

ディープラーニングで画像の認識・領域分割を高性能化

柳井 研究室



柳井 啓司
Keiji YANAI

コンピュータvs人間。その知性の対決は、科学技術の永遠のテーマと言ってもよいかもしれませ
ん。1997年にチェスの試合で人間がコンピュータに負けた時
は、世界中に衝撃が走りました。
2015年には、画像認識の分野
で象徴的な出来事がありました。
「1000種類の画像認識テスト」
において、人間が初めてコン
ピュータに敗北したのです。

ディープラーニングの登場

そのきっかけは、12年に人工知

能技術の一つであるディープラー
ニング(深層学習)を画像認識に応
用したアプローチが登場したこと
でした。ディープラーニングと
は、コンピュータ上に人間の脳を
模倣した学習型の多層ネットワー
クを構築する機械学習の一手法で
す。膨大なデータを学習させるこ
とで、極めて高い性能を引き出せ
るのが特徴です。

これを境に、画像認識の基本手
法はほぼすべて、ディープラーニ
ングを使った技術に置き換わりま
した。画像認識を研究する柳井研
究室では、これを「画像認識にお
けるパラダイムシフト」ととらえ
ています。コンピュータによる画
像認識性能はその後年々向上し、
15年にはついに95%以上の認識性

能を達成し、人間の能力を超える
に至ったのです。

リアルタイムで画像認識

このように飛躍的に発展する研
究分野において、柳井研究室で
は、画像や映像における一般的
な物体の認識の研究に取り組ん
でいます。ひとくちに画像認識とい
っても、文字認識や顔画像認識など
はすでに実用化されています。一
方で、画像の中から、例えば「空
や「山」、「ライオン」、「イス」など
の一般的な物体について、その位
置や名称を認識することはまだ難
しく、研究が必要な領域です。

柳井研究室では、この一般的な
物体の1000種類の画像認識に
おいて、スマートフォンなどのモ

バイル端末を使って、わずか0.03秒で高速に認識できる技術を開
発しました。ディープラーニング
技術を用いて、スマホ上で膨大な
画像をリアルタイムで認識するシ
ステムは世界初だそうです。
ディープラーニングは通常、膨大
な計算が必要ですが、高速化の工
夫とメモリー使用量の節約によ
り、リアルタイム処理を実現した
のです。

最近では、スマホ上で動画を撮
影しながら、リアルタイムに画像
のスタイル(色彩やタッチ)を変換
できるアプリケーションを開発し
ました。例えば、画像全体や特定
の部分だけを、ピカソやゴッホの
ような画風に即時に変えることが
できます。これは独自の技術で計

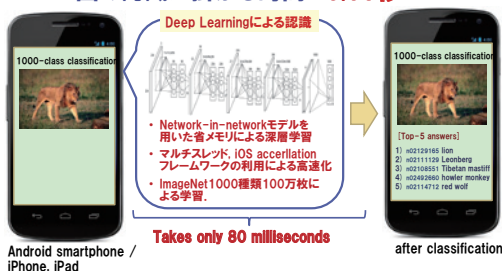
キーワード

画像認識、画像理解、物体認識、機械学習、情報検索

所属	大学院情報理工学研究所 情報学専攻
メンバー	柳井 啓司 教授
所属学会	情報処理学会、電子情報通信学会、 人工知能学会、 米電気電子学会コンピュータ学会 (IEEE Computer Society)、 米コンピュータ学会(ACM)
E-mail	yanai@cs.uec.ac.jp

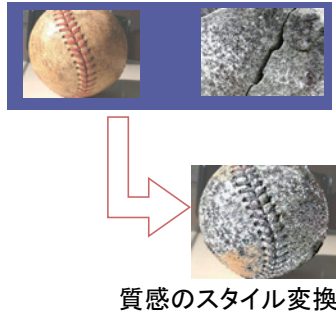
ディープラーニングによる 高速モバイル1000種類物体認識

・1回の認識に掛かる時間: **0.03秒**



算量を減らしたことで、初めて複
数のスタイル変換がリアルタイム
で可能になりました。そのほか、
画像の質感の変換などもできるよ
うになっています

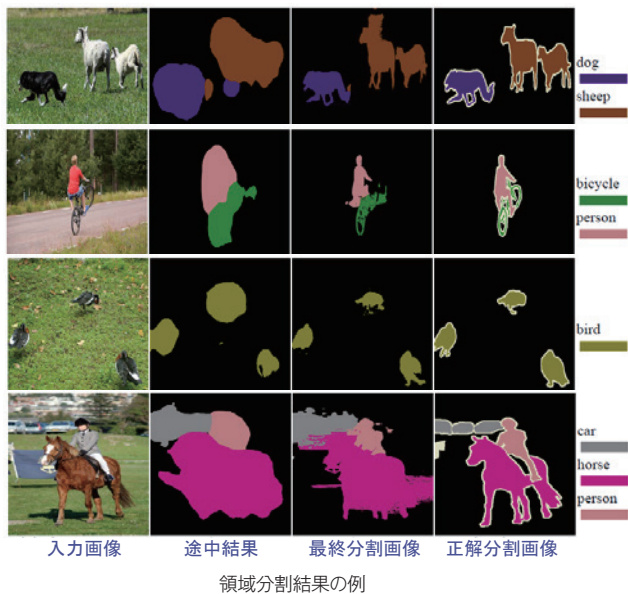
高精度の領域分割も
 ささらに進んで、画像の領域を「分割する」研究にも取り組んでいきます。画像認識でコンピュータが人間を超えたといっても、それはまだ限定的な課題にとどまります。高精度な領域分割は今なお人間だけがなせる技ですが、柳井研究室



質感のスタイル変換
 ボールを石のような質感に変える



動画を撮りながら、同時にスタイル変換ができる



領域分割結果の例



食事認識アプリの表示例

では、この作業をコンピュータに
 行わせる研究に乗り出しました。
 例えば、馬と人が一緒に写って
 いる一枚の画像があったとしてま
 す。そこで、事前に「この画像の
 中には馬と人がいる」という情報
 を与えたとすると、約70%という
 実用上十分な精度で馬と人が存在
 している場所をそれぞれ特定でき
 ました。事前に情報を与えなかつ
 た場合でも、50%程度の精度で特
 定できます。これは世界トップ級
 の性能です。ディープラーニング
 の適用により、精度が約2倍に向
 上したそうです。

**カロリー推定アプリや車載ナ
 ビなどに応用**
 こうした高性能な画像の認識、
 分割技術を使った応用システムも
 開発しています。代表的なもの
 は、企業と共同で開発した「101
 種類の食事認識アプリ」です。食
 堂などで毎日の食事メニューをス
 マホで撮影するだけで、個々の料
 理を認識し、その場で摂取カロ
 リーまで推定してくれる優れモノ
 です。リアルタイムの高速処理な
 がら、93・5%(候補を5個挙げ
 てその中で正解する確率)の精度

を達成しています。最近では、料
 理に使う食材や調味料の情報も同
 時に学習させることで、カロリー
 の推定精度をさらに高めました。
 ほかに、デジタルカメラで撮
 影した画像を全球測位システム
 (GPS)の位置情報によって管
 理し、自動で旅行アルバムなどを
 作成するシステムや、テレビ映像
 を自動でデータベース化し、特定
 のシーンを自動で検出するシステ
 ムなどを手がけています。車載カ
 メラの映像から道路標識を認識す
 ることで、車を運転しながら行き
 先をリアルタイムで確認できるナ

ビゲーションなども可能です。
グローバルに情報を収集
 画像認識には膨大なデータが不
 可欠なため、ツイッターやフェイ
 スブックなどのソーシャルメデ
 アを含めたウェブ上から、大量の
 画像や映像をどのように効率良く
 集めるかといった「マイニング技
 術」も一大テーマです。1000
 種類の画像認識を行うには、1種
 類当たり1000枚の画像の登録
 が必要として、単純に1000枚
 ×1000種類=100万枚の画
 像が必要です。

柳井研究室ではコンピュータに
 よる自動分類に加えて、「最後は
 やはり「人力」が要」として、世
 界中に仕事を発注してより安価な
 報酬で作業を集める「クラウド
 ソーシングサービス」という仕組
 みを積極的に活用しているそう
 です。さすが、世界で日々発信され
 る生きた情報を扱う分野だけあつ
 て、その規模、手法もグローバル
 です。

【取材・文】藤本信穂