センサベースの行動認識における CNN のカーネルサイズに関する一考察

清水椋右[1]　近藤和真[1,†1]　長谷川達人[1]

概要： スマートフォンやウェアラブルデバイスの普及に伴い，深層学習を用いたセンサベースの行動認識が盛んにおこなわれるようになった．しかし，現在は畳み込み層が 3 層程度のシンプルな CNN がよく用いられており，行動認識に特化した深層学習モデルの構造は明らかではない．ウェアラブルデバイスを用いて行動認識を行う場合，計算コストの削減は大きな課題である．モダンな深層学習モデルは一般に計算コストが高く，改変を行わず行動認識に適用するには不適当である．行動認識において深層学習モデルの軽量化についての議論は進んでいない．本研究では VGG 構造を対象として，畳み込み層のカーネルサイズに着目し，行動認識精度やモデルのパラメータ数に現れる影響を調査する．現在デファクトスタンダードである，小さいカーネルの多数積層構造を，これと同等の範囲の受容野を持つより大きいカーネルの単層の畳み込み層に変更することで，パラメータ数を削減しつつ，行動認識精度が向上する可能性があることが判明した．

Study on Kernel Size of CNN for Sensor-Based Activity Recognition

RYOSUKE SHIMIZU[1]　KAZUMA KONDO[1,†1]　TATSUHITO HASEGAWA[1]

1. はじめに

スマートフォンやウェアラブルデバイスの普及に伴い，センサベースの行動認識が盛んに行われるようになった．センサベースの行動認識では機械学習や深層学習が主に用いられている．中でも，畳み込み層が 3 層程度のシンプルな Convolutional Neural Network (CNN) がよく用いられている．より深層なネットワークでは VGG[1]，ResNet[2]，Inception[3] などの画像認識にて初期に開発されたネットワークを用いたものもあるが，行動認識の特徴抽出に適したネットワーク構造は明らかではない．またセンサベースの行動認識の多くが，ウェアラブルデバイスから得られたデータを用いて行動の分類を行っている．モバイルデバイス上での動作を考慮すると,モデルは軽量であるほどよい．行動認識モデルの軽量化に関して，Liu ら [4] は CNN の冗長性を軽減する新たな畳み込み層の提案をしているが，行動認識以外にも適用可能な手法であり，行動認識の特徴を活かした軽量化手法についての議論は進んでいない．

既存の深層学習ベースの行動認識手法は画像認識分野の知見を流用しているものが多い．Tuncer ら [5] は ResNet を用いてアンサンブル学習を行動認識に適用している．Xu ら [6] は Inception と GRU を組み合わせたモデルを行動認識に適用している．Zhao ら [7] は画像認識にて使用されている深層学習モデルを網羅的に検証し，Inception ベースのモデルの有効性を示している．一方で，センサベースの行動認識で使用されるデータは，主に加速度や角速度などの一次元時系列波形データである．画像認識で提案された深層学習モデルをセンサデータに適用する場合，二次元畳み込み (2D-Conv) を一次元畳み込み (1D-Conv) に変更するといった手続きにより容易に実現可能である．しかし，画像認識にて使用されるデータは画像という二次元データであるため，画像認識にて使用される技術に変更を加えずにセンサベースの行動認識に流用しても，必ずしも有効に働くとは限らない．

本研究では特に畳み込み層のカーネルサイズに焦点を当てる．CNN は畳み込み処理により特徴抽出と認識を End-to-End で行うことができ，人の認識を超える特徴を獲得できる可能性がある．しかし，畳み込み演算に深く関係するカーネルサイズへの議論は不十分である．例えば画像認識では小さいカーネルを多段に積層することが主流である．入出力チャネル数が等しい場合，小さいカーネルの畳み込み層を多段に積層することで，大きいカーネルの畳み込み層と受容野の範囲を同等にしつつパラメータ数を削減可能であるためである．加えて，非線形変換を行う活性化関数を多く挿入可能であるため，ネットワークを深層化しつつモデルの表現力を高めることが可能である．一方，行動認識のような一次元波形データを扱う場合，同等の範囲の受容野では単層の大きいカーネルの畳み込み層のほうがパラメータ数を削減可能である (詳細は後述する)．また，小林ら[8] が Neural Architecture Search (NAS) を用いて行

1 福井大学大学院工学研究科
Graduate School of Engineering, University of Fukui
†1 現在，NEC ソリューションイノベータ株式会社
Presently with NEC Solution Innovators, Ltd.

動認識に特化した CNN のネットワーク構造を探索した結果，比較的大きなカーネルサイズが選択されたという事例がある．以上を踏まえると，一次元波形データを対象とする行動認識で同等の範囲の受容野を持つ畳み込み層を考えるとき，パラメータ数を削減する大きいカーネルと，非線形変換を多段化可能な小さいカーネルの双方のメリットが均衡する範囲が存在する可能性がある．

本研究では，深層学習モデルにおいて，シンプルかつ頻繁に使用される VGG 構造について，畳み込み層のカーネルサイズによる行動認識精度やパラメータ数への影響を調査する．これにより，センサベースの行動認識に効果的な CNN 構造を明らかにすることを目的とする．本研究の貢献は以下の通りである．

- 1D-Conv において，畳み込みの受容野の範囲を同等にした際に，多段に積層した小さいカーネルの畳み込み層より単層の大きいカーネルの畳み込み層のほうがパラメータ数を削減可能であることを理論的に示した．
- VGG 構造において，複数の小さいカーネルの畳み込み層を単層の大きいカーネルの畳み込み層に変換した際に，パラメータ数を削減しつつ行動認識精度が向上することを実験的に示し，要因を考察した．

2. 畳み込み層の受容野とパラメータ数

2.1 ConvBlock

本研究で取り扱う VGG 構造について説明する．VGG は画像認識分野にて提案されたモデル構造である．VGG の特徴抽出器は 5 つの ConvBlock によって構成されている．各 ConvBlock は複数積層した小さなカーネルの畳み込み層と単層の Max Pooling によって構成されている．出力チャネル数を増減させる場合には，ConvBlock の最初の畳み込み層にて出力チャネル数を増減させ，後の畳み込み層では出力チャネル数を増減させることなく畳み込み演算を行う．従来は単層の大きいカーネルの畳み込み層と，単層の Max Pooling をひとまとまりとして CNN が構成されていた．しかし，入出力チャネルが等しい場合，単層の大きいカーネルの畳み込み層を，多数積層した小さいカーネルの畳み込み層に変更することで，受容野の範囲を同等にしつつパラメータ数を減少させることが可能である．また，非線形変換を多数挿入可能であるため，ネットワークの表現力が向上する．ConvBlock は ResNet に使用される ResidualBlock に応用されるなど，後の CNN 発展の礎となっている．ConvBlock を改善することは最新の深層学習モデルの改善につながる可能性がある．

2.2 ConvBlock の受容野

CNN の有効性に関わる重要な因子の一つに受容野がある．畳み込み層はカーネルを用いて空間方向に局所的な演算を行うため，全結合層とは異なり，演算後の 1 つのパラメータは特定の範囲（受容野）の情報しか有していない．一般的に CNN では 3 や 5 などの小さいカーネルサイズを用いられることが多い．畳み込み層は多段に積層することで，深層では広い受容野を持つことが可能になる特徴がある．したがって，深層 CNN は局所的な特徴を大域的な特徴へと変換することが可能である．

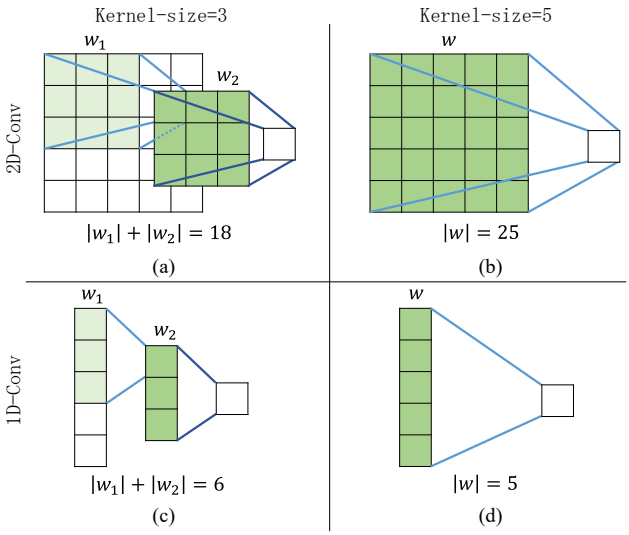


図 1：カーネルサイズが 3 と 5 の 2D, 1D-Conv の概要．wはパラメータを表す．

2D-Conv と 1D-Conv の受容野を数式で説明する．図 1 にカーネルサイズが 3 と 5 の 2D, 1D-Conv の概要を示す．2D-Conv の場合，縦横の受容野をそれぞれ求める必要がある．図 1(b)のように単層の 2D-Conv の場合は，カーネルサイズが受容野と同義であるため，カーネルサイズを k_{one} とすると受容野 (RF_{one}) は縦横それぞれ $RF_{one} = k_{one}$ となる．積層数が N の 2D-Conv の縦横の受容野 (RF_N) は，n 層目の畳み込み層のカーネルサイズを k_n とすると，$RF_N = k_1 + \sum_{n=2}^{N}(k_n - 1)$ で縦横それぞれ表せる．1D-Conv の場合，2D-Conv と同様にして表せる．$RF_{one} = RF_N$ のとき，k_{one} は以下のように表現可能である．

$$k_{one} = k_1 + \sum_{n=2}^{N}(k_n - 1) \tag{1}$$

畳み込み層の間に非線形変換を挿入しない場合，受容野の範囲が同等であるとき単層の畳み込み層と多数積層した畳み込み層は等価である．n 層目の畳み込み層の i 番目の入力とパラメータをそれぞれ $x_i^{(n)}$，$w_i^{(n)}$，カーネルサイズ k の CNN の出力を $z^{(KS:k)}$ とすると，図 1 (c) の場合，$z^{(KS:3)}$ は以下のように表せる．

$$z^{(KS:3)} = \sum_{l=1}^{3} w_l^{(2)} x_l^{(2)} \quad \left(\because x_i^{(2)} = \sum_{j=1}^{3} w_j^{(1)} x_{j+(i-1)}^{(1)} \right)$$

$$= \sum_{l=1}^{3} \left\{ w_l^{(2)} \sum_{j=1}^{3} w_j^{(1)} x_{j+(l-1)}^{(1)} \right\}$$

$$= \sum_{m=1}^{5} \hat{w}_m x_m^{(1)}$$

図 1 (d) の場合，$z^{(KS:5)}$ は以下のように表せる．

$$z^{(KS:5)} = \sum_{m=1}^{5} w_m x_m$$

$\hat{w}_m = w_m$の時，$z^{(KS:3)} = z^{(KS:5)}$となり，等価といえる．非線形関数を挿入する場合には等価とは言えず，非線形関数を挿入することで，表現力が向上する．

2.3 ConvBlock のパラメータ数

2D-Conv において同等の範囲の受容野を持つ CNN を比較した場合，図 1 上部に示すように，(3×3)×2=18 と (5×5)=25 となり，小さいカーネルサイズの畳み込み層を重ねることでパラメータ数を削減可能である．しかし，入出力チャネル数が 2 以上である場合を考えると，必ずしもパラメータ数を削減可能ではない．

1D-Conv の場合の VGG 構造の ConvBlock のパラメータ数を数式で説明する．入力チャネル数を C_{in}，出力チャネル数を C_{out}，積層数が N の ConvBlock を B_N，n 層目のカーネルサイズを k_n とすると，パラメータ数 $|B_N|$ は，

$$|B_N| = C_{out}\left(C_{in}k_1 + C_{out}\sum_{n=2}^{N} k_n\right)$$

と表せる．受容野の範囲が B_N と等価である単層の 1D-Conv のパラメータ数 $|B_{one}|$ は，式 (1) より，

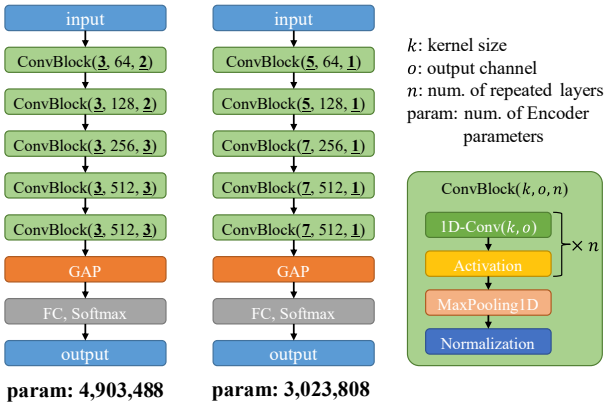
$$|B_{one}| = C_{in}C_{out}\left\{k_1 + \sum_{n=2}^{N}(k_n - 1)\right\}$$

と表せる．このとき，$|B_N| - |B_{one}|$ を考えると，

$$|B_N| - |B_{one}| = C_{out}\left\{C_{out}\sum_{n=2}^{N} k_n - C_{in}\sum_{n=2}^{N}(k_n - 1)\right\}$$

ここで，$N \geq 2$ より $\sum_{n=2}^{N} k_n > \sum_{n=2}^{N}(k_n - 1)$ であり，一般に $C_{out} \geq C_{in}$ であるため，$|B_N| - |B_{one}| > 0$ となり，単層の 1D-Conv のパラメータ数は，受容野の範囲が等価な ConvBlock のパラメータ数より少ない．したがって，1D-Conv では入出力チャネルに関わらず，単層の大きいカーネルの畳み込み層のほうがパラメータ数を削減可能である．

2D-Conv を用いる場合，出力チャネル数が入力チャネル数の倍程度より多ければ，B_N は B_{one} よりもパラメータ数を削減しつつ，非線形関数を多段に挿入可能である．ゆえに，多くの場合は単層の大きいカーネルの畳み込み層より小さいカーネルの畳み込み層を積層したものが採用される．深層学習モデルはパラメータ数が多いほどモデルの表現力が向上することが知られている．しかし，パラメータ数が増加することで深層学習モデルの訓練が困難になり，推論の計算コストも増加するため，多くの計算リソースが必要となる．人間行動認識ではプライバシ保護や推論用サーバが不要であることからスマートフォンなどのウェアラブル端末上での推論が選択されることがある．モバイルデバイス上で行動認識モデルを推論する場合，計算リソースが制限されるため，計算コストの小さいモデルが望まれる．



(a)VGG16-STD　(b)VGG16-CC　(c)ConvBlock

図 2：VGG16 の概要

2.4 ConvBlock の圧縮

VGG 構造を対象に，同等の範囲の受容野を持ちつつ，大きなカーネルを用いることでパラメータ数の削減を図った例を図 2 示す．特徴抽出器を 1D-Conv に変更したモデルを VGG16 Standard (VGG16-STD)，VGG16-STD の各 ConvBlock の受容野の範囲をそろえるように，カーネルサイズを大きくした単層の畳み込み層で置き換えたモデルを VGG16 Conv Compression (VGG16-CC) とする．VGG16-STD では最初の ConvBlock はカーネルサイズが 3 の 1D-Conv を 2 層積層しているが，VGG16-CC ではカーネルサイズが 5 の 1D-Conv 1 層に変更している．他の ConvBlock でも同様に変更を適用している．図 2 の (a), (b) の特徴抽出器のパラメータ数に着目すると，VGG16-CC は VGG16-STD と比較して約 38% パラメータ数を削減している．

VGG16-STD は，非線形変換を多数挿入可能であり，モデルの表現力が高いと考えられる．しかし，モデルの深層化とパラメータ数の増加により計算コストが大きくなる．VGG16-CC は VGG16-STD と比較して，非線形変換が少ないためモデルの表現力は落ちる．しかし，モデルが浅くパラメータ数が少ないため，計算コストが小さい．

3. 実験設定

3.1 データセット

データセットには HASC データセット [9] を用いる．HASC データセットはスマートフォンなどの端末を使用して，静止，歩行，走行，スキップ，階段上り，階段下りの 6 種類の行動が記録されているデータセットである．本研究では，サンプリング周波数 100Hz で iOS デバイスによって収集された 180 名の加速度データを使用する．各計測データから，前処理としてウィンドウサイズ 256 サンプ

表 1：全探索の結果

(a) Adam BN

LR	Opt	Norm	MT	Accuracy
1×10^{-2}	Adam	BN	STD	0.3898
1×10^{-3}	Adam	BN	STD	0.8859
1×10^{-4}	Adam	BN	STD	0.8595
1×10^{-5}	Adam	BN	STD	0.8552
1×10^{-6}	Adam	BN	STD	0.8417
1×10^{-2}	Adam	BN	CC	0.8202
1×10^{-3}	Adam	BN	CC	0.8781
1×10^{-4}	Adam	BN	CC	**0.8985**
1×10^{-5}	Adam	BN	CC	0.8585
1×10^{-6}	Adam	BN	CC	0.8017

(b) SGD BN

LR	Opt	Norm	MT	Accuracy
1×10^{-2}	SGD	BN	STD	0.8659
1×10^{-3}	SGD	BN	STD	0.8573
1×10^{-4}	SGD	BN	STD	0.6414
1×10^{-5}	SGD	BN	STD	0.2225
1×10^{-6}	SGD	BN	STD	0.2383
1×10^{-2}	SGD	BN	CC	**0.8752**
1×10^{-3}	SGD	BN	CC	0.8396
1×10^{-4}	SGD	BN	CC	0.6157
1×10^{-5}	SGD	BN	CC	0.3830
1×10^{-6}	SGD	BN	CC	0.2953

(c) RMSprop BN

LR	Opt	Norm	MT	Accuracy
1×10^{-2}	RMSprop	BN	STD	0.1763
1×10^{-3}	RMSprop	BN	STD	0.7535
1×10^{-4}	RMSprop	BN	STD	0.8707
1×10^{-5}	RMSprop	BN	STD	0.8449
1×10^{-6}	RMSprop	BN	STD	0.8349
1×10^{-2}	RMSprop	BN	CC	0.4982
1×10^{-3}	RMSprop	BN	CC	0.8698
1×10^{-4}	RMSprop	BN	CC	**0.8741**
1×10^{-5}	RMSprop	BN	CC	0.8583
1×10^{-6}	RMSprop	BN	CC	0.7898

(d) Adam LN

LR	Opt	Norm	MT	Accuracy
1×10^{-2}	Adam	LN	STD	0.2408
1×10^{-3}	Adam	LN	STD	0.8898
1×10^{-4}	Adam	LN	STD	0.9057
1×10^{-5}	Adam	LN	STD	0.8987
1×10^{-6}	Adam	LN	STD	0.8675
1×10^{-2}	Adam	LN	CC	0.5172
1×10^{-3}	Adam	LN	CC	0.8935
1×10^{-4}	Adam	LN	CC	0.9095
1×10^{-5}	Adam	LN	CC	**0.9102**
1×10^{-6}	Adam	LN	CC	0.8606

(e) SGD LN

LR	Opt	Norm	MT	Accuracy
1×10^{-2}	SGD	LN	STD	0.9034
1×10^{-3}	SGD	LN	STD	0.8843
1×10^{-4}	SGD	LN	STD	0.8071
1×10^{-5}	SGD	LN	STD	0.4413
1×10^{-6}	SGD	LN	STD	0.2392
1×10^{-2}	SGD	LN	CC	**0.9166**
1×10^{-3}	SGD	LN	CC	0.8841
1×10^{-4}	SGD	LN	CC	0.7496
1×10^{-5}	SGD	LN	CC	0.5517
1×10^{-6}	SGD	LN	CC	0.2289

(f) RMSprop LN

LR	Opt	Norm	MT	Accuracy
1×10^{-2}	RMSprop	LN	STD	0.1763
1×10^{-3}	RMSprop	LN	STD	0.7057
1×10^{-4}	RMSprop	LN	STD	0.8913
1×10^{-5}	RMSprop	LN	STD	0.9045
1×10^{-6}	RMSprop	LN	STD	0.8715
1×10^{-2}	RMSprop	LN	CC	0.2878
1×10^{-3}	RMSprop	LN	CC	0.8763
1×10^{-4}	RMSprop	LN	CC	0.9032
1×10^{-5}	RMSprop	LN	CC	**0.9162**
1×10^{-6}	RMSprop	LN	CC	0.8673

ル，ストライド 256 サンプルで時系列分割を行う．計測開始からデバイスの格納動作などの影響を取り除くために前後 5 秒をトリミングして除去している．

3.2 実験に使用する深層学習モデル

本研究では VGG 構造のなかでも頻繁に使用される VGG16 を対象として検証を行う．本来の VGG16 の分類器は 3 層の全結合層からなるが，本研究では畳み込み層のカーネルサイズによる影響を調査するため，分類器は 1 層の全結合層のみに変更する．また，特徴抽出器の出力を Global Average Pooling (GAP) により圧縮して分類器へ入力している．また，予備実験より Normalization を挿入することにより学習が安定化したため，挿入する．

3.3 共通する設定

180 名のデータを 36 名ずつ 5 つに分割し，5 交差検証を行う. 評価は 5 交差検証の Accuracy の平均値により行う．予備実験より，モデルの活性化関数は負の範囲の傾きが 0.01 の Leaky ReLU を使用した. 各畳み込み層の後には確率 0.1 の Dropout を挿入した．また，各畳み込み層のバイアス項は使用していない．ミニバッチサイズは 1024 であり，350 エポック訓練を行う．訓練時のデータ拡張として flipping と channel shuffling を適用した．

4. 実験・結果・考察

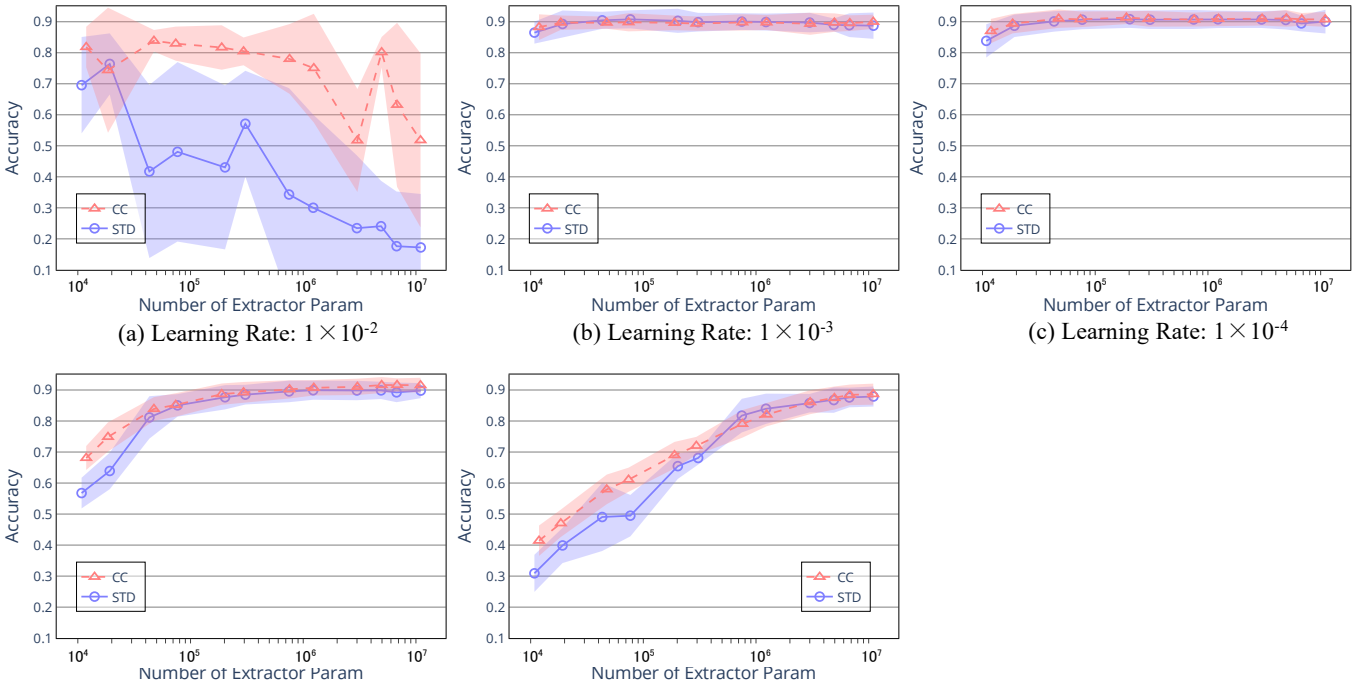
4.1 全探索

VGG16-STD と VGG16-CC がどの程度のポテンシャルを有しているかを検証するため，Learning Rate (LR)，Optimizer (Opt)，Normalization (Norm)，Model Type (MT)の全探索を行った．表 1 に全探索の結果を示す．各項目の探索範囲は以下の通りである．

- LR　：1×10^{-2}～1×10^{-6} (10^{-1} 倍ずつ 6 種)
- Opt　：Adam, SGD, RMSprop
- Norm：Batch Norm (BN), Layer Norm (LN)
- MT　：VGG16-STD (STD), VGG16-CC (CC)

Hammerla ら [10] は CNN を用いたセンサベースの行動認識においてハイパーパラメータが行動認識精度に与える影響を調査した結果，LR などの学習過程をコントロールするパラメータが最も行動認識精度への影響が高い可能性があると示した．したがって，本研究では広範囲の LR と深層学習において広く使用される Opt を選択した. 学習の安定化を図るために挿入した Norm においても同様に頻繁に使用される BN，LN を選択した．

表 1 より全探索の結果，LR : 1×10^{-2}，Opt : SGD，Norm : LN，MT : CC が 0.9166 で最高精度であることがわかる．表 2 上部と下部で比較すると，BN より LN を使用したほうが高精度であるとわかる．BN はミニバッチ内でチャネルごとに正規化を行う関数であり，LN はサンプルごとに正規化を行う関数である．センサデータセットは測定する機器によってデータの範囲などが異なる場合が多いため，サンプルごとに正規化を行うことで，データの規格がそろい, 精度が向上したと考えられる. Opt に着目すると, Adam, SGD，RMSprop では最高精度の差は小さく，精度への影響は小さいと考えられる．LR に着目すると，1×10^{-2} と $1\times$



(a) Learning Rate: 1×10^{-2} (b) Learning Rate: 1×10^{-3} (c) Learning Rate: 1×10^{-4}

(d) Learning Rate: 1×10^{-5} (e) Learning Rate: 1×10^{-6}

図 3：パラメータ数を変更した場合の結果．網掛けは 5 試行の標準偏差の範囲を表す．

10^{-6} において低精度であることがわかる．LR が高すぎると，最適解を飛び越えやすくなり収束しづらく，低すぎると局所解に陥るなどして最適解にたどり着きづらくなるためだと考えられる．MT に着目すると，おおむね VGG16-CC のほうが高精度であることがわかる．

以降の実験には Opt：Adam, Norm：LN を使用する．Adam は多くの研究にて使用されており，最高精度との差が小さく，複数の LR において高精度を達成しているため選択した．LN は高精度を達成可能であるため，選択した．

4.2 パラメータ数による影響

図 3 に横軸を特徴抽出器のパラメータ数，縦軸を精度とした，各 LR のパラメータ数を変更した場合の結果を示す．図 3 の網掛けは 5 試行の標準偏差を表している．モデルの構造を大きく変化させずにパラメータ数を変化させるため，畳み込み層の出力チャネル数を変更して検証を行う．深層学習モデルの性能はパラメータ数に関連があるとされ，一般にパラメータ数が多いモデルほど表現力が高いと知られている．同時に，パラメータ数が過剰であると学習が困難になりやすくなることも知られている．したがって，VGG16-STD の過剰なパラメータ数による精度低下が相対的に VGG16-CC の優位性につながっている可能性がある．本実験は恣意的に VGG16-CC が有利な条件で評価をおこなっていないことを検証するため，モデルの構造をそのままにパラメータ数を変化させ，行動認識精度への影響の調査を行う．

図 3 より，LR が 1×10^{-3} 以外の場合は，VGG16-STD より，VGG16-CC のほうがおおむね高精度であることがわかる．また，LR が 1×10^{-3} であっても精度にほとんど差は見られないことがわかる．したがって，VGG16-STD の精度低下は，過剰なパラメータ数による学習の難化によるものではないと考えられる．よって，VGG16-CC の精度向上はパラメータ数による影響ではないと考えられる．図 3 (b), (c)より，モデルパラメータ数が 5×10^{4} 程度で精度が頭打ちになり，モデルのパラメータ数を増加させても精度低下につながらないことがわかる．したがって，行動認識ではパラメータ数が少ないモデルでも十分な認識精度を得られる可能性がある．近年画像認識にて提案されている深層学習モデルはパラメータ数が膨大であるため，モバイルデバイス上での動作を考慮すると行動認識に向かないが．入出力チャネル数を変更するなどのモデル構造に大きな変更を加えずに，シンプルな変更によってモデルを軽量化しても，行動認識精度に悪影響を及ぼさない可能性があると考えられる．低パラメータ数の場合，VGG16-CC は VGG16-STD と比較して高精度であることがわかる．これより，VGG16-CC は少ないパラメータを効率的に使用できていると考えられる．図 3 (a) より LR が 1×10^{-2} の場合，パラメータ数が少ないモデルほど精度が向上していることがわかる．一般に高い LR の場合は，学習が不安定になる傾向があるが，パラメータ数が減少したことで，学習が容易になり精度が向上したと考えられる．図 3 (e) より，LR が 1×10^{-6} の場合，パラメータ数が多いモデルほど精度が向上していることがわかる．学習曲線を確認した結果，パラメータ数が多いモデルほど少ないエポック数で損失関数の値が収束していることが判明した．これはパラメータ数が多いモデルほど表現力が高いため，低い LR でも迅速に訓練データに適合するためであると考えられる．

表 2：Ablation Study

Model Type	Shallow Layer	Large Kernel	Encoder Params	Encoder FLOPs	Learning Rate				
					1×10^{-2}	1×10^{-3}	1×10^{-4}	1×10^{-5}	1×10^{-6}
STD			4.90M	144.85M	0.2408±0.1464	0.8898±0.0207	0.9057±0.0306	0.8987±0.0279	0.8675±0.0410
shallow	○		1.30M	34.75M	**0.8164**±0.0450	0.8838±0.0236	0.9038±0.0236	0.9040±0.0234	0.8082±0.0361
large		○	8.12M	241.42M	0.1723±0.0068	0.8912±0.0329	0.9022±0.0196	0.9024±0.0263	**0.8807**±0.0370
CC	○	○	3.02M	78.89M	0.5172±0.1660	**0.8935**±0.0356	**0.9095**±0.0218	**0.9102**±0.0263	0.8606±0.0386

4.3 Ablation Study

表 2 に浅層化とカーネル拡大についての Ablation Study の 5 試行平均精度と標準偏差を示す．Encoder Params は特徴抽出器のパラメータ数を表し，Encoder FLOPs は特徴抽出器の理論上の計算量を表す．VGG16-CC は VGG16-STD を浅層化し，大きいカーネルの畳み込み層に変更したモデルといえる．VGG16-CC の精度向上の要因がモデルの浅層化によるものか，カーネル拡大によるものかを検証するため，Ablation Study を行う．VGG16-STD の各 ConvBlock を単層の 1D-Conv に変更することを Shallow Layer とする．1D-Conv のカーネルサイズを 3 より大きくすることを Large Kernel とする．今回の実験では large の各畳み込み層のカーネルサイズはすべて 5 にしている．

表 2 より，VGG16-CC がほとんどの LR において最高精度であることがわかる．また，shallow，large は VGG16-STD と同程度の精度であることがわかる．したがって，VGG16-CC の精度向上は Shallow Layer と Large Kernel の 2 つの要素を同時に適用することによって生じると考えられる．したがって，Shallow Layer と Large Kernel の適切な組み合わせが存在する可能性がある．shallow は VGG16-STD と比較して，高い LR の場合高精度であることがわかる．これは図 3 (a) からもわかるように低パラメータ数による効果だと考えられる．また，shallow は VGG16-STD と比較してほぼすべての LR において標準偏差が低いことがわかる．これは shallow の学習が VGG16-STD よりも安定していることを示しており，Shallow Layer による効果だと考えられる．LR が 1×10^{-6} の場合に着目すると，shallow，VGG16-CC，VGG16-STD，large の順に精度が向上していることがわかる．各モデルの ConvBlock の受容野の範囲が shallow < VGG16-CC = VGG16-STD < large の順であることから，受容野が広いほど低い LR において高精度であると考えられる．パラメータ数と FLOPs に着目すると，VGG16-CC は shallow に次いで軽量なモデルであるとわかる．したがって VGG16-CC は高精度かつ，軽量なモデルであるため，4 つのモデルの中で最も行動認識に適したモデル構造をしているといえる．

5. まとめ

本研究では，VGG 構造を対象として，1D-Conv のカーネルサイズを変化させたときの行動認識精度やパラメータ数への影響を調査した．積層した 1D-Conv を，受容野の範囲が同等となるような，単層の大きいカーネルサイズの 1D-Conv に変更することにより行動認識精度が向上し，デフォルトの VGG 構造と比較して，パラメータ数を約 38% 軽減することが可能であることが判明した．センサベースの行動認識においては，積層した小さいカーネルの 1D-Conv を，単層の大きいカーネルの 1D-Conv に置き換えるというシンプルな変更で，モデルのパラメータ数を軽減しつつ，行動認識精度が向上する可能性があることが判明した．しかし，カーネルサイズの大きさと，モデルの深さには適切な組み合わせでないと精度向上につながらない可能性があることが示唆された．今後は VGG 構造以外の著名なモデル構造にも適用可能か検証していく．

謝辞

本研究の一部は，JSPS 科学研究費助成事業若手研究(19K20420)の助成によるものである．ここに謝意を表す．

参考文献

[1] Simonyan, K., and Zisserman, A.: Very deep convolutional networks for large-scale image recognition, Proc. *international conference on learning representations* (2015).
[2] He, K., Zhang, X, Ren, S, et al.: Deep Residual Learning for Image Recognition, Proc. *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778 (2016).
[3] Szegedy, C., Liu, W., Jia, Y., et al.: Going deeper with convolutions, Proc. *2015 IEEE Conference on Computer Vision and Pattern Recognition*, pp.1-9 (2015).
[4] Liu, T., Wang, S., Liu, Y. et al.: A lightweight neural network framework using linear grouped convolution for human activity recognition on mobile devices, Proc. *J Supercomput*, Vol.78, pp.6696–6716 (2022).
[5] Tuncer, T., Ertam, F., Dogan, S., et al.: Ensemble residual network-based gender and activity recognition method with signals, *J Supercomput*, Vol.76, pp.2119–2138 (2020).
[6] Xu, C., Chai, D., Zhang, X., et al.: InnoHAR: A Deep Neural Network for Complex Human Activity Recognition, *IEEE Access*, Vol.7, pp.9893-9902 (2019).
[7] Zhao, Z., Kobayashi, S., Kondo, K., et al.: A Comparative Study: Toward an Effective Convolutional Neural Network Architecture for Sensor-Based Human Activity Recognition, *IEEE Access*, Vol.10, pp.20547-20558 (2022).
[8] 小林慧，長谷川達人：Mobile-aware Convolutional Neural Network for Sensor-based Human Activity Recognition，情報処理学会第 84 回全国大会（2022）．
[9] Kawaguchi, N., Ogawa, N., Iwasaki, Y., et al.: HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-World Activity Understandings, Proc. *the 2nd Augmented Human International Conference* (2011)
[10] Hammerla, N.Y., Halloran, S. and Plötz, T.: Deep, convolutional, and recurrent models for human activity recognition using wearables, Proc. *the Twenty-Fifth International Joint Conference on Artificial Intelligence*, pp.1533–1540 (2016).