

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 —プロサッカー選手と大学生サッカー選手を対象として—

伊佐 稜 堀尾 郷介 平川 翼 山下 隆義 藤吉 弘亘

中部大学 〒487-8501 愛知県春日井市松本町 1200

E-mail: isaryo@mprg.cs.chubu.ac.jp

あらまし スポーツにおける視線行動は広く注目されている。本研究はサッカーにおけるトレーニング時の視線行動をプロサッカー選手と大学生サッカー選手で定量的に分析し比較すること、およびプロサッカー選手の視線を育成用に視線推定することを目的とする。視線の測定は Tobii Glasses 3 を使用した。調査対象はボールポゼッショントレーニング（以下 BPT）と、Small-sided games（以下、SSG）とした。測定された視線は発表者が目視で確認し、各トレーニングにおけるスキャン回数とスキャン中に何を注視しているかを攻撃と守備時に分けて分析した。また、視線推定は DINet を用いて行った。結果、BPT ではプロサッカー選手が大学生サッカー選手と比較して 2.82 倍スキャン回数が多く、敵により注視していた。また、SSG ではプロサッカー選手の方が大学生サッカー選手と比較して、攻撃時は 1.86 倍スキャン回数が多く、守備時は 2 倍スキャン回数が多かった。また、視線推定では攻守でデータを分けることで精度が向上した。

キーワード Deep Learning, Gaze Prediction, Sports, Data Analysis

Comparison of gazing points and gaze prediction of midfielders in soccer training —professional and college student soccer players—

Ryo ISA Kyosuke HORIO Tsubasa HIRAKAWA Takayoshi YAMASHITA
and Hironobu Fujiyoshi

Chubu University 1200 Matsumotocho, Kasugai, Aichi, 487-8501 Japan

E-mail: isaryo@mprg.cs.chubu.ac.jp

Abstract Gaze behavior in sports has received widespread attention. The purpose is to quantitatively analyze and compare the eye gaze behavior of professional soccer players and university students during training in soccer, and to prediction the gaze of professional soccer players for training. As a result, professional soccer players were scanned more often than college soccer players. In addition, the accuracy of gaze prediction was improved by separating the data for attackers and defenders.

Keyword Deep Learning, Gaze Prediction, Sports, Data Analysis

1. まえがき

スポーツにおける視線行動は広く注目されている。とりわけ、サッカーでは視線を用いた分析、研究が多く行われている。中でもスキャンと呼ばれる周囲を確認するために頭部を振る動作が重要視されている。

Phantak らの研究で、サッカー選手権ユーロ 2016 で、中盤選手(セントラルミッドフィルダー)においてスキャン回数が多い選手はパス成功率が高いことがわかっている[1]。また、Aksum らの研究で、エリートサッカー選手の 11 対 11 の試合において、スキャン中のビデオフレーム内にチームメイトより対戦相手が多く映っていることがわかっている[2]。これらの先行研究から、中盤選手のスキャン回数とスキャン

中の注視点は、技術を表す指標の 1 つとして比較される対象であると考えられる。

本研究の目的は 2 つある。1 つ目は、トレーニング中のプロサッカー選手と大学生サッカー選手の視線情報を定量的に比較することである。スキャン回数や注視点を指標として比較することで、プロサッカー選手の方がどのような点で優れているかを評価する。2 つ目は、プロサッカー選手の視線情報を AI モデルを通して模倣させ、大学生サッカー選手の育成に用いることである。大学生サッカー選手視点の画像を用いて AI モデルによりプロサッカー選手の視線を推定することで、より優れた選手はプレー中にどこを注視しているか視覚的にわかりやすくすることを目的とする。

表 2：各トレーニングにおける測定結果

		回数(回)	注視時間(sec)	注視位置(%)						
				Teammate	Opponent	Space	T/O	T/S	O/S	T/O/S
BPT	Pro	117	0.187	35.1	15.8	30.7	12.3	4.4	1.8	0
	Uni1	52	0.16	44.2	15.4	30.8	0	7.7	1.9	0
	Uni2	31	0.162	29.0	16.2	32.3	9.7	6.5	6.5	0
SSG	Att Pro	93	0.353	21.7	24.6	14.5	23.2	5.8	2.9	7.2
	Def Pro			25.0	37.5	25.0	4.2	4.2	0	4.2
	Att Uni	49	0.216	21.6	13.5	37.8	16.2	2.7	8.1	0
	Def Uni1			25.0	33.3	16.7	0	8.3	16.7	0



図 1：ビデオフレームの注視位置表示

表 1：各トレーニングの条件

	BPT	SSG
人数	7(攻)vs4(守)	11vs11
範囲	14m×14m	52.5m×40m
時間	10min	6min
条件	2 タッチ以下	フリータッチ

2. 測定概要

測定対象者は中盤選手のプロサッカー選手(18 歳/165cm/60kg)(以下 Pro)と大学生サッカー選手 2 名(19 歳/164cm/58kg)(以下 Uni1), (20 歳/179cm/65kg)(以下 Uni2)である。測定機器は Tobii Glasses 3 用いる。バッテリーは腰の位置にランニングベルトを使用して固定する。測定項目は、ボールポジショントレーニング(BPT)と Small-sided games(SSG)の 2 種類である。ここで、SSG は攻撃フェーズ(Att)と守備フェーズ(Def)に分けて分析を行った。それぞれのトレーニングの条件を表 1 に示す。また、Uni2 は BPT のみを測定した。測定変数は、スキヤニング回数とスキヤニング時の平均時間、スキヤニング時の注視位置とする。また、スキヤニング時の注視位置はフレーム内のボールがなくなったタイミングからボールが映るフレームまでとし、100msec 以上の時にフレームを目視で確認し、人手でクラスラベリングを行う。クラスは、Teammate(味方)、Opponent(敵)、Space(スペース)、T/O(味方と敵)、T/S(味方とスペース)、O/S(敵とスペース)、T/O/S(味方と敵とスペース)である。注視位置の例を図 1 に示す。

3. 測定結果

プロサッカー選手と大学生サッカー選手の各トレーニングにおける測定変数の結果を表 2 に示す。

表 2 より、BPT を見るとスキヤニング回数は Pro が Uni1, Uni2 と比べ最も高いことがわかる。先行研究[1]から、スキヤニング回数はパス成功率に比例することがわかっているため、大学生サッカー選手と比べプロサッカー選手の方が優れた中盤選手であると言える。また、スキヤニングの注視時間は各選手で大きな差はなかった。Pro の注視位置は、Uni1, Uni2 と比較し T/O の割合が高いことがわかる。BPT はボールを守備に取られないように味方同士でパスを回すトレーニングであるため、T/O の割合が高いことは正しいと考えられる。また、O/S は Uni2 が最も高くなったが、これは守備に対して逃げるような選択をしており、味方にパスを回すというトレーニングの趣旨から外れているため、プロサッカー選手の方が優れていると言える。

次に SSG を見ると、スキヤニング回数は Pro が Uni1 の倍ほどの回数を行っていることがわかる。SSG は BPT に比べ、より実際の試合に近い形式のトレーニングである。そのため先行研究[1]より、実際の試合を模したトレーニングにおいてもプロサッカー選手の方が大学生サッカー選手より優れていることがわかる。スキヤニングの注視時間は Pro の方が Uni1 よりも長いことがわかる。ここから、オフボール時に周りをよく見る時間が確保できていると考えられる。スキヤニング中の注視位置を見ると、各フェーズにおいて Opponent の割合が高いことがわかる。先行研究[2]から、大学生サッカー選手よりもプロサッカー選手の方が優れた選手であるという。また、Pro の T/O/S が攻撃フェーズは 7.2%、守備フェーズは 4.2%と一度のスキヤニングで複数の対象に注視していることがわかる。これらの結果から、スキヤニングの注視時間と合わせて、Pro は一回のスキヤニングにおける情報の処理能力が高いことがわかる。

BPT と SSG の結果を踏まえて、トレーニング中の注視点大学生サッカー選手よりもプロサッカー選手の方が優れていることがわかった。

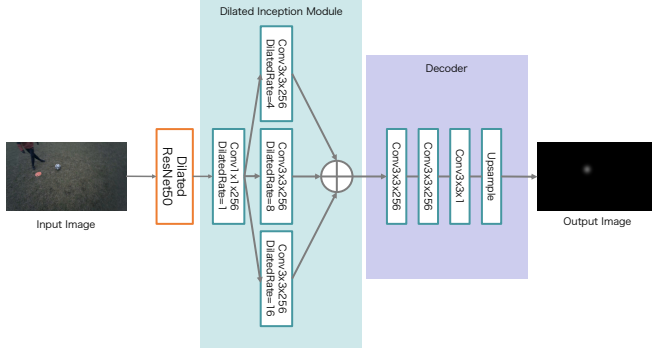


図 2 : DINet のモデル図

4. 視線推定

プロサッカー選手の各トレーニングにおける視線情報を大学生サッカー選手の育成に用いるため、AIモデルによるプロサッカー選手の視線推定を行う。

5. Dilated Inception Network

視線推定には Dilated Inception Network(DINet)[3]を用いる。DINet は Dilated Residual Network(DRN)[4]で特徴マップの解像度を落とさず Dilated Inception Module でマルチスケールな特徴を補足し、顕著性マップを出力する。DINet の概略図を図 2 に示す。

5.1. Dilated Residual Network

Dilated Residual Network(DRN)は、顕著性予測に使用される学習済みのネットワークの一つである ResNet-50 の拡張版である。ResNet-50 は大きく 5 つの畳み込みブロックがあり、出力される特徴マップは 7×7 である。DRN では 5 つの畳み込みブロックの内 2 つのダウンサンプリングを削除し、ブロック内の畳み込みをそれぞれ拡張率 2 と 4 の拡張畳み込みに置き換えることで出力される特徴マップを 28×28 まで大きくしている。

5.2. Dilated Inception Module

GoogleNet[5]で提案された Inception Module に拡張畳み込みを導入したモジュールである。拡張畳み込みとは、各空間の次元におけるフィルタ w 内で、連続する二つの空間の地点間に $r-1$ 個の 0 を挿入することで空間分解能を維持しつつ、プーリングをせずとも小さなフィルタで長距離の畳み込みができ、受容野の縮小を防ぐことが出来る。通常、畳み込みカーネルが $k \times k$ であると仮定すると、拡張畳み込みは $k_d \times k_d$ 、

$k_d = k + (k-1) \cdot (r-1)$ と表すことができる。

Inception Module に拡張畳み込みを導入することで 7×7 , 5×5 の畳み込みを、膨張率 3 と 2 の 3×3 の拡張畳み込みに置き換えることが可能である。この置換により、少ないパラメータ数、短い計算時間で Inception モジュールと同じまたはそれ以上の結果を出すことが出来る。

表 3 : 各トレーニングにおける定量的評価

	BPT	SSG	Att SSG	Def SSG
CC	0.41	0.4	0.49	0.51

5.3. Decoder

Dilated Inception Module で得た特徴マップを入力として、3 層の畳み込みを行い、最後にバイリニア法によるアップサンプリングを行う。

5.4. 損失関数

顕著マップは画像全体における人間の確率分布とみなすことができる。予測された顕著性マップと真値を確立分布に変換する場合、これらの値を正規化する必要があり、線形正規化を利用する。この線形正規化ベースの損失関数を以下の式 1 に示す。

$$L(p, g) = \sum_i |p_i - g_i|, \quad (1)$$

$$p_i = \frac{x_i^p}{\sum_{i=1}^N x_i^p}, \quad g_i = \frac{x_i^g}{\sum_{i=1}^N x_i^g}$$

ここで、 p は予測値、 g は真値、 x は非正規化された予測された顕著性マップの値と顕著性マップの真値のいずれかの顕著性を示す値である。

6. 評価実験

プロサッカー選手の視線を用いて模倣することを検証する。また比較実験として、SSG は攻撃フェーズと守備フェーズで分けた動画フレームを用意して、攻撃フェーズと守備フェーズに特化させた模倣を行う。

6.1. 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 10 とする。使用するデータセットはプロサッカー選手の各トレーニングの動画フレームとする。BPT は学習に 8,884 枚/評価に 1,121 枚、SSG は学習に 6,944 枚/評価に 1,029 枚用いる。

SSG の攻撃フェーズは学習に 3,718 枚/評価に 533 枚、守備フェーズは学習に 3,226 枚/評価に 496 枚用いる。

評価指標は The Linear Correlation Coefficient (CC) を用いる。CC は顕著性マップと真値の密度マップの線形相関である。CC を以下の式 2 に示す。

$$CC = \frac{cov(P, G)}{\sigma(P) \times \sigma(G)}$$

ここで、 P は予測した顕著性マップで、 G は真値の密度マップ、 $cov(\cdot)$ は共分散、 σ は標準偏差を表す。

6.2. 定量的評価

各トレーニングにおける定量的評価を表 3 に示す。表 3 を見ると SSG において、攻撃フェーズと守備フェーズに分けて学習を行う方が 0.1 ほど精度が良いことが確認できる。これは、3 章で行った測定結果分析の、プロサッカー選手のスキヤニング中の注視位置の割合が攻撃フェーズでは味方と敵を 1 回のスキヤニングで複合的に見ることが多かった。

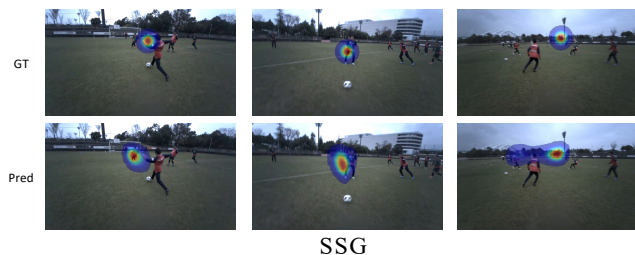


図 3：定性的評価

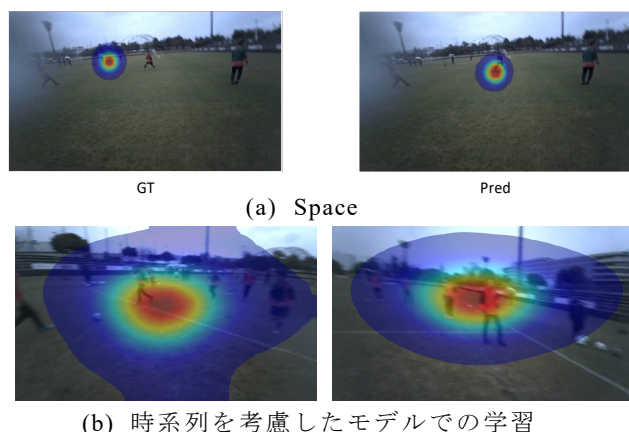


図 4：失敗例

それに対し、守備フェーズでは敵単体で見るものが多く割合が異なる。このことから、DINetで視線の注視領域等を捉えられたため、精度が向上したと考えられる。

6.3. 定性的評価

図 3 に定性的評価を示す。BPT を見ると、ボールへの注視は適切に行えていることがわかる。一方、敵の足元を注視する際に右足と左足で予測値がずれていることがわかる。これは同じ敵であるが、プロサッカー選手は相手の利き足などの情報からパス、ドリブル等の正しい選択をするための注視を行うため、BPT のような狭いスペースの激しいトレーニングに対しては視線推定は難しいと考える。一方、SSG を見ると、ボール保持者への注視、パスを出した先の味方への注視等が再現できていることがわかる。また、パスを出した味方に対しては右足を真値は見ているのに対して、予測結果も右足を見ている。これは、SSG の性質と中盤選手であることから、BPT と比較して人が密集しない。そのため、学習データ内のフレームに映る選手の割合が少ない。したがって、足に対する視線の学習がしやすく、右足を注視できたと考えられる。このことから、実際のゲームを模倣した SSG では、DINet を用いることでプロサッカー選手の視線を再現できていると考える。

6.4. 失敗例

図 4 に定性的評価の失敗例を示す。(a)を見ると、BPT におけるスキニング中において、真値は敵と味方の間を見ているのに対して、予測値では味方を見ていることがわかる。これは、特徴抽出を行う際、人以外特徴的なオブジェクトが少ないフレームにおいて、人に注視が寄ってしまうことが原因であると考えられる。また、学習枚数が 1 万枚を超えない動画像であるため、時系列の考慮が難しい問題も発生する。(b)を見ると、時系列を考慮したモデルで学習を行った際に、フレーム全体に緩やかに顕著性マップがかかっており、ほとんどフレームの中心点から動かない。

7. まとめ

今回は、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を行った。結果として、プロサッカー選手の方が大学生サッカー選手より優れていることが定量的に判明した。また、視線推定は DINet を用いることでプロサッカー選手の視線を模倣することができた。今後は、別のネットワークや新しいアプローチから時系列情報を考慮できるようにしていく。

文 献

- [1] A. Phantak and M. Gruber : “Keep your head up—Correlation between visual exploration frequency, passing percentage and turnover rate in elite football midfielders Sports”, MDPI Sports, 7(6), 139, (June 2019)
- [2] KM. Aksum, L. Magnaguagno, CT. Bjorndal and G. Jordet : “What do football players look at? An eye-tracking analysis of the visual fixations of players in 11 v 11 elite football match play”, Front Psychol, 11, pp.1-11(October 2020)
- [3] S. Yang, G. Lin, Q. Jiang and W. Lin : “A Dilated Inception Network for Visual Saliency Prediction”, TMM, 22, 8, pp.2163-2176(2019)
- [4] N. Liu and J. Ham, “A deep spatial contextual long-term recurrent convolutional network for saliency detection”, arXiv preprint arXiv:1610.01708(2016)
- [5] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. : “Going deeper with convolutions”, CVPR, pp.1-9(June 2015)