

1.はじめに

人物の動作識別は、画像ベースの識別法[1]と信号ベースの識別法[2]に大別できる。画像ベースの手法は、RGBカメラから取得した画像を用いるため、装着型計測デバイスを必要とせず、非接触で動作の識別が可能である。しかし、物を掴むなどの物体に対して手先のみで行う動作は、オクルージョンが発生すると識別が困難となる。一方、信号ベースの手法は、人体に装着したデバイスから筋電信号を用いるため、オクルージョンが発生しないというメリットがある。手先の動作識別を行うためには、手先の動作に深く関わる信号を計測する必要がある。そこで本研究では、人体の手首に位置し、手の筋肉を支配する手根管神経から計測可能な神経信号に着目する。神経信号の時系列データから動作固有の特徴を捉えるために、1D-Self-Attention Blockを提案する。

2.従来研究

信号ベースの従来手法として、Morbidoniらは表面筋電図(sEMG)を階層型ニューラルネットワークに入力して、人物のスタンスフェーズとスイングフェーズの動作識別を行う手法を提案している[3]。また、Ozalらは人間の心電図からフラグメントと呼ばれる小領域をサンプリングし、CNNを適応することで17クラスの心不全を検出する手法を提案している[2]。これらの従来手法は、クラス識別のための特徴抽出に多層ペーセプトロンや畳み込みニューラルネットワークを用いている。しかし、畳み込み層は、近傍の時刻に着目するため、離れた時刻の重要な変化を捉えることに適していない。

3.提案手法

人物の手先の動作は、その種類により動作時間が異なる。そのため、動作の短期的および長期的な変化を捉える事が重要となる。本研究では、時系列データにおいて、周辺時刻を考慮した特徴抽出が可能な1D-Self-Attention Block(1D-SAB)により構成される1D-Self-Attention Networks(1D-SAN)を用いた動作識別法を提案する。

3.1.手根管神経

本研究では、人物の手先の動作を対象とするため、人体の手首に位置し、手の筋肉を支配する手根管神経から計測可能な神経信号を用いる。手根管神経には、橈骨神経(Radial nerve), 正中神経(Median nerve), 尺骨神経(Ulnar nerve)がある。これらの神経信号をMudra wearable deviceにより計測する。サンプリング間隔は1msとする。

3.2.1D Self-Attention Networks(1D-SAN)

本研究では、Self-Attention Networks(SAN)[4]で用いられるSelf-Attention Block(SAB)を1次元データに対応させた、1D-SABを提案し、1D-SANを構成する。提案する1D-SABの構造を図1に示す。1D-SABには、手根管神経から取得した時系列信号を入力する。入力した信号は、1時刻ごとに処理し、対応する時刻のSelf-Attentionを算出する。図1における青色の値を処理の注目時刻とした時、赤色を近傍時刻1、緑色を近傍時刻2とする、各近傍時刻に対して、注目時刻(青)とのPointwise conv処理を行う。また、注目時刻と近傍時刻を、 ϕ および ψ の学習可能な関数(1D-CNN)へ入力し、関係関数 δ を求める。関係関数 δ の出力は、関数 ϕ の出力から関数 ψ の出力を減算したものであり、マッピング関数 γ によって1つ目の処理の出力とチャンネル数を合わせる。その後、2つの出力の要素積を算出する。この処理を近傍時刻分行い総和する。生成したSelf-Attention Map(SAM)はPointwise conv処理により入力チャネル数と同じチャネルにする。この出力に、ステップ機構として入力信号を加算し、最終出力とする。1D-SABを用いることで、神経信号内の重要な位置に大きな重みを与えることができ、離れた信号値との関係性を考慮できる。

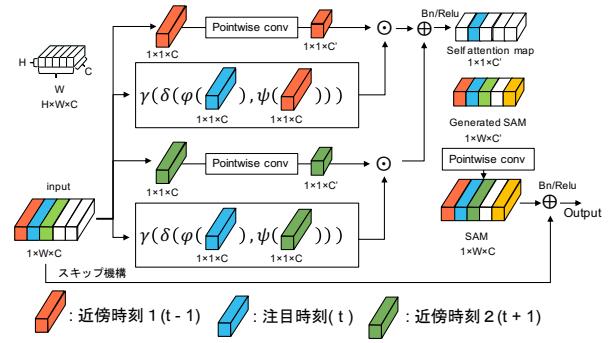


図1：1D Self-Attention Block の構造

3.3.ネットワークの構造

本研究では、3層の1D-SABからなる1D-SANを用いる。一層目の1D-SABにおける近傍時刻の間隔を3時刻、2、3層目の1D-SABにおける近傍時刻の間隔を5時刻とする。これにより、離れた時刻の信号との変化を考慮できる。1D-SANの構造と各1D-SABの近傍設定を図2に示す。

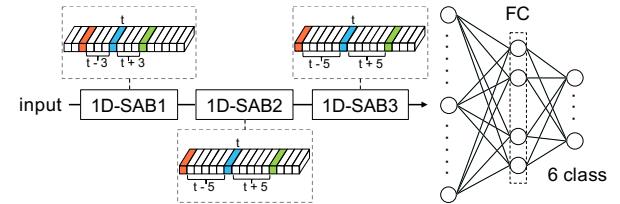


図2：ネットワーク構造と近傍

3.4.動作識別の流れ

本研究での動作識別は、3つのStepとなる。Step 1は、Mudra wearable deviceを用いて取得した神経信号を一定区間ごとに固定長でサンプリングし、1D-SANで推論ができるデータへ整形する。Step 2は、学習済み1D-SANに、識別対象のデータの0時刻を起点として、入力するデータの領域を10時刻ずつ繰り返し入力し、動作識別する。Step 3は、各区間の動作識別結果を統合して最終的な動作識別結果を出力する。

Step 1：前処理

本研究で扱う神経信号波形は、可変長のデータである。1D-SANで推論を行うために、64時刻分の1次元波形をサンプリングする。

Step 2：サンプリング時刻での動作識別

動作識別は、10時刻ずつサンプリング区間をずらしながら1D-SANに入力する。

Step 3：動作識別結果の統合

各時刻におけるネットワークの出力をクラス毎に累積する。累積値が最大のクラスを動作識別結果とする。識別結果累積処理を図3に示す。

4.評価実験

提案手法の有効性を示すために、1次元波形に対応した畳み込みニューラルネットワーク(1D-CNN)と認識精度の比較を行う。また、フーリエ変換し、周波数空間を入力とした際の精度とも比較する。

4.1.データセット

本研究では、6クラスの動作に対する神経信号を100回計測した600サンプルを用いる、これらを6:4で学習および評価データに分割して用いる。クラスのデータ取得時の

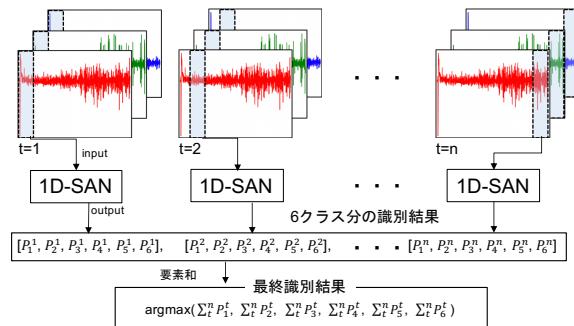


図 3 : 最終処理判定

画像を図 4 に示す。データは、Mudra wearable device を被験者の右手首に装着し、動作中の 1 次元波形を収集した。図 5 に動作 Push に対する神経信号を示す。

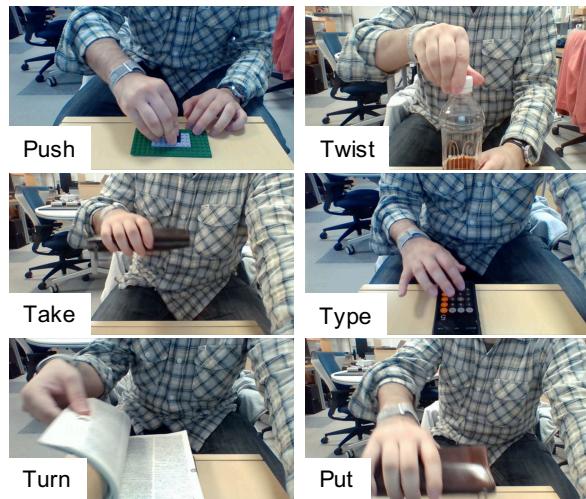


図 4 : 対象動作クラス

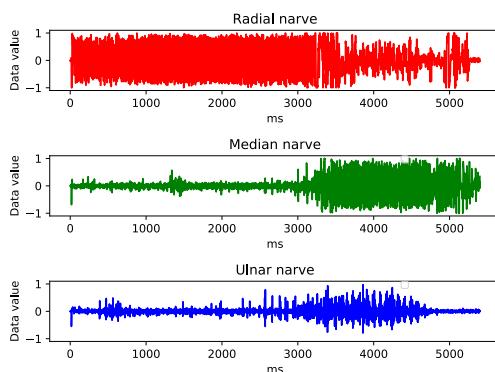


図 5 : 信号データサンプル (Push)

4.2. 実験概要

1D-CNN を用いたモデルと 1D-SAB を用いたモデルの精度比較を行う。1D-CNN は、2 層の畠み込み層と出力層の計 3 層の構造である。学習は、Epoch 数を 20000, BatchSize を 128, サンプリング幅を 64 とする。Optimizer は Momentum SGD とし、学習率を 0.001, Loss 関数に CrossEntropy Loss を用いる。

4.3. 実験結果

表 1 に各手法の Confusion Matrix を示す。比較実験の結果、1D-CNN の動作識別精度は 90.4 %、提案手法は 91.2 % となり、提案手法が動作識別で有効であることがわかった。提案手法は、Turn, Twist, Type の精度が向上し、Take, Push の精度が低下した。1D-CNN では、Twist を Turn と誤識別したが、提案手法では、Turn への誤識別

は減少した。一方で提案手法は、Push を Twist と誤識別することが多い。Push に対して提案手法の精度が低下した要因は、物を押し込む動作をする際に、力を込めて強く押し込むケースがあり、Push でながら Twist のように振動する信号が計測されたことが原因と考えられる。このようなケースを学習データにも追加する必要がある。また、Turn や Twist など、時系列的な変化が大きなクラスに対する精度が向上していることから、提案手法は信号値が時系列的に大きく変化するクラスに有効である。

表 1 : Confusion Matrix による識別結果の比較

1D-SAN_pair355_accuracy : 91.2%							1D-CNN_accuracy : 90.4%						
true label	Take	Put	Type	Turn	Twist	Push	Take	Put	Type	Turn	Twist	Push	
	90.0	0.0	0.0	5.0	2.5	2.5	97.5	0.0	0.0	2.5	0.0	0.0	
true label	Put	Take	Twist	Type	Turn	Push	Put	Take	Push	Twist	Type	Turn	
	7.5	85.0	0.0	5.0	0.0	2.5	12.5	85.0	0.0	0.0	2.5	0.0	
true label	Type	Push	Take	Put	Twist	Turn	Type	Push	Take	Put	Twist	Turn	
	0.0	0.0	92.5	7.5	0.0	0.0	0.0	0.0	90.0	10.0	0.0	0.0	
true label	Turn	Push	Push	Push	Push	Push	Turn	Turn	Push	Push	Push	Push	
	0.0	0.0	0.0	0.0	0.0	0.0	10.0	0.0	0.0	87.5	0.0	2.5	
true label	Twist	Push	Push	Push	Push	Push	Twist	Twist	Push	Push	Push	Push	
	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	7.5	92.5	0.0	
true label	Push	Push	Push	Push	Push	Push	Take	Take	Take	Take	Take	Take	
	0.0	0.0	0.0	0.0	0.0	0.0	5.0	0.0	5.0	0.0	5.0	90.0	

表 2 に、周波数領域における各手法の精度を示す。提案手法と 1D-CNN で比較を行った結果、周波数領域では、1D-SAN の動作識別精度が 80.0 %、1D-CNN の動作識別精度が 90.4 % となり、1D-SAN の精度が下回った。一方でフーリエ変換なしでは、1D-SAN の識別精度は 91.2 % と向上した。1D-SAN は、相対的な位置関係とその重要性を Self-Attention により表現できるためフーリエ変換なしで動作識別が可能である。

表 2 : 周波数領域における精度比較

ネットワーク	フーリエ変換あり (%)	フーリエ変換なし (%)
1D-SAN	80.0	91.2
1D-CNN	90.4	90.4

5. おわりに

本研究では、手先の筋肉を支配する手根管神経から取得した神経信号からの動作識別を行うために、1 次元波形に対応した 1D Self-Attention Network を提案した。精度比較実験では、従来手法の 1D-CNN に対して、提案手法が約 0.8 ポイント上回り、動作識別において、提案手法が有効であることを示した。また、Push のような神経信号値が連続的に大きくなるクラスでは、精度が低下したが、Turn や Twist など、時系列的な変化が捉えやすいクラスに関して精度が向上した。

参考文献

- [1] T. Xiao, et al., “Reasoning About Human-Object Interactions Through Dual Attention Networks”, ICCV, 2019.
- [2] O. Yildirim, et. al . , “Arrhythmia detection using deep convolutional neural network with long duration ECG signals”, Computers in Biology and Medicine, vol102, 2018, p411 420.
- [3] C.Moribidoni,et al., “A Deep Learning Approach to EMG-Based Classification of Gait Phases during Level Ground Walking” ,MDPI, 2019.
- [4] Hengshuang.Zhao,et al., “Exploring Self attention for Image Recognition” ,CVPR, 2020.

研究業績

- [1] 田邊稜等，“Self-Attention Networks による神経信号からの動作識別”，ビジョン技術の実利用ワークショッピング, 2020.