H-002

顕著性マップから抽出した関心領域に基づく視線遷移予測
Scanpath Prediction Based on Area-of-Interest Derived from Saliency Map

大野　永遠　　田中　雄一　　東　広志
Towa Ohno　　Yuichi Tanaka　　Hiroshi Higashi

1　はじめに

人は，眼球運動を通じて視覚情報を継続的に収集し，必要な情報を選択的に取得している．そのため，注視対象の選択メカニズムの理解は，ヒトの視覚的な情報処理を解明する上で重要である．

人の視線を直接的に計測する視線追跡技術は，このメカニズムを実証的に解明する手段となる．近年の技術進歩により高精度な視線データの取得が可能となり，その応用は認知科学，医療，教育，広告など多岐に渡る [1]．

人間の視線軌跡は，主に次の 2 種類の構成要素からなることが知られている [1]．

Fixation 特定の領域内にとどまっている注視点の集合：一般にその領域の中心に位置する点を指す．

Saccade Fixation 間を移動する視線遷移．

画像上の Fixation の空間分布は，各画素の注目度を輝度で表現する Saliency map（顕著性マップ）で可視化される．Saliency map は静的な視覚特性のモデル化に相当する．普通，人は Fixation と Saccade を交互に繰り返し，画像内の情報を逐次探索する．その軌跡を予測する Scanpath 予測は，動的な視覚特性のモデル化にあたる．

従来の Scanpath 予測モデルは，過去の注視点履歴から将来の注視点座標を予測する．しかし，人は特定の座標ではなく，座標周辺に含まれる物体に注意を向ける [2]．すなわち，現在の座標を予測単位とする Scanpath 予測は，人の視覚的認知を正しく反映しているとは言い難い．

本研究では，注視対象物体に対応する関心領域（Area of Interest: AOI）を予測単位とする Scanpath 予測モデルを提案する．本モデルは，画像から生成した Saliency map から抽出された AOI と過去の視線履歴を入力とし，AOI の視覚的特徴と視線文脈を統合して，各 AOI への視線遷移確率を予測する．提案モデルは，視線遷移先の領域の予測において従来手法を上回る性能を示した．

2　提案手法

データ準備の方法と，提案モデルについて述べる．

大阪大学大学院工学研究科
Graduate School of Engineering, The University of Osaka

2.1　入力データと前処理

入力画像と過去の注視履歴に対し，次の前処理を行う．

1. **AOI 特定**：Saliency map 上の局所ピーク点を中心に固定サイズ（200×200 pix）の AOI を設定する．画像あたり 3〜10 個の AOI が設定され，対応する注視点の約 90%は設定された領域内に含まれた．
2. **訪問統計量算出**：各 AOI の注視履歴から，累計注視回数と最終注視からの経過時間からなる訪問統計量 $\mathbf{h} \in \mathbb{R}^2$ を算出し，モデルの入力として用いる．

2.2　提案モデル

本項では図 1 に示すような，入力画像と過去の視線履歴から次に注視する AOI を予測するモデルを提案する．本モデルは，Saliency map から抽出した AOI，画像，注視履歴を入力とし，以下の 3 モジュールで構成される．

1. **AOIFeatureExtractor**：各 AOI の視覚特徴量 $\mathbf{f} \in \mathbb{R}^{d_{\text{aoi}}}$ を抽出する．
2. **ScanpathTCN**：過去の注視点系列から文脈情報を集約した注視履歴ベクトル $\mathbf{g} \in \mathbb{R}^{d_{\text{gaze}}}$ を生成する．
3. **NextAOIPredictionHead**：AOI 視覚特徴量 $\mathbf{f}$，注視履歴ベクトル $\mathbf{g}$，および AOI 訪問統計量 $\mathbf{h}$ を統合し，各 AOI への遷移スコアを出力する．

学習時には，注視点と AOI をソフトに対応付ける．具体的には，各 AOI 中心を平均とする二次元ガウス分布を定義し，各注視点位置での確率密度を取得し正規化することで，ターゲットラベルを生成する．損失関数には Kullback-Leibler divergence（KL-Div）を採用する．

3　実験

本節では，実験設定とモデルの性能評価方法および実験結果を述べる．

3.1　実験設定

実験には，700 枚の画像と 15 人の視線データから成る OSIE データセット [3] を使用する．データは 9:1 に分割し，学習と検証に用いる．

Saliency map 生成には DeepGaze II [4] を用いる．

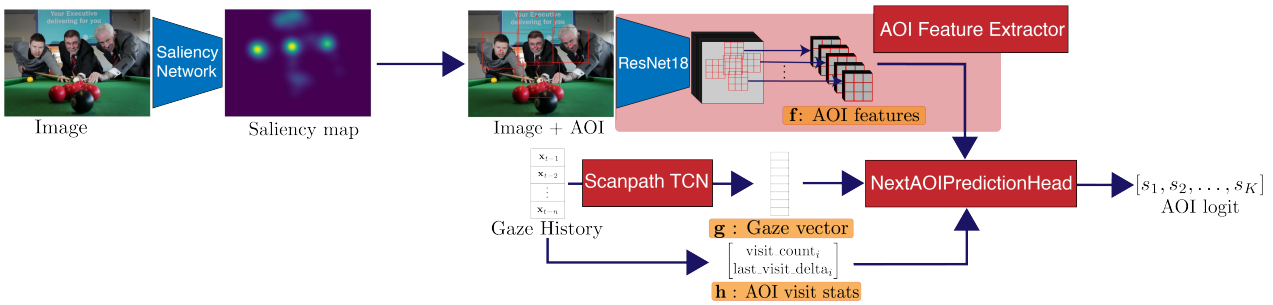

図 1: 提案モデルのアーキテクチャ.

表 1: 注視領域に基づく各モデルの性能評価

モデル名	KL-Div	Top-1	Top-3
提案モデル	**0.0085**	**51.57**	**86.02**
DeepGaze III	0.0099	48.20	84.85
Saliency map	0.0116	31.25	75.03
Centerbias	0.0126	23.26	63.26
Uniform	0.0120	16.50	54.97

表 2: 注視座標に基づく各モデルの性能評価

モデル名	LL ↑	IG ↑	AUC ↑	NSS ↑
提案モデル	1.564	1.306	0.886	2.255
DeepGaze III	**2.551**	**2.293**	**0.924**	**3.469**
Saliency map	1.865	1.607	0.900	2.581
Centerbias	0.258	0.000	0.705	0.635
Uniform	0.000	-0.258	0.500	1.000

3.2 評価方法

提案手法の予測性能を, 先行研究手法および複数のベースラインモデルと比較する．先行研究手法には, Saliency map を出力して次注視点予測を行う DeepGaze III [5] を再学習して用いる．ベースラインモデルは, 全画素均一確率（Uniform）, 画像中心への注視傾向モデル（Centerbias）[6]，被験者の Fixation の空間的分布を可視化した実測 Saliency map の 3 種である．

評価は次注視領域予測と次注視座標予測の 2 観点から異なる指標群を用いて行う．

注視領域に基づく指標 各 AOI への遷移スコア評価には，KL-Div と Top-k 精度（次注視 AOI が予測スコア上位 k 個に含まれる割合）を $k = 1, 3$ の場合で用いる．Saliency map 出力モデルでは，AOI 内の画素値平均を遷移スコアに変換し, 同指標を適用する．

注視座標に基づく指標 Saliency map の評価には，Log-Likelihood（LL），Information Gain（IG），AUC, NSS を用いる [7]．提案モデルでは，AOI 遷移スコアに基づく混合ガウス分布を用いて対数密度マップを生成し，同指標を適用する．

3.3 結果

注視領域に基づく評価 表 1 に示す通り，提案モデルは，KL-Div，Top-1 精度，Top-3 精度のいずれにおいても，他手法を上回った．

注視座標に基づく評価 表 2 に示す通り，提案モデルは全指標で Centerbias および Uniform を上回ったが，DeepGaze III，Saliency map には及ばなかった．

4 結論

DeepGaze III などの従来の視線遷移予測手法は，注視座標を予測単位とする評価において高い性能を示した．一方，予測単位を座標ではなく注視領域とした場合，提案手法は従来法を上回る予測精度を達成した．この結果は，人の視覚特性に合わせた予測，すなわち，注視領域に基づく視線予測の有効性を示すものである．

謝辞

本研究は JSPS 科研費 23H01415, 23K26110, 23K17461, 24K15047 の助成を受けた．

参考文献

[1] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka, and J. Van de Weijer, *Eye Tracking: A Comprehensive Guide to Methods and Measures*. oup Oxford, 2011.

[2] J. M. Henderson and A. Hollingworth, “Chapter 12 - Eye Movements During Scene Viewing: An Overview,” in *Eye Guidance in Reading and Scene Perception* (G. Underwood, ed.), pp. 269–293, Amsterdam: Elsevier Science Ltd, 1998.

[3] J. Xu, M. Jiang, S. Wang, M. S. Kankanhalli, and Q. Zhao, “Predicting human gaze beyond pixels,” *Journal of Vision*, vol. 14, no. 1, pp. 28–28, 2014.

[4] M. Kümmerer, T. S. Wallis, and M. Bethge, “DeepGaze II: Reading fixations from deep features trained on object recognition,” *arXiv preprint arXiv:1610.01563*, 2016.

[5] M. Kümmerer, M. Bethge, and T. S. Wallis, “DeepGaze III: Modeling free-viewing human scanpaths with deep learning,” *Journal of Vision*, vol. 22, no. 5, pp. 7–7, 2022.

[6] B. W. Tatler, “The central fixation bias in scene viewing,” *Journal of Vision*, vol. 7, no. 14, pp. 4–4, 2007.

[7] M. Kümmerer and M. Bethge, “State-of-the-art in human scanpath prediction,” *arXiv preprint arXiv:2102.12239*, 2021.