topic of computer vision from a young age, Ziv went on to do his fulfil his dream of studying the subject at the Ben-Gurion University of the Negev in Israel, completing a Masters degree and a PhD.

Ziv was drawn to the immense learning opportunities in this rapidly emerging area of research. "No one knows the answer to every problem in computer vision. There is always more to explore and learn."

One of Ziv's academic projects involved using drones to capture images of agricultural crops in Israel. He employed computer vision algorithms to detect, analyze and monitor nitrogen content in the land, uncovering opportunities for more efficient agricultural practices.

The technology breakthrough around neural network (a.k.a deep learning) in 2012 dramatically increased the capabilities of all AI related fields in general and the computer vision one specifically. While there will be more and more AI based products in the market, you will likely find computer vision and\or natural language processing technologies in most of them.

What are some things we can do with computer vision today that were not possible ten years ago?
As I mentioned, there was a deep learning revolution at the end of 2012, where research on CNN unlocked totally new levels of capabilities in the field of computer vision. Many applications were stuck before deep learning, with only very small improvements in accuracy like 0.3 percent or so every year. But deep learning pushed them forward and the field experienced a very big leap forward.

For example, in image recognition, the traditional benchmarks became easy to achieve and new benchmarks emerged in the market. Everything that we did before was replaced with deep learning.

It also made some applications of computer vision a commodity. Recognizing an object on your mobile phone is no longer something that only the big companies can do. Everyone can take open source code, public data sets and train a system quite easily. These can provide you with a very reasonable level of accuracy in object recognition.

Chapter 3: The State of Computer Vision Today

Computer Vision has come a long way from categorizing images into triangles and squares to recognizing faces and adding AR filters like those seen on popular apps like SnapChat.

But if you think all the painstaking research done by pioneers over decades is just being used for selfie masks, think again. Today, computer vision finds applications in nearly every major industry from security and defence to transportation, agriculture and retail. These are some of the ways in which everyday experiences are being disrupted through computer vision-powered solutions.

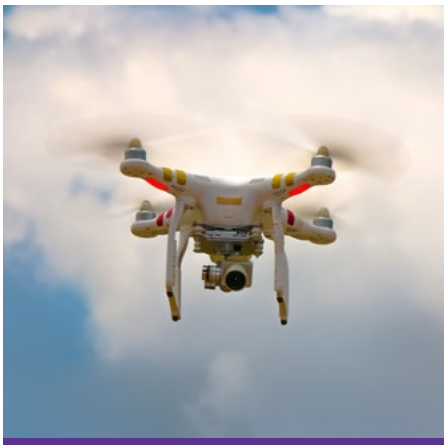
SHARPENING MILITARY AND DEFENCE STRATEGIES

Some of humanity's greatest technological breakthroughs have emerged from applications primarily motivated by war and defence. Even the Moon landing was a result of escalating space race between the US and the Soviet Union during the Cold War.

In recent decades, the Defense Advanced Research Project Agency (DARPA) has been one of the foremost promoters of AI research. In Israel, a hotbed of innovation in AI, it was the defence sector that provided a critical boost to the field of computer vision.

Using computer vision to search, filter and analyze visual media captured by cameras, and sensors in a battlefield environment helps develop safer combat strategies and protect soldiers. Missile guidance is another example where locally acquired image data is analyzed to select precise target areas.

The usage of drones or unmanned aerial vehicles (UAV) have sparked heated debates among governments. These are robots which collect images and video from in-built or mounted cameras to detect enemy objects and take offensive action on their own.



Drones can collect images and video from mounted cameras to detect enemy objects and initiate offensive action on their own.

REPLACING HUMAN DRIVERS

Some of the popular science fiction works in the last century told us we would have robots driving our cars. While they were right in that the actual task of driving is handed over to machines, it's the cars themselves that are intelligent today.

The autonomous vehicles industry is buzzing with activity. A number of major manufacturers and tech giants are entering the market, and receiving major media coverage. This is no surprise considering its rate of growth – the global market is projected to grow at a staggering rate of 39.47 percent between 2019 and 2026, according to Allied Market Research in August 2018.

Based on the level of autonomy offered, self-driving vehicles fall into five stages. Level 1 vehicles alert drivers as they drift lanes, enable emergency braking and cruise control, but require a significant involvement from the human driver. At the other end of the spectrum are Level 5 vehicles which are fully autonomous and require minimal input from the driver. Most of today's self-driving vehicles fall into Level 4, where self-driving is possible but within pre-mapped routes.

Leaders of the pack include Alphabet's Waymo which has run self-driving cars for over 5 million miles in 25 cities. General Motors plans to launch a driverless ride-hailing service next year with no steering wheels or pedals. Mercedez-Benz also has models which can sense when a car bears down too close to a rear bumper ahead, steer clear of pedestrians and avoid accidents. Others like Audi, BMW, Toyota, Ford and Volvo are also developing self-driving capabilities and testing models that work at Level 4 autonomy.

Elon Musk's Tesla, with the Autopilot feature, is capable of passing other cars and changing lanes without a need for a hand on the wheel. Tesla plans to improve their image recognition capabilities with cameras that can read signs and truly see the road ahead. Chinese search engine giant Baidu is also following their American counterpart in venturing into the self-driving space, aiming to have autonomous cars at Level 3 and 4 in the next four years.

The global market for autonomous vehicles is projected to grow at 39.47% between 2019 and 2026.

GIVING FARMERS INSIGHT INTO CROP HEALTH

A lesser known but massively impactful use of computer vision is to analyze and monitor crops in agriculture. Using camera-mounted drones, farmers can capture images of the field to detect the health of crops, pest infestations and other deficiencies that could impact harvest yield.

Slantrange is one firm which specialises in providing drones to remotely assess fields and provide farmers with data to improve their operations. Robots such as those developed by Blue River Technology use computer vision to monitor plants and control weed by spraying herbicides only where needed.

Aerial view of camera-mounted drone above a farm

A WHOLE NEW RETAIL EXPERIENCE

The announcement of Amazon Go in 2016 heralded the arrival of future of retail. When shoppers walk into an Amazon Go store, shopping for groceries means scanning their phone as they enter, picking up what they need, and walking out. There is no checkout; items chosen are tracked through RFID and cameras, and the shopper's Amazon account is billed automatically through a dedicated app. While it is a revolutionary concept that will make the shopping experience a breeze, the system is still not 100 percent accurate and experts foresee another decade or so before it becomes commonplace.

A more immediate application of computer vision in the retail sector is in capturing shelf images to analyze individual products. For example, Trax helps digitize the shelf to reduce audit times for sales reps, and translate the images to data for category management, shopper marketing and space planning teams to reduce out-of-stocks, improve distribution and gain market share over competitors.

Milions of product images are uploaded weekly to our Trax cloud

Fuelling the most advanced computer vision recognition for retail

THE UNIQUE CHALLENGE OF IMAGE RECOGNITION IN RETAIL

While today's advanced image recognition algorithms are capable of recognizing objects within an image with great accuracy, the process becomes much more complex in a retail setting.

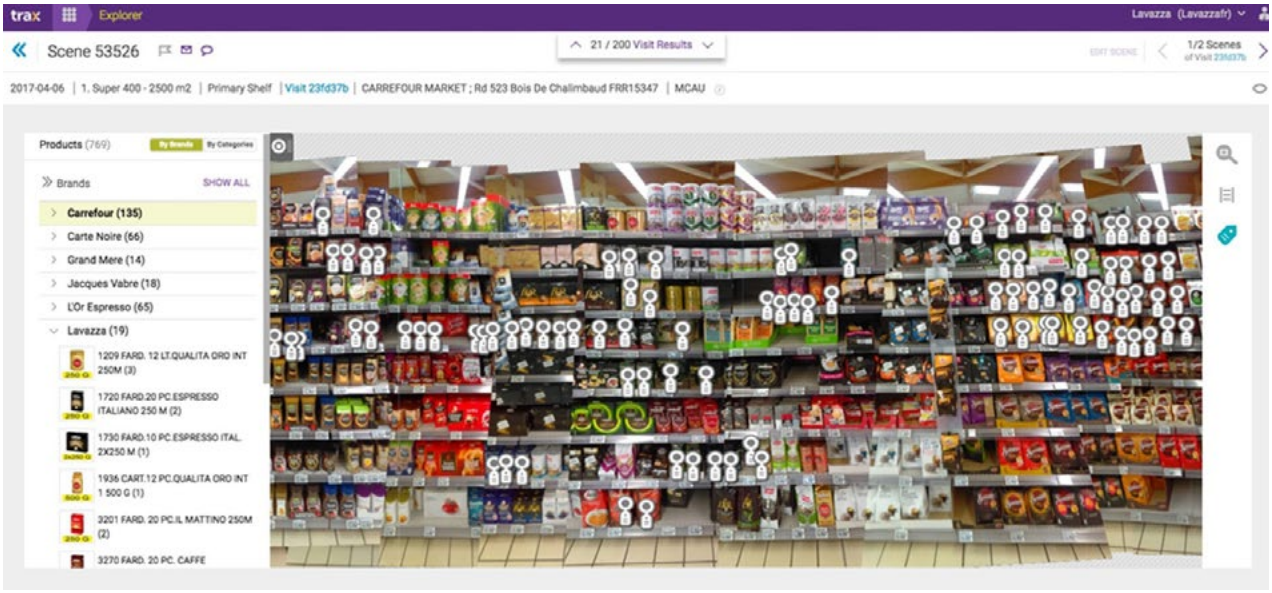
In the more commonplace situations, computer vision algorithms are tasked with identifying and distinguishing objects like chairs, tables, dogs, cars, pedestrians or road signs.

But consider a retail store scenario. There are over 30,000 SKUs in a typical grocery store and on average, 30% of the shelf changes each year with new products and packaging designs launched frequently. So an automated image recognition platform must be able to meet these key criteria to ensure a high level of accuracy:

- Distinguish multiple products that are nearly identical in appearance
- Overcome obscure and reflective packaging under poor lighting conditions
- Identify empty spaces and partially obstructed products
- Detect changes in the product life cycle like new design versions
- Know when a portion of the shelf is omitted from the image capture process

Example of differences in packaging on a Classic Coca-Cola 500ml bottle

Now scaling this across millions of products across multiple stores and getting complete, accurate data on true shelf conditions requires a highly sophisticated system. Deep-learning architectures trained on billions of shelf images can recognise fine-grained differences, such as a large and small version of the same product. Images taken from multiple viewpoints are stitched together and projected on to a single reference frame. Advanced geometrical techniques determine the size and location of every product on the shelf, and this provides the foundation for generating data across several shelf metrics like out of stocks and share of shelf.



Essential Reading on Computer Vision in Retail

Why accuracy is key in Computer Vision-powered retail execution

Is your in-store data giving you the complete picture?

How Computer Vision extracts data from in-store images

The Trax Computer Vision Platform

Expert Q&A: Computer Vision as a Commodity

Yair Adato, Chief Technology Officer at Trax on his own thoughts about where computer vision is headed, especially in the retail space.

What were some of the early applications of computer vision?
Barcode scanning was one of the early applications. But I know a colleague who says that the computer vision community ought to be ashamed about how this application is still around today. It's ugly and inelegant. In fact, barcodes originated from Japanese industrial engineering, when they looked for something they could recognize quickly on the production line.

Face recognition was a holy grail for a while, but it's something we are almost close to solving. Face detection was another big application, and having got better for over 20 years now, we know how to do that as well. In fact, we are at as stage where computer vision systems are actually better than humans in some cases, because they have access to enough information.

Is that true? Can computer vision systems be better than humans at face detection?
Absolutely. In fact, it's already happening. There are two ways that machines are better than humans.

Firstly, in some scenarios, automated machines are better than humans by definition. For example, for missions like validation, humans can make errors, tend to be tired before lunch and so on. But computers don't – they can do these tasks over and over again.

Secondly, if we have enough information, computers can do better than humans. We have various recognition challenges around the world, like ImageNet where if you have enough information, then your system could outperform humans. And it's not a surprise right? Computers have won at chess for a while now.

At the end of the day, it's a question of how much computing power we have access to and how much data we can feed the systems. And we have a lot of both today.

How did the arrival of deep learning revolutionize computer vision?
Firstly, deep learning just made things work better. So today, you can get reasonable results out of the box quite fast. In fact, there is a huge effort to make CV a commodity. Big players like Google, Facebook, Microsoft and Amazon want to give you CV solutions out of the box. And they will soon do that for mainstream applications. But niche applications, like what we do at Trax in recognizing on-shelf images, will still remain niche.

OpenCV is a big revolution that's happening today. It gives users a lot of algorithms – implemented and ready for use. Any engineer can work with just an internet camera and six to 10 lines of code.

In general, technology has become much easier than it used to be 20 years ago. A team of four people can have a reasonably big CV system. They don't need massive data centers or IT resources, they have the cloud. They can use open source tools for web and mobile development.

Computer vision will have the same effect. Soon you will be able to just upload your image to Google and it will do the recognition for you. We will soon have Computer Vision-as-a Service, and it will work, at least for 80 percent of applications.

But if you want to develop something new or niche or take an application's capabilities to the next level, then you need niche capabilities. For instance, in a niche domain like CV in retail, where you have properties that are not common – crowded environments, ever-changing SKUs, near-identical or similar products, you will need a dedicated CV solution.

Research at Trax:

Yair Adato

Chief Technology Officer

When asked about his motivation for going into computer vision, Yair jokes, "I got into this field by chance, like many other things in life". Aside from being a "cool research area" and one in which the results of an algorithm or experiment can be tested immediately, Yair also believes that computer vision is one field in academia that is bound to start a revolution, similar to what computer graphics did.

"In the 90s, there was tremendous progress in the field of computer graphics. The sentiment was that it would only take a few years for the field to venture out of academia into real world applications like better mobile screens, movie graphics and such. There is a similar feeling today about AI and ML being the next revolution."

Trained in mathematics and computer science, Yair went on to complete his Masters and PhD from the Ben-Gurion University of the Negev in Israel. One of his focus areas was the incredibly complex problem of reconstructing a 3D structure of specular shapes, or mirror-like objects which reflect some or all of the light falling on them. Yair and his team challenged the conventional understanding that this required a knowledge of the environment surrounding the object by discovering that one only needed to understand motion of the object.

He also worked on another fascinating project that involved fitting an owl with a head-mounted wireless micro camera to study "visual pop-out" – the animal behaviour where an eye fixates on a target irrespective of any distractions present in the view!

Chapter 4: What Lies Ahead

As with any ground-breaking technology, predicting what the future holds often involves a lot of educated guesswork. Even within computer vision, researchers and practitioners did not anticipate the sudden growth following the arrival of convolutional neural networks around 2012.

However, experts within the field point to some future scenarios where computer vision makes automation tasks easier. Nearly all processes which involve some level of computer vision to automate tasks are expected to get better. As the accuracy levels of image recognition go beyond what humans are capable of, machines will be entrusted to completely take over the "seeing" part from humans, only leaving the remaining to tasks for human involvement.

ROBOTS WILL TAKE OVER

One of the biggest disruptions will be seen in the field of autonomous vehicles. Today's self-driving vehicles at Level 4 can safely navigate through pre-mapped routes but as computer vision accuracy rises, we will have cars that can drive almost entirely on their own. Ziv Mhabary, VP of Computer Vision at Trax says, "we may not have our streets entirely navigated by autonomous cars in five to 10 years but they will definitely be more ready for usage soon."

Today we have robots capable of performing individual predefined tasks like folding clothes and cleaning floors. In the future, equipped with better AI, these will be expected to take on multiple tasks in the form of an autonomous assistants.

Robots equipped with AI will soon be able to take on multiple tasks as autonomous assistants.

THE INTERNET OF THINGS WILL BECOME MORE INTELLIGENT

The number of smart, interconnected devices in the market has been rising over the last few years. Gartner predicts that IoT endpoints will grow 32 percent every year to reach an installed base of 25.1 billion units by 2021. These include applications for both businesses and consumers.

64 percent of IoT devices shipped by 2021 would be for consumer usage. These would include home security systems equipped with cameras, wearable tech like smart watches and heart rate monitors and AI-powered smart fridges that could capture images inside to create inventories and alert users on missing items and expiry dates.

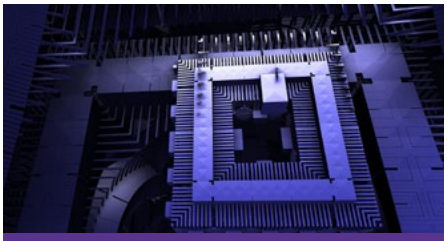
Similar IoT devices also find business use cases in the retail space in bars and convenience stores through door-mounted cameras on coolers, which enable brands to monitor out-of-stock levels, purity and compliance. Sensors and drones will also become widely used to check for gas leakages and carry out industrial inspections that otherwise pose safety hazards for humans. In the manufacturing sector, cameras will get better at catching anomalies in production facilities and identifying flaws in raw materials.

IoT endpoints will grow 32% every year to reach an installed base of 25.1 billion units by 2021

QUANTUM COMPUTING WILL ACCELERATE THE FIELD

Researchers in the field of computer vision and artificial intelligence are keenly following developments in quantum computing, which takes advantage of strange quantum phenomena like superposition and entanglement. While traditional computation relies on systems that store information in binary bits of either 1 or 0, quantum computers use what are called "qubits". Qubits can exist in multiple states of 1 and 0 at the same time, thereby increasing computing power exponentially.

Today's quantum computers are not yet practical enough to be useful, requiring extreme environments like near-zero temperatures to function. But if and when they become feasible, these are expected to speed up computation capabilities exponentially, allowing computer vision algorithms to sift through large data sets, identify patterns and carry out image recognition with even more accuracy.



Quantum computers, when they become feasible, will speed up our computation capabilities exponentially, making it easier for algorithms to sift through large data sets and recognize images more accurately.

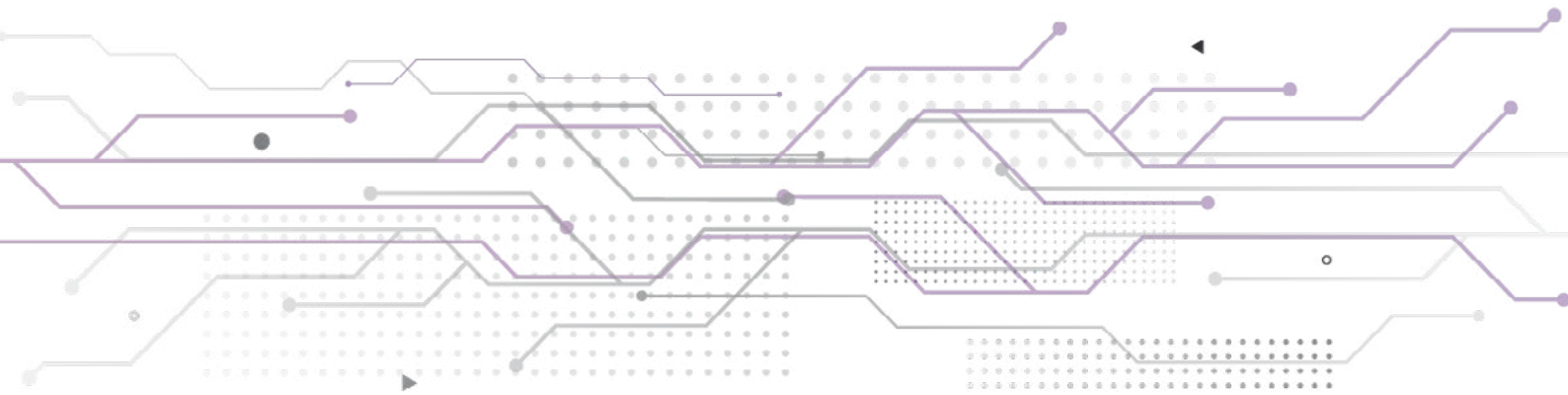
MORE INVENTIVE USES OF AUGMENTED REALITY

Augmented reality combines an understanding of 3D space and image recognition to help users understand and interact with the world, opening doors to various applications. There are several examples of augmented reality in action today at a basic level, like IKEA's app that lets shoppers see how a furniture would look like inside their homes. Even the native Measure app in iPhones uses augmented reality to help people do away with physical tapes for measuring dimensions of everyday objects.

But computer vision experts are excited about more game-changing applications of augmented reality in the future. Amazon Go's plans for frictionless retail has shown a lot of promise to change the entire shopping experience. But experts believe there is still some way to go for their vision to be realised.

"I think it is still at a proof-of-concept stage", says Ziv. "It is also very expensive so I doubt it will take off immediately. It may become a reality in 10 years but between now and then, what we will see are technologies that close the gap between very slow checkouts to no checkouts." Augmented reality is one such additional layer that moves the needle along the way towards frictionless checkouts. For example, it could allow shoppers to find the nearest retail store, be guided there and given assistance in finding exactly what they are looking for.

Augmented reality will help people interact with the real world in new ways and create many new applications.



INNOVATION @ TRAX

Trax has been at the forefront of innovations in computer vision for retail execution. It's technology stack includes state-of-the-art recognition capabilities and the largest retail image repository that helps in detecting products at brand and sub-brand level with over 96 percent accuracy.

In addition, our researchers have been hard at work devising new capabilities to prepare for the future of retail.

Menu Recognition in Bars and Restaurants

While supermarkets and convenience stores form a significant part of CPG brands' distribution, other channels like bars and restaurants will have their own unique needs. For example, beverage brands often look at "fair share of the menu" to see how often their brand shows up in the menus.

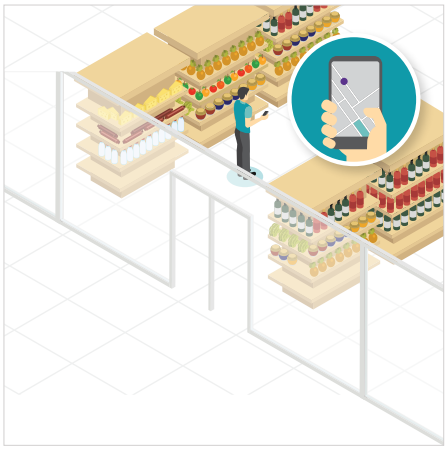
Trax is working on a solution that combines OCR and Natural Language Processing (NLP) to capture menus, understand and translate them into insights for brands.

Shopping Experience 2.0 with Store Mapper

As a consumer, shopping in a retail store can sometimes be an ordeal. Walking around the aisles to find the right categories, reading nutrition labels and product information, and constantly checking your shopping list can drain shoppers' time and energy. They may even run into unexpected out of stocks and miss valuable promotions in the process, making it an unproductive experience altogether.

To make shopping a breeze, Trax is currently researching a solution called Store Mapper, which will map physical retail stores and digitize them into a 2D map. Shoppers can open an app on their phones to be directed to the right aisles using AR-based location guides, targeted with location-based promotions and be alerted on any items that are running out. Scanning a product on the phone will open up its information and nutritional value while a virtual assistant helps shoppers add and track their lists. Last-minute requests or changes to the shopping list done by friends or family will also be reflected in the syncable shared lists.

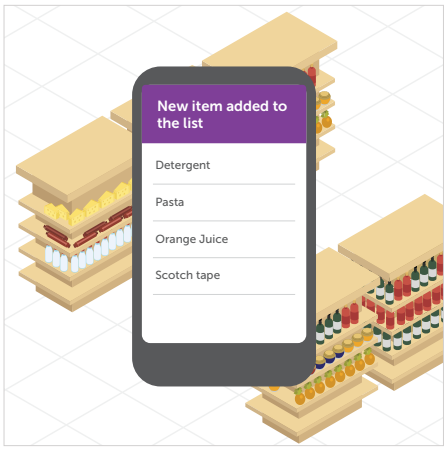
The Store of the Future



Shopper guided by AR on the Trax app

Scanning a product will reveal more info

Location-based promotions pop up on the app



Changes to shared shopping lists are

Dolev, Ziv, Yair and their colleagues at Trax are working on this very project for their clients. "One application we are quite excited about is the store mapper. It combines understanding of 3D space, mapping and image recognition to help the shopper understand what items are available in the store and where they are placed. "

"Frictionless retail is one thing" says Dolev, "but imagine if you can be guided as you go into the store to find exactly what you are looking for. Let's say you are searching for only organic or gluten-free products for instance. A store mapper powered by augmented reality could help you find them, compare prices and use a navigation route to get there instantly. This would make the shopping experience even more seamless."

About Trax

Trax is the world leader in computer vision solutions for retail, ranking in the top 25 Fastest Growing Companies on Deloitte's Technology Fast 500 list. The company enables tighter execution controls in-store and the ability to leverage competitive insights through their in-store execution tools, market measurement services and data science to unlock revenue opportunities at all points of sale. Trax does this using smartphones and tablets to gain actionable shelf analytics in real-time. With over 80 clients, in over 45 countries, top brands such as Coca-Cola, AB InBev, Nestle, Henkel, PepsiCo and many more, leverage Trax globally to manage their in-store execution and increase revenues at the shelf. Trax is headquartered in Singapore with offices worldwide. To learn more about Trax, please visit **www.traxretail.com/contact**

www.linkedin.com/company/trax-retail

www. twitter.com/TraxRetail

www.facebook.com/TraxRetailTech

trax