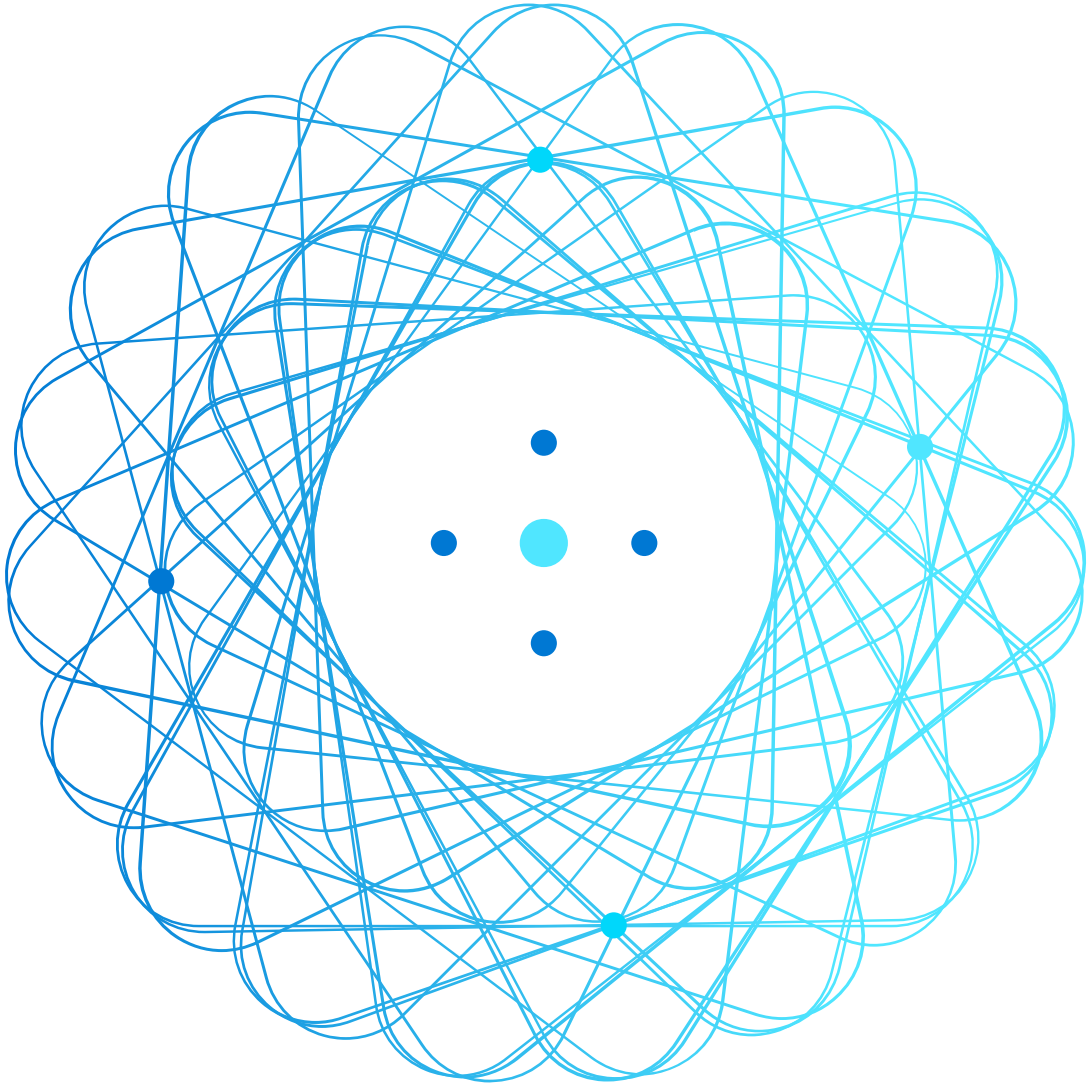
Microsoft Azure

Explore Fundamentals of Computer Vision



Agenda

Introduction

Image & image Processing

Machine Learning for Computer Vision

Azure AI Vision

Introduction

- *Computer vision* is one of the core areas of artificial intelligence (AI), and focuses on creating solutions that enable AI applications to "see" the world and make sense of it.

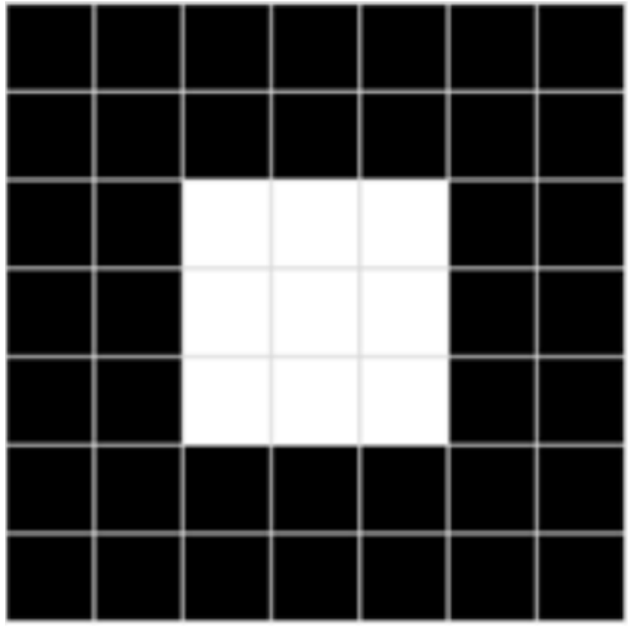
- Computers don't have biological eyes that work the way ours do, but they're capable of processing images; either from a live camera feed or from digital photographs or videos. This ability to process images is the key to creating software that can emulate human visual perception

Images & Image Vision

Images as pixel arrays

```
0   0   0   0   0   0   0
0   0   0   0   0   0   0
0   0 255 255 255   0   0
0   0 255 255 255   0   0
0   0 255 255 255   0   0
0   0   0   0   0   0   0
0   0   0   0   0   0   0
```

The array consists of seven rows and seven columns, representing the pixel values for a 7x7 pixel image (which is known as the image's *resolution*). Each pixel has a value between 0 (black) and 255 (white); with values between these bounds representing shades of gray. The image represented by this array looks similar to the following (magnified) image

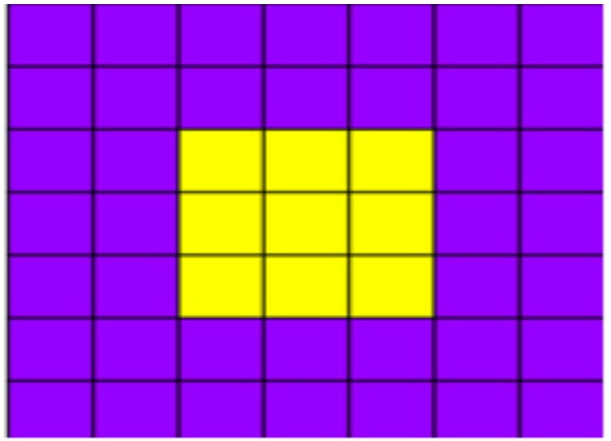


Example

```
Red:
 150  150  150  150  150  150  150
 150  150  150  150  150  150  150
 150  150  255  255  255  150  150
 150  150  255  255  255  150  150
 150  150  255  255  255  150  150
 150  150  150  150  150  150  150
 150  150  150  150  150  150  150

Green:
 0    0    0    0    0    0    0
 0    0    0    0    0    0    0
 0    0   255  255  255   0    0
 0    0   255  255  255   0    0
 0    0   255  255  255   0    0
 0    0    0    0    0    0    0
 0    0    0    0    0    0    0

Blue:
 255  255  255  255  255  255  255
 255  255  255  255  255  255  255
 255  255   0    0    0   255  255
 255  255   0    0    0   255  255
 255  255   0    0    0   255  255
 255  255  255  255  255  255  255
 255  255  255  255  255  255  255
```



The purple squares are represented by the combination:

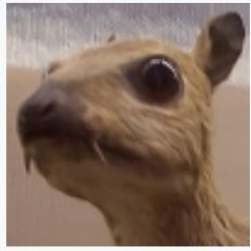
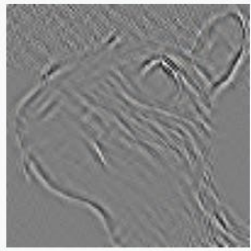
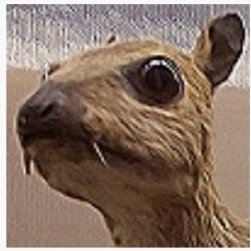
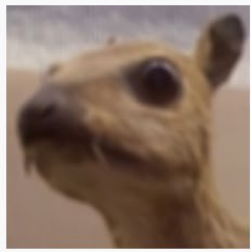
```
Red: 150
Green: 0
Blue: 255
```

The yellow squares in the center are represented by the combination:

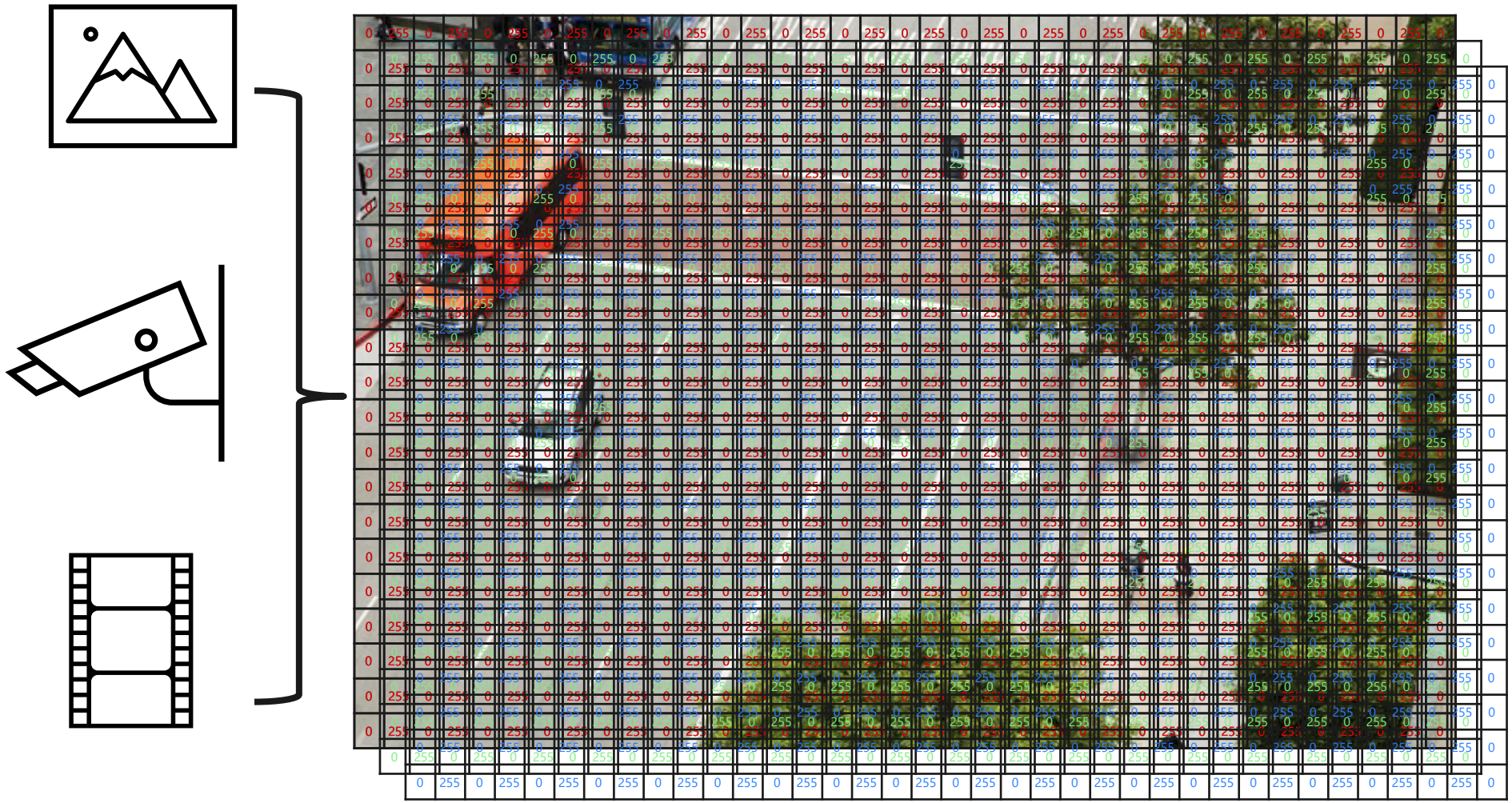
```
Red: 255
Green: 255
Blue: 0
```

Using Filters to Process Image

https://learn.microsoft.com/en-us/training/modules/analyze-images-computer-vision/2-understand-computer-vision

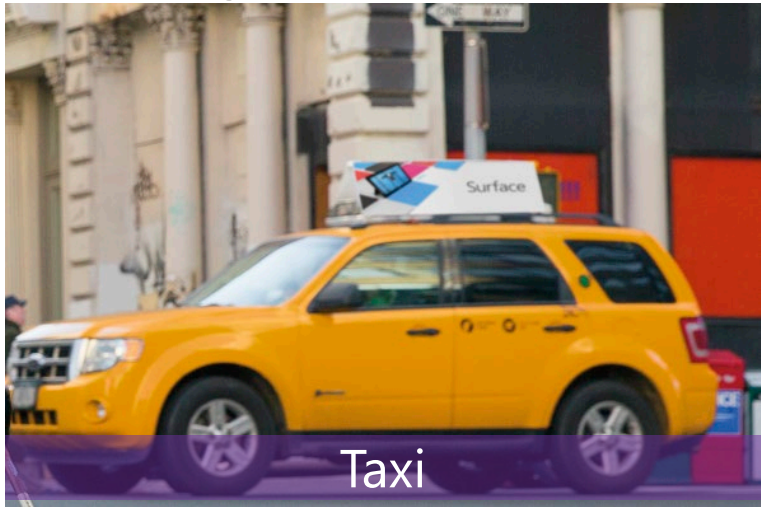
Operation	Kernel ω	Image resu g(x,y)
Identity	$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$	
Ridge or edge detection	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
	$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$	
Sharpen	$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$	
Box blur (normalized)	$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$	

What is Computer Vision?

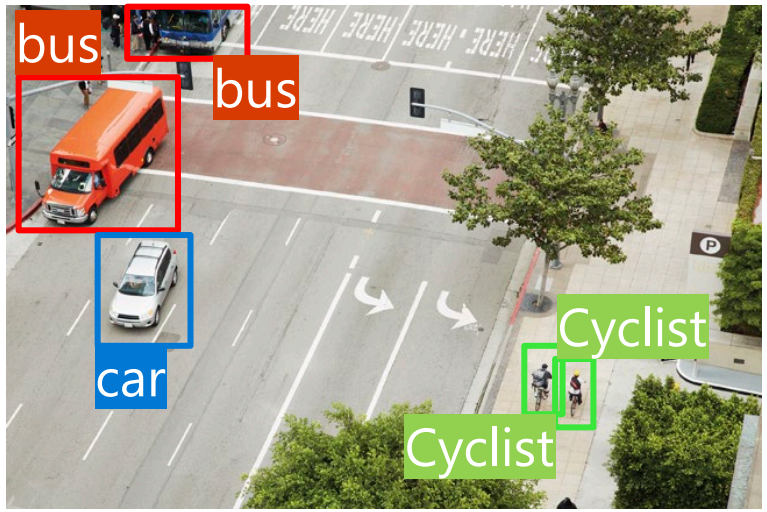


Applications of Computer Vision

Image Classification



Object Detection



Semantic Segmentation

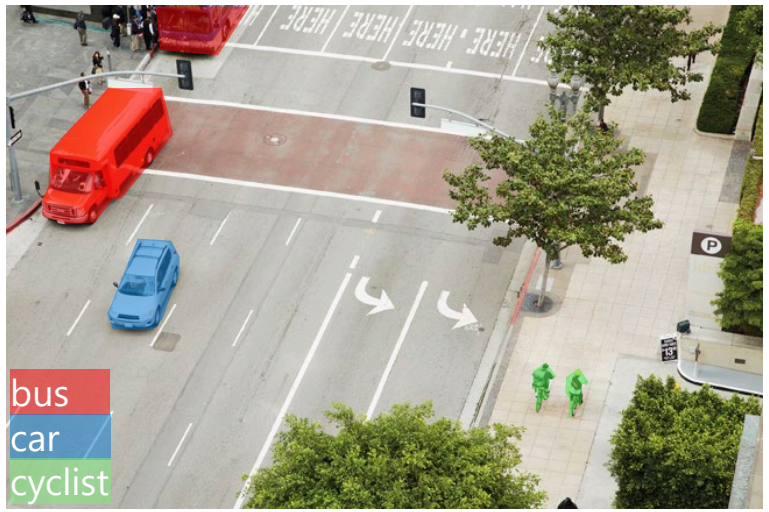


Image Analysis

Face Detection & Recognition

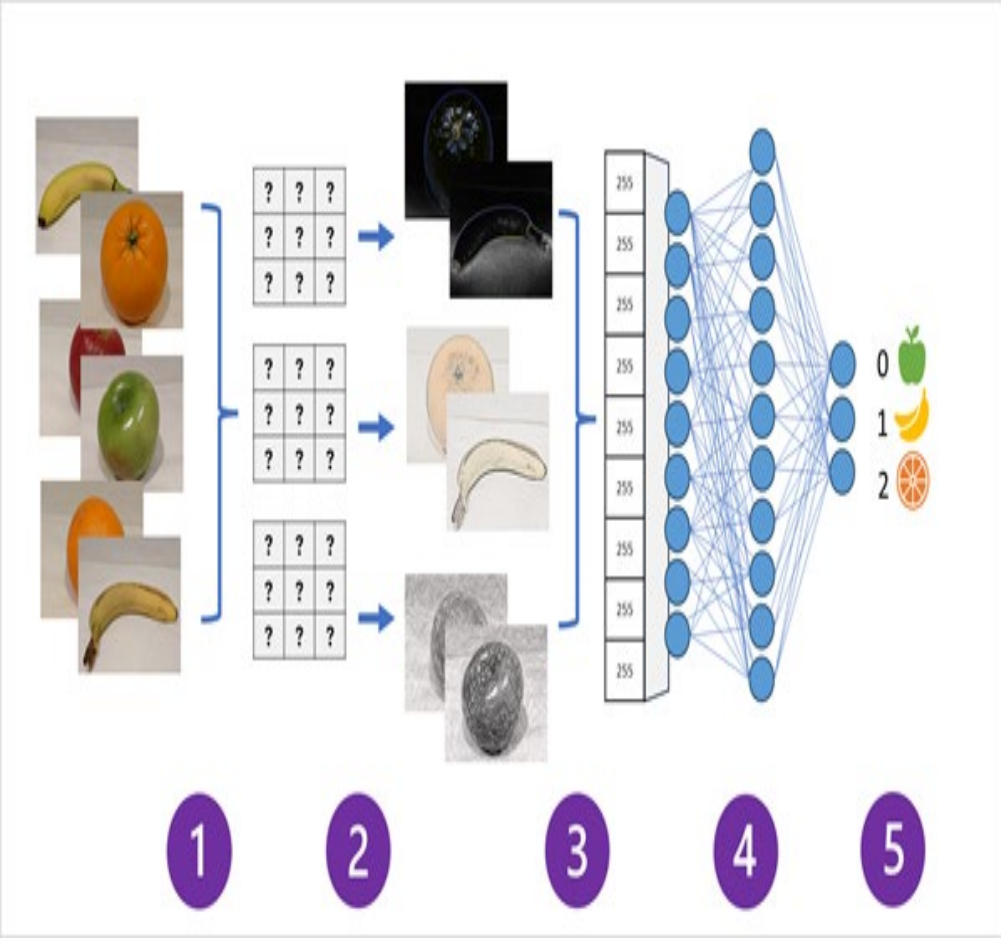
Optical Character Recognition

Machine learning for computer vision

The ability to use filters to apply effects to images is useful in image processing tasks, such as you might perform with image editing software. However, the goal of computer vision is often to extract meaning, or at least actionable insights, from images; which requires the creation of machine learning models that are trained to recognize features based on large volumes of existing images.

Convolutional neural networks (CNNs)

One of the most common machine learning model architectures for computer vision is a *convolutional neural network* (CNN), a type of deep learning architecture. CNNs use filters to extract numeric feature maps from images, and then feed the feature values into a deep learning model to generate a label prediction. For example, in an *image classification* scenario, the label represents the main subject of the image (in other words, what is this an image of?). You might train a CNN model with images of different kinds of fruit (such as apple, banana, and orange) so that the label that is predicted is the type of fruit in a given image.
During the *training* process for a CNN, filter kernels are initially defined using randomly generated weight values. Then, as the training process progresses, the models predictions are evaluated against known label values, and the filter weights are adjusted to improve accuracy. Eventually, the trained fruit image classification model uses the filter weights that best extract features that help identify different kinds of fruit.



1.Images with known labels (for example, 0: apple, 1: banana, or 2: orange) are fed into the network to train the model.
2.One or more layers of filters is used to extract features from each image as it is fed through the network. The filter kernels start with randomly assigned weights and generate arrays of numeric values called *feature maps*.
3.The feature maps are flattened into a single dimensional array of feature values.
4.The feature values are fed into a fully connected neural network.
5.The output layer of the neural network uses a *softmax* or similar function to produce a result that contains a probability value for each possible class, for example [0.2, 0.5, 0.3].
During training the output probabilities are compared to the actual class label - for example, an image of a banana (class 1) should have the value [0.0, 1.0, 0.0]. The difference between the predicted and actual class scores is used to calculate the *loss* in the model, and the weights in the fully connected neural network and the filter kernels in the feature extraction layers are modified to reduce the loss.
The training process repeats over multiple *epochs* until an optimal set of weights has been learned. Then, the weights are saved and the model can be used to predict labels for new images for which the label is unknown.

Transformers and multi-modal models

https://learn.microsoft.com/en-us/training/modules/analyze-images-computer-vision/2b-computer-vision-models

Azure AI Vision

While you can train your own machine learning models for computer vision, the architecture for computer vision models can be complex; and you require significant volumes of training images and compute power to perform the training process.
Microsoft's Azure AI Vision service provides prebuilt and customizable computer vision models that are based on the Florence foundation model and provide various powerful capabilities. With Azure AI Vision, you can create sophisticated computer vision solutions quickly and easily; taking advantage of "off-the-shelf" functionality for many common computer vision scenarios, while retaining the ability to create custom models using your own images.

To use Azure AI Vision, you need to create a resource for it in your Azure subscription. You can use either of the following resource types:
•**Azure AI Vision**: A specific resource for the Azure AI Vision service. Use this resource type if you don't intend to use any other Azure AI services, or if you want to track utilization and costs for your Azure AI Vision resource separately.
•**Azure AI services**: A general resource that includes Azure AI Vision along with many other Azure AI services; such as Azure AI Language, Azure AI Custom Vision, Azure AI Translator, and others. Use this resource type if you plan to use multiple AI services and want to simplify administration and development

Analyzing images with the Azure AI Vision service

https://learn.microsoft.com/en-us/training/modules/analyze-images-computer-vision/3-image-analysis-azure

Knowledege Check

Computer vision is based on the manipulation and analysis of what kinds of values in an image?

- Timestamp in Photograph metadata
- Pixels
- Image file names

Knowledege Check

Computer vision is based on the manipulation and analysis of what kinds of values in an image?

- Timestamp in Photograph metadata
- Pixels
  - Pixels are numeric values that represent shade intensity for points in the image
- Image file names

Knowledege Check

You want to use the Azure AI Vision service to analyze images. You also want to use the Azure AI Language service to analyze text. You want developers to require only one key and endpoint to access all of your services. What kind of resource should you create in your Azure subscription?

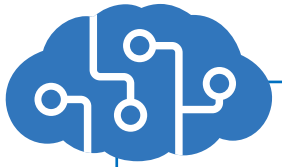
- Azure Ai vision
- Azure AI Services
- Azure Open AI service

Knowledege Check

You want to use the Azure AI Vision service to analyze images. You also want to use the Azure AI Language service to analyze text. You want developers to require only one key and endpoint to access all of your services. What kind of resource should you create in your Azure subscription?

- Azure Ai vision
- Azure AI Services
  - An Azure AI Services resource supports both Azure AI Vision and Azure AI Language.
- Azure Open AI service

Computer Vision Services in Azure

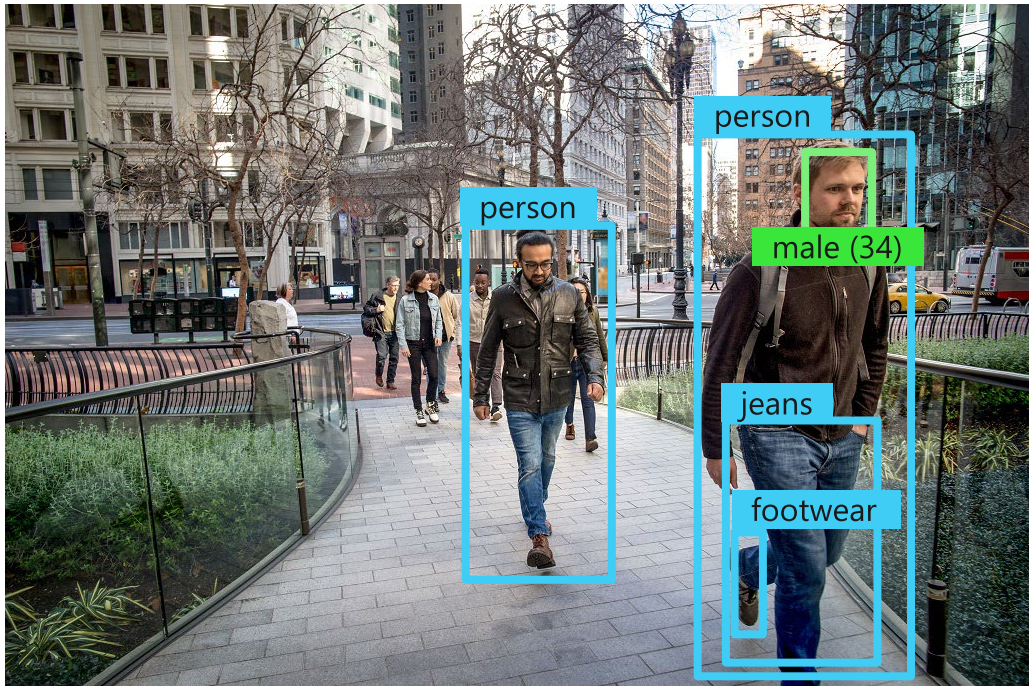


Computer Vision	• Image analysis – automated captioning and tagging • Common object detection • Face detection • Smart cropping • Optical character recognition
Custom Vision	• Custom image classification • Custom object detection
Face	• Face detection and analysis
Form Recognizer	• Data extraction from forms, invoices, and other documents

Creating Computer Vision solutions in Azure

Image Analysis with the *Computer Vision* Service

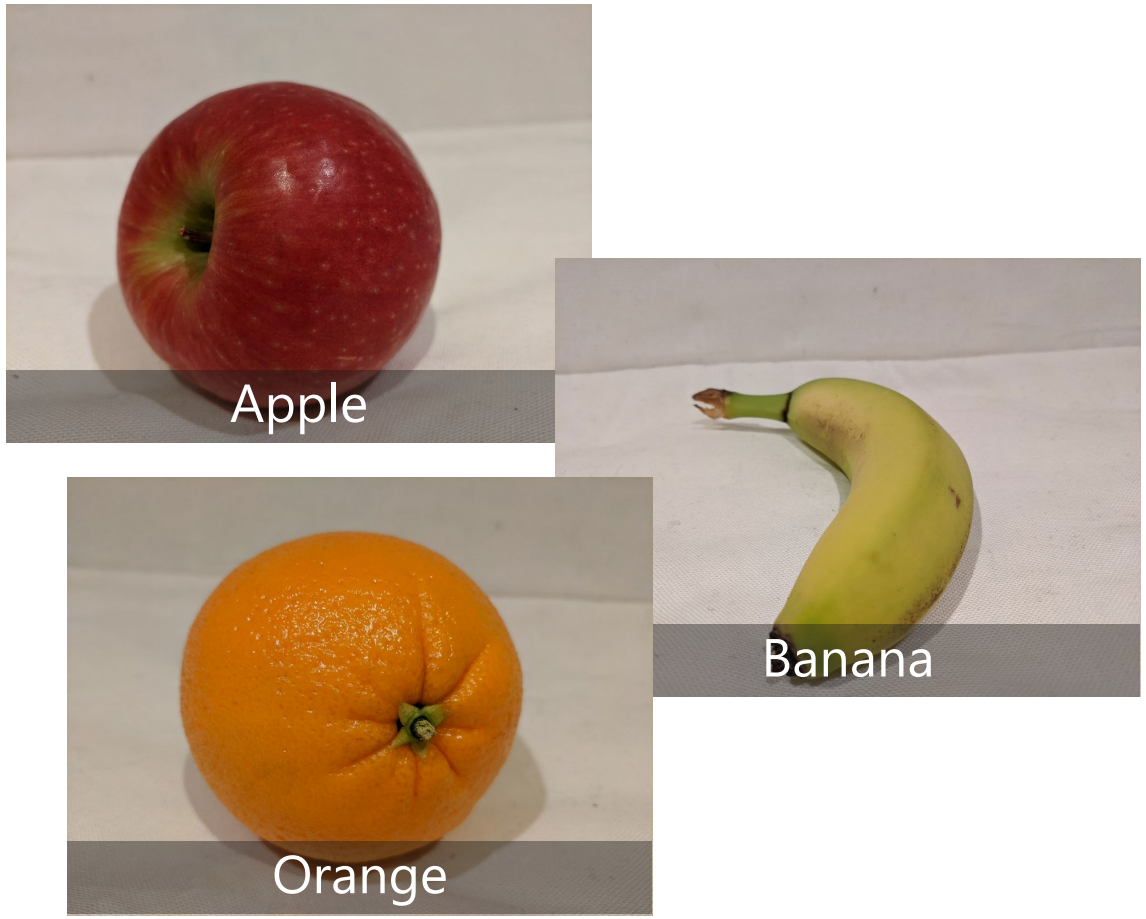
- Pre-trained computer vision model
- Object detection for over 10,000 predefined classes
- Image description and tag generation
- Face detection and analysis
- Content moderation
- Text detection and OCR



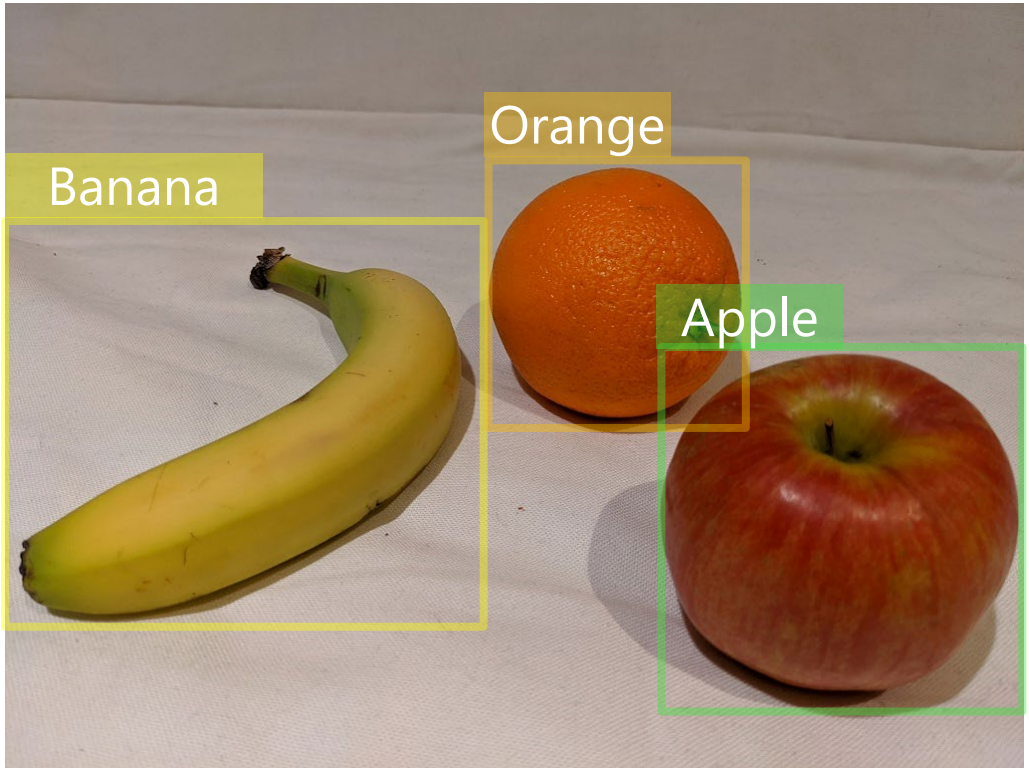
Caption: a group of people walking on a sidewalk
Tags: building, jeans, street, outdoor, jacket, city, person
Ratings: *Adult*: False, *Racy*: False, *Gore*: False

Training Models with the *Custom Vision* Service

Image Classification



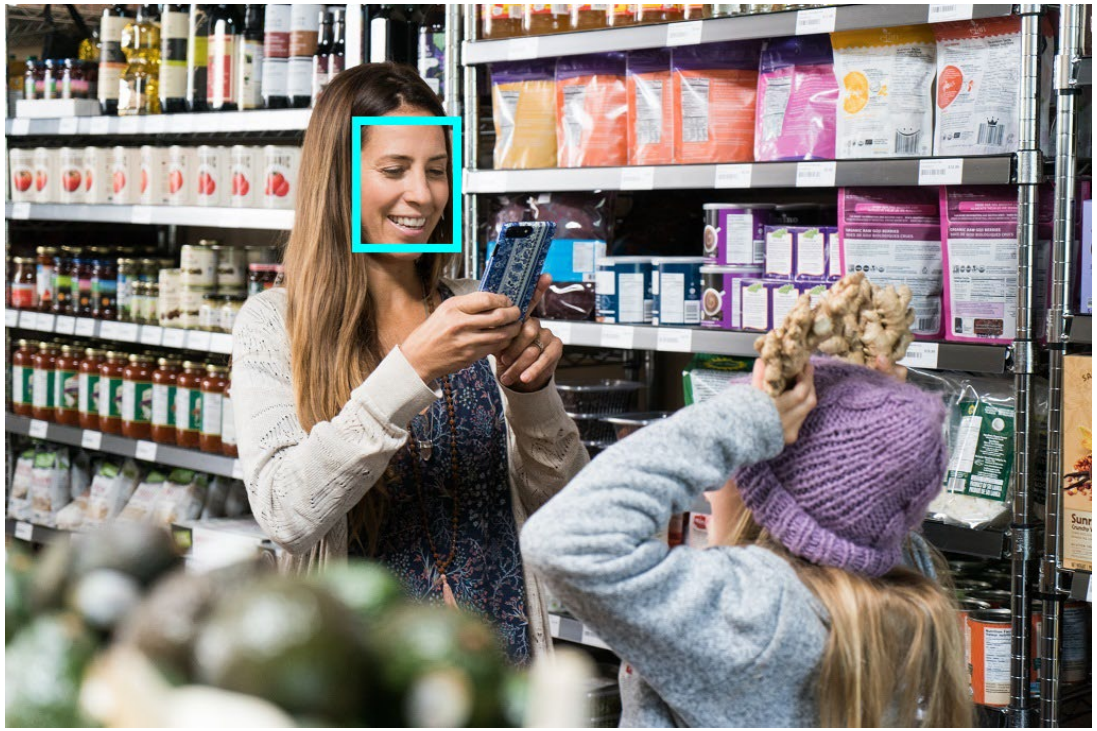
Object Detection



Detecting Faces with the *Face* Service

Anyone can use the Face service to detect:

- Blur
- Exposure
- Glasses
- Head pose
- Noise
- Occlusion



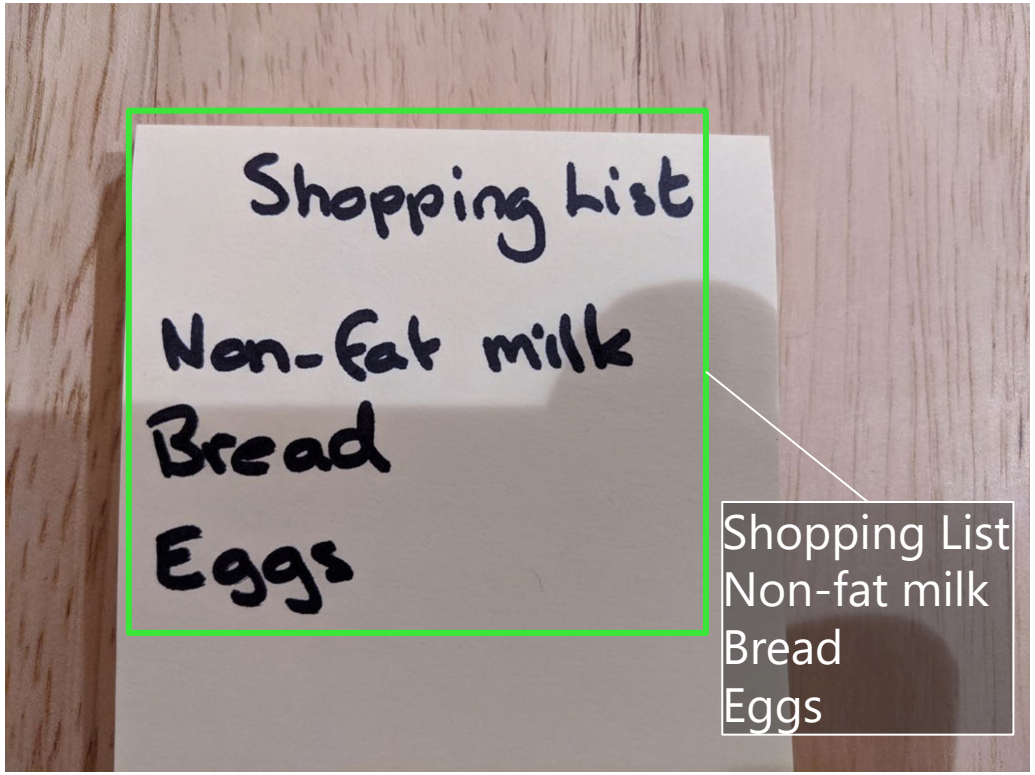
Only Managed Microsoft customers can access facial recognition capabilities:

- Similarity matching
- Identity verification

*To support Microsoft's Responsible AI Principles, Facial Recognition is under a Limited Access policy.

Reading Text with the *Computer Vision* Service

- Detect the location of text:
  - Printed
  - Handwritten
- Options for quick text extraction from images, or asynchronous analysis of larger scanned documents



Analyzing Forms with the *Form Recognizer* Service

- Extract information from scanned forms in image or PDF format
  - Use the pre-trained models for common document types
  - Train a custom model using your own forms
- Models perform *semantic recognition* of form fields – not just text extraction