

Azure で コンピューター ビジョン ソリューションを開発する



Azure でコンピューター ビジョン ソリ
ューションを開発する

目次

- 画像を分析する
- 画像内のテキストを読み取る (OCR)
- 顔を検出、分析、認識する (Face)
- 画像を分類する
- ビジョン対応のジェネレーティブ AI アプリケーションを開発する

目次



Azure AI
Computer Vision

- 画像を分析する
- 画像内のテキストを読み取る (OCR)
- 顔を検出、分析、認識する (Face)

- 画像を分類する

- ビジョン対応のジェネレーティブ AI アプリケーションを開発する



Azure AI
Custom Vision



+

生成AIモデル

Azure AI Foundry
プロジェクト

モジュール1

- 画像を分析する
- 画像内のテキストを読み取る (OCR)
- 顔を検出、分析、認識する (Face)
- 画像を分類する
- ビジョン対応のジェネレーティブ AI アプリケーションを開発する

「画像の分析」とは？

- 画像処理のAIサービス
- 画像からさまざまな視覚的特徴を抽出
- 画像のキャプション（説明文）の生成、オブジェクトの検出、人物の検出、画像へのタグ付けなどの機能を含む
- 現在2系統のAPIバージョンが提供されている
 - 4.0
 - より新しい機能が利用できる。基本的にはこちらを利用
 - ただし 3.2 で提供されていたすべての機能が利用できるわけではない
 - 3.2
 - 4.0で利用できない機能が必要な場合はこちらを利用
 - 例: 「クリップアート検出」（挿絵イラストの検出）



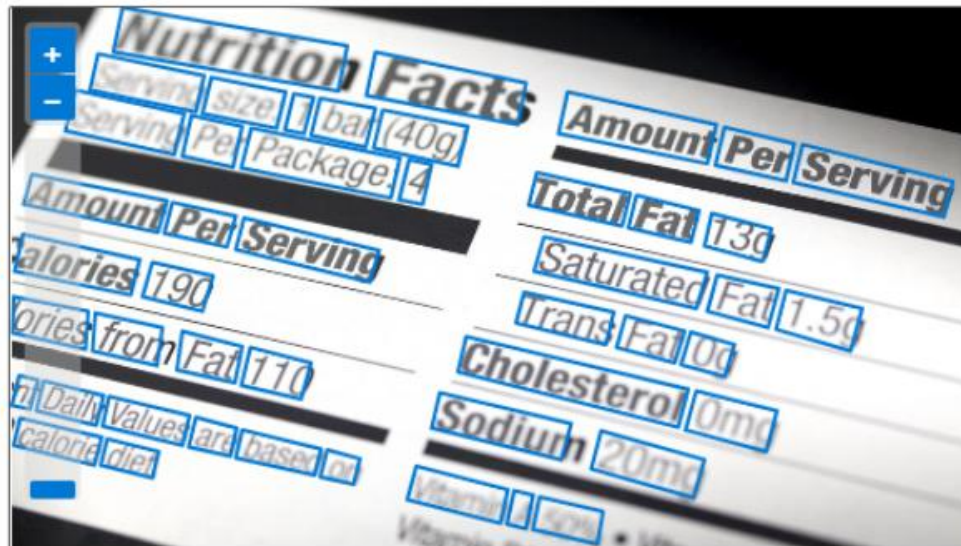
クリップアートの例

モジュール2

- 画像を分析する
- 画像内のテキストを読み取る（OCR）
- 顔を検出、分析、認識する（Face）
- 画像を分類する
- ビジョン対応のジェネレーティブ AI アプリケーションを開発する

Azure AI Computer Vision の「OCR」とは？

- Optical Character Reader（光学式文字読み取り）
- 画像に含まれる文字列を認識し、抽出する
- 手書き、印刷、彫刻文字などを読み取り可能
- 日本語を含む多数の言語に対応



Nutrition Facts Amount Per Serving
Serving size: 1 bar (40g)
Serving Per Package: 4
Total Fat 13g
Saturated Fat 1.5g
Amount Per Serving
Trans Fat 0g
calories 190
Cholesterol 0mg
calories from Fat 110
Sodium 20mg
nt Daily Values are based on

OCRはDockerコンテナでの運用も可能

※料金は通常通り発生

■ Dockerコンテナを使わない場合

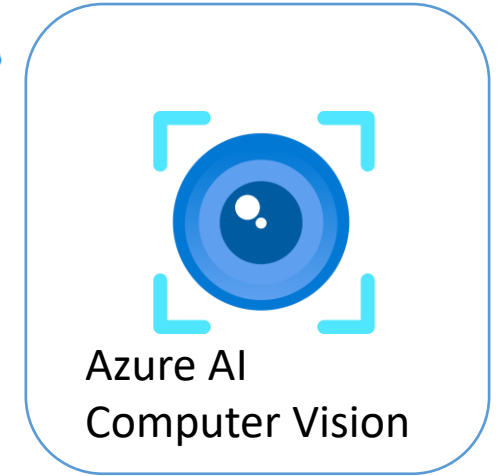


OCR処理をAzure上で実行
(画像がAzureに送信される)

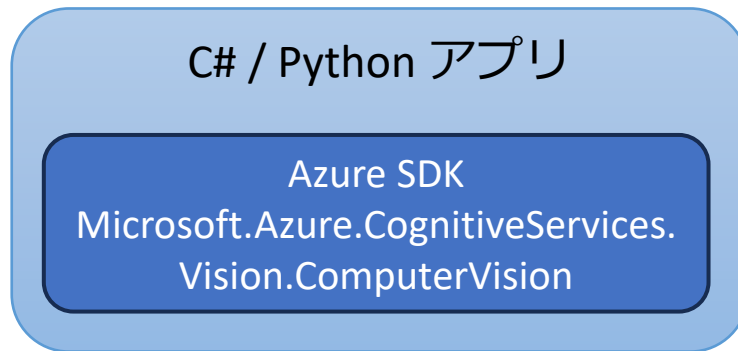
読み取り結果



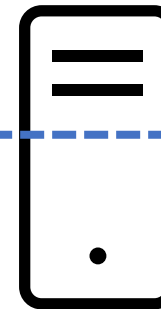
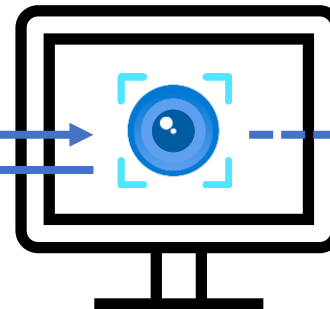
Azure



■ Dockerコンテナを使う場合



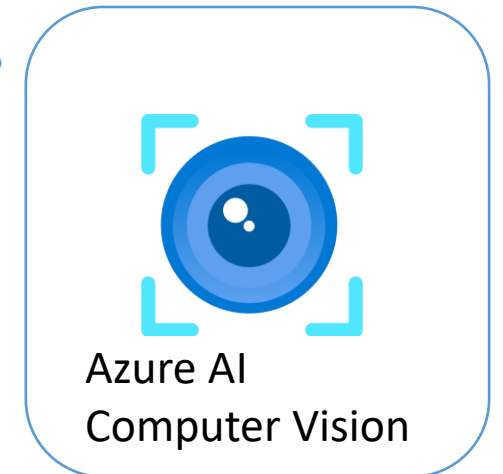
OCR処理をオンプレミスサーバーの
Dockerコンテナ内で実行
(画像はAzureに送信されない)



利用状況
データだけ
Azureに
送信される

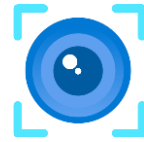


Azure



「Azure AI Computer Vision の OCR」 vs 「Azure AI Document Intelligence」

名刺のスキャン画像



Azure AI Computer Vision の OCR での読み取り結果

"CONTOSO", "Chris Smith", "Senior Researcher",
"Cloud & AI Department", "chris@contoso.com", ...

単に、認識できた文字列を返すだけ





Azure AI Document Intelligence での読み取り結果

CompanyName (社名) = "Contoso"
ContactName (氏名) = "Chris Smith"
といったように、読み取った文字列が
何の項目なのかという情報も得られる

Page # / Field name / Value	Confidence %
1 CompanyNames 76.8%	
CONTOSO	
1 ContactNames 99.0%	
Chris Smith	
1 JobTitles 99.0%	
Senior Researcher	
1 Departments 98.6%	
Cloud & AI Department	
1 Emails 99.0%	
chris.smith@contoso.com	

1 Websites 99.5%	
https://www.contoso.com/	
1 MobilePhones 99.5%	
+1 (987) 123-4567	
1 OtherPhones 98.8%	
+1 (987) 213-5674	
1 Faxes 98.9%	
+1 (987) 312-6745	
1 Addresses 98.7%	
4001 1st Ave NE Redmond, WA 98052	

「Azure AI Computer Vision の OCR」 vs 「Azure AI Document Intelligence」

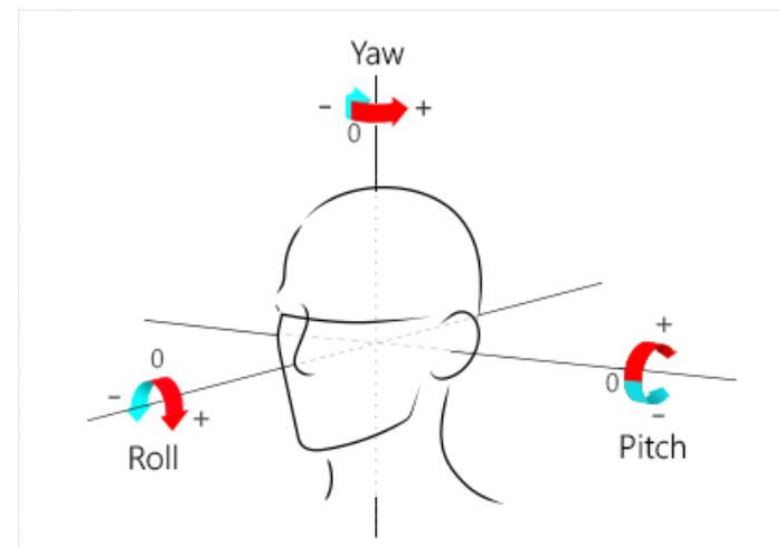
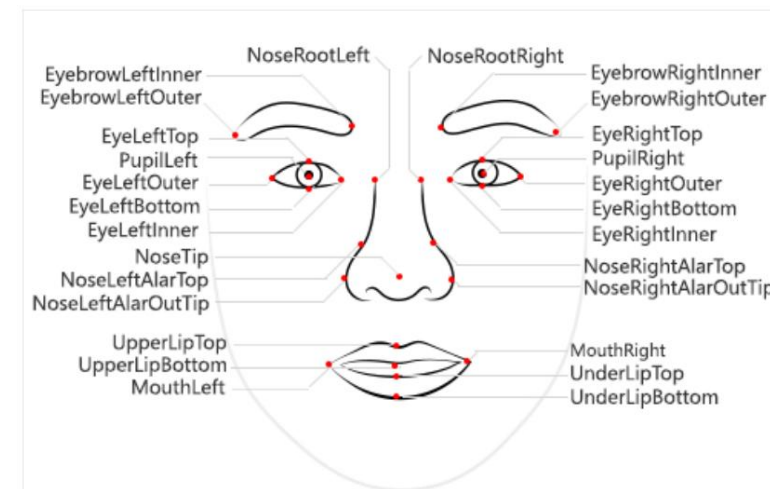
- 共通機能
 - 印刷テキストや手書きテキストの読み取りができる
- Azure AI Computer Vision の OCR 
 - 画像から文字列を読み取る
 - より基礎的なサービス
- Azure AI Document Intelligence 
 - 領収書・請求書などのスキャン画像などからの情報読み取りができる
 - PDF/Word/Excel/PowerPointなどからの情報読み取りができる
 - 大量のページの読み取りにも対応
 - **ドキュメント構造（表、項目など）の認識ができる**
 - より応用的なサービス

モジュール3

- 画像を分析する
- 画像内のテキストを読み取る (OCR)
- 顔を検出、分析、認識する (Face)
- 画像を分類する
- ビジョン対応のジェネレーティブ AI アプリケーションを開発する

Azure AI Computer Visionの「Face」とは？

- 画像に含まれる「顔」の情報を分析
- 目・鼻・口などの位置情報
- 顔の角度の情報
- メガネ・マスクの有無などの情報



モジュール4

- 画像を分析する
- 画像内のテキストを読み取る (OCR)
- 顔を検出、分析、認識する (Face)
- 画像を分類する
- ビジョン対応のジェネレーティブ AI アプリケーションを開発する

Azure AI Custom Visionとは？

- ユーザーが「コンピュータビジョンモデル」をトレーニングして利用することで、カスタムの画像の分類や検出を行うことができる
- 例えば Azure AI Computer Vision では「画像に猫が写っているかどうか」は判定できるが「猫の種類」までは判定できない
- Azure AI Custom Visionでは、ユーザーがたくさんの猫画像を用意してモデルをトレーニングすることで、猫画像から「猫の種類」を判定できるモデルを作成・運用できる



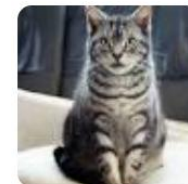
ブリティッシュ
ショートヘア



メインクーン



ペルシャ



アメリカンショ
ートヘア



パーミーズ



エキゾチックシ
ョートヘア

モジュール5

- 画像を分析する
 - 画像内のテキストを読み取る（OCR）
 - 顔を検出、分析、認識する（Face）
 - 画像を分類する
- ビジョン対応のジェネレーティブ AI アプリケーションを開発する

「ビジョン対応 生成AIアプリ」のイメージ

お手伝いできることはありますか？



この画像はなんですか？

+ ツール



マルチモーダル（モデル）

multimodal model

- 複数の種類の情報（モダリティ modality）を同時に扱えるAIモデル
- 「モダリティ」とは、モデルが扱う入力や出力の形式・種類のこと。
例:
 - テキスト（文章や言葉）
 - **画像（写真、イラストなど）**
 - 音声（話し声、環境音など）
 - 動画（動きと音を含む情報）
 - センサー情報（位置情報、温度など）

Phi-4-multimodal-instruct

- **マルチモーダル対応:** テキスト、ビジュアル、音声入力をサポートするマルチモーダルモデル
- **高度な推論能力:** 画像や音声入力を組み合わせて高度な推論を行う
- **エッジデバイスへの展開:** このモデルは、PC、モバイルデバイス、IoTシステムなどのエッジデバイスに展開可能
- **関数呼び出し機能:** 検索エンジンやさまざまなツールと連携するための関数呼び出し機能をサポート