

2020.3のブログ：「人工知能の今 応用：画像認識」を読んで、の詳細  
( → <http://www.1968start.com/M/blog/index2.html#2003> )

## 「人工知能の今 応用：画像認識」を読んで

中所武司

### ■この本の読書のきっかけ

人工知能学会誌 Vol.35 No.2 (2020/3)のレクチャーシリーズ：「人工知能の今」の「応用：画像認識」(p.262～p.270)という解説論文の冒頭に、『画像認識は、人間にとっては極めて簡単なタスクだが、計算機で実現するのは極めて困難』という表現があり、興味を持った。

### ■読後感

- ・1970年ごろからの画像認識技術の変遷および最近の深層学習に基づく画像認識技術がわかりやすく解説されている。さらに、その問題点についても適切に指摘されている。
- ・特定の条件下で特定の機能を果たす画像認識技術の実用化は、工学の分野であり、画像の意味内容を認識・理解した人工知能を実現したか否かは重要ではないと思う。

### ■断片的コメント

以下に、興味本位のコメントを述べる。  
→★の部分は、私のコメント

#### 【1. はじめに】

- ・最近、人工知能技術、特に深層学習の進展により、画像認識の性能が顕著に高まっており、自動運転のための視覚の実現、病理画像や内視鏡画像における病変部の自動検出など、実応用への展開が急速に進んでいる。
- ・そもそも画像情報は、計算機で扱うには困難な点が多く、かつ画像認識は、人間にとっては極めて簡単なタスクだが、計算機で実現するのは極めて困難である。

→★このフレーズが目にとまり、本文を読んでみることにした。

#### 【2. 視覚情報処理の難しさ】

- ・技術的難しさ：画像のデジタル化による記憶容量への対応
- ・本質的な難しさ：デジタル化された画像情報とその画像の有する意味情報との乖離  
(セマンティックギャップ)  
(例) 画像の中の犬の認識は、人間には容易で、計算機には難しい。

→★人間は画像をアナログ情報のまま処理している。

たとえば集合写真をみて友人Aを特定できるが、それを言葉では定義できない。  
(面長で坊主頭でメガネをかけている・・・という表現では定義できない)

→★「セマンティックギャップ」という表現は懐かしい。

計算機言語の世界で使っていた。

応用プログラムの機能を自然語で記述すると

人間には理解できるがコンピュータには理解できない。

逆に、プログラミング言語で記述したソースプログラムは

コンピュータには理解できるが人間には理解できない。

自然語とプログラミング言語の間にはセマンティックギャップがある。

→★私の研究（2000年前後）では、アプリケーションの表現レベルの説明に用いた。

ビジネスレベル（ビジネスモデル）とサービスレベル（ドメインモデル）の間に

セマンティックギャップがあり、その下のサービスレベル（ドメインモデル）と

ソフトウェアレベル（コンポーネント）の間を粒度的ギャップと呼んだ。

（参考ページ：研究紹介） <http://www.1968start.com/M/rd/rd/index.html>

（参考ページ：研究論文） <http://www.1968start.com/M/paper/compsac2000.pdf>

### 【3. 人工知能とニューラルネットワーク研究の歴史】

1940年代 ニューロンのモデル化

1956 ダートマス会議 →第一次A I ブーム

1980年代 エキスパートシステム →第二次A I ブーム

1986 誤差逆伝搬法

現在 第三次A I ブーム

### 【4. 画像認識研究の歴史】

#### ●4.1 黎明期

1970年代前半：人の画像認識をプログラム化：認識過程が不明で困難

→★1970年前後は、手書き文字認識がよく取り上げられていた。（例）郵便番号の自動認識

#### ●4.2 人工知能的アプローチ

1970年代後半：画像要素の認識ルールと相互関連ルールを十分な数だけ用意するのは困難

→★1980年代後半の第二次A I ブームの時は、

ルールベースなどの知識ベースシステムやエキスパートシステムが注目されていた。

ルール集合の充実に加えてルールの実行順が技術課題だった。

<拙著の「第9章 人工知能」参照>

「ソフトウェア危機とプログラミングパラダイム」啓学出版（1992発行）

<http://www.1968start.com/M/keigaku/sp09.pdf>

#### ●4.3 画像認識 冬の時代

## ●4.4 機械学習との出会い

1996：大量の顔写真：顔検出に成功：二並列ニューラルネットワーク

1995：黒い背景で、100種類の物体を72方向から撮影：画像認識は比較的簡単

2006：インターネットから集めた雑多な背景の画像の認識は困難

2004：画像中の小領域を表す特徴量（局所特徴量）（複数）に着目 →高い識別性能

## ●4.5 ニューラルネットワークの逆襲

2012：ILSVRC (ImageNet Large Scale Visual Recognition Challenge)

カテゴリー数の多い画像認識データセット ImageNet に関して、

数万のカテゴリーのデータセットのうちの、1000カテゴリーの画像認識のコンペにおいて一つの画像につき五つの答を出し、その中に正解があればよいという審査基準において、誤り率16%という好成績の手法が登場

システム構成は、8階層のニューラルネットワーク

→多数のニューロンを模擬 →多数の積和演算 →GPU使用

### ・深層学習などの機械学習アプローチ

入力データと対応する出力とが学習データとして与えられ、

未知のデータにも適切な出力を推定するように学習する。

ただし、推論プロセスはブラックボックスの end-to-end 学習なので、精度の高い推定が可能であるが、その理由を説明できない問題あり。

→★ルールベースシステムでは、実行されたルールのトレース情報を示すことができるが、多層のニューラルネットのニューロン間の結合係数やニューロン発火のトレース情報から結果の妥当性を検証することは難しい。

## 【5. 深層学習以降の昨今の流れ】

### ●5.1 画像認識の基本タスク

さまざまなタスクが高性能で実現可能：

- ・画像内の事物をカテゴリーに分類する画像識別タスク
- ・画像内のどこに何が写っているか特定する物体検出タスク
- ・物体検出は矩形で切り出すが、さらに物体の領域を正確に切り出す意味セグメンテーション
- ・さらに個別の物体を互いに区別するインスタンスセグメンテーション

### ●5.2 画像検索

2014：深層学習の途中層の出力を画像特徴量とし、それらの間の距離を画像類似度として有効な画像検索を可能とした

2016：高速化、10万画像からの物体検索を1秒以下で実現

### ●5.3 画像・言語連携タスク

- ・画像キャプションニング：入力画像に対して、説明文を出力
- ・画像問合せ：画像と問合せ文を入力し、回答を出力

(例) 家族写真 & 「何人写っていますか」 → 「4人」

→画像解析、自然言語解析、自然言語出力において、深層学習技術を用いて高い精度を実現

### ●5.4 チューリングテストとしての画像認識タスク

- ・画像のカテゴリ分類の精度が人間より優れていても  
人間のように画像の意味内容を認識・理解した、とは言えない
- ・画像の物体検出についても同様
- ・画像キャプションニングについても同様

→★賛成：特定の条件下で特定の機能を果たす画像認識技術の実用化は、工学の分野であり、  
画像の意味内容を認識・理解した人工知能を実現したか否かは重要ではないと思う。

### ●5.5 説明可能な画像認識に向けて

- ・大規模データベースと機械学習による画像認識の実現と引き換えに  
人に説明できる処理手順をあきらめた。
- ・その一方で、自動運転や医用画像に基づく自動診断の領域において、  
システムの動きを人間に説明し、システムの妥当性の判断根拠とする要望がある。
- ・2015：深層学習による高精度の画像認識システムにおいて、  
人間ならば誤らないような誤りをしでかす画像を生成できることが示され、  
システムの過信への警鐘となっている。

→★システムの信頼性については、

人工知能技術適用の有無に関係なく、誤りが無いことは証明できていない。

通常のアプリケーションの信頼性の場合も、開発の最終段階で行ったシステムテストや  
受け入れテストで使用されたテストデータセットの信頼性に依存する。

誤りのないことは保証しないので、不十分なテストデータセットが使用された場合は、  
稼働後に誤った結果を引き起こす可能性が高い。

## 【6. おわりに】

本稿では、画像認識の特性について概観し、画像認識・ニューラルネットワーク・  
人工知能の歴史を振り返り、今の深層学習に基づく画像認識技術について述べた。

以上