

平行平均畳み込み処理とドロップアウトによる 畳み込みニューラルネットワークの汎化能力向上

山森 一人^{a)}・長野 泰久^{b)}・相川 勝^{c)}

Generalization Ability Improvement by Parallel Convolution and Dropout

Kunihito YAMAMORI, Yasuhisa NAGANO, Masaru AIKAWA

Abstract

Convolutional neural networks can express models very well, but this ability gives us overfitting problem. To avoid overfitting problem, dropout technique are proposed. However, it is sometimes not enough because convolutional neural networks include a huge number of weights and biases. In this paper, we propose a combination of dropout technique and parallel convolution. In our method, neural network has two parallel convolution layer, and average value of these two convolution layer is used as the input to the pooling layer. Our proposed neural network could improve generalization ability about 4% for the unknown test patterns.

Keywords: Convolutional neural network, Over fitting, Dropout, Parallel convolution layer

1. はじめに

一般物体認識とは、画像内に含まれる物体を一般的な名称でコンピュータに認識させることをいう¹⁾。一般物体認識を行うためには画像内に含まれる物体を検出する必要があり、物体を検出・認識する手法としてテンプレートマッチングや階層型ニューラルネットワークがある。

テンプレートマッチングでは、テンプレートを画像の中で移動させながら、画像との類似度を計算して画像内の物体を認識する。テンプレートマッチングでの画像認識の精度は使用するテンプレートに大きく依存し、テンプレートからの変化が大きい物体に対しては認識精度が低下してしまう。

従来の階層型ニューラルネットワークを用いて画像認識を行う際、入力画像の特徴量を取得する必要がある。特徴量には、Histogram of Oriented Gradients (HOG) や Scale Invariant Feature Transform (SIFT) などがあるが、認識対象に有効な特徴量をこれらの中から取捨選択する必要がある、認識対象についての知識が必要とされる。

近年では、ディープラーニングと呼ばれる多層ニューラルネットワークモデルが注目されており、その一つに畳み込みニューラルネットワークがある²⁾³⁾。畳み込みニューラルネットワークは畳み込み層とプーリング層を複数持ち、認識対象に合った特徴を自動で取得できるため、画像を直接ニューラルネットワークへの入力とすることができる。つまり、事前に特徴量を求めることなく、認

識対象に適した特徴を自動で取得できる。

畳み込みニューラルネットワークはモデルの表現力が高く、学習誤差が小さくなりすぎて過学習に陥る場合がある。過学習を抑える手法の一つにドロップアウト⁴⁾がある。ドロップアウトは、ニューラルネットワークの順伝播の際に一部のユニットを無効にすることでニューラルネットワークの過学習を抑え、未知データに対する認識性能を高める役割を持つ。ドロップアウトは当初、認識を行う全結合層に適用されたが、畳み込み層に適用しても過学習を抑える効果があることが報告されている。

本研究では、畳み込みニューラルネットワークの過学習をより抑制し、汎化能力を向上させることを目的とする。そのため、平行平均畳み込み処理とドロップアウトを併用し、複数の特徴マップの平均値をプーリング層への入力に用いることで畳み込みニューラルネットワークの過学習を抑制し、認識精度を向上させることを目指す。

2. ニューラルネットワーク

2.1. ユニットの学習

図1のように、ニューラルネットワークを構成するユニットには重み w とバイアス θ と呼ばれるパラメータが存在する。これらのパラメータは学習を通じて適切に設定する必要がある。

式(1)に示す通り、ユニットは入力 x と重み w の内積からバイアス θ を減じた値に活性化関数 $f(u)$ を作用させた値 y を出力する。

a) 情報システム工学科教授

b) 情報システム工学科

c) 宮崎大学工学部教育研究支援技術センター技術職員

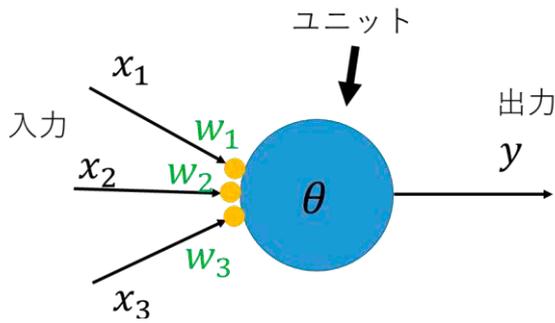


図1. ユニットのモデル.

$$y = f(x \cdot w - \theta). \tag{1}$$

ニューラルネットワークでは、式(1)の重み w とバイアス θ を更新することで学習を進める。更新後の重みを w_{new} 、更新前の重みを w_{old} 、更新後のバイアスを θ_{new} 、更新前のバイアスを θ_{old} 、更新量を調節するために用いられる学習率を η 、ニューラルネットワークの出力と望ましい出力値との誤差を E としたとき、重みとバイアスの更新式は式(2)、式(3)でそれぞれ表される。

$$w_{new} = w_{old} - \frac{\eta \partial E}{\partial w_{old}}. \tag{2}$$

$$\theta_{new} = \theta_{old} - \frac{\eta \partial E}{\partial \theta_{old}}. \tag{3}$$

2.2. 畳み込みニューラルネットワーク

畳み込みニューラルネットワークは、畳み込み層とプーリング層を持つニューラルネットワークである。畳み込み層では、図2に示す畳み込み層への入力と、図3に示す畳み込みフィルタの内積を取ることで特徴マップを得る。入力画像の要素を $x(i, j)$ 、畳み込みフィルタの要素を $h(p, q)$ としたとき、畳み込み処理は式(4)となる。

入力: $x(i, j)$ サイズ: $X \times X$

(0,0)	·	·	·	·	(0, X-1)
·	$x(i, j)$				
·					
·					
·					
(X-1,0)					

図2. 畳み込み層への入力.

サイズ: $H \times H$
 フィルタ: $h(p, q)$

(0,0)	·	(0, H-1)
·	$h(p, q)$	
(H-1,0)		

図3. 畳み込みフィルタの例.

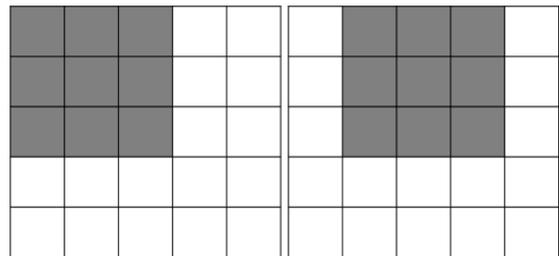


図4. ストライド1の例.

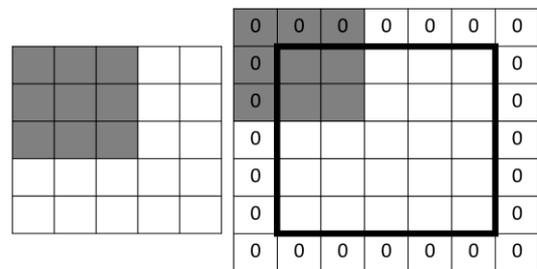


図5. ゼロパディング.

$$u(i, j) = \sum_{p=0}^{H-1} \sum_{q=0}^{H-1} x(i-p, j-q)h(p, q). \tag{4}$$

畳み込み処理ではストライドとパディングを設定する必要がある。ストライドとは、畳み込み処理を行う際のフィルタの移動量のことである。図4は、フィルタサイズが 3×3 で、ストライドが1の場合の例である。パディングとは、特徴マップの周りを任意の数値で埋めることである。ストライドを2以上に設定して畳み込み処理を繰り返すと、特徴マップのサイズが小さくなりすぎる場合がある。特徴マップのサイズが小さくなりすぎることを防ぐためにパディングを行う。図5の例では、左がパディングを行わなかった場合の畳み込み処理で、右がパディングを行った場合の畳み込み処理である。図5のように、特徴マップの周りを0で埋めることをゼロパディングと呼ぶ。プーリング層では特徴マップの指定した領域内から特定の値を取得し、新しい特徴マップを生成する。畳み込みニューラルネットワークでよく使用されるプーリングにマックスプーリングがある。マックスプーリングでは特徴マップの中の特定の領域内の最大値を出力する。入力内の要素 (i, j) を中心とする正方領域 $P(p, q)$ をとり、領域内の要素の集合を P_{ij} で表すと、プーリング

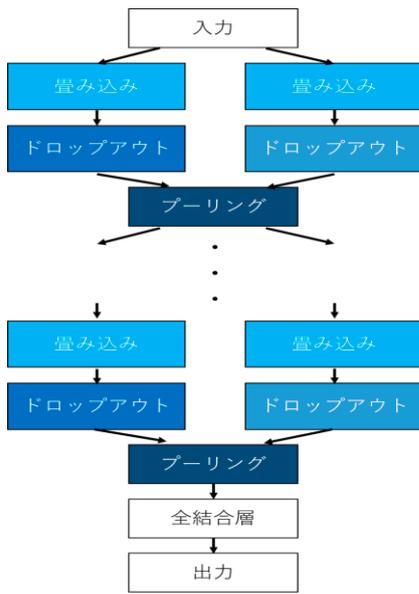


図 6. 提案手法の構成.

後の出力は式(5)で表される。

$$u(i, j) = \max_{(p, q) \in P_{ij}} P(p, q). \quad (5)$$

3. 提案手法

本研究では、ドロップアウトを畳み込み層にも適用するとともに、畳み込み層を平行して 2 つ設け、それらの各要素の平均値をプーリング層への入力とする手法を提案する。提案手法の構成を図 6 に示す。畳み込み層にドロップアウトを適用する例として、図 7 の特徴マップに

1.2	2.9	2.8	1.5
2.6	2.4	2.4	1.0
2.3	1.1	1.4	2.4
1.2	2.2	2.8	0

図 7. 特徴マップ.

2.5	2.5	2.5	0
2.2	2.1	2.1	2.1
2.1	2.2	0	2.1
2.4	2.4	2.5	0

図 8. ドロップアウト適用後の特徴マップ (削減率 20%) .

0	3.3	3.1	3.0
3.0	2.8	2.8	0
2.6	0	2.8	2.8
0	3.1	3.1	0

図 9. ドロップアウト適用後の特徴マップ (削減率 40%) .

1.2	2.9	2.8	1.5
2.6	2.4	2.4	1.0
2.3	1.1	1.4	2.4
1.2	2.2	2.8	0

図 10. 平均後の特徴マップ.

ユニットの削減率20%と40%のドロップアウトを適用する例を考える。削減率20%のドロップアウトでは、特徴マップの20%の要素をランダムに選び値を0にし、それ以外の要素の値を $\frac{1}{1-0.2}$ 倍する。図 7 の特徴マップに 20%のドロップアウトを適用した例を図 8 に、40%のドロップアウトを適用した例を図 9 に示す。提案手法では、図 8、図 9 のようなドロップアウト処理された特徴マップを 2 つ生成し、図 10 のように対応する要素の平均値を求めてプーリング層への出力とする。

4. 提案手法の評価

4.1. 評価方法

畳み込み層とプーリング層の組を 4 つ積層した後、全結合層、出力層と伝播する畳み込みニューラルネットワークにより提案手法を評価する。各層のユニット数は、予備実験により入力層 3,072、第一畳み込み層 8,192、第二畳み込み層 2,048、第三畳み込み層 512、第四畳み込み層 128、全結合層 80、出力層 10 とした。プーリング時のサイズは 2×2 、ストライドは 2 とした。畳み込み層と全結合層の活性化関数にはシグモイド関数を使用し、出力層の活性化関数にはソフトマックス関数、誤差関数にはクロスエントロピー関数を使用する。出力層 k 番目のユニットの出力を O_k 、同ユニットの教師信号を T_k で表した時の誤差関数を式(6)に表す。

$$E_n = - \sum_{k=1}^n T_k \log O_k. \quad (6)$$

実験に用いる画像には CIFAR-10⁵⁾を使用する。CIFAR-10 は 32×32 ピクセルの RGB 画像 6,000 枚 \times 10クラスのデータセットである。各クラス 5,000 枚を学習データ、1,000 枚をテストデータとする。出力層の 10 ユニットが認識画像 10 クラスに対応する。全ての学習データでの学習が一巡することを 1epoch として、学習を 60epoch 繰り返す。平行平均畳み込み層の 2 つの畳み込み層で削減率が同一にならないように、削減率は10%~50%でランダムとし、1epoch 毎に新たな削減率を設定し直す。

重み w は(-1,1)の乱数で、バイアス θ は 0 で初期化する。比較対象は平行平均畳み込み層を用いない畳み込みニューラルネットワークとする。

以上の条件で実験を10回行い、正解率の平均を求める。認識結果は、畳み込みニューラルネットワークの出力値が最も大きいユニットに対応するクラスとする。画像認識の精度は、認識結果が当該画像が属するクラスと一致した画像の枚数を正解数として、式(7)で示す正解率 (accuracy) で評価する。

$$\text{accuracy} = \frac{\text{正解数}}{\text{テストデータ総数}} \times 100(\%) \quad (7)$$

4.2. 実験結果

提案手法の正解率を図10、畳み込み層のドロップアウトのみを適用する場合の正解率を図11に示す。

図10から、学習データの正解率は80%、テストデータの正解率は75%であり、図11から学習データの正解率は77%、テストデータの正解率は71%となり、提案手法によってテストデータの正解率を4%向上できた。

図11から、正解率が急に落ちている epoch があることが分かる。これは、直前の epoch では認識に有効だった特徴がドロップアウトで失われたため、正解率が低下したと考えられる。

図12に、学習終了時の全結合層への入力のうち、値が0である要素と0以外の要素の数を示す。図12から、平行平均畳み込み処理を用いない場合は、値0の要素数が提案手法の約2倍となっているため、畳み込み層へのドロップアウトの適用により、有効な特徴量を失いやすく

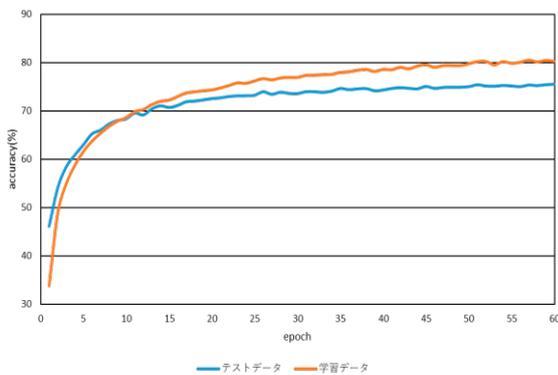


図11. 提案手法での正解率の推移.

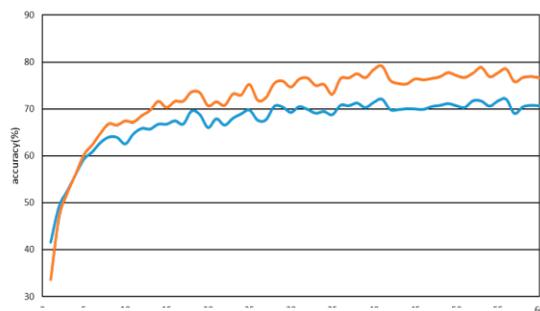


図12. 平行平均畳み込み処理を行わないときの正解率の推移.

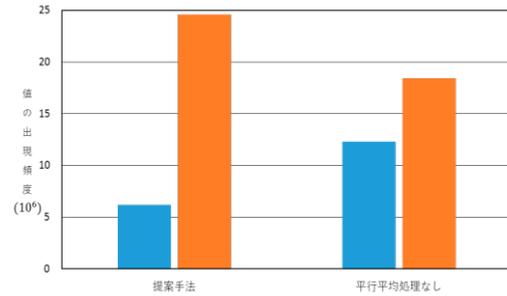


図13. 第四プーリング層の出力.

なつたと考えられる。提案手法では、ドロップアウト後の2つの特徴マップの平均値を使うため特徴を失いにくく、安定した認識を行うことができたと考えられる。

5. おわりに

畳み込みニューラルネットワークは大量の学習パラメータを持ち表現力が高く、学習データに対する認識精度が高くなりやすい。一方で、学習データを認識するために最適化されたパラメータでは、未知データに対する認識精度が低くなる過学習に陥ることがある。本研究では、畳み込みニューラルネットワークの過学習を抑制し、汎化能力を向上させることを目的とし、平行平均畳み込み処理とドロップアウトを併用する手法を提案した。実験の結果、提案手法は平行平均畳み込み処理を用いない場合と比べて、テストデータの正解率が4%向上し、畳み込みニューラルネットワークの汎化性能が向上したことを示した。

今後の課題として、誤差関数や活性化関数を変え、認識精度を向上させることが挙げられる。

参考文献

- 1) 柳井啓司, “一般物体認識の現状と今後”, 情報処理学会研究報告コンピュータビジョンとイメージメディア (CVIM), Vol.2006, No.93, pp. 121-134 (2006).
- 2) Q. V. Le, M. Ranzato, R. Monga, M. Devin, K. Chen, G. S. Corrado, J. Dean and A. Y. Ng, “Building high-level features using large scale unsupervised learning”, ArXiv-prints(2011).
- 3) 久保陽太郎, “ディープラーニングによるパターン認識”, 情報処理, Vol. 54, No. 5, pp. 500-508 (2013)..
- 4) N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever and R. Salakhutdinov, “Dropout: A simple way to prevent neural networks from overfitting”, Journal of Machine Learning Research, Vol. 15, pp. 1929-1958 (2014).
- 5) “cifar-10 データセット <http://www.cs.toronto.edu/kriz/cifar.html>” .