Modular Novelty Detection System for Driving Scenarios

Maike Rees[1,2,3], Melanie Senn[2], Pratik P. Brahma[2], and Astrid Laubenheimer[1]

[1] Karlsruhe University of Applied Sciences
[2] Volkswagen Group of America, Inc., Electronics Research Laboratory
[3] `maike.rees@gmx.de`

Abstract. Unsupervised novelty detection has many applications in various fields of current research. This work proposes a new combination of commonly used novelty detection techniques applied on an automotive dataset. The goal is to differentiate between known and novel driving scenarios. The presented method is unsupervised and combines a convolutional autoencoder, a principal component analysis and a non-linear one-class support vector machine. The strength of the presented approach is its modularity. Visualization and interpretation of lower dimensional features ensure transparency about what the model learns. A module can be derived from an existing function (e.g. a previous classification task) or specifically be designed for the application domain. Additionally, it can be replaced when the context changes. The approach is also implemented with respect to limited compute capabilities, allowing its application in an autonomous vehicle. The achieved results are satisfying, especially when compared to a similar supervised approach and the visualization complies to the intuitive expectations.

Keywords: novelty detection · convolutional autoencoder · principal component analysis · one-class support vector machine.

1 Introduction

Many supervised machine learning techniques require large labeled data sets. Novelty detection can reduce the amount of data that needs expensive hand-labeling by performing a binary classification into normal and novel data. Normal data belongs to classes that are known by a model and can therefore be automatically labeled. Novel data belongs to classes that this model has never seen during its training and therefore needs hand-labeling. This classification is also valuable for the retraining of a model: It already performs sufficiently well on normal data and thus only has to be adapted (retrained) to the novel samples. Both presented applications suggest a one class classification as solution approach: The model knows the normal data to such an extend that when presented a novel data sample, the output is significantly different to the output of normal inputs. This significant deviation can be measured as novelty score and a threshold determines the classification. For novelty detection, a high true positive rate (every novelty is detected) and a low false positive rate (not many normal samples are missclassified) are desired. Working with novelties poses multiple challenges. We assume that novelties are rare, resulting in heavily unbalanced data sets. Additionally, due to the curse of dimensionality, every data sample can be characterized as novel when enough dimensions are used for the detection. This also means that if an algorithm is trained on various kinds of novelties, most likely there will be another kind of novelty that the algorithm misclassifies. The following section gives an overview of related approaches. Section 3 describes the proposed approach in detail. Section 4 presents the experiments and Section 5 their results and discussion. Section 6 concludes.

2 Related Work

Novelty detection is related to anomaly, outlier and corner case detection and subject to research in various fields of studies like medicine [10, 8], robot systems [13] and image recognition [6]. The survey by Pimental [7] groups novelty detection approaches into probabilistic, distance based, reconstruction based and domain based approaches. The latter find a boundary around the known domain data and detect every sample outside this boundary as novelty. One-class support vector machines (1CSVMs), introduced by Schoelkopf et al. [9] for novelty detection, are an example of such an approach.

Marsland [5] presents different novelty detection approaches in learning systems, like neural networks. Generative adversarial networks (GANs) are a new type of neural networks that also find application for novelty detection [8]. GANs consist of a generator and a discriminator part that compete in the training process to improve the final result. Seeboeck et al. [10] combine three autoencoders with a linear 1CSVM for outlier detection to a modular novelty detection approach. Erfani et al. [2] combine a linear 1CSVM with a deep belief network. Utkin et al. [13] combine autoencoders and a siamese network to a siamese autoencoder for anomaly detection in multi robot systems. Nguyen and Vien [6] use an end-to-end approach including a convolutional autoencoder (ConvAE) and Fourier features where the training of each part is directly dependent of the other modules.

In safety critical applications like autonomous driving, it is important to know what a learning system actually learns. Saliency maps [12] are an often used technique to get an understanding of the intermediate steps in neural networks. They show which neurons in each layer are activated for a given input.

The approach presented in this work focuses on the advantages of a modular approach to include visualizations of intermediate results: A ConvAE extracts features from the input data, a principal component analysis (PCA) [4] reduces the dimension of resulting features which are then visualized. A non-linear 1CSVM detects the novelties based on these low-dimensional features. Due to the similarity of the input data, both in content and image sizes, the ConvAE used for feature extraction in this approach is inspired by Hasan et al. [3], who use the ConvAE to learn temporal regularities in videos.

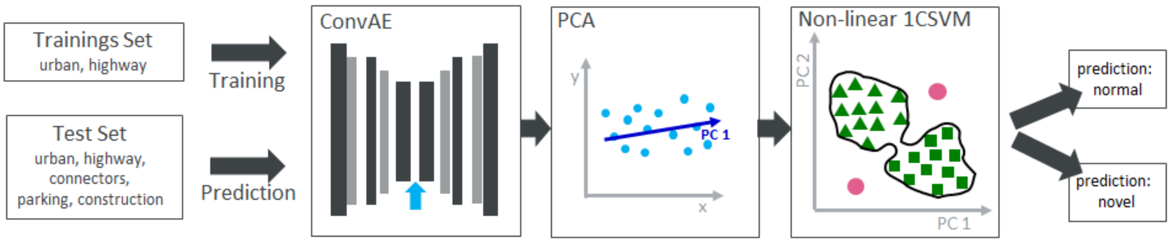

Fig. 1. Overview of the three modules that are combined.

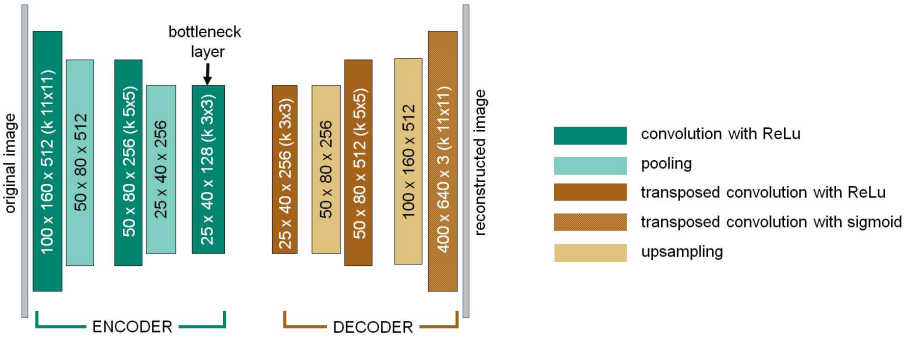

Fig. 2. The ConvAE is symmetric and consists of three convolution, two pooling, three transposed convolution and two upsampling layers.

3 System Overview

Three modules constitute the proposed novelty detection system: a ConvAE, a PCA and a 1CSVM, see Fig. 1. A ConvAE is a kind of neural network that consists of an encoding and a decoding part. The encoder extracts condensed high-level features from the input image. The decoder then reconstructs the image based on these features. The ConvAE is trained unsupervised, minimizing the reconstruction error (difference) between input and reconstructed image, the exact architecture is depicted in Fig. 2. Deep neural networks like ConvAEs tend to disentangle manifolds. Thus, the resulting feature space becomes more linear [1]. The PCA is a linear transformation that finds the directions with the highest variance in the data, the so-called principal components (PCs). The PCA is applied to reduce the dimensions of the features: They can be visualized more easily and the following 1CSVM is more efficient on lower input dimensions. A 1CSVM finds a hyperplane, defined by support vectors, that surrounds its training data [9]: data that is not similar to the training data lays outside of this hyperplane and is classified as novel. The features that characterize the novelties are expected to be highly non-linear, therefore a non-linear 1CSVM with a radial basis function (rbf) kernel is implemented. The computation of the 1CSVM requires the parameters ν and γ: ν is the lower bound on the fraction of support vectors, meaning that at least the ratio of ν samples of the training data are used as the support vectors. ν is also the upper bound on the fraction of outliers in the training data, meaning that a maximum of the ratio of ν samples of the training data lie outside the 1CSVM. γ is the kernel parameter of the rbf kernel, which uses a gaussian distribution to compute the similarity between two samples. γ can be interpreted as the inverse of the standard deviation. A high γ means that the samples need to be close together to be treated as similar, whereas a low γ means that they can be far away and still be treated as similar. If not otherwise specified, writing 1CSVM refers to the non-linear 1CSVM.

The major computation of all three modules takes place during training. The inference is very efficient since the input image is solely parsed through the network and the resulted features are inserted in the computed equations.

4 Experiments

The RGB images in the dataset have a size of 640x400 pixels. The images were taken by a front camera in a car and labeled according to five classes: urban street and highway images are normal classes. Connector (ramp on and off the highway), parking (open parking lot) and (urban) construction zones are novelties. The training dataset consists of 2028 urban and 2029 highway images. The test dataset consists of 688 urban, 512 highway, 204 parking, 43 construction and 736 connector images, see examples in Fig.3. The loss function of the ConvAE is the mean squared error. As the first layers of the ConvAE contain the simple features such as edges and the high-level features are desired to detect novelties, the bottleneck layer (3rd convolution layer) is the input to the PCA. The 1CSVM is computed on the PCs of the train images. 1CSVMs with different input dimensions and γ values are compared. Experiments showed, that varying the ν value has no noteworthy impact on the results.

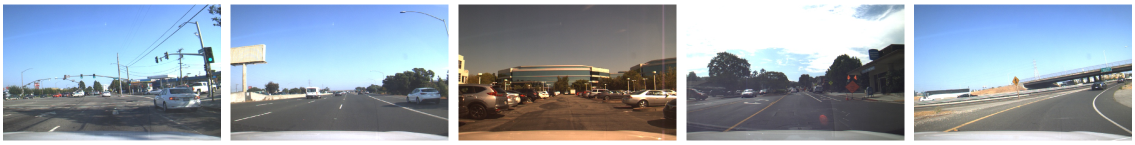

Fig. 3. Examples of the dataset, from left to right: The normal classes urban and highway, the novel classes parking, construction and connectors.

5 Results and Discussion

Fig. 4 shows the first two PCs of the extracted features of each test sample. The normal classes are not separable in this plot. This is expected, due to the way the ConvAE is trained: it is not discriminative and not using any information about the underlying class structure of the dataset. Highway and urban roads both have features in common that are significant for reconstruction. Fig. 4 also shows that parking and construction samples are not clearly separable from the normal data, but separable from each other. The PCA is only suitable for novelty detection, when the variation in the features that characterize the novelties

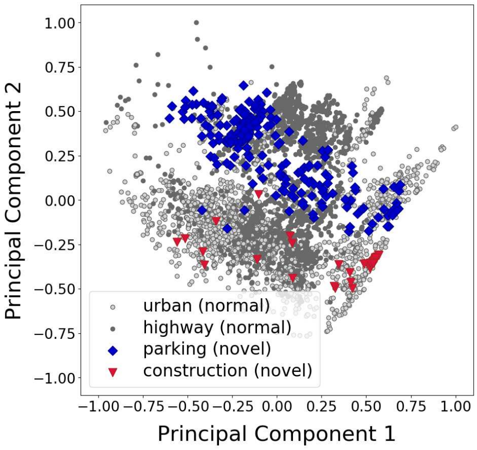

Fig. 4. The first two PCs of test images.

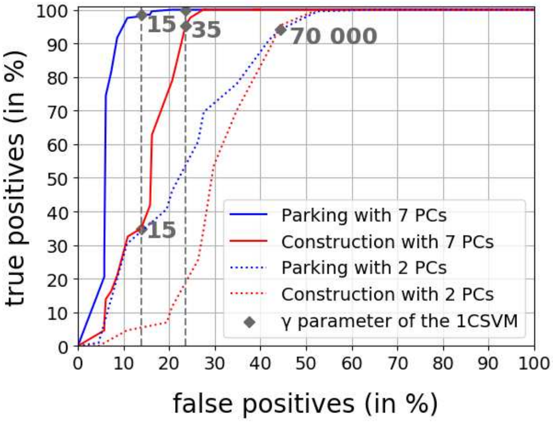

Fig. 5. ROC curves of the 1CSVM with varying γs and $\nu = 0.05$.

is not lost by computing the PCA. Shuy et al. [11] propose an approach using the first and the last PCs for novelty detection. Fig. 5 shows the Receiver Operator Characteristics (ROC) curves of the results of 1CSVMs. The 1CSVMs using seven input dimensions are better in detecting both types of novelties. The PCs are sorted in descending order of their variance on the training data. The results show, that adding the next PC to the input of the 1CSVM improves the final result. But the size of the improvement decreases per additional PC. Using more dimensions results in higher computing time, both at training and at inference. This poses a trade-off between the best results and a feasible time complexity. To achieve a good true positive rate the γ value has to be higher for construction than for parking. This means that the construction novelties are harder to separate from the normal data, compared to the parking data. This is in agreement with the results from the PCA, plotted in Fig. 4. Table 1 shows the exact detection results. It also shows the results using a supervised convolutional neural network (CNN) instead of the unsupervised ConvAE for feature extraction. The CNN was trained on the classification into urban and highway. It performs slightly better, but also needs labeled training data.

Table 1. Results of the unsupervised and supervised approach with different γs. All 1CSVMs are non-linear and use seven PCs.

	ConvAE + 1CSVM (unsup.)			CNN + 1CSVM (sup.)		
	tp (%)	fp (%)	γ	tp (%)	fp (%)	γ
parking	98,5	15,8	15	99	7	3.5
	100	24,7	35	100	23	35
construction	41,9	15,8	15	40	7	3.5
	97,7	24,7	35	99	23	35

The features of the input images are reduced by a factor of six in the ConvAE and by a factor of 18285 with the PCA: The higher dimension reduction takes place using a linear transformation instead of exploiting the non-linear power of the ConvAE.

6 Conclusion

The proposed novelty detection approach combines three modules: a convolutional autoencoder for feature extraction, a principal component analysis for dimension reduction and visualization and a non-linear one-class support vector machine for novelty detection. The results of the experiments comply with the intuitive expectations, achieve the goal of reducing the amount of data that needs hand-labeling and are close to a similar supervised approach. More complicated novelties are harder to detect, which can be seen in the increased false positive rate. Future work includes extensive hyperparameter tuning for the ConvAE, reducing the features in its bottleneck layer to observe the reciprocity between non-linear feature extraction and linear dimension reduction, automatically determining the ideal γ and ν parameters of the 1CSVM as well as model ensembles combining supervised and unsupervised approaches.

References

1. Brahma, P.P., Wu, D., She, Y.: Why deep learning works: A manifold disentanglement perspective
2. Erfani, S.M., Rajasegarar, S., Karunasekera, S., Leckie, C.: High-dimensional and large-scale anomaly detection using a linear one-class SVM with deep learning
3. Hasan, M., Choi, J., Neumann, J., Roy-Chowdhury, A.K., Davis, L.S.: Learning temporal regularity in video sequences
4. Jolliffe, I.T.: Principal Component Analysis. Springer New York (1986)
5. Marsland, S.: Novelty Detection in Learning Systems
6. Nguyen, M., Vien, N.A.: Scalable and interpretable one-class svms with deep learning and random fourier features
7. Pimentel, M.A., Clifton, D.A., Clifton, L., Tarassenko, L.: A review of novelty detection
8. Schlegl, T., Seebck, P., Waldstein, S.M., Schmidt-Erfurth, U., Langs, G.: Unsupervised Anomaly Detection with Generative Adversarial Networks to Guide Marker Discovery
9. Schölkopf, B., Williamson, R., Smola, A., Shawe-Taylor, J., Platt, J.: Support Vector Method for Novelty Detection
10. Seebck, P., Waldstein, S., Klimscha, S., Gerendas, B.S., Donner, R., Schlegl, T., Schmidt-Erfurth, U., Langs, G.: Identifying and Categorizing Anomalies in Retinal Imaging Data
11. Shyu, M.L., Chen, S.C., Sarinnapakorn, K., Chang, L.: A Novel Anomaly Detection Scheme Based on Principal Component Classifier
12. Simonyan, K., Vedaldi, A., Zisserman, A.: Deep inside convolutional networks: Visualising image classification models and saliency maps
13. Utkin, L.V., Zaborovsky, V.S., Lukashin, A.A., Popov, S.G., Podolskaja, A.V.: A Siamese Autoencoder Preserving Distances for Anomaly Detection in Multi-robot Systems