Geometric Transformed Anomaly Detection with Analytical Margin

4 Team Members

Abstract

Anomaly detection is a vigorously studied area in many industries, including the medical industry and construction industry. Classification-based techniques are one of the extensively-used for anomaly detection, but they have struggled with detecting anomalies in high-dimensional data. To detect anomalies in high-dimensional data (e.g., image and tabular data), geometric transformation-based and affine transformation-based approaches have been introduced. Despite the feasibility of transformation-based approaches, it is still challenging to identify crucial feature representations (i.e., transformations), thereby increasing model complexity. It is necessary to identify redundant features and reduce model complexity to process much larger-scale data. Therefore, this project aimed to follow a previous work on anomaly detection in high-dimensional data and improve it with the introduction of an analytical margin for effective model reduction. The authors proposed a geometric-transformation classification based anomaly detector combined with an analytic margin (GEOTADAM) by replacing the original margin term with an analytical margin. The receiver operating characteristic-area under the curve (ROC-AUC) and f1 score were used for the performance evaluation according to the data type. The authors compared the results with the previous one and found that GEOTADAM is better in both CIFAR-10 image data and the thyroid tabular data. Morevoer, the authors tested GEOTADAM with the few-shot condition (training with 16 images from the STL-10 dataset) and showed the better performance than the prior work. The main contribution of this research is to reduce model complexity of transformation-based anomaly detectors and to apply a feature space selection technique in order to enable end-users to detect anomalies in diverse datasets of varying sizes: superlarge-scale and small-scale.

1 Introduction

This project aimed to improve the classification of detecting anomaly data in the general dataset. This project was based on Bergman and Hoshen (2020)'s work introducing the classifier GOAD (GeOmetric-transformation classification-based Anomaly Detection). This method consisted of transforming data into M subspaces and then learning a feature space such that the inter-class separation is larger than the intra-class separation. Also, it expanded its method to classify anomaly data in non-image data, including tabular data through an affine transformation.

Classification-based technique is one of the extensively-used for anomaly detection. Many of prior works have focused on separating normal samples from anomalies or outliers. Nonetheless, it is still challenging to detect anomalies within high-dimensional data (e.g., image) due to high computational complexity. To identify anomalies in such high-dimensional data, Golan and El-Yaniv (2018) introduced the geometric transformation-based classification (GEOM) method and detected anomalies in images. However, this method only detected anomalies in one type of data, which indicates there is still a gap in figuring out anomalies in other types of data, such as tabular data. Therefore, Bergman and Hoshen (2020) extended Golan and El-Yaniv (2018)'s work to handle tabular data with the introduction of the affine transformation.

Submitted to 36th Conference on Neural Information Processing Systems (NeurIPS 2022). Do not distribute.

Despite the availability of GOAD in high-dimensional data, Bergman and Hoshen (2020) mentioned that choosing the margin parameter s is required to improve in future studies since GOAD is not sensitive to the choice of margin parameter s. Namely, finding an optimal transformation and selecting parameter s were required to be considered together before the process of anomaly detection. Therefore, this project improved the work by Bergman and Hoshen (2020) to find a desired transformation that matches the dataset while reducing the model complexity. As the L1 norm shows a good performance on feature selection compared to other p-norms, the authors proposed an improved model harnessing an analytical margin, L1 norm, on behalf of the parameter s. This work will contribute to detect anomalies in diverse datasets of varying sizes: superlarge-scale and small-scale.

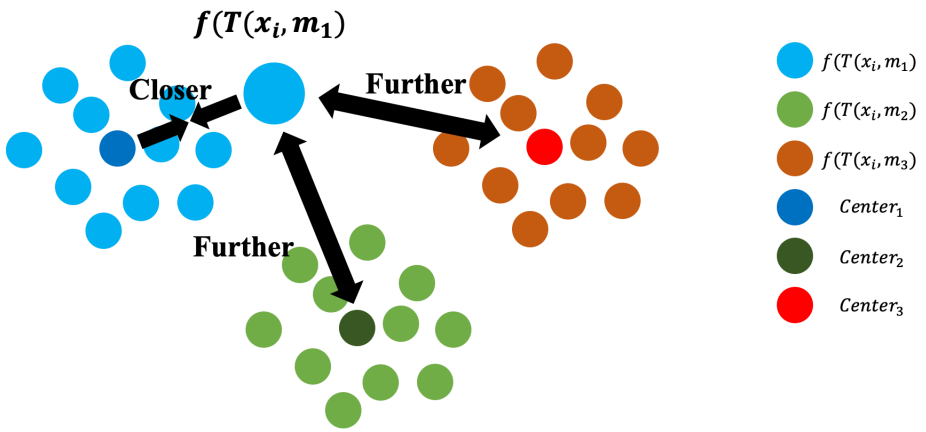
2 Related Work

There are four major techniques in anomaly detection (Chandola et al. (2009)): (1) nearest neighbor-based technique; (2) clustering-based technique; (3) statistical technique; and (4) classification-based technique. Among these techniques, a classification-based technique has garnered attention from researchers for its simplicity and intuitiveness (Chandola et al. (2009)). One-class classification is an example of classification-based techniques. Schölkopf et al. (1999) presented a support vector machine (SVM) returning 1 to normal data and -1 to outliers. They tested with the USPS dataset of handwritten digits, containing 9,298 small images of size $16 \times 16 = 256$. However, such a one-class SVM method can be inappropriate when identifying anomalies in high-dimensional and large data, such as high-resolution images, due to high computational complexity and the curse of dimensionality Ruff et al. (2018).

Recently, to overcome such issues, it has expanded to deep learning approaches from traditional classical kernel methods. Ruff et al. (2018) proposed a deep learning-based support vector data description (Deep SVDD), trained on an anomaly detection objective. Deep SVDD extracts the common features of variations of the data distribution from neural networks and detects outliers using SVDD, which finds a hypersphere enclosing normal samples and minimizes the distance between the center of the hypersphere and the boundary of the data samples. Ruff et al. (2018) tested Deep SVDD with MNIST and CIFAR-10 image datasets and found the most normal samples as well as the most anomalous samples. Another method for detecting anomalies in images is using geometric transformations. Golan and El-Yaniv (2018) described that a classifier, which can discriminate against many types of geometric transformation-applied-normal images, can detect an abnormal image. They first generated M transformed images, including flipping, rotation, reflection, and translation, from a single image and learned transformed features. This enabled them to identify non-geometric transformed images, which will be classified as abnormal images.

Although geometric-transformation classification (GEOM) by Golan and El-Yaniv (2018) performed well in computer vision tasks, it is still challenging to detect anomalies in other types of data, including tabular data. Yoon et al. (2021) presented a self-supervise, refine, repeat (SSR) method to detect anomalies in images as well as tabular data. Their method is based on the iterative training framework, which contains training a one class classifier for subsets of unlabeled data. By iterating training phases, it can be robust for all data types. However, this technique faces high computational complexity when handling a large dataset.

To expand GEOM to tabular datasets, Bergman and Hoshen (2020) developed an anomaly detector for both image and tabular datasets, called GOAD. Bergman and Hoshen (2020) harnessed an affine transformation which includes geometric transformation to transform tabular data. Although GOAD showed the potential of the affine transformation-based approach for anomaly detector in both image and tabular datasets, GOAD could not prove which transformation plays a significant role. Given that there will be a vast amount of computation loads when detecting anomalies in a large-scale dataset, it is necessary to reduce the number of transformations by identifying redundant ones. Muthukrishnan and Rohini (2016) used Least Absolute Shrinkage and Selection Operator (LASSO) regression to identify redundant features while reducing the model complexity. Moreover, the LASSO regression improved both prediction accuracy and model interpretation. Therefore, the authors applied the LASSO term to GOAD for model reduction. In Bergman and Hoshen (2020), GOAD was applied to detect anomalies in the large-scale datasets. However, it is not common to find large-scale datasets in the real world. The authors created the optimal model which can detect anomalies by using a few shots of the dataset.



Triplet Center Loss $L = \sum_i \max\left(\|f(T(x_i,m)) - c_m\|^2 + s - \min\|f(T(x_i,m) - c_{m^\star}\|^2_{m^\star \neq m}, 0\right)$

Modified Center Loss $L_{new} = \sum_i \max\left(\|f(T(x_i,m)) - c_m\|^2 + \lambda\|W\|_1 - \min\|f(T(x_i,m) - c_{m^\star}\|^2_{m^\star \neq m}, 0\right)$

Figure 1: Original loss and the proposed loss

3 Methodology

Bergman and Hoshen (2020) noticed that GOAD is not significantly sensitive to the fixed value of margin parameter s, indicating that there may be high correlations or orthogonality between the input features. Therefore, the proposed method utilized the LASSO term $\lambda||W||_1$, called an analytic margin, to impose high weights on crucial features in predicting anomalous data. The following paragraph describes the proposed method step by step.

Assuming that all data is located in the space R^L. In this space, normal data is positioned in the subspace $X \subset R^L$, while all anomalies lie outside X. The authors made a classifier such that $C(x) = 1$ if $x \in X$ and $C(x) = 0$ if $x \in R^L \backslash X$. Similarly to the prior anomaly detector, GOAD by Bergman and Hoshen (2020), X is transformed to $X_1, ..., X_M$ using M transformations. The authors used affine transformations, that is $T(x, m) = W_m x + b_m$. To generate feature representations from the input data, the authors employed two neural networks: (1) WideResNet by Zagoruyko and Komodakis (2017) for image datasets and (2) fully convolutional networks for tabular datasets. Each subspace X_m is projected to the feature space $f(x)|x \in X_m$ as a sphere with center c_m. The center of a sphere c_m can be computed by averaging features over the training set for each transformation i.e. $c_m = \frac{1}{N}\sum_{x \in X} f(T(x, m))$. The probability of data point x after transformation m is in $X^\star_m$ is parameterized by

$$P(T(x, m) \in X_{m^\star}) = \frac{1}{Z} e^{-||f(T(x,m)) - c_{m^\star}||^2}$$

where $m^\star$ and m are different transformations. Our classifier predicts that the transformed point $T(x, m)$ is transformed by $m^\star$ with the probability

$$P(m^\star|T(x, m)) = \frac{e^{-||f(T(x,m)) - c_{m^\star}||^2}}{\sum_{\hat{m}} e^{-||f(T(x,m)) - c_{\hat{m}}||^2}}$$

due to the Bayes' rule and the fact that every training sample is transformed exactly once by each transformation. The authors also used the triplet center loss function introduced by He et al. (2018) to minimize the low intra-class variation while maximizing inter-class variation, identical to GOAD's loss function. The triplet center loss applied in GOAD is as below.

$$L = \sum_i max(||f(T(x_i, m)) - c_m||^2 + s - min_{m^\star \neq m}||f(T(x_i, m)) - c_{m^\star}||^2, 0)$$

However, as there are potential correlations and orthogonality in the input features, the proposed method harnessed the LASSO term as the analytic margin parameter. Consequently, the modified

triplet center loss is the following.

$$L = \sum_i max(||f(T(x_i, m)) - c_m||^2 + \lambda||W||_1 - min_{m^\star \neq m}||f(T(x_i, m)) - c_{m^\star}||^2, 0)$$

where W denotes a vector whose elements are all parameters in neural networks related to feature representations and affine transformations.
The details of the triplet center loss and modified triplet center loss are visualized in Figure 1.

4 Experiments and Results

4.1 Code Implementation and Dataset

The authors aimed to detect anomalous data points in a dataset with k classes $C_1, C_2, \ldots C_k$, with a focus on identifying anomalies with respect to class C_1. To accomplish this, the authors utilized the sci-kit-learn library to randomly divide the dataset into a training set and a test set. The training set consisted of a random subset of images that belonged to class C_1, while the test set comprised a random subset of images drawn from all the classes $C_1, C_2, \ldots C_k$. For each image in the test set, the authors examined whether the model could correctly classify it as a member of C_1 or an anomaly. This process framed the supervised anomaly detection problem as an imbalanced binary classification task, with the authors seeking to identify the rare class of anomalous data points. The authors leveraged transformation functions based on prior work by Bergman and Hoshen (2020), which included rotations, reflections, and other transformations. However, unlike the previous study, the authors did not use transition-like transformations due to the limited computational capacity. The authors also extracted transformation features that were related to the dataset using an analytical margin. This approach allowed the authors to identify and utilize the most relevant transformation features for the given dataset, thereby improving the model's accuracy and reducing computational complexity. Overall, the authors' method demonstrated the effectiveness of using transformation functions and feature extraction to improve supervised anomaly detection performance in imbalanced datasets.

Algorithm 1: GEOTADAM: Training Algorithm

Input: Normal training data $x_1, x_2 ... x_N$, Transformations $T(,1), T(,2), \ldots, T(,M)$, the set of transformations $\mathcal{T}$
Output: Feature extractor f, centers $c_1, c_2 ... c_M$, $\mathcal{T} \leftarrow x_i$
Find $f, \lambda, c_1, c_2, ..., c_M$ that optimize the modified center triplet loss

---

Algorithm 2: GEOTADAM: Evaluation Algorithm

Input: Test sample x, feature extractor f, centers $c_1, c_2 ... c_M$, the set of transformations $\mathcal{T}$
Output: $Score(x)$, $\mathcal{T} \leftarrow x$
Derive T' // Find the transformation T' which corresponds to the data
$P(m|T'(x, m)) \leftarrow f(T'(x, m))$, λ, $c_1, c_2 ... c_M$ // Likelihood of predicting the correct transformation
$Score(x) \leftarrow P(1|T'(x, 1)), P(2|T'(x, 2)) ... P(M|T'(x, M))$ // Compute anomaly score

---

Algorithm 3: Selecting best λ For CIFAR-10 and STL-10

```
for lambda in [1.00E-06, 3.00E-06, 6.00E-06, 1.00E-05, 1.00E-04, 1.00E-03,
    1.00E-02, 1]
    roc_auc_scores_for_each_class = []
    for class i in [0...9]
        Consider class i as normal and train on class i
        Test on all classes and compute roc-auc score
        Append roc_auc_score_this_class to roc_auc_scores_for_each_class
    Compute average of roc_auc_scores_for_each_class-auc
Find lambda with highest average roc-auc score
```

Algorithm 4: Selecting best λ For Thyroid

```
for lambda in [1.00E-06, 3.00E-06, 6.00E-06, 1.00E-05, 1.00E-04, 1.00E-03,
    1.00E-02, 1]
    Consider healthy patients as normal and thyroid disease as anomalous
    Calculate f1 score
Find lambda with highest average f1 score
```

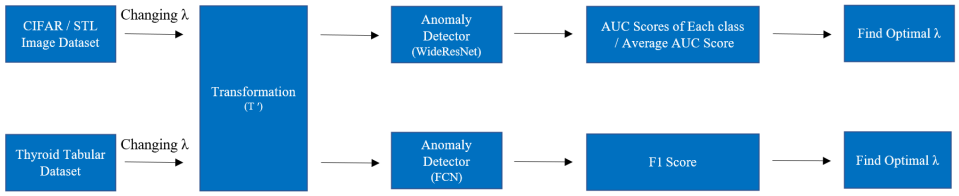


Figure 2: Overall research flow chart

4.2 CIFAR-10 Dataset

The authors used the CIFAR-10 dataset, which consists of small $32 \times 32 \times 3$ pixel images that are from one of these 10 classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck. For each of the 10 classes, the authors trained GEOTADAM on data from only one of the classes, and then the authors evaluated GEOTADAM on data from all classes. The single class that GEOTADAM is trained on is considered to be normal training data; normal data means there are no anomalies. The testing data consists of normal and anomalous data, and the authors evaluated whether GEOTADAM is able to detect the anomalies. The authors trained GEOTADAM for 16 epochs because training for more epochs did not yield improvements in accuracy and would be at risk of overfitting. The authors averaged the 10 ROC-AUC scores obtained when each of the 10 classes is considered to be normal. For each value of λ, the authors computed the average ROC-AUC score and selected the value of λ that maximizes the average ROC-AUC score.

4.3 Thyroid Tabular Dataset

The authors also evaluated our model on thyroid tabular data, which is from the UCI repository Dua and Graff (2017). The thyroid tabular data has 29 attributes for each patient such as age, gender, whether they have been prescribed thyroxine, TSH blood test reading, etc. Each patient also has a binary label indicating whether they have thyroid disease. For each value of λ, the authors computed the f1 score. The rationale behind selecting the f1 score instead of different metrics is that the authors aimed to minimize false negatives considering the nature of disease datasets. A false negative outcome signifies that a patient actually has thyroid disease but will not receive any treatments.

4.4 Novel STL-10 Dataset

For the new test environment, the authors compared our GEOTADAM with GOAD with the novel test dataset, STL-10. The STL-10 dataset was built by Coates et al. (2011) and contained 10 classes: airplane, bird, car, cat, deer, dog, horse, monkey, ship, and truck. Each class has 500 training images and 800 test images, and the image size is $96 \times 96 \times 3$. The authors followed the identical experiment procedure to the test with CIFAR-10. However, the authors only extracted 16 images for each class from the training dataset for training and resized images to $32 \times 32 \times 3$. The purpose of this test was to determine whether GEOTADAM could accurately detect anomalous samples despite being trained on a very limited sample size. Figure 3 displayed the examples of CIFAR-10 and STL-10 datasets.

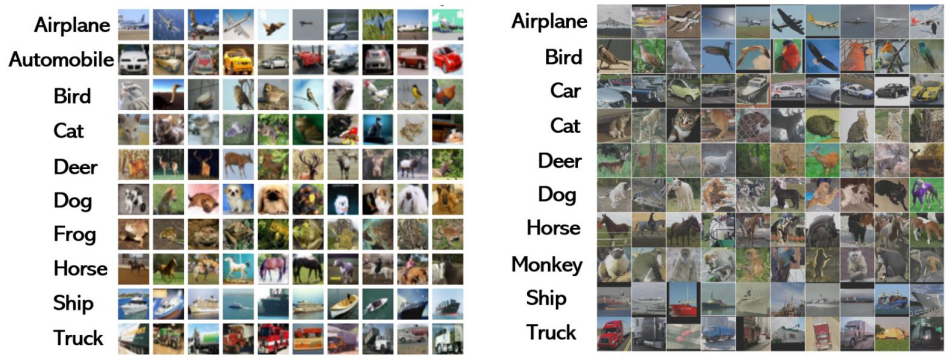


Figure 3: Examples of CIFAR-10 (left) and STL-10 (right)

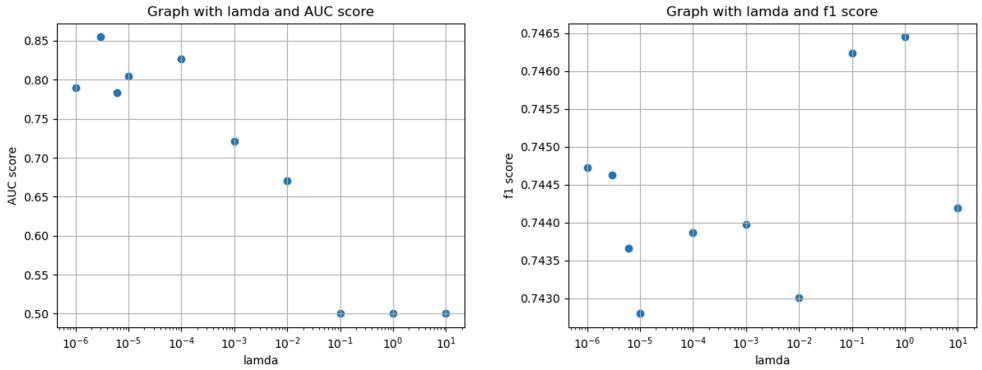


Figure 4: AUC scores and F1 scores depending on λ; the left graph shows the average ROC-AUC score depending on λ for CIFAR-10 dataset, while the right graph shows the average F1 score depending on λ for the thyroid tabular dataset.

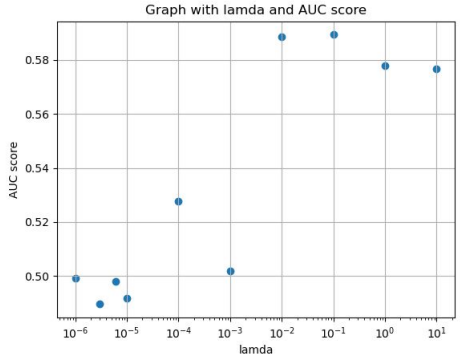


Figure 5: Average ROC-AUC score depending on λ for the novel STL-10 dataset.

Before training on a couple of datasets, the authors took a small subset of the training data consisting of about 100 examples to verify that our algorithms were implemented accurately. The authors also did not initially apply transformations on the data so that they could quickly test changes. Later, the authors used an expanded dataset and GPUs for efficient computation. The authors experimented with different λ values that are multiples of 3 or 10 higher than the previous value. Using this way,

the authors covered a sufficient spread of λ values and observed how λ affects the results. As the authors experimented with different lambda values, there is an increase in ROC-AUC or f1 score and then a decrease. The authors stopped experimenting with λ values when there was a clear trend that increasing λ would not improve the model's performance.

4.5 Evaluation

To determine whether a data point is anomalous, the authors calculated the total score as the follow formula.

$$Score(x) = -\log P(x \in X) = -\sum_m \log \tilde{P}(T(x,m) \in X_m) = -\sum_m \log \tilde{P}(m|T(x,m))$$

(x : Sample, $T(x,m)$: Transformed Sample, X_m : Subspace corresponding to Transformation m)

It gives the degree of anomaly of the data sample i.e. higher score indicate a more anomalous sample. Figure 4 visualizes the results when testing with the CIFAR-10 and thyroid tabular dataset. When testing with the CIFAR-10 dataset, GEOTADAM outperformed the target ROC-AUC 0.84, which was Bergman and Hoshen (2020)'s ROC-AUC averaged across all classes, where $\lambda = 3 \times 10^{-6}$. In contrast, GEOTADAM showed the better result than GOAD when testing with the thyroid tabular dataset where $\lambda = 1$. Figure 5 shows the ROC-AUC scores of GEOTADAM when testing with the STL-10 dataset. GOAD recorded less than the ROC-AUC score of 0.5, whereas our GEOTADAM showed the ROC-AUC score of approximately 0.6.

5 Discussion

The authors roughly found the optimal regularization parameter λ^* which gives the highest score (ROC-AUC score for CIFAR-10 data, STL-10 data and F1-score for the thyroid tabular data) after replacing the margin parameter s from the GOAD algorithm. λ^* is 1 for the thyroid tabular data, 3×10^{-6} for CIFAR-10 data, and 10^{-1} for the STL-10 data. By utilizing the λ^* and the corresponding LASSO regression, our model, GEOTADAM, showed the better average F1-score for the thyroid tabular data, and the better average ROC-AUC score for CIFAR-10 data and STL-10 data. The results showed that our GEOTADAM algorithm slightly outperformed the GOAD algorithm in detecting anomalous data for the CIFAR-10 data and thyroid tabular data, even though GOAD had already improved upon the strongest baseline methods such as Deep SVDD Ruff et al. (2018) and GEOM Golan and El-Yaniv (2018).

Additionally, our GEOTADAM algorithm outperformed the GOAD algorithm in detecting anomalous data for the novel STL-10 data, despite the fact that the GOAD algorithm produced an average ROC-AUC score of less than 0.5. GOAD's performance were worse than random guessing for anomaly detection. Also, as the authors learned that LASSO regression yields a sparse solution from EECS 553, GEOTADAM algorithm automatically performed the feature space selection (corresponding to a single affine transformation) and model reduction. Interestingly, λ^* is much larger for thyroid tabular data and STL-10 data than for CIFAR-10 data. This result indicated that the number of feature spaces, each of which corresponds to a single affine transformation directly associated with the anomaly detection in the thyroid tabular data is smaller than that in CIFAR-10 data. The authors conjectured that the feature space selection (corresponding to a single affine transformation) is one of the reasons why our GEOTADAM algorithm outperformed the GOAD algorithm in terms of ROC-AUC score for the STL-10 data, despite using only 16 extracted images.

6 Conclusions

In this paper, the authors improved the GOAD method by introducing L1 norm (LASSO) as an analytic margin. First, the authors proposed a novel anomaly detector, GEOTADAM, for both the image data and the tabular data with a boost from leveraging an analytic margin. Second, the authors tested GEOTADAM with a novel dataset with the few-shot condition. Third, the authors showed the potential of model reduction by the analytic margin. Consequently,
(1) GEOTADAM with $\lambda = 3 \times 10^{-6}$ slightly outperformed GOAD (the average ROC-AUC score 0.8538 > 0.8403) in the CIFAR-10 dataset,

(2) GEOTADAM with $\lambda = 1$ showed a better f1 score of 0.7465 compared to the f1 score of 0.7440 by GOAD, and
(3) GEOTADAM with $\lambda = 10^{-1}$ outperformed GOAD (the average ROC-AUC score 0.5894 > 0.4922).

The authors chose not to use the L2 norm (Ridge regression) in GEOTADAM since the authors anticipated that GEOTADAM would perform feature space selection (corresponding to a single affine transformation). Ridge regression was deemed unsuitable for this purpose as it is not well-suited for feature selection. Muthukrishnan and Rohini (2016) stated that the ridge regression only makes the solution approach zero, rather than precisely zero, which does not lend itself well to feature selection.

The limitation of this study is that the λ selection strategy for optimal transformation is significantly time-consuming and brute-force. The authors selected the optimal λ by grid search, which is one of the heuristic processes. Another way to find optimal λ is k-fold cross-validation choosing the value of λ that gives the best average performance across all folds. Also, the regularization path is the other way to find optimal λ by showing a plot according to λ. It can provide insights into how the model parameters change with different values of λ. As the above result showed, the value of λ significantly affects the accuracy of the model. Therefore, it is necessary to figure out an appropriate λ with less effort.

Apart from finding out λ quickly, the authors have the plan to apply GEOTADAM to other datasets for further research. In this research, the authors used the CIFAR-10 image dataset and the Thyroid tabular dataset to compare the performance of the GEOTADAM algorithm and the GOAD algorithm. Also, the authors used a few shots in the STL-10 dataset and showed that GEOTADAM algorithm worked well for not enough datasets compared to the initial GOAD algorithm. These better performances indicates our proposed algorithm, GEOTADAM can be applicable to other datasets. To demonstrate our algorithm to other types of dataset, it will be needed to detect anomalies in diverse types of datasets (e.g., text and graph dataset).

7 Contribution by Group Members

The authors discussed how to improve the former algorithm through model selection. They introduced L1 Lasso, which they learned in class. Authors 1 and 2 wrote the code according to the algorithm specified by the original paper. Authors 3 and 4 considered potential improvements to the original paper's algorithm and implemented the lasso modification. Authors 1 and 3 considered ways to improve the efficiency of the code and reduce runtime. Authors 2 and 4 debugged the code to represent the GEOTADAM algorithm. Authors 3 and 4 applied the original algorithm and the new algorithm to the STL-10 dataset, which is a novel dataset on which these algorithms have not previously been applied.

References

Bergman, L. and Hoshen, Y. (2020). Classification-based anomaly detection for general data. In *International Conference on Learning Representations (ICLR)*.

Chandola, V., Banerjee, A., and Kumar, V. (2009). Anomaly detection: A survey. *ACM Comput. Surv.*, 41(3).

Coates, A., Ng, A., and Lee, H. (2011). An analysis of single-layer networks in unsupervised feature learning. In Gordon, G., Dunson, D., and Dudík, M., editors, *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, volume 15 of *Proceedings of Machine Learning Research*, pages 215–223, Fort Lauderdale, FL, USA. PMLR.

Dua, D. and Graff, C. (2017). UCI machine learning repository.

Golan, I. and El-Yaniv, R. (2018). Deep anomaly detection using geometric transformations. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 9781–9791, Red Hook, NY, USA. Curran Associates Inc.

He, X., Zhou, Y., Zhou, Z., Bai, S., and Bai, X. (2018). Triplet-center loss for multi-view 3d object retrieval. *CoRR*, abs/1803.06189.

Muthukrishnan, R. and Rohini, R. (2016). Lasso: A feature selection technique in predictive modeling for machine learning. pages 18–20.

Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S. A., Binder, A., Müller, E., and Kloft, M. (2018). Deep one-class classification. In Dy, J. and Krause, A., editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 4393–4402. PMLR.

Schölkopf, B., Williamson, R. C., Smola, A., Shawe-Taylor, J., and Platt, J. (1999). Support vector method for novelty detection. In Solla, S., Leen, T., and Müller, K., editors, *Advances in Neural Information Processing Systems*, volume 12. MIT Press.

Yoon, J., Sohn, K., Li, C.-L., Arik, S. Ö., Lee, C.-Y., and Pfister, T. (2021). Self-supervise, refine, repeat: Improving unsupervised anomaly detection.

Zagoruyko, S. and Komodakis, N. (2017). Wide residual networks.