

令和 5 年 6 月 12 日現在

機関番号：82626

研究種目：基盤研究(C)（一般）

研究期間：2020～2022

課題番号：20K11900

研究課題名（和文）適応的事前確率モデルに基づく特徴表現学習に関する研究

研究課題名（英文）A study on feature representation learning based on adaptive prior models

研究代表者

小林 匠（Kobayashi, Takumi）

国立研究開発法人産業技術総合研究所・情報・人間工学領域・主任研究員

研究者番号：30443188

交付決定額（研究期間全体）：（直接経費） 3,300,000円

研究成果の概要（和文）：データの自動認識に有効なニューラルネットワーク（NN）を用いた深層学習において、その性能を向上させる新たな学習の枠組みを提案した。従来はNNパラメータを学習により直接的に最適化していたが、本提案法ではパラメータに事前モデルを新たに導入し、階層的なパラメータ表現を定式化した。これによりNNパラメータは事前モデルから導出される形式となり、学習は事前モデルのハイパーパラメータ最適化に置き換えられる。このような階層的表現により、過学習等を抑制した頑健かつ適応的な深層学習が可能となった。NNを構成する様々な計算過程（層）に事前モデルを導入する手法を開発することで、自動認識の高精度化を達成した。

研究成果の学術的意義や社会的意義

画像や音声など膨大なパターンデータが蓄積される現在、それらを有効に活用するために、AIによる自動認識技術への社会的ニーズが高まっている。本研究成果は自動認識・解析で特に効果の高いニューラルネットワークを用いた深層学習の性能改善に資するものである。ニューラルネットワークのパラメータ表現に着目することで、既存の様々なモデルへと適用できる汎用性をもちつつ、実世界の多様な訓練データセットでの学習を可能とする頑健性を与えることができるため、実用面での波及効果が期待できる。さらに、事前モデルを導入したパラメータの階層的表現は学術的意義も大きい数理的アプローチとなっている。

研究成果の概要（英文）：In this study, we have proposed effective approaches to train neural networks in a framework of deep learning. Neural networks are composed of plenty of parameters which are directly optimized through the learning process. We introduce prior models for those parameters to construct a hierarchical representation of the parameters. Thereby, the parameters in the neural networks are derived from the prior model in a hierarchical manner, and hyper-parameters in the prior models are optimized during training. The hierarchical representation in parameters contributes to robust and adaptive learning such as by suppressing over-fitting. We proposed several prior models applicable to layer modules of neural networks, demonstrating the promising performance improvement in classification tasks.

研究分野：パターン認識

キーワード：ニューラルネットワーク 深層学習 事前モデル 画像認識

## 様式 C - 19、F - 19 - 1、Z - 19 (共通)

### 1. 研究開始当初の背景

(1) カメラ等のセンサデバイスの普及により多種多様なデータが収集・蓄積され、それら膨大なパターンデータの自動認識や解析のニーズが急速に高まっている。そのため、データ処理・解析の自動化に資する、いわゆる AI (人工知能) 技術は社会的にも必要不可欠な技術と位置付けられる。

(2) 従来はパターン解析の専門知識に基づいて入力データを数値化 = 特徴として表現することで自動認識を実現していたが、近年では深層学習の枠組みでニューラルネットワーク (NN) の有効性が再認識され、特徴表現自体もデータからの NN 学習により最適化されるようになった。

(3) NN を用いた深層学習は多岐にわたる発展が見られるが、いずれも訓練データに基づく NN モデルパラメータの学習という最適化問題がその根幹にある。深層学習での特徴表現 NN モデルは膨大な数のパラメータで構成される点に特長があり、それ故に学習パラメータ数と訓練データ数の不均衡等に起因するモデルの過学習 (汎化性能の低下) が大きな問題となる。

### 2. 研究の目的

(1) そこで本研究では、パターンデータの特徴表現の汎化性を向上させるため、NN モデルを効率的かつ効果的に学習する新たな数理的アプローチを開発する。

(2) 本提案の枠組みでは、従来の学習対象であった NN のパラメータに対して、新たに事前モデルを導入する。それにより NN パラメータは事前モデルから導出されるものと定義され、パラメータ表現に階層構造を埋め込んでいる点が特長となる。

(3) 事前モデルからのパラメータのサンプリングは、学習過程において入力パターンの変動・摂動とも解釈でき、NN モデル学習の頑健化にも資する。つまり、事前モデルに基づく階層的表現により、データ数の少ない学習環境においても事前モデルの正則化により過学習を軽減し、頑健なモデルの学習が可能となる。

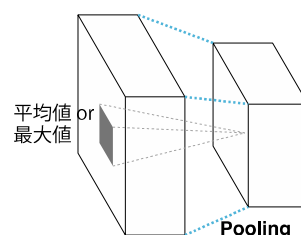
(4) NN モデルは多様な計算処理の層によって構成されており、そのような多様な層・パラメータに対して事前モデルを導入することが目的となる。同時に、計算量的観点からも効率が高く、従来の認識システムへ容易に導入可能となる汎用的な手法の構築を目指す。

### 3. 研究の方法

本研究では、NN モデルを構成する主要な層である、プーリングや畳み込みといった計算処理層に対して事前モデルを導入する技術を開発した。具体的な内容を以下に示す。

#### (1) プーリング層の事前モデル

プーリング層では一般に、入力となる特徴マップの空間次元を圧縮することで計算コストの削減と入力変動への頑健性の向上を同時に実現している。従来のプーリング処理では、特徴マップの局所的統計量である平均値や最大値によって特徴値を代表させることで、圧縮表現を得ていた (右図)。



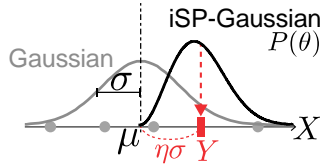
ここでは、局所特徴量に対して事前モデルを導入し、確率的なプーリングを行う手法を開発した。局所特徴分布をガウシアン分布に基づいてモデル化すると、局所特徴量  $X$  の平均  $\mu$  と分散を用いて以下式のように表せる。

$$X \sim \mathcal{N}(\mu, \sigma) \leftrightarrow X = \mu + \epsilon \sigma \quad \text{where } \epsilon \sim \mathcal{N}(0, 1).$$

プーリングにおいては平均値や最大値をとることによる有効性が知られており [1]、平均  $\mu$  以下に向かう変換は有効ではない。そのため、ここではプーリングに対する事前モデル  $P$  として正の値を定義域とする確率分布モデルを導入し、プーリング変換を次式のように定式化した。

$$Y = \mu + \eta \sigma \quad \text{where } \eta \sim P(\theta) \text{ and } \eta \geq 0.$$

これにより、圧縮表現  $Y$  としては入力の平均値  $\mu$  以上の変換値となることが保証される。また、事前モデルはハイパーパラメータを用いて定式化されている。例えば、事前モデルとして逆 softplus ガウシアン分布を組み込んだプーリング変換は次図のように表される。



さらに、事前モデル  $P$  のハイパーパラメータ を事前に設定した固定値ではなく、入力である特徴マップから推定する適応的な枠組みも開発した。

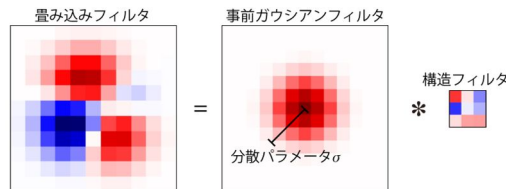
$$\theta = f(X)$$

これにより、入力のパターンが変動した場合でも適応的に事前モデル  $P$  が変化し、多様な入力に頑健な NN モデルを構築することができる。

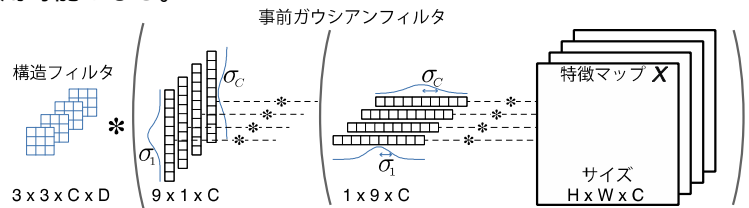
## (2) 畳み込み層の事前モデル

畳み込み (convolution) は画像など多次元データを効率的に処理する計算手法であり、深層学習で広く適用されている畳み込み NN (CNN) の根幹をなす処理となる。この畳み込み演算では局所フィルタを入力多次元データに畳み込み計算することで、入力の特性を効果的に抽出することができ、CNN モデルではそのフィルタパラメータが最適化対象となる。

入力データはある種の構造を内包しており、そのような構造的特性を抽出するためのフィルタ構造が CNN の学習を通して獲得される [2]。ここでは、そのようなフィルタ構造に対する事前モデルとして、ガウシアンモデルを導入した。ガウシアンモデルは例えば微分フィルタなど、様々な (画像) フィルタ形状の基盤となるモデルであり、事前モデルとして採用することは自然である。一般の畳み込みフィルタは、この事前ガウシアンモデルと入力特性に対応した構造フィルタとの畳み込みとして下図のように表現されることになる。



計算の効率性を考慮すると、従来の畳み込み演算は本提案法においては下図に示す形の演算処理となる。等方的な 2 次元ガウシアンフィルタは 1 次元フィルタに分割されることで、計算の効率化を図っている。任意の畳み込み演算は下図に置き換えられるため、提案法は既存の様々な CNN モデルへと適用可能となる。



事前ガウシアンモデルはハイパーパラメータとして分散パラメータ を持ち、この分散パラメータによりガウシアンのがりが制御される。そのため、これは畳み込みフィルタ、ひいては CNN モデルの受容野サイズを決定する重要なパラメータとなる。そのため、ここでは事前ガウシアンモデルの分散パラメータも構造フィルタと同時に学習することで、受容野サイズも適切に最適化した CNN モデルを学習することができる。

## 4. 研究成果

画像認識課題に提案手法を適用することで、その性能を評価した。

### (1) プーリング層の事前モデル

ベンチマークデータセットである ImageNet [3] を用いて、CNN モデルである MobileNet [4]、ResNet-50 [5]、ResNeXt-50 [6] を学習し、その画像認識性能 (1000 カテゴリ識別) を計測した。各 CNN モデルはパラメータ数や構造の違いから基本性能が異なるが、これらはいずれもプーリング層を有しており、そこに提案手法を組み込むことで性能を評価した。性能比較の結果は次表のようになり、提案法が性能向上に有効に貢献することが確認できる。また、ここでは様々な CNN

モデルの性能を向上させていることから、提案手法の汎用性も示されている。

表：ImageNet データセットにおける画像認識性能（正解率[%]）

CNN モデル	MobileNet [4]	ResNet-50 [5]	ResNeXt-50 [6]
Baseline	70.16%	76.47%	77.31%
提案法	72.67%	78.63%	79.34%

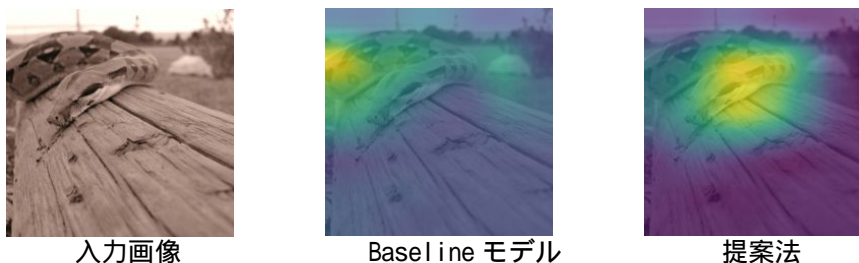
## (2) 畳み込み層の事前モデル

上記と同様に ImageNet データセットを用いて、CNN の畳み込み層に関する提案法の性能を検証した。CNN モデルとしては ResNet-34/50[5]と ResNeXt-50[6]を考え、これら CNN の畳み込み層に対して事前モデルを組み込み、画像識別性能を検証した結果が下表である。提案法により、CNN モデルの性能が向上していることを定量的に確認することができた。

表：ImageNet データセットにおける画像認識性能（正解率[%]）

CNN モデル	ResNet-34 [5]	ResNet-50 [5]	ResNeXt-50 [6]
Baseline	74.10%	76.19%	77.68%
提案法	74.64%	77.42%	78.07%

また、CNN の受容野が提案法によりどのように変化したかを定性的に示した図を以下に示す。ここでは GradCam[7]により算出される受容野重みマップを入力画像に重畳表示している。事前モデルとしてガウシアンモデルを導入し、その分散パラメータも最適化することにより、受容野は適切なサイズとなり、対象をよりよく表現する領域に受容野が反応していることも見て取れる。



入力画像

Baseline モデル

提案法

図：CNN 受容野の図示

## < 引用文献 >

- [1] Takumi Kobayashi, Global feature guided local pooling. In ICCV, pp.3365-3374, 2019.
- [2] Takumi Kobayashi, Analyzing Filters Toward Efficient ConvNet. In CVPR, pp.5619-5628, 2018.
- [3] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In CVPR, pp.248-255, 2009.
- [4] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, Mobilenets: Efficient convolutional neural networks for mobile vision applications. arXiv:1704.04861, 2017.
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition. In CVPR, pp.770-778, 2016.
- [6] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, Kaiming He, Aggregated residual transformations for deep neural networks. In CVPR, pp.5987-5995, 2017.
- [7] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, Dhruv Batra, Grad-CAM: visual explanations from deep networks via gradient-based localization. In: ICCV, pp.618-626, 2017.
- [8] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, Stella X. Yu, Large-scale long-tailed recognition in an open world. In: CVPR. pp.2537-2546, 2019.
- [9] iNaturalist: The inaturalist 2018 competition dataset. [https://github.com/visipedia/inat\\_comp/tree/master/2018](https://github.com/visipedia/inat_comp/tree/master/2018)

5. 主な発表論文等

〔雑誌論文〕 計1件（うち査読付論文 1件 / うち国際共著 0件 / うちオープンアクセス 0件）

1. 著者名 Kobayashi Takumi	4. 巻 169
2. 論文標題 Disentangled convolution for optimizing receptive field	5. 発行年 2023年
3. 雑誌名 Pattern Recognition Letters	6. 最初と最後の頁 67 ~ 74
掲載論文のDOI（デジタルオブジェクト識別子） 10.1016/j.patrec.2023.03.029	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -

〔学会発表〕 計1件（うち招待講演 0件 / うち国際学会 1件）

1. 発表者名 Takumi Kobayashi
2. 発表標題 Rotation Regularization Without Rotation
3. 学会等名 European Conference on Computer Vision (ECCV) (国際学会)
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

6. 研究組織

氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------