

動作センシング点群データを対象とした深層学習ベース識別モデルに関する研究

1X20C058-1 鈴木一央
指導教員 後藤正幸

1. 研究背景・目的

近年、高齢者施設では介護者の負担低減が求められており、特に入居者の安全を守るためのカメラやセンサを用いた見守り技術が注目されている。例えば、見守り技術によって入居者の転倒などの異常を自動で検知し、介護者に知らせることができれば、介護者が常に傍で見守っていても素早い対応が可能になる。このような見守り技術のためのセンシングとして、ミリ波レーダーにより取得された時系列情報を含む点群データの活用が注目されている。一方、点群データの活用に着目すると、深層学習を用いた手法が近年多く研究されている。例えば Obinata ら [1] は人間の骨格の動きを表すデータに深層学習を適用し、高い精度で動作の識別が可能であることを示している。

ここで、深層学習の適用に際しては、識別のために必要十分な情報を含む特徴空間を構成すると共に、大量の教師データを準備する必要がある。しかし研究対象であるミリ波レーダーから取得される点群データは取得時間（フレーム）ごとに点の数が異なり、これらの点群をそのまま深層学習モデルの入力とすることができない。また、動作情報が付与されたデータの取得にコストがかかるという課題がある。そこで本研究では、得られた点群データに対し統計量を用いた特徴量抽出を行うことで入力データの次元数を一定にし、かつデータ拡張手法により十分な学習データを用意することで深層学習を適用可能とする手法を提案する。具体的にはまず、フレームごとに動作識別に有用な特徴量を抽出したデータを作成し、Convolutional Neural Network[2]（以下、CNN）を用いて時系列方向の畳み込みを構成することで、動作識別を行うモデルを構築する。さらに、点群の水平方向への平行移動、任意の動作として切り取る時系列フレームの変更、抽出した特徴量への mixup[3] の適用の 3 つの手法を用いて、データ拡張を行い、データ数を確保することで精度を向上する。本稿では、実際の点群データに対して提案手法を適用し、動作識別が可能であること及びデータ拡張による精度向上を確認する。

2. 関連研究

2.1. Convolutional Neural Network (CNN)

CNN[2] は主に画像認識の分野で用いられている深層学習モデルであり、畳み込み層、プーリング層、全結合層から構成される。畳み込み層では、入力データの一部との積和演算により局所的な特徴を抽出できるフィルタを一定の間隔でスライドさせることで、特徴マップを形成する。プーリング層では、特徴マップ中の小領域の中から最大値を残して集約する Max Pooling(Pool) の処理を施すことによって次元を削減するとともに、特徴が出現する位置に対する頑健性を付与する。全結合層では、畳み込み層とプーリング層を経て

得られた特徴ベクトルを統合し、タスクに応じた最終的な結果を出力する。CNN の一般的な構成を図 1 に示す。

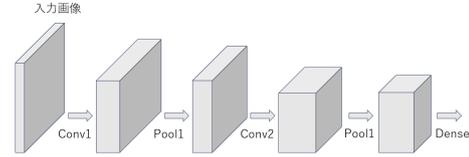


図 1: CNN の一般的な構成

2.2. mixup

mixup[3] は、画像分野において提案されたデータ拡張手法である。データ拡張とは、学習データに何らかの変換を施すことでデータを増やすことを指し、mixup では、2 つの学習データを混合することによって新たな学習データを作成する。具体的には、式 (1),(2) に示すように、ランダム抽出されたデータとラベルのペア (x_i, y_i) , (x_j, y_j) に対して混合処理を行う。

$$\tilde{x} = \lambda x_i + (1 - \lambda)x_j \quad (1)$$

$$\tilde{y} = \lambda y_i + (1 - \lambda)y_j \quad (2)$$

なお λ は混合比率であり、ベータ分布に従い生成する。この混合処理を行うことで、新たな学習データ (\tilde{x}, \tilde{y}) を作成する。この手法を用いることにより、学習データ内には存在していなかった中間的なデータを生成することができる。

3. 提案モデル

3.1. 着想

本研究で用いる点群データはフレームごとにそのデータ数が異なるため、取得された点群データをそのまま、入力次元数を一定にする必要がある通常の深層学習時系列モデルに入力することはできない。そこで、得られたデータから任意の次元数の各フレームに共通する特徴量を抽出し、モデルの入力とすることを考える。また、データ数不足を補うための手法としてデータ拡張を導入する。

3.2. 提案手法

まず、点群データからフレームごとの特徴量を抽出し、縦方向に特徴量、横方向に時系列を並べてデータを作成する。このデータを 1 次元方向にのみ畳み込みを行う CNN である 1 Dimensional-CNN（以下、1D-CNN）に入力し動作識別を行う。提案モデルでは、時系列方向にのみ畳み込むことによって時系列情報を考慮することができる。提案モデルの構造を図 2 に示す。

データ拡張に関してはまず、動作が行われる場所に対しての頑健性を高めるため、点群の座標を水平方向に平行移動することを考える。同様に、動作として切り取るフレームの位置を前後に変化させるデータ拡張を行うことで、データの中で実際に動作をしている時間に対して頑健になることが期待される。また、本来識別対象となるような各動作には、今回

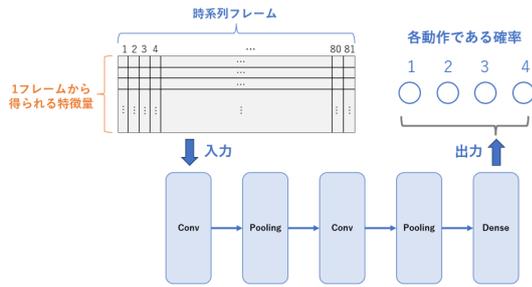


図 2: 提案モデルの構造

学習データとして扱った動作の中間のような動作も含まれていると考えられる。そこで、データの中間となる新規データを作成することができる mixup を特徴量に適用したデータ拡張を行うことを考える。具体的には同じ動作に関するデータの中間となるデータを作成するため、同じラベルを持つ学習データの中からランダムに選ばれた 2 組のデータを、ベータ分布から得られる λ を混合比率として混合処理を施し、新たな学習データとする。

4. 実データ分析

本提案モデルの有効性を検証するため、実際にミリ波レーダーを用いて観測した点群データに対して本提案手法を適用し、動作識別の実験を行う。本研究で用いるミリ波レーダーは、物体の動きを検出することが可能であり、物体の動きを点群データとして取得することが可能である。1 フレーム (0.05 秒) ごとに動きが生じた物体を示す点群の座標を観測し、点群の動きを観測することで物体の動きを捉えることができる。

4.1. 実験条件

本実験では、「転倒」、「座る」、「転倒から立ち上がる」、「座りから立ち上がる」の 4 動作について識別を行う。1 回の動作時間は動作の前後を含めて 81 フレーム (4 秒間) とした。各動作につき 50 回測定を行い、計 200 個のデータと対応する動作ラベルを持つデータセットを作成している。各動作 50 データのうち、40 データを学習用、10 データをテスト用データとして用いた。各フレームにおいて抽出する特徴量は、点の数、重心の xyz 座標、重心からの平均距離の 5 つである。1D-CNN ではエポック数を 50、学習率を 1.0×10^{-4} とし、活性化関数は ReLU 関数を用いた。データ拡張手法はそれぞれ学習データ 40 データに対して適用し、データ数を 5,10,20 倍に拡張して実験を行った。

4.2. 実験結果

4.2.1. 動作識別

表 1 に、提案モデルを適用して動作識別を行った際の学習データとテストデータに対する精度を示す。

表 1: 提案モデルによる動作識別の精度 [%]

学習データ	テストデータ
94.3	45.0

表 1 より、提案モデルは学習データに対しては高い精度を示す一方で、テストデータに対する精度は低いことがわかる。

これは、学習データに過度に適合する過学習が生じていることを示唆している。

4.2.2. データ拡張

表 2 に、提案する 3 つのデータ拡張手法を用いて学習データの拡張を行い、拡張後のデータセットを用いて動作識別を行った際のテストデータに対する精度を示す。

表 2: データ拡張後のテストデータに対する精度 [%]

データ拡張の倍率	水平方向	フレーム前後	mixup
5	47.5	52.5	47.5
10	57.5	60.0	47.5
20	60.0	62.5	50.0

表 1, 2 より、データ拡張をしない場合と比較して、いずれのデータ拡張手法を用いた場合も、データ数を増やすほどテストデータに対する精度が向上していることがわかる。また、3 つの手法を比較すると、切り取るフレームを前後させるデータ拡張が最も精度向上に有効であると考えられる。

5. 考察

実験結果から、提案手法により、点群データの動作識別が可能であること、及びデータ拡張の有効性が示された。表 2 から、mixup を用いた場合の精度が他の 2 手法と比較して低いことがわかる。これは、81 フレームのデータの中で実際に動作を行った時系列上の位置が異なるデータ同士を混ぜ合わせたため、元のデータから大きく乖離したデータが生成されたためと考えられる。したがって、時系列上の位置が近いデータ同士を選んで混ぜ合わせることにより、精度が改善する可能性がある。

6. まとめと今後の課題

本提案モデルでは、ミリ波レーダーにより得られた点群データから時系列間に共通な特徴量を抽出し、それらのデータに対して 1D-CNN を用いて動作識別を行った。これにより、時系列情報を保持した動作識別が可能になったと考えられる。さらに、学習データにデータ拡張を適用することにより、過学習を抑制することを試みた。提案手法をミリ波レーダーを用いて観測した点群データに適用して、時系列を考慮した動作識別ができることを示し、また、データ拡張に一定の効果があることを確認した。今後の課題としては、物体同士が重なることなどにより生じる一部劣化したデータへの適用や、センサ設置場所の違いへの対応などが挙げられる。

参考文献

- [1] Yuya Obinata and Yamamoto Takuma. Temporal extension module for skeleton-based action recognition. *International Conference on Pattern Recognition*, Vol. 25, pp. 534–540, 2021.
- [2] Yann LeCun, Bernhard Boser, John S Denker, Donnie Henderson, Richard E Howard, Wayne Hubbard, and Lawrence D Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, Vol. 1, No. 4, pp. 541–551, 1989.
- [3] Hongyi Zhang, Moustapha Cisse, Yann N Dauphin, and David Lopez-Paz. mixup: Beyond empirical risk minimization. *International Conference on Learning Representations*, pp. 1–13, 2018.