Deep Unsupervised Representation Learning for Abnormal Heart Sound Classification

Shahin Amiriparian[1,2], Maximilian Schmitt[1], Nicholas Cummins[1],
Kun Qian[1,2], Fengquan Dong[3], Björn Schuller[1,4]

1. Chair of Embedded Intelligence for Health Care and Wellbeing, University of Augsburg, Germany
2. Machine Intelligence & Signal Processing Group, Technische Universität München, Germany
3. Shenzhen University General Hospital, Shenzhen, P. R. China.
4. Group on Language, Audio, and Music, Imperial College London, UK

shahin.amiriparian@tum.de

Problem Description

- A myriad of acoustic sounds
  - Reflecting our **physiological** and **pathological states**
- Classification of **abnormal heart sounds**
- Feature **engineering** vs. deep representation **learning**

Research Aims of Paper

- **Extraction** of **expert-designed** features
- **Quantisation** of **expert-designed** features
- **Learning** task-dependent **deep representation**

Heart Sound Dataset

- Heart Sounds Shenzhen (HSS) corpus
- 845 recordings (30 seconds on average)
- Total length: 7 hours

Partition	normal	mild	moderate/severe	SUM
Train.	84	276	142	502
Devel.	32	98	50	180
Test	-	-	-	163

Heart Sound Dataset

- Recording equipment
  - electronic stethoscope
- Recording from one of the:
  - auscultatory mitral area
  - aortic valve auscultation area
  - pulmonary valve auscultation area, and
  - auscultatory area of the tricuspid valve.
- **170** independent subjects (**55** f and **115** m)
  - Mean age: **65.4** years
  - Standard deviation: **13.2** years

Feature Engineering

- ComParE feature set (**6373** dimensional):
  - Prosodic
  - Spectral
  - Cepstral, and
  - Voice quality low-level descriptors (LLDs)

Feature Engineering

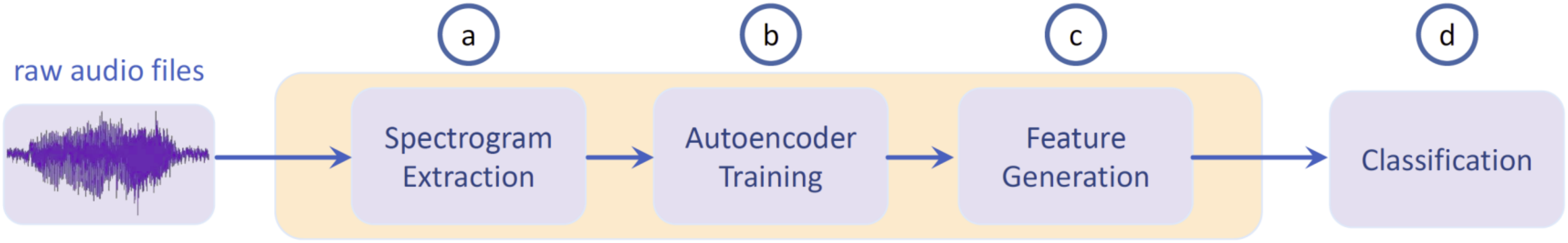
- Bag-of-Audio-Words:
  - Quantisation of ComParE features
  - openXBOW [1]
  - Forming **sparse** fixed-length **histogram** representation of an audio clip

---

- [1] Maximilian Schmitt and Björn Schuller: "openXBOW - Introducing the Passau Open-Source Crossmodal Bag-of-Words Toolkit", The Journal of Machine Learning Research, Volume 18, No. 96, pp. 1-5, October 2017.

Deep Representation Learning

- Using auDeep [2]:
  - Deep representation learning from raw audio

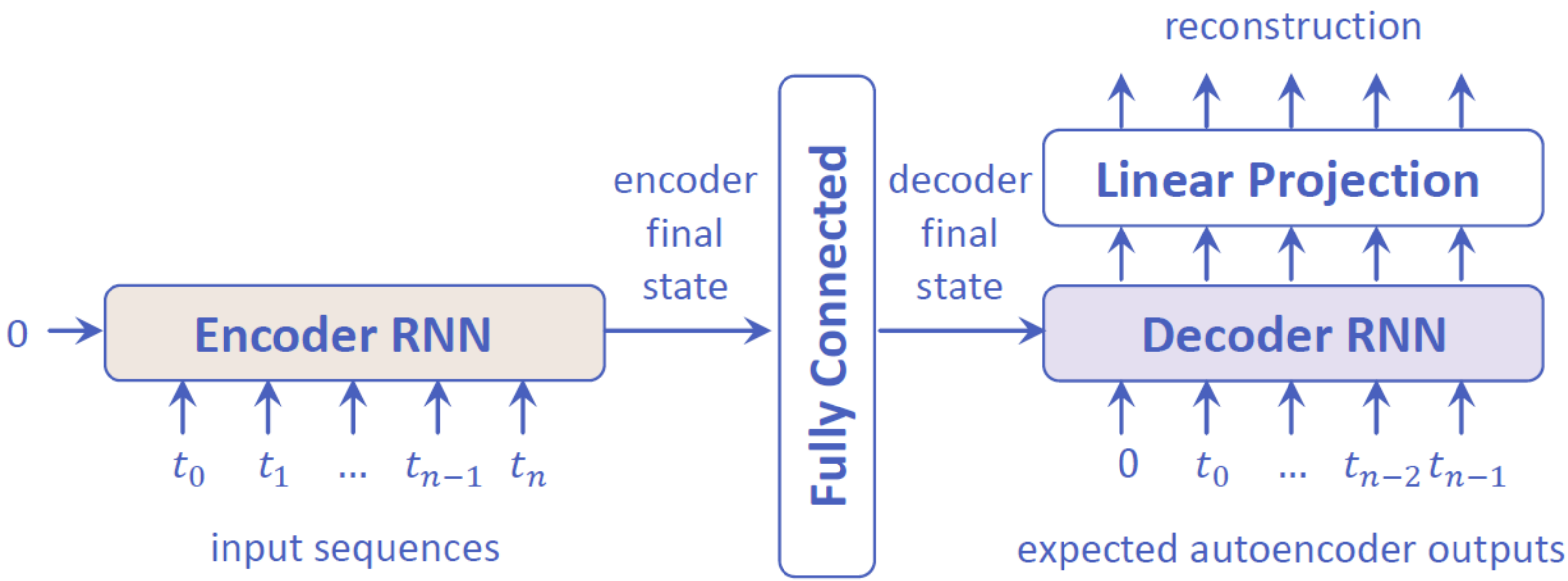


- [2] https://github.com/auDeep/auDeep
- [2] M. Freitag, S. Amiriparian, S. Pugachevskiy, N. Cummins, and B. Schuller,"audeep: Unsupervised learning of representations from audio with deep recurrentneural networks,"Journal of Machine Learning Research, vol. 18, no. 173, pp. 1–5,2018.

Spectrogram Extraction

- Hann windows with width $\boldsymbol{w}$ and overlap $\mathbf{0.5}\boldsymbol{w}$
- Computing $\boldsymbol{N}_{\{\boldsymbol{mel}\}}$ of log-scaled Mel frequency bands
- Normalising the Mel-spectra $[-\mathbf{1}, \mathbf{1}]$
- Amplitude clipping $\{-\mathbf{30}, -\mathbf{45}, -\mathbf{60}, -\mathbf{75}\}\boldsymbol{dB}$

Recurrent Sequence to Sequence Autoencoders



Autoencoder Hyperparameters

- Tested
  - $\boldsymbol{N_{\{layer\}}} \in \{\mathbf{2,3,4}\}$
  - $\boldsymbol{N_{\{unit\}}} \in \{\mathbf{64,128,256,512}\}$
  - All combinations of **unidirectional** and **bidirectional** encoder and decoder RNN

- Best configuration
  - $\boldsymbol{N_{\{layer\}}} = \mathbf{2}$
  - $\boldsymbol{N_{\{unit\}}} = \mathbf{256}$
  - Unidirectional encoder
  - Bidirectional decoder

Results

Engineered Features

System	Dimensionality	UAR [%] C	Devel.	Test
ComParE	6 373	10^{-6}	41.1	44.8
		10^{-5}	44.5	45.6
		10^{-4}	50.3	46.4
		10^{-3}	44.5	40.4
		10^{-2}	43.2	41.7
BoAW	250	10^{-3}	43.1	43.4
	500	10^{-3}	42.3	47.2
	1000	10^{-2}	43.7	41.0

Learnt Deep Representations

System	Dimensionality	UAR [%] C	Devel.	Test
auDeep: Individual Feature Sets				
-30 dB	1 024	$2 \cdot 10^{-2}$	32.8	40.0
-45 dB	1 024	$5 \cdot 10^{-4}$	38.4	40.6
-60 dB	1 024	$6 \cdot 10^{-2}$	39.6	45.2
-75 dB	1 024	$8 \cdot 10^{-3}$	36.9	41.7
Fused	4 096	$4 \cdot 10^{-3}$	35.2	**47.9**

Conclusions

- Promising results with **sequence to sequence autoencoders**
- Effective alternative to **expert-designed** feature sets
- Fully **unsupervised** autoencoder training
- **Variable-length** input

Future Research

- Applying **data augmentation** techniques
- Comparison and/or fusion with Deep Convolutional Generative **Adversarial Networks**
- Feature selection and **dimensionality reduction**