

Octave Mix : 行動認識における周波数分解を用いたデータ拡張 Octave Mix: Data augmentation using frequency decomposition for activity recognition

長谷川 達人[†]
Tatsuhito Hasegawa

1. はじめに

Society 5.0 の実現に向け、センサを用いたコンテキストウェアネスの研究が盛んに行われている。中でも行動認識は人が持つセンサから動作を計測し、人の行動の種類を認識する課題である。人の行動が認識できることで、ライフログや、行動に応じたサービスの提供が実現できる。集団の行動データが集まれば、マーケティングや渋滞緩和等への活用も見込める。

行動認識は機械学習で実現される事が多く[1]、近年では深層学習を応用した行動認識も活発に研究がなされている[2]。深層学習モデルは表現能力の高さ故に、過学習を回避した汎用的なモデルを獲得するために大量の訓練データを必要とする。この課題に対応するため、一般にデータ拡張 (DA : Data Augmentation) によりデータのかさ増しが行われる。特に、センサを用いたコンテキストウェアネスの分野では、教師ラベル付きのデータを拡充させることは容易ではなく、DAは重要な技術である。

本研究では、センサデータを用いた行動認識における新しい DA 手法として Octave Mix を提案する。DA には、単一データを変形するものと、複数データを合成するものがあり、Octave Mix は後者の複数データを合成するデータ拡張手法である。2 つのセンサデータに対して、片方に Low Pass Filter (LPF) を適用、もう片方に High Pass Filter (HPF) を適用後、両者を合成する。提案手法の評価を行うため、スマートフォンセンサを用いた基本行動認識のベンチマークデータセット HASC [3]を用いて評価実験を行い、複数データを合成する DA 手法である mixup [4]、及び、RICAP [5]等と比較を行うことで、提案手法の有効性を示す。また、これらを組合せた手法を新たに提案し、複数 DA 手法の組み合わせによる推定精度の向上についても議論を行う。

2. 関連研究

2.1 Data Augmentation (DA)

DA は得られている訓練データを幾何学的変換やノイズの付与により、擬似的にデータのバリエーションを増加させる手法である。深層学習の火付け役となった AlexNet [6]にて DA が用いられたこともあり、画像認識分野では DA を用いることが一般的である。同文献では画像の切り出し、水平反転、RGB チャネル強度の変化を行っている。画像認識でよく用いられる DA 手法としては、ノイズの付与、水平垂直反転、拡大縮小、色調反転、シアー等がある。

DA は手法自体がシンプルであるため、DA 自体に焦点を当てた研究は近年まであまりなされていなかった背景がある。DA 手法の提案に焦点を当てた論文としては Cutout [7] や Random Erasing [8]のように画像の一部矩形領域をラン

ムにマスクする手法がある。これらの手法は単一画像を幾何学的変換によって変形するものである。これに対して、複数の画像を合成することで DA を行う手法も提案されている。mixup[4]は複数の訓練用データを合成する手法である。2 つのラベル付きデータ (x_1, y_1) , (x_2, y_2) が与えられるとき、式(1)で拡張後のデータ (\tilde{x}, \tilde{y}) を生成する。

$$\begin{aligned}\tilde{x}_{mixup} &= \lambda x_1 + (1 - \lambda)x_2, \\ \tilde{y}_{mixup} &= \lambda y_1 + (1 - \lambda)y_2.\end{aligned}\quad (1)$$

ここで、 $\lambda \sim \text{Beta}(\alpha, \alpha)$, $\alpha \in (0, \infty)$ であり、 $\lambda \in [0, 1]$ となる。 α は mixup のハイパーパラメータである。上式より、mixup は 2 つの入力及び出力を、ベータ分布に基づく乱数 λ を重みとした加重平均によって合成する手法であるといえる。mixup の重要な点は、入力 x だけでなく出力 y も加重平均で合成するという点である。これによって、2 データの中間となるような訓練データに存在しないデータを生成しつつ、擬似的な教師ラベルを与えている。その他の合成系 DA 手法として、RICAP[5]がある。RICAP は 4 つの訓練用画像データから、それぞれランダムで決定した矩形領域を切り出し、並べて結合することで合成画像を生成する手法である。mixup 同様に出力 y も切り出した矩形領域のサイズに応じた加重平均で合成する点の特徴である。

近年では複数の DA 手法から最適な方策を強化学習により探索する AutoAugment [9]や、これをランダムに行うことで高速化した RandAugment [10]、敵対的訓練により高速化した Adversarial AutoAugment [11]なども提案されている。本研究では新しい DA 手法の提案に焦点を当てることから、これらの最適 DA 方策の探索手法については範囲外とする。

2.2 行動認識と DA

行動認識は機械学習を用いることが多い[1, 2]ため、実現するためにはモデルの訓練用データセットを準備する必要がある。行動認識の場合、入力は加速度センサ等のセンサデータであり、出力は行動ラベルとなる。例えば、看護師の業務行動認識を行う例を考えると、看護師に加速度センサを装着してもらい業務中のセンサデータを計測しつつ、同時に各時刻の業務を行動ラベルとして付与してもらう必要がある。もしくは常に行動中の看護師をビデオカメラ等で撮影し、後日動画を確認しながら行動ラベルを付与する必要がある。このように行動認識では訓練用データセットを収集するために多くの労力を要することが少なくない。また、行動認識では、認識対象者によって行動の動作が変わるため、多くの人からバリエーションのある行動データセットを蓄積できることが望ましい。特に深層学習を用いた行動認識の場合、過学習を回避した汎用的な特徴表現を獲得するためには、多くの訓練データを必要とする。すなわち、バリエーション豊かなデータセットが必要であるのに対して、データセットを充実させることは容易ではないという課題背景がある。

[†] 福井大学工学部電気電子情報工学科

Department of Electrical, Electronic and Computer Engineering, School of Engineering, University of Fukui.

この課題を解決する一手法として DA があるが、行動認識分野では、画像認識ほど DA が用いられていない。画像認識と異なり、見た目でもわかりやすいような DA を実装しづらい点や、分野としての議論がまだ浅い点などに起因すると考えられる。いくつかの研究では深層学習を用いた行動認識において DA を行っている[12-14]。これらの文献で採用されている DA は、

- **Rotation:** x, y, z 軸を回転させる。
- **Permutation:** 時系列の一部区間を入れ替える。
- **Scaling:** 振幅方向に値を拡張する。
- **Time-warping:** 時間軸方向に値を拡張する。
- **Magnitude-warping:** なめらかな曲線を畳み込む。
- **Jittering:** ノイズを付与する。
- **Cropping:** 時系列の一部区間をマスクする。

このように、行動認識における DA は限定的な幾何学的変換による議論にとどまっており、特に複数のデータを合成する手法の有効性については議論されていない。

2.3 本研究の位置づけ

以上を踏まえ、以下 4 点が本研究の主要な貢献である。

- 行動認識における DA の有用性は確認されているものの一部の幾何学的変換手法にとどまっている。本研究では、mixup や RICAP といった複数のデータを合成する合成系の DA 手法に関する有用性を実験により検証する。
- センサを用いた行動認識のための新しい合成系 DA 手法として Octave Mix を提案する。Octave Mix はハイパーパラメータによっては mixup を一部に含んでおり、周波数分解により mixup を拡張した手法とみなせる。これにより通常の mixup よりも高い認識精度を実現することができる。
- スマートフォンのセンサを用いた基本行動認識の大規模データセット HASC を用いて、上記を検証する実験を行い、提案手法の有効性を示す。
- 実験結果を深く考察した上で、複数 DA 手法の組み合わせにより精度向上を図る新たな行動認識手法を提案する。

3. Octave Mix

3.1 アルゴリズム

Octave Mix は低周波数成分と高周波成分に分解して畳み込みを行う Octave Convolution [15]から着想を得て、これを合成系の DA に応用したものである。Octave Mix の合成処理は以下の 3 手順で構成される。

1. 2つの入力に対して、それぞれ Low Pass Filter (LPF) と High Pass Filter (HPF) をかけ、低周波数成分 $LPF(x)$ と高周波成分 $HPF(x)$ に分解する。
2. 2つの入力を交差させるように、低周波数成分と高周波数成分を合成 ($LPF(x_1) + HPF(x_2)$ と $LPF(x_2) + HPF(x_1)$) する。
3. 2つの合成波形を、係数 λ を重みとした加重平均により合成する。

したがって、2つのラベル付きデータ (x_1, y_1) , (x_2, y_2) が与えられるとき、この手順で生成されるデータ $(\tilde{x}_{octmix}, \tilde{y}_{octmix})$ は式(2)で定式化される。

$$\begin{aligned}\tilde{x}_{octmix} &= \lambda\{LPF(x_1) + HPF(x_2)\} \\ &\quad + (1-\lambda)\{LPF(x_2) + HPF(x_1)\}, \\ \tilde{y}_{octmix} &= \lambda y_1 + (1-\lambda)y_2.\end{aligned}\quad (2)$$

Octave Mix のアイデアの要点は x の合成前に周波数分解を行っている点と、それに加えて、mixup と同様に y も合成している点である。

図 1 は、 y_1 ="歩行"、 y_2 ="走行"のデータを例に Octave Mix による合成波形の生成例を示している。歩行データである x_1 の方が、全体的に振幅が小さく周波数が低い波形が観測されている。 x_1 に LPF, HPF をかけると、図中青枠で示されるように緩やかな歩行の波形とノイズのような振動に分解される。同様に x_2 着目すると、 x_1 に比べて低周波成分はやや振幅と周波数が高くなっており、高周波成分も全体的に振幅が増している事がわかる。これらを交差させて合成することで、歩行の波形に走行の高周波成分の振動が加わったような波形と、走行の波形に歩行の高周波成分の振動が加わったような波形が生成される。最終的に両者を加重平均によって合成すると、図中の右のように重み係数 λ に応じた波形が生成される。

3.2 ハイパーパラメータと既存手法との関係

Octave Mix のハイパーパラメータは 2 つある。1 つは、重み係数 λ ~ Beta(α, α) を決定する際の α である。mixup にならば、ベータ分布のパラメータは α で統一している。もう 1 つは、LPF と HPF のカットオフ周波数 f_c である。特に行動認識においては、認識対象となる行動の種類によって観測される主な周波数が変化する。したがって、適用するタスクに応じて f_c は調整されることが望ましい。

ここで、各合成波形を g_1, g_2 とすると、Octave Mix の合成は式(3)に変形できる。

$$\begin{aligned}g_1 &= LPF(x_1) + HPF(x_2), \\ g_2 &= LPF(x_2) + HPF(x_1), \\ \tilde{x}_{octmix} &= \lambda g_1 + (1-\lambda)g_2, \\ \tilde{y}_{octmix} &= \lambda y_1 + (1-\lambda)y_2.\end{aligned}\quad (3)$$

式(3)は、 $g_1 = x_1, g_2 = x_2$ のときに mixup の式(1)と一致する。すなわち、カットオフ周波数 f_c を可能な限り大きくしたとき mixup と同じ処理となることを意味する。したがって、Octave Mix は mixup を一部に含む、mixup を拡張した手法であるといえる。

4. 実験設定

4.1 データセットと評価指標

スマートフォンセンサを用いた基本行動認識のベンチマークデータセット HASC [3]を用いて、提案手法の有効性評価を行う。基本行動 6 種類 (停止, 歩行, 走行, スキップ, 階段上り, 階段下り) のラベルがつけられた加速度センサ, 角速度センサ等の計測データがコーパスとして提供されている。このうち、2011 から 2013 年までのコーパスの BasicActivity よりサンプリング周波数が 100Hz のデータを抽出し、加速度センサの生データのみを用いることとした。前処理として、各計測ファイルから前後 5 秒を除去し、フレームサイズ 256 サンプル、ストライド 256 サンプルで時系列分割を行った。計測開始から端末の格納動作等の影響

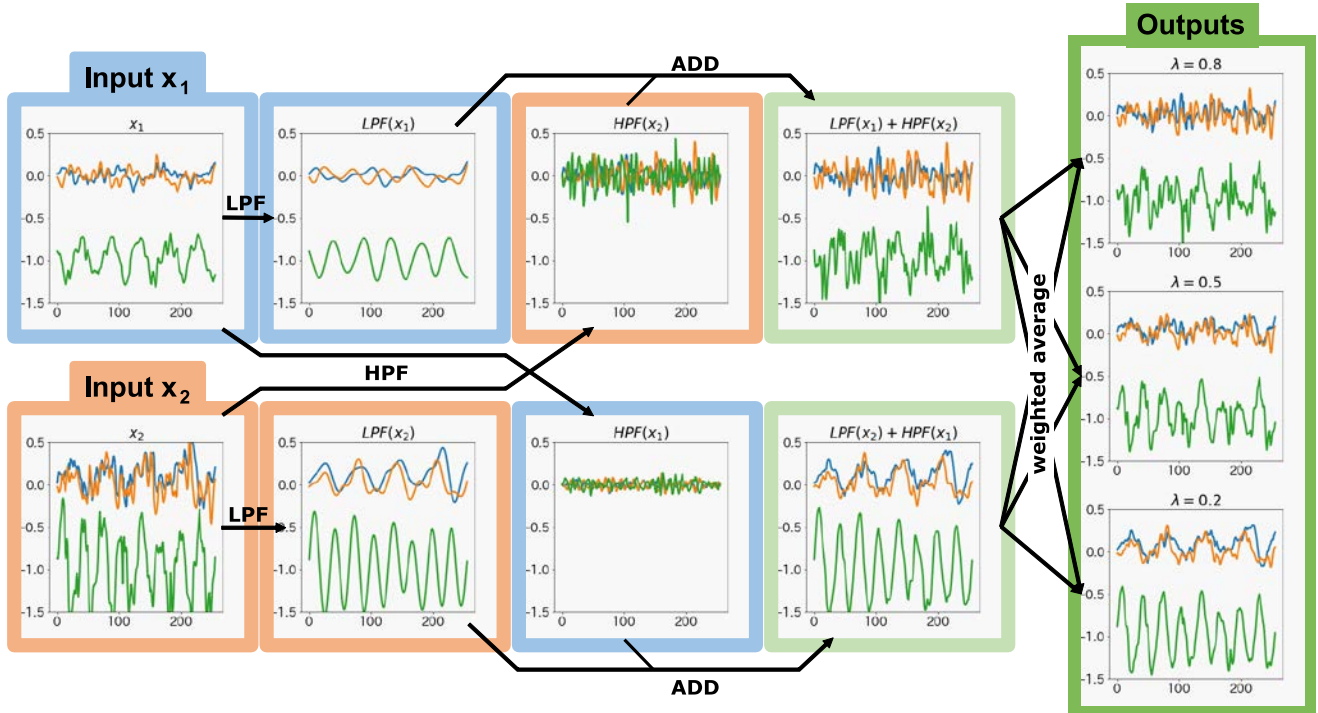


図 1 Octave Mix による 2 波形 (y_1 ="歩行", y_2 ="走行") の合成例

を取り除くため前後 5 秒でトリミングしている。計測機種や性別等のメタ情報は用いない。トリミング後に 1 フレーム以上データが取得できた 176 名のデータを採用した。

行動認識では、個人ごとに行動の動作が異なることから個人依存が問題になることが多い。実運用時には推定対象者のラベル付きデータは獲得できないことを想定し、データセットを被験者単位で分割して評価を行う。本研究では、10名のデータを訓練用に、50名のデータを検証用に、50名のデータをテスト用に分割し、計 110名のデータを使用した。訓練用データを 10人としたのは、DA による精度向上の影響を考察しやすくするためである。被験者はランダムサンプリングにより選択する。訓練用データの被験者数を 10名とすると、訓練データ数は 1500 前後となり多くはない。そのため、ランダムサンプリングで抽出された被験者によって検証精度が多少上下する。そこで、(1)データセットの分割、(2)各手法の訓練と精度評価までを 1 セットとして、30 セット試行した結果の平均で議論することで手法間の推定精度比較を行う。

ラベル間で偏りが少ないことから、評価指標は Accuracy とし、訓練中の各エポックで検証用精度を求め全エポックの最良を取る Best と、最終エポックでの検証用精度 Last で議論する。5.3 節以降は、最終エポック完了後にテスト用データで推定精度を算出した結果で議論する。

4.2 深層学習モデル

本研究では VGG16 [16] を 1 次元センサデータに適用した Convolutional Neural Network (CNN) モデルを用いて検証を行う。行動認識に適した CNN 構造の検証を行った先行研究[17]より、VGG16 は実装が容易かつ高い推定精度を達成するため採用した。構造とパラメータを図 2 に示す。本研究は DA 手法に焦点を当てているため、深層学習モデル構造の違いに関する議論は範囲外とする。

Conv(c, k, s)
 c: Number of channels
 k: Size of kernel
 s: Strides

MaxPool(k, s)
 k: Size of kernel
 s: Strides

ConvBlock(n, c)
 n: Number of repeat
 c: Number of channels

Conv(c, 3, 1)

ReLU

 : n times repeat

MaxPool(2, 2)

VGG16	
Input	(256, 3)
ConvBlock	(2, 32)
ConvBlock	(2, 64)
ConvBlock	(3, 128)
ConvBlock	(3, 256)
ConvBlock	(3, 256)
Flatten	
Dense	(4096)
ReLU	
Dropout (0.5)	
Dense	(4096)
ReLU	
Dropout (0.5)	
Output	(6)

図 2 実験に用いた VGG16 ベースの CNN 構造

本研究では、行動ラベル 6 種類に対して、カテゴリカルクロスエントロピー誤差を最小化する問題として、上記モデルを最適化する。最適化手法は Adam[18] を学習率 $3e-3$ 、その他はデフォルトパラメータで用いた。パラメータは全比較対象の予備実験として学習の収束具合を確認した結果、すべての手法が概ね収束することを確認している。ミニバッチサイズ 500 で、モデルの訓練は 300 エポック行った。

4.3 ベースライン

以下に示す 4 種類の手法をベースラインとして、提案手法と比較評価を行う。

Ref : DA を実施せず訓練データのみを用いる。Ref からの差分が、DA による精度向上度合いを意味する。

mixup [4] : 2 つの入出力ペアを与えて、入力及び出力を式(1)に基づく加重平均により合成する DA を行った。

RICAP [5]: 2つの入出力ペアを与えて、入力波形をベータ分布に基づくランダムな位置で時間方向に分割し、切り貼りすることで合成する DA を行った。出力は mixup と同様の手法で合成した。

LPF-Jit: 1つの入力を与えて、LPF 適用後、ランダムノイズの付与 (Jittering) を行った。LPF の f_c は提案手法同様ハイパーパラメータとした。Jittering は探索の結果良い精度を示した $z \sim N(0, 0.8)$ を入力に加算し、以下の式(4)で DA を行った。提案手法は LPF 適用後ノイズを合成することと同義なのではないかという可能性を検証するため、比較対象として採用した。

$$\tilde{x}_{LPF-jit} = LPF(x) + z \quad (4)$$

モデルの訓練をミニバッチ単位で実施する際に、選択されたミニバッチ X_{mini} と、そのミニバッチの各サンプルに対して 50% の確率で DA を行った \tilde{X}_{mini} を結合して入力とした。ミニバッチサイズ 500 の場合、 $|X_{mini}| = |\tilde{X}_{mini}| = 500$ で合計 1000 サンプルを訓練することになる。

5. 実験結果

5.1 単一 DA 評価結果

各 DA 手法を 1 種類だけ適用する場合の推定精度評価の結果を表 1, 2 に示す。今回は、各エポック終了時に検証用データセットに対して精度評価を行い全エポック中で最良だった精度 Best の 30 試行の平均 (表 1)、最終エポックでの精度 Last の 30 試行の平均 (表 2) を集約した。横軸は DA 手法であり、縦軸はハイパーパラメータ α である。

単一の DA の適用結果では、Best, Last いずれも提案手法の **OctMix** ($\alpha = 0.5, f_c = 4.1$ Hz) で最良の検証精度となった。DA を適用しない場合 (**Ref**) と比べると、6.2% (Best), 5.5% (Last) 精度向上に成功している。他の合成系 DA 手法と比較すると、**mixup** より 1.9% (Best), 2.3% (Last) 精度向上、**RICAP** より 3.2% (Best), 3.6% (Last) 精度向上となった。また、**OctMix** は LPF 適用後に Jittering をしているのと同義ではないかという疑問のもと **LPF-Jit** を比較対象とした。**LPF-Jit** と比較すると 0.4% (Best), 1.1% (Last) の精度向上となり、**OctMix** は **LPF-Jit** と類似した効果を発揮するものの、やや精度向上に貢献することを明らかにした。

OctMix のハイパーパラメータに着目すると、 f_c は 4.1 Hz が最良となった。今回は基本行動認識を対象としていることから比較的low周波で区切ることが効果を発揮したと思われる。また α は 0.5 か 1.0 が良さそうである。

OctMix 適用に際して、式(3)の \tilde{x}_{octmix} に対する λ を反転するという手法も考えられる。すなわち、

$$\begin{aligned} \tilde{x}_{octmix} &= (1 - \lambda)g_1 + \lambda g_2, \\ \tilde{y}_{octmix} &= \lambda y_1 + (1 - \lambda)y_2. \end{aligned} \quad (5)$$

とする手法である。 g_1 及び g_2 はそれぞれに x_1, x_2 を含んでいるため、LPF か HPF どちらを適用したものに、ラベルの比率を合わせるのかという違いがある。これを実験により評価した結果、表 1, 2 の精度よりも 5% 程度精度が低下した。したがって、**OctMix** は LPF を適用したデータのラベルの比率に合わせる式(3)を用いるほうが良いと結論づける。これは低周波成分側に基本行動を分類する上で重要な情報が含まれていることに起因すると考えられる。

5.2 複数 DA 評価結果

単一 DA の評価結果を確認すると、稀に学習が収束しないケースや同じデータを用いても精度が変動することが確認された。これは訓練データのバリエーションが少ないことに加えて、スマートフォンを用いた行動認識であるという課題背景によるものと考えた。スマートフォンは様々な向きで端末をポケット等に格納することから、センサの向きが計測時に不定である。そこで関連研究[12]でも有効性が示されている **Rotation** で DA を行うことで精度向上と安定化を図った。**Rotation** は加速度センサの x, y, z 軸の入れ替え及び正負反転をランダムに実施する DA 手法である。

Rotation 適用後に各 DA 手法を重ねがけした場合の推定精度評価を行った結果を表 3, 4 に示す。単一 DA 時と比べて特徴的な変化は以下の 3 点である。

1. 全体的な精度が大幅に底上げされている。
2. **OctMix** と **LPF-Jit** の差が少し広がっている。
3. **RICAP** の精度が大幅に向上している。

1 について、**Ref** で見ると **Rotation** 有無により 7.7% (Best), 4.8% (Last) 精度が向上している。2 について、表 3, 4 より **OctMix** は **LPF-Jit** よりも 0.9% (Best), 1.7% (Last) 精度が高く、単一 DA 時と比べやや差が広がっている。最も特徴的な 3 について、単一 DA 時に **RICAP** は全く精度向上が期待できなかったが、**Rotation** を事前に適用することで大幅に改善され、全体でも最良の精度を達成する結果となった。これらの結果を総合すると、**OctMix** や **mixup** は 2 つの波形を加算により合成することから、**Rotation** のような効果を内包していたと思われる。一方 **RICAP** は切り貼りによりデータを合成するため **Rotation** のような効果は含まれず、単体では効果を発揮しないが

表 1 単一 DA 評価結果 (最良エポックの検証用データ推定精度の 30 試行平均 : Best)

	Ref	mixup	RICAP	LPF-Jit ($f_c = 10.1$)	OctMix ($f_c = 2.1$)	OctMix ($f_c = 4.1$)	OctMix ($f_c = 6.1$)	OctMix ($f_c = 8.1$)	OctMix ($f_c = 10.1$)
$\alpha = 0.5$		71.9%	71.5%		74.2%	74.7%	73.2%	72.7%	73.8%
$\alpha = 1.0$	68.5%	72.0%	69.1%	74.3%	72.0%	74.6%	72.4%	73.8%	73.1%
$\alpha = 5.0$		72.8%	70.0%		73.8%	73.9%	73.7%	74.2%	72.9%

表 2 単一 DA 評価結果 (最終エポックの検証用データ推定精度の 30 試行平均 : Last)

	Ref	mixup	RICAP	LPF-Jit ($f_c = 10.1$)	OctMix ($f_c = 2.1$)	OctMix ($f_c = 4.1$)	OctMix ($f_c = 6.1$)	OctMix ($f_c = 8.1$)	OctMix ($f_c = 10.1$)
$\alpha = 0.5$		69.6%	68.3%		69.4%	71.9%	70.6%	69.9%	70.7%
$\alpha = 1.0$	66.4%	69.0%	65.5%	70.8%	68.0%	71.2%	69.4%	71.2%	70.3%
$\alpha = 5.0$		68.8%	66.5%		70.4%	70.7%	70.7%	71.0%	69.9%

表 3 Rotation 適用後の各 DA 評価結果 (最良エポックの検証用データ 推定精度の 30 試行平均 : Best)

	Ref	mixup	RICAP	LPF-Jit ($f_c = 8.1$)	OctMix ($f_c = 2.1$)	OctMix ($f_c = 4.1$)	OctMix ($f_c = 6.1$)	OctMix ($f_c = 8.1$)	OctMix ($f_c = 10.1$)
$\alpha = 0.5$		77.4%	78.8%		76.9%	79.0%	78.9%	77.5%	78.7%
$\alpha = 1.0$	76.2%	77.4%	79.2%	78.2%	78.7%	79.1%	78.5%	78.5%	77.1%
$\alpha = 5.0$		77.6%	78.9%		78.3%	78.2%	78.6%	76.3%	78.0%

表 4 Rotation 適用後の各 DA 評価結果 (最終エポックの検証用データ 推定精度の 30 試行平均 : Last)

	Ref	mixup	RICAP	LPF-Jit ($f_c = 8.1$)	OctMix ($f_c = 2.1$)	OctMix ($f_c = 4.1$)	OctMix ($f_c = 6.1$)	OctMix ($f_c = 8.1$)	OctMix ($f_c = 10.1$)
$\alpha = 0.5$		75.1%	76.8%		73.6%	76.8%	76.7%	74.3%	76.3%
$\alpha = 1.0$	71.2%	75.2%	77.6%	75.1%	76.3%	76.9%	76.4%	76.2%	74.0%
$\alpha = 5.0$		75.4%	77.0%		76.2%	76.2%	76.1%	73.9%	75.6%

表 5 各手法におけるテストデータに対するテスト精度と F 値

	accuracy	fall-avg	f _{stay}	f _{walk}	f _{jog}	f _{skip}	f _{stup}	f _{stdown}
(A) Rotation → Ref (None)	71.8%	71.5%	94.7%	53.2%	83.1%	80.4%	54.8%	62.7%
(B) Rotation → RICAP($\alpha = 1.0$)	78.1%	78.6%	98.6%	60.0%	90.6%	88.9%	62.9%	70.4%
(C) Rotation → OctMix($\alpha = 0.5, f_c = 4.1$)	77.5%	77.5%	92.9%	62.7%	89.1%	87.5%	61.9%	70.7%
(D) Rotation → OctMix → RICAP	78.2%	78.3%	92.2%	62.2%	90.3%	89.8%	64.2%	71.2%
(E) Rotation → RICAP → OctMix	75.9%	75.6%	89.1%	59.0%	87.3%	86.8%	61.9%	69.3%
(F) Combined (B) and (C)	78.3%	78.3%	93.2%	62.2%	90.5%	89.3%	62.7%	71.6%
(G) Two-stage Model (図 3)	79.0%	79.4%	98.5%	62.9%	90.7%	89.9%	63.0%	71.3%

Rotation と組み合わせることで十分なバリエーションを獲得できたと考えられる。

5.3 OctMix と RICAP の特性の考察

前節の評価実験により, Rotation を併用する際には RICAP 及び OctMix が高精度を達成できることを示した。本節では, 両者の違いを考察した上で, 両者を併用することで高精度を達成する新たな手法を模索する。

まず, Rotation を併用する場合において, ベースライン (Ref), RICAP, OctMix のテスト精度を表 5 (上部) に示す。前節では検証用データに対する検証精度で議論したが, 本節ではテスト用データに対するテスト精度で議論する。検証用データは各エポック終了後に精度検証を行うが, テスト用データは全ての訓練が終了した際に精度検証を行う。すなわち, 前節 Last と同じ意味合いを持つが, データセットを事前に分割している点が異なる。また, 予測結果を全て保存していることから F 値による議論が可能となる。表 4 と比較して精度が前後しているのは, 検証用とテスト用で被験者が異なるためである。

表 5 上部より F 値に着目すると, RICAP と OctMix の主な違いは f_{stay} と f_{walk} である。RICAP は f_{stay} が 98.6% と高精度を達成している。一方で OctMix は f_{stay} が 92.9% となっており, Ref と比べても低下している。手法の特性上, 加重平均で合成する OctMix は stay のような, ほぼ一定の値を出力するケースに悪影響を及ぼしたと考えられる。一方, RICAP のように切り貼りする手法により, stay の精度が改善される点は興味深い知見である。次に, f_{walk} に着目すると, RICAP は Ref と比べると精度向上ができていたことから悪影響は無いが, OctMix の方が効果は高いことがわかる。以上を踏まえると下記の 2 点が知見として得られる。

- Rotation → RICAP は "stay" の検出に特に優れている。
- RICAP と OctMix には, やや異なる分類傾向がある。

5.4 OctMix と RICAP の併用

前節の知見を踏まえ, OctMix と RICAP を併用する手法を新たに提案する。DA 適用順序が大きく影響する可能性を考慮し, (D) Rotation → OctMix → RICAP と (E) Rotation → RICAP → OctMix の順序で DA する手法を実装した。また, DA を重ねがけするのではなく, (B) と (C) を併用する手法として (F) Combined (B) and (C) を実装した。

検証実験の結果を表 5 中段に示す。表 5 中段の結果を見ると, (E) の順序で DA を適用すると RICAP の利点を活かしきれず精度が低下する結果となった。対して, (D) と (F) は accuracy でわずかに (B) RICAP を上回った。しかしながら, RICAP の利点である f_{stay} の高さを活かしきれず, f_{walk}, f_{skip}, f_{stup}, f_{stdown} をやや改善する結果となった。

5.5 Two-stage Model

これまでの結果を総合し, (B) Rotation → RICAP を "stay" の検出に用い, (D) Rotation → OctMix → RICAP で残る 5 つの行動を分類する (G) Two-stage Model を新たに提案する。処理概要を図 3 に示す。本モデルでは, Classifier₁ は (B) で DA され "stay" を検出する用途で用い, Classifier₂ は (D) で

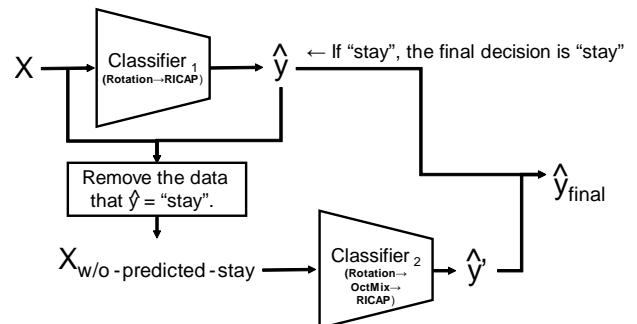


図 3 Two-stage Model の処理概要図

され、その他5行動の検出に用いられる。2段階のモデルとすることで、**RICAP**の強みを活かしつつ全体の高精度化を図った。

検証結果を表5の最下段に示す。accuracy及び平均F値で全手法最良の結果となっている。(A) ベースラインと比べるとaccuracyで7.2%、F値で7.9%の精度向上となった。

(B) **Rotation** → **RICAP** と比べてもaccuracyで0.9%、F値で0.8%の精度向上となった。

F値を見ると、 f_{stay} の高さは(B) **RICAP** と同等で活かされつつ、その他の分類でも高精度化が図られている。このことから、両者の特性を考慮した2段階の併用モデルが実現できているといえる。

6. おわりに

本研究では、スマートフォンセンシングによる基本行動認識に焦点を当て、新しいDA手法 **Octave Mix** を開発した。提案手法は複数の入力データを合成するDA手法であり、周波数分解を用いて従来手法の **mixup** を改良したものである。スマートフォンを用いた基本行動認識ベンチマークデータセット **HASC** を用いて、提案手法の有効性評価実験を行った結果、いくつかの有益な知見とともに提案手法の有効性を確認した。まず、単体のDAを用いる場合は **Octave Mix** が最も高い精度を実現したという点である。この結果から、周波数分解により **mixup** を改良することに成功したといえる。次に、スマートフォンが回転することを考慮し、**Rotation** によるDAを併用するケースでは、**RICAP** が大幅に精度向上に寄与するという点である。特に“stay”の検出精度が大幅に向上しF値98.6%を達成している。一方で、**RICAP** は **Octave Mix** とやや異なる分類傾向があることを考察した。最終的に、これらの特性を活かして **Octave Mix** と **RICAP** を併用した **Two-stage Model** を新たに提案し、最も高い推定精度が達成されることを確認した。

今回の実験では、スマートフォンセンシングによる基本行動認識を対象として検証を行っているため、提案手法の有効性検証は限定的である。今後の課題として、ウェアラブルセンシング全般に対する有効性検証実験を執り行い、**Two-stage Model** のような併用によるチューニング手法についても効率化を検討していく予定である。

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] O. D. Lara, M. A. Labrador, “A Survey on Human Activity Recognition using Wearable Sensors”, *IEEE Communications Surveys & Tutorials*, Vol. 15, No. 3, pp. 1192-1209 (2012).
- [2] J. Wang, Y. Chen, S. Hao, X. Peng, Lisha Hu., “Deep learning for sensor-based activity recognition: A survey”, *Pattern Recognition Letters*, Vol. 119, No. 1, pp. 3-11 (2019).
- [3] N. Kawaguchi, N. Ogawa, Y. Iwasaki, K. Kaji, T. Terada, K. Murao, S. Inoue, Y. Kawahara, Y. Sumi, N. Nishio, “HASC Challenge: Gathering Large Scale Human Activity Corpus for the Real-World Activity Understandings”, In *Proceedings of the 2nd Augmented Human International Conference*, pp. 1-5 (2011).
- [4] H. Zhang, M. Cisse, Y. N. Dauphin, D. Lopez-Paz, “mixup: Beyond Empirical Risk Minimization”, In *Proceedings of the International Conference on Learning Representations*, pp. 1-13 (2018).
- [5] R. Takahashi, T. Matsubara, K. Uehara, “RICAP: Random Image Cropping and Patching Data Augmentation for Deep CNNs”, In

Proceedings of the 10th Asian conference on Machine Learning, vol. 95, pp.786-798 (2018).

- [6] A. Krizhevsky, I. Sutskever, G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks”, In *Proceedings of the 25th International Conference on Neural Information Processing Systems*, Vol. 1, pp. 1097-1105 (2012).
- [7] T. DeVries, G. W. Taylor, “Improved Regularization of Convolutional Neural Networks with Cutout”, *arXiv preprint arXiv:1708.04552*, pp. 1-8 (2017)
- [8] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, “Random Erasing Data Augmentation”, *arXiv preprints arXiv:1708.04896*, pp. 1-10 (2017)
- [9] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, Q. V. Le, “AutoAugment: Learning Augmentation Policies from Data”, In *Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, (2019)
- [10] E. D. Cubuk, B. Zoph, J. Shlens, Q. V. Le, “RandAugment: Practical automated data augmentation with a reduced search space”, *arXiv preprints arXiv:1909.13719* (2019)
- [11] X. Zhang, Q. Wang, J. Zhang, Z. Zhong, “Adversarial AutoAugment”, In *Proceedings of the International Conference on Learning Representations*, pp. 1-13 (2020)
- [12] T. T. Um, F. M. J. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. M. Fietzek, D. Kulić, “Data augmentation of wearable sensor data for parkinson’s disease monitoring using convolutional neural networks”, In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*, pp. 216-220 (2017)
- [13] O. S. Eyobu, D. S. Han, “Feature Representation and Data Augmentation for Human Activity Classification Based on Wearable IMU Sensor Data Using a Deep LSTM Neural Network”, *Sensors*, vol. 18 (9), pp. 1-26 (2018)
- [14] G. Kalouris, E. I. Zacharaki, V. Megalooikonomou, “Improving CNN-based activity recognition by data augmentation and transfer learning”, In *Proceedings of the IEEE 17th International Conference on Industrial Informatics* (2019)
- [15] Y. Chen, H. Fan, B. Xu, Z. Yan, Y. Kalantidis, M. Rohrbach, S. Yan, J. Feng, “Drop an Octave: Reducing Spatial Redundancy in Convolutional Neural Networks with Octave Convolution”, In *Proceedings of the 2019 IEEE / CVF International Conference on Computer Vision* (2019)
- [16] K. Simonyan, A. Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, In *Proceedings of the International Conference on Learning Representations*, pp. 1-14 (2015)
- [17] T. Hasegawa, M. 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*, pp. 99-104 (2019)
- [18] D. P. Kingma, J. L. Ba, “Adam: a Method for Stochastic Optimization”, In *Proceedings of the International Conference on Learning Representations*, pp. 1-15 (2015).