

1. はじめに

Skill Assessment は、動画内の特定の動作に着目して、スキルの優劣を評価する手法である。Skill Assessment の手法として、Pairwise Deep Ranking (PDR) [1] や Attention Pairwise Ranking (APR) [2] が提案されている。APR は、入力する動画の中から重要となる領域を獲得し、スキルの優劣判定に用いる。しかし、スポーツのような複雑な動作の優劣判定には、人の姿勢も重要となる。そこで本研究では、骨格情報をグラフとして入力した Skill Assessment として、Attention Pairwise Ranking GCN を提案する。

2. Attention Pairwise Ranking

APR は、スキルの優劣判定だけでなく、PDR の出力結果に対する判断根拠を可視化できる手法である。動画を3つの Segment に等分し、各 Segment からランダムで1枚を選択する。そして、選択した画像を Feature extractor 入力する。APR では PDR に Attention 機構を導入し、獲得した注視領域を Ranking branch に与える。タスクを行う動作には、優れた動作と劣った動作がそれぞれ含まれている。このことから APR には、優れた動作を評価する Superior Network と、劣った動作を評価する Inferior Network の2つを用い、学習させることでそれぞれの動作に対する注視領域を別々に獲得する。

3. データセットの作成

PDR で提案された EPIC-Skills2018 Dataset では、スポーツのような複雑な動作を行うタスクを含まないため、スポーツにおける技術の優劣の判定を行うことは困難である。そこで新たに卓球のサーブを打ち返す動画のデータセットを作成する。被撮影者を正面から捉えるために撮影者の頭に付けたカメラを Head camera、天井に設置することで頭上から被撮影者を捉えるカメラを Top camera、横から被撮影者の動きを捉えるカメラを Side camera とする。それぞれのカメラの配置と各カメラでの撮影例を図1に示す。

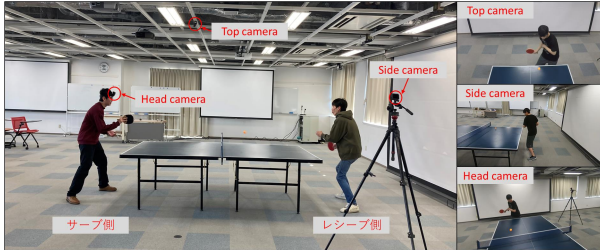


図1：カメラの配置場所と各カメラの視点

作成した卓球データセットの詳細を表1に示す。スポーツにおいて、専門的なスキルを持った人がいなければスキルの優れた動画を集めることは難しい。そのため、スキルの劣った動画の比率が高い不均一なデータセットになる可能性がある。よって、データの均一さによる影響を確認するため、データの不均一なデータセットとデータが均一なデータセットを作成した。サーブを打ち返す動作の動画を卓球経験者により上手、普通、下手に分類するアノテーションを行った。

表1：データセットの動画数

		上手	普通	下手	動画数	ペア数
不均一	train	13	70	178	261	14774
	test	4	23	61	88	1647
均一	train	65	65	65	195	12675
	test	25	25	25	75	1875

4. 提案手法

4.1. 姿勢情報

本データセットに対して、2次元の姿勢推定手法である Evopose2D [3] を使用し、姿勢情報を追加した。ここで姿勢情報は、鼻の推定座標からの12か所の関節位置の相対座標である。

4.2. APR-GCN

APR の Superior Network と Inferior Network の各 Ranking branch に Graph Convolution を追加する。そして図2のように、画像から取得した関節位置の座標からグラフを作成し、Graph Convolution に入力する。これにより、姿勢情報を考慮できるようになる。Graph Convolution から出力される特徴マップを FC 層へ入力する。その出力を ResBlock の Convolution layers から出力される特徴マップと結合し、FC 層へ入力する。

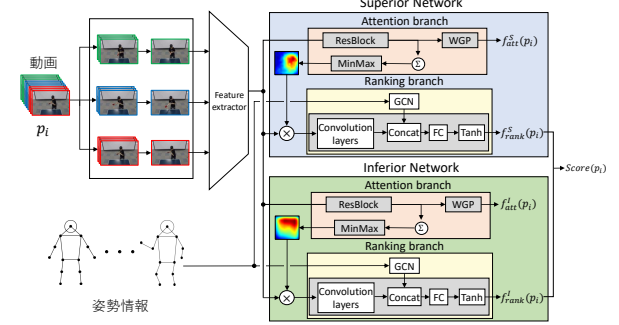


図2：APR-GCN の構造

4.3. 学習方法

損失関数には、APR と同じ Log Ranking Loss を使用する。提案手法の損失 L は式 (1) のようになる。Superior Network では、 $p_i - p_j$ を最大化するように、Inferior Network では、 $p_j - p_i$ を最大化するように学習する。

$$L = \log(1 + \exp(-f(p_i) + f(p_j))) \quad (1)$$

5. 実験結果

本研究で作成したデータセットを用いて APR を評価した結果を表2に示す。実験は Head, Side, Top の3視点の画像を使用し、各視点で5回ずつ学習して精度の平均を比較する。表2より被撮影者側が常に映っている Top の精度が比較的高い傾向にある。そして、カメラのブレが大きい Head や体で打つ際の右腕が隠れてしまう Side では、精度が Top に比べ低い。また、データセットのデータ数を不均一の場合よりも均一にした方が高い精度である。また、最も高精度である Top に姿勢情報を追加し APR-GCN で学習した結果、精度の向上は見られなかった。

表2：実験結果

train	test	姿勢情報	Head	Side	Top
不均一	不均一	なし	74.1	65.5	93.5
均一	不均一	なし	92.3	75.9	94.8
均一	均一	なし	92.3	85.2	93.5
均一	均一	あり	-	-	71.7

6. おわりに

本研究では、Skill Assessment タスクに適したのスポーツのデータセットの作成を行った。また、データセットにおけるデータの均一さによる精度の比較と姿勢情報の有無による精度の比較を行った。評価実験の結果、卓球におけるスキルの優劣を高い精度で判別することが可能であり、提案手法によりスポーツの技能獲得への利用が期待できる。

参考文献

- [1] H.Doughty, *et al.*, “Who’s better? who’s best? pairwise deep ranking for skill determination”, CVPR, 2018.
- [2] 高田雅之等, “Attention Pairwise Ranking によるスキル優劣判定における視覚的説明と高精度化”, MIRU, 2020.
- [3] W.McNally, *et al.*, “EvoPose2D: Pushing the Boundaries of 2D Human Pose Estimation Using Accelerated Neuroevolution With Weight Transfer”, IEEE Access, Vol. 9, pp. 139403-139414, 2021.