

2015年10月7日

人間の「動作」を理解する新しい人工知能(AI) 「時系列 Deep Learning」を開発、8割強の精度で識別に成功

～時系列データの解析によって、映像から人の動作を高精度に識別。
防犯など新たなビジネス領域への応用が可能に～

NTT コミュニケーションズ(略称：NTT Com)は、時系列データの解析が可能な Deep Learning 技術^{*1}を開発し、映像データから人間の動作を高精度に検知することに成功しました。

2015年10月に行った実験において、ネットワークカメラなどで撮影した人間がいる映像に対して、「しゃがんでいる」「きよろきよろしている」「ものを置いている」などの動作を検知させたところ、8割強の高い精度で正答しました。

物体や顔の認識に比べ、高い精度での解析が困難な時系列の映像認識に成功したことで、“人間の動き”を分析することが重要と考えられる防犯分野における活用をはじめ、工場での異常検知、店舗での購買行動分析、スポーツにおけるプレーの分析など、様々な領域への応用可能性が広がりました。

また NTT Com は、今後、パートナーとなる企業とともに商用化に向けた実証実験を検討していきます。

1. 背景

近年、人工知能(AI)^{*2}の重要な要素である Deep Learning 技術が発展した結果、静止画に写っている物体や人間に対する認識精度は飛躍的に向上しました。しかし、映像(特に人間の動作)など、連続した時間的変化を捉えてはじめて意味があるものに対する解析については、現在も高い精度での認識が困難な状況です。

また一方で、監視カメラをはじめとしたネットワークカメラの普及によって、収集可能な映像データが増大しており、さらにこのようなカメラや数多のセンサー機器などがネットワークに接続された IoT^{*3}が進展の兆しを見せています。このため、映像ビッグデータの解析精度が高まれば、新たなビジネス領域を開拓できる可能性があります。

NTT Com はこのような状況を踏まえ、人工知能(AI)による高精度な映像解析を可能にする「時系列 Deep Learning」技術を開発しました。

2. 「時系列 Deep Learning」技術の概要

従来の Deep Learning 技術は主に静止画を対象としており、各画像の縦横 2次元のデータを元に学習を行うことで、画像に含まれる物体を認識していました。本技術では、これに加え学習時および検知時に時間軸方向の情報も盛り込み 3次元とすることで、より精度を高めた動作の特定を可能にしています。また、画像だけでなく、時系列で変化する様々な IoT データ(温度、電圧などのセンサー情報)の高精度な解析が可能であり、汎用性に優れて

います。

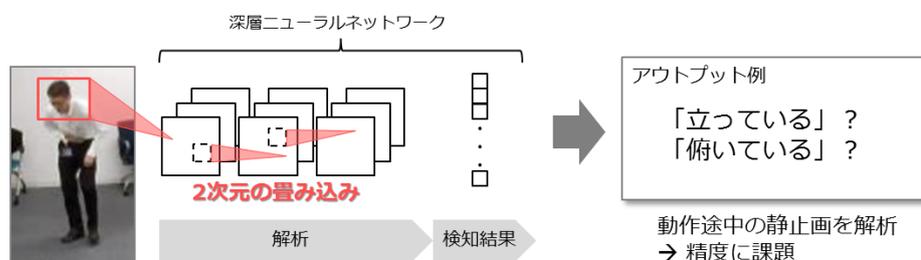
◎主な仕組みと特長

- ・映像をフレーム(コマ)ごとの静止画に分解、1フレームにおける分析対象範囲(フィルタ)内ピクセルのみではなく、近い時間軸フレームの範囲内ピクセルに対しても局所結合(畳み込み結合)^{*4}を行います。

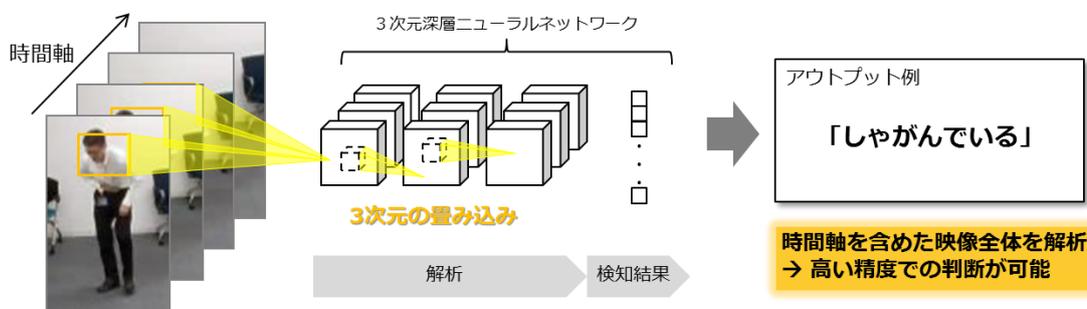
各与えられたデータの各位置/各箇所を学習して特徴を抽出する方法を「畳み込み(Convolution)学習」と言います。本技術は、畳み込み学習の際に、画面内のx軸、y軸だけでなくt軸(時間)についても考慮した3次元でのDeep Learning技術です。

- ・今回数種類の動作を対象とした検証を実施した結果、時系列を考慮しないCNN^{*5}による認識精度が60%弱程度であったのに対し、今回の手法を用いた場合には約85%の認識精度に達することを確認しています。

■従来のDeep Learning技術(2次元)による解析イメージ



■「時系列Deep Learning」技術(3次元)による解析イメージ



3. 今後の展開

NTT Com は、「時系列 Deep Learning」の技術を活用し、防犯やマーケティングなど様々な用途に応じた映像データの解析を可能とする「映像解析プラットフォームサービス」(仮称)の提供を検討しています。店舗や製造工場における不審行動の検出、2020年に向けて観光客の増加が見込まれる多くの施設における防犯対策などに応用することが可能です。

また今後は、映像データだけでなく、多様なIoT端末から収集されたセンサーデータ、端末ログなどを統合的に分析し、より高度な解析を行うことで、ビジネス領域への応用範囲を広げていきます。

なお、「時系列 Deep Learning」の仕組みを活用し、今後社外の様々なサービスとのコラボレーションによって、新しいビジネスを創出することを目指します。

(参考)

「時系列 Deep Learning」は、2015年10月8日～9日に開催する「NTT Communications Forum 2015」にて、展示を行う予定です。

- *1 : 深層学習と呼ばれる多層構造のニューラルネットワーク。従来型の機械学習と異なり、ルールを教え込むことなく、自動的に物事を解釈するために必要となる特徴・要素を抽出することが可能。人工知能(AI)の発展におけるブレークスルーとして注目されている。
- *2 : 人間が脳で行っている知的な作業をコンピュータで模倣したソフトウェアやシステム。
- *3 : Internet of Things の略。コンピュータなどの情報・通信機器だけでなく、世の中に存在する様々なモノに通信機能を持たせ、インターネット/クラウドに接続され、相互に通信・制御する仕組み。
- *4 : 画像内における各位置の特徴を要約するような形で取り出す仕組み。
- *5 : Convolutional Neural Network の略で、畳み込みニューラルネットワーク。