

H-023

## 事前教師あり学習を適用した Network in Network による画像認識の高精度化 Applying Supervised Pre-Training to Network in Network for Precise Image Recognition

古井 秀弥<sup>†</sup>  
Shuya Furui

黒木 修隆<sup>†</sup>  
Nobutaka Kuroki

廣瀬 哲也<sup>†</sup>  
Tetsuya Hirose

沼 昌宏<sup>†</sup>  
Masahiro Numa

### 1. はじめに

近年、画像から抽出した特徴量をもとに機械学習を用いる画像認識が様々な場面で用いられ、顔認識や歩行者認識、防犯カメラによる異常検出などが実用化されている。画像認識においては、対象物体の変形および、照明条件の変化に対する頑健性や、認識対象の変更に対して柔軟に対応できる高い汎用性が求められる。

機械学習を用いた画像認識において、特徴抽出処理にニューラル・ネットワークを用いることにより、高精度な認識を行う手法の研究が行われている。近年では、ニューラル・ネットワークの中でも畳み込みニューラル・ネットワーク [1] を用いて、画像認識を行う手法が注目されている。しかし、ニューラル・ネットワークを用いる手法では、良好な結果を得るための学習に、多くの計算リソースを必要とするといった課題がある。そこで本稿では、畳み込みニューラル・ネットワークの一手法である Network In Network [2] における学習処理を改良することによって、学習画像の増加を避けながらも画像認識の精度の向上を実現する手法を提案する。

### 2. Network In Network

#### 2.1 畳み込みニューラル・ネットワークの概要

畳み込みニューラル・ネットワークでは、各ニューロンが入力の全画素に接続されるのではなく、画像の一部の近傍領域のみに接続される。ここで、結合重みが画像の領域ごとに変化しない場合、この接続は画像の一部にフィルタを畳み込むのと同義である。このような構造が、全接続のネットワークよりも位置ずれに強い一因とされる。

図 1 に、入力の近傍画素にのみ接続される局所受容野をもつネットワークを示す。畳み込みニューラル・ネットワークでは、局所受容野によって入力画像の一部、または一部の特徴の出力にしか接続されていない複数のネットワークを、最終段において全接続のネットワークを用いて結合する。図 2 に、畳み込みニューラル・ネットワークにおける畳み込み層の構造を示す。畳み込み層では、最後に活性化関数を用いて出力が得られる。

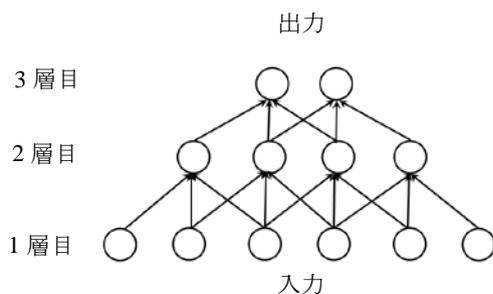


図 1 局所受容野をもつネットワーク

### 2.2 Network In Network の構造

従来の畳み込みニューラル・ネットワークにおいては、局所受容野層において一般に複数のフィルタを用いている。一方、Network In Network では、図 3 に示すように、単純なフィルタの代わりに小規模な多層パーセプトロン (MLP: Multi-Layer Perceptron) を組み込む。これにより、従来の畳み込みニューラル・ネットワークよりも複雑な表現が可能になり、画像の特徴抽出能力が向上すると期待できる。

### 3. 事前学習の検討

データを増加させずに教師あり学習を Network In Network に適用した場合、認識精度は著しく低下する。これは Network In Network では、同じ層数の畳み込みニューラル・ネットワークと比べて、学習パラメータの増加にともない、学習が困難になることが原因と考えられる。この問題への対処法として、主に学習データに対して幾何変換を行うことで、擬似的に大量増加させることが行われる。しかし、GPU を用いた場合はグラフィックカード上に搭載されるメモリの容量は数 GB 程度に限定され、メインメモリとの間で頻繁にスワップを行う必要があり、学習画像の増分以外に要する処理時間が問題となる。

そこで、学習データを増加させる代わりに、事前に教師あり学習を行う手法を提案する。提案手法の概要を図 4 に示す。この手法では、ネットワーク全体に対して教師あり学習させる前に、入力側から小規模 MLP を含む畳み込み層 (NIN 層) を各段階毎に 1 層ずつ識別器に接続し、学習用の入出力として画像とラベルデータを用いた教師あり学習を行う。そして、順に学習済みの NIN 層を積み上げながら、最上段の NIN 層まで 1 層ずつ学習を進行させる。これにより、ネットワーク全体の教師あり学習を行う際の、適切なパラメータの初期値を得ることができると考

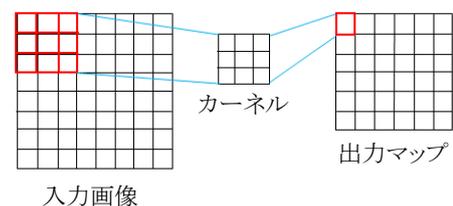


図 2 畳み込みニューラル・ネットワークの畳み込み層

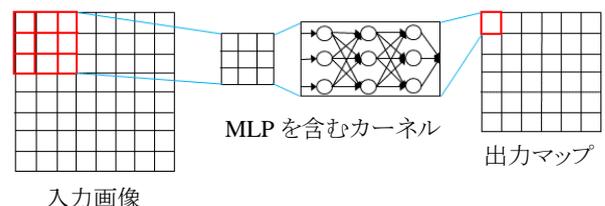


図 3 Network In Network の畳み込み層

<sup>†</sup>神戸大学, Kobe University

えられる。提案手法では図 4 に示す事前学習を行った後、最終段に全接続ネットワークと識別器を接続してネットワーク全体の教師あり学習を行う。

#### 4. 実験

本章では、Network In Network (NIN) を用いた一般物体分類に関する実験結果について述べる。

##### 4.1 実験内容

実験では、CIFAR-10 画像データベース [3] の  $32 \times 32$  pixel のカラー画像と 10 種類の分類クラスを合計 5 万枚学習した後、1 万枚のテスト画像を 10 種類のクラスに分類する。テスト画像 1 万枚を分類した際の精度を、評価項目とする。本実験では、Network In Network 層を 2 層または 3 層積層し、その出力を 2 層の全接続パーセプトロン層に接続し、ロジスティック回帰による分類器を用いた。Network In Network 層内の MLP は 3 層構成で、各層のユニット数は 100 の過剰基底を用いた。また、各 Network In Network 層内で Max-pooling とコントラスト正規化を行っている。活性化関数には ReLU を用いており、学習では全接続層にて Drop Out を行う。なお、入力画像に対して事前に ZCA による白色化とコントラストの正規化 [4] を行っている。また、学習時のアルゴリズムには ADADELTA [5] を用いている。

##### 4.2 結果と考察

表 1 に、通常の教師あり学習を行う NIN (従来手法) と、事前に教師あり学習を行う NIN (提案手法) の実験結果を示す。この結果から、提案手法を用いることで、特に NIN を 3 層用いた場合に認識精度が 40 pt. 以上向上していることが確認できる。その一方で、NIN を 3 積層する優位性は確認できなかった。これは、3 層目の NIN で十分な特徴抽出が行われていないのが原因と考えられ、パラメータ数に対して学習データが不足している可能性がある。また、提案手法では 2 層目の NIN に対して事前教師あり学習を十分に行ってから、3 層目の NIN の学習を同様に始める。

表 1 認識精度 (%)

構成	従来手法	提案手法
NIN-1 層	58.94	58.94
NIN-2 層	40.31	69.31
NIN-3 層	28.66	69.87

そのため、2 層目のパラメータ学習の時点で過学習を起し、3 層構成での認識精度の向上が制限された可能性も考えられる。以上の結果から、今後は NIN のみを積層させるのではなく、通常の畳み込み層も併用する手法や、学習データをリアルタイムで増加させる等の手法が考えられる。また、事前学習の段階で Drop Out を行うことや、NIN 内の MLP についても Drop Out を行うといった手法について検討の余地がある。

#### 5. まとめ

本稿では、事前学習を Network In Network に適用した手法を提案するとともに実装を行い、一般物体認識の実験を行った。その結果、大幅な精度向上が見られたものの、提案手法のみでは十分ではなく、学習データを増加させる必要があるということが確認できた。今後は、ネットワーク内の出力を拡張することにより、同一層内に複数のサイズのフィルタを用いた畳み込みニューラル・ネットワークについて検討を加える。

#### 参考文献

- [1] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," Proceedings of the IEEE, Vol. 86, No. 11, pp. 2278-2324, 1998.
- [2] M. Lin, Q. Chen, S. Yan, Network In Network, Cornell University Library, arXiv: 1312.4400, 2014
- [3] CIFAR-10 データベース, <http://www.cs.toronto.edu/~kriz/cifar.htm>
- [4] A. Krizhevsky, Learning Multiple Layers of Features from Tiny Images, Master's thesis, Department of Computer Science, University of Toronto, 2009.
- [5] M. D. Zeiler, ADADELTA: an adaptive learning rate method, Cornell University Library, arXiv: 1212.5701, 2012

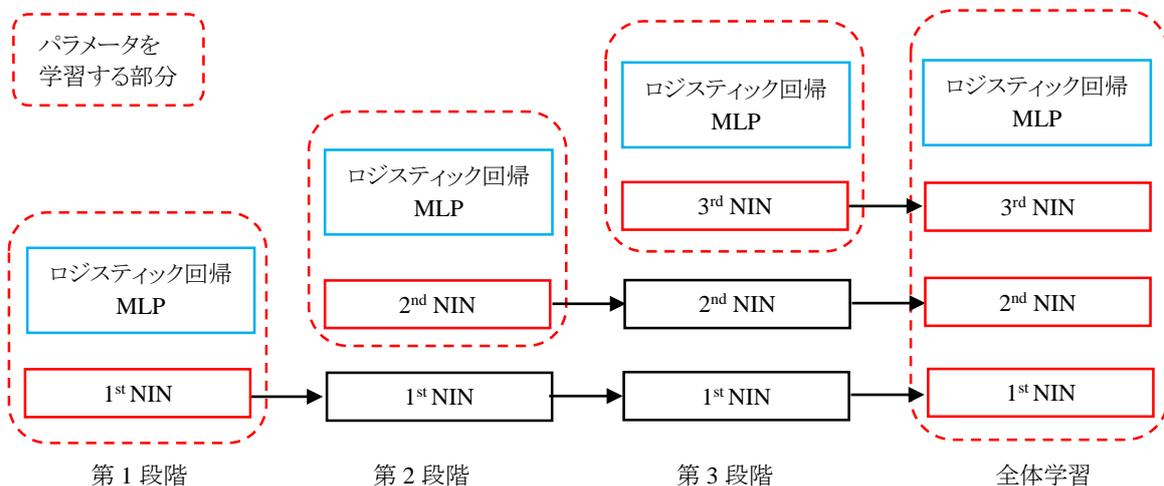


図 4 Network In Network における事前教師あり学習の手順