

## 深層学習と視覚的特徴の基底抽出

篠崎 隆志

国立研究開発法人 情報通信研究機構  
大阪大学 大学院情報科学研究科  
〒565-0871 大阪府吹田市山田丘1-4  
tshino@nict.go.jp

### 1. はじめに

深層学習の躍進の原動力のひとつに表現学習と呼ばれる特性がある。これは映像であればガボールパッチのような、入力信号を効率よく表現するための基底を、データからの学習によって獲得するものである。本講演では深層学習における画像の基底抽出について、畳み込みニューラルネットワーク (Convolutional Neural Network, 以下CNN) における表現学習を中心に、著者の行ってきた半教師あり学習<sup>1)</sup>を交えつつ紹介する。さらに生体の視覚処理系をヒントに構成した競合学習による視覚的運動の基底抽出についても紹介する。

### 2. 視覚的特徴の基底抽出

CNNとは福島によって提案された生体における視覚情報処理系をヒントに構築されたニューラルネットワーク (以下NN) のモデルである<sup>2)</sup>。その後LeCunらによって一般的なNNの学習法である誤差逆伝播法を適用可能なCNNが提案され<sup>3)</sup>、深層学習の中心的技術として画像認識をはじめとして様々な形態での応用がなされている。CNNの高性能の秘密は、識別のために必要な特徴の基底を学習によって獲得できるという点にある。従来機械学習ではこのような特徴量を人間が設計しており、例えば画像認識であればHistograms of Oriented Gradients (HOG) や Scale-Invariant Feature Transform (SIFT) などが知られている。このよ

うな人間が経験に基づき設計したEngineered Featureに対し、深層学習ではデータの識別のために最適な基底を学習によって自動的に抽出する。図1にImageNetの自然画像で学習されたネットワークの1層目の基底の一覧を示す<sup>4)</sup>。様々な方向、空間周波数のガボールや色パッチが視覚情報を表す基底として効率よく抽出されていることが見て取れる。深層学習の分野では、抽出された基底を学習表現 (Learning Representation), 抽出過程のことを表現学習 (Representation Learning) と呼び、深層学習を進展させるための最重要事項として様々な研究が進められている。

なお余談ではあるが、近年のCNNで用いられているRectified Linear Unit (ReLU) と呼ばれる活性化関数は実質的に側抑制と同様の効果をもたらす。ReLUは値が0以上の場合はその値、0以下の場合は0をとるが、これは実質的に閾値関数として働く。CNNでは実質的な閾値が学習によって高い点に設定されるため、最も適合するものだけが出力し、他は出力しないという、側抑制と同様の動作をする。識別性能を上げるために経験的に導入されたReLUが結果的に生体の視覚処理系と同様の動作を行っているという点で興味深いものである。

### 3. CNNでの半教師あり学習

様々な分野で画期的な成果を上げているCNNではあるが、一方で学習に大量のデータが必要であるという問題が存在する。特にCNNの学習においてはデータに正解が対応付けられたラベル付きデータが必要であり、例え

2017年冬季大会。シンポジウム講演。

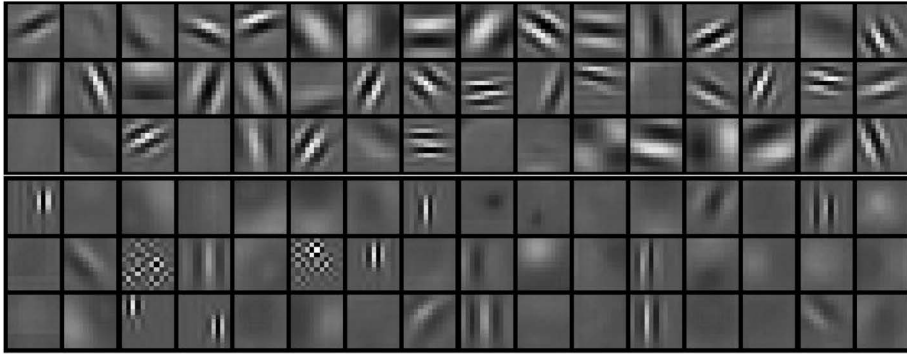


図1 CNNで獲得された画像の基底<sup>4)</sup>。

ばImageNetの自然画像の場合は1000カテゴリで約120万サンプル、ラベルあたり約1000サンプルを必要とする<sup>4)</sup>。このようなデータ量は一般的な利用には非現実的であり、実際には公開されている学習済みのCNNを個々の目的に合わせて学習しなおす転移学習ということが行われている。

このような転移学習が良く機能するという事は、表現学習された特徴の基底は必ずしも最終的な識別内容に縛られず、同じタイプのデータ（例えば画像）である程度の普遍性を持つことを示している。このことは特徴の基底の獲得には必ずしも正解の対応付けは必要でない可能性を示唆している。そこで我々は正解が対応付けられていないラベル無しデータから教師なし学習によって基底を抽出し、ラベル付きデータによる教師あり学習でファインチューニングを行うような半教師あり学習を提案した<sup>1)</sup>。提案した手法は、弱い自己肯定を教師信号に、マイルドな特性の損失関数と更新則を適用することで、単一のCNNにおける教師あり学習と教師なし学習のシームレスな適用を実現する。本手法を利用することによって、教師なし学習によってNN内に獲得された基底が、教師あり学習の効率を数倍に引き上げることが示された。

#### 4. 競合学習による運動の基底抽出

以上のように、CNNでの半教師あり学習によってラベルなしデータによる基底の獲得の効

果を部分的に示したが、誤差逆伝播法で学習するため、入力に近い層で学習の効果が十分でないという問題がある。これはCNN全般に共通した問題であり、ネットワーク構造や初期値、各種学習係数などのメタパラメーターがチューニングされていない場合は、抽出される基底はランダムに近い状態となることが知られている。図2(a)にCNNで獲得されたMNISTの手書き文字画像に対する基底の一覧を示す。図は識別精度が99%を超えている場合のものであるが、抽出されたフィルターはランダムに近い状態であることが見て取れる。

このようなCNNでの誤差逆伝播法に対して、福島によるネオコグニトロンやKohonenの自己組織化マップ<sup>5)</sup>と呼ばれるNNでは、競合学習という、入力されたデータの基底をより強力に抽出する学習法が用いられている。図2(b)は競合学習で獲得された基底の一覧で、CNNの場合と異なり画像の基底となる様々なエッジの断片が明瞭に獲得されている。

競合学習は大脳皮質における側抑制の仕組みをヒントに開発された学習法で、脳の仕組みとの整合性が高い。したがって競合学習に脳の仕組みを組み合わせることによって、従来のCNNでは不可能であった基底の抽出が可能となると考えられる。このようなものの一つに視覚的運動の基底があり、CNNにおいては3次元量み込みなどの手法が提案されているが、パラメータの自由度の高さから、十分な精度の基

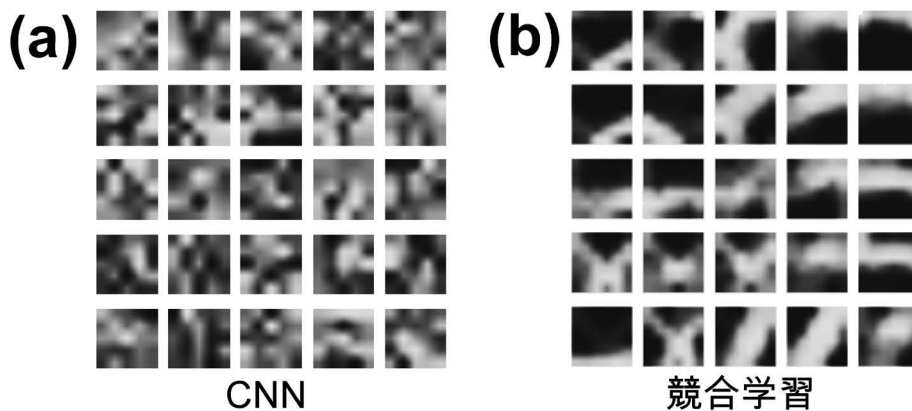


図2 学習によって抽出された基底の比較.

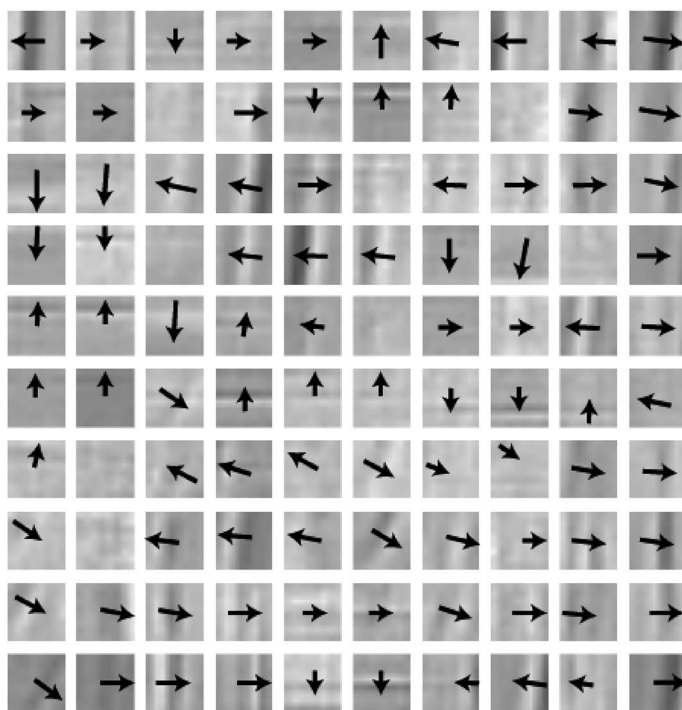


図3 競合学習によって抽出された視覚的運動の基底.

底を抽出するに至っていない。我々は神経生理学の知見<sup>6)</sup>に基づき、Parvo経路が遅延回路としてMagno経路の情報と統合することによって、視覚的運動の基底が抽出できるか検証した。その結果、時間的に変化があるときに反応するMagno経路を生成のトリガーとすることによって基底が効率よく抽出されることを確認した。図3に抽出された視覚的運動の基底を示

す。競合学習によって様々な方向、速度の基底が抽出されていることが見て取れる。

## 5. ま と め

近年の深層学習の躍進の秘密は、データから識別に必要な特徴の基底を抽出する表現学習にある。しかしながら現行の深層学習の基底抽出力はやや弱く、自由度の高い既定の抽出は困難

である。深層学習のさらなる発展のためにも、強力な基底抽出力を持つ競合学習を統合し、様々な脳の仕組みを取り入れていく研究が必要とされている。

## 文 献

- 1) T. Shinozaki: Semi-supervised learning for convolutional neural networks using mild supervisory signals. *Lecture Notes in Computer Science*, **9950**, 381–388, 2016.
- 2) K. Fukushima: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological Cybernetics*, **36**, 193–202, 1980.
- 3) Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard and L. D. Jackel: Backpropagation applied to handwritten zip code recognition. *Neural Computation*, **1**, 541–551, 1989.
- 4) A. Krizhevsky, I. Sutskever and G. E. Hinton: Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, 1097–1105, 2012.
- 5) T. Kohonen: Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, **43**, 59–69, 1982.
- 6) M. T. Schmolesky, Y. Wang, D. P. Hanes, K. G. Thompson, S. Leutgeb, J. D. Schall and A. G. Leventhal: Signal timing across the macaque visual system. *Journal of Neurophysiology*, **79**, 3272–3278, 1998.