Automatic gate-to-gate time recognition from audio recordings in alpine slalom skating using neural networks

Friedrich Menhorn[*1], Chris Hummel[2], Andreas Huber[3], Karlheinz Waibel[4], Hans-Joachim Bungartz[1], Peter Spitzenpfeil[2]

[1] Department of Computer Science, Technical University of Munich, Garching, Germany
[2] Department of Applied Sport Science, Technical University of Munich, Munich, Germany
[3] Olympiastützpunkt Bayern, Munich, Germany
[4] German Ski Federation (DSV), Planegg, Germany

[*] menhorn@in.tum.de

ORIGINAL ARTICLE

Submitted: 1 June 2023

Accepted: 3 January 2024

Published: 2 May 2024

Editor-in-Chief:

Claudio R. Nigg, University of Bern, Switzerland

Guest Editors:

Thomas Stöggl, Paris Lodron University Salzburg, Austria; Red Bull Athlete Performance Center, Austria

Hermann Schwameder, Paris Lodron University Salzburg, Austria

Hans-Peter Wiesinger, Paris Lodron University Salzburg, Austria

ABSTRACT

We introduce a novel approach for computing gate-to-gate time automatically from audio recordings. In slalom skiing, gate-to-gate timing is a valuable metric for athletes and trainers, capturing the time elapsed between slalom gates. The availability of these measurements immediately after each run allows for prompt feedback. This study specifically concentrates on gate-to-gate timing in alpine slalom skating, serving as a foundational step towards its future application in slalom skiing.

While existing methods for measuring gate-to-gate time vary in their feasibility, accuracy, and compliance with regulations, we propose a solution utilizing a convolutional neural network (CNN) to predict gate locations using the audio signals generated upon gate contact. By leveraging these predictions, we achieve fully automated computation of gate-to-gate timings.

We conduct a comparative analysis between the CNN's predictions and data obtained from an inertial measurement unit. Our findings reveal a strong predictive correlation between the two methods, with an R-squared value of 0.94 and a root mean squared error of 0.036. The majority of predictions demonstrate high accuracy, falling within a range of thousandths of a second. However, a few outliers negatively impact the overall performance. Notably, we observe no deterioration in predictive quality based on the distance between the camera and the gate.

Finally, we delve into the challenges and limitations associated with our approach and provide a comprehensive discussion. To conclude, we outline potential avenues for future research and extensions of our methodology to the realm of slalom skiing.

Keywords

performing analysis, inline slalom, alpine skiing, slalom, machine learning

Citation:

Menhorn, F., Hummel, C., Huber, A., Waibel, K., Bungartz, H.-J., & Spitzenpfeil, P. (2024). Automatic gate-to-gate time recognition from audio recordings in alpine slalom skating using neural networks. *Current Issues in Sport Science*, *9*(3), Article 003. https://doi.org/10.36950/2024.3ciss003

Introduction

In slalom skiing, as in any other alpine skiing competition, the victor is determined by the fastest recorded time. Athletes as well as trainers are therefore always interested in a detailed breakdown of the timings. Are there certain passages where time was lost or gained? What did the other athlete do differently to be faster?

The current metric used here are section times, which measure timings for consecutive sections of the course. This, however, only offers a coarse granularity with, usually, only a handful of sections for a full run. A more valuable metric in slalom skiing is the gate-to-gate time, which indicates the duration between consecutive slalom gates. This offers a performance analysis with a much finer resolution, than just section times. It is highly desirable for athletes and trainers to have access to these measurements immediately after each run, enabling prompt feedback. Such timely information proves beneficial during training sessions, allowing for the comparison of various runs by the same athlete. Additionally, in competitive scenarios, it facilitates the analysis of results from previous athletes, aiding in the identification of specific sections where time could potentially be gained or lost.

Different strategies to measure gate-to-gate time have been explored. However, they are either infeasible with respect to the amount of work needed, they lack accuracy or rules prohibit the use of required sensors. A common approach is using video analysis. However, this requires the manual collection of gate contacts in the video, which is a time consuming and tedious task. In the work by Swarén et al. (2021), the authors present the average time between gates as gate-to-gate times, i.e., they divided the total runtime by the number of gates. This is, of course, only an approximation and does not offer the accuracy required for detailed feedback on the run. The work by Fasel et al. (2019) used magnetic field sensors in the gloves and magnets in the gates, which showed great accuracy. This approach could be used in training but is rather tedious to set up and rules prohibit its application in competition.

In this work we present a new concept to measure gate-to-gate time in fully automated fashion from audio recordings. In a first explorative study of the approach, we look at gate-to-gate timings in inline slalom skating to reduce challenges that come in slalom skiing. We record runs consisting of 13 gates to create a data set where we train a convolutional neural network to automatically detect the location of gates only on the audio recording. From this prediction of gates, we can compute the gate-to-gate timing.

Neural networks have been well-established in the field of human activity recognition, especially in the field of health care. Murad & Pyun (2017) present an approach using neural networks to recognize a variety of human activity, using long short-term memory networks. Golestani & Moghaddam (2020) combine a system based on magnetic induction to recognize human activity using deep recurrent neural networks. Gupta et al. (2022) give a review about the general application of artificial intelligence in the field and predict that the industry will evolve even further in the future. In other

fields, similar approaches have been employed with respect to audio and vibration signals. As an example, Han et al. (2021) use audio and vibration signals for state monitoring in precision machining to detect tool wearing in the system.

To the best of our knowledge, no previous attempts have been made to employ machine learning techniques for calculating gate-to-gate timing in skiing, particularly by relying exclusively on audio data. Although we initially focus on the less complex scenario of inline skating, this study aims to establish a foundation for the future application of such an approach in alpine skiing.

Methods

We want to find the gate-to-gate timings automatically from audio recordings, which can be extracted from videos. For this, we use a neural network to automatically learn the sound of the pole-gate-contact. For the neural network to perform well, it must be trained on data where we know the timing of the gate contact. As a first step we talk about the data acquisition that we used to create the training data set.

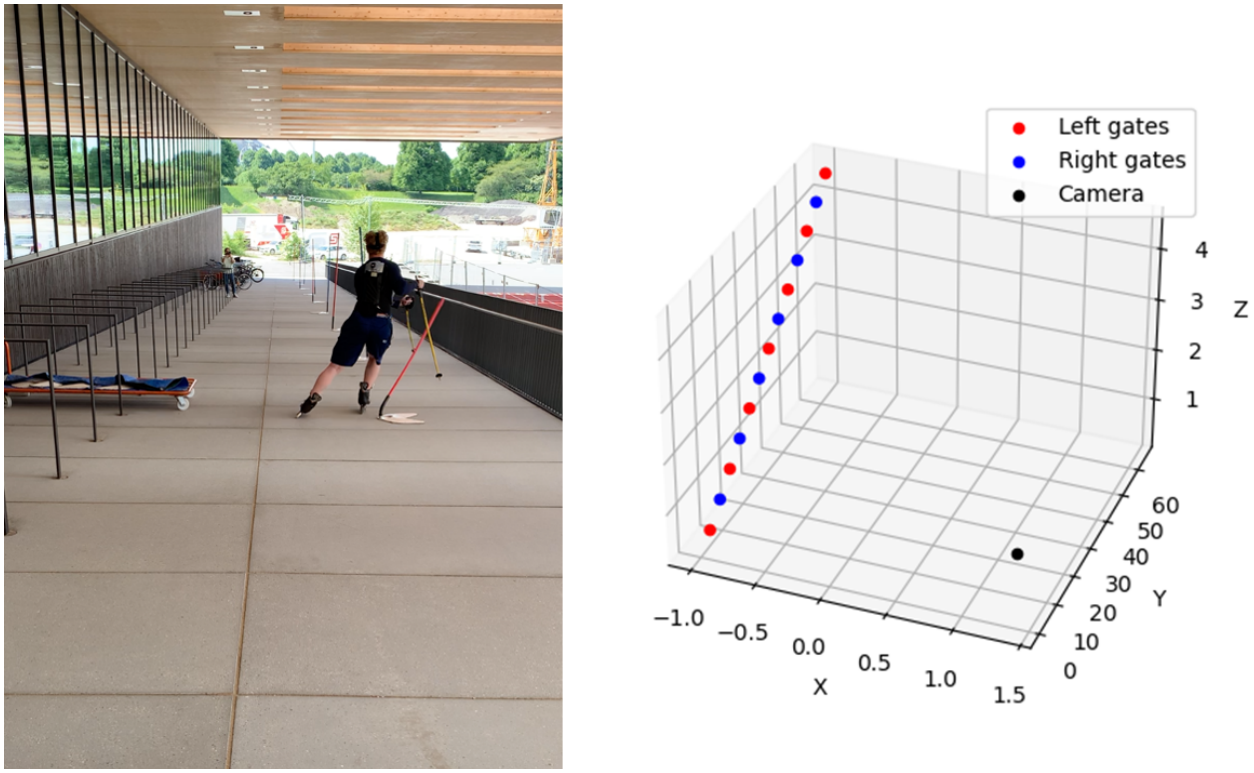

Figure 1 Image of a test run of the slalom course on the ramp at the west side of the sport faculty of TUM. We used 13 gates in total. The gates are five meters apart from each other. The ramp as a slope of 4 degrees. The camera is located 2.4 m to the right of the last gate at the bottom and positioned at a height of 1.6 m.

Data acquisition

We prepared a slalom course on the ramp at the west end of the facilities of the sport faculty of the Technology University of Munich (TUM). An image of the setup in a test run is shown in Figure 1 on the left and a plot of the locations of the 13 gates and camera is given on the right.

One subject drove the course in total 21 times. No additional acceleration was used apart from the con-

stant acceleration due to gravity. We placed six-axis inertial measurement unit (IMU, 1000 Hz, ± 16 g) sensors at the inside of the skiing poles on each hand to have a reference for the contact with the gate. The data logger and battery were fixed at the back using a back protector. For each run, we get seven contacts of the right hand with the gates during the left turns and six contacts of the left hand with the gates during the right turns, respectively.

For recording of the audio, we used two devices. The first was a Sony FDR-AX700. The second was a regular iPhone XR (iOS-Version 16.3.1). Both cameras where static and not moved during recordings. Since we are only interested in the audio recordings, it was not required to, e.g., ensure that the subject is perfectly visible in the frame.

Data Processing

Using two cameras for 21 runs, this results in a total of 42 recordings, each with a runtime of about 20 seconds up to 1 min. (depending on the start of the recording). To extract the audio from the recorded videos, we use FFmpeg (https://ffmpeg.org/), a software to extract and process audio (and videos). We convert the videos to audio with a sample rate of 48000 Hz.

After creating the audio files, we needed to manually label the data. We used a variety of scripts to visualize the audio files at different levels of magnification and computed gradient information since we looked for a steep increase in amplitude. Those gradients were computed using finite differences. A threshold on the amplitude was not used since we saw a varying amplitude with distance to the camera.

Based on this additional gradient information and magnified visualization, we marked the gate contact as the beginning of the clearly visible audio peaks in the data. We visualized one of the runs with 13 gate contacts in Figure 2 on the left. Although this can be misleading due to the line width in the figure, we labeled a single frame for the gate contact. In the figure, also note that we do not only see the first audio peak due to the pole hitting the gate, but also a second peak when the gate hits the ground.

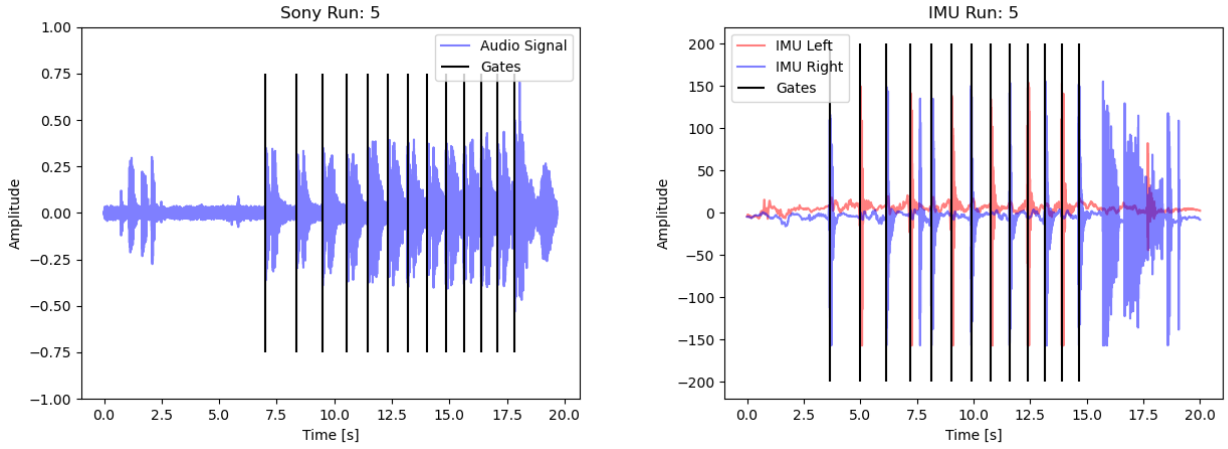

Figure 2 Labeled audio data where the gate contact is visualized as black vertical line on the left. Labeled IMU data on the right with gate contacts in black. The red and blue signals present the left and right IMU sensor, respectively.

We proceeded similarly for the IMU data, which is shown in Figure 2 on the right. Here, we see recordings from the left and right sensor for the acceleration in x-direction which we extracted and plot on top of each other. We used the acceleration in x-direction as indicator of movement in the pole and, thus, a contact

with the gate. Here, the x-direction is defined with respect to the standing pole as the horizontal direction, orthogonal to the direction of motion.

Neural network architecture

Before we dive into the details of the neural network architecture, we briefly summarize the idea, also visualized in Figure 3. We used a sliding window approach, where we used short snippets of the full audio as input for the network. We did not process these snippets in their original form, but transform them into their frequency domain, using a spectrogram. The neural network had then to classify if the window contains a gate contact or not. Hence, we used a neural network for classification where the output is a probability between 0 and 1 where 1 stands for a gate and 0 stands for no gate in the current window.

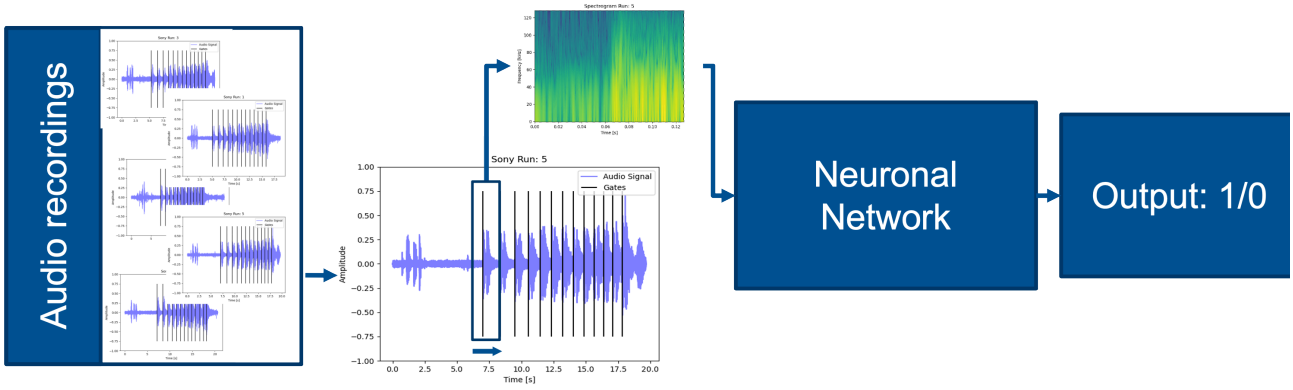

Figure 3 Neural network pipeline.

The advantage of using a spectrogram compared to the audio signal is due to its dimensionality. While the audio signal is only one dimensional, the spectrogram offers a two-dimensional visualization of the data. We can, therefore, rely on convolutional neural networks for image recognition, which are well-established in the field.

We selected a window size of 480 frames, which corresponds to 1/100 of the audio's sample rate. We aimed to strike a balance in choosing this size, as a larger window would limit the maximum accuracy of our approach. Conversely, opting for a smaller window size would create difficulties for the network in identifying gates effectively.

To construct the spectrogram, we divided the entire window into frames and performed a Fast Fourier Transform (FFT) on each frame. We utilized a frame length of 10 with a step size of 1 and employed 256 terms in the FFT calculation. Consequently, the resulting spectrogram image assumed a shape of (471, 129).

For the network architecture, we employed a combination of two-dimensional convolutional layers, pooling layers, and dense layers. Dropout layers were included to enhance performance and mitigate overfitting. In the convolutional layers, we utilized classical Rectified Linear Unit (ReLu) activation functions, while the final output layer employed the sigmoid function.

To optimize the network, we set the learning rate to 0.0001 and utilized the Adam's optimization method. During training, we employed binary cross-entropy as the loss function and measured overall performance using binary accuracy and false negatives as metrics.

The complete design of the neural network, along with the layer settings, is depicted in Figure 4. The network architecture was determined through an optimization process involving various architectures. The implementation of the network was carried out in Python (version 3.9.16), utilizing TensorFlow and Keras (both version 2.11.0).

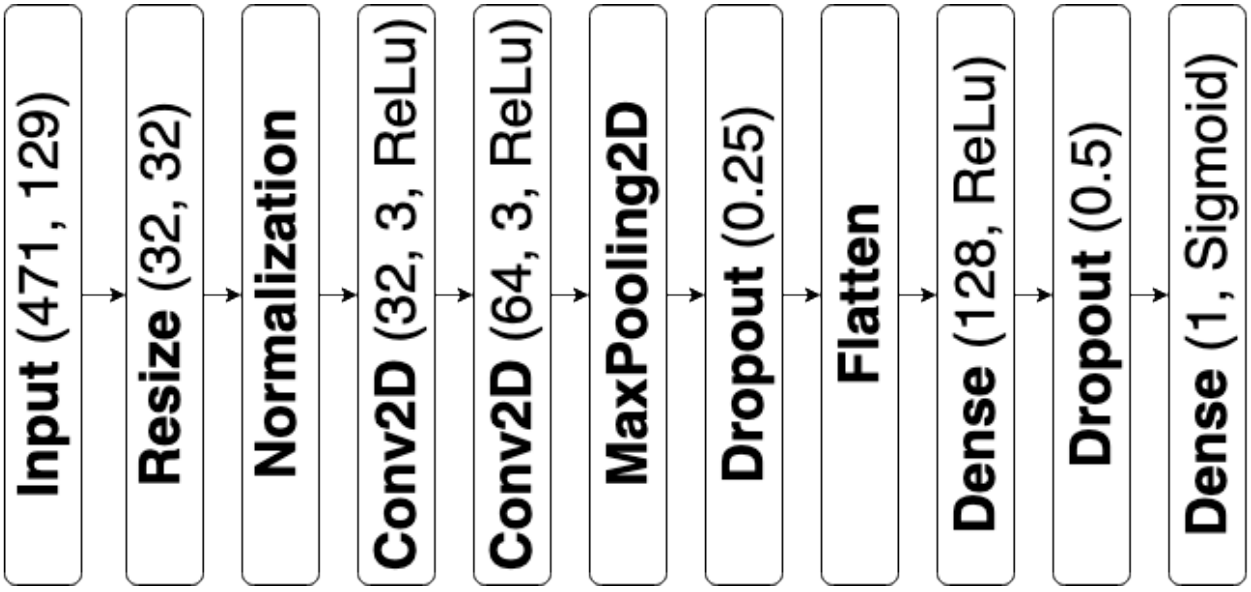

Figure 4 Neural network architecture. The numbers in brackets show the different settings for the individual layers following the nomenclature in TensorFlow.

Neural network training

For the training, we had in total 42 runs available (two cameras, 21 recordings per camera). The runs vary in length, but each contain the 13 gate contacts. This results in 546 gate contacts. Since we used a sliding window approach with a window size of 480 frames, we got 480 windows, that, theoretically, contain the gate contact for each gate. However, we used a step size of 5 frames between windows, which reduces the number to 96 windows per gate. We further reduced the amount since we only allow windows where the gate is located a maximum of 100 frames away from the centre of the window. This has the intention to have a clear separation to windows containing no gate. This reduced the number of windows per gate to 41. Considering these adjustments, the total count of windows containing a gate amounted to 22.386.

Of course, our training set also needs data containing no gate. For this, we picked random windows without a gate from the data set such that we got 100 times the number of windows with gate, in total 2,238,600 windows. The rationale behind this choice was to expose the neural network to a greater number of windows without gate contact, mirroring the distribution in real-world data. Accounting for both windows with and without gate contact, the complete training set comprised 2,260,986 windows. For the training, we used a batch size of 128 windows per batch.

Postprocessing neural network results

For a given audio signal, we wanted to compute the gate-to-gate times for the full signal, whereas the neural network expects an input of only 480 frames, which is transformed to a spectrogram. As we stated, the neural network is a classifier that tells us if the network thinks that the current window contains a gate. In this section, we discuss how we computed the gate-to-gate times from the neural network output.

We again used a sliding window approach. We subdivided the full audio signal into windows of length 480 and propagated each of the windows through the network. For each window, we received a classification from the neural network in form of a probability in [0, 1] for the location of a gate in the given window. For the distance between windows, we picked a step size of 100. This step size still showed good results while reducing the computational load.

We collected the results by summing up the probabilities of each window for its specific location in the full audio signal. In the end, we expected to see the highest values at locations where the neural network classifies a gate. We visualize this pipeline for an example in Figure 5.

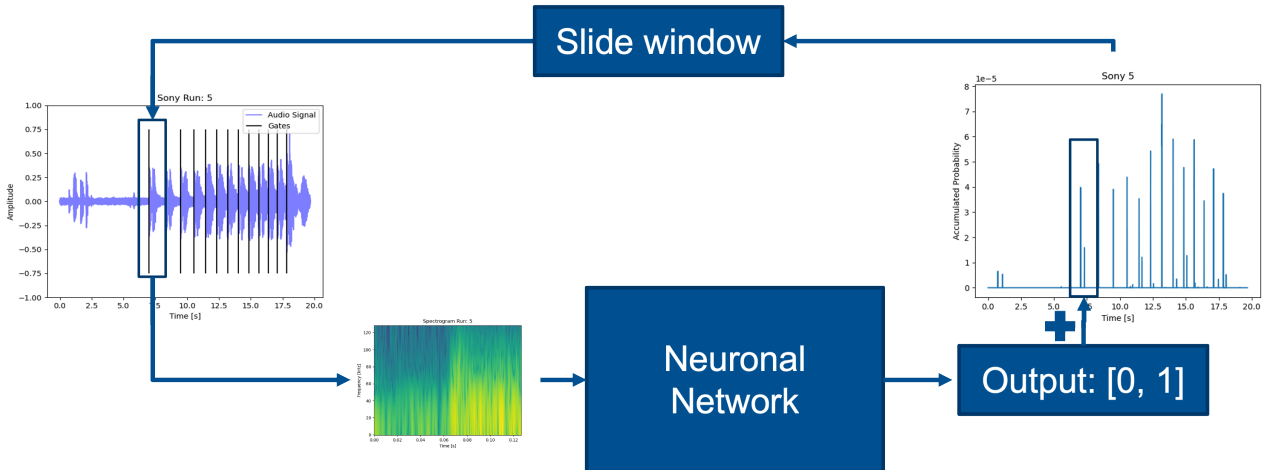

Figure 5 Neural network postprocessing to collect the estimated probabilities for containing a gate for each window.

Gate-to-gate time computation

We see an example of the resulting estimate for the location of gates in Figure 5 on the right. To find the gates, we located the maxima of the result. We observed that we also get smaller peaks for the impact of the gate with the ground. Hence, we also prescribed that we need at least 0.25 seconds between each peak, which is a reasonable difference, also in professional skiing (apart maybe from double gates, which we talk about in the discussion). Furthermore, we prescribed the exact number of peaks that we needed to find since we know the number of gates in the course.

After finding the peaks, we were able to compute the time of each peak. Here, we took the distance of the gate to the camera into account, correcting for the speed of sound. In our case, we used a speed of 340.2 m/s, which corresponds to conditions at 15°C. Finally, we computed the gate-to-gate time by taking the difference between the times of successive gates.

Results

We first discuss the results for the neuronal network training before we get to the gate-to-gate results and validation.

Neural network training

Figure 6 illustrates the training and validation loss and accuracy curves. It is evident that the neural network exhibits impressive performance right from the beginning. As the training progresses over multiple epochs, it continues to refine its results by reducing the loss and improving accuracy in the training data.

When examining the validation results, which represent data that the neural network has not encountered during training, we observe some fluctuations in both loss and accuracy. Nonetheless, there is a trend of decreasing loss and increasing accuracy as the model becomes more adept at generalizing to unseen data. Notably, the best validation results are achieved after seven epochs.

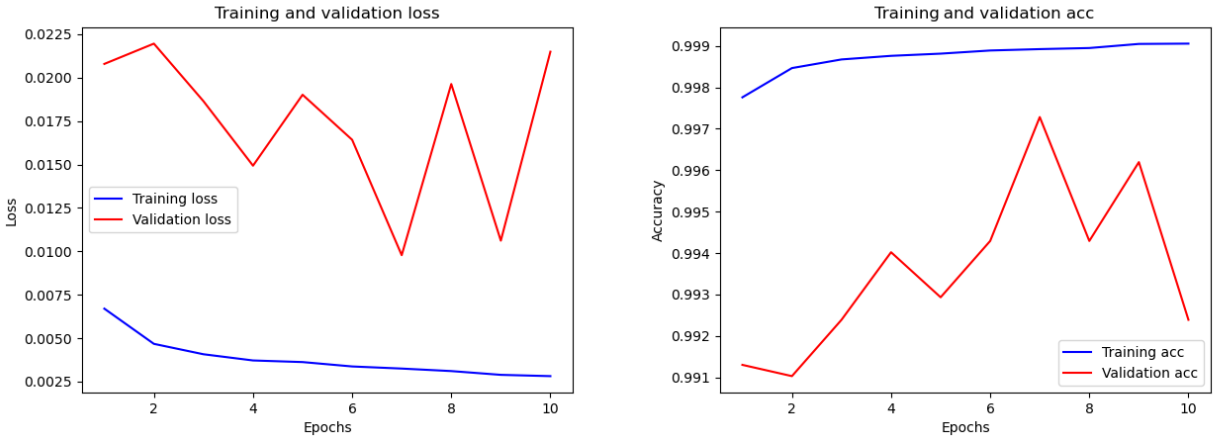

Figure 6 Performance of the neural network during training using different metrics over 10 epochs. Training and validation loss is given on the left. Training and validation accuracy is plotted on the right.

Gate-to-gate results

After having successfully trained the model, we were able to proceed to the gate-to-gate timing evaluation. For the computation of the results, we computed gate-to-gate timings for all 42 recordings, with twelve timings (for 13 gates) per recording. As reference, we used the IMU measurements, where we also computed the gate-to-gate differences visible in peaks in the acceleration in x-direction.

We plotted the gate-to-gate timings computed for each recording compared to the respective gate-to-gate timing from the IMU measurements in Figure 7. We additionally plotted the regression line. *R*-squared and the root mean squared error (RMSE) were given in the title of the figure. The RMSE is computed as $RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}\left(t_{IMU,i} - t_{NN,i}\right)^2}$, where $t_{IMU,i}$ and $t_{NN,i}$ describe the gate-to-gate time for the IMU and the neural network prediction for gate i, respectively.

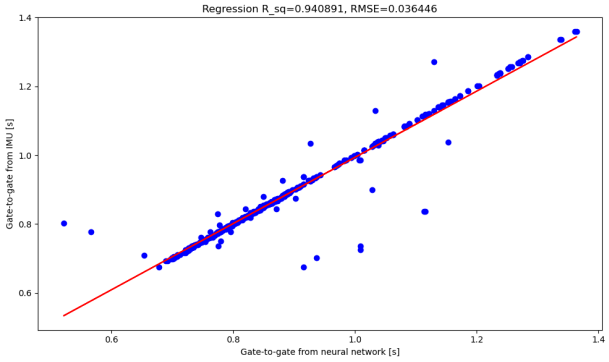

Figure 7 Gate-to-gate timing comparison with the prediction by the neural network on the x-axis and the prediction by the IMU measurement on the y-axis. We further show the linear regression line in red. The regression coefficient, *R*-squared, and the RMSE is given in the title.

We plotted the difference for pair of gates in Figure 8, where we visualize the absolute error of gate-to-gate timings between the IMU and the neural network in the form of a box plot. Here, the label $i - j$ refers to the gate-to-gate time from gate i to gate j. The plot further shows the RMSE for each pair of gates above the median and presents outliers, where we collect outliers larger than 0.01 in single data points for better presentation.

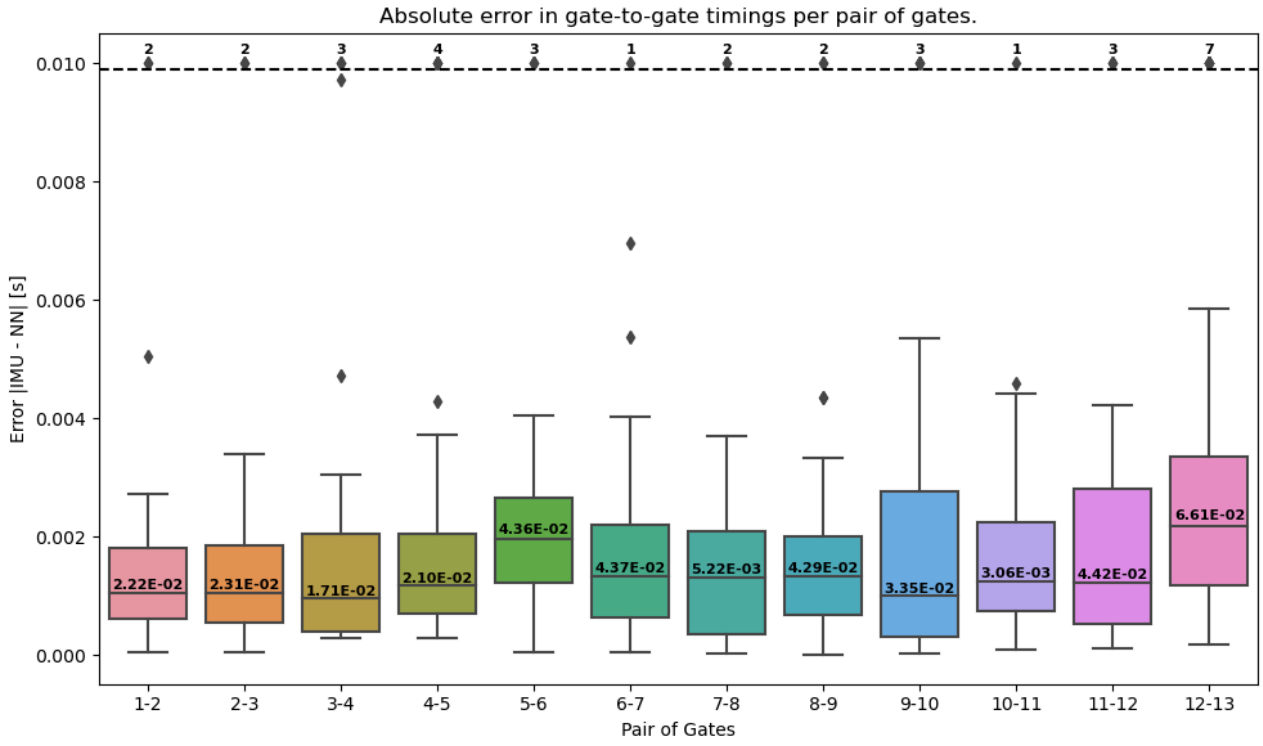

Figure 8 Boxplot showing the absolute error in the gate-to-gate time between neural network prediction (NN) and IMU measurement (IMU) for each pair of gates. The RMSE for each pair of gates is shown above the median. We collect all outliers equal or larger then 0.01 in a single data point. The number indicates the number of outliers above that line.

We also visualize this difference by comparing the performance of both forms of measurements in a Bland-Altman plot in Figure 9. Again, the label $i - j$ refers to the gate-to-gate time from gate i to gate j, which we sort by using the same colors as for Figure 8 and different marker styles. We further show the mean value of the difference and the 95% confidence interval.

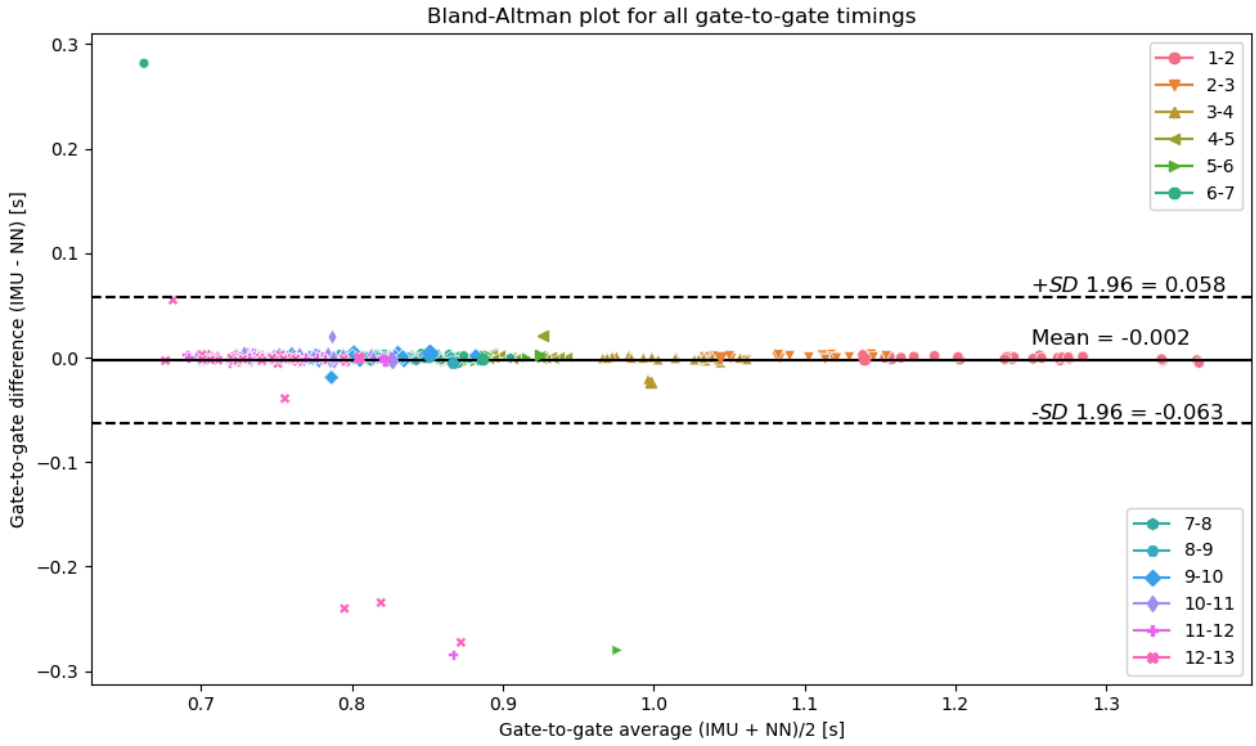

Figure 9 Bland-Altman plot for all gate-to-gate timings comparing the IMU measurement (IMU) with the neural network prediction (NN). The x-axis shows the average measurement while the y-axis shows the difference. The coloring and marker style differentiate the different pairs of gates. The mean and 95% confidence interval is visualized with horizontal lines and their given values.

Discussion

Figure 7 demonstrates a high level of prediction quality exhibited by the neural network when compared to the measurements obtained from the inertial measurement unit (IMU). The *R*-squared value of 0.940891 further reinforces our confidence in the effectiveness and overall performance of the approach. This finding suggests that the approach is viable and can be employed successfully.

Furthermore, the RMSE of 0.03661 provides additional encouragement, highlighting the potential of the approach. Given that the RMSE can be interpreted as a measure of accuracy, it serves as a significant metric for enhancing the prediction quality in skiing. While an accuracy in the range of hundredths of a second is advantageous, our aim is to achieve accuracy in the thousandths of a second, which would greatly enhance the precision of the predictions in the context of skiing.

In Figure 8, we observe that our main set of predictions indeed falls within the range of a thousandth of a second when examining the error for individual pairs of gates. This level of precision is highly encouraging and demonstrates the accuracy of our approach. Given that the IMU sensor has a sample rate of 1000 Hz, we might even reach the limits of measurable accuracy. Additionally, we note that there is no significant variation in the median gate-to-gate time throughout the course of the run. This finding is particularly promising considering that the first gate was located approximately 65 meters away from the camera.

The Bland-Altman plot in Figure 9 further solidifies the performance of the neural network where the mean difference to the IMU measurement is in the order

of milliseconds. Note here, that we clearly see the reduced time between gates for later gates since the speed of the skater increases. For later gates, we observe an increase in outliers. This might be due to wrong identification of peaks or more noise since the distance to the camera decreases. Nevertheless, we observe a good match between both methods.

The increase in outliers explains the slight increase in variance towards the end of the run, our approach demonstrates robustness in relation to distance. This flexibility in camera positioning is advantageous, especially in scenarios where specific locations might be constrained during training or competition. However, it is important to acknowledge the presence of outliers, which contribute to the overall increase in the RMSE. To address this issue, we intend to further refine our algorithm in future work, aiming to minimize the occurrence of such outliers and enhance the overall performance of our system.

Transitioning from inline skating to slalom skiing presents numerous challenges that must be addressed. Firstly, the presence of low temperatures and adverse weather conditions can adversely affect audio signals, potentially hindering the accuracy of our measurements. Moreover, ambient noise becomes a significant challenge, particularly noise generated by skis on snow, the presence of other athletes during training sessions, and the noise from crowds in competitive events. These factors can overshadow the crucial sound of gate contact, making it more difficult to extract accurate information. Additionally, the athlete usually hits the gate with his hand but occasionally might use his other body parts or not hit the gate at all. This is a factor we cannot control. Furthermore, slalom skiing necessitates a high level of predictive accuracy due to the short time scales involved. If we consider short gate distances, e.g., for double gates, we need to be able to capture even the smallest time differentials. Finally, to cover a full race might require multiple microphones (or cameras) along the course or moving cameras might be used. This necessitates to correct for different cameras. However, this is a postprocessing step and does not change the method, detecting gate contacts from audio recording, itself. Additionally, we do not need to correct for the speed of sound if we compare different runs of the same athlete or in between athletes since the delay is constant over different runs.

Conclusion

Despite these challenges, our approach demonstrates promising potential in accurately predicting gate-to-gate time. In future work, we aim to further enhance the performance of our system by exploring alternative neural network architectures and refining our training set.

One avenue for improvement is to investigate the use of long-short-term memory networks (LSTM) in our framework. LSTM networks have gained significant attention in signal processing applications, as demonstrated in studies such as the work by Murad & Pyun (2017). These networks aim to imitate the short-term memory of the brain to enhance the predictive quality of the network. Exploring the integration of LSTM architectures could improve the accuracy of our predictions.

Additionally, we intend to explore more complex image recognition networks, such as EfficientNet (Tan & Le, 2021) or CoAtNet (Dai et al., 2021), which have shown promising performance in various image-related tasks. Integrating these advanced network architectures could enhance the precision and robustness of our gate detection and timing predictions.

In terms of improving the training set, we aim to investigate the approach suggested in Fasel et al. (2019), which utilizes a magnet-based timing system for data labeling. This approach could provide more accurate and precise labeling, thereby enhancing the quality of our training data and ultimately improving the overall performance of our system.

While this work is a first prototype, exploring the validity of using techniques from machine learning for gate-to-gate time prediction, we showed that the approach has great potential and we lay the groundwork for future explorations on this topic. The final goal is to

give trainers and athletes immediate feedback after a run, just from video recordings. This has the potential to greatly improve the quality of feedback in training and competition, which makes this work quite valuable.

References

Dai, Z., Liu, H., Le, Q. V., & Tan, M. (2021). CoAtNet: Marrying convolution and attention for all data sizes. *Advances in Neural Information Processing Systems*, *34*, 3965–3977. https://proceedings.neurips.cc/paper_files/paper/2021/file/20568692db622456cc42a2e853ca21f8-Paper.pdf

Fasel, B., Spörri, J., Kröll, J., Müller, E., & Aminian, K. (2019). A magnet-based timing system to detect gate crossings in alpine ski racing. *Sensors*, *19*(4), Article 940. https://doi.org/10.3390/s19040940

Golestani, N., & Moghaddam, M. (2020). Human activity recognition using magnetic induction-based motion signals and deep recurrent neural networks. *Nature Commununications*, *11*, Article 1551. https://doi.org/10.1038/s41467-020-15086-2

Gupta, N., Gupta, S. K., Pathak, R. K., Jain, V., Rashidi, P., & Suri, J. S. (2022). Human activity recognition in artificial intelligence framework: A narrative review. *Artificial Intelligence Review*, *55*(6), 4755–4808. https://doi.org/10.1007/s10462-021-10116-x

Han, S., Mannan, N., Stein, D. C., Pattipati, K. R., & Bollas, G. M. (2021). Classification and regression models of audio and vibration signals for machine state monitoring in precision machining systems. *Journal of Manufacturing Systems*, *61*, 45–53. https://doi.org/10.1016/j.jmsy.2021.08.004

Murad, A., & Pyun, J.-Y. (2017). Deep recurrent neural networks for human activity recognition. *Sensors*, *17*(11), Article 2556. https://doi.org/10.3390/s17112556

Swarén, M., Gallagher, C., & Björklund, G. (2021). Impact on ski regulation changes on race and gate-to-gate times in world cup giant slalom skiing 2005-2020. *Research & Investigations in Sports Medicine*, *7*(5), 668–673. https://doi.org/10.31031/RISM.2021.07.000672

Tan, M., & Le, Q. (2021). EfficientNetV2: Smaller models and faster training. *Proceedings of Machine Learning Research*, *139*, 10096–10106. https://proceedings.mlr.press/v139/tan21a.html

Acknnowledgements

Funding

The authors thank the German Ski Federation (DSV) for their support

Competing interests

The author/s has/have declared that no competing interests exist.

Data availability statement

All relevant data are within the paper.