CS 195: One Class Classifier Report

Syed Ashal Ali

PAC Lab, Winter Quarter 2024 (Grade Received: A+)

1 Introduction: What is the Purpose of the One-Class Classifier

With a focus on establishing a data quality control pipeline before creating the VLM for "Echo-GPT," our team proposed building a one-class classifier (OCC) for anomaly detection. The use of OCCs for this domain is not alien: there are many pieces of scientific literature that support the efficacy of implementing an OCC for effective anomaly detection with the main purpose of data quality control. For example, *Bartkowiak AM* in **"Anomaly, Novelty, One-Class Classification: A Comprehensive Introduction"** examines the use of implementing a OCC for detecting anomalous patterns in system calls - Bartkowiak explains the usefulness of this method as "monitoring systems using non-invasive measurements able to signalize that something abnormal starts to happen." Another example of using OCC can be seen in a paper authored by *Mourão-Miranda J et al.*: **Patient Classification as an Outlier Detection Problem: an Application of the One-Class Support Vector Machine**. This study used the support-vector machine (SVM) algorithm in order to detect depressed patients (the outliers) using analyzed fMRI data. Through this literature review, our team was able to decide on using a one-class classifier for data quality control for the VLM by distinguishing scientific professional images (normal case) from non-professional images (anomalous case). The labels for the images were determined by medicine student Wasan Kumar, and cross-checked by Dr. Chieh-Ju Chao of the PAC lab.

2 Our Approach

The role that I played in the data quality control pipeline involved taking the labelled data from Wasan and Dr. Chao, and then using it to build the OCC. The original task I was designated was to fine-tune DINOv2 on the labelled data.

2.1 What is DINOv2?

DINOv2 (Self-distillation with No Labels v2) is a self-supervised vision transformer model that was released by Meta in 2023 and is considered to be state-of-the-art in producing universal features suitable for image-level visual tasks (image classification, instance retrieval, video understanding) as well as pixel-level visual tasks (depth estimation, semantic segmentation).

2.2 Implementation of the OCC

The labelled dataset consisted of 550 images of which 271 were "normal," or professional images, while 279 were "anomalous," or unprofessional images. While the original task involved fine-tuning DINOv2, I felt that this was computationally inefficient. Instead, I came up with the idea of using the base DINOv2 model for feature extraction on the entire professional images dataset (271 images), and then determining the centroid for these features. Then, I determined a distance threshold using the distances of the features from the centroid. This distance threshold was then used to classify whether an image is normal or anomalous. When I shared this approach with Dr. Chao, he seemed satisfied with my interpretation of the task, and said it was "clean and flexible, (with) very low cost." The most challenging part of this implementation was determining the optimal threshold for the classification. An important decision that I had to make was which metric I was going to use to evaluate the different thresholds. Since the original labelled dataset of 550 images had a roughly even split (271 professional vs 279 non-professional), I decided that accuracy was a more appropriate metric to evaluate the threshold by as opposed to evaluation by F1-score. In order to do this, I created a *countOnes* function to apply on the normal images (of all the normal [positive] images, how many were actually classified as normal [true positives]), and a *countZeroes* function to apply on the anomalous images(of all the anomalous [negative] images, how many were actually classified as anomalous [true negatives]). Ultimately, I used these functions and calculated the accuracy. Moreover, I also calculated the false positive rate (from anomalies, how many are marked as normal) because I considered that marking anomalies as normal could be more detrimental to data quality control than marking normal as anomalies and ignoring them (since our actual dataset is very large with 35,000 images). Finally, I intuitively weighed the false positive rate against the accuracy to determine the best threshold.

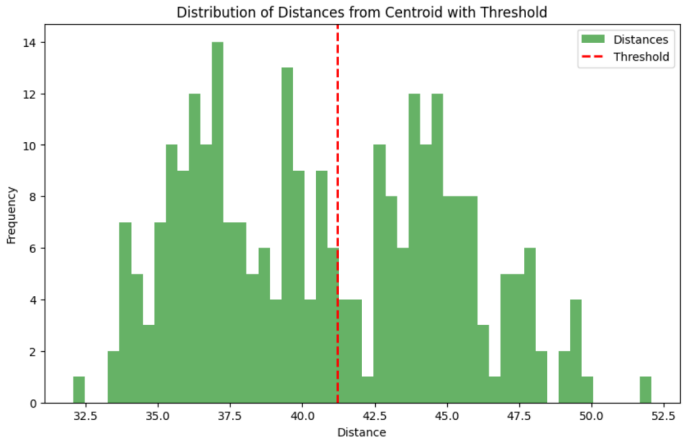

Figure 1: Feature Distances from Centroid

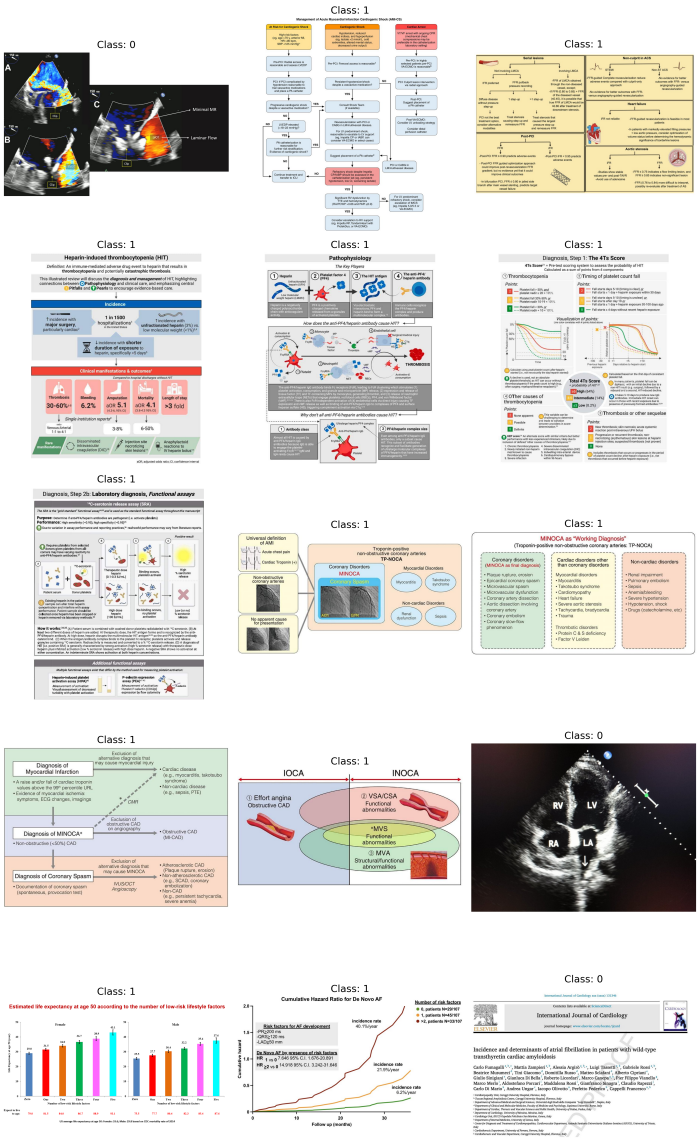

Figure 2: Test Classifications using OCC

2.3 Results

The percentile that gave me the most balanced trade-off between high accuracy and low false positive rate was percentile=55 with an accuracy of 74 percent and a false positive rate of 7.5 percent. This resulted in a threshold distance of 41.2 from the centroid, beyond which images were classified as anomalies. Fig. 1 shows these insights in a plot of the feature distances extracted from the professional images data. The centroid distance obtained after setting percentile to 55 is plotted in red. Moreover, Fig. 2 shows random images from the remaining unlabelled dataset that were unseen by the OCC, and the labels that the OCC determined for them with a class=0 representing an anomaly and a class=1 representing a normal image. I used this to visually test the OCC after I had performed my iterative optimization. In Fig. 2, I believe everything is correctly classified, except the image in the first row and first column, which should be labelled as class=1 (professional), and the image in the fourth row and the third column, which should be labelled as class=1 (professional).

3 Recommendations and Conclusion

While building the OCC, I realized that the training data was quite noisy (no clear delineation as per Fig. 1). My main recommendation to Dr. Chao is working with Wasan to label more data, so that there can be more training data for the model to learn from instead of just 271 professional images. I believe that this would normalize some of the noise in the data and make it easier to delineate a threshold that maximizes the accuracy and reduces the false positive rate. Additionally, I would love to revisit the OCC by discussing other threshold evaluation metrics and testing them out to see if they give better results. For next quarter, I plan to continue working on optimizing the OCC as soon as more data is available while working on the VLM with the rest of the team.