When more is less: Adverse Effects in Outlier Exposure

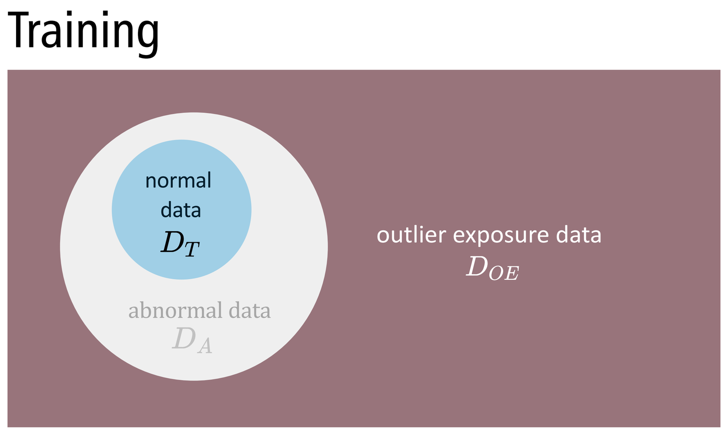
Jennifer Matthiesen, Ulf Brefeld

In **Anomaly detection** often only normal data samples are given during training. A recent strategy, named **Outlier Exposure (OE)**, tries to overcome the sparsity of the given data by including an auxiliary dataset of outliers. Experimentally we discover the impact of different classes, when used in the OE dataset. Further, we show that even using more classes in the OE dataset during training can result in decreasing the performance.

Methodology of Outlier Exposure

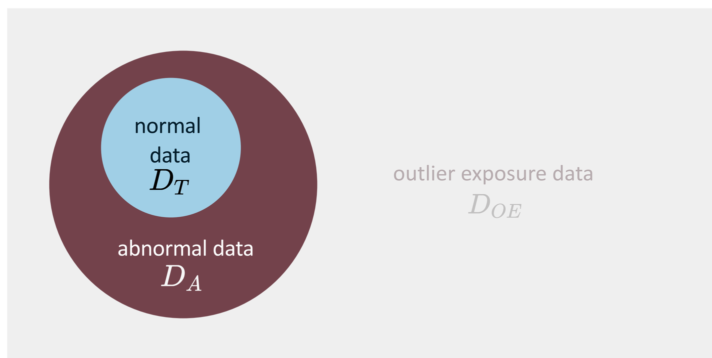
The original thought of the OE approach derived from the idea that, for a natural image AD problem, one has a huge amount of random natural images at hand which are likely not normal, which could be used as examples for anomalies (Hendrycks et al. 2019). **The idea is as follows:** Besides the samples from the normal class, we use additional data from another dataset.

Training

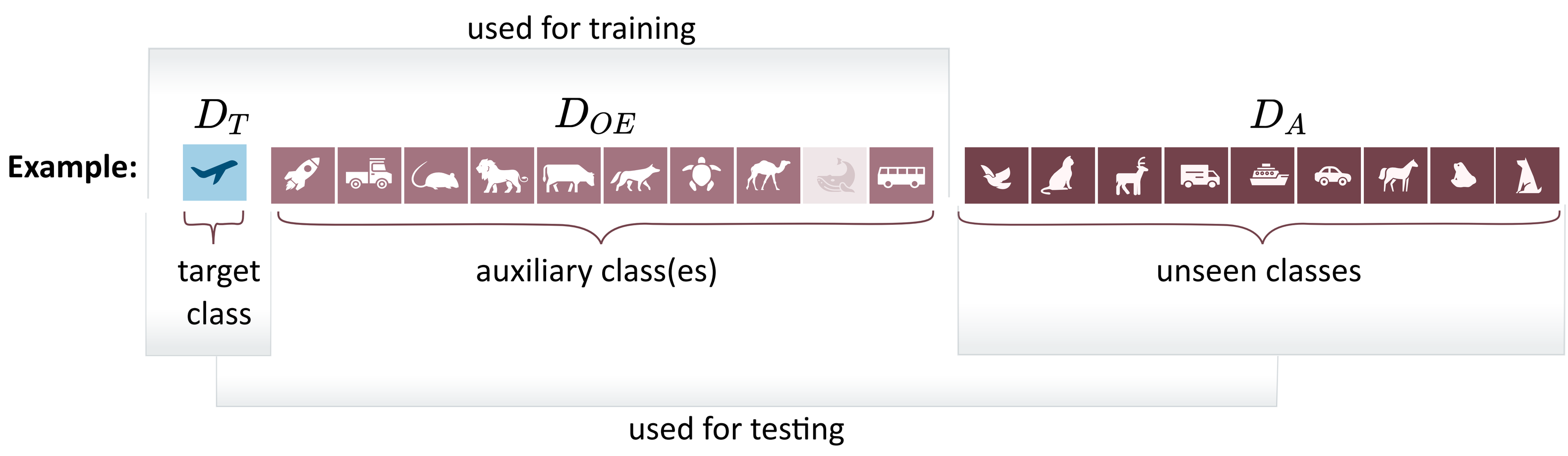


We train the models using the train set of the normal class and up to 10 classes of the outlier exposure dataset.

Testing



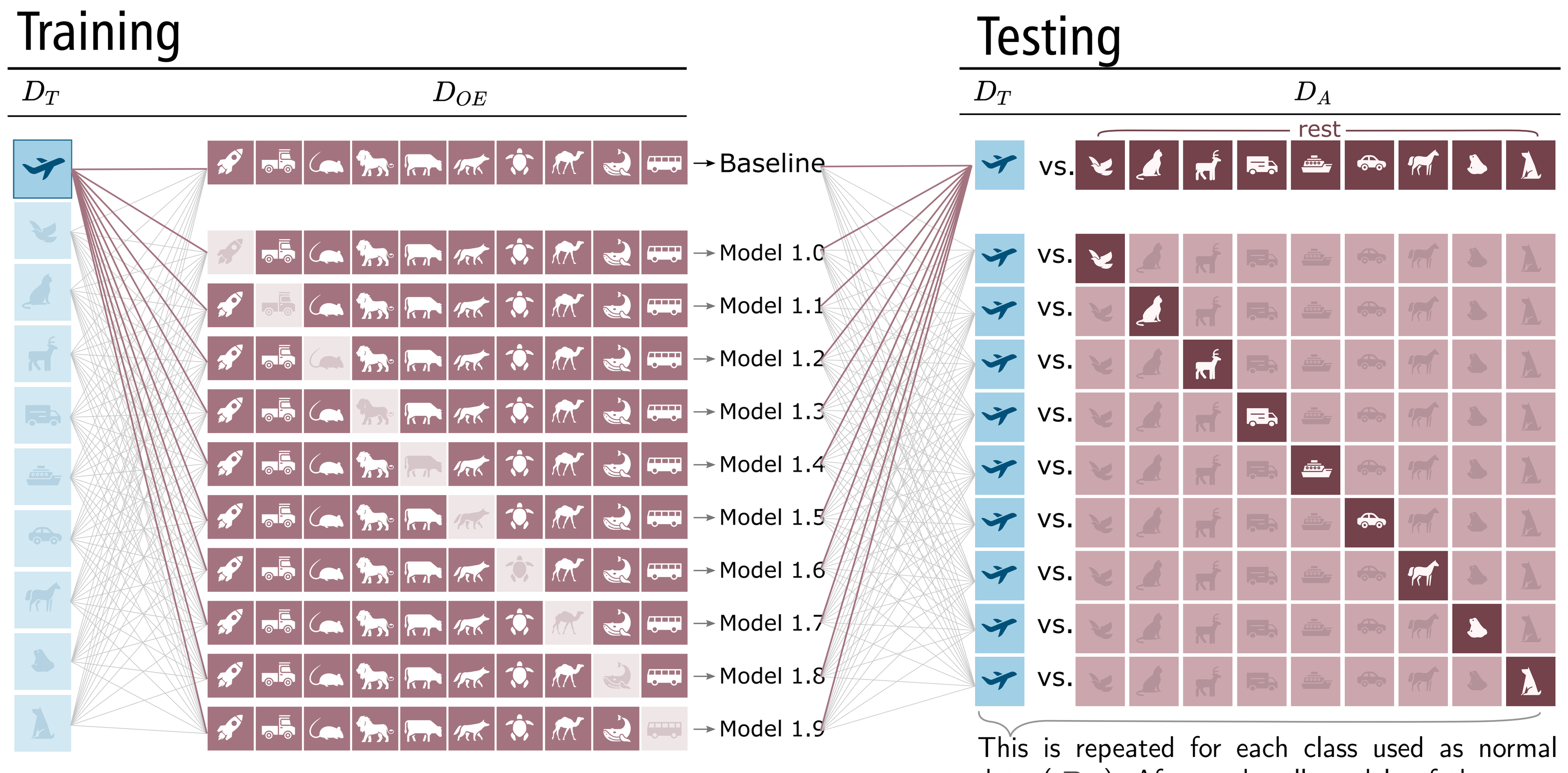
We test in a one vs. rest manner using the test set of the normal class and the 9 remaining classes as anomalous.



The models are trained on the target class as well as on the classes in the outlier exposure dataset, while test data includes the test set of the target class and the remaining classes of the considered dataset.

Influence of OE Classes: Training and Testing

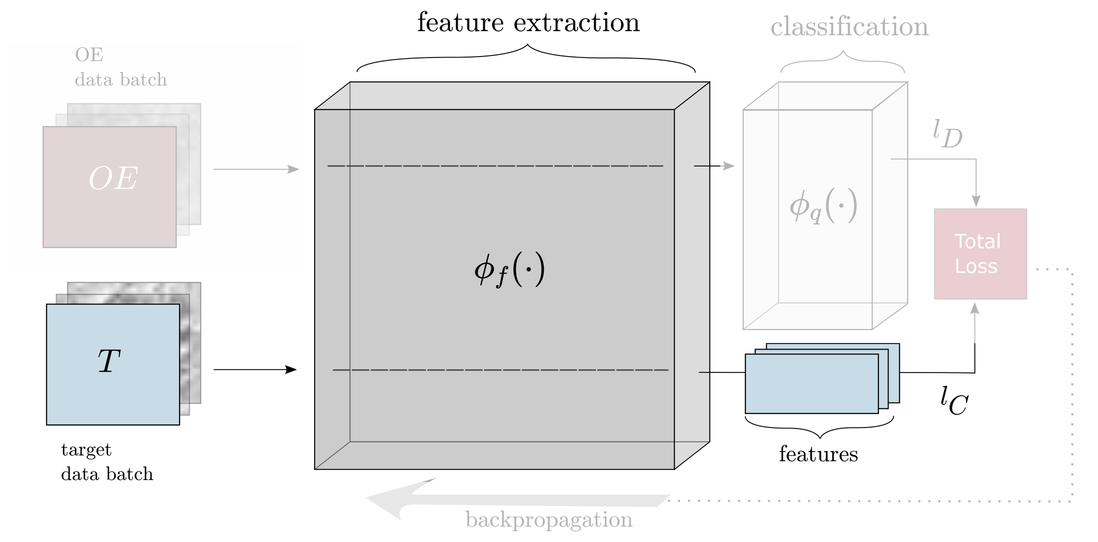
We study multiple variations of training data and its effects on the performance. We observe signicant dependencies of the target data and data used as OE that may either foster or prohibit predictive performance.



This is repeated for each class used as normal data (D_T). Afterwards, all models of the same test senarios are compaired against the baseline. This determines the influence the performance of the left out class

Network Architecture

The network $\phi_f(\cdot)$ extracts features from both datasets, while additional layers of $\phi_q(\cdot)$ are used for the auxiliary data to calculate l_D .



$$l_D = -\sum_{i=1}^{C} y_i log(\phi_q(x_i;\theta))^2$$

$$l_C = \frac{1}{n}\sum_{i=1}^{n}(\phi_f(x_i;\theta) - m_i)^2$$

$$l_D(D_{OE}) + \lambda l_C(D_T)$$

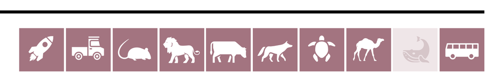
Using additional data during training, we propose a **composite loss** to address the two desired characteristics: **comparativeness and descriptiveness** (cf. Patel et al.).

- The **descriptive loss** l_D is calculated by the the cross entropy loss over all classes C.
- The **compactness loss** l_C is computed by the squared intra-batch distances. m_i is the mean vector of the rest of the features of the regarded sample.

Preliminary Results

Training with/ without whales

Cls. in training: D_{OE}



Cls. in testing: D_T	D_A	AUC	AUC_{all}	diff.
plane	rest	82.1	**79.6**	2.5
bird	rest	73.7	**71.3**	2.4
deer	rest	75.9	**73.3**	2.6
cat	rest	73.1	74.2	-1.1

Training with/ without wolves

Cls. in training: D_{OE}

Cls. in testing: D_T	D_A	AUC	AUC_{all}	diff.
plane	dog	83.9	87.1	- 3.2
bird	dog	60.5	60.6	- 0.5
deer	dog	68.1	72.2	- 4.1
cat	dog	54.3	54.9	- 0.6

- Using whales in D_{OE} results in a worse performance.
- When we try to distinguish between a target class and dogs, it is better when wolves are present in D_{OE}.
- When we try to distinguish between a target class and cats, it is better when wolves are present in D_{OE} .
- If both D_T and D_A are animals, it is good to have cattles in D_{OE} .

Conclusion

Increasing the variety of alternative classes (anomalies) should increase performance, but it does not.

Using more classes in the OE dataset during training can result in decreased performance.

Check out the abstract for more information!