Pattern Recognition Lab

FRIEDRICH-ALEXANDER UNIVERSITÄT ERLANGEN-NÜRNBERG

FACULTY OF ENGINEERING

Deep Learning for Orca Call Type Identification – A Fully Unsupervised Approach

Christian Bergler[1], Manuel Schmitt[1], Rachael Xi Cheng[2], Andreas Maier[1], Volker Barth[3], and Elmar Nöth[1]

[1] *Friedrich-Alexander University Erlangen-Nuremberg, Department of Computer Science – Pattern Recognition Lab, Martensstr. 3, 91058 Erlangen, Germany*

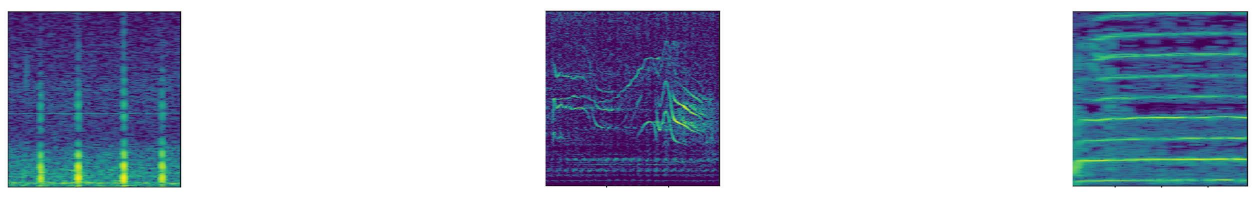
[2] *Leibniz Institute for Zoo and Wildlife Research (IZW) in the Forschungsverbund Berlin e. V., Alfred-Kowalke-Straße 17, 10315 Berlin, Germany*

[3] *Anthro-Media, Nansenstr. 19, 12047 Berlin, Germany*

Introduction – Killer Whale *(Orcinus Orca)* Communication

©*Volker Barth, DeepAL*

- **Largest member** of the **dolphin family** with complex and well-studied vocal structures [1] **producing three different sound types** [2]:



a) Echolocation Clicks **b)** Whistle **c)** Pulsed Call

- Pulsed calls are besides whistles and echolocation clicks the most common type of killer whale vocalization (discrete, variable, aberrant)
- **Pulsed calls (call types)** have sudden and patterned shifts in frequency with a pulse repetition rate between 250 and 2,000 Hz [2]
- Various pods (socializing matrilines) have **distinct vocal repertoires** (mixture of unique and shared discrete call types) $\rightarrow$ **Dialects**

Motivation – Fully Unsupervised Call Type Identification

- Current understanding of killer whale vocalizations refer to the **human classified killer whale sound type catalog** by Ford in **1987** [3]

a) A5 N09 **b)** A12 N09 **c)** A24 N09 **d)** A36 N09

- Huge **inter- but also intra-pod signal variations** even within one single human-labeled call type
- Fully unsupervised multi-step machine- and data-driven approach to address issues such as: (1) labor-intensive and **missing data annotation**, (2) **human** perception-based **classification**, (3) **human error-proneness**, (4) analysis of **large (bioacoustic) audio archives**

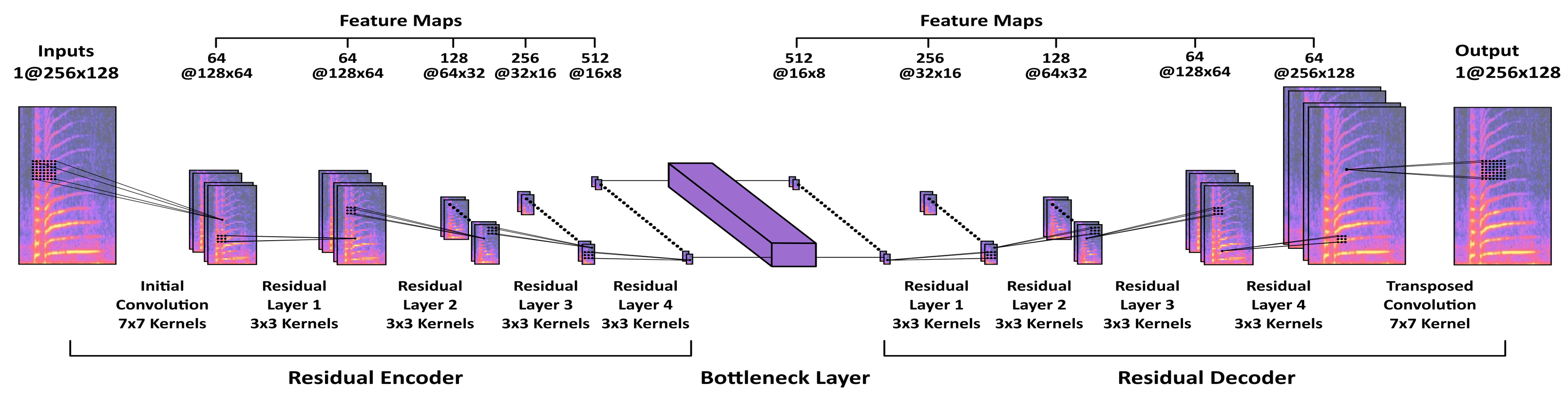
Methodology – Approach, Data Material, Network Models

- **Approach:** (1) Unsupervised killer whale **feature learning** using a **convolutional undercomplete ResNet18-autoencoder** trained on machine-annotated orca data, and (2) **Spectral clustering** of killer whale signals utilizing compressed **orca feature representations**
- **Orca Segmented Data (OSD):** 19,211 samples (100.0 %), Training: 13,443 samples (70.0 %), Validation: 2,902 samples (15.1 %), Test: 2,866 (14.9 %) (ResNet18-based orca/noise segmenter [4, 5])
- **Call Type Data:** 514 samples (100.0 %), Training: 363 samples (70.6 %), Validation: 72 samples (14.0 %), Test: 79 (15.4 %) [4]

Orca Call Type/ Corpus	N01	N02	N03	N04	N05	N07	N09	N12	N47	el	whistles	ns	SUM
CCS	33	10	—	21	14	18	26	16	—	—	—	—	**138**
CCN	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**

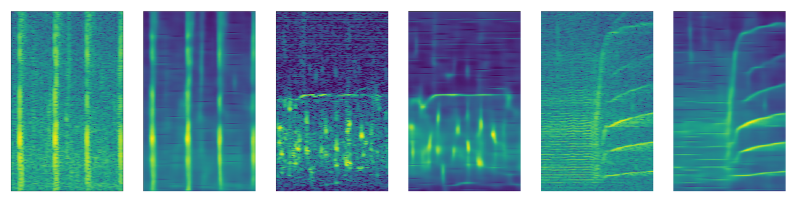
- **Network models:** (1) **Orca/Noise Segmenter** (CNN, 2-classes, cross-entropy loss) [4, 5], **Call Type Classifier** (CNN, 12-classes, cross-entropy loss) [4], and **convol. undercomplete Autoencoder** (mean squared error loss) are all based on ResNet18 [6]

Experiments – Data Preprocessing, Training, Setup

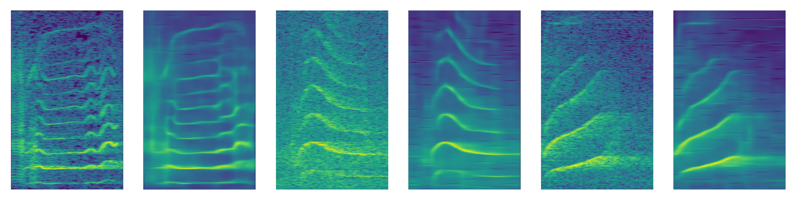


- **Data Preprocessing:** power spectrum, dB-conversion, augmentation, linear frequency compression (256 bins), noise augmentation, dB-normalization, subsampling/padding (1.28 s) $\rightarrow 1 \times 256 \times 128$
- **Network Training:** implemented in PyTorch, Adam optimizer, $\alpha = 10^{-5}$, $\beta_1 = 0.5$, $\beta_2 = 0.999$, batch size of 32 (Segmentation/Feature Learning) and 4 (Call Type Classification), α decay of 0.5 after 4 epochs and training stopped after 10 epochs without improvements on the validation set
- **Experimental Setup:** (1) Autoencoder feature learning using the automatic pre-segmented OSD dataset combined with a subsequent spectral clustering (gap statistic) using $4 \times 16 \times 8$ bottleneck features of the call type dataset, (2) Identifying potential call type sub-classes and human-misclassifications for all 514 human-labeled orca signals, and (3) Supervised [4] vs. Unsupervised Call Type Classification

Result – Reconstructions

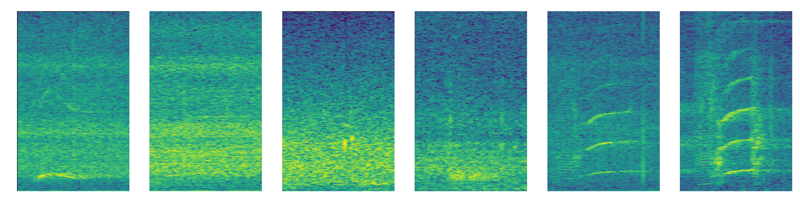


echo clicks whistle N09

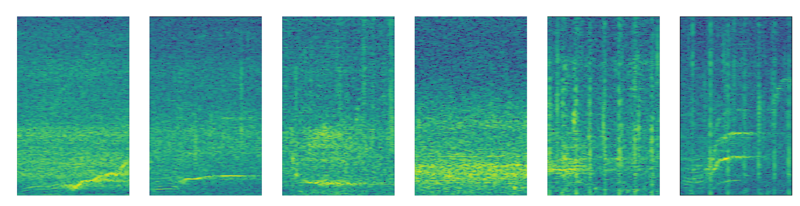


N01 N04 N47

Result – Misclassifications

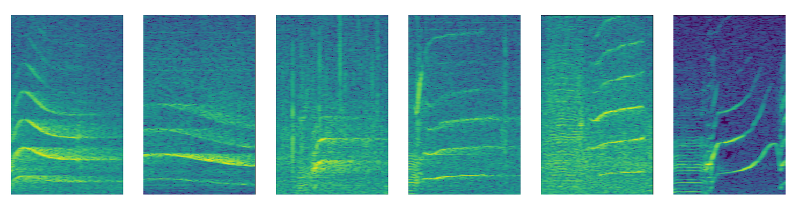


cl.a) N04/N09 **cl.b)** ns/el **cl.c)** N07/N09

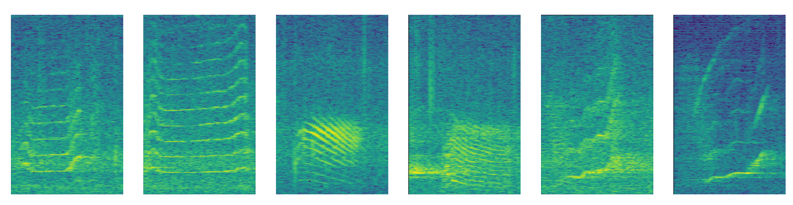


cl.d) N47/N09 **cl.e)** N03/ns **cl.f)** el/N07

Result – Sub-Call Types

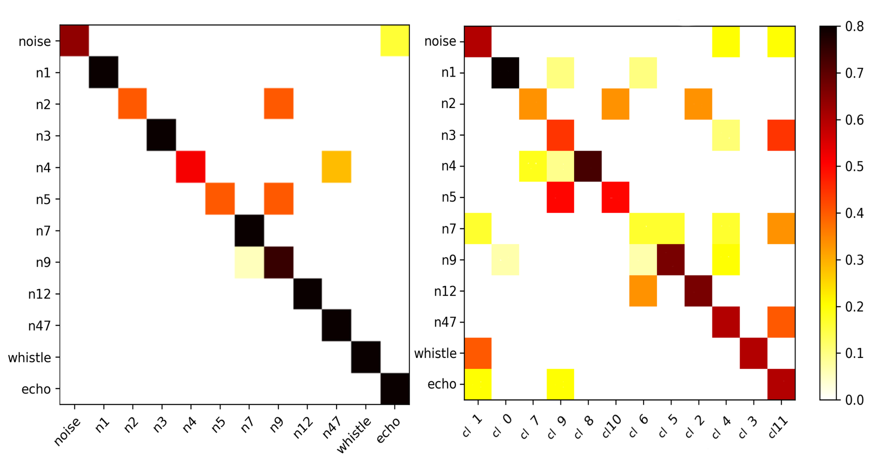


N04 cluster N07 cluster N09 cluster



N01 cluster N03 cluster N47 cluster

Result – Superv./Unsuperv.



hypothesis [4] hypothesis

Conclusion and Future Work

- Robust analysis of large datasets, no labeled data required, less susceptibility to human errors, human perception eliminated, derivation of new, previously unknown (sub-)call types
- Process entire Orchive [7] to derive totally new insights/possibilities

References

[1] O. A. Filatova, F. I. Samarra, V. B. Deecke, J. K. Ford, P. J. Miller, and H. Yurk, "Cultural evolution of killer whale calls: background, mechanisms and consequences," *Behaviour*, vol. 152, pp. 2001–2038, 2015.

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

[3] 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*, p. 165, January 1987.

[4] 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)*, May 2019.

[5] 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*, vol. 9, 12 2019.

[6] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778, 2016.

[7] S. Ness, *The Orchive : A system for semi-automatic annotation and analysis of a large collection of bioacoustic recordings.* PhD thesis, 2013.

Contact

Christian Bergler

Pattern Recognition Lab
Friedrich-Alexander University Erlangen-Nuremberg
Erlangen, Germany
☎ +49 9131 85 27872
✉ christian.bergler@fau.de

Acknowledgements

The authors would like to thank Helena Symonds and Paul Spong from OrcaLab and Steven Ness, formerly UVIC, for giving us permission to use the raw data and annotations from the orcalab.org, and the Paul G. Allen Frontiers Group for their initial grant for the pilot research. Moreover, the authors would like to thank Michael Weber for designing the autoencoder image.