**CHAPTER 3**

**Describe features of computer vision workloads on Azure**

Cognitive Services is a suite of prebuilt AI services that developers can use to build AI solutions. Cognitive Services meet common AI requirements that allow you to add AI to your apps more quickly with less expertise.

This chapter explains the pre-built AI provided in Azure: Cognitive Services. The chapter will begin with an overview of all Cognitive Services but then will focus on one of the major components of Cognitive Services, the Computer Vision service.

Computer vision is the processing of still images and video streams. Computer vision can interpret the image and provide detail and understanding about the image in computer readable form.

The concepts involved in computer vision will be outlined with use cases, followed by how to use the Azure Cognitive Services Computer Vision service.

This chapter provides an overview of Cognitive Services and the details of the Computer Vision service. [Chapter 4](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch04.xhtml#ch04) will explain the other major component of Cognitive Services, Natural Language Processing.

**Skills covered in this chapter:**

* [Skill 3.1: Identify common types of computer vision solution](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03lev1sec1)
* [Skill 3.2: Identify Azure tools and services for computer vision tasks](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03lev1sec2)

**Skill 3.1: Identify common types of computer vision solution**

*Computer vision* is the processing of still images and video streams and extracting information from those images. Computer vision can interpret the image and provide detail and understanding about the image in a computer-readable form. Computers can take this information and perform further processing and analysis. Many applications use computer vision to enhance user experience or to capture information about objects and people.

Microsoft Azure provides a set of services around computer vision as part of Azure Cognitive Services. You can also use Azure Machine Learning to create your own image-processing models.

A focus of the Microsoft Azure AI Fundamentals certification is on the capabilities and features of computer vision and how computer vision can be applied in solutions. This requires you to understand the use cases for computer vision and to be able to differentiate the various services for computer vision in Microsoft Azure.

This skill covers how to:

* Introduce Cognitive Services
* Understand computer vision
* Describe image classification
* Describe object detection
* Describe optical character recognition
* Describe facial detection, recognition, and analysis

**Introduce Cognitive Services**

Before we look at computer vision, we need to describe Cognitive Services and how you configure Cognitive Services for use.

Cognitive Services are prebuilt machine learning models trained by Microsoft with massive volumes of data that developers can use to build AI solutions without requiring ML skills. Cognitive Services are focused on a subset of common AI requirements around processing images and analyzing text.

Cognitive Services are available as a set of REST APIs that can easily be deployed and consumed by applications. Essentially, Cognitive Services are off-the-shelf services that help you develop an AI-based solution more quickly and with less specialist expertise.

**Overview of Cognitive Services**

Cognitive Services are a family of AI services and APIs that you can use to build intelligent solutions. Cognitive Services enable applications to see, hear, speak, search, understand, and begin with decision-making.

This family of AI services is categorized into five groups:

* Decision
* Language
* Speech
* Vision
* Web search

The group of services in the Decision group helps you make smarter decisions:

* **Anomaly Detector**   Quickly identify potential problems by detecting unusual data points or trends in time-series data.
* **Metrics Advisor**   Built on Anomaly Detector, this service identifies the key areas for root cause analysis. Metrics Advisor helps focus on fixing issues rather than monitoring.
* **Content Moderator**   Detect potentially offensive or undesirable text, image, and video content. Content Moderator provides a review tool, where a human can validate flagged content and improve the sensitivity of moderation.
* **Personalizer**   Creates a personalized experience for a user based on his/her behavior. This could be content shown on a website or providing a different layout. Personalizer is an example of reinforcement learning.

The group of services in the Language group extract meaning from unstructured text:

* **Immersive Reader**   Helps readers of all ages and abilities to comprehend text using audio and visual cues. Immersive Reader can be used to improve literacy.
* **Language Understanding**   Builds natural language understanding into apps, bots, and IoT devices. Language Understanding interprets the intent and extracts key information from supplied text.
* **QnA Maker**   Creates a conversational question and answer layer on your existing FAQ and company information. QnA Maker is explained in [Chapter 5](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch05.xhtml#ch05).
* **Text Analytics**   Discovers insights from textual data. Text Analytics is one of the most used Cognitive Services. You can detect the sentiment of sentences or whole paragraphs. You can extract key phrases from a piece of text, and extract entities such as people, places, and things from a piece of text. Text Analytics supports a wide range of languages.
* **Translator**   Detects and translates text in real-time or in batch across more than 90 languages.

The Language services are the focus of [Chapter 4](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch04.xhtml#ch04).

The group of services in the Speech group allows you to add speech processing into your apps:

* **Speech to Text**   Transcribes audio into readable, searchable text in real-time or from audio files.
* **Text to Speech**   Synthesizes text into lifelike speech.
* **Speech Translation**   Converts audio into text and translates into another language in real-time. Speech Translation utilizes the Translator services.
* **Speaker Recognition**   Identifies people from the voices in an audio clip.

The Speech services are covered in [Chapter 4](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch04.xhtml#ch04).

The group of services in the Vision group helps you extract information from images and videos:

* **Computer Vision service**   Analyzes content in images and video and extracts details from the images.
* **Custom Vision**   Trains computer vision with your own set of images that meets your business requirements.
* **Face**   Detects faces in images and describes their features and emotions. Face can also recognize and verify people from images.
* **Form Recognizer**   Extracts text, key-value pairs, and tables from documents.
* **Video Analyzer for Media**   Analyzes the visual and audio channels of a video and indexes its content.

The rest of this chapter will focus on these Vision services.

The group of services in the Web Search group allows you to utilize the Bing search engine to search millions of webpages for images, news, product, and company information. These services have been moved from Cognitive Services to a separate service, Bing Web Search.

As you can see, Cognitive Services consist of a broad, and growing, set of AI services. A common feature of these services is that they require no training and can easily be consumed by applications with a REST API call.

We will now look at how you can deploy Cognitive Services in Azure.

**Deploy Cognitive Services**

Cognitive Services are easily deployed in Azure as resources. You can use the Azure portal, the CLI, or PowerShell to create resources for Cognitive Services. There are even ARM templates available to simplify deployment further.

Once created, the APIs in Cognitive Services will then be available to developers through REST APIs and client library SDKs.

You have two options when creating resources for Cognitive:

* Multi-service resource
* Single-service resource

With a multi-service resource, you have access to all the Cognitive Services with a single key and https endpoint. Benefits of a multi-service resource are the following:

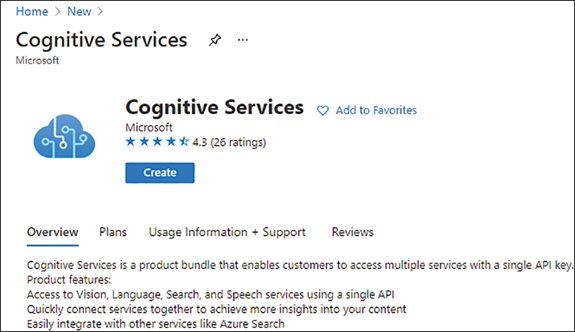
* Only one resource to create and manage
* Access to Vision, Language, Search, and Speech services using a single API
* Consolidates billing across all services

With a single-service resource, you access a single Cognitive Service with a unique key and https endpoint for each individual service. Benefits of a single-service resource are the following:

* Limits the services that a developer can use
* Dedicated API for each service
* Separate billing of services
* Free tier available

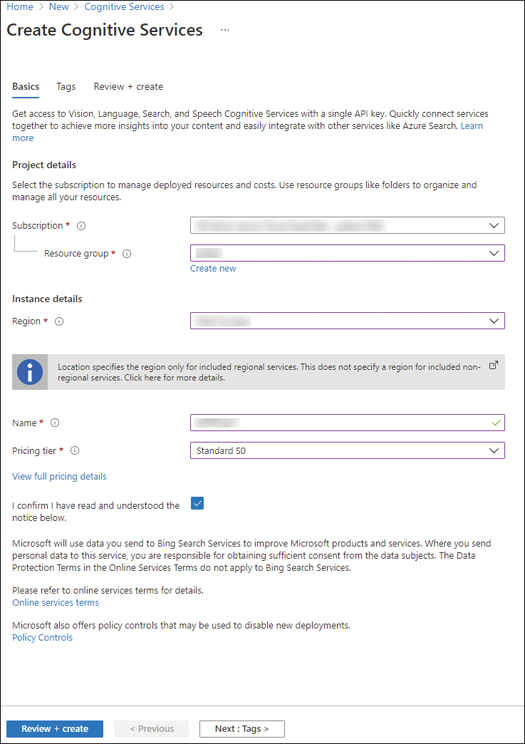
The multi-service resource is named Cognitive Services in the Azure portal. To create a multi-service Cognitive Services resource in the Azure portal, search for Cognitive Services and pick Cognitive Services by Microsoft.

[Figure 3-1](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig01) shows the service description for the Cognitive Services multi-resource service.



**FIGURE 3-1** Cognitive Services multi-resource service description

After clicking on the Create button, the Create Cognitive Services pane opens, as shown in [Figure 3-2](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig02).



**FIGURE 3-2** Creating a Cognitive Services resource

You will need to select the subscription, resource group, and region where the resource is to be deployed. You will then need to create a unique name for the service. This name will be the domain name for your endpoint and so must be unique worldwide. You should then select your pricing tier. There is only one pricing tier for the multi-service resource, Standard S0.

Clicking on Review + create will validate the options. You then click on Create to create the resource. The resource will be deployed in a few seconds.

You can create a Cognitive Services resource using the CLI as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-082pro01a)

az cognitiveservices account create --name <unique name> --resource-group <resource

group name> --kind CognitiveServices --sku S0 --location <region> --yes

To create a single-service resource for the Computer Vision service using the Azure portal, you should search for Computer Vision and create the resource. The options are the same as the multi-service resource, except you can select the Free pricing tier.

You can create a Computer Vision resource using the CLI as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-082pro02a)

az cognitiveservices account create --name <unique name> --resource-group <resource

group name> --kind ComputerVision --sku F0 --location <region> --yes

If you want to create other single-service resources, use the following CLI command to find the correct values for kind:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-082pro03a)

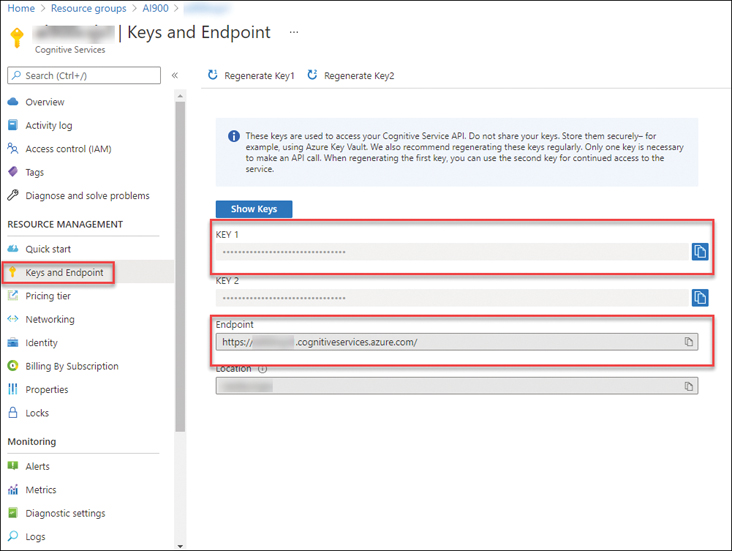
az cognitiveservices account list-kinds

Once your resource has been created, you will need to obtain the REST API URL and the key to access the resource.

**Use Cognitive Services securely**

Once created, each resource will have a unique endpoint for the REST API and authentication keys. You will need these details to use Cognitive Services from your app.

To view the endpoint and keys in the Azure portal, navigate to the resource and click on Keys and Endpoint, as shown in [Figure 3-3](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig03).



**FIGURE 3-3** Keys and Endpoint

You can access the keys using the CLI as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-083pro01a)

az cognitiveservices account keys list --name <unique name> --resource-group <resource

group name>

Image ***EXAM TIP***

Practice creating single- and multi-service resources in the Azure portal and make sure you know where the endpoint and keys can be found.

**Containers for Cognitive Services**

Cognitive Services are also available as Docker containers for deployment to IoT devices and to on-premises systems. Containers provide advantages with hybrid and disconnected scenarios and allow higher throughput with lower latency of data.

Only some Cognitive Services are available in containers.

***NEED MORE REVIEW?*   AZURE COGNITIVE SERVICES CONTAINERS**

For more information on using Cognitive Services with containers AI, see <https://docs.microsoft.com/azure/cognitive-services/cognitive-services-container-support>.

**Understand computer vision**

Computer vision is the interaction with the world through visual perception. Computer vision processes still images and video streams to interpret the images, providing details and understanding about the images.

A computer sees an array holding the color and intensity as number values. Computer vision analyzes these values using pre-built models to detect and interpret the image.

Computer vision makes it easy for developers to process and label visual content in their apps. The Computer Vision service API can describe objects in images, detect the existence of people, and generate human-readable descriptions and tags, enabling developers to categorize and process visual content.

**Key features of computer vision**

Some other key features of computer vision include the ability to:

* Categorize images
* Determine the image width and height
* Detect common objects including people
* Analyze faces
* Detect adult content

**Use cases for computer vision**

There are many uses for computer vision:

* In retail stores, a network of cameras can detect a shopper taking an object from the shelf and adding it to their basket.
* In vehicles, cameras can be used to detect pedestrians and cyclists, warning the driver of vulnerable road users.
* In healthcare, computer vision can analyze images of skin conditions to determine the severity with much higher accuracy than human specialists.
* In utilities, the positions of the panels on solar farms can be analyzed using cameras mounted on drones and the orientation changed to maximize efficiency.

**Describe image classification**

Image classification is a machine learning model that predicts the category, or class, the contents of the image belong to. A set of images is used to train the model. A new image can then be categorized using the model.

There are 86 standard categories that can be detected in an image. Categories are different to tags. Tags are based on the objects, people, and actions identified in the image.

Image classification can:

* Describe an image
* Categorize an image
* Tag an image

[Figure 3-4](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig04) shows an example of image classification with a textual description of the image added to the bottom of the image.



**FIGURE 3-4** Example of image classification

Detecting the color scheme in an image is an example of image classification. Colors are classified in the image: the dominant foreground color, the dominant background color, and the accent color, which is the most vibrant color in the image.

Identifying products on a warehouse shelf is an example of image classification. The model will check for products against trained images added to the model.

Quality control on a manufacturing line is another example of image classification. Product labels and bottle caps can be verified to be correctly attached using image classification against a set of trained images of correctly labeled and sealed products.

**Describe object detection**

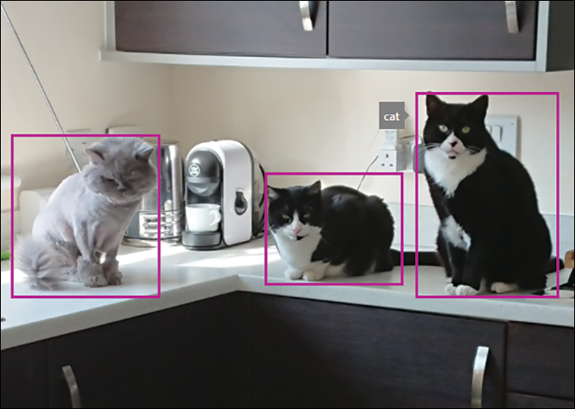
Object detection identifies and tags individual visual features (objects) in an image. Object detection can recognize many different types of objects.

Object detection will also return the coordinates for a box surrounding a tagged visual feature. Object detection is like image classification, but object detection also returns the location of each tagged object in an image.

Object detection can:

* Detect common objects
* Tag visual features
* Detect faces
* Identify brands and products
* Identify landmarks

[Figure 3-5](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig05) shows an example of object detection. Three cats have been identified as objects and their coordinates indicated by the boxes drawn on the image.



**FIGURE 3-5** Example of object detection

Object detection can be used to detect objects in an image. For example, you could train computer vision to detect people wearing face masks. Facial detection does not include the ability to recognize that a face is covered with a mask, and masks may prevent faces from being recognized.

Evaluating compliance with building safety regulations is another example of object detection. Images of a building interior and exterior can be used to identify fire extinguishers, doors, and other access and emergency exits.

**Describe optical character recognition**

Optical character recognition (OCR) extracts small amounts of text from an image. OCR can recognize individual shapes as letters, numerals, punctuation, and other elements of text.

OCR can:

* Extract printed text
* Extract handwritten text

Using OCR, you can extract details from invoices that have been sent electronically or scanned from paper. These details can then be validated against the expected details in your finance system.

[Figure 3-6](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig06) shows an example of using OCR to extract text from an image.



**FIGURE 3-6** Example of image classification

The OCR service extracted the following pieces of text from the image:

* 220-240V ~AC
* hp
* LaserJet Pro M102w
* Europe - Multilingual localization
* Serial No.
* VNF 4C29992
* Product No.
* G3Q35A
* Option B19
* Regulatory Model Number
* SHNGC-1500-01
* Made in Vietnam

**Describe facial detection, recognition, and analysis**

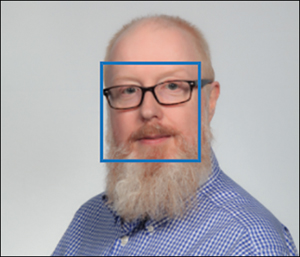
Facial detection can provide a series of attributes about a face it has detected, including whether the person is wearing eyeglasses or has a beard. Facial detection can also estimate the type of eye covering, including sunglasses and swimming goggles.

Facial detection and recognition can:

* Detect faces
* Analyze facial features
* Recognize faces
* Identify famous people

Object detection includes the detection of faces in an image but only provides basic attributes of the face, including age and gender. Facial detection goes much further in analyzing many other facial characteristics, such as emotion.

[Figure 3-7](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig07) shows an example of facial detection of the author.



**FIGURE 3-7** Example of facial detection

The facial detection identified the face, drew a box around the face, and supplied details such as wearing glasses, neutral emotion, not smiling, and other facial characteristics.

Customer engagement in retail is an example of using facial recognition to identify customers when they walk into a retail store.

Validating identity for access to business premises is an example of facial detection and recognition. Facial detection and recognition can identify a person in an image, and this can be used to permit access to a secure location.

Recognition of famous people is a feature of domain-specific content where thousands of well-known peoples’ images have been added to the computer vision model. Images can be tagged with the names of celebrities.

Face detection can be used to monitor a driver’s face. The angle, or head poise, can be determined, and this can be used to tell if the driver is looking at the road ahead, looking down at a mobile device, or showing signs of tiredness.

Now that you have learned about the concepts of computer vision, let’s look at the specific Computer Vision services provided by Azure Cognitive Services.

**Skill 3.2: Identify Azure tools and services for computer vision tasks**

Azure Cognitive Services provide pre-trained computer vision models that cover most of the capabilities required for analyzing images and videos.

This section describes the capabilities of the computer vision services in Azure Cognitive Services.

A focus of the Microsoft Azure AI Fundamentals certification is on the capabilities of the Computer Vision service. This requires you to understand how to use the Computer Vision service and especially how to create your own custom models with the Custom Vision service.

Image ***EXAM TIP***

You will need to be able to distinguish between the Computer Vision, Custom Vision, and Face services.

This skill covers how to:

* Understand the capabilities of the Computer Vision service
* Understand the Custom Vision service
* Understand the Face service
* Understand the Form Recognizer service

**Understand the capabilities of the Computer Vision service**

The Computer Vision service in Azure Cognitive Services provides a few different algorithms to analyze images. For instance, Computer Vision can do the following:

* Detect and locate over 10,000 classes of common objects.
* Detect and analyze human faces.
* Generate a single sentence description of an image.
* Generate a set of tags that relate to the contents of the image.
* Identify images that contain adult, racy, or gory content.
* Detect and extract the text from an image.

To use Computer Vision, you will need to create a Cognitive Services multi-service resource, or a Computer Vision single-service resource, as described earlier in this chapter.

The following sections describe the capabilities of the APIs in the Computer Vision service.

**Analyze image**

The analyze operation extracts visual features from the image content.

The image can either be uploaded or, more commonly, you specify a URL to where the image is stored.

You specify the features that you want to extract. If you do not specify any features, the image categories are returned.

The request URL is formulated as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-090pro01a)

https://{endpoint}/vision/v3.1/analyze[?visualFeatures][&details][&language]

The URL for the image is contained in the body of the request.

The visual features that you can request are the following:

* **Adult**   Detects if the image is pornographic (adult), contains sexually suggestive content (racy), or depicts violence or blood (gory).
* **Brands**   Detects well-known brands within an image.
* **Categories**   Categorizes image content according to a taxonomy of 86 categories.
* **Color**   Determines the accent color, dominant color, and whether an image is black and white.
* **Description**   Describes the image content with a complete sentence.
* **Faces**   Detects if there are human faces in the image with their coordinates, gender, and age.
* **ImageType**   Detects if the image is clipart or a line drawing.
* **Objects**   Detects various objects within an image, including their coordinates.
* **Tags**   Tags the image with a detailed list of words related to the content.

The details parameter is used to extract domain-specific details:

* **Celebrities**   Identifies celebrities in the image.
* **Landmarks**   Identifies landmarks in the image.

The language parameter supports a few languages. The default is en, English. Currently, English is the only supported language for tagging and categorizing images.

The Computer Vision service only supports file sizes less than 4MB. Images must be greater than 50x50 pixels and be in either of the JPEG, PNG, GIF, or BMP formats.

Below is the JSON returned for the image of the three cats used earlier in this chapter for these categories: adult, color, and imageType features.

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-091pro01a)

"categories": [{

"name": "animal\_cat",   "score": 0.79296875  }],

 "adult": {   "isAdultContent": false,   "isRacyContent": false,   "isGoryContent":

false,   "adultScore": 0.010710394941270351,   "racyScore": 0.01310222502797842,

"goreScore": 0.05890617147088051  },

 "color": {   "dominantColorForeground": "Black",   "dominantColorBackground": "Grey",

"dominantColors": ["Black", "Grey", "White"],   "accentColor": "635D4F",   "isBWImg":

false  },

 "imageType": {   "clipArtType": 0,   "lineDrawingType": 0  },

The category has been correctly identified with a confidence of 79.2%. There is no adult content in the image. The main colors are black, white, and grey.

The analyze operation provides a generic image analysis returning many different visual features. There are other operations that extract other information from the image or provide more detail than that provided by the analyze operation.

**Describe image**

The describe operation generates description(s) of an image using complete sentences. ­Content tags are generated from the various objects in the image.

One or more descriptions are generated. The sentences are evaluated, and confidence scores are generated. A list of captions is returned ordered from the highest confidence score to the lowest.

The request URL is formulated as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-091pro02a)

https://{endpoint}/vision/v3.1/describe[?maxCandidates][&language]

The parameter maxCandidates specifies the number of descriptions to return. The default is 1. The default language is English.

Following is the JSON returned for the image of the three cats used earlier in this chapter:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-091pro03a)

"description": {

"tags": ["cat", "sitting", "wall", "white", "indoor", "black", "sink", "counter",

"domestic cat"],

"captions": [{ "text": "a group of cats sitting on a counter top",  "confidence":

0.6282602548599243  }]  },

There are multiple tags related to the content in the image and a single sentence describing the image with a confidence of 62.8%.

**Detect objects**

The detect operation detects objects in an image and provides coordinates for each object detected. The objects are categorized using an 86-category taxonomy for common objects.

The request URL is formulated as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-092pro01a)

https://{endpoint}/vision/v3.1/detect

Following is the JSON returned for the image of the three cats used earlier in this chapter:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-092pro02a)

"objects": [{

"rectangle": {   "x": 556,   "y": 130,   "w": 190,   "h": 277   },   "object": "cat",

"confidence": 0.853,   "parent": {   "object": "mammal",   "confidence": 0.864,

"parent": {   "object": "animal",   "confidence": 0.865   }   }  }, {

 "rectangle": {   "x": 17,   "y": 183,   "w": 200,   "h": 216   },   "object": "cat",

"confidence": 0.831,   "parent": {   "object": "mammal",   "confidence": 0.839,

"parent": {   "object": "animal",   "confidence": 0.84   }   }  }, {

 "rectangle": {   "x": 356,   "y": 238,   "w": 182,   "h": 149   },   "object":

"cat",   "confidence": 0.81,   "parent": {   "object": "mammal",   "confidence": 0.816,

"parent": {   "object": "animal",   "confidence": 0.818   }   }  }]

The detect operation identified three cats with a high level of confidence and provided the coordinates for each cat.

**Content tags**

The tag operation generates a list of tags, based on the image and the objects in the image. Tags are based on objects, people, and animals in the image, along with the placing of the scene (setting) in the image.

The tags are provided as a simple list with confidence levels.

The request URL is formulated as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-092pro03a)

https://{endpoint}/vision/v3.1/tag[?language]

Following is the JSON returned for the image of the three cats used earlier in this chapter:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-092pro04a)

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-093pro01a)

"tags": [{

"name": "cat",   "confidence": 0.9999970197677612  }, {

"name": "sitting",   "confidence": 0.9983036518096924  }, {

"name": "wall",   "confidence": 0.9743844270706177  }, {

"name": "animal",   "confidence": 0.9706938862800598  }, {

"name": "white",   "confidence": 0.9519104957580566  }, {

"name": "indoor",   "confidence": 0.9119423627853394  }, {

"name": "black",   "confidence": 0.8455044031143188  }, {

"name": "kitty",   "confidence": 0.8295007944107056  }, {

"name": "small to medium-sized cats",   "confidence": 0.65200275182724  }, {

"name": "sink",   "confidence": 0.6215651035308838  }, {

"name": "feline",   "confidence": 0.5373185276985168  }, {

"name": "counter",   "confidence": 0.51436448097229  }, {

"name": "domestic cat",   "confidence": 0.2866966426372528  }],

The tag operation generated a list of tags in order of confidence. The cat tag has the highest confidence score of 99.9%, with domestic cat the lowest score of 28.7%.

**Domain-specific content**

There are two models in Computer Vision that have been trained on specific sets of images:

* **Celebrity**   Recognizes famous people.
* **Landmark**   Recognizes famous buildings or outdoor scenery.

The request URL is formulated as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-093pro02a)

https://{endpoint}/vision/v3.1/models/{model}/analyze[?language]

The model is either celebrities or landmarks. English is the default language.

[Figure 3-8](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig08) is a photograph of the Belém tower in Lisbon, Portugal. This is a famous ­sixteenth-century landmark, the place from where explorers set sail.



**FIGURE 3-8** Example of a landmark

The JSON returned includes the name of the celebrity or landmark, as shown next:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-094pro01a)

"landmarks": [{  "name": "Belém Tower",   "confidence": 0.9996672868728638   }]

These domain-specific models can also be used by the analyze operations by using the details parameter.

The analyze operation can also detect commercial brands from images using a database of thousands of company and product logos.

**Thumbnail generation**

The Get thumbnail operation generates a thumbnail image by analyzing the image, identifying the area of interest, and smart crops the image.

The generated thumbnail will differ depending on the parameters you specify for height, width, and smart cropping.

The request URL is formulated as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-094pro02a)

https://{endpoint}/vision/v3.1/generateThumbnail[?width][&height][&smartCropping]

Width and height are numeric values. SmartCropping is either 0 or 1.

The response contains a binary jpg image.

**Optical character recognition (OCR)**

OCR is the extraction of printed or handwritten text from images. You can extract text from images and documents.

There are two operations for extracting text from images:

* **Read**   The latest text recognition model that can be used with images and PDF documents. Read works asynchronously and must be used with the Get Read Results operation.
* **OCR**   An older text recognition model that supports only images and can only be used synchronously.

The request URL for the Read operation is formulated as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-094pro03a)

https://{endpoint}/vision/v3.1/read/analyze[?language]

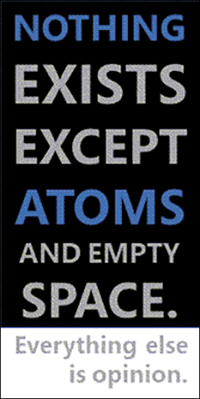
The request URL for the OCR operation is formulated as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-094pro04a)

https://{endpoint}/vision/v3.1/ocr[?language][&detectOrientation]

The default language is unknown, and the language will be detected from the text.

[Figure 3-9](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig09) is an image containing a quote from the Greek philosopher Democritus.



**FIGURE 3-9** Quote printed in an image

The JSON returned includes the pieces of text from the image, as shown next:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-095pro01a)

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-096pro01a)

{  "language": "en",  "textAngle": 0.0,  "orientation": "Up",

  "regions": [{

"boundingBox": "21,16,304,451",   "lines": [{   "boundingBox": "28,16,288,41",

"words": [{

"boundingBox": "28,16,288,41",   "text": "NOTHING"   }]   }, {

"boundingBox": "27,66,283,52",   "words": [{   "boundingBox": "27,66,283,52",   "text":

"EXISTS"   }]   }, {

"boundingBox": "27,128,292,49",   "words": [{   "boundingBox": "27,128,292,49",

"text": "EXCEPT"   }]   }, {

"boundingBox": "24,188,292,54",   "words": [{   "boundingBox": "24,188,292,54",

"text": "ATOMS"   }]   }, {

"boundingBox": "22,253,297,32",   "words": [{   "boundingBox": "22,253,105,32",

"text": "AND"   }, {

"boundingBox": "144,253,175,32",   "text": "EMPTY"   }]   }, {

"boundingBox": "21,298,304,60",   "words": [{   "boundingBox": "21,298,304,60",

"text": "SPACE."   }]   }, {

"boundingBox": "26,387,294,37",   "words": [{   "boundingBox": "26,387,210,37",

"text": "Everything"   }, {

"boundingBox": "249,389,71,27",   "text": "else"   }]   }, {

"boundingBox": "127,431,198,36",   "words": [{   "boundingBox": "127,431,31,29",

"text": "is"   }, {

"boundingBox": "172,431,153,36",   "text": "opinion."   }]   }]  }]

}

OCR only extracts the text it identifies. It does not provide any context to the text it extracts. The results are simply pieces of text.

**Content moderation**

The analyze operation can identify images that are risky or inappropriate. The Content Moderator service, although not part of Computer Vision (it is in the Decision group of APIs), is closely related to it.

Content Moderator is used in social media platforms to moderate messages and images. Content Moderator can be used in education to filter content not suitable for minors.

Content Moderator includes the ability to detect and moderate:

* **Images**   Scans images for adult or racy content, detects text in images with OCR, and detects faces.
* **Text**   Scans text for offensive or sexual content, profanity (in more than 100 languages), and personally identifiable information (PII).
* **Video**   Scans videos for adult or racy content.
* **Custom terms**   You can supply a set of terms that the Content Moderator can use to block or allow.
* **Custom images**   You can supply a set of custom images that the Content Moderator can use to block or allow.

Content Moderator includes a human review tool, a web portal where content that has been identified by the algorithms can be approved or rejected.

**Understand the Custom Vision service**

The Custom Vision service is an alternative to the pretrained Computer Vision service. Custom Vision enables you to build, train, and deploy a custom image recognition model based on images you provide.

In Custom Vision, you define the labels for your model and a set of sample images. You tag your images with your labels. The Custom Vision service uses a machine learning algorithm to analyze these sample images. Custom Vision trains and evaluates the custom model.

You can then deploy your model with an endpoint and key and consume this model in your apps in a similar way to the Computer Vision service.

Custom Vision supports two different types of mode:

* **Image classification**   Tags an image using the labels defined for the model.
* **Object detection**   Identifies objects using the tags and provides the coordinates of objects in an image. Object detection is a type of classification model.

A model can only be built for one of these two types.

Custom Vision uses a web portal (https://www.customvision.ai) where you can create your model, upload your images, label the images or the objects, train the model, test and evaluate the model, and finally deploy the model.

To use Custom Vision, you will need to create either a Cognitive Services multi-service resource, or a Custom Vision service resource, as described earlier in this chapter. There are two Custom Vision services: Training and Prediction. You will require both services.

**Creating a Custom Vision model**

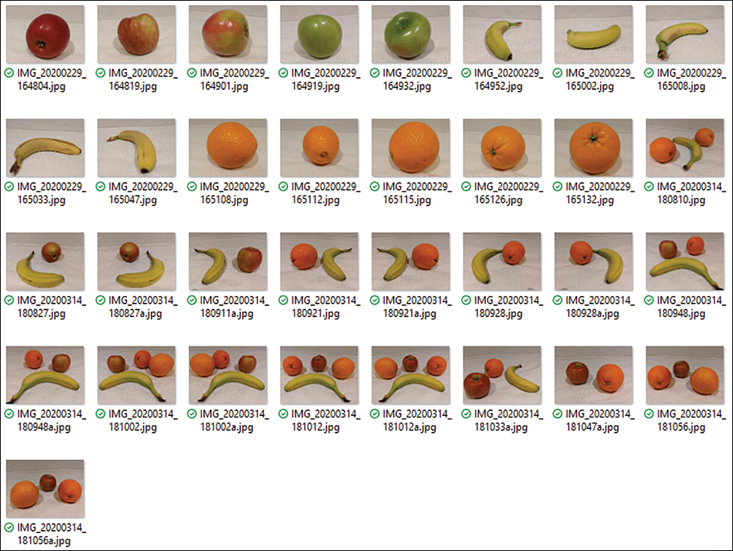
The process for creating a Custom Vision model is as follows:

1. Specify the model type.
2. Upload own images.
3. Define your labels.
4. Either
   1. Label images.
   2. Identify the object in the images.
5. Train the model.
6. Evaluate the model.
7. Deploy the model.

**Custom Vision exercise**

The following steps take you through creating a custom object detection model to identify fruit from images.

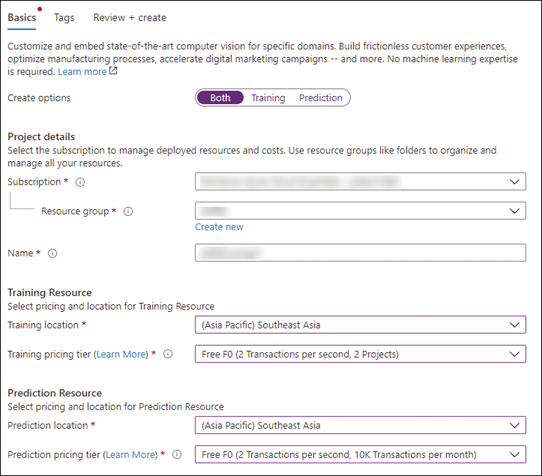
We will use the fruits dataset that you can download from <https://aka.ms/fruit-objects>. Extract the image files. There are 33 images, as shown in [Figure 3-10](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig10).



**FIGURE 3-10** Images of fruit

You will need to use 30 of the images to train your model, so keep three images for testing your model after you have trained it.

First, you need to create a Custom Vision service. [Figure 3-11](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig11) shows the pane in the Azure portal for creating a Custom Vision service.



**FIGURE 3-11** Creating a Cognitive Services resource

There is a toggle to choose which service(s) you require: Training and/or Prediction. You will need to select the subscription and resource group. You will then need to create a unique name for the service. This name will be the domain name for your endpoint and must be unique worldwide. For the Training resource, you should select the region where the Training resource is to be deployed and select your pricing tier: Free F0 or Standard S0. You then need to select the region and pricing tier for the Prediction resource.

Clicking on Review +Create will validate the options. You then click on Create to create the resource. If you selected Both, two resources will be deployed with the Training resource using the name you provided and the name of the Prediction resource with “-Prediction” appended.

You can create Custom Vision resources using the CLI as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-099pro01a)

az cognitiveservices account create --name <unique name for training> --resource-group

<resource group name> --kind CustomVision.Training --sku F0 --location <region>

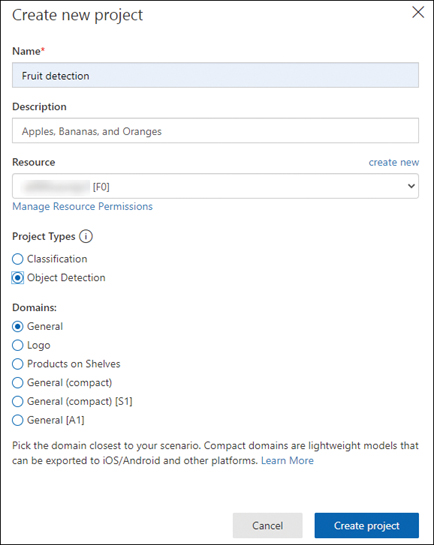
az cognitiveservices account create --name <unique name for prediction> --resource-group

<resource group name> --kind CustomVision.Prediction --sku F0 --location <region>

Next, you need to navigate to the Custom Vision web portal, [https://www.customvision.ai](https://www.customvision.ai/), and sign in with the credentials for your Azure subscription.

You will need to create a new project. You will need to name your project and select your Custom Vision training resource (or you can use a multi-service Cognitive Service resource).

Next, you should select Object Detection as the Project Type and General for the Domain, as shown in [Figure 3-12](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig12).



**FIGURE 3-12** New Custom Vision project

The domain is used to train the model. You should select the most relevant domain that matches your scenario. You should use the General domain if none of the domains are applicable.

Domains for image classification are as follows:

* General
* Food
* Landmarks
* Retail
* General (compact)
* Food (compact)
* Landmarks (compact)
* Retail (compact)
* General [A1]
* General (compact) [S1]

Domains for object detection are as follows:

* General
* Logo
* Products on Shelves
* General (compact)
* General (compact) [S1]
* General [A1]

Compact domains are lightweight models that are designed to run locally—for example, on mobile platforms.

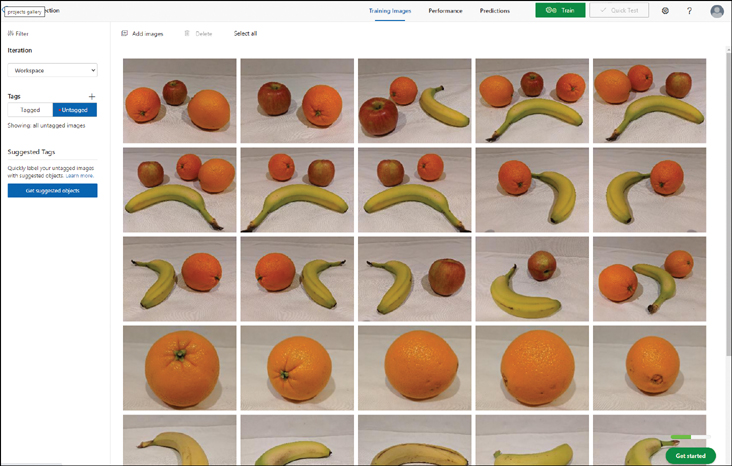
***NEED MORE REVIEW?*   DOMAINS**

For more explanation as to which domain to choose, see <https://docs.microsoft.com/azure/cognitive-services/custom-vision-service/select-domain>.

Once the project is created, you should create your tags. In this exercise, you will create three tags:

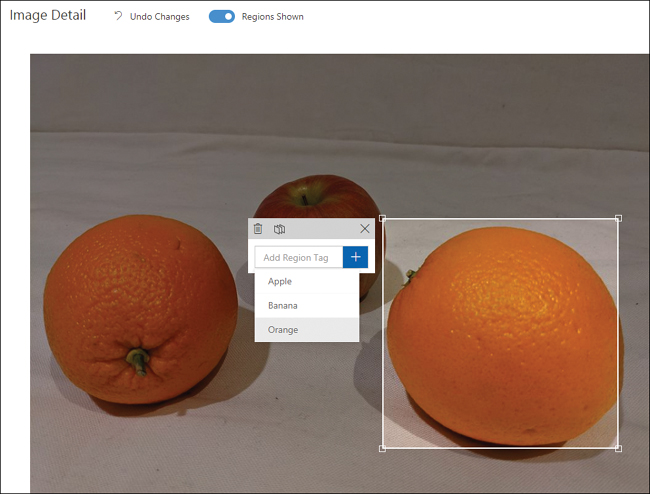
* Apple
* Banana
* Orange

Next, you should upload your training images. [Figure 3-13](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig13) shows the Custom Vision project with the images uploaded and untagged.



**FIGURE 3-13** Custom Vision project with uploaded images

You now need to click on each image. Custom Vision will attempt to identify objects and highlight the object with a box. You can adjust and resize the box and then tag the objects in the image, as shown in [Figure 3-14](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig14).



**FIGURE 3-14** Tagging objects

You will repeat tagging the objects for all the training images.

You will need at least 10 images for each tag, but for better performance, you should have a minimum of 30 images. To train your model, you should have a variety of images with different lighting, orientation, sizes, and backgrounds.

Select the Tagged button in the left-hand pane to see your tagged images.

You are now ready to train your model. Click on the Train button at the top of the project window. There are two choices:

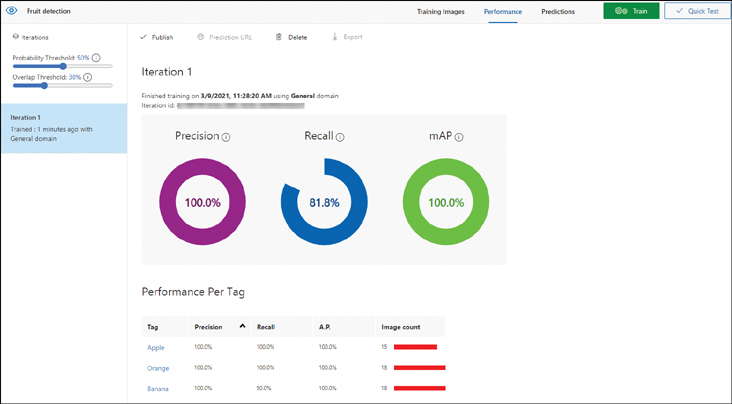
* **Quick Training**   Training will take a few minutes.
* **Advanced Training**   Specifies the amount of time to spend training the model from 1 to 24 hours.

Select the Quick Training option and click on Train.

When training has completed, the model’s performance is displayed. There are two key measures that indicate the effectiveness of the model:

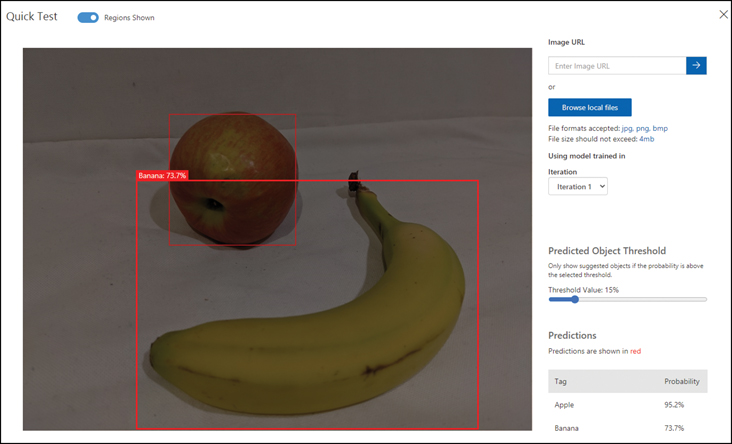
* **Precision**   The percentage of predictions that the model correctly detected. This is a value between 0 and 1 and is shown as a percentage (the higher the better).
* **Recall**   The percentage of the predictions that the model was correct. This is a value between 0 and 1 and is shown as a percentage (the higher the better).

[Figure 3-15](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig15) shows the results after training the model.



**FIGURE 3-15** Model performance

You can use the Quick Test option to check your model. You should upload one of the three images you put aside. The image will be automatically processed, as shown in [Figure 3-16](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig16).



**FIGURE 3-16** Quick Test

The model has identified both the apple and the banana and drawn boxes around the pieces of fruit. The objects are tagged, and the results have high confidence scores of 95.2% and 73.7%.

To publish your model, click on the Publish button at the top of the Performance tab shown in [Figure 3-16](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig16). You will need to name your model and select a Custom Vision Prediction resource.

***NOTE*   PUBLISHED ENDPOINT**

You cannot use a multi-service Cognitive Services resource for the published endpoint.

Publishing will generate an endpoint URL and key so that that your applications can use your custom model.

**Computer Vision vs. Custom Vision**

It is important that you understand the differences of capabilities of the prebuilt Computer Vision service compared with the capabilities of Custom Vision.

Computer Vision uses prebuilt models trained with many thousands of images. The Computer Vision service has the following capabilities:

* Object detection
* Image classification
* Content moderation
* Optical character recognition (OCR)
* Facial recognition
* Landmark recognition

Custom Vision uses images and tags that you supply to train a custom image recognition model. Custom Vision only has two of the capabilities:

* Object detection
* Image classification

**Understand the Face service**

While the Computer Vision service includes face detection, it provides only basic information about the person. The Face service performs more detailed analysis of the faces in an image. The Face service can examine facial characteristics, compare faces, and even verify a person’s identity. If you want to do analysis around the characteristics of faces or compare faces, you should use the Face service instead of Computer Vision.

Facial recognition has many use cases, such as security, retail, aiding visually challenged people, disease diagnosis, school attendance, and safety.

The Face service contains several advanced face algorithms, enabling face attribute detection and recognition. The Face service examines facial landmarks including noses, eyes, and lips to detect and recognize faces.

The Face service can detect attributes of the face, such as the following:

* Gender
* Age
* Emotions

The Face service can perform facial recognition:

* Similarity matching
* Identity verification

The Face service can be deployed in the Azure portal by searching for Face when creating a new resource. You must select your region, resource group, provide a unique name, and select the pricing tier: Free F0 or Standard S0.

You can create Face resources using the CLI as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-104pro01a)

az cognitiveservices account create --name <unique name> --resource-group <resource

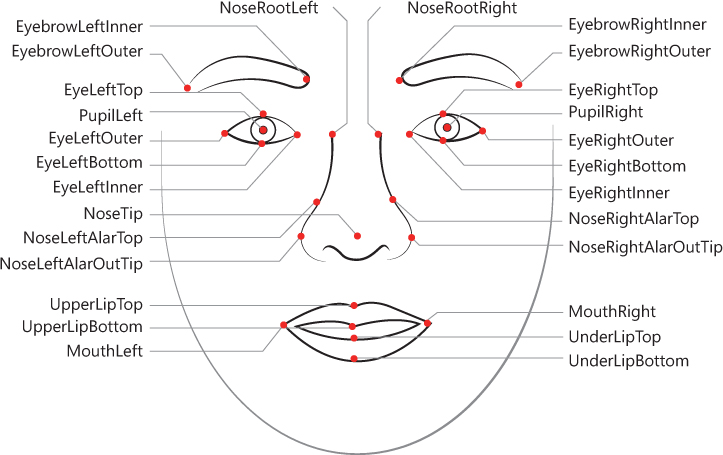
group name> --kind Face --sku F0 --location <region>

The Face service has several facial image-processing operations.

**Detection**

The Face service detects the human faces in an image and returns their boxed coordinates. Face detection extracts face-related attributes, such as head pose, emotion, hair, and glasses.

The Face service examines 27 facial landmarks, as shown in [Figure 3-17](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig17). The location of eyebrows, eyes, pupils, nose, mouth, and lips are the facial landmarks used by the Face service.



**FIGURE 3-17** Facial landmarks

Facial detection provides a set of features, or attributes, about the faces it has detected:

* **Age**   The estimated age in years.
* **Gender**   The estimated gender (male, female, and genderless).
* **Emotion**   A list of emotions (happiness, sadness, neutral, anger, contempt, disgust, surprise, and fear) each with a confidence score. The scores across all emotions add up to 1.
* **Glasses**   Whether the given face has eyeglasses and the type of eye covering (NoGlasses, ReadingGlasses, Sunglasses, or Swimming Goggles).
* **Hair**   Whether the face has hair, and the hair color, or is bald.
* **Facial hair**   Whether the face has facial hair.
* **Makeup**   Whether the eyes and/or lips have makeup as either true or false.
* **Smile**   Whether the face is smiling. A value of 0 means no smile and a value of 1 is a clear smile.
* **Occlusion**   Whether there are objects blocking parts of the face. True or false is returned for eyeOccluded, foreheadOccluded, and mouthOccluded.
* **Blur**   How blurred the face is in the image. This has a value between 0 and 1 with an informal rating of low, medium, or high.
* **Exposure**   The level exposure of the face between 0 and 1 with an informal rating of underExposure, goodExposure, or overExposure.
* **Noise**   The level of visual noise detected in the face image. This has a value between 0 and 1 with an informal rating of low, medium, or high.
* **Head pose**   The orientation of the face. This attribute is described by the pitch, roll, and yaw angles in degrees.

The request URL is formulated as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-106pro01a)

https://{endpoint}/face/v1.0/detect[?returnFaceId][&returnFaceLandmarks]

[&returnFaceAttributes][&recognitionModel][&detectionModel]

The parameters you can specify include the following:

* **returnFaceId**   True or false to indicate if the API should return IDs of detected faces.
* **returnFaceLandmarks**   True or false to indicate if the API should return facial landmarks.
* **returnFaceAttributes**   A comma-separated list of the attributes you want returned (age, gender, headPose, smile, facialHair, glasses, emotion, hair, makeup, occlusion, accessories, blur, exposure, and noise).
* **detectionModel**   There are three detection models you can use: detection\_01, detection\_02, and detection\_03. The default is detection\_01. The detection\_02 model should be used for images with small, side, and blurry faces. The detection\_03 model has better results on small faces. Facial attributes are not available for detection\_02 and detection\_03.
* **recognitionModel**   You should use the recognitionModel if you want to use the Recognition operations described in the next section. There are three recognition models you can use: recognition\_01, recognition\_02, and recognition\_03. The default model is recognition\_01. The latest model, recognition\_03, is recommended since its accuracy is higher than the older models.

The detection model returns a FaceId for each face it detects. This Id can then be used by the face recognition operations described in the next section.

The JSON returned using the detect operation on the image of the author in [Figure 3-7](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig07) is shown next:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-106pro02a)

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-107pro01a)

{   "faceId": "aa2c934e-c0f9-42cd-8024-33ee14ae05af",

 "faceRectangle": {   "top": 613,   "left": 458,   "width": 442,   "height": 442   },

 "faceAttributes": {   "hair": {   "bald": 0.79,   "invisible": false,   "hairColor":

[   {   "color": "gray",   "confidence": 0.98   },   {   "color": "brown",

"confidence": 0.7   },   {   "color": "blond",   "confidence": 0.47   },   {   "color":

"black",   "confidence": 0.45   },   {   "color": "other",   "confidence": 0.28   },

{   "color": "red",   "confidence": 0.04   },   {   "color": "white",   "confidence":

0.0   }   ]   },

 "smile": 0.011,

"headPose": {   "pitch": 2.9,   "roll": -2.2,   "yaw": -9.3   },

 "gender": "male",

 "age": 53.0,

 "facialHair": {   "moustache": 0.9,   "beard": 0.9,   "sideburns": 0.9   },

 "glasses": "ReadingGlasses",

 "makeup": {   "eyeMakeup": false,   "lipMakeup": false   },

 "emotion": {   "anger": 0.0,   "contempt": 0.0,   "disgust": 0.0,   "fear": 0.0,

"happiness": 0.011,   "neutral": 0.989,   "sadness": 0.0,   "surprise": 0.0   } }

As you can see, the attributes are mainly correct except for the hair color. This is expected as the image in [Figure 3-7](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig07) was a professionally taken photograph with good exposure and a neutral expression.

**Recognition**

The Face service can recognize known faces. Recognition can compare two different faces to determine if they are similar (Similarity matching) or belong to the same person (Identity verification).

There are four operations available in facial recognition:

* **Verify**   Evaluates whether two faces belong to the same person. The Verify operation takes two detected faces and determines whether the faces belong to the same person. This operation is used in security scenarios.
* **Identify**   Matches faces to known people in a database. The Identify operation takes one or more face(s) and returns a list of possible matches with a confidence score between 0 and 1. This operation is used for automatic image tagging in photo ­management software.
* **Find Similar**   Extracts faces that look like a person’s face. The Find Similar operation takes a detected face and returns a subset of faces that look similar from a list of faces you supply. This operation is used when searching for a face in a set of images.
* **Group**   Divides a set of faces based on similarities. The Group operation separates a list of faces into smaller groups on the similarities of the faces.

You should not use the Identify or Group operations to evaluate whether two faces belong to the same person. You should use the Verify operation instead.

Image ***EXAM TIP***

Ensure that you can determine the scenario for each of the four facial recognition operations.

**Computer Vision vs. Face service**

There are three services that perform an element of facial detection:

* Computer Vision
* Face
* Video Analyzer for Media

It is important that you understand the differences of capabilities of the prebuilt Computer Vision service compared with the capabilities of the Face service and the Video Analyzer for Media service.

Video Analyzer for Media, formerly Video Indexer, is part of Azure Media Services and utilizes Cognitive Services, including the Face service, to extract insights from videos. Video Analyzer for Media can detect and identify people and brands.

Image ***EXAM TIP***

You will need to be able to distinguish between Computer Vision, Face, and Video Analyzer for Media.

Computer Vision can detect faces in images but can only provide basic information about the person from the image of the face, such as the estimated age and gender.

The Face service can detect faces in images and can also provide information about the characteristics of the face. The Face service can also perform the following:

* Facial analysis
* Face identification
* Pose detection

The Video Analyzer for Media service can detect faces in video images but can also perform face identification.

Here are some examples of the differences between these services:

* The Face API can detect the angle a head is posed at. Computer Vision can detect faces but is not able to supply the angle of the head.
* Video Analyzer for Media can detect faces but does not return the attributes the Face API can return.
* The Face API service is concerned with the details of faces. The Video Analyzer for Media service can detect and identify people and brands but not landmarks.
* Custom Vision allows you to specify the labels for an image. The other services cannot.
* Computer Vision can identify landmarks in an image. The other services cannot.

**Understand the Form Recognizer service**

Optical character recognition (OCR) is an operation available in Computer Vision. As you will have seen, OCR simply extracts any pieces of text it can find in an image without any context about that text.

The Form Recognizer service extracts text from an image or a document using the context of the document.

***NOTE*   FORM RECOGNIZER**

Form Recognizer can extract text, key-value pairs, and tabular data as structured data that can be understood by your application.

Form Recognizer can extract information from scanned forms in images or PDF formats. You can either train a custom model using your own forms or use one of the pre-trained models.

There are three pre-trained models:

* Business cards
* Invoices
* Receipts

The Form Recognizer service can be deployed in the Azure portal by searching for Form Recognizer when creating a new resource. You must select your region, resource group, provide a unique name, and select the pricing tier: Free F0 or Standard S0. The free tier in the Form Recognizer service will only process the first two pages of a PDF document.

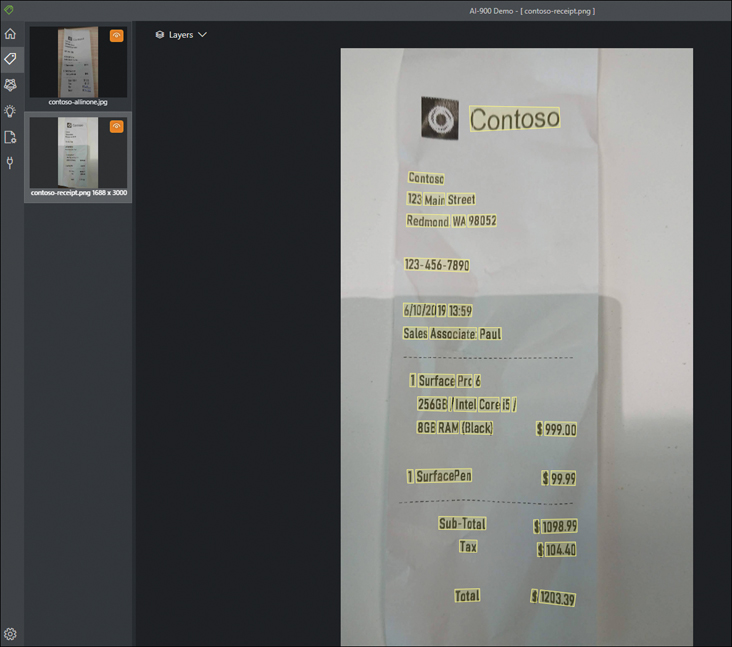
You can create Form Recognizer resources using the CLI as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-109pro01a)

az cognitiveservices account create --name <unique name> --resource-group <resource

group name> --kind FormRecognizer --sku F0 --location <region>

You can try out the Form Recognizer at [https://fott.azurewebsites.net](https://fott.azurewebsites.net/). [Figure 3-18](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03.xhtml#ch03fig18) shows a receipt in the Sample labeling tool.



**FIGURE 3-18** Form Recognizer tool

The text in the receipt is highlighted in yellow. The information generated from the receipt is as follows:

[Click here to view code image](https://learning.oreilly.com/library/view/exam-ref-ai-900/9780137358076/ch03_images.xhtml#a-110pro01a)

Receipt Type: Itemized

Merchant: Contoso

Address: 123 Main Street Redmond, WA 98052

Phone number: +19876543210

Date: 2019-06-10

Time: 13:59:00

Subtotal: 1098.99

Tax: 104.4

Total: 1203.39

Line items:

Item Quantity: 1

Item Name: Surface Pro 6

Total Price: 999.00

Item Quantity: 1

Item Name: Surface Pen

Total Price: 99.99

**Form Recognizer vs. OCR**

There are three services that perform an element of text extraction from images:

* OCR
* Read
* Form Recognizer

You should understand the differences between these services.

The older OCR operation can only process image files. OCR can only extract simple text strings. OCR can interpret both printed and handwritten text.

The Read operation can process images as well as multi-page PDF documents. Read can interpret both printed and handwritten text.

The Form Recognizer service can extract structured text from images and multi-page PDF documents. Form Recognizer will recognize form fields, and is not just text extraction.

**Chapter summary**

In this chapter, you learned some of the general concepts related to computer vision. You learned about the types of computer vision, and you learned about the services in Azure Cognitive Services related to computer vision. Here are the key concepts from this chapter:

* Cognitive Services are prebuilt Machine Learning models available through REST APIs.
* Cognitive Services enable applications to see, hear, speak, search, understand, and build intelligence into your applications quickly and easily.
* Cognitive Services can be deployed with either a multi-service resource or single-service resource.
* Cognitive Services can be deployed through the Azure portal or with the CLI.
* You need both the endpoint and key to use an Azure Cognitive Services resource.
* Computer vision analyzes still images and video streams and can detect and classify images.
* Image classification categorizes images based on the content of the image.
* Object detection identifies and tags individual visual features in an image.
* Optical character recognition (OCR) extracts text from an image.
* Facial detection uses the characteristics of a face to provide attributes about the face.
* Computer Vision service can perform many operations on an image. Analyzing the image can detect objects, describe the image in a single sentence, tag the objects in the image, identify brands and landmarks, extract text, and identify inappropriate content.
* Object detection is a type of classification model.
* Custom Vision enables you to build, train, and deploy a custom image recognition model based on images you provide when the prebuilt Computer Vision service does not explain your domain.
* Custom Vision creates a custom image recognition model.
* Custom Vision requires you to upload your images, tag the images, train, and evaluate your model.
* Custom Vision can only perform image classification or object detection.
* Computer Vision service can detect faces in images but only provides basic information about the person and face. The Face service provides more detailed facial analysis.
* The Face service uses facial landmarks to analyze and identify faces.
* The Face service can detect faces and extract attributes about the face.
* The Face service can perform facial recognition.
* Only the Verify operation should be used to identify a person from an image of their face.
* The Form Recognizer service extracts structured contextually aware information from images and documents.

***NEED MORE REVIEW?*   HANDS-ON LABS**

For more hands-on experience with Computer Vision, complete labs 1 to 6 at <https://github.com/MicrosoftLearning/mslearn-ai900>.

**Thought experiment**

Let’s apply what you have learned in this chapter. In this thought experiment, demonstrate your skills and knowledge of the topics covered in this chapter. You can find the answers in the section that follows.

You work for Fabrikam, Inc., a vehicle insurance company. Fabrikam is interested in ­processing the many images and documents that customers and assessors send to the company using AI.

Fabrikam wants to evaluate how Cognitive Services can improve their document processing time accuracy.

Fabrikam has recently created an app for customers to send in details of incidents and upload photographs of damage to their vehicles. Fabrikam wants the app to assess the level of damage from the photographs taken.

The app requests that customers take a photo of the driver after an incident. The app also requests that customers take several pictures of the scene of an incident, showing any other vehicles involved and the street. Customers are able upload dashcam videos as evidence for their claims. Customers can also upload scanned images of their claim forms that also contain a diagram explaining the incident.

Insurance adjustors have a mobile app where they can assess and document vehicle damage. Fabrikam wants the app to assess the cost of repairs based on photographs and other information about the vehicle.

Answer the following questions:

1. Assessing the damage to a vehicle from a photograph is an example of which type of computer vision?
2. You need to capture the license plates of the vehicles involved in an incident. Which type of computer vision should you use?
3. You need to confirm that the driver is insured to drive the vehicle. Which type of ­computer vision should you use?
4. You need to automatically identify the vehicles in the image. Which type of computer vision should you use?
5. Can you use the free tier to create a single resource for all these requirements?
6. You are unable to process some high-quality photographs that customers upload. Can you configure Computer Vision to process these images?
7. You need to prevent your employees from seeing inappropriate customer uploaded content. Which service should you use?
8. Which service should you use to assess the level of damage to a vehicle from a photograph?
9. Which model type should you use to assess the level of damage?
10. Which service should you use to process the scanned claim detail and diagram that the customer has uploaded?

**Thought experiment answers**

This section contains the solutions to the thought experiment. Each answer explains why the answer choice is correct.

1. Image classification is a machine learning model that predicts the category, or class, the contents of the image belong to. The categories are the level of damage involved.
2. Optical character recognition (OCR) extracts text from an image. You should use OCR to read the license plate of the vehicle but would not be able to assess the level of damage to the vehicle.
3. Identifying people in an image is an example of facial detection and recognition. Facial detection and recognition can identify people in an image.
4. Object detection will identify and tag the vehicle and may be able to identify the manufacturer and model but will not be able to assess the level of damage to the vehicle.
5. No, you cannot use the free tier with a multi-service resource. You must create resources for each Computer Vision service if you want to use the free tier.
6. No, the Computer Vision service only supports file sizes less than 4MB.
7. The Content Moderator service detects potentially offensive or undesirable content from both still images and video content. The Content Moderator provides a review tool where users can examine flagged content and approve or reject the content.
8. You should use the Custom Vision service to assess damage. A set of images can be used to train a custom model. A new image can then be categorized using the model. You would train the model with sets of vehicle images with differing levels of damage with categories (tags) that you define. The model will then be able to place any new image in one of the categories.
9. You should use the image classification model type rather than object detection. Image classification categorizes the images.
10. You should use the Form Recognizer service. This service can process both images and documents and is able to match form fields to data items, extracting the data in a structured format that your application can process.