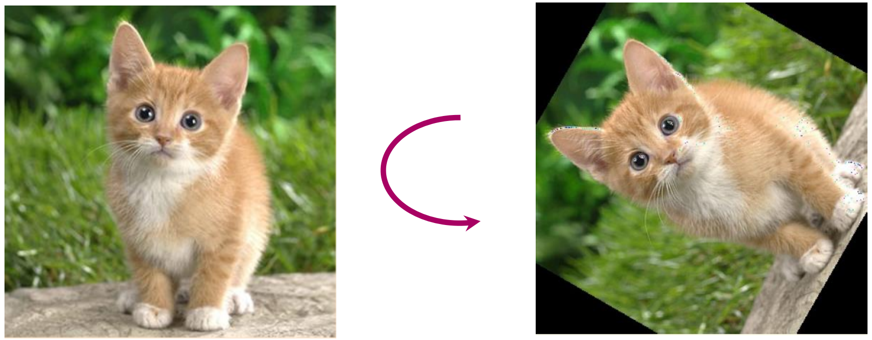
Neural Transformation Learning For Deep Anomaly Detection Beyond Images

Chen Qiu[1,2], Timo Pfrommer[1], Marius Kloft[2], Stephan Mandt[3], Maja Rudolph[1]

1: Bosch Center for AI, Germany
2: Technical University Kaiserslautern, Germany
3: University of California, Irvine, USA

BOSCH

Self-supervised anomaly detection
Works great for images



Good hand-crafted transformations help networks to learn good features for downstream tasks, including anomaly detection

[Golan & El-Yaniv, 2018; Wang et al., 2019; Chen et al., 2020; Sohn et al., 2021; Sehwag et al., 2021]

Figure from *Wang et al. Effective end-to-end unsupervised outlier detection via inlier priority of discriminative network.*

Chen Qiu, Timo Pfrommer, Marius Kloft, Stephan Mandt, Maja Rudolph

Self-supervised anomaly detection
Motivation

	value	unit
height	180	cm
weight	65	kg

value	unit	cm
	180	kg
height	weight	65

What transformations are best suited for other types of data

beyond images?

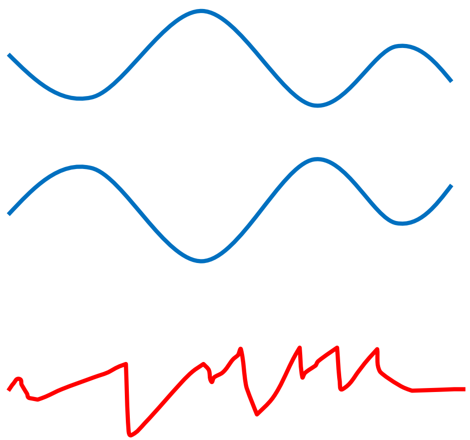
Chen Qiu, Timo Pfrommer, Marius Kloft, Stephan Mandt, Maja Rudolph

Neural Transformation Learning for Anomaly Detection

Goals & Tasks

- Anomaly detection on real-world time series data from various domains:

  detect abnormal time series on a *whole sequence level* .

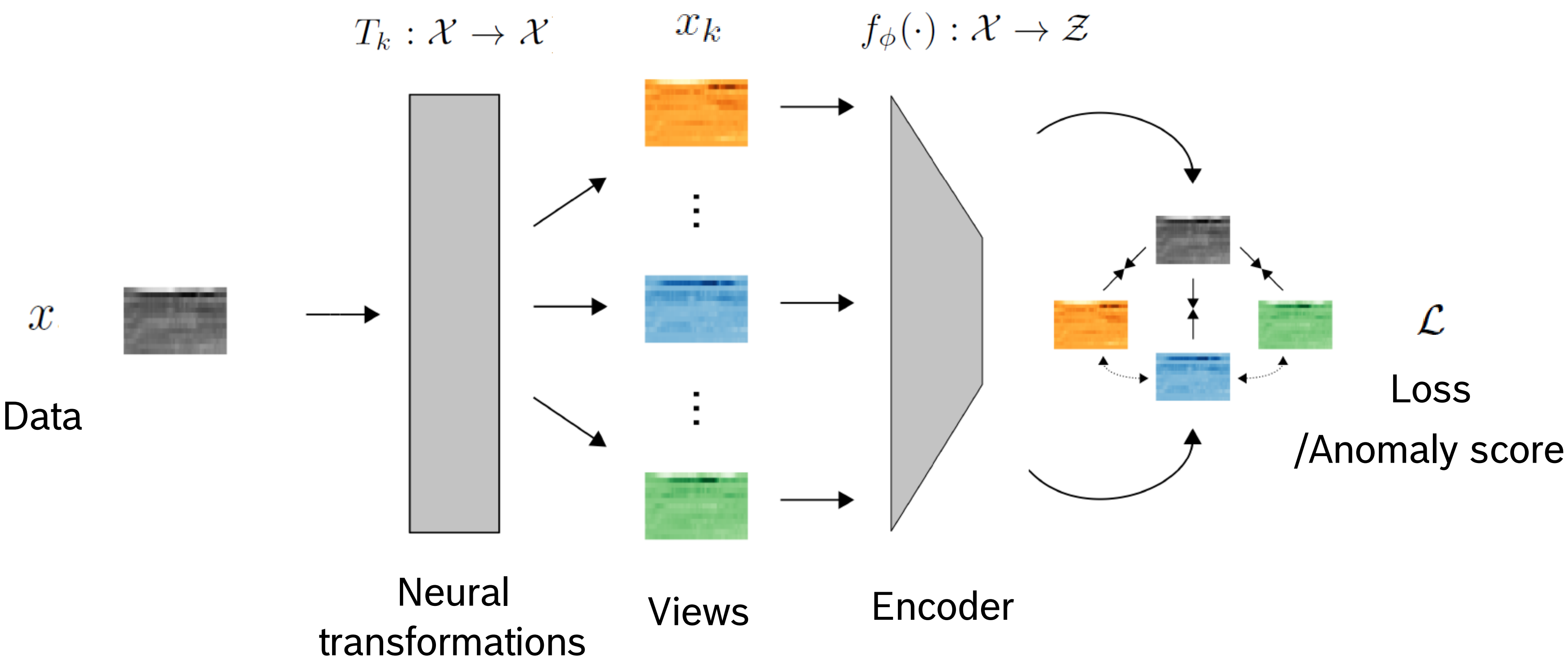


- Anomaly detection on tabular data from medical and cyber-security domains:
  detect abnormal samples in a tabular form.

	Test 1	Test 2	Test 3
P1	Normal	Normal	Normal
P2	Normal	Normal	Normal
P3	Low	High	High

Chen Qiu, Timo Pfrommer, Marius Kloft, Stephan Mandt, Maja Rudolph

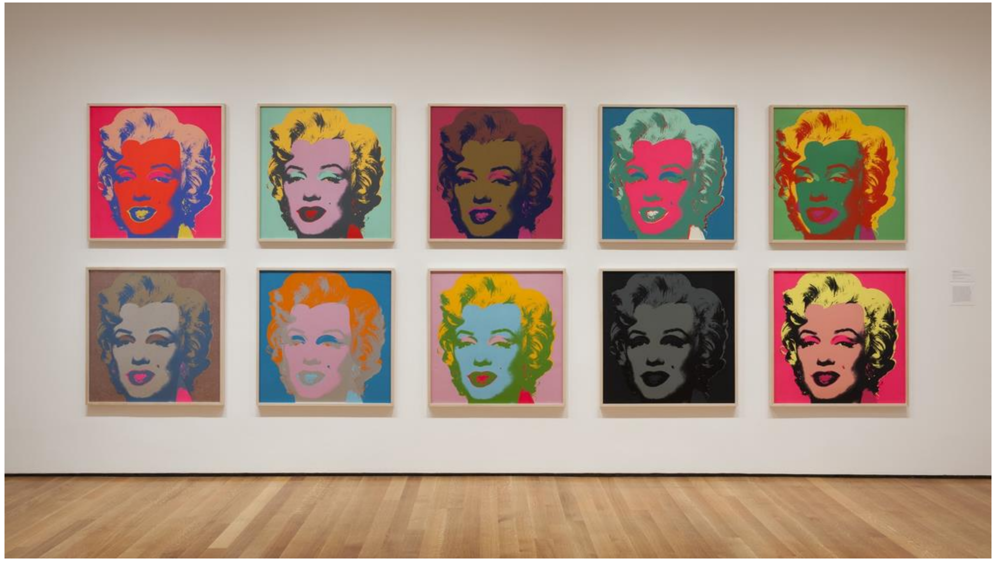
Neural Transformation Learning for Anomaly Detection
Overview



Chen Qiu, Timo Pfrommer, Marius Kloft, Stephan Mandt, Maja Rudolph

Neural Transformation Learning for Anomaly Detection
How to learn good transformations?

- Semantics: The transformations should produce views that share relevant semantic information with the original data.
- Diversity: The transformations should produce diverse views of each sample.



Photograph by Jonathan Muzikar, *"Andy Warhol: Campbell's Soup Cans and Other Works, 1953-1967"*

Chen Qiu, Timo Pfrommer, Marius Kloft, Stephan Mandt, Maja Rudolph

Neural Transformation Learning for Anomaly Detection
How to learn good transformations?

- Semantics: The transformations should produce views that share relevant semantic information with the original data.
- Diversity: The transformations should produce diverse views of each sample.
- Deterministic Contrastive Loss (DCL):

$$\mathcal{L} := \mathbb{E}_{x\sim\mathcal{D}}\left[-\sum_{k=1}^{K}\log\frac{h(x_k,x)}{h(x_k,x)+\sum_{l\neq k}h(x_k,x_l)}\right],$$

where $h(x_k,x_l) = \exp(\mathrm{sim}(f_\phi(T_k(x)), f_\phi(T_l(x)))/\tau)$

$$\mathrm{sim}(z,z') := z^T z'/\|z\|\|z'\|$$

Chen Qiu, Timo Pfrommer, Marius Kloft, Stephan Mandt, Maja Rudolph

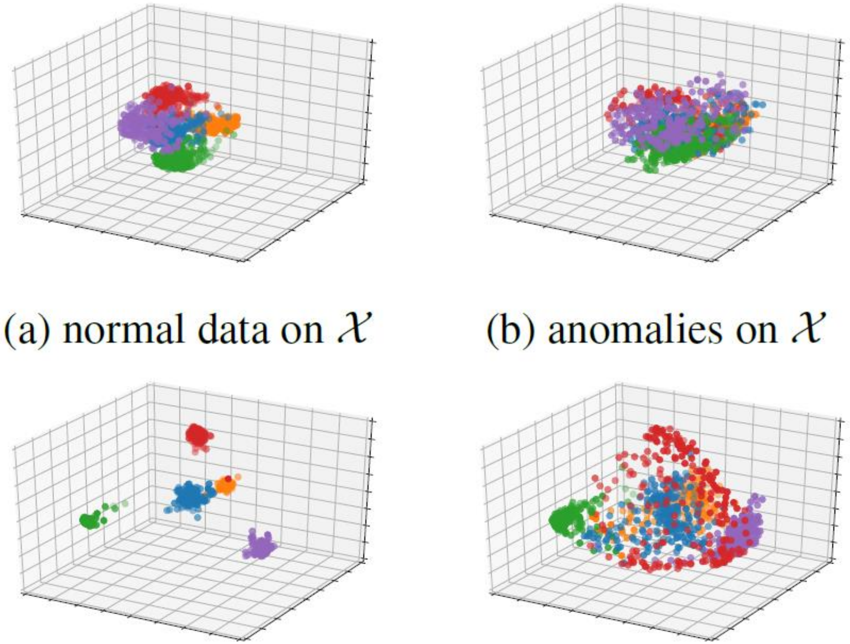
Neural Transformation Learning for Anomaly Detection
Visualization & Insights

- Visualization of original samples (blue) and different views in data space and in the embedding space.

Separation of Transformations

v.s.

Overlap between Transformations



(a) normal data on $\mathcal{X}$ (b) anomalies on $\mathcal{X}$

(c) normal data on $\mathcal{Z}$ (d) anomalies on $\mathcal{Z}$

- Data is from SAD dataset.
- Colors code for transformation (blue = untransformed)
- Data and embeddings are projected to 3D using PCA.

Chen Qiu, Timo Pfrommer, Marius Kloft, Stephan Mandt, Maja Rudolph

BOSCH

Neural Transformation Learning for Anomaly Detection
Theoretical analysis

Does an edge-case of unsuited transformations minimize the loss?

	'Constant' edge-case $f_\phi(T_k(x)) = Cc_k$	***'Identity' edge-case*** $T_k(x) = x$	***Suitable?***
Classification loss [1]	✓	x	x (Prop.1)
NCE loss [2]	x	✓	x (Prop.1)
DCL (our)	x	x	✓ (Prop.3)

[1] Golan, I. and El-Yaniv, R. Deep anomaly detection using geometric transformations. 2018
[2] Chen et al. A simple framework for contrastive learning of visual representations. 2020

Chen Qiu, Timo Pfrommer, Marius Kloft, Stephan Mandt, Maja Rudolph

Neural Transformation Learning for Anomaly Detection
Quantitative results on time series and tabular data

	SAD	NATOPS	CT	EPSY	RS
OC-SVM	95.3	86.0	97.4	61.1	70.0
IF	88.2	85.4	94.3	67.7	69.3
LOF	98.3	89.2	97.8	56.1	57.4
RNN	81.5±0.4	89.5±0.4	96.3±0.2	80.4±1.8	84.7±0.7
LSTM-ED	93.1±0.5	91.5±0.3	79.0±1.1	82.6±1.7	65.4 ±2.1
Deep SVDD	86.0±0.1	88.6±0.8	95.7±0.5	57.6±0.7	77.4±0.7
DAGMM	80.9±1.2	78.9±3.2	89.8±0.7	72.2±1.6	51.0±4.2
GOAD	94.7±0.1	87.1±1.1	97.7±0.1	76.7±0.4	79.9±0.6
DROCC	85.8±0.8	87.2±1.4	95.3±0.3	85.8±2.1	80.0±1.0
fixed Ts	96.7±0.1	78.4±0.4	97.9±0.1	80.4±2.2	**87.7**±0.8
NeuTraL AD	**98.9**±0.1	**94.5**±0.8	**99.3**±0.1	**92.6**±1.7	86.5±0.6

	Arrthythmia	Thyroid	KDD	KDDRev
OC-SVM	45.8	38.9	79.5	83.2
IF	57.4	46.9	90.7	90.6
LOF	50.0	52.7	83.8	81.6
Deep SVDD	53.9±3.1	70.8±1.8	99.0±0.1	98.6±0.2
DAGMM	49.8	47.8	93.7	93.8
GOAD	52.0±2.3	74.5±1.1	98.4±0.2	98.9±0.3
DROCC	46	27	-	-
NeuTraL AD	**60.3**±1.1	**76.8**±1.9	**99.3**±0.1	**99.1**±0.1

Chen Qiu, Timo Pfrommer, Marius Kloft, Stephan Mandt, Maja Rudolph

THANK YOU

Paper Link

Correspondence: Chen.Qiu@de.bosch.com