

Create computer vision solutions with Azure AI Vision

Introduction

Azure AI Vision is a branch of artificial intelligence (AI) in which software interprets visual input, often from images or video feeds.

In this module, we'll learn how to use the **Azure AI Vision** service to extract information from images.

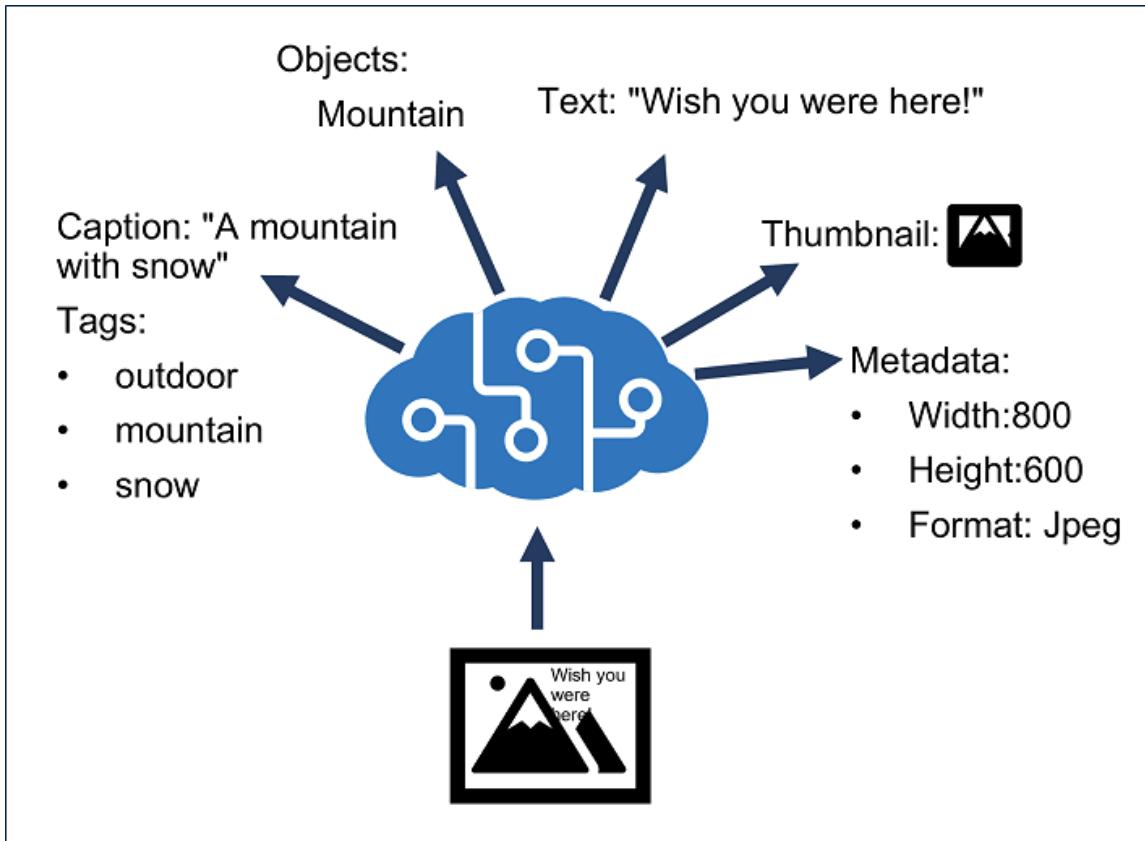
After completing this module, we'll be able to:

- Provision an Azure AI Vision resource.
- Analyze an image.
- Remove an image background.
- Generate a smart cropped thumbnail.

Provision an Azure AI Vision resource

The **Azure AI Vision** service is designed to help you extract information from images. It provides functionality that you can use for:

- *Description and tag generation* - determining an appropriate caption for an image, and identifying relevant "tags" that can be used as keywords to indicate its subject.
- *Object detection* - detecting the presence and location of specific objects within the image.
- *People detection* - detecting the presence, location, and features of people in the image.
- *Image metadata, color, and type analysis* - determining the format and size of an image, its dominant color palette, and whether it contains clip art.
- *Category identification* - identifying an appropriate categorization for the image, and if it contains any known landmarks.
- *Background removal* - detecting the background in an image and output the image with the background transparent or a greyscale alpha matte image.
- *Moderation rating* - determine if the image includes any adult or violent content.
- *Optical character recognition* - reading text in the image.
- *Smart thumbnail generation* - identifying the main region of interest in the image to create a smaller "thumbnail" version.



Analyze an image

To analyze an image, we can use the **Analyze Image** REST method or the equivalent method in the SDK for our preferred programming language, specifying the visual features we want to include in the analysis (and if we select categories, whether or not to include details of celebrities or landmarks). This method returns a JSON document containing the requested information.

```
from azure.ai.vision.imageanalysis import ImageAnalysisClient
from azure.ai.vision.imageanalysis.models import VisualFeatures
from azure.core.credentials import AzureKeyCredential

client = ImageAnalysisClient(
    endpoint=os.environ["ENDPOINT"],
    credential=AzureKeyCredential(os.environ["KEY"]))
```

```
)  
  
    result = client.analyze(  
        image_url=<url>,  
        visual_features=[VisualFeatures.CAPTION, VisualFeatures.OBJECTS],  
        gender_neutral_caption=True,  
        language="en",  
)
```

Available visual features are contained in the `VisualFeatures` enum:

- `VisualFeatures.TAGS`: Identifies tags about the image, including objects, scenery, setting, and actions
- `VisualFeatures.OBJECTS`: Returns the bounding box for each detected object
- `VisualFeatures.CAPTION`: Generates a caption of the image in natural language
- `VisualFeatures.DENSE_CAPTIONS`: Generates more detailed captions for the objects detected
- `VisualFeatures.PEOPLE`: Returns the bounding box for detected people
- `VisualFeatures.SMART_CROPS`: Returns the bounding box of the specified aspect ratio for the area of interest
- `VisualFeatures.READ`: Extracts readable text

Specifying the visual features we want analyzed in the image determines what information the response will contain. Most responses will contain a bounding box (if a location in the image is reasonable) or a confidence score (for features such as tags or captions).

The JSON response for image analysis looks similar to this example, depending on your requested features:

```
{  
    "apim-request-id": "abcde-1234-5678-9012-f1g2h3i4j5k6",  
    "modelVersion": "<version>",
```

```
"denseCaptionsResult": {
    "values": [
        {
            "text": "a house in the woods",
            "confidence": 0.7055229544639587,
            "boundingBox": {
                "x": 0,
                "y": 0,
                "w": 640,
                "h": 640
            }
        },
        {
            "text": "a trailer with a door and windows",
            "confidence": 0.6675070524215698,
            "boundingBox": {
                "x": 214,
                "y": 434,
                "w": 154,
                "h": 108
            }
        }
    ]
},
"metadata": {
    "width": 640,
    "height": 640
}
}
```

Generate a smart-cropped thumbnail and remove background

Thumbnails are often used to provide smaller versions of images in applications and websites. For example, a tourism site might display a list of tourist attractions in a city with a small, representative thumbnail image for each attraction; and only display the full image when the user selects the "details" page for an individual attraction.

The Azure AI Vision service enables us to create a thumbnail with different dimensions (and aspect ratio) from the source image, and optionally to use image analysis to determine the *region of interest* in the image (its main subject) and make that the focus of the thumbnail. This ability to determine the region of interest is especially useful when cropping the image to change its aspect ratio.



We can specify the aspect ratio of the cropped image (width / height), ranging from `0.75` to `1.80`.

Remove image background

The background removal feature can split the image into the subject in the foreground, and everything else that is considered background. Azure AI Vision achieves this feature by creating an *alpha matte* of the foreground subject, which is then used to return either the foreground or the background.

For example, take this image original of a skateboarder.



With the background removed, we get just the skateboarder on a transparent background.



When creating an alpha matte of an image, the result is the foreground in all white, with a black background.



Alpha matte images are helpful when client applications intend to do further processing of an image that requires separation of foreground and background objects.