Development of an AI-based bioacoustic wolf monitoring system

Olivier Stähli
College of Management of Technology
EPFL
Odyssea, Station 5
1015 Lausanne
Switzerland
olivier.staehli@epfl.ch

Thomas Ost
Life Science Dep.
Supercomputing Systems AG
Technoparkstrasse 1
8005 Zürich
Switzerland
thomas.ost@scs.ch

Thomas Studer
Institute of Computer Science
University of Bern
Neubrückstrasse 10
3012 Bern
Switzerland
thomas.studer@inf.unibe.ch

Abstract

Wolves are spreading in the Alpine region at an increasing rate, which leads to human–wolf conflicts. In order to reduce those and to perform an active wolf management, solid information about the presence of wolves is required. Getting this information is challenging since wolves are nocturnal, have sharp senses, and large territories. A monitoring method is to detect wolf howling, which can be heard over several kilometres and therefore simplifies finding them. Current acoustic methods, however, are very labour-intensive as a memory card has to be fetched from the device in the field and then the recordings have to be checked for howling manually. We present a novel approach to acoustic wolf monitoring using a convolutional neural network that runs on an embedded system in the wolf territory. Thus, we obtain accurate real-time information about the presence of wolves. On our data set, we achieve an F1-score of 0.61, thus outperforming previous systems by far. We develop prototypes and conduct two field test: first in a zoo, where we even achieve an F1-score of 0.8, and then in a wolf territory, where we successfully detect wolf howling.

Introduction

After the wolf (Canis lupus italicus) disappeared in the second half of the 19th century, it returned to Switzerland in 1995 (Breitenmoser and Breitenmoser-Würsten 2001) and has since been spreading increasingly in the Alpine region. This leads to many human–wolf conflicts because of the wolves' killing of livestock. Wolves are mostly shy and secretive animals. Finding out if and how many wolves are present in an area is a big challenge. However, a solid database on wolves is crucial for functioning wolf management.

There are various methods to detect wolves. The best known methods are DNA analysis, GPS tracking, searching for footprints, photo trap monitoring, and acoustic monitoring. While the first four methods require knowledge about the routes and movements of wolves, the acoustic method makes it possible to cover wolf territories over a large area. The howling of wolves can be heard over three kilometres (Suter et al. 2017). Thus, one microphone can theoretically cover an area of 28 km^2. A wolf territory in Switzerland is approximately 250 km^2 (Blanché, Jaeger, and Oppermann 2006). Hence, one can monitor a territory with less than ten microphones. The range also allows us to use microphones for opportunistic monitoring of wolves in areas with unknown wolf activity.

As wolves are spreading faster and faster, biologists and rangers in Switzerland face another problem: The human resources are soon no longer sufficient to generate a good database through extensive field work. Therefore, it is necessary to have a monitoring method that is cheap, automated, and scalable.

We present a novel method for automated wolf monitoring using deep learning on the edge. Our work substantially improves current methods of acoustic monitoring.

First, we developed a Convolutional Neural Network (CNN) and the corresponding feature preprocessing for detecting wolf howls. This CNN runs on an embedded system, which allows for real-time detections in the field.

Finally, we have developed a prototype, called *Wolf-Box*, and performed field tests in Switzerland. Our results show that our approach provides better detection results (F1-score) on the edge than existing off-line acoustic wolf howl detection algorithms. The results show that the system is a promising candidate to realize large-scale wolf monitoring.

Related Work

The *active* acoustic monitoring method of so-called invasive howling is controversial. It involves playing the howling of wolves in an attempt to artificially provoke a response (Reinhardt et al. 2015). Wolves not only communicate within the pack with the howl, but also to assert their territory against strangers. The artificial howls can thus cause unnecessary stress because the territorial fights of the wolves are often deadly (Halfpenny 2003). The *Howlbox* is a project from North America that relied on this kind of invasive howling. The box played a wolf howl at 6pm and 6am. Afterwards, a microphone recorded for two minutes. Since the *Howlbox* records only four minutes per night, automated analysis is not necessary. The researchers tested the setup in Idaho with two different wolf packs. Thanks to the fact that the wolves wore transmitting collars, it was possible to go directly to the rendezvous sites. After three days, twelve artificial howls elicited a response from the wolves four times(Ausband, Skrivseth, and Mitchell 2011).

In another experiment with seven wolf packs during the Wyoming winter, the response rate was only 1.1%. However, it had wolf tracks less than 50 meters from the Howlbox 14.8% of the time. Whether the artificial howl or the human scent attracted the wolves could not be conclusively determined. What is clear is that the wolves were present and did not respond (Brennan et al. 2013).

For our work, we use the *passive* method, where one waits for the wolves to start howling by themselves. A first feasibility study on non-invasive acoustic wolf monitoring has been carried out in (Suter et al. 2017). They showed that howling can be recorded on distances as far as 3 km. The devices used in their study recorded for 11 consecutive hours each night. Each audio recording was then visually scanned by the authors to detect wolves. During the biological year 2019/20, wolf monitoring was carried out in the Swiss Jura Mountains and adjacent areas in France where, among other techniques, the *Wolf Detection App* was employed (Zimmermann et al. 2020). To determine whether howling occurred, this app relies on mel-frequency cepstral coefficients. If howling is recognized, a sample of two seconds is recorded. This method has a high false-positive rate: about 200 sequences were classified as wolf each night and had to be analysed by the researchers. Moreover, it was necessary to go to the recording device to fetch the SD card with the recordings.

Requirements

In this section, we address the requirements placed on the monitoring method by the various stakeholders. The stakeholders are the wolf, the users of the equipment, mostly biologists and rangers, and nature itself.

Wolf The howling serves the wolves to know which individuals are where (Harrington 2015). This is especially important during the mating season in March. The highest howling activity occurs in late summer during the rearing of the young from the age of a few weeks (Nowak et al. 2007).

Wolf howls vary a lot. We differentiate between single howls (0.7–14 seconds long) and chorus howls, when the whole pack howls (on average 90 seconds long (Suter 2019)). We evaluated the data from the Swiss monitoring, where howling activities, not necessarily continuous, lasted for 4.56 minutes on average.

According to the literature, the frequency range in which mature wolves howl is 150–780 Hz. Juveniles howl at higher frequencies of around 1000 Hz (Nietlisbach 2014).

In zoos, we know that the wolves howl daily (Nietlisbach 2014). However, in nature, it is much harder to detect the howling due to their wide-ranging territories.

Moreover, wolves are predominantly nocturnal and therefore almost exclusively howl in darkness. The Swiss wolf monitoring has pursued evaluation of the howling activities by setting up recorders at different locations on wolves territories. Figure 1 shows an overview of the observed results.

Further evaluations on the recordings have shown that the wolves can be detected on average 19 times per season, which is about once a week, for each location.

Considering the microphones can be partly switched off during the day, we can conclude that, by splitting the recordings into 10 second sequences, we should expect about one wolf howling sequence for 2000 "non-wolf" sequences that do not contain any howling.

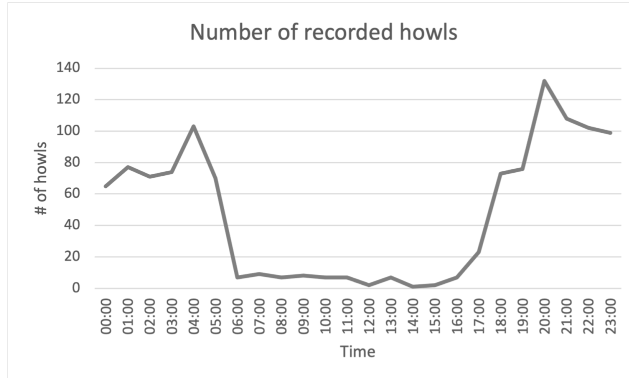

Figure 1: Wolves mostly howl before midnight and in the early morning.

User of Equipment The device should be easy to use for non-technical users. It should have at least a month of battery life and be light and small, since some monitoring locations are difficult to reach. The device shall transmit detections of the presence of wolves from the field to the user.

We prefer low false-negative rates to detect as many wolves as possible. This usually comes at the cost of more false-positive detections, which means more effort for the operators to evaluate the transmitted detection.

Nature The landscape presents challenges to our technology regarding acoustics. Weather and animals present challenges for robustness.

The landscape affects the soundscape, the attenuation of sound waves, and network coverage. Especially if there are many small side valleys or hills, the howling can be heard over smaller distances only. There is also the factor of network coverage. We need sufficient reception in order for the data to be sent from the field.

Weather requires that the installation is robust. Temperature, in particular, is a challenge. The electronics functions at temperatures between -20 C to 70 C. For the battery, however, temperatures close to freezing mean a reduction in runtime.

Different animal species place requirements on the hardware. Experience with camera traps has shown that, in particular, the marten is a challenge as it bites into cables and housings. Moreover, invertebrates could cause problems. They fit into almost all openings and could therefore enter the device.

Data

The wolf dataset consists mainly of data from the Swiss acoustic wolf monitoring, supplemented with recordings from Wild Sweden and the Cornell Lab of Ornithology. In total, it contains 1300 wolf howl recordings with a total length of 24 hours.

False-Positives

The non-wolf dataset consists of recordings of the natural environment of the monitoring locations and additional specific sounds that often trigger false positives. These can be categorized into three different classes:

- Background noises that were on wolf recordings. These noises include heavy winds, rain, and cow bells.
- Noise from vehicles like cars, planes, and motorcycles.
- Wildlife that makes similar sounds. This includes various bird species like pigeons, owls, and cuckoos (see Figure 2).

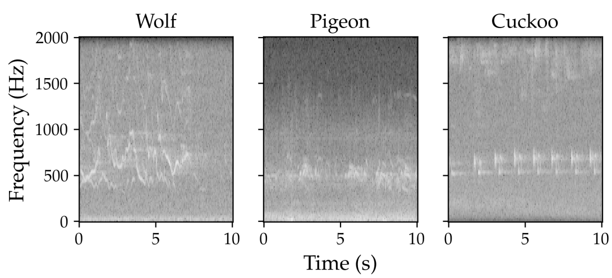

Figure 2: On the left, a spectrogram of a clear howling of two adult wolves can be seen. In the middle and on the right, we see the sounds of two animals that often trigger false positives.

Preprocessing

Preprocessing of the data, which is stored in waveform audio file format, consists of eight steps (Figure 3):

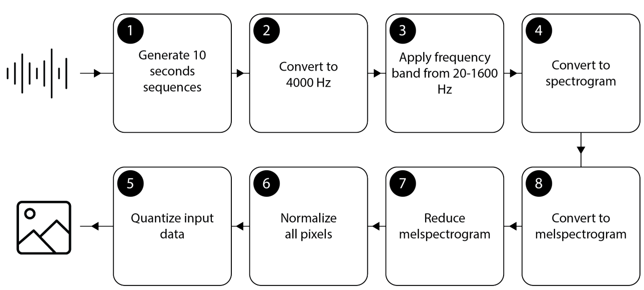

Figure 3: Preprocessing steps

1. Cut the recordings into 10 second sequences. With this length, we can do justice to the single howlers as well as the choir howlers.
2. Set the sample rate to 4000Hz.
3. Truncate the frequency below 20Hz and above 1600Hz with a Butterworth filter of order 6.
4. Convert audio files into spectrograms using the Short-time Fourier transform (STFT).
5. Map the frequency scale onto the Mel scale to form a Mel spectrogram.
6. Resize the images to 64 x 64 pixels.
7. Standardize the pixels with the means and standard deviations obtained in each time step.
8. Quantize the data during inference on the embedded system from 32-bit float to 8-bit integer values for compatibility with the edge neural network.

Deep Learning

We first looked at various architectures that delivered promising results for classification of bats, birds, and whales (Harvey 2018; Nanni et al. 2020). However, there, four input channels are used whereas we want to work with one channel, the spectrogram.

We then decided to use a custom architecture for the CNN after experiments with other models showed various problems. For instance, the ResNet architecture (He et al. 2016), which was successfully employed for acoustic orca detection (Bergler et al. 2019), was too complex to run efficiently on the Coral USB Accelerator whereas the MobileNet (Howard et al. 2017) architecture (which is particularly well-suited for embedded systems) had large accuracy losses after quantization.

Our model consists of nine convolutional layers and three max pooling layers in between. The number of filters was increased in each layer, except in the last two. A constant kernel size of three was chosen. With two dropout layers that randomly set inputs to 0, overfitting is reduced. After converting to an edge Tensor Processing Unit (TPU) model, this custom architecture gave the best results.

We used the following parameters:

Activation Function: We have chosen the usual Rectified Linear Unit (ReLU). Other options like Leaky ReLU or Parametric ReLu did not yield better results.

Optimizer Algorithms: The Adam algorithm has prevailed against RMSprop, Adadelta, and Adagrad.

Loss Function: We use cross entropy loss, which is a popular loss function in classification problems.

Class Weight: We assigned the non-wolf class 10 times more weight, which helps to reduce the number of false positives.

Batch size: We tested batch sizes of 16, 32, 64, 128, and 256. We achieved the best results with a batch size of 128.

Epochs: We trained the CNN for 30 epochs.

Edge TPU Compatibility

The target system for inference is an edge TPU. Hence, we need to convert the model for compatibility. This means to quantize the weights to 8-bit integer values. We also have to convert the model to a TensorFlow light model due to the reduced operation set allowed on the edge TPU. It also makes the model much smaller.

Running the CNN on the edge TPU yields a performance improvement of a factor 500 compared to running it on the CPU of the edge device (Raspberry Pi Zero). However, the conversion to the Lite model results in a loss of accuracy, see Table 1. Still, our results on the edge TPU are substantially better than the performance of the Wolf Detection App, which is the system currently used by Swiss wolf monitoring.

Method	TF	TF Lite	TPU	Wolf Det. App
Recall	0.77	0.70	0.70	0.69
Precision	0.95	0.54	0.54	0.038
F1-Score	0.85	0.61	0.61	0.073

Table 1: Comparison of architectures. TF stands for TensorFlow.

Embedded System

Hardware We built hardware prototypes to run the developed neural network on the edge in the field. Figure 4 shows its hardware components.

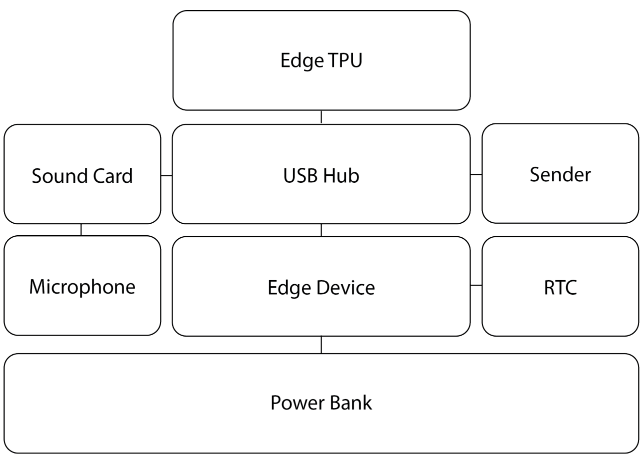

Figure 4: Hardware architecture

We use a Raspberry Pi Zero W. This fits the application purpose due to high performance (512 MB RAM, 1 GHz single kernel CPU) and low power consumption (0.7 W).

We use an MVL lavalier **microphone** from Sure with a range from 20 to 20,000 Hz and a signal-to-noise ratio of 65 dB. To digitize the microphone's recordings, a **sound card** (UGREEN USB) connects it to the edge device via a USB2 port.

On the **edge TPU**, the inference is performed. That is, an audio sequence is classified by the finished trained model. We used the Coral USB Accelerator from Google as the TPU. It performs the function of the CPU or GPU in terms of processing the neural network. A TPU is an application-specific chip specifically designed to efficiently process neural networks. Edge TPUs are characterized by low power consumption, compactness and fast processing (Jouppi et al. 2018).

With the **sender unit,** we want to make a status message whether the neural network has detected a wolf howl. The audio sequence in which the howl was detected will be transmitted. The concrete product is the 4G/3G LTE Base HAT (Hardware Attached on Top) from Sixfab. We use it to establish an Internet connection via a point-to-point protocol (PPP) over which e-mails can be sent.

We use the Witty Py to provide a **Real Time Clock** (RTC) and power management to the system.

The power supply is the limiting factor of the installation's autonomy. We **power banks** with 111 watt hours which offer the best price-performance ratio.

Software Figure 5 shows an overview of the software architecture.

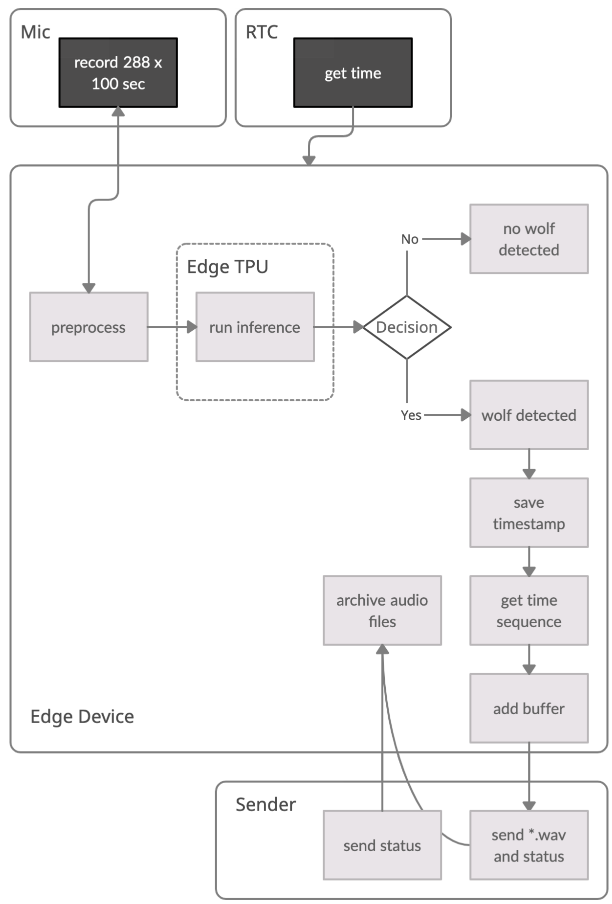

Figure 5: Software architecture

Recording is done only during the night. The sound is recorded for 100 seconds at a time. This is then written in parallel to the next 100 seconds recording in another thread as a *.wav file. This is repeated constantly during the whole recording time. For a recording time of eight hours, 288 times in total. This way we can record sound and have as little missing data between recordings as possible. The files are named with the timestamp of the howl.

The second step is the preprocessing as described in Section *Data*. After that, power is given to the Coral USB Accelerator, which performs the actual wolf detection. After all images have been run, a check is made to see if there are any wolf detections that are less than 40 seconds apart. If so, the sequences are merged. Then a ten-second snippet is added as a buffer at the beginning and end of the howling sequence.

Next, the sender is started and the Coral USB Accelerator is shut down. If no wolf was detected, an email containing status information is sent as a so-called heartbeat. If a wolf was detected, the same status information is sent, plus the number of detected sequences and of course the howl recordings themselves.

All recordings of the night are stored in an archive on the Raspberry Pi Zero. The recordings in which wolves were detected are stored separately. Once the files have been archived, the process is complete and the installation can be switched off by the Witty Pi.

Field Test

We performed two tests, first a proof of concept test in a wildlife park and then a field test in a wolf territory.

Proof of Concept

In Wildlife Park Goldau, located in central Switzerland, a pack of four European gray wolves lives in an almost two hectare large area together with two Syrian brown bears. The pack howls daily and is therefore ideally suited for our experimental setup.

The Wolf-Box was positioned southeast on the opposite slope of the enclosure at a distance of about 14 meters, facing north. The maximum distance to the microphone where the wolves could stay was 140 meters, see Figure 6. In total, the experiment was in operation for 13 days. Table 2 shows the results. During the subsequent evaluation, precision and

Figure 6: Situation at Wildlife Park Goldau

	Message	Recall	Precision
Night 1	34 wolf sequences	0.66	1.0
Night 2	8 wolf sequences	1.0	1.0
Night 2	6 wolf sequences	0.54	1.0
Night 4	no howling	-	-
Night 5	25 wolf sequences	0.92	1.0
Night 6	21 wolf sequences	1.0	0.95
Night 7	battery empty	-	-
Night 8	11 wolf sequences	1.0	1.0
Night 9	9 wolf sequences	1.0	1.0
Night 10	1253 wolf sequences	0.83	0.008
Night 11	548 wolf sequences	0.62	0.04
Night 12	system malfunction	-	-
Night 13	7 wolf sequences	0.41	1.0
Average		0.8	0.8

Table 2: Results Wildlife Park Goldau

recall were checked manually. The problem in nights 10 and 11 was that the creek between the microphone and the enclosure had more water than the days before. When the audio sequences were run on the non-quantized algorithm, it did not generate these false positives. This shows in practice that the accuracy of the quantized and converted algorithm is worse. Despite these false positives, the average F1 score, over all nights in which wolves were detected, is 0.8. This result is good enough to continue with the field test in a wolf territory.

Wolf Territory Test

For the field test, we monitored the Marchairuz pack in the Swiss canton of Vaud, which formed in 2019 and had offspring again in 2020 and 2021 (Gruppe Wolf Schweiz 2022). Three prototypes of the Wolf-Box were placed in the Marchairuz area. Two devices were in the valley and presumably close to the pack's core area. The third installation listened to the area more extensively from a hill. Unfortunately, one of the devices in the valley had a technical defect and did not deliver any data. The other device recorded a wolf howl that lasted more than two minutes in night 4. Detected were three of the fourteen parts of this sequence. On night 6, the noise of a machine was misclassified as a wolf, see Table 3. No wolf howl and no false positive sequence was detected by the installation on the hill.

	Message	Recall	Precision
Night 1	no howling	-	-
Night 2	no howling	-	-
Night 3	no howling	-	-
Night 4	3 wolf sequences (TP)	0.21	1.0
Night 5	no howling	-	-
Night 6	12 wolf sequences (FP)	-	-
Night 7	no howling	-	-
Night 8	battery empty	-	-

Table 3: Results Marchairuz pack

Conclusion

We developed the Wolf-Box, a system for passive acoustic wolf monitoring. It performs substantially better than previous systems. We obtain a precision of 0.54 on the TPU whereas, e.g., the existing wolf detection app only achieves 0.038 on our test data. Also in the field test, our prototype implementation reliably detected wolf howling. Moreover, with the current wolf detection app, the memory card has to be fetched from the device in the field, which can lead to a delay of several weeks. Our system, on the other hand, sends the results back each night. Having this real-time information is very beneficial for a successful wolf management.

For the future, we need to improve the power consumption of the Wolf-Box in particular, so that it can operate autonomously for a longer time. We are also experimenting with other CNN architectures. In the lab, we can achieve even better results than our current prototype. However, this approach does not yet run on the edge.

One could also push automation even further by having the algorithm filter out more accurate information, as suggested by (Passilongo et al. 2015): number of individuals, classification if juveniles are howling or not, identification of individuals.

Last but not least, our method can be applied to other species that communicate over long distances. The big cats lion, tiger, and leopard can also be heard over several kilometres. Further, canine species such as various hyenas would also be suitable.

Acknowledgments

We would like to thank the company Supercomputing Systems AG, which provided intensive support for the content of this work. In particular, we would like to thank Dr. Christof Bühler for his outstanding guidance. We would also like to thank the Institute for Computer Science at the University of Bern, for financing the field test. Many thanks also to Dr. Stefan Suter from ZHAW, who provided supported with his great biological expertise. Furthermore, we would like to thank Fabio Matti, Léon Delachaux, Marco Banzer, Michael Freuler, Samuel Gantner, Dr. Martin Wehrle, Beat Hörman, Pascal Kaiser, the Nature and Animal Park Goldau and the Macaulay Library at the Cornell Lab of Ornithology.

References

Ausband, D. E.; Skrivseth, J.; and Mitchell, M. S. 2011. An automated device for provoking and capturing wildlife calls. *Wildlife Society Bulletin* 35(4):498–503.

Bergler, C.; Schröter, H.; Cheng, R. X.; Barth, V.; Weber, M.; Nöth, E.; Hofer, H.; and Maier, A. 2019. ORCA-SPOT: An automatic killer whale sound detection toolkit using deep learning. *Scientific reports* 9(1):1–17.

Blanché, P.; Jaeger, R.; and Oppermann, V. 2006. *Wölfe*. Gesellschaft zum Schutz der Wölfe, 2nd edition.

Breitenmoser, U., and Breitenmoser-Würsten, C. 2001. Die ökologischen und anthropogenen voraussetzungen für die existenz grosser beutegreifer in der kulturlandschaft. *Forest Snow and Landscape Research* 76(1/2):23–39.

Brennan, A.; Cross, P. C.; Ausband, D. E.; Barbknecht, A.; and Creel, S. 2013. Testing automated howling devices in a wintertime wolf survey. *Wildlife Society Bulletin* 37(2):389–393.

Gruppe Wolf Schweiz. 2022. Rudel Marchairuz. *https://www.gruppe-wolf.ch/Rudel/Rudel-Marchairuz.htm*. Accessed: Feb. 11, 2022.

Halfpenny, J. C. 2003. *Yellowstone Wolves in the Wild*. Riverbend Publishing.

Harrington, F. H. 2015. What's in a howl?

Harvey, M. 2018. Acoustic detection of humpback whales using a convolutional neural network. *Google AI Blog*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Howard, A. G.; Zhu, M.; Chen, B.; Kalenichenko, D.; Wang, W.; Weyand, T.; Andreetto, M.; and Adam, H. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861*.

Jouppi, N. P.; Young, C.; Patil, N.; and Patterson, D. 2018. A domain-specific architecture for deep neural networks. *Communications of the ACM* 61(9):50–59.

Nanni, L.; Costa, Y. M.; Aguiar, R. L.; Mangolin, R. B.; Brahnam, S.; and Silla, C. N. 2020. Ensemble of convolutional neural networks to improve animal audio classification. *EURASIP Journal on Audio, Speech, and Music Processing* 2020:1–14.

Nietlisbach, S. 2014. Akustische Kommunikation bei Wölfen: Messungen an Gehegen.

Nowak, S.; Jkedrzejewski, W.; Schmidt, K.; Theuerkauf, J.; Myslajek, R. W.; and Jkedrzejewska, B. 2007. Howling activity of free-ranging wolves (Canis lupus) in the Białowieża Primeval Forest and the Western Beskidy Mountains (Poland). *Journal of Ethology* 25(3):231–237.

Passilongo, D.; Mattioli, L.; Bassi, E.; Szabó, L.; and Apollonio, M. 2015. Visualizing sound: counting wolves by using a spectral view of the chorus howling. *Frontiers in zoology* 12(1):1–10.

Reinhardt, I.; Kaczensky, P.; Knauer, F.; Rauer, G.; Kluth, G.; Wölfl, S.; Huckschlag, S.; and Wotschikowsky, U. 2015. *Monitoring von Wolf, Luchs und Bär in Deutschland*. BfN, Bundesamt für Naturschutz.

Suter, S. M.; Giordano, M.; Nietlispach, S.; Apollonio, M.; and Passilongo, D. 2017. Non-invasive acoustic detection of wolves. *Bioacoustics* 26(3):237–248.

Suter, S. 2019. Wolfsmonitoring mit bioakustischen Methoden. *Fauna Focus* (55).

Zimmermann, F.; Manz, R.; Chenesseau, D.; Suter, S.; Wider, S.; and Kunz, F. 2020. Monitoring du loup dans le sud du jura vaudois et la france limitrophe 2019/20. *KORA Bericht* 96.