Using UMAP to Inspect Audio Data for Unsupervised Anomaly Detection under Domain-Shift Conditions

Andres Fernandez, Mark D. Plumbley

Centre for Vision, Speech and Signal Processing – University of Surrey, UK

Introduction

In Unsupervised Anomaly Detection (UAD), only non-anomalous (i.e. normal) data is known beforehand. In UAD under domain-shift conditions (UAD-S), the data is also subject to potentially unknown changes.

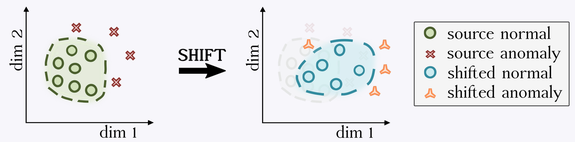


Figure 1: UAD-S illustration. Maintaining the green boundary would result in false positives/negatives.

The 2020 and 2021 DCASE editions presented UAD (2020, task 2 [3]) and UAD-S (2021, task 2 [2]) challenges. Both editions featured a broad variety of approaches, but the 2021 results were significantly lower, independently of the approach. Trends in the literature indicate a higher complexity of the 2021 task. In order to find out possible reasons, we propose to visually inspect the 2021 audio data. Our proposed contributions are:

- Methodology+software to visualize audio data, looking for separability and discriminative support.
- Insights on micro- and macrostructure.
- Formulation of hypotheses to direct future efforts.

Visualizing UAD-S Data

We project the data down to 2D using Uniform Manifold Approximations and Projections (UMAPs)[5]. We assume that if two regions appear separable on the projection, they are also separable in the original domain. This allows us to consider 2 beneficial visual qualities:

- **Separability (SEP):** If there is a simple boundary between normal and anomalous data.
- **Discriminative support (DSUP):** If the training data provides set support for all normal data, and is separable from anomalous data.

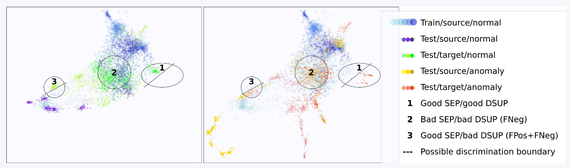
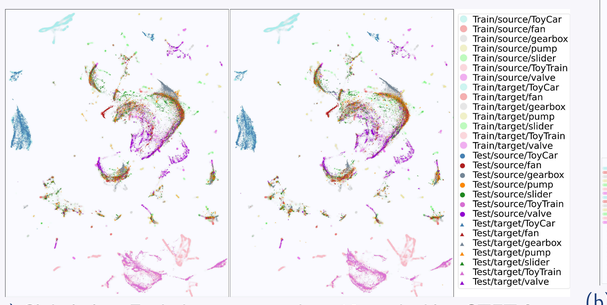


Figure 2: Excerpt from the UMAP plot for pump, illustrating different cases of SEP/DSUP. Test normal data is shown on the left. Test anomalous data on the right. Training data is shown on both sides as underlying shadows.

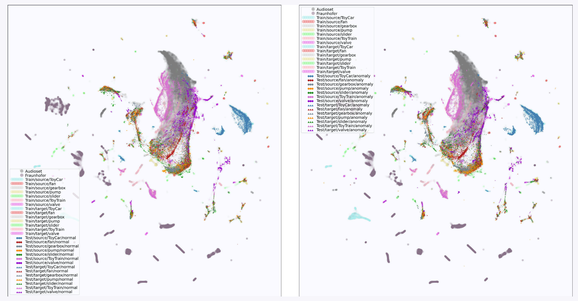
Experiments

The DCASE dataset features sounds from 7 machines during operation. The machines are then intentionally damaged to gather anomalous data. Further domain shifts are introduced (e.g. changes in load, speed), but only ∼0.3% of the shifted data is available before evaluation[6, 4, 7, 1].

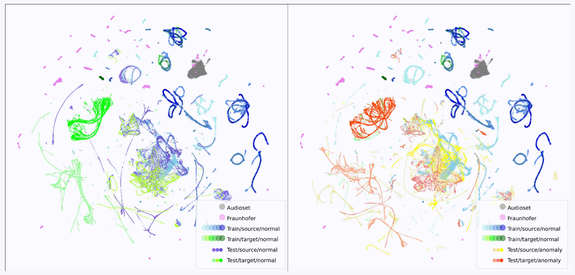
We start with 10s clips from 3 datasets: DCASE, AudioSet and IDMT-ISA-EE. Then we compute 3 representations: 1024-log-STFT spectrograms, 128-log-Mel spectrograms and L3 embeddings. To encode temporal relations, we concatenate consecutive frames (3 lengths: 1, 5 and 10). We then randomly sample the concatenated frames and compute the UMAPs. Finally, we plot the UMAPs with 3 different levels of detail: global, per-device and per-section, resulting in 198 plots in total.



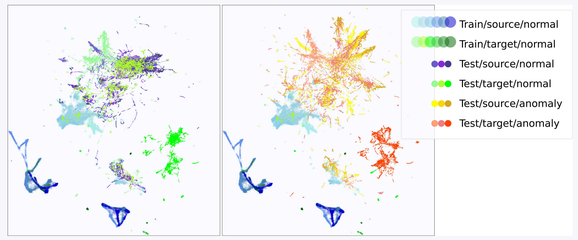
(a) Global plot. Each dot corresponds to 10 stacked log-STFT frames.



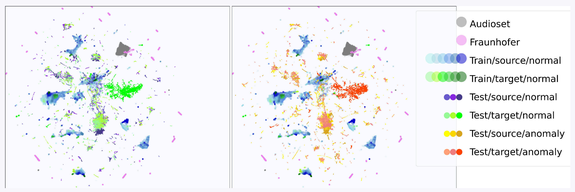
(b) Global plot (with external datasets). Each dot corresponds to 5 stacked log-mel frames.



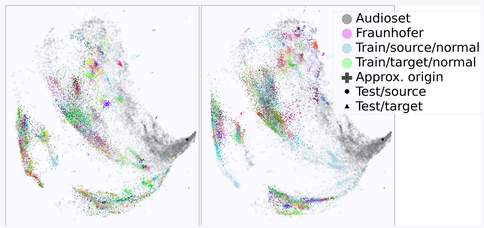
(c) Per-device plot (ToyTrain). Each dot corresponds to 5 stacked L3 frames.



(d) Per-device plot (ToyCar). Each dot corresponds to a single L3 frame.



(e) Per-device plot (fan). Each dot corresponds to a single L3 frame.



(f) Per-section plot (valve, section 0). Each dot corresponds to a single log-mel frame.

Figure 3: Selection of UMAP plots.

- Hard to find anom. patterns with good SEP+DSUP.
- ToyTrain and ToyCar clearly separated.
- ToyAdmos data presents simpler shapes than MIMII.
- UMAPs with/without external data behave differently.
- AudioSet "horn" → tip of the non-negative cone.
- AudioSet appears very concentrated in the L3 plots.
- Figure 3e: lack of DSUP with just one L3 frame.
- Rings in 3c may reflect the ToyTrain circular motion.

Hypotheses

1. **Mixing ToyAdmos2 and MIMII DUE data may hinder performance:** Trivially distinguishable categories may lead to inefficient boundaries for anomaly discrimination.
2. **Temporal context and pretraining regulate a tradeoff between SEP and DSUP:** We observe that longer stack sizes tend to "space out" the data. Similarly, the L3 embeddings tend to compact pre-training data and space out the rest. This generally improves SEP, but can hinder DSUP if training and test data get also separated.
3. **Normalization is a dominating factor for performance:** This was pointed out in several top-performing approaches.
4. **Incorporating domain-related priors may help performance:** While SEP and DSUP are beneficial, only SEP is necessary. The need for training support can be replaced with other priors. The dataset provides such priors, they could be used on scenarios with good SEP.

Discussion

The presented methodology has several shortcomings, which could be tackled via quantitative methods and interactive plots:

- We can only compute UMAPs for a data subset. We likely miss extreme outliers that may be relevant.
- Projections can confirm SEP/DSUP, not discard.
- Qualitative, visual inspection is subject to perceptual biases (e.g. to shape, color and distance).
- Encoding temporal relations by stacking consecutive frames ignores potentially relevant relations.

Still, it helped to expose potential issues in connection with the literature, and can complement well other quantitative forms of analysis. We hope that the software we provide can become a useful tool in the context of UAD-S.

Future work: plotting further representations, extending to supervised scenarios and adding interactivity (e.g. sonification, highlighting).

DCASE2021 WORKSHOP
15-19 November 2021

Acknowledgements: This work was supported by grant EP/T019751/1 from the Engineering and Physical Sciences Research Council (EPSRC) and made use of time on Tier 2 HPC facility JADE2, funded by EPSRC grant EP/T022205/1.

References: [1] Noboru Harada et al. "ToyADMOS2: Another dataset of miniature-machine operating sounds for anomalous sound detection under domain shift conditions". In: arXiv:2106.02369 (2021). [2] Yohei Kawaguchi et al. "Description and Discussion on DCASE 2021 Challenge Task 2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring under Domain Shifted Conditions". In: arXiv:2106.04492, 1–5 (2021). [3] Yuma Koizumi et al. "Description and Discussion on DCASE2020 Challenge Task2: Unsupervised Anomalous Sound Detection for Machine Condition Monitoring". In: DCASE2020 Proceedings. 2020. [4] Yuma Koizumi et al. "ToyADMOS: A Dataset of Miniature-machine Operating Sounds for Anomalous Sound Detection". In: WASPAA 2019. [5] Leland McInnes and John Healy. "UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction". In: arXiv:1802.03426 (2018). [6] Harsh Purohit et al. "MIMII Dataset: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection". In: DCASE 2019 Proceedings. [7] Ryo Tanabe et al. "MIMII DUE: Sound Dataset for Malfunctioning Industrial Machine Investigation and Inspection with Domain Shifts due to Changes in Operational and Environmental Conditions". In: arXiv:2006.05822, 1–4 (2021).

{andres.fernandez, m.plumbley}@surrey.ac.uk