One-Class Classification for Anomaly Detection on CIFAR-10

1st Ahsen Beyza Ozkul
dept.Artificial Intelligence and Data Engineering
Istanbul Technical University
Istanbul, Turkey
ozkula21@itu.edu.tr

2nd Tesnime Jemmazi
dept.Artificial Intelligence and Data Engineering
Istanbul Technical University
Istanbul, Turkey
jemmazi22@itu.edu.tr

***Abstract*—Anomaly detection identifies abnormal patterns based only on normal data. This project applies a one-class classification to the CIFAR-10 dataset, with airplane images treated as normal and others as anomalies. We train a ResNet18 model using deep SVDD to group normal data in a high-density feature space.**

***Index Terms*—Anomaly detection, One-class classification, Deep SVDD, CIFAR-10, One-Class SVM, Isolation Forest, ResNet18, Image classification.**

I. Team Work Distribution

Both team members were responsible for implementing the Deep SVDD model, conducting experiments, and analyzing results due to the long runtime of the notebooks, which necessitated trying different approaches simultaneously. Ahsen Beyza Ozkul additionally focused on data preprocessing and contributed to the literature survey. Tesnime Jemmazi also handled baseline model implementation (One-Class SVM and Isolation Forest) and contributed to report writing. Both collaborated on designing the experimental setup and interpreting the findings.

II. Problem Statement

Anomaly detection is one of the core machine learning tasks that involve identifying unusual patterns when only normal data is available for training. This is common in cases where deviations, like rare errors or rare events, must be signaled without examples. Given the CIFAR-10 dataset, we would like to model the distribution of airplane images as normal and recognize images from other classes (e.g., cars, dogs) as anomalies. CIFAR-10's simple structure makes it an excellent option for testing anomaly detection methods, simulating real-world usage where typical patterns are known but deviations are diverse and unknown.

III. Hypothesis

We hypothesize that thanks to its automatic feature extraction, Deep SVDD-trained ResNet18 will outperform One-Class SVM and Isolation Forest in detecting CIFAR-10 anomalies by effectively modeling airplane images as the normal class.

IV. Literature Survey

One-class classification (OCC) is a well-studied approach to anomaly detection. Deep SVDD learns a neural network that maps normal data near a center in feature space and achieves very strong performance on datasets like CIFAR-10. [1] One-Class SVM learns a boundary around normal data but struggles with high-dimensional, intricate images. [2] Isolation Forest detects anomalies by randomly partitioning features, which works well for less complex data but not as well for images without preprocessing. [3] Vision Transformers are promising for anomaly detection since they learn global image patterns but need large datasets and extra computation, so they are less suitable for CIFAR-10. [4] These techniques illustrate that deep learning techniques such as Deep SVDD cope with sophisticated data more effectively, whereas classical approaches are simpler but less versatile.

V. Method(s)

We employ Deep SVDD, an anomaly detection algorithm that integrates Support Vector Data Description with deep learning principles. Unlike traditional SVDD or One-Class SVM, which use kernel methods and require manual feature engineering, Deep SVDD learns a neural network to map input data into a latent space where normal data points cluster tightly around a fixed center, and anomalies lie farther away. This approach leverages representation learning to automatically extract features, making it well-suited for high-dimensional datasets like CIFAR-10 and real-world problems requiring only normal data for training.

The Deep SVDD objective is to minimize the following loss function:

$$\min_{W} \frac{1}{N} \sum_{i=1}^{N} \|\phi(x_i; W) - c\|^2 + \lambda \sum_{l=1}^{L} \left\|W^l\right\|_F^2, \qquad (1)$$

where $\phi(x_i; W)$ is the neural network mapping of input x_i, c is a fixed center in the latent space, W represents the network weights, λ is the regularization parameter, and $\|\cdot\|_F$ denotes the Frobenius norm. The first term encourages normal data to cluster around c, while the second term regularizes the network weights to prevent overfitting.

We use a modified ResNet18 as the feature extractor, pretrained on ImageNet and adapted by removing the final classification layer. The encoder outputs a 512-dimensional feature vector, which is passed through a custom fully connected layer to produce a 128-dimensional latent representation. The model is trained on airplane images (normal class) for 100 epochs using the Adam optimizer with a learning rate of 0.001 and weight decay of 10^{-6}. The center c is initialized as the mean of the network outputs over the training data, and the radius is set as the $(1-\nu)$-quantile of distances from c, where $\nu = 0.1$ represents the expected outlier fraction.

For comparison, we implement One-Class SVM and Isolation Forest using features extracted from the same ResNet18 encoder (before the final layer). Features are standardized using a StandardScaler. One-Class SVM uses an RBF kernel with $\nu = 0.1$ and gamma set to 'scale'. Isolation Forest is configured with a contamination parameter of 0.1 and a random seed for reproducibility. These baselines rely on pre-extracted features, unlike Deep SVDD's end-to-end learning approach.

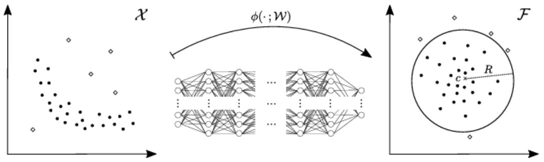

Figure 1. Deep SVDD learns a neural network transformation $\phi(\cdot\,;\mathcal{W})$ with weights $\mathcal{W}$ from input space $\mathcal{X} \subseteq \mathbb{R}^d$ to output space $\mathcal{F} \subseteq \mathbb{R}^p$ that attempts to map most of the data network representations into a hypersphere characterized by center c and radius R of minimum volume. Mappings of normal examples fall within, whereas mappings of anomalies fall outside the hypersphere.

Fig. 1: Illustration of Deep SVDD architecture, mapping normal data to a compact latent space around a center c. Adapted from [1].

VI. Data

The CIFAR-10 dataset comprises 60,000 32×32 color images across 10 classes: airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck, with 6,000 images per class. For this project, we designate the airplane class as normal, using 5,000 airplane images for training and 1,000 for testing. The remaining 9 classes (54,000 images) serve as anomalies during testing to evaluate the model's ability to distinguish normal from anomalous data. The dataset's balanced class distribution and moderate image complexity make it suitable for benchmarking anomaly detection algorithms.

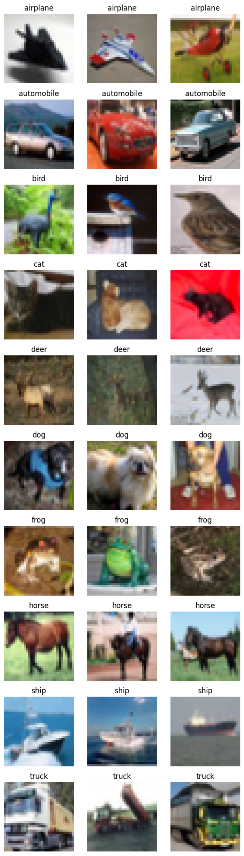

Fig. 3: sample images from CIFAR-10

Training data is preprocessed with random horizontal flips, random crops (padding=4), and normalization using mean (0.485, 0.456, 0.406) and standard deviation (0.229, 0.224, 0.225). Test data undergoes only normalization to maintain consistency. Data loaders are created with a batch size of 128, shuffling enabled for training, and disabled for testing to ensure reproducible evaluation.

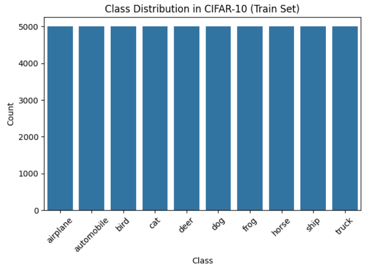

Fig. 2: class distribution in CIFAR-10

VII. Results

The performance of Deep SVDD, One-Class SVM, and Isolation Forest is evaluated using AUC-ROC on a test set of 1,000 normal (airplane) and 9,000 anomaly samples. Table I summarizes AUC-ROC, precision, and recall for anomalies (class 1.0).

TABLE I: Performance comparison of anomaly detection models on CIFAR-10.

Model	**AUC**	**Prec.**	**Rec.**
Deep SVDD	0.4715	0.61	0.10
One-Class SVM	0.7145	0.93	0.81
Isolation Forest	0.7057	0.93	0.73

One-Class SVM achieves the highest AUC-ROC (0.7145), followed by Isolation Forest (0.7057). Deep SVDD performs poorly (AUC-ROC: 0.4715), with low precision (0.61) and recall (0.10) for anomalies, indicating it struggles to distinguish anomalies effectively. One-Class SVM and Isolation Forest show higher precision (0.93) and recall (0.81 and 0.73, respectively), better handling the high-dimensional feature space.

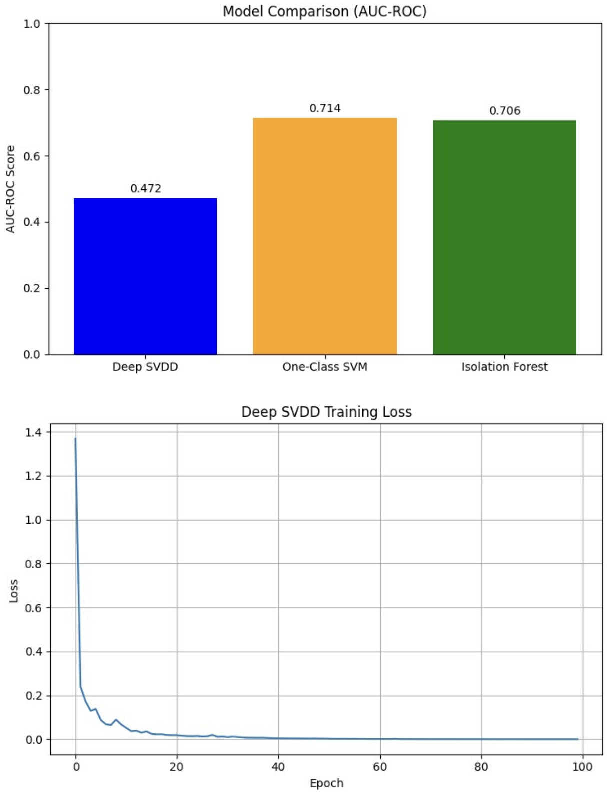

Fig. 4: (a) AUC-ROC scores comparing Deep SVDD, One-Class SVM, and Isolation Forest. (b) Deep SVDD training loss over 100 epochs, converging to a low value.

Sample predictions show Deep SVDD misclassifying many anomalies as normal, reflected in its low recall. One-Class SVM and Isolation Forest better discriminate between classes, though their performance remains moderate.

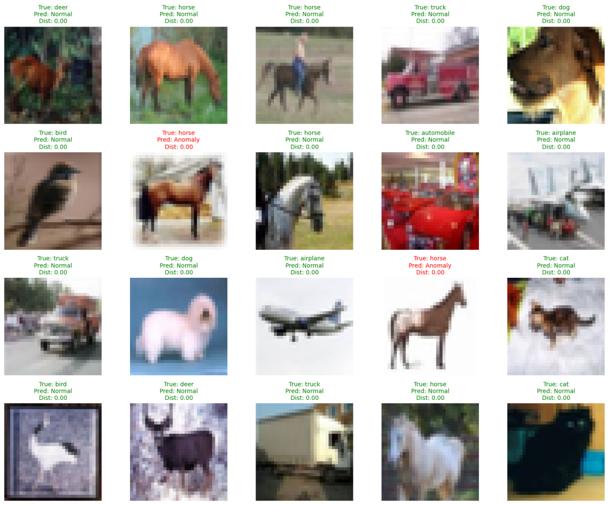

Fig. 5: Result Samples

VIII. Discussion and Conclusions

Contrary to our hypothesis, Deep SVDD with ResNet18 underperforms One-Class SVM and Isolation Forest on CIFAR-10, with an AUC-ROC of 0.4715 compared to 0.7145 and 0.7057, respectively. One-Class SVM's superior performance likely stems from its effective boundary definition in the pre-extracted feature space, despite relying on fixed features. Isolation Forest's random partitioning also outperforms Deep SVDD, suggesting robustness in high-dimensional spaces when features are preprocessed.

Deep SVDD's poor performance (low recall: 0.10) indicates issues in clustering normal data tightly in the latent space. Possible causes include insufficient training epochs, suboptimal hyperparameters (e.g., learning rate, ν), or overfitting to the pretrained ResNet18 weights, which may not adapt well to CIFAR-10's airplane class. The low AUC-ROC suggests the model fails to generalize to diverse anomalies, contrary to expectations of end-to-end feature learning superiority.

One-Class SVM and Isolation Forest benefit from standardized features extracted by ResNet18, which may capture more discriminative patterns than Deep SVDD's learned representations. However, their moderate AUC-ROC scores (0.7145, 0.7057) indicate room for improvement, as they still struggle with CIFAR-10's complexity.

Limitations include Deep SVDD's computational cost and sensitivity to hyperparameter tuning. The baseline methods, while simpler, are limited by fixed features, reducing adaptability.

In conclusion, One-Class SVM offers the best performance for anomaly detection on CIFAR-10 in this study, highlighting the effectiveness of traditional methods with pre-extracted features. Deep SVDD's underperformance underscores the need for careful tuning and adaptation of deep learning approaches for specific datasets.

IX. Future Work

Despite expectations, Deep SVDD underperformed compared to classical models like One-Class SVM and Isolation Forest. This outcome highlights several limitations and opportunities for improvement in future work. First, hyperparameter tuning for Deep SVDD was limited. Exploring a broader range of values for learning rate, regularization, and the v parameter could improve the model's capacity to cluster normal samples more effectively. Second, reliance on pretrained ResNet18 weights may have limited adaptability to the CIFAR-10 dataset. Fine-tuning more layers or training the feature extractor from scratch; even though it's computationally expensive, could yield latent representations more suited to the specific normal class (airplanes). Finally, augmenting the training set with semi-supervised labels or synthetic outliers may improve generalization to unseen anomalies. Exploring more expressive models while incorporating anomaly-aware loss functions could help bridge the performance gap between deep and classical one-class models.

References

[1] Ruff, L., et al. (2018). "Deep One-Class Classification." *International Conference on Machine Learning (ICML).*

[2] Schölkopf, B., et al. (2001). "Estimating the Support of a High-Dimensional Distribution." *Neural Computation.*

[3] Liu, F. T., et al. (2008). "Isolation Forest." *IEEE International Conference on Data Mining (ICDM).*

[4] Dosovitskiy, A., et al. (2020). "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale." *International Conference on Learning Representations (ICLR).*