

画像分析

(Image Analysis)

画像分析 (Image Analysis) の位置付け

Azure AI services

- └ Azure AI Vision (Computer Vision)

- └ Image Analysis service (画像分析) 3.2

- └ Image Analysis service (画像分析) 4.0

- └ Image captions (画像キャプション)

- └ Dense Captions (高密度キャプション)

- └ Image tagging (画像のタグ付け)

- └ Object detection (オブジェクト検出)

- └ People detection (人物検出)

- └ OCR (文字読み取り)

画像分析の概要

- 公式ドキュメント
- 画像からさまざまな視覚的特徴を抽出する
- 新機能が追加される場合がある
 - マルチモーダル埋め込み（画像/テキストのベクトル化）
- 提供されていた機能が「非推奨」（廃止）となる場合がある
 - 廃止済みとなった機能
 - 背景削除
 - 製品認識
 - モデルのカスタマイズ
 - 廃止機能の利用が必要な場合は「[Azure AI Custom Vision](#)」に移行するか、オープンソース・サードパーティの製品を使用する。

画像分析のAPIのバージョン

- Image Analysis APIには、古い [3.2 系](#)と、新しい [4.0 系](#)がある
 - 本資料では4.0系について解説
- [2022/10/12 Computer Vision Image Analysis 4.0 プレビュー開始](#)
- 4.0系の特徴
 - 3.2系よりも優れたモデルを使用
 - 1 回の API 呼び出しで、画像のキャプション、タグ付け、読み取り (OCR)、人の検出、スマートトリミング、オブジェクト検出などを一度に実行できる
 - OCR では、キリル語、アラビア語、ヒンディー語を含む 164 の言語がサポートされるようになった

画像分析 4.0 に含まれる主な機能(1)

- **画像キャプション**: 画像全体に対して1つの説明文を生成（座標なし）
- **高密度キャプション**: 画像全体に対する1つの説明文（座標なし）と、画像の最大 10 個の領域それぞれの説明文（座標あり）を生成
 - 座標: Bounding Box (x, y, w, h) 、対象領域の左、上、幅、高さ（ピクセル）。
- **画像のタグ付け**: 画像に含まれるモノ、人、動物などの情報（タグ）を返す（座標なし）
- **オブジェクト検出**: 画像の中の各物体の情報（タグ）を返す（座標あり）
- **人物検出**: 画像に含まれる各人物の座標を返す

画像分析 4.0 に含まれる主な機能(2)

- 文字読み取り(OCR): 画像に含まれる単語や文章を読み取りテキストを返す
(座標あり)
 - ※OCR: Optical Character Reading、光学式文字読み取り
- マルチモーダル埋め込み (画像とテキストのベクトル化)
 - VectorizeImage API (画像のベクトル化) / VectorizeText API (テキストのベクトル化) によるベクトル化、ベクトルの「類似度」比較に利用できる

画像分析を行うためのAzureリソース

画像分析には以下のいずれかのリソースが必要。

- 「Computer Vision」リソース
 - 価格レベル: Free (F0) / Standard (S1)
- 「Azure AI services マルチサービスアカウント」リソース
 - 価格レベル: Standard (S1)
- 「Azure AI services」リソース
 - 価格レベル: Standard (S1)

※リージョンに注意: 「画像キャプション」や「高密度キャプション」を使用するためには、GPUが使用できるリージョン（米国東部、フランス中部、韓国中部、北ヨーロッパ、東南アジア、西ヨーロッパ、米国西部、東アジア）にリソースを作成する必要がある。

画像分析の料金（コスト） (1)

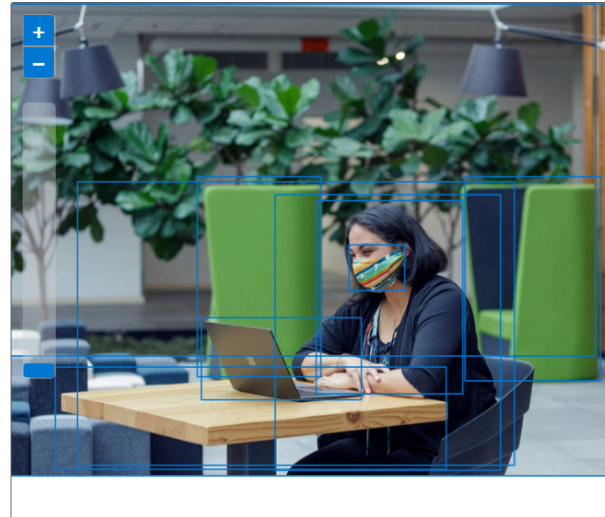
- 価格のページ(Computer Vision)
- トランザクション
 - それぞれの機能の利用を「トランザクション」としてカウント
 - ※ transaction = 取引、ITの処理単位
 - 画像分析バージョン 4.0 では、1回のAPI呼び出しで複数の画像分析機能が実行できる
 - たとえば「画像キャプション」と「オブジェクト検出」を1回のAPI呼び出しで実行した場合、2つのトランザクションとしてカウント

画像分析の料金（コスト） (2)

- 価格レベル Free (F0) のリソース
 - 無料で利用できるが、1 か月あたり 5,000 件まで（クォータ）、1 分あたり 20 件まで（レート制限）といった制限がある
- 価格レベル Standard (S1) のリソース
 - クォータ制限なし。より高いレートでの利用が可能
 - 機能により単価が異なる
 - 例: 「画像キャプション」: \$1.50/1,000 トランザクション など

- Computer Visionの機能を簡単に試することができるサイト。
- <https://portal.vision.cognitive.azure.com/>
- 「高密度キャプション」の利用例

Sample image 2



Detected attributes JSON

```
A woman wearing a mask sitting at a table with a laptop
A person using a laptop
A woman wearing a face mask
A close-up of a green chair
A person wearing a colorful cloth face mask
A person using a laptop
A woman wearing a colorful fabric face mask
A close-up of a laptop on a table
A woman wearing a face mask
A woman wearing a face mask sitting at a table with a laptop
```

- SDK = Software Development Kit
- C# / Python / Java / JavaScript の4言語のSDKが提供されている。
- SDKの概要
 - 各言語でのSDKの提供状況など
- クイックスタート
 - リソースの作成、プロジェクトの作成、SDKの導入、コードの記述と実行などの具体的な手順

SDKの利用例(C#)

```
// キーとエンドポイントを環境変数から読み取り
string endpoint = Environment.GetEnvironmentVariable("VISION_ENDPOINT");
string key = Environment.GetEnvironmentVariable("VISION_KEY");

// クライアントを作成
ImageAnalysisClient client = new ImageAnalysisClient(
    new Uri(endpoint),
    new AzureKeyCredential(key));

// 「画像キャプション」と「OCR（文字読み取り）」を同時に実行
ImageAnalysisResult result = client.Analyze(new Uri("画像のURL"),
    VisualFeatures.Caption | VisualFeatures.Read);

// キャプションを表示
Console.WriteLine($"Caption: {result.Caption.Text}");

// 読み取った文字を表示
foreach (DetectedTextBlock block in result.Read.Blocks)
{
    foreach (DetectedTextLine line in block.Lines)
    {
        foreach (DetectedTextWord word in line.Words)
        {
            Console.WriteLine($"Word: '{word.Text}'");
        }
    }
}
```

実行結果例

<https://learn.microsoft.com/en-us/azure/ai-services/computer-vision/media/quickstarts/presentation.png>



Caption: a person pointing at a screen

Word: 9:35

Word: AM

Word: Conference room 154584354

Word: 555-123-4567

...