

Self-Attention Networks による神経信号からの動作識別

○田邊 稜[†], 平川 翼[‡], 山下 隆義[‡], 藤吉 弘亘[‡]

[†]: 中部大学工学研究科ロボット理工学専攻

[‡]: 中部大学

naber@mprg.cs.chubu.ac.jp

概要: 人物の動作識別は、工場での作業監督及び、セル生産に対応したファクトリーオートメーションの実現にとって重要である。画像を入力とした動作識別手法の場合、手先付近が物によって隠れる場合があり、動作識別が困難となる。また、画像からは、物体を扱う作業において重要となる、力加減の情報を計測できない。そこで本研究では、手先の隠れを考慮する必要がない信号ベースの手法に着目する。入力として用いる信号は、人体の手首に位置し、手先の動きを支配する手根管神経（橈骨神経、正中神経、尺骨神経）である。これは、人物の意図した動きを取得できるため、手先の動作を識別するデータとして適している。動作識別には、Self-Attention Network を用いる。本研究では、入力として神経信号を用いるため、Self-Attention Network を 1 次元に適応した 1D-Self-Attention Network を提案する。また、比較手法として、1 次元の畳み込み手法である 1D-CNN と比較実験を行い、1D-Self-Attention Network の精度が 1D-CNN の精度を約 1.2% 上回ることができた。

神経信号, Self-Attention Network, 動作識別

1. はじめに

Convolutional Neural Network(CNN)を用いた人物の動作識別には、画像ベースの識別方法と信号ベースの識別方法があり、それぞれにメリットとデメリットがある。画像ベースの方法は、RGB カメラから人物の状態を取得できるため、動作識別の対象となる人物がデバイスをつける必要がなく、非接触で識別できるというメリットがある。しかしながら、物を掴むなど、物体に対して、手先を使って行う動作を識別する際には、手先付近に物によってオクルージョンが発生する場合があります、識別が困難である。また、画像から得ることの出来ない力加減などの情報を学習に反映させることが出来ないというデメリットがある。信号ベースの方法は、画像ベースのようにオクルージョンが発生せず、筋電信号や神経信号などの人体から直接収集した情報を用いることができるため、力加減などの情報を考慮でき、動作識別精度を向上させることができるメリットがある。しかしながら、人物に専用のデバイスを装着させる必要がある。本研究では、動作識別精度を向上させるという観点をもとに、信号ベースの方法に着目する。人物に装着したデバイスから取得することのできる信

号には、筋電信号と神経信号がある。本研究では、人物が手先を使って行った動作を識別することを目的としているため、人体の手首に位置し、手の筋肉を支配する、手根管神経（橈骨神経、正中神経、尺骨神経）から収集可能な神経信号を用いる。また、神経信号は人物の意図しない動作を反映しない性質もある。

2. 従来研究

画像ベースの動作識別として、Spatial Temporal Graph Convolutional Networks[1]が提案されている。この手法は、RGB 動画の 1 フレーム毎に人体の骨格モデルを作成し、作成した骨格モデルの同一関節を時系列方向に接続する。同一フレームの空間的なグラフと同一関節の空間的なグラフを ST-GCN に入力することで動作識別を行う。しかし、人物の骨格モデルのみを考慮しているため、物体把持のような手元の細かな動作を識別することはできない。Dual Attention Networks[2]では、人間と物体の相互作用について重要な機能に重み付けをする、Dual Attention Framework を提案し、物体と人間の関係を考慮した動作識別を行った。信号ベースの従来手法としては、筋電信号を用いた研究では、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

別を行うと共に、足と床の接触信号を予測した。心電図を用いた研究では、Ozal Yildirim ら[3]が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した。

3. 信号データ

本研究では、手根管神経(橈骨神経, 正中神経, 尺骨神経)から収集した神経信号データを用いる。これらの神経信号を Mudra wearable devise[4]により取得する。サンプリング間隔は1msである。

4. 提案手法

従来手法では、様々な信号を入力として人物の動作識別に取り組んでいる。これらの信号からの動作識別では、短期および長期の時間的な変化を捉える事が重要である。動作識別に広く活用されている畳み込みニューラルネットワークでは、入力に近い畳み込み層で短期の特徴、深い層で長期の時間的な特徴を捉えることができる。一方で、畳み込み層のフィルタは全ての時刻で同じであるため、ある時刻の周辺信号に対する重要性を考慮できていない。そこで、本研究では、Self-Attention Block(SAB)により構成される Self-Attention Networks(SAN)[5]を用いて動作識別を行う。

SANは、CNNと異なり画素ごとに重みを決定する仕組みが導入されている。しかし、SANを構成する Self-Attention Block(SAB)は画像から特徴を抽出する仕組みとなっているため、本研究で扱う神経信号(1次元データ)に対応していない。そのため、本研究では、SABを1次元データに対応させた、1D-SABにより、1D-SANを構成する。1D-SAB

を用いることで、神経信号波形内で信号値の相対位置を重み決定に反映させることができ、周囲の信号値との関係性を考慮できる。

4.1. 1D-Self-Attention Networks(1D-SAN)

1D-SANは、1D-SABを多層化して構成する。1D-SABの構造を図1に示す。1D-SABには、手根管神経(橈骨神経, 正中神経, 尺骨神経)から取得した信号をチャンネル方向に連結した3チャンネルの1次元信号を入力する。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、赤色の近傍時刻1と緑色の近傍時刻2、それぞれと注目時刻(図中青色)で処理を行う。近傍時刻ごとに行う処理は、2つに分岐する。1つ目は、注目時刻(青)と近傍時刻を Pointwise conv して出力する。2つ目は、注目時刻と近傍時刻を、 ϕ, ψ の学習可能な関数(1D-CNN)へ入力し、関係関数 δ を構築する。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の数だけ行い、総和する。生成した Self-Attention Map(SAM)は Pointwise conv により入力チャンネル数と同じチャンネルに戻す。この出力に、ステップ機構による、3チャンネルの入力を加算し、最終出力となる。

4.2. 時系列処理

本手法では、Mudra wearable devise を用いて取得

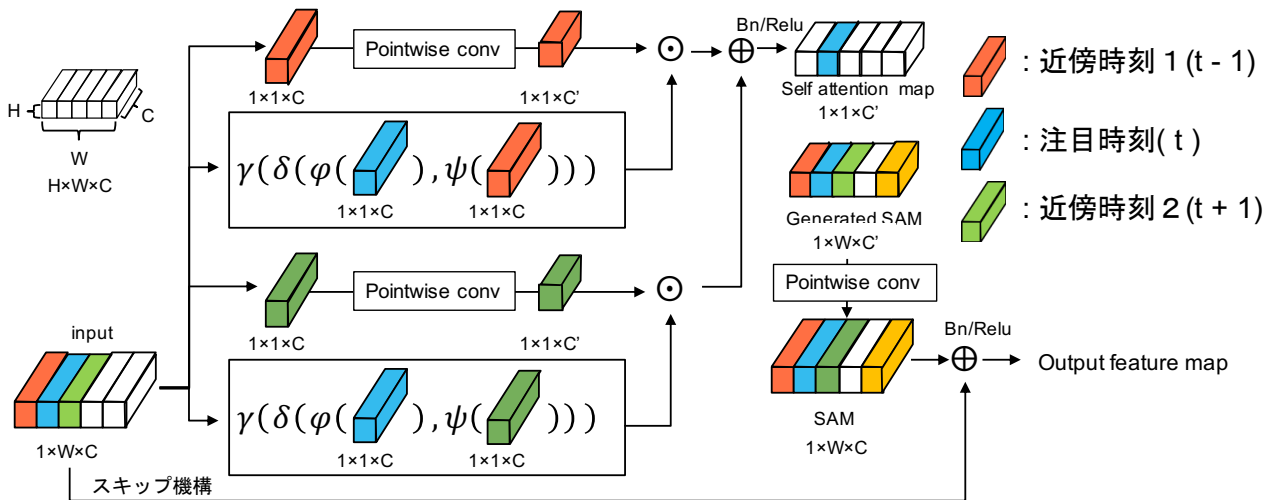


図 1 1D-Self-Attention Block の構造

した神経信号を一定区間ごとにサンプリングして64時刻分を1次元波形データとして1D-SANに入力し、動作識別する。そして、1D-SANに入力する区間を10時刻ずらして繰り返し動作識別を行う。その後、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。

4.2.1. 学習データの準備

本研究で扱う神経信号波形は、可変長のデータであるためSANで学習を行うために、一定の幅でサンプリングを行う。また、神経信号は、橈骨神経、正中神経、尺骨神経の3種類が同時に取得されるため、図2に示すように3チャンネルの64時刻分の1次元波形がサンプリングできる。

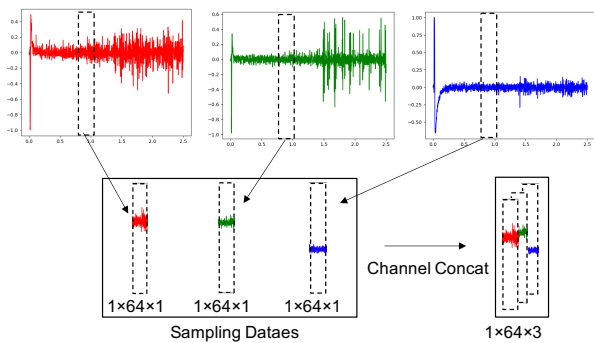


図2 データからのサンプリング

学習時は、サンプリング位置をランダムに決めて入力する。これにより、動作に含まれる多様な信号を学習することができ、汎化性能を向上させることができる。

4.2.2. 1D-SANの構造

本研究では、3層の1D-SABからなる1D-SANを用いる。各1D-Self-Attention Blockにおけるアテンションペアの間隔は5時刻としている。これによ

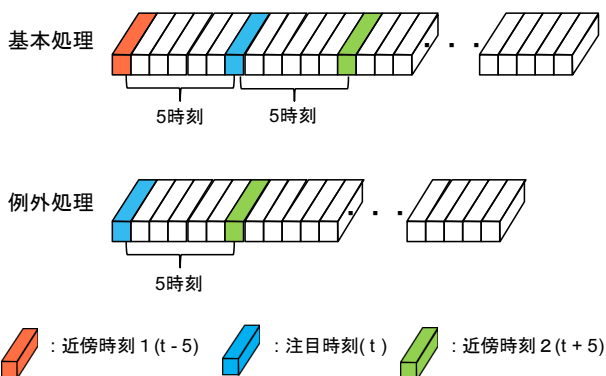


図3 1D-SABにおける近傍の設定

り、周辺時刻の信号との変化を考慮できる。アテンションペアの決定処理を図3に示す。基本処理では、注目時刻(t)を基準として($t-5$), ($t+5$)の近傍時刻をアテンションペアとして決定する。また、($t-5$), ($t+5$)に近傍時刻が存在しない場合は、例外処理として、近傍時刻が1時刻分となる。

4.2.3. 動作識別結果の統合

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。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。識別結果累積の処理を図4に示す。

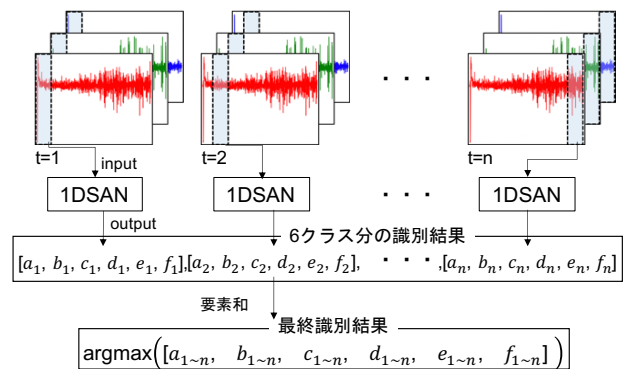


図4 識別結果累積処理

4.2.4. 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取得し、学習および評価データに分割して用いる。収集データの詳細とクラスの内訳を表1に示す。データは、Mudra wearable deviceを被験者の右手首に装着し、動作を行ってもらい収集した。8人の被験者から収集した収集データ長の最大値と最小値を表2に示す。

表1 収集データ数とクラスの内訳

クラス	動作内容	学習データ	評価データ
Take	物を掴む	60	40
Put	物を置く	60	40
Type	指で電卓を叩く	60	40
Twist	キャップを捻る	60	40
Push	物を押し込む	60	40
Turn	ページを捲る	60	40

表 2 収集データの長さ

クラス	最大データ長(ms)	最小データ長(ms)
Take	5634	1386
Put	10908	1566
Type	7596	1998
Twist	6174	1530
Push	5994	1188
Turn	7200	1656

5. 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。

5.1. 実験概要

1D-CNNを用いたモデルと1D-SABを用いたモデルで精度比較を行う。1D-CNNは、2層の畳み込み層と出力層の計3層の構造である。1D-SANは、4層の畳み込み層と3つの1D-SABから構成されている。推論時には、どちらのネットワークも出力層にSoftmax関数を用いる。

5.2. 学習の設定

実験時のパラメータを表3に示す。

表 3 実験時パラメータ

Epoch	20000
Batch Size	128
サンプリング幅	64
Optimizer	Momentum SGD
学習率	0.001
Scheduler	[3000, 5000, 7500]epoch
gamma	0.5
Loss 関数	CrossEntropy Loss

5.3. 実験結果

表4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90.8%、提案手法は92.0%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、Twistに対して大きく精度が向上した他、Put、Turnで精度が向上したが、Takeの精度が僅かに低下、Pushの精度が大きく低下した。1D-CNNでは、TwistをTake、Pushと誤識別しているが提案手法では、Take、Pushへの誤識別は削減された。一方で、提案手法は、Pushにおいて1D-CNNでは、誤識別しなかった、Take、Twistと誤識別した。また、Pushにおいて提

案手法の精度が大きく低下した要因としては、物押し込むという動作では、神経信号の値が連続的に大きな値をとることから、前後の神経信号値との関係性を考慮した、Self-Attentionが常に大きな値となってしまったと考察する。また、TurnやTwistなど、時系列的な変化が顕著なクラスに関して、精度が向上していることから、提案手法が時系列的に大きく変化するクラスに対して、有効だとわかった。

表 4 Confusion Matrix による識別結果の比較

1DSAN_result ac : 92.083							1DCNN_result ac : 90.833						
true label	Take	Put	Type	Turn	Twist	Push	true label	Take	Put	Type	Turn	Twist	Push
Take	38	1	0	0	1	0	Take	39	0	0	1	0	0
Put	2	36	0	1	0	1	Put	4	35	0	0	0	1
Type	0	0	36	3	0	1	Type	0	1	36	3	0	0
Turn	0	0	0	38	1	1	Turn	3	0	0	36	0	1
Twist	0	0	0	1	39	0	Twist	2	0	0	2	35	1
Push	3	0	0	1	2	34	Push	0	0	3	0	0	37
	Take	Put	Type	Turn	Twist	Push		Take	Put	Type	Turn	Twist	Push

6. 考察

実験の結果より、提案手法がTurn、Twistのような時系列的に大きく変化するクラスに対して、有効だとわかった。実際の信号とクラス尤度を比較し、1D-SANの優位点を考察する。Turnの信号、尤度比較を図5、Twistの信号尤度比較を図6に示す。比較図は4段になっており、上から、生の橈骨神経信号(Radial Nerve)、正中神経信号(Median Nerve)、尺骨神経信号(Ulnar Nerve)を重ね合わせたもの、図4の統合処理を行う際にネットワークに入力されたデータを時間方向につなげたもの、統合処理の際に1D-SANが出力するクラス尤度と正解ラベル領域を重ねたもの、1D-CNNで同様のものとなっている。

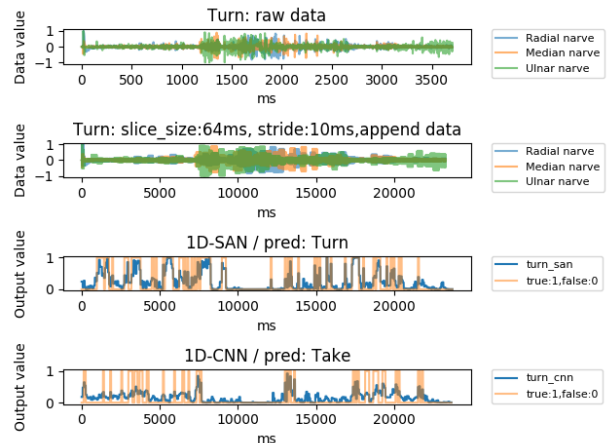


図 5 Turn の信号尤度比較

図 5 から 1D-SAN と 1D-CNN のクラス尤度を比較すると提案手法の方が全体的にクラス尤度が高くなる傾向となっている。また、本のページを捲るという動作に対して、尤度が高い領域が動作の最初と最後に来ていることから、ページを掴むという動作とページを離すという動作に着目できていると考える。

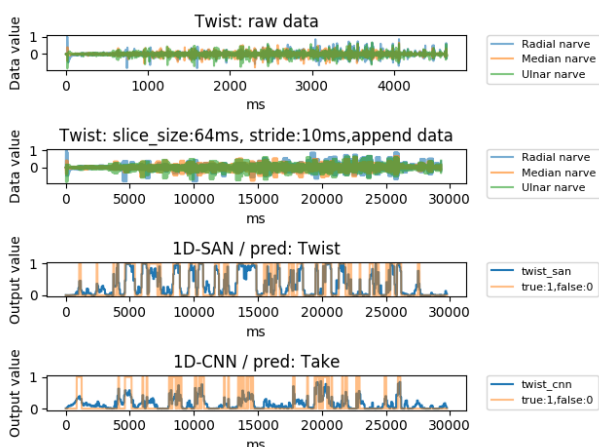


図 6 Twist の信号尤度比較

図 6 から 1D-SAN と 1D-CNN のクラス尤度を比較すると提案手法の方が全体的にクラス尤度が高くなる傾向となっている。また、電卓を叩くという動作に対して、信号の中間付近に細かく尤度が高い領域があることから、電卓を叩く際に細かく動く指に着目できていると考える。

7. おわりに

本研究では、手先の筋肉を支配する手根管神経(橈骨神経, 正中神経, 尺骨神経)から取得した神経信号を教師データとして、人物の手先を使った動作の識別を行い、一次元波形の特徴抽出方法として、Self-Attention Network(SAN)を一次元波形に適応した、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 1.2 ポイント上回り、動作席別において、提案手法が有効であるとわかった。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、Turn や Twist など、時系列的な変化が顕著なクラスに関して、精度が向上した。

今後は、神経信号以外の音などの一次元波形にも適応し、精度実験を行うとともに、1D-Self-Attention の推論時における判断根拠の可視化などを行う。

8. 参考文献

- [1] S. Yan, et. al. , “Spatial Temporal Graph Convolutional Networks for Skeleton-Based Action Recognition”, AAAI, 2018.
- [2] T. Xiao, et. al. , “Reasoning About Human-Object Interactions Through Dual Attention Networks”, ICCV, 2019.
- [3] 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.
- [4] <https://www.getmudra.com>
- [5] Hengshuang.Zhao, et. al. , “Exploring Self attention for Image Recognition”, CVPR, 2020.

田邊 稜：現在中部大学大学院修士課程在学中，神経信号を用いた動作識別に関する研究に従事。

平川 翼：2013 年広島大学大学院博士課程前期終了，2014 年 広島大学大学院博士課程後期入学，2017 年中部大学研究員（～2019 年），2017 年広島大学大学院博士後期課程修了．2019 年中部大学特任助教. 2014 年独立行政法人日本学術 振興会特別研究員 DC1. 2014 年 ESIEEParis 客員研究員（～ 2015 年）. コンピュータビジョン，パターン認識，医用画像 処理の研究に従事。

山下隆義：2002 年奈良先端科学技術大学院大学博士前期課程修了. 2002 年オムロン株式会社入社，2009 年中部 大学大学院博士後期課程修了(社会人ドクター)，2014 年中 部大学講師，2017 年より同大学准教授. 人の理解に向けた 動画画像処理，パターン認識・機械学習の研究に従事。

藤吉弘亘：1997 年中部大学大学院博士後期課程修了. 1997 ～2000 年米カーネギーメロン大学ロボット工学研究所 Postdoctoral Fellow. 2000 年中部大学講師. 2004 年より同大 学教授. 2005～2006 年米カーネギーメロン大学ロボット 工学研究所客員研究員，計算機視覚，動画画像処理，パター ン認識・理解の研究に従事。