

人体姿勢推定技術を用いたカーリング動作解析の一考察

Consideration of Curling Motion Analysis Using Human Pose Estimation Technology

黒田 英慈¹ 林 将寛¹ 榎井 文人² 柳 等²

Eiji Kuroda¹, Masahiro Hayashi¹, Fumito Masui², and Hitoshi Yanagi²

¹株式会社アイエンター

i-enter Corporation, Ltd.

²北見工業大学冬季スポーツ科学研究推進センター

² Research Center for Winter Sports Science, Kitami Institute of Technology

Abstract: We aim to develop a system for judging the delivery form of curling using human pose estimation technology in the image processing field. As a preliminary study, we investigated the feasibility of 3D human pose estimation from camera images of delivery form.

1. はじめに

カーリングは、冬季スポーツの1つであり、ストーンと呼ばれるハンドルの付いた石を数メートル離れた同心円上の的(ハウス)に目掛けて投球を行うスポーツである。またカーリングは、氷上のチェスとも呼ばれるように、戦術が勝敗に関わるスポーツである。戦術を遂行するためには、ストーンを狙った場所に投石しなければならない。カーリングの投石はデリバリーと呼ばれており、陸上競技のスターティングブロックのような足掛け(ハック)を蹴ることで勢いをつけて氷上を滑り、ストーンに回転をつけて石を離す一連の動作。デリバリー動作においては、正確な投石が求められるため、投石精度を高めるためのトレーニングが重要である。

しかしながら、従来のデリバリートレーニングは目視による確認が一般的であり、経験の浅い競技者には判断が難しかったり、目視判断の結果の一般化が難しいといった問題があった。

一方、近年野球やサッカーといったスポーツでは、フォームをAIで分析フィードバックするようなコーチング方法も取り入れつつある。このような取り組みを行うことにより、人材不足の改善や競技力の向上につながる。

AIによるフォームの分析では、主に(1)センシング機器での取得や、(2)画像解析による方法が考えられる。(1)は精度が良いものの、導入にコストがかかることや競技者にセンシング機器を取り付けることでプレーに支障が出る可能性がある。

一方、(2)は、精度がセンシング機器に劣るものの、低コストであることやカメラ撮影のみで処理可

能であることから、通常のパフォーマンスでプレーを行うことができる。

画像解析による人体の骨格情報解析方法として、姿勢推定がある。この技術を用いることにより、カメラ映像や画像から被写体となる人物の骨格情報の取得が可能となり、効率良く動作解析が行える。

我々は、カーリングのデリバリー動作に着目し、被写体となるカーリング競技者を撮影して、その選手のデリバリー動作の良し悪しを判定するシステムの開発を目指している。本稿では、事前研究として、姿勢推定技術を用いてカーリング競技者の骨格位置の取得が可能であるか、動作解析が可能であるかどうかの実現性調査を行ったので報告する。

2. 関連研究

本章では、関連研究としてスポーツの動作解析に関して、センシング機器や姿勢推定を用いた例について述べる。

桂ら[1]は、ボルダリングを対象に姿勢推定ライブラリを用いて被写体がクライミングウォールを登り終えたどうかの完登判定の研究を行っている。クライミングウォール上のホールド(手をかける石)を配列に登録し、ゴールとなるホールドに推定された骨格が範囲内に入ることによって判定している。研究の課題として、手首をゴール判定のトリガーとしているため、ホールドに手が付いている状態でも手首の座標がゴールの範囲内に入らず失敗判定となることや、胴体などで対象となる骨格が遮断され推定が行えず失敗することであったと言及している。中村ら[2]は、サッカーのシュート動作を対象として、姿勢

推定の精度向上について研究をしている。サッカーのシュート動画に対して一度姿勢推定を行い、目視で推定結果から失敗した動画フレーム画像を抽出して、抽出した画像に正解データを付与して、再度学習を行うことで、精度向上を目指している。失敗フレームには、やはり正解を付与したい骨格が隠れていることや、服の色などで付与が難しいといったことを言及している。上記に2件について、姿勢推定を行う上で対象となるスポーツの特色を把握した上で学習データやそのスポーツに特化した学習モデルの生成が必要であることが窺える。

スポーツの動作解析においては、その競技の未経験者、経験者に分類して分析されることがしばしばある。清水ら[3]は可視光センシング機器を用いてバドミントンのスマッシュの解析を行っており、佐藤ら[4]も、可視光のセンシング機器を用いて、サッカーのリフティング動作についての研究を行っている。上記のような研究では、そのスポーツの動作で気を付けることが明確になるため、カーリングのデリバリ動作の動作解析でも応用したい。

一方カーリングのデリバリ動作について、姿勢推定を用いた動作解析の研究は事例が見当たらず、新規性があると言える。

3. 人体姿勢推定について

本章では、画像解析分野の人体姿勢推定について概略を述べる。なお具体的なアルゴリズムや考え方が知りたい場合には、こちらのサーベイ論文を参照されたい[5]。

人体姿勢推定とは、画像解析の一分野であり、画像上の被写体の肩や肘といった骨格点を AI で認識する技術である。姿勢推定には画像上の被写体の骨格を画像上の座標位置で推定する 2D 姿勢推定と、3次元空間上での座標位置を推定する 3D 姿勢推定がある。

2D 姿勢推定には、Top-Down アプローチと、Bottom-Up アプローチの2つの手法が存在する。前者は、姿勢推定を行う前処理として画像中の人物の位置を特定するために物体検出を行う必要がある。物体検出後、人物のみが写った画像を姿勢推定 AI に入力することで被写体の画像中の骨格座標を得ることができる。後者は、物体検出で画像中から被写体の位置を取得する前処理が不要であり、画像中に存在するそれぞれの肩や肘、膝といった骨格の座標を推定し、その後それぞれの骨格位置から人物の骨格を再構成することで全身の骨格情報が得られる。両者を比較すると、Top-Down アプローチは、精度が良いものの、画像上の人物位置の特定が必要であり、

Bottom-up アプローチは、ワンショットでの推定や、複数人推定に強い一方、Top-Down アプローチに比べ精度が落ちる。

3D 姿勢推定にも、大まかに2種類の推定手法が存在する。前者は、予め推定を行った 2D 骨格座標を用いて 3D 骨格を推定する手法で、前処理が必要となるが、精度自体は良いとされている。後者は、画像上を直接入力して 3D 骨格を推定する手法で、ワンショットで推定できるが、2D 骨格から 3D 骨格を推定する手法に比べ精度が低いとされる。

本研究においては、精度を重視するため、2D 姿勢推定では、Top-Down アプローチの human-pose-estimation.pytorch というライブラリ[6, 7]と、3D 姿勢推定では、2D 骨格から 3D 骨格の推定を行う GAST-Net というライブラリ[8, 9]を選択した。

4. 実験

本章では、姿勢推定を用いてカーリングのデリバリ動作について骨格情報の取得が可能であるかどうかの実現性調査のため、いくつかの学習データを用意し、学習を行い予め撮影したデリバリ動画に対して精度向上が見込めるか確認を行う。

4.1 多方向からの姿勢推定アプローチ

一般的な姿勢推定においては、単眼カメラで撮影した RGB 画像、映像に対して被写体の骨格推定を行う。しかし、被写体の映る角度によっては体の一部が死角となり精度低下を招く可能性がある。

この点の解決策として、3D の姿勢推定について2方向からの撮影を行い、2つの骨格情報を入力して1つの 3D 姿勢推定結果を出力する新たなネットワークを検討した。図1に処理の流れのイメージ図を示す。これが実現できれば、姿勢推定で2台のカメラが必要となるコストは発生するが、死角がなくなることにより精度が向上すると仮定している。

以後上述したネットワークを多方向姿勢推定と呼称する。

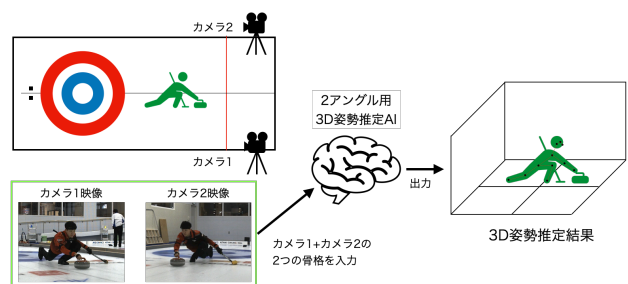


図1：多方向姿勢推定のイメージ

4.2 取り扱う骨格について

姿勢推定で取り扱う骨格は、以下の通りである。2D は一般的な姿勢推定データセットである coco-dataset[10]形式に倣い、鼻、左目、右目、左耳、右耳、左肩、右肩、左肘、右肘、左手首、右手首、左腰、右腰、左膝、右膝、左足首、右足首の計 17 骨格を扱う。3D もまた、一般的な 3D 姿勢推定データセットである human-3.6m[11]形式に倣い、腰の中心、右腰、右膝、左足首、右腰、左膝、左足首、背骨、胸骨、首、鼻、頭頂部、右肩、左肘、左手首、左肩、右肘、右手首の計 17 骨格を取り扱う。

4.3 評価指標

本節では 2D、3D 姿勢推定の精度についてそれぞれ記述する。

2D 姿勢推定の精度評価には、姿勢推定分野で一般的に用いられる PCK (Percentage of Correct Keypoints) を選択した。PCK は、推定された骨格と正解骨格との差を算出し、任意の閾値以内であれば、正しく推定が行えているとする手法である。本実験では、人体矩形の対角線を 0.05 倍した値をシビアな閾値 (@0.05) とし、0.1 倍した値をルーズな閾値 (@0.1) とした。

3D 姿勢推定については、LIU ら[8]の手法に従い、P-MPJPE (Procrustes analysis MPJPE) で評価を行う。元となる MPJPE (Mean Per Joint Position Error) は 2D の評価と同様に推定結果と正解データの差の平均値を算出するが、3D の推定では、正解データ必ずしも出力結果の角度や出力座標が一致するとは限らないため、回転角や並行移動などが考慮された P-MPJPE を用いる。算出される数値は、精度が高いほど 0 に近い値となる。

4.5 実験設定

本節では 2D と 3D の姿勢推定の学習を行う上で用意したデータについて述べる。

2D 姿勢推定用学習データ

2D の学習データは、インターネット上からダウンロードしたデリバリー画像と、インターネット上とそれから実際に撮影を行った動画をフレームに分割した画像、計 2707 枚用意した。上記画像に対して、人手で骨格情報を付与するアノテーション作業を行い、学習用 2437 枚、検証用 270 枚として学習させた。また coco データセットを学習した事前学習済みモデルをベースラインとして比較対象とした。

3D 姿勢推定用学習データ

3D 姿勢推定の学習データを表 1, 2 に示す。

データセット 1 は、3D 姿勢推定分野で一般的に用いられる Human3.6m データセットを学習させたベースラインとなる学習モデルである。データセット 2, 3 は Noitom 社のパーセプションニューロンプロ [12] というセンシング機器を用いて収集する。パーセプションニューロンは、無線式のモーションキャプチャーであり、15 個のセンサーを体に装着することで、3D のモーションデータを取得することが可能となる。データセット 2 は、パーセプションニューロンを用いて実際に氷上でデリバリー動作を行った際に得られた 3D のモーションデータを元に GAST-Net 用に整形したデータで学習を行ったモデルである。データセット 3 については、データセット 2 と同様にパーセプションニューロンを用いて取得したモーションデータであるが、こちらは氷上での動作ではなく陸上でデリバリーの動作を疑似的(素振り)に再現した動作である。

上記 3 つのデータセットをそれぞれ学習させて生成した学習モデル A~C と、多角方向 GAST-Net 用学習モデル D~F に対して、事前に撮影した動画で性能の評価を行った。

表 1 : 3D 姿勢推定の学習データ一覧

学習モデル名	データセット概要
学習モデル A	データセット 1
学習モデル B	データセット 2
学習モデル C	データセット 3

表 2 : 多方向姿勢推定の学習データ一覧

学習モデル名	データセット概要
学習モデル D	データセット 1
学習モデル E	データセット 2
学習モデル F	データセット 3

※

- データセット 1 : ベースライン
(一般的な 3D 姿勢推定データセット, Human 3.6m)
- データセット 2 : パーセプションニューロンによる、
デリバリーのデータセット(299 ショット分)
- データセット 3 : パーセプションニューロンによる、
疑似的なデリバリーのデータセット(177 ショット分)

5. 結果

本章では、上述した実験の結果について述べる。2D 姿勢推定の精度検証の結果を表 3 に、推定され



図 2 : 学習モデル別 2D 姿勢推定結果例

画像上部は、事前学習済みモデルの推定結果、画像下部は、カーリング画像で学習を行ったモデルの推定結果。足以外は精度に大きな差は無いが、足の推定結果は大きく異なり、事前学習済みモデルは推論が困難であるが、カーリングモデルではデリバリー動画を学習したことにより正確に推論が行えている。

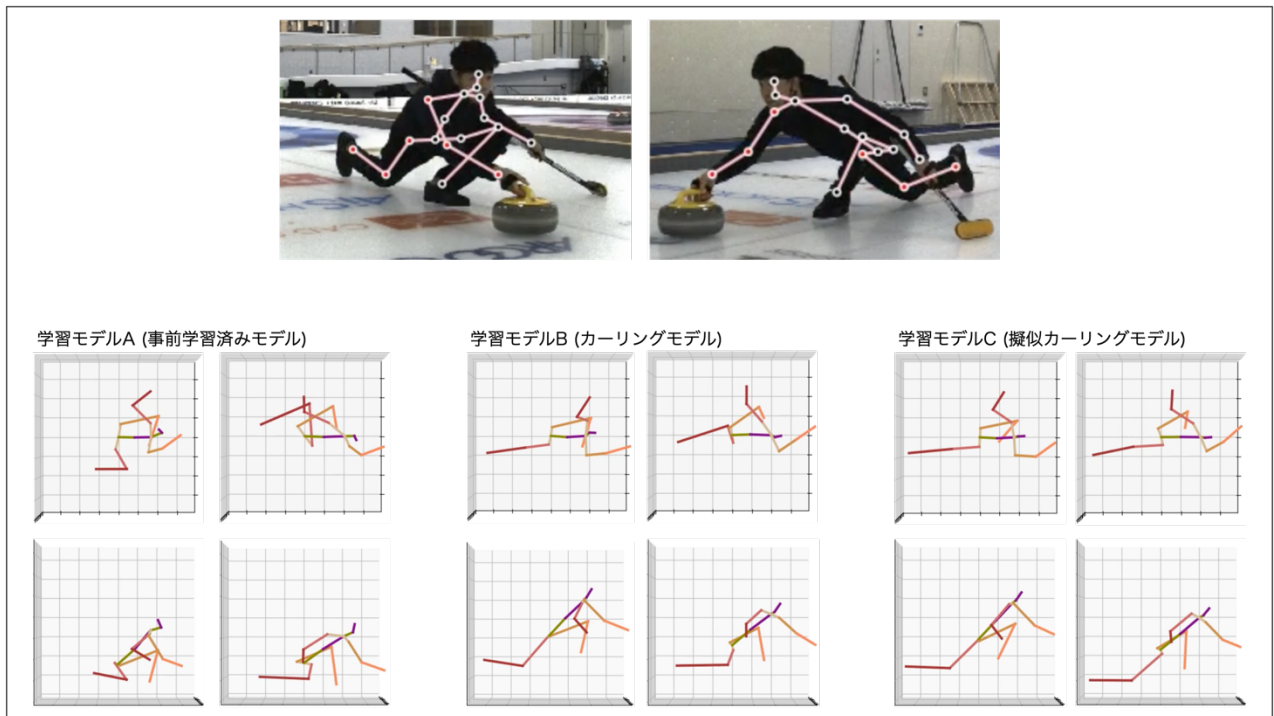


図 3 : 学習モデル別 3D 姿勢推定結果例

画像上部は精度検証用の動画の 1 フレームを抜き出し、2D 姿勢推定結果をそれぞれの学習モデルに入力し、出力された 3D 骨格座標を画像下部に示している。画像下部の 3 つのモデルに対して、上部が上からのアングル、下が真横からのアングルとなっている。また、ブロック左側が画像上部左の画像、ブロック右側が画像上部右の画像に対応する

た画像の例を図 2 に示す。

まず、デリバリー画像を学習したカーリングモデルと事前学習済みモデルの平均精度を比較すると、@0.05 は 15.7%、@0.1 では 8.1% と共に精度が向上していることが分かる。特に右足、左足に関しては、大きく精度が向上していることが確認できる。図 2 上部の推定例を確認すると、事前学習済みモデルでは両足の推定がうまく行えておらず、カーリングモデルでは正しい場所にプロットされている。カーリングのような一般的なデータセットに内包されない

特殊な動作については、新たに学習を行うことで姿勢推定が可能になることが示唆される。

続いて 3D 姿勢推定の精度検証の結果を表 4 に、推定された画像の例を図 3,4 に示す。3D の姿勢推定については、ベースラインである学習モデル A の P-MPJPE が 0.1549 に対して、学習モデル B は 0.0906、学習モデル C は 0.0894 となり、数値では擬似的なデリバリー動作を学習させた学習モデル C が最も高性能となった。図 3 を確認すると、学習モデル A については、デリバリフォームの特徴とも言える右足が全

表 3：2D 姿勢推定の学習結果の骨格別精度評価

学習モデル	顔 (%)	左腕 (%)	右腕 (%)	左足 (%)	右足 (%)	平均 (%)
事前学習済み モデル @0.05	89.5	72.2	74.7	45.6	38.7	64.1
事前学習済み モデル @0.1	94.9	88.8	89.4	69.7	67.9	82.2
カーリング モデル @0.05	89.4	79.6	79.8	76.0	74.4	79.8
カーリング モデル @0.1	92.2	90.1	90.9	89.6	88.8	90.3

表 4：3D 姿勢推定のモデル別 P-MPJPE 値

学習モデル	カメラ 1	カメラ 2
学習モデル A	0.1549	0.1794
学習モデル B	0.0906	0.0838
学習モデル C	0.0894	0.0791
学習モデル D		0.2551
学習モデル E		0.0746
学習モデル F		0.0753

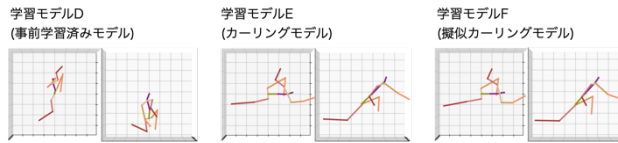


図 4：多方向姿勢推定の結果

く伸びていない。これは、2D の事前学習済みモデルと同様の問題として考えられる。学習モデル B, C については、右足が伸びており学習が成功していると判断できるが、B のカメラ 2 の右足においては想定し得ない方向に存在した。

多角の推定についても、同様に学習モデル D よりデリバリーモーションを学習させたモデル E, F の方が精度が良い結果となった。総合的に判断すると、学習モデル E, F についていずれも単眼のモデルより P-MPJPE 値が低くなり、多角での学習の有用性が証明できた。また、カーリングのデリバリー動作については、学習を行うことで姿勢推定を行うが確認できた。これに加え、仮にその他のスポーツの姿勢推定を行うことを考慮すると、カーリングと同様に専用の学習モデルを作成する必要があることが示唆された。

6. 考察

本章では、実験に対する考察と、本実験を踏まえ

た上でのデリバリーの動作解析についての検討を行う。

まず、学習モデル B と C について、氷上での動作と擬似的な動作の学習結果に差異が生じた点について考察を行う。本実験ではデータ収集にパーセプションニューロンを用いていたが、この機器は電波や、機器による電波干渉を受けやすい性質があり、またデリバリー動作は移動を伴う動作であるため、少なからずモーションデータに悪影響を及ぼしてしまう。対して、擬似的な動作については、移動を伴わずに動作を収集できるため、電波影響を受けづらいと言える。上記が影響したため、単眼の 3D 姿勢推定モデルでは実際のデリバリー動作を用いた学習モデル B より擬似的な動作を用いた C の精度が良くなったと考えられる。また、多方向姿勢推定においては、モデル E と F で同等の推定精度となったため、場所や時間に囚われずデータ収集が行えることが可能となり、素振りといった類似したデータでの学習の有用性も示唆された。さらに、より高精度なモデルを作成するためには、コストがかかるが無線式のモーションキャプチャではなく光学式のものを用いてモーションを取得して学習を行うことで、さらに精度が向上するのではないかと考えている。

本実験では、3D 姿勢推定の精度指標として P-MPJPE を用いたが一般的な姿勢推定ではなくデリバリー動作に特化した姿勢推定を行う上では、指標として不十分な可能性がある。デリバリー動作は一般的にストーンを持つ手と、左足の爪先か踵、伸ばした脚が進行方向に対して真っ直ぐな状態が正しい姿勢とされている。一方本実験の推定結果を確認すると、画像では真ん中の線に従ってストーンと両足が一直線に伸びているが、推定結果は上記 3 点が一直線上に存在しない。そのため、現状の学習モデルの状態では十分な精度があるとは言えない。他にもカーリングにおけるフォームの特性が存在する可能性があるため、その点の調査を行い、精度評価の検討や学習損失として用いることで精度の向上が期待できる。

次に動作解析についての検討を行う。

本実験の評価指標として用いた P-MPJPE は本研究の最終目的である良し悪し判定にも適用できるのではないかと考えられる。例えば、初級中級者であれば、目標としている選手の動作を正解データとして自分の推定結果との P-MPJPE を算出することで、どのくらい似ているフォームを行えているが数値として確認することができる。また、上級者であれば、自分の中で一番良いプレーを正解データとして姿勢推定を行えばその日のプレーの良し悪しが判断できると考えられる。さらに、P-MPJPE を時系列とする

ことで時間軸上での動作の良し悪しの判断も可能となる。上記を実現するためには、デリバリー姿勢推定の精度の更なる向上はもちろんであるが、他選手との姿勢推定結果との比較の際は、デリバリーの時間が異なるため、時系列を揃える手段を検討が必要になる。

7. 画像解析システム

本章では、2020年10月にオープンしたアルゴグラフィックス北見カーリングホールに導入された画像解析システムについて述べる。

画像解析システムでは、2方向のカメラを用いてカーリング競技者の動作を推定するシステムである。クライアントPCで2方向のデリバリー動画を撮影し、姿勢推定サーバへ送信。サーバからの解析結果をクライアントPCで受け取り結果を可視化できる。可視化できる内容として、2Dの姿勢推定結果と、3Dの姿勢推定結果については、360°から骨格の確認が可能である。また、それぞれの推定された骨格の時系列データをグラフとして出力しているため、不自然な動作等あれば動画と照らし合わせいろいろ確認できる。さらに、2つのデリバリー動画の比較を行うことができ、3D骨格を用いて比較が行える。

現時点(2021年10月23日)では、目視での解析しか行っていないため、解析方法を検討して結果の出力を行えるようにしたい。

8. おわりに

本稿では、カーリングのデリバリー動作の良し悪しを判定するシステムの構築を目指し、予備調査として、画像処理技術の一分野である人体姿勢推定を用いてデリバリー動作の動画から骨格の推定を行えるかどうかの実現性調査を行った。結果として、用意した姿勢推定モデルについて2Dと3Dのデリバリー学習データを用いてデリバリーに特化した学習モデルを作成することで骨格の推定が行えることが確認できた。

今後は、考察で検討した内容を踏まえて動作解析を行い、客観的にデリバリーがショットに影響があるかの調査や、主目的である良し悪し判定のアルゴリズムの検討を行いたい。

参考文献：

- [1] 桂大地, et al. 単眼カメラを用いたボルダリング完登判定システムの提案. エンタテインメントコンピューティングシンポジウム 2019 論文集, 2019, 2019: 145-153.
- [2] 中村拓馬; 森裕一; 矢入郁子. 動画からのサッカーシュート動作の姿勢推定の精度改善. In: 人工知能学会全国大会論文集 第34回全国大会 (2020). 一般社団法人 人工知能学会, 2020. p. 3M5GS1204-3M5GS1204.
- [3] 清水幹弥, et al. バドミントン競技経験者と未経験者のスマッシュ動作中の運動連鎖に関する研究. 東海大学スポーツ医科学雑誌, 2021, 33: 7-14.
- [4] 佐藤幸一郎, et al. 熟練度の違いによるサッカーボールリフティングの運動制御機構. バイオメカニズム学会誌, 2021, 45.2: 95-101.
- [5] DANG, Qi, et al. Deep learning based 2d human pose estimation: A survey. Tsinghua Science and Technology, 2019, 24.6: 663-676.
- [6] XIAO, Bin; WU, Haiping; WEI, Yichen. Simple baselines for human pose estimation and tracking. In: Proceedings of the European conference on computer vision (ECCV). 2018. p. 466-481.
- [7] GitHub - microsoft/human-pose-estimation.pytorch: The project is an official implement of our ECCV2018 paper "Simple Baselines for Human Pose Estimation and Tracking(<https://arxiv.org/abs/1804.06208>)" (<https://github.com/microsoft/human-pose-estimation.pytorch>) (2021/10/15)
- [8] LIU, Junfa, et al. A Graph Attention Spatio-temporal Convolutional Network for 3D Human Pose Estimation in Video. arXiv preprint arXiv:2003.14179, 2020.
- [9] GitHub - fabro66/GAST-Net-3DPoseEstimation: A Graph Attention Spatio-temporal Convolutional Networks for 3D Human Pose Estimation in Video (GAST-Net) (<https://github.com/fabro66/GAST-Net-3DPoseEstimation>) (2021/10/15)
- [10] LIN, Tsung-Yi, et al. Microsoft coco: Common objects in context. In: European conference on computer vision. Springer, Cham, 2014. p. 740-755.
- [11] IONESCU, Catalin; LI, Fuxin; SMINCHISESCU, Cristian. Latent structured models for human pose estimation. In: 2011 International Conference on Computer Vision. IEEE, 2011. p. 2220-2227.
- [12] PN Pro Landing Page (<https://neuronmocap.com/ja/node/5199>) (2021/10/17)