Chenning Li and Hanqing Guo *Michigan State University, MI* **Shuai Tong** *Tsinghua University, China*
Zhichao Cao, Mi Zhang, Qiben Yan and Li Xiao *Michigan State University, MI*
Jiliang Wang and Yunhao Liu *Tsinghua University, China*

Editors: Nicholas D. Lane and Xia Zhou

NELoRa: Neural-enhanced Demodulation for Low-Power WANs

Excerpted from "NELoRa: Towards Ultra-low SNR LoRa Communication with Neural-enhanced Demodulation," from *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems* with permission. https://dl.acm.org/doi/10.1145/3485730.3485928 ©ACM 2021

Low-Power Wide-Area Networks (*LPWANs*) has emerged as a promising mechanism to connect billions of low-cost Internet of Things (IoT) devices for wide-area data collection. Long Range (LoRa) [1] is a commercialized and widely deployed wireless technology that facilitates the establishment of LPWANs. As illustrated in Figure 1, a LoRaWAN consists of end nodes, gateways, a network server, and an application server. The collected sensory data (e.g., temperature, humidity) transmitted from the distributed end nodes are relayed by several gateways to the network server.

To enable such long-distance communication, the physical layer of LoRa employs Chirp Spread Spectrum (CSS) modulation at the end nodes and dechirp at the gateways. By modulating data via CSS, LoRa allows sensor nodes to send data at low data rates to gateways several or even tens of miles away. Unfortunately, recent studies [2, 3] show that the communication range of LoRa is far from the expectation in complex real-world environments (e.g., urban areas, campuses). The blockage attenuation could severely degrade the Signal-to-Noise Ratio (SNR) of LoRa packets, causing decoding failures even at a sub-kilometer distance. Consequently, a LoRa node has to adapt its configuration with more energy consumption to compensate for the SNR degradation, reducing its battery life drastically.

Status Quo and their Limitations. Status quo approaches [4-7] obtain extra SNR gains by enhancing the observed SNR over a weak LoRa link. Such extra SNR gains are obtained by leveraging information from multiple LoRa nodes or gateways. Although these approaches have achieved impressive SNR gains, such gains are costly if the LoRa nodes and gateways are not densely deployed.

The root cause of the limitations shared across the status quo approaches described above is that they are all designed based on dechirp, which decodes a chirp symbol by only relying on its energy in the spectrum. Such a design choice, though simple, is coarse-grained: it ignores fine-grained information embedded inside the chirps, which can be helpful in chirp symbol decoding.

Overview of the Proposed Approach. The limitation of dechirp motivates us to rethink the design of the LoRa demodulation method. To this end, we present *NELoRa*, a neural-enhanced demodulation method that achieves ultra-low SNR LoRa communication with a single gateway. The key idea of NELoRa is to use Deep Neural Networks (DNN) to extract the fine-grained information embedded inside the chirps for decoding. Compared to the single-dimension energy information used in dechirp, the extracted fine-grained information contains robust and consistent multi-dimension patterns across the chirps' time, frequency, phase, and energy information. By doing this, NELoRa breaks the SNR threshold of dechirp and obtains extra SNR gains by lowering the SNR threshold. As a result, NELoRa can enlarge the LoRa communication range and reduce the energy consumption at a single gateway.

CREATING THE DUAL-CHANNEL SPECTROGRAM

Given a chirp symbol, we first divide it into a sequence of short chirp segments with equal length. Then we compute the spectrum of each short chirp segment separately as dechirp does to generate an amplitude spectrogram. In addition, we extract the phase of each short chirp segment's spectrum, which leads to a dual-channel spectrogram. To demonstrate the feature space of our dual-channel spectrogram, we collect four chirp symbols representing four different but very close data bits (e.g., 0x40, 0x41, 0x43, 0x45) at a high SNR level. The dual-channel spectrogram creates the desired feature space only if the spectrograms of the four chirp symbols are distinct enough. Then, we take 0x40 chirp symbol as a reference to compute the dual-channel spectrogram differences with others.

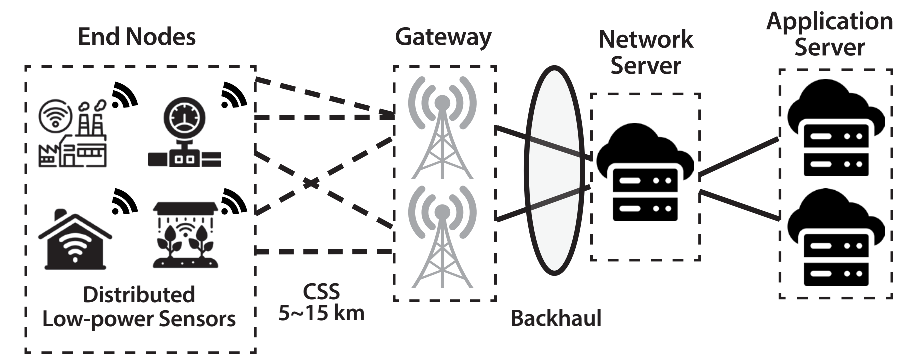


FIGURE 1. Illustration of LoRaWAN architecture.

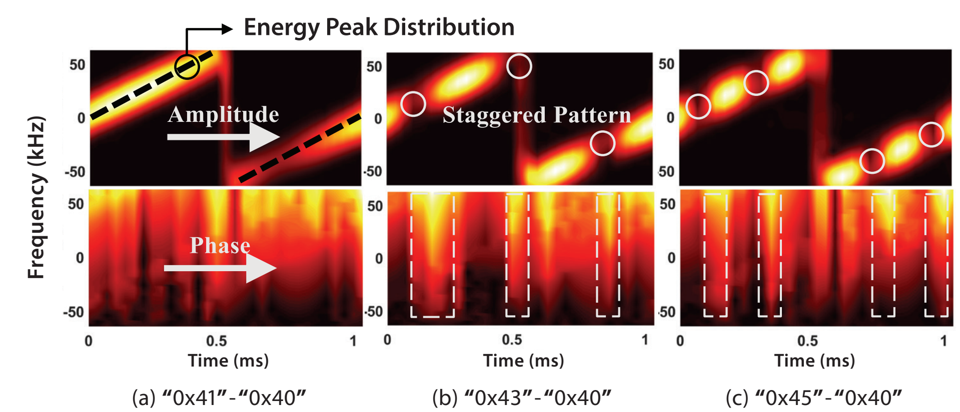
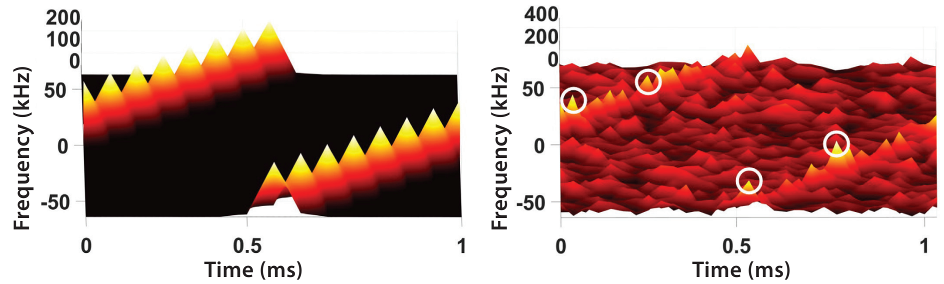


FIGURE 2. Dual-channel spectrogram.

The results are shown in Figure 2. The spectrograms of amplitude and phase are on the top and bottom, separately. For a chirp symbol, the spectrum energy peaks derived by continuous short chirp segments form a linearly increasing energy peak distribution in its amplitude spectrogram. The initial frequency of the energy peak distribution is determined by the initial frequency of the chirp symbol, which corresponds to the data bits it represents. As shown at the top of Figure 2, although the encoded initial frequencies among the four chirps are close, we can clearly observe the energy peak distributions from the amplitude spectrogram differences. This indicates that energy peak distribution is a useful feature dimension. Additionally, as shown in Figure 2b and Figure 2c, we can see peaks (e.g., bright areas) and valleys (e.g., dark areas) appear alternatively in both amplitude (e.g., circle) and phase (e.g., dashed rectangle) spectrograms. Specifically, the patterns observed by amplitude and phase are correlated, but different chirp symbols exhibit diverse patterns. Hence, the staggered pattern is another feature dimension of the dual-channel spectrogram to distinguish different chirp symbols.

When SNR is getting low, however, the dual-channel spectrogram will be polluted by noise. To illustrate this, we collect a chirp symbol to calculate its amplitude spectrogram under different SNR levels. As shown in Figure 3, when SNR is 35 dB, we can see the spectrum energy peaks of all short chirp segments. When SNR drops to -10 dB, Figure 3 shows only several short chirp segments' energy peaks (e.g., white circles) that can be explicitly observed compared to surrounding noise energy. Facing the seriously polluted dual-channel spectrogram, a DNN can succeed in recognizing chirp symbols due to the noise-resilient patterns obtained from both amplitude and phase spectrograms. Specifically, the energy peak distribution exhibits a linear pattern, which can still be observed with several explicit energy peaks in Figure 3. Moreover, the staggered pattern exists in both amplitude and phase spectrograms. Since the amplitude and phase of a short chirp segment's spectrum are affected by the noise independently, the staggered pattern has the potential to tolerate specific noise. A well-designed DNN is good at learning these patterns. Although random noises may be much stronger than chirp symbols, it is hard to simultaneously form similar patterns in multi-dimensional feature space to mislead the DNN. Hence, we feed the dual-channel spectrogram of a chirp symbol to our DNN.



(a) SNR = 35 (b) SNR = -10

FIGURE 3. Features at different SNR levels.

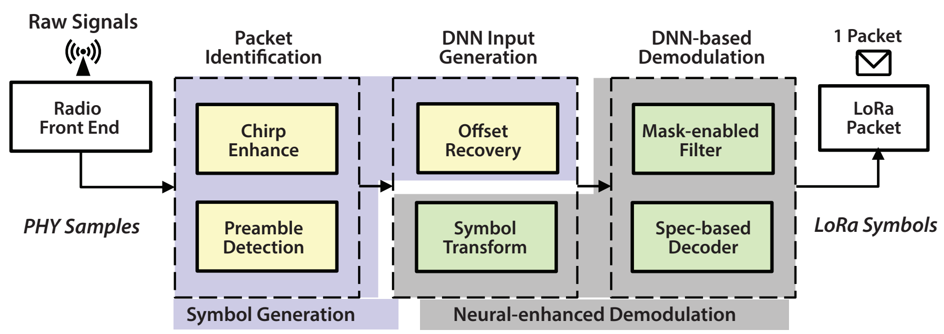


FIGURE 4. Overall architecture of NELoRa.

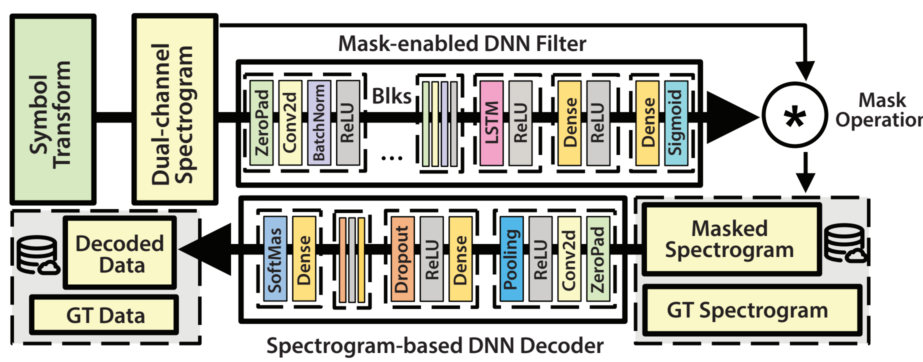


FIGURE 5. Architecture of the dual-DNN model.

NELoRa OVERVIEW

Figure 5 illustrates the overall architecture of NELoRa. NELoRa consists of three stages to achieve reliable symbol generation and neural-enhanced demodulation. In the Packet Identification stage, a LoRa packet is first detected from raw signal samples via the Chirp Enhance and Preamble Detection modules. The detected packet is then imported into the DNN Input Generation stage. The Offset Recovery module exploits the redundant chirp symbols in packet preamble to compensate offsets in frequency and time domains to generate the time-aligned and offset-free chirp symbols in packet payload. Each extracted chirp symbol is then transformed by the Symbol Transform module into a dual-channel spectrogram. The final stage is DNN-based Demodulation. Given the dual-channel spectrogram, the Mask-enabled Filter module alleviates the channel noise to obtain a masked spectrogram, which is then decoded by the Spectrogram-based Decoder module to generate the packet.

Packet Identification. The default packet detection method utilizes the preamble of a LoRa packet, which consists of multiple continuous base up-chirps. To tolerate a lower SNR threshold than the one used in dechirp, instead of using the energy peak of a window chirp, we sum up multiple continuous window chirps to form an enhanced window chirp, in which the window chirps are added up coherently, but the random noise is not. We apply dechirp on the enhanced window chirp to obtain an accumulated energy peak, surpassing the randomly increased noise energy. In theory, when we sum up eight window chirps coherently, the resulting SNR gains will be 9 dB. If the energy peak of the enhanced window chirp is higher than the average noise energy, a LoRa packet is detected. Due to the existence of carrier frequency offset (CFO) and sampling frequency offsets (SFO), which introduce phase shifts onto the window chirps that accumulate over time, different window chirps may have different initial phases. To take advantage of coherently overlapping, we use a greedy search method to find the clock drift, which can accurately estimate the phase offset.

Neural-enhanced Demodulation. As shown in Figure 5, our dual-DNN model includes two modules, the noise filter and the spectrogram-based decoder for noise reduction and chirp symbol decoding, respectively. The first module aims to preserve the primary spectrogram features of a chirp symbol by masking the raw dual-channel spectrogram. In a conceptual sense, the noise filter is more like an enhanced shortcut connection in the ResNet block [8] by transforming the shortcut from layers into ends. It contains multiple blocks of CNN and one LSTM to fully exploit the spatial and temporal features of the input, followed by two dense layers to output a well-matched mask. Moreover, a four-layer CNN-based decoder is designed to fully capture the spatial energy peak distribution and temporal staggered pattern in the masked spectrogram.

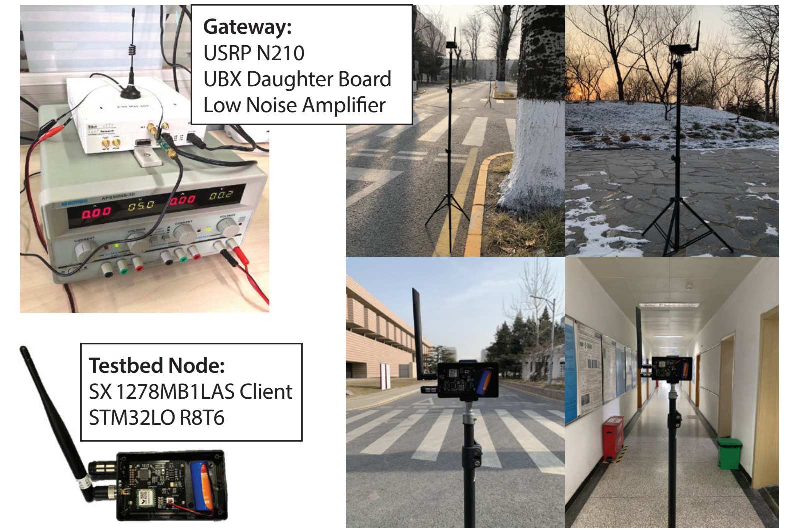
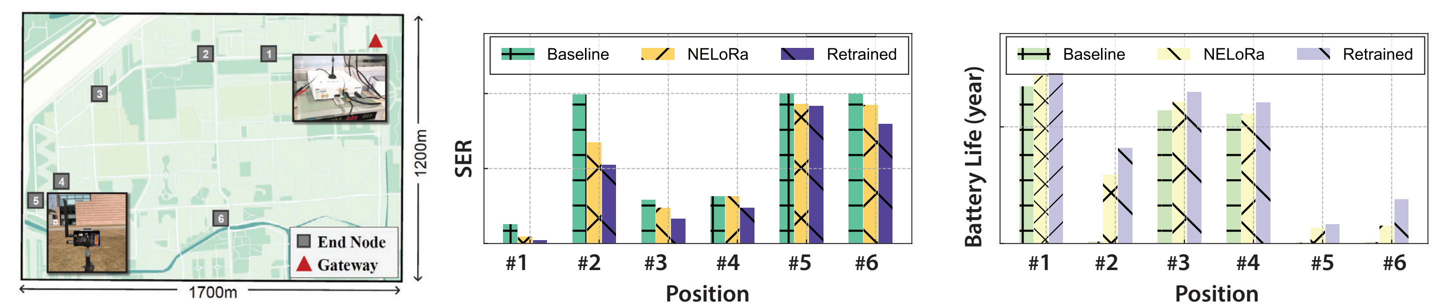


FIGURE 6. NELoRa gateway and node prototype.

WE PRESENT NELoRa, A NEURAL-ENHANCED DEMODULATION METHOD THAT ACHIEVES ULTRA-LOW SNR LoRa COMMUNICATION WITH A SINGLE GATEWAY



(a) Deployment (b) SER at different locations (c) BL at different locations

FIGURE 7. Outdoor deployment and performance.

Data Augmentation. We improve the generalization of our DNN model by training it with millions of syntheses LoRa chirp symbols, which cover different SNR levels with diverse random noise patterns. Specifically, we collect each type of chirp symbol at high SNR using an indoor testbed. To achieve fine-grained SNR control, we add various Gaussian white noises with controlled amplitude on the collected I and Q traces [9-11] to generate new chirp symbols.

DNN Model Compression. We adopt the structured pruning [12] to compress the original model for efficient inference. Specifically, we calculated the L1-norm of weights in each filter of CNN and dense layer and preserved those with the largest L1-norm. Besides, we also replace the LSTM layer with the GRU layer, which is a more computation-efficient version of RNN, while achieving similar performance when the input sequence is not too long.

SYSTEM IMPLEMENTATION

We have implemented NELoRa and evaluated its performance with commercial LoRa nodes. Figure 6 illustrates the system prototype of NELoRa. Specifically, we use the USRP N210 software-defined radio (SDR) platform for capturing over-the-air LoRa signals, operating on a UBX daughter board at the 470MHz bands. The captured signal samples are then delivered to a back-end host for pre-processing and demodulation. Note that demodulation methods of NELoRa are hardware-independent, so they can be implemented on any other commercial LoRa gateways as long as the signal samples can be obtained. On the transmitter side, we use SX1278 client radio-based commodity LoRa nodes for transmitting LoRa packets.

PERFORMANCE EVALUATION

To examine the performance in the outdoor environment, we deployed our testbed on a university campus covering various land cover types (e.g., trees, buildings, roads, and pond) as illustrated in Figure 7(a). Specifically, we deployed the LoRa nodes at six different locations. Each LoRa node transmits 15 packets, each of which contains 188 chirp symbols. Besides the pre-trained DNN decoder, we use the newly collected symbols in the campus environment to fine-tune our DNN model to achieve higher performance. We use the standard dechirp as the baseline. As shown in Figure 7(b), in general, SER is increased as the distance between the gateway and the LoRa node increases. Although location 2 is near the gateway, it has a high SER due to low SNR caused by the building blockage. Compared with the baseline, the original NELoRa decreases SER by 5.51% to 31.9%, and the re-trained NELoRa further reduces the SER by 7.72% to 46.9%. We estimate the battery life of the operating LoRa node at each location. Figure 7(c) shows the battery life can be extended by 1.32 to 5.73 years with the original NELoRa. The maximum battery life gain can reach 8.06 years by using the re-trained NELoRa at location 2. The SER of location 2, 5, and 6 reach 100% using the dechirp. The gateway cannot demodulate any data from those locations even if the LoRa nodes drain out their battery. With NELoRa, the SER is lowered, and the battery life is increased significantly.

CONCLUSION AND FUTURE DIRECTIONS

Neural-enhanced LoRa demodulation is a promising way to break the SNR threshold of the standard dechirp approach. In this work, we propose NELoRa based on such neural-enhanced demodulation design, and demonstrate that the obtained SNR gains enable longer communication distance and battery lifetime in LoRa. For a detailed evaluation of NELoRa, please refer to the original paper [13]. To improve the performance of NELoRa, a DNN model optimization is needed when the length and types of chirp symbols are getting larger. Moreover, an online DNN model adaption scheme is needed to cope with the environment dynamics in real-world deployment. We leave these as our future works. ■

Chenning Li is a graduate student at the Department of Computer Science and Engineering, Michigan State University, working with Prof. Yunhao Liu and Prof. Zhichao Cao. His research focuses on wireless networks, machine learning systems, edge computing.

Hanqing Guo received his B.Eng. degree from Chongqing University of Posts and Telecommunications, China, in 2015. He is currently working toward a Ph.D. degree in the Department of Computer Science and Engineering, Michigan State University, USA. His research interests include signal processing, speech recognition, machine learning, and AI-enabled mobile systems.

Shuai Tong received his B.E. degree from the College of Computer Science, Nankai University, in 2019. He is currently a Ph.D. candidate with the School of Software, Tsinghua University. His research areas include low-power wide-area networks, Internet of Things, and mobile computing.

Zhichao Cao is currently an assistant professor in Department of Computer Science and Engineering, Michigan State University. He received his Ph.D. degree in the Department of Computer Science and Engineering of Hong Kong University of Science and Technology, and a B.E. degree in the Department of Computer Science and Technology of Tsinghua University. His research interests lie broadly in IoT systems, edge computing and mobile computing.

Mi Zhang is an associate professor and the director of the Machine Learning Systems Lab at Michigan State University. He received his Ph.D. from University of Southern California and B.S. from Peking University. Before joining MSU, he was a postdoctoral scholar at Cornell University. His research lies at the intersection of mobile/edge/IoT systems and machine learning.

Qiben Yan is an assistant professor in the Department of Computer Science and Engineering at Michigan State University. From 2015 to 2019, he was an assistant professor in the Department of Computer Science and Engineering at the University of Nebraska-Lincoln. He received his Ph.D. in Computer Science from Virginia Tech in 2014, and his M.S. and a B.S. degree in Electronic Engineering from Fudan University in Shanghai, China. His current research interests include cyber-physical systems, wireless network security and privacy, mobile and IoT security, and big data privacy.

Li Xiao received her B.S. and M.S. degrees in computer science from the Northwestern Polytechnic University, China, and her Ph.D. degree in computer science from the College of William and Mary, Virginia. She is a professor of computer science and engineering at Michigan State University. Her research interests are in the areas of distributed and networking systems, overlay systems and applications, and wireless networks.

Jiliang Wang is an associate professor in the School of Software, Tsinghua University, P.R. China. His research interests include Internet of things, LPWANs, and mobile computing. Jiliang Wang received his B.E. degree in computer science and technology from University of Science and Technology of China and his Ph.D. degree in computer science and engineering from Hong Kong University of Science and Technology, respectively.

Yunhao Liu is a professor and the dean at GIX, Tsinghua University, China. He received his B.S. degree in Automation from Tsinghua University, and an M.S. and a Ph.D. degree in Computer Science and Engineering at Michigan State University. His research interests include networking systems, mobile computing and cloud computing. He is a Fellow of ACM and IEEE.

REFERENCES

[1] Chenning Li, and Zhichao Cao. 2022. LoRa networking techniques for large-scale and long-term IoT: A down-to-top survey. *ACM Computing Survey (CSUR)* 55(3), 52.

[2] Li Liu, Yuguang Yao, Zhichao Cao, and Mi Zhang. 2021. DeepLoRa: Learning accurate path loss model for long distance links in LPWAN. *Proceedings of IEEE INFOCOM.*

[3] Jansen C. Liando, Amalinda Gamage, Agustinus W. Tengourtius, and Mo Li. 2019. Known and unknown facts of LoRa: Experiences from a large-scale measurement study. *ACM Transactions on Sensor Networks.*

[4] Rashad Eletreby, Diana Zhang, Swarun Kumar, and Osman Yağan. 2017. Empowering low-power wide area networks in urban settings. *Proceedings of ACM SIGCOMM.*

[5] Adwait Dongare, Revathy Narayanan, Akshay Gadre, Anh Luong, Artur Balanuta, Swarun Kumar, Bob Iannucci, and Anthony Rowe. 2018. Charm: Exploiting geographical diversity through coherent combining in low-power wide-area networks. *Proceedings of ACM/IEEE IPSN.*

[6] Artur Balanuta, Nuno Pereira, Swarun Kumar, and Anthony Rowe. 2020. A cloud-optimized link layer for low-power wide-area networks. *Proceedings of ACM MobiSys.*

[7] Akshay Gadre, Revathy Narayanan, Anh Luong, Anthony Rowe, Bob Iannucci, and Swarun Kumar. 2020. Frequency configuration for low-power wide-area networks in a heartbeat. *Proceedings of USENIX NSDI.*

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. *Proceedings of the IEEE CVPR.*

[9] Shuai Tong, Jiliang Wang, and Yunhao Liu. 2020. Combating packet collisions using non-stationary signal scaling in LPWANs. *Proceedings of ACM MobiSys.*

[10] Shuai Tong, Zhenqiang Xu, and Jiliang Wang. 2020. CoLoRa: Enabling multi-packet reception in LoRa. *Proceedings of IEEE INFOCOM.*

[11] Chenning Li, Xiuzhen Guo, Longfei Shangguan, Zhichao Cao, and Kyle Jamieson. 2022. CurvingLoRa to boost LoRa network throughput via concurrent transmission. *Proceedings of USENIX NSDI.*

[12] Biyi Fang, Xiao Zeng, and Mi Zhang. 2018. NestDNN: Resource-aware multi-tenant on-device deep learning for continuous mobile vision. *Proceedings of ACM MobiCom.*

[13] Chenning Li, Hanqing Guo, Shuai Tong, Xiao Zeng, Zhichao Cao, Mi Zhang, Qiben Yan, Li Xiao, Jiliang Wang, and Yunhao Liu. 2021. NELoRa: Towards ultra-low SNR LoRa communication with neural-enhanced demodulation. *Proceedings of ACM Sensys.*