

API (Application Programming Interface)

The textbook definition for API is: An application programming interface (API) is a set of routines, protocols, and tools for building software applications. An API expresses a software component in terms of its operations, inputs, outputs, and underlying types.

APIs often come in the form of a library that includes specifications for routines, data structures, object classes, and variables. An API is the messenger that runs and delivers your request to the provider you're requesting it from, and then delivers the response back to you.

For example:

Without API:

An app finds the current weather in London by opening <http://www.weather.com/> and reading the webpage like a human does, interpreting the content.

With API:

An app finds the current weather in London by sending a message to the weather.com API (in a structured format like JSON). The weather.com API then replies with a structured response.

Specifically, we will research Microsoft Computer Vision API.

Microsoft Computer Vision API

Microsoft Computer Vision API can:

- Tag images based on content.
- Categorize images.
- Identify the type and quality of images.
- Detect human faces and return their coordinates.
- Recognize domain-specific content.
- Generate descriptions of the content.
- Use optical character recognition to identify text found in images.
- Distinguish color schemes.
- Flag adult content.
- Crop photos to be used as thumbnails.

Analyze Image Method

Two input methods are supported -- (1) Uploading an image or (2) specifying an image URL. Within your request, there is an optional parameter to allow you to choose which features to return. By default, image categories are returned in the response.

Such responses can be:

1. **Categories** - categorizes image content according to a taxonomy defined in documentation.
2. **Tags** - tags the image with a detailed list of words related to the image content.
3. **Description** - describes the image content with a complete English sentence.
4. **Faces** - detects if faces are present. If present, generate coordinates, gender and age.
5. **ImageType** - detects if image is clipart or a line drawing.
6. **Color** - determines the accent color, dominant color, and whether an image is black&white.
7. **Adult** - detects if the image is pornographic in nature (depicts nudity or a sex act). Sexually suggestive content is also detected.
8. **Celebrities** - identifies celebrities if detected in the image.

Optical Character Recognition

Optical Character Recognition (OCR) detects text in an image and extracts the recognized characters into a machine-usable character stream.

Responses for OCR are:

1. **TextAngle** - The angle, in degrees, of the detected text with respect to the closest horizontal or vertical direction.
2. **Orientation** - Orientation of the text recognized in the image.
3. **Language** - The BCP-47 language code of the text detected in image.
4. **Regions** - An array of objects, where each object represents a region of recognized text.
5. **Lines** - An array of objects, where each object represents a line of recognized text.
6. **Words** - An array of objects, where each object represents a recognized word.
7. **BoundingBox** - Bounding box of a recognized region, line, or word, depending on the parent object.

Thumbnail Generator

This operation generates a thumbnail image with the user-specified width and height. By default, the service analyzes the image, identifies the region of interest (ROI), and generates smart cropping coordinates based on the ROI. Smart cropping helps when you specify an aspect ratio that differs from that of the input image.

The width and height of the generated thumbnail and if the smart cropping is enabled or disabled can be requested from API.

In response, a thumbnail image with specified width and height will be returned.