Vol.2021-NL-248 No.36
Vol.2021-CVIM-226 No.36
2021/5/21

スポーツ選手のマーカレスモーションキャプチャーのための効率的なOpenpose再学習

北村　卓弥[1]　トマ　ディエゴ[1]　川崎 洋[1]

概要：モーションキャプチャを行う際、2 次元の姿勢推定後、三角測量を用いてスケルトンの合成を行うことで、マーカレスでモーションキャプチャを行うことができる。モーションキャプチャにより得られる 3 次元のスケルトンの解析や可視化により、スポーツ分野における運動動作の最適化などへの応用が期待される。2 次元の姿勢推定手法として OpenPose と呼ばれる手法が存在し、画像から高精度で 2 次元のスケルトンを生成可能である。一方で、OpenPose には倒立等の複雑な姿勢においては姿勢推定に失敗が生じる問題が存在する。そこで、本研究では、このような複雑な姿勢における姿勢推定精度を向上するために、複雑な姿勢を多く含むスポーツデータセットを作成し、それを用いてネットワークを再学習することで、精度の向上を図る手法について研究を行った。

1. はじめに

近年、最先端のテクノロジーがスポーツの分野に取り入れられる場面が増えている。サッカーにおけるハイスピードカメラや磁場センサーを用いたゴール判定システム等の判定への活用、ドローンによる選手の追跡、多人数スポーツにおける戦術の AI を用いた解析など、活用分野は多岐に渡る。トレーニングにおいても活用が見られ、様々なセンサーや機器を用いて選手の動きや力の強さなどを測定、解析することにより、選手の動きの最適化や、体の状態の管理が行われている。特に姿勢推定を行い、人体のスケルトンを生成することができれば、スケルトンの角度を用いて運動の解析を行うことができる。

現在、姿勢推定を行う手法は複数存在し、モーションセンサーや画像から検出を行う手法が存在する。モーションセンサーを用いる手法は人体にマーカーを付けることで、人間の骨格の動作の検出などを行うことができる。一方で、モーションセンサーなどの専用機材は高価な場合が多く、測定する環境も限られてしまう。また、人体にマーカーが必要になり、運動の妨げになる可能性も存在する。そこで、複数の RGB カメラで撮影を行い、各画像ごとに 2 次元の姿勢推定を行った後、三角測量により 2 次元スケルトンの合成を行い、3 次元のスケルトンを作ることで、マーカーレスで場所を選ばない、安価なシステムを構築することができる。

三角測量を用いて 3 次元のスケルトンを合成するにあたり、まず 2 次元のスケルトンを生成することが必要となる。OpenPose[1] は 1 枚の RGB 画像から 2 次元のスケルトンを推定可能な畳み込みニューラルネットワークであり、今日幅広く用いられている。OpenPose は複数人が含まれる画像においても高い精度を発揮する。しかし、スポーツの分野においては、倒立のような通常の生活では見られない姿勢が多く存在しており、OpenPose はこのような画像を入力とした姿勢推定においての失敗が見られる。原因としては、OpenPose の学習に使われている COCO データセットには倒立のような姿勢が希少であるためだと考えられる。そこで、本研究ではこのような推定困難な画像を多数含むスポーツデータセットを作成し、このデータセットを用いて OpenPose の再学習を行うことで、複雑な姿勢に対しての精度の向上を図り、スポーツ分野での活用を目的とした。

また、再学習によるネットワークの更新を行う際、スポーツデータセットの情報に過学習してしまい、元の OpenPose の精度が低下してしまう問題が生じた。そこで、Augmentation や COCO データセットを混ぜて再学習を行うことにより、従来姿勢、複雑な姿勢の両方のデータでの精度の改善及び、それらの効果の検証を行った。さらに、OpenPose の推論時に回転させた画像を入力させた場合の精度との比較も行った。

本研究での貢献は以下である。(1) 本研究では 2 次元の姿勢推定手法である OpenPose の再学習及び事前回転手法を行い、評価を行った。(2) 再学習した OpenPose から生成されるアノテーションを用いて, 通常のデータセットに

[1] 九州大学

稀有な複雑な姿勢を多く含む、スポーツデータセットを作成した。(3) 過学習対策及び推定精度の向上のために、Augmentation や COCO データセットを混ぜる手法についての検証、比較を行った。

2. 関連研究

2.1 3 次元姿勢推定

近年、3 次元の姿勢推定を行う研究はさかんに行われている。3 次元の関節座標に関するアノテーションは主にモーションキャプチャを用いて行われており、大規模なデータセットが用意できないという問題が存在する。[9] は畳み込みニューラルネットワーク (CNN) を用いる手法であり、自動エンコーダーを使った潜在表現の学習を行い、姿勢の構造を考慮することで性能の向上を図っている。[10] は人体の周囲の 3 次元空間を細かく離散化し、各ボクセルにおける各関節の尤度を示す 3 次元ヒートマップを採用し、CNN を用いて学習を行うことで、2 次元のみでなく、3 次元においてもヒートマップが活用できることを示した。[11] では 2 次元の姿勢推定とデプス推定のタスクを 2 次元、3 次元それぞれのアノテーションを用いて同時に学習することで、3 次元データセットに不足している屋外画像の 3 次元姿勢推定の性能を向上させている。[12] は 3 次元の姿勢データとそれに対応する複数方向からの 2 次元姿勢をセットで学習に用いており、2 次元の姿勢推定を行った後、3 次元に復元する際、デプスを 3 次元の姿勢ライブラリの中で 2 次元推定結果にマッチングするものから選択しており、オクルージョン領域の部位についても正確な姿勢推定を行っている。[8] は複数の 2 次元画像から三角測量を行うことで 3 次元のスケルトンを生成している。Algebraic Triangulation、Volumetric Triangulation の 2 種類の方法を用いて 3 次元ポーズを生成しており、既存手法の性能を大きく上回っている。本研究では [8] のように三角測量により、2 次元のスケルトンを合成することで、3 次元のスケルトンを生成し、スポーツ分野での活用が可能であると考えている。

2.2 2 次元姿勢推定

2 次元の姿勢推定手法として、[3] は人物の姿勢推定に初めてディープラーニングを用いた手法である。人間検出器で人間を検出した後、人物ごとに姿勢推定を行うトップダウン型の手法であり、ディープラーニングが姿勢推定に有効であること、及び、カスケードが有効であることも示した。同じくトップダウン型の手法である [5] は第 1 ステージで各部位についての信頼度マップを作成し、その後のステージでは一つ前のステージの信頼度マップを参考に部位間の関連性を学習しつつ、関連性の受容野を広げることで信頼度マップを洗練している。一方で、人数の増加に比例して人物ごとに行う姿勢推定のタスクが増えるので、人数が多い場合には不向きである。OpenPose[1] は画像上に存在するキーポイントを抽出した後、キーポイントをマッチングさせるボトムアップ型の手法を採用しており、複数人の姿勢推定の精度を向上した。ボトムアップ型は人数の増加によって、計算量が増加しにくいが、マッチングには時間がかかってしまう問題が存在する。そこで OpenPose は PAF(Part Affinity Field) と呼ばれる関節間ベクトルを採用することでマッチングの問題を解決し、リアルタイムで高性能な姿勢推定を行うことに成功した。

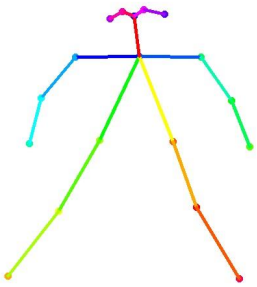

図 1 OpenPose から出力される 2 次元スケルトンのフォーマット

3. スポーツ動画での OpenPose の利用

3.1 OpenPose の問題点

OpenPose はリアルタイムで複数人の姿勢検出を高精度で行うことができる (図 2)。一方で、人物が倒立しているような動作に対しては、手と足が逆転して推定されるなど、姿勢推定に失敗する。これは、学習に用いられた COCO データセット [2] には倒立のような非日常的なポーズが少ないためだと考えられる。OpenPose は部位を示す信頼度マップと部位間のベクトルである PAF(Part Affinity Field) を学習する手法であり、直立のポーズと逆立ちのポーズでは関節の位置と関節間の関連性などが大きく異なっているため、倒立等のポーズが持つ関節の位置と関節間の関連性についての情報を改めて OpenPose の学習に取り込む必要があると考えられる。本研究では倒立等のポーズが持つ情報を OpenPose に学習させ、推定可能にすることを目的として研究を行った。図 3 に OpenPose の失敗例を示す。図 1 に示すように、スケルトンの赤色は左足部分、薄緑は右足部分、濃緑は左手部分、青色は右手部分を示すが、図 3 では手と足の色が入れ替わっており、姿勢推定が失敗していることが分かる。

3.2 スポーツ動画用の学習データ作成

OpenPose の学習には、人物を含む画像 64115 枚とアノテーションからなるトレーニングデータセットと 2693 枚の画像とアノテーションからなるバリデーションデータセットで構成されたバリデーションデータセットから構成される COCO データセット [2] が用いられている。これ

Vol.2021-NL-248 No.36
Vol.2021-CVIM-226 No.36
2021/5/21

図 2 OpenPose の成功例

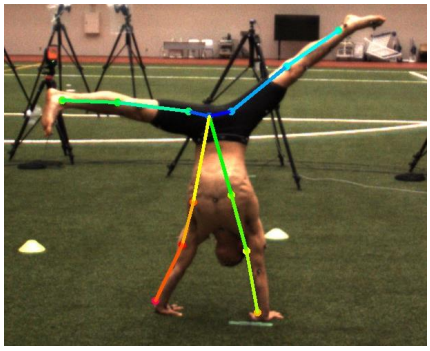

図 3 OpenPose の失敗例

らのデータセットには倒立のような複雑な姿勢とそのアノテーションが希少である。本研究では体操選手の一連のスポーツシーンの映像を用いて、そのような希少なシーンが多く含まれるトレーニングデータセットを作成した。映像は選手が倒立や後方転回 (バク転、バク宙) を複数回行う 25 秒間のものである。これらの動画は複数方向から同時に撮影されている。正面と背後からの動画をそれぞれ 750 フレーム (計 1500 フレーム) に分割し, 各フレームに対して作成したアノテーションツール (図 4) を用いて 18 個キーポイントについてアノテーション付けを行った。アノテーションツールは OpenPose の出力などにより得られるアノテーションと画像を入力とし、GUI 上でキーポイントの位

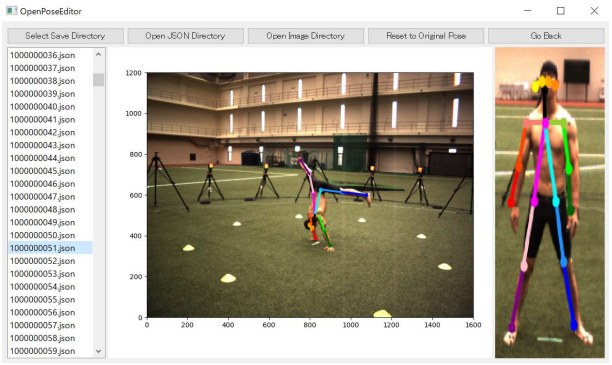

図 4 手動アノテーションツールを用いた正しいアノテーション付け

置を修正することが可能である。OpenPose が必要とするアノテーションが首を除いた 17 個であったため、GUI での修正後に変換を行った。アノテーションを修正した 1500 枚のうち、1200 枚をトレーニング用として、300 枚をバリデーション用とした。

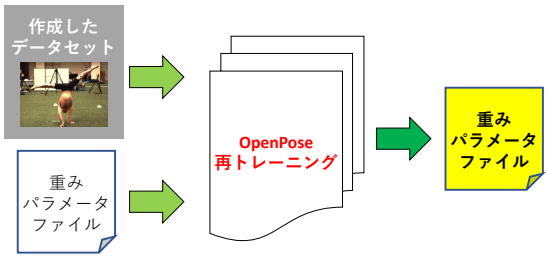

図 5 スポーツデータセットと学習済みの重みを用いた再学習手法

3.3 OpenPose の再学習

図 5 に示すように OpenPose の既存の重みファイルと、前節で作成したスポーツデータセットを入力として再学習を行った。出力として、再学習されたネットワークの重みパラメータファイルが生成される。出力される Loss 情報,heatmap(図 6),PAF(Part Affinity Field)(図 7)、推論結果を元に、学習のハイパーパラメータを調整することで、ネットワークの最適化を行った。

3.4 過学習対策

上記の手法を用いて、OpenPose の再学習を行うことで、倒立等に対する推定精度の向上が見られたが、一方で、過学習により従来の推定精度が低下してしまう問題が生じた。図 8 に示すように通常の姿勢が上下反転してしまう。この問題に対処するために、本研究では 2 つの手法、及びその両方を組み合わせた。

1 つ目は作成したデータセットに COCO データセットを混ぜて再学習を行う手法 (Mix と呼ぶこととする) である。COCO データセットは OpenPose の学習に使われた

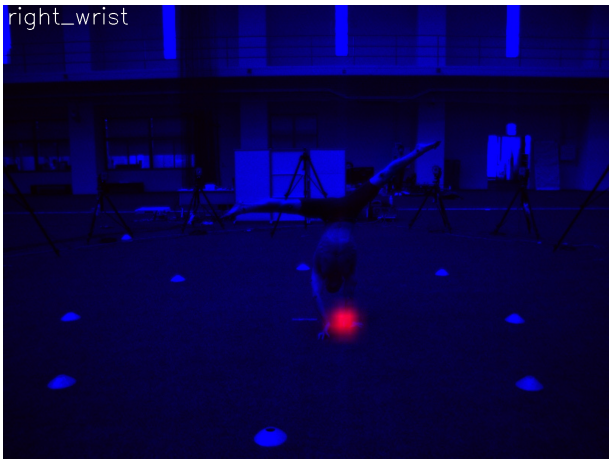

図 6 heatmap の例

図 7 PAF(Part Affinity Fields) の例

データセットであるため、このデータセットを混ぜて学習を行うことで、従来の推定精度を維持しつつ再学習をすすめることができないか検証を行った。

2 つ目に Augmentation を行った。従来のデータセット (約 65000 データ) に比べ、作成されたスポーツデータセット (1500 枚) は大幅に数が少ない。そこで、スポーツデータセットを Augmentation により、データ数を増加させることで、OpenPose の精度向上を図った。Augmentation としては 8 方向 (45°,90°,135°,,,) の回転と、水平方向の反転、スケールの調整を行っており、様々な角度のポーズを学習させることで、姿勢の方向についてのロバスト性の確保についても検証を行った。また、上記の手法を組み合わせる手法についても検証を行った。

3.5 事前回転手法

OpenPose が倒立等に失敗する原因は、OpenPose の学習に使われた COCO データセットが、倒立等の画像を含んでいないためである。そこで再学習手法とは別に、OpenPose の推論時に入力画像を予め回転させ、画像上向きに頭部が位置するように、画像下部に足が位置するようすること

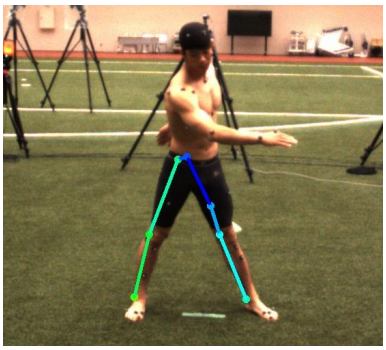

図 8 過学習による失敗例

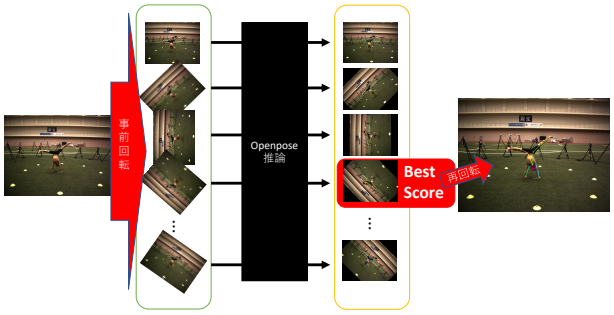

図 9 事前回転手法

で、OpenPose が十分な推定を行うことができると考え、実行した。手法としては、予め 8 方向に回転させた画像を OpenPose の推論の入力とし、8 枚の出力結果を得た。OpenPose の出力の一つであるキーポイントの信頼度スコアの合計値を用いて、8 枚の中で最も適当なものを最終的な出力とした。

4. 実験

4.1 検証方法

検証に用いるテストデータセットとして 2 つのデータセットを用いた。1 つ目は COCO データセットのバリデーションデータセットを用いた。このデータセットは OpenPose の従来の学習のバリデーションに用いられており、このデータセットを用いて検証を行うことで、再学習後も倒立等のポーズに過学習を起こさず、従来の OpenPose の精度を維持しているかの確認を行った。2 つ目のデータセットとして、スポーツテストデータセットを作成した。トレーニングのために作成したスポーツデータセットと同様の手法で作成を行い、トレーニングに用いた動画とは別の動画から抽出した 81 枚の画像とアノテーションで構成されている。このデータセットを用いることで、OpenPose が倒立のような複雑な姿勢に対してどの程度推定できるようになったかの確認を行った。

評価の指標として、COCO データセットより用意されている AP スコアを用いた。COCO データセットで用意されている AP スコアは、キーポイントの推定値と真値の関連度

Vol.2021-NL-248 No.36
Vol.2021-CVIM-226 No.36
2021/5/21

	COCOバリデーションデータセット			Kanoyaテストデータセット		
	AP	AP50	AP75	AP	AP50	AP75
オリジナル	0.457	0.712	0.475	0.194	0.427	0.081
①：再学習	0.245	0.478	0.225	0.461	0.828	0.446
②：①+Augmentation	0.143	0.304	0.116	**0.521**	**0.903**	**0.545**
③：①+Mix (COCO+スポーツデータセット)	0.378	0.645	0.37	0.29	0.725	0.165
④：②+③	**0.413**	**0.675**	**0.419**	0.483	0.892	0.412
事前回転	0.374	0.639	0.37	0.221	0.627	0.061

図 10 OpenPose の結果

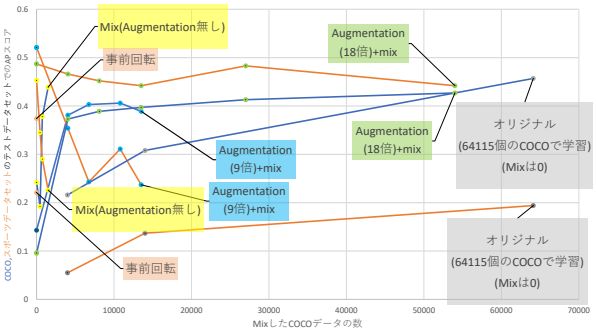

図 11 COCO のデータ数による結果の推移

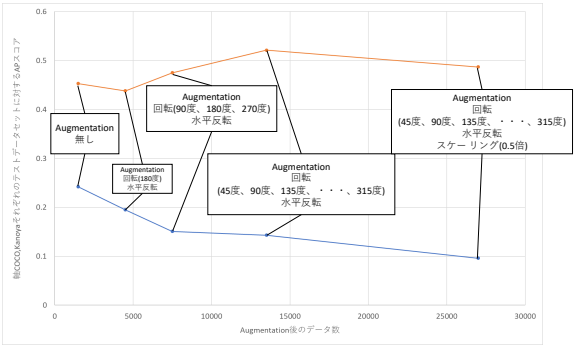

図 12 Augmentation 後のデータ数による結果の推移

を示す値である Oks を複数の閾値を用いて真偽を判定し、その真偽による Precision と Recall の積分値を AP スコアとしている。閾値ごとに、一般的に 3 つの指標での比較が行われており、AP50 は閾値 0.5 の場合の AP スコアを、AP75 は閾値 0.75 の場合を、AP は AP50,AP55,AP60,・・・,AP95 の平均値である。AP スコアをそれぞれのテストデータセットに対して計算し、比較を行うことで検証を行った。Oks の計算において、人物部分の領域面積が必要となるが、COCO バリデーションデータセットは領域面積のアノテーション情報も保持しているのに対し、スポーツデータセットでは保持していない。そのため、代わりに Bounding box の 1/2 の値を領域面積として用いることで Oks の計算を行った。そのため、2 つのテストデータセット間の AP スコアを単純に比較することは困難である。

4.2 結果

オリジナルの OpenPose、スポーツデータセットを用いた再学習、Augmentation を用いた手法、スポーツデータセットに COCO データセットを混ぜる手法 (Mix)、Augmentation+Mix(スポーツデータセットの Augmentation 後に COCO データセットを混ぜる)、事前回転手法の 6 つを比較し、2 つのテストデータセット (スポーツデータセット、COCO データセット) での AP,AP50,AP75 での結果を図 10 に示す。2 つのデータセットはデータ数の規模が異なり、AP スコアの算出に必要となる人物領域の面積の算出方法も異なるため、2 つのデータセット間の値の比較は困難である。

結果からスポーツデータセットを用いて再学習を行うことで、スポーツテストデータセットの値が 0.194 から 0.461 と大幅に改善しており、倒立等に対する推定精度が向上していることが分かる。一方で COCO テストデータセットに対する結果としては 0.475 から 0.225 と下がってしまっている。これは、OpenPose が倒立等のデータに対して過学習を起こした結果だと考えられる。2 つの過学習対策を行った結果、Augmentation を行うことで倒立等に対しての精度がますます向上することが分かった。また、Mix の手法には従来の OpenPose の精度を保つ効果が見られ、2 つの過学習対策を組み合わせることで、スポーツテストデータセットに対して 0.483,COCO テストデータセットに対して 0.413 という結果を出しており、従来の精度を十分保ったまま、逆立ちのような複雑の姿勢を推定することができるようになった。事前回転手法の精度は再学習を用いた手法よりも劣っている。原因として、回転時の姿勢は足が 180° 開いているなどの回転時のみ可能なポーズが存在し、単純に回転させ画像上部に上半身が位置するようにするだけでは、地面に足がついていないなど不自然なものとなってしまい、推定精度が低下してしまうと考えられる。また、推定にかかる時間も 8 倍になってしまう。そのため、スポーツのようなリアルタイムでの活用を考えた場合、事前回転手法よりも再学習を用いる手法が有効だと考えられる。

また、スポーツデータセットに COCO データセットを混ぜる手法 (Mix) について、混ぜる COCO データ数による過学習抑制効果の変化を図 11 に示す。横軸を混ぜた COCO データセットの数 (ただし、オリジナルは従来の学習時に用いた COCO データセットの数)、縦軸を AP スコアとして、様々な条件下で実験を行った。オレンジの線はスポーツテストデータセットに対する結果を示し、青の線は COCO テストデータセットに対する結果を示す。混ぜる COCO データセットの数が増えるほど、COCO テストデータセットに対する結果が向上していることが分かる。また、Augmentation と組み合わせることで、学習に用いるスポーツデータセットのデータを増加させ、混ぜること

Vol.2021-NL-248 No.36
Vol.2021-CVIM-226 No.36
2021/5/21

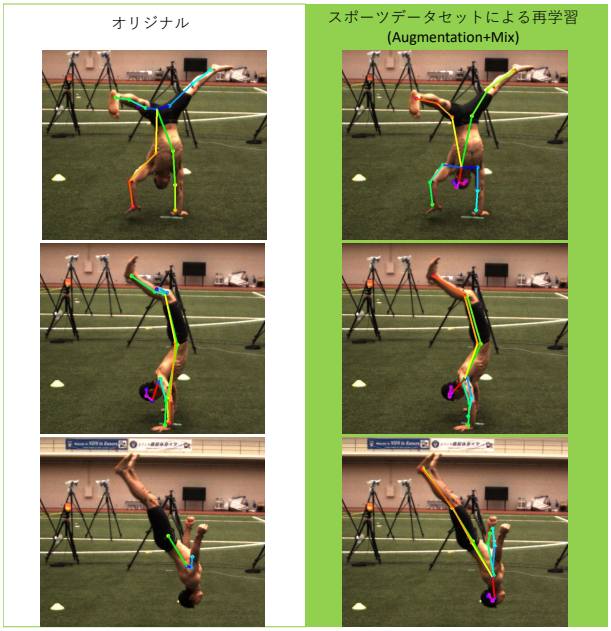

図 13 オリジナルと再学習の比較結果 (スポーツテストデータセット)

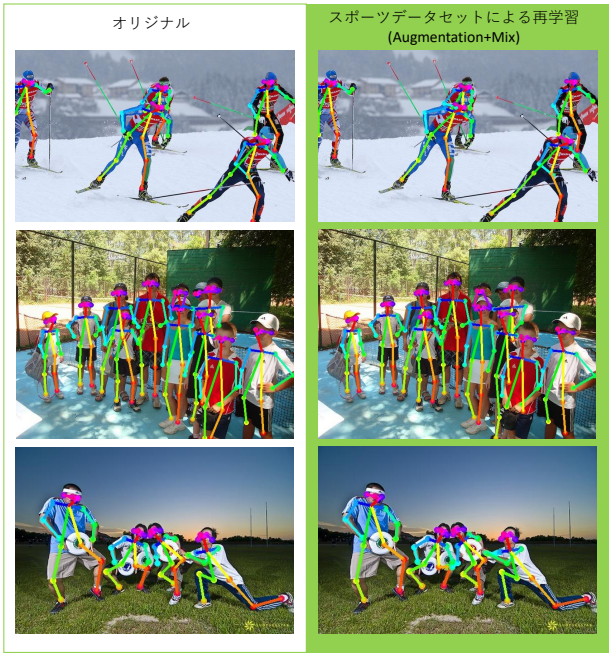

図 14 オリジナルと再学習の比較結果 (coco バリデーションデータセット)

のできる COCO データセットのデータ数も増やすことができるため、両方の結果を向上することが分かった。

本研究では Augmentation で増やす量についても検証を行った。図 12 に示すように、Augmentation を行うことで、再学習に用いるデータのバリエーションが増え、倒立等の複雑な姿勢の推定精度が向上した。COCO テストデータセットに対する精度は低下する傾向が見られる。本研究で用いたスポーツデータセットは、画像中の人物が一人であり、COCO テストデータセットでは複数の人物が画像中に存在する。そのため、COCO テストデータセットに対する精度が低下したのではないかと見られる。

また、図 13,14, に、実際に画像上での結果を示す。オリジナルの OpenPose と図 10 で 2 つのデータセットに対して高い精度を示した Augmentation + Mix の手法の比較を行った。図 13 に示すように倒立等のポーズにおいて、従来の OpenPose は失敗しているが、本研究の手法を用いることで推定が可能となった (オリジナルも成功しているように見えるが、スケルトンの色を図 1 と比較すると間違っていることが分かる)。また、図 14 に示すように、Mix の手法により過学習を抑制し、従来の OpenPose の精度を十分維持できていることが分かる。

5. まとめ

本研究では OpenPose の再学習を行うことで、従来の OpenPose が推定できない倒立等の複雑な姿勢を推定する手法を提案した。また、単純に倒立等のデータを混ぜて学習を行うだけでは、過学習による推定精度の低下が生じてしまうため、Augmentation 及び COCO のデータを一定の割合を混ぜて学習を行うことで、従来の OpenPose の精度を維持しつつ、倒立等の姿勢を推定可能であることを示した。 提案手法を用いることで、スポーツの分野などで登場する複雑な姿勢を推定することが可能となる。本手法を用いたスポーツ選手の測定データは、マーカーや特別な機材を使わずないため、より手軽に運動動作の改善を行うことが期待できる。

6. 謝辞

本研究は JSPS 科研費 JP20H00611,JP18K19824 ,JP18H04119 の助成を受けたものです。また、スポーツデータの取得にあたり、鹿屋体育大学のスポーツパフォーマンス研究センターを使用し、和田智仁先生、本嶋良恵先生および学生の皆様に協力頂きました。ここに感謝申し上げます。

参考文献

[1] Cao, Zhe and Simon, Tomas and Wei, Shih-En and Sheikh, Yaser : Realtime multi-person 2d pose estimation using part affinity fields . Proceedings of the IEEE conference on computer vision and pattern recognition(2017) .

[2] Lin, Tsung-Yi and Maire, Michael and Belongie, Serge

and Hays, James and Perona, Pietro and Ramanan, Deva and Dollar, Piotr and Zitnick, C Lawrence: Microsoft coco :Common objects in context . European conference on computer vision(2014).

[3] Alexander Toshev and Christian Szegedy : DeepPose: Human Pose Estimation via Deep Neural Networks . CoRR(2013) .

[4] Tompson, Jonathan and Jain, Arjun and LeCun, Yann and Bregler, Christoph : Joint training of a convolutional network and a graphical model for human pose estimation(2014). arXiv preprint arXiv:1406.2984.

[5] Shih-En Wei and Varun Ramakrishna and Takeo Kanade and Yaser Sheikh : Convolutional Pose Machines. CoRR(2016).

[6] Leonid Pishchulin and Eldar Insafutdinov and Siyu Tang and Bjoern Andres and Mykhaylo Andriluka and Peter V. Gehler and Bernt Schiele : DeepCut: Joint Subset Partition and Labeling for Multi Person Pose Estimation. CoRR(2015).

[7] Eldar Insafutdinov and Leonid Pishchulin and Bjoern Andres and Mykhaylo Andriluka and Bernt Schiele : DeeperCut: A Deeper, Stronger, and Faster Multi-Person Pose Estimation Model. CoRR(2016).

[8] Iskakov, Karim and Burkov, Egor and Lempitsky, Victor and Malkov, Yury : Learnable triangulation of human pose. Proceedings of the IEEE/CVF International Conference on Computer Vision(2019).

[9] Tekin, Bugra and Katircioglu, Isinsu and Salzmann, Mathieu and Lepetit, Vincent and Fua, Pascal : Structured prediction of 3d human pose with deep neural networks. arXiv preprint arXiv:1605.05180(2016)

[10] Georgios Pavlakos and Xiaowei Zhou and Konstantinos G. Derpanis and Kostas Daniilidis : Coarse-to-Fine Volumetric Prediction for Single-Image 3D Human Pose. CoRR(2016)

[11] Zhou, Xingyi and Huang, Qixing and Sun, Xiao and Xue, Xiangyang and Wei, Yichen : Towards 3D Human Pose Estimation in the Wild: A Weakly-Supervised Approach. The IEEE International Conference on Computer Vision (ICCV)(2017)

[12] Chen, Ching-Hang and Ramanan, Deva : 3d human pose estimation= 2d pose estimation + matching. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(2017)