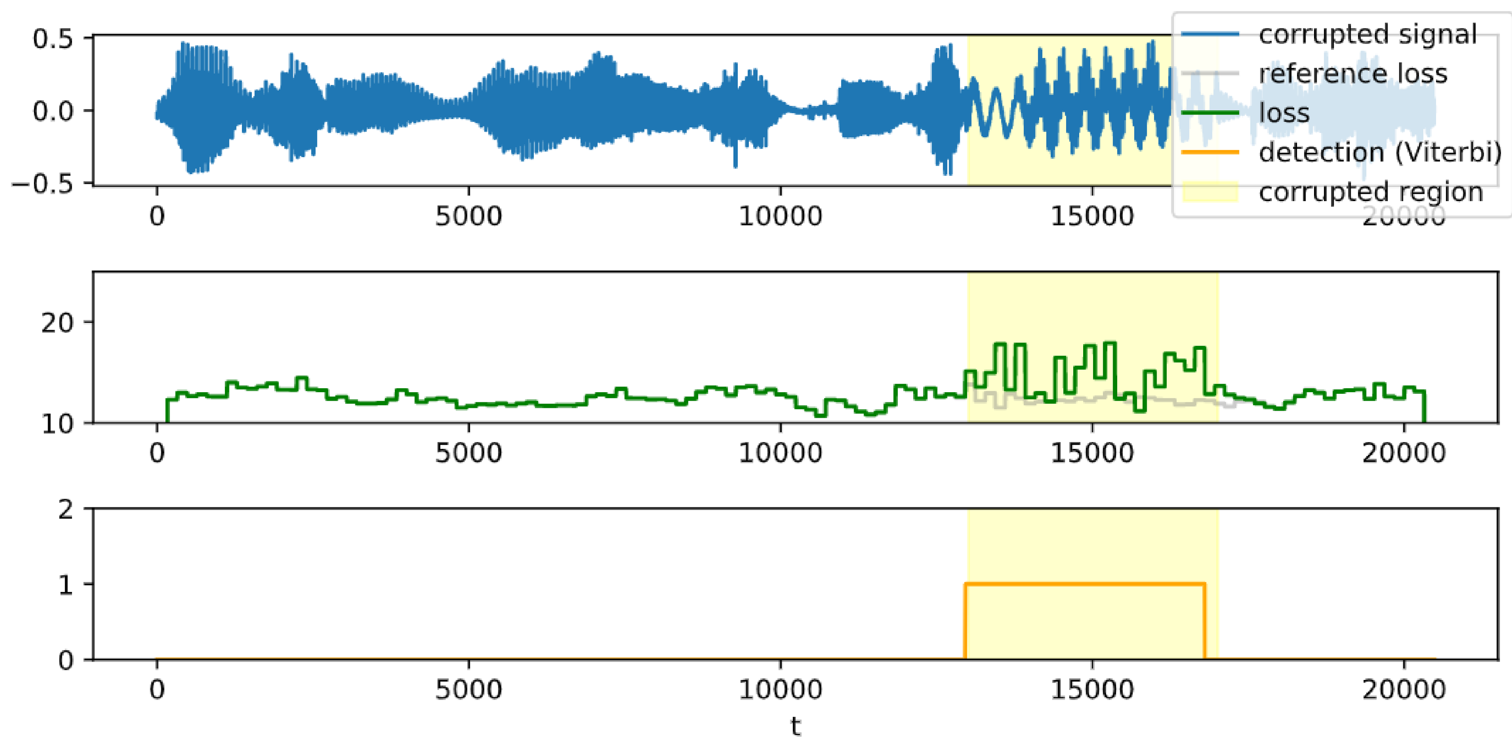
Anomalous Region Detection in Time Series with Local Neural Transformations (LNT)

Tim Schneider, Chen Qiu, Maja Rudolph

BOSCH · University of Stuttgart Germany · Technische Universität Kaiserslautern

Task & Goals

- Find anomalous regions within a time series (sub-sequence level)
- Many important applications in health and industry
- **Goal:** use performance on self-supervision task to score anomalies
- Very successful on image domain, e.g. transformation prediction with rotations or color distortions



„How to find good transformations for time series anomaly detection?"
➔ learn from data

Transformation Learning Principles

Not every function learned from data is a valid transformation / view. We follow / extend ideas of *Qiu et. al (2021)* to guide the learning with principles.

Propose **three principles** for transformation learning:

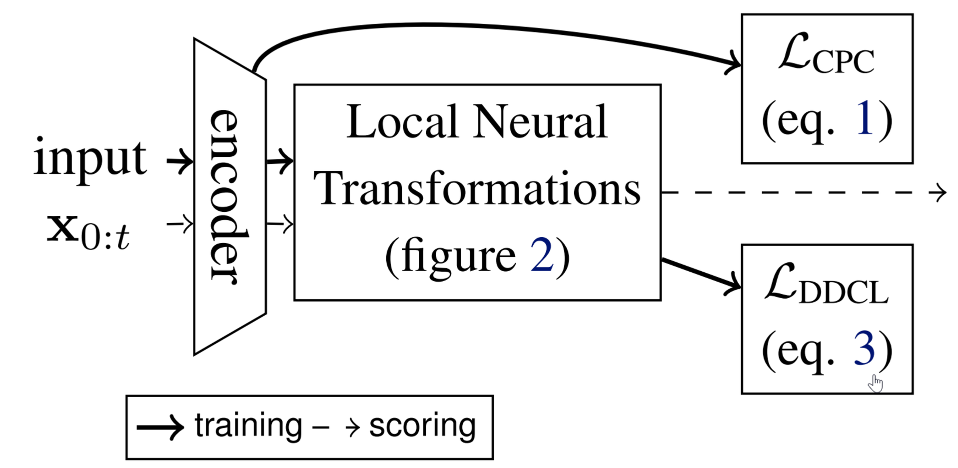
- **Semantics:** Views produced by learned transformations should share significant semantic information with the original sample.
- **Diversity:** Learned transformations should produce diverse views for each sample, resulting in varietal and difficult self-supervision tasks that need strong semantic features to be solved.
- **Locality:** The transformations should affect the data only in a local neighbourhood, while respecting the global context of the series.

References

[1] Qiu, C., Pfrommer, T., Kloft, M., Mandt, S., and Rudolph, M. Neural transformation learning for deep anomaly detection beyond images. arXiv preprint arXiv:2103.16440, 2021

[2] Oord, A. v. d., Li, Y., and Vinyals, O. Representation learning with contrastive predictive coding. arXiv preprint arXiv:1807.03748, 2018.

Architecture Overview of LNT



$$\mathcal{L} = \mathcal{L}_{\text{CPC}} + \lambda \cdot \mathcal{L}_{\text{DDCL}}$$

How transformations are trained?

Joint training of encoder (CPC) and Local Neural Transformations with two contrastive losses and exponentiated similarity measures f,h:

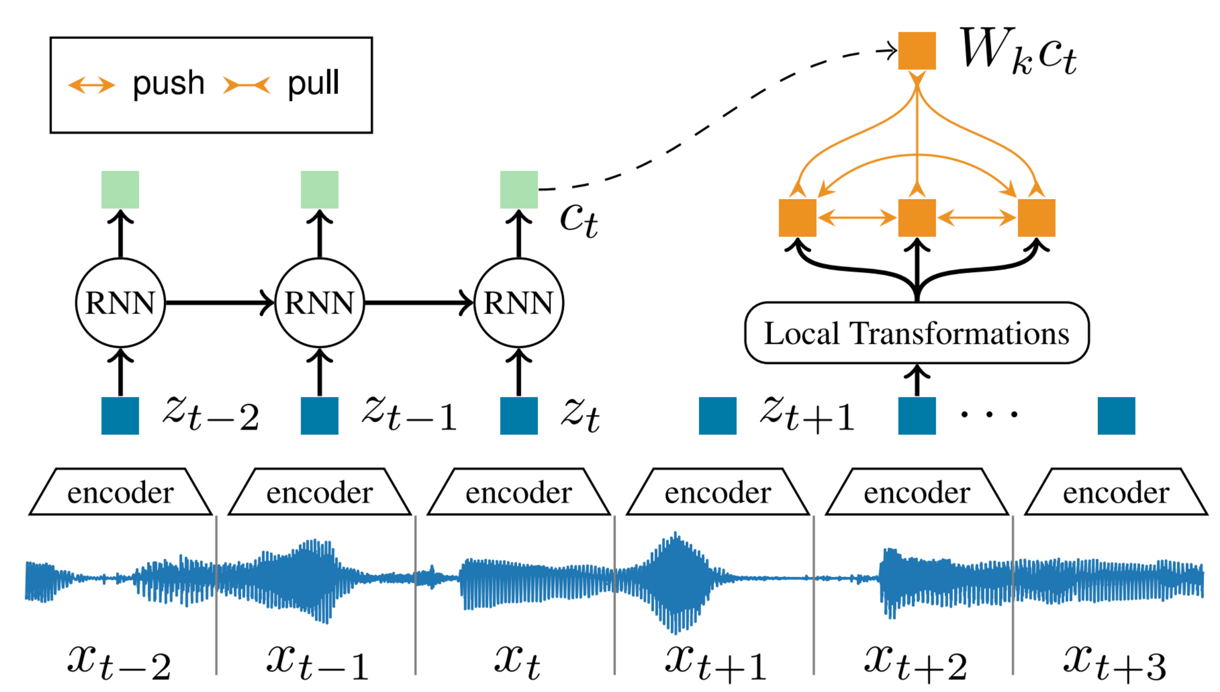
$$\mathcal{L}_{\text{CPC}} = -\mathop{\mathbb{E}}_{Z}\left[\log \frac{f_k(z_{t+k}, c_t)}{\sum_{z_j \in Z} f_k(z_j, c_t)}\right]$$

following *Oord et. al (2018)* and a new **Deterministic Dynamic Contrastive Loss (DDCL)** for training transformations

$$\mathcal{L}_{\text{DDCL}} = -\mathop{\mathbb{E}}_{Z}\left[\sum_{k=1}^{K}\sum_{t}^{T}\sum_{l \in L} \ell_t^{(k,l)}\right] \qquad z_t^{(l)} := \mathcal{T}_l(z_t)$$

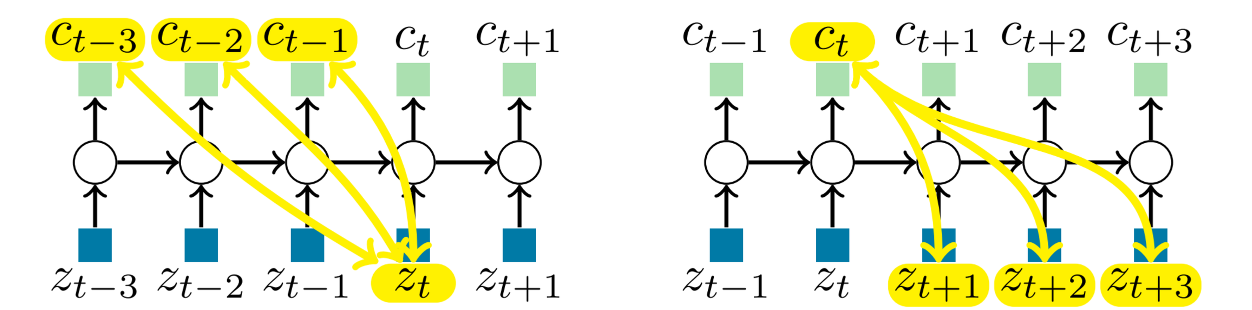
$$\ell_t^{(k,l)} = \log \frac{h(z_{t+k}^{(l)}, W_k c_t)}{h(z_{t+k}^{(l)}, W_k c_t) + \sum_{m \neq l} h(z_{t+k}^{(l)}, z_{t+k}^{(m)})}$$

Principles can be interpreted as "pushing" and "pulling" different representations in latent space:



How anomalies are scored / predicted ?

➔ Reuse DDCL loss for scoring (performance measure). Different loss contribution need to be combined:



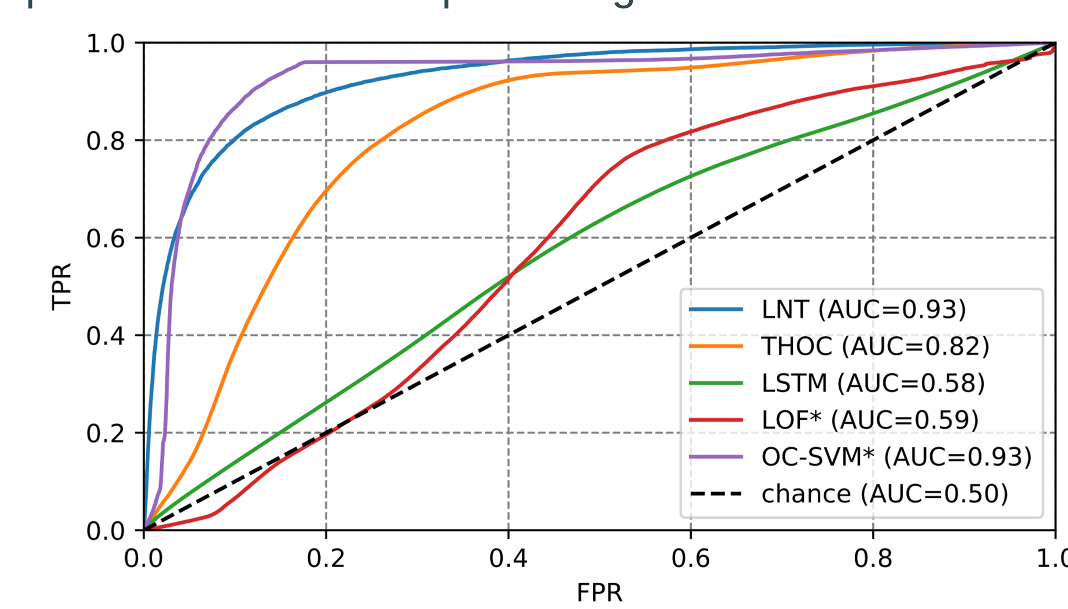
(a) Forward Scoring (b) Backward Scoring

$$\ell_t = \sum_{k=1}^{K}\sum_{l \in L} \ell_{t-k}^{(k,l)} \qquad \ell_t = \sum_{k=1}^{K}\sum_{l \in L} \ell_t^{(k,l)}$$

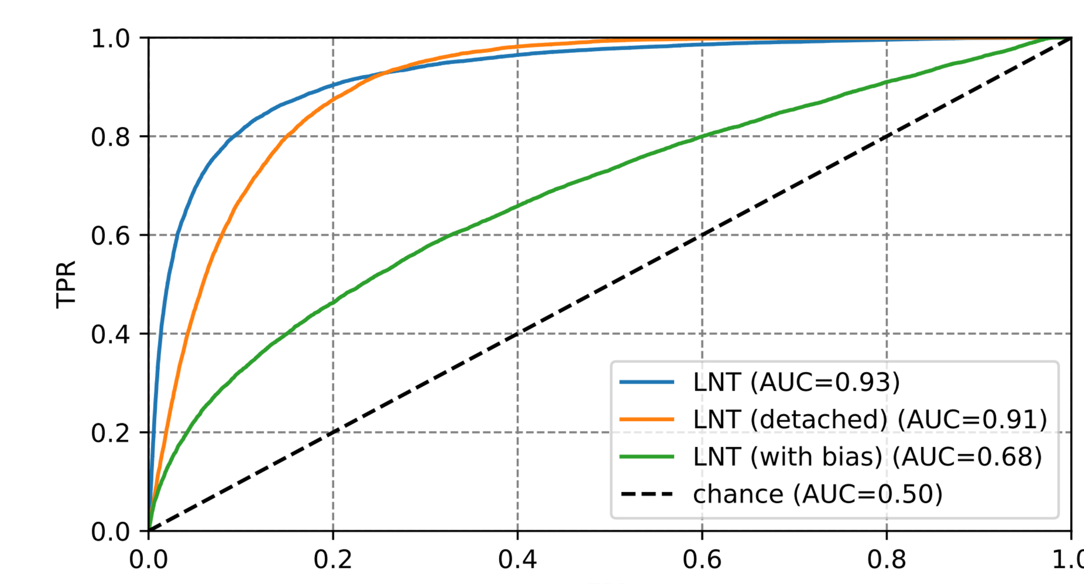
➔ Backward scoring smoothes and displaces scores: choose forward scoring
➔ Extract the maximum likelihood state sequence in an **HMM** for binary decision

First Experimental Results with LibriSpeech

LibriSpeech dataset with artificial anomaly regions by randomly placed additive pure sine tones of varying frequency and length.
➔ LNT outperforms several deep learning baselines:



➔ Dropping biases and joint training are beneficial for the performance:



Next Steps / Current Work:

- Domain transfer to datasets with (real) annotated outliers / anomalies
- Visualization / Interpretation of learned transformations