Computer Vision in Azure

Microsoft Azure AI

CO-PRODUCED BY

Microsoft & DEVintersection

Azure Cognitive Services

- **“Cognitive Services” multi-service resource for:**
  - Vision(Computer Vision, Custom Vision, Face)
  - Speech (Speech-to-text, Text-to-speech, Speech Translation)
  - Language (LUIS, Language/Text Analytics, Translator)
  - Decision (Content Moderator)
- **Each is also available as a single service**
- **Consume cognitive services with**
  - Endpoint URI
  - Subscription Key
  - (sometimes API version)

Image Analysis

- **Pre-trained model**
- **Images are required to be:**
  - JPEG, PNG, GIF, or BMP
  - Less than 4MB in size
  - Greater than 50x50 pixels
- **Accessible via the “Analyze Image” API**
- **Customize results by including additional “visualFeatures”**

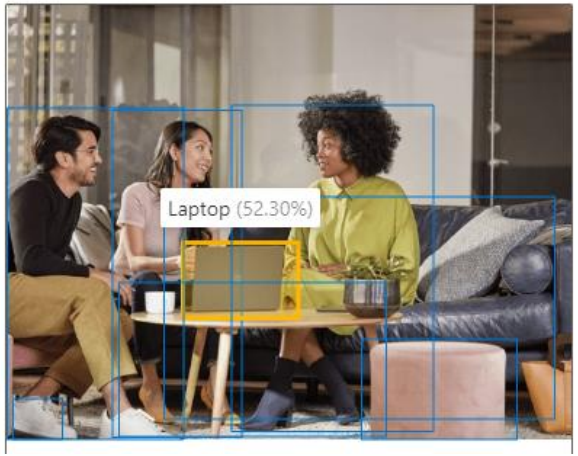
Image Analysis - Tagging

- **Model is pre-trained to identify over 10K known objects**
  - Living things, scenery, and actions
- **Tags include foreground and background elements**
- **Results include a confidence score**
- **In visualFeatures include: “Tags”**

sport (99.60%)
person (99.56%)
footwear (98.05%)
skating (96.27%)
boardsport (95.58%)
skateboarding equipment (94.43%)
clothing (94.02%)
wall (93.81%)
skateboarding (93.78%)
skateboarder (93.25%)
individual sports (92.80%)
street stunts (90.81%)
balance (90.81%)

Image Analysis – Detect objects

- Similar to tagging, only it includes a bounding box indicating the coordinates of detected objects
- Informs you if there are multiple instances of the same tag in the image
- In visualFeatures include: “Objects”



footwear (52.50%)
person (76.50%)
Laptop (52.30%)
seating (53.30%)
person (85.60%)
person (72.30%)
seating (67.80%)
table (61.30%)

Image Analysis – Detect brands

- Model is pre-trained with logos of thousands of commercial brands
- In visualFeatures include: “Brands”

JSON Copy

```
"brands":[
   {
      "name":"Microsoft",
      "rectangle":{
         "x":20,
         "y":97,
         "w":62,
         "h":52
      }
   }
]
```

Image Analysis – Categorize image

- Identify and categorize an entire image, using a category taxonomy with parent/child hereditary hierarchies.
- In visualFeatures include: “Categories”

outdoor_mountain

Image Analysis – Categorize image - Categories

Image Analysis – Describe an image

- Generates human-readable captions for images.
- In visualFeatures include: “Description”


```
{
    "description": {
        "tags": ["outdoor", "building", "photo", "city", "white", "black",
        "captions": [
            {
                "text": "a black and white photo of a city",
                "confidence": 0.95301952483304808
            },
            {
                "text": "a black and white photo of a large city",
                "confidence": 0.94085190563213816
            },
            {
                "text": "a large white building in a city",
                "confidence": 0.93108362931954824
            }
        ]
    },
    "requestId": "b20bfc83-fb25-4b8d-a3f8-b2a1f084b159",
    "metadata": {
        "height": 300,
        "width": 239,
        "format": "Jpeg"
    }
}
```

Image Analysis – Detect faces

- A lighter-weight subset of the Face service
- In visualFeatures include: “Faces”


```
{
    "faces": [
        {
            "age": 23,
            "gender": "Female",
            "faceRectangle": {
                "top": 45,
                "left": 194,
                "width": 44,
                "height": 44
            }
        }
    ],
    "requestId": "8439ba87-de65-441b-a0f1-c85913157ecd",
    "metadata": {
        "height": 200,
        "width": 300,
        "format": "Png"
    }
}
```

Image Analysis – Detect image types

- **Indicates if an image is clip art or a line drawing**
  - 0 == Non-clip-art
  - 1 == Ambiguous
  - 2 == Normal-clip-art
  - 3 == Good-clip-art
- **In visualFeatures include: “ImageType”**


```
{
    "imageType": {
        "clipArtType": 3,
        "lineDrawingType": 0
    },
    "requestId": "88c48d8c-80f3-449f-878f-6947f3b35a27",
    "metadata": {
        "height": 225,
        "width": 300,
        "format": "Jpeg"
    }
}
```

Image Analysis – Detect domain-specific content

- Identify celebrities or landmarks
- In “details” query parameter include: “Celebrities” or “Landmarks”


```
{
  "result": {
    "celebrities": [{
      "faceRectangle": {
        "top": 391,
        "left": 318,
        "width": 184,
        "height": 184
      },
      "name": "Satya Nadella",
      "confidence": 0.99999856948852539
    }]
  },
  "requestId": "8217262a-1a90-4498-a242-68376a4b956b",
  "metadata": {
    "width": 800,
    "height": 1200,
    "format": "Jpeg"
  }
}
```

Image Analysis – Detect color scheme

- Identifies the dominant foreground color, the dominant background color, and the larger set of dominant colors in the image, as well as an accent color returned in hexadecimal
- Possible colors: black, blue, brown, gray, green, orange, pink, purple, red, teal, white, and yellow
- In visualFeatures include: “Color”


```
{
    "color": {
        "dominantColorForeground": "Black",
        "dominantColorBackground": "Black",
        "dominantColors": ["Black", "White"],
        "accentColor": "BB6D10",
        "isBwImg": false
    },
    "requestId": "0dc394bf-db50-4871-bdcc-13707d9405ea",
    "metadata": {
        "height": 202,
        "width": 300,
        "format": "Jpeg"
    }
}
```

Image Analysis – Detect adult content

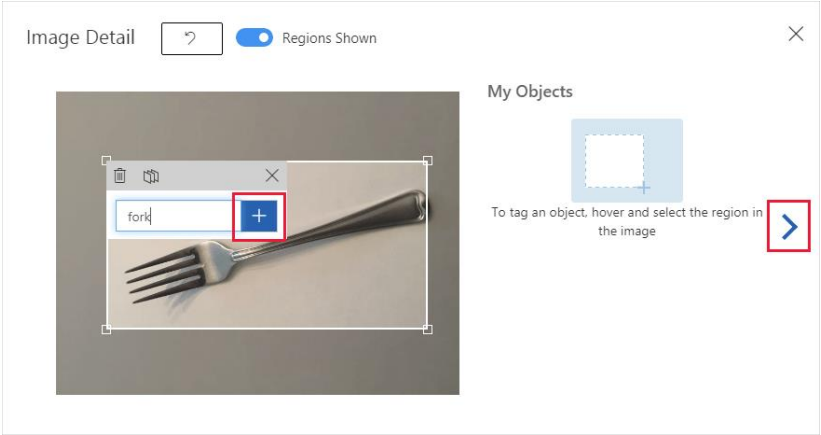
- **In visualFeatures include: “Adult”**
- **Returns boolean values for:**
  - isAdultContent
  - isRacyContent
  - isGoryContent
- **As well as their associated scores:**
  - adultScore
  - racyScore
  - goreScore

Get Area of Interest / Generate a Thumbnail

- **“Get Area of Interest” API**
  - Determine the main object of the image
- **“Generate Thumbnail” API**
  - Smart-crop based on the are of interest

Custom Vision Service

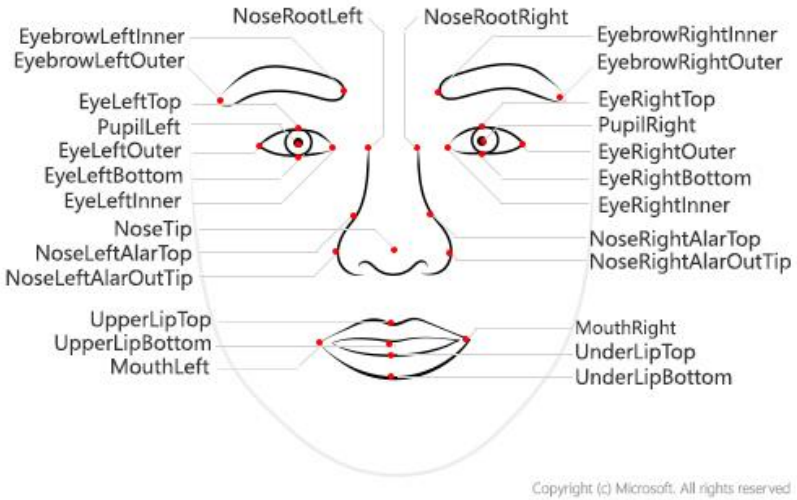
- **Train vision models by uploading and tagging your own images**
  - JPG, PNG, BMP, or GIF
  - Less than 6MB in size (4MB for prediction images)
  - No less than 256px on the shortest edge
  - Min. 15 images per tag
- **https://www.customvision.ai**
- **Image classification and object detection**



Consuming the Custom Vision Service

- **To consume the custom vision service, you will need:**
  - Project ID
  - Model Name
  - Prediction Endpoint
  - Prediction Key

Face Service



- **Face detection (Detect API)**
  - Bounding box
- **Facial landmark location (Detect API)**
  - Coordinates of pupils, tip of nose, etc.
- **Facial attribute analysis (Detect API)**
  - Used to infer age, emotional state,etc.
- **Facial comparison (Find Similar API)**
  - Identify similar faces
- **Facial recognition (Identify &Verify API)**
  - Identify returns closest matches (1-many)
  - Verify determines if two faces belong to the same person (1-1)

Consuming the Face Service

- **Consume the face service using**
  - key
  - Endpoint
- **Image requirements**
  - JPG, PNG, GIF, BMP
  - 4 MB or smaller
  - Face size range 36x36 to 4096x4096 px

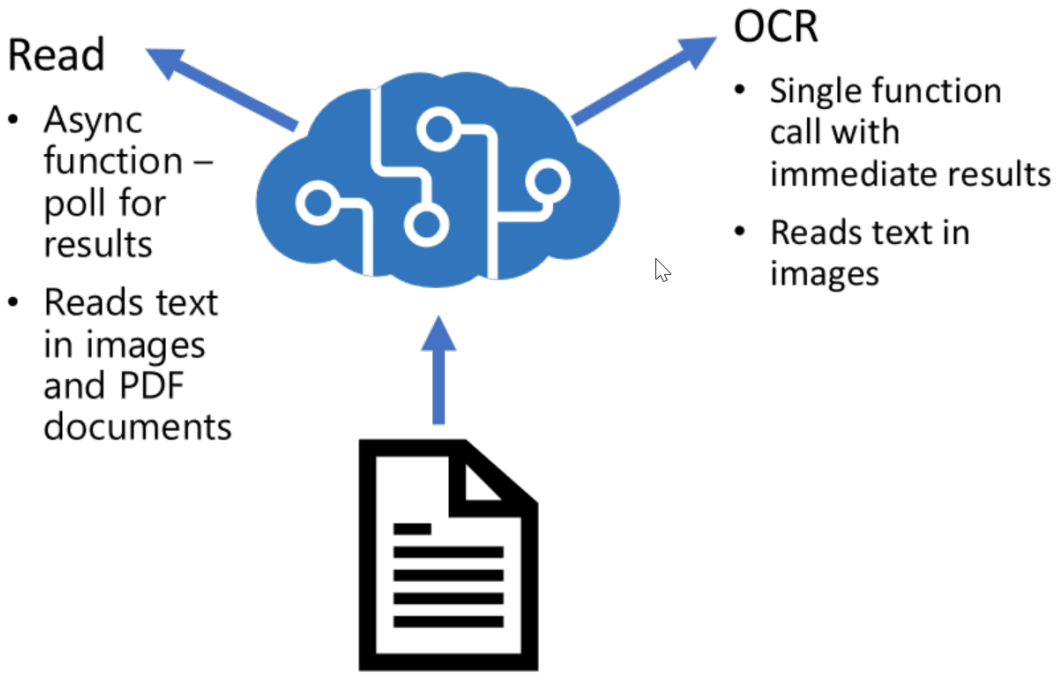
OCR

- **OCR API**
  - Meant for quick extraction of small amounts of text from images
  - Operates synchronously
  - Returns
    - Regions (location in the image)
    - Lines of text
    - Words in each line of text

(OCR) Read API

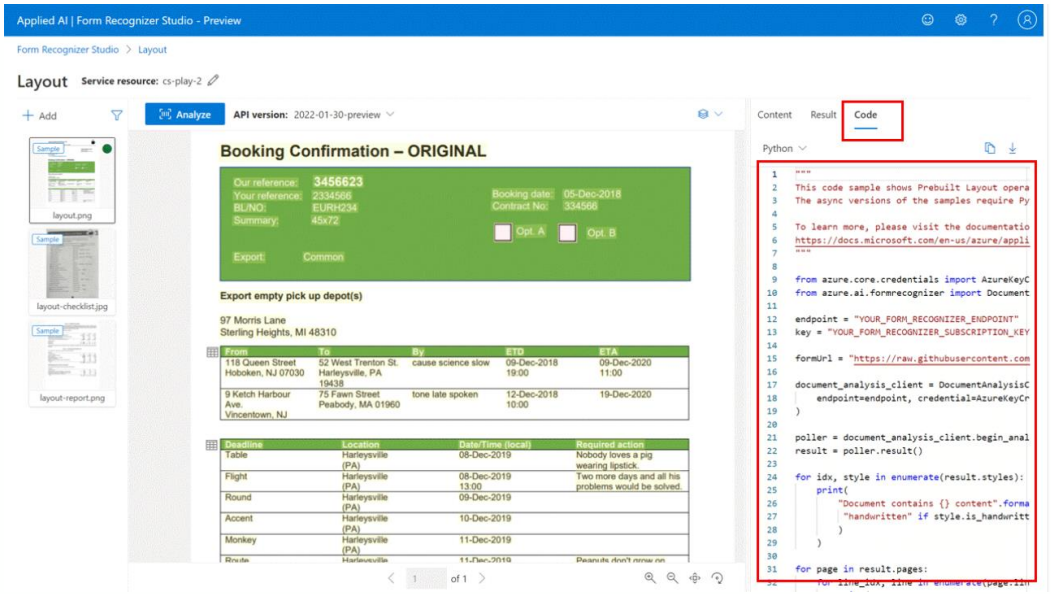
- Extracts text from images and PDF documents
- Greater accuracy than OCR API
- Can extract handwritten text (English only)
- Asynchronous

OCR vs Read API



Form Recognizer

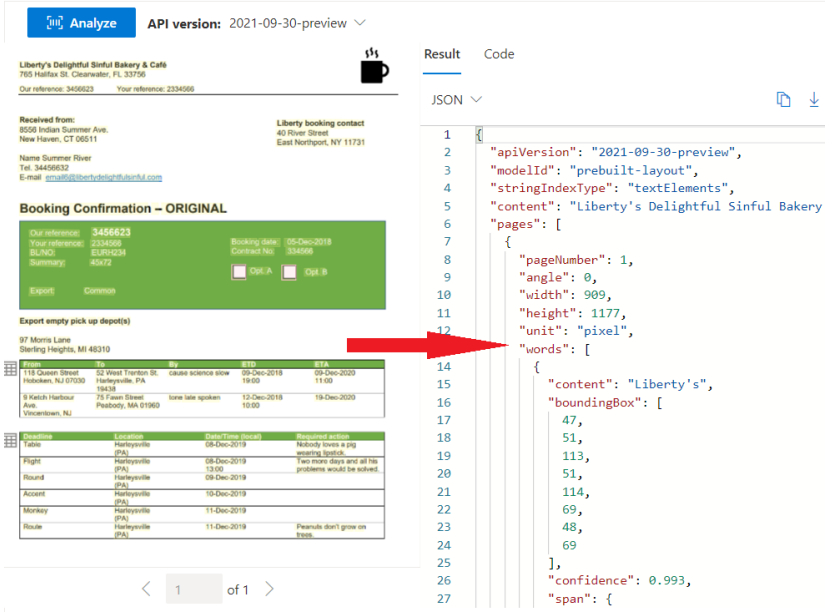
- **Automated processing of data**
  - Hand-filled/digital scan or image
- **Pre-built models available for**
  - Receipts
  - W-2*
  - ID Document
  - Business Card
  - Invoice
  - General document*
  - Read*



* Currently in preview

Form Recognizer – Layout API

- **Another pre-built model**
  - Introspects form, identifies tables, labels, and input automatically



Form Recognizer – Custom model

- **Interpret specific forms or documents**
- **Train with as few as 5 documents**
- **Requirements**
  - JPG, PNG, BMP, TIFF, PDF
  - Less than 2K pages (free max 2 pages)
  - PDF Max 17x17”
  - Total size of training data 500 pages or less
  - Unsupervised
    - Data must have keys and values
    - Keys need to be above or to the left of values
  - PDF can’t be password protected/locked

Custom form models

Custom form models work well when the target documents share a common visual layout. Training only takes a few minutes, and more than 100 languages are supported.

IN-PERSON EVENT

FOR INFORMATION ABOUT OUR NEXT IN PERSON EVENT, VISIT OUR WEBSITE AT:

APRIL 5-7, 2022

AzureAIConf.com

DECEMBER 7–9 2021

Microsoft Azure + AI Conference

CO-PRODUCED BY Microsoft & DEVintersection

Microsoft Azure + AI Conference

CO-PRODUCED BY Microsoft & DEVintersection

LAS VEGAS, NV

MGM GRAND