

Octave Convolutionを用いたセンサによる行動認識手法

近藤 和真¹ 長谷川 達人¹

概要:

人間行動認識 (Human Activity Recognition: HAR) は人が身に付けたセンサや空間に設置されているセンサから人間の行動を観測し, 得られたデータから行動や活動の種類を推定する課題である. 近年では, Convolutional Neural Network(CNN) などの深層学習を用いた HAR 手法が多数提案されている. HAR では加速度センサ, ジャイロセンサ, 磁力センサなど様々なセンサが用いられるが, 特に加速度センサは多くの手法で用いられているセンサである. 人間の行動を観測した加速度センサデータは一般に時系列データであるが, 周波数解析により異なる行動には異なる周波数成分が含まれていることがわかる. 従って, HAR において周波数空間に着目することは重要であると考えられる. 本研究では, 入力データを高周波成分と低周波成分に分離しながら畳み込み処理を行う Octave Convolution を CNN ベースの HAR モデルに適用し認識精度と計算コストの2点に関して検証を行った. その結果, 被験者数が多く十分に訓練データが確保できる状況では Octave Convolution を用いることで推定精度が向上することを明らかにした. 一方で, 1次元の Octave Convolution を用いた CNN モデルでは訓練データの被験者数によらず計算コストが増加することが明らかとなった.

Sensor-based Human Activity Recognition using Octave Convolution

KAZUMA KONDO¹ TATSUHITO HASEGAWA¹

1. はじめに

人間行動認識 (Human Activity Recognition: HAR) は観測されたセンサデータから人の行動や状態を予測する課題である. HAR では主に, スマートフォンやウェアラブルデバイスなどの人が常に身に着けているデバイスから観測されるデータを用いて, 人の行動や状態を推定する. 推定結果から得られた情報はデバイスユーザの日常生活のサポートや組織の意思決定に用いられる. 例えば, スマートフォンから得られたセンサデータを用いてユーザの行動を推定し, 推定された行動情報を用いて, ユーザへ運動を促すことでユーザのヘルスケアをサポートすることができる. また軍事においては, 現場の兵士の状態をセンサデータから把握することでその後の作戦行動に役立てることができる [1]. 近年では, 様々な分野で深層学習の適用が盛んに行われており, 深層学習を用いた HAR に関する研究も数多く行われている. Convolutional Neural Network(CNN)

は深層学習を用いた HAR でよく用いられるモデルであり, 「静止」・「歩く」・「走る」などの基本行動の推定では 93.8% の精度を達成している [2].

HAR ではモバイルデバイスに広く搭載されている加速度センサのデータを用いることが多い. 加速度センサから得られるデータは時系列データであるが, 周波数解析によってパワースペクトルに変換することで, 行動の種類によってセンサデータに含まれる周波数成分に違いが出ることがわかる. 例えば, 「歩行」のように動きの変化が比較的小さい行動は高周波成分が小さくなる傾向があり, 「走行」のように動きの変化が大きい行動は全体を通して高周波数成分が大きくなる傾向がある. ゆえに, 加速度センサデータによる行動認識において周波数成分に着目することは重要であると考えられるが, 深層学習を用いた HAR 手法において時系列センサデータに含まれる周波数成分に着目した手法はあまり検討されていない. また, HAR では推定モデルの計算コスト削減が研究課題の一つとなっている. HAR を実用化する際, HAR モデルはスマートフォンのようなモバイルデバイスに搭載されることが想定されている

¹ 福井大学工学研究科知識社会基礎工学専攻
Graduate School Engineering, University of Fukui

ため、HAR モデルは計算コストを抑えつつ、高精度の行動認識を行うことが求められる。

コンピュータビジョンの分野では、画像の高周波成分と低周波成分を分離しながら畳み込み処理を行う Octave Convolution[3] という手法が提案されている。Octave Convolution は入力画像を高周波成分と低周波成分に分離しながらそれぞれに対し畳み込み処理を行い、それぞれを交差させながら合成したものを出力する。低周波成分はオリジナルの入力画像をダウンサンプリングすることで抽出を行うため、低周波成分の空間冗長性を削減することができ、深層学習モデルの計算コストを削減することができる。また、この手法は通常の畳み込み層と同じパラメータ数で精度を向上させることができることが報告されている。

以上を踏まえて、本研究では行動認識の改善と消費電力の削減を目的として Octave Convolution を用いた HAR 手法を提案する。一般的な CNN モデルに対し、畳み込み層を Octave Convolution に置き換えたモデルを実装し、比較実験を行う。実験では画像認識分野で提案されている Octave Convolution が HAR に与える影響を明らかにする。

2. 関連研究

2.1 深層学習を用いた HAR 手法

深層学習を用いた HAR では CNN をベースとしたモデルが多く提案されている。モデルの多くは観測されたセンサデータをそのままモデルの入力とし、特徴抽出と行動分類を自動で行う。従って、深層学習を用いた HAR モデルではセンサデータから行動認識に有用な特徴表現を獲得することが重要となる。CNN を用いた HAR モデルでは Zeng ら [4] のように畳み込み層を 2, 3 層用いるシンプルモデルが多いが、最近ではより複雑な手法も提案されている。

Zeng ら [4] は 3 軸の加速度センサデータを用いた CNN モデルを提案し、HAR における CNN の有効性を検証した。提案された CNN モデルでは畳み込み層と MaxPooling 層を順に 1 層ずつ接続したモデルを採用している。このモデルを各軸に対して独立に用いることによって特徴抽出を行っている。その後、結合された各軸の特徴量を用いて全結合層によって行動を推定する。また、HAR では観測されたセンサデータから一定のサイズのフレームを切り出し、切り出されたフレームをモデルの入力として扱う。そのため、モデルに入力されるフレームはフレームごとに位相が異なってしまう。彼らはこの問題に対処するために "Partial Weight Sharing" のという手法も提案している。通常の CNN では全ての領域に対して、同じフィルタを用いた畳み込み処理を行うが、提案された手法は異なる領域に対して異なるフィルタを用いて畳み込み処理を行う。領域ごとに異なるフィルタを用いることで全体から一つの特徴を抽出するのではなく、各領域ごとに異なる特徴を抽出することができるため、入力データの位相差に頑健になる

と考えられる。

Yang ら [5] は 1 種類のセンサデータだけではなく、複数種類のセンサのデータを用いた CNN モデルを提案した。提案されたモデルは Zeng ら [4] が提案したモデルよりも深いモデル構造となっており、より高レベルの特徴表現を獲得することを目的としている。このモデルでは複数のセンサによって観測されたの時系列データを並列に連結し画像のようなデータを作成し、モデルの入力としている。従って、CNN モデルへの入力は 2 次元データとなるが、CNN モデル内部で行われる畳み込み処理や Pooling 処理は画像の場合とは異なり、2 次元方向に行うのではなく、各時系列データに対して平行な方向 (時間方向) にのみ行う。

Xu ら [6] は GoogleNet[7] で用いられる Inception Module と Gated Recurrent Unit (GRU)[8] を組み合わせた Inno-HAR を提案している。Inception Module による特徴抽出を行い、GRU を用いてセンサデータの時系列性を考慮した行動分類を行う。

2.2 HAR システムのコスト削減

HAR では行動推定モデルの計算コストを削減することが重要である。行動認識システムは人間が普段身に付けることができるようなモバイルデバイスに搭載されることが想定されるため、ハードウェアの制約により HAR システムのコスト削減は重要な課題である。

Krause ら [9] は自身らが開発した eWatch[10] を用いて、サンプリングレートと機械学習モデルに入力する特徴ベクトルの選択による行動認識精度と消費電力の変化を調査し、選択的なサンプリング戦略を提案した。その結果、1-10Hz といった低周波領域においても十分な分類性能を持つモデルにより、分類精度を改善することができることがわかった。これはサンプリングレートを下げ、消費電力を抑えたとしても、分類精度の低下を抑えることができることを示している。

中島ら [11] はサンプリング周波数を制御し、消費電力を低減させる手法を提案している。彼らは 2 種類のサンプリング周波数制御手法を提案しているが、そのうちの一つである "バースト抜き" は観測されたデータ系列の分散値を用いる手法で、サンプリング周波数を動的に制御できる。行動は繰り返しのため、行動を観測したデータは一定時間で同じようなパターンを繰り返すと考えられる。そこで観測された長さ N のデータ系列の分散値に変化がなければ、その時点から N サンプルのデータを観測を行わないことで、実質的にサンプリングレートを下げたことになり消費電力を減らすことができる。

2.3 本研究の立ち位置

深層学習を用いた HAR の先行事例では行動認識精度を向上させるためにセンサデータの時系列性を考慮する手

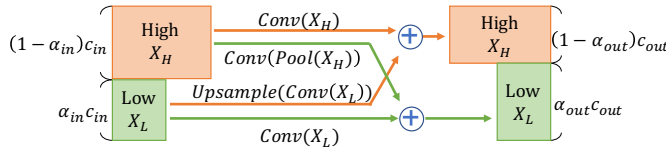


図 1: Octave Convolution の処理概要

法など様々な手法が提案されている。しかし、時系列センサデータに含まれる周波数成分に着目した手法はあまり検討されていない。また、HAR システムのコスト削減については主にサンプリング周波数の動的な調節による消費電力削減手法が提案されており、特に深層学習を用いた HAR ではモデルそのものの計算コストに着目した取り組みは少ない。一方、画像認識分野では画像の周波数成分に着目した Octave Convolution が提案されている。Octave Convolution は通常の畳み込み層を置き換える形で用いることができ、Octave Convolution を用いることで認識精度の向上と計算コスト削減を同時に達成できることが示されている。そこで本研究では、画像認識分野で提案されている Octave Convolution を用いた CNN ベースの HAR モデルを実装し、行動認識精度と HAR モデルの計算コストの 2 点に関して、Octave Convolution がどのように作用するかを明らかにする。

3. Octave Convolution

3.1 Octave Convolution の概要

本節では画像認識手法として Chen ら [3] が提案した Octave Convolution について説明する。画像はオブジェクトのエッジのように値の変化が大きい高周波成分と色の濃淡のような隣接する領域が比較的に近い値をとる低周波成分に分解することができる。低周波成分では隣接する値が近い値をとることが多いため、隣接する領域を一つの値にまとめて圧縮することで情報量を減らすことなく、データサイズを小さくすることができると考えられる。また、低周波成分の圧縮によりメモリコストを減らし、低周波成分に対する処理における計算コストを削減することが可能である。

Octave Convolution ではこのような仮定の下で、Pooling による低周波成分の抽出により高周波成分と低周波成分の分解を促し、各成分に対して別の畳み込み処理を行う。高周波成分と低周波成分は畳み込み処理後にそれぞれを交差させながら合成される。これにより高周波成分と低周波成分の関係性も考慮することができる。また、Octave Convolution を用いたモデルは通常の畳み込み層と置き換える形で構成することができるため、元のモデルを大きく変更することがなく実装することが可能である。

3.2 HAR への適用

ここでは HAR における Octave Convolution の具体的

な処理を説明する。図 1 に HAR 向けに調整した Octave Convolution の処理概要を示す。Octave Convolution ではチャンネルで高周波成分と低周波成分を区別し、各成分の割合はハイパーパラメータ α によって決定される。例えば、バッチサイズが 100、チャンネル数が 64 である中間表現を考える。このときハイパーパラメータ $\alpha = 0.5$ である場合、64 のチャンネルのうち連続する $64 * \alpha = 32$ チャンネルの特徴マップを低周波成分、残りの 32 チャンネルを高周波成分として扱う。このとき高周波成分の特徴マップサイズを 128 とすると、高周波成分の形はバッチサイズ、チャンネル数、特徴マップサイズの順に (100, 32, 128) と表される。低周波成分は Pooling によって半分にダウンサンプリングされたものであるため、特徴マップのサイズは高周波成分の半分となる。よって、低周波成分の形は (100, 32, 64) となる。一方で、通常の畳み込み層では高周波成分と低周波成分を分解せずに畳み込み処理をおこなうため低周波成分の形は (100, 32, 128) となり、Octave Convolution ではメモリコスト削減とそれに伴う畳み込み処理の計算コスト削減が実現されていることがわかる。

畳み込み処理では、入力チャンネル数、出力チャンネル数をそれぞれ c_{in}, c_{out} 、入力データと出力データの低周波成分の割合をそれぞれ $\alpha_{in}, \alpha_{out}$ とすると、図 1 のように高周波成分から高周波成分 ($(1 - \alpha_{in})c_{in} \rightarrow (1 - \alpha_{out})c_{out}$)、高周波成分から低周波成分 ($(1 - \alpha_{in})c_{in} \rightarrow \alpha_{out}c_{out}$)、低周波成分から高周波成分 ($\alpha_{in}c_{in} \rightarrow (1 - \alpha_{out})c_{out}$)、低周波成分から低周波成分 ($\alpha_{in}c_{in} \rightarrow \alpha_{out}c_{out}$) の 4 種類の畳み込み処理を行う (カッコ内は入力と出力のチャンネル数を表している)。高周波成分から低周波成分に変換する際は MaxPooling を用い、低周波成分から高周波成分に変換する際は線形補完 (Upsample) を用いている。

4. 検証実験

本章では、長谷川ら [12] によって行動認識における有効性が示されている VGG モデル [13] に Octave Convolution を適用し、行動認識精度と計算コストの 2 点について検証を行う。行動認識精度の検証に関しては訓練に用いる被験者数と Octave Convolution のパラメータの 2 点と認識精度の関係性を調査する。

4.1 データセット

本研究ではすべての検証実験において HASC データセット [14] を使用している。HASC データセットには被験者ごとに、「stay(静止)」、「walk(歩行)」、「jog(走行)」、「skip(スキップ)」、「stUp(階段上り)」、「stDown(階段下り)」の 6 種類の行動データが様々な条件下で記録されている。HASC データセットには 1 種類の行動のみを 1 つセッションで記録したデータと様々な行動を 1 つのセッションで記録したデータが含まれているが、本研究では前者のデータのみ

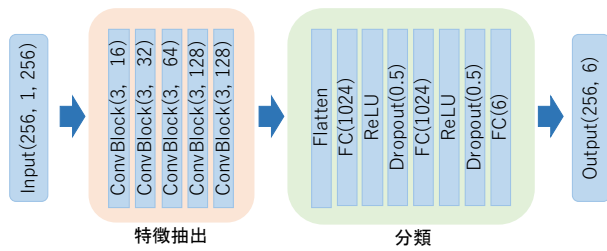


図 2: 検証に用いた VGG11 モデルの概要

を用いている。検証実験ではサンプリング周波数が 100Hz である y 軸の加速度センサデータを用いており、他の条件の制約はない。Octave Convolution ではチャンネルで高周波成分と低周波成分を区別する。本稿では周波数成分の分解による行動認識精度の変化を検証したいため、事前の検証で最も精度が高かった y 軸の加速度センサデータのみを用いている。モデルへの入力是一定サイズのフレームであり、フレームサイズ 256(2.56 s)、ストライド幅 256(2.56 s) として入力フレームを作成した。

4.2 検証用モデルの実装

本稿では、Octave Convolution を用いた VGG11 を構成し、検証実験を行う（以降このモデルを OctaveVGG11 と呼ぶ）。図 2 は検証に用いた VGG11 モデルの概要を示している。図中の ConvBlock は Conv, ReLU, MaxPool を順に接続した構造となっており、ConvBlock(3, 128) は畳み込み層のカーネルサイズが 3、出力チャンネル数が 128 であることを表している。ただし、出力チャンネル数が 64 および 128 の ConvBlock では畳み込み層が 1 層増え、Conv, Conv, ReLU, MaxPool という構造になっている。また、ConvBlock で用いられる MaxPool のカーネルサイズはすべて 2 である。OctaveVGG11 は VGG11 モデルに含まれる全ての畳み込み層を Octave Convolution に置き換えたモデルである。

検証用モデルへの入力はセンサデータから切り出した一定サイズのフレームである。実際にモデルに入力する際は複数のフレームからなるミニバッチ構成し、それを 1 テンソルとして扱う。しかし、Octave Convolution では高周波成分と低周波成分をそれぞれ別のテンソルとして扱うため、高周波成分と低周波成分の 2 つのテンソルがモデル上を伝搬していくことになる。そこで今回の検証実験では Chen ら [3] の方法にならい、入力されるフレームを高周波成分として扱い、モデルへの入力は高周波成分のみとしている。つまり、最も入力層に近い Octave Convolution では高周波成分のみが入力され、ハイパーパラメータ α で定められた割合に従って、高周波成分と低周波成分を出力する。また、最も出力層に近い Octave Convolution では高周波成分のみを出力するようにハイパーパラメータを $\alpha = 0$ と設定し、Octave Convolution 内部で高周波成分と低周波成分

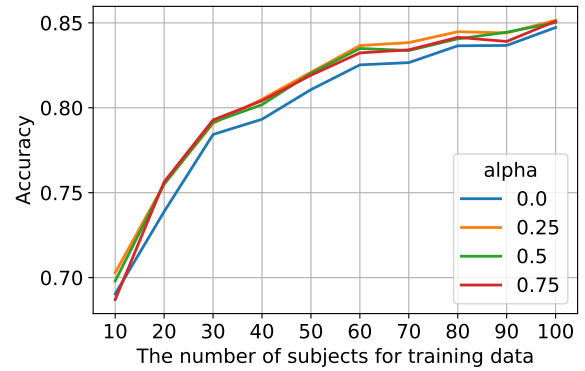


図 3: Octave Convolution を用いた CNN モデルにおける訓練用データの被験者数による行動認識精度の変化

を結合した。

4.3 行動認識精度の検証

ここでは訓練に用いる被験者数および Octave Convolution のハイパーパラメータ α による行動認識精度の変化を検証する。訓練に用いる被験者数は 10 人から 100 人まで 10 人ずつ増やしながらか比較検証を行う。訓練データに含まれない 83 人のデータで構成される。検証ではランダムに被験者を選択することで訓練用データセットを作成し、作成したデータセットを用いてハイパーパラメータ α を変えながら精度検証を行う。これを各被験者数で 10 回ずつの精度検証を行い、その平均値で評価を行う。Octave Convolution のハイパーパラメータ α は $\alpha \in \{0.25, 0.5, 0.75\}$ とし、Octave Convolution 内で高周波成分から低周波成分に変換する際のダウンサンプリングのスケール値 r は 2 とする。本研究では 1 次元のセンサデータを用いるため、それに合わせて VGG11 の畳み込み層には 1 次元畳み込み層を使用し、出力フィルタ数をすべてオリジナル VGG11 モデルから $\frac{1}{4}$ に減らしている。

図 3 に訓練データの被験者数を変化させ、推定精度を検証した結果を示す。 $\alpha = 0$ は Octave Convolution を用いない VGG11 における結果を表している。 $\alpha = 0$ の場合、Octave Convolution 層に入力されるデータは全て高周波成分として扱われるため、実質的に通常の畳み込み層と同等の処理となる。

この結果から、被験者数に関わらず、OctaveVGG11 は通常の VGG11 モデルよりも約 1% の精度向上が見られる。Octave Convolution のハイパーパラメータ α による精度変化はほぼないが、 $\alpha = 0.25$ のときが最も精度が良い傾向がある。訓練データに含まれる被験者数に関しては被験者数を増やすことで通常の VGG11 および OctaveVGG11 ともに精度の向上が見られるが、被験者数が増加するにつれて精度の増加率は小さくなっている。

図 4 は被験者数を 1 から 10 人まで 1 ずつ増やし、同様の

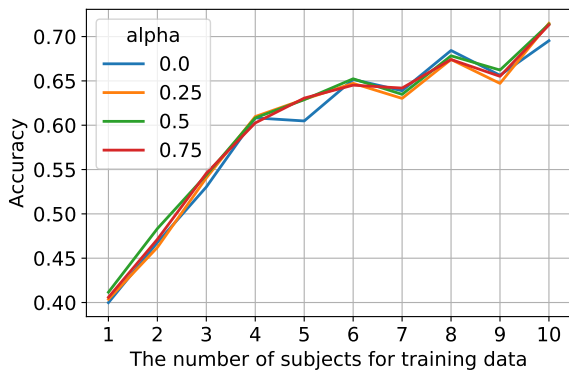


図 4: Octave Convolution を用いた CNN モデルにおける
訓練用データの被験者数による行動認識精度の変化
(1 から 10 人の場合)

表 1: 実験環境

	Env1	Env2
CPU	Core i9-990KF	Core i5-7500
CPU Memory	16GB	8GB
GPU	RTX 2080 Ti	GTX 1070
GPU Memory	11GB	8GB

検証を行った結果である。被験者数が 1 から 10 人の場合、通常の VGG11 と OctaveVGG11 の精度差は被験者数・ハイパーパラメータ α によってバラバラであるが、被験者数が 5 人の場合を除いて、通常の VGG11 と OctaveVGG11 で精度に大きな差は見られない。

これらのことから、行動認識において Octave Convolution は訓練データが十分にある状況下で、通常の畳み込み層よりも有用な特徴表現を獲得できたと考えられる。Octave Convolution では入力されたデータを高周波成分と低周波成分に分解しそれぞれに対し別の畳み込み処理を行うため、Octave Convolution は高周波成分と低周波成分のそれぞれから特徴抽出を行うことを促すことができ、より詳細に特徴表現を獲得できたと考えられる。

4.4 計算コストの検証

ここでは Octave Convolution を用いた際の計算コストを検証する。VGG11 に対し Octave Convolution を適用し、それぞれのモデルで計算コストを計測する。計算コストは 1 つのミニバッチを入力した際の推論時間で評価を行う。推論時間は 100 回の計測し、その平均値を計算コストの指標とする。計算コストの検証は 2 種類の環境で CPU と GPU(CUDA) で行った。実験環境は表 1 の通りである。入力するデータは乱数を用いたダミーデータであり、バッチサイズ、チャンネル数、フレームサイズはそれぞれ 256, 1, 256 である。

計算コストの検証結果を表 2 に示す。この結果から、VGG11 において、Octave Convolution を用いることで計

表 2: Octave Convolution を用いた VGG11 の計算コスト

alpha	Env1		Env2	
	cpu	cuda	cpu	cuda
0	4.070E-02	1.961E-03	7.138E-02	3.534E-03
0.25	7.862E-02	3.056E-02	1.280E-01	3.758E-02
0.5	6.590E-02	1.917E-02	1.130E-01	2.828E-02
0.75	5.464E-02	1.290E-02	9.542E-02	1.675E-02

表 3: 2 次元の Octave Convolution を用いた VGG11 の
計算コスト

alpha	Env1		Env2	
	cpu	cuda	cpu	cuda
0	7.112E-01	9.309E-03	1.223E+00	1.761E-02
0.25	9.255E-01	1.459E-02	1.321E+00	2.567E-02
0.5	6.881E-01	8.878E-03	9.606E-01	1.862E-02
0.75	4.289E-01	6.919E-03	6.410E-01	1.500E-02

算コストが増加していることが分かる。Octave Convolution は通常の畳み込み層よりも畳み込み処理の回数が 4 倍となる。そのため、Octave Convolution によるメモリコストの削減よりも畳み込み処理の増加による計算コストの増加が支配的になり、全体として計算コストが増加したと考えられる。

表 3 は 2 次元の Octave Convolution を VGG11 に適用し、同様の検証を行った結果を示している。入力の形はバッチサイズ 8、チャンネル数 1、画像サイズ 256x256 である。ハードウェアの制約により、バッチサイズを小さくしている。2 次元の畳み込み処理を行う Octave Convolution では環境によらず、推論時間は $\alpha = 0.25$ で一旦増加するが以降は減少し $\alpha = 0$ の場合より小さくなっている。 $\alpha = 0.25$ で推論時間が増加するのは 1 次元の Octave Convolution の場合と同様に畳み込み処理回数の増加による計算コストの増加が支配的になったためであると考えられる。

この結果を踏まえると、波形のような 1 次元データは画像などの 2 次元データと比べて、データサイズが 1 オーダー少なくなり、Octave Convolution によるメモリコストの削減量も相対的に小さくなってしまいうことで、結果として計算コストが増加したと言える。

5. まとめ

本研究では、画像認識分野で提案された Octave Convolution を行動認識に適用し、精度検証を行った。検証では VGG11 モデルをベースとして、訓練データに含まれる被験者数と Octave Convolution のハイパーパラメータ α を様々に変えながら、通常の畳み込み層と Octave Convolution で比較を行った。検証の結果、訓練データに含まれる被験者数が多く十分な訓練データが存在する場合において、Octave Convolution による精度向上が見られた。Octave Convolution は 1 つのデータを Pooling によって高周波成

分と低周波成分に分け、それぞれに対し、畳み込み処理を行うため、通常の畳み込み層よりも詳細な特徴表現を獲得できたと考えられる。

また、画像認識において Octave Convolution は低周波成分空間冗長性を除去し、計算コストを削減できることが示されており、本研究では行動認識における計算コスト削減効果についても検証を行った。今回は1軸のセンサデータを入力としているため、1次元畳み込み処理を行う Octave Convolution における計算コスト削減効果について検証した。検証の結果、低周波成分のメモリコストの削減よりも畳み込み処理回数の増加が支配的になり、1次元の Octave Convolution の計算コストは通常の畳み込み処理よりも増加することが明らかとなった。

Octave Convolution は画像認識分野において提案された手法で、この手法をそのまま行動認識に適用しても劇的な効果は見られなかった。これは画像とセンサデータでの周波数の扱いの違いが関係していると見られる。画像における周波数は空間周波数を指し、センサデータでは時間周波数を指す。Octave Convolution は空間周波数に着目した手法であり、時間周波数に着目した手法を検討することで更なる精度向上を見込むことができる。よって、今後は時間周波数に着目した手法を検討するとともに、様々なモデルで Octave Convolution の検証を行いたい。

謝辞

本研究の一部は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] Oscar Lara and Miguel Labrador. A Survey on Human Activity Recognition Using Wearable Sensors. *Communications Surveys & Tutorials, IEEE*, Vol. 15, pp. 1192–1209, 01 2013.
- [2] Yuqing Chen and Yang Xue. A Deep Learning Approach to Human Activity Recognition Based on Single Accelerometer. pp. 1488–1492, 10 2015.
- [3] Yunpeng Chen, Haoqi Fan, Bing Xu, Zhicheng Yan, Yanis Kalantidis, Marcus Rohrbach, Shuicheng Yan, and Jiashi Feng. Drop an Octave: Reducing Spatial Redundancy in Convolutional Neural Networks With Octave Convolution. In *The IEEE International Conference on Computer Vision (ICCV)*, October 2019.
- [4] Ming Zeng, Le T. Nguyen, Bo Yu, Ole Mengshoel, Jiang Zhu, and Pang Wu. Convolutional Neural Networks for Human Activity Recognition using Mobile Sensors. 11 2014.
- [5] Jian Bo Yang, Minh Nhut Nguyen, Phyto Phyto San, Xiao Li Li, and Shonali Krishnaswamy. Deep convolutional neural networks on multichannel time series for human activity recognition. In *Proceedings of the 24th International Conference on Artificial Intelligence, IJCAI'15*, p. 3995–4001. AAAI Press, 2015.
- [6] Cheng Xu, Duo Chai, Jie He, Xiaotong Zhang, and Shihong Duan. Innohar: A deep neural network for complex human activity recognition. *IEEE Access*, Vol. 7, pp. 9893–9902, 2019.
- [7] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*, 2015.
- [8] Kyunghyun Cho, Bart van Merriënboer, Caglar Gulcehre, Dzmitry Bahdanau, Fethi Bougares, Holger Schwenk, and Yoshua Bengio. Learning phrase representations using RNN encoder–decoder for statistical machine translation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pp. 1724–1734, Doha, Qatar, October 2014. Association for Computational Linguistics.
- [9] Andreas Krause, Matthias Ihmig, Edward Rankin, Derek Leong, Smriti Gupta, Daniel Siewiorek, Asim Smailagic, Michael Deisher, and Uttam Sengupta. Trading off prediction accuracy and power consumption for context-aware wearable computing. *2012 16th International Symposium on Wearable Computers*, Vol. 0, pp. 20–26, 10 2005.
- [10] Asim Smailagic, Daniel P. Siewiorek, Uwe Maurer, Anthony Rowe, and Karen P. Tang. ewatch: context sensitive system design case study. In *IEEE Computer Society Annual Symposium on VLSI: New Frontiers in VLSI Design (ISVLSI'05)*, pp. 98–103, May 2005.
- [11] 中島悠貴, 村尾和哉, 寺田努, 塚本昌彦. サンプリング制御とデータ補完による行動認識システムの省電力化手法. 情報処理学会論文誌, Vol. 52, No. 8, pp. 2455–2466, aug 2011.
- [12] Tatsuhito Hasegawa and Makoto Koshino. Representation learning by convolutional neural network for smartphone sensor based activity recognition. In *Proceedings of the 2019 2nd International Conference on Computational Intelligence and Intelligent Systems, CIIS 2019*, p. 99–104, New York, NY, USA, 2019. Association for Computing Machinery.
- [13] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *International Conference on Learning Representations*, 2015.
- [14] Haruyuki Ichino, Katsuhiko Kaji, Ken Sakurada, Kei Hiroi, and Nobuo Kawaguchi. HASC-PAC2016: large scale human pedestrian activity corpus and its baseline recognition. pp. 705–714, 09 2016.