

Feature Extraction by Denoising Autoencoders for Image Classification

長峯 脩兵[†] 湧井 力[†] 日高 章理[†] 狩野 弘之[†]

[†] 東京電機大学大学院 理工学研究科

1. はじめに

2012 年頃よりニューラルネットワークの応用である Deep Neural Network が注目され、画像認識の分野においても近年広く応用されている。

本研究では、Deep Learning の 1 つである Denoising Convolutional Autoencoder によって画像の特徴を抽出し、より識別精度の高いニューラルネットワークの構築を行う。

2. Deep Neural Network

Deep Learning とは、複数の層をもつニューラルネットワークで学習することである。ここでは本研究で用いた画像認識のための 2 つの手法について述べる。

2.1 Convolutional Neural Network (CNN)

CNN は画像認識のためのニューラルネットワークであり、生物の脳の視覚認知処理過程をモデル化した手法である[1][2]。一般に畳み込み層とプーリング層のペアを何層も積み重ねた構造を持つ。一般的なニューラルネットワークは隣接する層のすべてのユニットが全結合しているのに対し、CNN では隣接する層のユニットが特定のユニットとのみが結合している。

2.2 Denoising Convolutional Autoencoder

Autoencoder (AE, 自己符号化器) は一般に入力データよりも低次元の中間層によって次元圧縮を行ってから入力データそのものを復号する教師なしニューラルネットワークである[3]。Convolutional Autoencoder (CAE) は CNN のフィルタを AE と同じく入力データ自身を符号化・復号化するように学習したものである。

Denoising Convolutional Autoencoder (dCAE) は CAE の自己符号化学習の際に、入力画像にノイズを加え、出力(復号)画像がノイズの加えられていない入力画像と同じになるように学習をおこなう。

3. 提案手法

本研究では、CNN で得られた特徴と dCAE で得られた特徴を組み合わせ、ノイズに対してより頑健なネットワークを構築する。2 つのニューラルネットワークの組み合わせ方には様々な方法が検討される。本研究では 2 つのニューラルネットワークで得られた画像フィルタを新たな CNN の初期の重みとして利用する。

4. 実験

本研究では KITTI 画像データセット[4]による画像認識実験を行った。本データセットは乗用車、バス、トラック、歩行者、サイクリスト等の 7 クラスの画像から成る。本研究ではこのうち遮蔽のない車画像 8000 枚を学習とテストに用いる。また車以外の画像として、KITTI の評価用道路画像のうち、車が含まれていないフレームからランダムに切り出した 8000 枚を用いた。上記はすべて 64 × 64 ピクセルにリサイズして用いた。これらの画像のうち、車と非車それぞれ 5000 枚ずつを学習に用い、残りの 3000 枚ずつをテストに用いる。

本実験における CNN は入力層の後に畳み込み層とプーリング層のペアが 3 組続き、その後に 2 クラス識別用の出力層が続く構造となる。各畳み込み層はフィルタ数が 6 個で、それぞれカーネルサイズは 5 × 5 ピクセルである。プーリング層は 1 層目のみ 3 × 3、2 層目以降は 2 × 2 のマックスプーリングを用いた。

上記の CNN に対し、1 層目のフィルタのみを dCAE で事前教師なし学習し、CNN の 1 層目のフィルタの初期値としたものを dCAE-CNN とする。

提案手法では、上記の CNN の各層のフィルタ 6 個と dCAE-CNN の各層のフィルタ 6 個を同じ層で横に並べ、結果として各層が 12 個のフィルタを持つ CNN (CNN+dCAE-CNN と置く) を構築した。

下表に従来の CNN, dCAE による事前教師なし学習を行った CNN および提案手法の識別性能(テスト画像に対する識別率の 3 試行分の平均値)を示す。提案手法により識別性能が向上したことが確認できる。

ネットワーク	平均誤識別率
CNN	5.96%
dCAE-CNN	5.68%
CNN + dCAE-CNN	4.14%

参考文献

- [1] Fukushima, Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position, 1980
- [2] Yann LeCun, convolutional networks for images speech and time-series, 1995.
- [3] 岡谷貴之, 深層学習, pp75-76, 2015
- [4] The KITTI Vision Benchmark Suite. <http://www.cvlibs.net/datasets/kitti/>