Improved Domain Generalization via Disentangled Multi-task Learning in Unsupervised Anomalous Sound Detection

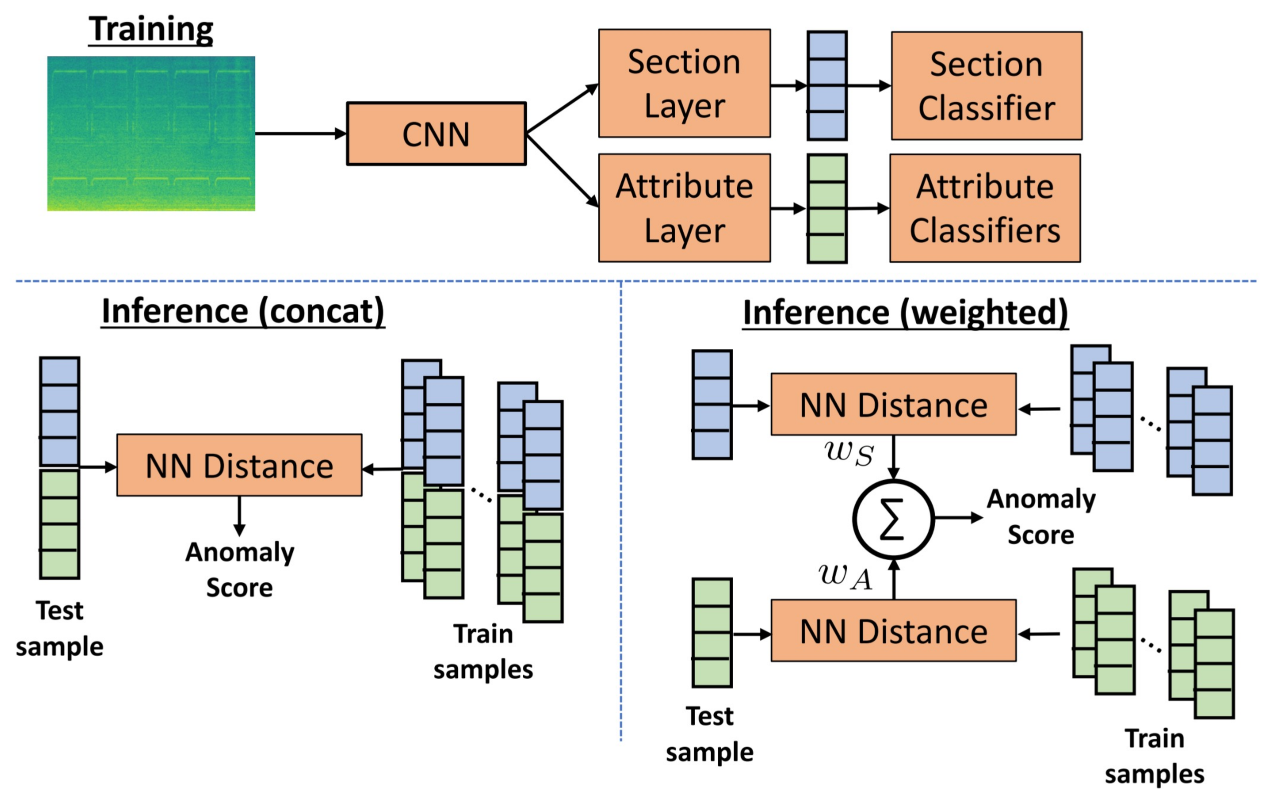
Satvik Venkatesh, Gordon Wichern, and Jonathan Le Roux (MERL, U. Plymouth)

Overview

Unsupervised anomalous sound detection
- Automatic machine condition monitoring
- Algorithms are trained using only normal data

Domain Generalization
- Domain shift between between samples, and the domain is unknown at inference time
- Disentangle domain-shared and domain-specific features

Experiments & Results
- Train on DCASE 2022 Task 2 dataset [1] for seven machine types
- Obtained an overall harmonic mean of 67.57% on the blind evaluation set
- Ranked 5th out of 32 teams in the challenge



Surrogate Task Training

Definitions

Section: main data partition (domain-shared)

Attribute: machine state or condition (domain-specific)

$X \in \mathbb{R}^{F \times T}$: Magnitude Spectrogram

$y = [y_s, y_{a_1}, \ldots, y_{a_M}] \in \mathbb{N}^{M+1}$

y_s: represents machine section

y_{a_m}: represents the categorical label of the m-th attribute

Disentangled Embeddings

$$z_S = \Phi^{\text{Sec}}[\text{CNN}(X)] \in \mathbb{R}^{D_S}$$

$$z_A = \Phi^{\text{Att}}[\text{CNN}(X)] \in \mathbb{R}^{D_A}$$

Cross Entropy Loss

$$\mathcal{L}^{\text{Sec}} = \log \frac{\exp(w_{0,y_s} \cdot z_S + b_{0,y_s})}{\sum_{c=1}^{C} \exp(w_{0,c} \cdot z_S + b_{0,c})}$$

$$\mathcal{L}^{\text{Att}} = \sum_{m=1}^{M} \log \frac{\exp(w_{m,y_m} \cdot z_A + b_{m,y_m})}{\sum_{c_m=1}^{C_m} \exp(w_{m,c_m} \cdot z_A + b_{m,c_m})}$$

$$\mathcal{L} = \mathcal{L}^{\text{Sec}} + \mathcal{L}^{\text{Att}}$$

Experimental Setup

Disentangled Anomaly Detector

- Time average pooling of embeddings across sample
  - Standard-deviation pooling for valve [2]
- Nearest neighbor cosine distance as anomaly score
- Compare different inference approaches:
  - Disentangled Concatenated
  - Disentangled Weighted
  - Disentangled Sections
  - Disentangled Attributes

Other Baselines

- ArcFace
- Entangled multi-task learning (MTL)
- Attentive neural process (ANP-Boot) [3]

Machine Specific Loss (MSL)

- Use the best development set training objective for each machine
- Perform grid search for ensemble weights (MSL+ANP) and disentanglement weights

Table 3. Ensemble Details

		Ens. wt. (S2)		Disent. wt. (S4)	
Machine	**MSL (S1)**	**MSL**	**ANP**	w_S	w_A
ToyCar	Disent_Cat	0.60	0.40	0.90	0.10
ToyTrain	MTL	0.70	0.30	0.00	1.00
Bearing	Sections only	1.00	0.00	1.00	0.00
Fan	ArcFace	0.95	0.05	0.15	0.85
Gearbox	Adversarial	0.65	0.35	0.80	0.20
Slider	Disent_Cat	0.70	0.30	0.90	0.10
Valve	Disent_Split	0.80	0.20	0.90	0.10

Results

Table 1. Performance on development set

System	ToyCar	ToyTrain	Bearing	Fan	Gearbox	Slider	Valve	AUC (S)	AUC (T)	pAUC	Overall
MSL+ANP (S2)	76.43	**59.96**	**73.93**	**68.89**	**85.37**	**85.93**	**95.83**	**87.55**	**73.43**	**70.36**	**76.43**
MSL (S1)	76.43	59.37	73.93	68.85	83.03	85.37	95.63	86.78	73.34	69.68	75.93
Disent_Wt (S4)	**76.95**	59.74	72.07	63.91	81.38	85.14	94.50	86.09	71.65	68.21	74.57
Disent_Cat (S3)	76.43	58.67	67.09	63.18	80.99	85.37	95.01	84.61	70.59	67.02	73.34
Disent_Sec	76.84	56.64	72.07	62.35	81.04	84.84	94.42	86.43	68.64	68.01	73.45
Disent_Att	75.26	59.74	60.82	63.02	78.86	78.88	92.72	82.12	69.31	64.09	71.08
MTL	75.61	59.37	68.24	59.14	80.63	83.51	94.49	81.35	70.72	66.83	72.47
Sec_ArcFace	72.31	58.09	71.30	68.85	79.37	82.50	92.87	86.50	71.19	66.04	73.62
Sec_Softmax	76.20	52.85	**73.93**	64.39	81.43	85.89	90.11	86.05	67.34	67.92	72.82
ANP-Boot	59.84	50.87	55.54	55.31	64.38	64.11	52.63	69.26	50.87	54.24	57.10
AE Baseline	51.06	39.61	54.80	58.54	63.07	57.99	50.59	68.74	41.91	53.76	52.62
MN Baseline	54.23	51.18	59.16	57.21	59.91	50.26	62.42	63.87	50.14	55.69	56.01

- Performance metrics
  - AUC (Area under the ROC curve)
  - pAUC (partial AUC at low-false alarm rates)
  - Source (S) and Target (T) domains
- Disentanglement outperforms multi-task learning
- ArcFace performs well on sections, but didn't work with sparse attribute labels
- Attribute learning helps for all machines except fan and bearing

Table 2. Performance on evaluation set

System	ToyCar		ToyTrain		Fan		Gearbox		Bearing		Slider		Valve		Overall
	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	AUC	pAUC	
Top rank [28]	88.45	**81.83**	**70.46**	**61.14**	**57.34**	**57.33**	86.04	64.22	68.85	54.45	**78.26**	**66.39**	**83.87**	**75.22**	**70.97**
Disent_Cat (S3)	**93.88**	78.67	58.23	54.73	48.17	50.34	**86.76**	**79.43**	**72.54**	61.86	73.64	60.70	83.72	62.93	67.57
Disent_Wt (S4)	93.30	75.47	57.30	54.93	46.93	50.33	86.34	78.47	71.96	**64.26**	75.94	64.29	83.05	64.01	67.49
MSL (S1)	**93.88**	78.67	55.53	54.33	44.50	50.84	86.47	68.54	69.94	61.64	73.64	60.70	78.51	66.08	65.66
MSL+ANP (S2)	**93.88**	78.67	54.92	54.22	44.29	50.97	82.37	70.76	69.94	61.64	75.96	62.40	77.69	65.39	65.57
MN Baseline	42.79	53.44	51.22	50.98	50.34	55.22	51.34	48.49	58.23	52.16	62.42	53.07	72.77	65.16	54.02
AE Baseline	61.18	60.21	43.14	49.36	41.16	50.12	61.92	51.95	59.93	53.95	58.95	54.16	54.26	51.30	52.94

- Surpassed the best baseline by 13.5% and trailed the top ranking system by 3.4%.
- Outperformed the top rank for bearing, gearbox, and ToyCar (AUC)
- Ensemble weights overfit to the dev set
- Disentangled concatenated inference (S3) was our best performing system

Future Work

- Explore other anomaly detection backends in addition to NN, e.g., GMM
- Improve training pipeline: pre-train on all seven machines and use average model weights from multiple epochs at inference time
- Use disentangled embeddings to *explain* anomaly decisions

References

[1] Dohi, et al. "Description and discussion on DCASE 2022 challenge task 2: Unsupervised anomalous sound detection for machine condition monitoring applying domain generalization techniques," arXiv 2022.

[2] Morita, et al, "Anomalous sound detection using cnn-based features by self supervised learning," DCASE Challenge Tech Report 2021.

[3] Wichern et al, "Anomalous sound detection using attentive neural processes," in Proc. WASPAA 2021.