

## AI 特許紹介(4)

～空間変換ネットワーク STN～

2019年7月10日

河野特許事務所

所長 弁理士 河野英仁

「AI 特許紹介」シリーズは、注目すべき AI 特許のポイントを紹介します。熾烈な競争となっている第4次産業革命下では AI 技術がキーとなり、この AI 技術・ソリューションを特許として適切に権利化しておくことが重要であることは言うまでもありません。

AI 技術は Google, Microsoft, Amazon を始めとした IT プラットフォーマ、研究機関及び大学から毎週のように新たな手法が提案されており、また AI 技術を活用した新たなソリューションも次々とリリースされています。

本稿では米国先進 IT 企業を中心に、これらの企業から出願された AI 特許に記載された AI テクノロジー・ソリューションのポイントをわかりやすく解説致します。

### 1.概要

特許権者 Deep Mind Technologies

出願日 2016年6月6日

登録日 2018年7月24日

公開番号 US10,032,089

発明の名称 空間変換モジュール

089 特許は、ニューラルネットワークに空間変換モジュールを組み込んだものである。空間変換モジュールを学習させることで認識対象が傾斜している、一部にしか存在しない場合でも、精度良く分類処理を行うことができる。

### 2.特許内容の説明

画像処理ニューラルネットワークシステム 100 は、ニューラルネットワークコンポーネント 104、空間変換モジュール 108 及びニューラルネットワークコンポーネント 114 により構成される。

ニューラルネットワークコンポーネント 104 は、入力画像に対する入力特徴マップ 106 を出力する。入力特徴マップは、空間変換モジュール 108 に入力される。空間変換

モジュール 108 は、空間変換パラメータに従い変換特徴マップ 112 を生成する。

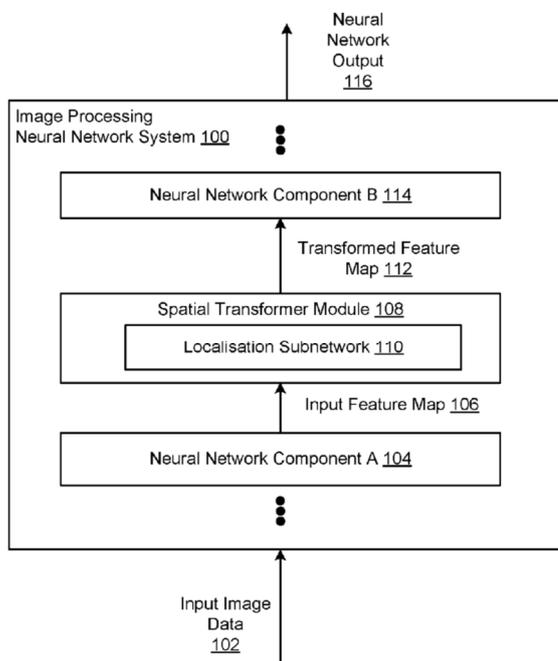
空間変換モジュール 108 は、パラメータ  $\theta$  の下記行列で示される二次元アフィン変換を実行する。

$$\begin{bmatrix} \theta_{11} & \theta_{12} & \theta_{13} \\ \theta_{21} & \theta_{22} & \theta_{23} \end{bmatrix}$$

このパラメータ  $\theta$  に応じて入力特徴マップのクロッピング、移動、回転、スケーリング、スキュー等が行われる。また下記式のパラメータ  $s, t$  等で表される変換も可能である。

$$\begin{bmatrix} s & 0 & t_x \\ 0 & s & t_y \end{bmatrix},$$

その他、平面射影変換、分割アフィン変換、スプライン変換等が行われる。変換特徴マップは最終的にニューラルネットワークコンポーネント 114 を経て変換後の画像データとして出力される。



上述した空間変換パラメータは、トレーニング画像用の画像処理ニューラルネットワークによって生成されたニューラルネットワーク出力と、トレーニング画像用の既知の出力との間の誤差に対し、バックプロパゲーションすることで算出される。

### 3. クレーム

089 特許のシステムクレーム 1 及び方法クレーム 20 は以下のとおりである。

- 1つ以上の入力画像を受信し、該1つ以上の入力画像からニューラルネットワーク出力を生成すべく、該1つ以上の入力画像を処理するように構成された、1以上のコンピュータにより実装される画像処理ニューラルネットワークシステムにおいて、  
以下の処理を実行するよう構成された空間変換モジュールを備え、  
前記1つ以上の入力画像から導出された入力特徴マップを受信し、  
入力特徴マップに空間変換を適用して変換特徴マップを生成し、  
入力特徴マップに基づいて、入力特徴マップに適用される空間変換を定義する空間変換パラメータを生成するために、前記入力特徴マップを処理し、  
変換特徴マップを生成するために、入力特徴マップに基づいて生成された空間変換パラメータに従って、入力特徴マップからサンプリングする。

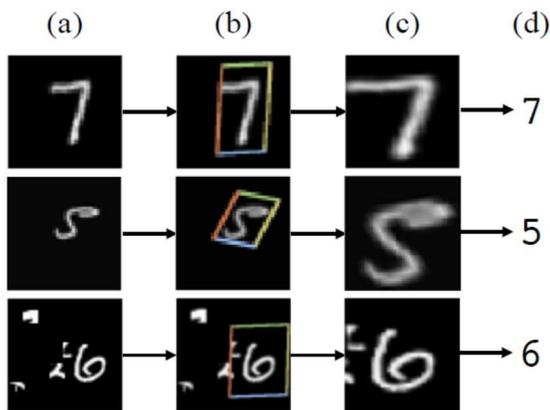
### 20.

- トレーニング画像上で画像処理ニューラルネットワークをトレーニングし、  
画像処理ニューラルネットワークは、1つまたは複数の入力画像を受信し、1つまたは複数の入力画像からニューラルネットワーク出力を生成するために1つまたは複数の入力画像を処理するように構成され、  
画像処理ニューラルネットワークは、空間変換モジュールを含み、  
空間変換モジュールは、以下を含む動作を実行するように構成され、  
1つまたは複数の入力画像から導出された入力特徴マップを受け取り、  
変換特徴マップを生成するために入力特徴マップに空間変換を適用し、  
入力特徴マップに基づいて、入力特徴マップに適用される空間変換を定義する空間変換パラメータを生成すべく、入力特徴マップを処理し、  
変換特徴マップを生成すべく、入力特徴マップに基づいて生成された空間変換パラメータに従って入力特徴マップからサンプリングし、  
画像処理ニューラルネットワークをトレーニングするにあたり、  
空間変換モジュールをトレーニングするために、トレーニング画像用の画像処理ニューラルネットワークによって生成されたニューラルネットワーク出力と、トレーニング画像用の既知の出力との間の誤差をバックプロパゲーションする方法。

### 4. 空間変換ネットワーク STN

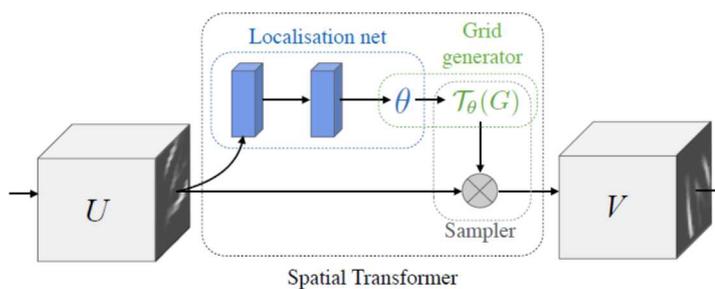
本特許で紹介した空間変換ネットワークは、NIPS(Neural Information Processing Systems)2015にて採択されている技術である。

Jaderberg 氏らの論文<sup>1</sup>によれば、本技術により学習された空間変換ネットワークを用いることで、認識対象の画像に歪みが生じている、背景が写り込んでいる等の場合でも、精度良く認識を行うことができる。



(a)は、ランダム並進、スケール、回転およびクラッタで歪んだ MNIST 画像である。  
 (b)は空間変換ネットワーク内のローカリゼーションネットワークが推定した入力画像である。  
 (c)は空間変換を行った後の空間変換ネットワークからの出力である。

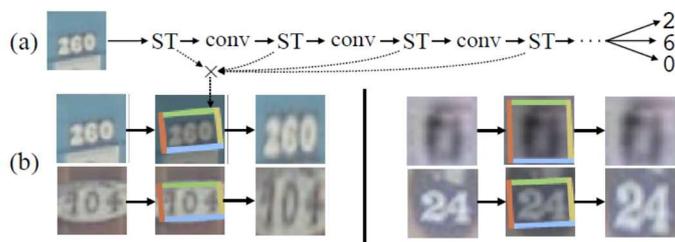
上記写真からわかるように、認識対象の文字が歪んでいる、回転している、他のノイズを含んでいる場合でも空間変換を行い、適切に文字を分類している。



上記図はネットワーク構成であるが、089 特許と同様の構成である。

<sup>1</sup> Max Jaderberg, Karen Simonyan, Andrew Zisserman, Koray Kavukcuoglu  
 「Spatial Transformer Networks」

Model	Size	
	64px	128px
Maxout CNN [10]	4.0	-
CNN (ours)	4.0	5.6
DRAM* [1]	3.9	4.5
ST-CNN	Single	3.7
	Multi	<b>3.9</b>



他の手法との対比実験も行われた。(a)は、ST-CNN マルチモデルの概略図である。各空間変換モジュールの変換は、前の層によって生成された畳み込み特徴マップに適用される。(b)は、ST-CNN Multi の4つの空間変換器によって予測されたアフィン変換の合成の結果を示す。空間変換ネットワークの誤答率が、他の手法よりも低い結果となっている。

## 著者紹介

河野英仁

河野特許事務所、所長弁理士。立命館大学情報システム学博士前期課程修了、米国フランクリンピアースローセンター知的財産権法修士修了、中国清華大学法学院知的財産夏季セミナー修了、MIT(マサチューセッツ工科大学)コンピュータ科学・AI研究所 AIコース修了。

[AI 特許コンサルティング](#)の他、米国・中国特許の権利化・侵害訴訟を専門としている。著書に「世界のソフトウェア特許(共著)」、「FinTech 特許入門」、「[AI/IoT 特許入門](#)」がある。

以上