Detecting Climbing Form in Bouldering Using Transfer Learning

Kenny Tran
University of Central Florida
ke219417@ucf.edu

Matthew Park
University of Central Florida
ma675447@ucf.edu

Abstract

Climbing technique plays a critical role in climbing performance, injury prevention, and fatigue prevention. Evaluating form is a very complex process and often times requires expert observation. This project utilizes computer vision and classification models to identify whether a climber is over-gripping on a bouldering hold or not. This is done using transfer learning on a variety of climbing poses. A custom dataset of manually labeled climbing images were taken from Central Rock Gym by the authors of this paper. A ResNet-18 CNN was used for this binary classification by altering the final layer for classification. This model was trained and validated using a 70/15/15 split between training, validation, and testing data respectively. The dataset was also taken from multiple angles of the same climbing positions, improving the robustness of variety and angles. The results of the experiment show that the model can successfully classify over-gripping and not over-gripping in various climbing positions. The results were strong despite the limited data. This work demonstrates how deep learning can be used to analyze climbing technique serves as a foundation for future innovations that could involve full videos and movement sequences.

ACM Reference Format:
Kenny Tran and Matthew Park. 2025. Detecting Climbing Form in Bouldering Using Transfer Learning. In . ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/nnnnnnn.nnnnnnn

1 Introduction

Good climbing performance relies heavily on efficient, strong, and stable movement patterns. Small differences in technique may allow beginner climbers to improve their skills and strength drastically. However, evaluating technique typically requires someone with great experience. Poor form, such as over-gripping, tapping feet, or unbalanced body position can lead to faster fatigue, higher risk of injury, and limit the room for improvement. As climbing (specifically bouldering) increases in popularity every year, utilizing computer vision and AI to assess climbing technique has become a point of interest.

Convolution Neural Networks (CNNs), are the standard when it comes to image classification. This is why they are continually used in classifying posture, movements, and technique across many sports. This is why the ResNet-18 model was used for this project. Using a modified pre-trained model allowed the team to easily integrate it with the dataset. The results demonstrate that transfer learning is a strong foundation for automating climbing form analysis and contributes to the future development of more complex models that will be able to analyze videos, dynamic movements, and full movement sequences.

Conference'17, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YYYY/MM
https://doi.org/10.1145/nnnnnnn.nnnnnnn

2 Problem Statement

The goal of this project is to develop a computer vision model that can classify over-gripping vs not over-gripping based on a climber's body position on the wall. The model should be able to predict whether a climbers body position reflects over-gripping cues or not over-gripping when given an input image. Over-gripping cues involve tight elbow angles, being excessively close to the wall, and foot misplacement.

3 Motivation

Strong climbing technique is the foundation of any climbers performance. Good technique is particularly important when climbs are long, thus demanding efficiency. Beginners have a hard time learning and fixing bad form. This can manifest as sagging hips, imprecise foot placement, and over-gripping. Fixing these issues can be difficult when someone doesn't have access to good feedback. They often have to rely on online tutorials and film of themselves which is a relatively inefficient way to learn.

Having this system that automates identifying a key component of bad form (Over-gripping) would be useful for:

- Self-Coaching: Climbers can record their session and look at their various body position and receive feedback without needing to hire a coach.
- Training Tool: Climbing gyms could implement this as part of their membership so basic members can have access to basic "coaching".
- Injury Prevention: Climbers can use this to spot weakness in their form which could lead to injury over time.

4 Method

The core model our project is based on is the ResNet-18 CNN architecture. We chose this because it's a well established model for image classification and met the performance and efficiency needs of our project and dataset. Since our dataset is smaller, it made sense to use ResNet-18 rather than ResNet-50 for example. ResNet is also in Torchvision which allowed us to easily use a pretrained model which we could fine tune to meet our binary classification needs.

4.1 Loss Function

For optimization, we used cross-entropy loss. This loss function is the standard when it comes to binary classification. The loss function takes the predicted probability and compares it with the ground truth label which penalizes the model when it makes an incorrect prediction. Then, we used Adam to update the weights of our model based on our loss function.

4.2 Training

Our over-gripping classification model was trained using a transfer learning approach based on the ResNet-18 backbone pre-trained on the ImageNet dataset. Before training the full model, we froze early convolutional layers and kept layers 3 and 4 as well as the convolutional head unfrozen for fine-tuning. This allowed us to reduce overfitting and adapt high level feature representations based on our climbing dataset.

The model was trained based on two classes of "over-gripping" or "not over-gripping", using the Adam optimizer. We used a learning rate of $1x10^{-3}$ for the classification head and a learning rate of $1x10^{-4}$ for the backbone layers. Training was conducted for around 8 epochs with a batch size of 16.

The epoch number was due to early stopping being implemented to reduce overfitting after a certain number of epochs. The checkpoint with the highest validation accuracy was selected as the final model for the results. Cross-entropy loss was used as the objective function, given its suitability for two-class classification.

Sample predictions by the climbing classification model were also displayed. Figure 1 shows several examples of the model's predictions along with the confidence scores. The results indicate that the model learns meaningful visual cues associated with overgripping, and correctly identifies most cases of "not over-gripping" with high confidence. Misclassifications do occur with poses in strange angles, hard to discern hand placements, or even with an out of focus image. The sample predictions show that the trained model performs reasonably well on unseen test samples, though additional data and refinement may further improve the performance of the model.

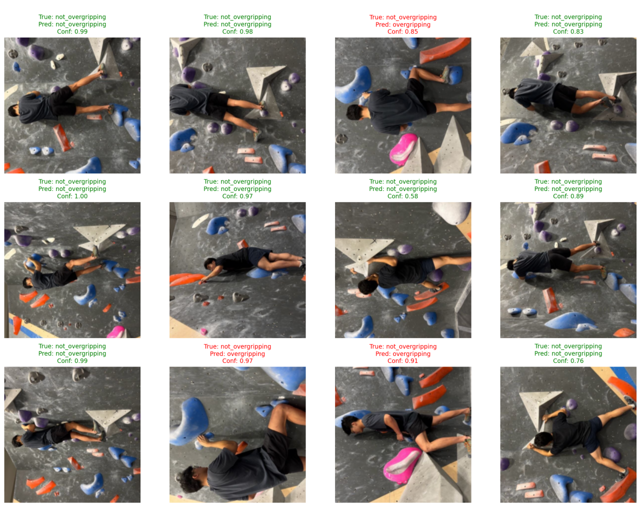

Figure 1: Sample Predictions

5 Experiments

5.1 Dataset

There were a handful of datasets online that we found. After carefully combing through each dataset, we found that the data collected by the publishers was not reliable and mostly redundant. For example, we found an extensive dataset containing over 800 climbing position images. The publishers claimed they had 97% accuracy when they trained their model. Upon further investigation, most of this dataset was inflated. We came to the conclusion that they recorded a climb and split the video into frames and used the frames as images. Because of this, a significant portion of the images essentially represented the same body position because they were one frame apart.

Therefore, we decided to create our own dataset. We went to our local climbing gym (Central Rock Gym - Orlando) and took 500 images. 250 of these images represented over-gripping, while the other 250 were not or "good" form. These images spanned over 150 body positions with three angles of each body position. One angle from the left, middle, and right side of the active climber.

Each image was standardized to be 224x224 because of its general use in deep learning models. Our data was split to be 70/15/15 with training, validation, and testing data. This equates to roughly 350 training images, 75 validation and testing images.

We recognize that our dataset is relatively small and foresaw how this could affect the training and testing of our model which will be discussed later.

5.2 Data Pre-Processing

To add an extra layer of robustness to our data, we preprocessed our images. This was done by cropping, flipping, rotating, erasing, color jittering to reduce overfitting and improve overall the generalization capabilities of our model. These were only done to our training data. Our validation and testing data was only cropped and centered without any addition augmentation to ensure consistent evaluation.

5.3 Evaluation Metrics

To assess the performance of the overgripping classification model, we used accuracy, precision, recall, F1-score, and a confusion matrix. These metrics provide a comprehensive view of how well the classifier distinguished between overgripping and not overgripping.

In our experiments, precision and recall were computed separately for each class to understand whether the model was more prone to false positives or false negatives. Recall was used to measure how well the classifier detected all true instances of each class, especially when it came to overfitting, which is what we were focused on. Precision measured how reliable the model was when predicting the grip.

F1 score provided a balanced measurement of performance by combining both precision and recall, allowing us to assess how well the model handled differences between overgripping and not overgripping.

Finally, we included a confusion matrix to aid in the visualization of true vs predicted labels. This representation highlights misclassification patterns, and if the model was likely to over predict one class or struggle with another.

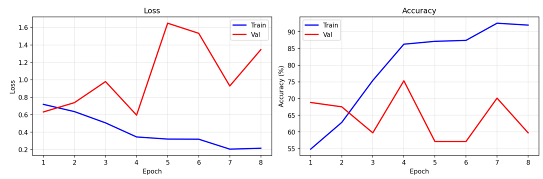

Figure 2: Training Loss vs Accuracy.

5.4 Results

We tracked the training and validation loss and accuracy to evaluate the performance of our ResNet-18 model. This was tracked across 8 epochs as shown in Figure 2. Training loss decreased from 0.72 in epoch 1 to roughly 0.21 by epoch 8. Meanwhile, training accuracy rose from 55% to 92%. Based on these results, the model successfully learned the features of over-gripping and not over-gripping and was able to distinguish between the two features within the training set.

Although our model learned on the training data, the validation and testing data told a different story. Validation accuracy had more variability than the training data. Validation accuracy fluctuated between 57% and 75. The highest accuracy occurring at epoch 4 with an accuracy of 75.32%. This gap between training and validation suggests that our model was overfit. Our main suspicion is that our dataset was simply too small, leading the model to excessively learn based on the training data, unable to generalize to unseen data. Despite this, our model displayed strong performance suggesting that it was able to learn some features of over-gripping even when data was limited.

The final version of our model scored an accuracy of 66.22% when tested against our testing data. This indicated that it was overall moderately able to classify over-gripping and not over-gripping. We can see how the model did exactly in Figure 3. We can see that the model is better at predicting not over-gripping with 28 labeled correctly while only 21 for over-gripping. We also see that there are 16 false negatives for over-gripping. This means that the model fails to label the position as over-gripping in these instances. The model is also conservative when it comes to labeling positions as over-gripping as there were only 9 false positives.

Overall, this suggests that our model is better at detecting not over-gripping than it is at detecting over-gripping.

6 Analysis and Discussion

Our main concern for this project was gathering data. We underestimated how much data we would need and generally assumed it would be publicly available. Even though there were some public datasets, it was unfortunate that they did not meet our standards of quality. Our hope was that when we created our own dataset, it would be enough for a model to perform over 80% accuracy in detecting over-gripping vs not over-gripping. Evidently we undershot our goal by roughly 10% but it was to be expected given the depth of our data.

The first model we trained performed very poorly, hovering around 40-50%. After making some changes such as image preprocessing, unfreezing the third and fourth layer of the ResNet-18 CNN, and early stopping when validation accuracy began to fluctuate and decrease (usually around epoch 7-8), to improve the accuracy and reduce overfitting of our model.

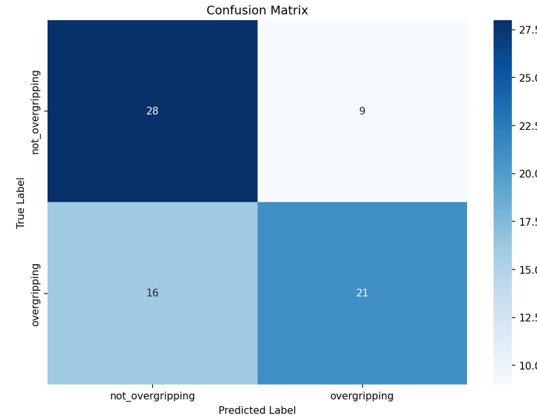

Figure 3: Confusion Matrix

7 Conclusion

Our project explored how image classification could be used on climbing form, specifically distinguishing between over-gripping and not over-gripping based on body position images on a bouldering wall. By tuning a ResNet-18 model on our custom dataset collected from Central Rock Gym, we demonstrated that deep learning can extract important body positional cues even when the dataset is small. Although our model performed well on training data, our validation and testing sets exposed the limitations of our model, proving that it could only make accurate predictions 66.22% of the time. Our confusion matrix revealed that our model is conservative when classifying over-gripping and is better at detecting not over-gripping positions.

Despite the limitations of our model, the results showcase the potential for using computer vision to assess climbing technique. The model demonstrated that it can learn essential body position features like elbow angle and posture.

There's substantial room for improvement, including increasing the dataset, having more diverse climbers, more angles, incorporating videos, and utilizing a deeper model to increase accuracy and reliability.

Overall, this project demonstrates that computer vision can help climbers receive automated feedback on their technique. Should this project or field be developed further, it has the potential to build climbing skills, prevent injury, build engagement within the climbing community.

8 Contribution

8.1 Code

I wrote the model_.py, train.py, and plotting.py. These files initialized the ResNet-18 model, defines the parameters and how training is done, and how results are plotted respectively.

8.2 Report

I wrote the abstract, problem statement, motivation, method, dataset, and results sections of this report.

8.3 Takeaways

This project taught me the importance of data collection and how data is the center of all machine learning. I have a new found respect for machine learning researchers who collect and label data and the sheer amount of data that must be processed. This project also taught me a lot about how image classification works and what type of classification models are best suited for each use case. Overall I had a good experience with this project. This project felt like a little bit of every assignment this semester. I see a clear vision of what kind of app or company this could be used for in the climbing community and hope to see it fully built out one day.

References

[1] Roboflow. 2025. What Is ResNet-18? How to Use the Lightweight CNN Model. https://blog.roboflow.com/resnet-18/.

[2] (Adhis Workspace). 2023. Climbing-Form Classification Dataset and Model. https://universe.roboflow.com/adhis-workspace-bwqgy/climbing-form-classification-hpfpl.