

End-to-End Move Prediction System for Indoor Rock Climbing: Beta Caller

Kevin Cardenas
University of Colorado
Colorado Springs
1420 Austin Bluffs Pkwy
Colorado Springs, CO 80907
kcarden4@uccs.edu

Sudhanshu Semwal
University of Colorado
Colorado Springs
1420 Austin Bluffs Pkwy
Colorado Springs, CO 80907
ssemwal@uccs.edu

James Maher
United States Air Force
Academy
2345 Fairchild Dr
USAFA, CO 80840
james.maher@afacademy.af.edu

ABSTRACT

We developed Beta Caller, an end-to-end system supporting the sport of rock climbing for climbers with visual impairment. Beta Caller provides real-time, audible instructions containing a prediction for the climber's next move while they are actively climbing a rock wall. This system leverages computer vision techniques to collect key information about the climber's environment, enabling Beta Caller to make move predictions on climbing walls it has never encountered before. Neural networks are used to predict where the climber should move next, based on information provided by the computer vision models. The predicted move is translated into a verbal message guiding the climber to the next hold and then transmitted via wireless headphones using a text-to-speech model. This novel idea makes one of the fastest growing sports in the world even more appealing and approachable to climbers with visual impairment, however, this tool can be utilized by all climbers to improve their climbing skills. Beta Caller achieved 80.08% accuracy predicting which limb the climber should move next and, when predicting the location of the next hold, Beta Caller achieved a bounding box error of only 6.79%. These results pioneer a strong foundation shaping the future landscape of rock climbing prediction tools for visually impaired climbers.

Keywords

artificial intelligence, computer vision, human computer interaction, object detection, pose estimation, human motion prediction, rock climbing

1 INTRODUCTION

Rock climbing is one of the fastest growing sports in the world, exploding in popularity in the last few years. With inspiring and alluring documentaries like *The Dawn Wall* [Low17a], *Free Solo* [Vas18a], *Valley Uprising* [Mor14a], and many more, coupled with the sport's Olympic debut in the 2020 Summer Olympics, rock climbing grew exponentially in the last few years and captivated the interest of millions of people worldwide. Even before these catalyzing events, the sport of rock climbing garnered interest across the globe due to the sport's ability to concurrently challenge participants physically and mentally. The sport initially interested a niche group of outdoorsy, free-spirited people, but now captivates the attention of a wide-spread, diverse audience. For those less familiar with this sport, definitions

of several rock climbing terms whose meaning may not be innately understood are provided in Table 1.

Hold	3D plastic grip for a climber to grab with their hand or step on with their foot
Move	Transition of hand or foot to the next hold
Route	A sequence of moves, typically using a single hold color, to get to the top of the wall
Beta	Information about how to complete a route
Caller	Someone providing instructions to a climber

Table 1: Rock Climbing Terminology

A large population of athletes with varying disabilities competitively participate in sports, and rock climbing is no exception. Climbers with disabilities have found many innovative and unique ways to push and pull their body up a wall to get to the top. There have been many technological advancements in prosthetics and various rock climbing gear to make the sport more approachable, inclusive and enjoyable to a wider scope of abilities. However, technology assisting visually impaired climbers remains grossly under-researched. The most common and "advanced" way for a visually impaired climber to participate in the sport is to have someone, who is typically a specially trained coach, on the ground verbally coaching the climber (known as the "caller")

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

by providing them information (commonly referred to as “beta”) about how to ascend the wall. In addition to hearing the commands for the next move, visually impaired climbers use their palms to scan the wall for rock holds. This current approach is exceedingly challenging, time consuming and may be intimidating to a new climber and to the caller who is coaching the climber up the wall. The caller’s task is arguably the most important and requires significant climbing experience, spatial estimation skills and quick thinking in order to process the climber’s body position, what holds are available to them, where they should move next, and finally call out a command to guide the climber to the next hold.

This paper introduces the application of computer vision and mobile media technology to a revolutionary sensory substitution device, Beta Caller, that leverages Artificial Intelligence (AI) to aid rock climbers of all abilities to better enjoy and compete in the sport. Beta Caller combines two computer vision models to understand key information about the climber’s environment (where holds are located on the wall and the climber’s body pose), as well as a combination of neural networks to provide a real-time prediction of the climber’s next move. Beta Caller uses a text-to-speech model to transmit an audible message containing coaching information in a timely and useful manner to enable the climber to quickly and successfully complete the next move and ascend to the top of the wall. In addition to this novel system, our paper contributes a rock climbing movement dataset that, to the best of our knowledge, stands as the sole resource of its kind.

2 RELATED WORK

Despite the recent interest in rock climbing, research within the sport is fairly limited in quantity, depth and scope. A small amount of rock climbing research focuses on better understanding how rock climbers move their bodies [Pfe11a][Ouc10a][Sib07a][Wei14a], and additional research describes a handful of real-time climbing tools to make the sport more enjoyable by helping climbers efficiently move through a route [Kos17a][Kos17b]. A majority of climbing research, however, aims to provide performance feedback to the climber after their workout session concludes in order to evaluate how well they trained instead of providing feedback while the climber is on the rock wall [Lad13a][Bre23a][Kos15a].

One example of a performance feedback tool was created by Ekaireb et al. to assess video footage of a climber using computer vision and machine learning to provide a report of how well the person climbed a route [Eka22a]. Their research conducts image segmentation creating a mask of the raw image only containing the climbing wall and excluding the background. Additionally, information about the climber’s environment is

gathered using an object detection model to locate all of the holds on the climbing wall and then a pose estimation model identifies all of the climber’s joint keypoints. This information is collected for each video frame and, after the climb is completed, these frames are evaluated to provide the climber with feedback about how well they climbed the route. Beta Caller improves the system created by Ekaireb et al. by collecting information about the climber’s environment and training a model to provide climbers with real-time movement predictions on any indoor rock climbing wall.

2.1 Assistive Technology for Visually Impaired Rock Climbers

Richardson et al. developed a real-time, assistive climbing system to guide visually impaired climbers [Ric22a]. Climb-o-Vision uses a helmet-mounted, computer vision system that identifies where holds are located on the climbing wall and provides tactile feedback on the climber’s tongue to guide them to a hold. This system is able to identify holds only where the climber’s helmet is facing and does not provide guidance to the climber where the next best hold is located. Electrotactile tongue interfaces have advanced the capabilities of vision substitution, however this research did not provide any conclusions evaluating the effectiveness of the tongue interface during climbing to guide a climber’s hand to holds.

Ramsay and Chang [Ram20a] developed a real-time climbing tool to assist visually impaired climbers as they are actively climbing using a body pose sonification system. They developed a tool to provide the climber with an auditory command containing the location of the next hold on the wall. Additionally, the tool produces a tone for the climber that changes pitch based on the distance from the climber’s hand to the next hold. However, this tool is limited to a specific, standardized climbing wall called a MoonBoard. A MoonBoard is a small (7.5ft wide by 10.5ft tall) climbing wall used with the sole purpose of training. This board is filled with evenly-spaced holds each with an LED light. When lit, the LED lights identify the holds the climber is allowed to use creating the intended route. The primary benefit of this board is that there are over 6,000 pre-programmed routes in a small space. The MoonBoard was created to provide advanced rock climbers with a training tool. The easiest route on a MoonBoard is significantly more challenging than the most routes in an indoor rock climbing gym, therefore it is not intended for use by a majority of the climbing community.

Ramsay and Chang’s research [Ram20a] capitalized on the known, uniform placement of holds on the MoonBoard, eliminating the need for object detection to identify the location of holds on the rock climbing wall. Additionally, the use of a MoonBoard automatically provides known move sequences, so there is no need to

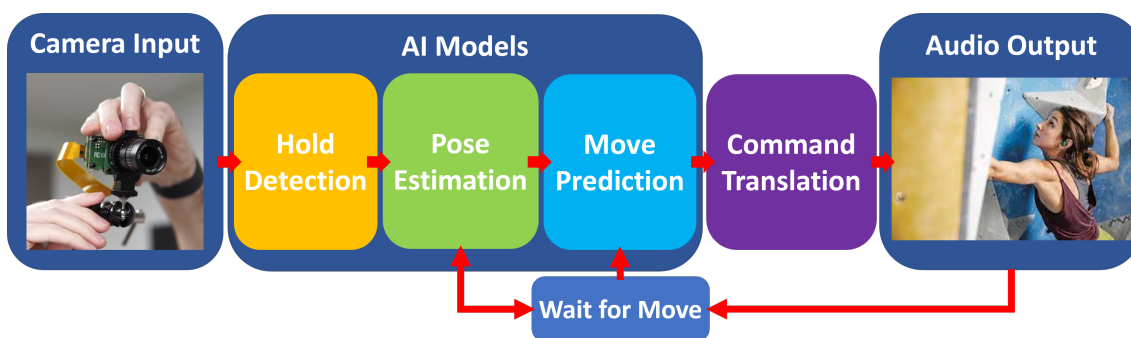


Figure 1: Beta Caller Architecture

create a move prediction model to intelligently predict where the climber should move next or which limb they should move. Ramsay and Chang solely focused on using a pose estimation model to identify the climber’s joint keypoints in order to provide the climber with an auditory command of how to move to the next known hold. Their results were successful when demonstrated on a vertical Twister game mat representing a simulated climbing wall. Beta Caller significantly augments this work outside of the confines of the MoonBoard by generalizing the nature of the system to work on any indoor rock climbing wall. This requires Beta Caller to be able to accurately identify hold locations in addition to the climber’s pose, as well as predict the next best hold for the climber to move to and which hand should move to that hold. To the best of our knowledge, there does not exist any research into real-time systems that predict where a climber should move next.

3 METHODOLOGY

Beta Caller gathers information about the climber’s environment and provides real-time, audible commands containing a prediction for the climber’s next move while they are actively climbing. In practice, the proposed system allows a climber to point a camera at the climbing wall to guide them to the top by providing a series of commands to accomplish each next move.

3.1 System Architecture

Beta Caller comprises three primary components: video camera input, a suite of AI models, and a text-to-speech model providing an audible command to the climber via wireless headphones (Figure 1).

Beta Caller runs on a laptop which is stationed on the ground with a camera facing the climber on the climbing wall. The first two models utilize computer vision techniques to gather key information about the climber’s environment. The first model employs object detection to identify where the holds are located on the climbing wall. The second model uses pose estimation to locate the climber’s pose by identifying joint keypoints on the climber’s body. The last model, the core

of Beta Caller, uses information about the climber’s environment to predict where the climber should move next. Specifically, a combination of neural networks were built to predict which limb the climber should move and, if either hand is predicted to move next, to which hold the climber should move that limb.

After a move is predicted, simple trigonometry is used to calculate the direction the climber should move their hand to the next hold. Additionally, the distance between the climber’s wrist and the next hold is calculated. Using this information, Beta Caller calls out which hand the climber should move, the angle which the climber should move their hand, and the distance from their wrist to the predicted hold (e.g. “Right hand. Two o’clock. About two feet.”). This command is played using a text-to-speech engine, *pyttsx3*, and enables the climber to listen to the output using wireless headphones connected to the laptop. Once the climber receives this command, Beta Caller waits for the climber to complete the move by continuously tracking the climber’s pose. Once the move is completed, Beta Caller makes another prediction where the climber should move to and transmits that command to the climber. This process is repeated until the climber finishes the entire route.

3.2 Data Pipeline

To the best of our knowledge, there does not exist a rock climbing dataset suitable for predicting climbing movement, so we constructed a dataset with over 4,100 images collected as image sequences from over 250 videos where each frame is saved after a completed move. The images are named according to the video they correspond to followed by the time sequence (e.g. 01-01.jpg). These images are sent through a data pipeline to gather features and labels to later be used to train the move prediction model.

The first step in the data pipeline is to run inference on each image using two computer vision models to gather key information about the climbing wall and the climber’s body position. Beta Caller uses one of the leading and most widely used object detection models, YOLOv8 [Red16a], to train a custom object detection

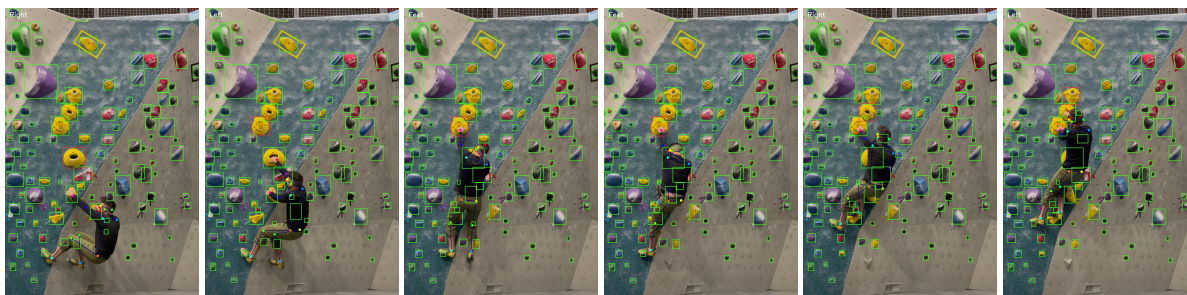


Figure 2: Results of Data Pipeline for Image Labeling

model to identify indoor rock climbing holds. In addition to object detection, Beta Caller utilizes the state of the art pose estimation model, ViTPose [Xu22a]. ViTPose is a vision transformer model that has an encoder to extract image features and a lightweight decoder to conduct pose estimation. Beta Caller uses ViTPose to get the x - y coordinates of 17 joint keypoints from the climber's body. The data collected from each of these computer vision models is illustrated in Figure 3.



Figure 3: Data Output from Computer Vision Models

Hold predictions can be drawn with bounding boxes around each hold found in the image and joint keypoints are drawn as dots accompanied by lines connecting the keypoints, providing a stick figure of the climber. The illustrations of the computer vision models' results shown in Figure 4 are used solely for demonstrative and validation purposes because Beta Caller will not need to visually display the holds and pose, rather use that information to provide an audible command to the climber.



Figure 4: Computer Vision Models' Results

After running inference on an image, the output data is added to a CSV file. The data output from these models are two data structures containing normalized x - y coordinates of the holds within the image and the climber's joint keypoints. The holds data structure contains a row for each hold detected and four columns representing the bounding box coordinates ($x_{min}, y_{min}, x_{max}, y_{max}$). The pose data structure contains a row for each of the 17 joint keypoints and two columns representing the keypoint's x and y coordinates.

The data pipeline uses these two data structures for both the current image and the previous image from the image sequence to compute and identify the labels for the previous image. There are five labels in the dataset: which limb moved and the bounding box coordinates of the hold the climber moved their limb to. Beta Caller combines both feet together into a single class because the primary focus of calling moves to a climber is to direct them to the best hand holds. Focusing on accurate hand predictions provides the climber more time, using less energy, to find the best holds for their feet to prepare them for the next hand movement. Additionally, providing specific guidance for four limbs instead of two might easily become overwhelming for a climber. Previous experience climbing with people with visual impairment confirms the necessity of predicting the correct hand to move as well as when the climber is required to move their feet up to prepare for the next hand movement. Therefore, the limb label will be assigned the value left hand, right hand or feet.

In order to validate the data pipeline, labels are drawn on each image. The top left corner of the image contains text for which limb will move and a red bounding box is drawn around which hold the climber will move their limb to. For example, the first image in the Figure 2 sequence is labeled indicating the climber will move their right hand to the red box because this outcome is observed in the subsequent image from the sequence.

The final dataset prepared for the move prediction models is in CSV format containing 34 features, representing the normalized x - y coordinates for each of the 17 joint keypoints, and 5 labels ($limb, x_{min}, y_{min}, x_{max}, y_{max}$) for over 4,100 images. The dataset was split into two subsets: training and testing. Specifically, 90% of the

data was used for training the model and 10% was reserved for testing its performance. During training, 30% of the training data was used for validation.

3.3 Move Prediction

Two dense neural networks are trained to predict the climber's next move. The first is a multi-class classification network used to predict which hand the climber should move or if they should move their feet. The model architecture is illustrated in Figure 5. The input layer contains 34 input neurons matching the number of features in the dataset. The first hidden layer contains 256 neurons and the second contains 128 neurons and they both use the ReLU activation function. Finally, the output layer contains three neurons (left, right, feet) and uses the Softmax activation function to create simulated prediction probabilities for each class. This model uses the Adam optimizer, the Categorical Cross-Entropy loss function, as well as prediction accuracy for the performance metric.

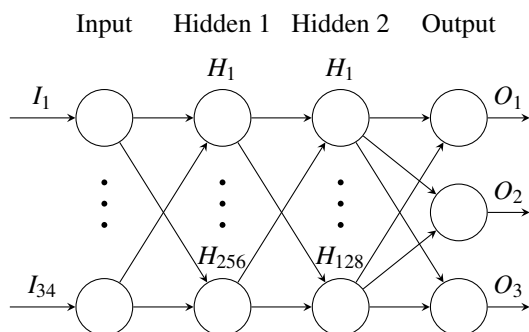


Figure 5: Limb Prediction Neural Network

Combining both the left and right foot movement into a single class causes a class imbalance where each hand represents 25% of the training dataset and the feet class contains 50%. In order to maintain the largest number of data points, Synthetic Minority Over-sampling Technique (SMOTE) [Cha02a] is used to balance the three classes. This technique creates additional, synthetic samples similar to the left and right hand observations to match the number of feet observations.

The second neural network is a multi-output regression network used to predict a bounding box ($x_{min}, y_{min}, x_{max}, y_{max}$) where the climber should move their hand next. Similar to the limb prediction model, the input layer contains 34 input neurons and this model uses five hidden layers. The hidden layers contain 256, 128, 64, 32 and 16 neurons, all using the ReLU activation function. The output layer consists of four neurons ($x_{min}, y_{min}, x_{max}, y_{max}$) activated by the Sigmoid function. This model uses the Adam optimizer, the Mean Squared Error (MSE) loss function, and Root Mean-Squared Error (RMSE) as the performance metric.

3.4 Command Translation

After the next move is predicted, Beta Caller continues in one of two ways depending on which limb is predicted. If the climber's feet are predicted to move, Beta Caller immediately communicates to the climber to move their feet upwards to a nearby hold. If one of the climber's hands is predicted to move, a series of computations must occur first to translate the move predictions into a usable command. For a majority of visually impaired climbers, a usable command contains which hand they should move along with what direction and how far they should move that hand.

To translate a prediction, Beta Caller first maps the predicted hold location to the nearest actual hold on the climbing wall found by the object detection model. Using this hold's center point and the climber's nose keypoint, the angle between the two coordinates is calculated and converted into an hour hand on a clock. The 12 o'clock position represents 0 degrees and each additional 30 degrees increments the hour hand by one.

In addition to the direction the climber should move their hand, the distance between the climber's predicted hand and the next hold is calculated using the climber's wrist keypoint and the hold's center point. The hold location and climber's joint keypoints are all represented by pixel values, so the pixel distance can be calculated easily. However, physical distance is required. In order to convert the pixel distance into physical distance, a known physical distance within the image must be used to create a conversion factor.

The distance between the climber's eyes, known as pupillary distance, is a distance on the human body that has the least amount of variance between human subjects. Limb length, waist or shoulder width, and essentially all other human body measurements vary greatly from human to human. However, the average pupillary distance is 2.5 inches and only ranges from 2-3 inches for all humans [Whi22a]. Therefore, the average pupillary distance is used as a known physical distance to create a conversion factor for calculated pixel distances. Despite the climber facing away from the camera, ViT-Pose is able to make an accurate prediction of where the person's eyes are located.

The exact distance, in feet and inches, is not the easiest for a climber to process, understand and move their body. Beta Caller converts the distance to the next hold to a more usable approximation of distance in order to simplify the command and not overwhelm the climber. Table 2 provides examples of these simplifications. Once the distance is known, all three parts of a command (hand, direction, and distance) are transmitted to the climber using a text-to-speech conversion library, *pyttsx3*. For example, a climber might hear "Right hand. Two o'clock. About two feet."

Feet	Inches	Move Distance
0	6	6 inches
2	0 - 3	About 2 feet
2	4 - 9	2 and a half feet
2	10 - 11	About 3 feet

Table 2: Distance Simplification Examples

4 RESULTS

An experiment was conducted for both the limb prediction and hold prediction neural networks with the goal of creating the best network architecture by trying three pose estimation models and changing the number of hidden layers and neurons.

The three pre-trained pose estimation models used in the experiment are MediaPipe [Goo24a], YOLOPose [Maj22a], and ViTPose [Xu22a]. The accuracy of the pose estimation model is of utmost importance as its outputs serve as crucial inputs for both the limb and hold prediction models. In order to find the best network architecture, the number of hidden layers was varied from one to five layers. Each additional hidden layer contains half of the number of neurons from the preceding hidden layer. The experiment includes networks where the first hidden layer contains 256 neurons, as well as networks that start with 128 neurons in the first hidden layer. In total, thirty different configurations were tested to find the best limb prediction model and ten configurations for the hold prediction model.

The results of the experiment are summarized in Table 3. The most accurate configurations obtain 80.08% limb prediction accuracy and both use ViTPose joint keypoints for limb prediction. One network contains two hidden layers with 256 and 128 neurons and the other network has four hidden layers containing 128, 64, 32, and 16 neurons. With the exception of one model architecture, using ViTPose for features for limb prediction always led to the highest accuracy. As a result, ViTPose data was used for all models in the hold prediction experiment. The best hold prediction model configuration reached a bounding box RMSE of 0.0679. The input and output data use normalized pixels between 0 and 1, so this RMSE represents the percentage of error from the predicted bounding box to the actual hold. The most accurate neural network contains five hidden layers with 256, 128, 64, 32 and 16 neurons and achieves a bounding box prediction error of 6.79%.

In order to validate the move predictions made by both neural networks, the ViTPose data from an image is used for inference with the two trained models. The predicted limb is written in the top left corner of the image and an yellow box is drawn to show the predicted coordinates where the climber should move their hand. A red box is drawn to indicate the closest hold to the predicted coordinates. The final inference results can be visualized in Figure 6. In the left image, the move

prediction models predict the climber will move their left hand to the yellow box which is mapped to the hold outlined with a red box. The right image confirms an accurate prediction that the climber did move their left hand to that predicted hold.

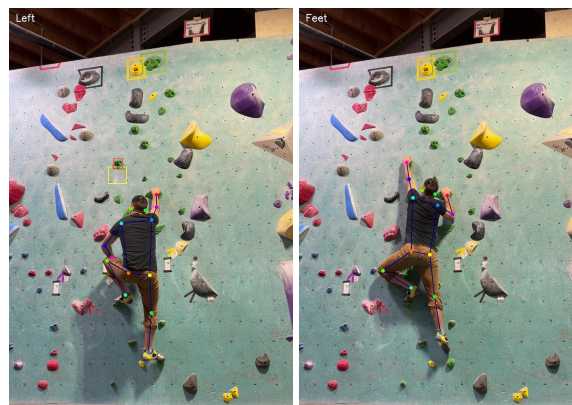


Figure 6: Beta Caller Prediction Results

Quantitative metrics describing the models' ability to accurately predict which limb the climber should and to where the climber should move their limb should not be the primary focus of evaluating the effectiveness of Beta Caller. Rock climbing, like dancing, can be more of an art than a science at times and provides climbers the option of approaching a route differently depending on the climber's body type, climbing style, flexibility, and numerous other factors. This results in a single route being climbed in innumerable diverse ways and establishes the fact that a correct sequence of moves does not exist, rather there are better singular moves than others.

In practice, a human caller will provide an "incorrect" command for which limb the climber should move or where they should move that limb, or both. However, this is merely the result of the caller's climbing style and how they would approach the route. Fortunately, the "incorrect" command will not impede the climber from continuing to move up the wall. Therefore, when Beta Caller's limb prediction model incorrectly predicts a limb to move, it is no different than the human caller and the climber will still be able to find a hold to move to even with the incorrectly predicted limb.

The hold prediction model's error is greatly reduced when Beta Caller maps the predicted bounding box to the closest hold found by the object detection model. However, even with the known location of a hold it is challenging for *any* climber to hear the command "Right hand. Two o'clock. About two feet." and move their hand to the exact location of the hold because the clock direction is rounded to the nearest hour and the distance is rounded to the nearest half foot. This style of command provides an appropriate amount of information without overwhelming the climber with too much

Hidden Layers	Hidden Neurons	Limb Accuracy			Hold RMSE
		MediaPipe	YOLOPose	ViTPose	ViTPose
1	256	69.74	68.94	73.18	0.0696
	128	69.30	68.32	67.05	0.0702
2	256, 128	72.81	72.67	80.08	0.0686
	128, 64	71.93	68.94	73.95	0.0687
3	256, 128, 64	69.30	69.57	77.78	0.0680
	128, 64, 32	72.81	68.94	79.31	0.0684
4	256, 128, 64, 32	71.49	69.57	79.69	0.0682
	128, 64, 32, 16	72.37	65.97	80.08	0.0690
5	256, 128, 64, 32, 16	71.49	65.97	77.78	0.0679
	128, 64, 32, 16, 8	71.93	64.60	78.16	0.0692

Table 3: Limb Prediction and Hold Prediction Experiment Results

detail. Fortunately, visually impaired climbers are especially talented at scanning the wall with their palms to find a hold nearby.

Without existing systems for comparison, these results provide a solid foundation, establish strong baselines, and pave the way for future rock climbing prediction systems.

5 CONCLUSION

This paper provided an operational prototype of an end-to-end system to revolutionize rock climbing for people of all skill levels, especially those with visual impairment. Beta Caller leverages a compilation of AI models to assist rock climbers up a wall to better enjoy and compete in the sport. Beta Caller successfully combines two state-of-the-art computer vision models for object detection and pose estimation and uses two neural networks to predict which limb the climber should move, as well as the next best hold for a climber to move to. Achieving limb prediction accuracy of 80.08% and predicting the next hold with only 6.79% error, Beta Caller creates a vital audible command containing the predicted move's information in a timely and useful way to enable the climber to make the next move and ascend to the top of the rock climbing wall.

6 FUTURE WORK

Current accuracy for the limb prediction and hold prediction neural networks are sufficient to demonstrate proof of concept and create a robust starting point for rock climbing movement prediction. However, these results have the potential to be more accurate. Dense neural networks are the simplest of feed-forward networks due to their connections from all neurons to all neurons between each layer. The images used as input for the dense neural networks are currently treated as individual, independent observations and they exclude the information observed from the previous frame or multiple frames. The sport of rock climbing is inherently sequential. For example, if a climber moves both

of their feet, the next move will likely be one of their hands, not their feet again. In order to more accurately model this sequential nature of rock climbing, a recurrent neural network or transformer network can be used to leverage the capability of making strong sequence-to-sequence predictions.

Beta Caller was built on a smaller dataset, so a promising avenue for future work involves collecting more data to fine-tune the object detection, pose estimation, and move prediction neural networks. This iterative fine-tuning process holds the potential to significantly enhance the accuracy and versatility of these models, ensuring their effectiveness across various climbing environments and accommodating diverse body types and climbing styles.

In order to establish a conversion ratio for estimating physical distance from pixel distance, Beta Caller utilizes the climber's pupil distance. It is imperative to validate the accuracy of these estimates through empirical testing. While this conversion methodology holds promise for its simplicity and applicability, its real-world performance should be evaluated across various rock climbing walls. Testing should assess the accuracy of the derived physical distances compared to ground truth measurements. Additionally, the robustness of the conversion ratio should be tested under different lighting conditions, camera angles, and camera distances away from the wall.

Another opportunity for future research could involve a comparative study evaluating the efficacy of Beta Caller, contrasted with the performance of a human caller providing beta to the same climber on the same routes. Speed, accuracy, and ease of understanding could be used as metrics to assess the climber's response to Beta Caller predictions compared to human caller predictions. Such a study would not only offer insights into the system's effectiveness but also shed light on the climber's interaction with inclusive technologies and enhance our understanding of the capabilities and adaptations within the visually impaired community.

7 ACKNOWLEDGMENTS

Our thanks to the United States Paralympic Triathlon Team for teaching us how to guide a climber with visual impairment up a wall and to CityROCK for providing a safe and welcoming gym space to collect data and for allowing one of the authors to climb amazing routes.

8 REFERENCES

- [Bre23a] Breen, M. Reed, T., Nishitani, Y., Jones, M., Breen, H.M., and Breen, M.S.. Wearable and Non-Invasive Sensors for Rock Climbing Applications: Science-Based Training and Performance Optimization. *Sensors*, Vol. 23, No. 11, pp. 5058, MDPI, 2023.
- [Cha02a] Chawla, N., Bowyer, K., Hall, L., and Kegelmeyer, W.P.. SMOTE: Synthetic Minority Over-Sampling Technique. *Journal of Artificial Intelligence Research*, Vol. 16, pp. 321-357, 2002.
- [Eka22a] Ekaireb, S., Khan, M.A., Pathuri, P., Bhatia, P.H. and Sharma, R. and Manjunath-Murkal, N.. Computer Vision Based Indoor Rock Climbing Analysis. 2022.
- [Goo24a] Google. Pose Landmark Detection Guide. MediaPipe, https://developers.google.com/mediapipe/solutions/vision/pose_landmarker. 2024.
- [Kos15a] Kosmalla, F., Daiber, F., Krüger, A.. Climb-sense: Automatic Climbing Route Recognition using Wrist-Worn Inertia Measurement Units. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pp. 2033-2042, 2015.
- [Kos17a] Kosmalla, F., Zenner, A., Speicher, M., Daiber, F., Herbig, N., Krüger, A.. Exploring Rock Climbing in Mixed Reality Environments. *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*, pp. 1787-1793, 2017.
- [Kos17b] Kosmalla, F., Daiber, F., Wiehr, F., Krüger, A.. Climbvis: Investigating In-Situ Visualizations for Understanding Climbing Movements by Demonstration. *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, pp. 270-279, 2017.
- [Lad13a] Ladha, C., Hammerla, N.Y., Olivier, P., and Plötz, T.. ClimbAX: Skill Assessment for Climbing Enthusiasts. *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pp. 235-244, 2013.
- [Low17a] Lowell, J. and Mortimer, P.. *The Dawn Wall*. Red Bull Media House and Sender Films, 2017.
- [Maj22a] Maji, D., Nagori, S., Mathew, M., and Poddar, D.. YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2637-2646, 2022.
- [Mor14a] Mortimer, P. and Rosen, N.. *Valley Uprising*. Sender Films, 2014.
- [Ouc10a] Ouchi, H., Nishida, Y., Kim, I., Motomura, Y. and Mizoguchi, H.. Detecting and Modeling Play Behavior using Sensor-Embedded Rock-Climbing Equipment. *Proceedings of the 9th International Conference on Interaction Design and Children*, pp. 118-127, 2010.
- [Pfe11a] Pfeil, J., Mitani, J. and Igarashi, T.. Interactive Climbing Route Design using a Simulated Virtual Climber. *SIGGRAPH Asia 2011 Sketches*, pp. 1-2, 2011.
- [Ram20a] Ramsay, J. and Chang, H.J.. Body Pose Sonification for a View-Independent Auditory Aid to Blind Rock Climbers. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 3414-3421, 2020.
- [Red16a] Redmon, J., Divvala, S., Girshick, R., and Farhadi, A.. You Only Look Once: Unified, Real-Time Object Detection. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.
- [Ric22a] Richardson, M., Petrini, K., and Proulx, M.. Climb-o-Vision: A Computer Vision Driven Sensory Substitution Device for Rock Climbing. *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, pp. 1-7, 2022.
- [Sib07a] Sibella, F., Frosio, I., Schena, F. and Borghese, N.A.. 3D Analysis of the Body Center of Mass in Rock Climbing. *Human Movement Science*, Vol. 26, No. 6, pp. 841-852, 2007.
- [Vas18a] Vasarhelyi, E. and Chin, J.. *Free Solo*. National Geographic Documentary Films, 2018.
- [Wei14a] Wei, E.. *Indoor Rock Climbing Wall Route Displayer*. 2014.
- [Whi22a] Whitten, C. Pupillary Distance: Types and How to Measure. *WebMD*. 2022.
- [Xu22a] Xu, Y., Zhang, J., Zhang, Q., and Tao, D.. ViTPose: Simple Vision Transformer Baselines for Human Pose Estimation. *Advances in Neural Information Processing Systems*, Vol. 35, No. 6, pp. 38571-38584, 2022.