2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)
October 1-5, 2023. Detroit, USA

Learning Representation for Anomaly Detection of Vehicle Trajectories

Ruochen Jiao[1], Juyang Bai[1], Xiangguo Liu[1], Takami Sato[2], Xiaowei Yuan[1], Qi Alfred Chen[2], Qi Zhu[1]

***Abstract*— Predicting the future trajectories of surrounding vehicles based on their history trajectories is a critical task in autonomous driving. However, when small crafted perturbations are introduced to those history trajectories, the resulting anomalous (or adversarial) trajectories can significantly mislead the future trajectory prediction module of the ego vehicle, which may result in unsafe planning and even fatal accidents. Therefore, it is of great importance to detect such anomalous trajectories of the surrounding vehicles for system safety, but few works have addressed this issue. In this work, we propose two novel methods for learning effective and efficient representations for online anomaly detection of vehicle trajectories. Different from general time-series anomaly detection, anomalous vehicle trajectory detection deals with much richer contexts on the road and fewer observable patterns on the anomalous trajectories themselves. To address these challenges, our methods exploit contrastive learning techniques and trajectory semantics to capture the patterns underlying the driving scenarios for effective anomaly detection under supervised and unsupervised settings, respectively. We conduct extensive experiments to demonstrate that our supervised method based on contrastive learning and unsupervised method based on reconstruction with semantic latent space can significantly improve the performance of anomalous trajectory detection in their corresponding settings over various baseline methods. We also demonstrate our methods' generalization ability to detect unseen patterns of anomalies.**

I. Introduction

Tremendous progress has been made for autonomous driving in recent years. The autonomous driving pipeline typically consists of several modules, such as sensing, perception, prediction, planning, and control. In particular, the prediction module encodes other vehicles' past trajectories along with map context and decodes them into potential future trajectories of surrounding vehicles to facilitate the planning module. Recent works [1]–[3] have developed various deep learning-based models for trajectory prediction and achieved great performance in terms of the average error between predicted trajectories and ground truth. However, only improving average performance is not enough for autonomous driving systems, where system robustness, safety and security are critical [4]–[9].

Due to the complexity of real traffic situations and limited coverage of training data, the trajectory prediction task suffers from the "long-tail" scenarios. The work in [10] further demonstrates that state-of-the-art trajectory prediction models can be significantly misled by natural-looking but carefully-crafted past trajectory of a certain surrounding vehicle, and discusses several defense methods such as smoothing and SVM-based detection. [11] shows that adversarial training techniques can mitigate the effect of adversarial trajectories. However, few works focus on advanced *online anomaly detection methods for vehicle trajectories*. We believe that it is crucial to detect anomalous trajectories and scenarios in the prediction stage during runtime, as online anomalous trajectory detection will not only help monitor the prediction module but also enhance the safety of downstream modules in planning [12] and control [13], [14].

In this work, we consider two different settings – supervised and unsupervised, based on whether we have prior knowledge of patterns of anomalous trajectories during the training. Both scenarios are possible in real road situations, but they may make a significant difference to methods of learning representations. Thus, we focus on detecting various patterns of anomalous vehicle trajectories in both *supervised* and *unsupervised* settings, and investigate what kinds of representations and corresponding learning techniques are most effective for this safety-critical task.

The representation for anomalous vehicle trajectory detection is more complicated than that for general time-series anomaly detection because that 1) the driving scenarios contain rich contexts such as road maps and interactions between agents and 2) the anomalous or adversarial trajectories may be associated with specific driving behavior that is difficult to model. To tackle these challenges, an ideal anomaly detector should be able to effectively represent the driving scenario at a single sample level and also model the patterns underlying all normal and anomalous trajectories at the distribution level. Therefore, we first apply a state-of-the-art feature extractor based on graph neural networks to represent the trajectories as well as the road contexts, which is trained on a normal trajectory prediction dataset. Based on the extracted feature, we further add an encoder to capture the distribution-level patterns underlying the anomalous and normal trajectories. In the supervised setting, we add a contrastive-learning-based encoder to separate the two patterns in the representation space. In the unsupervised setting, we introduce semantics of driving behavior to learn a general and effective latent space for anomaly detection in complex scenarios without labels.

We extensively compare the anomaly detection performance of different representations under various kinds of anomalies and test scenarios and demonstrate that our proposed representations significantly enhance anomalous tra-

[1] Ruochen Jiao, Juyang Bai, Xiangguo Liu, Xiaowei Yuan, and Qi Zhu are with the Department of Electrical and Computer Engineering, Northwestern University, IL. {ruochen.jiao, juyangbai2023, xg.liu,xiaoweiyuan2023}@u.northwestern.edu, qzhu@northwestern.edu.

[2] Takami Sato and Qi Alfred Chen are with the Department of Computer Science, University of California, Irvine, CA. {takamis,alfchen}@uci.edu.

2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) | 978-1-6654-9190-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/IROS55552.2023.10342070

jectory detection performance over baseline methods. The contributions of our work are summarized as follows:

- We design a supervised contrastive learning-based method and an unsupervised method with semantics-guided reconstruction for the anomaly detection of vehicle trajectories and demonstrate their effectiveness in different settings.
- We explore and compare various representations and architectures for anomalous trajectory detection under supervised and unsupervised settings. We evaluate their performances with three metrics in two different datasets.
- We further demonstrate the algorithms' generalization ability to detect unseen patterns of anomalies and provide a detailed study to analyze the effectiveness of different modules in our methods.

II. Background

A. Anomalous and Adversarial Trajectories

Recent work [10] shows that the trajectory prediction module in autonomous driving pipelines can be easily misled by adversarial (history) trajectories of a surrounding vehicle. In a white-box setting, an anomalous trajectory is optimized with Projected Gradient Decent (PGD) [15] and the maximum deviation between benign and anomalous trajectories is limited to 1 m. There are different patterns of anomalous trajectories – random anomalies and directional anomalies. As shown in Fig. 1, both kinds of anomalous trajectories can effectively interfere with the prediction module and may lead to dangerous scenarios. The random anomaly is a generated trajectory that maximizes the average of the root mean squared error between the predicted and the ground-truth trajectory waypoints. The directional anomalous trajectory is to deliberately mislead the prediction of the surrounding vehicle's future trajectory to a wrong direction. In this work, we apply the lateral directional anomalous as a targeted anomalous pattern and the random attack as a general anomalous pattern, so that we can evaluate the anomaly detection algorithm comprehensively and study their generalization ability to previously-unseen patterns of anomalous trajectories. The detailed metrics for the optimization of targeted attacks are shown in Eq. (1):

$$D(\alpha, R) = (p_\alpha - s_\alpha)^T \cdot R\left(s_{\alpha+1}, s_\alpha\right), \tag{1}$$

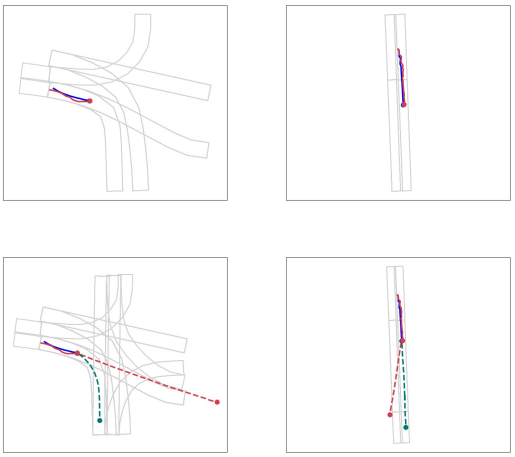
where α denotes the time frame, R is a function to generate the unit vector to a specific direction (lateral direction in our setting), and p and s are vectors denoting predicted and ground-truth vehicle locations, respectively.

B. Anomaly Detection

Anomaly detection refers to the problem of finding patterns in data that do not conform to expected behavior [16]. Supervised approaches, unsupervised approaches, and semi-supervised approaches have been applied to anomaly detection in different scenarios.

Supervised approaches generally have better performance on classification tasks, but require prior knowledge of both normal and anomalous samples. KNN-based methods [17], [18] capture nominal data patterns from the local interaction of nominal data points, and anomalous instances are expected to lie further away from nominal data patterns. The support vector machine (SVM) and neural networks are commonly used to project the input to a feature space and then detect the anomalies from normal data. Some other methods, such as Bayesian networks [19] and inverse reinforcement learning [20], are also effective in supervised anomaly detection.

When labels of anomalies are limited or even unavailable, we have to utilize semi-supervised [21] methods and unsupervised methods for the anomaly detection tasks. Reconstruction-based methods assume that anomalies are not compressible and thus cannot be reconstructed from low-dimensional projections [22]. Deep generative models, such as variational autoencoder (VAE) [23], [24], Generative Adversarial Networks [25], [26] and adversarial autoencoder [27], are commonly used to perform reconstruction-based anomaly detection. One-class classification methods including one-class SVM (OC-SVM) [28], [29] and one-class neural network (OCNN) [30] are designed to learn a discriminative boundary surrounding the normal samples.



(a) (b)

Fig. 1: Different patterns and corresponding effects of anomalous trajectories. The figures in the top row are anomalous (red line) and benign (blue line) input history trajectories for prediction and they look very close to human eyes. The figures in the bottom row show the corresponding affected future trajectory prediction (red dashed line) and the ground truth trajectory in the normal scenario (green dashed line). The differences between the two are clearly visible, showing the great influence caused by anomalous trajectories on prediction modules. Figure (a) is the random anomalous trajectory, which will randomly lead to maximum average deviation. Figure (b) is the lateral directional anomaly, which mainly leads the vehicle to deviate to the left or right.

C. Contrastive Learning

Contrastive learning [31] learns representations by contrasting positive pairs against negative pairs. Generally, the

augmented versions of the original samples are regarded as positive pairs, and a memory bank is used to stabilize the learning process. Recent works show that contrastive learning techniques can benefit representation learning significantly and there are also some advances in enhancing anomaly detection by utilizing the idea of contrastive learning. For instance, [32] proposes an unsupervised method TS2Vec for learning representations of time series. The TS2Vec method captures the contextual representation by leveraging both instance-wise and temporal contrastive loss, and the method shows great performance in time-series anomaly detection. Under a supervised setting, [33] demonstrates that the intermediate features of anomaly and normal data can be considered as negative pairs and help learn an effective representation based on contrast.

D. Adversarial Autoencoder

The variational autoencoder (VAE) [34] provides a principled method for jointly learning deep latent-variable models and corresponding inference models using stochastic gradient descent [35], which is commonly used to generate samples in the target space from pre-defined latent distribution. Training a VAE model consists of two kinds of loss: regularization and reconstruction. The regularization is aimed to encode the input as certain distributions over the latent space using Kullback-Leibler (KL) divergence, while the reconstruction is to decode the latent variables to the target or original space. In contrast to VAE that uses KL divergence and evidence lower bound, adversarial autoencoder (AAE) [36] uses adversarial learning to impose a specific distribution on the latent variables, making itself superior to VAE in terms of imposing complicated distributions and shaping the latent space. In our work, we utilize an AAE architecture to model the semantics in the driving scenario.

III. Our Methods

A. Feature Extractor

The anomalous trajectory detection task is more complex compared to general time-series anomaly detection because anomalous trajectory detection highly depends on rich contexts, such as the road map and the behavior of surrounding vehicles. We first feed the map information and the trajectories of vehicles into a feature extractor. Similar to [1], we apply a one-dimensional convolutional network to model history trajectories and utilize graph-neural networks to represent map contexts and interactions between agents. In the training process, we first train a feature extractor in a trajectory prediction pipeline and fix the extractor in the anomaly detection task.

B. Supervised Contrastive Learning-based Method

As shown in Fig. 2, after the feature extractor outputs a representation combining vehicle trajectories and contexts, we further develop a contrastive learning (CL) based encoder to obtain a compact representation for anomalous trajectory detection. Different from instance-wise contrastive learning, the proposed method compares the patterns of two different classes. This CL-based method is considered supervised because we build the negative pairs by contrasting the anomaly and normal data. The CL-based encoder is designed to maximize the similarity between benign scenarios and minimize the similarity between benign and anomalous scenarios. Finally, a simple binary classifier based on the encoded representations will generate the decision on whether the input is an anomalous scenario or not.

In every training mini-batch, we have M scenarios containing anomalous trajectories and N normal scenarios, so that we have $N(N-1)$ positive pairs and MN negative pairs, as demonstrated in Fig. 2. We use the inner product of two vectors to measure the cosine similarity between encoded features and we set τ to control the concentration of samples' distribution [37]. More negative pairs will generally improve the performance of the learned representation, but it is difficult to calculate and optimize such a large model with MN-way softmax vector. To fully utilize the labeled data and keep the model efficient, we apply the idea of Noise Contrastive Estimation (NCE) [38] to the optimization. We have an $(M+1)$-way softmax classifier (one way for a certain positive pair and M ways for negative pairs) to learn a 32-dimensional representation. The loss function is shown in Eq. (2), where $\boldsymbol{s_n}$ and $\boldsymbol{s_a}$ are the feature vectors of normal and anomalous scenarios, respectively. We define the overall loss function $\mathcal{L}$ for a mini-batch, as shown in Eq. (3):

$$\mathcal{L}_{ij} = -\log \frac{\exp\left(\boldsymbol{s_{ni}}^{\mathrm{T}} \boldsymbol{s_{nj}}/\tau\right)}{\exp\left(\boldsymbol{s_{ni}}^{\mathrm{T}} \boldsymbol{s_{nj}}/\tau\right) + \sum_{m=1}^{M} \exp\left(\boldsymbol{s_{ni}}^{\mathrm{T}} \boldsymbol{s_{am}}/\tau\right)}, \tag{2}$$

$$\mathcal{L} = \frac{1}{N(N-1)} \sum_{i=1}^{N} \sum_{j=1}^{N} \mathbb{1}_{j \neq i} \mathcal{L}_{ij}. \tag{3}$$

During the test time, we add a classifier after the CL-based encoder to produce the detection result based on the 32-dimensional vector. It is feasible to set a threshold of the distance between test samples and the average normal vector to distinguish anomaly from normal data, but in this work, we apply an SVM classifier for all kinds of representations so that we can compare their results fairly. The overall pipeline is demonstrated in Algorithm 1.

C. Unsupervised Method using Reconstruction with Semantic Latent Space

In real road scenes, it may be challenging for us to get valid negative labels for training due to the difficulty of obtaining prior knowledge of surrounding vehicles' potential anomalous trajectories, which motivates us to explore unsupervised detection algorithms. The unsupervised methods are aimed to learn the representation underlying normal driving scenarios and then detect unseen patterns of anomalous trajectories at runtime. Most previous works directly use VAE-based reconstruction, one-class SVM, or contrastive learning (unsupervised) to detect anomalies for time-series data. For vehicle trajectories, however, we can utilize more contexts and domain knowledge to enhance unsupervised

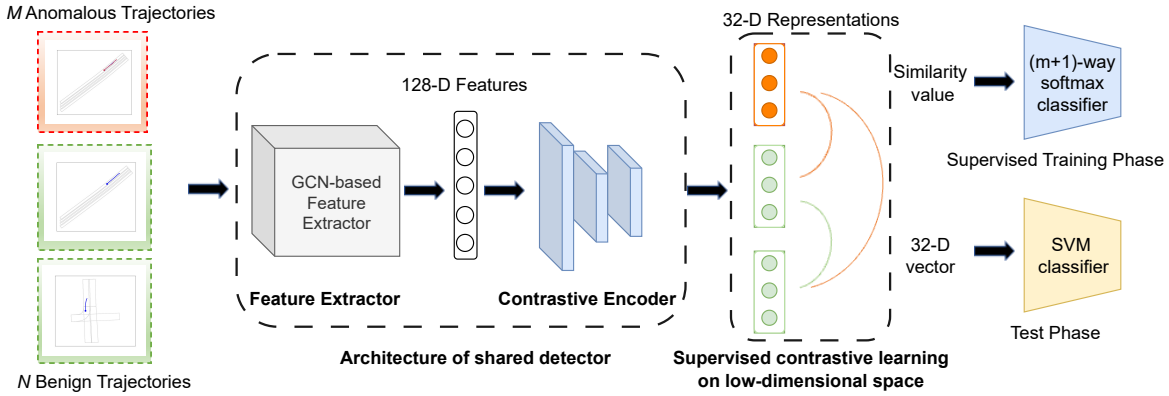


Fig. 2: Our supervised contrastive learning-based anomalous trajectory detector. The features of normal and anomalous scenarios are regarded as negative pairs (red ones) and the normal scenarios form positive pairs (green ones).

Algorithm 1 Supervised Contrastive Learning Method

Initialize: feature extractor F, CL-based encoder E, Cosine similarity S_c Softmax classifier C, and SVM-detector D.
Input: past trajectories t and map graph g.
for each mini-batch **do**
  128-D Features $x = F(t, g)$.
  32-D latent vectors $z = E(x)$.
  N benign trajectories and M anomalous trajectories generate N(N-1) positive pairs and MN negative pairs in the latent space.
  Similarity scores of pairs $s = S_c(z)$.
  for each positive pair (i, j) **do**
    Calculate NCE softmax loss $L_{ij}(s)$ as in (2).
  end for
  Update the encoder E by the CL loss as in (3).
end for
Based on the learned 32-D representation, train an SVM classifier for online anomaly detection.

anomaly detection. In this work, we propose an unsupervised detection method based on the adversarial autoencoder architecture and semantics modeling in the latent space. The encoder takes 128-dimensional features from the extractor as inputs and projects them into a low-dimensional latent space that is divided into three separate parts – a three-dimensional vector representing lateral intention, a one-dimensional vector representing longitudinal aggressiveness, and a six-dimensional remaining latent vector. Here, we introduce domain knowledge into latent space modeling. We apply *time headway* to extract the longitudinal feature, which measures the time difference between two successive vehicles when crossing a given point. We assume that the time headway follows a log-normal distribution, based on the statistics in urban transportation systems [3], [39]. The lateral intention is modeled by three simple but reasonable classes that follow categorical distribution: moving forward, turning/changing lanes to the left, and turning/changing lanes to the right. All this semantic information can be collected from benign input trajectory and no knowledge of anomaly is required. For the remaining variables in the latent space, we assume that they follow Gaussian distributions.

To optimize the latent space, we conduct a two-fold modeling in both overall distributions and semantics of a single vehicle's trajectories. We apply the *adversarial autoencoder* architecture to regularize these distributions of the latent space. Specifically, for each latent vector, a discriminator is trained to distinguish the generated latent vector from the sample in real targeted distribution (log-normal, categorical, or Gaussian). At the same time, we use behavior information such as values of time headway and lateral intention to further render the latent vectors with specific semantics. Thereby, the model can further disentangle the latent space and embed domain knowledge into it. The semantic representation will benefit both unsupervised and supervised anomalous trajectory detection. The loss for semantic latent space modeling is shown below in Eq. (4):

$$Loss_{sem}(z, g) = -\sum_{i=1}^{3} g_{intent} \log z_{intent} + (g_{agg} - z_{agg})^2, \tag{4}$$

where z represents the predicted semantic vectors and g represents the reference collected from the input trajectory.

The overall pipeline for our unsupervised method is shown in Fig. 3. In addition to the semantic latent space modeling, we use a decoder to reconstruct the input trajectories with a smooth L1 loss as shown below in Eq. (5):

$$Loss_{recon}(y_i, \hat{y}_i) = \begin{cases} 0.5\,(y_i - \hat{y}_i)^2 & \text{if } \|y_i - \hat{y}_i\| < 1, \\ \|y_i - \hat{y}_i\| - 0.5 & \text{otherwise}, \end{cases} \tag{5}$$

where y and $\hat{y}$ represent the input trajectories and reconstructed trajectories, respectively. Both the encoder and the decoder will be optimized by the reconstruction loss. The overall optimization pipeline is shown in Algorithm 2.

Note that in an unsupervised setting, we use the error between the input trajectory and reconstructed trajectory as

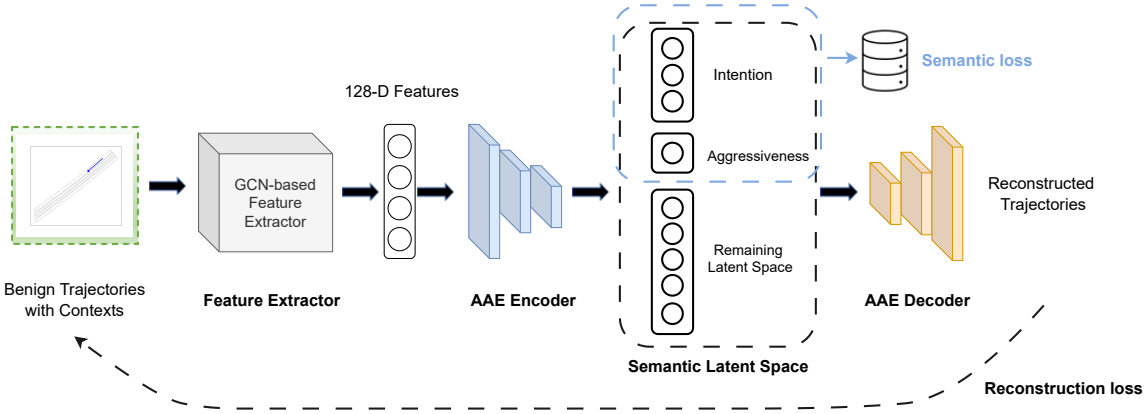


Fig. 3: Our unsupervised reconstruction method based on semantic latent space. The latent space contains three kinds of vectors – intention, aggressiveness, and the remaining vectors. The latent vectors are modeled by semantic labels and target distributions in the adversarial autoencoder.

a signal for anomaly detection. We consider the input as an anomalous trajectory if the reconstruction error is larger than a threshold. The learned representation can also be used in a supervised setting by adding common binary classifiers after the latent space.

Algorithm 2 Unsupervised Semantic Reconstruction Method

Initialize: feature extractor F, AAE encoder E, decoder R, discriminator D_i, target distribution p_i, i = 1,2,3.
Input: past trajectories t and map graph g.
for each batch **do**
  Features $x = F(t, m)$.
  Let latent vectors $z = E(x)$.
  Sample s_i from target distribution p_i and calculate $D_i(z_i)$ and $D_i(s_i)$.
  Update E and D_i by discrimination loss and generation loss as in [36].
  Calculate the true value for intention and aggressiveness, respectively.
  Update E by semantic loss $Loss_{sem}$ as in (4).
  Concatenate the latent vectors and feed them to the detector R $\hat{y} = R(z)$.
  Update G, R by reconstruction loss $Loss_{recon}(y, \hat{y})$ as in (5).
end for

IV. Experiments

In this section, we present the anomaly detection results of six methods under two patterns of anomalous scenarios. Each method is tested in two commonly-used datasets with three different metrics so as to comprehensively evaluate the performance, especially under imbalanced data distribution. The results show that the representation learned by our supervised contrastive learning can significantly improve detection performance. Moreover, the semantic latent space we construct can effectively model the context and explicitly encode driving behavior, enhancing anomaly detection in both supervised and unsupervised settings. We conduct a study to show to what extent the 'semantics' and 'contrast' can benefit the representation learning for anomaly trajectory detection. In addition, we evaluate the generalization ability of learned representations, which is critical for detecting unseen patterns of anomalies.

A. Experiment Setup

1) Data Collection: We conduct experiments with both random and directional anomalous trajectories on two datasets: Argoverse 1 [40] and Argoverse 2 [41]. The Argoverse 1 motion forecasting dataset has more than 30K driving scenarios collected in Miami and Pittsburgh, while Argoverse 2 collects longer and more complicated driving scenarios in six cities. Each scenario used in this work consists of a road graph and trajectories of multiple agents. The history trajectories are 20 waypoints collected in the past 2 seconds.

To collect the anomalous trajectories, we apply the attack methods mentioned in Sec. II-A to generate different patterns of anomalies. For lateral directional anomalous trajectories, we consider past trajectories that can lead to a prediction error of more than 1.5 meters in a lateral direction as anomalies. For random anomalous trajectories, the threshold is set as 5-meter average displacement error (ADE).

2) Evaluation Metrics: We utilize three metrics to evaluate the performance of anomaly detection approaches – ROC AUC (area under the receiver operating characteristic curve), PR AUC (area under the precision-recall curve), and F1 score. The ROC AUC is a general metric to evaluate the binary classification ability at all classification thresholds, but it can be overly optimistic on severely imbalanced classification problems. For imbalance datasets in anomaly detection, the PR AUC is a more powerful metric as both precision and recall are focused on the anomaly class and unconcerned with the majority class. The F1 score is the harmonic mean of precision and recall. In the anomaly detection task, the recall (detected anomalies over all anomalies) is expected to be high. Thus we find the point where recall is fixed as 0.8 and calculate its corresponding F1 score.

B. Effectiveness of Our CL-based Supervised Method

In the supervised setting, the labels of anomalous driving scenarios are available. We use SVM as a fixed classifier to compare the results of different learned representations. The naive SVM method builds an SVM directly on the acceleration series of the input trajectories. For the methods with semantic latent space and contrastive learning encoder, we use the 128-dimensional feature produced by the feature extractor as input. The results in Tables I and II show that **our contrastive learning-based supervised method greatly outperforms other supervised methods in both directional and random anomaly patterns**. Compared to the method using semantic latent space ('Semantics + SVM'), our CL-based supervised method ('Sup-CL + SVM') can effectively model the distribution of normal and abnormal trajectories and separate them in the CL-based representation space, making it easy for a simple SVM classifier to detect anomalies. The results of naive SVM ('Naive SVM') demonstrate that it is difficult to directly distinguish anomaly from normal data in the trajectory space, even with enough labels.

TABLE I: Results of supervised anomaly detection methods for the Argoverse 1 dataset.

Methods	Anomaly Pattern	F1 score	ROC AUC	PR AUC
Naive SVM [10]	Random	0.51	0.63	0.43
Naive SVM [10]	Lateral	0.49	0.75	0.51
Semantics + SVM	Random	0.68	0.85	0.74
Semantics + SVM	Lateral	0.84	0.96	0.92
Sup-CL + SVM	**Random**	**0.81**	**0.93**	**0.86**
Sup-CL + SVM	**Lateral**	**0.87**	**0.98**	**0.96**

TABLE II: Results of supervised anomaly detection methods for the Argoverse 2 dataset.

Methods	Anomaly Pattern	F1 score	ROC AUC	PR AUC
Naive SVM [10]	Random	0.42	0.59	0.31
Naive SVM [10]	Lateral	0.47	0.64	0.40
Semantics + SVM	Random	0.62	0.84	0.65
Semantics + SVM	Lateral	0.84	0.95	0.89
Sup-CL + SVM	**Random**	**0.80**	**0.94**	**0.87**
Sup-CL + SVM	**Lateral**	**0.99**	**0.98**	**0.99**

C. Effectiveness of Our Unsupervised Method with Semantic Reconstruction

In the unsupervised case, without prior knowledge of anomaly patterns, it is more difficult to learn an informative representation. We utilize the one-class SVM on trajectory space [28] and a state-of-the-art unsupervised contrastive learning method – TS2Vec [32] as baselines. As shown in Tables III and IV, the one-class SVM on trajectory space method ('OC-SVM') and the unsupervised contrastive learning method ('Unsup-CL + OC-SVM') have relatively poor performance and can hardly detect the anomalous trajectories. **Our unsupervised method with semantic reconstruction ('Semantic Recon') has much better performance on every metric over other unsupervised methods.** In addition, when detecting random anomalous trajectories, we find that our unsupervised method has close performance to its corresponding supervised version, which further demonstrates the effectiveness of our semantic latent representation. The ROC curves of all supervised and unsupervised methods are shown in Fig. 4.

TABLE III: Results of unsupervised anomaly detection methods for the Argoverse 1 dataset.

Methods	Anomaly Pattern	F1 score	ROC AUC	PR AUC
OC-SVM [28]	Random	0.51	0.58	0.36
OC-SVM [28]	Lateral	0.47	0.64	0.30
Unsup-CL + OC-SVM [32]	Random	0.41	0.54	0.30
Unsup-CL + OC-SVM [32]	Lateral	0.40	0.56	0.29
Semantic Recon	**Random**	**0.65**	**0.81**	**0.61**
Semantic Recon	**Lateral**	**0.60**	**0.83**	**0.56**

TABLE IV: Results of unsupervised anomaly detection methods for the Argoverse 2 dataset.

Methods	Anomaly Pattern	F1 score	ROC AUC	PR AUC
OC-SVM [28]	Random	0.42	0.55	0.28
OC-SVM [28]	Lateral	0.40	0.54	0.26
Unsup-CL + OC-SVM [32]	Random	0.41	0.54	0.30
Unsup-CL + OC-SVM [32]	Lateral	0.38	0.57	0.27
Semantic Recon	**Random**	**0.53**	**0.77**	**0.47**
Semantic Recon	**Lateral**	**0.53**	**0.79**	**0.48**

D. Effectiveness of Components in Our Proposed Methods

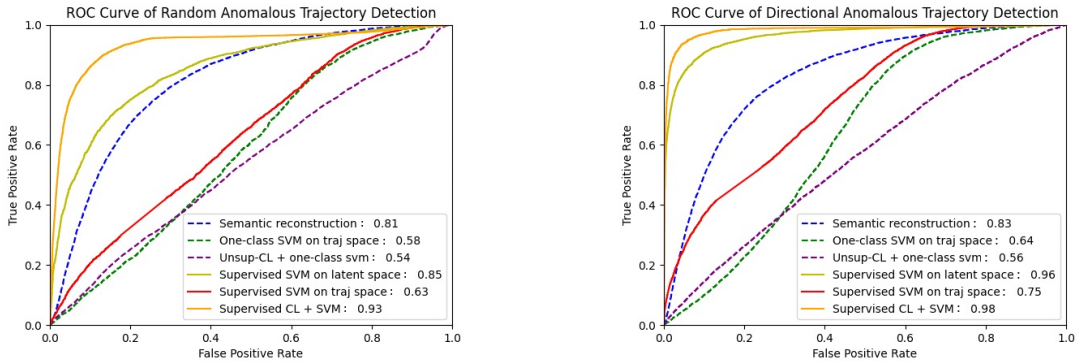
We conduct more experiments to study to what extent different components of our methods benefit the overall performance improvement. In this study, we test all representations in a supervised manner for comparison. We evaluate how much the feature extractor (embedding the contexts) and the following encoder (semantics or CL-based) contribute to the performance, respectively. The 'NN + SVM' stands for the method of directly adding an SVM classifier after the GNN-based feature extractor. In Table V, we find that using such representation directly from the feature extractor has a much poorer detection performance than our CL-based representation ('Sup-CL + SVM'). The latent space modeling methods ('Semantics + SVM' and 'Naive Latent + SVM') also outperform the pure feature extractor ('NN + SVM'). Moreover, Table VI reveals that our semantic latent space modeling ('Semantics + SVM') significantly improves the generalization ability to unseen patterns of anomalies when compared to the naive latent space modeling without any semantics ('Naive Latent + SVM') and the pure feature extractor ('NN + SVM').

TABLE V: Effectiveness of components: trained and tested on random anomalies.

Methods	Anomaly Pattern	F1 score	ROC AUC	PR AUC
NN + SVM	Random	0.66	0.61	0.68
Naive Latent + SVM	Random	0.68	0.86	0.74
Semantics + SVM	Random	0.68	0.85	0.74
Sup-CL + SVM	Random	0.81	0.93	0.86

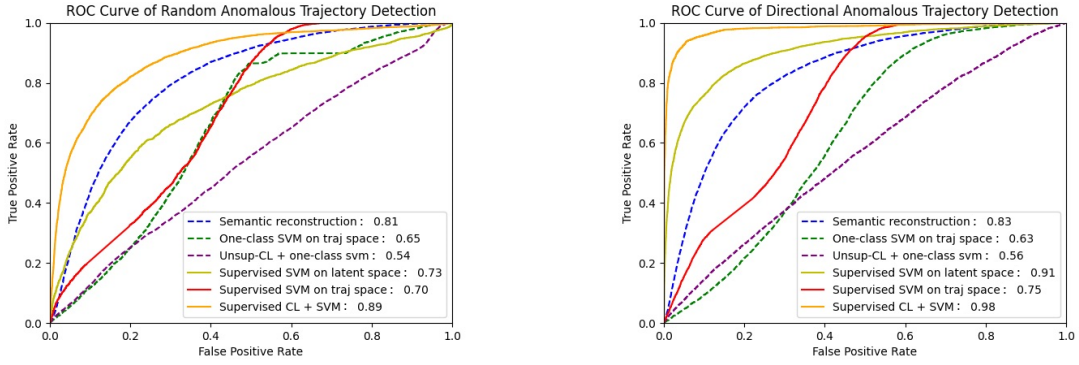
E. Evaluation on Generalization Ability

In the supervised setting, one critical aspect is how the representation learned from normal samples and a certain pattern of anomalies can be generalized to other unseen



(a) (b)

Fig. 4: ROC curves of both supervised (solid lines) and unsupervised (dashed lines) anomalous trajectory detection approaches The AUC values are shown in legends. The left figure shows the results of detecting **random** anomalous trajectories when all supervised methods are trained on the **random** anomalies. The right figure shows the results of detecting **lateral directional** anomalous trajectories when all supervised methods are trained on the **directional** anomalies.



(a) (b)

Fig. 5: ROC curves of both supervised (solid lines) and unsupervised (dashed lines) anomalous trajectory detection approaches under **unseen** patterns of anomalies. The AUC values are shown in legends. The left figure shows the results of detecting **random** anomalous trajectories when all supervised methods are trained on the **lateral directional** anomalies. The right figure shows the results of detecting **lateral directional** anomalous trajectories when all supervised methods are trained on the **random** anomalies.

TABLE VI: Effectiveness of components: trained on directional anomalies and tested on random anomalies.

Methods	Anomaly Pattern	F1 score	ROC AUC	PR AUC
NN + SVM	Random	0.44	0.60	0.35
Naive Latent + SVM	Random	0.44	0.58	0.45
Semantics + SVM	Random	0.65	0.81	0.58
Sup-CL + SVM	Random	0.70	0.89	0.75

patterns of anomalies. Fig. 5 shows the results when the supervised methods are trained and tested on different patterns of anomalies. Compared to Fig. 4, we find that the lateral directional anomalies are relatively easy to detect, even when the models are trained on another kind of anomaly. However, when the models are trained on lateral directional anomalies but tested on the random anomalous trajectories, the performances of supervised methods drop significantly, although the supervised CL-based method is still the best, which reveals overfitting and a lack of ability to generalize. In this setting, the unsupervised reconstruction with semantic latent space even outperforms its corresponding supervised version. Table VI further shows our semantics modeling can help in learning a more generalized representation and mitigating overfitting to a certain pattern of anomalies, compared to the naive latent space.

V. Conclusions

We present novel contrastive learning-based supervised method and semantic reconstruction-based unsupervised method for anomalous vehicle trajectory detection. We embed driving contexts and distributions underlying the normal (and anomalous) trajectories into the representation by various methods. Experiments demonstrate that our methods can significantly improve the detection performance over baseline methods in supervise and unsupervised settings, respectively. We also demonstrate that our methods have better generalization ability to address unseen attack patterns.

ACKNOWLEDGMENT

We gratefully acknowledge the support from NSF grants 1834701, 1724341, 2038853, 1929771, 2145493, 1932464, and ONR grant N00014-19-1-2496.

REFERENCES

[1] M. Liang, B. Yang, R. Hu, Y. Chen, R. Liao, S. Feng, and R. Urtasun, "Learning lane graph representations for motion forecasting," in *European Conference on Computer Vision*. Springer, 2020, pp. 541–556.

[2] Y. Yuan, X. Weng, Y. Ou, and K. M. Kitani, "Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 9813–9823.

[3] R. Jiao, X. Liu, B. Zheng, D. Liang, and Q. Zhu, "Tae: A semi-supervised controllable behavior-aware trajectory generator and predictor," *arXiv preprint arXiv:2203.01261*, 2022.

[4] Z. Wang, C. Huang, and Q. Zhu, "Efficient global robustness certification of neural networks via interleaving twin-network encoding," in *DATE'22: Proceedings of the Conference on Design, Automation and Test in Europe*, 2022.

[5] X. Liu, C. Huang, Y. Wang, B. Zheng, and Q. Zhu, "Physics-aware safety-assured design of hierarchical neural network based planner," in *2022 ACM/IEEE 13th International Conference on Cyber-Physical Systems (ICCPS)*. IEEE, 2022, pp. 137–146.

[6] X. Liu, B. Luo, A. Abdo, N. Abu-Ghazaleh, and Q. Zhu, "Securing connected vehicle applications with an efficient dual cyber-physical blockchain framework," in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, pp. 393–400.

[7] B. Zheng, C.-W. Lin, *et al.*, "Delay-Aware Design, Analysis and Verification of Intelligent Intersection Management," in *2017 SMART-COMP*.

[8] ——, "Design and Analysis of Delay-Tolerant Intelligent Intersection Management," *ACM T-CPS*, vol. 4, no. 1, pp. 3:1–3:27, Nov. 2019.

[9] B. Zheng, H. Liang, Q. Zhu, H. Yu, and C. W. Lin, "Next generation automotive architecture modeling and exploration for autonomous driving," in *2016 IEEE Computer Society Annual Symposium on VLSI (ISVLSI)*, July 2016, pp. 53–58.

[10] Q. Zhang, S. Hu, J. Sun, Q. A. Chen, and Z. M. Mao, "On adversarial robustness of trajectory prediction for autonomous vehicles," *arXiv preprint arXiv:2201.05057*, 2022.

[11] R. Jiao, X. Liu, T. Sato, Q. A. Chen, and Q. Zhu, "Semi-supervised semantics-guided adversarial training for trajectory prediction," *arXiv preprint arXiv:2205.14230*, 2022.

[12] X. Liu, R. Jiao, B. Zheng, D. Liang, and Q. Zhu, "Safety-driven interactive planning for neural network-based lane changing," in *Proceedings of the 28th Asia and South Pacific Design Automation Conference*, 2023, pp. 39–45.

[13] R. Jiao, H. Liang, T. Sato, J. Shen, Q. A. Chen, and Q. Zhu, "End-to-end uncertainty-based mitigation of adversarial attacks to automated lane centering," in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, pp. 266–273.

[14] Q. Zhu, C. Huang, R. Jiao, S. Lan, H. Liang, X. Liu, Y. Wang, Z. Wang, and S. Xu, "Safety-assured design and adaptation of learning-enabled autonomous systems," in *Proceedings of the 26th Asia and South Pacific Design Automation Conference*, 2021, pp. 753–760.

[15] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," *arXiv preprint arXiv:1706.06083*, 2017.

[16] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM computing surveys (CSUR)*, vol. 41, no. 3, pp. 1–58, 2009.

[17] S. Ramaswamy, R. Rastogi, and K. Shim, "Efficient algorithms for mining outliers from large data sets," in *Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, 2000, pp. 427–438.

[18] K. Doshi and Y. Yilmaz, "Online anomaly detection in surveillance videos with asymptotic bound on false alarm rate," *Pattern Recognition*, vol. 114, p. 107865, 2021.

[19] S. Mascaro, A. E. Nicholso, and K. B. Korb, "Anomaly detection in vessel tracks using bayesian networks," *International Journal of Approximate Reasoning*, vol. 55, no. 1, pp. 84–98, 2014.

[20] D. Li, M. L. Shehab, Z. Liu, N. Aréchiga, J. DeCastro, and N. Ozay, "Outlier-robust inverse reinforcement learning and reward-based detection of anomalous driving behaviors," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 4175–4182.

[21] L. Ruff, R. A. Vandermeulen, N. Görnitz, A. Binder, E. Müller, K.-R. Müller, and M. Kloft, "Deep semi-supervised anomaly detection," *arXiv preprint arXiv:1906.02694*, 2019.

[22] T. Li, Z. Wang, S. Liu, and W.-Y. Lin, "Deep unsupervised anomaly detection," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 3636–3645.

[23] J. Wiederer, A. Bouazizi, M. Troina, U. Kressel, and V. Belagiannis, "Anomaly detection in multi-agent trajectories for automated driving," in *Conference on Robot Learning*. PMLR, 2022, pp. 1223–1233.

[24] Y. Tang, L. Zhao, S. Zhang, C. Gong, G. Li, and J. Yang, "Integrating prediction and reconstruction for anomaly detection," *Pattern Recognition Letters*, vol. 129, pp. 123–130, 2020.

[25] T. Li, M. Shang, S. Wang, M. Filippelli, and R. Stern, "Detecting stealthy cyberattacks on automated vehicles via generative adversarial networks," in *2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2022, pp. 3632–3637.

[26] T. Schlegl, P. Seeböck, S. M. Waldstein, G. Langs, and U. Schmidt-Erfurth, "f-anogan: Fast unsupervised anomaly detection with generative adversarial networks," *Medical image analysis*, vol. 54, pp. 30–44, 2019.

[27] S. Pidhorskyi, R. Almohsen, and G. Doretto, "Generative probabilistic novelty detection with adversarial autoencoders," *Advances in neural information processing systems*, vol. 31, 2018.

[28] S. M. Erfani, S. Rajasegarar, S. Karunasekera, and C. Leckie, "High-dimensional and large-scale anomaly detection using a linear one-class svm with deep learning," *Pattern Recognition*, vol. 58, pp. 121–134, 2016.

[29] M. Amer, M. Goldstein, and S. Abdennadher, "Enhancing one-class support vector machines for unsupervised anomaly detection," in *Proceedings of the ACM SIGKDD workshop on outlier detection and description*, 2013, pp. 8–15.

[30] R. Chalapathy, A. K. Menon, and S. Chawla, "Anomaly detection using one-class neural networks," *arXiv preprint arXiv:1802.06360*, 2018.

[31] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, vol. 2. IEEE, 2006, pp. 1735–1742.

[32] Z. Yue, Y. Wang, J. Duan, T. Yang, C. Huang, Y. Tong, and B. Xu, "Ts2vec: Towards universal representation of time series," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, no. 8, 2022, pp. 8980–8987.

[33] O. Kopuklu, J. Zheng, H. Xu, and G. Rigoll, "Driver anomaly detection: A dataset and contrastive learning approach," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2021, pp. 91–100.

[34] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[35] D. Kingma and M. Welling, "An introduction to variational autoencoders," *arXiv preprint arXiv:1906.02691*, 2019.

[36] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," *arXiv preprint arXiv:1511.05644*, 2015.

[37] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," *arXiv preprint arXiv:1503.02531*, 2015.

[38] M. Gutmann and A. Hyvärinen, "Noise-contrastive estimation: A new estimation principle for unnormalized statistical models," in *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, 2010, pp. 297–304.

[39] D.-H. Ha, M. Aron, and S. Cohen, "Time headway variable and probabilistic modeling," *Transportation Research Part C: Emerging Technologies*, vol. 25, pp. 181–201, 2012.

[40] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, *et al.*, "Argoverse: 3d tracking and forecasting with rich maps," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 8748–8757.

[41] B. Wilson, W. Qi, T. Agarwal, J. Lambert, J. Singh, S. Khandelwal, B. Pan, R. Kumar, A. Hartnett, J. K. Pontes, *et al.*, "Argoverse 2: Next generation datasets for self-driving perception and forecasting," *arXiv preprint arXiv:2301.00493*, 2023.