Player Pose-Driven Handedness Prediction for Ice Hockey

Kevin Wei
University of Waterloo VIP Lab
200 University Ave. West, Waterloo, Ontario Canada
kevin.wei1@uwaterloo.ca

Evan Iaboni
evan.iaboni@uwaterloo.ca

Amir Nazemi
amir.nazemi@uwaterloo.ca

David Clausi
dclausi@uwaterloo.ca

Abstract

We present a pose-driven framework for handedness prediction in ice hockey, designed to support reliable player identification in broadcast video. Building on extended 2-D human-stick pose representations, we introduce a lightweight multilayer perceptron (MLP) that classifies handedness from torso-normalized keypoint coordinates and confidence features. To improve stability over time, we further extend frame-level predictions to tracklet-level inference through a signed-confidence aggregation method that amplifies reliable predictions while suppressing ambiguous ones. Experiments on large-scale pose and handedness datasets demonstrate that our approach achieves accurate pose estimation and handedness classification. The resulting system provides appearance-agnostic discrimination of player handedness, which can be integrated to support identity preservation in hockey player tracking systems.

1. Introduction

Computer vision has become a central tool for sports analytics, where automated systems aim to identify and track individual athletes directly from broadcast videos. Reliable player identity maintenance is critical for higher-level tasks in sports analytics. However, in a visually homogeneous sport like ice hockey, factors such as uniforms, equipment, and fast motions can cause maintaining unique player identities to be a challenge. While end-to-end learning approaches can associate visual appearance or pose directly with player identity, such methods often require large amounts of labeled data and struggle to generalize across games with differing camera setups. Rather than replacing end-to-end models, we seek to complement existing tracking pipelines by providing an appearance-agnostic constraint that helps prevent identity switches between visually similar players.

Within this context, we focus on the task of predicting player handedness. Unlike visual appearance features, which fluctuate due to body orientation and occlusion, handedness provides a discriminative trait to assist in obtaining consistent identity assignments over time.

Accurately detecting handedness in broadcast footage presents several challenges. Sticks can often be obscured from view or blurred by fast motion, due to the lower quality of typical broadcasts. Furthermore, handedness classification is affected by asymmetries in the available data, where left-handed players are more frequent, stick visibility is low, and pose estimation producing lower confidence predictions for stick keypoints compared to torso ones.

Prior work in pose estimation and hockey analysis has largely focused on general 2-D human keypoint detection [2, 3], using the standard 17-point human configuration. Recent hockey-specific systems have introduced additional stick keypoints [9], which augments human pose with explicit stick annotations to improve action understanding.

While prior approaches are effective for modeling body posture, these methods do not explicitly address handedness classification, particularly with the incorporation of stick geometry. Unlike appearance-based cues, handedness is a static player attribute that must remain invariant to viewpoint changes, skating direction, and frequent occlusion in broadcast footage. In particular, torso-normalized coordinates provide a stable reference frame for comparing arm and stick configurations across frames, enabling handedness inference that is robust to camera motion and player rotation. From a tracking perspective, this design allows for handedness to function reliably as an identity-preserving attribute rather than a transient visual signal. The preservation of this attribute can facilitate more accurate player tracking downstream.

To address these gaps and incorporate our idea of emphasizing handedness as a critical identity attribute, we introduce a pose-driven handedness prediction framework that integrates extended 2-D body-and-stick keypoints with confidence-aware reasoning. Our contributions focus on two core areas:

- Develop a dedicated handedness classification module based on a lightweight multilayer perceptron (MLP) that consumes torso-normalized pose coordinates together with body and stick keypoint confidences.
- Extend handedness prediction from a frame-level task to a tracklet-level inference procedure by introducing a signed-confidence formulation that aggregates handedness labels across each tracklet.

A tracklet is a folder containing a continuous sequence of frames of a single player over a time period.

2. Methodology

2.1. Problem Definition

The objective of this work is to determine a player's handedness directly from broadcast video. Handedness is a deterministic player attribute within ice hockey, where approximately two-thirds of NHL players are left-handed and one-third right-handed, making it invaluable for preserving identity within multi-player tracking systems [4]. We formulate handedness prediction as a supervised classification problem, where the input is a set of 2-D pose keypoints and the output is a binary handedness label.

2.2. Handedness Definition

In ice hockey, handedness is defined by the orientation of the stick and the placement of the dominant upper hand. A player is considered left-handed if the left hand is positioned lower on the stick shaft and the blade points toward the player's left side. Conversely, they are considered right-handed if the right hand is positioned lower on the stick shaft and the blade points toward the player's right side.

2.3. Classification-Based Approach

Figure 2 shows the complete pipeline for frame to handedness label. First, the image goes through the pose estimation model to produce a set of 20 keypoints. Given a set of 20 detected pose keypoints per frame, as shown in Figure 1, we construct a geometric representation that encodes the relative position of the torso, arms, and stick. Handedness prediction is then modeled as a binary classification problem. We employ a MLP classifier, trained on large-scale frame-level samples.

Because frame-level predictions may be affected by noise and occlusion, we extend the classifier to operate over entire tracklets. Frame-level probabilities are aggregated into a tracklet-level decision, yielding a temporally stable handedness estimate for each consecutive sequence of frames.

Figure 1. Example player crop (left) and corresponding 20-keypoint pose estimate (right) used as input to the handedness classifier.

2.4. Tracklet-Level Aggregation Methods

Given a tracklet containing T frames, the model outputs for each frame t a probability $(p_t^{\text{left}}, p_t^{\text{right}})$ with $p_t^{\text{left}} + p_t^{\text{right}} = 1$. We define a hard frame label

$$m_t = p_t^{\text{right}} - p_t^{\text{left}}, \qquad \hat{y}_t = \begin{cases} 1, & m_t \geq 0, \\ 0, & m_t < 0. \end{cases}$$

Majority Voting This baseline selects the class occuring most frequently among the hard labels:

$$\hat{y}_{\text{maj}} = \begin{cases} 1, & \sum_{t=1}^{T} \hat{y}_t > T/2, \\ 0, & \text{otherwise}, \end{cases}$$

where 1 denotes right-handed and 0 left-handed.

Mean Probability This method averages the per-frame right-handedness probabilities and thresholds at 0.5:

$$\bar{p} = \frac{1}{T}\sum_{t=1}^{T} p_t^{\text{right}}, \qquad \hat{y}_{\text{mean}} = \begin{cases} 1, & \bar{p} \geq 0.5, \\ 0, & \bar{p} < 0.5. \end{cases}$$

Weighted Confidence We first compute a signed margin $m_t = p_t^{\text{right}} - p_t^{\text{left}} \in [-1, 1]$. We then apply

$$c_t = \operatorname{sign}(m_t)\big(\exp(k|m_t|) - 1\big), \quad k = 3.$$

The exponential form is chosen to nonlinearly emphasize frames with large signed margins while remaining insensitive to low-confidence predictions near zero. Linear weight was found to insufficiently suppress ambiguous frames, whereas higher-order polynomials introduce sharper transitions that can overemphasize isolated predictions. The parameter k controls the strength of this emphasis and was

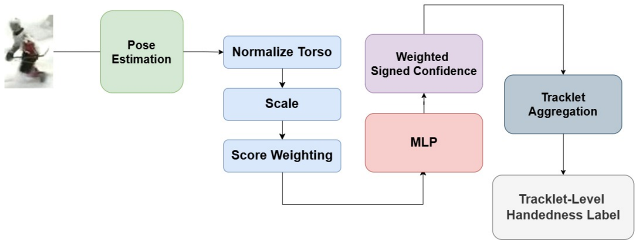

Figure 2. Figure showing handedness classification model architecture.

set to $k = 3$ as a moderate, fixed value that amplifies confidence predictions without requiring tuning or dataset-specific optimization.

The tracklet-level score is the mean signed confidence $\bar{c} = \frac{1}{T}\sum_{t=1}^{T} c_t$, and the final prediction is

$$\hat{y}_{\text{wconf}} = \begin{cases} 1, & \bar{c} > 0, \\ 0, & \bar{c} \le 0. \end{cases}$$

that is, right-handed if $\bar{c} > 0$ and left-handed otherwise.

3. Experiment

3.1. Experimental Setup

All experiments are conducted using PyTorch and the public MMPose library [3]. All experiments are conducted with a single RTX 6000 GPU. Only keypoint coordinates and confidences are used for handedness prediction, with no input regarding RGB appearance.

Although pre-trained pose estimation models are available, we trained a dedicated one for pose estimation to ensure consistent keypoint definitions aligned with the handedness goal. In particular, hockey-specific stick keypoints were often missing in other generic pose estimation models.

3.1.1. Pose Estimation

RTMPose is trained to provide consistent 20-keypoint detections for downstream handedness prediction. The network uses a CSPNeXt backbone with an RTMCC head and SimCC encoding [1, 6]. Training runs for 50 epochs using AdamW [8] with a cosine-annealing learning-rate. A linear warm-up phase initializes the schedule, and standard top-down data augmentations are applied. All preprocessing steps are matched between training and testing.

3.1.2. Handedness Prediction

Handedness prediction is performed using a multilayer perceptron. Each frame-level feature vector is constructed from 20 pose keypoints after torso normalization. The torso is defined by the shoulders, with fallback to the hips when shoulder confidence is low. Coordinates are centered, scaled by the torso width, and multiplied by their respective keypoint confidences. Two global confidence statistics, mean body-joint confidence and stick-join confidence, are appended as additional dimensions in training. During MLP training, light Gaussian jitter is applied to pose coordinates to increase robustness [5].

The MLP applies LayerNorm, two SiLU activated hidden layers (256 and 128 units), and dropout. Training uses class-weighted cross-entropy, Adam with a learning rate of 10^{-3}, and 50 epochs. Preprocessing is identical between training and inferencing.

3.2. Experiment Details

3.2.1. Datasets

We evaluate our pipeline using two datasets supporting pose estimation and handedness classification respectively.

Table 1. Datasets in experiments.

Section	Train / Val / Test Frames
Pose Estimation	2,800 / 300 / 0
Handedness	405,000 / 45,000 / 50,000

The pose estimation dataset is made up of broadcast frames containing bounding boxes and 20 keypoint annotations. The dataset contains of folders, each containing a selection of continuous frames from a specified game. Each folder contains a CSV file with all the x- and y-coordinates for the keypoints, the frame and player it belongs to, as well as the score for each point, most of which have an 100% accuracy. It is used exclusively for training and testing the pose estimation model. This dataset is used over the much larger handedness dataset due to it not missing any player or stick parts in its field of view, a problem that is present with the handedness dataset.

The handedness dataset is made up of player crops grouped by tracklets, each labeled with left/right handedness. The number of frames per tracklet varies widely, reflecting the diverse sources of footage used to construct the dataset. This dataset is derived from the one used in Player Tracking and Identification in Ice Hockey [10]. To prepare it for the purpose of our research, we removed referees from the dataset and replaced jersey number labels with handedness labels by referencing online player databases. This dataset is fed through the pose estimation model to produce 20 keypoints, which are then fed through the MLP layer for handedness prediction.

3.2.2. Splits and Training Parameters

The dataset split details are outlined in Table 1. For pose estimation, each frame is an image capturing the entire rink, with multiple player inside. For handedness, each frame is a tight crop of a single player.

3.2.3. Pose Estimation

The pose estimation model accuracies are evaluated with OKS [7]. OKS measures how close each predicted keypoint lies to its groundtruth position while accounting for person scale and per-join uncertainty. Following COCO convention, AP summarize performance across OKS thresholds from 0.50 to 0.95, whereas AP_{50} and AP_{75} denote accuracy at individual OKS thresholds of 0.50 and 0.75 respectively. The threshold determines the margin of error that is allowed, with higher thresholds leaving less margin for error. The complete model precision details are shown in Table 2

Table 2. Pose Estimation Model Precision.

Method	AP	AP_{50}	AP_{75}
MMPose	79.7%	91.0%	84.8%

Pose estimation accuracy is reported to contextualize the quality of the geometric inputs used for handedness prediction. Since the handedness classifier operates solely on pose coordinates and confidence values, its performance is inherently bounded by keypoint localization accuracy. Reporting pose estimation metrics therefore clarifies that downstream handedness results are obtained under realistic, non-ideal condition.

3.3. Results

The pose estimation model attains an AP of 79.7%, providing sufficiently accurate keypoints for downstream classification. The MLP achieves reliable frame-level handedness predictions, but temporal aggregation significantly improves stability. Among the aggregation strategies, the weighted confidence method achieves an accuracy of 95.03%, outperforming majority voting and mean probability as shown in Table 3. These results demonstrate that pose-derived geometric features, combined with confidence-weighted aggregation, yield robust handedness estimates suitable for unique identity tracking in multi-player systems.

Table 3. Comparison of tracklet-level aggregation approaches.

Method	Accuracy
$Majority\,Voting$	83.21%
$Mean\,Probability$	91.30%
$Weighted\,Confidence$	95.03%

Analysis of incorrectly classified tracklets suggests that residual errors are primarily attributable to persistent ambiguity at the frame level rather than limitations of the aggregation strategy. In such cases, the majority of frames exhibited low-confidence or inconsistent pose estimates due to stick occlusion, low resolution frames, or truncated player crops. When confident frames are sparse or absent, aggregation cannot recover a reliable handedness signal, indicating that further gains are likely to come from improved pose quality or richer geometric features rather than more complex aggregation methods.

4. Conclusion

This work introduces a pose-based approach for handedness prediction in broadcast hockey video, relying solely on 2-D keypoints derived from a top-down pose estimation model. By normalizing torso geometry, incorporating keypoint confidences, and applying a compact MLP classifier, the system produces frame-level predictions which are then aggregated to enhance stability. The proposed weighted confidence method achieves a 95% tracklet-level accuracy.

The results demonstrate that handedness is a strong, pose-expressible attribute that can be extracted in challenging broadcast conditions. This makes it well suited for unique identity tracking in multi-player settings. Future work may integrate handedness predictions directly into tracking frameworks and incorporate other attributes to jointly reason the identity of players.

Acknowledgments

The author would like to thank Stathlete for their support in this research.

References

[1] Xinyu Chen, Cheng Yang, Jiarui Mo, Yu Sun, Hamza Karmouni, Yifei Jiang, and Zhaohui Zheng. Cspnext: A new efficient token hybrid backbone. *Engineering Applications of Artificial Intelligence*, 2024. https://ouci.dntb.gov.ua/en/works/735PZvol/. 3

[2] MMPose Contributors. Rtmpose: Real-time multi-person pose estimation, 2023. GitHub Repository: https://github.com/open-mmlab/mmpose/tree/main/projects/rtmo. 1

[3] OpenMMLab Contributors. Mmpose: A flexible toolkit for human pose estimation, 2023. GitHub Repository: https://github.com/open-mmlab/mmpose. 1, 3

[4] François Grondin et al. Hockey intelligym: A deep learning system for tracking players in broadcast hockey video. *Scientific Reports*, 2024. https://www.nature.com/articles/s41598-024-70480-w. 2

[5] Trung-Hieu Hoang et al. Improving the robustness of 3d human pose estimation: A benchmark. In *CVPR Workshops*, 2024. https://openaccess.thecvf.com/content/CVPR2024W/TCV2024/papers/Hoang_Improving_the_Robustness_of_3D_Human_Pose_Estimation_A_Benchmark_CVPRW_2024_paper.pdf. 3

[6] Yanjie Li, Shuai Yang, Pan Liu, Shaowei Zhang, Yu Wang, Zhe Wang, Wei Yang, and Shu-Tao Xia. Simcc: A simple coordinate classification perspective for human pose estimation. In *European Conference on Computer Vision (ECCV)*, 2022. https://link-springer-com.demo.remotlog.com/chapter/10.1007/978-3-031-20068-7_6. 3

[7] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft coco: Common objects in context. *arXiv preprint*, 2014. https://arxiv.org/abs/1405.0312. 4

[8] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint*, 2019. https://arxiv.org/abs/1711.05101. 3

[9] Helmut Neher. Hockey player tracking thesis. https://uwspace.uwaterloo.ca/items/04139f2e-c1eb-48f9-97d7-e6cfeb6057e1, 2024. Advisor: David Clausi and Alexander Wong. 1

[10] Jonathan Tseng et al. Hake: A dataset for human activities and kinematic evaluations. *arXiv preprint*, 2021. https://arxiv.org/abs/2110.03090. 3