Cichlid Computer Vision Project – Weekly Progress

Week ending Friday, January 17th, 2025

Time Log

Charlie Clark

What progress did you make?

- **Ran YOLO-SORT on lindenthal videos to generate tracks during the winter break.**
- **Attended weekly admin meetings since the beginning of the winter break.**
- **Assisted with registration questions and needs during the first week of classes.**
- **Completed various managerial tasks related to the smooth and successful launch of the Spring 2025 cohort.**
- **Started collaborating with members of higher ed to successfully run the McGrath and Freeman projects.**
- **Attended multiple meetings associated with the projects and teams I'm serving as point person for.**
- **Submitted an initial methods report for review by Bree.**

What are you planning on working on next?

- **Look into fixing my personal computer for continued use on my BioBoost tasks this weekend, otherwise going to resume research work on Monday using PACE.**
- **Continue serving as point person for McGrath and Freeman projects, helping when necessary and delegating when possible.**
- **Continue literature review.**
- **Attend all necessary weekly meetings (and some non-required ones when I have free time).**
- **Re-submit a methods proposal for review by Bree and Jeanette.**

Is anything blocking you from getting work done?

- **My personal computer's WSL2 environment seems to be having some issues.**
  - **Can't access the Ubuntu distro's CLI (hanging cursor issue).**
  - **Can't restart the WSL2 environment (also hanging cursor issue).**
  - **Was able to copy data from the WSL2 file system to my Windows environment as well as the Dropbox.**
  - **Going to try and debug WSL2 this weekend if I get the chance; if it doesn't work, resuming work next week using PACE.**

Kailey Quesada

1. **What did you do this week?**
   1. **I attended the following meetings:**
      a. **Higher Ed Intro Meeting on January 7th. Got to answer a couple of questions about Higher Ed.**
      b. **Cichlid Team Meeting on January 13th. I am meeting manager, so I created slides.**
      c. **BioBoost Meeting on January 13th. Learned about the new direction of this publication.**
   2. **I registered for the class and submitted the first part of my petition to apply my CS 8903 credits to the CS 6999 requirements.**
   3. **I provided research information from last semester to members of the lizard behavioral analysis project. I messaged a few new and returning members so that we could answer each other's questions.**
   4. **I worked on starting to rewrite the WACV BioBoost paper for Ecological Informatics. Specifically, I converted the references and equations to Word from Overleaf and coppied over the text. I began to carefully read the formatting requirements for Ecological Informatics and examine similar papers provided by Bree.**

1. **What are you going to do next week?**
   1. **I need to work on restructuring the paper to fit**

Eric Iamarino

What did you do this week?

- **Registered for CS 8903**
- **Had initial meetings with Cichlid ReID group**
- **Helped draft first iteration of our groups methods document**
- **Reviewed FaceNet paper that introduced triplet learning**
- **Reviewed two papers that propose different triplet selection strategies than FaceNet**
- **Drafted document outlining 3 approaches Cichlid ReID project can take for their triplet selection strategy**
- **Scheduled meeting for January 17th to correct methods document, and present triplet selection findings**

What are you going to do next week?

- **Look for preexisting implementations of proposed triplet selection strategies (finding linked Githubs, pulling code locally, etc)**
- **Begin processing video footage of cichlids into cropped frames that can be formed into triplets**
- **Confirm I have access to necessary tools for project (PACE, Dropbox, website, etc)**

Blockers, things you want to flag, problems, etc.

- **No current blockers**

Abstracts

Charlie Clark

“”, Shi et al. (submitted to WACV in Fall 2024, no public copies available)

- Abstract: “Object detection is crucial in many domains, from animal tracking to autonomous vehicles. However, certain object classes exhibit substantial visual similarity, posing challenges for accurate classification from single images. Standard object detection models may struggle to achieve high precision in such tasks especially when some classes are underrepresented. We propose a novel annotation methodology and temporal model that combine detection with tracking to generate robust annotation sets from videos for training object detection and classification models with minimal additional annotation effort. The semiautomatic annotation methods leverages tracking to identify multiple instances of individual objects to enrich the annotation data efficiently and achieves a frame accuracy of 96% compared to the 89% accuracy achieved with manual annotation methods. The BioBoost model combines a temporal classifier with our semi-automatic model to achieve a track accuracy of 99.6%, demonstrating superior performance compared to the 93% accuracy achieved when using manual annotation methods, without demanding significant supplementary annotation time. This strategy is particularly advantageous for datasets where object tracking can be used to significantly enhance annotations while minimizing manual effort.”
- AI Summary by ChatGPT (4o): “This paper presents BioBoost, a novel framework for improving object detection and classification in imbalanced datasets with visually similar classes. BioBoost combines semi-automatic annotation and temporal trajectory analysis to enhance dataset quality and model performance. By leveraging object tracking to generate annotations across video frames, the proposed method significantly reduces manual annotation efforts while improving classification accuracy. BioBoost demonstrates superior performance in experimental settings, particularly in applications like animal behavior studies, with a focus on Lake Malawi cichlid fish.”
    - Key Contributions
        - “Semi-Automatic Annotation: Utilizes Simple Online and Realtime Tracking (SORT) with YOLOv5-based object detection to generate annotations efficiently from videos. This process increases the

number of labeled images significantly while maintaining high accuracy."

    - "Temporal Metrics Integration: Introduces a decision tree-based temporal classifier to analyze motion trajectories, incorporating features like speed, acceleration, and travel distance. These temporal metrics help refine classification for ambiguous cases."
    - "BioBoost Framework: Combines the semi-automatic and temporal pipelines to selectively apply trajectory-based corrections in cases of high uncertainty, achieving improved classification performance and balancing imbalanced datasets."
    - "Superior Performance: BioBoost achieves 99.6% track accuracy, outperforming manual annotation and standalone semi-automatic methods. It effectively handles class imbalances and improves robustness to visual variability in datasets."
  - Contributions to Knowledge:
    - "Trajectory-Aided Annotation: By integrating object tracking with temporal features, BioBoost demonstrates how trajectory data can be used to enhance object classification, especially in fine-grained and imbalanced datasets."
    - "Efficiency in Dataset Creation: The semi-automatic approach reduces reliance on manual annotations, addressing scalability issues in domains requiring large-scale labeled data, such as wildlife monitoring and ecological studies."
    - "Handling Class Imbalance: BioBoost provides a practical solution for datasets with imbalanced class distributions by leveraging trajectory data to boost the representation and accuracy of minority classes."
  - Future Research Directions:
    - "Scaling to Diverse Applications: Applying BioBoost to other domains, such as human activity recognition, autonomous vehicles, or medical imaging, could test its adaptability and scalability."
    - "Enhanced Temporal Analysis: Future work could explore more advanced temporal models, such as recurrent neural networks or transformers, for richer trajectory-based insights."
    - "Real-Time Implementation: Developing real-time versions of BioBoost could expand its applicability in dynamic settings, such as live wildlife monitoring or surveillance systems."

- “Generalization to Unseen Classes: Exploring zero-shot or few-shot learning extensions for BioBoost might enable it to generalize better to novel classes or species.”

Kailey Quesada

Zheng, T., et al. “A video object segmentation-based fish individual recognition method for underwater complex environments.” Ecological Informatics, 2024. https://www.sciencedirect.com/science/articl e/pii/S1574954124002310.

Modern aquaculture combines large-scale operations with intelligent technology to reduce labor and increase the survival rates of fish. Accurate recognition of individual fish is crucial for monitoring their health, feeding habits, and overall condition, but current methods struggle in complex underwater environments where visibility is poor and fish may overlap. This paper presents a new approach using video object segmentation and deep learning to improve the accuracy of fish recognition, achieving over 96% accuracy in tests and outperforming existing methods. This model uses Additive Angular Margin Loss. Future work includes validation and optimization of the created method, as well as testing it on more complex scenarios.

Eric Iamarino

Sumbul, G., Ravanbakhsh, M., & Demir, B. (2021). Informative and representative triplet selection for multilabel remote sensing image retrieval. IEEE Transactions on Geoscience and Remote Sensing, 60, 1-11. https://doi.org/10.48550/arXiv.2105.03647

Learning the similarity between remote sensing (RS) images forms the foundation for content-based RS image retrieval (CBIR). Recently, deep metric learning approaches that map the semantic similarity of images into an embedding (metric) space have been found very popular in RS. A common approach for learning the metric space relies on the selection of triplets of similar (positive) and dissimilar (negative) images to a reference image called as an anchor. Choosing triplets is a difficult task particularly for multi-label RS CBIR, where each training image is annotated by multiple class labels. To address this problem, in this paper we propose a novel triplet sampling method in the framework of deep neural networks (DNNs) defined for multilabel RS CBIR problems. The proposed method selects a small set of the most representative and informative triplets based on two main steps. In the first step, a set of anchors that are diverse to each other in the embedding space is selected from the current mini-batch using an iterative algorithm. In the second step, different sets of positive and negative images are chosen for each anchor by evaluating the relevancy, hardness and diversity of the images among each other based on a novel strategy. Experimental results obtained on two multi-label benchmark archives show that the selection of the most informative and representative triplets in the context of DNNs results in: i) reducing the computational complexity of the training phase of the DNNs without any significant loss on the performance; and ii) an increase in learning speed since informative triplets allow fast convergence. The code of the proposed method is publicly available at https://git.tu-berlin.de/rsim/image-retrieval-from-triplets.

Documentation of Work

Charlie Clark

- Ran YOLO-SORT on lindenthal videos to generate tracks during the winter break.
    - See Figure 1 and Section 4, Results Visualization for proof.
- Attended weekly admin meetings since the beginning of the winter break.
- Assisted with registration questions and needs during the first week of classes.
- Completed various managerial tasks related to the smooth and successful launch of the Spring 2025 cohort.
- Started collaborating with members of higher ed to successfully run the McGrath and Freeman projects.
- Attended multiple meetings associated with the projects and teams I'm serving as point person for.
- Submitted an initial methods report for review by Bree.

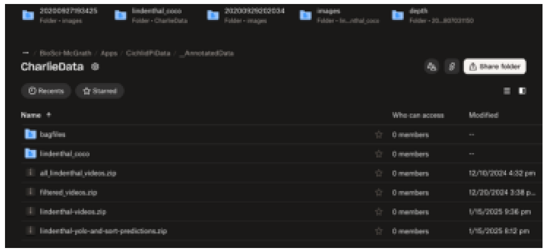


Figure 1: the location in the McGrath Dropbox where the YOLO and SORT predictions are stored.

4. *Results Visualization:*

A link to a YOLO-processed video from the lindenthal dataset: [20200623153306_Color_0.mp4]

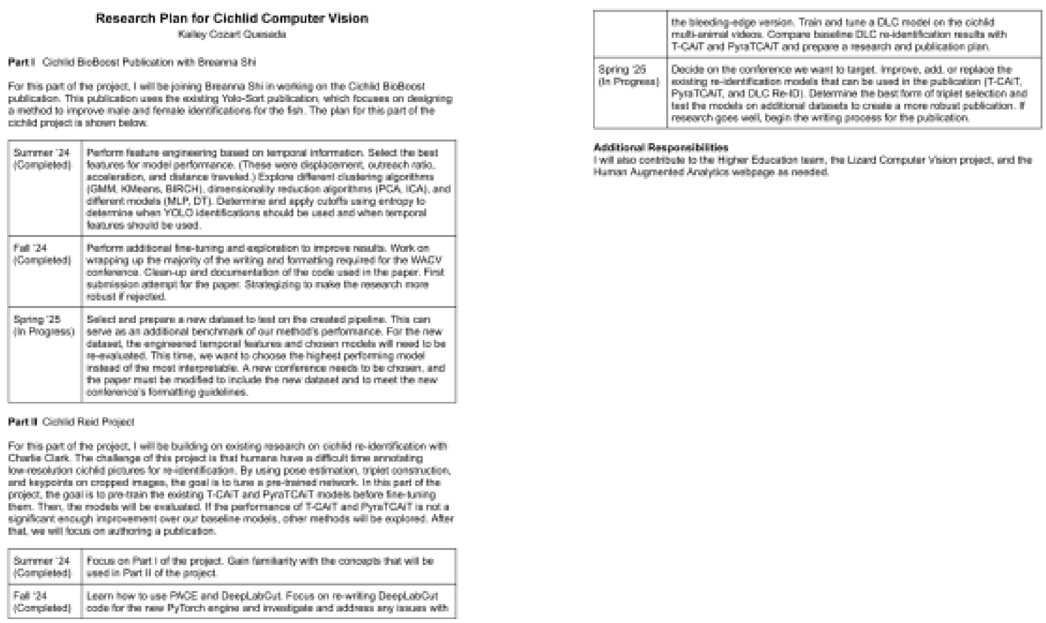
frame	track_id	x_min	y_min	x_max	y_max
1	1	0.557813	0.4888884	0.607031	0.55
2	2	0.954688	0.6236109	1	0.695833
3	2	0.9538394	0.6239711	0.9992847	0.6913072309334158
5	2	0.9519379	0.6223524	1.0000476	0.6942761475629281
6	2	0.9519841	0.6221908	1.0001616	0.6945081293847224
7	2	0.9525966	0.623793	1.0001968	0.6949750445087401
8	2	0.9528691	0.6244608	1.0001507	0.6949824187235241
9	2	0.9530857	0.6233510	1.0000140	0.6924223026854663
10	2	0.9531177	0.6236010	1.0000094	0.6927409621492244
11	2	0.9534627	0.6242002	1.0002032	0.6942377829086245
12	2	0.9540506	0.6240648	1.0003377	0.6942102427191388
13	2	0.9535885	0.6264582	1.0000179	0.696386
14	2	0.9538614	0.625324	0.9999672	0.6927456743983298
15	2	0.952743	0.6229397	0.9996130	0.6914791896152492
16	2	0.9533836	0.6260461	0.9999664	0.6943421650165169
20	2	0.9533858	0.625807	1.0001447	0.6958322686935977
24	2	0.9526850	0.6235537	0.9998587	0.6914976806097919
25	2	0.9533028	0.6224353	1.0001136	0.6899348892971224
26	2	0.9532034	0.623488	1.0000327	0.6915116298269568
27	2	0.952451	0.6248581	0.9996951	0.6920600998524296
28	2	0.9528155	0.6260332	0.9999388	0.6934335159078673
29	2	0.9529840	0.6234649	1.0000031	0.6906129817516881
30	2	0.9530655	0.6216256	1.0000104	0.688593
31	2	0.9531005	0.6230362	1.0000091	0.6902691246136694
32	2	0.9527201	0.6270209	0.9998924	0.6956375825381517

20200622220342_Color_0

Figure 2: an sample of one video's SORT predictions, as stored in a CSV file.

Kailey Quesada

For a full list of what I did, see the time log above. The most important thing I worked on this week was starting to rewrite the WACV BioBoost paper. I started by looking over the general formatting requirements and the three reference papers that Bree provided. From there, I started making the headers and then the sub-headers, copying over the relavent text from the Overleaf. After that, I realized that, in order to use the full bibliography and equation settings in Word, I had to download an use the local desktop version of Word instead of the online version. After installing the free Georgia Tech office suite, I added the equations and the references. As requested by Bree, I removed the fine-grained classification section. We will have to decide on another sub-section for the related work. To see the start of the publication re-write, go to the Publication.docx in the BioBoost folder of Projects: https://gtvault.sharepoint.com/:w:/s/HAAG/ES4iZkwggI1LslNfPp9_oVUBDmWtV6CyYaRwi2s0cuU29A?e=QLG2Zr. An updated copy of the project plan for Cichlid CV that I created for registration and a screenshot of this week's slides are below.



Research Plan for Cichlid Computer Vision
Kailey Cozart Quesada

Part I Cichlid BioBoost Publication with Breanna Shi

For this part of the project, I will be joining Breanna Shi in working on the Cichlid BioBoost publication. This publication uses the existing Yolo-Sort publication, which focuses on designing a method to improve male and female identifications for the fish. The plan for this part of the cichlid project is shown below.

Summer '24 (Completed)	Perform feature engineering based on temporal information. Select the best features for model performance. (These were displacement, outreach ratio, acceleration, and distance traveled.) Explore different clustering algorithms (GMM, KMeans, BIRCH), dimensionality reduction algorithms (PCA, ICA), and different models (MLP, DT). Determine and apply cutoffs using entropy to determine when YOLO identifications should be used and when temporal features should be used.
Fall '24 (Completed)	Perform additional fine-tuning and exploration to improve results. Work on wrapping up the majority of the writing and formatting required for the WACV conference. Clean-up and documentation of the code used in the paper. First submission attempt for the paper. Strategizing to make the research more robust if rejected.
Spring '25 (In Progress)	Select and prepare a new dataset to test on the created pipeline. This can serve as an additional benchmark of our method's performance. For the new dataset, the engineered temporal features and chosen models will need to be re-evaluated. This time, we want to choose the highest performing model instead of the most interpretable. A new conference needs to be chosen, and the paper must be modified to include the new dataset and to meet the new conference's formatting guidelines.

Part II Cichlid Reid Project

For this part of the project, I will be building on existing research on cichlid re-identification with Charlie Clark. The challenge of this project is that humans have a difficult time annotating low-resolution cichlid pictures for re-identification. By using pose estimation, triplet construction, and keypoints on cropped images, the goal is to tune a pre-trained network. In this part of the project, the goal is to pre-train the existing T-CAiT and PyraTCAiT models before fine-tuning them. Then, the models will be evaluated. If the performance of T-CAiT and PyraTCAiT is not a significant enough improvement over our baseline models, other methods will be explored. After that, we will focus on authoring a publication.

Summer '24 (Completed)	Focus on Part I of the project. Gain familiarity with the concepts that will be used in Part II of the project.
Fall '24 (Completed)	Learn how to use PACE and DeepLabCut. Focus on re-writing DeepLabCut code for the new PyTorch engine and investigate and address any issues with the bleeding-edge version. Train and tune a DLC model on the cichlid multi-animal videos. Compare baseline DLC re-identification results with T-CAiT and PyraTCAiT and prepare a research and publication plan.
Spring '25 (In Progress)	Decide on the conference we want to target. Improve, add, or replace the existing re-identification models that can be used in the publication (T-CAiT, PyraTCAiT, and DLC Re-ID). Determine the best form of triplet selection and test the models on additional datasets to create a more robust publication. If research goes well, begin the writing process for the publication.

Additional Responsibilities
I will also contribute to the Higher Education team, the Lizard Computer Vision project, and the Human Augmented Analytics webpage as needed.

Figure 1: Research Plan Screenshot

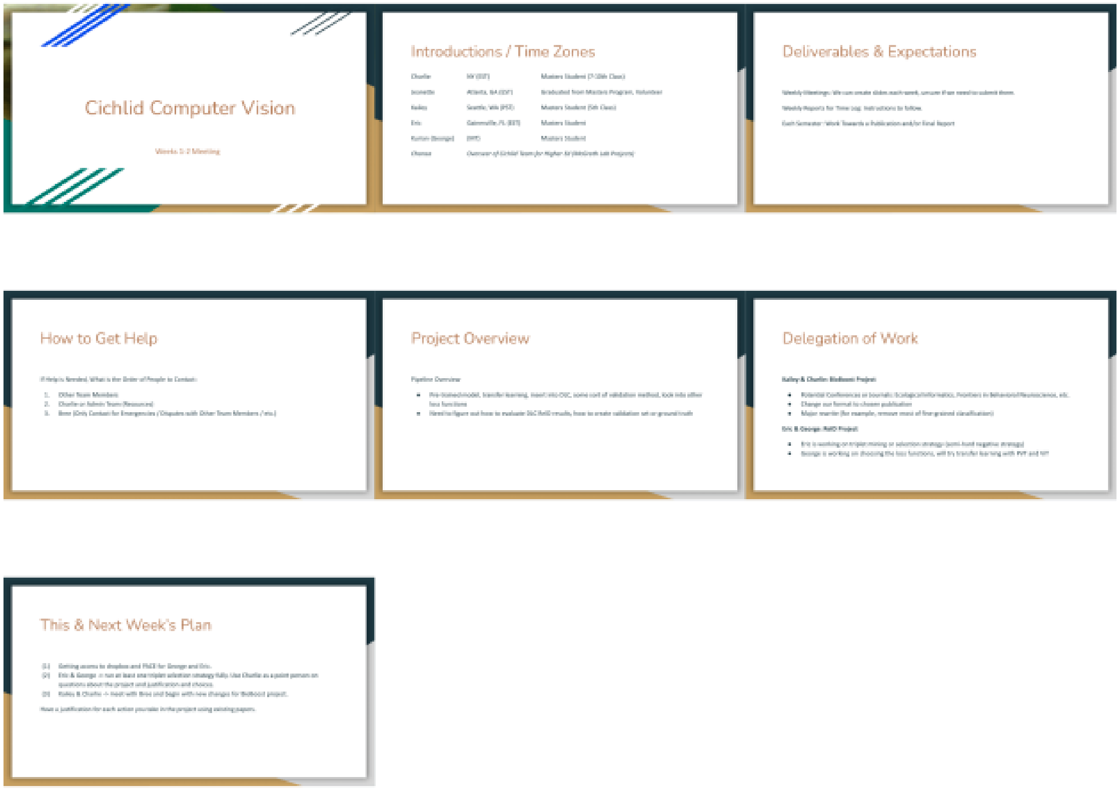


Figure 2: Weekly Meeting Slides Screenshot

Eric Iamarino

This week, I did a deep dive into three research papers (FaceNet, Distance Weighted Sampling, and Relevant and Informative Sampling), which all explored different triplet selection strategies alongside their benefits. From this, I proposed three different triplet selection strategies: two derived directly from the papers and a custom third strategy combining the two approaches. For each strategy, I outlined their goal, provided their mathematical implementations, and included justifications supported by the papers in the form of a document. Most of my time was spent writing this document, as I wanted to ensure we had multiple options to try for the Cichlid ReID project. Screenshots of the document are provided below. Additionally, there were other smaller tasks completed this week that are listed in the Time Log.

Introduction
The goal of Cichlid ReID is to distinguish visually similar cichlids without any sort of ground truth for identifying individual fish. The objective of this report is to select a triplet mining strategy that maximizes the models ability to learn effectively and not plateau.

In this report I will present two papers that showcase triplet selection strategies that show promising results (see Option 1 and Option 2). Additionally, I will propose a custom third strategy that combines the previous two strategies.

Option 1 - Informative & Representative Sampling - https://arxiv.org/pdf/2105.03647
Representative anchor sampling with relevant, hard and diverse positive-negative selection

Anchor Selection

- Goal: Select a small set of the most representative anchors from our samples.
- Proposition: Implement the Diverse Anchor Selection (DAS) strategy proposed in [1], which aims to find the most diverse anchors in the embedding space.
- Implementation: The first anchor is randomly selected from a mini-batch B, and subsequent anchors are selected to maximize their minimum distance from already chosen anchors in the embedding space.

$$X_k = \arg \max_{X_k \in B \setminus A} \left[\max_{X_a \in A} D(X_k, X_a) \right]$$

- Justifications
  - This strategy provided the highest scores under all the metrics compared to *Random Anchor Selection* and *Batch Anchor Selection* [1]
  - Selecting a large and redundant set of triplets increases the computational complexity of the training, which is something this strategy reduces without sacrificing results [1]

Positive-Negative Selection

- Goal: Select novel relevant, hard and diverse positive and negative images
- Proposition: Select positives that are of the same class type as the anchor, but are far a distance in the embedding space. Select negatives that are of a different class type than the anchor, but are closer in the embedding space.
- Implementation: Positive samples are passed through informativeness function which looks at how similar the class and distance in the embedding space of the samples are to one another. Only take positives that are close in class similarity, but far in embedding space distance. Negative samples are passed through a separate informativeness function which is essentially inverted from the positive informativeness function. We want negatives that are in a different class but closer distance to the anchor

$$I_p(X_a, X_b) = \beta \cdot S(X_a, X_b) + (1 - \beta) \cdot D(X_a, X_b)$$

$$I_n(X_a, X_b) = \beta \cdot [1 - S(X_a, X_b)] + (1 - \beta) \cdot [1 - D(X_a, X_b)]$$

- Justifications
  - Ensures the selected positives and negatives are challenging and representative. This accelerates training convergence.
  - Has higher recall compared to random positive-negative sampling [1]

Option 2: Distance Weighted Sampling - https://arxiv.org/pdf/1706.07567
Focus on sampling negatives that are uniformly distributed across a range of distances. Anchor & Positive Selection strategies are referenced from FaceNet paper [2]

Anchor Selection

- Proposition: No explicit anchor selection strategy, anchors are chosen as all available samples in a batch

Positive Selection

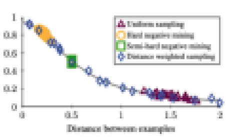
- Proposition: No explicit positive selection strategy, all anchor-positive pairs in a batch are selected.

Negative Selection

- Goal: Sample negatives based on distances from anchors more uniformly than other strategies, increasing variety of sample types and reducing clustering
- Proposition: Sample negatives inversely proportional to their density in the embedding space.
- Implementation: *q(d)*: density of points at a distance d, λ: cap for the sampling weight

$$P(n \mid a) \propto \min(\lambda, q(D_{an})^{-1})$$

- Justifications
  - Prevents oversampling of hard-negatives which destabilizes training, and easy-negatives which do not contribute to training
  - Provides true uniform distribution and variety of samples which leads to improved convergence [3]
  - Semi-hard mining has a thin set of data that can train a model which leads to plateaus in learning, this model overcomes it by providing variety in triplets



Option 3: Custom Mixed Approach
Representative anchor sampling with *RHDIS* positive sampling, and *distance weighted* sampling for negatives

Anchor Selection

- Goal: Select a small set of the most representative anchors from our samples.
- Proposition: Implement the Diverse Anchor Selection (DAS) strategy proposed in [1], which aims to find the most diverse anchors in the embedding space.
- Implementation: The first anchor is randomly selected from a mini-batch B, and subsequent anchors are selected to maximize their minimum distance from already chosen anchors in the embedding space.

$$X_k = \arg \max_{X_k \in B \setminus A} \left[\max_{X_a \in A} D(X_k, X_a) \right]$$

- Justifications
  - This strategy provided the highest scores under all the metrics compared to *Random Anchor Selection* and *Batch Anchor Selection* [1]
  - Selecting a large and redundant set of triplets increases the computational complexity of the training, which is something this strategy reduces [1]

Positive Selection

- Goal: Select novel relevant, hard and diverse positive images
- Proposition: Search for positives that are close in class similarity but far in embedding space distance
- Implementation: Positive samples are passed through informativeness function which looks at how similar the class and embedding space distance of the samples are to one another. Only take positives that are close in class similarity, but far in distance

$$I_p(X_a, X_b) = \beta \cdot S(X_a, X_b) + (1 - \beta) \cdot D(X_a, X_b)$$

- Justifications
  - Ensures positives are not trivially similar, which would cause model to not effectively learn

Negative Selection

- Goal: Sample negatives based on distances from anchors *more* uniformly than other strategies, increasing variety of sample types and reducing clustering
- Proposition: Sample negatives inversely proportional to their density in the embedding space.
- Implementation: *q(d)*: density of points at a distance d, λ: cap for the sampling weight

$$P(n \mid a) \propto \min(\lambda, q(D_{an})^{-1})$$

- Justifications
  - Prevents oversampling of hard-negatives which destabilizes training, and easy-negatives which do not contribute to training
  - Provides true uniform distribution and variety of samples which leads to improved convergence

References
[1] https://arxiv.org/pdf/2105.03647
[2] https://arxiv.org/pdf/1503.03832
[3] https://arxiv.org/pdf/1706.07567