Project MVTec Anomaly Detection

Report 2: Modeling

### Overview

* **Train- / Test- / Validation-Sets**
  + Split proportions of train/test/validation/crossvalidation
  + synthetic validation data
* **Data Preprocessing**
  + Feature Selection?
  + Standardization / Normalization
  + PCA
  + ResNet for feature extraction
* **Machine Learning Models and results on different datasets**
  + One Class SVM
  + Isolation Forest
  + Local Outlier Factor
  + Elliptic Envelope
  + DBSCAN
* **Deep Learning Models and results on different datasets**
  + Autoencoder
  + Variational Autoencoder
  + CNN
* **Conclusion**

### Datasets for Training, Testing and (Cross-)Validation

**[Mici]**

* which sets have been divided into which subsets, at which proportion?
* creation of synthetic data vor (cross-)validation

### Data Preprocessing

**“Standard” features**

* which columns have been dropped?
* sorting of datasets by categories
* Standardization and Normalization
* PCA

**ResNet Features**

* which Nets
* which Layers

### Machine Learning Models

**One Class SVM [peppe: template for other models]**

* short model description
* hyperparameter grid
* results on different datasets

**Isolation Forest [Mici & Majid, according to peppes template]**

* short model description
* hyperparameter grid
* results on different datasets

**Local Outlier Factor [Mici & Majid, according to peppes template]**

* short model description
* hyperparameter grid
* results on different datasets

**Elliptic Envelope [Mici & Majid, according to peppes template]**

* short model description
* hyperparameter grid
* results on different datasets

**DBSCAN [Mici, optional]**

* short model description
* hyperparameter grid
* results on different datasets

### Hybrid Approach (optional)

**ResNet50 with KNN [peppe]**

### Deep Learning Models

**Autoencoder + ResNet50 for FE**

* short model description
* hyperparameter grid
* results on different datasets

**Autoencoder + custom CNN for FE**

* short model description
* hyperparameter grid
* results on different datasets

### Conclusion

* best Model
* outlook: what could/should be done next?