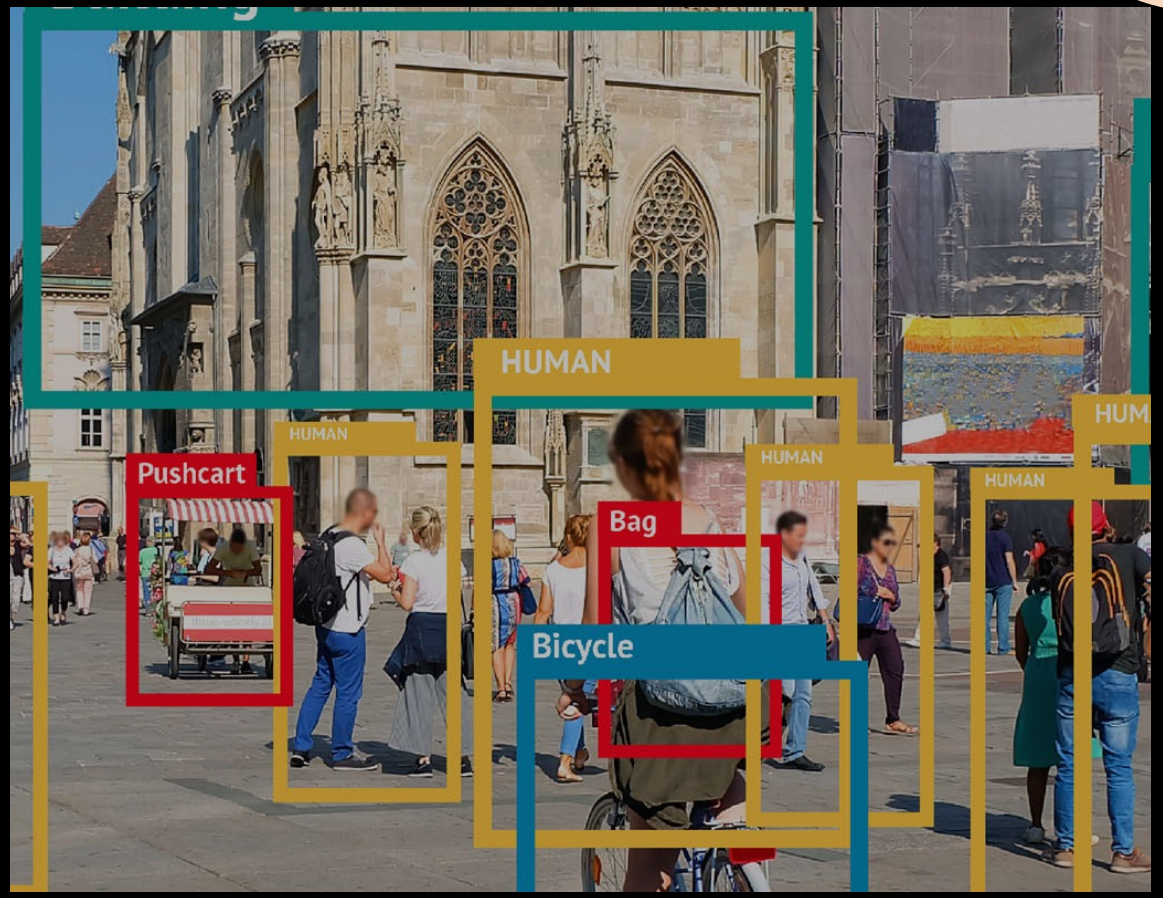
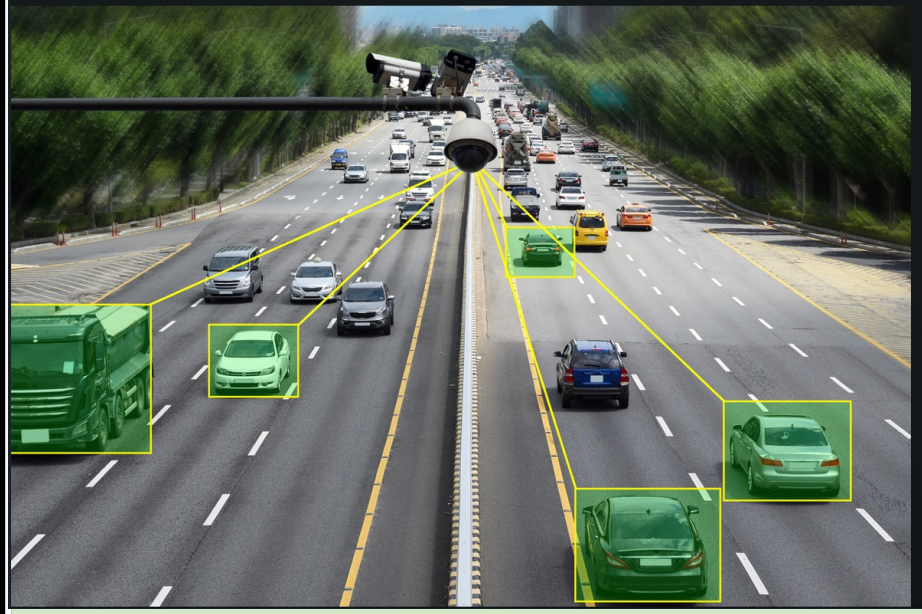
Chapter 9
Vision Systems



Agenda

- Computer Vision
- CNN
- Object Detection using YOLO Algorithm
- Azure Custom Vision for Object Detection

What is it Computer Vision ?



Computer vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos.

Question: What are the applications of Computer Vision?

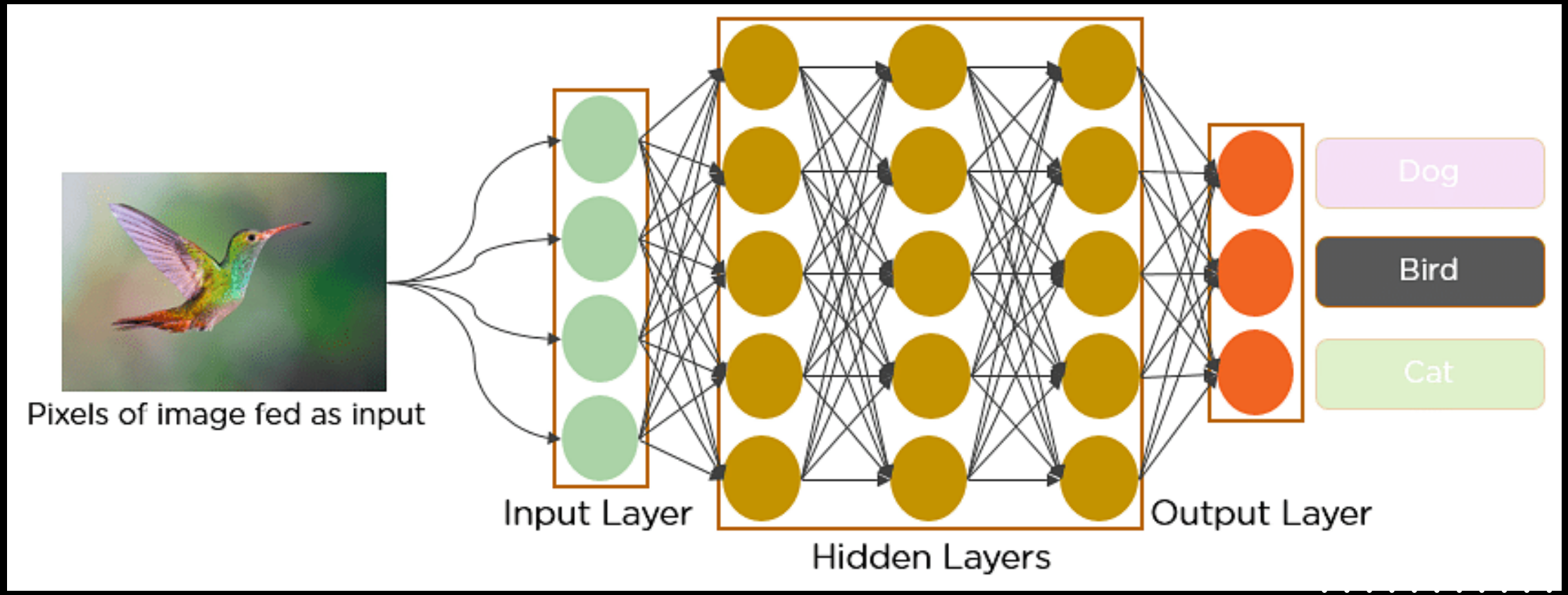
What is it Computer Vision ?

Different Computer Vision Tasks:

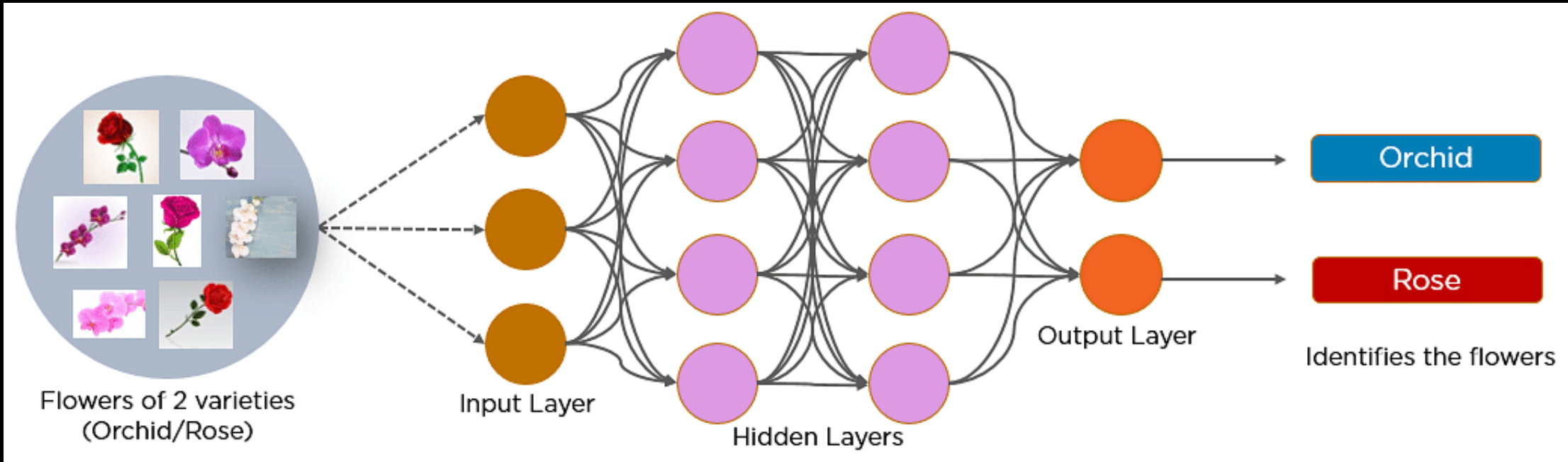
- Object Detection
- Image Classification
- Image Captioning
- Image Reconstruction

Part A: What are convolutional neural networks?

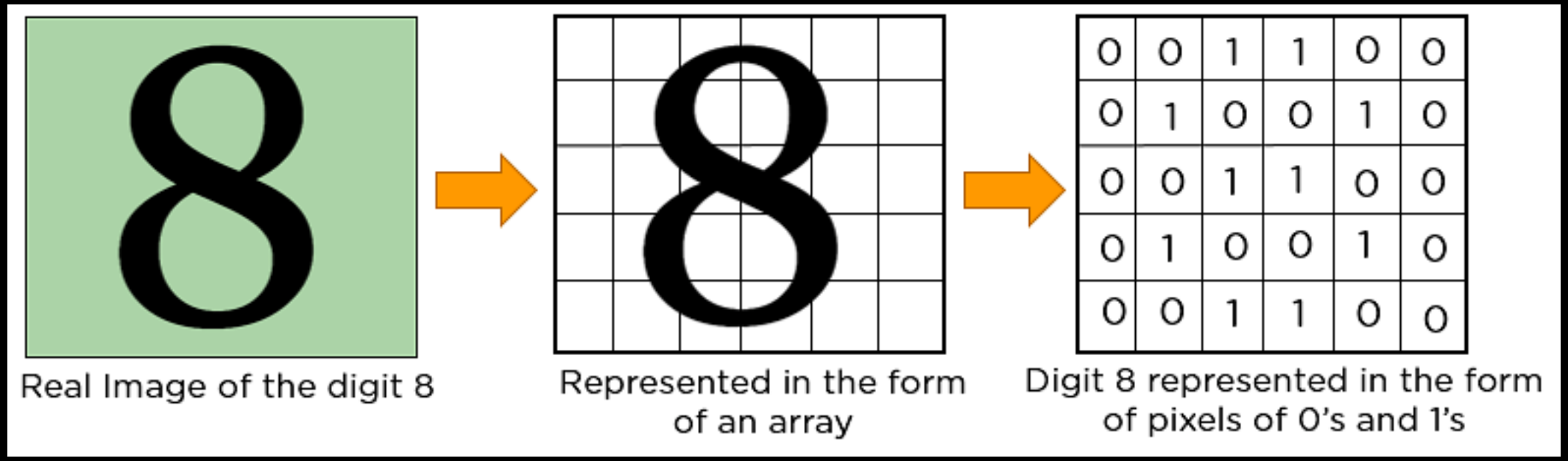
What is Convolutional Neural Network?



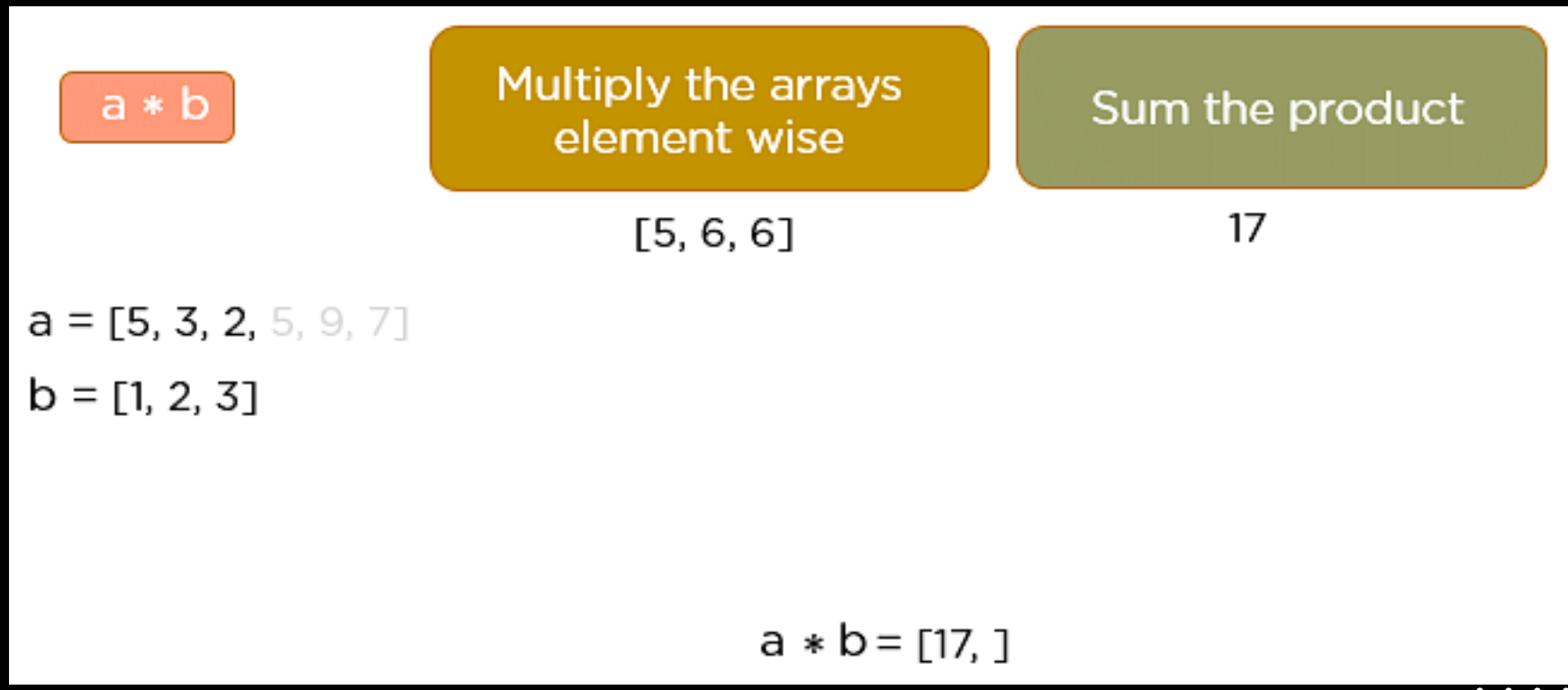
What is Convolutional Neural Network?



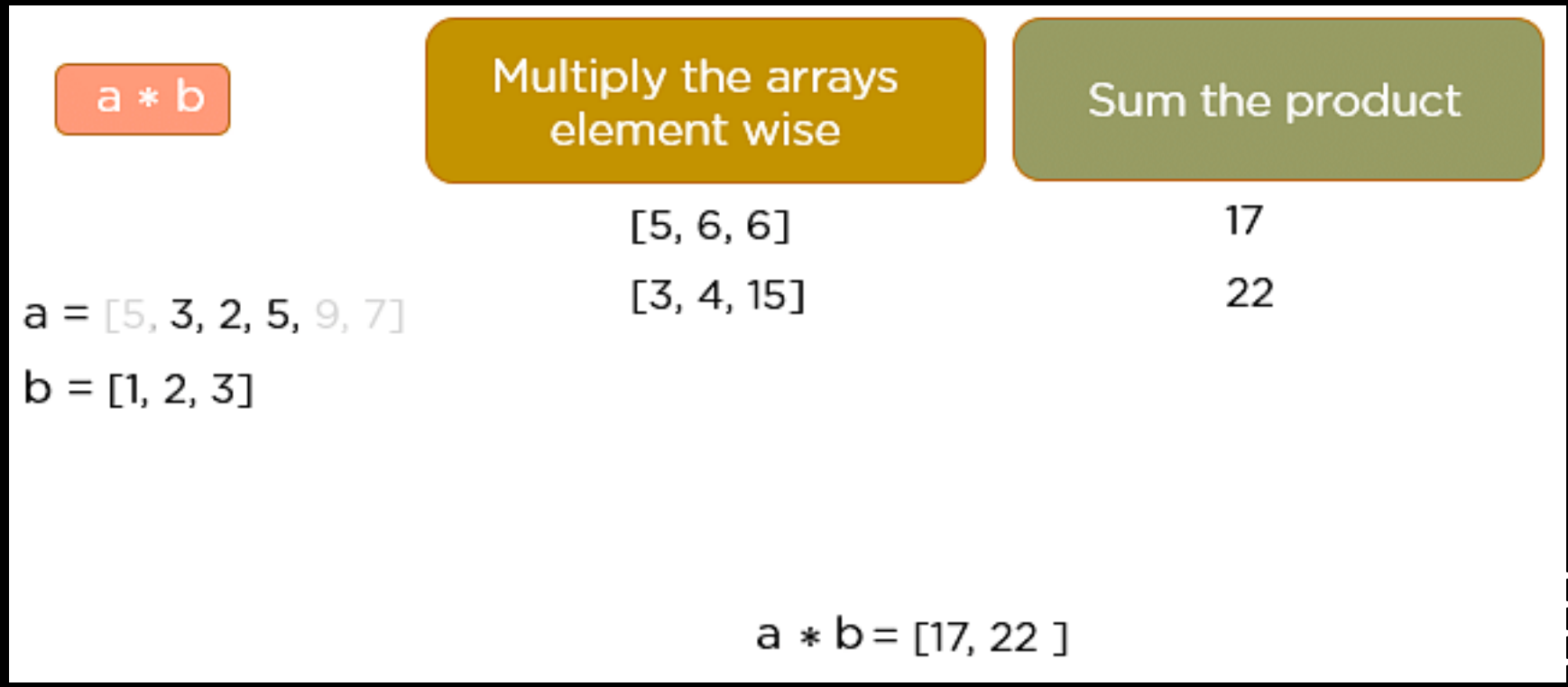
What is Convolutional Neural Network?



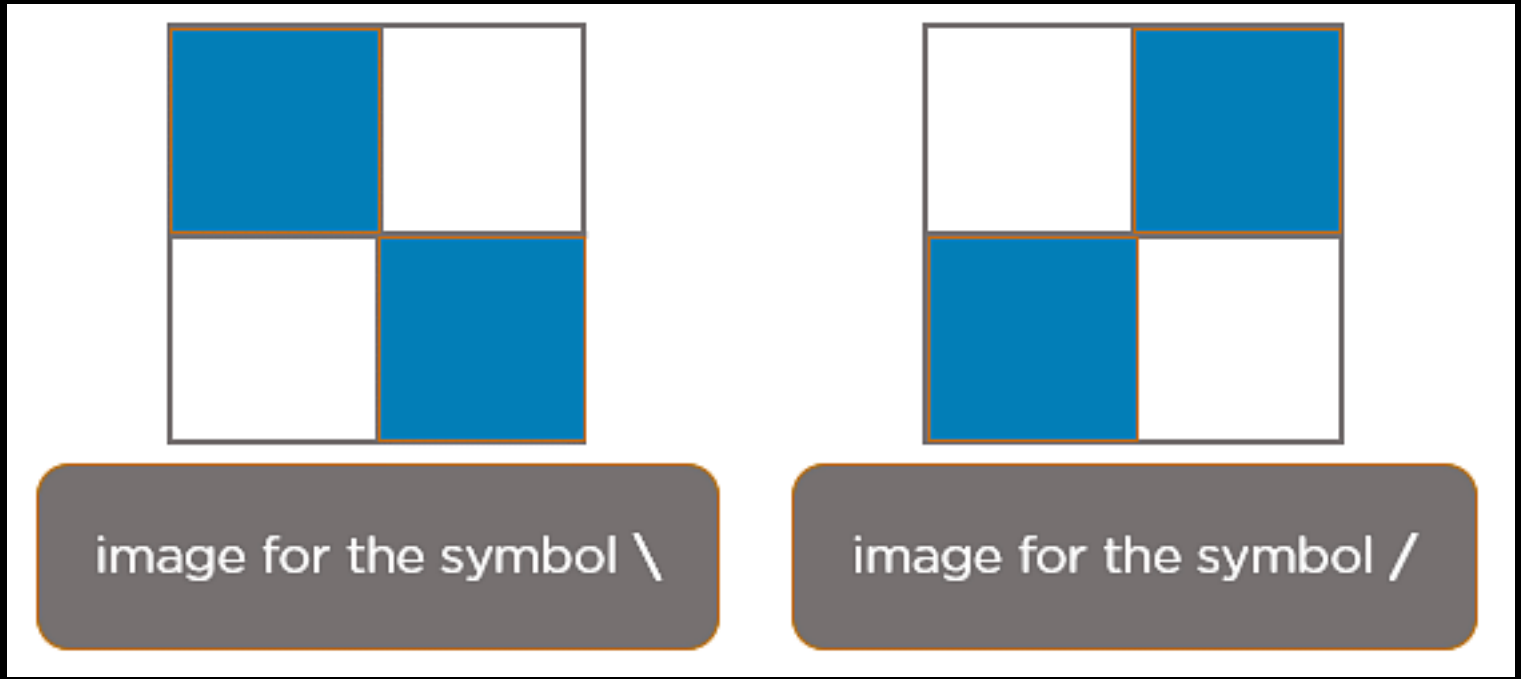
What is Convolutional Neural Network?



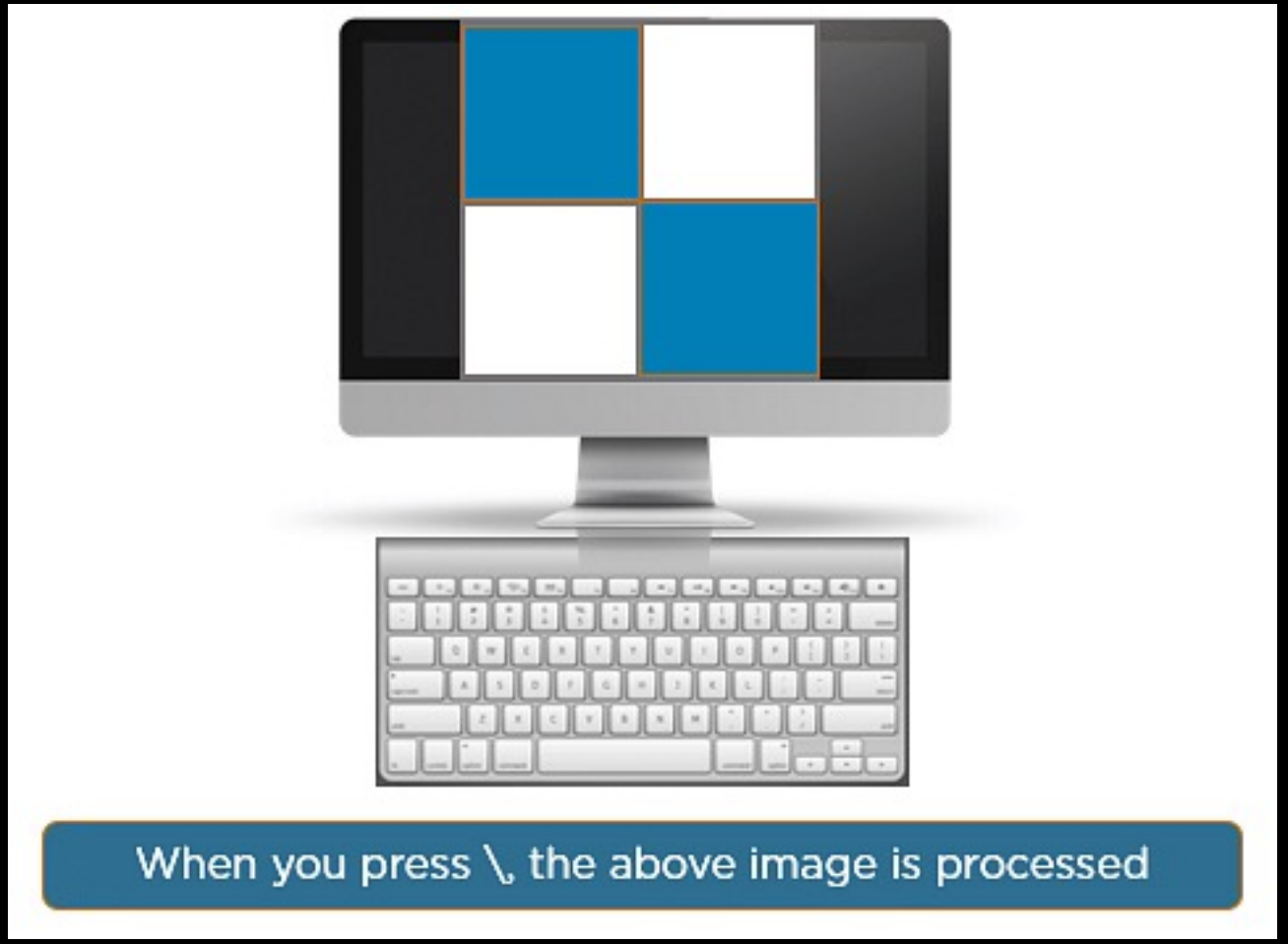
What is Convolutional Neural Network?



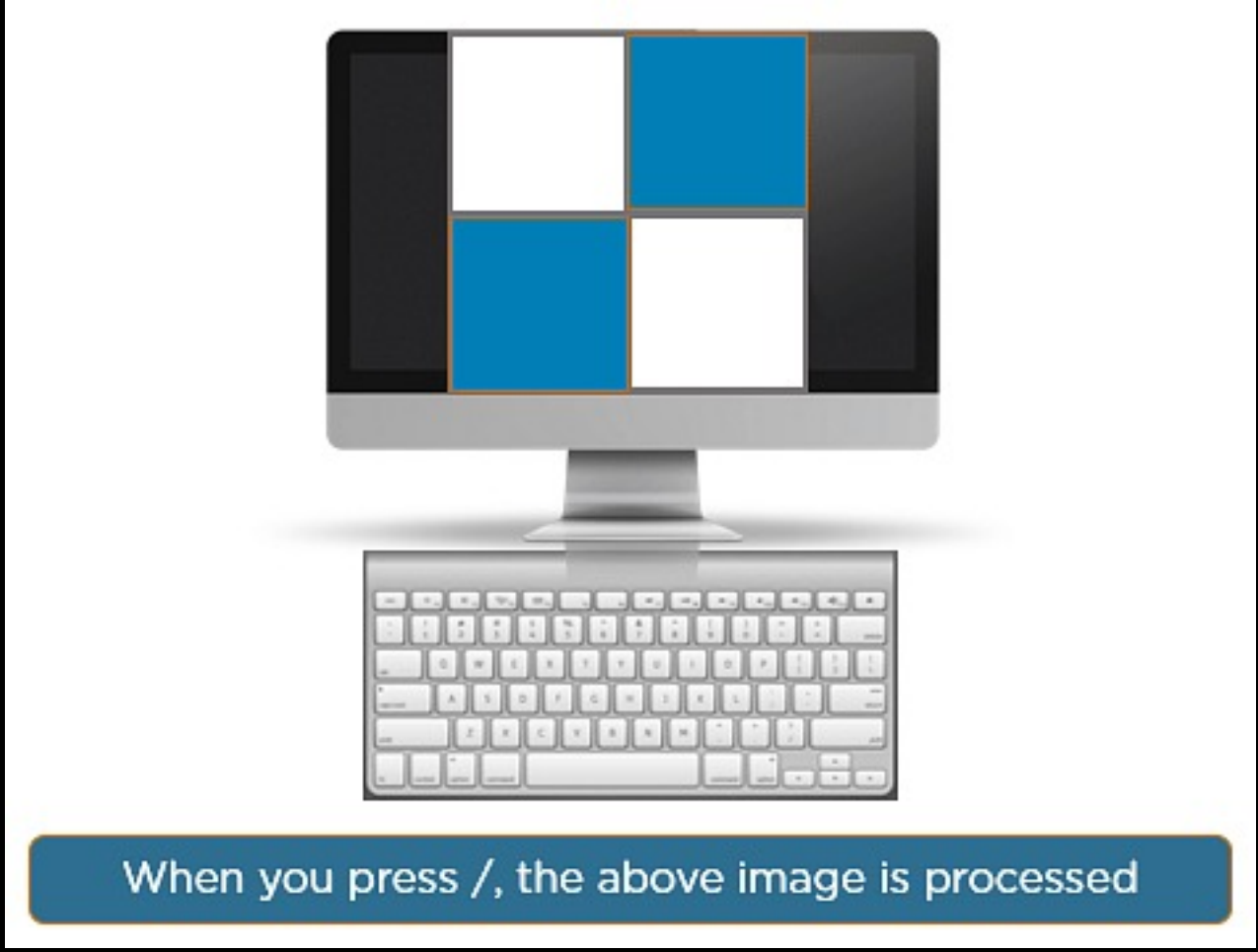
How Does CNN Recognize Images?



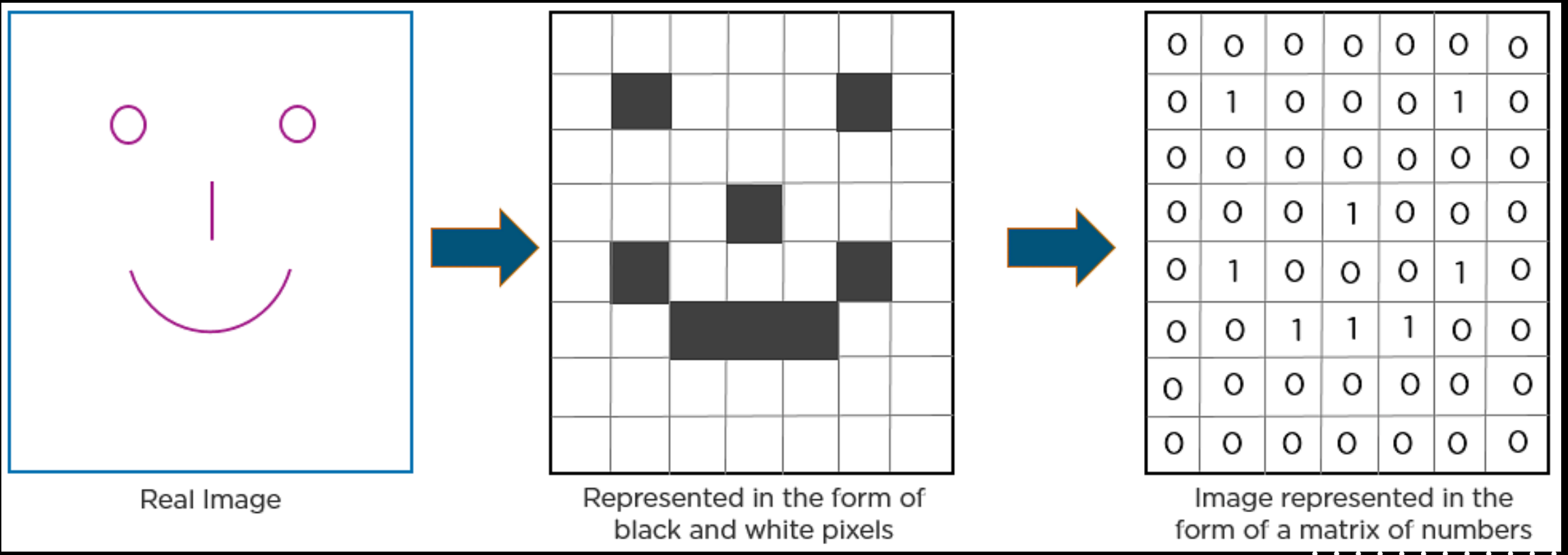
How Does CNN Recognize Images?



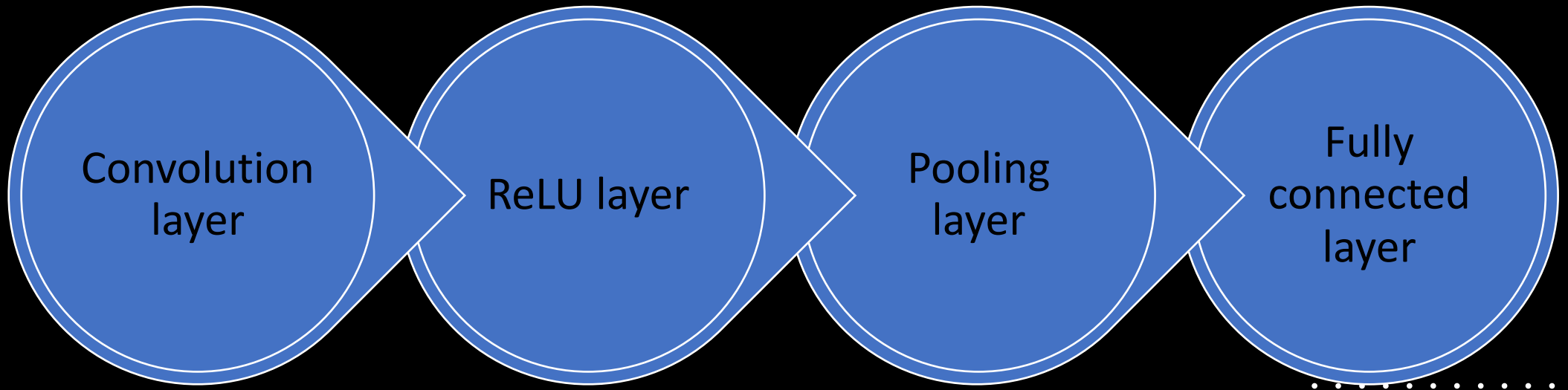
How Does CNN Recognize Images?



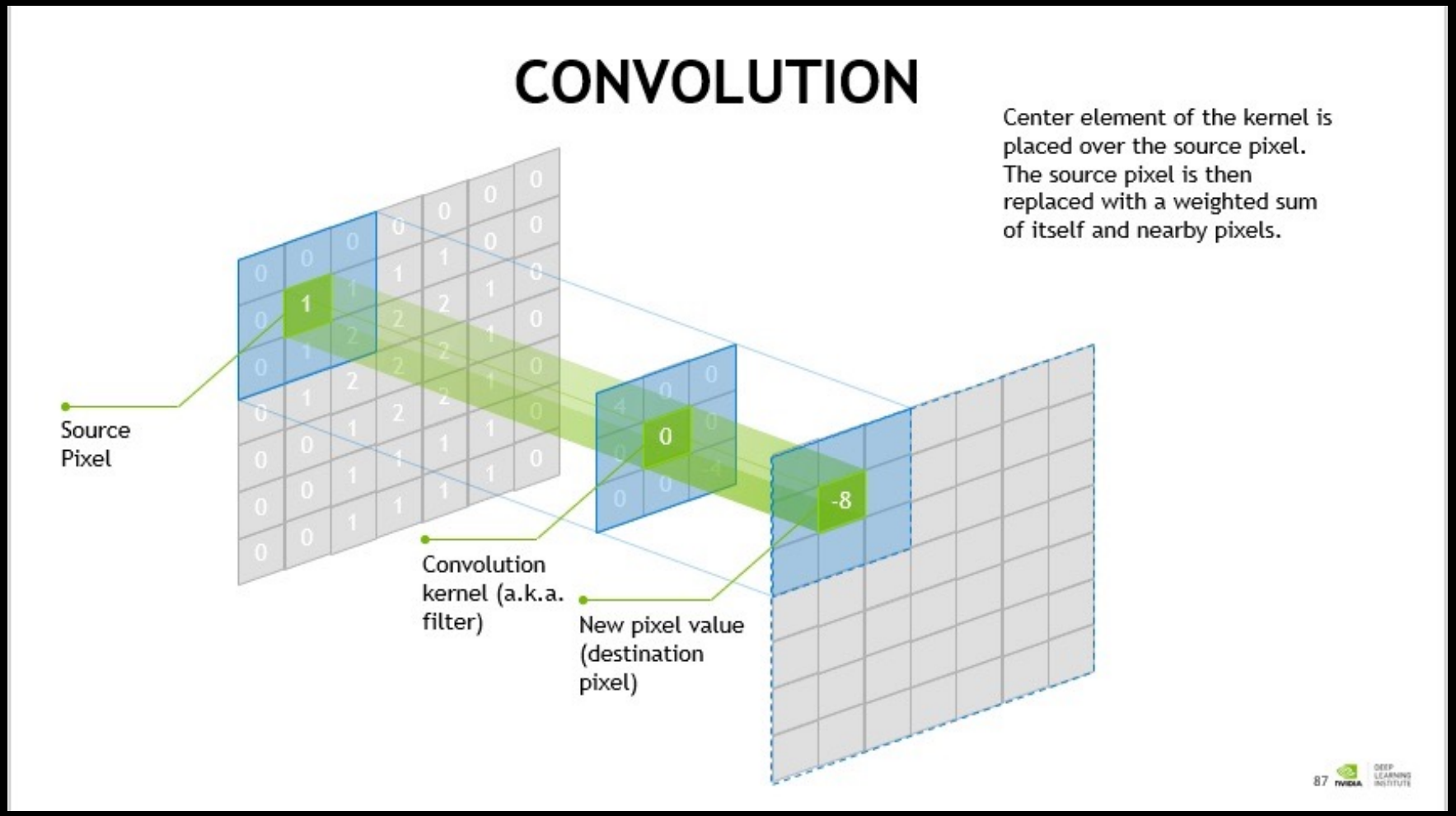
How Does CNN Recognize Images?



Layers in a Convolutional Neural Network

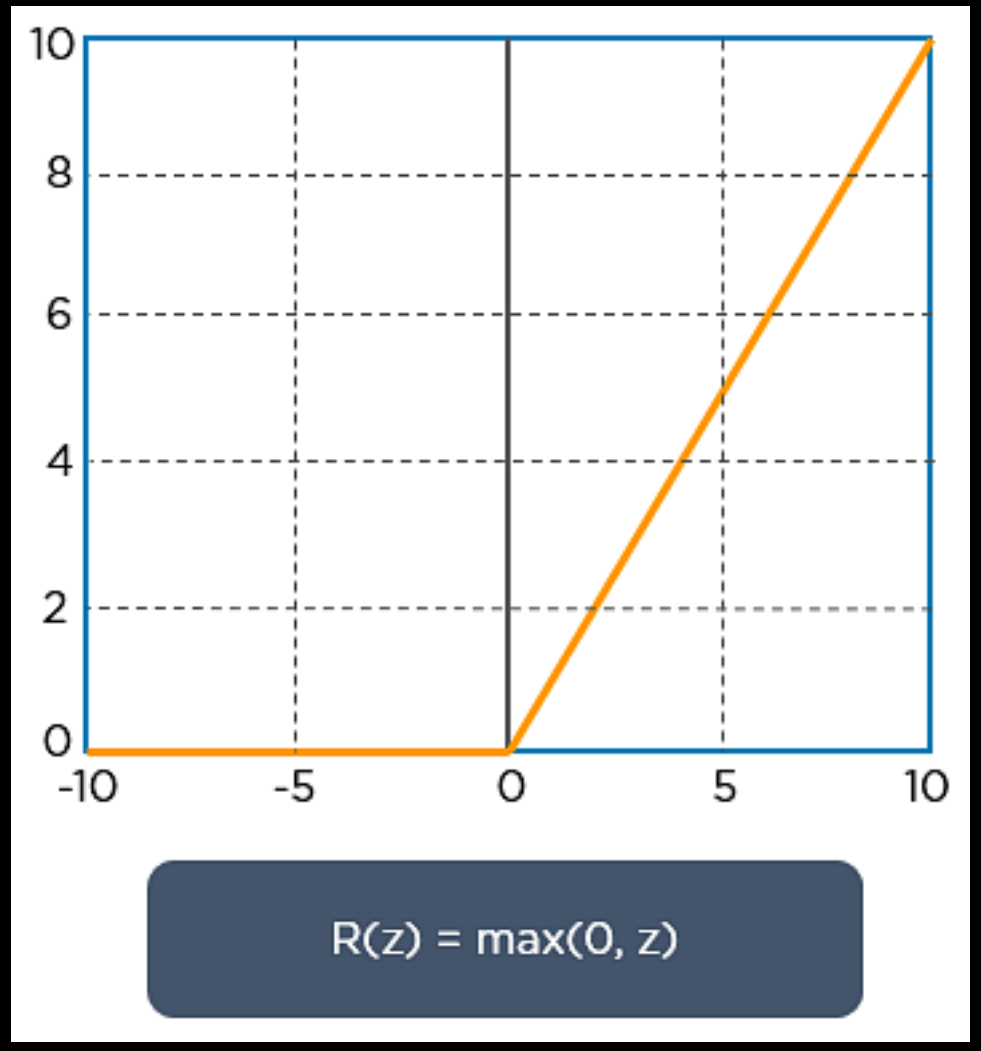


Convolution Layer

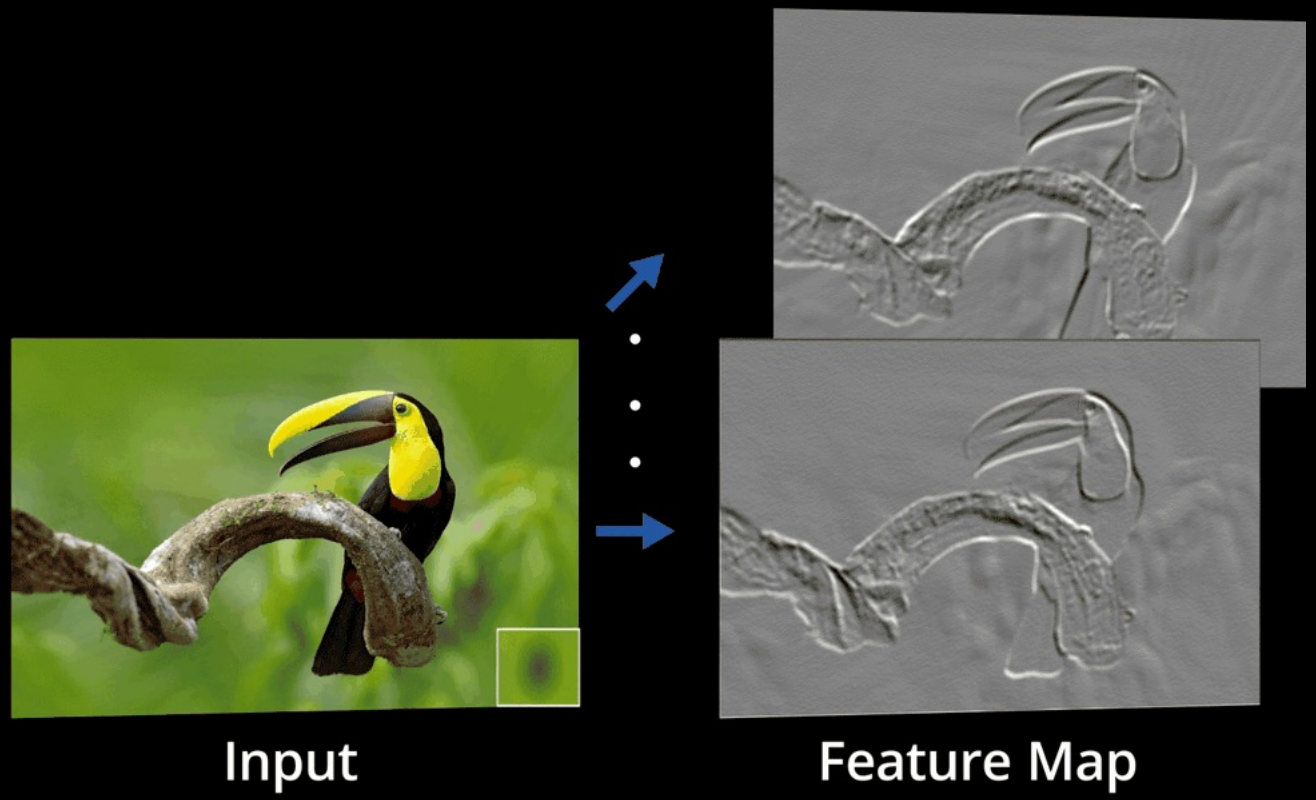


Convolutional layer (source: https://www.embedded-vision.com)

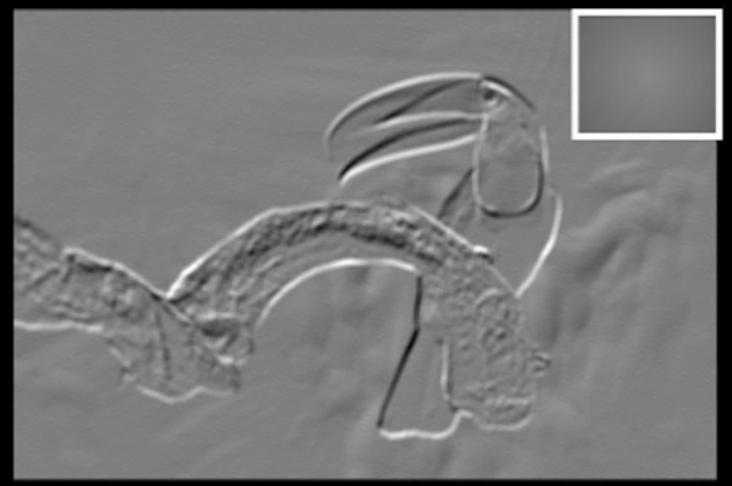
ReLU layer



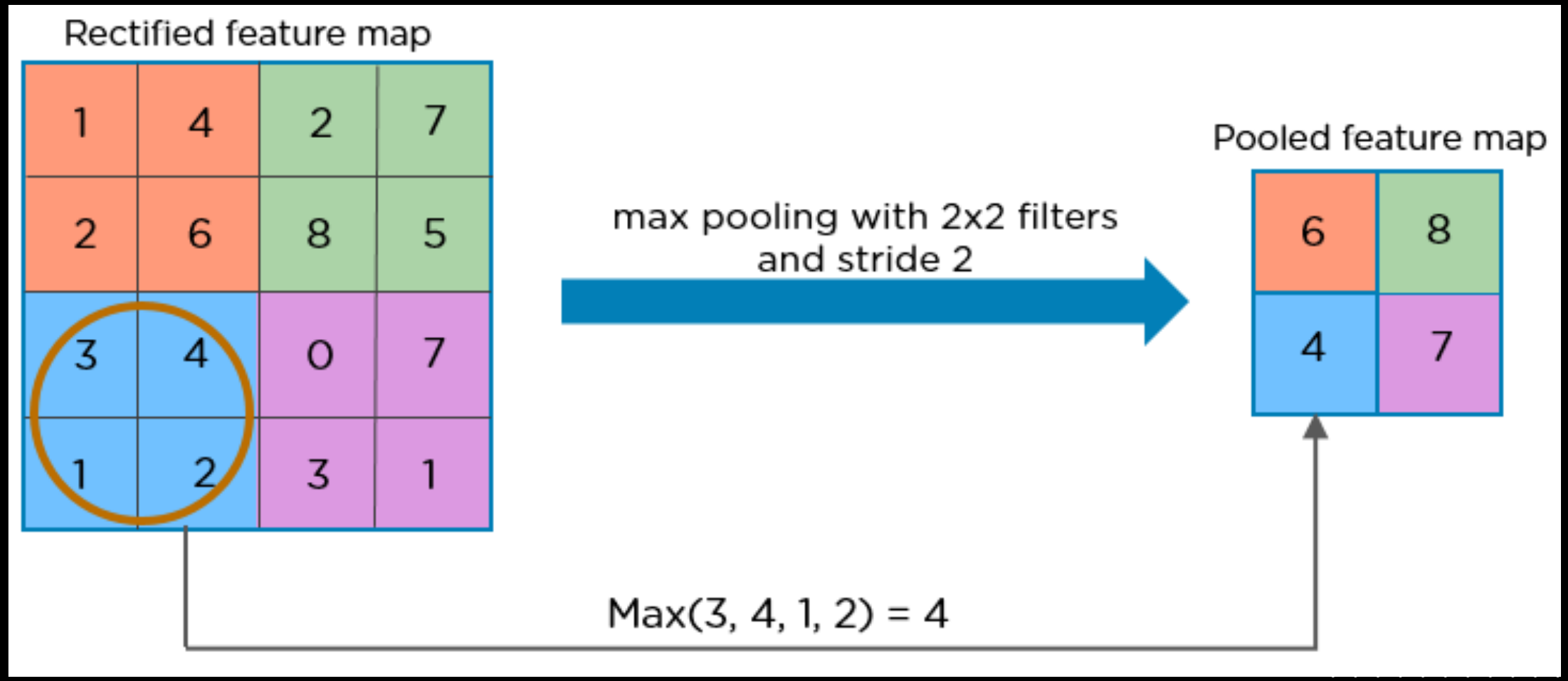
ReLU layer



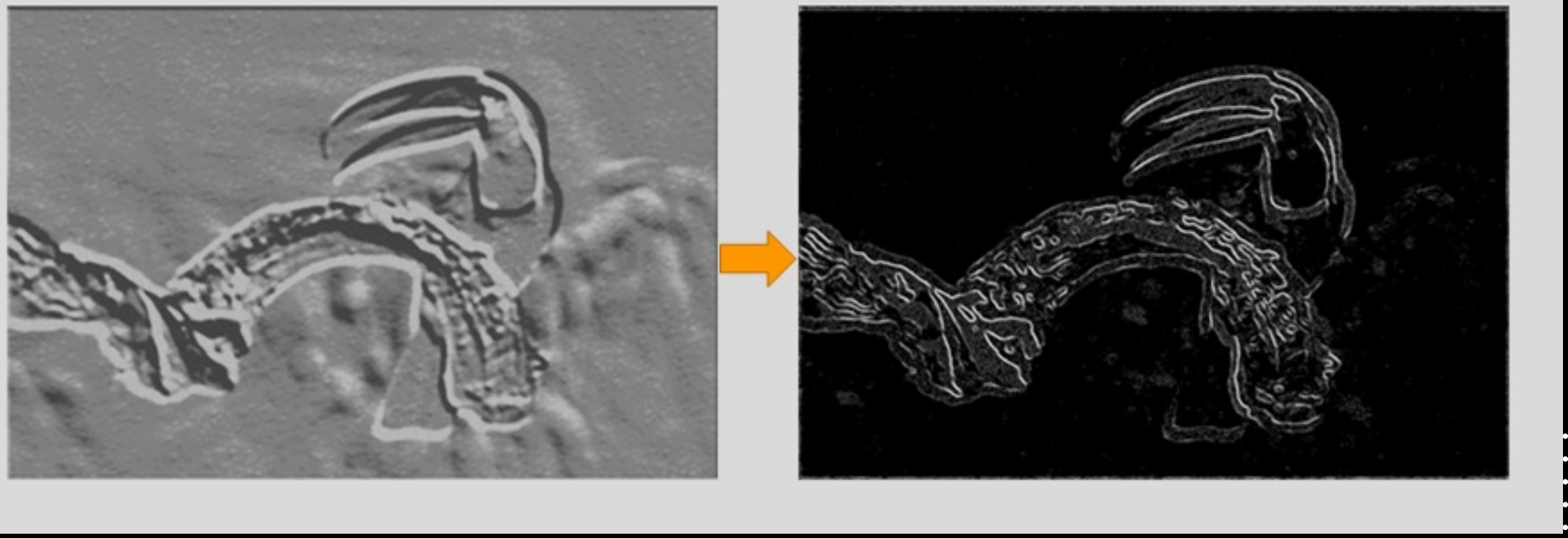
ReLU layer



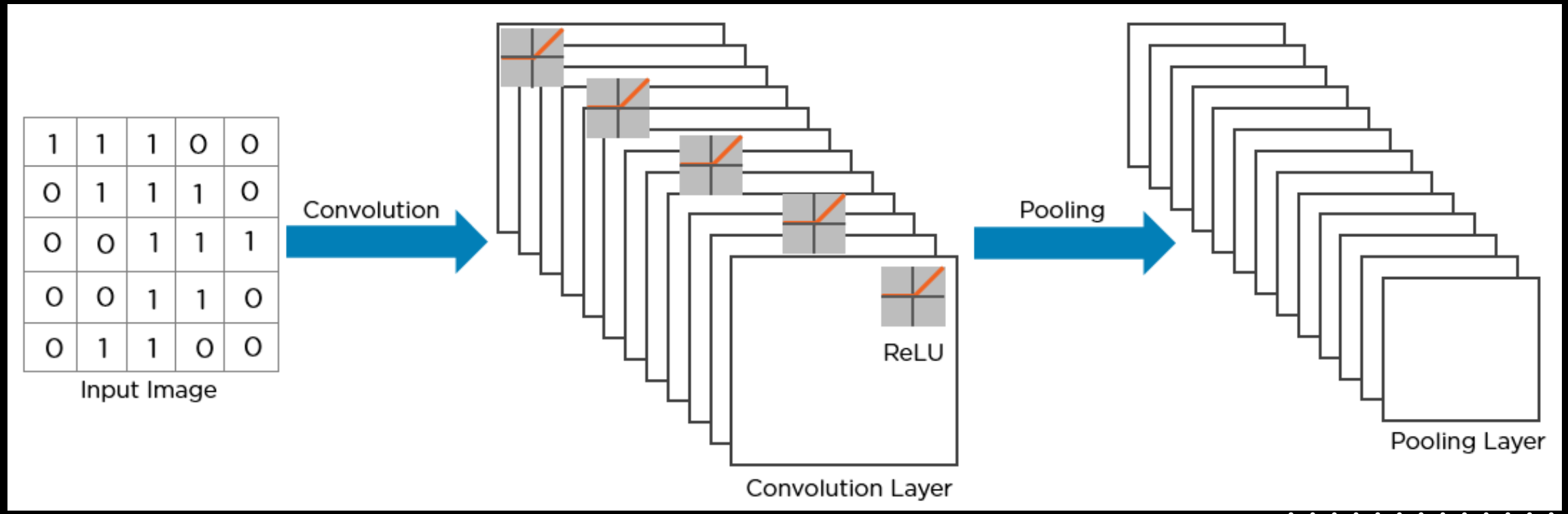
Pooling Layer



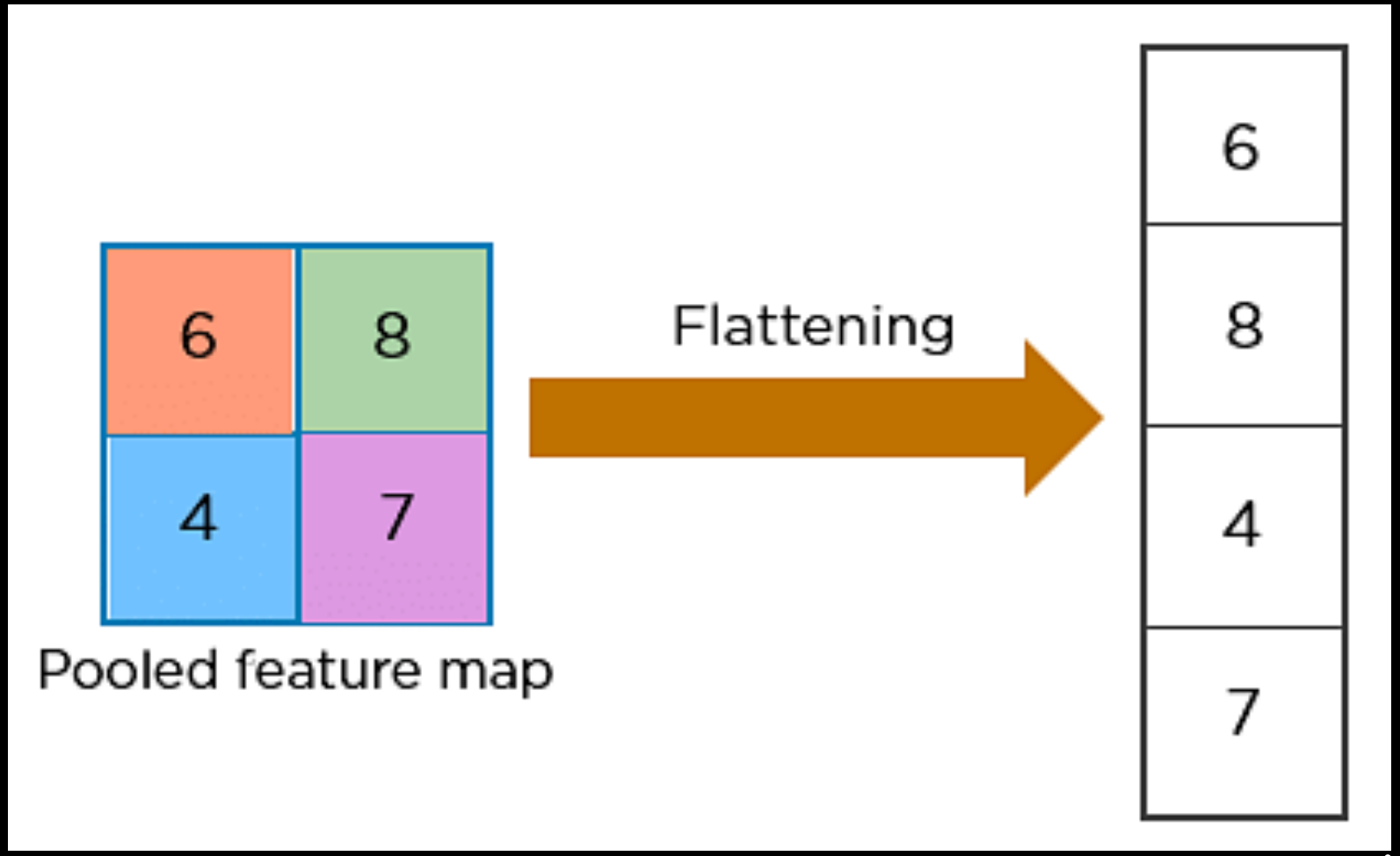
Pooling Layer



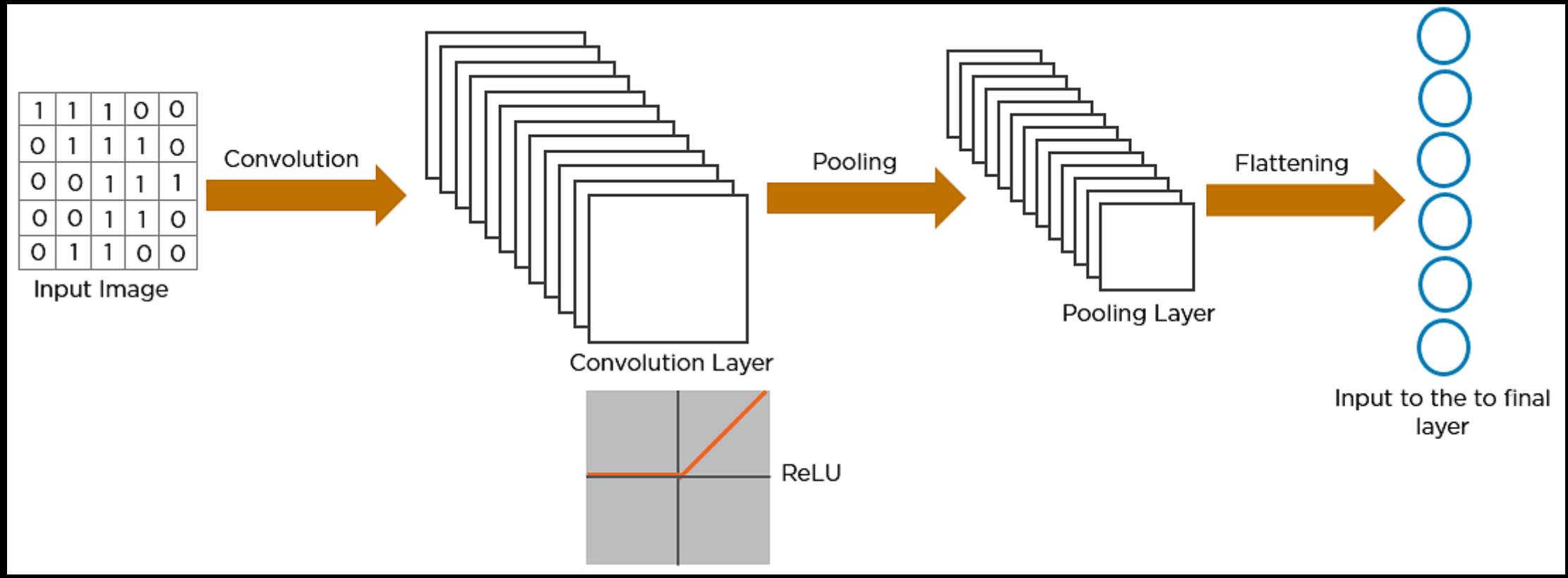
Pooling Layer



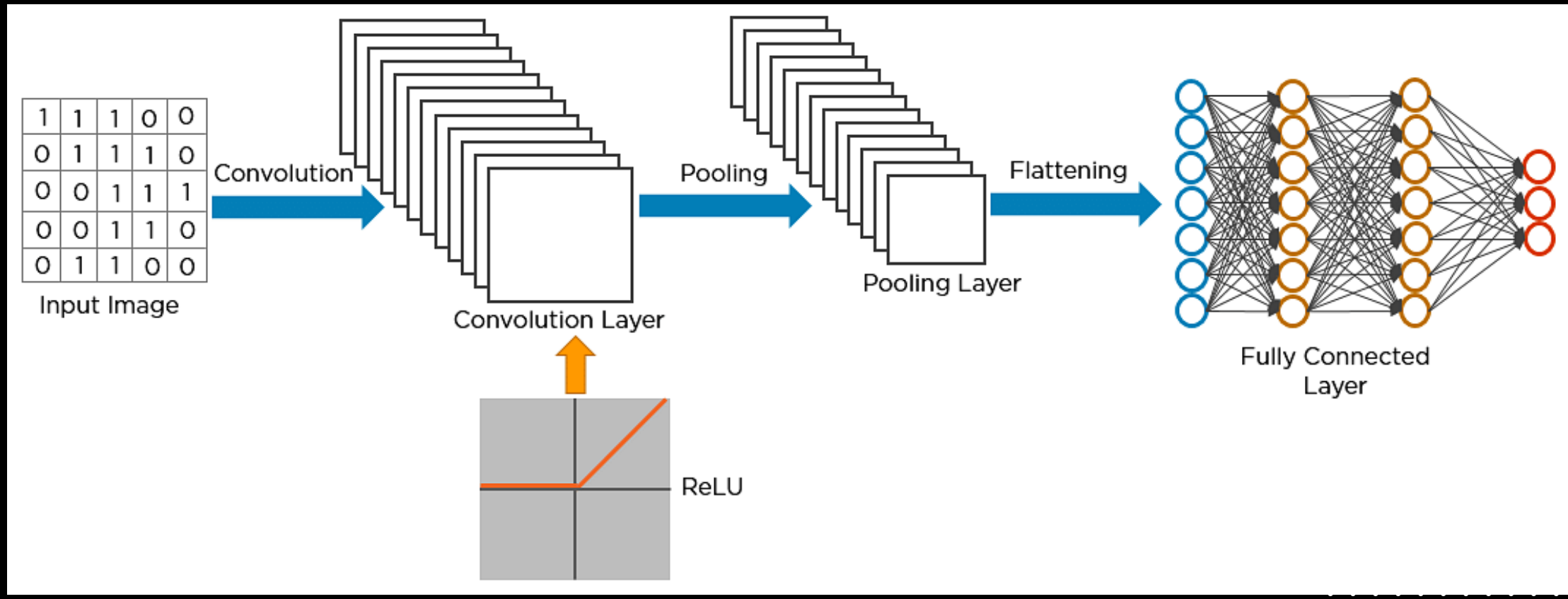
Flattening



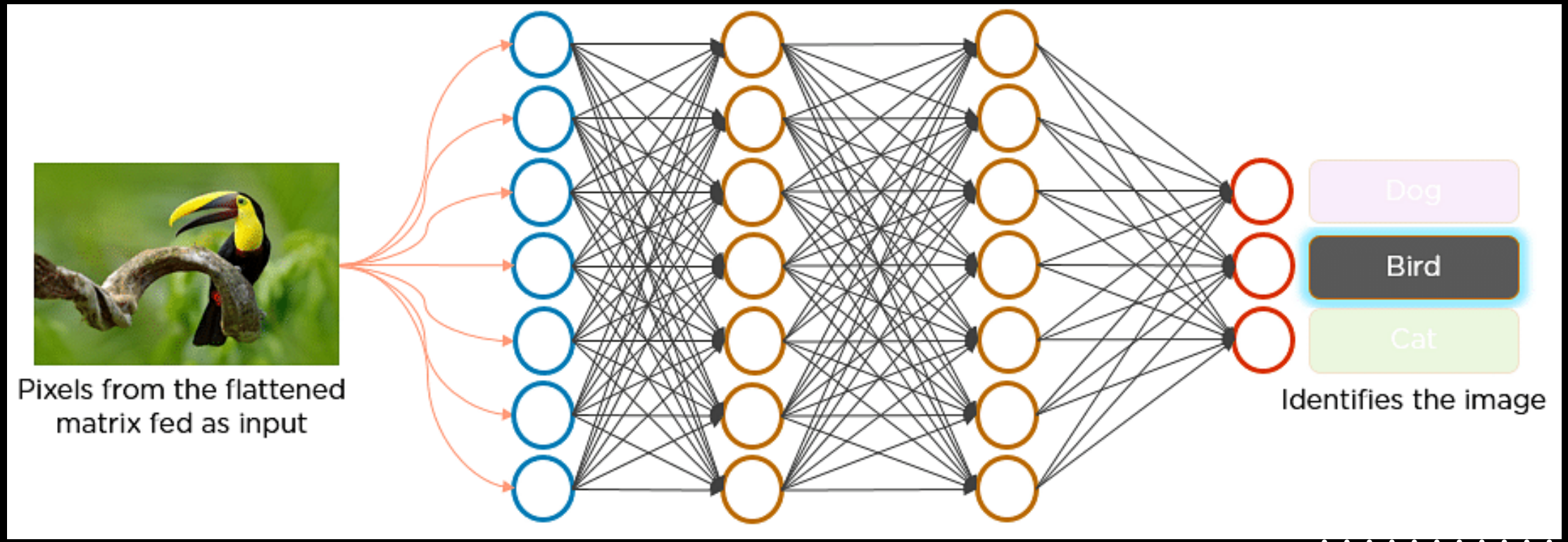
Final Layer


1 1 1 0 0
0 1 1 1 0
0 0 1 1 1
0 0 1 1 0
0 1 1 0 0
Input Image
Convolution
Convolution Layer
ReLU
Pooling
Pooling Layer
Flattening
Input to the to final layer

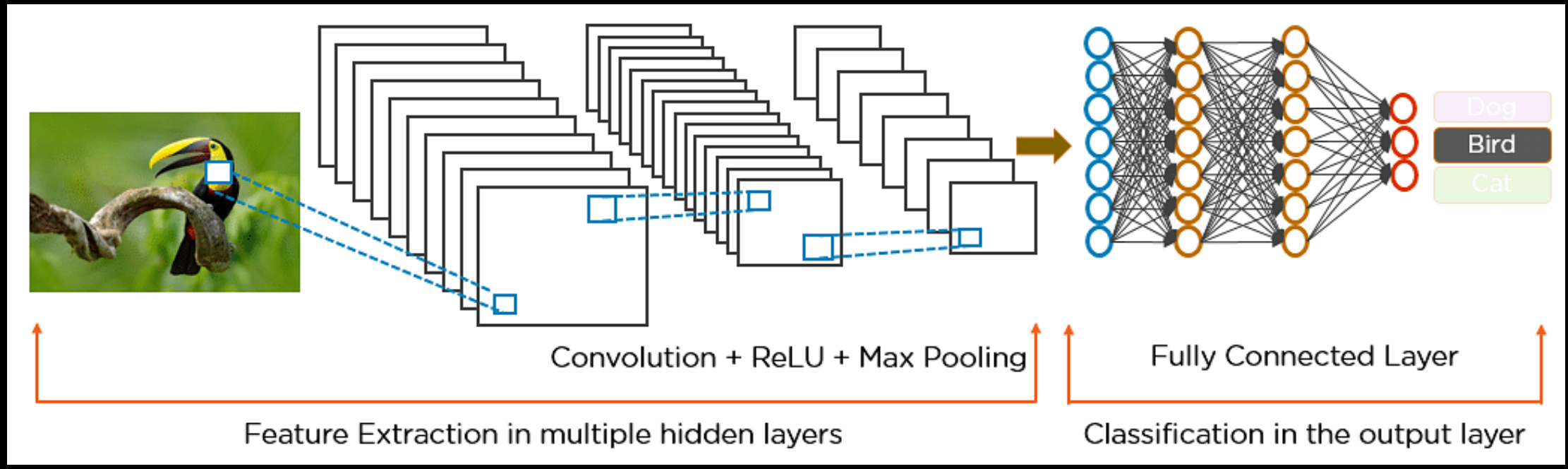
Fully Connected Layer



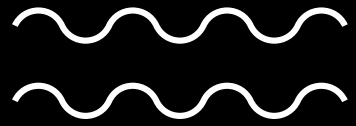
Fully Connected Layer

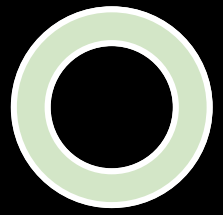


CNN

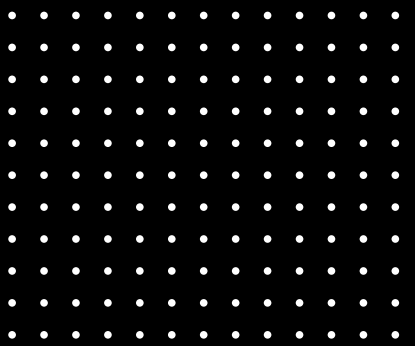

Dog
Bird
Cat
Convolution + ReLU + Max Pooling
Fully Connected Layer
Feature Extraction in multiple hidden layers
Classification in the output layer

Part B





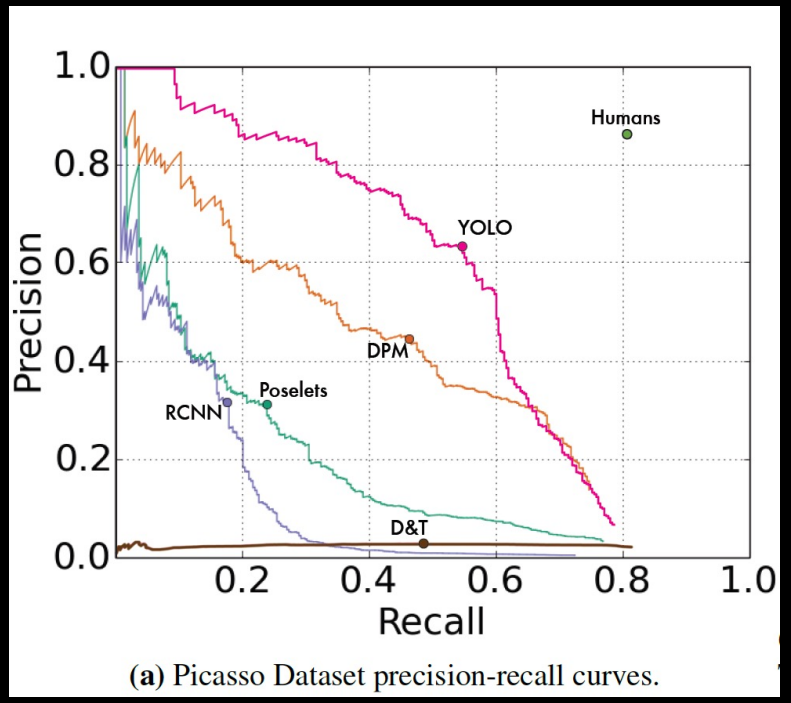
Object Detection



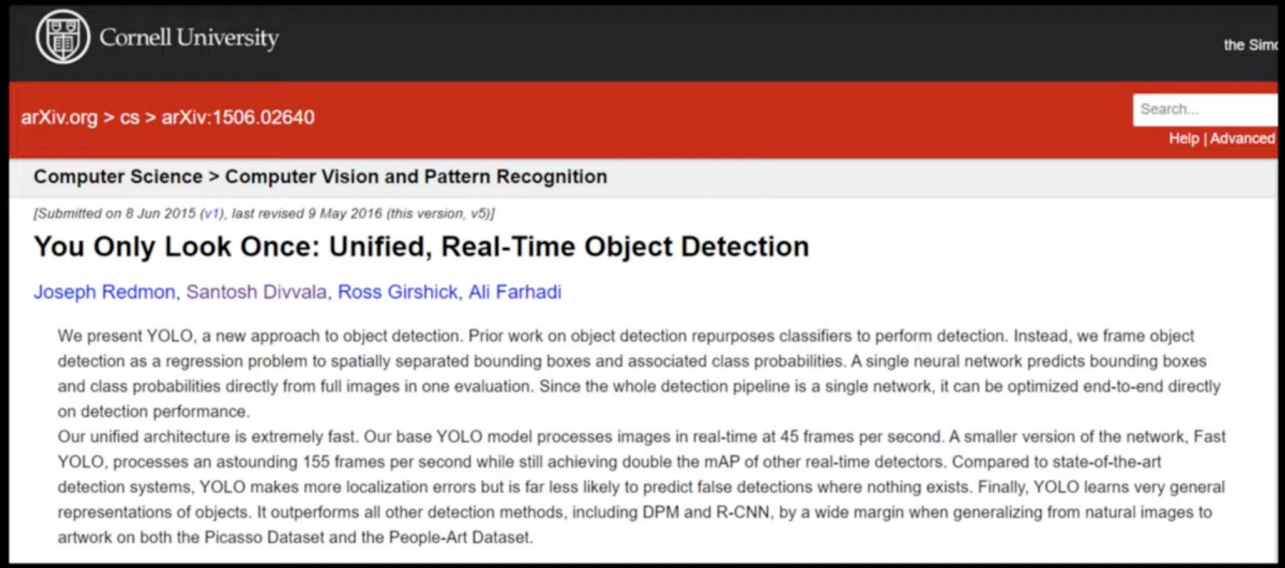
Object Detection using YOLO Algorithm

2016

YOLO V1



(a) Picasso Dataset precision-recall curves.


Cornell University

arXiv.org > cs > arXiv:1506.02640

Search...

Help | Advanced

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 8 Jun 2015 (v1), last revised 9 May 2016 (this version, v5)]

You Only Look Once: Unified, Real-Time Object Detection

Joseph Redmon, Santosh Divvala, Ross Girshick, Ali Farhadi

We present YOLO, a new approach to object detection. Prior work on object detection repurposes classifiers to perform detection. Instead, we frame object detection as a regression problem to spatially separated bounding boxes and associated class probabilities. A single neural network predicts bounding boxes and class probabilities directly from full images in one evaluation. Since the whole detection pipeline is a single network, it can be optimized end-to-end directly on detection performance.

Our unified architecture is extremely fast. Our base YOLO model processes images in real-time at 45 frames per second. A smaller version of the network, Fast YOLO, processes an astounding 155 frames per second while still achieving double the mAP of other real-time detectors. Compared to state-of-the-art detection systems, YOLO makes more localization errors but is far less likely to predict false detections where nothing exists. Finally, YOLO learns very general representations of objects. It outperforms all other detection methods, including DPM and R-CNN, by a wide margin when generalizing from natural images to artwork on both the Picasso Dataset and the People-Art Dataset.

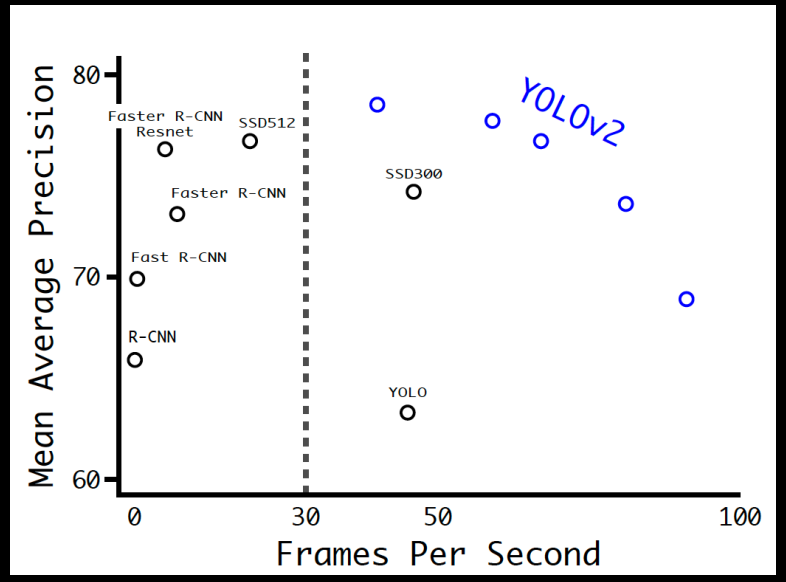
YOLO V1 - https://arxiv.org/abs/1506.02640

You Only Look Once

Object Detection using YOLO Algorithm

2016 YOLO V1

2017 YOLO V2



Cornell University

arXiv.org > cs > arXiv:1612.08242

Search... Help | Ad

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 25 Dec 2016]

YOLO9000: Better, Faster, Stronger

Joseph Redmon, Ali Farhadi

We introduce YOLO9000, a state-of-the-art, real-time object detection system that can detect over 9000 object categories. First we propose various improvements to the YOLO detection method, both novel and drawn from prior work. The improved model, YOLOv2, is state-of-the-art on standard detection tasks like PASCAL VOC and COCO. At 67 FPS, YOLOv2 gets 76.8 mAP on VOC 2007. At 40 FPS, YOLOv2 gets 78.6 mAP, outperforming state-of-the-art methods like Faster RCNN with ResNet and SSD while still running significantly faster. Finally we propose a method to jointly train on object detection and classification. Using this method we train YOLO9000 simultaneously on the COCO detection dataset and the ImageNet classification dataset. Our joint training allows YOLO9000 to predict detections for object classes that don't have labelled detection data. We validate our approach on the ImageNet detection task. YOLO9000 gets 19.7 mAP on the ImageNet detection validation set despite only having detection data for 44 of the 200 classes. On the 156 classes not in COCO, YOLO9000 gets 16.0 mAP. But YOLO can detect more than just 200 classes; it predicts detections for more than 9000 different object categories. And it still runs in real-time.

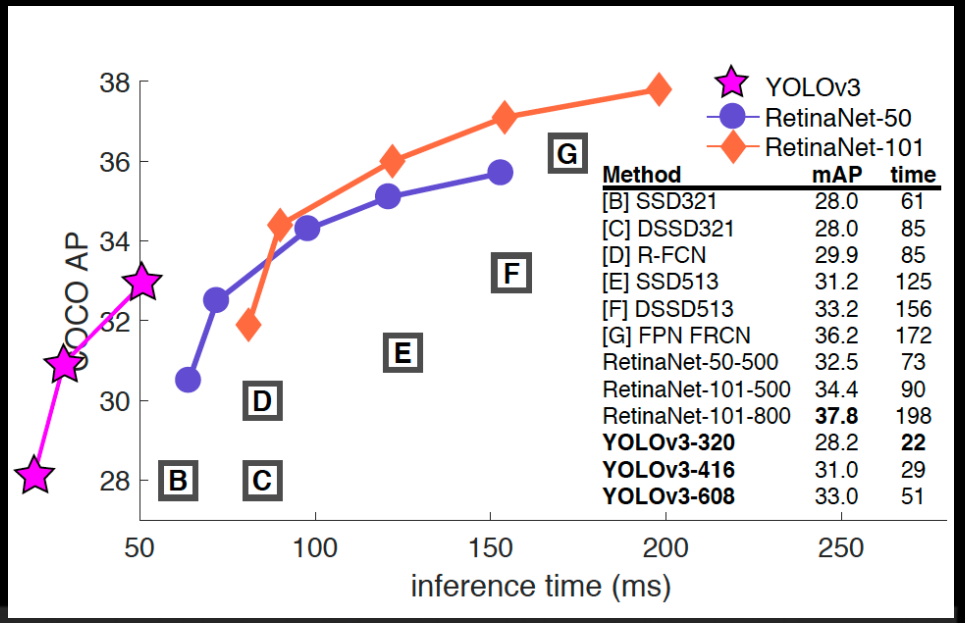
YOLO V2 - https://arxiv.org/abs/1612.08242v1

Object Detection using YOLO Algorithm

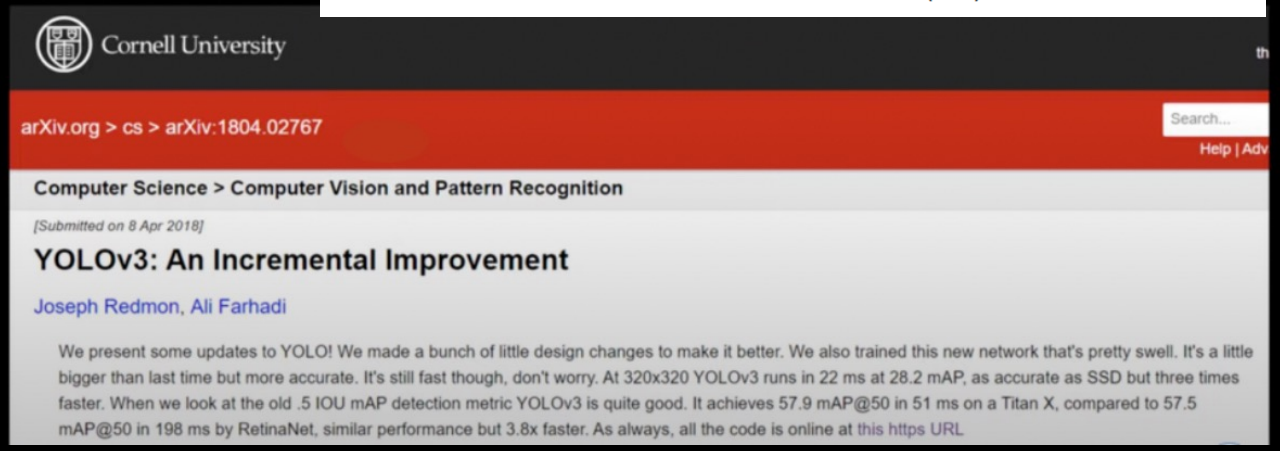
2016 YOLO V1

2017 YOLO V2

2018 YOLO V3



Method	mAP	time
[B] SSD321	28.0	61
[C] DSSD321	28.0	85
[D] R-FCN	29.9	85
[E] SSD513	31.2	125
[F] DSSD513	33.2	156
[G] FPN FRCN	36.2	172
RetinaNet-50-500	32.5	73
RetinaNet-101-500	34.4	90
RetinaNet-101-800	**37.8**	198
YOLOv3-320	28.2	**22**
YOLOv3-416	31.0	29
YOLOv3-608	33.0	51



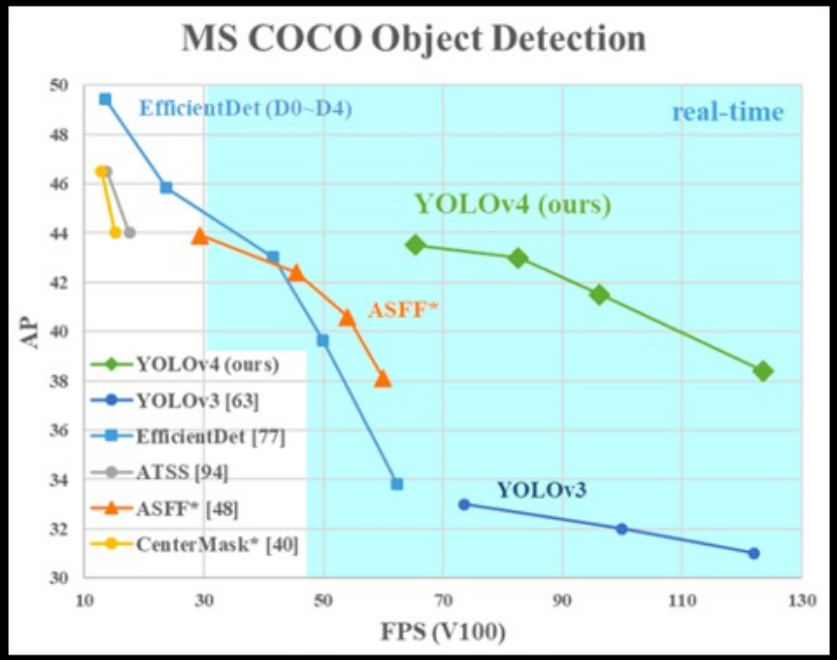
YOLO V3 - https://arxiv.org/abs/1804.02767v1

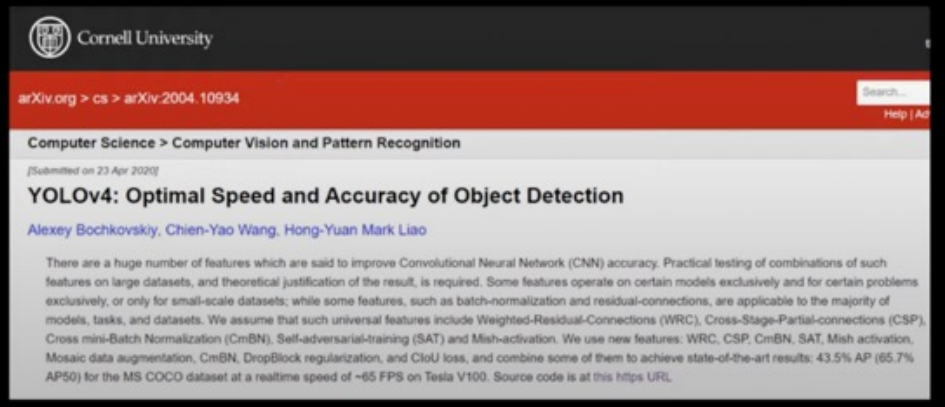
Object Detection using YOLO Algorithm

2016 YOLO V1

2017 YOLO V2

2018 YOLO V3




Cornell University

arXiv.org > cs > arXiv:2004.10934

Computer Science > Computer Vision and Pattern Recognition

[Submitted on 23 Apr 2020]

YOLOv4: Optimal Speed and Accuracy of Object Detection

Alexey Bochkovskiy, Chien-Yao Wang, Hong-Yuan Mark Liao

There are a huge number of features which are said to improve Convolutional Neural Network (CNN) accuracy. Practical testing of combinations of such features on large datasets, and theoretical justification of the result, is required. Some features operate on certain models exclusively and for certain problems exclusively, or only for small-scale datasets; while some features, such as batch-normalization and residual-connections, are applicable to the majority of models, tasks, and datasets. We assume that such universal features include Weighted-Residual-Connections (WRC), Cross-Stage-Partial-connections (CSP), Cross mini-Batch Normalization (CmBN), Self-adversarial-training (SAT) and Mish-activation. We use new features: WRC, CSP, CmBN, SAT, Mish activation, Mosaic data augmentation, CmBN, DropBlock regularization, and CIoU loss, and combine some of them to achieve state-of-the-art results: 43.5% AP (65.7% AP50) for the MS COCO dataset at a realtime speed of ~65 FPS on Tesla V100. Source code is at this https URL

2020 YOLO V4

2020 YOLO V5

YOLO V4 - https://arxiv.org/abs/2004.10934v1

YOLO V5 - https://pytorch.org/hub/ultralytics_yolov5/

Image Classification

Is this a dog or a person?

Neural Network Output

Dog = 1
Person = 0

Object Localization

Where exactly is the dog in this image?

Neural Network Output

Dog = 1
Person = 0
+
Bounding Box

Object Localization

$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 50 \\ 70 \\ 60 \\ 70 \\ 1 \\ 0 \end{bmatrix}$$

C_1 = *Dog class*
C2 = Person Class

$$\begin{bmatrix} 1 \\ 30 \\ 28 \\ 28 \\ 82 \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ - \\ - \\ - \\ - \\ - \\ - \end{bmatrix}$$

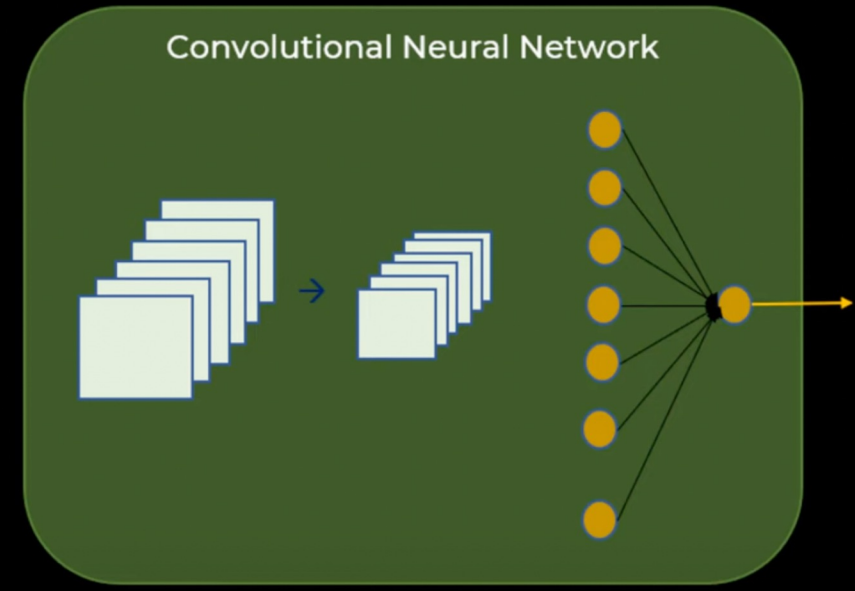
X_train

y_train

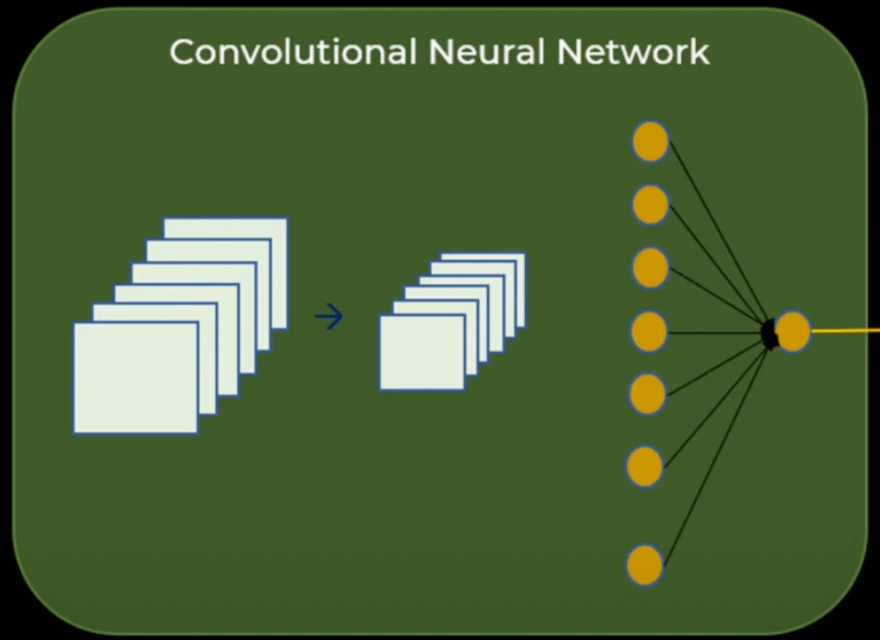
$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix} \quad \begin{bmatrix} 1 \\ 50 \\ 70 \\ 60 \\ 70 \\ 1 \\ 0 \end{bmatrix}$$

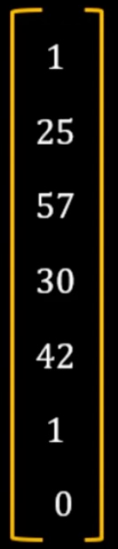
$$\begin{bmatrix} 1 \\ 30 \\ 55 \\ 28 \\ 82 \\ 0 \\ 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 \\ - \\ - \\ - \\ - \\ - \\ - \end{bmatrix}$$

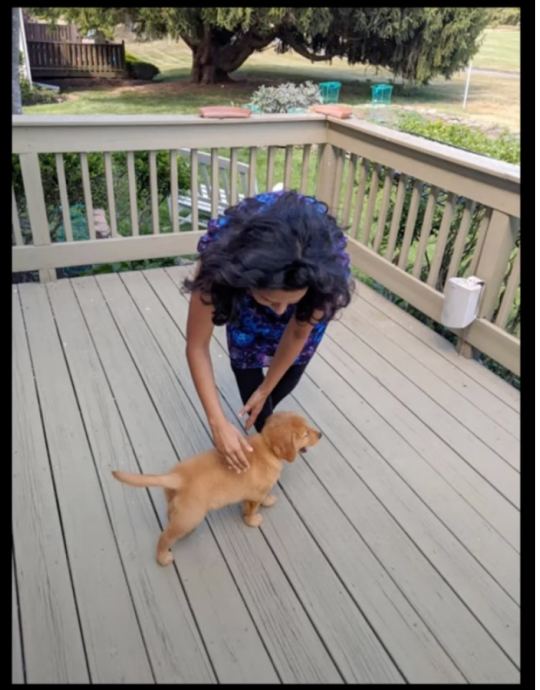


$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$

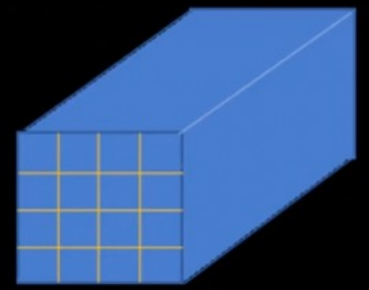

Convolutional Neural Network


1
25
57
30
42
1
0

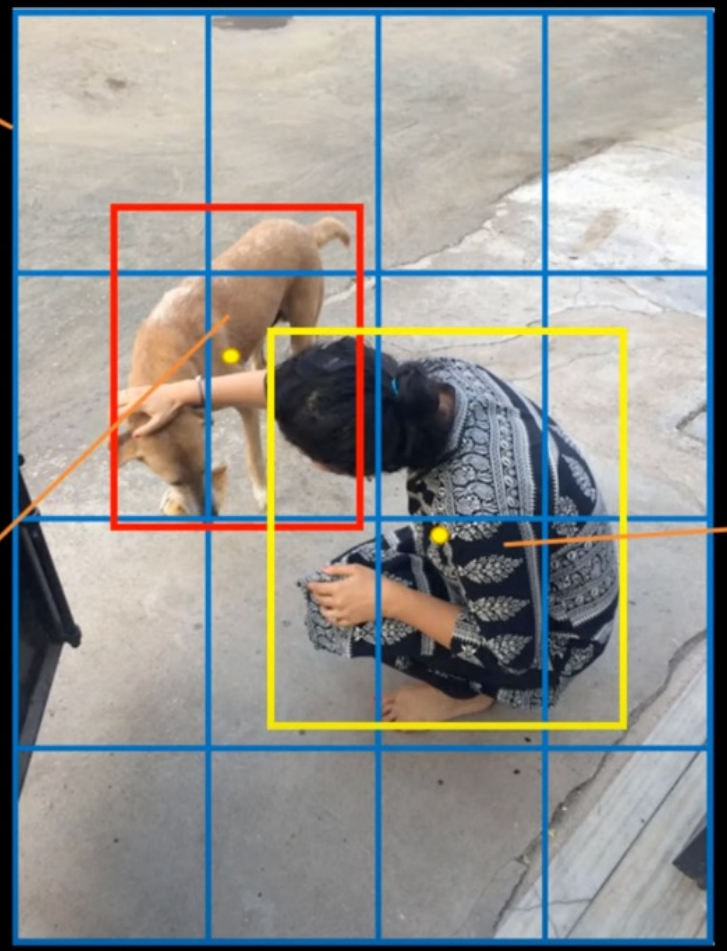
This works ok only for single object. What about multiple objects in an image?



4 by 4 by 7

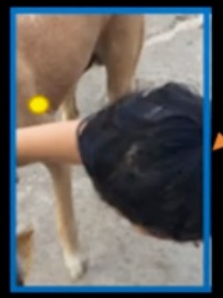


$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix} \quad \begin{bmatrix} 0 \\ - \\ - \\ - \\ - \\ - \\ - \end{bmatrix}$$

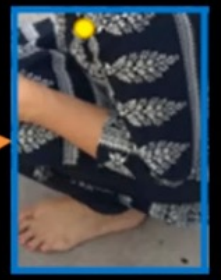


$$\begin{bmatrix} 1 \\ 0.05 \\ 0.3 \\ 2 \\ 1.3 \\ 1 \\ 0 \end{bmatrix}$$

(0,0)



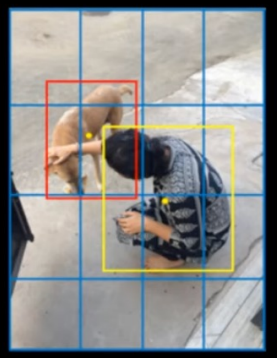
(1,1)



$$\begin{bmatrix} 1 \\ 0.32 \\ 0.02 \\ 3 \\ 2 \\ 0 \\ 1 \end{bmatrix}$$

Training

X_train



y_train

16 such vectors

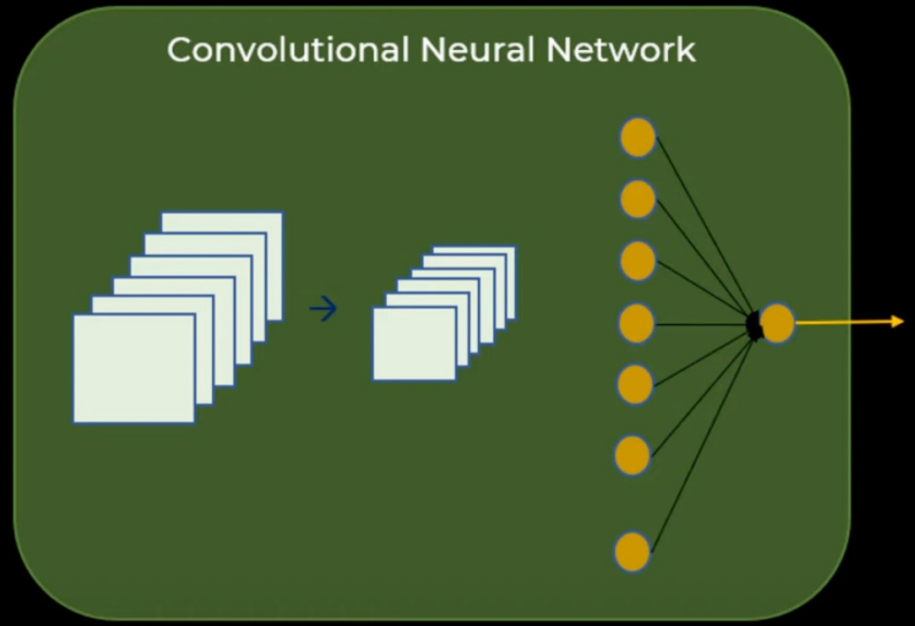
$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$

16 such vectors

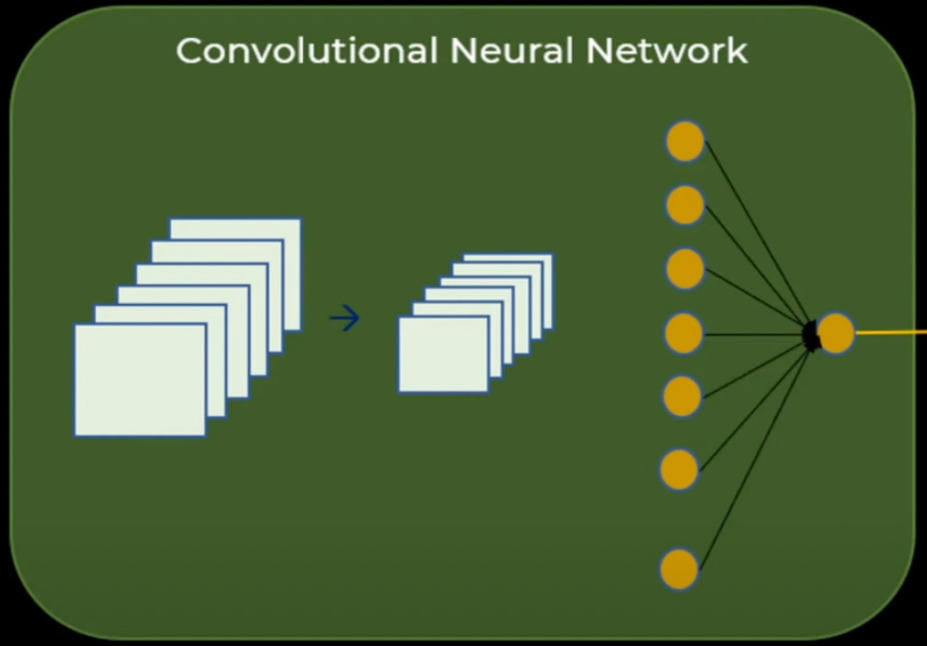
$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$

16 such vectors

$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$

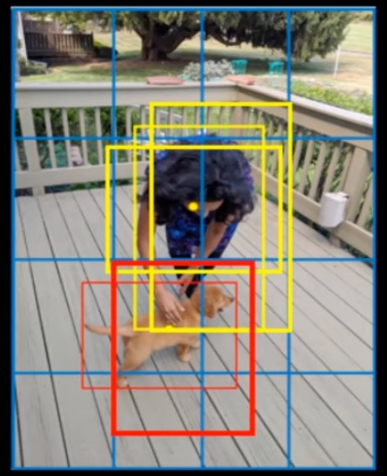


Prediction

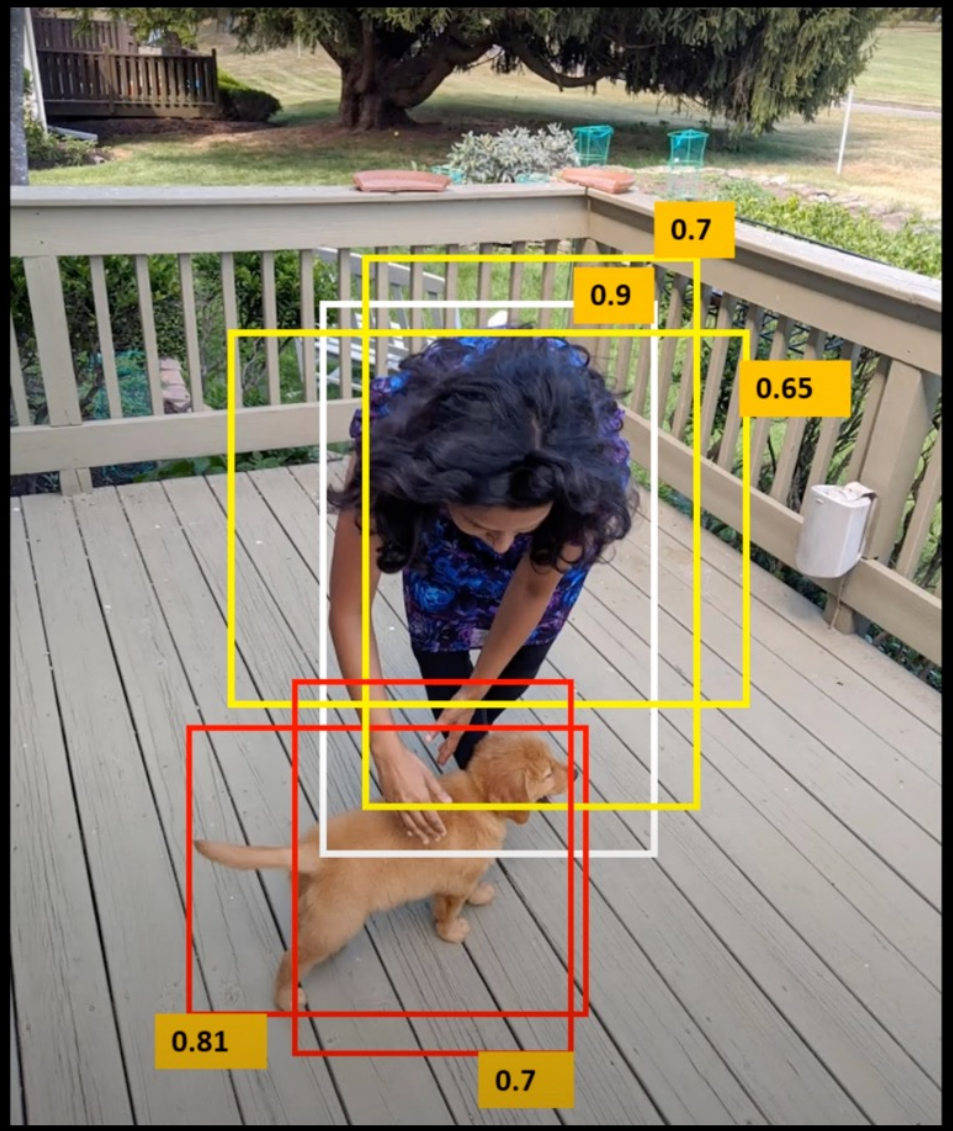


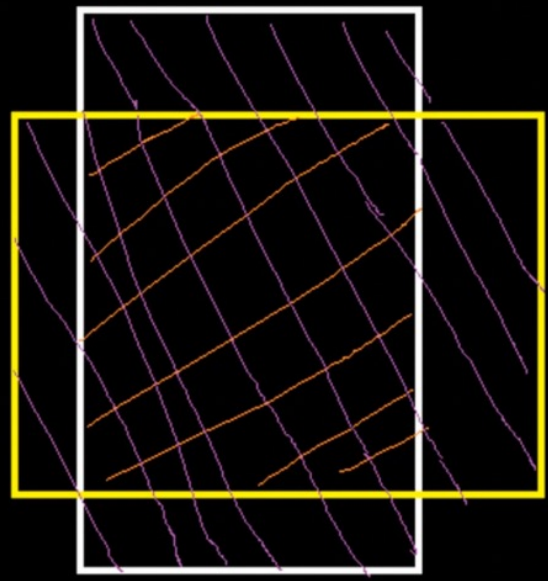
16 such vectors

$$\begin{bmatrix} P_c \\ B_x \\ B_y \\ B_w \\ B_h \\ C_1 \\ C_2 \end{bmatrix}$$



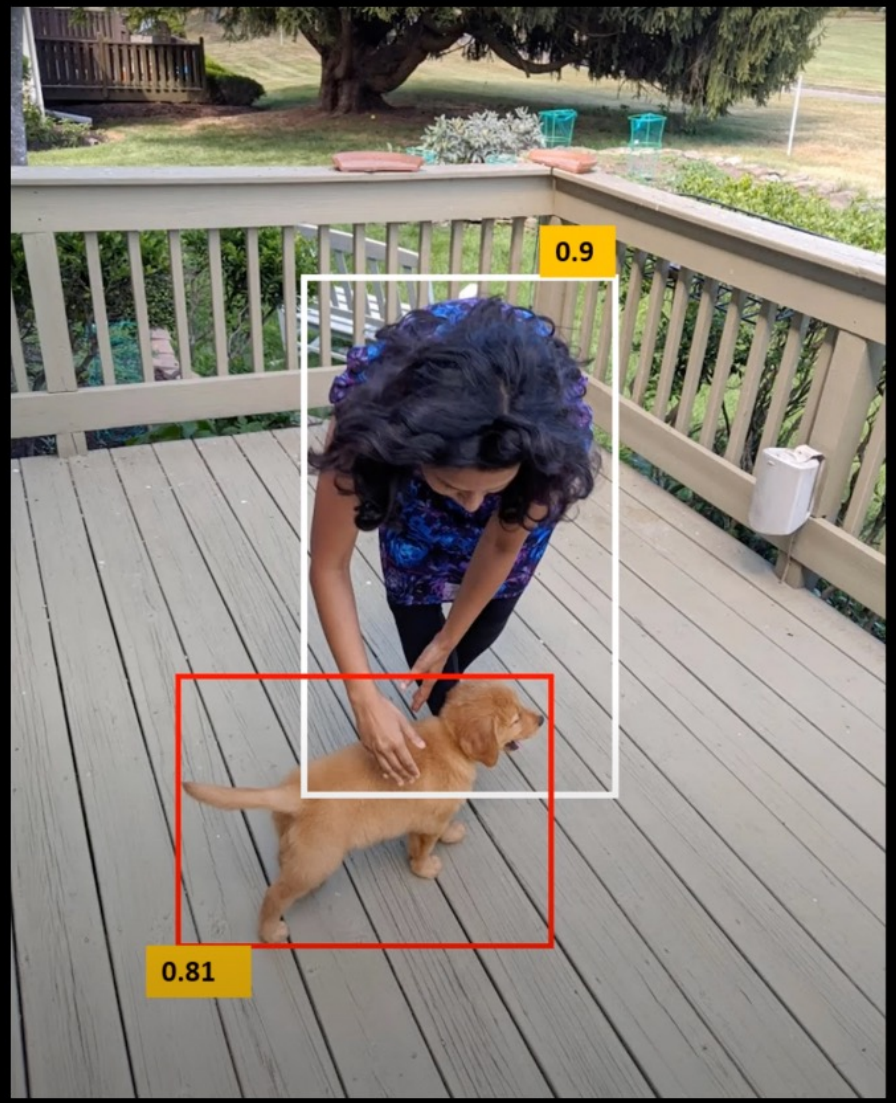
Multiple bounding boxes





Intersection over union = intersect area / union area

Intersection over union : IOU



Non max suppression

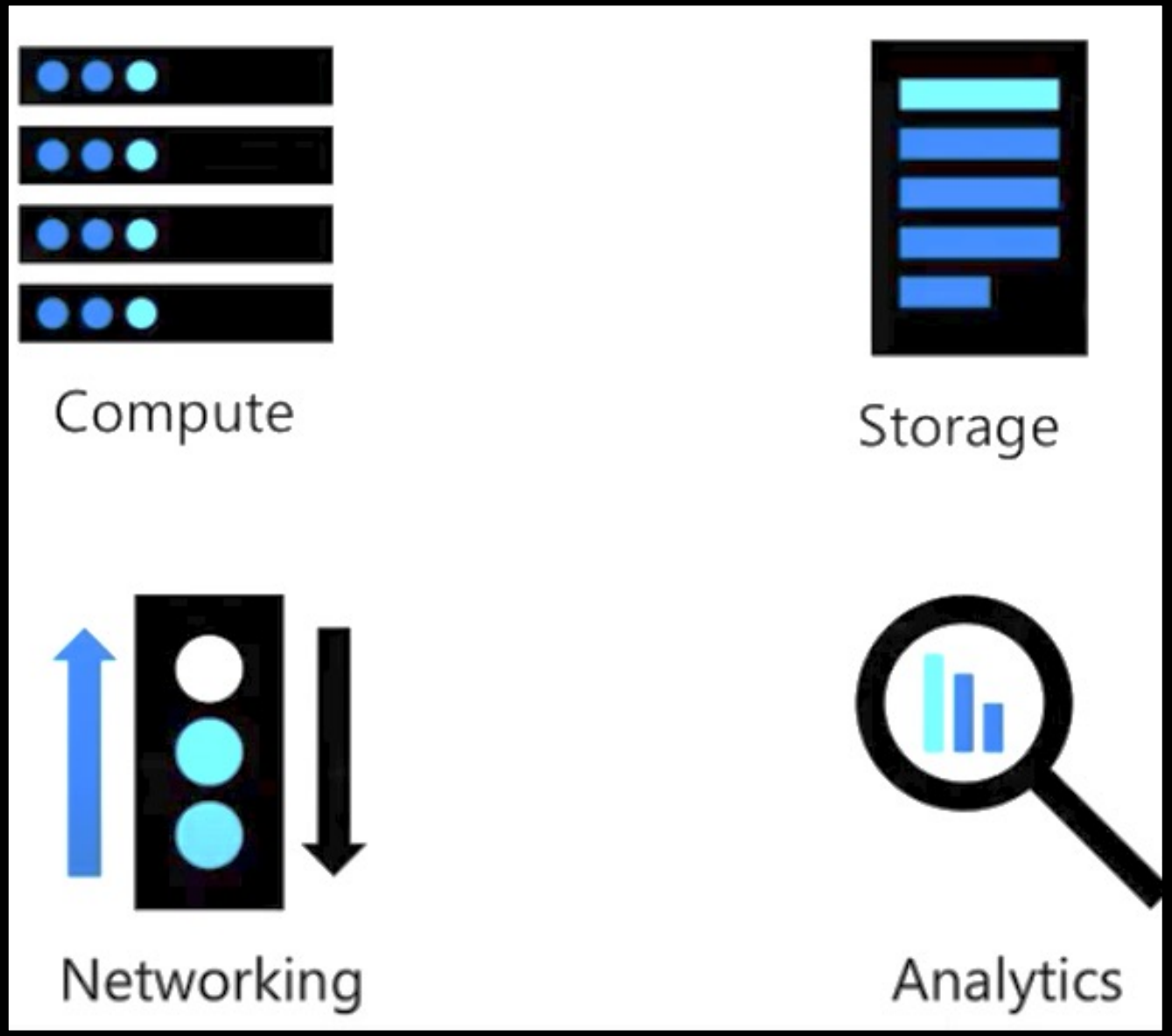
Part C

Object Detection with Azure Cloud

Why cloud services?

Cloud Services

- Cloud Providers include Microsoft, Amazon, Google



What is Cloud Computing

- Simply put, cloud computing is the delivery of computing services—including servers, storage, databases, networking, software, analytics, and intelligence—over the Internet (“the cloud”) to offer faster innovation, flexible resources, and economies of scale. You typically pay only for cloud services you use, helping you lower your operating costs, run your infrastructure more efficiently, and scale as your business needs change.
- https://azure.microsoft.com/en-us/overview/what-is-cloud-computing/
- On-premise
  - You own the servers, hire IT staff, you pay or rent real-estate and you administrate services and take all the risks
- Cloud Providers
  - Someone else owns the servers, they hire IT staff, they pay or rent real-estate, they administrate services, and you are only responsible for configuring your cloud services and code

Cloud computing characteristics

Ref: The NIST Definition of Cloud Computing
https://csrc.nist.gov/publications/detail/sp/800-145/final

On-demand self-service

Ubiquitous network access

Location transparent resource pooling

Rapid elasticity

Measured service: Consumption Based

Types of Cloud Computing

SaaS Software as a Service **For Customers**

A product that is run and managed by the service provider
Don't worry about how the service is maintained.
It just works and remains available.

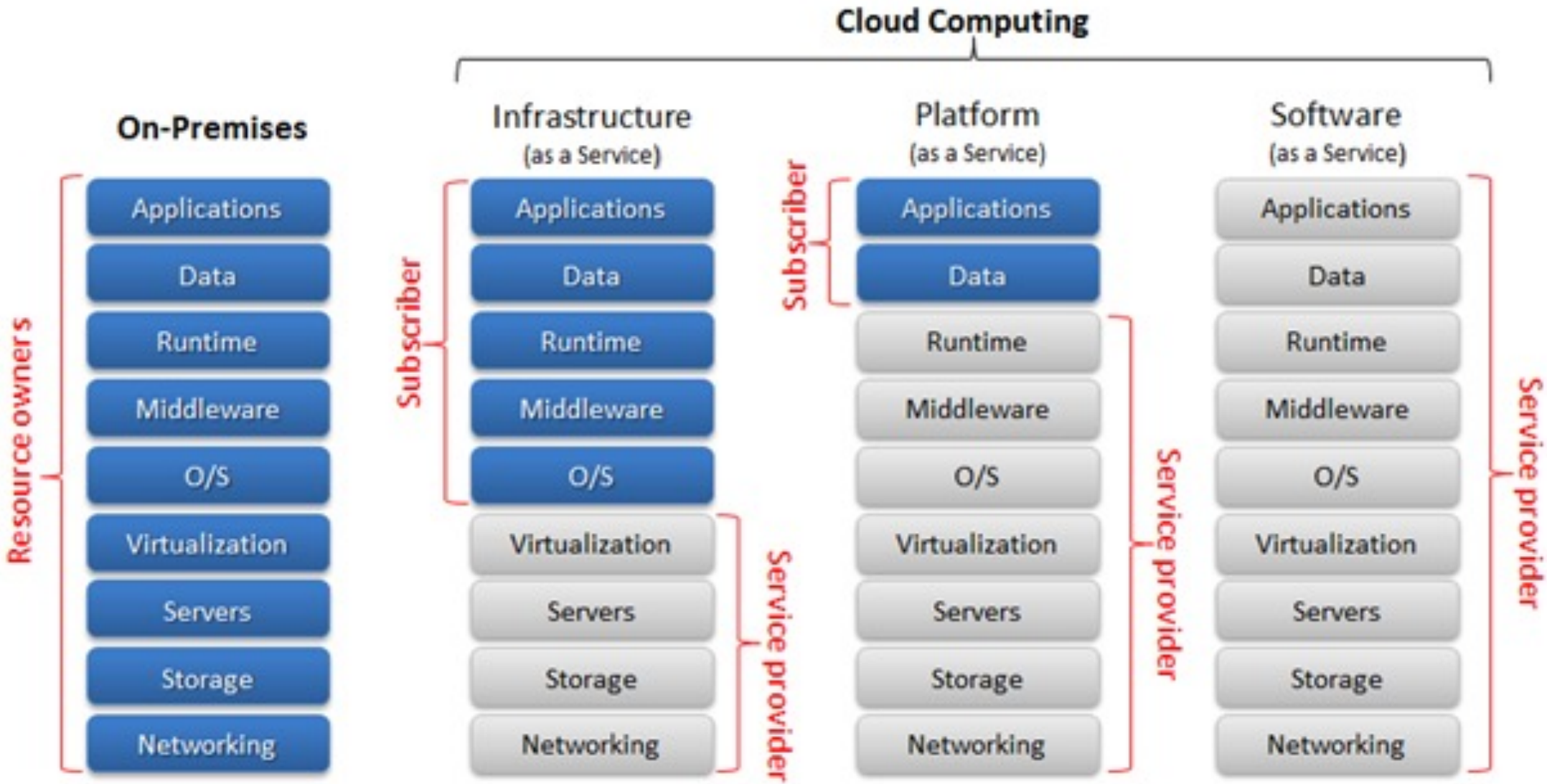
PaaS Platform as a Service **For Developers**

Focus on the deployment and management of your apps.
Don't worry about, provisioning, configuring or understanding the hardware or OS.

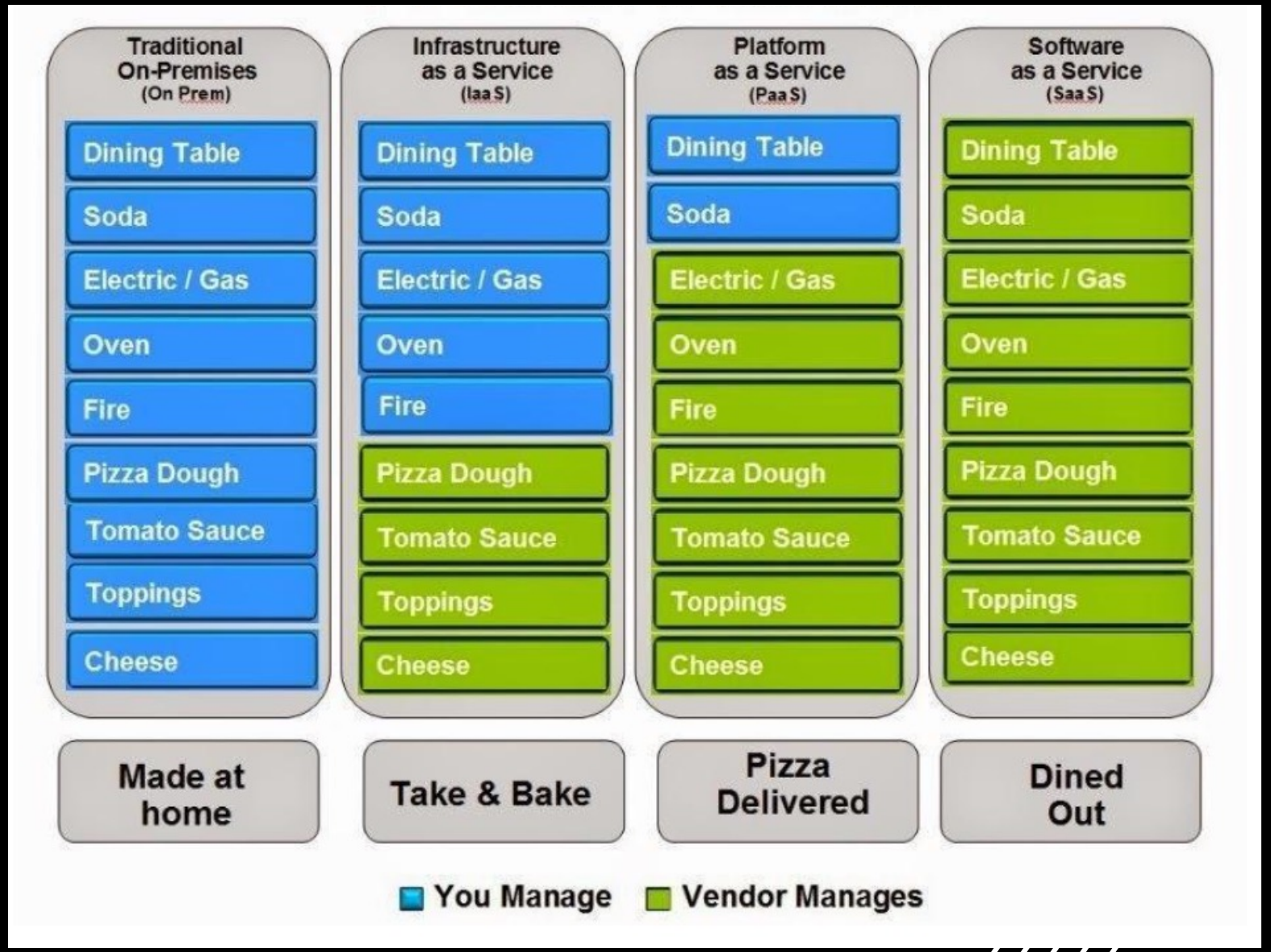
IaaS Infrastructure as a Service **For Admins**

The basic building blocks for cloud IT. Provides access to networking features, computers and data storage space.
Don't worry about IT staff, data centers and hardware.

Separation of Responsibilities



Think of it as “Pizza as a Service”



Benefits of Cloud Computing

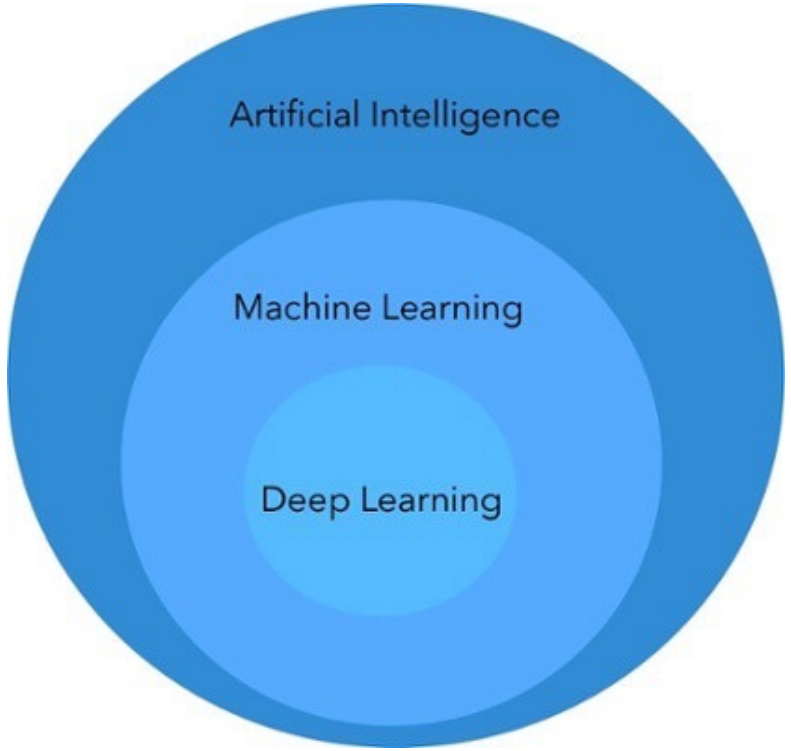
Cost-effective	You **pay for what you consume**, **no up-front cost**. Pay-as-you-go (PAYG) thousands of customers sharing the cost of the resources
Global	Launch workloads **anywhere in the world**, Just choose a region
Secure	Cloud provider takes care of physical security. **Cloud services can by secure by default** or you have the ability to configure access down to granular level.
Reliable	data backup, disaster recovery, and data replication, and fault tolerance
Scalable	Increase or decrease resources and services based on demand
Elastic	**Automate** scaling during spikes and drop in demand
Current	The underlying hardware and managed software is patched, upgraded and replaced by the cloud provider without interruption to you.

AI and ML Services

What is Artificial Intelligence (AI)?
Machines that perform jobs that mimic human behavior

What is Machine Learning (ML)?
Machines that get better at a task without explicit programming

What is Deep Learning (DL)?
Machines that have an artificial neural network inspired by the human brain to solve complex problems.



Azure Machine Learning Service
A service for that simplifies running AI/ML related workloads allowing you to build flexible Pipelines to automate workflow. Use Python an R, Run DL workloads such as Tensorflow

Azure Machine Learning Studio (classic)
An older service that manages AI/ML workloads. Does not have a pipeline and other limitations. Workloads are not easily transferable to from classic to the new service.

Azure Machine Learning Studio is a web portal *in* Azure Machine Learning that contains low-code and no-code options for project authoring and asset management.

AI and ML Services

Personalizer Deliver rich, personalised experiences for every user.

Translator Add real-time, multi-language text translation to your apps, website and tools.

Anomaly detector Detect anomalies in data to quickly identify and troubleshoot issues.

Azure Bot Service Intelligent, serverless bot service that scales on demand

Form Recogniser Automate the extraction of text, key/value pairs and tables from your documents.

Computer Vision Easily customise computer vision models for your unique use case.

Language Understanding Build natural language understanding into apps, bots and IoT devices.

AI and ML Services

QnA Maker Create a conversational question-and-answer bot from your existing content.

Text Analytics Extract information such as sentiment, key phrases, named entities and language from your text.

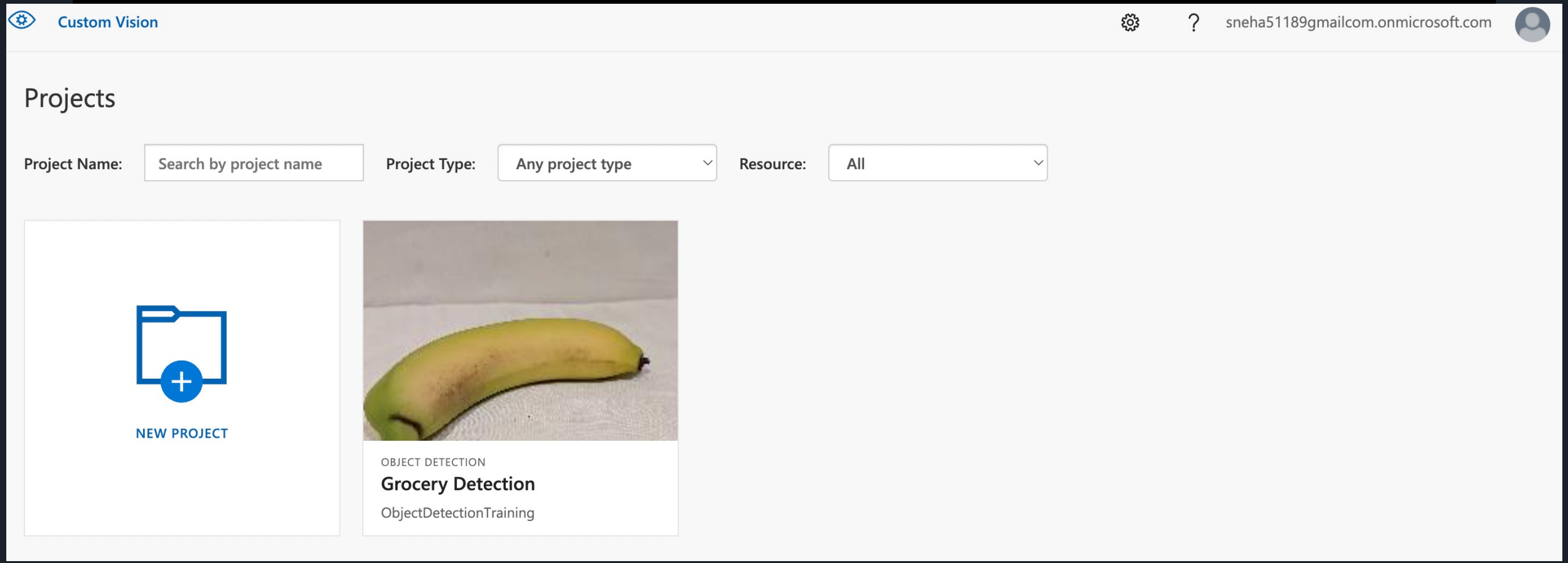
Content moderator Moderate text and images to provide a safer, more positive user experience.

Face Detect and identify people and emotions in images.

Ink Recogniser Recognise digital ink content, such as handwriting, shapes and document layout.

Demo: Object Detection using Azure Custom Vision

Object Detection using Azure Custom Vision



Grocery Detection

Training Images | Performance | Predictions | Train | Quick Test

Filter

Iteration

Workspace

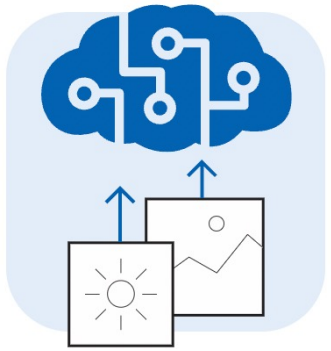
Tags

Tagged | Untagged

Showing: all tagged images

Search For Tags:

Add images | Delete | Select all



Looks like you don't have any images here!

Go ahead and browse for images to upload to your project, tag them, and they will be ready to be trained.

Add images

.JPG, .PNG, .BMP format, up to 6 MB per image

Grocery Detection

Training Images | Performance | Predictions

Train | Quick Test

Filter

Iteration

Workspace

Tags

Tagged | Untagged

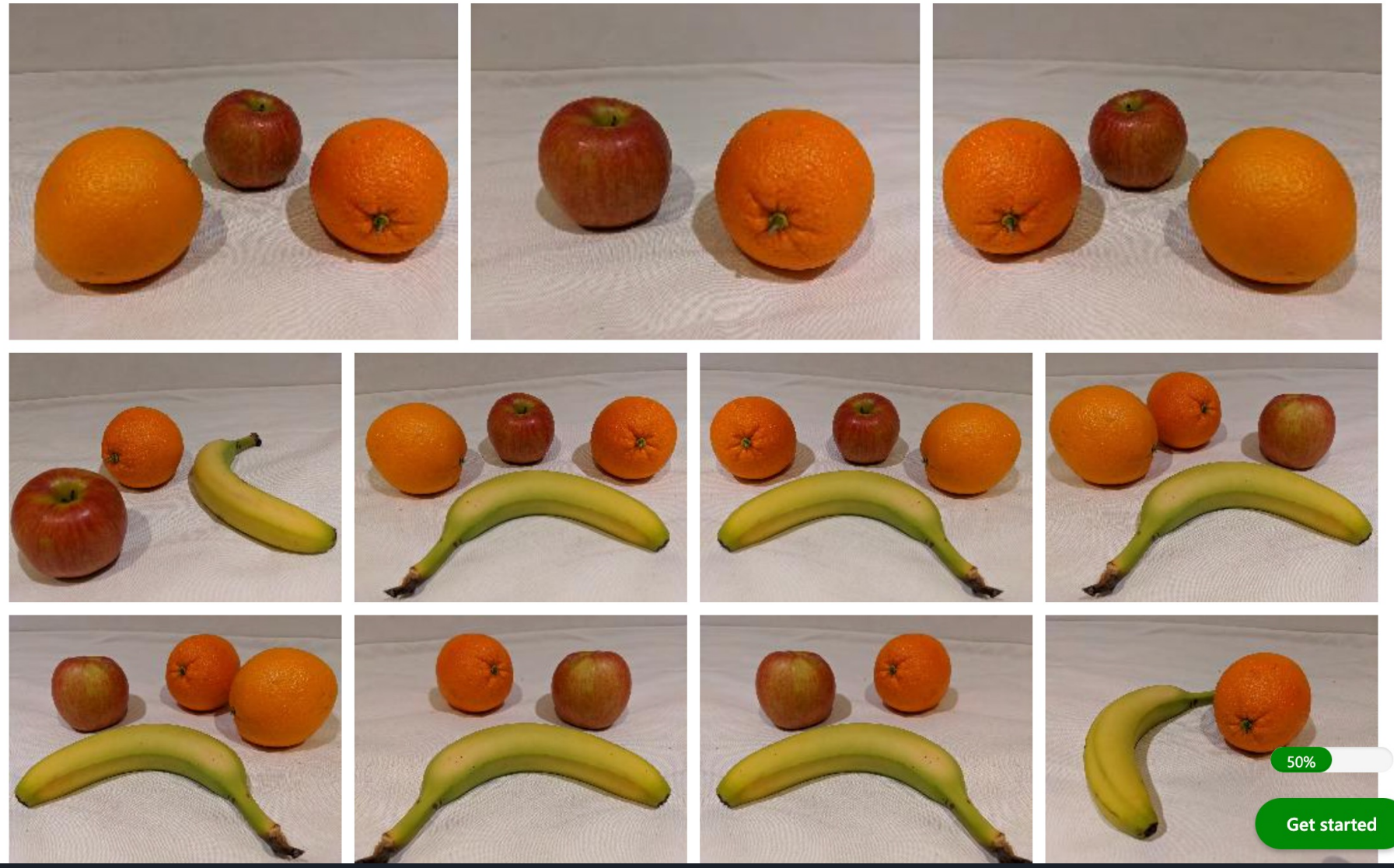
Showing: all untagged images

Suggested Tags

Quickly label your untagged images with suggested objects. Learn more.

Get suggested objects

Add images | Delete | Select all



50%

Get started

Grocery Detection

Training Images | Performance | Predictions | Train | Quick Test

Filter

Iteration

Workspace

Tags

Tagged

Showing: all untag

Suggested Tags

Quickly label your u
with suggested obje

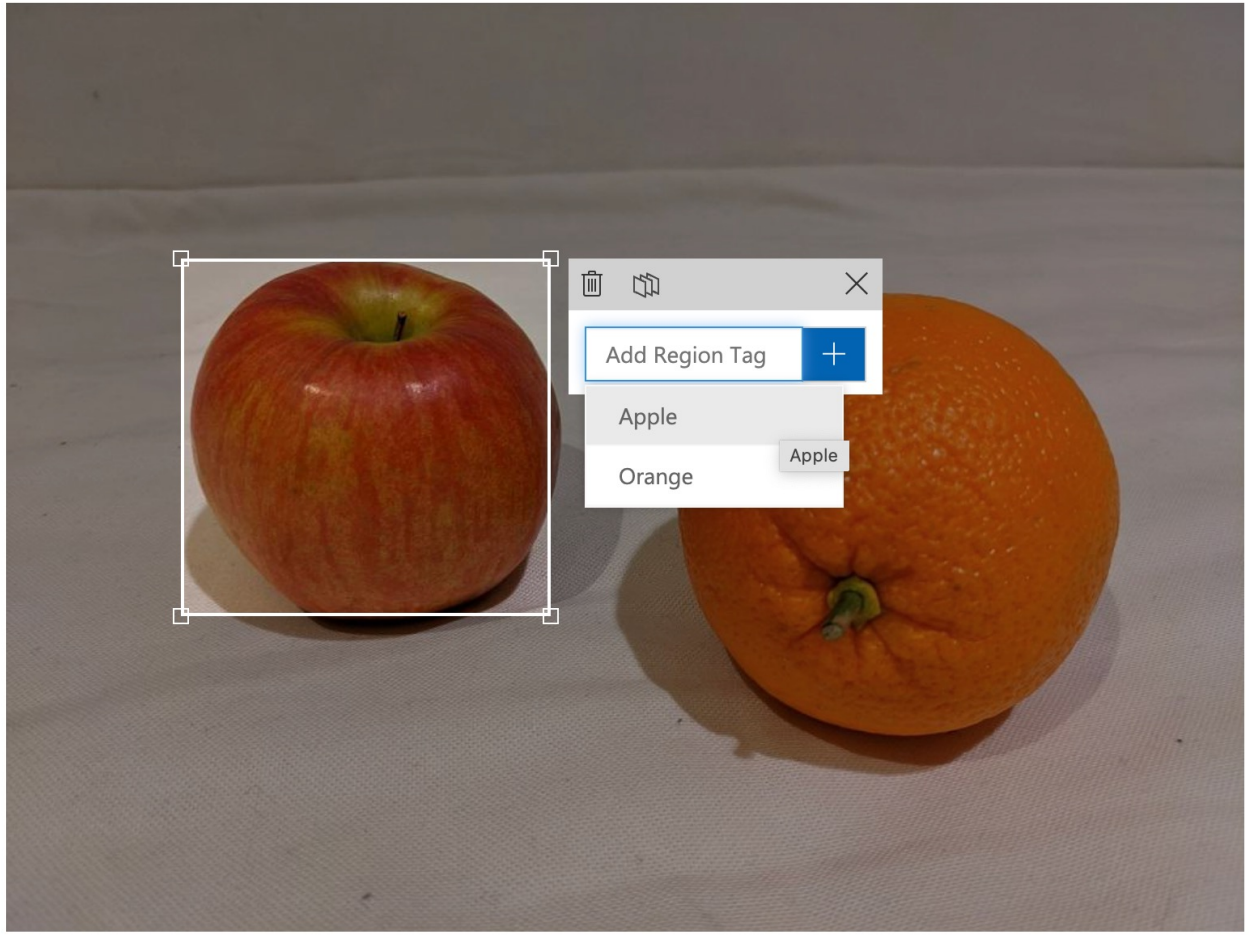
Get suggeste

50%

Get started

Image Detail

Undo Changes | Regions Shown



My Objects

To create an object, hover and select the region in the image

Grocery Detection

Training Images | Performance | Predictions

Train | Quick Test

Iterations

Probability Threshold: 50%

Overlap Threshold: 30%

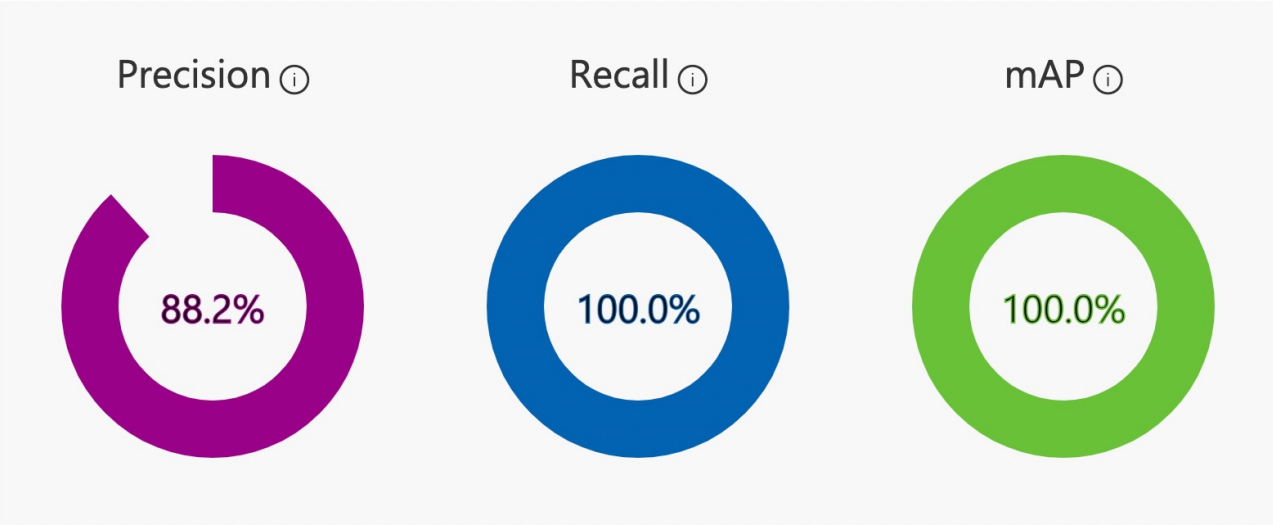
Iteration 1

Trained : moments ago with General [A1] domain

Publish | Prediction URL | Delete | Export

Iteration 1

Finished training on **12/4/2022, 7:15:00 PM** using **General [A1]** domain
Iteration id: **50f8dbbb-72eb-4e17-b689-b946d3d5e030**



Performance Per Tag

Tag	Precision	Recall	A.P.	Image count
Apple	100.0%	100.0%	100.0%	18
Orange	85.7%	100.0%	100.0%	20

100%

Get started

Grocery Detection | Training Images | Performance | Predictions | Train | Quick Test

Iterations

Probability Thresh

Overlap Threshold:

Iteration 1

Trained : 7 minutes

General [A1] domai

Quick Test

Regions Shown

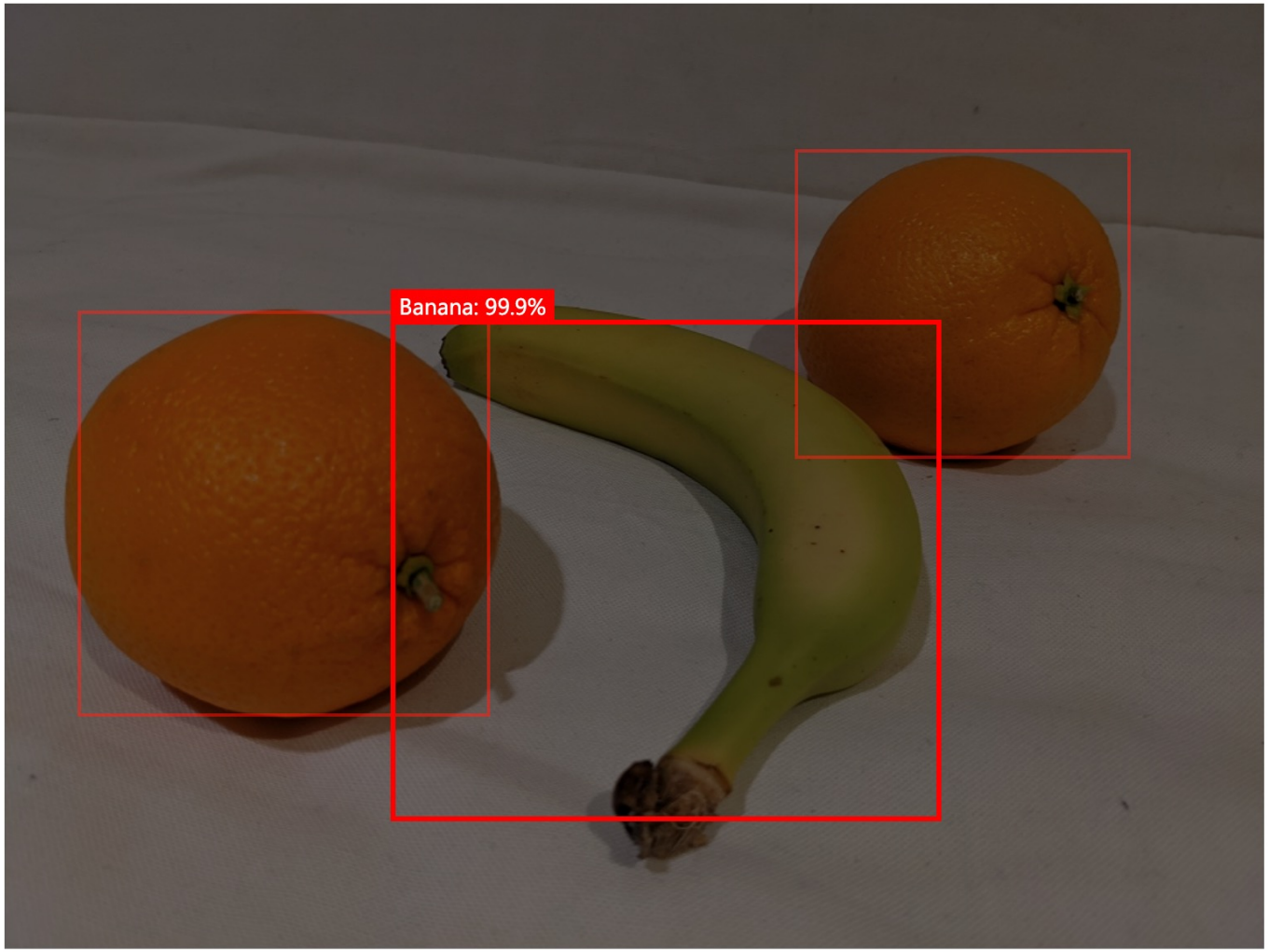


Image URL

https://aka.ms/apple-orang

or

Browse local files

File formats accepted: jpg, png, bmp
File size should not exceed: 4mb

Using model trained in

Iteration

Iteration 1

Predicted Object Threshold

Only show suggested objects if the probability is above the selected threshold.

Threshold Value: 91%

Predictions

Predictions are shown in red

Tag	Probability

Microsoft Azure | Search resources, services, and docs (G+/) | sneha51189@gmail.com DEFAULT DIRECTORY

Home > Resource groups >

Resource groups
Default Directory

Create | Manage view | ...

Filter for any field...

< Page 1 of 1 >

ComputerVision
Resource group

Search

Overview
Activity log
Access control (IAM)

Create | Manage view | Delete resource group | Refresh | Export to CSV | Open query | ...

Name	Type	Location
ObjDetectTraining	Custom vision	East US 2
ObjectDetectionPrediction	Custom vision	East US 2

PowerShell

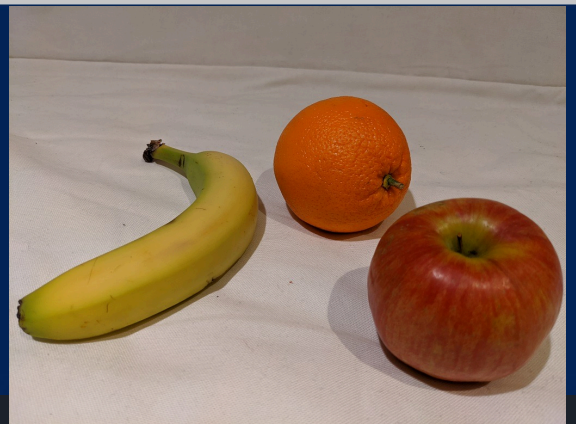
detect-objects.ps1

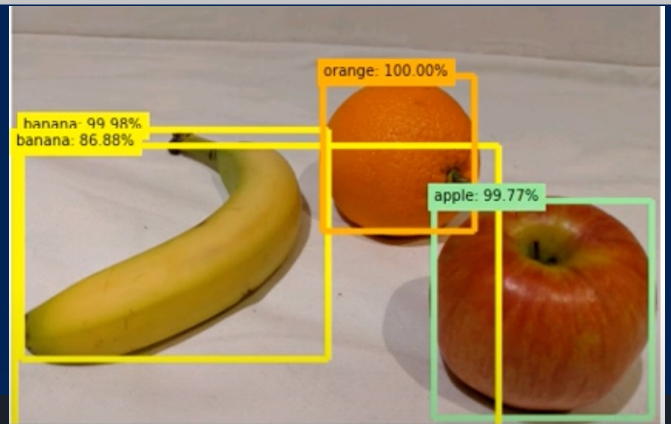
FILES

- analyze-text.ps1
- classify-image.ps1
- detect-anomalies.ps1
- detect-objects.ps1
- find-faces.ps1
- form-recognizer.ps1
- index.md
- LICENSE
- mapping.md
- ocr.ps1

```
1  $predictionUrl = "https://eastus2.api.cognitive.microsoft.com/customvision/v3.0/Prediction/72473436-af2b-4612-8b33-70e2b73587ca/d
2  $predictionKey = "6048cb5062d6418dbbd2cf6d9bad52ed"
3
4
5  # Code to call Custom Vision service for image detection
6
7  $img = "https://raw.githubusercontent.com/MicrosoftLearning/AI-900-AIFundamentals/main/data/vision/produce.jpg"
8
9  $headers = @{}
10 $headers.Add( "Prediction-Key", $predictionKey )
11 $headers.Add( "Content-Type","application/json" )
12
13 $body = "{'url' : '$img'}"
14
```

```
PS /home/sneha/ai-900>  ./detect-objects.ps1
Analyzing image...

 Apple


 Banana


 Orange

PS /home/sneha/ai-900>
```





Summary

- Object Detection
  - Defining the object detection problem and a naïve solution
- YOLO Algorithm
  - YOLO Algorithm Steps
  - Bounding Boxes
  - Measuring Performance (IoU)
- Azure Custom Vision for Object Detection
  - Train and Test Model
  - Publish it using the endpoint in your application

References

- Azure Custom Vision - https://azure.microsoft.com/en-us/products/cognitive-services/custom-vision-service/
- Azure Custom Vision - https://learn.microsoft.com/en-us/azure/cognitive-services/custom-vision-service/overview
- YOLO V1 - https://arxiv.org/abs/1506.02640
- YOLO V2 - https://arxiv.org/abs/1612.08242v1
- YOLO V3 - https://arxiv.org/abs/1804.02767v1
- YOLO V4 - https://arxiv.org/abs/2004.10934v1
- YOLO V5 - https://pytorch.org/hub/ultralytics_yolov5/
- YOLO V5 Documentation - https://docs.ultralytics.com/