画像認識 Deep Learning ニューラルネットワーク

Technology Reports

Deep Learningによって広がる画像認識アプリケーション

近年，機械学習を用いた画像認識サービスの実用化が加速しているが，従来技術では，画像に写っている物のカテゴリ（“料理”や“花”など）のような，抽象的な概念を画像から認識する事が困難という課題があった．一方，機械学習分野ではDeep Learningの実用化が進んでいる．そこでドコモは，本技術を用いた画像認識システムを開発し，認識機能をAPIとして公開した．本システムでは，学習用画像データを準備し，学習させるだけで，画像へのさまざまなタグ付けが可能な画像認識モデルを精度よく作成する事が可能となる．

サービスイノベーション部 酒井 俊樹（さかい としき） 郭 心語（かく しんご）

1. まえがき

近年，Deep Learningを用いた機械学習*1技術のさまざまな分野への実用化が進んでいる．米Google社，米Facebook社，中国Baidu社などが，2013年ごろから研究所の設立，スタートアップ企業の買収を進めており，例えば米Google社では，2015年3月時点で画像認識*2，音声認識など47のサービスでDeep Learningを活用している[1].

特に画像認識分野においては，Deep Learningを用いる事で大きな精度向上が見込める事が示され[2]，さまざまな画像認識課題で適用が進んでいる（**図1**）.

ドコモではすでに，従来の画像認識技術を用いた画像認識のAPI（Application Programming Interface）*3を提供してきており[3] [4]，それにより画像に写った“商品パッケージ”などの形の決まった物体の認識が可能であった．しかし，この画像認識APIでは，スマートフォンでユーザが撮影したさまざまな画像をタグ付けするための認識ができなかった．そこで，画像のシーン（“結婚式”や“運動会”など）や物体のカテゴリ（“料理”や“花”など）のような抽象的な概念の認識，パンやカレーライスのように形の決まっていない物体の名称の認識，ファッションアイテムの色や柄のような，より人の感覚的な判断に依存する特徴の認識を可能とする画像認識技術をDeep Learningを用いて開発した．この画像認識技術に，独自に収集した大量の画像データを学習データとして用いることで，シーン，ファッション，料理，花などの画像に高精度でタグ付け可能な認識エンジンをそれぞれ開発し，2015年11月よりAPIとして公開した[3].

本稿では，このDeep Learning技術の概要と，従来技術とDeep Learningを用いた画像認識の違い，Deep Learningによって克服された課題とドコモが開発・公開した画像認識APIサービスの特徴とAPIを用いたアプリケーションについて解説する．

2. Deep Learningの概要

Deep Learningは多層ニューラルネットワークを用いた機械学習技術の一種である．ニューラルネットワークは，生物の脳神経での情報処理メカニズムを参考に作られた機械学習手法で，1950年頃から利用され

*1 **機械学習**：人間が，知覚，経験から知識や判断基準，動作などを獲得していくように，コンピュータにデータから知識や判断基準，動作などを獲得させる技術.

*2 **画像認識**：画像処理技術や機械学習技術を用いて，画像を機械に理解させ，意味（画像に写っている物体の名称や，シーンなど，画像から人間が受け取る“意味”）を取り出す技術.

*3 **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.

ている[5]．ベクトルデータや画像などの多次元のデータを複数のクラスに分ける分類問題の解決などに利用されてきた（**図2**(a)）．このニューラルネットワークの層を増やし，より複雑な学習・分類・認識を可能にしたものが，多層ニューラルネットワークであり（図2(b)），1980年代から90年代にかけて流行した．多層

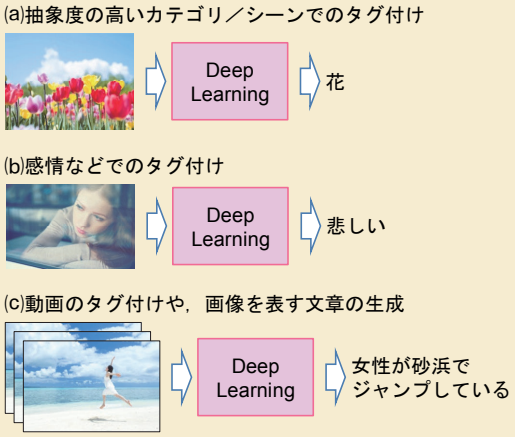

図1　Deep Learningを用いた画像認識

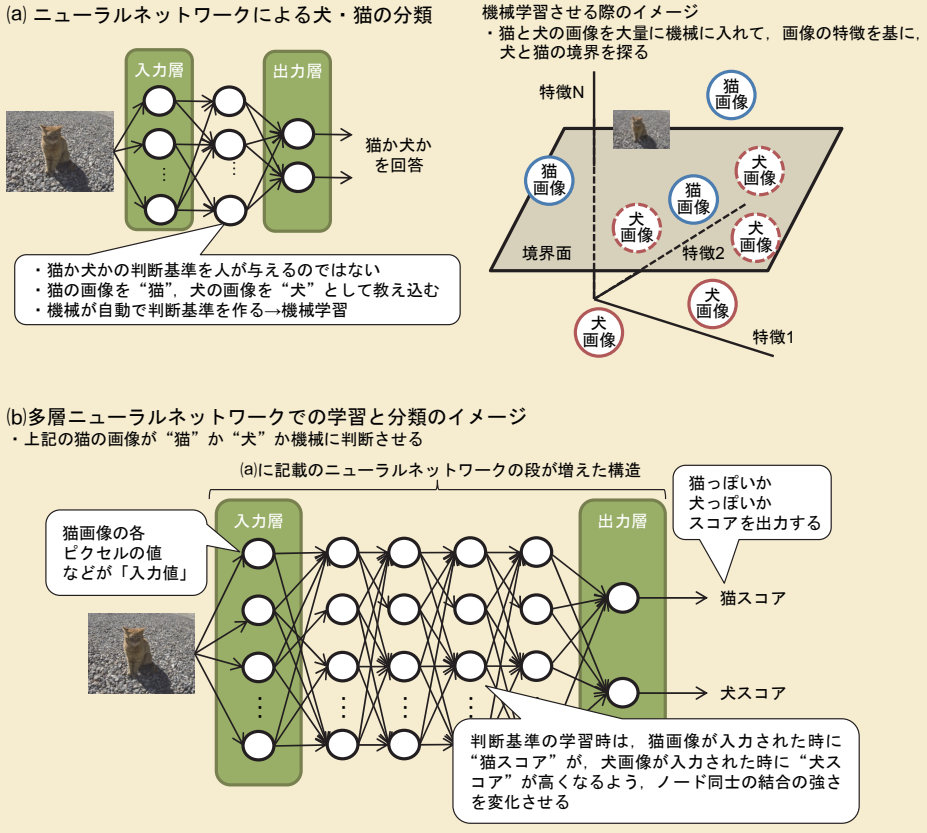

図2　ニューラルネットワークによる機械学習技術を用いた分類・認識のイメージ

ニューラルネットワークは画像認識分野でも古くから利用が試みられてきており，1979年に手書き数字画像の認識で98.6％の認識率を示している[6]．しかし，古典的な多層ニューラルネットワークでは，層の数が増えるほど，学習が困難になる，膨大な時間がかかってしまうといった問題があり，多層化が必要になる複雑な認識課題をニューラルネットワークで解く事は困難で，サービスを提供できるレベルには至らなかった．

この多層ニューラルネットワーク技術は，学習の困難さを解決するためのパラメータの初期化手法の開発や学習の汎用性を増すための手法の導入など技術改良がなされ，またGPGPU（General Purpose computing on Graphics Processing Units）*4を用いた並列分散処理の一般化によって学習速度が飛躍的に向上した．それにより，より深い層を持つネットワークであっても学習が可能となり，2000年代後半からDeep Learningとして再び注目を集める事となった．特に画像認識分野では，2012年に画像認識精度を競う大規模コンペティション（ILSVRC 2012）にて，画像内の被写体を判別する課題（被写体を認識し“ペルシャ猫”などのタグを付ける課題）において，Deep Learningを用いた認識手法が従来の画像認識技術を用いた前年度の手法から約10％もの認識率の改善を示し（2010年から2011年の間では約2％しか精度向上していなかった），画像認識分野での流行のきっかけとなった[2]．

3. 従来技術とDeep Learningを用いた画像認識技術との違い

(1)従来技術

Deep Learning以前の画像認識技術は，主に**図3**(a)に示すような，2段階の構成となっている．まず1段階目では，画像をそのまま利用するのではなく，画像の特徴を数値化した特徴量（例えば，画像内にどの色がどの程度の頻度で出現するかを示したヒストグラムや，画像内の輝度の分布を表したものなど）に変換する．そして，その特徴量を基に分類・認識を行う．特徴量からの分類・認識の判断基準は，機械学習により機械に獲得させる事が一般的である（以下，この判断基準を学習した分類・認識を行う機械を認識器と呼ぶ）．この認識器に画像特徴量を入力すると，認識器が分類結果・認識結果を返却する．

前半の，画像からどのような特徴量を抽出するかに関しては，人手で，認識の課題ごとに設計されており，例えば人の検出に適した画像特徴量，人の顔の認識に適した画像特徴量な

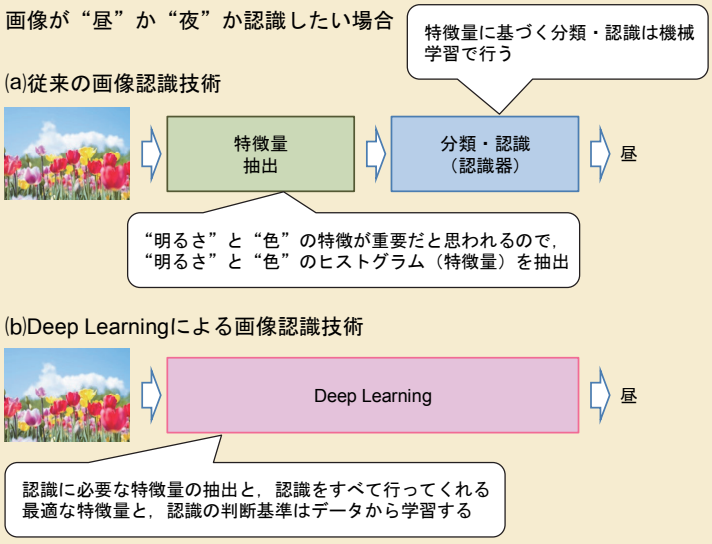

図3　従来技術とDeep Learningによる画像認識技術との違い

*4 **GPGPU**：一般にコンピュータにおける画像の描画などの画像処理に用いられるGPUを画像処理以外の用途に転用する事．並列分散処理に優れる．

どを抽出するアルゴリズムが開発されてきた．

この特徴量を人手で設計する方法では，分類に適切な特徴量を人が考えるため，画像から，シーン（“結婚式のシーン”など）や，画像に写っている物のカテゴリ（“料理”や“花”など）のような抽象的な概念を認識したい場合においては，画像のどのような特徴に着目し，特徴量を算出すればうまく分類できるかわからない，特徴量の算出アルゴリズムが最適化されていないという事態が生じ，認識精度の向上が困難であった．

⑵Deep Learning

Deep Learningを用いた画像認識では，図3⑹に示すようにDeep Learningが学習と認識の両方を行う．認識のために利用する特徴量の最適化と，特徴量による認識基準の作成に相当する作業が学習の過程で自動に行われる．そのため，前述のようにどのような特徴に注目すればよいかわからない抽象的な概念においても認識が可能となる．

一方で，Deep Learningを用いた画像認識では，最後の分類部分だけでなく，特徴量抽出方法もデータから学習をするため，学習用のデータが膨大に必要という欠点がある．近年では，この欠点に対処するため，Deep Learningの認識器をImageNet[7]などの一般的な大規模画像データベースであらかじめ学習させておく事前学習や，人工的に学習データを増やすData Augmentationと呼ばれる手法が一般的に用いられている．

4. 画像認識APIとアプリケーションでの利用

前述のような，適切な特徴量が何か判断しづらい課題においても，あらかじめデータを大量に集めて学習する事で，高精度での認識が期待できるDeep Learningの特性を利用し，シーン認識や，ファッションアイテムの柄や色，アイテムの種類を判別できるファッション認識が可能な画像認識機能を，2015年11月より画像認識API（カテゴリ認識）としてdocomo Developer support[2]で公開した．docomo Developer supportはアプリ／サービスの開発に役立つ機能を提供するサービスで，会員登録および利用申請を行うことで誰でも，Deep Learningを用いた画像認識機能をはじめとするAPIが利用可能となる．

docomo Developer supportで公開した画像認識API（カテゴリ認識）の利用イメージを**図4**に，認識可能な画像の例を**図5**に示す．

シーン，ファッションなどの「認識種別」ごとに学習済みのDeep Learningのネットワークを準備しており，そのAPIを提供している．アプリ／サービスの開発者は，画像を

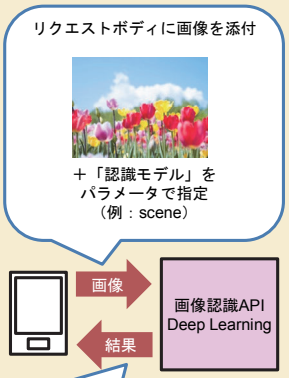

図4　APIをサービス（アプリなど）に組み込んだ際の利用イメージ（例：シーン認識）

図5　APIをサービス（アプリなど）に組み込んだ際の認識種別例

認識したいとき，どのネットワークを使うか選択する．ネットワークを準備するにあたっては，1つのタグ（ここでは，例えば“結婚式”のような，画像認識の結果返却される名称やカテゴリ名をタグと呼ぶ）当り，1,000枚以上の画像をドコモで独自に集め，学習を行った．

docomo Developer supportのユーザは，これら大規模画像データですでに学習済みのモデルを用いて，自身で学習用のデータを集めなくても，すぐにDeep learningによる画像認識機能をアプリなどに組み込む事が可能である．

4.1　シーン認識を活用したアプリケーション

シーン認識では，画像から，その画像に写っているシーン（結婚式や，運動会，誕生日など）と，一部の物体のカテゴリ（花や料理など）を認識できる．

この認識機能のアプリケーションとして，まず想定されるのが，画像のクラウドストレージでの保管アプリ／スマートフォン内の画像管理アプリ／アルバムの自動生成アプリなどである．ユーザが撮影した画像を認識し自動でタグ付けを行う事で，ユーザの画像管理が簡便になる．

また，画像投稿サイトやSNSにおいて本認識機能を用いる事で，ユーザの画像投稿の際のタグ付けの手間を低減する事も可能である．

4.2　ファッション認識を活用したアプリケーション

ファッション認識では前述のDeep Learningを用いた画像認識技術を使い，ユーザによって入力されたファッション画像から，ファッションアイテムのカテゴリを高速で認識でき，画像にタグをつける機能を実現している．

現在は以下の4つのファッション認識モデルを提供している．

①種類：コート，カーディガンなど
②柄：無地，ボーダーなど
③色：ピンク，イエローなど
④スタイル：ビジネス，カジュアルなど

本ファッション認識技術を利用する事で，ユーザからの質問画像を上記のモデルでタグ付けし，タグ情報（種類や柄，色など）を利用して類似なアイテムの画像を探し出す機能（類似検索）を実現できる（図6）．開発者はあらかじめ類似検索結果として表示するためのファッションアイテムの画像群と，各画像に色や柄などのタグ情報を付与しておき，本機能は，それらとユーザが撮影したファッションアイテムの画像のファッション認識結果（色や柄などのタグ）を照合することで似たアイテムを探す．

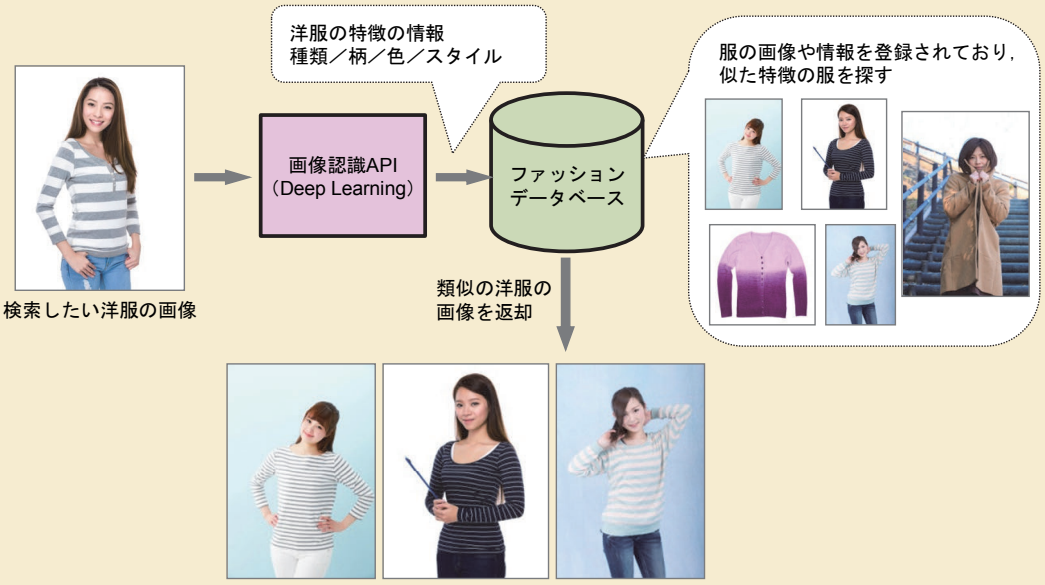

図6　類似画像検索

類似検索を利用すれば，雑誌やカタログなどに掲載された服，スマートフォン向け写真共有アプリで見た服，ドラマで主人公が着ていた服などが気になる時，アイテムの詳細が不明であっても，欲しい服の写真を撮影すれば類似度が高い服の画像やアイテムを検索するサービスが実現できる．購入可能なEC（Electronic Commerce）サイト*5などの情報を添える事で，ユーザ自身でアイテムを探す時間を大幅に節約できる．

また，新たなコーディネートを発見する機会を提供する機能も実現できる．現在手持ちの服の中から合った服を選べない人や，いつも同じような服を買う人，また，買った服の着こなし方で困っていて，持っている服を着るチャンスがなかなかない人を対象者とできるだろう．本機能は，ユーザが撮影した服に類似な洋服とそれに合ったさまざまな全身コーディネートを提示する．これにより無駄な服を買う事が減り，相性のよいアイテムを探す時間も大幅に節約でき，ユーザが自宅や，ショッピングモール，電車などでの服選びも楽しくできるだろう．

5. Deep Learningによる画像認識の適用先の拡大

Deep Learningの画像認識分野での適用先は，今回APIとして公開した単純なタグ付け以外にも広がりを見せている．例えば，画像を見た際に想起される感情（“怒り”，“悲しみ”など）の予測の試みが行われている（図1(b)）[8]．また，動画認識の試みも進んでおり，動画に文章でタグ付けする手法が提案されている（図1(c)）[9]．

6. あとがき

本稿では，Deep Learning技術の概要と，従来の画像認識技術との違い，ドコモが開発・公開した画像認識APIサービスの特徴と，それを用いたサービスについて解説した．

Deep Learningは画像認識以外の適用先でも研究が進んでおり，Deep Learningを用いた自然言語処理や機械翻訳や，マーケティング，Web上でのコンテンツ推薦への活用が検討されている．Deep Learningは今後，あらゆるデータ解析・活用の場面で必要不可欠な技術となっていくと考えられる．

ドコモは今後も，Deep Learningを用いた画像認識のAPIで，認識可能な対象物を順次拡大させる取組みを進めていくと同時に，画像以外のデータに対する認識や，画像と他の

*5 ECサイト：商品やサービスを販売するウェブサイト．

データを組み合わせた新たな認識技術の開発を進めていく.

文　献

[1] J. Dean: “Large Scale Deep Learning.” http://on-demand.gputechconf.com/gtc/2015/presentation/S5817-Keynote-Jeff-Dean.pdf

[2] A. Krizhevsky, I. Sutskever and G. E. Hinton: “ImageNet Classification with Deep Convolutional Neural Networks,” Advances in Neural Information Processing Systems, 25, pp.1097-1105, 2012.

[3] NTTドコモ：“画像認識｜docomo Developer support｜NTTドコモ.” https://dev.smt.docomo.ne.jp/?p=docs.api.page&api_docs_id=102

[4] 赤塚，ほか：“高速大規模画像認識エンジンの開発とAPIの提供,” 本誌, Vol.23, No.1, pp.14-20, Apr. 2015.

[5] F. Rosenblatt: “The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain,” Psychological Review, Vol.65 (6), pp.386-408, Nov.1958.

[6] 福島 邦彦：“位置ずれに影響されないパターン認識機構の神経回路のモデル—ネオコグニトロン—,” 電子通信学会論文誌A, Vol.J62-A, No.10, pp.658-665, Oct.1979.

[7] Stanford Vision Lab, Stanford University, Princeton University: “ImageNet.” http://image-net.org/

[8] K. Peng, T. Chen, A. Sadovnik and A. Gallagher: “A mixed bag of emotions: Model, predict, and transfer emotion distributions,” in Computer Vision and Pattern Recognition (CVPR), 2015 IEEE Conference on, pp.860-868, 2015.

[9] J. Donahue, L. A. Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko and T. Darrell: “Long-term Recurrent Convolutional Networks for Visual Recognition,” CoRR, abs/1411.4389, 2014.