

科学研究費助成事業 研究成果報告書

平成 28 年 5 月 27 日現在

機関番号：12601

研究種目：若手研究(B)

研究期間：2014～2015

課題番号：26730085

研究課題名(和文) 畳み込み層の識別的初期化に基づく少サンプル深層学習法の構築

研究課題名(英文) Small-sample Deep Learning Method Based on Analytic Initialization of Convolutional Layers

研究代表者

中山 英樹 (Nakayama, Hideki)

東京大学・情報理工学(系)研究科・講師

研究者番号：00643305

交付決定額(研究期間全体)：(直接経費) 2,900,000円

研究成果の概要(和文)：本研究では、少数の教師付学習サンプルから高速かつ安定な学習を可能とする新しい深層学習の枠組を提案した。まず、フィッシャー重みマップ法による識別的な解析解により畳み込み層を算出し、これを順次積み上げることで多層畳み込みニューラルネットワーク(CNN)をボトムアップに構築する。さらに、これを初期状態として誤差逆伝播法による学習を進めることで、より優れた局所解に高速に到達させることを狙う。提案手法は、MNIST・STL-10などの本分野におけるいくつかの代表的なベンチマークにおいてstate-of-the-artの識別精度を達成した。

研究成果の概要(英文)：We proposed a new deep learning methodology that enable fast and stable learning from limited amount training examples. We stacked convolutional layers using discriminative analytic solutions obtained by Fisher weight map to build multi-layer convolutional neural networks. Then we further fine tune the entire network by means of the standard backpropagation to quickly reach better local minima. Proposed method achieved state-of-the-art classification accuracy on some standard benchmarks, namely MNIST and STL-10.

研究分野：知能情報処理

キーワード：深層学習 ディープラーニング 画像認識 畳み込みニューラルネットワーク 表現学習 人工知能

1. 研究開始当初の背景

計算機の著しい進歩やデータ量の指数的な増加を背景に、多層ニューラルネットワークを用いた深層学習(deep learning) が注目を集めている(引用文献)。特に画像認識の研究分野においては、畳み込みニューラルネットワーク(convolutional neural networks, CNN) が多くのベンチマークやコンペティションにおいて大差で従来の手法を上回っている(引用文献)。

しかしながら、CNN は極めて大きな自由度を持つモデルであるため、過学習を抑え十分な認識精度を得るためには非常に多くの教師付学習サンプルを要するという欠点がある。実応用を考えた場合、一般に教師付サンプルの作成は常に高コストであり数は限られているため、この問題が最大のボトルネックとなる。教師付サンプルが少ない場合は、これと別に用意する多数の教師なしサンプルを用いたネットワークの初期化(教師なし事前学習)を行うアプローチが現在の主流であるが、計算量が膨大であると同時に、職人芸的なハイパーパラメータの調整が必要であることが指摘されている(引用文献)。これらの理由により、深層学習は誰もが容易に使える枠組としては未だ確立していないのが現状と言える。

2. 研究の目的

本研究では、少数の教師付学習サンプルから高速かつ安定な学習を可能とする新しい深層学習の枠組を提案する。まず、フィッシャー重みマップ法による識別的な解析解により畳み込み層を算出し、これを順次積み上げることで多層畳み込みニューラルネットワーク(CNN)をボトムアップに構築する。さらに、これを初期状態として誤差逆伝播法による学習を進めることで、より優れた局所解に高速に到達させることを狙う。

3. 研究の方法

研究は基本的に当初計画通りに進行でき、主に以下の項目を順に実施した。

(1) 本研究開始以前に提案した単層畳み込み手法(引用文献)を詳細に分析し、他の多変量解析手法と体系的に比較を行った。また、これを多段階で用いることで多層の畳み込みネットワークを識別的かつ高速に初期化するアルゴリズムを導出した。さらに、複数のネットワークをアンサンブルする方法についても詳細に検討を行った。

(2) (1)で得たネットワークを初期状態とし、誤差逆伝播手法を適用することでネットワーク構造の全体最適化を行った。また、最適化における最新の知見を実装し、提案手法における効果について検証した。

(3) オープンソースフレームワークを用い提案手法のGPU実装を行った。



図 1: STL-10, CIFAR-10/100, MNIST の各データセットの画像例。

表 1: 提案手法(Fine-tuning 未実施時)と先行研究との識別精度比較(%)。

	提案手法	先行研究
STL-10	66.9	62.3 (引用文献) 64.5 (引用文献)
CIFAR-10	81.9	88.3 (引用文献) 89.6 (引用文献)
CIFAR-100	62.05	63.15 (引用文献) 64.32 (引用文献)
MNIST	99.60	99.53 (引用文献) 99.55 (引用文献)

(4) 画像クラス分類やセマンティックセグメンテーションのベンチマークを数種類用い、提案手法の評価実験を行った。

4. 研究成果

(1) ネットワーク初期化の効果

まず、3(1)で開発した初期化手法により得られる多層CNNの初期状態を評価するために、最後の全結合層のみ学習させた場合の性能を評価する。すなわち、ネットワークは特徴抽出器として機能することになり、一般的なニューラルネットワークのようなend-to-endの学習は未実施の状態である。ここでは、STL-10(引用文献)、CIFAR-10/100(引用文献)、MNIST(引用文献)データセットを用い評価を行った(図1)。いずれも、本分野において標準的に用いられるベンチマークである。実装したネットワークのパラメータ等の詳細は、学会発表を参照されたい。

表1に、各データセットにおける先行研究と提案手法の識別精度の比較を示す。ここでは、先行研究のスコアのうち、データセット拡張を行っていない場合で最もよいものをまとめた。提案手法は、STL-10, MNISTにおいて先行研究を上回る識別精度を達成した(公表時)。また、CIFAR-100においても、最新手法に匹敵する良好な精度を達成した。

CIFAR-10, CIFAR-100は似た性質を持つデータセットであるが、提案手法は相対的に見てCIFAR-100においてより良好な結果を示していると言える。例えば、提案手法はCIFAR-100においてはMaxoutを上回った

が, CIFAR-10 では及ばない結果となっている。この理由として, CIFAR-10 では一クラスあたりの学習サンプル数が 5000 個であるのに対し, CIFAR-100 では 500 個と少ないため, 少サンプルからも安定に学習可能である提案手法の性質が奏功したものと思われる。これは, STL-10 において提案手法が非常に良好な性能を示していることから裏付けられる。また, 別の要因として, FWM は判別分析と同様にクラス数が増えるほど固有値問題のランクが増し多くの特徴がとれるようになるため, よりクラス数の多い CIFAR-100 において有利に働いた可能性がある。今後, よりクラス数の多い大規模なデータセットを用い, この点の検証を進めたい。

いずれにしても, ここでの提案手法は, 一切全体の学習を行わない完全にボトムアップなアプローチをとっているにも関わらず, 最新の深層学習に匹敵する識別精度を得たことは興味深い結果であると言える。

(2) End-to-end 学習時の効果

次に, 初期化されたネットワークに対して誤差逆伝播法により fine-tuning を行った場合の最終的な識別精度および学習速度について比較評価を行った。ここでは, 細胞画像識別タスクのデータセットを用いて検証した。画像パッチを入力とし, 中心のピクセルが細胞, 背景, 境界のいずれであるかを判定する 3 クラス識別タスクとなっている。実験セットアップの詳細については学会発表を参照されたい。

図 2 に, 畳み込み層を FWM, CWM(後述), ランダム値で初期化した場合(Rand)のそれぞれの訓練誤差・テスト誤差の推移を示す。FWM が提案手法, Rand が一般的な CNN に相当する結果, 一般的なランダム初期化による CNN と比較し, 約半分程度の反復回数で良好な識別精度を達成できることが示された。また, 最終的な識別精度もランダム初期化を上回る結果となり, おおむね本研究計画における目的通りの効果を得た。一方で, 他の初期化手法と比較した場合の少サンプル時における精度・速度面での優位性に関しては, より多くのデータセットを用いた慎重かつ網羅的な検証が必要である。

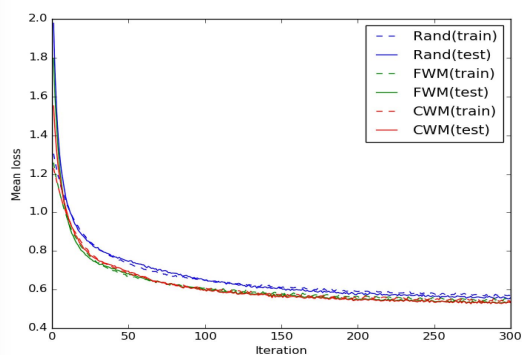


図 2: 細胞画像識別データセットにおける, 学習の進行に伴う訓練・テスト誤差の推移。

(3) その他の成果

当初計画には直接含まれないが, 研究の進展に伴い以下のような派生的な成果を得た。まず, FWM を用いた畳み込みを部位特徴の抽出に応用することで, 詳細画像識別タスクにおいて良好な識別精度を得た。本成果は国際論文誌に採択済みである(発表論文)。さらに線形判別分析を規範とする FWM に対し, 正準相関分析を規範とした正準重みマップ法(CWM)を提案し, 画像のセマンティックセグメンテーションタスクにおいてより良好な結果を得た(学会発表)。

<引用文献>

- G. E. Hinton and R. R. Salakhutdinov, Reducing the dimensionality of data with neural networks, *Science*, 313:504–507, 2006.
- A. Krizhevsky, I. Sutskever, and G. E. Hinton, ImageNet classification with deep convolutional neural networks, In Proc. NIPS, 2012.
- Y. Bengio. Practical recommendations for gradient-based training of deep architectures, *Neural Networks: Tricks of the Trade*, 2012.
- H. Nakayama, Efficient discriminative convolution using Fisher weight map, In Proc. BMVC, 2013.
- A. Coates, H. Lee, and A. Y. Ng, An analysis of single layer networks in unsupervised feature learning, In Proc. AISTATS, 2011.
- A. Krizhevsky, Learning multiple layers of features from tiny images. Master's thesis, Toronto University, 2009.
- Y. LeCun, The MNIST database of handwritten digits. <http://yann.lecun.com/exdb/mnist/>.
- R. Gens and P. Domingos, Discriminative learning of sumproduct networks, In Proc. NIPS, 2012.
- L. Bo, X. Ren, and D. Fox, Unsupervised feature learning for RGB-D based object recognition, In Proc. ISER, 2012.
- I. J. Goodfellow, D. Warde-farley, M. Mirza, A. Courville, and Y. Bengio, Maxout networks, In Proc. ICML, 2013.
- M. Lin, Q. Chen, and S. Yan, Network in network, In Proc. ICLR, 2014.
- N. Srivastava and R. Salakhutdinov, Discriminative transfer learning with tree-based priors, In Proc. NIPS, 2013.

5. 主な発表論文等

(研究代表者, 研究分担者及び連携研究者には下線)

〔査読付き雑誌・国際会議論文〕(計3件)

Hideki Nakayama and Tomoya Tsuda, Efficient two-step middle-level part feature extraction for fine-grained visual categorization, IEICE Transactions on Information and Systems, Vol.E99-D, 2016. (in press)

DOI: 10.1587/transinf.2015EDP7358

Noriki Nishida and Hideki Nakayama, Multimodal gesture recognition using multi-stream recurrent neural network, Image and Video Technology, Lecture Notes in Computer Science (LNCS), Vol.9431, 2015, pp682-694.

Ruka Funaki and Hideki Nakayama, Image-mediated learning for zero-shot cross-lingual document retrieval, Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP), 2015, pp.585-590.

〔学会発表〕(計8件)

福田竣, 指田岳彦, 中山英樹, 正準重みマップを用いた CNN による画像領域分割, 第22回画像センシングシンポジウム(SSII), 2016年6月10日(予定), パシフィコ横浜(神奈川県横浜市).

舟木類佳, 中山英樹, Zero-Shot 言語横断文書検索のための画像媒介学習, 情報処理学会第78回全国大会, 2016年3月10日, 慶應義塾大学矢上キャンパス(神奈川県横浜市).

岩瀬智亮, 中山英樹, 深層一般化正準相関分析, 情報処理学会第78回全国大会, 2016年3月12日(慶應義塾大学矢上キャンパス).

中山英樹, 深層畳み込みニューラルネットによる画像特徴抽出と転移学習と応用研究, 情報処理学会音声言語情報処理研究会, 2015年7月17日, 諏訪湖ホテル(長野県諏訪市).(招待講演)

Hideki Nakayama, Stacked local autocorrelation features, Meeting on Image Recognition and Understanding (MIRU), 2014年7月29日, 岡山コンベンションセンター(岡山県岡山市)(査読有)

中山英樹, 階層フィッシャー重みマップを用いた識別的初期化による深層畳み込みニューラルネットワーク構築法, 第21回画像センシングシンポジウム(SSII), 2014年6月11日. パシフィコ横浜(神奈川県横浜市).

岩本恵太, 中山英樹, Extreme Learning Machine による特徴抽出を用いた決定木

とセマンティックセグメンテーションへの応用, 第21回画像センシングシンポジウム(SSII), 2014年6月11日, パシフィコ横浜(神奈川県横浜市).

中山英樹, Deep Learning による画像認識革命 - 歴史・最新理論から実践応用まで - 第21回画像センシングシンポジウム(SSII) チュートリアルセッション, 2014年6月10日, パシフィコ横浜(神奈川県横浜市).(招待講演)

〔産業財産権〕

出願状況(計1件)

名称: 画像を媒介した異言語文書間の学習法及び装置, 言語横断文書検索方法及び装置

発明者: 中山英樹, 舟木類佳

権利者: 東京大学

種類: 特許

番号: 特願 2015-220107

出願年月日: 平成 27 年 11 月 10 日

国内外の別: 国内

〔その他〕

ホームページ等

<http://www.nlab.ci.i.u-tokyo.ac.jp/projects/deeplearn.html>

6. 研究組織

(1) 研究代表者

中山 英樹 (NAKAYAMA, Hideki)

東京大学・大学院情報理工学系研究科・講師

研究者番号: 00643305