

人間行動認識における セグメント情報を用いた自己教師あり学習

近藤圭¹ 武仲紘輝¹ 長谷川達人¹

概要：深層学習を用いたセンサによる行動認識では、頑健な深層学習モデルを作成するために大量のアノテーションされた時系列データが必要となる。ウェアラブルデバイスの普及により大規模なデータセットの収集は可能となったが、それらのデータのアノテーションは依然として課題となっている。この問題に対して、画像認識分野では自己教師あり学習を用いた手法が多く提案されている。しかし、行動認識分野においては活発に研究が行われていない。そこで本研究ではセンサによる行動認識において、画像認識分野で提案されている自己教師あり学習手法の実装、検証を行う。また、センサデータのセグメントの情報を使った新たな自己教師あり学習手法を提案する。既存の自己教師あり学習と比較した結果、ラベルありデータによる Fine Tuning において、提案手法の有効性が確認された。

Self-Supervised Learning using Segment Information for Human Activity Recognition

KEI KONDO¹ KOKI TAKENAKA¹ TATSUHITO HASEGAWA¹

1. はじめに

センサを用いた行動認識ではデバイスから取得したセンサデータを用いて、そのデバイスを装着しているユーザの行動を推定する。ユーザの行動の推定はユーザ活動の補助や個人認識に利用することができるため、幅広い分野で応用が期待されている。近年では深層学習が様々な分野で使われるようになり、行動認識においても深層学習を用いた手法が多く提案されている。一般的に深層学習において、頑健な深層学習モデルを作成するためには大量のアノテーションされたデータが必要となる。行動認識においてはスマートフォンやウェアラブルデバイスの普及により大規模なデータセットの収集が可能となったが、それらのデータのアノテーションは未だ人為的に行う必要がある。

アノテーションされていないデータから深層学習モデルを学習する手法として、半教師あり学習や自己教師あり学習がある。半教師あり学習は少量のラベルありデータと大量のラベルなしデータを組み合わせてモデルを訓練する。自己教師あり学習は大量のラベルなしデータに対してモデルが自動的にラベル付けを行い、そのラベルを元にモデルを訓練する。どちらの学習法もラベルなしデータからモデルを学習させるという点で同じだが、半教師あり学習はラベルありデータが必要となる。画像などと異なり、センサデータは人の目でアノテーションを行うのが難しく、専門

的な知識も必要となってくる。そのため、本研究ではアノテーションされていないデータから深層学習モデルの学習が行える自己教師あり学習を中心に研究を行う。

近年では多くの自己教師あり学習手法が提案されている。Chen ら [1] は深層学習モデルから得られた特徴ベクトルを用いて対照学習を行う手法 SimCLR を提案している。また、Grill ら [2] はネガティブサンプルを使わずに対照学習を行う手法を提案している。その他多くの自己教師あり学習手法が提案されているが、多くは画像認識分野であり、行動認識分野においてはあまり活発に研究されていないのが現状である。

本研究では画像認識分野において提案されている自己教師あり学習手法を行動認識に適用し、有効性を評価する。その上で SimCLR を行動認識向けに改良した手法を提案する。SimCLR では同じデータ同士の特徴ベクトルが似るように対照学習を行う。一方で提案手法では同一ユーザの時間的に近傍なウィンドウは関連性が高いという仮説のもと、時間的に近傍な異なるデータ同士で対照学習を行う。時間的に近傍なデータを用いた新たな自己教師あり学習手法を提案し、有効性を示すことが本研究の目的である。本研究の貢献は以下の4点である。

- SimCLR にセグメント情報を取り入れた手法を提案した。
- 画像認識分野などで提案されている自己教師あり学習手法を行動認識に適用し、提案手法が有効であることを確認した。

¹ 福井大学大学院工学研究科
Graduate School of Engineering University of Fukui

- ラベルありデータによる Fine Turning において、提案手法が教師あり学習やベースとなった SimCLR よりも高精度であることを確認した。
- 異なるデータセットによる転移学習において、提案手法が教師あり学習より高精度であることを確認した。

2. 関連研究

2.1 深層学習を用いたセンサによる行動認識

センサによる行動認識はセンサデータから特徴抽出を行い、それらを分類することで行われる。従来は人の手で特徴量の設計が行われていたが [3] [4], 深層学習により特徴量の設計が自動で行われるようになった。近年では深層学習を用いた行動認識手法として, Convolutional Neural Network (CNN) を用いた手法が多数提案されている [5] [6] [7]. 行動認識に特化したモデルも提案されており, Long ら [8] はセンサデータに対して異なる時間窓を適用し, それぞれを別々のモデルに入力して特徴量を合わせることで空間的な特徴と時間的な特徴を捉える CNN モデルを提案している。また, Ma ら [9] はマルチモーダル行動認識に対して, Attention ベースの CNN と Gated Recurrent Unit (GRU) を組み合わせたモデルを提案している。

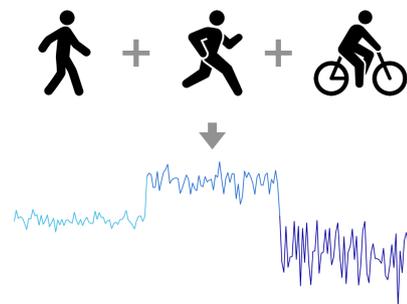
2.2 自己教師あり学習

自己教師あり学習は事前タスクとターゲットタスクの 2 つのプロセスから構成される。事前タスクではラベルなしデータから自動的にラベルを生成し, そのラベルを用いてモデルの学習を行う。ターゲットタスクでは事前タスクで学習したモデルの重みを使い, ラベルありデータで Fine Tuning を行う。事前タスクによって, データの異なる部分がどのように相互作用するのかを学習させることでデータの表現を学習することができ, その後のターゲットタスクの精度を向上させることができる。

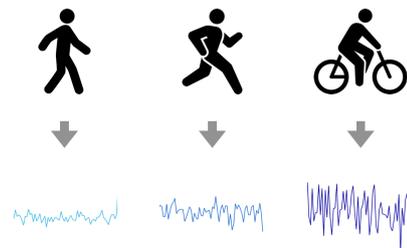
画像認識分野では多くの自己教師あり学習手法が提案されており, その中でも対照学習を使った手法が高精度を達成している。Chen ら [1] は入力データに対して異なるデータ拡張を適用した 2 つのデータを作成し, それらの特徴ベクトルが似るように, また他のデータの特徴ベクトルとは異なるように対照学習をする SimCLR を提案している。Dwibedi ら [10] は SimCLR にメモリバンクを適用した SimCLRv2 [11] を改良し, NNCLR を提案している。NNCLR では入力データの特徴ベクトルとメモリバンク内の似た特徴ベクトルとで対照学習を行うことで既存手法より精度を向上させた。また SimCLR のようなネガティブサンプルを使うような手法ではなく, ポジティブサンプルのみで対照学習を行う手法も提案されている。Grill ら [2] は 2 つのモデルと Predictor という特徴ベクトルを予測する機構を使い, ポジティブサンプルのみで対照学習を行う

BYOL を提案している。BYOL では 2 つのモデルを用い, 一方のモデルから得られた特徴ベクトルをもう一方のモデルが予測することで高精度を達成している。Chen ら [12] は BYOL に似た SimSiam を提案している。SimSiam では BYOL を簡素化したアーキテクチャであり, Predictor と勾配停止を組み合わせることでシンプルながらも高精度を達成している。

行動認識分野では画像認識分野や音声認識分野の手法を適用した研究がいくつか報告されている。武仲ら [13] は個々のデータを個別のクラスとみなして対照学習を行う Instance Discrimination (ID) [14] を行動認識に適用し, さらに ID をセンサデータのセグメントごとに行う Segment Discrimination (SD) を提案している。Khaertdinov ら [15] や Tang ら [16] は SimCLR を行動認識に適用している。Qian ら [17] は BYOL や SimSiam, NNCLR を行動認識に適用し, 有効なデータ拡張やモデルを検証している。Saeed ら [18] はマルチタスクで入力データにどのデータ拡張が適用されているかを推測する手法 [19] を行動認識に適用し, 少数ラベルありデータによる Fine Turning における精度や有効なモデル構造を検証している。Harish ら [20] は一部分をマスクしたデータをモデルに入力し再構築を行う手法 [21] を行動認識に適用している。Haresamudram ら [22] はデータの特徴ベクトルから次のタイムスタンプの特徴ベクトルを推測する手法である CPC [23] を行動認識に適用している。



(a) 複数の行動を連続して計測



(b) 行動ごとに計測

図 1: センサを用いた行動認識の計測手法

3. 提案手法

3.1 問題設定

センサを用いた行動認識ではデータを計測する方法として、行動ごとに計測する方法と複数の行動を連続して行い計測する方法がある(図1). 通常の自己教師あり学習ではラベルのないデータを用いることを前提とする. 本研究ではこの制約を緩和し、行動の種類を示すラベルは分からないが、図1bのように行動の切り替わるタイミングはアノテーションを行えるという環境を想定する. 実運用を考えると、行動の種類をアノテーションすることはユーザに一定のワークロードを強いることになり、実用的ではない. 一方、行動の切り替わりのみをアノテーションする場合は、行動が変わるタイミングでワンタップなどの単純な操作を行うだけで良いため、ユーザの負担を大幅に軽減することが可能となる. また、本研究では行動ごとに計測された一連のデータをセグメントと定義する.

3.2 セグメントを用いた対照学習

提案手法の概要を図2に示す. 提案手法は SimCLR (図2a) にセンサデータのセグメント情報を適用した手法である. SimCLR では1つの入力データから異なるデータ拡張を適用した2つのデータを作成し、それらの特徴量と学習バッチ内の他のデータの特徴量とを対照学習することでデータの表現を学習している. すなわち、データのインスタンス空間で学習を行う ID [14] に基づいた手法であるといえる. SimCLR の損失関数を式1に示す.

$$\ell_i = -\log \frac{\exp(\text{sim}(z_i, z_j)/\tau)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp(\text{sim}(z_i, z_k)/\tau)} \quad (1)$$

z_i と z_j は1つの入力データに異なるデータ拡張を適用して得られた特徴ベクトル、 z_k はバッチ内の他のデータの特徴ベクトルである. sim はベクトルの類似度を計算する関数であり、本研究では cosine 類似度を使用している. $\mathbf{1}_{[k \neq i]} \in \{0, 1\}$ は $k \neq i$ のときに1になる指示関数であり、 N はバッチサイズ、 τ は温度パラメータである.

提案手法(図2b)では同一セグメント内の異なるデータ同士で対照学習を行う. センサを用いた行動認識では一連のデータを sliding-window 方式などを使い、一定のウィンドウ幅で分割して、モデルの訓練に用いる. そのため、同一セグメント内のデータ同士は関連性が高い. しかし、SimCLR ではデータのインスタンス同士で対照学習を行うため、同一セグメント内のデータ同士を遠ざけるように学習してしまい、有用な特徴表現を学習できないことが考えられる. そこで提案手法では同一セグメント内のデータ同士の特徴ベクトルが似るように、異なるセグメントのデータの特徴ベクトルとは異なるように学習を行う. 通常の SimCLR は ID に基づいた手法であるが、提案手法は SimCLR をセンサデータのセグメントベースに学習を行う

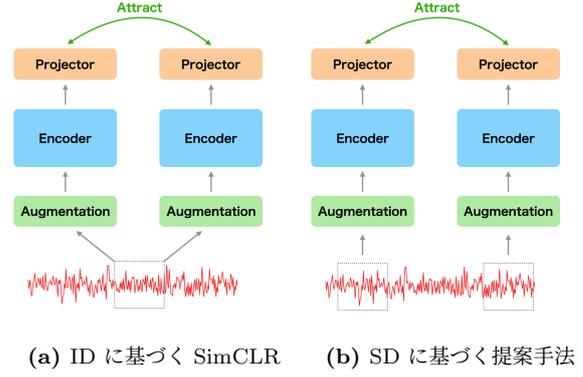


図2: 提案手法概要

SD [13] に基づいた手法に改良した点で異なる. 提案手法の損失関数を式2に示す.

$$\ell_i = -\log \frac{\exp(\text{sim}(z_1^{s_i}, z_2^{s_i})/\tau)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} \exp(\text{sim}(z_1^{s_i}, z^{s_k})/\tau)} \quad (2)$$

$z_1^{s_i}$ と $z_2^{s_i}$ は同一セグメント s_i 内の異なるデータにデータ拡張を適用して得られた特徴ベクトル、 z^{s_k} は異なるセグメントのデータにデータ拡張を適用して得られた特徴ベクトルである.

4. 実験設定

提案手法の有効性を検証するために既存の自己教師あり学習手法と比較して検証を行った. 本章では実験設定の詳細を述べる.

4.1 データセット

提案手法の有効性を HASC [24], UCI-HAR [25], USC-HAD [26] の3種類のデータセットを用いて評価する.

HASC[24] はスマートフォンなどのウェアラブルデバイスを用いて人間の基本行動を計測したベンチマークデータセットである. 基本行動は、静止、歩行、走行、スキップ、階段上り、階段下りの6種類あり、加速度センサやジャイロセンサ、磁気センサなどで計測されている. 本実験では Basic Activity 内からサンプリング周波数が 100Hz かつ iPhone で計測されたデータを抽出し、加速度センサで計測されたデータを使用した. 前処理としてデータの前後5秒をトリミングし、ウィンドウサイズを256サンプル、ストライド幅を256サンプルとして、sliding-window 方式でデータを分割した.

UCI-HAR[25] はスマートフォンを用いて日常生活動作を計測したベンチマークデータセットである. 30名の被験者によって収集された加速度データやジャイロデータによって構成されている. 日常生活動作は、歩く、階段上り、階段下り、着席、静止、臥床の6種類である. データセットはサンプリング周波数 50Hz で計測されており、各データは128サンプルのウィンドウに区切られた状態で提供されている. 本実験では加速度データを使って検証を行った.

加速度データには total acceleration と estimated body acceleration があるが, total acceleration を使用した。

USC-HAD[26] はウェアラブルデバイスを用いて日常生活動作を計測したベンチマークデータセットである。14名の被験者によって収集された加速度データやジャイロデータによって構成されている。日常生活動作は, 前を歩く, 反時計回りに歩くなどの12種類である。本実験では加速度データを使って検証を行った。

いずれのデータセットも, 被験者ごと, 行動ごとにセンサデータが時系列で計測されている。本研究では, これらのデータセットのように被験者ごと, 行動ごとに計測ファイルが分割されている環境下で, 行動ラベルがアノテーションされていないという設定で実験を行う。

4.2 モデル構造

本実験では Qian ら [17] の検証結果に基づき, Encoder に3層の畳み込みブロックを用いた。それぞれのブロックは畳み込み層, Batch Normalization, ReLU 関数, Max-Pooling 層で構成されており, 1つ目の畳み込みブロックにのみ Dropout 層を追加した。また, Classifier を Global Average Pooling (GAP) と1層の全結合層とした。Qian らのモデルでは Classifier に Flatten を用いていたが, 事前検証により GAP の方が高精度であったため採用し, また SimCLR などの対照学習を行う手法は Encoder からの出力を損失関数を適用する空間に投影するために Projector (図2上部) を用いる。本実験で用いる Projector は Qian [17] らと同様に手法ごとに異なり, 数層の全結合層, Batch Normalization, ReLU 関数とした。

4.3 モデルの訓練

本実験では提案手法及びその他の自己教師あり学習手法に適用するデータ拡張を統一して実験を行った。Khaertdinov ら [15] の検証結果に基づき, 事前タスクに用いるデータ拡張をデータセットごとに設定し, UCI-HAR と USC-HAD には {Jittering, Scaling, Permutation}, HASC には {Jittering, Scaling, Rotate} の3種をそれぞれ適用した。適用する際には Jittering のみ確率 $p = 1$ で適用し, それ以外のデータ拡張は確率 $p = 0.5$ で適用した。用いたデータ拡張の詳細について以下に示す。

- **Jittering:** 信号にランダムなガウシアンノイズを加算する。
- **Scaling:** 信号に正規分布からサンプリングした値を乗算する。
- **Rotation:** 信号にランダムな三次元回転を適用する。
- **Permutation:** 信号を一定の区間に分割し, ランダムに並べ替える。

訓練時の最適化関数には Adam を用い, 学習率を 0.001

表 1: 各データセットの被験者数の内訳

データセット	訓練	検証	テスト
HASC	80	20	30
UCI-HAR	19	5	6
USC-HAD	8	2	4

表 2: 各手法の推定精度

	HASC	UCI-HAR	USC-HAD
TPN	79.48	95.65	51.05
ID	74.90	96.25	46.06
SD	75.62	96.01	43.82
SimCLR	80.20	92.07	48.41
NNCLR	81.85	95.51	45.24
BYOL	80.59	85.54	35.48
Proposed	86.07	95.95	51.36
Supervised	85.25	96.62	50.26

とした。ID, SD 以外の手法は事前タスク時のエポック数を 150, ID, SD は 300 とし, ターゲットタスク (行動分類) 時のエポック数を 300 とした。SimCLR などのネガティブサンプルを使う対照学習手法はバッチサイズによって精度が変化するため, バッチサイズを HASC は 512, UCI-HAR と USC-HAD は 256 に統一して訓練を行った。

4.4 評価方法

本実験では Hold-out 法によって各手法の評価を行うため, データセットを被験者単位で分割した。各データセットの訓練, 検証, テストデータの内訳を表1に示す。事前タスク時には訓練データを, ターゲットタスクには訓練, 検証データを用い, テストデータで評価した。ただし, 事前タスク時には訓練データのラベルは使わずにデータのみを用いた。評価指標には各試行のテストデータの平均マイクロ F 値を用いた。また, 本実験ではベースラインとして, 教師あり学習 (事前タスクを行わずにターゲットタスクのみを行う) も行い評価した。さらに, 各手法がデータから有効な特徴表現を学習しているかを評価するために, 事前タスク後のターゲットタスク時には Backbone のパラメータを固定して検証を行った。

5. 実験結果

5.1 推定精度

事前タスクによる影響をターゲットタスクにおける F 値を用いて評価する。本実験では同じデータセットで事前タスクとターゲットタスクを行った。

表2に各手法のターゲットタスクにおける精度を示す。提案手法は HASC, USC-HAD において最も精度が高かった。ベースとなった SimCLR よりも精度が高いことから, セグメント情報を使った対照学習は有効であることが分かる。また UCI-HAR, USC-HAD においては HASC と比べ

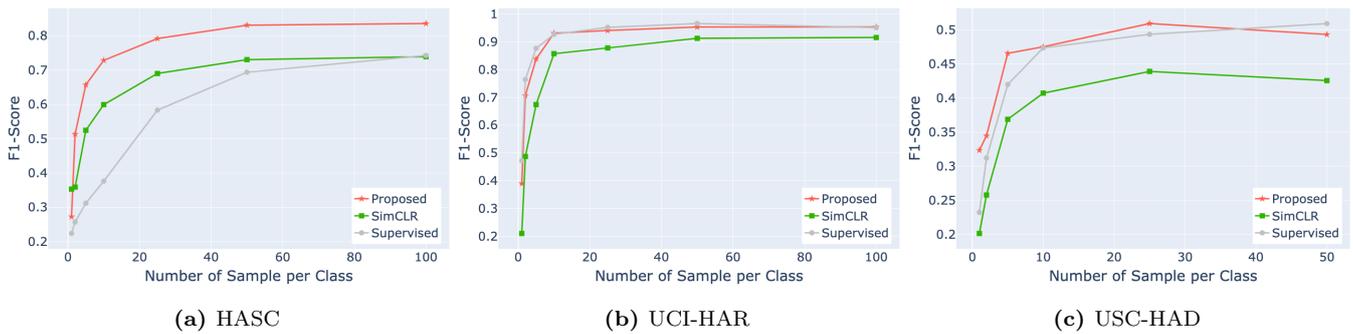


図 3: 半教師あり学習シナリオにおける精度

て、あまり精度が高くなかった。これは提案手法がセグメント同士で対照学習をしているために、HASC よりセグメントの数が十分でない UCI-HAR, USC-HAD では十分な対照学習ができなかったことが考えられる。

5.2 半教師あり学習

センサデータのアノテーションは画像などより困難であるため、少量のラベルありデータかつ大量のラベルなしデータという半教師あり学習シナリオにおいて、提案手法の性能を評価する。本実験ではターゲットタスクにおいて、クラスごとのデータ数を制限して検証を行った。

図 3 に半教師あり学習シナリオにおける各手法の精度を示す。HASC では提案手法が教師あり学習やベースとなった SimCLR よりも高精度であった。USC-HAD では各クラスのデータ数が 10 以下の場合において、提案手法が教師あり学習よりも高精度であった。UCI-HAR では提案手法は教師あり学習とあまり差がなかった。また、いずれのデータセットにおいても提案手法はクラスごとのデータ数が 50 または 25 程度で表 2 と同程度の精度を達成しており、提案手法の半教師あり学習シナリオにおける有効性が示された。

5.3 転移学習

提案手法の転移学習における性能を評価する。転移学習ではあるデータセットで事前タスクを行ったモデルに対して、別のデータセットでターゲットタスクを行う。本実験では HASC で事前タスクを行い、USC-HAD でターゲットタスクを行った。また、ベースラインとしての教師あり学習は HASC で行動分類を行ったモデルを用い、USC-HAD でターゲットタスクを行った。

表 3 に各手法の転移学習での精度を示す。SimCLR, BYOL 以外の手法は教師あり学習より精度が上回っているため、転移学習において提案手法は有効であることがわかる。また表 2 と比較すると、ID, SD, BYOL は同一データセットで事前タスクとターゲットタスクを行うよりも精度が高くなっている。ID, SD, BYOL の共通点として、ネガティブサンプルを使わない対照学習であることが挙げられ

表 3: 転移学習での精度

	USC-HAD
TPN	48.18
ID	50.39
SD	52.75
SimCLR	40.36
NNCLR	42.56
BYOL	40.26
Proposed	45.93
Supervised	42.74

る。このことから SimCLR などの手法はネガティブサンプルを使うために、訓練データに特化した特徴表現が得られ、一方で BYOL などの手法はより汎用的な特徴表現が得られることが考えられる。

6. まとめ

本研究では人間行動認識におけるセグメント情報を用いた自己教師あり学習手法の提案を行った。既存の手法と比較した結果、HASC, UCI-HAR データセットにおいて最も精度が高かった。また半教師あり学習シナリオにおいても高精度を達成しており、提案手法の有効性が示された。一方で提案手法はセグメント同士で対照学習させているため、セグメントの数が少ない場合は高い効果が得られない可能性が考えられる。また、転移学習においてはネガティブサンプルを使わない対照学習手法が有効であることが示唆された。今後はセグメント以外の情報を使う、損失関数を変えるなどして、提案手法の改良を行いたい。

謝辞

本研究の一部は、JSPS 科学研究費助成事業基盤研究 (C) (20K03084) の助成によるものである。ここに謝意を表す。

参考文献

- [1] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In Hal Daumé III and Aarti Singh, editors, *In International Conference on Machine Learning*, Vol. 119

- of *Proc. Machine Learning Research*, pp. 1597–1607. PMLR, 2020.
- [2] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, Carl Doersch, Bernardo Avila Pires, Zhaohan Guo, Mohammad Gheshlaghi Azar, Bilal Piot, koray kavukcuoglu, Remi Munos, and Michal Valko. Bootstrap your own latent - a new approach to self-supervised learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, Vol. 33, pp. 21271–21284. Curran Associates, Inc., 2020.
 - [3] Ferhat Attal, Samer Mohammed, Mariam Dedabrishvili, Faicel Chamroukhi, Latifa Oukhellou, and Yacine Amirat. Physical human activity recognition using wearable sensors. *Sensors*, Vol. 15, No. 12, pp. 31314–31338, 2015.
 - [4] Akram Bayat, Marc Pomplun, and Duc A. Tran. A study on human activity recognition using accelerometer data from smartphones. *Procedia Computer Science*, Vol. 34, pp. 450–457, 2014.
 - [5] Ming Zeng, Le T. Nguyen, Bo Yu, Ole J. Mengshoel, Jiang Zhu, Pang Wu, and Joy Zhang. Convolutional neural networks for human activity recognition using mobile sensors. In *6th International Conference on Mobile Computing, Applications and Services*, pp. 197–205, 2014.
 - [6] Song-Mi Lee, Sang Min Yoon, and Heeryon Cho. Human activity recognition from accelerometer data using convolutional neural network. In *2017 IEEE International Conference on Big Data and Smart Computing (BigComp)*, pp. 131–134, 2017.
 - [7] Francisco Javier Ordonez and Daniel Roggen. Deep convolutional and lstm recurrent neural networks for multimodal wearable activity recognition. *Sensors (Basel, Switzerland)*, Vol. 16, , 2016.
 - [8] Jun Long, Wuqing Sun, Zhan Yang, and Osolo Ian Raymond. Asymmetric residual neural network for accurate human activity recognition. *Information*, Vol. 10, No. 6, 2019.
 - [9] HaoJie Ma, Wenzhong Li, Xiao Zhang, Songcheng Gao, and Sanglu Lu. Attnsense: Multi-level attention mechanism for multimodal human activity recognition. In *Proc. Twenty-Eighth International Joint Conference on Artificial Intelligence, IJCAI-19*, pp. 3109–3115. International Joint Conferences on Artificial Intelligence Organization, 7 2019.
 - [10] D. Dwibedi, Y. Aytar, J. Tompson, P. Sermanet, and A. Zisserman. With a little help from my friends: Nearest-neighbor contrastive learning of visual representations. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9568–9577, Los Alamitos, CA, USA, oct 2021. IEEE Computer Society.
 - [11] Ting Chen, Simon Kornblith, Kevin Swersky, Mohammad Norouzi, and Geoffrey E Hinton. Big self-supervised models are strong semi-supervised learners. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, Vol. 33, pp. 22243–22255. Curran Associates, Inc., 2020.
 - [12] Xinlei Chen and Kaiming He. Exploring simple siamese representation learning. In *Proc. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15750–15758, June 2021.
 - [13] 紘輝武仲, 達人長谷川. 加速度センサーデータを用いた深層学習モデルの教師なし表現学習手法. Technical report, 第 101 回 MBL・第 87 回 ITS 合同研究発表会, nov 2021.
 - [14] Zhirong Wu, Yuanjun Xiong, Stella X. Yu, and Dahua Lin. Unsupervised feature learning via non-parametric instance discrimination. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
 - [15] Bulat Khaertdinov, Esam Ghaleb, and Stylianos Asteriadis. Contrastive self-supervised learning for sensor-based human activity recognition. In *2021 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 1–8, 2021.
 - [16] Chi Ian Tang, Ignacio Perez-Pozuelo, Dimitris Spathis, and Cecilia Mascolo. Exploring contrastive learning in human activity recognition for healthcare. *ArXiv*, Vol. abs/2011.11542, , 2020.
 - [17] Hangwei Qian, Tian Tian, and Chunyan Miao. What makes good contrastive learning on small-scale wearable-based tasks?, 2022.
 - [18] Aaqib Saeed, Tanir Ozcelebi, and Johan Lukkien. Multi-task self-supervised learning for human activity detection. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, Vol. 3, No. 2, jun 2019.
 - [19] C. Doersch and A. Zisserman. Multi-task self-supervised visual learning. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2070–2079, Los Alamitos, CA, USA, oct 2017. IEEE Computer Society.
 - [20] Harish Haresamudram, Apoorva Beedu, Varun Agrawal, Patrick L. Grady, Irfan Essa, Judy Hoffman, and Thomas Plötz. Masked reconstruction based self-supervision for human activity recognition. In *Proc. 2020 International Symposium on Wearable Computers, ISWC '20*, pp. 45–49, New York, NY, USA, 2020. Association for Computing Machinery.
 - [21] Weiran Wang, Qingming Tang, and Karen Livescu. Unsupervised pre-training of bidirectional speech encoders via masked reconstruction. In *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6889–6893, 2020.
 - [22] Harish Haresamudram, Irfan Essa, and Thomas Ploetz. Contrastive predictive coding for human activity recognition. *Proc. ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, Vol. 5, pp. 1 – 26, 2021.
 - [23] Aäron van den Oord, Yazhe Li, and Oriol Vinyals. Representation learning with contrastive predictive coding. *CoRR*, Vol. abs/1807.03748, , 2018.
 - [24] 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.
 - [25] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, and Jorge Luis Reyes-Ortiz. A public domain dataset for human activity recognition using smartphones, 2013.
 - [26] Mi Zhang and Alexander Sawchuk. Usc-had: A daily activity dataset for ubiquitous activity recognition using wearable sensors. In *ACM International Conference on Ubiquitous Computing (UbiComp) Workshop on Situation, Activity and Goal Awareness (SAGAware)*, September 2012.