# Pose Estimation of Players in Hockey Videos using Convolutional Neural Networks


| **Helmut Neher** | **Mehrnaz Fani** | **David A. Clausi** | **Alex Wong** | **John Zelek** |
|---|---|---|---|---|
| University of Waterloo | Shiraz University | University of Waterloo | University of Waterloo | University of Waterloo |
| hneher@uwaterloo.ca | fani.mehrnaz@shirazu.ac.ir | dclausi@uwaterloo.ca | a28wong@uwaterloo.ca | jzelek@uwaterloo.ca |



## Abstract

*Traditional hockey scouting procedures for evaluating player performance is based on visual monitoring of hockey videos and statistics. However, that evaluation is time consuming and prone to human bias. In addition, current research within hockey analytics quantifies player performances by employing statistical models on common hockey statistics. To improve statistical models and increase the precision of evaluations, gathering data that is specific to each individual player, as well as, evaluating a player based on the capability of a player's performance is crucial. Techniques of skating, shooting, stick handling, passing, and speed of a player, are clues for identifying the performance of a player. In order to gather these key pieces of information from video data, the position of player limbs and joints should be found in each frame of the video. By knowing the location of joints in all video frames, one can easily track all body movements of players and use them for calculating desired information. In this work, we use a novel approach that determines a hockey player's body placement in video frames, also known as pose estimation via a convolutional neural network-a computer vision algorithm that imitates the learning processing of a brain. The proposed method provides a tool to analyze the pose of a hockey player via broadcast video which aids in the eventual assessment of a hockey player's speed, shooting etc., which, are alternatives to goal orientated statistics. The algorithm proves to be successful since it identifies on average 81.56% of the joints of a hockey player on a set of test images.*


Key words- hockey analytics, pose estimation, video analytics, convolutional neural networks, stacked hourglass network, compute vision.

## 1. Introduction

Hockey analytics is important for hockey players, coaches and analysts alike, because this field of research provides a scientific way to evaluate team and player performances. Various statistics can be extracted from league, team and players in each game or season. Typically, such data is derived manually by watching live or recorded hockey games, and gathering statistics determined across players, teams, and/or seasons. The statistics developed for hockey analytics, which are centered on goal specific statistics, provides a greater understanding of the *effects* of a player's performance. The statistics, however, do not assess the *capability* of a hockey player, which can be measured through assessing speed and technique.

The goal of this research is to automatically determine the action of a hockey player from game video in order to eventually assess the *capabilities* of that player. To accomplish this goal, we utilize a convolutional neural network (CNN), a computer vision algorithm that is able to learn patterns based on game video. The algorithm is designed to estimate the "pose" information of the player, namely the joint (e.g., wrist, shoulder, pelvis, knees, elbows neck) locations and associated limb positions. More information concerning recent trends of hockey analytics, and information concerning pose estimation and CNNs are found within the Background (Section 2) of this research paper. Pose estimation provides valuable information since it can be continuously derived from game video as opposed to goal or shot statistics that only occur periodically during the game for only the individual with the puck. Also, pose information can be derived from all players viewed in the video as opposed to the limited gathering of statistics based only on the player shooting.

The authors are not aware of any other research that focuses on pose estimation of hockey players and CNNs offer a novel approach to perform this task. Extracting the pose of a player is a gateway to provide analysts and coaches with the ability to automatically assess player performance via performance statistics based on speed and technique to evaluate the capability of a player. The first step is to use a CNN to estimate hockey player pose from still images and video as described in the Methodology (Section 3). Testing based on a selection of hockey images and video is presented in the Results and Discussion (Section 4).

## 2. Background

Hockey analytics is a growing research field that analyzes the characteristics of hockey players and teams through the use of statistics and other tools to gain a greater understanding of the effects of their performance. Fundamental statistics used for hockey analytics can be found at http://nhl.com and this includes goals, assists, plus-minus, etc. Statistics and models have recently been proposed to take into account shortcomings of goal orientated statistics. Examples include adjusting the plus/minus approach through the implementation of probability [1] and creates additional statistics such as Corsi to determine a shot differential between teams or the DeltaSOT statistic to adjust the delta statistic via teammate adjustment [2]. Two other recently developed statistics are THoR and DIGR [3][4] which provides alternative ratings for player performance. Thomas et al. [5], modeled the scoring rate for each team and adapts the model for each team based on statistically inferred additional variable effects such as home-ice advantage and game score. Gramacy et al. [6] implements a model to predict scoring based on players on the ice and the game situation (e.g. power-play or full-strength). Peel and Clauset [7] create different prediction models to determine who will score next and who will win the game, using information such as scoring, time of score, and the team which scored. Chan and Novati [8] utilizes common NHL statistics to determine and classify players into various levels of play. The disadvantage of some methods is that they span the whole duration of a game or even a season rendering the statistics and models ineffective for determining player performances based on a single play within a game [3-8]. All of the statistics and models mentioned above are evaluated based on shot orientated statistics; other valuable statistics such as player speed and technique are not evaluated. These statistics and models mentioned do not assess the abilities and capabilities of a player, which, do not explain how a player can improve; the statistics only portray the effects of the player with respect to a goal. The need for new methods to evaluate player performance in addition to explaining the capabilities of a hockey player is paramount.

Another approach to evaluate player performance is via video analytics. Video analytics to automatically assess hockey player abilities include tracking player movement [9][10][11] and recognizing player actions (e.g., skating, shooting, passing) [12][13][14] [15].Previous methods to analyze hockey players via video analytics was by determining the position of a player in a hockey rink (tracking a player) or through recognizing a player's actions (i.e., skate, shoot, or pass) by leveraging computer algorithms. Some methods to track players use the motion trajectory of a hockey player, or a sequence of images along with camera parameters to estimate the location of players [9][10]. A tracking system using the evaluation of color gradients of pixels either using a sequence of images or one image is conducted [11][12]. In addition to tracking, classifying basic actions of a hockey player (e.g. shooting, passing, skating, skating left, skating right) was also experimented [12]. In some instances, tracking multiple players occurred with or without the viewpoint of the camera and camera motion information to improve tracking [13][14]. Another method for video analytics is tracking player and analyzing data from other sensor technology such as inertial measurement units (IMUs). These sensors have been added to clothing and equipment of hockey players and used to determine a player's skating and shooting motion [15]. The sensors is expensive and the set-up time to mount additional hardware to register the sensors and collect data from the sensors is costly and time consuming. Video is preferred due to low cost, unobtrusive nature, and relatively easy capture.

Within the context of hockey analytics, tracking hockey players and player action recognition have both been researched. A new method that can be used to support hockey analytics from game video is pose estimation. A most common method to determine the pose of a player via video is to take a sequence of images, and using computer algorithms, determine the joints and limb based on relative movements. A CNN is a successful software algorithm that is able to recognize patterns in data such as still images or video. The CNN mimics the structure of the brain in order to learn from such data.

In our brain, there are billions of neurons, which are interconnected to each other to form our brain's neural network. Neurons receive input signals from sensory organs and combines this information to recognize patterns. This signal is finally used by our brain to make a decision. In the brain, based on the strength of a neuron (its weight), and the inter-connections between neurons, a person can learn. The bigger the network, the 'deeper' it is, the more expected the algorithms learns the patterns. Therefore, to mimic the brain's learning mechanism, artificial networks are developed and formed that mimic the capabilities of the human brain. CNNs are a particular type of neural network for interpreting images that are comprised of many layers of artificial neurons, where each artificial neuron is associated with a weight indicating its strength and with a mathematical model. By iteratively modifying weights of the network using data with the correct output (ie., training data), the computer algorithm strengthens what it learns. Hence, pose estimation using a CNN operates by using many images with known joint locations as training input, learns the weights that can be used to produce accurate outputs for this training data, and then applies this model to unknown input data to estimate pose automatically.

Statistics within the field of hockey analytics are limited to evaluate player performance and player ability due to the goal orientated nature of statistics. Video analytics provides an alternative view. The use of pose estimation for hockey players provides a tool used within the hockey analytics field to later assess player abilities and performance by assessing player statistics such as speed and technique.

## 3. Methodology

As seen in Figure 1, to estimate the pose of a hockey player, six processes occur. An image is first captured from a broadcast hockey video, then the location of the player's body center is automatically determined through a computer vision algorithm that tracks the body center. The image and the coordinate of the body center are then rescaled to an optimal resolution for the CNN (i.e., 720 x 1280 pixels) and cropped to isolate the desired player. Next, the scaled and cropped image, along with body center point is delivered into the hourglass network for pose estimation. The output of the network are 16 joint coordinates that indicated the predicted place for a joint. Afterwards, the body limbs are obtained by connecting related joints together. Finally, an output is visually seen with color-coded limbs in the image.

The output of joint coordinates for players in each frame are valuable pieces of information that can later be used to evaluate player performance such as a player's position, movement, speed, and reaction to a play for a whole broadcasted hockey game. At this aim, the vehicle to determine the joint placement is a convolutional neural network for finding places of joints of a hockey player in each frame of a hockey video or in a single image of a hockey player.

The employed convolutional neural network, shown in Figure 2(a), is called the stacked hourglass network [16], and is composed of several layers of artificial neurons that are interconnected to each other. The weights of the interconnections, are previously learned from a dataset is the MPII Human Pose Dataset, which is composed of 40,000 images, annotated for 16 body-joints [17]. This network is trained to learn the place of joints for images of people doing daily activities. The output of the network is a set of heatmaps. Each heatmap is an image which gives the predicted probability of a joint's presence at each and every image pixel [17]. In Figure 2(b) an example of a heatmap is provided that predicts the location of left elbow in the image. Where, pixels with blue color are less probable to be assigned to the left elbow, while for the pixels that are near the elbow location, the heatmap color gradually grows from blue to yellow and finally to red. The 16 body-joints that are spotted by this network are listed in Table I. The experimental results in the next section show the proposed methodology for hockey pose estimation.

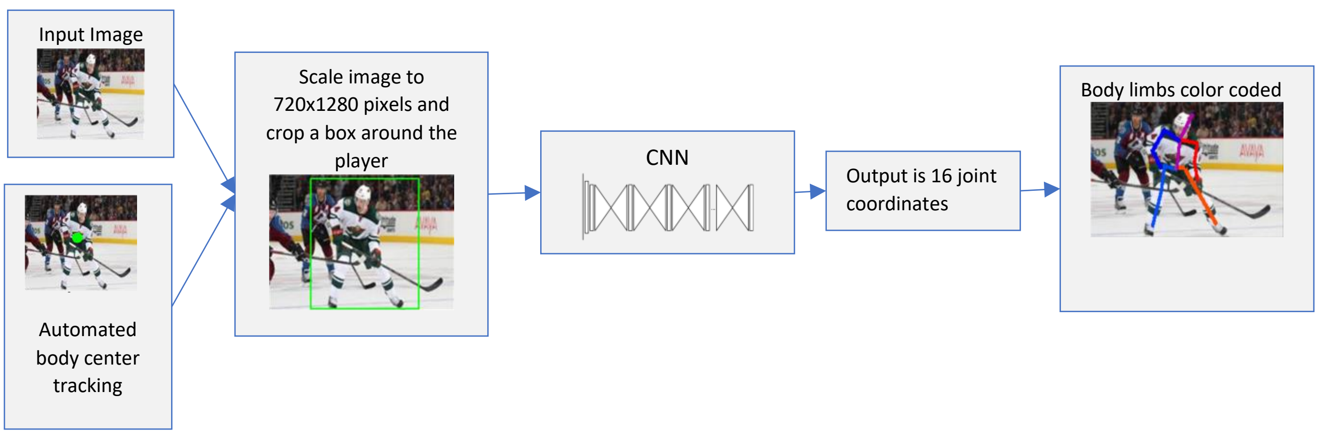


Figure 1 The proposed framework for hockey pose estimation. The input to the CNN is a set of trained properly scaled images and a known tracked body center for the player. The CNN output is the 16 joint coordinates color coded for ease of viewing

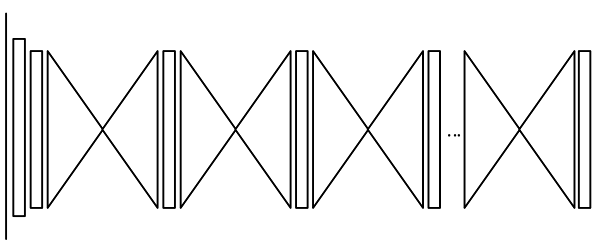

Figure 2. (a) The stacked hourglass network which is the employed CNN for this research.

(b) Instance of a heatmap that shows the predicted joint placements for the left shoulder of a hockey player

Table I. List of 16 body joints

| # | **Body-Joints** | # | **Body-Joints** |
|---|---|---|---|
| 1 | Right ankle | 9 | Upper neck |
| 2 | Right knee | 10 | Head top |
| 3 | Right hip | 11 | Right wrist |
| 4 | Left hip | 12 | Right elbow |
| 5 | Left knee | 13 | Right shoulder |
| 6 | Left ankle | 14 | Left Shoulder |
| 7 | Pelvis | 15 | Left elbow |
| 8 | Thorax | 16 | Left wrist |

## 4. Experimental Results and Discussion

In this part, different experiments are conducted to show the accuracy and effectiveness of the proposed methodology in determining the estimated pose of a hockey player. The results are categorized in three groups: results on still images, results on videos, and finally numerical results.

### A. Test Results on Still Images

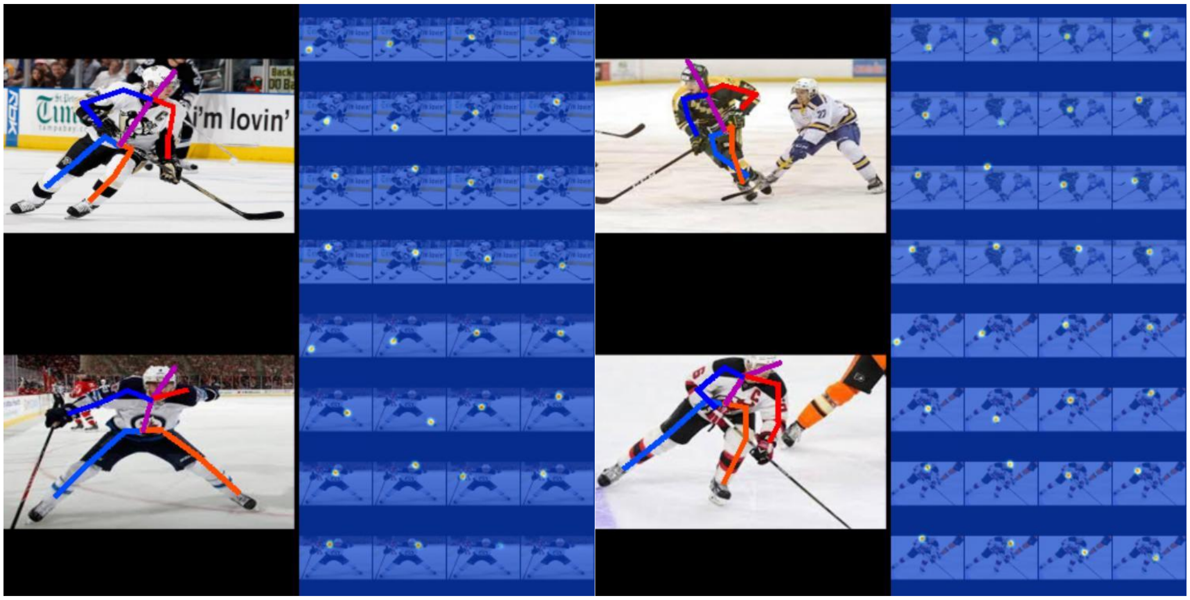

Figure 3 Visual demonstration of hockey pose estimation along with corresponding heatmaps for four still images of hockey players from a corpus of images collected.

In the first part of experiments, the framework of Figure 1 is applied on 20 still images of hockey players from a corpus of images with various body poses. Four of the output images along with their corresponding heatmaps are given in images of Figure 3. From these images, joints have been found successfully in each. The 16 heatmaps that are provided for each image, show the predicted probability for location of each joint, with the same order that is given in Table I. In the heatmaps blue color correspond to low probability, while growing to yellow and then red shows the increment of probability for a joint to be situated in a specific position in the image.

From visual inspection, the predictions are accurate; each image demonstrates by the color-coded limbs the general pose of the hockey player. In fact, hockey sticks and bodies of other players do not affect or skew the accuracy of the results; the limbs and hockey sticks of other players aren't recognized as the limbs of the player being evaluated. Some joints, however, are not accurately identified such as the ankle joints of the player in question; the equipment of a hockey player may obscure the joints from the CNN. Although, some of the joints may not be accurately located, the CNN successfully identifies the general pose of the hockey player being evaluated without interference from other players within the still image.

### B. Test Results on Video

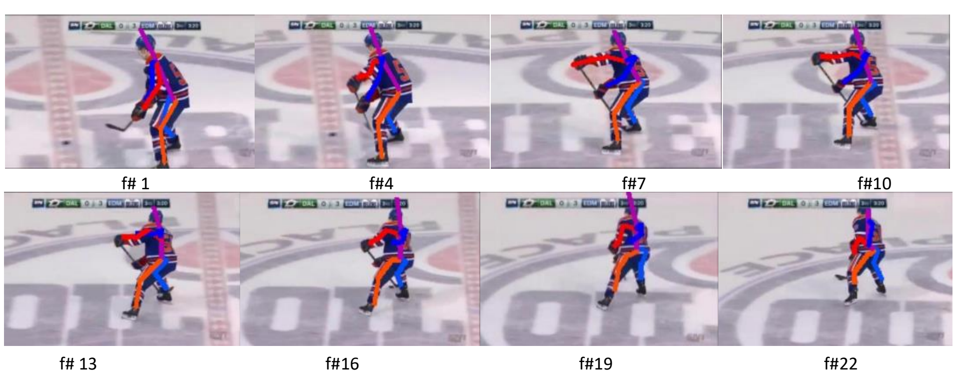


Fig. 4 Visual demonstration of hockey pose estimations for eight frames of an 114 frame hockey video.

In the second experiment conducted, Figure 1 is applied on a video of 114 frames, (with frame size of 640 x 480, and frame rate of 25 f/sec). Some of the output frames are illustrated in images of Figure 4, for visual assessment. In Figure 4, most of the body joints are detected with high accuracy even in the frames where the right arm is hidden behind the player's body (i.e., frames # 13, 16, 19, and 22). There are still some inaccurate joint locations, but in general, the player's pose is accurately assessed. Therefore, the test video demonstrates the capability of the proposed procedure in pose estimation.

### C. Numerical results

The percentage of correctly identified joints for 20 still images of hockey players are reported in Table II. According to Table II, the highest accuracies of detection achieved are for head and ankle joints, which are both 90%, while, wrists and elbows are detected with lowest precisions, i.e., 65% and 70% respectively. Generally, the equipment a player wears, namely skates and gloves, hides the joints from the algorithm that adversely affects the precision of the estimated pose. However, as given in the last column of Table II, the average of correct detections for all body-joints is 81.56%. Since the overall percentage indicates that most of the limbs are detected per image, pose estimation is a valuable tool to extract data from video and still images.

Table II. Accuracy of joint detection for 20 Hockey Player Test Images

| **Joint name** | Head | Shoulder | Elbow | Wrist | Hip | Knee | Ankle | Total |
|---|---|---|---|---|---|---|---|---|
| **Accuracy of detection** | 90% | 87.5% | 70% | 65% | 80% | 82.5% | 90% | 81.56% |

## 5. Conclusion and Future Research

In this work, an automated method to determine the pose of a hockey player from broadcast game video is achieved. The successful numerical and visual outcomes prove that the employed convolutional neural network for estimating the pose is excellent even in challenging scenarios like hockey videos, where severe occlusions (i.e., bulky clothing and high speed of players due to skating), exist.

The results of this research provide a contribution to hockey analytics. By estimating the pose of a player through a convolutional neural network, coaches and analysts can utilize the joint data received to evaluate a hockey player based on their capability as a player rather than based on the effects of a hockey player's performance. Information like velocity and technique of shooting, action recognition, and skating speed for each player can be developed via pose estimation using hockey videos and CNNs. Further research can be done by improving the accuracy of pose estimation for hockey players. Joints such as hips, ankles and wrists could be improved by collecting more data to be used in the training of the CNN. In addition, the pose estimation of hockey goalies is another area of further research. Estimating a goalie's pose is quite challenging since a goalie's gloves, helmet and pads of a goalie hide the joints of the player; the CNN is unable to predict knee, ankle, wrists and sometimes joints around the head. Another avenue of research, is to implement of the joint coordinates in assessing a player's performance.

References

[1] Schuckers, M., Lock, D., Wells, C., Knickerbocker, C., and Lock, R. (2010). “National Hockey League Skater Ratings Based upon All On-Ice Events: An Adjusted Minus/Plus Probability (AMPP) Approach.” Tech. rep., St. Lawrence University

[2] Vollman, R. (2010). “Howe and Why: Ten Ways to Measure Defensive Contributions.” Hockey Prospectus, March 04, 2010.

[3] Schuckers, M, Curro, J “Total Hockey Rating (THoR): A comprehensive statistical rating of National Hockey League forwards and defensemen based upon all on-ice events”, in Proceedings of the 2013 MIT Sloan Sports Analytics Conference, March 2013

[4] Schuckers, M “DIGR: A Defense Independent Rating of NHL Goaltenders using Spatially Smoothed Save Percentage Maps” Proceedings of the MIT Sloan Sports Analytics Conference, March 2011.

[5] Thomas, A., Ventura, S., Jensen, S., and Ma, S. (2012). “Competing Process Hazard Function Models for Player Ratings in Ice Hockey.” Tech. rep., ArXiv:1208.0799.

[6] Gramacy, R., Taddy, M., Tian, S. "Hockey Player Performance via Regularized Logistic Regression." *arXiv preprint arXiv:1510.02172* (2015).

[7] Peel, L., Clausie, A. “Predicting sports scoring dynamics with restoration and anti-persistence.” Available online at http://arxiv.org/abs/1504.05872, 2015.

[8] Chan, T., Novati, David. “Split personalities of NHL players: Using clustering, projection and regression to measure individual point shares”. In: MIT Sloan Sports Analytics Conference. 2012. URL: http : / / www . sloansportsconference .com /wp-content/uploads/2012/02/59-Chan_Novati_Split-personalities-of-NHL-players.pdf.

[9] Okuma, K., Little, J., Lowe, D. “Automatic Acquisition of Motion Trajectories: tracking Hockey Players”, Proceedings of the SPIE, vol. 5304, pp. 202-213, 2003.

[10] Okuma, K., Taleghani, A., Freitas, N., Little, J., Lowe, David. “A boosted particle filter: multitarget detection and tracking”, in: European Conference on Computer Vision, Prague, Czech Republic, May 11–14, 2004.

[11] Lu, W.-L., Okuma, K., Little, J. “Tracking and recognizing actions of multiple hockey players using the boosted particle filter”, Image and Vision Computing, Volume 27, Issues 1–2, 1 January 2009, Pages 189-205, ISSN 0262-8856, http://dx.doi.org/10.1016/j.imavis.2008.02.008. (http://www.sciencedirect.com/science/article/ pii/S0262885608000462)

[12] Lu, W.-L., Little, J. “Tracking and recognizing actions at a distance”, in: ECCV Workshop on Computer Vision Based Analysis in Sport Environments, 2006, pp. 49-60.

[13] Lu, W.-L., Little, J. “Simultaneous tracking and action recognition using the PCA-HOG descriptor”, in: The Third Canadian Conference on Computer and Robot Vision, 2006.

[14] Cai, Y., de Freitas, N., Little, J. 2006. “Robust Visual Tracking for Multiple Targets”. Springer, Berlin Heidelberg, pp. 107-118.

[15] Hardegger, M., Ledergerber, B., Mutter, S., Vogt, C., Seiter, J., Cala-troni, A., Tröster, G. "Sensor technology for ice hockey and skating", Wearable and Implantable Body Sensor Networks (BSN) 2015 IEEE 12th International Conference on, pp. 1-6, June 2015.

[16] Newell A., Yang K., Deng J. (2016) “Stacked Hourglass Networks for Human Pose Estimation”. In: Leibe B., Matas J., Sebe N., Welling M. (eds) Computer Vision – ECCV 2016. ECCV 2016. Lecture Notes in Computer Science, vol 9912. Springer, Cham

[17] Andriluka, M., Pishchulin, L., Gehler, P., and Schiele, B. “2d human pose estimation: New benchmark and state of the art analysis.” In CVPR, 2014