

## 1. はじめに

Convolutional Neural Network (CNN) は、コンピュータビジョンや自然言語処理の研究分野で広く用いられている。CNN は、中間層において多数のパラメータを持つことで、学習データに対する複雑な関数近似が可能となり、高精度な表現学習が可能となる。一方で、パラメータ数が多い場合、学習データへの過剰適合が発生する可能性や、複数の計算機を用いた分散学習時にデータ転送が遅くなるなどの問題点がある。そこで本研究では、Sparse Coding を CNN に適応することで、Sparse Coding が持つ正則化効果により、少数のパラメータで高精度な特徴抽出を実現する。研究のアプローチは、Sparse Coding における最適化の反復演算をレイヤ方向に重み共有することで、CNN をベースとする新たな学習モデル CSC-Network を提案する。Sparse Coding に基づき CNN を再考することで、工学上 3 つのメリットが得られる。1 つ目は、正則化効果による精度向上である。2 つ目は、Sparse Coding の理論背景に基づいた学習が可能である。3 つ目は、レイヤを多層にすることで、少数のパラメータで高い非線形性を持った特徴表現が得られる。

## 2. Sparse Coding

Sparse Coding [1] は、可能な限り少ない基底の線形和を用いて入力データを表現する手法である。Sparse Coding は、データ圧縮、ノイズ除去、特徴抽出など幅広く応用されている。入力データ  $y \in \mathbb{R}^m$ 、基底行列 (Dictionary)  $D \in \mathbb{R}^{m \times n}$ 、係数ベクトル  $x \in \mathbb{R}^n$  とすると、Sparse Coding における条件式は次式である。

$$\min_x \|x\|_1 \quad \text{subject to } y = Dx \quad (1)$$

$\min \|x\|_1$  は、 $x$  をスパース (要素がほとんどゼロ) なベクトルにするための制約である。 $y$  のみが既知である場合、 $Dx$  の解が無限に存在するという問題を正則化効果により防ぐことが可能となる。

## 3. 提案手法

本研究では、Sparse Coding を最適化手法の近接勾配法 [2] を用いて Unroll することで、誤差逆伝播が可能な Convolutional Sparse Coding-Network (CSC-Network) を提案する。

### 3.1 CSC-Network の設計

Sparse Coding で用いられる一般的なコスト関数は、次式である。ここで、 $\lambda$  は正則化効果の強さを定めるパラメータである。第 1 項目は、入力データと  $Dx$  で再構成したベクトルの誤差、第 2 項目は、 $x$  のスパース性を表している。

$$\min_{D, x} E(x) + \lambda \|x\|_1, \quad E(x) = \frac{1}{2} \|y - Dx\|_2^2 \quad (2)$$

上記の Sparse Coding のコスト関数を画像問題に適応させるため、コスト関数を以下のように定義する。

$$\min_{D^c, X^c} E(X^c) + \lambda \|X^c\|_1 \quad (3)$$

$$E(X^c) = \frac{1}{2} \|Y^c - D^c \otimes X^c\|_2^2 \quad (4)$$

ここで、 $Y^c \in \mathbb{R}^{h \times w \times n^2}$  及び  $X^c \in \mathbb{R}^{h \times w \times d}$  は、図 1 で示すように入力画像を変形したテンソル及び係数行列とし、 $D^c \in \mathbb{R}^{n^2 \times d}$  を辞書とする。ただし、 $h$  及び  $w$  を入力画像の縦及び横のサイズとし、 $d$  を出力チャンネル数となる辞書サイズ、 $n$  を Sparse Coding を適応する正方形フィルタの一辺のサイズとする。

式 (3) のコスト関数における  $X^c$  及び  $Y^c$  の最適化を行う。はじめに、 $X^c$  の最適化について述べる。式 (3) は、L2

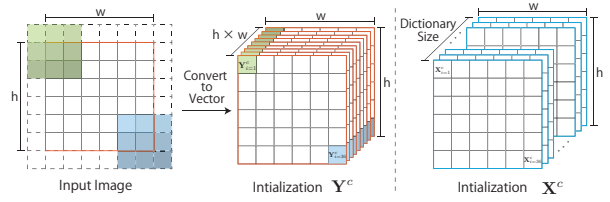


図 1: CSC-Network に対する入力データの初期化

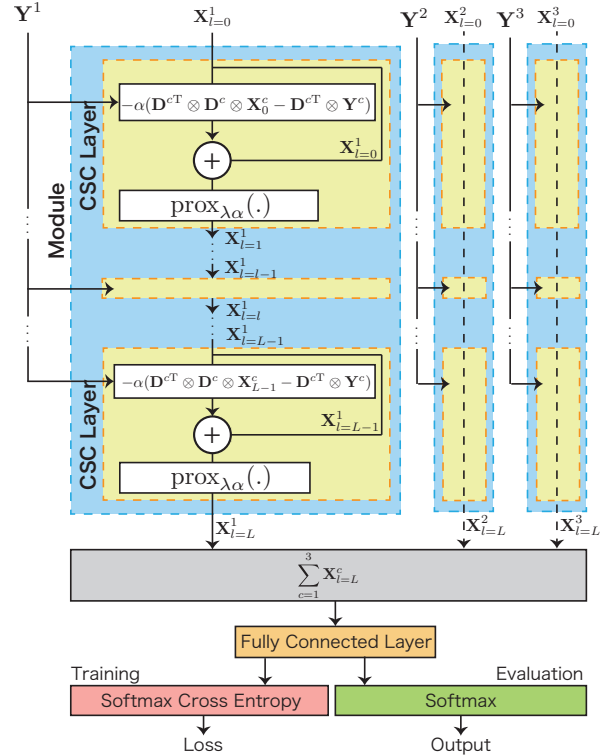


図 2: CSC-Network の全体構造

ノルムと L1 ノルムの 2 項で構成されている。そのため、滑らかな凸関数と滑らかではない凸関数の合成である。上記の合成関数を最適化する手法として、次式の近接勾配法を用いる。ここで、 $\alpha$  は近接勾配法における学習率である。

$$x_{l+1} = \text{prox}_{\lambda\alpha}(x_l - \alpha \frac{dE(x_l)}{dx_l}) \quad (5)$$

また、近接勾配法における prox 作要素は、次式で表す通りである。ただし、 $\beta = \lambda\alpha$  とする。

$$\text{prox}_{\beta}(z) = \begin{cases} z + \beta & (z < -\beta) \\ 0 & (-\beta \leq z \leq \beta) \\ z - \beta & (z > \beta) \end{cases} \quad (6)$$

次式は、コスト関数の第 1 項目に対する  $x_l$  での微分を示している。

$$\frac{dE(x_l)}{dx_l} = D^{cT} \otimes D^c \otimes X^c - D^{cT} \otimes Y^c \quad (7)$$

続いて、CSC-Network 構築のための近接勾配法の Unroll を行う。式 (5) は、 $X^c$  についての漸化式であり、 $l$  層目の更新式を表している。そのため、式 (5) を  $l$  層目の CSC Layer とし、多層に連結することで、図 2 で示す Module を構成できる。ここで、辞書  $D^c$  は、全 CSC Layer において共有する。近接勾配法の Unroll により構成した Module

の後に, Fully Connected Layer (FC Layer) を接続することで, 識別対象のクラス数に合わせて次元圧縮を行う. 対象のタスクは, 識別問題であるため, 出力層において Softmax 関数を用いる. 本研究では, Module, FC Layer, Softmax 関数で構成されるネットワークを CSC-Network として提案する.

### 3.2 CSC-Network の学習

はじめに, FC Layer の学習について述べる. FC Layer のパラメータの初期値は, 平均 0, 標準偏差 0.05 の正規乱数とする. FC Layer の学習誤差を導出するため, CSC-Network の学習誤差を誤差逆伝播する. ここで, CSC-Network の学習誤差は, Softmax Cross Entropy 関数の出力から得る. FC Layer の学習誤差より更新量を求め, 勾配法に従いイテレーション毎にパラメータを更新する.

続いて, Module 内部の  $X^c$  及び  $D^c$  の最適化・学習について述べる.  $X^c$  及び  $D^c$  は逐次更新する.  $X^c$  の最適化は, 近接勾配法に基づき展開された CSC Layer においてレイヤ毎に行う.  $X^c$  の初期値は, 平均 0, 標準偏差 0.05 の正規乱数とする.  $X^c$  は近接勾配法に従い更新するため, レイヤが深くなるにつれ更新量が収束する.  $D^c$  の学習は, FC Layer の学習と同様にイテレーション毎に行う.  $D^c$  の初期値は, 平均 0, 標準偏差 0.1 の正規乱数とする.  $D^c$  の学習誤差を導出するために, FC Layer の学習誤差を CSC-Network の 1 レイヤ目まで誤差逆伝播する.  $D^c$  の学習誤差より更新量を求め, 勾配法に従いパラメータを更新する.

### 3.3 CSC-Network の推論処理

CSC-Network の推論は, 学習により最適化された  $D^{c*}$  及び FC Layer を用いる. また, 近接勾配法の収束レートを改善するため, Accelerated proximal gradient algorithm [3] を用いる. 入力データ  $Y^c$  及び  $X_0^c$  を Module に入力し, CSC-Network の学習時と同様に  $X^c$  をレイヤ毎に更新する. Module からの特徴マップ  $X_L^c$  は, チャンネル毎に加算を行い統合する. 統合後は, FC Layer 及び Softmax 関数の順に適用し, クラス毎の確率値を算出する. 最も確率値の高いクラスを推論結果とする.

## 4. 評価実験

CSC-Network の有効性を検証するため, CNN と性能比較を行う.

### 4.1 実験概要

CIFAR-10 データセットを用いて識別精度を評価する. CIFAR-10 は, 10 種類の識別タスク用のデータセットである. 実験で使用する画像は, 学習用に 50,000 枚, 評価用に 10,000 枚とする. CNN は, CSC-Network の CSC Layer を Convolution, 活性化関数 ReLU に置き換えたネットワークとする. CSC-Network と CNN のレイヤ数を 1~6 の間で変更し, 各レイヤ数において評価する. フィルタサイズは, CNN が  $3 \times 3$ , CSC-Network が  $9 \times 9$  とする. CSC-Network の辞書及び CNN のフィルタの最適化は Adam とする. また, CSC-Network の  $\lambda$  は, レイヤ数 1:5, レイヤ数 2:3, レイヤ数 3:2.5, レイヤ数 4:2, レイヤ数 5:1.5, レイヤ数 6:1 とする.

### 4.2 実験結果

CSC-Network と CNN の Error Rate を図 3 に示す. レイヤ数が 1~3 において, CSC-Network の方が CNN よりも良い識別精度となった. レイヤ数 4 では, CSC-Network と CNN の識別精度の差がなくなり, レイヤ数 5 以上では, CNN の方が良い識別精度となった. CSC-Network は, レイヤ数 3 において最も良い精度となった. レイヤ数 4 以降で精度が悪化した原因は, 近接勾配法による更新量が収束した後に, 設定した学習率  $\alpha$  の影響により, 最適解の近くで振動したためと考えられる.

CSC-Network と CNN のパラメータ数を図 4 に示す. CNN は, レイヤ数の増加に伴いパラメータ数も増加していることがわかる. これは, 各レイヤにおいて他のレイヤと独立した学習パラメータを持ち, レイヤ数の増加に従いパラメータ数も単調増加するためである. 一方で, CSC-

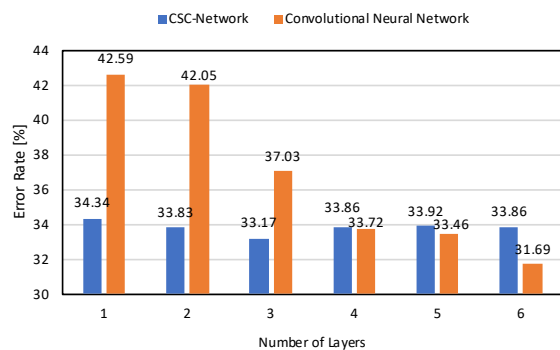


図 3: CSC-Network と CNN の Error Rate [%] の比較

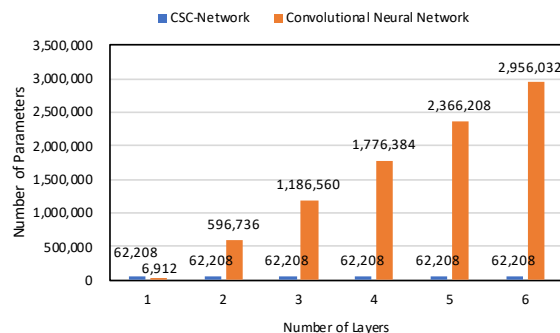


図 4: CSC-Network と CNN のパラメータ数の比較

Network のパラメータ数は, 全レイヤ数において 62,208 である. CSC-Network は, レイヤ方向に共有されているため, レイヤ数が増加した場合においてもパラメータ数は一定となる. レイヤ数 1 の CSC-Network のパラメータ数が多い要因は, フィルタサイズの違いによるものである.

以上の結果から, CSC-Network は, 3 層の CNN より高精度な識別が可能であることがわかる. 4 層以降で CNN の精度を下回っている要因として, CSC-Network は漸化式に基づき特徴を更新しており, レイヤが深くなるほど新たな特徴を獲得することが困難であるためと考えられる.

## 5. おわりに

本研究では, Sparse Coding を CNN に適応することで, 少ないパラメータ数において有効な特徴表現が可能なネットワークを提案した. CSC-Network は, レイヤ数 3 まで識別精度が単調増加し, CNN の約 19 分の 1 のパラメータ数で 2.77 ポイント高い精度を実現した. 今後の課題としては, CSC-Network における Module 部分の多層化を行い, 識別精度の向上に取り組む.

## 参考文献

- [1] 富岡亮太, “スパース正則化学習の理論とアルゴリズム”, 日本応用数理学会論文誌, Vol. 23, No. 3, pp. 485–515, 2013.
- [2] S. Mark, *et al.*, “Convergence Rates of Inexact Proximal-Gradient Methods for Convex Optimization”, NIPS, 2011.
- [3] A. Beck, *et al.*, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems”, SIAM J. Imaging Science, Vol. 2, No. 1, pp. 183–202, 2009.

## 研究業績

- [1] 古川弘憲 等, “2D-QRNN を導入した DCNN によるセマンティックセグメンテーションの高精度化と高速化”, MIRU, 2017.
- [2] 古川弘憲 等, “Dilated 2D-QRNN を導入した DCNN によるセマンティックセグメンテーションの高精度化と高速化”, WiNF, 2017.
- [3] 古川弘憲 等, “Residual Network へのスパースコーディング導入”, MIRU, 2018.