Deep Neural Strategies for Uncovering Climbing Elements and Mapping Route Patterns

Srinivasa Sai Abhijit Challapalli*

Research Scholar and PhD Student, The University of Texas, Arlington, Texas USA

abhijitchallapalli99@gmail.com

Abstract: This work presents a new deep learning architecture specifically designed to extract and structure climbing features in indoor settings. By building on two neural models, one pair wise similarity and one triplet comparison, the approach is learned to discriminate between climbing holds belonging to the same route and those that do not. In contrast with more conventional color clustering methods, our method is superior in accuracy and robustness, albeit with the requirements of complex manual annotation and higher computational demands. The findings emphasize the promise of deep neural networks to transform automated route mapping within climbing environments.

Keywords: neural models, conventional color clustering methods, accuracy, robustness, deep learning architecture.

1 Introduction

Indoor climbing has experienced a significant surge in popu- larity, particularly after its debut at the Tokyo 2020 Olympics. This increased interest has spurred a range of technological innovations in climbing analysis and management. In partic- ular, computer vision techniques that focus on segmenting climbing holds and identifying routes have become integral in both commercial gyms and competitive events.

A. Applications in Climbing Analysis

Several practical applications emerge from advanced image processing in climbing facilities:

- **Competition Evaluation** – Real-time pose tracking sys- tems can monitor climber performance by recording attempts and successful completions, reducing the depen- dency on large judging teams.
- **Climber Log Systems** – After route modifications, fa- cilities can capture images that feed into mobile or web applications. These platforms enable climbers to record ascents, provide feedback through likes or comments, and propose difficulty ratings, akin to existing solutions for standardized boards such as Kilterboard (1) and Moonboard (2).
- **Difficulty Assessment** – Techniques to automatically predict the challenge level of a route using established grading standards.

Beyond these targeted uses, any vision-based system de- signed for indoor climbing must accurately delineate the segments of an image (or textures in 3D models) that corre- spond to climbing holds and determine how these holds form coherent routes.

Prior investigations have addressed some of these chal- lenges. For example, one study employed machine learning to predict the difficulty of a standardized training board (3) while another focused on post-climb analysis for performance evaluation in commercial gyms (4). Additional work has examined hold segmentation through both conventional and learning-based approaches (5; 6); however, many of these models and datasets remain restricted in accessibility.

2 Method

Climbing walls are outfitted with holds and volumes that vary in shape, size, and color. Typically, a single climbing route is defined by holds sharing a common color, with the initial and final holds highlighted by distinctive tape. The volumes—large fixtures integrated into the wall—are designed for multiple uses across different routes, though their arrangement may vary with each setting to avoid repetitive patterns.

A. Problem Formulation

Under these assumptions, the challenge of hold segmentation is defined as follows: Given a 2D RGB image, generate two sets of non-overlapping polygons that pinpoint the locations of holds and volumes. (Note: For simplicity, holds mounted in a stacked arrangement are treated as a single structure.)

Similarly, the task of route segmentation is described by the following formulation: Given a collection of hold images (obtained through segmentation), identify clusters of holds that represent distinct climbing routes. Since volumes are universally present across routes, they are excluded from this grouping process, thereby simplifying the segmentation task. Although certain gyms may allocate volumes to specific routes, the prevailing approach treats them as a common feature.

B. Acquisition of Data

Datasets for training and evaluation were compiled from images taken at two distinct climbing facilities—one located in the Czech Republic (Sm´ıchoff) and the other in Heidelberg (Boulderhaus). To assess performance across varying image qualities, both a high-end DSLR camera and a modern smart- phone were utilized. Table I summarizes the key parameters of the dataset.

Capturing images from multiple gyms is crucial, as variations in hold manufacturing and wall backgrounds (e.g., dark)

TABLE I: Summary of dataset parameters

Location	Device	Images	Annotated
Boulderhaus	Nikon Z 50	1062	15 hold (8 route)
Boulderhaus	Samsung A51	244	2 hold (0 route)
			oute)

Fig. 1. Sm´ıchoff facility

Fig. 2. Boulderhaus facility

Fig. 3. Comparison of different climbing facilities

A subset of the collected images was manually annotated using the VGG Image Annotator (VIA) (7; 8). This process accounted for varying numbers and dimensions of holds and volumes, diverse lighting conditions, multiple perspectives, and assorted backgrounds. Each hold was delineated with a polygon and classified as either a hold or a volume, with holds that belonged to the same climbing route receiving a common route identifier.

C. *Annotation Statistics*

Across the collected datasets, a total of 1688 regions were manually delineated, with 1597 corresponding to climbing holds and 91 to wall volumes. The relative sizes of these annotations (as a percentage of the total image area) and the frequency per image are illustrated in Figure 4 and Figure 5, respectively

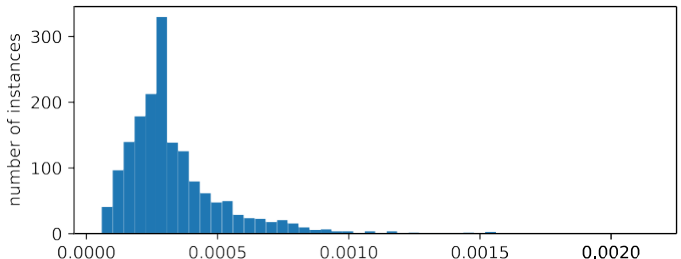

Fig. 4. Relative size of holds as a percentage of the image area

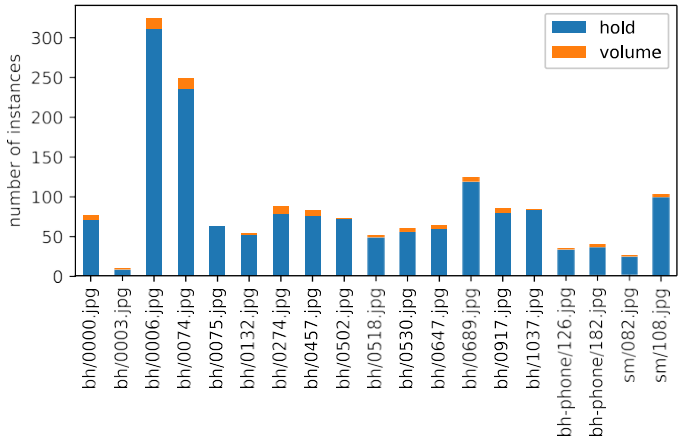

Fig. 5. Count of hold and volume instances per category

The entire dataset, comprising both the images and corresponding annotations, is distributed freely under the CC NC- SA 4.0 license via Kaggle (9).

D. Hold Segmentation

1) Conventional Technique

In a classical approach, two primary features are utilized to identify a hold: its edges and color. These attributes, however, are not unique to holds and may describe other elements within the scene, which introduces ambiguity in discerning object semantics.

The conventional method was executed using Python and OpenCV (10), drawing inspiration from (4). This strategy enhances basic blob and edge detection by incorporating hold- specific heuristics and integrating both approaches into a unified process

a) Edge Extraction.

The Canny edge detector, with hysteresis thresholds set at 20 and 25, is applied to a Gaussian-blurred version of the image to suppress noise. Unwanted edges—such as those from bolt holes—are eliminated by filtering based on minimum area coverage. Additionally, a straightness filter, implemented via linear regression on edge coordinates and thresholding the average squared distance, is used to remove linear artifacts such as wall seams. Sample outcomes of this process are depicted in Figure 9.

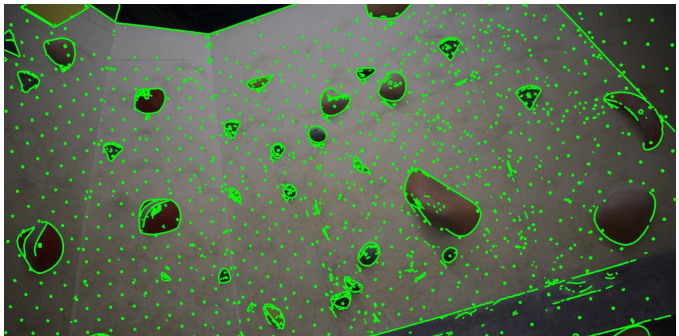

Fig. 6. Canny edge extraction

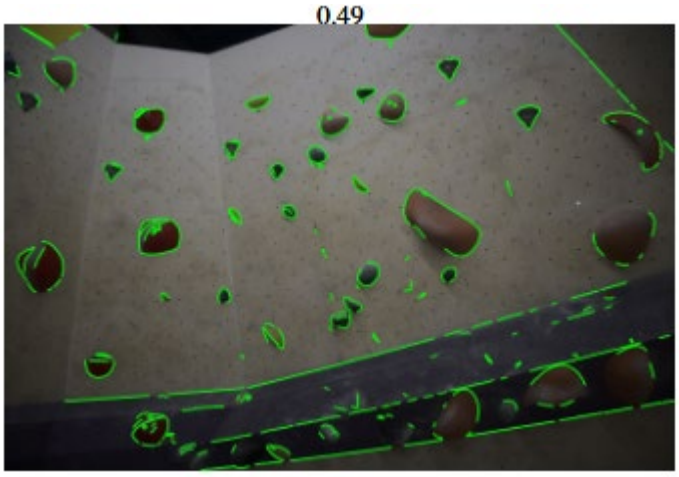

Fig. 7. Area-based filtering

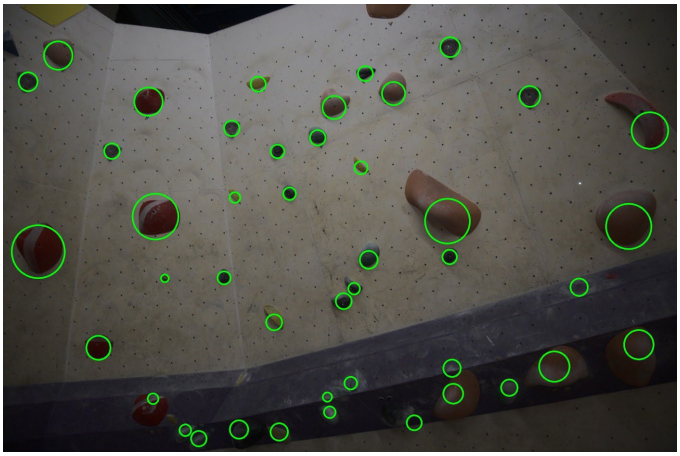

Fig. 8. Straightness-based filtering

Fig. 9. Results from the edge detection process.

b) Blob Detection.

A simple blob detector is employed with an area filter. The method converts the image to binary across a range of bright- ness thresholds (minimum 1, maximum 200, with increments of 10), identifies connected components to determine centers, and groups nearby centers into unified blobs. Blobs exhibiting more than a 15% overlap are merged, a necessary step to handle dual-texture holds where chalk creates a visual split. An example of blob detection is shown in Figure 10.

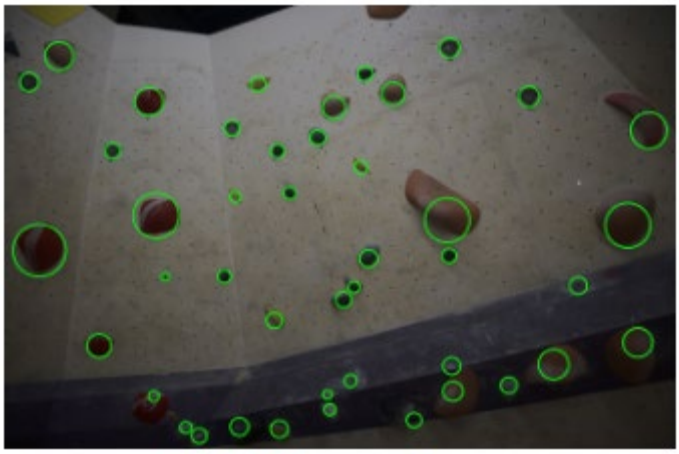

Fig. 10. Result of the blob detection procedure.

c) Fusion of Techniques

The two methods are combined by using the blobs as initial indicators of hold positions, while the edges serve as guides for defining their boundaries. To refine these boundaries into precise masks, binary thresholding is applied to the RGB channels at incremental steps. The Canny detector then reprocesses these thresholds, selecting the contour that best approximates the original outline based on a metric (for instance, the squared distance between corresponding contour points). A sample output of this combined strategy is provided in Figure 11.

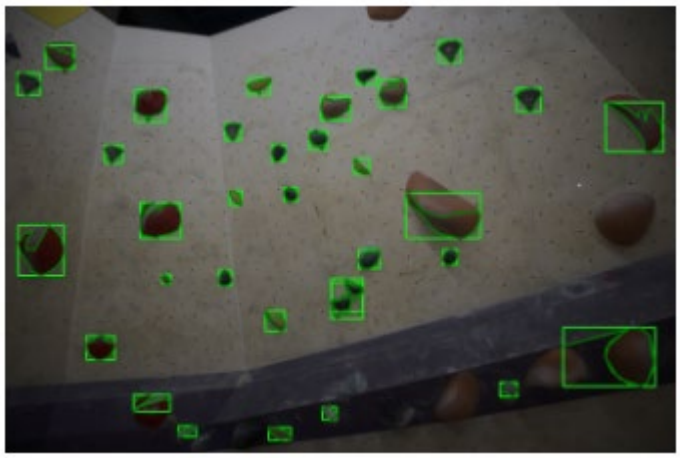

Fig. 11. Result of the combined hold segmentation method

A significant limitation of this traditional approach is its sensitivity to parameter variations. Optimal settings for edge and blob detection differ from one image to another. Parameter estimation can be automated using heuristic functions (e.g., determining minimum areas via histogram spikes of contour sizes), yet this may fail on datasets lacking common features such as bolt holes, particularly in images from competitive settings

2) Learning-Based Technique

a) Overview.

A learning-based strategy employs the Mask R-CNN ar- chitecture (11), implemented via the detectron2 library (12). This approach leverages pretrained weights from the COCO dataset (13), which are then fine-tuned for the specialized task of climbing hold segmentation.

b) Data Augmentation and Training.

Given the limited size of the dataset, extensive data aug- mentation is performed. Operations include random cropping to half the original height and width, resizing such that the shortest edge measures 640 pixels, random horizontal flipping, and minor random adjustments to rotation, brightness, contrast, and saturation. The dataset is partitioned into 80% for training and 20% for testing. Multiple model variations were trained on a GTX 1060 6GB for 1000 iterations with a learning rate of $\epsilon = 0.0006$, using the default configuration of the mask_rcnn_R_50_FPN_3x model

c) Optimization for Small Objects.

Due to the small size and high count of objects in each image, modifications were necessary. Initially, the number of proposals from the Region Proposal Network (RPN) was dou- bled (from 12,000 to 24,000 pre-NMS and from 2000 to 4000 post-NMS) to capture more instances. Further adjustments included limiting the model to utilize only P_2 (features at 1/4 scale) and P_3 (features at 1/8 scale) from the Feature Pyramid Network (FPN) to enhance detection of small objects.

d) Evaluation Metrics.

For assessment, the maximum detections per image were increased to 300. The evaluation metric employed is the Intersection over Union (IoU) between the predicted mask (mask_p) and the ground truth mask (mask_{gt}):

$$\text{IoU} = \frac{\text{area}(\text{mask}_p \cap \text{mask}_{gt})}{\text{area}(\text{mask}_p \cup \text{mask}_{gt})}$$

A detection is deemed correct if the IoU exceeds a threshold t and the class prediction is accurate; otherwise, it is marked as a false positive. The average precision (AP) is derived from the area under the precision-recall curve for each class, and the mean average precision (mAP) is calculated as

$$\text{mAP} = \frac{1}{|K|} \sum_{i=1}^{|K|} \text{AP}_i$$

This computation is repeated over multiple IoU thresholds $t \in$

$\{0.50, 0.55, \ldots, 0.95\}$, consistent with the COCO evaluation protocol.

e) Results.

Due to the limited number of volume instances, emphasis is placed on the AP for holds. Among the various configurations, the Augmented-P2P3-DoubleTopK-NMS-2K variant achieved the highest AP on both the Boulderhaus and Sm´ıchoff datasets (65.99 and 63.67, respectively). On the Boulderhaus- Phone dataset, the Augmented-DoubleTopK-NMS variant outperformed others with an AP of 81.23. The improved performance on the phone dataset is attributed to the larger apparent size of holds, facilitating easier segmentation. A representative comparison of the segmentation results is displayed in Figure 14.

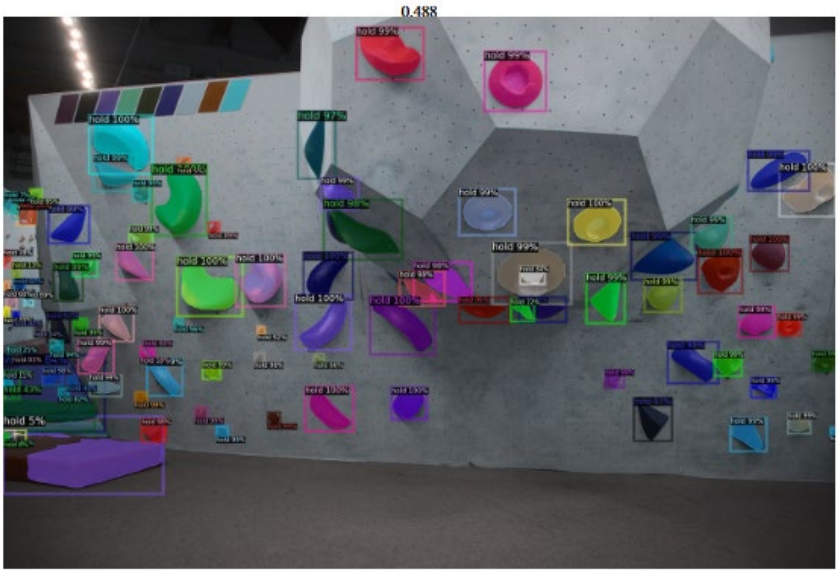

Fig. 12. Sm´ıchoff facility

Fig. 13. Boulderhaus facility (camera)
Fig 14. comparison of the segmentation results

TABLE II: SUMMARY OF MODEL VARIATIONS EXPLORED FOR HOLD DETECTION.

Basic	Default configuration without any augmentations.
Augmented	Includes a suite of augmentation techniques.
Augmented-NoRandomCrop	Augmentations applied, excluding random cropping.
Augmented-DoubleTopK-NMS	Augmentations applied with a doubled count of top-k proposals pre- and post-NMS.
Augmented-P2P3	Augmentations with ROI and segmentation heads utilizing only P_2 and P_3 features.
Augmented-P2P3-DoubleTopK-NMS	Combines P_2/P_3 feature utilization with a doubled top-k proposal count.
Augmented-P2P3-DoubleTopK-NMS-2K	Same as the previous model, but trained for 2000 iterations.

f) Model Descriptions.

A variety of model configurations were explored, as summa- rized in Table II. These configurations differ in augmentation strategies, the inclusion or exclusion of random cropping, adjustments in the number of top-k results in non-maximum suppression (NMS), and modifications in the feature layers used by the region-of-interest (ROI) and segmentation heads

E. Route Segmentation

1) Conventional Methodology

After the identification and annotation of climbing holds, the next step is to assemble routes based on holds sharing similar color properties. This traditional method employs OpenCV

(10) alongside k-means and k-medoids clustering available in the scikit-learn toolkit (14) using various feature representa- tions. Due to the similar mechanics behind these clustering algorithms, the outcomes tend to be alike.

a) Mean RGB Feature Clustering.

One approach involves computing the average RGB val- ues for each hold (15). In bright and clear images—where shadows and chalk artifacts are minimal—this method clusters holds with comparable colors into distinct routes. Figure 15 illustrates a scenario where ten out of twelve red holds were correctly grouped, although there can be misclassifications when clusters inadvertently merge holds from separate routes (see Figure 16).

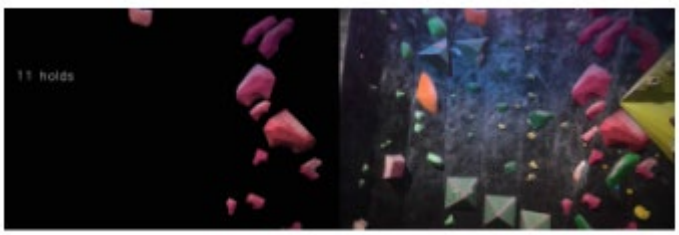

Fig. 15. Clustering based on mean RGB values: a red route example.

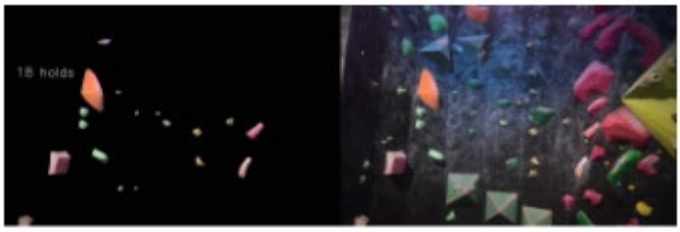

Fig. 16. Clustering based on mean RGB values resulting in an incorrect grouping.

b) Color Moment Features.

An alternative feature set employs color moments, which encapsulate the color distribution of each hold much like central moments describe probability distributions. Although this method tends to reduce false positives, it may incorrectly split holds of identical colors across multiple clusters, as depicted in Figure 17.

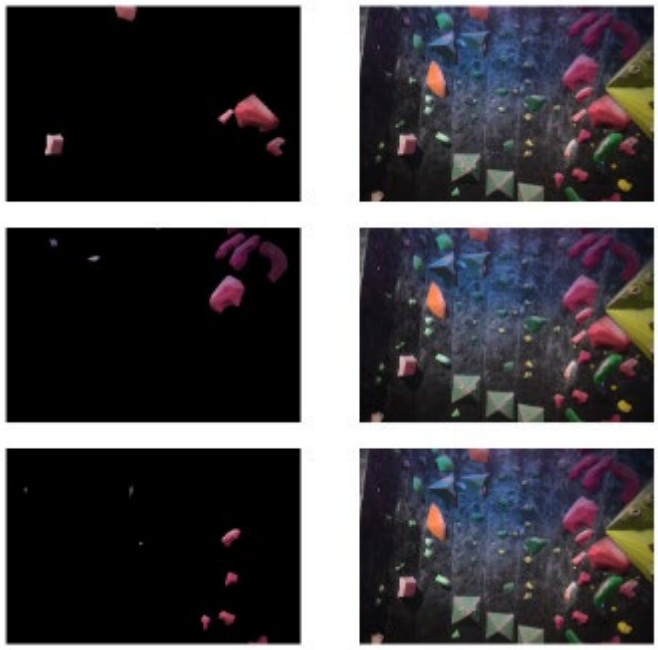

Fig. 17. Clustering using color moments, resulting in multiple red route groupings

c) Histogram-Based Features.

Additionally, histograms computed via OpenCV's calcHist function on the RGB channels were tested. Contrary to expectations, this method tended to aggregate most holds into a single cluster, leaving other clusters sparsely populated, as shown in Figure 18.

Fig. 18. Histogram-based clustering resulting in one dominant cluster.

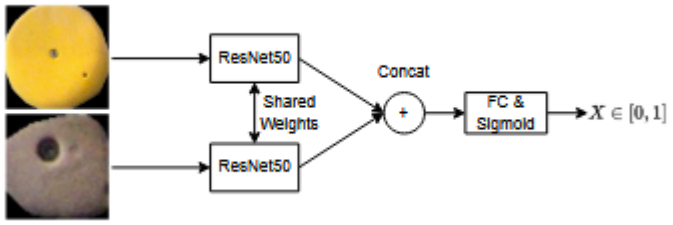

Fig. 19. Example of a Siamese network used for comparing hold images from different routes

Fig. 20. Visualization of triplet loss

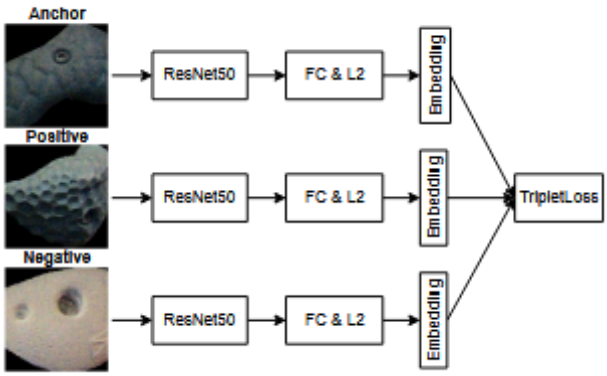

Fig. 21. Illustration of the triplet network, where anchors and positive samples belong to the same route while negatives come from a different route

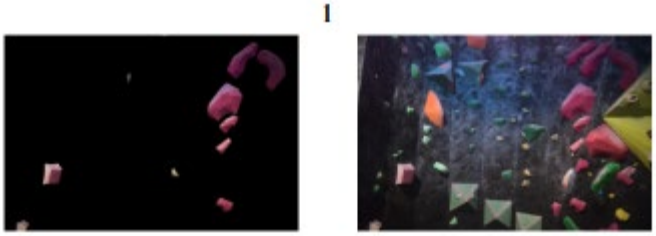

Fig. 22. Siamese network producing a small cluster

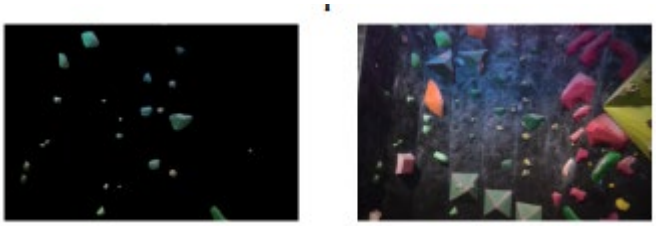

Fig. 23. Triplet network combining light and dark green holds

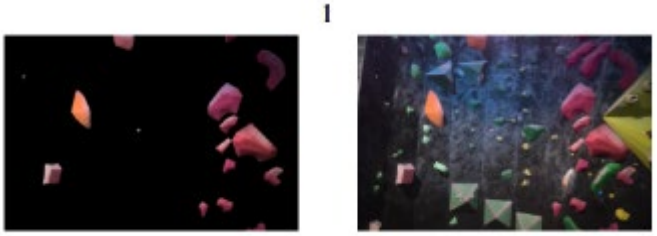

Fig. 24. Triplet network merging holds with subtle color variations

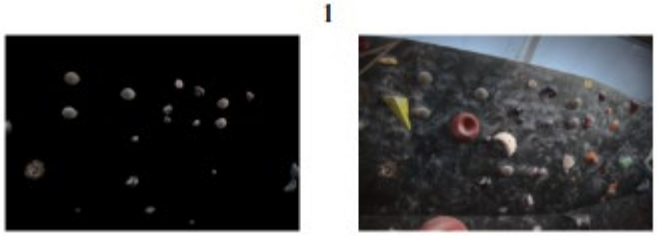

Fig. 25. Triplet network clustering black/gray holds.

Fig. 26. Clustering results for route segmentation using the Siamese and Triplet network approaches (tmed = 0.7, tmax = 2.65)

2) Learning-Based Strategy

a) Reframing the Problem.

Identifying routes on a climbing wall can be recast as a similarity assessment challenge. Instead of treating the task as a straightforward classification based solely on color—a method that fails when colors vary between gyms or holds exhibit multiple hues—the focus shifts to determining whether two hold images belong to the same category based on their visual characteristics.

b) Siamese Network Implementation.

The first network architecture adopts a Siamese configuration (16), which utilizes two parallel pretrained ResNet50 (17) branches with shared weights. The final classification layer is removed, and the outputs from the last layers are concatenated and fed through several fully connected layers, culminating in a single neuron output that is squashed to the range [0, 1] via a Sigmoid activation function. Figure 19 displays an example of the network structure used to compare pairs of holds.

The network is optimized using binary cross-entropy loss with the AdamW optimizer (18). Initially, the network is trained for 100 epochs at a learning rate of $\epsilon = 0.001$, while keeping the weights of the ResNet backbone fixed. Subsequently, the backbone is unfrozen and fine-tuned for an additional 100 epochs at a lower learning rate of $\epsilon = 0.00001$. Sampling is conducted by randomly selecting an annotated image and then choosing holds from the same or different routes, ensuring that holds are not inadvertently mixed with similar-colored holds from other routes.

c) Triplet Network Approach.

A second network design is inspired by FaceNet (19). This approach embeds hold images into a d-dimensional space (with d = 256) while normalizing the embeddings such that $\| f(x) \|_2 = 1$. The triplet loss function is minimized to enforce that the Euclidean distance between an anchor and a positive sample (from the same route) is smaller than that between the anchor and a negative sample (from a different route) by at least a margin α, as described by

$$\|a_i - p_i\|_2 + \alpha < \|a_i - n_i\|_2,$$

where ai, pi, and ni represent the embeddings of the anchor, positive, and negative samples, respectively. A visualization of this loss mechanism is presented in Figure 20.

In a manner similar to the Siamese network, the triplet-based model is trained with the same hyperparameters. Sampling involves choosing two holds from a common route (anchor and positive) and one hold from a distinct route (negative). Triplet mining is applied to focus on challenging examples, although this refinement yielded only incremental benefits. Figure 21 illustrates the overall architecture of the triplet network.

d) Augmentation and Evaluation.

Heavy use of data augmentation (random rotations, horizon- tal and vertical flips, and color adjustments) is crucial given the limited number of annotated training images. For evaluation, the Siamese network is assessed by measuring the accuracy in determining whether pairs of holds belong to the same route. Meanwhile, the triplet network's performance is gauged by counting the proportion of triplets that satisfy the constraint $\|a_i - p_i\|_2 < \|a_i - n_i\|_2$. Table III summarizes the accuracy metrics for both approaches

TABLE III: Accuracy in determining whether two hold images belong to the same route.

Model	Accuracy

Siamese Network	66.95%
Triplet Network	81.74%

e) Route Formation Algorithm.

Routes are constructed by initializing a cluster with a randomly selected hold and then iteratively comparing each unassigned hold against holds in existing routes. If the median distance between the candidate hold and all holds in a route is below a threshold tmed, and the maximum distance does not exceed tmax, the hold is appended to that route. If no matching route is found, a new route is initiated. The resultant clusters are visualized in Figure 26, which demonstrates differences in clustering behavior between the Siamese and Triplet models. Notably, the Siamese network often produces smaller, more homogeneous clusters, while the triplet network tends to generate larger clusters that may inadvertently merge holds of similar hues across routes

3 Findings

A. Hold Segmentation

Conventional hold segmentation delivers acceptable outcomes only when the parameters are meticulously calibrated for each specific dataset. This approach, although free from the burden of extensive manual annotation required by learning- based methods, suffers from the limitation of having to adjust parameters for each new background or hold color variation, often resulting in suboptimal segmentation performance.

In contrast, the learning-based strategy for hold segmentation exhibits markedly superior accuracy and robustness. This method, despite necessitating a manually annotated dataset, specialized hardware (e.g., a CUDA-enabled GPU), and longer training durations, consistently outperforms traditional tech- niques. The increased precision of the deep learning frame- work suggests that the higher resource investment is justified by the significant improvements in detection and delineation of holds.

B. Route Segmentation

The conventional route segmentation method relies on pre- defined clustering parameters, including the predetermined number of clusters, which must be set in advance. Among the various techniques experimented with, mean color clustering emerged as the most effective when handling holds with uniform colors under controlled lighting conditions. However, the conventional methods can struggle when faced with over- lapping routes or inconsistent color distributions.

The machine learning-based approach, leveraging similarity metrics through Siamese and triplet network architectures shows only a modest improvement over traditional clustering methods. Despite the slight edge in performance, there is clear potential for further enhancement by expanding the annotated dataset and extending the training period. Additionally, integrating the spatial context of the entire image—beyond the individual hold—may offer a more nuanced determination of route membership, further refining the accuracy of route segmentation.

4 Conclusion

Both conventional and deep learning-based methodologies were implemented for the dual tasks of hold and route segmentation in indoor climbing environments. For hold segmentation, the traditional technique capitalizes on blob detection for pinpointing hold locations and edge detection for outlining their contours, implemented using OpenCV (10). In contrast, the deep learning approach utilizes the detectron2 implementation of Mask R-CNN (11), which was specifically fine-tuned to address the nuances of climbing hold detection.

Regarding route segmentation, a conventional strategy based on clustering algorithms provided baseline results by grouping holds using features such as mean colors, histograms, and color moments, all processed with tools from OpenCV (10) and scikit-learn (14). The learning-based method, on the other hand, applies a ResNet-based model to generate feature embeddings, ensuring that holds from the same route exhibit minimal L2 distances in the embedding space. Although the improvement in route segmentation accuracy was incremental, it underscores the potential of deep learning methods in capturing subtle visual similarities between holds.

The comparative analysis reveals that while both tasks benefit from deep learning approaches, the tradeoff involves increased demands in terms of manual dataset annotation and computational resources. Nevertheless, the significant improvements in segmentation accuracy justify these investments. The detailed experiments indicate that refining network parameters, expanding the dataset, and incorporating full- image spatial context may further enhance performance.

References

[1] "Kilter Board Website," https://kilterboardapp.com/, ac- cessed: 2023-03-24.

[2] Srinivasa Sai Abhijit Challapalli, Bala kandukuri, Hari Bandireddi, & Jahnavi Pudi. "Profile Face Recognition and Classification Using Multi-Task Cascaded Convolutional Networks," Journal of Computer Allied Intelligence, 2(6), 65-78, 2024. https://doi.org/10.69996/jcai.2024029

[3] A. Dobles, J. C. Sarmiento, and P. Satterthwaite, "Ma- chine Learning Methods for Climbing Route Classifica- tion," Stanford, Tech. Rep., 2017, https://cs229.stanford.edu/proj2017/final-reports/5232206.pdf.

[4] S. Ekaireb, M. A. Khan, P. Pathuri, P. H. Bha- tia, R. Sharma, and N. Manjunath-Murkal, "Com- puter Vision Based Indoor Rock Climbing Analy- sis," University of California San Diego, Tech. Rep., 2022, https://kastner.ucsd.edu/ryan/wp-content/uploads/ sites/5/2022/06/admin/rock-climbing-coach.pdf.

[5] B. Murphy, "CLIMBNET – CNN for detecting + segmenting indoor climbing holds," https://github.com/ cydivision/climbnet, 2020, accessed: 2023-03-23.

[6] E. Wei, "Indoor Rock Climbing Wall Route Displayer," Stanford University, Tech. Rep., 2014, https://stacks. stanford.edu/file/druid:bf950qp8995/Wei.pdf.

[7] A. Dutta, A. Gupta, and A. Zissermann, "VGG im- age annotator (VIA)," http://www.robots.ox.ac.uk/∼vgg/ software/via/, 2016, version: 2.0.12, Accessed: 2023-03-16.

[8] A. Dutta and A. Zisserman, “The VIA annotation software for images, audio and video,” in Proceedings of the 27th ACM International Conference on Multimedia, ser. MM ’19. New York, NY, USA: ACM, 2019. [Online]. Available: https://doi.org/10.1145/3343031.3350535

[9] K. Kisin, P. Go¨ldner, T. Sla´ma, and V. S. Sanjaybhai, “Kaggle: Indoor climbing gym hold segmentation dataset,” https://www.kaggle.com/datasets/tomasslama/indoor-climbing-gym-hold-segmentation, 2023, accessed: 2023-03-29.

[10]Srinivasa Sai Abhijit Challapalli. “Sentiment Analysis of the Twitter Dataset for the Prediction of Sentiments,” *Journal of Sensors, IoT & Health Sciences*, *2*(4), 1-15, 2024. https://doi.org/10.69996/jsihs.2024017

[11]K. He, G. Gkioxari, P. Dolla´r, and R. B. Girshick, “Mask R-CNN,” CoRR, vol. abs/1703.06870, 2017. [Online]. Available: http://arxiv.org/abs/1703.06870

[12]Y. Wu, A. Kirillov, F. Massa, W.-Y. Lo, and R. Girshick, “Detectron2,” https://github.com/facebookresearch/ detectron2, 2019.

[13]T. Lin, M. Maire, S. J. Belongie, L. D. Bourdev, R. B. Girshick, J. Hays, P. Perona, D. Ramanan, P. Dolla´r, and C. L. Zitnick, “Microsoft COCO: common objects in context,” CoRR, vol. abs/1405.0312, 2014. [Online]. Available: http://arxiv.org/abs/1405.0312

[14]Massoud Qasimi. “Personalized Recommendation Intelligent Fuzzy Clustering Model for the Tourism,” *Journal of Computer Allied Intelligence*, *2*(5), 42-53, 2024. https://doi.org/10.69996/jcai.2024024

[15]G. Koch, R. Zemel, and R. Salakhutdinov, “Siamese neural networks for one-shot image recognition,” 2015. [Online]. Available: https://www.cs.cmu.edu/∼rsalakhu/papers/oneshot1.pdf

[16]K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” CoRR, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385

[17]I. Loshchilov and F. Hutter, “Fixing weight decay regu- larization in adam,” CoRR, vol. abs/1711.05101, 2017. [Online]. Available: http://arxiv.org/abs/1711.05101

[18]F. Schroff, D. Kalenichenko, and J. Philbin, “Facenet: A unified embedding for face recognition and clustering,” *CoRR*, vol. abs/1503.03832, 2015. [Online]. Available: http://arxiv.org/abs/1503.03832

[19]Srinivasa Sai Abhijit Challapalli. “Optimizing Dallas-Fort Worth Bus Transportation System Using Any Logic,” *Journal of Sensors, IoT & Health Sciences (JSIHS,ISSN: 2584-2560)*, *2*(4), 40-55, 2024. https://doi.org/10.69996/jsihs.2024020