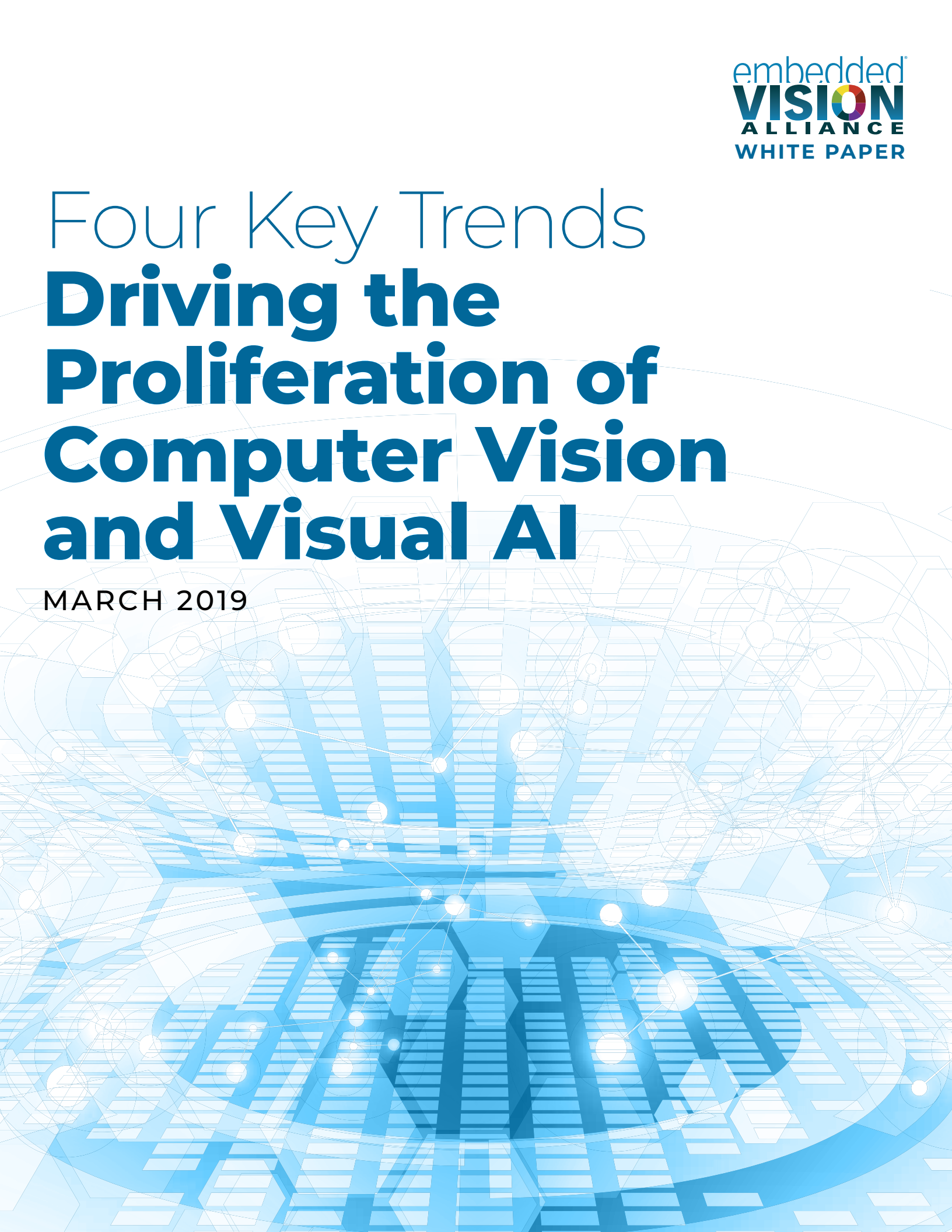
embedded VISION ALLIANCE

WHITE PAPER

Four Key Trends
Driving the Proliferation of Computer Vision and Visual AI

MARCH 2019



With so much happening in computer vision applications and technology, and happening so fast, it can be difficult to see the big picture. This white paper discusses the key trends driving the proliferation of vision applications and influencing the future of the industry, explaining what's fueling each of these trends, and highlighting key implications for technology suppliers, solution developers and end-users.

Join computer vision industry leaders at the 2019 Embedded Vision Summit!

The Embedded Vision Summit is the only conference focused on practical computer vision and deep learning for visual AI. Coming up May 20-23, 2019, you'll learn the latest in this rapidly changing field through:

Inspiring **keynotes**

90+ presentations on **applications, trends, technologies and business opportunities** for vision-based products

100+ demos from 60+ exhibitors in the Vision Technology Showcase

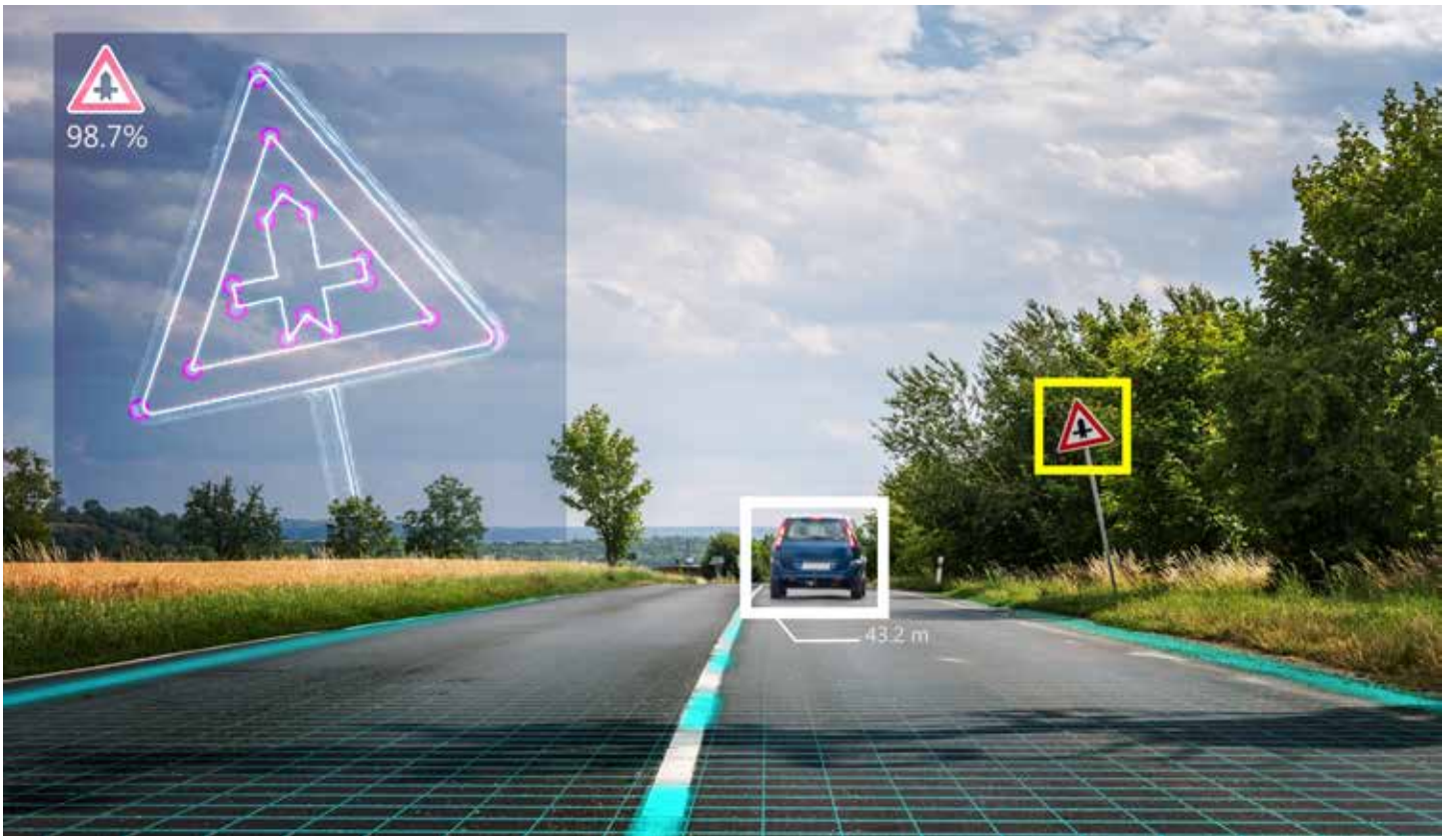
Hands-on trainings on TensorFlow 2.0 and OpenCV

Full-day Vision Technology Workshops provided by Alliance Member companies

To register visit **embedded-vision.com/summit**!

0319R3

Introduction



Back in 2011, when the Embedded Vision Alliance launched, its founding companies believed that there would soon be unprecedented growth in investment, innovation and deployment of practical computer vision technology and solutions across a broad range of markets. Less than a decade later, that prediction has come to pass in earnest. Investments in both U.S. and China-based computer vision companiels have accelerated over the past six years, translating into 100x investment growth across that time frame, and growth that shows no sign of slowing down any time soon (Fig. 1a)

Those investments are spurring these companies, along with their partners and customers, to accelerate their vision-related

Investment in Computer Vision-Related Companies in China vs. US

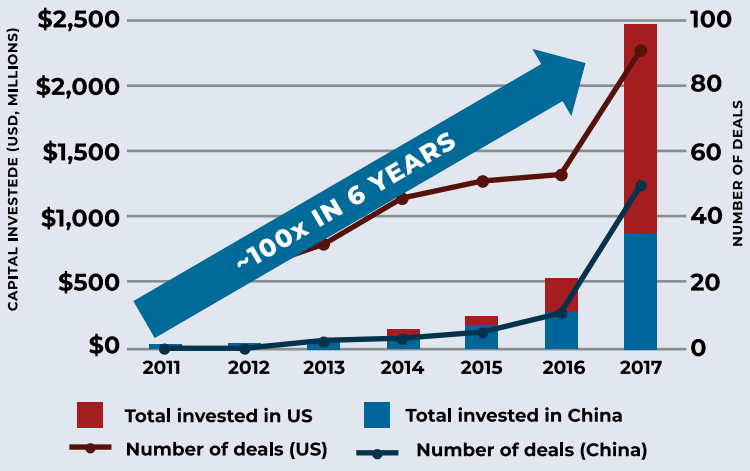


Figure 1a. Worldwide investments in computer vision-related companies are significant and growing.

Source: Woodside Capital/Crunchbase

research, development and deployment activities. The Alliance regularly surveys the vision developer community on a variety of subjects, and the most recent results indicate that 93% of the organizations surveyed report an increase (61% reporting a large increase, i.e., >10%) in vision-related activity over the coming year (Fig. 1b). And those increased activities are predicted to translate into disproportionately increased revenue returns; for example, a recently published market research report by Tractica forecasts a 25x increase in revenue for the computer vision market (encompassing hardware, software and services, Fig. 1c) between now and 2025, a year when it will surpass $26 billion.

The four key trends underlying these factors are the focus of the remainder of this white paper:

1. Deep learning
2. 3D perception
3. Fast, inexpensive, energy-efficient processors
4. The democratization of hardware and software

Growth Predicted in Developers' Vision-Related Activity

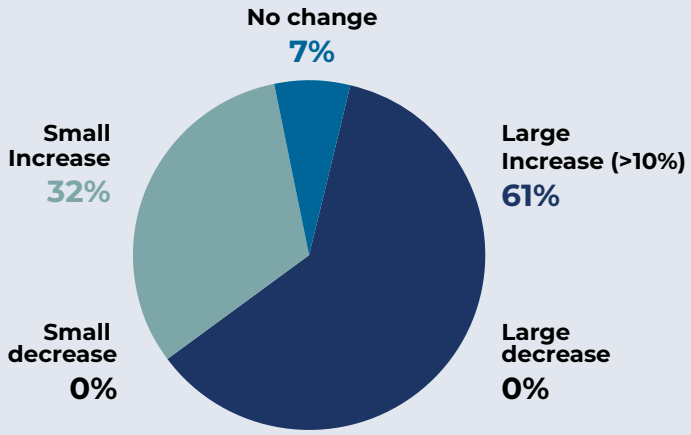


Figure 1b. Developers anticipate an increase in their organization's vision-related activity in 2019.
Source: Embedded Vision Alliance's November 2018 Computer Vision Developer Survey.

Projected Total Revenue for Vision-Related Companies

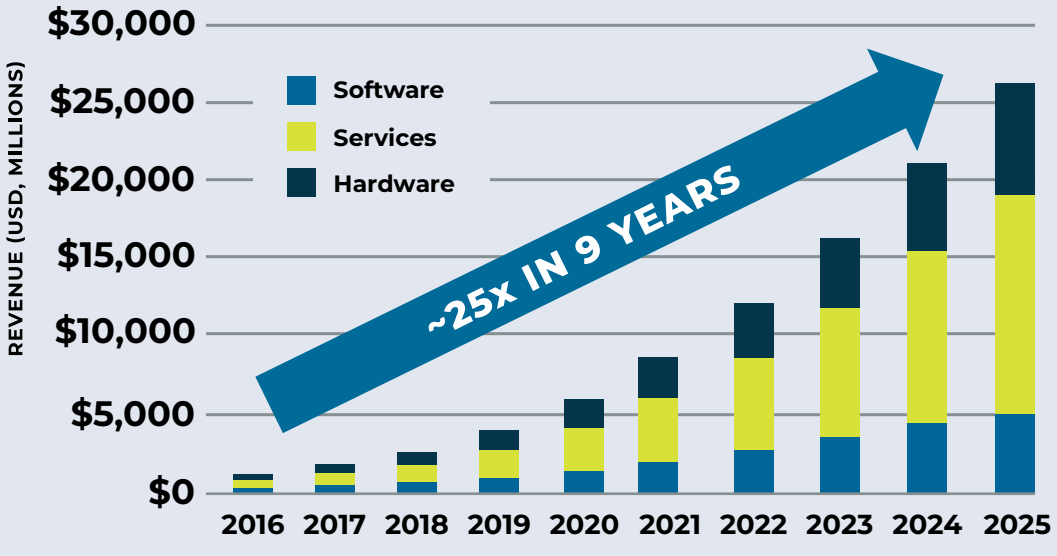


Figure 1c. Worldwide investments in computer vision-related companies are leading to dramatic revenue growth for those companies going forward.
Source: Tractica.

Deep Learning

TRADITIONALLY, COMPUTER VISION applications have relied on special-purpose algorithms that are painstakingly designed to recognize specific features (e.g., edges, corners, objects). Recently, however, convolutional neural networks (CNNs) and other deep learning approaches have been shown to be superior to traditional algorithms on a variety of image understanding tasks. In contrast to traditional algorithms, deep learning approaches are generalized learning algorithms trained through examples to recognize specific features, including object types and locations. Deep neural networks (DNNs) have transformed the field of computer vision, delivering superior results on functions such as recognizing objects, localizing objects within a frame, and determining which pixels belong to which object. Even problems like optical flow and stereo correspondence, which had been solved quite well with conventional techniques, are now finding even better solutions using deep learning techniques.

Deep neural networks have been trained to fill in the missing patches in photographs, matching the skill sets of adept operators of photo editing software packages—not to mention delivering comparable results far faster than a skilled human being could.

And not only are deep learning-based vision processing approaches proving superior to traditional computer vision algorithms in solving many problems, they're also beginning to give humans a serious "run for their money." Imagenet image recognition challenge winners' results show that, beginning a couple of years ago, their accuracy at identifying objects began to exceed typical humans' performance at the same task and on the same dataset (Fig. 2a). Deep neural networks have also been trained to fill in the missing patches in photo-

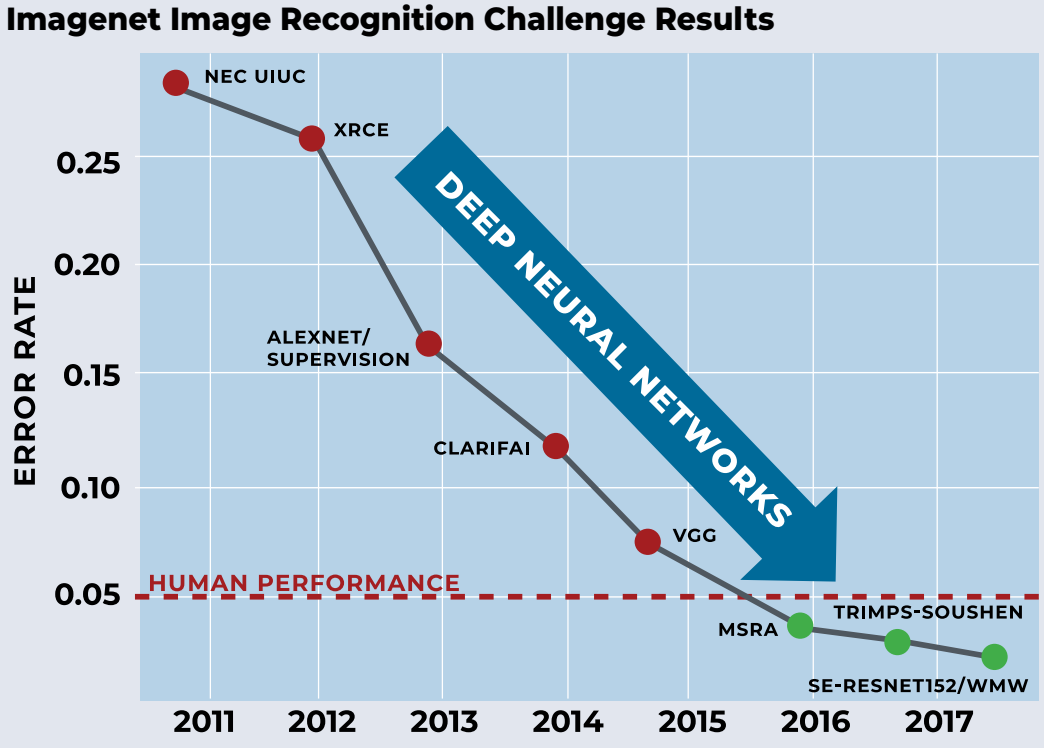


Figure 2a. Deep learning algorithms are now capable of matching humans' capabilities to accurately identify objects in images.
Source: www.eff.org/ai/metrics

1. Camera output with ISO 8,000

2. Camera output with ISO 409,600

3. Image result from algorithm trained on low-light dataset.

Figure 2b. Deep learning algorithms can, in some cases, produce results that exceed humans' capabilities.
Source: Learning to See in the Dark, Chen Chen, Qifeng Chen, Jia Xu and Vladlen Koltun, CVPR 2018.

graphs, matching the skill sets of adept operators of photo editing software packages, not to mention delivering comparable results far faster than a skilled human being could. And well-trained neural networks are even starting to deliver achievements that go well beyond what even skilled humans can do, such as producing acceptable images from very poor-exposure source photos (Fig. 2b).

Unsurprisingly, therefore, computer vision developers are increasingly adding deep learning techniques to their tool chests (Fig. 3). In the Alliance's most recent survey results, 59% of vision system and solution developers are already using DNNs, an increase from 34% just two years ago. Another 28% are planning to use DNNs for visual intelligence in the near future. In total, 87% of surveyed developers already use or plan to use neural networks to perform computer vision functions.

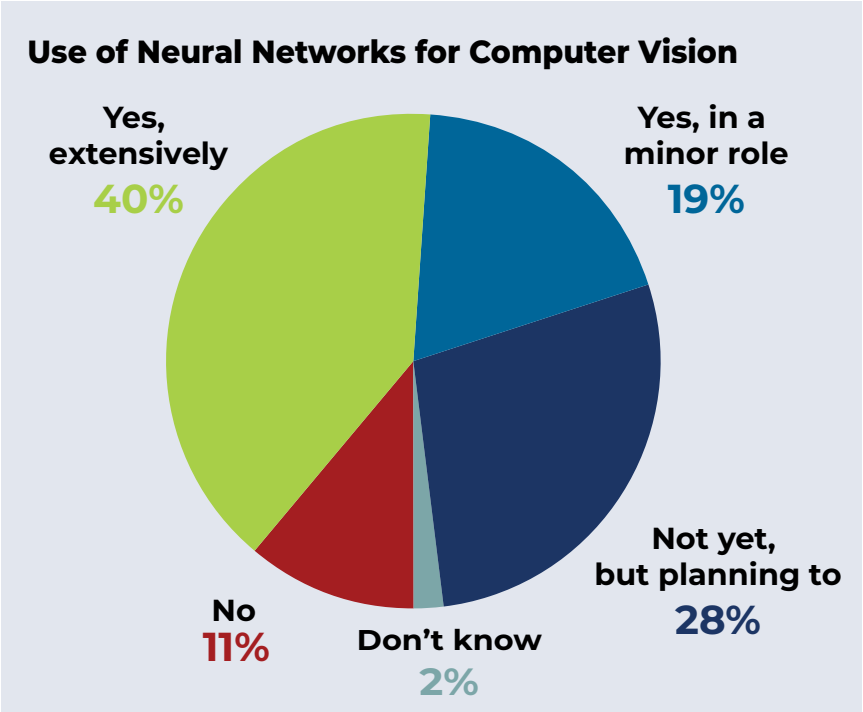


Figure 3. 87% of developers surveyed already use or plan to shortly begin using neural networks to perform computer vision functions.
Source: Embedded Vision Alliance's November 2018 Computer Vision Developer Survey

THE 2D IMAGE SENSORS found in many embedded vision system designs enable tremendous vision capabilities. However, their inability to discern an object's distance from the sensor can make it difficult or impossible to implement some vision functions. Consider, for example, a gesture interface implementation. The ability to discern motion not only up-and-down and side-to-side but also front-to-back greatly expands the variety, richness and precision of the suite of gestures that a system can decode. Or, consider face recognition (Fig. 4): depth sensing is valuable in determining that the object being sensed is an actual person's face, versus a photograph of that person's face.

ADAS (automotive advanced driver assistance system) and other semi- and fully-autonomous device applications that benefit from 3D sensors are abundant. You can easily imagine, for example, the added value of being able to determine not only that another vehicle or object is in the roadway ahead of or behind you, but also to accurately discern its distance from you. Precisely determining the distance between your vehicle and a speed-limit-change sign is equally valuable in ascertaining how much time you have to slow down in order to avoid getting a ticket.

The ability to discern motion—not only up-and-down and side-to-side but also front-to-back—greatly expands the variety, richness and precision of the suite of gestures that a system can decode.

Similarly, 3D scanning of objects for 3D printing is an important use case, as are applications such as manufacturing line "binning." Fortunately, the recent introduction of 3D optical sensors into high-volume applications like mobile phones and automobiles has catalyzed a dramatic acceleration in innovation, collapsing the size, cost and complexity of 3D perception (Fig. 5, next page). 3D camera modules often include some form of infrared illumination, which has similarly benefited from recent significant

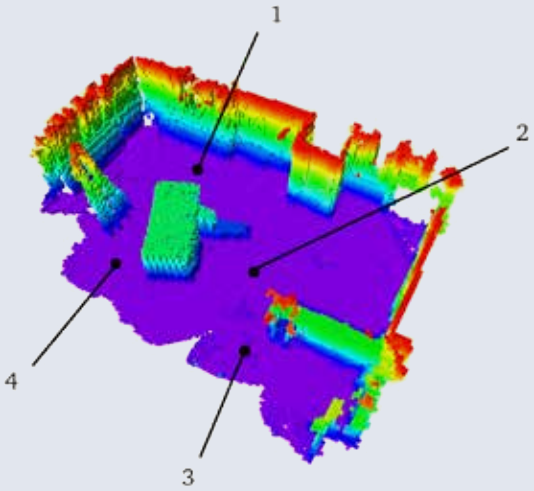


Figure 4. Face recognition (above) and visual simultaneous localization and mapping, (vSLAM, below) are two of the many capabilities enabled by 3D image sensing.
Top: appleinsider.com, bottom: pcc.disam.etsii.upm.es

cost reduction trends and is useful in low ambient light environments as well as, for example, when monitoring the attentiveness of a vehicle driver who is wearing sunglasses.

Eight years after the debut of the Microsoft Kinect game console peripheral, 3D camera modules are now ready for deployment in cost- and power-sensitive applications. And, as with previously described deep learning, computer vision developers are responding with aggressive adoption actions and forecasts. Almost 30% of developers who participated in the Alliance's most recent survey are already using 3D perception, with another 26% planning to incorporate it in near-future projects (Fig. 6).

Camera Technologies

Figure 5. The latest generation of small, low-cost, and low-power 3D cameras enable robust vision deployments (above); their infrared illumination modules are similarly becoming increasingly cost-effective (below).

Top, clockwise from top left, Microsoft, Intel, and Occipital. Below: Yole Développement

Cost Comparison of Infrared Cameras for both 2D & 3D Imaging

	Processor	Optics	Manufacturing	Total
Active Stereo	$8 *(40% of cost)*	$6 *(30% of cost)*	$6 *(30% of cost)*	$20
Structured Light	$9 *(45% of cost)*	$4.5 *(22.5% of cost)*	$6.5 *(32.5% of cost)*	$20
Time-of-Flight	$12.50 *(41.2% of cost)*	$4.50 *(15% of cost)*	$13 *(43.3% of cost)*	$30

Use of 3D Perception in Vision-Related Projects

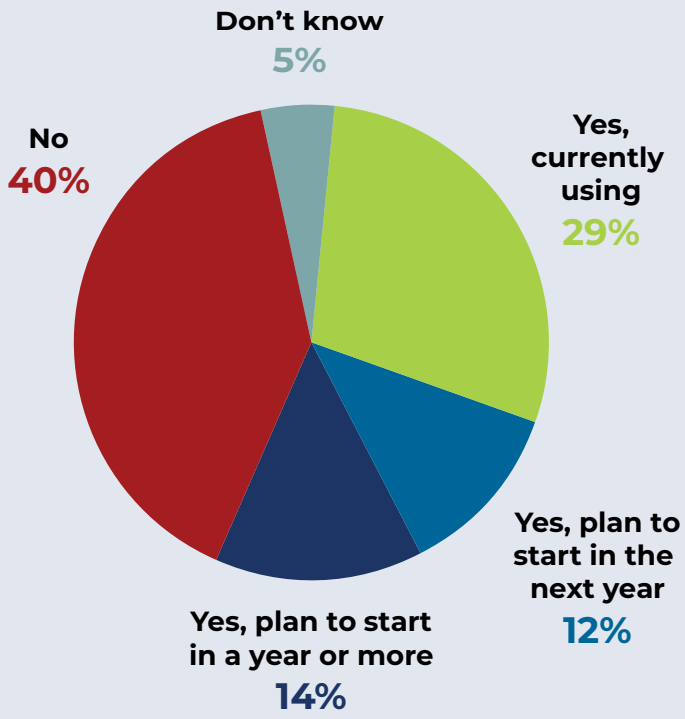


Figure 6. 55% of developers surveyed already use or plan to shortly begin incorporating 3D perception in their computer vision projects—a 4% increase since last year.

Source: Embedded Vision Alliance's November 2018 Computer Vision Developer Survey

Better Processors

THE MOST IMPORTANT ingredient driving the deployability of robust and widespread visual perception is better processors. By "better" we mean higher performance, lower cost, lower power consumption, and improvements in other key factors. Vision algorithms are very demanding of compute performance and embedded systems of all kinds are usually required to fit into tight cost and power consumption envelopes. In other application domains, such as digital wireless communications and compression-centric consumer video equipment, chip designers achieve this challenging combination of high performance, low cost, and low power by using specialized coprocessors and accelerators to implement the most demanding processing tasks in the application. These co-processors and accelerators are, however, typically not programmable by the chip user.

This tradeoff is often acceptable in standards-based applications, where there is strong commonality among algorithms used by different designers. In vision applications, however, there are no standards constraining the choice of algorithms. Moreover, vision algorithms are developing rapidly and change frequently.

Achieving the combination of high performance, low cost, low power, and programmability is therefore challenging, and often accomplished by combining multiple processor types (CPU, GPU, FPGA, DSP, etc.) in a heterogeneous computing architecture.

Machine learning-based vision processing is particularly resource-intensive for both upfront training and subsequent inference tasks, as measured by its compute and memory requirements. Fortunately, vision processors are improving at an astounding rate, a reflection both of the rapid overall pace of ongoing development and the competitive pressures derived from the large and still-growing number of technology suppliers. Right now, for example, there are more than 50 companies simultaneously developing processors for deep learning inference and/or training. In the last few years there have been two orders of magnitude improvements in processing power for deep learning acceleration. Compounding this performance increase over multiple generations yields an exponential increase in processing power.

Data collected in the most recent Embedded Vision Alliance developer survey reveals dramatic adoption of deep learning-specific processors; nearly 1/3 of respondents are using them now, versus only 19% just two years ago (Fig. 7). This trend is particularly astounding considering that the deep learning-specific processor category didn't even exist a few years ago. Note, too, the other processing architectures commonly used for various vision tasks. (Keen-eyed readers will notice that the total adds up to more than 100%; this is because survey respondents were instructed to mark all processing options that applied to their projects. As previously noted, it's common in vision-based products to leverage both a CPU and a GPU, for example, in a heterogeneous computing arrangement.)

Types of Processors Used for Vision Tasks, Ranked as One of Top Three

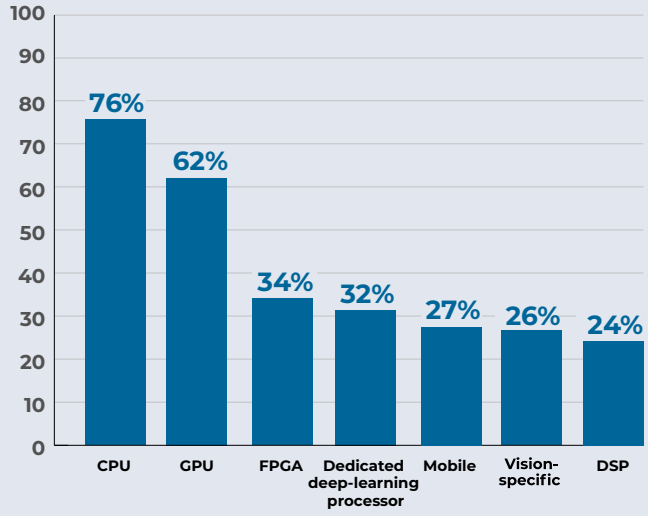


Figure 7. Developers surveyed leverage diverse processing architectures, both standalone and in heterogeneous combinations, in their computer vision design.
Source: Embedded Vision Alliance's November 2018 Computer Vision Developer Survey

The Democratization of Hardware and Software

"DEMOCRATIZATION" MEANS THAT it is rapidly getting easier to develop effective computer vision-based systems and applications, as well as to deploy these solutions at scale. Why? Three reasons:

1. Deep learning makes it easier for non-experts to create functional vision systems using sample image data (vs. hand-engineered code)

2. Higher-performance, lower-cost processors, with effective development tools, are now available

3. Cloud computing is becoming increasingly common as an adjunct or substitute for edge-based processing

The first two points have been covered already but the third point deserves some attention. Cloud computing is becoming increasingly common as an adjunct to, if not a substitute for, historical edge-based vision processing approaches.

The cloud-vs-edge-vs-both (i.e., hybrid) topology decision is often not straightforward, and the "right" answer will vary from one application to another, one company to another, and even one project to another within the same company (Fig. 8). Factors that favor the cloud include:

Time-to-market: Software development for the cloud is usually faster and easier than developing for an embedded platform.

Upgradability: Within limits, you can easily upgrade to higher performance processors, more memory, more hard drive storage, newer operating system and middleware versions, and the like. But you obviously can't upgrade everything in the cloud – e.g., you can't upgrade your image sensor.

Accuracy: You have access to massive compute power in the cloud, so you can run larger neural networks and, more generally, more complex algorithms, including being able to draw on bursts of as-needed extra processing power.

Coordination among distributed devices: If you are tracking vehicles moving throughout a city, for example, there's an inherent need to pool information over a geographical area. While the cloud isn't the only way to accomplish this, it can be a convenient mechanism to combine information from numerous distributed edge nodes.

Device cost: Less comparative processing power in the edge device translates into lower BOM (bill of materials) costs, smaller batteries, etc.

Conversely, an edge-centric approach has compelling advantages of its own.

No recurring costs: You're not paying for cloud compute processing, memory and storage resources on a per-use basis

Network connectivity: Often not needed at all, otherwise not often needed consistently

Comparing Cloud- and Edge-Based Vision Processing Topologies

	Edge	Cloud
Time to market		★★★
Upgradability		★★
Accuracy		★★★
Coordination among distributed devices		★★★
Device cost		★★
Recurring Costs	★★★	
Internet connectivity, bandwidth required	★★★	
Response time	★★★	
Privacy and security	★	

Figure 8. Edge vs. cloud benefits. More stars implies a greater advantage

The Democratization of Hardware and Software

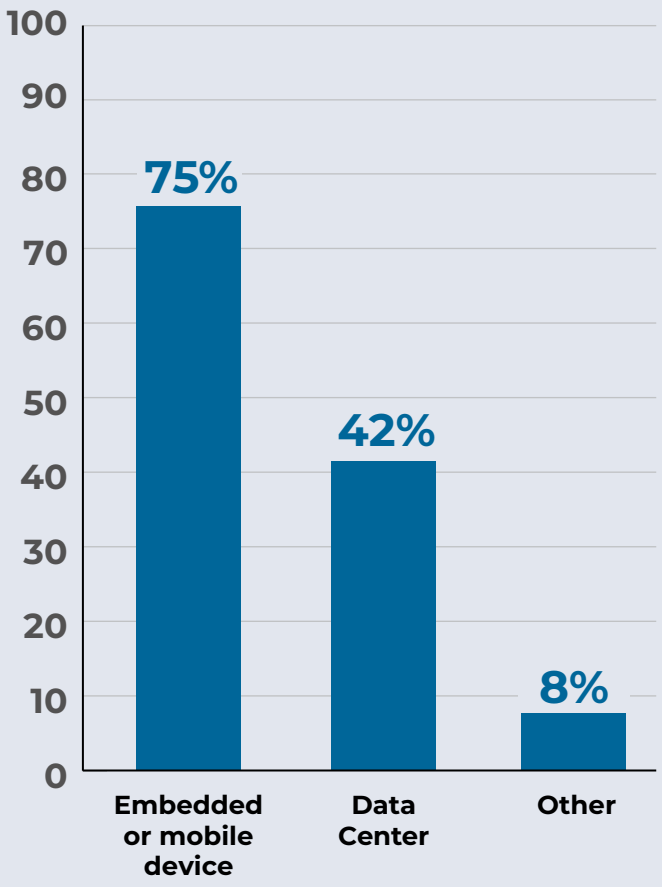


Figure 9. The majority of developers surveyed do at least some neural network inference at the edge; however, nearly 50% also do part or all of the inference functions in the cloud.
Source: Embedded Vision Alliance's November 2018 Computer Vision Developer Survey

Bandwidth and latency: When network connectivity is required at all, its bandwidth and latency requirements are reduced in comparison to a cloud-centric approach since a large percentage of the data processing is done in situ on the edge device prior to cloud transfer.

Privacy and security: Again, raw source data is processed on the edge device and often immediately discarded. The only information that makes it to the cloud is metadata, and is often also anonymized.

The most recent Embedded Vision Alliance developer survey results unsurprisingly show that most respondents do at least some neural network inference at the edge (Fig. 9). However, nearly half of them also do inference in part or in full in the cloud. As with the earlier vision processor discussion, the total adds up to more than 100%, because survey respondents were instructed to mark all options that applied to their projects.

The Key Takeaways

COMPUTER VISION HAS TREMENDOUS POTENTIAL to improve an incredible diversity of applications, from automotive safety to photography. We will see more progress in deployment of practical computer vision in the next five years than we've seen in the past 50 years, because:

Computer vision is increasingly effective—i.e., it "just works";

Computer vision is increasingly accessible to non-specialists, to incorporate into their solutions;

And, computer vision is increasingly affordable to deploy.

Also, the rate of innovation is accelerating, because volume markets are emerging, and awareness is growing, both factors driving rapid expansion in investment of capital and talent.

About the Embedded Vision Alliance

THE EMBEDDED VISION ALLIANCE® is a worldwide organization of technology developers and providers working to empower product creators to transform the potential of vision processing into reality. The Alliance's mission is to provide product creators with practical education, information and insights to help them incorporate vision capabilities into new and existing products. To execute this mission, the Alliance maintains a website providing tutorial articles, videos, code downloads and a discussion forum staffed by technology experts. Registered website users can also receive the Embedded Vision Alliance's twice-monthly email newsletter, Embedded Vision Insights, among other benefits.

The Embedded Vision Alliance also offers a free online training facility for vision-based product creators: the Embedded Vision Academy. This area of the Embedded Vision Alliance website provides in-depth technical training and other resources to help product creators integrate visual intelligence into next-generation software and systems. Course material in the Embedded Vision Academy spans a wide range of vision-related subjects, from basic vision algorithms to image pre-processing, image sensor interfaces, and software development techniques and tools such as OpenCL, OpenVX and OpenCV, along with Caffe, TensorFlow and other machine learning frameworks. Access is free to all through a simple registration process.