Adapting XAI methods for multi-target tasks

Addressing challenges and Inter-Keypoint dependencies in Cricket Pose Analysis

Author: Atanas Semov [1]

Responsible Professor and Supervisors: Ujwal Gadiraju, Danning Zhan

[1]Delft University of Technology

Introduction

As AI systems are increasingly deployed in domains such as healthcare, sports analytics, and human-computer interaction, the need to explain their decisions has become more pressing than ever. While existing explainable AI (XAI) methods have shown success in single-output tasks like image classification, many real-world applications involve **multi-target models**—systems that generate multiple, interrelated predictions simultaneously. This research addresses the urgent need for **interpretable, structure-aware XAI techniques** tailored to multi-target models, offering insight into how complex outputs are computed and how keypoint dependencies in pose estimation and cricket can be effectively revealed.

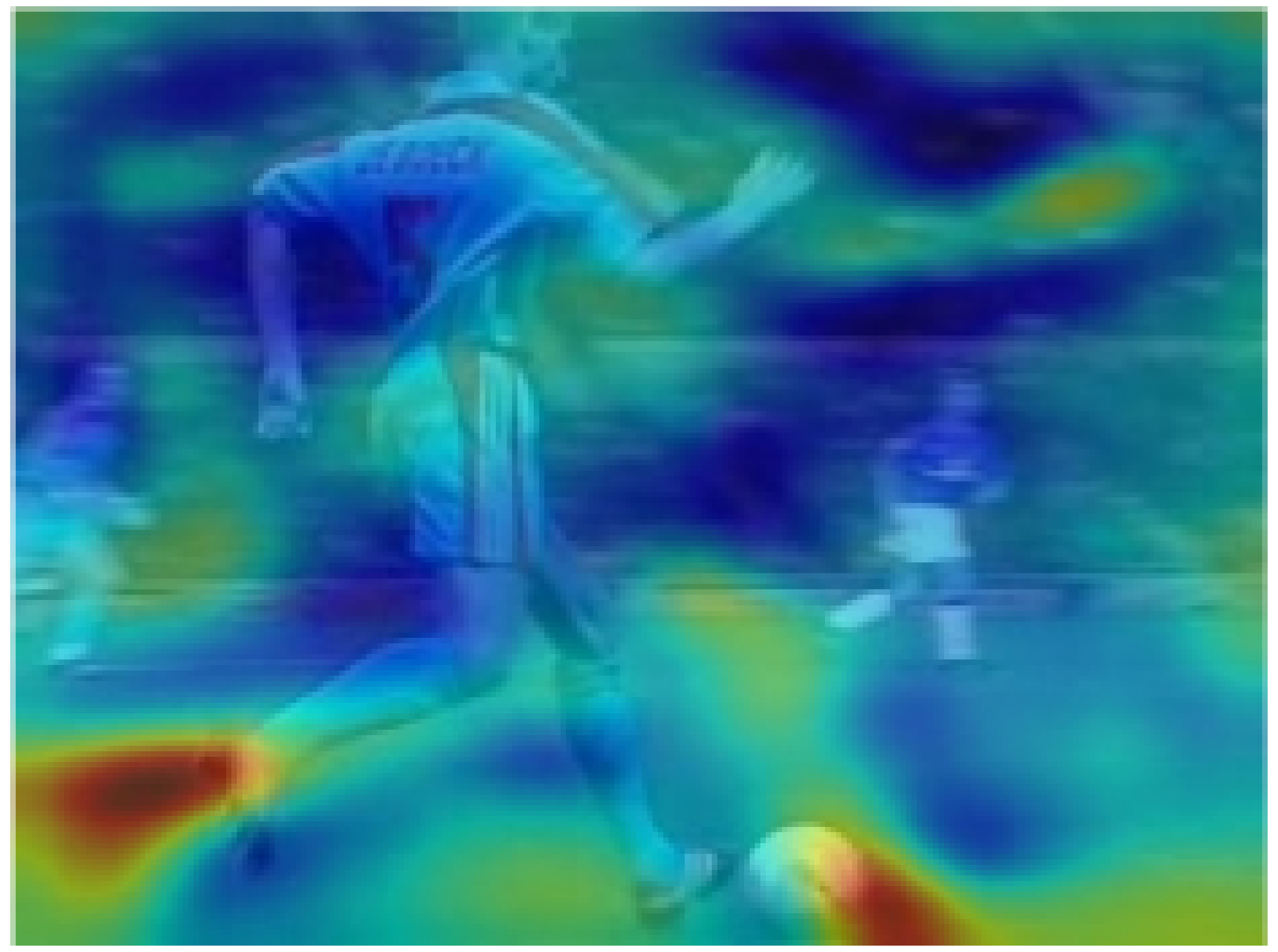

Figure 1. Representation of right ankle seen by a pose estimation model

Related XAI methods

SHAP (SHapley Additive exPlanations) quantifies how each input feature contributes to a model's prediction using cooperative game theory [1].

Grad-CAM (Gradient-weighted Class Activation Mapping) visualizes the important regions of an input image based on gradient information from convolutional layers [2].

Research Question

How can XAI methods be adapted for multi-target tasks like pose estimation?

- *What are the specific challenges in applying XAI methods to multi-target tasks?*
- *How can interdependencies between keypoints be identified and explained?*
- *How can XAI be used to understand the behavior of the model in relation to the estimation of the cricket pose?*

Methodology

Challenges in Multi-Target XAI: Traditional XAI tools like SHAP and Grad-CAM are optimized for single-output tasks. In pose estimation, however, predictions are structured and interdependent (e.g., the wrist's position depends on the shoulder). This makes standard scalar-based explanations insufficient for interpreting human motion.

Modeling Keypoint Dependencies: To understand how pose estimation models capture joint relationships, we analyze pairwise heatmap correlations between keypoints. By quantifying overlap in predicted spatial distributions across samples, we generate a *Keypoint Dependency Matrix*, which visualizes how strongly different joints are connected in the model's internal logic.

Cricket Shot Analysis: We apply a framework to compare the biomechanics of two cricket shots: the pull shot and the cover drive. We track the wrist-to-shoulder motion across time using 2D vectors, and fit polynomial curves to describe wrist trajectories. Additionally, we analyze heatmap activation shifts to assess the model's attention throughout the motion.

XAI in Multi-Target Tasks

Several methods have been proposed to address the challenge of explaining models with multiple, interdependent outputs especially in the healthcare domain:

SHAP extensions such as hierarchical and clustered SHAP enable interpretation in structured prediction tasks by attributing importance to groups of outputs, rather than treating each individually. These adaptations have proven useful in clinical settings where predictions are organized by systems or diagnosis categories [3].

Deep-SHAP for neuroimaging has been used to connect brain region activations with multiple cognitive targets, effectively modeling the influence of a single feature across several outputs in Alzheimer's disease prediction [4].

Grad-CAM in multi-label vision models enables visual comparison of model focus. In retinal disease classification, Grad-CAM has been integrated during training to emphasize lesion regions relevant to each diagnosis, improving both accuracy and interpretability [5].

Keypoint Interdependence Analysis

To understand how the model internally relates different body keypoints, we compute pairwise similarities between the predicted heatmaps produced by the pose estimation model.
Each heatmap reflects the spatial probability distribution of a keypoint's location. Each cell quantifies the average spatial co-activation between two keypoints. Darker regions indicate stronger similarity between the respective heatmaps, suggesting a higher degree of model-level dependency.

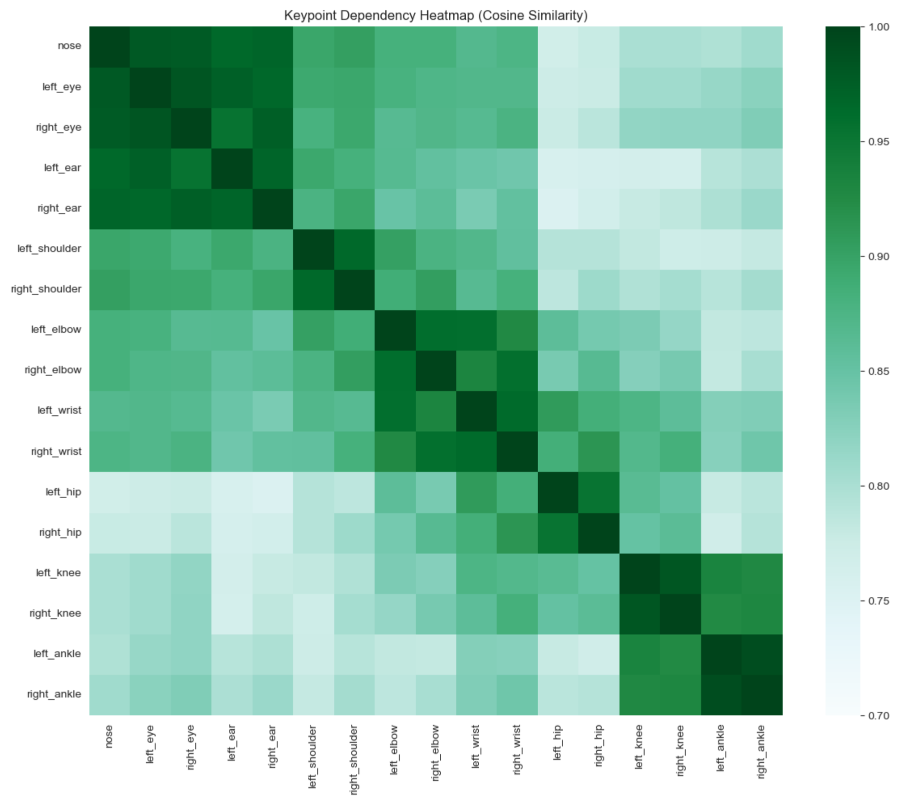

Figure 2. Keypoint Dependency Matrix

Analysis of Keypoint Behavior Across Cricket Shots

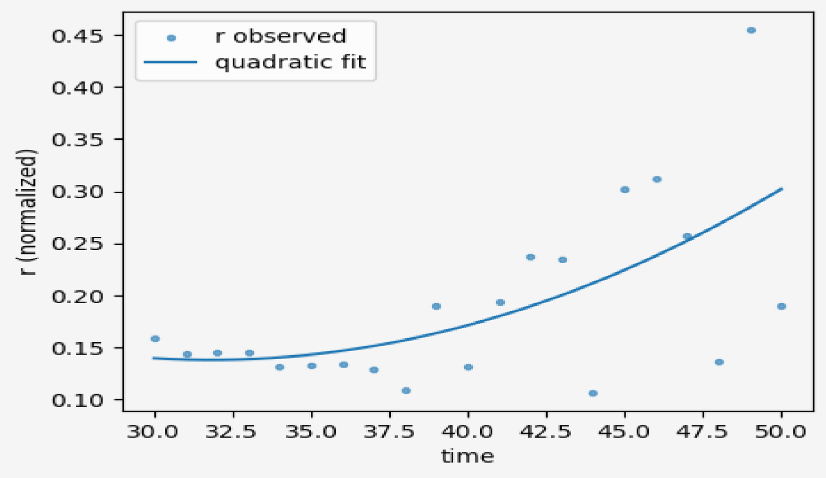

Figure 3. Distance between wrist and shoulder for cover drive

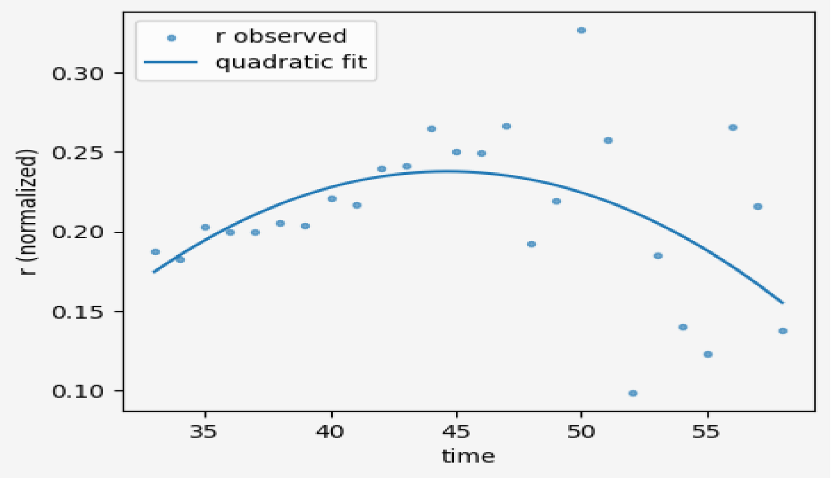

Figure 4. Distance between wrist and shoulder for pull shot

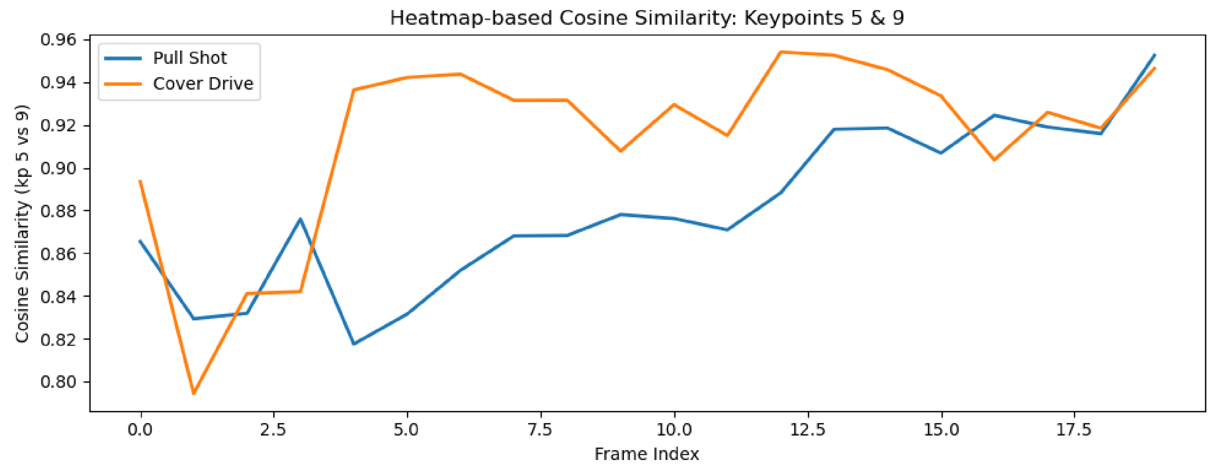

Figure 5. Comparison of the correlations between wrist and shoulder for different shots

Conclusion & Future Work

Contributions: This work introduced a heatmap-based method to reveal inter-keypoint dependencies in pose estimation models, enabling clearer insight into model behavior. A polynomial-based geometric framework was used to compare wrist-shoulder trajectories across cricket shots. Together, these methods combined spatial attention with movement analysis to produce interpretable, task-specific explanations.

Future Work: Future work includes extending the analysis to additional keypoint pairs—such as hips and knees or across bilateral limbs—to uncover broader biomechanical relationships. Furthermore, incorporating temporal modeling could reveal how keypoint dependencies evolve over the course of an action, offering deeper insight into motion dynamics.

References

[1] S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in *Advances in Neural Information Processing Systems*, 2017, pp. 4765–4774.

[2] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, "Grad-cam: Visual explanations from deep networks via gradient-based localization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2017, pp. 618–626.

[3] E. Tjoa and C. Guan, "A survey on explainable artificial intelligence (xai): Toward medical xai," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 32, no. 11, pp. 4793–4813, Nov 2021.

[4] P. Bhattarai, D. S. Thakuri, Y. Nie, and G. B. Chand, "Explainable AI-based Deep-SHAP for mapping the multivariate relationships between regional neuroimaging biomarkers and cognition," *European Journal of Radiology*, vol. 174, p. 111403, 2024.

[5] Z. Li, M. Xu, X. Yang, Y. Han, and J. Wang, "A multi-label detection deep learning model with attention-guided image enhancement for retinal images," *Micromachines*, vol. 14, no. 2, p. 371, 2023.

A.M.Semov@student.tudelft.nl