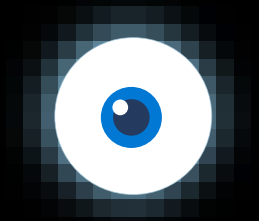
Azure AI Vision

Microsoft

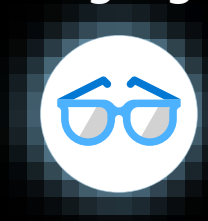
Accelerating meaningful AI innovation

Vision



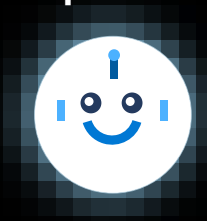
2016
Object recognition human parity

Language



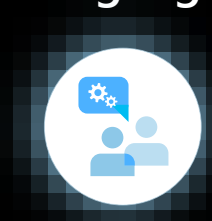
2018
Reading comprehension human parity

Speech



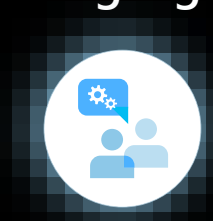
2018
Speech synthesis near-human parity

Language



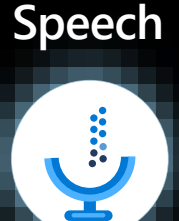
2020
Document summary at human parity

Language



2021
Natural language understanding human parity

Speech



2017
Speech recognition human parity

Language

2018
Machine translation human parity

Language

2019
General Language Understanding human parity

Vision

2020
Image captioning human parity

Decision

2021
CommonsenseQnA

Azure AI

Applications

Microsoft 365

Microsoft Dynamics 365

Partner Solutions

Application Platform

AI Builder

Power BI

Power Apps

Power Automate

Power Virtual Agents

Business Users

Customizable Models and Scenario-Based Services

Azure AI services

Vision

Speech

Language

Decision

Azure OpenAI Service

Bot Service

Cognitive Search

Document Intelligence

Video Indexer

Metrics Advisor

Immersive Reader

Developers & Data Scientists

ML Platform

Azure Machine Learning

A Glimpse of Diverse Azure AI Vision Tasks

Image Classification

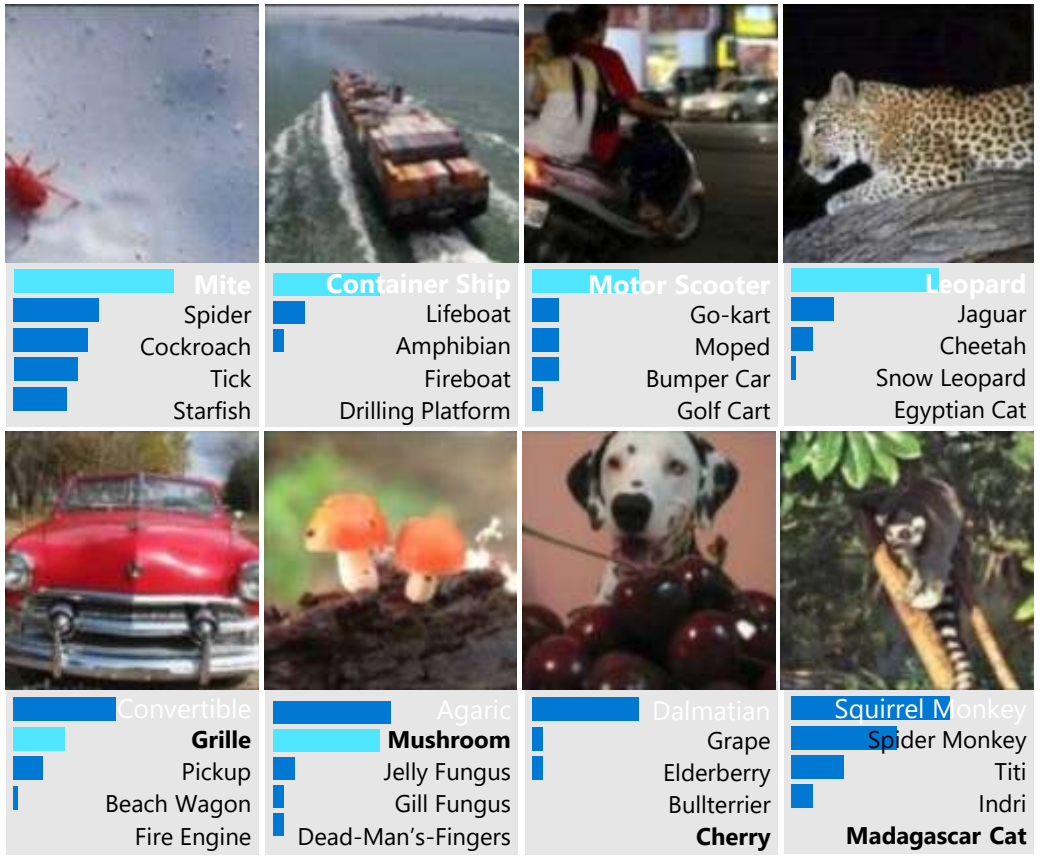
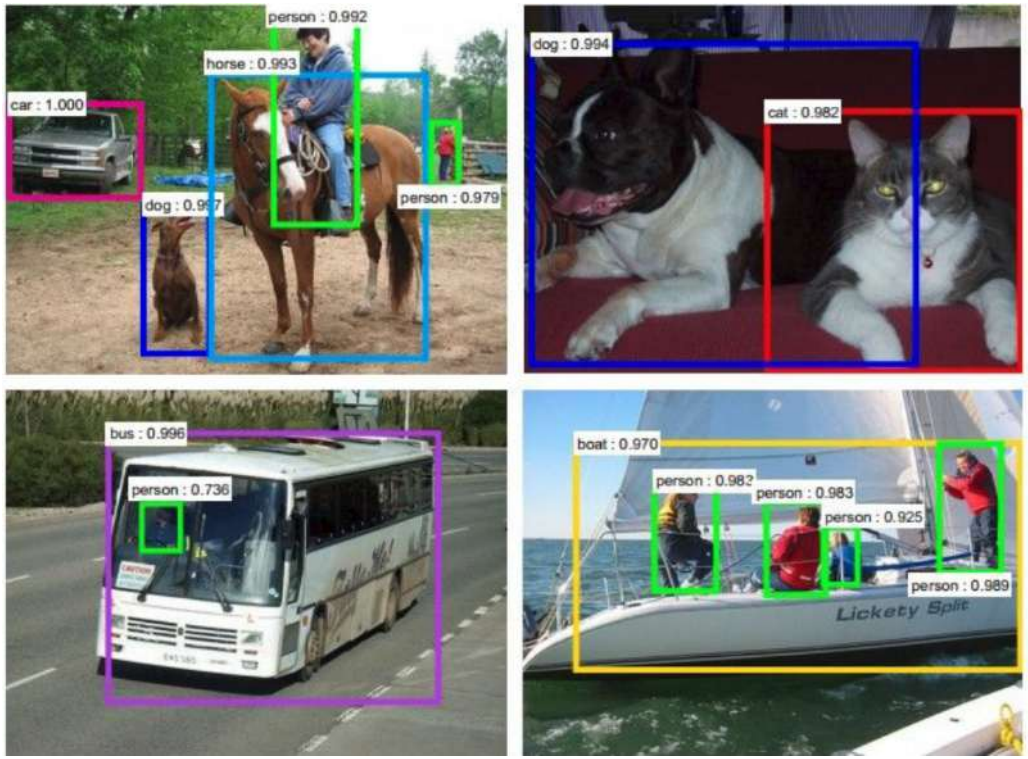


Image Retrieval

Krizhevsky, Sutskever, and Hinton, 2012

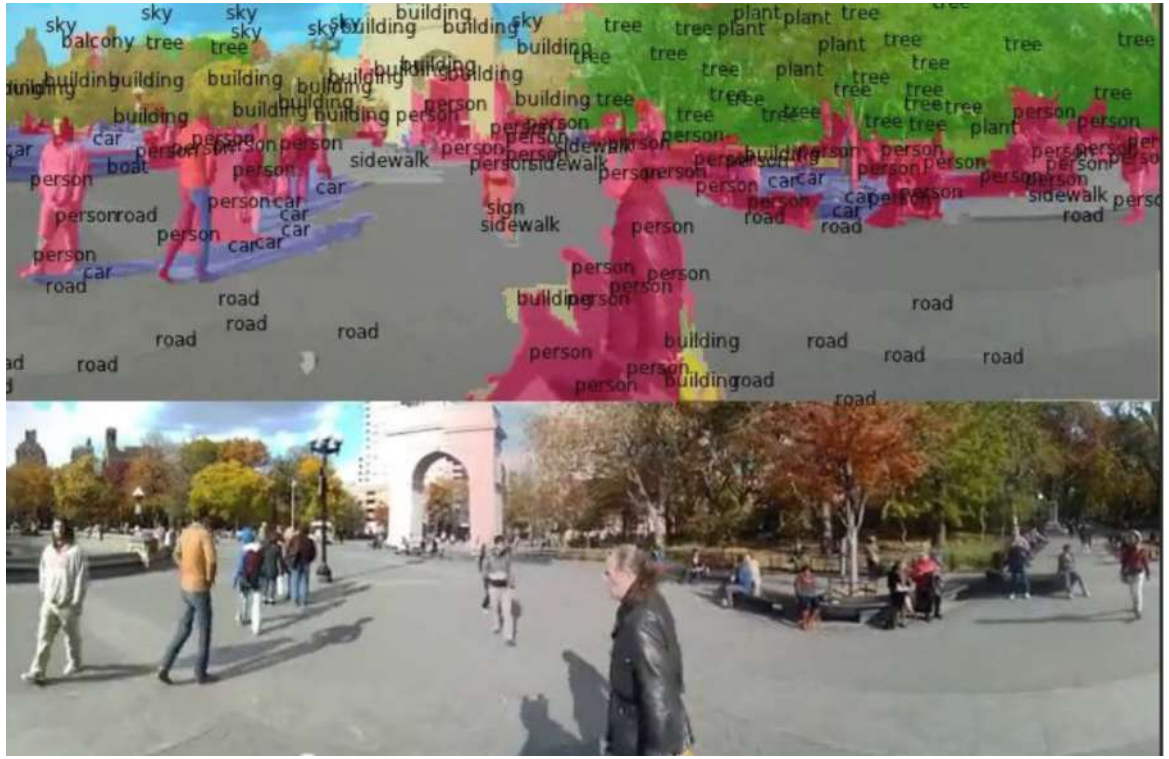
A Glimpse of Diverse Azure AI Vision Tasks

Object Detection



Ren, He, Girshick, and Sun 2015

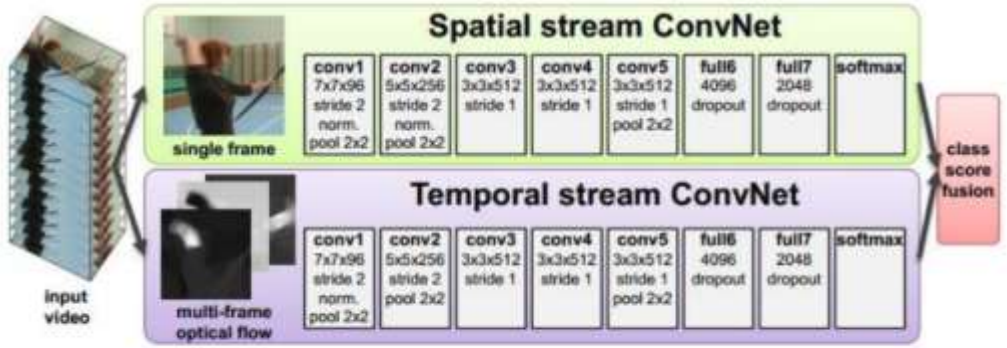
Image Segmentation



Fabaret et al, 2012

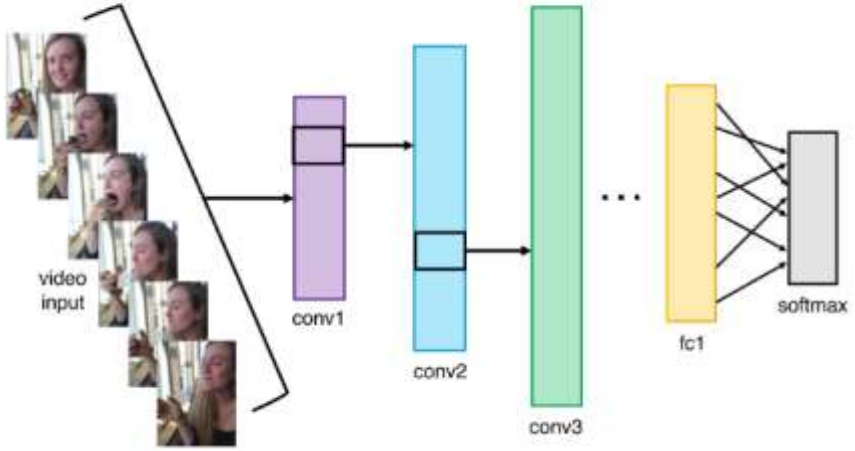
A Glimpse of Diverse Azure AI Vision Tasks

Video Classification



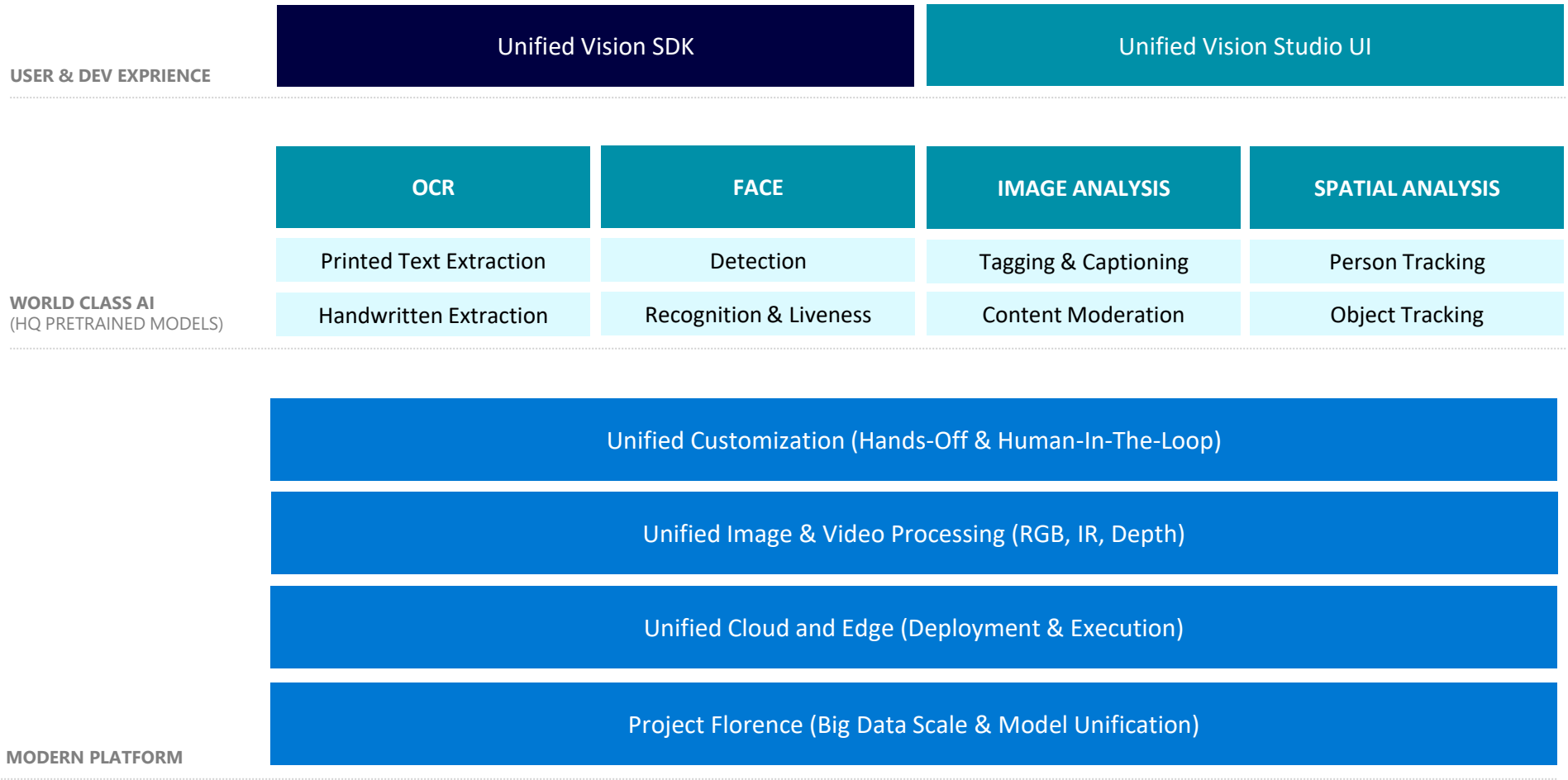
Simonyan et al, 2014

Activity Recognition



Pose Recognition (Toshev and Szegedy, 2014)

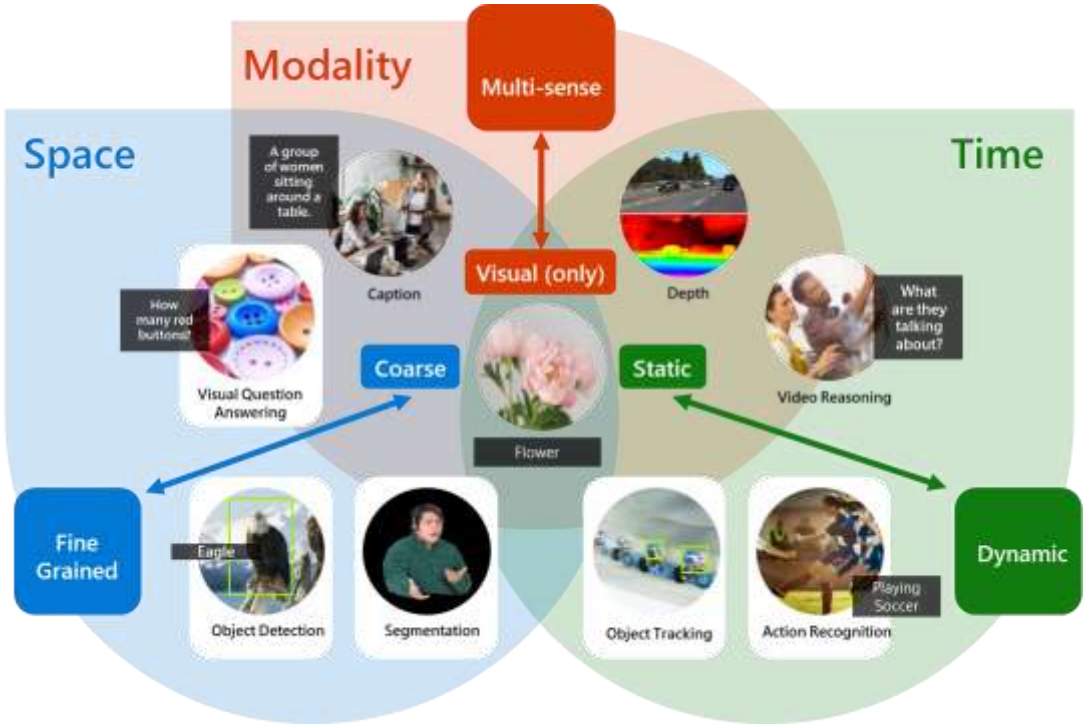
Azure AI Vision

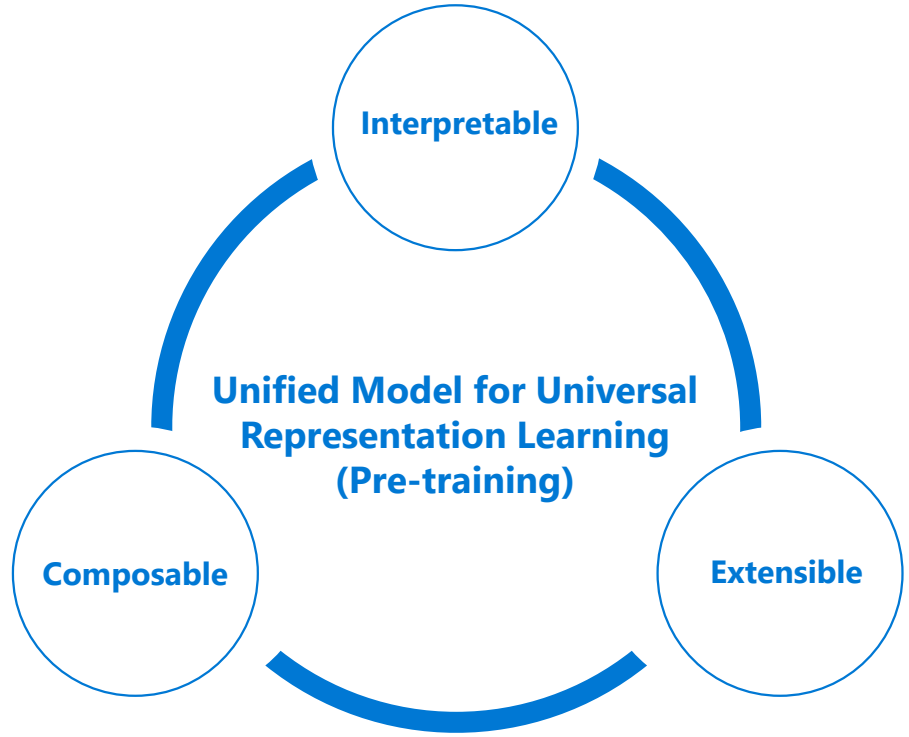


Microsoft

Florence: Towards Unified Large Scale Multi-task Multi-modality Vision Representation

End: multimodality video understanding
(ultimate of Florence)





Origin: image classification
(core vision problem)


a green vehicle with the word elephant on the door
a man is driving a car and the driver is holding a microphone.
a rhino is standing on the road and the other one is on the
a dog
a black and white image of three elephants with the word s on it.
a man
a black and white image of a horse with the word elephant on it
Tel: +27 (82) 4288 223
Africa
Adventures

IMAGE ANALYSIS

Image Analysis with Azure AI Vision

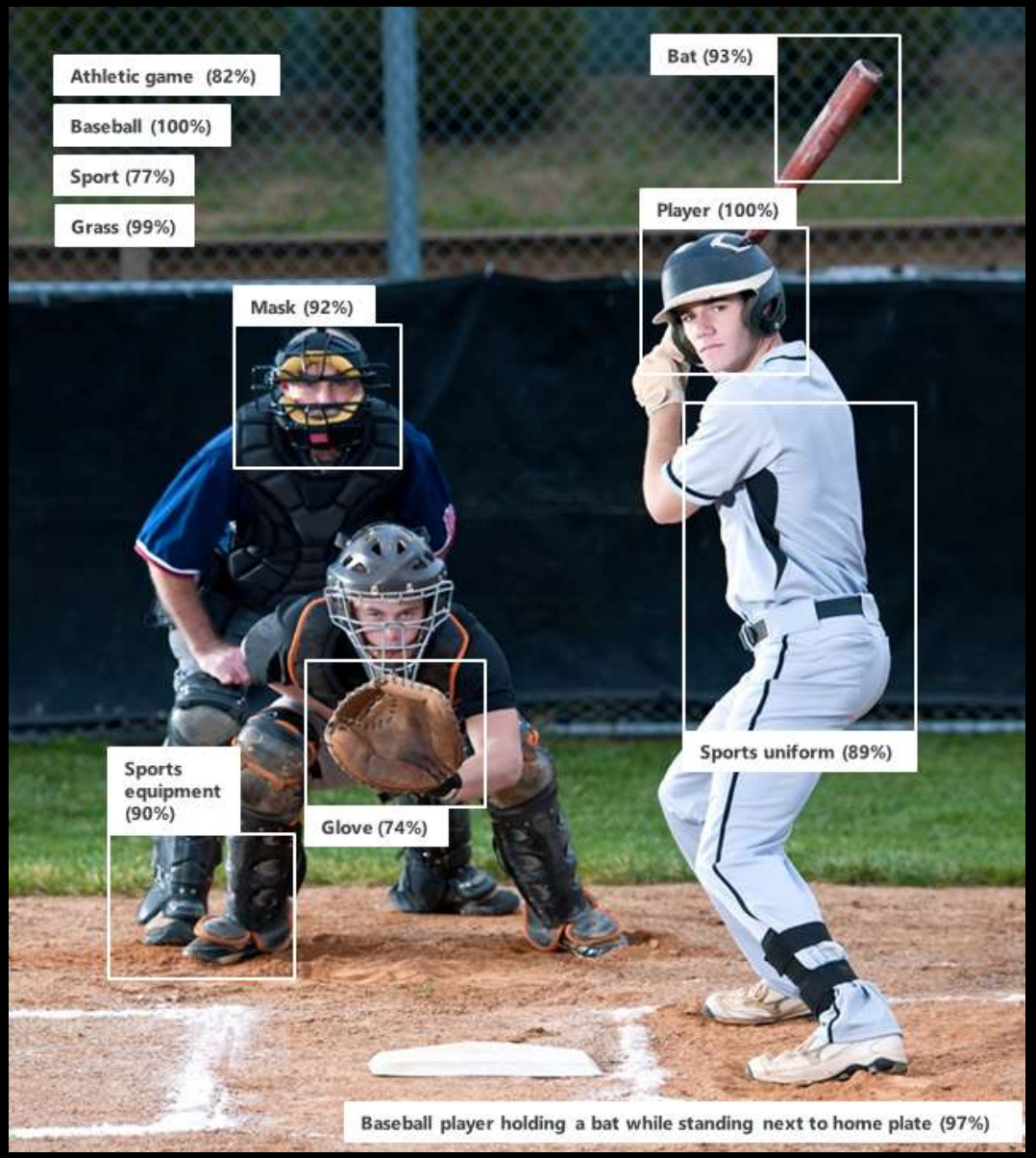


Image Tagging:

- Returns tags based on thousands of recognizable objects, living beings, scenery, and actions
- Significant quality improvements generally available since April 2021

Image Captioning:

- Automatically extracts contents in the image and generates human-readable sentences as description
- Language support expansion coming by fall 2021

Adult Content Detection:

- Detects presence of adult, racy and gore contents in the images and provides a flag for content moderation
- Model updates with accuracy improvements generally available since April 2021

Object Detection:

- Detects objects in images

Image Captioning with Azure AI Vision

Previous Production
A man in a blue shirt

Latest 3.2 GA
A few people wearing surgical masks.

Previous Production
A close-up of a person cooking hot dogs on a cutting board.

Latest 3.2 GA
A person making bread.

Previous Production
A close-up of a plant.

Latest 3.2 GA
A close-up of wheat in a field.

Image Tagging with Azure AI Vision

Previously
flower; nature; grass; plant

3.2 GA
plant; dandelion; green; field; grass; sky; sun; day; outdoor; flower; grassy; spring; landscape; nature

Previously
person; screenshot; text

3.2 GA
person; health care; medical; clothing; medical equipment

Previously
Person; man; clothing

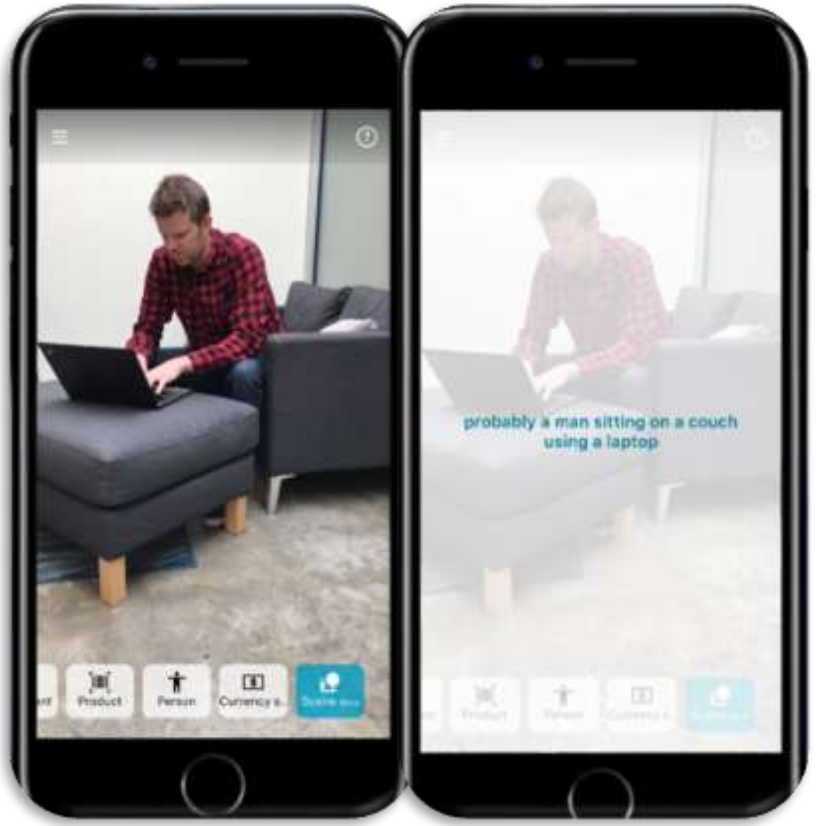
3.2 GA
clothing; man; human face; person; screenshot, metalworking, welding, welder, metalsmith, factory

Microsoft Azure

Microsoft Seeing AI App

Hear descriptive audio everywhere

Seeing AI brings together the power of the cloud and AI to deliver an intelligent app that helps you navigate your day with the help of narration describing people, text, and objects.



CV 3.2 API: a person wearing red shoes

New version: a child playing hopscotch

CV 3.2 API: a person in a boat on water

New version : a woman in a canoe touching water with a man in the back

CV 3.2 API: a person holding a pole

New version: a group of football players holding a football

CV 3.2 API: a close-up of a pipe

New version: a sewing machine with a needle and thread

CV 3.2 API: a woman writing on a whiteboard

New version: a woman pointing at a screen

CV 3.2 API: a group of hands holding a stick with food on it

New version: a table with pasta and a rolling

CV 3.2 API: a large collection of batteries

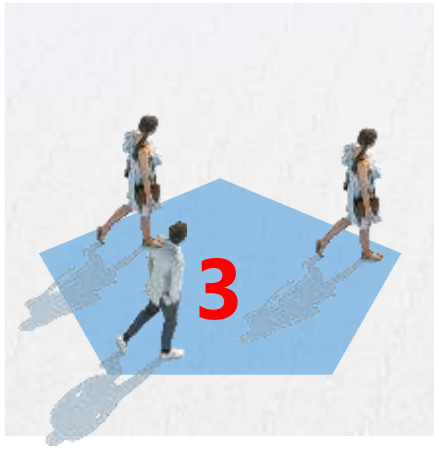
New version: a parking lot full of cars

CV 3.2 API: a woman and a man holding hands

New version: a woman sitting in a yoga pose

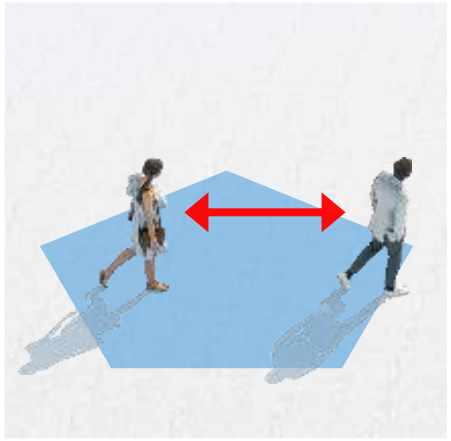
SPATIAL ANALYSIS – PERSON UNDERSTANDING

Vision Platform Primitives

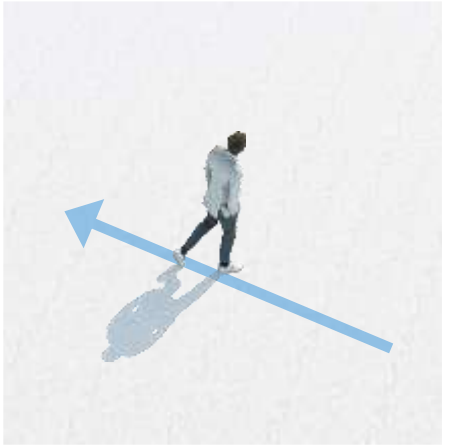


Person Count in a Polygon

Person Crossing In/Out of a Polygon



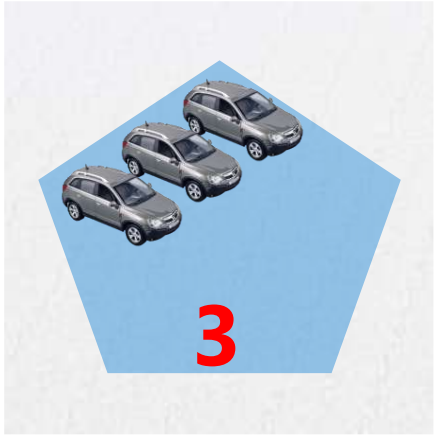
Social Distance Threshold



Person Crossing Directional Line - Entry/Exit

Person Classification (PPE or Uniform Classification)

Vehicle in Polygon



Vehicle Count

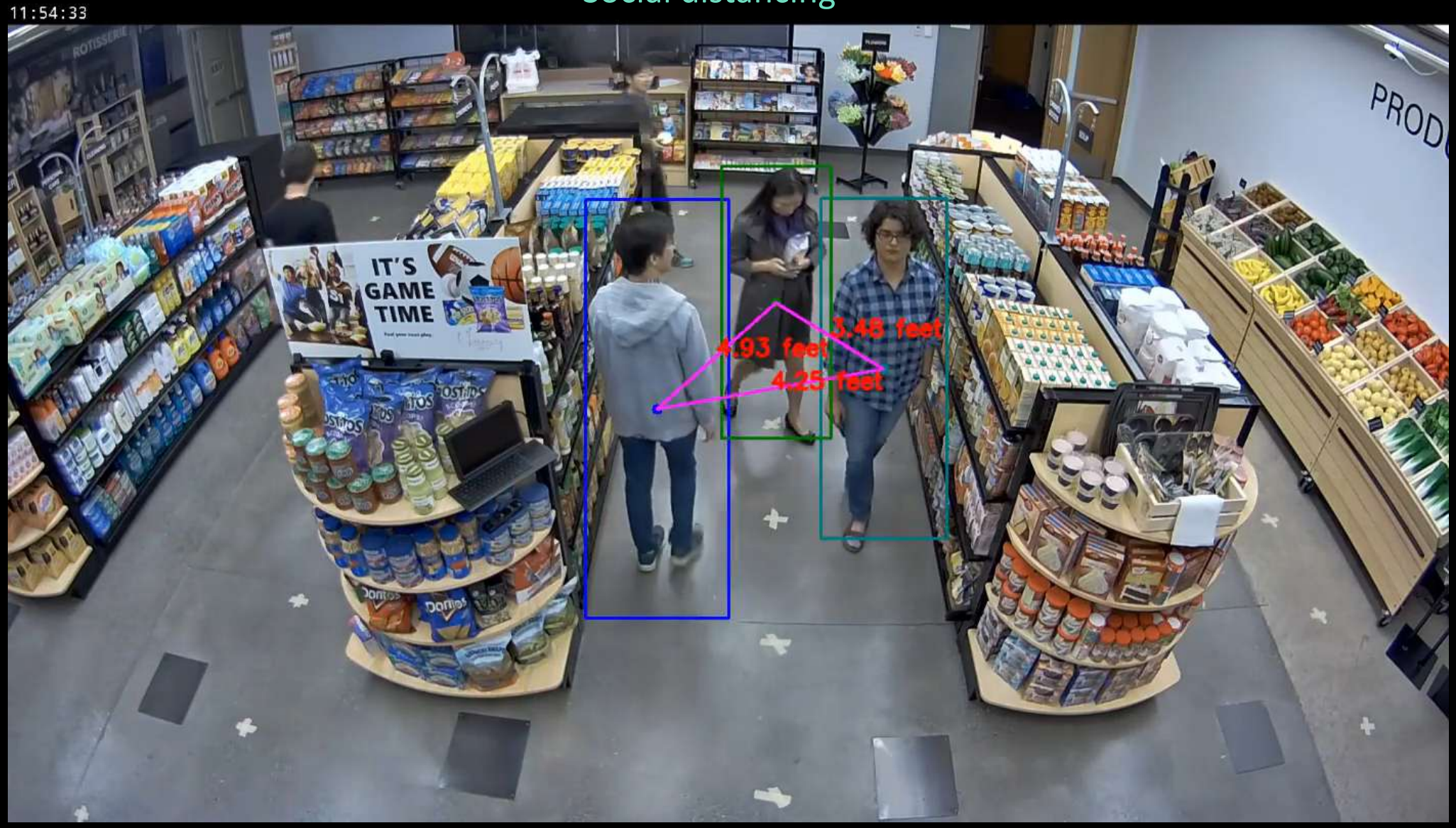
Person Near Moving Vehicle

Vehicle Type Classification

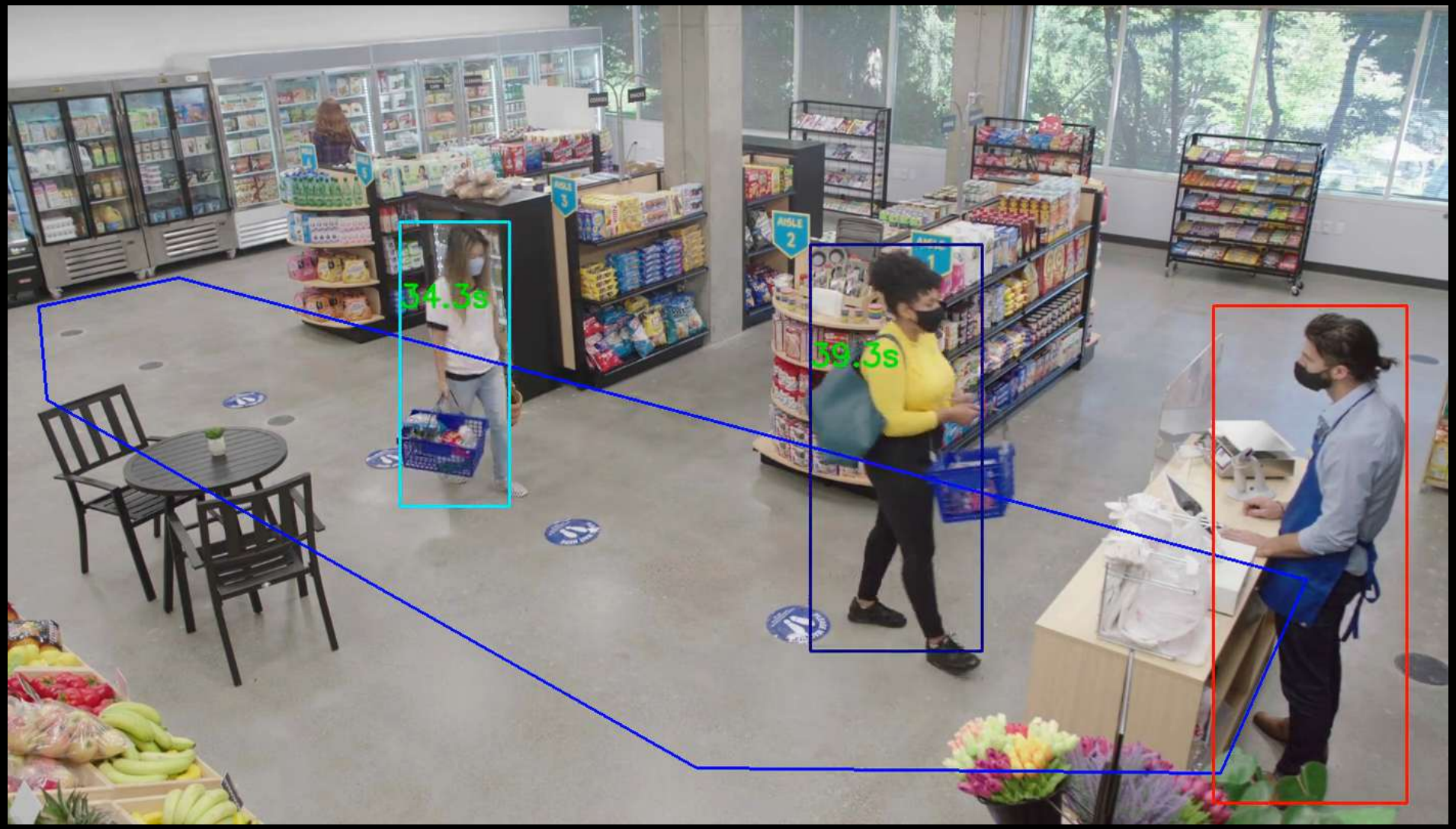
People Detection and Tracking

People Detection and Tracking

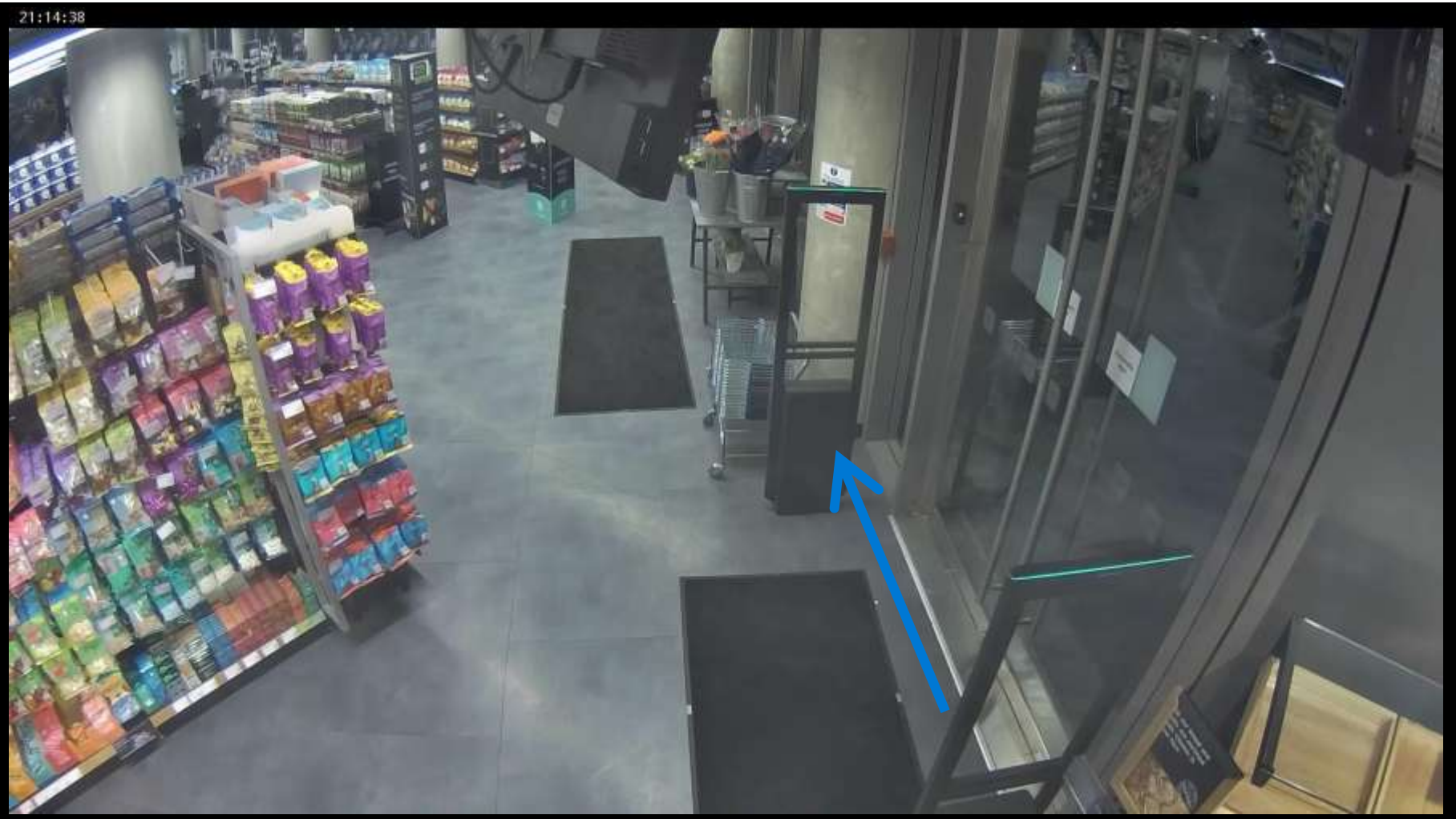
Social distancing



Entry/Exit Spaces



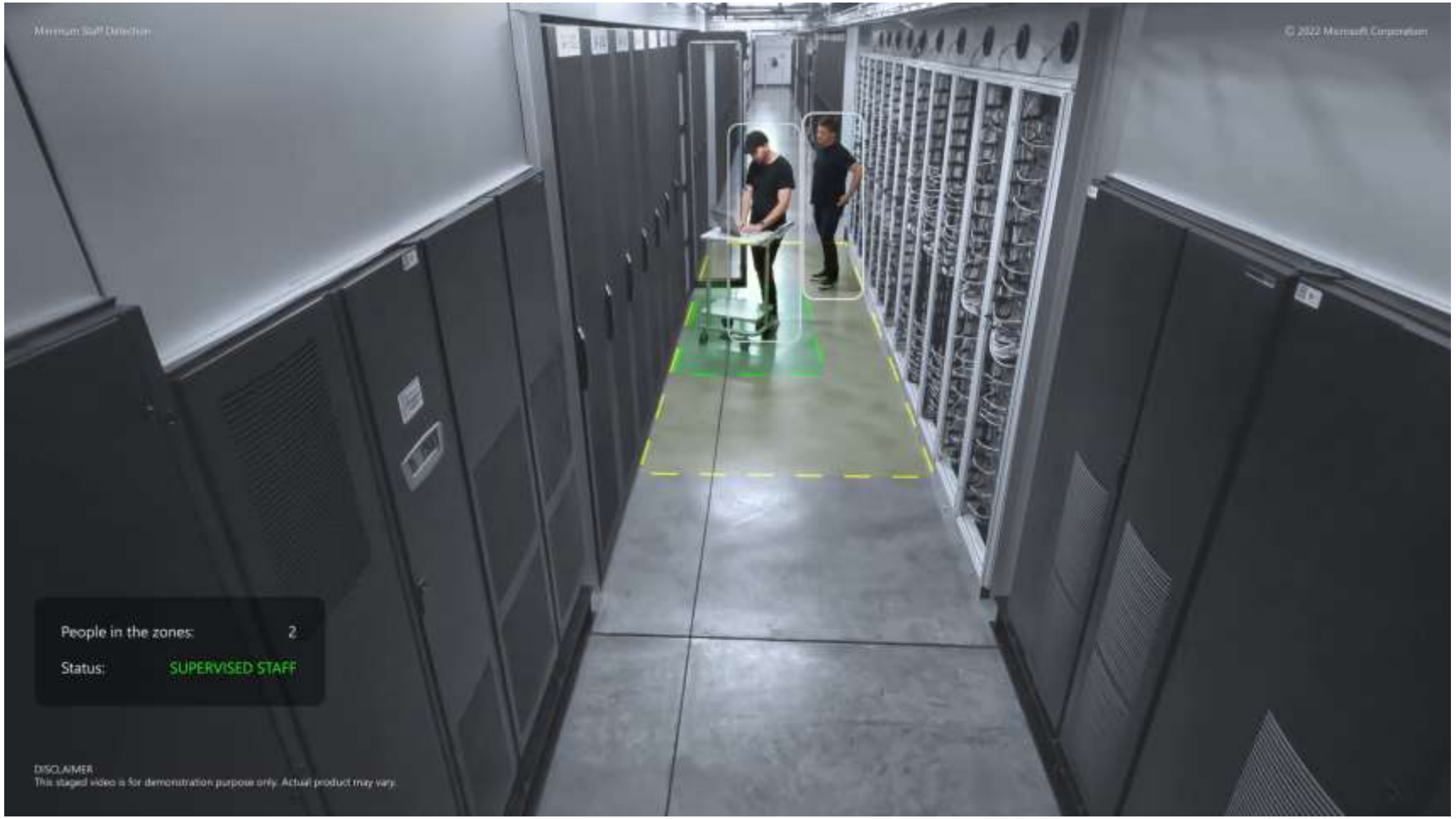
Line Configuration



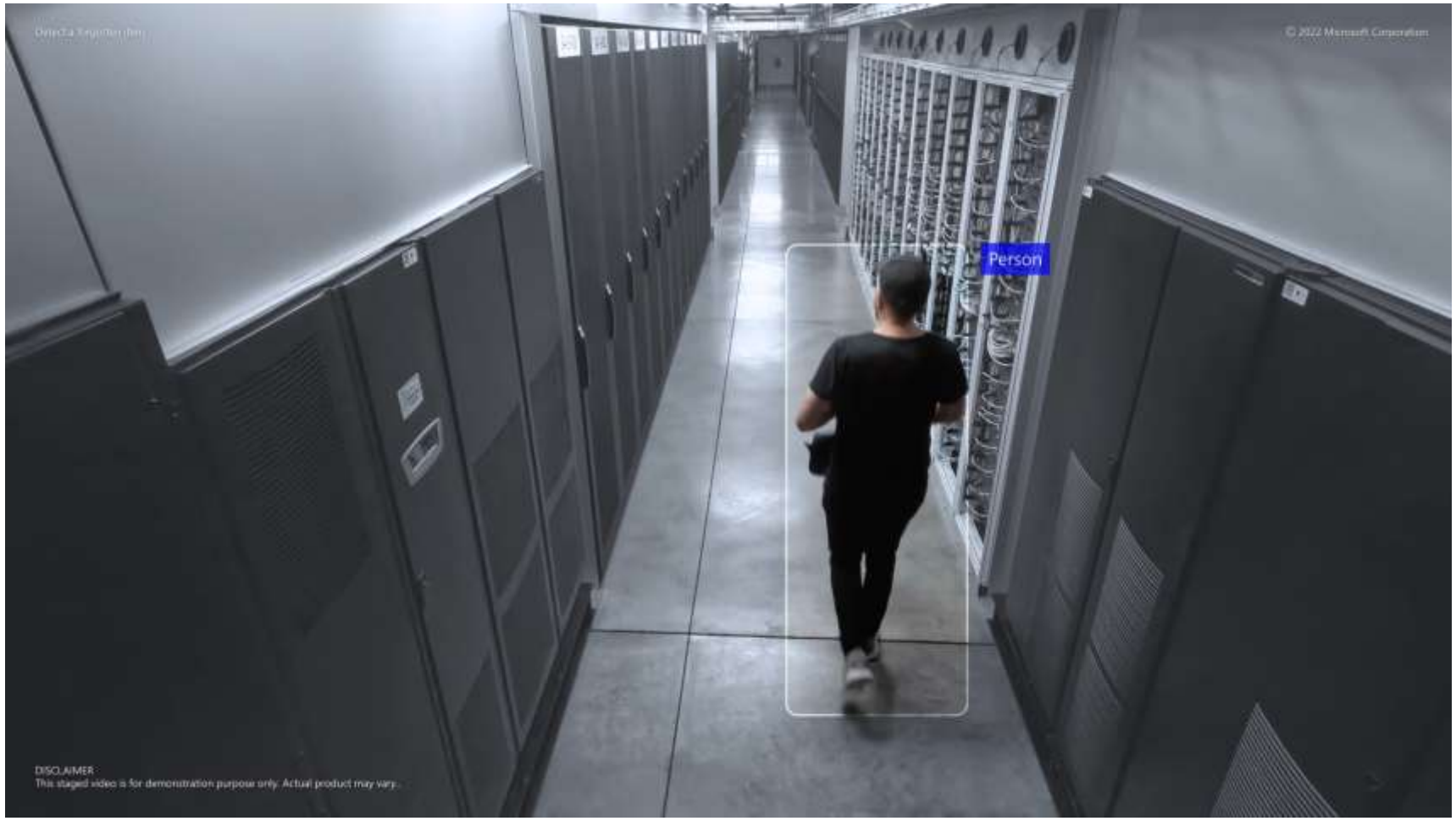
Enter/Exit through doorways

Tailgating

Minimum Staff Detection



Forgotten Item

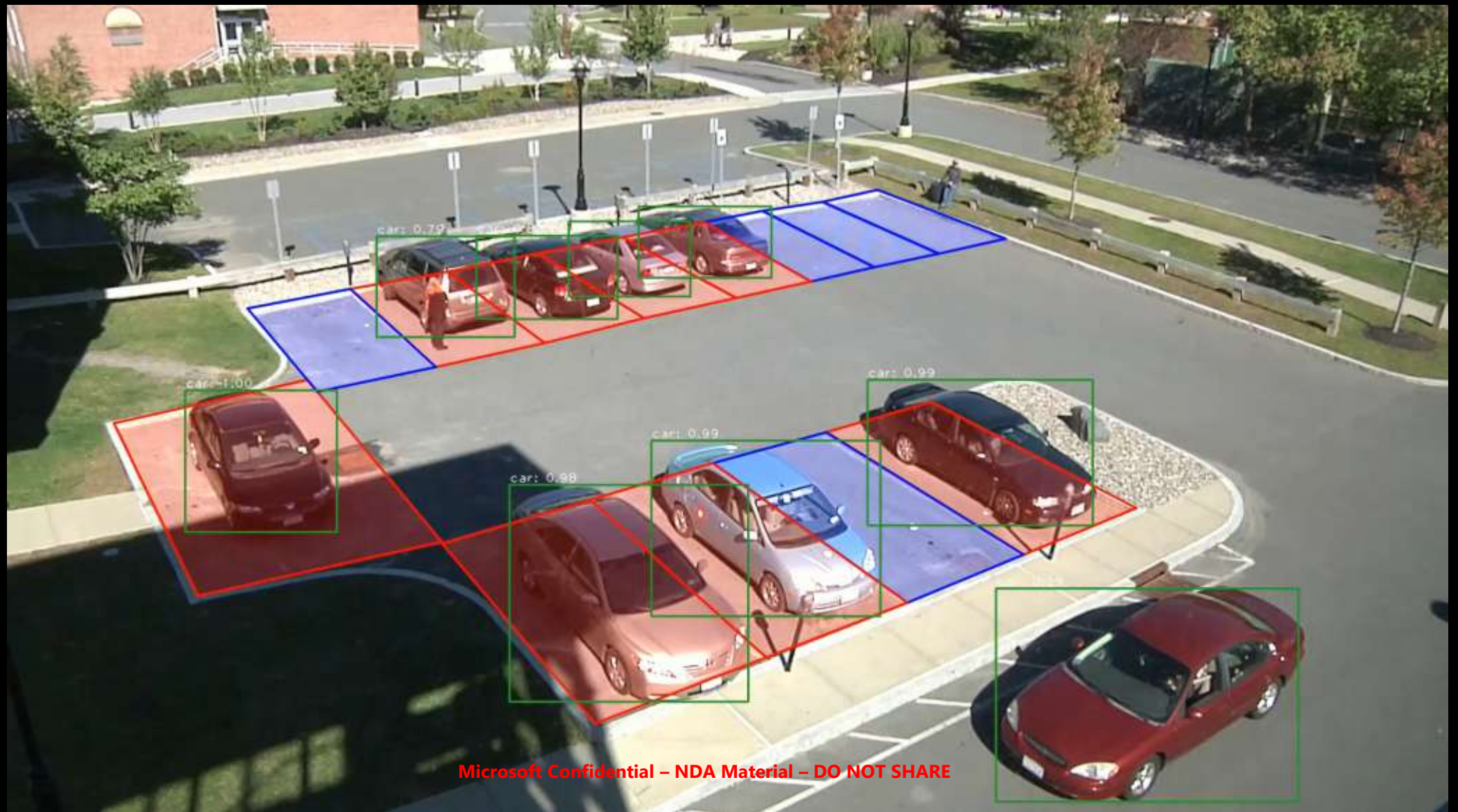


Collision Detection

Running In A Hazardous Area

Mapping Capabilities

Spatial Analysis for Vehicles at the Edge Public Preview March 2022



Person Detection for Workplace safety

Robust AI under harsh conditions

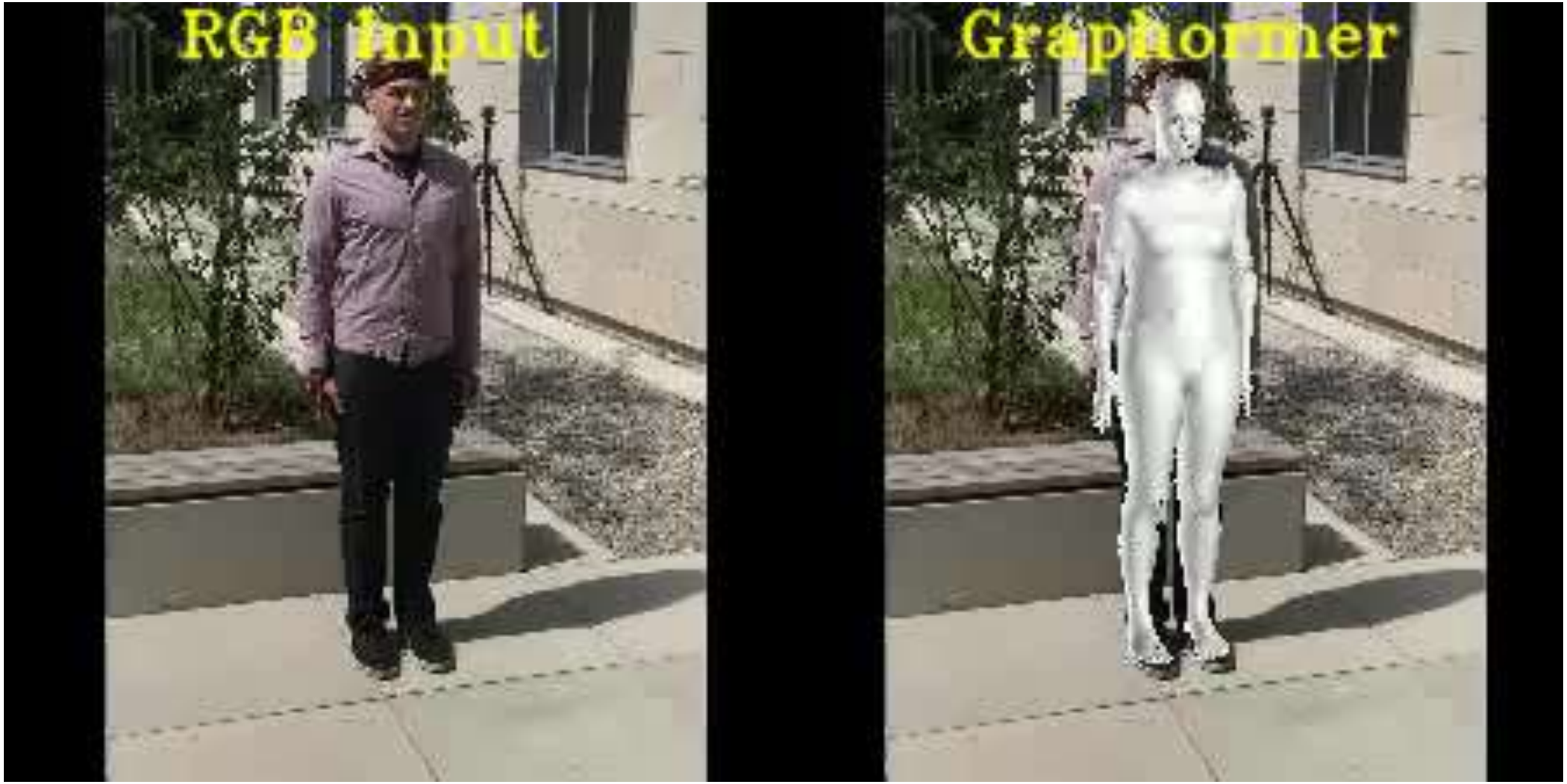
Multiple **moving object** detection

Fine detail detection

SPATIAL ANALYSIS - BODY TRACKING

Body Tracking running in Azure Cloud for Depth cameras

3D Body Tracking from Single RGB Camera



IDENTITY ANALYSIS

Microsoft Azure

Azure Facial Biometrics Products & Services

Part of Windows 10 & 11

Gives users a personal, secured experience where the device is authenticated based on their presence. Users can log in with a look or a touch, with no need for a password

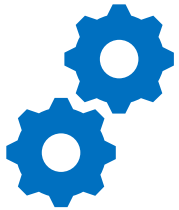
Azure Cloud Service Face APIs

Face detection
Detect one or more human faces along with attributes such as pose, face coverings, and face location, including 27 landmarks for each face in the image.

Face verification
A "one-to-one" matching of a face in an image to a single face from a secure repository or photo to verify they are the same individual, using unmanipulated images.

Face identification
"One-to-many" matching of a face in an image to a set of faces in a secure repository.

Vision Edge SDK

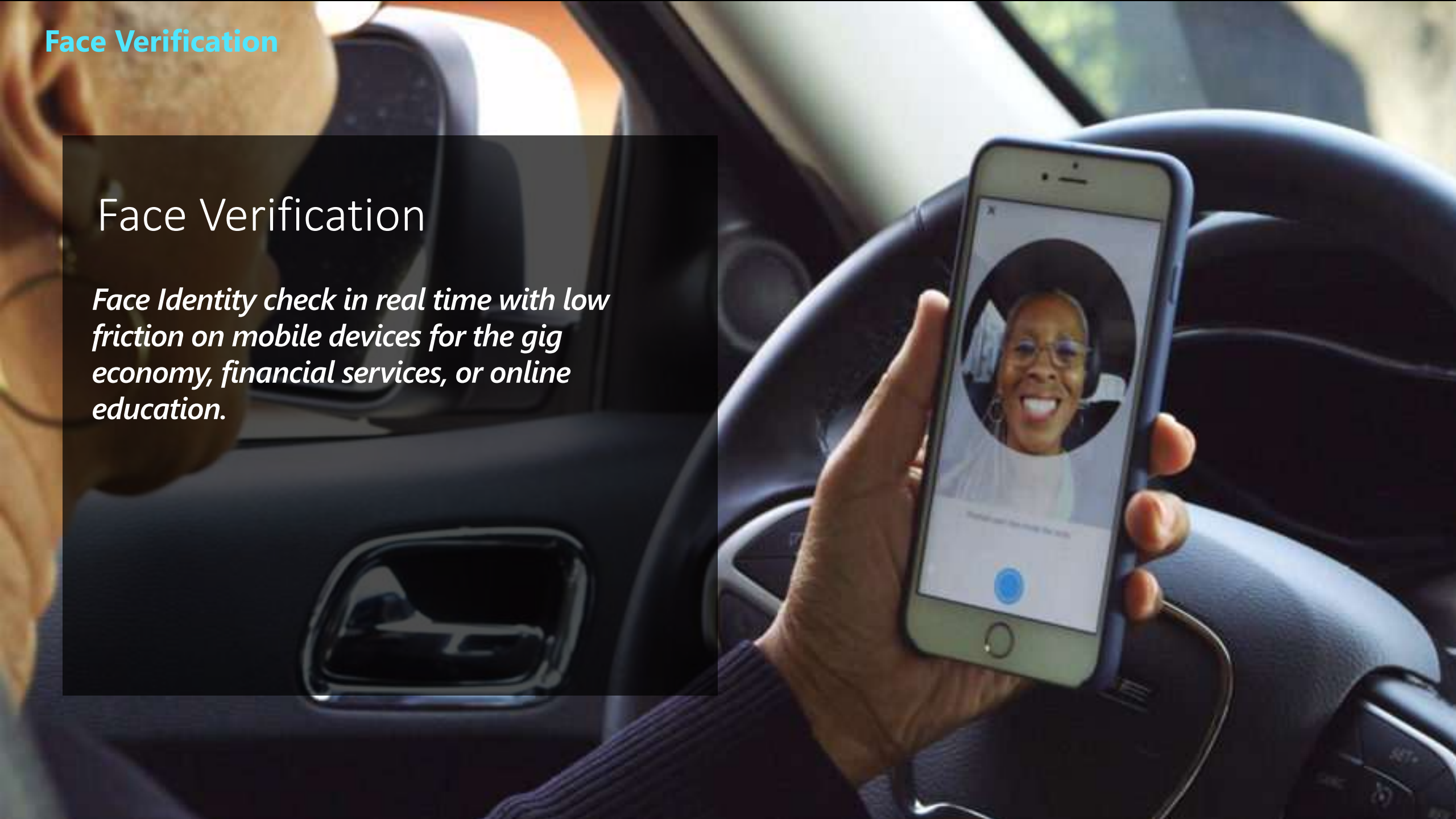


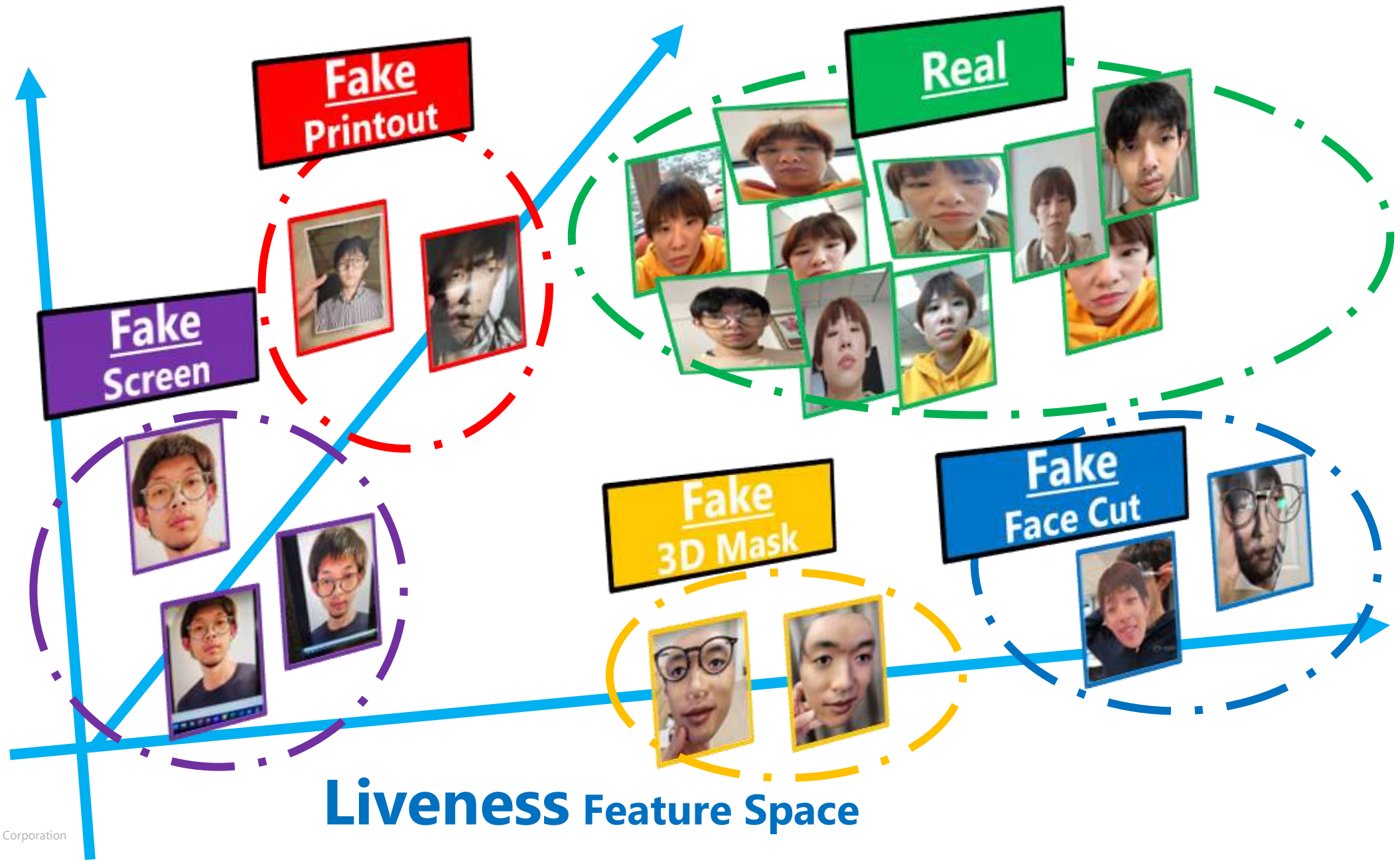
Liveness (requires Cloud)
works with facial recognition to determine if a biometric sample is being captured from a living subject who is present at the point of capture.

Edge Components
Face detection, face tracking, quality filter and camera connection support to enable developers to put their sole focus on core business logic

Face Verification

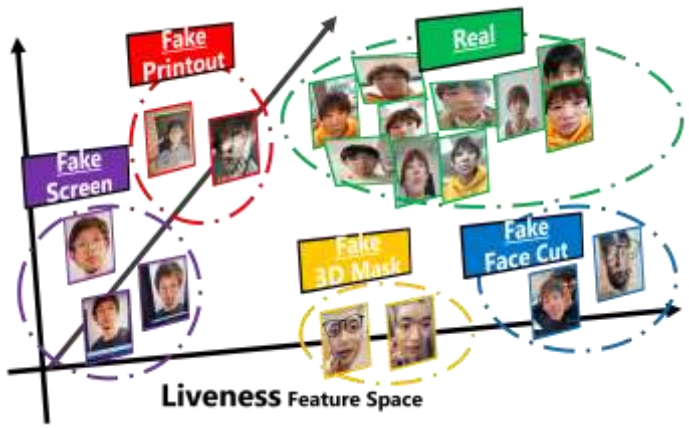
Face Identity check in real time with low friction on mobile devices for the gig economy, financial services, or online education.

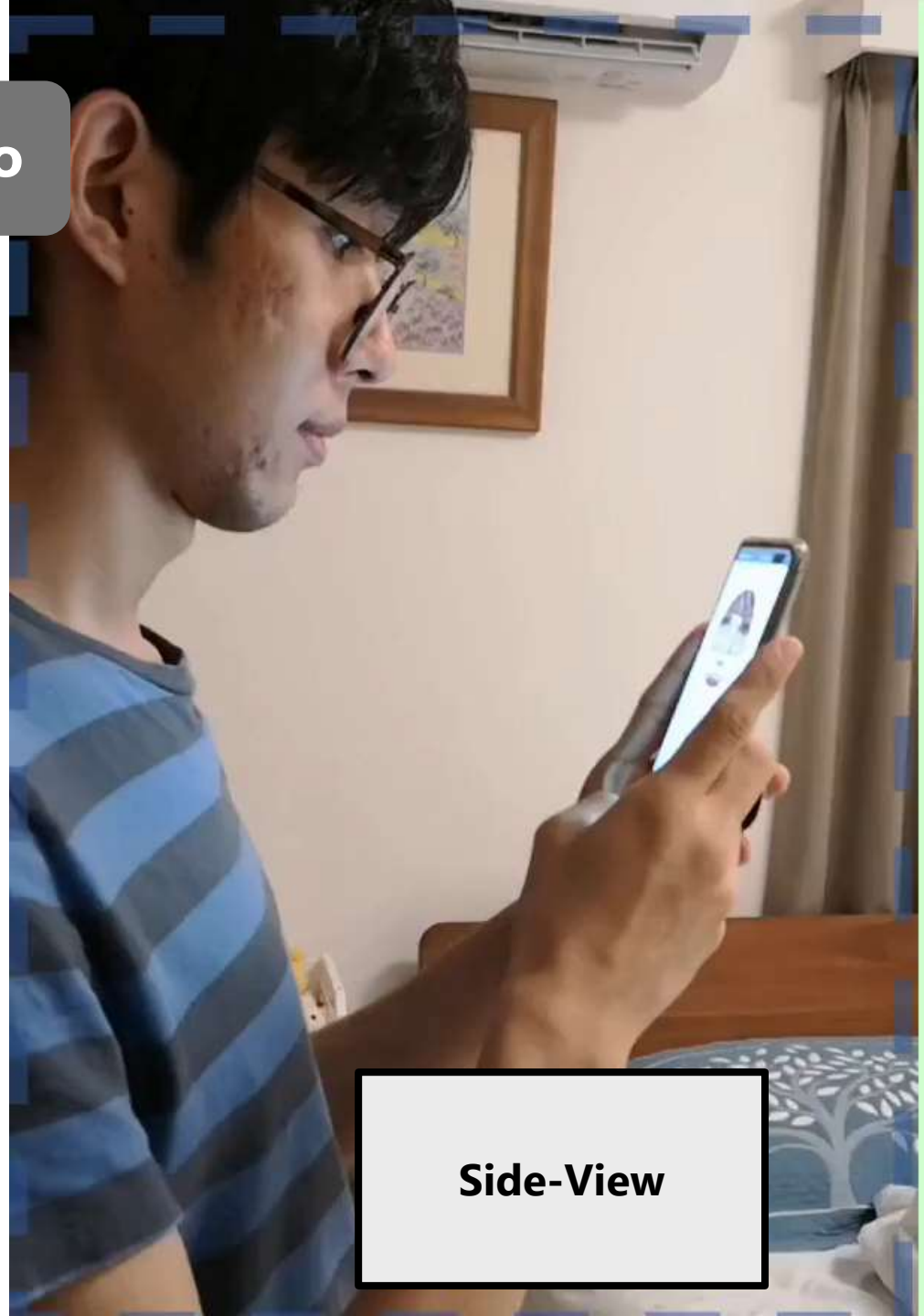




Liveness Feature Space

Liveness Service Demo





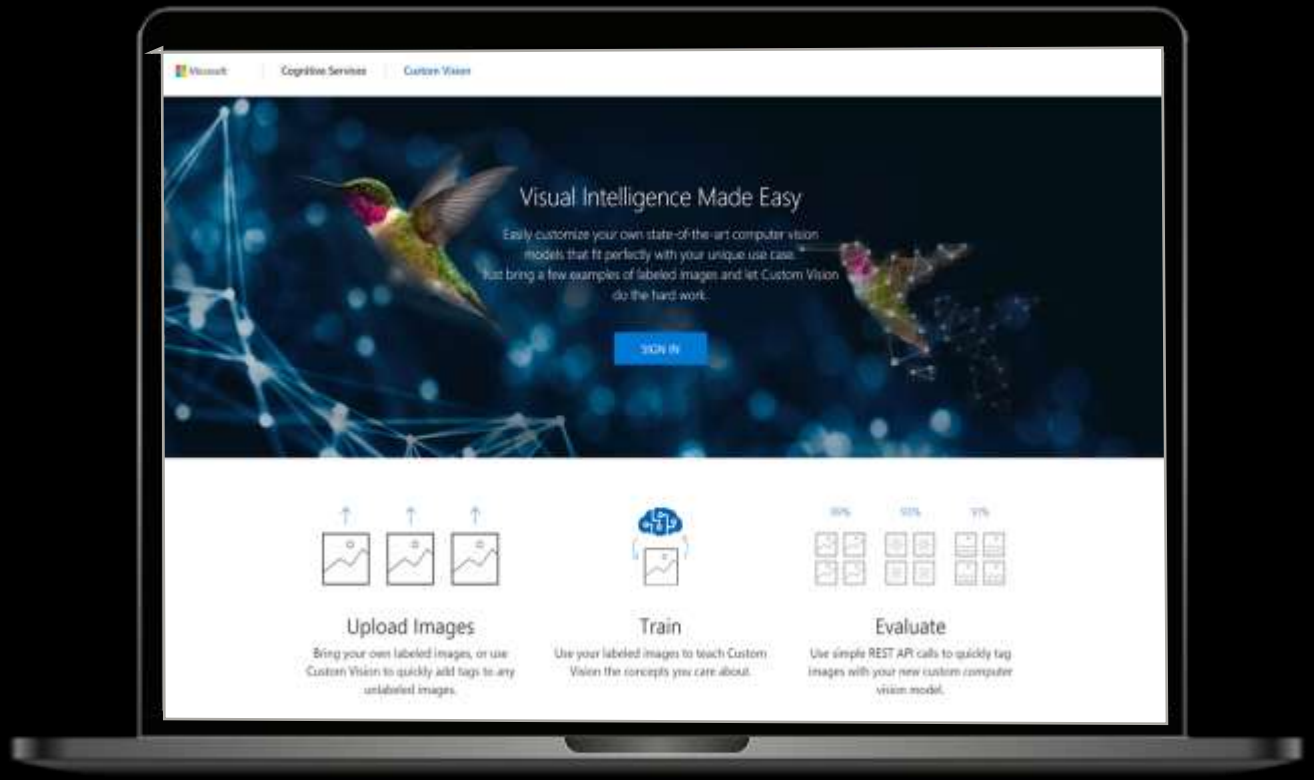
Side-View

Completed !

Liveness Service
iOS/Android
Client + Cloud

MODEL CUSTOMIZATION

AZURE AI CUSTOM VISION



Azure AI Custom Vision Web Portal

- No code

Submit groups of images, label images, model trains data

AZURE AI CUSTOM VISION

Image classification vs Object Detection

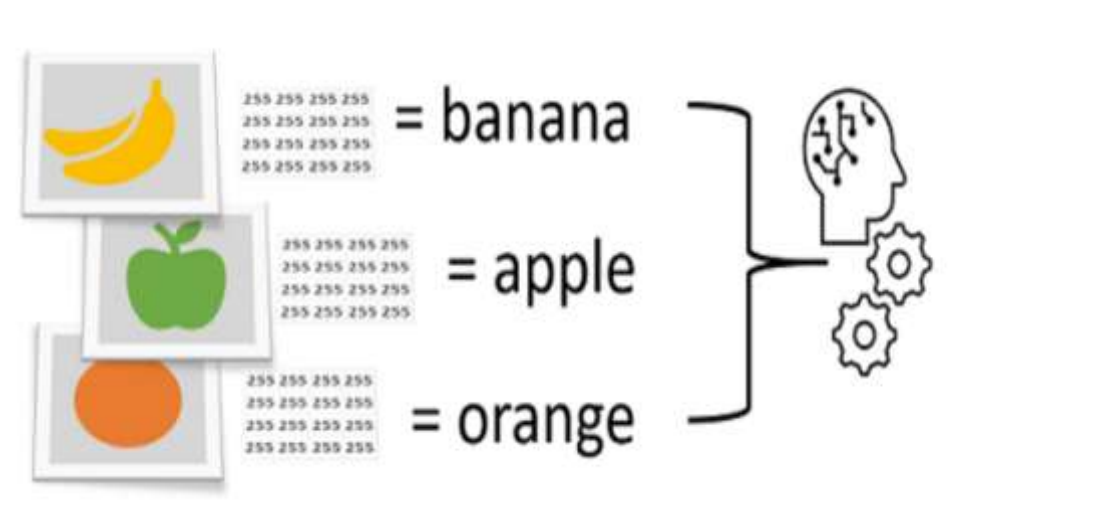


Image classification – a ML technique in which the object being classified is an image such as a photograph.

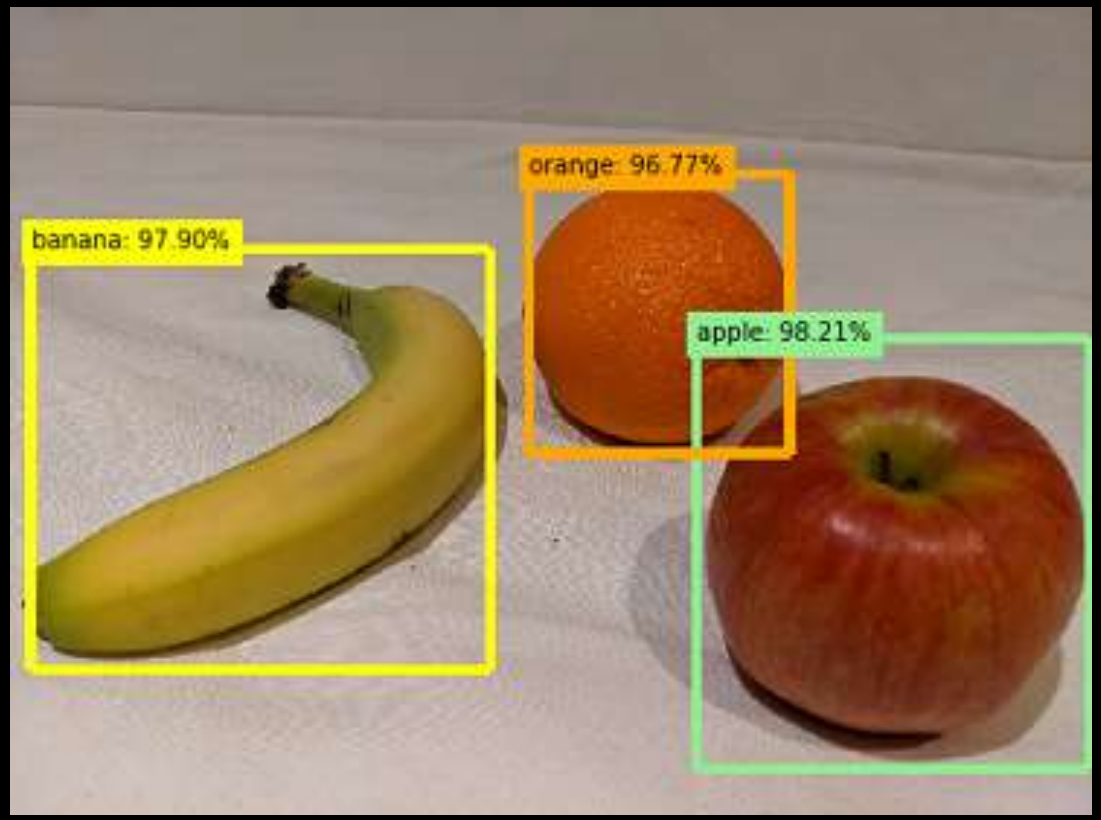
Training a Model

Upload images to your training resource and label them with the appropriate class labels.
Then, train the model and evaluate the training results.

Result returns a *probability score* and *class* for the image

AZURE AI CUSTOM VISION

Image classification vs Object Detection



Object Detection– a ML technique in which a model is trained to recognize individual types of objects in an image, and to identify their location in the image.

Training a Model

Upload images to your training resource and label them with the appropriate class labels.
Then, train the model and evaluate the training results.

Results return *class*, *probability score* and *bounding box* for each object

AZURE AI CUSTOM VISION

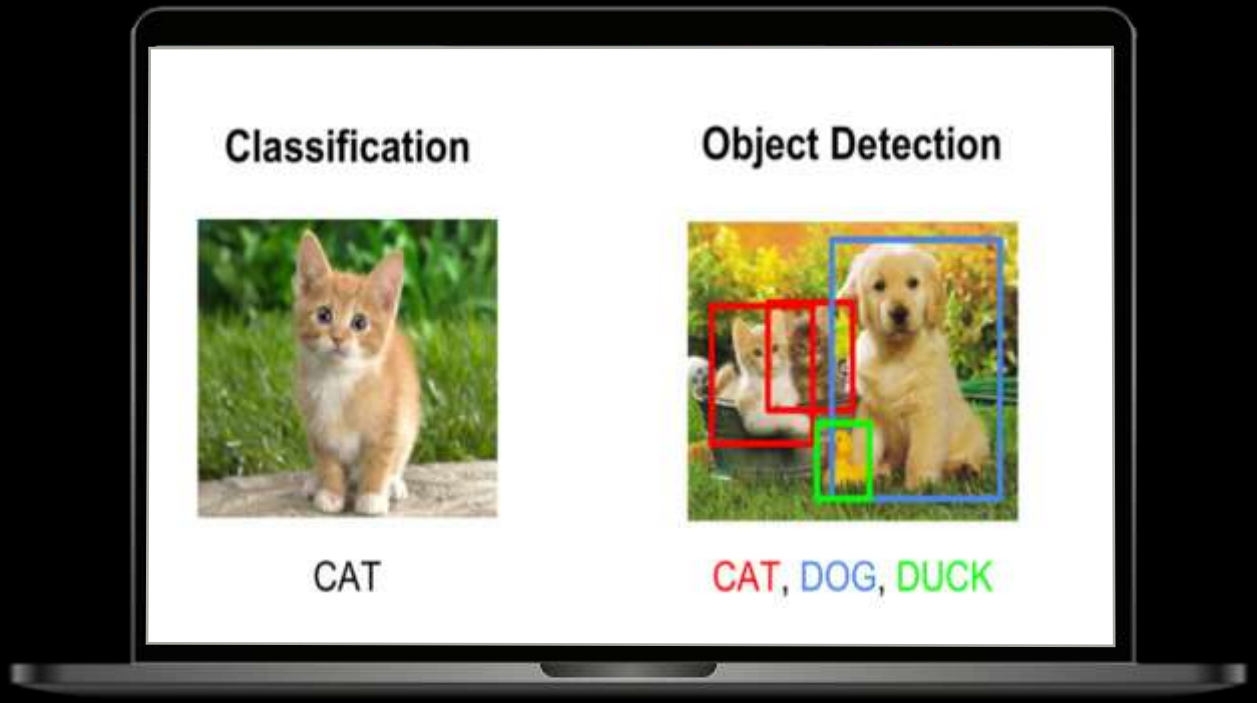


Image classification is best suited for classifying images into a certain category.

Object Detection is best suited for identifying the location of objects in an image.

AZURE AI CUSTOM VISION

Hosting a Model- Deploying to Cloud

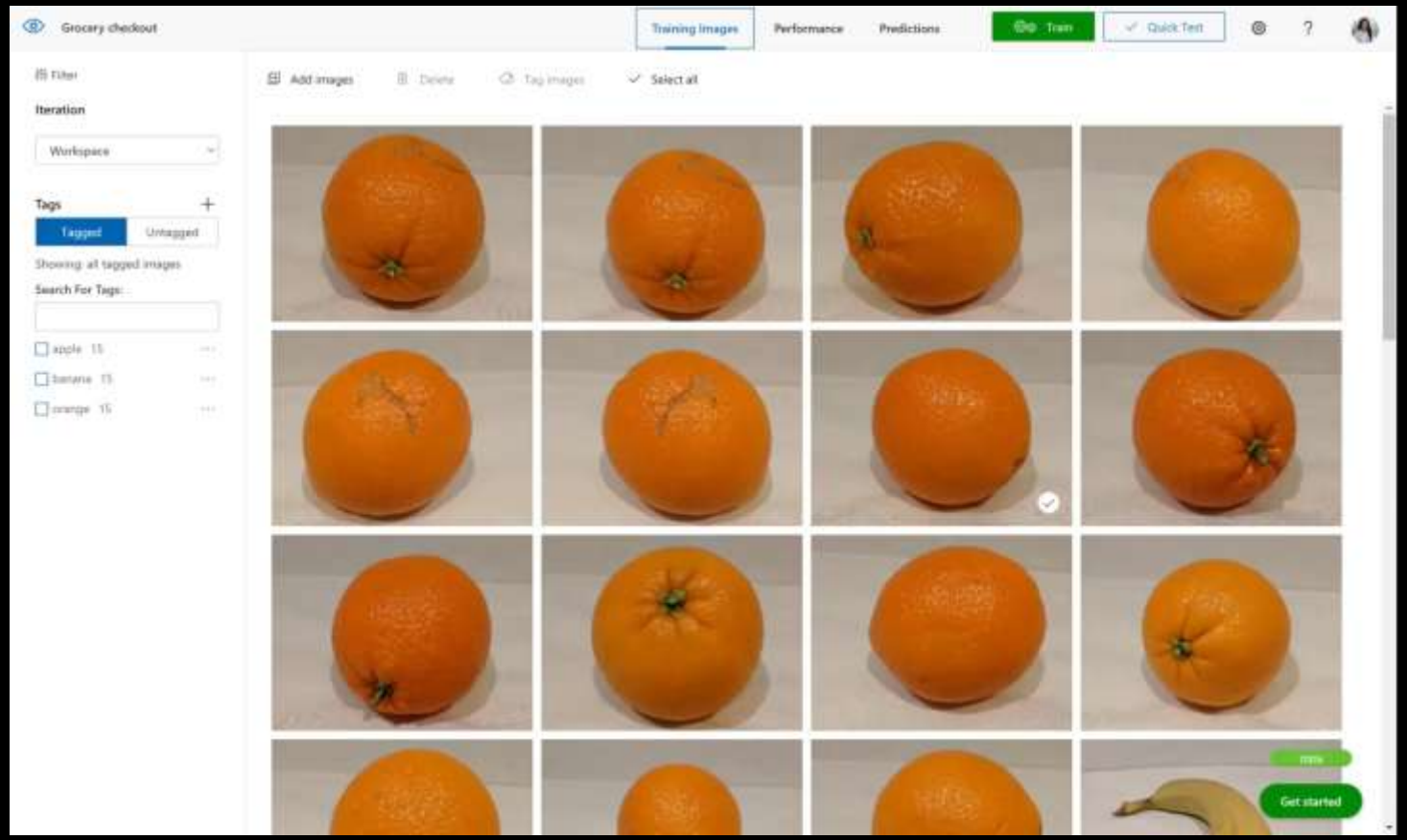
To use your model, client application developers need

Project ID

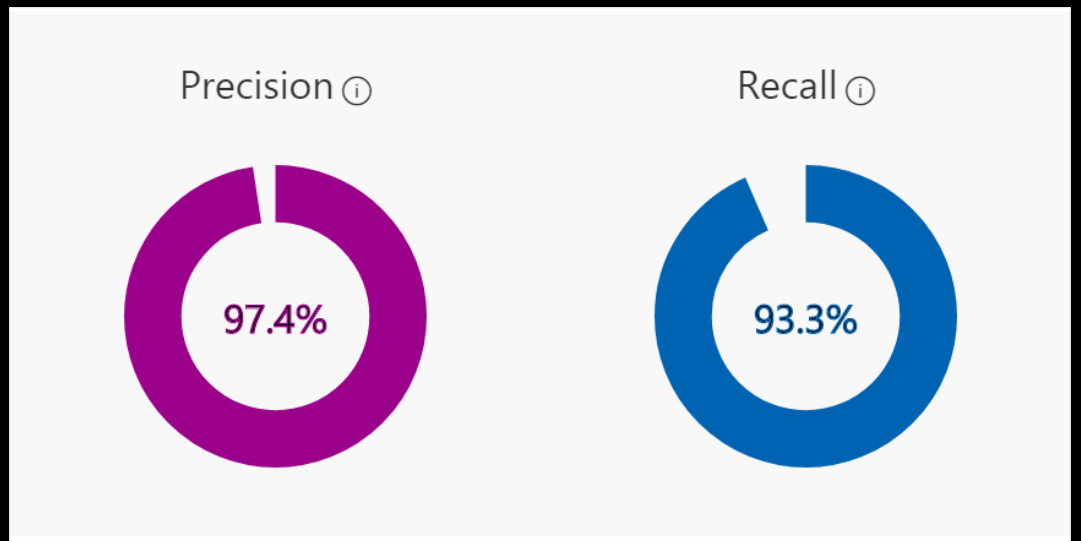
Model name

Prediction endpoint

Prediction key



Azure AI Custom Vision



Performance Per Tag

Tag	Precision	Recall
strawberry	99.2%	99.2%
Banana	99.1%	97.2%
Pineapple	98.9%	95.2%
Apple	98.4%	89.5%
Orange	98.3%	94.1%
Passionfruit	96.8%	85.1%
Coconut	91.1%	92.0%

Azure AI Custom Vision models for image classification:

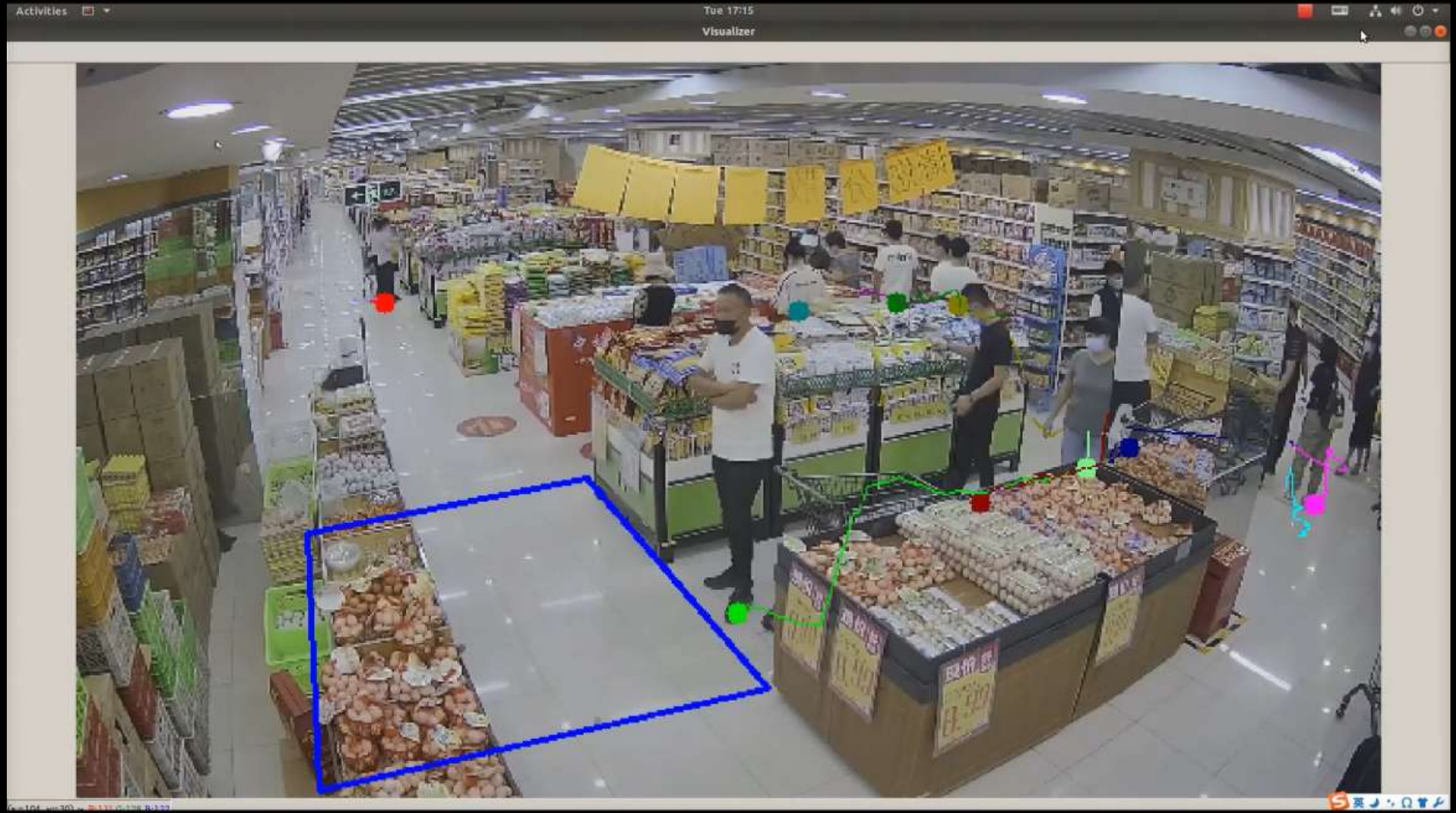
- New **General [A2]** domain added for better accuracy and shorter training time
- Common use case is training custom models for **defect detection** in various industries
- **Custom Image Analysis** - Docker Container running at the edge with full model download available in Private Preview

Azure AI Custom Vision models for object detection:

- New **General [A1]** domain for better accuracy

Availability of **Products on Shelves** domain to apply transfer learning with customer's data for **retail** scenarios

Private Preview - Model Customization for Person Attribute



Private Preview Model Customization for Personal Protection Equipment – Reflective Vest

New Vision Services

Video Summarization

Algorithm Output

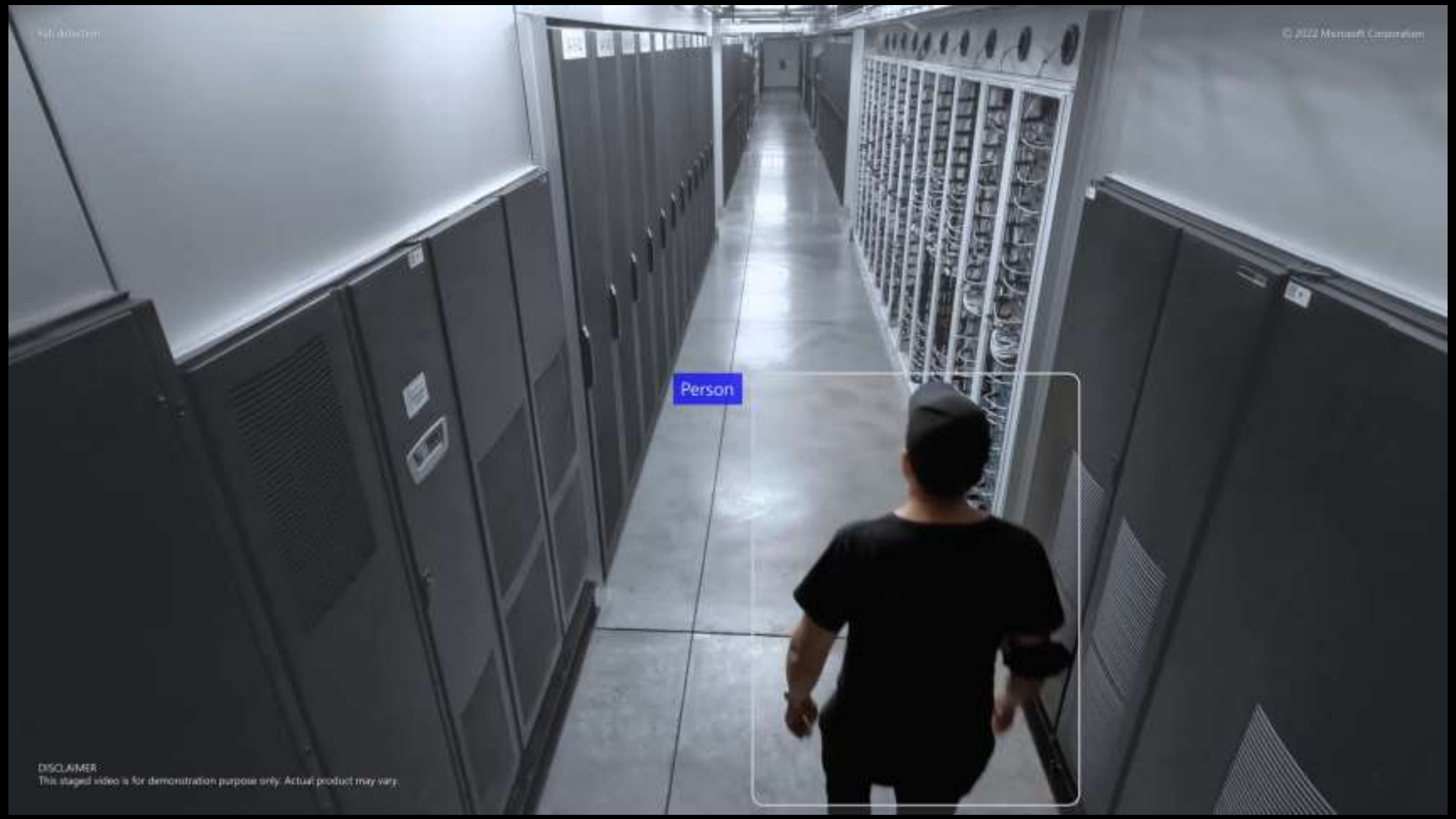
Flying Kite

Braiding Hair

Changing Oil

Action Recognition - Person Taking Photos

Action Recognition - Person Falling



Product Recognition

Shelf Analytics KPIs

- Planogram compliance

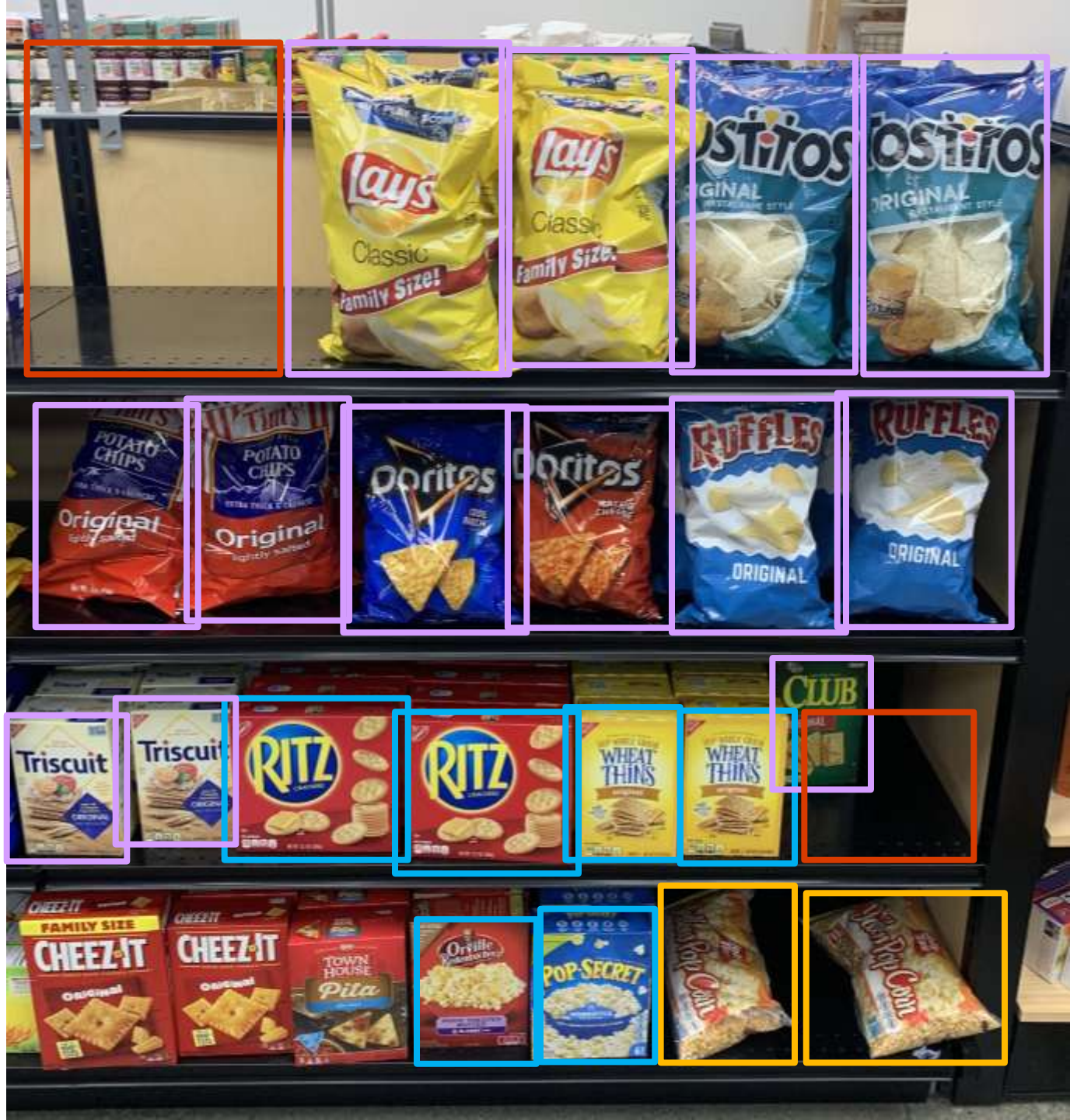


Image Stitching API

Up to 10 images to be stitched, RGB, 12 Mega Pixels resolution
Capture photos from left to right, top to bottom fashion
Recommended: 50% overlap
No panorama, no motion/blur/glare, no fisheye, no filters

Thank you!