ニューラルネットワークによる身体動作分析に関する研究

M2022SE008 小関翔也
指導教員：野呂昌満

1　はじめに

近年，スポーツ分野において機械学習を応用する研究が盛んに行われている．以下の観点で，様々なシステムが提案されている．

-アナリストの観点

アナリストを支援するという観点では，密集するラグビー選手の検出するための技術や，バドミントンのショットを分類する技術の研究がされている．

-スポーツエンタテイメントの観点

非接触センシングデータを用いた，競泳における泳速・ストローク解析のシステムが提案されている．

コーチングの観点から研究がされており，実際にサッカーやアメリカンフットボール，ラグビーなどのプロスポーツでは機械学習を用いた ICT による指導が浸透している．

一方で，日本のカレッジスポーツの現場においては機械学習を用いたツールの普及が顕著ではなく，プレイをビデオカメラで録画して目視で確認をする程度にとどまっている．コーチングに必要なスポーツのフォームの熟練度別の分類には，複数のセンサや，それらを搭載した特殊な道具が必要であり，それらを用意するにはコストがかかるからである．現在の日本においてはカレッジスポーツ市場に割かれるコストが少ないことを一因として，機械学習を用いた ICT の普及が阻害されている．コーチによるコーチングは脆弱であり，その点を解決するために普及が望まれている．

現在，単一視点のスポーツ画像の分類は可能とされていない．少ない情報量から分類することが困難となっているからである．先行研究 [1] ではラクロスのショットフォームの画像を用いて熟練度別の分類を試みた．画像分類には CNN が適しているので，自身で設計した CNN と，学習済みニューラルネットワークである VGG16 の 2 つの CNN を用いて熟練度別の分類を行った．分類精度は 60% 程度であった．60% ほどの精度では実用には至らないので，分類精度の向上が問題点となっている．

本研究の目的は，卒業論文で残された課題を解決し，単一視点のスポーツ画像の分類を可能とすることである．課題はデータセットの質の向上，データセット不足の解決，ニューラルネットの構造の最適化である．これらの課題を解決し，少ない情報量から分類を高精度にするモデルを設計することで，コーチングのために必要な機材を最小限にすることが可能となる．これにより，カレッジスポーツの現場における ICT の普及を実現を目指す．

本研究の目的を達成するために以下の技術課題を達成する必要がある．

RQ1　データ加工の手法の提案

卒業論文時に使用したデータの前処理を行う．課題であったデータの質の向上とデータセット不足の解決をし，分類精度の向上を図る．

RQ2　分類を可能とする予測モデルの提案

ラベル付けされたデータを正確に分類できるモデルを設計する．その設計をもとにプロトタイプを作成する．

RQ3　妥当性の評価

設計したモデルの妥当性の評価を行う．分類精度を卒業論文時の分類精度と定量的に比較して，提案モデルの妥当性の評価を行う．

2　既存研究ならびに関連技術

機械学習を用いて，スポーツ画像から熟練度を分類する研究がされている．以下に先行研究を示す．

2.1　金子らの研究 [3]

金子らは熟練度別に分類することが困難であることを問題視し，サッカーのショットフォームを，機械学習を用いて分類することを目的とした研究を行った．決定木，SVM，ロジスティック回帰の手法を用いて，分類を行っている．

使用したデータは，10 名のショットフォームを Openpose によって関節点を抽出した画像である．ショットフォームは心臓と利き足首の座標の差が最も大きくなる時点の画像である．心臓と利き足首の座標の変化率が最も高く，それぞれのフォームの特徴が発現しやすいので，このシーンが選択された，

この研究では，どの手法による分類でも 80% 以上の精度を出した．Openpose で関節点を抽出し，それぞれの座標によって分類が出来るということが示された．

2.2　渥美らの研究 [4]

渥美らはスキーは冬にしかできないスポーツであり，コーチングの機会が他のスポーツより少ないことを問題視した．冬以外の屋内でもコーチングが可能となるように，CNN を用いてスキーシミュレータ利用時の初心者と熟練者の分類することを目的として研究を行った．

この研究では実際のスキー状況を再現するために，複数のデバイスやセンサを使用している．スキーシミュレータや VR デバイスを用いることで，室内での姿勢データの取得を可能としている．荷重データの取得のために靴のインソールに圧力センサを取り付けている．

YOLOv3 を参照した CNN を用いて，熟練度を 4 段階に分けた分類がされている．分類精度は 90% 以上であり，姿勢データと荷重データによる分類の有効性が示された．

2.3 Libin らの研究 [5]

Libin らは深層学習の浸透の結果，分類器として CNN を用いるようになったことを疑問視した．それまで分類器として使われていた SVM との比較を目的として研究が行われた．分類の 3 つのセンサを使用して，ゴルフスイングの一連の動作から得たデータによって分類をしている．プレイヤーは 3 つのセンサが付いているゴルフクラブを使用してスイングを行った．センサはそれぞれ，ゴルフクラブの歪み，加速度，角速度を計測するものである．

分類には各年の ILSVRC で優勝した 3 つのニューラルネットを参照に設計したモデルを用いている．参照した 3 つの NN はそれぞれ，VGGNet [6]，GoogLeNet [7]，ResNet [8] である．最新の NN になるにつれて、層の数は増加し，勾配消失問題が解決された構造になっている．

全てのデータを合わせたデータ，歪みデータ，加速度データを用いた分類において分類精度が 90% を上回った．センサを取り付けて得たデータからゴルフスイングの分類が可能であった．SVM との定量的な比較を行い，分類器として CNN を用いることの有効性が示された．

3 技術課題の解決

本研究では単一視点のスポーツ画像の熟練度別分類精度を向上させるために，新たな CNN を設計する．問題解決へのアプローチを図 1 に示す．

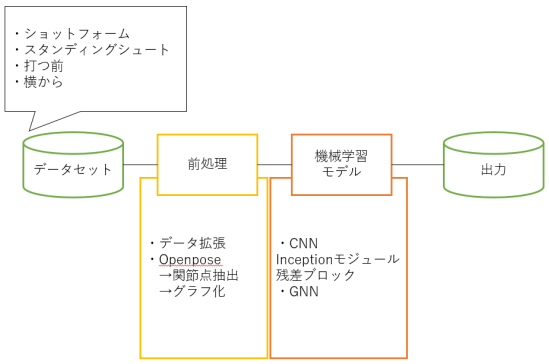

図 1 技術課題解決のアプローチ

データセットは卒業論文時に使用したものを拡張・加工したものを用いる．先行研究の考察では，分類精度の向上のために，データセット不足・質の悪さの改善が挙げられた．それらの改善を図るために前処理を行う．データセットの前処理の手法については節 4.1 で述べる．

卒業論文時には自作した CNN と VGG16 を用いたが，層の深さは最大で 16 層であった．画像分類に用いる CNN では一般に，層が多ければ多いほど細かな特徴抽出が可能となる．新たなモジュールを用いることで，NN の層の数を増やすことが可能となる．層を増やすことによって過学習や勾配消失問題が発生しやすくなるが，それらについても別のモジュールを用いることで解決が可能である．本研究ではそれらのモジュールを使用して層の深さを深くした CNN を設計する．用いるモジュールについては節 4.2 で述べる．

出力は教師ラベルに則って，初心者と熟練者の二種類に分類する．二値分類では一般に sigmoid 関数が使われるので，本研究においても sigmoid 関数を用いる．

4 データセットの前処理及び提案モデル

4.1 データセットの前処理

先行研究では分類精度向上のための課題として，データセットの質の向上が挙げられた．本研究では Openpose[9] を用いて画像の前処理を行い，データセットの質の向上を行う．Openpose による関節点抽出とデータ拡張をデータセットの前処理とする．

Openpose は Cao らが発表した DeepLearning を活用した姿勢推定ツールである ．画像を分析することで首，肩，肘など 18 点の関節点を抽出する．Openpose を用いて関節点を抽出することでフォーム分類に不必要であった背景などの情報を削減することができる．

Openpose を用いて前処理を施した画像を図 2 に示す．図 2 は関節点を青色，それらを繋ぐ線を緑色，背景と黒色としている．背景や体形，道具の色や形によらず，関節の情報のみからの分類を行えるような加工を施した．

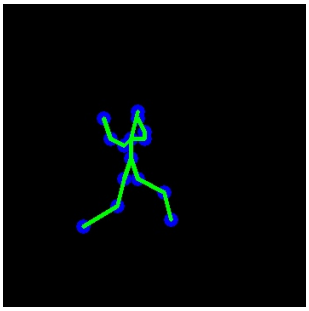

図 2 前処理を施した画像

次に，OpenPose を適用したデータの拡張を行い，データセット不足を解決する．本研究ではデータの RGB 値を変更してデータ拡張をした．使用した色は，赤，青，緑，白の 4 種類である．関節点の色とそれらをつなぐ線の色を，4 種類の色のうちから 2 色選択し，1 つの画像から 12 種類のデータに拡張した．上級生のデータに関しては，データ量の関係からグレースケール化した画像も用意した．図 3 は拡張したデータの図である．

拡張を行ったことで 1 クラス 2000 枚のデータを確保した．学習において十分な量のデータであるとし，データセット不足の解決を試みた．

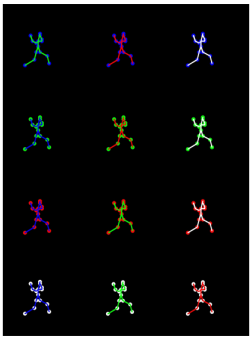

図 3　拡張したデータ

4.2　提案モデル

提案するモデルに新たなモジュールを追加する．Inception モジュール，残差ブロック (Residual-block) を組み込むことで CNN の層を深くすることが実現出来る．本研究ではこれらのモジュールを組み込んだ CNN に入力する．層を深くすることで上級生と下級生の特徴の僅かな差異の抽出を可能とし，分類精度の向上を目標とする．

4.2.1　Inception モジュール

Inception モジュールは 2014 年に発表された GoogLeNet に導入されている．それ以前に発表された VGGNet は層が 16 層であったのに対し，GoogLeNet は 22 層である．このモジュールを用いることで勾配消失を起こすことなく，層の増加を実現した．Inception モジュールのアーキテクチャを図 4，図 5 に示す．

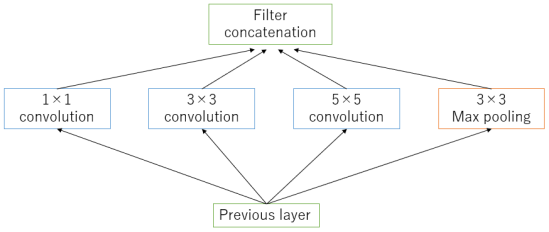

図 4　inception モジュールの基本構造 [7] の図を再作図

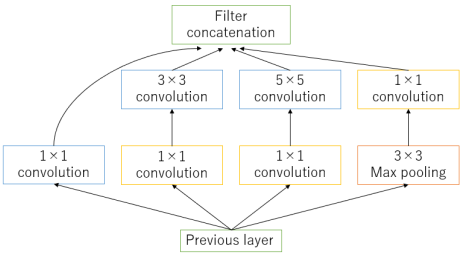

図 5　inception モジュールの応用構造 [7] の図を再作図

これまでの CNN のアーキテクチャは、畳み込み層を単純に繋げていた．単純に繋げて，畳み込み層を増やしていくと画像サイズが小さくなる．これが原因で層を増やすことが出来なかった．この問題を解決するために Inception モジュールが追加された．図 4 に inception モジュールのアーキテクチャを示す．1 つの画像を 1 × 1，3 × 3，5 × 5 の畳み込み層と 3 × 3 のプーリング層に入力し，それぞれの層の出力を結合して 1 つの出力とする．図 5 は基本構造に畳み込み層を追加したモジュールのアーキテクチャである．複数の Inception モジュールをアーキテクチャに組み込むことで，パラメータの数が多くなる．この問題を解決するためのアーキテクチャが図 5 である．

4.2.2　残差ブロック

残差ブロックは 2015 年に発表された ResNet に導入されている．GoogLeNet が 22 層であるのに対し，ResNet は 152 層からなっている CNN である．残差ブロックの導入によって層の数を増加させることを可能とし，層の増加によって発生する勾配消失問題を解決した．残差ブロックのアーキテクチャを図 6 に示す．

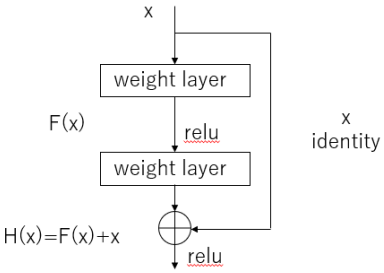

図 6　残差ブロックのアーキテクチャ [8] の図を再作図

残差ブロックは畳み込み層とスキップ接続からなる．いくつかの畳み込み層とスキップ接続を 1 つのブロックとしている．残差ブロックを経た出力は以下の計算式 (1) で表される．

$$H(x) = F(x) + x \tag{1}$$

F(x) は複数の畳み込み層を通して出力された結果である．x はスキップ接続を経た数値である．残差ブロックの最後に F(x) と x を加算する．

畳み込み層の活性化関数に ReLU 関数を使用する．ReLU 関数は勾配消失問題の解決に影響を与えた関数であり，現在，一般的に使用されている関数である．

5　実験結果

加工を施したデータセットを，上記のモジュールを持つ GoogLeNet と ResNet に入力する．

VGG16，GoogLeNet，ResNet の層の数はそれぞれ 16 層，22 層，152 層である．ハイパーパラメータとしてバッ

チサイズを 32，学習回数を 80 回，学習率を 0.0001 で統一している．入力サイズはそれぞれのニューラルネットワークで決まっているサイズで入力する．それぞれ，(224, 224)，(299，299)，(224，224) である．

上記の条件で学習を行った結果を図 7，図 8 に示す．青色の線がトレーニングセットに対する結果を示し，橙色の線がテストセットに対する結果を示す．左図は正答率を示し，右図は損失率である．

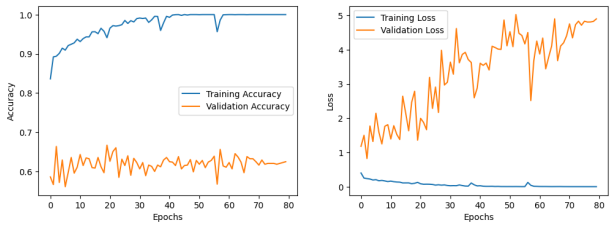

図 7　GoogLeNet の結果

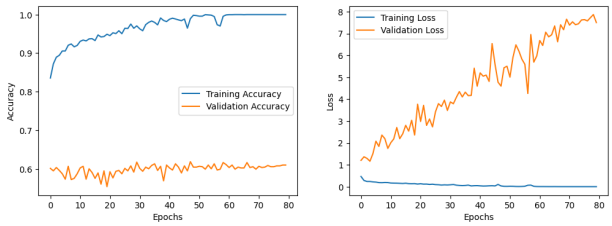

図 8　ResNet の結果

VGG16 での実験では，テストセットに対する正答率は 62% 程に収束している．テストセットに対する損失率は右肩上がりとなっている．80 回の学習を終えた後では 8 近くの値を出している．

GoogLeNet の実験では，テストセットに対する正答率は変動が激しく，58% から 63% までの変化がある．損失率は 80 回の学習を終えた後では 5 近くの値が出ている．3 つの学習モデルの中で最も損失率の値が低くなっている．

ResNet での実験では，テストセットに対する正答率は変動が少なく，61% 程になっている．損失率は，他と同様にすぐに 0% に収束するが，テストセットに対する損失は右肩上がりとなっている．学習終了時の値は 8 である．

6　考察

本研究では，先行研究で課題であったデータセット不足の解決し，入力する CNN の構造を変化させた．一般的に，層の数を増やすことで画像の細かい特徴や複数の特徴を抽出できる．層を増やすことで過学習が抑制できているので，CNN の構造は適切であると考える．

加工したデータを CNN に入力して結果は，60% 程度の正答率であった．先行研究において，加工する前のデータを VGG16 に入力したときの正答率は 30% であった．正答率を比較すると本研究のアプローチは妥当性があると考える．

以上より，作成したデータセットが適切にラベル付けがされていないと考える．データの選別の基準として経験年数を用いているが，データを取った人数が少なかったので，外れ値のデータが多くなったと考えられる．

熟練度による分類を行うための今後の課題は以下のとおりである．

- データセットの適切なラベル付け
- 元データの外れ値の削除

画像分類においてこの課題を解決することで，熟練度分類の実現が可能であると考える．

7　おわりに

本研究では，分類精度を高めるために，Openpose を用いてデータの前処理・拡張を行った．加工したデータを卒業論文時に用いた CNN と別のアーキテクチャを持つ GoogLeNet や ResNet に入力し，実験を行った．今後の課題を解決することで，熟練度分類の精度の向上を目指す．

参考文献

[1] 小関翔也, 野呂昌満, “ニューラルネットワークによる身体動作分析に関する研究,” 南山大学卒業論文要旨集，南山大学, 2022.

[2] 青木義満，“スポーツ競技を対象とした画像 AI 技術,” 人工知能，34 巻 4 号 p. 497-502, 2019.

[3] 金子和樹，中村拓真，矢入郁子，平田均，“OpenPose を用いたサッカー熟練度の分類,” 2020.

[4] 渥美亮祐，金井秀明，“深層学習を用いたスキーシミュレータ利用時における 初心者と熟練者の運動特徴抽出手法,” Vol.2020-GN-110 No.15, 2020.

[5] Libin Jiao Rongfang Bie, “Golf swing classification with multiple deep convolutional neural networks,” International Journal of Distributed Sensor Networks 2018, Vol. 14, 2018.

[6] Simonyan K and Zisserman A. “Very deep convolutional networks for large-scale image recognition, ” Chornell University archive, arXiv:14091556, 2014

[7] Szegedy C, Liu W, Jia Y, et al. “Going deeper with convolutions,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1-9

[8] He K, Zhang X, Ren S, et al. “Deep residual learning for image recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 770-778

[9] CAO, Zhe, et al. “realtime multi-person 2D pose estimation using Part Affinity Fields,” Chornell University archive, arXiv:1812.08008, 2018.