Pattern Recognition Lab

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG
FACULTY OF ENGINEERING

Deep Representation Learning for Orca Call Type Classification

Christian Bergler, Manuel Schmitt, Rachael Xi Cheng, Hendrik Schröter, Andreas Maier, Volker Barth, Michael Weber, and Elmar Nöth
Pattern Recognition Lab, Friedrich-Alexander University Erlangen-Nuremberg
Text, Speech, and Dialogue (TSD), 22nd International Conference
September 11th – 13th 2019, Ljubljana, Slovenia

Introduction: Killer Whale Research

The Killer Whale

- Largest member of the dolphin family
- Distinct communication system
- Vocalization not only for mating or alarm calls also for orientation and hunting
- Orcas have well marked social behavior and have highly social interactions
- Complex social, communicative, and cognitive capacities

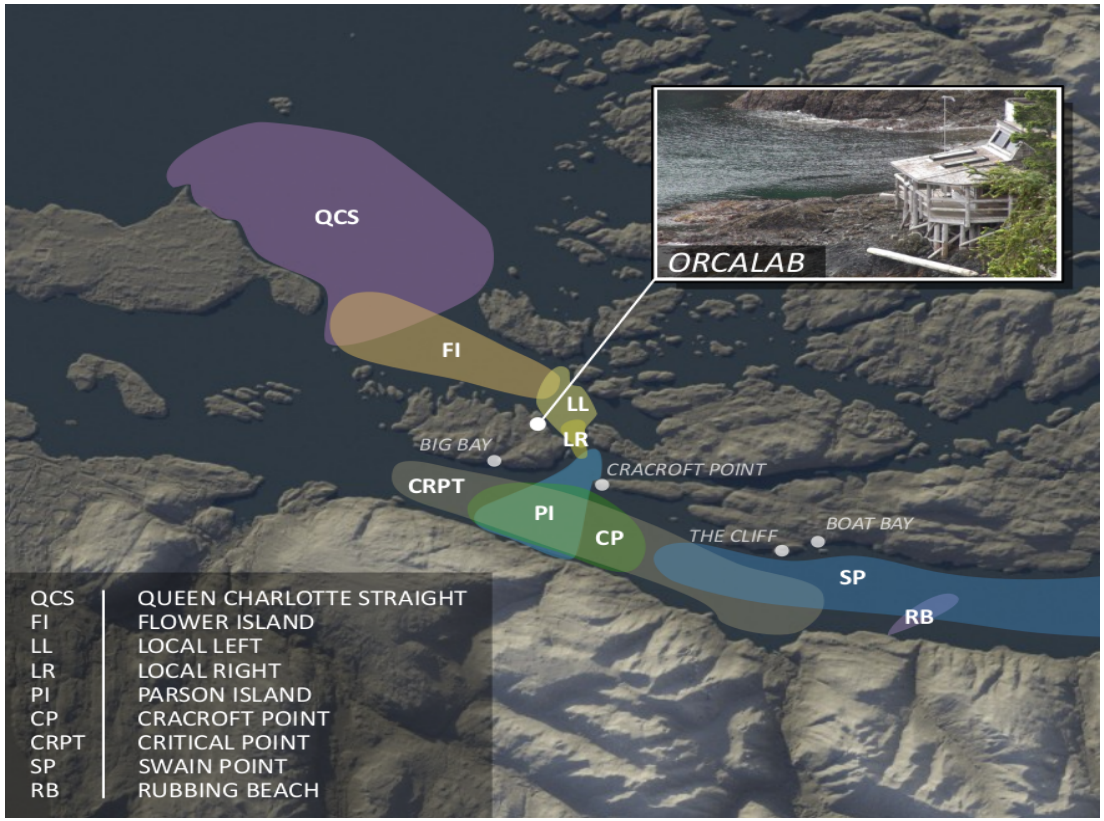
©Volker Barth, DeepAL

The Killer Whale during fieldwork expedition 2017/2018 in northern Vancouver Island, British Columbia, Canada

Introduction: Killer Whale Research

Existing Bioacoustic Archive – The Orchive

- collected by the Orcalab [2] and Steven Ness [4]
- 20,000 hours of underwater recordings by using 6 stationary hydrophones (1985–2010)
- 23,511 digitized audio tapes each ∼45 min.
- Orchive Annotation Catalog (OAC) [4] comprises 15,480 orca/noise labels



The Orcalab on Hanson Island (northern Vancouver Island, British Columbia, Canada) and its recording environment

Introduction: Killer Whale Research

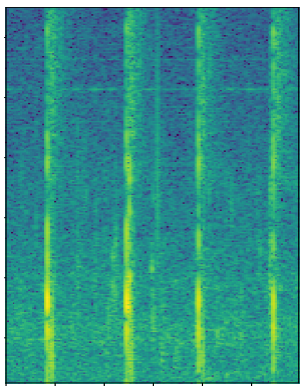
DeepAL Fieldwork Expedition - The DeepAL Fieldwork Data (DLFD) 2017/2018

- collected via a 15-meter research trimaran
- 1,007 hours of multi-channel underwater recordings
- 89 hours video footage about behavioral data
- Interdisciplinary team consisting of marine biologists, computer scientists, and psychologists

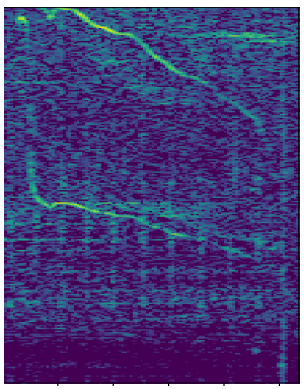
DeepAL 2017/2018 Expedition Route (British Columbia) [1]

Introduction: Killer Whale Research

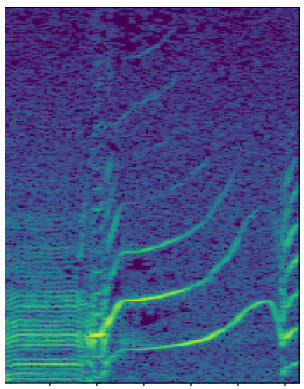
Killer whale sound type segmentation



Echolocation Click



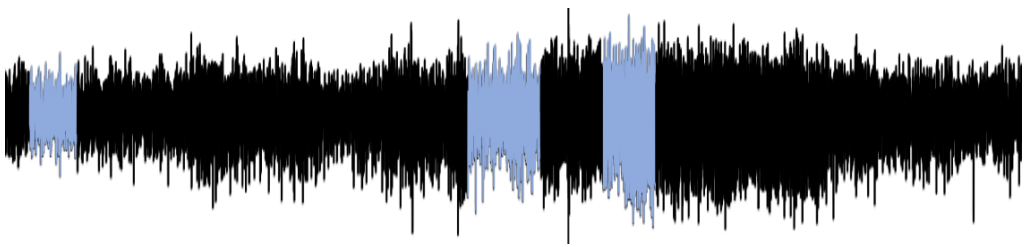
Whistle



Pulsed Call

Spectrograms from three characteristic killer whale sounds (see [3])

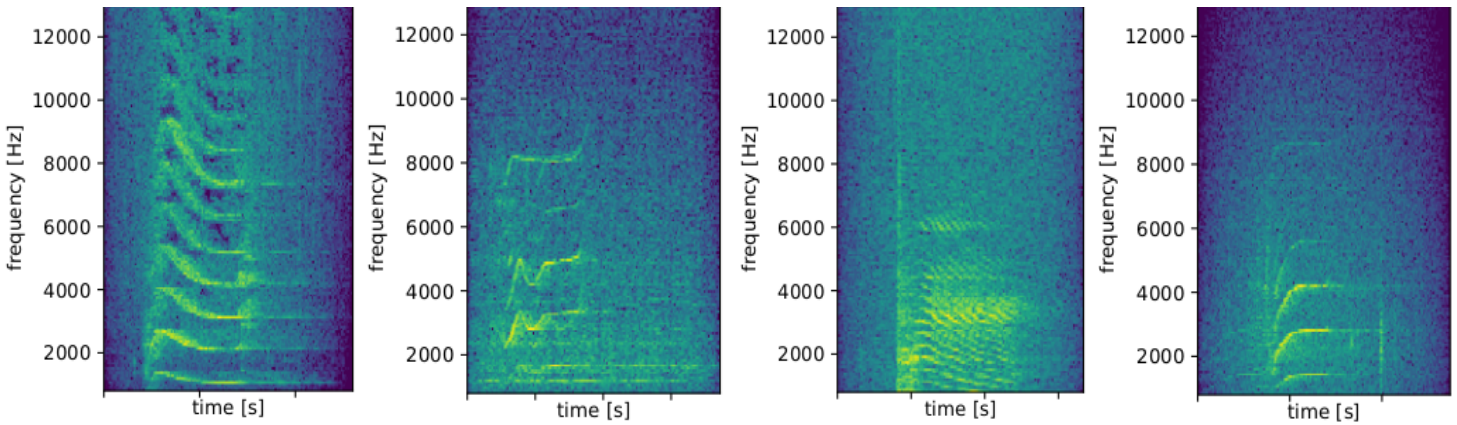
- Goal: Robust and accurate segmentation of orca sounds within noise-heavy underwater recordings (ORCA-SPOT [3])



Introduction: Killer Whale Research

Discrete Pulsed Calls (Call Types)

- Pulsed calls are classified into discrete (call types), variable, and aberrant calls [5]



a) A4 N04 **b)** A5 N02 **c)** A12 N03 **d)** A36 N09

Various pod-specific discrete call types

- Northern residents (killer whale population in northern British Columbia) vocal repertoire of discrete calls consists of more than 40 different types [8, 9]
- Huge inter- but also intra-pod signal variations even within one single human-labeled call type

Outline

Data Corpora and Preprocessing

Deep Representation Learning – Network Architecture, Training, and Results

Call Type Classification – Network Architecture, Training, and Results

Conclusion

Pattern Recognition Lab

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

FACULTY OF ENGINEERING

Data Corpora and Preprocessing

Data Corpora – Deep Representation Learning

Representation learning datasets

Split/ Datasets		train		val		test	
		smp	%	smp	%	smp	%
OAC[1]	**7,903**	5,832	73.8	1,171	14.8	900	11.4
AEOTD[2]	**1,667**	1,172	70.3	260	15.6	235	14.1
DLFD[3]	**3,331**	1,384	41.5	1,171	35.2	776	23.3
OSD[4]	**19,211**	13,493	70.2	2,863	14.9	2,855	14.8
SUM	**32,112**	21,881	68.1	5,465	17.0	4,766	14.8

[1] Orchive Annotation Catalog (OAC) [2]
[2] Automatic Extracted Orchive tape data (AEOTD) [3, 4]
[3] DeepAL Fieldwork Data (DLFD) [1]
[4] Orca Segmented Data (OSD) [3]

Training, validation, and test distribution for deep representation learning

Data Corpora – Call type classification

Call type datasets

Split/ Datasets		train		val		test	
		smp	%	smp	%	smp	%
CCS[1]	**138**	102	73.9	19	13.8	17	12.3
CCN[2]	**286**	198	69.2	41	14.4	47	16.4
EXT[3]	**90**	63	70.0	12	13.3	15	16.7
SUM	**514**	363	70.6	72	14.0	79	15.4

[1] Call Catalog Symonds (CCS) [2]
[2] Call Catalog Ness (CCS) [4]
[3] Orchive Extension Catalog (EXT)

Training, validation, and test distribution for call type classification

Data Preprocessing

Preprocessing and Augmentation

- Power-Spectrogram
- Augmentation
  - Amplitude scaling
  - Frequency shift
  - Time stretch
- Linear frequency compression (256 bins)
- Addition of noise spectrograms (only for call type classification)
- dB-Normalization
- Trimming / Padding to fixed length (1.28 s)

Pattern Recognition Lab

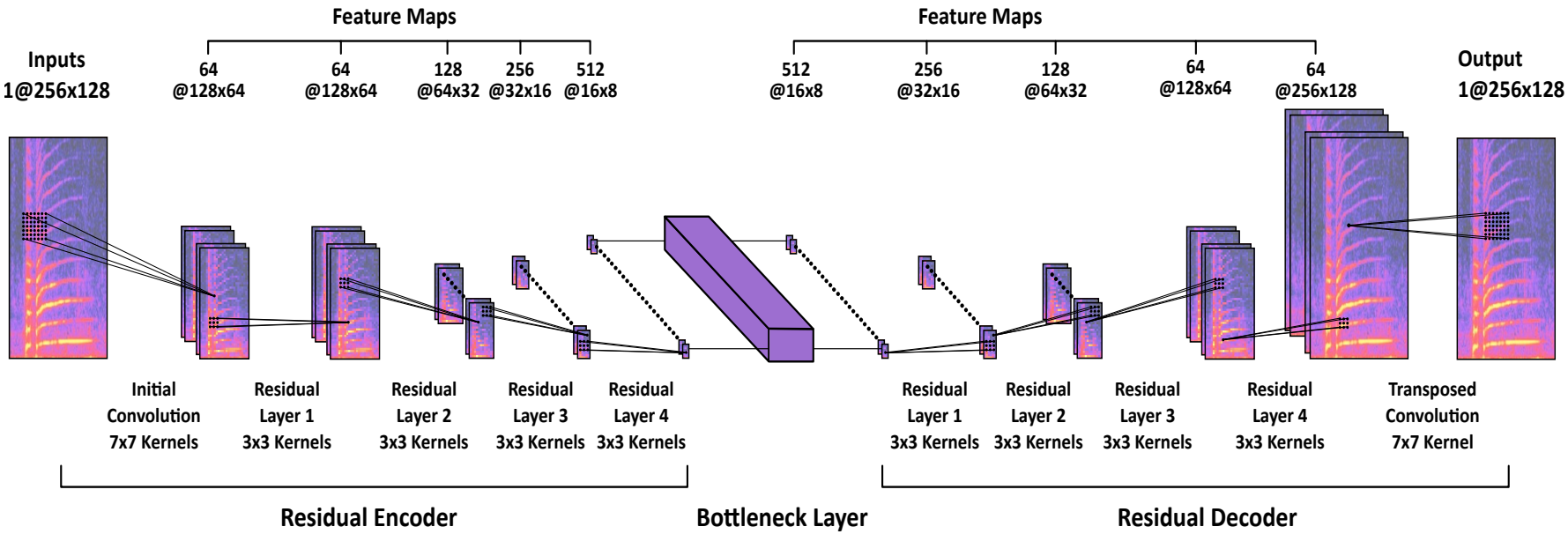
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

FACULTY OF ENGINEERING

Deep Representation Learning – Network Architecture, Training, and Results

Network Architecture and Training

Architecture



ResNet18-based undercomplete convol. autoencoder for deep representation learning of killer whale signals

- **Covolutional bottleneck layer:** Convolutional layer (1×1 kernel, no stride) compressing $512 \times 16 \times 8$ to $4 \times 16 \times 8$ (512 features) and back
- **Linear bottleneck layer:** Max-pooling $512 \times 16 \times 8$ to 512 features, fully-connected latent layer (512-D), max-unpooling $512 \times 16 \times 8$

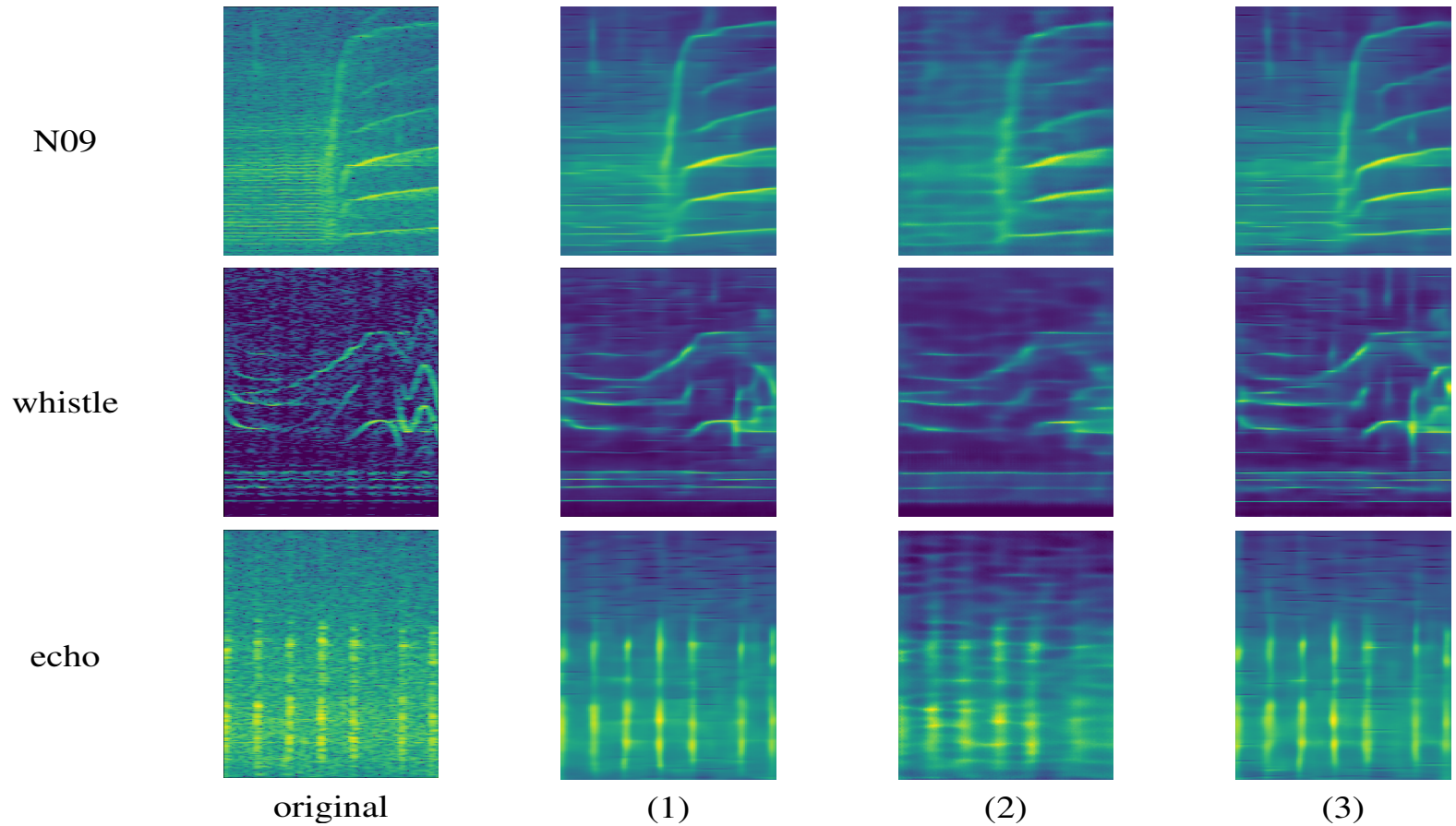
Network Architecture and Training

Training

- Implemented in PyTorch [21]
- Adam optimizer together ($\alpha = 10^{-5}$, $\beta_1 = 0.5$, and $\beta_2 = 0.999$)
- α was decayed by 1/2 after 4 epochs, training stopped after 10 epochs without having any improvements on the validation set
- Batch size of 32 together with a **mean squared error (MSE) loss**
- Lowest validation loss was selected as criterion for the best autoencoder model
- Various data combinations: **semi-automatic labeled data** (entire representation corpora), **fully hand-labeled data** (only OAC corpus), and **automatic labeled data** (only OSD dataset)

Network Results

Reconstruction results



Top 3 autoencoder (all using bottleneck layer: 1×1 convolution) reconstructions of killer whale sound types from the call type test set; **(1)** semi-automatic labeled data, **(2)** hand-labeled data **(3)** automatic machine-labeled data

Pattern Recognition Lab

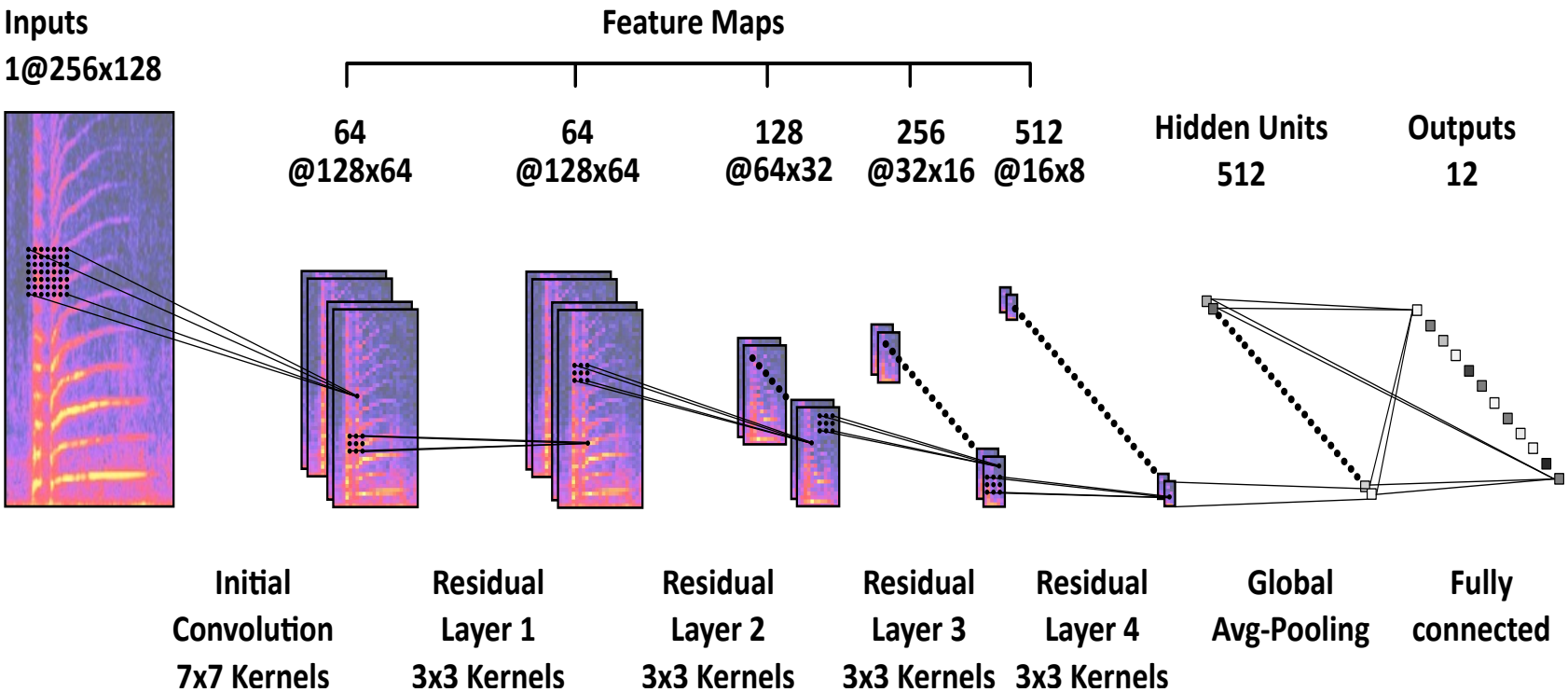
FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

FACULTY OF ENGINEERING

Call Type Classification – Network Architecture, Training, and Results

Network Architecture and Training

Architecture



ResNet18-based Convolutional Neural Network (CNN) without max-pooling in the first residual layer for a 12-class problem [10]

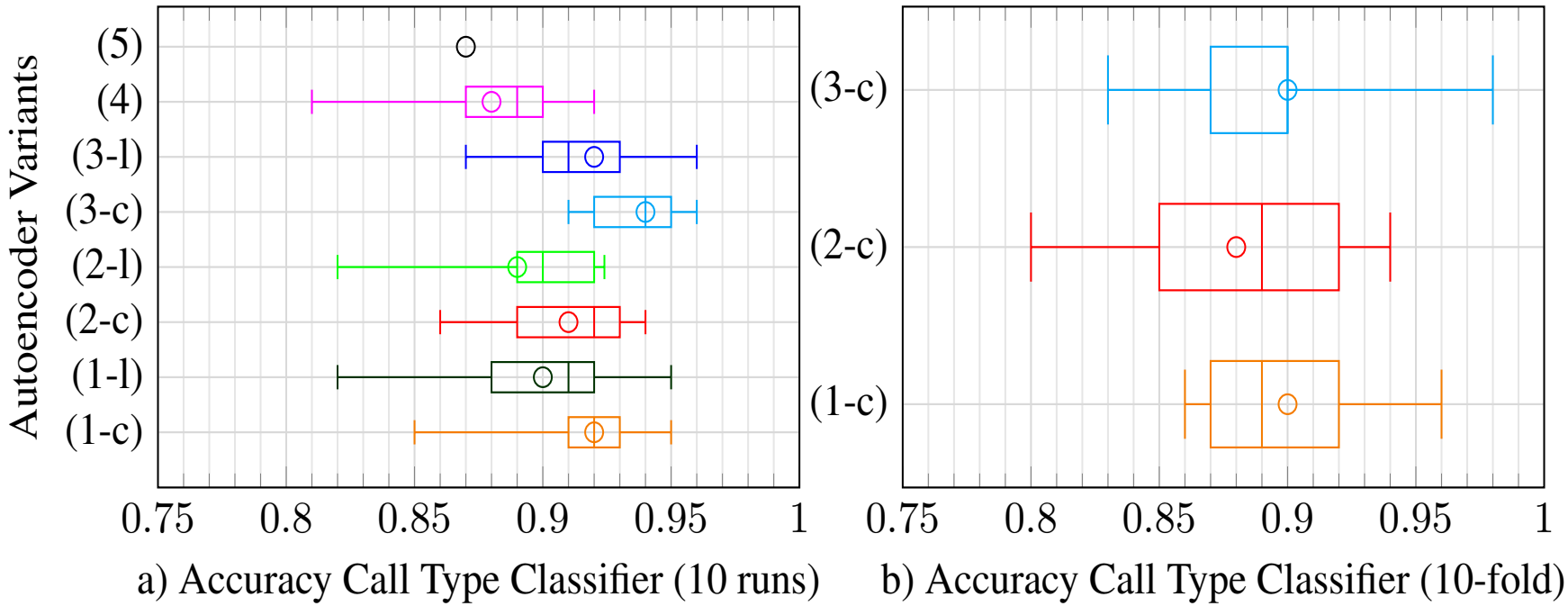
Network Architecture and Training

Training

- Implemented in PyTorch [21]
- Adam optimizer together ($\alpha = 10^{-5}$, $\beta_1 = 0.5$, and $\beta_2 = 0.999$)
- α was decayed by 1/2 after 4 epochs, training stopped after 10 epochs without having any improvements on the validation set
- Batch size of 4 together with a **12-class cross entropy loss**
- Highest validation accuracy was selected as criterion for the best classification model
- **Pretrained autoencoder encoder parts** (6 different variants) were separately **used for weight initialization** of the call type classifier (except last residual layer)

Network Results

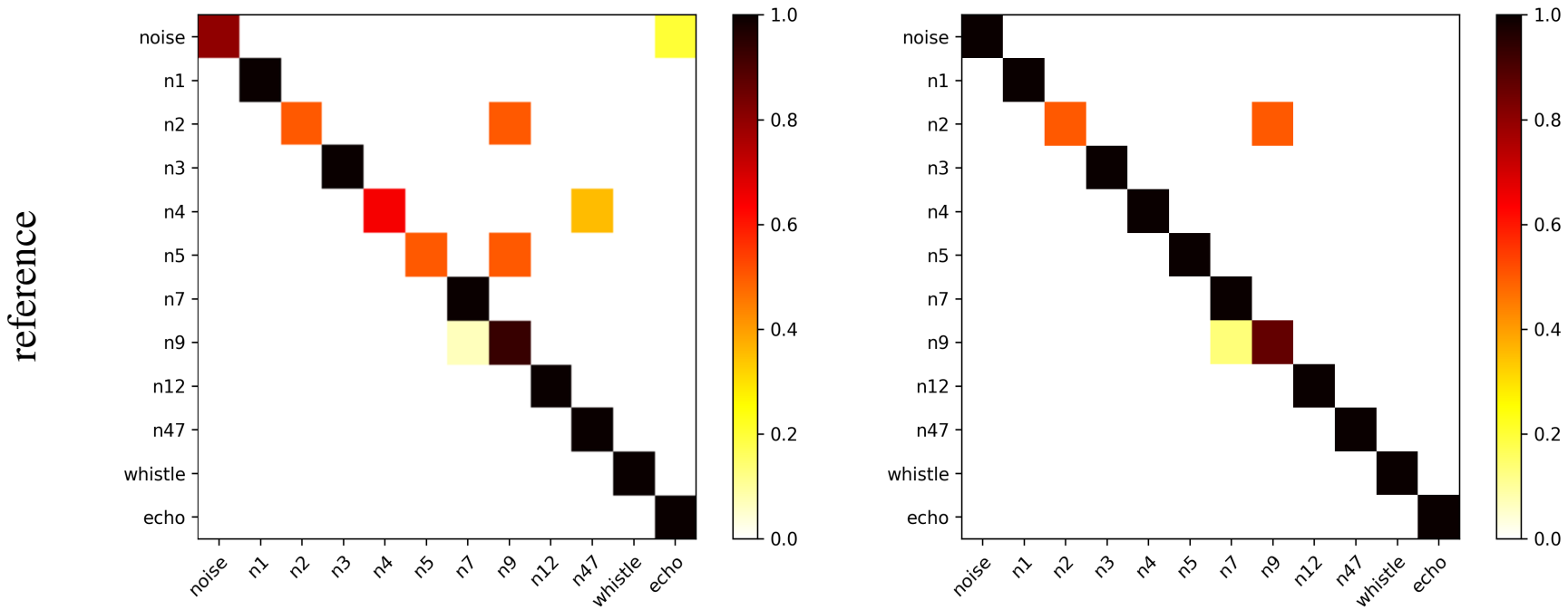
Pretrained call type classification results



a) Mean test accuracy of 10 train/evaluation runs:
(1-c)/(1-l) conv./linear AE on the entire representation learning data, (2-c)/(2-l) conv./linear AE on the OAC corpus, (3-c)/(3-l) conv./linear AE on the OSD dataset, (4) no pretrain (5) mean test accuracy of [10]
b) Classifier accuracy in a 10-fold cross validation for the top 3 AEs (1-c), (2-c), (3-c)

Network Results

Non-pretrained vs. pretrained call type classification results



hypothesis – non-pretrained classifier

hypothesis – best pretrained classifier

- 87 % test accuracy (without pretraining) [10] vs. 96 % test accuracy (best pretrained model)

Conclusion

Conclusion

- Deep representation learning (any form) has a significant positive influence on killer whale call type classification
- Pretraining on our fully automatic machine-labeled OSD corpus led to the best performance (robust and reliable segmentation process [3])
- Fully unsupervised methods (feature learning and clustering on fully automatic segmented orca data) to machine-identify finer and potential undiscovered call types by segmenting/clustering the entire Orchive (20,000 h of underwater recordings)
- Using various cluster outputs for a more robust supervised call type classification by removing human-perception
- Deriving sequential ordered call type structures (syntactic patterns)

Thank you for your attention.

Questions?

©Volker Barth, DeepAL

References I

[1] C. Bergler, *Deepal fieldwork data 2017/2018 (dlfd)*, https://www5.cs.fau.de/research/data/ (September 2019).

[2] ORCALAB, *Orcalab - a whale research station on hanson island*, http://orcalab.org (September 2019).

[3] C. Bergler, H. Schröter, R. Xi Cheng, V. Barth, M. Weber, E. Noeth, H. Hofer, and A. Maier, "Orca-spot: an automatic killer whale sound detection toolkit using deep learning", Scientific Reports **9** (2019) 10.1038/s41598-019-47335-w.

[4] S. Ness, "The orchive : a system for semi-automatic annotation and analysis of a large collection of bioacoustic recordings", PhD thesis (Department of Computer Science, University of Victoria, 3800 Finnerty Road, Victoria, British Columbia, Canada, V8P 5C2, 2013), p. 228.

[5] J. K. B. Ford, "Acoustic behaviour of resident killer whales (Orcinus orca) off Vancouver Island, British Columbia", Canadian Journal of Zoology **67**, 727–745 (1989).

References II

[6] M. A. Bigg, P. F. Olesiuk, G. M. Ellis, J. K. B. Ford, and K. C. Balcomb, "Organization and genealogy of resident killer whales (Orcinus orca) in the coastal waters of British Columbia and Washington State", International Whaling Commission, 383–405 (1990).

[7] T. Ivkovich, O. Filatova, A. Burdin, H. Sato, and E. Hoyt, "The social organization of resident-âăřtype killer whales (Orcinus orca) in Avacha Gulf, Northwest Pacific, as revealed through association patterns and acoustic similarity", Mammalian Biology **75**, 198âĂŘ210 (2010).

[8] J. K. B. Ford, "A catalogue of underwater calls produced by killer whales (Orcinus orca) in British Columbia", Canadian Data Report of Fisheries and Aquatic Science, 165 (1987).

[9] J. K. B. Ford, "Vocal traditions among resident killer whales (Orcinus orca) in coastal waters of British Columbia", Canadian Journal of Zoology **69**, 1454–1483 (1991).

References III

[10] H. Schröter, E. Nöth, A. Maier, R. Cheng, V. Barth, and C. Bergler, "Segmentation, classification, and visualization of orca calls using deep learning", in International conference on acoustics, speech, and signal processing, proceedings (ICASSP) (2019).

Data Distribution

Call Type Label Distribution

Orca Call Type/ Corpus	N01	N02	N03	N04	N05	N07	N09	N12	N47	echo	whistles	noise	SUM
CCS [2]	33	10	—	21	14	18	26	16	—	—	—	—	**138**
CCN [4]	36	—	56	60	—	31	70	—	33	—	—	—	**286**
EXT	—	—	—	—	—	—	—	—	—	30	30	30	**90**
SUM	**69**	**10**	**56**	**81**	**14**	**49**	**96**	**16**	**33**	**30**	**30**	**30**	**514**

Orca call type, echolocation, whistle, and noise label distribution of the CCS, CCN, and EXT data corpus

Data Preprocessing

Preprocessing and Augmentation

Data: Training Input Audio $\mathcal{A}_{inp}$

Result: Trainable Spectrogram $\mathcal{S}_{train}$

1 $\mathcal{S}_{inp} \leftarrow 10 \cdot log_{10}\left(\left|\mathcal{FFT}\left(resamp\left(mono\left(\mathcal{A}_{inp}\right), 44.1\,\text{kHz}\right), \text{ffts} = 4096, \text{hop} = 441\right)\right|^2\right)$

2 $\mathcal{S}_{train} \leftarrow scaleAmplitude\left(\mathcal{S}_{inp}, \alpha_{dB} = sample\left(\left[-6\,\text{dB}, 3\,\text{dB}\right]\right)\right)$

3 $\mathcal{S}_{train} \leftarrow shiftPitch\left(\mathcal{S}_{train}, \alpha = sample\left(\left[0.5, 1.5\right]\right)\right)$

4 $\mathcal{S}_{train} \leftarrow stretchTime\left(\mathcal{S}_{train}, \alpha = sample\left(\left[0.5, 2\right]\right)\right)$

5 $\mathcal{S}_{train} \leftarrow compressFrequencies\left(\mathcal{S}_{train}, \text{f}_{min} = 500\text{Hz}, \text{f}_{max} = 10\,000\,\text{Hz}, \text{bins} = 256\right)$

6 $\mathcal{S}_{train} \leftarrow addNoise\left(\mathcal{S}_{train}, sample\left(\mathcal{S}_{noise}\right), \text{SNR} = sample\left(\left[12\,\text{dB}, -3\,\text{dB}\right]\right)\right)$

7 $\mathcal{S}_{train} \leftarrow normalize\left(\mathcal{S}_{train}, \text{dB}_{min} = -100\,\text{dB}, \text{dB}_{ref} = 20\,\text{dB}\right)$

8 $\mathcal{S}_{train} \leftarrow trimPad\left(\mathcal{S}_{train}, \text{length} = sample\left(128\right)\right)$

9 **return** $\mathcal{S}_{train}$