Everything You Ever Wanted To Know About Computer Vision

Computer vision is the field of computer science that focuses on replicating parts of the complexity of the human vision system and enabling computers to identify and process objects in images and videos in the same way that humans do. Until recently, computer vision only worked in limited capacity.

Early experiments in computer vision started in the 1950s and it was first put to use commercially to distinguish between typed and handwritten text by the 1970s, today the applications for computer vision have grown exponentially. By 2022, the computer vision and hardware market is expected to reach $48.6 billion.

One of the driving factors behind the growth of computer vision is the amount of data we generate today that is then used to train and make computer vision better.
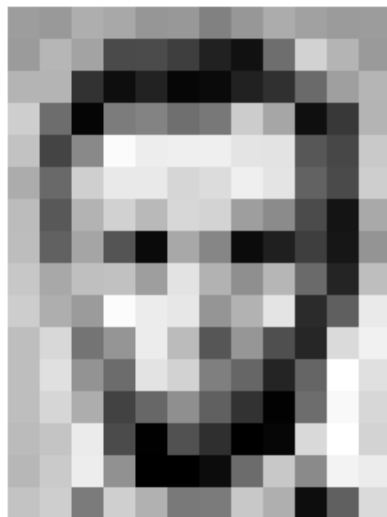

Along with a tremendous amount of visual data (more than 3 billion images are shared online every day), the computing power required to analyze the data is now accessible. As the field of computer vision has grown with new hardware and algorithms so has the accuracy rates for object identification. In less than a decade, today's systems have reached 99 percent accuracy from 50 percent making them more accurate than humans at quickly reacting to visual inputs.

How Does Computer Vision Work?

On a certain level Computer vision is all about pattern recognition. So one way to train a computer how to understand visual data is to feed it images, lots of images thousands, millions if possible that have been labeled, and then subject those to various software techniques, or algorithms, that allow the computer to hunt down patterns in all the elements that relate to those labels.

it will subject them all to algorithms that let them analyze the colors in the photo, the shapes, the distances between the shapes, where objects border each other, and so on, so that it identifies a profile of what "cat" means. When it's finished, the computer will (in theory) be able to use its experience if fed other unlabeled images to find the ones that are of cat.

Below is a simple illustration of the grayscale image buffer which stores our image of Abraham Lincoln. Each pixel's brightness is represented by a single 8-bit number, whose range is from 0 (black) to 255 (white):

157	153	174	168	150	152	129	151	172	161	155	156
155	182	163	74	75	62	33	17	110	210	180	154
180	180	50	14	34	6	10	33	48	106	159	181
206	109	5	124	131	111	120	204	166	15	56	180
194	68	137	251	237	239	239	228	227	87	71	201
172	105	207	233	233	214	220	239	228	98	74	206
188	88	179	209	185	215	211	158	139	75	20	169
189	97	165	84	10	168	134	11	31	62	22	148
199	168	191	193	158	227	178	143	182	106	36	190
205	174	155	252	236	231	149	178	228	43	95	234
190	216	116	149	236	187	86	150	79	38	218	241
190	224	147	108	227	210	127	102	36	101	255	224
190	214	173	66	103	143	96	50	2	109	249	215
187	196	235	75	1	81	47	0	6	217	255	211
183	202	237	145	0	0	12	108	200	138	243	236
195	206	123	207	177	121	123	200	175	13	96	218

In point of fact, pixel values are almost universally stored, at the hardware level, in a one-dimensional array. For example, the data from the image above is stored in a manner similar to this long list of unsigned chars:

```
{157, 153, 174, 168, 150, 152, 129, 151, 172, 161, 155, 156,
 155, 182, 163,  74,  75,  62,  33,  17, 110, 210, 180, 154,
 180, 180,  50,  14,  34,   6,  10,  33,  48, 106, 159, 181,
 206, 109,   5, 124, 131, 111, 120, 204, 166,  15,  56, 180,
 194,  68, 137, 251, 237, 239, 239, 228, 227,  87,  71, 201,
 172, 105, 207, 233, 233, 214, 220, 239, 228,  98,  74, 206,
 188,  88, 179, 209, 185, 215, 211, 158, 139,  75,  20, 169,
 189,  97, 165,  84,  10, 168, 134,  11,  31,  62,  22, 148,
 199, 168, 191, 193, 158, 227, 178, 143, 182, 106,  36, 190,
 205, 174, 155, 252, 236, 231, 149, 178, 228,  43,  95, 234,
 190, 216, 116, 149, 236, 187,  86, 150,  79,  38, 218, 241,
 190, 224, 147, 108, 227, 210, 127, 102,  36, 101, 255, 224,
 190, 214, 173,  66, 103, 143,  96,  50,   2, 109, 249, 215,
 187, 196, 235,  75,   1,  81,  47,   0,   6, 217, 255, 211,
 183, 202, 237, 145,   0,   0,  12, 108, 200, 138, 243, 236,
 195, 206, 123, 207, 177, 121, 123, 200, 175,  13,  96, 218};
```

This way of storing image data may run counter to your expectations, since the data certainly appears to be two-dimensional when it is displayed. Yet, this is the case, since computer memory consists simply of an ever-increasing linear list of address spaces.

How the pixels look:

H	E	L	L	O
O	P	E	N	F
R	A	M	E	W
O	R	K	S	!

How the pixels are numbered:

0	1	2	3	4
5	6	7	8	9
10	11	12	13	14
15	16	17	18	19

How the pixels are stored in computer memory:

H	E	L	L	O	O	P	E	N	F	R	A	M	E	W	O	R	K	S	!

↗ ↗ ↗ ↗
0 1 2 3...

Let's go back to the first picture again and imagine adding a colored one. Now things start to get more complicated. Computers usually read color as a series of 3 values — red, green, and blue (RGB) — on that same 0–255 scale. Now, each pixel actually has 3 values for the computer to store in addition to its position. If we were to colorize President Lincoln, that would lead to 12 x 16 x 3 values, or 576 numbers.

The Evolution Of Computer Vision

Before the advent of deep learning, the tasks that computer vision could perform were very limited and required a lot of manual coding and effort by developers and human operators. For instance, if you wanted to perform facial recognition, you would have to perform the following steps:

Create a database: You had to capture individual images of all the subjects you wanted to track in a specific format.

Annotate images: Then for every individual image, you would have to enter several key data points, such as distance between the eyes, the width of nose bridge, distance between upper-lip and nose, and dozens of other measurements that define the unique characteristics of each person.

Capture new images: Next, you would have to capture new images, whether from photographs or video content. And then you had to go through the measurement process again, marking the key points on the image. You also had to factor in the angle the image was taken.

After all this manual work, the application would finally be able to compare the measurements in the new image with the ones stored in its database and tell you whether it corresponded with any of the profiles it was tracking. In fact, there was very little automation involved and most of the work was being done manually. And the error margin was still large.

Machine learning provided a different approach to solving computer vision problems. With machine learning, developers no longer needed to manually code every single rule into their vision applications. Instead they programmed "features," smaller applications that could detect specific patterns in images. They then used a statistical learning algorithm such as linear regression, logistic regression, decision trees or support vector machines (SVM) to detect patterns and classify images and detect objects in them.

Machine learning helped solve many problems that were historically challenging for classical software development tools and approaches. For instance, years ago, machine learning engineers were able to create a software that could predict breast cancer survival windows better than human experts. However, building the features of the software required the efforts of dozens of engineers and breast cancer experts and took a lot of time develop.
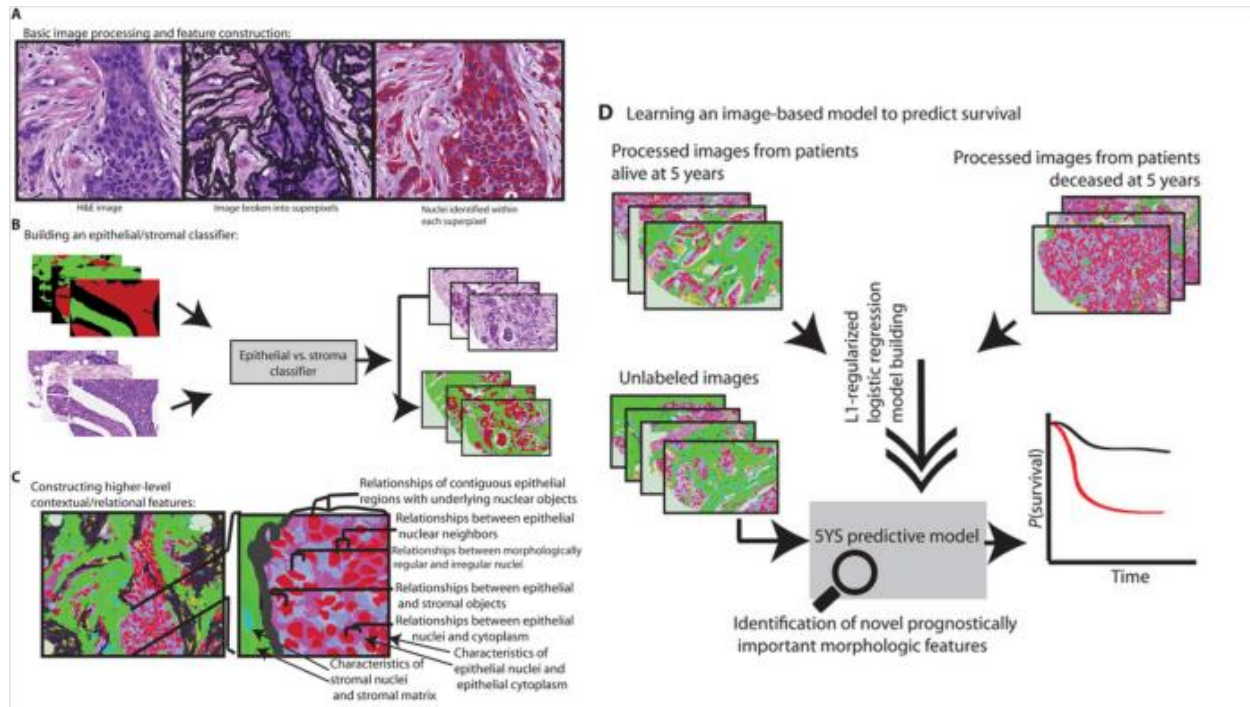
Deep learning provided a fundamentally different approach to doing machine learning. Deep learning relies on neural networks, a general-purpose function that can solve any problem representable through examples. When you provide a neural network with many labeled examples of a specific kind of data, it'll be able to extract common patterns between those examples and transform it into a mathematical equation that will help classify future pieces of information.

Most of current computer vision applications such as cancer detection, self-driving cars and facial recognition make use of deep learning. Deep learning and deep neural networks have moved from the conceptual realm into practical applications thanks to availability and advances in hardware and cloud computing resources.

How Long Does It Take To Decipher An Image:

The AI industry is cooking along, and experiments that not long ago took weeks to run might take 15 minutes today. And for many real-world applications of computer vision, this process all happens continuously in microseconds, so that a computer today is able to be what scientists call "situationally aware."

CV In Self-Driving Cars

Computer vision enables self-driving cars to make sense of their surroundings. Cameras capture video from different angles around the car and feed it to computer vision software, which then processes the images in real-time to find the extremities of roads, read traffic signs, detect other cars, objects and pedestrians. The self-driving car can then steer its way on streets and highways, avoid hitting obstacles, and (hopefully) safely drive its passengers to their destination.

CV In Facial Recognition

Computer vision also plays an important role in facial recognition applications, the technology that enables computers to match images of people's faces to their identities. Computer vision algorithms detect facial features in images and compare them with databases of face profiles. Consumer devices use facial recognition to authenticate the identities of their owners. Social media apps use facial recognition to detect and tag users. Law enforcement agencies also rely on facial recognition technology to identify criminals in video feeds.

CV In Augmented Reality & Mixed Reality

Computer vision also plays an important role in augmented and mixed reality, the technology that enables computing devices such as smartphones, tablets and smart glasses to overlay and embed virtual objects on real world imagery. Using computer vision, AR gear detect objects in real world in order to determine the locations on a device's display to place a virtual object. For instance, computer vision algorithms can help AR applications detect planes such as tabletops, walls and floors, a very important part of establishing depth and dimensions and placing virtual objects in physical world.

CV In Healthcare

Computer vision has also been an important part of advances in health-tech. Computer vision algorithms can help automate tasks such as detecting cancerous moles in skin images or finding symptoms in x-ray and MRI scans.
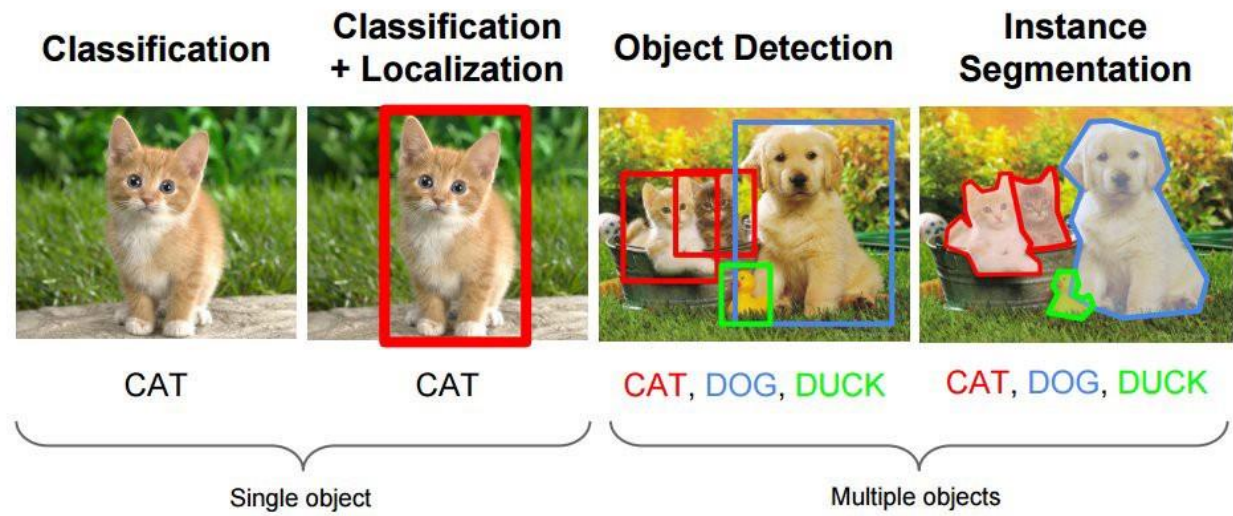
Challenges of Computer Vision

Helping computers to see turns out to be very hard.

Inventing a machine that sees like we do is a deceptively difficult task, not just because it's hard to make computers do it, but because we're not entirely sure how human vision works in the first place.

Studying biological vision requires an understanding of the perception organs like the eyes, as well as the interpretation of the perception within the brain. Much progress has been made, both in charting the process and in terms of discovering the tricks and shortcuts used by the system, although like any study that involves the brain, there is a long way to go.

Computer Vision Tasks

Classification | Classification + Localization | Object Detection | Instance Segmentation

CAT | CAT | CAT, DOG, DUCK | CAT, DOG, DUCK

Single object | Multiple objects

Examples of image processing include:

1. Normalizing photometric properties of the image, such as brightness or color.
2. Cropping the bounds of the image, such as centering an object in a photograph.
3. Removing digital noise from an image, such as digital artifacts from low light levels.

Many popular computer vision applications involve trying to recognize things in photographs; for example:

The goal of computer vision is to understand the content of digital images. Typically, this involves developing methods that attempt to reproduce the capability of human vision.

Object Classification: What broad category of object is in this photograph?

Object Identification: Which type of a given object is in this photograph?

Object Verification: Is the object in the photograph?

Object Detection: Where are the objects in the photograph?

Object Landmark Detection: What are the key points for the object in the photograph?

Object Segmentation: What pixels belong to the object in the image?

Object Recognition: What objects are in this photograph and where are they?

Outside of just recognition, other methods of analysis include:

Video motion analysis uses computer vision to estimate the velocity of objects in a video, or the camera itself.

In image segmentation, algorithms partition images into multiple sets of views.

Scene reconstruction creates a 3D model of a scene inputted through images or video.

In image restoration, noise such as blurring is removed from photos using Machine Learning based filters.

Any other application that involves understanding pixels through software can safely be labeled as computer vision.

The 2010 textbook on computer vision titled "Computer Vision: Algorithms and Applications" provides a list of some high-level problems where we have seen success with computer vision.

- Optical character recognition (OCR)
- Machine inspection
- Retail (e.g. automated checkouts)
- 3D model building (photogrammetry)
- Medical imaging
- Automotive safety
- Match move (e.g. merging CGI with live actors in movies)
- Motion capture (mocap)
- Surveillance
- Fingerprint recognition and biometrics