Hold On to Data

Patrick Brophy, Michael Roytman, Prabvir Kukreja
Date of submission: 12/14/2023
Project Repository: https://github.com/mroytman83/CS539_Final_Project

Abstract:

This paper presents an in-depth investigation into the efficacy of data augmentation techniques in improving rock climbing hold detection, utilizing the YOLOv8 architecture. Given the unique challenges posed by the variability in rock climbing hold shapes, sizes, and colors, coupled with diverse lighting conditions and backgrounds, conventional object detection models often struggle to achieve high accuracy. To address these challenges, we implement various data augmentation strategies, such as rotation, scaling, and color adjustment, to enrich our dataset. The YOLOv8 model, known for its speed and accuracy in object detection, is then trained on this augmented dataset. Our results demonstrate a significant improvement in the model's ability to detect climbing holds under varying conditions, compared to training with non-augmented data. This advancement not only enhances the safety and planning aspects of rock climbing but also contributes to the broader field of object detection in dynamic and unstructured environments.

Introduction:

The group's shared interest in rock-climbing was foundational to creating a project which resonated with our passion. The initial goal of the project was to employ a vision-based neural network for the identification of distinct rock climbing holds on a wall. Given a few categories of rock climbing holds, we sought to collect a large dataset of walls from different rock climbing gyms which contained different categories of holds and their corresponding labels. The intuition is that once the model is trained, it can be given an image of a wall and then it will output labels and bounding boxes for the different types of holds. Additionally, we experimented on the quality of predictions made by the model, by examining the model's performance from a lens of data augmentation. The experiment would segment the inference into two parts, by creating a model which will be trained on un-transformed data, and a model which will be trained on augmented data. We hypothesize that transforming the data would improve object recognition by increasing the diversity of the training data, helping the model generalize better to out of distribution samples. The use of an optimal pre-processing pipeline for the construction of rock-climbing hold detection would allow for more robustness in out of distribution performance, which could be then applied to applications within other gyms to serve as an educational resource for planning climbing strategies according to appropriate holds. This project took inspiration from CLIMBNET-an implementation of a CNN that detects holds on climbing gym walls and returns the appropriate boundary mask for use in instance segmentation. CLIMBNET [5] was limited to solely detecting volumes versus holds, so we expanded the number of labels.

Methods:

In the following sections we describe the methods used for our experiment.

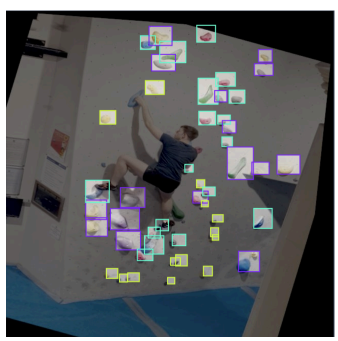
Data Collection and Augmentation:

We used this dataset [1] from RoboFlow which contains several hundred labeled images of climbing holds. This dataset was used for training our base model. To build our augmented dataset used the original dataset and performed the following augmentations:

- Rotation: Between -15 degrees and 15 degrees
- Shear: 15 degree horizontal and 15 degree vertical
- Brightness: Between -25% and 25%
- Exposure: Between -25% and 25%
- Blur: Up to 1.25 pixels

These augmentations capture the diversity of both rock climbing holds, cameras, camera angles, and climbing gym environments. To build a comprehensive out of sample dataset for evaluating our rock climbing hold detection model, we collected images from our local climbing gym, Boulders. To ensure the dataset's quality and relevance, team members Patrick Brophy, Michael Roytman, and Prabvir Kukreja manually labeled the images, specifying the types of climbing holds and corresponding bounding boxes. Our out of sample dataset can be found on Roboflow[6]. Figure 1 shows examples of images of our base dataset, the augmented dataset, and our out of sample evaluation set.

Figure 1

Base Dataset	Augmented Dataset	Out of Sample Evaluation Set
		

This table shows samples from the different datasets we used for training and evaluating the base and augmented models.

Algorithm and Program:

We selected the YOLOv8 architecture as the foundation for our rock climbing hold detection model, motivated by its enhanced ability to classify objects of various sizes, a notable advancement over its predecessors. Initially, we fine-tuned this model with a collection of non-augmented images of climbing holds, aiming to optimize its baseline performance and improve its precision in detecting climbing holds. Subsequently, we applied further fine-tuning to this pre-adjusted model using an augmented dataset of climbing hold images. This additional phase of fine-tuning was designed to expose the model to a broader and more varied range of images, thereby enhancing its overall robustness and accuracy. This training process allowed us to investigate the differences in performance between a model fine-tuned solely on non-augmented data and the same model extended to accommodate augmented data.

Platform:

The experiments will be conducted on a high-performance computing platform equipped with GPUs to accelerate the training process. We will utilize popular deep learning frameworks such as PyTorch to implement and fine-tune the YOLOv8 model.

Experimentation:

We orchestrated the ClimbingHoldDetection from Roboflow to serve as the fine-tuned dataset for our custom object detection models using two .yaml. The base pre-trained Yolov8 model was fine-tuned with the original dataset over 40 epochs, and arrived at a precision of .856, recall of .713 and mAP of .7935. The augmented Yolov8 model was fine-tuned with the same dataset which underwent the data augmentation methods listed in *Data Collection and Augmentation*. The augmented Yolov8 model achieved a precision of .816, recall of .744 and mAP of .799. For visualization purposes we created a function which would map the predicted bounding boxes onto objects within the image and the confidence of the prediction.

Next, we used a custom dataset from Roboflow, composed of well curated images from a pristine local rock climbing venue. The out-sample dataset is comprised of 12 images, and the reference for the process of its collection can be found in the *Data Collection and Augmentation* section, and its source Roboflow[6]. This out-of-sample data was then used to evaluate the performance of the augmented fine-tuned model vs the base fine-tuned model.

Evaluation Metrics:

The out-of-sample testing dataset contained text documents with 8-tuples of ground-truth coordinates for the bounding boxes of each labeled rock-climbing hold, and its corresponding image. The 8-tuples were inserted into a dictionary which contained ground truth data for each image, with its corresponding “box” key, as well as the label, and confidence score. In order to compare the predicted bounding boxes, and and the ground truth bounding boxes, we created a function which transforms the 8-tuple to 4-tuple by calculating the minimum and maximum x and y coordinates among these eight values to form a new bounding box class in the format (x1, y1, x2, y2).

To quantify the performance of the two models, we needed to calculate the precision, recall, and F1 score for both models on out-of-sample data. The IOU(Intersection over Union) helped us compare how the predicted bounding box aligned with the actual object annotation, furthermore enabling us to fulfill the assessment of model accuracy with our selected metrics.

The function which accomplished that calculation first computed the height and the width of the intersection between the two bounding boxes. Then the function verifies non-negative/non-zero values. Finally the function computes the areas of intersection, the area of each bounding box, and the areas of the union between the two bounding boxes, and then returns the ratio.

Calculating precision and recall was designated within its own function. First, we initialize d counters for true positives, false positives, and false negatives. Then we extracted the bounding box information and labels from the ground truth and predicted results. We iterated over each detected result, and for each result, then iterated over the ground truth data to find a matching ground truth bounding box with the same label and an Intersection over Union (IoU) greater than the specified threshold which we set at .5. Given the match of either condition we updated the true positives and false positives counters based on whether a match is found or not. Finally, we calculated the number of false negatives as the difference between the total ground truth instances and true positives and then delivered the final calculations for precision, recall and F1.

The testing loop consisted of evaluating the predicted bounding boxes and labels for each testing image, and then comparing it with the corresponding 8-tuple for each image's precision and recall, which were then averaged out over the testing images.

Figure 2

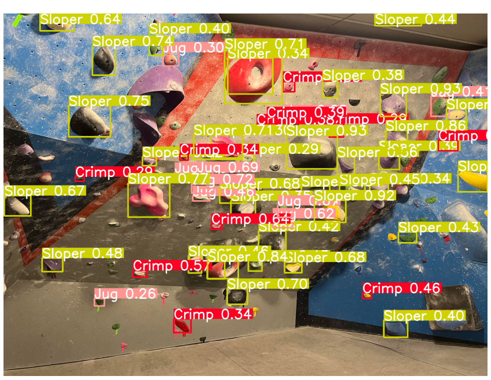
	Precision	Recall	F1
Base Model	0.11	0.14	0.12
Augmented Model	0.14	0.20	0.16

This table shows the performance between the Base Model and the Augmented Model on the Out of Distribution Evaluation Set.

Results:

After meticulously fine-tuning both the base and the augmented models, we conducted evaluations on an out-of-distribution dataset. The results from these evaluations clearly demonstrate that the augmented model exhibits superior performance in out-of-distribution scenarios. This enhanced performance is particularly evident from the augmented model's higher precision, recall, and F1 scores. These findings robustly validate our hypothesis that applying data augmentation techniques would enhance the efficacy of rock climbing hold detection in out-of-sample environments.

Figure 3

Base Model	Augmented Model
	

Discussion:

In the study "Hold On to Data," the use of data augmentation techniques, including rotation, shear, brightness, exposure, and blur adjustments, enhanced the performance of the YOLOv8 model for rock climbing hold detection. This approach addressed the challenges posed by the variability in hold shapes, sizes, and environmental conditions, leading to improved precision, recall, and F1 scores. While focusing on climbing hold detection, the study's findings have broader implications for object detection in dynamic environments, suggesting the potential for these methods to be applied in other complex domains and showcasing the prospect of improving the data label and collection process. The research highlights the importance of robust and adaptable object detection methods, offering insights for future advancements in the niche of computer vision and rock climbing.

References

[1] Robertson, Shane. "ClimbingHoldDetection Computer Vision Project." *Roboflow*, universe.roboflow.com/shane-robertson-9azgn/climbingholddetection. Accessed 14 Dec. 2023.

[2] Murryv, Gabriel. "Climbing Holds Detector Object Detection Dataset and Pre-Trained Model by Gabriel Murryv." *Roboflow*, universe.roboflow.com/gabriel-murry-hbznu/climbing-holds-detector. Accessed 14 Dec. 2023.

[3] Wei, Eric. "Indoor Rock Climbing Wall Route Displayer - Stanford University." *Indoor Rock Climbing Wall Route Displayer*, stacks.stanford.edu/file/druid:bf950qp8995/Wei.pdf. Accessed 15 Dec. 2023.

[4] Yan, Ming. "CIMI4D: A Large Multimodal Climbing Motion Dataset under Human-Scene." *CIMI4D: A Large Multimodal Climbing Motion Dataset under Human-Scene Interactions*, openaccess.thecvf.com/content/CVPR2023/papers/Yan_CIMI4D_A_Large_Multimodal_Climbing_Motion_Dataset_Under_Human-Scene_Interactions_CVPR_2023_paper.pdf. Accessed 15 Dec. 2023.

[5] Murphy, Cebas. "Climbnet - Detect and Segment Indoor Climbing Holds." *GitHub*, github.com/cydivision/climbnet. Accessed 14 Dec. 2023.

[6] Kukreja, Prabvir, et al. "Rock-Climbing-Hold-Classifier Object Detection Dataset by Rockclimbingholdclassifier." *Roboflow*, universe.roboflow.com/rockclimbingholdclassifier-2yrpe/rock-climbing-hold-classifier. Accessed 14 Dec. 2023.