チャネル間関係を考慮した ResNetLSTM による行動識別

榎堀 優^{1,a)} 間瀬 健二^{1,b)}

概要:行動識別においても深層学習による研究が盛んに行われている.本稿では、チャネル間関係を考慮した ResNetLSTM による行動認識を提案する.提案するネットワークは、DeepConvLSTM のコンボリューション層を ResNet に置き換え、ResNet 層および LSTM 層のそれぞれにおいてチャネル間の関係を考慮する拡張を施した物である.HASC コーパスに収録されている静止、ジョギング、走行、スキップ、階段の昇り・下りの6 種類の行動データを用いて検証した結果、全ての拡張を導入した DNN は、行動データのサンプル数によって補正した Weighted F¹-score において、同一被験者の他の試行を学習データとして含む交差検証において 99.1、単一のデータ収集グループ内における1人抜き交差検証において 89.7、複数のデータ収集グループを含む交差検証において 87.5 の精度を示し、DeepConvLSTM を用いた場合からそれぞれ 0.7%、8.5%、6.7% 向上した.

A Study of Activity Recognition Classification with ResNetLSTM including Inter-channel Relationship Considerations

Yu Enokibori^{1,a)} Kenji Mase^{1,b)}

1. はじめに

行動識別においても深層学習 (Deep Neural Network (DNN)) を用いた研究が盛んに行われている。特に,学習 に Long Short Term Memory (LSTM, [1]) を用いた識別機 が成果を上げている [2]. また,画像認識の分野で成果を上 げているコンボリューション層を用いた DNN[3], [4] を行 動認識に応用する研究 [5] もなされており, LSTM と組み 合わせた DeepConvLSTM[6] も提案されている.

しかし, DeepConvLSTM は多数のセンサを用いたデー タセットにおいて評価されていたためか,単一の3軸加 速度センサのみを用いた行動認識へ適用してみたところ, 期待した精度が得られなかった.そこで本研究では Deep-ConvLSTM のネットワーク構造において表現が不足して いると考えられる「センサチャネル間の関係性」および「上 位層の出力チャネル間の関係性」を考慮する拡張を加える ことによって精度の向上を試みた.また, DeepConvLSTM

 1 名古屋大学情報学研究科 Graduate School of Informatics, Nagoya University

^{a)} enokibori@i.nagoya-u.ac.jp ^{b)} mase@nagoya-u.jp で用いている4層のコンボリューション層を,画像認識 で成果を上げてるResidual Network (以下 ResNet [4]) に 置き換えることによる精度の向上も試みた.上記3つの 拡張の効果をHASCコーパス [7] に収録されている静止, 歩行,ジョギング,スキップ,階段の昇り・下りの6種類 の行動データを用いて検証した.結果,全ての拡張を導入 した DNN は,行動データのサンプル数によって補正した Weighted F¹-score において,同一被験者の他の試行を学 習データとして含む交差検証において 99.1,単一のデータ 収集グループ内における1人抜き交差検証において 87.5の 精度を示し,DeepConvLSTMを用いた場合からそれぞれ 0.7%, 8.5%, 6.7% 向上した.

2. 関連研究

LSTM を用いて行動認識を行った研究としては Inoue ら の研究がある [2].本研究は、本稿と同様に HASC コーパ スを用いて評価を行っており、7名の540トライアルを432 の学習データ、108の評価データに分割した場合のベスト モデルで95.8%の精度を達成している.





コンボリューション層を用いた行動認識用 DNN として は Zeng らの研究がある [5].本研究の DNN では、3 軸加 速度センサの x, y, z 軸に個別にコンボリューション層を 適用し、全結合層へ繋げている.Skoda[8]を用いた評価 で 88.19%, Opportunity[9]を用いた評価で 76.83%, Actitracker[10]を用いた評価で 96.88%の精度を達成している. この中で、本稿で用いた HASC コーパスに最も近いのは Actitracker である.取得している行動も "Jog", "Walk", "Up stair", "Down stair", "Stand", "Sit" と Sit 以外は 共通している.36名のデータの単純 10分割交差検証で精 度を評価しているため、上記の結果は4章で述べる評価 1 に相当すると考えられる.

コンボリューション層とLSTM を併用した行動認識用 DNN としては、Ordóñez らの研究がある [6]. 本研究で 提案された DeepConvLSTM は、Opportunity データセッ トにおいて、行動データのサンプル数によって補正した Weighted F¹-score で 91.5 の高い精度を示した. しかしな がら、DeepConvLSTM は多数のセンサを用いたデータセッ トにおいて評価されていたためか、単一の3軸加速度セン サのみを用いた行動認識へ適用してみたところ、期待した 精度が得られなかった. 詳しい数値は4章にて示す.

図1に DeepConvLSTM の DNN 構造を文献 [6] から引 用する.図1から分かるように,DeepConvLSTM では時 間軸方向のたたみ込みのみを行っておりセンサチャネル間 の関係は LSTM 層まで考慮されていない.また,上位層で あるコンボリューション層の出力チャネル間の関係も十分 に考慮されていない.また,上位層に用いられているネッ トワークは単純な4層コンボリューション層であり,画像 認識で用いられるネットワークとしては古いものである. 本稿ではこれらの点について拡張の余地があると考えた. 詳しくは次節以降で述べる.

3. チャネル間関係を考慮した ResNetLSTM

前章で述べたように DeepConvLSTM は拡張の余地があ

る.本稿では「拡張 1: 上位層の ResNet 化」「拡張 2: セン サチャネル間関係の考慮」「拡張 3: 上位層の出力チャネル 間関係の考慮」の 3 つの拡張を提案する.本稿で提案する 拡張を全て実装した DNN の構造を図 2 に示す.以下,そ れぞれについて詳しく述べる.

3.1 拡張 1: 上位層の ResNet 化

画像認識の分野で高い精度を示したネットワーク構造 に Residual Net (以下, ResNet) がある. DeepConvLSTM の上位層は単純な4層のコンボリューション層であり, ResNet 化することで性能向上が期待できる. 図 2 では, DeepConvLSTM の4段のコンボリューション層を1層飛 びの ResNet に置き換えている. 図 2 中の L*-1 及び L*-3 である. たたみ込みカーネルは DeapConvLSTM と同じく $n \times 1$ の形とした.

3.2 拡張 2: センサチャネル間関係の考慮

DeepConvLSTM ではセンサチャネル間の関係は後段の LSTM 層で学習されている.コンボリューション層でも2 次元 (例えば3×3)のたたみ込みを行えば、センサチャネ ル間の関係性も学習できると考えられるが、3チャネルし かない加速度センサなどでは、すぐに1×1の1次元まで たたみ込まれ、ネットワークが維持できなくなる.そこで 我々は、1×m (m は導出元の層の列数)のコンボリュー ション層を別途に設けてセンサチャネル間の関係を表現 し、ResNetの層へ連結する形で表現した.図2中のL*-2 及びL*-4である.本結合も ResNet の実装と合わせ、1層 飛びの結合とした.

3.3 拡張 3: 上位層の出力チャネル間の関係性を表現する 拡張

DeepConvLSTM では上位層であるコンボリューション 層の出力を LSTM 層へ直接入力していた。例えば、セン サチャネル数が 3, たたみ込みカーネル数が 64, 時系列方



Fig. 2 ResNetLSTM with consideration of inter channel relationships.

向のサンプル数が 30 の場合, 3×64 のデータを 30 回入力 する形となる. しかし, この場合, 時刻 t におけるコンボ リューション層の第 n カーネルの出力と, 時刻 q における コンボリューション層の第 m カーネルの出力との関係は, 明示的に学習されない. そこで本提案では, 上位層の出力 チャネルを一列に連結する形で LSTM 層へ入力すること で,上記のような関係を明示的に LSTM 層で学習できる ように拡張した. 図 2 中の L15 である. これにより, 例え ば,時刻 t における第 n カーネルに存在して欲しい出力が 時刻 t+1 の第 n カーネルの出力や時刻 q の第 m カーネル に出現しても LSTM 層によって学習が可能となる.

3.4 提案ネットワークにおけるマジックナンバーについて 図 2 に示したネットワークにおいて「時間軸方向へのた たみ込みカーネルのサイズ」「ResNet の段数」「ResNet に おいて何層前の出力と合成するか」などは事前実験により 探索的に精度の良くなる数値を求めた.結果,時間軸方向 のたたみ込みカーネルのサイズは 7×1, ResNet の段数は L2 から L13 の 12 段, ResNet において合成する出力は最 初段を除いて1段飛びに2層前の出力を用いることとした.

4. HASC コーパスを用いた各拡張の評価

本節では提案の拡張がどのように精度に影響するかについて, HASC コーパスを用いて検証した結果について述べる.本研究では以下の3パターンに分けて検証した.

- (1) 学習・評価データに同一被験者の試行が含まれる場合
- (2) 学習・評価データが同一収集グループである場合

(3)学習・データに複数の収集グループが含まれる場合 検証1から、同一被験者の試行間の動作の違いに影響されることなく行動が識別できるかが分かる。検証2から、 被験者間の動作の違いや、センサの装着の差異に影響され ることなく行動が識別できるかが分かる。検証3から、収 集環境や利用したセンサ、試行のコントロールの差異など に影響されることなく行動が識別できるかが分かる。検証 1 > 2 > 3 の順で学習データと評価データの間の差異、母 集団の複雑さが増加する。

以下,まず HASC コーパスの概要を述べた後に,全検証 に共通するデータの前処理やその他の関連事項について述 べ,その後,上記の3つの検証結果について詳しく述べる.

4.1 HASC ⊐−パス

HASC コーパスには、静止・歩行・ジョギング・スキッ プ・階段上り・階段下りの6行動のデータが含まれている。 複数の団体が、それぞれ異なるセンサ、異なるサンプリン グレート、異なる場所、異なる条件下で収集したデータが 混在しており、多様性に富むデータセットとなっている。 本稿では HASC2010 コーパス、および、HASC2014 コー パスからデータを選出して用いた。選出の詳細は 4.2.2 に て述べる.

4.2 全検証に共通する事項

本節では細かな共通事項について述べる.

4.2.1 DNN 構造の呼称

本評価では,以下に示す6つのDNN構造の精度を比較し,本提案による拡張効果について検証した.本稿では,

- それぞれの DNN 構造ついて, 呼称を以下のとおりとする. (1) DeepConvLSTM: [6] で提案された DNN
- (2) ResNetLSTM: (1) の4層のコンボリューション層を ResNet 構造に置換した DNN (拡張1:図2のL*-1及 びL*-3のみを導入した構造)
- (3) ResNetLSTM+: (2) へ 3.2 節で示した拡張 2 を導入した DNN (拡張 1+2: 図 2 の L14 まで実装し, 2 層の 128 カーネル LSTM へ接続した構造)
- (4) DeepConvLSTM-: (1) の LSTM 層へ 3.3 節で示した 拡張 3 を導入した DNN (拡張 3: 4 層のコンボリュー ション層の後に図 2 の L15 以降を接続した構造)
- (5) ResNetLSTM-: (2) へ 3.3 節で示した拡張 3 を導入した DNN (拡張 1+3:図 2 からの L*-2 及び L*-4 を取り除いた構造)
- (6) ResNetLSTM±: 全ての拡張を導入した DNN (拡張 1+2+3: 図 2 に示した構造)
- 4.2.2 データの前処理

検証に用いたデータセットは HASC2010 コーパスと HASC2014 コーパスの 3 軸加速度センサデータである. 50Hz 以上のサンプリングレートがあり, 腰に固定されたセ ンサのみを対象とし, 50Hz を超えるセンサデータは 50Hz までサブサンプリングした. DNN へ投入するデータは, 1 秒のスライディングウィンドウ (50 samples) と1 sample の移動幅で抜き出し, 次に示す前処理を適用した.

まず,加速度センサの装着方向と動作の方向は被験者毎 に異なると考え,n=3のPCAにより主動作軸を特定し, 3次元の回転を行った.PCAの結果により右手系と左手 系の入れ替わりが発生することがあるが,これらは個別に 元の系へ修正した.続いて,スライディングウィンドウ内 のデータの最大値を1,最小値を0となるように正規化し た.最後に,センサ出力値を128段階に離散化した.これ は事前実験において,センサ出力値をそのまま用いるより も離散化した方が精度が向上したためである.事前実験で は64,128,256段階の離散化を検討し,本稿ではもっと精 度が向上した128段階の離散化を採用した.

4.2.3 学習における Dropout 層と BatchNormalization 層の導入

本評価における学習では,最終の出力層の直前 (図 2 であ れば L17 と L18 の間) に 50%の Dropout 層 [11] と Batch-Normalization 層 [12] を導入して,過学習の抑制と性能の 向上を図った.以後の議論における学習では,全て上記 2 層が導入されている.なお、DeapConvLSTM において上 記2層を導入した結果においても、評価1における交差検 証の平均精度が0.937±0.047から0.983±0.016へ向上し ており、提案手法にのみ有利に働く事項ではない.

4.2.4 学習結果の選択

本稿の評価では交差検証を用いている.交差検証に含ま れる各試行ではそれぞれ10回の学習を実施し、その中で 最も検証ロスが少なくなったエポックの学習結果を、その 試行における学習結果として採択し、評価に用いた.

4.2.5 Weighted F¹-score

用いたデータセットに含まれる行動は、行動毎にサンプ ル数が異なる.そのため、例えば、最もサンプル数の多い 行動を優先的に提示するような識別機は、本来の性能より も不当に高い精度が示される可能性がある.そこで、行動 別に F^{1} -score (= $2 \frac{\text{precision-recall}}{\text{precision} + \text{recall}}$, $\text{precision} = \frac{\text{TP}}{\text{TP}+\text{FP}}$, $\text{recall} = \frac{\text{TP}}{\text{TP}+\text{FN}}$)を導出し、各行動のサンプル数によって 補正を掛けた Weighted F^{1} -score にを評価に用いた.以下 に Weighted F^{1} -score の定義を示す.

Weighted
$$F^1$$
-score $(F_t^1) = \sum_i w_i F_i^1$
 $w_i = n_i/N$
where *i*: i-th class,

 \mathbf{F}_i^1 : \mathbf{F}^1 -score of i-th class,

 $n_i:$ sample size of i-th class,

N: total size of samples.

4.3 検証 1: 学習・評価データに同一被験者の試行が含まれる場合

本検証では、同一被験者の試行間の動作の違いに影響されることなく行動が識別できるかどうかを、学習データと 評価データに同一被験者の試行が含まれる場合の学習結果 を用いて議論する.なお文献 [6] における評価は、本検証 に相当する.

本検証では、HASC2014 コーパスに含まれる person1001 から person1005 の5名のデータを用いた。4名のデータ を学習データとし、1名のデータの第1・2 試行を学習デー タに、第3試行を検証データに、第4・5試行を評価に用い て交差検証を行った。従って、交差検証の総検証数は5回 である。結果を表1に示す。

本検証結果では全 DNN 構造において 100%に迫る高い 精度が得られた. どの DNN 構造においても同一被験者の 試行間の動作の違いに影響されることなく行動が識別で きていると考えられる. 微差ではあるが最も高い精度を 示したのは ResNetLSTM+であり, Weighted F¹-score で 0.993 を示した. 全ての拡張を導入した ResNetLSTM± も 0.991 と同程度の精度を示している.

ResNetLSTM は DeepConvLSTM と同程度の結果であ

表 1 検証 1 の結果 Table 1 Result of evaluation 1.

| | CV 平均精度 ± 標準偏差 | Weighted F^1 |
|------------------------------|-------------------|----------------|
| DeepConvLSTM | 0.983 ± 0.016 | 0.984 |
| ResNetLSTM | 0.983 ± 0.013 | 0.984 |
| $\operatorname{ResNetLSTM}+$ | 0.993 ± 0.007 | 0.993 |
| DeepConvLSTM- | 0.973 ± 0.021 | 0.974 |
| ResNetLSTM- | 0.988 ± 0.010 | 0.988 |
| ${\rm ResNetLSTM}\pm$ | 0.991 ± 0.010 | 0.991 |

表 2 検証 2 の結果 Table 2 Result of evaluation 2.

| | CV 平均精度 ± 標準偏差 | Weighted F^1 |
|------------------------------|-------------------|----------------|
| DeepConvLSTM | 0.797 ± 0.175 | 0.827 |
| ResNetLSTM | 0.833 ± 0.151 | 0.878 |
| $\operatorname{ResNetLSTM}+$ | 0.872 ± 0.110 | 0.905 |
| DeepConvLSTM- | 0.805 ± 0.114 | 0.826 |
| ResNetLSTM- | 0.858 ± 0.126 | 0.896 |
| ${\rm ResNetLSTM}\pm$ | 0.858 ± 0.127 | 0.897 |

り, 拡張1による精度向上は見受けられなかった. ResNetL-STM で精度が向上せず, ResNetLSTM+で高い精度が出 たため, 3.2 節で示したセンサチャネル間関係を考慮する 拡張2が一定の効果が示したものと考えられる. 一方で, 3.3 節で示した上位層の出力チャネル間関係を考慮する 拡張3を加えた DeepConvLSTM-は精度が低下していお り, 拡張の悪影響が出ている可能性が示唆された. 一方で, ResNetLSTM に拡張3を導入した ResNetLSTM-の精度 は向上しており, 上位層の出力チャネル間の関係を考慮す る拡張は一概に悪影響のみがあるとは言えないことも分 かった.

4.4 検証 2: 学習・評価データが同一収集グループである 場合

本検証では、被験者間の動作の違いやセンサの装着の差 異に影響されることなく行動が識別できるかどうかを、学 習データと評価データが、同一収集グループである場合の 学習結果を用いて議論する.

本検証では、前節に引き続き、HASC2014 コーパスに含 まれる person1001 から person1005 の 5 名のデータを用い た.4名のデータを学習データとし、1 名のデータの第1・ 2・3 試行を検証データに、第4・5 試行を評価に用いて交 差検証を行った.従って、総検証数は5 回である.結果を 表2に示す.

微差ではあるが最も高い精度を示したのは検証1と同様 に ResNetLSTM+であり, Weighted F¹-score で 0.905 を 示した. 全ての拡張を導入した ResNetLSTM± ならびに ResNetLSTM-も 0.897 と 0.896 と同程度の精度を示して いる.

ResNetLSTM は Weighted F¹-score で DeepConvLSTM

| | CV 平均精度 ± 標準偏差 | Weighted \mathbf{F}^1 |
|-------------------------------|-------------------|-------------------------|
| DeepConvLSTM | 0.822 ± 0.023 | 0.820 |
| $\operatorname{ResNetLSTM}$ | 0.846 ± 0.026 | 0.849 |
| $\operatorname{ResNetLSTM}+$ | 0.854 ± 0.028 | 0.854 |
| DeepConvLSTM- | 0.835 ± 0.023 | 0.836 |
| $\operatorname{ResNetLSTM}$ - | 0.873 ± 0.033 | 0.872 |
| ${\rm ResNetLSTM}\pm$ | 0.876 ± 0.030 | 0.875 |

より高い0.878 を示しており, ResNet による拡張が機能し た結果と言える.検証1において悪影響を示した拡張3に ついては, DeepConvLSTM-において0.826とDeepConvLSTMと同等の精度を示しており,本検証では悪影響は 見られなかった.逆に拡張3を導入した ResNetLSTM-で は,導入前の ResNetLSTM から精度が0.018 ポイント向 上している.従って,3.3節で示した上位層の出力チャネ ル間の関係性を表現する拡張が一定の効果を示したものと 考えられる.

4.5 検証 3: 学習・評価データに複数の収集グループが含 まれる場合

本検証では、収集環境や利用したセンサ、試行のコント ロールの差異などに影響されることなく行動が識別できる かどうかを、学習データと評価データに複数の収集グルー プが含まれる場合の学習結果を用いて議論する.

本検証では HASC2010 コーパスに含まれる person102, 105-107, 141-150, 162-171, 177-186 の計 34 名のデータを 用いた.上記 34 名に 1 から連番を振り, 奇数番号の被験 者を固定学習データとし, 偶数番号の被験者の 1/3 ずつに 分け, 学習・検証・評価名のデータとして交差検証を行っ た.従って, 総検証数は 6 回である.結果を表 3 に示す.

微差ではあるが最も高い精度を示したのは全ての拡張 を導入した ResNetLSTM± であり, Weighted F^1 -score で 0.875を示した. ResNetLSTM-がほぼ同値の0.872で続く.

ResNetLSTM は Weighted F¹-score で DeepConvLSTM より高い 0.849 を, ResNetLSTM+はより高い 0.854 を示 しており,それぞれの拡張が機能した結果と言える. 評 価 1 において精度が低下した DeepConvLSTM-において も 0.836 と DeepConvLSTM より高い精度を示している. また,同様に ResNetLSTM と ResNetLSTM-の間でも精 度が向上している. 従って,本検証では提案の全ての拡張 が一定の効果を示したものと考えられる.

4.6 全体考察

検証1より検証2,検証2より検証3とデータの母集団 の複雑さが増すにつれて,提案している全拡張: ResNet 化,センサチャネル間関係を考慮する拡張,上位層の出力 チャネル間関係を考慮する拡張のそれぞれが効果を発揮し



図 3 検証 3 における ResNetLSTM± の第 3 試行のベストモデル
学習過程

Fig. 3 The best training of ResNetLSTM \pm in 3rd validation in evaluation 3.

たように見受けられる。特に最もデータの母集団が複雑な 検証3において、その傾向が顕著である。

拡張の組み合わせから見ると,拡張1(ResNet 化)と拡張 3(上位層の出力チャネル間の関係性を表現する拡張)を組み 合わせた ResNetLSTM-と,それに拡張2(センサチャネル 間の関係を考慮する拡張)を組み合わせた ResNetLSTM± の精度は,全ての検証においても大きな違いが無いことが 分かる.これは「拡張2と拡張3の組み合わせに問題があ る」「表現対象の複雑さに対して DNN 構造が潤沢すぎる」 などの原因が考えられる.今後,検証数を増やす,より複 雑な動作やデータセットを対象として評価を行う,などの 方法により詳細な検証が必要である.

精度面から見ると、学習データと評価データに同一の 被験者の試行を含めて学習した検証1で99.1%の精度が実 現できており、学習済みモデルを利用者自身のデータで Fine-Tuning するなどの使い方で高い精度が示されると考 えられる.一方で、データ母集団のセンサ種や計測環境な どが異なる検証3における精度は最高で0.875であり、更 なる向上が必要であると考えられる.図3に検証3にお ける ResNetLSTM±の第3試行のベストモデルの学習過 程を示す.なお、当該ベストモデルを用いた評価の精度は 0.911であった.図3に示すように学習ロスと学習精度は 継続して収束していくが、検証ロスと検証制度は早期に頭 打ちとなっている.これはネットワークの表現力はあるが、 過学習が発生していると考えられる.従って、Dropout 層 や BatchNormalization 層の適切な挿入などで精度が向上 できる可能性がある.

表 4, 表 5, 表 6 に ResNetLSTM±の検証 1, 2, 3 に おける混合行列を示す.いずれにおいても混合行列から導 出された Total Accuracy と Weighted F^1 -score がほぼ同値 であり,サンプル数の多い行動のみが優先的に学習された わけではないことが分かる.表5と表6より,階段の昇降と歩行の間で誤認識が多いことが分かる.本稿の学習では,階段昇降のデータは全ての時間にわたって学習しており,踊り場などにおける歩行データが混在していた可能性はある.スライディングウインドウのサイズを1秒から拡張し,より長時間のデータから行動を判断する事とすれば,数秒の踊り場での歩行などは排除できる可能性はある.

5. まとめ

本稿では、行動識別における DNN として、チャネル間 関係を考慮した ResNetLSTM による行動認識を提案した. 提案したネットワークは DeepConvLSTM のコンボリュー ション層を ResNet に置き換え、ResNet および LSTM 層 のそれぞれにおいてチャネル間の関係を考慮する拡張を 施した物である。HASC コーパスに収録されている静止、 歩行、ジョギング、スキップ、階段の昇り・下りの6 種類 の行動データを用いて検証した結果、全ての拡張を導入 した DNN は、行動データのサンプル数によって補正した Weighted F¹-score において、同一被験者の他の試行を学 習データとして含む交差検証において 99.1、単一のデータ 収集グループ内における 1 人抜き交差検証において 89.7、 複数のデータ収集グループを含む交差検証において 87.5 の 精度を示し、DeepConvLSTM を用いた場合からそれぞれ 0.7%、8.5%、6.7% 向上した.

一方で,提案した3つの拡張を全て組み合わせた場合の 精度の向上が,拡張を二つ組み合わせた場合と比較してご く少数に留まった.また,学習時の評価ロスの低下が,学習 ロスの低下に比べて早期に頭打ちになる問題も残っている. 今後は,検証数を増やす,より複雑な動作やデータセットを 対象として評価を行う,Dropout 層や BatchNormalization 層の適切に挿入して過学習を抑制するなどの方法により, 詳細な検証と精度の向上を目指す.

謝辞 本研究の一部は, JSPS 科研費 15H02736, および, 公益財団法人立石科学技術振興財団の助成を受けて行われ ました.ここに記して感謝の意を表します.

参考文献

- Hochreiter, S. and Schmidhuber, J.: Long short-term memory, *Neural computation*, Vol. 9, No. 8, pp. 1735– 1780 (1997).
- [2] Inoue, M., Inoue, S. and Nishida, T.: Deep Recurrent Neural Network for Mobile Human Activity Recognition with High Throughput, arXiv preprint arXiv:1611.03607 (2016).
- [3] Krizhevsky, A., Sutskever, I. and Hinton, G. E.: Imagenet classification with deep convolutional neural networks, Advances in neural information processing systems, pp. 1097–1105 (2012).
- [4] He, K., Zhang, X., Ren, S. and Sun, J.: Deep residual learning for image recognition, *Proceedings of the IEEE* conference on computer vision and pattern recognition,

pp. 770–778 (2016).

- [5] Zeng, M., Nguyen, L. T., Yu, B., Mengshoel, O. J., Zhu, J., Wu, P. and Zhang, J.: Convolutional neural networks for human activity recognition using mobile sensors, 6th International Conference on Mobile Computing, Applications and Services (MobiCASE), IEEE, pp. 197–205 (2014).
- [6] Ordóñez, F. J. and Roggen, D.: Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition, *Sensors*, Vol. 16, No. 1, p. 115 (2016).
- [7] Kawaguchi, N., Ogawa, N., Iwasaki, Y., Kaji, K., Terada, T., Murao, K., Inoue, S., Kawahara, Y., Sumi, Y. and Nishio, N.: HASC Challenge: gathering large scale human activity corpus for the real-world activity understandings, *Proceedings of the 2nd augmented human international conference*, ACM, p. 27 (2011).
- [8] Zappi, P., Lombriser, C., Stiefmeier, T., Farella, E., Roggen, D., Benini, L. and Troster, G.: Activity recognition from on-body sensors: accuracy-power trade-off by dynamic sensor selection, *Lecture Notes in Computer Science*, Vol. 4913, p. 17 (2008).
- [9] Chavarriaga, R., Sagha, H., Calatroni, A., Digumarti, S. T., Tröster, G., Millán, J. d. R. and Roggen, D.: The Opportunity challenge: A benchmark database for onbody sensor-based activity recognition, *Pattern Recognition Letters*, Vol. 34, No. 15, pp. 2033–2042 (2013).
- [10] Lockhart, J. W., Weiss, G. M., Xue, J. C., Gallagher, S. T., Grosner, A. B. and Pulickal, T. T.: Design considerations for the WISDM smart phone-based sensor mining architecture, *Proceedings of the Fifth International Workshop on Knowledge Discovery from Sensor Data*, ACM, pp. 25–33 (2011).
- [11] Srivastava, N., Hinton, G. E., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting., *Journal of Machine Learning Research*, Vol. 15, No. 1, pp. 1929–1958 (2014).
- [12] Ioffe, S. and Szegedy, C.: Batch normalization: Accelerating deep network training by reducing internal covariate shift, *International Conference on Machine Learning*, pp. 448–456 (2015).

| | | 識別結果 | | | | | | | |
|----|-----------|-------|-------|-------|-------|--------------------------------|-------|--------|----------------|
| | | 静止 | 歩行 | ジョギング | スキップ | 階段上り | 階段下り | recall | \mathbf{F}^1 |
| | 静止 | 6827 | 5 | 0 | 0 | 0 | 0 | 0.999 | 1.000 |
| | 步行 | 0 | 6004 | 9 | 0 | 6 | 27 | 0.993 | 0.986 |
| 真値 | ジョギング | 0 | 9 | 6901 | 21 | 0 | 5 | 0.995 | 0.996 |
| | スキップ | 0 | 1 | 0 | 6793 | 3 | 0 | 0.999 | 0.995 |
| | 階段上り | 0 | 94 | 0 | 0 | 6718 | 82 | 0.974 | 0.983 |
| | 階段下り | 0 | 14 | 10 | 38 | 43 | 6734 | 0.985 | 0.984 |
| | precision | 1.000 | 0.980 | 0.997 | 0.991 | 0.992 | 0.983 | | |
| | | | | | | Total Accuracy | | 0.9 | 991 |
| | | | | | | Weighted F ¹ -score | | 0.9 | 991 |

表 4 検証 1 の混合行列 (ResNetLSTM±) Table 4 Confusion matrix of ResNetLSTM± on evaluation 1.

表 5 検証 2 の混合行列 (ResNetLSTM±) Table 5 Confusion matrix of ResNetLSTM± on evaluation 2.

| | 識別結果 | | | | | | | | |
|----|-----------|-------|------|-------|-------|--------------------------------|-------|--------|----------------|
| | | 静止 | 歩行 | ジョギング | スキップ | 階段上り | 階段下り | recall | \mathbf{F}^1 |
| | 静止 | 6826 | 0 | 0 | 6 | 0 | 0 | 0.999 | 0.999 |
| | 步行 | 6 | 4684 | 0 | 98 | 461 | 797 | 0.775 | 0.854 |
| 真値 | ジョギング | 0 | 0 | 5540 | 712 | 208 | 476 | 0.799 | 0.886 |
| | スキップ | 0 | 12 | 18 | 6665 | 0 | 102 | 0.981 | 0.932 |
| | 階段上り | 0 | 87 | 0 | 0 | 6016 | 791 | 0.873 | 0.874 |
| | 階段下り | 0 | 146 | 18 | 31 | 184 | 6460 | 0.945 | 0.835 |
| | precision | 0.999 | 0.95 | 0.994 | 0.887 | 0.876 | 0.749 | | |
| | | | | | | Total Accuracy | | 0.8 | 397 |
| | | | | | | Weighted F ¹ -score | | 0.8 | 397 |

表 6 検証 3 の混合行列 (ResNetLSTM±) Table 6 Confusion matrix of ResNetLSTM± on evaluation 3.

| | | 識別結果 | | | | | | | |
|----|-----------|--------|--------|--------|--------|-----------------------|----------|----------|----------------|
| | | 静止 | 歩行 | ジョギング | スキップ | 階段上り | 階段下り | recall | \mathbf{F}^1 |
| | 静止 | 145014 | 0 | 0 | 0 | 4202 | 0 | 0.972 | 0.983 |
| | 步行 | 0 | 104315 | 27 | 113 | 15460 | 14223 | 0.778 | 0.773 |
| 古店 | ジョギング | 0 | 2217 | 128077 | 919 | 40 | 527 | 0.972 | 0.981 |
| 具個 | スキップ | 0 | 6 | 115 | 129296 | 27 | 1370 | 0.988 | 0.98 |
| | 階段上り | 875 | 20153 | 0 | 263 | 99285 | 18300 | 0.715 | 0.738 |
| | 階段下り | 0 | 9116 | 1039 | 2368 | 11353 | 108470 | 0.820 | 0.788 |
| | precision | 0.994 | 0.768 | 0.991 | 0.972 | 0.762 | 0.759 | | |
| | | | | | | Total | Accuracy | cy 0.874 | |
| | | | | | | Weighted F^1 -score | | 0.8 | 375 |