

1. はじめに

スポーツ選手は、スキルレベルにより視線の動きに差があることが知られている [1]。そのため、プロ選手の視線情報を活用したアマチュア選手の育成が注目されている。伊佐らはサッカーのプロ選手の視線データを収集し、任意のシーンにおけるプロ選手の視線を推定する手法を提案した [2]。しかし、RGB 画像から視線を予測するため、選手やボールとの関係が十分に活用されていない。そこで、本研究では RGB 画像に加え、選手とボールの位置、選手の姿勢をシーンのコンテキスト情報として入力することで、視線推定精度の向上を図る。このとき、どのような詳細情報の組み合わせが適切か定量的、定性的に分析する。

2. Dilated Inception Network

本研究では、Dilated Inception Network(DINet)[3]を用いて視線推定を行う。図 1 に DINet のモデルアーキテクチャを示す。DINet は、特徴抽出器に拡張畳み込み処理を用いた Dilated Residual Network(DRN), Dilated Inception Module(DIM)を採用することで、高い精度を保ったまま計算量を減少させたモデルである。

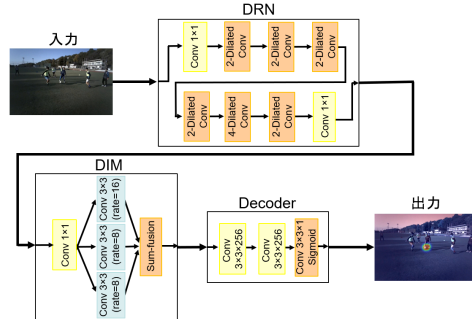


図 1: DINet のモデルアーキテクチャ

3. シーンのコンテキスト情報の作成

RGB 画像のみでは獲得しにくい情報を補完するため、選手とボールの位置をバイナリマスク (位置データ)、選手の姿勢を骨格画像として作成し、入力データに追加する。

3.1. 作成方法

サッカー中継映像データで学習した YOLOv7 を用いて選手及びボールの位置を検出する。そして、検出した物体領域内を1、背景を0としたバイナリマスクを作成する。また、各選手の骨格を Keypoint Detection で検出して骨格データを作成する。

3.2. シーンのコンテキスト情報の処理方法

位置データおよび骨格データをそれぞれの特徴抽出器に入力し、特徴量を連結した合成特徴量をデコーダに入力する。図 2 に変更後の DINet のモデルアーキテクチャを示す。

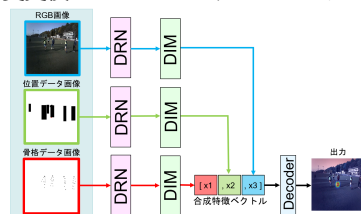


図 2: 変更後の DINet のモデルアーキテクチャ

4. 評価実験

シーンのコンテキスト情報の組み合わせによる精度変化を比較し、本手法の有効性を検証する。

4.1. 実験条件

最適化手法を Adam, 学習率を 0.0001, エポック数を 10, バッチサイズを 8 として DINet を学習した。評価指標は The Linear Correlation Coefficient (CC), Normalized Scanpath Saliency (NSS) を用いる。また、先行研究 [2]

と同様にボールのパスとキープの練習シーンから構成する Ball Position Training(BPT) のフレーム画像を学習データに用いる。

4.2. 定量的評価

定量的評価結果を表 1 に示す。これより、選手とボールの位置、選手の姿勢情報を用いることで従来手法より精度が向上することが確認できる。また、NSS の値は位置データのみを用いた場合が最も高く、CC の値は位置データのみを用いた場合が最も高い。NSS は注視点の位置を評価できるため、位置データはボールや選手の正確な位置を獲得することに寄与していると考えられる。また、CC は Ground Truth と推定結果を確率分布と見なした際の相関を評価できるため、骨格データはシーン状況と視線の関係性の獲得に寄与していると考えられる。しかし、位置・骨格データを併用した場合、いずれかのデータのみを追加した場合と比べ精度向上がわずかである。これは、位置データと骨格データの認識精度の差異から、同時に処理した際に不適合が発生したためだと考えられる。

表 1: コンテキスト情報を追加した時の精度

位置データ	骨格データ	CC ↑	NSS ↑
		0.378	5.048
✓		0.408	5.537
	✓	0.409	5.445
✓	✓	0.401	5.383

4.3. 定性的評価

図 3 に視線データの正解画像と、それぞれの条件での視線推定結果を示す。位置データのみを用いた場合、図 3(a)(b)(c) で推定結果が正解に近いことが確認できる。また、骨格データのみを用いた場合、図 3(a)(b)(d) で推定結果が正解に近いことが確認できる。一方で、全データを用いた場合、図 3(a) および (c) では正解と同様にボールを注視しているが、図 3(b) および (d) では、正解とは異なり、手前側の選手に注目していることが確認できる。これは、全ての選手に対して Keypoint が設定していないため、誤った入力になったことが原因と考える。

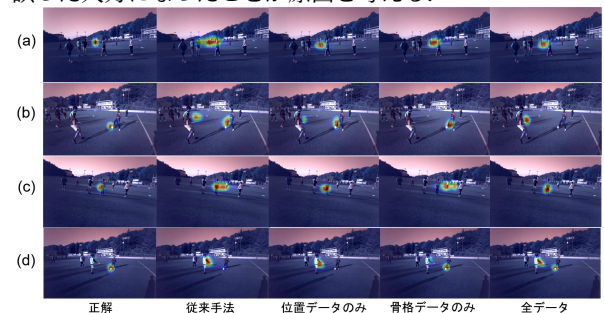


図 3: 定性的評価

5. おわりに

本研究では、RGB 画像に加え、選手とボールの位置、選手の姿勢をシーンのコンテキスト情報として用いる視線推定手法を提案した。評価実験より、位置データと骨格データは視線推定に有効であることが確認できた。今後は更なる精度向上に向け、シーンのコンテキスト情報の活用方法を考える。

参考文献

- [1] A.Phantak, *et al.*, “Keep your head up—Correlation between visual exploration frequency, passing percentage and turnover rate in elite football midfielders Sports”, MDPI Sports, 2019.
- [2] 伊佐 稜, *et al.*, “サッカーのトレーニングにおける中盤選手の注視点比較と視線推定”, メディア工学研究会, 2022.
- [3] Sheng Yang, *et al.*, “A dilated inception network for visual saliency prediction”, TMM, 2019.