Neural Enhanced Underwater SOS Detection

Qiang Yang[†‡], Yuanqing Zheng[†]
[†]*The Hong Kong Polytechnic University, Hong Kong, China*
[‡]*University of Cambridge, Cambridge, United Kingdom*
qiang.yang@connect.polyu.hk, csyqzheng@comp.polyu.edu.hk

***Abstract*—Every day, one person loses his life due to drowning in swimming pools, even with professional lifeguards present. Contrary to what the public might assume, drowning swimmers can hardly splash or yell for help. This life-threatening situation calls for a robust SOS channel between the swimmers and the lifeguards. This paper proposes Neusos, a neural-enhanced underwater SOS communication system based on commercial wearable devices and low-cost hydrophones deployed in the swimming pool. Specifically, we repurpose popular wearable devices (*e.g.*, smartwatches) as SOS transmitters, which can send a distress signal when the user is in an emergency. In response, an underwater hydrophone in the swimming pool can detect SOS signals and make alerts immediately to facilitate a timely rescue. The main technical challenge lies in reliably detecting weak SOS signals in non-stationary underwater scenarios. To achieve so, we thoroughly characterize the properties of underwater channels and examine the limitations of the traditional correlation-based signal detection method in underwater communication scenarios. Based on our empirical findings, we developed a robust SOS detection method enhanced with deep learning. By fully embedding signal characteristics into networks, Neusos outperforms traditional signal processing-based underwater SOS detection methods. In particular, our experiments in a real swimming pool show that Neusos can detect SOS signals with a detection rate of 98.2% under various underwater conditions. Given the increasing popularity of smartwatches among swimmers, our system holds immense potential to enhance their safety in swimming pools.**

***Index Terms*—Underwater communication, Signal detection, Wearable devices, Machine learning**

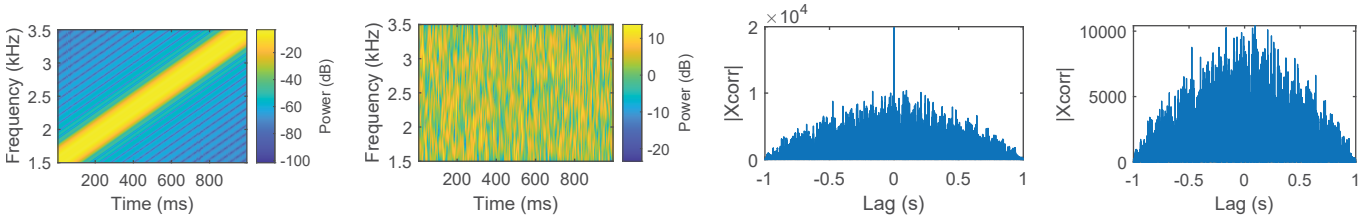
I. INTRODUCTION

Drowning remains a serious public health threat, with an average of 390 people losing their lives in swimming pools every year in the USA alone [1]. Surprisingly, even with professional lifeguards present, roughly one-third of drowning deaths still occur [2]. This is partly due to a common misconception among the public that drowning swimmers would wave their hands and call for help, while lifeguards could notice and intervene in a timely manner [3]. In practice, however, some drowning swimmers in panic often struggle to reach the surface and cannot call for help, while others could experience sudden cramps or severe discomfort and sink underwater quickly and quietly [4]. Moreover, lifeguards can be distracted and unable to stay alert and continuously monitor every individual in a large swimming pool [3]. To prevent these tragedies, it is crucial to equip swimmers with an effective method to seek timely help in emergencies. By doing so, lifeguards can be promptly alerted and respond swiftly to rescue those in distress.

To provide timely assistance, international rescue standards require lifeguards to quickly identify drowning swimmers within 10 seconds [5]. However, the average detection time of a drowning person is unfortunately around 69 seconds [6]. To address this critical issue, several existing solutions rely on multiple cameras to monitor swimmers and alert lifeguards when a swimmer sinks to the bottom of a swimming pool or remains still for an abnormally long period of time [7–10]. Nevertheless, these solutions face significant practical challenges, including poor performance in weak lighting conditions, swimmer occlusions, and highly dynamic backgrounds [10], which could lead to delays in alerting lifeguards or even false negatives. As a result, relying solely on vision-based systems, many swimmers in distress can miss the critical window for rescue. Therefore, we believe that a more effective and responsive SOS communication system could complement existing drowning detection and rescue systems.

The prevalence of waterproof wearable devices, such as smartwatches, has become increasingly popular among swimmers for exercise assessment and health monitoring. This prompts us to ask: *can we harness these wearable devices as underwater SOS transmitters?* We consider acoustics as the SOS signal medium due to the significant attenuation of RF signals in water [11], and the susceptibility of light to occlusion and limited view field. Thus, we give an affirmative answer by designing and implementing Neusos, an acoustic underwater SOS transmission and detection system. Neusos can offer an API that integrates with existing wearable-based drowning detection systems [12–14], enabling the transmission of SOS signals as soon as a swimmer is detected in a state of drowning. In addition, a swimmer can also manually activate the SOS transmission by pressing a button on the smartwatch in the event of an emergency like cramps. The smartwatch will immediately broadcast acoustic SOS signals, which will be captured by underwater hydrophones deployed in the swimming pool. Receiving the SOS signals, Neusos will promptly alert lifeguards on duty or the emergency medical center through alarm sounds and messages, facilitating a swift and effective rescue.

Transforming Neusos from a concept into a practical system involves several technical challenges. First, lightweight smartwatches are limited by their transmission power due to small form factors. Substantial attenuation of underwater signals leads to a low signal-to-noise ratio (SNR) [11]. Second, underwater environments are characterized by dynamic noise and varying channel conditions over time [11]. Third, underwater



(a) Chirp signals (b) A chirp in noise (c) Xcorr with the noisy chirp (d) Xcorr with Gaussian noise

Fig. 1: Cross-correlation-based signal detection with Gaussian noise.

noise, such as water flow and bubbles, combined with channel fading effects, introduces fluctuations in received signals. Consequently, traditional threshold-based cross-correlation methods, which rely on detecting peaks above a hand-crafted threshold, are not suitable for practical implementation [15]. In specific, setting the threshold too low makes the detection system over-sensitive and susceptible to noise, while setting it too high may miss SOS signals, which can have severe consequences in emergencies. Furthermore, swimming pool environments are prone to various sources of interference, such as pump operation and intense splashing, which generate wideband disturbances that severely hinder signal detection.

To achieve reliable SOS detection in low-SNR underwater scenarios, we utilize acoustic chirps as SOS signals, which are robust to channel selective fading and Doppler effects. Moreover, we develop a short-time dechirp transform to enhance the SOS signal. By doing so, the linear pattern of the SOS chirp becomes more prominent in the frequency domain, making it easier to distinguish it from background noise. Furthermore, instead of detecting SOS signals with traditional correlation methods, we build an explicit deep-learning model incorporating the prior knowledge of chirps to learn the distinctive linear pattern of SOS signals in the spectrogram. This enables automatic detection at low SNR and eliminates the requirement for a fixed threshold. Benefiting from this signal characteristic-aware design, the model can achieve good generalization performance by training only on synthesized data, which alleviates the workload of data collection. Upon detection of the SOS signal, Neusos will immediately alert lifeguards or the emergency center by alarms to facilitate faster rescue operations. By integrating these methods, Neusos balances the false positive and false negative rates, making underwater acoustic SOS detection more practical and robust. In summary, this paper makes the following contributions:

- We propose Neusos, a robust underwater SOS detection system that provides an underwater communication channel between swimmers and lifeguards, effectively making alerts in early response time and enhancing the safety of swimmers.
- We develop an explicit signal-aware deep learning model to effectively capture the features of SOS signals, which balances the system reliability and sensitivity, thereby improving the practicality of the SOS detection system.
- Real-world evaluation in a swimming pool shows that Neusos can achieve a detection rate of 98.2%, outperforming the traditional correlation-based baseline by 9.2% with the same false positive rate.

The rest of this paper is organized as follows. Sec. II introduces the challenges and opportunities of SOS detection in underwater scenarios. We provide a comprehensive description of our proposed solutions and the design of Neusos in Sec. III. The implementation and evaluation of Neusos are detailed in Sec. IV. We give a literature review in Sec. V, and conclude this paper in Sec. VI.

II. Problem and Motivation

This section first examines the traditional cross-correlation based detection method widely adopted in previous works in the context of underwater communication [11]. We then summarize its pros and cons in our target problem of underwater SOS detection, and motivate our design of neural enhanced detection.

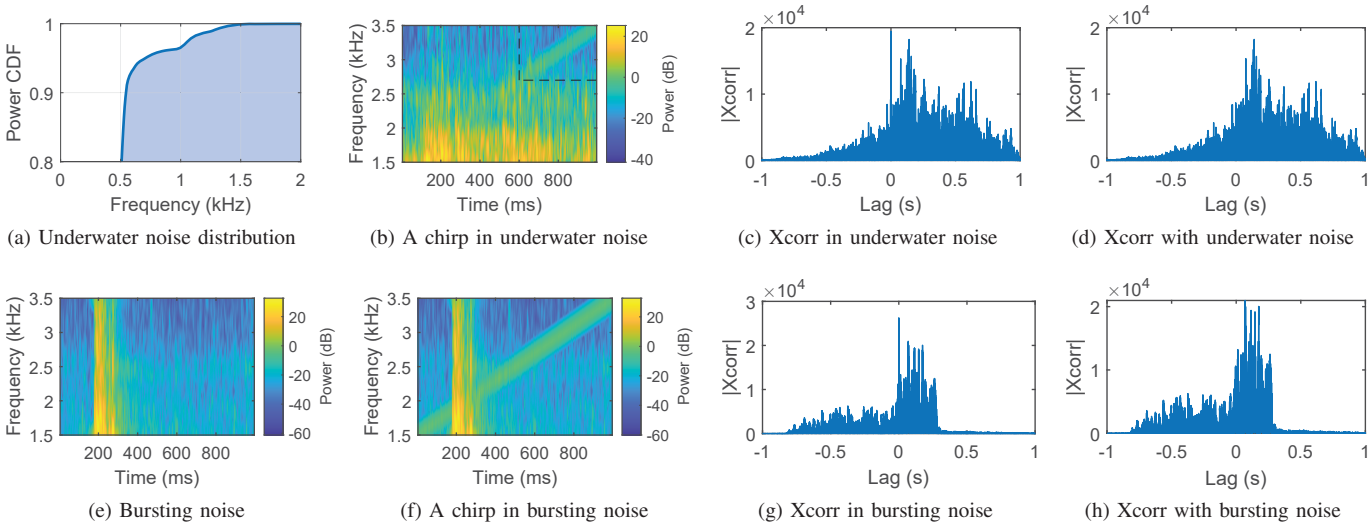
A. Signal Detection with Cross-correlation

Weak signal detection in the presence of complex background noise has been extensively explored in wireless communication systems. For example, weak signals can be detected with matched filter [16] which essentially performs cross-correlation (xcorr). This technique involves comparing the received signal with the known transmitted signal and assessing their similarity [17]. Due to the excellent correlation property, chirp signals are widely used for underwater communication [18–20]. As depicted in Fig. 1(a), chirp signals spread across a frequency band with a linearly increasing frequency and offer good correlation properties. Chirp signals are also robust to frequency-selective fading and Doppler shift [21].

The cross-correlation $r_{xy}(n)$ between the known signal $x(n)$ and the received signal $y(n)$ is defined as:

$$r_{xy}(n) = \sum_{\ell=0}^{N-1} y[n+\ell]s[\ell] \tag{1}$$

where ℓ denotes the sample shift. When the received signal $y(n)$ aligns with $x(n)$ at time n', the cross-correlation produces a maximum peak with height of $\sqrt{BT}$ in theory [15], where B is the bandwidth and T is the chirp duration. Fig. 1(b) illustrates a chirp signal overwhelmed by strong Gaussian noise with an SNR of -20dB. After performing



(a) Underwater noise distribution (b) A chirp in underwater noise (c) Xcorr in underwater noise (d) Xcorr with underwater noise

(e) Bursting noise (f) A chirp in bursting noise (g) Xcorr in bursting noise (h) Xcorr with bursting noise

Fig. 2: Cross-correlation-based signal detection with underwater noise (the upper row) and bursting noise (the lower row).

cross-correlation (Fig. 1(c)), a peak stands out from the noise floor, indicating the presence of a chirp signal.

B. Challenges for Underwater Channels

In previous works, the wireless channel is commonly modelled as an Additive White Gaussian Noise (AWGN) channel for simplicity, where the noise follows a Gaussian distribution [22]. In such cases, as illustrated in Fig. 1(c) and Fig. 1(d), a prominent peak appears when correlating the received signal with the known signal. In contrast, when correlating the chirp with Gaussian noise, there is no peak clearly standing out from the noise floor. As a result, the signal can be detected by examining if a peak exceeds a preset threshold $u = \bar{r} + \alpha\sigma(r)$, where $\bar{r}$ and $\sigma(r)$ are the mean and standard variation of $r(n)$. α is the threshold parameter. However, unlike the assumption of the AWGN channel, the unique properties of underwater channels introduce significant challenges for SOS detection.

Unbalanced noise distribution. In underwear scenarios, the noise distribution is unbalanced, with most of the noise concentrated in the low frequency band. Fig. 2(a) illustrates the power Cumulative Distribution Function (CDF) of underwater noise in a typical swimming pool, revealing that 95% of underwater noise, such as water flow and air bubbles, is distributed below 1 kHz, and 99% of noise is less than 1.5 kHz. Moreover, underwater noise is much stronger than the weak SOS chirp signals. As depicted in Fig. 2(b), the residue underwater noise remains more significant than the chirp signals at frequencies higher than 1.5 kHz. Although one might consider using chirp signals in higher frequency bands less affected by noise, the frequency band above 3.5 kHz experiences severe underwater attenuation [23]. This limitation confines the SOS chirp signals to the frequency band of [1.5, 3.5] kHz [23]. As such, with the same SNR level (*i.e.*, -20 dB), the signal peaks resulting from cross-correlation will be heavily influenced by noise interference, making it challenging to distinguish the SOS signals from noise, as illustrated in Fig.2(c). Moreover, strong interference and noise can also generate high correlation peaks, possibly leading to false positives (Fig. 2(d)).

Low SNR. The signal strength of received chirps can become significantly weakened due to the joint effects of low transmission power and severe channel attenuation. As a result, the correlation peak generated by cross-correlation can be very weak and overwhelmed by noise. As mentioned in Sec. II-A, the height of the signal peak after cross-correlation is directly related to the bandwidth and chirp duration. In the context of underwater scenarios, the bandwidth is limited to 2 kHz due to the unbalanced noise distribution and severe attenuation. Enlarging the chirp duration is one way to increase the SNR. However, this comes at the cost of a longer detection time, which is critical for the timely detection and rescue of drowning swimmers. To balance efficiency and effectiveness, we have set the chirp duration to one second. This duration enhances the SNR while maintaining a reasonable detection time in real-world rescue scenarios.

Non-stationary channel state. The underwater channel is subject to dynamic water flow, rendering it non-stationary and possessing a very short coherent time [11]. Consequently, the SNR of received signals fluctuates over time, making the use of a pre-defined threshold impractical. Moreover, cross-correlation requires a precise alignment between the known and received signals. However, the presence of strong noise can heavily corrupt a fragment of the chirp signal and render the chirp signals undetected as illustrated in Fig. 2(b).

Bursting noise. The underwater environment, such as swimming pools, is characterized by the presence of strong noise sources, including pump operations and water splashing, which can cause a sudden burst of energy within a very short time. As shown in Fig. 2(e), the bursting noise spreads across the spectrum with much higher power than the chirp signal. Fig. 2(f) illustrates a chirp signal that interfered with

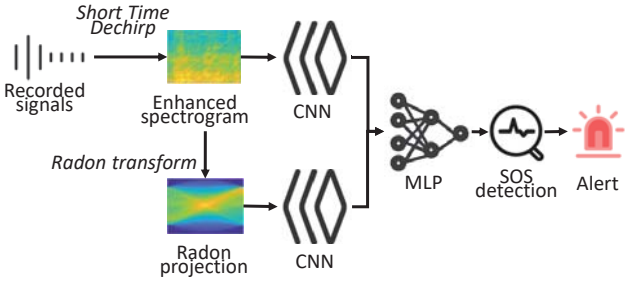
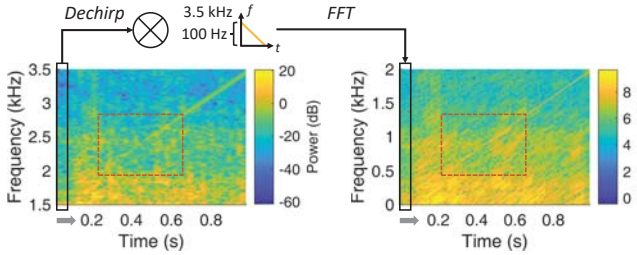


Fig. 3: Workflow of Neusos.



(a) STFT spectrogram. (b) Short time dechirp spectrogram.

Fig. 4: SOS pattern enhancement.

the bursting noise at an SNR of -20 dB. Similarly to strong underwater noise, the signal peak in the cross-correlation between the chirp signal and the received signal can hardly be distinguished from the noise (Fig. 2(g)). Figure 2(h) shows the correlation result between the bursting noise and the clean chirp signal, and we can observe that the powerful bursting noise can cause false alarms if a preset threshold is adopted.

C. Opportunities for Underwater SOS Detection

The challenges discussed above make traditional cross-correlation unsuitable for reliable underwater SOS detection. However, despite the powerful nature of underwater noise, *it does not completely overwhelm the entire chirp signal* as illustrated in Fig. 2(b) and Fig. 2(f). This observation provides us with an opportunity to detect weak SOS chirp signals. For example, as illustrated in Fig. 2(b), a fragment of the SOS chirp remains distinguishable from the noise, particularly in the high-frequency band (within the dashed box). This is due to the dynamic nature of underwater noise and the short duration of bursting noise events. Consequently, we can exploit these dynamic underwater characteristics as an opportunity rather than a hindrance to detect weak SOS signals. As such, rather than detecting the signal with cross-correlation, we cast this task as *line pattern detection in the spectrogram* of the received signal, inherently corresponding to *chirp signal detection*. This approach aligns well with the prominent strength of deep learning methods, which have demonstrated their efficacy in feature extraction and pattern recognition in computer vision tasks.

Compared with traditional correlation-based detection methods, deep learning-based SOS detection can potentially enjoy several benefits. First, the deep learning model can discern subtle patterns and distinguish chirp signals with a very low SNR, which makes it robust to unbalanced and dynamic underwater noise. Second, deep learning methods can detect the fragment of a chirp by examining the linear pattern as long as it appears in the spectrogram, which reduces the crucial rescue time. Third, deep learning methods can adapt to different SNR levels and automatically examine the presence of SOS chirps. This eliminates the need for the predefined thresholds α, enhancing the reliability of SOS detection while minimizing unnecessary false alarms.

However, translating these potential opportunities into concrete gains entails two primary challenges. Chirp fragments in the received signal may be significantly weaker than the surrounding strong noise interference, making it difficult for models to accurately extract the line pattern from the spectrogram. Second, training an effective deep-learning model requires a large dataset of SOS chirp signals in various underwater conditions. However, collecting such a dataset can be labor-intensive and time-consuming. To address these challenges, we propose several techniques to enhance linear pattern extraction and detection in the spectrogram. In addition, we introduce the chirp characteristic as a prior knowledge into the deep learning model to ensure the network can utilize the linear pattern to perform signal detection. Finally, we perform data augmentation to synthesize a large amount of training data based on the noise data collected in swimming pools to bootstrap model training and enhance their performance in chirp signal detection.

III. System Design

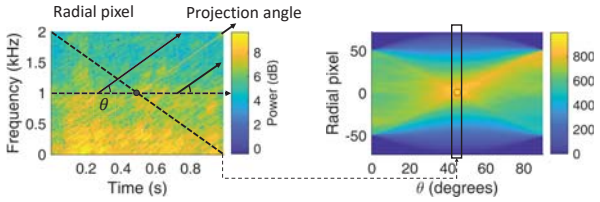
In this section, we introduce Neusos, a neural-enhanced underwater SOS system that utilizes deep learning to overcome practical challenges and achieve reliable SOS signal detection. In the following, we first present the system overview of Neusos before delving into the details of key technical components.

A. Neusos Overview

Figure 3 illustrates the system overview of Neusos. The process starts with short-time dechirp on the received audio, which enhances the spectrogram of SOS signals. The enhanced spectrogram is then fed into a Convolutional Neural Network (CNN) to extract high-level features. Additionally, a Radon transform is applied to extract chirp shape-related features from the spectrogram, which are forwarded to another CNN. The outputs from the dual-CNN branch are concatenated to form the input to a Multi-Layer Perception (MLP), which detects the presence of SOS signals as a binary classification task. We synthesize massive training data with variations in time shifts, SNR levels, and noise types to train the deep learning model, so that the trained model can adapt to diverse channel conditions while minimizing the workload involved in data collection.

B. SOS Signal Enhancement

As depicted in Fig. 4(a), the SOS chirp signal appears very weak when submerged in underwater noise, posing a challenge



(a) Original spectrogram. (b) Radon transform projection.

Fig. 5: Radon transform.

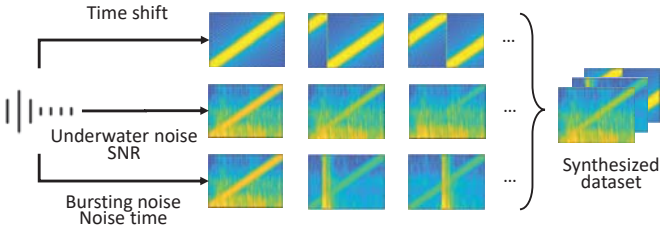


Fig. 6: SOS signal synthesis process.

for deep learning models to learn the linear chirp patterns in the spectrogram. This is primarily because the spectrogram is calculated using Short Time Fourier Transform (STFT). If a long window is used for STFT, the frequency resolution is higher, but the power of the chirp signals gets dispersed into multiple frequency bins. Consequently, the chirp pattern in the spectrogram appears blurry and is easily confused with the noise. Conversely, using a smaller window for STFT reduces the frequency resolution, leading to fewer pattern observations in the frequency domain.

To enhance the clarity of the chirp pattern, we propose a short-time dechirp transform (STDT) approach to enhance the original linear chirp pattern in the spectrogram. Similar to STFT, STDT operates on a sliding window of the received audio. For each window, the dechirp process is performed, which concentrates the signal energy into a single tone [24]. As a result, the chirp pattern becomes more discernible and focused, appearing stronger and thinner in the spectrogram. As shown in Fig. 4(b) (especially in the red box), the enhanced spectrogram exhibits a clearer and more pronounced chirp pattern compared to the original STFT spectrogram, although they have the same parameters (*i.e.*, window size, overlap, and FFT point). This improvement in clarity enables deep learning models to track and identify the linear chirp pattern more effectively, facilitating reliable SOS signal detection.

The dechirp process is a frequency domain equivalent of cross-correlation and has distinct advantages over the latter due to its insensitivity to different chirp time shifts [24]. In other words, the dechirp can be performed effectively even when the sliding window is not aligned with the chirp. Mathematically, a chirp with a time shift can be represented as follows:

$$S(f', t) = C(t) \cdot e^{j2\pi f' t} = e^{j2\pi(f_0 + \frac{k}{2}t)t} \cdot e^{j2\pi f' t} \quad (2)$$

where $C(t) = e^{j2\pi(f_0+\frac{k}{2}t)t}$ denotes the base chirp, starting with a frequency of f_0. $k = \frac{B}{T}$ is the increasing rate of the frequency. $S(f', t)$ is a time-shifted chirp, starting from $f_0 + f'$. When performing dechirp, we multiply the received chirp $S(f', t)$ by the conjugate of the base chirp, denoted by $C^{-1}(t)$:

$$S(f', t) \cdot C^{-1}(t) = C(t) \cdot e^{j2\pi f' t} \cdot C^{-1}(t) = e^{j2\pi f' t} \quad (3)$$

Consequently, we can observe that all samples' power will be accumulated at the initial frequency f' by performing an FFT operation, irrespective of the time shift.

In our setting, the sampling rate Fs is 48 kHz. We set the size of the sliding window to 2400 with an overlap of 80%. The starting frequency f_0 and the duration t of the base chirp used for dechirp are 3.4 kHz and $\frac{2400}{48000} = 50ms$ (corresponding to a bandwidth of 100 Hz), respectively. The FFT point is equal to the window size, resulting in an enhanced spectrogram of the size $\mathbb{R}^{100\times100}$. This STDT approach enhances the visual representation of the chirp pattern, enabling deep-learning models to effectively learn and identify the linear chirp pattern in the spectrogram.

C. SOS Signal Transformation

To further assist models in detecting the linear chirp pattern, we perform the Radon transform on the enhanced spectrogram to explicitly extract the linear features. The Radon transform is commonly used for line detection in images by projecting the image from Cartesian coordinates to the (θ, ρ) parameter space, where θ represents the projection angle, and ρ is the radial axis orthogonal to the projection direction [25]. As shown in Fig. 5(a), the Radon transform is the projection of the image intensity along a radial line oriented at a specific angle θ. Figure 5(b) shows the Radon transform of the enhanced spectrogram with a size of $\mathbb{R}^{145\times91}$. We can observe that the linear chirp pattern in the enhanced spectrogram is transformed into the brightest point $(45°, 0)$ at the red circle in the Radon space. In this way, the Radon projection serves as a pattern indicator, guiding the deep learning model to detect the presence of SOS chirps.

D. SOS Signal Synthesis

Training an effective deep learning model requires a large amount of diverse data to generalize to real-world scenarios. However, collecting such a diverse dataset can be labor-intensive and time-consuming. To address this challenge, Neusos utilizes a large amount of synthesized data with signal varieties and performs pattern-aware model training to ensure system generalization. Figure 6 illustrates the data synthesis process. To create the synthesized data set, we collected a variety of underwater noises and swimming interference by placing hydrophones in swimming pools. This data augmentation process does not require SOS signal transmission or user participation. Subsequently, we can synthesize SOS chirps with different time shifts between $[0s, 1s]$, add different underwater noises at various SNR levels between $[-20dB, 0dB]$, and randomly add the chirp with or without bursting noise at different time instances. In total, we synthesize a dataset with 20,000 SOS chirps for model training.

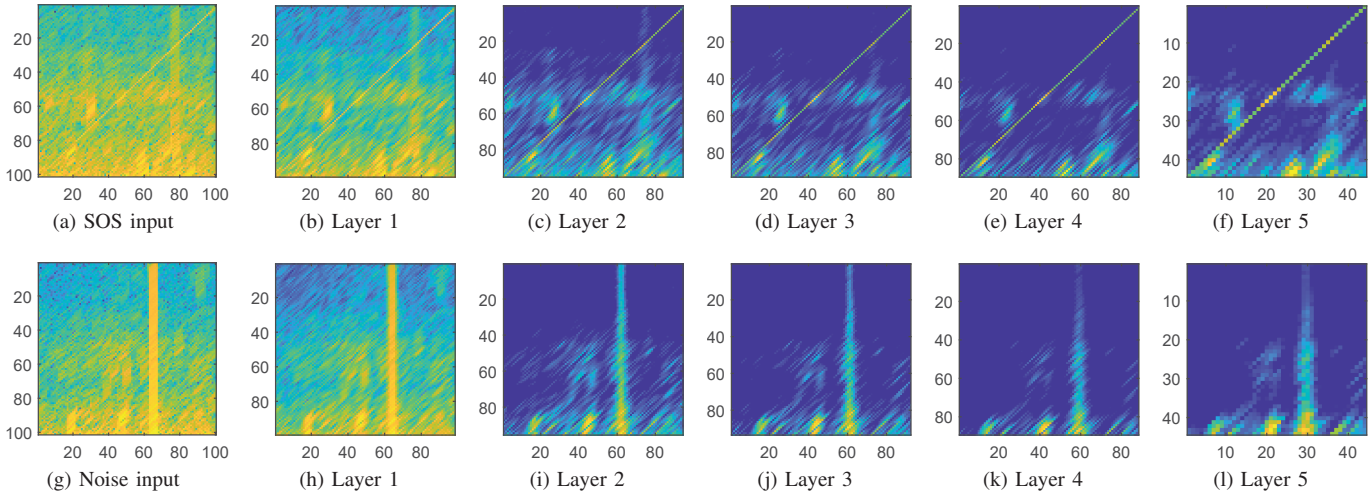


Fig. 7: Feature map visualization of the SOS chirp (the upper row) and noise (the lower row) from different conv layers.

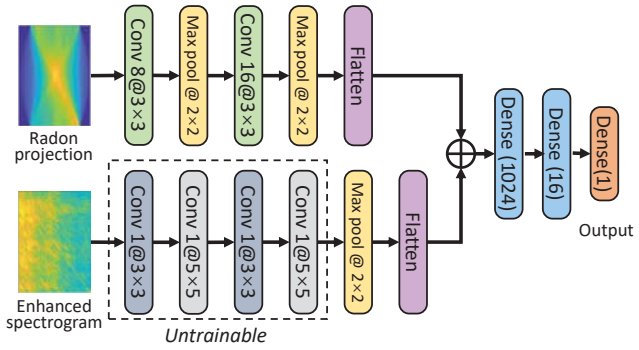


Fig. 8: Model architecture of Neusos.

E. SOS Signal Detection

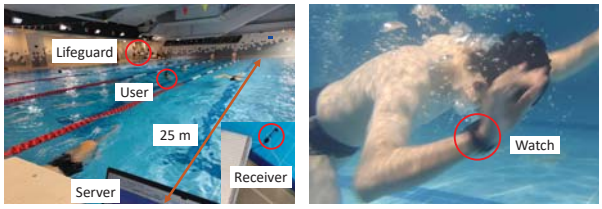
As illustrated in Figure 8, a dual-CNN backbone is used to build the deep-learning model. The enhanced spectrogram of received signals $\boldsymbol{s} \in \mathbb{R}^{100\times100}$ is fed into a four-layer CNN branch to extract the high-level chirp features. To ensure that CNN can effectively capture linear patterns in the spectrogram, we explicitly provide the network with prior knowledge about the expected patterns. Specifically, we initialize the kernels of these layers with identity matrices, which naturally correspond to the linear chirp pattern[1]. The 3×3 and 5×5 kernels are used to capture the fine-grained and coarse-grained linear patterns, which are repeated two times. After that, we freeze these layers to force the network to magnify the chirp pattern and suppress the impact of noise in the spectrogram. A pooling layer is used to reduce the feature size, and the following dense layers are trainable to learn the extracted chirp patterns to perform classification. As mentioned in Sec. III-C, the Radon transform of the spectrogram also explicitly contains the linear pattern. Therefore, we adopt a two-layer CNN to encode the Radon transform and concatenate its output with the high-level representations from the bottom-branch encoder. The concatenated features are then passed through an MLP consisting of two dense layers to predict whether the SOS chirp exists or not (a binary result). A batch normalization layer is attached to each convolution layer in this network. The loss function used for training the model is binary cross-entropy.

Figure 7 provides insights into the feature maps obtained after applying different convolution layers to the spectrogram input. It demonstrates that the identity kernels employed in the convolution layers can effectively capture and enhance the chirp patterns while significantly suppressing the noise in the spectrogram. This pattern-aware network design is crucial in distinguishing SOS chirps from interference and noise, allowing the model to generalize to a variety of unseen conditions and achieve robust SOS detection. As shown in Fig. 7(g), the received signal contains high-energy bursting noise, which would cause a false positive peak with cross-correlation method (Sec. II-B). After convolution with identity kernels, the bursting noise is effectively eliminated, since it does not follow a diagonal pattern in the spectrogram. This demonstrates the power of the pattern-aware convolution layers in suppressing unwanted noise while enhancing the linear chirp patterns. While theoretically stacking multiple convolution layers could further enhance the chirp pattern detection and extraction, our experiments have not shown substantial performance improvement with an increased number of layers. In practice, we use four convolution layers to strike a balance between model size and performance.

IV. Implementation and Evaluation

Implementation. We implemented Neusos and conducted experiments using two commercial smartwatches: Huawei Watch 3 and OPPO Watch 3 Pro. The SOS chirps will be transmitted repeatedly once activated. We attempted to imple-

[1]The chirp linearly increasing with the frequency shows a back-diagonal shape in visualization, but the frequency order is reversed in the actual spectrogram matrix, which is a diagonal pattern instead.



(a) Swimming pool (b) Swimmer

Fig. 9: Experiment setting.

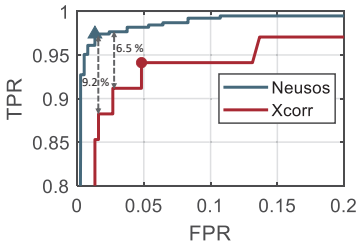


Fig. 10: ROC curve.

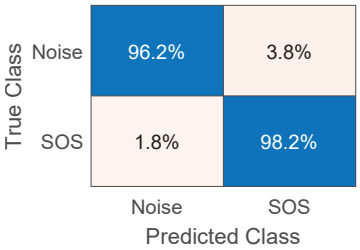


Fig. 11: Confusion matrix.

ment Neusos on an Apple Watch 6, but encountered an issue where audio playback was automatically interrupted when the watch was submerged in water. For our experimental setup, we deployed a Shuimi SN005 hydrophone (28 USD per unit) around a swimming pool to receive SOS signals transmitted by the smartwatch. The underwater receiver was connected to Raspberry Pi devices, which forwarded the collected data to a PC for further post-processing. The deep learning model is implemented with TensorFlow running on this PC. With an average processing time of approximately 60.7 ms per detection window, Neusos achieved a maximum detection rate of 16 Hz, which is sufficient for timely SOS signal detection.

Experiment setting. As shown in Fig. 9(a), the experiments were conducted in a 6-lane, $25m \times 15m$ swimming pool. A user wears smartwatches and simulates drowning movements, initiating SOS signals at different locations underwater (Fig. 9(b)). Several swimmers swim freely nearby. Throughout the experiments, professional lifeguards closely monitored the process, and the study was approved by the university authority. We evaluated the SOS detection performance of Neusos under various conditions, including different distances, depths, orientations, and potential interference scenarios.

Evaluation Metric. We evaluate SOS detection performance of Neusos with the metrics of $recall = \frac{TP}{TP+FN}$ and $specificity = \frac{TN}{TN+FP}$. TP, FN, TN, and FP are true positives, false negatives, true negatives, and false positives, respectively. $Recall$ is also known as $Sensitivity$, which indicates detection ability of SOS signals. $Specificity$ quantifies detection ability of noise. For both metrics, the higher, the better. We also use the Receiver Operating Characteristic (ROC) curve to illustrate the trade-off between these two metrics [26]. The AUC (Area Under the ROC Curve) is also used to quantify the overall performance of our binary classifier, where a higher AUC value indicates better performance [26].

A. Overall performance

We first compare the overall performance between Neusos and cross-correlation based detection method. Figure 10 shows the ROC curve of these two methods, plotted with the true positive rate (TPR) on the y-axis and the false positive rate (FPR) on the x-axis. The data for testing is collected in the swimming pool and does not overlap with the training data. The ROC curve of cross-correlation can be plotted by varying the detection threshold α from 3 to 10 as introduced in Sec. II-B. We can find the Optimal Operating Point (OOP) of the two methods at (0.013, 0.974) and (0.048, 0.941). This result indicates that Neusos outperforms cross-correlation by 3.3% in detection recall with only a false positive rate of one-third [26]. As shown in the dashed lines, with the same FPR of 2%, Neusos has a 6.5% detection recall higher than cross-correlation. Moreover, Neusos outperforms cross-correlation by 9.2% with the same FPR of 1%. Overall, the AUC of Neusos is 0.995, which is 1. 2% higher than the cross-correlation (0.983).

For an SOS detection system, the mission-critical nature prefers more reliable SOS detection at the cost of slightly increased false positive rates. Therefore, we set a higher penalty for misclassifying a positive class (SOS) as a negative class (noise) when finding the OOP for Neusos to reduce the missing rate below 2%. Figure 11 shows the classification confusion matrix in this case. We can see that Neusos has a detection recall and specificity of 98.2% and 96.2%, respectively. The classification accuracy is 97.2%, which is efficient for SOS detection in a continuous manner.

B. Impact of Distance

We conducted the experiment with varying distances ranging from $5m$ to $29m$ between the receiver and the transmitter in the swimming pool. The total number of chirps is about 1200. As shown in Fig. 12, the performance of Neusos at $5m$ is 100%, but decreases slightly at a distance greater than 15m. This is because the chirp energy attenuates fast with the increasing distance, making it challenging to identify the chirp pattern in the spectrogram. The result shows that Neusos can still achieve 97.3% detection recall at the farthest distance in the swimming pool (*i.e.*, the diagonal line), with the capability to detect most SOS chirp signals. We note that only a single microphone is used in our evaluation. To ensure the full coverage of a swimming pool, we can deploy a few more underwater microphones. We leave the deployment planning of multiple underwater microphones for future work.

C. Impact of SNR

To evaluate the performance of Neusos at different levels of SNR, we conducted a trace-based experiment with data collected at a distance of $1m$ from the transmitter in the swimming pool and then synthesized the SOS chirps with collected underwater noise with SNR from -14 dB to -24 dB. As shown in Fig. 13, the performance of Neusos gradually decreases along with the SNR. The detection accuracy is 100% at -14 dB and decreases to 93.5% at -20 dB. A performance drop of 19.5% occurs when the SNR reduces to -22 dB.

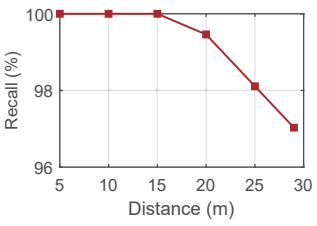


Fig. 12: Distance impact.

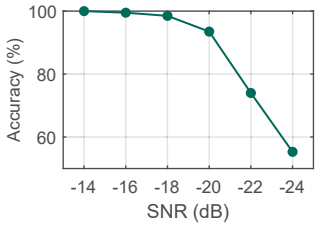


Fig. 13: SNR impact.

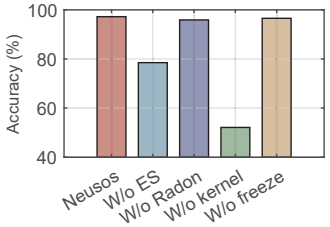


Fig. 14: Ablation study.

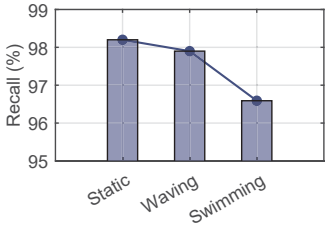


Fig. 15: Movement impact.

This result aligns with our expectation since the noise may overwhelm the chirp pattern when SNR is very slow. Another reason may be that the SNR of these data is even lower than the lowest SNR included in our training data. When the SNR decreases to -24 dB, the model nearly deteriorates to a random guess and only has an accuracy of 55.3%. Including more low-SNR data in model training may improve the system performance, but it is also limited by the visibility of chirp patterns in the spectrogram.

D. Ablation Study

We conducted ablation studies to evaluate the contributions of different components of Neusos. Figure 14 illustrates the evaluation result. SOS detection accuracy decreases to 78.5% if we train Neusos on the original spectrogram directly (w/o ES). This is because the chirp energy in each short-time window spreads into multiple frequency bins in the STFT spectrogram, which is hard to capture and magnified by the identity kernel. By contrast, the STDT-enhanced spectrogram concentrates the chirp energy into a single frequency bin, which aligns with the small pixel-level identity kernels. Without Radon projection, the performance slightly reduces by 1.3%, which indicates that Radon projection can also encode the chirp patterns and help boost the accuracy. We also found that when training networks with enhanced spectrograms and Radon projection, the performance will converge four epochs earlier than using the former alone. This observation confirms the guidance effect of the Radon projection.

The system performs worst with a random kernel initialization. The kernel parameters are trainable. The performance is close to a random guess with an accuracy of 52.1%. Moreover, the network cannot converge, although we have tried different hyper-parameters and training strategies. We suspect this is because it is hard for the network to accurately find the pattern formed by a series of pixels in a 100×100 image. To verify this, we initialize the kernel with the identity matrix and also unfreeze all convolution layers. As shown in Figure 14 (w/o freeze), the network converges and achieves slightly lower (0. 7%) but comparable performance. This result shows that identity kernel parameters are the key enabler of a network to detect linear chirp patterns and achieve robust SOS detection in various channel conditions. We did not observe a performance gain using trainable kernels and even noticed that the performance fluctuates occasionally, so we froze the convolution layers to reduce the training workload and obtain a stable performance.

E. Impact of Movement

Drowning swimmers struggling in the water will inevitably introduce smartwatch movement, which may distort dechirp patterns due to the Doppler effect. We conducted the experiment to evaluate Neusos performance under different kinds of movements. As shown in Fig. 15, the detection recall is 97.9% where a swimmer waves his hands intensely and arbitrarily to simulate drowning conditions, which is slightly lower but comparable to the static cases (98.2%). The reason behind this is the low center frequency of the SOS signals. The frequency resolution of the enhanced spectrogram is $\Delta f = \frac{B}{100} = \pm 10\ Hz$. According to the Doppler theory, we can infer that the maximum speed that Neusos can tolerate is $\frac{\Delta f v_s}{f_0 + kt} = 4.3\ m/s$, where v_s is the sound speed in water ($1500\ m/s$). Therefore, the general smartwatch movement can hardly distort the linear chirp pattern. An interesting finding is that the detection recall when the user swims (96.6%) is slightly lower than hand waving. We found that this is because the arm of the swimmer will surface out of the water for a short period, leading to fragment chirp patterns in the spectrogram.

F. Impact of Interference

We evaluated the impact of interference from other nearby swimmers on the system recall. In this setting, the user wearing the smartwatch is $10m$ away from the receiver, while a swimmer (not wearing the smartwatch) as the interferer swims and splashes water at a distance varying from $1m$ to $5m$. As shown in Fig. 16, Neusos achieves 100% detection recall when the distance between the interferer and the receiver is $5m$, indicating that the impact of far interference was negligible. When this interferer moves to $1m$, the detection recall reduces to 98.9% due to strong splashing interference. This result indicates that Neusos can still detect SOS signals under splash interference.

We also evaluated the impact of interference on system specificity. In this experiment, the smartwatch does not transmit signals, while the interferer keeps splashing. Figure 16 shows that Neusos achieves a false positive rate of 99.01% when the distance between the interferer and the receiver is $5m$. As the distance decreases to 1m, the false positive rate decreases to 98.6%. One reason is that intense splash can produce high-energy noise and another is that we trade the specificity for a higher detection recall, because it is of utmost importance to save a life. We believe occasional false alarms are acceptable in practice. In real-world scenarios, users can adjust the misclassification cost to find a proper OOP.

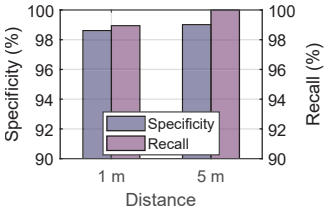


Fig. 16: Interference impact.

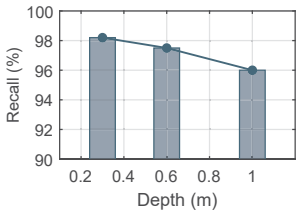


Fig. 17: Depth impact.

G. Impact of Depth

We evaluated Neusos at different depths with 0.3 m to 1 m in the water. The swimmer wears the transmitter at $15m$ away from the receiver. Figure 17 shows performance under different depths. Neusos can achieve a detection recall of 100% at a depth of $0.3m$. This value decreases to 98.2% as the depth increases to $0.6m$. When the depth continues to increase to $1m$, the detection recall drops to 96%. One reason may be that the speakers of commercial smartwatches are not designed for use in the water and have low transmission power. Another reason is that the water pressure becomes higher as the depth increases. Consequently, the vibration of the speaker membrane will be confined, leading to a much lower SOS signal strength. This result suggests that Neusos can effectively detect SOS signals at the early stages of drowning. As the depth increases, Neusos's SOS detection ability will slightly decrease. To mitigate this problem, we can deploy a few more hydrophones at the bottom of the swimming pool to reduce the communication range and achieve reliable SOS detection.

V. Related work

A. Drowning Monitoring

Vision-based drowning detection has been extensively explored in both industry [7, 8, 27] and academia [9, 28, 29]. Many of these methods rely on identifying if a person remains stationary in the water for an extended period to detect potential drowning incidents. Before such detection methods recognize an incident, the drowning person could have been suffocated for a long time. Some machine learning-based approaches [30–32] extract motion features, such as speed and posture to build supervised classifiers and recognize drowning events. Nonetheless, the complex nature of drowning behavior and the scarcity of training samples pose difficulties in achieving high detection performance [3]. Additionally, vision-based methods are affected by various factors such as poor lighting conditions, highly dynamic backgrounds, and limited visibility of targets in the water [10].

Wearable sensor-based approaches [33–35] offer greater resilience to lighting conditions. Some works [12–14, 33, 36] monitor sensor readings related to human motion or vital signs (*e.g.*, oxygen level, respiration) to detect potential drowning events and activate an airbag for safety. Another approach involves utilizing water pressure sensors [37, 38] or RFID technology [39] to monitor the duration a swimmer remains underwater, triggering an alert if it surpasses a predefined threshold. While these systems can detect the risk of drowning, they typically require swimmers to wear bulky life kits [13, 36] equipped with powerful acoustic transmitters, a setup that isn't practical or viable for the generic swimmer.

Neusos can offer an API that integrates with existing wearable-based drowning detection systems, providing them with a robust underwater communication channel to transmit SOS signals. In this case, Neusis can serve as a complementary solution to vision-based methods and facilitate early-stage rescue operations. Moreover, Neusos empowers swimmers to actively call for help in case of emergencies, such as severe discomfort. In this research, we focus on SOS transmission and detection, leaving the integration with wearable-based drowning detection for future work.

B. Underwater Acoustic Communication

Acoustic signals have been widely used for human-device interaction [40, 41] and localization [42–44] in the air. In fact, underwater is where sound signals hold the greatest advantage and find widespread application. In recent years, many researchers aim to enable underwater acoustic communication using commercial devices [11, 23, 45]. AquaApp [11], for example, has designed an underwater messaging system that utilizes commercial smartphones, providing a communication range of up to $100m$. However, AquaApp replies on traditional cross-correlation to detect signal preambles, which requires smartphones with strong transmission power, much higher than that of commercial smartwatches. AquaHelper [23] achieves reliable underwater SOS transmission and detection with wearable devices. However, AquaHelper requires multiple chirp aggregation in the time domain to enhance the SNR and detect weak SOS signals reliably. As a result, this process may result in a longer duration for SOS detection in low-SNR scenarios before initiating rescue operations. In contrast, Neusos addresses these challenges by efficiently training and deploying a signal-characteristic-aware deep learning-based detector. This innovation enables reliable underwater SOS detection using lightweight smartwatches within a single detection window.

VI. Conclusion

Drowning incidents pose a critical public safety concern worldwide. Many drowning swimmers have no effective way to ask for help and get drowned, even with lifeguards present. To remedy this life-threatening issue, we introduce Neusos, an underwater SOS system with commercial smartwatches and low-cost hydrophones. The unique property of the underwater environment poses significant challenges in detecting SOS signals with the traditional cross-correlation method. We carefully observe this property and propose an explicit deep-learning based detection method to achieve reliable underwater SOS detection. Considering the increasing popularity of waterproof smartwatches, we envision that this life-saving functionality can potentially become indispensable for swimmers and swimming pools in the future.

Acknowledgments

This work is supported by the Hong Kong GRF under grant 15206123. Yuanqing Zheng is the corresponding author.

REFERENCES

[1] "Swimming injury statistics - swimming pool accidents," https://www.edgarsnyder.com/statistics/swimming-pool-statistics.html, 2022, (Accessed on 08/20/2022).

[2] D. C. Schwebel, H. N. Jones, E. Holder, and F. Marciani, "Lifeguards: A forgotten aspect of drowning prevention," *Journal of injury and violence research*, vol. 2, no. 1, p. 1, 2010.

[3] A. Carballo-Fazanes, J. J. Bierens, and I. E. G. to Study Drowning Behaviour, "The visible behaviour of drowning persons: A pilot observational study using analytic software and a nominal group technique," *International journal of environmental research and public health*, vol. 17, no. 18, p. 6930, 2020.

[4] "Facts & stats about drowning - stop drowning now," https://www.stopdrowningnow.org/drowning-statistics/, 2022, (Accessed on 08/20/2022).

[5] "A short history of the 10:20 protection standard," https://www.thewirh.com/blog/10-20, 2023, (Accessed on 03/17/2023).

[6] J. Ellis, "Lifegaurds may look but they don't always see: A study points to causes contributing to more than 400 deaths annually in u.s. public swimming facilities," https://www.lib.niu.edu/2002/ip020538.html, 2002, (Accessed on 08/21/2022).

[7] "Swimeye - a drowning detection system for swimming pools," https://swimeye.com/, 2022, (Accessed on 08/21/2022).

[8] "Poseidon - drowning detection system for swimming pools," https://poseidon-tech.com/en-GB/technology-2/, 2022, (Accessed on 08/21/2022).

[9] C. Zhang, X. Li, and F. Lei, "A novel camera-based drowning detection algorithm," in *Chinese Conference on Image and Graphics Technologies*. Springer, 2015, pp. 224–233.

[10] H.-L. Eng, K.-A. Toh, A. H. Kam, J. Wang, and W.-Y. Yau, "An automatic drowning detection surveillance system for challenging outdoor pool environments," in *Computer Vision, IEEE International Conference on*, vol. 2. IEEE Computer Society, 2003, pp. 532–532.

[11] T. Chen, J. Chan, and S. Gollakota, "Underwater messaging using mobile devices," in *Proceedings of the ACM SIGCOMM 2022 Conference*, 2022, pp. 545–559.

[12] M. Kharrat, Y. Wakuda, N. Koshizuka, and K. Sakamura, "Automatic waist airbag drowning prevention system based on underwater time-lapse and motion information measured by smartphone's pressure sensor and accelerometer," in *2013 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2013, pp. 270–273.

[13] J. Wen, D. Zhou, H. Feng, Y. Wang, X. Geng, H. Ma, and Z. Yang, "Lifetag: Vital sign detection for drowning people in sea accidents by wearable device," in *Proceedings of the 2019 8th International Conference on Networks, Communication and Computing*, 2019, pp. 57–64.

[14] A. Kulkarni, K. Lakhani, and S. Lokhande, "A sensor based low cost drowning detection system for human life safety," in *2016 5th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*. IEEE, 2016, pp. 301–306.

[15] F. Steinmetz, J. Heitmann, and C. Renner, "A practical guide to chirp spread spectrum for acoustic underwater communication in shallow waters," in *Proceedings of the 13th International Conference on Underwater Networks & Systems*, 2018, pp. 1–8.

[16] M. I. Skolnik, "Introduction to radar systems," *New York*, 1980.

[17] N. Shankari, "A review on different weak signal detection technologies."

[18] E. Demirors and T. Melodia, "Chirp-based lpd/lpi underwater acoustic communications with code-time-frequency multidimensional spreading," in *Proceedings of the 11th International Conference on Underwater Networks & Systems*, 2016, pp. 1–6.

[19] W. Lei, D. Wang, Y. Xie, B. Chen, X. Hu, and H. Chen, "Implementation of a high reliable chirp underwater acoustic modem," in *2012 Oceans-Yeosu*. IEEE, 2012, pp. 1–5.

[20] K. G. Kebkal and R. Bannasch, "Sweep-spread carrier for underwater communication over acoustic channels with strong multipath propagation," *The Journal of the Acoustical Society of America*, vol. 112, no. 5, pp. 2043–2052, 2002.

[21] M. Stojanovic and J. Preisig, "Underwater acoustic communication channels: Propagation models and statistical characterization," *IEEE communications magazine*, vol. 47, no. 1, pp. 84–89, 2009.

[22] N. Hou, X. Xia, and Y. Zheng, "Don't miss weak packets: Boosting lora reception with antenna diversities," *ACM Transactions on Sensor Networks*, vol. 19, no. 2, pp. 1–25, 2023.

[23] Q. Yang and Y. Zheng, "Aquahelper: Underwater sos transmission and detection in swimming pools," in *Proceedings of the 21th ACM Conference on Embedded Networked Sensor Systems*, ser. SenSys '23. Association for Computing Machinery, 2023.

[24] Z. Xu, S. Tong, P. Xie, and J. Wang, "From demodulation to decoding: Toward complete lora phy understanding and implementation," *ACM Transactions on Sensor Networks*, vol. 18, no. 4, pp. 1–27, 2023.

[25] J. S. Lim, "Two-dimensional signal and image processing," *Englewood Cliffs*, 1990.

[26] T. Fawcett, "Roc graphs: Notes and practical considerations for researchers," *Machine learning*, vol. 31, no. 1, pp. 1–38, 2004.

[27] "About coral manta - coral drowning detection systems," https://coraldrowningdetection.com/about-coral-manta/, 2022, (Accessed on 08/21/2022).

[28] X. He, F. Yuan, and Y. Zhu, "Drowning detection based on video anomaly detection," in *Image and Graphics: 11th International Conference, ICIG 2021, Haikou, China, August 6–8, 2021, Proceedings, Part III 11*. Springer, 2021, pp. 700–711.

[29] F. Lei, H. Zhu, F. Tang, and X. Wang, "Drowning behavior detection in swimming pool based on deep learning," *Signal, Image and Video Processing*, pp. 1–8, 2022.

[30] S. Hasan, J. Joy, F. Ahsan, H. Khambaty, M. Agarwal, and J. Mounsef, "A water behavior dataset for an image-based drowning solution," in *2021 IEEE Green Energy and Smart Systems Conference (IGESSC)*. IEEE, 2021, pp. 1–5.

[31] H.-L. Eng, K.-A. Toh, W.-Y. Yau, and J. Wang, "Dews: A live visual surveillance system for early drowning detection at pool," *IEEE transactions on circuits and systems for video technology*, vol. 18, no. 2, pp. 196–210, 2008.

[32] L. He, H. Hou, Z. Yan, and G. Xing, "Demo abstract: An underwater sonar-based drowning detection system," in *2022 21st ACM/IEEE International Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 2022, pp. 493–494.

[33] B. K. Nuhu, B. U. Umar, A. K. Abu, G. Blessed, and K. Jinsul, "Development of a smart wearable antidrowning system for swimmers," 2022.

[34] J. G. Ramani, J. Gayathri, R. Aswanth, and M. Gunasekaran, "Automatic prevention of drowning by inflatable wrist band system," in *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*. IEEE, 2019, pp. 346–349.

[35] M. Ramdhan, M. Ali, S. Ali, M. Kamaludin *et al.*, "An early drowning detection system for internet of things (iot) applications," *TELKOMNIKA (Telecommunication Computing Electronics and Control)*, vol. 16, no. 4, pp. 1870–1876, 2018.

[36] S. Winkler, "Development of a mobile device to prevent accidental drowning by monitoring respiratory activity," in *Proceedings of the MCI Medical Technologies Master's Conference*. MCI the Entrepreneurial University Group, 2022, pp. e8–e8.

[37] H. Liu, M. B. H. Frej, and B. Wen, "A novel method for recognition, localization, and alarming to prevent swimmers from drowning," in *2019 IEEE Cloud Summit*. IEEE, 2019, pp. 65–71.

[38] A. Roy and K. Srinivasan, "A novel drowning detection method for safety of swimmers," in *2018 20th National Power Systems Conference (NPSC)*. IEEE, 2018, pp. 1–6.

[39] F. Dehbashi, N. Ahmed, M. Mehra, J. Wang, and O. Abari, "Swimtrack: drowning detection using rfid," in *Proceedings of the ACM SIGCOMM 2019 conference posters and demos*, 2019, pp. 161–162.

[40] C. Cai, R. Zheng, and J. Luo, "Ubiquitous acoustic sensing on commodity iot devices: A survey," *IEEE Communications Surveys & Tutorials*, vol. 24, no. 1, pp. 432–454, 2022.

[41] K. Wu, Q. Yang, B. Yuan, Y. Zou, R. Ruby, and M. Li, "Echowrite: An acoustic-based finger input system without training," *IEEE Transactions on Mobile Computing*, vol. 20, no. 5, pp. 1789–1803, 2020.

[42] Y. He, W. Wang, L. Mottola, S. Li, Y. Sun, J. Li, H. Jing, T. Wang, and Y. Wang, "Acoustic localization system for precise drone landing," *IEEE Transactions on Mobile Computing*, 2023.

[43] C. Cai, H. Pu, P. Wang, Z. Chen, and J. Luo, "We hear your pace: Passive acoustic localization of multiple walking persons," *Proceedings of the ACM IMWUT*, vol. 5, no. 2, pp. 1–24, 2021.

[44] T. Zheng, C. Cai, Z. Chen, and J. Luo, "Sound of motion: Real-time wrist tracking with a smart watch-phone pair," in *IEEE INFOCOM 2022-IEEE Conference on Computer Communications*. IEEE, 2022, pp. 110–119.

[45] T. Chen, J. Chan, and S. Gollakota, "Underwater 3d positioning on smart devices," in *Proceedings of the ACM SIGCOMM 2023 Conference*, 2023, pp. 33–48.