TS-TranSalNet によるサッカー選手の視線推定の高精度化

ER21082 村上和樹　　指導教授：藤吉弘亘

1. はじめに

スポーツ選手のプレイスキル向上を目的とし，視線や位置情報のトラッキングデータを用いた分析や研究が注目されている．サッカーにおいては視野の広さがパフォーマンスに深く関与する．先行研究 [1] では，プロサッカー選手の視線を再現する視線推定モデルを提案し，大学生サッカー選手との視線の比較を行っている．プロ選手の視線を再現するには，より高精度な視線推定モデルが必要である．そこで本研究では，時系列の考慮した高度な視線推定モデルの構築を目標とする．

2. 視線推定タスクにおける代表的なモデル

視線推定は，入力画像に対して一人称視点における注視領域を出力するタスクである．視線推定モデルの多くは，入力画像から特徴抽出を行うエンコーダと，特徴から視線推定を行うデコーダで構成される．

視線推定の代表的な手法である DINet[2] は，拡張畳み込みを用いて画像全体の特徴を捉えて視線を推定する．また，CNN と Transformer を組み合わせた TranSalNet[3] は，画像全体の特徴と局所的な特徴を組み合わせて高精度な視線推定をする．Temporal Shift Dilated Inception Network (TS-DINet) は Temporal Shift Module (TSM)[4] と DINet を組み合わせた手法である．TSM は，特徴量の一部を前後のフレームにシフトし，次の畳み込み層へ伝達することで，時系列特徴を保持することができる．

3. 提案手法

本研究では，視線推定の高精度化を目指し，時系列情報を考慮した視線推定モデルを提案する．提案モデルは，TranSalNet を TS-TranSalNet のベースとする．TranSalNet の CNN エンコーダの 3 層の出力に TSM を導入する．これにより，時系列情報が統合され考慮可能となる．TS-DINet と違い，TSM で前後のフレーム特徴を付与した特徴マップを Transformer エンコーダに入力する．本手法の概略図を図 1 に示す．

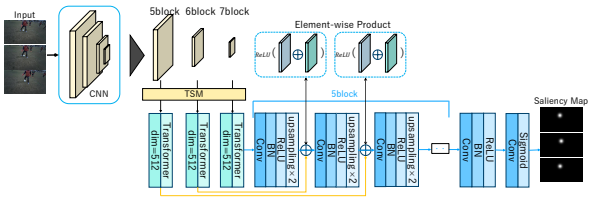

図 1：TS-TranSalNet のモデル構造

4. データセットの構築

本研究では，視線計測機器である Tobii Glasses3 で計測したサッカー選手の視線データを学習および評価用のデータからなるデータセットを構築する．収集したデータは，Ball Posession Training (BPT) と Small Sided Games (SSG) の 2 種類のトレーニングデータで構成される．BPT は，ボールを相手に取られないようにパスを回すトレーニングである．SSG は，狭いコートで少人数による実際のサッカーの試合形式で行うトレーニングである．表 1 に収集したデータ等の条件を示す．

表 1：データセットの概要

トレーニング名	BPT	SSG
人数	7（攻）vs 4（守）	11 vs 11
範囲	14 × 14 [m]	52.5 × 40 [m]
時間	10 [min]	6 [min]
画像サイズ	1920 × 1080	
学習用データ数	9120 枚	5976 枚
評価用データ数	530 枚	692 枚

5. 評価実験

前述したデータセットを用いて評価実験を行う．従来手法である DINet，TS-DINet，TranSalNet をベースライン手法とし，提案手法である TS-TranSalNet と比較することでその有効性を示す．

本研究では，先行研究と同様に BPT，SSG のそれぞれで個別に学習，評価を行う．学習時のエポック数は 10，最適化手法に AdamW，損失関数に The Linear Correlation Coefficient (CC)，Kuller Leibler Divergence (KLdiv)，Total Validation Distance (TVDist) を用いる．損失関数以外のその他の条件は提案手法と同じである．評価指標は，CC，Normalized Scanpath Saliency (NSS)，Mean Absolute Error（MAE）を用いる．

5.1. 定量的評価

各モデルの精度を表 2 に示す．BPT では CC の評価指標で提案手法が最も高く，SSG では全体的に提案手法の精度が最も高い．一方で，BPT と SSG の両方で MAE の評価指標において，提案手法は TS-DINet を下回った．これは，MAE は予測値と真値の間の誤差の絶対値を計算する指標であり，視線マップの形状や分布の適切さは評価しない．そのため，視線マップの最大値のずれが，誤差を大きくする要因と考えられる．

表 2：各モデルにおける視線推定の精度比較

モデル	TSM	BPT			SSG		
		CC ↑	NSS ↑	MAE ↓	CC ↑	NSS ↑	MAE ↓
DINet		0.518	4.924	0.017	0.390	3.512	0.021
TS-DINet	✓	0.530	4.801	**0.012**	0.398	3.542	**0.014**
TranSalNet		0.560	**5.410**	0.013	0.442	4.039	0.020
TS-TranSalNet	✓	**0.571**	5.309	0.015	**0.455**	**4.088**	0.023

5.2. 定性的評価

BPT の視線推定結果を図 2，SSG の視線推定結果を図 3 に示す．BPT はスキャニング中の画像のため，従来手法による視線推定はオブジェクトにほとんど当たっていない．一方で，提案手法は GT に近い視線推定結果である．また，SSG の相手選手のボールを取ろうとするシーンにおいて，従来手法はボールの付近を推定しているが，提案手法は BPT と同様に GT に近い位置を推定している．

図 2：BPT における各モデルの推定結果

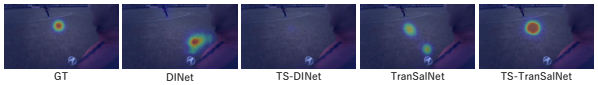

図 3：SSG における各モデルの推定結果

6. おわりに

本研究ではサッカー選手のより高精度な視線推定を行うため，時系列を考慮した視線推定手法を提案した．評価実験より，提案手法は従来手法よりも高い精度で視線推定が行えることを確認した．今後はデータセットの規模を拡大して検証を行う．

参考文献

[1] 伊佐，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.

[2] S.Yang，et al.“A Dilated Inception Network for Visual Saliency Prediction”，IEEE Transactions on Multimedia，2019.

[3] J.Lou，et al.“TranSalNet: Towards perceptually relevant visual saliency prediction”，Neurocomputing，2021.

[4] J.Lin，et al.“TSM: Temporal Shift Module for Efficient Video Understanding”，ICCV，2019.