Video-Based Ski Jump Style Scoring from Pose Trajectory

Dejan Štepec[1,2], Danijel Skočaj[1]
[1]University of Ljubljana, Faculty of Computer and Information Science, Ljubljana, Slovenia
[2]XLAB d.o.o., Ljubljana, Slovenia
dejan.stepec@xlab.si

Abstract

Ski jumping is one of the oldest winter sports and takes also part in the Winter Olympics from the very start in 1924. One of the components of the final score, which is used for ranking the competitors, is the style score, given by five judges. The goal of this work was to develop a prototype for automatic style scoring from videos. As the main source of information, the proposed approach uses the detected locations of the ski jumper body parts and his skis to capture a full-body movement through the entire ski jump. We extended a method for human pose estimation from images to detect also the tips and the tails of the skies and adapted it to the domain of ski jumping. We proposed a method to utilize the detected trajectories along with the scores given by real judges to build a model for predicting the style scores. The experimental results obtained on the data that we had available show that the proposed computer-vision-based system for automatic style scoring achieves an error comparable to the error of real judges.

1. Introduction

In recent years, technology has been entering the world of sports at a great pace. In addition to the traditional applications used for improving the performance of athletes, and the experience of spectators, very recently, various technological aids have been developed for increasing the fairness of competitions by helping the referees to bring better decisions. In this paper, we focus on the application of such technological solutions in winter sports, more specifically in ski jumping (Figure 1).

In this sport, the competitors are ranked according to the score they achieve, which is compound of three components: (i) the jump length, (ii) the jump style, and (iii) the compensation for inrun length and wind conditions. The latest component was introduced in 2009, and the wind speed meters are used, along with the starting gate information, to compensate for variable outdoor conditions when calculating the ski jumper's score. Also for measuring the jump length various technological aids can be used. It is quite a well-defined problem; once the landing point of a ski jumper is determined, the jump length can be calculated using a suitable measuring device. Explicit rules exist on how to perform this measurement, which can be straightforwardly implemented, providing that a (imaging) sensor system enables sufficiently accurate measurement.

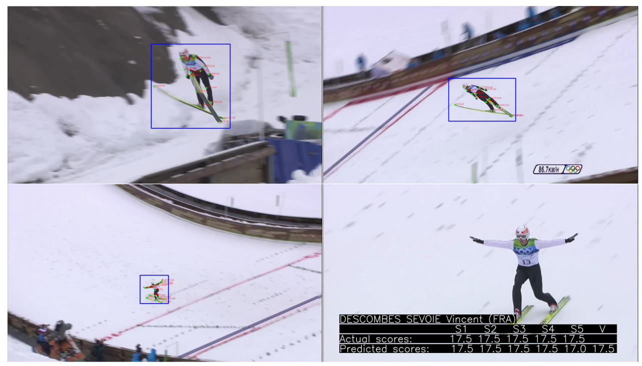


Figure 1: Proposed automated system for style scoring based on evaluating the body and ski parts trajectories during the ski jump.

The remaining component, judging the jump style, is significantly more difficult to automatize. The International ski competition rules, defined by FIS [8], specify how the judges should score the points for the jump style. However, these rules are quite vaguely defined; with a few exceptions, they do not contain clear quantitative descriptions, which could be explicitly implemented in terms of, e.g., measuring specified angles or distances. They are rather mostly given qualitatively; i.e., the judging criteria include "Actively utilisation of the air pressure", "A smooth movement from the flight position to landing by straightening the upper body", and "With equal weight on both legs in upright body position skiing safely through the fall line with arms and legs in any relaxed position." For many of these criteria, it is therefore not possible to program the measuring technique. How to explicitly program the measurement of "utilisation of air pressure" or "skiing safely"? It is something that humans, i.e., experienced judges can implicitly asses easily. We, therefore, opted to learn this judging technique from real judges by observing the athlete's performance and the style scores they were given. In addition, we wanted to avoid any sensors that would have to be attached to the ski jumpers or the skis. We wanted to develop a completely non-intrusive technique for ski jump style scoring, by using only information provided by a camera.

In this paper, we, therefore, present a computer vision approach for style scoring. We utilise regular video sequences of ski jumps as recorded during the TV broadcasts. In every frame, the ski jumper and the individual body parts, as well as skies, are detected, as depicted in Figure 1. The trajectories of the detected parts through time (image coordinates in the video sequence) are then used to describe the jump. They are then utilised to predict the score of the individual judges, based on the models build in the learning stage. Several deep learning methods are used to perform these tasks. By combining the predicted scores a virtual score is calculated. Experimental results show that the obtained virtual scores are in the range of the scores given by the real judges, and demonstrate the potential of the proposed approach for automating the style scoring.

2. Related Work

There has been a steady increase in research in the ski jump domain in recent years, including works, that use computer vision directly. To the best of our knowledge, there is no research work published, except our prior work [17], that would directly tackle the problem of style scoring with computer vision methods, but similar problems of distance measurement [6] and jump parameters estimation [22, 15] are present in the literature.

The works [22] and [15] represent the closest contributions to our work. In [22] they automate the computation of jump forces solely based on cameras, predominately based on ski jumper's posture, while in [15] they automate estimation of different flight parameters (in terms of body-ski-trajectory angles). The architecture of the system in [22] is very similar to our work, by first detecting a ski jumper using a MobileNet [10] and then estimating the pose with Convolutional Pose Machines (CPM) [19]. The pose estimation was split into separate steps of body pose estimation (CPM) and the detection of ski parts (ski tips and ski tails) using a Hough transformation. In comparison, in our work, we extend the CPM [19] method to directly detect body and ski parts. This was done in a similar fashion in a recent work [15], where they have instead used Mask R-CNN method [9]. The dataset used in [15] consisted out of 10,070 annotated frames from 290 jumps, obtained via professional camera setup.

Style scoring for ski jumping was already tackled in the literature [2, 3, 14], but all the solutions are built around inertial sensors and are thus not based on imaging data. In [2] they used a similar approach of using CNNs to process obtained trajectories, but the problem was simplified to multiple binary classification tasks of predicting whether a certain trajectory contains a specific error, which could cause style point deductions. The work in [3] also focuses on estimating deductions points, but more accurately, by comparing the trajectory of a specific jump to the set of trajectories in training data, which were free of that specific error. Works that include inertial sensors [2, 3, 14] performed small-scale experiments on at most 6 participants, but included professional FIS accredited judges, thus making available point deductions per separate flight stages.

Pose estimation represents the main input data for many downstream tasks in ski jumping, in our work, as well as in related work [22, 15], 2D pose estimation is used, but 3D pose estimation from a single monocular RGB image represents another field of research, not yet applied to ski jumping domain. The recent work in 3D pose estimation [5, 18] has significantly improved the real-world applicability of such methods, including with the real-time capability on mobile devices [5] and with data captured in the wild [18]. 3D pose information should significantly boost the performance of downstream ski jump analysis tasks and presents an open research problem. Our proposed proof-of-concept modular architecture enables future adaptability to these new advancements.

3. Methodology

In this section, we present the main building blocks of the system for an automatic scoring of the ski jump style. The system is designed in a modular fashion, which enables the use of different methods. The architecture of the system is presented in Figure 2 and consists out of three main modules, namely: (1) Ski jumper detector, (2) body and ski parts detector, and (3) the method for scoring the style of a jump. The input is represented by the jump video and the output represents the style score of the jump. The separate modules are presented in detail in the following subsections.

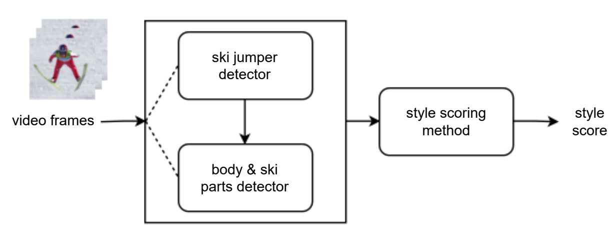


Figure 2: The architecture of the system for an automatic ski jump style scoring.

3.1. Ski Jumper Detection

The detection of the ski jumper serves for estimating his location and size, the information which is used in the pose estimation module. We utilised Faster R-CNN [16], which we adapted to our custom dataset for ski jumper detection. We initialized the method on the COCO [13] dataset and adapted only the last fully connected layer to support ski jumper detection. The obtained location and the size of the detected ski jumper are then used to significantly speed up the pose estimation method, by applying it only to the detected region, on a single scale. The proposed approach applies detection on every video frame separately, which could be also replaced with a short-term visual tracker in a production system.

3.2. Detection of Body and Ski Parts

The detection of body and ski parts represents one of the main contributions of our work. We used the Convolutional Pose Machines (CPM) [19] implementation from OpenPose [4] as a baseline method for pose estimation. The CPM [19] method is built around multi-stage CNN architecture, where the input into the first stage is just the image, while in later stages, both the image, as well as intermediate detections are used as an input. This enables the method to model the interactions between different body parts, in order to resolve the potential ambiguities and fine-tune the positional accuracy of the detections. We used 6-stage CPM [19] implementation and the outputs of different stages for some of the body parts are depicted in Figure 3. We can see that the symmetric body parts are poorly separated after the first stage, as the topological order of the body parts and their interdependence is still not inferred. One obvious example is the right ankle in Figure 3, which is at first detected as the left ankle, but it then gets slowly properly detected in the 4th and later stages, due to inference from the locations of the other body parts.

The body pose in a ski jump is quite unique and not frequently encountered in the everyday life, therefore the pretrained models, such as the ones trained on MPII [1] or COCO [13], are not adequate for an accurate detection. We, therefore, created a specific ski jump dataset, presented in Section 4.1, and also extended the CPM [19] method with the capability of detecting the tail and the front points of the skies. The method takes input image regions of the size of 368 x 368 pixels, centered around the detected ski jumper, which also enables to scale the input, such that the ski jumper represents 70% of the frame size input. The ground truth data is represented by the location of the body parts, presented in the top row of Figure 5. The CPM [19] method augments this, by adding Gaussian peaks with small variance (bottom row in Figure 5), which represents an ideal belief map, to be learned by the method.

We extended the CPM [19] method with the capability of detecting the ski parts, by treating them as additional body parts. We can see in Figure 5 that the positions of skies differ in different stages of the flight and that they also directly influence the pose of the ski jumper, thus making it reasonable to threat the ski parts as additional body parts.

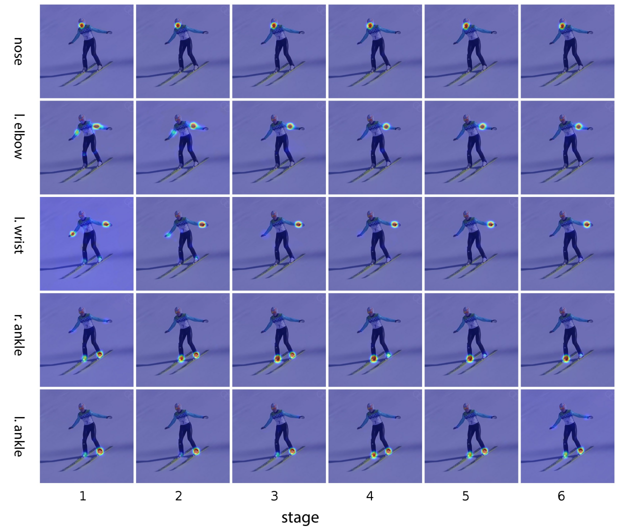


Figure 3: Output probability maps for individual body parts (nose, left shoulder, left wrist, right ankle, left ankle) on each stage (n=6) of the CPM [19] method.

3.3. Ski Jump Style Scoring

The style of the ski jump is judged by 5 judges, where each judge can give up to 20 points. The lowest ad the highest scores are eliminated, thus style points may reach a maximum of 60 points. The jump is evaluated from the end of the take-off to the passing of the fall line in the outrun and the jump must be judged based on the outer appearance of the succession of the jumpers movements, from the aspect of precision (timing), perfection (carrying out the movements), stability (flight-position, outrun) and general impression [8]. According to the FIS rule book [8], the ideal performance is concerned with utilisation of the aerodynamic efficiency of the body and ski, the posture of arms, legs, as well as ski position during the flight, and their succession of the movements during landing and outrun. The point deductions for the faults are specified in [8] and should be reported separately for the flight, landing, and outrun stages.

From the above description, we can notice that the rules are vaguely defined and carry a lot of subjectivity. Nevertheless, the posture of the body and ski parts plays a crucial role in scoring and is thus also used in our work as the primary input. To obtain a style score, we use the information of the locations of the body and ski parts during the ski jump. To achieve that, we encode the information about the

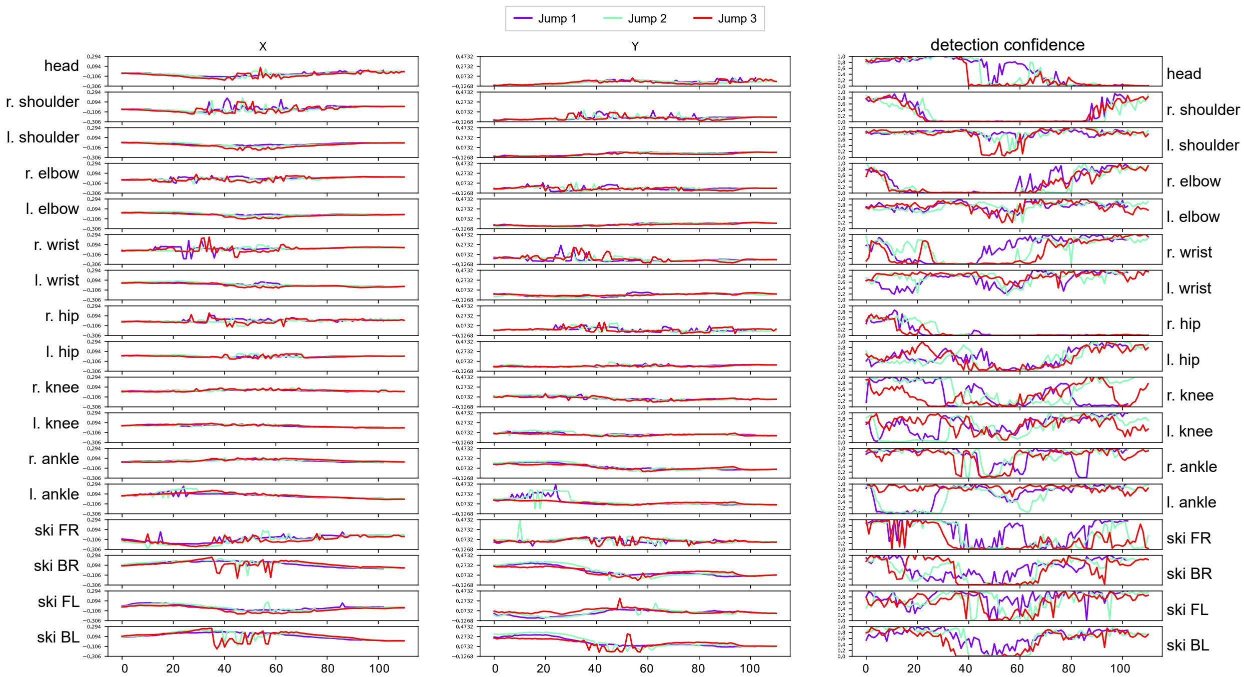


Figure 4: Visualization of the raw input into the method for the style scoring. The visualization shows the trajectories and their confidence for three different jumps, where the locations are normalized according to the body mass center.

locations of the body and ski parts (X and Y coordinates), along with their confidence estimation (C) in a 3D image-like structure with 3 slices (X, Y, C), where the width of the structure equals the number of body and ski parts and height to the number of the image frames captured for the jump. The coordinates of the body and ski parts are normalized according to the center of mass of the body. The encoding in the form of 3D tensors enabled the use of CNN architectures for style scoring, which is widely used in action recognition [7, 12, 21].

For style scoring, we used a similar approach to action recognition work [7], based on 2D CNNs, but with a simplified architecture consisting out of two convolutional layers, each followed by a max pooling layer, and two fully-connected layers at the end. We cast the problem as a regression problem, using an L2 loss, to predict the scores of 5 judges. The actual raw input into the method with the trajectories and their confidence for the three example jumps is presented in Figure 4. We can notice how detection confidence nicely spots spurious detections (e.g. Y-axis for l. ankle in jumps 1 and 2), effectively pushing confidence to zero, which is also learned by the style scoring method.

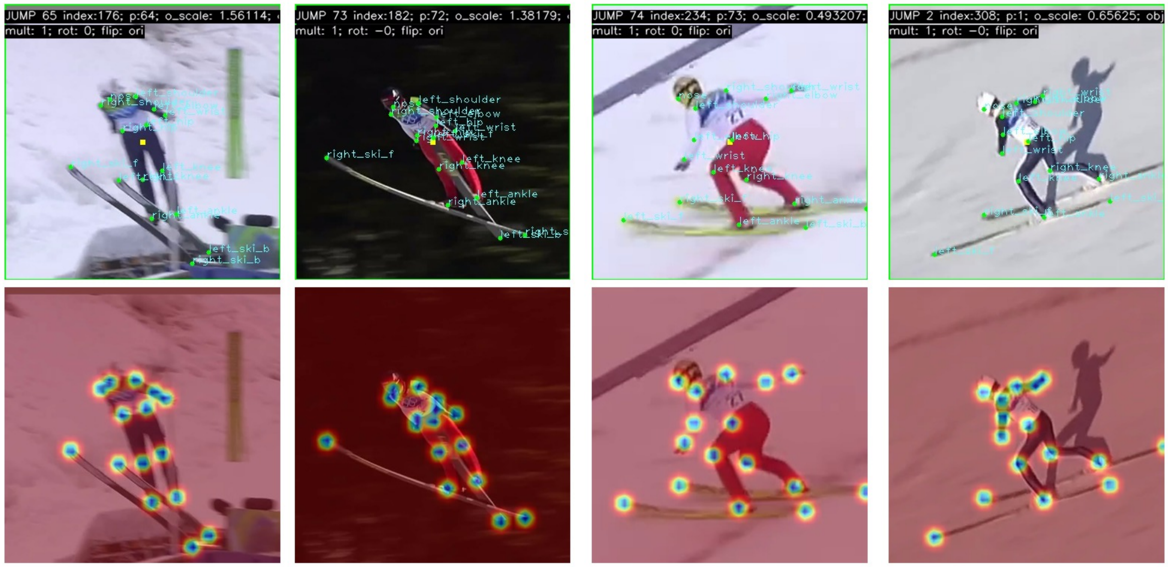
4. Experiments and Results

In this section, we present the performed experiments and results on a newly created dataset for ski jump style scoring. We first present the dataset and the labels provided in Section 4.1 and then the results for each of the solution's building blocks - ski jumper detection and pose estimation results in Section 4.2 and style scoring in Section 4.3.

4.1. Ski Jump Dataset

A specific ski jump dataset for human pose estimation was constructed for successful domain adaptation and inclusion of ski parts, not included in existing pose estimation datasets [1, 11, 13]. We have also included the data that is needed for ski jumper detection (bounding box) for each of the frames and ski jump style scoring and other metadata related to the particular competition and ski jumper. The dataset represents the first of such kind in the ski jump domain and contains almost 1800 annotated images from the 2010 Winter Olympics in Vancouver. The prototype was developed on the TV footage available on YouTube and we have chosen a particular competition and footage due to consistent camera movements. A production setup would allow for (and probably require) a professional dedicated camera setup which would significantly constrain the environment and our prototype can be viewed as the lower bound of what is possible to achieve in a constrained, professional environment, if available.

The distribution of annotated frames and jumps across the ski jump hill types and train/test sets is presented in Table 1. The training part of the dataset was constructed from the first rounds of the competitions and the test set from



(a) take-off **(b)** flight **(c)** landing **(d)** outrun

Figure 5: Example annotations for each part of the flight (take-off, flight, landing, outrun) - upper row and the actual input into the pose estimation method CPM [19], with added 2D Gaussian peaks with small variance on the body and ski parts locations.

the final rounds. For training (first rounds) we labeled 50 jumps from small and big ski jump hill types (100 jumps all-together) and, for the test set, 30 jumps in the final rounds (60 all-together), both using the footage from the main competition. Additionally, we labeled 50 jumps from the first round of the Nordic combined discipline (denoted with †) on the small ski jump hill type in order to increase the variety of the ski jump style scores. We have provided 17 labels for body and ski parts for each of the frames, as presented in Figure 5 (upper row) and the distribution of frames over the ski jump is presented in Figure 6. We see from the distribution that most of the frames were labeled in the first and last phases of the jump, due to large body and ski parts movements. During the flight, the position of the ski jumper is much more stable, thus fewer annotations are needed to achieve the satisfying performance of pose estimation.

Table 1: Distribution of annotated frames (detection) and jumps (scoring) across jumping hills and train/test sets.

Ski jump hill	Detection train	Detection test	$\sum$	Scoring train	Scoring test	$\sum$
big	679	116	795	50	30	80
small	868	98	966	$50^* + 50^\dagger$	30^*	130
$\sum$	1547	214	1761	$100 + 50^\dagger$	60	210

*Main competition data used for style scoring (Section 4.3)
†Additional data from Nordic combined discipline

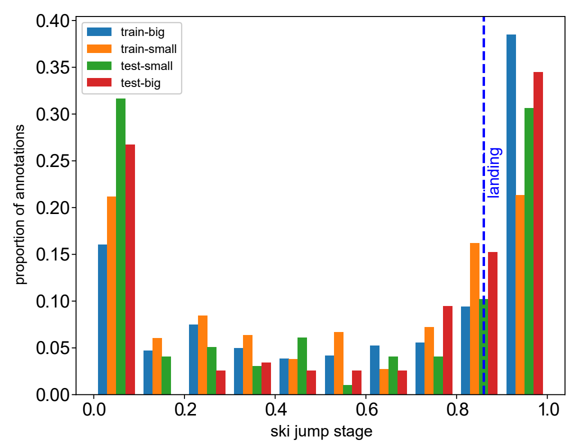


Figure 6: Distribution of annotations along the ski jump and ski jump hill types. We divided the jump into 10 parts, with specifically marked landing moment (averaged across all the jumps).

We have also included additional information about the competition, such as individual performances of the ski jumpers using a unique FIS code, length of the jump, and its associated style scores given by 5 judges.

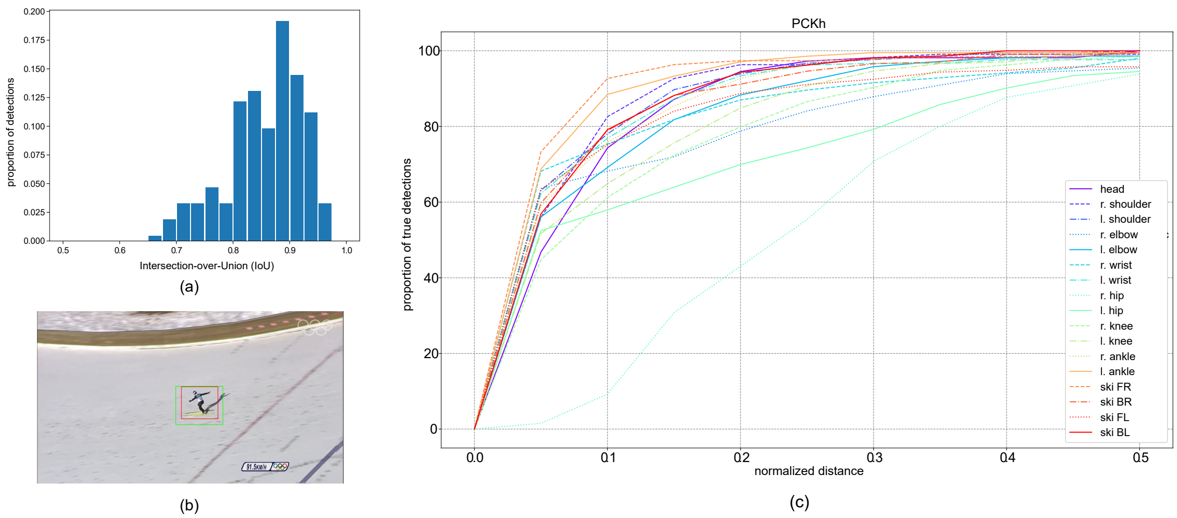


Figure 7: (a) The distribution of IoU values across the test set of frames for ski jumper detection. (b) An example of the worst detection ($IoU = 0.65$). The red rectangle represents the ground truth location, while the green one represents the predicted one. (c) PCKh value for different values of α for all body and ski parts.

4.2. Ski Jumper Detection and Pose Estimation

We evaluate the method for ski jumper detection and pose estimation on our test set of frames (214 frames) on small and big ski jump hills of the Vancouver Winter Olympics. We have used Intersection-over-Union (IoU) as a measure of our detector's performance. The distribution of IoU scores across test frames is presented in Figure 7(a). The histogram shows that more than 80% of detections are detected with $IoU \geq 0.8$, with $IoU = 1$ representing the perfect detection. The worst detection in terms of IoU score is presented in Figure 7(b). We can see that the performance of the detector sufficiently fulfills the performance requirements of the pose estimation method.

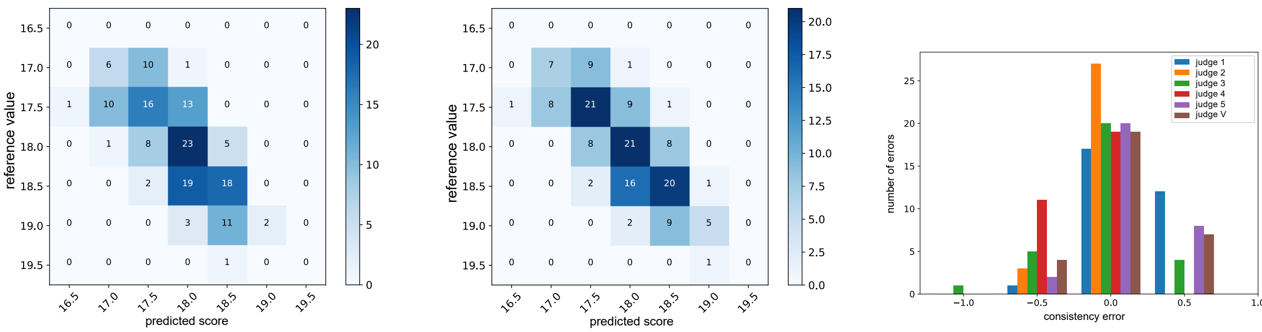
For pose estimation performance evaluation, we used the PCK metric, first presented in [20]. The detection of the individual body or ski part is deemed correct if it lies on a distance from the ground truth location, that is less than $\alpha * max(h, w)$, where h, w represents the height and width of the frame around the person that we want to detect the pose. α represents the threshold for that distance. MPII [1] dataset uses a slightly different implementation of the PCK metric (i.e. PCKh), which we also use in our work. PCKh metric [1] uses the head of the person as a reference frame, in order to reduce the influence of different body pose constellations on the metric. The maximum distance for the detection to be deemed correct is presented with the Equation 1. X_1 and X_2 are the locations of the diagonally opposite rectangular corners of the head annotation. The actual evaluation script provided for MPII [1] applies an additional scaling factor of 0.6 and the threshold factor α set to 0.5.

$$max\_dist = \alpha * 0.6 * \|X_1 - X_2\|_{L_2} \tag{1}$$

Body and ski parts detection results are presented in Figure 7(c). We can see that the vast majority of the limbs are detected in over 75% of the cases already at the threshold value of $\alpha = 0.1$, which is roughly equal to an error grade of the 5% of the size of the ski jumper's head. We can also notice that the right part of the body is slightly less accurately detected, which is due to the camera view from the left side. Importantly, we can see that the detection of the ski parts is equally accurate to the other limbs, which implies that the method successfully modeled the newly designed constellation of body and ski parts. Figure 9 presents the qualitative result of limb and ski parts detection. We similarly notice slightly worse detection of the right-hand side body and ski parts and the ones, that are not so visible, due to the camera setup and were also not labeled in a larger quantity. Such false detections are learned to be filtered by using the confidence information in the style scoring method.

4.3. Ski Jump Style Scoring

We evaluate the performance of style scoring by directly predicting the scores of the 5 judges and then measuring the absolute error against the scores given by them, as well as by creating an additional virtual judge out of predictions and then measuring its consistency with other judges. We perform all of the evaluations for style scoring on a small ski jump hill, due to consistent camera setting across jumpers. Initially, we only use the training data from the first round



(a) Main competition only*   **(b)** Main + Nordic combined competitions†   **(c)** CS on Main + Nordic combined†

Figure 8: Confusion matrix, obtained by predicting style scores that would be given by real-world judges (1-5), trained solely on main competition data (a) and when combined with Nordic combined competition data (b). Histogram of the consistency scores (CS) for our best performing model (Judge V†) and five real judges (c).

of the main competition (n=50 jumpers, marked with * in Table 1) and perform the evaluation on the final round of the same competition (n=30 jumpers).

In Table 2 we present the average absolute error scores for predicting the scores of the individual judges (1-5) of the proposed, as well as of the baseline approach (both rounded to the closest half). As there is no other system available to compare with, we compare it against the baseline average predictor, i.e. the predictor that always predicts the average of the scores given in the first round by a particular judge. We can see that our method performs roughly twice as good, while similarly only being based on the data from the first round of the competition.

Table 2: Average absolute errors and standard deviations of the proposed and a baseline approach against real judges (1-5) and the measured consistency scores (CS) of the real and virtual judges (Judge V).

Judge/Predictor	Baseline	**Ours***	CS
Judge 1	0.60 ± 0.35	$\mathbf{0.30 \pm 0.28}$	0.22 ± 0.19
Judge 2	0.66 ± 0.48	$\mathbf{0.27 \pm 0.28}$	0.10 ± 0.13
Judge 3	0.69 ± 0.44	$\mathbf{0.33 \pm 0.32}$	0.17 ± 0.21
Judge 4	0.54 ± 0.44	$\mathbf{0.35 \pm 0.32}$	0.22 ± 0.23
Judge 5	0.64 ± 0.52	$\mathbf{0.32 \pm 0.27}$	0.16 ± 0.17
Judge V*			0.27 ± 0.21
Judge V†			$\mathbf{0.21 \pm 0.20}$

*Only main competition data used for training (n=50)
†Additional training data from Nordic combined discipline (n=50+50)

For obtaining a virtual judge grade (Judge V), we first remove the lowest and the highest predictions out of the predictions of the scores for 5 judges and then use the closest-half rounded average as the grade for the virtual judge - u_v. To evaluate the performance of the virtual judge, we measure the consistency of the virtual judge with other (real) judges. Let it be $\hat{y} = (y_1, y_2, y_3, y_4, y_5)$ the ground-truth judge scores for one particular jump. We first remove the lowest and the highest scores then calculate an average score $\overline{y}$ representing a reference score. For each of the judges, we then calculate the absolute distance to the reference score $d_i = |y_i - \overline{y}|$, which can be thought of as a consistency score with other judges. Then we also calculate the consistency score of the virtual judge $d_v = |u_v - \overline{y}|$. We average the consistency scores for all the real judges and a virtual one across all the jumps. We want the average consistency score of the virtual judge to be similar or smaller, compared to real judges.

The consistency scores (CS) are also presented in Table 2. We can see that the virtual judge (Judge V*) achieved a CS score, which is slightly worse, in comparison with real judges. We then extended the training data of the virtual judge (Judge V*) with the data from the Nordic combined competition, which has much larger variability in terms of style score ranges. We see that the performance significantly improved (Judge V†), slightly overreaching 2 out of 5 real judges. This can be also seen in Figure 8c, where we present the distribution of CS scores in the histogram. We can notice that the performance of Judge V† is consistent with other judges, while Judge 2 clearly represents the most consistent judge.

In Figure 8a we present the predictions for the real-world judges in a confusion matrix, where we can notice that in the majority of the cases, the error is at most half a point. We can also notice, that our score is usually a bit higher in the case of lower ground truth style score and vice-versa, which we attribute to the lack of border cases in our training data. These results improve significantly when we introduce Nordic combined training data in Figure 8b. This is due to the higher inclusion of border case examples, as the average style score in the Nordic combined competition was approximately 1 point lower. This clearly demonstrates that our prototype system clearly benefits from the introduction of additional data.

5. Conclusion

In this work we presented a prototype computer vision system for automatic ski jump style scoring using the TV video footage. We created and annotated the image dataset that was used for training the deep learning model to detect the body and ski parts throughout the video sequence. We proposed a method to utilize the detected trajectories of the body and ski parts locations, along with the detection confidences, to build a model for predicting the ski jump style scores. This presents the first such implementation in the research community and broader.

Despite vaguely and subjectively defined judging rules, we demonstrate that the proposed system successfully learns a model, based solely on time-series data of the body and ski parts locations and existing reference style scores, which performs on par with real-world judges on our experimental data. We showed that the results significantly improve when the training set increases. We experimented with the data in the dataset containing 210 jumps that we were able to collect and annotate. If a large-scale data were available, including several hills, competitions, judges, etc., the reliability of the obtained results would even increase.

The system is built in a modular fashion, which enables future development and integration of newer or better-performing methods. It could also utilize 3D pose estimation methods, or even 3D pose information obtained using a professional dedicated camera setup, placed around the ski jump hill, if available, to make the system more view-point invariant.

The system would also benefit from a more granular scoring information in terms of deduction points per different jump stages, which are currently not publicly available. This would enable learning the models of the individual jump stages, thus increasing the amount of data available and the robustness and accuracy of the system. This would also enable improved explainability of the given score and serve also as a training system for younger generations of ski jumpers.

Acknowledgment

This work was in part supported by the Slovenian Research Agency project J2-9433 and program P2-0214.

Figure 9: Example output detection heatmaps of body and ski parts for different images (columns). Body and ski parts are presented in rows for the same image, wherein in the case of symmetry, the right part of the limb or ski is first presented.

References

[1] Mykhaylo Andriluka, Leonid Pishchulin, Peter Gehler, and Bernt Schiele. 2d human pose estimation: New benchmark and state of the art analysis. In *Proceedings of the IEEE Conference on computer Vision and Pattern Recognition*, pages 3686–3693, 2014.

[2] Heike Brock, Yuji Ohgi, and James Lee. Learning to judge like a human: convolutional networks for classification of ski jumping errors. In *Proceedings of the 2017 ACM international symposium on wearable computers*, pages 106–113, 2017.

[3] Heike Brock, Yuji Ohgi, and Kazuya Seo. Development of an automated motion evaluation system from wearable sensor devices for ski jumping. *Procedia engineering*, 147:694–699, 2016.

[4] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. Openpose: realtime multi-person 2d pose estimation using part affinity fields. *IEEE transactions on pattern analysis and machine intelligence*, 43(1):172–186, 2019.

[5] Sangbum Choi, Seokeon Choi, and Changick Kim. Mobilehumanpose: Toward real-time 3d human pose estimation in mobile devices. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2328–2338, 2021.

[6] Tadej Ciglarič, Karmen Gostiša, Timotej Kovač, Darja Peternel, Manja Pograjc, Nac Stoklas, Blaž Štempelj, Gašper Vodan, Robert Rozman, and Matjaž Kukar. Video meritve dolzin smucarskih skokov. In *Zbornik šestindvajsete mednarodne Elektrotehniške in racunalniške konference ERK*, pages 337–340, 2017.

[7] Yong Du, Yun Fu, and Liang Wang. Skeleton based action recognition with convolutional neural network. In *2015 3rd IAPR Asian Conference on Pattern Recognition (ACPR)*, pages 579–583. IEEE, 2015.

[8] FIS. The International Ski Competition Rules (ICR), Book 3, Ski Jumping (2019). `https://assets.fis-ski.com/image/upload/v1572940099/fis-prod/assets/ICR_Ski_Jumping_2019_clean.pdf`. [Access: October 19th, 2021].

[9] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross Girshick. Mask R-CNN. In *Proceedings of the IEEE international conference on computer vision*, pages 2961–2969, 2017.

[10] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*, 2017.

[11] Sam Johnson and Mark Everingham. Clustered pose and nonlinear appearance models for human pose estimation. In *Proceedings of the British Machine Vision Conference*, 2010. doi:10.5244/C.24.12.

[12] Chao Li, Qiaoyong Zhong, Di Xie, and Shiliang Pu. Skeleton-based action recognition with convolutional neural networks. In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 597–600. IEEE, 2017.

[13] 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. In *European conference on computer vision*, pages 740–755. Springer, 2014.

[14] Grega Logar and Marko Munih. Estimation of joint forces and moments for the in-run and take-off in ski jumping based on measurements with wearable inertial sensors. *Sensors*, 15(5):11258–11276, 2015.

[15] Katja Ludwig, Moritz Einfalt, and Rainer Lienhart. Robust estimation of flight parameters for ski jumpers. In *2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pages 1–6. IEEE, 2020.

[16] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28:91–99, 2015.

[17] Dejan Štepec. Razvoj postopka za avtomatsko ocenjevanje sloga smučarskih skokov. Master's thesis, Univerza v Ljubljani, 2017.

[18] Bastian Wandt, Marco Rudolph, Petrissa Zell, Helge Rhodin, and Bodo Rosenhahn. Canonpose: Self-supervised monocular 3d human pose estimation in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 13294–13304, 2021.

[19] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. Convolutional pose machines. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pages 4724–4732, 2016.

[20] Yi Yang and Deva Ramanan. Articulated human detection with flexible mixtures of parts. *IEEE transactions on pattern analysis and machine intelligence*, 35(12):2878–2890, 2012.

[21] Guangle Yao, Tao Lei, and Jiandan Zhong. A review of convolutional-neural-network-based action recognition. *Pattern Recognition Letters*, 118:14–22, 2019.

[22] Dan Zecha, Christian Eggert, Moritz Einfalt, Stephan Brehm, and Rainer Lienhart. A convolutional sequence to sequence model for multimodal dynamics prediction in ski jumps. In *Proceedings of the 1st International Workshop on Multimedia Content Analysis in Sports*, pages 11–19, 2018.