

# 深層行動認識モデルのユーザ特化圧縮

Personalized Model Compression: Decomposing Large User-Independent Action Recognition Model into Small User-Specific Recognition Models

岩澤 有祐<sup>\*1</sup>      矢入 郁子<sup>\*2</sup>      松尾 豊<sup>\*1</sup>

Yusuke Iwasawa      Ikuko Eguchi Yairi      Yutaka Matsuo

<sup>\*1</sup>東京大学工学系研究科技術経営戦略学専攻

Graduate School of Technology Management for Innovation, The University of Tokyo

<sup>\*2</sup>上智大学理工学研究科理工学専攻

Graduate School of Science and Engineering, Sophia University

Deep nets are increasingly used in application on sensor based human action recognition, which improve recognition performance and enable various applications; however, discrepancy between hardware requirements of deep nets and capacity of current mobile devices appeared a major bottleneck to make it practice. This paper proposes novel concept for compressing deep action recognition models, i.e., personalized model compression, considering unique characteristics of sensor based action recognition. While existing model compression method aim to preserve all information or function of referenced model, the proposed concept purpose to decompose one large user-independent network to small user-specific networks. This paper explains the concept, discuss the relation between existing research areas, and show the empirical validation with a public sensor based action recognition dataset.

## 1. はじめに

スマートフォンのようにユーザの動きを阻害しないデバイスを利用した人間行動の認識技術は、パーベイシブ・ヘルスケアやコンテキストウェアシステムなどの次世代アプリケーションの中心的な技術要素である。センサによる行動認識は、多次元の系列データ  $X_t$  を入力としてある時刻  $t$  における行動  $y_t$  を予想する問題と考えられ、教師あり学習を利用した方法により定式化することができる。近年の研究では、画像認識や音声認識と同様に、センサベース行動認識においても深層学習を利用した手法（深層行動認識モデル）が高い精度を示すことが報告されている [Plötz 11, Yang 15]。本稿では、深層行動認識モデルを実用的に利用する上での課題の 1 つであるモデルサイズに関する問題を取り扱う。一般に深層学習によるモデルは大量のパラメータを持ち必要なメモリサイズも莫大となるため、スマートフォンのようにハードウェア制限がある端末上で直接利用することは難しい。スマートフォン上で深層学習を利用できるようにすることは行動認識に限らず音声認識や物体認識などの領域でも重要であり、学習済みのニューラルネットワークを小型化するさまざまな手法が提案されている [Gupta 15, Sainath 13, Chen 15, Hinton 15]。

本稿では、深層行動認識モデルを圧縮する方法として、行動認識の特徴を利用したユーザ特化圧縮という枠組みを提案しその有効性について議論する。ユーザ特化圧縮とは、ユーザ非依存なモデルを複数のユーザ特化したモデルに分解し圧縮する枠組みであり、1) 訓練データがないユーザに対して認識を行うことができる、2) モデルサイズの圧縮が可能である、という深層行動認識モデルに必要な 2 つの要件を実現することができる。2 章では、まずユーザ特化圧縮の有用性をセンサベース行動認識における背景を整理しながら述べたあと、提案した枠組みの新規性について関連研究について言及しながら論じる。3 章では、ユーザ特化圧縮を実現する上で最も単純な方法とし

て蒸留 [Hinton 15] を利用した方法について説明し、4 章で実験により提案した枠組みの有効性について実データを用いて検証する。最後に実験結果をふまえて今後の方向性について言及し本論文の結びとする。本研究の貢献は次の 3 点である。

- 深層行動認識モデルを圧縮する新たな方法として、ユーザ非依存なモデルを複数の小規模なユーザ特化モデルに分解圧縮するユーザ特化圧縮の枠組みを提案した。
- ユーザ特化圧縮を実現する最も単純な方法として、ユーザ非依存なモデルを模倣するように複数のユーザ特化モデルを学習させる方法を提案した。
- センサベース行動認識で最もよく利用される Opportunity データセットによる実験を行い、ユーザ特化圧縮が有効な場面があることを検証した。

## 2. ユーザ特化圧縮

行動認識問題において問題を難しくする特徴の 1 つが、学習データを取得できるユーザが限られているということである。例えば、あるユーザの転倒を教師あり学習により検出する問題を考えてみる。この場合、安全な環境においてデータ取得実験を行うことで擬似的な教師データを得るという方法は認識精度を高めるための 1 つの方向性である。しかし、アプリケーションを利用する可能性がある全ユーザに対してデータ取得実験を行うのは非現実的である。そのため、限られたユーザの学習データからより共通する特徴を取り出し、学習データがないユーザに対しても認識できる必要がある。通常、もっとも単純にはなるべく大規模なユーザのデータを実験的に収集し利用して学習を行うことで、新しいユーザに対しても適用可能なモデル（ユーザ非依存モデル）を学習する方法がとられている。

行動認識のもう 1 つの特徴が、このように学習したモデルをスマートフォンなどに搭載して実際に利用する場合では、ある特定のユーザに対しての認識精度があれば十分である点である。なぜならば、あるスマートフォンに搭載されたモデルが実際に認識する必要があるのは通常そのデバイスを持ち歩く特定のユーザと考えて良いためである。直感的に、ユーザに特化し

連絡先: 岩澤有祐, 東京大学工学系研究科技術経営戦略学専攻, 〒 113-8656 東京都文京区本郷 7-3-1, iwawasa@weblab.t.u-tokyo.ac.jp

たモデルはユーザ非依存なモデルと比較してより単純であり少ないパラメタで表現することができると考えられる．そのため、学習した非依存モデルの単純に圧縮するよりもむしろ一部の機能を取り出すように圧縮することで効率的な圧縮可能と予想される．

ユーザ特化圧縮は上記のような行動認識の特徴を活用し、学習済みのユーザ非依存モデルの情報を複数のユーザ特化モデルで代替することで、クライアント側で実行されるモデルが少ないパラメタで表現できるようにする枠組みである．ここで、各特化モデルは各クライアント端末で認識を行い、それゆえモデルの総パラメタ数よりもむしろこの特化モデルのパラメタ数のみが重要である．図 1 にユーザ特化圧縮の枠組みを示す．ユーザ特化圧縮は次の 2 つのステップから構成される．1) 教師データのあるユーザのデータを利用して 1 つのユーザ非依存モデルを利用する、2) ユーザ非依存モデルの情報を利用して、教師データの無いユーザを含むすべてのユーザごとに特化モデルを作成する．この際、各特化モデルのパラメタ数  $\|\theta_s\|$  がユーザ非依存モデルのパラメタ数  $\|\theta_e\|$  より小さくなるように学習する．このように、一旦ユーザ非依存なモデルを学習した後に特化させる 2 段階の学習により、次のような利点がある．

- ユーザ特化モデルを単に学習する場合と比較すると豊富な情報を利用可能であり、特に教師データの無いユーザに対して高い予測精度が期待できる
- ユーザ非依存モデルを単純に圧縮する場合と比較し、 $\|\theta_s\|$  をより減らすことができる

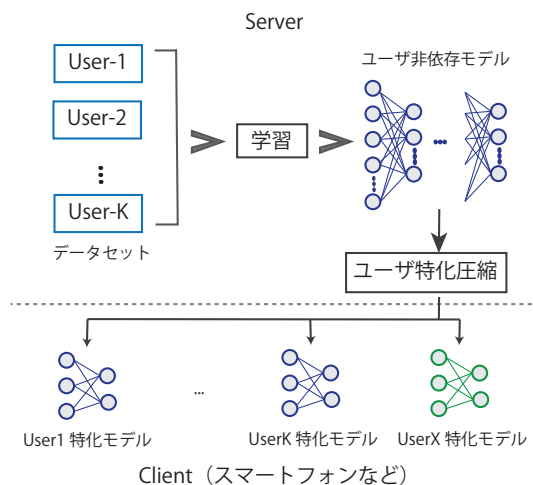


図 1: ユーザ特化圧縮の枠組み．

## 2.1 関連研究

### 2.1.1 センサベース行動認識における深層学習とその課題

画像認識や音声認識の領域などと同様に、センサベース行動認識の分野においても深層学習を利用した方法が良い認識精度を達成することがさまざまな研究で示されている [Plötz 11, Yang 15]．センサの系列データは高次元・非線形かつ画像などと比較しても人間が解釈しにくいいため、深層学習のようにデータから特徴を獲得するアプローチに関する研究が重要性を増すと考えられる．深層学習のモデルが必要とするメモリサイズと実際のモバイル機器の性能とのギャップは、深層行動認識モデルを実用的に利用する上で重大な障壁の 1 つである．スマートフォン上で直接処理するのではなくサーバ上にデータを送信しサーバ側で処理することでメモリサイズの問題を解消することはできるが、1) サーバ側の負荷の大きさ、2) オフラインで

の認識の難しさ、3) プライバシー、4) 通信によるバッテリー消費、などの理由から望ましくない場面がある．特に行動認識の場合、連続的にセンシングされる大量のデータを処理する必要があり、サーバ側での負荷および通信によるバッテリー消費が大きな問題となりうる．モバイル機器でも利用可能なようにモデルを圧縮する技術が重要となる．

### 2.1.2 ニューラルネットワークの圧縮

センサによる行動認識以外にも、ハードウェア制約をもつ端末上で深層学習のモデルを利用することの需要は大きく、ニューラルネットワークを圧縮する方法は数多く提案されている．代表的なものとしては、パラメタあたりのビット数を下げる [Gupta 15]、重みを低ランク近似する [Sainath 13]、ランダムに重みを共有させる [Chen 15]、あらかじめ豊富なパラメタで学習した教師モデルを模倣するように少ないパラメタ数のモデルを学習させる [Hinton 15] などの手法がある．いずれの方法も圧縮前のモデルのもつ情報や機能をすべて保存した上でモデルサイズを小さくすることを目指したものであり、本研究のように元のモデルがもつ機能を複数のモデルに分解することで担保する研究は筆者の知る限り存在しない．

### 2.1.3 ユーザ適応

特定の 1 つのモデルで認識精度を担保するのではなく、ユーザごとに特化した複数のモデルで認識精度を担保するという観点から見ると、ユーザ特化圧縮はユーザ適応 [Zhao 11] と関連がある．ユーザ適応では、学習データからなんらかのモデルを学習した後に、ユーザごとに用意した適応用の学習データを使いパラメタを修正したり特徴空間を変換したりすることでより特定のユーザに特化したモデルを構築する方法である．しかし、ユーザ適応が特定のユーザのデータを使って微調整することでモデルの認識精度を向上させることを主な目的とするのに対してユーザ特化圧縮はモデルサイズを小さくしようとする点で異なっている．

## 3. 蒸留によるユーザ特化圧縮

2 章で説明したユーザ特化圧縮の枠組みを実現する方法として、蒸留 [Hinton 15] を応用した方法を提案しその有効性を検証する．提案手法では、1 つの大規模なユーザ非依存モデル (教師モデル) を、教師モデルの活性状態を真似るように学習した複数のユーザ特化モデルへと分解する．蒸留によるユーザ特化圧縮は次の 2 つのステップから構成される．ここで、各関数  $f_t$  および  $f_s$  はそれぞれ異なるパラメタ  $\theta$  を持つニューラルネットワークで近似されたものであるとする．また、 $\{X_j^i, y_j^i\}$  をユーザ  $i$  の入出力のペア、 $n_i$  をユーザ  $i$  のサンプル数、 $\sigma$  をソフトマックス関数、 $L$  を損失関数とする．

ステップ 1: ユーザ非依存モデル (教師モデル) の学習

教師データがいたすべてのユーザのデータを利用し、ユーザに依存せずに認識できる関数  $f_t$  を学習する．定式的には、

$$f_t = \arg \min_f \sum_{i=1}^k \frac{1}{n_i} \sum_{j=1}^{n_i} L(y_j^i, \sigma(f(X_j^i))) + \Omega(\|f\|) \quad (1)$$

を最適化する．第 2 項目はなんらかの正則化項である．

ステップ 2: 教師モデルを真似るユーザ特化モデルの学習

学習した  $f_t$  から蒸留した知識と、特定ユーザのデータを利用してそれぞれのユーザに特化した認識精度を持つ関数  $f_s$  (ユーザ特化モデル) をユーザ数だけ学習する．定式的には、各ユー

が  $i$  に対して、

$$f_i = \arg \min_f \frac{1}{n_i} \sum_{j=1}^{n_i} [\lambda L(y_j^i, \sigma(f_i(X_j^i))) + (1 - \lambda) L(s_j^i, \sigma(f_i(X_j^i)))] \quad (2)$$

を最適化する．ここで、 $\lambda$  はどの程度教師を強く模倣するかを制御するパラメタである． $s_j^i$  はソフトターゲットと呼ばれ、

$$s_j^i = \sigma\left(\frac{f_i(x_j^i)}{T}\right) \quad (3)$$

で計算される、教師モデルの出力層の活性状態を滑らかにしたものである．式 2 にあるように、蒸留は生徒モデルが通常の入力出力のペアから定義される損失の度合い（第 1 項）に加えて教師モデルの出力を模倣するような正則化（第 2 項）を加えたものとして解釈することができる．このように単純に入力出力の間で定義される損失関数を最小化するのではなく、学習済みのモデルを真似るように制約を加えた上で学習することでよりパラメタの少ないモデルや少ないサンプルで認識精度を維持できることが報告されている [Hinton 15]．また、教師データがないユーザに対しては第 1 項目を単純に無視した上で、元のモデルを真似るように学習することが可能である．そのため、教師なしユーザに対してもモデルサイズを小さくした上で元の精度を保つことができるという利点も期待できる．

一方で、明らかな欠点はユーザ非依存モデルの学習だけでなくユーザ依存モデルを再学習するため学習時間が増加する点である．特に、上記の方法ではユーザ非依存モデルのために学習したパラメタ  $\theta_t$  は直接利用されず、ソフトターゲットを介してのみ利用されることになる．通常ユーザごとに最適なパラメタは異なるが何らかの関係はあると考えられ、学習済みの  $\theta_t$  を利用できれば特化モデルを効率的に学習できると考えられる．この点に関しては、4.2 節で議論する．

## 4. 実験

### 4.1 実験設定

提案手法の有効性検証のため、Opportunity Recognition Dataset<sup>\*1</sup>（以下 Opportunity データセット）による検証を行った．Opportunity データセットとはセンサベース行動認識において最もよく利用されるデータセットの 1 つであり、朝食を作るシナリオにおける 18 種類<sup>\*2</sup> の行動データが 4 人分提供されている．先行研究に従い、ウィンドウサイズを 30、重なり率を 0.5 としたスライディングウィンドウ方式によりサンプルを獲得し、4 人合計で 57,790 のサンプルを得た．

教師として利用するユーザ非依存モデルは [Yang 15] に従い、1) 入力層、2)3 回の畳み込み、プーリング、ReLU による活性化の繰り返し、3)400 個のユニットを持つ全結合層、4) 出力層とした．特に言及がない限り、教師モデル、特化モデル群、およびベースライン手法のすべてで学習率を 0.1 とした確率的降下法で 100 エポック学習した．また、模倣パラメタ  $\lambda = 0.3$ 、 $T = 5$  として実験を行った．

精度検証は、教師データがあるユーザ教師データがないユーザそれぞれに対しての認識精度を次の方法で検証した．1) ユー

\*1 <http://opportunity-project.eu/challenge>

\*2 open door 1, open door 2, close door1, close door 2, open fridge, close fridge, open dishwasher, close dishwasher, open drawer 1, close drawer 1, open drawer 2, close drawer 2, open drawer 3, close drawer 3, clean table, drink cup, toggle switch, and Null

ザ数 4 人のうち 3 人のデータを使い学習した教師モデルを式 1 を最適化するように学習する．2) 教師モデルに対して、それぞれのデータだけを利用したモデルを式 2 を最適化するように 4 個学習し、各モデルを対象ユーザのテストデータにより精度検証する．ここで、学習時に存在していたユーザに関しては教師データがあるものとして学習を行い、存在していなかったユーザについては教師データがないものとして学習した．また、テストデータは各ユーザの全データのうち 20% を利用し、残りの 80% を学習データとした．上記の方法を 1) の教師モデル学習に利用するユーザを 4 通り変更して繰り返し、教師データがあるユーザに関する評価を合計で  $4 \times 3$  回、教師データがないユーザに関する評価を 4 回行い、評価値の平均値を最終的なスコアとして用いた．

ベースライン手法として提案手法と同じ構造の 3 つのニューラルネットワークをそれぞれ次の方法で学習したものを利用し、モデルサイズを小さくした際の提案手法の有効性を定量的に検証した．1) ユーザ特化モデル：特定のユーザのデータのみを利用して式 1 を最適化、2) ユーザ非依存モデル：3 人のデータを利用して式 1 を最適化、3) DK：提案手法と同じ教師モデルで計算した  $s$  を使い、教師モデル作成と同じ 3 人のユーザのデータで式 2 を最適化．また、モデルサイズを小さくする際には簡単化のため畳み込みの圧縮は行わず、全結合層のユニット数を減らして実験を行った．例えば、圧縮率 0.05 とはもともと 400 個であったユニット数を 20 個まで減らして学習を行ったことを意味するものとする．

### 4.2 実験結果・考察

表 1 はモデルの圧縮率を 0.05 およびに 1.00 とした際の提案手法とベースライン手法の認識精度を比較したものである．各評価指標にはクラスごとの F 値の平均値 (F1-Macro) を利用し、最も精度が高かったものを太字で示した．結果として、教師ありユーザについては、1) モデルサイズを小さくした際に、ユーザ非依存モデルよりユーザ特化モデル、蒸留よりも提案手法が高精度である、2) ユーザ特化モデルより提案手法、ユーザ非依存モデルよりも蒸留を用いた方法が高精度であるという結果が得られた．それぞれの結果は、a) ユーザに特化させることでモデルサイズを小さくすることができる、b) 特化時に蒸留を利用することが有効であるということを示唆している．一方で、教師なしユーザに対しては蒸留を利用する有効性は確認できたもののユーザに特化させることで認識精度が向上することは確認できなかった．今回のデータの場合 4 人と人数が少なく特化させる有効性が少ない可能性がある、教師としたモデルが十分にユーザ非依存でなかった（表 2）などの原因が考えられるが、今後より検証する予定である．

図 2 は、圧縮率を変化させた際の認識精度の変化を示している．各線はそれぞれ、青：通常の蒸留による圧縮、赤：提案手法、点線：教師モデルの認識精度を示している．結果として、提案手法 ( $DK_p$ ) を利用することで元のモデルの 30% 程度のパラメタ数で同程度の精度を達成できていることがわかる．蒸留 ( $DK$ ) を用いた場合にはおよそ 50% 程度のときに精度が教師モデルと同程度になっており、提案手法により効率的に圧縮できていることがわかる．また、どちらの手法でも圧縮率が 20%（ユニット数が 80）を超えたあたりから精度が急激に低下しているが、これは今回の問題が 18 クラスであるのに対してユニット数が少なくなりすぎたことが要因と考えられる．今後はより大きなモデルで検証を行いたいと考えている．

実験結果より、1) ユーザ特化させることでより小さなモデルサイズで全体としての認識精度を保つことができる場面があること、2) ユーザ非依存モデルの知識を蒸留して特化モ

表 1: 提案手法とベースライン手法の精度比較

	教師あり		教師なし	
	0.05	1.00	0.05	1.00
ユーザ特化モデル	0.78	0.85	-	-
ユーザ非依存モデル	0.76	0.85	0.47	0.52
蒸留 [Hinton 15] (DK)	0.79	<b>0.88</b>	<b>0.49</b>	<b>0.53</b>
提案手法 ( $DK_p$ )	<b>0.81</b>	<b>0.88</b>	<b>0.49</b>	<b>0.53</b>

表 2: 教師モデルの認識精度

	mean	std
訓練に含まれるユーザのテストデータ	0.87	0.01
訓練に含まれないユーザのテストデータ	0.54	0.08

ルの学習に利用することでより小さなモデルサイズで認識精度を維持することができることの2点が確認できた。一方で、蒸留を用いてユーザ特化圧縮する際にユーザ非依存モデルの情報で利用されるのは出力の確率のみであり、重みや中間層の情報をしていないのは非効率的であると考えられる。図3は、特化モデルのパラメタの初期値を教師モデルの学習結果とした場合 ( $DK_p^{copy}$ ) とせずにランダムに初期化した場合 ( $DK_p$ ) で必要なデータ数を比較したものである。結果から、ユーザ非依存モデルの学習結果を初期値とすることでより少ない訓練データ数で認識精度向上が見込めることがわかる。特化モデルのパラメタを減らす場合にも、学習済みの重みを低ランク近似 [Sainath 13] した上で再学習を行う等の工夫の導入することで効率的にユーザ特化モデルを学習できると考えられる。また、一般に深層モデルでは入力層に近い層ほどドメインに依存した活性が、出力層に近い層ほどタスクに特化した活性が起こりやすいことが知られている [Yosinski 14]。そのため、出力層の活性を模倣するように学習するよりも、入力に近い層の活性を模倣するように学習の方がユーザ (すなわちドメイン) に特化した学習を行う上では適切である可能性がある。Romeroらは出力層だけでなく中間層の活性状態も教師モデルを模倣するように学習する Fitnets を提案し、より深いモデル学習に有効であることを示している [Romero 15]。今後は同様の手法の本問題に対する有効性を検証予定である。

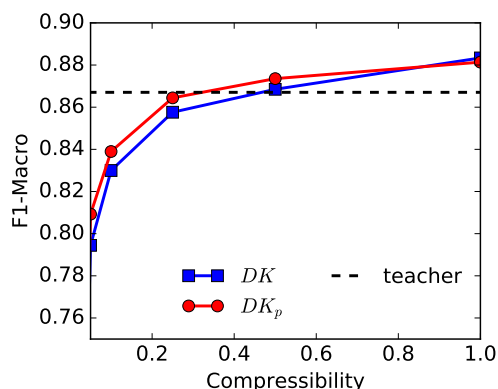


図 2: 圧縮率を変化させた際の認識精度の変化。

## 5. おわりに

本稿では、深層学習により学習したユーザ非依存な行動認識モデルを圧縮させる新たな方法として、大規模なユーザ非依存モデルを小規模な複数のユーザ特化モデルへと圧縮するユー

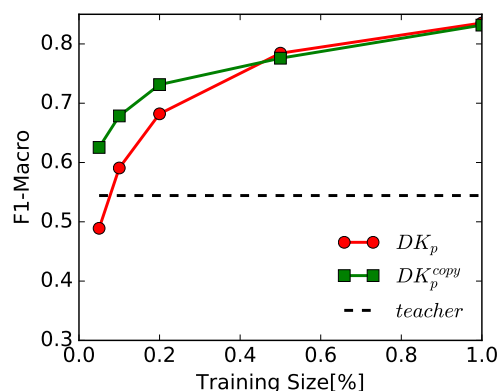


図 3: 圧縮に利用するデータ数を変化させた際の認識精度。

ザ特化圧縮提案しその周辺分野との関連を論じた。また、最も単純な実現方法として蒸留を用いた手法を提案しユーザ特化によりモデル圧縮が見込める場合があることを確認した。今後は、より大規模なデータでの評価や必要なデータ数およびモデルサイズの観点からより効率の良い手法を検討していく予定である。

## 参考文献

- [Chen 15] Chen, W., Wilson, J. T., Tyree, S., Weinberger, K. Q., and Chen, Y.: Compressing neural networks with the hashing trick, in *Proc. of ICML* (2015)
- [Gupta 15] Gupta, S., Agrawal, A., Gopalakrishnan, K., and Narayanan, P.: Deep learning with limited numerical precision, *arXiv preprint arXiv:1502.02551* (2015)
- [Hinton 15] Hinton, G., Vinyals, O., and Dean, J.: Distilling the knowledge in a neural network, *arXiv preprint arXiv:1503.02531* (2015)
- [Plötz 11] Plötz, T., Hammerla, N. Y., and Olivier, P.: Feature learning for activity recognition in ubiquitous computing, in *Proc. of IJCAI*, No. 1, pp. 1729–1734 (2011)
- [Romero 15] Romero, A., Ballas, N., Kahou, S. E., Chas-sang, A., Gatta, C., and Bengio, Y.: Fitnets: Hints for thin deep nets, in *Prog. of ICLR* (2015)
- [Sainath 13] Sainath, T. N., Kingsbury, B., Sindhwani, V., Arisoy, E., and Ramabhadran, B.: Low-rank matrix factorization for deep neural network training with high-dimensional output targets, in *Proc. of ICASSP, IEEE International Conference on*, pp. 6655–6659 IEEE (2013)
- [Yang 15] Yang, J. B., Nguyen, M. N., San, P. P., Li, X. L., and Krishnaswamy, S.: Deep Convolutional Neural Networks On Multichannel Time Series For Human Activity Recognition, in *Proc. of IJCAI* (2015)
- [Yosinski 14] Yosinski, J., Clune, J., Bengio, Y., and Lipson, H.: How transferable are features in deep neural networks?, in *Proc. of NIPS*, pp. 3320–3328 (2014)
- [Zhao 11] Zhao, Z., Chen, Y., Liu, J., Shen, Z., and Liu, M.: Cross-people mobile-phone based activity recognition, in *Proc. of IJCAI*, Vol. 11, pp. 2545–250 Citeseer (2011)