Bird@Edge:
Bird Species Recognition at the Edge

NETYS 2022: The International Conference on Networked Systems

Jonas Höchst[1], Hicham Bellafkir[1], Patrick Lampe[1], Markus Vogelbacher[1], Markus Mühling[1], Daniel Schneider[1], Kim Lindner[2], Sascha Rösner[2], Dana G. Schabo[2], Nina Farwig[2] and Bernd Freisleben[1]

[1] Dept. of Mathematics & Computer Science
[2] Dept. of Biology

May 17th-19th, 2022

Motivation: Change in Bird Populations and Biodiversity

Continuous loss of biodiversity

- There is a sharp decline of bird populations in recent decades
- Birds are important - they interconnect habitats, resources, and ecological processes
- Birds are early warning bioindicators of an ecosystem's health

Bird species monitoring in time / space

- Traditionally achieved by human experts, acoustic or visual observation
- More recently: microphones recording audio for later analysis;
  Drawback: huge data amounts, time delay until results become available

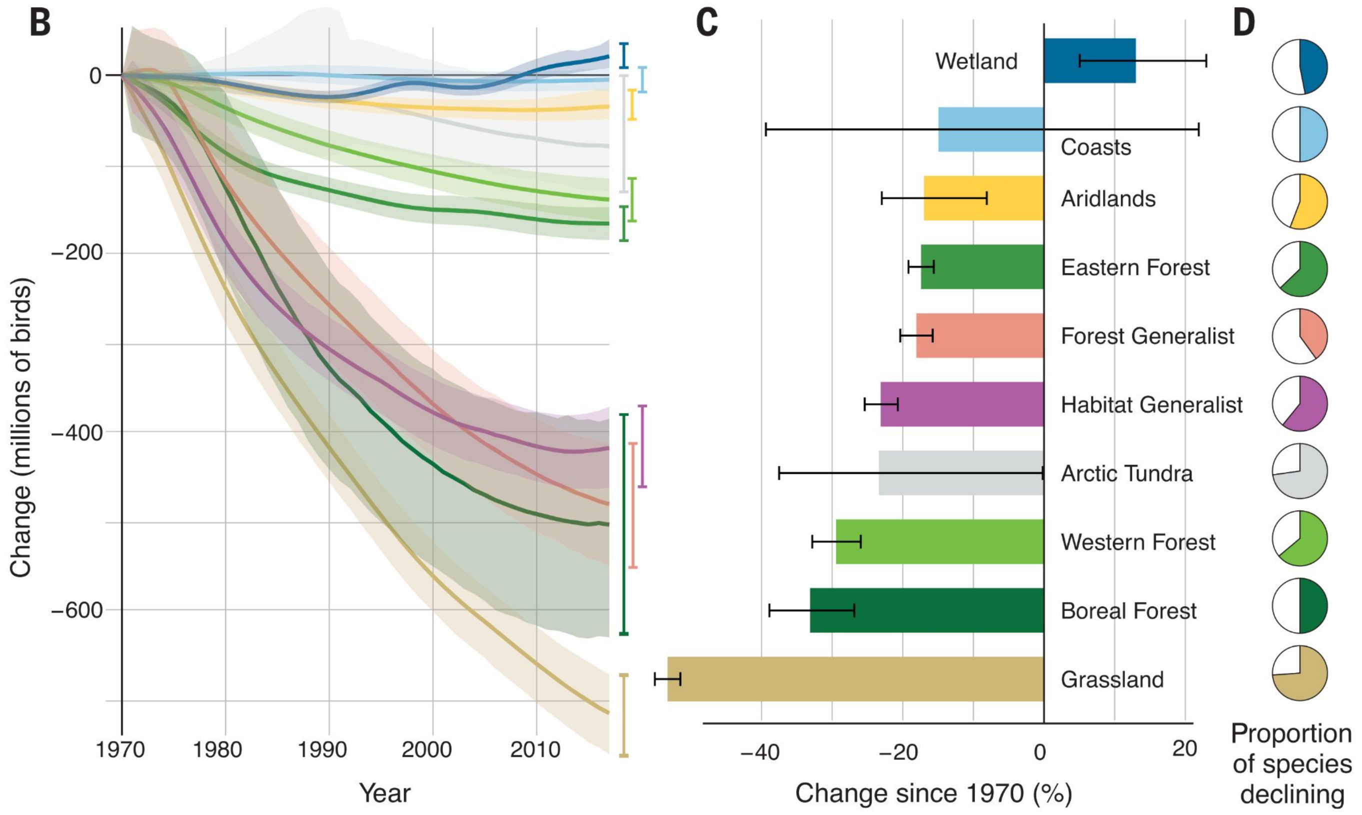


Fig.: Net loss of 2.9 billion breeding birds in North America

Souce: Rosenberg, K. V., Dokter, A. M., Blancher, P. J., Sauer, J. R., Smith, A. C., Smith, P. A., ... & Marra, P. P. (2019). Decline of the North American avifauna. Science, 366(6461), 120-124. https://www.science.org/doi/10.1126/science.aaw1313

Bird@Edge: Edge AI System

- **Bird@Edge Mic**
  - Record audio in local proximity
  - Stream audio wirelessly
- **Bird@Edge Station**
  - Create audio chunks from incoming stream
  - Perform bird species recognition for multiple Bird@Edge Mics
  - Transmit results to backend for further analysis
- **Bird@Edge Server**
  - Grafana-based Dashboards
  - Dynamic visualization of recognition results

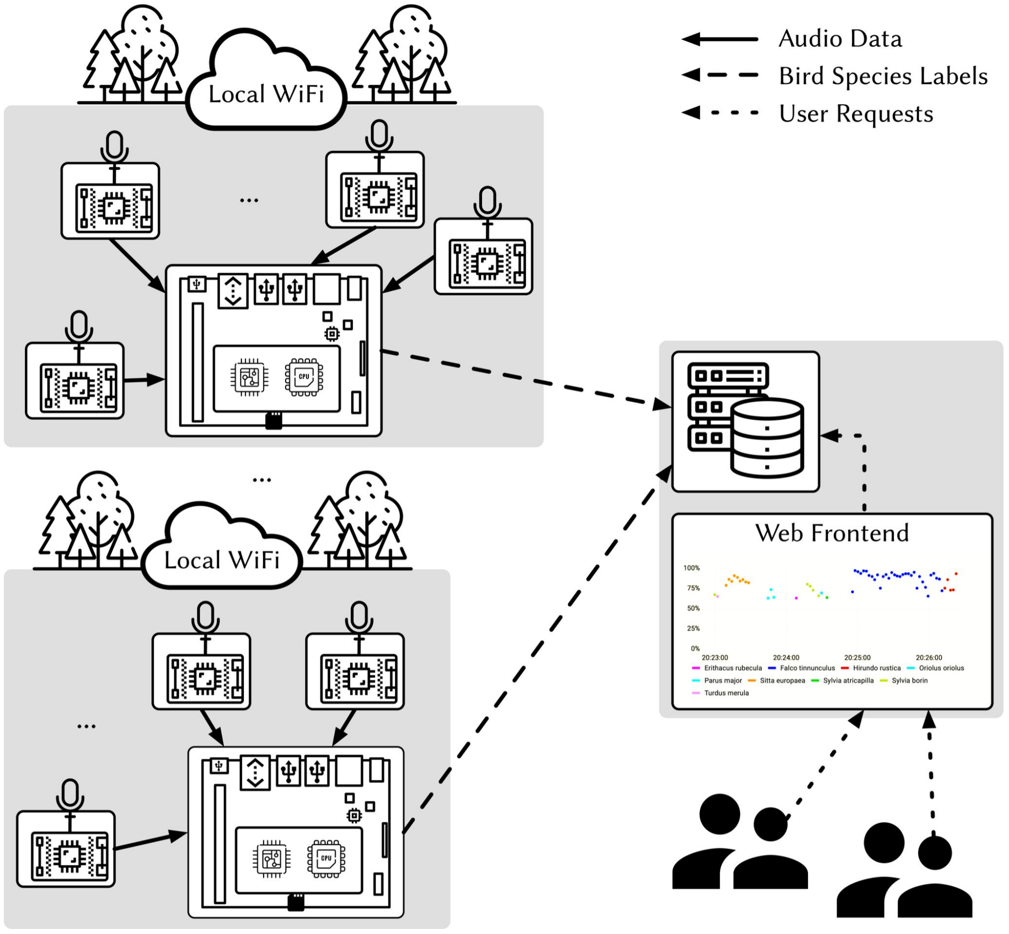


Fig. 1: Overview over the Bird@Edge system

Bird@Edge Hardware: Components and Features

Bird@Edge Mic

- ESP32 @ 80 MHz, Bluetooth, WiFi
- Knowles I²S SPH0645LM4H Microphone
- Step Down Converter, 18650 Li-Ion Cell
- 22€ - 50€, <= 500 grams

Bird@Edge Station

- NVIDIA Jetson Nano
- RTL8812BU-based WiFi
- Huawei E3372H LTE modem
- 12V / 5V Step-Down Converter
- ~110€, <= 1.5 kilograms

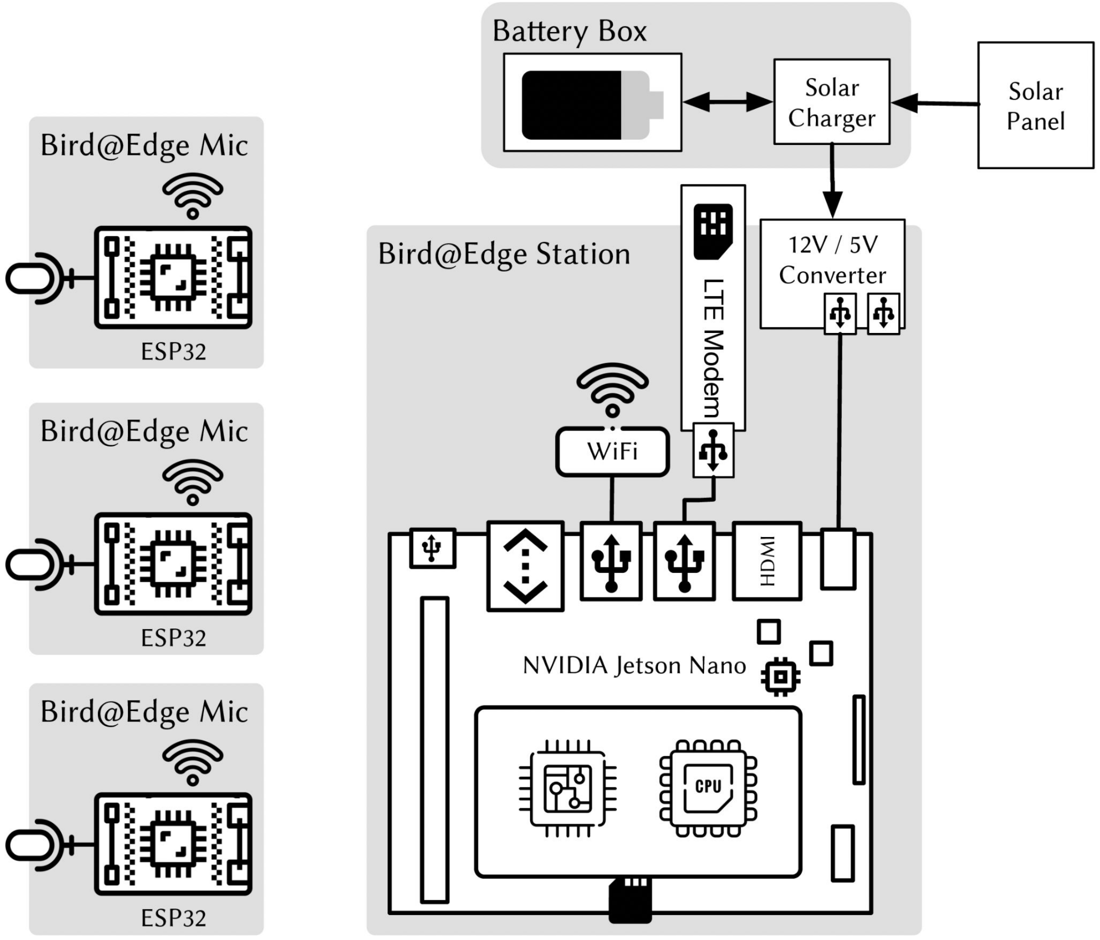
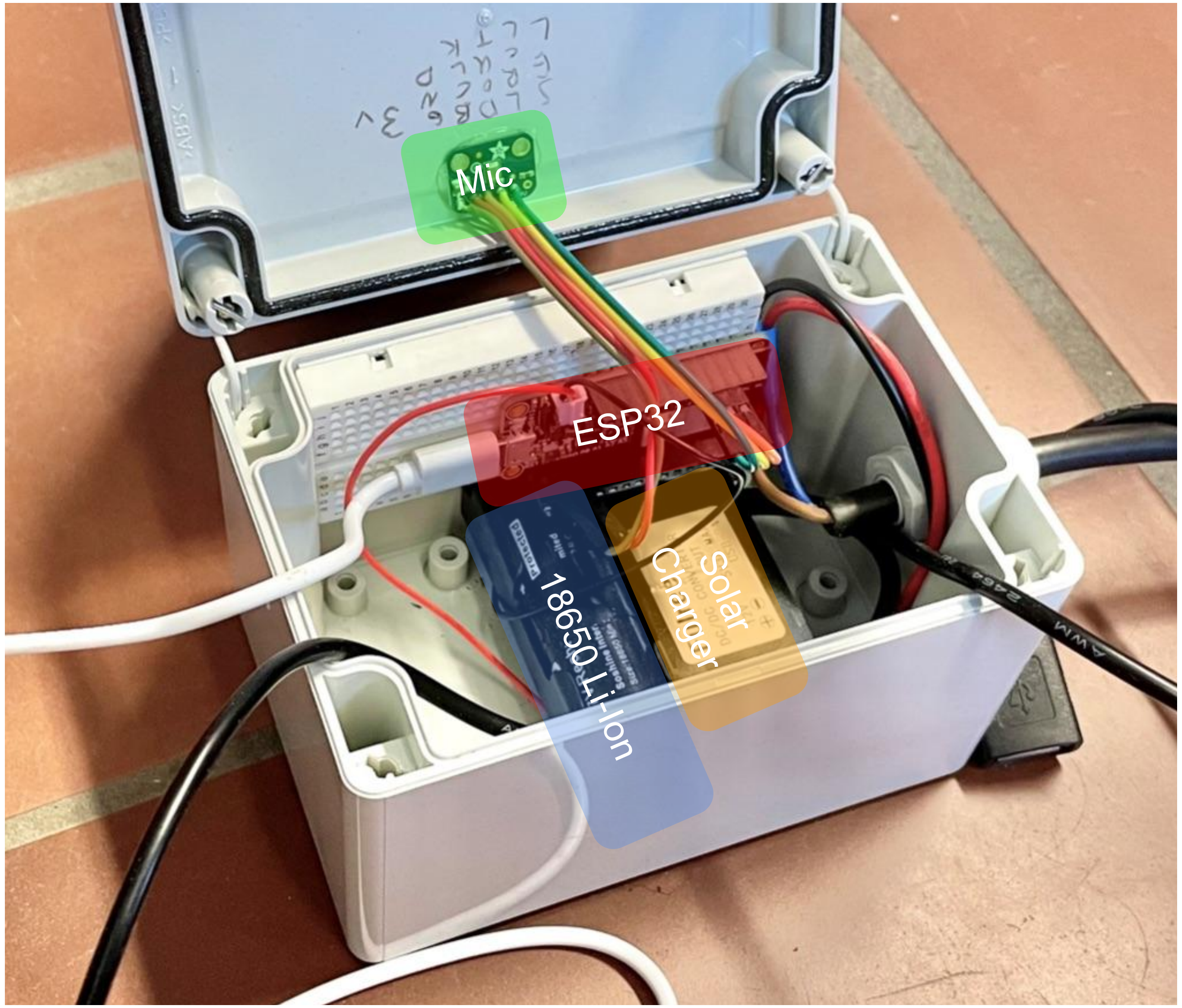
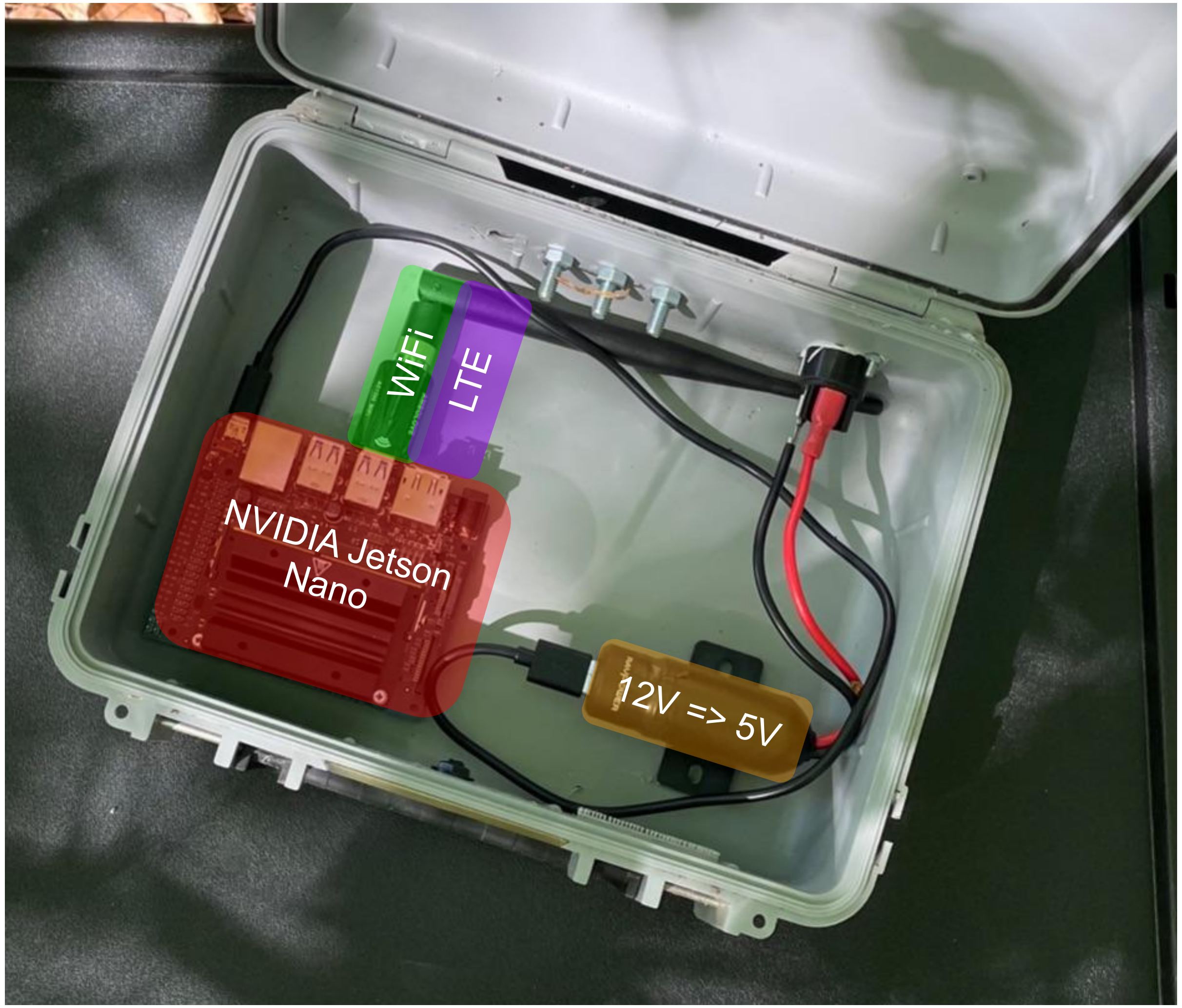


Fig. 2: Bird@Edge hardware components

Bird@Edge Hardware: Prototypes for Field Experiments





Bird@Edge Software

Bird@Edge Mic

- Built using the ESP-IDF Framework
- Connects to WiFi SSID *BirdEdge*
- Announces service via mDNS
- WiFi Connection check via ICMP (Ping)

Bird@Edge Station

- Based on Jetson Nano Development Kit OS (Ubuntu Linux Distribution)
- Runs Bird@Edge Daemon
- Searches for new Bird@Edge Mics, manages processing pipeline

Bird@Edge Server

- InfluxDB as time series database
- Grafana for visualization

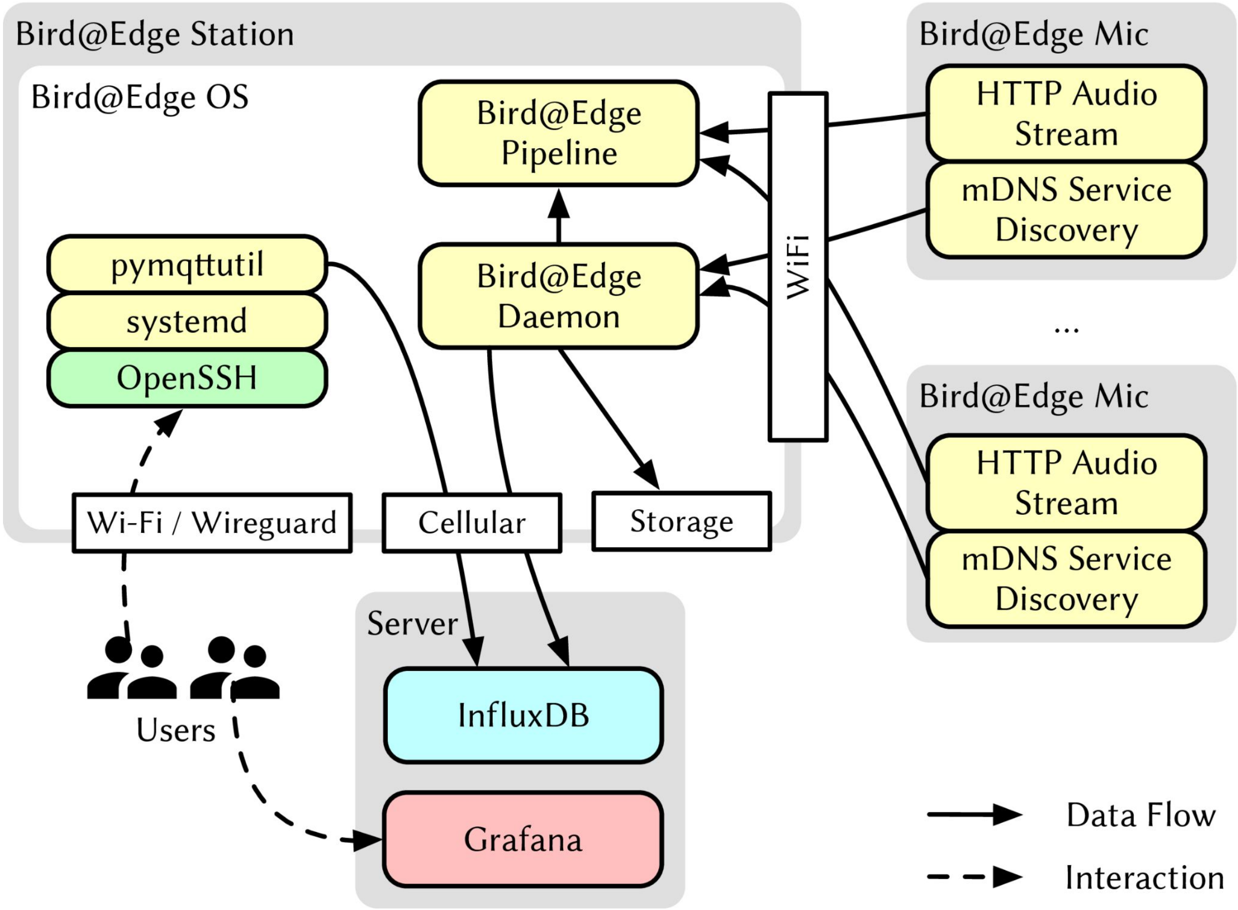


Fig. 3: Bird@Edge software components

Bird@Edge Model: Recognizing Bird Species in Soundscapes

Data Sets (82 Central European bird species)

- Marburg Open Forest
- Xeno-Canto
- iNaturalist

Data Set	MOF	Xeno-Canto	iNaturalist
Training	4,294	104,989	30,631
Test	913	2,144	1,365

Table 1: Overview of the training and test data.

Pre-processing

- Select 44.1 kHz as the sampling rate
- Random selection of 5 second snippets
- Noise augmentation with realistic background sounds
- Transformation of audio snippets into visual representations via Short-time Fourier transform (STFT)

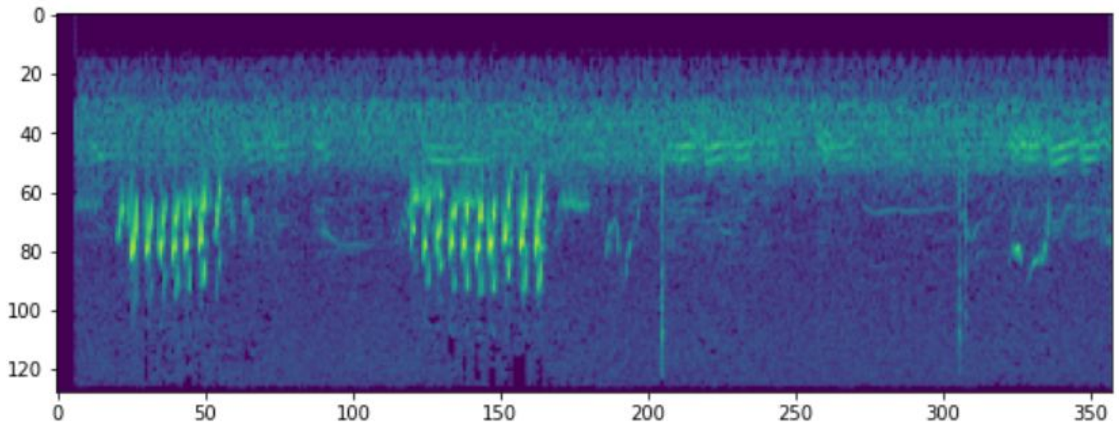


Fig.: Log-Mel-Spectrogram

Bird@Edge Model: Recognizing Bird Species in Soundscapes

Philipps Universität Marburg

Model Architecture

- Deep Convolutional Neural Networks (CNNs)
- EfficientNet-B3 (pre-trained on ImageNet)
- Trade-off between runtime and performance

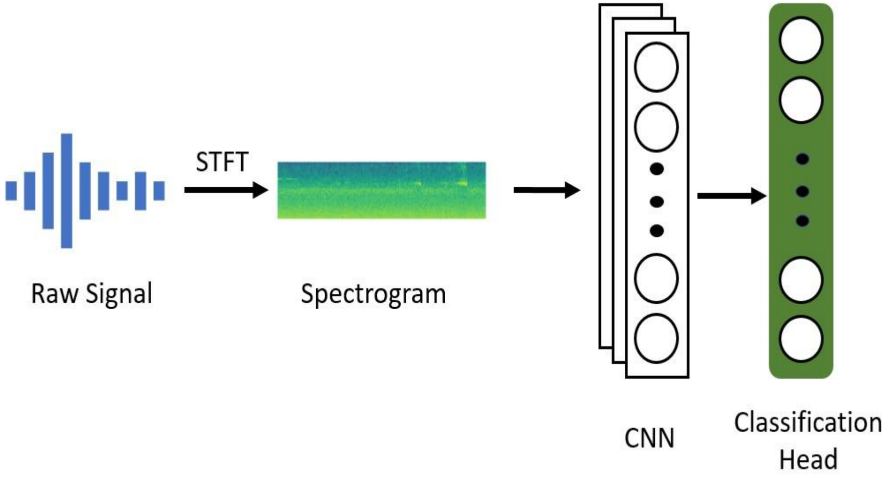


Fig.: Model architecture

Training

- Two-phase training on GPU server
- Supervised optimization
- Optimization using focal loss
- TensorFlow deep learning framework

$$L = \sum_{k=1}^{K} l(y_k, p_k),$$

$$l(y,p) = \begin{cases} -\alpha_{pos}(1-p)^\gamma \log(p) & \text{if } y \text{ is positive} \\ -\alpha_n p^\gamma \log(1-p) & \text{if } y \text{ is negative} \\ -\alpha_{hn} p^\gamma \log(1-p) & \text{if } y \text{ is hard negative} \end{cases}$$

Bird@Edge Model: Optimization and Deployment

Optimization

- Nvidia TensorRT[1]
- Apply quantization (FP32 → FP16)
- Reduce inference time while maintaining accuracy

Inference

- Nvidia DeepStream SDK[2]
- Upscale to multiple live input streams
- Apply highpass filter to reduce noise
- Inference on input data using the NvInferaudio Gstreamer plug-in

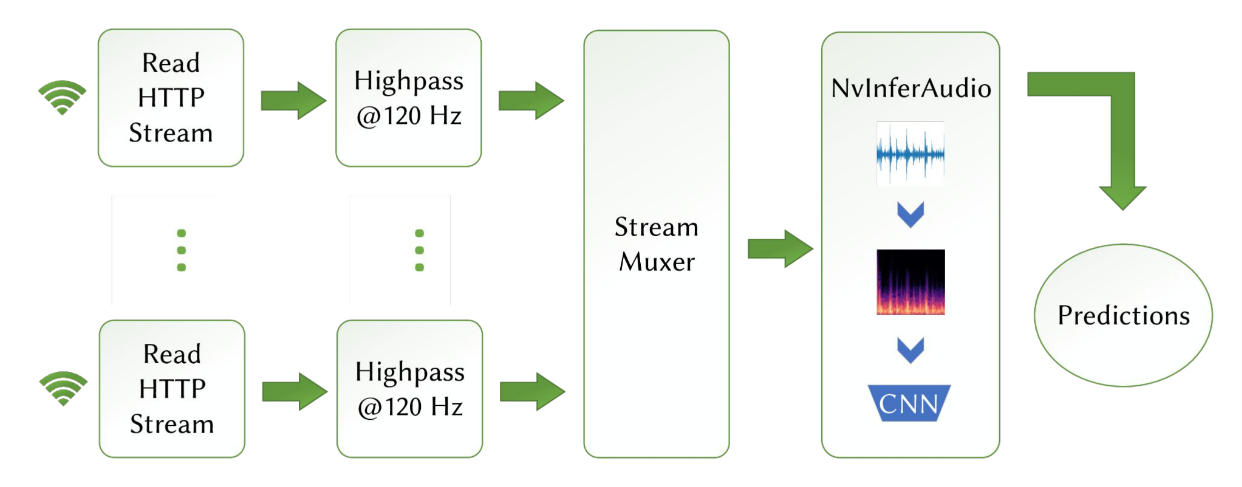


Fig. 4: Overview of the Bird@Edge processing pipeline

1 https://developer.nvidia.com/tensorrt
2 https://developer.nvidia.com/deepstream-sdk

Bird@Edge Model: Experimental Evaluation

Quality

- Similar performance for server & edge model
- Better average precision (AP) results compared to BirdNET and BirdNET-Lite[1] on species that occur in Germany

Method	MOF	XC	iNat
BirdNET	0.833	0.725	0.725
BirdNET-Lite	0.859	0.737	0.714
EfficientNet-B3	0.952	0.820	0.811
Bird@Edge	0.952	0.816	0.819

Table 2: Results (mAP)

Runtime

- Low inference runtime using Jetson Nano GPU
- 10 ms runtime reduction using the optimized CNN model

Model	Device	Inference time (ms)
BirdNET-Lite	Raspberry Pi-4B	279
Bird@Edge (FP32)	Jetson Nano	64
Bird@Edge	Jetson Nano	54

Table 3: Model inference runtimes

[1] https://github.com/kahst/BirdNET-Lite

Experimental Evaluation: Visualization of Bird Species Recognition Results

Automatically generated graph of recognized bird species

- Data from a single Bird@Edge Mic
- x: Time / y: Confidence of recognition

Observations:

- Several occurrences of *Coccothraustes coccothraustes* (hawfinch) in two clusters
- *Picus canus* (grey-headed woodpecker) detected in one cluster in the middle
- *Sitta europaea* (Eurasian nuthatch) is detected in two clusters (start and end)

More complex analysis can be performed in the cloud, e.g., on multiple stations.

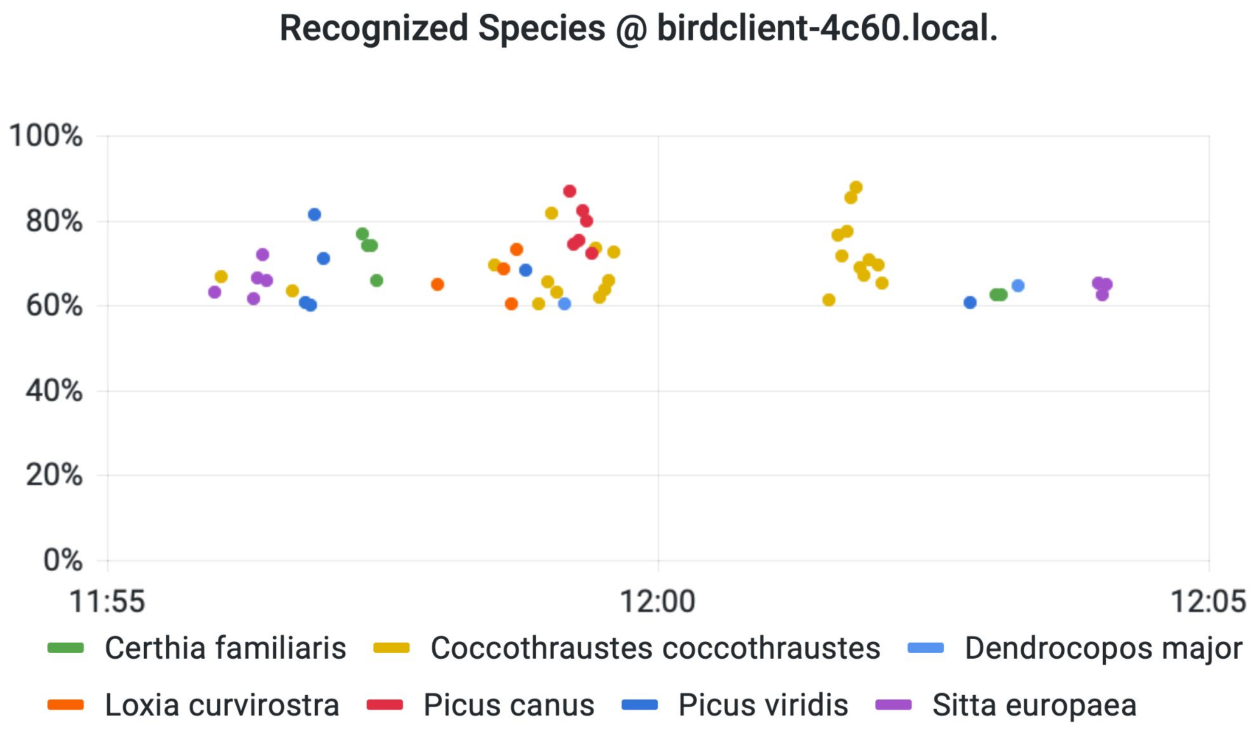


Fig. 5: Grafana panel (x-axis: clock time; y-axis: recognition confidence) showing recognized bird species of a certain Bird@Edge Mic, based on Xeno-Canto file XC706150, recorded by user *brickegickel*

Experimental Evaluation: Power Consumption

Applicability of system is limited by power consumption aspects

Bird@Edge Station

- Custom low power profile for Jetson
- Power requirement: ~3.16 W
- 12V Battery @ 100 Ah: ~14 days
- Continuous operation:: 50-100 W solar panel

Bird@Edge Mic

- Power requirement: ~0.49 W
- 3.3V Li-Ion Battery @ 3500 mAh: ~27.6 hours
- Continuous operation: 10 W solar panel

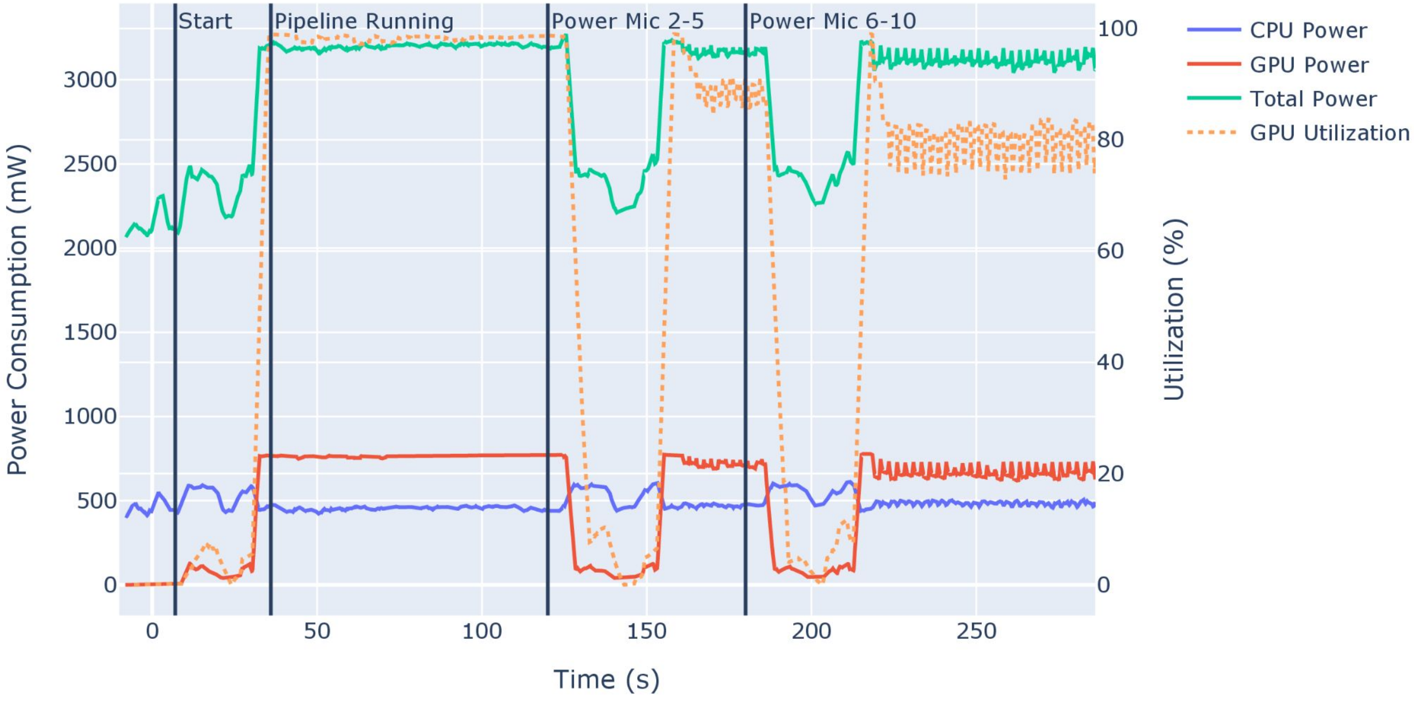


Fig. 6: Power consumption of a Bird@Edge Station in a dynamic scenario.

Conclusion

Bird@Edge: Edge AI system for recognizing bird species in audio recordings to support real-time biodiversity monitoring

- EfficientNet-B3 architecture, optimized for execution on NVIDIA Jetson Nano
- 95.2% mean average precision, outperforms state-of-the-art BirdNET
- Low power demand of 3.18 W (Station) and 0.492 W (Mic)
- Software components are open source: https://github.com/umr-ds/BirdEdge

Future Research

- Self-supervised learning to leverage unlabeled data recorded in the fields
- Real-world long-term test of Bird@Edge

Bird@Edge:
Bird Species Recognition at the Edge

Philipps Universität Marburg

Thank You!

Questions & Discussion:

- Jonas Höchst <hoechst@uni-marburg.de>
- Hicham Bellafkir <hicham.bellafkir@uni-marburg.de>

https://github.com/umr-ds/BirdEdge

May 17th-19th, 2022

References

1. Abadi, M., Agarwal, A., Barham, P., Brevdo, E., Chen, Z., Citro, C., Corrado, G.S., Davis, A., Dean, J., Devin, M., Ghemawat, S., Goodfellow, I., Harp, A., Irving, G., Isard, M., Jia, Y., Jozefowicz, R., Kaiser, L., Kudlur, M., Levenberg, J., Mané, D., Monga, R., Moore, S., Murray, D., Olah, C., Schuster, M., Shlens, J., Steiner, B., Sutskever, I., Talwar, K., Tucker, P., Vanhoucke, V., Vasudevan, V., Viégas, F., Vinyals, O., Warden, P., Wattenberg, M., Wicke, M., Yu, Y., Zheng, X.: TensorFlow: Large-scale machine learning on heterogeneous systems (2015), https://www.tensorflow.org/
2. Darras, K., Batáry, P., Furnas, B., Celis-Murillo, A., Van Wilgenburg, S.L., Mulyani, Y.A., Tscharntke, T.: Comparing the sampling performance of sound recorders versus point counts in bird surveys: A meta-analysis. Journal of Applied Ecology 55(6), 2575–2586 (2018). https://doi.org/10.1111/1365-2664.13229
3. Disabato, S., Canonaco, G., Flikkema, P.G., Roveri, M., Alippi, C.: Birdsong detection at the edge with deep learning. In: 2021 IEEE International Conference on Smart Computing (SMARTCOMP). pp. 9–16. IEEE (2021)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., Houlsby, N.: An image is worth 16x16 words: Transformers for image recognition at scale. In: 9th Int. Conference on Learning Representations, ICLR 2021, Austria (2021)
5. Gallacher, S., Wilson, D., Fairbrass, A., Turmukhambetov, D., Mac Aodha, O., Kreitmayer, S., Firman, M., Brostow, G., Jones, K.: Shazam for bats: Internet of things for continuous real-time biodiversity monitoring. IET Smart Cities (2021)
6. Gong, Y., Chung, Y., Glass, J.R.: AST: audio spectrogram transformer. In: Interspeech 2021. pp. 571–575 (2021). https://doi.org/10.21437/Interspeech.2021-698
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016). https://doi.org/10.1109/CVPR.2016.90.
8. Henkel, C., Pfeiffer, P., Singer, P.: Recognizing bird species in diverse soundscapes under weak supervision. In: Faggioli, G., Ferro, N., Joly, A., Maistro, M., Piroi, F. (eds.) Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, Bucharest, Romania, September 21-24, 2021. CEUR Workshop Proceedings, vol. 2936, pp. 1579–1586. CEUR-WS.org (2021), http://ceur-ws.org/Vol-2936/paper-134.pdf
9. Hill, A.P., Prince, P., Snaddon, J.L., Doncaster, C.P., Rogers, A.: Audiomoth: A low-cost acoustic device for monitoring biodiversity and the environment. Hardware X 6, e00073 (2019). https://doi.org/10.1016/j.ohx.2019.e00073
10. Höchst, J., Penning, A., Lampe, P., Freisleben, B.: PIMOD: A Tool for Config- uring Single-Board Computer Operating System Images. In: 2020 IEEE Global Humanitarian Technology Conference (GHTC 2020). pp. 1–8. Seattle, USA (Oct 2020). https://doi.org/10.1109/GHTC46280.2020.9342928
11. iNaturalist: A community for naturalists, https://www.inaturalist.org/
12. Kahl, S., Clapp, M., Hopping, W.A., Goëau, H., Glotin, H., Planqué, R., Vellinga, W., Joly, A.: Overview of birdclef 2020: Bird sound recognition in complex acoustic environments. In: Cappellato, L., Eickhoff, C., Ferro, N., Névéol, A. (eds.) Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Forum, Thessaloniki, Greece, September 22-25, 2020. CEUR Workshop Proceedings, vol. 2696. CEUR-WS.org (2020), http://ceur-ws.org/Vol-2696/paper_262.pdf
13. Kahl, S., Denton, T., Klinck, H., Glotin, H., Goëau, H., Vellinga, W., Plan- qué, R., Joly, A.: Overview of birdclef 2021: Bird call identification in sound- scape recordings. In: Faggioli, G., Ferro, N., Joly, A., Maistro, M., Piroi, F. (eds.) Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, Bucharest, Romania, September 21-24, 2021. CEUR Workshop Proceedings, vol. 2936, pp. 1437–1450. CEUR-WS.org (2021), http://ceur-ws.org/Vol-2936/paper-123.pdf
14. Kahl, S., Wood, C.M., Eibl, M., Klinck, H.: Birdnet: A deep learning solution for avian diversity monitoring. Ecological Informatics 61, 101236 (2021). https://doi.org/10.1016/j.ecoinf.2021.101236
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: Bengio, Y., LeCun, Y. (eds.) 3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings (2015), https://arxiv.org/abs/1412.6980
16. Lin, T.Y., Goyal, P., Girshick, R., He, K., Dollar, P.: Focal loss for dense object detection. 2017 IEEE International Conference on Computer Vision (ICCV) (Oct 2017)
17. McFee, B., Raffel, C., Liang, D., Ellis, D.P., McVicar, M., Battenberg, E., Nieto, O.: librosa: Audio and music signal analysis in python. In: Proceedings of the 14th python in science conference. vol. 8 (2015)
18. Merenda, M., Porcaro, C., Iero, D.: Edge Machine Learning for AI-enabled IoT devices: A Review. Sensors 20(9), 2533 (2020)
19. Mühling, M., Franz, J., Korfhage, N., Freisleben, B.: Bird species recognition via neural architecture search. In: Cappellato, L., Eickhoff, C., Ferro, N., Névéol, A. (eds.) Working Notes of CLEF 2020 - Conference and Labs of the Evaluation Fo- rum, Thessaloniki, Greece, September 22-25, 2020. CEUR Workshop Proceedings, vol. 2696. CEUR-WS.org (2020), http://ceur-ws.org/Vol-2696/paper_188.pdf
20. Puget, J.F.: Stft transformers for bird song recognition. In: Working Notes of CLEF 2021 - Conference and Labs of the Evaluation Forum, Bucharest, Romania, September 21-24, 2021. CEUR Workshop Proceedings, vol. 2936. CEUR-WS.org (2021), http://ceur-ws.org/Vol-2936/paper-137.pdf
21. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. International journal of computer vision 115(3), 211–252 (2015)
22. Tan, M., Le, Q.V.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: Chaudhuri, K., Salakhutdinov, R. (eds.) Proceedings of the 36th International Conference on Machine Learning, ICML 2019, 9-15 June 2019, Long Beach, California, USA. Proceedings of Machine Learning Research, vol. 97, pp. 6105–6114. PMLR (2019). https://doi.org/1905.11946
23. Xeno-canto: Sharing bird sounds from around the world, https://www.xeno-canto.org/
24. Zoph, B., Le, Q.V.: Neural architecture search with reinforcement learning. In: 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings (2017)
25. Zualkernan, I., Judas, J., Mahbub, T., Bhagwagar, A., Chand, P.: An aiot system for bat species classification. In: 2020 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS). pp. 155–160 (2021). https://doi.org/10.1109/IoTaIS50849.2021.9359704