

深層零平均正規化相互相関ネットワーク

荒井 敏^{1,a)} 長尾 智晴¹

受付日 2017年11月13日, 再受付日 2017年12月25日,
採録日 2018年1月26日

概要: 本稿では, 新たな深層ネットワークのモデルを提案する. 提案モデルはテンプレートマッチングの分野で広く用いられている零平均正規化相互相関を基本要素とし, これを多層に組み合わせたものである. 画像分類のタスクにおいて既知の画像データセットを用いた精度比較を行い, 従来手法と同等以上の分類精度が得られることを示す. また, 入力画像の明るさが変動した場合を想定した精度検証を行い, 提案手法がそれらの変動に対して従来手法よりも頑健であることを示す.

キーワード: 深層学習, テンプレートマッチング, 零平均正規化相互相関 (ZNCC), 正規化, 画像分類

Deep Zero-mean Normalized Cross Correlation Networks

SATOSHI ARAI^{1,a)} TOMOHARU NAGAO¹

Received: November 13, 2017, Revised: December 25, 2017,
Accepted: January 26, 2018

Abstract: In this paper, we propose a new deep network model. Our proposed model consists zero-mean normalized cross correlation which is widely used in the field of template matching as basic elements and it is a multi-layered combination of those. We show that our method performs equally or better than conventional methods in accuracy comparison for image classification tasks using known image datasets. Moreover, in accuracy verification assuming that the brightness of input images vary, we show that our method is more robust than conventional methods for the variation.

Keywords: deep learning, template matching, zero-mean normalized cross correlation, normalization, image classification

1. はじめに

深層学習 [1], [2], [3], [4] の目覚ましい発展にともない, 産業界においても広く活用の機運が高まっている. そのような中で, 深層学習あるいは深層ネットワークの動作機序について非専門家からも説明を求められることがあるが, 数学理論として「正しい」説明ではまず理解を得られない. では一体どこまで噛み砕けば誤解を含まず納得してもらえるかといえば, これは非常に難しい問題である.

テンプレートマッチングくらいのシンプルさであれば非専門家にも理解してもらえると期待してよいであろう. そ

れならばテンプレートマッチングのように機序を説明しやすく, そのうえで実用的な分類精度が得られるようなモデルは構築できないだろうか.

このような着想をふまえ, 本稿では, 零平均正規化相互相関 (Zero-mean Normalized Cross Correlation ; ZNCC) [5] を用いた新たな深層ネットワークモデル (ZNCC Net) を提案する.

提案モデルの特徴は以下のとおりである.

- テンプレートマッチングの分野で広く用いられている ZNCC をベースとする深層ネットワークであり, 比較的クラス数の少ない画像分類タスクにおいて, 従来手法と同等以上の分類精度が得られる.
- 深層ネットワークの新たな正規化手法であり, Batch Normalization などのようにバッチ統計量を用いることなく, 従来手法と同等以上の分類精度が得られる.

¹ 横浜国立大学大学院環境情報学府
Graduate School of Environment and Information Sciences,
Yokohama National University, Yokohama, Kanagawa 240–
8501, Japan

^{a)} arai-satoshi-rw@ynu.jp

2. 関連研究

2.1 テンプレートマッチング

テンプレートマッチングは最も単純な画像認識手法の1つであり、距離や類似度を用いて、記憶されたパターンと入力パターンとの直接的な照合を行うものである。

パターン間の距離を算出する手法として、L1 距離ベースの差分絶対値総和 (SAD)、L2 距離ベースの差分二乗総和 (SSD) などが知られている。また、パターン間の類似度を算出する手法として、ベクトルの成角のコサインを類似度とする正規化相互相関 (NCC)、さらに中心化を組み合わせた零平均正規化相互相関 (ZNCC) などが広く用いられている。

2.2 畳み込みニューラルネットワーク

LeCun ら [6] によって提案された畳み込みニューラルネットワーク (CNN) は、画像を取り扱うタスクにおいて最も標準的に用いられるネットワーク構成である。畳み込み演算は局所的に内積を算出し、これを空間方向に反復するものであり、これは Balestrierio [7] によれば内積による類似度算出をベースとしたテンプレートマッチングととらえられる。しかし実際には内積演算による類似度算出結果にバイアスが加算されるため、パターンの類似度は出力値とは直接対応していない。

2.3 正規化

Batch Normalization (BN) は Ioffe ら [8] によって提案された手法であり、バッチ統計量を用いた中心化・正規化を行うことでより大きな学習率に対しても学習を安定化させ、その結果、より少ない学習回数で効率的な学習を可能とするものである。

Weight Normalization (WN) は Salimans ら [9] によって提案され、フィルタを単位ベクトルとスカラー倍率に分解することで簡便な正規化を行うものである。

Cosine Normalization (CN) は Chunjie ら [10] によって提案され、入力パターンとフィルタの内積に代えて正規化相互相関 (NCC) を用いる。バイアス加算は不要であるが、認識精度は他の正規化手法と比べて見劣りする。

3. 提案手法

以下、提案手法について詳説する。まず基本的な考え方を述べたのち、ZNCC Net の構成要素を説明する。さらに改良点と実装方法について解説し、最後に他の正規化手法とともに特性を比較する。

3.1 基本的な考え方

筆者らの考え方は以下のようにきわめて単純である。

1. パターン認識は (k-近傍法のように) 記憶されたパター

ンとの照合で解決できる。特に画像の照合であればテンプレートマッチングが最もシンプルで分かりやすい。

2. しかし現実にはすべてのパターンをテンプレートとして記憶することは不可能であり、効率化のためにパターンの分割や共有、そして階層化が必要となる。
3. そのようにして得られたモデルは階層型のネットワークそのものであり、End-to-End 学習によって最適化が可能であろう。

このような発想に基づき、ZNCC を基本要素とする深層ネットワークを構築するに至った。非専門家に対しては「少し入り組んだテンプレートマッチング」として十分説明可能であると期待される。

3.2 零平均正規化相互相関 (ZNCC)

零平均正規化相互相関 (Zero-mean Normalized Cross Correlation; ZNCC) [5] は、バイアスの変動に頑健な系列データの類似度算出手法として広く知られている。特に画像処理の分野では、輝度の変化に頑健な画像間類似度算出手法として、テンプレートマッチングの際によく用いられる。

入力 $X \in \mathbb{R}^{R \times C \times D}$ とテンプレート $W \in \mathbb{R}^{U \times V \times D \times K}$ が与えられた場合、両者の ZNCC のインデックス (r, c, k) の要素は式 (1) で与えられる。ただし、 R, C は入力空間サイズ、 D はチャンネル数を表す。 U, V はテンプレートの空間サイズ、 K はテンプレート数を表す。小文字は各次元に対応するインデックスである。また、簡単のため $\sum \bullet = \sum_{u=-U'}^{U'} \sum_{v=-V'}^{V'} \sum_{d=0}^{D-1} \bullet$, $U' = (U-1)/2$, $V' = (V-1)/2$ と表記している。

$$\begin{aligned} \text{ZNCC}(X, W)(r, c, k) &= \frac{\sum \{X(r+u, c+v, d) - \bar{X}(r, c)\} \times \{W(u+U', v+V', d, k) - \bar{W}(k)\}}{\sqrt{\sum \{X(r+u, c+v, d) - \bar{X}(r, c)\}^2} \times \sqrt{\sum \{W(u+U', v+V', d, k) - \bar{W}(k)\}^2}} \quad (1) \end{aligned}$$

$$\bar{X}(r, c) = \frac{1}{\sum 1} \sum X(r+u, c+v, d) \quad (2)$$

$$\bar{W}(k) = \frac{1}{\sum 1} \sum W(u+U', v+V', d, k) \quad (3)$$

しかしこれらの記法はきわめて煩雑であるため、簡略化する。 X, W の要素のうち式 (1) に関与するものだけを抜き出してそれぞれ一列に並べたベクトルを $\mathbf{x}, \mathbf{w}$ とすると、式 (1) は式 (4) のように簡潔に記述できる。以降この形式を用いる。ただし、 $(\mathbf{x}, \mathbf{w})$ はベクトル $\mathbf{x}$ と $\mathbf{w}$ の内積を表し、 $\|\mathbf{x}\|$ は $\mathbf{x}$ の L2 ノルムを表す。また、 $\bar{\mathbf{x}}$ はベクトル $\mathbf{x}$ の全要素の平均値 (スカラー) を表し、 $\mathbf{x} - \bar{\mathbf{x}}$ はベクトル $\mathbf{x}$ の全要素から同一のスカラー値 $\bar{\mathbf{x}}$ を減算する操作を表す。

ZNCC は式 (4) が表すように、内積演算に平均値の減算 (中心化) とノルムの除算 (正規化) を組み合わせた処理であると理解される。

$$\text{ZNCC}(x, w) = \frac{(x - \bar{x}, w - \bar{w})}{\|x - \bar{x}\| \|w - \bar{w}\|} \quad (4)$$

実際には0による除算を防ぐため、式(4)に代えて式(5)を用いる。ここで ε は微小定数であり、本稿では特に断らない限り 10^{-5} を用いる。

$$\text{ZNCC}(x, w) = \frac{(x - \bar{x}, w - \bar{w})}{\sqrt{\|x - \bar{x}\|^2 + \varepsilon} \sqrt{\|w - \bar{w}\|^2 + \varepsilon}} \quad (5)$$

ZNCCの微分は式(6)および式(7)で与えられる。

$$\begin{aligned} & \frac{\partial}{\partial x} \text{ZNCC}(x, w) \\ &= \frac{1}{\|x - \bar{x}\|} \left(\frac{w - \bar{w}}{\|w - \bar{w}\|} - \frac{x - \bar{x}}{\|x - \bar{x}\|} \text{ZNCC}(x, w) \right) \end{aligned} \quad (6)$$

$$\begin{aligned} & \frac{\partial}{\partial w} \text{ZNCC}(x, w) \\ &= \frac{1}{\|w - \bar{w}\|} \left(\frac{x - \bar{x}}{\|x - \bar{x}\|} - \frac{w - \bar{w}}{\|w - \bar{w}\|} \text{ZNCC}(x, w) \right) \end{aligned} \quad (7)$$

なお W について、以降ニューラルネットワークの用語に従って(テンプレートではなく)フィルタと呼ぶ。

3.3 ZNCC Networks

本節ではZNCCを基本要素とする深層ネットワークZNCC Networks(以後ZNCC Netと呼称)について、特に画像分類タスクのための構成を説明する。

ZNCC Netの基本的な構成要素(ZNCC Block)は図1(a)に示すとおりで、ZNCCとRectified Linear Unit (ReLU) [11]を順に適用するものである。その意味は本手法の基本的考え方に基づき、図1(b)のように自然に解釈される。ここで負の類似度をクリップするためのReLUは、記憶と大きく異なる雑音パターンを除外するために必要と考えられる。

複数のZNCC Blockを適用することで、ZNCCの出力にさらにZNCCを適用することになり、部分の照合から全体の照合に向かうような階層的なテンプレートマッチングが構成される。その後、Global Average Poolingを用いてZNCCの出力を集約する。この流れは通常のCNNと同じである。その後Full Connect層(FC)に入力するが、ここでも同様に式(5)に従って中心化と正規化を行う。これをZNCC FCと呼称する。最後にReLUを適用しネットワーク全体の出力とする(図2)。通常のCNNとは異なり、Softmax回帰[12]は使用しない。この点については3.4節で再度触れる。

なお、複数のZNCC Blockを重ねる際は、通常のCNNと同様0-paddingを行って空間サイズを調節する。ZNCCにおいて値0はフィルタと無相関であることを表すため、追加された画素は前層の出力と見なした場合、すべてのフィルタと無相関の扱いとなり、空間範囲外という特殊条

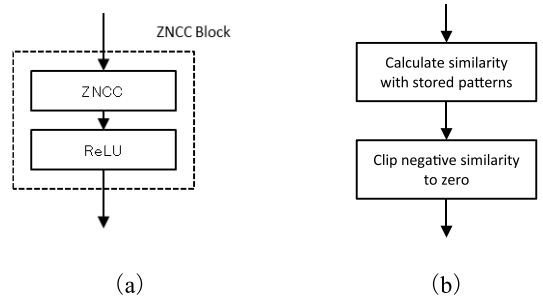


図1 ZNCC Netの基本構成要素

Fig. 1 Basic components of ZNCC Net. (a) ZNCC block consists of ZNCC layer and ReLU. (b) Concept diagram of ZNCC block. ZNCC layer calculates similarity with stored patterns then ReLU clips negative similarity values to zero.

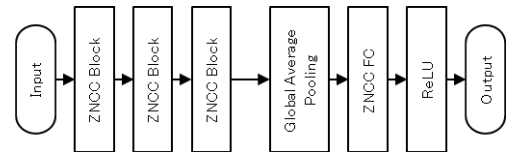


図2 クラス分類用 ZNCC Net の基本構成

Fig. 2 Typical architecture of ZNCC Net for classification. It consists plural ZNCC blocks followed by global average pooling and ZNCC full connect layer. We use pair of ReLU and Weighted Mean Squared Error (WMSE) instead of pair of softmax and categorical cross entropy widely used in CNNs.

件を表す値としてリーズナブルである。

3.4 損失関数

深層ニューラルネットワークを用いた画像分類のタスクでは、最終層の活性化関数および損失関数として Softmax 回帰と Categorical Cross Entropy (CCE) の組合せが広く用いられている。しかし筆者らの実験では、これらを ZNCC Net に適用した場合は学習が十分に進まない現象が見られた。そこで筆者らはいくつかの試行の結果、出力関数および損失関数として ReLU と Weighted Mean Squared Error (WMSE) をそれぞれ用いることにした。

WMSE はネットワークからの予測値 y と正解値 t が与えられた場合、式(8)で定式化される。ここで x^2 はベクトル x の要素ごとの二乗を表す。 y と t は同じサイズのベクトルであり、特に t は要素の1つだけが1で他の要素が0であるようなベクトル、いわゆる 1-hot ベクトルである。WMSE では t の値1の要素の重みと値0の要素の重みをバランスさせるので、これによって予測値のすべての要素が0という自明解に陥るのを抑制する。

$$\text{WMSE}(y, t) = \left((y - t)^2, \left(\frac{t}{\|t\|} + \frac{1 - t}{\|1 - t\|} \right) \right) \quad (8)$$

ZNCC Net において Softmax + CCE がうまく機能しない理由として、筆者らは ZNCC と Softmax それぞれの持つ

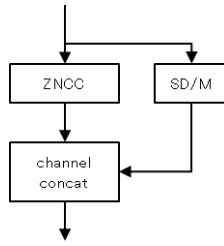


図 3 標準偏差・平均値 (SD/M) チャンネル付き ZNCC

Fig. 3 ZNCC with Standard Deviation and Mean (SD/M) channels. Though ZNCC discards some information during centering and normalization, SD/M channels compensates similar information.

正規化作用が干渉しているためではないかと考えているが、現時点では確証は得られていない。

3.5 ZNCC Net の改良

提案手法の分類精度を向上させるため、以下のような改良を加えた。

3.5.1 標準偏差チャンネル・平均値チャンネル

ZNCC は局所的な中心化と正規化を行いながらパターンの類似度を算出するため、バイアスの変動に対して頑健である。しかし一方、中心化や正規化で削除される情報が有益なものであっても、これらを活用することができない。この弱点を解決するため、筆者らは中心化と正規化で削除される情報と同様の情報を ZNCC とは別に伝達するための経路を用意した。

経路は 2 つあり、それぞれ標準偏差 (SD) チャンネルおよび平均値 (M) チャンネルと呼称する。これらはそれぞれ ZNCC と同じ空間サイズでチャンネル数 1 のマップを出力する。出力されたマップは ZNCC の出力とチャンネル連結され、次層に送られる。その結果、ZNCC の出力チャンネル数は 2 増加する (図 3)。

3.2 節の記法に従えば、標準偏差チャンネルと平均値チャンネルはそれぞれ式 (9) および式 (10) で定式化される。ただし、 N は $\mathbf{x}$ の要素数を表す。

$$\text{StdDev}(\mathbf{x}) = \frac{1}{\sqrt{N}} \|\mathbf{x} - \bar{\mathbf{x}}\| \quad (9)$$

$$\text{Mean}(\mathbf{x}) = \bar{\mathbf{x}} \quad (10)$$

3.5.2 ショートカット経路との組合せ

He ら [13], [14] によって提案された加算型のショートカット経路は ZNCC Net にもそのまま取り入れることができる。さらに標準偏差・平均値 (SD/M) チャンネルと組み合わせることで ZNCC の弱点を補いつつ、より深層のネットワークを構築することが可能となる。この構成を ZNCC ResBlock と呼称する (図 4)。ただし、図 4 の “Nop or Conv” は、ブロック内の ZNCC によって空間解像度とチャンネル数の少なくとも一方が変更される場合は Conv を用いてサイズ合わせの変換を行い、それ以外では何もせずに素

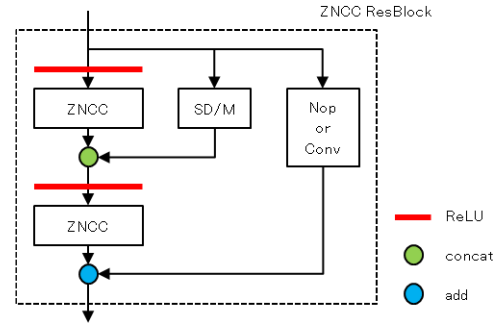


図 4 ZNCC ResBlock

Fig. 4 Block diagram of ZNCC ResBlock. ZNCC ResBlock is a combination of pre-activated ZNCC blocks and SD/M channels. “Nop or Conv” means convolution when spatial and/or channel size changes in the block, otherwise it means identity mapping.

通しすることを表す。

3.6 ZNCC の実装

深層学習の一般的なフレームワークを用いて ZNCC を実装する方法について述べる。

式 (4) は式 (11) の形式に変形できることが知られている [5]。ただし、 N はベクトル $\mathbf{x}$ の要素数を、 $\mathbf{x}^2$ は $\mathbf{x}$ の要素ごとの二乗を表す。また、 $\mathbf{1}$ は $\mathbf{w}$ とサイズが等しく全要素が 1 であるようなベクトルを意味する。

$$\begin{aligned} \text{ZNCC}(\mathbf{x}, \mathbf{w}) &= \frac{N(\mathbf{x}, \mathbf{w}) - (\mathbf{x}, \mathbf{1})(\mathbf{w}, \mathbf{1})}{\sqrt{N(\mathbf{x}^2, \mathbf{1}) - (\mathbf{x}, \mathbf{1})^2} \sqrt{N(\mathbf{w}^2, \mathbf{1}) - (\mathbf{w}, \mathbf{1})^2}} \quad (11) \end{aligned}$$

これを利用すると式 (1) の X と W の ZNCC (全要素) は式 (12) として計算できる。ここで N は W の要素数、 $\mathbf{1}$ は W とサイズが等しく全要素が 1 であるようなテンソル、 $\text{conv}(X, W)$ は X と W の畳み込み、 $\text{sum}(W)$ は W の要素のチャンネルごと総和を表す。また、べき乗演算はすべて要素ごとに行う。サイズの異なるテンソルどうしの演算ではサイズが揃うように要素を反復拡張 (broadcast) し、テンソルとスカラ間の演算ではテンソルの要素ごとに同じスカラ演算を行う。

$$\begin{aligned} \text{ZNCC}(X, W) &= \frac{N \cdot \text{conv}(X, W) - \text{conv}(X, \mathbf{1}) \cdot \text{sum}(W)}{\sqrt{N \cdot \text{conv}(X^2, \mathbf{1}) - \{\text{conv}(X, \mathbf{1})\}^2} \times \sqrt{N \cdot \text{sum}(W^2) - \{\text{sum}(W)\}^2}} \quad (12) \end{aligned}$$

式 (12) の実装では、ZNCC を計算するために 3 回の畳み込み演算が必要である。これは同じメモリ領域に複数のアクセスが発生することを意味し、パフォーマンスを低下させる。ただし、実装の工夫によってメモリへのアクセス回数を減らすことは可能であり、高速化の余地が残されているといえる。

表 1 正規化手法の特徴比較

Table 1 Feature comparison of normalization method. Only ZNCC performs centering on both training and test. This feature makes ZNCC Net robust against luminance variation.

手法	中心化		正規化		ゲイン補正	バイアス加算	バッチ統計
	学習時	テスト時	学習時	テスト時			
Conv	—	—	—	—	—	—	—
Conv+Bias	—	—	—	—	—	○	—
BN [8]	○	—	○	—	○	○	○
WN [9]	—	—	○	—	○	○	—
WN+mean only BN[9]	○	—	○	—	○	○	○
NCC (CN) [10]	—	—	○	○	—	—	—
ZNCC	○	○	○	○	—	—	—

3.7 関連手法との比較

ZNCC にはテンプレートマッチング手法としての側面と正規化手法としての側面があるが、ここでは後者に注目し、深層ネットワーク向けの他の正規化手法との比較を行う。

手法を整理・分類するために、(1) 中心化・正規化、(2) ゲイン補正・バイアス加算、(3) バッチ統計、の3点に注目する (表 1)。

(1) 中心化・正規化

平均値を減算する中心化とノルムを除算する正規化は、それぞれ学習時のみ行うか、テスト時にも行うかが手法によって異なる。ZNCC は学習時とテスト時の両方で中心化と正規化を行う。特にテスト時の中心化は他の手法にはない特徴であり、輝度変動に対する頑健性が期待される。

(2) ゲイン補正・バイアス加算

中心化・正規化を行った後にゲイン補正やバイアス加算を行うことで、中心化や正規化という強すぎる制約を緩和する効果が見込まれる。しかし NCC や ZNCC ではゲイン補正やバイアス加算なしでも学習が可能なので、必須というわけではない。

(3) バッチ統計

バッチ統計量はサンプル統計量の近似値として安価に計算可能であり、中心化・正規化の効果を高めると考えられる。しかし一方、バッチ統計量を算出するために適度なサイズのミニバッチを用いる必要があり、学習設定の自由度が制約される。ZNCC ではバッチ統計量は使用せずに単一画像における近傍領域の情報のみを用いて中心化・正規化を行っている。

4. 実験

提案手法の有効性を検証するため、ベンチマーク用画像データセットを用いて実験を行った。

4.1 実験設定

本稿の実験に共通して用いた条件設定を以下に示す。

4.1.1 画像データセット

CIFAR-10 [15] は一般物体を被写体とする 10 クラスの

表 2 実験に使用した環境 [17], [18]

Table 2 Environment commonly used in our experiments.

OS	Windows 10 Pro
CPU	Intel Core i7-6700K 4GHz
GPU	NVIDIA Titan X
開発環境	Visual Studio 2013, CUDA 8.0, cuDNN 6 Python 3.6, TensorFlow 1.3, Keras 2.0.6

画像分類用データセットである。学習用に 50,000 枚、テスト用に 10,000 枚の画像が用意されており、各画像には 10 クラスのいずれかのラベルが付与されている。各画像は 32×32 画素の RGB 画像である。

CIFAR-100 [15] は CIFAR-10 と同様に一般物体の画像分類用データセットであるが、全 100 クラスである点が異なる。学習用・テスト用の画像数と画像サイズは CIFAR-10 と同じであるため、1 クラスあたりの画像数は 1/10 になっている。

SVHN Format 2 [16] は Google Street View から抽出された家屋番号画像のデータセットである。画像中央の数字に 0~9 のラベルが与えられており、これを推定する 10 クラス分類のタスクに使用する。学習用とテスト用の画像はそれぞれ 73,257 枚と 26,032 枚であり、各画像は CIFAR-10 と同様 32×32 画素の RGB 画像となっている。

4.1.2 実験環境

本稿の実験に使用したソフトウェアおよびハードウェア環境を表 2 に示す。

4.2 他の類似度算出手法との比較

深層ネットワークに ZNCC を用いることの有効性を検証するため、ZNCC を他の類似度算出手法に入れ換えた場合との比較を行った。比較対象は、バイアス項なしの畳み込み (Conv)、バイアス項ありの畳み込み (Conv + Bias)、正規化相互相関 (NCC) である。ZNCC に関しては、前処理を適用しない場合 (no prep) についても実験した。

実験には Salimans ら [9] で使用されたものと同じ構成のネットワークを用いた。類似度算出に ZNCC を使用する場合のネットワーク構成を表 3 に示す。パラメータ欄の “3x3, c96, s1” はフィルタサイズ 3x3, 出力チャネル数 96, スライド 1 を表し (他の場合も同様)、“ReLU” は活性化関数として Rectified Linear Unit を適用することを意味する。最終層 (層番号 10) の活性化関数と損失関数の組合せとして、Conv および Conv + Bias の場合は Softmax + CCE, NCC および ZNCC の場合は ReLU + WMSE を用いる。

学習時の条件設定を表 4 に示す。学習率の初期値として全手法共通で 0.25 を使用したが、学習開始から 20 epoch 経過しても教師データに対する分類精度の上昇が見られない場合、初期値を 0.1 に切り下げて実験をやり直した。したがって、手法によって実際の学習率の初期値は異なる。また ZNCC に前処理を適用しない場合、学習の発散を抑え

表 3 ネットワーク構成 (1)

Table 3 Network architecture used for experiment 4.2, 4.3 and 4.5. The network consists of ten ZNCC layers. c96 and s2 mean 96 channels and stride of 2 respectively.

層番号	種別	パラメータ
1	ZNCC	3x3, c96, s1, ReLU
2	ZNCC	3x3, c96, s1, ReLU
3	ZNCC	3x3, c96, s1, ReLU
	MaxPool	2x2, s2
4	ZNCC	3x3, c96, s1, ReLU
5	ZNCC	3x3, c96, s1, ReLU
6	ZNCC	3x3, c96, s1, ReLU
	MaxPool	2x2, s2
7	ZNCC	3x3, c192, s1, ReLU
8	ZNCC	1x1, c192, s1, ReLU
9	ZNCC	1x1, c192, s1, ReLU
	Global Average Pooling	
10	ZNCC-FC	c10, ReLU

表 4 学習設定

Table 4 Setting for training process commonly used in our experiments.

エポック数	200
バッチサイズ	128
最適化手法	SGD
学習率	初期値0.25から終了値0.0005までをコサイン関数に沿って減衰
モーメント	0.9
Weight Decay	0.0001
前処理	教師データ全サンプルの画素ごとの平均値を減算して中心化
データ拡張	画像の上下左右に4画素0 padding. $\pm$ 4画素のランダムシフト切り出し+水平方向のランダムフリップ

表 5 類似度算出手法の比較結果

Table 5 Comparison result of similarity calculation methods. Bold underlined means the best result for each dataset. ZNCC outperforms other methods for CIFAR-10 and SVHN. “no prep” means no preprocessing.

手法	CIFAR-10		CIFAR-100		SVHN
	学習パラメータ数	分類精度	学習時間	分類精度	分類精度
Conv	659K	91.90%	1	68.11%	95.44%
Conv+Bias	660K	91.98%	1.07	69.24%	94.83%
NCC	659K	91.06%	2.07	66.69%	95.78%
ZNCC	659K	92.69%	2.60	68.92%	96.40%
ZNCC (no prep)	659K	92.71%	2.60	69.18%	96.24%

るため式 (5) の ε を 5×10^{-5} とした。

各手法によるテストデータの分類結果を表 5 に示す。各画像データセットに対して最良の結果を太字下線で示した。ここで学習時間は CIFAR-10 で 1 epoch 学習するのに要した時間であり、最も学習時間の短かった Conv を 1 とした場合の比で表した。学習パラメータ数は CIFAR-10 および SVHN を対象とした場合の値を掲載した。

表 6 正規化手法の比較結果 (CIFAR-10)

Table 6 Comparison result of normalization methods. ZNCC outperforms other methods with/without batch statistics except WN + mean only BN.

手法	分類精度	
	バッチ統計なし	バッチ統計あり
NCC (CN) [10] (ours)	89.60%	
mean only BN [9]		91.48%
WN [9]	91.54%	
Conv + Bias [9]	91.57%	
Conv + BN [9]		91.95%
WN + mean only BN [9]		92.69%
ZNCC (ours)	92.00%	

なお実際には、この他の類似度算出手法として SSD についても評価を試みた。具体的には、SSD によって算出された二乗距離に指数関数 $e^{-\kappa x}$ を適用して類似度に変換し (κ はパラメータ)、学習可能となるパラメータを探索しながら他の手法と同様に学習を試みた。しかし筆者らが試行したすべての条件設定において学習に成功しなかったため、今回の比較対象からは除外した。

4.3 他の正規化手法との比較

正規化手法としての ZNCC の有効性を検証するため、他の深層ネットワーク向け正規化手法と比較を行った。

実験には文献 [9] と同じネットワーク構成を使用した。各層の演算を ZNCC に、最終層の活性化関数を ReLU に変更している。これは 4.2 節の構成と同一であり、詳細は表 3 に記載のとおりである。

条件を揃えて直接比較を行うため、学習には文献 [9] と同様の設定で Adam を用いた。ただし、学習中盤で教師データに対する分類精度が低下しさらに停滞する現象が見られたため、学習率の初期値を 0.0005 に変更した。

CIFAR-10 テストデータの分類結果を表 6 に示す。ただし、Cosine Normalization (CN) は正規化相互相関 (NCC) と同一の処理であるため、NCC (CN) のように記載している。なお、表 6 における WN + mean only BN の分類精度が表 5 における ZNCC の分類精度と同値 (92.69%) になっているが、これは偶然の一致である。

4.4 超深層畳み込みニューラルネットワークとの比較

3.5 節で提案した ZNCC Net の改良手法を用いて深層ニューラルネットワークを構成し、画像分類タスクにおいて代表的な超深層畳み込みニューラルネットワークとの比較を行った。比較対象とするネットワークは ResNet-110 および ResNet-164 [14] である。

なお、本実験は提案手法を用いてある程度実用的なモデルが構築可能であることを検証するためのものであり、state-of-the-art を目指したものではない。

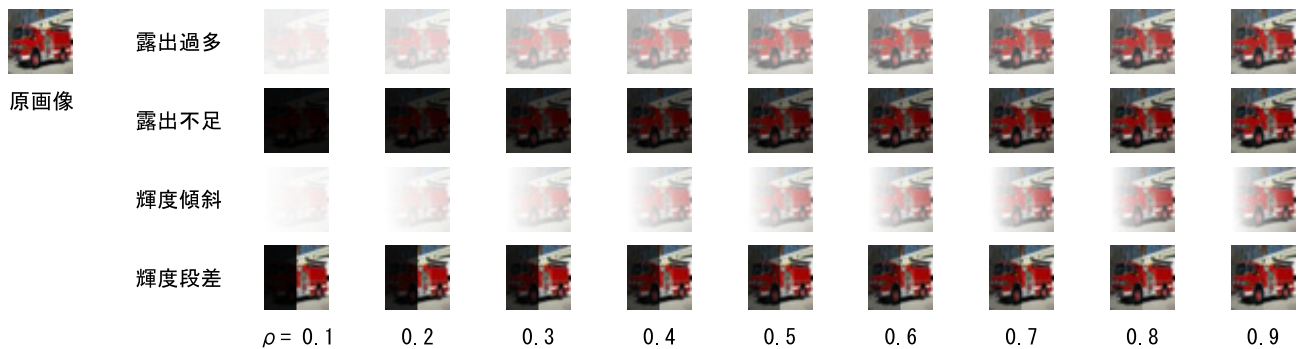


図 5 悪照明条件の画像サンプル (1)

Fig. 5 First sample images of poor lighting condition. From top to bottom, each row imitates over-exposure, under-exposure, gradient transition of luminance, and stepwise transition of luminance respectively. The smaller ρ , the worse lighting condition. The original image is shown in the leftmost column.



図 6 悪照明条件の画像サンプル (2)

Fig. 6 Second sample images of poor lighting condition. White spot which imitates specular reflection exists at the center of images.

表 7 ネットワーク構成 (2)

Table 7 Network architecture used for experiment 4.4. The network consists of fourteen ZNCC layers. Detail of ZNCC ResBlock is depicted in Fig. 4.

層番号	種別	パラメータ	Nop or Conv
1	ZNCC	3x3, c16, s1	Nop
2	ZNCC ResBlock	3x3, c48, s1, ReLU	Conv(1x1)
3		3x3, c48, s1, ReLU	
4	ZNCC ResBlock	3x3, c48, s1, ReLU	Nop
5		3x3, c48, s1, ReLU	
6	ZNCC ResBlock	3x3, c96, s2, ReLU	Conv(2x2)
7		3x3, c96, s1, ReLU	
8	ZNCC ResBlock	3x3, c96, s1, ReLU	Nop
9		3x3, c96, s1, ReLU	
10	ZNCC ResBlock	3x3, c192, s2, ReLU	Conv(2x2)
11		3x3, c192, s1, ReLU	
12	ZNCC ResBlock	3x3, c192, s1, ReLU	Nop
13		3x3, c192, s1, ReLU	
	ReLU		
	Global Average Pooling		
14	ZNCC-FC	c10, ReLU	

本実験で用いたネットワーク構成を表 7 に示す。ただし，“Nop or Conv”は図 4 に記載された同名項目の実際の動作内容を表す。

CIFAR-10 および CIFAR-100 のテストデータを分類した結果を表 8 に示す。

4.5 悪照明条件に対する頑健性

ZNCC は輝度の変動に頑健な類似度算出手法であるため、ZNCC Net においても同様の特性を有していることが期待される。これを検証するために実験を行った。

表 8 超深層 CNN との比較結果

Table 8 Comparison result with extremely deep CNNs. Fourteen-layer ZNCC Net outperforms CNNs of more than hundred layers for both CIFAR-10 and CIFAR-100.

手法	層数	CIFAR-10		CIFAR-100	
		パラメータ数	分類精度	パラメータ数	分類精度
ResNet-110 [14]	110	1.7M	93.63%	—	—
ResNet-164 [14]	164	1.7M	94.54%	1.7M	75.67%
ZNCC Net (ours)	14	1.63M	94.66%	1.65M	73.48%

ここでは照明条件の不適正な状態（悪照明条件）を模した補正を画像に加え、その結果画像分類精度がどの程度の影響を受けるかを観察した。

本実験では以下の 5 つの悪照明条件のシナリオを想定した。画像サンプルを図 5, 図 6 に示す。また、本来の画像の輝度を I ($0 \leq I \leq 1$) とした場合のテスト用画像の生成方法をあわせて示す。

- 露出過多：自動露出補正の失敗などで画像全体が白っぽくなった状態 ($I * \rho + (1 - \rho)$)。
- 露出不足：自動露出補正の失敗などで画像全体が暗く沈んだ状態 ($I * \rho$)。
- 輝度傾斜：強い光がレンズに斜めに入射するなどして画像の一部が白っぽくなった状態（画像の左端 1 から右端 $I * \rho + (1 - \rho)$ に傾斜）。
- 輝度段差：日向と日陰のように明るさの大きく異なる領域が同一画像中に同時に現れた状態（画像の左半分のみ $I * \rho$)。
- 白飛び：照明光が被写体表面で正反射し、それを受光

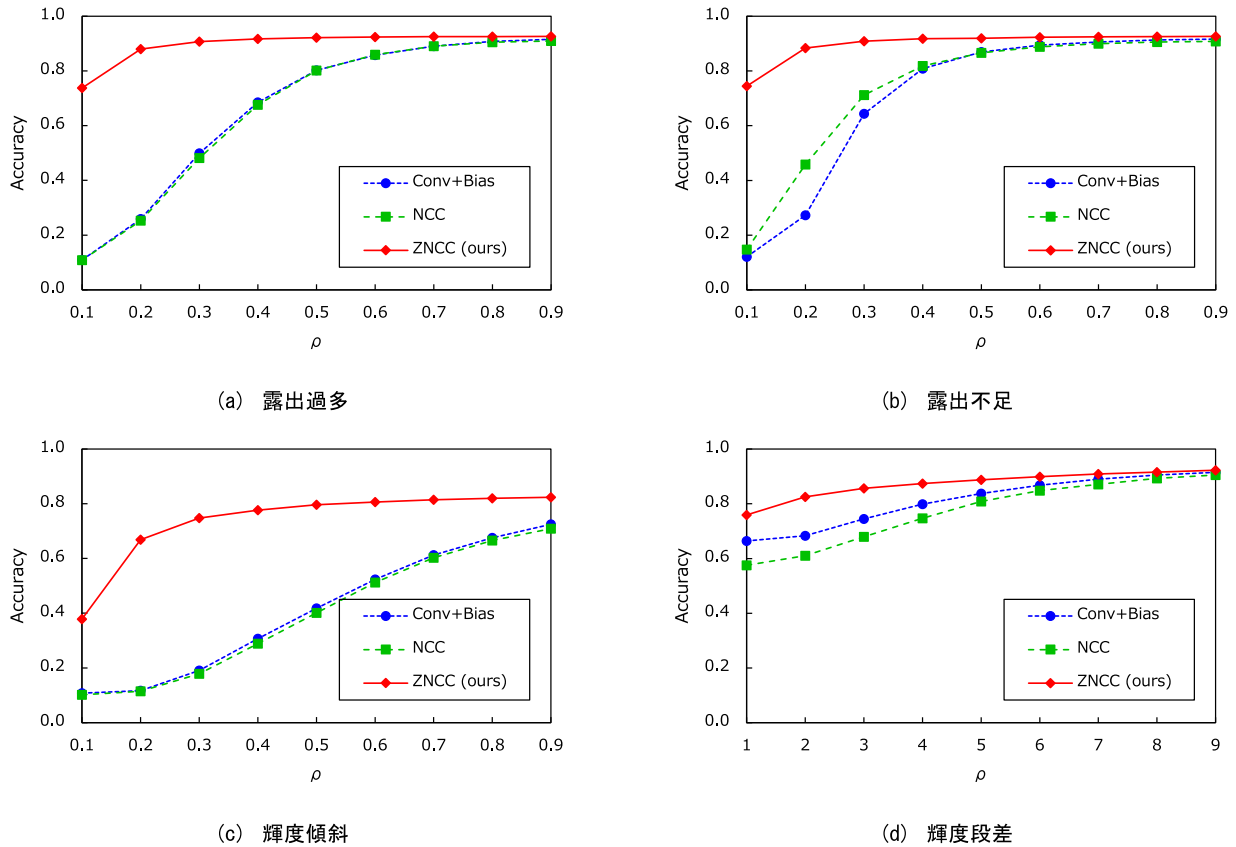


図 7 悪照明条件画像における画像分類精度

Fig. 7 Image classification accuracy on poor lighting condition. (a) over-exposure, (b) under-exposure, (c) gradient transition of luminance, and (d) stepwise transition of luminance. ZNCC is more robust than other methods against lighting condition variations.

した撮像素子の輝度値が白色に飽和して観察された状態（画像中央 size $\times$ size 画素を 1 に）。

ここで ρ ($0 \leq \rho \leq 1$) は照明条件を調整するパラメータであり、値が小さいほど条件が悪いことを表す。

CIFAR-10 テストデータに上記悪照明条件を模した修正を適用し、4.2 節で学習済みのモデルを用いてクラス分類する実験を行った。Conv + Bias, NCC, ZNCC の各手法を用いて CIFAR-10 のテストデータを分類した結果を図 7 および表 9 に示す。

5. 考察

これまでの実験をふまえて、以下の 4 項目について考察を行った。

5.1 類似度算出手法との比較

CIFAR-10 および SVHN の分類において、ZNCC の分類精度は他の類似度算出手法を上回っている。しかし CIFAR-100 に関しては Conv + Bias の分類精度が ZNCC を上回っており、一概に ZNCC が優れているとはいえない。また、ZNCC は処理時間が最も長く、Conv と比較して 2.6 倍もの時間を要している点はデメリットである。ただし、これ

は実装の最適化によって今後改善される可能性があり、現時点では判断が難しい。

ZNCC の場合、前処理の有無にかかわらず同程度の分類精度が得られており、前処理による中心化は必須ではないことが分かる。

5.2 正規化手法との比較

CIFAR-10 の分類において、ZNCC の分類精度はバッチ統計量を使用しない他の正規化手法を上回っている。バッチ統計量を使用する正規化手法のうち Conv + BN とは同程度の分類精度であるが、WN + meanonlyBN とは約 0.7% の差がついている。バッチ統計量を使用する場合ミニバッチの構成などで学習設定に制約が生じるので、バッチ統計量を用いずに比較的高い分類精度が得られる点は提案手法のメリットである。

5.3 超深層畳み込みニューラルネットワークとの比較

CIFAR-10 の分類において、提案手法はわずか 14 層でありながら、ResNet-110, ResNet-164 をともに上回った。ネットワークの層数はそのまま処理の待ち時間につながるため、浅い層で比較的高い分類精度が得られる点はメリッ

表 9 悪照明条件画像における画像分類精度 (詳細)

Table 9 Image classification accuracy on poor lighting condition. Our method (ZNCC) outperforms other methods in all of poor lighting conditions in our experiments.

照明条件		Conv+Bias (基準精度 : 91.98%)	NCC (基準精度 : 91.06%)	ZNCC (提案手法) (基準精度 : 92.69%)
		分類精度	分類精度	分類精度
露出過多	$\rho = 0.2$	25.87%	25.27%	87.96%
	$\rho = 0.5$	80.21%	80.14%	92.19%
	$\rho = 0.8$	90.77%	90.47%	92.58%
露出不足	$\rho = 0.2$	27.25%	45.80%	88.36%
	$\rho = 0.5$	86.91%	86.61%	91.96%
	$\rho = 0.8$	91.37%	90.60%	92.59%
輝度傾斜	$\rho = 0.2$	11.71%	11.43%	66.89%
	$\rho = 0.5$	41.79%	40.11%	79.67%
	$\rho = 0.8$	67.52%	66.53%	82.01%
輝度段差	$\rho = 0.2$	68.36%	61.06%	82.57%
	$\rho = 0.5$	83.78%	80.87%	88.79%
	$\rho = 0.8$	90.54%	89.33%	91.62%
白飛び	size = 4	84.80%	84.39%	87.33%
	size = 2	89.09%	89.36%	90.79%
	size = 1	89.52%	89.70%	91.62%

トといえる。

一方, CIFAR-100 に関しては ResNet-164 を下回った。5.1 節の結果もあわせて考えると提案手法は CIFAR-100 の分類を苦手としていると思われるが, 現時点で原因は判明しておらず, 今後さらなる研究が必要である。

5.4 悪照明条件に対する頑健性

悪照明条件を模した CIFAR-10 テストデータの分類において, 提案手法は他の手法と比較して悪照明条件による分類精度の低下が小さく, 相対的に高い分類精度を維持する結果となった。ZNCC の持つ輝度変動に対する頑健性がそのまま活かされているといえる。このような悪条件照明はデータ拡張によって学習することも可能だが, すべてのパターンを尽くすことは難しく, また手間も膨大になると考えられるため, 提案手法が元より備える頑健性は大きなメリットといえる。

6. まとめ

本稿では零平均正規化相互相関 (ZNCC) を基本要素とする深層ネットワークを提案した。提案手法を画像分類タスクに適用したところ, 他の手法と比較しておおむね高い分類精度が得られ, 悪照明条件に対しても頑健であった。これらの検討によって本手法の有効性を示した。

参考文献

- [1] LeCun, Y. et al.: Deep Learning, *Nature*, Vol.521, pp.436-444 (2015).
- [2] 中山英樹: ディープラーニングの発展と最新動向, 画像

- 電子学会 (2016/06/01) (2016).
- [3] 山下隆義: ディープラーニングによる画像認識と応用事例, *DeepLearningDay2016* (2016).
- [4] 岡谷貴之: ディープラーニング, 映像情報メディア学会誌, Vol.68, No.6, pp.466-471 (2014).
- [5] Lewis, J.P.: Fast Normalized Cross-Correlation, *Vision Interface* (1995).
- [6] LeCun, Y. et al.: Backpropagation Applied to Handwritten Zip Code Recognition, *Neural Computation*, Vol.1, No.4, pp.541-551 (1989).
- [7] Balestrieri, R.: Adaptive Partitioning Spline Neural Networks: Template Matching, Memorization, Inhibitor Connections, Inversion, Semi-Sup, Topology Search, ..., arxiv:1710.09302 (2017).
- [8] Ioffe, S. and Szegedy, C.: Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift, arxiv:1502.03167 (2015).
- [9] Salimans, T. and Kingma, D.: Weight Normalization: A Simple Reparameterization to Accelerate Training of Deep Neural Networks, *Advances in Neural Information Processing Systems 29*, pp.901-909 (2016).
- [10] Chunjie, L. et al.: Cosine Normalization: Using Cosine Similarity Instead of Dot Product in Neural Networks, arxiv:1702.05870 (2017).
- [11] Nair, V. and Hinton, G.E.: Rectified Linear Units Improve Restricted Boltzmann Machines, *Proc. 27th International Conference on Machine Learning* (2010).
- [12] Heckerman, D. and Meek, C.: Models and Selection Criteria for Regression and Classification, *Proc. 13th Conference on Uncertainty in Artificial Intelligence* (1997).
- [13] He, K. et al.: Deep Residual Learning for Image Recognition, *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp.770-778 (2016).
- [14] He, K. et al.: Identity Mappings in Deep Residual Networks, *Computer Vision - ECCV 2016*, pp.630-645 (2016).
- [15] Krizhevsky, A.: Learning Multiple Layers of Features

from Tiny Images, Technical Report, University of Toronto (2009).

- [16] Netzer, Y. et al.: Reading Digits in Natural Images with Unsupervised Feature Learning, *NIPS Workshop on Deep Learning and Unsupervised Feature Learning* (2011).
- [17] Abadi, M. et al.: TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems, arxiv: 1603.04467v2 (2016).
- [18] Chollet, F. et al.: Keras: <https://github.com/fchollet/keras> (2015).



荒井 敏 (正会員)

1995 年東京工業大学大学院総合理工学研究科博士課程前期修了。精密機器メーカー勤務。横浜国立大学大学院環境情報学府に在学。



長尾 智晴 (正会員)

1985 年東京工業大学大学院総合理工学研究科博士課程後期中退。同年同大学助手。同大学助教授を経て、2001 年横浜国立大学大学院環境情報研究院教授。工学博士。画像処理，進化計算法等の知能情報学の研究に従事。電子

情報通信学会，人工知能学会，進化計算学会，IEEE 等各会員。