

ディープラーニングを用いたオブジェクト認識における 整流・マックスプーリング効果

Effect of Rectification and Max-pooling at Deep Learning for Object Recognition

張 傑[†]
Jie Zhang[†]

石井 大祐[†]
Daisuke Ishii[†]

渡辺 裕[‡]
Hiroshi Watanabe[‡]

早稲田大学大学院国際情報通信研究科[†]
Graduate School of Global Information and
Telecommunication Studies, Waseda University[†]

早稲田大学基幹理工学部情報通信学科[‡]
Department of Communications and Computer
Engineering, Waseda University[‡]

概要: 画像解析技術の進歩に伴い、物体検出や認識における強力な手法が次々と提案されてきた。特に、近年ではディープラーニングというアプローチが提案され、更に良い結果が示されている。我々は畳み込みニューラルネットワーク (CNN: Convolutional Neural Networks) と呼ばれる認識手法を用いて、鉄道の乗客認識を行っている。畳み込みニューラルネットワークにおいては、絶対値もしくは二乗による整流及びマックスプーリングが施される。しかし、これらの処理が実際の認識において、どのような影響を与えているか明確ではない。本稿では人物認識のための畳み込みニューラルネットワークにおいて、整流とマックスプーリングの影響を調査する。

キーワード: ディープラーニング, 畳み込みニューラルネットワーク, 人物認識, 整流, マックスプーリング

1. はじめに

近年、ディープラーニングが人工知能、機械学習の分野で注目を集めている。ディープラーニングは、多層ニューラルネットワークの一手法であり、人間の神経回路の様な深い構造を作り出し、最適な特徴量を自主的な学習により獲得することで、精度を大幅に向上させる手法である。

特に、畳み込みニューラルネットワークは2次元の物体認識においてよく用いられる手法となっている[1]。一つの畳み込みニューラルネットワークは多段接続された複数のレイヤーを通じて処理される。畳み込みニューラルネットワークにおいて、絶対値もしくは二乗による整流層及びマックスプーリング層は典型的な二つのオプションレイヤーである。しかし、この二つのレイヤーが実際の認識に対して、どのような影響を与えているか明確ではない。本稿では人物認識のための畳み込みニューラルネットワークにおいて、整流とマックスプーリングの影響を調査する。

2. 整流とマックスプーリング

2.1 整流

整流は、入力 x に対し $R(x) = \max(0, x)$ を取る関数である。近年の研究では、整流層が多層ニューラルネットワークの学習において広く用いられている[2]。我々の実験では、整流層を双曲線関数 \tanh で活性化した畳み込み層の後に配置した。

2.2 マックスプーリング

マックスプーリングは、局所領域内におけるニューロンの反応の最大値を抽出し、要約する操作である。マックスプーリングは段階的に平行移動不変性を与えるため、画像認識の精度へ大きな影響を与える。我々の実験では、マックスプーリング層を、畳み込み層の後に配置する場合と、整流層の後に配置する場合について調査した。

3. 実験

人認識のための畳み込みニューラルネットワークにおいて、整流とマックスプーリングの影響を解明するために以下に示す4種類の組み合わせについてニューラルネットワークを構築し、精度を調査する。1) 特徴抽出部が畳み込み層だけで構成される畳み込みニューラルネットワーク。2) 特徴抽出部が畳み込み層と整流層で構成される畳み込みニューラルネットワーク。3) 特徴抽出部が畳み込み層とマックスプーリング層で構成される畳み込みニューラルネットワーク。4) 特徴抽出部が畳み込み層と整流層とマックスプーリング層で構成される畳み込みニューラルネットワーク。

3.1 実験概要

本実験では、人物認識を対象とする。データセットには INRIA Person Dataset を使用する。学習画像は3500枚、評価サンプルに1000枚を使用する。

3.2 実験条件

実験1～4に対応するパラメタを表1～4に示す。

表1 実験1の条件

特徴量マップ	横*縦*枚数	フィルター	ストライド
入力画像	64*128*1	—	—
C1	60*124*4	5*5	1*1
C2	29*61*8	4*4	2*2
C3	27*59*16	3*3	1*1
C4	13*29*16	3*3	2*2
C5	11*27*16	3*3	1*1
C6	5*13*8	3*3	2*2

表2 実験2の条件

特徴量マップ	横*縦*枚数	フィルター	ストライド
入力画像	64*128*1	—	—
C1	60*124*4	5*5	1*1
R1	60*124*4	—	—
C2	29*61*8	4*4	2*2
R2	29*61*8	—	—
C3	27*59*16	3*3	1*1
R3	27*59*16	—	—
C4	13*29*16	3*3	2*2
R4	13*29*16	—	—
C5	11*27*16	3*3	1*1
R5	11*27*16	—	—
C6	5*13*8	3*3	2*2

表3 実験3の条件

特徴量マップ	横*縦*枚数	フィルター	ストライド
入力画像	64*128*1	—	—
C1	60*124*4	5*5	1*1
M1	30*62*4	2*2	2*2
C2	28*60*8	3*3	1*1
C3	26*58*16	3*3	1*1
M2	13*29*16	2*2	2*2
C4	11*27*16	3*3	1*1
C5	5*13*8	3*3	2*2

表4 実験4の条件

特徴量マップ	横*縦*枚数	フィルター	ストライド
入力画像	64*128*1	—	—
C1	60*124*4	5*5	1*1

特徴量マップ	横*縦*枚数	フィルター	ストライド
R1	60*124*4	—	—
M1	30*62*4	2*2	2*2
C2	28*60*8	3*3	1*1
R2	28*60*8	—	—
C3	26*58*16	3*3	1*1
R3	26*58*16	—	—
M2	13*29*16	2*2	2*2
C4	11*27*16	3*3	1*1
R4	11*27*16	—	—
C5	5*13*8	3*3	2*2

3.3 実験結果

各実験の結果を表5に示す。整流とマックスプーリングを施すことにより、識別率が向上する事が確認された。また、整流による識別率向上は小さく、マックスプーリングが効果的であることが分かる。

表5 識別結果

	True Positive	True Negative	Average
実験1	0.952	0.972	0.966
実験2	0.949	0.976	0.968
実験3	0.962	0.986	0.979
実験4	0.969	0.984	0.98

4. おわりに

本稿では、ディープラーニングを用いたオブジェクト認識における整流とマックスプーリング効果を解明するため、人認識向け畳み込みニューラルネットワークを導入し、4種類の組み合わせによりその精度への影響を調べた。同一データセットで学習を行い、同じ未知サンプルに対する識別実験を行った結果、整流とマックスプーリングを施すことにより、識別率が向上する事が確認された。

謝辞

本研究成果は、独立行政法人情報通信研究機構 (NICT) の委託研究「ソーシャル・ビッグデータ利活用・基盤技術の研究開発」により得られたものである。

参考文献

- (1) A. Krizhevsky.: “Convolutional Deep Belief Networks on CIFAR-10,” (<http://www.cs.utoronto.ca/~kriz/conv-cifar10-aug2010.pdf>), 2010
- (2) Y. Sun, X. Wang, and X. Tang. Deep Convolutional Network Cascade for Facial Point Detection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013