Flow-based SVDD for anomaly detection

Marcin Sendera [1] Marek Śmieja [1] Łukasz Maziarka [1] Łukasz Struski [1] Przemysław Spurek [1] Jacek Tabor [1]

Abstract

We propose FlowSVDD – a flow-based one-class classifier for anomaly/outliers detection that realizes a well-known SVDD principle using deep learning tools. Contrary to other approaches to deep SVDD, the proposed model is instantiated using flow-based models, which naturally prevents from collapsing of bounding hypersphere into a single point. Experiments show that FlowSVDD achieves comparable results to the current state-of-the-art methods and significantly outperforms related deep SVDD methods on benchmark datasets.

1. Introduction

Anomaly (novelty/outlier) detection refers to the identification of novel or abnormal patterns embedded in a large amount of typical (normal) data (Miljković, 2010). Anomaly detection algorithms find application in fraud detection systems, discovering failures in the industrial domain, detection of adversarial examples, etc..

In contrast to typical binary classification problems, where every class follows some probability distribution, an anomaly is a pattern that does not conform to the expected behavior. In consequence, a completely novel type of anomalies can occur at test time, which is not similar to any known anomalies. Moreover, in most cases, we do not have access to any anomalies at training time. In consequence, novelty detection is usually solved using unsupervised approaches, such as one-class classifiers, which focus on describing the behavior of available data (inliers). Any observation, which deviates from this behavior, is labeled as an outlier.

Our research is motivated by the idea of Support Vector Data Description (SVDD) (Tax & Duin, 2004), which obtains a spherically shaped boundary around a dataset by usage of soft margin and penalization of data points from outside the bounding region. We propose FlowSVDD – a one-class classifier based on flow-based models (Dinh et al., 2014), which finds a hypersphere with a minimal volume that encloses data. Since flow-based models are commonly used in the context of generative models, we redefine their cost function to minimize the volume of the bounding hypersphere instead of maximizing the log-likelihood function. On one hand, flow-based models allow us to calculate a Jacobian of a neural network at every point. In consequence, minimizing the volume of the hypersphere in the feature space leads to the minimization of the volume of the corresponding bounding region in the input space. On the other hand, since flow-based models give an explicit formula for the inverse mapping, we automatically get a parametric form for the corresponding bounding region in the input space. In contrast to deep SVDD models, our approach eliminates the problem of hypersphere collapse, which makes it easy to use.

Extensive experiments performed on typical benchmark datasets show that our method significantly outperforms the deep SVDD model while being comparative to state-of-the-art models for anomaly detection.

Our contribution is summarized as follows:

1. We propose an adaptation of the SVDD method to deep neural networks with the use of flow models.
2. We show that the realization of the SVDD loss function on flow-based models prevents from hypersphere collapse.
3. We experimentally compare FlowSVDD with Deep SVDD and current state-of-the-art methods.

2. Proposed model

Preliminaries: SVDD. Our approach is motivated by a classical Support Vector Data Description (SVDD) (Tax & Duin, 2004), which tries to find a minimal hypersphere to enclose the data. To allow the possibility of outliers in the training set, SVDD uses a soft margin and penalizes data points that lie outside the bounding hypersphere. If f maps input data to the output kernel space, then SVDD loss

[*] Equal contribution [1] Faculty of Mathematics and Computer Science, Jagiellonian University, Kraków, Poland. Correspondence to: Marcin Sendera <marcin.sendera@gmail.com>.

Second workshop on *Invertible Neural Networks, Normalizing Flows, and Explicit Likelihood Models* (ICML 2020), Virtual Conference

equals:

$$F(R,c;f) = R^2 + \frac{1}{\nu n}\sum_i \max(0, \|f(x_i) - c\|^2 - R^2) \quad (1)$$

where $c \in \mathbb{R}^D, R \in \mathbb{R}$ is the center and the radius of the hypersphere, respectively, and ν is the trade-off between the volume and boundary violations of the hypersphere, i.e. fraction of outliers.

The realization of SVDD using deep neural networks was presented in (Ruff et al., 2018) (it was termed DSVDD). However, direct minimization of the SVDD loss may lead to a trivial solution, i.e. the hypersphere collapses to a single point c. To avoid this negative behavior, it has been recommended that the center c must be something other than the all-zero-weights solution, and the network should use only unbounded activations and omit bias terms. While the two first conditions can be accepted, omitting bias terms in a network may lead to a sub-optimal feature representation due to the role of bias in shifting activation values.

To eliminate the above restrictions a recent work (Chong et al., 2020) proposes two regularizers, which prevent hypersphere collapse, and uses an adaptive weighting scheme to control the amount of penalization between the SVDD loss and the respective regularizer.

Flow-based SVDD. As an alternative to DSVDD, we realize the SVDD objective using flow models. Let us recall that a neural network $f : \mathbb{R}^D \to \mathbb{R}^D$ is a flow model if the inverse mapping f^{-1} is given explicitly and the Jacobian determinant $w(x) = \det df(x)$ can be easily calculated. In our approach, we use a special class of flow models, in which Jacobian determinant is constant at every point, i.e. $w(x) = w$, such as NICE (Dinh et al., 2014). In this case, we get a natural correspondence between the volume of the bounding hypersphere in the output space and the volume of a bounding region in the input space, see below.

Let us first consider the simplest situation when $w = 1$. In such a scenario, the volume of any shape in the input space equals the volume of its image in the output feature space. In consequence, a direct minimization of the SVDD objective does not lead to the hypersphere collapse.

In a more general scenario, when $w \neq 1$, we need to include w in the SVDD objective. Observe that the Jacobian determinant of the mapping $f/w^{1/D}$ equals 1. Thus to get the equality of the volume in the input and output space, we redefine the SVDD loss (1) as follows:

$$F(R,c;f) = R^2 + \frac{1}{\nu n}\sum_{i=1}^{n} \max(0, \|f(x_i)/w^{1/D} - c\|^2 - R^2). \quad (2)$$

In a test phase, a given example x is deemed as an outlier if:

$$\|f(x)/w^{1/D} - c\| > R,$$

which is equivalent to

$$\|f(x) - w^{1/D}c\| > Rw^{1/D}.$$

In other words, inliers lie inside the ball $B(w^{1/D}c; Rw^{1/D})$.

3. Experiments

In this section, we experimentally examine FlowSVDD and compare it with several state-of-the-art approaches. FlowSVDD is implemented using the architecture of the NICE flow model (4 coupling layers – each consisted of 4 layers and 256 hidden dimensions) with constant Jacobian determinant and $\nu = 0.05$.

Illustrative example. To get the intuition behind FlowSVDD, we first consider 2-dimensional examples, which are easy to visualize. The results presented in Figure 1 show the resulting hyperspheres in the latent space and the corresponding bounding regions in the input space. At first glance, we can observe that the bounding region in the original space is close to the structure of inliers. In the latent space, FlowSVDD finds the center point c and radius R to enclose $(1-\nu)$ percentage of data inside the ball $B(c; R)$. Observe that, unlike the density-based flow models, FlowSVDD does not transform data into Gaussian distribution in a latent space.

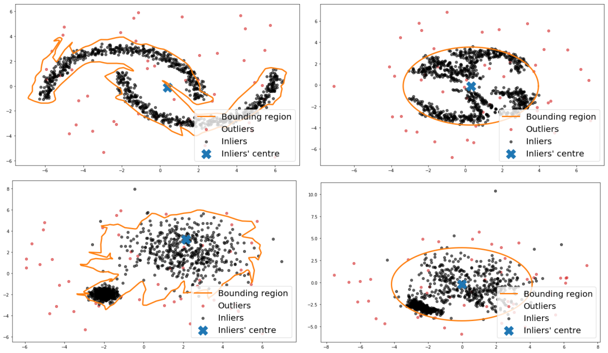

Figure 1. Enclosing hyperspheres in the latent space of FlowSVDD (right) and the corresponding bounding regions in the input space (left).

Benchmark data for anomaly detection. To provide quantitative assessment, we take into account Thyroid[1] and KDDCUP[2] datasets, which are typically used for anomaly

[1] http://odds.cs.stonybrook.edu/thyroid-disease-dataset/

[2] http://kdd.ics.uci.edu/databases/kddcup99/kddcup.testdata.unlabeled_10_percent.gz

detection. We use the standard training and test splits and follow exactly the same evaluation protocol as in (Wang et al., 2019). In particular, we use the F1 score and the Area Under Receiver Operating Characteristic curve (AUC).

Our model is compared with the following algorithms: (1) One-class SVM (OC-SVM) (Schölkopf et al., 2001), (2) Deep structured energy-based models (DSEBM) (Zhai et al., 2016), (3) Deep autoencoding Gaussian mixture model (DAGMM) (Zong et al., 2018), (4) variants of MQT – multivariate quantile map (NLL, TQM_1, TQM_2, TQM_∞) (Wang et al., 2019) and (5) Deep Support Vector Data Description (DSVDD) (Ruff et al., 2018) - another implementation of SVDD cost function in deep neural networks.

The results presented in Table 1 show that FlowSVDD model performs better than most methods on the Thyroid dataset and is significantly better than DSVDD in terms of the AUC metric. In the case of KDDCUP, FlowSVDD achieves a score in between the classical methods and current state-of-the-art.

Image datasets. To provide further experimental verification, we use two image datasets: MNIST and Fashion-MNIST. In contrast to the previous comparison, these two datasets are usually used for multiclass classification and thus need to be adapted to the problem of anomaly detection. For this purpose, each of the ten classes is deemed as the nominal class while the rest of the nine classes are deemed as the anomaly class, which results in 10 scenarios for each dataset.

We additionally compare FlowSVDD with the following models: (1) Geometric transformation (GT) (Golan & El-Yaniv, 2018), Variational autoencoder (VAE) (Kingma & Welling, 2013), Denoising autoencoder (DAE) (Vincent et al., 2008), Generative probabilistic novelty detection (GPND) (Pidhorskyi et al., 2018), Latent space autoregression (LSA) (Abati et al., 2019). In contrast to previous experiment, we only use TQM_2 and NLL as the only implementations of MTQ, because they output the highest value of AUC (Wang et al., 2019).

To present the results, we compute the ranking on each of 10 scenarios and summarize it using a box plot, see Figure 2. The results show that FlowSVDD significantly outperforms DSVDD in both datasets. In the case of the MNIST dataset, we observe that FlowSVDD is almost as good as the current state-of-the-art methods, like GT and NLL.

Finally, we analyze, which samples are localized close to or furthest from the center of bounding hypersphere. Results in Figure 3 shows that FlowSVDD maps regular images in the hypersphere center. Contrary, examples localized far from the center, could be hard to identify. It means that FlowSVDD gives results consistent with our intuition.

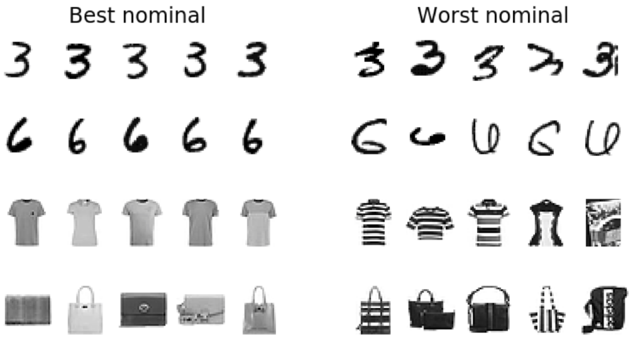

Figure 3. Best nominal (left) and worst nominal (right) examples determined by FlowSVDD for MNIST (top) and Fashion-MNIST (bottom).

4. Conclusion

The paper introduced FlowSVDD, which realizes the SVDD paradigm in the case of neural networks. Making use of flow-based models and an appropriate SVDD-like cost function, we find a minimal bounding region for a majority of data. Unlike other deep SVDD realizations, FlowSVDD does not change the determinant of a Jacobian matrix, which means that the resulting hypersphere cannot collapse in a latent space. The experimental results demonstrate that FlowSVDD presents a very good performance in the case of both artificial and real-world one-class settings.

Acknowledgements

The work of M. Śmieja was supported by the National Centre of Science (Poland) Grant No. 2018/31/B/ST6/00993. The work of P. Spurek was supported by the National Centre of Science (Poland) Grant No. 2019/33/B/ST6/00894. The work of J. Tabor was supported by the National Centre of Science (Poland) Grant No. 2017/25/B/ST6/01271.

References

Abati, D., Porrello, A., Calderara, S., and Cucchiara, R. Latent space autoregression for novelty detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 481–490, 2019.

Chong, P., Ruff, L., Kloft, M., and Binder, A. Simple and effective prevention of mode collapse in deep one-class classification. *arXiv preprint arXiv:2001.08873*, 2020.

Dinh, L., Krueger, D., and Bengio, Y. Nice: Non-linear independent components estimation. *arXiv preprint arXiv:1410.8516*, 2014.

Golan, I. and El-Yaniv, R. Deep anomaly detection us-

Table 1. Performance on two anomaly detection datasets.

Thyroid

	OC-SVM	DSEBM	DAGMM	NLL	TQM_1	TQM_2	TQM_∞	DSVDD	FlowSVDD
F1	.3887	.0403	.4782	.7312	.5269	.5806	.7527	-	.7097
AUC	-	-	-	-	-	-	-	0.749	.9797

KDDCUP

	OC-SVM	DSEBM	DAGMM	NLL	TQM_1	TQM_2	TQM_∞	DSVDD	FlowSVDD
F1	.7954	.7423	.9369	.9622	.9621	.9622	.9622	-	.9030
AUC	-	-	-	-	-	-	-	-	.9384

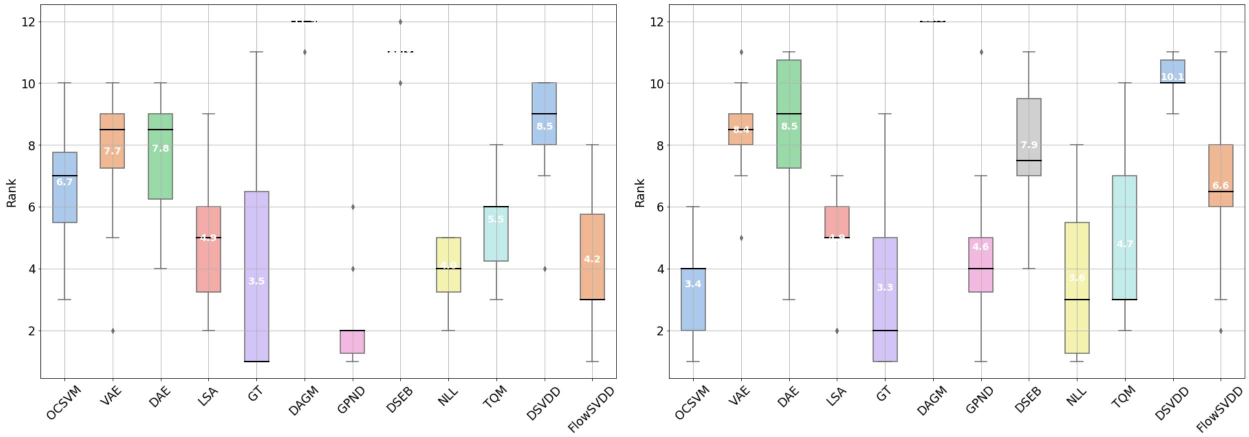

Figure 2. Box plots for rankings calculated on MNIST (left) and Fashion-MNIST (right). The median ranking is marked by a line, while the average ranking is marked with a number.

ing geometric transformations. In *Advances in Neural Information Processing Systems*, pp. 9758–9769, 2018.

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

Miljković, D. Review of novelty detection methods. In *The 33rd International Convention MIPRO*, pp. 593–598. IEEE, 2010.

Pidhorskyi, S., Almohsen, R., and Doretto, G. Generative probabilistic novelty detection with adversarial autoencoders. In *Advances in neural information processing systems*, pp. 6822–6833, 2018.

Ruff, L., Vandermeulen, R., Goernitz, N., Deecke, L., Siddiqui, S. A., Binder, A., Müller, E., and Kloft, M. Deep one-class classification. In *International conference on machine learning*, pp. 4393–4402, 2018.

Schölkopf, B., Platt, J. C., Shawe-Taylor, J., Smola, A. J., and Williamson, R. C. Estimating the support of a high-dimensional distribution. *Neural computation*, 13(7): 1443–1471, 2001.

Tax, D. M. and Duin, R. P. Support vector data description. *Machine learning*, 54(1):45–66, 2004.

Vincent, P., Larochelle, H., Bengio, Y., and Manzagol, P.-A. Extracting and composing robust features with denoising autoencoders. In *Proceedings of the 25th international conference on Machine learning*, pp. 1096–1103, 2008.

Wang, J., Sun, S., and Yu, Y. Multivariate triangular quantile maps for novelty detection. In *Advances in Neural Information Processing Systems*, pp. 5061–5072, 2019.

Zhai, S., Cheng, Y., Lu, W., and Zhang, Z. Deep structured energy based models for anomaly detection. *arXiv preprint arXiv:1605.07717*, 2016.

Zong, B., Song, Q., Min, M. R., Cheng, W., Lumezanu, C., Cho, D., and Chen, H. Deep autoencoding gaussian mixture model for unsupervised anomaly detection. 2018.