

ドコモ画像認識プラットフォームを用いた画像認識AI提供

サービスイノベーション部 **さかい としき** 酒井 俊樹 **いわた もとき** 岩田 基暉

近年、AI技術が普及し、画像認識の分野においてAIを用いた、人の作業の代替、自動化、省人化が図られているが、その導入には、高性能なサーバ環境の準備、ソフトウェアの導入といったAI開発環境の整備が必要であった。そこでドコモでは、画像認識AI開発用のデータを準備すれば、さまざまな種類の画像認識AIの開発、デプロイが可能となる「ドコモ画像認識プラットフォーム」を開発した。これにより、AIの利用企業は自前のAI開発・運用環境を用意しなくても、画像認識AI技術を導入することが可能となった。

1. まえがき

近年、Deep Learning^{*1}を用いたAI技術の発展、普及は目覚ましく、特に画像認識AI^{*2}の産業応用が進んでいる。例えば、工場における検品の補助、防犯カメラ映像での人や車両の検知、医療における画像診断の補助など、今までは人が行っていた作業の補助、代替として活用されている。しかし、画像認識AIを用いて、既存の作業の効率化、自動化、DX (Digital Transformation)^{*3}を行う際、各企業が解決したい課題をそのまま解決できる画像認識AIが提供されていることは少ない。このため各企業が、保持している画像／動画データと、当該デー

タに対して画像認識AIにどのような認識処理を行わせたいか、その答えを記録したアノテーションデータを準備した上で、企業独自の画像認識AIを開発する「学習」と、学習した画像認識AIを使える状態にする「デプロイ^{*4}」の作業が必要となる。デプロイにおいては、作成した画像認識AIをサーバ上に展開し、画像の入力を受け付け、結果をテキストなどで返却するAPI (Application Programming Interface)^{*5}を構築する。しかし、学習とデプロイをAIの利用企業が自身で行うには、Deep Learningを扱うためのフレームワーク^{*6}の理解、高速な学習、認識を行うためのハードウェアの準備などの課題がある。

©2022 NTT DOCOMO, INC.

本誌掲載記事の無断転載を禁じます。

本誌に掲載されている社名、製品およびソフトウェア、サービスなどの名称は、各社の商標または登録商標。

^{*1} Deep Learning：人の神経処理メカニズムを模して作られたニューラルネットワークをより複雑にすることによって、より複雑な概念の学習や、判断、推定の模倣を可能とした機械学習手法。

^{*2} 画像認識AI：画像入力に対して、機械が判断、推定などを行い、結果を返却するAI。

このような課題に対処すべく、画像認識の学習、デプロイをクラウド上で手軽に行えるようにするサービスが提供されている。例えば、Amazon Rekognition Custom Labels [1]、AutoML Vision [2] などがある。ドコモでも、自社で開発してきた画像認識技術*7を、より多くのユーザに手軽に使っていただくべく、ドコモ画像認識プラットフォーム [3] を2020年に提供開始した。

画像認識AIの学習・デプロイ機能の提供にあたっては、ユーザごとに必要とする画像認識機能が異なるため、それらをより効率的に提供する仕組みが求められる。本稿では、ドコモ画像認識プラットフォームにおいて、複数の画像認識機能を提供するための仕組みについて解説する。また、ドコモ画像認識プラットフォームでは、ドコモの閉域網内で画像認識を行うことで、よりセキュアな画像認識環境を実現しており、このメリットについても解説する。

2. ドコモ画像認識プラットフォームの概要

ドコモ画像認識プラットフォームは、クラウド上で画像認識AIの学習および画像認識AIを用いた認

識（以下、「推論」）を行える状態にするデプロイの機能を提供する。

(1)提供される画像認識機能

ドコモ画像認識プラットフォームが提供する画像認識機能の一覧を図1に示す。①物体検出は、画像の中から特定の物体を見つけ、画像内でのその物体の座標を推定する機能である。②一般物体認識は、画像に写っている物、風景、シーン、画像全体の特徴から、カテゴリに分類する機能である。③文字認識とは、画像内に写っている文字を認識する機能である。④類似画像検索は、画像と似た画像を探す機能である。⑤特定物体認識は、あらかじめ用意した多数の物体の画像と認識したい物体の画像とを照合することで、認識したい物体が何かを特定する機能である。⑥姿勢推定とは、画像内に写っている人の骨格／関節の位置を推定する機能である。

(2)想定ユースケース

各画像認識機能の想定ユースケースを図2に示す。物体検出は、ドローン映像からの雑草検出や、監視カメラ映像からの人や車の検出に活用できる。検出した植物や、車の画像領域に対して、一般物体認識を組み合わせることで、植物の生育度や車種の判別が可能となる。また、物体検出と文字認識を組み合

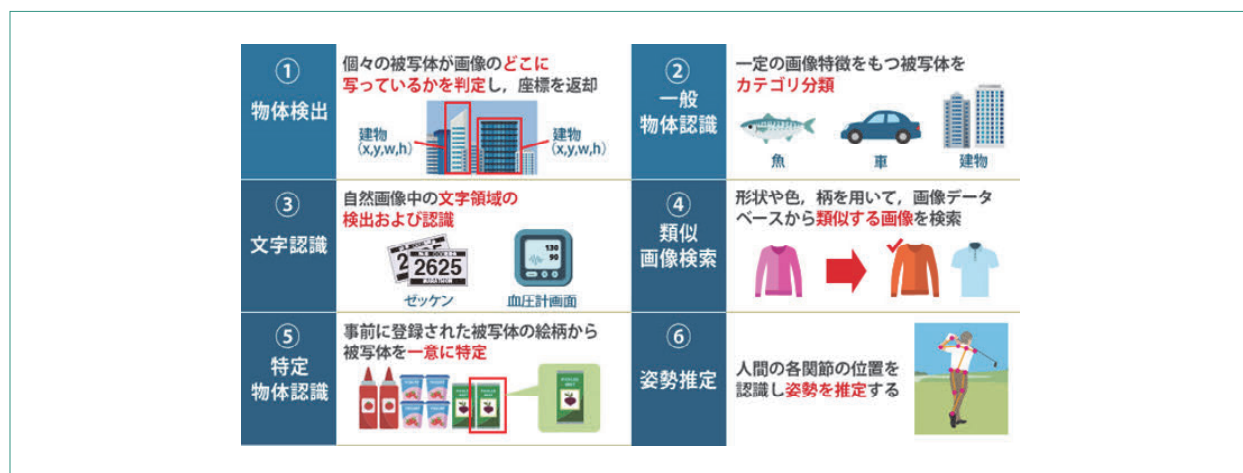


図1 ドコモ画像認識プラットフォームで提供する画像認識機能

*3 DX：ITの浸透が人々の生活をあらゆる面で良い方向に変化させること。
*4 デプロイ：アプリケーションをそれらの実行環境に配置して展開すること。
*5 API：ソフトウェアの機能を他のプログラムから利用できるように切り出したインタフェース。

*6 フレームワーク：ある領域のソフトウェアに必要とされる汎用的な機能や基本的な制御構造をまとめたもの。ライブラリでは、開発者が個別の機能呼び出す形となるが、フレームワークでは、全体を制御するのはフレームワーク側のコードで、そこから開発者が個別に追加した機能呼び出す形となる。



図2 画像認識の想定ユースケース

わせた伝票や書類から文字を読み込む利用方法や、物体検出と特定物体認識を組み合わせた商品棚の画像を撮影するだけで、どこに何の商品が置いてあるかデータ化する利用方法などが想定される。

さらに、類似画像検索では、似たファッションアイテムを画像ベースで探す、姿勢推定ではスポーツの動作のフォームを確認するなどの用途で活用できる。

(3)機能構成

ドコモ画像認識プラットフォームの仕組みを図3に示す。ドコモ画像認識プラットフォームでは、「物体検出」「一般物体認識（カテゴリ分類）」においては、ユーザのデータで独自のAIを提供できる学習機能と推論機能の両方をカスタム学習モデルとして提供している。また、「類似画像検索」「特定物体認識」は、ユーザ独自の画像データを基に検索・認識ができる「ユーザ辞書の作成機能」を提供している。「物体検出」「一般物体認識（カテゴリ分類）」「文字認識」「姿勢推定」においては、ユーザが独自に学習用のデータを用意しなくても推論が可能なよ

う、ドコモにて独自に学習した画像認識AI（共通学習済み学習モデル）も提供している。

(4)Webサイト／Webコンソールの提供

ドコモ画像認識プラットフォームでは、ユーザ向けにWebサイト／Webコンソールを提供している。本コンソールにアクセスすることで、学習、学習結果の確認（「評価」）、学習して作成した画像認識AIのデプロイまでが実施できる。デプロイされたAIにはWebAPIを通して、認識リクエスト（推論リクエスト）を行うことができる。画像データをWebAPIに入力すると、認識結果がテキスト形式で返却される。

3. ドコモ画像認識プラットフォームにおける2つの取組み

ドコモ画像認識プラットフォームでは、2つの取組みでユーザの利便性の向上を図っている。1つ目が、画像認識機能をコンテナ化することであり、こ

*7 画像認識技術：画像処理技術や機械学習（*8参照）技術を用いて、画像を機械に理解させ、意味を取り出す技術。

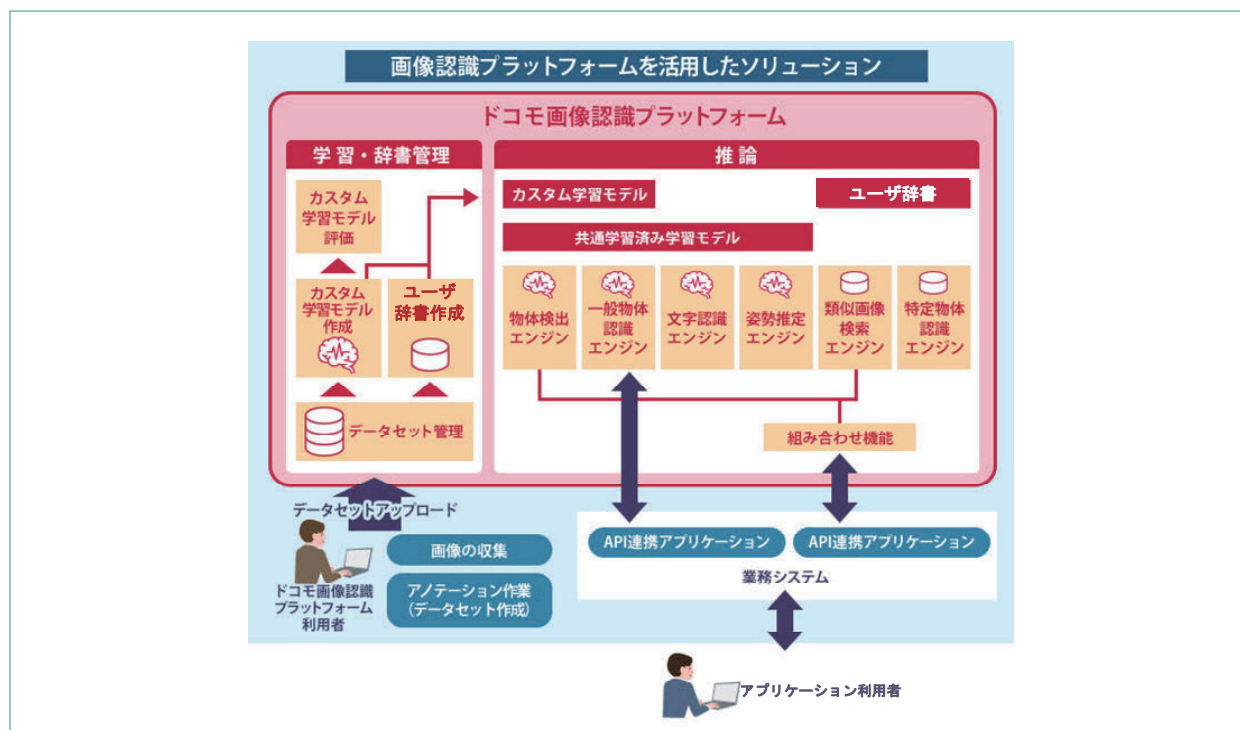


図3 ドコモ画像認識プラットフォームの仕組み

れにより画像認識機能の提供／開発／更新スピードが増加する。充実した画像認識機能のラインナップが可能となり、ユーザの課題を解決できる画像認識機能を提供できる。2つ目は、ドコモ画像認識プラットフォームをドコモの閉域網内のデータセンタ上に構築し、インターネットを介さずに接続できるようにすることである。これにより、セキュアな画像認識AIの提供を実現している。

3.1 コンテナを活用した画像認識AIのWebAPI化

前述のとおり、ドコモ画像認識プラットフォームでは、複数の画像認識機能を提供している。ユーザが必要とする画像認識機能は、ユーザの解決したい課題によって異なり、今後も増加することが予想される。また、画像認識AI技術は日々向上しており、より高精度／高速なものをユーザに提供するために

は、すでに提供している画像認識機能の逐次的なアップデートが必要となる。そこで、ドコモ画像認識プラットフォームでは、各画像認識機能の提供／開発／更新スピードの向上のため、コンテナを用いた機能の仮想化と、コンテナ内での機能のモジュール化／共通化を実施している。概念図を図4に示す。

コンテナを用いた仮想化とは、画像認識機能のようなアプリケーションとその起動に必要なライブラリなどを「コンテナ」と呼ばれるパッケージにまとめ、サーバのOS／環境に可能な限り依存せずに、アプリケーションを実行できるようにする技術である。ドコモ画像認識プラットフォームでは、各画像認識機能のコアプログラムと、画像認識機能をWebAPIとして提供するためのインタフェースを1つのコンテナとしてまとめている。これにより、ドコモ画像認識プラットフォーム内では、コンテナを追加／削除することで、新しい画像認識機能の追加

や、既存の画像認識機能のアップデート、認識リソースの増強（より多くの画像を学習／推論できるようにするための変更）が可能となっている。

また、画像認識のコンテナ内では、前述のインタフェースを司るプログラムと、画像認識のコアプログラムは独立であり、インタフェースとなるプログラムを共通化することで、新規の画像認識機能を開発する際の開発範囲を削減している。インタフェースを司る前者のプログラムと、コアプログラムとの間のインタフェースの入出力形式を、表1のとおり各コアプログラム側で柔軟に変更できるようにし、画像認識機能の拡張のための柔軟性を確保している。

加えて、画像認識AIのうちユーザごとに学習した画像認識モデルをコンテナ外に独立して保存している。これにより、ユーザの作成した特定の画像認識AIのリソースのみを増強することや、コンテナの更新時にコンテナのみを入れ替え、モデルを引き継ぐことを可能としている。

3.2 ドコモ・オープンイノベーションクラウドを活用したセキュアな画像認識

画像認識機能を利用する際には、画像というセンシティブな情報の入力が必要となるため、セキュリティの確保も課題となる。画像認識の過程で、画像

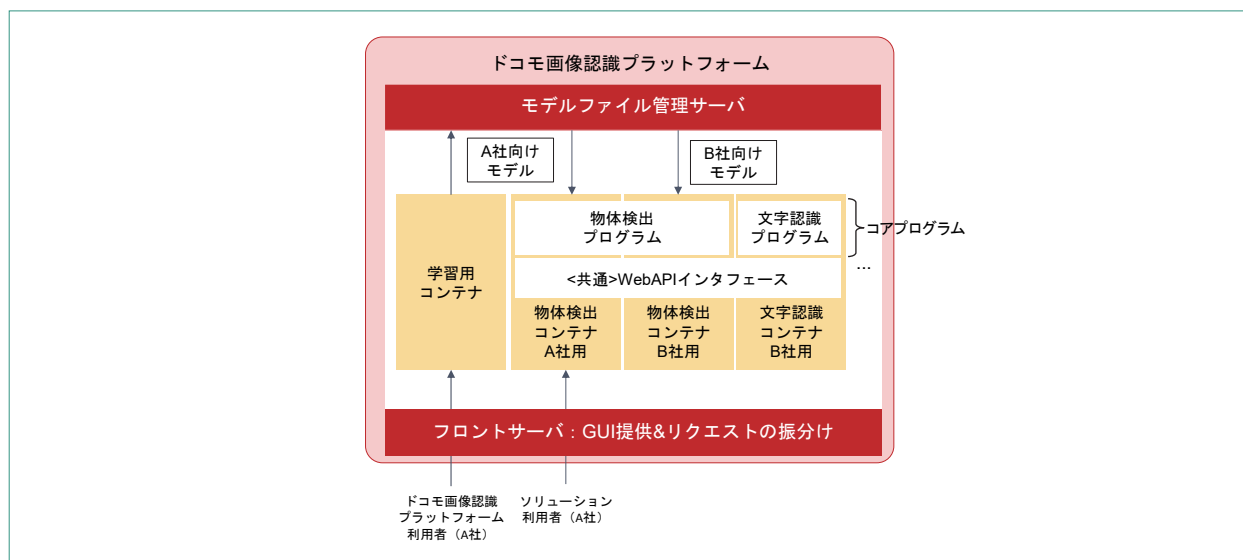


図4 コンテナ化を用いた画像認識機能提供

表1 画像認識のインタフェースプログラムとコアプログラムの役割分担

	共通化	各コアプログラムにて定義
コンテナの起動時	<ul style="list-style-type: none"> 学習した画像認識モデル ファイルの入力方法 	<ul style="list-style-type: none"> 学習した画像認識モデルのロード方法
推論時の入力	WebAPIの入力形式 <ul style="list-style-type: none"> JSON形式 multipart/form-data形式 	WebAPIの入力の詳細 <ul style="list-style-type: none"> 画像をどのようにWebAPI入力のリクエスト内に格納するかなど
推論時の出力	WebAPIの出力形式 <ul style="list-style-type: none"> JSON形式 	WebAPIの出力の詳細 <ul style="list-style-type: none"> 画像認識結果をどのようにwebAPI出力のJSON形式内に格納するか

情報が外部に出ることがないように、画像認識機能の利用にあたっては、画像の伝送経路、および画像認識を行うサーバ側で留意が必要となる。

画像認識を行う際には、画像認識システムをクラウド上や、自前のデータセンタ上で構築することにより、処理を集約して画像認識処理を行うことができる。このため、クラウド／データセンタ上の計算リソースの利用率を上げることができ、計算リソースを効率的に利用して画像認識処理が可能となる。その際システムのセキュリティ対策を適切に実施することでリスクを低減することが考えられる。しかし、この場合も画像をクラウド／データセンタに送信する経路のセキュア化が別途必要であり、インターネットを介さずセキュア化するためには専用線を利用するなどの対策が必要となる。

この課題を解決するために、画像認識処理をローカルのPCや端末、エッジ側のデバイスで行う方法がある。これにより、画像のネットワーク上での伝送を無くすことができ、画像はデバイス内にとどまるため、セキュリティを確保することができる。一方で、エッジ側のデバイス、PCや端末は画像認識を行うには十分な性能が無く、認識処理に時間がかかるといった問題がある。また性能が十分な場合でも、画像認識を個別のデバイスで行う際には、クラ

ウド／データセンタで処理を行う場合と比べ、画像認識処理を集約してリソースの利用効率を上げることができないという課題がある。

そこでドコモ画像認識プラットフォームでは、両者の中間として、ドコモの通信網内にデータセンタ（ドコモ・オープンイノベーションクラウド）を設置し、そこで画像認識処理を行う構成をとることで、これらの問題を解決した。概念図を図5に示す。画像認識処理は、ドコモの通信網内のデータセンタで行われるため、ドコモ回線（第5世代移動通信システム（5G）もしくは第4世代移動通信システム（4G））を通じたアクセスの場合、インターネットに出ることなく通信することが可能となる。これにより、画像情報が外部に流出するリスクを減らすことができる。一方で、画像認識機能はデータセンタ側に集約されるため、計算コストの最適化も図ることが可能となる。

ドコモ画像認識プラットフォームでは、この閉域網を使った接続をクラウドダイレクト [4] 接続オプションとして提供している。この閉域網を使ったアクセスには、インターネットを経由しないため接続までの遅延が少ないというメリットもある。また、接続に5Gを用いることで、さらなる遅延の低下を図ることができる。

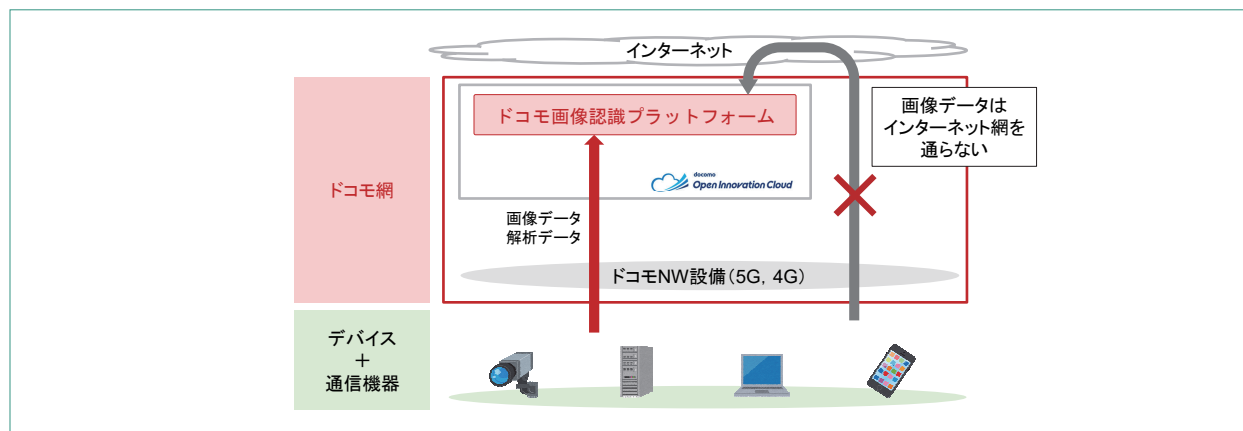


図5 閉域網接続を用いた画像認識のセキュア化

4. ドコモ画像認識プラットフォーム上の画像認識APIの構築・利用方法

ドコモ画像認識プラットフォームでは、APIを利用するまでに、大きく分けて学習とデプロイという2つの工程を経る。これらの作業をWebブラウザ上（クラウド上）で、シンプルなユーザインタフェースで可能としたのが本サービスである（図6）。学習をすることで個々のタスクにあったモデルの作成を実施し、デプロイをすることで学習したモデルや汎用的に提供されているモデルが利用可能となる。また、類似画像検索、特定物体認識では、ユーザ独自の辞書の作成ができる。ドコモ・オープンイノベーションクラウドの開発者ポータル内にて各工程の詳細なマニュアルが公開されている。

4.1 「学習」の実施

現在のドコモ画像認識プラットフォームでは、物体検出と一般物体認識の機能において学習を実施す

ることができる。具体的な学習の手順は、開発者ポータル内のチュートリアルに記載されている。

学習には画像データに加えて、アノテーションデータが必要である。アノテーションデータとは、本画像認識技術を含めた一般的な機械学習^{*8}の学習および評価の工程にて使用される、必要な情報をタグ付けされたデータのことである。学習の工程では、アノテーションデータを正解の情報としてモデルの学習を行い、評価の工程では、アノテーションデータと推論結果を比較して、作成したモデルの評価を実施する。アノテーションデータの作成のためのツールは、ドコモからサンプルツールとして提供されている。

ユーザは、ローカル環境にて、画像データとそれに対応するアノテーションデータを仕様に従ってまとめたデータセットを作成し、ドコモ画像認識プラットフォームにアップロードする。同様に検証用のデータセットも作成・アップロードを実施し、学習用データセット、検証用データセット、そして学



図6 サービス利用画面一例

*8 機械学習：サンプルデータから統計処理により、有用な判断基準をコンピュータに学習させる技術。

習アルゴリズム（物体検出or一般物体認識）を設定して学習を実施する。

学習したモデルはその精度評価をプラットフォーム上で実施することが可能である。学習・検証用のデータセットと同様に、評価用のデータセットを作成・アップロードしたのち、サービス利用画面の「学習モデル評価」から学習したモデルと評価用データセットを選択し評価を実施する。評価結果はプラットフォームからダウンロードでき、画像分類・物体検出とともに、Accuracy（正解率）*9、Precision（適合率）*10、Recall（再現性）*11、およびF1スコア*12の精度を示す各指標を確認することができる。

4.2 「デプロイ」の実施

デプロイを実施することで、学習して作成したモデルおよびドコモにて独自に学習し提供しているモデルが、Web APIを通じて利用（推論）可能となる。デプロイした画像認識AIは、APIキーを払い出して割り当てることで利用可能となる。APIの方式としてREST API (REpresentational State Transfer API)*13を採用し、シンプルな設計でユーザー側システムとの連携が可能である。

デプロイした画像認識AIは、サービス利用画面の「API管理」および「API認証キー設定」にて管理の実施ができる。API管理では、学習したモデルの管理および画像認識AIの増設・削除を実施することができ、簡単にスケールさせることができる。API認証キー設定では、デプロイした画像認識AIに対する認証キーの割当て・変更が可能である。

4.3 「辞書の作成」の実施

ドコモ画像認識プラットフォームの類似画像検索と特定物体認識では、独自の辞書を作成することが可能である。類似画像検索では、リクエストされた

画像が、事前に作成した辞書画像内のどの画像に類似しているかを判定できる。また、特定物体認識では、リクエストされた画像を、作成した辞書内の画像と照らし合わせることで、リクエストされた画像が何なのかを特定できる。

どちらの辞書作成においても、画像の収集およびアノテーションの実施が必要となる。辞書作成後は、他の画像認識機能同様にデプロイを実施することで各機能を利用することができる。

5. あとがき

本稿では、画像認識サービスを手軽に利用できるようにするにあたっての背景や課題と、ドコモ画像認識プラットフォームにおいて提供している画像認識機能について解説した。また、ドコモ画像認識プラットフォームにおいて、さまざまな画像認識機能を提供可能とするシステムの特性と、画像認識においてセキュリティ向上を図るための閉域網内での認識処理について述べた。また、実際にドコモ画像認識プラットフォームを利用する際のイメージについても解説した。今後もドコモでは、ユーザの必要とする画像認識機能を提供し続けるべく、継続的な機能改善、アップデートを行っていく予定である。

文 献

- [1] AWS：“Amazon Rekognition カスタムラベル。”
<https://aws.amazon.com/jp/rekognition/custom-labels-features/>
- [2] Google Cloud：“AutoML Vision.”
<https://cloud.google.com/vision/automl/docs>
- [3] NTTドコモ：“ドコモ画像認識プラットフォーム。”
<https://www.nttdocomo.co.jp/biz/service/dirp/>
- [4] NTTドコモ：“クラウドダイレクトとは。”
<https://developer.dev-portal.d-oi.com/document/docs/cloud-direct/concepts/overview.html>

*9 Accuracy（正解率）：推論したデータのうち、正しく分類・検出できた割合。

*10 Precision（適合率）：物体検出においては、あるラベルにおいて、検出した物体のうち、確かにそれが該当の物体である割合。一般物体認識においては、あるラベルと推論したデータのうち、その推論結果が正しい割合。

*11 Recall（再現性）：あるラベルの全データのうち、正しくそのラベルと分類（物体検出）できた割合。

*12 F1スコア：PrecisionとRecallの調和平均。

*13 REST API：RESTの制約に従ったAPI。RESTはRoy Fielding氏が2000年に提唱した設計原則を基に発展した、ソフトウェアアーキテクチャのスタイル。