

# 畳込みニューラルネットワークのための新しい半教師あり学習法

A novel semi-supervised learning method for convolutional neural networks

篠崎隆志 \*1\*2

Takashi SHINOZAKI

\*1 国立研究開発法人 情報通信研究機構 脳情報通信融合研究センター

CiNet, National Institute of Information and Communications Technology

\*2 大阪大学大学院 情報科学研究科

Graduate School of Information Science and Technology, Osaka University

We propose a novel semi-supervised learning method for convolutional neural networks (CNNs). CNN is the most popular deep learning method, and successes among various types of applications: image and speech recognition, image captioning and the game of go etc. On the other hand, requirements of a huge amount of labeled data for the supervised learning are serious problem. Unsupervised learning which utilizes the information of non-labeled data might be one of the key to address the problem, however, has not been investigated enough in a CNN regime. The proposed method involves both supervised and unsupervised learning in an identical feedforward network, and enables seamless switching among them. We validate the method by a image recognition task, resulting that the learning by non-labeled data dramatically improves the efficiency of supervised learning.

## 1. はじめに

深層学習は脳における情報処理の仕組みを模した超多層ニューラルネットワークによる機械学習の仕組みである。深層学習のためのネットワークは数多く提案されているが、近年では特に畳込みニューラルネットワーク (Convolutional Neural Network: 以下 CNN)[LeCun 89] が画像や音声の認識をはじめとする広範な分野で目覚ましい成果を上げている [Krizhevsky 12, Dahl 12]。

このような巨大なネットワークを学習させるためには極めて大量の学習データが必要とされ、特に教師あり学習に必要なラベルありデータの準備の困難さが深層学習のより広い分野への応用の妨げとなっている。この問題を解決するために、比較的準備しやすいラベルなしデータで学習可能な教師なし学習の研究が必要とされている。

深層学習における教師なし学習の先行研究としては、古くは Restricted Boltzmann Machine (RBM)[Hinton 06] や、Google Brain [Le 12] などが報告されているが、ネットワークの動作原理が根本的に異なるために CNN には適用できなかった。Radford らは画像生成に深層学習の教師なし学習を適用としたモデルを提案しているが識別課題への応用は困難であった [Radford 16]。これに対して Goroshin らは層数の少ない CNN を用いた教師なし学習を提案しているが [Goroshin 15]、学習が動画の時間連続性に強く依存しているために静止画像や医療データ等、他の一般的なデータへの適用が困難であった。またこの教師なし学習は最終的に課題を適用し教師あり学習を行うネットワークと異なる浅い簡易型の CNN によって行うため、学習結果を識別用の深いネットワークに転移学習して使用する必要があった。

これらの問題を解決するために、教師なしと教師ありの2つの学習を同一のネットワークで適用し、シームレスな学習を実現するような半教師あり学習の手法が必要とされている。本研究では現在の深層学習の主流である CNN に適用可能な半教師あり学習の手法を提案し、画像認識課題による検証を行った。

## 2. 半教師あり学習法

提案する半教師あり学習法はラベルなしデータによる教師なし学習とラベルありデータによる教師あり学習とを同一の CNN に適用することによって半教師あり学習を実現する。

教師なし学習においては、ラベルによる同一性の情報が入力されないため、同一の意味を持つ学習表現が多数の部分学習表現に分散して表現されることが想定される。従って本手法を適用する CNN のネットワーク構造は、基本的には通常の識別課題に用いるものと同様であるが、分散した学習表現に対応するために、特に出力層においてユニット数に十分な余裕を持たせる必要がある。そして教師あり学習におけるラベルからの情報によって、この分散した部分学習表現を束ね直すことにより効率的な学習を実現する。

学習は教師なし、教師ありともに通常の CNN と同様の誤差逆伝播法によって行う。教師なし学習においては入力に対して最大値をとった出力層のユニットをラベルとして、教師あり学習においては与えられたラベルを用いて学習する。

教師なし学習においては正解の情報は確度の低いものとして取り扱う。通常の学習では最終層に提示される正解を表す教師ベクトルには、正解のみ 1 でそれ以外はすべて 0 を持つものが用いられるが、本学習においては正解以外にバイアスを与えた上で規格化したものを用いる。例えば出力層のユニット数が 100、バイアスが 0.8 の場合、正解のみ 0.0125、それ以外は 0.0100 の値を持つ。さらに損失関数にも識別特性がマイルドな平均 2 乗誤差 (Mean Squared Error (MSE)) を用い、最適化についてもマイルドな特性をもつ通常の Stochastic Gradient Descent (SGD)[Bottou 98] を用いる。

これに対して教師あり学習においては学習はラベルからの確度の高い情報によって行われることから、教師ベクトルには通常の正解のみ 1 でそれ以外はすべて 0 を持つものを用い、損失関数には識別特性がシャープな平均交差エントロピー (Averaged Cross Entropy (ACE)) を、最適化についてはシャープな特性を持つ Adam [Kingma 15] を用いる。

新規に学習する場合はまず教師なし学習を行い、ラベルに

連絡先: 篠崎隆志, 〒 565-0871 大阪府吹田市山田丘 1-4

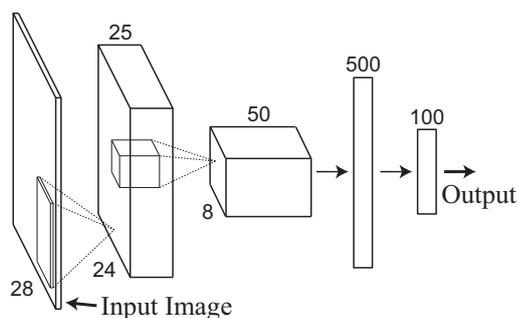


図 1: 検証実験に用いた CNN の構造の概要。2 セットの畳込み層と 2 層の全結合層から構成される。

よらないデータの基本的な構造 (例えば画像であれば 2 次元空間での連続性, 音声であれば倍音構造など) をネットワークに学習する。そして教師あり学習では獲得した学習表現をラベルありのデータで結びつけることによって迅速な学習を行う。

新たにラベルのない入力を学習する場合はまた教師なし学習を行い, ラベルがあるデータに対してのみ教師あり学習を行うことによって持続的な半教師あり学習を実現する。

### 3. 数値モデルによる検証

提案する学習法の効果を確認するために MNIST の手書き文字セット [LeCun 98] を用いた画像識別課題による検証を行った。ネットワークには通常の CNN を用い, 28x28 ピクセルの画像を入力とし, 2 セットの畳込み層と, さらに 2 層の全結合層からなる (図 1)。畳込み層のユニット数は 24x24 と 8x8, 受容野はそれぞれ 5x5 と 2x2 で, カーネル数は 25 と 50, max-pooling はともに 2x2 とした。全結合層のユニットはそれぞれ 500 と 100 とし, 特に出力層は識別するラベル数である 10 に対して十分な余裕をもたせた。

ネットワークは乱数で初期化され, 半教師あり学習を行ったものと通常の教師あり学習のみを行ったものについて教師あり学習における学習速度の比較を行った。学習は 60,000 サンプルの訓練用セットによって行い, 10,000 サンプルのテスト用セットによってエラー率を確認した。教師なし学習は訓練用サンプルを 2 回適用して合計で 120,000 試行とし, 教師ベクトルのバイアスは 0.8 とした。

図 2 に教師あり学習における識別のエラー率の変化を示す。エラー率が 1 割を切るために必要なサンプル数は教師なし学習をしない場合が約 10,000 回であるのに対して, 本手法を適用し場合は約 4,500 回となっており, 学習速度が 2 倍以上に高速化された。さらに中程度の学習状態 (エラー率 0.3-0.6) におけるエラー率の分散が小さくなっており, 半教師あり学習によって学習が安定していることが見て取れる。

### 4. まとめ

本研究では識別課題用の CNN において, 教師なし学習と教師あり学習とをシームレスに適用可能な半教師あり学習の手法を提案した。提案した学習法は大量のラベルなしデータを用いた教師なし学習によって, データの持つ特徴の学習表現を獲得し, 少量のラベルありデータを用いた教師あり学習における効率を大幅に向上させた。

本手法では教師なし学習を適用するにあたって, ネットワークのユニット数を特に出力層で余裕を持たせたものとしているが, これはネットワークパラメータとしての余裕を意味し, 結果としてパラメータチューニングの負担を軽減する。

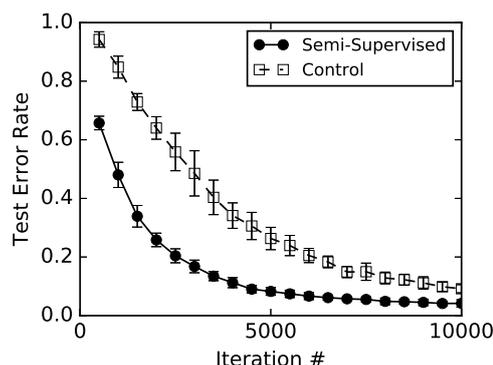


図 2: 教師あり学習によるエラー率の変化。黒丸に実線が半教師あり学習の場合。エラーバーは標準偏差 ( $n=10$ )。

本手法は一般的な CNN 全般に適用可能であることから, 画像認識や音声認識にとどまらず, IoT をはじめとして, ラベルありデータの確保が困難な医療データなど多方面での応用が期待される。

### 参考文献

- [Bottou 98] Bottou, L.: Online algorithms and stochastic approximations. In Online Learning and Neural Networks *Cambridge University Press*, (1998)
- [Goroshin 15] Goroshin, R., Bruna, J., Tompson, J., Eigen, D. & LeCun, Y.: Unsupervised Learning of Spatiotemporally Coherent Metrics, *arXiv:1412.6056* (2015)
- [Radford 16] Radford, A. & Metz, L.: Unsupervised representation learning with deep convolutional generative adversarial networks, *IPLR 2016*, (2016)
- [Kingma 15] Kingma, D.P. & Ba, J.L.: Adam: a Method for stochastic optimization *Proc. 4th Int. Conf. on Learning Representations* (2015)
- [Le 12] Le, Q.V., Ranzato, M.A., Monga, R., Devin, M., Chen, K., Corrado, G.S., Dean, J. & Ng, A. Y.: Building high-level features using large scale unsupervised learning, *Proc. 29th Int. Conf. on Machine Learning* (2012)
- [Dahl 12] Dahl, G.E., Yu, D., Deng, L. & Acero, A.: Context-dependent pre-trained deep neural networks for large vocabulary speech recognition, *IEEE Trans. on Audio, Speech, and Lang. Process.*, **20**(1), pp. 30-42 (2012)
- [LeCun 89] LeCun, Y., Boser, B., Denker, J.S., Henderson, D., Howard R.E. & Jackel, L.D.: Backpropagation applied to hand-written zip code recognition, *Neural Computation*, **1**(4), pp. 541-551 (1989)
- [Krizhevsky 12] Krizhevsky, A., Sutskever, I & Hinton, G.E.: ImageNet classification with deep convolutional neural networks, *Advances in Neural Information Processing Systems 25*, pp. 1106-1114 (2012)
- [Hinton 06] Hinton, G.E., & Salakhutdinov, R.: Reducing the dimensionality of data with neural networks, *Science*, **313**, pp. 504-507 (2006)
- [LeCun 98] LeCun, Y., Cortes, C. & Barges, C.J.C.: *The MNIST database of handwritten digits*. (1998)